Search for gluino production in final states with an isolated lepton and b-tagged jets using 13 TeV collisions at CMS

Dissertation

zur Erlangung des Doktorgrades

des Fachbereichs Physik

der Universität Hamburg

vorgelegt von

Artur Lobanov

aus St. Petersburg

CERN-THESIS-2016-125



Hamburg

2016

Gutachter der Dissertation:	Dr. Isabell-Alissandra Melzer-Pellmann Prof. Dr. Erika Garutti
Gutachter der Disputation:	Prof. Dr. Dieter Horns Prof. Dr. Erika Garutti Prof. Dr. Johannes Haller Dr. Isabell-Alissandra Melzer-Pellmann Prof. Dr. Gudrid Moortgat-Pick
Datum der Disputation:	11. Juli 2016

Abstract

A search for supersymmetric particles in proton-proton collisions at a centre-of-mass energy of 13 TeV is presented in this thesis. The data were recorded by the CMS experiment during *Run 2* of the LHC, corresponding to an integrated luminosity of 2.3 fb^{-1}. The signal signature comprises pair produced gluon superpartners, gluinos, decaying each to a top-antitop quark pair and a neutralino – the lightest supersymmetric particle. Since the probability, that only one W boson from the four initial top-quarks decays to a lepton-neutrino pair, is about 40%, events with a single isolated lepton, electron or muon, accompanied by several high-energy hadronic jets and at least one b-quark tagged jet are selected.

Signal-enriched search regions are defined based on the azimuthal angle between the lepton and four-vector sum of the missing energy and lepton. The dominant Standard Model backgrounds tend towards lower angles, while the investigated signals are expected to show little dependence on the angle. Further sensitivity to different signal mass scenarios is increased by considering events in categories with different jet and b-tag multiplicities, hadronic and leptonic scales, which are defined as the scalar sums of the jet transverse momenta, and the missing energy and lepton momentum, respectively.

The background is estimated by a data-driven approach using background-enriched control regions. Low jet multiplicity data sidebands are used to obtain respective signal-to-control region transfer factors. Contributions from multijet events, which arise from misidentified leptons, are subtracted from these control regions using misidentification probabilities measured in data.

No significant deviation from the predicted Standard Model background is observed. This result is interpreted in the framework of simplified models corresponding to the aforementioned signal signature, resulting in exclusion limits for particular gluino-neutralino mass scenarios. Gluinos below 1575 GeV are excluded for light neutralinos, while for scenarios with low neutralino-gluino mass splitting, neutralinos are excluded up to 850 GeV. This extends the limits obtained in *Run 1* by about 200 and 250 GeV, respectively.

In addition, the upgrade of the CMS outer hadron calorimeter readout system with silicon photomultipliers is presented, describing its motivation and the commissioning of the new hardware. An initial energy calibration with cosmic muons is performed afterwards.

Kurzfassung

In dieser Arbeit wird eine Suche nach supersymmetrischen Teilchen in Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 13 TeV vorgestellt. Die Daten wurden vom CMS-Experiment während der zweiten Betriebsperiode des LHC aufgenommen und entsprechen einer integrierten Luminosität von 2.3 fb^{-1}. Die untersuchte Signatur ist die Paarproduktion von Gluinos, der Superpartner des Gluons, wobei diese jeweils in ein Top-Antitop-Quark-Paar zerfallen, sowie in ein Neutralino, das leichteste supersymmetrische Teilchen.

Da die Wahrscheinlichkeit, dass nur ein W-Boson aus dem Zerfall der vier Top-Quarks in ein Lepton-Neutrino-Paar zerfällt, bei rund 40% liegt, werden Ereignisse mit einem isolierten Lepton (Elektron oder Myon) selektiert, das von mehreren hadronischen Jets begleitet ist. Mindestens ein Jet muss als b-Quark-Jet identifiziert sein.

Die Suchbereiche, in denen ein höherer Signalanteil erwartet wird, werden mit Hilfe des Azimuthwinkels zwischen dem Lepton und der Summe aus Leptonimpuls und dem fehlenden Transversalimpuls definiert. Der Untergrund durch Prozesse des Standardmodells liegt typischerweise bei kleinen Winkeln, während die untersuchten supersymmetrischen Signale gleichmässig über dem Winkel verteilt sind.

Für verschiedene Szenarien bezüglich der Masse der gesuchten supersymmetrischen Teilchen kann die Empfindlichkeit durch Definition von mehreren Suchkategorien erhöht werden. Dazu werden die Jet- und b-Quark-Jetmultiplizität benutzt, sowie verschiedene Variablen, die ein Maß für die leptonische und hadronische Aktivität im Ereignis sind. Verwendet werden hierzu die skalare Summe der Jet-Transversalimpulse, sowie die Summe aus fehlendem Transversalimpuls und dem Transversalimpuls des Leptons.

Der zu erwartende Standardmodelluntergrund wird aus den Daten abgeschätzt. Hierzu werden Kontrollregionen definiert, in denen Untergrundprozesse überwiegen und Seitenbänder mit wenigen Jets werden verwendet, um die Transferfaktoren zu bestimmen, die das Verhältnis von Signal- zu Kontrollregion beschreiben. Dabei werden Beiträge aus Multijetereignissen, die durch falsch identifizierte Leptonen entstehen, berücksichtigt.

Es wird keine signifikante Abweichung von dem abgeschätzten Standardmodelluntergrund beobachtet. Dieses Ergebnis wird im Rahmen von sogenannten vereinfachten Modellen, entsprechend der oben genannten Signalsignatur, interpretiert, was in Ausschlussgrenzen für bestimmte Gluino-Neutralino-Massenszenarien resultiert. Für leichte Neutralinos werden Gluinos unter 1575 GeV ausgeschlossen, während für Szenarien mit niedriger Neutralino-Gluino-Massendifferenz Neutralinos bis 850 GeV ausgeschlossen werden. Dies erweitert die in der ersten LHC Betriebsperiode erhaltenen Ausschlussgrenzen um etwa 200 und 250 GeV.

Darüber hinaus wird das Upgrade des Auslesesystems des äußeren CMS-Hadronkalorimeter mit Silizium-Photomultipliern vorgestellt. Die entsprechende Motivation sowie die Inbetriebnahme der neuen Hardware werden beschrieben und die erste Energiekalibrierung mit kosmischen Myonen wird durchgeführt.

Contents

Abbreviations vii

Introduction 1

1 Supersymmetry: an extension of the Standard Model of Particle Physics 5
1.1 Current status of the Standard Model . . . 6
1.1.1 Particles of the Standard Model . . . 6
1.1.2 The Standard Model as a quantum field theory . . . 7
1.2 Unsolved issues of particle physics . . . 9
1.2.1 Experimental puzzles . . . 9
1.2.2 Theoretical questions . . . 10
1.3 Supersymmetry . . . 11
1.3.1 Closing the hierarchy problem . . . 11
1.3.2 Minimal Supersymmetric Standard Model . . . 12
1.3.3 Other extensions beyond the Standard Model . . . 14
1.4 Searches for SUSY at colliders . . . 15
1.4.1 Simplified models . . . 15
1.4.2 Status of SUSY searches after *Run 1* . . . 16
1.4.3 Outlook for *Run 2* . . . 19

2 Experimental Setup 21
2.1 The Large Hadron Collider . . . 21
2.1.1 The CERN accelerator complex . . . 24
2.1.2 Outlook towards the future of the LHC . . . 25
2.1.3 LHC experiments . . . 26
2.2 The Compact Muon Solenoid experiment . . . 29
2.2.1 Magnet . . . 31
2.2.2 Inner Tracker . . . 31
2.2.3 Electromagnetic Calorimeter . . . 33
2.2.4 Hadron Calorimeter . . . 34
2.2.5 Muon System . . . 38
2.2.6 Trigger and Data Acquisition Systems . . . 40
2.2.7 Luminosity measurement . . . 41
2.2.8 Upgrades of CMS . . . 42
2.3 Event simulation . . . 43
2.3.1 Event generators . . . 43
2.3.2 Detector simulation . . . 44

3 Upgrade of the CMS outer hadron calorimeter 47
3.1 The HO upgrade . . . 47
3.1.1 HO design . . . 48
3.1.2 Silicon photomultipliers . . . 50

3.2 Commissioning . . . 51
3.3 Calibration with cosmic muons . . . 53
3.3.1 Local runs . . . 53
3.3.2 Global runs . . . 59
3.3.3 Muon detection efficiency . . . 63

4 Object reconstruction and identification **65**
4.1 Data flow at CMS . . . 65
4.2 Particle-Flow algorithm . . . 66
4.3 Physics objects . . . 69
4.3.1 Primary vertices . . . 70
4.3.2 Leptons . . . 70
4.3.3 Jets . . . 74
4.3.4 Missing transverse energy . . . 78

5 Event Selection **79**
5.1 Search signature . . . 79
5.2 Data samples . . . 80
5.2.1 Event cleaning . . . 81
5.2.2 Trigger . . . 82
5.3 Simulated samples . . . 84
5.3.1 Background samples . . . 85
5.3.2 Signal samples . . . 85
5.3.3 Scale factors . . . 87
5.4 Baseline selection . . . 88
5.4.1 Control distributions . . . 89
5.5 Search regions . . . 91

6 Background Estimation **97**
6.1 Background composition . . . 98
6.2 R_{CS} method . . . 98
6.2.1 R_{CS} composition . . . 101
6.2.2 R_{CS} in $t\bar{t}$ events . . . 101
6.2.3 R_{CS} in different kinematic regions . . . 106
6.3 QCD background estimation . . . 112
6.3.1 QCD estimation in the electron channel . . . 114
6.3.2 QCD estimation in the muon channel . . . 124
6.4 Systematic uncertainties . . . 128
6.4.1 Uncertainties on background estimation . . . 128
6.4.2 Uncertainties on background and signal prediction . . . 130
6.4.3 Uncertainties on signal prediction . . . 133
6.5 Prediction in data . . . 136

7 Results and Interpretation **139**
7.1 Results . . . 139
7.2 Limit setting procedure . . . 143
7.2.1 The CL_s method for statistical hypothesis testing . . . 143

7.2.2 Statistical interpretation of results . . . 145
7.3 Interpretation . . . 148
7.3.1 T1tttt cross section upper limit . . . 148
7.3.2 Projections for *Run 2* . . . 150
7.4 Comparison to other *Run 1* and *Run 2* results . . . 151

Conclusions **153**

A Backup trigger strategy **157**

B Lepton isolation **161**

C Additional control plots after baseline selection **165**

D Additional plots for the QCD determination **167**
D.1 L_{p} and $\Delta\varphi$ shape comparison . . . 167
D.2 L_{p} template fit in simulation . . . 169
D.3 Trigger efficiency plots for anti-selected leptons . . . 170

E B-tagging scale factor methods **173**

Bibliography **175**

Abbreviations

ADC: analogue-to-digital converter

AGR: April Global Run

ALICE: A Large Ion Collider Experiment

APD: avalanche photodiode

ASIC: application-specific integrated circuit

ATLAS: A Toroidal LHC Apparatus

BCM1f: Fast Beam Condtions Monitore

BSM: beyond the Standard Model

CASTOR: Centauro And Strange Object Research Calorimeter

CCM: clock and control module

CERN: European Organization for Nuclear Research – *Conseil Européen pour la Recherche Nucléaire* –

C.L.: confidence level

CMB: Cosmic Microwave Background

CMS: Compact Muon Solenoid

cMSSM: constrained MSSM

CR: control region

CSC: cathod strip chamber

CSV: Combined Secondary Vertex

CU: calibration unit

DAQ: Data Acquisition

DCS: Detector Control System

DQM: Data Quality Monitoring

DT: drift tube

EB: ECAL barrel

ECAL: electromagnetic calorimeter

ECR: Extended Cosmis Run

EE: ECAL endcap

ES: ECAL preshower

EWSB: electroweak symmetry breaking

FEE: front-end electronics

FPGA: field-programmable gate array

FSR: final-state radiation

GEANT4: GEometry ANd Tracking

GSF: Gaussian-Sum Filter

GUT: Grand Unified Theory

HB: HCAL barrel

HCAL: hadron calorimeter

HE: HCAL endcap

HF: forward calorimeter

HL-LHC: High-Luminosity LHC

HLT: High-Level Trigger

HO: outer hadron calorimeter

HPD: hybrid photodiode

IP: interaction point

ISR: initial-state radiation

IVF: Inclusive Vertex Finder

JEC: Jet Energy Calibration

JES: Jet Energy Scale

L1A: Level-1 accept
L1 Trigger: Level-1 Trigger
LEP: Large Electron-Positron Collider
LHC: Large Hadron Collider
LHCf: Large Hadron Collider forward
LO: leading order
LS1: first Long Shutdown
LS2: second Long Shutdown
LSP: lightest supersymmetric particle
LUT: look up table

MB: mainband
MC: Monte Carlo
ME: matrix element
MIP: minimum ionizing particle
MoEDAL: Monopole & Exotics Detector At the LHC
MPPC: Multi-Pixel Photon Counter
MPV: most probable value
MSSM: Minimal Supersymmetric extension of the Standard Model
MWGR: Midweek Global Run

NbTi: Niobium-Titanium
NLO: next-to-leading order

ODU: optical decoding unit

PbPb: lead-lead
$PbWO_4$: lead tungstate
PCC: Pixel Cluster Counting
PD: Primary Dataset
PDF: parton distribution function
PF: Particle-Flow Algorithm
PLT: Pixel Luminosity Telescope
pMSSM: phenomenological MSSM
pp: proton-proton
PS: Proton Synchrotron
PSB: PS Booster
PV: primary vertex

QCD: quantum chromodynamics
QED: quantum electrodynamics
QIE8: charge integration and encoding chip
QPS: Quench Protection System

RBX: readout box
RICH: Ring-imaging Cherenkov detector
RM: readout module
RPC: resistive plate chamber

SB: sideband
SiPM: silicon photomultiplier
SL: SuperLayer
SM: Standard Model of particle physics
SMS: simplified model spectra
SPE: single photo-electron
SPS: Super Proton Synchrotron
SR: search region
SUSY: Supersymmetry
SV: secondary vertex

TEC: Tracker Endcap
TIB: Tracker Inner Barrel
TID: Tracker Inner Disk
TOB: Tracker Outer Barrel
TOTEM: TOtal Elastic and diffractive cross section Measurement
TriDAS: Trigger and Data Acquisition Systems
TRT: Transition Radiation Tracker
TS: time slice
TTP: Technical Trigger Processor

VELO: Vertex Locator
VPT: vacuum phototriode

WLS: wavelength shifting fibre
WP: working-point

YB: barrel return yoke
YE: endcap return yoke

ZDC: Zero-Degree Calorimeter

Introduction

"To boldly go where no man has gone before . . . "

— James T. Kirk

Symmetry and compositeness – these are two keywords that can be used to describe Nature. Ever since humankind started to exist, symmetries in the shapes of natural phenomena and organisms are reflected and repeated in man-made objects. The observation of Nature also lead to the idea of object compositeness, i.e. the property of being composed of smaller entities, which manifested itself in ancient traditions of *atomism*. Both ideas became important concepts in ancient natural philosophy.

However, as the name συμμετρία suggests, the perception of the measure (μέτρον) of symmetry is quite subjective. The idea of indivisible ἄτομον objects also depends on the perspective. Modern science provides more objective and robust definitions of these terms based on deductive and experimental methods.

Symmetries are understood as invariance of objects with respect to transformations, which results in conservation of corresponding quantities according to Noether's theorem [1, 2]. Together with the successive experimental discovery of atoms, nuclei and elementary particles this has lead to the creation of particle physics.

The Standard Model of particle physics (SM) is a theoretical framework that incorporates the knowledge about fundamental particles and their interactions accumulated throughout the last century. It describes the particles that form matter – the three generations of leptons and quarks – and the fields which serve as mediators for the respective interactions – the gauge bosons – governed by a local gauge symmetry and global Poincaré symmetry [3]. This constructions allows the Standard Model of particle physics to describe three of four fundamental interactions – the electromagnetic, weak and strong forces – in a wide range of energies for various physics phenomena.

Throughout the last decades, the SM has proven to be a very precise and robust theory, almost in complete agreement with Nature. An apogee in testing the SM has been achieved with the Large Hadron Collider (LHC) – the most powerful collider up to date – which is designed to collide protons at a centre-of-mass energy ($\sqrt{s}$) of 14 teraelectronvolt (TeV). After the recent discovery of the Higgs boson [4,5] by the ATLAS and CMS experiments using LHC data with $\sqrt{s} = 8\,\text{TeV}$, the Standard Model particle content seems to be finalized.

Nevertheless, the mission of particle physics has not yet arrived at its end. Gravity poses a large discomfort to the SM, which tries to be an universal theory but seems ignorant to this fourth force. Moreover, observations at cosmological scales have shown, that the ordinary matter, described by the SM, accounts only for 5% of the whole energy budget in the Universe. Together with several aesthetic problems, these are strong indications, that the SM cannot be the final theory, but rather has to be part of some more universal framework.

At this point the idea of symmetries could again help to improve the understanding of Nature. Supersymmetry (SUSY) represents a hypothesized extension of the SM, which could circumvent several of its shortcomings by introducing an additional space-time symmetry [6]. In general, this theory attributes to each SM particle a superpartner with exactly the same quantum numbers except spin, which differs by 1/2. The new particle and field content of SUSY results in a rich phenomenological background for searches at such machines as the LHC. Unfortunately, no evidence for such particles has been found up to now [7, 8].

It is possible, that supersymmetric particles are much heavier or produced in more complex states than expected. The second run of the LHC at an unseen centre-of-mass energy of 13 TeV should help reveal any physics beyond the Standard Model. Along the works for the enhancement of the accelerator performance, the LHC experiments also underwent a vast majority of improvements and consolidations resulting in an increased efficiency and sensitivity.

This thesis presents a search for Supersymmetry targeting models where pair produced gluon superpartners, gluinos $\widetilde{g}$, decay each to a top-antitop quark pair $t\bar{t}$ and neutralino $\widetilde{\chi}_1^0$ – the lightest supersymmetric particle, and one of the top-quarks further decays leptonically via a W boson. Therefore, the event selection requires final states with a single isolated lepton, electron or muon, accompanied by several high-energy hadronic jets and at least one b-quark tagged jet. Data used for this analysis were recorded by the CMS experiment during *Run 2* of the LHC in 2015, corresponding to an integrated luminosity of $2.3\,\text{fb}^{-1}$.

The Standard Model background is dominated by $t\bar{t}$+jets and W+jets events, where the isolated lepton is originating from a leptonic W-boson decay, and therefore aligned with the mother particle. The W-decay neutrino creates an energy imbalance in the detector, which together with the lepton corresponds to the W-boson. In case of supersymmetric gluino production, additional momentum imbalance is created by the neutralinos escaping detection, which cancel the angular relation of the lepton and reconstructed W boson. This feature allows to define signal-enriched regions with suppressed background expectation. Further sensitivity to different signal mass scenarios is increased by considering events in categories with different jet and b-tag multiplicities, hadronic and leptonic scales, which are defined as the scalar sums of the jet transverse momenta, and the missing energy and lepton momentum, respectively. Similar searches have been performed on 7 and 8 TeV data [9–11] using single search categories.

The background is estimated by a data-driven approach from background-enriched control regions. Data sidebands with low jet multiplicity are used to obtain respective signal-to-control-region transfer factors. Contributions from multijet events, which arise from misidentified leptons, are subtracted from these control regions using misidentification probabilities measured in data. The results are interpreted in the framework of simplified models corresponding to the aforementioned signal model of gluino production.

This thesis is organized as follows:

- Chapter 1 briefly introduces the Standard Model of particle physics (SM). Some experimentally and theoretically motivated shortcomings are discussed further, providing a motivation for an extension beyond the Standard Model (BSM) with Supersymmetry (SUSY) being one of the most appealing of such theories. A short review of searches for SUSY at colliders is given.
- Chapter 2 describes the experimental setup used to obtain data for this analysis. A brief overview of the LHC is followed by a more elaborate description of the CMS experiment. Furthermore, simulation programmes which are used to complement and verify the experimental data are reviewed.
- The upgrade of the CMS outer hadron calorimeter is presented in Chapter 3, describing the motivation, procedure and commissioning of the new hardware. The first energy calibration of the detector using cosmic muons is finalizing this chapter.
- Chapter 4 reviews the reconstruction and identification of physics objects and respective performance in the CMS experiment.
- The data and simulated samples are described in Chapter 5. Motivation is provided for the event selection based on the signal signature. Search regions dominated by hypothetical signals are defined.
- Chapter 6 presents the data-driven estimation of the SM background. The R_{CS} method allows to obtain signal region predictions based on the complementary control regions using respective transfer factors measured in data sidebands. Contributions from multijet events are subtracted from the CRs.
- The predicted SM background event yields are compared to actual data in Chapter 7. These results are interpreted within the framework of simplified models using a frequentist hypothesis testing method. Upper limits on signal cross sections are obtained allowing to extend the *Run 1* exclusions on the gluino and neutralino masses by up to 200 and 250 GeV, respectively.
- The final Chapter presents a summary of this thesis as well as an outlook on the prospects of searches for BSM physics.

Chapter 1

Supersymmetry: an extension of the Standard Model of Particle Physics

1.1 Current status of the Standard Model . . . 6
1.1.1 Particles of the Standard Model . . . 6
1.1.2 The Standard Model as a quantum field theory . . . 7
1.2 Unsolved issues of particle physics . . . 9
1.2.1 Experimental puzzles . . . 9
1.2.2 Theoretical questions . . . 10
1.3 Supersymmetry . . . 11
1.3.1 Closing the hierarchy problem . . . 11
1.3.2 Minimal Supersymmetric Standard Model . . . 12
1.3.3 Other extensions beyond the Standard Model . . . 14
1.4 Searches for SUSY at colliders . . . 15
1.4.1 Simplified models . . . 15
1.4.2 Status of SUSY searches after *Run 1* . . . 16
1.4.3 Outlook for *Run 2* . . . 19

The Standard Model of particle physics (SM) is nowadays the established foundation of modern particle physics and one of the most elaborated and precise theories of fundamental physics. It incorporates knowledge accumulated over almost one century, successfully predicting many observed phenomena at the particle level. This chapter will commence with a short review of the SM, listing its main theoretical and experimental aspects in Sec. 1.1.

Despite its huge success, the SM is unable to answer several major physics problems, which are described in Sec. 1.2. Different extensions BSM provide solutions for some of these problems, with one of them – SUSY – being one of the most promising candidates. Section 1.3 will give a general introduction to SUSY and present the simplest possible supersymmetric extension of the SM.

This chapter is finalised with Section 1.4, reviewing the searches for SUSY at colliders and presenting the framework of simplified models within which the analysis presented in this thesis is conducting the search for new physics BSM.

A variety of sources has been used for the preparation of this chapter, including textbooks and reviews [12–15]. The most recent experimental results are obtained from Ref. [15].

1.1 Current status of the Standard Model

In the middle of the XIXth Century a ground-braking step in the understanding of fundamental forces was taken: the formerly separately treated electric and magnetic forces were elegantly combined by a single theory of electromagnetism, ultimately described by Maxwell's equations. It was not until the middle of the next, the XXth Century that, following a similar evolution of thought, the Standard Model of particle physics (SM) was created merging three of the four currently known fundamental forces – the electromagnetic, the weak and the strong forces – into a unique framework. Since then the SM has proven to be a very effective and accurate tool to describe the phenomena and observations of the microcosm of particle physics in a very wide range of energies.

The only exception is the gravitational interaction of matter, which fails to be included in the SM and therefore is one of the biggest open questions of modern physics. However, given its very weakness, this force has no impact on the interaction of single particles.

In analogy with the wave-particle duality, the SM can be approached from two sides: on one hand are the particles, which can be observed in experiments, and on the other there is the quantum field theory dictating their interactions and properties.

1.1.1 Particles of the Standard Model

The SM comprises three groups of fundamental particles: the fermions with spin one-half form the matter, interactions of those fermions are mediated by spin-one bosons, which are quanta of gauge fields, and finally, the only spin-zero particle is the Higgs boson, which is responsible for the mass of particles.

The "particle zoo" of the SM can be further categorized when looking at the matter particles and bosons separately:

Fermions The fermions can be divided into two groups of six particles each: quarks and leptons. Both these groups consist of three so-called generations, which tend to have successively increasing masses. The quarks are electrically charged particles that also interact weakly and strongly. On contrast, the leptons which comprise three electrically charged and three neutral particles do not take part in strong interactions, but all experience the weak force. For each fermion there is a corresponding antipartcle with the same set of quantum number having an opposite sign.[1]

Bosons The strong interactions are carried out by eight massless gluons, usually depicted by a single symbol g. There are four gauge bosons for the electroweak force: the massless photon is responsible for the electromagnetic exchanges, while the massive neutral Z^0 and charged $W^{\pm}$ bosons mediate the weak interaction. Gauge boson masses arise in the electroweak symmetry breaking, while for fermions the masses are generated through Yukawa interactions with the field associated to the Higgs boson.

A visual representation of this classification is shown in Fig. 1.1: for each particle the spin and electrical charge are provided with the most recent measured mass (or limit), the shaded areas indicate the coupling of the fermions to the corresponding force mediators. More detailed properties of these particles can be found in the "Particle Listings" of Ref. [15].

[1] Throughout this thesis, no differentiation between particles and antipartcles will be made given the identity of their interactions and properties.

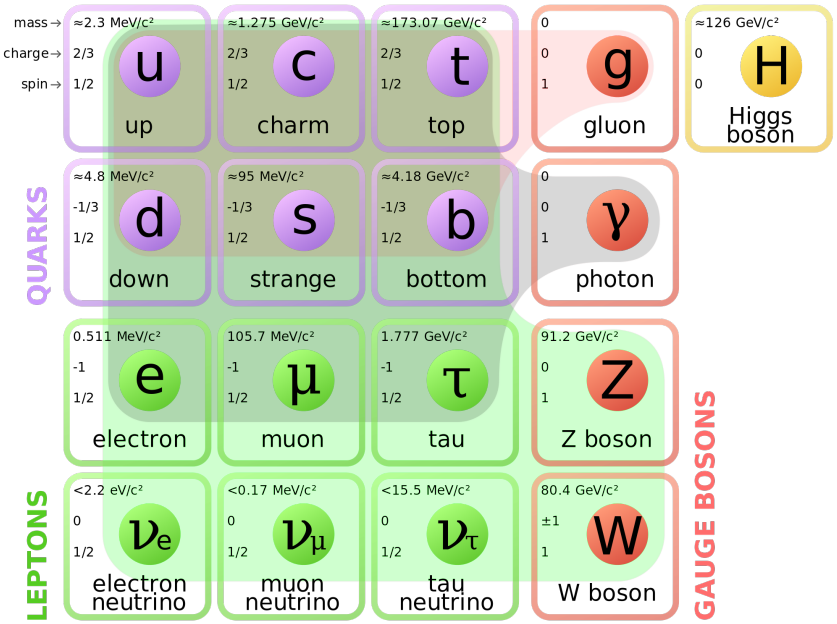


Fig. 1.1: All fundamental particles of the Standard Model with three generations of matter, the gauge and Higgs bosons; the electric charge, spin and mass (or limit for neutrinos) are given in the corners of the boxes. The shaded areas indicate which particles couple to which gauge boson according to the different interactions [16].

1.1.2 The Standard Model as a quantum field theory

Based on the principle of local gauge invariance, the SM is a renormalizable quantum field theory, which can be represented by the local gauge group $SU(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. This product represents the inclusion of the three fundamental forces into a single framework.

Strong interaction

The *colour* group $SU(3)_C$ is responsible for the strong interaction described by quantum chromodynamics (QCD). Each quark can have one of three colours as a quantum number, while anti-quarks hold an anti-colour. The corresponding triplet representation results in eight gluons, which are the force carries of the strong interaction and carry a colours and anti-colour. All leptons are $SU(3)_C$ singlets and do not couple to gluons.

The strong coupling constant α_s increases with separation distance, but is weak at very small distances, or high energies. This results in a phenomenon called *asymptotic freedom*, where quarks in a bound state, like the proton, do not interact much. On the contrary, if one tries to remove a quark from a proton, the accumulating energy rather creates new bound states (i.e. hadronization) than freeing the quark. This *confinement* implies the impossibility of observing a single free coloured quark. Therefore quarks can only form colourless states of two or three particles called *mesons* and *baryons*, respectively, or *hadrons* in general.

Hypothesized four- and five-quark states (called tetra and pentaquarks) were thoroughly searched for in the recent decades. Only recently the Belle and LHCb experiments succeeded in the experimental observation of such rare states [17, 18].

Electroweak interaction

At energies below the W-boson mass, the electromagnetic interaction of charged particles is described by quantum electrodynamics (QED) governed by the local gauge group $\mathrm{U}(1)_{\mathrm{EM}}$. In this case only a massless field exist, which is associated with the photon.

Initially also the weak interaction was considered independently with its massive gauge bosons $W^{\pm}$ and Z^0 under the group $\mathrm{SU}(2)_L$. All left-handed fermions[2] undergo weak interactions either through *charged currents* mediated by the $W^{\pm}$ bosons, or *neutral currents* via the Z boson. A straightforward example is the β-decay, where the neutron effectively decays through the W boson to a proton, electron and corresponding antineutrino.

However, as shown by Glashow, Salam and Weinberg [19–21], the electromagnetic and weak interactions are *de jure* parts of a unified electroweak force corresponding to the local gauge group $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$, where Y stands for the *hypercharge*, which any fermion owns independent of its chirality.

The actual gauge fields of $\mathrm{SU}(2)_L$ are arranged in a weak isospin triplet: $W^1_\mu, W^2_\mu, W^3_\mu$, while a single gauge field B_μ is associated to the $\mathrm{U}(1)_L$ group. From these fields the actual observable mass eigenstates are obtained through linear combinations:

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \tag{1.1}$$

$$W^{\pm} = W^1 \pm iW^2 \tag{1.2}$$

where θ_W is the electroweak mixing angle, or Weinberg angle, which can be measured experimentally since it relates the heavy boson masses: $M_W^2 = \cos^2_W M_Z^2$. In this elegant way the quanta of the electromagnetic and weak fields are combined into a single theory.

However, the fact that the two forces can *de facto* be observed separately, implies that this symmetry must be broken. The implementation of this electroweak symmetry breaking (EWSB) was found by Brout, Englert [22] and Higgs [23] and is referred to as the Brout-Englert-Higgs (BEH, or just Higgs) mechanism. It introduces a scalar potential, which permeates the whole universe and is symmetric except in its ground state: $V(\Phi) = m^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$ where m and λ are the bare mass and self-coupling parameters of the corresponding Higgs field Φ.

This field is a complex doublet with weak hypercharge $Y = 1$, that acquires a non-zero vacuum expectation value (VEV) v, which induces the spontaneous breaking of the SM symmetry. The Higgs boson mass is therefore given as $m_H = \sqrt{2\lambda}v$.[3] This VEV also provides mass to the heavy Z and $W^{\pm}$ bosons, while fermions obtain their mass through direct Yukawa couplings to the Higgs boson.

With the discovery of the top quark by the CDF and DØ experiments in 1995 [24,25], the SM fermionic content was finalized with the heaviest quark. It was not until 2012, when the ATLAS and CMS collaborations announced the discovery of particle similar to the Higgs boson predicted by the SM [4,5]. With this observation and further studies of the Higgs boson parameters the SM seems to have culminated in a fully consistent theory of particles and interactions of the microcosm.

[2] And right-handed antifermions. Handedness, or chirality describe the direction of the particle spin relative to its momentum.

[3] The expectation value of the Higgs field is fixed by the Fermi coupling $v = (\sqrt{2}G_F)^{-1/2} \approx 246\,\mathrm{GeV}$, but the self-coupling λ is a free parameter of the SM.

1.2 Unsolved issues of particle physics

As discussed before, the SM is an incredibly robust theory that is successfully verified in experiments since decades. Especially in the area of high-precision quantities (such as particle properties) as well as in the predictions of particles, the SM did stand all the tests. However, there are several reasons – some of experimental nature, others more from a theorist's aesthetic point of view – that require to assume the limitedness of the SM [15].

1.2.1 Experimental puzzles

The following experimental observations either do not fit to the Standard Model, or cannot be explained by it. Interestingly, most of these phenomena are observed on a global (cosmological) scale, i.e. at much larger distances, than common in particle interactions.

- **Neutrino oscillations**
  The observation of neutrino oscillations, i.e. changes of the lepton flavour over time, are a striking evidence for neutrinos to have mass. This, however, is in strong contradiction to the SM, which expects massless neutrinos. No approach compatible with the BEH-mechanism, which provides fermions with masses, exists to circumvent this discrepancy.

- **Baryon asymmetry**
  The standard *cosmological* model predicts, that equal amounts of baryonic matter and antimatter should have been produced after the Big Bang. However, the absence of antimatter in everyday life is obvious, while cosmic ray experiments report only a small fraction of antipartcles bombarding Earth and no explicit signs of annihilation regions. Current estimates show, that antimatter is nine orders of magnitude less abundant than matter. The mechanisms present in the SM to cause such an asymmetry (e.g. CP-violation) do not suffice to describe this observed imbalance.

- **Dark Matter**
  Measurements of the Cosmic Microwave Background (CMB) indicate, that the aforementioned matter (including antimatter) is estimated to contribute only about 5% to the total energy budget of the Universe [26]. Another 27% are attributed to so-called *Dark Matter*, which is supposed to be some gravitationally interacting substance pervading the galaxies.[4]

  Several independent phenomena like the CMB, galaxy rotation curves, gravitational lensing, etc., provide a motivation for the actual existence of such matter, but none of them can be explained with conventional matter described by the SM.

- **Gravity**
  Undoubtedly, gravity does exist and also acts on any massive particle, but its effect would be only visible at energies around the Planck scale $M_{\mathrm{Pl}} \approx 10^{19}$ GeV. Moreover, the integration into the SM would require to quantize the gravitational field, which however does not allow any predictions due to non-renormalizability.

[4] The remaining 68% are attributed to so-called *Dark Energy*, which represents even a bigger mystery than Dark Matter.

1.2.2 Theoretical questions

In addition to the experimental reasons, there are several deficits of the SM, that are considered by theorists more from an aesthetic point of view, or do not fit into the instinctive perception of the SM as a comprehensive theory of particle physics.

- **Grand Unification**
  The consecutive steps of unifying the electric and magnetic forces, and thereafter the electromagnetic with weak forces into the electroweak, gives rise to an idea of also merging the electroweak and strong interactions. Indeed, the fact that the running coupling constants of the electromagnetic, weak and strong forces seem to converge at about 10^{16} GeV feeds this hope further.

  At this scale, the SM would cease to be valid and should be replaced by a more general theory. The plethora of such theories is referred to as Grand Unified Theories (GUT), which also gives a name to this unification energy as the *GUT scale*. It is shown, that such a force unification could be possible if building a SM-inspired theory based on local gauge groups of higher orders, such as $SU(5)$ or $SO(10)$.

- **Hierarchy problem, naturalness and fine-tuning**
  As seen before, the masses of the electroweak gauge bosons lie at a scale of $O(100)$ GeV, which defines the electroweak energy scale. If new physics shows up only at the GUT scale, this would imply that in the interval of $O(10^2)$ to $O(10^{16})$ GeV no new particles exist. This large difference in the force hierarchy is considered as *unnatural* from an aesthetic point of view [27].

  In addition to this aesthetic discomfort, this problem can be anticipated when considering the composition of the Higgs boson mass.

  Given that the Higgs boson couples to each massive SM fermion f via a Yukawa term λ_f, one-loop radiative corrections (cf. Fig. 1.2, *left*) are affecting its bare mass as:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + \dots \tag{1.3}$$

  where Λ_{UV} is an ultraviolet (upper) momentum cutoff, which is used to restrict the loop integral, and can be interpreted as a lower estimate for a new energy scale at which the SM would cease to be valid and new physics should be expected.

  A natural assumption for the cutoff would be either the Planck scale, or the slightly lower GUT scale, with $\Lambda_{\mathrm{UV}} \approx 10^{16}$ GeV, since the unification of forces would imply new conditions not covered by the SM. This discrepancy between the scale of these corrections and the actual observed Higgs boson mass of 125 GeV again poses the *hierarchy problem* mentioned before.

  To comply with the Higgs mass observed at the electroweak scale, its bare mass should compensate for the large corrections from the one-loop diagrams. This however requires a large *fine-tuning* of the SM parameters, which is not desirable in this kind of theory. Such a parameter fine-tuning can only be avoided, if new physics phenomena are present at much lower scales of about $O($ TeV).

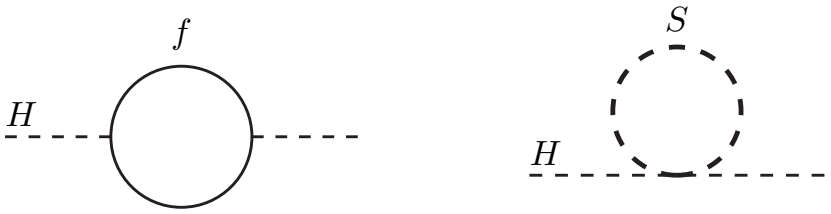


Fig. 1.2: Feynman diagrams corresponding to the one-loop quantum corrections to the Higgs boson mass due to a Dirac fermion f (*left*) or a scalar S (*right*) [6].

1.3 Supersymmetry

The aforementioned problems naturally require either an extension of the existing SM, or a more general descriptive theory-of-everything. A plethora of such theories beyond the Standard Model (BSM) were proposed over the last decades, but only few are of phenomenological interest, i.e. are predicting phenomena, which might be observed experimentally in the near future. Within these, Supersymmetry (SUSY) – the supersymmetric extension of the SM – is considered as one of the most promising theories, that would solve several shortcomings of the SM and provide a phenomenological background for experimental physics.

This discussion is based mostly on the reviews in Refs. [15] and [6].

1.3.1 Closing the hierarchy problem

Even though the original motives for SUSY theories were of different nature [28–32], the hierarchy problem provides a perfect justification for a supersymmetric extension of the SM.

In the previous section, the quantum corrections to the Higgs mass were considered only from fermions. If one would instead assume a similar contribution from a heavy scalar particle S like in Fig. 1.2 (*right*), the correction could be written as:

$$\Delta m_H^2 = +\frac{|\lambda_S|^2}{16\pi^2}\Lambda_{\mathrm{UV}}^2 + \dots \tag{1.4}$$

where λ_S represents the respective coupling of the scalar to the Higgs field and terms logarithmically diverging in Λ_{UV} are omitted. This way, the fermion one-loop contributions can be cancelled if scalar particles exist, such as $2|\lambda_S|^2 = |\lambda_f|^2$.

In order to satisfy this requirement, a new symmetry between bosons and fermions is proposed, called Supersymmetry (SUSY). The supersymmetric transformation turns fermionic states into bosonic states, and vice versa, such that each SM particle receives a so-called *superpartner* from the SUSY side. Since this transformation requires the change of the spin angular momentum (exactly by $1/2$), this must be a space-time symmetry. SM particles and their SUSY partners form *supermultiplets* in a way, that the only differing quantum number is the spin.

As SM fermions are left- and right-handed (chiral), each of those states receives a separate scalar partner bound into *chiral supermultiplets*. This allows for an exact cancellation of the aforementioned leading order one-loop contributions from Eqs.1.4 and 1.4. Additional logarithmic contributions would also be cancelled, if the supersymmetry is unbroken, since this implies the superpartnes to have the same mass.

Table 1.1: Chiral supermultiplets in the Minimal Supersymmetric extension of the Standard Model [6]. The spin-0 fields are complex scalars, and the spin-1/2 fields are left-handed two-component Weyl fermions. Symbols in the second column denote each of the chiral supermultiplets.

Names & Multiplets		spin 0	spin 1/2	$SU(3)_C$, $SU(2)_L$, $U(1)_Y$
squarks, quarks	Q	$(\widetilde{u}_L \;\; \widetilde{d}_L)$	$(u_L \;\; d_L)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
(3 generations)	$\overline{u}$	$\widetilde{u}_R^*$	$u_R^\dagger$	$(\mathbf{3}, \mathbf{1}, -\frac{2}{3})$
	$\overline{d}$	$\widetilde{d}_R^*$	$d_R^\dagger$	$(\mathbf{3}, \mathbf{1}, \frac{1}{3})$
sleptons, leptons	L	$(\widetilde{\nu} \;\; \widetilde{e}_L)$	$(\nu \;\; e_L)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
(3 generations)	$\overline{e}$	$\widetilde{e}_R^*$	$e_R^\dagger$	$(\mathbf{1}, \mathbf{1}, 1)$
Higgs, higgsinos	H_u	$(H_u^+ \;\; H_u^0)$	$(\widetilde{H}_u^+ \;\; \widetilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, +\frac{1}{2})$
	H_d	$(H_d^0 \;\; H_d^-)$	$(\widetilde{H}_d^0 \;\; \widetilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

No scalars with the electron mass or any other superpartners have been found yet, which indicates that *supersymmetry is broken*. Several approaches like the gravity or gauge mediated supersymmetry breaking [33] exist, but are out of scope of this thesis. However, if the superpartners have masses of the order $O(\text{TeV})$, the required fine-tuning is still acceptable, and SUSY can be considered natural. This *soft-supersymmetry breaking* is assumed in the following.

1.3.2 Minimal Supersymmetric Standard Model

The Minimal Supersymmetric extension of the Standard Model (MSSM) represents the simplest possible way of constructing a SUSY model using the least number of particles [6].

A peculiarity of the MSSM is the requirement of two Higgs chiral supermultiplets, dictated by the conditions of gauge anomaly cancellation. Their only difference has to be in the weak hypercharge, which automatically allows the H_u multiplet with $Y = +1/2$ to give masses to up-type quarks, while only the H_d with $Y = -1/2$ has the necessary Yukawa couplings for down-type quarks and charged leptons. In fact, each such doublet contains a neutral and charged Higgs field, like in the SM case.

By convention, all superpartners of SM particles are indicated by the same symbols with a tilde ($\widetilde{\ }$), and the names of SM fermion partners are constructed by prepending an "s" for scalar (e.g. *selectron*), while the fermionic gauge counterparts names are formed by appending "ino" (e.g. *higgsino*). Table 1.1 summarizes all the chiral supermultiplets of the MSSM. Note, that the subscripts L and R of the scalar sleptons correspond only to the helicity of their SM partners.

The SM gauge bosons together with their fermionic supersymmetric partners, or *gauginos*, form gauge supermultiplets, as indicated in Table 1.2. The gluino is the spin-1/2 colour octet superpartner of the $SU(3)_C$ gauge mediator. As the electroweak gauge symmetry $SU(2)_L \otimes U(1)_Y$ is associated to the vector bosons $W^\pm$, W^0 and B^0, their corresponding superpartners are $\widetilde{W}^\pm$, $\widetilde{W}^0$ and $\widetilde{B}^0$, called *winos* and *binos*, respectively.

In contrast to the SM particles, which obtain their mass exclusively through the electroweak symmetry breaking, all MSSM particles can have a Lagrangian mass term also in absence of EWSB. Nevertheless, an important feature of the MSSM is that the aforementioned superpartners do not necessarily correspond to the mass eigenstates. Mixing between the electroweak

Table 1.2: Gauge supermultiplets in the Minimal Supersymmetric Standard Model. [6]

Names	spin 1/2	spin 1	$SU(3)_C$, $\mathrm{SU}(2)_L$, $\mathrm{U}(1)_Y$
gluino, gluon	$\widetilde{g}$	g	(**8**, **1** , 0)
winos, W bosons	$\widetilde{W}^{\pm}$ $\widetilde{W}^{0}$	$W^{\pm}$ W^{0}	(**1**, **3** , 0)
bino, B boson	$\widetilde{B}^{0}$	B^{0}	(**1**, **1** , 0)

gauginos and higgsinos (which now have the same spin) results in electrically charged and neutral mass eigenstates called *charginos* and *neutralinos*, respectively, while same-charge scalars mix within the chiral supermultiplets. The only exception is the gluino, which carries only colour and therefore does not mix with other particles.[5]

The mass eigenstates that differ from the SUSY fields after the mixing are as follows:

squarks	$\widetilde{t}_L, \widetilde{t}_R, \widetilde{b}_L, \widetilde{b}_R \rightarrow \widetilde{t}_1, \widetilde{t}_2, \widetilde{b}_1, \widetilde{b}_2$	**staus**	$\widetilde{\tau}_L, \widetilde{\tau}_R, \rightarrow \widetilde{\tau}_1, \widetilde{\tau}_2$
neutralinos	$\widetilde{B}^0, \widetilde{W}^0, \widetilde{H}_u^0, \widetilde{H}_d^0 \rightarrow \widetilde{\chi}_1^0, \widetilde{\chi}_2^0, \widetilde{\chi}_3^0, \widetilde{\chi}_4^0$	**charginos**	$\widetilde{W}^{\pm}, \widetilde{H}_u^{+}, \widetilde{H}_d^{-} \rightarrow \widetilde{\chi}_1^{\pm}, \widetilde{\chi}_2^{\pm}$

Interestingly, given these features and experimentally measured SM properties, some MSSM scenarios require at least one neutral Higgs boson to be lighter than 135 GeV. The observation of a Higgs boson at 125 GeV harmonically fits into this prediction.

In total, the minimal version of supersymmetry contains 124 independent parameters, of which 18 are attributed to the pure SM, and one is a Higgs sector parameter corresponding to its SM mass. The 105 new parameters are split as follows: eight plus 40 CP-violating phases in the gaugino/higgsino and squark/slepton sectors, respectively, and 21 squark and slepton masses with 36 mixing angles for defining their mass eigenstates.

Answers to other particle physics problems

Together with a possible solution to the hierarchy problem, the MSSM also offers compelling answers to other phenomena, which cannot be explained within the SM (cf. Sec. 1.2).

- The running gauge couplings get affected by the plethora of new supersymmetric particles in such a way that the unification at the GUT scale becomes more precise.
- The intrinsic mechanisms of CP-violation of the SM are not sufficient to satisfy Sakharov's conditions for the observed baryon asymmetry [34]. In SUSY models, new sources of CP-violation and supersymmetric loops provide a possibility to satisfy those conditions.
- Unlike in the SM, baryon and lepton number conservation is generally not imposed in the MSSM. However, the lower bound on the proton lifetime ($\approx 10^{32}$ years) strongly disfavours baryon number violation. A new symmetry, called R-parity, has to be introduced to eliminate the corresponding violation:

$$R = (-1)^{3B+L+2s} \tag{1.5}$$

[5] The gluino mass itself is related to a single parameter of the MSSM, $m_{\widetilde{g}} = |M_3|$.

where B and L are the baryon and lepton numbers, respectively, and s is the spin. This implies, that all SM particles have positive R-parity, while their supersymmetric partners have negative R-parity.

Together with prohibiting proton decays, the conservation of R-parity in particle scattering and decay processes implies that SUSY particles can only be produced in pairs, and that the lightest supersymmetric particle (LSP) is absolutely stable.

Provided that the LSP is a massive particle without colour or electric charge, it becomes a perfect candidate for the omnipresent Dark Matter.

- If SUSY is a local symmetry, the theory is forced to include gravity by introducing a spin-2 massless *graviton* accompanied by its superpartner, the spin-3/2 *gravitino*. If the gravitino is light enough, it can play the phenomenologically important role of the LSP. Such *supergravity* theories are still non-renormalizable, but at least provide a link from quantum field theories to General Relativity.

1.3.3 Other extensions beyond the Standard Model

Supersymmetry is not the only BSM extension, that provides solutions to the aforementioned SM shortcomings. Given the recent discovery of a SM-like Higgs boson, these theories also have to account for this new particle. Correspondingly, models related to the Higgs physics seem to be very theoretically appealing, but as well models with additional dimensions could provide an insight to many current questions. In the following, these two classes of BSM models will be discussed in short.

Composite Higgs

Theories, where the Higgs boson is considered as a bound state of new dynamics that becomes powerful at the weak scale, are called *composite Higgs* models. In analogy to the pion, which is the lightest bound particle of QCD, the SM Higgs can be made much lighter than the other strong resonances, if it acts as a pseudo Goldstone boson (the *Little Higgs* model [15]).

A new set of particles around the Fermi scale – gauge bosons, vector-like quarks and massive scalars – allows the cancellation of the SM quadratic divergences similar to supersymmetric theories, however the new particles have the same sign as their SM partners. Such particles provide a phenomenological base for searches at the colliders.

Extra dimensions

Ideas, that the surrounding space-time is only a subset of a higher-order dimension have been already proposed by Kaluza and Klein in the early XX$^{\text{th}}$ Century [35, 36] in an attempt to unite electromagnetism and gravity. String theories revived this formalism in the 1980's by predicting extra dimensions that are compactified at a scale close to M_{Pl}, and therefore not testable experimentally.

In an approach of Arkani-Hamed, Dimopoulos and Dvali [37] the observed weakness of gravity is explained by postulating at least two extra dimensions in which only gravity could propagate. The size of these extra dimensions is predicted to range from a few millimetres to $\sim 1/$ TeV, which could have an effect on current observables. Randal-Sundrum (RS) models [38] extend this approach into a five-dimensional warped space-time with a compactification scale of the order TeV. In this case not only gravitation, but also the SM gauge fields are permitted to propagate into the additional dimensions.

1.4 Searches for SUSY at colliders

From a phenomenological point of view, the MSSM offers a wide variety of possibilities to produce supersymmetric particles in collisions, provided the centre-of-mass energy is sufficiently high. The subsequent decay chains through intermediate SUSY and SM particles even further enhance the possible final states, that can be used to for a search of new BSM physics. Moreover, the long decay chains are normally accompanied by strong production of quarks, which ultimately hadronize into particle jets.

Of particular interest are models with R-parity conservation, since this implies, that the decay chain of any SUSY particle should eventually terminate in the production of an LSP. Since the lightest supersymmetric particle is supposed to be only weakly interacting, it can be seen as an analogue to the SM neutrino, which escapes detection in typical collider experiments. Correspondingly, it creates an imbalance in the (transverse) energy of the collisions at a magnitude proportional to its mass.

Overall, the final state signature of events in many SUSY scenarios might become truly spectacular, featuring multiple jets already from the hard process, as well as large missing energy in case of R-parity conservation.

Despite (or owing to) such striking signatures, no hints of SUSY have been found yet by the previous generations of particle colliders, like the Large Electron-Positron Collider (LEP) or Tevatron. The experiments at the LHC (cf. Chapter 2) are now taking over the baton in searching for new physics at the TeV scale.

1.4.1 Simplified models

From the experimental perspective, such a variety of possible final states and model parameters provides a certain challenge, since the amount of work-force and time is finite. To restrict the possible phase space for searches and concentrate on particular phenomenologically sensible signatures, either constraints or simplifications on the SUSY models have to be introduced.

One such model, the constrained MSSM (cMSSM), is an attempt to limit the number of arbitrary parameters in the MSSM by assuming universal scalar and gaugino masses at the GUT scale. In this case the number of free parameters is reduced to five[6]. The spectrum of supersymmetric particles still remains sophisticated, as shown in the left plot of Fig. 1.3. Before the start of the second LHC run, the cMSSM was one of the primary tools to guide and interpret searches for SUSY.

A less constrained phenomenological MSSM (pMSSM), featuring 19 additional parameters, is also used to interpret results of SUSY searches [39, 40]. In this case, a multidimensional fit is performed over all parameters, allowing for a more general interpretation of the MSSM.

An alternative to the use of constrained SUSY models is implemented by the framework of so-called *simplified model spectra (SMS)* [41, 42]. Within this effective approach only a small number of supersymmetric particles is considered, of which some are produced in pairs and decay exclusively through a limited number of channels. This simplification allows to study SUSY models with only a few phenomenologically important parameters, such as the cross section, branching ratios and masses, while staying independent of theory-related details. In addition, this facilitates the reinterpretation of results within other (non-)SUSY theories.

[6] The respective scalar and gaugino masses, the trillinear coupling, the ratio of the vacuum expectations values of the two Higgs doublets and the sign of the higgsino mass term.

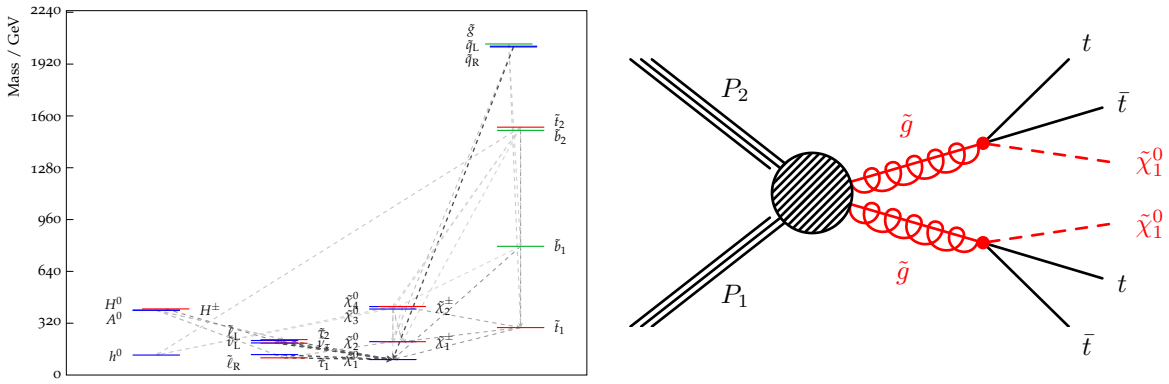


Fig. 1.3: *Left*: SUSY particle spectrum for a pMSSM model [43]. The dashed lines indicate possible decay channels, with darker colours representing higher branching ratios.
Right: Feynman diagram for the T1tttt simplified SUSY model of gluino pair production with subsequent decays as $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$, where the neutralino $\tilde{\chi}_1^0$ is the LSP.

The approach of SMS is used in the present thesis, and a variety of searches performed by the ATLAS and CMS experiments at the LHC within this framework will be discussed in the following.

1.4.2 Status of SUSY searches after *Run 1*

Given that the LHC is a proton collider, the majority of particles are produced in the sector of strong interactions. For the case of SUSY models this corresponds to the supersymmetric partners of the gluon and quarks, the gluino and squarks, respectively. In particular, the largest cross sections are expected for gluino pair and squark–antisquark production.

Gluino searches

Within models used by the ATLAS and CMS experiments (cf. Sec. 2.2), the gluino pair production is followed by the subsequent decay of each gluino, modelled as a three-body decay to a SM quark-antiquark pair[7] together with the lightest neutralino, which is considered as the LSP: $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$ [44]. The decay channels are further categorised depending on the quarks being light-flavoured ($q = u, d, s, c$), or heavy-flavoured (bottom $q = b$, or top $q = t$).

The decay through top-quark pairs, for which the model is called **T1tttt** in CMS, is illustrated in the right plot of Fig. 1.3. A distinguishing feature of this channel, w.r.t. to the other quark flavours, is that it complements the "usual" SUSY signature of many jets and missing energy with the presence of four top quarks, which subsequently decay to a pair of W boson and b quark.

The different decay channels of the W boson allow to search for this signal model in various final states. When all four W bosons decay into quarks (*hadronically*), there are eight hard jets in the event, while the most probable case of one *leptonically* decaying W boson produces an event with at least six jets accompanied by a lepton.

[7] The corresponding squark partners are assumed to be *decoupled*, i.e. their masses are much higher than $m_{\tilde{g}}$, such that these squarks are virtual.

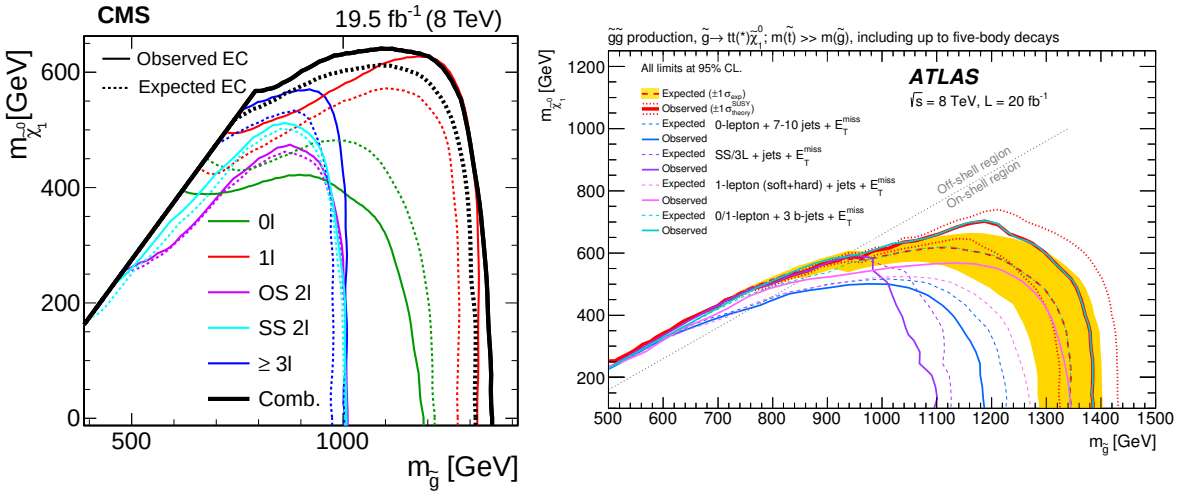


Fig. 1.4: Exclusion contours at 95% C.L. in the gluino-neutralino mass plane of $\sqrt{s} = 8\,\text{TeV}$ analyses of CMS [45] (*left*) and ATLAS [46] (*right*) searching for simplified gluino pair production with subsequent decays to $t\bar{t}$ and a neutralino, which is the LSP.

Both, the ATLAS and CMS collaborations, performed searches in the different channels during the first run of the LHC. Figure 1.4 shows exclusions contours[8] in the gluino-neutralino mass plane obtained from proton collisions at a a centre-of-mass energy of 8 TeV. The contours in the left plot show the results obtained by the CMS experiment of the individual final state searches (in colour) compared to their combination (in black). The right plot shows the corresponding results of the ATLAS experiment.

When interpreting these results, two regions in the mass plane can be identified by different kinematic properties:

$m_{\tilde{g}} < m_{\tilde{\chi}_1^0} + 2m_t$ When the gluino-neutralino mass splitting is smaller than double the top-quark mass, at least one of these top quarks is produced off-shell. This results in softer decay products, but large missing energy and indicates a *compressed mass spectrum*.

In this case the experimental sensitivity to potential signal is roughly independent of the gluino mass and signals are excluded with $m_{\tilde{\chi}_1^0}$ up to 650 GeV.

$m_{\tilde{g}} \geq m_{\tilde{\chi}_1^0} + 2m_t$ The gluino is much heavier than the neutralino, and both top quarks decay on-shell to high energy products, which characterizes an *open spectrum*.

The experimental sensitivity to such events is mostly independent of the neutralino-LSP mass, and the exclusion reaches up to gluino masses of about 1400 GeV.

For both experimental results the combined limit is mainly driven by the single-lepton channel, where one of the four W bosons decays leptonically[9]. The reasons for such a high sensitivity are the relatively high branching ratio of about 40% and the final state signature with one lepton and several jets, which is more easy to distinguish from SM backgrounds, than in

[8] At 95% confidence level (C.L.), see details in Section 7.2

[9] In general, τ leptons are considered only if they also decay leptonically into an electron or muon.

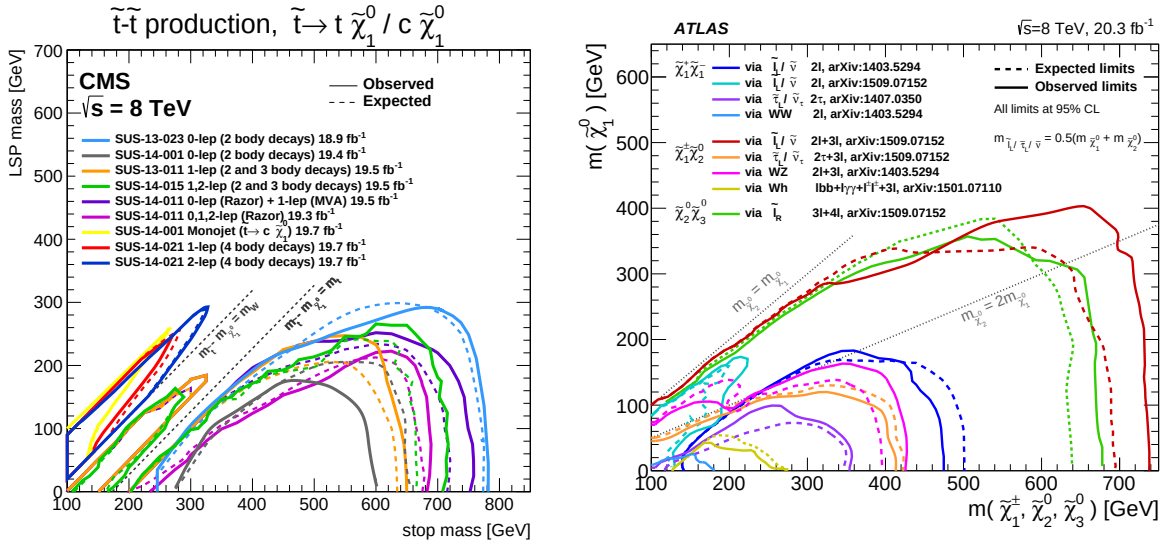


Fig. 1.5: Exclusion limits at 95% C.L. for searches of simplified pair production of stop quarks by CMS [49] (*left*) and different neutralinos or charginos by ATLAS [50] (*right*).

the full hadronic case. These properties of the search topology serve as a motivation for the analysis presented in this thesis (cf. Chapter 5).

Alternatively to the gluino decay into the top-quark pair, searches were performed and interpreted for the respective decays into b quarks or light-flavour quark pairs. In addition to the different final state signatures, these searches can probe even more compressed scenarios in terms of the gluino-neutralino mass splitting, since these quarks are much lighter than the top quark. In this case the limit on the neutralino mass reaches up to 750 GeV, while gluino masses are also excluded up to 1400 GeV [47], which is similar to the T1tttt limit.

Squark searches

The search for direct production of supersymmetric quark partners allows to test specific assumptions about the corresponding squark masses. Since the partners of the third-generation quarks are of great importance in the cancellations of divergences to the Higgs mass, the masses of at least the top and bottom squarks are expected to be small, below 1 TeV [48].

Similar to the gluino searches, here simplified models are considered with production of squark-antisquark pairs, which are either the light-flavour partners $\widetilde{q} = (\widetilde{u}, \widetilde{d}, \widetilde{s}, \widetilde{c})$, bottom $\widetilde{b}$ or top $\widetilde{t}$ squarks. Typically, subsequent two-body decays to the SM quark partner and a neutralino LSP are modelled $\widetilde{q} \to q\widetilde{\chi}^0_1$.

The *left* plot of Fig. 1.5 shows 8 TeV results from CMS searches for direct top-squark production. Regions where the squark-neutralino mass difference is close to the top-quark mass become especially hard for searches, since in this case the signal is almost indistinguishable from the SM $t\bar{t}$ background. In this phase space the sensitivity is very irregular and neutralino masses can be only excluded up to 300 GeV, while for large mass splitting top squarks are excluded below 800 GeV.

EWKino searches

In addition to searches for superpartners of strongly interacting SM particles, various simplified scenarios of chargino or neutralino (*EWKino*) production were also considered during the first run of the LHC. In all cases the $\widetilde{\chi}_1^0$ is considered as the lightest supersymmetric particle.

Although their production is heavily suppressed due to the proton content, these final states typically involve multiple leptons, which can be used to discriminate these SUSY signals from the SM background.

Figure 1.5 (*right*) shows the results obtained by the ATLAS experiment from the 8 TeV LHC data. Depending on the specific type of the produced SUSY particles, the mass limits extend to 500, 690 and 720 GeV for $\widetilde{\chi}_1^+\widetilde{\chi}_1^-$, $\widetilde{\chi}_2^0\widetilde{\chi}_3^0$ and $\widetilde{\chi}_1^\pm\widetilde{\chi}_2^0$ production, respectively. The $\widetilde{\chi}_1^0$ LSP masses are excluded above 400 GeV.

A more complete review of SUSY results from the first LHC run is presented in Refs. [7, 8], covering as well searches for R-parity violating SUSY or other more exotic scenarios.

1.4.3 Outlook for *Run 2*

The results of the first LHC run can be shortly summarised as follows. A (the) Higgs boson was discovered at 125 GeV, finalizing the particle content of the Standard Model. No significant deviations from the SM predictions have been observed in its overall wide field of validity. Moreover, no hints of new physics beyond the SM, and in particular compatible with SUSY theories, were found. These results allow to refine the strategy of SUSY searches for *Run 2*.

Constraints on SUSY

Apart from the negative results of SUSY searches, the phase space of supersymmetric models is even further constrained by the discovery of the SM Higgs boson, which plays a crucial role in the SM and SUSY sectors. Within the minimal SUSY model this results in the following requirements on sparticle masses to comply with the hierarchy / naturalness / fine-tuning problem resolution [51]:

- third generation quarks superpartners ($\widetilde{t}$ and $\widetilde{b}$) must be lighter than the TeV scale,
- gluinos cannot be much heavier than 1 TeV,
- the higgsino mass is favoured around 200 GeV.

Clearly, the results of SUSY searches from the ATLAS and CMS experiments challenge these requirements in a strong way, despite the simplistic assumptions on the search models. In more realistic models the decay chains might feature several decay modes with different SUSY particles, resulting in smaller sensitivity to the aforementioned channels.

However, there is still room for expanding the searches towards larger masses, e.g. with higher centre-of-mass energy, and/or in the regions with compressed spectra, which are less trivial to examine. The prospects of searches at higher energies are discussed in the following.

SUSY at $\sqrt{s} = 13$ TeV

The profit from a new centre-of-mass energy is twofold and connected to the change in the proton parton distribution function (PDF), in which the contribution of gluons increases significantly with energy.

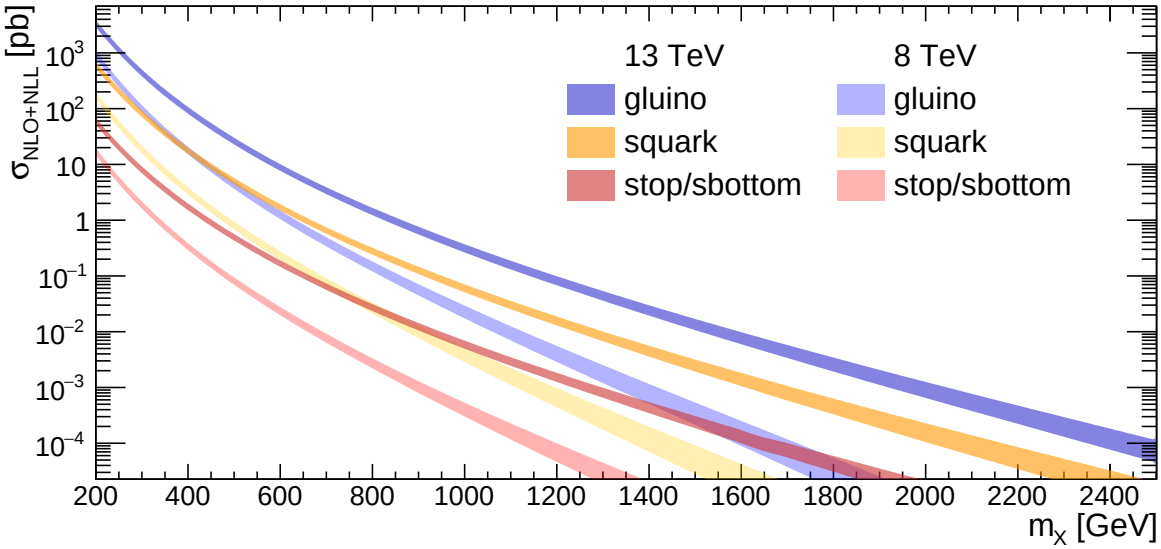


Fig. 1.6: SUSY particle pair production cross-sections for different masses of gluinos, squarks and stops/sbottoms in proton collisions at $\sqrt{s} = 8$ and 13 TeV. Based on Refs. [52–55].

Since the strong SUSY sector is mainly produced in gluon and quark annihilation, the increase in centre-of-mass energy results in correspondingly larger cross sections. This effect is illustrated in Fig. 1.6, where the production cross sections for gluon, squark and stop/sbottom pair production are plotted with respect to the mass of the corresponding sparticles.

In general, for a given mass the largest section is achieved for gluino production due to the dominant gluons in the proton. The change in the centre-of-mass energy from 8 to 13 TeV increases the cross section of 1400 GeV gluinos by a factor of 30, whereas for 800 GeV squarks the increase is only tenfold.

In contrast, the production cross sections for such typical SM backgrounds as $t\bar{t}$+jets or W+jets increase by a factor of 3 and 2, respectively [56–59]. This implies, that the SUSY-signal–to–SM-background ratio is increasing tenfold for the case of gluino searches. Such a large increase in sensitivity makes the search for strong superpartner production, and in particular for gluino pairs, a very promising analysis at $\sqrt{s} = 13$ TeV.

Analysis strategy

Based on the aforementioned arguments, the analysis presented in this thesis is concentrating on the search of SUSY in gluino-pair production using the new *Run 2* data of the LHC at $\sqrt{s} = 13$ TeV. This search is optimized and interpreted within the simplified SUSY model T1tttt, where the gluinos decay directly into a top-antitop SM quark pair and a neutralino, which is considered as the lightest supersymmetric particle. The final state of interest features a single isolated lepton stemming from one of the top-quark decays, while the remaining three hadronically decaying tops provide a large amount of jets.

Chapter 2

Experimental Setup

2.1 The Large Hadron Collider 21
2.1.1 The CERN accelerator complex 24
2.1.2 Outlook towards the future of the LHC 25
2.1.3 LHC experiments 26
2.2 The Compact Muon Solenoid experiment 29
2.2.1 Magnet 31
2.2.2 Inner Tracker 31
2.2.3 Electromagnetic Calorimeter 33
2.2.4 Hadron Calorimeter 34
2.2.5 Muon System 38
2.2.6 Trigger and Data Acquisition Systems 40
2.2.7 Luminosity measurement 41
2.2.8 Upgrades of CMS 42
2.3 Event simulation 43
2.3.1 Event generators 43
2.3.2 Detector simulation 44

Modern high-energy physics requires a large variety of tools for research at the leading edge. This chapter gives an overview of the experimental setup used to collect data for this analysis. The first section introduces the most powerful particle accelerator to date – the Large Hadron Collider (LHC) – including the pre-accelerator complex and the major experiments.A detailed overview of the CMS experiment will be given in Sec. 2.2, where the different detector components and subsystems will be presented. Finally, the event simulation tools used to validate and complement the experiment are described in Sec. 2.3.

2.1 The Large Hadron Collider

The LHC [60, 61] is currently the largest particle accelerator worldwide. It is primarily designed to collide protons at a centre-of-mass energy of 14 TeV. Operated by the European Organization for Nuclear Research – *Conseil Européen pour la Recherche Nucléaire* – (CERN) and located in the suburbs west of Geneva, Switzerland the accelerator complex lies between the Jura mountains and Lake Geneva at depths ranging from 50 to 150 m. Originally build for the LEP between 1984 and 1989, the tunnel consists of eight arcs and eight straight sections, with a total circumference of 27 km.

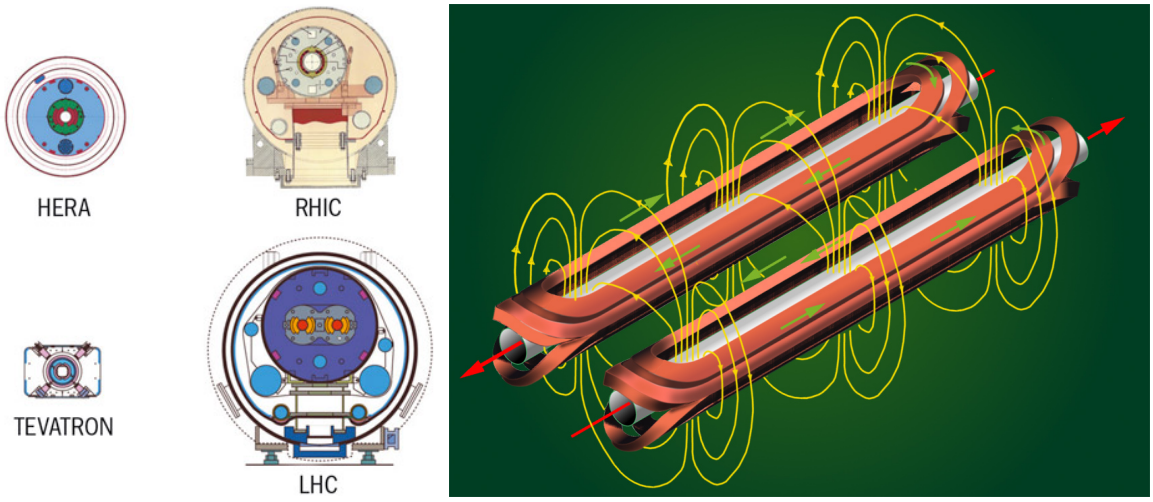


Fig. 2.1: *Left*: cross-section comparison of dipole magnets from the Tevatron to the LHC [62]. *Right*: schematic view of the two dipole coils of the LHC cryodipole [63]. The red arrows indicate the beam direction, the yellow show the magnetic field lines.

The construction of the LHC began in 2000 after the LEP shutdown, and the first beams were circulated in September, 2008. During this initial commissioning of the machine an electrical fault in the magnet system caused a severe damage of several components including the cryogenic cooling system. The recovery delayed the first collisions until November, 2009.

The LHC can be considered as the cutting edge of particle accelerator technology: it incorporates various methods already probed at different synchrotrons worldwide, but pushes the limits even further. Due to the limitations coming from the re-use of the LEP tunnel, this proton collider had to be fit into the same space, but now hosting the more energetic same-sign proton beams. Therefore the traditional two-ring solution like was used at the Tevatron (Fermilab, Batavia, USA) and HERA (DESY, Hamburg, Germany) colliders could not made to be use of. Instead, the LHC beam pipes and magnet systems are placed into a single cryomodule (twin-bore solution): the antiparallel magnetic fields are generated by independent coils which are nevertheless magnetically coupled due to their proximity (see Fig.2.1).

As in the previous generation of colliders, the LHC uses Niobium-Titanium (NbTi) Rutherford cables as superconductors, but the temperature of the Helium is reduced from 4.2 K to 1.8 K bringing it into the superfluid state. This allows to increase the operational magnetic field strength to 8.33 T with respect to the 5 T of HERA and the Tevatron.

Altogether more than 9000 magnets are used in the LHC, of which only 1232 are the main superconducting dipoles. However, being the longest modules with 15 m length, they make up 70% of the total 27 km circumference. The beams are focused by superconducting quadrupole magnets, while the remaining bulk of magnets is used to correct the main fields using sextupole and higher order magnets.

The beam acceleration takes place at the so-called Point 4 (between the ALICE and CMS experiments), where 16 superconducting cavities are installed. With a operational frequency of 400 MHz the radio frequency cavities provide an acceleration of 485 keV per turn, allowing to accelerate the protons from the initial 450 GeV to the nominal 7 TeV in about 20 minutes.

Table 2.1: LHC beam peak parameters during the major runs over the years of operation in comparison with the design values [60, 64–66]. For pileup (PU) the average is given.

Period		Energy	Bunch spacing	N_b	N_p	β^* [m]	$\mathcal{L}$ [$cm^{-2}s^{-1}$]	PU
pp	Design	7 TeV	25 ns	2808	1.2×10^{11}	0.55	1.0×10^{34}	19
	2015	6.5 TeV	25 ns	2244	1.2×10^{11}	0.8	5.0×10^{33}	15
	2012	4 TeV	50 ns	1374	1.6×10^{11}	0.6	7.7×10^{33}	35
	2011	3.5 TeV	50 ns	1374	1.5×10^{11}	1.0	3.6×10^{33}	10
PbPb	Design	2.76 TeV/u	100 ns	592	7×10^{7}	0.5	1.0×10^{27}	< 1
	2015	2.76 TeV/u	150 ns	518	2×10^{8}	0.8	3.0×10^{27}	< 1

LHC parameters

A key parameter for the experiments is the instantaneous luminosity $\mathcal{L}(t)$ delivered by the collider, which can then be used to calculate the event rate for a given physical process X with a corresponding production cross-section σ_X:

$$\frac{dN_X}{dt} = \sigma_X \times \mathcal{L}(t). \tag{2.1}$$

Since the LHC is a synchrotron performing head-on collisions of particle bunches, the luminosity depends on the number of bunches per beam N_b, number of particles per beam N_p and revolution frequency f_{rev}:

$$\mathcal{L} = \frac{N_p^2 N_b f_{\text{rev}} \gamma_r}{4\pi\varepsilon_n \beta^*} F. \tag{2.2}$$

More subtle parameters include the relativistic gamma factor γ_r, the normalized transverse beam emittance ε_n, the beta function β^* at the interaction point (IP), and the reduction factor F due to the beam crossing angle.

By design, the two LHC beams would be capable of carrying 2808 bunches containing 1.15×10^{11} protons each. The maximal center-of-mass energy for the proton-collisions is envisaged to be 14 TeV. At that energy and with a total bunch crossing rate of 40 MHz, a peak luminosity of $\mathcal{L} = 10^{34} cm^{-2}s^{-1}$ would be achieved. The corresponding time interval between the single bunches – called bunch spacing – would be 25 ns.

However, these extreme conditions can not be reached without prior studies and commissioning of the machine at lower performance levels. Therefore the first proton-proton (pp) run of the LHC was performed at a centre-of-mass energy of 7 TeV with a lower peak luminosity of 3.6×10^{33}. The total integrated luminosity delivered to the major ATLAS and CMS experiments until the end of 2011 was around 5 fb^{-1} each [64].

In the following "production" year 2012 improvements allowed a slight increase of the centre-of-mass energy to 8 TeV with an instantaneous luminosity peaking at $7.7 \times 10^{33} cm^{-2}s^{-1}$. This allowed to deliver a record sample of more than 23 fb^{-1} to both major experiments at the end of *Run 1* [65], which was followed by the first Long Shutdown (LS1) for 1.5 years.

During LS1 consolidation work on the LHC was carried out, that allowed to accelerate protons to the record energy of 6.5 TeV, which is close to the design. In 2015 the second LHC run (*Run 2*) has begun with a peak luminosity of $5 \times 10^{33} cm^{-2}s^{-1}$ at a bunch spacing of 25 ns, which allowed to keep the number of simultaneous pp-interactions per crossing (pileup) low.

Complementary to the proton only operation mode, the LHC is able to accelerate fully-

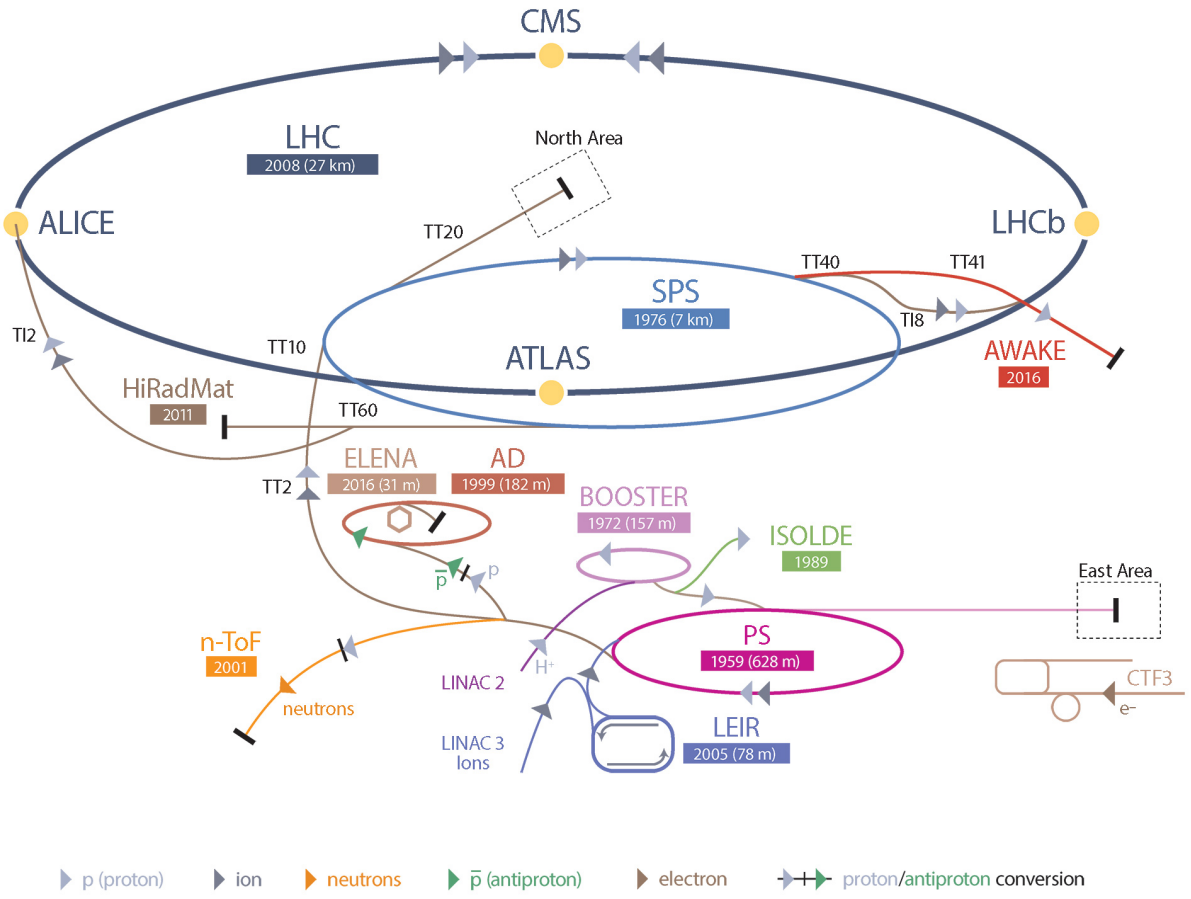


Fig. 2.2: The accelerator complex of CERN [67]: a schematic overview of the operating accelerators, storage rings and beam transfer lines. The location of the major LHC experiments is also indicated.

stripped lead ions $^{208}Pb^{82+}$ to an energy of 2.76 TeV per nucleon. The corresponding centre-of-mass energy of such lead-lead (PbPb) collisions is slightly above 5 TeV/u. This heavy ion programme also includes several proton-lead runs at a centre-of-mass energy of 5 TeV, which are used as reference for the lead-lead ones [66].

Table 2.1 summarizes the typical values of the main beam parameters of the LHC during the different runs in the recent years.

2.1.1 The CERN accelerator complex

Protons injected into the LHC are pre-accelerated by several stages of other accelerators. Figure 2.2 shows the whole CERN accelerator complex including the storage rings, beam transfer lines and major LHC experiments. The journey of the protons starts in a hydrogen bottle at the beginning of the LINAC2 which brings the particles to an energy of 50 MeV over its 30 m lenght. This linear accelerator is used to provide protons to all experiments at CERN, but will be succeeded by the newly commissioned LINAC4 in 2020. The linac beams are injected into

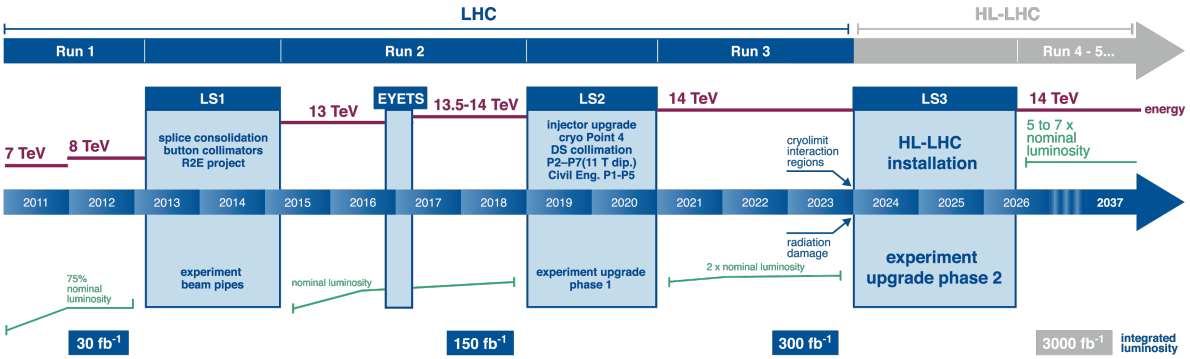


Fig. 2.3: The LHC and HL-LHC operation and upgrade timeline [68]. Projections for the luminosity and pp centre-of-mass are indicated by the green and red lines, respectively.

the PS Booster (PSB) where the particles are further accelerated to 1.4 GeV. With its radius of 25 m the PSB was the smallest synchrotron at CERN until recently.

With the Proton Synchrotron (PS), CERN is still operating its oldest synchrotron nowadays as the third stage of the LHC acceleration chain. The 628 m long ring boosts the protons to an energy of 25 GeV. In addition to protons, it has been used to accelerate alpha particles (helium nuclei), oxygen and sulphur nuclei, electrons, positrons and antiprotons.

The last machine in the LHC pre-acceleration chain is the Super Proton Synchrotron (SPS), where the protons reach the energy of 450 GeV before finally being injected into the LHC. From 1981 to 1984 the SPS operated as a proton-antiproton collider, which resulted in the discovery of the W and Z bosons.

As with most accelerator complexes, the intermediate injectors are not filled continuously, but rather by complex schemes with varying number of bunches and fills per accelerator. With the current injection scheme, the total time for the protons to reach the LHC is 16 minutes [60].

2.1.2 Outlook towards the future of the LHC

The research programme of the current LHC is planned until after 2020, with a target integrated luminosity of about 500 fb^{-1} spread over three run periods. After *Run 1*, where over 29 fb^{-1} were delivered to both, ATLAS and CMS, a major consolidation campaign was undertaken during the LS1. The main work was devoted to improvements of the electrical interconnections of the magnets and installation of additional emergency systems to prevent situations like the incident in 2009. Additionally, 15 dipole and 4 quadrupole magnet modules were replaced completely.

In preparation of *Run 2* the magnets were trained to allow higher magnetic fields, hence increasing the energy of the particles. The training follows a delicate procedure: the main dipole circuit of a given sector is ramped up until a quench (a transition to the resistive state) of a single dipole occurs; the Quench Protection System (QPS) immediately extracts the energy from the circuit and the current is ramped down. After sanity checks the training is repeated usually yielding a higher current for the magnets. For the target energy of 6.5 TeV an average of 20 training quenches per sector was required during LS1.

Investigations are currently ongoing to estimate the number of training quenches required to reach the design energy of 7 TeV. An average number of several hundred cycles is predicted from the experience of the last runs. These training are scheduled to be completed by the beginning of 2021, the start of *Run 3* (Fig.2.3).

Given that the LHC ring size is fixed, and hence the proton energy can't be increased much further, the LHC could expand in the direction of the intensity frontier. Such a proposal is considered with the ambitious **High-Luminosity LHC (HL-LHC)** project, which could reach luminosities of more than $10^{35} cm^{-2} s^{-1}$ – a tenfold increase with respect to the LHC design. The main challenge will be the development of new magnet systems to cope with such a high beam intensity of about 2×10^{11} protons per beam.

In addition, the whole pre-accelerator chain will have to be upgraded as well to allow the injection of high intensity beams to the LHC. This process is coordinated by the LHC Injectors Upgrade (LIU) effort, which covers upgrades of the linac, PSB, PS and SPS. The main future linac (LINAC4) is already in the commissioning stage, while for the rest the main upgrades will take place during the second Long Shutdown (LS2) in 2019-2020.

2.1.3 LHC experiments

Like most circular colliders, the LHC has multiple regions where the counter-rotating beams are intersecting, allowing several experiments to collect data in parallel using the same beams. While two IPs of the LEP were reused for smaller detectors, the new general-purpose experiments ATLAS and CMS required larger caverns, which were enlarged during the LHC installation. The numeration of the IPs starts from the one located nearest to the CERN main site and increases clock-wise.

- **Interaction point IP1**

  The first interaction point (IP) hosts the largest particle detector build to date – **A Toroidal LHC Apparatus (ATLAS)** [69], which is a general-purpose detector capable to digest all types of the LHC collisions and perform analysis from QCD and Standard Model high-precision measurements to searches for new physics beyond the SM, as well as the study of the recently discovered Higgs boson [70].

  For this purpose, the A Toroidal LHC Apparatus (ATLAS) experiment utilises a wide variety of detectors: starting from a mix of silicon based inner tracker and a Transition Radiation Tracker (TRT), over a sampling calorimeter system consisting of an inner liquid Argon electromagnetic calorimeter, and an outer steel-tile calorimeter. The extensive muon system covers the whole detector and is reason to its huge dimensions of 25 m diameter with a length of 46 m. The inner sub-detectors are interleaved with a sophisticated magnet system, which includes a solenoid magnet around the IP providing an axial field of 2 T and several coils of the name-giving toroid magnets creating up to 4 T in the vast space around the solenoid.

  Just at 140 meters away from each side of the ATLAS collision point another experiment has placed its detectors: the **Large Hadron Collider forward (LHCf)** [71] is a special-purpose experiment aimed at measuring the energy and multiplicity of neutral pions produced in the very forward region by pp-collisions. This knowledge allows to calibrate hadron interaction models used in the study of ultra-high-energy cosmic rays through atmospheric showers.

- **Interaction point IP2**

  Taking the place of the former L3 experiment at LEP, **A Large Ion Collider Experiment (ALICE)** [72] also reuses the characteristic red octagonal magnet of it. This detector is specially designed to record the high multiplicity and high energy lead-lead collisions of the LHC. Within this extreme environment a state of matter – called the quark-gluon plasma (or soup) – is expected to reproduce the condition of the Universe momentarily after the Big Bang. Hence, it is of high importance to get a handle on such collisions.

  Despite of not being fully hermetic, the ALICE experiment features numerous high precision sub-detectors allowing to cope with those extreme conditions. The Inner Tracking System is complimented by a Time Projection Chamber (TPC) and Transition Radiation Detector. Several calorimeter systems allow the measurement of the particles' energies, while additional particle identification information is provided by Time Of Flight (TOF) and Ring-imaging Cherenkov detector (RICH) units. A muon spectrometer is build in the forward region, which allows to study the decay products of heavy quarkonium states, which are particularly sensitive to quark-gluon plasma.

- **Interaction point IP5**

  The second general-purpose experiment is the **Compact Muon Solenoid (CMS)** [73], located just on the opposite side of the LHC ring – like a metaphor for being the counterpart to the ATLAS experiment. Indeed, the physics programmes of both experiments are very alike, while the detectors differ in a great way. Since this experiments' data is used for this thesis, a more detailed description will be given in the next section (cf. Sec. 2.2).

  The **TOtal Elastic and diffractive cross section Measurement (TOTEM)** [74] experiment is not only sharing the same LHC collision point than CMS, but is also partially placed just within the bigger detector. As stated in its name, the primary goal of this collaboration is the measurement of the total cross section, elastic scattering, and diffractive processes in the pp collisions. These results are used as a reference for the pp-interaction measurements at the other experiments.

  In addition to the tracking telescopes within the CMS detector volume, there are several Roman pot stations installed at distances of 147 m and 220 m on both sides of IP5.

- **Interaction point IP8**

  The last of the four biggest LHC experiments is the so called LHCb, or **LHC beauty**, which specializes on measurements related to particles containing the b-quark [75]. In particular those are the decays of B-mesons, which in some channels lead to direct hints of CP violation. These precision measurements are also highly sensitive to contributions from new physics beyond the SM. The HERAb experiment at the HERA collider of DESY in Hamburg can be considered as a predecessor of LHCb in may ways.

  The layout of the detector is perfectly aligned to the physics it is built to study: since b-hadrons are predominantly produced in the forward region, i.e. along the proton beam axis, a detector can be restricted to the very same region. Therefore the LHCb represents a typical single-arm forward spectrometer, with a large variety of sub-detectors placed along the beam-pipe up to 20 m from the IP.

  Most notable is the Vertex Locator (VELO) – an array of half-moon shaped silicon detectors reaching up to 7 mm close to the beam-pipe. This proximity to the interaction point

allows to detect the secondary vertices of the decaying B-hadrons produced in the collision. Another consequence is an enormous risk to the sensors due to stray radiation, and therefore the VELO modules are retracted from the beampipe until the beams are considered stable by the LHC. The tracking system makes use of a 4 T normal conducting dipole magnet located between the outer tracker stations.

Finally, the **Monopole & Exotics Detector At the LHC (MoEDAL)** is a small experiment dedicated to the direct search for the magnetic monopole, or Dyon, and other highly ionizing stable (or pseudo-stable) massive particles [76]. It's nuclear track detectors are placed around the VELO of LHCb, and would experience a characteristic damage due to highly ionizing particles.

Unlike the ATLAS and CMS experiments, ALICE and LHCb can not sustain the same huge instantaneous luminosities and vertex multiplicities provided by the LHC. For ALICE the limiting factor is the TPC which has a long dead-time, and accumulates tracks from consecutive bunch crossings (out of time pileup). The LHCb acceptance implies a large particle flux through the detector, which drives the data acquisition to its limits. Additionally, the assignment of tracks to specific interaction vertices also becomes problematic at such angles.

Therefore, a so-called luminosity levelling is performed, where the beams are automatically separated when the target luminosity is reached. This way the experiments can operate with a comfortable and stable instantaneous luminosity of around $\mathcal{L} = 10^{27} cm^{-2} s^{-1}$ for ALICE and $\mathcal{L} = 10^{32} cm^{-2} s^{-1}$ for LHCb.

2.2 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) experiment [73] is located next to the french village of Cessy, on the opposite side of the main CERN premises. The cavern is about 100 m deep underground at the *Point 5* of the LHC ring. As a multi-purpose experiment, the detector covers most of the solid angle around the interaction point and has an onion-shaped structure with the central component being the superconducting 4 T solenoid magnet (Fig. 2.4).

With its maximum length of 29 m and overall diameter of 15 m CMS is almost twice as small as the ATLAS detector, but in terms of weight is it double as heavy: 14.000 tonnes. This makes CMS the heaviest particle detector at colliders worldwide, while also explains the "compact" in its name. Finally, the "muon" part of the names stems from the extensive muon system surrounding the solenoid, which makes up more than 80% of the detector volume.

The initial Letter of Intent for the CMS detector was published back in 1992 [77], while the construction of the individual components commenced soon afterwards. Building on the experience from the L3 experiment, it was decided to split the whole detector in several major parts – five barrel "wheels" and six "endcaps" – so that each of them could be constructed overground and lifted into the cavern after completion. Nowadays the CMS Collaboration involves over 4000 scientists, engineers and students from 192 institutes of 43 countries all around the world [78].

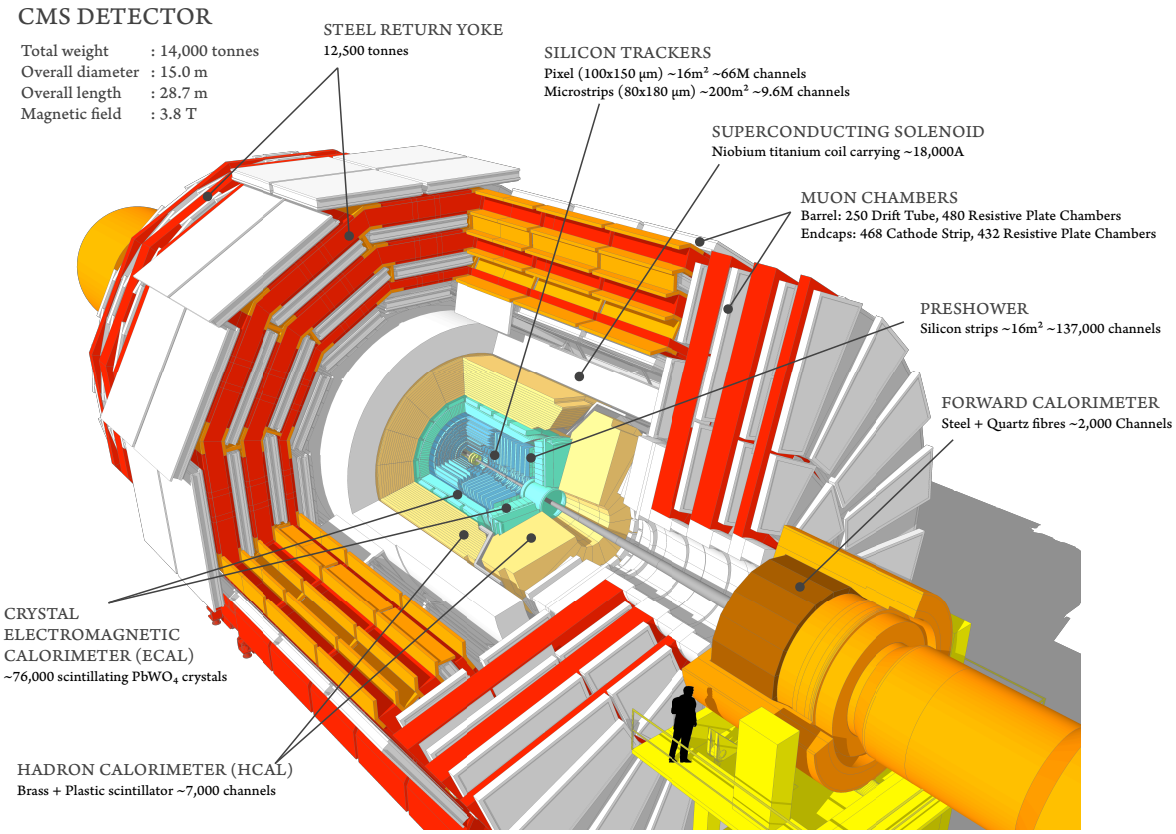


Fig. 2.4: Schematic view of the CMS detector with a cut-out quadrant [79]. The individual subdetectors are shown with some highlights pointed out.

The principal design of the CMS detector is as follows: within the 6 m wide solenoid magnet the inner tracking system is surrounding the very proximity of the interaction point; thereafter the electromagnetic and hadron calorimeters fill the space around the tracker. Outside of the magnet an extensive system of muon trackers is placed that expands up to 7 additional meters above the coil, and an extra 12 m from either side of the barrel. Additional forward calorimeters are placed along the beam-pipe just after the muon detectors.

Coordinates and kinematic variables

The coordinate system used within the CMS experiment at is the following: the middle of the detector is taken as the origin of the right-handed coordinates. In Cartesian coordinates the x-axis points towards the center of the LHC, which is almost straight to South, while the y-axis is directed towards the sky, i.e. the zenith. Therefore, the z-axis coincides with the beam axis, following the direction of the counter-clockwise rotating beam.

Given the shape of the detector, which is almost symmetric with respect of the beam axis, a cylindrical coordinate system is also used often. Naturally, the same z-axis is taken, while the radial distance r and azimuthal angle φ are defined within the xy plane, which marks the middle of the CMS detector. While $\varphi = 0$ corresponds to the x-axis, the polar angle θ is measured from the z-axis.

Another widely used variable is the pseudorapidity η, which characterizes the boost of the particles:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.3}$$

For massless particles, or in the relativistic limit it converges with the rapidity:

$$y = \ln\sqrt{\frac{E + p_z}{E - p_z}} \tag{2.4}$$

The advantage of the pseudorapidity over the polar angle is that the difference in rapidity is Lorentz invariant under boosts along the longitudinal axis, which is the beam axis. Similarly, an invariant spacial distance between two Lorentz vectors can be constructed as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} \tag{2.5}$$

Due to the compositeness of the proton, the proton-proton collisions are actually interactions of the corresponding partons. Therefore, a single parton interaction must not necessarily be energy symmetric, i.e. the parton energies might differ substantially. De facto this renders the total center-of-mass energy an unpredictable quantity for each event. Nevertheless, since the incoming partons are aligned along the beam axis, the total momentum in the transverse plane is fixed to be zero, which has to be followed also by the collision remnants given the momentum conservation. This makes the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ of the particles an extremely useful characteristic.

Additionally, the measured imbalance of the total p_T sum of all particles in a collisions – called missing transverse energy $\not{E}_T$ – provides a handle on the particles, that don't leave any trace in the detectors, such as neutrinos or any neutral weakly interacting particles predicted by BSM theories.

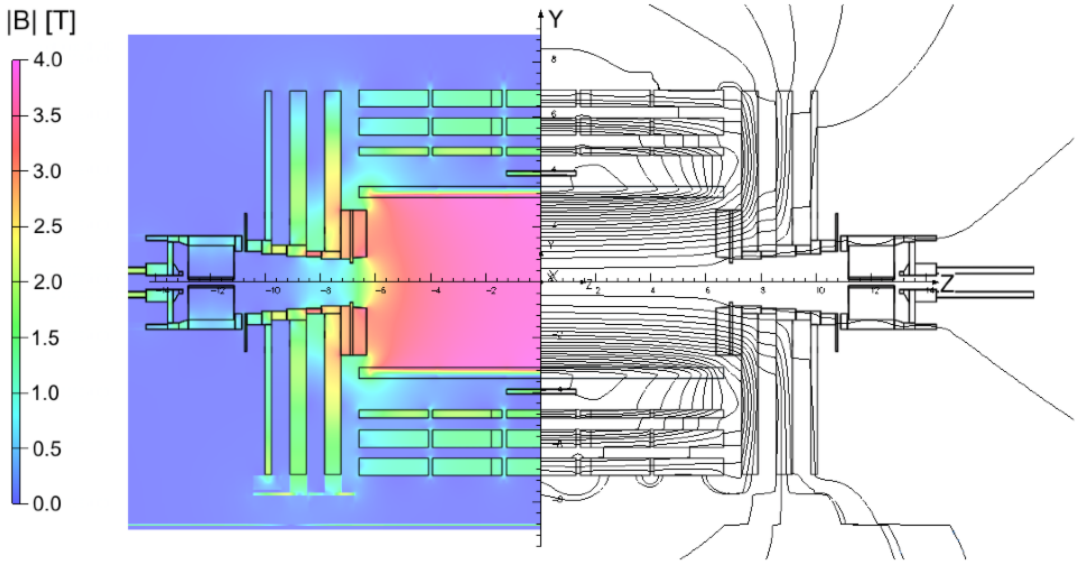


Fig. 2.5: Magnetic field strength (left) and field lines (right) in a longitudinal section of CMS as predicted at a central magnetic flux density of 3.8 T. The magnetic flux changes between the lines by 6 Wb. From Ref. [80].

2.2.1 Magnet

The important role of the solenoid magnet is already stressed out in the name of the CMS experiment. Alike, it is the central element of the whole detector layout. It is designed to host the tracker and calorimeter systems providing a uniform magnetic field of 4 T within its volume. The muon system surrounds the magnet within a steel flux-return yoke.

The magnet free bore is 12.5 m long with a diameter of 6.3 m and designed to reach a 4 T axial magnetic field. It is placed within a cryostat ensuring the necessary operating temperature of 4.5 K. The winding is composed of NbTi superconducting cables. Due to the large required current of 41.7 MA per turn, it was decided to use four layers of winding, instead of the usual one or two. In total, the cold mass weights 220 t and the stored energy-to-mass ratio of 11.6 KJ/kg is the largest up to date.

Since this magnet has a unique design and its behaviour and ageing is not yet well understood, the CMS collaboration decided to operate it at a slightly lower field strength of 3.8 T. The corresponding magnetic field strength and lines from a simulation are shown in Fig.2.5.

2.2.2 Inner Tracker

The innermost part of the CMS experiment is comprised of the inner tracking system, whose main purpose is the identification and measurement of tracks stemming from charged particles originating from the interaction vertices. Due to the extremely large collision rates and high particle multiplicities, this detector system has one of the most stringent requirements in terms of detection efficiency, spatial resolution and radiation safety.

To cope with this challenging environment, the tracking system is split into two sub-detectors: the pixel detector and the silicon strip tracker (Fig.2.6) [73].

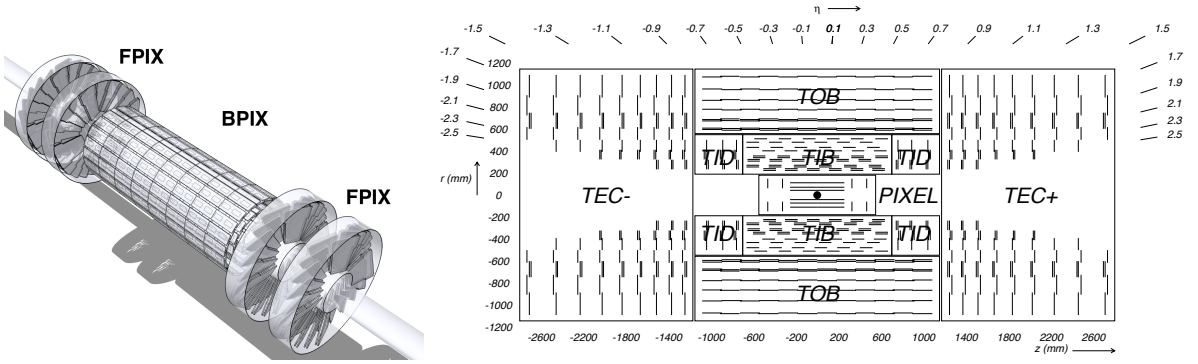


Fig. 2.6: The inner silicon tracking system of CMS. *Left*: the pixel barrel (BPix) and forward (FPix) detectors depicted by a 3D model [81]. *Right*: schematic view of the tracker in the r-z plane, with the silicon strip tracker segments labelled individually. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits [73].

Pixel The Silicon Pixel detector is situated just around the interaction region and effectively covers a pseudorapidity range of $|\eta| < 2.5$. Silicon sensors with a pixel size of $100 \times 150\,\mu\text{m}^2$ are arranged in 3 barrel layers (BPix) and 2 endcap discs (FPix) on either side of the IP (Fig. 2.6, left). The barrel layers are placed at radii of 4.4, 7.3 and 10.2 cm, while the FPix starts at 6 cm from the beampipe center up to 15 cm. This arrangement allows to provide three tracking points over most of the η range. Due to the alignment of the sensors relative to the vertex and magnetic field, a final spatial resolution is achieved in the range of 15–20 μm. In total, the pixel detector covers an area of about 1 m^2 and has 66 million pixels.

The precise measurements of hits in the $r - \varphi$ and z coordinates are required to seed the further reconstruction of tracks (cf. Sec.4), and also to determine secondary vertices of decaying particles like B-mesons and τ leptons. Prompt electrons can be separated from converted photons based on the hit pattern. Additionally, the sensor occupancy information allows to precisely determine the luminosity of the collisions using the Pixel Cluster Counting (PCC) method [82].

Strips The Silicon Strip tracker surrounds the pixel detector providing track measurements in the same pseudorapidity coverage (Fig. 2.6, right). It occupies the radial region between 20 and 116 cm, while extending in z between ± 282 cm. The increased distance to the IP yields a lower flux of particles, and thus silicon micro-strip sensors can be used instead of pixels. There are three different subsystems with varying strip pitches and sensor thickness depending on the location of the modules.

Innermost are the Tracker Inner Barrel (TIB) and Tracker Inner Disk (TID) consisting of four barrel layers and three disks on either side extending to a radius of 55 cm. While having a varying strip pitch from 80 to 120 μm and sensor thickness of 320 μm, the single point resolution reaches from 23 to 35 μm. Up to four r-φ measurements are provided by the TIB and TID.

Enclosing the inner system is the Tracker Outer Barrel (TOB) with an outer radius of

116 cm and length of 236 cm corresponding to TIB/TID. 500 μm thick sensors are used with a wider pitch of 122-183 μm yielding a resolution of 35 to 53 μm for the six r-φ measurements.

The forward coverage is provided by the Tracker Endcaps (TEC) lying within a radius of 22.5 cm to 113.5 cm and 124 cm $< |z| <$ 282 cm. Each endcap comprises nine disks with up to seven rings covered by radial micro-strip detectors with a varying sensor thickness (320-500 μm). With an average pitch of 97 μm to 184 μm the detector provides up to nine φ measurements.

Several layers and rings of all those subsystems also include a second micro-strip sensor in addition to the aforementioned (Fig. 2.6, double lines). These are mounted back-to-back with a stereo angle of 100 $m$rad to allow the measurement of the second coordinate (r in the rings and z in the barrel). A single-point resolution of up to 230 μm is achieved in the TIB. This layout allows to record up to four two-dimensional measurements of the nine hits in the tracker for an acceptance of $|\eta| < 2.4$.

The whole silicon strip tracker has a total of 9.3 million strips and 198 m^2 of active area, being the largest silicon detector worldwide.

The total material budget of the whole inner tracking system is between 0.4 and 1.8 radiation lenghts [1] X_0 over the covered η range.

An important method to prevent silicon detectors from radiation defects is the continuous cooling during its operation. By design, the whole tracker is supposed to be cooled down by dry gas to -20 C°, but for the *Run 1* a working temperature of only 4 C° could be achieved. During the LS1 maintenance period, the tracker was refurbished, including it's cooling system and thermal isolation, which allowed a step towards the designed operating conditions. Starting from *Run 2* the pixel detector has been cooled down to -10 C°, while the rest of the tracker operated at -15 C°. This will allow for an efficient operation until LS2, when the pixel detector will be replaced during the CMS Phase I Upgrade (cf. Sec.2.2.8).

2.2.3 Electromagnetic Calorimeter

The inner tracking system is followed by a hermetic homogeneous electromagnetic calorimeter (ECAL), which consists of almost 70 000 lead tungstate ($PbWO_4$) crystals surrounding the strip tracker [83]. Its main purpose is to accurately measure the energy of electrons and photons coming from the tracking system, while also registering the start of hadronic showers. The ECAL consists of the ECAL barrel (EB), covering a range of $|\eta| < 1.479$ and an ECAL endcap (EE) on each side, which extends the range further until $|\eta| = 3.0$ (Fig. 2.7). A preshower detector (ES) is placed additionally in front of the EE.

The main reason for the choice of $PbWO_4$ crystals for the ECAL was their fast light response, radiation hardness and high density, which allows for a fine granularity calorimeter: together with the rather small radiation length of 0.89 cm and Molière radius of 2.2 cm, about 80% of the scintillating light is emitted within the LHC bunch crossing time of 25 ns. The detection of the emitted blue-green light is performed by avalanche photodiodes (APD) in the barrel part, and vacuum phototriodes (VPT) in the endcap. At the operating temperature of 18 C° those detectors collect around 4.5 photoelectrons per MeV of deposited energy.

[1] The radiation length X_0 is the mean distance for which a high-energy electron loses $1/e$ of its energy by bremsstrahlung.

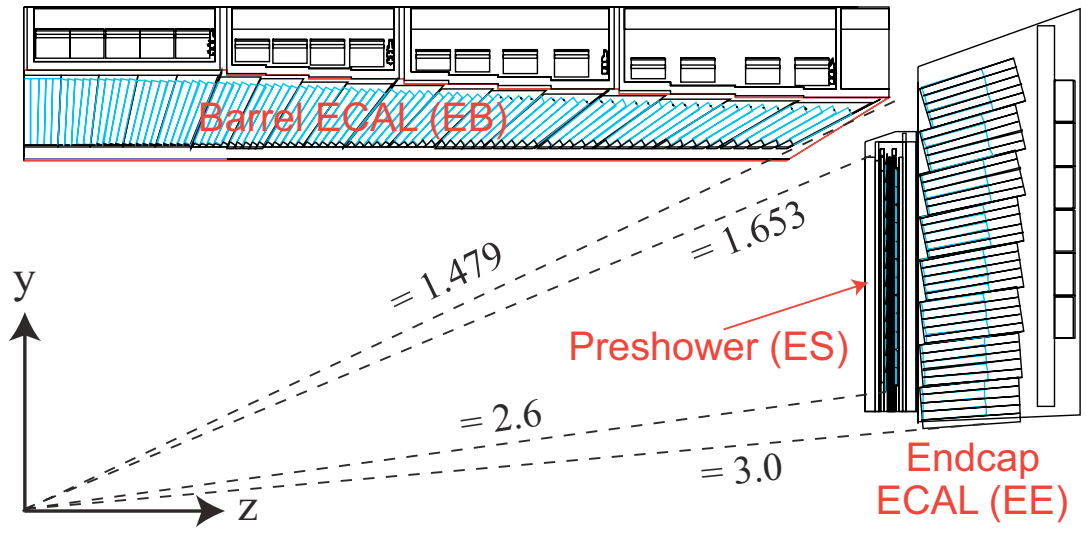


Fig. 2.7: Layout of an ECAL quarter with its subsystems indicated [83], their geometric acceptances are shown with the dashed lines. The $PbWO_4$ crystals are highlighted in blue.

The granularity in the barrel and endcaps varies due to increasing flux towards the beam axis. In the EB there are 360 ϕ and 2×85 η segments, resulting in 61 200 crystals. To avoid cracks aligned with particle tracks, the crystals are not pointing directly towards the interaction region, but have a slight angle of 3° in the ϕ and η projections. A single crystal occupies approximately 0.0174×0.0174 in η-ϕ, or $22 \times 22\,\mathrm{mm}^2$ with its front face. To avoid gaps between the adjacent crystals, they have a tapered shape resulting in a slightly wider back face of $26 \times 26\,\mathrm{mm}^2$. With a length of 230 mm, corresponding to 25.8 X_0, the electromagnetic showers of most energies are fully contained in the ECAL.

For the ECAL endcaps identically shaped crystals are grouped in units of five by five, so-called supercrystals.The crystals are pointing at a focus 1.3 m away from theIP. The area facing towards the IP has a size of $28.62 \times 28.62\,\mathrm{mm}^2$, becoming $30 \times 30\,\mathrm{mm}^2$ in the back, while the length of the crystals is 220 mm (24.7 X_0). A total of 7 324 crystals are installed in the EE. Combined, the crystals of EB and EE occupy a volume of $11\,\mathrm{m}^3$ and weigh more than 90 tons.

Just before the ECAL endcaps, on their inner surface, 20 cm thick preshower detectors are attached, which cover a fiducial volume of $1.653 < |\eta| < 2.6$. The primary purpose of the ECAL preshower (ES) is to identify neutral pions which enter the endcaps, while also helping to separate electrons from minimum ionizing particles (MIP) and increasing the granularity of the position measurement of photons and electrons.

The ES comprises two layers of a lead absorber followed by a silicon sensor plane, which make up around 1 X_0. Since 2 X_0 are already achieved within the tracker, this causes most of the single-incident photons to start showering between the sensors.

2.2.4 Hadron Calorimeter

The next step in the particle detection chain is performed by the hadron calorimeter (HCAL) system [84]. Its principal aim is the measurement of hadronic jets, especially of the neutral

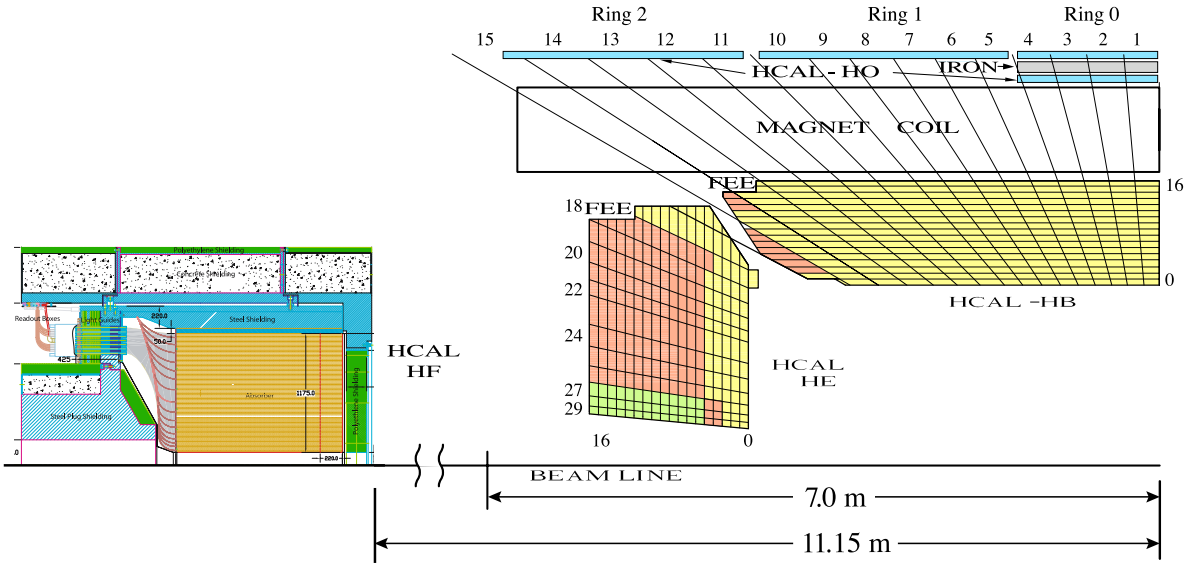


Fig. 2.8: Schematic view of a quarter of the CMS HCAL subsystems. For the central HCAL detectors (right) the tiles with the same color are read out by a single channel [85]. The HF (left) segmentation is explained in the text [86].

constituents that cannot be detected in the previous systems. Therefore, the destructive power of a huge amount of material is taken use of.

The size of the magnet cryostat allows for the placement of the major central part directly after the ECAL without any dead material in between. As for the ECAL there are two components placed inside the solenoid: the HCAL barrel (HB) extends the EB in the barrel part, while the EE is continued by the HCAL endcap (HE). The lower thickness[2] of the HB is compensated by the outer hadron calorimeter (HO), which is placed just outside the magnet and acts as a tail-catcher.

In addition to the central part of the HCAL the coverage is extended by a very forward calorimeter (HF), which is placed outside the CMS endcaps at a distance of $|z| = \pm 11.2\,\mathrm{m}$ from the collision vertex. A close to natural representation of the HCAL subsystems is given in figure 2.8.

HB and HE In contrast to the ECAL, the central HCAL is based on a sampling technique with heavy absorber and plastic scintillator layers interleaved. The barrel part covers the pseudorapidity range $|\eta| < 1.3$ and is fixed directly inside the cryostat, while the endcaps extend further ($1.3 < |\eta| < 3$) and are mounted on the CMS endcap iron yokes.

The HB is split into 36 identical azimuthal (in ϕ) wedges aligned parallel to the beam axis, where 18 wedges each form a half-barrel (HB+ and HB-). Each wedge is further segmented into four parts, yielding a total of 72 ϕ sectors per half barrel. There are 15 layers of flat brass absorber surrounded by the innermost and outermost plates that are made of steel for structural reasons. The plastic scintillator trays placed between the absorbers are divided into 16 η sectors, resulting in a fine granularity of 0.087×0.087 in

[2] In terms of travelled path from the IP

the η-ϕ plane. This (smallest) segment is called a tile, and there are about 35 000 tiles in the whole HB. A stack of scintillator tiles from different layers, but same η-ϕ coordinates, is called an HCAL "projective" tower, since it points towards the IP.

Each tile is inlaid with a wavelength shifting fibre (WLS), which is supposed to collect the scintillating light, and afterwards connects to a clear fibre, which transmits the optical signal output of the tray. The light is further led to the read-out electronics, where the signal is digitized. Within this front-end electronics (FEE) the fibres are bundled by an optical decoding unit (ODU) to shine on a single hybrid photodiode (HPD).

The first 15 η towers have a single longitudinal read-out, while the last two are split into two separate segments as shown by the color coding in Figure 2.8.

For the HCAL endcap, a very similar approach is taken in terms of material, scintillator segmentation and read-out. The HE has a tapered shape to interlock with the barrel part, as well as to overlap with the last HB tower. It consists of up to 18 brass layers, which are matching the ϕ segmentation of the barrel. Thus, the wedges have an angular size of 20°.

The light is collected by 19 scintillator layers interleaving the absorber. For $1.3 < |\eta| < 1.74$ the η-ϕ tower matches the one in the barrel, while for the outer part the ϕ segmentation is reduced by a factor of two to 10°. Like for HB, the different towers have varying read-out segmentation with up to three separate depths (see Fig 2.8).

HO As already mentioned before, the barrel part of HCAL is complemented with a tail catcher just outside the magnet cryostat. This outer hadron calorimeter (HO) represents an additional layer of plastic scintillator in the outer barrel region ($0.4 < |\eta| < 1.26$), and two layers in the central part. The whole magnet coil is used as an absorber, while for the central region an 19.5 cm thick additional steel plate is placed between the scintillators.

In contrast to HB, the outer hadron calorimeter (HO) is not mounted on the cryostat, but on the rings of the barrel return yoke (YB), also referred to as wheels. In total, there are five of those rings, with the central one holding the magnet cryostat including the inner detectors. The wheels are segmented in 12 sectors and have similar dimensions: around 2.5 m wide along the z-axis and with an inner and outer radius of 4 and 8 m, correspondingly. Since the YB mainly hosts the muon system, it will be described in more detail in the next section.

Following the dimensions of the barrel yoke, the HO scintillator panels are 2.5 m long and 2.5 m wide. In the central ring the two layers are placed at radial distances of 3.82 m and 4.07 m, respectively, while the single layer of the other rings is placed at 4.07 m from the beam axis. Combined with the previous sub-detectors, the total barrel depth in terms of interaction lengths[3] reaches at least $10\lambda_I$, as shown in Fig. 2.9 (left).

The production, transmission and read-out of the light signal is similar to the HB and HE. Initially, the light was also detected by HPDs, but during LS1 they were replaced by silicon photomultipliers (SiPM) to improve the performance of HO. A detailed description of the upgrade will follow in the next chapter 3.

[3]The nuclear interaction length λ_I is the mean path after which the number of relativistic charged particles is reduced by a factor of e.

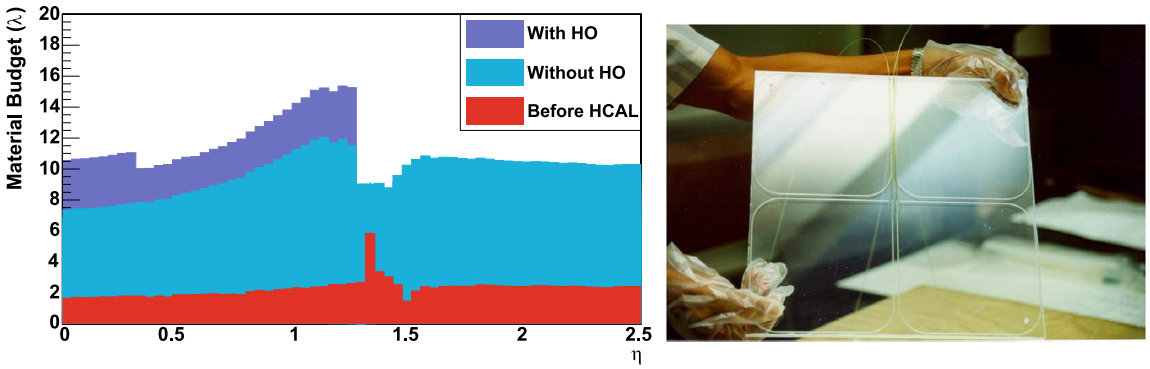


Fig. 2.9: *Left*: material budget (in units of nuclear interaction length) with and without the HO, as well as before the HCAL [87]. *Right*: an HO scintillator tile with four embedded fibres ("sigma grooves"); this group corresponds to a single tower in the HB and is read out together by a single channel.

HF The central HCAL detector is complemented with a forward calorimeter (HF) system, which covers the range of $2.85 < |\eta| < 5.19$, and comprises two detector blocks attached to the yoke endcaps from each side (Fig. 2.8, left).

Given the extreme particle flux at such high pseudorapidities, a radiation-hard technology was chosen for HF: a steel absorber with the external size of 1.65 m along the z-axis, and extending from 12.5 cm to 1.3 m in the radial distance, is interleaved with hundreds of quartz-fibres as active material. Passing charged particles radiate Cherenkov light, which is then transmitted to photo-detectors behind the absorber.

There are two types of fibres: the long ones extend from the front to the very back of the absorber (165 cm $\approx 10\lambda_I$), while the short fibres start at a depth of 22 cm and are read out separately. The fibres are placed alternately with a separation of 5 mm. This configuration allows to identify showers stemming from electrons and photons, since they are usually absorbed within the first 22 cm. Hadronic showers, in contrast, spread through the full depth, and therefore leave similar signals in both fibre types.

In the transverse plane, each of the two HF modules is subdivided into 18 wedges with an angular size of 20°. The fibres are bundled into towers of 0.175×0.1754 η-ϕ, which are then read out individually by phototubes.

There are two calorimeter systems placed even further in the forward region: the Centauro And Strange Object Research Calorimeter (CASTOR) and the Zero-Degree Calorimeter (ZDC). CASTOR is a single Cherenkov sampling calorimeter that covers the pseudorapidity range of $-6.6 < \eta < -5.2$. It is designed to search for new particles, but also provides a unique possibility to study processes in this very forward region [88]. The ZDC has two calorimeter stations located at $z = \pm 140$ m, which also allows for a wide physics programme in proton or ion collisions [89].

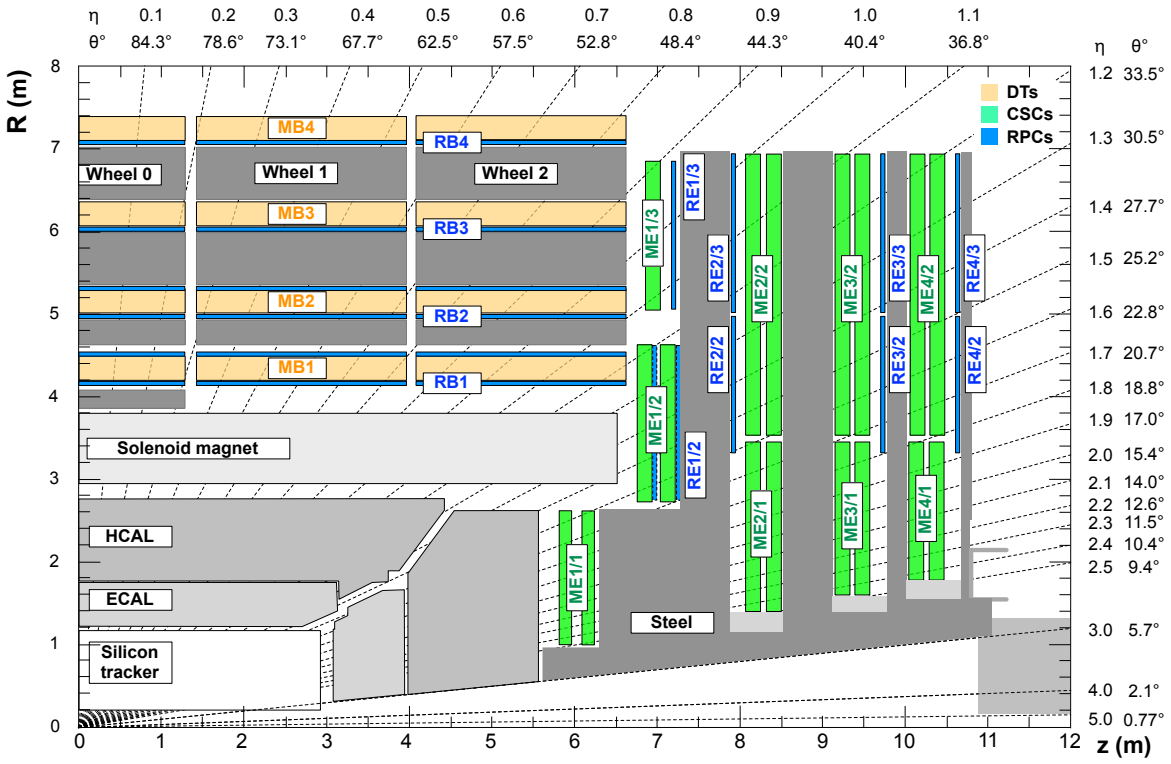


Fig. 2.10: Layout of a quadrant of the CMS Muon system after the LS1 [90]. The barrel wheels host the DT and RPC chambers in four stations parallel to the beam-pipe. Another four stations of CSC and RPC provide coverage in the endcaps.

2.2.5 Muon System

Highlighted already in the middle name of the experiment, the muon system is of a high importance for CMS. Provided the large amount of material within the cryostat, mostly muons manage to escape the inner part of the detector. To ensure further stopping of particles other than muons, as well as to return the field flux, the magnet is surrounded by the steel flux-return yoke [73]. The total material depth a particle has to travel through to reach the end of the muon system is at least 20 λ_I for low pseudorapidities.

The three main functions of the muon system are identification and momentum measurement, as well as triggering of muons. To achieve these goals, three types of gaseous particle detectors are used in an optimal combination to cover the wide η range up to 2.5. The barrel part is covered with drift tubes (DT), while the endcap is equipped with cathod strip chambers (CSC). For triggering fast resistive plate chambers (RPC) are used throughout the detector. In total there are about 25 000 m^2 of detection planes.

As shortly mentioned in Sec. 2.2.4, the return yoke is constructed of five barrel wheels (YB), which have a dodecagonal shape, i.e. consist of 12 equal sectors. The barrel is enclosed by the four endcap return yoke (YE) discs from each side, which are numbered in a similar fashion as the YBs.

DT Given the low muon rate and moderate magnetic field in the central region, the barrel muon system is using drift chambers with rectangular drift cells. These cover the region of $|\eta| < 1.2$ as pictured in Fig. 2.10. Overall four stations per sector are interleaving with the layers of the yoke plates. The first three stations are equipped with 8 chambers that measure the muon coordinate in the r-ϕ plane, and 4 chambers detecting the z coordinate of the tracks. The last station lacks the z measurement chamber. In total there are 60 drift chambers in the first three layers, and 70 in the last.

The tubes within a station are grouped into SuperLayers (SL) and are staggered by half a tube to improve the detection efficiency and allow for a more precise timing measurement. A honeycomb structure is placed between the SLs to provide a lever arm for the track reconstruction. Each tube spans for the whole length of the chamber, thus the wires are around 2.5 m long. In total there are 172 000 sensitive wires in the DT.

CSC Outside the barrel yoke the magnetic field becomes larger and non-uniform (cf. Fig.2.5). In addition, the particle flux increases towards the forward region. Therefore, the muon detectors in the two endcaps have to be fast and radiation-hard. This challenge is confronted with finely segmented multiwire proportional cathod strip chambers, which cover the pseudorapidity range of $0.9 < |\eta| < 2.4$.

Similarly to the DT, the CSCs have four stations placed between the yoke plates. The r-ϕ coordinates are provided by cathode strips extending radially outward, while perpendicular anode wires are also read out to extract the η, as well as the crossing time of the muon.

During the *Run 1*, the system consisted of 468 CSCs, but during LS1 it was upgraded, reaching the baseline design of 540 chambers. The grouping and labelling of the CSCs can be obtained from Fig.2.10. All chambers are trapezoidal and cover either a 10° or 20° angle in ϕ. Most chambers additionally overlap in ϕ to provide a continuous coverage.

The full area covered by the CSCs alone is about $5000\,\mathrm{m}^2$ with a gas volume of $50\,\mathrm{m}^3$. A total of around 2 million wires are installed, of which approximately 220 000 cathode strips and 180 000 anode wires are read out.

RPC The precise DT and CSC muon chambers are also able to trigger on muons with a good efficiency and purity, but the uncertainty on the performance required to have a backup solution. Therefore, a complementary dedicated trigger system of resistive plate chambers (RPCs), that cover a similar η range in the barrel and endcaps, was installed. The RPCs provide a fast and independent trigger with a sharp p_T threshold independently of the two other muon systems.

There are 6 layers of RPCs installed in the barrel muon system. The first two DT stations have RPCs attached from the inner and outer side to allow the measurement of low p_T muons. For the outer two stations, only a single RPC is placed in front of the drift chamber.

In the endcaps, the RPC detectors are placed in parallel to the CSCs in the four YE stations. Like for the latter, the RPC coverage was extended during the LS1 to reach up to an η of 1.9. Before, only the first three discs were equipped with RPCs.

The principal technology used for all chambers is a double-gap module with two gaps operated in avalanche mode with common read-out strips in between.

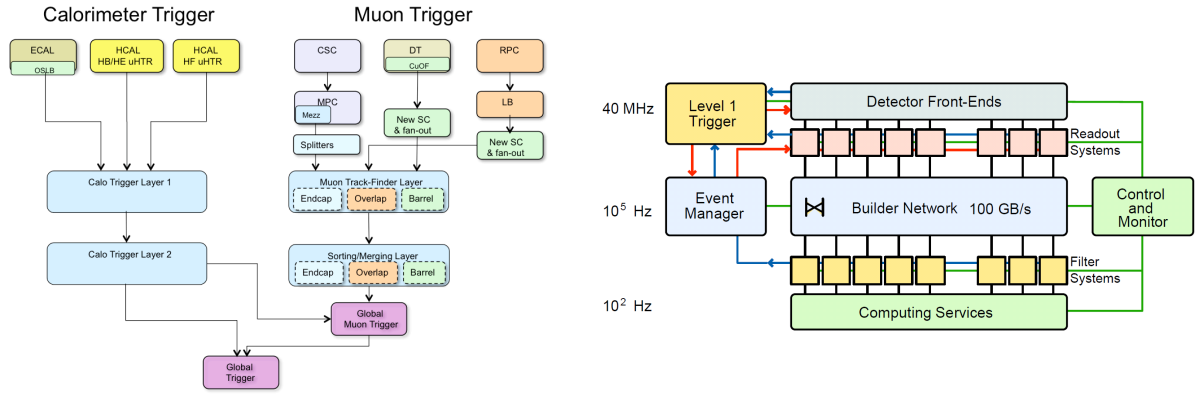


Fig. 2.11: Dataflow of the upgraded Level-1 Trigger [93] (left). The general architecture of the CMS Data Acquisition (DAQ) system [73] during *Run 1* (right) shows the key building blocks for a single slice.

2.2.6 Trigger and Data Acquisition Systems

Apart from the challenging detection of the particles, the next big hurdle is the processing and filtering of the huge amounts of data provided by the particle detectors. During nominal operation the LHC delivers collisions every 25 ns, i.e. with a frequency of 40 MHz. With rising instantaneous luminosity the number of interacting protons also grows, thus increasing the number of particles in the detector.

The total amount of data from the millions of channels is then of the order of 1 MByte per single bunch crossing. The Trigger and Data Acquisition Systems (TriDAS) are laid out to cope with this extreme flow of data [91, 92].

The main restriction in handling such amounts of data is the ability to write it to a permanent storage. The typical limit is of around a hundred of events per second, thus the reduction factor from the collisions to disk is 10^6. CMS adopts a two-stage filtering mechanism with a hardware based Level-1 (L1) Trigger, and an High-Level Trigger (HLT) running on a big computer farm. The design output rate of the L1 Trigger, which henceforth is the input for the HLT, is 100 kHz.

L1 Trigger The first stage of the CMS trigger has to take decisions very fast, and thus relies on a logic performed on a low hardware level. When possible, field-programmable gate arrays (FPGA) are used to allow for a flexible adjustment to the operational conditions. For certain components more simple and reliable application-specific integrated circuits (ASIC) and programmable memory look up tables (LUT) are used.

The inputs are provided by the calorimeter and muon systems in a coarsely segmented format to speed up the calculations. A schematic view of the dataflow within the L1 Trigger is shown in Fig. 2.11 (left).

The L1 Trigger system comprises local, regional and global components which are running in a sequential order, and allow for the muon and calorimeter parts to exchange relevant information with each other. For example, the occupancy from the ECAL or HCAL can be used in the identification of isolated muons.

The exact interaction of the systems has been upgraded in the year-end technical stop 2015/206, which is the second stage after first modifications during the LS1. The main

change is the switch to a new-generation hardware as well as the usage of fast optical connections [94]. In the muon calorimeter the three separate systems will not be handled independently, but rather in spatial categories for the barrel, endcap and overlap regions [95].

The full information from the Global Muon and Global Calorimeter triggers is combined and evaluated in the Global Trigger, which ultimately gives an Level-1 accept (L1A) signal to the FEE and DAQ, such that the event can be processed further.

DAQ and HLT The reduced event rate of 100 kHz is then sent to a computer farm on the surface of *Point 5*. Here, the full event information is scrutinized by a sophisticated program running on a single processor. This corresponds to an approximate transfer rate of 200 GByte/s. The current number of CPUs in the HLT farm exceeds 13 000, but with the increase in event size and processing time it will be extended.

A schematic overview of the DAQ system in Fig. 2.11 shows the dataflow within the TriDAS chain for a single slice.

For each event, the individual fragments are assembled by an Event Builder, and then the Event Filter reconstructs the event objects and decides whether it should be kept according to the programmed "trigger rules". These rules have to be fed in within a so-called trigger menu, which stores the settings for various trigger paths.

Depending on the instantaneous luminosity, some trigger paths may be "prescaled", i.e. throttled in a way, that only a fraction of all triggers are accepted. This system allows to control the total rate as well as to keep the trigger path active for studies on parts of the data.

The events are categorized in so-called Primary Datasets (PD) depending on the trigger they fired, e.g. all events that triggered an electron path are stored in the "electron" PD. This allows to spread the IO load on separate storages and transfer nodes. Firstly, the events of the PDs are stored locally in a "RAW" format, before they are transferred to the main Tier-0 at CERN and other Tiers for further reconstruction (cf. Sec.4).

Data Quality Monitoring (DQM)

The quality of the recorded data is constantly monitored on various levels. A dedicated PD selects a small subset of all events for verification by the DQM teams. In the first "online" step, the basic detector properties are checked: channel qualities, timing synchronisation and data integrity. This monitoring is performed in situ in the control centre of CMS at *Point 5*. For the "offline" validation the full reconstruction chain is run, and high-level objects such as momentum spectra and particle multiplicities are inspected by shifters around the world.

The list of "certified" events is then provided to the analyzers in form of a data range list.

2.2.7 Luminosity measurement

The instantaneous luminosity delivered by the LHC is of extreme importance for the operation and physics analysis of the CMS experiment. On one hand, the luminosity information is crucial for the tuning of the beam parameters, which affect the number and quality of the collisions. Otherwise, the DAQ and trigger systems have to adapt their behaviour to the rate and

intensity of the interactions. This can have an effect on the prescales of single trigger paths, as mentioned before.

For the physics analyses, the luminosity integrated over the time of data recording provides a direct connection to the number of events of a certain hard process (cf. Eq.2.1).

The CMS experiment uses several independent detector systems to monitor and measure the instantaneous luminosity at *Point 5* [73, 96]. The generic pixel tracker, DT and HF systems are used together with dedicated luminosity detectors: the Fast Beam Condtions Monitore (BCM1f) [97] and the recently installed Pixel Luminosity Telescope (PLT).

There are two basic modes for the luminosity measurement: the "online" monitoring is performed with the HF, BCM1f and PLT based on basic fast-computable variables. Those systems provide in-situ feedback about the instantaneous luminosity for the operation of the LHC and CMS. The read-out is decoupled from the global CMS DAQ for the sake of speed and reliability.[4]

The more accurate "offline" measurement is obtained from the fully reconstructed information of the remaining detectors. The aforementioned PCC method yields the highest precision result, since the pixel detector is located directly next to the interaction point.

In general, both methods require an a priori calibration of the detectors. Initially based on Monte Carlo (MC) simulations, it is usually improved using dedicated LHC runs. For the *Run 2* in 2015 the final uncertainty on the integrated luminosity of the dataset used in this thesis is 2.7% [96].

2.2.8 Upgrades of CMS

As previously discussed in Sec. 2.1.2, the physics programme of the LHC is planned for at least two decades ahead. The increase in beam energy and intensity will represent a major challenge to the experiments, which will have to upgrade and adapt their detector components accordingly. Similar to the timeline of LHC operation, the CMS experiment foresees two major stages of its detector consolidation.

Phase-I

The first stage of upgrades, called **Phase-I** [98], spans the period from the first to the second long shutdown of the LHC. Since the beam parameters are going to change only moderately in this time interval, the upgrade mainly concerns the maintenance and improvement of current subdetector systems.

The upgrades performed during LS1 where already shortly mentioned in this section, while a more elaborate description of the HO readout electronics upgrade will be given in the next Chapter 3. The HB and HE subsystems will undergo a similar upgrade until the end of LS2, which will allow a finer segmentation of the calorimeter tower readout using SiPMs.

During the same time period, the full inner pixel tracker will be replaced by a new detector with 4 and 3 layers in the barrel and endcap regions, respectively. The muon and calorimeter trigger systems are going to be improved to allow for a more efficient selection of events. In addition, the data transfer and handling infrastructure of most subsystems will be upgraded to modern standards to allow for higher data rates and quality.

[4] Here, a separate DAQ system is used for the HF calorimeter; it runs in parallel to the global DAQ.

Phase-II

The second upgrade campaign, **Phase-II** [99], is foreseen during the High-Luminosity LHC installation. Operational parameters of the HL-LHC will exceed the LHC design resulting in a more challenging environment for the detectors, which have to cope with the increased instantaneous luminosity and pileup interactions. In addition, the radiation ageing of the detectors will become an important factor, since the annual dose of the HL-LHC will correspond to the total LHC operation period with about $300\,\text{fb}^{-1}$.

Given these operational challenges, major detector upgrades and replacements are planned for the Phase-II period. The whole inner tracking system will be replaced with a more granular and radiation tolerant detector with increased coverage of up to $|\eta| < 3.8$. Moreover, special data links will allow to include tracking information already at the L1 Trigger stage. The coverage of the muon system and trigger will also be extended to higher pseudorapidities, complemented with an upgrade of the corresponding front- and back-end electronics. A decrease of the operating temperature of the barrel ECAL will allow for a more stable response of the crystals.

Due to the increased particle flux in the forward regions causing severe radiation damage, the EE and HE endcaps will be replaced by a new high-granularity calorimeter. Multiple layers of absorber will be interleaved with silicon pads and plastic scintillators that will allow to reconstruct the electromagnetic and hadronic shower shapes in great detail providing an increased energy resolution.

In addition to the detector upgrades, the bandwidth of both trigger levels will be increased by a factor of 75 w.r.t. the nominal design to 750 and 7.5 kHz, respectively, using corresponding upgrades of the DAQ system and other detector infrastructure.

Overall, the CMS experiment is envisaged to successfully operate at the LHC and subsequent HL-LHC for about two more decades accumulating at least $3000\,\text{fb}^{-1}$ integrated luminosity of data.

2.3 Event simulation

Nowadays, Monte Carlo (MC) simulation techniques are an essential part of both, experimental and theoretical, particle physics research, since they allow to study close-to-reality events and develop the theoretical frameworks or detector methods accordingly.

In this chapter a short overview of simulation methods used in CMS, and in particular in this analysis, will be given. First, the idea of event generators will be shown together with a list of several major packages used. Afterwards, the CMS simulation workflow will be presented.

2.3.1 Event generators

Due to the composite nature of the proton, the simulation of pp-collisions becomes an extremely difficult task. Figure 2.12 shows a schematic view of a $t\bar{t}H$ event produced in such a collision. The hard scattering itself is shown as a red blob with the solid and dashed lines indicating the resulting particles, which themselves decay further. Independently happening multi-particle interactions are indicated by the violet blob. The initial- and final-state radiation is shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particles are dark green.

Following this complexity, a paradigm is adopted, where the events are generated in a sequence of stages with decreasing energy scale.

ME First, the scattering amplitude is calculated using the matrix element (ME). Here, the proton PDFs are used to sample the initial state partons' momenta, and the final particles are distributed accordingly. Typically, this creates the highest energy particles of the collision.

Decay of unstable heavy particles such as the gauge bosons is using the matrix element similarly.

PS The parton shower (PS) models the initial-state radiation (ISR) and final-state radiation (FSR) through a successive process of parton emission from the highest to lowest scale.

HAD The hadronization finalizes the generation with the formation of stable hadrons, that can be detected in an experiment. This step might also include the simulation of additional multi-particle interactions, which are generally at a lower scale as the hard interaction.

A big variety of event generators capable of performing those different steps exist. The MC simulation programmes used for this analysis are:

PYTHIA is a general-purpose MC event generator, that is able to perform all the above mentioned generation steps [101]. For the hard process only tree-level 2 to 1 and 2 to 2 can be calculated at the leading order (LO), while the higher order is approximated with parton showers using the Lund string model [102].

MADGRAPH allows to compute hard processes with up to ten final particles in the final state using the ME [103]. Radiation of gluons in the initial or final state can be obtained from the ME as well. Starting from MADGRAPH5, next-to-leading order (NLO) calculations are implemented using the AMC@NLO framework.

Since parton showering and hadronization is not implemented in MADGRAPH, it is usually interfaced to a PS generator like PYTHIA. In this case double counting of radiation from the ME and PS steps is possible. Therefore, an energy threshold is assigned to separate the jets to the generators accordingly. The MLM scheme [104] is used to perform the corresponding matching.

POWHEG incorporates a different approach than MADGRAPH to calculate the ME at NLO [105]. This allows to avoid double counting while coupling to parton shower programmes.

For this analysis, the vast majority of MC samples was simulated at LO with the MADGRAPH5_AMC@NLO 2.2.2 generator interfaced to PYTHIA 8.205. More details will be given in Chapter 5.

2.3.2 Detector simulation

After the generation, showering and hadronization of the events, the final state particles have to be propagated through the detectors to obtain the corresponding hardware responses. This process involves the simulation of the interaction of those particles with all the material of the various subsystems of CMS.

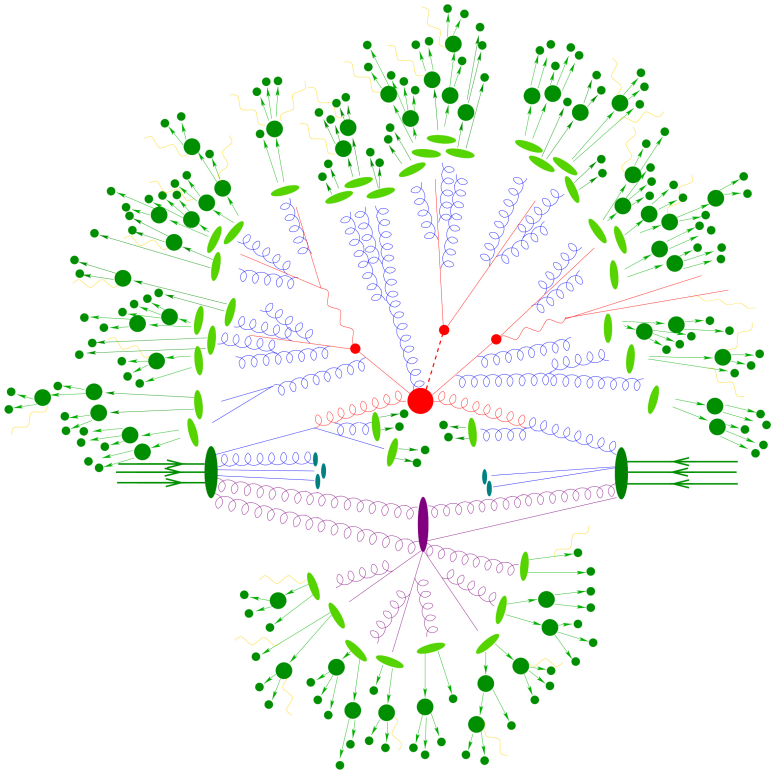


Fig. 2.12: Schematic view of a $t\bar{t}H$ event produced in a pp-collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles [100]. Independently happening multi-particle interactions are indicated by the violet blob. The initial- and final-state radiation is shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particle are dark green.

The main method used by CMS, referred to as *Full Simulation*, makes use of the GEometry ANd Tracking (GEANT4) toolkit [106]. Here, a precise 3D-model of the CMS detector is implemented including not only the sensitive detectors, but also the corresponding infrastructure. Apart from the detailed MC simulation of the passage of the particles though this material, the behaviour of the readout electronics is also modelled.

Naturally, the above detailed simulation requires huge computational resources. In case of large-scale MC production, like in the case of SUSY parameter phase scans, a more time-effective method is needed.

Subsequently, CMS developed a *Fast Simulation* (FastSim) software as an alternative to the full GEANT4-based simulation [107]. The increase in computational speed is achieved through the simplification of various aspects of the detector geometry and particle-matter interactions. The FastSim detector responses are tuned to correspond to the full simulation.

Regardless of the simulation type used, the data output format is designed to match the one coming from the detector DAQ system; this allows to employ the same reconstruction techniques for simulation as for real data. Additional MC-related information such as from the original event generator and detector simulation is included to facilitate various studies.

Chapter 3

Upgrade of the CMS outer hadron calorimeter

3.1 The HO upgrade 47
3.1.1 HO design 48
3.1.2 Silicon photomultipliers 50
3.2 Commissioning 51
3.3 Calibration with cosmic muons 53
3.3.1 Local runs 53
3.3.2 Global runs 59
3.3.3 Muon detection efficiency 63

In this chapter the upgrade of the CMS outer hadron calorimeter (HO) during the first Long Shutdown (LS1) of the LHC is described. First, a more detailed description of the HO together with the motivation for the upgrade is given in Sec. 3.1. Then, Sec. 3.2 will elaborate the upgrade workflow and commissioning of the new hardware. Finally, the first calibration of the HO using cosmic muons will be presented in Sec. 3.3.

3.1 The HO upgrade

A general introduction to the HO tail-catcher was already provided in the previous chapter in Sec. 2.2.4. In the original detector design the read out of wavelength shifted scintillator light was performed by hybrid photodiodes (HPDs). However, during operation in *Run 1*, these photo-detectors proved to be sub-optimal for the operational conditions in HO. The main reasons were:

- discharge caused by the fringe field of the CMS magnet
- low gain and photo detection efficiency
- lower than expected lifetime.

The CMS HCAL group developed a drop-in replacement for the design front-end based on silicon photomultipliers (SiPMs) as photo-sensors [108, 109]. The main advantages of SiPMs are the magnetic field insensitivity, the relatively high photon-detection efficiency, and the high gain. SiPM boards are compact enough to easily fit into the limited space of the existing readout modules. In addition, these detectors can operate at lower voltages, than HPDs.

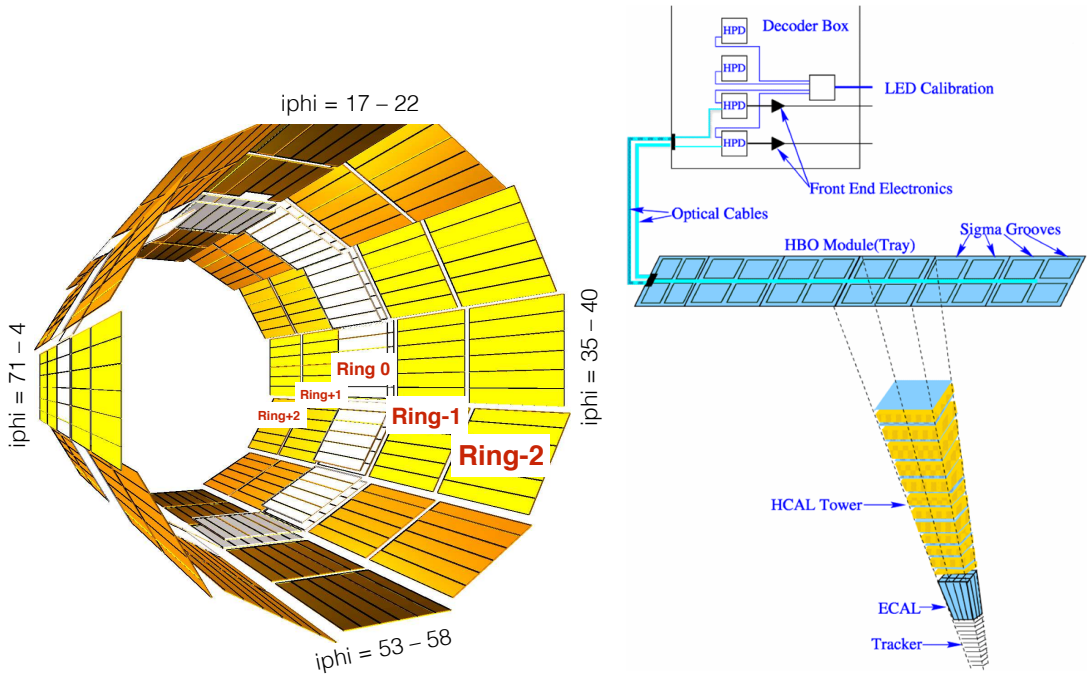


Fig. 3.1: *Left*: layout of the HO scintillator trays. There are five rings with twelve sectors each. The $i\phi$ coordinates are shown for the horizontal and vertical sectors. *Right*: HO scintillator tray and corresponding HB projective tower [87]. The light from the four fibres in the HO tower is combined in a decoder box and then measured by the photodetectors.

This upgrade design has been validated using laboratory measurements, test-beam data, and within the detector itself during *Run 1*. The upgrade took place along the LS1 in 2013-2014: first, the existing readout modules (RM) were extracted from the detector, then refurbished with SiPMs, and burned-in in a test stand. After verifying that the RMs were working properly, they were installed into HO. Additionally during the installation, the commissioning of the upgraded parts took place [110].

3.1.1 HO design

Figure 3.1 (left) shows the layout of all HO scintillator trays in the five rings of the flux-return yoke. The central ring is labelled *Ring 0*, while the two rings from the plus/minus side along the z axis are *Rings*$\pm$*1,2* (or YB instead of *Ring*). The two layers of *Ring 0* are depicted in white.

In *Ring 0* the trays consist of eight $i\eta$ segments – the scintillator tiles. Each tile is split into four "sigma grooves" (Fig.3.1, right) inlaid with a single WLS fibre. Those fibres then extend until one end of the tray, where optical connectors are mounted as shown in Fig.3.1 (right). Similarly, in *Rings*$\pm$*1* there are five tiles in a tray, and six in *Rings*$\pm$*2*. In total, there are 30 longitudinal $i\eta$ segments in HO.

Each tray represents a single unit in the transverse view. These segments are enumerated in integer $i\phi$ pseudo-coordinates, while covering an angle of about 5° each. Six trays make up a single sector, and there are 12 identical sectors within each ring, resulting in a total of 72 $i\phi$ sections per ring.

The $i\eta$ coordinate ranges from -15 to 15, excluding 0, since the middle of CMS corresponds to the tile boundary in the *Ring 0* trays. The numbering of the sectors starts from the vertical one nearest the counting room (the opposite side of the LHC center) and increases clockwise as seen from the minus side of the z-axis; correspondingly, the $i\phi$ counting originates from the middle tray in this sector. Figure 3.1 (left) shows the $i\phi$ coordinates for the horizontal and vertical sectors.

Due to two service passages for the magnet system (also called chimneys), 6 $i\eta$-$i\phi$ positions are not equipped with scintillators. Four empty positions exist on the minus side of the detector ($i\eta$ =-5, $i\phi$ =11-14), and two on the plus side of the detector ($i\eta$ =5, $i\phi$ =18-19). In total, there are 2154 $i\phi$-$i\eta$ positions instrumented with tiles.

Readout modules

At one end of the trays the WLS fibres are coupled to clear fibres that transmit the light to the readout box (RBX) placed nearby. The RBX hosts the front-end electronics, which consists of the RMs, the clock and control module (CCM) and low/high voltage supplies. In the RM the light is digitized and the signal forwarded to the back-end. This process is controlled by the CCM module responsible for the communication, timing synchronisation and monitoring of the RMs. An additional LED calibration unit (CU) provides test signals used to monitor and calibrate the photo-detectors.

Since the HO tiles are designed to match the projective towers of the barrel HCAL, the light from each of the four grooves has to be bundled. For the central ring there are two scintillator layers, therefore eight fibres have to be combined. This happens in the optical decoding unit (ODU) inside the RM shown in Fig. 3.2 (right). Finally, the ODU is coupled to an 18-channel photo-sensor mounted on top of it.

The electrical signal of the photo-sensors is processed by an ASIC, specifically version 8 of the charge integration and encoding chip (QIE8). The chip integrates the analogue signal in time slices (TS) corresponding to the LHC bunch spacing of 25 ns and encodes it into a 7-

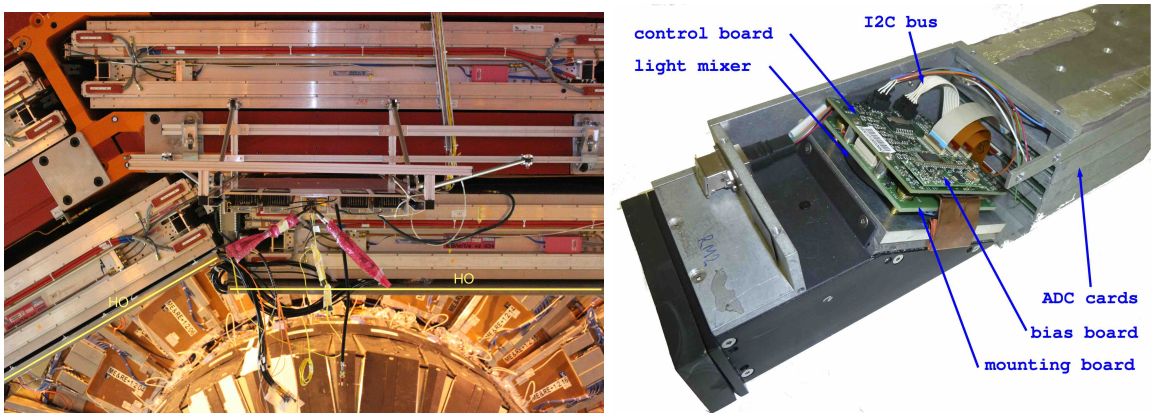


Fig. 3.2: *Left*: an HO readout box extracted from the barrel yoke. The black rectangles are the readout modules with the optical connectors. The scintillator planes of two sectors are indicated by yellow lines. *Right*: internal components of an HO readout module [109]. The SiPMs are fixed to the mounting board and coupled to the ODU with a light mixer.

bit word. An offset can be introduced to adjust the baseline signal of the analogue-to-digital converter (ADC), i.e. the *pedestal*. For HO, like for all HCAL, four TS are summed to obtain the full signal.

Due to geometrical constraints in *Ring 0*, each single sector is equipped with an RBX, containing three RMs. In the other rings six RBXs with four RMs are installed. Therefore, there are 132 RMs necessary to read all the 2154 HO channels. The additional 222 channels that do not receive any scintillation light can be used to monitor the performance of the readout system.

3.1.2 Silicon photomultipliers

SiPMs are essentially arrays of APDs placed on a common silicon substrate. Each diode operates in Geiger mode and is connected in parallel, such that the SiPM produces an analogue signal proportional to the number of firing pixels. The main operational parameters are as well similar to the ones of APDs:

Breakdown Voltage indicates the threshold after which the diodes switch to avalanche mode.

Bias Voltage is the operational voltage of the diode. The difference in bias and breakdown voltage is referred to as **over-voltage**.

Gain corresponds to the charge the SiPM produces for a single photo-electron (SPE).

Dark Rate reflects the thermal noise of the pixels.

Cross-talk in pixels is induced either through avalanches in neighbouring pixels, or through afterpulsing from a trapped charge.

The SiPM chosen for HO is a *Hamamatsu* Multi-Pixel Photon Counter (MPPC) in a surface-mounted device housing. It has a cell pitch of 50 μm and an active area of 3×3 mm, yielding a total of 3 600 pixels.The required operating voltage is of the order of 70 V with a gain that is typically around 6×10^5 fC/photo-electron when operated 1 V above breakdown voltage. At this over-voltage the typical dark current rate is of the order of a few hundred kHz.

One important property of the SiPM is the change of the breakdown voltage and the gain with temperature. As the gain depends linearly on the over-voltage, the change of the breakdown voltage with temperature translates into a change of the gain with temperature. This change has been measured to be about -8 $\frac{\%}{\mathrm{K}}$ at an operating point of 1.5 V over-voltage [111]. This behaviour has important consequences for the operations of the HO detector system, as the SiPM temperature must be well controlled to achieve stable operating parameters.

SiPM supply and control electronics

A compact supply and control system has been developed to facilitate the drop-in replacement of the HPDs with SiPMs. It consists of a stack of three printed circuit boards: a mounting board holding the SiPM and a Peltier element for active temperature control of the SiPM, a bias board generating the SiPM operation voltage, and a control board that regulates and monitors the operation of the individual SiPMs as well as the Peltier element. The control board also connects the SiPMs electrically to the HCAL readout electronics. This architecture allows to reduce the number of newly introduced components and the work needed for installation to a minimum. Figure 3.2 (right) shows those components in an opened RM.

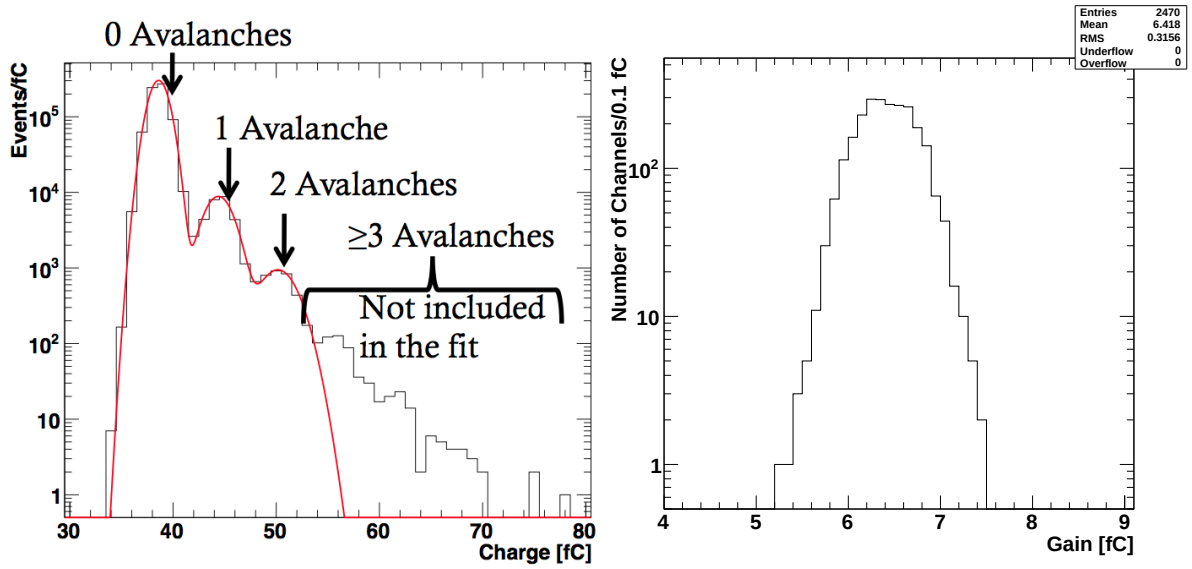


Fig. 3.3: *Left*: Charge distribution in dark events for one SiPM. The integration time here is 100 ns (4TS). The first peak corresponds to the pedestal distribution, where no pixel fires, followed by the SPE peaks. A three-peak fit (red line) is performed to extract the gain [110]. *Right*: Distribution of gain for all SiPMs installed in HO as measured from the dark spectrum [112].

3.2 Commissioning

The hardware is commissioned during the installation in order to identify problems with the new and existing hardware, to validate the installation, and, in case of malfunctions, to make repairs. This is of critical importance, as the barrel part of CMS would remain closed until the next extensive upgrade of CMS in the end of 2016.

The first commissioning step is a communication test with the readout system and the verification of slow control operation and channel response. Next, the measurement and optimization of SiPM operational variables is done as follows:

Temperature The SiPM gain depends linearly on the temperature with a relative dependence of 8% gain shift per $^\circ$K at a foreseen operating point of 1.5 V over-voltage [111]. This temperature dependence requires active control of the temperature of the SiPM with better than 0.1°K stability. Therefore, the SiPM temperature is stabilized by a Peltier element mounted on the back of the SiPM mounting board.

Instead of operating all SiPMs at the same temperature,the Peltier voltage is chosen to be fixed at around 0.3 V. This minimizes the power consumption while keeping a large range of cooling margin.

Pedestal As mentioned before, the baseline signal from the ADC can be tuned. To have a uniform response from all channels, the pedestal is set to 9 ADC counts per 1 TS. For a 4 TS sum, which is used in HO, this results in $\approx$ 36 ADC counts (0 avalanches in Fig. 3.3).

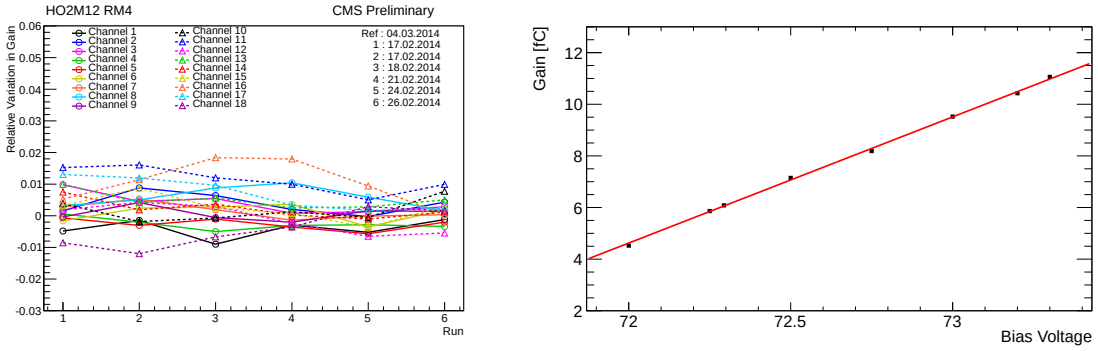


Fig. 3.4: *Left*: Relative variation of the SiPM gain over time for a single RM with 18 channels [113]. *Right*: Gain vs. bias voltage distribution for one SiPM fitted with a linear function [112]. The extrapolation of the line to zero gain provides the breakdown voltage.

Gain The SiPM gain is adjusted by the operating voltage and can be measured from the SPE spectrum as in Fig. 3.3. The distance between the SPE peaks (avalanches) is the gain value, which is obtained from a three-peak fit. All channels are set up to have a gain of around 6 fC/photo-electron to prevent the ADC from saturating, but still have a good signal-to-noise ratio.

The stability of the SiPM gain for all channels of a single RM is shown in Fig 3.4 (left).

Breakdown Voltage Once the operating temperature is adjusted and the system gets into thermal equilibrium, the breakdown voltage is determined: a scan of the bias voltage is performed and the gain measured; this gain-bias voltage dependence (Fig. 3.4, right) is fitted with a linear function, and the extrapolation to zero gain provides the breakdown voltage value.

LED An LED system is used to monitor the stability of the SiPMs outside of collisions. The light intensity is adjusted such that it is sufficiently low to avoid saturation of the SiPMs, and the average signal from all channels is uniform. Because of different SiPM-fiber light coupling in *Ring 0* and *Rings 1-2*, the LED amplitude is set to different values.

A complimentary method allows to determine the SiPM gain using the spread of the registered LED signal distribution as described in Ref. [112].

After the successful completion of these commissioning steps, the new readout hardware is ready for data taking.

For an efficient operation of the upgraded HO, like for any other CMS subdetector, the commissioning setup is migrated to the operational procedures. Therefore, the SiPM parameters are constantly monitored and verified with the reference values. The CMS Detector Control System (DCS) [114] allows the online shifters at *Point 5* to check the status of the detector, as well as to adapt changes if necessary.

3.3 Calibration with cosmic muons

For the usage of any calorimeter it is vital to obtain the absolute energy scale, which connects the electrical output signal to the energy response of the detector. In CMS the measurement of these dependencies is usually performed at test beams prior to the installation on the detector.

Additional inter-calibrations are afterwards performed using reference objects from the proton collisions. In absence of pp-collisions most of the HCAL sub-detectors – HB, HE and HF – may be surveyed and calibrated using radioactive sources placed temporarily next to the active material.

In contrast, the energy calibration of the HO readout system can only be obtained from muons traversing the detector. Those relativistic muons have a very characteristic signal, since they are typically minimum ionizing particles (MIPs). Therefore, the energy response for an HO channel can be factorized as:

$$E\ [\mathrm{GeV}] = \underbrace{Q\ [\mathrm{fC}]}_{\text{SiPM}} \times \underbrace{I\ [\mathrm{ADC/fC}]}_{\text{QIE}} \times \underbrace{M\ [\mathrm{MIP/ADC}]}_{\text{Muons}} \times \underbrace{ES\ [\mathrm{GeV/MIP}\ ]}_{\text{Energy Scale}} \tag{3.1}$$

where E is the final energy response of the HO, Q is the corresponding charge collected by the photo-sensors, I is the ADC transfer factor, M indicates the characteristic MIP signal in ADC units and ES is a relative energy scale factor obtained from test beam measurements. The ADC calibrations I were obtained for every QIE8 card previous to the installation into HCAL [115].

After the commissioning of the new readout electronics, data with cosmic muons was taken to obtain the MIP calibration of the HO. There are two different sets of runs used for the measurement: local HO only runs and global CMS runs involving the muon system.

3.3.1 Local runs

During the early installation and commissioning phase of LS1 each detector system had to perform data taking in a decoupled regime, i.e. independent of other sub-detectors – so-called *local runs*. Therefore, the HO system had to rely on itself to collect and evaluate the data from cosmic muons.

A coincidence HO self-trigger is implemented within the L1 Trigger using the Technical Trigger Processor (TTP) mechanism. The HO rings are split into four quadrants of 3 sectors each as shown in Fig. 3.5 (left). The TTP constantly searches for channels over a threshold of 40 fC in a single time slice. If two channels from different quadrants are above threshold, the trigger fires an L1A command and the event is written to disk. A time delay of 1 TS (25 ns) between the top and bottom quadrants is required to account for the time of flight through a diameter of approximately 8 m.

For a single L1A the buffer of ten TS is written to disk. In Fig. 3.5 (right) two averaged TS profiles are shown for a single channel of the top and bottom sectors, respectively. The first bins reflect the SiPM signal shape, while the last ones are zero-suppressed[1]. The shifted peaks of the signal distributions origin from the arrival time of the muon, and the top sectors receive an earlier signal.

[1] Zero-suppression is a hardware-level mechanism that reduces the stored information by rejecting noise. In HO, a channel is not zero-suppressed, if the sum of any two consecutive TS within the first six exceeds 21 ADC counts.

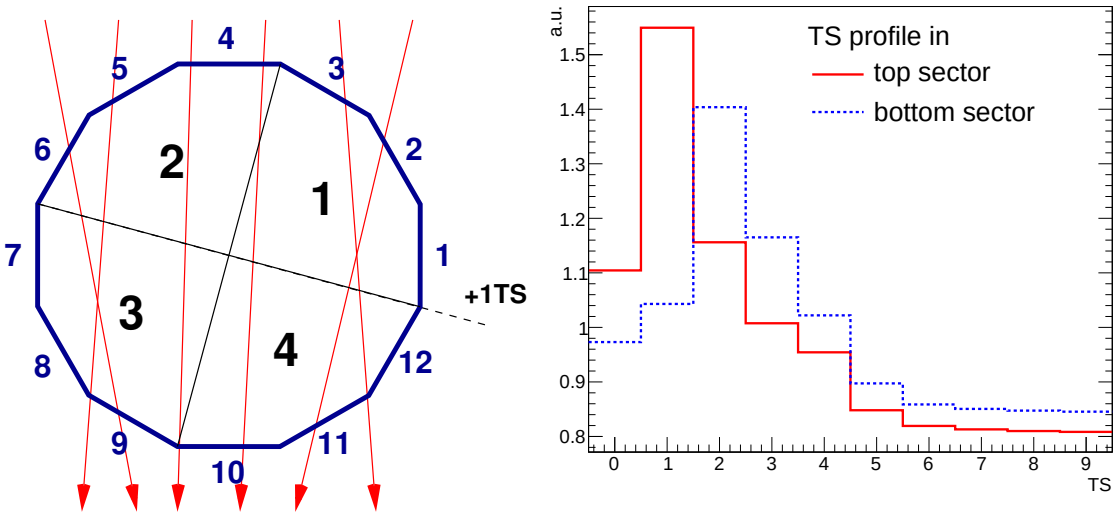


Fig. 3.5: *Left*: Configuration of the TTP quadrants in HO. The coincidence trigger is fired if a signal is seen in two separate quadrants. A time delay of 1 TS is required between the top and bottom sectors. *Right*: Comparison of averaged TS profiles for an up and down sector channel. The first bins show the signal shape, while the last ones are zero-suppressed.

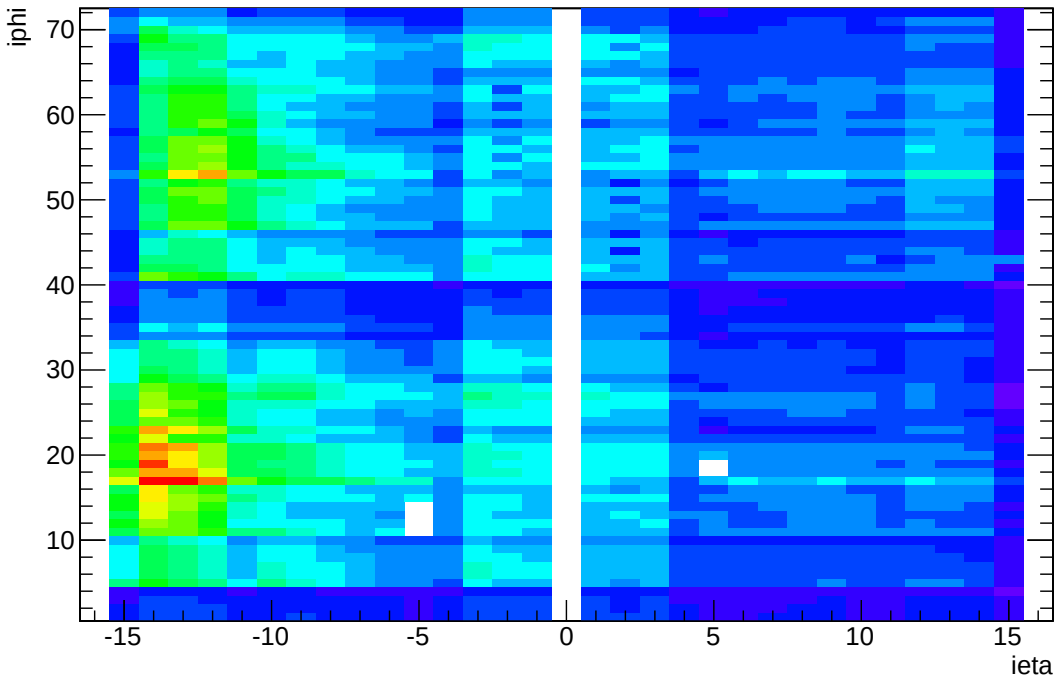


Fig. 3.6: Occupancy of the individual HO channels in a self-triggered run. The tiles with $i\eta < -10$ are close to the cavern shaft. Horizontal sectors are within $10 < i\phi < 30$ and $45 < i\phi < 65$. The white spots at $i\eta = \pm 5$ correspond to the magnet chimneys and are not instrumented with tiles.

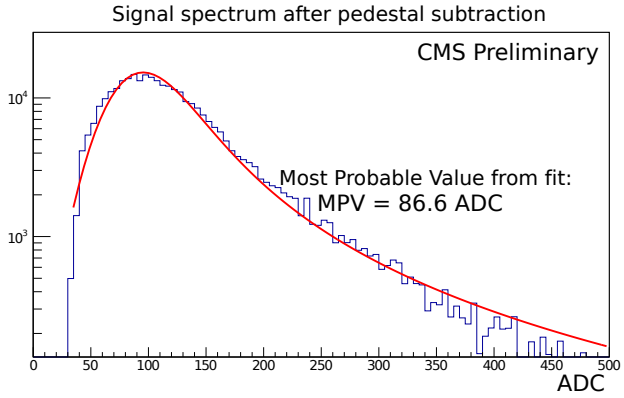


Fig. 3.7: Single channel signal spectrum with coincidence trigger. The fit with a Landau-Gaussian function is shown in red.

Despite the underground location of the CMS detector, approximately 100 m below the surface, a 20 m wide access shaft allows for a high muon flux in certain parts of the underground cavern. During the shut-down, some of the rings were moved in different positions, including directly under the shaft. This fact introduces differences in the occupancy of the different scintillator tiles, but yields a high rate needed for calibration measurements.

Figure 3.6 shows the occupancy of all individual HO channels, obtained with this coincidence trigger configuration. The empty strip at $i\eta = 0$ is an artefact of the HCAL coordinate system rather than a physical gap in the detector. The white regions at $i\eta$ ± 5 correspond to the magnet chimneys, where HO is not instrumented with scintillator tiles.

As expected, the highest rates are observed in the horizontal tiles ($i\phi$ = 15-25, 55-65) in the wheels under the shaft (YB-2: $i\eta$-15,-11; YB-1: $i\eta$-10,-5). The larger rate in the central *Ring 0* compared to the adjacent rings is due to the presence of two scintillating layers, yielding a larger efficiency. The finer structure of outliers along $i\phi$ lines is due to the different scintillator tile sizes, which also leads to different light collection efficiencies and acceptances. For example, $i\eta = \pm 15$ show lower rates because these tiles are half the size of the neighbours.

Events with a number of triggered channels between two and six are selected for the analysis; this reduces the probability of shower events. In total, more than 70M events are selected from coincidence-triggered HO runs.

Local calibration

For the signal reconstruction only four of the ten TS are used: starting from the one before until two after the maximum TS. This time window of 100 ns allows to contain most of the SiPM pulse length. In addition, the corresponding pedestal value is subtracted. Figure 3.7 shows the signal distribution for a single HO channel from local runs.

Theoretically, the energy deposit of a MIP traversing a thin layer follows a Landau distribution. However, due to the thickness of the scintillator and electronics noise the signal size experiences additional Gaussian fluctuations. Therefore, the signal distribution of each HO channel is fitted with a Landau convoluted Gauss function, that accounts for the smearing. The most probable value (MPV) of the Landau fit function is then defined as the MIP value of that channel. An example of the fit including the MPV is shown in Fig. 3.7.

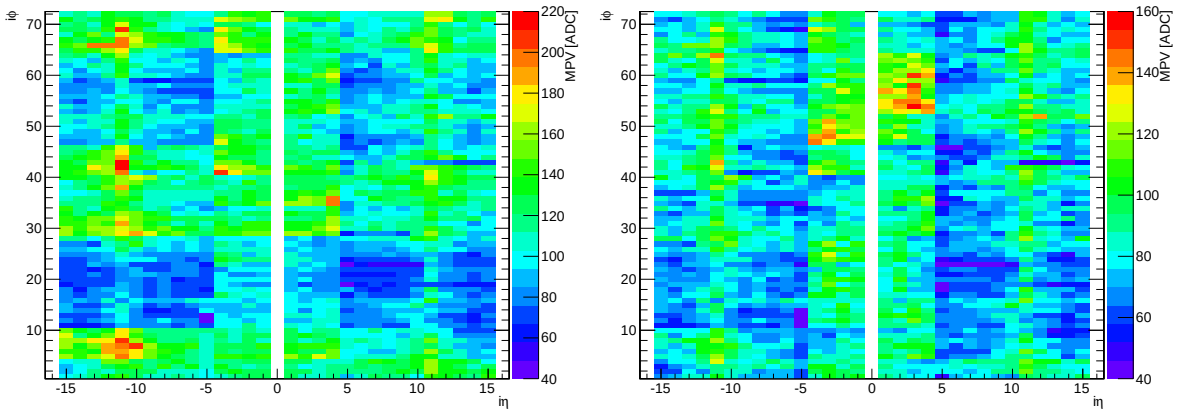


Fig. 3.8: Landau fit most probable values (MPVs) of nominal (*Left*) and path-length corrected signal (*Right*) obtained from HO local run data. The uncorrected signal MPVs show an $i\phi$ modulation due to the incident angle of the muon.

Figure 3.8 (left) shows the MPV parameters for the MIP energy deposits in all individual HO channels. Besides slight variances in SiPM gain and pedestal values of the channels, variations in these measured MPVs are expected due to several factors:

- Ring 0 has two scintillator layers, while *Rings ±1,2* have only one layer.
- Non-uniform light coupling: *Ring 0* uses light mixers to distribute the light evenly on the SiPM surface, thus lowering the light yield.
- Different sizes of scintillators (from 20×40 to $40\times40\,\mathrm{cm}^2$) give varying light collection efficiencies.
- Optical fibres connecting SiPMs and tiles vary in length depending on the distance between scintillator and readout module.
- Variation in $i\phi$ because of sector alignment and muon path-length in the scintillator

From all the listed variations only the last one has to be accounted for: the muon path length within the tile is in direct proportion to the amount of scintillating light emitted. Since the muons hit the tiles from different angles, the signal distribution is smeared out accordingly. In pp-collisions muons originate mostly from the IP, such that the angles are quite small. Therefore, it is more convenient to provide the calibration for muons passing the tile perpendicular.

Since the CMS muon reconstruction is unavailable in local runs, a provisional algorithm was developed to obtain an estimated muon track and correct for the muon path-length within the scintillator:

For each event the HO channels with the largest signals are selected. The pair is rejected if the corresponding tiles are too close to each other (within the same or adjacent sectors), and the next combination is taken. Further, a muon track is built by connecting the two selected hits; and then the angles between the track and the two scintillator planes are calculated. Finally,

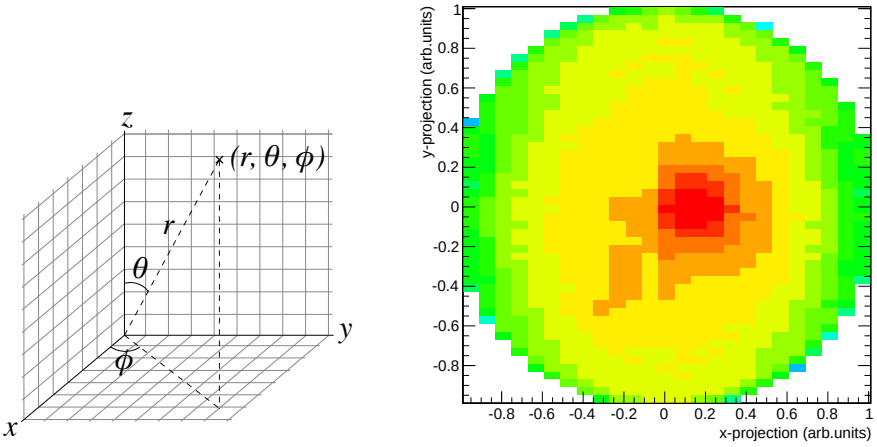


Fig. 3.9: *Left*: Coordinate system used for cosmic muons. For the entry point validation the track length is assumed one [116]. *Right*: Projection if the HO track entry points. Most tracks hit the center of the detector, while a slight shift towards the access shaft is seen. Another hot spot caused by the secondary shaft is seen at $x \approx -0.2, y \approx -0.4$.

the recorded signal is multiplied by the cosine of this angle to correct the energy to the value that a muon at normal incidence to the scintillator tile would produce:

$$\text{Signal}_{MIP} = \text{Signal}_{Muon} \times \cos(\text{tile, muon track}) \; . \tag{3.2}$$

In order to verify the method, i.e. that the reconstructed track is the result of a cosmic muon, the entry point of the trajectories is evaluated. Here, another coordinate system than the CMS one is more convenient: since the muons originate from the sky, the z axis is pointing towards the zenith, and the others are parallel to the surface as in Fig. 3.9 (left). The x and y entry coordinates are computed as the projection of the muon momentum direction on the horizontal plane:

$$x = \sin\phi\cos\theta, \; y = \sin\phi\sin\theta \; . \tag{3.3}$$

The spatial distribution of the HO muon tracks in Fig. 3.9 (right) shows the expected performance, as the majority of the tracks originate from a central position directly above the CMS detector. The deviations are a result of the cavern shafts, which enhance the muon flux from their directions, respectively.

The left plot in Fig. 3.10 shows the MIP energy spectra with and without path-length correction for a side tile of HO, which has a large angle to the abundantly vertical muons. As expected, the uncorrected distribution is shifted to higher values and exhibits a broader peak than the corrected one. In contrast, for a horizontally aligned tile, which is mostly hit at normal incidence by the cosmic-ray muons, the path-length correction does not introduce any visible shift as seen in Fig. 3.10 (right). These results also indicate a sane behaviour of the HO muon track reconstruction.

After the path-length correction, the MPVs of different channels become equalized with the variations remaining due to differences in the read-out hardware (cf. Fig 3.8). An $i\phi$ profile of the nominal and corrected MPVs averaged over the $i\eta$ positions in *Rings*$\pm$*1,2* is shown in Fig. 3.11 (left). On one hand, those rings have all a single layer of scintillator, on the other,

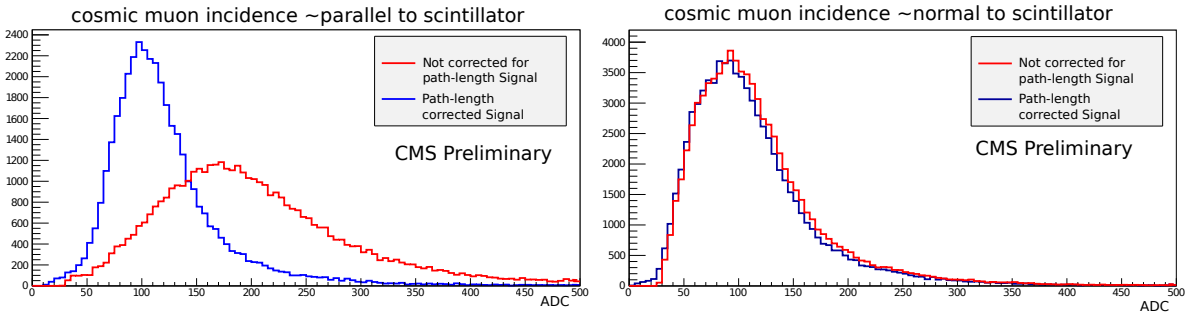


Fig. 3.10: Muon signal spectra for vertically aligned (*Left*) and horizontal (*Right*) tiles with and without the HO muon path-length correction. The signal distribution of the side channel is narrowed and shifted towards a smaller MPV, while for the top channel the changes are marginal.

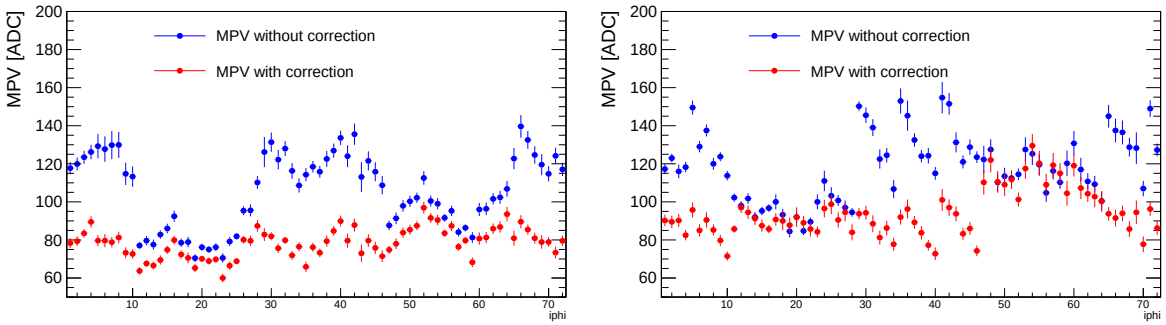


Fig. 3.11: Profile of nominal and path-length corrected MPVs averaged over the $i\eta$ axis for *Rings*$\pm$*1,2* (*Left*) and *Ring 0* (*Right*). The nominal values (blue) show a modulation along the $i\phi$ axis, while the muon path-length correction equalizes the MPVs. The remaining structure is due to hardware differences between the channels.

the data sample is large due to the proximity of the access shafts. In the nominal MPV profile (blue) a modulation is seen, which corresponds to the alignment of the HO sectors forming a circle. The correction flattens out the general modulation, but leaves local variations, which are related to the light guides.

A similar profile is shown in the right plot of Fig. 3.11 for *Ring 0* only. Similar effects of the muon path-length correction can be seen here as well. A big problem with the calibration of the central ring using HO-only runs is due to the double layer of scintillators: a muon from the IP would cross both tiles, but this might not be the case for some cosmic muons, which come from all over the sky.

Therefore, the local track reconstruction experiences an ambiguity for *Ring 0* and cannot select tracks traversing both layers simultaneously. This yields a lower signal as expected from double the tile thickness with respect to the other rings.

3.3.2 Global runs

Within CMS, a global run indicates data taking with several sub-detectors using the central DAQ system. Outside collisions, those runs are mainly taken for commissioning, calibration and time synchronisation between the subsystems.

Almost a year before the start of *Run 2* CMS conducted its first global run of LS1. During the April Global Run (AGR) in 2015 most of the muon systems as well as parts of the HCAL, including HO, were participating in data taking at zero magnetic field. Like for local runs, dedicated cosmic trigger paths were used to collect muons.

For the HO calibration the special cosmic DT trigger was used, where the chambers were triggering either on tracks passing the upper or bottom stations. The side sectors were not included in the trigger due to their low acceptance for muons originating from the sky.

In addition to the AGR, HO also participated in the weekly following Midweek Global Runs (MWGR) as well as the Extended Cosmis Run (ECR) in the end of 2015. Table 3.1 shows the data taking periods including the number of recorded events, as well as the number of tracks selected for the calibration of HO (as shown further).

Table 3.1: Datasets used for the HO global run calibration and performance measurement. The DT trigger type is indicated, as well as the corresponding run ranges and number of selected muons.

Period	Trigger condition	Runs	Number of muons
AGR	Bottom half DT trigger	220744 - 220755 220896 - 220900 220961 - 220966	7.7M (5.0M hits)
	Top half DT trigger	220853 - 220889	3.9M (1.8M hits)
MWGR	Bottom half DT trigger	227323 - 228782	10.7M (6.8M hits)
ECR	Bottom half DT trigger	228783 - 229461	10.6M (3.7M hits)
Total		220744 - 229461	33M (17.3M hits)

Event reconstruction

Like in the local runs, the signal time profile depends on the location of the tile, but as well on the trigger. Since in global runs either top or bottom DT chambers are used, the HO reconstruction window is been adjusted to fit the trigger timing. Among ten time slices, the 3-6 (2-5) TS are integrated for the top (bottom) half trigger runs, respectively.

The data of the muon chambers allows to use reconstructed single-leg stand-alone muons (cf. Sec. 4). Events are selected, that satisfy the DT trigger condition and contain a reconstructed muon matching the trigger object. Similar to Fig. 3.9, the origin of the global muons is verified with the track entry point projection. The resulting distribution in Fig. 3.12 agrees with the expectations as well as the local HO runs. The increased resolution of the muon reconstruction allows to see more details, especially regarding the location of the access shafts.

Next, the global muon tracks are extrapolated with a straight line to the HO plane, and the cross-point determines the scintillator tile which is hit. In order to ensure a good quality of the selected muon as well as a central hit in the HO tiles, the following requirements on the track are applied in addition:

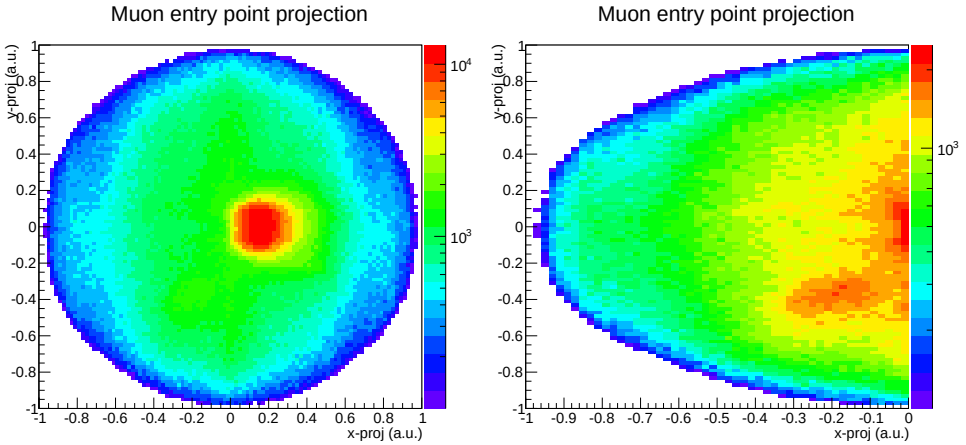


Fig. 3.12: Projections of the entry point for the global muons collected during the AGR (*left*). A zoom into the left half is shown on the *right*. The majority of tracks are coming vertically down, while also the effect of the access shafts is visible.

- Muon track fit: reduced $\chi^2/\text{n.d.f} < 5$
- ΔR between the closest muon hit position in the DT station and the HO hit position (cross-point of the HO tile and the extrapolated muon) to be less than 20 cm for YB $\pm 1, 2$ and less than 50 cm for YB 0.
- The extrapolation point of the muon on the HO tile should be away from the edge by more than 5 cm.

Figure 3.13 shows the above variables together for all global muons and the corresponding hit HO tiles. After this selection, the tracks are more similar to muons originating from the IP.

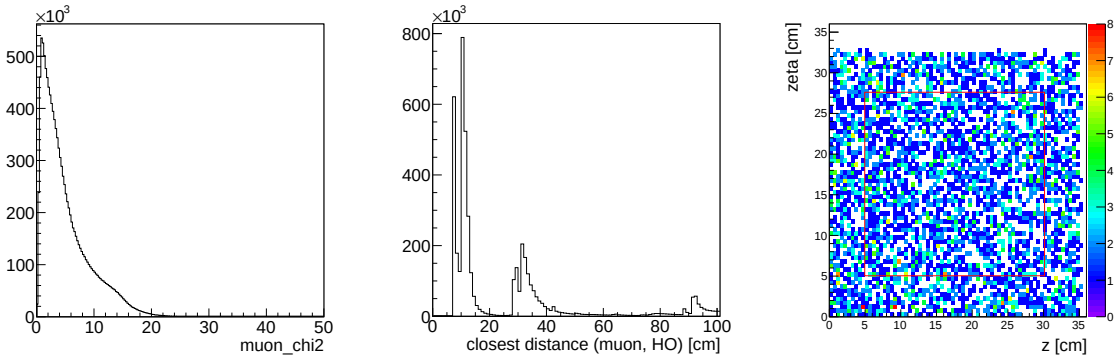


Fig. 3.13: Muon variables used for the selection of global run events. From *left* to *right*: muon $\chi^2/\text{n.d.f}$, ΔR (closest muon station hit, HO hit position), and the HO tile hit map with the selection border indicated in red. In the middle plot, the first peak corresponds to the HO layer close to the first muon station, while the second peak arises from hits into the inner HO layer in *Ring 0* which is farther away from the DT station.

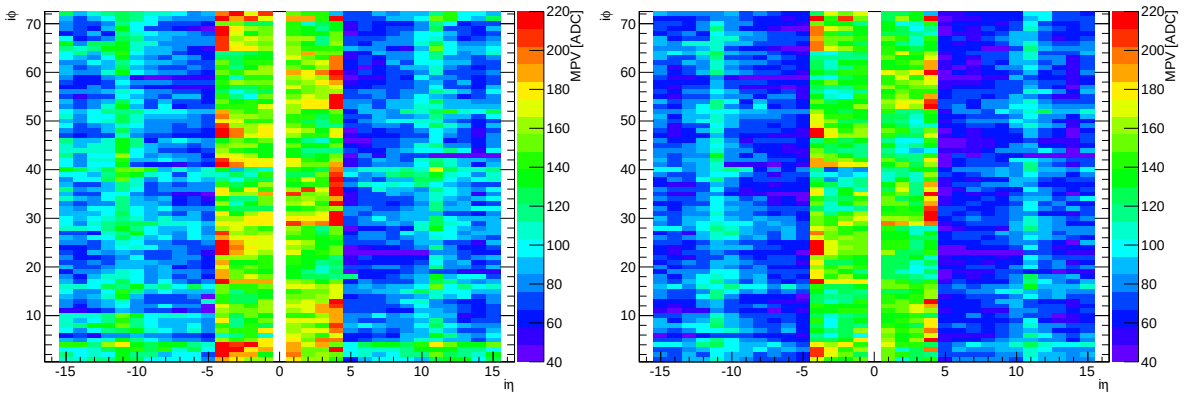


Fig. 3.14: Landau fit MPVs of nominal (*left*) and path-length corrected signal (*right*) obtained from HO global run data. The uncorrected signal MPVs show an $i\phi$ modulation due to the incident angle of the muon. The central ring signal is twice as high due to the double layer of scintillators. Compare to the local run results in Fig. 3.8.

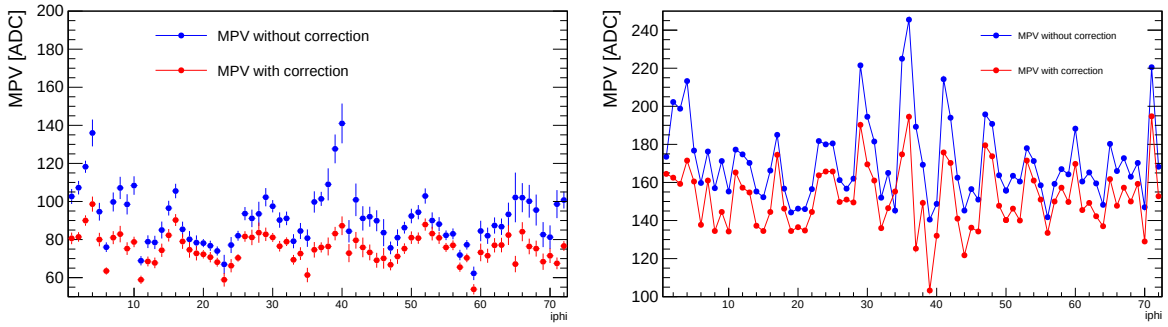


Fig. 3.15: Profile of nominal and path-length corrected MPVs averaged over the $i\eta$ axis for *Rings*$\pm$*1,2* (*left*) and *Ring 0* (*right*) obtained from global runs. The nominal values (blue) show a modulation along the $i\phi$ axis, while the muon path-length correction equalizes the MPVs. The remaining structure is due to hardware differences between the channels. *Ring 0* MPVs are higher due to the two layers.

Global calibration

The calibration in global cosmic runs resembles the local one, with the difference of using reconstructed muon tracks for the path-length correction of the signal. Here the calibration profits from the increased accuracy of the track reconstruction using the muon system.

Overall, this angle correction is smaller than for the local runs, since the DT trigger is limited to the sectors' acceptance. This also affects the statistics for the vertical HO tiles, which lie next to non-triggering DT stations. The resulting MPV from the global runs are shown in Fig. 3.14 for the nominal signal (left) and the path-length corrected one (right).

The biggest advantage of global runs is the more precise calibration of the *Ring 0* channels, which read out two HO scintillating tiles each. In local runs the track reconstruction is too

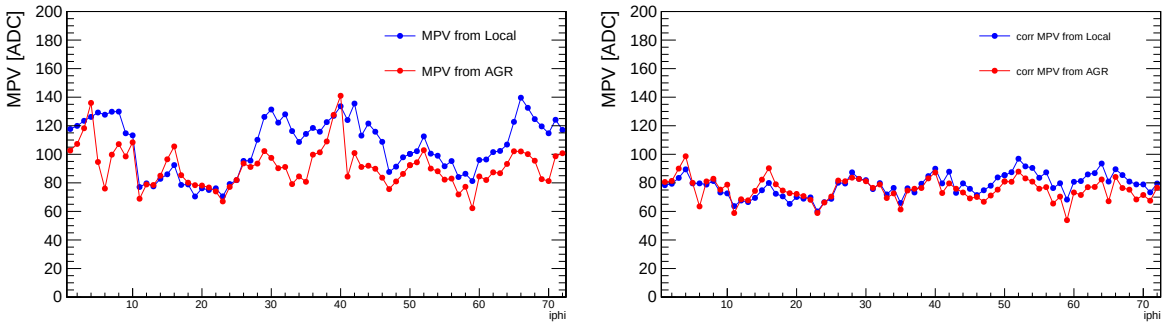


Fig. 3.16: Comparison of averaged MPV profiles from the local and global calibrations in *Rings*±*1,2*. The values for the uncorrected signal (*Left*) are similar only in the horizontal sectors, where the trigger acceptances are similar. After the path-length correction (*Right*) the overall agreement is improved over the whole $i\phi$ range.

inaccurate to distinguish whether only one or both tiles were hit, whereas the globally reconstructed track allows to select muons passing through both layers. The expected doubling in signal amplitude can be also seen in Fig. 3.15 comparing the average MPV profiles along $i\phi$ in the central and outside rings.

Nevertheless, the two calibrations show good agreement in the *Rings*±*1,2* with a single scintillator layer, see Fig. 3.16. For the signals without the path-length correction in the left plot of Fig. 3.16 the best agreement is observed in the horizontal layers, where most muons are normal to the tile and the HO/DT trigger acceptances are similar. After the correction is applied (Fig. 3.16, right) both local and global calibrations agree well in the whole $i\phi$ range. Only for $45 < i\phi < 70$ a small systematic shift is visible, which is due to the different timing of the trigger configurations, while the bottom HO tiles, which are next to the triggering DT chambers agree well between the local and global calibrations.

In global runs the variance in the muon arrival time for the different HO tiles is decreased: more similar time profiles of the signal lead to similar energy containment in the TS sum window. Table 3.2 compares the main features of the local and global run calibrations.

The agreement of the calibrations from local and global cosmic runs verifies both methods and allows to use the obtained values for the absolute energy scale in the HCAL reconstruction, as well as for further studies like the CMS simulation.

Table 3.2: Comparison of local and global run main features.

Local	**Global**
Independent of other sub-detectors	Needs DT to reconstruct muons
Independent of detector configuration	Reco of muons works best with closed barrel
Inaccurate track reco, i.e. angle correction	Standalone reco muon
Non-synchronized timing in different tiles	Stable timing wrt muon trigger
Large acceptance angles	Low acceptance (within DT sector)
Fast, independent calibration	Slow, precise calibration

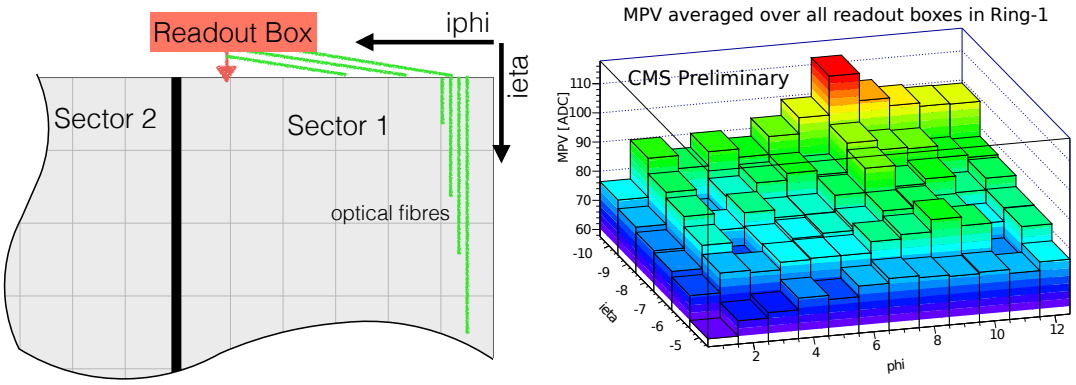


Fig. 3.17: *Left*: HO sector layout in *Rings*±*1,2*, only few fibres are drawn.
Right: MPVs of an averaged sector showing the signal increase towards the RBX.

Fibre light attenuation

One of the remaining sources for the signal variation across the tiles is the fibre light attenuation. From the optical cabling point of view, all sectors within a wheel are symmetric and should show the same behaviour of MPV variation with respect to the scintillator-SiPM distance.

In *Rings*±*1,2* each RBX is connected to 12 $i\phi$ trays of two HO sectors, which is illustrated in the left sketch of Fig. 3.17. The right plot of this figure shows the MPVs averaged over all readout boxes of *Ring-1* corresponding to $i\eta$ [-10,-5]. One can clearly see the tile closest to the readout modules with SiPMs and the decrease to the sides.

The decrease along $i\eta$ is due to the WLS fibre length inside the scintillator tray, while along $i\phi$ the MPV falls with the length of the clear fibres connecting the trays with the RBXs. For *Rings*±*1,2* the RBXs are located between 1 and 2 (at $i\eta = \pm$ 10-11), but for the central ring the readout is located at both edges of the ring ($i\eta = \pm 4$). Figure 3.18 shows the corresponding MPV behaviour along the $i\eta$ coordinate in all rings.

3.3.3 Muon detection efficiency

Following the evaluation of the signal response of the individual channels, the HO muon detection efficiency can be measured as well. The efficiency is defined as the fraction of muons traversing an HO tile, that are detected. Based on the result of Fig. 3.7, a signal of > 45 ADC counts, before subtracting the pedestal, is considered as detection in an HO channel.

With the use of global runs, the denominator is the number of muons, that are projected to cross an HO tile $(i\eta, i\phi)$, while the numerator would be the number of those muons that produce a signal of at least 45 ADC counts:

$$\epsilon(i\eta, i\phi) = \frac{\#\text{ muons passing through HO } (i\eta, i\phi) \;\&\; \text{ADC} > 45}{\#\text{ muons passing through HO } (i\eta, i\phi)} \tag{3.4}$$

The muon reconstruction and track requirements from Sec. 3.3.2 are applied here as well to improve the track quality.

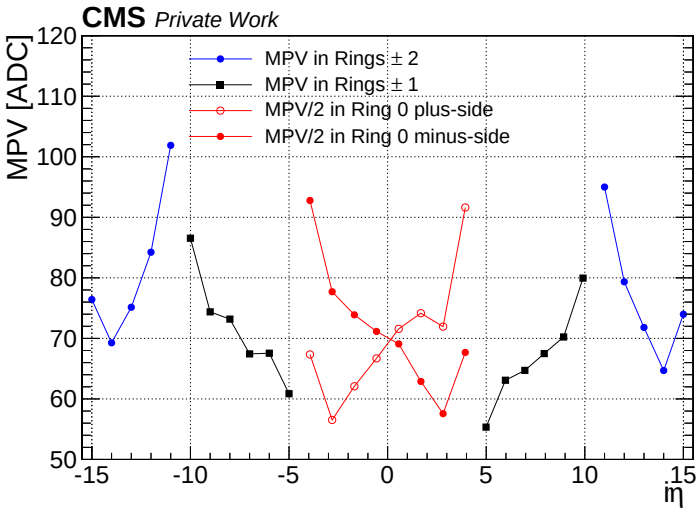


Fig. 3.18: Profile of the path-length corrected MPVs averaged over the $i\phi$ axis in all HO tiles. For *Ring 0* the values are divided by two because of the double layers. The highest signal corresponds to the position closest to the RBX.

Figure 3.19 shows the detection efficiency map in the $i\eta$,$i\phi$ plane (left) and its distribution for all HO tiles (right). In the missing region around $i\eta = \pm 5$, the vacuum chimney is installed and thus there are no scintillator tiles. In general, the detection efficiencies are found to be above 90% in most of the channels, with the median at 98.8%.

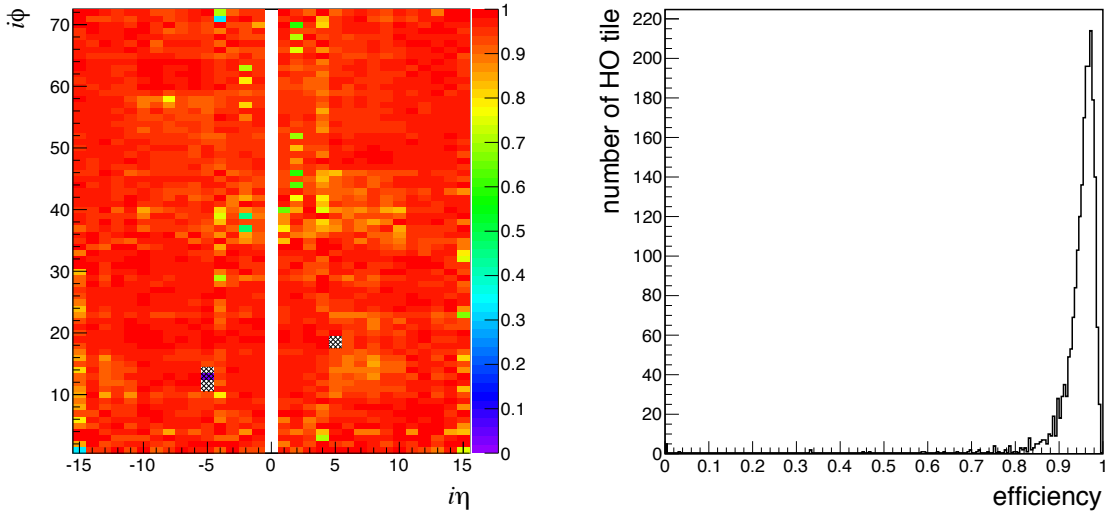


Fig. 3.19: *Left*: HO muon detection efficiency with a signal requirement above 45 ADC counts obtained from cosmic muons in global runs. The dashed regions correspond to not instrumented channels in the magnet chimney. *Right*: distribution of the efficiency for all HO channels. The median is 98.8%.

Chapter 4

Object reconstruction and identification

4.1 Data flow at CMS **65**
4.2 Particle-Flow algorithm **66**
4.3 Physics objects **69**
4.3.1 Primary vertices 70
4.3.2 Leptons 70
4.3.3 Jets 74
4.3.4 Missing transverse energy 78

In this chapter the object reconstruction employed by the CMS experiment is discussed. First, the general processing steps of data and simulation samples will be shown. The second section will give an overview of the Particle-Flow algorithm, that allows to construct physical object candidates based on a combination of the various subdetector systems. In the end, the reconstruction of actual physics objects used in this analysis will be presented.

4.1 Data flow at CMS

Similar to the two-step CMS trigger system, the event processing is performed in multiple stages until the data can be used by the end-users in physical analyses. For this purpose, CMS developed several data formats (tiers) that each serve the purpose of storing the relevant information of the reconstruction step. A short description of the main tiers is given in Tab. 4.1.

As mentioned in Sec. 2.3.2, the MC simulation takes two steps – the event generation and detector simulation –, which is reflected in the *GEN* and *SIM* formats used by CMS. After the digitization of the detector signals (*DIGI*) the MC format copies the real data obtained from the detector after trigger processing (*RAW*).

The next steps involve the reconstruction of the physical objects such as tracks, jets and leptons and are similar for data and MC. These algorithms will be presented in the next section (cf. Sec. 4.2). For the actual analysis, samples with skimmed information are used, since the reduction in event data size plays a considerable role in the computing performance. The final *miniAOD* format [117] provides the most common object variables requiring 40 times less space than the fully reconstructed event information.

Table 4.1: Data tiers used for the processing of the real and simulated CMS data [117, 118].

Usage	Tier	Purpose	MB/ev
MC	*GEN*	Event information as provided by the MC generator. Contains kinematic variables as well as auxiliary strings such as event weights and model parameters.	$O(0)$
	SIM	Energy depositions of the particles using the GEANT4-based simulation (sim hits).	2
	DIGI	Detector responses corresponding to the simulated hits. Similar to the real detector output.	2
Data/MC	*RAW*	Detector data after the L1 Trigger and HLT processing, includes higher level quantities calculated for the trigger decisions.	2
	RECO	Reconstructed detector hits and clusters, as well as objects like tracks, vertices, jets, leptons.	2
	AOD	Analysis Object Data: subset of *RECO* including reconstructed objects and few hit information.	0.5
	miniAOD	Skimmed *AOD* information with only the most relevant quantities available.	0.05

4.2 Particle-Flow algorithm

A typical pp-collision results in a large variety of (semi-) stable particles, that can be registered in several independent detector systems of CMS simultaneously. Figure 4.1 shows the signatures of some characteristic particles while traversing the subdetector layers.

Muons, being mainly minimum ionizing particles, are the only particles to cross the muon system, but also leave tracks in the inner tracking detectors, whereas electrons get absorbed in the ECAL after leaving a track. Charged hadrons behave similarly to electrons, but deposit most of their energy in the HCAL. Neutral particles, such as photons or pions, can only be identified as a calorimeter cluster without a corresponding track.

Given the complementarity of the measurements in the tracking detectors and the calorimeters, it seems natural that CMS has developed the Particle-Flow Algorithm (PF) [120–122] to combine this information for a more accurate reconstruction of the physical objects. A great benefit is achieved for such complex objects as jets, since they consist of multiple particles, with two thirds being charged, thus measurable in the tracker.

Fundamental elements

The building blocks for the algorithm are composed of *elements* from the charged particle tracks, calorimeter clusters and muon tracks obtained from the individual subsystems. Those elements have to be reconstructed with a high efficiency and low fake rate given any complex environment of the collision.

Therefore, two main concepts are adopted: the *iterative tracking*, which ensures a proper track finding even for low p_{T} objects, and the *calorimeter clustering*, that pursues several aims from the detection and measurement of stable neutral particles, as well as the separation of those from charged objects, over to complementing the electron and low momentum charged hadron measurement.

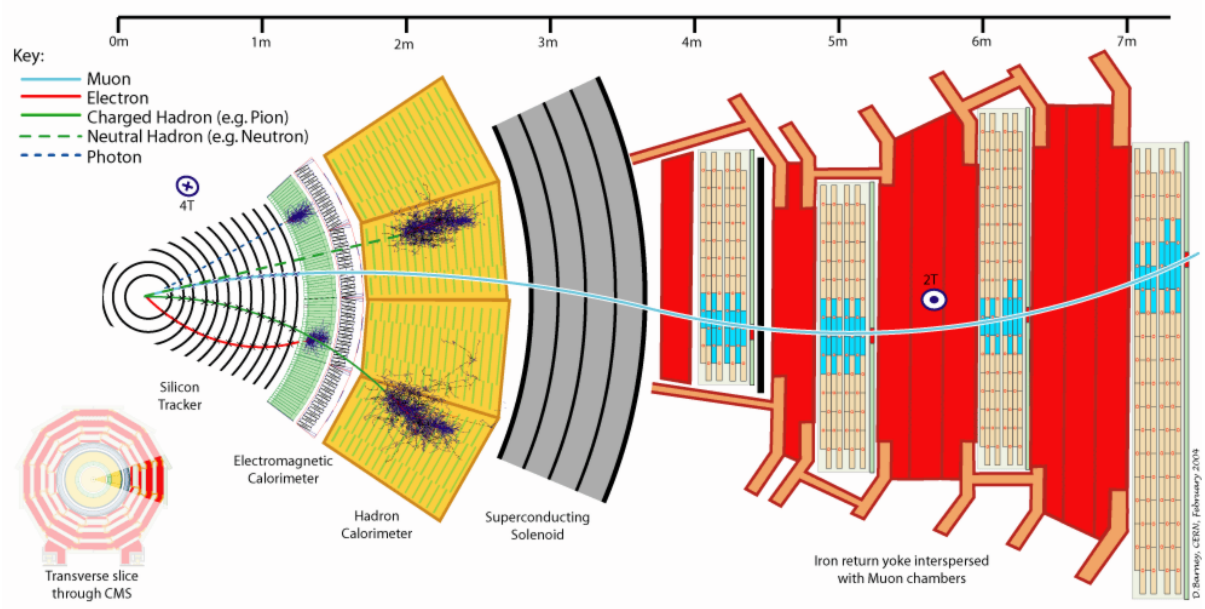


Fig. 4.1: Slice of the CMS detector in the transverse plane with some characteristic particles crossing the subsystems indicated [119]. For simplicity, only energy deposits in the most relevant subdetectors are shown.

Iterative Tracking The Kalman filter [123] method is used to recognize single particle tracks. Starting from seeds selected with very tight criteria, the found tracks are removed from the next iterations. The following iterations use progressively loosened seeding criteria to increase the tracking efficiency. After the third step the constraints on the origin vertex are even further reduced to allow the reconstruction of tracks from secondary vertices.

Calorimeter Clustering This algorithm identifies energy clusters in the independent calorimeter subsystems separately, except for HF, where each cell is assigned to a single cluster. Three steps are performed: first, local energy maxima in the cells are identified as "cluster seeds", from which further "topological clusters" are accumulated from the adjacent cells. Finally, the energy of each cell is shared among all clusters according to the cell-cluster distance.

Link algorithm

After the basic elements are created, those corresponding to the same particle have to be connected into a single object called *block*. A certain measure of closeness is used to quantify the probability of two elements to belong to the same object. Given the granularity of the CMS detectors, these object blocks typically contain no more than three elements.

For the simple case of a track reconstructed in the muon system, it is linked to a charged particle track from the inner tracker, if the fit of the combined tracks results in an acceptable χ^2. In the case where several of such *global muons* can be fitted between a particular muon track and several tracker tracks, the muon with the smallest χ^2 is preferred. Hence, the distance here is defined by the goodness-of-fit value χ^2.

More complex algorithms are required for the connection of charged-particle tracks with calorimeter clusters. A given track is extrapolated from the last measured hit towards the calorimeters, and if it falls within the boundaries of any adjacent cluster, these two blocks are linked together. The distance in the η-ϕ plane ΔR is used here as a metric of the link quality.

In a similar fashion, links are established between bremsstrahlung photons emitted by an electron and the corresponding track, since the photons should be radiated from within the tracker layers tangentially to the track.

ECAL clusters are linked to elements in the HCAL, if they lie within the boundaries of the less granular calorimeter and are within an acceptable distance in the η-ϕ plane. The same applies for links between the preshower ES and ECAL.

Particle reconstruction and identification flow

After the blocks are assembled from the single elements, the actual particle reconstruction and identification is carried out by the PF algorithm. Following a sequence from simple to more complex objects, a list of particle candidates is created from all blocks in each event. The same procedure is followed for each block:

- A *particle-flow muon* is created for each global muon, if its momentum is compatible with the tracker-only measurement within three standard deviations. The corresponding track is removed from the block afterwards, but an estimated deposit of 0.5 (3) GeV in the ECAL (HCAL) is kept for later steps.
- Since electrons tend to radiate bremsstrahlung leading to short tracks while traversing the tracker volume, the assumed electron hits are refit with a Gaussian-Sum Filter (GSF) [124] to account for those possible energy losses. After certain identification criteria on the track-cluster compatibility are satisfied, a *particle-flow electron* is created, and the corresponding footprints are removed from the block.
- At this point, tighter quality criteria are applied to the remaining tracks, to select only presumably charged hadron tracks that are measured more precisely in the tracker, than in the calorimeter. The required energy calibrations are obtained beforehand [120].
- The detection of neutral particles within a block involves the comparison of the track momenta and corresponding cluster energy. If there are several tracks linked to a single cluster, the combined momenta of the tracks are compared to the calibrated calorimeter energy. Vice versa, only the cluster closest to the track is selected. Depending on the track momenta-cluster energy balance, those additional links might be preserved for more accurate hadron energy determination, or removed in case of overlapping photons.
- Remaining tracks give rise to *charged particle-flow hadrons*, for which the momentum and energy measurement is mainly driven by the track, rather than the corresponding cluster.
- In case of a significant excess of the cluster energy over the matching track momenta, neutral objects are created. Particularly, if the excess is larger than the total ECAL energy, a *particle-flow photon* is created from the ECAL cluster, and the remaining part of the excessive energy is calibrated and assigned to a *neutral particle-flow hadron*. Otherwise, the uncalibrated ECAL energy fully translates to the photon.

  Similarly, the remaining unlinked calorimeter clusters are assigned to neutral particles.

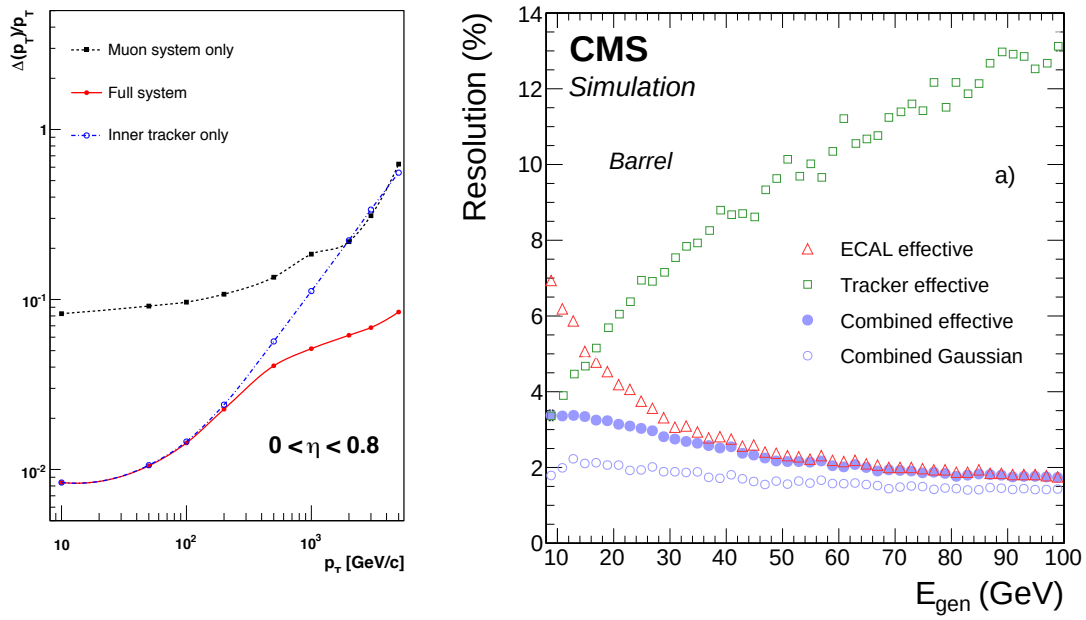


Fig. 4.2: *Left*: comparison of the muon transverse momentum resolution for global muons (red), tracker-only muons (blue) and tracks reconstructed only in the muon system (black) [73]. The low p_T resolution is mainly driven by the inner tracker.
Right: energy resolution for electrons reconstructed using only the track (squares) or the ECAL (triangles) in comparison to the combined measurement as used by the PF algorithm (circles) [125].

Figure 4.2 compares the muon momentum and electron energy resolution reconstructed in the individual CMS subdetectors with the combined measurements. In the case of muons, the tracker resolution dominates throughout a wide p_T range due to the precise track measurement and larger magnetic field. At high p_T, both measurements complement each other decreasing the combined momentum resolution. For electrons, the difference between the tracker-only and ECAL-based measurement arises from the contribution of bremsstrahlung photons. Since the fraction of those photons is relatively independent of the electron energy, the low momentum resolution is driven by the tracker measurement. At higher energies, already more energy is carried away by the photons, such that the ECAL resolution takes over.

4.3 Physics objects

The PF algorithm creates candidates for most of the basic physics objects required for analysis. Those are then typically stored in the aforementioned *RECO* data format. Nevertheless, physics analyses require refinement and identification of those objects, as well as additional quantities not covered by the algorithm.

This section discusses the physics objects used for this analysis together with the corresponding identification criteria. A description of the reconstruction for photons and τ-leptons, which are not used in this analysis, can be found elsewhere [120, 122, 126, 127].

4.3.1 Primary vertices

Undoubtedly, the central quantity of each event is the primary vertex (PV), where the parton collision takes place. This point will be later used for the reconstruction and selection of tracks. As already shortly mentioned in Chapter 2, the LHC beams are steered in such a way as to overlap in a small area around the nominal IP. Typically, this interaction region is constrained to about 50 cm along the z-axis, and not further than a few cm away from the beam axis.

During operation at such enormous instantaneous luminosities as that of the LHC, the probability of multiple pp interactions per bunch crossing – pileup – is quite high. Therefore, the primary vertex of the most interesting collision has to be identified.

The basic idea for the reconstruction is as follows. First, a set of tracks is selected according to some quality criteria, like the impact parameter, or number of hits in the tracker. Those tracks are then clustered based on their z-coordinate at the point of closest approach to the beam axis. Clusters that more than 1 cm apart and contain at least two tracks are selected and an adaptive vertex fit [128] is performed. Finally, the cluster with the largest sum of the squared track momenta $\sum p_{\mathrm{T}}^2$ is selected as the primary vertex of interest.

The PF algorithm works independently of the vertex reconstruction, but assigns each particle candidate to a particular PV wherever possible. Candidates not associated to the good PV are used to estimate the average effect of pileup in the event, as discussed further.

In addition to the multiple interactions within the same bunch crossing, or *in-time pileup*, particles from the adjacent collisions might also affect the object reconstruction. This *out-of-time pileup* mainly interferes with the pulse shapes of the calorimeter sensors, rendering the reconstruction less accurate. Special methods based on pulse shape fits have been developed in CMS to mitigate those effects and maintain the nominal resolution [129].

As common for CMS, the vertex for this analysis is required to be less than 24 (2) cm away from the nominal IP in the longitudinal direction (transverse plane). In addition, the number of degrees of freedom of the fit result should exceed four to ensure a good quality vertex.

4.3.2 Leptons

Since this search focuses on final states with a single lepton, i.e. electron or muon, those objects are of great importance for the analysis. The *particle-flow* candidates for these leptons are used as a baseline, but additional requirements are imposed to ensure a good quality of the reconstructed object, as well as the rejection of fake leptons or those not coming from the primary vertex. In general, similar criteria on the kinematic acceptances are imposed for electrons and muons and both are required to be isolated.

Muons

To improve the quality of the *particle-flow muon* candidates, additional selection criteria, so-called *muon identification (ID) working-points (WP)* [130], are defined. The three WPs – *loose*, *medium* and *tight* – are based on successively tightened requirements, leading to a decreasing fake rate at the cost of falling efficiency [131]. While the *loose* WP is basically a PF muon, that was reconstructed either as a global or tracker-only muon, the *tight* criteria include additional requirements on the track and hit parameters.

For this analysis the *medium* WP is used, which yields a relative high efficiency with a good fake rejection power [132] as shown in Fig. 4.3 (*left*). The candidate is required to be identified

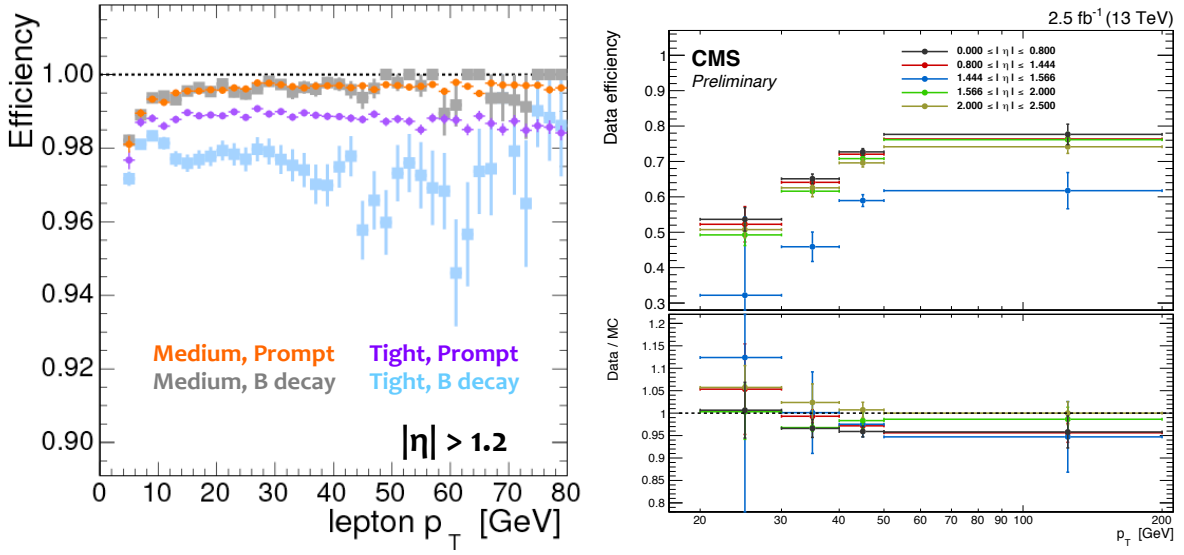


Fig. 4.3: *Left*: muon identification efficiency in simulated $t\bar{t}$ events for the *medium* and *tight* criteria [134]. The *medium* WP shows a high and stable efficiency for prompt muons, as well as from b-decays. *Right*: electron identification efficiency with the *tight* WP in data and a comparison to the simulation expectation [135]. Except for the barrel-endcap overlap region, the efficiency is similar over the whole pseudorapidity range, and in good agreement with simulation.

as a *loose* muon with a fraction of valid tracker hits above 0.8. Further, the segment compatibility[1] is evaluated. If this value exceeds 0.451, the muon is accepted, otherwise, if it is below 0.303, the object is rejected. For candidates with intermediate compatibility, additional requirements are imposed:

- the candidate is a global muon
- the global track fit gives a χ^2 below three
- the local position match between the tracker and muon satisfies $\chi^2 < 12$
- the kink finder returns a value below 20

On top of these CMS-wide ID recommendations, additional analysis-specific criteria are imposed on the origin of the muon in order to select only muons stemming from the primary vertex (i.e. prompt leptons). The track distance of closest approach to the vertex should be within 0.5 mm in the transverse plane, and not farther than 1 mm along the z-axis. Moreover, the significance of the three-dimensional impact parameter is required to be above four.

Finally, only candidates within the muon system acceptance $|\eta| < 2.4$ and with a transverse momentum $p_T > 10\,\text{GeV}$ are selected to ensure a good identification efficiency (Fig. 4.3, *left*). The corresponding muons will be further referred to as *good muons*.

[1]Segments are tracks found within a single station of the DT or CSC [130]. The segment compatibility is a continuous variable to quantify the compatibility of a tracker-muon object with the muon hypothesis [133].

Table 4.2: Variables used in the CMS electron ID. The upper limits corresponding to the *tight* WP are shown for the barrel ($|\eta| < 1.479$) and endcap ($1.479 < |\eta| < 2.5$) regions separately [136].

Variable	Definition	Barrel	Endcap
$\sigma_{i\eta i\eta}$	ECAL crystal-based shower covariance in the η direction	0.0101	0.0283
$\Delta\eta_{in}$	Difference in η between the energy-weighted supercluster[2] position in the ECAL and the track direction at the innermost trackerposition	0.00926	0.00724
$\Delta\phi_{in}$	Same for ϕ	0.0336	0.0918
H/E	Ratio of energy measured in the HCAL, in a $\Delta R = 0.15$ cone behind the electron seed, over the energy measured in the ECAL	0.0597	0.0615
$\lvert 1/E - 1/p \rvert$	Absolute difference between the inverse electron energy measured in the ECAL and the inverse momentum measured in the tracker	0.012	0.00999
Δxy	Track-vertex closest approach in the transverse plane	0.0111	0.0351
Δz	Same along the beam axis	0.0466	0.417
N_{hits}^{miss}	Number of missing inner layer hits in the electron track	≤ 2	≤ 1
No associated photon conversion vertex			

Electrons

Since *particle-flow electrons* are constructed from matched tracks and calorimeter clusters, the identification procedure is more elaborate [125]. The WPs are defined in a similar fashion as for muons, but now all categories include the same variables with tightening requirements [136]. Those criteria mostly comprise the shower shape quality and compatibility of the cluster energy and track momentum [135]. In addition, misidentification of jets or photons converted in the tracker is suppressed.

In this analysis the *tight* WP of the ID is used to select electron candidates. Table 4.2 summarizes the selection variables as well as the upper limits. Given the decreasing accuracy of track reconstruction towards higher pseudorapidities, as well as different geometries of the EE and EB, separate requirements are applied in the barrel and endcap regions. Figure 4.3 (*right*) shows the identification efficiency for electrons in different pseudorapidity intervals.

Despite the large acceptance of the ECAL, PF electrons can only be identified within the silicon tracker coverage of $|\eta| < 2.5$. In this analysis the pseudorapidity is restricted even further to 2.4 to allow for the measurement of activity in a cone around the electron. Due to the identification inefficiency in the overlap region between the ECAL barrel and endcaps, at $1.44 < |\eta| < 1.56$, this region is excluded as well.

Like in the case of muons, the electron transverse momentum is required to be above 10 GeV for an efficient identification. Those selected electrons will be further referred to as *good electrons*.

[2]Superclusters are collections of ECAL clusters from the electromagnetic shower and bremsstrahlung photons associated with a single electron object.

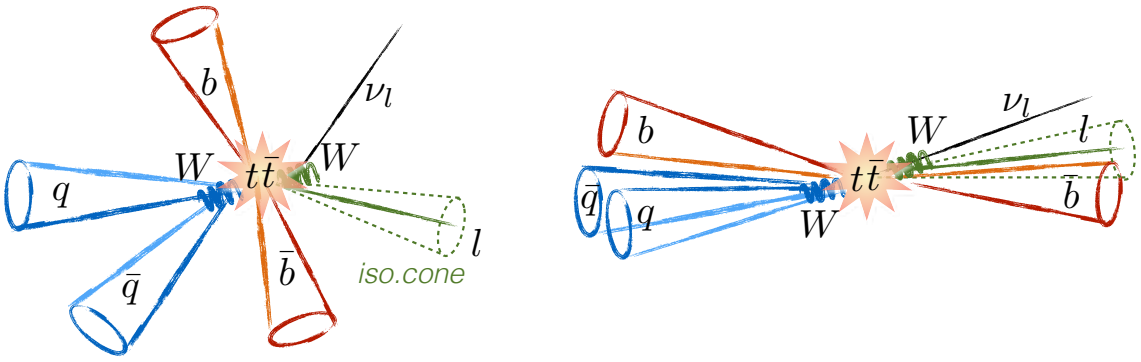


Fig. 4.4: Topology of a $t\bar{t}$ event in case of low (*left*) and high (*right*) boost, where the isolation cone of the lepton can overlap with other decay products [137].

Isolation

This analysis concentrates on primary (*prompt*) electrons and muons, that are produced in the decays of W-bosons, but in pp-collisions charged leptons can also originate from b- or c-quark decays. Those secondary, or *non-prompt*, leptons, as well as misidentified jets are the main background to the leptons from boson decays. Nevertheless, the distinct feature of non-prompt or fake leptons is the significant energy flow near their trajectories, due to the accompanying hadronic activity. The absence of such a surrounding activity, which is typical for prompt leptons, is characterized as isolation.

To quantify the degree of isolation for a particular lepton, the transverse momenta of all the PF particle candidates in a ΔR-cone around the trajectory are summed up. Then, the ratio of these sums and the lepton p_T define the *relative isolation* for a given cone size R:

$$I_R = \frac{1}{p_\mathrm{T}^\mathrm{lep}} \Bigg\{ \sum_{\Delta R<R} p_\mathrm{T}(\text{charged hadrons from the PV}) + \max\left[0, \sum_{\Delta R<R} p_\mathrm{T}(\text{photons}) + \sum_{\Delta R<R} p_\mathrm{T}(\text{neutral hadrons}) - \frac{1}{2}\sum_{\Delta R<R} p_\mathrm{T}(\text{ch. had. not from PV})\right] \Bigg\} \tag{4.1}$$

The first term accounts for all charged hadrons whose tracks are identified to be originating from the PV. Contributions from photons and neutral hadrons are corrected for the surrounding pileup energy. This energy is estimated from the identified charged hadrons, that do not originate from the PV. A factor of one-half is used to account for the average fraction of charged and neutral pileup contributions. In case the pileup correction exceeds the neutral sums, only the first term from charged hadrons is kept.

Normally, the isolation cone size is fixed to a value, which is comparable to the jet clustering radius (cf. Sec. 4.3.3). In the case of final states at the TeV scale, as those in the studied SUSY scenarios, the top quarks are produced with a high Lorentz boost [138]. As a result, the top decay products are highly collimated as illustrated in Fig. 4.4. In addition, the high jet multiplicity environment increases the probability of an accidental overlap of the lepton and a jet. This renders the fixed-cone isolation criteria less efficient in events with high jet activity [137].

A solution to reduce this effect was proposed in Refs. [138, 139]. Based on the assumption,

that the main source of overlap is the b-jet from the top-quark decay, the cone size of this jet is estimated as:

$$\Delta R_{\text{b-jet}} \approx \frac{2m_{\text{mother}}}{p_{\text{T mother}}} = \frac{2m_{\text{b}}}{p_{\text{T}}^{b}} \propto \frac{10\ \text{GeV}}{p_{\text{T}}^{\text{lep}}} \tag{4.2}$$

This approximation motivates the following $p_{\text{T}}^{\text{lep}}$-dependent isolation cone size:

$$R_{\text{iso}} = \begin{cases} 0.2, & p_{\text{T}}^{\text{lep}} \leq 50\ \text{GeV} \\ \frac{10\ \text{GeV}}{p_{\text{T}}^{\text{lep}}}, & p_{\text{T}}^{\text{lep}} \in (50, 200)\ \text{GeV} \\ 0.05, & p_{\text{T}}^{\text{lep}} \geq 200\ \text{GeV} \end{cases} \tag{4.3}$$

which is still sufficiently large to identify non-prompt leptons from b-quark decays, but small enough to avoid accidental overlap of the lepton and jets. The upper and lower bounds ensure some safety margin for those two cases.

Given this new cone size dependency, the now called *mini-isolation* I_{mini} for electron (muon) candidates is required to be below 0.1 (0.2) to satisfy the isolation requirement. A comparison of the fixed-cone and $p_{\text{T}}^{\text{lep}}$-dependent relative isolation performances is discussed in the Appendix B.

4.3.3 Jets

Jets, i.e. ensembles of particles created during the hadronization of partons, are produced in abundance in the decays of supersymmetric gluinos. In general, the multiplicity and energy of the jets are a measure of the energy scale, and thus an important property of the events.

First, the jets are reconstructed from the detector responses, whereupon corrections to the measured energies can be applied [140]. Additional identification criteria allow to distinguish jets originating from b-hadrons [141], as well as from pileup interactions [142].

Clustering

The typical jet energy fractions carried by charged particles, photons and neutral hadrons are 65%, 25% and 10%, respectively (Fig. 4.5, *left*). Therefore, the PF algorithm allows to reconstruct about 90% of the jet constituents with good precision, whereas only 10% rely on the less accurate response of the HCAL.

In the present analysis, the *particle-flow* candidates are clustered into jets using the anti-k_T algorithm [143], designed to reproduce the original parton energy prior to hadronization. This algorithm belongs to the family of sequential recombination algorithms, where particles are clustered into subjets based on a defined distance parameter proportional to the particles' transverse momentum k_T; a particle i is merged to the entity j, which can be either another particle or a jet, if:

$$\min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta R_{ij}}{R^2} < k_{Ti}^{2p} \tag{4.4}$$

where ΔR_{ij} indicates the spatial distance between the entities in the η-ϕ plane and R is the radius parameter related to the area of the jet. The parameter p defines the power of the energy

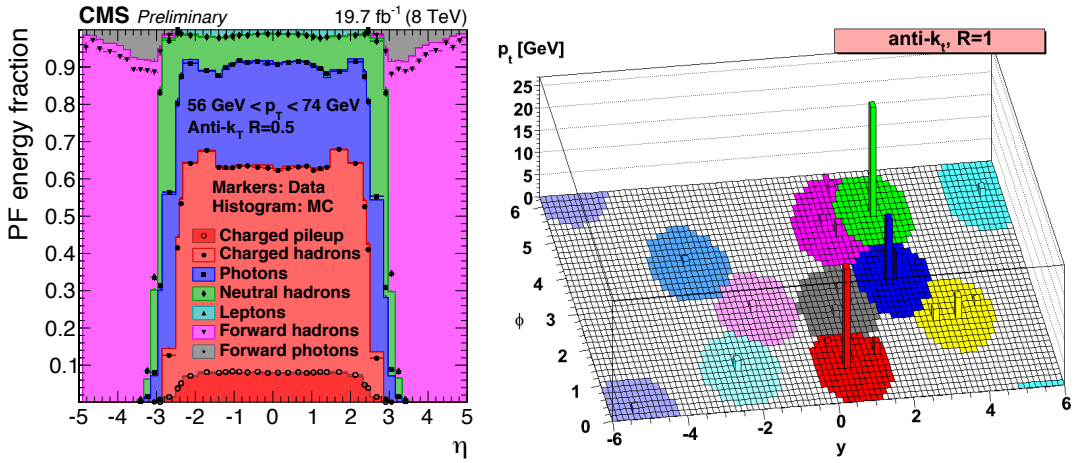


Fig. 4.5: *Right*: energy fraction of individual components of PF jets clustered with R = 0.5 using the anti-k$_T$ algorithm [140]. In the central region $|\eta| < 2.4$ the tracker information allows to distinguish all components of a typical jet.
Left: jets clustered with the anti-k$_T$ algorithm and R = 1 show a round shape [143].

relative to the geometrical scale. Moreover, its actual absolute value is less important, than its sign.

In case of the anti-k$_T$ algorithm, p is set to -1, which implies, that the clustering commences with the hardest particles in the event and clusters them to the closest entities. This distance metric results in almost ideally round shaped jets in the ΔR-plane, as shown in Fig 4.5 (*right*). The parameter R translates into the radius of those clusters.

For this analysis, *particle-flow candidates*[3] are clustered into $R = 0.4$ jets, which are required to be within the tracker acceptance $|\eta| < 2.4$ to profit from the PF precision. Further, a jet is rejected, if the energy contribution from either the neutral, charged electromagnetic or hadronic constituents exceeds 99% (jet ID).

Energy corrections

The measured jet energy is naturally not the same as of the corresponding parton. Differences arise mainly due to the non-linearity and non-uniformity of the calorimeter response, but as well from electronics noise and pileup energy contribution. Most of those effects can be compensated with an appropriate Jet Energy Calibration (JEC).

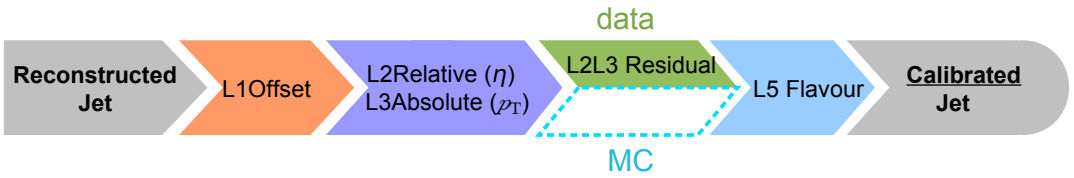


Fig. 4.6: Factorization of the CMS Jet Energy Calibration corrections. Except for the residual corrections, all factors are applied for data and simulated jets [144].

[3]Charged hadrons originating from identified pileup PVs are removed beforehand.

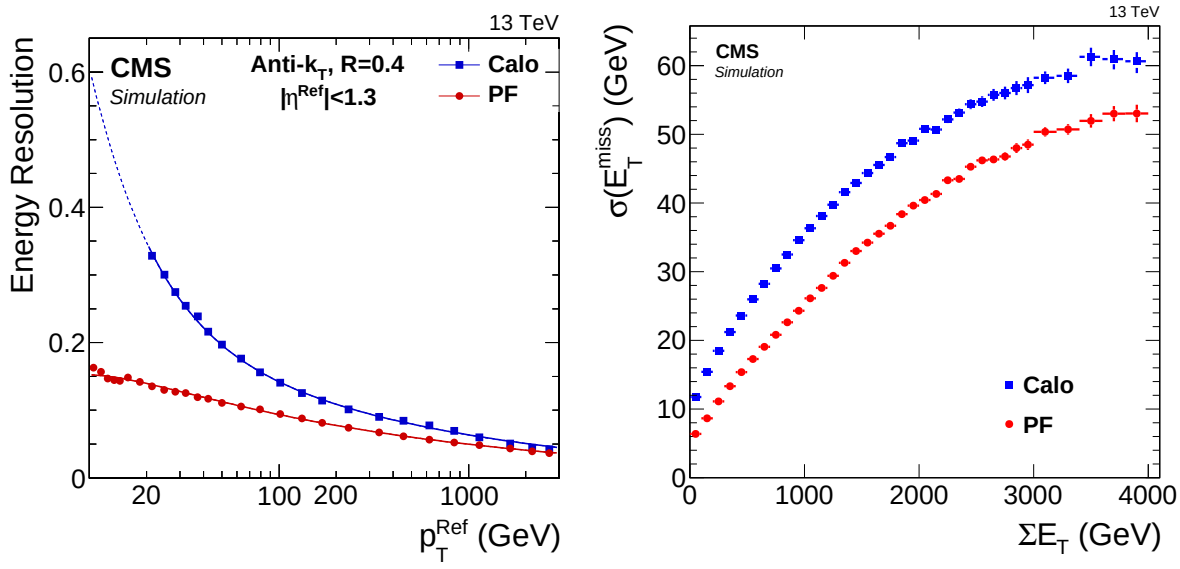


Fig. 4.7: Resolutions of the jet transverse momentum (*left*) and absolute missing transverse energy (*right*) [122]. E_T is the transverse energy sum of all objects that enter the calculation of $\not{E}_T$. The red lines indicate the resolutions for the *particle-flow* objects, while the blue ones correspond to jets and $\not{E}_T$ reconstructed using only the information from the calorimeters [145]. The PF quantities have systematically better energy resolutions.

In CMS, the procedure of JEC [140] involves multiple correction factors, that are sequentially applied to the reconstructed jet to obtain the calibrated energy (Fig. 4.6):

Offset Subtracts the electronics noise and pileup-induced contributions [142]. The pileup is estimated from an average energy density for each event, which for R = 0.4 jets is typically about 10 GeV during those runs considered in the present analysis.

Relative Corrects the relative jet response, i.e. the expected reconstructed energy over the original, along η to match the reference jets from $|\eta| < 1.3$. These corrections are obtained from simulation, but verified in data control samples.

Absolute Similarly, the jet response is made uniform in p_T by adjusting the absolute scale.

Residual Compensates remaining data-simulation discrepancies for jets in data only. Data driven corrections are derived from events with transverse momentum balance.

(Flavour) Systematic uncertainty not used in this analysis.

After all correction factors are applied to the reconstructed jet, the calibrated transverse momentum is required to be above 30 GeV for the jet to qualify for the analysis. Such selected jets will be referred to as *good jets* in the following. According to Fig. 4.7 (*left*), the energy resolution of these PF jets (red) is below 10% for $p_T > 30$ GeV, which is more than two times better, than the for jets reconstructed only based on the calorimeter information (red).

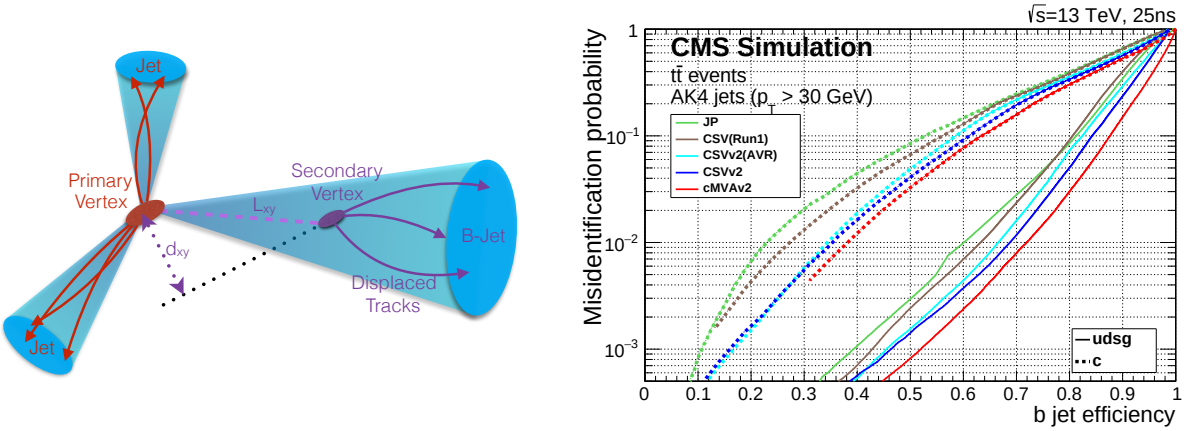


Fig. 4.8: *Left*: transverse view of a event containing two jets from the PV, and a b-jet composed of tracks originating from a secondary vertex. The SV is displaced by L_{xy}, and the transverse impact parameter d_{xy} is shown for a single displaced track.
Right: performance of the CMS b-jet identification algorithms[4]; the fraction of correctly identified b-jets (efficiency) is shown versus the probability to accept a non-b jet (misidentification). CSVv2, shown in blue, corresponds to the tagger used in the present analysis; the other algorithms are described elsewhere [141].

Identification of b-quark jets

Many processes within the SM and SUSY involve the production of top quarks, which immediately decay to a W-boson and b quark pair. Therefore, the identification of jets arising from b-quark hadronization, or *b-tagging*, is a crucial task for many analyses, including the presented. These techniques [141, 146] typically exploit the relatively long lifetime and high mass of the b quark, which results in a displaced vertex for the decay products (Fig 4.8, left).

The first step is to identify those secondary vertex (SV) candidates, that are likely to correspond to b-quark decays. The Inclusive Vertex Finder (IVF), introduced in Ref. [147], allows to reconstruct SVs based only on the available tracks, i.e. without requiring jets. Tracks with significant transverse impact parameters are used as clustering seeds for the adjacent tracks. The resulting clusters are fitted with the adaptive vertex fitter [128], and unique vertices are identified [141].

The information of displaced tracks and SVs associated to a jet is united by the Combined Secondary Vertex (CSV) [141] b-jet identification algorithm, or *tagger*. Since this is an optimized version of the *Run 1* CSV b-tagger [146], it is also referred to as *CSVv2*.

A multivariate technique is used to evaluate the probability for a jet to originate from b-quark hadronization. The input variables include parameters of the SV like its mass, the number of tracks, distance to the PV, but also the overall number of identified SVs and displaced tracks. Moreover, the total number of tracks within the jet is used. As a result, the CSV algorithm returns a single discriminating value (score). Figure 4.8 (right) shows the performance of the CMS b-taggers for various requirements on the score.

For the present analysis, the *medium* WP of the CSVv2 tagger is used, which corresponds to a misidentification probability of about 1% for good jets originating from light u,d,s quarks or

[4]For jets clustered with $R = 0.4$ using the anti-k$_T$ algorithm (AK4).

gluons. At this point, the b-tagging efficiency is about 69%.

4.3.4 Missing transverse energy

As soon as any weakly interacting neutral particle like a neutrino or neutralino is produced in the collision, it escapes the detector without depositing any energy. This creates an imbalance in the total momentum of the collisions, the magnitude of which is referred to as missing energy (cf. Sec. 2.2). Naturally, in searches for R-parity conserving SUSY, the missing energy is the key quantity of the event, and a direct handle to the undetectable particles.

In proton collisions, as already mentioned, the initial state has no transverse momentum, which has to be conserved by the interaction products. Consecutively, only the missing transverse momentum $\vec{\not{E}}_{\mathrm{T}}$ and its magnitude $\not{E}_{\mathrm{T}}$ are measurable at the LHC.

Like in the case of jets, the reconstruction of $\not{E}_{\mathrm{T}}$ relies on the particle-flow algorithm [148], which provides the most detailed description of the event. The missing transverse energy is simply defined as the negative vectorial sum of all PF candidates, except for those charged particles which are not associated with the PV:

$$\vec{\not{E}}_{\mathrm{T}} \equiv -\sum_{\mathrm{PF}} \vec{p}_{\mathrm{T}} \tag{4.5}$$

In addition to the aforementioned sources of *real* missing transverse energy, a momentum imbalance can be introduced through mismeasurements in the various subdetectors. A reduction of these effects is possible by replacing the energy contribution of PF candidates, that are clustered to a jet, with the calibrated energy of that jet. This correction can be summarized as:

$$\vec{\not{E}}_{\mathrm{T}}{}^{\mathrm{corr}} = \vec{\not{E}}_{\mathrm{T}}{}^{\mathrm{raw}} - \vec{\Delta}_{\mathrm{jets}} = -\sum_{\mathrm{jets}} \vec{p}_{\mathrm{T}}{}^{\mathrm{JEC}} - \sum_{\mathrm{uncl.}} \vec{p}_{\mathrm{T}} \tag{4.6}$$

where $\vec{\not{E}}_{\mathrm{T}}{}^{\mathrm{corr}}$ and $\vec{\not{E}}_{\mathrm{T}}$ denote the corrected and original missing transverse energy, $\vec{p}_{\mathrm{T}}{}^{\mathrm{JEC}}$ are the jet transverse moments with JEC corrections applied, and the last sum runs over all unclustered PF candidates. The jet sum includes only those with a corrected p_{T} above 15 GeV to avoid those possibly originating from pileup interactions. With those corrections, the $\not{E}_{\mathrm{T}}$ reconstruction profits from the jet energy calibration, reducing its uncertainty.

The *right* plot in Fig. 4.7 shows the expected missing transverse energy resolution as a function of the transverse energy sum of all objects entering the $\not{E}_{\mathrm{T}}$ calculation, E_{T}. The PF quantity (red) is compared to the missing energy reconstructed using only the calorimeter based information (blue).

Alternative methods for the missing energy reconstruction and corrections exist [148], but have not been used yet for *Run 2* data.

Chapter 5

Event Selection

5.1 Search signature 79
5.2 Data samples 80
5.2.1 Event cleaning 81
5.2.2 Trigger 82
5.3 Simulated samples 84
5.3.1 Background samples 85
5.3.2 Signal samples 85
5.3.3 Scale factors 87
5.4 Baseline selection 88
5.4.1 Control distributions 89
5.5 Search regions 91

In this chapter, the event selection for the analysis is presented. First, the target event signature is discussed and the key kinematic variables are introduced, motivating the use of the data samples and triggers shown in Sec. 5.2. The simulated samples used for the main SM background processes and signal models are discussed afterwards. Section 5.4 will discuss the selection criteria used to suppress the background events. Motivated by the kinematic and topological properties of the signal events, search regions are defined in the last section.

5.1 Search signature

As discussed earlier in Sec. 1.4.3, this analysis targets final states with four top quarks, where one of these quarks decays leptonically via a W boson, producing an isolated lepton and an undetectable neutrino. Together with the elusive neutralinos, an additional momentum imbalance at the scale of the gluino-neutralino mass difference is created. Since all top quarks decay to a b quark and a W boson, the final state also features multiple b-quark jets. Altogether, this provides a very distinctive signature for the event selection.

Key variables

Figure 5.1 illustrates the gluino production and decay within the discussed benchmark SUSY model. Two main quantities characterize the energy scale of the event:

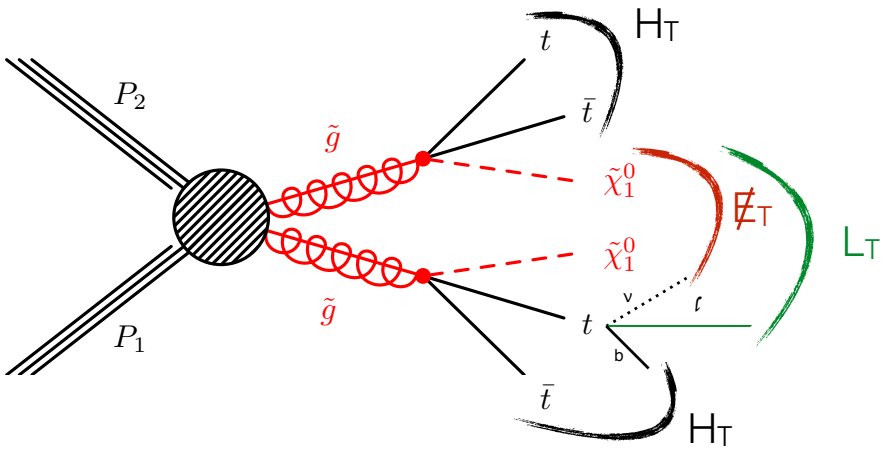


Fig. 5.1: Illustration of gluino pair production and decay within the T1tttt model. The energy scale of the event is characterized by: the multitude of jets creating a large hadronic energy H_T, while the missing transverse energy $\not{E}_T$ from the neutrino and neutralinos add up with the lepton momentum into the leptonic energy scale L_T.

- the multitude of jets, originating from top-quark decays, creates a large hadronic energy, which is measured by the scalar sum of the jet transverse momenta:

$$H_T = \sum_{\text{jets}} p_T \tag{5.1}$$

- the missing transverse energy from the neutrino and neutralinos together with the lepton momentum represents the leptonic energy scale:

$$L_T = \not{E}_T + p_T^{\text{lep}} \tag{5.2}$$

In typical SM events containing a single isolated lepton, the hadronic and lepton energy scales tend to be low in comparison to the considered SUSY scenarios. Since leptons and jets are relatively distinct objects and reasonably fast to reconstruct, the trigger selection should rely on them.

5.2 Data samples

The data used for this analysis were collected in pp-collisions at a centre-of-mass energy of 13 TeV during the *Run 2* of the LHC using the CMS detector. Figure 5.2 (left) shows the cumulative distribution of integrated luminosity delivered to and recorded by CMS in the year of 2015. Due to a malfunction of the CMS cryogenic system, not all of the data were recorded at the nominal magnetic field. Given the importance of the transverse momentum measurement for this analysis, only data collected at the full field strength $B = 3.8\,\text{T}$ is used.

An additional filtering of events based on their quality was performed by the DQM experts (cf. Sec 2.2.6). In the present analysis only certified data ranges were used, as indicated by

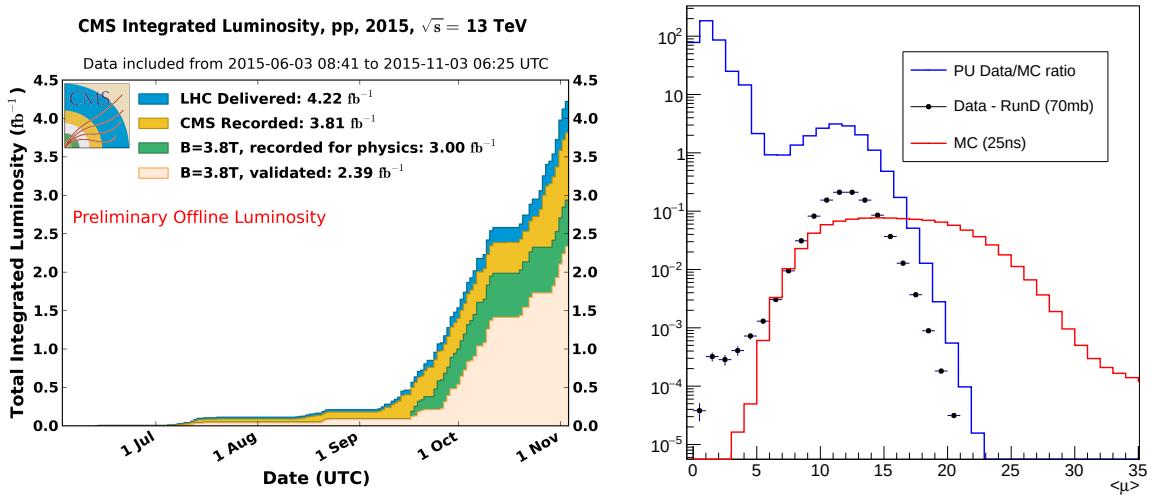


Fig. 5.2: *Left*: Cumulative integrated luminosity delivered to (blue), and recorded by CMS (orange) during pp-collisions at a centre-of-mass energy of 13 TeV in 2015 [149]. The green histogram shows the amount of data collected at the full magnetic field, while the fraction of data validated by the DQM experts is indicated in beige.
Right: Mean number of interactions per bunch crossing during the analysed data runs (black) and as simulated in the MC samples (red). For the data distribution the latest luminosity calibrations [96] and an inelastic pp cross-section of 70 mb were used.

the beige histogram in Fig. 5.2 (right). The mean number of interactions per bunch crossing during the analysed runs is shown with black dots in Fig. 5.2 (left). On average, there were about eleven interactions per bunch crossing.

5.2.1 Event cleaning

The DQM certification handles the data in time ranges rather than in single events. Therefore, a small fraction of single events with collective detector noise or contamination from beam stray radiation may still remain in the validated dataset. So-called *event filters* are applied to reject those abnormal events [150].

Beam Halo Particles from the beam halo might interact with the accelerator instruments or shielding. The resulting muons can pass through CMS in parallel to the beam, depositing non-characteristic energies in the calorimeters, which negatively affect the reconstruction algorithms. Those beam halo muons are identified by the CSC system in the endcaps.

Detector noise Electronics noise can arise in any readout channel anytime. However, if this happens in an ensemble of adjacent channels, the reconstruction is not able to filter out such noise. Special algorithms can identify spurious noise of calorimeter clusters in the cells of HB, HE or EB, as well as abnormal tracks.

Overall, the effect of those filters is relatively small with respect to the total amount of analysed events.

5.2.2 Trigger

Events are selected by two HLT paths requiring a loosely isolated muon or electron with p_T above 15 GeV in addition to an H_T greater than 350 GeV:

- `HLT_Ele15_IsoVVVL_PFHT350` **OR** `HLT_Mu15_IsoVVVL_PFHT350`

where H_T is computed from central PF jets ($|\eta| < 2.4$) with $p_T > 40$ GeV at the HLT level. The lepton candidates are required to pass very loose identification requirements and the relative isolation has to satisfy $I_{R=0.2} < 0.8$ (1.2) for electrons (muons). These HLT paths are seeded by a pure $H_T > 175$ GeV L1 trigger, which exhibits a better performance, than leptonic seeds at the same level. Events collected with these trigger paths are allocated in the `SingleElectron` and `SingleMuon` datasets, respectively.

Out of precaution, these low-threshold trigger paths were initially prescaled during the early stages of *Run 2* data-taking. A backup trigger strategy was developed and evaluated as presented in Appendix A. For this search, only the data with no prescaled triggers are used, corresponding to the LHC run period with a 25 ns bunch crossing interval. The resulting total integrated luminosity of this dataset is $\mathcal{L} = 2.3\,\text{fb}^{-1}$, as calculated using the latest calibration of the luminosity instruments [96].

Trigger efficiency

The measurement of trigger efficiencies is performed with a "cut-and-count" method using independent reference trigger paths. A baseline selection is defined using these reference triggers with an additional preselection. The efficiency is then defined as the fraction of events that pass the analysis trigger with respect to the baseline selection:

$$\epsilon = \frac{N(\text{all events passing analysis trigger + preselection + reference trigger})}{N(\text{all events passing preselection + reference trigger})} \tag{5.3}$$

In general, the baseline selection requires the presence of a single muon or electron as described in Sec. 4.3.2, which will be called *good lepton*. Since the analysis triggers contain two different physics objects, or legs, – the lepton and the jet momentum sum – the efficiency can be factorized for the two legs and measured independently.

First, the lepton p_T leg efficiency is measured with the pure H_T trigger `HLT_PFHT800`, which shares the same L1 Trigger seed as the analysis paths and requires $H_T > 800$ GeV. Given the same variable but higher threshold, the reference trigger cancels out the hadronic leg of the analysis triggers: `PFHT350`. Therefore, the remaining inefficiency is only due to the leptonic leg. No additional requirements except the baseline selection with a good lepton are used.

Figure 5.3 shows the trigger paths' efficiencies versus the p_T of the corresponding lepton object. A plateau of 94% is reached for muons with $p_T > 18$ GeV, which is 3 GeV above the HLT requirement of 15 GeV. For electrons the turn-on is slightly slower due to the more complex reconstruction, but still only 7 GeV above the trigger threshold. Overall, both trigger paths show a similar efficiency in the plateau region over a wide p_T range.

For the H_T leg efficiency measurement the reference trigger is chosen to be a single isolated electron or muon trigger: `HLT_Ele23_WPLoose_Gsf` or `HLT_IsoMu20`. In addition, the good lepton p_T is required to be above 25 GeV to ensure that the lepton legs are in their plateau regions. Figure 5.4 shows the H_T leg efficiency of the two analysis trigger paths. Both triggers become

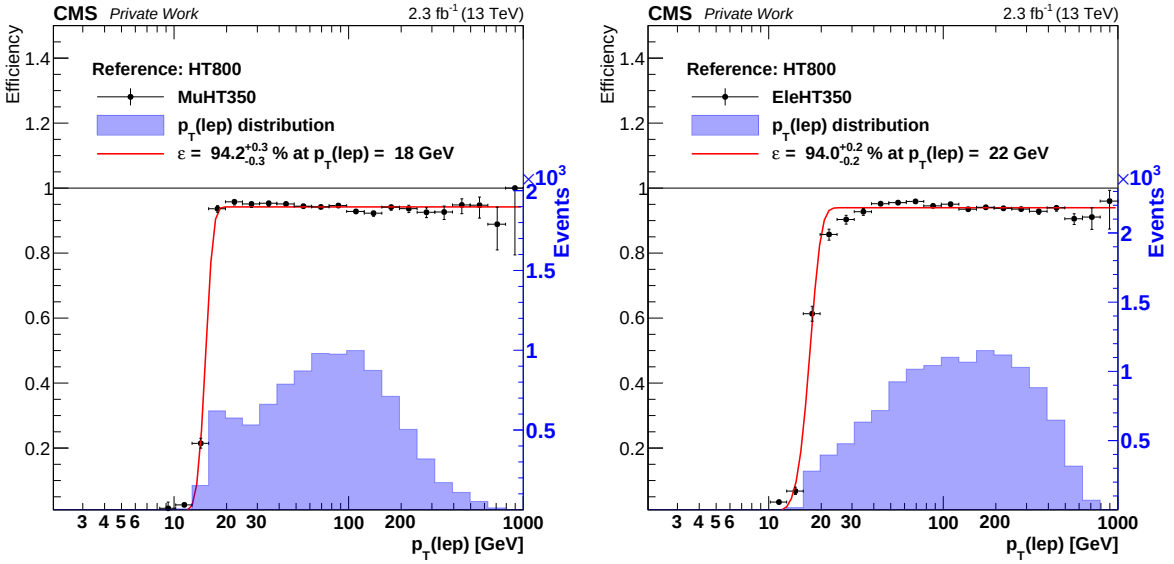


Fig. 5.3: Efficiency of the lepton p_T leg for the analysis HLT paths including a muon `HLT_Mu15_IsoVVVL_PFHT350` (left) or an electon `HLT_Ele15_IsoVVVL_PFHT350` (right). The baseline selection requires the events to pass a pure H_T trigger `HLT_PFHT800` and contain a single good lepton with $p_T > 10\,\text{GeV}$.
The number of events passing the triggers is shown in the blue histograms.

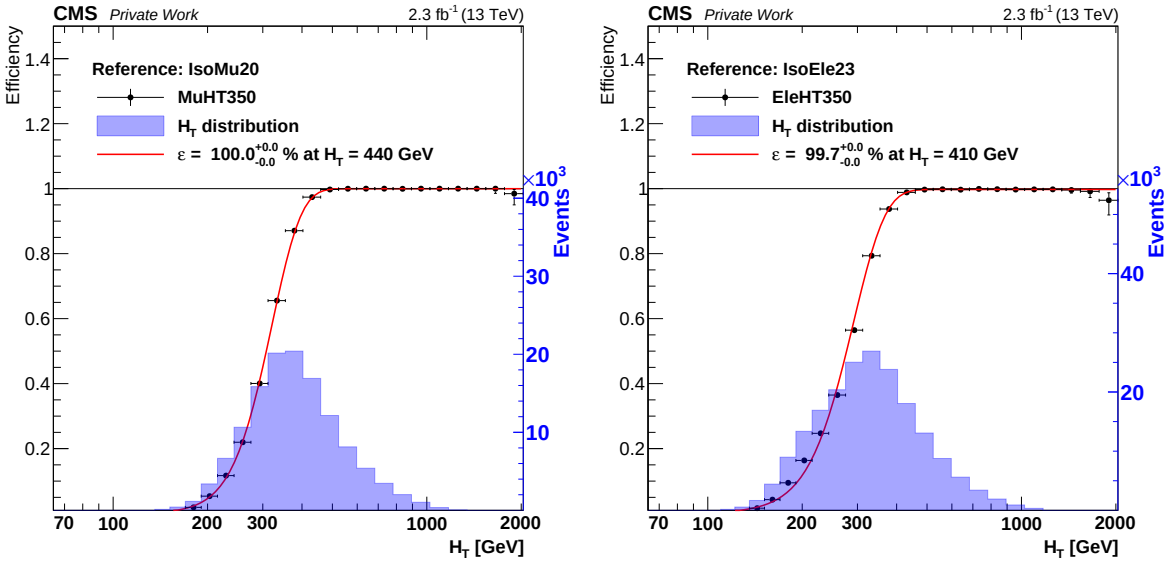


Fig. 5.4: Efficiency of the H_T leg for the analysis HLT paths including a muon `HLT_Mu15_IsoVVVL_PFHT350` (left) or an electon `HLT_Ele15_IsoVVVL_PFHT350` (right). The baseline selection requires the events to pass a corresponding single isolated lepton trigger (`HLT_IsoMu20` or `HLT_Ele23_WPLoose_Gsf`) and contain a single good lepton with $p_T > 25\,\text{GeV}$.
The number of events passing the triggers is shown in the blue histograms.

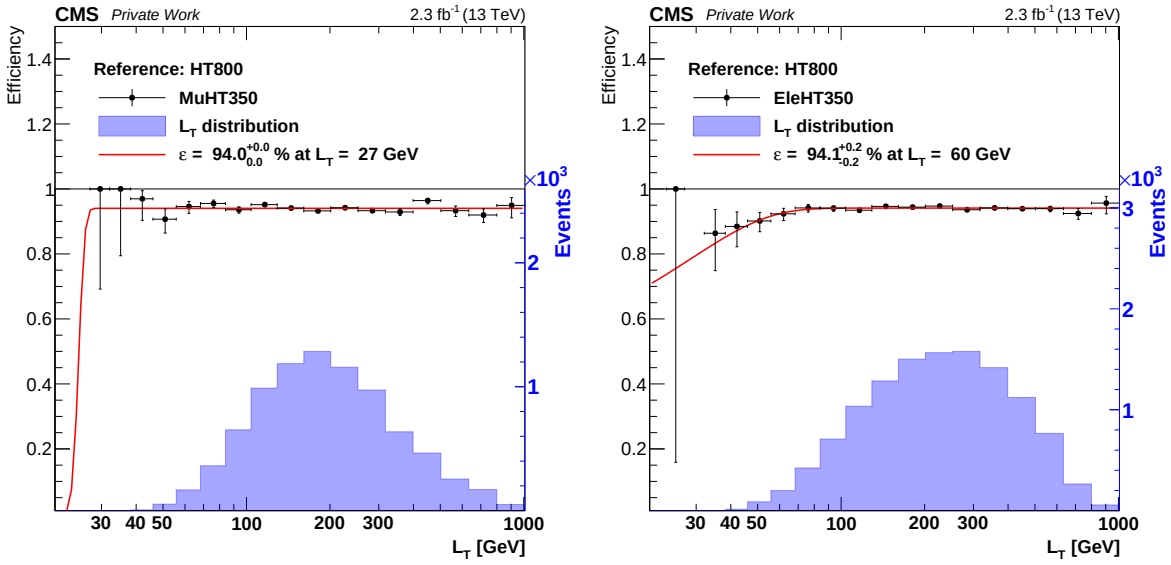


Fig. 5.5: Event selection efficiency as a function of the L_T variable for the analysis HLT paths including a muon `HLT_Mu15_IsoVVVL_PFHT350` (left) or an electon `HLT_Ele15_IsoVVVL_PFHT350` (right). The baseline selection requires the events to pass a pure H_T trigger `HLT_PFHT800` and contain a single good lepton with $p_T > 25\,\text{GeV}$. The number of events passing the triggers is shown in the blue histograms.

fully efficient after $H_T > 410$ (440) GeV for the electron (muon) path, while the turn-on of the electron path is faster, since the HLT algorithms also include the electron p_T into the calculation of H_T.

Finally, since the analysis also relies on the L_T variable, which is the scalar sum of $\not{E}_T$ and p_T^{lep}, the selection efficiency is also evaluated with respect to this quantity.

Here, the same reference trigger as for the lepton p_T leg measurement is used. In addition, the good lepton is required to satisfy $p_T > 25\,\text{GeV}$ to avoid the inefficient region. The final selection efficiency as a function of L_T is shown in Fig. 5.5. Since the variable incorporates two different quantities – the transverse lepton momentum and event missing energy – the efficiency turn-on differs slightly from the lepton leg p_T. Nevertheless, a plateau of 94% is reached for L_T well below 100 GeV for both trigger paths. A conservative uncertainty of 1% is assumed for the total trigger efficiency in both cases.

5.3 Simulated samples

This section introduces the simulated samples used in the present analysis. First, the samples for SM processes that represent the main backgrounds are listed. The signal samples are produced for a variety of different gluino-neutralino mass points and are shown afterwards. The parton showering and hadronization are always simulated with the PYTHIA 8.2 software package using the CUETP8M1 parameter tune [151].

The pileup of the 25 ns LHC running is modelled with minimum-bias events, that are mixed with the hard scattering process. The number of additional interactions is sampled from a pre-

Table 5.1: 13 TeV cross-sections of SM background processes versus two signal points [52, 152].

Label	Explanation	σ [pb^{-1}]
QCD	QCD multijet events	$O(10^9)$
W+jets	leptonic W boson with jets	61526
DY	Drell-Yan production of Z/$\gamma*$ ($M_{ll} > 50$ GeV)	6025
$t\bar{t}$ + jets	Top quark pair production	831
single top	Single top quark in the t-, s- and tW channels	145
$t\bar{t}V$	$t\bar{t}$ in association with a W or Z boson	1.4
$m_{\tilde{g}} = 1200$	Gluino pair production as in T1tttt	0.086
$m_{\tilde{g}} = 1500$		0.014

defined pileup distribution with a mean of 20 as shown in Fig. 5.2 (right).

5.3.1 Background samples

The SM processes are mainly produced using the MADGRAPHv5 event generator in the LO approximation. For the backgrounds the full CMS detector simulation as explained in Sec. 2.3.2 is used. Several samples are generated in H_T ranges to enhance statistics in the high H_T phase-space. This H_T is computed at the generator level using the transverse momenta of partons originating from the hard process. The full list of samples is as follows:

$t\bar{t}$ MADGRAPH: several samples binned in H_T plus dedicated semi- and dileptonic samples to enhance the statistics.

W+jets MADGRAPH: leptonically decaying W, binned in H_T.

single top The tW-channel samples are produced with POWHEG, the t- and s-channels with AMC@NLO.

$t\bar{t}V$ AMC@NLO: $t\bar{t}$ associated with W or Z in leptonic and hadronic decays.

Drell-Yan MADGRAPH: leptonic Z/$\gamma*$ decays in bins of H_T.

QCD MADGRAPH: QCD multijet events in bins of H_T.

Table 5.1 summarizes the selected SM background processes together with the theoretical cross sections at a centre-of-mass of 13 TeV. For each process (or sub-process), the corresponding cross section is calculated at the highest available precision. Overall, the equivalent luminosity of all samples exceeds the one of the used data at least six times, while the $t\bar{t}$ sample has hundred times more statistics.

5.3.2 Signal samples

The signal samples are simulated according to the simplified SUSY model T1tttt discussed in Sec 1.4.1. Since this model contains two free parameters – the masses of the gluino $m_{\tilde{g}}$ and neutralino $m_{\tilde{\chi}_1^0}$ – a separate sample is produced for each pair. This *scan* of the parameter space normally requires more computational resources than standard SM processes, hence the CMS fast simulation (cf.. Sec 2.3.2) is used to produce the signal samples.

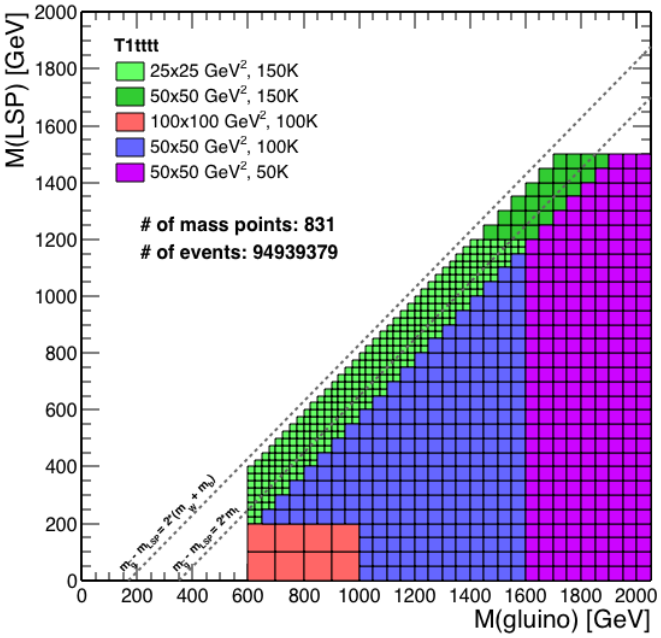


Fig. 5.6: Produced signal mass points for the simplified T1tttt model from Ref. [153]. This model features a 3-body decay of the gluino $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$, where $\tilde{\chi}_1^0$ is the LSP. For points between the dashed lines one of the top quarks is produced off-shell.

The MADGRAPHv5 event generator is used for the modelling of the hard scattering as well as the subsequent decays. The samples are normalized to the most accurate cross section available to date [55], and the equivalent luminosity of the signal samples is at least hundred times that of the data.

For this simulation the directly produced gluinos decay through a 3-body decay to a top-quark pair together with a neutralino: $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$. Further, the top quarks decay according to the SM, i.e. in the majority of cases to a b quark and W boson. For $m_{\tilde{\chi}_1^0}$ above $m_{\tilde{g}} - 2m_t$ one of the top quarks is produced off-shell. Therefore, ths neutralino mass is restricted to be less than $m_{\tilde{g}} - 2(m_b + m_W)$, which allows for two off-shell top quark decays.

Figure 5.6 shows the mass-point grids of these two simplified models with the mass boundaries indicated by the dashed lines. The finest binning is chosen for the diagonal mass points, where the spectrum is *compressed* and one of the top quarks is produced off-shell. For the masses already excluded during *Run 1* (cf. Sec. 1.4.2) are more coarse binning is chosen.

Throughout this thesis, two mass points of the T1tttt model are used as benchmark scenarios:

T1tttt (1200,800) with $m_{\tilde{g}} = 1200\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 800\,\text{GeV}$ represents a point in the *compressed* parameter phase-space, where the top quarks are produced at the mass shell boundary with low boost. This results in a softer spectrum of the respective decay products.

T1tttt (1500,100) with $m_{\tilde{g}} = 1500\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 100\,\text{GeV}$ covers the *non-compressed* phase-space, where both on-shell top quarks have high energy decay products.

The corresponding cross-sections are shown in Tab. 5.1 in comparison with the main SM backgrounds. Both points are beyond the reach of *Run 1* analyses.

5.3.3 Scale factors

Naturally, the modelling of real physics processes as well as the simulation of the detector responses is not perfect. The study and understanding of the sources of those discrepancies helps to improve the data-MC agreement. For this analysis event-by-event weights, so-called *scale factors*, are used to correct various differences between the simulation and data. In case of full or fast simulation, the corresponding corrections are used.

Top quark p_{T} spectrum

As observed during the 8 TeV LHC run [154], the current MC generators tend to overestimate the number of $\mathrm{t\bar{t}}$ events for high transverse momenta of the top quark. Since no detailed studies have been performed yet at a centre-of-mass energy of 13 TeV, the *Run 1* results are taken to correct the simulated top quark p_{T} spectrum in events with a $\mathrm{t\bar{t}}$ pair [155].

B-tagging efficiency

The efficiency of detecting jets from b-quark hadronization differs in the simulation and real samples for various reasons related to the simulation and reconstruction accuracy. Several methods exist to correct for such discrepancies [156] based on data-driven measurements. A comparison of methods is discussed in Appendix E.

In this analysis, for each event a weight is calculated, which reflects the probability of the observed number of b-tags given the efficiency to identify b jets and the total number of jets in the event.

Lepton identification and reconstruction efficiency

For similar reasons, the identification and reconstruction efficiencies of leptons has to be corrected in simulated samples. Generally, p_{T} and η dependent scale factors are applied depending on the good lepton [157].

Pileup

As already mentioned, the number of additional interactions is sampled based on a predefined distribution. Since the data and MC distributions differ, the simulated distribution is rescaled to match the one obtained in data. Fig. 5.2 (right) shows the pileup distributions for the analysed data and MC samples, together with their ratio as used for the event weights.

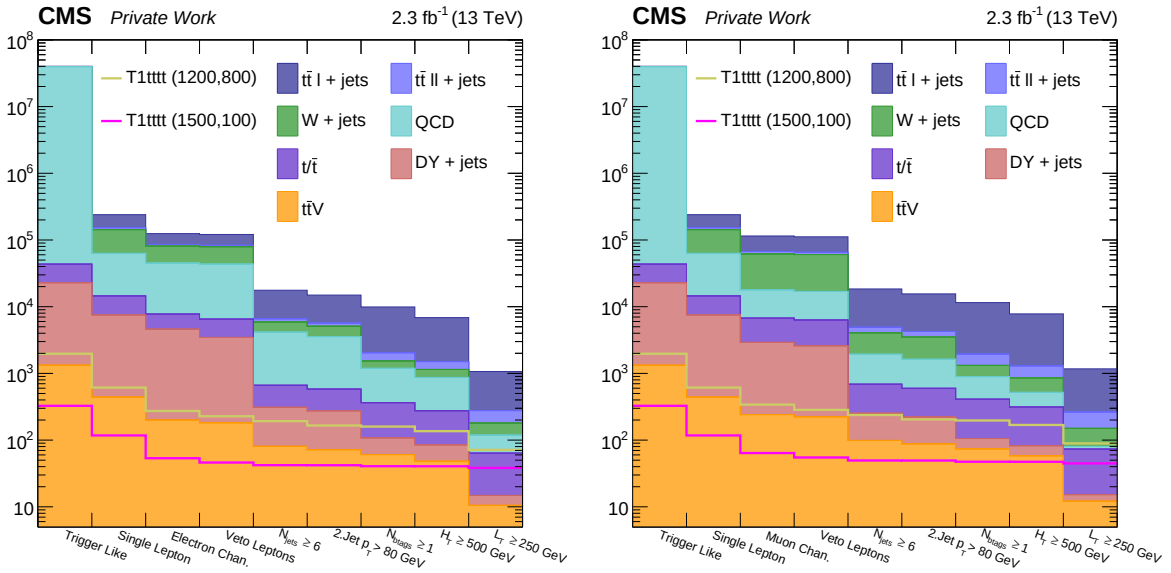


Fig. 5.7: Number of events for each sample, scaled to the luminosity by the respective cross section, after each requirement step of the baseline selection in the electron (left) and muon (right) channel. While the SM backgrounds get substantially suppressed, the two signal benchmark scenarios loose only a fraction of events.

5.4 Baseline selection

The basic signature of the considered SUSY models were already discussed previously in Sec. 5.1. To comply with the trigger thresholds and achieve an even further rejection of SM backgrounds, the following baseline event selection is adopted. An illustration of the requirements' impact on the number of selected events is shown in Fig. 5.7.

Leptons

Events are required to contain exactly one good lepton as described in Sec. 4.3.2: electrons and muons must pass the *tight* and *medium* WP of the respective ID criteria and lie within the tracker acceptance $|\eta| < 2.4$. The isolation is enforced with an upper limit on the mini-isolation of 0.1 (0.2) for electrons (muons). The transverse momentum of the leptons must be above 25 GeV to ensure an efficient trigger selection. Leptons, that satisfy these requirements will be labelled as *selected* in the following.

To ensure that the final state contains exactly one prompt lepton, events with additional leptons passing looser criteria are vetoed. PF lepton candidates with $p_T > 10$ GeV, looser identification criteria and $I_{mini} < 0.4$ are counted as veto leptons.

Jet selection

Jets are selected with $p_T > 30$ GeV and $|\eta| < 2.4$ as described in Sec. 4.3.3. In order to avoid double counting of objects, jets are removed, if they are close[1] to either a veto or selected

[1] $\Delta R < 0.4$ between the lepton and jet candidates.

Table 5.2: Event selection criteria and object requirements. The definition of good leptons, jets and missing transverse energy is provided in Sec. 4.3.

Requirement	Objects
single isolated lepton	Good leptons with $p_T > 25\,\text{GeV}$, $\lvert\eta\rvert < 2.4$ and $I_{\text{mini}} < 0.1$ (0.2) for electrons (muons).
no veto leptons	Loose leptons with $p_T > 10\,\text{GeV}$, $\lvert\eta\rvert < 2.4$ and $I_{\text{mini}} < 0.4$.
$N_{\text{jets}} \geq 6$	Good jets with $p_T > 10\,\text{GeV}$, $\lvert\eta\rvert < 2.4$ cleaned from lepton overlap.
p_T (2. jet) $> 80\,\text{GeV}$	
$N_{\text{bjets}} \geq 1$	Good jets satisfying the *medium* WP of the CSVv2 tagger.
$H_T > 500\,\text{GeV}$	$H_T \equiv \sum_{\text{jets}} p_T$
$L_T > 250\,\text{GeV}$	$L_T \equiv \not{E}_T + p_T^{\text{lep}}$

lepton. Given the intrinsically high jet multiplicity of signal events, at least six good jets are required. Except for the FastSim samples, the jet ID (cf. Se. 4.3.3) is applied. In this phase-space SM processes must already contain additional jets from ISR.

Since the signal features three hadronically decaying top quarks, the resulting jets must be energetic. Therefore, the transverse momenta of the first two hardest jets must satisfy $p_T > 80\,\text{GeV}$. In addition, at least one b-tagged jet (cf. Sec 4.3.3) is required to filter events containing top quarks.

Energy scale

As already stated in the beginning of the section (cf. Sec. 5.1), the large leptonic and hadronic energy scales are a key feature of the signal models. The hadronic scale H_T, which is calculated from the selected jets' transverse momenta, should be at least 500 GeV. A requirement of $L_T > 250\,\text{GeV}$ is imposed on the leptonic scale, which is the scalar sum of the lepton p_T and the magnitude of the missing transverse energy $\not{E}_T$. Above those thresholds, as shown previously, the triggers show a flat efficiency with respect to the variables.

A summary of the event selection criteria and object requirements is shown in Tab. 5.2.

5.4.1 Control distributions

Distributions of the main kinematic variables H_T and L_T are shown in Fig. 5.8 (top) and show a reasonable MC-data agreement. The two selected signal benchmark scenarios exhibit differences in the distributions, that are related to the gluino-neutralino mass splitting. In case of the non-compressed scenario T1tttt (1500,100) the created top quarks have a large boost, resulting in high leptonic and hadronic energy scales. For the compressed region with T1tttt (1200,800) no such effect is observed, and the distributions look very similar to the SM $t\bar{t}$ production.

The bottom plots of Fig. 5.8 compare the multiplicities of jets and b-tagged jets in data and simulated events. As expected, no difference between the signal scenarios is observed, since the mass splitting has no impact on the decay topology. Additional control distributions are shown in Appendix C. Overall, a good modelling of the data by the simulation is observed.

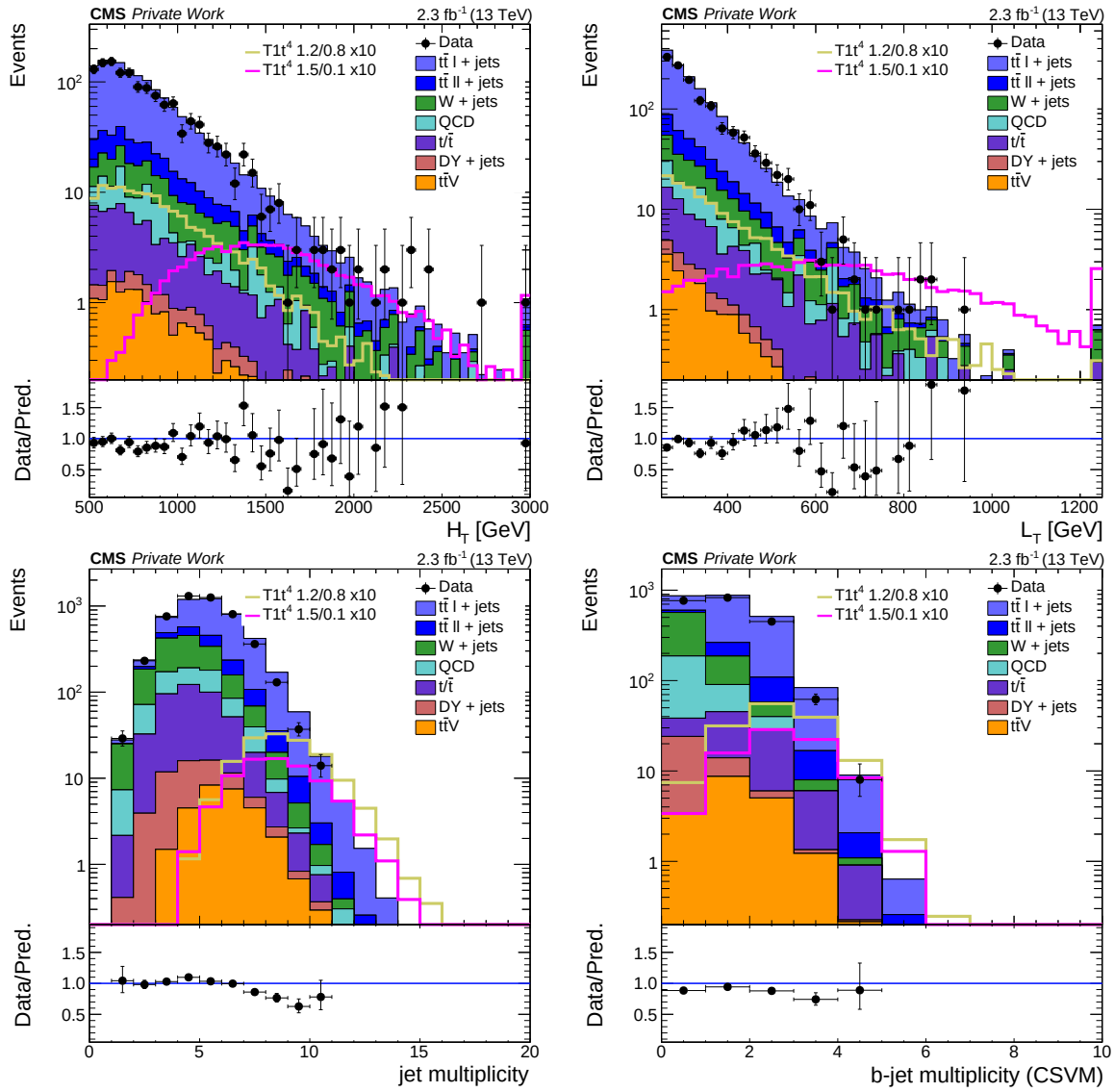


Fig. 5.8: Distributions of the H_T (left) and L_T (right) variables after the baseline selection in the top row. The bottom plots show the jet (left) and b-tag (right) multiplicities without the respective criteria applied besides the preselection. In the top plots the long tails of the non-compressed T1tttt (1500,100) signal distributions are visible, while the more compressed T1tttt (1200,800) scenario rather resembles the SM $t\bar{t}$ process. The jet and b-tag multiplicities are independent of the gluino-neutralino mass splitting, and generally higher than of the background.
The data is shown in black dots and the shaded histograms represent the simulated SM background processes stacked on top of each other. Two benchmark signal points are scaled up by a factor of ten and shown by the coloured lines. The bottom pad shows the ratio of the data to the total background.

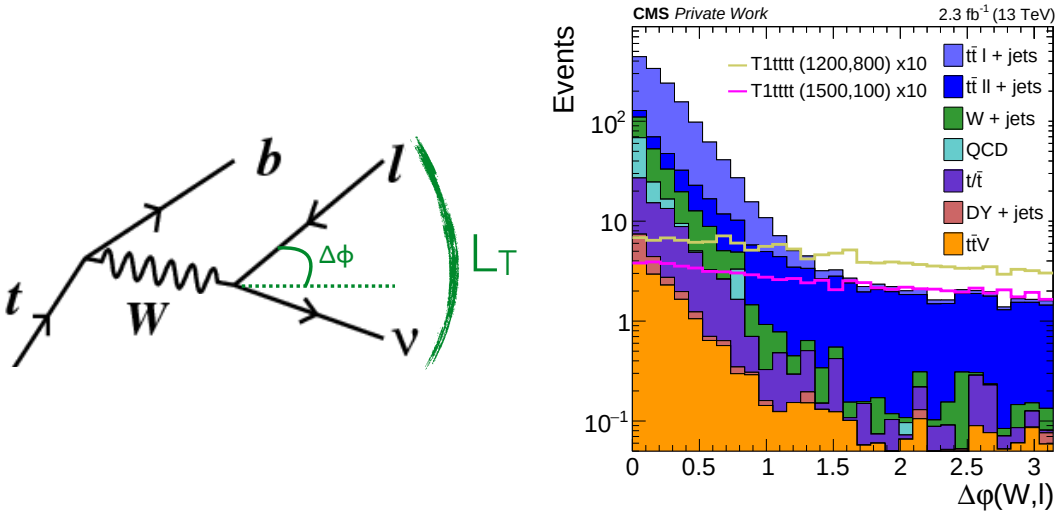


Fig. 5.9: *Left*: topology of a top quark and subsequent leptonic W-boson ecay. $\Delta\varphi$ shows the azimuthal angle between the reconstructed boson and lepton. L_T is the scalar sum of the missing energy from the neutrino and the lepton p_T.*Right*: distribution of the $\Delta\varphi$ variable after the baseline selection. Due to the large and not lepton-aligned $\not{E}_T$, SUSY events give a flat $\Delta\varphi$ distribution. The filled histograms of the SM backgrounds are stacked on top of each other, and the signal benchmarks scaled by ten are shown in the solid lines.

5.5 Search regions

After the preselection, the main background events are the semileptonic $t\bar{t}$+jets followed by dileptonic $t\bar{t}$+jets and leptonically decaying W+jets. Common for the SUSY and SM events are the W bosons decaying into a lepton and neutralino pair as shown in Fig. 5.9 (left). In absence of any other undetectable particles, the missing energy provides an estimate of the neutrino momentum. As a result, the kinematics of the W-boson candidate can be partially reconstructed as the vectorial sum of the lepton p_T and $\not{E}_T$.

It was previously observed [10], that the azimuthal angle $\Delta\varphi$ between the lepton and W is small, since the daughter particle is aligned with its mother. For events with large $\not{E}_T$ from additional neutralinos, like in the signal discussed, the reconstructed W-boson p_T does not represent the boson any more. Due to the random direction of the neutralinos, the resulting angle $\Delta\varphi$ distribution becomes flat for SUSY events. This makes the $\Delta\varphi$ angle a powerful discriminating variable for the signal events.

Signal region definition

Figure 5.9 (right) shows the simulated $\Delta\varphi$ distribution for the main SM processes in comparison to the two signal benchmarks. As expected, most background events tend towards low values of the angle, while the signals have a flat distribution. This difference allows to use the large-$\Delta\varphi$ events to search for these hypothetical SUSY models. The remaining SM background, which is dominated by dileptonic $t\bar{t}$ events with two neutrinos, has to be estimated as shown in the next Chapter 6.

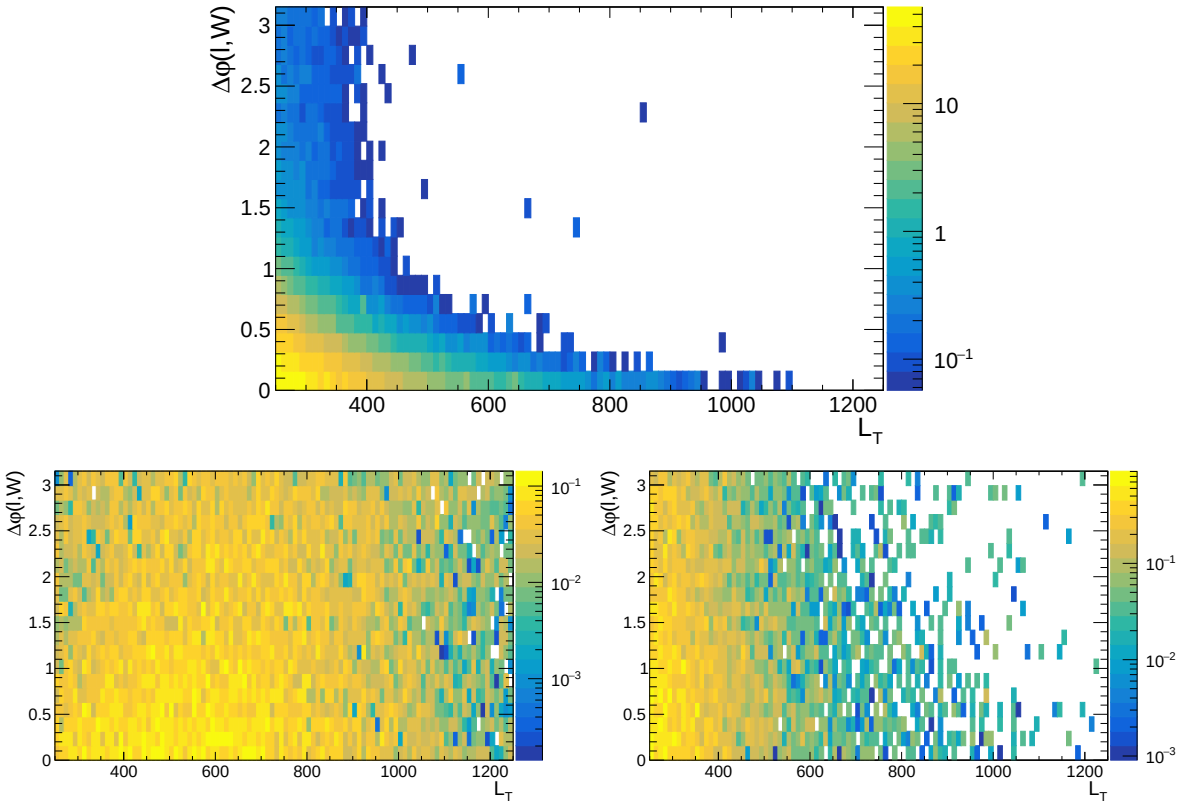


Fig. 5.10: Distribution of simulated *EWK* background (*top*) and signal T1tttt (1500,100) (*left*), T1tttt (1200,800) (*right*) events in the the $\Delta\varphi$ vs. L_T plane after the preselection. While the signal shows no correlation of those variables, the $\Delta\varphi$ of background events decreases towards large values of L_T.

Since the $\Delta\varphi$ and L_T variables originate from the same two reconstructed objects – p_T^{lep} and $\not{E}_T$ – a certain correlation has to be assumed. In SM events with a single leptonically decaying W boson the lepton becomes more aligned with increasing boost of the mother particle. Therefore, the $\Delta\varphi$ distribution squeezes towards zero with increasing L_T, while the signal shape does not follow such a trend, since the $\not{E}_T$ stems mostly from the neutralinos. Figure 5.10 confirms this assumption by comparing the background shape in the $\Delta\varphi$-L_T plane with the two benchmark scenarios.

This behaviour allows to define the signal region requirements depending on the L_T interval such, as to keep a good signal efficiency, while rejecting more background. In the following, the term search region (SR) will denote events with a $\Delta\varphi$ above a certain threshold value, defined as:

$$L_T < 350\,\text{GeV} \quad \Delta\varphi\,(\text{SR}) > 1.0$$
$$350 < L_T < 600\,\text{GeV} \quad \Delta\varphi\,(\text{SR}) > 0.75$$
$$L_T > 600\,\text{GeV} \quad \Delta\varphi\,(\text{SR}) > 0.5$$

Events, that fail these requirements lie within the control region (CR).

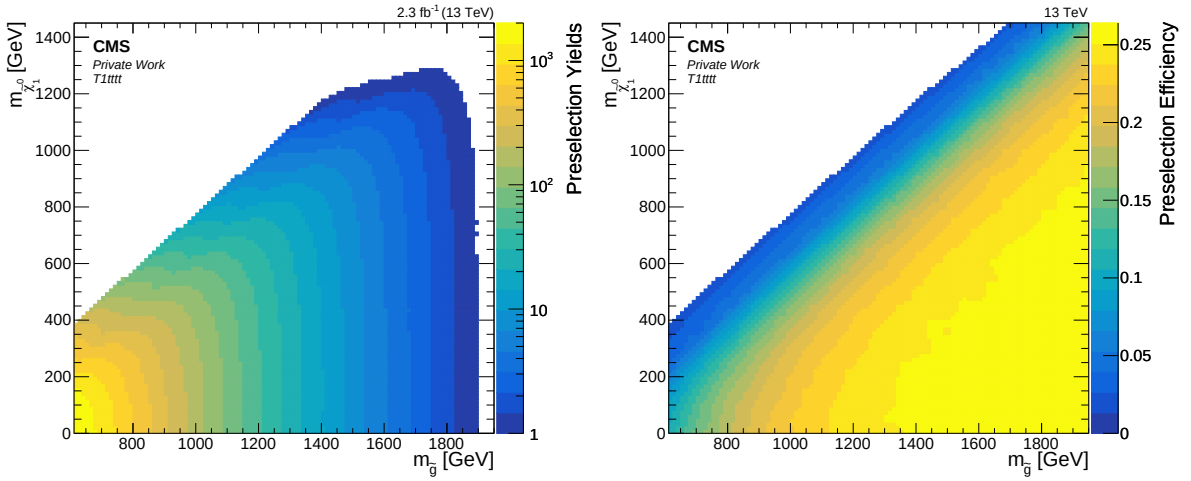


Fig. 5.11: T1tttt signal yields (*left*) and selection efficiency (*right*) after the baseline requirements. The missing mass points are interpolated from the existing yields.

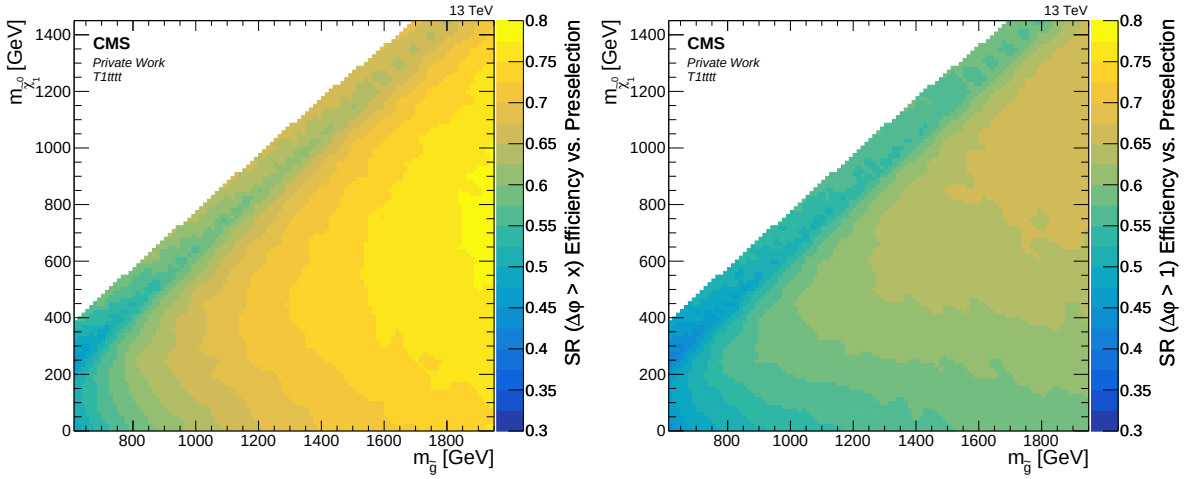


Fig. 5.12: Signal region selection efficiency with respect to the preselection for the L_T-dependent $\Delta\varphi$ threshold x (*left*) and constant $\Delta\varphi > 1$ (*right*) for the T1tttt model. The control region efficiencies are respectively inverse.

Signal acceptance

The number of signal events after the baseline selection is shown in Fig. 5.11 (*left*) in the gluino-neutralino mass plane for the T1tttt model. At least one event at the given luminosity would be accepted up to a $m_{\tilde{g}} = 1850\,\text{GeV}$ for $m_{\tilde{\chi}_1^0} < 1200$ GeV. The corresponding selection efficiency, i.e. the fraction of selected with respect to the total number of produced events is shown in the *right* plot.

The SR efficiency relative to the preselection is shown in Fig. 5.12 for a constant and L_T-depending $\Delta\varphi$ threshold. While the $\Delta\varphi > 1$ requirement yields a uniform efficiency around 50-60%, the variable threshold allows for an increased acceptance reaching up to 80% for high gluino masses.

Table 5.3: Signal search region boundaries and the corresponding lower $\Delta\varphi$ thresholds.

n_{jet}	$n_{\text{b-tag}}$	L_{T} [GeV]	H_{T} [GeV]	$\Delta\varphi$
[6,8]	$= 1, = 2, \geq 3$	$[250, 350]$	$[500, 750]$, ≥ 750	1.0
		$[350, 450]$	$[500, 750]$, ≥ 750	0.75
	$= 1, \geq 2$	$[450, 600]$	$[500, 1250]$, ≥ 1250	
		≥ 600	$[500, 1250]$, ≥ 1250	0.5
≥ 9	$= 1, = 2$	$[250, 350]$	$[500, 1250]$, ≥ 1250	1.0
	≥ 3		≥ 500	
	$= 1, = 2, \geq 3$	$[350, 450]$	≥ 500	0.75
	$= 1, \geq 2$	≥ 450	≥ 500	

Search categories

In order to further increase sensitivity to the various signal mass scenarios, the search is performed in multiple orthogonal search regions, or bins. As already shown in the control distributions of Sec. 5.4.1, the compressed and non-compressed models populate different regions in H_{T} and L_{T} (Fig. 5.8, *top*). In addition, the different jet and b-tag multiplicities allow to discriminate regions that are enriched in signal events (Fig. 5.8, *bottom*).

Based on these ideas, the binning is chosen as follows:

$$L_{\mathbf{T}} \in \text{LT1: } [250, 350], \text{ LT2: } [350, 450], \text{ LT3: } [450, 600], \text{ LT4: } \geq 600 \text{ GeV}.$$

$$H_{\text{T}} \in \text{HT0: } [500, 750], \text{ HT1: } [750, 1250], \text{ HT2: } \geq 1250 \text{ GeV}.$$

$$n_{\mathbf{jet}} \in \text{6-8, } \geq 9$$

$$n_{\mathbf{b\text{-}tag}} \in \text{1,2, } \geq 3$$

where LT1-4 and HT0-2 are the labels for the corresponding L_{T} and H_{T} ranges.

The final categories are formed to avoid extreme combinations such as high H_{T} with low or medium L_{T} and vice versa. In addition, the bins should contain enough events at the given data luminosity of 2.3 fb^{-1}. Table 5.3 illustrates the total of 30 search regions together with the corresponding variable boundaries.

A comparison of the simulated background and signal event yields in all search bins is presented in Fig. 5.13 for the SR and CR independently. The bin labels show the corresponding H_{T}, L_{T} labels as well as the jet and b-tag multiplicities. In general, the signal contamination in the CRs is small relative to the backgrounds. For the SR multiple categories show high sensitivity to the signal, while the compressed and non-compressed scenarios populate different bins. This confirms the importance of the various search categories for the sake of signal sensitivity.

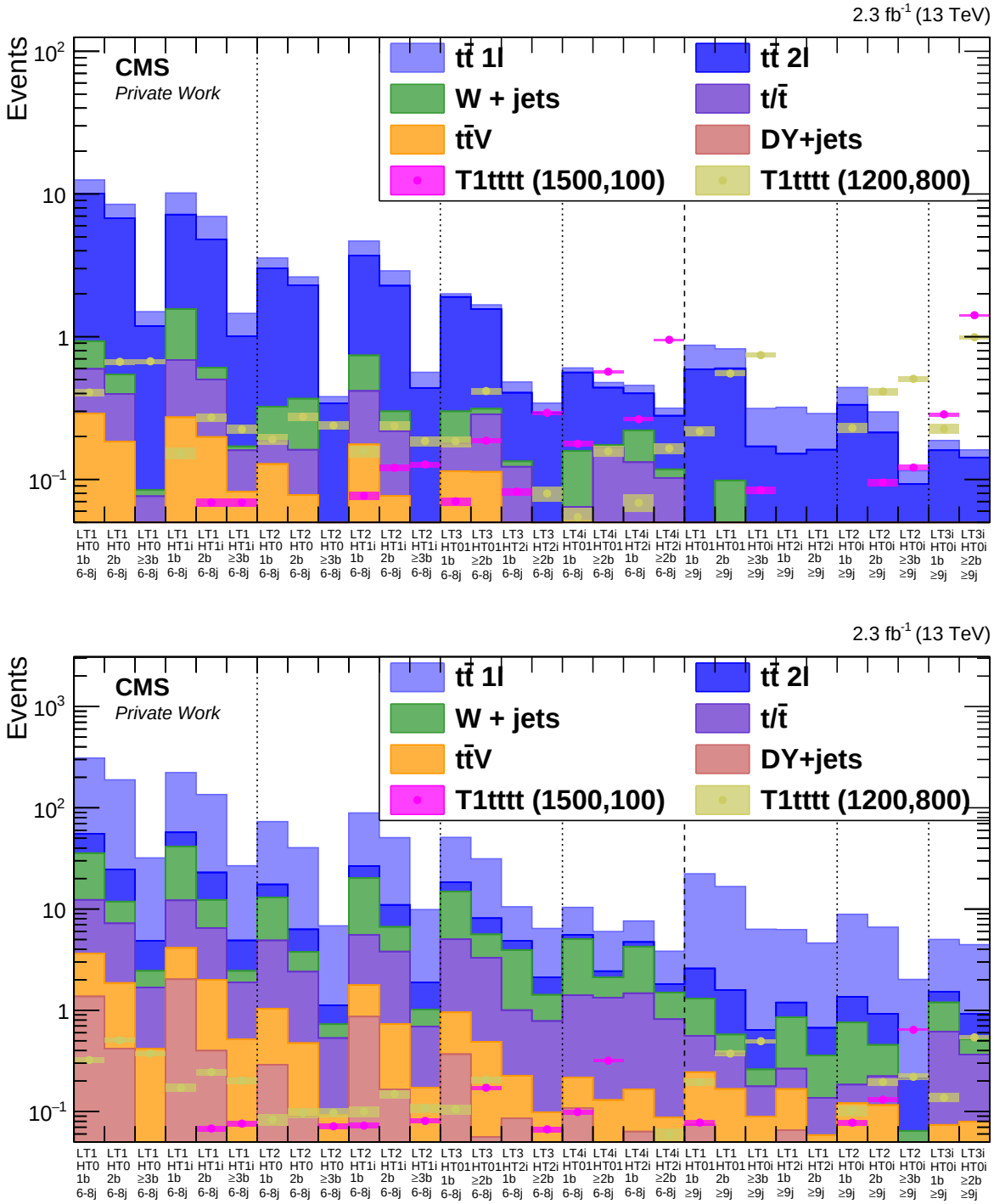


Fig. 5.13: Simulated event yields for the *EWK* backgrounds and two signal benchmark models in all 30 search categories for the signal (*top*) and control (*bottom*) regions. The non-compressed T1tttt (1500,100) signal mostly populates the bins with large L_T, while the compressed T1tttt (1200,800) scenario contribution is spread over the whole search phase-space. The signal contamination in the control regions is always less than one event, and hence negligible.

Chapter 6

Background Estimation

6.1 Background composition 98
6.2 R_{CS} method 98
6.2.1 R_{CS} composition 101
6.2.2 R_{CS} in $t\bar{t}$ events 101
6.2.3 R_{CS} in different kinematic regions 106
6.3 QCD background estimation 112
6.3.1 QCD estimation in the electron channel 114
6.3.2 QCD estimation in the muon channel 124
6.4 Systematic uncertainties 128
6.4.1 Uncertainties on background estimation 128
6.4.2 Uncertainties on background and signal prediction 130
6.4.3 Uncertainties on signal prediction 133
6.5 Prediction in data 136

In order to quantify a possible new physics effect in the search regions (SRs), the knowledge of the standard model background contributions in the respective phase-space is required. Due to the new centre-of-mass energy, possible event mismodelling, and not well studied detector performance, the usage of MC simulation to predict those backgrounds is not favourable. Therefore, a data-driven approach has been developed.

This chapter presents the data-driven methods used for the estimation of the most important SM backgrounds. All processes are predicted from control regions (CRs) using transfer factors, measured in lower jet multiplicity regions in data (cf. Sec. 6.2). For the *QCD* background a fake-lepton enriched data sample is used to obtain the prediction as described in Sec. 6.3. These methods are validated and verified using MC simulation.

Finally, the simulation accuracy is evaluated for various parameters and propagated to the background estimation as systematic uncertainties (cf. Sec. 6.4). The uncertainties on the SUSY signal expectations are obtained in a similar way.

6.1 Background composition

As discussed previously in Chapter 5, after the baseline selection the largest SM background comes from $t\bar{t}$ contributing at least 80% of the events, followed by W+jets. The SR is dominated by dileptonic $t\bar{t}$, where one of the leptons is outside acceptance, does not satisfy the identification criteria, or is a hadronic decaying τ. In this case, the two neutrinos render the W-boson reconstruction impossible. As a result, the $\Delta\varphi$ distribution looks very different from that of semileptonic $t\bar{t}$, with almost a flat shape similar to the studied SUSY signals (cf. Fig. 5.9, right).

The low $\Delta\varphi$ CR comprises mostly events with a single lepton, that is aligned to the reconstructed W-boson, or events with low $\not{E}_T$. Since QCD events have no intrinsic missing energy, they also do not populate the high $\Delta\varphi$ signal region. The $t\bar{t}V$ processes contribute to the whole phase-space, but due to the low cross section, they are negligible with respect to the plain $t\bar{t}$ background.

6.2 R_{CS} method

Since the signal contamination in the control region is negligible, the number of background events in the signal region can be determined from the control region, if the relation between these regions is known. The ratio of the number of events in the signal region to the event yield in the control region defines this translation factor R_{CS}:

$$R_{CS} = \frac{N^{SR}}{N^{CR}} \ , \tag{6.1}$$

which allows to estimate the number of SR events for a certain search bin from the corresponding CR event yield. As QCD events do not populate the high $\Delta\varphi$ tail, this background is excluded from the transfer factor. The combination of the other SM backgrounds will be further denoted as EWK. Using the previous definitions, the R_{CS} factor for simulation events is then defined as:

$$R_{CS}^{MC} = \frac{N^{\text{all backgrounds except QCD}}(\Delta\varphi > x)}{N^{\text{all backgrounds except QCD}}(\Delta\varphi < x)} = \frac{N_{SR}^{EWK}}{N_{CR}^{EWK}} \ . \tag{6.2}$$

Naturally, the use of this prediction method requires the a priori knowledge of this translation factor. An orthogonal sample with similar kinematics and no signal contamination would allow to measure the R_{CS} in data. Since the $\Delta\varphi$ value explicitly depends on the lepton momentum and missing energy, it is assumed to be not correlated to the jet multiplicity. Therefore, a sideband (SB) sample with four or five jets is defined, which is used to determine the R_{CS} factor in data.

Given that QCD events are present only in the CRs, their contribution has to be subtracted from the R_{CS} calculation. This finally yields the following definition for the determination of the transfer factor in data:

$$R_{CS} = \frac{N_{SB}^{data}(\Delta\varphi > x))}{N_{SB}^{data}(\Delta\varphi < x) - N_{SB}^{\text{QCD estimate}}} \ , \tag{6.3}$$

The measurement of transfer factors for the QCD estimation is performed in the zero $n_{\text{b-tag}}$ sideband with three or four jets, as explained later in Sec. 6.3. Higher jet multiplicity bins are not used due to potential signals without b jets [158]. Table 6.1 shows the event regions used

Table 6.1: Overview of the event categories used for the background prediction in the low jet multiplicity sideband (SB). The signal search region is located in the high n_{jet} mainband (MB).

n_{jet}	$n_{b\text{-tag}} = 0$	$n_{b\text{-tag}} \geq 1$
$= 3$	QCD Fit (el. sample)	
$= 4$		R_{CS} determination
$= 5$		
≥ 6		Signal region

for the background prediction as well as the search region. The region with at least six jets will be labelled as the mainband (MB).

Ideally, each signal search category should have a corresponding SB bin with matching kinematic properties. Due to the lack of data statistics in some bins with extreme conditions, such as three b-tags or large H_T, the requirements are loosened for the matching SB bin. As a result, several MB bins share the same SB.

The sideband region is split into 1 $n_{b\text{-tag}}$, and ≥ 2 $n_{b\text{-tag}}$, and the appropriate L_T and H_T bins. In case of very high H_T events with ≥ 1 $n_{b\text{-tag}}$ are used. Figure 6.1 shows the control region composition in the main- and sideband bins, with each simulated SM process given as a fraction of the total background.

Residual differences

Despite not being explicitly used in the $\Delta\varphi$ calculation, the jet multiplicity has an indirect impact on the R_{CS} value, since the reconstructed missing energy is correlated with the hadronic activity of the event, and thus, also to the jet multiplicity. The residual difference between the R_{CS} in the SB and MB is evaluated in simulation as a correction factor κ:

$$\kappa = \frac{R_{CS}^{MC}(\text{MB})}{R_{CS}^{MC}(\text{SB})} \ , \tag{6.4}$$

where $R_{CS}^{MC}(\text{MB})$ is the actual R_{CS} in a main band region from simulation and $R_{CS}^{MC}(\text{SB})$ is the simulated R_{CS} in the corresponding sideband region. The final prediction of the SM background accounts for these differences using the κ factor:

$$N_{SR}^{SM} = R_{CS}^{SB} \cdot \kappa \cdot (N_{CR}^{obs} - N_{CR}^{\text{QCD estimate}}) \ , \tag{6.5}$$

where N_{SR}^{SM} is the number of predicted SM background events in the signal region, R_{CS}^{SB} is the transfer factor determined from data in the sideband region with a lower number of jets, and N_{CR}^{obs} – the number of counted data events in the control region with $\Delta\varphi < x$. The κ factor is determined for each search bin separately from MC simulation. As will be discussed further, κ is similar for bins that differ only by the b-tag multiplicity, which allows to apply the same correction factor in bins with high b-jet multiplicity to reduce statistical uncertainties.

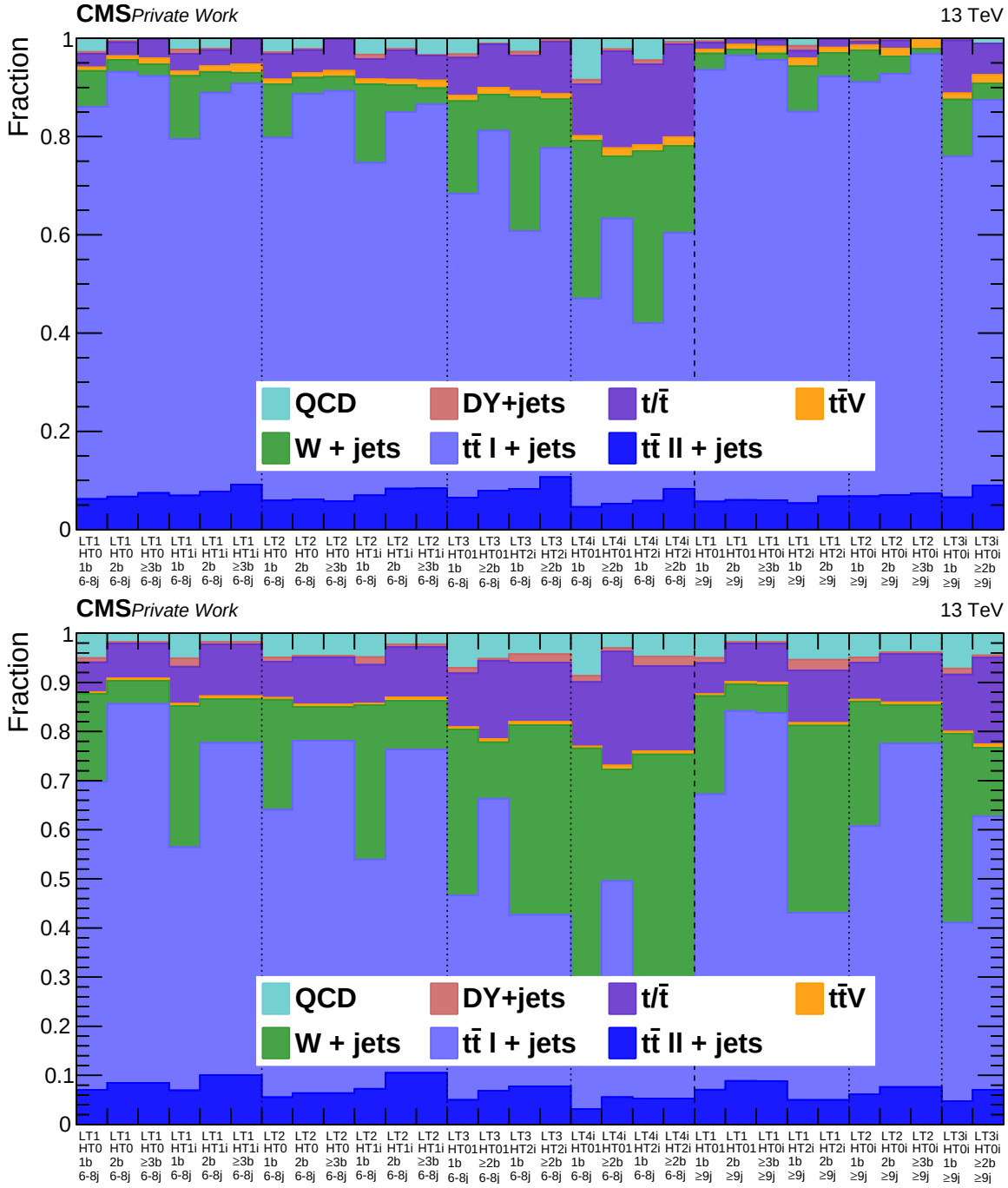


Fig. 6.1: Fractions of the simulated SM backgrounds in the control regions of the different mainband (*top*) and sideband (*bottom*) bins. The largest change is seen for the dominating W+jets and semileptonic $\mathrm{t\bar{t}}$+jets. The corresponding H_T, L_T, $n_\text{b-tag}$ and n_jet ranges are indicated in the bin labels.

6.2.1 R_{CS} composition

As the prediction extends from $n_{jet} \in [4,5]$ to two search bands with $n_{jet} \in [6,8]$ and $n_{jet} \geq 9$, it is important to understand the origin of all residual differences in detail.

Two important quantities play a role when the R_{CS} is compared between the side and mainband: the relative fraction of a certain background (e.g. semi- or dileptonic $t\bar{t}$) and the corresponding R_{CS} value of this component. The combined R_{CS} of all background components can then be expressed as:

$$\begin{aligned} R_{CS} = \frac{N_{SR}}{N_{CR}} = \frac{\sum_i N^i_{SR}}{\sum_j N^j_{CR}} = \frac{1}{N_{CR}} \sum_i N^i_{SR} \cdot \frac{N^i_{CR}}{N^i_{CR}} = \\ = \sum_i \frac{N^i_{CR}}{N_{CR}} \frac{N^i_{SR}}{N^i_{CR}} = \sum_i f^i_{CR} \cdot \frac{N^i_{SR}}{N^i_{CR}} = \sum_i f^i_{CR} \cdot R^i_{CS} \,, \end{aligned} \tag{6.6}$$

where f^i_{CR} is the fraction relative to the total number of events from a background process i in the control region $\Delta\varphi < x$. R_{CS} thus is proportional to the R^i_{CS} of the backgrounds as well as their corresponding fraction f^i_{CR}.

As discussed earlier, the $\Delta\varphi$ shape and, hence, R_{CS} is especially different for events with one W-boson that decays leptonically and those, where more hard neutrinos are involved, e.g. for $t\bar{t}$ events in which both W bosons decay leptonically. A larger R_{CS} is also expected for events with three neutrinos as e.g. $t\bar{t}Z$, when the $t\bar{t}$ system decays semileptonically and the Z boson decays to two neutrinos. If the R_{CS} value of a given background component does not change when going from sideband to mainband region, then the mixture of all components determines the change in the total R_{CS} from between the two bands, hence κ.

In the following, two effects are investigated. First, the composition of the different $t\bar{t}$ decay channels and the compositions of different background processes (i.e. the fraction of W+jets, $t\bar{t}$, $t\bar{t}V$), especially when the low jet multiplicity sideband regions are compared to the mainband with $n_{jet} \geq 6$. Second, the effect of different kinematic regions in L_T and H_T in the R_{CS} values in general.

first, the background composition of similar ($t\bar{t}$ decay channels) and completely different physics processes (i.e. the fraction of W+jets, $t\bar{t}$, $t\bar{t}V$), especially when the low jet multiplicity sideband regions are compared with regions of $n_{jet} \geq 6$; and second, the effect of different kinematic regions in L_T and H_T in the R_{CS} values in general.

6.2.2 R_{CS} in $t\bar{t}$ events

Dileptonic $t\bar{t}$ events might be misidentified as semileptonic events when one of the leptons is a hadronically decaying τ lepton, or if one of the leptons (electron or muon, possibly from a τ decay) is not identified. In both cases, the relation between the lepton and $\not{E}_T$ is not given as for the semileptonic decay, where the real $\not{E}_T$ originates from a single neutrino, such that more events are expected at large $\Delta\varphi$ values.

Apart from the varying sources of missing energy, the $t\bar{t}$ decay channels also differ in the number of jets. While semileptonic decays are always accompanied by four jets from the hard interaction, the dileptonic channels might have between two jets (of the b-quark decay for $t\bar{t} \to \ell\ell$, with $\ell = \mu$,e) and four jets, with two b jets and two from the hadronic τ decay. As a result, based on the number of jets produced from the matrix element, the semi- and dileptonic

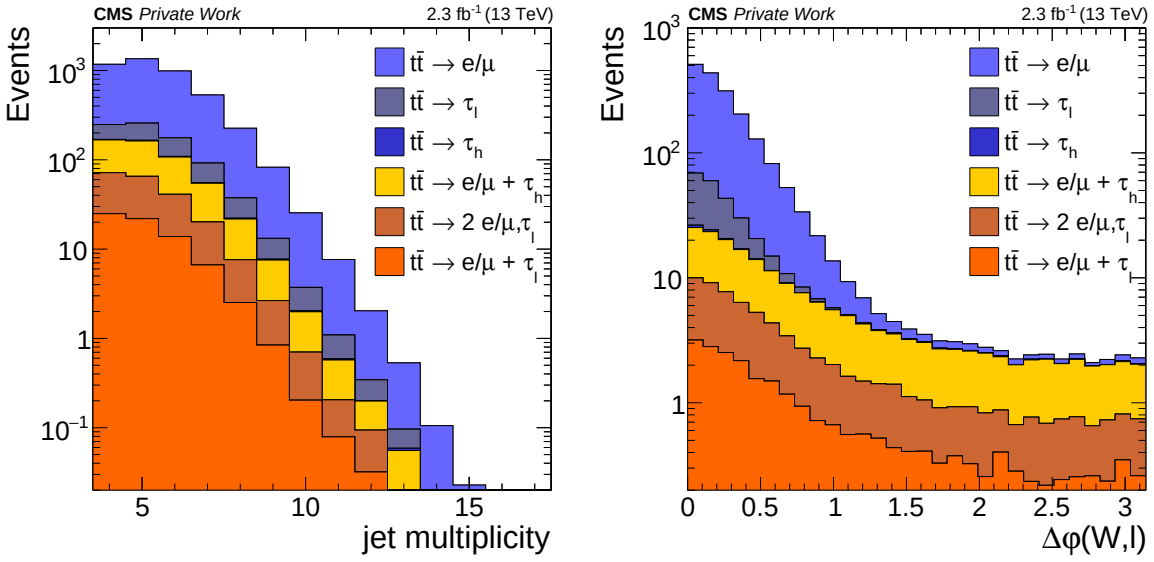


Fig. 6.2: Jet multiplicity and $\Delta\varphi$ distributions for the different components of the $t\bar{t}$ decay after the baseline selection. Only the most contributing channels are chosen. For the *left* (*right*) plot an n_{jet} requirement of ≥ 4 (6) jets is applied.

$t\bar{t}$ events should exhibit varying n_{jet} distributions. This translates into a different evolution of the R_{CS} factor as a function of the jet multiplicity.

However, the phase-space after the baseline selection with $H_{\text{T}} > 500\,\text{GeV}$ often also contains events with more than four jets, where additional jets arise from initial-state radiation. This results in a more similar n_{jet} distribution tail for both decay modes as shown in the left plot of Fig. 6.2. After requiring 6 or more jets, the main contribution in the high $\Delta\varphi$ region stems from dileptonic $t\bar{t}$ events, where one of the leptons is a hadronic τ, as shown in in the right plot.

Due to the similar shape of the $\Delta\varphi$ distributions, the R_{CS} of the different dileptonic decay modes is expected to be very similar. This allows to study the dileptonic decays as a combined category, as adopted further. Similarly, the semileptonic $t\bar{t}$ channels, which are dominated by direct W boson decays to an electron or muon, are considered as a single sample. The negligible contribution from the hadronic channels is included in the semileptonic $t\bar{t}$.

R_{CS} for the $t\bar{t}$+jets background

As mentioned earlier, when several processes with different $\Delta\varphi$ distributions are combined, the total R_{CS} depends on the fractions of the single backgrounds. Here, a simple example for the two $t\bar{t}$ channels will be considered to demonstrate this dependence.

Figure 6.3 (*left*) shows the individual R_{CS} values of the semi- and dileptonic $t\bar{t}$ samples as a function of the jet multiplicity. As expected, no significant change in the translation factor can be observed. The increase in jet activity leads to a certain worsening of the $\not{E}_{\text{T}}$ resolution, thus to a longer $\Delta\varphi$ tail, which mostly affects the semileptonic $t\bar{t}$ channel with an intrinsically low R_{CS} of about 0.01. In dileptonic events with multiple neutrinos and a high $R_{\text{CS}} \approx 0.45$ this effect is less visible.

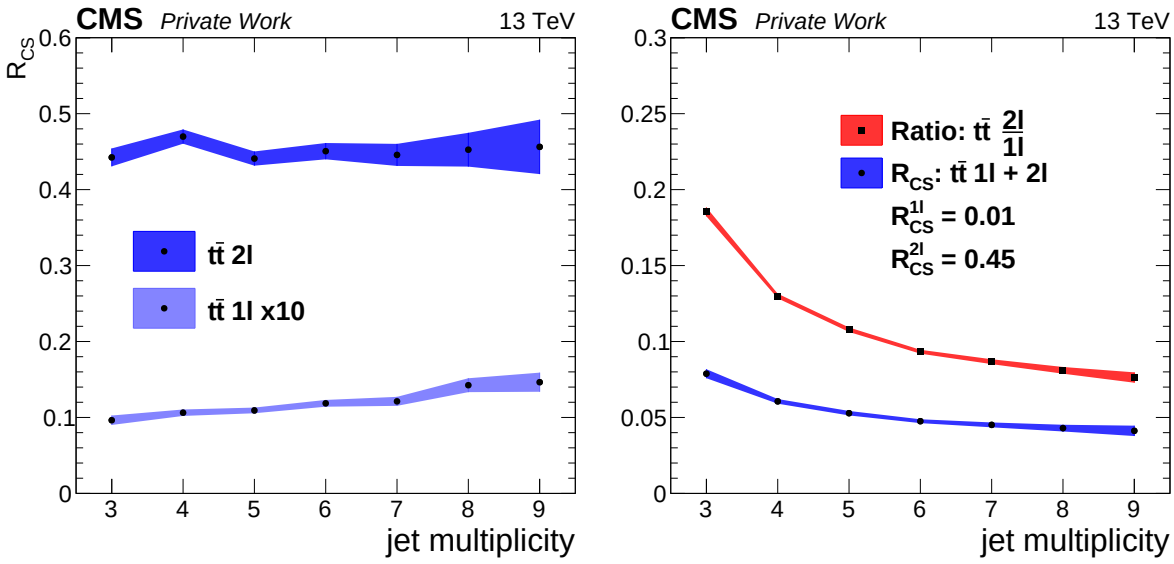


Fig. 6.3: *Left*: R_{CS} of semi- and dileptonic $t\bar{t}$ values versus the jet multiplicity. No significant trend is observed between the side- and mainband regions. *Right*: ratio of di- to semileptonic $t\bar{t}$ events in the CR (blue) and the combined R_{CS} as calculated with Eq. 6.7 (red). Although the relative composition changes, the combined R_{CS} value varies only slightly from side- to mainband.

The ratio of the number of events in dileptonic to semileptonic $t\bar{t}$ is shown in Fig. 6.3 (*right*) by the blue band. The total $t\bar{t}$ R_{CS} is then computed as:

$$R_{CS}^{t\bar{t}} = f_{1\ell} \cdot R_{CS}^{1\ell} + f_{2\ell} \cdot R_{CS}^{2\ell} = f_{1\ell} \cdot 0.01 + f_{2\ell} \cdot 0.45 \,. \tag{6.7}$$

The evolution of this combined R_{CS} value with the jet multiplicity is shown in the same plot (Fig. 6.3, *right*) by the red band. Since the dileptonic $t\bar{t}$ contribution decreases with growing n_{jet}, the total R_{CS} is moving towards the semileptonic value. Nevertheless, the change between the side- and mainband regions is small, resulting in a total $t\bar{t}$ κ factor slightly below one. The combined R_{CS} of around 0.05 is about ten times larger (smaller) than the semileptonic (dileptonic) $t\bar{t}$ component.

Dilepton control sample in data

The actual R_{CS} values in data might differ from simulation due to resolution effects of the various reconstructed objects, notably $\not{E}_T$. Since the κ factor rather reflects the underlying physics, it is less affected by detector or simulation misbehaviour. Sources of such systematic uncertainties will be discussed later in Sec. 6.4.

Given that κ is determined from simulation, it is nevertheless important to understand how well it reflects the real data. In the following, the focus will be put on the comparison of the n_{jet} distribution for semi- and dileptonic $t\bar{t}$ events.

For this purpose, a dilepton data control sample is defined as follows:

- Exactly two selected opposite-sign leptons are required similarly to the normal selection.

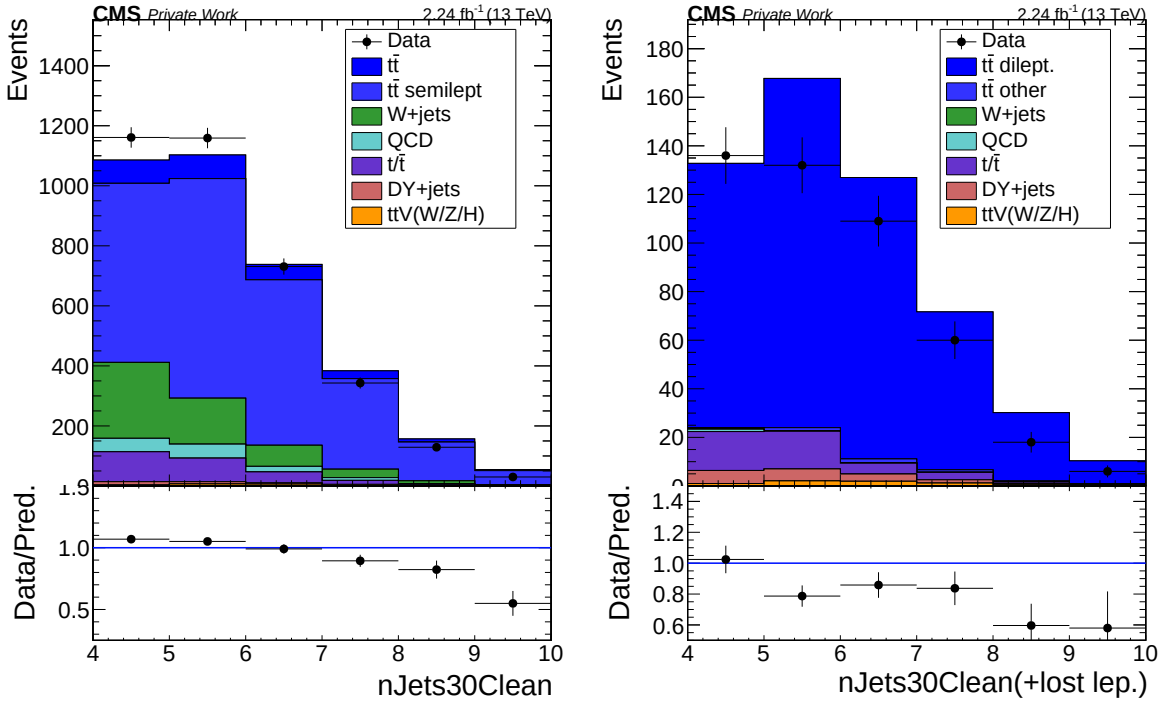


Fig. 6.4: Jet multiplicity distribution after the single-lepton baseline event selection with $\Delta\varphi < x$ (*left*) and in the dilepton control region (*right*), where one of the leptons is replaced with a τ-dacay like jet [159]. In both cases the simulation tends to overestimate the data towards high n_{jet}.

- For same-flavour leptons the invariant mass of the lepton pair is required to lie outside the Z boson mass peak: $81\,\text{GeV} < m_{\ell\ell} < 101\,\text{GeV}$.

  This results in a sample with typically ∼80-100% of the events having two real leptons.

- In order to mimic semileptonic events, one of the leptons is deleted from the event and replaced by a jet-like object. The jet momentum is set to 2/3 of the lepton p_{T} as a crude estimate for a hadronic τ decay.

- The new jet is added to the jet collection, and all involved variables are recalculated: n_{jet}, H_{T}, $\not{E}_{\text{T}}$, L_{T} and, finally, $\Delta\varphi$.

To enrich the statistics, each event is used twice, with both leptons once being replaced by the τ-like jet. No $\Delta\varphi$ requirement is imposed for the same reason.

The jet multiplicity distribution in the CR after the single-lepton baseline selection is shown in the left plot of Fig. 6.4. For the right plot the aforementioned dilepton selection is used. In both cases, $\text{t}\bar{\text{t}}$ events are the dominating background, and the simulation generally overestimates the data towards higher jet multiplicities. The difference of these trends in the semileptonic and dileptonic background distributions can be evaluated through their double ratio.

Fig. 6.5 shows the data-MC ratios in the single-lepton (*left*) and dilepton (*middle*) samples and the resulting ratio of those trends in the *right* plot. This double ratio is fitted with a linear

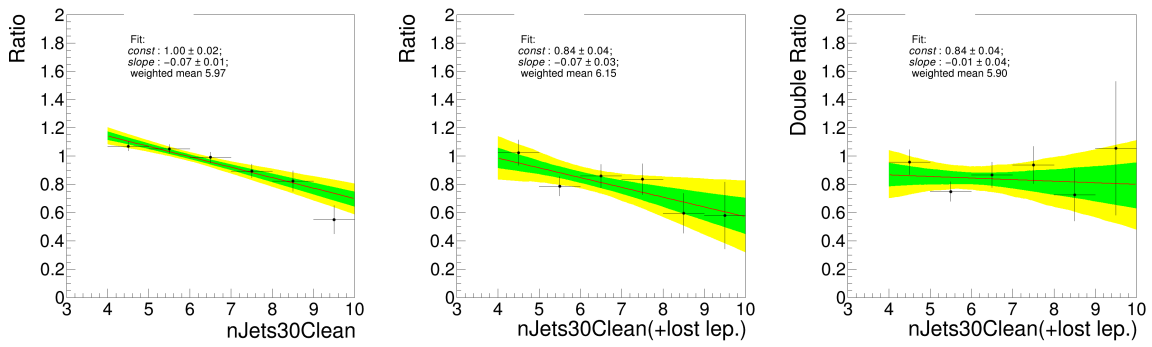


Fig. 6.5: Ratio of data and simulated events for the jet multiplicity distribution of single lepton events (*left*), dileptonic events (*middle*), and the resulting double ratio (dileptonic/semileptonic; *right*) [159]. The ratios are fitted with a linear parametrisation and the fit parameters are indicated in the plots.

parametrization:

$$f(n_{\text{jet}}) = a + b \cdot (n_{\text{jet}} - \langle n_{\text{jet}} \rangle) \ , \tag{6.8}$$

where n_{jet} is the jet multiplicity, $\langle n_{\text{jet}} \rangle$ is the weighted mean value of n_{jet}; a and b determine the offset and slope, respectively. This parametrization decorrelates the parameters a and b.

The double ratio provides an estimate of the n_{jet}-dependent relative composition of the di- and semileptonic $\mathrm{t\bar{t}}$ decays in data, as discussed before. Since the κ factor is a ratio of R_{CS} values, it is mostly insensitive to the constant offset of this double ratio trend. However, a slope indicates, that the di- and semileptonic $\mathrm{t\bar{t}}$ events do not have similar n_{jet} distributions, which is expected from simulation. Therefore, the MC-based κ translation factor is determined with a greater uncertainty.

These fit results can be later on used as a systematic variation of the dileptonic fraction of events in the κ determination. For the offset, the fit parameter deviation from one (15%) is added in quadrature to its uncertainty (6%). This can be further applied as a constant weight on the dileptonic background as:

$$w(\text{DL}_{\text{Const}}) = 1 \pm 16\% \ . \tag{6.9}$$

The variation from the double ratio slope is obtained in a similar way: the deviation from zero (1%) and the fit error (4%) add up to a conservative 4.5% uncertainty, which can be applied as a n_{jet}-dependent weight to the dileptonic sample normalization:

$$w(\text{DL}_{\text{Slope}}) = 1 \pm (n_{\text{jet}} - \langle n_{\text{jet}} \rangle) \cdot 4.5\% \ . \tag{6.10}$$

A limitation to this method is the fact, that the data samples used for the double ratio determination, are not purely consisting of $\mathrm{t\bar{t}}$+jets events. In particular, the single-lepton selection shows a considerable contribution from the W+jets background (Fig 6.5, *left*). Those events could be suppressed by requiring at least two b-tagged jets, however, the resulting statistics in the corresponding dilepton sample would not suffice to draw a significant conclusion.

6.2.3 R_{CS} in different kinematic regions

As the SM background does not only consist of $t\bar{t}$ events, the total R_{CS} and κ values also get affected by the other backgrounds. Despite the increased number of contributing processes, the factorization of the R_{CS} as in Eq. 6.6 still holds. Therefore, changes of the relative contribution of backgrounds with similar R_{CS} values do not change the combined R_{CS}.

Individual R_{CS}

The two plots in Fig. 6.6 show the fraction of the individual SM processes relative to the total background in simulated events; the baseline selection as discussed before is applied, and only mainband search bins with an $n_{b\text{-tag}} = 1$ requirement are shown. On the *left*, the MB is dominated by $t\bar{t}$+jets events, followed by W+jets, while in the sideband (on the *right*) these backgrounds contribute on par. The other processes' shares are below 10% with only small bin-to-bin variations. As discussed before, the *QCD* contamination will be handled separately as shown in Sec. 6.3.

Figure 6.7 shows the individual R_{CS} values in the MB and SB and the corresponding κ factor for the most important backgrounds: semi- and dileptonic $t\bar{t}$, W+jets and $t\bar{t}V$. The behaviour for the two $t\bar{t}$ decay channels agrees with what was already discussed previously in Sec. 6.2.2: while the R_{CS} values differ between the various categories, the resulting κ factor stays close to unity. This confirms the correctness of the sideband requirement definition. Since W+jets events are similar to the semileptonic $t\bar{t}$ in respect of the $\Delta\varphi$ distribution, the R_{CS} values are also small ~ 0.02 and do not change between the side- and mainband. The same is valid for the single-top quark background, which is not shown due to this similarity and larger statistical uncertainties.

As for $t\bar{t}V$ events, this background can be considered as a composition of different subprocesses, like in the case of $t\bar{t}$+jets. Due to the stronger SB to MB change in the contributions, the κ factors lie below unity at around 0.7. Nevertheless, the tininess of the overall $t\bar{t}V$ fraction renders its impact on the total R_{CS} and κ values negligible.

In general, a similar picture is seen for the higher b-tag multiplicity categories in Figs. 6.8 and 6.9 for $n_{b\text{-tag}} \geq 2$ and in Figs. 6.10 and 6.11 for $n_{b\text{-tag}} \geq 3$. With increasing $n_{b\text{-tag}}$ requirements the contribution from W+jets gradually decreases, such that in the $n_{b\text{-tag}} \geq 3$ region 90% of the background stems from $t\bar{t}$+jets events. The individual R_{CS} and κ values, however, do not change much for the different $n_{b\text{-tag}}$ categories.

EWK R_{CS}

After the background composition and the behaviour of the individual processes is studied in simulation, the R_{CS} transfer factors and related κ values are determined. First, purely MC-based quantities are evaluated. As discussed, the κ factor is obtained from all SM background processes, except for *QCD*, denoted as *EWK*.

Figure 6.12 illustrates the total *EWK* R_{CS} in the side- and mainband regions, as well as the corresponding κ factors. Overall, the SB R_{CS} follows closely the MB value, with a mean of about 0.05, which corresponds to the dominating $t\bar{t}$ component. As a result, the κ factors lie around unity, with no visible systematic trends, as shown in Table 6.2.

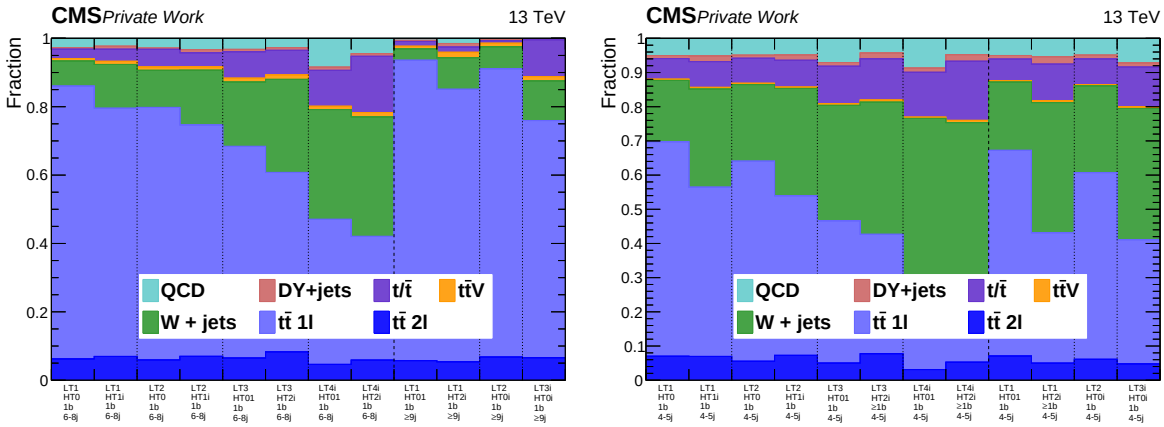


Fig. 6.6: Fractions of the SM backgrounds in the control regions of the mainband (*left*) bins with $n_{\text{b-tag}} = 1$ and corresponding sideband (*right*). While the MB is dominated by $\mathrm{t\bar{t}}$, the SB has also a similar contribution from W+jets.

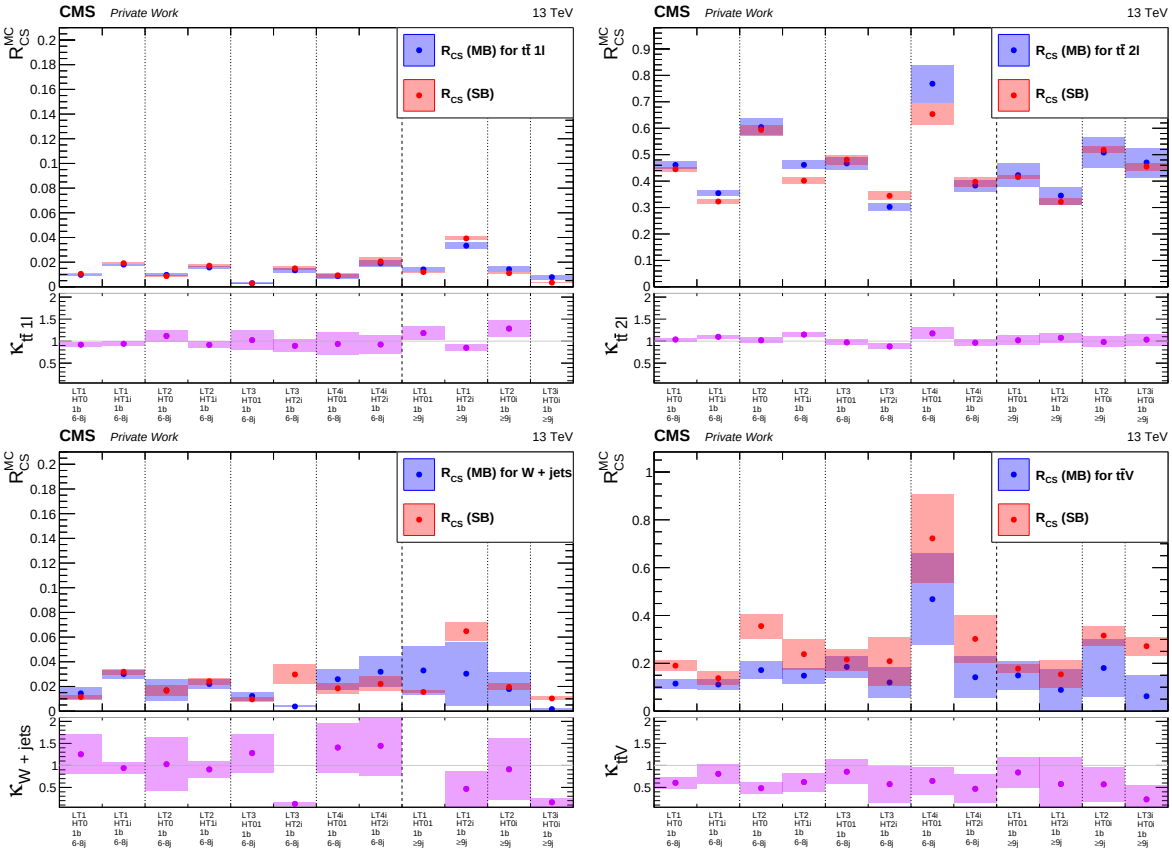


Fig. 6.7: R_{CS} in the SB and MB regions with $n_{\text{b-tag}} \geq 1$, and κ values of the corresponding bins for the most important backgrounds. The individual R_{CS} of the $\mathrm{t\bar{t}}$ components and W+jets are similar between the n_{jet} bands, thus yielding a κ factor around one. The $\mathrm{t\bar{t}V}$ R_{CS} values resemble the one of combined $\mathrm{t\bar{t}}$ channels with a κ below one.

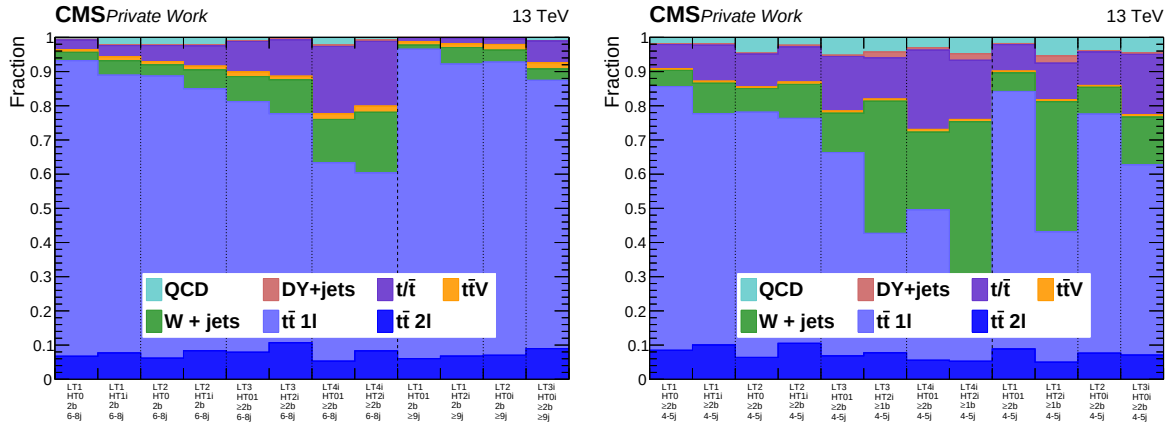


Fig. 6.8: Fractions of the SM backgrounds in the control regions of the mainband (*left*) bins with $n_{\text{b-tag}} \geq 2$ and corresponding sideband (*right*). The harder $n_{\text{b-tag}}$ requirement further suppresses the W+jets contributions in both bands.

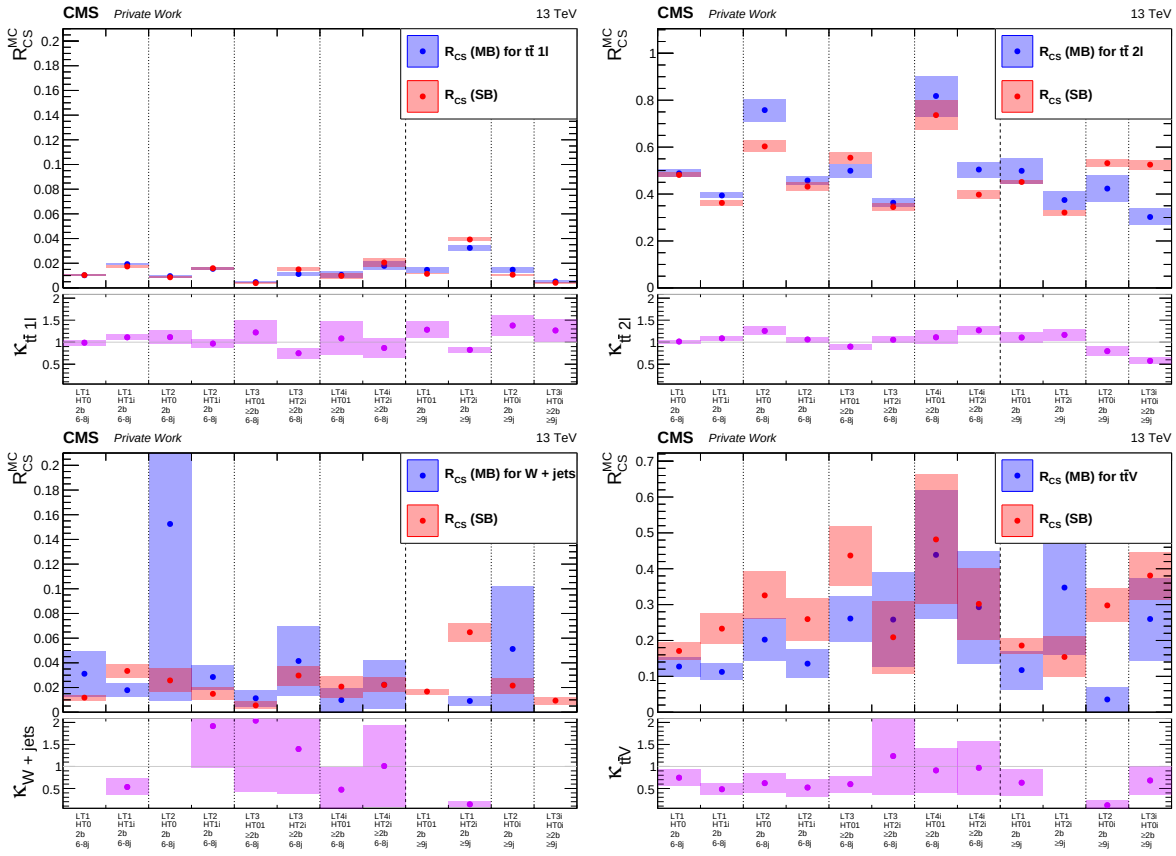


Fig. 6.9: R_{CS} in the SB and MB regions with $n_{\text{b-tag}} \geq 2$, and κ values of the corresponding bins for the most important backgrounds. The individual R_{CS} of the $\text{t}\bar{\text{t}}$ components are similar between the n_{jet} bands, thus yielding a κ factor around one. The reduced statistics of W+jets events results in large statistical uncertainties. The $\text{t}\bar{\text{t}}\text{V}$ R_{CS} values resemble the one of combined $\text{t}\bar{\text{t}}$ channels with a κ below one.

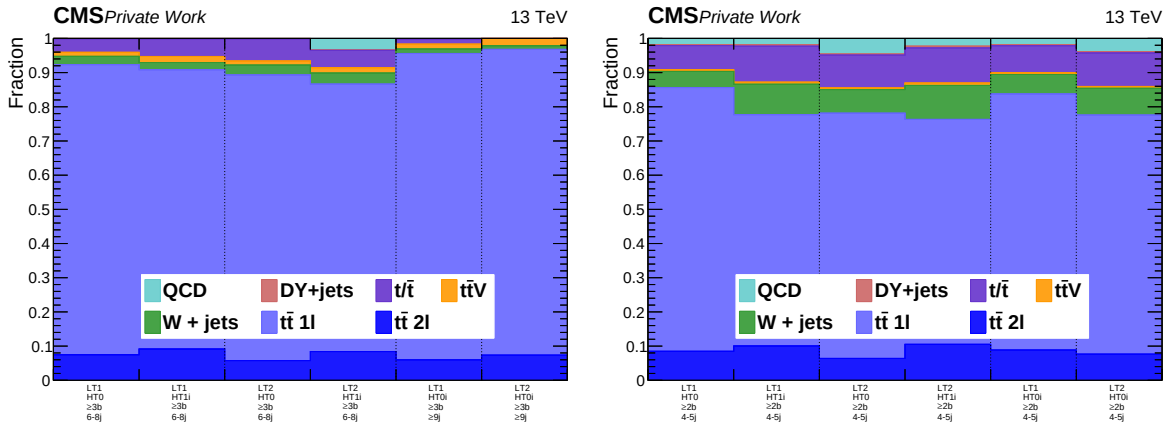


Fig. 6.10: Fractions of the SM backgrounds in the control regions of the mainband (*left*) bins with $n_{\text{b-tag}} \geq 3$ and corresponding sideband (*right*). Both bands are dominated by tt̄+jets events at around 90%.

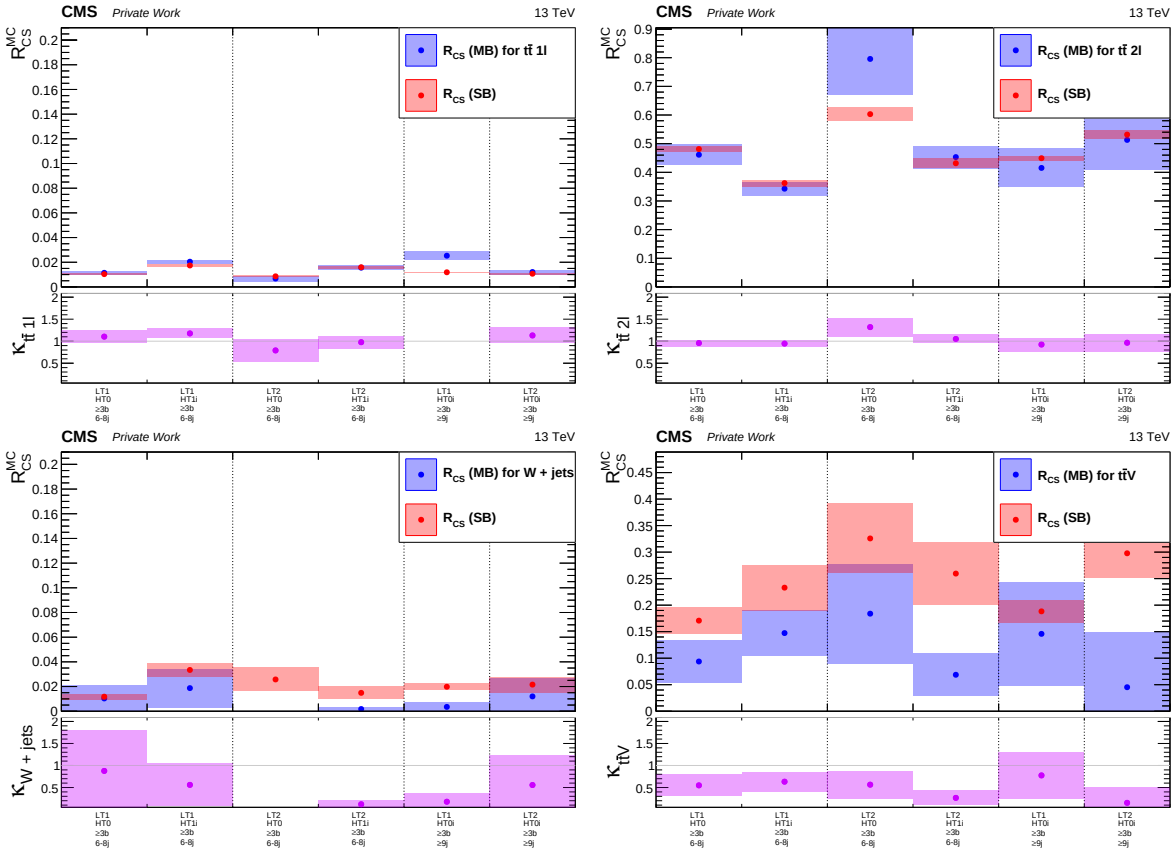


Fig. 6.11: R_{CS} in the SB and MB regions with $n_{\text{b-tag}} \geq 3$, and κ values of the corresponding bins for the most important backgrounds. Similar to the $n_{\text{b-tag}} \geq 2$ behaviour except for increased statistical fluctuations due to a harder $n_{\text{b-tag}}$ requirement in the mainband.

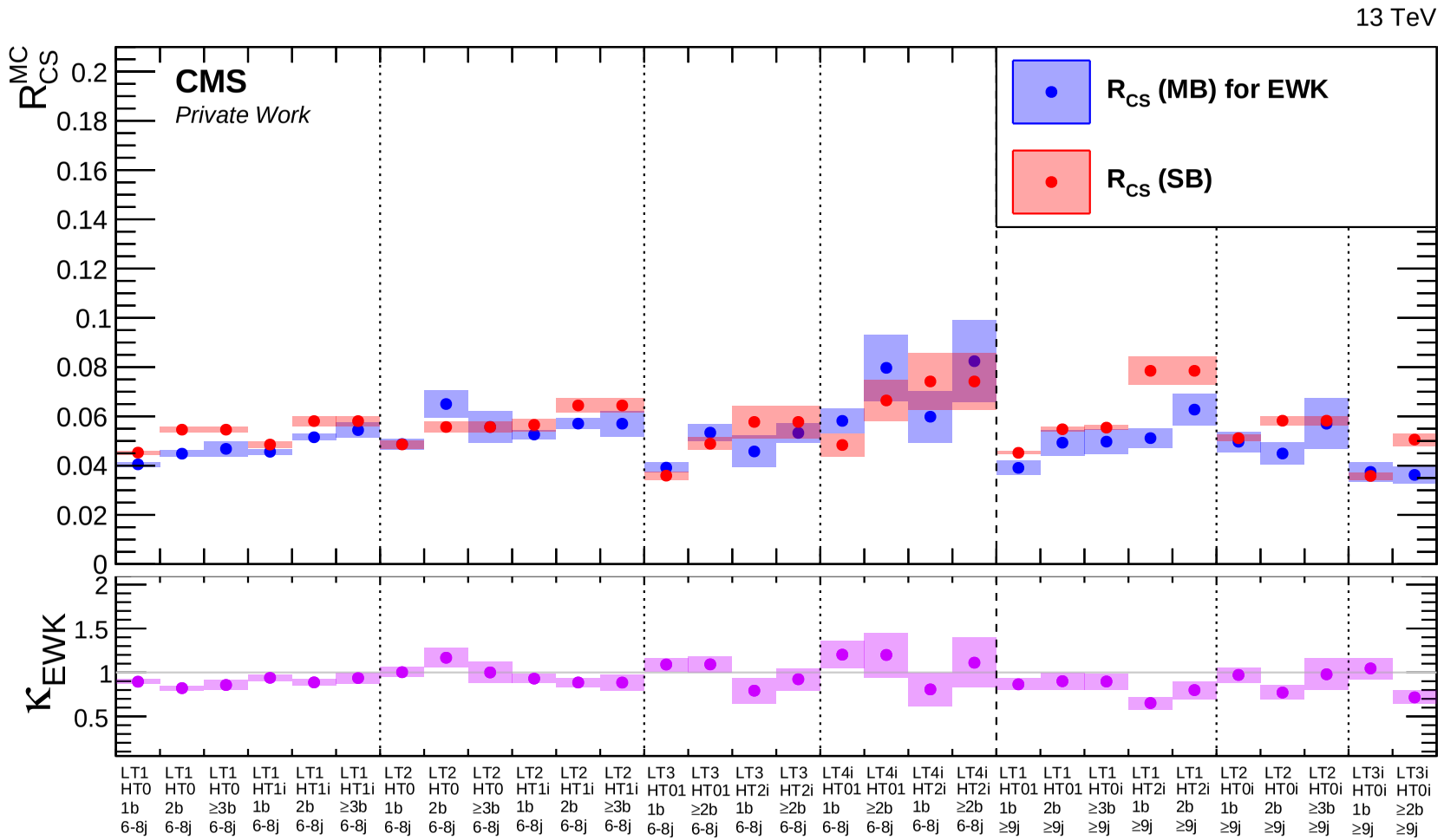


Fig. 6.12: R_{CS} values in the SB and MB regions, including the corresponding κ factors for the simulated total *EWK* background, consisting of all discussed SM processes, except for *QCD* multijet events. The sideband R_{CS} follows closely the mainband value, resulting in a κ factor close to one over most of the bins.

Table 6.2: R_{CS} values in the SB and MB regions, including the corresponding κ factors for the simulated total *EWK* background, consisting of all discussed SM processes, except for *QCD* multijet events.

n_{jet}	L_T [GeV]	H_T [GeV]	$n_{b\text{-}tag}$	R_{CS} MB	R_{CS} SB	κ^{EWK}
[6, 8]	[250, 350]	[500, 750]	$= 1$	0.0405 ± 0.0010	0.0453 ± 0.0008	0.8946 ± 0.0270
			$= 2$	0.0448 ± 0.0013	0.0546 ± 0.0011	0.8213 ± 0.0288
			≥ 3	0.0468 ± 0.0030	0.0546 ± 0.0011	0.8573 ± 0.0579
		≥ 750	$= 1$	0.0456 ± 0.0011	0.0486 ± 0.0014	0.9396 ± 0.0350
			$= 2$	0.0515 ± 0.0014	0.0581 ± 0.0020	0.8870 ± 0.0390
			≥ 3	0.0544 ± 0.0031	0.0581 ± 0.0020	0.9371 ± 0.0624
	[350, 450]	[500, 750]	$= 1$	0.0488 ± 0.0022	0.0486 ± 0.0017	1.0030 ± 0.0572
			$= 2$	0.0650 ± 0.0056	0.0557 ± 0.0023	1.1674 ± 0.1116
			≥ 3	0.0557 ± 0.0064	0.0557 ± 0.0023	1.0003 ± 0.1218
		≥ 750	$= 1$	0.0526 ± 0.0019	0.0566 ± 0.0025	0.9304 ± 0.0533
			$= 2$	0.0571 ± 0.0022	0.0644 ± 0.0029	0.8856 ± 0.0523
			≥ 3	0.0570 ± 0.0053	0.0644 ± 0.0029	0.8849 ± 0.0906
	[450, 600]	[500, 1250]	$= 1$	0.0392 ± 0.0020	0.0359 ± 0.0017	1.0905 ± 0.0760
			≥ 2	0.0534 ± 0.0035	0.0489 ± 0.0026	1.0921 ± 0.0925
		≥ 1250	$= 1$	0.0458 ± 0.0065	0.0577 ± 0.0066	0.7926 ± 0.1443
			≥ 2	0.0533 ± 0.0040	0.0577 ± 0.0066	0.9229 ± 0.1254
	≥ 600	[500, 1250]	$= 1$	0.0582 ± 0.0051	0.0484 ± 0.0047	1.2024 ± 0.1563
			≥ 2	0.0797 ± 0.0134	0.0665 ± 0.0084	1.1991 ± 0.2530
		≥ 1250	$= 1$	0.0599 ± 0.0105	0.0742 ± 0.0116	0.8070 ± 0.1892
			≥ 2	0.0824 ± 0.0166	0.0742 ± 0.0116	1.1110 ± 0.2832
≥ 9	[250, 350]	[500, 1250]	$= 1$	0.0391 ± 0.0028	0.0452 ± 0.0007	0.8648 ± 0.0642
			$= 2$	0.0493 ± 0.0051	0.0548 ± 0.0010	0.9004 ± 0.0947
		≥ 500	≥ 3	0.0497 ± 0.0050	0.0554 ± 0.0010	0.8973 ± 0.0913
		≥ 1250	$= 1$	0.0512 ± 0.0040	0.0785 ± 0.0058	0.6519 ± 0.0703
			$= 2$	0.0627 ± 0.0065	0.0785 ± 0.0058	0.7994 ± 0.1015
	[350, 450]	≥ 500	$= 1$	0.0497 ± 0.0042	0.0511 ± 0.0014	0.9717 ± 0.0860
			$= 2$	0.0449 ± 0.0045	0.0582 ± 0.0018	0.7712 ± 0.0813
			≥ 3	0.0570 ± 0.0104	0.0582 ± 0.0018	0.9790 ± 0.1807
	≥ 450	≥ 500	$= 1$	0.0375 ± 0.0039	0.0358 ± 0.0015	1.0458 ± 0.1186
			≥ 2	0.0362 ± 0.0034	0.0506 ± 0.0026	0.7163 ± 0.0771

6.3 QCD background estimation

The R_{CS} prediction method relies on the fact, that the background events populate both, the low and high $\Delta\varphi$ regions. Since *QCD* multijet events do not have intrinsic missing energy and prompt leptons, the $\Delta\varphi$ tends to be around zero, and therefore can not be accounted for in the R_{CS} determination. For this reason, the contribution of the *QCD* background has to be subtracted before calculating the R_{CS} factor in the data control region, as in Eq. 6.3.

In this section, the contribution of *QCD* events will be reviewed and a data-driven estimation method presented for both, the electron and muon, channels.

QCD contribution

Figure. 6.13 shows the simulated number of events for the *QCD* background in the electron and muon selections in comparison to the total SM background. While both lepton channels show in general similar total yields, the electron channel has systematically more events mainly due to the contribution from *QCD* background, which is shown in the bottom panels of the plots.

Overall, the fraction of *QCD* events does not exceed 10% for electron events in the sideband, whereas the mainband shows even a lower share. *QCD* yields in the muon channel are systematically around one order lower, than in the case of electrons.

These simulation predictions give reason to expect a low contamination from *QCD* events in the analysis search region. Nevertheless, this contributions have to be accounted for, since uncertainties in the R_{CS} control regions translate into the signal region estimation.

QCD estimation method

Due to the nature of quantum chromodynamics, a precise simulation of *QCD* multiet events is not achievable. To circumvent the inaccuracy of *QCD* simulation predictions, data-driven approaches have to be used to estimate the *QCD* contribution to the search phase-space.

Intrinsically, *QCD* events do not contain heavy vector bosons, and therefore, also no prompt leptons can be produced. However, the enormous cross section and different reconstruction inefficiencies allow for such events to pass the analysis selection requirements.

The majority of reconstructed electrons in *QCD* events originate from misidentified jets or from photons, that convert in the inner tracker layers. This explains the larger *QCD* event yields in the electron channel with respect to the muon channel, and gives another motivation for a data-driven estimation, since the modelling of such processes in simulation is poor.

The identification criteria discussed in Sec. 4.3.2 ensure a certain rejection of those misidentified, or "fake", leptons when selecting the good electrons or muons. On the other hand, the inversion of those criteria also allows to obtain a sample, that is enriched with *QCD* events. Provided there exists a relation between the fake and selected leptons, it becomes possible to estimate the fraction of fake events, that pass the analysis selection.

An event sample has to be defined to measure this relation quantitatively. To ensure orthogonality to the R_{CS} determination regions and also no potential signal contamination, a *QCD* enriched sample is defined by requiring three to four selected jets and vetoing events with b-tagged jets (cf. Table 6.1). The low jet multiplicity and zero $n_{\text{b-tag}}$ selection are close to the *QCD* phase-space.

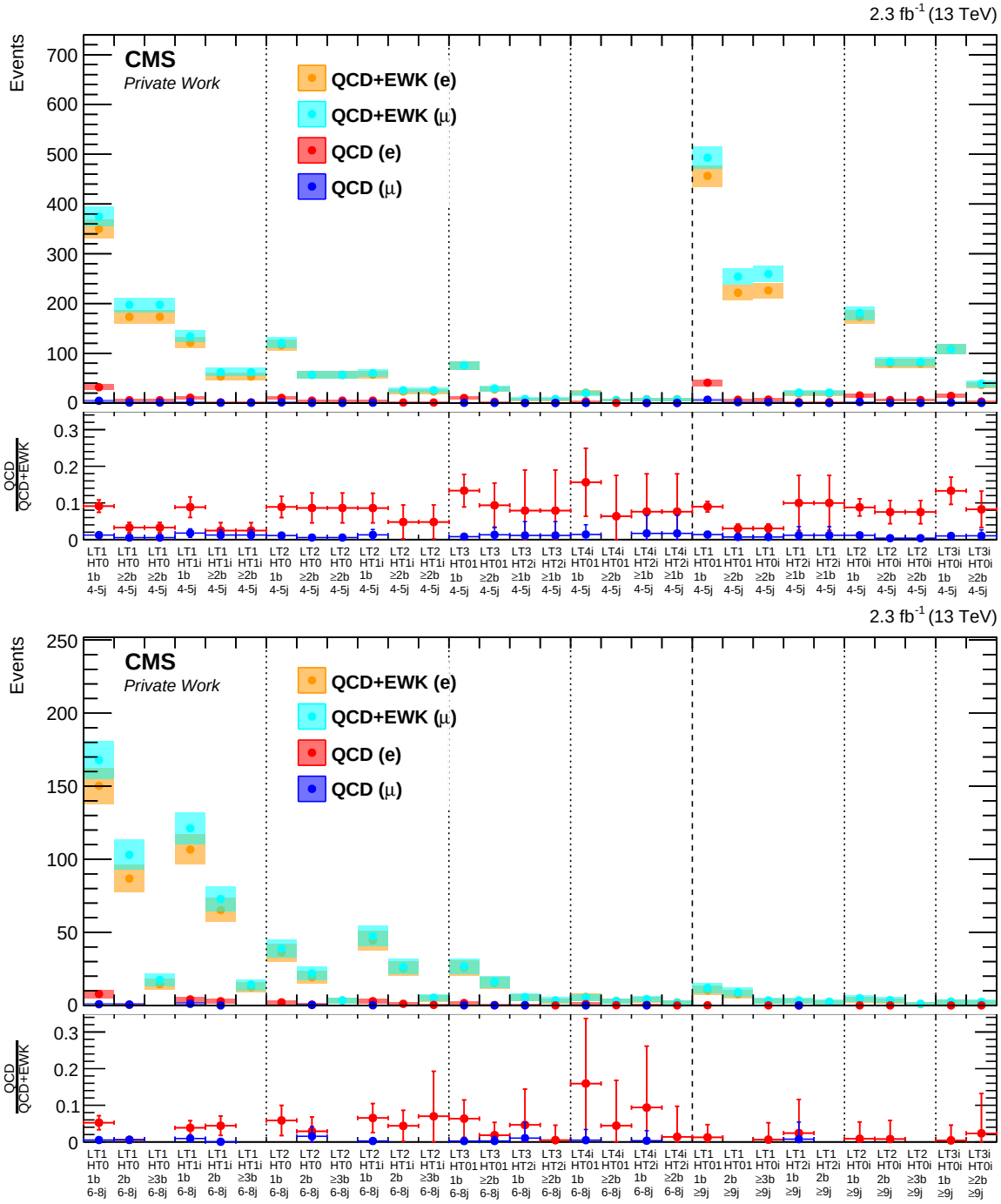


Fig. 6.13: Comparison of the simulated total background and QCD-only yields for electron and muon events in the sideband (*top*) and mainband (*bottom*) control regions. The error bars indicate the expected statistical uncertainty as for Poissonian distributed data.The bottom panels show the fraction of the *QCD* events relative to the total SM (i.e. *QCD* and *EWK*) background.

In addition to the analysis lepton events, an orthogonal selection is created by modifying the lepton identification criteria as shown later on separately for the electron and muon channels. These fake-lepton enriched samples will be further called "anti-selected", while the normal selection will be simply "selected" events.

Further, the relation between the selected and anti-selected events is established. The present method, which was already introduced by the W-polarization measurement [160] and other SUSY searches [161], relies on the L_{p} variable:

$$L_{\mathrm{p}} = \frac{p_{\mathrm{T}}(l)}{p_{\mathrm{T}}(W)} \cos(\Delta\varphi) \ , \tag{6.11}$$

that reflects the effective lepton polarization from the W boson decay. For this reason, the shape of the L_{p} distribution is well known for SM backgrounds like $\mathrm{t\bar{t}}$+jets and W+jets, where it shows a characteristic fall from 0 towards 1. This variable is also chosen, as QCD events have a different distinctive shape with a peak at $L_{\mathrm{p}} \approx 1$.

If the L_{p} distribution is similar for selected and anti-selected events, its shape can be used to obtain the respective fraction from a template fit, using also the shape of the *EWK* backgrounds. The ratio of selected to anti-selected events will characterize the probability of a selected lepton to actually originate from a fake object:

$$F_{\text{sel-to-anti}} = \frac{N_{\text{QCD selected}}}{N_{\text{QCD anti-selected}}} \tag{6.12}$$

The R_{CS} method requires the knowledge of the *QCD* contribution in the control regions of both, the side- and mainbands. Assuming that the $F_{\text{sel-to-anti}}$ is similar in these regions, the number of *QCD* events can be then estimated as:

$$N^{\text{estimation}}_{\text{QCD selected}} = F_{\text{sel-to-anti}} \times N^{\text{data}}_{\text{QCD anti-selected}} \tag{6.13}$$

In the following, the approach in defining the anti-selected sample for the electron and muon channels will be discussed, followed by the determination of the $F_{\text{sel-to-anti}}$ and final prediction of the *QCD* contribution in the search regions.

6.3.1 QCD estimation in the electron channel

As previously mentioned, the *QCD* contamination in the electron channel exceeds the muon one by almost an order of magnitude. The dominant sources of fake electrons are misidentified jets or converted photons. This motivates the requirements for the anti-selected sample.

Similar to the selected leptons described in Sec. 4.3.2, electron PF candidates are chosen. To ensure orthogonality of the samples, the anti-selected electrons are required to fail the *medium* WP criteria, but to satisfy the looser *veto* WP of the electron ID. Moreover, to enhance the fraction of misidentified jets, the ratio of the hadronic energy over the electromagnetic cluster energy (H/E) is required to be above 0.01. The additional requirements on the IP parameters d_{xy} and d_z as well as the photon conversion veto are removed, which increases the probability for the electron candidate to originate from a photon. The compatibility with the isolated electron trigger is ensured with a $\mathrm{I}_{\mathrm{mini}} < 0.4$ selection on the mini-isolation.

It should be noted, that events are first checked for an existing lepton satisfying the selected requirements, and in case none is found, the lepton candidates are evaluated in terms of the anti-selected criteria. This ensures, that anti-selected events do not contain good leptons.

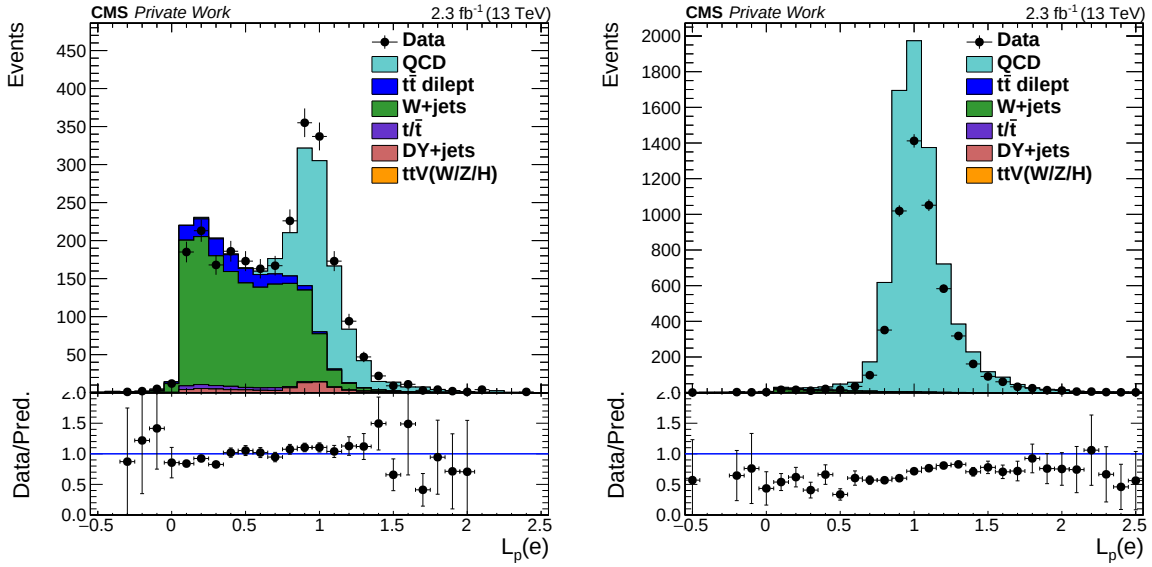


Fig. 6.14: L_p distributions for selected (*left*) and anti-selected (*right*) electron events after the baseline selection with modified requirements of $n_{jet} \in [3, 4]$ and zero b-tags. In this phase-space the selected events are composed of similar contributions from the *EWK* and *QCD* backgrounds, however the former is modelled better, than the latter. The anti-selected sample is dominated by *QCD* events, as intended for a fake electron enriched selection.

Figure 6.14 shows the L_p distribution for the nominal selected electron events on the *left* and for the anti-selection in the *right* plot. The baseline selection is modified according to the lepton type, and three or four jets are required together with zero tagged b jets. According to the simulation, the selected lepton events comprise a mixture of *EWK* and *QCD* backgrounds, with the former describing the data better than the latter.

In contrast, the anti-selection is dominated by *QCD* events, as intended by the electron identification requirements. Although the simulation shape is similar to data, the normalization does not agree, which is however expected from the nature of *QCD* processes.

A direct comparison of the L_p distribution shapes in selected and anti-selected samples is shown in Fig. 6.15 (*left*) for the corresponding events in simulation. Despite the discussed inaccuracy of the *QCD* modelling, the agreement of these shapes allows to assume a certain similarity as well for data. Therefore, it is possible to use the anti-selected data to obtain the L_p shape of selected *QCD* events in data.

L_p shape template fit

Using the L_p shape templates for *QCD* events together with an *EWK* template from simulation, it is possible to fit the data distribution in the selected sample, The *right* plot of Fig. 6.15 illustrates this fit for pseudo-data, that are obtained from a template of the combined expectation of simulated *EWK* and *QCD* events, shown by the shaded histograms.

The fit model[1] consists of the two p.d.f. templates of the L_p shapes (f_{EWK} and f_{QCD}), where

[1] Implemented in the RooFit [162] package.

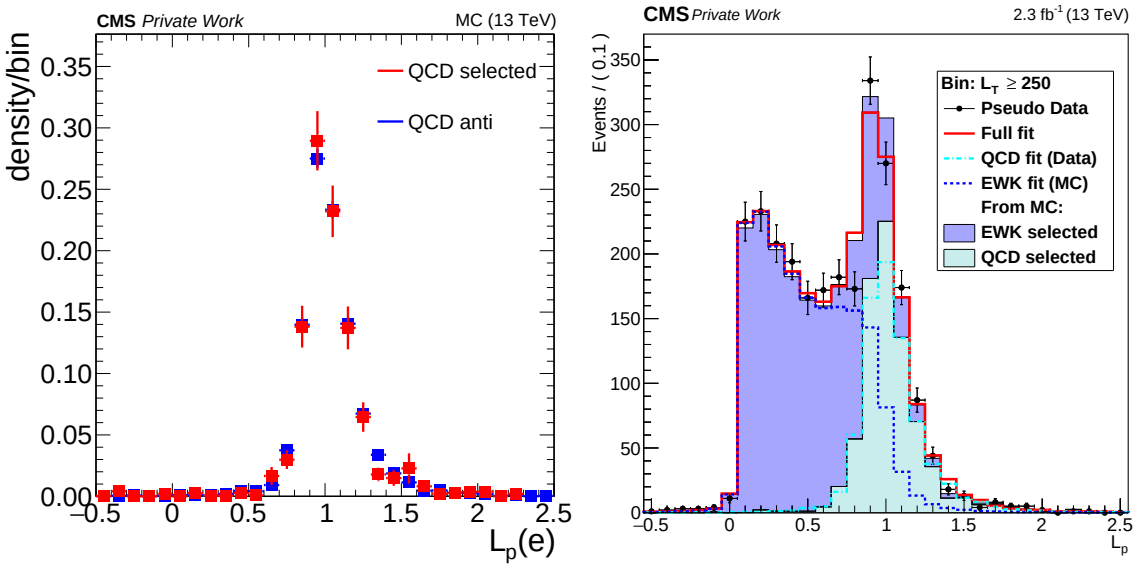


Fig. 6.15: *Left*: Comparison of normalized L_p shapes for simulated selected (red) and anti-selected (blue) electron events after the baseline selection with modified requirements of $n_{jet} \in [3, 4]$ and zero b-tags. *Right*: L_p shape fit for pseudo-data after the same selection as on the *left*. The simulated distributions of *EWK* and *QCD* events for the given luminosity are shown by the shaded histograms, which, combined, serve as a template for the pseudo-data (black) generation. The *EWK* shape (blue dashed) is taken from the same selected events, while the *QCD* template (cyan dashed-dotted) is obtained from the anti-selected sample.

both can float freely according to the corresponding normalization parameter (N_{EWK} or N_{QCD}):

$$f_{\text{Data}}(L_p) = N_{QCD} \cdot f_{QCD}(L_p) + N_{EWK} \cdot f_{EWK}(L_p) \, . \tag{6.14}$$

The resulting fit template is shown in Fig. 6.15 (*right*) by the red solid line, while the two components are indicated by the dashed lines for *EWK* (blue) and *QCD* (cyan).

The best agreement is naturally found in the lower L_p region, which is populated mainly by *EWK* events: the template used to generate pseudo-data and the fit is identical. However, also the *QCD* dominated region around $L_p \approx 1$, where the fit template stems from the anti-selected sample, shows a good performance of the fit.

The extracted number of selected *QCD* events allows to determine the $F_{\text{sel-to-anti}}$ according to Eq. 6.12. Since no shape uncertainties are propagated into the fit, only the fit parameter error provides an uncertainty on the number of selected *QCD* events N_{QCD}.

$F_{\text{sel-to-anti}}$ in kinematic regions

Until now, the $F_{\text{sel-to-anti}}$ was considered only in the phase-space after the modified baseline selection with three to four jets and zero b-tags. As the search is performed is several kinematic categories, and the ratio can be only measured in the dedicated sideband, the behaviour of the $F_{\text{sel-to-anti}}$ ratio has to be studied beforehand.

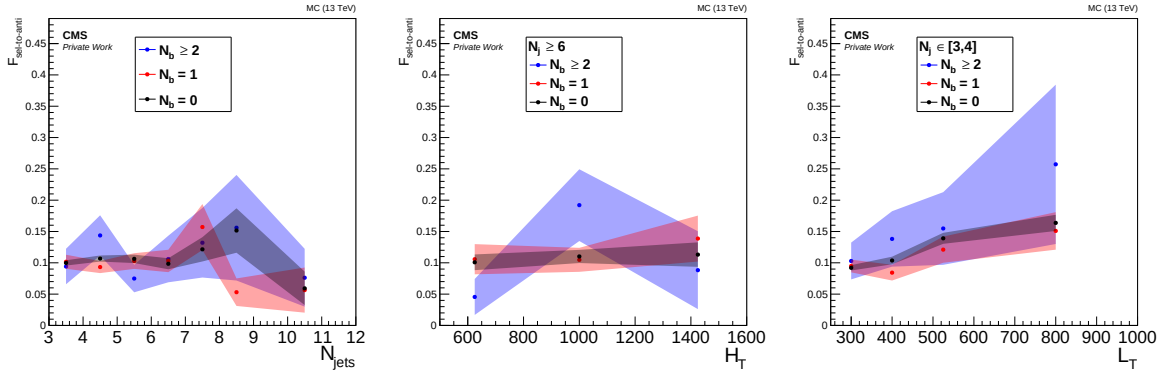


Fig. 6.16: Ratio $F_{\text{sel-to-anti}}$ of $N_{\text{QCD selected}}$ to $N_{\text{QCD anti-selected}}$ from simulated *QCD* yields for different $n_{\text{b-tag}}$ requirements after the baseline selection. The H_{T} (*left*) and n_{jet} (*middle*) plots show no dependence of the ratio, except for large statistical fluctuations in the kinematic tails. The L_{T} trend (*right*) is expected from the increasing misidentification probability with growing lepton p_{T}.

As discussed previously, the ratio $F_{\text{sel-to-anti}}$ is related to the probability of a fake lepton to pass the analysis identification criteria. Therefore, this probability should be independent of the hadronic activity of the event, quantified by the variables H_{T}, n_{jet}, $n_{\text{b-tag}}$ as well as the missing energy $\not{E}_{\text{T}}$. However, since the probability to misidentify a jet or converted photon increases with energy, the ratio is expected to change accordingly. The search categories are binned in L_{T}, thus the lepton p_{T} has to be considered together with $\not{E}_{\text{T}}$.

Figure 6.16 shows the $F_{\text{sel-to-anti}}$ from simulated *QCD* events for different b-tag multiplicities in various ranges of H_{T}, n_{jet} and L_{T}. As expected, the ratio seems to not depend on the jet multiplicity or H_{T} and also $n_{\text{b-tag}}$ within the statistical uncertainties. In general, the event samples are very small in the extreme kinematic tails with $n_{\text{b-tag}} \geq 2$, while lower b-tag multiplicities show similar values. Regarding the L_{T} dependence, here the $F_{\text{sel-to-anti}}$ is increasing with growing L_{T}, which is correlated with the increase in lepton momentum.

According to these findings, $F_{\text{sel-to-anti}}$ depends only on the L_{T} range. Therefore, it is sufficient to measure this ratio in bins of L_{T} for the baseline selection with $n_{\text{jet}} \in [3,4]$, $n_{\text{b-tag}} = 0$ and inclusive $H_{\text{T}} > 500\,\text{GeV}$ as:

$$F_{\text{sel-to-anti}}(L_{\text{T}}, n_{\text{b-tag}} = 0, n_{\text{jet}} \in [3,4]) = \frac{N^{\text{fit}}_{\text{QCD selected}}(L_{\text{T}}, n_{\text{b-tag}} = 0, n_{\text{jet}} \in [3,4])}{N^{\text{data}}_{\text{QCD anti-selected}}(L_{\text{T}}, n_{\text{b-tag}} = 0, n_{\text{jet}} \in [3,4])}\,, \tag{6.15}$$

where $N^{\text{fit}}_{\text{QCD selected}}$ is the number of selected *QCD* events as from the L_{p} template fit, and $N^{\text{data}}_{\text{QCD anti-selected}}$ the corresponding data yield in the anti-selected sample. $F_{\text{sel-to-anti}}(L_{\text{T}}, n_{\text{b-tag}} = 0, n_{\text{jet}} \in [3,4])$ will be shortly referred to as $F_{\text{sel-to-anti}}(L_{\text{T}})$.

Due to the low data event yield in the high L_{T} bins of the anti-selected sample, the quality of the *QCD* L_{p} shape template degrades by suffering from statistical fluctuations. But since the L_{p} shape of *QCD* events does not depend on the L_{T} interval, the fit is performed with the *QCD* template from the inclusive region $L_{\text{T}} > 250\,\text{GeV}$. Moreover, this allows to suppress the contribution from *EWK* events, which becomes more significant in the high L_{T} bins, altering the shape of the anti-selected L_{p} distribution. For the *EWK* template there is enough MC statistics

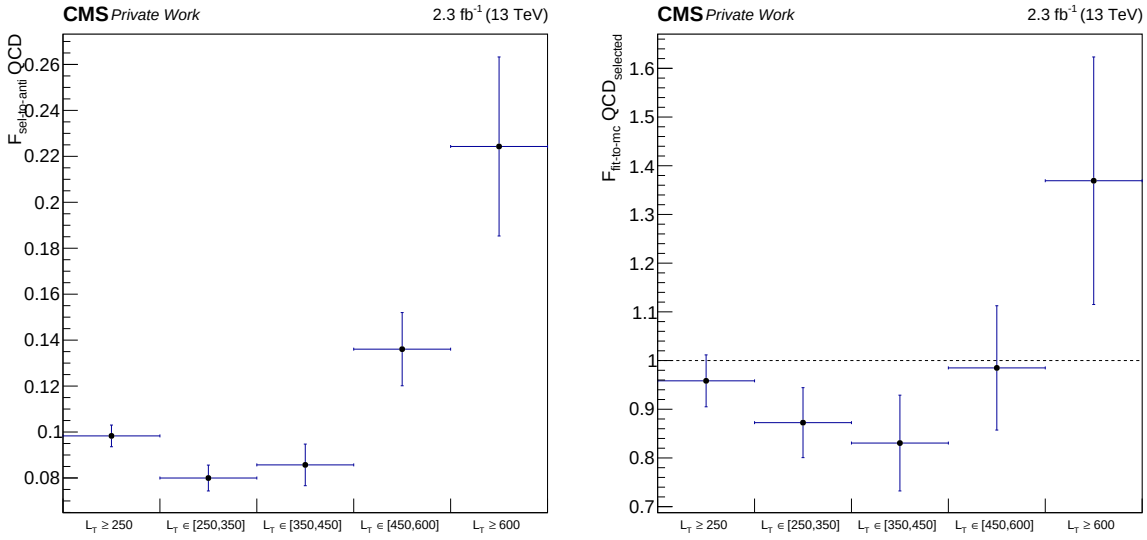


Fig. 6.17: The *left* plot shows the values of $F_{\text{sel-to-anti}}$ from the fits to pseudo-data in different L_{T} bins with $n_{\text{b-tag}} = 0, n_{\text{jet}} \in [3, 4]$. The measured ratios are in agreement with the values based on the MC yields. Due to the decreasing sample, the statistical uncertainties rise with L_{T}. The comparison of the number of selected QCD events from the fit and MC yields in the corresponding bins is shown on the *right*.

to obtain the templates for each L_{T} range.

The $F_{\text{sel-to-anti}}(L_{\text{T}})$ obtained from the L_{p} template fit of pseudo-data are shown in Fig. 6.17 (*left*) for different ranges in L_{T}. In general, the ratios are found to be in agreement with the expected (true) ones from the simulated QCD yields in Fig. 6.16, also showing the increasing $F_{\text{sel-to-anti}}$ with growing L_{T}. A closure test for the fit – the direct comparison of number of selected QCD events obtained from the fit to the event yield from MC – is shown in the right plot of Fig. 6.17. This closure test yields good agreement for the given H_{T}, $n_{\text{b-tag}}$ and n_{jet} requirements.

Systematic uncertainty on $F_{\text{sel-to-anti}}$

As previously shown in Fig. 6.16, the $F_{\text{sel-to-anti}}$ does not show any significant dependence on the b-tag and jet multiplicities throughout the H_{T} range. Hence, it is measured inclusively in H_{T}, and a systematic uncertainty can be derived based on the deviation from the central value in different H_{T}, n_{jet} and $n_{\text{b-tag}}$ bins.

Figure 6.18 shows the ratio of selected to anti-selected QCD yields in simulation for bins corresponding to the kinematic regions used for the R_{CS} prediction (cf. Sec. 6.2). The first bin represents the region, where the template fit is performed on data to obtain the selected to anti-selected transfer factors.

Overall, no trend is seen here as well, and except for high b-jet multiplicities, the deviations from the fit value are well below 75%. The grey mesh shows the chosen uncertainty values for each H_{T}-n_{jet} bin, while for $n_{\text{b-tag}} > 2$ an error of 100% is assumed, which is acceptable because of the low QCD contamination in this phase-space. The systematic uncertainty is added in quadrature to the error of the $F_{\text{sel-to-anti}}$ measurement.

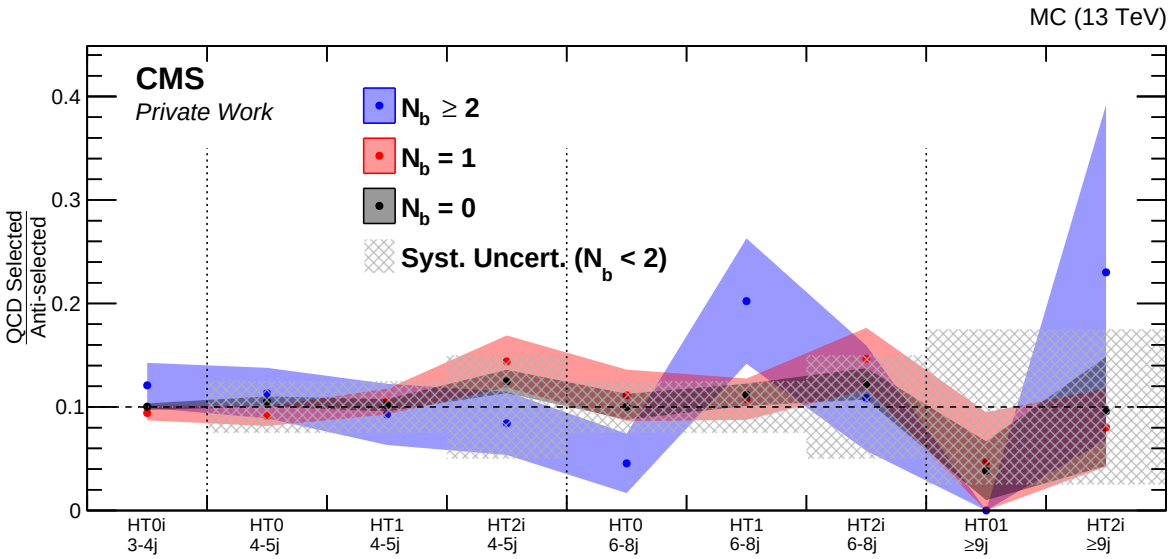


Fig. 6.18: $F_{\text{sel-to-anti}}$ from simulated QCD event yields for different n_{jet}, H_{T} and $n_{\text{b-tag}}$ bins with $L_{\text{T}} > 250\,\text{GeV}$. The grey mesh is showing the chosen systematic uncertainties in the corresponding H_{T} and n_{jet} bins for $n_{\text{b-tag}} < 2$. For higher b-jet multiplicities an uncertainty of 100% is assigned.

Prediction and closure test

After the $F_{\text{sel-to-anti}}$ (L_{T}) is measured in the $n_{\text{jet}} \in [3,4]$, $n_{\text{b-tag}} = 0$ and $H_{\text{T}} > 500\,\text{GeV}$ region, the same ratio is used for higher (b-) jet multiplicities and exclusive H_{T} bins to predict the QCD background contamination in selected events as:

$$N^{\text{pred}}_{\text{QCD selected}}(L_{\text{T}}, H_{\text{T}}, n_{\text{b-tag}}, n_{\text{jet}}) = F_{\text{sel-to-anti}}(L_{\text{T}}) \times N^{\text{data}}_{\text{QCD anti-selected}}(L_{\text{T}}, H_{\text{T}}, n_{\text{b-tag}}, n_{\text{jet}})\,. \tag{6.16}$$

where $N^{\text{data}}_{\text{QCD anti-selected}}(L_{\text{T}}, H_{\text{T}}, n_{\text{b-tag}}, n_{\text{jet}})$ represents the number of events in the anti-selected sample of the corresponding bin. In case of MC simulation, all SM backgrounds are considered in the anti-selected region.

A closure test allows to evaluate the QCD prediction method in simulated events. Figure 6.19 shows the comparison of the QCD yields in the selected sample as expected by the MC simulation and as predicted from the anti-selected yields. The prediction uses the $F_{\text{sel-to-anti}}$ (L_{T}) and systematic uncertainties as described before.

Overall, a good agreement between the estimated and real simulated QCD event yields is observed, where all bins are consistent within the uncertainties. The sideband closure is slightly better than the one in the mainband because of two reasons: *i)* less statistical fluctuations due to the larger event yield in the low jet multiplicity region, and *ii)* the high n_{jet} regions have a non-negligible contribution from EWK backgrounds in the anti-selected sample, which leads to a systematic overestimation of the QCD yield. However, since the QCD contamination in the mainband region is small, the lower accuracy of the QCD estimate has no significant effect on the R_{CS} determination.

After validating the QCD prediction method in simulation, the steps are repeated in data.

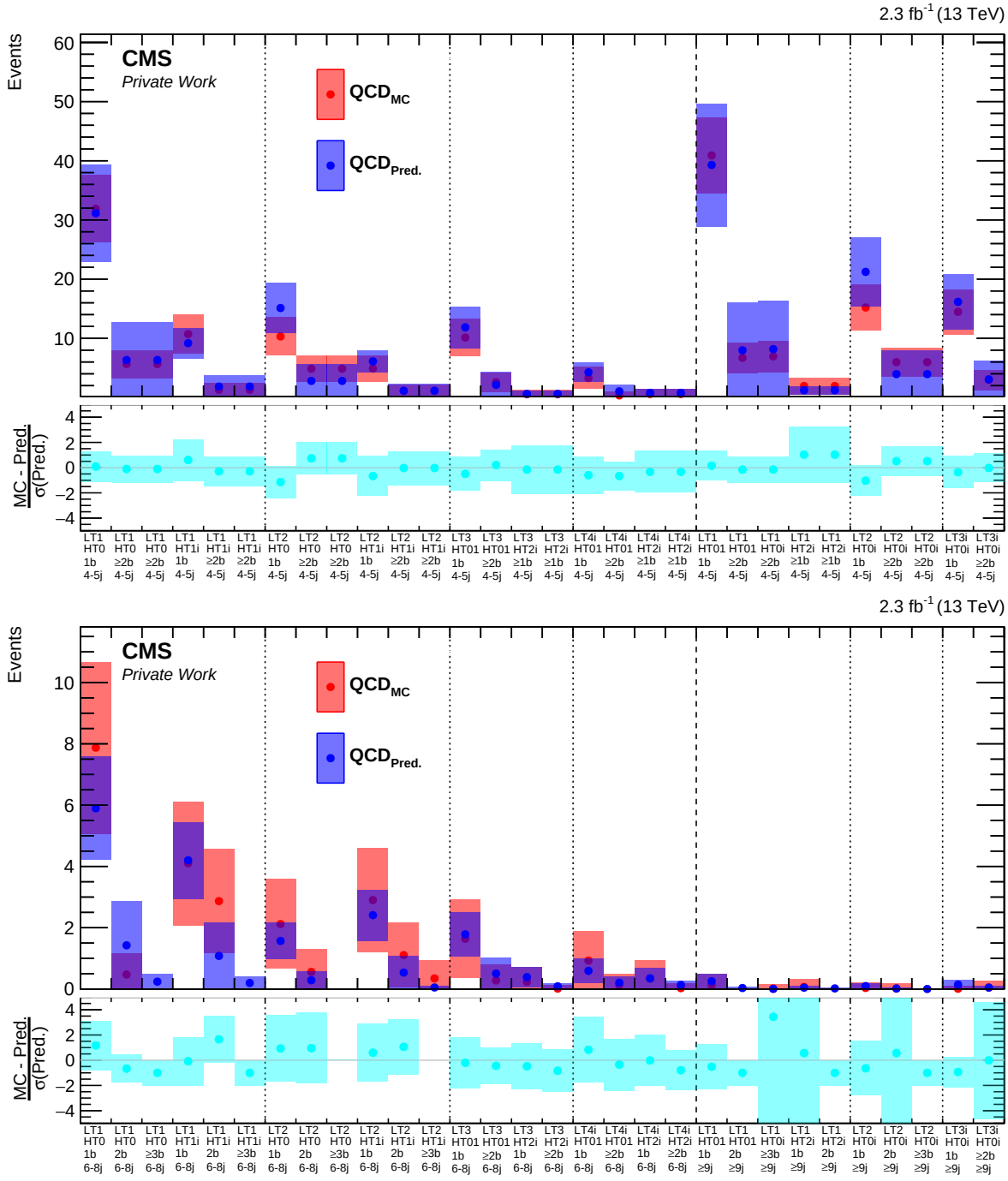


Fig. 6.19: Comparison of the number of selected *QCD* events in the control region as expected from simulation (red) with predicted yields estimated from the anti-selected sample for electrons. The $F_{\text{sel-to-anti}}$ is obtained in bins of L_{T} from simulation, and a systematic uncertainty is added as explained in the text. For the sideband (*top*) the agreement is better than for the corresponding mainband bins (*bottom*), due to higher statistics and negligible contribution from *EWK* events in the anti-selected sample.

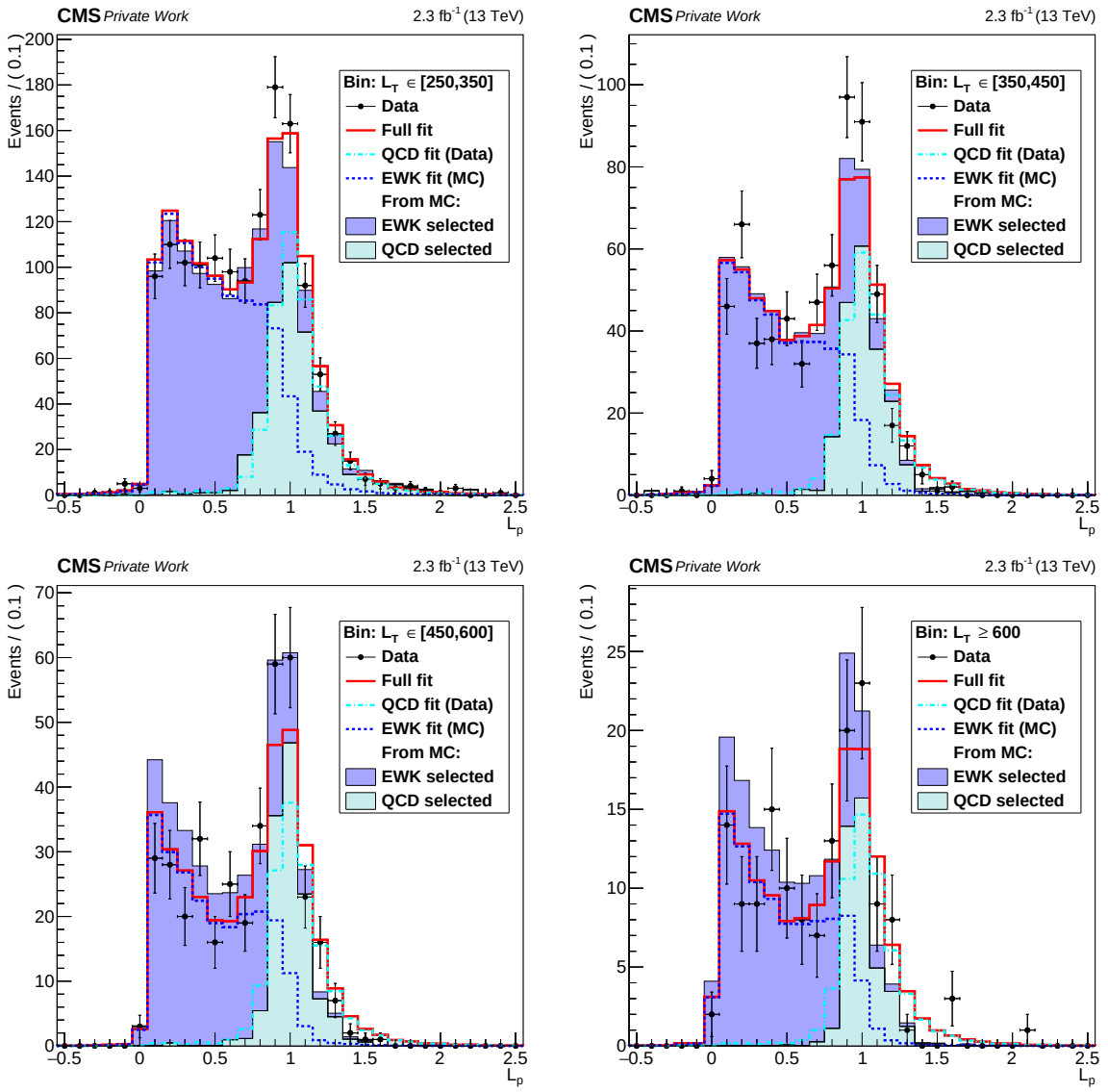


Fig. 6.20: L_{p} shape fit results in the data with $n_{\text{b-tag}} = 0$, $n_{\text{jet}} \in [3,4]$ for different L_{T} ranges. The black dots represent data. *EWK* and *QCD* selected events from simulation normalized to the luminosity are shown with the blue and cyan filled histograms. The full fit result is shown with the red solid line, while the separate *EWK* and *QCD* components are depicted by the blue and cyan dashed lines, respectively.

The results of the L_{p} shape fit performed on data is shown in Fig. 6.20, using the *QCD* template from the inclusive L_{T} region with anti-selected electrons and the *EWK* shape from simulation. Corresponding ratios of selected to anti-selected events for each L_{T} bin are given in the left plot of Fig. 6.21. The data repeats the L_{T}-dependence observed in Fig 6.17 (*left*). Nevertheless, the $F_{\text{sel-to-anti}}$ are higher than in simulation, since the trigger efficiency for anti-selected electrons with loose identification requirements is lower than for selected (as the trigger successfully rejects *QCD* events, cf. App. D.3).

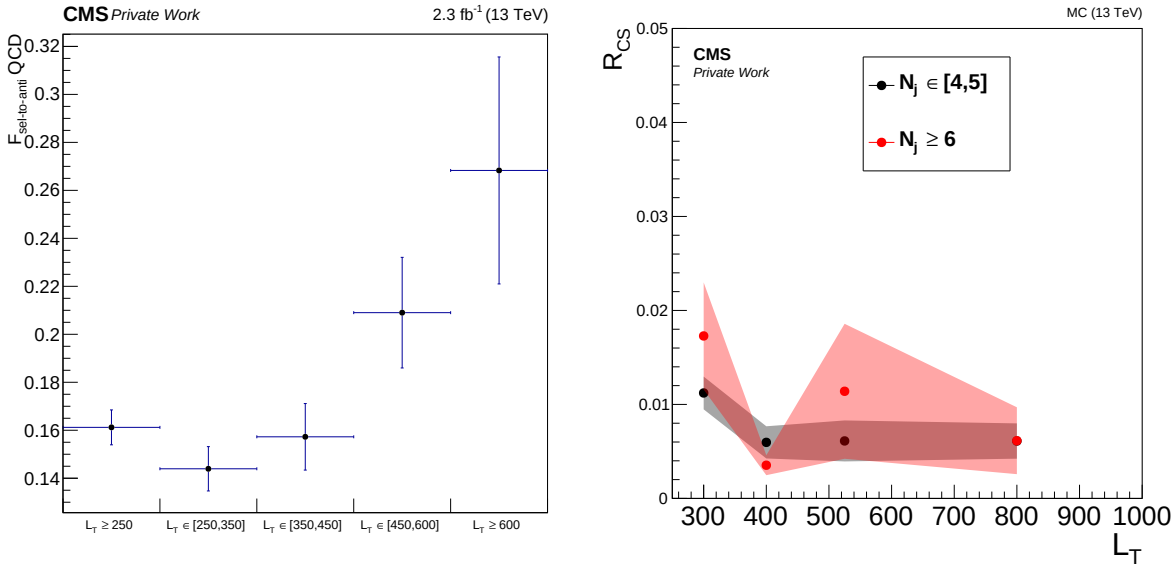


Fig. 6.21: Left: the corresponding ratio of selected to anti-selected events $F_{\text{sel-to-anti}}$ in L_T bins. Right: R_{CS} values for anti-selected electron events in different L_T and $n_{\text{b-tag}}$ bins for the n_{jet} = 4,5 and $\geq$6 in QCD MC.

Signal region contamination

As discussed earlier, selected *QCD* events mainly populate the control regions of the search bins and therefore do not have to be accounted for in the R_{CS} signal regions. The anti-selected electron sample allows to verify this hypothesis with simulation and data.

The right plot of Fig. 6.21 shows the R_{CS} values of simulated anti-selected *QCD* electron events for different L_T intervals. Within uncertainties, no significant dependence is observed, and the average R_{CS} of about 0.02 can be used throughout the whole L_T range as a conservative upper limit.

A comparison of the $\Delta\varphi$ shapes between selected and anti-selected electron events in simulation, as shown in Fig. 6.22 (*top*), allows to conclude, that the R_{CS} transfer factors are similar for both categories. The bottom plots of the same figure show the $\Delta\varphi$ distributions in data in comparison to the simulated SM backgrounds for the low and high jet multiplicity regions.

Given the good agreement of the shapes and the dominating contribution from *QCD* events, the aforementioned $R_{CS} \approx 0.02$, obtained from simulation, provides a good estimate for the *QCD* background in data. Since, as discussed in Sec. 6.3, the fraction of *QCD* events is below 10% in the control regions, the signal region contamination has to be at the per mille level, and therefore can be neglected.

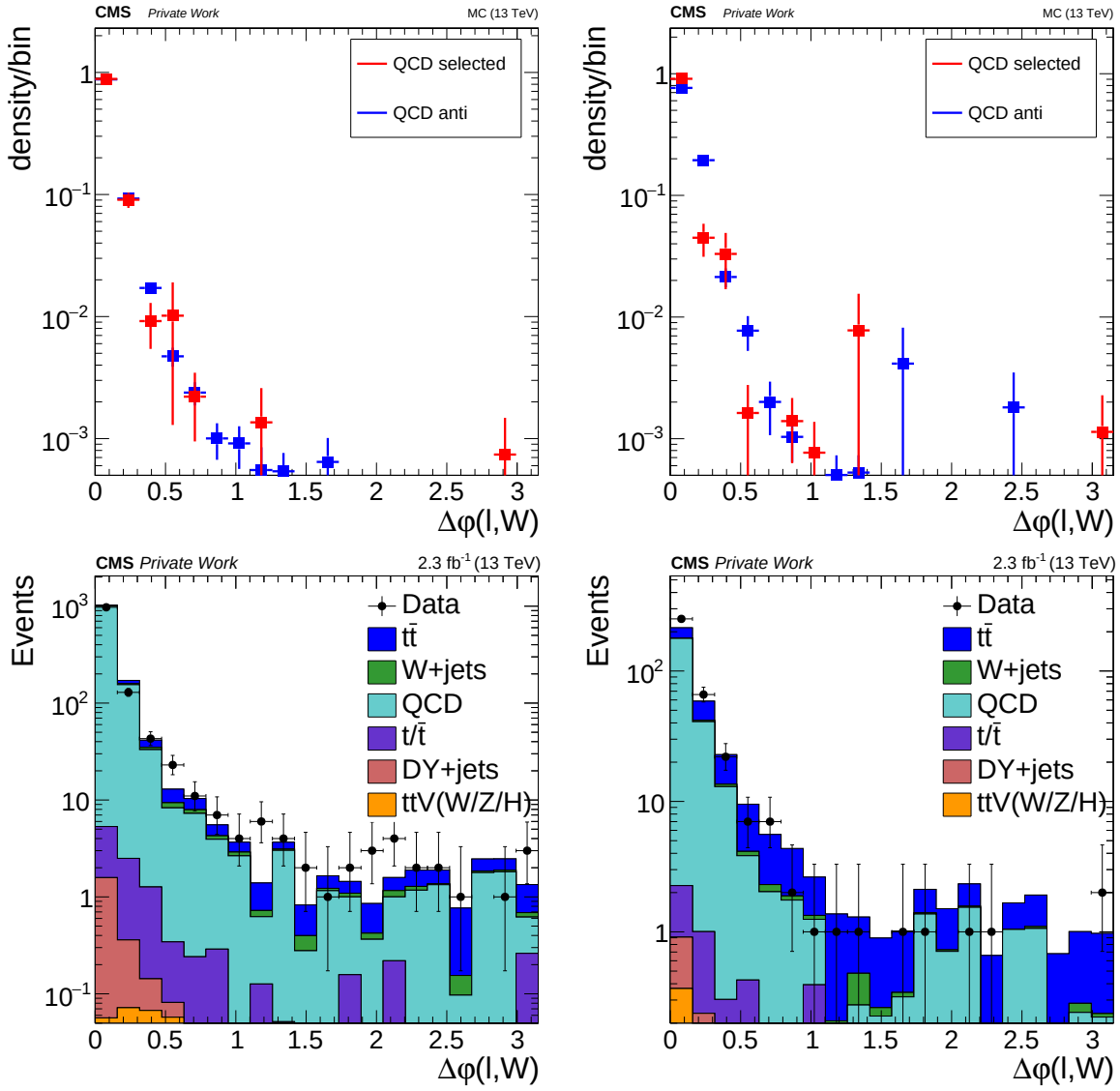


Fig. 6.22: *Top*: normalized $\Delta\varphi$ distributions for selected and anti-selected electron *QCD* events for $n_{\text{jet}} \in [3,4]$ (*left*) and $n_{\text{jet}} \geq 6$ (*right*) after the baseline selection with $L_{\text{T}} > 250\,\text{GeV}$ and $H_{\text{T}} > 500\,\text{GeV}$. The agreement in shapes indicates a similar R_{CS} transfer factor for both samples. *Bottom*: distribution of $\Delta\varphi$ for anti-selected electrons in data for $n_{\text{jet}} \in [3,4]$ (*left*) and $n_{\text{jet}} \geq 6$ (*right*) compared to the simulated backgrounds. The sample is mostly dominated by *QCD* events, and the good dataMC agreement allows to use the simulated R_{CS} factor for data as well.

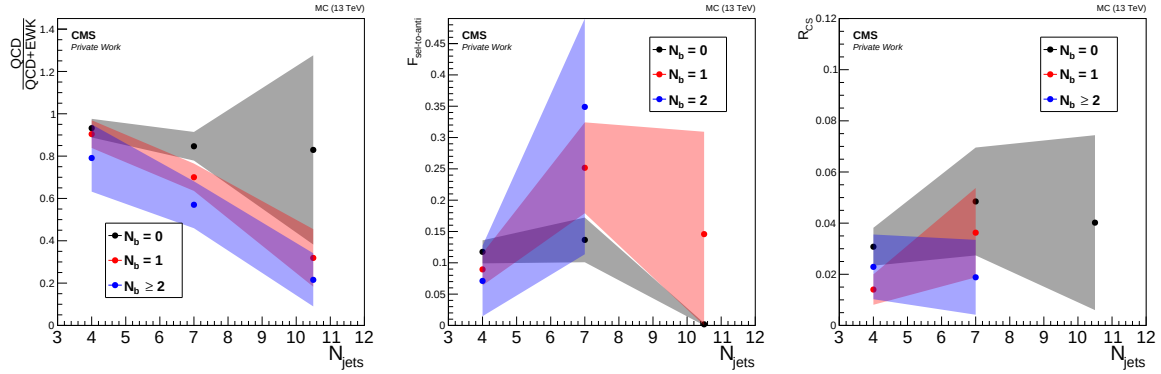


Fig. 6.23: Properties of the *QCD* background in simulated anti-selected muon events versus n_{jet} for $L_{\text{T}} > 250\,\text{GeV}$ and $H_{\text{T}} > 500\,\text{GeV}$: fraction of multijet events over the sum of QCD, $\text{t}\bar{\text{t}}$+jets and W+jets (*left*), the selected to anti-selected ratio $F_{\text{sel-to-anti}}$ (*middle*) and the R_{CS} factor (*right*).

6.3.2 QCD estimation in the muon channel

The main source of high p_{T} muons in the analysis selection are b-hadron decays, where the muon is not properly identified as isolated. This is a consequence of the mini-isolation with decreasing cone size. However, as already shown in Sec. 6.3, the overall *QCD* contamination in the muon channel is almost an order of magnitude lower than in the electron channel.

Nevertheless, to ensure a proper measurement of R_{CS}, especially in the low n_{jet} sideband region, the contribution of muons from multijet events is also taken into account. A procedure similar to the electrons is used, but due to the lower number of events, the estimation relies more on the simulation. This way, the overall effect of the *QCD* contamination will be treated in a controlled way and in the same fashion as for electrons.

Like in the case of electrons in Sec. 6.3.1, a fake muon enriched sample is created with a looser selection of candidates. The *anti-selected* muons are required to only pass the *loose* WP of the identification criteria, and the IP requirements on d_{xy} and d_z are relaxed. Moreover, the orthogonality of the sample is ensured by requiring the muons to be non-isolated with $0.2 < \text{I}_{\text{mini}} < 0.8$. An upped threshold is needed to comply with the trigger isolation, such that the MC selection is similar to the data.

As the fake muons mostly stem from b-quark jets, other SM backgrounds that produce b-quarks can also pass the anti-selection requirements. The contribution from the *QCD* events in the anti-selected sample is evaluated in simulation. The *left* plot of Fig. 6.23 shows the fraction of multijet events with respect to the total background from *QCD*, $\text{t}\bar{\text{t}}$+jets and W+jets. The decrease of the *QCD* contribution at high n_{jet} and $n_{\text{b-jet}}$ is expected since multijet events tend to have lower jet multiplicities than these *EWK* backgrounds. Nevertheless, the *QCD* fraction is conservatively assumed to always be 100% independent of the $n_{\text{b-tag}}$ and n_{jet} for the muon anti-selected sample.

Given the low *QCD* event yield in the muon channel, no trustworthy fit can be performed in the L_{p} distribution to determine the ratio of selected to anti-selected events in data. The middle plot of Fig. 6.23 shows the corresponding $F_{\text{sel-to-anti}}$ for muons in the *QCD* simulation. Any trend in this dependency is diminished by large statistical uncertainties, that are mainly caused by the small sample of selected muons. Taking the electron channel as a reference, no

large variation on the $F_{\text{sel-to-anti}}$ is expected and, hence, the value in the bin with the smallest uncertainty is taken as the average: 0.1 for $n_{\text{jet}} \in [3, 5]$.

Like in the case of electrons, the $F_{\text{sel-to-anti}}$ in data explicitly depends on the differences of the trigger efficiency for the (anti-)selected leptons. However, the muon trigger efficiency for these two categories is found to be fairly similar (see Appendix D.3: Fig. D.5), and therefore the estimated MC-based ratio can be used for the data prediction of the *QCD* background in the muon channel.

The total *QCD* yield for selected muons is estimated similar to Eq. 6.16 as:

$$N^{\text{pred}}_{\text{QCD selected}}(L_{\text{T}}, H_{\text{T}}, n_{\text{b-tag}}, n_{\text{jet}}) = F_{\text{sel-to-anti}} \times f_{\text{QCD/MC}} \times N^{\text{data}}_{\text{QCD anti-selected}}(L_{\text{T}}, H_{\text{T}}, n_{\text{b-tag}}, n_{\text{jet}}) \quad (6.17)$$

where the $F_{\text{sel-to-anti}}$ is assumed to be around 0.1 and the total *QCD* fraction $f_{\text{QCD/MC}}$ is conservatively set to 100% based on the simulation. Given the low event yields in the muon channel, as well as the simplistic estimation method, an uncertainty of 100% is assigned to the total predicted yield.

The contribution of muon *QCD* events in the signal regions is studied in analogy to the electron channel. According to simulation, selected muons have a similar $\Delta\varphi$ distribution as of the anti-selected sample, hence the R_{CS} transfer factor can be estimated from simulation. According to the *right* plot of Fig. 6.23, the R_{CS} value is well below 0.05, which can be taken as an upper limit. Taking into account the overall low *QCD* event yields in the muon channel, such a low R_{CS} renders the signal region contribution negligible.

Prediction in data

Figure 6.24 shows the final predictions of the selected *QCD* background based on the anti-selected samples in data, compared to the total data event yields in the control regions of the sideband and mainband. Similar to the simulation (cf. Fig. 6.13), the contribution of *QCD* events in the electron channel is estimated to be about 10%, and that of the muon channel at the percent level. Table 6.3 compares the total data yield to the number of estimated *QCD* events in the combined lepton channel. These predicted *QCD* yields are further used to correct the control region observation for the determination of the R_{CS} values according to Eq. 6.3.

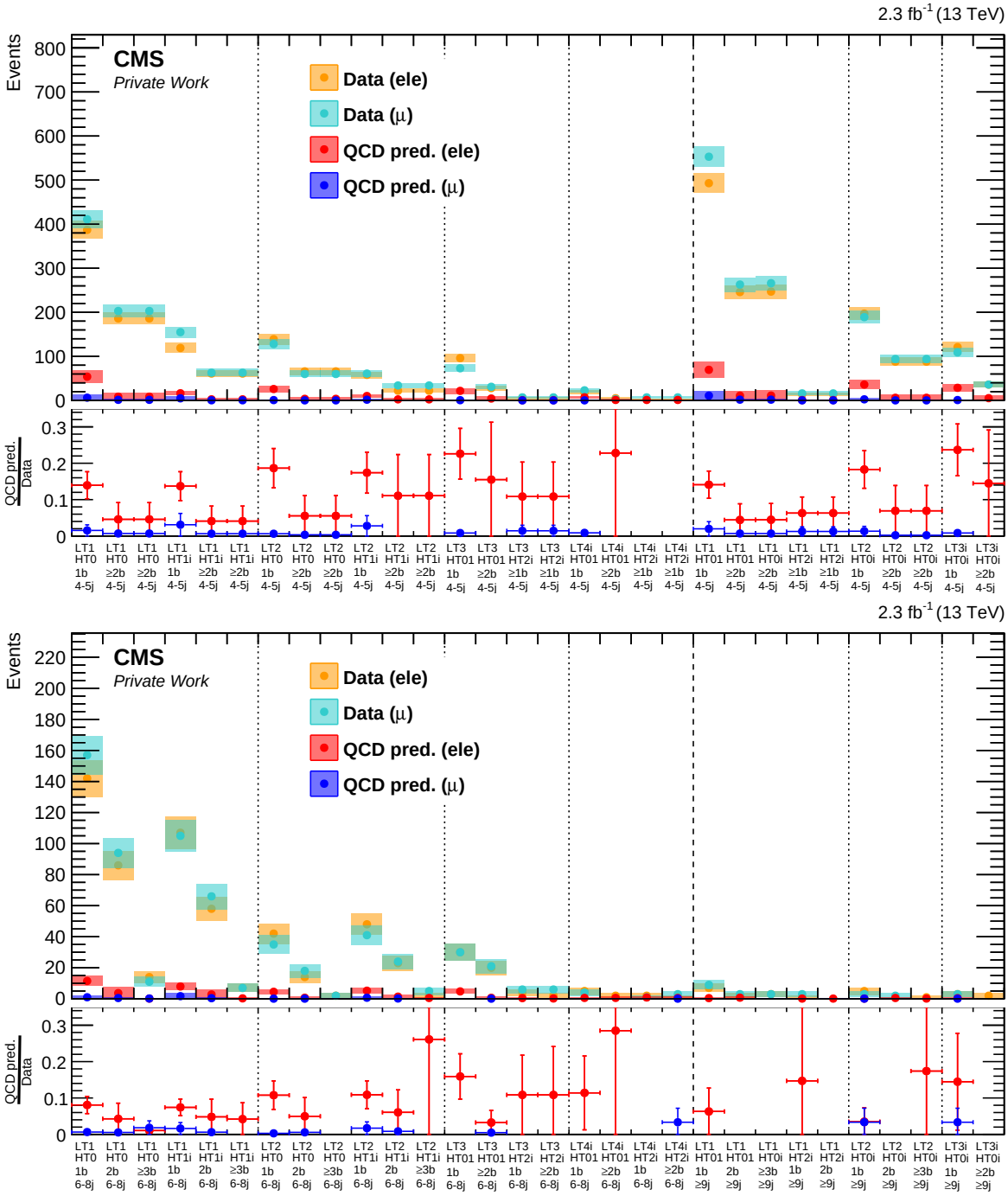


Fig. 6.24: Comparison of yields for selected electrons and muons with the data-driven *QCD* predictions in the sideband (*top*) and mainband (*bottom*) control regions. The average estimated contributions from the electron channel at about 10%, and in the muon channel at the percent level, agree with the expectation from simulation.

Table 6.3: Comparison of total data yields with the data-driven *QCD* predictions in the sideband and mainband control regions.

n_{jet}	L_{T} [GeV]	H_{T} [GeV]	$n_{\text{b-tag}}$	Data SB	*QCD* SB	Data MB	*QCD* MB
[6, 8]	[250, 350]	[500, 750]	$= 1$	798 ± 28.25	60.25 ± 15.51	299 ± 17.29	12.47 ± 3.38
			$= 2$	389 ± 19.72	9.93 ± 8.64	180 ± 13.42	4.17 ± 3.71
			≥ 3	389 ± 19.72	9.93 ± 8.64	25 ± 5	0.35 ± 0.25
		≥ 750	$= 1$	274 ± 16.55	21.12 ± 6.56	212 ± 14.56	9.64 ± 2.87
			$= 2$	124 ± 11.14	2.90 ± 2.53	124 ± 11.14	3.19 ± 2.82
			≥ 3	124 ± 11.14	2.90 ± 2.53	14 ± 3.74	0.29 ± 0.29
	[350, 450]	[500, 750]	$= 1$	267 ± 16.34	26.73 ± 7.22	77 ± 8.77	4.62 ± 1.49
			$= 2$	126 ± 11.22	3.85 ± 3.66	32 ± 5.66	0.80 ± 0.70
			≥ 3	126 ± 11.22	3.85 ± 3.66	4 ± 2	0 ± 0
		≥ 750	$= 1$	119 ± 10.91	11.79 ± 3.43	89 ± 9.43	5.92 ± 1.82
			$= 2$	56 ± 7.48	2.44 ± 2.44	47 ± 6.86	1.59 ± 1.41
			≥ 3	56 ± 7.48	2.44 ± 2.44	7 ± 2.65	0.52 ± 0.52
	[450, 600]	[500, 1250]	$= 1$	169 ± 13	22.30 ± 6.35	60 ± 7.75	4.77 ± 1.66
			≥ 2	59 ± 7.68	4.34 ± 4.34	41 ± 6.40	0.75 ± 0.66
		≥ 1250	$= 1$	13 ± 3.61	0.75 ± 0.51	10 ± 3.16	0.43 ± 0.38
			≥ 2	13 ± 3.61	0.75 ± 0.51	8 ± 2.83	0.22 ± 0.22
	≥ 600	[500, 1250]	$= 1$	41 ± 6.40	7.32 ± 2.61	9 ± 3	0.57 ± 0.44
			≥ 2	8 ± 2.83	1.14 ± 1.14	2 ± 1.41	0.57 ± 0.57
		≥ 1250	$= 1$	10 ± 3.16	1.14 ± 0.83	3 ± 1.73	0.85 ± 0.67
			≥ 2	10 ± 3.16	1.14 ± 0.83	4 ± 2	0.67 ± 0.58
≥ 9	[250, 350]	[500, 1250]	$= 1$	1046 ± 32.34	80.43 ± 21.20	16 ± 4	0.44 ± 0.42
			$= 2$	509 ± 22.56	12.68 ± 11.03	5 ± 2.24	0.73 ± 0.73
		≥ 500	≥ 3	513 ± 22.65	12.82 ± 11.17	6 ± 2.45	0 ± 0
		≥ 1250	$= 1$	30 ± 5.48	1.08 ± 0.61	4 ± 2	0.15 ± 0.15
			$= 2$	30 ± 5.48	1.08 ± 0.61	0 ± 0	0.15 ± 0.15
	[350, 450]	≥ 500	$= 1$	386 ± 19.65	38.52 ± 10.16	8 ± 2.83	0.27 ± 0.20
			$= 2$	182 ± 13.49	6.29 ± 6.09	2 ± 1.41	0.52 ± 0.52
			≥ 3	182 ± 13.49	6.29 ± 6.09	1 ± 1	0.17 ± 0.17
	≥ 450	≥ 500	$= 1$	230 ± 15.17	29.54 ± 8.27	6 ± 2.45	0.53 ± 0.33
			≥ 2	72 ± 8.49	5.21 ± 5.21	2 ± 1.41	0 ± 0

6.4 Systematic uncertainties

The data-driven estimation of the SM backgrounds in the signal regions (Eq. 6.3) consists of three types of components: *i)* the data observation in the sideband CR and SR, and the mainband CR yield, *ii)* the corresponding R_{CS} transfer factor κ for the *EWK* background calculated from simulation, and *iii)* the data-driven *QCD* estimate in the control regions.

As the uncertainty on the *QCD* estimate is already discussed in the previous section, the only remaining systematic uncertainty stems from the κ modelling by the simulation. In this section, the various sources that have an impact on the κ determination are discussed.

In addition to the κ factor, the modelling of signal processes also depends on the different aspects of simulation, affecting the expected event yield in the search regions. The sources of systematic uncertainties, that are common for signal and background modelling will be discussed together, followed by uncertainties, that are sensible only for the signal prediction.

6.4.1 Uncertainties on background estimation

As discussed in Sec. 6.2.1, the R_{CS} transfer factors, and hence the κ values, depend on the R_{CS} and fractions of the individual backgrounds. Uncertainties in the determination of the κ for all *EWK* backgrounds together arise from the mismodelling of those factors.

Since the data $\Delta\varphi$ distribution changes in the sideband in a similar way as in the mainband, the R_{CS} variations are already accounted for in the data-driven R_{CS} measurement. Therefore, the change of the κ value is evaluated only for deviations from the expected background composition, which reflect the individual fractions.

- **n_{jet} extrapolation for $t\bar{t}$+jets:**
  The impact of the dileptonic $t\bar{t}$ background, which has a large R_{CS} factor, is discussed in Sec. 6.2.2. An uncertainty on the dileptoninc fraction was derived from the discrepancy of the jet multiplicity distributions in the single-lepton and dileptonic samples. It is applied as a rescaling of the dileptonic contribution by a n_{jet}-dependent (slope) and constant weight:

  $$w(DL_{Const}) = 1 \pm 16\% \quad \text{and} \quad w(DL_{Slope}) = 1 \pm (n_{jet} - \langle n_{jet} \rangle) \cdot 4.5\% \tag{6.18}$$

  where n_{jet} is the jet multiplicity and $\langle n_{jet} \rangle$ is the weighted mean. All κ factors are calculated for the configurations with the plus and minus weight, and then compared to the nominal values. The largest deviation in the respective search bins is taken as the systematic uncertainty.

  As the offset variation mostly has no effect on κ, the slope yields the largest uncertainties with respect to other sources, due to the large statistical fluctuations in the dilepton control region. In the $n_{jet} \in [6, 8]$ region, the relative uncertainty on κ is found to be below 10%, while for $n_{jet} \geq 9$ it reaches up to 20% in the single search bins.

- **W+jets cross section:**
  Although the inclusive *W* boson production cross section is predicted at a percent level precision, the uncertainty becomes larger in the analysis phase space. Since W+jets events are the second most important background, this can lead to changes in the R_{CS} at low jet multiplicities of the sideband.

To account for this uncertainty, the inclusive W+jets cross section is conservatively varied by 30% above and below the nominal value. The impact on κ is evaluated and the largest deviation in each bin is assigned as the systematic uncertainty.

Despite the large variation of the W+jets cross section, the uncertainties on the total EWK κ do not exceed a few percent. Due to the similarity of the R_{CS} of W+jets and semileptonic $t\bar{t}$, their combined R_{CS} does not change much with this variation.

- W **boson polarization:**
  The angular distribution of the W boson decay products is correlated with the polarization of the mother particle. Hence, the $\Delta\varphi$ distribution, which reflects the angle between the lepton and reconstructed W boson, is also affected by the polarization. To account for this effect, events containing a W boson are scaled according to a weight:

$$w = 1 + x \cdot (1 - \cos(\theta^*))^2 \tag{6.19}$$

  where $\theta^*$is the azimuthal angle between the charged lepton and W boson in the W rest-frame, and x is the variation in the polarization as approximated by this formula. The uncertainty on κ is evaluated with $x = \pm 10\%$ for W+jets events, while for $t\bar{t}$+jets the variation is 5%, motivated by theory uncertainty predictions [163] and measurements [160, 164, 165].

  Provided the large leptonic scale requirement $L_T > 250\,\text{GeV}$, the W boson becomes boosted, hence rendering any angular dependence negligible. For most bins this uncertainty on κ is found to be below 1%.

- **Top quark** p_T **spectrum:**
  As discussed in Sec. 5.3.3, the top quark p_T spectrum in simulation is corrected to match the data. The weights for events containing a $t\bar{t}$ pair are calculated as:

$$w = \max\left[0.5, \sqrt{SF(t) \cdot SF(\bar{t})}\right] \text{ with } SF = \exp\left[0.156 \pm 0.00137 \cdot p_T(t/\bar{t})\right] \tag{6.20}$$

  The difference in κ between the nominal and rescaled events is taken as uncertainty.

  Since the $t\bar{t}$+jets background constitutes the largest fraction of all SM processes, this rescaling of its contribution does not yield major changes of the κ extrapolation factor. Overall, the systematic uncertainties are found to be within a few percent, not exceeding 4% in most cases.

- $t\bar{t}V$ **cross-section:**
  The contribution of $t\bar{t}V$ events is conservatively varied by 100% to account for the uncertainty on the theoretical prediction. In general, the contribution from $t\bar{t}V$ events is low in most of the search bins. For the bins, where this background becomes significant (with $n_{\text{b-tag}} \geq 2$ and $n_{\text{jet}} \geq 9$), the uncertainty does not exceed 5%.

Overall, the largest systematic uncertainty on κ stems from the difference in the n_{jet} distribution for single- and dileptonic events, measured in data. All uncertainties are considered uncorrelated, and therefore added in quadrature. An overview of the systematic uncertainties on the κ measurement in simulation due to the individual sources is shown in Fig. 6.25 and Table 6.4.

6.4.2 Uncertainties on background and signal prediction

In addition to the mismodelling of the background composition, systematic uncertainties may arise from the mismeasurement and misidentification of particular objects in the single events. Since these inefficiencies are a general feature of the event reconstruction, they are evaluated for all MC samples, i.e. the SM backgrounds as well as for the signal.

For the background samples, these uncertainties lead to changes in the individual $\Delta\varphi$ shapes, thus R_{CS} values, and, in less extend, to the total background composition. In the case of signal, these uncertainties affect the acceptance and selection efficiency, and therefore are evaluated for the signal region predictions.

- **Jet Energy Scale:**
  As discussed in Sec. 4.3.3, the reconstructed jet energy is corrected for different mismeasurement effects by the JEC. The accuracy of these corrections is evaluated in dedicated studies [140, 145], and uncertainties are obtained on the Jet Energy Scale (JES), i.e. the relative difference between the true and measured energies.

  These uncertainties are propagated into the jet energy measurement as additional correction factors corresponding to the $\pm 1\sigma$ variations depending on the jet p_T and η. For each single event this leads to the re-evaluation of all kinematic values related to the jet activity. Effected are the number of good jets and b-tagged jets, as well as the hadronic scale H_T. Since the missing energy is computed from JEC calibrated jets, the changes are also propagated to the corrected $\not{E}_T$.

  These variations lead to a change in the selection efficiency dictated by the threshold values for the analysis values like H_T, L_T and the jet multiplicities, as well as to migration of events between the different search bins. In case of the background estimation, the systematic uncertainties on κ vary from as low as 1% up to 10% in a single bin with $L_T > 600\,\text{GeV}$. For the signal, the relative uncertainty on the event yield prediction in the signal region varies in a range from 5% to 10%.

- **B-tagging efficiencies**:
  The b-jet identification efficiency in simulation is corrected to match the one in data using dedicated event weights as explained in Sec. 5.3.3. Like in the case of JEC corrections, uncertainties are measured for these b-tagging scale factors [141].

  The b-quark and c-quark jet efficiencies are varied simultaneously, while the efficiencies for jets from light quarks and gluons are varied independently. For each configuration the simulated event yields are evaluated, and the changes in the background and signal prediction are propagated as systematic uncertainties.

  The uncertainties for the heavy and light flavour quark variations are treated as uncorrelated and found to be around 1% and 5% for the background κ factor. For the signal predictions this systematic uncertainties are on average 6% and 2%, respectively.

- **Lepton identification and reconstruction efficiencies**:
  The differences in the identification and reconstruction efficiencies between the MC simulation and data are evaluated using "tag-and-probe" methods [125, 157]. For the full simulation of the CMS detector those discrepancies are usually below 2% for the leptons considered in this analysis, and depend on the lepton p_T and η.

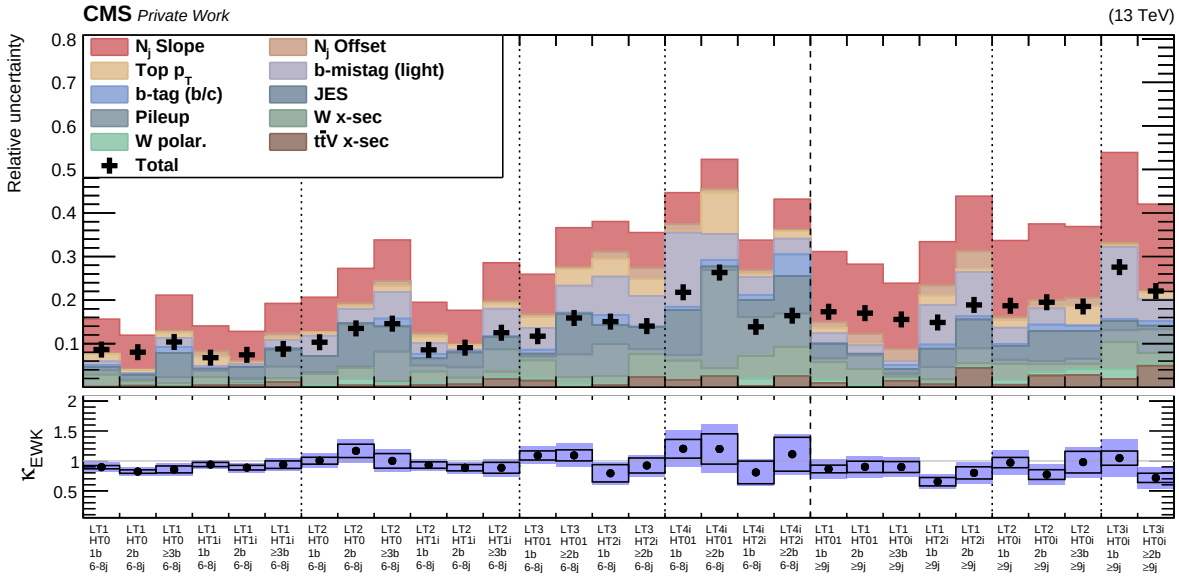


Fig. 6.25: Systematic uncertainties on the κ extrapolation for the simulated *EWK* background. *Top*: relative uncertainties for the individual sources stacked linearly on top of each other. The black crosses represent the total uncertainty. *Bottom*: nominal κ values with statistical uncertainties represented by the black box and the combined systematic and statistical uncertainties shown by the blue area.

Since most SM backgrounds have similar η profiles, and the search is performed in bins of L_T, these uncertainties are cancelled out when determining the κ factor. However, in the case of fast simulation (cf. Sec. 2.3.2), which is used to generate the signal samples, the differences to data are larger, and estimated with a fixed uncertainty of 5%. This variation directly translates to a systematic uncertainty of 5% on the signal expectations.

- **Pileup**:
  The average number of interactions in data and simulation differs, such that the simulated samples have to be corrected to match the pileup distribution in data (cf. Sec. 5.3.3).

  However, the large difference in the average number of interactions leads to high event weights, which enhance statistical fluctuations. To avoid these artificial uncertainties, no pileup rescaling is applied to obtain the background and signal predictions.

  Nevertheless, the possible effect of pileup is evaluated by comparing the predictions obtained with the nominal data distribution, assuming an inelastic cross section of 69 mb, and alternative pileup distributions, where this value is changed by 5% up and down. Overall, the uncertainties on the κ factor and signal prediction remain moderate below 5%, but are mostly driven by statistical fluctuations.

The total systematic uncertainty on the background prediction, or κ, is shown in Fig. 6.25. In general, the systematic uncertainty reaches from 10% for bins with low L_T to 20%, and is mainly driven by the n_{jet} extrapolation uncertainty, but also by the low event yield in extreme regions.

Table 6.4: Systematic uncertainties (in %) on the *EWK* κ factor from simulation.

n_{jet}	L_{T} [GeV]	H_{T} [GeV]	$n_{\text{b-tag}}$	$\bar{\text{t}}\text{tV}$ cross section	W polarization	W cross section	Pileup	JEC	b-tag (b,c)	b-mistag (light)	Top quark p_{T}	n_{jet} Offset	n_{jet} Slope	Total
[6, 8]	[250, 350]	[500, 750]	$= 1$	0.13	0.01	2.77	0.97	0.76	0.48	0.83	1.71	0.12	7.85	8.64
			$= 2$	0.36	0.13	0.90	0.23	1.25	0.23	0.10	0.74	0.18	7.84	8.04
			≥ 3	0.13	0.26	0.56	1.33	5.55	1.42	2.04	1.29	0.39	8.18	10.38
		≥ 750	$= 1$	0.43	0.18	1.78	1.44	0.28	0.28	0.46	2.16	1.22	5.83	6.78
			$= 2$	0.48	0.16	0.31	1.04	2.64	0.09	0.61	0.29	0.40	6.79	7.43
			≥ 3	1.15	0.16	0.74	2.71	3.93	0.35	1.86	0.56	0.88	6.92	8.78
	[350, 450]	[500, 750]	$= 1$	0.02	0.51	2.49	0.45	3.68	0.06	4.63	0.75	0.06	8.02	10.32
			$= 2$	0.41	1.49	2.52	0.31	9.90	0.07	3.31	0.70	0.60	7.98	13.50
			≥ 3	0.33	0.68	0.30	6.86	5.79	1.79	6.08	1.39	1.20	9.39	14.59
		≥ 750	$= 1$	0.49	0.33	2.79	1.38	1.64	0.96	2.48	1.59	0.72	7.11	8.57
			$= 2$	0.65	0.09	1.50	2.32	3.44	0.34	0.28	1.05	0.18	7.82	9.07
			≥ 3	1.84	0.38	1.37	5.11	2.80	0.25	6.26	1.22	0.61	8.72	12.51
	[450, 600]	[500, 1250]	$= 1$	1.52	0.64	4.00	0.91	0.62	0.91	4.96	2.61	0.56	9.20	11.71
			≥ 2	0.10	0.74	1.47	5.23	9.35	0.22	6.21	3.89	0.41	9.01	15.89
		≥ 1250	$= 1$	0.44	0.32	1.79	7.35	4.37	2.31	8.81	4.09	1.65	6.91	15.06
			≥ 2	2.31	0.02	5.24	1.19	5.06	0.13	7.04	3.79	2.59	8.14	14.03
	≥ 600	[500, 1250]	$= 1$	1.68	0.05	4.38	1.29	10.32	0.76	17.00	0.32	1.64	7.24	21.79
			≥ 2	2.53	0.38	1.44	22.56	0.83	1.51	5.97	9.52	0.83	6.76	26.33
		≥ 1250	$= 1$	0.30	1.69	5.14	8.98	3.92	1.18	4.13	0.65	0.93	6.85	13.86
			≥ 2	2.55	0.36	6.37	7.64	8.62	4.96	3.60	1.78	0.30	7.00	16.42
≥ 9	[250, 350]	[500, 1250]	$= 1$	0.95	0.80	3.98	0.90	3.36	0.14	2.26	1.29	1.29	16.19	17.32
			$= 2$	0.08	0.40	3.70	3.13	0.33	0.06	1.91	0.29	2.34	16.01	17.00
		≥ 500	≥ 3	1.47	0.25	0.75	0.63	1.10	0.85	0.93	0.05	2.73	15.12	15.56
		≥ 1250	$= 1$	0.65	0.19	0.98	2.79	4.24	0.94	9.05	2.21	2.33	10.03	14.86
			$= 2$	4.45	0.24	0.89	3.34	6.64	0.73	10.16	0.68	4.08	12.65	18.89
	[350, 450]	≥ 500	$= 1$	0.57	1.13	3.70	0.92	3.09	0.55	3.70	1.82	0.70	17.50	18.70
			$= 2$	2.68	0.89	1.69	0.73	6.86	1.52	3.73	0.26	1.72	17.40	19.51
			≥ 3	2.78	1.33	1.29	1.16	6.34	1.21	0.20	4.34	1.90	16.34	18.53
	≥ 450	≥ 500	$= 1$	1.88	2.33	6.10	2.75	2.11	0.47	16.55	0.35	0.68	20.67	27.57
			≥ 2	4.94	0.63	2.24	5.47	0.85	1.01	5.01	1.73	0.21	19.95	22.09
n_{jet}	L_{T} [GeV]	H_{T} [GeV]	$n_{\text{b-tag}}$	$\sigma(\bar{\text{t}}\text{tV})$	W pol.	$\sigma(W)$	PU	JEC	b-tag(b)	b-tag(l)	Top p_{T}	n_{jet} Off.	n_{jet} Slope	Total

6.4.3 Uncertainties on signal prediction

The following uncertainties are applied only to the simulated signal, since they directly affect the normalization of the prediction, or are specific to the physical process of the gluino pair production. Since these effects can have different impacts depending on the gluino-neutralino mass difference, the uncertainties are evaluated separately for each mass point. In the case of simulated background samples, those uncertainties are either not applicable, or cancel out in the R_{CS} or κ ratios.

- **Initial-state radiation**

  During *Run 1* it was observed, that the MADGRAPH event generator does not accurately describe the hadronic recoil from ISR for boosted heavy particle pairs, like the $t\bar{t}$ system [166]. The gluino pair of the discussed SUSY model can be considered as a very similar system, and therefore discrepancies have to be expected.

  However, since those measurements have not been repeated yet for the *Run 2* data, only a rough estimate of the uncertainty is taken. A per-event weight is varied based on the combined p_T of the gluino pair system according to:

  $$\begin{aligned} p_T \text{ (gluino pair)} &\leq 400\,\text{GeV}: && \text{no uncertainty} \\ 400 < p_T \text{ (gluino pair)} &< 600\,\text{GeV}: && 15\% \text{ uncertainty} \\ p_T \text{ (gluino pair)} &\geq 600\,\text{GeV}: && 30\% \text{ uncertainty} \end{aligned}$$

  The largest difference in the event yield prediction of the signal in each bin is extracted for each mass scenario. Overall, this systematic uncertainty becomes more prominent for compressed signal scenarios with a small gluino-neutralino mass splitting, where it reaches up to 20% in high L_T regions.

- **Signal cross section**

  Despite being calculated at the highest available order[2], the cross section of the gluino pair production shows large theoretical uncertainties, which increase with growing gluino mass. Naturally, this has a direct impact on the signal normalization, and hence the predicted event yield.

  However, as this is a purely theoretical uncertainty, it is not used directly for the search experiment, but only for the signal model hypothesis testing (cf. next Chapter, Sec. 7.2).

- **Renormalization and factorization scale**:

  The renormalization and factorization scales, that were discussed in Sec. 2.3.1, are to some extend arbitrary parameters used for the calculation of the matrix element. To account for the impact of the choice of these scales on the signal acceptance, the predictions are evaluated for their variations (0.5 and 2 times of the original parameter), excluding the non-physical anti-correlations. Since these variations are correlated with the production cross section, the impact is normalized to the nominal value, keeping the cross section constant.

  An envelope of all individual variations is computed and a systematic uncertainty derived for each search bin. The overall impact is below 2% for most mass points throughout the search phase space.

[2] NLO+NNL, cf. Fig. 1.6

- **Luminosity**

  The precision on integrated luminosity measurement of the CMS data used in this analysis is 2.7% [96]. This systematic uncertainty is propagated directly to the expected signal yields.

- **Trigger efficiency**

  As discussed in Sec. 5.2.2, the uncertainty on the trigger selection efficiency is measured to be 1%, and also accounted for in the signal prediction.

In general, the systematic uncertainties on the signal prediction are similar between the different mass scenarios and lie around 15%. In certain search bins, like those with large L_T, the expected mismodelling of the recoil from ISR drives the systematic uncertainty up to 30% for compressed scenarios.

Figure 6.26 shows the individual contributions from the various uncertainty sources on the predictions for the two T1tttt benchmark scenarios. Except for the few bins with expected yields close to one, in the majority of the bins the prediction is dominated by statistical uncertainties as shown by the black boxes in the lower pads. Table. 6.5 shows the average relative systematic uncertainties for the individual sources for the same model points.

Table 6.5: Average relative systematic uncertainties (in %) on the simulated event yields for the signal benchmark scenarios T1tttt (1200,800) and T1tttt (1500,100).

Uncertainty source	T1tttt (1200,800)	T1tttt (1500,100)
Trigger	1	1
Pileup	5	5
Lepton efficiency	5	5
Luminosity	2.7	2.7
Initial-state radiation	3-20 (high L_T)	2
b-tag (b,c)	6	6-10
b-mistag (light)	2	2
Jet Energy Scale	5-10	5-10
Factorization / renormalization scale	< 2	<2
Total	12-25	11-16

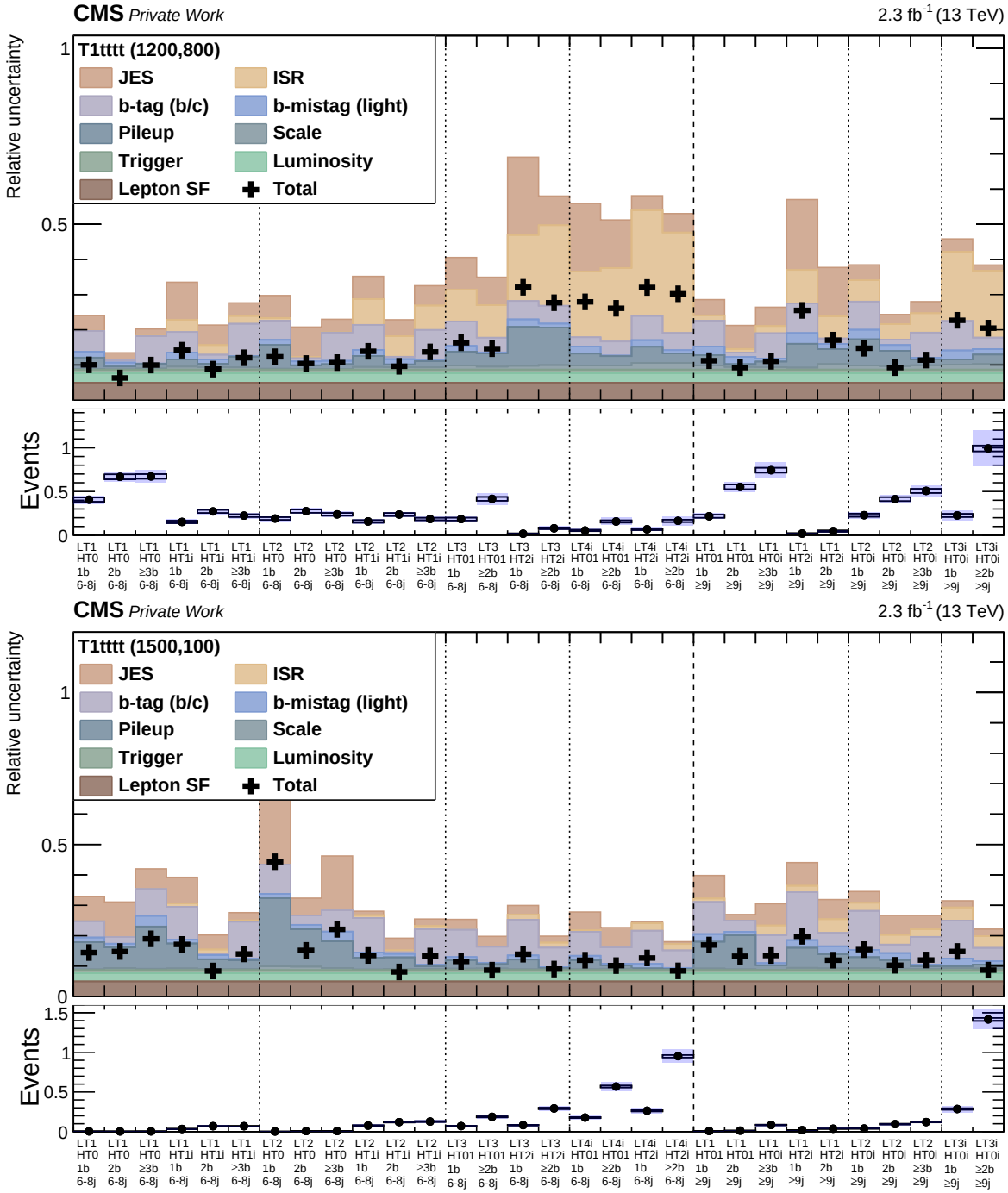


Fig. 6.26: Systematic uncertainties on the signal yields for the T1tttt (1200,800) (*top*) and T1tttt (1500,100) (*bottom*) signal points. While the non-compressed scenario shows an overall flat total uncertainty of about 15%, the compressed scenario suffers from the increased ISR uncertainty in the high L_T bins increasing the total up to 30%.
The *top* part of the plot shows the relative uncertainty on the nominal yield for the different sources of uncertainties. The various uncertainties are stacked linearly on top of each other, while the black crosses represent the total. The *bottom* plot shows the nominal signal yields with the statistical uncertainty represented by the black box and the combined systematic and statistical uncertainties shown by the blue area.

6.5 Prediction in data

After the background estimation method was discussed, and the systematic uncertainties were evaluated, the final prediction of the SM backgrounds in the signal regions is obtained. The estimation of the QCD background in data, as required by the R_{CS} method, was already discussed in Sec. 6.3.2.

R_{CS} data-MC comparison

Figure 6.27 shows the R_{CS} values in the sideband as expected from the EWK background simulation and as measured in data. Even tough the R_{CS} transfer factors are not required to be exactly the same, an agreement in the bin-to-bin trends allows to assume a realistic representation of the data by the MC simulation. This provides confidence in the correctness of the EWK κ factor modelling. However, the amount of currently available data in several bins leads to dominating statistical, rather than systematic, uncertainties.

Signal region prediction

The full data-driven background prediction is shown in Fig. 6.28, compared to the expected SM background event yield from simulation. The red mesh corresponds to the combined statistical and systematic uncertainties of the prediction. Overall, the prediction agrees with the expectations from the simulation without any systematic over- or underestimation. The categories with $n_{jet} \in [6, 8]$ have naturally higher event yields, while the bins with at least nine jets show big statistical fluctuations with no single bin expecting more than one event.

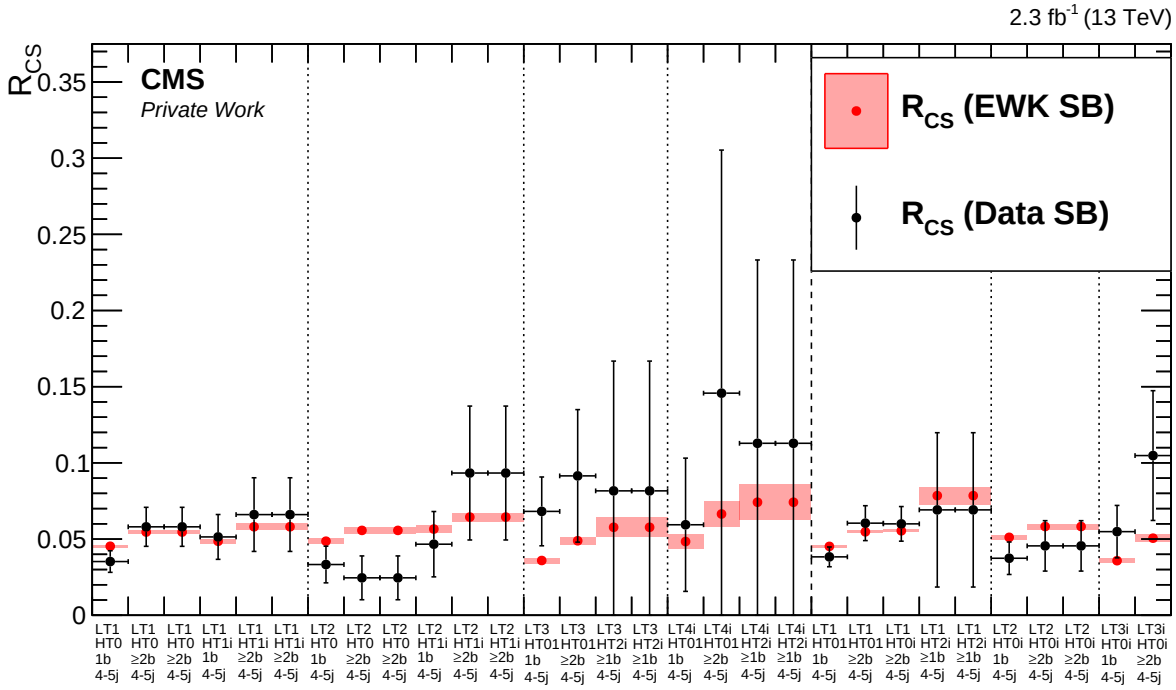


Fig. 6.27: R_{CS} values in the sideband regions as expected from the *EWK* backgrounds in simulation (red) and as observed in data (black). The error bands correspond to the statistical uncertainties only. Despite some large statistical uncertainties, the agreement is satisfactory in the majority of the categories.

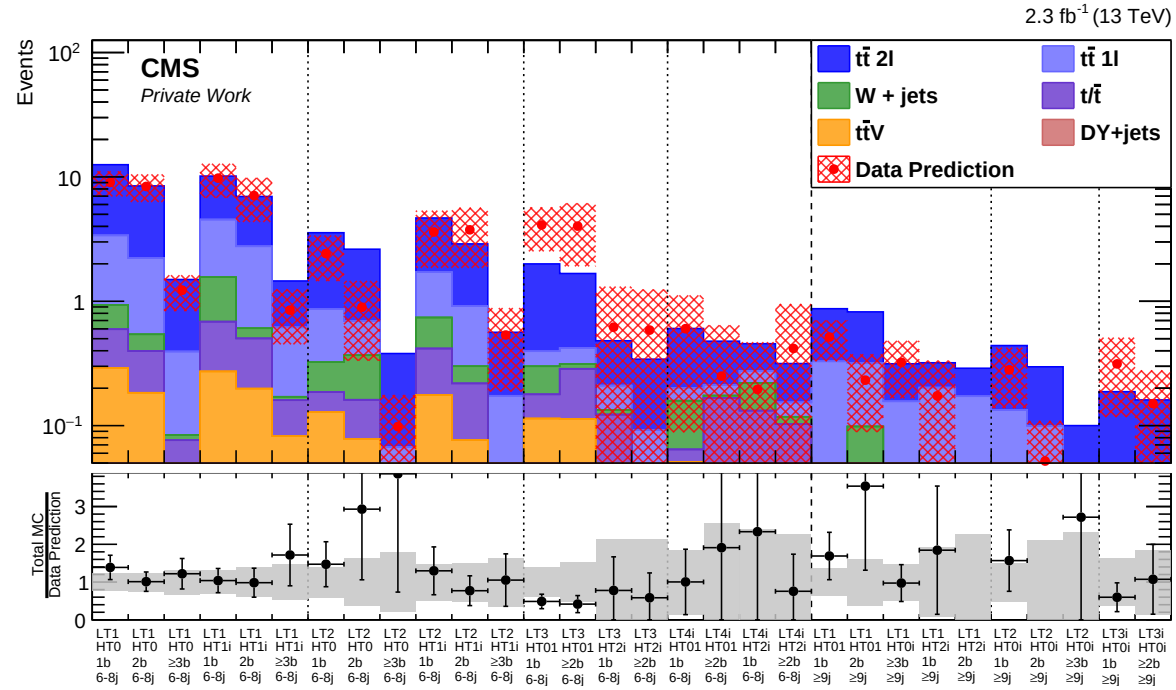


Fig. 6.28: Comparison of simulated event yields and the data-driven predictions in the signal regions of the mainband. The lower panel shows the ratio of the total background from simulation over the data prediction. The grey band shows the total uncertainty on the prediction. In general, a good agreement is observed for the majority of the bins, with no systematic over- or underestimation.

Chapter 7

Results and Interpretation

7.1 Results **139**
7.2 Limit setting procedure **143**
7.2.1 The CL_s method for statistical hypothesis testing 143
7.2.2 Statistical interpretation of results 145
7.3 Interpretation **148**
7.3.1 T1tttt cross section upper limit 148
7.3.2 Projections for *Run 2* 150
7.4 Comparison to other *Run 1* and *Run 2* results **151**

In this chapter the results of the search for supersymmetric gluino pair production are presented. First, the data-driven background predictions in the signal regions are compared to the observed event yields in Sec. 7.1. These results are then interpreted in the framework of simplified SUSY models.

Section 7.2 describes the procedure used to test the data compatibility with different hypotheses, and to obtain observed and expected upper limits on the signal cross-sections. Thereafter, the limit setting is performed in Sec. 7.3.1 for the SUSY model T1tttt using the observed data yields. An outlook towards higher integrated luminosities is given. Finally, a comparison to other similar analyses performed during both runs of the LHC is provided in Sec. 7.4.

7.1 Results

Figure 7.1 shows the observed event yields in data for the different search bins, compared to the data-driven SM background expectation as discussed in Sec. 6.5. The prediction is represented by the stacked shaded histograms corresponding to the individual background contributions from simulation normalized to the predicted number of events.

Table 7.1 summarizes the values used for the background prediction in each bin, comparing them to the final observed yields. Based on Eq. 6.5, the data yields in the sideband SR and CR, mainband CR and corresponding estimates of the *QCD* background are given. The only simulation-based values are the *EWK* κ factors.

In the $n_{\text{jet}} \in [6, 8]$ search regions, where for several bins up to ten events are expected, the agreement between the data and expectation is satisfactory, with no systematic misprediction observed. For the bins with at least nine good jets, only a single event is observed, while in

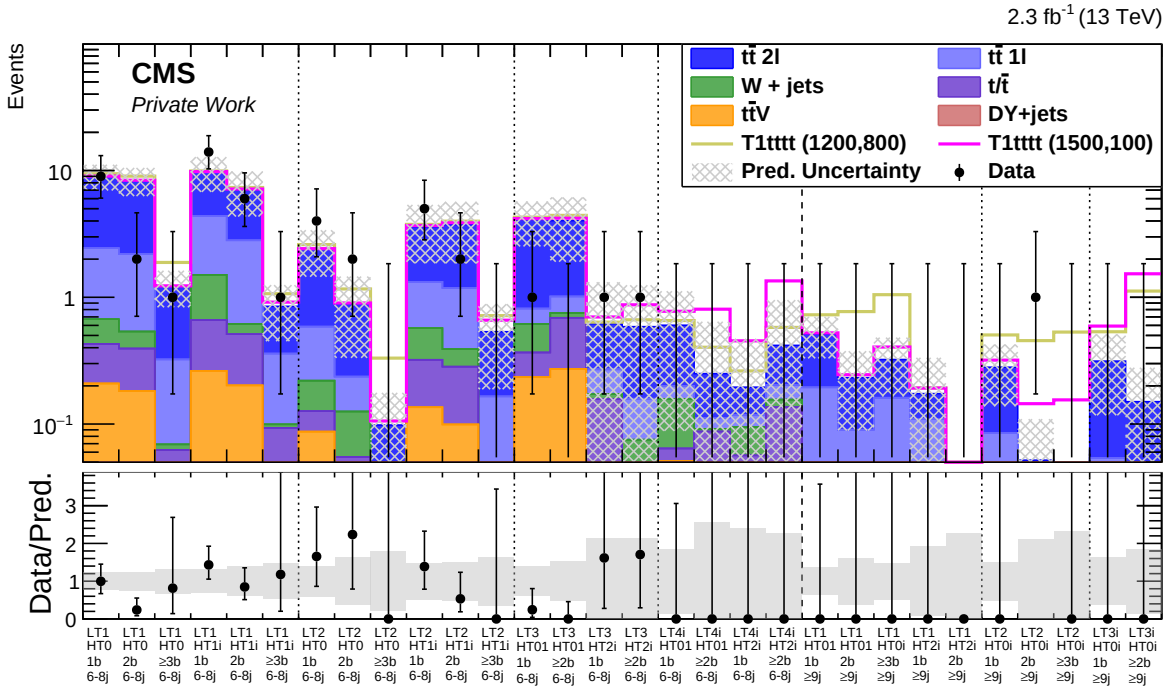


Fig. 7.1: Observed and predicted event yields in the 30 search regions. The black points show the number of events observed in data. The data-driven prediction of the SM background is represented by the filled histograms, preserving the individual contributions from simulation. The coloured lines illustrate the expectations for two benchmark scenarios of the T1tttt model, which are stacked on top of the data prediction. The lower panel shows the ratio of data to the prediction, where the error bars indicate the total statistical and systematic uncertainty on the ratio and the grey band shows the prediction uncertainty.

none of the bins more than one event is expected within the uncertainties. However, for the whole $n_{\text{jet}} \geq 9$ region about two events are predicted.

A comparison of the $\Delta\varphi$ distributions in data and simulation is shown in Figure 7.2 for the four L_{T} search ranges after the baseline selection with inclusive requirements on $n_{\text{jet}} \geq 6$, $n_{\text{b-tag}} \geq 1$ and $H_{\text{T}} > 500\,\text{GeV}$. While the signals become more prominent at higher L_{T} and $\Delta\varphi$ values, the background expectation from simulation agrees well with the observation. This agreement also verifies the reliability of the use of simulation to obtain the κ transfer factors.

Overall, no significant deviation from the SM expectation is observed. The compatibility of the data with the predicted yields is tested with statistical methods, which are described in the next section. Moreover, since no excess is found, the absence of signal is interpreted using the discussed SUSY signal samples, allowing to exclude particular mass scenarios of these simplified models.

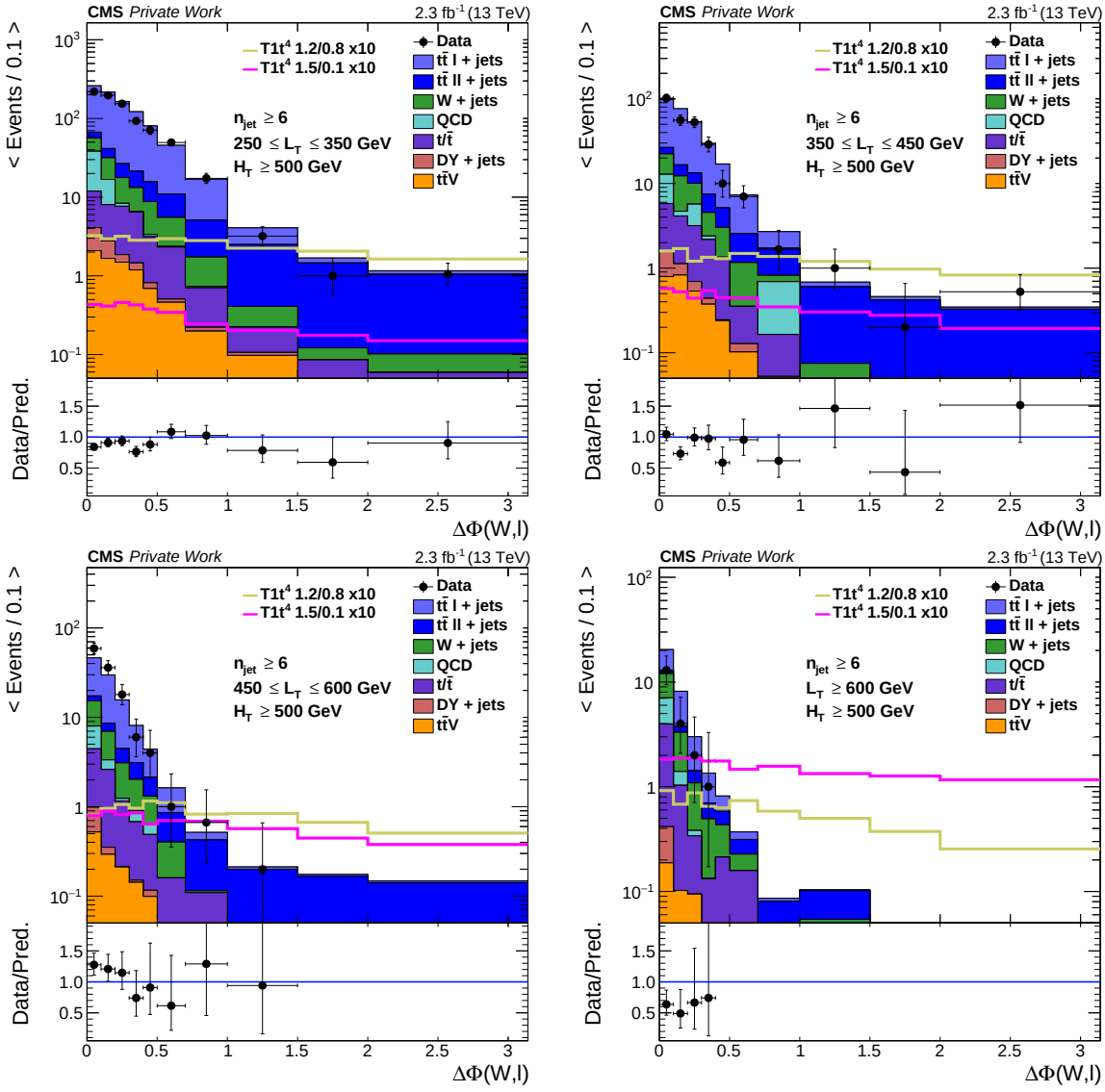


Fig. 7.2: $\Delta\varphi$ distributions in different L_T bins in data (black) and simulated samples after the baseline selection with $n_{\text{jet}} \geq 6$, $n_{\text{b-tag}} \geq 1$ and $H_T > 500\,\text{GeV}$. While the SM background samples are stacked on top of each other, the two signal benchmark points are scaled up by a factor of ten and drawn separately. The data-MC agreement is satisfactory throughout the $\Delta\varphi$ tails, and no significant excess is visible.

Table 7.1: Event yields as observed in data and predicted using the R_{CS} method: $N_{SR}^{pred.} = R_{CS}^{SB} \cdot \kappa \cdot (N_{CR}^{obs} - N_{CR}^{QCD\ estimate})$. The *QCD* contamination is estimated from the anti-selected data sample, and the κ factor is taken from simulation.

				Sideband $n_{jet} \in [4,5]$					Mainband $n_{jet} \geq 6$			
				CR		SR	data	Simulation	CR		SR	
n_{jet}	L_T [GeV]	H_T [GeV]	$n_{b\text{-tag}}$	data	*QCD*	data	R_{CS}^{EWK}	κ^{EWK}	data	*QCD*	Predicted	Obs.
[6, 8]	[250, 350]	[500, 750]	$= 1$	798	35.51 ± 9.27	26	0.04 ± 0.01	0.89 ± 0.03	299	6.55 ± 1.75	9.03 ± 2.07	9
			$= 2$	389	7.13 ± 6.38	22	0.06 ± 0.01	0.82 ± 0.03	180	1.45 ± 1.43	8.38 ± 2.10	2
			≥ 3	389	7.13 ± 6.38	22	0.06 ± 0.01	0.86 ± 0.06	25	0.35 ± 0.27	1.23 ± 0.39	1
		≥ 750	$= 1$	274	11.87 ± 3.62	13	0.05 ± 0.01	0.94 ± 0.03	212	5.01 ± 1.39	9.77 ± 2.99	14
			$= 2$	124	2.19 ± 1.87	8	0.07 ± 0.02	0.89 ± 0.04	124	1.23 ± 1.09	7.08 ± 2.74	6
			≥ 3	124	2.19 ± 1.87	8	0.07 ± 0.02	0.94 ± 0.06	14	0.24 ± 0.20	0.85 ± 0.40	1
	[350, 450]	[500, 750]	$= 1$	267	15.77 ± 4.15	8	0.03 ± 0.01	1.00 ± 0.06	77	1.57 ± 0.49	2.42 ± 0.96	4
			$= 2$	126	2.83 ± 2.80	3	0.02 ± 0.01	1.17 ± 0.11	32	0.29 ± 0.29	0.89 ± 0.56	2
			≥ 3	126	2.83 ± 2.80	3	0.02 ± 0.01	1.00 ± 0.12	4	0.00 ± 0.00	0.10 ± 0.08	0
		≥ 750	$= 1$	119	6.95 ± 1.87	5	0.05 ± 0.02	0.93 ± 0.05	89	2.66 ± 0.73	3.61 ± 1.75	5
			$= 2$	56	1.27 ± 1.16	5	0.09 ± 0.04	0.89 ± 0.05	47	0.59 ± 0.54	3.75 ± 1.90	2
			≥ 3	56	1.27 ± 1.16	5	0.09 ± 0.04	0.88 ± 0.09	7	0.11 ± 0.08	0.54 ± 0.35	0
	[450, 600]	[500, 1250]	$= 1$	169	12.07 ± 3.30	10	0.07 ± 0.02	1.09 ± 0.08	60	1.84 ± 0.54	4.11 ± 1.59	1
			≥ 2	59	2.19 ± 2.14	5	0.09 ± 0.04	1.09 ± 0.09	41	0.54 ± 0.51	4.02 ± 2.11	0
		≥ 1250	$= 1$	13	0.72 ± 0.34	1	0.08 ± 0.09	0.79 ± 0.14	10	0.42 ± 0.21	0.62 ± 0.69	1
			≥ 2	13	0.72 ± 0.34	1	0.08 ± 0.09	0.92 ± 0.13	8	0.09 ± 0.09	0.59 ± 0.66	1
	≥ 600	[500, 1250]	$= 1$	41	4.34 ± 1.37	2	0.06 ± 0.04	1.20 ± 0.16	9	0.60 ± 0.21	0.60 ± 0.51	0
			≥ 2	8	1.05 ± 1.04	1	0.15 ± 0.16	1.20 ± 0.25	2	0.21 ± 0.20	0.25 ± 0.39	0
		≥ 1250	$= 1$	10	0.82 ± 0.42	1	0.11 ± 0.12	0.81 ± 0.19	3	0.36 ± 0.19	0.20 ± 0.27	0
			≥ 2	10	0.82 ± 0.42	1	0.11 ± 0.12	1.11 ± 0.28	4	0.14 ± 0.13	0.42 ± 0.53	0
≥ 9	[250, 350]	[500, 1250]	$= 1$	1046	45.99 ± 12.27	37	0.04 ± 0.01	0.86 ± 0.06	16	0.26 ± 0.21	0.52 ± 0.19	0
			$= 2$	509	9.05 ± 8.05	30	0.06 ± 0.01	0.90 ± 0.09	5	0.04 ± 0.03	0.23 ± 0.14	0
		≥ 500	≥ 3	513	9.32 ± 8.25	30	0.06 ± 0.01	0.90 ± 0.09	6	0.00 ± 0.00	0.32 ± 0.16	0
		≥ 1250	$= 1$	30	1.66 ± 0.76	2	0.07 ± 0.05	0.65 ± 0.07	4	0.05 ± 0.04	0.17 ± 0.16	0
			$= 2$	30	1.66 ± 0.76	2	0.07 ± 0.05	0.80 ± 0.10	0	0.03 ± 0.02	0.24 ± 0.31	0
	[350, 450]	≥ 500	$= 1$	386	22.72 ± 5.89	13	0.04 ± 0.01	0.97 ± 0.09	8	0.10 ± 0.03	0.28 ± 0.14	0
			$= 2$	182	4.10 ± 3.95	8	0.05 ± 0.02	0.77 ± 0.08	2	0.01 ± 0.01	0.05 ± 0.06	1
			≥ 3	182	4.10 ± 3.95	8	0.05 ± 0.02	0.98 ± 0.18	1	0.00 ± 0.00	0.04 ± 0.05	0
	≥ 450	≥ 500	$= 1$	230	16.57 ± 4.46	11	0.05 ± 0.02	1.05 ± 0.12	6	0.15 ± 0.05	0.31 ± 0.20	0
			≥ 2	72	3.15 ± 3.07	7	0.10 ± 0.04	0.72 ± 0.08	2	0.05 ± 0.05	0.15 ± 0.13	0

7.2 Limit setting procedure

In this section the statistical framework used to derive limits on signal cross sections is presented. First, the general statistical procedure for testing signal hypotheses and obtaining exclusion limits is discussed, followed by the description of the implementation of these methods together with the data-driven background estimation. This method will be illustrated on a simplified example before being applied to the full analysis chain. Finally, upper limits on the cross section of the investigated SUSY models will be derived.

7.2.1 The CL_s method for statistical hypothesis testing

The approach described here is based on the CL_s method [167–169] and was refined by the LHC Higgs Combination Group consisting of members of both, the ATLAS and CMS collaboration, in an effort to prepare for the combination of the Higgs search results [170].

Following the general convention, the event yields for signal and background are denoted s and b, respectively. In the case of this analysis, these numbers correspond to the event yields in the signal regions of the different search bins.

Systematic uncertainties on the signal or background expectations are handled by so-called *nuisance parameters* θ, which become arguments for these expectations as: $s(\theta)$ and $b(\theta)$. All sources of uncertainties are assumed to be either independent (uncorrelated) or fully (100%) correlated, to allow for a factorization in the likelihood functions. If $\tilde{\theta}$ represents an estimate of the nuisance parameter set, the true value of θ can be estimated by probability density functions (*p.d.f.*) of the uncertainties, $\rho(\theta|\tilde{\theta})$. These *p.d.f.*s can be reinterpreted as posteriors $p(\tilde{\theta}|\theta)$ obtained from a measurement of $\tilde{\theta}$ based on Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) \propto p(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta) \,, \tag{7.1}$$

where $\pi_\theta(\theta)$ are prior *p.d.f.*s for the measurement. This reinterpretation allows the treatment of all systematic uncertainties in a fully frequentist approach.

Observed and expected limits

Limits are obtained for the ratio of tested signal cross section over the nominal theory value, which is called *signal strength modifier* μ. The level of incompatibility of a signal hypothesis with data is expressed as a C.L.. Within high-energy physics it is a convention to require a 95% C.L. for excluding a signal, i.e. when the 95% C.L. limit on μ reaches one, $\mu^{95\%\mathrm{CL}} = 1$.

In the general case, the probability of observing *data* in a measurement provided the parameters μ and θ defines the likelihood function:

$$L(\mathrm{data}|\mu,\theta) = \mathrm{Poisson}(\mathrm{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \,, \tag{7.2}$$

where $\mathrm{Poisson}(\mathrm{data}|\mu \cdot s(\theta) + b(\theta))$ stands for the Poisson probability corresponding to the data, and $p(\tilde{\theta}|\theta)$ is the aforementioned nuisance posterior.

A *test statistic* is introduced to discriminate between the the *signal+background* and *background-only* hypotheses. According to the Neyman-Pearson lemma [171], the most powerful discrimination is achieved by using the ratio of likelihoods. The test statistic used here is based on the

profile likelihood ratio [172] and can be written as[1]:

$$\tilde{q}_\mu = -2 \ln \frac{L(\text{data}|\mu, \hat{\theta}_\mu)}{L(\text{data}|\hat{\mu}, \hat{\theta})} \text{, while } 0 \leq \hat{\mu} \leq \mu \text{ ,} \tag{7.3}$$

where $\hat{\theta}_\mu$ is the maximum likelihood estimator for the *signal+background* hypothesis provided the signal strength μ and the observed data, while the parameter set of $\hat{\mu}$ and the nuisances $\hat{\theta}$ provide a global maximum of the likelihood. Unphysical negative signals are avoided by requiring a positive μ, while the upper constraint ensures a one-sided confidence interval.

After the test statistic is defined, it is possible to find the observed value of the test statistic $\tilde{q}_\mu^{\text{obs}}$ as a function of the μ value. Correspondingly, the nuisance parameters $\tilde{\theta}_\mu^{\text{obs}}$ and $\tilde{\theta}_0^{\text{obs}}$ that maximize the likelihood in Eq. 7.2, thus best describing the observed data, are found, for the *signal+background* and *background-only* hypotheses, respectively.

Next, the observed and expected limits for both hypotheses are calculated. For this matter, it is necessary to construct the corresponding test statistic *p.d.f.*s $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu)$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0)$. The methods of constructing these functions are discussed later on.

The probability for the observed value $\tilde{q}_\mu^{obs}$ to be as or less compatible with the *signal+background* (*background-only*) hypothesis, denoted as CL_{s+b} (CL_b), defines the corresponding *p-value*[2]:

$$\text{CL}_{s+b} = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|\text{signal+background}) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu) d\tilde{q}_\mu \text{ ,} \tag{7.4}$$

$$\text{CL}_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|\text{background}) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0) d\tilde{q}_\mu \text{ .} \tag{7.5}$$

The C.L. is then defined as the ratio of these two probabilities:

$$\text{CL}_s(\mu) = \frac{\text{CL}_{s+b}}{\text{CL}_b} \tag{7.6}$$

With this definition, a signal hypothesis would be excluded at $(1-\alpha)$ CL_s confidence level, if CL_s $(\mu = 1) \leq \alpha$. In general, the 95% C.L. upper limit on μ will be further denoted as $\mu^{95\%\text{CL}}$.

In addition to the observed limits, the same framework can be used to calculate possible limits on μ based on the *background-only* hypothesis. By definition, this expected limit has to be derived from a large set of background-only pseudo-data, for which the same procedure as for real data is applied. The cumulative distribution of the resulting $\mu^{95\%\text{CL}}$ values provides the 95% C.L. expected median upper limit by its 50% quantile, while the $\pm 1\sigma$ band is taken from the 16% and 84% quantiles.

Asymptotic approximation of the profile likelihood

The construction of the test statistic *p.d.f.*s $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu)$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0)$ generally requires the generation of toy MC pseudo-data samples with a signal strength of μ and $\mu = 0$, respectively. However, since the *signal+background* hypothesis test requires to scan a range of μ values, this results in a large number of different MC samples. Moreover, the calculation of expected limits also requires the generation of a large set of background-only samples. In the large

[1] Computationally, it is easier to minimize a logarithm, rather than to maximize a ratio. The factor two comes for convenience when determining the half minimum.

[2] In the classical frequentist approach, the signal is assumed to be excluded at 95% C.L., if $\text{CL}_{s+b} = 0.05$.

sample limit, the so-called asymptotic approximation allows to circumvent this computational challenge by using analytical formulas instead of toy MC.

In the asymptotic limit, according to the Wald approximation [173], the test statistic $\tilde{q}_\mu$ becomes equivalent to q_μ, which is the same but has no lower constraint on $\hat{\mu}$ [172]. Moreover, Wilks theorem [174] predicts, that in this regime q_μ is expected to have a half χ^2 distribution for one degree of freedom:

$$f(q_\mu|\mu') = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2} \tag{7.7}$$

One can then retrieve the observed upper limits on the *signal+background* hypothesis by setting $\mu' = \mu$, or for the *background-only* hypothesis with $\mu' = 0$, by using:

$$\mathrm{CL}_s = \frac{1-\Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}}-\sqrt{q_\mu})} \tag{7.8}$$

where $q_{\mu,A}$ is the test statistic of the representative *Asimov* data set with the observation set to the expected background and nominal nuisance parameters, and Φ is the cumulative distribution of the standard Gaussian. The quantile Φ^{-1} allows to compute the median upper and expected (N standard deviation) error bands limit for a $1-\alpha$ C.L. as:

$$\mu_{up+N} = \sigma(\Phi^{-1}(1-\alpha\Phi(N)) + N) \text{ , where } \sigma^2 = \mu^2/q_{\mu,A} \text{ .} \tag{7.9}$$

7.2.2 Statistical interpretation of results

Although the search discussed in this thesis is performed only in the dedicated signal regions of the search bins, the information from the control regions and sidebands also plays an important role in the statistical interpretation of the results. This information includes the amount of expected signal contamination, the estimated *QCD* contribution as well as the statistical errors of the event yields. In total, for each search category there are four bins, that have to be accounted for: the respective signal and control regions of the side- and mainbands.

Therefore, this search can be interpreted in a simplified way as a matrix (*ABCD*) method, where A defines the actual search region SR MB, while the auxiliary regions B: CR MB, C: SR SB and D: CR SB are required for the SM background estimation. As discussed in Chapter 6, the *QCD* contribution, estimated from data, has to be subtracted from the control regions B and D, and the residual differences in R_{CS} between the sideband and mainband have to be corrected for by the κ factor, obtained from simulation. This results in the following relation of the four *ABCD* regions:

$$A = \frac{C}{D - D_{QCD}} \times (B - B_{QCD}) \times \kappa\,, \tag{7.10}$$

which in fact repeats Eq. 6.5 for the background prediction described in Sec. 6.2.

The statistical treatment is implemented in the Higgs Combined Limit Tool of CMS [175, 176], which is based on the RooStat [177] and RooFit [162] packages and incorporates the recommendations from Ref. [170]. The additional functionality which allows to handle non-linear relations between different regions (e.g. an *ABCD* method) is explained in Ref. [178].

A likelihood function is constructed to incorporate the *ABCD* regions and the explicit relation between them as given in Eq. 7.10. This allows to account for the statistical Poisson

```
#Datacard to demonstrate the ABCD method with combine using the rateParam
#directive.
#name rateParam bin process init      --> floating multiplicative factor initialised to "init"
#name rateParam bin process expr args  --> scaling factor given by expression "expr" (function of "args")

imax 4  number of channels
jmax 1  number of processes -1
kmax *  number of nuisance parameters (sources of systematical uncertainties)
-------
bin               A    B    C    D
observation       0    2    7    72
-------
bin               A    B    C    D    A    B    C    D
process           bkg  bkg  bkg  bkg  sig  sig  sig  sig
process           1    1    1    1    0    0    0    0
rate              1    1    1    1    1.4  0.6  0.2  0.08
-------
SigSys    lnN     -    -    -    -    1.2  1.2  1.2  1.2
BkgSys    lnN     1.2  -    -    -    -    -    -    -

alpha     rateParam     A bkg (@0*@1/(@2-@3)*@4) beta,gamma,delta,QCDdelta,kappa
beta      rateParam     B bkg 2   [0,6]
gamma     rateParam     C bkg 7   [0,21]
delta     rateParam     D bkg 67  [0, 216]

QCDdelta param   5.2 5.2
kappa     param   0.716 0.077
```

Fig. 7.3: Example datacard for the *ABCD* method for one bin in the Limit Tool. For simplicity only two nuisance parameters representing systematic uncertainties are given.

uncertainty of the data in each of the regions separately. Moreover, the estimated *QCD* contribution in the control regions and the κ transfer factor are included as nuisance parameters. As a result, the likelihood function for each separate search bin can be written as:

$$L = \prod_{i}^{ABCD} \text{Poisson}(\text{data}_i|\mu \cdot s_i + b_i) \times \prod_{j}^{\text{nuisances}} \text{Constraints}(\theta_j, \hat{\theta}_j)\,, \tag{7.11}$$

where the four *ABCD* regions are included with Poisson *p.d.f.*s for the observed data given the predicted signal+background, and the term Constraints refers to the nuisance parameters.

Apart from the *QCD* estimates and κ factor, the nuisance parameters also include the effect of the statistical and systematic uncertainty sources discussed in Sec. 6.4 as constraints on the region A. These nuisances are represented by *p.d.f.*s following the log-normal distribution.[3] The systematic uncertainties on κ are treated as correlated between the different search bins. In addition to the accounting of signal contamination in the regions, the corresponding statistical and systematic uncertainties are included (cf. Sec. 6.4.3).

In practice, this whole information is provided to the Limit Tool using a so-called *datacard* text file. For simplicity, Fig. 7.3 shows an example datacard with only two nuisance parameters introduced for the signal and background systematic uncertainties, and only one *QCD* estimate in region D. A graphical representation of the likelihood model that is build as a workspace of RooFit in the Limit Tool with this datacard is shown in Figure 7.4.

The ultimate signal hypothesis testing requires the combination of the full datacards of all 30 search bins. The CL_s criterion is used to obtain the upper limits on the signal strengths.

[3]The log-normal *p.d.f.* is similar to the Gaussian, but can describe positively defined uncertainties, such as obtained for κ. In addition, this *p.d.f.* can be easily reinterpreted as shown in Eq. 7.1 with a flat prior $\pi_\theta(\theta)$.

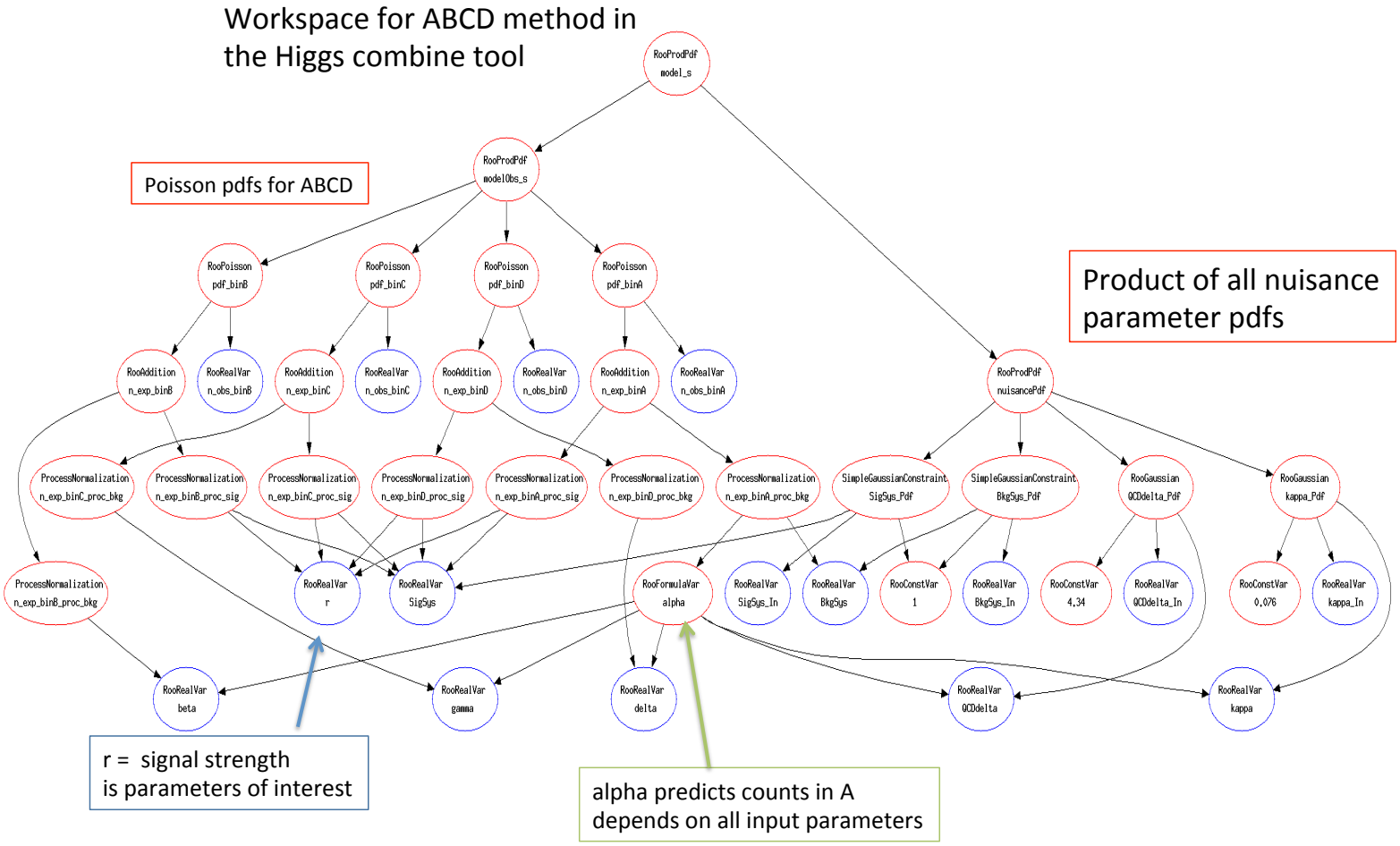


Fig. 7.4: Graphical representation of the likelihood implemented with an *ABCD* method for one bin in the Limit Tool [159].

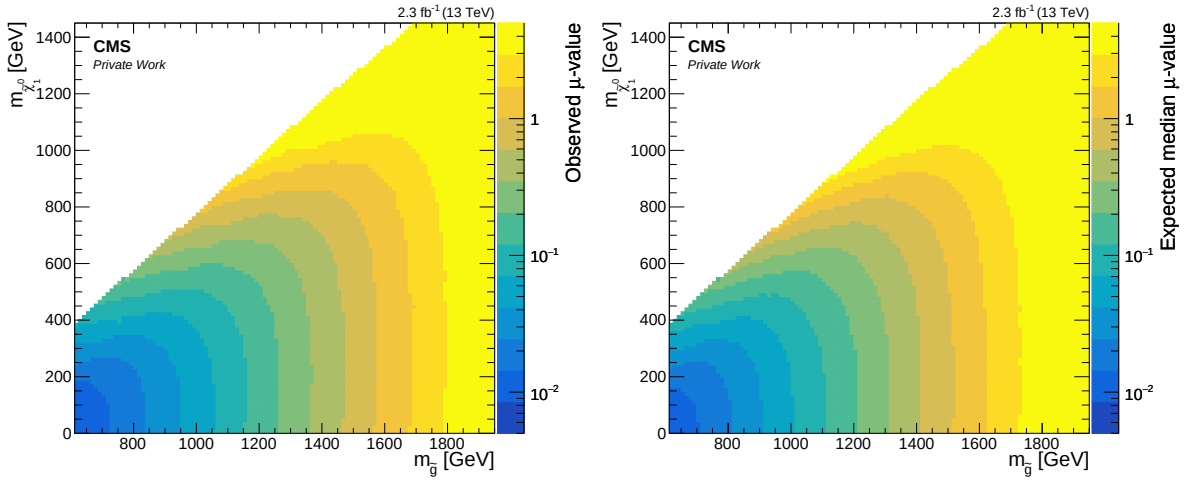


Fig. 7.5: 95% CL upper limits on the signal strength modifier μ for the T1tttt model as observed (*left*) and for the median expected (*right*). The values for the missing mass scenarios are obtained from interpolating the adjacent results.

7.3 Interpretation

As discussed previously, the results of this analysis – the absence of any significant deviation from the SM prediction – are interpreted within the framework of supersymmetric simplified models (cf. Sec. 1.4.1). In this section the upper limits on the cross section of the T1tttt model scan will be presented. The obtained exclusion limits are compared to other analysis performed within CMS, and also to similar searches by the ATLAS collaboration.

7.3.1 T1tttt cross section upper limit

For each mass scenario of the T1tttt signal sample the hypothesis testing is performed as presented in Sec. 7.2. Using the CL_s criterion, 95% C.L. upper limits on the observed and median expected signal strength are obtained. Figure 7.5 shows the corresponding results for the whole mass scan of the T1tttt model. To compensate for the varying granularity of the mass scenarios used in the simulation (cf. Fig. 5.6), the missing points are obtained from interpolating between adjacent μ values[4].

As discussed before, a signal hypothesis is excluded, if the signal strength corresponding to the 95% CL_s drops below one. Therefore, the contour corresponding to $\mu = 1$ defines the exclusion limit. In the case of the observed limit this results in a single contour, while for the expected limit the median exclusion is complemented with the respective statistical $\pm 1\sigma$ error band, as explained previously in Eq. 7.9.

The upper limit on the production cross section of the T1tttt model can be derived straightforward by multiplying the excluded signal strength with the theoretical predicted cross section. However, the uncertainty on this prediction has also to be accounted for when quoting the observed limit. Therefore, additional exclusion contours are obtained for observed μ limits evaluated at the $\pm 1\sigma$ values of the signal cross sections.

[4] Using the Delaunay triangles interpolation [179] with a 25 GeV mass step.

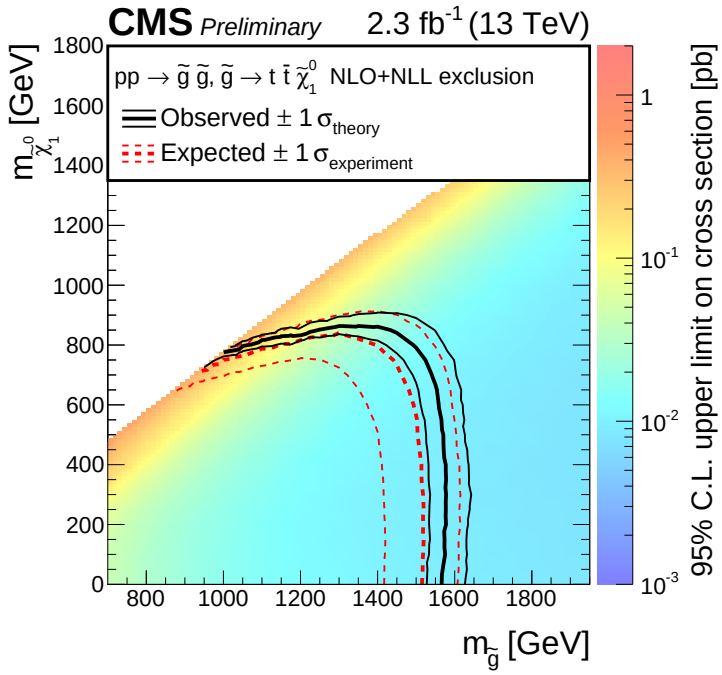


Fig. 7.6: 95% C.L. cross section limits for the T1tttt model as obtained using the CL_s criterion. The observed contour for the nominal theory cross section is shown in black, complemented with the exclusions originating from the corresponding theory uncertainty. In red (dashed) the expected exclusion contours are shown for the median and $\pm 1\sigma$ error bands corresponding respectively to the 50%, 16% and 84% quantiles of the $\mu^{95\%\mathrm{CL}}$ distribution.

Figure 7.6 shows the full set of limits for the T1tttt model. The observed upper limit on the expected (theoretically predicted) cross section is shown by the histogram, while the corresponding exclusion contour is given by the black thick line, and the $\pm 1\sigma$ contours due to the theoretical uncertainty are shown by the thinner lines. The expected exclusion limits for the median and $\pm 1\sigma$ error bands are shown by the dashed red thick and thin lines, respectively. In general, the observed limit follows closely the expected median. The systematically larger observed exclusion can be explained by the deficit in observed data relative to the prediction in several bins (cf. Fig. 7.1).

In the non-compressed region, where $m_{\tilde{g}} - m_{\tilde{\chi}_1^0} \gg 2m_t$, the limit is almost independent of the neutralino mass, resulting in the exclusion of signals with $m_{\tilde{g}}$ up to 1575 GeV. The typical 95% C.L. upper limit on the cross section in this phase-space is about 10 fb. In this region the gluino and subsequent top decay products have high energies leading to large hadronic and leptonic scales, which provides the largest signal sensitivity.

Towards the more compressed mass scenarios the behaviour of the limit contour changes to the opposite, becoming almost independent of the gluino mass and excluding signals with $m_{\tilde{\chi}_1^0}$ up to 850 GeV. The highest upper limit for these scenarios is of the order of 1 pb in the region, where the gluinos decay to two off-shell top quarks, leading to very soft decay products and decreased acceptance.

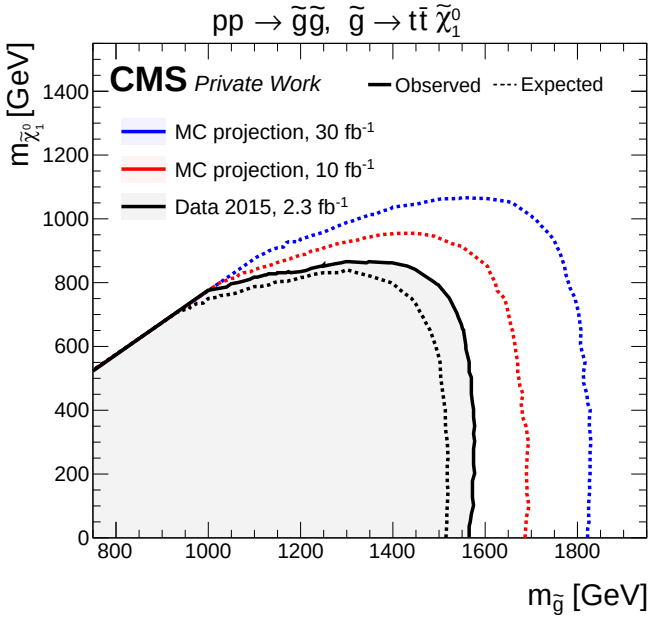


Fig. 7.7: T1tttt model exclusion limit for the current analysis with $\mathcal{L} = 2.3\,\text{fb}^{-1}$ data (black) compared to the expected limits with $10\,\text{fb}^{-1}$ (red) and $30\,\text{fb}^{-1}$ (blue) based on simulation.

7.3.2 Projections for *Run 2*

As stated previously in Sec. 7.1, one of the challenges for this analysis are the large statistical uncertainties of the observed event yields in the four *ABCD* regions, especially in the sidebands. In this case the systematic uncertainties of the prediction become less important. Therefore, it is of high interest to evaluate the expected sensitivity of this analysis for larger integrated luminosities.

Figure 7.6 shows to the observed and expected limits as obtained from the current analysis for $\mathcal{L} = 2.3\,\text{fb}^{-1}$, compared to the projections of the expected exclusion limit using the simulated background and signal samples scaled to 10 and $30\,\text{fb}^{-1}$, respectively. The scenario of $30\,\text{fb}^{-1}$ corresponds to the expectation for the total integrated luminosity delivered to CMS in the whole year of 2016 [180].

In the case of the projections, the simulated background is taken instead of the data observation, and the whole prediction chain is executed as would be for real data[5]. For the intermediate scenario with $\mathcal{L} = 10\,\text{fb}^{-1}$, which is four times that of the current data, the increase in exclusion reach for the gluino and neutralino mass is expected to be about 100 GeV. The triple amount of data with $\mathcal{L} = 30\,\text{fb}^{-1}$ would further push the exlusion limits by a similar increment up to 1825 GeV for $m_{\tilde{g}}$ and 1075 GeV for $m_{\tilde{\chi}_1^0}$.

This way, the fast increase in available data will not be paid off by a similar expansion of the exclusion limit.The reason for this is the logarithmic decrease of the gluino production cross section with its mass. Nevertheless, the improvement of the theoretical cross section precision and reduction of the systematic uncertainties will provide stronger results in the future.

[5]The simulated pseudo-data is taken as Poissonian.

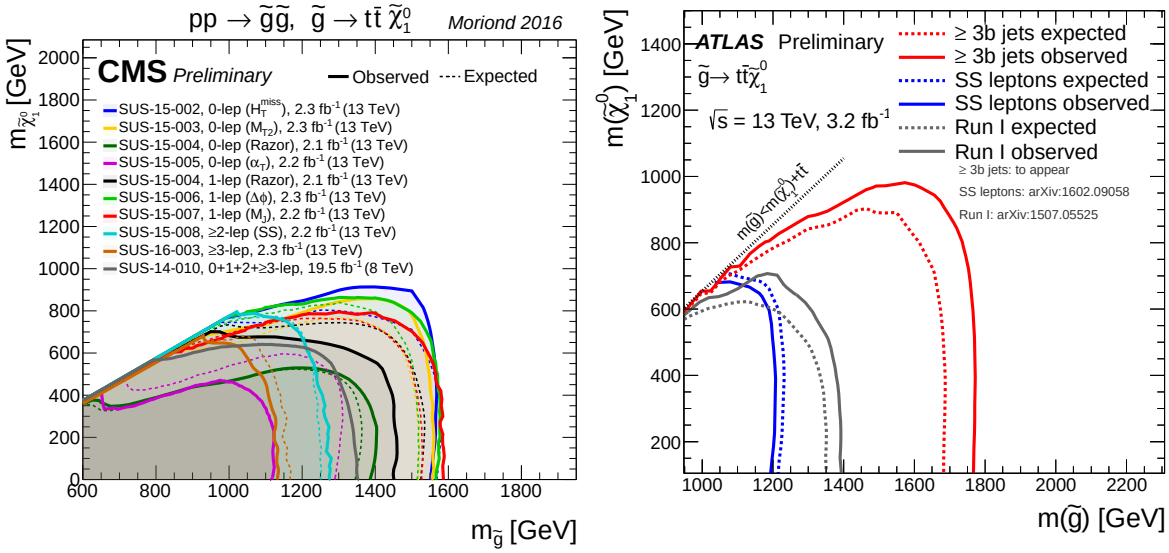


Fig. 7.8: Exclusions limits at 95% C.L. on the simplified SUSY model of gluino pair production with subsequent decay $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ obtained by the CMS [181] (*left*) and ATLAS [182] (*right*) collaborations. The *Run 1* results demonstrate the respective strongest exclusion achieved. *SUS-15-006* corresponds to results obtained by the present analysis.

7.4 Comparison to other *Run 1* and *Run 2* results

As discussed in Sec. 1.4.2, the T1tttt simplified SUSY model was already studied at CMS during the first run of the LHC. Similar to these analyses, the search was performed in several exclusive final states using different discriminating variables. Naturally, many of those searches were repeated at the new centre-of-mass energy.

The left plot of Figure 7.8 shows the obtained exclusion limits on the T1tttt model by the new *Run 2* analyses of CMS compared to the strongest search from *Run 1* (in grey), which incorporates various final states [45]. The limits presented in this thesis correspond to the green light line labelled *SUS-15-006* ($\Delta\varphi$) and are made public in Ref. [158].

The analysis with the largest "covered" exclusion area (*SUS-15-002*) is based on a search in the full-hadronic final state, i.e. where all four W bosons decay hadronically, using the variables H_T and missing H_T [183].[6] An analysis similar to the present one in terms of the single lepton final state is *SUS-15-007* [184], where the search is performed using the sum of the masses of large-radius jets and requiring large missing energy.

Overall, the present analysis is very competitive to these other CMS searches, excluding a similar phase-space of the T1tttt model. Especially close to the diagonal the sensitivity becomes one of the strongest due to the absence of an explicit requirement on $\not{E}_T$. At low neutralino masses the full-hadronic searches suffer from the indistinguishable (irreducible) *QCD* background. This analysis extends the CMS *Run 1* limits in terms of gluino and neutralino mass exclusion by additional 200 GeV and 250 GeV, respectively.

[6] The missing H_T is computed similar to $\not{E}_T$, but using only the transverse momentum of jets.

The ATLAS collaboration has also preformed several searches for direct gluino pair production in a simplified model similar to the T1tttt of CMS. However, in this case only the decays of gluinos to two on-shell top quarks are considered. The right plot of Figure 7.8 presents the results obtained by the two *Run 2* analyses of ATLAS [185,186] in comparison to the respective *Run 1* result [46].

The strongest exclusion is obtained from the analysis requiring large missing transverse energy, at least three b-tagged jets and either none or one single isolated lepton [186]. This combination of the full-hadronic and single-leptonic final state searches allows to increase the signal sensitivity. Moreover, large-radius jets are used to identify highly boosted top quarks similarly as in the *SUS-15-007* [184] analysis by CMS.

Due to operational reasons, the total integrated luminosity of the ATLAS and CMS experiments in the year of 2015 differ significantly, being respectively $\mathcal{L} = 3.2\,\text{fb}^{-1}$ and $2.3\,\text{fb}^{-1}$ providing ATLAS 50% more data than CMS. Together with the final state combination, this results in an increased sensitivity of the ATLAS search relative to the CMS results.

For low neutralino masses the ATLAS results exclude gluinos below 1760 GeV, while neutralinos are excluded up to 975 GeV for $m_{\tilde{g}} \approx 1600\,\text{GeV}$. However, since ATLAS does only consider on-shell top quarks for their searches, the present analysis extends the exclusion further into the more compressed mass scenarios along the diagonal.

Conclusions

"I know you're living in my mind
It's not the same as being alive"

— *Supersymmetry*
Arcade Fire

Symmetries are seldom exact – especially in nature. The highly successful Standard Model of particle physics (SM) is not only a theory of symmetries, but also of broken symmetries. Without the Higgs boson that causes electroweak symmetry breaking none of the particles would have masses, rendering the Universe completely different from what we are used to. Now the SM seems to be finalized after decades of theoretical and experimental work, but some questions do still remain. What is Dark Matter, how can gravity be included, and why did Nature choose the vacuum it has?

A new symmetry between fermionic and bosonic states, called Supersymmetry, would provide appealing answers to some of these fundamental questions. If SUSY is not a broken symmetry, it should have already been observed a long time ago, since the supersymmetric particles would have the same masses as their SM partners. Therefore, SUSY has to be searched for at higher energy scales.

The Large Hadron Collider (LHC) is an ultimate tool to not only further test the validity of the SM, but also to search for new physics phenomena at the multi-TeV scale. A plethora of searches for different flavours of SUSY theories has been performed during the first run of the LHC at $\sqrt{s} = 7$-$8\,\text{TeV}$. None of them yielded positive results, further constraining SUSY.

Prior to data taking at a new centre-of-mass energy, refurbishment works have been performed at the LHC complex as well as on the side of the experiments in order to increase their performance and efficiency. In this thesis the particular upgrade of the CMS outer hadron calorimeter (HO) readout electronics was presented. Silicon photomultipliers have proven to be superior to hybrid photodiodes previously used in the HO, and therefore installed into CMS with a drop-in replacement during the first long shutdown of the LHC.

After the installation, the new hardware was commissioned in several steps to verify its performance. The first energy calibration of the silicon photomultipliers was performed using cosmic muons recorded in local HO and global CMS runs. The observed high muon detection efficiency allows to extend the use of the HO detector beyond calorimetry, e.g. for the purpose of muon triggering or identification. These results were presented at the CALOR2014 conference [110], and documented in Ref. [113].

With the second run of the LHC at a new centre-of-mass energy of 13 TeV the sensitivity to production of supersymmetric gluon partners, gluinos, has increased significantly, while for most SM backgrounds the increase was moderate. In the case where each of the gluinos decays to a $\mathrm{t\bar{t}}$ pair and the lightest supersymmetric particle, the neutralino $\widetilde{\chi}_1^0$, the final state signature can be truly spectacular, featuring multiple high-energy jets out of which several originate from b-quarks. The presence of four top-quarks results in a 40% possibility for a charged lepton (e,μ or τ) to appear in the final state.

This thesis has presented a search for gluino pair production in final states with several high-energy jets, b-tagged jets and an isolated lepton, using data recorded by the CMS experiment in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. Even though the analysed data corresponds to an integrated luminosity of $2.3\,\text{fb}^{-1}$, which is ten times less than during *Run 1*, this analysis was able to outperform previous searches of similar models conducted at $\sqrt{s} = 7$ and 8 TeV.

The sensitivity to different signal mass scenarios was enhanced by using 30 exclusive search categories based on the jet and b-tag multiplicities, as well as the leptonic and hadronic scales, L_T and H_T, respectively. Signal-enriched regions were created using the $\Delta\varphi$ variable – the azimuthal angle between the reconstructed lepton and W boson candidates, represented by the transverse vectorial sum of the missing energy and lepton momentum.

The resulting Standard Model background is dominated by $\text{t}\bar{\text{t}}$+jets and W+jets processes. A data-driven approach was used to predict the background contamination in the signal regions. Here, data sidebands with lower jet multiplicities n_jet were used to measure the respective signal-to-control region transfer factors R_CS. Simulation was used to estimate residual differences in the R_CS values between the n_jet categories.

Multijet events from *QCD* tend to be selected in the control regions due to misidentified leptons. Therefore, the *QCD* contamination was estimated from an orthogonal sample with looser lepton selection criteria, and thus higher misidentification probability. This probability was measured in a *QCD*-enriched sideband without b-tagged jets and further used to predict the multijet event contribution to the control regions.

The largest uncertainty on the background prediction stems from the difference in the n_jet distributions observed in dileptonic and semileptonic events, resulting in a total systematic uncertainty in the range from 10 to 30%. However, the statistical uncertainties on the R_CS factors dominate in a large number of search bins, resulting from the low event yields in the sideband regions in data.

The resulting event yield predictions of the background were found to be in good agreement with the observation, indicating no deviation from the SM. These results were interpreted within the framework of simplified models, which comprise only very basic SUSY spectra, facilitating the re-interpretation of the results in more detailed theories. For the T1tttt model investigated in this thesis upper limits on the production cross sections were obtained for certain gluino-neutralino mass scenarios at 95% C.L. using a frequentist approach.

In case of light neutralinos, the sensitivity of this analysis is similar to other 13 TeV searches and results in gluinos excluded below 1575 GeV. When the gluino-neutralino mass gap becomes small, the present analysis profits from the absence of a strict missing energy requirement and is able to exclude models with neutralinos lighter than 850 GeV. These results extend the *Run 1* limits by about 200 and 250 GeV, respectively, and were presented at the MORIOND2016 conferences accompanied by Ref. [158].

Outlook

After the first few inverse femtobarns of 13 TeV data the SM still seems to hold and no hints of SUSY have been found. SUSY in its most simple implementation – the MSSM – becomes more and more restricted. But SUSY is "difficult to kill" [187], even in its more constrained form, given the variety of possibilities this symmetry between bosons and fermions offers.

The journey of particle physics has not yet come to an end. The unanswered questions and demands for a more universal theory remain, while the exact properties of the discovered Higgs boson are yet to be measured.

Nevertheless, during the last decades the progress in theoretical physics methods allowed more for theory-guided research, forcing the experimental community to catch up with testing theory predictions. The SM is a very bright example for this trend, but also the discovery of gravitational waves [188] can be considered as a long anticipated confirmation of a fundamental theory.

It is possible, that particle physics has arrived at a point, where only an unexpected discovery will help to guide the development of theories. Whether the recent excesses in the diphoton spectrum [189, 190] will result in such a change of theorists' doctrines will only become clear with the further operation of the LHC.

Appendix A

Backup trigger strategy

This Appendix reviews the backup trigger strategy discussed in Sec. 5.2.2.

Given the new beam conditions of the LHC at the beginning of *Run 2*, it was not determined, whether the main HLT trigger paths used in this analysis – `HLT_Lep15_IsoVVVL_PFHT350` – would remain prescaled after the initial runs in 2015. The other existing lepton triggers, which were not prescaled, either had to strong isolation requirements, or also thresholds on the missing transverse energy. Since this analysis relies on the L_T variable, an alternative trigger strategy was developed, which would allow for an efficient selection of events in the intended search region.

For both kinds of leptons a combination of two trigger paths could be used:

- `HLT_Lep15_IsoVVVL_PFHT350_PFMET50`, where `Lep` referres to either an electron or muon, is similar to the final analysis trigger `HLT_Lep15_IsoVVVL_PFHT350` except for an additional requirement of $\not{E}_T$ above 50 GeV.
- `HLT_Mu50` and `Ele105_CaloIdVT_GsfTrkIdT` select muons and electrons with p_T above 50 and 105 GeV, respectively. No isolation is required in this case, thus these paths are referred to as "non-isolated lepton" triggers.

This combinations allows to have a high efficiency for large lepton momenta independent of $\not{E}_T$ using one path and vice versa. Since L_T is calculated as the scalar sum of the lepton p_T and $\not{E}_T$, the combination of these two different kinds of trigger paths would allow to have an efficient selection above an L_T corresponding to the sum of the respective turn-on thresholds.

First, the individual trigger efficiencies of these paths are measured to identify the thresholds. Figure A.1 shows the efficiencies of the non-isolated lepton triggers in comparison to the analysis triggers[1] as a function of the lepton p_T. Overall, the turn-on shapes agree with the online thresholds, resulting in a turn-on of about 60 GeV for `HLT_Mu50`, and 130 GeV for `Ele105`.

In comparison to the analysis triggers, the electron path shows a similar plateau at about 95%, while the non-isolated muon path is about 3% lower than the isolated. One reason for such a discrepancy is the L1 Trigger seed, which in the latter case is based on the fully efficient H_T path, and for the former a muon seed is used, which is known to have a lower efficiency.

Figure A.2 shows the efficiencies of the backup paths `HLT_Lep15_IsoVVVL_PFHT350_PFMET50` versus the missing transverse energy. Despite having high $\not{E}_T$ turn-on thresholds, both lepton paths achieve full efficiency in the plateau.

[1] The efficiencies are measured in a similar way to Sec. 5.2.2 using the full 2.3 fb^{-1} dataset.

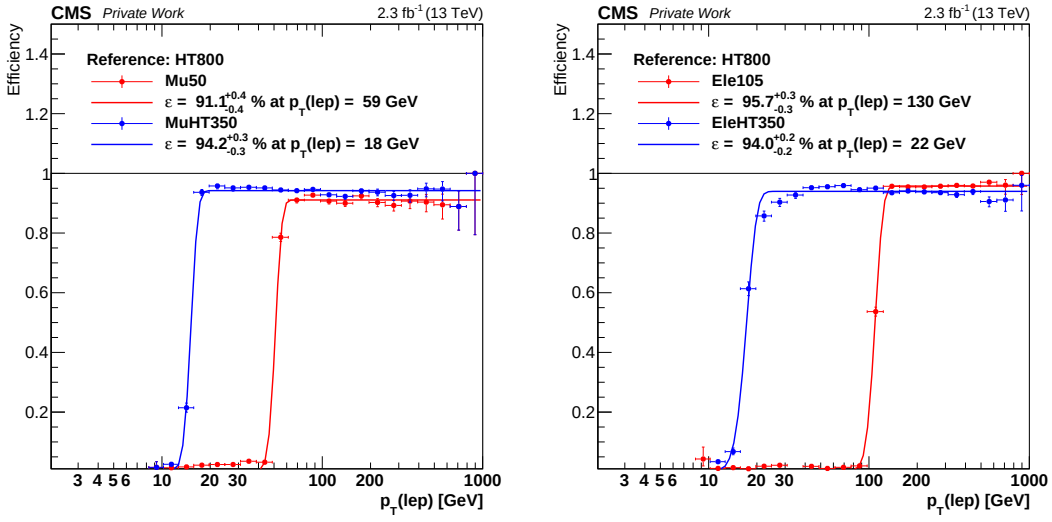


Fig. A.1: Efficiency of the lepton p_{T} leg for the main and backup analysis HLT paths with isolated and non-isolated leptons, respectively. The reference selection requires the events to pass a pure H_{T} trigger `HLT_PFHT800` and contain a single lepton with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$.

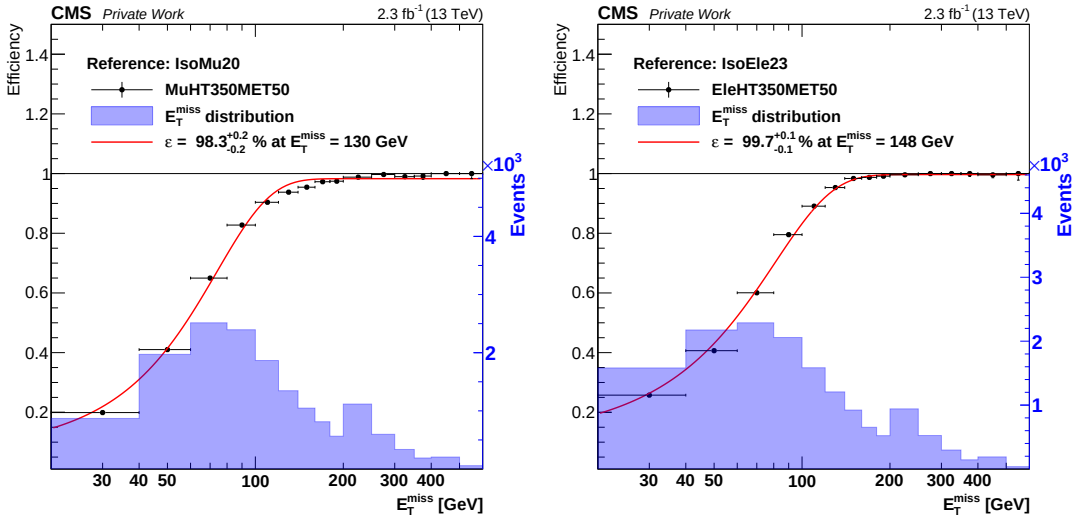


Fig. A.2: Efficiency of the $\not{E}_{\mathrm{T}}$ leg for the backup HLT paths including an isolated muon (*left*) or electon (*right*). The reference selection requires the events to pass an isolated lepton trigger, contain a single lepton with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and satisfy $H_{\mathrm{T}} > 500\,\mathrm{GeV}$.

A two-dimensional representation of the efficiency of the combined backup trigger requirement is shown in Fig. A.3 in the lepton p_{T}-$\not{E}_{\mathrm{T}}$ plane. The red regions represent the phase space with highest efficiency, while their horizontal and vertical borders correspond to the turn-on thresholds of the individual trigger paths. The baseline selection threshold of $L_{\mathrm{T}} = 250\,\mathrm{GeV}$ is represented by the black dashed lines. For both lepton types the search region above the lines

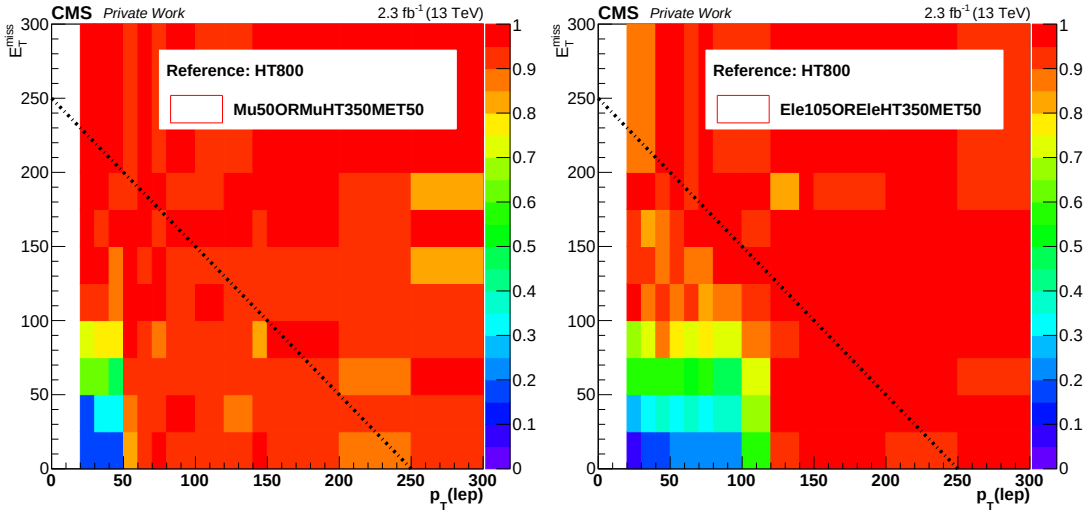


Fig. A.3: Efficiency of the combined isolated and non-isolated lepton HLT trigger paths for muons (*left*) and electrons (*right*). The black dashed lines indicate the analysis selection threshold $L_\mathrm{T} = 250\,\mathrm{GeV}$. The reference selection requires the events to pass a pure H_T trigger `HLT_PFHT800` and contain a single lepton with $p_\mathrm{T} > 10\,\mathrm{GeV}$.

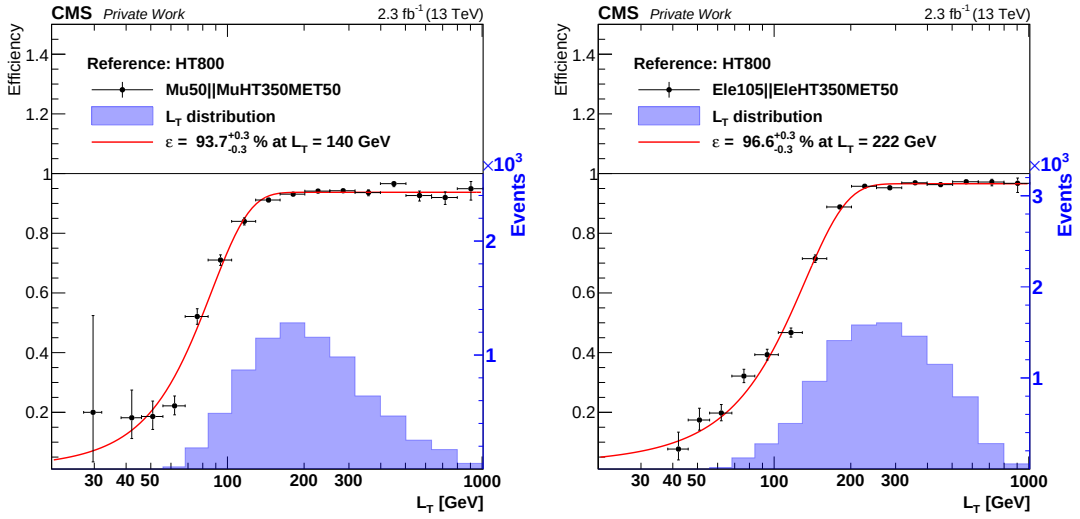


Fig. A.4: Selection efficiency versus the L_T variable for the combined backup trigger strategy with muons (*left*) and electrons (*right*). The reference selection requires the events to pass a pure H_T trigger `HLT_PFHT800` and contain a single lepton with $p_\mathrm{T} > 10\,\mathrm{GeV}$. The number of events passing the triggers is shown by the blue histograms.

is at the maximum efficiency.

The selection efficiency w.r.t. the single variable L_T can be seen in Fig. A.4, also confirming the sufficient efficiency of this trigger strategy. The plateau is reached for muon (electron) p_T above 140 GeV (222 GeV).

Appendix B

Lepton isolation

The isolation of leptons is an important criteria for their identification as discussed in Sec. 4.3.2. During the LHC *Run 1* with $\sqrt{s} = 8\,\text{TeV}$ the hadronic activity in events was lower compared to with $\sqrt{s} = 13\,\text{TeV}$. This justified the usage of fixed-cone isolation variables, like the relative isolation I_{rel03} with $R = 0.3$. In this section the justification for the use of variable-cone isolation, the mini-isolation I_{mini} will be provided.

The main criteria for an identification variable is the efficiency with which it allows to select objects of interest, and the rejection power of unwanted objects. As this analysis targets prompt electrons or leptons from W-boson decays, the ID has to efficiently select these *prompt* leptons, and reject others.

In simulated events lepton candidates can be classified as follows:

Prompt Leptons directly from the hard scattering, e.g. from W-boson decays

Non-prompt Real electrons or muons from other processes such as b-jets or hadronization

Fake Lepton candidates from misidentified jets or photons

Unmatched In case the origin of the lepton candidate cannot be identified

For this study simulated samples for the signal benchmark T1tttt (1500,100) and dominant $\bar{\text{t}}\text{t}$+jets background are used. The efficiency is defined as:

$$\epsilon_{\text{Iso}} = \frac{N(\text{passing ID and isolation requirements})}{N(\text{passing only ID requirements})} \tag{B.1}$$

The plots in Figs. B.1 and B.3 show the respective isolation efficiencies separately for the different lepton candidate categories. As can be seen for both sample types, the efficiency of prompt leptons with fixed-cone relative isolation I_{rel03} decreases with growing lepton p_{T} or H_{T}, since the hadronic activity surrounding the lepton also rises. In case of mini-isolation the respective efficiency is in general much higher and also stable at large p_{T} and H_{T}, while for non-prompt leptons the efficiency does not change with respect to I_{rel03}.

The plots in Figs. B.2 and B.4 show the efficiencies obtained with different isolation thresholds for both types. In general, mini-isolation performs similar or better than the fixed-cone isolation. A summary of these plots for a combined background lepton category is shown in Fig. B.5.

Overall, the variable-cone mini-isolation provides a higher prompt lepton efficiency even at growing lepton p_{T} and H_{T}, while showing a similar background rejection power as the standard fixed-cone isolation.

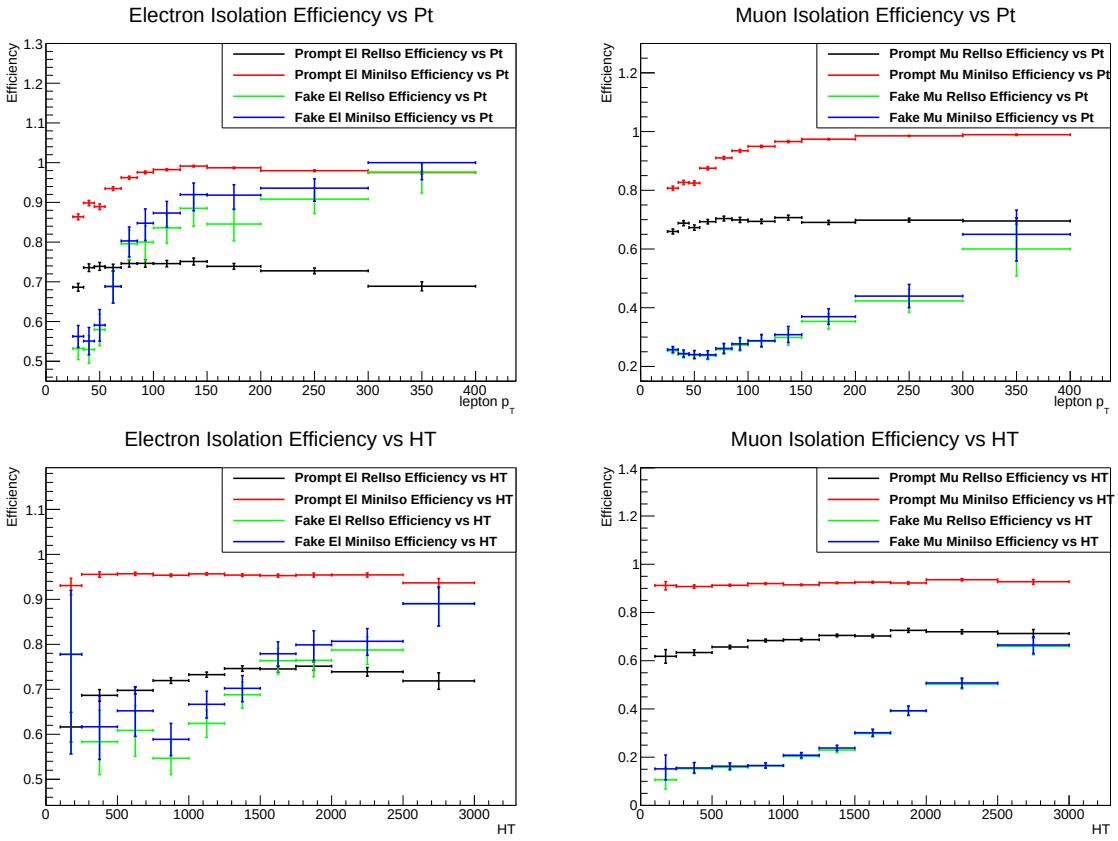


Fig. B.1: Isolation efficiency for T1tttt (1500,100) events with upper requirements of $I < 0.1$ for both isolation types. Mini-isolation improves the efficiency of prompt leptons in general w.r.t. the fixed-cone isolation, and also remains efficient at high p_{T} and H_{T}.

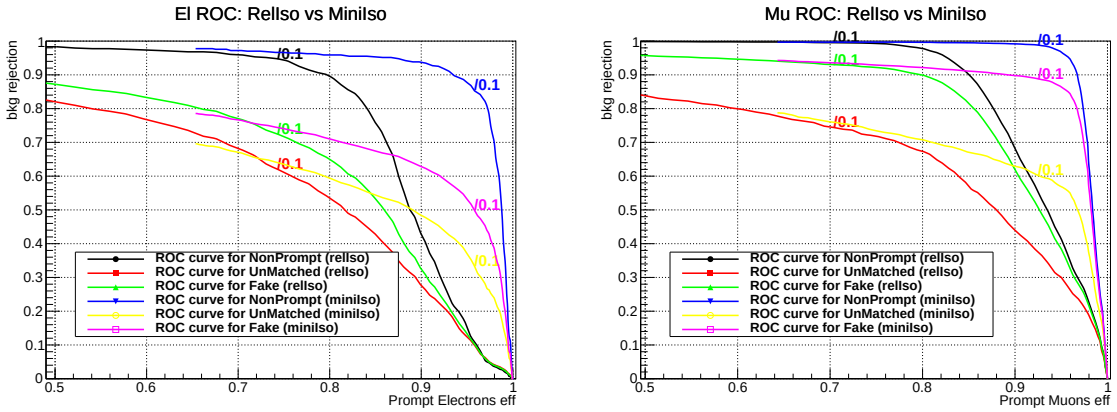


Fig. B.2: Comparison of isolation efficiencies for prompt electrons (*left*) and muons (*right*) versus the rejection power of other leptons for simulated T1tttt (1500,100) events. The points corresponding to the $I < 0.1$ requirement are indicated.

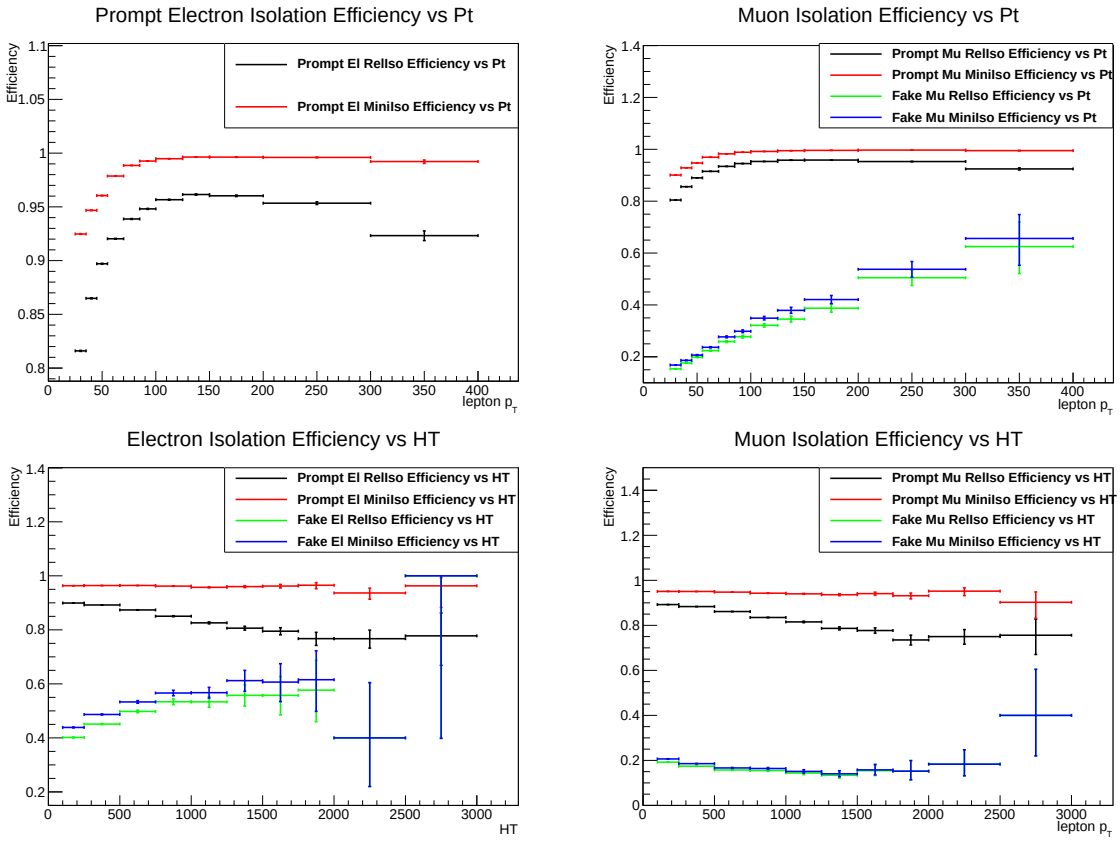


Fig. B.3: Isolation efficiency for $\mathrm{t\bar{t}}$+jets events with upper requirements of $I < 0.1$ for both isolation types. Mini-isolation improves the efficiency of prompt leptons in general w.r.t. the fixed-cone isolation, and also remains efficient at high p_{T} and H_{T}.

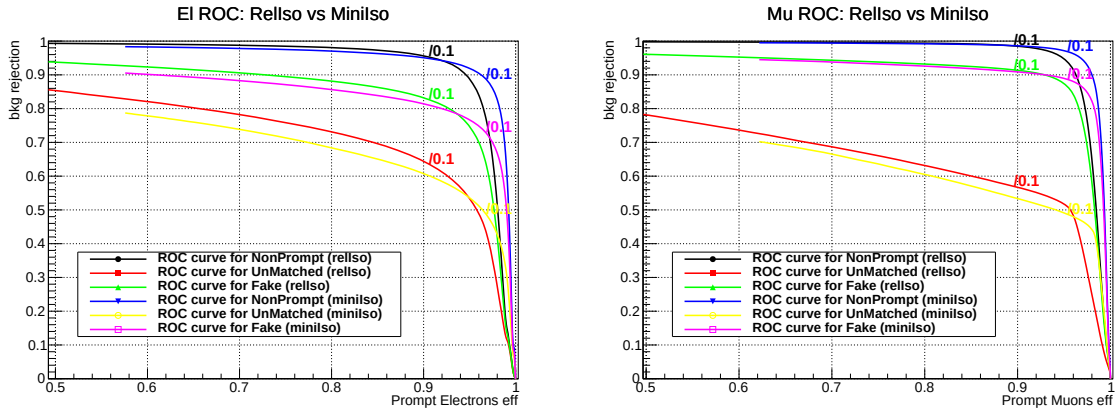


Fig. B.4: Comparison of isolation efficiencies for prompt electrons (*left*) and muons (*right*) versus the rejection power of other leptons for simulated $\mathrm{t\bar{t}}$+jets events. The points corresponding to the $I < 0.1$ requirement are indicated.

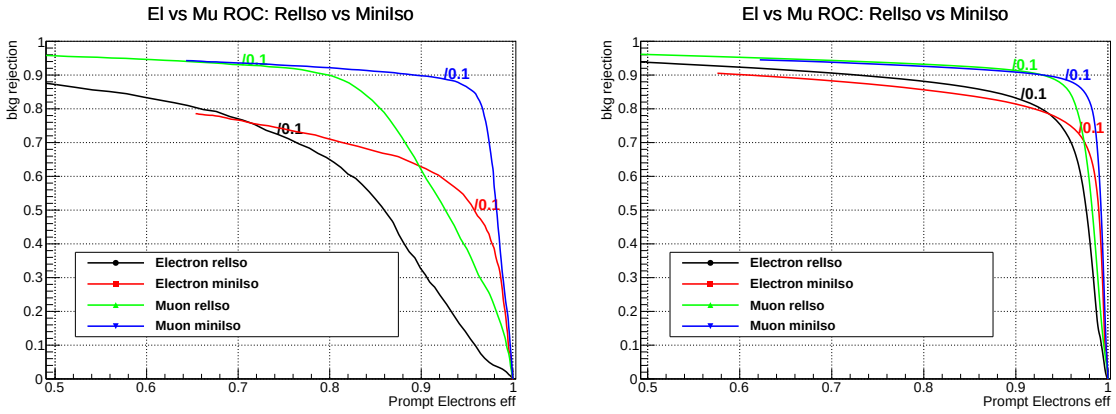


Fig. B.5: Comparison of isolation efficiencies for prompt leptons versus the rejection power of other leptons for simulated T1tttt (1500,100) events (*left*) and $t\bar{t}$+jets background (*right*).

Appendix C

Additional control plots after baseline selection

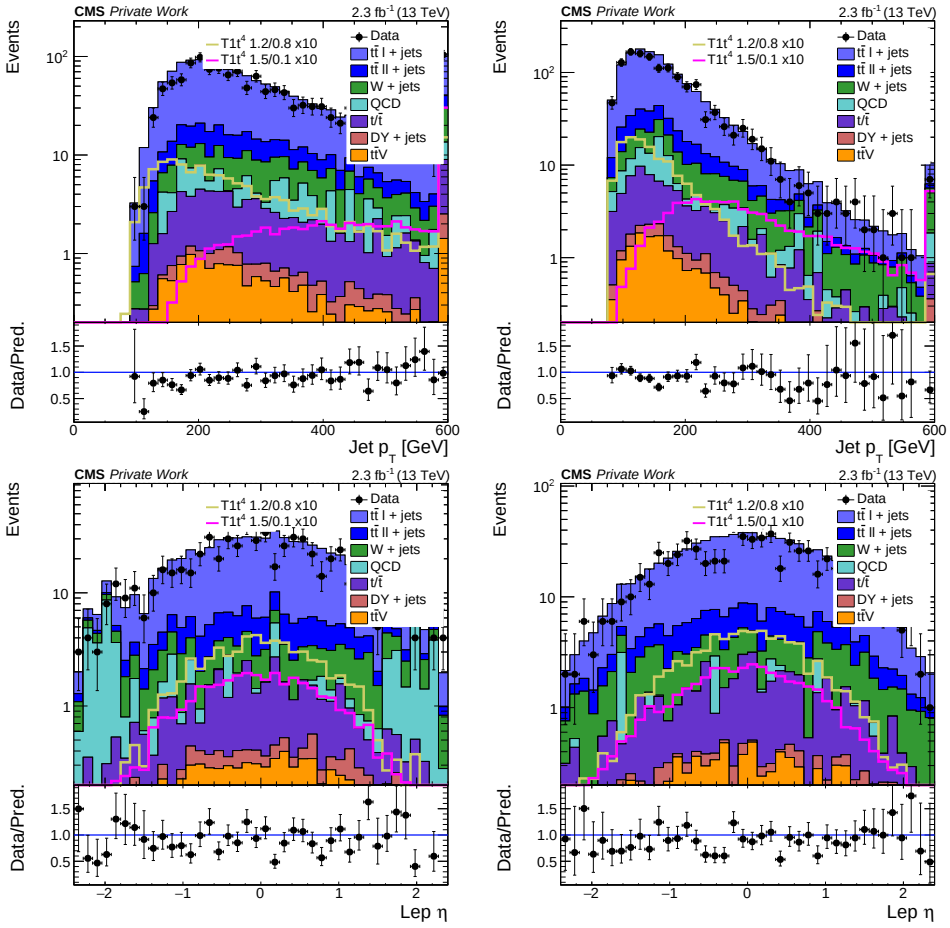


Fig. C.1: Distributions of the leading (top, left) and subleading (top, right) jet transverse momenta and electron (bottom, left) and muon(bottom, right) pseudorapidity after the baseline selection. Similar to the H_T distribution, the non-compressed T1tttt (1500,100) scenario dominates the high values of the jet momentum spectra in the top figures. The pseudorapidity distributions allow to see the different contribution of the *QCD* background for electron and muon events.

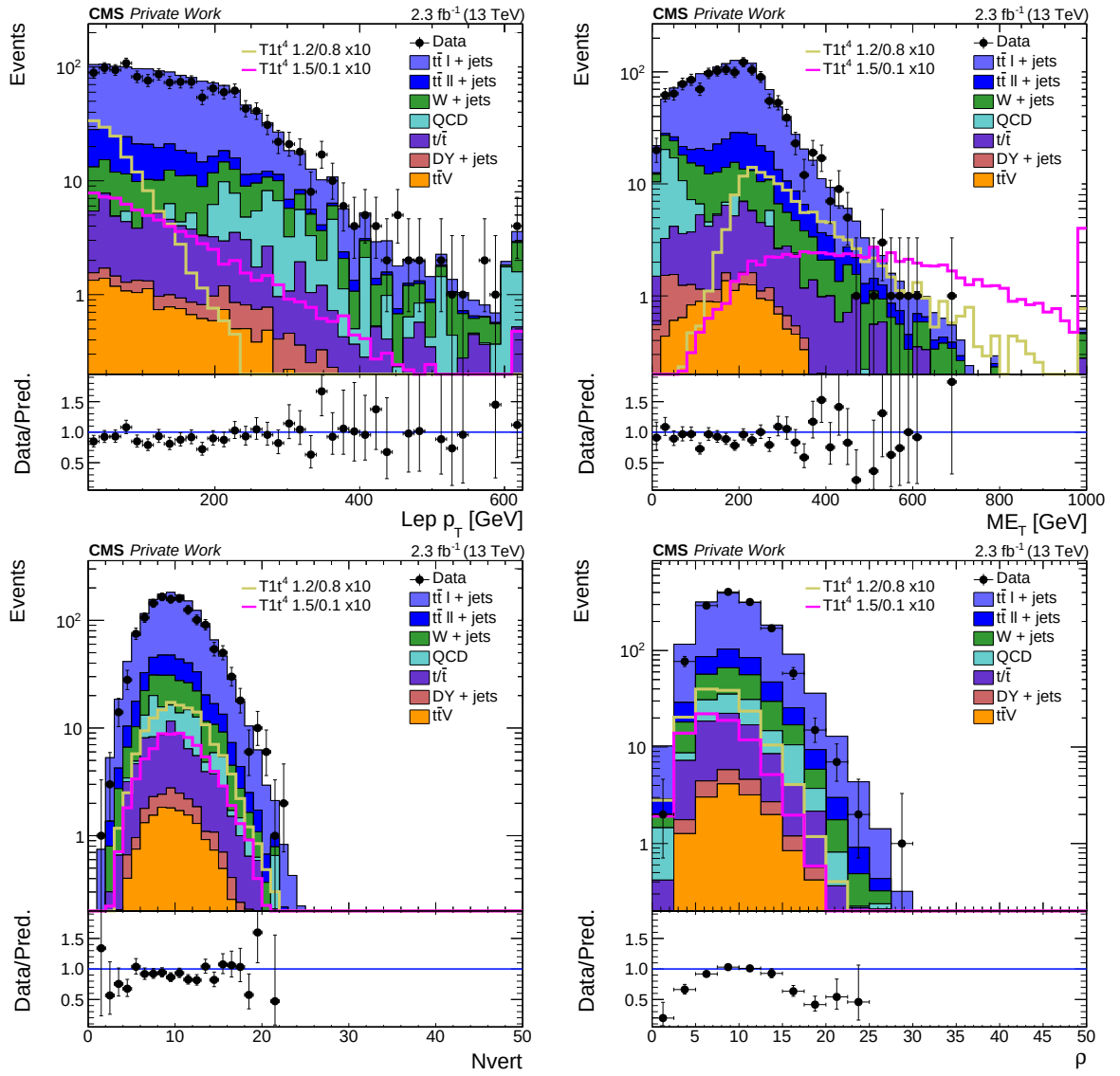


Fig. C.2: Distributions of the lepton p_T (top, left), $\not{E}_T$ (top, right) and number of reconstructed PV (bottom, left) and average hadronic activity (bottom, right) after the baseline selection. The shapes of the lepton p_T and $\not{E}_T$ distributions are affected by the L_T threshold, but also show a good data-MC agreement. The pileup distribution shows that the previously discussed corrections provide a good agreement, as well as the average hadronic activity, that is used to subtract the effect of pileup energy from the jet energy.
The data is shown in black dots and the shaded histograms represent the simulated SM background processes stacked on top of each other. Two benchmark signal points are scaled up by a factor of ten and shown by the coloured lines. The bottom pad shows the ratio of the data to the total background.

Appendix D

Additional plots for the QCD determination

D.1 L_p and $\Delta\varphi$ shape comparison

As the *QCD* shape is determined for the fit in L_p in the anti-selected sample (cf. Sec. 6.3), the similarity of the shapes for selected and anti-selected electrons has to be verified. A comparison of the L_p shapes is shown in Fig. D.1.

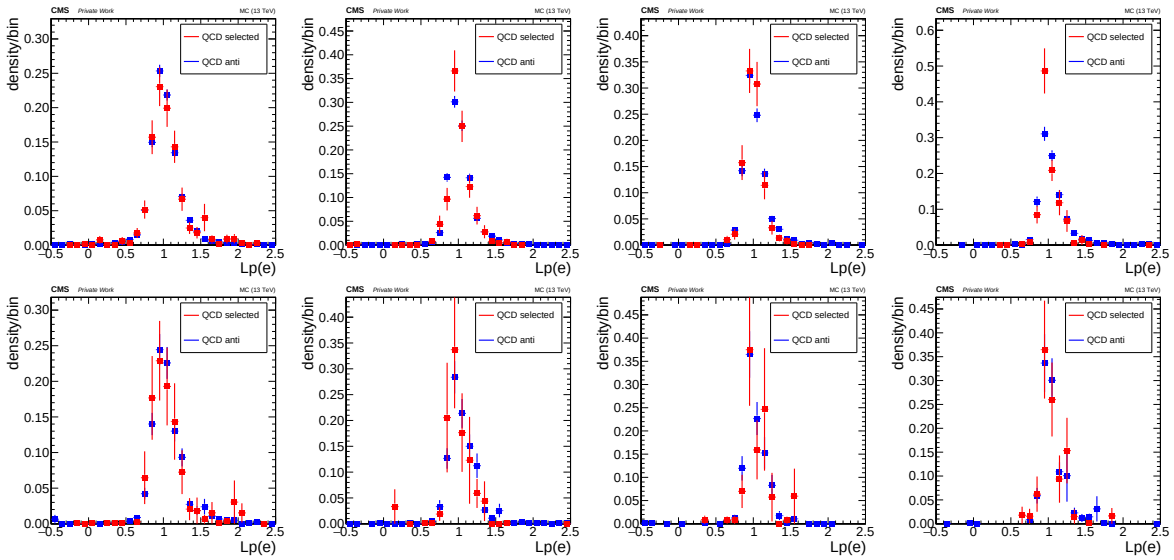


Fig. D.1: Shapes of L_p for selected and anti-selected *QCD* events in $n_{jet} \in [3,4]$ (top) and $n_{jet} \geq 6$ (bottom) for different L_T bins in simulation: $[250,350]$, $[350,450]$, $[450,600]$ and $> 600\,\text{GeV}$.

The R_{CS} method requires the knowledge of the *QCD* contribution in the signal and search regions separately. Since the *QCD* background estimation is performed inclusively with respect to $\Delta\varphi$, the R_{CS} factor for *QCD* multijets is determined as well in data in order to verify the absence of such events in the signal regions (cf. Sec. 6.3.1). The corresponding $\Delta\varphi$ shapes are shown in Fig. D.2 inclusive in H_T in the sideband ($n_{jet} \in [4,5]$) and in the main band ($n_{jet} \geq 6$). Given the good data-MC agreement, the R_{CS} obtained in simulated anti-selected *QCD* provide a good estimate on the selected *QCD* in data.

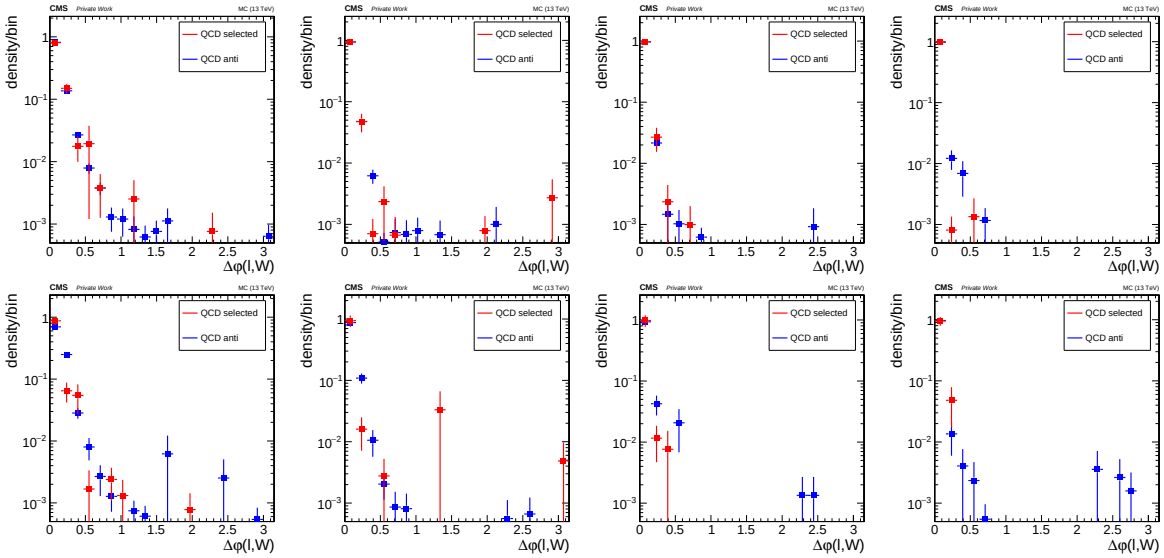


Fig. D.2: Shapes of $\Delta\varphi$ for selected and anti-selected *QCD* events in $n_{\text{jet}} \in [4,5]$ (top) and $n_{\text{jet}} \geq 6$ (bottom) for different L_{T} bins in simulation: $[250,350]$, $[350,450]$, $[450,600]$ and $> 600\,\text{GeV}$.

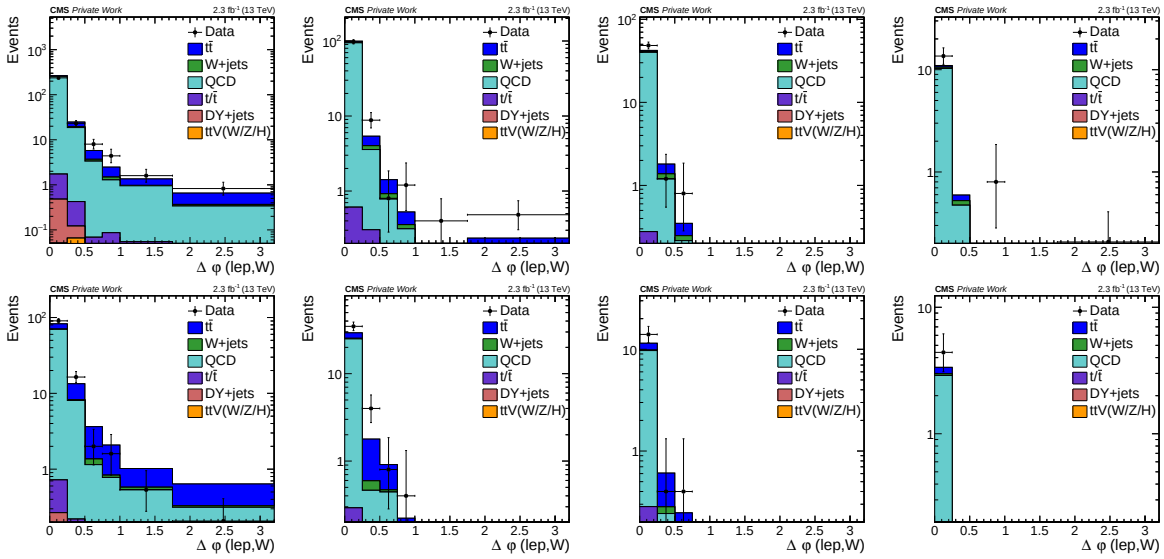


Fig. D.3: $\Delta\varphi$ for anti-selected events in data compared to simulation for n_{jet} = 4,5 (top) and n_{jet} = 6-8 (bottom) for different L_{T} bins with $n_{\text{b-tag}} \geq 1$: $[250,350]$, $[350,450]$, $[450,600]$ and $> 600\,\text{GeV}$. The content is normalized to the bin width.

D.2 L_p template fit in simulation

The results of the *QCD* fit with pseudo-data discussed in Sec. 6.3 is given in Fig. D.4. The best agreement is naturally found in the lower L_p region, which is populated mainly with *EWK* events: the template used to generate pseudo-data and to do the fit are identical. Moreover, the *QCD* dominated region around $L_p \approx 1$ shows also a good performance of the fit. The usage of *QCD* template from the inclusive $L_T > 250\,\text{GeV}$ bin allows to minimize the uncertainty of the shape, and reduce the number of outliers.

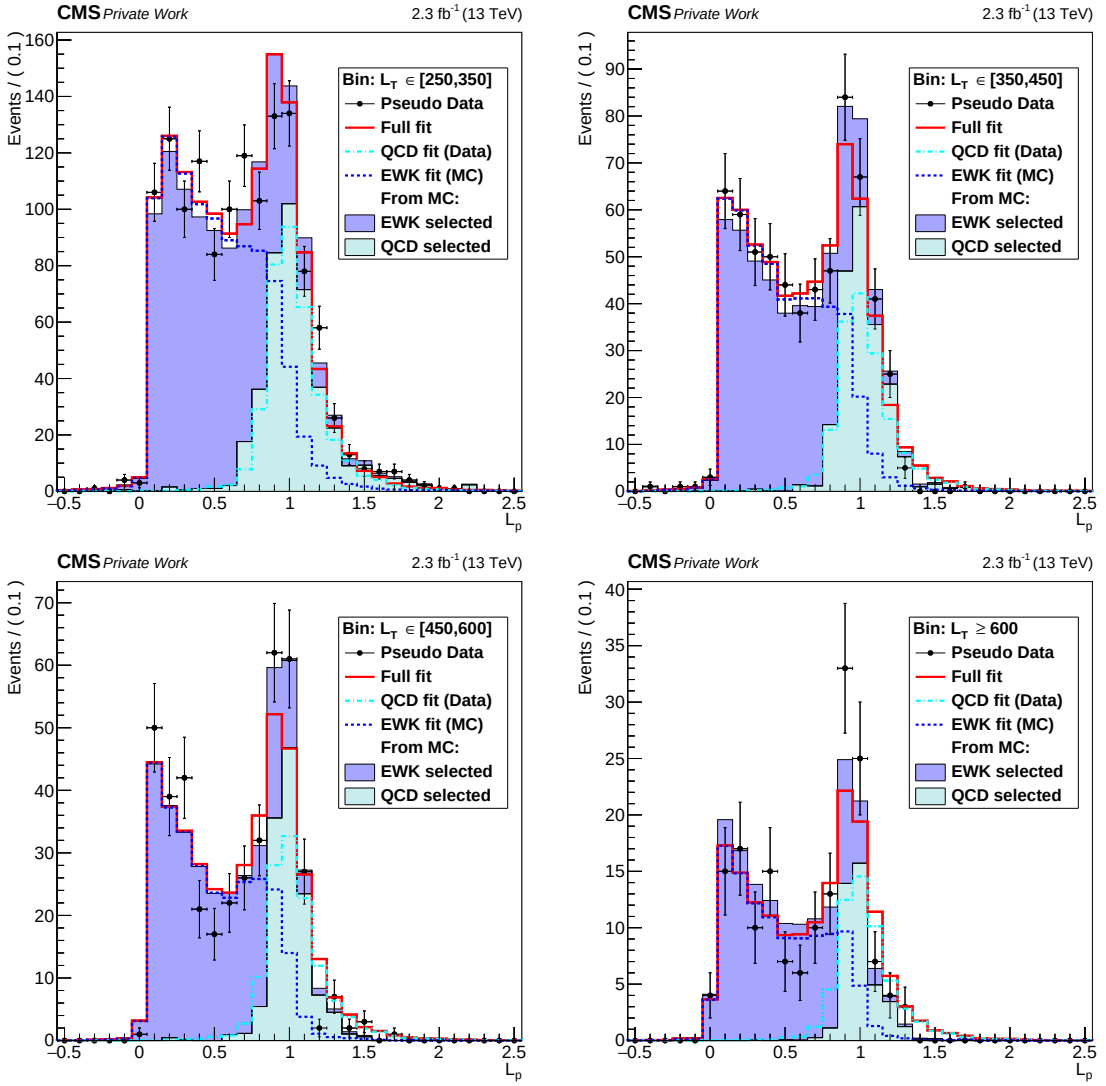


Fig. D.4: Fit results in the zero b-tag bin with n_{jet} = 3,4 for all four L_T bins. The black dots represent the pseudo-data obtained from the template of *EWK* and *QCD* selected events, which are shown with the blue and green dashed histograms. The full fit result is shown with the solid red line, while the separate *EWK* and *QCD* components are depicted by the blue and green dashed lines, respectively.

D.3 Trigger efficiency plots for anti-selected leptons

As mentioned in Sec. 6.3.2, the difference in trigger efficiency for selected and anti-selected leptons has an impact on the $F_{\text{sel-to-anti}}$ as:

$$F_{\text{sel-to-anti}} = \frac{N_{\text{QCD selected}} \times \epsilon^{\text{trig}}_{\text{sele}}}{N_{\text{QCD anti-selected}} \times \epsilon^{\text{trig}}_{\text{anti}}} \tag{D.1}$$

where ϵ^{trig} denotes the corresponding trigger efficiency for the anti/selected leptons.

For the electron channel these efficiencies cancel out since the ratio is using data for the number of anti-selected electrons, and afterwards it is multiplied by $N_{\text{QCD anti-selected}}$ in the side and main bands. Therefore we don't take into account these efficiency differences for the prediction. But if one actually compares the obtained $F_{\text{sel-to-anti}}$ from MC (Fig. 6.17) and Data (Fig. 6.20), one can observe, that the values show the same dependency of L_{T}, except for a systematic shift. This can be exactly explained by the trigger efficiencies in data, which are absent in MC. Figure D.5 shows the lepton and hadron leg turn-ons for the analysis triggers like in Sec. 5.2.2 but now for anti-selected leptons. Although no strict plateau is observed for electrons, the trigger efficiency can be estimated from the right end of the curves: since *QCD* mostly resides at low $\not{E}_{\text{T}}$, the high L_{T} stems from high p_{T} leptons.

For anti-selected muons the turn-ons (Fig. D.5, right) look very similar to the selected muons, but with an approximately 10% drop which can be expected from the relaxed ID criteria. This allows to claim the $F_{\text{sel-to-anti}}$ taken from MC can be well used for the data *QCD* prediction in the muon channel.

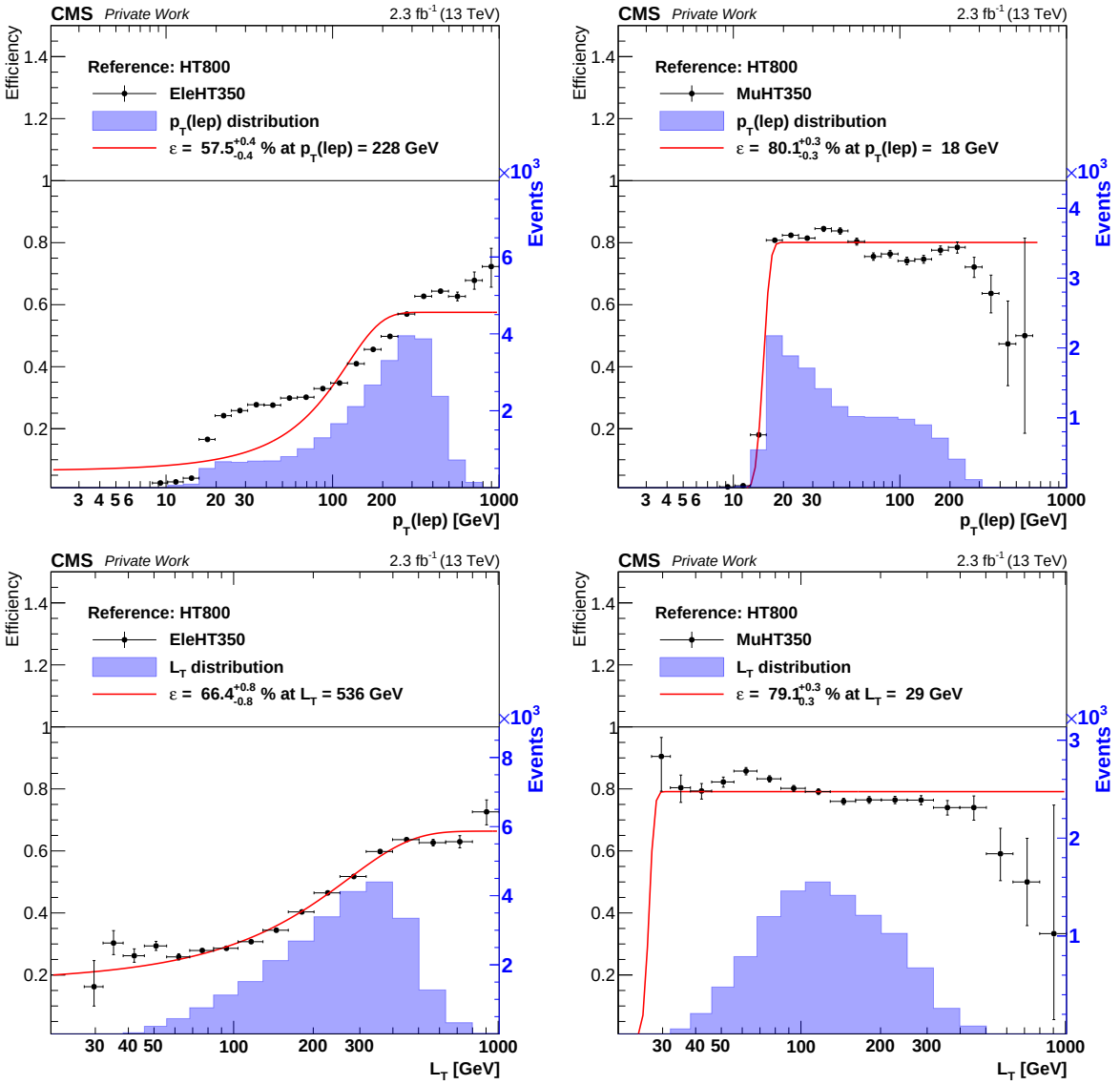


Fig. D.5: Measurement of the `Lep15_IsoVVVL_PFHT350` trigger efficiency for the electron (*left*) and the muon (*right*) paths. The top plots show the turn-on of the leptonic leg, while the bottom show the turn-on of the hadronic leg. The number of events passing the trigger is given as shaded histogram.

Appendix E

B-tagging scale factor methods

As discussed in Sec. 5.3.3, the simulated b-jet identification efficiency is corrected to match the one in data. Multiple methods exist to obtain such corrections, and those adopted in the CMS experiment can be found in Ref [156]. In the following two different approaches will be reviewed briefly and their effect on the simulated κ transfer factors will be compared.

1A For each simulated event a single weight is computed, which represent the ratio of probabilities to measure the observed $n_{\text{b-tag}}$ in data and simulation given the particular event content. The total number of simulated events used later on remains unchanged.

1B For each simulated event multiple weights are computed corresponding each to a particular value of $n_{\text{b-tag}}$. The event content and b-tagging efficiencies in data and simulation are used.

While method 1A represents an event-by-event scaling like most other scale factors, method 1B allows to reuse each simulated event multiple times. The main advantage of the latter method is the increased available sample size, since no generated events are rejected by the $n_{\text{b-tag}}$ requirements. On the other hand, bin-to-bin correlations are induced, which are difficult to handle when evaluating systematic uncertainties. Therefore, the main method used in this thesis is 1A, and 1B is used to provide a cross-check for the κ evaluation.

The plots in Fig. E.1 show the κ factors obtained from simulation using both b-tag scale factor methods. While the values obtained by the 1B method in general do follow those of 1A, less fluctuations between the bins are found. Since the dileptonic $\mathrm{t\bar{t}}$+jets process has larger R_{CS}, it is less affected by statistical fluctuations as e.g. the semileptonic sample. A similar picture was found for the other simulated SM backgrounds.

The bottom plot shows the κ values for the combined *EWK* backgrounds, where also reduced statistical uncertainties are observed with method 1B, while the overall agreement with the first approach is satisfactory.

Given the increased complexity of method 1B and similar results to 1A, the latter method is found to be satisfactory.

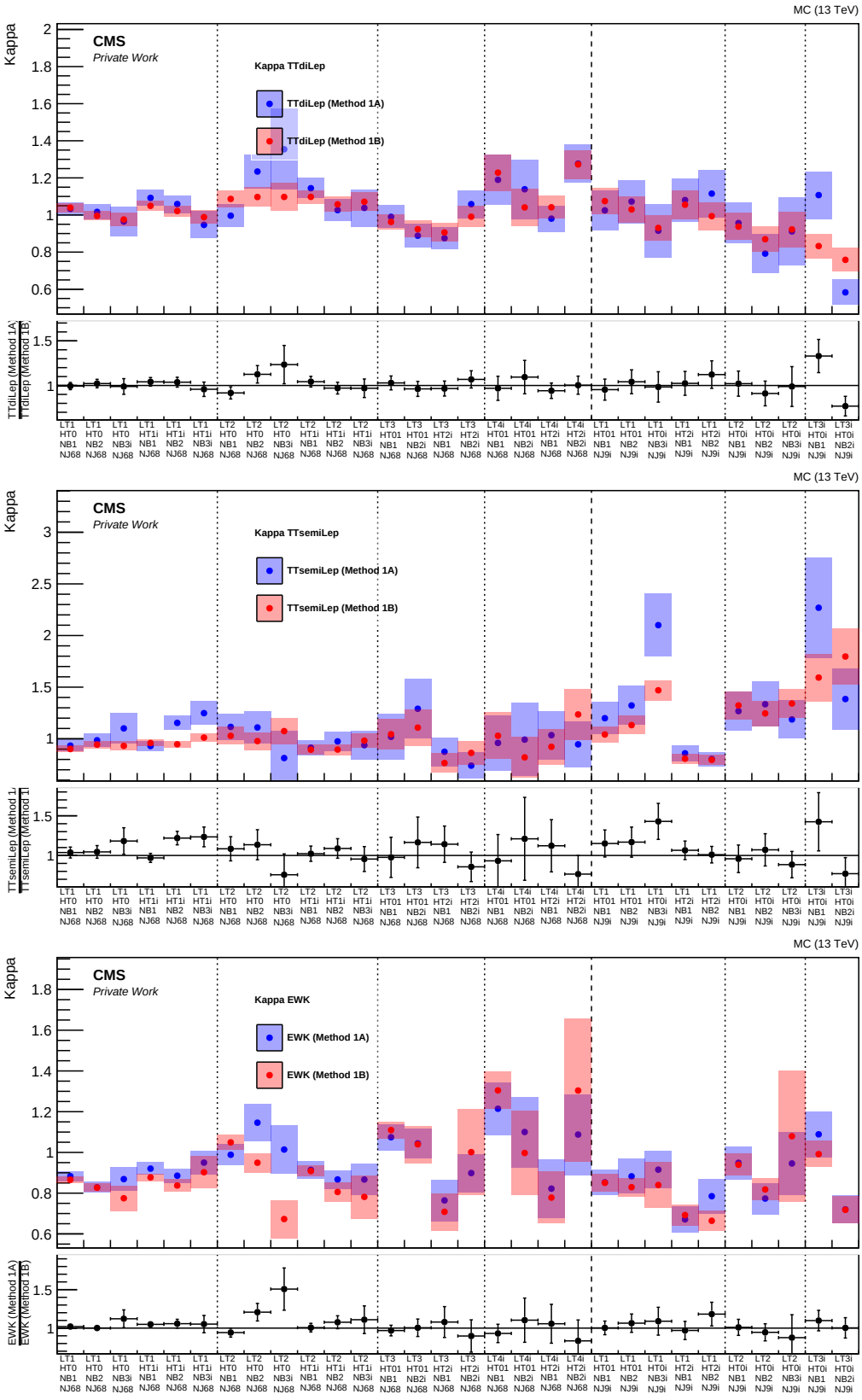


Fig. E.1: Comparison of simulated κ values computed with two different b-tagging scale factor approaches. The *top* plot corresponds to dileptonic $t\bar{t}$+jets, where no significant difference between the methods is found. For semileptonic $t\bar{t}$ events the increased event sample allows for lower statistical uncertainties using method 1B. In the combined *EWK* sample both approaches provide consistent κ values.

Bibliography

[1] E. Noether, "Invarianten beliebiger Differentialausdrücke.", *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* **1918** (1918) 37–44.

[2] E. Noether, "Invariante Variationsprobleme.", *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* **1918** (1918) 235–257.

[3] M. H. Poincaré, "Sur la dynamique de l'électron", *Rendiconti del Circolo Matematico di Palermo (1884-1940)* **21** (1906), no. 1, 129–175, doi:10.1007/BF03013466.

[4] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[5] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[6] S. P. Martin, "A Supersymmetry primer", doi:10.1142/9789812839657_0001,10.1142/9789814307505_0001, arXiv:hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[7] I. Melzer-Pellmann and P. Pralavorio, "Lessons for SUSY from the LHC after the first run", *Eur. Phys. J.* **C74** (2014) 2801, doi:10.1140/epjc/s10052-014-2801-y, arXiv:1404.7191.

[8] P. Bechtle, T. Plehn, and C. Sander, "The Large Hadron Collider: Harvest of Run 1", ch. Supersymmetry, pp. 421–462. Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-15001-7_10.

[9] CMS Collaboration, "Search for supersymmetry in final states with a single lepton, b-quark jets, and missing transverse energy in proton-proton collisions at $\sqrt{s} = 7$ TeV", *Phys.Rev.* **D87** (2013), no. 5, 052006, doi:10.1103/PhysRevD.87.052006, arXiv:1211.3143.

[10] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s}$=8 TeV in events with a single lepton, large jet multiplicity, and multiple b jets", *Phys.Lett.* **B733** (2014) 328–353, doi:10.1016/j.physletb.2014.04.023, arXiv:1311.4937.

[11] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ TeV in events with a single lepton, jets, and missing transverse momentum", *Eur.Phys.J.* **C73** (2013) 2404, doi:10.1140/epjc/s10052-013-2404-z, arXiv:1212.6428.

[12] L. Okun, "Leptons and Quarks". North-Holland Personal Library. Elsevier Science, 2013.

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

[14] D. Griffiths, "Introduction to Elementary Particles". Physics textbook. Weinheim, Germany: Wiley-VCH 454 p, 2014.

[15] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

[16] MissMJ, "Standard Model of Elementary Particles". https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. (Accessed on 05/18/2016).

[17] Belle Collaboration, "Observation of a resonance-like structure in the pi+- psi-prime mass distribution in exclusive B —> K pi+- psi-prime decays", *Phys. Rev. Lett.* **100** (2008) 142001, doi:10.1103/PhysRevLett.100.142001, arXiv:0708.1790.

[18] LHCb Collaboration, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, doi:10.1103/PhysRevLett.115.072001, arXiv:1507.03414.

[19] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[20] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[21] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367–377.

[22] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[23] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[24] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[25] D0 Collaboration, "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* **74** (1995) 2422–2426, doi:10.1103/PhysRevLett.74.2422, arXiv:hep-ex/9411001.

[26] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", arXiv:1502.01589.

[27] G. 't Hooft, "Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking", *NATO Sci. Ser. B* **59** (1980) 135.

[28] P. Ramond, "Dual Theory for Free Fermions", *Phys. Rev.* **D3** (1971) 2415–2418, doi:10.1103/PhysRevD.3.2415.

[29] Yu. A. Golfand and E. P. Likhtman, "Extension of the Algebra of Poincare Group Generators and Violation of p Invariance", *JETP Lett.* **13** (1971) 323–326. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].

[30] A. Neveu and J. H. Schwarz, "Factorizable dual model of pions", *Nucl. Phys.* **B31** (1971) 86–112, doi:10.1016/0550-3213(71)90448-2.

[31] D. V. Volkov and V. P. Akulov, "Possible universal neutrino interaction", *JETP Lett.* **16** (1972) 438–440. [Pisma Zh. Eksp. Teor. Fiz.16,621(1972)].

[32] J. Wess and B. Zumino, "Supergauge Transformations in Four-Dimensions", *Nucl. Phys.* **B70** (1974) 39–50, doi:10.1016/0550-3213(74)90355-1.

[33] G. F. Giudice and R. Rattazzi, "Theories with gauge mediated supersymmetry breaking", *Phys. Rept.* **322** (1999) 419–499, doi:10.1016/S0370-1573(99)00042-3, arXiv:hep-ph/9801271.

[34] A. D. Sakharov, "Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35, doi:10.1070/PU1991v034n05ABEH002497. [Usp. Fiz. Nauk161,61(1991)].

[35] T. Kaluza, "Zum Unitätsproblem der Physik", *Sitzungsber. Preuss. Akad. Wiss.* (1921) 966–972.

[36] O. Klein, "Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)", *Z.Phys.* **37** (1926) 895–906, doi:10.1007/BF01397481.

[37] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, "The Hierarchy problem and new dimensions at a millimeter", *Phys. Lett.* **B429** (1998) 263–272, doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.

[38] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[39] S. Sekmen et al., "Interpreting LHC SUSY searches in the phenomenological MSSM", *JHEP* **02** (2012) 075, doi:10.1007/JHEP02(2012)075, arXiv:1109.5119.

[40] CMS Collaboration, "Phenomenological MSSM interpretation of CMS results at $\sqrt{s} = 7$ and 8 TeV", Technical Report CMS-PAS-SUS-15-010, CERN, Geneva, 2015.

[41] B. Knuteson and S. Mrenna, "BARD: Interpreting new frontier energy collider physics", arXiv:hep-ph/0602101.

[42] N. Arkani-Hamed et al., "MARMOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories", arXiv:hep-ph/0703088.

[43] M. Berggrenet al., "Non-Simplified SUSY: stau-Coannihilation at LHC and ILC", in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*. 2013. arXiv:1307.8076.

[44] CMS Collaboration, "Interpretation of Searches for Supersymmetry with simplified Models", *Phys. Rev.* **D88** (2013), no. 5, 052017, doi:10.1103/PhysRevD.88.052017, arXiv:1301.2175.

[45] CMS Collaboration, "Searches for supersymmetry based on events with b jets and four W bosons in pp collisions at 8 TeV", *Phys. Lett.* **B745** (2015) 5–28, doi:10.1016/j.physletb.2015.04.002, arXiv:1412.4109.

[46] ATLAS Collaboration, "Summary of the searches for squarks and gluinos using $\sqrt{s} = 8$ TeV pp collisions with the ATLAS experiment at the LHC", *JHEP* **10** (2015) 054, doi:10.1007/JHEP10(2015)054, arXiv:1507.05525.

[47] CMS Collaboration, "Exclusion limits on gluino and top-squark pair production in natural SUSY scenarios with inclusive razor and exclusive single-lepton searches at 8 TeV.", Technical Report CMS-PAS-SUS-14-011, CERN, Geneva, 2014.

[48] M. Papucci, J. T. Ruderman, and A. Weiler, "Natural SUSY Endures", *JHEP* **09** (2012) 035, doi:10.1007/JHEP09(2012)035, arXiv:1110.6926.

[49] CMS Collaboration, "Summary of comparison plots in simplified models spectra for the 8TeV dataset". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SUSYSMSSummaryPlots8TeV. (Accessed on 05/23/2016).

[50] ATLAS Collaboration, "Summary plots from the ATLAS Supersymmetry physics group". https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SUSY/. (Accessed on 05/23/2016).

[51] E. Bertuzzo, "SUSY after LHC8: a brief overview", *EPJ Web Conf.* **60** (2013) 18001, doi:10.1051/epjconf/20136018001, arXiv:1307.0318.

[52] C. Borschensky et al., "Squark and gluino production cross sections in pp collisions at $\sqrt{s}$ = 13, 14, 33 and 100 TeV", *Eur. Phys. J.* **C74** (2014), no. 12, 3174, doi:10.1140/epjc/s10052-014-3174-y, arXiv:1407.5066.

[53] W. Beenakker et al., "Supersymmetric top and bottom squark production at hadron colliders", *JHEP* **08** (2010) 098, doi:10.1007/JHEP08(2010)098, arXiv:1006.4771.

[54] W. Beenakker et al., "Squark and Gluino Hadroproduction", *Int. J. Mod. Phys.* **A26** (2011) 2637–2664, doi:10.1142/S0217751X11053560, arXiv:1105.1110.

[55] C. Borschensky et al., "LHC SUSY Cross Section Working Group". `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections`. (Accessed on 05/23/2016).

[56] CMS Collaboration, "Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at 8 TeV", Technical Report CMS-PAS-TOP-12-041, CERN, Geneva, 2013.

[57] CMS Collaboration, "Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{(s)} = 13$ TeV", *Phys. Rev. Lett.* **116** (2016), no. 5, 052002, `doi:10.1103/PhysRevLett.116.052002`, `arXiv:1510.05302`.

[58] CMS Collaboration, "Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. Lett.* **112** (2014) 191802, `doi:10.1103/PhysRevLett.112.191802`, `arXiv:1402.0923`.

[59] CMS Collaboration, "Measurement of inclusive W and Z boson production cross sections in pp collisions at sqrt(s)=13 TeV", Technical Report CMS-PAS-SMP-15-004, CERN, Geneva, 2015.

[60] O. S. Brüning et al., "LHC Design Report". CERN, Geneva, 2004.

[61] L. R. Evans and P. Bryant, "LHC Machine", *J. Instrum.* **3** (2008) S08001. 164 p. This report is an abridged version of the LHC Design Report (CERN-2004-003).

[62] "Superconductivity and the LHC: the early days - CERN Courier". `http://cerncourier.com/cws/article/cern/47504`. (Accessed on 03/14/2016).

[63] "LHC Inauguration - LHC Exhibition". `https://lhc2008.web.cern.ch/lhc2008/inauguration/lhcexpo.html`. (Accessed on 03/14/2016).

[64] D. Fournier, "Performance of the LHC, ATLAS and CMS in 2011", *EPJ Web Conf.* **28** (2012) 01003, `doi:10.1051/epjconf/20122801003`, `arXiv:1201.4681`.

[65] M. Lamont, "Status of the LHC", *Journal of Physics: Conference Series* **455** (2013), no. 1, 012001.

[66] R. Alemany-Fernandez, "Plans for the Ion Run in 2016", in *Proceedings, LHC Performance Workshop (Chamonix 2016)*. 2016.

[67] C. De Melis, "The CERN accelerator complex. Complexe des accélérateurs du CERN",. General Photo.

[68] "The HL-LHC project | The HL-LHC Project". `http://hilumilhc.web.cern.ch/about/hl-lhc-project`. (Accessed on 03/15/2016).

[69] ATLAS Collaboration Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *J. Instrum.* **3** (2008) S08003. 437 p. Also published by CERN Geneva in 2010.

[70] "Publications < AtlasPublic < TWiki". `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/Publications`. (Accessed on 03/15/2016).

[71] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006, `doi:10.1088/1748-0221/3/08/S08006`.

[72] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, `doi:10.1088/1748-0221/3/08/S08002`.

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

[74] TOTEM Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007, `doi:10.1088/1748-0221/3/08/S08007`.

[75] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[76] MoEDAL Collaboration, "Technical Design Report of the MoEDAL Experiment", Technical Report CERN-LHCC-2009-006. MoEDAL-TDR-001, CERN, Geneva, Jun, 2009.

[77] CMS Collaboration, "Letter of intent: by the CMS for a general purpose detector at LHC", Technical Report CERN-LHCC-92-003. CERN-LHCC-92-3. LHCC-I-1, CERN, Geneva, 1992. Open presentation to the LHCC 5 November 1992, M. Della Negra/CERN, CMS Spokesman.

[78] "CMS Collaboration | CMS Experiment". http://cms.web.cern.ch/content/cms-collaboration. (Accessed on 03/15/2016).

[79] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032, arXiv:1311.4942.

[80] CMS Collaboration, "Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays", *JINST* **5** (2010) T03021, doi:10.1088/1748-0221/5/03/T03021, arXiv:0910.5530.

[81] "CMS-doc-11806-v2: Images for US LHC CMS Detector Upgrade Project". https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=11806. (Accessed on 03/15/2016).

[82] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[83] CMS Collaboration, G. L. Bayatian, S. Chatrchyan, and G. Hmayakyan, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.

[84] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[85] CMS HCAL Collaboration, "Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges", Technical Report CMS-NOTE-2006-138, CERN, Geneva, May, 2007.

[86] CMS HCAL Collaboration, "Design, Performance and Calibration of the CMS Forward Calorimeter Wedges", Technical Report CMS-NOTE-2006-044, CERN, Geneva, Feb, 2006.

[87] CMS HCAL Collaboration, "Design, performance, and calibration of the CMS Hadron-outer calorimeter", *Eur. Phys. J.* **C57** (2008) 653–663, doi:10.1140/epjc/s10052-008-0756-6.

[88] CMS Collaboration, P. Gunnellini, "The CASTOR calorimeter at the CMS experiment", in *16th International Moscow School of Physics and 41st ITEP Winter School of Physics Moscow, Russia, February 12-19, 2013*. 2013. arXiv:1304.2943.

[89] CMS Collaboration, "Status of zero degree calorimeter for CMS experiment", *AIP Conf. Proc.* **867** (2006) 258–265, doi:10.1063/1.2396962, arXiv:nucl-ex/0608052. [,258(2006)].

[90] CMS Collaboration, "CSC local efficiency measurements from run II data",.

[91] CMS Collaboration, G. L. Bayatyan et al., "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems". Technical Design Report CMS.

[92] CMS Collaboration, S. Cittolin, A. Rácz, and P. Sphicas, "CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project". Technical Design Report CMS. CERN, Geneva, 2002.

[93] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", Technical Report CERN-LHCC-2013-011. CMS-TDR-12, CERN, Geneva, Jun, 2013. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[94] CMS Collaboration, "Operation and Performance of the Upgraded CMS Calorimeter Trigger in LHC Run 2", Technical Report CMS-CR-2015-317, CERN, Geneva, Nov, 2015.

[95] M. Jeitleret al., "Upgrade of the Global Muon Trigger for the CMS experiment", *Journal of Instrumentation* **8** (2013), no. 12, C12017.

[96] CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", Technical Report CMS-PAS-LUM-15-001, CERN, Geneva, 2016.

[97] A. Bell et al., "Fast beam conditions monitor BCM1F for the CMS experiment", *Nuclear Instruments and Methods in Physics Research A* **614** (March, 2010) 433–438, doi:10.1016/j.nima.2009.12.056, arXiv:0911.2480.

[98] CMS Collaboration, "Technical proposal for the upgrade of the CMS detector through 2020", Technical Report CERN-LHCC-2011-006. LHCC-P-004, CERN, Geneva, Jun, 2011.

[99] D. Contardo et al., "Technical Proposal for the Phase-II Upgrade of the CMS Detector", Technical Report CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, Geneva, Jun, 2015. Upgrade Project Leader Deputies: Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts: Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch.

[100] "MonteCarloGenerators – Sherpa – Hepforge". https://sherpa.hepforge.org/trac/wiki/MonteCarloGenerators. (Accessed on 03/29/2016).

[101] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

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

[103] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[104] M. L. Mangano, M. Moretti, and R. Pittau, "Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study", *Nucl. Phys.* **B632** (2002) 343–362, doi:10.1016/S0550-3213(02)00249-3, arXiv:hep-ph/0108069.

[105] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[106] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[107] A. Giammanco, "The Fast Simulation of the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022012, doi:10.1088/1742-6596/513/2/022012.

[108] J. Freeman, "Silicon photomultipliers for the CMS hadron calorimeter", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **617** (2010), no. 1–3, 393 – 395, doi:http://dx.doi.org/10.1016/j.nima.2009.10.132. 11th Pisa Meeting on Advanced DetectorsProceedings of the 11th Pisa Meeting on Advanced Detectors.

[109] CMS Collaboration, B. Lutz, "Upgrade of the CMS Hadron Outer Calorimeter with SiPM sensors", in *15th International Conference on Calorimetry in High Energy Physics (CALOR 2012)*, volume 404, p. 012018. 2012. doi:10.1088/1742-6596/404/1/012018.

[110] CMS Collaboration, A. Lobanov, "The CMS Outer HCAL SiPM Upgrade", in *16th International Conference on Calorimetry in High Energy Physics (CALOR 2014)*, volume 587, p. 012005. 2015. doi:10.1088/1742-6596/587/1/012005.

[111] J. Anderson, J. Freeman, S. Los, and J. Whitmore, "Upgrade of the CMS Hadron Outer Calorimeter with SIPMs", in *Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011)*, volume 37, pp. 72 – 78. 2012. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011). doi:10.1016/j.phpro.2012.02.358.

[112] CMS Collaboration, A. Künsken, "MPPC photon sensor operational experience in CMS", in *Proceedings of 16th International Conference on Calorimetry in High Energy Physics (CALOR 2014)*, volume 587, p. 012022. 2015. doi:10.1088/1742-6596/587/1/012022.

[113] CMS Collaboration, A. Lobanov, B. Lutz, K. Kaadze et al., "Commissioning and performance of the CMS Hadron Outer Calorimeter", 2014. Unpublished. Available at http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=DN2014_019.pdf.

[114] G. Bauer et al., "Status of the CMS Detector Control System", *J. Phys. Conf. Ser.* **396** (2012) 012023, doi:10.1088/1742-6596/396/1/012023.

[115] CMS Collaboration, P. de Barbaro, D. Vishnevsky, A. Bodek et al., "Validation of HCAL energy calibration using cosmic ray muon data from Global Run at the end of November 2007 (GREN)", June, 2008. Unpublished. Available at http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=IN2008_025.pdf.

[116] "File:3D Spherical.svg - Wikimedia Commons". https://commons.wikimedia.org/wiki/File:3D_Spherical.svg. (Accessed on 03/24/2016).

[117] G. Petrucciani, A. Rizzi, and C. Vuosalo, "Mini-AOD: A New Analysis Data Format for CMS", *J. Phys. Conf. Ser.* **664** (2015), no. 7, 072052, doi:10.1088/1742-6596/664/7/072052.

[118] "WorkBookDataFormats < CMSPublic < TWiki". https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookDataFormats. (Accessed on 03/30/2016).

[119] "CMS-doc-4172-v2: Interactive Slice of the CMS detector". https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4172. (Accessed on 03/30/2016).

[120] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", technical report, 2009.

[121] CMS Collaboration, F. Beaudette, "The CMS Particle Flow Algorithm", in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*, pp. 295–304. 2013. arXiv:1401.8155.

[122] CMS Collaboration, "Particle Flow and Global Event Description in CMS", Technical Report CMS-PAS-PRF-14-001, CERN, Geneva, 2014. Unpublished.

[123] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track Reconstruction in the CMS tracker", Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec, 2006.

[124] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *eConf* **C0303241** (2003) TULT009, doi:10.1088/0954-3899/31/9/N01, arXiv:physics/0306087. [J. Phys.G31,N9(2005)].

[125] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[126] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV", *JINST* **10** (2015), no. 08, P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[127] CMS Collaboration, "Performance of τ-lepton reconstruction and identification in CMS", *Journal of Instrumentation* **7** (January, 2012) 1001, doi:10.1088/1748-0221/7/01/P01001, arXiv:1109.6034.

[128] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys.* **G34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.

[129] A. Apresyan, "First Results and Prospects from CMS at 13 TeV". http://theory.fnal.gov/jetp/talks/Apresyan_W&C.pdf. (Accessed on 03/31/2016).

[130] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[131] CMS Collaboration, "First Muon Identification Efficiencies with 13 TeV, 50 ns Data",.

[132] CMS Collaboration, "SWGuideMuonIdRun2 < CMS < TWiki". https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2. (Accessed on 04/01/2016).

[133] CMS Collaboration, "Muon Identification in CMS", 2008. Unpublished. Available at http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2008/098.

[134] "Proposal of a new"Medium" Muon Id". https://indico.cern.ch/event/357213/contributions/1769745/attachments/710701/975626/muonid-pog081214.pdf. (Accessed on 05/09/2016).

[135] CMS Collaboration, "Electron and Photon performance using data collected by CMS at $\sqrt{s}$ = 13 TeV and 25ns",.

[136] CMS Collaboration, "CutBasedElectronIdentificationRun2 < CMS < TWiki". https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2#Spring15_selection_25ns. (Accessed on 04/02/2016).

[137] M. F. Sevilla, "Alternatives to Standard Isolation for 13 TeV". https://indico.cern.ch/event/360073/contribution/2/attachments/717288/984663/15-01-20_ucsb_isolation_ra2b.pdf. (Accessed on 04/02/2016).

[138] K. Rehermann and B. Tweedie, "Efficient Identification of Boosted Semileptonic Top Quarks at the LHC", *JHEP* **03** (2011) 059, doi:10.1007/JHEP03(2011)059, arXiv:1007.2221.

[139] CMS Collaboration, "Search for supersymmetry in pp collisions at sqrt (s) = 13 TeV in the single-lepton final state using the sum of masses of large radius jets", Technical Report CMS-PAS-SUS-15-007, CERN, Geneva, 2015.

[140] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", Technical Report CERN-PH-EP-2015-305. CMS-JME-13-004-003. arXiv:1607.03663, CERN, Geneva, Jul, 2016. Comments: Submitted to J. Instrum.

[141] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[142] CMS Collaboration, "Pileup Jet Identification", Technical Report CMS-PAS-JME-13-005, CERN, Geneva, 2013.

[143] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[144] CMS Collaboration, "8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb^{-1} of data in CMS",.

[145] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[146] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[147] CMS Collaboration, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 136, doi:10.1007/JHEP03(2011)136, arXiv:1102.3194.

[148] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[149] "DataQuality < CMSPublic < TWiki". https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality. (Accessed on 04/06/2016).

[150] "MissingETOptionalFiltersRun2 < CMS < TWiki". https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2. (Accessed on 04/08/2016).

[151] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[152] "SummaryTable1G25ns < CMS < TWiki". https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns. (Accessed on 04/08/2016).

[153] "SUSYMCProduction13TeV < CMS < TWiki". https://twiki.cern.ch/twiki/bin/view/CMS/SUSYMCProduction13TeV. (Accessed on 04/08/2016).

[154] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J.* **C75** (2015), no. 11, 542, doi:10.1140/epjc/s10052-015-3709-x, arXiv:1505.04480.

[155] "TopPtReweighting < CMS < TWiki". https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting. (Accessed on 04/10/2016).

[156] "BTagSFMethods < CMS < TWiki". https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods. (Accessed on 04/10/2016).

[157] "SUSY Data/MC Lepton Scale Factors". https://twiki.cern.ch/twiki/bin/view/CMS/SUSLeptonSF. (Accessed on 05/15/2016).

[158] CMS Collaboration, "Search for supersymmetry in events with one lepton in proton-proton collisions at sqrt(s)=13 TeV with the CMS experiment", Technical Report CMS-PAS-SUS-15-006, CERN, Geneva, 2016.

[159] CMS Collaboration, "Search for Supersymmetry with single-lepton events at 13 TeV using 2015 data", 2015. Unpublished. Available at http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/207.

[160] CMS Collaboration, "Measurement of the Polarization of W Bosons with Large Transverse Momenta in W+Jets Events at the LHC", *Phys. Rev. Lett.* **107** (2011) 021802, doi:10.1103/PhysRevLett.107.021802, arXiv:1104.3829.

[161] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ TeV in events with a single lepton, jets, and missing transverse momentum", *Eur. Phys. J.* **C73** (2013) 2404, doi:10.1140/epjc/s10052-013-2404-z, arXiv:1212.6428.

[162] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf* **C0303241** (2003) MOLT007, arXiv:physics/0306116. [,186(2003)].

[163] Z. Bern et al., "Left-Handed W Bosons at the LHC", *Phys. Rev.* **D84** (2011) 034008, arXiv:1103.5445.

[164] CMS Collaboration, "Angular coefficients of Z bosons produced in pp collisions at $\sqrt{s}$ = 8 TeV and decaying to $\mu^+\mu^-$ as a function of transverse momentum and rapidity", *Phys. Lett.* **B750** (2015) 154–175, doi:10.1016/j.physletb.2015.08.061, arXiv:1504.03512.

[165] ATLAS Collaboration, "Measurement of the polarisation of W bosons produced with large transverse momentum in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS experiment", *Eur. Phys. J.* **C72** (2012) 2001, doi:10.1140/epjc/s10052-012-2001-6, arXiv:1203.2165.

[166] CMS Collaboration, "Hadronic Recoil Studies of Heavy Boosted Systems", 2013. Unpublished. Available at http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2013/059.

[167] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[168] A. L. Read, "Modified frequentist analysis of search results (The CL(s) method)", in *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. 2000.

[169] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313. [,11(2002)].

[170] The ATLAS Collaboration, The CMS Collaboration, 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.

[171] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Philosophical Transactions of the Royal Society of London Series A* **231** (1933) 289–337, doi:10.1098/rsta.1933.0009.

[172] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].

[173] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large", *Transactions of the American Mathematical Society* **54** (1943), no. 3, 426–482.

[174] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Annals Math. Statist.* **9** (1938), no. 1, 60–62, doi:10.1214/aoms/1177732360.

[175] "GitHub CMS Higgs Combination toolkit". https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit. (Accessed on 05/12/2016).

[176] "SWGuideHiggsAnalysisCombinedLimit < CMS < TWiki". https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit. (Accessed on 05/12/2016).

[177] L. Moneta et al., "The RooStats Project", *PoS* **ACAT2010** (2010) 057, arXiv:1009.1003.

[178] "SWGuideNonStandardCombineUses < CMS/HiggsWG < TWiki". https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsWG/SWGuideNonStandardCombineUses#Rate_Parameters. (Accessed on 05/12/2016).

[179] "ROOT: TGraph2D Class Reference". https://root.cern.ch/doc/master/classTGraph2D.html. (Accessed on 05/13/2016).

[180] P. Collier and G. Papotti, "LHC Performance Workshop, Summary Session 5, 2016 & Run 2". https://indico.cern.ch/event/486336/contributions/2006459/attachments/1237723/1818251/160303-Chamonix-Sumary-session5.pdf. (Accessed on 05/13/2016).

[181] "PhysicsResultsSUS < CMSPublic < TWiki". https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS. (Accessed on 05/13/2016).

[182] "SupersymmetryPublicResults < AtlasPublic < TWiki". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults. (Accessed on 05/13/2016).

[183] CMS Collaboration, "Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV", *Phys. Lett. B* (2016) doi:10.1016/j.physletb.2016.05.002, arXiv:1602.06581.

[184] CMS Collaboration, "Search for supersymmetry in pp collisions at sqrt(s) = 13 TeV in the single-lepton final state using the sum of masses of large-radius jets", arXiv:1605.04608.

[185] ATLAS Collaboration, "Search for supersymmetry at $\sqrt{s} = 13$ TeV in final states with jets and two same-sign leptons or three leptons with the ATLAS detector", *Eur. Phys. J.* **C76** (2016), no. 5, 259, doi:10.1140/epjc/s10052-016-4095-8, arXiv:1602.09058.

[186] ATLAS Collaboration, "Search for pair production of gluinos decaying via stop and sbottom in events with b-jets and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", arXiv:1605.09318.

[187] P. Bechtle et al., "Killing the cMSSM softly", *Eur. Phys. J.* **C76** (2016), no. 2, 96, doi:10.1140/epjc/s10052-015-3864-0, arXiv:1508.05951.

[188] Virgo, LIGO Scientific Collaboration, "Observation of Gravitational Waves from a Binary Black Hole Merger", *Phys. Rev. Lett.* **116** (2016), no. 6, 061102, doi:10.1103/PhysRevLett.116.061102, arXiv:1602.03837.

[189] ATLAS Collaboration, "Search for resonances decaying to photon pairs in 3.2 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector", Technical Report ATLAS-CONF-2015-081, CERN, Geneva, Dec, 2015.

[190] CMS Collaboration, "Search for new physics in high mass diphoton events in proton-proton collisions at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-EXO-15-004, CERN, Geneva, 2015.

Eidesstattliche Versicherung

Hiermit erkläe ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Declaration on oath

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den 2 Juni 2016 Artur Lobanov

Acknowledgements

At this point, I would like to thank everyone who supported me in any way during my time in and around Hamburg, and made this years such a fruitful and unforgettable experience!

First of all, my gratitude goes to my supervisor Dr. Isabell Melzer-Pellmann, who introduced me to the world of LHC physics and lead the various projects I worked on within the SUSY Group. Her management and guidance were invaluable for the success of my PhD project.

I would also like to thank Prof. Dr. Erika Garutti for being my co-supervisor and also "secret" mentor. I've always enjoyed our discussions very much, and I am grateful for her advice and support.

I am also thankful to Isabell and Erika for promptly reviewing my thesis and also joining my disputation committee together with Prof. Dr. Dieter Horns, Prof. Dr. Johannes Haller and Prof. Dr. Gudrid Moortgat-Pick, whom I thank for dedicating their time just before the summer break.

To Prof. Dr. Kerstin Borras goes my sincere gratitude for bringing me to DESY and supporting me in any matter throughout all these years.

I would like to specially thank Dr. Benjamin Lutz for supervising my activities within the HO upgrade project. Working with Beni was always very instructive and efficient, but it also made lots of fun.

Furthermore, I would like to thank the fellow HCAL colleagues and CMS technical coordination for the truly inspiring time I spent working at *Point 5* during LS1.

Many thanks to the current and former members of the SUSY Group at DESY – Alexis, Altan, Claudia, Dirk, Elias, Francesco, Hannes and Özgür – for the helpful and enjoyable environment in the group. To Claudia I express my gratitude for the guidance and support in every step of the analysis.

Apart from work, many colleagues turned into friends, and so I would like to thank them for sharing joyful moments at work and beyond.

Thank you, Francesco and Özgür, for making office life so much fun and exploring Hamburg's greatest drinking and eating places (including their flats).

Thanks to Ali, Engin, Jasone, Juan, Paolo, Simon and Tuğba for the joyful atmosphere in the CMS Group (corridors).

I would also like to thank Carsten and Oskar for restoring my German self, and the Russian-speaking photon scientists – Aram, Dima, Dina, Masha, Natasha, Oleg, Sasha, Sergey, Vanya, Svitozar – for taking over the Russian stuff.

Finally, I would like to thank my relatives all over the globe for their support and care.

My deepest gratitude goes to Ksu.