

ERSTE MESSUNG DES WIRKUNGSQUERSCHNITTS
DER TOP-QUARK-PAARPRODUKTION
BEI $\sqrt{s} = 7$ TEV
IM ELEKTRON+JETS-KANAL MIT DEM
CMS-EXPERIMENT

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Phys. Manuel Renz
aus Karlsruhe

Tag der mündlichen Prüfung: 04.07.2011

Referent: Prof. Dr. Th. Müller, Institut für Experimentelle Kernphysik

Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

*Dem Licht der Sonne folgend,
verließen wir die „Alte Welt“
(Christoph Kolumbus)*

Zusammenfassung

Am 3. August 1492 brach Christoph Kolumbus zu einer Entdeckungsreise in unbekannte Gewässer auf. Westwärts über den Atlantischen Ozean segelnd, wollte er einen neuen Seeweg nach Indien erschließen. Am Ende dieser Reise stand jedoch die Entdeckung eines bislang unbekanntes Kontinents, welche die Menschheitsgeschichte in den folgenden Jahrhunderten entscheidend prägen sollte.

Mancher Physiker mag sich vielleicht wie Christoph Kolumbus am Beginn seiner Reise gefühlt haben, als der Large Hadron Collider (LHC) am 30. März 2010 in Betrieb genommen wurde. Mit einer Schwerpunktsenergie von 7 TeV stellt der LHC Teilchenkollisionen bei einer Energie zur Verfügung, die bisher in keinem Beschleunigerexperiment erreicht werden konnte. Die ersten Proton-Proton-Kollisionen können daher als der Beginn einer neuen Ära im Forschungsbereich der Teilchenphysik angesehen werden. Am Anfang dieser Ära steht die Untersuchung des Standardmodells der Teilchenphysik (SM) bei bisher unerreichten Energien. In ihrer Blütezeit könnte die Entdeckung des im SM vorhergesagten Higgs-Bosons und die Entdeckung neuer physikalischer Phänomene, die nicht innerhalb der Theorie des SM erklärbar sind, folgen.

Das Standardmodell der Teilchenphysik beschreibt zwölf fundamentale Fermionen sowie deren Wechselwirkungen und basiert auf einer relativistischen Quantenfeldtheorie. Die Fermionen können in die Familien der Quarks und Leptonen unterteilt werden. Jede dieser Familien besteht weiterhin aus drei Generationen mit jeweils zwei Teilchen. Gewöhnliche Materie besteht ausschließlich aus Quarks und Leptonen der ersten Generation, die Teilchen der höheren Generationen können nur in hochenergetischen Teilchenkollisionen erzeugt werden. Das SM umfasst drei Wechselwirkungen, welche durch Austauschteilchen, die sogenannten Eichbosonen, übertragen werden. Die elektromagnetische Kraft wirkt auf Teilchen mit elektrischer Ladung und wird durch Photonen übertragen. Die Austauschteilchen der starken Kraft, die ausschließlich auf Teilchen mit Farbladung wirkt, sind acht verschiedene Gluonen. Teilchen mit schwachem Isospin unterliegen der schwachen Wechselwirkung, welche durch geladene W^\pm -Bosonen und die elektrisch neutralen Z -Bosonen übertragen wird. In der Quantenfeldtheorie sind alle Eichbosonen masselos, experimentelle Resultate zeigten jedoch, dass W -Bosonen und Z -Bosonen ca. 80 bzw. 90 mal so schwer sind wie Protonen [1–4]. Ein Prozess, Higgsmechanismus [5–7] genannt, kann diesen Widerspruch auflösen und die Massen dieser Bosonen innerhalb der Theorie des SM erklären. Allerdings wird durch diesen Prozess die Existenz eines weiteren, bisher unentdeckten Teilchens, des

sogenannten Higgs-Bosons vorhergesagt. Die Suche nach diesem noch unentdeckten Teilchen ist einer der Hauptgründe für den Bau von Beschleunigern wie dem LHC. Das schwerste der bislang entdeckten Quarks ist das Top-Quark. Es hat eine Masse von $173.3 \text{ GeV}/c^2$ [8] und wurde 1995 zum ersten Mal am Tevatron-Beschleuniger nahe Chicago nachgewiesen [9, 10]. Im SM werden Top-Quarks hauptsächlich als Paare von Top- und Antitop-Quarks über die starke Wechselwirkung produziert. Die Produktion einzelner Top-Quarks durch die schwache Wechselwirkung ist jedoch auch möglich. Theoretischen Vorhersagen, basierend auf MCFM [11] und HATHOR [12] zufolge, beträgt der Wirkungsquerschnitt für Top-Quark-Paarproduktion am LHC $157.5_{-24.4}^{+23.2} \text{ pb}$ in nächst führender Ordnung Störungstheorie (NLO) bzw. $163_{-10}^{+11} \text{ pb}$ in „approximate NNLO“. Aufgrund ihrer hohen Masse zerfallen Top-Quarks nahezu instantan nach ihrer Erzeugung in ein W -Boson und ein Bottom-Quark. Je nach Zerfall der W -Bosonen, lässt sich der Zerfall eines Top-Quark-Paares als voll-hadronisch (beide W -Bosonen zerfallen in zwei Quarks), di-leptonisch (beide W -Bosonen zerfallen in ein Lepton-Neutrino-Paar) oder semi-leptonisch (ein W -Boson zerfällt in Quarks, das andere in ein Lepton-Neutrino-Paar) kategorisieren. Der letztgenannte Zerfallskanal wird oft auch als Lepton+Jets Kanal bezeichnet. In der vorliegenden Arbeit wird der Wirkungsquerschnitt für Top-Quark-Paarproduktion bei einer Schwerpunktsenergie von 7 TeV im Elektron+Jets Kanal gemessen. Diese Messung ist eine wichtige Überprüfung der Vorhersagen des SM in einem neuen Energiebereich. Des Weiteren bildet die präzise Kenntnis des Wirkungsquerschnitts die Grundlage für interessante Messungen im Top-Quark-Sektor. Da Top-Quark-Paarproduktion der dominierende Untergrundprozess in Suchen nach bisher unbekanntem physikalischen Prozessen darstellt, ist die Messung des Wirkungsquerschnitts weiterhin ein wichtiger Bestandteil dieser Analysen.

Der LHC der Europäischen Organisation für Kernphysik (CERN) in Genf nahe der schweizerisch-französischen Grenze, ist ein 26,7 km langer Ringbeschleuniger in dem Protonenstrahlen auf eine Energie von 3.5 TeV beschleunigt und schließlich zur Kollision gebracht werden. In einer Tiefe von 45 bis 170 Metern unter der Erdoberfläche liegend benutzt der LHC den bereits für seinen Vorgängerbeschleuniger LEP gegrabenen Tunnel. Das Herz des LHC sind 1232 supraleitende Dipolmagneten mit einer Länge von 15 m, die ein starkes Magnetfeld erzeugen und so die Protonenstrahlen auf ihren Kreisbahnen halten. Bevor die Protonenstrahlen im LHC zur Kollision gebracht werden können, müssen die Protonen zunächst durch die Ionisation von Wasserstoffatomen gewonnen werden. Hierauf durchlaufen sie diverse lineare und kreisförmige Vorbeschleuniger bevor sie mit einer Energie von 450 GeV in den LHC geschossen werden. Für die Beschleunigung der Protonenstrahlen kommen alternierende elektrische Felder in geeigneten Hohlraumresonatoren zum Einsatz. Da nur eine Phase dieses elektrischen Feldes die Protonen beschleunigt, befindet sich im LHC kein kontinuierlicher Protonenstrahl, sondern einzelne Pakete mit jeweils $1 \cdot 10^{11}$ Protonen. Im nominellen Betrieb kann der Beschleunigerring bis zu 2808 Pakete mit einem zeitliche Abstand von 25 ns enthalten. Ein Maß für die Kollisionsrate in einem Beschleuniger ist die sogenannte Luminosität. Im Jahre 2010 konnte der LHC eine maximale Luminosität

von $2 \cdot 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ erreichen. Die integrierte Luminosität, d.h. das zeitliche Integral über die Luminosität, ist ein Maß für die Größe eines aufgezeichneten Datensatzes. Während des Betriebes des LHC im Jahr 2010 wurde eine integrierte Luminosität von 47 pb^{-1} ($1 \text{ b} = 1 \cdot 10^{-24} \text{ cm}^2$) den beiden Vielzweckdetektoren ATLAS und CMS zur Verfügung gestellt.

Um die in Proton-Proton-Kollisionen erzeugten Teilchen nachzuweisen, kommen große Teilchendetektoren zum Einsatz. Entlang des LHC-Rings sind vier dieser Detektoren an den Kollisionpunkten der beiden Protonenstrahlen aufgebaut. ALICE und LHCb sind Detektoren, die auf die Analyse bestimmter physikalischer Prozesse spezialisiert sind, ATLAS und CMS sind hingegen Vielzweckdetektoren, die ein breites Spektrum physikalischer Prozesse untersuchen können. Die Daten, die in der vorliegenden Arbeit analysiert werden, wurden mit dem Compact Muon Solenoid (CMS) Detektor aufgezeichnet. CMS ist ein 12500 Tonnen schwerer und 21,6 m langer Detektor mit einem Durchmesser von 14,6 m. Er befindet sich in einer Kaverne ca. 100 m unter der Erdoberfläche. Hermetisch und symmetrisch um den Kollisionspunkt aufgebaut, folgt CMS dem typischen, zwiebelschalenförmigen Design eines großen Teilchendetektors. In seinem Innern befindet sich der weltgrößte auf Silizium basierende Spurdetektor in dem elektrisch geladene Teilchen entlang ihrer Spur durch Ionisation Elektron-Loch-Paare erzeugen. Der Spurdetektor ist von einem aus Bleiwolframatkristallen (PbWO_4) bestehenden, homogenen elektromagnetischen Kalorimeter (ECAL) umgeben, in dem primär elektromagnetisch wechselwirkende Teilchen wie Elektronen und Photonen absorbiert werden. Durch das dabei entstehende Szintillationslicht kann ebenfalls die Energie der Teilchen gemessen werden. Außerhalb des ECALs befindet sich das hadronische Kalorimeter (HCAL). Hier wechseln sich aus Messing bestehende Absorberschichten, in denen durch inelastische Kernreaktionen wechselwirkende Hadronen absorbiert werden, mit aktiven Szintillatorschichten ab. Die bisher beschriebenen Subdetektoren befinden sich innerhalb eines supraleitenden Solenoidmagneten der ein Magnetfeld von 3,8 T erzeugt. Elektrisch geladene Teilchen bewegen sich daher auf gekrümmten Bahnen aus deren Radius der Teilchenimpuls bestimmt werden kann. Außerhalb des Solenoiden befindet sich das Eisenjoch, das für die Rückführung des magnetischen Flusses sorgt. In Hohlräumen innerhalb des Jochs sind die sogenannten Myonkammern platziert, welche dem Nachweis von Myonen dienen, die im Allgemeinen die inneren Detektorschichten durchqueren ohne absorbiert zu werden. Im nominellen Betrieb des LHC finden Proton-Proton-Kollisionen mit einer Rate von 40 MHz statt. Diese Rate ist jedoch sehr viel größer als die Rate, mit der die Detektorsignale prozessiert und schlussendlich gespeichert werden können. CMS hat daher ein zweistufiges Triggersystem entwickelt, welches nur interessante Ereignisse passieren können. In einem ersten Schritt wird dazu die Rate mit Hilfe des aus einer geeigneten Elektronik bestehenden Level-1 Trigger auf ca. 100 kHz reduziert, bevor ein auf Software basierender High Level Trigger (HLT) die Ereignisrate auf prozessierbare 100 Hz reduziert.

Während einer typischen Proton-Proton-Kollision entstehen hunderte von neuen Teilchen. Um diese mit einem interessanten physikalischen Prozess in Verbindung brin-

gen zu können, ist ein tiefes Verständnis der zugrundeliegenden Physik notwendig. Protonen sind zusammengesetzte Objekte und daher wechselwirken bei hochenergetischen Kollisionen nicht die Protonen als Ganzes sondern nur ihre Konstituenten (Partonen), d.h. Quarks und Gluonen miteinander. Während die Streuung von Partonen bei hohen Impulsüberträgen theoretisch gut berechenbar ist, kann das Verhalten der Protonreste sowie die Hadronisierung der in den Kollisionen entstandenen Quarks und Gluonen nur durch phänomenologische Modelle beschrieben werden. Physiker setzen daher Simulationsprogramme ein, die auf theoretischen Rechnungen und phänomenologischen Modellen basieren, um eine Vorhersage über die Art, die Anzahl und die Kinematik der in den Kollisionen entstandenen Teilchen zu bekommen. Der quantenmechanischen Natur der Prozesse wird dabei durch sogenannte Monte Carlo Techniken Rechnung getragen. In der Detektorsimulation wird schließlich die Wechselwirkung der produzierten Teilchen mit dem Detektormaterial simuliert. Neben diesen Energiedepositionen in den verschiedenen Subdetektoren, werden auch die Ausleseelektroniken simuliert, was zu Detektorsignalen vergleichbar mit denen echter Kollisionen führt.

Um aus Detektorsignalen auf bestimmte Teilchen rückschließen zu können, kommen Rekonstruktionsalgorithmen zum Einsatz. Im Spurdetektor werden die Bahnen geladener Teilchen rekonstruiert, aus denen wiederum der Ort der primären Partonstreuung bestimmt werden kann. Elektronen lassen sich unter Kombination von Informationen aus dem Spurdetektor und dem ECAL rekonstruieren. Ebenso kann die Flugbahn und der Impuls von Myonen mit Hilfe der Energiedepositionen in den Myonkammern und Informationen aus dem Spurdetektor gemessen werden. Da Quarks und Gluonen dem „Confinement“ der starken Wechselwirkung unterliegen, können sie nicht als freie Objekte, sondern nur als Bündel kollimierter Hadronen im Detektor nachgewiesen werden. Die Energie und die Richtung dieser sogenannten Jets ist im Allgemeinen ähnlich der Energie und Flugrichtung des ursprünglichen Quarks oder Gluons. Eine wichtige Voraussetzung dafür ist, dass alle Energiedepositionen dem richtigen Jet zugewiesen werden, wobei spezielle Jet-Clustering-Algorithmen zum Einsatz kommen.

Aufgrund des relativ kleinen Wirkungsquerschnitts für Top-Quark-Paarproduktion entsteht ca. nur in jeder milliardsten Proton-Proton-Kollision ein Top-Quark-Paar. Um trotzdem den Wirkungsquerschnitt messen zu können, wird eine Ereignisselektion basierend auf der Charakteristik von Top-Quark-Ereignissen durchgeführt, um somit deren Anteil im selektierten Datensatz zu erhöhen. Wie bereits beschrieben, wird für die Messung der Elektron+Jets Kanal verwendet. Die Signatur selektierter Ereignisse ist daher ein hochenergetisches, isoliertes Elektron, vier hochenergetische Jets und fehlende Transversalenergie (\cancel{E}_T). Obwohl diese Ereignissignatur sehr charakteristisch für den Zerfall eines Top-Quark-Paares ist, gibt es dennoch andere physikalische Prozesse, die eine ähnliche Signatur im Detektor aufweisen. Dies ist zum Einen die Produktion von W - und Z -Bosonen in Verbindung mit zusätzlichen Jets aus der Abstrahlung von Quarks und Gluonen. Zum Anderen aber auch Multijet-Ereignisse in denen es zu semi-leptonischen Zerfällen von B - oder C -Hadronen kommt. Einen kleinen Beitrag liefert auch die Produktion einzelner, leptonisch zerfallender Top-Quarks. Zerfälle von π^0 -Mesonen aus Jets in zwei Photonen, die ihrerseits Konversionselektronen erzeugen,

sind eine weitere Quelle von Untergrundereignissen. Um Untergrundereignisse effizient zu unterdrücken und dabei möglichst wenig Signalereignisse zu verlieren, wird unter der Verwendung simulierter Daten eine Ereignisselektion entwickelt. Die hierfür notwendigen Simulationen wurden mit MADGRAPH/MADEVENT [13] in Verbindung mit PYTHIA [14] durchgeführt. Die entwickelte Ereignisselektion fordert genau ein isoliertes Elektron mit einer Transversalenergie größer als 30 GeV, welches weitere Qualitätskriterien erfüllt. Zusätzlich darf sich kein isoliertes Myon im Ereignis befinden um statistisch unabhängig von einer ähnlichen Analyse im Myon+Jets Kanal zu sein. Ebenfalls darf die invariante Masse des selektierten Elektrons und eines weiteren, schwächere Kriterien erfüllenden Elektrons nicht in einem schmalen Fenster um die Z -Boson-Masse liegen, um diesen speziellen Untergrund zu unterdrücken. Schließlich muss das selektierte Elektron zwei Algorithmen, die auf die Identifikation von Konversionselektronen ausgelegt sind, erfolgreich passieren. Weiterhin wird eine effiziente Untergrundunterdrückung erreicht, indem ein großer Transversalimpuls von mehr als 30 GeV für die begleitenden Jets gefordert wird. Aufgrund der Beobachtung, dass durch diese strenge Forderung Top-Quark-Paarereignisse auch nur drei selektierte Jets aufweisen können und um im weiteren Verlauf der Analyse eine Abschätzung des Untergrundes erhalten zu können werden Ereignisse mit mindestens drei Jets ausgewählt.

Durch unvollkommene Simulationen können die Effizienzen der angesprochenen Selektionskriterien in Simulation und realen Daten unterschiedlich sein. Da die vorhergesagte Anzahl an Top-Quark-Paarereignisse aber die Grundlage für die Messung des Wirkungsquerschnitts bildet, muss die Simulation auf mögliche Unterschiede korrigiert werden. Zu diesem Zweck wird die Effizienz der oben beschriebenen Selektionskriterien mit Hilfe der sogenannten „Tag-and-Probe Methode“ in der Simulation und in realen Daten bestimmt. Hierfür wird die wohlbekannte Masse der Z -Boson Resonanz ausgenutzt und die notwendige Untergrundabschätzung mit Hilfe gleichgeladener Elektronenpaare erzielt. Unterschiedliche Effizienzen der einzelnen Selektionschritte in Simulation und realen Daten führen zu einem Gesamtkorrekturfaktor von $0,933 \pm 0,025$, welcher durch eine Umgewichtung der simulierten Ereignisse berücksichtigt wird.

Nach dieser Korrektur kann die von der Simulation vorhergesagte Anzahl an Ereignissen, die alle Selektionskriterien erfüllen, mit der Anzahl selektierter Datenereignisse im untersuchten Datensatz von $35,9 \text{ pb}^{-1}$ verglichen werden. Hierbei fällt auf, dass von der Simulation zu wenig Ereignisse vorhergesagt werden. Da während der Simulation allerdings verschiedene Annahmen gemacht wurden, die mitunter einen großen Einfluss auf die Anzahl vorhergesagter Ereignisse haben können, kann eine solche Differenz erklärt werden. Es kann allerdings gezeigt werden, dass die Simulation in der Lage ist die Form verschiedener kinematischer Verteilungen zu beschreiben. Zusammenfassend kann die Anzahl an Top-Quark-Paarereignissen in Daten nicht durch simple Subtraktion der vorhergesagten Untergrundereignisse bestimmt werden sondern die relativen Anteile der verschiedenen Prozesse am selektierten Datensatz müssen in geeigneter Weise bestimmt werden.

In der aus Simulationen bestimmten Anzahl an erwarteten Ereignissen sind bereits

die limitierte Detektorakzeptanz und die begrenzte Selektionseffizienz berücksichtigt. Daher ist das Verhältnis aus gemessener und vorhergesagter Anzahl an Top-Quark-Paarereignissen gleich dem Verhältnis aus dem gemessenen und dem theoretisch vorhergesagten Wirkungsquerschnitt. Die relativen Anteile der verschiedenen Prozesse am selektierten Datensatz wird durch einen simultanen Maximum-Likelihood-Fit an die Verteilungen der fehlenden Transversalenergie für Ereignisse mit genau drei Jets und die M3 Verteilung für Ereignisse mit vier oder mehr Jets bestimmt. M3 ist die invariante Masse der drei Jets, die vektoriell addiert den höchsten Transversalimpuls aufweisen. Diese Variable ist somit eine einfache Rekonstruktion der Masse des hadronisch zerfallenden Top-Quarks. Der selektierte Datensatz kann in zwei Untermengen unterteilt werden, Ereignisse mit genau drei Jets und Ereignisse mit vier oder mehr Jets. Da der Datensatz mit Ereignissen mit genau drei Jets hauptsächlich Untergründereignisse enthält, kann hier eine Bestimmung der relativen Anteile der Untergrundprozesse durchgeführt werden. Mit Hilfe des vorhergesagten Verhältnisses der Ereignisse in den beiden Untermengen und dem mit Signal angereicherten Datensatz mit vier und mehr Jets, wird eine möglichst genaue Bestimmung des relativen Anteils von Top-Quark-Paarereignissen erzielt.

Die Simulation von Multijet-Ereignissen mit isolierten Elektronen im Endzustand ist äußerst schwierig. Die für den Fit benötigten Formen von \cancel{E}_T und M3 für diese Prozesse werden daher aus einem Seitenband direkt aus Daten bestimmt. Hierfür werden ausdrücklich Ereignisse mit schlecht isolierten Elektronen, die nur schwache Qualitätskriterien erfüllen, selektiert. Basierend auf Simulationen kann überprüft werden, dass dieses Seitenband von Multijet Ereignissen dominiert ist und dass durch die Veränderungen der Selektion keine entscheidenden Verformungen der \cancel{E}_T und M3 Verteilungen entstehen.

Zur Bestimmung der Unsicherheiten auf die Messung des Wirkungsquerschnitts der Top-Quark-Paarerezeugung werden Ensembletests mit jeweils 50000 Pseudoexperimenten durchgeführt. Unter der Verwendung von Monte-Carlo-Techniken wird dabei für jedes Pseudoexperiment eine Pseudodatenverteilung für \cancel{E}_T und M3 entsprechend der Vorhersagen der Simulation erstellt. Mit Hilfe des Maximum-Likelihood-Fits wird dann die Anzahl von Top-Quark-Paarereignissen im jeweiligen Pseudoexperiment bestimmt. Aus der Verteilung dieser Werte für alle Pseudoexperimente eines Ensembles kann schließlich mit Hilfe des Medians und der 2.5%, 16%, 84% und 97.5% Quantile der Zentralwert und die zentralen 68% und 95% Vertrauensintervalle bestimmt werden. Aus Ensembletests mit verschiedenen angenommenen Wirkungsquerschnitten wird eine sogenannte Neyman-Konstruktion gebildet aus der schlussendlich die Unsicherheiten auf einen bestimmten Messwert abgelesen werden können.

Verschieden Quellen systematischer Unsicherheiten können das Messergebnis beeinflussen. Zum Einen verändern systematische Effekte die Anzahl an vorhergesagten Ereignissen, zum Anderen können sie zur Verformung der \cancel{E}_T und M3 Verteilungen führen. In der Analyse werden die Einflüsse von zwölf systematischen Unsicherheiten, wie die Unsicherheit auf die Jet-Energie-Korrekturfaktoren oder die Unsicherheit auf die Modellierung des Multijet-Untergrundes, untersucht. Diese Unsicherheiten wer-

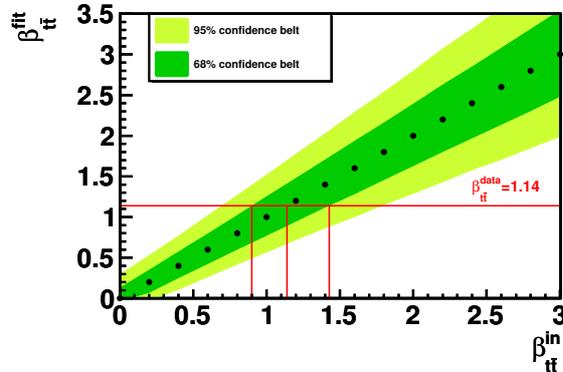


Abbildung 1: Neyman-Konstruktion zur Bestimmung der Unsicherheiten auf den gemessenen Wirkungsquerschnitt der Top-Quark-Paarzeugung. $\beta_{t\bar{t}}^{\text{in}}$ gibt dabei das Verhältnis aus dem in den Pseudodaten verwendeten und dem theoretisch vorhergesagten Wirkungsquerschnitt der Top-Quark-Paarzeugung an. Die Zentralwerte und Vertrauensintervalle der verschiedenen Ensemble werden durch $\beta_{t\bar{t}}^{\text{fit}}$ repräsentiert. $\beta_{t\bar{t}}^{\text{data}}$ gibt hingegen den im Maximum-Likelihood-Fit an Daten bestimmten Wert an.

den in den Pseudodatenverteilungen der verschiedenen Ensembles berücksichtigt, was hauptsächlich zu einer Verbreiterung der zentralen 68% und 95% Vertrauensbänder in der Neyman-Konstruktion führt. Wie in Abbildung 1 gezeigt, wird durch Anwendung des Maximum-Likelihood-Fits auf reale Daten der Messwert bestimmt und als horizontale Linie in die Neyman-Konstruktion eingezeichnet. Aus den Schnittpunkten dieser Linie mit dem 68% Vertrauensband wird schließlich die Unsicherheit auf den Messwert bestimmt. Die Messung im Elektron+Jets Kanal ergibt

$$\sigma_{t\bar{t}} = 180_{-38}^{+45} (\text{stat.} + \text{syst.}) \pm 7 (\text{lumi.}) \text{ pb},$$

was in sehr guter Übereinstimmung mit der theoretischen Vorhersage ist. Der rein statistische Fehler beträgt hierbei $_{-22}^{+23}$ pb. Es zeigt sich somit, dass die Analyse bereits durch systematische Unsicherheiten limitiert ist, wobei die Unsicherheit auf die Jet-Energie-Korrekturfaktoren bei weitem für die größte Unsicherheit verantwortlich ist. Die erwähnte Diskrepanz zwischen der Anzahl an vorhergesagten und beobachteten Ereignissen wird im Fit hauptsächlich durch einen bis zu 40% größeren Anteil von Untergrundereignissen mit W -Bosonen kompensiert. Die Fitresultate können benutzt werden um die Vorhersage auf die in Daten gemessenen Anteile der verschiedenen Prozesse zu skalieren, um somit die Fitresultate mit den Daten vergleichen zu können. Ein solcher Vergleich ist in Abbildung 2 für \cancel{E}_T und M3 gezeigt.

Neben der Messung im Elektron+Jets-Kanal wird auch eine Messung im kombinierten Elektron+Jets und Myon+Jets-Datensatz durchgeführt. Dabei werden zusätzlich die Ereignisse, welche in der in [15] beschriebenen Analyse selektiert wurden, verwendet und ein simultaner Fit an vier Variablen, \cancel{E}_T und M3 separat für Elektron+Jets- und Myon+Jets-Ereignisse, durchgeführt. Da beide Analysen von systematischen Unsicherheiten limitiert sind, kann nur eine kleine Verbesserung der statistischen Unsicherheit

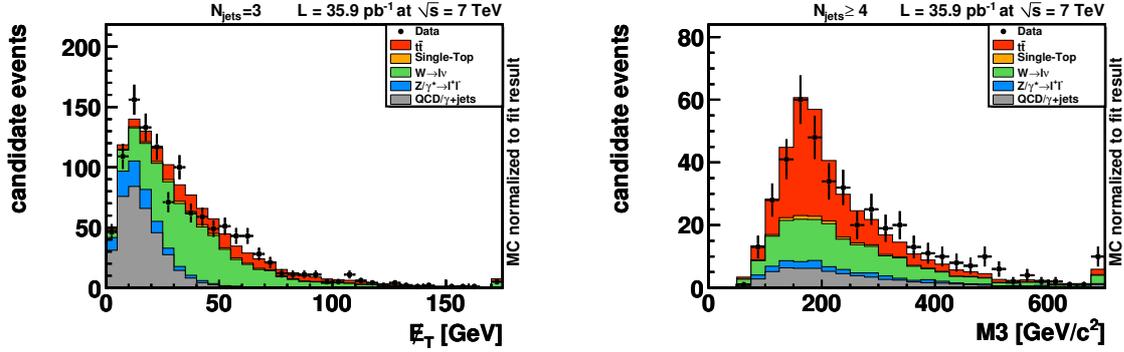


Abbildung 2: Vergleich der auf das Fitresultat normierten Simulation und der in Daten beobachteten Verteilung von \cancel{E}_T (links) und $M3$ (rechts).

erzielt werden. Der gemessene Wirkungsquerschnitts der Top-Quark-Paarzeugung ist

$$\sigma_{t\bar{t}} = 173_{-32}^{+39} \text{ (stat. + syst.)} \pm 7 \text{ (lumi.) pb},$$

was ebenfalls in sehr guter Übereinstimmung mit den theoretischen Vorhersagen ist.

Schon mit einem Datensatz von ungefähr 36 pb^{-1} kann der Wirkungsquerschnitt für Top-Quark-Paarzeugung bei einer Schwerpunktsenergie von 7 TeV gemessen werden. Bis zum jetzigen Zeitpunkt sind alle Messungen dieser Größe mit den Vorhersagen des SM verträglich. Im Jahr 2011 hat das CMS Experiment bereits einen Datensatz von über 500 pb^{-1} aufgezeichnet. Bis zum Ende dieses Jahres soll der Datensatz Schätzungen zufolge sogar bis auf $2\text{-}3 \text{ fb}^{-1}$ anwachsen. Dies impliziert, dass Ende 2011 erste Anzeichen des Higgs-Bosons oder neuartiger physikalischer Prozesse gefunden werden können. Im Top-Quark-Sektor wird hauptsächlich die Messung der Ladungsasymmetrie und die Suche nach schweren, in Top-Quark-Paare zerfallende, Resonanzen von höchstem Interesse sein.

FIRST MEASUREMENT OF THE TOP-QUARK
PAIR PRODUCTION CROSS-SECTION
AT $\sqrt{s} = 7$ TEV
IN THE ELECTRON+JETS CHANNEL
WITH THE CMS-EXPERIMENT

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Phys. Manuel Renz
aus Karlsruhe

Tag der mündlichen Prüfung: 04.07.2011

Referent: Prof. Dr. Th. Müller, Institut für Experimentelle Kernphysik

Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

*Following the light of the
sun, we left the Old World*
(Christopher Columbus)

Introduction

On August 3rd, 1492 Christopher Columbus started his journey searching for a new route to India while sailing westwards across the Atlantic ocean — a journey that finally led to the discovery of a thitherto unknown continent and heavily affected the history of mankind during the following centuries. Some physicists might have felt just like Columbus when the Large Hadron Collider (LHC) of the European Organization for Nuclear Research (CERN) started its operation on March 30th, 2010 at a center-of-mass energy of 7 TeV, a collision energy so far unrivaled by any laboratory based collider experiment. With these first collisions a new era in the research field of particle physics has begun. At its beginning, the predictions of the Standard Model (SM) of particle physics are tested at a new energy regime. During its heyday, the discovery of the yet unobserved Higgs boson as well as the discovery of phenomena which cannot be explained within the framework of the SM may follow.

The LHC is a 26.7 km long circular accelerator, located near Geneva, Switzerland, roughly 100 m below the surface. Two proton beams moving in opposite directions are accelerated to 3.5 TeV and brought to collision in the center of four large particle detectors. The collision data analyzed in this thesis has been collected with the Compact Muon Solenoid (CMS) detector. CMS is a multi-purpose particle detector, hermetically built around the collision point of the two proton beams. It is constructed with typical onion-like design of collider detectors: a silicon-based tracking system surrounded by a homogeneous electromagnetic calorimeter and a sampling hadronic calorimeter. These subdetectors are enclosed by a superconducting solenoid providing a 3.8 T magnetic field which is necessary to bend the trajectories of charged particles and thus allows a precise measurement of their momenta. The solenoid is surrounded by an iron yoke, responsible for returning the magnetic flux, as well as by the muon chambers used to identify muons traversing all of the inner detector regions without being absorbed.

The SM is a relativistic quantum field theory describing twelve fundamental fermions and the interactions between them. These fermions can be grouped into six quarks and six leptons which can be further divided into three generations. All ordinary matter is made of fermions from the first lepton and quark generations while the fermions of the remaining generations can solely be produced in highly energetic particle collisions. Moreover, there are four forces acting between the fundamental particles. Besides gravitation which is not part of the SM, these are the electromagnetic, the strong and weak

interaction. Designed in the late 1960s and early 1970s, the SM and the predictions deduced from it have undergone a variety of experimental tests. So far, almost all predictions have been found to be fulfilled.

With a mass of $173.3 \text{ GeV}/c^2$ [8], the top quark is the heaviest particle among the fundamental fermions. It has been discovered in 1995 by the CDF [9] and DØ [10] experiments at Fermilab's proton anti-proton collider Tevatron near Chicago, USA. At the Tevatron and the LHC, top quarks are mainly produced pairwise via the strong interaction. The measurement of the top-quark pair production cross-section at the LHC is one of the first important steps when testing the predictions of the SM at the new energy regime. Moreover, a precise knowledge of the $t\bar{t}$ production cross-section is the basis for interesting measurements in the top-quark sector as well as for searches for physics which cannot be explained in the context of the SM.

In this thesis, the first measurement of the top-quark pair production cross-section at $\sqrt{s} = 7 \text{ TeV}$ using events with an electron+jets final state is described. The production of top-quark pairs is distinguished from other physics processes by solely employing the kinematic properties of $t\bar{t}$ events. The analyzed dataset has been recorded with the CMS detector during the 2010 operation of the LHC and corresponds to an integrated luminosity of 35.9 pb^{-1} .

Contents

1	Theoretical Principles	1
1.1	The Standard Model of Particle Physics	1
1.1.1	Qualitative Introduction	1
1.1.2	Mathematical Formulation	5
1.1.3	Feynman Diagrams and Cross Sections	12
1.2	Top-Quark Physics	15
1.2.1	Top-Quark Production	16
1.2.2	Top-Quark Decay	20
2	Experimental Setup	23
2.1	The Large Hadron Collider	24
2.1.1	Production of Proton Beams and Preacceleration	24
2.1.2	The LHC Main Ring	26
2.2	The Compact Muon Solenoid Detector	29
2.2.1	The Inner Tracking System	33
2.2.2	The Electromagnetic Calorimeter	35
2.2.3	Hadron Calorimeter and Solenoid	36
2.2.4	The Muon System	38
2.2.5	Trigger, Data Acquisition, and Computing Model	40
3	Generation, Simulation, and Reconstruction	43
3.1	Event Generation	44
3.1.1	MC Event Generators	47
3.2	Detector Simulation	49
3.3	Event Reconstruction	50
3.3.1	Track Reconstruction	50
3.3.2	Primary-Vertex Reconstruction	51
3.3.3	Electron Reconstruction	53
3.3.4	Reconstruction of Muons	56
3.3.5	Jet Reconstruction	57
3.3.6	Missing Transverse Energy	63

4	Selection of $t\bar{t}$ Candidates	65
4.1	Signal and Background Events	66
4.1.1	Background Processes in the Electron+Jets Final State	67
4.1.2	Signal and Background Modelling	70
4.2	Event Selection	75
4.2.1	Analyzed Collision Data	75
4.2.2	Selection Criteria	78
4.3	Electron Selection and Trigger Efficiency	88
4.3.1	Electron Selection-Efficiency	90
4.3.2	Electron Trigger-Efficiency	99
4.4	Comparison of Simulation and Collision Data	100
5	Measurement of the $t\bar{t}$ Cross-Section	105
5.1	Analysis Strategy	106
5.1.1	Modelling of the QCD Multi-Jet Background	106
5.1.2	Discriminating Variables	108
5.2	Statistical Method	110
5.2.1	Binned Likelihood Template Fit	111
5.2.2	Ensemble Tests	112
5.2.3	Neyman Construction	113
5.3	Measurement in the Electron+Jets Final State	114
5.3.1	Sources of Systematic Uncertainties	115
5.3.2	Result	130
5.4	Combined Electron+Jets and Muon+Jets Measurement	133
5.4.1	The Muon+Jets Analysis	133
5.4.2	Statistical Method and Systematic Uncertainties	135
5.4.3	Result	137
	Summary and Conclusion	139
A		143
A.1	Dataset Names of CMS Fall10 Production	143
A.2	Dataset Names of CMS Collision Data	143
A.3	Same-Sign/Opposite-Sign Method	144
A.4	Datasets with Systematic Variations	146
	List Of Figures	146
	List Of Tables	148
	Bibliography	150

Chapter 1

Theoretical Principles

Over the last five decades, the Standard Model (SM) of particle physics [16–27] has been very successful in describing the fundamental building blocks of matter, i.e. quarks and leptons, as well as the interactions between them. It is based on the concept of quantum field theory which describes continuous systems with an infinite number of degrees of freedom, so-called fields, and the application of quantum mechanics to them. Until today, there have been various experimental tests of different predictions deduced from the SM, and all of the predictions have been found to be fulfilled. However, it is known, that the SM cannot be the final theoretical answer since it involves 19 free parameters which are not predicted by the theory and can only be derived experimentally. Moreover it does not provide a way to incorporate gravitation into the theory which is thus the only known fundamental force that is not part of its framework and is instead described by the theory of general relativity [28]. Nevertheless, the SM is the most reliable theory describing subatomic particles and their interactions that has been developed so far.

In the following sections, a short introduction to the Standard Model of particle physics is given in a phenomenological way and with a deeper look into the mathematical formalism. In this thesis a measurement of the top-quark pair production cross-section is performed. Therefore, also the production and decay of top quarks at hadron colliders will be discussed in more detail.

1.1 The Standard Model of Particle Physics

1.1.1 Qualitative Introduction

The Standard Model of particle physics comprises twelve fundamental fermions without any substructure that can be grouped into six quarks and six leptons as shown in table 1.1. The quantum state of a given particle is described by so-called quantum numbers. For each of the mentioned fermions there is also a corresponding anti-fermion which differs from the fermion only by its opposite charges. All of the quarks and leptons carry spin $\frac{1}{2}\hbar$ and thus obey the Pauli exclusion principle [30], which states

Name	Fermion-Type	Symbol	El. Charge[e]	Mass [MeV/ c^2]
1st generation:				
electron neutrino	lepton	ν_e	0	$< 2 \cdot 10^{-6}$
electron	lepton	e	-1	0.511
up quark	quark	u	$+\frac{2}{3}$	1.7 – 3.3
down quark	quark	d	$-\frac{1}{3}$	4.1 – 5.8
2nd generation:				
muon neutrino	lepton	ν_μ	0	< 0.19
muon	lepton	μ	-1	105.66
charm quark	quark	c	$+\frac{2}{3}$	$(1.27^{+0.09}_{-0.07}) \cdot 10^3$
strange quark	quark	s	$-\frac{1}{3}$	101^{+29}_{-21}
3rd generation:				
tau neutrino	lepton	ν_τ	0	< 18.2
tau	lepton	τ	-1	1776.8 ± 0.16
top quark	quark	t	$+\frac{2}{3}$	$(173.3 \pm 1.1) \cdot 10^3$
bottom quark	quark	b	$-\frac{1}{3}$	$(4.19^{+0.18}_{-0.06}) \cdot 10^3$

Table 1.1: The fundamental fermions with their electric charge given in terms of the elementary charge and their masses [29]. The top quark mass is based on recent Tevatron results [8].

that two fermions cannot exist in exactly the same quantum state.

The six leptons can be further sub-divided into three generations with different lepton flavors. The first generation consists of the electron which carries one negative elementary electric charge (e) and the electron neutrino which is charge neutral. Similarly, the second and third generations are made up of the more massive brothers of the electron, the muon and the tau lepton, and their corresponding neutrinos. The SM predicts the neutrinos to be massless, however, experimental results [31–33] obtained within the last decade showed oscillations from one neutrino flavor into another. Since this is only possible for neutrinos with a non-zero mass, the Standard Model has to be extended for a proper inclusion of neutrino masses [34, 35]. Limits on the neutrino masses obtained in direct measurements are given in table 1.1. However, using cosmological models and data, the upper boundary on the sum of all three neutrino masses is found to be between 0.3 and 2.0 eV, depending on the used parameters and models [36].

Quarks carry fractional electric charge. Similar to the leptons, the six quarks can be arranged in three generations. Each generation consists of an up-type quark with an electric charge of $+\frac{2}{3}e$ and a down-type quark carrying $-\frac{1}{3}e$. Quarks appear in one of six different flavors: up, down, strange, charm, bottom and top. In nature, quarks cannot be observed as free particles. Due to a mechanism called confinement, they are always bound in quark-antiquark states (mesons) or states of three quarks or antiquarks (baryons), respectively. The existence of baryons with half-integer spin

Name	Symbol	Force	El. Charge[e]	Mass [GeV/c ²]
photon	γ	electromagnetic	0	$\leq 1 \cdot 10^{-27}$
gluon	g	strong	0	0
W boson	W^\pm	weak	± 1	80.399 ± 0.023
Z boson	Z^0	weak	0	91.1876 ± 0.0021

Table 1.2: The Gauge Bosons of the Standard Model with their electric charges and their masses [29]. The zero-mass of the gluon is taken from the theory prediction.

consisting of three identical quarks, which clearly violates the Pauli exclusion principle, led to the introduction of a further quantum number called color [37, 38]. There are three possible types of color: red, green, and blue. Quarks carry color while antiquarks carry anti-color. However, only colorless objects can be observed and thus a meson consists of a quark and an antiquark which carry the same type of color or anti-color, respectively. On the other hand, baryons contain all three types of color which also results in colorless objects.

All ordinary matter we know from our daily life is made up of quarks and leptons of the first generation. The particles from the other generations can only be produced in highly energetic reaction. In nature, those reactions happen when cosmic ray particles collide with the atoms of the earth's atmosphere. Under laboratory conditions, the particles of the higher generations can be produced in collider experiments, as will be described in chapter 2. Once produced, these particles rapidly decay into particles of the first generation. The Standard Model cannot predict the masses of the mentioned fermions. Hence, the nine masses enter the SM as free parameters.

Three types of forces are described by the Standard Model: the electromagnetic, the strong and the weak force. These interactions can be described by so-called gauge theories which predict massless gauge bosons being the mediators of the forces. Together with some of their properties the force-mediating bosons which all carry spin $1\hbar$ are shown in table 1.2.

The electromagnetic interaction is mediated by massless and charge neutral photons. This force is, for example, responsible for binding electrons and nuclei into atoms. Theoretically the interaction is described by Quantum Electrodynamics (QED) [39–42] and every particle carrying electric charge is subject to it. Quantum Chromodynamics (QCD) [16–22] is the gauge theory describing the interactions of particles carrying color charge. The strong force is responsible for binding quarks to baryons and mesons as well as for the formation of nuclei out of protons and neutrons. This force is mediated by massless gluons which carry color charges themselves. This feature of the strong interaction entails the possibility of gluon self-interactions which are the reason for a surprising behavior of the interaction strength. At short distances, the field strength of the strong interaction is small, a phenomenon called asymptotic freedom. For large distances, however, the field strength of the interaction increases until there is enough

energy available to produce quark-antiquark pairs from the vacuum which leads to the formation of colorless mesons and baryons. This effect is the already mentioned confinement and it is the reason why quarks cannot be observed as free particles.

The weak interaction is somehow special in the sense that the mediating bosons, the electrically charged W^\pm and the electrically neutral Z^0 , are found to have masses roughly 80 to 90 times higher than the proton mass. Thus, the weak interaction does not seem to fit into the concept of describing the SM interaction through gauge theories which predict massless bosons as the mediators of the forces. A mechanism to include the weak interaction into a gauge theory and to create the observed boson masses will be briefly explained later. An example for a process that occurs through weak interactions is the β -decay of a radioactive nucleus. The charge of the weak interaction is the weak isospin which is carried by all leptons and quarks. Since neutrinos neither obtain electric nor color charge, they can only be influenced by the weak force. Also the exchange of a W boson is the only possibility to change a quark's flavor. This is possible, because the eigenstates of the weak interaction (d' , s' , b') are not the same as the mass eigenstates (d , s , b). The transformations between those two bases are described by a 3×3 matrix which is by convention unitary to ensure that all transition probabilities sum up to one:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (1.1)$$

The matrix is called Cabibbo-Kobayashi-Maskawa (CKM) matrix [43, 44] and the strength of the coupling between the W boson and the quarks q_i and q_j is proportional to the element $V_{q_i q_j}$. Using all available measurements and imposing SM theoretical constraints, the following magnitudes for the absolute values of the CKM matrix elements can be determined [29]:

$$\tilde{V}_{\text{CKM}} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & (3.47_{0.12}^{+0.16}) \cdot 10^{-3} \\ 0.2252 \pm 0.0007 & 0.97345_{-0.00016}^{+0.00015} & (41.0_{-0.7}^{+1.1}) \cdot 10^{-3} \\ (8.62_{-0.20}^{+0.26}) \cdot 10^{-3} & (40.3_{-0.7}^{+1.1}) \cdot 10^{-3} & 0.999152_{-0.000045}^{+0.000030} \end{pmatrix}. \quad (1.2)$$

Here, \tilde{V}_{CKM} denotes the matrix containing the absolute values of the CKM matrix elements.

For a long time, the results of an experiment have been expected to be unchanged by the reversal of all spatial axes which is called the law of parity conservation. Hence, another surprising feature of the weak interaction is its maximal parity violating structure which is closely linked to the chirality of particles. An object is called chiral if it cannot be transformed into its mirror image through rotations or translations alone. In analogy to two well-known chiral objects, human hands, particles can have left-handed or right-handed chirality. Transitions between left-handed and right-handed states are possible via parity transformations. The maximal parity violating structure of the weak interactions arises from the W boson exchange, because it couples solely to left-handed particles and right-handed anti-particles. For massless particles, the chirality

is equal to the helicity which is the projection of a particle's spin onto its momentum direction. Thus, massless particles are only left-handed or only right-handed. However, the wave-functions of massive particles consist of left-handed and right-handed parts.

In the late 1960s, Glashow, Salam, and Weinberg showed how one can unify the electromagnetic and the weak interaction into one single gauge theory [25–27]. This new interaction is called the electroweak force. The pure electromagnetic or weak effects described above are assumed to be different aspects of the electroweak force for low energies. To explain the different behavior of the electromagnetic and the weak force, the electroweak symmetry has to be broken below a certain energy threshold. A method called Higgs mechanism [5–7] includes spontaneous breaking of the electroweak symmetry into the Standard Model by adding the so-called Higgs field. In this way, the masses of the W and Z bosons can be explained, but additionally a quantum of the Higgs field, the so-called Higgs boson, is predicted. The Higgs boson is the only particle within the SM that has not been discovered so far. From direct searches it is known that its mass has to be larger than $114.4 \text{ GeV}/c^2$ [45]. Recently also the mass window between 158 and $175 \text{ GeV}/c^2$ could be excluded [46].

1.1.2 Mathematical Formulation

The Standard Model is a gauge theory that is based on the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry group. Due to the axiomatic requirement of local gauge invariance of the Lagrangian density describing a particular system, new vector-like gauge fields are introduced into the theory. These new fields describe massless bosons which mediate forces between the particles subject to the force. First a short introduction to the least action principle and the Lagrangian formalism will be given. Thereafter, the Lagrangian density for QED, QCD and the electroweak theory will be derived. Finally, a brief explanation of the Higgs mechanism and the origin of the masses of the electroweak gauge bosons and the fundamental fermions will be given.

Lagrangian Formalism

In classical mechanics the equation of motion of a given system can be derived from the least action principle which states, that a particle will follow that particular path, for which the so-called action is minimized. Generalizing this approach also to relativistic fields, their dynamics can be expressed through a relativistically invariant action (if not explicitly mentioned otherwise, natural units with $\hbar = c = 1$ are used from now on in this chapter):

$$S = \int dt \int d^3x \mathcal{L} = \int d^4x \mathcal{L}. \quad (1.3)$$

In this equation, \mathcal{L} denotes the Lagrangian density which is a functional of the field $\phi(x)$ and its derivatives $\partial_\mu \phi(x)$. By requiring that the action is left unchanged under infinitesimal changes of the fields, what is called the principle of stationary action, a

relativistic form of the Euler-Lagrange equation can be derived:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) = 0. \quad (1.4)$$

Inserting the Lagrangian density of a given system into the Euler-Lagrange equation yields the equation of motion of the system. Hence, by using the Klein-Gordon Lagrangian for real scalar fields (spin-0 particles)

$$\mathcal{L} = \frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2, \quad (1.5)$$

the Klein-Gordon wave equation [47, 48],

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0, \quad (1.6)$$

can be obtained from the Euler-Lagrange equation. Similarly, the Dirac equation [49]

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \quad (1.7)$$

which describes massive spin- $\frac{1}{2}$ particles can be obtained by substituting the Dirac Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.8)$$

into equation 1.4. Here ψ is a four-component Dirac spinor and $\bar{\psi} \equiv \psi^\dagger \gamma^0$ its adjoint spinor.

Local Gauge Invariance and QED

As an example, and also to introduce the concept of local gauge invariance of a Lagrangian density, Quantum Electrodynamics are discussed at this point. Starting from the free Dirac Lagrangian density in equation 1.8, the theory is required to be invariant under local phase transformations of the form

$$\psi(x) \rightarrow \psi'(x) = e^{iQ\chi(x)}\psi(x), \quad (1.9)$$

where Q is an arbitrary real constant, $\chi(x)$ is a space-time dependent phase factor and $\psi(x)$ denotes the fermionic fields. The set of all such phase transformations forms the so-called $U(1)$ group. Since the elements of the $U(1)$ group are commuting, the group is called abelian. It is obvious that equation 1.8 is not invariant under $U(1)$ transformations, because of the extra term appearing in the derivative:

$$\partial_\mu \psi(x) \rightarrow e^{iQ\chi(x)} [\partial_\mu + iQ\partial_\mu \chi(x)] \psi(x). \quad (1.10)$$

However, if the derivative in equation 1.8 is replaced by the so-called covariant derivative

$$D_\mu \psi(x) \equiv [\partial_\mu - ieQA_\mu(x)] \psi(x), \quad (1.11)$$

where $A_\mu(x)$ is a new spin-1 field that transforms under the local phase transformation like

$$A_\mu(x) \rightarrow A'_\mu(x) \equiv A_\mu(x) + \frac{1}{e} \partial_\mu \chi(x), \quad (1.12)$$

the Lagrangian density can be made invariant. The modified Lagrangian thus gained an interaction term of the form

$$\mathcal{L}_{\text{int}} = eQ\bar{\psi}\gamma^\mu\psi A_\mu, \quad (1.13)$$

which specifies a unique form of the interaction between the Dirac spinor and the gauge field. If a further gauge-invariant term describing the kinematics of the gauge field is added to the Lagrangian, A_μ can be interpreted as a true propagating field. Therefore, in analogy to the first term in equation 1.5 (called kinetic term),

$$\mathcal{L}_{\text{kin}} \equiv -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \quad (1.14)$$

is added to the Lagrangian density. Here, the abbreviation $F_{\mu\nu}(x) \equiv \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x)$ has been used. A mass term similar to the one in equation 1.5 cannot be added, since it would spoil the local gauge invariance. Hence, the quanta of A_μ , the photons, are predicted to be massless.

Non-Abelian Gauge Theories and QCD

If a transformation like equation 1.9 acts on more than one field at the same time, e.g. on a vector of fields, the simple phase factor χ turns into matrices U which in general do not commute. Therefore, the related group is called non-abelian. To derive the Lagrangian density of Quantum Chromodynamics, the Dirac Lagrangian in equation 1.8 for quark fields $q_f^T \equiv (q_f^1, q_f^2, q_f^3)$, with f representing the quark flavor and the superscripts denote the three different colors, is required to be invariant under rotations in color space. The Lagrangian density thus reads

$$\mathcal{L} = \sum_f \bar{q}_f(x)(i\gamma^\mu\partial_\mu - m_f)q_f(x), \quad (1.15)$$

and the local phase transformation is given by

$$q_f(x) \rightarrow Uq_f(x) \equiv e^{i\alpha_a(x)\frac{\lambda_a}{2}}q_f(x). \quad (1.16)$$

The matrices U belong to the $SU(3)$ group. They can be parameterized in their general form by a set of eight linearly independent and traceless 3×3 matrices $\frac{\lambda_a}{2}$ and eight space-time dependent real parameters $\alpha_a(x)$, where a summation over the index a , running from 1 to 8, is implied. The λ_a matrices are commonly referred to as the Gell-Mann matrices and fulfill the following commutation relation:

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2} \right] = if_{abc} \frac{\lambda_c}{2}. \quad (1.17)$$

Here, the f_{abc} are real constants that are antisymmetric under exchange of any two indices. The $SU(3)$ group is an example for a so-called Lie group. Those groups possess the property, that their physical impact can already be obtained from infinitesimal changes to the identity. Therefore, the phase transformation of equation 1.16 can be written as

$$\begin{aligned} q_f(x) &\rightarrow \left[1 + i\alpha_a(x)\frac{\lambda_a}{2} \right] q_f(x), \\ \partial_\mu q_f(x) &\rightarrow \left[1 + i\alpha_a(x)\frac{\lambda_a}{2} \right] \partial_\mu q_f(x) + i\frac{\lambda_a}{2} q_f(x) \partial_\mu \alpha_a. \end{aligned} \quad (1.18)$$

As before, by replacing the derivative by the covariant derivative

$$D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} G_\mu^a, \quad (1.19)$$

and requiring that the new eight gauge fields transform in an appropriate way, the Lagrangian density can be made invariant under the transformation of equation 1.16. However, the transformation rule for the G_μ^a is not as simple as in QED, since the $\frac{\lambda_a}{2}$ matrices do not commute. This more complicated transformation is given by

$$G_\mu^a \rightarrow G_\mu^a - \frac{1}{g_s} \partial_\mu \alpha_a - f_{abc} \alpha_b G_\mu^c. \quad (1.20)$$

As in QED, the modified Lagrangian density exhibits an interaction term describing the interactions between the quark fields and the new gauge field. To obtain a true propagating field, an additional kinetic term for the G_μ^a is added to the Lagrangian and the gauge field quanta are identified as the massless gluons already mentioned before. The final Lagrangian density is given by

$$\begin{aligned} \mathcal{L}_{\text{QCD}} &= \mathcal{L} + \mathcal{L}_{\text{int}} + \mathcal{L}_{\text{kin}} \\ &= \sum_f \left[\bar{q}_f(x) (i\gamma^\mu \partial_\mu - m_f) q_f(x) - g_s (\bar{q}_f(x) \gamma^\mu \frac{\lambda_a}{2} q_f(x)) G_\mu^a \right] \\ &\quad - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \end{aligned} \quad (1.21)$$

where the abbreviation $G_{\mu\nu}^a \equiv \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c$ has been used. The unusual structure of the kinetic term is also a consequence of the non-abelian character of $SU(3)$ and features self-interactions between gluons through terms proportional to third and fourth order of the gauge field.

Electroweak Unification

As described above, the charge of the weak interaction is the weak isospin. Since only left-handed fermions as well as right-handed anti-fermions take part in the weak interaction, the particles can be grouped into weak isospin doublet states with $T = \frac{1}{2}$

and $T_3 = \pm\frac{1}{2}$. In contrast, the right-handed fermions as well as the left-handed anti-fermions can be treated as weak isospin singlet states with $T = 0$. As an example, this arrangement is shown for the first generation of leptons and quarks:

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, u_R, d'_R, \nu_{eR}, e_R^- \quad (1.22)$$

In principle the right handed neutrino, ν_{eR} , could be omitted here, because it neither carries electric charge nor weak isospin nor color charge and is thus not taking part in any of the SM interactions. For any fermion f , the left- and right-handed components can be extracted by using the chirality projection operators:

$$\frac{1}{2}(1 - \gamma^5)f = f_L, \quad \frac{1}{2}(1 + \gamma^5)f = f_R \quad (1.23)$$

In order to construct a gauge theory for the weak interaction it has to be required that the corresponding Lagrangian density is invariant under local $SU(2)_L$ transformations. However, it turned out that it is helpful to treat electromagnetic and weak phenomena together as a single interaction [25–27]. Hence, the Lagrangian is required to be invariant under combined $SU(2)_L \otimes U(1)_Y$ transformations.

The following discussions will be limited to only one set of leptons which are $\psi_1(x) = \begin{pmatrix} u \\ d \end{pmatrix}_L$, $\psi_2(x) = u_R$ and $\psi_3(x) = d_R$. Since it is not possible to include a fermion mass term, which would mix left-handed and right-handed fermion parts and thus spoil local gauge invariance, into the Lagrangian density, the common Dirac Lagrangian reduces to

$$\mathcal{L} = \sum_{j=1}^3 i\bar{\psi}_j(x)\gamma^\mu\partial_\mu\psi_j(x). \quad (1.24)$$

This Lagrangian density is then required to be invariant under local $SU(2)_L \otimes U(1)_Y$ transformations of the form

$$\begin{aligned} \psi_1(x) &\rightarrow \psi'_1(x) = e^{i\frac{Y_1}{2}\beta(x)} e^{i\frac{\sigma_k}{2}\alpha_k(x)}\psi_1(x), \\ \psi_{2/3}(x) &\rightarrow \psi'_{2/3}(x) = e^{i\frac{Y_{2/3}}{2}\beta(x)}\psi_{2/3}(x), \end{aligned} \quad (1.25)$$

where the Y_i are called hypercharge in analogy to QED, the index k runs from 1 to 3, the σ_k are the Pauli matrices and $\alpha_k(x)$ and $\beta(x)$ are space-time dependent factors. Following the same procedure as for QED and QCD the derivatives are replaced by the covariant derivatives,

$$\begin{aligned} D_{\mu,1} &= \partial_\mu + ig\frac{\sigma_k}{2}W_\mu^k(x) + ig'\frac{Y_1}{2}B_\mu(x), \\ D_{\mu,2/3} &= \partial_\mu + ig'\frac{Y_{2/3}}{2}B_\mu(x), \end{aligned} \quad (1.26)$$

and the four new gauge fields are required to transform in an appropriate way:

$$\begin{aligned} B_\mu(x) &\rightarrow B'_\mu(x) = B_\mu(x) - \frac{1}{g'}\partial_\mu\beta(x), \\ W_\mu^k(x) &\rightarrow W'^k_\mu(x) = W_\mu^k(x) - \frac{1}{g}\partial_\mu\alpha_k(x) - \epsilon_{mlk}\alpha_m(x)W_\mu^l(x). \end{aligned} \quad (1.27)$$

Here ϵ_{mlk} is the Levi-Civita tensor in three dimensions. It can be seen, that $B_\mu(x)$ transforms like $A_\mu(x)$ in the QED case. Therefore, it is reasonable to introduce also a kinetic term similar to equation 1.14 for the $B_\mu(x)$ fields. The kinetic term for the $W_\mu^k(x)$ has also the usual structure of $-\frac{1}{4}W_{\mu\nu}^k W_k^{\mu\nu}$, however, the $W_{\mu\nu}^k$ contain quadratic terms which give rise to gauge field self-interactions:

$$W_{\mu\nu}^k = \partial_\mu W_\nu^k - \partial_\nu W_\mu^k - g\epsilon_{mlk}W_\mu^m W_\nu^l. \quad (1.28)$$

The modified Lagrangian density, \mathcal{L}_{EWK} , contains interaction terms which couple the fermion fields to the gauge fields

$$\mathcal{L}_{\text{int}} = -g\bar{\psi}_1\gamma^\mu\frac{\sigma^k}{2}W_\mu^k\psi_1 - g'B_\mu\sum_{j=1}^3\frac{Y_j}{2}\bar{\psi}_j\gamma^\mu\psi_j. \quad (1.29)$$

Rewriting this equation using $W_\mu \equiv (W_\mu^1 + iW_\mu^2)/\sqrt{2}$ and its complex conjugate counterpart, these terms can be identified as the W^\pm bosons mediating weak charged current interactions. However, W_μ^3 and B_μ cannot directly be related to the Z boson and the photon. To reproduce the correct coupling of the photon known from QED, the following combinations of W_μ^3 and B_μ have to be used:

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix}. \quad (1.30)$$

Additionally $g\sin\theta_W = g'\cos\theta_W = e$ has to be fulfilled and the hypercharge is consequently defined as $\frac{Y}{2} = Q - T_3$, where Q is the electromagnetic charge and T_3 is the third component of the weak isospin. The relation between the hypercharge and the electromagnetic charge is called the Gell-Mann-Nishijima [50] formula and θ_W is known as the Weinberg angle.

Spontaneous Symmetry Breaking and Higgs Mechanism

The unification of the electromagnetic and the weak interaction into a gauge theory of the electroweak force has been described above. However, the gauge bosons appearing in this theory are still massless and thus in contradiction with the experimental results of massive W bosons [1, 2] and Z bosons [3, 4]. The first step in the mechanism of giving masses to the weak gauge bosons is the so-called Goldstone theorem. It states that whenever a continuous symmetry of a physical system is spontaneously broken, i.e. the symmetry holds for the system but not for its ground state, a massless spin-0 Goldstone boson occurs. Finally, the so-called Higgs mechanism explains how the choice of an appropriate gauge can be used to eliminate the unwanted massless Goldstone bosons, while at the same time creating masses for the gauge fields of the system.

Starting from a simple Lagrangian density describing a doublet of complex scalar fields ϕ in a potential $V(\phi)$

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi(x))^\dagger D^\mu\phi(x) - V(\phi(x)), \quad (1.31)$$

where D_μ is the covariant derivative of the $SU(2)_L \otimes U(1)_Y$ group defined in equation 1.26 and $V(\phi(x))$ is given by

$$V(\phi(x)) = \mu^2 \phi^\dagger(x)\phi(x) + \lambda(\phi^\dagger(x)\phi(x))^2 \text{ with } \mu^2 < 0, \lambda > 0, \quad (1.32)$$

it can be seen that the theory is invariant under local $SU(2)_L \otimes U(1)_Y$ transformations. In the end only the weak gauge bosons should acquire masses, but the photon has to remain massless. Hence, the scalar doublet is already chosen in an appropriate way as an weak isospin doublet with $T = \frac{1}{2}$ and hypercharge $Y = 1$:

$$\phi(x) \equiv \begin{pmatrix} \phi^{(+)}(x) \\ \phi^{(0)}(x) \end{pmatrix}. \quad (1.33)$$

Therefore, the $\phi^{(+)}(x) \equiv (\phi_1(x) + i\phi_2(x))$ field is electrically charged while the $\phi^{(0)}(x) \equiv (\phi_3(x) + i\phi_4(x))$ is charge neutral. The potential $V(\phi)$ exhibits an infinite number of states with minimum energy. By choosing a particular ground state, for example

$$\phi_0(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \text{ with } v \equiv \sqrt{\frac{-\mu^2}{\lambda}}, \quad (1.34)$$

the $SU(2)_L \otimes U(1)_Y$ symmetry gets spontaneously broken, because the new ground state is not invariant under these kinds of transformations. But since $\phi^{(0)}(x)$ is charge neutral, the ground state is still invariant under $U(1)_{EM}$ transformations. Hence, according to the Goldstone theorem, it is expected to find three new massless scalar Goldstone bosons when inserting $\phi(x)$ expanded around its new vacuum state $\phi_0(x)$ into equation 1.31.

A general expansion of $\phi(x)$ around $\phi_0(x)$ is given through

$$\phi(x) = e^{i\frac{\sigma_k}{2}\theta_k(x)} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (1.35)$$

where $\theta_k(x)$ and $H(x)$ are four real fields and $v \approx 246$ GeV is the vacuum expectation value of the Higgs field. The Lagrangian in equation 1.31 is still invariant under local $SU(2)$ transformations. Thus, the particular gauge of $\theta_k(x) = 0$ can be chosen. When inserting $\phi(x)$ from equation 1.35 into the Lagrangian density and using the definitions of W_μ and Z_μ from above, the following terms can be found:

$$\begin{aligned} (D_\mu \phi(x))^\dagger D^\mu \phi(x) &\rightarrow \frac{1}{2} \partial_\mu H(x) \partial^\mu H(x) \\ &+ (v + H(x))^2 \left[\frac{g^2}{4} W_\mu^\dagger(x) W^\mu(x) \right. \\ &\left. + \frac{g^2}{8 \cos^2 \theta_W} Z_\mu(x) Z^\mu(x) \right]. \end{aligned} \quad (1.36)$$

First it is obvious that the weak gauge fields have become massive with $M_W = \frac{1}{2}vg$ and $M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}$. Also the relation between the Weinberg angle and the masses

of the weak gauge bosons can be derived:

$$\cos \theta_W = \frac{M_W}{M_Z}. \quad (1.37)$$

Furthermore, it can be seen that a new massive spin-0 field, $H(x)$, has appeared. The quanta of this field are the still not discovered Higgs bosons. The interactions between the Higgs boson and the gauge bosons are always proportional to the square of the gauge boson mass, as can be seen in equation 1.36.

The only thing still missing, is a mechanism that gives masses also to fermions, because a fermion mass term could not be introduced in equation 1.24. Using the same Higgs field doublet as above, the following $SU(2)_L \otimes U(1)_Y$ invariant term can be added to the Lagrangian density:

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}} = & - c_1 (\bar{u}, \bar{d}')_L \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} d'_R - c_2 (\bar{u}, \bar{d}')_L \begin{pmatrix} -\bar{\phi}^{(0)} \\ \phi^{(-)} \end{pmatrix} u_R \\ & - c_3 (\bar{\nu}_e, \bar{e})_L \begin{pmatrix} \phi^{(+)} \\ \phi^{(0)} \end{pmatrix} e_R + \text{h.c.} \end{aligned} \quad (1.38)$$

Again the discussion is limited to the first fermion generation and the c_1 , c_2 , and c_3 represent the coupling constants for the coupling of the Higgs field to these fermion fields. In the second term, the charged conjugate scalar field $\phi_C = -i\sigma_2\phi^* = \begin{pmatrix} -\bar{\phi}^{(0)} \\ \phi^{(-)} \end{pmatrix}$ has been used and the abbreviation h.c. stands for hermitian conjugate. After the spontaneous symmetry breaking the Lagrangian turns into

$$\mathcal{L}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}}(v + H) [c_1 \bar{d}' d' + c_2 \bar{u} u + c_3 \bar{e} e], \quad (1.39)$$

and the fermion mass terms can be identified as $m_d = c_1 \frac{v}{\sqrt{2}}$, $m_u = c_2 \frac{v}{\sqrt{2}}$ and $m_e = c_3 \frac{v}{\sqrt{2}}$. Since the values of the coupling constants c_1 , c_2 , and c_3 are not predicted by the SM, the fermion masses enter the theory as free parameters. Furthermore the couplings between the fermions and the Higgs boson are found to be proportional to the fermion mass divided by the vacuum expectation value of the Higgs field and are thus small for all fermions except for the top quark. This type of interaction between Dirac fields and a scalar field is called Yukawa coupling. Finally, the Lagrangian density of the Standard Model of particle physics can be written as

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EWK}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \quad (1.40)$$

1.1.3 Feynman Diagrams and Cross Sections

In Quantum Mechanics, the transition rate from a given initial state into a particular final state in the presence of an interaction potential can be obtained from Fermi's Golden Rule [51]. This rule states, that the transition rate is proportional to the square of the Lorentz invariant matrix element describing the interaction in terms of

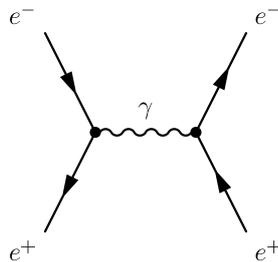


Figure 1.1: Example of a leading order Feynman Diagram showing the annihilation of an electron-positron pair into a virtual photon which finally converts into an electron-positron pair.

an infinite perturbation series. The different terms of the perturbation series are proportional to different powers of the interaction's coupling constant. Here, the coupling constant of the strong interaction plays a special role, because it can be larger than one for small momentum transfers and thus perturbation theory cannot be applied. Matrix elements including only terms proportional to the smallest possible power of the coupling constant are called leading order (LO). The matrix elements including also the following powers of the coupling constant are called next-to-leading order (NLO) or even next-to-next-to-leading-order (NNLO). For coupling constants much smaller than one, the LO matrix element will have the largest contribution to the perturbation series, however, higher orders might also have significant contributions. There are also effects which are not possible at LO and do only occur at higher orders, like, for example, the oscillation from K^0 mesons into \overline{K}^0 mesons or vice versa via the weak interaction. The matrix elements depend on the kinematics of the incoming and outgoing particles as well as on the kinematics of the bosons which mediate the interaction. A graphical representation of the different terms in the perturbation series is given through so-called Feynman Diagrams which also give a schematic view of the mentioned kinematics. A set of dedicated Feynman Rules allows to translate the graphical representation into mathematical expressions for the calculation of the corresponding matrix element. A typical Feynman Diagram describing the annihilation of an incoming electron and positron into a virtual photon which propagates and finally converts into an outgoing electron and positron is shown in figure 1.1. For the description of diagrams like figure 1.1 with two incoming and two outgoing particles, it is useful to introduce the Lorentz invariant Mandelstam variables [52]. Assuming that the incoming and outgoing electrons have momenta of p_A and p'_A , respectively, while the incoming and outgoing positrons obtain momenta of p_B and p'_B , the Mandelstam variables are defined as

$$\begin{aligned}\hat{s} &= (p_A + p_B)^2 = (p'_A + p'_B)^2, \\ \hat{t} &= (p_A - p'_A)^2 = (p_B - p'_B)^2, \\ \hat{u} &= (p_A - p'_B)^2 = (p_B - p'_A)^2.\end{aligned}\tag{1.41}$$

These definitions hold for any process with two particles in the initial and final state.

In particle physics, the cross section is a measure for the probability of a given process to occur. It can also be regarded as the effective area over which particles are interacting and is thus frequently given in terms of barn [b], where $1 \text{ b} = 1 \cdot 10^{-24} \text{ cm}^2$. The cross section of a given process is proportional to the transition rate mentioned above. Hence, it is also proportional to the square of the invariant matrix element and can be calculated by summing over all contributing Feynman Diagrams.

When colliding composite objects like hadrons at high energies, not the full hadrons themselves but only their constituents which are called partons will interact with each other. Thus, to derive cross sections in hadronic interactions correctly, the interaction is typically broken down to the partonic level. Using the factorization theorem, the hadronic cross section can be expressed through the partonic cross section convoluted with the parton distribution functions (PDF) $f(x_i, \mu_F^2)$, which describe the probability of finding a parton carrying a particular hadron momentum fraction x_i inside the hadron. The PDFs depend on the factorization scale μ_F^2 , which can be thought of as the boundary between the non-perturbative long-range and the perturbative short-range QCD processes. Therefore, partons with momenta smaller than μ_F are treated as part of the hadron structure and are absorbed within the PDF. Experimentally, the PDFs can be extracted from experiments probing the substructure of protons with point-like particles like electrons or neutrinos (deep inelastic scattering). An overview of the measurements performed at the electron-proton collider HERA (occasionally also positron-proton collisions) is given in [53]. There are different PDF parameterizations available, however, in this thesis mainly the CTEQ6L1 [54] PDF has been used. In figure 1.2, the proton CTEQ6L1 PDFs of u and d quarks (valence and sea quarks combined), \bar{u} and \bar{d} sea quarks and gluons at $\mu_F = 172.5 \text{ GeV}$ are shown.

Using the PDFs the factorization ansatz is given by

$$\sigma(AB \rightarrow cd) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_{i,A}(x_i, \mu_F^2) f_{j,B}(x_j, \mu_F^2) \cdot \hat{\sigma}_{ij}(ij \rightarrow cd; \hat{s}, \mu_R^2, \mu_F^2). \quad (1.42)$$

Here A and B represent the original hadrons, $f_{i,A}(x_i, \mu_F^2)$ and $f_{i,B}(x_j, \mu_F^2)$ are the PDFs for the two hadrons and the $x_{i/j}$ denote the hadron momentum fractions carried by the partons i and j , respectively. The partonic cross section $\hat{\sigma}_{ij}$ depends on μ_F and also on the renormalization scale μ_R which is introduced in the procedure of regulating the divergences appearing through loop corrections during the calculation of $\hat{\sigma}_{ij}$. Up to a certain degree, μ_F and μ_R are arbitrary constants and thus often $\mu_F = \mu_R$ is chosen for simplicity. The partonic cross section also depends on the square of the center-of-mass energy of the colliding partons which is given through the already introduced Mandelstam variable $\hat{s} = (p_i + p_j)^2 = (x_i p_A + x_j p_B)^2$. Assuming symmetric collisions, i.e. $p_A = p_B$, the partonic center-of-mass energy is related to the center-of-mass energy of the colliding hadrons through $\hat{s} = x_i x_j s$.

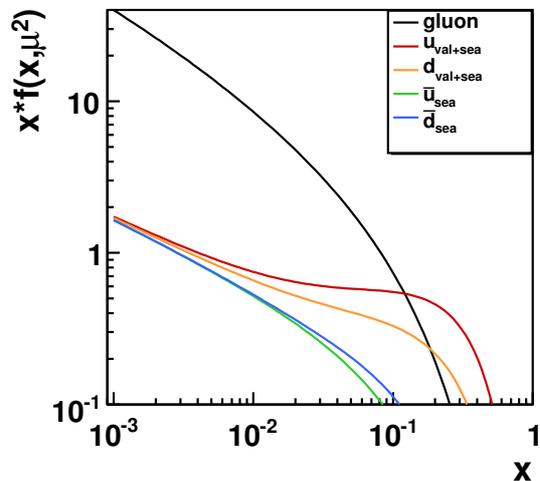


Figure 1.2: CTEQ6L1 proton PDF for different quark flavors. For u and d quarks the combined valence+sea quark PDF is shown. For the factorization scale $\mu_F = 172.5$ GeV has been chosen.

1.2 Top-Quark Physics

The discovery of the top quark in 1995 by the CDF [9] and DØ [10] experiments at Fermilab’s proton anti-proton collider Tevatron was one of the outstanding confirmations of the Standard Model predictions. Since the discovery of the bottom quark in 1977 at Fermilab [55], the top quark has been predicted to be its missing weak isospin partner. Precise electroweak measurements on the Z boson resonance summarized in [56] encouraged this expectation. Also the top quark mass could be predicted with high accuracy [57] even before the top quark discovery. The most recent combination of top quark mass measurements performed at the Tevatron yields a mass of 173.3 ± 1.1 GeV [8].

Due to its much higher mass compared to all of the other known quarks, the top quark plays a special role within the Standard Model. First of all, the large mass is the reason for the very small lifetime of the top quark of $\tau_t = 1/\Gamma_t \approx 5 \cdot 10^{-25}$ s [58, 59], where Γ_t is the top quark decay width. This lifetime is smaller than the typical time in which quarks are arranged in colorless hadrons which is of the order $\tau_{\text{had}} \approx 1 \text{ fm}/c \approx 3 \cdot 10^{-24}$ s. Consequently, the top quark decays before top mesons or top baryons can be formed. Therefore, the top quark spin is transferred to its decay products without being depolarized by the strong interaction. Hence, top quarks provide the unique possibility to study a quasi-free quark through its decay products.

At the time of its discovery, it was not clear if the observed particle was really the SM predicted top quark, since its charge has not been measured directly and thus also an exotic quark with charge $-\frac{4}{3}e$ would have been possible. However, recent measurements of the CDF collaboration [60] exclude this exotic quark hypothesis at the 95% confidence level. A further interesting feature in the production of top quark

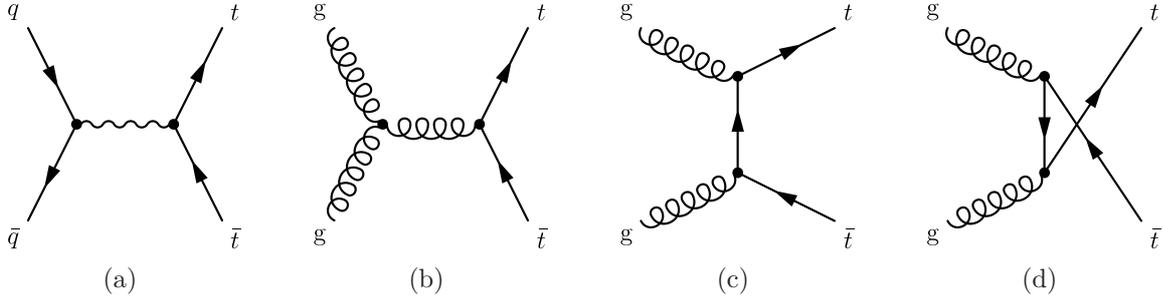


Figure 1.3: Feynman diagrams for the LO production of top-quark pairs via quark-antiquark annihilation (a) and gluon-gluon fusion (b)-(d).

pairs at the Tevatron is an asymmetry in the rapidity distributions of the top and anti-top quarks which results in a forward-backward asymmetry with respect to the beam direction. For high invariant masses of the two top quarks, $M_{t\bar{t}} > 450$ GeV, the Standard Model predicts an asymmetry of roughly 9%. Recently, the CDF experiment measured an asymmetry of $(47.5 \pm 11.4)\%$ [61] in this special region of phase space. Possibly, this could be a hint for heavy particles with an asymmetric coupling to quarks and anti-quarks which are decaying into top-quark pairs. Thus, top-quark physics are also providing a window to phenomena which are not described within the framework of the Standard Model. A more detailed review on top quarks, their properties, and experimental results can be found in [62].

1.2.1 Top-Quark Production

At hadron colliders, top quarks can be produced in two different processes. Firstly, they can be produced via the strong or electroweak interaction in pairs of top and anti-top quarks, where the contribution of the electroweak process is negligible. Secondly, they can be produced singly via the electroweak interaction. For collisions energies well above $2 \cdot m_t$, the top quark pair production is the dominant process, while the production cross-section of single top-quarks is roughly two times smaller.

Top-Quark Pair-Production

The top-quark pair-production cross section can be derived from equation 1.42. An estimation of the cross section at LO perturbation theory in the strong coupling constant α_s (i.e. terms proportional to α_s^2) can be calculated from the LO Feynman diagrams shown in figure 1.3. As can be seen, top-quark pairs can be produced by quark-antiquark annihilation (figure 1.3 (a)) or by gluon-gluon fusion processes (figure 1.3 (b)-(d)). Using the Feynman rules, the LO differential cross section can be

calculated as given in [63]:

$$\begin{aligned}
\frac{d\hat{\sigma}}{d\hat{t}}(q\bar{q} \rightarrow t\bar{t}) &= \frac{4\pi\alpha_S^2}{9\hat{s}^4} \cdot [(m_t^2 - \hat{t})^2 + (m_t^2 - \hat{u})^2 + 2m_t^2\hat{s}] & (1.43) \\
\frac{d\hat{\sigma}}{d\hat{t}}(gg \rightarrow t\bar{t}) &= \frac{\pi\alpha_S^2}{8\hat{s}^2} \cdot \left[\frac{6(m_t^2 - \hat{t})(m_t^2 - \hat{u})}{\hat{s}^2} - \frac{m_t^2(\hat{s} - 4m_t^2)}{3(m_t^2 - \hat{t})(m_t^2 - \hat{u})} \right. \\
&\quad + \frac{4}{3} \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) - 2m_t^2(m_t^2 + \hat{t})}{(m_t^2 - \hat{t})^2} \\
&\quad + \frac{4}{3} \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) - 2m_t^2(m_t^2 + \hat{u})}{(m_t^2 - \hat{u})^2} \\
&\quad - 3 \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) + m_t^2(\hat{u} - \hat{t})}{\hat{s}(m_t^2 - \hat{t})} \\
&\quad \left. - 3 \cdot \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u}) + m_t^2(\hat{t} - \hat{u})}{\hat{s}(m_t^2 - \hat{u})} \right]. & (1.44)
\end{aligned}$$

Here \hat{s} , \hat{t} , and \hat{u} are the Lorentz invariant Mandelstam variables introduced before and m_t is the top-quark mass. The production of top-quark pairs is kinematically only possible if $\hat{s} \geq 4m_t^2$. Hence, when assuming $x_i \approx x_j$, the minimal hadron momentum fraction can be derived via $x_{\min} = \frac{2m_t}{s}$. For a hadronic center-of-mass energy of 7 TeV the minimal x is thus roughly 0.05. At such a low hadron momentum fraction, the gluon-gluon fusion processes are dominating and the ratio of gluon-gluon fusion to quark-antiquark annihilation processes is roughly 4 : 1. Calculations of the partonic cross section at NLO are available since the end of the 1980s [64, 65]. At a center-of-mass energy of 7 TeV the NLO top-quark pair production cross section can be calculated using the MCFM [11] package version 5.8:

$$\sigma_{t\bar{t}, \text{NLO}} = 157.5^{+18.0}_{-19.5} (\text{scale}) \pm 14.7 (\text{PDF}) \text{ pb}. \quad (1.45)$$

For this calculation a top-quark mass of 172.5 GeV and $\mu = \mu_F = \mu_R = m_t$ has been used. The PDF parameterization used is CTEQ6M [54] and the corresponding uncertainties are evaluated by using the corresponding 90% confidence intervals. The uncertainties arising from the chosen scale are estimated by recalculating the cross section for $\mu = 2 \cdot m_t$ and $\mu = 0.5 \cdot m_t$.

Currently, the most precise determinations of the $t\bar{t}$ production cross section are approximate NNLO calculations using next-to-next-to-leading-logarithm (NNLL) accuracy in the soft-gluon resummation [66, 67]. The effects of soft-gluon emissions from the external partons can be evaluated by applying a Mellin transformation [68] to the cross section. In Mellin space, the corrections to the cross section due to soft-gluon radiations can then be expressed in a power series of logarithms and everything up to the second leading term is used for the NNLL approximation. The resulting cross section at 7 TeV center-of-mass energy derived in [66] for a top-quark mass of 173 GeV, $\mu = m_t$ and the MSTW2008 [69] NNLO PDF is

$$\sigma_{t\bar{t}, \text{approxNNLO}} = 163^{+7}_{-5} (\text{scale}) \pm 9 (\text{PDF}) \text{ pb}. \quad (1.46)$$

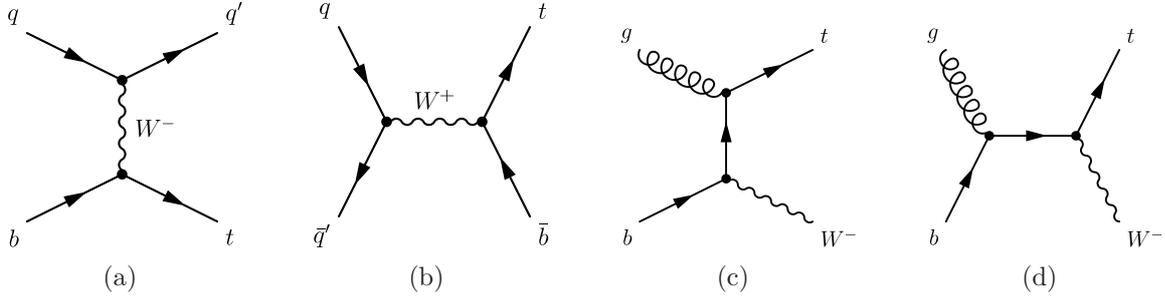


Figure 1.4: Feynman diagrams for the LO production of single top-quarks. In (a) the t-channel production process is shown, while in (b) the s-channel production mode is shown. The two diagrams (c) and (d) show the single top-quark production in association with a real W boson.

The scale uncertainty is again evaluated by taking twice or half the original scale and the PDF uncertainties are derived from the 90% confidence level of the MSTW2008 NNLO PDF set. Furthermore, the HATHOR package [12] can be used to derive the cross section based on the results given in [67] while using the same assumptions for the top-quark mass, the scale, and the PDF as for [66]:

$$\sigma_{t\bar{t}, \text{approxNNLO}} = 164_{-9}^{+5} (\text{scale}) \pm 9 (\text{PDF}) \text{ pb}. \quad (1.47)$$

The uncertainties due to the scale and the PDF are derived in the same way as explained above. Although there are subtle differences in the way the two calculations are done, both results are in very good agreement.

Single Top-Quark Production

The production of single top-quarks via the weak interaction was first discovered at the Tevatron in 2009 [70, 71]. In the Standard Model, there are three different production modes: the t-channel, the s-channel and the associated production. The Feynman diagrams shown in figure 1.4 (a)-(d) illustrate the LO matrix elements for the mentioned production modes. The different modes can be distinguished by the virtuality Q^2 of the participating W boson ($Q^2 = -q^2$, where q is the four momentum of the W boson). Since all of these processes obtain a Wtb vertex, the corresponding cross sections are proportional to the CKM matrix element V_{tb} . Therefore, the production of single top-quarks provides the unique opportunity of a direct determination of $|V_{tb}|$.

t-channel Production ($q^2 = \hat{t}$)

In the t-channel production a virtual, space-like ($q^2 < 0$) W boson hits a bottom sea-quark which arises from a gluon splitting. That is the reason why this production mode is also often called W -gluon fusion. In proton-proton collisions the cross section for producing single top-quarks is roughly two times higher than the cross section for producing single anti-top quarks. This is a direct consequence of the proton's quark

content where the number of u valence quarks is twice the number of d quarks. The Feynman diagram illustrating the LO matrix element is shown in figure 1.4 (a). The combined t-channel single top- and anti-top quark production cross section at NLO for a center-of-mass energy of 7 TeV can be calculated using the MCFM package:

$$\sigma_{t+\bar{t}} = 64.6_{-1.1}^{+1.5} (\text{scale}) \pm 3.0 (\text{PDF}) \text{ pb}. \quad (1.48)$$

For this calculation the top-quark mass is assumed to be 172.5 GeV, $\mu = m_t$, and the CTEQ6M PDF have been used.

s-channel Production ($q^2 = \hat{s}$)

In the s-channel production mode, the fusion of a quark and an antiquark leads to the production of a time-like W boson ($q^2 \geq (m_t + m_b)^2$) which decays into a bottom and a top quark. For proton-proton collisions this production is almost negligible, because any antiquark in the initial state has to be a sea-quark. As for the t-channel mode, the number of produced single top-quarks is twice the number of single anti-top quarks. The LO Feynman diagram is shown in figure 1.4 (b). Using approximate NNLO calculations including NNLL soft-gluon resummation, the s-channel production cross sections at $\sqrt{s} = 7$ TeV for $\mu = m_t = 173$ GeV and the MSTW2008 NNLO parton distribution function is [72]:

$$\sigma_t = 3.17 \pm 0.06 (\text{scale})_{-0.10}^{+0.13} (\text{PDF}) \text{ pb}. \quad (1.49)$$

tW Production ($q^2 \approx m_W$)

This production mode is characterized by the production of a single top-quark and a real W boson at the same time. The tW production is less probable than the t-channel mode but still exceeds the s-channel production at $\sqrt{s} = 7$ TeV. When calculating NLO corrections to the cross section a difficulty arises since the tW production interferes with the LO $gg \rightarrow t\bar{t}$ process with a subsequent top decay. Different approaches are available to define the tW production and to keep the interference small. To define the tW process, the method described in [73] uses an upper boundary on the momentum of the additional b quark appearing at NLO. This approach is implemented in MCFM which can be used to derive the NLO cross section for $\sqrt{s} = 7$ TeV:

$$\sigma_{tW, \text{NLO}} = 10.6 \pm 0.8 (\text{scale}) \text{ pb}. \quad (1.50)$$

Again $\mu = m_t = 172.5$ GeV and the CTEQ6M PDF have been used. Two further methods are explained in [74]. The first one is called diagram removal (DR) and simply removes all diagrams shown in figure 1.3 from the NLO tW calculations. The second one, called diagram subtraction (DS), modifies the NLO tW cross section by implementing a subtraction term designed to cancel locally the $t\bar{t}$ contribution. A recent approximate NNLO calculation with NNLL soft-gluon resummation uses the

DS method. The resulting cross section at $\sqrt{s} = 7$ TeV for $\mu = m_t = 173$ GeV and the MSTW2008 NNLO parton distribution function is [75]:

$$\sigma_{tW^-} = \sigma_{\bar{t}W^+} = 7.8 \pm 0.2 (\text{scale})_{-0.6}^{+0.5} (\text{PDF}) \text{ pb}. \quad (1.51)$$

As stated in [74], a rough estimate of the systematic uncertainty on the tW cross section for a particular set of selection cuts can be obtained by comparing the results from the DR and the DS methods.

1.2.2 Top-Quark Decay

In the Standard Model, the values of the CKM matrix elements V_{ts} and V_{td} are very small (see equation 1.2). Hence, the top quark decays with almost 100% into a W boson and a bottom quark and top-quark decays into down-type quarks of a different generation are highly suppressed. In contrast to weak decays of all of the other quarks, in top-quark decays real W bosons can be produced due to the high top-quark mass. This is the reason for the very small top-quark lifetime mentioned earlier.

The decays of top-quark pairs can be subdivided into three different categories depending on the subsequent decay of the W bosons. In case both W bosons decay into a pair of quarks, the top-quark pair decay is said to be fully-hadronic. If both W bosons decay into a charged lepton and the corresponding neutrino, instead, the top-quark pair decay is called di-leptonic. The remaining possibility, for which one W boson decays into quarks while the other decays into a lepton and a neutrino, is called semi-leptonic or lepton+jets channel.

Because of the universal coupling of the W boson to quarks and leptons, all W decay channels exhibit in principle the same probability when differences in the available phase space due to the different quark and lepton masses are neglected. However, since quarks can appear in three different colors, each hadronic W decay is enhanced by a factor of three compared to any of the leptonic decay channels. In summary, this leads to a probability of 1/9 for any leptonic decay channel and a probability of 3/9 for any hadronic decay channel. Based on these W boson decay probabilities the branching ratio for the different $t\bar{t}$ decay channels can be calculated. The results are illustrated in the pie chart shown in figure 1.5.

The measurement of the top-quark pair production cross section described in this thesis is performed in the electron+jets and the muon+jets channel. Top-quark pairs decay with a probability of roughly 30% into these final states. Besides the branching ratio, the expected contributions from background processes exhibiting the same final state, have to be considered. In this sense, the chosen decay channels are a compromise between the all-hadronic channel, which has a very high branching ratio but also a very high background contamination, and the nearly background-free di-lepton channel, which suffers from the disadvantage of a small branching ratio.

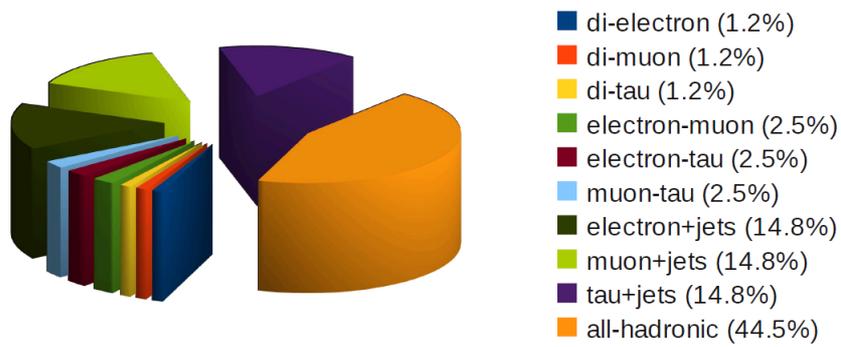


Figure 1.5: Illustration of the probabilities for the different $t\bar{t}$ decay channels. The branching ratio for each channel is given in brackets and is derived based on the probabilities for the different W boson decay channels. For a simple estimation of these probabilities, the differences in the available phase space due to the different quark and lepton masses have been neglected.

Chapter 2

Experimental Setup

To probe the fundamental building blocks of matter, physicists build large machines, so-called colliders, where particles are accelerated almost to the speed of light and finally collided at high center-of-mass energies. According to Einstein's famous relation between mass and energy, $E = m \cdot c^2$ [76], new particles can be produced in these highly energetic collisions. The particles and their characteristics are the subjects of current research in high-energy physics.

The Large Hadron Collider (LHC) [77] of the European Organization for Nuclear Research (CERN) in Geneva, Switzerland, is currently the world's most energetic collider. Installed in the tunnel of the former Large Electron Positron collider (LEP) which is located between 45 m and 170 m below the surface, the LHC is designed to accelerate single proton beams up to an energy of 7 TeV and then provide proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 14$ TeV. Superconducting dipole magnets, designed to provide a magnetic field of 8.33 T, are used to guide the proton beams around the 26.7 km long ring. Figure 2.1 shows an aerial view of the Geneva countryside. The white lines indicates the location of the LHC main ring and its preaccelerators, the CERN area can be seen on the bottom right hand side of the main ring.

Four large particle detectors are installed along the ring, recording the products of the collisions that take place in their centers. The ALICE (A Large Ion Collider Experiment) detector [79] focusses on the strong-interaction sector of the Standard Model and is designed to address the physics of strongly interacting matter at extreme values of energy density and temperature in nucleus-nucleus collisions. Two general purpose detectors, ATLAS (A Torroidal LHC ApparatuS) [80] and CMS (Compact Muon Solenoid) [81, 82], are mainly searching for the yet to be discovered Higgs boson and hints for phenomena which cannot be explained within the framework of the Standard Model. Although both experiments pursue the same physics goals, they have developed different detection strategies and different detector designs allowing to independently cross-check each others results. The primary goal of the fourth experiment, LHCb [83], is the search for phenomena which could explain the imbalance between baryonic and antibaryonic matter in the universe. The rare decays of heavy beauty



Figure 2.1: Aerial view of the Geneva countryside [78]. The white lines indicate the location of the LHC main ring and its preaccelerators below the surface. The CERN (Meyrin Site) can be found at the bottom right of the main ring.

and charm hadrons are also of special interest to this experiment. In addition, there are two smaller detectors, LHCf [84] and TOTEM [85], which are installed close to ATLAS and CMS, respectively.

The analysis described in this thesis uses data recorded with the CMS detector in 2010. Therefore, in the following sections only the main components of CERN's accelerator chain and the different components of the CMS detector will be discussed in more detail.

2.1 The Large Hadron Collider

2.1.1 Production of Proton Beams and Preacceleration

Figure 2.2 schematically illustrates the different parts of CERN's accelerator chain from the production of the proton beams to their final injection into the LHC main ring. The proton beams are produced in a duoplasmatron source where hydrogen atoms are ionized by a beam of electrons emitted from a thermionic cathode. A further cathode is used to extract the positively charged hydrogen ions which leads to a proton beam with an average energy of 90 keV leaving the duoplasmatron source. The proton beam then enters a Radio Frequency Quadrupole (RFQ) which is a 1.75 m long linear accelerator where the beam is focussed, divided into groups of protons, so-called bunches, and further accelerated to an energy of 750 keV. The concept of arranging protons into bunches instead of having a continuous proton beam is an elementary prerequisite for the technique which is used to increase the energy of particle beams in modern accelerators. The devices responsible for the acceleration are radio frequency (RF) cavities producing oscillating electrical fields. The oscillation frequency of these fields

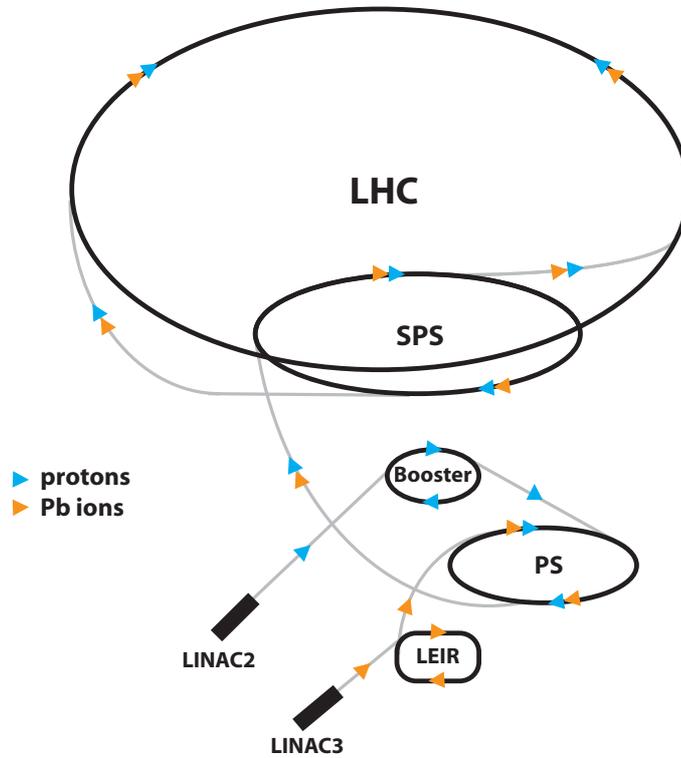


Figure 2.2: Schematical view of the different parts of CERN’s accelerator chain (not to scale). The proton beams produced in a duoplasmatron source are preaccelerated within the LINAC2, the Booster, the PS and the SPS before entering the LHC main ring. Lead ion beams are first accelerated by LINAC3 and LEIR before entering the PS, SPS and finally the LHC.

is chosen in such a way, that the beam particles are always passing an accelerating potential difference. Since only one phase of the electrical field is adequate for this — the other would lead to a deceleration — the oscillation frequency of the fields also determines the minimal distance between two adjacent bunches.

After leaving the RFQ the proton bunches are entering the LINAC2, a 30 m long linear accelerator which increases their energy up to 50 MeV. Thereafter, the proton beam is inserted into the first circular accelerator, the Proton Synchrotron Booster (PSB) [86] which consists of four rings mounted on top of each other allowing to use the same magnets and RF cavities for all four accelerator rings at the same time. In each of these rings the energy of the proton bunches is increased to 1.4 GeV. The advantage of having circular accelerators, so-called synchrotrons, instead of linear ones is that the particles can pass the same RF cavities many times and thus, their energy is increased in every turn. However, the particles have to be kept on circular orbits using magnetic fields commonly produced by dipole magnets. Also the frequency of the RF cavities and the strength of the magnetic field has to be increased simultaneously with the increase of the beam energy. Hence, one limiting factor for the maximum energy that can be achieved with a synchrotron of a given circumference is the maximal

magnetic field that can be reached by the dipole magnets.

From the PSB the proton beam is sent to the Proton Synchrotron (PS) [87] which has a circumference of 628.3 m and further accelerates the protons to an energy of 25 GeV. The final stage of preacceleration is done in the Super Proton Synchrotron (SPS) [88]. After passing this accelerator with a circumference of roughly 6.9 km the proton beam exhibits an energy of 450 GeV. The beam is then extracted from the SPS at two different points and sent via the transfer lines TI 2 and TI 8 to the LHC main ring where it is injected clockwise or counter-clockwise, respectively.

At nominal operation the LHC main ring will be filled with 2,808 proton bunches per beam, each consisting of $1.15 \cdot 10^{11}$ protons. Most of the bunches are separated by 25 ns where this bunch spacing time originates from the 40 MHz RF cavity which is used for the acceleration in the PS. However, since the injection of the beams into the different accelerators or transfer lines is done with fast ramping, so-called kicker-magnets with a finite rise time, also longer bunch spacing intervals are inevitable. The LHC is filled within 12 SPS cycles where in nine cycles three so-called batches and in the remaining three cycles four batches are injected. Each of these batches consists of 72 bunches separated by 25 ns and eight missing bunches due to the SPS injection kicker rise time of 220 ns. Because of the LHC injection kicker rise time of 0.94 μ s, there is a gap of 38 or 39 missing bunches after a three or four batch injection, respectively. For safety reasons, it is necessary that the beam can be extracted from the LHC ring within one revolution. Therefore, a 3 μ s gap is included which accounts for the rise time of the LHC dump kicker magnet.

Besides colliding protons at $\sqrt{s} = 14$ TeV the LHC is also designed to collide lead ions at a center-of-mass energy of 5.52 TeV per nucleon. The lead ions are produced in an Electron Cyclotron Resonance Ion Source (ECRIS) [89]. The resulting lead ion beam mainly consists of Pb^{27+} and is accelerated within a RFQ, the LINAC3 and LEIR (Low Energy Ion Ring) [90] before being injected into the PSB and the following accelerators already mentioned above. After LINAC3 the Pb ions are passing a carbon foil where further electrons are stripped resulting in a beam of Pb^{54+} ions. All the remaining electrons are stripped after the PS where the lead ion beam is sent through a sequence of copper and nickel foils leaving Pb^{84+} ions. When reaching the LHC main ring the lead ion beam exhibits an energy of 177 GeV per nucleon.

2.1.2 The LHC Main Ring

The LHC main ring with a radius of roughly 4.25 km is installed in the tunnel that has been built for the predecessor-collider LEP and consists of eight arcs and eight straight sections. This layout, also illustrated in figure 2.3, was mainly driven by the high synchrotron radiation losses occurring in an electron-positron collider which had to be compensated with RF cavities installed in each of the straight parts. In general, synchrotron radiation is emitted when a relativistic charged particle is exposed to a magnetic field and forced to travel on a curved path. The energy loss per turn due to the emission of synchrotron radiation of a proton or an electron in a synchrotron is

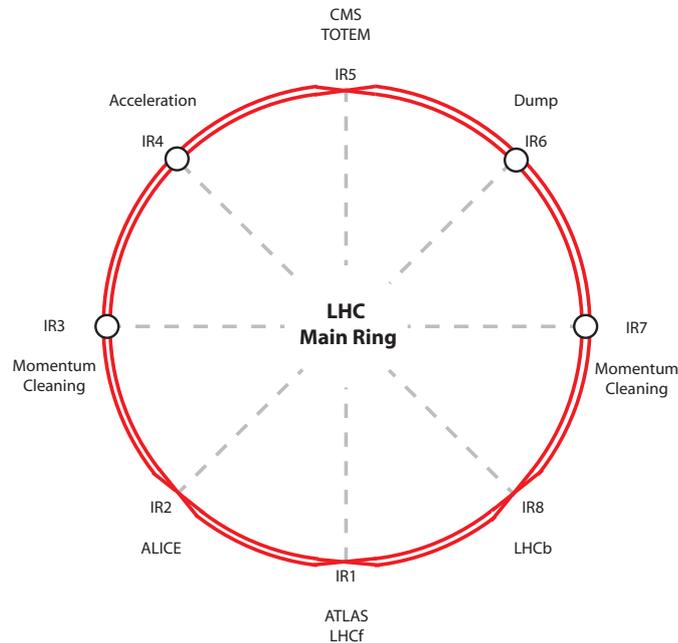


Figure 2.3: Layout of the LHC main ring consisting of eight arcs and eight straight sections. Each straight section provides a possible interaction point, labeled IR1-IR8, where the two proton beams could be brought to collision. In practice, the beams are only collided in the interaction regions 1, 2, 5, and 8. The utilization of the remaining interaction regions is also mentioned.

given by

$$\Delta E = \frac{e^2 \cdot \beta^3 \cdot \gamma^4}{\epsilon_0 \cdot 3R}, \quad (2.1)$$

where e represents the electric charge, $\beta = \frac{v}{c}$ is the particle velocity divided by the speed of light, $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{E}{m_0 c^2}$ is the relativistic Lorentz factor, ϵ_0 is the vacuum permittivity and R is the radius of the accelerator. As can be seen, this energy loss is roughly 13 orders of magnitude smaller for protons compared to electrons because of the nearly 1850 times higher proton rest mass. Hence, a proton accelerator like the LHC would ideally have larger arcs and smaller straight sections, however re-using the already built tunnel was the most cost-effective solution.

In principle, each of the straight sections serves as a possible interaction region (IR), labeled IR1-IR8, in which the two proton beams could be brought to collision. However, this is only done in IR1, IR2, IR5, and IR8 where the four already mentioned detectors ATLAS, ALICE, CMS, and LHCb are installed. The straight sections at the interaction regions 2 and 8 are also used for the injection of the two proton beams into the LHC ring. Momentum cleaning insertions are installed at IR3 and IR7, where particles with a large momentum offset or high betatron amplitudes are scattered and absorbed by collimators. At the straight section around IR4 in total 16 RF superconducting cavities — eight at each beam pipe — are installed which are used to capture, accelerate and store the proton beams. These cavities are built of copper sputtered with niobium

and are operated at a temperature of 4.5 K. They are delivering an electric field with a strength of 5.5 MV/m and an oscillation frequency of 400 MHz. In principle this frequency would allow for 35,640 possible RF buckets. However, due to the already mentioned bunch spacing time of at least 25 ns, at most every tenth bucket could be filled with a proton bunch. IR6 is devoted to the LHC beam dumping system which allows for the extraction of both beams within one revolution. In case of a dump, both beams are sent to special transfer lines and finally guided on a roughly 7 m long absorber mainly consisting of graphite.

The LHC accommodates 1,232 uniform dipole magnets with a length of 15 m and a diameter of 0.57 m, most of them are located in the eight arcs. They are providing the magnetic field which is necessary to keep the proton beams on their curved track along the ring. Due to the special layout of the accelerator with arcs and straight sections the bending radius of the dipoles, 2,804 m, is much smaller than the radius of the ring itself. Hence, for the nominal beam energy of 7 TeV a magnetic field strength of 8.33 T is needed. Such high magnetic fields can only be achieved using superconducting magnets. The LHC dipole magnets are built from NbTi superconducting coils which are cooled with superfluid helium to their operating temperature of 1.9 K. In addition, the space limitations in the already existing tunnel made it hard to install two separate beam pipes, which lead to the so-called two-in-one design of the magnets. Consequently both beam pipes are installed in the same magnet cold mass and the same cryostat leading to a complicated structure of the magnetic field. Furthermore thousands of multipole magnets, i.e. quadrupoles, sextupoles, etc., are installed along the LHC main ring to focus the beams and to correct their trajectories.

Luminosity

Besides the center-of-mass energy, the luminosity is the second very important characteristic of a collider. It is defined through the expected rate $\frac{dN_{\text{proc}}}{dt}$ and the cross-section σ_{proc} of a given process via

$$\frac{dN_{\text{proc}}}{dt} = \sigma_{\text{proc}} \cdot L , \quad (2.2)$$

where L represents the luminosity. Since most of the interesting processes at the LHC are expected to encounter small cross-sections, it is necessary to reach the maximally attainable luminosity. The luminosity of the LHC can be calculated through

$$L = \gamma \cdot f_{\text{rev}} \cdot \frac{N_p^2 N_{\text{bunch}}}{4\pi\epsilon_n\beta^*} \cdot F , \quad (2.3)$$

where γ is the relativistic Lorentz factor, f_{rev} is the revolution frequency, N_p is the number of particles per bunch and N_{bunch} is the number of bunches per beam. The normalized transverse beam emittance ϵ_n is a measure for the phase space volume which is associated with the transverse degrees of freedom of the beam. A factor β^* accounts for the amplitude of the betatron oscillation at the interaction point, and F is a geometric reduction factor which incorporates the beam crossing angles at the

interaction point. Such a crossing angle is needed to avoid parasitic collisions in regions that are not the nominal interaction point. At nominal proton operation the LHC is aiming for a peak luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ provided to the two large experiments ATLAS and CMS. Frequently, the performance of a collider is also represented by the amount of delivered, time integrated luminosity $L_{\text{int}} = \int L dt$, which is a measure for the amount of collision data that has been provided to the experiments.

LHC Incident and 2010 Operation

The LHC started its operation at September 10th in 2008 when first protons were injected and successfully sent around the ring. However, it had to be stopped again already nine days later after a serious incident had happened [91]. During powering tests of the main dipole circuit in sector 3-4, i.e. the region between IR3 and IR4, a resistive zone developed in the electrical bus between a dipole and a quadrupole. As a consequence an electrical arc developed which punctured the helium enclosure of the magnet, leading to a large release of helium into the insulation vacuum of the cryostat. Self-actuated relief valves on the helium enclosure were opened but they were not able to keep the pressure rise below a critical value, which led to large pressure forces displacing and damaging the magnets also in the neighboring subsectors. Hence, roughly 50 main LHC magnets had to be replaced and the electrical interconnections had to be repaired. Also a new protection system, consisting of thousands of detectors monitoring the resistance of the interconnection between each of two magnets and new helium pressure relief valves have been installed.

After the repair, the LHC resumed its operation in November 2009 and provided first proton-proton collisions at the SPS injection energy of 450 GeV per beam. However, for safety reasons it was decided to further increase the beam energy only to 3.5 TeV, i.e. half of the design energy. First proton-proton collisions at a center-of-mass energy of 7 TeV were achieved at the end of March 2010. In table 2.1 the design values of some LHC machine parameters are compared with their values at the time when most of the data analyzed in this thesis has been recorded. Until mid November 2010, when the proton operation finished and a short heavy ion operation period was started, the LHC delivered about 47 pb^{-1} to the two main detectors ATLAS and CMS. Figure 2.4(a) shows the development of the instantaneous luminosity with time during the 2010 proton run. In figure 2.4(b) the integrated luminosity delivered by the LHC as well as the amount of data recorded by the CMS experiment are depicted as a function of time.

2.2 The Compact Muon Solenoid Detector

The CMS experiment is a typical multi-purpose detector which is installed at the LHC IR5 in a cavern roughly 100 m below surface. In order to be able to reconstruct new particles created in the proton-proton collision in its center, CMS is built as hermetical as possible around the interaction point and aims for the detection of all

LHC machine parameter	design value	value in 2010
beam energy [TeV]	7.0	3.5
number of bunches	2808	368
number of protons per bunch	$1.15 \cdot 10^{11}$	$1.15 \cdot 10^{11}$
bunch spacing time [ns]	25	150
β^* [m]	0.55	3.5
ϵ_n [μm]	3.75	2.5
instantaneous luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	$1 \cdot 10^{34}$	$2 \cdot 10^{32}$
stored energy per beam [MJ]	362	24.4

Table 2.1: Comparison of the design values of the main LHC machine parameters with the values achieved in the 2010 operation.

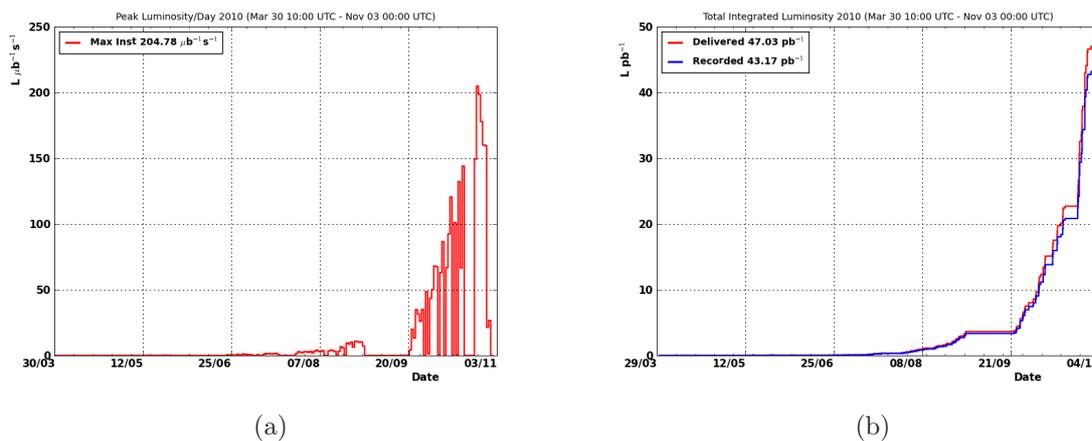


Figure 2.4: Luminosity profile in the 2010 proton operation. On the left hand side, the development of the instantaneous luminosity with time is shown. On the right hand side, the integrated luminosity delivered by the LHC (red line) and the integrate luminosity recorded by the CMS experiment (blue line) are compared.

the particles originating from a collision. With a total length of 21.6 m and a diameter of 14.6 m it is more compact than its competitor, the ATLAS detector, however with a weight of about 12,500 t it is almost as twice as heavy as ATLAS. The general design of CMS, depicted in figure 2.5, was mainly driven by its performance in detecting the SM Higgs boson but also by its ability to reconstruct the event signatures of decays of supersymmetric particles or new massive vector bosons. In summary, to meet the goals of the broad LHC physics program the following requirements have to be fulfilled: good charged particle momentum resolution and reconstruction efficiency, good muon identification and dimuon mass resolution, good electromagnetic energy resolution leading to a good diphoton and dielectron mass resolution, and hadron calorimeters with a large geometric coverage and a fine lateral segmentation. All this

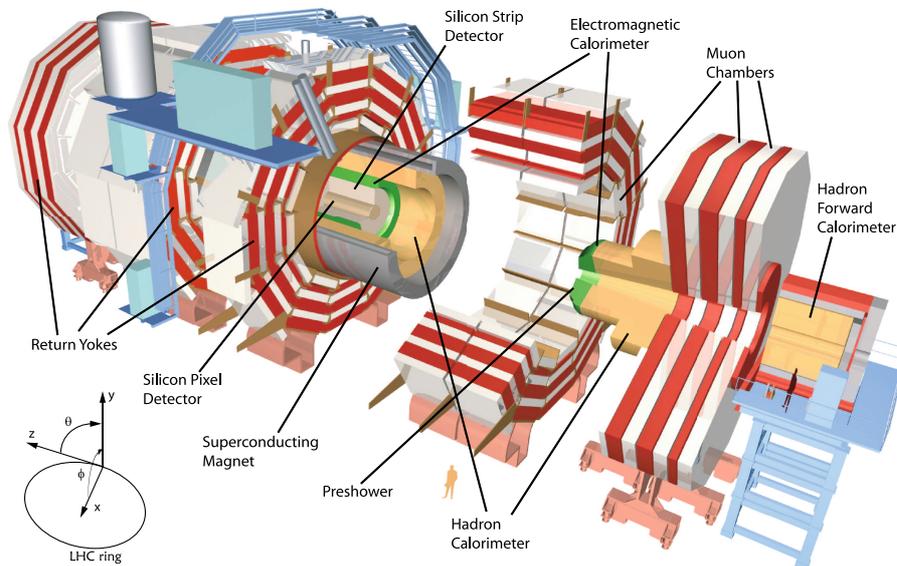


Figure 2.5: Overview of the CMS detector [92] and its subdetectors. The typical onion-like design with the tracking detectors close to the beam pipe surrounded by the calorimetry and finally the muon system is illustrated. On the lower left hand side the CMS coordinate system with respect to the LHC main ring is shown.

led to the onion-like layout of CMS where different subdetectors, each specialized in detecting different types of particles or in measuring their energies, are placed in a forward-backward and azimuthal symmetric way around the interaction point.

The heart of the CMS detector is a silicon-based tracking detector installed next to the beam pipe. When traversing the layers depleted of free charge carriers in the silicon pixel or strip detectors, charged particles are creating electron-hole pairs through ionization along their trajectory. An applied high voltage forces electrons and holes to drift towards the cathode and anode, respectively, which results in a measurable electrical signal. The tracking detector is surrounded by the calorimetry where almost all types of particles are absorbed completely and their energies are measured. One can distinguish between the electromagnetic calorimeter (ECAL), mainly absorbing particles subjected to electromagnetic interactions like e^+ , e^- , and photons, and the hadronic calorimeter (HCAL), where charged and neutral hadrons, which are primarily interacting via inelastic nuclear processes, are absorbed.

In the electromagnetic calorimeter, electrons and positrons lose their energy mainly due to the emission of Bremsstrahlung. Their energy after passing an absorber material of length x can be described by

$$E(x) = E_0 \cdot e^{-x/X_0} , \quad (2.4)$$

where E_0 is the initial energy and X_0 is the so-called radiation length which is defined as the distance a particle has passed when its energy is reduced to E_0/e . In contrast, high energetic photons exposed to the Coulomb field of the nucleus of the absorber

material are producing electron-positron pairs which are again losing their energy via Bremsstrahlung. The mean free path for photon induced pair production in matter is $\frac{9}{7}X_0$. Hence, the size of an electromagnetic calorimeter is frequently given in multiples of the radiation length. In the hadronic calorimeter, charged and neutral hadrons interact inelastically with the nuclei of the absorber material due to the strong force, first leading to excitation and finally to the break-up of the nucleus. Similarly to the radiation length a hadronic interaction length λ_I can be defined representing the mean free path of hadrons in matter. Commonly, calorimeters are built as so-called sampling calorimeters where layers of absorber material are altered with layers of scintillator that measure the energy of the particles traversing them. The scintillator material is excited by ionizing particles and emits photons when returning into the ground state. These photons are then recorded by photodetectors, e.g. so-called photo multiplier tubes (PMTs), which translate the light signal into an electric pulse which is a measure for the energy of the particle that has traversed the scintillator. However, as for the CMS ECAL, electromagnetic calorimeters can also be built in a homogeneous way, i.e. all of the calorimeter material acts as absorber and detector material at the same time which leads to a better energy resolution compared to sampling calorimeters.

Both, the tracking detectors and the calorimetry are embedded in a superconducting solenoid which is designed to produce a homogeneous magnetic field of 3.8 T pointing along the beam axis. Therefore, charged particles moving transverse to the beam axis are forced on curved trajectories and the radius of these trajectories can be used to determine the momentum of the particles as well as the sign of their charge. The magnetic flux is returned in a large iron yoke in which also the muon system is installed. Muons are minimum ionizing particles and due to their higher rest mass compared to electrons, they are losing only very little energy due to Bremsstrahlung. Thus, if neglecting neutrinos which do only interact weakly with the detector material, the only particles which can escape from the calorimeters are muons. A signal in any of the three types of gaseous detectors of the muon system is therefore a clear hint for the presence of a muon.

The CMS experiment is described by a right-handed coordinate system which is centered at the nominal interaction point in the middle of the detector. The y -axis is pointing vertically upwards and the x -axis is pointing radially inwards to the center of the LHC ring. Consequently the z -axis is pointing along the beam pipe. In the x - y -plane, the azimuthal angle ϕ is defined with respect to the x -axis and the radial coordinate in this plane is called r . The polar angle θ is measured from the z -axis in the y - z -plane. Typically, the kinematics of particles is described by a dimensionless variable, the so-called rapidity, which is defined through

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (2.5)$$

In this equation, E represents the energy of the particle and p_z the z component of its momentum. However, the variable frequently used to describe the geometry of the

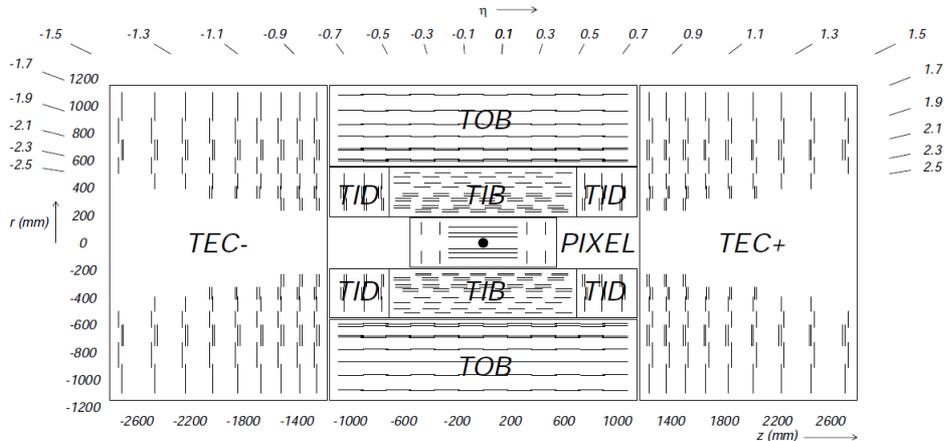


Figure 2.6: Schematic view of the CMS inner tracking system [81], consisting of silicon pixel detectors very close to the beam pipe and strip detectors at larger radii where the particle flux is reduced.

detector, is the pseudorapidity, which is defined as

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right). \quad (2.6)$$

2.2.1 The Inner Tracking System

The innermost part of the CMS detector is the inner tracking system [93,94] on which very challenging demands are made. On one hand side, it has to provide a very precise and efficient measurement of the tracks of charged particles as well as an accurate reconstruction of secondary vertices which are necessary for the identification of decays of long-lived particles, e.g. hadrons containing bottom or charm quarks. On the other hand, it is installed very close to the interaction point and thus has to withstand the very large flux of roughly 1,000 particles per bunch crossing assuming the LHC design luminosity. Moreover, this large particle flux will cause serious radiation damage to the detector material. All the mentioned requirements led to the design of a tracking system completely based on silicon detector technology. With an active area of roughly 200 m² it is the largest silicon tracking detector ever built and operated.

An illustration of the CMS tracking system is shown in figure 2.6. It has a length of 5.8 m, a diameter of 2.5 m and covers the region up to $|\eta| = 2.5$. The tracking system can be divided into two main parts: the pixel detector consisting of 1,440 silicon pixel modules mounted close to the beam pipe and the strip detector built from 15,148 silicon strip modules installed at larger radii where the particle flux is already reduced. Furthermore the strip detector can be divided in different subdetectors: The central part, also called barrel, consists of the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), and the Tracker Outer Barrel (TOB) which is surrounding the latter two. In the forward region, commonly referred to as the endcap, the Tracker EndCap

modules (TEC) are installed. In units of radiation lengths, the material budget of the CMS tracking detector is $0.4 X_0$ at $|\eta| \approx 0$, increases to about $1.8 X_0$ at $|\eta| \approx 1.4$ and finally decreases again to roughly $1 X_0$ at $|\eta| \approx 2.5$.

The Pixel Detector

The pixel detector is the part of the CMS detector which is installed closest to the interaction point. It is built from 66 million pixel cells with a size of $100 \times 150 \mu\text{m}^2$. Three barrel layers (BPix) with a length of 53 cm are located at radii of 4.4 cm, 7.3 cm and 10.2 cm. Two endcap disks (FPix) ranging from about 6 cm to 15 cm in radius are mounted at both sides at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm. With this design, for each trajectory at least three tracking points can be provided over almost the whole $|\eta| < 2.5$ region where in the outer part the information from the to FPix layers is combined with innermost BPix layer at $r = 4.4$ cm.

The expected spatial hit resolution is about $10 \mu\text{m}$ in the r - ϕ -plane and about $20 \mu\text{m}$ for the z direction. A reasonable agreement between these expectations and the measurement results have been found in cosmic muon [95] as well as in first collision data [96].

The Strip Detector

The CMS strip detector is built from about 9.6 million single-sided silicon microstrip sensors. As already outlined above, it is divided in different barrel and endcap parts. The TIB consists of four layers installed at radial distances of 25.5 cm, 33.9 cm, 41.9 cm and 49.8 cm covering the $z = \pm 70$ cm region. The first two layers are made from so-called stereo layers which consist of two single-sided sensors mounted back-to-back on each other with a stereo angle of 100 mrad. With stereo modules both, r - ϕ and also r - z measurements can be provided. In the TIB a single-point resolution between 23-34 μm in the r - ϕ -plane and 23 μm in the z direction is expected.

The TID consists of three small disks each housing three rings, centered on the beam pipe, on which silicon strips pointing towards the beam line are installed. On the first two rings again stereo modules are installed and as for the TIB the sensor thickness is 320 μm . The disks are mounted between $z = \pm 80$ cm and $z = \pm 90$ cm and the rings cover radii from roughly 20 to 50 cm.

The TOB is constructed from a wheel of six layers mounted at average radii of 60.8 cm, 69.2 cm, 78 cm, 86.8 cm, 96.5 cm and 108 cm and covers the region $z = \pm 109$ cm. Again the first two layers are equipped with stereo modules providing r - ϕ as well as r - z measurements. The expected single-point resolutions are varying between 32-52 μm in the r - ϕ direction and 52 μm in the z direction. Because of the reduced particle flux in the TOB region a larger sensor thickness of 500 μm is used.

Both TECs are consisting of nine layers on which the strip modules are mounted. The fifth layer is equipped with stereo modules and, similar to the TID, each module's front faces is pointing towards the beam line. They extend radially from 22 cm to

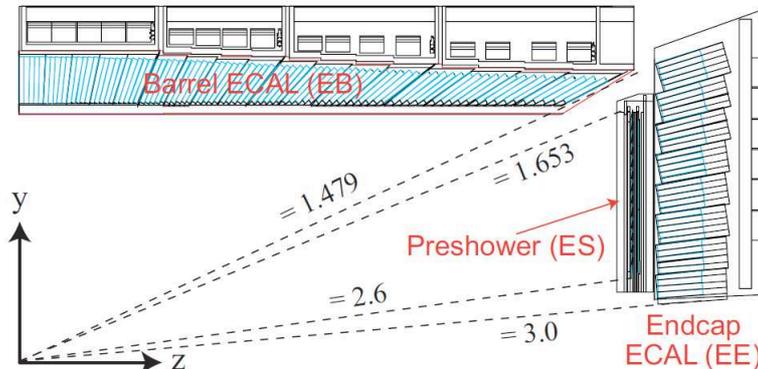


Figure 2.7: Overview of the CMS electromagnetic calorimeter [81] showing the ECAL barrel, one ECAL endcap and the preshower detector.

113.5 cm and cover the region between 124 cm and 280 cm along the z direction. As for the TOB, silicon sensors with a thickness of 500 μm have been used.

As for the pixel detector, the performance of the strip detector in first collision data can be found in [96].

2.2.2 The Electromagnetic Calorimeter

The CMS ECAL [97] is a homogeneous calorimeter which uses 75,848 lead tungstate (PbWO_4) crystals as absorber and active material at the same time. Light signals are recorded by avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) in the barrel part and the endcaps, respectively. One major criteria that led to the ECAL design was its performance in detecting possible $H \rightarrow \gamma\gamma$ decays as well as its energy resolution which is limiting the precision of the Higgs boson mass reconstruction. Hence, a homogeneous calorimeter design with its enhanced energy resolution compared to sampling calorimeters was a reasonable choice.

Several characteristics of the PbWO_4 crystals make them suitable for an operation in an LHC experiment [98]. Their high density of 8.28 g/cm^3 allows for the construction of a fast, radiation resistant calorimeter with a fine granularity. Furthermore the small radiation length of $X_0 = 0.89$ cm and the small Molière radius of $R_M = 2.2$ cm permitted a very compact calorimeter layout where R_M represents the radius of a cylinder in which 90% of the shower energy is contained. Moreover, at the 18 $^\circ\text{C}$ operating temperature of the ECAL, about 80% of the blue-green (wavelength 420-430 nm) scintillation light is emitted in 25 ns which corresponds to the nominal LHC bunch spacing time. However, the light output is small which means that only roughly 4.5 photoelectrons per MeV deposited energy can be collected in the APDs or VPTs.

An overview of the layout of the CMS electromagnetic calorimeter can be seen in figure 2.7. The pseudorapidity range $|\eta| < 1.479$ is covered by the ECAL barrel (EB). It consists of 61,200 PbWO_4 crystals with a length of 230 mm corresponding to 25.8 X_0 , where each crystal front face covers an area of 22×22 mm^2 . The front of

the crystals is mounted at a radial distance of 1.29 m with respect to the beam line. Two ECAL endcaps (EE) are covering the pseudorapidity range $1.479 < |\eta| < 3.0$ and reside at a longitudinal distance of 3.15 m away from the interaction point. Each endcap consists of so-called Dees equipped with 3,662 crystals which are grouped in mechanical units of 5×5 crystals, so-called super-crystals. Every crystal has a length of 220 mm, i.e. $24.7 X_0$, and covers an area of $28.62 \times 28.62 \text{ mm}^2$ at its front face. In front of the EEs a preshower detector (ES) is installed covering $1.653 < |\eta| < 2.6$. It is a two-layered sampling calorimeter using lead as absorber material and silicon strips to measure the energy deposits. Its principle aim is to identify neutral pions and also helps to distinguish electrons from minimum ionizing particles. The first absorber layer has a thickness of $2 X_0$ while the second layer exhibits only one radiation length.

The ECAL energy resolution is given by

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2, \quad (2.7)$$

where a is the stochastic term, b is the noise term, c is the constant term, and the energy E is given in GeV. The stochastic term mainly depends on fluctuations of the lateral shower containment, different photostatistics, and fluctuations of the energy deposited in the ES absorber with respect to what is actually measured in the silicon strip detectors. Noise from the electronics, the digitization, and from additional collisions in the same bunch crossing, so-called pile-up events, is contributing to the noise term. The constant term, which is dominant for large energies, is affected by non-uniformities of the longitudinal light collection, intercalibration uncertainties and energy leakage. In test beam measurements using electrons between 20 and 250 GeV, these terms have been found to be $a = 0.028 \sqrt{\text{GeV}}$, $b = 41.5 \cdot 10^{-3} \text{ GeV}$, and $c=0.003$ [99].

2.2.3 Hadron Calorimeter and Solenoid

The CMS hadronic calorimeter [100] consists of four main parts: the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeter. The HB and HE are placed within the magnet coil and thus the total amount of material which could be used to absorb hadronic showers is limited. Therefore, brass has been chosen as absorber material, because it has a short interaction length of 16.42 cm and is also non-magnetic. Plastic scintillator tiles in conjunction with wavelength-shifting fibers are used as active material and the photodetection is based on hybrid photodiodes (HPDs). The HO, consisting of a further layer of scintillator, resides on the outside of the CMS solenoid. In the very forward region between $|\eta| = 3$ and $|\eta| = 5.3$ the HF detector is installed. It is based on steel absorbers in conjunction with quartz fibers which provide a fast collection of the emitted Cherenkov light. The conceptual design of the CMS HCAL can be seen in figure 2.8.

The HB consists of two half-barrels each containing 18 azimuthal wedges which are constructed from brass absorber plates. Stainless steel is used for the first and last plate of the wedges to provide a larger structural strength. In total 17 layers of active

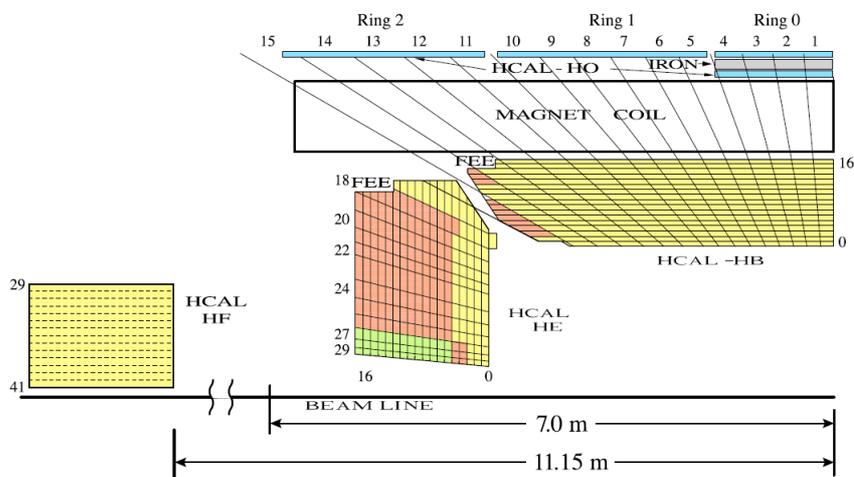


Figure 2.8: Drawing of the CMS HCAL [101] showing the arrangement of the HB and the HE within the magnet coil as well as the HO scintillators outside the magnet’s cryostat. On the left hand side, the very forward part of the CMS HCAL, the HF, is depicted.

material are installed between the different absorber plates where the first scintillator layer is placed already in front of the steel plate to be sensitive to hadronic showers developing in the region between the EB and the HB. With this design a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ could be reached. The HB ranges from a radial distance of 1.77 m to 2.95 m and covers the pseudorapidity region of $|\eta| < 1.3$. Its effective thickness in terms of interaction length is $5.82 \lambda_I$ at $\theta = 90^\circ$ which increases with decreasing angle θ like $1/\sin\theta$. The design of the HE is similar to the HB design explained above. It covers the pseudorapidity region $1.3 \leq |\eta| < 3.0$ and its thickness is roughly $10 \lambda_I$ (including the parts of the electromagnetic calorimeter installed in front of it). The granularity of the HE is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in the region $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ for $1.6 \leq |\eta| < 3.0$.

All the detector components explained so far are surrounded by a superconducting solenoid magnet whose design was mainly driven by the momentum resolution of $\Delta p/p \approx 10\%$ for muons with a momentum of $p = 1 \text{ TeV}/c$ which should be achieved with the CMS detector. The 12.5 m long solenoid has a diameter of 6 m and consists of a 220 t cold mass with a 4-layer winding of NbTi conductor operated at 4.45 K. It provides a magnetic field of 3.8 T and its thickness in terms of interaction length is given through $1.4/\sin\theta$.

Especially in the central region, $|\eta| < 1.3$, the combined EB plus HB absorber material is not sufficient to stop all of the hadrons. Hence the HO, a further layer of scintillator, is mounted on the outside of the solenoid’s vacuum tank benefitting from its additional material budget of at least $1.4 \lambda_I$. At the center around $|\eta| = 0$ an additional iron plate with a thickness of 19.5 cm is installed behind the first scintillator layer and

is followed by a second layer with active material. Consequently the calorimetry of the CMS detector has a minimum depth of roughly $11 \lambda_I$ over the whole pseudorapidity range.

The design of the HF was mainly driven by the necessity of residing in a region with a very large particle flux. Its modules are located at a longitudinal distance of $|z| = 11.2$ m and it consists of cylindrical steel structure with an inner radius of 12.5 cm and an outer radius of 130.0 cm which contains fibers as active material. A signal is generated when charged particles traversing the quartz fibers produce Cherenkov light which is then transported to photomultiplier tubes.

The performance of the CMS hadronic calorimeter has been studied with test beam measurements, cosmic muon data as well as first LHC beams [101]. Its energy resolution is given by

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + c^2, \quad (2.8)$$

where a represents the stochastic term, c the constant term and the energy E is measured in GeV. Both values have been measured: $a = 0.847 \pm 0.016 \sqrt{\text{GeV}}$ and $c = 0.074 \pm 0.008$ for the central detector parts and $a = 1.98 \sqrt{\text{GeV}}$ and $c = 0.09$ for the HF [101].

2.2.4 The Muon System

The CMS muon system [102] is built from three different types of gaseous detectors: drift chambers (DTs) used in the central barrel region, cathode strip chambers (CSCs) assembled in the endcap region, and resistive plate chambers (RPCs) employed in both detector regions. It is housed between the different parts of the solenoid's flux-return iron yoke which also serves as hadron absorber. The choice of different detector types in the mentioned detector regions was mainly driven by the strength of the expected, neutron-induced background. This background arises from spallation neutrons which are created during the absorption of hadron showers in the calorimeter and is thus expected to be larger in the endcap regions. Therefore the installation of CSC detectors with their fast response time, their fine segmentation, and their high radiation resistance was a reasonable choice in this detector region. In the barrel region however, it is possible to employ DTs because of the lower expected muon rate and the low local magnetic field strength. Overall, the muon system comprises an effective detector area of 25,000 m². Hence, it also has to be built in a cost-effective, reliable, and robust way. The design of the muon system is illustrated in figure 2.9.

Each of the three parts of the return yoke surrounding the inner barrel part of CMS consists of 12 sectors in the azimuth angle ϕ . In every sector, there are four layers of drift-tube chambers, so-called stations, which are labeled MB1-MB4: The first one resides outside the magnet cryostat, the second and third layer are placed in gaps between the three parts of the yoke and the fourth is installed on the outside of the third part of the yoke resulting in four concentric cylinders of muon detectors. Each drift-tube chamber is built of three superlayers (SLs) which can be further divided

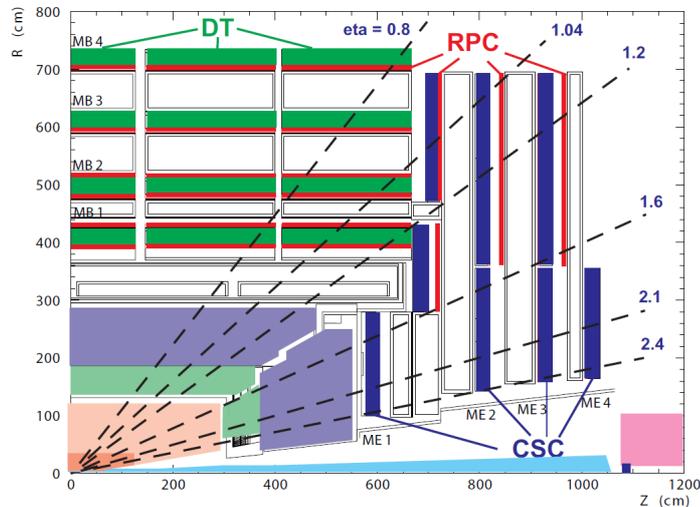


Figure 2.9: Layout of one quarter of the muon system [81] showing the arrangement of the DT in the barrel region, the CSCs in the endcap region and the RPCs in the region $|\eta| < 1.6$.

into four layers of rectangular drift cells. These drift cells are covering an area of $13 \times 42 \text{ mm}^2$ and are filled with a gas mixture of 85% Ar and 15% CO_2 yielding a maximum drift time of 380 ns. The anode wires of the drift cells in the two outer SLs are arranged parallel to the beam line, hence providing measurements of ϕ . To also provide a measurement in the z direction, the wires in the innermost SL are installed perpendicular to the beam line. However, in MB4 those SLs are missing which therefore only provides a ϕ measurement.

In the endcap region, $0.9 < |\eta| < 2.4$, the muon systems consists of 468 cathode strip chambers which are arranged in eight groups labeled ME1/1-3, ME2/1-2, ME3/1-2 and ME4/1. The CSCs have a trapezoidal shape and are covering 10° or 20° in ϕ , respectively. Except for ME1/3, they are installed in an overlapping geometry to provide a full coverage in the azimuthal angle. Each CSC is a multiwire proportional chamber built from six planes of anode wires which are enclosed between seven layers of cathode panels. The cathode strips are milled into the cathode panels each covering a fixed $\Delta\phi$ region. An r measurement is provide by the anode wires which are installed at fixed radii. As explained in more detail in [103], the measurement of the ϕ coordinate can be obtained by measuring the center of gravity of the pulses induced on different cathode strips.

The additional installation of RPCs is motivated by their good time resolution which is comparable to that of scintillators. Since these chambers are able to determine the timing of a detector signal in a much shorter time than the LHC nominal bunch spacing of 25 ns, RPCs are a powerful tool for the unambiguous assignment of detector signals to a specific bunch crossing. These characteristics are very well-suited for the construction of fast muon triggers (see 2.2.5 for more information on the CMS trigger system). In the barrel region, two RPC arrays are surrounding the DTs of the first and

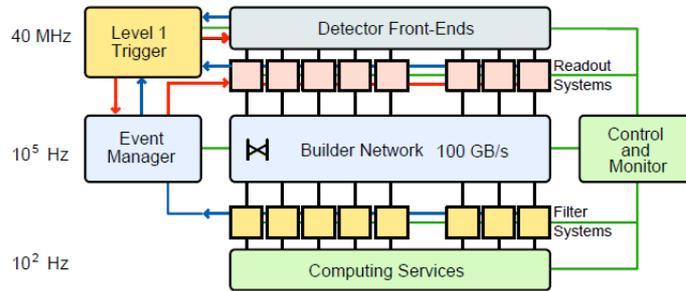


Figure 2.10: Architecture of the CMS DAQ system [82] showing the two event rate reduction steps. Roughly 100,000 events per second are passing the L1-trigger requirements and are further processed in the event builder where the information of different front-end drivers (FEDs) is combined into a single event. From the event builder the events are handed over to the HLT where the final decision on writing an event to mass storage is taken based on complex event reconstructions.

second muon station. At the third and fourth station, both RPC arrays are installed in front of the drift chambers. As can be seen in figure 2.9, four RPC arrays are installed together with the CSCs in the endcap region covering the pseudorapidity range up to $|\eta| = 1.6$.

With cosmic muon data recorded in autumn 2008, the performance of the CMS muon system has been measured [104,105]. For the DTs, a single hit resolution of the order of 260 μm and a hit reconstruction efficiency of close to 100% has been found. The CSCs are found to have a coordinate resolution between 47 μm and 243 μm and single hit reconstruction efficiencies above 99%. Also first experiences on the operation and the performance of the RPCs in the barrel region could be gained [106].

2.2.5 Trigger, Data Acquisition, and Computing Model

At nominal operation, the LHC will provide proton-proton collisions with a frequency of 40 MHz. However, this rate is much higher than the rate at which the detector signals can be read out, the data can be processed, and finally written to mass storage. Consequently, this high rate has to be reduced to a processable rate of roughly 100 Hz by selecting only events with an interesting physics content. This selection is done by the CMS trigger system and is also the first step in any data analysis that might follow. Since a reduction of the rate of five orders of magnitude is not feasible in only one step, CMS is using a two-stage procedure. The first stage is called L1-trigger [107] and is responsible for a first rate reduction to roughly 100 kHz. It is built from custom-designed electronics which can be programmed flexibly. The second stage is the High Level Trigger (HLT) [108] which is based on a software system running on a farm of roughly thousand commercial processors. Together with the whole set of detector front-end electronics and many controlling and monitoring devices, the L1-trigger and the HLT represent the CMS Data Acquisition (DAQ) system. An overview of its architecture is shown in figure 2.10.

There are two types of L1-trigger, muon trigger and calorimeter trigger, each consisting of local, regional, and global components. The local components are sensitive to energy deposits in the calorimeters, track segments, or hits in the muon chambers. Out of these information, the regional components build trigger objects like electron or muon candidates in restricted spatial regions. The Global Muon Trigger (GMT) or the Global Calorimeter Trigger (GCT), respectively, are determining those trigger objects, which exhibit the highest level of confidence attributed to the measured parameters over the whole detector region. These objects are then handed over to the Global Trigger that makes the final decision on keeping or rejecting a specific event at the L1 stage.

The maximum allowed latency due to the L1-trigger decision is 3.2 μs . During this time, the data is kept in dedicated front-end buffers. If an event has passed the L1-trigger, the data of the different subdetectors is extracted from the buffers and sent via front-end drivers (FEDs) to the readout units. An event builder is in charge for the combination of the data from the different FEDs into one single event which is then sent to a single filter unit where, amongst others the HLT event selection is run. The HLT has access to all detector signals of one event and is able to perform complex event reconstruction procedures by running specialized fast versions of the off-line reconstruction software. The implementations of the different HLT algorithms and selections is a rapidly evolving field. For example, they are depending on the instantaneous luminosity that determines the rate at which a given physics process will occur, but also on the definition of an event with an interesting physics content mentioned above. In any case, the maximum HLT output rate is determined by the rate at which the data can be archived and written to mass storage.

Even after the immense event rate reduction achieved by the CMS trigger system, the CMS experiment alone will have to store and process roughly 150-200 MB/s of data. Together with the three other detectors, around 700 MB/s or roughly $15 \cdot 10^6$ GB per year have to be stored safely. Moreover the data have to be made available to thousands of physicists from all around the world who are depending on a fast access to it while doing their analyses. The mentioned requirements led to the usage of decentralized mass storage and computing resources.

The Worldwide LHC Computing Grid (WLCG) project [109] is a collaboration of the LHC experiments, computing centers, and middleware providers which facilitates a four-tiered hierarchical computing structure. The Tier-0 center is located at CERN and is responsible for securing a copy of the raw data, performing a prompt reconstruction of the incoming data, and spreading the datasets to the corresponding Tier-1 centers. The duties of Tier-1 centers are the safe storage of reconstructed data and simulated event samples. They also provide computing resources for the re-reconstruction of collision data with an improved detector calibration or reconstruction software and large scale data preselections, so-called skimming jobs for different physics groups. The responsibility of Tier-2 centers is to provide computing resources for the data analysis in a given geographic region and the production of simulated events for the whole collaboration. Local Tier-3 centers are also connected to the grid, however, they

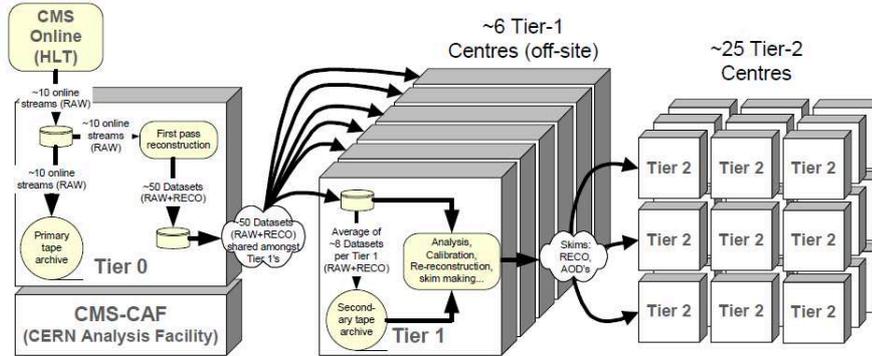


Figure 2.11: Overview of the layout of the Worldwide Computing Grid and the CMS Computing Model [82]. The raw format data from the detector is stored at the Tier-0 and a first reconstruction is done. RAW and RECO datasets are then transferred to Tier-1 sites. Datasets in the AOD format which occupy only small amounts of storage are finally distributed to Tier-2 centers where the computing resources for final data analyses are available.

are designed to provide computing and storage resources only for local analysis groups.

The CMS Computing Model [110] is based on the WLCG layout explained above. Data events passing the HLT are grouped into several, so-called primary datasets according to the trigger requirement they have fulfilled. These datasets of raw format data events (RAW) contain all the information which have been read-out from the detector as well as a record of the L1-trigger and HLT decisions. At the Tier-0, the CMS reconstruction software is run which produces datasets in the so-called RECO format. The RECO datasets contain the reconstructed physics objects as well as the inputs which have been used to produce them. A copy of each RAW dataset is kept at the Tier-0 site and several copies of RAW and RECO samples are distributed to Tier-1 centers. The third data format, the Analysis Object Data (AOD), contains only the informations essential for physics analysis. With an event size of roughly 0.05 MB, it is small enough to be widely spread over different Tier-2 sites. Figure 2.11 schematically summarizes the WLCG layout and the CMS Computing Model.

Chapter 3

Generation, Simulation, and Reconstruction

In order to be able to assign the hundreds of particles originating from a typical proton-proton collision at the LHC to an interesting physics process, a deep understanding of the hard scattering process, the decay and the hadronization of the produced particles, as well as their interaction with the detector material is necessary. While the interaction between two partons with a large momentum transfer can be calculated with the perturbative methods explained in chapter 1, the description of the behavior of proton remnants depends on phenomenological models. After their production in the hard process, the decay of unstable particles and the hadronization of quarks has to be taken into account correctly. Physicists therefore employ program packages based on Standard Model calculations and phenomenological models, so-called Monte Carlo (MC) generators, which are able to simulate particular physics processes and make predictions on the number of produced particles and their kinematics.

Based on MC generator predictions on the type and the kinematics of the produced particles, the detector simulation software models the interaction of the particles with the detector material. Besides the energy deposition in the different sub-detectors also the detector read-out electronics are simulated in this step, resulting in a set of electric signals similar to the signature observable in real collision events.

For any analysis of collision events it is a crucial prerequisite to reconstruct high-level physics objects like electrons, muons, and jets from the detector output signals. For this reconstruction dedicated algorithms are used. The reconstruction software is applied to real collision data as well as simulated events. Afterwards, the simulated datasets can be used to get a first impression on how a particular physics process would look like in the detector and can help to identify these processes in real proton-proton collisions. Moreover, not only the reconstructed objects but also the generated particles are accessible in a simulated event.

3.1 Event Generation

The idea of using Monte Carlo techniques [111] for solving physical problems goes back to Enrico Fermi who applied them in his calculation of neutron diffusion in the 1930's. In the following years, many Monte Carlo algorithms have been developed by Stan Ulam and John von Neumann who realized the great potential of this idea [112]. It was von Neumann who finally gave the technique its name Monte Carlo, after the famous quarter of Monaco where the casino is located.

In high energy physics, the basic idea is to simulate the randomness in the behavior of quantum mechanical particles, by using Monte Carlo techniques to draw the values of all relevant observables according to their probability density distributions. On average, the distributions and fluctuations in the simulated events should thus be the same as in real collision data. Since it is difficult to simulate the whole quantum mechanical process in one step, in general, a factorized approach is used. First the radiation of particles, such as gluons or photons, from the partons of the two beam particles are considered. Then the hard interaction of two partons is calculated. If resonances with a short life time have been produced, also their decay is simulated. Following this, the radiation of particles in the final state is taken into account. In the next step, the softer interactions of other partons of the beam particles and the behavior of the beam particle remnants are simulated. Finally, the hadronization of quarks, i.e. the formation of baryons and mesons as well as their decays, is computed. The particles arising from the hadronization of a quark form a stream of collimated hadrons, which is called a jet. An illustration of the different steps in the event generation process can be seen in figure 3.1.

Hard Scattering Process

In case of the LHC, the two incoming protons are characterized by PDFs, describing the momentum fraction carried by the different partons. Since the hard scattering between two partons occurs with a large momentum transfer, the partonic cross-section can be calculated perturbatively. If heavy resonances, such as W bosons or Z bosons for example, are produced in the interaction, their decay into quarks and leptons is also treated in the simulation of the hard process.

Initial- and Final-State Radiation

All processes containing electric or color charged particles, give rise to the radiation of gluons or photons in the initial- or final-state, i.e. before or after the hard scattering process. Since these radiation processes might lead to large corrections to the overall event topology, it is crucial to treat the effect in a reliable way. Two different approaches exist for the handling of initial- and final-state radiation processes. The first approach which is called the matrix element technique, computes all relevant Feynman diagrams, including real gluon or photon radiations at LO. The second approach, called

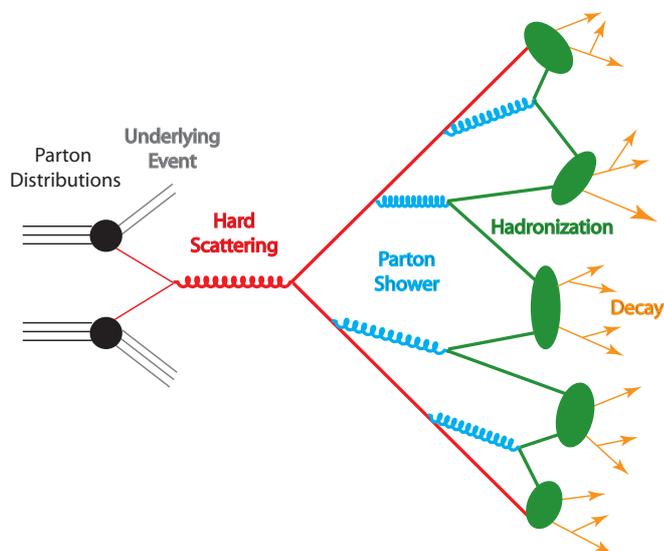


Figure 3.1: Illustration of the different steps of a typical event generation process. The characteristics of the beam particles, hadrons in this case, are described by parton distribution functions. Two partons interact in the hard interaction, while the beam particle remnants give rise to the underlying event. In this illustration, initial- and final-state radiation processes are modelled through parton showers. After the simulation of hadronization process, unstable particles are decayed.

parton shower, implements an arbitrary number of branchings of one parton into two particles.

In principle the matrix element approach is the more convenient one, because the exact kinematics, the full interference, and the helicity structure are taken correctly into account. However, for higher orders in perturbation theory, the calculations of the Feynman diagrams become more and more difficult. Furthermore, the matrix element method diverges for the radiation of soft or collinear partons. It is therefore only reliable in the prediction of states with well-separated, highly energetic parton radiations.

Due to the simplifications made in the derivations of the kinematics, the interference, and the helicity structure within the parton shower approach, the number of involved partons is in principle not limited. The parton radiation is modelled through successive parton emissions which is only fully correct for the limit of soft and collinear emissions. Hence the parton shower approach provides a good approximation for these kinds of emissions, but the model cannot describe hard and widely separated gluon radiations.

It has been shown, that both mentioned approaches provide reliable results in different phase space regions. Therefore, it is reasonable to combine the different methods to obtain a valid description of initial- and final-state radiation processes. The combination of both approaches avoiding double counting or uncovered phase space regions is challenging. Nevertheless, several matching algorithms have been proposed, where the CKKW [113,114] and the MLM [115] matching schemes are the frequently

used ones.

In the CKKW matching prescription, the domains of applicability of matrix element and parton shower simulations are clearly separated by the relative transverse momentum of two final-state particles. Two particles are clustered into the same jet, if their relative transverse momentum is smaller than a predefined critical value. Additionally, the transverse momenta of each jet also have to be larger than the critical value. After the clustering, appropriate weights are applied to the matrix elements to take into account terms which would appear in the corresponding parton shower evolution. Finally, every emission harder than the critical value is rejected during the parton shower evolution.

The MLM matching procedure which is used for most of the simulated samples used in this thesis, defines events at parton level by requiring a minimal parton E_T and a minimal separation (R_{\min}) between two partons in the η - ϕ -plane. The events are passed over to the parton shower where, in contrast to the CKKW prescription, no hard emission veto is applied. After the shower has been performed, all of the final-state particles which are found within a circle of radius R_{\min} in the η - ϕ -plane are combined into one jet. If the distance between an original parton and a jet is smaller than R_{\min} the parton and the jet are said to match. An event is kept for further processing, if all partons are matched to a jet and no additional jets are present. All other events are rejected.

A comparison of the two matching procedures for different processes and different observables relevant for LHC physics can be found in [116].

Underlying Event and Pile-Up

If composite objects like protons are collided, not only the partons which take part in the hard scattering process have to be taken into account. There are also contributions from the beam particle remnants which are color-connected to the hard scattering process and thus belong to the same hadronization system. Here, hadronization denotes the process of producing colorless objects from colored partons. Furthermore, there is the possibility of interactions between partons involved in the hard scattering process and partons from the beam particle remnants, a process called multiple interactions. Also the interactions between two partons from the beam particle remnants have to be taken into account. All the mentioned processes contribute to the so-called underlying event. At colliders with high instantaneous luminosities like the LHC, there is also the possibility of multiple interactions between different beam particles in the same bunch crossing. These additional events are called pile-up.

Hadronization

Unfortunately, the process of forming observable colorless objects from colored partons is a long range interaction occurring at low momentum transfers. Hence, the QCD coupling constant α_S is large, and this process cannot be calculated using perturbative methods. Instead, the process has to be described using phenomenological

models. There are two main hadronization models, the string fragmentation or Lund model [117, 118] and the cluster fragmentation model [119].

In the Lund string model a quark-antiquark pair is connected through a massless relativistic string describing the QCD field. A constant tension is given to this string which leads to a linear potential. When the two quarks are moving apart, the string gets stretched and thus the potential energy stored in the string increases. If the energy is large enough to produce a new quark-antiquark pair, the string can break up into smaller pieces connecting, for example, one of the original quarks with a newly produced one. This interplay of quark-antiquark pairs moving apart and the break up of strings connecting them evolves until only color-singlet hadrons remain. In this model, gluons are treated as internal excitations of the string. The production of baryon-antibaryon pairs is included through a so-called Popcorn model [120].

The cluster fragmentation model is based on a property of perturbative QCD which is called preconfinement [121]. The basic characteristic of preconfinement is that partons which are close in color are found to be close in coordinate and momentum space, too. Partons generated in the different branching processes are thus forming color-singlet clusters. Furthermore, preconfinement states that the masses of these clusters are independent of the hard interaction processes and its center-of-mass energy. Based on this, the cluster fragmentation model regards these clusters as highly excited hadrons which decay subsequently into the observed hadrons.

Most of the hadrons produced in the fragmentation process have only short lifetimes and their decay into the hadrons which finally interact with the detector material has to be modelled. It is therefore necessary to include all particles with a proper mass distribution and decay properties into the hadronization framework. In general this is very challenging, because different experimental results have to be combined. Different event generator packages implement different solutions for the description of these decays and depending on that, for example the spin information is preserved or not.

3.1.1 MC Event Generators

The event generation explained above is implemented in different Monte Carlo event generator programs. The available packages can be subdivided into three different groups. The first set of programs, like PYTHIA [14] and HERWIG [122], simulate the whole evolution of an event, i.e. the hard scattering process, the parton shower, and the hadronization and decay. Matrix element generators, such as MADGRAPH/MADEVENT [13] and ALPGEN [123] for example, include initial- and final-state radiations in LO matrix elements but do not provide any simulation of the hadronization process. In general, these types of programs have to be interfaced to packages of the first category for the hadronization step to obtain valid predictions. The challenges arising through the combination of parton shower and matrix element generators have already been discussed above. A third type of generators, like MC@NLO [124] and POWHEG [125] for example, comprise NLO calculations for the

hard scattering process but also do not provide a model for the hadronization step.

The simulated samples used in this thesis are almost exclusively produced with MADGRAPH/MADEVENT interfaced to PYTHIA. Hence, only these two generator packages are discussed in more detail in the following.

MadGraph/MadEvent

MADGRAPH/MADEVENT [126] is a multi-purpose LO event generator which is powered by the matrix element creator MADGRAPH (MG) [127]. For a given process, MG automatically calculates the amplitudes for all relevant subprocesses. The results of these calculations are then handed over to MADEVENT which provides the integration of the squared amplitudes over the whole phase space using Monte Carlo techniques. In this manner, the cross-section for a given process can be calculated and unweighted MC events can be obtained.

The resulting events are stored in a common event format, the Les Houches Event file (LHEF) [128], which allows for further processing the events with parton shower and hadronization program packages like PYTHIA. To avoid double counting of particular phase space regions, the MLM matching procedure is applied when combining the matrix element generator output with the parton shower simulations.

Pythia

The PYTHIA package is a program for the generation of events originating from the collisions of highly energetic particles like e^+ , e^- , p , and \bar{p} in various combinations. It contains theories and models for hard and soft interactions, parton distributions, initial- and final-state radiation, underlying event, and hadronization. In PYTHIA the initial- and final-state radiation processes are modelled by parton showers which are treated as a sequence of subsequent branchings of one original particle into two. The probability for parton to branch is given through the DGLAP QCD evolution equations [129–131]. During the showering process, each parton is characterized by a virtuality scale Q^2 which defines in some sense also a time ordering for the shower cascade. The choice of Q^2 is arbitrary in some sense, but for the simulated events used in this thesis, a p_T -ordered shower with $Q^2 = p_T$ has been used. Here p_T denotes the transverse momentum of the particles after the branching with respect to the momentum direction of the original parton.

In final-state showers, the Q^2 values are subsequently decreasing. Starting from a maximum value, the partons are evolved downwards in Q^2 (forwards in time) until a branching occurs. The current Q^2 values then describes the p_T of the branching and the resulting particles may now, in turn, evolve downwards until a lower boundary for Q^2 is reached. For QCD branchings this boundary is typically set to 1 (GeV/c)².

For initial-state showers, a method called backward-evolution is utilized. First the hard scattering process is selected, then the initial-state radiation is reconstructed by moving backwards in time. Such an approximation is possible, if evolved parton distribution functions are employed, containing the inclusive summation over all potential

initial-state shower processes that could lead to selected hard interaction. A particular initial-state radiation process is then chosen by selecting an exclusive shower evolution from the inclusive set.

In PYTHIA the hadronization process is modelled using the string fragmentation model. The parameters modelling the underlying event in PYTHIA are adjusted to describe the underlying event measured in real collision data. Different so-called tunes are available which have been derived using data from the Tevatron experiments (tune DWT and D6T [132]) as well as first data from the LHC experiments (tune Z1 and Z2 [133]).

Tauola

TAUOLA [134] is a program package for the precise simulation of τ -lepton decays which can be interfaced to MC event generators. It provides the possibility of correctly taking into account the effects of τ spin correlations in the distributions of the τ decay products. For most of the simulated events analyzed in this thesis, the TAUOLA package has been used for the simulation of τ -lepton decays.

3.2 Detector Simulation

Two different approaches exist for the simulation of the interaction of particles with the CMS detector material. First, there is a detailed full detector simulation that is based on the GEANT4 [135] toolkit. The second type of simulation, called fast simulation [136], achieves smaller computing times by using a simplified version of the detector material geometry as well as simplified models to describe the interaction of particles with the detector material.

GEANT4 is a simulation package which is able to model the traverse of particles through matter. It is thus well-suited for a detailed simulation of electromagnetic and hadronic interactions of generated particles with the CMS detector material. Moreover, GEANT4 provides the possibility for a detailed implementation of the CMS sub-detector geometries and the magnetic field which is essential for obtaining reliable predictions.

The events that have been produced using MC generators are handed over to one of the mentioned CMS detector simulations. Considering the magnetic field, the final-state particles are then propagated to the different sub-detectors where they might interact. During this propagation, particles are allowed to decay according to their known lifetimes, branching fractions, and decay kinematics. Secondary particles, produced in an interaction of an original particle with the detector material or stemming from particle decays, are also propagated and decayed, respectively.

The fast simulation approach achieves a roughly 100 times smaller computing time compared to the full simulation and reaches an almost as good as accuracy. Hence, it provides a useful alternative for analyses depending on large amounts of simulated

events. A comparison of simulated events passed through the CMS fast simulation with real collision data at $\sqrt{s} = 7$ TeV can be found in [137].

3.3 Event Reconstruction

Ideally, the hard scattering process can be identified from the topology of the particles which finally interact with the detector material. Hence, the starting point of any analysis is the reconstruction of high-level physics objects from the raw detector signals. For the CMS detector, this process starts with the reconstruction of tracks of charged particles based on signals in the silicon pixel and strip detectors. The combination of reconstructed tracks with detector signals from the electromagnetic calorimeter and the muon chambers allows for the reconstruction of electrons and muons, respectively. In the last step, jets of collimated hadrons arising from the hadronization of quarks or gluons are reconstructed. Traditionally, the jet reconstruction is based on signals from the calorimeters only. However, for the CMS detector also a new method is used which includes information from the tracking detectors.

3.3.1 Track Reconstruction

Charged particles traversing the CMS tracking system deposit energy in the different detector layers. Each energy deposit is called a hit in the following. By grouping the hits on the different layers, the trajectories of charged particles can be reconstructed. At the CMS detector, this reconstruction is performed by the combinatorial track finder (CTF) [138] which is based on a Kalman filter (KF) [139] for the track building and the estimation of the track parameters [140].

Tracks are described by five different parameters which are defined at the point of closest approach of the track to the beam axis, i.e. at the impact point. Two parameters are given by ϕ and $\cot \theta$, where ϕ and θ are the azimuthal and polar angle of the track momentum at the impact point, respectively. A further parameter is the transverse momentum p_T of the track. The two remaining parameters describe the coordinates $d_0 = y_0 \cdot \cos \phi - x_0 \cdot \sin \phi$ and z_0 of the impact point in the transverse and the longitudinal plane, where x_0 and y_0 are the transverse coordinates of the impact point. The distances between these coordinates and the beam axis are called the longitudinal and the transverse impact parameters (IPs), respectively.

The CTF algorithm consists of four logical parts: the generation of track seeds, the trajectory building, the resolution of ambiguities, and the final track fit. Track seeds exclusively built from hits in the three pixel layers, are necessary to provide starting values, sufficiently close to the true values. Furthermore, they are used to obtain reasonable uncertainties for the track parameters which allow for the use of a linear fitting algorithm like the Kalman filter. Seeds can be constructed from three hits in the pixel detector alone or from two hits in the pixel detector by using the beam spot as a third point [141]. Here, the beam spot is defined as the profile of the luminous region where the LHC beams collide within the center of the CMS detector.

The Kalman Filter based trajectory building in the strip detector starts from the coarse estimate of the track parameters provided by the track seed and includes iteratively the subsequent detection layers. With each inclusion of a hit the uncertainties on the track parameters become smaller. First the layers compatible with the track seed are identified and the track is then extrapolated to this layer by taking into account the magnetic field, multiple scattering processes, and the energy loss in the traversed detector layers. Since it is possible to find multiple hits, compatible with the predicted hit within the uncertainties of the track parameters, each of these hits is used to create a new trajectory candidate. It is also possible, that a charged particle does not produce a hit in a particular detector layer. Therefore, also a track candidate without a measured hit is created. All track candidates are then extrapolated to the subsequent layer. A track candidate is removed if on two subsequent detector layers no matching hits can be found. This algorithm depends on different parameters which can be adjusted to provide a highly efficient or a less computing time intensive track finding. Thus, the algorithm can also be used for a fast track reconstruction which is essential for the usage of tracks in the HLT decisions.

Ambiguities in the reconstruction of tracks might arise, if a particular trajectory is reconstructed from different seeds, or if a given seed leads to the reconstruction of multiple tracks. If the number of hits common to any two tracks divided by the number of hits of the track with the smaller number of hits is larger than 50%, the track with the larger χ^2 in the track fit is removed.

Finally, to avoid impacts on the trajectory reconstruction which might arise from the seeding step, all tracks are refitted using a least square approach. Here the Kalman filter starts at the location of the innermost hit and iteratively includes all hits associated to the track. Using the results of this filter, a second Kalman filter is initialized at the outermost detection layer which propagates towards the beam line. For each hit, the track parameters are combined with the track parameters predicted from the first filter, yielding an optimal estimate for the parameters.

All steps from the track seeding to the final track fit are repeated different times in an iterative tracking procedure. First, tracks are seeded and reconstructed using stringent quality requirements which leads to a moderate efficiency in combination with a negligible mis-identification of tracks. The hits associated to tracks are removed and the quality requirements are iteratively reduced. During the last two iterations the constraints on a primary vertex are loosened to allow for the reconstruction of tracks originating from the decay of heavy hadrons or conversion electrons, for example.

Studies showing the performance of the CMS track reconstruction in the first proton-proton collisions at $\sqrt{s} = 7$ TeV can be found in [142].

3.3.2 Primary-Vertex Reconstruction

The identification of the primary vertex (PV), i.e. the point where the hard scattering process has occurred, is an important step in the reconstruction of a collision event. As will be explained in more detail later, the compatibility of a track or particle with

originating from the primary vertex plays a key role in the rejection of pile-up as well as background events and helps in the identification of jets originating from b quarks. At the LHC, bunches of protons are collided and it is therefore not a priori clear, where exactly a collision will take place. Hence, the primary vertex has to be derived on an event-by-event basis.

Depending on the further usage of the obtained primary vertex candidates, different reconstruction algorithms are employed. For the primary vertices used for HLT decisions, a fast processing time is essential while the vertex candidates used in offline analyses are required to exhibit high precision.

For further usage in the HLT, primary vertices are exclusively reconstructed by using hits in the pixel detector. In a first step, three pixel hits are grouped, forming so-called pixel tracks. These pixel tracks are then ordered according to increasing longitudinal impact parameter values, which are calculated for pixel tracks as given in [143]. Adjacent tracks are grouped into the same PV cluster, if the difference in their longitudinal impact parameter is smaller than 0.6 mm. Otherwise, the track with the larger IP is the starting point of a new PV cluster. The center of each PV cluster is calculated by averaging over the IP values of all associated tracks. If the IP of an associated track is not consistent with the PV cluster center within the uncertainties on both points, the track is removed from the cluster and a new center is calculated until a stable set of associated tracks is found. The whole procedure is repeated using the tracks which have been removed in the step of finding the PV cluster center until at most two tracks remain. Finally, the cluster with the largest p_T^2 sum of the associated tracks is chosen as the primary vertex of an event.

If the reconstruction of primary vertex candidates is not limited by the computing time, more elaborate algorithms are used. In a first step, all tracks reconstructed as explained in subsection 3.3.1 are selected according to their number of hits in the pixel and strip detectors, their χ^2 from the track fit, and their transverse impact parameter significance, d_0/σ_{d_0} . Primary-vertex candidates are then formed by grouping tracks adjacent in z , which distance is smaller than 1 cm. In order to derive the best estimate of the vertex parameters, like the position and the corresponding covariance matrix, all vertex candidates with at least two tracks are fitted using an adaptive vertex fit procedure [144]. During the adaptive vertex fit, a specific weight is assigned to each track which represents the compatibility of the track with the vertex. This technique has the advantage of being less sensitive to mis-associated tracks or mis-estimated track uncertainties than standard fit techniques purely based on least-squares estimators. For each vertex candidate, the number of degrees of freedom is given by $n_{\text{dof}} = 2 \cdot (\sum_{i=1}^{N_{\text{tracks}}} w_i) - 3$ [142], where w_i is the weight of a particular track. The number of degrees of freedom is therefore correlated with the number of tracks compatible with the found vertex. In this thesis, the vertex candidate with the largest p_T^2 sum of the associated tracks is chosen as the primary vertex of an event.

Beam Spot Reconstruction

The center of the luminous region, where the two LHC beams collide within the CMS detector, is called the beam spot. It can be used as a coarse estimate for the primary vertex at HLT even before any primary vertex reconstruction has been performed. In contrast to the event-by-event PV reconstruction the beam spot is determined by averaging over many events. The beam spot can be reconstructed in two ways. It can be obtained from a likelihood fit to the three-dimensional primary vertex distribution as well as from a correlation between the transverse IP and the azimuthal angle of tracks, which is present if the LHC beam line is displaced from its expected position. Both methods are found to yield comparable results [142].

3.3.3 Electron Reconstruction

The reconstruction of electron candidates [145] is based on detector signals from the CMS tracking system and the electromagnetic calorimeter. Since the material budget of the tracking detector in front of the ECal is between 0.4 and 1.8 radiation lengths, most electrons radiate a large fraction of their energy in form of bremsstrahlung photons before reaching the calorimeter. Due to the bending magnetic field of 3.8 T, typically the energy deposits of an electron are thus spread in ϕ , which complicates the reconstruction process.

There are two complementary strategies used for the reconstruction of electron candidates. The ECal-driven seeding starts with the formation of clusters of ECal clusters, so-called superclusters, to account for the radiation of bremsstrahlung photons. This strategy is well-suited for isolated electrons above $p_T \approx 5 \text{ GeV}/c$. For non-isolated and low- p_T electrons, the so-called tracker-driven seeding is more convenient. In this subsection, only the ECal-driven seeding will be discussed. The explanation of the tracker-driven seeding is given in subsection 3.3.5 within the discussion of the Particle Flow algorithm.

Two different algorithms are used for the formation of superclusters in the ECal barrel and the ECal endcap region, respectively. The Hybrid algorithm which is used in the barrel region, uses the most energetic crystal in a predefined search region as the seed crystal if its transverse energy is larger than 1 GeV. Starting from this seed crystal, dominoes with an extension of one crystal in ϕ and three or five crystals in η are built, where the η extension depends on the transverse energy of the crystal in the domino center. Up to 17 dominoes on each side of the seed crystal are added, where the center of a domino is aligned with the seed crystal. If the energy contained in a domino is above 0.1 GeV, the domino is kept. The kept dominoes are clustered in ϕ into sub-clusters and different sub-clusters are finally combined into a supercluster.

The Island algorithm, used for the supercluster formation in the ECal endcap, also starts from a list of seed crystals. Further crystals are included into an island cluster if their energy is smaller than the energy of the previous crystal but still significantly larger than zero. Starting from the seed crystal, the algorithm first explores the adjacent crystal in the positive η region. If the crystal is included, all relevant crystals in

both ϕ directions from this new crystal are added and the algorithm again proceeds in the η direction. If all crystals in the positive η direction are added, also the region with negative η with respect to the seed crystal are explored. If a second seed is added to an island cluster, it is removed from the list of seeds to avoid double counting of energies. Starting from the most energetic island cluster, all island clusters found within a particular ϕ window around the first cluster are combined into a supercluster. A more detailed description of both algorithms can be found in [146].

In order to reject jets erroneously identified as electrons, some quality requirements are applied to all superclusters. The transverse energy of a supercluster has to be larger than 4 GeV. Furthermore, the ratio of the energy deposited in the HCal within a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.15$ around the supercluster position over the supercluster energy has to be smaller than 0.15. Since the supercluster resides on the helix of the initial electron, the innermost hits in the tracking detector can be identified by back-propagating the helix parameters through the magnetic field towards the pixel detector. Dedicated windows in ϕ and z direction in the barrel as well as ϕ and r direction in the endcap region are used to match the hits of track seeds to the remaining superclusters.

The track seeds matching a supercluster are used as input for a dedicated electron track reconstruction algorithm. In the standard Kalman filter technique the energy loss is modelled by a single Gaussian distribution. However, the electron energy loss due to bremsstrahlung emission, as it is given by the Bethe-Heitler model [147], shows a strong non-Gaussian behavior. Therefore, a non-linear generalization of the Kalman filter, the Gaussian-sum filter (GSF) [148], is employed for the reconstruction of electron tracks. The basic idea of the GSF is to model the bremsstrahlung energy loss by a mixture of Gaussian distributions rather than using a single Gaussian distribution.

Preselection of Electron Candidates

Besides the selection criteria already applied during the electron seeding, two further geometrical requirements are imposed to reduce the number of jets erroneously identified as electrons. The requirements are based on the compatibility of the supercluster and its associated GSF track. For this purpose, the azimuthal angle and the pseudorapidity of the GSF track at the innermost track position, ϕ_{in} and η_{in} , are compared with the supercluster's energy weighted position, represented by ϕ_{SC} and η_{SC} . The absolute value of the difference in the pseudorapidity, $|\Delta\eta_{\text{in}}|$, is required to be smaller than 0.02 and the difference in the absolute value of the azimuthal angle, $|\Delta\phi_{\text{in}}|$, is required to be smaller than 0.15. In a further selection step, ambiguous electron candidates that arise from the reconstruction of conversion legs from photons radiated by the primary electrons, are removed.

After the preselection, the energy and the momentum measurement, i.e. the information from the supercluster and the GSF track, can be combined to improve the estimation of the electron's momentum at the interaction point. As explained in more

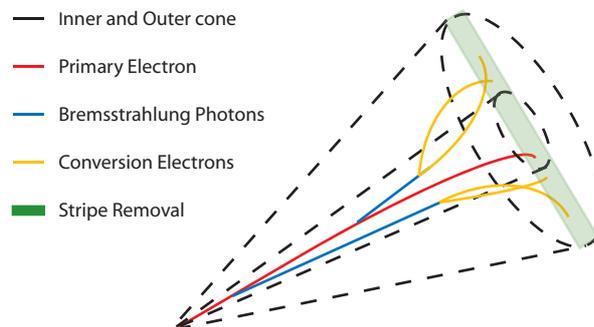


Figure 3.2: Schematical illustration of the derivation of the isolation energy in the tracking detector. The transverse momenta of all tracks within a cone in the η - ϕ -plane are summed. Afterwards the electron footprint is removed by an inner veto cone and a veto stripe which accounts for conversion electrons arising from bremsstrahlung photons.

detail in [145], the electrons are categorized according to the amount of bremsstrahlung they have radiated. In case, that the ratio of the measured supercluster energy and the GSF track momentum is close to unity, the electron's momentum is derived as the weighted mean between the supercluster energy and the track momentum. If the mentioned ratio shows significant deviations from unity, most of the times, solely the information from the subdetector offering a priori the most accurate measurement is used.

Electron Isolation

The amount of energy deposited in the vicinity of a given object is a measure for its degree of isolation. Electrons originating from W boson decays, for example, tend to be isolated in the detector, while electrons stemming from the decay of a hadron are frequently surrounded by various particles. Hence, the amount of energy deposited around an electron can be used to distinguish electrons arising from the hard scattering process from electrons produced in the decay of secondary particles. Isolation energies are derived in the tracking detector, the ECal, and the HCal. In the tracking detector, the isolation energy is defined as the sum of the transverse momenta of all tracks with p_T larger than $0.7 \text{ GeV}/c$ which are present in a cone with radius $\Delta R = 0.3$ around the electron's track. The energy of the electron track itself is removed by subtracting the energy in a concentric inner veto cone with radius $\Delta R = 0.015$. Additionally the energy contained in a stripe in ϕ with a width of $\Delta\eta = 0.015$ is removed, to account for secondary electrons arising from the conversion of bremsstrahlung photons. An illustration, schematically showing this isolation approach is depicted in figure 3.2. A similar approach is used to derive the isolation energy in the ECal, but energy deposits are summed rather than tracks and the radius of the inner veto cone and the η -width of the veto stripe are slightly different. To ensure that the HCal isolation energy is independent of the requirement on the hadronic-over-electromagnetic fraction applied during the supercluster preselection, the radius of the inner veto cone is set to

$\Delta\eta = 0.15$. In the HCal no further stripe removal is applied.

3.3.4 Reconstruction of Muons

An efficient reconstruction and identification of muons produced in proton-proton collisions has been one of the main design criteria of the CMS detector. Three different types of muon candidates are reconstructed: stand-alone muons, global muons and tracker muons. While stand-alone muons are identified by solely using information from the muon chambers, global muons, and tracker muons are reconstructed from both, muon chamber and tracking detector information. A detailed description of the muon reconstruction can be found in [149].

Stand-Alone Muons

The reconstruction algorithm starts with the identification of the hit positions in the DT, CSC, and RPC chambers. These hits are matched to form track stubs, so-called segments. Using rough geometrical criteria, compatible segments are grouped into track seeds which are the starting point of the actual track reconstruction. The track reconstruction algorithm is again based on the Kalman filter technique. Two filters are applied, a loose pre-filter evolving from the innermost muon detector layer towards the outside, and a final filter with tighter requirements, starting from the outermost hit, moving inwards. In the propagation step between two different detector layers the material effects like multiple scattering and energy losses due to ionization and bremsstrahlung emission are taken into account. At each layer, the most compatible hit is searched using a χ^2 , and the track parameters are updated by including the new hit if the χ^2 value is smaller than 100 for the pre-filter and 25 for the final filter, respectively. If no compatible hit is found, the track is extrapolated to the subsequent detector layer. Finally, a track is kept as stand-alone muon candidate if it consists of at least two hits, of which one is found in the DT and one in the CSC detectors.

Global Muons

To obtain a more reliable measurement of a muon candidate, the information from the muon chambers and the silicon tracking detector can be combined. For this purpose, it is necessary to match the stand-alone muon tracks with tracks reconstructed in the tracking system. In a first step, a so-called region of interest is defined, which depends on the primary vertex or the beam spot as well as an extrapolation of the stand-alone muon track to the tracking detector. All tracks in this region are treated as potential muon track candidates. In a further step, these candidates as well as the stand-alone muon track are propagated onto a common surface and the five parameters which describe the particular trajectory are compared. For muons with a small transverse momentum, a matching based on the comparability of the momentum parameters shows the best performance, while for high- p_T muon tracks, a matching based on the comparability of spatial parameters is more accurate. If the matching requirements

are not fulfilled by any candidate, the best candidate is chosen according to a loose requirement on the direction at the interaction point. In the end, the hits belonging to the stand-alone muon track and the associated track in the tracking detector are re-fitted to obtain a global muon track.

Tracker Muons

Especially for muons with low transverse momentum, it is possible that not enough hits are found within the muon chambers to reconstruct a stand-alone muon. Hence, an algorithm based on a complementary approach helps to increase the muon identification efficiencies in this phase space region. The algorithm uses all tracks reconstructed in the tracking system as input and searches for compatible signals in the muon chambers. For this purpose, the tracks are extrapolated to the muon chambers and loose matching requirements are applied to identify compatible hits. If at least two hits in the muon chambers are associated to a track, the track is used as a tracker muon candidate.

As for electrons, the degree of isolation of a muon candidate is a powerful quantity to distinguish prompt muons from muons originating from hadron decays. A similar approach as in the electron case is chosen, however, the muon footprint can be removed much easier than the electron footprint because of its minimal ionizing character. The energy contained in an isolation cone with radius $\Delta R = 0.3$ is summed and the energy of the muon is excluded by subtracting the energy in a veto cone with radius $\Delta R = 0.01$ and $\Delta R = 0.1$ in the tracking detector and the ECal/HCal, respectively.

3.3.5 Jet Reconstruction

Due to the confinement of the strong interaction, quarks and gluons cannot be observed as free particles. Through the hadronization process explained in section 3.1, quarks and gluons instead manifest as streams of collimated hadrons, so-called jets, in the detector. In order to reconstruct the energy of the parton that initiated the hadronization, the energies of all particles arising from the parton have to be summed. This summation is done using dedicated jet clustering algorithms which group close-by energy deposits in the detector into a jet object. Traditionally, only energy deposits in the calorimeters are accounted for. However, it has been shown that the usage of information from the tracking detector can help to significantly improve the energy measurements of jets, because especially for low- p_T particles the tracking detector provides a much more accurate p_T measurement compared to the calorimeters. Within the CMS reconstruction software, two different approaches are available, which include information from the tracking detector: The jet-plus-tracks approach [150], which exclusively aims for a more accurate measurement of the jet energy, and the more general Particle Flow (PF) approach [151], which attempts to reconstruct and calibrate each particle arising from proton-proton collisions. Since the jets employed in this thesis have been clustered using Particle Flow objects as input, the Particle Flow reconstruction is explained in more detail below. After the reconstruction, different corrections

have to be applied to achieve a jet energy measurement which is as close as possible to the energy of the original quark or gluon. In a process, which is called b -tagging, special algorithms are employed to identify jets originating from b quarks.

Jet Clustering Algorithms

The iterative cone algorithm (ICone) [149] provides a simple method for the reconstruction of jets. Starting from an energy ordered list of seed objects passing a minimum p_T threshold, a cone with a given radius in the η - ϕ -plane is defined around the most energetic object. The four-vectors of all objects belonging to this cone are added to define the new jet axis, and the procedure is repeated until a stable jet axis has been found. The objects finally associated to the jet are removed from the list of input objects and the algorithm proceeds with the most energetic object that remains. Algorithms of this kind suffer from two main drawbacks. First they are collinear unsafe, i.e. different jet configurations can be found if different p_T thresholds are used to generate the list of seed objects. Moreover, the addition of a very soft particle can result in a new jet configuration, a characteristic that is called infrared instability.

Although there are improved cone algorithms like the Seedless Infrared Safe Cone (SIS Cone) [152] available, often sequential cluster algorithms are used which are by construction collinear and infrared safe. A frequently used sequential cluster method is called k_T algorithm [153, 154]. However, it has the drawback that the area covered by the jet might have a very complicated shape. This is a crucial disadvantage when it comes to the correction of the jet energy from underlying event and pile-up contributions. The jets used in this analysis are clustered with the Anti- k_T algorithm [155] which combines the advantages of a sequential clustering method and a circle-like jet area. The Anti- k_T approach starts with the calculation of the distances d_{ij} between two input objects i and j represented by their four-vectors. Furthermore, the distance d_{iB} between each object i and a fixed point called beam B is derived. Within the Anti- k_T algorithm d_{ij} and d_{iB} are defined as

$$\begin{aligned} d_{ij} &= \min(k_{T,i}^{-2}, k_{T,j}^{-2}) \frac{\Delta_{ij}^2}{D^2} \quad \text{and} \\ d_{iB} &= k_{T,i}^{-2}. \end{aligned} \tag{3.1}$$

In this equations, $k_{T,i}$ and $k_{T,j}$ denote the transverse momentum of the objects i and j , and D is the so-called radius parameter which is to some extent similar to the cone radius in a cone algorithm. For the jets used in this thesis $D = 0.5$ has been chosen. The $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the quadratic sum of the differences in the rapidity y and the azimuthal angle ϕ of the two objects i and j . If the smallest observed distance is between to input objects, their four-vectors are merged and the calculation of the distances is repeated. In case, that a distance between an object and the beam is the smallest observed one, the object is called a jet and removed from the list of input objects. The algorithm proceeds until no further input object remains.

The Particle Flow Approach

In this thesis, the input objects which are clustered with the Anti- k_T algorithm are particles reconstructed using the Particle Flow approach. Particle Flow is a new type of reconstruction algorithm, which tries to identify all stable particles, i.e. electrons, muons, photons, charged hadrons, and neutral hadrons separately and to derive their energies as accurate as possible. For this purpose, it takes advantage of the information from all CMS sub-detectors.

The heart of the PF reconstruction is an efficient tracking with a low rate of erroneously identified tracks. This can be achieved by the iterative tracking procedure described in section 3.3.1. In contrast to the standard electron reconstruction described in subsection 3.3.3, in the Particle Flow approach, all tracks with a small number of associated hits as well as tracks with large χ^2 values are treated as possible electron candidates with large bremsstrahlung emission and are thus refitted with a Gaussian-sum filter.

In order to be able to separate the energy deposits of photons, electrons, as well as charged and neutral hadrons, a dedicated cluster algorithm has been developed for the PF reconstruction aiming for a high detection efficiency also for low energetic particles. This clustering is performed separately in the barrel and the endcap of the ECal and HCal, respectively, as well as in the first and second layer of the PS. First, every calorimeter cell above a given energy threshold gives rise to a cluster seed. Starting from the cluster seeds, all cells adjacent to a cluster are included in it, if their energy is larger than a given value. This leads to the formation of so-called topological clusters. Each seed contained in a topological cluster is the origin of a new PF cluster. The energy of the cells contained in a topological cluster is shared between the existing PF clusters taking into account the distance between the cell and the PF cluster. Afterwards, the energies and positions of the PF clusters are iteratively adjusted until a stable configuration is reached.

To associate the energy deposits in the different CMS sub-detectors to a particular particle, the tracks and PF clusters have to be linked. The linking algorithm extrapolates the tracks from their last hit to the two PS layers, the ECal — to a depth which is characteristic for electromagnetic showers originating from electrons — and the HCal, also to a depth which is characteristic for hadronic showers. A track is linked to a PF cluster, if its extrapolated position is within the cluster boundaries. Similarly, a PF cluster of the more granular calorimeter, i.e. the PS or the ECal, is linked to a PF cluster of the less granular calorimeter, i.e. the ECal or the HCal, if its position fits into the boundaries of the PF cluster in the less granular calorimeter. A link between a track reconstructed in the tracking detector and a track reconstructed in the muon system is found, if a combined fit fulfills the quality requirements applied for global muons as explained earlier. A special treatment is performed for the tracks, refitted with the Gaussian-sum filter. At each layer of the tracking detector a tangent to the track is extrapolated to the ECal. In this way, the clusters produced by bremsstrahlung photons can be identified and correctly linked to the corresponding GSF track.

From the linked tracks and clusters the individual particles have to be extracted.

In a first step, global muons are also PF muons, if the momentum resulting from the global fit agrees within the uncertainties with the momentum solely measured in the tracking detector. All energy deposits compatible with a PF muon are removed and the algorithm proceeds. Electrons with low bremsstrahlung radiation can directly be identified by a track linked to a PF cluster in the ECal which energy agrees with the track momentum within the uncertainties on both values. For electrons with a refitted GSF track a Boosted Decision Tree, based on different track and calorimeter-specific variables, is used to distinguish electrons from charged hadrons. Again, the tracks and calorimeter energies associated to electrons are removed. The remaining tracks and clusters give rise to photons as well as charged and neutral hadrons. If the momentum of all tracks pointing to a cluster is significantly smaller than the cluster energy, photons or neutral hadrons are reconstructed. If instead, the energy contained in the cluster is significantly smaller than the sum of the track momenta, a search for muons passing less stringent quality requirements is performed. The tracks are the source of charged PF hadrons which momentum and energy are directly taken from the track momentum using a charge pion mass hypothesis. PF clusters without any linked track are also treated as photons or neutral hadrons.

In a final step the energies of the different particle types are calibrated to account for the influences of the mentioned energy thresholds during the clustering as well as the different response of the ECal and HCal for different particle types.

Jet Energy Calibration and Resolution

Typically, jets reconstructed from the energy deposits in the detector (referred to as detector jets), show significant deviations in their associated energy when being compared with the corresponding particle jet. Particle jets are obtained from simulations by running the jet clustering algorithm on all stable MC particles. This energy difference is a direct consequence of the non-uniform and non-linear response of the CMS calorimeters. The purpose of jet energy calibrations, often also referred to as the jet energy scale (JES), is to account for these effects on average. Figure 3.3 shows the typical evolution of a jet from the parton to the energy deposits in the detector. In simulated events, the energies of the parton, the particle jet, and the reconstructed jet can be compared and calibrations can be derived. In general, the goal is to measure these calibrations from data where possible.

Depending on their influence on a specific physics analysis, different jet energy calibrations are necessary, while others are optional. Within the CMS collaboration, a factorized calibration approach [156] is used which consists of seven different levels. The first six levels are used to correct the energy of the reconstructed jet to the energy of the particle jet, the remaining level can be used to correct the jet energy to the energy of the parton.

- **Level 1: Offset Correction**

Offset corrections are used to subtract the energy arising from pile-up events and calorimeter noise. For this purpose the average energy contained in a cone,

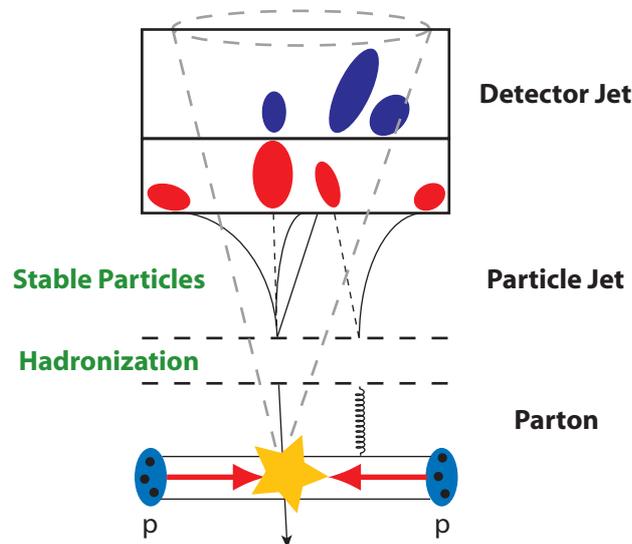


Figure 3.3: Typical evolution of a jet. A quark is produced in the hard scattering process and initiates the hadronization process. After the hadronization, stable or almost stable particles remain and interact with the detector material. A jet reconstructed from the energy deposits in the detector is called a detector jet, while a jet built from stable particles (only possible in simulations) is called a particle jet.

randomly spanned in the η - ϕ -plane, is removed from the jet energy.

- **Level 2: η -Dependence Correction**

For a particular particle jet, the p_T of the corresponding reconstructed jet shows an η -dependence. The level 2 calibration corrects for this effect. In first collision data, the level 2 calibration has been derived using a di-jet balance technique [157].

- **Level 3: p_T -Dependence Correction**

At a fixed η position, the p_T -difference between a particle jet and the corresponding reconstructed jet shows a dependency on the p_T of the particle jet. The purpose of the level 3 correction is to remove this dependency. In first collision data, the level 3 calibrations have been measured using γ +jets events [157].

- **Level 4: Electromagnetic Energy Fraction (EMF) Correction**

For a particular particle jet, the p_T of the corresponding reconstructed jet depends on the fraction of energy which has been deposited in the ECal. This correction is used to correct for this effect.

- **Level 5: Flavor Correction**

This calibration is intended to correct for the different behavior of jets originating from different parton flavors.

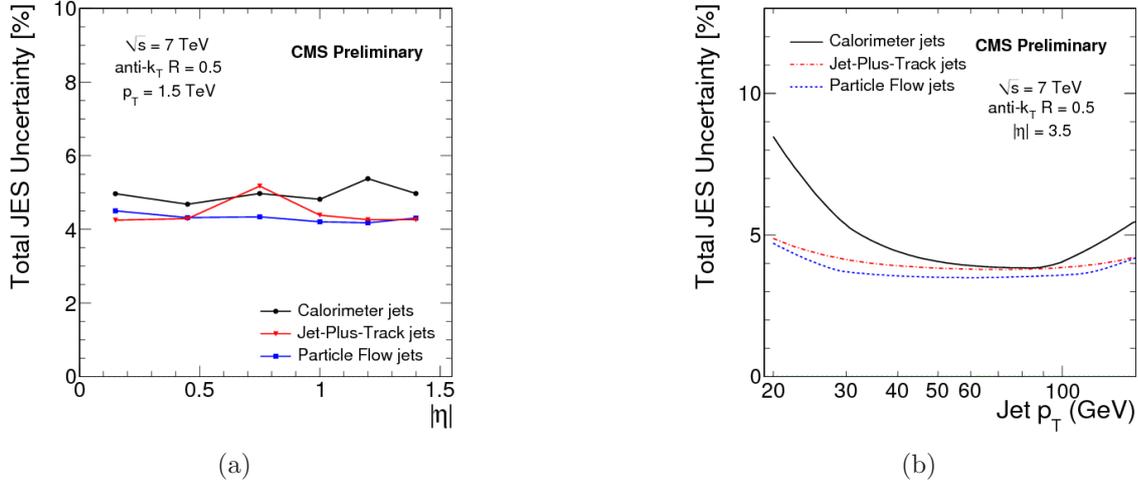


Figure 3.4: Dependency of the relative uncertainty of the jet energy calibration factors on η (a) and p_T (b) of the reconstructed jet taken from [158].

- **Level 6: Underlying Event Correction**

The additional energy originating from the interactions of the beam particle remnants is accounted for in this step.

- **Level 7: Parton Correction**

These corrections can be applied to correct the energy of a given particle jet back to the energy of the parton that initiated the hadronization process.

Level 2 and level 3 corrections are applied to the jets used in this thesis. All subsequent calibration levels are optional and are not necessary for significant impact on the presented analysis. Moreover, only the correction levels 2 and 3 have been derived from data, so far. Thus, the further calibrations would introduce an unnecessary dependency on the simulation into the analysis. The relative uncertainties on the jet energy calibration factors depend on η and on p_T of the reconstructed jet. The measured η and p_T dependencies for the combined level 2 plus level 3 corrections can be seen in figure 3.4 (a) and (b), respectively.

Another interesting characteristic of jets is their energy resolution. First, the so-called jet response is defined as the p_T of the reconstructed jet divided by the p_T of the corresponding particle jet. For particle jets with a particular p_T and η , the width of this response distribution is called the jet energy resolution (JER). The jet energy resolution of the CMS detector has been measured with collision data using a di-jet asymmetry method as well as a photon plus jet balance method [158]. In summary, the jet energy resolution has been found to be roughly 10% worse compared with the prediction obtained from simulations.

Identification of b -Quark Induced Jets

The identification of jets originating from bottom quarks, commonly referred to as b -tagging, is a powerful tool for the selection of physics processes containing b quarks in the final state. Most of the algorithms used to identify b -quark jets, take advantage of the long B -hadron lifetime of roughly 1.6 ps. The long lifetime, together with a large relativistic boost, allows B -hadrons to move several millimeters within the detector before they decay. The tracks originating from these decays are thus incompatible with stemming from the primary vertex and can be used to reconstruct secondary vertices. Moreover, B -hadrons can decay semileptonically, resulting in electrons or muons contained in the jet. Also this specific signature can be used to identify jets originating from b quarks.

Within the CMS collaboration, several b -tagging algorithms are available [159]. There are algorithms simply using the transverse impact parameters of tracks associated to a jet, to judge if the jet arises from a b quark or not. Furthermore, algorithms reconstructing secondary vertices as well as algorithms sensitive to leptons contained in jets are available. The commissioning of these algorithms using first collision data is documented in [160].

3.3.6 Missing Transverse Energy

Although, the CMS detector is built as hermetic as possible, there are uninstrumented regions in the very forward regions where the proton beams are inserted into the detector. Consequently, particles produced with momenta almost parallel to the beam axis, can escape the detector without being detected. Thus, the total energy balance is not an appropriate variable for describing a collision event. However, particles moving almost parallel to the beam line cannot carry a large amount of transverse momentum. Hence, any significant imbalance in the distribution of the momentum perpendicular to the beam line is an evidence for the production of a hardly interacting particle. Historically, the absolute value of the two-dimensional vector pointing in the direction of the missing transverse momentum is called missing transverse energy, \cancel{E}_T . In the Standard Model, a significant amount of \cancel{E}_T is assigned to a neutrino, but there are several extensions of the SM which also predict other weakly interacting stable particles. Unfortunately, the missing transverse energy distribution is sensitive to all kind of detector malfunctions and poorly instrumented regions that lead to an artificial momentum imbalance. A very good knowledge of the detector is thus needed for a deep understanding of this variable.

Traditionally, the missing transverse energy is calculated from the negative vectorial sum of all energy deposits within the calorimeter:

$$\begin{aligned}\vec{\cancel{E}}_T &= - \sum_n (E_n \sin \theta_n \cos \phi_n \hat{\mathbf{x}} + E_n \sin \theta_n \sin \phi_n \hat{\mathbf{y}}) \\ &= \cancel{E}_x \hat{\mathbf{x}} + \cancel{E}_y \hat{\mathbf{y}}.\end{aligned}\tag{3.2}$$

In this equation, n runs over all calorimeter energy deposits, $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit

vectors in the x and y direction, and θ and ϕ are the polar and the azimuthal angle, respectively. This missing transverse energy is finally corrected to account for jet energy calibrations and muons (if present in the event). With the development of the Particle Flow approach mentioned above, a new method for calculating \vec{E}_T has become available. Here the missing transverse energy is defined as the negative sum of the transverse momenta of all Particle Flow candidates. The missing transverse energy employed in this thesis has been calculated using Particle Flow Candidates and no further correction has been applied to it. The performance and a comparison of both algorithms using collision data at $\sqrt{s} = 7$ TeV is given in [161].

Chapter 4

Selection of $t\bar{t}$ Candidates

With the beginning of the operation of the Large Hadron Collider at a center-of-mass energy of 7 TeV in March 2010, the predictions deduced from the Standard Model of particle physics can be tested in an as yet unexplored energy regime. Besides the measurement of the production cross-sections for W and Z bosons, the measurement of the top-quark pair production cross-section is one of the first important steps. On one hand, a precise knowledge of the $t\bar{t}$ production cross-section is a crucial prerequisite for many interesting analyses which are carried out in the top-quark sector, like the derivation of the charge asymmetry [162], for example. On the other hand, $t\bar{t}$ production is predicted to be one of the most important background processes in the search for phenomena which are not described within the framework of the SM.

As depicted in figure 4.1, the total proton-proton cross-section at the LHC is roughly 120 mb. Comparing this value with the theoretically predicted $t\bar{t}$ production cross-section given in section 1.2.1, it can be seen, that on average only in every 750 millionth collision event a top-quark pair is produced. It is obvious, that selection criteria have to be applied to reject the large amount of events produced through different physics processes, called background processes in the following. Although the decay of top-quark pairs provides a very characteristic event topology, there are still some background processes left, which feature almost the same topology.

In order to estimate the number of expected events after a dedicated event selection has been applied, simulated datasets of all relevant processes are employed. These simulated events are also used to predict the shape of the distributions of different kinematic observables. However, there are also some background processes which are hard to simulate correctly. In such cases, a method has to be developed to model this background using collision data directly. Moreover, it has to be proven that the simulation is able to model real collision data and potential deviations have to be understood.

The aim of the described analysis is the measurement of the top-quark pair production cross-section using top-quark pairs decaying into an electron, a neutrino, and four quarks. For this purpose, an integrated luminosity of 35.9 pb^{-1} of data recorded with the CMS detector during the 2010 operation of the LHC, is analyzed. A study

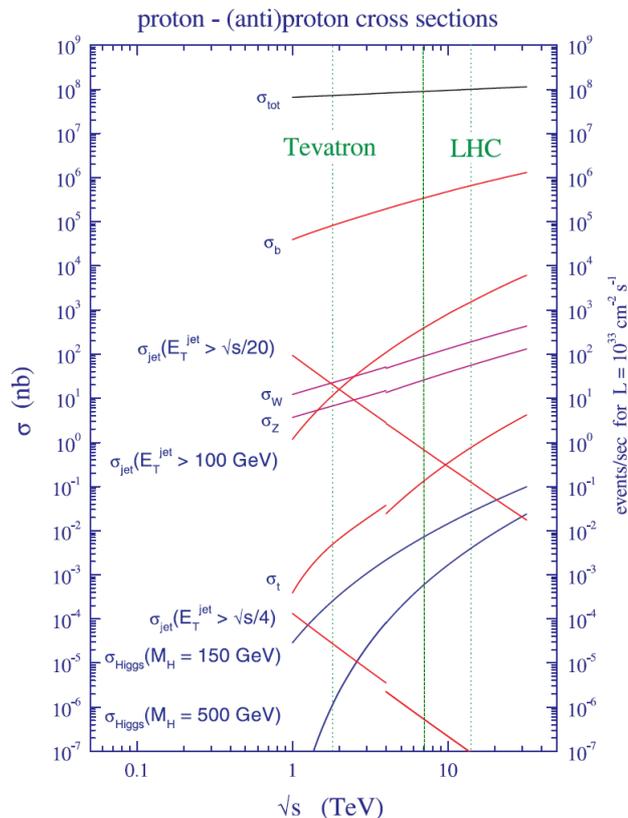


Figure 4.1: Cross-sections of different physics processes as a function of the center-of-mass energy for proton-antiproton as well as for proton-proton collisions [165]. The center-of-mass energies of the Tevatron (1.96 TeV), the center-of-mass energy of the current LHC operation (7 TeV), and the LHC design energy (14 TeV) are highlighted. It can be seen, that the top-quark production cross-section (σ_t) at $\sqrt{s} = 7$ TeV is roughly nine orders of magnitude smaller than the total proton-proton cross-section.

showing the prospects of such an analysis for $\sqrt{s} = 10$ TeV, is described in [163, 164]. An analysis, very similar to the one described in the following, has been performed using $t\bar{t}$ pairs decaying into the muon+jets final state [15]. As a consequence of this similarity, both analyses can be combined in a straight-forward way.

4.1 Signal and Background Events

In this analysis, the top-quark pair production cross-section at $\sqrt{s} = 7$ TeV is measured using events of the electron+jets channel. Within the framework of the SM, top quarks decay with a probability of almost 100% into a W boson and a bottom quark. For top-quark pairs decaying into the electron+jets final state, one of the W bosons decays leptonically into an electron and the corresponding neutrino, while the other W boson decays hadronically into two quarks. This decay signature is shown in figure 4.2. The

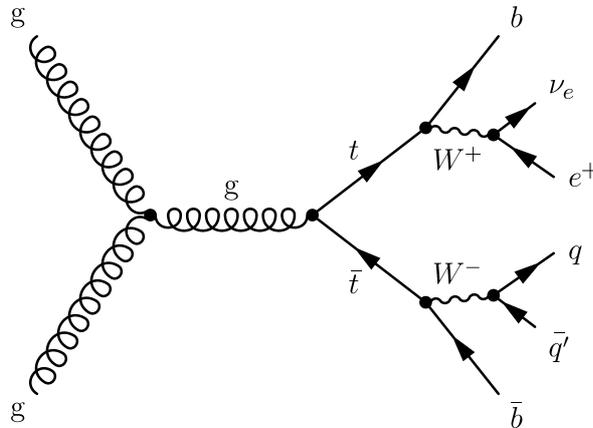


Figure 4.2: A top-quark pair produced in the LO gluon fusion process decays into the electron+jets final state. In this illustration, the top quark decays into a b quark and a W^+ boson, which further decays into a positron and an electron neutrino. The anti-top quark decays into a b quark and a W^- boson, that further decays hadronically into two light quarks. The decay channel, in which the top quark decays hadronically and the anti-top quark semileptonically into a \bar{b} quark, an electron, and a electron anti-neutrino, occurs with the exactly same probability.

resulting event topology in the detector is thus one highly energetic electron, four highly energetic jets, and a significant amount of missing transverse energy from the neutrino, which escapes undetected. Although this event topology is very characteristic for the decay of top-quark pairs, there are nevertheless several background processes which can mimic this signature. These background processes will be discussed in the following paragraphs.

4.1.1 Background Processes in the Electron+Jets Final State

The background processes in the electron+jets final state can be divided into two main categories. Firstly, there are processes with prompt electrons originating from the decay of W or Z bosons. Furthermore, there are background processes containing secondary electrons from the decay of heavy hadrons (dominantly B -hadrons). Secondly, electrons can also originate from the conversion of photons and their signature can be mimicked by charged pions or kaons, as well.

W +Jets Events

The background process with an event topology very close to that of the decay of a top-quark pair is the production of a W boson in association with jets. A valence or sea quark from one of the protons can annihilate with an anti-quark from the quark-sea of the other proton, leading to the production of a W boson. The W boson might decay into an electron and the corresponding neutrino. This LO process is illustrated in figure 4.3(a). In case the W boson production is accompanied by the radiation of gluons or the quarks annihilating to the W boson originate from a gluon splitting

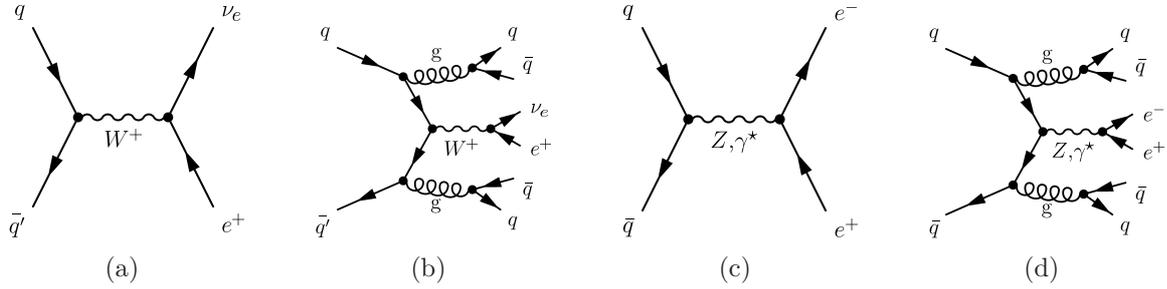


Figure 4.3: Feynman diagrams representing the LO production of a W boson (a) as well as the production of a W boson in association with jets (b). The latter final state is very similar to the decay of a $t\bar{t}$ pair. In (c) the LO Drell-Yan process is shown. If Drell-Yan production is accompanied by the radiation of additional gluons (d) and either the electron or the positron is not adequately reconstructed, the resulting event signature is similar to a $t\bar{t}$ decay, except for the missing transverse energy which is artificial in Drell-Yan events.

process, the final state of an electron, four jets, and substantial amount of missing transverse energy might be found. A Feynman diagram illustrating the mentioned higher-order process is shown in figure 4.3(b).

Drell-Yan+Jets Production

A quark and an anti-quark of the same flavor might annihilate and produce a Z boson or a virtual photon which decay into a pair of oppositely-charged electrons. This reaction is commonly referred to as Drell-Yan process [166] and is depicted in figure 4.3(c). If either the electron or the positron is not detected, because of inefficiencies in the reconstruction or limited detector acceptances, and the Drell-Yan process is accompanied by an additional radiation of gluons, the electron+jets final state can be counterfeited as shown in figure 4.3(d). However, these kinds of events are not expected to have a significant amount of missing transverse energy.

Photon+Jets Production

In contrast to the Drell-Yan process, a real photon is present in the final-state signature of photon+jets production. Photons can be produced in the scattering of a gluon and a quark as well as in the annihilation of a quark and an anti-quark. In both cases, the photon is produced in association with at least one jet. A Feynman diagram representing the LO quark-gluon scattering process is shown in figure 4.4(a). The photon might convert into an electron-positron pair in the detector material. If only one of these particles is reconstructed and the photon production is accompanied by additional jets, the same final state as for the decay of a $t\bar{t}$ pair is possible as shown figure 4.4(b).

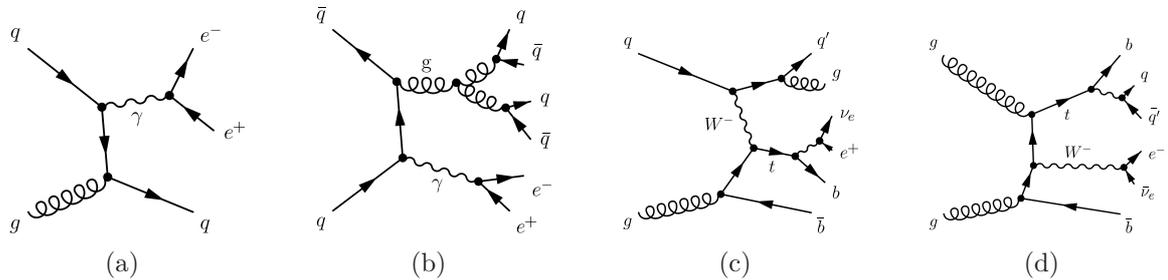


Figure 4.4: Feynman diagrams representing the LO production of real photons (a) and photon production with additional jets (b). If the photon converts into an electron-positron pair and one of these particles is not detected, the final state of a top-quark pair decay final state can be mimicked. In (c) the t-channel production of a single top quark with additional jets is shown. The associated production of a top quark and a W boson accompanied by additional jets is shown in (d).

Single Top-Quark Production

The production of single top quarks via the electroweak interaction has been discussed in section 1.2.1. Since the s-channel production exhibits only a small cross-section compared to the other two production processes, it can be neglected in the measurement described in this thesis. In order to counterfeit the final state of a top-quark pair decay, the top quark originating from the t-channel production has to decay semileptonically and has to be accompanied by additional jets. One of these jets might originate from a gluon splitting in the initial-state leading to the b quark shown in the diagrams in section 1.2.1. An example for this process is shown in figure 4.4(c). In events where the top quark is produced in association with a real W boson, either the top quark can decay semileptonically and the W boson hadronically, or the W boson decays into an electron and the corresponding neutrino, while the top quark decays hadronically. As for the t-channel production, an additional jet might arise from a gluon splitting in the initial-state. An example for such a process is given in figure 4.4(d).

QCD Multi-Jet Events

The background which comprises by far the largest cross-section is the production of multi-jet events via the strong interaction. In general, this final state only consists of jets originating from quarks and gluons. In such events, electrons can be produced through the semileptonic decay of B - or C -hadrons contained in jets. Moreover, π^0 mesons produced in jets can decay into two photons which might convert into electron-positron pairs. In some rare cases also charged pions or kaons originating from the hadronization of quarks and gluons could erroneously be identified as electrons. Since most of the electrons are produced in jets, QCD multi-jet events typically do not counterfeit the signature of an isolated electron originating from the decay of a prompt W boson. Despite the large production cross-section, the number of QCD multi-jet background events can thus be reduced to a manageable level by requiring a dedicated

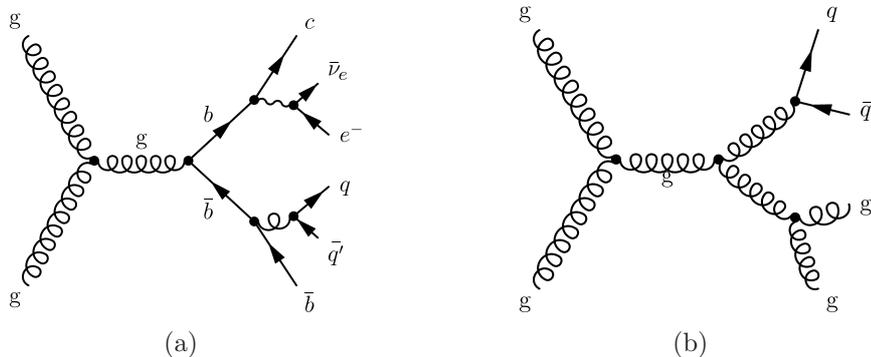


Figure 4.5: Two examples for the production of QCD multi-jet events via the strong interaction. A final state comprising electrons can be produced through the decay of heavy hadrons (a) or by the conversion of photons which have been produced in the decay of π^0 mesons. In rare cases, charged pions or kaons contained in a pure multi-jet final state (b) can be erroneously identified as electrons.

degree of electron isolation. In figure 4.5(a) the production of a $b\bar{b}$ pair is shown which gives rise to semileptonic B -hadron decays leading to electrons. In figure 4.5(b) a pure multi-jet final state is illustrated.

4.1.2 Signal and Background Modelling

In order to be able to develop an event selection which keeps as much of the $t\bar{t}$ signal as possible while rejecting background events as efficiently as possible, all the mentioned processes are simulated using MC generators. Each of these MC samples is passed through the full detector simulation to achieve a precise simulation of the interactions of the produced particles with the CMS detector material. High-level physics objects are reconstructed within the 3.8 version of the CMS software framework (CMSSW). The calibration constants for the CMS detector are taken from the so-called global tag, version START38_V14.

After dedicated event selection criteria have been developed, simulated datasets are used to get an estimation for the number of events passing the event selection. Moreover, the MC samples can be used to model the shape of different kinematic distributions. To get reliable estimations and to be independent of statistical fluctuations, the number of simulated events passing the selection criteria have to be sufficiently large. Due to the huge differences in the cross-sections of the processes mentioned above and the different probabilities to pass the event selection, the amounts of simulated events differ between the processes. Therefore, each MC event has to be given a weight ω , to enable the prediction for a certain amount of integrated luminosity. This weight can be calculated by

$$\omega = \frac{L_{\text{int}} \cdot \sigma \cdot \epsilon_{\text{F}}}{N_{\text{gen}}}. \quad (4.1)$$

Here, L_{int} represents the integrated luminosity, i.e. 35.9 pb^{-1} in this thesis, σ is the

Process	$\sigma \cdot \epsilon_F$ [pb]	N_{gen}	ω
$t\bar{t}$, inclusive	157.50	1,306,182	0.00433
t-channel, $t \rightarrow b\nu_l$	20.93	484,060	0.00155
tW -channel, inclusive	10.60	494,961	0.00077

Table 4.1: Details on the generation of the $t\bar{t}$ signal and the single top-quark background process. In case of $t\bar{t}$, all SM decay channels are simulated and the NLO production cross-section calculated with MCFM (see section 1.2.1) has been used in the calculation of the event weights. For the t-channel process, only semileptonic top quark decays are allowed. Hence, the effective cross-section is reduced by a factor of ~ 3 compared to the NLO value calculated with MCFM (see section 1.2.1). In the simulation of the tW -channel all SM decay channels are allowed and therefore the NLO prediction calculated with MCFM (see section 1.2.1) is used for calculating the event weights.

cross-section of the process, N_{gen} the number of generated events, and ϵ_F the filter efficiency. Some processes exhibit large cross-sections but only a very special final state is likely to pass the event selection. Therefore, a filter can be run during the event generation to select only events with this special final state. The efficiency of this filter is accounted for in ϵ_F . The product of cross-section and filter efficiency is often referred to as the effective cross-section. Although all the MC samples are produced with LO generators, the cross-sections used during the event weighting are typically the results of higher-order calculations. The MC samples used in this thesis belong to the so-called CMS Fall10 production and the official CMS name of all samples are listed in appendix A.1.

Modelling of Top-Quark Pair and Single Top-Quark Production

The $t\bar{t}$ signal process and the single top-quark background process are both generated using MADGRAPH/MADEVENT. For $t\bar{t}$ production, matrix elements with up to three additional jets are generated. The top-quark mass is set to $m_t = 172.5$ GeV, where again natural units with $\hbar = c = 1$ are used from now on in section 4.1. For each event, the renormalization and factorization scale is dynamically set to the quadratic sum of the top-quark mass and the p_T of the additional jets. For the simulation of the parton shower and the hadronization, the matrix elements are interfaced to PYTHIA employing the MLM matching prescription with a matching threshold of 20 GeV. The decay of τ -leptons is simulated using TAUOLA and for the modelling of the underlying event the PYTHIA D6T tune is used.

There are two possibilities for the LO generation of the single top-quark production in the t-channel and the tW -channel. First, the initial-state b quark can be modelled using the b quark PDF (see figure 1.4(a) and (c)). Moreover, the b quark might be treated as arising from an initial-state gluon splitting, i.e. the gluon-splitting is treated as a part of the hard process. An illustration of this type of modelling is shown in figure 4.4(c) and (d). Both approaches are believed to make reliable predictions in different phase space regions. Thus, a matching procedure for the two components has been developed to achieve an as appropriate as possible modelling for the whole

Process	$\sigma \cdot \epsilon_F$ [pb]	N_{gen}	ω
W +jets, $W \rightarrow l\nu$	31314	14,805,546	0.07591
Z/γ^* , $Z/\gamma^* \rightarrow l^+l^-$, $m(l^+l^-) > 50 \text{ GeV}/c^2$	3048	2,543,727	0.04301

Table 4.2: Details on the generation of W +jets and Drell-Yan+jets events. In the W +jets simulation, the W boson is forced to decay leptonically to ensure the presence of a lepton in the final state. For the Drell-Yan process, the Z bosons and virtual photons are only allowed to decay into an oppositely charged lepton pair. Moreover, the invariant mass of the lepton pair has to be larger than 50 GeV. The effective cross-sections for both processes are calculated with FEWZ at NNLO.

phase space. For the t-channel production, this matching procedure is described in detail in [167]. For both processes, the factorization scale μ_F and the renormalization scale μ_R are set to the top-quark mass and additional jets arise only from the parton shower simulated using PYTHIA. The underlying event is modelled using the PYTHIA Z2 tune.

In table 4.1 the number of generated events, the employed effective cross-sections, and the corresponding event weights are summarized.

Modelling of W +Jets and Drell-Yan+Jets Production

W +jets and Drell-Yan+jets events are generated with MADGRAPH/MADEVENT. The matrix elements of both processes are calculated for up to four additional jets. The parton shower and hadronization step of both samples is performed with PYTHIA, where the MLM matching procedure with a threshold of 10 GeV is applied. In both datasets decays of τ -leptons are simulated employing TAUOLA and the underlying event is modelled using the PYTHIA D6T tune.

In the production of W +jets events, μ_F and μ_R are dynamically set for each event to the quadratic sum of $m_W = 80.398 \text{ GeV}$ and the p_T of the additional jets. To ensure the presence of a lepton in the final state, the W boson is only allowed to decay into a lepton and the corresponding neutrino. The theoretical calculation of the effective cross-section for inclusive W boson production including a leptonic W boson decay, is carried out at NNLO with the FEWZ package [168]:

$$\sigma_W(W \rightarrow l\nu) = 31314 \pm 407 \text{ (scale)} \pm 1504 \text{ (PDF)} \text{ pb}. \quad (4.2)$$

The uncertainties due to the chosen renormalization and factorization scale is derived by re-calculating the cross-section using twice or half the original scale. The PDF parameterization used is CTEQ6M, and the uncertainties due to this choice are evaluated by using the corresponding 90% confidence interval values.

Drell-Yan+jets events are forced to decay into pairs of oppositely charged leptons. Additionally, the invariant mass of the lepton pair is required to be larger than 50 GeV. Similar to the W +jets production, μ_F and μ_R are dynamically set for each event to the quadratic sum of $m_Z = 91.1876 \text{ GeV}$ and the p_T of the additional jets. The effective

Process	$\sigma \cdot \epsilon_F$ [pb]	N_{gen}	ω
γ +jets, $40 \text{ GeV} \leq H_{T,\text{had}} < 100 \text{ GeV}$	23,620	2,217,101	0.38237
γ +jets, $100 \text{ GeV} \leq H_{T,\text{had}} < 200 \text{ GeV}$	3,476	1,065,691	0.11707
γ +jets, $H_{T,\text{had}} \geq 200 \text{ GeV}$	485	1,142,171	0.01524
BCtoE, $20 \text{ GeV} \leq \hat{p}_T < 30 \text{ GeV}$	132,160	2,243,439	2.1143
BCtoE, $30 \text{ GeV} \leq \hat{p}_T < 80 \text{ GeV}$	136,804	1,995,502	2.4606
BCtoE, $80 \text{ GeV} \leq \hat{p}_T < 170 \text{ GeV}$	9,360	1,043,390	0.32197
EMEnriched, $20 \text{ GeV} \leq \hat{p}_T < 30 \text{ GeV}$	2,454,400	37,169,939	2.3700
EMEnriched, $30 \text{ GeV} \leq \hat{p}_T < 80 \text{ GeV}$	3,866,200	71,845,473	1.9314
EMEnriched, $80 \text{ GeV} \leq \hat{p}_T < 170 \text{ GeV}$	139,500	8,073,559	0.62015
inclusive QCD, $15 \text{ GeV} \leq \hat{p}_T < 30 \text{ GeV}$	815,900,000	5,454,640	5368.6
inclusive QCD, $30 \text{ GeV} \leq \hat{p}_T < 50 \text{ GeV}$	53,120,000	3,264,660	583.99
inclusive QCD, $50 \text{ GeV} \leq \hat{p}_T < 80 \text{ GeV}$	6,359,000	3,191,546	71.511
inclusive QCD, $80 \text{ GeV} \leq \hat{p}_T < 120 \text{ GeV}$	784,300	3,208,299	8.7739
inclusive QCD, $120 \text{ GeV} \leq \hat{p}_T < 170 \text{ GeV}$	115,100	3,045,200	1.3566
inclusive QCD, $170 \text{ GeV} \leq \hat{p}_T < 300 \text{ GeV}$	24,260	3,220,080	0.27040

Table 4.3: Details on the generation of photon+jets and QCD multi-jets events. The γ +jets dataset is divided into three sub-samples each covering different regions in $H_{T,\text{had}}$. The corresponding effective cross-sections are derived using MADGRAPH/MADEVENT. Different types of QCD multi-jet events are simulated. Final states containing electrons are modelled with the BCtoE and EMEnriched datasets. For events with non-isolated electrons, the inclusive QCD datasets are used. For the three datasets, the effective cross-sections are derived using PYTHIA. Especially for the low \hat{p}_T sub-samples large weights have to be applied. However, since these kinds of events have only a negligible probability for populating the phase space with high jet multiplicities, the large weights can be accepted.

cross-section for this process is calculated with the FEWZ package at NNLO:

$$\sigma_{Z/\gamma^*}(Z/\gamma^* \rightarrow l^+l^-, m(l^+l^-) > 50 \text{ GeV}) = 3048 \pm 34 (\text{scale}) \pm 128 (\text{PDF}) \text{ pb}. \quad (4.3)$$

The uncertainties due to the choice of μ_F and μ_R as well as the uncertainties due to the PDF parameterization are derived similar to the uncertainties on σ_W . More details on the production of W +jets and Drell-Yan+jets events are summarized in table 4.2.

Modelling of Photon+Jets and QCD Multi-Jet Production

The photon+jets background is simulated with MADGRAPH/MADEVENT. Up to four additional jets can be present in the employed matrix elements. Furthermore, additional requirements on the produced events are applied. The p_T of the photon has to be larger than 20 GeV and it has to be produced in a pseudorapidity region $|\eta| < 2.5$. Moreover, the photon has to be separated in the η - ϕ -plane ($\Delta R > 0.3$) from any final-state quark or gluon. In order to provide a sufficient amount of events with a high jet multiplicity, the dataset is divided into three subsamples, each covering a particular $H_{T,\text{had}}$ region. Here $H_{T,\text{had}}$ is defined as the scalar sum of the p_T of the quarks and leptons in the final state. For each event, μ_F and μ_R are fixed to the Z -boson mass. The parton shower and hadronization step is simulated using PYTHIA, where the matching between the matrix elements and the parton shower is done with the

MLM approach using a matching threshold of 5 GeV, except for the sample with the largest $H_{T,\text{had}}$, where a 10 GeV threshold is employed. The modelling of the underlying event is performed based on the PYTHIA D6T tune.

Since the production cross-section of QCD multi-jet events is extremely large, only final states that have the possibility of mimicking an electron+jets final state are considered. For this purpose, two different datasets in two non-overlapping phase space regions are produced. In the first dataset, only electrons with $E_T > 10$ GeV and $|\eta| < 2.5$ originating from a b - or c -hadron are considered (BCtoE dataset). Such kind of events are explicitly vetoed in the second dataset which consists of two components. The first component clusters the energy from photons, electrons, charged pions, and charged kaons ($E_T > 1$ GeV, $|\eta| < 2.5$) in stripes of 0.4 in ϕ and 0.06 in η in the barrel region. In the endcap region, circles with radii of 15 cm in the x - y -plane are used. Each of these clusters has to have $E_T > 20$ GeV and is required to contain an electron or photon with $E_T > 5$ GeV. Furthermore, two isolation variables are defined using solely charged particles ($I_{\text{gen,tracker}}$) as well as charged and neutral particles, except photons ($I_{\text{gen,calo}}$). The isolation energies are computed by summing the energy contained in a cone with radius $\Delta R = 0.2$ around a cluster, where the cluster's energy itself is not included. $I_{\text{gen,tracker}}$ is required to be smaller than 5 GeV while $I_{\text{gen,calo}}$ is required to be smaller than 10 GeV. Moreover, the ratio of the transverse energy of hadronic particles other than charged kaons or charged pions within a cluster and the E_T of the cluster itself, has to be smaller than 0.5. Events with at least one cluster passing all these requirements are part of the first component. The second component searches for single electrons, charged pions, and charged kaons with $E_T > 20$ GeV and $|\eta| < 2.5$ which fulfill $I_{\text{gen,tracker}} < 4$ GeV and $I_{\text{gen,calo}} < 7$ GeV. However, in this second component the radius of the isolation cone is reduced to $\Delta R = 0.1$. Both of the mentioned components are present in the so-called EMEnriched dataset. The BCtoE and EMEnriched datasets are generated using PYTHIA and the Z2 tune is used for the underlying event modelling. To ensure a sufficient amount of events with a high jet multiplicity, each dataset is divided into three sub-samples, where every sample covers a particular $\hat{p}_T = \sqrt{\frac{\hat{t}\cdot\hat{u}}{\hat{s}}}$ region with the Mandelstam variables \hat{t} , \hat{u} , and \hat{s} .

As explained above, the EMEnriched dataset contains only events with isolated electron signatures. Hence, if the sideband region of non-isolated electrons is studied, these samples cannot be employed. Therefore, also a set of QCD multi-jet samples without the explicit requirement for an electron in the final state is used. Each of these datasets contains only events within a particular \hat{p}_T window. In order to obtain a modelling of the whole phase space, the different dataset have to be combined. A summary of the datasets used to model the photon+jets and the QCD multi-jet background is given in table 4.3.

4.2 Event Selection

In order to select a dataset enriched in top-quark pair events decaying into the electron+jets final state, dedicated event selection criteria have to be applied. These criteria are optimized for keeping as much events of the signal process as possible, while reducing the amount of background events to a manageable level. The different selection criteria are developed based on simulated datasets and are finally applied to real collision data. However, there are also requirements which are solely applied to real data, like the rejection of anomalous calorimeter events for example.

4.2.1 Analyzed Collision Data

The analysis described in this thesis uses all collision data recorded with the CMS detector during the LHC operation in 2010. The collected data is divided into different runs representing uninterrupted time periods in which the CMS detector has recorded data. Each run is labeled with a unique run-number and all of the events, passing the HLT during a particular run, are labeled with a unique event-number. Each run is further divided into equidistant time periods, where the instantaneous luminosity is assumed to be constant. These time periods, called luminosity sections (LS), correspond to roughly 2^{18} turns of the proton beams in the LHC, i.e. 23 s. The data analyzed in this thesis covers the run-numbers from 136035 to 149294. It is re-reconstructed using release 3_8_7 of the CMS software framework and the calibration constants for the CMS detector are taken from the GR_R_38X_V15 global tag. The official dataset names are given in appendix A.2.

Luminosity Measurement

There are two different approaches for the measurement of the luminosity of the collisions taking place in the center of the CMS detector [169]. The first method provides a luminosity measurement in real time exploiting the linear dependency between the average transverse energy per HF tower and the luminosity. A further approach is used to provide a luminosity measurement after a first reconstruction of the collected data is available. This method has of course a larger latency than the real time method, however, it allows for a better background rejection. The absolute normalization for both approaches is obtained from Van Der Meer scans [170], where the size and the shape of the interaction region of the two beams is derived by measuring the relative interaction rate as a function of the transverse separations of the beams. The systematic uncertainty on this measurement is evaluated to be $\pm 4\%$ [171].

Monitoring of Data Quality

The Data Quality Monitoring (DQM) system of the CMS detector [172] is of central importance for obtaining reliable certifications of the recorded data for any physics analyses. During the data taking process, its purpose is to spot any sub-detector

Trigger Name	E_T threshold [GeV]	Run-Range	L_{int} [pb $^{-1}$]
HLT_Ele10_LW_L1R	> 10	136035-140040	0.1
HLT_Ele15_SW_L1R	> 15	140041-143962	2.4
HLT_Ele15_SW_CaloEleId_L1R	> 15	143963-146427	0.7
HLT_Ele17_SW_CaloEleId_L1R	> 17	146428-147116	5.0
HLT_Ele17_SW_TightEleId_L1R	> 17	147117-148818	9.4
HLT_Ele22_SW_TighterEleId_L1R_v2	> 22	148819-149180	10.3
HLT_Ele22_SW_TighterEleId_L1R_v3	> 22	149181-149294	8.0

Table 4.4: Electron HLT triggers used to select the data analyzed in this thesis with their corresponding run-ranges. Due to the rapidly increasing instantaneous luminosity during the 2010 LHC operation, the electron requirements at HLT level had to be tightened accordingly to keep the trigger at a bearable rate. In the last column, the amount of integrated luminosity collected with a particular trigger is shown. Each quoted integrated luminosity has an uncertainty of $\pm 4\%$.

Startup Window (SW)	Large Window (LW)
$-0.035 < (\phi_{h1}^e - \phi_{SC}) < 0.025$	$-0.045 < (\phi_{h1}^e - \phi_{SC}) < 0.030$
$-0.025 < (\phi_{h1}^p - \phi_{SC}) < 0.035$	$-0.030 < (\phi_{h1}^p - \phi_{SC}) < 0.045$
$ \phi_{h2} - \phi_{h1} < 0.005$	$ \phi_{h2} - \phi_{h1} < 0.01$
$ z_{h2} - z_{h1} < 0.060$	$ z_{h2} - z_{h1} < 0.20$
$ r_{h2} - r_{h1} < 0.096$	$ r_{h2} - r_{h1} < 0.30$

Table 4.5: Size of the two windows used for the search for pixel hits which correspond to a super-cluster at the HLT. The ϕ position of the SC is given by ϕ_{SC} while ϕ_{h1}^e and ϕ_{h1}^p represent the position of the first pixel hit using the electron and positron hypothesis, respectively. After a first hit is found the search window for a second hit is given by ϕ_{h2} and z_{h2} (for ECal barrel) or r_{h2} (for ECal endcap), where z_{h2} and r_{h2} are the z - and r -coordinates of the second hit, respectively.

malfunction and to record which sub-detectors have been operated. In a second step after a first reconstruction of the collected data, the sub-detector operation and physics object quality are evaluated in more detail. Before the data is signed as good for physics analysis, the results of this evaluation are handed over to the responsible detector and physics objects groups for a final validation. For the analysis of top-quark events, all CMS subdetector systems has to be fully operational. Hence, only runs and LS where all sub-detectors are signed as good for physics analyses are taken into account. A list of these special runs and LS is given in a so-called JSON file. The JSON file used in this analysis is named `Cert_136033-149442_7TeV_Nov4ReReco_Collisions10_JSON`. The certified amount of data corresponds to an integrated luminosity of 35.9 ± 1.4 pb $^{-1}$.

Trigger

At the beginning of each physics analysis, it has to be ensured that the final state the analysis is interested in, passes the trigger requirements. All data analyzed in this thesis have passed an HLT trigger sensitive to the presence of at least one electron. As can be seen in figure 2.4(a), the LHC's instantaneous luminosity increased rapidly during the 2010 operation. As a consequence, also the rate of processes containing electrons increased accordingly. Thus, the trigger requirements for electrons had to be adjusted several times to avoid too large trigger rates. In summary, this led to the usage of seven different HLT triggers for different run-ranges. An overview of the seven HLT triggers, the corresponding run-ranges and the amount of integrated luminosity collected with a particular trigger is shown in table 4.4. Except for the triggers of the first two run-ranges which are based on a L1 electromagnetic object with $E_T > 5$ GeV, all of these HLT triggers are seeded by a L1 electromagnetic object with $E_T > 8$ GeV. Each HLT trigger first requires the presence of an electron supercluster with $E_T > 5$ GeV. Moreover, the energy deposited in the HCal in the region rear side of the ECal area attributed to the supercluster, divided by the energy of the supercluster itself, is smaller than 20%. This energy ratio is often abbreviated as H/E . If at least one supercluster is present in the event, matching hits in the pixel detector are searched for using the hypothesis that the supercluster originates from an electron or positron. The SW (Startup Window) and LW (Large Window) in the trigger names represent the size of the matching window used for the search of associated pixel hits. In table 4.5 the size of these windows is given in more detail. Four different observables are used to identify electrons based on the shape of the electromagnetic shower: the $|\Delta\eta_{\text{in}}|$ and $|\Delta\phi_{\text{in}}|$ variables introduced in section 3.3.3, the H/E ratio mentioned above, and $\sigma_{\eta\eta}$. The last variable is a measure for the η extension of an electron supercluster and is defined as

$$\sigma_{\eta\eta} = \sqrt{\frac{\sum_i^{5 \times 5} (S_i^\eta \cdot \Delta\eta_{\text{xtal}} + \eta_{\text{seed}} - \bar{\eta}_{5 \times 5}) w_i}{\sum_i^{5 \times 5} w_i}} \quad \text{with } w_i = 4.2 + \ln(E_i/E_{5 \times 5}). \quad (4.4)$$

In this equation the index i runs over all ECal crystals contained in a 5×5 block of crystals centered at the supercluster seed crystal. Moreover, S_i^η denotes an integer running from -2 to 2 accounting for the η difference between the crystal i and the seed crystal. The average η -size of a crystal is represented by $\Delta\eta_{\text{xtal}}$ which is set to 0.01745 and 0.0447 for the ECal barrel and endcap, respectively. The supercluster's energy weighted mean η -position is given by $\bar{\eta}_{5 \times 5}$ and E_i as well as $E_{5 \times 5}$ represent the energy deposited in the crystal i and the 5×5 block, respectively. In the trigger names, CaloEleId, TightEleId, and TighterEleId represent different requirements on these electron identification observables which have to be fulfilled in addition. The difference between V2 and V3 of the TighterEleId trigger is an algorithm which removes the contributions from anomalous signals in the ECal barrel already at the HLT stage included in V3. The exact thresholds are listed in table 4.6. If an electron candidate has passed all the mentioned requirements the final decision on whether the event passes

Label	H/E	$ \Delta\eta_{\text{in}} $ (EB)	$ \Delta\eta_{\text{in}} $ (EE)	$ \Delta\phi_{\text{in}} $	$\sigma_{i\eta i\eta}$ (EB)	$\sigma_{i\eta i\eta}$ (EE)
CaloEleId	< 0.15	-	-	-	< 0.014	< 0.035
TightEleId	< 0.15	< 0.010	< 0.010	< 0.08	< 0.012	< 0.032
TighterEleId	< 0.15	< 0.008	< 0.007	< 0.10	< 0.012	< 0.032

Table 4.6: Requirements applied on the electron identification observables at the HLT. The first column represents the labelling present in the different trigger names. For $|\Delta\eta_{\text{in}}|$ and $\sigma_{i\eta i\eta}$ slightly different thresholds are used in the ECal barrel (EB) and endcap (EE).

the HLT or not is taken based on the transverse energy of the electron candidate. The numbers behind Ele in the trigger names represent the lower E_{T} -boundary measured in GeV for an HLT object.

4.2.2 Selection Criteria

The selection criteria used to select a dataset enriched in $t\bar{t}$ events, exploit the characteristic event signature of top-quark pairs decaying into the electron+jets final state. The presence of highly energetic and well-isolated electrons are required to reject background contributions from QCD multi-jet production. The limitation to exactly one electron in the final state rejects Drell-Yan events as well as the background contribution from the production of real photons. Requiring additionally a high jet multiplicity helps to reduce the large amount of background events, stemming from the production of W bosons. Except for the rejection of anomalous calorimeter signals which is only applied to collision data, the described selection criteria are applied to simulated as well as real collision events.

Corrections to MC Events and Collision Event Quality

In the generation of $t\bar{t}$ MC events, the LO branching fractions for W boson decays into leptons and hadrons have been used, i.e $\text{BR}(W \rightarrow l\nu) = 0.111$ and $\text{BR}(W \rightarrow q\bar{q}') = 0.667$. By assigning an additional weight to each $t\bar{t}$ MC event, these branching ratios are corrected to their measured values of $\text{BR}(W \rightarrow l\nu) = 0.1080 \pm 0.0009$ and $\text{BR}(W \rightarrow q\bar{q}') = 0.676 \pm 0.027$ [29].

As explained in section 3.3.5, the jet energy resolution has been found to be roughly 10% worse in collision data compared to the JER in simulated events. To take into account the effects of a worse JER, the JER is increased by 10% in all simulated datasets. For this purpose, a matching between detector jets and particle jets is performed by requiring that the detector and the particle jet are close by in the η - ϕ -plane ($\Delta R < 0.4$) and that the difference in their p_{T} is smaller than three times the p_{T} of the detector jet. If the transverse momentum of the matched particle jet is smaller than 15 GeV or if no matching particle jet could be found, the JER of the detector jet is left unchanged.

All other detector jets are corrected by multiplying their four-vector with

$$C_{\text{JER}} = \frac{(p_{\text{T}}^{\text{det}} - p_{\text{T}}^{\text{part}}) \cdot 1.1 + p_{\text{T}}^{\text{part}}}{p_{\text{T}}^{\text{det}}}, \quad (4.5)$$

where $p_{\text{T}}^{\text{det}}$ and $p_{\text{T}}^{\text{part}}$ are the transverse momenta of the detector jet and the matched particle jet, respectively. This correction of the detector jet four-vector is consistently propagated to the missing transverse energy of an event by the following approach: First the raw (no jet energy corrections applied) p_{T} of each jet, before the JER correction is applied, is added to $\vec{\cancel{E}}_{\text{T}}$. Afterwards, the JER correction is applied and the JER corrected raw p_{T} of each jet is subtracted from $\vec{\cancel{E}}_{\text{T}}$. In this case raw jet energies are used, since the missing transverse energy has not been corrected for jet energy correction effects.

To ensure, that all selected data events originate from a proton-proton collision, a filter is applied to reject beam induced background events. This filter removes events with more than ten tracks, if the fraction of tracks passing high quality requirements is smaller than 25%. Moreover, collision events with anomalous signals in the HB or HE are removed. These signals have first been observed in cosmic muon data and originate from the hybrid photodiodes and the readout boxes used to convert the HPD scintillation light into electric pulses. The anomalous signals are identified based on the shape of the corresponding electric pulses, hit multiplicities, and timing measurements.

In this analysis, the vertex candidate with the largest p_{T}^2 -sum of the tracks associated to the vertex is chosen as the primary vertex of the event. For both, simulated and real collision events, this primary vertex is required to pass different quality criteria. First, the maximal allowed distance between the PV and the nominal interaction point is 24 cm in the z -direction and 2 cm in the plane transverse to the beam. Furthermore, the number of degrees of freedom in the vertex fit has to be larger than four, and the vertex must not originate from a spurious vertex fit, i.e. $\chi^2 = 0$, $N_{\text{dof}} = 0$, and $N_{\text{tracks}} = 0$.

Electron Selection Requirements

In a first step, the electron track is required to originate from the utilized primary vertex which ensures that it does not stem from additional proton-proton interactions within the same bunch crossing. Therefore, the z distance between the PV and the origin of the electron track has to be smaller than 1 cm.

The online electron reconstruction at the HLT and the offline reconstruction described in section 3.3.3 are different in some aspects. As a consequence, the E_{T} of an electron at the HLT and at the analysis level can be significantly different. To avoid an inefficiency due to the applied trigger E_{T} -threshold and to additionally reject electrons originating from the decays of heavy hadrons, the transverse energy of an electron is required to be larger than 30 GeV. Moreover, electrons are required to lie within the acceptance of the tracking detector, i.e. $|\eta| < 2.5$, and the electron supercluster must

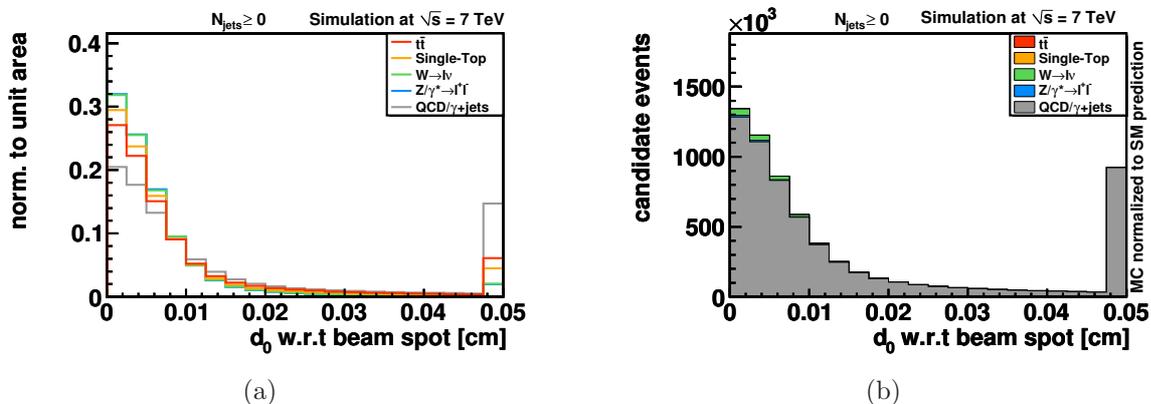


Figure 4.6: Normalized distributions of the transverse impact parameter (d_0), calculated with respect to the beam spot for different processes (a). In (b) the expected d_0 distribution for 35.9 pb^{-1} is shown.

not reside in the transition region between the ECal barrel and endcap, i.e. electrons with $1.4442 < |\eta_{\text{SC}}| < 1.566$ are rejected.

As explained in section 3.3.5, tracks with large transverse impact parameters are used to identify jets originating from b quarks. The tracks stem from decays of heavy hadrons and are thus incompatible with arising from the primary interaction. Therefore, the transverse impact parameter d_0 of the electron track with respect to the beam spot is required to be smaller than 200 μm . The normalized distribution of the transverse impact parameter for the different processes and the expected d_0 distribution for the selected dataset, are shown in figure 4.6(a) and (b). In all figures shown in the following, the contributions from the single top t -channel and tW -channel processes as well as the contribution of QCD multi-jet and γ +jet events are combined into a common background process.

In order to distinguish electrons from charged pions and kaons, the electron identification observables explained above are employed. These variables exploit the differences in the shape of the electromagnetic shower for the different particle types. A set of working points (WP) are defined, which implement thresholds on the values of the different identification observables. Each WP represents a particular selection efficiency for $W \rightarrow e\nu$ events, e.g. WP95 is fulfilled by 95% of all electrons originating from a W boson decay. All electrons are required to pass WP70 of the electron identification observables to ensure an appropriate rejection of erroneously identified electrons. The thresholds on the electron identification observables for WP95 (used in the Z boson veto described below) and WP70 are given in table 4.7. The distributions of the electron identification observables for different processes are shown in figure 4.7 for electrons reconstructed in the ECal barrel and in figure 4.8 for endcap electrons. For each variable, also the expected distribution for the selected dataset is shown.

A very powerful criterion for the reduction of the large QCD multi-jet background is the requirement for an isolated electron. Electrons produced in jets are typically

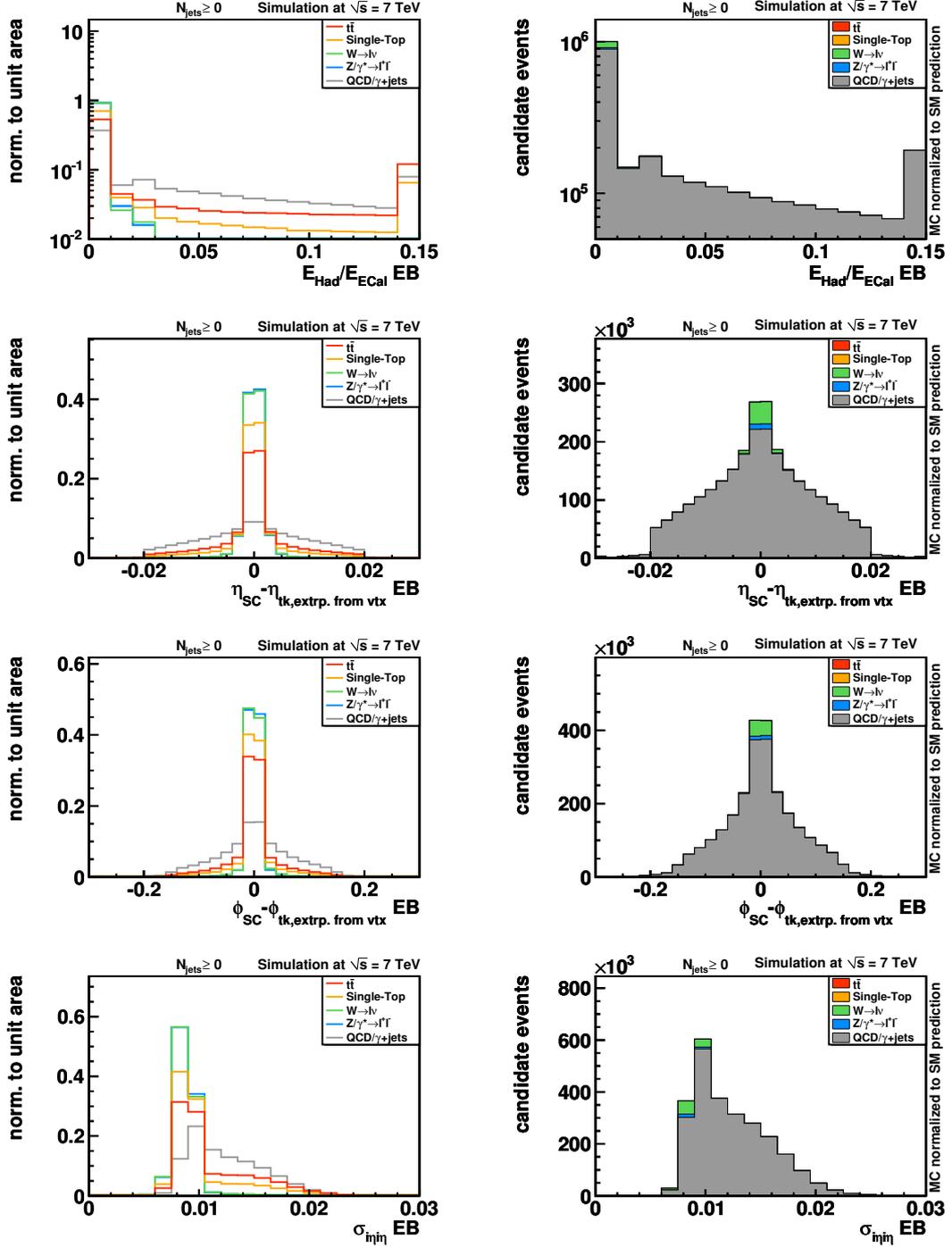


Figure 4.7: Normalized distributions of the electron identification observables in the ECAL barrel. For each variable, also the expected distribution for a dataset with 35.9 pb^{-1} is shown. From top to bottom, the H/E ratio, the $|\Delta\eta_{in}|$ distribution, the $|\Delta\phi_{in}|$ variable, and $\sigma_{i\eta i\eta}$ are depicted.

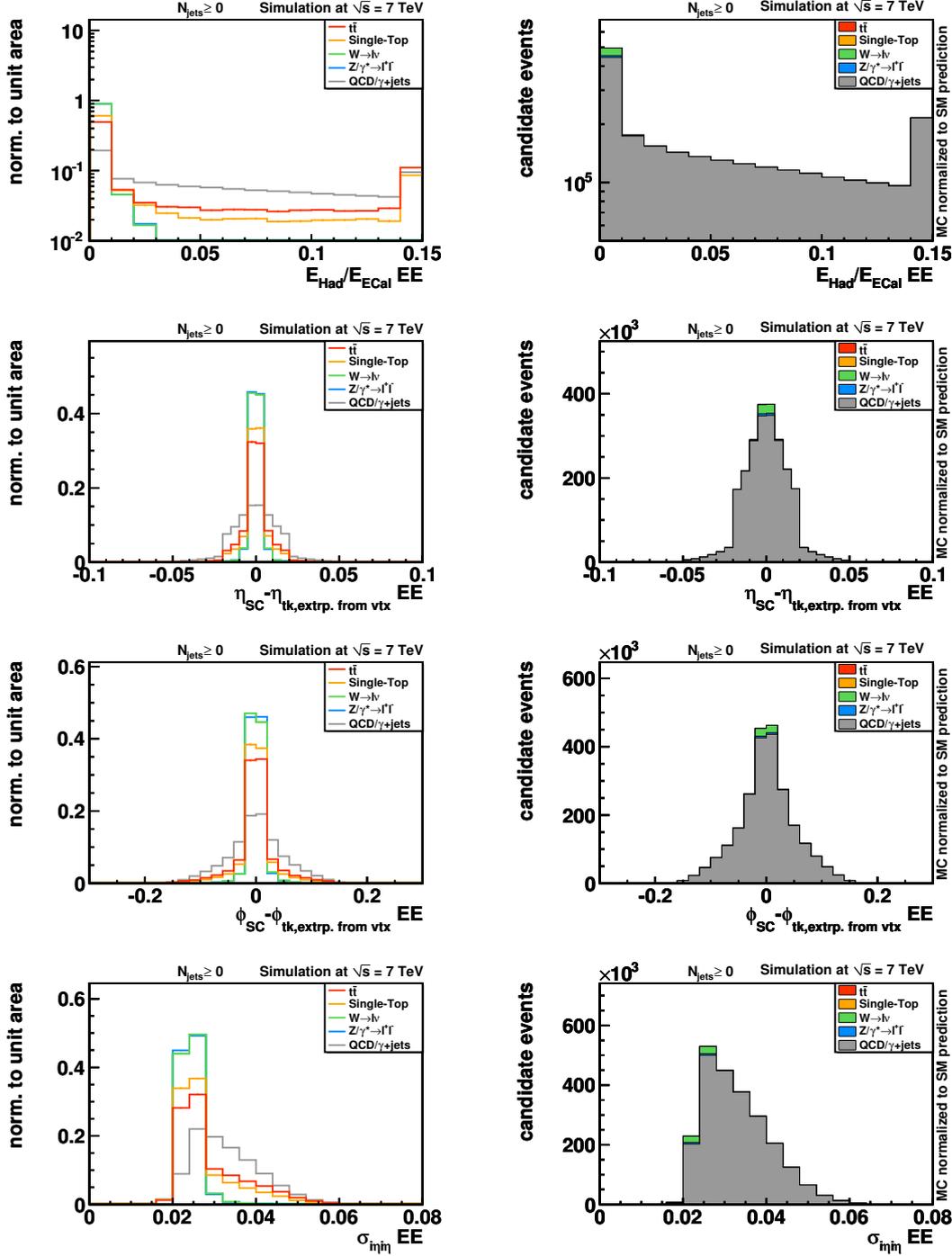


Figure 4.8: Normalized distributions of the electron identification observables in the ECal endcap. For each variable, also the expected distribution for a dataset with 35.9 pb^{-1} is shown. From top to bottom, the H/E ratio, the $|\Delta\eta_{in}|$ distribution, the $|\Delta\phi_{in}|$ variable, and $\sigma_{in\eta}$ are depicted.

Working Point	H/E		$ \Delta\eta_{\text{in}} $		$ \Delta\phi_{\text{in}} $		$\sigma_{i\eta i\eta}$	
	EB	EE	EB	EE	EB	EE	EB	EE
WP95	< 0.150	< 0.070	< 0.007	< 0.010	< 0.80	< 0.70	< 0.01	< 0.03
WP70	< 0.025	< 0.025	< 0.004	< 0.005	< 0.03	< 0.02	< 0.01	< 0.03

Table 4.7: Thresholds on the electron identification observables for the two working points employed in this analysis.

surrounded by further particles, while electrons originating from the decay of prompt W bosons are frequently isolated. A measure for the degree of isolation of a specific particle is the amount of energy which is deposited in the detector in its vicinity. As explained in section 3.3.3, isolation energies are derived in the tracking detector (I_{Tk}), the ECal (I_{ECal}), and the HCal (I_{HCal}), separately. Instead of defining absolute upper boundaries for each of these isolation energies, it turned out to be more efficient to set an upper boundary on the combined isolation energy relative to the electron's transverse energy. The relative isolation variable is defined as

$$I_{\text{rel}} = \frac{I_{\text{Tk}} + I_{\text{ECal}} + I_{\text{HCal}}}{E_{\text{T}}}. \quad (4.6)$$

The distributions of the single isolation energies and the distribution of the relative isolation for the $t\bar{t}$ signal and different background processes are shown in figure 4.9. For each variable, also the expected composition of the selected dataset using 35.9 pb^{-1} is shown. Based on the distribution of the relative isolation, each electron is required to fulfill $I_{\text{rel}} < 0.1$.

The presence of more than one electron passing all the mentioned selection criteria is a clear hint for a top-quark pair decaying into the di-electron final state or the decay of a Z boson or a virtual photon into an electron-positron pair. Therefore, only events with exactly one electron passing the mentioned selection criteria are subject to further considerations.

Muon and Z boson vetos

In order to minimize the selection of top-quark pairs decaying into the electron-muon final state, events containing additional isolated muons are rejected. Moreover, the rejection of events containing muons ensures disjoint datasets for the measurements of the $t\bar{t}$ production cross-section in the electron+jets and muon+jets final states. This independence of the two datasets is of central importance for a combined measurement in both channels, since no statistical correlations between the datasets have to be taken into account which would complicate the derivation of the uncertainties on the measurement. Compared to the muon selection criteria described in [15], the muon definition employed in this analysis is less stringent. Events containing global muons with $p_{\text{T}} > 10 \text{ GeV}/c$ within the acceptance of the tracking detector ($|\eta| < 2.5$) and a relative isolation smaller than 0.2 are rejected.

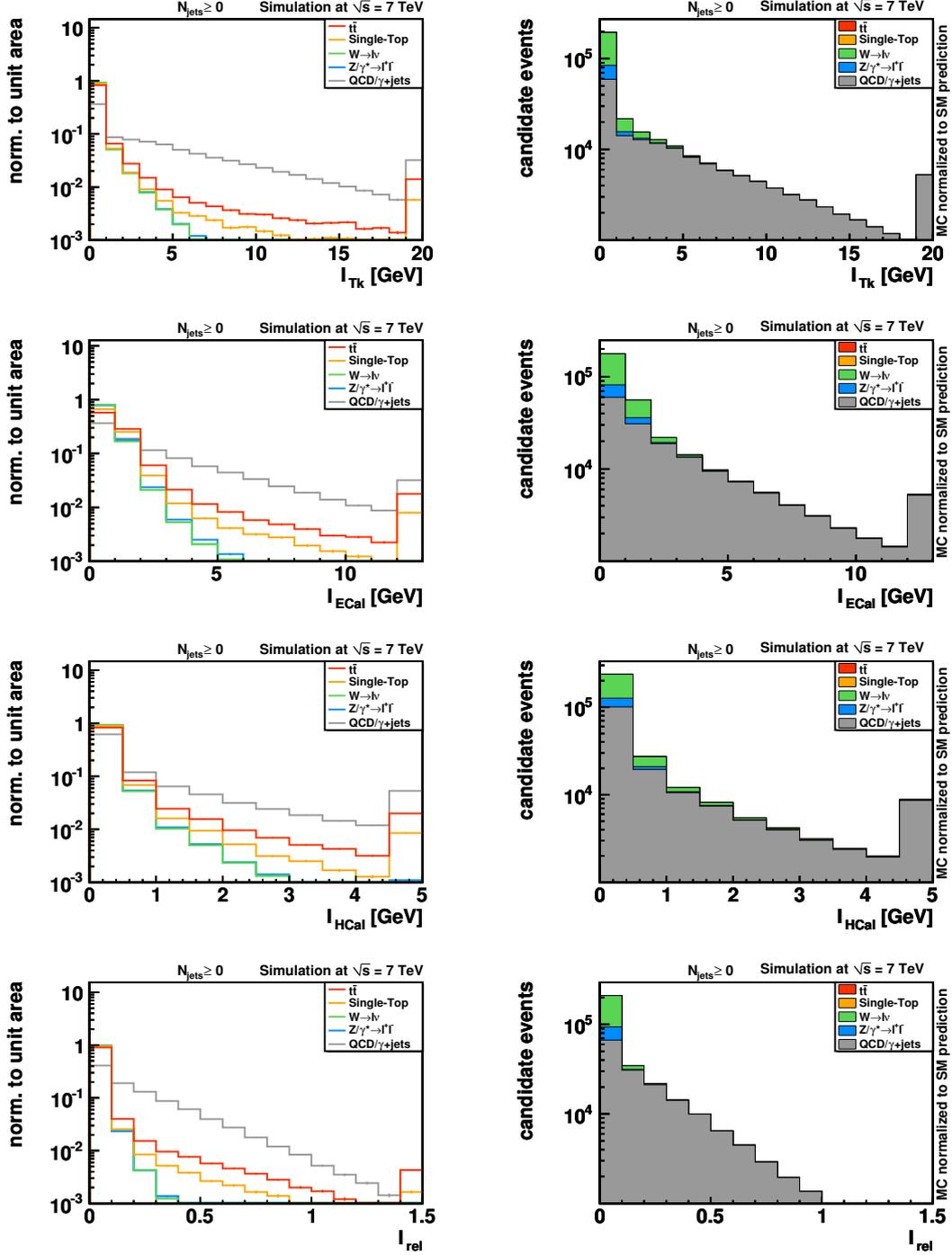


Figure 4.9: Electron isolation energies computed in the tracking detector (first row), the electromagnetic calorimeter (second row), and the hadronic calorimeter (third row). The combined relative isolation variable used to select electrons is shown in the last row. Only events with an electron fulfilling $I_{rel} < 0.1$ enter the selected dataset.

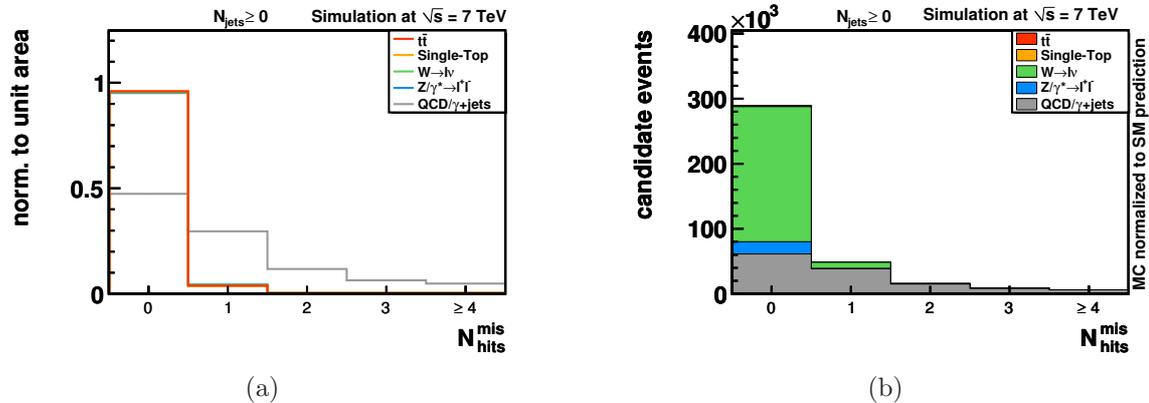


Figure 4.10: Normalized distribution of the number of missing hits variable (a) and the expected $N_{\text{hits}}^{\text{mis}}$ (b) distribution for the selected dataset. As can be seen, only tracks of electrons originating from photon conversions are likely to have no associated hit on the first pixel layer.

Although events with two or more electrons are rejected, the number of background events originating from Drell-Yan production processes is still significant. One reason for this is the high electron E_T -threshold, since electrons stemming from Z boson decays have on average a smaller transverse energy than electrons originating from the decay of a top quark. A second reason is the tight working point used for the electron identification, because only 70% of all electrons are expected to pass WP70. Thus, electrons passing less stringent selection criteria are defined, i.e. $E_T > 20$ GeV and WP95 of the electron identification (requirements on η and η_{SC} are still the same as mentioned above). If the invariant mass M_{ee} of the primary selected electron and a less stringently defined electron is close to the Z -boson mass, i.e. $76 \text{ GeV}/c^2 \leq M_{ee} \leq 106 \text{ GeV}/c^2$, the event is rejected.

Conversion Rejection

Electrons originating from photon conversions are likely to pass the event selection criteria mentioned so far, if their transverse energy is sufficiently large. Two algorithms are used to further suppress these background events. The first algorithm takes advantage of photon conversions occurring after the photon has traversed some detector material. Electrons originating from those photon conversion processes are likely to have the first hit associated to their track not on the first pixel layer. Therefore, a large amount of events, where the electron stems from a photon conversion, can be rejected by requiring that the number of missing hits $N_{\text{hits}}^{\text{mis}}$ is exactly zero. This quantity is defined as the number of tracking detector layers that are closer to the interaction point than the layer on which the innermost hit associated to the electron track is found. The $N_{\text{hits}}^{\text{mis}}$ distribution for the different processes is shown in figure 4.10(a) and the expected composition of the selected dataset as a function of this variable is shown in figure 4.10(b).

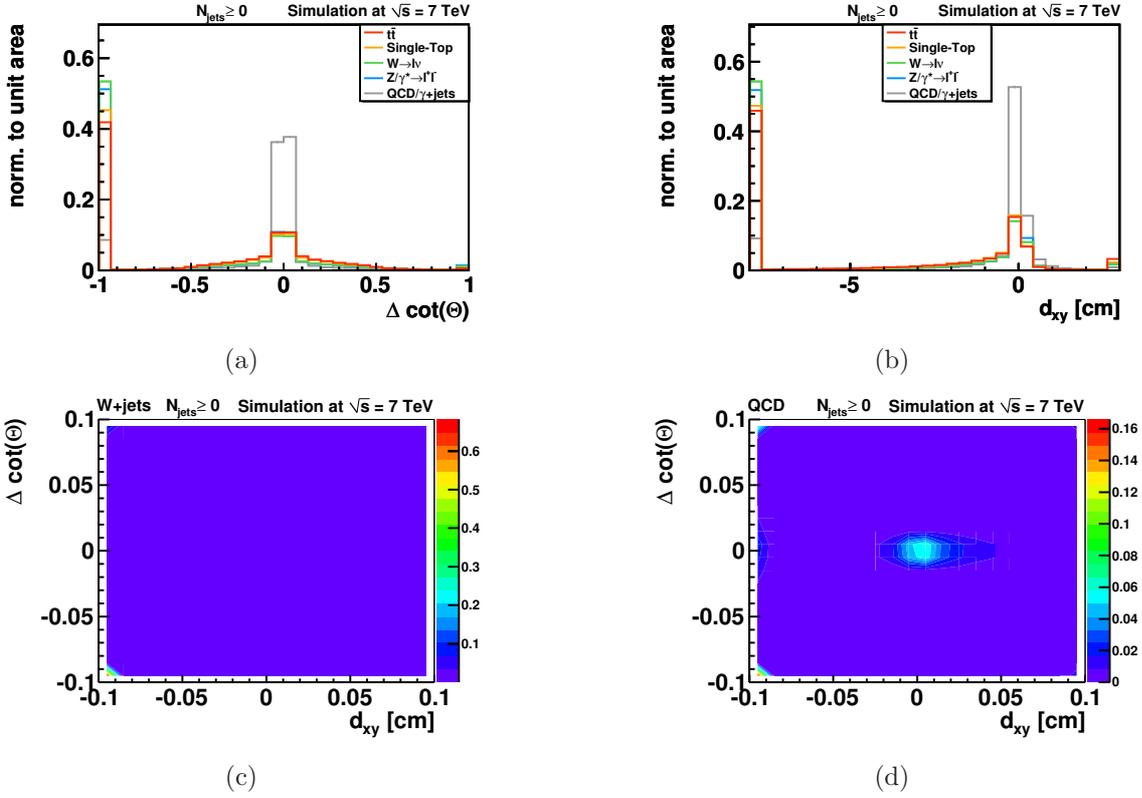


Figure 4.11: Normalized distribution of the variables used in the partner track conversion rejection algorithm. In (a) the $\Delta \cot \theta$ variable is shown, while the distribution of d_{xy} is shown in (b). In both variables large negative values represent the case in which no partner track for the electron GSF track could be found. In (c) and (d) the $\Delta \cot \theta$ - d_{xy} -plane is shown for W +jets and QCD/ γ +jets events, respectively. The different colors represent different relative event fractions which are given in the column on the right hand side of each figure. Only for QCD/ γ +jets events a significant accumulation of events in the region $|\Delta \cot \theta| < 0.2$ and $|d_{xy}| < 0.2$ is observed.

The second algorithm exploits the unique feature, that the two tracks originating from a photon conversion are parallel in the r - z -plane. For this purpose, all tracks within a cone of radius $R = 0.3$ around the GSF track of the selected electron, which exhibit an opposite curvature than the electron track, are selected. For each of these tracks, two geometric variables, the difference in the inverse tangent of the polar angle of the track and the GSF track ($\Delta \cot \theta$) and the two-dimensional distance of the two tracks in the x - y -plane at the point where the tracks are parallel (d_{xy}), are calculated. If the two tracks subtend each other, d_{xy} is defined to be negative. In case of $|\Delta \cot \theta| < 0.02$ and $|d_{xy}| < 0.02$ cm, a so-called partner track is found and the two tracks are likely to arise from a photon conversion. Events for which there is a partner track for the electron GSF track are rejected. The distributions of $\Delta \cot \theta$ and d_{xy} are shown in figure 4.11(a) and (b). Two-dimensional distributions illustrating the $\Delta \cot \theta$ - d_{xy} -plane for W +jets events and the combined QCD multi-jet production and photon+jets

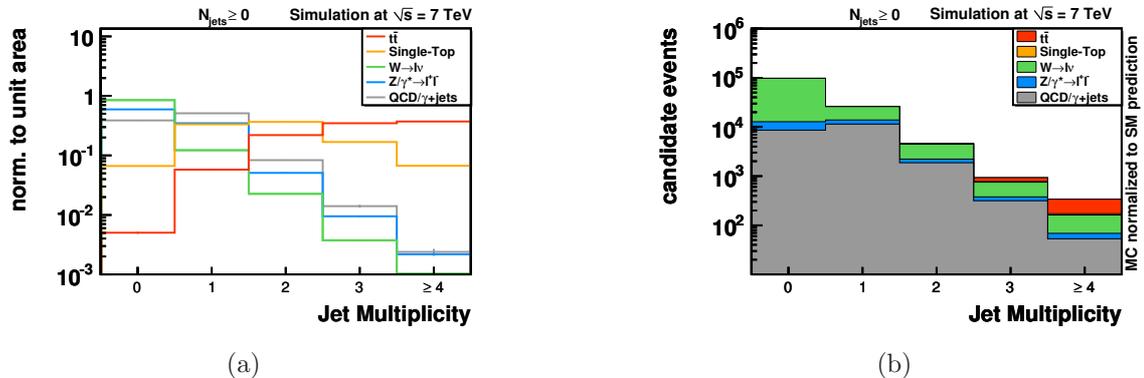


Figure 4.12: Jet multiplicity distribution for events passing the electron selection, the muon and Z boson vetos, and the conversion rejection algorithm. It can be seen, that the background processes tend to have a lower number of jets per event compared to $t\bar{t}$ production. Only events with at least three jets are used for the measurement of the $t\bar{t}$ production cross-section.

production processes are shown in figure 4.11(c) and (d), respectively.

Jet Selection

After selecting events containing well-identified and isolated electrons, the dataset is dominated by events stemming from W boson production. In order to reduce these background events, the characteristic four-jet topology of semileptonic $t\bar{t}$ decays is exploited. In a first step, jets solely originating from calorimeter noise are rejected by applying a set of jet identification criteria. For this purpose, the number of Particle Flow objects forming a jet is required to be larger than one. Moreover, the fraction of a jet's energy arising from charged and neutral electromagnetic objects, as well as neutral hadrons, has to be smaller than 99%. Furthermore the fraction of jet energy carried by charged hadrons has to exceed zero and at least one charged hadron has to be present within a jet. After removing jets stemming from anomalous signals in the calorimeters, a jet within a cone of radius $R = 0.3$ around the selected electron is removed from the event. The reason for this is that isolated electrons have not been excluded from the jet clustering process, and are thus also reconstructed as jets. To avoid counting their energy twice, the jets likely to originate from isolated electrons are identified by the mentioned geometrical criterion and finally removed from the list of jets.

On average, the transverse momentum of jets stemming from quark or gluon radiation processes is smaller than the transverse momentum of jets originating from the decay of a top-quark pair. Therefore, a rather large jet- p_T threshold of 30 GeV is chosen to obtain an efficient rejection of W +jets and QCD multi-jet events. Moreover, the jets stemming from $t\bar{t}$ decays are on average more central than jets originating from radiation processes. Thus, jets are required to be within $|\eta| < 2.4$. The number of selected jets (N_{jets}) for the different processes is shown in figure 4.12(a) and the expected

Selection	$t\bar{t} \rightarrow e+\text{jets}$	$t\bar{t}$ other	Single-Top	$W+\text{jets}$	$Z/\gamma^*+\text{jets}$	QCD/ $\gamma+\text{jets}$
1 electron	38%	4%	14%	10%	11%	0.03%
μ veto	38%	3%	13%	10%	11%	0.03%
Z boson veto	38%	3%	13%	10%	7%	0.03%
conv. rej.	35%	2%	12%	10%	7%	0.01%
$\geq 1\text{jet}$	35%	2%	12%	1%	3%	0.01%
$\geq 2\text{jets}$	34%	2%	7%	0.3%	0.4%	$1 \cdot 10^{-3}\%$
$\geq 3\text{jets}$	28%	1%	3%	0.04%	0.08%	$2 \cdot 10^{-4}\%$
$\geq 4\text{jets}$	15%	0.5%	1%	0.01%	0.01%	$2 \cdot 10^{-5}\%$

Table 4.8: Efficiencies for the different event selection criteria. For events with at least three jets, the efficiency for selecting $t\bar{t}$ pairs decaying into the electron+jets final state is still 28%, while the efficiencies for selecting events without top quarks is at the order of 0.1% or even lower.

composition of the selected dataset for 35.9 pb^{-1} , as a function of N_{jets} , is shown in figure 4.12(b). In contrast to the naive expectation of finding four or more jets for top-quark pairs decaying into the electron+jets final state, $t\bar{t}$ events often exhibit only three jets in the final state. This is mainly due to the high jet- p_T threshold of 30 GeV needed for an efficient rejection of background events which sometimes also rejects jets truly originating from a $t\bar{t}$ decay. Based on this observation, all events with at least three jets are considered for the measurement of the $t\bar{t}$ production cross-section.

Selection Efficiency Derived from Simulations

The efficiencies of the different selection criteria for top-quark pairs decaying into the electron+jets final state, $t\bar{t}$ events decaying into any final state except electron+jets ($t\bar{t}$ other), and all background processes are given in table 4.8. These numbers are derived by applying the event selection criteria to the simulated datasets introduced in section 4.1.2. For $W+\text{jets}$ events with at least three jets in the final state, the efficiency for selecting events where the produced W -boson has decayed leptonically, is smaller than 0.05 %. Also the efficiency for selecting photon+jets events and QCD multi-jet events containing electrons from b - and c -hadron decays or isolated electromagnetic signatures is almost negligible. The efficiency for selecting top-quark pairs decaying into the electron+jets final state is still 28%. Only 1% of $t\bar{t}$ events decaying into a different final state are selected, where most of these events arise from the electron-tau, the tau+jets, and the di-electron final state. The exact composition of the selected $t\bar{t}$ final states is illustrated in figure 4.13.

4.3 Electron Selection and Trigger Efficiency

So far, the event selection efficiencies presented in table 4.8 are derived based on simulated datasets only. Due to an imperfect modelling of the detector, the selection efficiencies derived from simulation might differ from the efficiencies observed in real

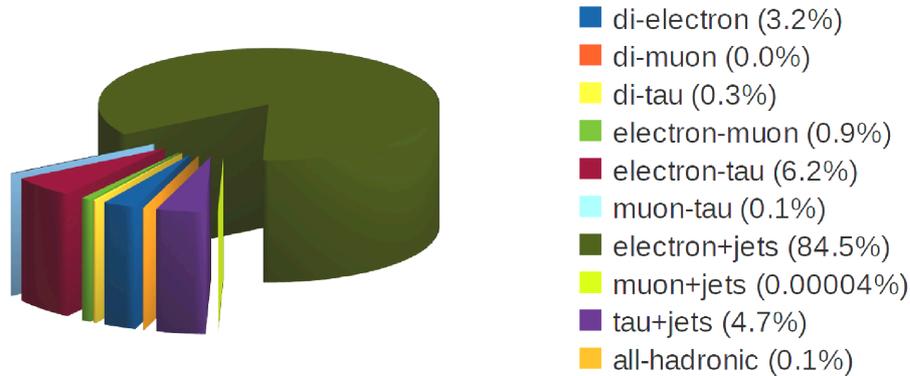


Figure 4.13: Selected $t\bar{t}$ events split into the different decay channels. As intended, the largest contribution arises from top-quark pairs decaying into the electron+jets final state. Moreover, the electron-tau, the tau+jets, and the di-electron final state contribute significantly.

collision data. Since the predictions obtained from MC simulations form the basis for the measurement of the $t\bar{t}$ production cross-section, any mis-modelling would lead to an incorrect interpretation of the obtained results. As a consequence, the estimation of selection efficiencies using real collision data, is a key-point for obtaining reliable results for the $t\bar{t}$ production cross-section. Moreover, the rapidly increasing instantaneous luminosity of the LHC during the 2010 operation necessitated the use of the quickly changing HLT trigger thresholds given in table 4.4. Unfortunately, most of these triggers could not be implemented in MC simulations due the vastly changing conditions. Therefore, a direct measurement of the trigger efficiency from collision data is essential.

An approach frequently used to estimate the reconstruction, selection, and trigger efficiencies of leptons is the so-called tag-and-probe method. It takes advantage of the known invariant mass of narrow resonances decaying into oppositely charged lepton pairs, like J/Ψ mesons and Z bosons for example. In the analysis described in this thesis, the Z boson resonance is used, because of the similar kinematics of electrons originating from Z boson and top-quark decays. The method starts by selecting events containing at least one electron fulfilling stringent quality requirements, to ensure that the selected electron is well-reconstructed and not a different particle that mimics an electron's signature. These electrons are called tag-electrons in the following. In all selected events, a second electron passing less stringent quality requirements is required. If the invariant mass of the tag-electron and the second electron is found to be within a narrow window around the Z boson mass, the second electron is called a probe-electron and the combination of the two is called a tag-and-probe pair. Although only less stringent quality requirements are applied to the probe, the employed invariant mass window ensures that also the probe-electron is very likely to be an electron. The probe-electron thus provides the unique feature of studying the efficiency of a particular selection criterion at an object that has been truly identified of being an electron. For

this purpose, the number of events before and after applying the selection criterion on the probe-electron have to be counted. The corresponding efficiency can then be calculated by

$$\epsilon = \frac{2 \cdot N_{\text{TT}} + N_{\text{TP}}}{2 \cdot N_{\text{TT}} + N_{\text{TP}} + N_{\text{TP}}} = \frac{N_{\text{sel}}^{\text{TaP}}}{N_{\text{all}}^{\text{TaP}}}. \quad (4.7)$$

In this equation N_{TT} represents the number of events where both electrons fulfill the stringent tag requirements and N_{TP} represents the number of events where the probe-electron passes the studied selection criterion. Instead, N_{TP} represents the number of events where the probe-electron fails the studied selection criterion. If the tag-and-probe pairs where both electrons fulfill the tag requirement are counted twice (tag-electron and probe-electron could be exchanged), the equation can be simplified by just counting the number of tag-and-probe pairs before ($N_{\text{all}}^{\text{TaP}}$) and after the selection criterion is applied ($N_{\text{sel}}^{\text{TaP}}$).

In the following measurements, the window around the Z boson mass is set to the same size as the window used in the Z boson veto, i.e. $76 \text{ GeV}/c^2 \leq M_{ee} \leq 106 \text{ GeV}/c^2$. This choice of the window size ensures that only events which are not present in the dataset intended for the measurement of the $t\bar{t}$ production cross-section are used for the estimation of the efficiencies. Moreover, events with more than one (disparate) tag-and-probe pair are rejected. The uncertainties on the derived efficiencies are quoted using Clopper-Pearson [173] 68% confidence intervals.

4.3.1 Electron Selection-Efficiency

For the measurement of the electron selection-efficiency, the tag-electron is required to pass all electron selection criteria mentioned in section 4.2.2, including the conversion rejection algorithms. In the beginning, the probe-electron is required to have a super-cluster with a transverse energy larger than 20 GeV which is not located in the ECal barrel-endcap transition region ($1.4442 < |\eta_{\text{SC}}| < 1.566$), and the electron is required to be in the acceptance of the tracking detector, i.e. $|\eta| < 2.5$. All of the remaining selection criteria are subsequently applied to the probe-electron and the corresponding efficiencies are derived. Since the number of jets in an event might have an influence on the electron selection-efficiency, it would be ideal to use events with at least three jets for the measurement. However, to avoid large statistical uncertainties on the derived efficiencies, the requirement on the number of jets has to be lowered to ≥ 1 jets. In the future, when a larger amount of collision data will have been collected, the electron selection-efficiency will be derived using events with a larger number of jets.

Since it is impossible to directly measure the electron selection-efficiency in events with at least three jets, the strategy pursued in this analysis, is to use correction factors C_{eff} to account for possible differences between the simulation and real collision data. The correction factors are derived using the following equation:

$$\epsilon^{\text{data}} = \frac{\epsilon_{\text{TaP}}^{\text{data}}}{\epsilon_{\text{TaP}}^{\text{MC}}} \cdot \epsilon^{\text{MC}} = C_{\text{eff}} \cdot \epsilon^{\text{MC}}. \quad (4.8)$$

Selection Criteria	$Z/\gamma^*+\text{jets}$	$t\bar{t}$	Single-Top	$W+\text{jets}$	QCD/ $\gamma+\text{jets}$
$E_T^{\text{SC}} > 20 \text{ GeV}$	2366	50	5	89	62
$E_T > 30 \text{ GeV}$	1849	31	3	48	29
$z \text{ dist. to PV} < 1 \text{ cm}$	1849	31	3	47	29
$d_0 \text{ (beam spot)} < 200 \mu\text{m}$	1810	28	3	44	24
WP70 fulfilled	1600	9	1	0	0
$N_{\text{hits}}^{\text{mis}} = 0$	1532	9	0	0	0
partner track veto	1504	8	0	0	0
$I_{\text{rel}} < 0.1$	1444	7	0	0	0

Table 4.9: Expected composition of the dataset selected for the measurement of the electron selection-efficiency. Especially before the WP70 requirement a substantial amount of background events are expected, which could lead to a significantly lower efficiency.

Here ϵ^{data} and ϵ^{MC} denote the electron selection-efficiencies in the dataset used to measure the $t\bar{t}$ cross-section in real collision data and simulation, respectively. The correction factor is given by the ratio of the efficiency derived from data and simulation using the tag-and-probe method in a dataset containing mostly Z bosons. Equation 4.8 is clearly based on the assumption that any imperfect modelling of the variables used in the electron selection criteria, is present in the simulation of Z events and, with a similar magnitude, present in all other simulated datasets, too. Since an imperfect modelling of these variables is mainly due to an imperfect detector calibration used for the detector simulation, this assumption is reasonable.

The data-to-MC correction factor for the efficiency of reconstructing electrons passing the initial probe requirement has been derived in [174] and used for the measurements described in [175]. Therein, two separate correction factors for electrons in the ECal barrel and endcap have been measured: $C_{\text{reco}}^{\text{EB}} = 1.00 \pm 0.01$ and $C_{\text{reco}}^{\text{EE}} = 1.00 \pm 0.02$. In this analysis, no distinction is made between EB and EE electrons. Hence, the two results are combined into a single overall correction factor $C_{\text{reco}} = 1.00 \pm 0.02$, where the larger uncertainty has been used. This result which implies that no correction has to be applied to account for differences in the electron reconstruction efficiency is the starting point for the efficiency measurements described in the following. The electron selection-efficiency correction factor derived in this thesis, is factorized into three independent data-to-MC correction factors accounting for differences in the acceptance of electrons with a transverse energy larger than 30 GeV (C_{acc}), differences in the efficiency of the isolation criterion (C_{iso}), and differences in the efficiency of the remaining selection criteria (C_{id}). The combined data-to-MC correction factor is then given by:

$$C_{\text{eff}} = C_{\text{reco}} \cdot C_{\text{acc}} \cdot C_{\text{id}} \cdot C_{\text{iso}}. \quad (4.9)$$

Background Estimation Technique

Despite the requirement that the invariant mass of the tag-and-probe pair is close to the Z boson mass, there might still be background contributions for the cases where only loose quality requirements are applied to the probe. Since the probability for passing the more stringent selection criteria is much smaller for objects erroneously identified as electrons than for real electrons, the presence of background electrons reduces the measured efficiency significantly. For this purpose, it is necessary to subtract background electrons from the dataset used to measure the efficiencies. In table 4.9, the simulation-based expectation for the composition of the dataset is shown, assuming an integrated luminosity of 35.9 pb^{-1} . It can be seen, that especially prior to the application of the WP70 selection requirement, background electrons are present which mainly arise from charged pions and kaons erroneously identified as electrons.

For the estimation of the backgrounds, the same-sign/opposite-sign method is used. This method is based on the assumption, that the two electrons from all relevant background processes are produced with identical fractions with same charges and opposite charges. Since Z bosons always decay into pairs of oppositely-charged (opposite-sign) electrons, the number of events with equally-charged (same sign) electrons is a good approximation for the number of background events in the selected dataset. Unfortunately, there is a non-negligible fraction of events where the electron charge is measured with the wrong sign. This charge misidentification effect has to be taken into account, too. The only background where the mentioned assumption is not valid, is the decay of $t\bar{t}$ pairs into the di-electron final state. However, as can be seen in table 4.9 their contribution is very small and can be neglected.

When averaging over all electrons the number of true Z boson events (N_Z) can be calculated according to

$$N_Z = \frac{N_{OS} - N_{SS}}{(1 - 2q^{\text{misid}})^2}, \quad (4.10)$$

where N_{OS} is the number of events with oppositely charged electrons, N_{SS} is the number of events with equally-charged electrons and q^{misid} is the electron charge-misidentification probability. This charge misidentification can be measured in an almost background free dataset containing solely Z boson events, according to

$$q^{\text{misid}} = \frac{1 - \sqrt{|\langle q_1 q_2 \rangle|}}{2}, \quad (4.11)$$

where $\langle q_1 q_2 \rangle$ is the averaged product of the two electron charges. A derivation of the equations 4.10 and 4.11 is given in the appendix section A.3.

In order to check after which probe selection criterion the selected dataset is sufficiently background free to obtain a reliable measurement of the charge misidentification probability, the q^{misid} values are derived from Z/γ^* +jets MC truth, from pure Z/γ^* +jets MC, from a dataset with a realistic mixture of all relevant processes (scaled to 35.9 pb^{-1}) referred to as **A11 MC**, and from collision data. For the derivation from MC truth the common tag-and-probe selection is applied, requiring in addition

Selection Criteria	Z/γ^* +jets MC truth	Z/γ^* +jets MC	A11 MC	data
$E_T^{\text{SC}} > 20$ GeV	0.0082 ± 0.0003	0.0092 ± 0.0003	0.028 ± 0.003	0.031 ± 0.003
$E_T > 30$ GeV	0.0078 ± 0.0003	0.0083 ± 0.0003	0.021 ± 0.003	0.023 ± 0.003
z dist. to PV < 1 cm	0.0078 ± 0.0003	0.0084 ± 0.0003	0.021 ± 0.003	0.023 ± 0.003
d_0 (beam spot) < 200 μm	0.0063 ± 0.0003	0.0069 ± 0.0003	0.018 ± 0.002	0.020 ± 0.002
WP70 fulfilled	0.0042 ± 0.0002	0.0042 ± 0.0002	0.004 ± 0.001	0.007 ± 0.001
$N_{\text{hits}}^{\text{mis}} = 0$	0.0041 ± 0.0002	0.0041 ± 0.0002	0.004 ± 0.001	0.007 ± 0.002
partner track veto	0.0036 ± 0.0002	0.0036 ± 0.0002	0.004 ± 0.001	0.007 ± 0.002
$I_{\text{rel}} < 0.1$	0.0036 ± 0.0002	0.0036 ± 0.0002	0.004 ± 0.001	0.007 ± 0.002

Table 4.10: Charge misidentification probabilities (q^{misid}) derived from the MC truth information, pure Z/γ^* +jets MC, a realistic mixture of all relevant processes (A11 MC) and real collision data. After applying the WP70 criterion the dataset is sufficiently background free, and the values derived from Z/γ^* +jets MC and A11 MC agree with the value obtained from MC truth information. The q^{misid} values derived from data are higher than what is expect from simulation, but both values still agree within the uncertainties.

that both electrons match with a generated electron from the Z boson decay within $\Delta R < 0.1$. The results for q^{misid} are summarized in table 4.10. The tables shows, that after applying the WP70 criterion the dataset is sufficiently background free, and consequently, the q^{misid} values derived from Z/γ^* +jets MC and A11 MC is in good agreement with the MC truth value. The table also shows, that the q^{misid} value measured in data is higher than what is expected from the simulation, however both values are still in agreement within the uncertainties.

Unfortunately, the charge-misidentification probability depends on the selection criteria applied to the electrons. In addition, the measurement of q^{misid} from collision data is only reliable after the WP70 criterion has been applied. For this reason, the value measured in the most background free dataset, i.e. after the isolation requirement, is used as the reference q^{misid} value in each analyzed scenario. The charge-misidentification probability for all the other selection steps is derived by scaling this reference value with the ratio of the corresponding MC truth value and the MC truth reference value. Following this procedure, the q^{misid} values listed in table 4.11 are used to finally derive the efficiencies of the different electron selection criteria. For the selection criteria applied after the WP70 requirement, the scaled q^{misid} values for data can be compared with the ones directly measured. It can be seen that any systematic effect which could arises from the usage of the MC truth information for the derivation of the proper q^{misid} values, is covered by the quoted statistical uncertainties.

To ensure the validity of the method outlined so far, a closure test is performed based on simulated events. For this purpose, the efficiency for the different selection criteria are derived using the MC truth information, a pure Z/γ^* +jets sample, and a realistic mixture of all relevant processes (scaled to 35.9 pb^{-1}). For this closure test, the scaled charge misidentification probabilities shown in table 4.11 are used for the derivation of the background contribution in the selected dataset. A comparison of

Selection Criteria	MC truth scaling factor	Z/γ^* +jets MC (scaled)	All MC (scaled)	data (scaled)
$E_T^{\text{SC}} > 20$ GeV	2.28 ± 0.15	0.0082 ± 0.0007	0.009 ± 0.002	0.016 ± 0.005
$E_T > 30$ GeV	2.17 ± 0.15	0.0078 ± 0.0007	0.009 ± 0.002	0.015 ± 0.004
z dist. to PV < 1 cm	2.17 ± 0.15	0.0078 ± 0.0007	0.009 ± 0.002	0.015 ± 0.004
d_0 (beam spot) < 200 μm	1.75 ± 0.13	0.0063 ± 0.0006	0.007 ± 0.002	0.012 ± 0.004
WP70 fulfilled	1.17 ± 0.09	0.0042 ± 0.0004	0.005 ± 0.001	0.008 ± 0.002
$N_{\text{hits}}^{\text{mis}} = 0$	1.14 ± 0.08	0.0041 ± 0.0004	0.005 ± 0.001	0.008 ± 0.002
partner track veto	1.00 ± 0.08	0.0036 ± 0.0003	0.004 ± 0.001	0.007 ± 0.002
$I_{\text{rel}} < 0.1$	-	0.0036 ± 0.0002	0.004 ± 0.001	0.007 ± 0.002

Table 4.11: Charge misidentification probabilities used to derive the electron selection efficiency. In each scenario (Z/γ^* +jets MC, All MC, and data) the q^{misid} value after the isolation requirement is chosen as the reference value. To calculate q^{misid} for all other selection steps, this reference value is scaled by the ratio of the corresponding MC truth q^{misid} value and the MC truth reference value. The scaling factors are shown in the second column, while the scaled q^{misid} values for the different scenarios are shown in the remaining columns.

the selection efficiencies obtained for the different scenarios is given in table 4.12. As can be seen, the efficiencies derived from the pure Z/γ^* +jets sample and the realistic mixture are in very good agreement with the efficiencies derived from the MC truth information. Consequently, the described method is likely to give reliable results. The uncertainties given in table 4.12 consist of two different parts. The first part is the statistical uncertainty on the central efficiency value which is derived using Clopper-Pearson 68% confidence intervals. The second part arises from the uncertainty on the used q^{misid} value. Due to the consecutively applied selection criteria, the q^{misid} values are highly correlated. To account for this, all q^{misid} values are fluctuated up or down within their uncertainties at the same time. The difference between the efficiency calculated with the shifted q^{misid} values, and the central efficiency is used as the corresponding uncertainty. Finally, both uncertainties are added in quadrature to obtain the uncertainties given in table 4.12.

Correction Factor Determination

In order to derive the data-to-MC correction factor accounting for the differences in the acceptance of electrons with $E_T > 30$ GeV, tag-and-probe pairs where the tag-electron passes all electron selection criteria and the probe-electron has a supercluster with $E_T > 20$ GeV which is not located in the ECal barrel-endcap transition region, are employed. The acceptance for the selection criterion is derived by applying it also to the probe and counting the number of tag-and-probe pairs passing this additional requirement. To account for background contamination, the number of true Z boson events is calculated using equation 4.10. Before the additional criterion is applied to the probe, the charge misidentification probability for simulated Z boson events is 0.0082 ± 0.0007 while it is 0.016 ± 0.005 for real collision data, see table 4.11. In the

Selection Criteria	MC truth	Z/γ^* +jets MC	All MC
$E_T > 30$ GeV	0.783 ± 0.002	0.783 ± 0.002	0.786 ± 0.009
z dist. to PV < 1 cm	0.9996 ± 0.0001	0.9997 ± 0.0001	$1.000 - 0.001$
d_0 (beam spot) < 200 μm	0.9798 ± 0.0007	0.980 ± 0.001	0.977 ± 0.004
WP70 fulfilled	0.887 ± 0.002	0.887 ± 0.002	0.885 ± 0.008
$N_{\text{hits}}^{\text{mis}} = 0$	0.958 ± 0.001	0.958 ± 0.001	0.958 ± 0.006
partner track veto	0.9815 ± 0.0007	0.982 ± 0.001	0.979 ± 0.004
$I_{\text{rel}} < 0.1$	0.960 ± 0.001	0.960 ± 0.001	0.960 ± 0.006

Table 4.12: In order to proof the reliability of the background estimation method, the efficiencies for the different selection criteria are derived from pure Z/γ^* +jets MC as well as a realistic MC mixture containing all relevant processes and compared with the efficiencies derived from MC truth. In general a good agreement is found which shows that the method is likely to give reliable results.

dataset, where the additional requirement has been applied to the probe, q^{misid} values of 0.0078 ± 0.0007 and 0.015 ± 0.004 have to be used for simulated Z boson events and real collision data, respectively. The derived acceptances are:

$$\begin{aligned} Z+\text{jets MC} : \epsilon_{E_T} &= 0.783 \pm 0.002, \\ \text{data} : \epsilon_{E_T} &= 0.775 \pm 0.009. \end{aligned} \tag{4.12}$$

Hence, the data-to-MC correction factor accounting for differences in the acceptance of electrons with $E_T > 30$ GeV is given by:

$$C_{\text{acc}} = 0.99 \pm 0.01.$$

The correction factor is very close to unity, representing a very good agreement between data and simulation. To ensure that this correction factor is independent of the detector region in which the electron has been reconstructed and also insensitive to the presence of additional jets in the event, the correction factor is additionally calculated as a function of the electron's azimuthal angle ϕ , the η of the electron, and the distance in the η - ϕ -plane between the electron and the closest jet. In figure 4.14 these differential correction factors are shown together with the overall correction factor and its uncertainties. With the current precision, no significant deviations between the overall and the differential correction factors can be found.

For the derivation of the data-to-MC correction factor which accounts for differences in the efficiencies for all selection criteria except for the isolation requirement tag-and-probe pairs where the tag-electron passes all electron selection criteria and the probe-electron has a transverse energy larger than 30 GeV are used. The q^{misid} values used in the background estimation for this dataset, have been already given above. In a second step, the z distance between the electron and the primary vertex, the impact parameter with respect to the beam spot, the WP70, and both conversion rejection selection criteria are applied to the probe, and the number of remaining tag-and-probe

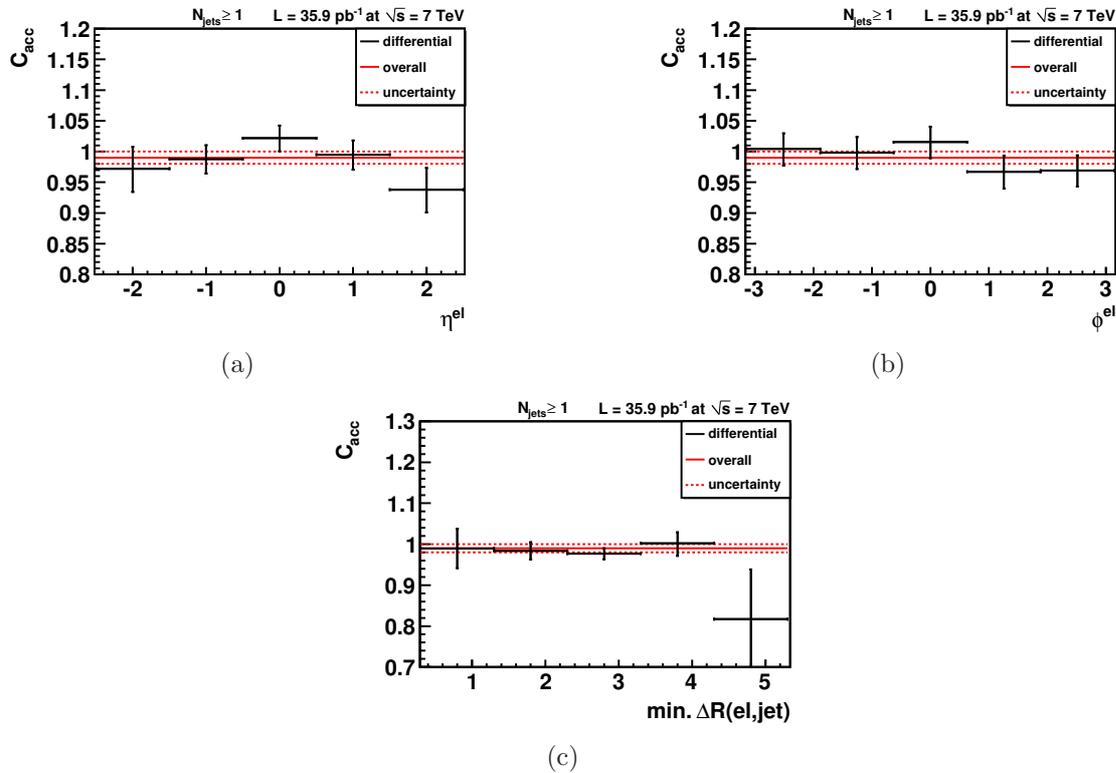


Figure 4.14: Differential distributions for the data-to-MC correction factor accounting for differences in the acceptance of electrons with $E_T > 30 \text{ GeV}$. The correction factor C_{acc} is given as a function of the η of the electron (a), the ϕ of the electron (b), and the distance of the electron and the closest jet in the η - ϕ -plane (c). For all three variables, the differential correction factor is consistent with the overall correction factor within the given uncertainties. The uncertainty on the overall correction factor is represented by the red dotted lines.

pairs is derived. For the estimation of the number of true Z boson events, q^{misid} values of 0.0036 ± 0.0003 and 0.007 ± 0.002 are used for simulated Z boson events and real collision data, respectively. The resulting efficiencies are:

$$\begin{aligned} Z+\text{jets MC} : \epsilon_{\text{id}} &= 0.816_{-0.003}^{+0.002}, \\ \text{data} : \epsilon_{\text{id}} &= 0.78 \pm 0.01. \end{aligned} \quad (4.13)$$

This leads to a data-to-MC correction factor accounting for differences in the electron identification process of

$$C_{\text{id}} = 0.96 \pm 0.01.$$

In this case, the correction factor is significantly smaller than one. Most of the observed differences arise from the WP70 requirement, which indicates an imperfect modelling of the variables describing the shape of the electromagnetic shower initiated by an electron. However, neither a specific detector region nor a particular electron phase space could be identified which led to the observed inefficiency. Moreover, the behavior of C_{id}

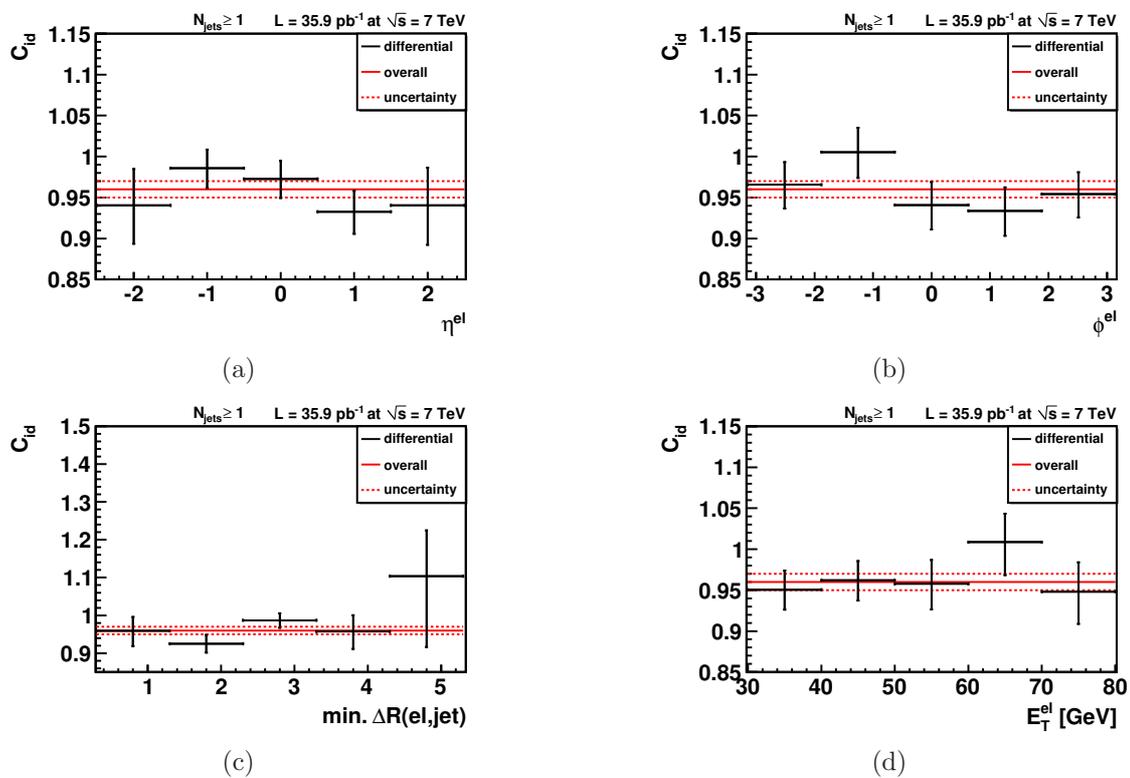


Figure 4.15: Differential distributions for the data-to-MC correction factor accounting for differences in the efficiencies for all selection criteria except for the isolation requirement. In addition to figure 4.14, the correction factor is shown also as a function of the electron’s transverse energy (d). In each differential distribution, the correction factor is consistent with the overall correction factor within the given uncertainties.

as a function of the electron transverse energy and of the variable already used above is studied. The resulting differential correction factors are compared with the overall correction factors in figure 4.15. Taking into account the uncertainties on both factors, no significant deviations between the differential and the overall correction factor is observed.

The data-to-MC correction factor accounting for differences between simulation and collision data in the electron isolation is derived using tag-and-probe pairs where the tag-electron passes all electron selection criteria and the probe-electron passes all selection criteria except that no isolation requirement is applied. The q^{misid} values for this dataset have already been mentioned above. The efficiency of the isolation criterion is derived by applying the isolation requirement also to the probe, and counting the surviving tag-and-probe pairs. The charge misidentification probabilities used in the background estimation for this dataset are 0.0036 ± 0.0002 for $Z/\gamma^* + \text{jets}$ simulation

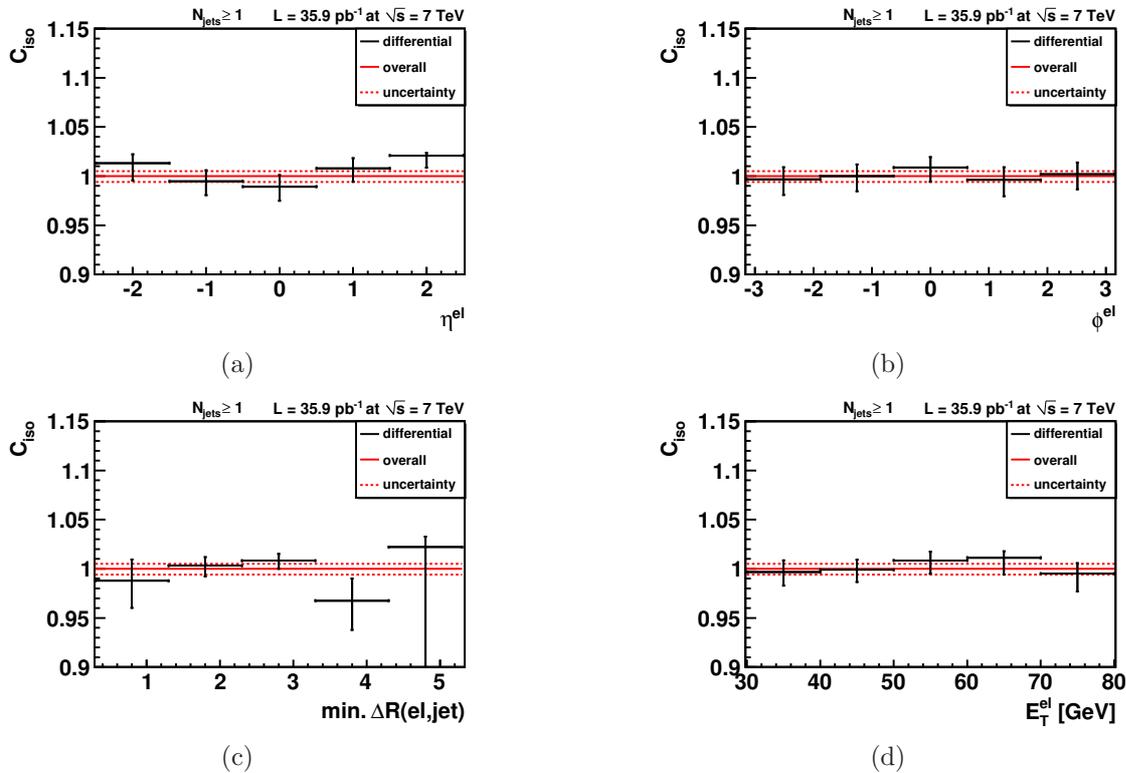


Figure 4.16: Differential distributions for the data-to-MC correction factor accounting for differences in the efficiencies for the electron isolation requirement. In each differential distribution, the correction factor is consistent with the overall correction factor within the given uncertainties.

and 0.007 ± 0.002 for collision data. In summary, the following efficiencies are derived:

$$\begin{aligned} Z+\text{jets MC} : \epsilon_{\text{iso}} &= 0.960 \pm 0.001, \\ \text{data} : \epsilon_{\text{iso}} &= 0.960^{+0.005}_{-0.006}. \end{aligned} \quad (4.14)$$

A very good agreement between simulation and data is found, leading to a data-to-MC correction factor of

$$C_{\text{iso}} = 1.000^{+0.005}_{-0.006}.$$

As for the correction factors derived above, the C_{iso} is additionally measured as a function of four variables. The resulting differential correction factors are shown in figure 4.16 and are compared with the corresponding overall correction factor. Within the uncertainties, all differential correction factors are in agreement with the overall correction factor.

The three correction factors derived so far, are combined with the data-to-MC correction factor accounting for differences in the electron reconstruction into a final correction factor for the electron selection:

$$C_{\text{eff}} = C_{\text{reco}} \cdot C_{\text{acc}} \cdot C_{\text{id}} \cdot C_{\text{iso}} = 0.950 \pm 0.025. \quad (4.15)$$

Trigger Name	Trigger Efficiency
HLT_Ele15_SW_L1R	$0.991^{+0.003}_{-0.005}$
HLT_Ele17_SW_CaloEleId_L1R	$0.994^{+0.002}_{-0.003}$
HLT_Ele17_SW_TightEleId_L1R	0.975 ± 0.003
HLT_Ele22_SW_TighterEleId_L1R_v2	$0.984^{+0.002}_{-0.003}$
HLT_Ele22_SW_TighterEleId_L1R_v3	0.977 ± 0.003

Table 4.13: Electron trigger-efficiencies derived using the tag-and-probe method. Although all efficiencies are close to unity, an inefficiency due to applying thresholds on the electron identification observables already at the HLT level, is clearly visible.

This correction factor is used as an additional weight which is applied to all simulated events passing the electron selection criteria.

4.3.2 Electron Trigger-Efficiency

Due to the absence of a proper simulation of the employed HLT triggers, there is no trigger requirement imposed to simulated events. Rather than deriving a data-to-MC correction factor, the efficiencies of the different triggers is directly measured using collision data. Since all of the employed triggers are only sensitive to the presence of at least one electron, again the tag-and-probe method is well-suited for the derivation of the efficiencies. The efficiency values, determined in the following, always represent the efficiency for an electron passing all selection criteria to additionally pass the trigger requirement. Since there is no isolation requirement applied at HLT, the trigger efficiencies are unlikely to depend on the jet activity present in the event. Therefore, no selection is made based on the jet multiplicity in contrast to the measurement of the electron selection efficiency discussed in the previous section.

For each HLT trigger, a list of objects fulfilling the trigger requirements is present in an event. If, for example, an event contains two electron HLT objects which both fulfill the trigger requirements, both objects are present in the mentioned list. In order to judge, if an offline selected electron would have passed the trigger, a matching between the electron and these HLT objects is performed. All electrons found to match an HLT object within $\Delta R < 0.2$ are thus identified as passing the trigger.

For the measurement of the trigger efficiencies, tag-and-probe pairs where the tag-electron and the probe-electron are both fulfilling all electron selection criteria and the tag-electron is as well matched to an HLT object are employed. Since the selected dataset is almost background free, no dedicated background estimation is used. To finally derive the trigger efficiencies, the number of tag-and-probe pairs where also the probe-electron could be matched to an HLT object (different from the tag-electron HLT object) are counted. The measured efficiencies are given in table 4.13. The efficiency of all triggers is close to unity, however, it can be seen that the application of thresholds on electron identification observables already at HLT level, led to small inefficiencies

Selection	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets	QCD/ γ +jets
1 electron	549 ± 87	145 ± 9	$106,928 \pm 7,422$	$11,480 \pm 719$	$62,421 \pm 3,069$
μ veto	488 ± 77	140 ± 9	$106,914 \pm 7,422$	$11,431 \pm 716$	$62,404 \pm 3,068$
Z boson veto	481 ± 76	139 ± 9	$106,892 \pm 7,420$	$7,500 \pm 470$	$62,356 \pm 3,066$
conv. rej.	453 ± 71	131 ± 8	$100,390 \pm 6,969$	$6,969 \pm 437$	$22,321 \pm 1,147$
≥ 1 jet	451 ± 71	122 ± 8	$14,874 \pm 1,033$	$2,853 \pm 179$	$13,664 \pm 717$
≥ 2 jets	424 ± 67	79 ± 5	$2,716 \pm 189$	436 ± 28	$2,230 \pm 150$
≥ 3 jets	325 ± 51	31 ± 2	468 ± 33	81 ± 5	367 ± 41
≥ 4 jets	168 ± 26	9 ± 1	94 ± 7	15 ± 1	53 ± 12

Table 4.14: Number of expected events per process after applying a particular selection criterion. The quoted uncertainties are due to the employed theoretical cross-section, the luminosity measurement, the finite number of simulated events passing a particular requirement, and the electron selection and trigger-efficiency measurements.

of the triggers. Since the amount of data collected with the HLT_E1e10_LW_L1R and HLT_E1e15_SW_CaloEleId_L1R are too small to obtain reliable results for the corresponding trigger efficiencies, the same trigger efficiencies as for HLT_E1e15_SW_L1R and HLT_E1e17_SW_CaloEleId_L1R are used. The differences between these sets of triggers are mainly due to the larger E_T -thresholds applied to the HLT objects. Since these thresholds are well below the final electron E_T -threshold, no differences in the trigger efficiencies are expected due to that.

The combined electron trigger-efficiency is derived by calculating the luminosity-weighted mean of the efficiencies of the single HLT triggers. This results in an efficiency of:

$$\epsilon_{\text{HLT}} = 0.982 \pm 0.001. \quad (4.16)$$

In order to account for the finite trigger efficiency, each simulated event passing all electron selection criteria is weighted accordingly.

4.4 Comparison of Simulation and Collision Data

In high-energy physics, the current knowledge about the Standard Model of particle physics and the interactions of particles with the detector used to collect real collision data is implemented in MC simulations. Deviations observed between the simulation and the recorded data are often speculated to arise from phenomena which are not described by the SM. However, the simulation of the hard interaction, the parton shower, the hadronization, as well as the simulation of the interactions of particles with the detector material have known weaknesses. It is therefore crucial to compare the predictions of the simulation with real data and to judge if possible deviation arise from an imperfect MC simulation or really from yet-unknown physics processes.

The data-to-MC correction factor accounting for differences in the electron selection-efficiency between data and simulation, as well as the trigger efficiency derived in the previous section are used to correct the predictions of the MC simulation. In table 4.14,

Selection	MC Sum	data
1 electron	$181,523 \pm 9,042$	175,625
μ veto	$181,377 \pm 9,036$	175,428
Z boson veto	$177,368 \pm 8,901$	171,077
conv. rej.	$130,264 \pm 7,248$	128,435
≥ 1 jet	$31,964 \pm 1,500$	34,850
≥ 2 jets	$5,885 \pm 291$	7,003
≥ 3 jets	$1,272 \pm 80$	1,611
≥ 4 jets	339 ± 31	428

Table 4.15: Number of events after applying a particular selection criterion in real collision data and simulation. The quoted uncertainties are due to the employed theoretical cross-section, the luminosity measurement, the finite number of simulated events passing a particular requirement, and the electron selection and trigger-efficiencies.

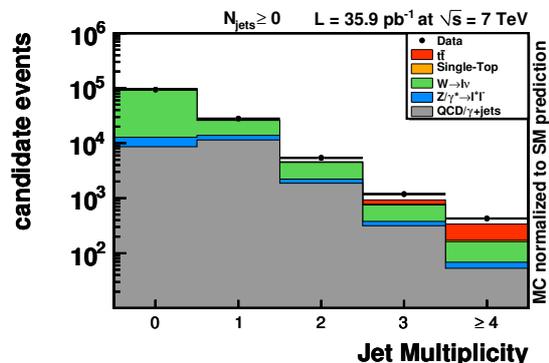


Figure 4.17: Comparison of the predicted jet multiplicity distribution with the observed distribution from real collision data.

the predicted number of events per process after applying a particular selection criterion are shown. A comparison between the predicted number of events and real data is given in table 4.15. At first sight, the prediction and the observation seem to agree well up to the requirements on a specific number of jets. Beginning from the requirement of at least one jet, there are slight differences between data and simulation, which grow with an increasing number of jets. The jet multiplicity distribution as predicted from simulation is compared with the observed jet multiplicity distribution in figure 4.17. There are several reasons why there could be an imperfect modelling of the jet multiplicity distribution. First, all employed simulated datasets are generated with LO event generators. Although additional radiations of gluons are already present in the matrix element for most of the generated processes, a simulation based on NLO calculations is likely to improve the modelling of the jet multiplicity distribution. Moreover, the choice of the factorization and renormalization scale in the generation of the samples is to some extent arbitrary. However, variations of these two scales can lead to large differences in the jet multiplicity distribution. Finally, the simulation of QCD multi-jet events with isolated electrons in the final state is very challenging, since the production and decay of heavy hadrons and π^0 -mesons occurs during the hadronization process which cannot be calculated in perturbative QCD. In addition, isolated electrons in the final state are very rare in QCD multi-jet processes, leading to an insufficient number of simulated events passing all selection criteria.

In figure 4.18 a comparison of data and simulation for events with exactly three jets is shown for different kinematic observables. Besides the transverse energy of the electron and the missing transverse energy, the transverse W boson mass, $M_T(W)$, and

the $H_{T,\text{lep}}$ variable are depicted. The transverse W boson mass is defined as

$$M_T(W) = \sqrt{2 \cdot \frac{E_{T,\text{ele}} \cdot \cancel{E}_T}{c^4} (1 - \cos \Delta\phi)}, \quad (4.17)$$

where $\Delta\phi$ is the difference in the azimuthal angle of the electron and the missing transverse energy direction. This variable is very useful, since it provides a measure for the invariant mass of the electron and the potential neutrino without explicitly reconstructing the amount of neutrino momentum parallel to the beam line, which would only be possible by employing the W boson mass as constraint. If $M_T(W)$ is close to the W boson mass, the \cancel{E}_T is likely to stem from an undetected neutrino and the electron and the neutrino are likely to originate from a W boson decay. The $H_{T,\text{lep}}$ variable is defined as the scalar sum of the electron's transverse energy and \cancel{E}_T and can be used to distinguish W boson and $t\bar{t}$ events from QCD multi-jet processes. In figure 4.18 each variable is shown twice, once the MC simulation is scaled to the SM expectation and once to the number of events observed in data. In the latter case, the overall difference between data and simulation is corrected. It is visible, that in these illustrations the agreement between data and simulation is much improved. Hence, the shape of the distributions observed in data, are sufficiently reproduced by the simulation. Though, care has still to be taken for the overall normalization and relative fractions of the different processes.

In figure 4.19 similar comparisons are shown for events with at least four jets. Besides the transverse momenta of the jet with largest and the second largest p_T , the M3 and H_T variables are illustrated. The M3 variable is defined as the invariant mass of those three jets which exhibit the largest vectorial-summed transverse momentum. For $t\bar{t}$ events, M3 is a very simple estimator for the mass of the hadronically decaying top quark. In about 30% of all cases, the three jets arising from the top-quark decay are correctly identified. The H_T variable is defined as the scalar sum of the electron's E_T , the \cancel{E}_T , and the p_T of all jets.

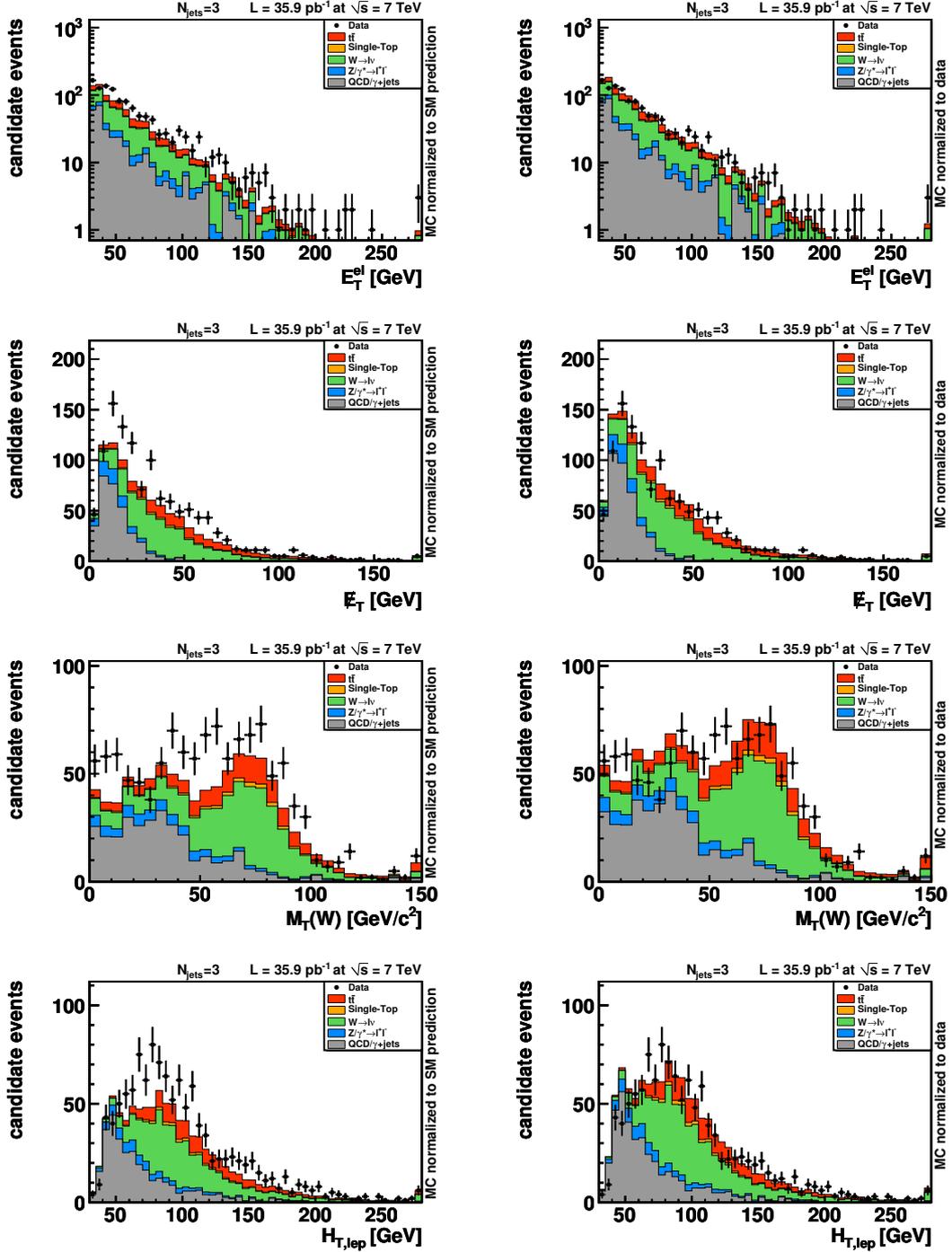


Figure 4.18: Kinematic observables for events with exactly three jets. For each variable, the MC simulation is scaled to the expectation for an integrated luminosity of 35.9 pb^{-1} (left hand side) and scaled to the number of events observed in data (right hand side).

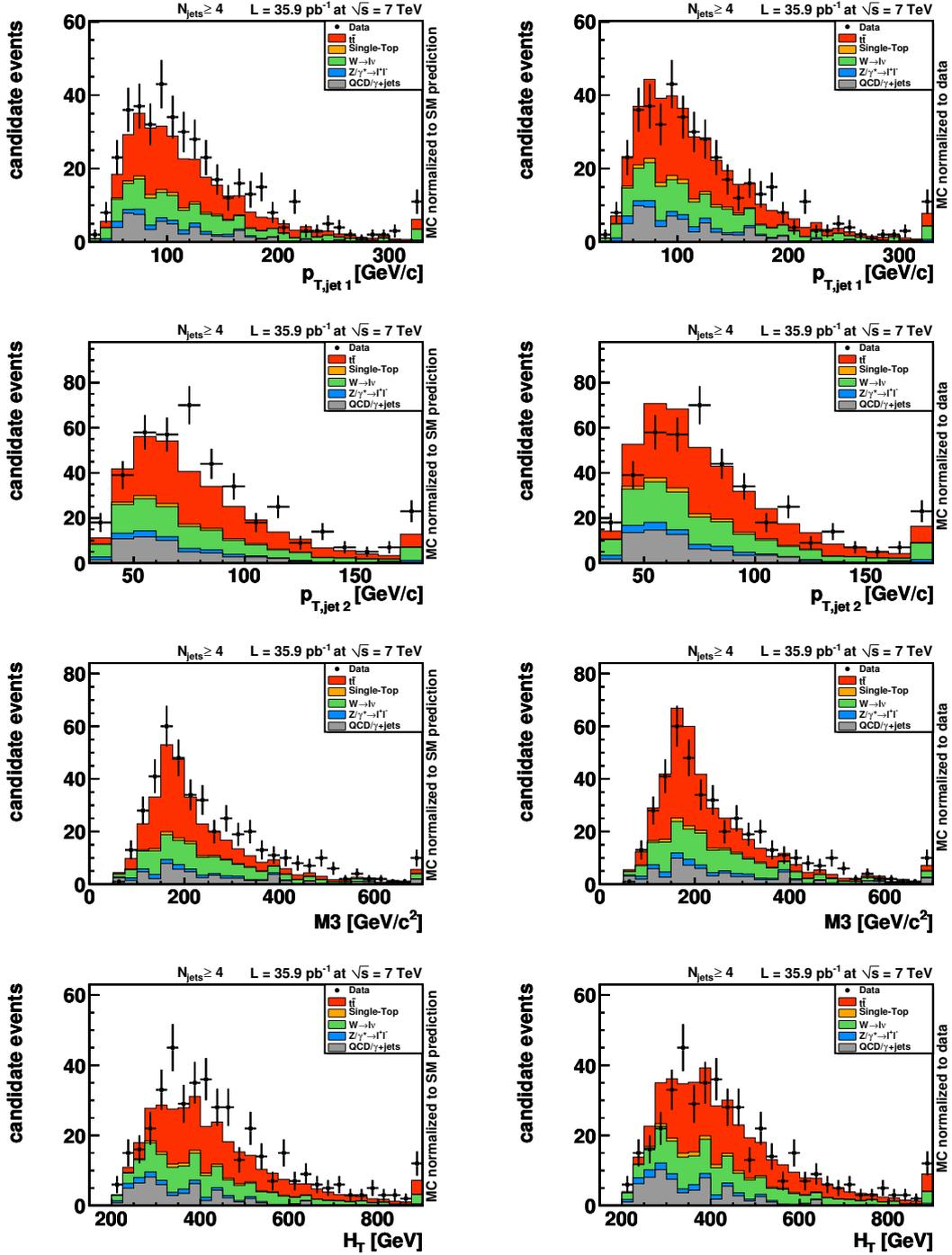


Figure 4.19: Kinematic observables for events with at least four jets. For each variable, the MC simulation is scaled to the expectation for an integrated luminosity of 35.9 pb^{-1} (left hand side) and scaled to the number of events observed in data (right hand side). The number of simulated QCD multi-jet events passing all selection criteria is too small to obtain reliable predictions for the shape of the shown distributions.

Chapter 5

Measurement of the $t\bar{t}$ Cross-Section

This thesis describes the measurement of the top-quark pair production cross-section at a center-of-mass energy of 7 TeV, using first LHC data recorded with the CMS experiment in the year 2010. For this purpose, the number of observed $t\bar{t}$ events is derived in a dataset which contains events exhibiting an electron+jets signature. The selection of these events has been discussed in the previous chapter, while the methods used to distinguish $t\bar{t}$ signal from the remaining background events will be explained in the following sections. After the number of observed $t\bar{t}$ events, $N_{t\bar{t}}^{\text{obs}}$, has been derived, the top-quark pair production cross-section can be calculated using

$$\frac{N_{t\bar{t}}^{\text{obs}}}{N_{t\bar{t}}^{\text{pred}}} = \frac{\sigma_{t\bar{t}}^{\text{obs}}}{\sigma_{t\bar{t}}^{\text{pred}}}. \quad (5.1)$$

In this equation $N_{t\bar{t}}^{\text{pred}}$ denotes the number of predicted $t\bar{t}$ events derived from Monte Carlo simulations corrected for differences in the selection efficiencies in simulations and real collision data. Since all effects due to limited detector acceptances and finite selection efficiencies are included in this prediction, the ratio of the number of observed and the number of predicted events is equal to the ratio of the measured ($\sigma_{t\bar{t}}^{\text{obs}}$) and the predicted $t\bar{t}$ production cross-section ($\sigma_{t\bar{t}}^{\text{pred}}$).

The number of observed $t\bar{t}$ events is extracted by employing a fitting procedure. Two variables appropriate to discriminate signal from background events are fit to derive the relative fractions of all processes contributing to the selected dataset. A sufficient modelling of the discriminating variables is an important prerequisite for this approach. However, the simulation of the QCD multi-jet background is challenging and suffers from different weaknesses. It is therefore necessary to develop an approach which can be used to obtain a model for the QCD multi-jet background using collision data directly.

Moreover, the number of expected signal and background events and the modelling of the two fit variables depend on several assumptions made during the generation of signal and background processes as well as on experimental determinations of different

quantities. The uncertainties on these assumptions and measurements are the sources of systematic uncertainties on the measured $t\bar{t}$ production cross-section and are studied in detail in this chapter.

Finally, the events used to measure the $t\bar{t}$ production cross-section in the muon+jets final state [15] are also taken into account, and a simultaneous measurement of $\sigma_{t\bar{t}}$ in both final states is performed.

5.1 Analysis Strategy

The selection of $t\bar{t}$ candidate events in the electron+jets final state has been described in section 4.2.2. Based on the observation that $t\bar{t}$ events might also exhibit only three jets passing the stringent jet selection criteria, events with at least three selected jets are used for the measurement of the $t\bar{t}$ production cross-section. As a consequence, the signal to background ratio is smaller in the selected dataset with respect to the more signal-enhanced dataset of events with at least four jets. This seems to be a drawback at first sight. However, the selected dataset can be split into two orthogonal subsets containing events with exactly three jets and at least four jets. Doing so, it can be taken advantage of the large background contamination in events with exactly three jets to obtain an estimate on the background contribution in the subset with at least four jets. For this purpose, the predicted jet multiplicity distribution, i.e the ratio of the number of events with exactly three jets to the number of events with at least four jets, is used and the amount of signal and background events is simultaneously estimated in both subsets.

In this analysis, the estimation of the relative fractions of contributing processes is done implementing a simultaneous maximum likelihood fit to the distributions of two kinematic variables which are able to discriminate the $t\bar{t}$ signal from background events. For all contributing processes the shape of these variables thus has to be well-modeled. Since the number of simulated QCD multi-jet events passing all selection requirements is too small, and the modelling of isolated leptons produced in jets is also very challenging, a model for this background has to be extracted using collision data directly.

5.1.1 Modelling of the QCD Multi-Jet Background

QCD multi-jet events typically pass the event selection when π^0 -mesons produced in jets decay into two photons which further convert into electron-positron pairs. In order to pass the event selection, electrons produced in such events have to be isolated and have to pass both conversion rejection algorithms. Although this happens only rarely, the large cross-section of QCD multi-jet events leads to a substantial amount of events passing the selection criteria. On the other hand, the rejection of QCD multi-jet events with leptonic decays of B -hadrons is very successful. Based on simulations, the fraction of these events is derived to be only about 7% of all QCD multi-jet events passing the selection criteria. Furthermore, γ +jets events enter the selected dataset also due to

conversion electrons passing the isolation requirement and both conversion rejection algorithms. Their relative contribution to the amount of selected QCD multi-jet and γ +jets events is derived from simulations to be about 17%.

Consequently, for the modelling of the QCD multi-jet background based on collision data, a sideband dominated by events containing electrons which originate from conversions has to be used. However, it is almost impossible to distinguish conversion electrons from QCD multi-jet events and γ +jets events. As a consequence, both processes are combined into a common background process and the behavior of this background is modelled from a data sideband.

For the definition of the data sideband different limitations have to be taken into account. First, the selected sideband has to be almost free of processes other than QCD multi-jet and γ +jets. Moreover, the phase space in which the sideband resides has to be close to the phase space of the selected data. This is important for ensuring that the shape of kinematic distributions for the combined QCD/ γ +jets process in the sideband is similar to the shape in the selected dataset.

Three different selection criteria are appropriate to reject QCD/ γ +jets events: the requirement on the combined relative isolation, the impact parameter with respect to the beam spot, and the WP70 for electron identification. Inverting those requirements allows for the selection of a sideband region which is enriched with QCD/ γ +jets events. The sideband is obtained by first selecting electrons with a transverse energy larger than 30 GeV, $|\eta| < 2.5$, a supercluster which does not reside in the ECal barrel-endcap transition region, and combined relative isolation smaller than 0.5. Finally, only electrons which do not fulfill at least two of the three mentioned selection criteria appropriate to reject QCD/ γ +jets events are used for the sideband. Similar to the requirements applied to select the dataset used to measure the $t\bar{t}$ production cross-section, the events in the sideband region are required to exhibit exactly one electron passing the sideband selection. Furthermore, they have to pass the muon veto, the Z boson veto, and both conversion rejection algorithms. The latter requirement might look odd at first sight because the aim is to select a sideband populated with conversion electrons. However, the focus solely lies on conversion electrons exhibiting appropriate kinematics allowing to pass the conversion rejection algorithms. Furthermore, the list of selected jets is cleaned from any selected electron using the same geometrical requirement as for the standard selection. In the end, all events with at least three jets are used for the sideband. In the following, this sideband is called the Background-Electron sideband (Bkg-Ele).

The sideband selection is applied to collision data and simulation. In the latter case, the inclusive QCD multi-jet samples have been used for the simulated dataset, since no isolation requirement has been applied during their generation. Moreover, the simulated dataset contains all the remaining relevant processes, i.e. $t\bar{t}$, Single-Top, W +jets, Z/γ^* +jets, and γ +jets events. From this simulation, the relative fraction of QCD/ γ +jets events in the Bkg-Ele is derived to be larger than 99%. Also if the sideband is split into events with exactly three jets and events with at least four jets, the purity in terms of QCD/ γ +jets in both subsets is larger than 99%. In figure 5.1

the shapes of different kinematic variables are compared for simulated QCD/ γ +jets events passing the standard selection and all simulated events passing the sideband selection (Bkg-Ele MC). This is important for ensuring that the sideband can be used to appropriately model the QCD/ γ +jets events passing the standard selection. It can be seen, that both shapes agree well within their uncertainties, however, the number of simulated events with at least four jets passing the standard selection is too small for drawing a final conclusion. Also there seems to be a slight overestimation of the transverse energy of the electron which leads also to slight differences in the $H_{T,\text{lep}}$ variable. This difference might arise from the different requirement on the relative isolation variable which is highly correlated with the electron's E_T . Beside the shapes for the two simulated datasets, additionally the shape for the sideband selected in collision data (Bkg-Ele data) is shown for each variable.

In summary, the selected sideband seems to be able to suitably model the shape of QCD/ γ +jets events for many, but not all kinematic variables. In the following, the shape of QCD/ γ +jets events for a particular variable is modelled using the shape extracted from the data sideband.

5.1.2 Discriminating Variables

In order to measure the $t\bar{t}$ production cross-section, the number of observed $t\bar{t}$ events in the selected dataset has to be derived and compared with the prediction obtained from MC simulations that are based on theoretical calculations. In section 4.4 a disagreement between the number of predicted events and the number of observed events in the selected dataset has been found. However, the simulation seems to sufficiently predict the shape of different kinematic distributions. Possible reasons for this slight mis-modelling have already been discussed and will be revisited when systematic uncertainties on the measurement are discussed. Because of the described disagreement between the number of predicted and observed events, the number of $t\bar{t}$ events in the selected dataset cannot be derived by simply subtracting the number of predicted background events from the number of observed events. Instead, the relative fraction of each process in the selected dataset has to be extracted.

One frequently used method for deriving the relative contributions of different processes to a particular dataset is the template fit procedure. A variable that exhibit a different shape for the contributing processes is an important prerequisite for this approach. Moreover, to obtain precise results, the number of events in the selected dataset has to be sufficiently large to minimize the effects of statistical fluctuations. In this analysis, the selected dataset is divided into two subsets, a set containing events with exactly three jets which is dominantly populated by background events, and a set containing events with at least four jets where the predicted number of signal and background events is almost equal. Now, the basic idea is to use the predicted ratio of events with exactly three jets and at least four jets, and derive the relative contributions of background events mainly from the background dominated subset, and the relative contribution of signal events from the more background free subset. For

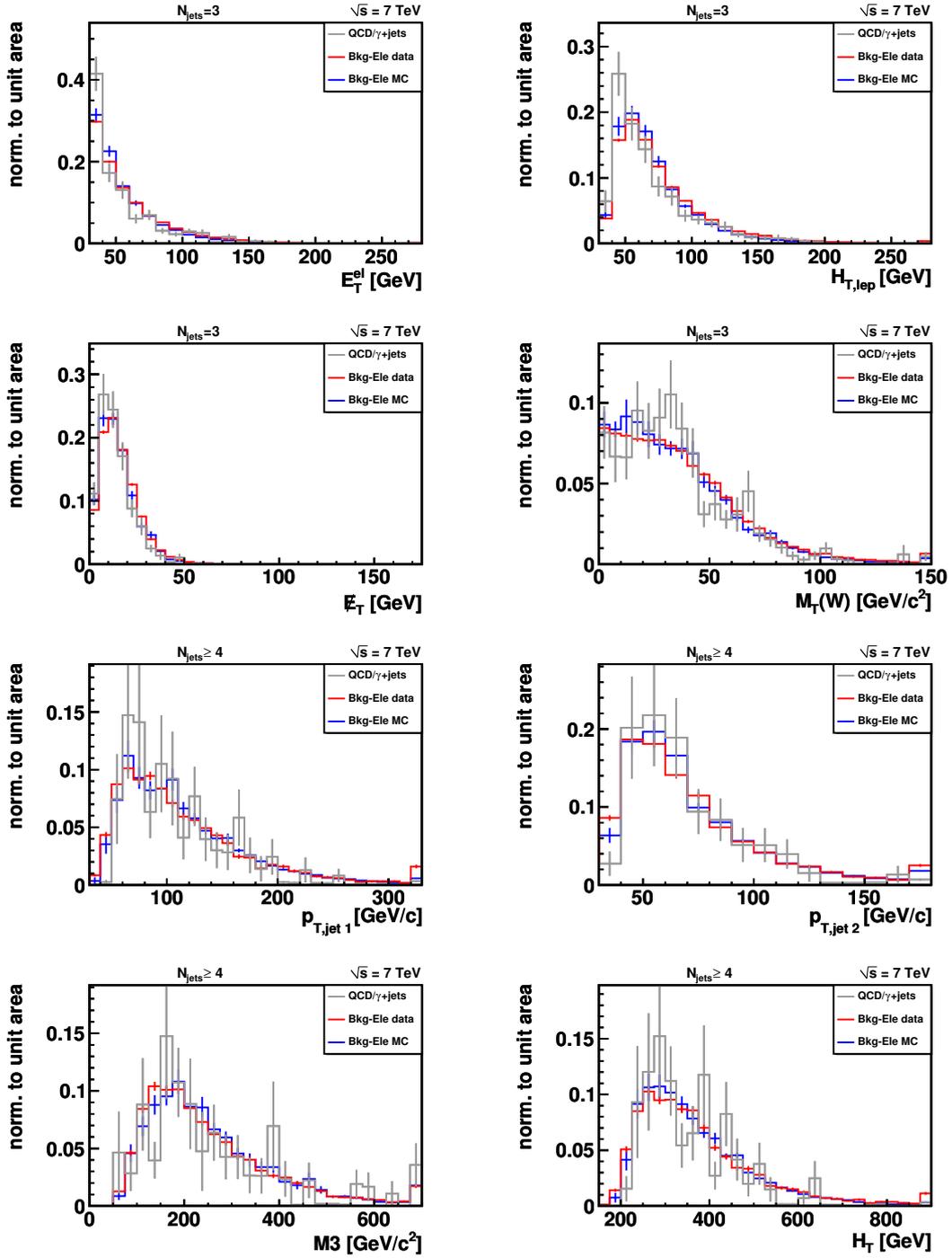


Figure 5.1: Comparison of the shape of different kinematic variables for QCD/ γ +jets events in the selected dataset with events passing the sideband selection in simulation and real collision data. In most distributions a good agreement is found between QCD/ γ +jets events in the selected dataset and the simulated events passing the sideband selection which indicates similar kinematics in both phase space regions. The upper four illustrations include only events with exactly three jets while the lower four illustrations show solely events with at least four selected jets.

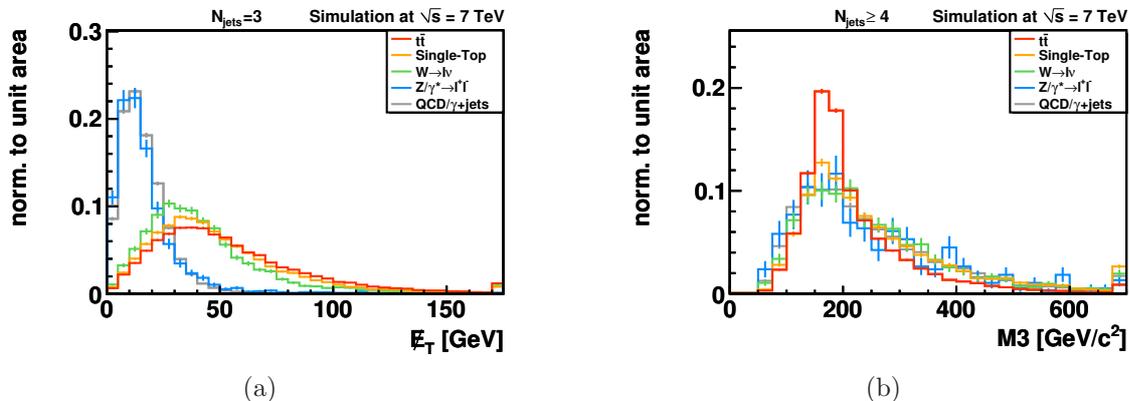


Figure 5.2: Distributions of the two discriminating variables used to distinguish between signal and background processes. The missing transverse energy for events with exactly three jets (a) is used to separate between processes with artificial and true \cancel{E}_T . A separation between $t\bar{t}$ and all background events is achieved using the M3 variable (b).

this purpose, two discriminating variables are fitted simultaneously. For events with exactly three jets the missing transverse energy distribution shown in figure 5.2(a) is employed. It can be seen, that this variable provides a good separation between processes with artificial \cancel{E}_T like QCD multi-jet, γ +jets and Z/γ^* +jets events and processes with true \cancel{E}_T originating from an undetected neutrino like $t\bar{t}$, Single-Top, and W +jets events. For events with at least four jets, the M3 variable, which is a simple reconstruction of the mass of the hadronically decaying top quark, is used to distinguish between top-quark pairs and background processes. While the background processes show a broad M3 distribution, the $t\bar{t}$ signal has a narrow peak around the top-quark mass. The M3 variable is depicted in figure 5.2(b). It is visible, that both variables provide only a small separation between QCD/ γ +jets and Z/γ^* +jets events, as well as Single-Top and W +jets events. This insight has to be taken into account in the implementation of the fitting procedure described in the following section.

5.2 Statistical Method

For the measurement of the relative contributions of all relevant processes to the selected dataset, a simultaneous binned likelihood template fit to the missing transverse energy distribution (events with exactly three jets) and the M3 distribution (events with at least four jets) is used. Within the template fit procedure, a dedicated likelihood function which, in principle, contains the number of events of each process as a parameter is maximized with respect to these parameters. In this way, the number of signal and background events in the selected dataset is derived and the $t\bar{t}$ production cross-section is measured. For the actual measurement and the derivation of the corresponding uncertainties, a Neyman construction [176] of central intervals using the $t\bar{t}$ production cross-section as test statistic is employed. The Neyman construction is

based on ensemble tests consisting of a dedicated number of pseudo-experiments. For each of these pseudo-experiments a possible \mathbb{E}_T and M3 distribution is derived from the expected signal and background contributions through the application of Monte Carlo techniques. The technical implementation of the fitting procedure and the ensemble tests is done using the theta framework [177].

5.2.1 Binned Likelihood Template Fit

The binned normalized distributions of \mathbb{E}_T and M3 for each process, commonly called templates, the expected number of events of each process, and the \mathbb{E}_T and M3 distributions observed in collision data are the ingredients for the likelihood fit employed in this analysis. Using a likelihood function containing the actual number of events of each contributing processes as parameter, the expected and observed number of events in each bin are compared and the combination with the largest value of the likelihood function is chosen as fit result.

The processes used to model the observed \mathbb{E}_T and M3 distributions are $t\bar{t}$, Single-Top (combined t-channel and tW -channel), W +jets, and Z/γ^* +jets derived from simulations, as well as the QCD/ γ +jets model derived from data. In the following, these processes are represented by an index $k \in \{t\bar{t}, t, W, Z, \text{QCD}\}$, the fitted variable is given by the index $j \in \{\mathbb{E}_T, \text{M3}\}$, and a particular bin is represented by the index i . Thus the expected number of events μ in a particular bin i of the fit variable j can be expressed by:

$$\mu_{ij} = \sum_k \beta_k \cdot \alpha_{ijk}. \quad (5.2)$$

In this equation, α_{ijk} denotes the fraction of events in bin i of the template α_{jk} for process k and the fit variable j . The α_{jk} are normalized to the predicted number of events of process k , $\hat{\nu}_{jk}$, in the datasets containing events with exactly three jets ($j = \mathbb{E}_T$) and with at least four jets ($j = \text{M3}$). The parameter β_k in equation 5.2 is defined via

$$\beta_k = \frac{\sum_j \nu_{jk}}{\sum_j \hat{\nu}_{jk}} = \frac{\sigma_k}{\hat{\sigma}_k} \quad (5.3)$$

and thus represents the ratio of the observed number of events, $\sum_j \nu_{jk}$, and the predicted number of events, $\sum_j \hat{\nu}_{jk}$, for a particular process k . Since all effects from limited detector acceptances and finite selection and reconstruction efficiencies are already taken into account in the prediction, β_k also represents the ratio of the measured cross-section (σ_k) and the predicted cross-section ($\hat{\sigma}_k$) of a specific process k . To prevent unphysical results, any β_k -parameter is strictly required to be non-negative during the fitting procedure.

The μ_{ij} are assumed to follow a Poisson distribution. Hence, the employed likelihood function can be written as

$$L(\vec{\beta}) = \prod_j \prod_{i=1}^{N_{\text{bins}}^j} \frac{\mu_{ij}^{n_{ij}} \cdot e^{-\mu_{ij}}}{n_{ij}!}, \quad (5.4)$$

where n_{ij} represents the number of observed data events in bin i of the fit variable j and N_{bins}^j denotes the number of bins for the variable j and is equal to 35 for \cancel{E}_T and 28 for M3. In this notation $\vec{\beta}$ is an abbreviation for the list of all included β_k -parameters. However, this definition of the likelihood function might lead to numerical instabilities in the fitting process because of the mentioned similarities between QCD/ γ +jets and Z/γ^* +jets events, as well as Single-Top and W +jets events. To avoid these kinds of instabilities, two further terms are added to the likelihood function which implement prior knowledge on β_t and the ratio of β_W and β_Z . The intension of these terms is to inhibit that the mentioned fit parameters (or their ratio) move too far away from their expected values of 1.0. Such an limitation can be achieved using Gaussian factors:

$$G(\beta_t; \Delta_t) = \frac{1}{\sqrt{2\pi\Delta_t^2}} \cdot e^{-\frac{(\beta_t-1)^2}{2\Delta_t^2}}$$

$$G(\beta_W, \beta_Z; \Delta_{W/Z}) = \frac{1}{\sqrt{2\pi\Delta_{W/Z}^2}} \cdot e^{-\frac{(\beta_W/\beta_Z-1)^2}{2\Delta_{W/Z}^2}}. \quad (5.5)$$

Both Δ -parameters represent the width of the corresponding Gaussian distribution and the larger the difference between a β -parameter and its expectation the smaller is the value of the Gaussian function which leads to a reduced likelihood for a particular hypothesis. The width of both Gaussian distributions is chosen to be 30%. These assumptions are reasonable, since the production of single top quarks is theoretically well understood and the measured uncertainty on the t-channel cross-section is roughly 30% [178]. Moreover, the ratio of the inclusive production cross-sections for W bosons and Z bosons is also theoretically well known and precisely measured [179], however, since the uncertainty on the measured $\sigma(W + \geq 3\text{jets})/\sigma(Z + \geq 3\text{jets})$ ratio is roughly 30% as can be derived using the results from [179] and [180], $\Delta_{W/Z}$ is set to 30%. The finally employed likelihood function is thus given by:

$$L(\vec{\beta}) = \prod_j \prod_{i=1}^{N_{\text{bins}}^j} \frac{\mu_{ij}^{n_{ij}} \cdot e^{-\mu_{ij}}}{n_{ij}!} \cdot G(\beta_t; \Delta_t) \cdot G(\beta_W, \beta_Z; \Delta_{W/Z}). \quad (5.6)$$

5.2.2 Ensemble Tests

In order to test the validity of the explained statistical method and to derive the expected uncertainties of the measurement, the approach of ensemble testing is used. Each ensemble consists of a set of pseudo-experiments which are based on the same prior knowledge. Monte Carlo techniques are used to draw potential \cancel{E}_T and M3 distributions according to the expectations. These distributions are then evaluated using the likelihood fitting procedure explained above, leading to results for each β_k -parameter. By evaluating a large number of pseudo-experiments the expected distribution of the β_k -parameters can be obtained and, for example, their mean and width can be derived.

The elementary step for the construction of an ensemble is the definition of the corresponding model which is common to all pseudo-experiments belonging to the same

ensemble. Assuming, for example, that collision data behaves exactly like predicted from simulation, the chosen model would be that all β_k -parameter are set to 1.0 and that the \cancel{E}_T and M3 distributions for the different processes are exactly the α_{jk} derived from simulation or the data-driven model, respectively. The model used in this thesis, however, additionally takes into account an uncertainty on the predicted number of events for each background process. Thus, for each pseudo-experiment the number of predicted events for a particular background process k is modified by multiplying $\sum_j \hat{\nu}_{jk} \equiv \hat{\nu}_k$ with $\delta\hat{\nu}_k$ which is drawn from a Gaussian distribution, which is centered at 1.0 and has a width of $\Delta\hat{\nu}_k$. Except for the QCD/ γ +jets background, where the MC prediction is less trustworthy and consequently $\Delta\hat{\nu}_k = 50\%$ is used, the uncertainty on the predicted number of background events is chosen to be 30%. The actual number of events of process k used in a particular pseudo-experiment is then drawn from a Poisson distribution with mean $\hat{\nu}_k \cdot \delta\hat{\nu}_k$. According to the predicted ratio of events with exactly three jets and at least four jets for each process, the number of events contributing to the different phase space regions are calculated. This number of contributing events for each process is then drawn from the corresponding \cancel{E}_T and M3 template and thus the pseudo-data distribution for both variables is obtained. Applying the likelihood fit to the pseudo-data distribution of each pseudo-experiment belonging to a given ensemble, the distribution of the β_k -parameters and their average values for a particular ensemble are derived.

5.2.3 Neyman Construction

For the measurement of the top-quark pair production cross-section and the derivation of the corresponding uncertainties, a Neyman construction for central intervals is employed. The Neyman construction is a frequentist method which allows to derive an interval for a particular confidence level with the correct coverage. To build the Neyman construction, ensemble tests assuming different $t\bar{t}$ production cross-sections are performed and $\beta_{t\bar{t}}^{\text{fit}}$, obtained from the maximization of the likelihood function, is used as test statistic. In summary, 16 different ensemble tests, each consisting of 50,000 pseudo-experiments are carried out. For the different ensembles, $\beta_{t\bar{t}}^{\text{in}}$ which represents the ratio of the $t\bar{t}$ production cross-section used to produce the pseudo-data distributions and the SM predicted cross-section, is varied between 0.0 and 3.0 in steps of 0.2. In contrast, the β_k -parameters of all background processes are always kept at 1.0 for each ensemble. From the resulting $\beta_{t\bar{t}}^{\text{fit}}$ distribution, for each ensemble the median is used as an estimate for the central $\beta_{t\bar{t}}^{\text{fit}}$ value and the 2.5%, 16%, 84% and 97.5% quantiles are employed to construct central 68% and 95% confidence belts.

In the following it is explained, how the Neyman construction is obtained and used to derive the expected statistical uncertainty on the $t\bar{t}$ production cross-section measurement in the electron+jets final state. In figure 5.3(a) the $\beta_{t\bar{t}}^{\text{fit}}$ distribution for the ensemble with $\beta_{t\bar{t}}^{\text{in}} = 1.0$ is shown. In addition, the 68% and 95% central confidence intervals are depicted. As expected, the $\beta_{t\bar{t}}^{\text{fit}}$ distribution has a Gaussian shape and is centered at 1.0 which is equal to the employed $\beta_{t\bar{t}}^{\text{in}}$. To derive the Neyman construction

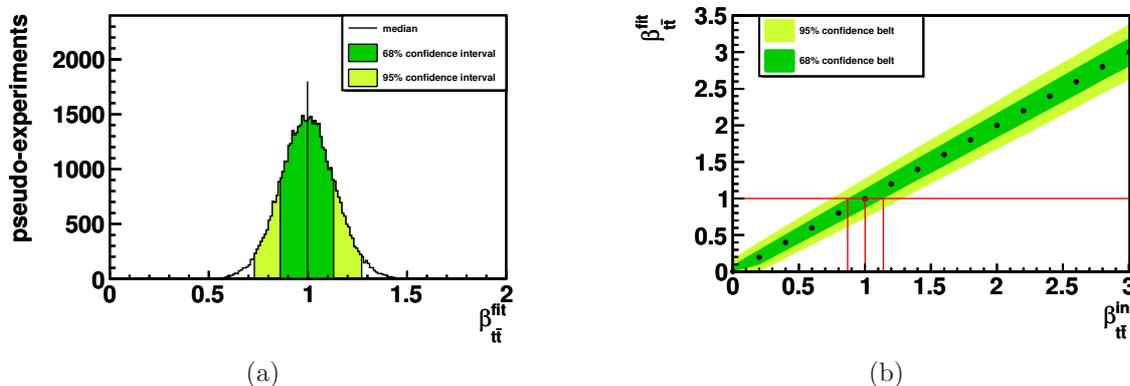


Figure 5.3: In (a) the $\beta_{t\bar{t}}^{\text{fit}}$ distribution for the ensemble with $\beta_{t\bar{t}}^{\text{in}} = 1.0$ is shown. Its median value is in very good agreement with the expected value of $\beta_{t\bar{t}}^{\text{in}}$ and its 2.5%, 16%, 84% and 97.5% quantiles are used to derive the central 68% and 95% confidence intervals. The Neyman construction used to derive the expected statistical uncertainty on the $t\bar{t}$ production cross-section is shown in (b). Drawing a horizontal line at $\beta_{t\bar{t}}^{\text{fit}} = 1.0$ and reading off the $\beta_{t\bar{t}}^{\text{in}}$ values at the intersections of this line and the 68% confidence belt, a statistical uncertainty of -13.1% and +14.0% is derived.

illustrated in figure 5.3(b), the median and the 68% and 95% confidence intervals are drawn as a function of $\beta_{t\bar{t}}^{\text{in}}$. For this purpose the $\beta_{t\bar{t}}^{\text{fit}}$ distributions for the employed $\beta_{t\bar{t}}^{\text{in}}$ values are exploited. The 68% and 95% confidence belts are obtained by connecting the values for the discrete $\beta_{t\bar{t}}^{\text{in}}$ points with a linear interpolation. This is reasonable, since the method features a linear behavior, i.e. the median of each $\beta_{t\bar{t}}^{\text{fit}}$ distribution is equal to the used $\beta_{t\bar{t}}^{\text{in}}$ value. Assuming a measured $\beta_{t\bar{t}}^{\text{fit}}$ value of 1.0, the expected statistical uncertainty (corresponding to one standard deviation) on the $t\bar{t}$ production cross-section measurement is derived by drawing a horizontal line at $\beta_{t\bar{t}}^{\text{fit}} = 1.0$ and evaluating the $\beta_{t\bar{t}}^{\text{in}}$ values at its intersections with the boundaries of the 68% confidence belt. In this way, the expected statistical uncertainty on the $t\bar{t}$ production cross-section using 35.9 pb^{-1} is derived to be -13.1% and +14.0%.

5.3 Measurement in the Electron+Jets Final State

In the following, the results of the first measurement of the top-quark pair production cross-section in the electron+jets final state using collision data collected with the CMS experiment are discussed. Besides the pure statistical uncertainty which has been derived above, also the effects of different sources of systematic uncertainties have to be taken into account. For this purpose, ensemble tests including systematic effects in the pseudo-data distributions are employed. The final cross-section measurement is done by applying the likelihood fitting procedure to the observed data distributions. The resulting $\beta_{t\bar{t}}^{\text{fit}}$ value as well as the Neyman construction built from ensembles taking into account all sources of systematic uncertainties are used to derive the $t\bar{t}$ production cross-section and the uncertainties on the measurement.

5.3.1 Sources of Systematic Uncertainties

Within the process of generating simulated signal and background events, different assumptions have been made that might have an influence on the obtained predictions. Two prominent examples are the choice of the factorization and renormalization scale as well as the employed matching threshold. Moreover, the predictions depend on experimental measurements of different quantities, such as jet energy correction factors or the jet energy resolution. To account for the uncertainties on the different assumptions or measurements, each particular quantity is varied within its uncertainty and the influence on the measured $t\bar{t}$ production cross-section is evaluated. In some cases, this is done by analyzing special simulated samples where one quantity is varied with respect to its nominal value, in other cases, the usual simulated datasets are used and specific kinematic values are varied in an appropriate way. The uncertainties on the measured $t\bar{t}$ production cross-section arising due to the mentioned variations are commonly referred to as systematic uncertainties. In section A.4 of the appendix, the details on the used special MC samples are given.

In general, the incorporation of a particular systematic uncertainty affects both, the number of predicted events for a given process as well as the shape of the templates for the \cancel{E}_T and M3 distributions. Usually, upwards and downwards fluctuations of the studied source of a systematic uncertainty are evaluated and treated as ± 1 standard deviation, or $\pm 1\sigma$. Consequently, there are three different templates, the nominal template α_{jk} and two systematically varied templates Ξ_{ujk}^{\pm} for each source of a systematic uncertainty u which might differ from the nominal templates in their normalization as well as their shape.

Variation of the Jet Energy Resolution

As discussed briefly in section 3.3.5, the jet energy resolution measured from data [158] is roughly 10% worse compared to the prediction obtained from simulations. In section 4.2.2 it has been explained how the simulated events are corrected for this effect. However, the described measurement also has an uncertainty of roughly 10%. To estimate the influence of this uncertainty on the measurement of the $t\bar{t}$ production cross-section, each previously modified detector jet four-vector is again multiplied with the factor defined in equation 4.5 to account for an upwards fluctuation. Obviously the factor of 1.1 in equation 4.5 is replaced by 0.9 to account for a downwards fluctuation. These variations are consistently propagated to the missing transverse energy of an event by employing the method described in section 4.2.2. The relative changes in the expected event yield of each process (except QCD/ γ +jets which is modeled using collision data) are shown in table 5.1. Variations in the shapes of the employed templates are illustrated in figure 5.4, where the nominal templates of all contributing processes are compared with the corresponding Ξ_{jk}^{\pm} templates.

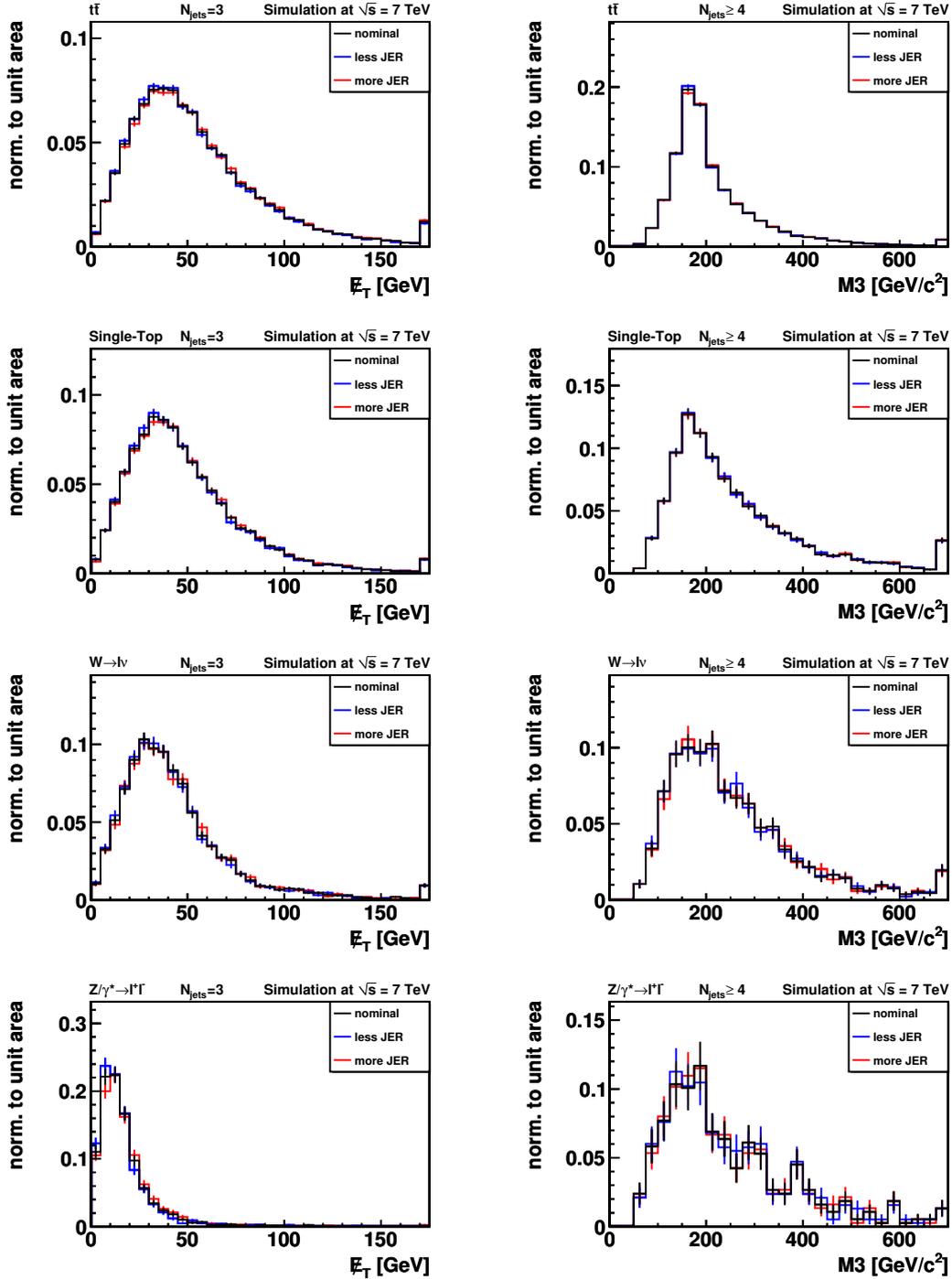


Figure 5.4: Comparison of the nominal templates with the templates obtained after JER variations have been applied. Illustrations for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes are shown for the two fit variables E_T and $M3$.

	Variation	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets
= 3 jets	less JER	+0.4%	+0.5%	-0.5%	+0.2%
= 3 jets	more JER	-0.4%	-0.6%	+0.5%	+1.1%
≥ 4 jets	less JER	+0.9%	+0.7%	-0.5%	+1.3%
≥ 4 jets	more JER	-0.9%	-0.7%	+0.2%	-0.8%

Table 5.1: Relative changes of the expected number of events for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes due to variations of the jet energy resolution.

Variation of the Jet Energy Corrections

In order to estimate the effect of a variation of the applied jet energy corrections, frequently called the jet energy scale (JES) uncertainty, the jet energy correction factors are varied within their uncertainties. First, for each jet the relative uncertainty ($\pm 1\sigma$) on the applied p_T - and η -dependent jet energy correction factor is derived. An additional p_T - and η -dependent uncertainty is added in quadrature to this uncertainty leading to a combined relative uncertainty of U_{JES} . The added extra uncertainty results from the quadratic sum of an 1.5% uncertainty accounting for calibration changes between the dataset used to measure the jet energy correction and the dataset used in this analysis, a p_T -dependent uncertainty accounting for the effects of pile-up events, as well as an p_T - and η -dependent uncertainty accounting for the different energy corrections for b -quark-induced jets. The relative uncertainty due to pile-up events is derived by

$$\frac{\bar{E}_{\text{pu}} \cdot A_{\text{jet}} \cdot \bar{N}_{\text{pu}}}{p_T}, \quad (5.7)$$

where $\bar{E}_{\text{pu}} = 0.75$ GeV is the average energy of a pile-up event, $A_{\text{jet}} = 0.8$ is the area covered by a jet in the η - ϕ -plane, and $\bar{N}_{\text{pu}} = 2.2$ is the average number of pile-up events per bunch crossing in the analyzed dataset. Since the majority of the jets used to measure the jet energy correction originates from gluons, also an uncertainty on the correction of b quark induced jets is applied. This uncertainty is only added if a jet is identified to originate from a b quark by matching the detector jet to the generated bottom quark applying similar requirements as have been used above to match detector jets to particle jets. For jets with a transverse momentum between 50 and 200 GeV/c and $|\eta| < 2.0$ the incorporated uncertainty is 2%, while it is 3% for all remaining jets.

The four-vector of each jet is finally scaled with $(1 \pm U_{\text{JES}})$ and the event selection procedure is repeated. The varied jet energies are also propagated to the missing transverse energy of an event employing a similar approach as has been used during the correction of the JER in simulated datasets, described in section 4.2.2. The relative changes in the expected event yields of all affected processes is shown in table 5.2. A comparison of the nominal template with the systematically altered templates for both fit variables and all affected processes is depicted in figure 5.5.

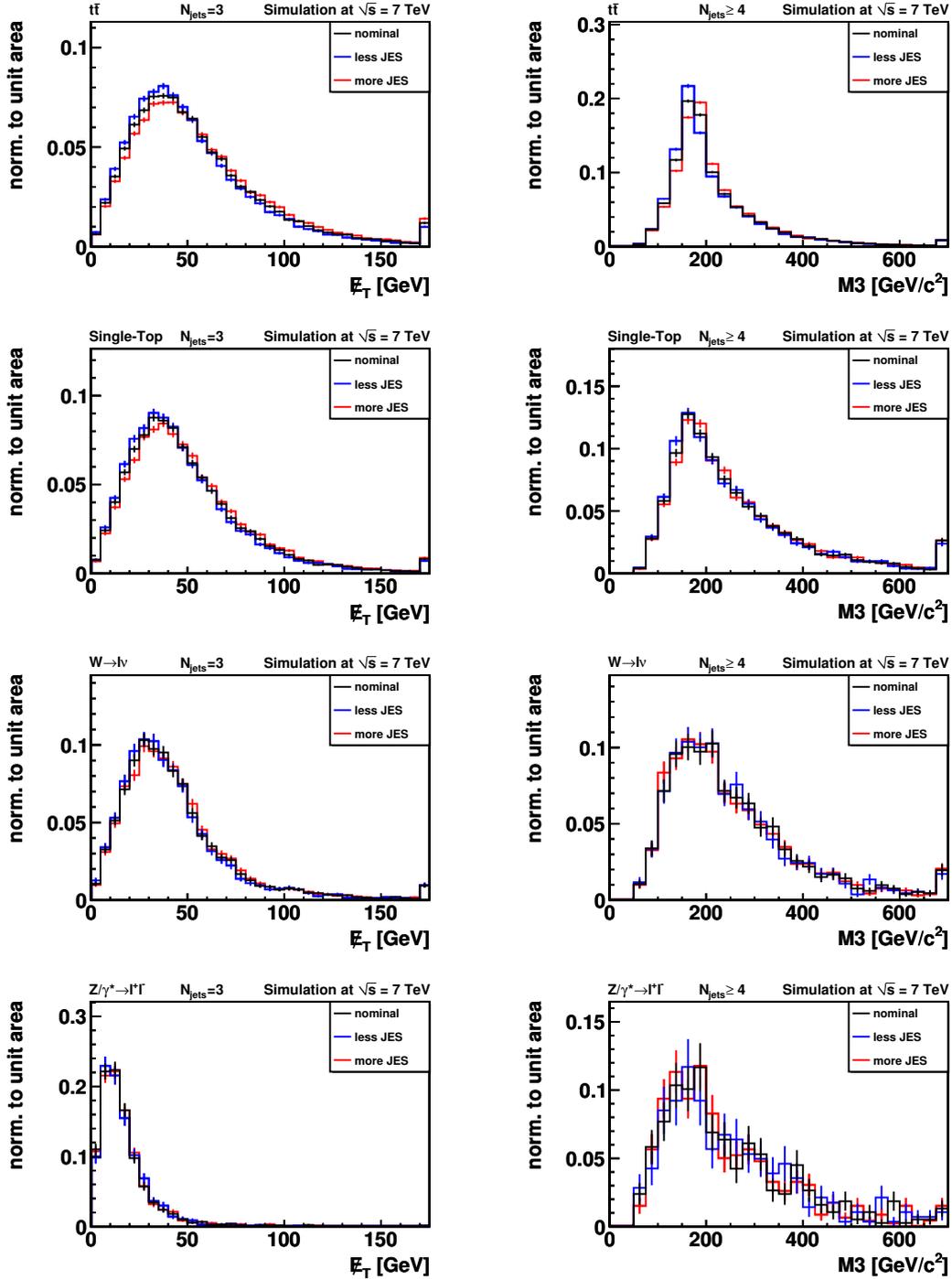


Figure 5.5: Comparison of the nominal templates with the templates obtained after JES variations have been applied. Illustrations for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes are shown for the two fit variables E_T and $M3$.

	Variation	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets
= 3 jets	less JES	+1.0%	-7.5%	-14.3%	-23.3%
= 3 jets	more JES	-1.9%	+6.5%	+16.8%	+20.2%
≥ 4 jets	less JES	-10.2%	-14.2%	-16.5%	-25.2%
≥ 4 jets	more JES	+10.0%	+15.3%	+21.7%	+21.8%

Table 5.2: Relative changes of the expected number of events for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes due to variations of the jet energy correction factors (JES uncertainty).

	Variation	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets
= 3 jets	Electron Energy Scale down	-0.5%	-0.4%	-0.9%	-0.2%
= 3 jets	Electron Energy Scale up	+0.4%	+0.4%	+0.4%	+0.4%
≥ 4 jets	Electron Energy Scale down	-0.5%	-0.4%	-0.4%	+0.5%
≥ 4 jets	Electron Energy Scale up	+0.4%	+0.4%	+0.7%	+0.0%

Table 5.3: Relative changes of the expected number of events for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes due to variations of the electron energy scale.

Variation of the Electron Energy Scale

To account for the uncertainty on the measurement of the electron's transverse energy, the four-vector of each electron is scaled by an appropriate factor. The uncertainty on E_T of electrons reconstructed in the ECal barrel is $\pm 0.5\%$, while it is $\pm 2.5\%$ for electrons reconstructed in one of the ECal endcaps. The four-vector of barrel electrons is thus multiplied by a factor of 1.005 or 0.995 to account for upwards or downwards fluctuations in the energy measurement, respectively. In a similar way, the four-vector of each endcap electron is multiplied by a factor of 1.025 or 0.975, respectively. The varied electron energies are propagated to the missing transverse energy of an event by adding the nominal four-vector and subtracting the varied four-vector of each electron. Afterwards, the event selection procedure is repeated. The relative changes in the expected number of events for all affected processes is shown in table 5.3, while a comparison of the nominal and the varied templates for \cancel{E}_T and M3 is depicted in figure 5.6. As can be seen, the templates for the different processes are almost unaffected by the variation of the electron energy scale.

Variation of the Unclustered Energy

So far, the contributions to the missing transverse energy of all energy deposits associated with high-level physics objects like jets and electrons have been varied. In order to account for the uncertainty on the measurement of the remaining, so-called unclustered energy, this energy is varied by $\pm 10\%$. For this purpose, the transverse energy components of all raw jets and electrons are added to \cancel{E}_T , the resulting energy is scaled by a factor of 0.9 or 1.1, respectively, and the transverse energies of the men-

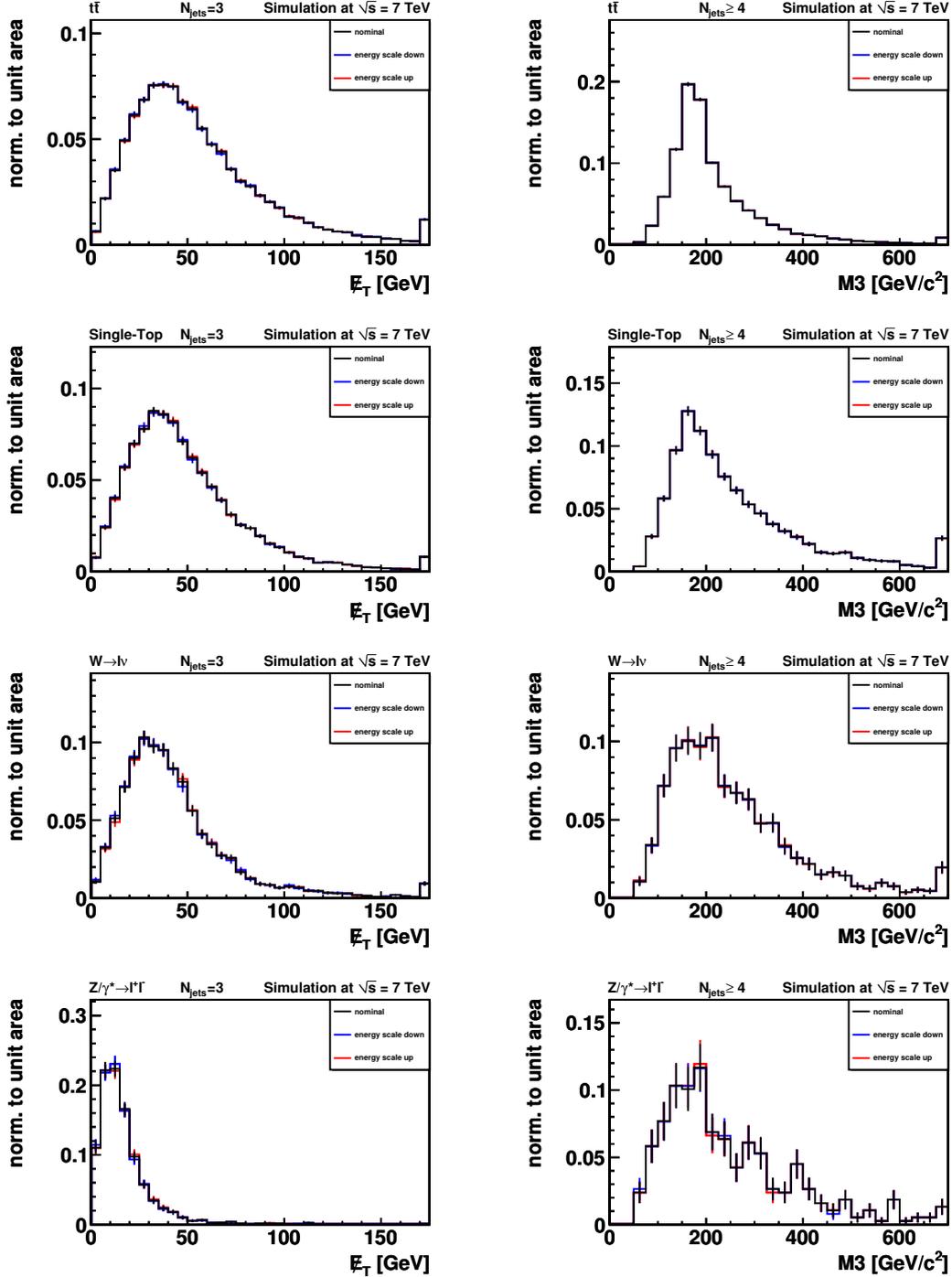


Figure 5.6: Comparison of the nominal templates with the templates obtained after electron energy scale variations have been applied. Illustrations for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes are shown for the two fit variables E_T and $M3$.

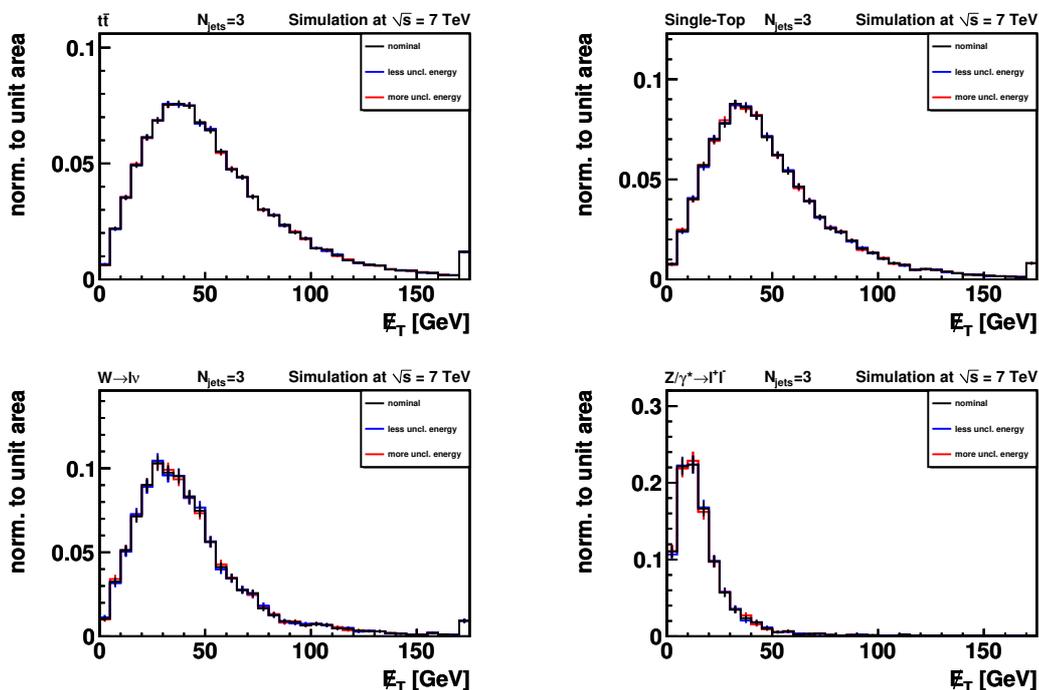


Figure 5.7: Comparison of the nominal templates with the templates obtained after a variation of the unclustered energy has been applied. Illustrations of the \cancel{E}_T distribution for $t\bar{t}$, Single-Top, W +jets and Z/γ^* +jets processes are shown.

tioned objects are removed again. A variation of the unclustered energy only effects the shape of the templates describing the \cancel{E}_T distribution. Comparisons between the nominal and the systematically varied templates are shown in figure 5.7.

Variation of the Electron Reconstruction and Trigger Efficiency

The data-to-MC correction factor accounting for differences in the electron selection efficiencies as well as the finite electron trigger efficiency has been derived in section 4.3. Consequently, all simulated events passing the electron selection requirements have been scaled with $0.95 \cdot 0.982 = 0.933$ to obtain reliable predictions for the number of selected events. The uncertainties on the two correction factors are added in quadrature leading to a combined uncertainty of ± 0.025 . Therefore, all simulated events are scaled by 0.958 or 0.908 instead of 0.933 to account for upwards or downwards fluctuations, respectively. Obviously, the predicted number of selected events changes according to the difference between the nominal and the systematically varied correction factors. Furthermore, the shape of the templates used to model the \cancel{E}_T and M3 distributions are not effected by this re-weighting.

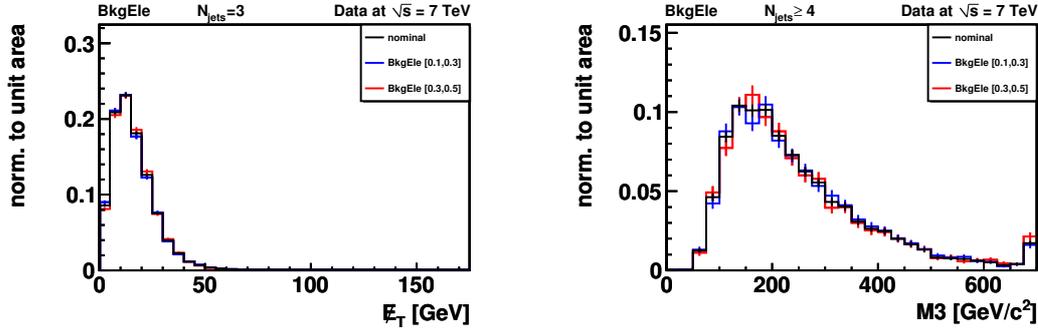


Figure 5.8: Comparison of \cancel{E}_T and M3 distributions for the nominal sideband used to model QCD/ γ +jets with two subsets defined via the relative combined isolation of the selected electron: $0.1 \leq I_{\text{rel}} < 0.3$ and $0.3 \leq I_{\text{rel}} \leq 0.5$.

Uncertainty due to the QCD/ γ +jets Model

The definition of the sideband region used to model the combined QCD multi-jet and γ +jets background process using collision data directly, has been described in section 5.1.1. Uncertainties on the obtained templates might arise from different kinematics in the selected dataset and the sideband region. It has been shown, that possible differences in the shape of kinematic variables arise from their correlation to the combined relative isolation variable. To study the influence of this variable on the modelling of the \cancel{E}_T and M3 variable, the sideband region with electrons fulfilling $0.1 \leq I_{\text{rel}} \leq 0.5$ is split into two disjoint subsets with $0.1 \leq I_{\text{rel}} < 0.3$ and $0.3 \leq I_{\text{rel}} \leq 0.5$. The subset with the lower I_{rel} values is considered as the 1σ downwards fluctuation, while the subset with larger I_{rel} values is used as the upwards fluctuation. A comparison of \cancel{E}_T and M3 for the three different sideband regions is shown in figure 5.8.

For the modelling of the ratio of the number of QCD/ γ +jets events with exactly three jets and at least four jets, purely the information from MC simulations is used. Comparing this ratio for simulated events of the selected dataset and simulated events in the nominal sideband region, it can be seen that both ratios are close to each other, but not the same. Hence, the ratio of the number of events with exactly three jets and at least four jets observed in the sideband obtained from data, is not used within the standard likelihood fit, but can be used to estimate the systematic uncertainty due to the assumed ratio. In simulated events this ratio is 5.9 ± 1.3 , while in the sideband selected in data, a slightly different value of 4.4 ± 0.1 is found. In order to estimate the systematic effect, the predicted number of QCD/ γ +jets with at least three jets is distributed over the different fit variables according to the ratio obtained from the data sideband. This variation is considered as a 1σ downwards fluctuation (it also could have been considered as an upwards fluctuation without any changes) and no upwards variation is taken into account.

	Variation	$t\bar{t}$	W +jets	Z/γ^* +jets
= 3 jets	scale \times 0.5	-3.5%	+97.8%	+76.7%
= 3 jets	scale \times 2.0	+3.6%	-36.2%	-25.7%
\geq 4 jets	scale \times 0.5	+8.3%	+84.8%	+92.2%
\geq 4 jets	scale \times 2.0	-5.4%	-41.9%	-32.5%

Table 5.4: Relative changes of the expected number of events for $t\bar{t}$, W +jets, and Z/γ^* +jets processes due to variations of the chosen factorization and renormalization scale during the event generation. The quoted values are obtained from dedicated simulated events where twice the nominal scale and half the original scale have been chosen.

Variation of the Factorization and Renormalization Scale

Dedicated simulated datasets have been generated to estimate the systematic uncertainty on the measured $t\bar{t}$ production cross-section due to the choice of the factorization and renormalization scale during the generation process of the nominal MC datasets. For this purpose, μ_F and μ_R in the generation of $t\bar{t}$, W +jets and Z/γ^* +jets processes are chosen to be twice the nominal scale, considered as upwards fluctuation, and half the original scale which is considered as the corresponding downwards fluctuation. Due to this variation, the number of predicted W +jets and Z/γ^* +jets is seriously affected as can be seen in table 5.4. Therefore, the observed discrepancy between the number of predicted events from simulation and the number of observed events might be caused by a slight overestimation of the employed factorization and renormalization scale during the production of the nominal MC datasets. A comparison of the nominal \cancel{E}_T and M3 templates and the systematically varied templates is illustrated in figure 5.9. Since the production of W bosons and Z bosons in association with jets is based on similar physics processes, it is reasonable to treat effects due to the scale variation in both datasets as a common systematic uncertainty. In contrast, the production of top-quark pairs via the strong interactions is based on a different physics process and the scale variations has to be treated independently from the scale variation for W boson and Z boson production, leading to two separate sources of systematic uncertainties.

Variation of the Matching Threshold

During the generation of the simulated datasets a minimal parton- E_T threshold is applied within the MLM matching prescription, see section 3.1. A variation of this threshold leads to different predictions for the number of events in the selected dataset and for the shape of the templates used to model the observed \cancel{E}_T and M3 distributions. Consequently, dedicated datasets are generated for $t\bar{t}$, W +jets, and Z/γ^* +jets processes, where the matching threshold is chosen to be twice or half of its nominal value. The relative changes to the event yield of the selected dataset due to the variation of the matching threshold are given in table 5.5. The differences between the nominal template and the systematically varied templates are depicted in figure 5.10.

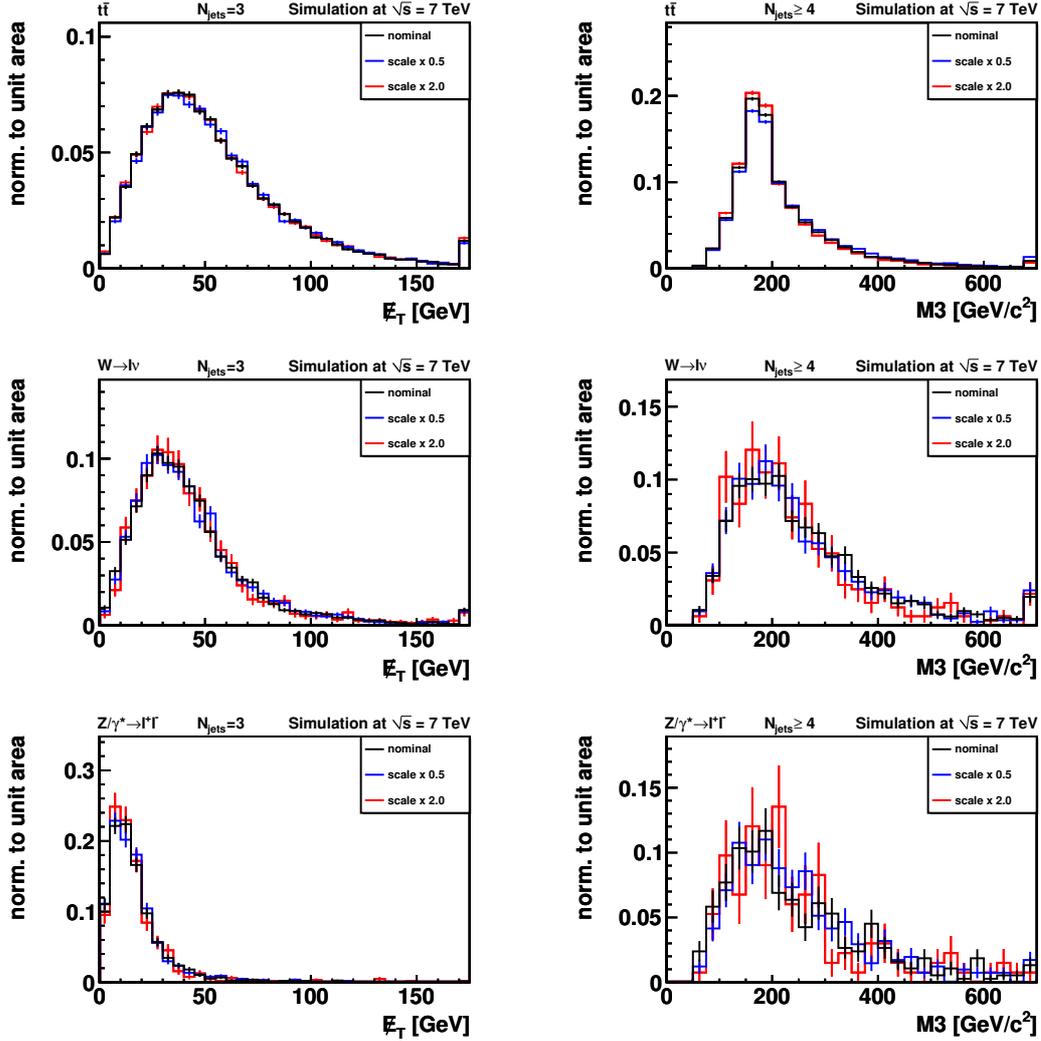


Figure 5.9: Comparison of the nominal templates with the templates obtained from dedicated MC datasets generated with twice the nominal factorization and hadronization scale and half the original scale. Illustrations for $t\bar{t}$, W +jets, and Z/γ^* +jets processes are shown for the two fit variables E_T and $M3$.

As for the variations of the factorization and renormalization scale, the variation of the matching threshold is split into two source of systematic uncertainties, one for the $t\bar{t}$ production process as well as one for W boson and Z boson production.

	Variation	$t\bar{t}$	W +jets	Z/γ^* +jets
= 3 jets	Matching Threshold $\times 0.5$	-2.8%	-2.2%	+0.0%
= 3 jets	Matching Threshold $\times 2.0$	+1.3%	+3.9%	+7.6%
≥ 4 jets	Matching Threshold $\times 0.5$	+0.9%	-4.4%	+5.9%
≥ 4 jets	Matching Threshold $\times 2.0$	-1.7%	-9.2%	+3.6%

Table 5.5: Relative changes of the expected number of events for $t\bar{t}$, W +jets, and Z/γ^* +jets processes due to variations of the chosen matching threshold in the MLM matching prescription. The quoted values are obtained from dedicated simulated events where twice the nominal threshold and half the original threshold have been chosen.

Uncertainty due to the Presence of Pile-Up

As stated above, the average number of additional soft proton-proton interactions per bunch crossing is 2.2 in the analyzed dataset. However, the employed simulation does not include any pile-up events. Hence, the energy of jets observed in collision data might be larger than in simulation because of the additional 0.75 GeV which are on average deposited in the detector per pile-up event. Moreover, the isolation energies might differ between collision data and simulation for the same reason. To account for these effects and to estimate their influence on the measured $t\bar{t}$ production cross-section, dedicated MC datasets have been generated for the $t\bar{t}$ signal and the dominating W +jets background. In these samples the hard scattering process is overlaid by pile-up events. Unfortunately, the average number of pile-up events is slightly overestimated and thus these samples could not be used as the nominal MC datasets for this analysis. Nevertheless, they are used to obtain a conservative estimate on the influence of additional soft proton-proton interactions on the $t\bar{t}$ production cross-section measurement. In table 5.6 the relative changes to the number of expected events for $t\bar{t}$ signal and the W +jets background are shown. It can be seen, that pile-up seriously affects the number of expected W +jets events, and could thus be a further reason for the observed discrepancy between the predicted and the observed number of events. A comparison between the nominal template and the template obtained from simulations including pile-up events is depicted in figure 5.11 for both fit variables. Since it is not possible to have less than zero pile-up events included in the simulation, the systematic variation due to pile-up is solely treated as an upwards fluctuation.

Variation of the Initial- and Final-State Radiation

The process of initial- and final-state radiation (ISR/FSR) in the generation of simulated events has been discussed in section 3.1. To account for the effects on the

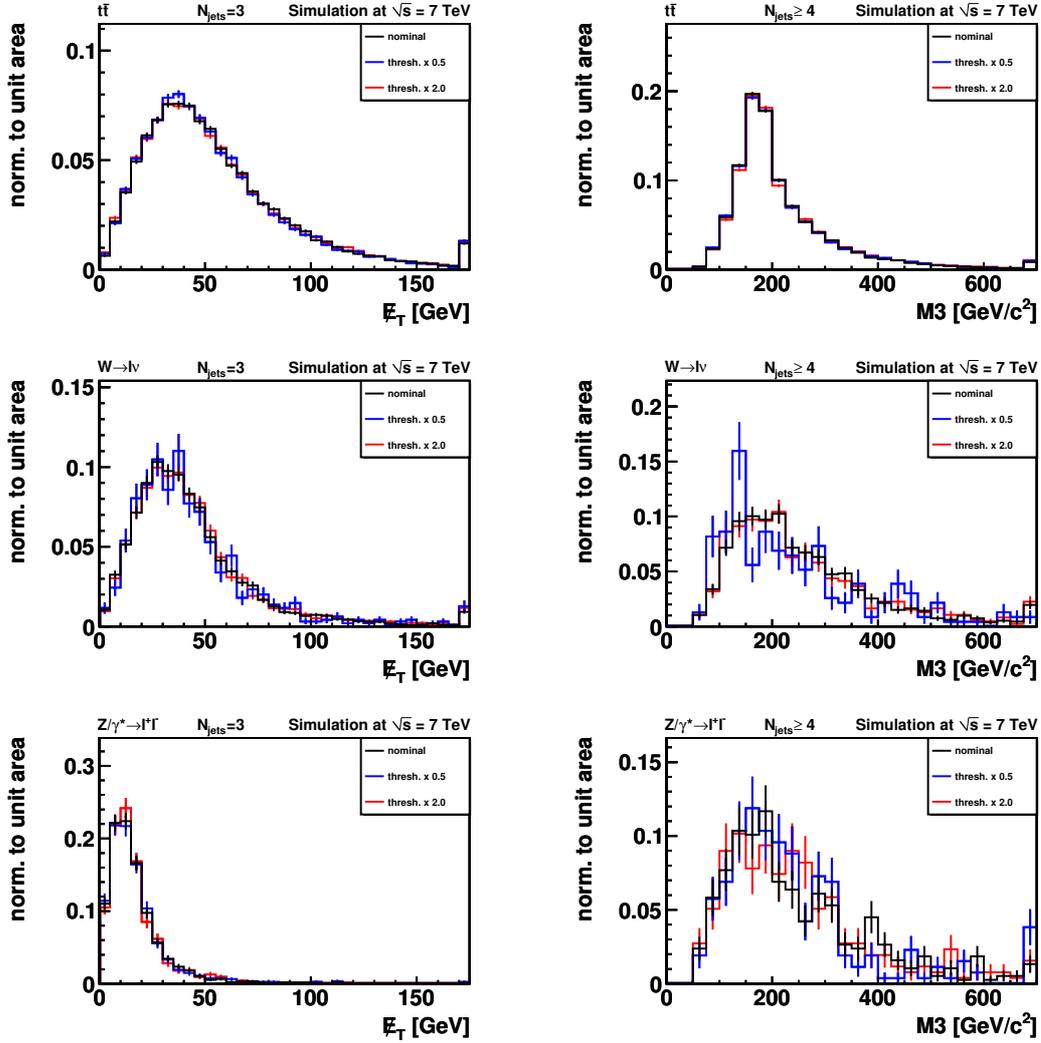


Figure 5.10: Comparison of the nominal templates with the templates obtained from dedicated MC datasets generated with twice the nominal matching threshold and half the original threshold. Illustrations for $t\bar{t}$, W +jets, and Z/γ^* +jets processes are shown for the two fit variables E_T and $M3$.

	Variation	$t\bar{t}$	W +jets
$= 3$ jets	with Pile-Up	-1.7%	+15.7%
≥ 4 jets	with Pile-Up	+5.1%	+19.6%

Table 5.6: Relative changes of the expected number of events for $t\bar{t}$ and W +jets processes when including pile-up events in the simulation. The quoted values are obtained from dedicated simulations where the number of average pile-up events is slightly overestimated.

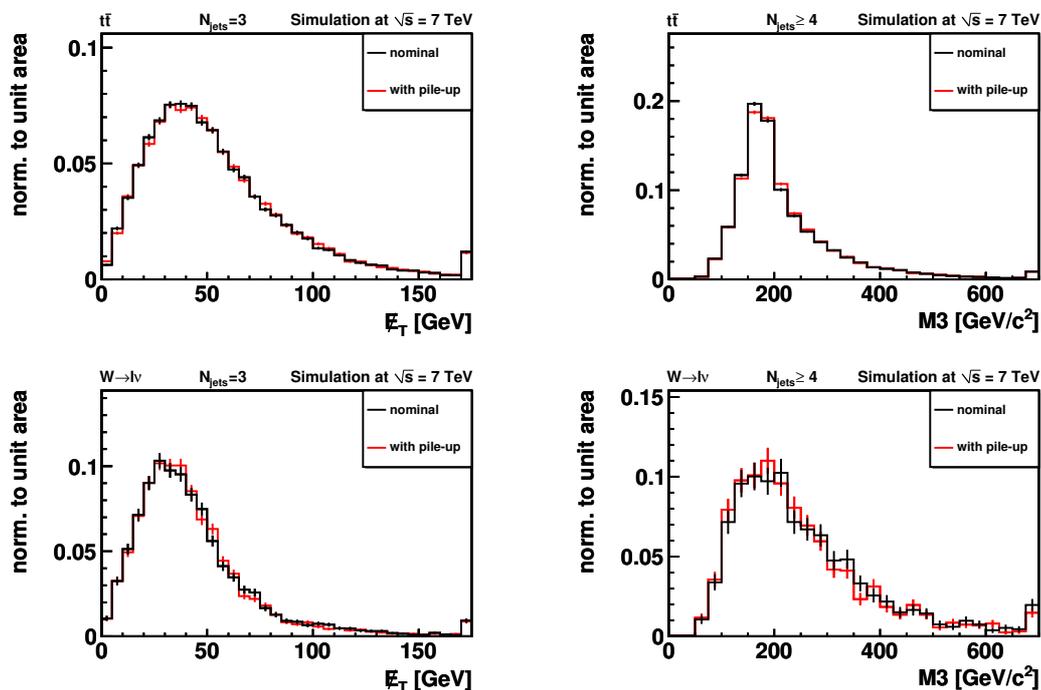


Figure 5.11: Comparison of the nominal template with the template obtained from dedicated MC datasets where the hard interaction has been overlaid with pile-up events. Illustrations for $t\bar{t}$ and W +jets processes are shown for the two fit variables \cancel{E}_T and M_3 .

measurement of $t\bar{t}$ production cross-section arising from a different modelling of these processes, the amount of ISR and FSR in the parton shower is once increased and once decreased during the simulation of the $t\bar{t}$ signal. Using this MC dataset with less ISR/FSR, the number of predicted events with exactly three jets increases by 0.5%, while the number of predicted events with at least four jets decreases by 0.4%. When using the MC dataset with more ISR/FSR, the number of predicted events with exactly three jets decreases by 1.5% and the expected number of events with at least four jets decreases by 1.9%. In figure 5.12 a comparison of the nominal $t\bar{t}$ templates with the $t\bar{t}$ templates obtained from MC datasets with a varied amount of ISR/FSR is illustrated.

Variation of the W Boson Branching Ratio

In section 4.2.2 it has been explained, that the $W \rightarrow l\nu$ branching ratio in the simulation of $t\bar{t}$ has been corrected to its measured value of $\text{BR}(W \rightarrow l\nu) = 0.1080$ by reweighting each simulated event. The hadronic branching fraction has been corrected accordingly to $\text{BR}(W \rightarrow q\bar{q}') = 0.676$. The uncertainty on the leptonic branching ratio is roughly 1% and thus the branching ratio is taken to be 0.109 and 0.107 to account for upwards and downwards fluctuations, respectively. The hadronic branching ratio is simultaneously changed to 0.673 and 0.679. Accordingly, for the smaller leptonic

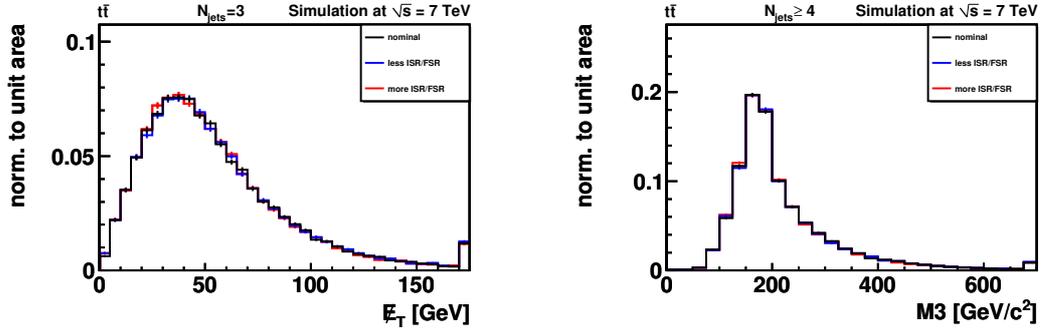


Figure 5.12: Comparison of the nominal $t\bar{t}$ template with the templates obtained from dedicated MC datasets with a varied amount of ISR/FSR in the parton shower simulation.

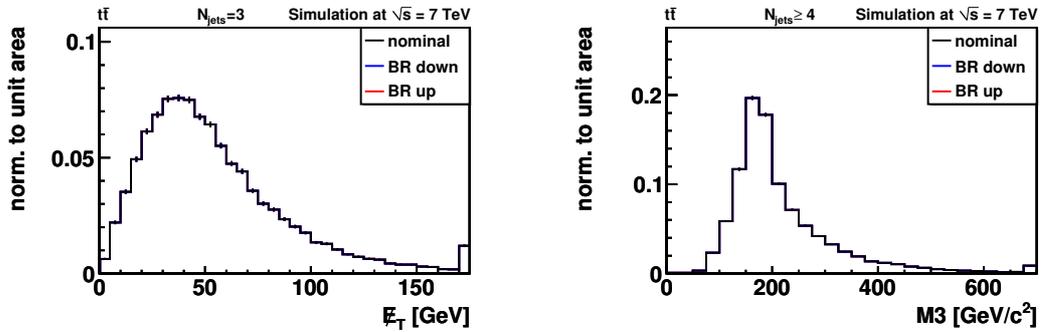


Figure 5.13: Comparison of the nominal $t\bar{t}$ template with the templates obtained after varying $\text{BR}(W \rightarrow l\nu)$ within its uncertainties where $\text{BR}(W \rightarrow q\bar{q}')$ have been varied accordingly. Nearly no differences are visible in the distributions of E_T and M_3 .

branching ratio the number of expected events in both phase space regions decreases by 0.9% while it increases by the same amount for the larger leptonic branching ratio. In figure 5.13 the varied templates are compared with the nominal $t\bar{t}$ template, but differences are only hardly visible.

Variation of the Parton Distribution Function

The uncertainty on the $t\bar{t}$ production cross-section measurement due to the imperfect knowledge on the parton distribution function of the colliding protons, is evaluated using the CTEQ6.6 PDF set [181] and its $2 \cdot 22$ eigenvector basis. Using these eigenvectors, the uncertainty on any physical quantity due to the uncertainties on the proton PDF can be evaluated. For this purpose, each of the 22 pairs of eigenvectors is considered as separate upwards and downwards fluctuation, i.e. there are in principle 22 separate systematic uncertainties due to the imperfect knowledge on the proton PDF. The fluctuations correspond to 90% confidence intervals. Their effect is evaluated by reweighting each event, represented by the fraction of the proton momentum carried by the two partons participating in the hard interaction as well as the chosen factor-

	Variation	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets
= 3 jets	PDF max. downwards fluc.	-0.5%	-0.5%	-2.9%	-2.4%
= 3 jets	PDF max. upwards fluc.	+0.3%	+0.7%	+1.5%	+1.2%
≥ 4 jets	PDF max. downwards fluc.	-0.5%	-0.4%	-3.4%	-2.9%
≥ 4 jets	PDF max. upwards fluc.	+0.6%	+0.4%	+3.3%	+2.9%

Table 5.7: Maximum relative changes of the expected number of events for $t\bar{t}$, Single-Top, W +jets, and Z/γ^* +jets processes due to variations of the employed parton distribution function.

ization scale, with the product of the PDF values obtained from the eigenvector PDF divided by the product of the nominal PDF values. The reweighting is done with the LHAPDF package [182].

In this way, the effect of variations of the employed PDF are evaluated for $t\bar{t}$, Single-Top, W +jets and Z/γ^* +jets processes. In table 5.7 the maximum relative changes in the number of expected events found for the different eigenvectors are shown. In summary, the observed differences are only small and also only hardly visible in the distributions of \cancel{E}_T and M3. Thus, for this source of systematic uncertainty no comparison between the nominal and the 44 systematically varied templates is shown.

Expected Effects on the $t\bar{t}$ Production Cross-Section Measurement

In order to estimate the influence of all mentioned sources of systematic uncertainties, ensemble tests are employed, where the systematic effects are included in the pseudo-data distribution. For this purpose, a parameter ξ^u representing the strength of a particular systematic uncertainty u is defined. To derive the pseudo-data distribution, first a ξ^u value for each incorporated systematic uncertainty is drawn from a Gaussian distribution centered at zero with a width of 1.0, except for ξ^{PDF} where the width of the Gaussian distribution is only 0.608 to account for the quoted 90% confidence intervals of this uncertainty. The nominal α_{ijk} of equation 5.2 are then replaced by $\alpha_{ijk}^{\text{syst}}$ defined as:

$$\alpha_{ijk}^{\text{syst}}(\vec{\xi}) = \alpha_{ijk} + \sum_u |\xi^u| \cdot (\Xi_{ujk}^{\text{sign}(\xi^u)} - \alpha_{ijk}). \quad (5.8)$$

In this equation, $\vec{\xi}$ is an abbreviation for all contributing ξ^u and the sum runs over all incorporated systematic uncertainties and $\text{sign}(\xi^u)$ denotes the sign of the drawn ξ^u value. To simplify the discussion of equation 5.8 only one contributing systematic uncertainty is assumed exemplarily. In this case, in a scenario where $\xi^u = 0$ is drawn, $\alpha_{ijk}^{\text{syst}}$ is equal to the nominal α_{ijk} and no effect from the systematic uncertainty is included in the pseudo-data distribution. In case that ξ^u is equal to ± 1.0 , $\alpha_{jk}^{\text{syst}}$ is simply the systematically varied template Ξ_{ujk}^{\pm} corresponding to a fluctuation of ± 1 standard deviation. For all cases in which ξ^u is different from 0.0 and ± 1.0 and $|\xi^u| < 1.0$, $\alpha_{ijk}^{\text{syst}}$ represents a linear interpolation between the nominal and the systematically varied template, while for $|\xi^u| > 1.0$ $\alpha_{jk}^{\text{syst}}$ represents a linear extrapolation.

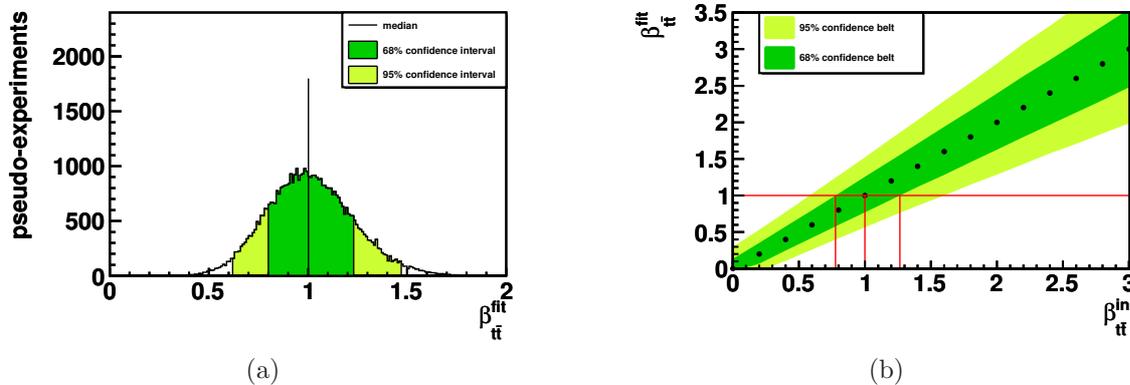


Figure 5.14: In (a) the $\beta_{t\bar{t}}^{\text{fit}}$ distribution for $\beta_{t\bar{t}}^{\text{in}} = 1.0$ when solely varying the jet energy scale is shown for illustration purposes. Compared to figure 5.3(a) the distribution has become broader leading to larger 68% and 95% confidence intervals. The Neyman construction derived from ensembles where all systematic uncertainties have been taken into account is depicted in (b). The expected uncertainty on the measurement is derived by drawing a horizontal line at $\beta_{t\bar{t}}^{\text{fit}} = 1.0$ and evaluating the intersections with the 68% confidence belt.

Although systematic effects are included in the pseudo-data distribution, the model used to describe this distribution remains unchanged. Therefore, still the nominal templates are employed in the likelihood fitting procedure. For illustration purposes, the $\beta_{t\bar{t}}^{\text{fit}}$ distribution for $\beta_{t\bar{t}}^{\text{in}} = 1.0$ if solely the JES uncertainty is included is shown in figure 5.14(a). Compared to the scenario without any incorporated systematic uncertainty (figure 5.3(a)) the distribution has become broader which results in broader 68% and 95% confidence intervals. The Neyman construction derived from ensembles where all sources of systematics have been taken into account is shown in figure 5.14(b). Again, the expected uncertainty on the measured $t\bar{t}$ production cross-section is derived by drawing a horizontal line at $\beta_{t\bar{t}}^{\text{fit}} = 1.0$ and reading off the $\beta_{t\bar{t}}^{\text{in}}$ values at the intersections with the boundaries of the 68% confidence belt. In table 5.8, the expected uncertainties for including only a particular systematic uncertainty as well as for incorporating all systematic uncertainties at the same time are given. Since all uncertainties have been derived using a Neyman construction, the listed values correspond to the combined statistical and systematic uncertainty. It can be seen, that when analyzing 35.9 pb^{-1} of collision data the measurement is already limited by systematic effects. The by far dominating systematic uncertainty is introduced by the uncertainty on the jet energy scale.

5.3.2 Result

In order to measure the $t\bar{t}$ production cross-section, the maximum likelihood fit procedure is applied to the dataset corresponding to 35.9 pb^{-1} of real collision data. The resulting values for the β_k -parameters which are called β_k^{data} , are summarized in table 5.9 together with their statistical uncertainties that are derived from the covariance

Variation	Combined Stat.+Syst. Uncertainty
Jet Energy Resolution	+14.0% -13.1%
Jet Energy Scale	+23.5% -20.4%
Unclustered Energy	+14.0% -13.1%
Electron Energy Scale	+14.0% -13.1%
Trigger and Selection Eff.	+14.5% -13.6%
QCD/ γ +jets Model	+14.7% -14.8%
Factorization and Renormalization Scale	+15.5% -14.3%
Matching Threshold	+15.0% -14.0%
Pile-Up	+14.4% -13.8%
Initial- and Final-State Radiation	+14.0% -13.3%
W Boson Branching Ratio	+14.0% -13.2%
Parton Distribution Function	+14.0% -13.1%
Total	+26.6% -22.2%

Table 5.8: Results for the expected ($\beta_{t\bar{t}}^{\text{fit}} = 1.0$) uncertainties on the $t\bar{t}$ production cross-section arising from a particular source of a systematic uncertainty. In the last row, the combined uncertainty due to all systematic variations is shown. All values have been extracted from a Neyman construction, hence, the quoted values are combined statistical and systematic uncertainties. The largest uncertainty arises from the variation of the jet energy scale while some sources of systematic uncertainties are shown to be negligible.

	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets	QCD/ γ +jets
β_k^{data}	1.14	1.05 ± 0.3	1.43 ± 0.13	1.43 ± 0.43	1.15 ± 0.14

Table 5.9: Results of the maximum likelihood fit to collision data. The uncertainties quoted for the β_k^{data} of the background processes are purely statistical and are derived from the covariance matrix at the maximum of the likelihood function.

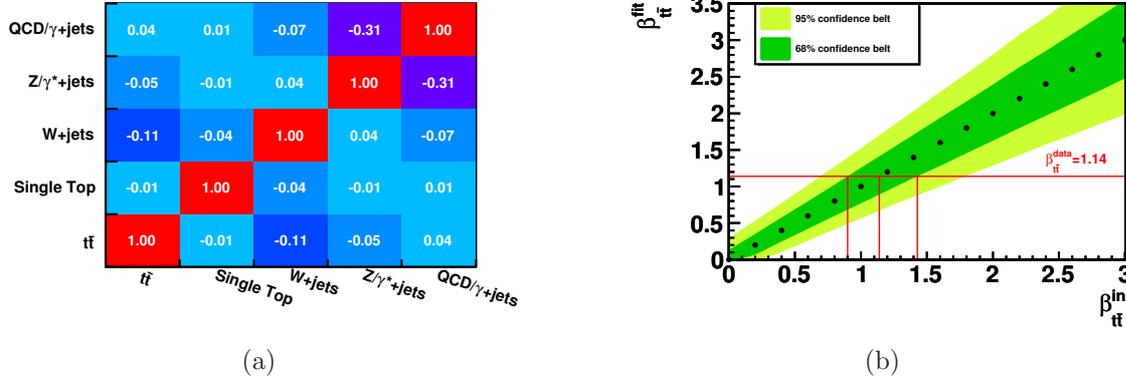


Figure 5.15: In (a) the correlation matrix at the maximum of the likelihood function is shown. The quoted values correspond to the correlation coefficients. The evaluation of the Neyman construction is shown in (b). A horizontal line is drawn at $\beta_{t\bar{t}}^{\text{data}} = 1.14$ and the intersections of this line with the boundaries of the 68% confidence belt are used to derive the combined stat.+syst. uncertainty on the measurement of the $t\bar{t}$ production cross-section.

matrix at the maximum of the likelihood function. In figure 5.15(a) the correlation matrix for the likelihood fit is shown. As expected, a large anti-correlation between the fit result for the Drell-Yan process and the QCD/ γ +jets process is found because of the large similarity of their \cancel{E}_T and M3 distributions. Moreover, a slight anti-correlation is found between the fit results for $t\bar{t}$ and W +jets processes.

Using the Neyman construction shown in figure 5.14(b), the observed $t\bar{t}$ production cross-section and the uncertainties of the measurement are derived by drawing a horizontal line at $\beta_{t\bar{t}}^{\text{data}} = 1.14$ and evaluating its intersection with the boundaries of the 68% confidence belt. This evaluation of the Neyman construction is depicted in figure 5.15(b) and leads to

$$\sigma_{t\bar{t}} = 180_{-38}^{+45} (\text{stat.} + \text{syst.}) \pm 7 (\text{lumi.}) \text{ pb}. \quad (5.9)$$

The additional uncertainty of ± 7 pb is due to the uncertainty on the luminosity measurement of $\pm 4\%$. The result of the measurement is in good agreement with the theoretical predictions at NLO and approximate NNLO of $\hat{\sigma}_{t\bar{t}} = 157.5_{-24.4}^{+23.2}$ pb and $\hat{\sigma}_{t\bar{t}} = 163_{-10}^{+11}$ pb, respectively. The discrepancies between the number of predicted events and the number of observed events is mainly absorbed in the β_k^{data} -values of the W +jets and Z/γ^* +jets processes, which are found to be roughly 40% larger than predicted. As discussed above this might be due to an imperfect choice of the factorization and renormalization scale during the event generation process or a consequence of pile-up events which have not been included in the simulation.

The β_k^{data} -values from table 5.9 can be used to scale the predicted number of events for the different processes to their observed numbers. Afterwards, the modelling of different kinematic variables can be re-examined by comparing the simulation with observed collision data. Such comparisons are shown in figure 5.16 for the two fit

variables \cancel{E}_T and M3 as well as three further kinematic variables for events with exactly three and at least four jets. In the distribution of the M3 variable, a region of ten adjacent bins where the number of observed data events is larger than the fitted number of events is observed. This issue is investigated using an ensemble of 50,000 pseudo-experiments including all discussed sources of systematic uncertainties within the pseudo-data distributions. For each pseudo-experiment the maximum likelihood fit is performed and the Kolmogorov-Smirnov (KS) distance is derived comparing the pseudo-data and the fitted distributions. In 10% of all cases the KS value observed in a particular pseudo-experiment is larger than the KS value observed comparing the true data distribution with the simulated distribution normalized to the fit result. Consequently, the observed deviation is consistent with reasonable statistical fluctuations and systematic variations of the templates. As can be seen in figure 5.16, the distribution of all the other shown variables is modelled very well.

5.4 Combined Electron+Jets and Muon+Jets Measurement

Besides using events with an electron+jets final state, the $t\bar{t}$ production cross-section is also measured using events with a muon+jets signature. Naturally, both final states can be used to obtain a combined measurement with smaller uncertainties. Roughly twice the number of events are available for the analysis which is expected to lead to an about $1/\sqrt{2}$ smaller statistical uncertainty. However, since each single measurement is already limited by systematic uncertainties, the overall reduction of the uncertainties is expected to be small. In the following, a short overview of the analysis in the muon+jets final state is given and a measurement of the top-quark pair production cross-section is performed in the combined electron+jets and muon+jets final state. The relative fraction of $t\bar{t}$ in the combined dataset is derived through a simultaneous fit to four observables. As before, for events with an electron or muon and exactly three jets the \cancel{E}_T distribution is employed, while for events with an electron or muon and at least four jets the M3 variable is chosen. The final $t\bar{t}$ production cross-section and the uncertainties on the measurement are extracted from a Neyman construction.

5.4.1 The Muon+Jets Analysis

The measurement of the top-quark pair production cross-section in events with a muon+jets signature is described in detail in [15]. Only events with exactly one muon with $p_T > 20$ GeV/c, $|\eta| < 2.1$, and $I_{\text{rel}} < 0.05$ fulfilling additional requirements on the quality of the reconstructed muon are selected. Solely events with at least three jets, where jets are similarly defined as in the electron+jets analysis, enter the final selected dataset. The muon selection efficiency as well as the trigger efficiency have been measured using a tag-and-probe method and the simulation has been corrected using dedicated data-to-MC correction factors. At the time this combined measurement is

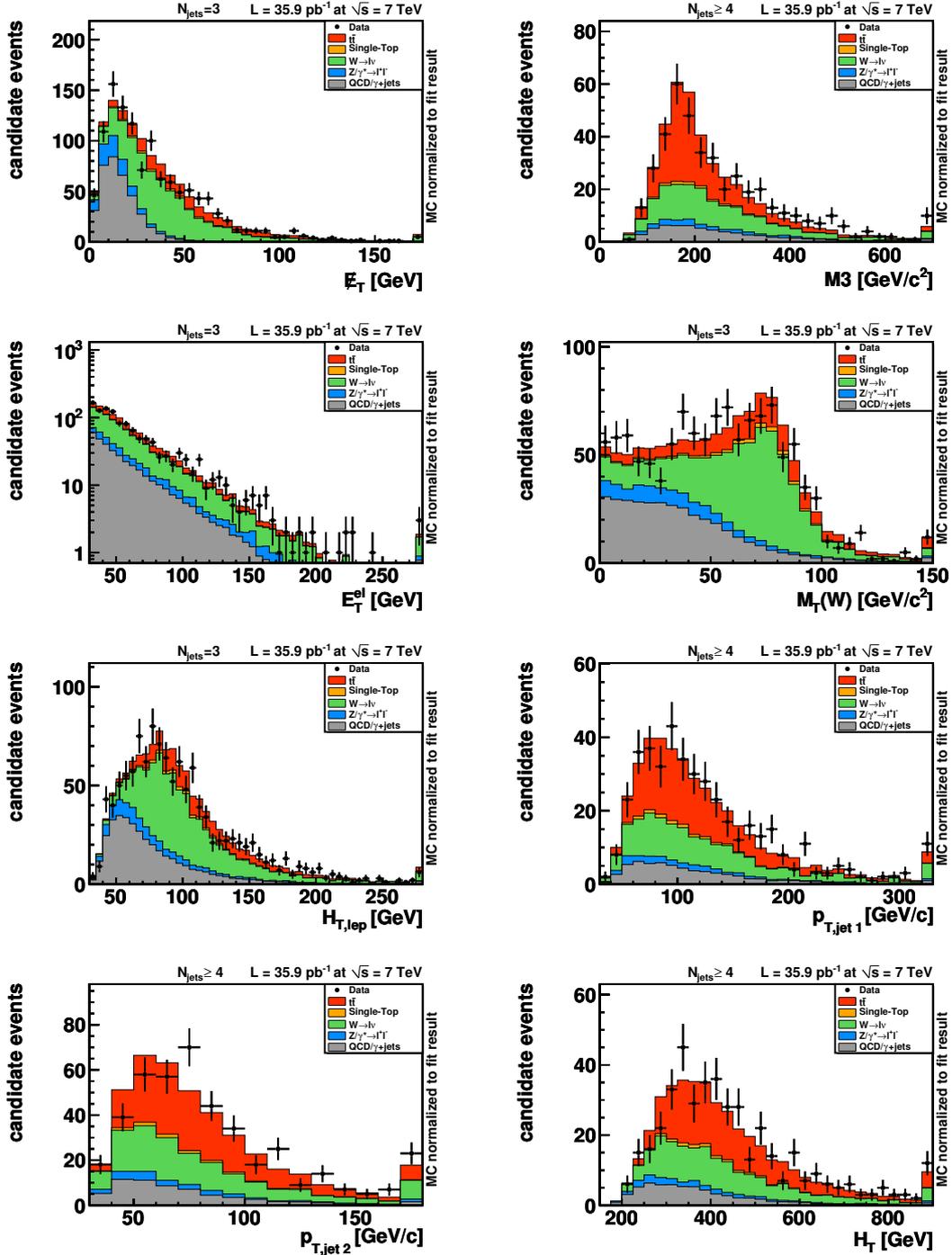


Figure 5.16: Comparison between observed and simulated distributions for different kinematic variables. For each process, the MC template is normalized to the number of events resulting from the maximum likelihood fit. In the upper row the two fit variables E_T and $M3$ are shown. Below, kinematic variables for events with exactly three jets and at least four jets are illustrated.

Selection	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets	QCD multi-jet	Sum (MC)	data
= 3jets	197 ± 31	30 ± 1	486 ± 34	46 ± 3	49 ± 3	807 ± 53	1064
≥ 4 jets	211 ± 33	11 ± 1	115 ± 9	11 ± 1	9 ± 1	358 ± 37	423

Table 5.10: Expected and observed number of events in the muon+jets analysis taken from [15]. The quoted expectations have been scaled to take into account a slight change in the amount of collected data resulting from a more precise luminosity measurement.

performed, a more precise luminosity measurement is available than in [15]. Thus, the integrated luminosity of $36.1 \pm 4.0 \text{ pb}^{-1}$ quoted in [15] is reduced to $35.9 \pm 1.4 \text{ pb}^{-1}$. Taking this difference into account, the expected number of events for all relevant processes is shown in table 5.10 together with the observed number of events in collision data. It can be seen, that also in events with a muon+jets final state a discrepancy between the number of events expected from simulation and the number of events observed in data is present.

In order to derive the relative fraction of $t\bar{t}$ events in the selected dataset, a simultaneous maximum likelihood fit to the observed distributions of \cancel{E}_T (events with exactly three jets) and M3 (events with at least four jets) is performed. For all contributing processes except QCD multi-jet production, the templates used to model these distributions are taken from MC simulations. A template describing the shape of \cancel{E}_T and M3 for QCD multi-jet events is directly extracted from collision data using events containing a muon with $0.2 < I_{\text{rel}} < 0.5$. The β_k^{data} resulting from the likelihood fit are given in table 5.11. It is visible, that also in the muon analysis the main discrepancy between the number of predicted events and the number of observed events is absorbed in larger expectation for W +jets and Z/γ^* +jets events. Furthermore, the number of QCD multi-jet events is found to be more than twice the predicted number of events. The result of the $t\bar{t}$ production cross-section measurement and the corresponding uncertainties are extracted from a Neyman construction [15]:

$$\sigma_{t\bar{t}} = 168_{-35}^{+42} (\text{stat.} + \text{syst.}) \pm 7 (\text{lumi.}) \text{ pb}. \quad (5.10)$$

This result is in very good agreement with the theoretical predictions as well as with the measurement in the electron+jets final state.

5.4.2 Statistical Method and Systematic Uncertainties

The observed fraction of $t\bar{t}$ events in the combined electron+jets and muon+jets dataset is derived by applying a simultaneous maximum likelihood fit to four variables. For electron+jets as well as for muon+jets events, the \cancel{E}_T distribution for events with exactly three jets and the M3 distribution for events with at least four jets are used. The templates used to model these four distributions are extracted from MC simulations, except for the templates describing the QCD multi-jet background in the muon+jets dataset and the QCD/ γ +jets background in the electron+jets dataset

	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets	QCD/ γ +jets
β_k^{data}	1.07	1.0 ± 0.3	1.36 ± 0.10	1.33 ± 0.41	2.12 ± 0.58

Table 5.11: Results of the maximum likelihood fit to collision data in the muon+jets analysis taken from [15]. For the quoted results the slight change in the expectation due to a more precise luminosity measurement has been taken into account. Moreover, the expectation for $t\bar{t}$ events in [15] has been derived using the approximate NNLO calculations for the $t\bar{t}$ production cross-section. In this thesis, the expectations are derived using the NLO predictions, thus the $\beta_{t\bar{t}}^{\text{data}}$ value changes from 1.03 to 1.07. The uncertainties on the derived β_k^{data} for the background processes are purely statistical and are derived from the covariance matrix at the maximum of the likelihood function.

which are derived directly from collision data. QCD multi-jet background contributing to the muon+jets dataset mainly arises from the semileptonic decays of B - and C -hadrons as well as the decays of charged pions and kaons into muons. Since the QCD/ γ +jets contribution in the electron+jets final state arises mainly from conversion electrons, both backgrounds enter the selected dataset for completely different reasons. In the following, they are thus treated independent of each other. The likelihood function given in equation 5.6 is thus extended by a further fit parameter, i.e. $k \in \{t\bar{t}, t, W, Z, \text{QCD}^e, \text{QCD}^\mu\}$ and the product is extended to include all four fit variables, i.e. $j \in \{\cancel{E}_T^e, M3^e, \cancel{E}_T^\mu, M3^\mu\}$. The \cancel{E}_T^μ distribution is divided into 36 bins, while the $M3^\mu$ is divided into 24 bins. Similar to the separate analyses, a Neyman construction is used to derive the measured $t\bar{t}$ production cross-section and the corresponding uncertainties for the combined dataset. The Neyman construction is built based on ensemble tests which consists of 50,000 pseudo experiments, by following the procedure described in section 5.2.2. From these ensembles, a linear behavior of the method is found, i.e. for each ensemble the average $\beta_{t\bar{t}}^{\text{fit}}$ is equal to the particular $\beta_{t\bar{t}}^{\text{in}}$ value used to create the pseudo-data distributions. Furthermore, the Neyman construction built from the results of the different ensemble tests is used to derive the expected ($\beta_{t\bar{t}}^{\text{fit}} = 1.0$) statistical uncertainty on the measurement, which is +8.7% and -8.4%.

In order to evaluate the influence of systematic uncertainties on the $t\bar{t}$ production cross-section measurement in the combined electron+jets and muon+jets final state, ensemble test where the influences of all sources of systematic uncertainties are taken into account in the pseudo-data distributions are used. In principle, the same procedure explained at the end of section 5.3.1 is employed and almost all systematic uncertainties are treated as completely correlated, i.e there is a common ξ^u -parameter drawn for a particular systematic uncertainty for both final states. However, there are also systematic uncertainties, namely the uncertainties due to the imperfect measurement of the lepton selection and trigger efficiencies as well as the uncertainties due to the model for the QCD multi-jet background, which obviously have to be treated as totally uncorrelated. In these cases two separate ξ^u -parameter are drawn. Furthermore, the uncertainty due to an imperfect measurement of the lepton transverse momentum is negligible in case of muons. Therefore this uncertainty is only evaluated for events

Variation	Combined Stat.+Syst. Uncertainty
Jet Energy Resolution	+8.8% -8.4%
Jet Energy Scale	+20.3% -17.6%
Unclustered Energy	+8.7% -8.4%
Electron Energy Scale	+8.7% -8.4%
Trigger and Selection Eff.	+9.2% -8.7%
QCD Multi-Jet Model	+9.1% -8.9%
Factorization and Renormalization Scale	+11.2% -10.6%
Matching Threshold	+10.5% -9.8%
Pile-Up	+9.3% -9.3%
Initial- and Final-State Radiation	+9.0% -8.6%
W Boson Branching Ratio	+8.9% -8.6%
Parton Distribution Function	+8.7% -8.5%
Total	+23.5% -19.3%

Table 5.12: Results for the expected ($\beta_{tt}^{\text{fit}} = 1.0$) uncertainties on the $t\bar{t}$ production cross-section arising from a particular source of a systematic uncertainty in the combined measurement. In the last row, the combined uncertainty due to all systematic variations is shown. All values have been extracted from a Neyman construction, hence, the quoted values are combined statistical and systematic uncertainties. The values for the uncertainties which are treated as uncorrelated between the two final states are included at the same time into the pseudo-data distributions. As expected, the improvement between the uncertainties of the combined measurement and the separate measurements merely is due to the improved statistical uncertainty.

containing electrons. The uncertainties when solely including the effects of a particular source of a systematic uncertainty in the pseudo-data distributions and the uncertainty due to all sources of systematic uncertainties are summarized in table 5.12.

5.4.3 Result

The relative fractions of all processes contributing to the combined electron+jets and muon+jets dataset are derived by applying the maximum likelihood fit to collision data. The resulting β_{tt}^{data} values are given in table 5.13 where the uncertainties on the background parameters are purely statistical and derived from the covariance matrix at the maximum of the likelihood function. As for the results of the separate measurements, most of the discrepancy between the number of predicted events and the number of observed events is absorbed in an almost 40% larger fit value of W +jets and Drell-Yan events. Overall, a good agreement between the results of the separate measurements and the measurement in the combined electron+jets and muon+jets

	$t\bar{t}$	Single-Top	W +jets	Z/γ^* +jets	QCD^e	QCD^μ
β_k^{data}	1.10	1.06 ± 0.30	1.38 ± 0.08	1.33 ± 0.37	1.20 ± 0.12	1.94 ± 0.54

Table 5.13: Results of the maximum likelihood fit to the combined electron+jets and muon+jets dataset. The uncertainties quoted for the β_k^{data} of the background processes are purely statistical and are derived from the covariance matrix of the likelihood fit.

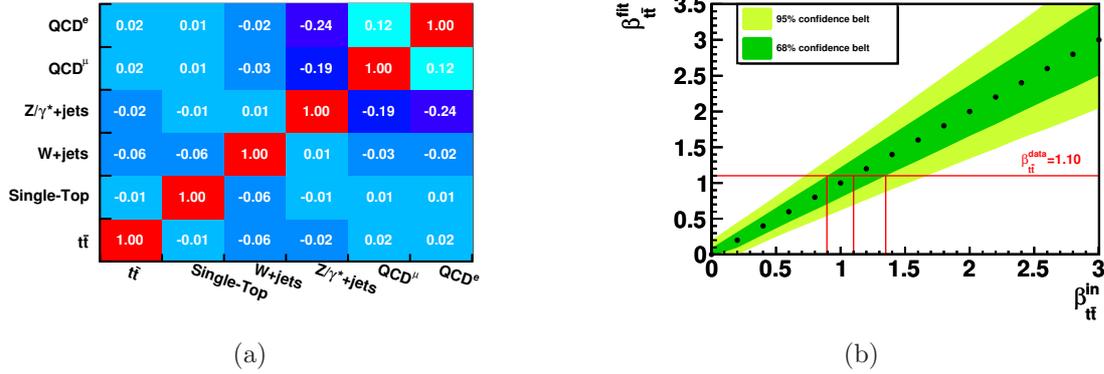


Figure 5.17: In (a) the correlation matrix for the different fit parameters obtained from the maximum likelihood fit to the combined electron+jets and muon+jets dataset is shown. The quoted values correspond to the correlation coefficients. The evaluation of the Neyman construction is shown in (b). A horizontal line is drawn at $\beta_{t\bar{t}}^{\text{data}} = 1.10$ and the intersections of this line with the 68% confidence belt are used to derive the combined stat.+syst. uncertainty on the measurement of the $t\bar{t}$ production cross-section.

final state is found. In figure 5.17 (a) the correlation matrix for the different fit parameters is shown. Besides the already mentioned anti-correlation between QCD multi-jet and Z/γ^* +jets processes a correlation between the QCD/ γ +jets process in the electron+jets and the QCD multi-jet process in the muon+jets dataset is observed. In figure 5.17 (b) the Neyman construction derived from ensembles including the effects of all sources of systematic uncertainties in the pseudo-data distribution is illustrated. The top-quark pair production cross-section and the uncertainties on the measurement are derived by drawing a horizontal line at $\beta_{t\bar{t}}^{\text{data}} = 1.10$ and evaluating the intersection of this line with boundaries of the 68% confidence belt. This results in

$$\sigma_{t\bar{t}} = 173_{-32}^{+39} (\text{stat.} + \text{syst.}) \pm 7 (\text{lumi.}) \text{ pb}, \quad (5.11)$$

which is in very good agreement with the theoretical predictions at NLO and approximate NNLO. As expected, only a slight reduction of the uncertainties on the measured $t\bar{t}$ production cross-section is achieved through combining the electron+jets and muon+jets datasets.

Summary and Conclusion

Starting the operation of CERN's Large Hadron Collider (LHC) at a center-of-mass energy of 7 TeV at the end of March 2010, a new era in the research field of high energy particle physics has begun. The LHC enables particle physicists to study the collisions of protons with a roughly 3.5 times larger center-of-mass energy than has ever been reached in a laboratory based experiment before. During its 2010 operation the LHC delivered about 47 pb^{-1} of collision data to each of the two multi-purpose particle detectors which are installed along the 26.7 km long ring. For the measurement described in this thesis, the collision data collected by the Compact Muon Solenoid (CMS) detector has been analyzed.

The LHC allows to test the predictions deduced from the Standard Model of particle physics (SM) at a new energy regime. With a mass of $173.3 \text{ GeV}/c^2$ [8], the top quark is the heaviest particle among the fundamental fermions. It has been discovered in 1995 by the CDF [9] and DØ [10] experiments at Fermilab's proton anti-proton collider Tevatron. At the Tevatron and the LHC, top quarks are dominantly produced in pairs of top and anti-top quarks via the strong interaction, however, also the recently observed production of single top quarks [70, 71] via the weak interaction is possible. The measurement of the $t\bar{t}$ production cross section at the LHC is one of the first important milestones for testing the SM at the new energy regime. Besides being interesting itself, a precise knowledge of the $t\bar{t}$ production cross section is a crucial prerequisite for many interesting measurements in the top-quark sector. Furthermore, top-quark events are the main background contribution in many searches for phenomena which are not described by the framework of the SM.

Within the SM, top quarks decay with a probability of almost 100% into a W boson and a bottom quark. According to the decay of the W boson, the decay channels of top quark pairs can be identified as being fully-hadronic (both W bosons decay into quarks), di-leptonic (both W bosons decay into a charged lepton and the corresponding neutrino), or semi-leptonic (one W boson decays into quarks while the other decays into a charged lepton and the corresponding neutrino). The latter channel is also called the lepton+jets channel and events with an electron+jets final state are used to measure the $t\bar{t}$ production cross section in this thesis. Although the event signature of a semi-leptonically decaying top-quark pair with an isolated electron accompanied by four highly energetic jets is very characteristic, there are nevertheless different physics processes which occupy the same final state. To develop an event selection strategy, keeping as much of the $t\bar{t}$ signal events as possible while efficiently rejecting the

expected background events, all relevant physics processes have been simulated using dedicated Monte Carlo event generators. Finally, only events with exactly one highly-energetic and isolated electron passing several electron identification criteria and at least three jets have been selected. In this thesis, explicitly no b -tagging algorithms have been used to identify jets originating from bottom quarks. This is reasonable, since it is important to understand the different background contributions in the dataset with and without b -tagging requirements applied. Furthermore, it is possible that signs of physics which are not explained in the context of the SM might contribute differently to both datasets.

Since the efficiencies for the several selection requirements applied to electrons might differ between simulation and real collision data, these efficiencies have been measured in collision data directly and data-to-MC correction factors have been applied to correct the simulation. For this purpose, a tag-and-probe method exploiting the well-known mass of the Z boson resonance has been used. To estimate the contribution of background events to the $Z \rightarrow ee$ dataset a same-sign/opposite-sign method has been employed which estimates the background contribution to the dataset with oppositely charged electrons from the dataset where both electrons have the same charge. The derived data-to-MC correction factor that has been used to weight each simulated event is 0.933 ± 0.025 .

The analyzed dataset corresponds to an integrated luminosity of $35.9 \pm 1.4 \text{ pb}^{-1}$ which has been certified by the CMS collaboration for physics analysis, i.e. all CMS subdetectors were operational and no detector malfunctions were observed during the data taking process. From simulations which already include the data-to-MC correction factor mentioned above, the number of predicted events with an electron and at least three jets is 1272 ± 80 . However, in collision data 1611 events have been observed. Several reasons for this deviation have been discussed, like a slight over-estimation of the factorization and renormalization scale used during the generation process of the simulated samples as well as the effects of additional soft proton-proton collisions within the same bunch crossing which have not been taken into account in the simulations. Nevertheless, normalizing the predicted number of events to the number of events observed in data, it has been shown, that the MC simulations are able to sufficiently model the shape of different kinematic distributions. Hence, it was not possible to simply subtract the number of predicted background events from the number of observed events to obtain the number of $t\bar{t}$ events in the selected dataset but the relative fraction of all contributing processes had to be derived using more elaborate statistical methods.

Since the predicted number of $t\bar{t}$ events in the selected dataset already takes into account the effects of limited detector acceptances and finite selection efficiencies, the ratio of the observed and the predicted number of $t\bar{t}$ events is equal to the ratio of the measured and the predicted top-quark pair production cross-section ($\beta_{t\bar{t}}$). To extract the observed number of $t\bar{t}$ events, the selected dataset has been divided into two subsets, one containing events with exactly three jets in the final state and one with an event signature of at least four jets. A simultaneous maximum likelihood fit to

two kinematic variables, i.e. the missing transverse energy for events with exactly three jets and the M3 variable for events with at least four jets, has been performed to derive the relative fractions of all contributing processes. M3 is a simple estimator for the mass of the hadronically decaying top quark. Based on the predicted jet multiplicity distribution, the fit takes advantage of the large background contamination in events with exactly three jets to derive the relative fractions of the contributing background processes. The larger signal purity in events with at least four jets, is employed to get an estimate for the relative fraction of signal events.

In order to estimate the expected uncertainties on the measurement of the $t\bar{t}$ production cross-section, ensemble tests have been performed. For these ensemble tests, MC techniques have been used to obtain pseudo-data distributions for both fit variables according to the expectations and the maximum likelihood fit has been applied to this pseudo-data. The results of these fits are used to derive an estimation for the central value of the $t\bar{t}$ production cross-section and to construct central 68% and 95% confidence intervals, respectively. From ensemble tests with different assumed $t\bar{t}$ production cross-sections a Neyman construction of central intervals has been built, from which the uncertainty on the final measurement has been derived. The top-quark pair production cross-section measured in the electron+jets final state (assuming a top-quark mass of 172.5 GeV/c²) is

$$\sigma_{t\bar{t}} = 180_{-38}^{+45} (\text{stat.} + \text{syst.}) \pm 7 (\text{lumi.}) \text{ pb}, \quad (5.12)$$

which is in good agreement with the theoretical predictions calculated at NLO and approximate NNLO of $\hat{\sigma}_{t\bar{t}} = 157.5_{-24.4}^{+23.2}$ pb and $\hat{\sigma}_{t\bar{t}} = 163_{-10}^{+11}$ pb, respectively. The observed discrepancy between the number of predicted and observed events for the selected dataset is mainly absorbed in a roughly 40% larger observed number of W +jets and Z/γ^* +jets events. The presented measurement is already limited due to systematic uncertainties, where the uncertainty on the employed jet energy correction factors has been shown to have by far the largest impact. In order to reduce the influence of the JES uncertainty on future measurements, different strategies might be pursued. First, the uncertainties on the measured jet energy correction factor will decrease if a larger amount of collision data can be used to derive them. Furthermore, an additional fit parameter might be included in the likelihood function which takes into account the effects of a possible JES variation within the fitting procedure. In this way, information on the strength of the JES uncertainty might be directly extracted from collision data.

Besides the measurement in the electron+jets channel, the $t\bar{t}$ production cross-section has also been measured in the combined electron+jets and muon+jets dataset. For this purpose, the events selected in a similar analysis performed in the muon+jets channel [15] have been included and the relative fraction of $t\bar{t}$ events is extracted by a simultaneous fit to four kinematic variables, the \cancel{E}_T distributions for events with exactly three jets and the M3 distributions for events with at least four jets, separately for both event signatures. The $t\bar{t}$ production cross-section measured in the combined dataset is

$$\sigma_{t\bar{t}} = 173_{-32}^{+39} (\text{stat.} + \text{syst.}) \pm 7 (\text{lumi.}) \text{ pb}, \quad (5.13)$$

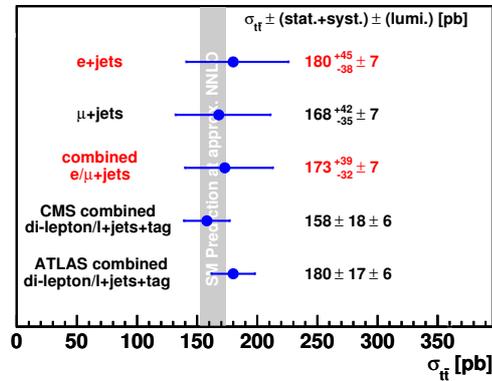


Figure 5.18: Overview of different top-quark pair production cross-section measurements. Besides the three measurements discussed in this thesis, the result of a measurement in the combined di-lepton and lepton+jets final state using b -tagging algorithms [183] reported by the CMS collaboration and the combination of results from the ATLAS collaboration [184] are shown. For a comparison with the SM prediction, the theoretical calculation (assuming a top-quark mass of $172.5 \text{ GeV}/c^2$) at approximate NNLO of $\hat{\sigma}_{t\bar{t}} = 163^{+11}_{-10} \text{ pb}$ is depicted.

which is also in very good agreement with the theoretical predictions. As expected, the improvement on the uncertainties on the measured production cross-section is small, since both separate measurements are already limited due to systematic uncertainties. The observed improvements only results from the larger amount of analyzed data leading to a smaller statistical uncertainty.

In figure 5.18 an overview of different measurements of the top-quark pair production cross-section is given. Besides the three measurements discussed in this thesis, the result of a measurement in the combined di-lepton and lepton+jets final state using b -tagging algorithms [183] reported by the CMS collaboration and the combination of results from the ATLAS collaboration [184] are shown.

Already with the small dataset of roughly 36 pb^{-1} , the top-quark pair production cross-section has been measured at a center-of-mass energy of 7 TeV. So far, all measurements are in good agreement with SM predictions. During the 2011 operation of the LHC, a dataset with more than 1 fb^{-1} has been collected up to July and it is expected that the complete 2011 dataset will be $2\text{-}3 \text{ fb}^{-1}$. This implies that at the end of this year first hints for the Higgs boson or physics which cannot be explained in the framework of the SM might show up. In the top-quark sector, especially the precise measurement of the charge asymmetry [162] and the search for heavy resonances decaying into top-quark pairs will be of particular importance. The basis for many of these measurements will be a precise determination of the top-quark pair production cross-section.

Appendix A

A.1 Dataset Names of CMS Fall10 Production

Process	CMS Dataset Name
$t\bar{t}$, inclusive	/TTJets_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v2/AODSIM
t-channel, $t \rightarrow b\bar{\nu}_l$	/TToBLNu_TuneZ2_t-channel_7TeV-madgraph/Fall10-START38_V12-v2/AODSIM
$t\bar{W}$ -channel, inclusive	/TToBLNu_TuneZ2_tW-channel_7TeV-madgraph/Fall10-START38_V12-v2/AODSIM
W +jets, $W \rightarrow l\nu$	/WJetsToLNu_TuneD6T_7TeV-madgraph-tauola/Fall10-START38_V12-v1/GEN-SIM-REC
Z/γ^* , $Z/\gamma^* \rightarrow l^+l^-$, $m(l^+l^-) > 50 \text{ GeV}/c^2$	/DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola/Fall10-START38_V12-v2/AODSIM
γ +jets, $40 \text{ GeV}/c \leq H_{T,\text{had}} < 100 \text{ GeV}/c$	/GJets_TuneD6T_HT-40to100_7TeV-madgraph/Fall10-START38_V12-v1/AODSIM
γ +jets, $100 \text{ GeV}/c \leq H_{T,\text{had}} < 200 \text{ GeV}/c$	/GJets_TuneD6T_HT-100to200_7TeV-madgraph/Fall10-START38_V12-v1/AODSIM
γ +jets, $H_{T,\text{had}} \geq 200 \text{ GeV}/c$	/GJets_TuneD6T_HT-200_7TeV-madgraph/Fall10-START38_V12-v1/AODSIM
BCtoE, $20 \text{ GeV}/c \leq \hat{p}_T < 30 \text{ GeV}/c$	/QCD_Pt-20to30_BCtoE_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
BCtoE, $30 \text{ GeV}/c \leq \hat{p}_T < 80 \text{ GeV}/c$	/QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
BCtoE, $80 \text{ GeV}/c \leq \hat{p}_T < 170 \text{ GeV}/c$	/QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
EMEnriched, $20 \text{ GeV}/c \leq \hat{p}_T < 30 \text{ GeV}/c$	/QCD_Pt-20to30_EMEnriched_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
EMEnriched, $30 \text{ GeV}/c \leq \hat{p}_T < 80 \text{ GeV}/c$	/QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
EMEnriched, $80 \text{ GeV}/c \leq \hat{p}_T < 170 \text{ GeV}/c$	/QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
inclusive QCD, $15 \text{ GeV}/c \leq \hat{p}_T < 30 \text{ GeV}/c$	/QCD_Pt-15to30_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
inclusive QCD, $30 \text{ GeV}/c \leq \hat{p}_T < 50 \text{ GeV}/c$	/QCD_Pt-30to50_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
inclusive QCD, $50 \text{ GeV}/c \leq \hat{p}_T < 80 \text{ GeV}/c$	/QCD_Pt-50to80_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
inclusive QCD, $80 \text{ GeV}/c \leq \hat{p}_T < 120 \text{ GeV}/c$	/QCD_Pt-80to120_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
inclusive QCD, $120 \text{ GeV}/c \leq \hat{p}_T < 170 \text{ GeV}/c$	/QCD_Pt-120to170_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM
inclusive QCD, $170 \text{ GeV}/c \leq \hat{p}_T < 300 \text{ GeV}/c$	/QCD_Pt-170to300_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/AODSIM

Table A.1: Official dataset names of the CMS Fall10 production.

A.2 Dataset Names of CMS Collision Data

Run Range	CMS Dataset Name
135821-144114	/EG/Run2010A-Nov4ReReco_v1/AOD
146240-149711	/Electron/Run2010B-Nov4ReReco_v1/AOD

Table A.2: Official data set names of the CMS 2010 Collision Data.

A.3 Same-Sign/Opposite-Sign Method

In order to derive the electron selection-efficiencies in a proper way, a background estimation technique is necessary. With the same-sign/opposite-sign method the number of true Z boson events can be estimated directly from data. The technique is based on the assumption, that the two electrons from all relevant background processes are produced with identical fractions with same charges and opposite charges. The number of events with opposite electron charge observed in data, can be written as [185]

$$\begin{aligned} N_{\text{OS}} &= N_{\text{OS}}^{\text{S}}(1 - q_1^{\text{misid}})(1 - q_2^{\text{misid}}) + N_{\text{OS}}^{\text{S}} \cdot q_1^{\text{misid}} \cdot q_2^{\text{misid}} \\ &+ N_{\text{OS}}^{\text{B}}(1 - q_1^{\text{misid}})(1 - q_2^{\text{misid}}) + N_{\text{OS}}^{\text{B}} \cdot q_1^{\text{misid}} \cdot q_2^{\text{misid}} \\ &+ N_{\text{SS}}^{\text{B}} [q_2^{\text{misid}}(1 - q_1^{\text{misid}}) + q_1^{\text{misid}}(1 - q_2^{\text{misid}})] . \end{aligned} \quad (\text{A.1})$$

In this equation, N_{OS}^{S} and N_{OS}^{B} denote the number of signal and background events with oppositely charged electrons, N_{SS}^{B} is the number of background events with equally charged electrons, and q_1^{misid} and q_2^{misid} are the charge-misidentification probabilities for the first and the second electron, respectively. Similarly, the number of events with equally charged electrons in data is given by

$$\begin{aligned} N_{\text{SS}} &= N_{\text{SS}}^{\text{B}}(1 - q_1^{\text{misid}})(1 - q_2^{\text{misid}}) + N_{\text{SS}}^{\text{B}} \cdot q_1^{\text{misid}} \cdot q_2^{\text{misid}} \\ &+ N_{\text{OS}}^{\text{B}} [q_2^{\text{misid}}(1 - q_1^{\text{misid}}) + q_1^{\text{misid}}(1 - q_2^{\text{misid}})] \\ &+ N_{\text{OS}}^{\text{S}} [q_2^{\text{misid}}(1 - q_1^{\text{misid}}) + q_1^{\text{misid}}(1 - q_2^{\text{misid}})] . \end{aligned} \quad (\text{A.2})$$

Using the assumption that the background is charge-symmetric, i.e. $N_{\text{OS}}^{\text{B}} = N_{\text{SS}}^{\text{B}}$, and subtracting equation A.2 from equation A.1,

$$N_{\text{OS}} - N_{\text{SS}} = N_{\text{OS}}^{\text{S}}(1 - 2 \cdot q_1^{\text{misid}})(1 - 2 \cdot q_2^{\text{misid}}) \quad (\text{A.3})$$

is obtained. Solving this equation for N_{OS}^{S} and averaging over all electrons, i.e. $q_1^{\text{misid}} = q_2^{\text{misid}}$, equation 4.10 can be derived:

$$N_{\text{OS}}^{\text{S}} \equiv N_Z = \frac{N_{\text{OS}} - N_{\text{SS}}}{(1 - 2q^{\text{misid}})^2} . \quad (\text{A.4})$$

Finally, an expression for the charge-misidentification probability has to be derived. For $Z \rightarrow e^+e^-$, the product of the signs of the electron and the positron charge, $q_1 \cdot q_2$, can be written as [185]

$$\begin{aligned} q_1 \cdot q_2 &= q_1^{\text{misid}} q_2^{\text{misid}}(-1) + (1 - q_1^{\text{misid}})q_2^{\text{misid}}(1) + (1 - q_2^{\text{misid}})q_1^{\text{misid}}(1) \\ &+ (1 - q_1^{\text{misid}})(1 - q_2^{\text{misid}})(-1) - (1 - 2 \cdot q_1^{\text{misid}})(1 - 2 \cdot q_2^{\text{misid}}) \end{aligned} \quad (\text{A.5})$$

where the numbers in brackets represent the sign of the product of both charges. Assuming that both, electron and positron exhibit an equal probability that their charge is misidentified, i.e. $q_1^{\text{misid}} = q_2^{\text{misid}}$, equation A.5 can be re-written as

$$4(q^{\text{misid}})^2 - 4q^{\text{misid}} + (1 + q_1 q_2) = 0 . \quad (\text{A.6})$$

This equation can only be solved if the product of the two charges is smaller than one. This is true if $q_1 q_2$ is replaced by the average product $\langle q_1 q_2 \rangle$ found in an almost background free Z boson dataset. In this case, the average misidentification probability is given by

$$q^{\text{misid}} = \frac{1 - \sqrt{|\langle q_1 q_2 \rangle|}}{2}. \quad (\text{A.7})$$

A.4 Datasets with Systematic Variations

Process	variation	N_{gen}	Dataset Name
$t\bar{t}$	scale \times 0.5	1,098,971	/TTJets_TuneD6T_scaledown_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	scale \times 2.0	1,153,236	/TTJets_TuneD6T_scaleup_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	thresh. \times 0.5	938,005	/TTJets_TuneD6T_matchingdown_7TeV-madgraph-tauola/Fall10-START38_V12-v2
	thresh. \times 2.0	1,036,492	/TTJets_TuneD6T_matchingup_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	less ISR/FSR	1,221,664	/TTJets_TuneD6T_smallerISRFSR_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	more ISR/FSR	1,394,010	/TTJets_TuneD6T_largerISRFSR_7TeV-madgraph-tauola/Fall10-START38_V12-v2
	with pile-up	1,281,237	/TTJets_TuneD6T_7TeV-madgraph-tauola/Fall10-E7TeV_ProbDist_2010Data_BX156_START38_V12-v1
W +jets	scale \times 0.5	5,042,219	/WJets_TuneD6T_scaledown_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	scale \times 2.0	6,218,255	/WJets_TuneD6T_scaleup_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	thresh. \times 0.5	2,706,986	/WJets_TuneD6T_matchingdown_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	thresh. \times 2.0	10,370,368	/WJets_TuneD6T_matchingup_7TeV-madgraph-tauola/Fall10-START38_V12-v1
	with pile-up	14,766,396	/WJetsToLLNu_TuneD6T_7TeV-madgraph-tauola/Fall10-E7TeV_ProbDist_2010Data_BX156_START38_V12-v1
Z/γ^* +jets	scale \times 0.5	1,436,150	/DYJetsToLL_TuneD6T_scaledown_7TeV-madgraph-tauola/Fall10-START38_V12-v2
	scale \times 2.0	1,329,028	/DYJetsToLL_TuneD6T_scaleup_7TeV-madgraph-tauola/Fall10-START38_V12-v2
	thresh. \times 0.5	1,662,884	/DYJetsToLL_TuneD6T_matchingdown_7TeV-madgraph-tauola/Fall10-START38_V12-v2
	thresh. \times 2.0	1,667,367	/DYJetsToLL_TuneD6T_matchingup_7TeV-madgraph-tauola/Fall10-START38_V12-v1

Table A.3: Details on MC samples with systematic variations. The number of produced events as well as the official CMS dataset names are given.

List of Figures

1.1	Example for Feynman Diagram	13
1.2	CTEQ6L1 PDF	15
1.3	Feynman diagrams for LO $t\bar{t}$ production	16
1.4	Feynman diagrams for LO single top-quark production	18
1.5	Probabilities for different top-quark pair decay channels	21
2.1	Aerial view of the Geneva countryside	24
2.2	Schematic view of CERN's accelerator chain	25
2.3	Layout of the LHC main ring	27
2.4	Instantaneous and integrated luminosity in 2010 proton operation	30
2.5	Layout of the CMS detector	31
2.6	Illustration of the CMS tracking system	33
2.7	Overview of the CMS ECAL	35
2.8	Overview of the CMS HCAL	37
2.9	Layout of the CMS Muon system	39
2.10	Architecture of the CMS DAQ system	40
2.11	Overview of the WLCG and the CMS Computing Model	42
3.1	Event Generation Process	45
3.2	Illustration of the Isolation Cone	55
3.3	Typical Jet Evolution	61
3.4	Uncertainties on the Jet Energy Correction Factors	62
4.1	Cross-Sections of Different Physics Processes	66
4.2	Top-quark Pair Decay to Electron and Jets	67
4.3	Feynman Diagrams for W Boson and Drell-Yan Production	68
4.4	Feynman Diagrams for Photon and Single Top Production	69
4.5	Feynman Diagrams for QCD Multi-Jet Production	70
4.6	Transverse Impact Parameter with Respect to the Beam Spot	80
4.7	Electron Identification Observables in the ECal Barrel	81
4.8	Electron Identification Observables in the ECal Endcap	82
4.9	Electron Isolation Observables	84
4.10	Number of Missing Hits Distribution	85
4.11	Variables for Partner Track Conversion Rejection	86

4.12	Jet Multiplicity Distribution	87
4.13	Composition of Selected $t\bar{t}$ Events	89
4.14	Differential Distributions of C_{acc}	96
4.15	Differential Distributions of C_{id}	97
4.16	Differential Distributions of C_{id}	98
4.17	Jet Multiplicity Distribution with Data	101
4.18	Kinematic Observables for $N_{\text{jets}} = 3$	103
4.19	Kinematic Observables for $N_{\text{jets}} \geq 4$	104
5.1	QCD/ γ +jets Sideband for Data and Simulation	109
5.2	Discriminating Variables	110
5.3	$\beta_{t\bar{t}}^{\text{fit}}$ Distribution and Neyman Construction	114
5.4	Template Shapes for JER Variations	116
5.5	Template Shapes for JES Variations	118
5.6	Template Shapes for Electron Energy Scale Variations	120
5.7	Template Shapes for Unclustered Energy Variations	121
5.8	Template Shapes for Different QCD/ γ +jets Sidebands	122
5.9	Template Shapes for μ_F/μ_R Variations	124
5.10	Template Shapes for Matching Threshold Variations	126
5.11	Template Shapes With Pile-Up Included	127
5.12	Template Shapes for ISR/FSR Variations	128
5.13	Template Shapes for ISR/FSR Variations	128
5.14	$\beta_{t\bar{t}}^{\text{fit}}$ Distribution and Neyman Construction Including Systematic Uncertainties	130
5.15	Correlation Matrix and Final Neyman Construction	132
5.16	Kinematic Variables for the Comparison of Data and Fit Result	134
5.17	Correlation Matrix and Final Neyman Construction	138
5.18	Overview of Different $t\bar{t}$ Production Cross-Section Measurements	142

List of Tables

1.1	The Fundamental Fermions	2
1.2	The Standard Model Gauge Bosons	3
2.1	LHC machine parameters	30
4.1	Details on the Generation of $t\bar{t}$ and Single Top-Quark Events	71
4.2	Details on the Generation of W +Jets and Drell-Yan+Jets Events	72
4.3	Details on the Generation of γ +Jets and QCD Multi-Jets Events	73
4.4	Run-Ranges for Different Electron Triggers	76
4.5	Pixel Matching Windows for the HLT	76
4.6	Electron Identification at HLT	78
4.7	Electron ID Working Points	83
4.8	Efficiencies of Selection Criteria	88
4.9	Expected Composition of the Tag-and-Probe Dataset	91
4.10	Charge Misidentification Probabilities	93
4.11	Scaled Charge Misidentification Probabilities	94
4.12	Electron Selection Efficiencies from Simulation	95
4.13	Electron Trigger-Efficiencies	99
4.14	Number of Expected Events per Process	100
4.15	Number of Events in Data and Simulation	101
5.1	Relative Event Yield Changes for JER Variations	117
5.2	Relative Event Yield Changes for JES Variations	119
5.3	Relative Event Yield Changes for Electron Energy Scale Variations	119
5.4	Relative Event Yield Changes due to μ_F/μ_R Variations	123
5.5	Relative Event Yield Changes due to Matching Threshold Variations	125
5.6	Relative Event Yield Changes due to Pile-Up Events	126
5.7	Maximum Relative Event Yield Changes for PDF Variations	129
5.8	Systematic Uncertainties on the Measurement	131
5.9	Results of the Maximum Likelihood Fit to Data	131
5.10	Number of Expected Events in the Muon+Jets Dataset	135
5.11	Results of the Maximum Likelihood Fit to Data in the Muon+Jets Analysis	136
5.12	Systematic Uncertainties on the Combined Measurement	137

5.13 Results of the Maximum Likelihood Fit to the Combined Dataset . . .	138
A.1 Dataset Names of CMS Fall10 Production	143
A.2 Dataset Names of CMS 2010 Collision Data	143
A.3 Details on MC Samples with Systematic Variations	146

Bibliography

- [1] G. Arnison et al., “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $s^{1/2} = 540$ GeV”, Phys. Lett. B122, 103–116 (1983).
- [2] M. Banner et al., “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider”, Phys. Lett. B122, 476–485 (1983).
- [3] G. Arnison et al., “Experimental Observation of Lepton Pairs of Invariant Mass around 95 GeV/c² at the CERN SPS Collider”, Phys. Lett. B126, 398–410 (1983).
- [4] P. Bagnaia et al., “Evidence for $Z^0 \rightarrow e^+ e^-$ at the CERN anti-p p Collider”, Phys. Lett. B129, 130–140 (1983).
- [5] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, Phys. Rev. Lett. 13, 321–322 (1964).
- [6] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, Phys. Rev. Lett. 13, 508–509 (1964).
- [7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, Phys. Rev. Lett. 13, 585–587 (1964).
- [8] CDF and D0, “Combination of CDF and D0 Results on the Mass of the Top Quark”, [arXiv:hep-ex/1007.3178](https://arxiv.org/abs/hep-ex/1007.3178) (2010).
- [9] F. Abe et al., “Observation of Top Quark Production in $\bar{p}p$ Collisions”, Phys. Rev. Lett. 74, 2626–2631 (1995).
- [10] S. Abachi et al., “Observation of the Top Quark”, Phys.Rev.Lett. 74, 2632–2637 (1995).
- [11] “MCFM Users Guide.” <http://mcfm.fnal.gov/mcfm.pdf> (March 2011).
- [12] M. Aliev et al., “– HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR”, Comput. Phys. Commun. 182, 1034–1046 (2011).

-
- [13] J. Alwall et al., “MadGraph/MadEvent v4: The New Web Generation”, JHEP 09, 028 (2007).
- [14] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, JHEP 05, 026 (2006).
- [15] J. Gruschke, “Nachweis von Top-Quarks und erste Messung des $t\bar{t}$ -Wirkungsquerschnittes bei einer Schwerpunktsenergie von 7 TeV mit dem CMS-Experiment am LHC”, IEKP-KA/2011-4 (2011).
- [16] M. Gell-Mann, “A Schematic Model of Baryons and Mesons”, Phys. Lett. 8, 214–215 (1964).
- [17] G. Zweig, “An SU(3) Model for Strong Interaction Symmetry and its Breaking”, CERN-TH-401 (1964).
- [18] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories”, Phys. Rev. Lett. 30, 1343–1346 (1973).
- [19] D. J. Gross and F. Wilczek, “Asymptotically Free Gauge Theories. 1”, Phys. Rev. D8, 3633–3652 (1973).
- [20] D. J. Gross and F. Wilczek, “Asymptotically Free Gauge Theories. 2”, Phys. Rev. D9, 980–993 (1974).
- [21] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, Phys. Rev. Lett. 30, 1346–1349 (1973).
- [22] H. D. Politzer, “Asymptotic Freedom: An Approach to Strong Interactions”, Phys. Rept. 14, 129–180 (1974).
- [23] G. 't Hooft and M. J. G. Veltman, “Regularization and Renormalization of Gauge Fields”, Nucl. Phys. B44, 189–213 (1972).
- [24] G. 't Hooft and M. J. G. Veltman, “Combinatorics of Gauge Fields”, Nucl. Phys. B50, 318–353 (1972).
- [25] A. Salam and J. C. Ward, “Weak and Electromagnetic Interactions”, Nuovo Cim. 11, 568–577 (1959).
- [26] S. Weinberg, “A Model of Leptons”, Phys. Rev. Lett. 19, 1264–1266 (1967).
- [27] S. L. Glashow, “Partial Symmetries of Weak Interactions”, Nucl. Phys. 22, 579–588 (1961).
- [28] A. Einstein, “Die Grundlage der Allgemeinen Relativitätstheorie”, Annalen der Physik 354, 769–822 (1916).
- [29] K. Nakamura et al., “Review of Particle Physics”, J. Phys. G37, 075021 (2010).

-
- [30] W. Pauli, “Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren”, *Zeitschriften für Physik* 31, p. 765–783 (1925).
- [31] Y. Fukuda et al., “Evidence for Oscillation of Atmospheric Neutrinos”, *Phys. Rev. Lett.* 81, 1562–1567 (1998).
- [32] Q. R. Ahmad et al., “Measurement of the Charged Current Interactions Produced by B_8 Solar Neutrinos at the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* 87, 071301 (2001).
- [33] T. Araki et al., “Measurement of Neutrino Oscillation with KamLAND: Evidence of Spectral Distortion”, *Phys. Rev. Lett.* 94, 081801 (2005).
- [34] P. Minkowski, “ $\text{Mu} \rightarrow \text{E Gamma}$ at a Rate of One Out of 1-Billion Muon Decays?”, *Phys. Lett.* B67, 421 (1977).
- [35] J. Schechter and J. W. F. Valle, “Neutrino Masses in $\text{SU}(2) \times \text{U}(1)$ Theories”, *Phys. Rev.* D22, 2227 (1980).
- [36] S. Hannestad, “Neutrino Physics From Precision Cosmology”, *Prog. Part. Nucl. Phys.* 65, 185–208 (2010).
- [37] M. Y. Han and Y. Nambu, “Three-triplet model with double $\text{SU}(3)$ symmetry”, *Phys. Rev.* 139, B1006–B1010 (1965).
- [38] W. A. Bardeen, H. Fritzsch, and M. Gell-Mann, “Light-cone current algebra, π^0 decay, and $e^+ e^-$ annihilation”, [arXiv:hep-ph/0211388](https://arxiv.org/abs/hep-ph/0211388) (1972).
- [39] S. Tomonaga, “On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields”, *Prog. Theor. Phys.* 1, 27–42 (1946).
- [40] J. S. Schwinger, “Quantum Electrodynamics. III: The Electromagnetic Properties of the Electron: Radiative Corrections to Scattering”, *Phys. Rev.* 76, 790–817 (1949).
- [41] R. P. Feynman, “Space-time Approach to Quantum Electrodynamics”, *Phys. Rev.* 76, 769–789 (1949).
- [42] F. J. Dyson, “The Radiation Theories of Tomonaga, Schwinger, and Feynman”, *Phys. Rev.* 75, 486–502 (1949).
- [43] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* 10, 531–533 (1963).
- [44] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* 49, 652–657 (1973).

-
- [45] R. Barate et al., “Search for the Standard Model Higgs Boson at LEP”, *Phys. Lett. B* 565, 61–75 (2003).
- [46] CDF and D0, “Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-1} of Data”, [arXiv:hep-ex/1007.4587](https://arxiv.org/abs/hep-ex/1007.4587) (2010).
- [47] W. Gordon, “Der Comptoneffekt nach der Schrödingerschen Theorie”, *Zeitschrift für Physik A: Hadrons and Nuclei* 117–133 (1926).
- [48] O. Klein, “Elektrodynamik und Wellenmechanik vom Standpunkt des Korrespondenzprinzips”, *Zeitschrift für Physik A — Hadrons and Nuclei* 407–442 (1927).
- [49] P. A. M. Dirac, “The Quantum Theory of the Electron”, *Proc. Roy. Soc. Lond. A* 117, 610–624 (1928).
- [50] T. Nakano and K. Nishijima, “Charge Independence for V-particles”, *Prog. Theor. Phys.* 10, 581–582 (1953).
- [51] P. A. M. Dirac, “Quantum Theory of Emission and Absorption of Radiation”, *Proc. Roy. Soc. Lond. A* 114, 243 (1927).
- [52] S. Mandelstam, “Determination of the Pion - Nucleon Scattering Amplitude from Dispersion Relations and Unitarity. General theory”, *Phys. Rev.* 112, 1344–1360 (1958).
- [53] F. D. Aaron et al., “Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA”, *JHEP* 01, 109 (2010).
- [54] J. Pumplin et al., “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”, *JHEP* 07, 012 (2002).
- [55] S. W. Herb et al., “Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton - Nucleus Collisions”, *Phys. Rev. Lett.* 39, 252–255 (1977).
- [56] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, and SLD Heavy Flavour Group, “Precision Electroweak Measurements on the Z Resonance”, *Phys.Rept.* 427, 257–454 (2006).
- [57] R. D. Peccei, S. Peris, and X. Zhang, “Nonstandard Couplings of the Top Quark and Precision Measurements of the Electroweak Theory”, *Nucl. Phys. B* 349, 305–322 (1991).
- [58] J. H. Kühn, “Theory of Top Quark Production and Decay”, [arXiv:hep-ph/9707321](https://arxiv.org/abs/hep-ph/9707321) (1996).

-
- [59] W. Wagner, “Top-Antitop-Quark Production and Decay Properties at the Tevatron”, *Mod. Phys. Lett. A*25, 1297–1314 (2010).
- [60] T. Aaltonen et al., “Exclusion of an Exotic Top Quark with $-4/3$ Electric Charge Using Soft Lepton Tagging”, *Phys. Rev. Lett.* 105, 101801 (2010).
- [61] T. Aaltonen et al., “Evidence for a Mass Dependent Forward-Backward Asymmetry in Top Quark Pair Production”, [arXiv:hep-ex/1101.0034](https://arxiv.org/abs/hep-ex/1101.0034) (2011).
- [62] J. R. Incandela, A. Quadt, W. Wagner, and D. Wicke, “Status and Prospects of Top-Quark Physics”, *Prog. Part. Nucl. Phys.* 63, 239–292 (2009).
- [63] V. Barger and R. Phillips, “Collider Physics”. *Frontiers in physics*. Addison-Wesley, Redwood City, Calif. [u.a.], 1987.
- [64] P. Nason, S. Dawson, and R. K. Ellis, “The Total Cross-Section for the Production of Heavy Quarks in Hadronic Collisions”, *Nucl. Phys. B*303, 607 (1988).
- [65] W. Beenakker, W. L. van Neerven, R. Meng, G. A. Schuler, and J. Smith, “QCD Corrections to Heavy Quark Production in Hadron Hadron Collisions”, *Nucl. Phys. B*351, 507–560 (1991).
- [66] N. Kidonakis, “Next-To-Next-To-Leading Soft-Gluon Corrections for the Top Quark Cross Section and Transverse Momentum Distribution”, *Phys. Rev. D*82, 114030 (2010).
- [67] U. Langenfeld, S. Moch, and P. Uwer, “Measuring the Running Top-Quark Mass”, *Phys. Rev. D*80, 054009 (2009).
- [68] H. Mellin, “Abriß einer einheitlichen Theorie der Gamma- und der hypergeometrischen Funktionen”, *Mathematische Annalen* 305–337 (1910).
- [69] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton Distributions for the LHC”, *Eur. Phys. J. C*63, 189–285 (2009).
- [70] T. Aaltonen et al., “First Observation of Electroweak Single Top Quark Production”, *Phys. Rev. Lett.* 103, 092002 (2009).
- [71] V. M. Abazov et al., “Observation of Single Top-Quark Production”, *Phys. Rev. Lett.* 103, 092001 (2009).
- [72] N. Kidonakis, “NNLL Resummation for s-Channel Single Top Quark Production”, *Phys. Rev. D*81, 054028 (2010).
- [73] J. M. Campbell and F. Tramontano, “Next-To-Leading Order Corrections to Wt Production and Decay”, *Nucl. Phys. B*726, 109–130 (2005).

-
- [74] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, “Single-Top Hadroproduction in Association with a W Boson”, JHEP 07, 029 (2008).
- [75] N. Kidonakis, “Two-Loop Soft Anomalous Dimensions for Single Top Quark Associated Production with a W^- or H^- ”, Phys. Rev. D82, 054018 (2010).
- [76] A. Einstein, “Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?”, Annalen der Physik 18, p. 639–643 (1905).
- [77] L. Evans and P. Bryant, “LHC Machine”, JINST 3, S08001 (2008).
- [78] “CERN Aerial View.” <http://visits.web.cern.ch/visits/guides/tools/presentation/vite/images/aerial-view.jpg> (February 2011).
- [79] K. Aamodt et al., “The ALICE Experiment at the CERN LHC”, JINST 3, S08002 (2008).
- [80] G. Aad et al., “The ATLAS Experiment at the CERN Large Hadron Collider”, JINST 3, S08003 (2008).
- [81] “CMS Physics Technical Design Report Volume I : Detector Performance and Software”, CERN-LHCC-2006-001 (2006).
- [82] R. Adolphi et al., “The CMS Experiment at the CERN LHC”, JINST 3, S08004 (2008).
- [83] A. A. Alves et al., “The LHCb Detector at the LHC”, JINST 3, S08005 (2008).
- [84] O. Adriani et al., “The LHCf Detector at the CERN Large Hadron Collider”, JINST 3, S08006 (2008).
- [85] G. Anelli et al., “The TOTEM Experiment at the CERN Large Hadron Collider”, JINST 3, S08007 (2008).
- [86] K. Schindl, “The PS Booster as Pre-Injector for LHC”, Part. Accel. 58, 63–78 (1997).
- [87] R. Cappi, “The PS in the LHC Injector Chain”, Part. Accel. 58, 79–89 (1997).
- [88] T. Linnecar, “Preparing the SPS for LHC”, Part. Accel. 58, 91–101 (1997).
- [89] R. Geller, “Ecris: The Electron Cyclotron Resonance Ion Sources”, Ann. Rev. Nucl. Part. Sci. 40, 15–44 (1990).
- [90] M. Chanel, “LEIR: The Low Energy Ion Ring at CERN”, Nucl. Instrum. Meth. A532, 137–143 (2004).

-
- [91] “Interim Summary on the Analysis of the 19 September 2008 Incident at the LHC.” https://edms.cern.ch/file/973073/1/Report_on_080919_incident_at_LHC__2_.pdf (February 2011).
- [92] CMS Collaboration, “The Compact Muon Solenoid Experiment: How was CMS designed.” <http://cms.web.cern.ch/cms/Detector/Designed/index.html> (February 2011).
- [93] CMS Collaboration, “The CMS Tracker System Project: Technical Design Report”, CERN-LHCC-98-006 (1998).
- [94] CMS Collaboration, “The CMS Tracker: Addendum to the Technical Design Report”, CERN-LHCC-2000-016 (2000).
- [95] S. Chatrchyan et al., “Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons”, JINST 5, T03007 (2010).
- [96] V. Khachatryan et al., “CMS Tracking Performance Results From Early LHC Operation”, Eur. Phys. J. C70, 1165–1192 (2010).
- [97] CMS Collaboration, “The CMS Electromagnetic Calorimeter Project : Technical Design Report”, CERN-LHCC-97-033 (1997).
- [98] P. Lecoq et al., “Lead Tungstate (PbWO_4) Scintillators for LHC EM calorimetry”, Nucl. Instrum. Meth. A365, 291–298 (1995).
- [99] P. Adzic et al., “Energy Resolution of the Barrel of the CMS Electromagnetic Calorimeter”, JINST 2, P04004 (2007).
- [100] CMS Collaboration, “The CMS Hadron Calorimeter Project : Technical Design Report”, CERN-LHCC-97-031 (1997).
- [101] S. Chatrchyan et al., “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data”, JINST 5, T03012 (2010).
- [102] CMS Collaboration, “The CMS Muon project : Technical Design Report”, CERN-LHCC-97-032 (1997).
- [103] G. Charpak and F. Sauli, “High-Accuracy, Two-Dimensional Read-Out in Multiwire Proportional Chambers”, Nucl. Instrum. Meth. 113, 381–385 (1973).
- [104] S. Chatrchyan et al., “Performance of the CMS Drift Tube Chambers with Cosmic Rays”, JINST 5, T03015 (2010).
- [105] S. Chatrchyan et al., “Performance of the CMS Cathode Strip Chambers with Cosmic Rays”, JINST 5, T03018 (2010).
- [106] S. Chatrchyan et al., “Performance Study of the CMS Barrel Resistive Plate Chambers with Cosmic Rays”, JINST 5, T03017 (2010).

-
- [107] CMS Collaboration, “The TriDAS Project, Technical Design Report. Volume 1: The Level-1 Trigger”, CERN-LHCC-2000-038 (2000).
- [108] CMS Collaboration, “The TriDAS Project, Technical Design Report. Volume 2: Data Acquisition and High-Level Trigger Technical Design Report”, CERN-LHCC-2002-026 (2002).
- [109] CMS Collaboration, “LHC Computing Grid : Technical Design Report”, CERN-LHCC-2005-024 (2005).
- [110] CMS Collaboration, “CMS Computing : Technical Design Report”, CERN-LHCC-2005-023 (2005).
- [111] N. Metropolis and S. Ulam, “The Monte Carlo Method”, *Journal of the American Statistical Association* 44, 335–341 (1949).
- [112] R. Eckhardt, “Stan Ulam, John von Neumann, and the Monte Carlo method”, *Los Alamos Science, Special Issue* 15, 131–137 (1987).
- [113] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, “QCD Matrix Elements + Parton Showers”, *JHEP* 11, 063 (2001).
- [114] F. Krauss, “Matrix Elements and Parton Showers in Hadronic Interactions”, *JHEP* 08, 015 (2002).
- [115] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching Matrix Elements and Shower Evolution for Top-Quark Production in Hadronic Collisions”, *JHEP* 01, 013 (2007).
- [116] S. Hoeche et al., “Matching Parton Showers and Matrix Elements”, [arXiv:hep-ph/0602031](https://arxiv.org/abs/hep-ph/0602031) (2006).
- [117] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* 97, 31–145 (1983).
- [118] B. Andersson, “The Lund Model”, Cambridge University Press (1998).
- [119] B. R. Webber, “A QCD Model for Jet Fragmentation Including Soft Gluon Interference”, *Nucl. Phys.* B238, 492 (1984).
- [120] B. Andersson, G. Gustafson, and T. Sjöstrand, “Baryon Production in Jet Fragmentation and Upsilon Decay”, *Phys. Scripta* 32, 574 (1985).
- [121] D. Amati and G. Veneziano, “Preconfinement as a Property of Perturbative QCD”, *Phys. Lett.* B83, 87 (1979).
- [122] G. Corcella et al., “HERWIG 6.5 release note”, [arXiv:hep-ph/0210213](https://arxiv.org/abs/hep-ph/0210213) (2002).

-
- [123] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, “ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions”, *JHEP* 07, 001 (2003).
- [124] S. Frixione and B. R. Webber, “Matching NLO QCD Computations and Parton Shower Simulations”, *JHEP* 06, 029 (2002).
- [125] 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, 043 (2010).
- [126] F. Maltoni and T. Stelzer, “MadEvent: Automatic Event Generation with MadGraph”, *JHEP* 02, 027 (2003).
- [127] T. Stelzer and W. F. Long, “Automatic Generation of Tree Level Helicity Amplitudes”, *Comput. Phys. Commun.* 81, 357–371 (1994).
- [128] J. Alwall et al., “A Standard Format for Les Houches Event Files”, *Comput. Phys. Commun.* 176, 300–304 (2007).
- [129] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys.* B126, 298 (1977).
- [130] V. N. Gribov and L. N. Lipatov, “Deep Inelastic Electron Scattering in Perturbation Theory”, *Phys. Lett.* B37, 78–80 (1971).
- [131] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics”, *Sov. Phys. JETP* 46, 641–653 (1977).
- [132] R. Field, “Studying the Underlying Event at CDF and the LHC.”
http://www.phys.ufl.edu/~rfield/cdf/Toronto_RickField_3-18-08.pdf
(April 2011).
- [133] R. Field, “Early LHC Underlying Event Data - Findings and Surprises”,
[arXiv:hep-ph/1010.3558](https://arxiv.org/abs/hep-ph/1010.3558) (2010).
- [134] S. Jadach, J. H. Kühn, and Z. Was, “TAUOLA: A Library of Monte Carlo Programs to Simulate Decays of Polarized Tau Leptons”, *Comput. Phys. Commun.* 64, 275–299 (1990).
- [135] S. Agostinelli et al., “GEANT4: A Simulation Toolkit”, *Nucl. Instrum. Meth.* A506, 250–303 (2003).
- [136] CMS Collaboration, “Fast Simulation of the CMS Detector at the LHC”,
CERN-CMS-CR-2010-297 (2010).

-
- [137] CMS Collaboration, “Comparison of the Fast Simulation of CMS With the First LHC Data”, CERN-CMS-DP-2010-039 (2010).
- [138] CMS Collaboration, “Track Reconstruction in the CMS Tracker”, CERN-CMS-NOTE-2006-041 (2006).
- [139] R. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *Transaction of the ASME, Journal of Basic Engineering* 35–45 (1960).
- [140] R. Frühwirth, “Application of Kalman Filtering to Track and Vertex Fitting”, *Nucl. Instrum. Meth. A262*, 444–450 (1987).
- [141] CMS Collaboration, “Track Reconstruction, Primary Vertex Finding and Seed Generation with the Pixel Detector”, CERN-CMS-NOTE-2006-026 (2006).
- [142] CMS Collaboration, “Tracking and Primary Vertex Results in First 7 TeV Collisions”, CMS-PAS-TRK-10-005 (2010).
- [143] CMS Collaboration, “Track Parameter Evaluation and Primary Vertex Finding with the Pixel Detector”, CERN-CMS-NOTE-2006-026 (2006).
- [144] CMS Collaboration, “Adaptive Vertex Fitting”, CERN-CMS-NOTE-2007-008 (2007).
- [145] S. Baffioni et al., “Electron Reconstruction in CMS”, *Eur. Phys. J. C49*, 1099–1116 (2007).
- [146] CMS Collaboration, “Electron Reconstruction in the CMS Electromagnetic Calorimeter”, CERN-CMS-NOTE-2001-034 (2001).
- [147] H. Bethe and W. Heitler, “On the Stopping of Fast Particles and on the Creation of Positive Electrons”, *Proc. Roy. Soc. Lond. A146*, 83–112 (1934).
- [148] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at LHC”, *ECONF C0303241, TULT009* (2003).
- [149] CMS Collaboration, “CMS Physics Technical Design Report Volume I : Detector Performance and Software”, CERN-LHCC-2006-001 (2006).
- [150] CMS Collaboration, “The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS”, CMS-PAS-JME-09-002 (2009).
- [151] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS-PAS-PFT-09-001 (2009).
- [152] G. P. Salam and G. Soyez, “A Practical Seedless Infrared-Safe Cone Jet Algorithm”, *JHEP* 05, 086 (2007).

-
- [153] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, “Longitudinally Invariant k_T Clustering Algorithms for Hadron Hadron Collisions”, Nucl. Phys. B406, 187–224 (1993).
- [154] S. D. Ellis and D. E. Soper, “Successive Combination Jet Algorithm for Hadron Collisions”, Phys. Rev. D48, 3160–3166 (1993).
- [155] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti- k_T Jet Clustering Algorithm”, JHEP 04, 063 (2008).
- [156] CMS Collaboration, “Plans for Jet Energy Corrections at CMS”, CMS-PAS-JME-07-002 (2007).
- [157] CMS Collaboration, “Determination of the Jet Energy Scale in CMS with pp Collisions at $\sqrt{s} = 7$ TeV”, CMS-PAS-JME-10-010 (2010).
- [158] CMS Collaboration, “Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV ”, CMS-PAS-JME-10-014 (2010).
- [159] CMS Collaboration, “Algorithms for b Jet Identification in CMS”, CMS-PAS-BTV-09-001 (2009).
- [160] CMS Collaboration, “Commissioning of b Jet Identification with pp Collisions at $\sqrt{s} = 7$ TeV”, CMS-PAS-BTV-10-001 (2010).
- [161] CMS Collaboration, “Missing Transverse Energy Performance in Minimum-Bias and Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV”, CMS-PAS-JME-10-004 (2010).
- [162] CMS Collaboration, “Measurement of the Charge Asymmetry in Top Quark Pair Production with the CMS Experiment”, CMS-PAS-TOP-10-010 (2010).
- [163] CMS Collaboration, “Plans for an Early Measurement of the $t\bar{t}$ Cross Section in the Electron+Jets Channel at $\sqrt{s} = 10$ TeV”, CMS-PAS-TOP-09-004 (2009).
- [164] T. Chwalek, “Messung der W-Boson-Helizitätsanteile in Top-Quark-Zerfällen mit dem CDF II Experiment und Studien zu einer frühen Messung des $t\bar{t}$ -Wirkungsquerschnitts mit dem CMS Experiment”, IEKP-KA/2010-5 (2010).
- [165] J. M. Campbell, J. W. Huston, and W. J. Stirling, “Hard Interactions of Quarks and Gluons: A Primer for LHC Physics”, Rept. Prog. Phys. 70, 89 (2007).
- [166] S. D. Drell and T.-M. Yan, “Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies”, Phys. Rev. Lett. 25, 316–320 (1970).
- [167] J. Bauer, “Perspektiven zur Beobachtung der elektroschwachen Produktion einzelner Top-Quarks mit dem CMS-Experiment”, IEKP-KA/2010-8 (2010).

-
- [168] K. Melnikov and F. Petriello, “Electroweak Gauge Boson Production at Hadron Colliders Through $O(\alpha_s^2)$ ”, *Phys. Rev. D* 74, 114017 (2006).
- [169] CMS Collaboration, “Measurement of CMS Luminosity”, CMS-PAS-EWK-10-004 (2010).
- [170] S. van der Meer, “Calibration of the Effective Beam Height in the ISR”, CERN-ISR-PO-68-31 (1968).
- [171] CMS Collaboration, “Absolute Luminosity Normalization”, CMS-DP-2011-002 (2011).
- [172] L. Tuura, A. Meyer, I. Segoni, and G. Della Ricca, “CMS Data Quality Monitoring: Systems and Experiences”, *J. Phys. Conf. Ser.* 219, 072020 (2010).
- [173] C. Clopper and S. Pearson, “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial”, *Biometrika* 26, 404–413 (1934).
- [174] J. Berryhill et al., “Electron Efficiency Measurements with 2.88/pb of pp Collision Data at $\sqrt{s} = 7$ TeV”, Internal CMS Analysis Note CMS-AN-2010/349 (2010).
- [175] V. Khachatryan et al., “Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”, *JHEP* 01, 080 (2011).
- [176] J. Neyman, “Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability”, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236, no. 767, 333-380, (1937).
- [177] Th. Müller, J. Ott and J. Wagner-Kuhr, “theta – A Framework for Template-Based Modeling and Inference.”
<http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/files/theta.pdf>
(May 2011).
- [178] CMS Collaboration, “Measurement of the Single-Top t-Channel Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV”, CMS-PAS-TOP-10-008 (2011).
- [179] CMS Collaboration, “Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”, CMS-PAS-EWK-10-005 (2011).
- [180] CMS Collaboration, “Rates of Jets Produced in Association with W and Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV”, CMS-PAS-EWK-10-012 (2011).
- [181] P. M. Nadolsky et al., “Implications of CTEQ Global Analysis for Collider Observables”, *Phys. Rev. D* 78, 013004 (2008).

-
- [182] M. R. Whalley, D. Bourilkov, and R. C. Group, “The Les Houches Accord PDFs (LHAPDF) and LHAGLUE”, [arXiv:hep-ph/0508110](https://arxiv.org/abs/hep-ph/0508110) (2005).
- [183] CMS Collaboration, “Combination of Top Pair Production Cross Sections in pp Collisions at 7 TeV and Comparisons with Theory”, CMS-PAS-TOP-11-001 (2011).
- [184] ATLAS Collaboration, “A Combined Measurement of the Top Quark Pair Production Cross-Section Using Dilepton and Single-Lepton Final States”, ATLAS-CONF-2011-040 (2011).
- [185] D. Bandurin et al., “Data Driven Techniques to Estimate the Background in the $Z \rightarrow ee$ Events”, Internal CMS Analysis Note CMS-AN-2010/277 (2010).

Danksagung

Am Ende dreier sehr interessanter aber auch arbeitsintensiver Jahre möchte ich mich herzlichst bei allen Personen bedanken, die maßgeblich zum erfolgreichen Gelingen dieser Arbeit beigetragen haben.

Ich danke Herrn Prof. Dr. Thomas Müller, der mir nach Abschluss meiner Diplomarbeit die Möglichkeit gegeben hat meine Doktorarbeit in seiner Arbeitsgruppe anzufertigen und der mich in den vergangenen drei Jahren durch hilfreiche Kommentare angeleitet hat. Ebenfalls hat er mir ermöglicht ein Jahr meiner Doktorandenzeit am CERN zu verbringen in dem ich wichtige Erfahrungen sammeln konnte, die mich sowohl fachlich als auch persönlich voran gebracht haben.

Für die Übernahme des Korreferats möchte ich mich gerne bei Herrn Prof. Dr. Günter Quast bedanken, der sich trotz zweier eigener Doktoranden mit ähnlichen Abgabeterminen die Zeit genommen hat meine Arbeit zu begutachten.

Des Weiteren gilt meiner großer Dank allen Mitgliedern meiner Arbeitsgruppe für die hilfreichen Vorschläge und Anregungen, das äußerst kollegiale Verhältnis sowie die stets gute Stimmung. Insbesondere danke ich Dr. Thorsten Chwalek und meiner „Tischnachbarin“ Dr. Jasmin Gruschke mit denen ich intensiv zusammenarbeiten durfte und auf deren Unterstützung ich mich immer verlassen konnte. Ebenso geht ein herzliches Dankeschön an meinen langjährigen Mitstreiter Dr. Thomas Peiffer, der mich vom Beginn des Studiums bis zum finalen Lernen auf die Promotionsprüfung begleitet hat. Meinem Zimmerkollegen Dr. Jan Lück danke ich für seine vielen nützlichen Ratschläge, die er aufgrund seiner großen CDF Erfahrung immer dann anbringen konnte, wenn gerade Not am Mann war. Zusätzlich zu den bereits genannten Personen möchte ich mich bei Frau Dr. Jeannine Wagner-Kuhr und Herrn Dr. Frank-Peter Schilling für das Korrekturlesen dieser Arbeit bedanken. Weiterhin danke ich Jochen Ott für seine kompetente Beratung bei Fragen zur statistischen Auswertung und seine Hilfe bei der Benutzung seines theta-Frameworks.

Für die schöne Zeit am CERN bedanke ich mich herzlichst bei meinen Karlsruher Kollegen, die zeitgleich mit mir in Genf weilten. Neben den bereits erwähnten Dr. Thomas Peiffer, Dr. Frank-Peter Schilling und Jochen Ott waren dies Dr. Danilo Piparo, Dr. Klaus Rabbertz, Dr. Philipp Schieferdecker, Hauke Held, Fred

Stober und Manuel Zeise.

Ebenso gilt mein Dank meinen Kollegen vom Administratoren-Team die für das reibungslose Funktionieren der Computing-Infrastruktur gesorgt haben sowie der guten Seele des Instituts, unserer Sekretärin Frau Bärbel Bräunling. Des Weiteren bedanke ich mich bei allen Mitarbeitern des Instituts für Experimentelle Kernphysik für die oftmals nahezu familiäre Arbeitsatmosphäre.

Abschließend danke ich meiner Schwester Yvonne, meinen Eltern Andrea und Eugen sowie meinen Großeltern, die mich immer bedingungslos unterstützt haben und gerade in schwierigen Phasen mein grenzenloser Rückhalt waren. Weiterhin bedanke ich mich bei allen, die ich zu meinen Freunden zählen darf, einfach dafür, dass sie so sind, wie sie sind.