Masterarbeit

Zur Erlangung des akademischen Grades Master of Science

Studien zur Rekonstruktion der invarianten Masse von Top-Antitop-Paaren im dileptonischen Kanal mit Hilfe eines neuronalen Netzwerkes am ATLAS Detektor

Eingereicht von:	B.Sc. Lukas Roscher
Erstgutachter:	PD Dr. Klaus Mönig
Zweitgutachter:	Prof. Dr. Heiko Lacker

Eingereicht am Institut für Physik der Humboldt Universität zu Berlin am: 18.07.2024

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und noch nicht für andere Prüfungen eingereicht habe. Sämtliche Quellen einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, sind als solche kenntlich gemacht. Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung eingeleitet wird.

Berlin, ...

Kurzfassung

Das Standardmodell (SM) der Teilchenphysik bildet die Grundlage für unser Verständnis der fundamentalen Kräfte und Teilchen. Dennoch existieren Phänomene, wie die dunkle Materie oder die Beobachtung der Neutrinooszillation, die gute Gründe dafür liefern, dass das SM keine vollständige Theorie ist. Präzisionsmessungen des SM an Kollisions-Experimenten sind entscheidend, um Abweichungen zu identifizieren, die auf neue Physik hinweisen könnten. Die Untersuchungen dieser Arbeit sind durch die Messung der Top-Quark-Yukawa Kopplung Y_t im Verhältnis zum SM-Wert im dileptonischen Zerfallskanal von Top-Anti-Top ($t\bar{t}$)-Paaren motiviert. Die verwendeten (simulierten) Daten entsprechen den zwischen 2015 und 2018 aufgenommenen Proton-Proton-Kollisionen bei $\sqrt{s} = 13$ TeV am ATLAS-Experiment mit einer integrierten Luminosität von 140 fb^{-1}. Im dileptonischen Kanal zerfällt das $t\bar{t}$-Paar in zwei b-Quarks und zwei W-Bosonen, die wiederum in jeweils ein Lepton und Neutrino zerfallen. Virtuelle Korrekturen durch den Austausch eines Higgs-Bosons zwischen den produzierten Top-Quarks modifizieren das Spektrum der invarianten Masse des Top-Anti-Top Paares $m_{t\bar{t}}$ im Bereich der Produktionsschwelle ($\approx 2m_t$). Eine genaue Untersuchung dieser Verteilung gibt Aufschluss über die Stärke der Top-Yukawa Kopplung. Die Rekonstruktion von $m_{t\bar{t}}$ ist aufgrund der beiden Neutrinos im Endzustand nur durch Näherungsmethoden möglich, die kinematische Beschränkungen der W-Boson- und Top-Quark-Masse sowie den rekonstruierten fehlenden Transversalimpuls einbeziehen. Daher wird ein neuronales Netzwerk aus dem Bereich des Deep-Learning in dieser Arbeit motiviert und vorgestellt, welches $m_{t\bar{t}}$ durch eine Regression unter Eingabe von *high-level* Observablen rekonstruiert. Das Modell wird mit Monte-Carlo simulierten Ereignissen trainiert. Dabei werden die invarianten Massen $m_{e\mu}$, m_{eb_1}, m_{eb_2}, $m_{\mu b_1}$, $m_{\mu b_2}$ und $m_{b_1 b_2}$ der Elektronen, Myonen und b-Jets im Endzustand, zusammen mit dem Betrag des fehlenden Transversalimpulses $E_{\mathrm{T}}^{\mathrm{miss}}$ als Eingangsvariablen verwendet. Die ausgegebene Verteilung des neuronalen Netzwerks wird für die durch PYTHIA8 und HERWIG7 generierten Datensätze analysiert, um den Einfluss der verschiedenen Ereignismodellierungen zu untersuchen. Außerdem wird die Übereinstimmung von Simulation und experimentellen Daten untersucht. Ein Maximum-Likelihood-Fit für eine zukünftige Messung von Y_t wird vorgestellt. Dabei werden die experimentelle Daten und das erwartete $t\bar{t}$-Signal für verschiedene Werte von Y_t verglichen und die beste Übereinstimmung unter der Berücksichtigung von systematischen Unsicherheiten als Schätzung für Y_t angegeben. Da die systematischen Unsicherheiten aus zeitlichen Gründen nicht implementiert wurden, wird ein erster Anpassung durch die präsentierte Fit-Routine anhand von Asimov-Daten vorgenommen, um die Präzision der Messung abzuschätzen. Im Vergleich zur einfach zu rekonstruierenden invarianten Masse m_{bbll}^{reco} der Leptonen und b-Jets zeigt die Verteilung des neuronalen Netzwerks eine um etwa 3,5% geringere Unsicherheit in der Schätzung von Y_t (ohne Berücksichtigung systematischer Unsicherheiten). Die systematische Unsicherheit der $t\bar{t}$-Modellierungsunterschiede durch PYTHIA8 und HERWIG7 wird ebenfalls untersucht.

Abstract

The Standard Model (SM) of particle physics forms the foundation of our understanding of the most fundamental forces and particles. Nevertheless, phenomena such as dark matter or the observation of neutrino-oscillation provide good reasons to believe that the SM is not a complete theory. Precision measurements of the SM in collision experiments are crucial for identifying deviations that could point to new physics. The studies in this thesis are motivated by the measurement of the top quark Yukawa coupling Y_t relative to the SM value in the dileptonic decay channel of top-antitop ($t\bar{t}$) pairs at the ATLAS experiment. The used (simulated) data correspond to the proton-proton collision data recorded between 2015 and 2018 at $\sqrt{s} = 13$ TeV with an integrated luminosity of 140 fb^{-1}. In the dileptonic channel, the $t\bar{t}$ pair decays into two b quarks and two W bosons, which then decay into one lepton and neutrino each. Virtual corrections through the exchange of a Higgs boson between the produced top quarks modify the spectrum of the invariant mass of the top-antitop pair $m_{t\bar{t}}$ in the region of the production threshold ($\approx 2m_t$). A precise analysis of this distribution provides insights into the strength of the top-Yukawa coupling . Due to the two neutrinos in the final state, the reconstruction of $m_{t\bar{t}}$ is only possible through approximation methods that incorporate kinematic constraints on the W mass and top quark mass as well as the reconstructed missing transverse momentum. Therefore, a neural network using techniques from the field of deep learning is motivated and presented in this thesis, which reconstructs $m_{t\bar{t}}$ through regression using high-level observables as input. The model is trained on Monte-Carlo simulated events. As inputs, the invariant masses $m_{e\mu}$, m_{eb_1}, m_{eb_2}, $m_{\mu b_1}$, $m_{\mu b_2}$, and $m_{b_1 b_2}$ of the electrons, muons, and b-jets in the final state, together with the magnitude of the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, are used. The output distribution of the neural network is analyzed for datasets generated by PYTHIA8 and HERWIG7 to investigate the influence of different $t\bar{t}$ modeling. Additionally, the agreement between simulation and experimental data is examined. A maximum-likelihood fit for a future measurement of Y_t is presented. Here, the experimental data and the expected $t\bar{t}$ signal for different values of Y_t are compared, and the best agreement is given as an estimate for Y_t, considering systematic uncertainties. Since the systematic uncertainties were not implemented due to time constraints, an initial fit based on Asimov data is performed to estimate the measurement precision. Compared to the easily reconstruced invariant mass m_{bbll}^{reco} of the leptons and b-jets, the distribution from the neural network shows an approximately 3.5% lower uncertainty in the estimation of Y_t (excluding systematics.) The systematic uncertainty of the $t\bar{t}$ modeling differences by PYTHIA8 and HERWIG7 is also examined.

Inhaltsverzeichnis

1 Motivation und Einführung

Die Hochenergiephysik (HEP) untersucht die elementaren Bausteine der Materie und die Wechselwirkungen zwischen diesen Elementarteilchen. Die zugrunde liegende Theorie ist das Standardmodell (SM) der Elementarteilchenphysik, welches die fundamentalen Teilchen in drei Familien von Fermionen einteilt, die jeweils aus zwei Quarks oder einem Lepton und einem Neutrino bestehen. Die fundamentalen Kräfte - elektromagnetische, schwache und starke Wechselwirkung - werden durch die jeweiligen Eichbosonen vermittelt. Der Large Hadron Collider (LHC) [1] erlaubt Präzisionstests des Standardmodells mittels Proton-Proton Kollisionen bei einer Schwerpunktenergie von 13 TeV. Durch Wechselwirkungen der Quarks und Gluonen in den Protonen entstehen bei den Kollisionen zahlreiche weitere Teilchen, anhand derer grundlegende Eigenschaften und Kopplungen untersucht werden können.

Mit der Entdeckung des Higgs-Bosons [2, 3] an den ATLAS [4] und CMS [5] Experimenten am LHC wurde der Teilcheninhalt des Standardmodells vervollständigt. Der Higgs-Mechanismus des Standardmodells erzeugt die Massen aller Fermionen über deren Wechselwirkung mit dem Higgs-Feld. Genauer ist die Masse m_f eines Fermions gegeben durch eine Yukawa Wechselwirkung mit dem Higgs-Feld mit einer Kopplungsstärke $g_f = \sqrt{2}m_f/\nu$, wobei ν der Vakuumerwartungswert des Higgs-Felds ist [6]. Da die Kopplungsstärke proportional zur Fermionenmasse ist, hat das 1995 am Tevatron [7] entdeckte Top-Quark [8, 9], das schwerste Elementarteilchen mit einer Masse von etwa $m_t = 172{,}5$ GeV [6], die stärkste Kopplung mit einem Wert von $g_t \approx 1$.

Eine Messung der Top-Yukawa Kopplung ist ein wichtiger Test des Standardmodells, da die Yukawa-Kopplungen durch (z.B.) alternative Higgs-Modelle [10, 11] als potenzielle Erweiterungen des Standardmodells modifiziert werden können. Außerdem gibt eine präzise Bestimmung der Top-Yukawa Kopplung Aufschluss über potenzielle Energieskalen neuer Physik und die Stabilität des elektroschwachen Vakuums [12–14]. Das Top-Quark, welches am LHC in hohen Raten in Top-Anti-Top ($t\bar{t}$) Paaren produziert wird, ist einzigartig in der Hinsicht, dass es im Gegensatz zu den anderen Quarks keine Bindungszustände eingeht, sondern aufgrund seiner hohen Masse über die schwache Wechselwirkung fast ausschließlich in ein b-Quark und ein W-Boson zerfällt [6]. Je nach Zerfall der beiden W-Bosonen kann der Zerfall eines Top-Quark-Paares in den vollhadronischen Kanal (zwei Quark-Antiquark-Paare), den semileptonischen Kanal (ein Quark-Antiquark-Paar und ein Lepton mit Neutrino) und den dileptonischen Kanal (zwei Leptonen mit Neutrinos) kategorisiert werden. Eine Messung der Top-Yukawa Kopplung kann auf direkter Weise in der Produktion eines Higgs-Bosons in Verbindung mit einem $t\bar{t}$-Paar vorgenommen werden [15, 16]. Diese Art von Messungen beinhaltet jedoch Annahmen über die Yukawa-Kopplung zu den jeweiligen Zerfallsprodukten, da die Kopplung im Top-Loop des Anfangszustands anders ist, als im Endzustand. Alternativ kann die Top-Yukawa Kopplung indirekt in Ereignissen der Vier-Top-Quark-Produktion [17, 18] untersucht werden, wobei das Higgs-Boson den Produktionswirkungsquerschnitt direkt beeinflusst. Auch anhand von virtuellen Korrekturen zwischen den produzierten Top-Anti-Top-Paaren sind Messungen möglich [19, 20], wie in dieser Arbeit thematisiert wird.

Elektroschwache Korrekturen durch den Austausch eines virtuellen Higgs-Bosons zwischen den produzierten Top- und Anti-Top-Quarks in $t\bar{t}$-Ereignissen können differentielle Verteilungen, wie die invariante Masse des $t\bar{t}$-Paares oder den Streuwinkel des Top-Quarks relativ zur Strahlachse $\cos\theta^*$, im Energiebereich der Produktionsschwelle ($\approx 2m_t$) erheblich modifizieren. Demnach sind präzise Untersuchungen dieser Verteilungen sensitiv auf die Top-Yukawa Kopplung, die mit Hilfe von statistischen Methoden abgeschätzt bzw. eingeschränkt werden kann. In die-

ser Arbeit wird eine indirekte Messung der Top-Yukawa Kopplung im Verhältnis zum SM-Wert $Y_t = g_t/g_t^{\mathrm{SM}}$ motiviert. Dabei wird die differentielle Verteilung der invarianten Masse des $t\bar{t}$-Paares für verschiedene Werte von Y_t^2 parametrisiert und Anpassungswerte für Y_t^2 in einem Maximum-Likelihood-Fit untersucht. Die Parametrisierung in Y_t^2 erlaubt die Untersuchung eines linearen Problems, da die elektroschwachen Korrekturen genau quadratisch in den Wirkungsquerschnitt eingehen. Die Messung wird im dileptonischen Zerfallskanal der analysierten $t\bar{t}$-Paare durchgeführt, d.h. beide W-Bosonen zerfallen leptonisch in ein Lepton (Elektron oder Myon) und ein Neutrino. Mit einem Verzweigungsverhältnis (engl. *Branching Ratio* BR) von etwa BR=10,5% [6] ist die Ereignisrate im Vergleich zum hadronischen Kanal ($BR = 45{,}7\%$ [6]) und semi-leptonischen Kanal ($BR = 43{,}8\%$ [6]) am geringsten, jedoch liefert er durch die klar identifizierten Leptonen und die geringe Rate an Untergrundprozessen eine klare Signatur. Neutrinos wechselwirken nur durch die schwache Wechselwirkung und hinterlassen keine Spur im Detektor. Die kollidierenden Protonen stoßen frontal in longitudinaler Richtung aufeinander, sodass der Transversalimpuls aller produzierten Teilchen summiert Null ergeben muss. Daraus wird der fehlende Transversalimpuls $\vec{p_{\mathrm{T}}}^{\mathrm{miss}}$ als die negative Summe aller detektierten Transversalimpulse definiert, der die Information der nicht beobachteten Teilchen (z.B. Neutrinos) beinhaltet. Da der longitudinale Impuls nicht messbar ist, ist die vollständige Rekonstruktion kinematischer Verteilungen wie die invariante Masse der $t\bar{t}$-Paare in dileptonischen Ereignissen durch die beiden Neutrinos im Endzustand nur bedingt möglich. Typischerweise wird die Rekonstruktion durch Methoden genähert, die versuchen eine Aufteilung von $\vec{p_{\mathrm{T}}}^{\mathrm{miss}}$ für die beiden Neutrinos vornehmen und fordern, dass die beiden Lepton-Neutrino Paare der invarianten Masse des W-Bosons und die Lepton-Neutrino-Jet Kombination der invarianten Masse des Top-Quarks entsprechen [21, 22].

Alternative Methoden bietet der Bereich des Maschinellen Lernens. Durch die rasante Entwicklung der Rechenleistung von Computern haben Techniken des Tiefen Lernens (engl. *Deep Learning* DL) insbesondere in den letzten 15 Jahren vermehrt Anwendung in Analysen der Hochenergiephysik gefunden [23]. In dieser Arbeit wird die Rekonstruktion der invarianten Masse der $t\bar{t}$-Paare durch ein tiefes neuronales Regressionsnetzwerk motiviert und untersucht. Das grundlegende Prinzip besteht darin, dass aus großen Datenmengen Strukturen und Muster erkannt und präzise Vorhersagen erstellt werden können. Im Falle der Regression wird eine sog. Verlustfunktion, die den Unterschied zwischen den eingegebenen Daten und einer vorgegebenen Zielfunktion angibt, iterativ in einem Trainingsprozess mit Hilfe von Optimierungsalgorithmen minimiert. Die Eingebedaten können dabei rekonstruierte kinematische Observablen der untersuchten Teilchen im dileptonischen Endzustand sein, während die Zielverteilung im Kontext dieser Arbeit die von einem Ereignisgenerator mit einer bestimmten Genauigkeit berechnete Verteilung der invarianten Masse der $t\bar{t}$-Paare ist. Das neuronale Netzwerk wird mit simulierten $t\bar{t}$-Daten, die durch Monte-Carlo Generatoren generiert werden, trainiert und ausgewertet und die ausgegebene Verteilung der invarianten Masse der $t\bar{t}$-Paare letztlich für die Analyse der Top-Yukawa Kopplung verwendet.

Die Arbeit gliedert sich wie folgt. Zunächst wird in Kapitel 2 eine Einführung in das Standardmodell gegeben und auf die Eigenschaften des Top-Quarks sowie die Produktion und Simulation von Top-Quarks am LHC eingegangen. Kapitel 3 erläutert den Aufbau des ATLAS-Experiments am LHC sowie die Rekonstruktion von Teilchen aus Signalen im Detektor. In Kapitel 4 wird auf

die Grundlagen des maschinellen Lernens und die Funktionsweise künstlicher neuronaler Netzwerke eingegangen sowie eine Beschreibung der Optimierungsalgorithmen gegeben. Kapitel 5 gibt einen Überblick zu den verwendeten Datensätzen und der Ereignisselektion. In Kapitel 6 wird detailliert auf die systematische Optimierung des neuronalen Netzwerks und die ausgegebene Verteilung der invarianten Masse der $t\bar{t}$-Paare eingegangen. Der Einfluss der Top-Yukawa Kopplung auf differentielle Verteilungen, sowie die statistische Methode zur Schätzung der Top-Yukawa Kopplung anhand der vom Netzwerk rekonstruierten invarianten Masse der $t\bar{t}$-Paare ist in Kapitel 7 erläutert. In Kapitel 8 werden die Ergebnisse zusammengefasst und ein Ausblick gegeben.

Die dargestellten Studien wurden im Rahmen der offiziellen ATLAS-Analysegruppe „Extraction of the top quark Yukawa coupling from the top quark pair production cross section in the dilepton final state with the complete run 2 data set“ in einer laufenden Zusammenarbeit von ATLAS-Arbeitsgruppen der Universität Siegen, University of Cape Town (Südafrika) und DESY-Zeuthen angefertigt.

2 Theoretische Grundlagen

Die Hochenergiephysik (HEP) beschäftigt sich mit der Zusammensetzung der Materie und ihrer fundamentalen Wechselwirkungen. Das aktuelle Verständnis wird im sogenannten Standardmodell (SM) zusammengefasst, dessen Genauigkeit durch experimentelle Bestätigungen an Teilchenbeschleunigern wie LEP [24] und Tevatron [7] bewiesen und insbesondere durch die Vielfalt der Messungen am LHC [1] (s. Kap. 3) weiter untermauert wird. Dennoch bestehen ungelöste Probleme und Fragen, auf die das SM keine Antworten liefert. Beispiele hierfür sind die dunkle Materie, die Baryonenasymmetrie oder die nicht-verschwindende Masse von Neutrinos. Im folgenden Kapitel soll ein Überblick über das Standardmodell (nach [25–27]) und insbesondere die für diese Arbeit relevanten Theorien und Prozesse gegeben werden. Zudem wird genauer auf die besondere Bedeutung und Physik des Top-Quarks, sowie auf die Produktion und Simulation von Top-Quark-Paaren am LHC eingegangen.

2.1 Das Standardmodell der Teilchenphysik

Das Standardmodell der Teilchenphysik beschreibt die Elementarteilchen und deren Wechselwirkungen untereinander. Im SM werden drei der vier Grundkräfte der Physik beschrieben: die elektromagnetische, die schwache und die starke Wechselwirkung. Einzig die Gravitation als verbleibende Grundkraft wird im Standardmodell nicht untergebracht. Sie kann jedoch aufgrund ihrer vergleichsweise schwachen Wirkung in Bezug auf die Teilchenphysik (besonders am LHC) vernachlässigt werden. Eine Vereinigung aller vier Grundkräfte ist eine der bekanntesten offenen Fragen der Teilchenphysik.

Das Standardmodell ist eine relativistische Quantenfeldtheorie, in dessen Rahmen Einsteins spezielle Relativitätstheorie mit den Prinzipien der Quantenmechanik vereint wird. Genauer wird jede Wechselwirkung im SM durch eine Quantenfeldtheorie beschrieben. Beispiele hierfür sind die Quantenelektrodynamik (QED), welche die Interaktion von Licht mit Materie beschreibt oder die Quantenchromodynamik (QCD), die die Interaktion stark wechselwirkender Teilchen charakterisiert.

2.1.1 Die Elementarteilchen

Wie in Abbildung 2.1 zu erkennen ist, werden die als punktförmig angenommenen Konstituenten des Standardmodels, die Elementarteilchen, in drei Gruppen unterteilt: Fermionen mit Spin-1/2, die wiederum in sechs Leptonen (grün) und sechs Quarks (violett) aufgeteilt sind, vier Austauschbosonen (rot), mit Spin-1 und das skalare Higgs-Boson (gelb) mit Spin-0. Zu jedem Fermion existiert ein Anti-Teilchen, welches jeweils die gleiche Masse, aber entgegengesetzte Ladung besitzt.

Das Elektron und das Elektron-Neutrino sind als *erste Generation* der Leptonen bekannt. Dazu kommt das im Vergleich zum Elektron schwerere Myon ($m_\mu \approx 200 m_e$) mit Myon-Neutrino der *zweiten Generation.* Zuletzt existiert das noch massereichere Tau-Lepton ($m_\tau \approx 3500 m_e$) inklusive Tau-Neutrino der *dritten Generation.*

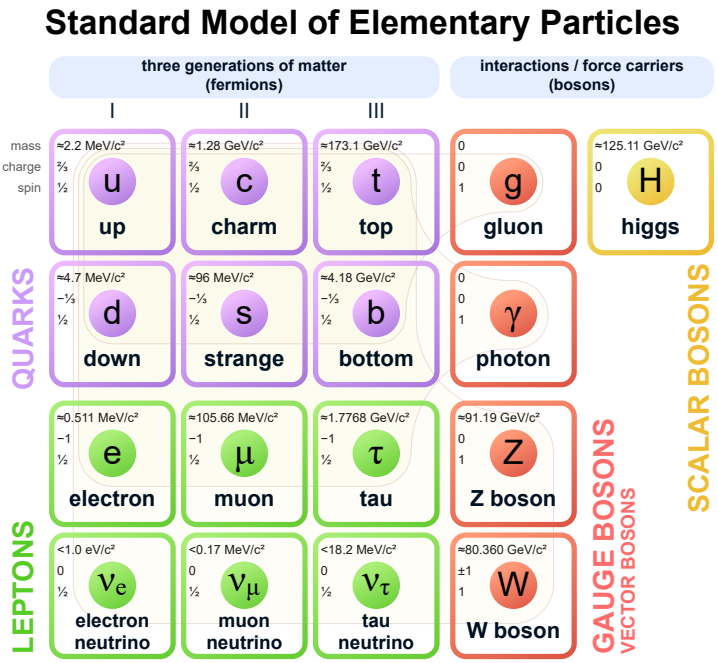

Abbildung 2.1: Übersicht der Elementarteilchen des Standardmodells. Zusätzlich sind die Masse, Ladung und Spin gegeben. Die dünnen hellbraunen Linien geben an, welche Fermionen (Quarks und Leptonen) mit den Eichbosonen wechselwirken. [28]

Typischerweise werden die Lepton-Paare in Dubletts angeordnet:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}. \tag{2.1}$$

Ein Grund für die Anordnung der Leptonen in Dubletts liefert die Erhaltung der sogenannten Leptonenfamilienzahl, welche besagt, dass sich Leptonen nur innerhalb eines Dubletts (unter Abstrahlung eines $W^\pm$ Bosons) ineinander umwandeln können. Außer den Unterschieden in der Masse sind die physikalischen Eigenschaften der drei geladenen Leptonen identisch und sie nehmen an der elektromagnetischen als auch an der schwachen Wechselwirkung teil. Die neutralen Neutrinos wechselwirken mit der Materie nur durch die schwache Wechselwirkung.

Auch Quarks kommen in drei Generationen vor. Diese bestehen aus dem Up- und Down-Quark, aus denen Proton (uud) und Neutron (udd) gebildet werden, dem Charm- und Strange-Quark sowie dem Top- und Bottom-Quark und werden ebenfalls in Teilchendubletts angeordnet:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}. \tag{2.2}$$

Auch im Quarksektor steigt die Masse der Quarks über die Generationen an. Die Ladung aller „Up-type“ -Quarks (u,c,t) beträgt $+\frac{2}{3}$, die aller „Down-type“ Quarks (d,s,b) $-\frac{1}{3}$. Übergänge

zwischen den Dubletts sind im Gegensatz zu den Leptonen möglich, allerdings sind diese gegen Übergänge innerhalb eines Dubletts unterdrückt. Dies wird durch die Cabibbo-Kobayashi-Maskawa Matrix (auch CKM-Matrix) [29], [30] beschrieben. Auch Quarks koppeln an die elektromagnetische und die schwache Wechselwirkung.

Quarks haben jedoch eine zusätzliche Quantenzahl, die sogenannte *Farbladung*. Diese kann *rot*, *grün* oder *blau* sein (für Anti-Quarks dementsprechend die Anti-Farbe) und repräsentiert die Kopplung an die starke Wechselwirkung und die damit verbundenen acht masselosen Austauschbosonen, den Gluonen g. Zu den weiteren Austauschbosonen des SM gehören das masselose Photon (γ) des Elektromagnetismus und die drei massiven Bosonen der schwachen Wechselwirkung Z^0 mit $m_Z \approx 90$ GeV[1)] [6] und $W^{\pm}$ mit jeweils $m_{W^{\pm}} \approx 80$ GeV [6]. Abschließend gibt es das skalare Higgs-Boson H, welches das als letztes entdeckte Teilchen des Standardmodells ist und eine wichtige Rolle im Zusammenhang mit den Massen aller anderen Elementarteilchen spielt. Auf diesen Mechanismus wird unter anderem im nächsten Abschnitt eingegangen.

2.1.2 Die fundamentalen Wechselwirkungen

Im Formalismus der Quantenfeldtheorie werden Teilchen als Anregungen eines Quantenfeldes beschrieben. Dessen Dynamik und Form ist durch eine Lagrangedichte $\mathcal{L}$ bestimmt. Dem Standardmodell liegt die Eichsymmetrie $SU(3)_C \times SU(2)_L \times U(1)_Y$ zugrunde, wobei die Austauschbosonen die Anregungen der jeweiligen Eichfelder sind. Nach dem Noether-Theorem [31] geht jede kontinuierliche Symmetrie mit einer erhaltenen Ladung einher. Die starke Wechselwirkung wird durch die $SU(3)_C$-Symmetrie beschrieben und die erhaltene Ladung als *Farbladung* (engl. *colour*) definiert. Die schwache sowie die elektromagnetische Wechselwirkung werden im SM zu der Elektroschwachen Theorie zusammengefasst, welche durch die Symmetriegruppen $SU(2)_L \times U(1)_Y$ beschrieben werden. Die erhaltenen Quantenzahlen dieser Symmetrien sind die dritte Komponente des Isospins I_3 und die Hyperladung Y. Die elektrische Ladung folgt aus der linearen Kombination ($Q = I_3 + \frac{1}{2}Y$). Der Higgs-Mechanismus des SM erklärt die Massen der Elementarteilchen durch die Interaktion mit dem Higgs-Feld. Im Folgenden soll nur kurz auf die grundlegenden Aussagen der Theorien eingegangen werden.

Starke Wechselwirkung
Die Quantenchromodynamik (QCD) ist eine durch die $SU(3)$-Symmetriegruppe beschriebene Theorie der starken Wechselwirkung zwischen Quarks und Gluonen. Quarks tragen die Farbladung rot, grün oder blau. Eine Besonderheit der QCD ist, dass die Gluonen als Austauschbosonen selbst auch eine Farbladung tragen (im Vergleich, das Photon der QED ist elektrisch neutral). Dies erlaubt Gluon-Selbstwechselwirkungen in Form von 3-Gluon und 4-Gluon Vertices.

Quarks sind als einzelne, freie Teilchen nicht in Experimenten zu beobachten. Sie kommen gebunden in Form von Farb-Singuletts vor[2)], welche als Hadronen bezeichnet werden. Dieses Phä-

[1)] In der Hochenergiephysik wird das natürliche Einheitensystem verwendet, sodass die Lichtgeschwindigkeit $c = 1$, und das plancksche Wirkungsquantum $\hbar = 1$ sind. Damit haben Energie, Impuls und Masse die Dimension einer Energie, typischerweise in GeV.

[2)] Eine Ausnahme bildet das Top-Quark, was wegen seiner geringen mittleren Lebensdauer nicht hadronisiert. Für Details siehe 2.2

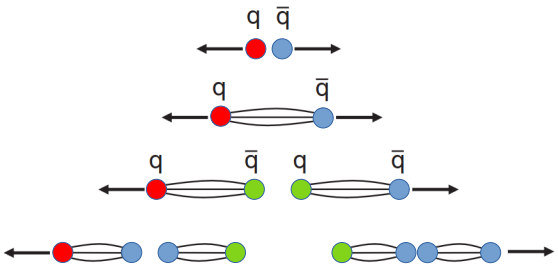

Abbildung 2.2: Schema der Hadronisierung von Quarks. Bei der Auseinanderbewegung zweier Quarks erhöht sich die Feldenergie, bis ein neues Quark-Anti-Quark-Paar produziert wird. Dieser Prozess wird fortgeführt bis gebundene Zustände (Hadronen) entstehen. Angepasst aus [25]

nomen wird als Quark-Einschluss (engl. *quark confinement*) bezeichnet. Abbildung 2.2 zeigt anschaulich was passiert, wenn zwei freie Quarks voneinander weg bewegt werden. Die Wechselwirkung der Quarks über den Austausch virtueller Gluonen bildet Feldlinien, die sich aufgrund anziehender Wechselwirkungen zwischen den Gluonen zu einer (räumlich begrenzten) Röhre bilden. Bei großen Abständen ist die gespeicherte Energiedichte proportional zum Abstand r der beiden Quarks, was zu einem Potentialterm der Form:

$$V(r) \sim \kappa r \tag{2.3}$$

führt. Die Konstante $k = 1$ GeV/fm kann experimentell bestimmt werden [6]. Dies zeigt, dass eine hohe Energie nötig wäre, die beiden Quarks weiter auseinanderzubringen. Während keine genaue analytische Erklärung existiert, lässt sich sagen, dass bei der Auseinanderbewegung zweier Quarks ab einem bestimmten Punkt genug Energiedichte in der Feldröhre vorhanden ist, um neue Quark-Antiquark Paare zu bilden. Dieser Prozess wird solang fortgeführt, bis die Energie der $q\bar{q}$-Paare gering genug ist, um Hadronen zu bilden. Als Resultat der *Hadronisierung* werden aus dem ursprünglichen Quark und Anti-Quark zwei Hadronen-Jets gebildet, die in Richtung der Ausgangsquarks zeigen. Hadronen unterteilen sich in zwei Konfigurationen, den *Mesonen* ($q\bar{q}$), sowie den *Baryonen* ($\bar{q}\bar{q}\bar{q}$ oder qqq).

Eine weitere Besonderheit der QCD ist das Verhalten der Kopplungskonstante α_s in Abhängigkeit der Energie des betrachteten Prozesses. So ist $\alpha_s(1\text{GeV}) \approx 1$ bei niedrigen Energien, jedoch bei höheren Energien $\alpha_s(100\text{GeV}) \approx 0{,}1$, was störungstheoretische Rechnungen erlaubt und bedeutet, dass Quarks und Gluonen bei hohen Energien als quasi-freie Teilchen behandelt werden können. Diese Eigenschaft der QCD wird als *asymptotische Freiheit* bezeichnet. Im Vergleich, die Kopplungskonstante der elektromagnetischen Wechselwirkung α_{EM} nimmt für steigende Energie leicht zu.

Elektroschwache Theorie und Higgs-Mechanismus

Die elektroschwache Wechselwirkung ist eine Eichtheorie mit der Symmetriegruppe $SU(2)_L \times U(1)_Y$ und vereint die schwache und elektromagnetische Wechselwirkung, welche durch die Eichbosonen $W^{\pm}$ und Z^0, sowie durch das Photon γ übertragen werden. Dies geht auf Glashow, Salam und Weinberg in den 1960er Jahren zurück [32–34]. Die schwache Wechselwirkung ist paritätsverletztend [35] und wirkt ausschließlich auf linkshändige Fermionen, was den Index

L erklärt. Durch die Symmetriegruppe $SU(2)$ der schwachen Wechselwirkung wird der sogenannte Isospin I definiert, welcher in der Form $I_i = \frac{\sigma_i}{2}(i = 1,2,3)$ geschrieben werden kann, wobei σ_i die Pauli-Matrizen als Generatoren der Gruppe darstellen. So werden linkshändige Fermionen als Isospin-Dubletts ($I = \frac{1}{2}$) und rechtshändige Fermionen als Isospin-Singuletts ($I = 0$) angeordnet. Durch die Symmetriegruppe $U(1)_Y$ des Elektromagnetismus wird die Hyperladung Y eingeführt. Die elektrische Ladung Q folgt aus der linearen Kombination $Q = I_3 + \frac{1}{2}Y$.
Aus der Forderung nach lokaler Eichinvarianz der Lagrangedichte entstehen vier masselose Eichfelder: Die Felder (W^1,W^2,W^3) der $SU(2)_L$ sowie das Feld B der $U(1)_Y$. Aus diesen ergeben sich die physikalischen Felder der bekannten Eichbosonen wie folgt:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.4}$$

$$\begin{pmatrix} Z \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & -\cos\theta_W \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix}. \tag{2.5}$$

Dabei bezeichnet $\sin^2\theta_W \approx 0.23$ [6] den schwachen Mischungswinkel (auch Weinbergwinkel genannt). Bisher wurden Fermionen und Bosonen als masselos angenommen. Das Hinzufügen eines Massenterms in den Lagrangian würde die Eichinvarianz verletzen. Dieses Problem wird mit dem sogenannten Higgs-Mechanismus gelöst. In dessen Rahmen wird ein neues skalares Feld eingeführt, das *Higgs-Feld*, was aus zwei komplexen skalaren Feldern wie folgt zu einem Isospin-Dublett gebildet wird:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.6}$$

wobei ϕ^0 elektrisch neutral und $(\phi^+)^* = \phi^-$ gilt. Das zugehörige Potential ist das *Higgs-Potential*:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 = \mu^2\phi^2 + \lambda\phi^4. \tag{2.7}$$

Das Potential ist abhängig von den Parametern λ und μ, wobei $\lambda > 0$ sein muss, um ein endliches Minimum zu gewährleisten. Für $\lambda > 0$ und $\mu > 0$ ist das Potential symmetrisch mit genau einem Minimum bei $\phi = 0$ (s. Abb. 2.3 links). Für $\mu < 0$ ergibt sich die Form gezeigt in Abbildung 2.3 rechts, mit unendlich vielen Minima bei $\phi = \pm\left|\sqrt{\frac{-\mu^2}{\lambda}}\right| = \pm\nu$. Da das Minimum nicht bei $\phi = 0$ ist, wird von einem Vakuumerwartungswert $\nu \approx 246$ GeV [6] gesprochen. Die Wahl eines bestimmten Zustands als Vakuum wird als spontane Symmetriebrechung bezeichnet. Da das neutrale Photon nach der Symmetriebrechung erhalten bleibt, wird das Minimum als von Null verschiedener Vakuumerwartungswert des neutralen Feldes ϕ^0 gewählt (in anderen Worten: das Universum ist elektrisch neutral). Damit wird im SM die $SU(2)_L \times U(1)_Y$ Symmetrie in eine $U(1)_{EM}$ des Elektromagnetismus gebrochen. Dabei entstehen Massenterme für die

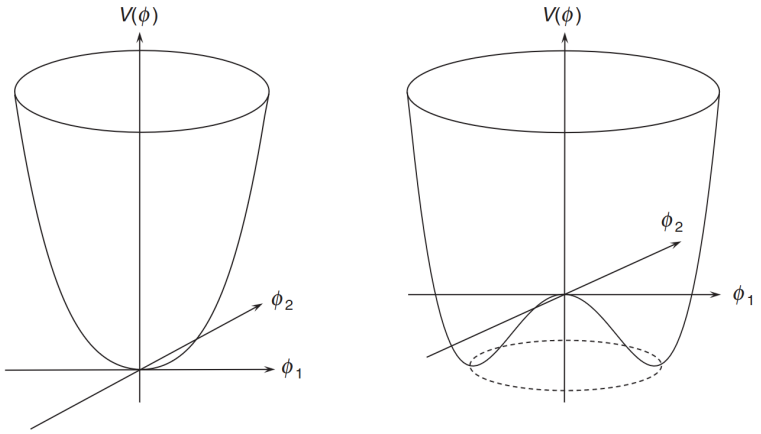

Abbildung 2.3: Die Form des eindimensionalen Higgs-Potentials nach Gl. 2.7 für $\lambda > 0$ und die beiden Fälle $\mu > 0$ (links) und $\mu < 0$ (rechts). Verwendet aus [25]

drei elektroschwachen Bosonen (das masselose Photon wird erhalten) und ein neues skalares, massives Teilchen mit Spin-0; das Higgs-Boson H:

$$m_{W^\pm} = g\frac{\nu}{2}, \qquad m_Z = \sqrt{g^2 + g'^2}\,\frac{\nu}{2}, \qquad m_\gamma = 0, \qquad m_H = 2\lambda\nu^2, \tag{2.8}$$

wobei g und g' jeweils die Kopplungsstärken der $SU(2)$ und $U(1)$ Theorien sind. Das Higgs-Boson wurde 2012 von den ATLAS und CMS Kollaborationen am LHC unabhängig voneinander experimentell bestätigt [2, 3] als ein skalares Teilchen mit einer Masse von $m_H = 125 \pm 0{,}17$ GeV [6]. Der Higgs-Mechanismus ist zudem verantwortlich für die Massen aller Fermionen im Standardmodell. Dies passiert, in dem sogenannte Yukawa-Wechselwirkungstherme zwischen dem Higgs-Feld und den Fermionen zur Lagrangedichte hinzugefügt werden, wobei die Kopplung gegeben ist durch:

$$g_f = \sqrt{2}\frac{m_f}{\nu}. \tag{2.9}$$

Die Kopplung selbst ist also abhängig von der Masse der Fermionen und skaliert mit dem Vakuumerwartungswert ν des Higgs, wie es analog auch bei den Bosonen zu sehen ist (s. Gl. 2.8). Das Top-Quark als schwerstes Elementarteilchen mit einer Masse von $m_t = 172{,}5$ GeV [6] koppelt somit am stärksten an das Higgs, insbesondere ergibt die Kopplung unter Einsetzen der Top-Masse einen interessanten Wert von $g_t \approx 1$.

2.2 Top-Quark Physik

Bereits im Jahre 1973 wurde die Existenz einer dritten Quark-Generation vorhergesagt [30]. Vier Jahre später wurde am Fermilab E228 Experiment das Bottom-Quark gefunden [36], was die Suche nach einem Partnerteilchen weiter motivierte. Ab 1983 wurde am Fermilab Tevatron [7], ein Proton-Antiproton-Beschleuniger, mit einer späteren Schwerpunktenergie von $\sqrt{s} = 1{,}8$ TeV

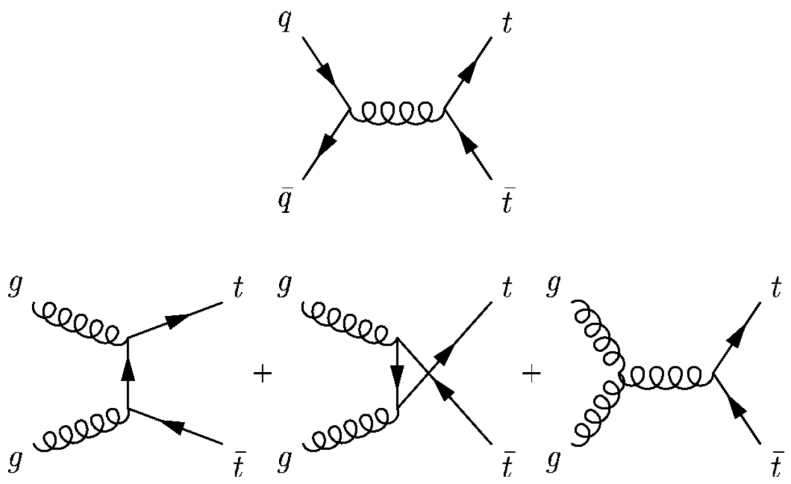

Abbildung 2.4: Feynman Diagramme der Top-Quark-Paarproduktion in führender Ordnung: Quark-Anti-Quark Vernichtung (oben) und Gluon-Fusion (unten). Verwendet aus [39]

in Betrieb genommen. Es dauerte noch bis 1995, ehe die Experimente DØ und CDF die Entdeckung eines neuen Teilchens [8, 9] vermeldeten, wobei es sich um das Top-Quark handelte. Das Top-Quark ist das schwerste Elementarteilchen mit einer Masse von $m_t = 172{,}5$ GeV [6], was zu einer mittleren Lebensdauer von nur rund $5 \cdot 10^{-25}$ s [6] führt. Damit hadronisiert das Top-Quark nicht, bildet also keine Bindungszustände mit anderen Quarks und zerfällt über die schwache Wechselwirkung. Dies erlaubt die Untersuchung der Eigenschaften eines „reinen" Quarks über seine unmittelbaren Zerfallsprodukte. Weiterhin ist das Top-Quark ein wesentlicher Bestandteil in Suchen nach Erweiterungen des Standardmodells [37, 38]. Eine allgemeine Übersicht über die Top-Quark Physik sowie Erkenntnisse vom Tevatron gibt [39], eine neuere Version zur Top-Physik am LHC gibt [40]. Beide dienen als Referenzen für die Informationen des folgenden Abschnitts.

2.2.1 Top-Quark-Paarproduktion und Zerfall

Das Top-Quark wird primär durch zwei Mechanismen produziert. Die Einzelproduktion von Top-Quarks über die schwache Wechselwirkung ist die deutlich seltenere Produktionsart und wird in dieser Arbeit nicht weiter behandelt. Viel häufiger werden Top-Quarks am LHC in Paaren ($t\bar{t}$-Paare) durch die starke Wechselwirkung produziert. In führender Ordnung gibt es für die Produktion zwei Prozesse: die Gluon-Fusion (Abb. 2.4 unten) und die Quark-Antiquark-Vernichtung (Abb. 2.4 oben). Ein Proton beispielsweise ist aus drei *Valenzquarks* (uud) zusammengesetzt, welche dessen Quantenzahlen bestimmen und typischerweise den größten Anteil des Gesamtimpulses tragen. In hochenergetischen Kollisionen können weitere, niederenergetische *Seeqaurks* und Gluonen als Bestandteile des Protons aufgelöst werden. Im Parton-Modell [41] werden alle im Hadron existierenden Elementarteilchen als *Partonen* zusammengefasst, die einen gewissen Anteil x des Gesamtimpulses des Hadrons besitzen. Im Schwerpunktsystem (engl. *center of mass system*) einer Kollision zweier Protonen findet der harte Streuprozess (also hohe Impulsüberträ-

ge) zwischen zwei Partonen der jeweiligen Protonen statt. Dies geschieht in einer Zeitspanne, die deutlich kürzer ist als die typische Fluktuationsdauer der QCD. Dies erlaubt eine Faktorisierung [42] des gesamten Wirkungsquerschnitts der Wechselwirkung in einen (pertubativ beschreibbaren) Wirkungsquerschnitt des Parton-Streuprozesses gewichtet mit den Partondichtefunktion (engl. *parton distribution functions* PDFs). Die Faktorisierung erfolgt anhand einer Faktorisierungsskala μ_F. In Berechnungen höherer Ordnungen der Störungstheorie treten Divergenzen auf, die mit Hilfe einer Renormierungsprozedur beseitigt werden können. Dabei wird eine Renormierungsskala μ_R eingeführt. Folgend wird $\mu = \mu_F = \mu_R = \sqrt{m_{t\bar{t}}^2 + p_{\mathrm{T},t}^2 + p_{\mathrm{T},\bar{t}}^2}$ verwendet, wobei p_T der Transversalimpuls des Top- und Anti-Top-Quarks ist. Die PDFs geben die Wahrscheinlichkeit an, ein Parton i mit Impulsbruchteil x des ursprünglichen Hadrons zu finden, sodass $p_i = x_i\, p_\mathrm{hadron}$ gilt. Für die $t\bar{t}$-Produktion in pp-Kollisionen in führender Ordnung der Störungstheorie ergibt sich damit ein Wirkungsquerschnitt in folgender Form [43]:

$$\sigma^{t\bar{t}}\left(\sqrt{s}\right) = \sum_{i,j=q,\bar{q},g} \int \mathrm{d}x_i\, \mathrm{d}x_j f_i\left(x_i,\, \mu\right) f_j\left(x_j,\, \mu\right) \cdot \hat{\sigma}^{ij\to t\bar{t}}\left(\hat{s},\, p_i,\, p_j,\, \mu\right), \tag{2.10}$$

wobei die Summe über alle Quarks und Gluonen der Protonen läuft, $f_i(x_i,\mu)$ und $f_j(x_j,\mu)$ die PDFs und $\hat{s}$ die partonische Schwerpunktenergie sind. Für den Wirkungsquerschnitt bei $\sqrt{s} = 13$ TeV und einer Top-Quark-Masse von $m_t = 172{,}5$ GeV lautet der gemessene Wert $\sigma^{t\bar{t}} = 833{,}9^{+37}_{-43}$ pb[3)], wobei die Unsicherheiten aus der Unsicherheit der Faktorisierungs- und Renormierungsskala, Top-Masse, Partondichtefunktion und Kopplungskonstante α_s stammen und in Quadratur aufaddiert werden (dies sind die empfohlenen Werte für die Run-2 Datenanalysen am ATLAS-Experiment, siehe [44–48]). Die partonische Schwerpunktenergie $\sqrt{\hat{s}} \approx \sqrt{x_i x_j s}$ muss für die Produktion eines $t\bar{t}$-Paares mindestens $\sqrt{\hat{s}} \geq 2m_t$ sein. Wird vereinfacht angenommen, dass die zwei Partonen den gleichen Impulsbruchteil $x_i = x_j = x$ tragen folgt:

$$x \geq \frac{2m_t}{\sqrt{s}}. \tag{2.11}$$

Daraus ergibt sich mit einer Top-Quark-Masse von $m_t \approx 175$ GeV und der LHC Schwerpunktenergie während Run-2 von $\sqrt{s} = 13$ TeV ein Impulsbruchteil von $x \approx 0{,}03$. Bei diesen x-Werten ist die Dichtefunktion der Gluonen deutlich größer als die Dichtefunktion der leichten Quarks. Rund 90% der $t\bar{t}$-Paare werden am LHC durch Gluon-Fusion und 10 % durch die $q\bar{q}$-Vernichtung erzeugt. Am Tevatron [7] kann ein $q\bar{q}$-Paar direkt aus den Valenzquarks der kollidierenden Protonen und Anti-Protonen ausgewählt werden, womit $t\bar{t}$-Paare hauptsächlich aus der Vernichtung von Quark-Paaren entstehen.

Zerfall von $t\bar{t}$-Paaren

Wie bereits erwähnt hadronisiert das Top-Quark nicht. Stattdessen zerfällt es über die schwache Wechselwirkung durch das CKM-Matrix-Element $|V_{tb}| \approx 0.999$ [29], [30] fast ausschließlich zu einem $W^\pm$-Boson und einem Bottom-Quark b. In jedem Endzustand sind also mindestens zwei b-Quarks vorhanden. $t\bar{t}$-Ereignisse werden nach dem Zerfall der jeweiligen W-Bosonen klas-

[3)] Die Einheit *Barn* wird für die Angabe von Wirkungsquerschnitten verwendet und entspricht einer Fläche von 1 b = $10^{-28}\mathrm{m}^2$ = 100 fm^2. Eine weitere wichtige Kenngröße von Teilchenbeschleunigern, die integrierte Luminosität L_int, wird damit in inversen Pikobarn oder inversen Femtobarn angegeben.

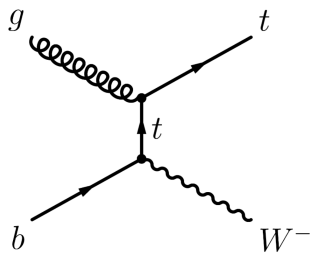

Abbildung 2.5: Feynman-Graph des dominanten Untergrundprozesses tW bei der $t\bar{t}$-Produktion im dileptonischen Kanal. Verwendet aus [49]

sifiziert. Es werden folgende drei Endzustände mit jeweiligen Zerfallswahrscheinlichkeiten [6] unterschieden:

A: $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'b \;\; q''\bar{q}'''\bar{b}$, (45,7%)
B: $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'b\ell^-\bar{\nu}_\ell\bar{b} + \ell^+\nu_\ell bq''\bar{q}'''\bar{b}$, (43,8%)
C: $t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\nu_\ell b \;\; \ell^-\bar{\nu}_{\ell'}\bar{b}$. (10,5%),

wobei $\ell = (e^-,\mu^-,\tau^-)$ die Leptonen und ν_ℓ die dazugehörigen Neutrinos sind. Die Quarks q, $\bar{q}'$, q'' und $\bar{g}'''$ sind fast ausschließlich $(u\bar{d})$ und $(c\bar{s})$. **A** bezeichnet den (voll-)hadronischen Kanal, bei dem kein Lepton im Endzustand vorkommt. **B** ist der semihadronische Kanal mit einem Lepton im Endzustand und **C** der dileptonische Kanal, bei dem zwei Leptonen und zwei Neutrinos im Endzustand vorkommen. Die Datenanalyse dieser Arbeit findet im dileptonischen Zerfallskanal statt. Der dileptonische Kanal hat den Vorteil, dass er eine geringe Rate an Untergrundereignissen anderer Physikprozesse aufweist und damit eine ziemlich reine Produktion von $t\bar{t}$-Ereignissen erlaubt. Der dominante Untergrundprozess ist die Einzel-Top Produktion in Verbindung mit einem $W^\pm$-Boson durch die schwache Wechselwirkung. Der zugehörige Feynman-Graph ist in Abbildung 2.5 dargestellt. Ein Nachteil des dileptonischen Kanals ist, dass er i.A. weniger Statistik liefert und die zwei nicht im Detektor beobachtbaren Neutrinos die vollständige Rekonstruktion der Ereignisse, um beispielsweise die invariante Masse der $t\bar{t}$-Paare zu berechnen, erschwert. Der transversale Impuls von Teilchen, die im Detektor kein Signal hinterlassen (wie die beiden Neutrinos), wird unter dem fehlenden Transversalimpuls $E_\mathrm{T}^\mathrm{miss}$ als negative Summe der Transversalimpulse aller gemessenen Teilchen zusammengefasst:

$$E_\mathrm{T}^\mathrm{miss} = \left| - \sum_{\text{detektiert}} \begin{pmatrix} p_x \\ p_y \end{pmatrix} \right|. \tag{2.12}$$

Eine Messung für die longitudinale Komponente existiert nicht, da die kollidierenden Partonen verschiedene longitudinale Impulsanteile besitzen und die Ruhesysteme der Kollisionen unterschiedliche Lorentz-Boosts aufweisen. Die Rekonstruktion ohne diese Information durch ein neuronales Netzwerk wird in dieser Kapitel 6 präsentiert.

2.2.2 Ereignissimulation

Für die Datenanalyse dieser Arbeit werden u.a. Top-Quark-Paarereignisse verwendet, die durch Monte-Carlo Simulationen generiert werden. Die Ereignisgeneration imitiert die Proton-Proton

Kollisionen mit dem Ziel, die entstehenden Teilchen samt ihrer kinematischen Eigenschaften und Wechselwirkung im Detektor zu simulieren. Wie im vorherigen Abschnitt angesprochen lässt sich die Berechnung des Wirkungsquerschnitts (s. Gl. 2.10) faktorisieren in die Partondichtefunktionen und den Wirkungsquerschnitt des Parton-Streuprozesses, die über die Integration der Matrix-Elemente der jeweiligen Feynman-Graphen über den Phasenraum berechnet werden. In dieser Arbeit wird der POWHEG-Generator [50–52] verwendet, der die partonischen Wirkungsquerschnitte der $t\bar{t}$ (s. Abb. 2.4) und tW (s. Abb. 2.5) Produktion in der nächst-höheren Ordnung (engl. *Next-to-Leading-Order* NLO) der QCD-Störungstheorie berechnet.

Effekte höherer Ordnungen wie die Abstrahlung weiterer Partonen oder sogenanntes Parton-Splitting in weitere Partonen werden durch Partonschauer-Algorithmen berücksichtigt. Diese Emissionen passieren vorrangig unter geringem Winkel und niedriger Energie und können im Anfangszustand (engl. *initial state radiation* ISR) der wechselwirkenden Teilchen (im Falle der Gluon-Fusion der $t\bar{t}$-Produktion sind dies zwei Gluonen) auftreten und beeinflussen die Schwerpunktenergie des harten Streuprozesses. Weiterhin treten die Effekte im Endzustand (engl. *final state radiation* FSR) der produzierten Teilchen (hier ein $t\bar{t}$-Paar) auf, wodurch die Impulsverteilung der produzierten Teilchen verändert wird. Emissionen und Parton-Splitting ausgehend von einem Gluon oder Quark der Skala Q^2 (aus dem harten Streuprozess) werden einem Partonschauer anhand einer Evoltuionsskala hinzugefügt. Dabei wird für jedes entstehende Teilchen aus Gluon-Abstrahlung oder Splitting eine Skala q^2 geschätzt [53]. Ist diese Skala oberhalb einer gesetzten Cut-Off Skala von etwa 1 GeV, werden die Teilchen dem Partonschauer hinzugefügt. Unterhalb diesem Cut-Off können störungstheoretische Rechnungen nicht mehr vorgenommen werden, da die Kopplungskonstante bei diesen Energien zu groß wird. Da jedoch aufgrund des Confinements der QCD (s. 2.1.2) alle Partonen zu Hadronen geformt werden, müssen phänomenologische Modelle dieser Fragmentation verwendet werden.

Anschaulich wird im sogenannten Lund-String-Modell [54] die Wechselwirkung zwischen zweier Quarks mit einer Saite beschrieben, die zwischen den Quarks aufgespannt wird. Die Feldenergie (siehe Abs. 2.1.2 Gl. 2.3) nimmt mit dem Abstand zwischen den Quarks zu und sobald die Energie der Saite groß genug wird, bricht sie und bildet Saiten mit den jeweils neuen $q\bar{q}$-Paaren. Gluonen werden als „Knick“ in das $q\bar{q}$-System aufgenommen. Dieser Prozess findet statt, bis die invariante Masse der produzierten $q\bar{q}$-Paare in der Größenordnung eines Hadrons ist. Ein weiteres Fragmentationsmodell ist die Cluster-Fragmentation [55, 56]. Dabei werden alle Gluonen am Ende des Partonschauers gezwungen, in $q\bar{q}$-Paare zu zerfallen. Aus den nun entstandenen benachbarten Quarks (bzw. Anti-Quarks) werden innerhalb bestimmter Radien farbneutrale Cluster gebildet, welche in weitere Cluster und schließlich Hadronen zerfallen.

In dieser Arbeit werden zwei verschiedene Generatoren für Partonschauer und Hadronisierung verwendet. PYTHIA8 [57] verwendet einen p_T-geordneten Partonschauer, sodass die Emissionen von Partonen, beginnend mit den energiereichsten Emissionen, in einer Reihenfolge durch zunehmend niedrigere Werte vom Transversalimpuls p_T charakterisiert werden. Die anschließende Fragmentation basiert in PYTHIA8 auf dem Lund-String-Modell. HERWIG7 [58, 59] verwendet einen Winkel-geordneten Partonschauer, wobei mit den größten Emissionswinkeln begonnen wird und aufeinanderfolgende Emissionen bei kleiner werdenden Winkeln stattfinden. Für die Fragmentation wird die Cluster-Fragmentation [55, 56] verwendet.

3 Das ATLAS Experiment am LHC

Die in dieser Arbeit verwendeten (simulierten) Daten von Proton-Proton-Kollisionen wurden vom ATLAS-Experiment am Large Hadron Collider (LHC) zwischen 2015 und 2018 aufgenommen. Im folgenden Kapitel wird ein Überblick über den LHC und den Aufbau des ATLAS-Experiments gegeben. Zudem wird die Rekonstruktion von Teilchen beschrieben.

3.1 Der Large Hadron Collider

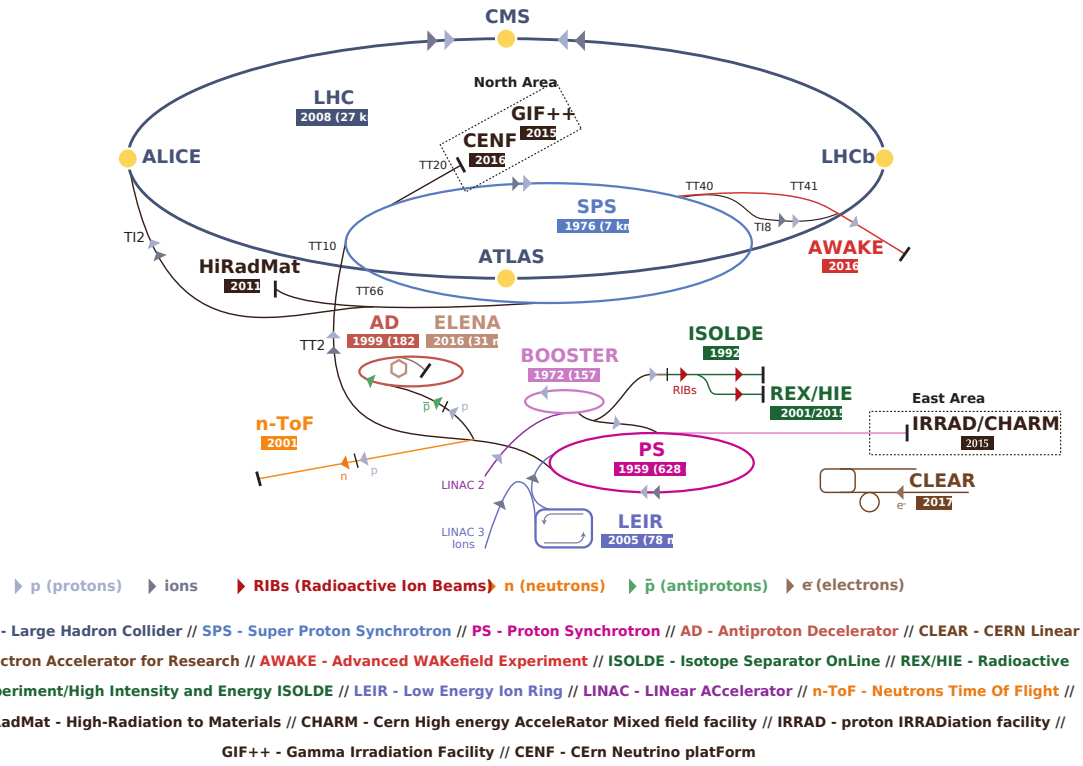

Abbildung 3.1: Schema des Beschleunigerkomplexes am CERN. Grafik aus [60]

Der LHC [1] ist der energetisch leistungsstärkste Teilchenbeschleuniger der Welt. Er befindet sich am CERN[1)] in der Nähe von Genf in einem bis zu 175 m unter der Erde liegenden, 26,659 km langen ringförmigen Tunnel. Der LHC wurde in den bereits vorhandenen Tunnel des damaligen Elektron-Positron Beschleunigers LEP [24] gebaut. Am LHC werden je nach Betriebsmodus Protonen oder schwere Ionen gegenläufig in zwei Strahlrohren durch supraleitende Hochfrequenz-Hohlraumresonatoren auf nahezu Lichtgeschwindigkeit beschleunigt. Genauer durchgehen die Protonenstrahlen eine schrittweise Beschleunigung: Zunächst werden sie im linearen Beschleuniger LINAC 2 auf 50 MeV beschleunigt, bis sie über den Proton Synchrotron Booster, den Proton Synchrotron (PS) und den Super Proton Synchrotron (SPS) eine kinetische Energie von 450 GeV erreichen und in den LHC-Ring eingespeist werden. Eine Übersicht aller vorhandenen Beschleuniger gibt Abbildung 3.1. Um die geladenen Teilchen auf ihrer Bahn entlang des Ringes zu halten, werden 1232 supraleitende Niob-Titan Dipol-Magnete verwendet, die eine Feldstärke von bis zu

[1)] CERN, die Europäische Organisation für Kernforschung

8,33 T erreichen und durch eine Kühlung mit flüssigem Helium auf etwa 1,9 K gekühlt werden.

Es gibt vier Kollisionspunkte (siehe Abb. 3.1), an dem die vier großen Experimente stationiert sind: ATLAS [4] und CMS [5], welche die beiden Universaldetektoren sind und ein möglichst breites physikalisches Spektrum abdecken, LHCb [61], dessen Schwerpunkt Präzisionsmessungen im Quark-Flavour-Sektor ist und ALICE [62], dessen Ziel u.a. die Erzeugung und Untersuchung eines Quark-Gluonen-Plasmas bei der Kollisionen schwerer Ionen ist. Im Rahmen dieser Arbeit wird mit (simulierten) Daten vom ATLAS-Detektor gearbeitet, dessen Aufbau im nächsten Abschnitt 3.2 beschrieben wird.

Ein Protonenstrahl besteht aus 2808 Bündeln (engl. *bunches*) von etwa 115 Milliarden Protonen, die zeitlich um 25 ns separiert sind. Dies erlaubt bis zu 1 Milliarde Kollisionen pro Sekunde. Die Ereignisrate $\dot{N}$ eines bestimmten Prozesses (bzw. eines bestimmten Endzustands) ist mit dem Wirkungsquerschnitt σ und der Luminosität L verbunden:

$$\dot{N} = \frac{\mathrm{d}N}{\mathrm{d}t} = L\sigma(\sqrt{s}). \tag{3.1}$$

Der Wirkungsquerschnitt ist abhängig von der Schwerpunktenergie $\sqrt{s}$ des Colliders, die eine entscheidende Rolle dabei spielt, wie wahrscheinlich es ist, einen Endzustand zu produzieren und welche Teilchen überhaupt erzeugt werden können. Die Luminosität L wird bei einem ringförmigen Beschleuniger beschrieben durch:

$$L = f\frac{nN_1N_2}{4\pi\Sigma_x\Sigma_y}, \tag{3.2}$$

wobei n die Anzahl der Bündel, N_1 und N_2 die Anzahl der Protonen im jeweiligen Bündel, Σ_x und Σ_y die transversalen Strahlbreiten und f die Frequenz ist, mit der die Bündel kollidieren. Die integrierte Luminosität $L_{\text{int}} = \int L\mathrm{d}t$ gibt die Anzahl an Ereignissen in einem bestimmten Zeitraum an. Über die Run-2 (2015-2018)-Datennahmeperiode wurden etwa $L_{\text{int}} = 147\,\text{fb}^{-1}$ [63] aufgenommen. Die für Physikanalysen verwendbaren Daten sind aufgrund von Detektordefekten während bestimmter Zeiträume verringert, sodass eine nutzbare integrierte Luminosität von etwa $L_{\text{int}} = 140\,\text{fb}^{-1}$ [64] für die Analysen dieser Arbeit verwendet werden. Bei der Kollision der gegenläufigen Protonenbündel entstehen mehrere Proton-Proton Wechselwirkungen. Eine höhere Luminosität entspricht gleichzeitig einer höheren Anzahl an Wechselwirkungen. Die Verteilung der mittleren Anzahl der Proton-Proton Wechselwirkungen μ ist in Abbildung 3.2 gezeigt und beträgt im Durchschnitt $\langle\mu\rangle = 33{,}7$ [63].

3.2 Der ATLAS-Detektor

ATLAS[2)] [4] ist ein Allzweck-Experiment der Teilchenphysik zur Untersuchung des Standardmodells und dessen möglichen Erweiterungen. Der ATLAS-Detektor ist 46 m lang, 25 m hoch, 25 m breit und wiegt etwa 7000 Tonnen. Weiterhin deckt er durch die zylindrische Bauweise fast den ganzen Bereich um den Kollisionspunkt ab. Abbildung 3.3 gibt einen Überblick des Experi-

[2)] ATLAS steht für **A** **T**orodial **L**HC **A**pparatu**S**.

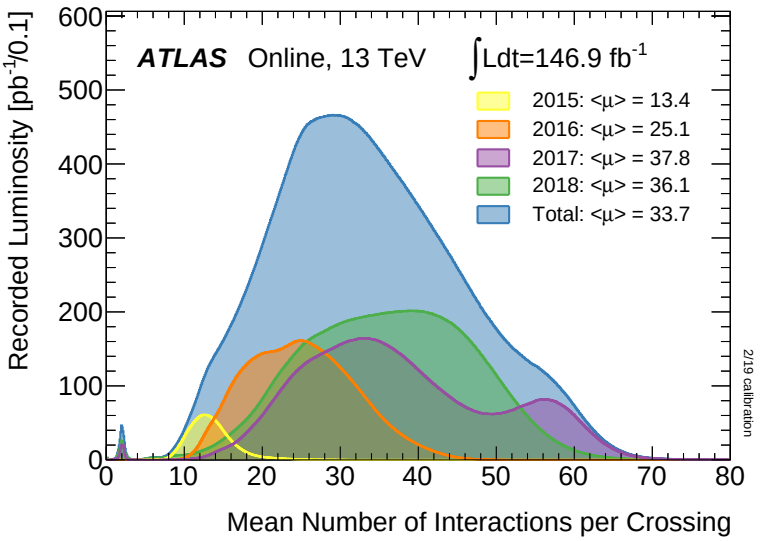

Abbildung 3.2: Die durchschnittliche Anzahl an Wechselwirkungen μ pro Zusammenstoß der Protonenbündel gemessen vom ATLAS Detektor zwischen 2015 und 2018 (Run-2). Grafik aus [63]

ments. Der ATLAS-Detektor besteht aus drei Hauptkomponenten: dem inneren Detektor (ID), dem Kalorimetersystem, bestehend aus dem elektromagnetischen Kalorimeter (ECAL) und dem hadronischen (HCAL) Kalorimeter sowie dem Myon-System. Etwa ± 17 m vom Kollisionspunkt befindet sich der Luminositäts-Detektor LUCID [65] und 240 m vom Interaktionspunkt in Vorwärtsrichtung ein weiterer Luminositäts-Detektor ALFA [66].

Das ATLAS-Koordinatensystem [4] ist speziell an den Aufbau des Experiments angepasst, um die Physik im Detektorvolumen effizient zu beschreiben. Der Koordinatenursprung befindet sich am Interaktionspunkt, in dem die Teilchen kollidieren. Die z-Achse zeigt in Richtung der Strahllinie. Die x-y-Ebene bildet die transversale Ebene, wobei die x-Achse in die Mitte des Beschleunigerrings und die y-Achse in den Himmel zeigt. Aufgrund der zylindrischen Anordnung wird der Abstand r von der Strahlachse und der Azimuthwinkel ϕ definiert, der gegen den Uhrzeigersinn um die Strahlrichtung von der x-Achse nach oben zeigt, sowie die Pseudorapidität $\eta = -\ln \tan(\theta/2)$ ($\eta = \infty$ für Teilchen parallel zur Strahlrichtung und $\eta = 0$ für Teilchen in der transversalen Ebene). Die Pseudorapidität wird häufig als Parametrisierung des Polarwinkels θ benutzt, da Differenzen $\Delta\eta$ für masselose Teilchen invariant unter Lorentz-Boosts entlang der z-Achse sind. Ein häufig benutztes Abstandsmaß zweier Teilchen in der Winkelebene ist

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.3}$$

3.2.1 Der innere Detektor

Nah am Kollisionspunkt befindet sich der innere Detektor (ID). Seine größte Aufgabe ist es, die Spuren (engl. *tracks*) von geladenen Teilchen zu bestimmen. Da er von einem Solenoidmagneten mit einer Feldstärke von 2 T parallel zur Strahlachse umgeben ist, kann der Impuls des Teilchens

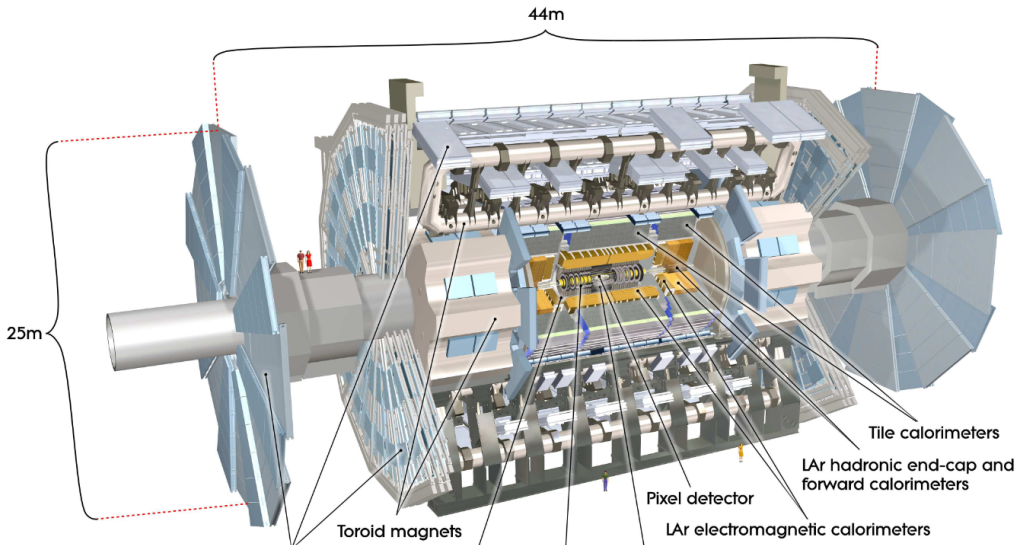

Abbildung 3.3: Überblick des ATLAS-Detektors, bestehend aus dem inneren Detektor, Kalorimeter- und Myonsystem. Grafik aus [4]

unter Kenntnis der Spur ermittelt werden. Weiterhin kann dadurch der Entstehungsort von geladenen Teilchen durch Rekonstruktion der Interaktions- und Zerfallsvertizes bestimmt werden. Der ID besteht (von innen nach außen) aus dem Pixel-Detektor, dem Silizium-Streifendetektor (engl. *semiconductor tracker* SCT) und dem Übergangsstrahlungsspurdetektor (engl. *transition radiation tracker* TRT). Der genaue Aufbau ist in Abbildung 3.4 veranschaulicht.

Pixel-Detektor
Die erste Schicht, noch vor dem genannten Pixel-Detektor, ist das Insertable B-Layer (IBL) [67], welches etwa 3,3 cm von der Strahllinie entfernt ist und damit der erste Messpunkt des ATLAS-Detektors ist. Das IBL besteht aus etwa 12 Millionen Silizium-Pixelsensoren mit einer Fläche von jeweils 50 µm × 250 µm (in $(r - \phi) \times z$). Das IBL deckt den Bereich bis $|\eta| < 2{,}5$ ab. Der Pixel-Detektor [4] besteht aus insgesamt drei zylindrisch angeordneten Schichten von Siliziumsensoren im Zentralbereich (auch *Barrel* genannt), mit 3 Scheiben in beiden Endkappen, in der Pixel in der transversalen Ebene angeordnet sind. Die insgesamt rund 80 Millionen 50 µm × 400 µm großen Pixel erfassen einen Bereich von $|\eta| < 2{,}5$ mit einer Genauigkeit von etwa 10 µm in radialer Richtung und 115 µm in Strahlrichtung.

SCT
Der SCT [4] besteht im Barrelbereich aus vier Lagen Siliziumstreifen-Detektoren, wobei die Siliziumstreifen rechteckig mit einer Länge bis zu 12 cm sind und einen Streifenabstand von 80 µm besitzen. Dazu gibt es jeweils 9 Scheiben in den Endkappen mit radial ausgerichteten Streifen unterschiedlicher Abstände. Mit dieser Anordnung deckt der SCT den Bereich $|\eta| < 2{,}5$.

TRT
Die äußerste Komponente des ID ist der TRT [4]. Er besteht aus sogenannten *Straw-Tubes*, wel-

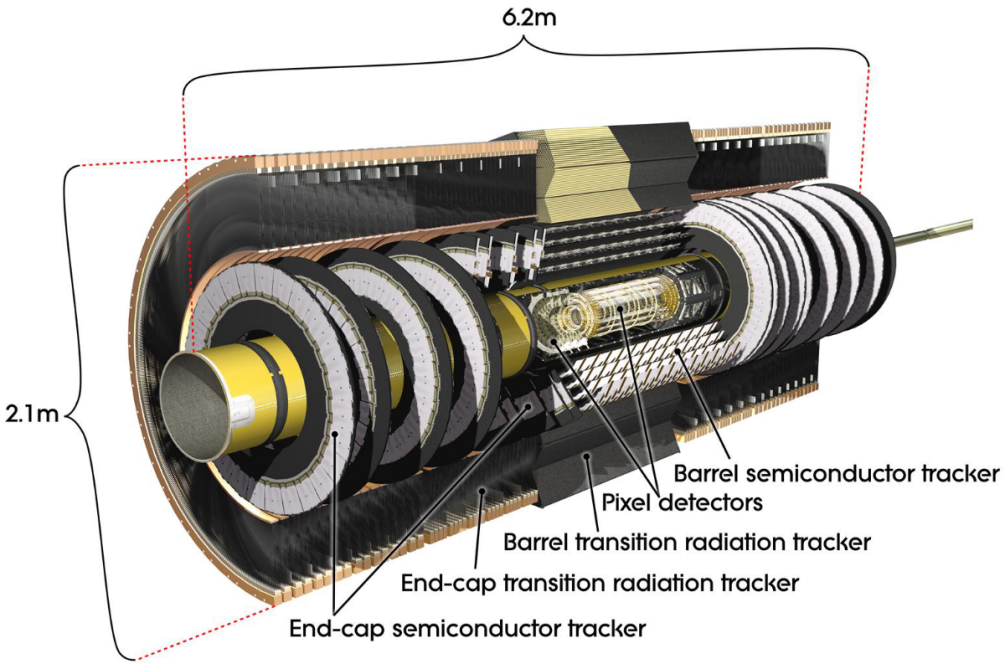

Abbildung 3.4: Aufbau des inneren Detektors bestehend aus dem Pixeldetektor, SCT-Detektor und TRT-Detektor. Grafik aus [4]

che Gas-gefüllte Röhren mit einem Draht in der Mitte sind und nach dem Prinzip der Proportionalkammer funktionieren. Die Straw-Tubes haben einen Durchmesser von 4 mm und sind im Barrelbereich axial ausgerichtet und 144 cm lang bzw. in den Endkappen radial ausgerichtet und 37 cm lang. Der TRT umfasst den Bereich $|\eta| < 2$ und ist so aufgebaut, dass ein geladenes Teilchen durchschnittlich etwa 30 Straw-Tubes passiert. Die Auflösung einer einzelnen Straw-Tube ist dafür nur rund 130 µm. Jedoch kann der TRT unter Benutzung der Übergangsstrahlung auch für Teilchenidentifikation benutzt werden. Die mit Xenon gefüllten Straw-Tubes sind im Zentralbereich in Schichten von Polypropylen-Fasern (bzw. -Folie in den Endkappen) eingebettet, wodurch passierende Elektronen Übergangstrahlung im Röntgenbereich abgeben, dessen Intensität linear vom Lorentzfaktor $\gamma = \frac{E}{mc^2}$ abhängt. Durch Messung dieser Übergangsstrahlung kann die Masse des Teilchens berechnet und anschließend identifiziert werden.

3.2.2 Das Kalorimetersystem

Der innere Detektor liefert keinerlei Informationen über die Energie eines Teilchens. Im Kalorimetersystem [4], bestehend aus elektromagnetischem Kalorimeter (ECAL) sowie dem hadronischen Kalorimeter (HCAL), wird ein Teilchenschauer aus der Kollision des hochenergetischen Teilchens mit dichter Materie produziert. Die Teilchen verlieren Energie beispielsweise durch Ionisation und werden gebremst. Durch die gemessene Energieablagerung kann somit auf die Energie des eintreffenden Teilchens zurückgeschlossen werden. Für eine Beschreibung der Rekonstruktion siehe Abschnitt 3.3. Realisiert wird dies durch abwechselnde Schichten aktivem Material und Absorbermaterial bzw. Szintillatoren. Eine schematische Übersicht gibt Abbildung

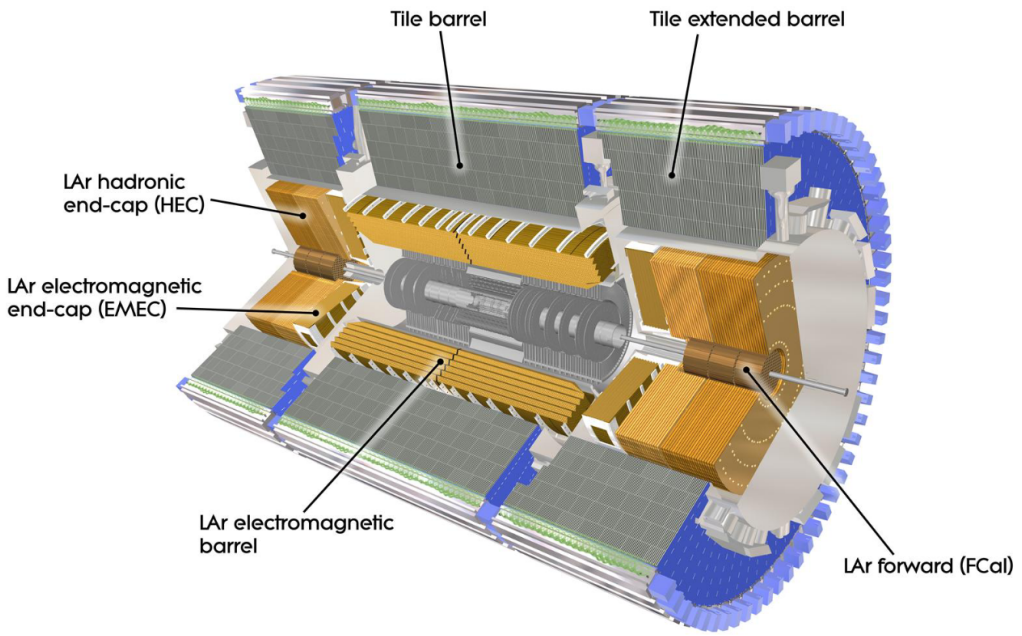

Abbildung 3.5: Übersicht zum Aufbau des Kalorimetersystems bestehend aus elektromagnetischem Kaloriemeter (ECAL) und hadronischen Kalorimeter (HCAL). Grafik aus [4]

3.5. Das Kalorimetersystem umfasst den Bereich $|\eta| < 4{,}9$.

ECAL
Das Ziel des elektromagnetischen Kalorimeters [4] ist die Energiemessung von Photonen und Elektronen (bzw. Positronen). Fällt ein Elektron (Photon) in das ECAL ein, werden Bremsstrahlungs-Photonen (e^+e^--Paare) erzeugt, welche wiederum e^+e^--Paare (Bremsstrahlung) produzieren. Dies passiert so lang, bis die mittlere Energie einen Wert erreicht, bei der Elektronen und Photonen bevorzugt Energie durch Streuprozesse verlieren, bis sie schließlich von Atomen absorbiert werden. Diese Vervielfachung der genannten sekundären Teilchen nennt man elektromagnetischen Schauer. Dementsprechend dick muss das ECAL mindestens gebaut sind, um Schauer mit verschiedenen Energien auflösen zu können.

Das elektromagnetische Kalorimeter besteht aus 1,9 mm dicken Absorberlagen Blei, abwechselnd mit Schichten aus flüssigem Argon, in denen die Energiedeposition gemessen wird. Der zylinderförmige Zentralbereich des ECAL umfasst $|\eta| < 1{,}475$, die beiden Endkappen erweitern dies auf $1{,}375 < |\eta| < 3{,}2$. Die Übergangsregion zwischen Barrel und Endkappen (etwa $1{,}375 < |\eta| < 1{,}52$) wird für Analysen üblicherweise nicht verwendet. Ein Teilchen kann schon vor dem Eintritt in das ECAL schauern, weswegen eine Presampler-Schicht aus flüssig-Argon in der Region $|\eta| < 1{,}8$ verbaut ist, um dies zu korrigieren.

HCAL
Die Aufgabe des HCAL [4] ist es, die Energie von Hadronen zu messen. Im Vergleich zu Teilchen

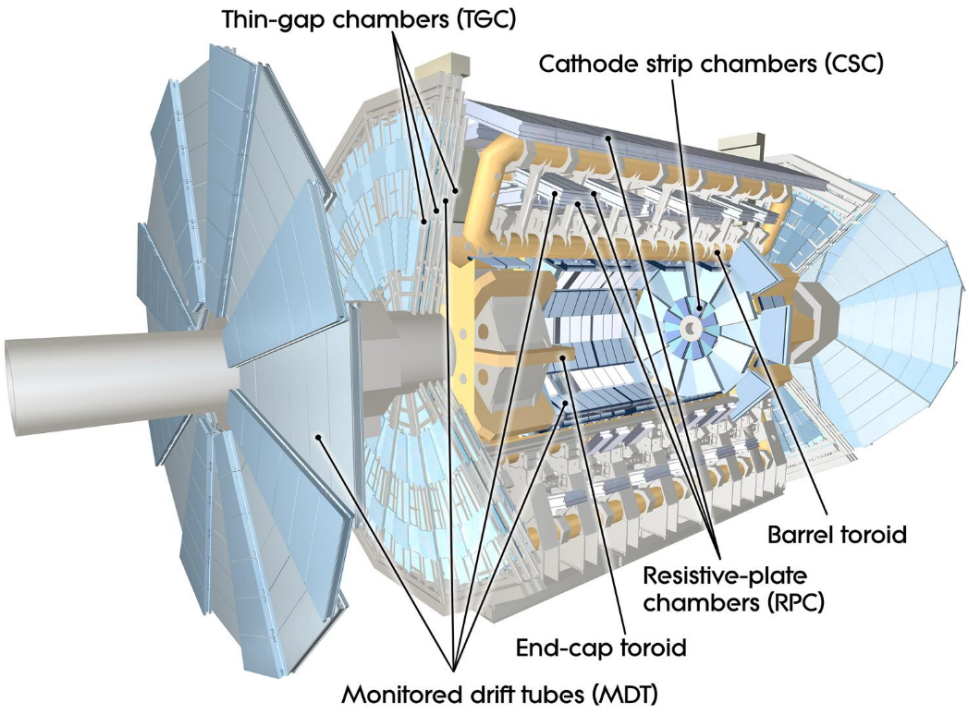

Abbildung 3.6: Schema des ATLAS Myon-Detektors. Grafik aus [4]

in elektromagnetischen Schauern können Hadronen stark wechselwirken, was einen hadronischen Teilchenschauer deutlich komplexer macht, da es beispielsweise die Produktion von anderen Hadronen oder die Anregung von Atomkernen beinhaltet. Im Barrelbereich, sowie im sog. tile-extended-Barrel, werden Stahl und Plastik-Szintillatoren als Absorbermedium und aktivem Material verwendet. Zusammen decken sie den Bereich von $|\eta| < 1{,}7$. In den Endkappen des HCAL befinden sich jeweils ein hadronisches Endkappen-Kalorimeter mit Kupfer als Absorbermaterial und flüssig-Argon als Nachweismedium, was die Region $1{,}5 < |\eta| < 3{,}2$ umfasst. Weiterhin befindet sich in den Endkappen jeweils ein drei-schichtiges Vorwärts-Kalorimeter, wobei die erste Schicht aus Kupfer und die anderen beiden Schichten aus Wolfram bestehen. Flüssig-Argon bildet das aktive Material in den drei Schichten. Die Vorwärts-Kalorimeter decken den Bereich $3{,}1 < |\eta| < 4{,}9$ ab.

3.2.3 Der Myon-Detektor

Der Myon-Detektor [4] bildet das äußerste System des ATLAS-Experiments und ist in Abbildung 3.6 dargestellt. Er wird benötigt, um den Impuls der schwach mit Materie wechselwirkenden Myonen zu messen. Dazu wird ein torodiales Magnet-System verwendet, dessen Magnetfeld weitgehend senkrecht zu den Myon-Trajektorien ausgerichtet ist. Außerdem besitzt der Myon-Detektor ein eigenständiges Trigger-System. Über den abgedeckten Bereich von $|\eta| < 2{,}7$ wird die Impulsmessung hauptsächlich von sogenannten *Monitored Drift Tubes* (MDT) vorgenommen, welche dreilagig im Barellbereich bzw. vierlagig in den Endkappen verbaut sind. Im Vorwärtsbe-

reich sowie nahe dem Kollisionspunkt sind *Cathode Strip Chambers* (CSC) platziert, welche aus streifenförmigen Kathoden aufgebaut sind, eine höhere Granularität aufweisen und für höhere Strahlungsintensitäten ausgelegt sind. Zur Komplementierung der Positions-Informationen und für das Triggersystem sind zwischen den MDTs im Barrelbereich *Resistive Plate Chambers* (RPC) und in den Endkappen eine Schicht von *Thin Gap Chambers* (TGC) verbaut. Dies erlaubt eine Auswahl relevanter Ereignisse anhand (z.B.) festgelegter Transversalimpuls-Schwellenwerten, was bei den hohen Datenmengen am LHC nötig ist. Dazu wird die räumliche Auflösung und damit die Genauigkeit der Bestimmung der Myon-Trajektorien verbessert.

3.2.4 Triggersystem

Am LHC werden die Protonenbündel mit einer Frequenz von 40 MHz zur Kollision gebracht. Die dabei produzierte Datenmenge kann als Ganzes nicht gespeichert werden. Da von den Wechselwirkungen zwischen den kollidierenden Protonen viele für Physikanalysen uninteressant sind, wird eine erste Selektion der Ereignisse getroffen, die durch das Trigger-System implementiert wird. Das hardwarebasierte Level-1-Triggersystem (L1) reduziert die aufgenommene Triggerrate auf 100 kHz. Diese Ereignisse werden dann durch Softwarealgorithmen im High-Level-Triggersystem (HLT) auf eine Rate von etwa 1 kHz reduziert und letztendlich zur Weiterverarbeitung gespeichert.

3.3 Rekonstruktion von Teilchen

Folgend wird näher auf die Rekonstruktion der relevanten Teilchen und Objekte eines dileptonischen $t\bar{t}$-Endzustands eingegangen, die durch eine Kombination von Signalen einzelner Detektor-Komponenten rekonstruiert werden.

3.3.1 Spur- und Vertexrekonstruktion

Teilchenspuren werden durch Treffer (auch „Hits“) im inneren Detektor (ID) rekonstruiert. Die Spurrekonstruktion beginnt mit der Erzeugung von „Seeds“, bestehend aus drei Hits im IBL, Pixel-Detektor oder SCT [68]. Diese Seeds werden aus der Kombination mit Hits aus den jeweils anderen Detektorkomponenten zu möglichen Spurkandidaten erweitert. Ein Qualitätsscore wird definiert, welcher von der Anzahl der Hits in den ID-Komponenten und dem Impuls der Spurkandidaten abhängt. Anhand dieses Scores werden Ambiguitäten zwischen Spurkandidaten gelöst, indem geteilte Hits den Kandidaten mit der niedrigeren Punktzahl entfernt werden [69]. Abschließend werden die besten Spurkandidaten mit Treffern im TRT erweitert und einem endgültigen Spur-Fit unterzogen.

Neben dem typischerweise interessanten harten Streuprozess treten weitere Proton-Proton Interaktionen auf. Um diese zu identifizieren, werden die Vertices der Wechselwirkungen eines Ereignisses rekonstruiert. Durch rekonstruierte Tracks, die aus der Wechselwirkungszone (engl. *beam spot*) stammen und bestimmte Kriterien erfüllen, werden die Positionen der Vertices letztendlich durch einen Vertex-Fit-Algorithmus berechnet [70, 71]. In dieser Arbeit verwendete Ereignisse erfordern mindestens einen Vertex, der aus mindestens zwei Tracks mit einem Trans-

versalimpuls von $p_T > 500\,\text{MeV}$ rekonstruiert ist. Der Vertex mit der größten Summe der quadrierten Trackimpulse $\sum p_T^2$ wird als primärer Vertex ausgewählt.

3.3.2 Elektronen

Elektronen werden am ATLAS-Experiment durch eine Zuordnung von Energieablagerungen im elektromagnetischen Kalorimeter (ECAL) und Spuren im Inneren Detektor (ID) rekonstruiert. Zunächst werden benachbarte Kalorimeterzellen im ECAL durch Algorithmen basierend auf Energieablagerungen und geometrischen Kriterien zu einem sogenannten EM-Topo-Cluster zusammengefasst [72]. Aus den verschiedenen Tracks, die einem EM-Topo-Cluster unter Erfüllung bestimmter Kriterien zugeordnet werden können, wird der beste Track anhand eines Scores basierend auf der Anzahl der Hits im ID und dem Abstandsmaß ΔR (s. Gl. 3.3) zwischen den Tracks und dem Cluster ausgewählt. Besitzt der Cluster eine Transversalenergie von $E_T > 1\,\text{GeV}$ und verzeichnet ein dazugehöriger Track mindestens vier Hits im Pixel-Detektor, kommt er als Supercluster-Seed in Frage. Algorithmen erweitern den Supercluster-Seed mit Tracks, die analog zu dem Verfahren bei den EM-Topo-Clustern ausgewählt werden und weiteren naheliegenden EM-Topo-Clustern, um den Supercluster zu bilden, der letztendlich das rekonstruierte Elektron definiert [72].

Die für die Analyse relevanten Elektronen entstehen aus dem harten Streuprozess oder Zerfällen von z.B. $W^{\pm}$-Bosonen unmittelbar danach. Jedoch können Elektronen auch aus hadronischen Zerfällen oder Energieablagerungen anderer Teilchen, wie z.B. Hadronen (sog. Fake-Elektronen) stammen [72]. Die Identifikation interessanter Elektronen erfolgt mit einer Likelihood-Methode (LH), die verschiedene Eigenschaften der Tracks, den Energieablagerungen und der Abstimmung zwischen Track und Cluster [73] einbezieht. Drei Betriebspunkte Loose(80%), Medium(88%) und Tight(93%) werden definiert und geben die durchschnittliche Effizienz an, mit der ein interessantes Elektron die Identifikationskriterien erfüllt. Für diese Arbeit relevante Elektronen müssen die TightLH-Kriterien erfüllen. Zusätzlich wird ein LooseAndBLayerLH-Kriterium erfordert, was eine Loose-Selektion mit der Anforderung eines potentiellen Hits im IBL (s. Abs. 3.2.1) verbindet, um die Wahrscheinlichkeit von Fehlidentifikation zu reduzieren. Außerdem zeigen die interessanten Elektronen üblicherweise weniger zusätzliche Tracks oder Energieablagerungen in ihrer unmittelbaren Umgebung [72], weswegen die Elektronkandidaten das PromptLeptonImprovedTight-Isolationskrietrium [75] erfüllen müssen. Weiterhin müssen die Elektronen dem primären Wechselwirkungsvertex entstammen, was durch zusätzliche Bedingungen an die Trackparameter gewährleistet wird. Für Monte-Carlo Simulationen werden Skalierungsfaktoren (SF) verwendet, um die Rekonstruktions-, Identifikations- und Isolationseffizienzen an die Effizienzen der experimentellen Daten anzupassen.

3.3.3 Myonen

Myonen sind minimal ionisierende Teilchen mit einer mittleren Lebensdauer von etwa $2\,\mu\text{s}$, sodass sie den ATLAS-Detektor vollständig durchdringen. Die in dieser Arbeit verwendeten Myonen sind sogenannte Combined-Myonen, da sie aus dem Abgleich von Spuren aus dem inneren Detektor und Myon-System unter Einbeziehung des Energieverlusts im Kalorimeter rekonstruiert werden [76]. Wie bei den Elektronen werden die Myon-Kandidaten anhand von

Betriebspunkten mit unterschiedlicher Effizienz selektiert, um zwischen interessanten Myonen und solchen aus hadronischen Zerfällen zu unterscheiden. Für diese Arbeit wird der Medium-Betriebspunkt für die Identifikation gewählt. Als Isolationskriterium wird der Betriebspunkt PromptLeptonImprovedTight [75] verwendet. Analog zu den Elektronen werden Skalierungsfaktoren (SF) ermittelt, um die Rekonstruktions-, Identifikations- und Isolationseffizienzen der Monte-Carlo Simulationen an die Effizienzen der experimentellen Daten anzupassen.

3.3.4 Jets

Jets sind gebündelte Strahlen von Hadronen, die aus einzelnen Quarks und Gluonen erzeugt werden. Zunächst rekonstruiert der Particle-Flow Algorithmus [77] die einzelnen Teilchen der primären Wechselwirkung anhand einer Kombination von Tracks, die bestimmte Qualitätskriterien erfüllen [77], und den schon für die Elektron-Rekonstruktion beschriebenen Energieclustern der Kalorimeter. Zudem wird ein Subtraktionsschema angewendet, um Überschneidungen der Impulsmessung im ID und Energiemessung in den Kalorimetern zu verhindern [77]. Die ausgegebenen ParticleFlow Objekte werden durch den anti-k_t Algorithmus [78] mit einem Radiusparameter von $R = 0{,}4$ zu Jets gruppiert. Energieablagerungen durch Pile-Up Teilchen, die auf der Fläche des Jets erwartet werden, werden abgezogen. Anschließend werden die Energieskala (engl. *jet energy scale* JES) und Energieauflösung (engl. *jet energy resolution* JES) der Jets kalibriert [79] und weitere Kriterien definiert, um Untergrund-Jets abzuweisen, die nicht aus der Proton-Proton-Kollision entstehen oder aus Detektorrauschen hervorgehen [80]. Pile-up Jets, die nicht aus der primären Kollision stammen, werden mit dem Jet-Vertex-Tagger (JVT) [81] unterdrückt.

3.3.5 b-tagging

Das b-tagging ist eine Methode zur Identifikation von Jets, die aus einem b-Quark entstehen. Sogenannte b-Hadronen haben eine vergleichsweise lange Lebensdauer von etwa 1,5 ps, wodurch sie eine Flugstrecke von einigen Millimetern im Detektor zurücklegen können. In dieser Arbeit wird der DL1r Algorithmus [82] für das b-tagging verwendet. Darin werden Informationen aus Trackparametern von Jets und der Rekonstruktion der Zerfallskette des b-Quarks in ein tiefes neuronales Netzwerk eingegeben. Für jeden Eingabe-Jet gibt DL1r die Wahrscheinlichkeit aus, dass ein Jet ein b-Jet, c-Jet oder ein Jet leichter Quarks (u,d) ist. Diese Wahrscheinlichkeiten gehen in die DL1r-b-tagging Diskriminante ein. Daraus werden verschiedene Betriebspunkte definiert, die die durchschnittliche Effizienz angeben, mit der ein b-Jet identifiziert wird [83]. In dieser Arbeit wird der Betriebspunkt mit einer durchschnittlichen Effizienz von 77% verwendet, was einer Fehlidentifikation von etwa 20% der c-Jets und weniger als 1% der leichten Jets als b-Jets in $t\bar{t}$-Ereignissen entspricht. Auch hier werden Skalierungsfaktoren berechnet, die die Effizienz der simulierten Daten an die experimentellen Daten anpassen.

3.3.6 Fehlender Transversalimpuls

Der fehlende transversale Impuls ist ein Messwert für den transversalen Impuls von Teilchen bei Proton-Proton Kollisionen, die nicht im Detektor gemessen werden können. Beispiele hierfür sind Neutrinos oder hypothetische Teilchen in Erweiterungen des Standardmodells. Bei einer Kollision zweier Protonen ist der Impuls aller produzierten Teilchen in der transversalen Ebene

erhalten. Der Anteil des fehlenden Transversalimpulses ist dann definiert als die negative Vektorsumme über alle rekonstruierten Myonen μ, Elektronen e und Jets j, sowie ein $soft$-Term, welcher alle Particle-Flow Tracks beinhaltet, die zum primären Vertex gehören und keinem rekonstruierten Objekt zugeordnet werden kann [84]:

$$\vec{p_{\mathrm{T}}}^{\,\mathrm{miss}} = -\sum_{\mu} \vec{p_{\mathrm{T}}}^{\,\mu} - \sum_{e} \vec{p_{\mathrm{T}}}^{\,e} - \sum_{j} \vec{p_{\mathrm{T}}}^{\,j} - \vec{p_{\mathrm{T}}}^{\,soft}. \tag{3.4}$$

Im weiteren Verlauf dieser Arbeit wird der Betrag des fehlenden Transversalimpulses $E_{\mathrm{T}}^{\mathrm{miss}}$ verwendet, die unter Vernachlässigung der Masse auch fehlende Transversalenergie bezeichnet wird.

3.3.7 Ambiguitäten in der Rekonstruktion

Die Rekonstruktion der beschriebenen physikalischen Objekte ist nicht in jedem Fall eindeutig. Elektronen und Jets werden im elektromagnetischen Kalorimeter rekonstruiert, sodass Elektronen im Jet-Container und Jets, die die Elektron-Rekonstruktionskriterien erfüllen, im Elektron-Container doppelt gezählt werden. Weiterhin können naheliegende Elektronen und Jets nicht ausreichend in Kalorimeter-Cluster isoliert werden, sodass es zu fehlerhafter Rekonstruktion kommt. Daher werden sog. *overlap removal*-Algorithmen verwendet.

Teilen ein Elektron und ein Myon eine Spur im inneren Detektor, wird das Elektron entfernt. Ein (nicht b-tagged) Jet innerhalb eines Abstandes von $\Delta R \leq 0{,}2$ zu einem Elektron wird entfernt. Elektronen in einem Abstand $\Delta R \leq 0{,}4$ zu Jets werden entfernt, um Elektronen aus hadronischen Zerfällen zu reduzieren. Jets, die weniger als drei Spuren im inneren Detektor aufweisen und vom einem gleichen Vertex wie ein Myon stammen, werden entfernt. Myonen in einem Abstand $\Delta R \leq 0{,}4$ zu den übrigen Jets werden entfernt.

4 Grundlagen des Maschinellen Lernens

Die Entwicklung des Maschinellen Lernens (engl. *Machine Learning* ML) hat in unserem täglichen Leben eine große Bedeutung. Die Grundidee des maschinellen Lernens ist, dass ein Algorithmus Muster bzw. Zusammenhänge aus einem gegebenen Datensatz lernt und anschließend eine korrekte Ausgabe für neue, zuvor unbekannte Datensätze liefert. Von den unzähligen Einsatzmöglichkeiten [85,86] seien Bild-, Sprach- und Texterkennung oder verschiedenste medizinische Anwendungen genannt. Auch ist das heute sehr populäre Gebiet der künstlichen Intelligenz (KI) nah verwandt mit dem maschinellen Lernen.

Experimente der Hochenergiephysik (HEP) produzieren eine große Anzahl komplexer und hochdimensionaler Daten. Um diese effizient auszuwerten, werden fortgeschrittene Analysemethoden benötigt. Maschinelles Lernen und insbesondere Tiefes Lernen (engl. *Deep Learning* DL) haben sich über die Jahre durch den Fortschritt leistungsfähigerer Computer sowie kontinuierlicher Forschung als fundamentale Werkzeuge etabliert. Im folgenden Kapitel soll kurz auf die grundlegenden Konzepte des maschinellen Lernens eingegangen werden. Anschließend werden neuronale Netzwerke mit dem Feld des Tiefen-Lernens erläutert. Zuletzt soll kurz auf die Herausforderungen des Maschinellen Lernens hingewiesen werden. Die Grundlagen, sofern nicht explizit angegeben, basieren auf den Quellen [87–89].

4.1 Konzepte des maschinellen Lernens

Algorithmen des maschinellen Lernens kann man grob in die beiden Kategorien *überwachtes Lernen* und *unüberwachtes Lernen* einteilen. Beim unüberwachten Lernen erkennt ein Modell Regelmäßigkeiten bzw. Strukturen innerhalb der Eingabedaten, ohne eine konkrete Ausgabe (Lösung) vorgegeben zu bekommen. Beim überwachten Lernen hingegen gibt es eine Eingabe X sowie eine Ausgabe Y und das Ziel besteht darin, die Abbildung von $X \rightarrow Y$ zu lernen. Es gibt weitere Techniken, wie das *teilweiseüberwachte Lernen*, *aktive Lernen* oder *bestärkende Lernen*. Das in dieser Arbeit benutzte Modell wird dem überwachtem Lernen zugeordnet, weshalb einige Grundlagen im Folgenden geklärt werden.

Das Konzept des überwachten Lernens kann als eine Suche nach einer Funktion $f : \boldsymbol{X} \rightarrow \boldsymbol{Y}$, welche die Eingabedaten $\boldsymbol{X}$ auf die Ausgabedaten $\boldsymbol{Y}$ abbildet, formuliert werden. Dabei sind:

$$\boldsymbol{X} = \begin{bmatrix} x_1^{(1)} & \dots & x_m^{(1)} \\ \vdots & \ddots & \\ x_1^{(N)} & & x_m^{(N)} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} y_1^{(1)} & \dots & y_l^{(1)} \\ \vdots & \ddots & \\ y_1^{(N)} & & y_l^{(N)} \end{bmatrix} \tag{4.1}$$

Matrizen aus N Zeilen, was die Anzahl der Ereignisse (Events) darstellt und damit der Größe des Datensatzes entspricht. Im Kontext dieser Arbeit wären dies z.B. Ereignisse, in denen ein $t\bar{t}$-Paar aus einer Teilchenkollision erzeugt wird und über einen bestimmten Zerfallskanal zerfällt (s. Abs. 2.2.1). Jede der individuellen Ereignisse in $\boldsymbol{X}$ ist ein Vektor aus m Merkmalen bzw. Eigenschaften, welche die erforderlichen Informationen zur Lösung der Lernaufgabe darstellen. Im eben genannten Beispiel könnten dies etwa rekonstruierte kinematische Eigenschaften der Teilchen im Endzustand des Zerfalls sein. Parallel dazu setzt sich die Zielmatrix $\boldsymbol{Y}$ aus N sogenannten *Labeln* zusammen, die die Zielwerte der jeweiligen Ereignisse darstellen (im Kontext dieser Arbeit könnten dies Verteilungen auf dem Teilchen-Level sein). In der sogenannten *Trainingsphase* optimiert der Algorithmus eine vorab gewählte Metrik $\mathcal{L}(f(\boldsymbol{X}),\boldsymbol{Y})$: die sogenannte

Verlustfunktion (auch Kostenfunktion oder Fehlerfunktion genannt). Diese ist ein Maß für die Genauigkeit der Vorhersage des Modells aus den m Merkmalen im Vergleich zu den vorgegebenen Zielwerten und wird über den gesamten Datensatz während der Trainingsphase berechnet. Letztendlich ist das Ziel des Trainings, genaue Vorhersagen auf unbekannte Daten zu treffen. Die Generalisierungsfähigkeit ist eine zentrale Voraussetzung für den Einsatz eines ML Modells.

Die bekanntesten zwei Formen des überwachten Lernens sind die *Klassifikation* und die *Regression*. Erstere probiert Datenpunkte einer bestimmten Gruppe zuzuordnen, indem aus einer festgelegten Anzahl von Kategorien die zutreffende gewählt wird. Dies findet in der Hochenergiephysik beispielsweise Verwendung in der Teilchenidentifikation [23]. Die Regression hingegen liefert einen kontinuierlichen Wert bzw. eine quantitative Ausgabe und wird bei dem Modell dieser Arbeit verwendet. Die am häufigsten verwendeten Verlustfunktionen in einer Regressionsaufgabe sind der mittlere absolute Fehler (engl. *mean absolute error* MAE) und die mittlere quadratische Abweichung (engl. *mean squared error* MSE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{true} - y_{pred}|, \tag{4.2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_{true} - y_{pred})^2, \tag{4.3}$$

wobei N wieder die Anzahl der Daten im Trainingssatz ist, über welche die Verlustfunktion berechnet wird und „pred“ für die Vorhersage des Modells steht. Beide Funktionen liefern positive Werte, wobei 0 den „Optimalwert“ , also eine völlige Übereinstimmung von Vorhersage und Vorgabe angibt. Die Wahl der Verlustfunktion ist ein entscheidender Schritt für die Optimierung des Models und im Allgemeinen abhängig von der Art der Testdaten und Anforderungen des Problems.

4.2 Neuronale Netzwerke

Die Idee künstlicher neuronaler Netzwerke, inspiriert von der Funktionsweise des menschlichen Gehirns, geht auf McCulloch und Pitts in 1943 zurück [90]. In neuronalen Netzwerken werden die Neuronen (auch Knoten oder engl. *nodes* genannt) in Schichten (engl. *layers*) zusammengefasst, wobei die Eingabeschicht die erste Ebene ist, die die Daten aufnimmt und die Ausgabeschicht die Letzte ist und die Ergebnisse liefert. Die Zahl der Neuronen pro Schicht sowie die Gesamtzahl der Schichten in einem neuronalen Netzwerk kann je nach Aufgabenstellung stark variieren. Allgemein können künstliche neuronale Netzwerke in Feedforward-Nerzwerke (auch vorwärtsgerichtete Netzwerke genannt), bei denen die Informationen von der Eingabe in Richtung Ausgabe geleitet werden und in rekurrente Netzwerke, bei denen Schleifen enthalten sind und damit Informationen über die Zeit gespeichert und verarbeitet werden können, eingeteilt werden. Das einfachste Modell ist das in 1957 vorgestellte Perzeptron [91], was aus einer einzelnen Schicht mit genau einem Neuron besteht.

In diesem mathematischen Modell empfängt das Neuron m Eingabewerte x_i und gewichtet diese mit w_i. Ist die Summe der Gewichte größer als ein Schwellenwert b (auch Bias genannt), feuert das Neuron. Dies kann man auch so formulieren:

$$\hat{y} = H(\sum_{i=1}^{m} x_i w_i + b), \tag{4.4}$$

wobei $\left(H(z) = \begin{cases} 1, & z > 0 \\ 0, & \text{sonst} \end{cases} \right)$ die Heaviside-Funktion ist und die beiden Zustände des Neurons - **An** oder **Aus** - angibt. In diesem simplen Modell führen jedoch kleine Änderungen der Eingangssignale zu großen Änderungen in der Ausgabe, da die Stufenfunktion nur zwei Werte annehmen kann. Zudem kann die Funktion bei Null nicht differenziert werden, was Probleme für einige Lern-Algorithmen (siehe Abs. 4.2.2) ergibt. Um diesem Problem entgegen zu wirken, ersetzt man $H(z)$ mit einer nicht-linearen, sogenannten Aktivierungsfunktion. Damit kann nach dem universellen Approximationstheorem [92] eine beliebig komplexe Funktion durch ein neuronales Netzwerk mit einer gegebenen Struktur von tiefen Schichten und Neuronen approximiert werden. Die Wahl der passenden Aktivierungsfunktion ist u.a. abhängig vom Wertebereich der Eingabedaten und wird i.A. empirisch ermittelt. Eine gängige Wahl der Aktivierungsfunktion bei Regressionsproblemen stellt die *rectified linear unit* ReLU dar:

$$ReLU(z) = \begin{cases} z, & z \geq 0 \\ 0, & \text{sonst} \end{cases}, \tag{4.5}$$

welche linear für positive Argumente und sonst Null ist. Andere populäre Aktivierungsfunktionen sind die Tangens-Hyperbolicus-Funktion $\tanh(z)$ oder die Sigmoid-Funktion $\sigma(z)$, gegeben durch:

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \tag{4.6}$$

Im Gegensatz zur ReLU-Funktion sind die Ausgaben dieser Funktionen beschränkt. Ein Vorteil der ReLU-Funktion ist, dass sie rechnerisch effizient berechnet werden kann und eine schnelle Konvergenz beim Training erreicht. Das Problem verschwindender Gradienten in der Nähe der Extrempunkte von $\sigma(z)$ und $\tanh(z)$ kann mit der ReLU-Funktion vermieden werden. Insbesondere in tiefen Netzwerken (siehe Abs. 4.2.1) kann ReLU eine vergleichsweise bessere Leistung erzielen [93]. Ein Nachteil kann das sogenannte *dying ReLU*-Problem sein, wobei Neuronen, welche negative Eingaben bekommen und damit Null ausgeben, somit nicht zum Lernprozess beitragen. Dies berücksichtigen Erweiterungen der ReLU-Funktion, wie z.B. *leaky ReLU*:

$$leaky\ ReLU(z) = \begin{cases} z, & z \geq 0 \\ \omega \cdot z, & \text{sonst} \end{cases}, \tag{4.7}$$

welche für Werte $z < 0$ einen vorab festgelegten, kleinen linearen Anstieg ω besitzt, sodass es Ausgaben von ungleich Null für negative Aktivierungen gibt.

4.2.1 Deep Learning

Das tiefe Lernen (engl. *Deep Learning* DL) [94] ist eine Form des maschinellen Lernens, welche auf künstlichen neuronalen Netzwerken mit mehreren „verborgenen Schichten“ (engl. *hidden layers*) basiert. Über die letzten 15 Jahre hat das Feld durch leistungsstärkere Computer eine große Bedeutung in der Wissenschaft gewonnen und in vielen Anwendungen der Hochenergiephysik Benutzung gefunden. So beispielsweise sind die DL1-Algorithmen [95] für das Tagging von b-Quarks basiert auf vorwärtsgerichteten tiefen neuronalen Netzen und gehören zu den leistungsfähigsten Taggern von Run-2-Analysen der ATLAS Kollaboration. Vielversprechende Entwicklungen in diesem Gebiet zeigen die GN-b-Tagger [96], welche auf *graph neural networks* (GNN) [97] basieren.

Je nach Komplexität kann ein einziges tiefes Netzwerk aus Millionen von freien (internen) Parametern Θ, bestehend aus Gewichten und Biases, aufgebaut sein. Ziel ist, diese Parameter im Modell mit etwa gleich vielen Beispielen im Trainingsdatensatz zu optimieren. Beginnend mit den Eingabedaten können in jeder verborgenen Schicht die Werte der vorherigen Schicht kombiniert werden, um immer komplexere Funktionen der Eingabe zu lernen. Aufeinander folgende Schichten entsprechen zunehmend abstrakteren Repräsentationen. Betrachtet wird ein tiefes neuronales Netz, wobei die Aktivierung y_j^i des i-ten Neuron in der j-ten Schicht gegeben ist durch:

$$\hat{y}_j^i = f\left(\sum_k w_{jk}^i \hat{y}_{j-1}^k + b_j^i\right), \tag{4.8}$$

(vgl. Gl. 4.4), wobei f die Aktivierungsfunktion, w_{jk}^i das Gewicht, das das k-te Neuron der $j-1$-ten Schicht mit dem i-ten Neuron der j-ten Schicht verbindet, $\hat{y}_{j-1}^k$ die Aktivierung des k-ten Neurons der $(j-1)$-ten Schicht und b_j^i der Bias des Neurons ist. Es gibt mehrere Arten von Algorithmen, um das dargestellte Problem zu lösen, was im folgenden Abschnitt thematisiert wird.

4.2.2 Lern-Algorithmen

In diesem Kapitel wird näher auf den Lernprozess eines neuronalen Netzwerks, insbesondere auf das Lösen von Gleichung 4.8 eingegangen, wie es im neuronalen Netzwerk dieser Arbeit verwendet wird. Wie im vorherigen Abschnitt erläutert, erfolgt das Lernen durch die Anpassung einer Reihe von Netzwerk-internen Parametern Θ, die eine ausgewählte Verlustfunktion $\mathcal{L}(\Theta)$ minimieren. Auch wenn dies analytisch nur selten möglich ist, gibt es verschiedene Optimierungsstrategien, welche eine gute Approximation liefern. Diese können je nach Art der berücksichtigten Informationen für die Parameteranpassung kategorisiert werden in Algorithmen nullter, erster und zweiter Ordnung, benannt nach der Ordnung der verwendeten Ableitung der Verlustfunktion. Die am häufigsten verwendete Optimierungsmethode ist das Gradientenabstiegsverfahren (engl. *gradient descent* GD), was einem Algorithmus erster Ordnung entspricht. Hier wird eine Funktion $\mathcal{L}(\Theta)$ in Bezug auf die Modellparameter Θ minimiert, in dem schrittweise in Richtung des negativen Gradienten (also in Richtung des steilsten Abstiegs)

der Verlustfunktion vorangeschritten wird. Wie in Abschnitt 4.1 beschrieben, handelt es sich bei den Verlustfunktionen oft um Summen der Form:

$$\mathcal{L}(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i(\Theta), \tag{4.9}$$

wobei N die Gesamtzahl der eingegebenen Trainingsdaten ist, beispielsweise die Anzahl an $t\bar{t}$ Ereignissen eines simulierten Datensatzes. Die Gesamtzahl der anpassbaren Parameter Θ in einem neuronalen Netzwerk mit L vollständig miteinander verbundenen, verborgenen Schichten lässt sich berechnen durch:

$$N_\Theta = \sum_{i=1}^{L}((n_{i-1}\cdot n_i) + n_i), \tag{4.10}$$

wobei n_i die Anzahl der Neuronen in der $i-$ten Schicht sind. Die Funktion $\mathcal{L}_i$ in Gleichung 4.9 wird für jedes Ereignis $i = 1,...N$ der Trainingsdaten berechnet. Um eine solche Funktion zu minimieren, wendet ein GD-Algorithmus folgenden Minimierungsschritt iterativ an, um die Verlustfunktion entlang des steilsten Abstiegs zu optimieren:

$$\Theta^{n+1} = \Theta^n - \alpha\cdot\frac{1}{N}\sum_{i=1}^{N}\nabla\mathcal{L}_i(\Theta^n). \tag{4.11}$$

Ein vollständiger Durchlauf des Trainingsdatensatzes wird als *Epoche* bezeichnet. Typischerweise ist ein Training etwa 10-500 Epochen lang. α wird als Schrittweite bzw. oft als *learning rate* bezeichnet. Sie ist ein vorab gewählter Parameter des neuronalen Netzwerks und ist i.A. während der Trainingsphase nicht konstant. Eine zu hohe Lernrate könnte ein gutes lokales Minimum überspringen bzw. kann zu divergierendem Verhalten führen. Eine zu niedrige Lernrate kann den Trainingsprozesse verlangsamen oder in ungewollten lokalen Minima stecken bleiben.

Da in Gleichung 4.11 N Gradienten berechnet werden, wird mit hoher Genauigkeit in Richtung eines Minimums geschritten, jedoch kann dies für große Datensätze (mit vielen Ereignissen) rechnerisch sehr aufwendig werden. Um diesem Problem entgegenzuwirken, wurde das stochastische Gradientenverfahren (engl. *stochastic gradient descent* SGD) [98, 99] entwickelt. Hier wird ein einzelnes Ereignis aus dem Datensatz zufällig bei der Anpassung gewählt, um den Gradienten der Verlustfunktion zu berechnen. Dies ist rechnerisch sehr effizient, jedoch entsteht dadurch eine hohe Varianz in der Anpassung und das Modell „wandert“ nicht gleich in die Richtung eines Minimums. Eine hilfreiche und heute fast ausschließlich benutzte Variante ist, dass statt über nur einem Datenpunkt (Ereignis) über eine Gruppe von zufällig gewählten Ereignissen iteriert wird. Im Vergleich zu Gleichung 4.11 wird damit folgende Näherung vorgenommen:

$$\frac{1}{N}\sum_{i=1}^{N}\nabla\mathcal{L}_i(\Theta^n) \approx \frac{1}{|\mathcal{B}|}\sum_{j\in\mathcal{B}}\nabla\mathcal{L}_j(\Theta^n). \tag{4.12}$$

Hier bezeichnet $\mathcal{B} \subset \{1,...,N\}$ die zufällig zusammengestellte Indexmenge aus Ereignissen, welche als *Batch* bezeichnet wird. Werte für diese sind üblicherweise $N_B = 2^n$, wobei n eine

natürliche Zahl ist, und liegen oft im Bereich $N_B = (32 \text{ bis } 1024)$. Dies erlaubt eine effiziente und meistens genaue Approximation. Es existieren zahlreiche Erweiterungen des SGD. Häufig verwendete sind beispielsweise RMSPRop [100] und ADAM (Adaptive Moment Estimation) [101], welche adaptive Lernraten verwenden, um ein schnelleres und stabileres Training zu erreichen.

Für das Modell dieser Arbeit wird der ADAM-Algorithmus [101] verwendet. ADAM beruht auf Gleichung 4.12, nimmt jedoch Erweiterungen vor. ADAM aktualisiert den exponentiell gewichteten, gleitenden Durchschnitt[1)] des Gradienten der Verlustfunktion m_Θ^n und des quadrierten Gradienten v_Θ^n für jede Iteration n. Zwei Abklingraten β_1 und β_2 werden definiert, die steuern, wie stark frühere Gradienteninformationen bei der Berechnung von m_Θ^n und v_Θ^n gewichtet werden:

$$m_\Theta^{n+1} = \beta_1 m_\Theta^n + (1-\beta_1)\nabla\mathcal{L}(\Theta^n) \tag{4.13}$$

$$v_\Theta^{n+1} = \beta_2 v_\Theta^n + (1-\beta_2)(\nabla\mathcal{L}(\Theta^n))^2 \tag{4.14}$$

Die Aktualisierung der Parameter Θ pro Iteration n ist damit:

$$\Theta^{n+1} = \Theta^n - \alpha \cdot \frac{\hat{m}_\Theta}{\sqrt{\hat{v}_\Theta + \epsilon}}, \qquad \text{mit } \hat{m}_\Theta = \frac{m_\Theta^{n+1}}{1-\beta_1^{n+1}},\ \hat{v}_\Theta = \frac{v_\Theta^{n+1}}{1-\beta_2^{n+1}}. \tag{4.15}$$

Standardwerte für den ADAM-Algorithmus sind eine Lernrate von $\alpha = 0{,}001$, sowie die Abklingraten $\beta_1 = 0{,}9$ und $\beta_2 = 0{,}999$. Der Wert $\epsilon = 10^{-8}$ verhindert eine Division durch Null.

Variablen, die den Aufbau eines neuronalen Netzwerks bestimmen und den Trainingsprozess steuern, werden *Hyperparameter* genannt. In diesem Kapitel wurden Verlustfunktion, Anzahl tiefer Schichten und Neuronen, Aktivierungsfunktion, Anzahl der eingegebenen Trainingsdaten, Batchgröße und die Wahl des Optimierungsalgorithmus mit zugehöriger Lernrate als Hyperparameter beschrieben. Eine Optimierung dieser Hyperparameter ist die zentrale Aufgabe bei der Erstellung eines Deep-Learning Modells. In Kapitel 6 wird das in dieser Arbeit verwendete neuronale Netzwerk anhand dieser Hyperparameter optimiert.

4.3 Herausforderungen des Maschinellen Lernens

Wie bei klassischen Analysemethoden kommt auf das maschinelle Lernen als datengestützte Technik eine Reihe von u.a. neuen Herausforderungen zu [23, 102]. Im Allgemeinen gilt, dass ein Lernalgorithmus genauere Vorhersagen liefert, umso mehr Lerndaten es zur Verfügung gestellt bekommt. Somit hängt die Qualität des Modells auch von der Qualität der Trainingsdaten ab. In der Hochenergiephysik dienen simulierte Kollisionsdaten standardmäßig als Eingabedaten für ein ML-Modell. Dabei muss vorausgesetzt werden, dass die Simulationen und die experimentellen Kollisionsdaten eine gute Übereinstimmung aufweisen und Modellierungsfehler in den

[1)] Ein exponentiell gleitender Durchschnitt gewichtet die Gradienten einer Folge von Iterationen so, dass die Gewichte exponentiell mit der Anzahl der Iterationen abnehmen, sodass neuere Gradienten stärker gewichtet werden als ältere.

Simulationen erkannt und als systematische Unsicherheiten berücksichtigt werden.

Auch wenn ein befriedigendes Ergebnis durch ein Modell erzielt wurde, ist durch die intrinsische Komplexität künstlicher neuronaler Netzwerke (insbesondere tiefer Netzwerke) eine direkte Einsicht in die Lösungsstruktur verhindert. Damit ist die Interpretierbarkeit und das Verständnis der Entscheidungsfindung innerhalb des Modells beeinträchtigt, was insbesondere wichtig ist, wenn das gleiche Modell für mehrere Anwendungen genutzt wird. Weiterhin ist die Optimierung der in Abschnitt 4.2.2 beschriebenen Hyperparameter eines Modells oft ein **trial and error**-Prozess ohne direkte Einsicht, wie diese die Leistung des Modells genau beeinflussen.

Bei der Integration von maschinellem Lernen in die HEP-Analyse muss das Modell, insbesondere die Verlustfunktion und Lern-Algorithmen, an die physikalischen Ziele, für die das Modell verwendet wird angepasst werden. Dazu muss die Behandlung der relevanten systematischen Unsicherheiten, die die Messung beeinflussen, gut verstanden sein. Da das maschinelle Lernen eine vielversprechende Erweiterung für viele Analysen und zukünftig unersetzlich wird, müssen die genannten Herausforderungen sorgfältig berücksichtigt werden.

5 Methodik

Im folgenden Kapitel werden die Methoden und Techniken beschrieben, die zur Untersuchung des neuronalen Netzwerks in Kapitel 6 und zur Analyse der Top-Yukawa-Kopplung in Kapitel 7 verwendet werden. Dabei wird zunächst kurz auf die verwendete Software eingegangen. Anschließend werden die in dieser Arbeit verwendeten Datensätze beschrieben. Nach einer Beschreibung der Ereignisselektion wird näher auf die Gewichtung der Monte-Carlo Ereignisse eingegangen.

5.1 Datensätze und Eventselektion

Software
Alle produzierten Datensätze für die Analyse dieser Arbeit werden in der Programmiersprache Python [103] verarbeitet. Diese werden in pyroot RDataFrame [104] eingelesen und die Ereignisse anhand bestimmter Kriterien selektiert. Die Eingabedaten für das neuronale Netzwerk (s. Kap. 6) werden direkt daraus gespeichert. Keras [105] ist eine Open Source Deep-Learning-Bibliothek zur Erstellung von tiefen neuronalen Netzwerken und wird zusammen mit der Berechnungsplattform TensorFlow [106] für die Erstellung des Modells benutzt. Der Umgang mit den Ausgabedaten des Modells erfolgt mit numpy [107] und die Erstellung der Diagramme mit pyroot [108] und matplotlib [109].

Hauptdatensätze
In dieser Arbeit werden simulierte Datensätze von dileptonischen $t\bar{t}$-Ereignissen verwendet. Ein zugehöriges Schema des Zerfallskanals ist in Abbildung 5.1 dargestellt. Mit geladenen Leptonen werden im weiteren Text Elektronen (Positronen) sowie Myonen (Anti-Myonen) gemeint.

Die experimentellen Kollisionsdaten basieren auf Proton-Proton-Kollisionen bei einer Schwerpunktenergie von $\sqrt{s} = 13$ TeV, welche im Zeitraum zwischen 2015 und 2018 (Run-2) vom ATLAS-Detektor aufgenommen wurden. Diese Daten entsprechen einer integrierten Luminosität von $140{,}1$ fb^{-1}. Die folgend beschriebenen Monte-Carlo-Simulationen werden für die Analyse von Signal- und Untergrundprozessen sowie zum Vergleich mit den experimentellen Daten produziert. Zusätzlich werden simulierte Daten sowohl für die Eingabedaten des Trainings, als auch für die Validierungs- und Testdaten des neuronalen Netzwerks verwendet. Der $t\bar{t}$-Datensatz, sowie der dominante Untergrundprozess der Einzel-Top-Produktion tW (siehe. Abb. 2.5) wird unter Verwendung von powheg [50–52] in NLO QCD-Genauigkeit simuliert und mit Pythia8 [57] für die Simulation von Parton-Schauer und Hadronisierung gekoppelt (s. Abs. 2.2.2 für Grundlagen der Ereignissimulation). Dies wird in Kombination mit der vollständigen ATLAS-Detektorsimulation basierend auf geant4 [110] generiert. Die simulierten Signal- und Untergrunddatensätze werden daher im Folgenden $t\bar{t}$ (PP8$_{\text{full}}$) und tW (PP8$_{\text{full}}$) genannt. Es wird sich auf den tW-Untergrund beschränkt, da andere Untergrundprozesse einen geringen Beitrag leisten und $t\bar{t} + tW$ die Datenstatistik ungefähr beschreibt (s. Tab 5.2). Die Verwendung offizieller ATLAS-Simulationen gewährleistet eine zuverlässige Modellierung der Eingabedaten für das neuronale Netzwerk und die daraus resultierende Vorhersage.

Alternative $t\bar{t}$-Datensätze
Alternative $t\bar{t}$-Datensätze werden produziert, um den Einfluss von Modellierungsunsicherheiten bei Partonschauer und Hadronisierung auf kinematische Variablen und die Verteilung des neuronalen Netzwerks zu untersuchen. Für die alternativen Datensätze wird die Berechnung des Matrix-Elements ebenfalls durch powheg [50–52] vorgenommen, jedoch wird dies mit zwei

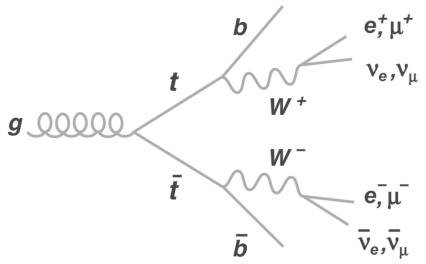

Abbildung 5.1: Feynman-Diagramm in führender Ordnung für den dileptonischen $t\bar{t}$-Zerfall. Endzustände mit Tau-Leptonen werden in dieser Arbeit nicht behandelt. Adaptiert von [111]

unterschiedlichen Generatoren für Partonschauer und Hadronisierung gekoppelt. Ein Datensatz wird wie der Hauptdatensatz mit POWHEG+PYTHIA8 [57] und ein weiterer mit POWHEG+HERWIG7 [58,59] generiert. Die Unterschiede der verwendeten Algorithmen sind in Abschnitt 2.2.2 diskutiert. Beide werden jedoch mit einer schnellen Simulation des ATLAS-Detektors durch ATLASFASTII [112] produziert. Damit kann zusätzlich ein Vergleich zwischen schneller ATLAS-Simulation und vollständiger ATLAS-Simulation des Hauptdatensatzes vorgenommen werden. Die beiden alternativen Datensätze werden von hier an mit $t\bar{t}$ (PP8_{fast}) und $t\bar{t}$ (PH7_{fast}) bezeichnet.

Eventselektion
In Abschnitt 3.3 wird die Rekonstruktion der physikalischen Objekte beschrieben, die den dileptonischen Endzustand definieren, doch nicht alle rekonstruierten Objekte werden für die Analyse verwendet. Um das Volumen der eingelesenen rekonstruierten Daten zu reduzieren, wird die Vorselektion TOPQ1 verwendet. Dabei werden nur Ereignisse ausgesucht, die mindestens einem Medium oder LHMedium (Likelihood basiert) identifiziertem Elektron (s. Abs. 3.3.2) mit einem Transversalimpuls von $p_T > 15$ GeV und Pseudorapidität $|\eta| < 2{,}5$ oder mindestens einem Myon mit $p_T > 15$ GeV und Pseudorapidität $|\eta| < 2{,}5$, was aus der kombinierten Spurinformation aus innerem Detektor und Myon-Detektor rekonstruiert ist, entsprechen. Die detaillierten Informationen der aus TOPQ1-produzierten xAOD[1)] Dateien sind über die sogenannten Produktionstags der verwendeten Datensätze p4514 (Monte-Carlo Simulationen) und p4513 (Daten) einsehbar.

Für die Eventselektion wird das *AnalysisTop*-Framework verwendet. Die Rekonstruktion eines dileptonischen Ereignisses (siehe Abb. 5.1) erfolgt unter diesen Bedingungen:

- zwei entgegengesetzt geladene isolierte Leptonen (genau $1e$,1μ), wobei das höherenergetische Lepton eine transversalen Impuls von $p_T > 28$ GeV und das verbleibende Lepton $p_T > 25$ GeV besitzen muss. Elektronen müssen im Bereich von $|\eta| < 2{,}47$ und außerhalb der Übergangsregion $1{,}37 < |\eta| < 1{,}52$ des ECALs liegen, Myonen im Bereich $|\eta| < 2{,}5$.
- zwei als b-Jets identifizierte Jets mit $p_T > 25$ und innerhalb $|\eta| < 2{,}5$.

Ereignisse mit genau einem Elektron und einem Myon, welche die oben beschriebenen Voraussetzungen erfüllen und durch den entsprechenden Elektron- und Myontrigger ausgelöst wer-

[1)] xAOD steht für Analysis Object Data und sind Dateien, die durch ROOT-Integration effizient bearbeitet werden können.

Tabelle 5.1: Übersicht der verwendeten Trigger für die Aufnahme von Elektron- und Myonereignissen in den Jahren 2015-2018. Für Details siehe [113–115].

Jahr	Elektron Trigger	Myon Trigger
2015	e24_lhmedium_L1EM20VH OR e60_lhmedium OR e120_lhloose	mu20_iloose_L1MU15 OR mu50
2016-18	e26_ltight_nod0_ivarloose OR e60_lhmedium_nod0 OR e140_lhloose_nod0	mu26_ivarmedium OR mu50

den, werden für die Analyse ausgewählt. Die verwendeten Trigger sind in Tabelle 5.1 genannt. Für eine genaue Beschreibung siehe [113–115]. b-Jets werden mit dem b-tagging Algorithmus DL1r [82] von Jets leichterer Quarks unterschieden und identifiziert (s. Abs. 3.3.5). Dies erfolgt mit einer Effizienz von 77%. Durch die beiden Neutrinos im Endzustand des Zerfalls existiert zusätzlich ein fehlender Transversalimpuls $E_{\mathrm{T}}^{\mathrm{miss}}$ (s. Abs. 3.3.6). Eine Identifizierung des Top- und Anti-Top-Quarks ist nicht nötig, da in dieser Arbeit nur die invariante Masse der $t\bar{t}$-Paare von Bedeutung ist. Deswegen werden die beiden energiereichsten b-Quarks als die Quarks des zerfallenden $t\bar{t}$-Systems gewählt.

5.2 Gewichtung der Ereignisse

5.2.1 Monte-Carlo Gewichtung

Es werden üblicherweise so viele Monte-Carlo Ereignisse simuliert, wie es die Rechenkapazität bzw. die Einschränkung der Rechenzeit erlaubt. So kann die Zahl der generierten Ereignisse je nach physikalischem Prozess von der Anzahl der gemessenen Kollisionsdaten abweichen. Für die oben genannten Datensätze werden jeweils etwa acht Millionen Ereignisse simuliert, die den definierten Selektionskriterien entsprechen. Um die Simulationensereignisse mit den Daten vergleichen zu können, muss man diese nach Gleichung 3.1 auf die Luminosität und den Wirkungsquerschnitt der verwendeten Daten normieren. Pro simuliertem Ereignis i wird ein Monte-Carlo Gewicht w_i wie folgt berechnet:

$$wi^{\mathrm{sim}} = w_i \cdot \sigma_{MC} \cdot L_{\mathrm{int}} \cdot k_{\mathrm{faktor}}. \tag{5.1}$$

Dabei ist σ_{MC} der von POWHEG [50–52] berechnete Wirkungsquerschnitt berechnet in NLO, der k_{faktor} das Verhältnis vom Wirkungsquerschnitt berechnet bei NNLO+NNLL[2)] zu NLO, L_{int} die integrierte Luminosität der in der Analyse benutzen Daten und w_i der Anteil der individuellen Ereignisgewichte des Generators vom gesamten Datensatz. Bei der Modellierung eines Ereignisses durch den Generator sowie bei der Modellierung des Detektors können Fehler bei (z.B.) der Identifikation oder beim Triggern von Teilchen auftreten. Um diese auszugleichen, kann

[2)] NNLO berücksichtigt Korrekturen der zweitnächsten führenden Ordnung (engl. *Next-to-Next-to-Leading Order*). NNLL bezeichnet die *Next-to-Next-to-Leading-Logarithmic*-Ordnungskorrekturen.

Tabelle 5.2: Übersicht der erwarteten $t\bar{t}$ Ereignisse für Monte-Carlo Simulation und Kollisionsdaten für die Jahre 2015-2018. Die Simulation wird mit POWHEG+PYTHIA8 (PP8) generiert. Dazu ist der dominante Untergrundprozess tW dargestellt. Der Index „full" steht für die volle Detektorsimulation mit GEANT4. Für alle simulierten Ereignisse ist die statistische Unsicherheit gegeben.

Datensatz	2015/16 (36,5 fb^{-1})	2017 (44,5 fb^{-1})	2018 (59 fb^{-1})	Run-2 (140 fb^{-1})
$t\bar{t}$ (PP8$_{\text{full}}$)	81 350±55	96 550±60	127 040±70	304 950±110
tW (PP8$_{\text{full}}$)	2590±15	3130±20	4100±25	9810±35
Daten	83 950	98 550	129 730	312 230

die Effizienz des betreffenden Parameters in Monte-Carlo-Simulationen und Daten verglichen werden. Die daraus berechneten Verhältnisse für die Modellierungsfehler - Skalierungsfaktoren (SF) genannt - werden auf die Selektion der Monte-Carlo Ereignisse (die Ereignisse auf dem Generator-Level bleiben unverändert) angewendet, um eine bessere Übereinstimmung mit den realen Daten zu erreichen. Im Fall der hier benutzten simulierten Daten werden Skalierungsfaktoren für Leptonen (s. Abs. 3.3.2), b-Tagging (s. Abs. 3.3.5) und dem Jet-Vertex-Tagger (s. Abs. 3.3.4)berücksichtigt. Diese müssen mit in die Gewichtung eingehen und werden multiplikativ zu w_i^{sim} hinzugefügt. Das simulierte *pile-up*-Profil [116] wird an die gemessenen Daten angepasst und geht außerdem in die folgende Gesamtgewichtung ein:

$$w_i^{\text{ges}} = w_i^{\text{sim}} \cdot \prod_k \text{SF}_k \cdot w_i^{\text{pile-up}}. \tag{5.2}$$

Die erwartete Anzahl an Ereignissen aller Datensätze nach dieser Normierung ist in Tabelle 5.2 für die einzelnen Jahre des Run-2-Zeitraums (2015-2018) und der Gesamtzahl gegeben. Es wird neben den $t\bar{t}$-Daten der dominante Untergrundprozess der Einzel-Top-Produktion tW angegeben (für Details siehe Abs. 5.1), welcher nur etwa 3% der gesamten Ereignisse ausmacht. Die statistischen Unsicherheiten der simulierten Daten berechnen sich durch $\delta w_i = \sqrt{\sum_n (w_i^{\text{ges}})^2}$ und sind ebenfalls angegeben.

5.2.2 Elektroschwache Korrektur und Top-Yukawa Kopplung

Teil dieser Arbeit ist, eine Motivation für die Messung der Top-Yukawa-Kopplung im dileptonischen Kanal zu geben (S. Kap. 7). Dies wird anhand einer Analyse an der vom neuronalen Netz berechneten $m_{t\bar{t}}$-Verteilung veranschaulicht (s. Abs. 7.2). Dafür müssen zunächst Korrekturterme der elektroschwachen Wechselwirkung bei der $t\bar{t}$-Produktion berücksichtigt werden. Diese sind in den oben genannten Ereignis-Generatoren nicht implementiert, da sie erst in Ordnung $\alpha_s^2\alpha_{\text{weak}}$ merklich in den Wirkungsquerschnitt der $t\bar{t}$-Produktion eingehen und damit einen geringen Beitrag zum Gesamtwirkungsquerschnitt leisten. Die dazugehörigen Feynman-Diagramme sind in Abbildung 5.2 gezeigt, wobei Γ in diesem Kontext für ein virtuelles Higgs-Boson steht. Die elektroschwachen Korrekturen werden mit dem HATHOR v2.1-b3 Paket [117] als Funktion der invarianten Masse des Top-Quark-Paares $m_{t\bar{t}}$ und des Streuwinkels des Top-Quarks relativ zur Strahllinie $\cos\theta^*$ berechnet. Genauer wird ein Skalierungsfaktor $Y_t = g_t/g_t^{\text{SM}}$ aus den Wirkungsquerschnitten $\sigma(q\bar{q} \to t\bar{t})$ und $\sigma(g\bar{g} \to t\bar{t})$ als ein freier Parameter

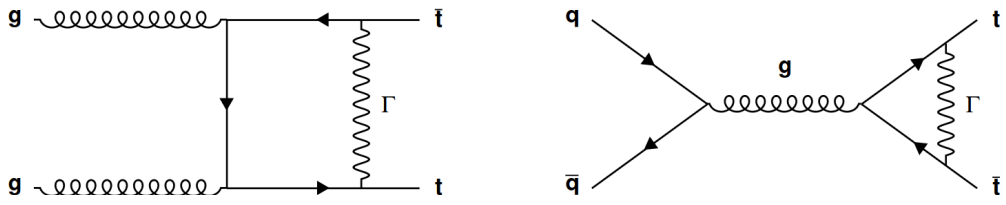

Abbildung 5.2: Beispieldiagramme für die $t\bar{t}$-Produktion durch Gluon-Fusion und Quark-Paarvernichtung mit den virtuellen Korrekturen zwischen den Top-Quarks. Im Kontext dieser Arbeit entspricht Γ dem Austausch eines virtuellen Higgs-Bosons. Grafik aus [19]

(z.B. für die Werte $Y_t = 0,1,2,...$) bestimmt, wobei g_t^{SM} für den Wert des Standardmodells nach Gleichung 2.9 steht. Die durch POWHEG [50–52] generierten Ereignisse werden durch das Verhältnis:

$$R(m_{t\bar{t}},\ \cos\theta^*, p) = \frac{\sigma_{\mathrm{H}}(m_{t\bar{t}},\ \cos\theta^*, p, \mathrm{Born} + Y_t + \mathrm{EW})}{\sigma_{\mathrm{H}}(m_{t\bar{t}},\ \cos\theta^*, p, \mathrm{Born})} \tag{5.3}$$

umgewichtet für verschiedene Werte von Y_t, wodurch kinematische Verteilungen in Abhängigkeit vom Wert für Y_t analysiert werden können. Weiterhin stellen $p = (gg,u\bar{u},d\bar{d})$ die verschiedenen Anfangszustände da. „Born“ steht für die führende Ordnung der Störungsrechnung. Der Austausch eines Higgs-Bosons zwischen dem Top- und Anti-Top-Quark ($\Gamma = H$ in Abb. 5.2) führt zu einer quadratischen Abhängigkeit der Top-Yukawa-Kopplung vom Wirkungsquerschnitt, was die Form von kinematischen Verteilungen erheblich verändern kann und u.a. für die statistische Analyse in Abschnitt 7.2 ausgenutzt wird. Terme, die linear von Y_t abhängen sind vernachlässigbar klein.

6 Aufbau und Optimierung des neuronalen Netzwerks

Die Rekonstruktion der invarianten Masse $m_{t\bar{t}}$ von $t\bar{t}$-Systemen ist in vielen Analysen der Top-Quark-Physik von zentraler Bedeutung. Hier ist sie motiviert durch die Bestimmung der Top-Yukawa Kopplung, die anhand der rekonstruierten Verteilung $m_{t\bar{t}}$ bestimmt werden kann (s. Kap. 7). Eine vollständige analytische Rekonstruktion der invarianten Masse im dileptonischen Zerfallskanal (s. Abb. 5.1) ist aufgrund der zwei Neutrinos im Endzustand nicht möglich. Es existieren Näherungsmethoden, welche unter Beschränkung kinematischer Größen wie der invarianten Massen des W-Bosons und der Top-Quarks lösbare Gleichungssysteme aufstellen [21, 22]. Dies motiviert dennoch die Möglichkeit, die Massenrekonstruktion als Regressionsproblem aufzufassen und unter Benutzung von *Deep-Learning*-Techniken (s. Abs. 4.2.1) aus dem Bereich des maschinellen Lernens zu lösen.

In Abschnitt 6.1 wird zunächst das optimierte neuronale Netzwerk vorgestellt, an dem die Hyperparameter (s. Ende von Abs. 4.2.2) des Modells systematisch optimiert werden. Die Ergebnisse dieser Optimierung sind in Abschnitt 6.2 erläutert. Die ausgegebene Massenverteilung des Modells wird in Abschnitt 6.3 analysiert. Die physikalischen Eingabegrößen des Modells werden in Abschnitt 6.3.1 untersucht. Weiterhin wird der Unterschied der schnellen und vollständigen ATLAS-Detektorsimulation, sowie der Einfluss der verschiedenen Partonschauer- und Hadronisierungsalgorithmen (s. Abs. 5.1 für Details der verschieden generierten Datensätze) in Bezug auf die invariante Massenverteilung des Modells in Abschnitt 6.3.2 untersucht.

6.1 Entwurf des neuronalen Netzwerkes

Im folgenden Abschnitt wird die Struktur, sowie die Wahl der Hyperparameter des Regressions-Modells vorgestellt. Dieses Netzwerk wird weitergehend das Grundnetzwerk darstellen, an dem alle Hyperparameter (s. Ende von Abs. 4.2.2) systematisch optimiert werden. Die Ergebnisse der Trainings- und Testphase, die zur Wahl dieser Parameter geführt haben, sind in den späteren Abschnitten detailliert erläutert.

Das neuronale Netzwerk ist ein vorwärts gerichtetes Regressionsmodell, bestehend aus einer Eingabeschicht, drei verborgenen Schichten mit jeweils 30 Neuronen und einer Ausgabeschicht mit einem Neuron und ist schematisch in Abbildung 6.1 dargestellt. Das Netzwerk wird mit bestimmten Eingabeobservablen trainiert. Hier werden Paare von invarianten Massen der Elektronen, Myonen und b-Jets - $m_{e\mu}$, m_{eb_1}, m_{eb_2}, $m_{\mu b_1}$, $m_{\mu b_2}$ und $m_{b_1 b_2}$ - sowie die fehlende Transversalenergie $E_{\mathrm{T}}^{\mathrm{miss}}$ verwendet. In den drei tiefen Schichten sind die Neuronen mit jedem Neuron der vorherigen Schicht verbunden. Damit fließen die Berechnungen jedes vorherigen Neurons in die Berechnung des aktuellen Neurons ein. Man spricht dann von *fully connected layers*. Als Aktivierungsfunktionen werden in den drei verborgenen Schichten ReLU (s. Gl. 4.5) und in der Ausgabeschicht eine lineare Funktion verwendet. ADAM (s. 4.15) [101] wird mit einer Ausgangslernrate von $\alpha = 0{,}0003$ als Lern-Algorithmus (s. Abs. 4.2.2) verwendet. Als Verlustfunktion wird der mittlere absolute Fehler aus Gleichung 4.2 verwendet, der über eine Batch-Größe von $N_B = 128$, d.h. über 128 Ereignisse gemittelt (siehe Abs. 4.2.2), berechnet wird. Etwa vier Millionen Ereignisse werden für die Trainingsphase verwendet. Die Wahl aller Hyperparameter ist in Tabelle 6.1 zusammengefasst. Dieses Modell ist das Ergebnis einer systematischen Optimierung dieser Hyperparameter, die in den folgenden Kapiteln erläutert wird.

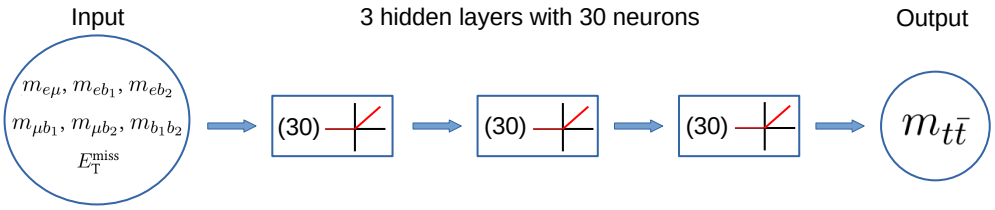

Abbildung 6.1: Schematische Struktur des neuronalen Netzwerks. Das Netzwerk wird mit invarianten Massen von Teilchenpaaren (Elektronen, Myonen und *b*-Jets) und der fehlenden Transversalenergie $E_{\mathrm{T}}^{\mathrm{miss}}$ trainiert. Nach der Eingabeschicht folgen drei vollständig verbundene, verbogene Schichten mit jeweils 30 Neuronen und ReLU Aktivierungsfunktion (s. Gl. 4.5), sowie die Ausgabeschicht, in der die Verteilung $m_{t\bar{t}}$ ausgegeben wird.

Tabelle 6.1: Festgelegte Wahl der Hyperparameter für das Regressions-Netzwerk. Die Auswahl wird in Abschnitt 6.2 begründet.

Hyperparameter	Einstellung
Anzahl verborgener Schichten	3
Neuronen pro Schicht	30, 30, 30
Verlustfunktion	Mittlerer absoluter Fehler (MAE)
Optimierungs-Algorithmus	ADAM mit Lernrate $\alpha = 0{,}0003$
Aktivierungsfunktion	ReLU
Batch-Größe	128
Trainingsbeispiele	≈ 4 Millionen

6.2 Trainingsphase des neuronalen Netzwerkes

Das in Abschnitt 6.1 beschriebene neuronale Netzwerk ist das Ergebnis einer systematischen Optimierung aller in Tabelle 6.1 aufgeführten Hyperparameter. Dieses Netzwerk dient folgend als Grundlage, von dem die Werte der einzelnen Hyperparameter optimiert werden und deren Einfluss auf den Lernprozess des neuronalen Netzwerks untersucht wird.

Wie in Abschnitt 5.1 genannt, stehen etwa acht Millionen simulierte Ereignisse zur Verfügung, welche aus drei Datenerfassungsperioden (2015/16, 2017 und 2018) vereint werden und die leicht veränderten experimentellen Bedingungen beinhalten. Es soll hier genannt sein, dass die einzelnen Perioden keinen Unterschied im Training und der Leistung des Netzwerks verursachen. Der in Abschnitt 5.1 beschriebene $t\bar{t}$ ($\mathrm{PP8}_{\mathrm{full}}$)-Datensatz wird mit der *train-test-split*-Funktion von SCIKIT-LEARN [118] im Verhältnis 50:25:25 in Trainings-:Validierungs:Test-Datensätze aufgeteilt. Die Trainingsdaten dienen als Input für den Optimierungsalgorithmus des neuronalen Netzwerks. Die Validierung erfolgt während des Trainingsprozesses anhand der (für das Modell vorher „ungesehenen“) Validierungsdaten. Die Leistung des trainierten Modells kann dann auf (z.B. verschieden generierten) Testsätzen verglichen werden. Die *train-test-split*-Funktion hat den Vorteil, dass der Datensatz mit der *stratify*-Option vor jedem Split so zufällig gemischt wird, dass die simulierten Daten aus allen drei Datennahmeperioden in jeder Teilmenge berück-

sichtigt werden. Es wird ein *random-seed* angegeben, welcher die Zufälligkeit der Aufteillung steuert, sodass die Ergebnisse reproduzierbar sind. Da der dominante Untergrundprozess der tW-Produktion (s. Abb. 2.5) nur etwa 3% der gesamten Signalstärke ausmacht, wird er für das Training des Modells gänzlich vernachlässigt und nur das Signal $t\bar{t}$ (PP8_{full}) (s. Abs. 5.1) verwendet.

Die Trainingsphase eines neuronalen Netzwerks beinhaltet das Finden der optimalen Hyperparameter und einen Optimierungsprozess mit dem Ziel, den Unterschied zwischen der Netzwerkausgabe und der Zielverteilung zu minimieren. Dabei wird die Verlustfunktion durch den Optimierungsalgorithmus während jeder Epoche minimiert und anhand der Trainingsverlustkurve (engl. *training loss*, also der Wert der Verlustfunktion bei einer bestimmten Epoche) ausgewertet. Der Validierungsverlust (engl. *validation loss*) wird unter Verwendung einer gewählten Metrik (z.B. MAE (Gl. 4.2) oder MSE (Gl. 4.3)) am Ende jeder Trainingsiteration aus den Vailidierungsdaten berechnet (die für die Optimierung unbekannt sind) und wird anschließend mit dem *training loss* verglichen. So kann die Generalisierungsfähigkeit des Modells während des Trainings abgeschätzt werden.

Optimierung der Hyperparameter
Folgend wird die Optimierung der in Abschnitt 6.1 genannten Hyperparameter vorgestellt. Dabei dient das in Tab. 6.1 zusammengefasste Modell als Grundlage, von dem die Werte der Parameter systematisch angepasst werden.

Hyperparameter werden vor dem Trainingsprozess festgelegt und bleiben währenddessen unverändert. Es gibt a priori keinen theoretisch „besten“ Satz an Hyperparametern, stattdessen müssen mehrere Kombinationen getestet und validiert werden. Mit Hilfe der *GridSearch*-Methode aus dem KERAS-Paket [105] kann ein Gitter (Grid) von Hyperparameter-Kombinationen erstellt und anhand der Leistung des Trainingsprozesses und der Zeit bis zur Konvergenz (d.h. der Bereich, wo der Verlustwert nicht mehr merklich verkleinert wird) bewertet werden. Dies kann je nach Anzahl der Kombinationen eine lange Zeit in Anspruch nehmen. Im Folgenden werden die Ergebnisse der GridSearch zusammengefasst.

Nicht jeder Hyperparameter hat einen großen Einfluss auf die Verlustfunktion. Der Unterschied zwischen den Optimierungsalgorithmen RMSProp und ADAM (s. Abs. 4.2.2), sowie zwischen MAE und MSE (s. Gl. 4.2 und Gl. 4.3) als Verlustfunktion ist marginal. Auch die Wahl der Aktivierungsfunktion aus ReLU (s. Gl. 4.5) oder *leaky*-ReLU (s. Gl. 4.7) beeinflussen die Leistung des Modells, sowie die Geschwindigkeit der Konvergenz nicht erheblich. Weitere Verlustfunktionen wie ELU [119] werden getestet, liefern jedoch keine besseren Ergebnisse. Somit werden ADAM, MAE und ReLU für die Variation der verbleibenden Parameter verwendet.

Die Batchgröße N_B, über dessen Anzahl von Ereignissen die Verlustfunktion in jeder Trainingsiteration berechnet wird, sowie die initiale Lernrate α des ADAM-Optimierers (s. Gl. 4.15) beeinflussen hauptsächlich die Konvergenzgeschwindigkeit und werden mit der Anzahl der Trainingsepochen so abgestimmt, dass eine effiziente Konvergenz im Bereich von 50-150 Epochen stattfindet. In der GridSearch werden Werte von $N_B = (32,...,1024)$ und $\alpha = (0{,}0001,...,0{,}01)$ (Der Standardwert des ADAM-Algorithmus ist $\alpha = 0{,}001$) untersucht und auf $N_B = 128$ und

Tabelle 6.2: Übersicht der getesteten Netzwerkstrukturen inklusive der nach Gleichung 4.10 berechneten anpassbaren Modellparameter.

Verborgene Schichten	Neuronen	Anzahl freier Parameter N_Θ
1	8	73
2	10, 10	201
2	100, 100	10901
3	20, 20, 20	1021
3	30, 30, 30	2131

$\alpha = 0{,}0003$ festgelegt. Die Wahl der physikalischen Größen für die Eingabe des neuronalen Netzwerks wird in Abschnitt 6.3.1 behandelt.

Was jedoch die Leistung des Modells stärker beeinflusst, ist der strukturelle Aufbau und die damit verbundene Anzahl an freien bzw. anpassbaren Parametern Θ (siehe Abs. 4.2.2). Mit den bereits bestimmten Werten für die Hyperparameter werden folgend die in Tabelle 6.2 genannten Strukturen untersucht. Die Anzahl der während des Trainings anpassbaren Parameter ermitteln sich aus Gleichung 4.10. Die Verlustfunktionen für 100 Trainingsiterationen (Epochen) der jeweiligen Modellstrukturen sind in Abbildung 6.2 (links) dargestellt. In allen gezeigten Lernkurven ist ein schneller Abfall des Verlustes in den ersten Epochen zu erkennen. Weiterhin ist beobachtbar, dass gerade die weniger tiefen und simpel aufgebauten Netzwerke mit $N_\Theta = 73$ (pinke Kurve) oder $N_\Theta = 201$ (orange Kurve) einen erwartet höheren Verlustwert nach 100 Epochen aufweisen. Ein begrenzter Parameterraum kann die Kapazität des Modells einschränken, komplexe Datenstrukturen zu erkennen, was oft dazu führt, dass nach anfänglichen Verbesserungen keine signifikante Entwicklung im Trainingsprozess stattfindet und die Verlustkurve weitgehend konstant bleibt. In vielen Anwendungen erreichen tiefe Netzwerke (viele Schichten) im Vergleich zu flachen (wenige Schichten), aber „breiten“ (viele Neuronen) Netzwerken eine bessere Leistung [120]. Dafür kann in Abbildung 6.2 (links) die rote Kurve ($N_\Theta = 2131$) mit der blauen Kurve ($N_\Theta = 10901$) verglichen werden. Bei beiden Kurven ist zu erkennen, dass die Parameteranzahl bzw. die Komplexität des Modells groß genug ist, um nach etwa 80 Epochen (blau) bzw. 90 Epochen (rot) ein lokales Minimum der Verlustfunktion zu finden, was an der deutlichen Abnahme des Verlustes erkennbar ist. Diese Abnahme verändert die ausgegebene Massenverteilung $m_{t\bar{t}}$ des Modells stark, worauf in Abschnitt 6.3 eingegangen wird. Aufgrund der deutlich höheren Paramteranzahl ist zu erwarten, dass das Modell der blauen Kurve schneller in dieses Minimum findet. Zwar passiert dies etwa 10 Epochen früher, jedoch ist die Zeit für die Berechnung einer Epoche fast genau doppelt so lang wie die der roten Verlustkurve.

In Abbildung 6.2 (rechts) ist die Lernkurve dieser beiden Modelle inklusive der jeweiligen Validierungskurven (gestrichelt) für weitere 50 Epochen zu sehen. Die beiden Lernkurven stagnieren nach 120 Epochen auf den etwa gleichen Verlustwert. Aus der Validierungskurve (validation loss) ist zu erkennen, dass das Model mit der größeren Anzahl an anpassbaren Parametern (in blau) eine schlechtere Generalisierungsfähigkeit besitzt, da der Validierungsverlust stark schwankt. Das Modell kann so keine zuverlässigen Aussagen auf vorher unbekannte Daten treffen. Dagegen zeigt die rot-gestrichelte Validierungskurve einen deutlich beständigeren Verlauf, doch auch

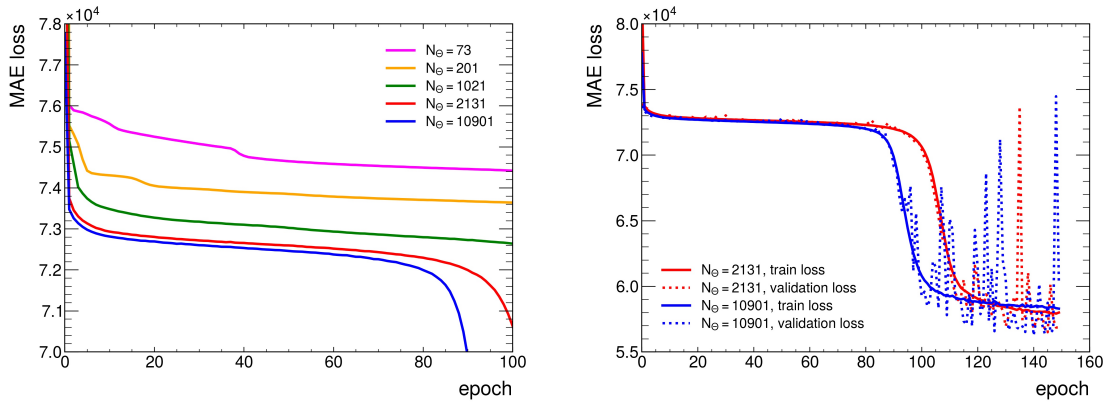

Abbildung 6.2: Vergleich der Lernkurven verschiedener Regressionsmodelle mit unterschiedlicher Parameteranzahl N_Θ. Die Graphen zeigen die Abnahme des mittleren absoluten Fehlers (MAE) über die Anzahl der Trainingsdurchläufe (Epochen) und beschreiben den Einfluss der Modellkomplexität auf das Lernverhalten.

hier gibt es bei Epoche 136 eine große Schwankung der Vorhersage. Schwankungen in den Verlustkurven können ein Anzeichen auf eine sogenannte Überanpassung (engl. *overfitting*) sein. Dabei gibt das Modell die Trainingsdaten gut wieder, scheitert jedoch auf vorher ungesehenen (Validierungs-)Daten. Dies kann durch eine zu große Komplexität des Modells verursacht werden, d.h. eine Wahl aus mehr freien Parametern, als durch die physikalischen Eingabeobservablen gerechtfertigt ist. Diese Fluktuation wird für verschiedene Hyperparameter-Kombinationen beobachtet und kann bei jedem Training vorkommen. Daraus kann geschlossen werden, dass es sich dabei um die stochastische Natur des Optimierungsprozesses handelt bzw. um Artefakte in den Optimierungsalgorithmen, wie der automatischen Anpassung der Lernrate des ADAM-Algorithmus (s. Abs. 4.2.2) in Verbindung mit der Wahl der verwendeten Verlustfunktion. ADAM bezieht alle vorherigen Iterationen mit exponentiell abnehmender Gewichtung in die Anpassung ein. So könnte das Modell dazu tendieren, aus dem aktuellen Minimum „herauszuspringen“ , um ein nahegelegenes, möglicherweise tieferes Minimum zu finden. Da insbesondere die letzten Iterationen mit höherer Gewichtung in die nächste Berechnung einbezogen werden, ist die Fluktuation nicht über mehrere Epochen verteilt, sondern nur bei genau einer. Die Generalisierungsfähigkeit adaptiver Lern-Algorithmen (s. Abs. 4.2.2) wie ADAM ist Bestandteil aktueller Forschung im Gebiet des Deep-Learnings [121, 122].

Trotzdem kann damit die vorher getroffene Aussage bestätigt werden, dass ein Netzwerk mit insgesamt weniger Parametern, aber einer größeren Anzahl von tiefen Schichten eine deutlich effektivere Lösung liefert, als ein Netzwerk mit mehr Neuronen in insgesamt weniger Schichten. Insgesamt werden Modellstrukturen mit bis zu 10 tiefen Schichten und jeweils maximal 100 Neuronen getestet. Aufgrund der insgesamten Konvergenzzeit und der Leistung wird ein Modellaufbau aus 3 tiefen Schichten mit jeweils 30 Neuronen verwendet.

Die beschriebenen Lernkurven sind weiterhin abhängig von der Größe des verwendeten Trainingssatzes. Werden beispielsweise 15000 simulierte $t\bar{t}$-Trainingsereignisse verwendet, sind über 4000 Epochen (die aber einzeln schneller durchlaufen werden) nötig, um auf die gleichen Verlust-

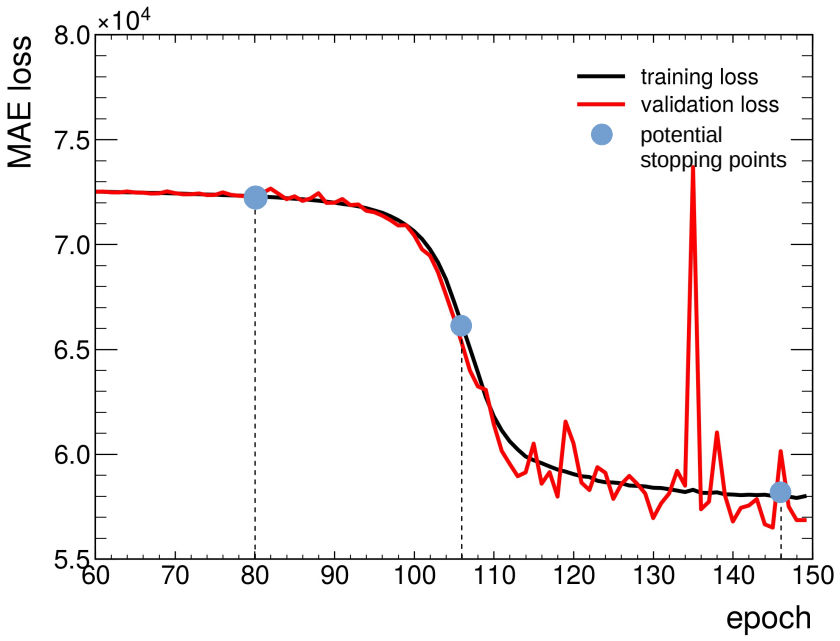

Abbildung 6.3: Kurven für den Trainingsverlust (schwarz) und Validierungsverlust (rot) des neuronalen Netzwerks. Drei Kontrollpunkte sind nach Trainingsiterationen (Epoche) 80, 106 und 146 eingezeichnet, an denen die Modellvorhersage analysiert wird. Für Details siehe Text.

werte wie in Abbildung 6.2 (rechts) zu kommen. Im Kontext dieses Regressionsmodells gilt, dass immer mindestens so viele Trainingsbeispiele gewählt werden müssen, wie das Modell anpassbare Parameter hat. Sonst kann es zu dem oben angesprochenem *overfitting* kommen, wobei das Modell die Trainingsdaten „auswendig lernt“ und eine schlechte Leistung auf neue Daten zeigt. Deshalb werden in der Praxis deutlich mehr Ereignisse verwendet als anpassbare Parameter vorhanden sind. In dieser Arbeit entspricht dies etwa vier Millionen Ereignissen. Viel mehr Trainingsereignisse würden die Konvergenzzeit unnötigerweise erhöhen und ab einem bestimmten Punkt keine Verbesserung der Modellleistung zeigen. Kennzeichen von *overfitting* müssen während des Trainings genau beobachtet werden und können mit verschiedenen Regulierungsmethoden [123] verringert werden.

6.3 Testphase des neuronalen Netzwerks

Im vorherigen Abschnitt wurde die Struktur des verwendeten neuronalen Netzwerks beschrieben. Folgend soll die Leistung des Modells auf dem Testsatz analysiert werden, welcher aus etwa 2 Millionen Ereignissen des $t\bar{t}$ ($\mathrm{PP8_{full}}$) Datensatzes (s. Abs. 5.1) besteht. Zunächst zeigt Abbildung 6.3 wiederholend die Lern- (schwarz) und Validierungskurve (rot) des Modells für 150 Epochen. Zwischen den Epochen 95-110 sieht man eine starke Abnahme beider Kurven. So kann der Bereich bis etwa 90 Epochen als „Nebenminimum“ betrachtet werden, aus dem sich das Modell bei längerem Training in ein „tieferes“ Minimum ab Epoche 120 bewegt. In diesem neuen Minimum beginnt die Validierungskurve zu fluktuieren. Kleine Schwankungen sind zu erwar-

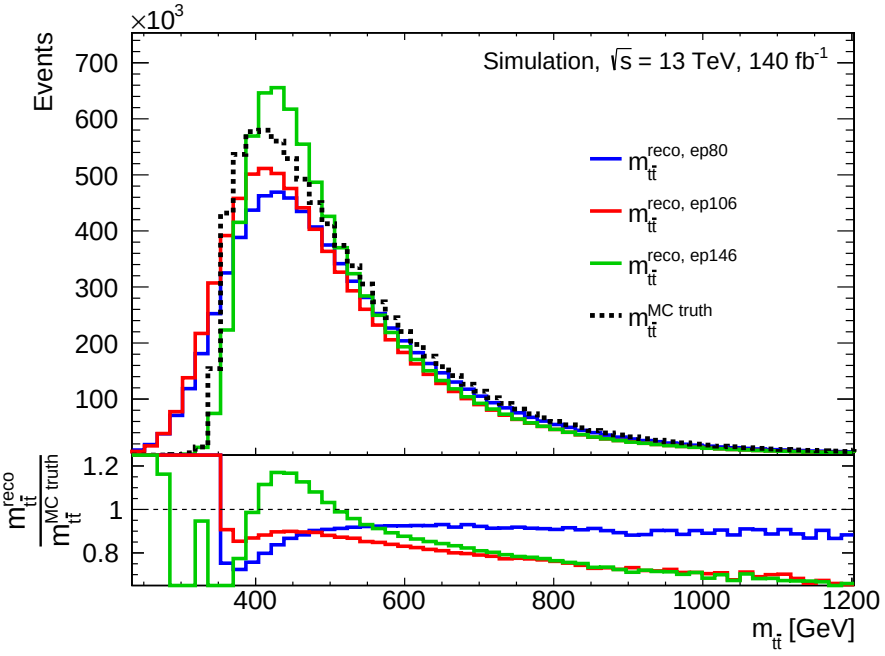

Abbildung 6.4: Die rekonstruierte invariante Masse des Top-Quark-Paares $m_{t\bar{t}}^{\text{reco}}$ des neuronalen Netzwerks für drei verschiedene Zeitpunkte des Trainings: Epoche 80 (blau), 106 (rot), 146 (grün). Dazu ist die Zielverteilung $m_{t\bar{t}}^{\text{MC truth}}$ (schwarz gepunktet) als Vergleich angegeben, inklusive der jeweiligen Verhältnisse im unteren Bereich.

ten, da der Validierungssatz nur halb so viele Ereignisse wie der Trainingssatz aufweist. Derart große Fluktuationen weisen jedoch auf eine unzuverlässige Vorhersage auf neue Datensätze hin, was nicht wünschenswert ist. Besonders in Epoche 136 ist eine große Schwankung der Validierungskurve zu sehen, worauf zum Ende des vorherigen Abschnitts 6.2 detailiert eingegangen wird. Diese Fluktuation ist auf die Stochastik des Optimierungsprozesses zurückzuführen. Der Validierungsverlustwert springt dabei in etwa auf den Wert der 80. Epoche, womit auch die ausgegebene Massenverteilung gleich der in Epoche 80 ist. Die ausgegebene Massenverteilung des Modells ist stark abhängig vom Zeitpunkt des Lernprozesses und dem zugehörigen Verlustwert. Dafür sind in Abbildung 6.3 drei Kontrollpunkte bei Epoche 80 (vor Abfall der Kurve), Epoche 106 (während des Abfalls) und Epoche 146 (im neuen Sättigungswert) eingezeichnet, für die die Netzwerkausgabe folgend verglichen wird.

In Abbildung 6.4 sind die dazugehörigen Verteilungen der invarianten Masse der $t\bar{t}$-Paare für die drei Epochenpunkte [$m_{t\bar{t}}^{\text{reco, ep80}}$ (blau), $m_{t\bar{t}}^{\text{reco, ep106}}$ (rot), $m_{t\bar{t}}^{\text{reco, ep146}}$ (grün)][1)] dargestellt. Als Vergleichswert ist die schwarz gepunktete Massenverteilung $m_{t\bar{t}}^{\text{MC truth}}$, welche die theoretisch berechnete[2)] (zur Genauigkeit des verwendeten Monte-Carlo-Generators POWHEG, s. 5.1)

[1)] Dabei steht „reco“ kurz für „reconstruction“, da es sich um die vom neuronalen Netzwerk rekonstruierte invariante Masse handelt.

[2)] Gemeint ist hier eine Berechnung am Matrix-Element, bevor Algorithmen der Partonschauer und Hadronisierung angewendet werden. Diese Größen werden üblicherweise *Monte-Carlo truth*-Variablen genannt.

Massenverteilung des Top-Quark-Paars darstellt, die außerdem als vorgegebene Zielverteilung während des Modelltrainings definiert ist. Die truth-Verteilung $m_{t\bar{t}}^{\text{MC truth}}$ (schwarz) zeigt an der Produktionsschwelle des $t\bar{t}$-Paars (bei $\approx 2m_t \approx 345$ GeV) einen steilen Anstieg. Der Großteil der Ereignisse konzentriert sich um 400 GeV, woraufhin die Verteilung exponentiell abnimmt. Die Netzwerk-Vorhersagen $m_{t\bar{t}}^{\text{reco, ep80}}$ (blau) und $m_{t\bar{t}}^{\text{reco, ep106}}$ (rot) weisen im Gegensatz dazu eine flachere linke Flanke auf, wodurch die Produktionsschwelle „verschmiert“ wird und auch Ereignisse unterhalb dieser Grenze bis etwa 300 GeV existieren. Zudem liegen die Maxima der beiden Kurven unterhalb der theoretischen Vorhersage. Die Modellvorhersage für den geringsten Verlustwert $m_{t\bar{t}}^{\text{reco, ep146}}$ (grün) modelliert die Produktionsschwelle deutlich präziser und zeigt einen vergleichbaren Anstieg. Hierbei werden jedoch mehr Ereignisse an der Produktionsschwelle verzeichnet, was zu einem höheren Maximum führt. Die Verteilung $m_{t\bar{t}}^{\text{reco, ep80}}$ (blau) beschreibt den exponentiellen Abfall am genausten und ist ab etwa 500 GeV über den restlichen Wertebereich konstant. Die Verteilungen $m_{t\bar{t}}^{\text{reco, ep106}}$ (rot) und $m_{t\bar{t}}^{\text{reco, ep146}}$ (grün) fallen schneller ab als die blaue Kurve und die theoretische Vorhersage. Um zu entscheiden, welche Modellvorhersage verwendet werden soll (bzw. an welcher Epoche das Training beendet wird), wird die Auflösung bzw. der relative Fehler

$$\frac{m_{t\bar{t}}^{\text{reco, ep } i} - m_{t\bar{t}}^{\text{MC truth}}}{m_{t\bar{t}}^{\text{MC truth}}}, \qquad i = 80{,}106{,}146 \tag{6.1}$$

in verschiedenen Intervallen (Bins) von $m_{t\bar{t}}^{\text{MC truth}}$ betrachtet. Eine perfekte Übereinstimmung von Vorhersage und Zielwert wäre somit eine scharfe, symmetrische Verteilung um den Nullpunkt, da dann $m_{t\bar{t}}^{\text{reco, ep } i} = m_{t\bar{t}}^{\text{MC truth}}$ im jeweiligen Bin entspricht. Dies ermöglicht einen genaueren Einblick in die verschiedenen Spektren und die Umverteilung der Ereignisse zwischen den Epochen 80 und 146. Auch kann ein möglicher Bias des Modells, also eine Bevorzugung bestimmter Bereiche der Zielverteilung $m_{t\bar{t}}^{\text{MC truth}}$, untersucht werden.

Dies ist für die drei Modellvorhersagen in den Intervallen [300-400 GeV, 400-500 GeV, 500-600 GeV, 600-700 GeV, 700-800 GeV und >800 GeV] von $m_{t\bar{t}}^{\text{MC truth}}$ in Abbildung 6.5 gezeigt. Die größten Unterschiede sind in den ersten zwei Bins, also im Bereich der Produktionsschwelle, zu beobachten. Die Verteilungen $m_{t\bar{t}}^{\text{reco, ep80}}$ (blau) und $m_{t\bar{t}}^{\text{reco, ep106}}$ (rot) weisen eine ähnlich hohe Symmetrie um den Nullpunkt auf, während die Verteilung $m_{t\bar{t}}^{\text{reco, ep146}}$ (grün) im Bereich von 300-400 GeV eine positive Verschiebung in Richtung des Kurvenmaximums von $m_{t\bar{t}}^{\text{MC truth}}$ um 400 GeV aufweist. Die leichte Asymmetrie ist für alle Verteilungen vorhanden und ist damit eine feste Eigenschaft des gewählten Modells. Insbesondere im Bereich ab $m_{t\bar{t}}^{\text{MC truth}} > 500$ GeV ist ein leichter Bias in Richtung geringerer $m_{t\bar{t}}^{\text{MC truth}}$-Werte zu beobachten. Dies ist für die grüne Kurve am stärksten ausgeprägt, gefolgt von der roten Kurve. Die blaue Kurve weist allgemein einen niedrigen Bias auf. Das Modell scheint Ergebnisse im Lernprozess zu präferieren, die im Bereich des Maximums der Zielverteilung $m_{t\bar{t}}^{\text{MC truth}}$ liegen. Das ist eine ungewollte Eigenschaft und Zeichen eines klar vorhandenen Biases, wobei der allgemeine Trend ist, dass der Bias für eine zunehmende Epochenzahl anwächst. Für alle gezeigten Bereiche von $m_{t\bar{t}}^{\text{MC truth}}$ (und insbesondere für kleine Werte) besitzt die Verteilung $m_{t\bar{t}}^{\text{reco, ep80}}$ (blau) den niedrigsten Bias. Deshalb wird für alle weiteren Analysen die Ausgabe des neuronalen Netzwerks nach 80 Trainingsepochen verwendet.

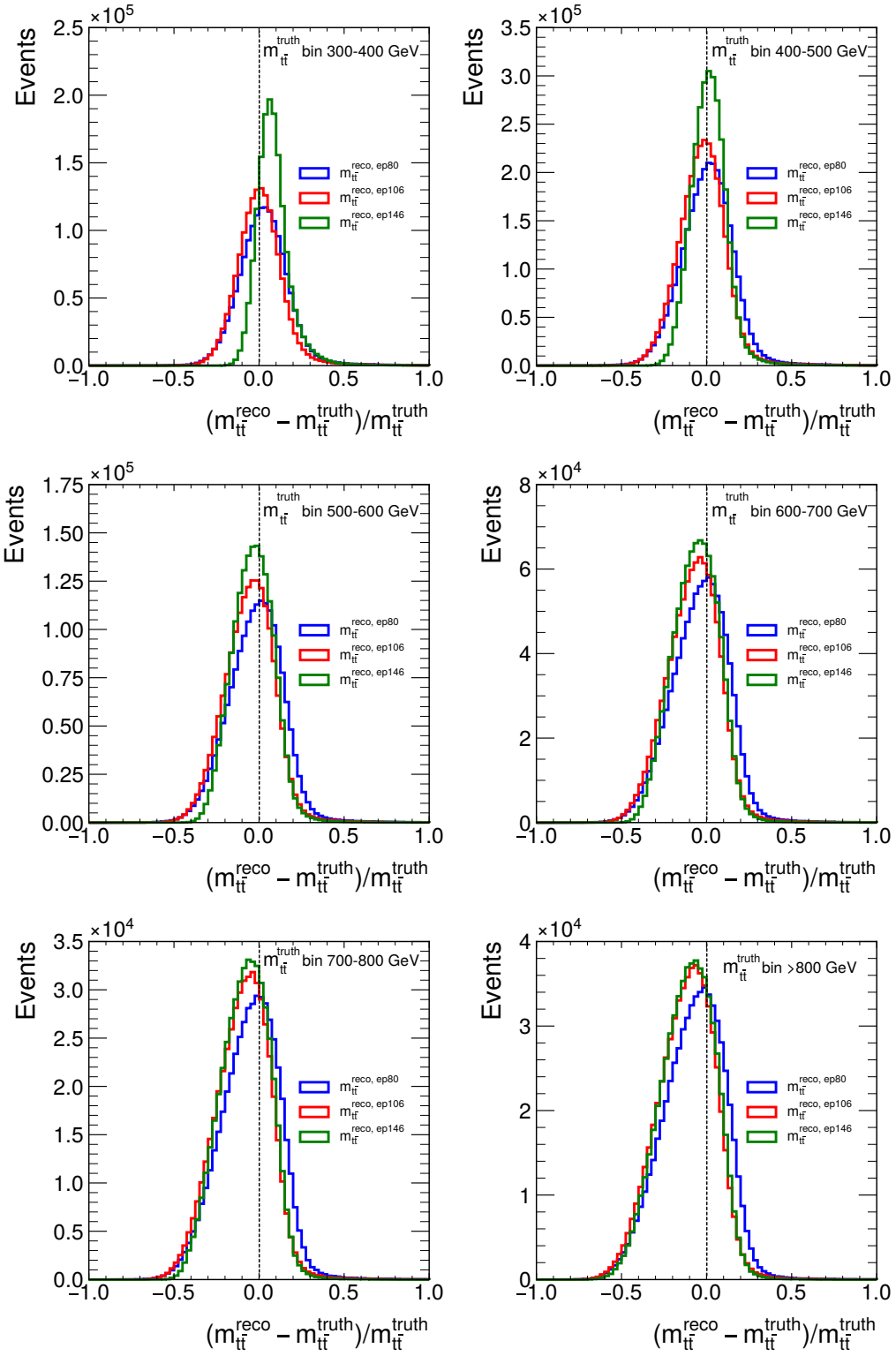

Abbildung 6.5: Vergleich der nach Gleichung 6.1 berechneten relativen Fehler der Netzwerkausgaben für drei verschiedene Trainingszeitpunkte (Epoche 80 (blau), 106 (rot), 146 (grün)) in verschiedenen Energieintervallen (Bins) der Zielverteilung $m_{t\bar{t}}^{\text{MC truth}}$.

Der Trainingsprozess wird über eine sogenannte *callback*-Funktion von KERAS [105] im Bereich von 80 Epochen gestoppt, indem der Verlustwert am Ende jeder Trainingsepoche überwacht und das Training beendet wird, sobald dieser einen bestimmten Wert unterschreitet. Um im Bereich von Epoche 80 zu stoppen, wird in diesem Fall ein MAE-Verlust von 72500 festgelegt (vgl. Abb. 6.3).

In den gezeigten Untersuchungen liefert ein geringerer Verlustwert, was eigentlich einer positiven Optimierung der Verlustfunktion entspricht, eine ungewollte Eigenschaft in Form einer Bevorzugung für Ereignisse im Bereich der Produktionsschwelle der $t\bar{t}$-Produktion. Damit kommt der Wahl der Verlustfunktion eine besondere Bedeutung zu. Die verwendeten Funktionen MAE (s. Gl. 4.2) und MSE (s. Gl. 4.3) weisen äquivalente Ergebnisse auf. Auch wurde eine Verlustfunktion wie Gleichung 6.1 getestet, um die Auflösung direkt zu optimieren, wobei jedoch keine Verbesserung beobachtet wurde. Eine gezielte Suche nach potentiellen Verlustfunktionen ohne einen beobachtbaren Bias ist eine wichtige Aufgabe von Untersuchungen, die auf diese Arbeit folgen könnten.

6.3.1 Vergleich von Eingabeobservablen

Im Folgenden wird untersucht, wie die Leistung des Modells von den gewählten physikalischen Eingabegrößen (folgend auch Feature oder Merkmal genannt) abhängt. Intuitiv kommen zunächst die unmittelbar messbaren kinematischen Variablen der Zerfallsprodukte (1 Elektron, 1 Myon, 2 b-Jets) wie der Transversalimpuls p_T, die Pseudorapidität η, der Azimutwinkel ϕ, die Energie E oder die gesamte fehlende Transversalenergie $E_{\mathrm{T}}^{\mathrm{miss}}$ in Frage. Dies sind sogenannte *low-level features*, da sie direkt aus der Simulation und Rekonstruktion der Ereignisse stammen und keiner weiteren Vorverarbeitung unterzogen sind. Aus ihnen können *high-level features* konstruiert werden. Einerseits werden die invarianten Massen von Teilchen im Endzustand (1 Elektron, 1 Myon und 2 b-Jets) berechnet. Genauer wird dies für die Kombinationen aus zwei dieser Teilchen aus der Summe der Impuls-Vierervektoren[3)] bestimmt, was insgesamt sechs Paare von Massen ergibt: $m_{e\mu}$, m_{eb_1}, m_{eb_2}, $m_{\mu b_1}$, $m_{\mu b_2}$ und $m_{b_1 b_2}$. Weiterhin wird der Abstand in der Winkelebene $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ der selben Teilchenpaare berechnet.

In der folgenden Untersuchung wird das neuronale Netzwerk mit verschiedenen Eingabegrößen trainiert und die Leistung des Modells verglichen. Dafür wird ein Algorithmus verwendet, welcher die Wichtigkeit einer Eingabegröße misst, indem die Auswirkung einer zufälligen Permutation der Werte eines Features auf die Vorhersageleistung gemessen wird. In anderen Worten: Wie wirkt sich das Herausnehmen einer Größe aus dem Datensatz auf die Modellleistung aus? Dazu wird der Unterschied des erreichten Verlustwertes bei der Vorhersage des unveränderten Datensatzes (genannt S_{baseline}) mit dem Verlustwert, bei dem die Werte eines Features zufällig verändert werden ($S_{\mathrm{permutation}}$), verglichen. Berechnet wird also folgende relative Empfindlichkeit für jede Eingabegröße x:

$$S_x = \frac{S_{\mathrm{baseline}} - S_{\mathrm{permutation}}}{S_{\mathrm{baseline}}}. \tag{6.2}$$

[3)] An Hadron-Kollidierern ist es üblich, den Viererimpuls durch die messbaren Größen (p_T, η, ϕ, E) auszudrücken.

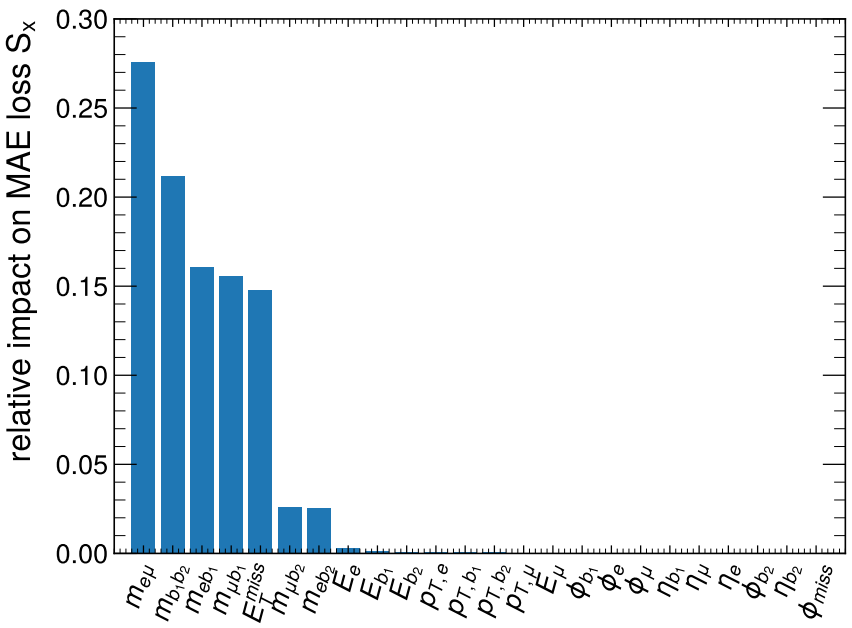

Abbildung 6.6: Die Eingabegrößen sind gemäß ihrer Bedeutung für den Trainingsprozess geordnet. Genauer wird nach Gleichung 6.2 der Einfluss auf den Verlustwert berechnet, wenn für die Größe zufällige Wert angenommen werden.

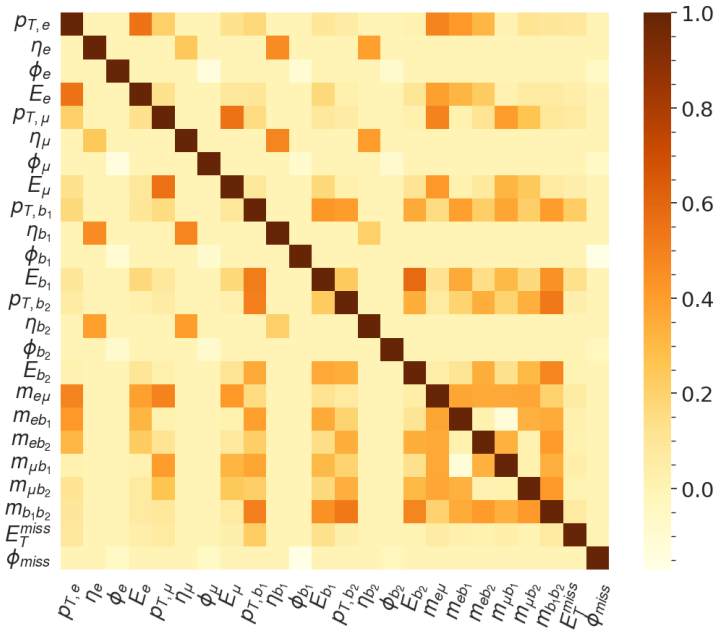

Abbildung 6.7: Die lineare Korrelationsmatrix der möglichen kinematischen Eingabegrößen der Elektronen, Myonen und *b*-Jets für das neuronale Netzwerk.

Die Ergebnisse sind in Abbildung 6.6 dargestellt. Auf der y-Achse ist der Wert S_x für verschiedene Eingabeparameter gezeigt. Die höchsten Punktzahlen werden von den invarianten Massenpaaren sowie der fehlenden Transversalenergie erzielt, was bedeutet, dass diese für das verwendete Modell von größerer Bedeutung für die Optimierungsaufgabe sind. Dieser Schluss kann nur in Kombination mit dem Wissen über die Korrelation der Eingabegrößen gezogen werden. Sind zwei Eingaben stark miteinander korreliert, kann die Permutation eines dieser Features keinen großen Einfluss auf die Modellleistung haben, da die relevante Information über das jeweils andere korrelierte Feature erhalten bleibt. Die lineare Korrelationsmatrix aller Eingabevariablen ist in Abbildung 6.7 gezeigt. Die Transversalimpulse und Energien der jeweiligen Leptonen und b-Jets (z.B. E_e und $p_{T,e}$) sind am stärksten korreliert, mit Werten $\approx 0{,}5$. Die invarianten Massen sind untereinander etwa so stark korreliert, wie mit den Energien und Transversalimpulsen der Leptonen und b-Jets. Dazu wird beispielsweise getestet, $m_{e\mu}$ mit $p_{T,e}$, $p_{T,\mu}$, E_e oder E_μ im Trainingssatz zu ersetzen. Jedoch liefert dies ein schlechteres Ergebnis, was aus den deutlich geringeren Werten in Abbildung 6.6 und der relativ geringen Korrelation zu erwarten war und bestätigt, dass die invarianten Massen bessere Eingabegrößen für das gewählte neuronale Netzwerk sind, als die kinematischen Größen der einzelnen Teilchen. Es ist zu beachten, dass nur die lineare Korrelation dargestellt ist. Insbesondere tiefe neuronale Netzwerke besitzen die Fähigkeit, nicht-lineare Zusammenhänge zwischen den Eingangsgrößen und der Zielverteilung herzustellen.

Die geringsten Werte erzielen die Winkelinformationen η und ϕ aller Teilchen. Die daraus berechneten Werte für ΔR sind ebenfalls schlecht bewertet und deshalb der Übersicht halber rausgelassen. Eine mögliche Erklärung für den geringen Einfluss der Winkelvariablen ist, dass Werte der Winkel in einer anderen Größenordnung liegen (z.B. η $\mathcal{O}(1)$), als beispielsweise ein Transversalimpuls (p_T $\mathcal{O}(10^9$ eV)). Dies kann im Modell zu verschiedenen Gewichtungen in der Anpassung führen, wobei Werte mit größerem Betrag bevorzugt werden. Daher ist es gängige Praxis, die Eingaben vor dem Training zu normalisieren. Dafür können verschiedene Funktionen von Scikit-learn [118] genutzt werden. Der MinMaxScaler skaliert die jeweiligen Eingaben gleichmäßig auf einen Bereich zwischen 0 und 1, und der StandardScaler auf einen Mittelwert von 0 mit Standardabweichung 1. Beide Funktionen werden für das Training vor dem oben beschriebenen Permutationstest verwendet, was zu keiner Änderung der Ergebnisse führt. Auch für einen größeren Datensatz werden die invarianten Massen (und $E_{\mathrm{T}}^{\mathrm{miss}}$) vom neuronalen Netzwerk bevorzugt.

Die Normalisierung der Eingaben auf eine gemeinsame Skala beschleunigt den allgemeinen Trainingsprozess, sodass nach nur etwa 30 Epochen der gleiche Verlustwert wie in Abbildung 6.3 nach 130 Epochen erreicht wird. Jedoch konvergiert das Modell direkt in dieses Minimum und „überspringt“ das erste Minimum bis Epoche 90. Da die entsprechende Ausgabe im neuen Minimum wie im vorherigen Abschnitt und Abbildung 6.5 beschrieben nicht wünschenswert ist, wird auf eine Normalisierung der Daten verzichtet. Dies ist dann gerechtfertigt, wenn nur die invarianten Massen der Teilchenpaare und die fehlende Transversalenergie als Eingabeparameter verwendet werden. Da sich diese in der selben Skala befinden und die verwendete Aktivierungsfunktion der Neuronen ReLU (s. Gl. 4.5) nicht nach oben beschränkt ist, ist keine Normalisierung notwendig.

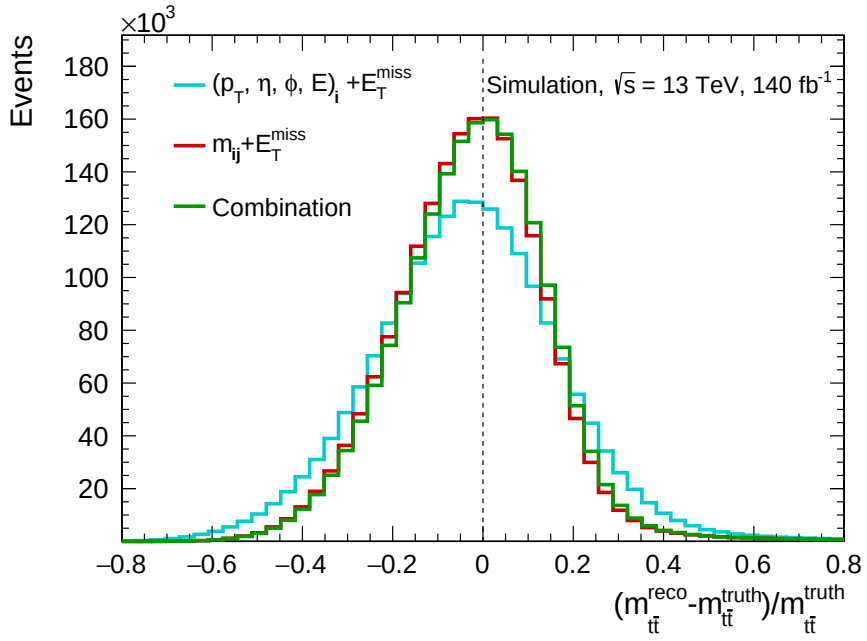

Abbildung 6.8: Vergleich des relativen Fehlers nach Gleichung 6.1 des neuronalen Netzwerks für verschiedene Eingabegrößen: Die invariante Masse von zwei unterschiedlichen Leptonen oder *b*-Jets im Endzustand m_{ij} und fehlender Transversalenergie (rot), die kinematischen Größen der einzelnen Leptonen oder *b*-Jets mit der fehlenden Transversalenergie (hellblau) und deren Kombination (grün).

Abbildung 6.8 zeigt abschließend den relativen Fehler, also die durchschnittliche Abweichung des vorhergesagten Wertes im Gegensatz zum vorgegebenen Wert, für verschiedene Kombinationen der Eingabefeatures. Die rote Kurve zeigt die Leistung des Modells unter Verwendung der invarianten Massen m_{ij} (wobei i,j für Leptonen und b-Jets stehen) und der fehlenden Transversalenergie als Eingabegrößen und weist eine recht symmetrische Verteilung auf. Die hellblaue Kurve dagegen zeigt die Leistung bei einem Training mit den kinematischen Größen der einzelnen Leptonen und b-Jets. Es ist eine breite und asymmetrischere Verteilung zu erkennen, was auf eine schlechtere Übereinstimmung hinweist. Die Kombination der beiden (grüne Kurve) gibt gegenüber der roten Kurve keine merkliche Verbesserung. Die höhere Anzahl der Eingabevariablen erhöht jedoch die Zeit für einen Trainingsdurchlauf, da mehr anpassbare Parameter in das Modell hinzugefügt werden. Deshalb wird auf die *low-level*-Features gänzlich verzichtet und das neuronale Netzwerk nur mit den invarianten Massen und der fehlenden Transversalenergie trainiert.

In vielen Analysen am ATLAS-Experiment werden bessere Ergebnisse unter der Verwendung von *low-level* Variablen erzielt, die jedoch häufig eine andere Deep-Learning Struktur verwenden [124, 125]. Auch der häufig verwendete b-Tagger (s. Abs. 3.3.5) DL1r [82] beruht auf der Eingabe von *low-level* Features. Letztendlich sind die verwendeten Deep-Learning Techniken abhängig von den Besonderheiten der jeweiligen Analyse.

Tabelle 6.3: Übersicht der erwarteten $t\bar{t}$ Eregnisse für verschiedene Monte-Carlo Simulationen für die Jahre 2015-2018. Diese setzen sich aus POWHEG+PYTHIA8 und POWHEG+HERWIG7 zusammen. Der Index „full“ steht für die volle Detektorsimulation mit GEANT4 und „fast“ für die schnelle Simulation mit ATLASFASTII. Für alle simulierten Ereignisse ist die statistische Unsicherheit gegeben.

Datensatz	2015/16 (36,5 fb^{-1})	2017 (44,5 fb^{-1})	2018 (59 fb^{-1})	Run-2 (140 fb^{-1})
$t\bar{t}$ (PH7_{fast})	80 250±60	95 690±65	126 350±75	302 300±115
$t\bar{t}$ (PP8_{fast})	81 870±60	97 700±65	129 140±75	308 710±115
$t\bar{t}$ (PP8_{full})	81 350±55	96 550±60	127 040±70	304 950±110

6.3.2 Vergleich unterschiedlich generierter Datensätze

Die Details der verwendeten Simulationsdaten sind in Abschnitt 5.1 erklärt. Im folgenden Abschnitt werden die unterschiedlich generierten Datensätze anhand von kinematischen Variablen verglichen und ihr Einfluss auf die Verteilung $m_{t\bar{t}}^{\text{reco}}$ des neuronalen Netzwerks untersucht.

In Tabelle 6.3 ist die erwartete Anzahl an Ereignissen der drei generierten $t\bar{t}$ Datensätze, gewichtet nach Gleichung 5.2 für die Jahre 2015-2018 samt der statistischen Unsicherheit dargestellt. Die verschiedene Modellierung der Partonschauer und Hadronisierung von PYTHIA8 und HERWIG7 führt zu einem abweichendem Teilcheninhalt und leicht unterschiedlicher Kinematik. Dies beeinflusst die Effizienz, mit der verschiedene Objekte rekonstruiert oder identifiziert werden und führt zu einer Abweichung in der erwarteten Anzahl von Ereignissen eines bestimmten Endzustands. Der Unterschied in der Ereigniszahl zwischen schneller und vollständiger Simulation kann an verschiedenen Effizienzeffekten liegen, die durch die unterschiedliche Simulation der Wechselwirkungen in den Kalorimetern zustande kommt. Davon kann beispielsweise die Energieauflösung betroffen sein (für Details der Unterschiede in schneller und vollständiger Simulation siehe [112]).

Vergleich von schneller und vollständiger ATLAS-Detektorsimulation
Folgend werden die Datensätze $t\bar{t}$ (PP8_{full}) und $t\bar{t}$ (PP8_{fast}) (s. Abs. 5.1), also die vollständige ATLAS-Simulation durch GEANT4 [110] mit der schnellen ATLAS-Simulation durch ATLASFASTII [112] verglichen. Das neuronale Netzwerk wird mit der Eingabe der fehlenden Transversalenergie $E_{\text{T}}^{\text{miss}}$ und den invarianten Massen, welche aus den kinematischen Variablen der einzelnen Leptonen und b-Jets im dileptonischen Endzustand berechnet werden, trainiert (s. 6.3.1). In Abbildung 6.9 ist der Transversalimpuls der Leptonen und b-Jets sowie die fehlende Transversalenergie $E_{\text{T}}^{\text{miss}}$ und die invariante Masse des Elektron-Myon-Paares $m_{e\mu}$ für die vollständige ATLAS-Simulation mit GEANT4 (blau) und die schnelle ATLAS-Simulation mit ATLASFASTII (rot) gezeigt. Zudem sind die Verhältnisse von vollständiger zu schneller Simulation gegeben, wobei „PP8“ für die Ereignisgeneration von POWHEG+PYTHIA8 steht, die in beiden Fällen verwendet wird. Im Allgemeinen zeigt sich eine hohe Übereinstimmung. Für den Transversalimpuls der Leptonen und b-Jets tritt eine Abweichung bis zu 4% für höhere Energiebereiche auf.

Vergleich der Modellierung von Partonschauer und Hadronisierung
Analog wird nun der Unterschied zwischen dem POWHEG+PYTHIA8 (PP8_{fast}) [57] generiertem $t\bar{t}$-Datensatz und dem POWHEG+HERWIG7 (PH7_{fast}) [58, 59] generiertem $t\bar{t}$-Datensatz (s. Abs. 5.1)

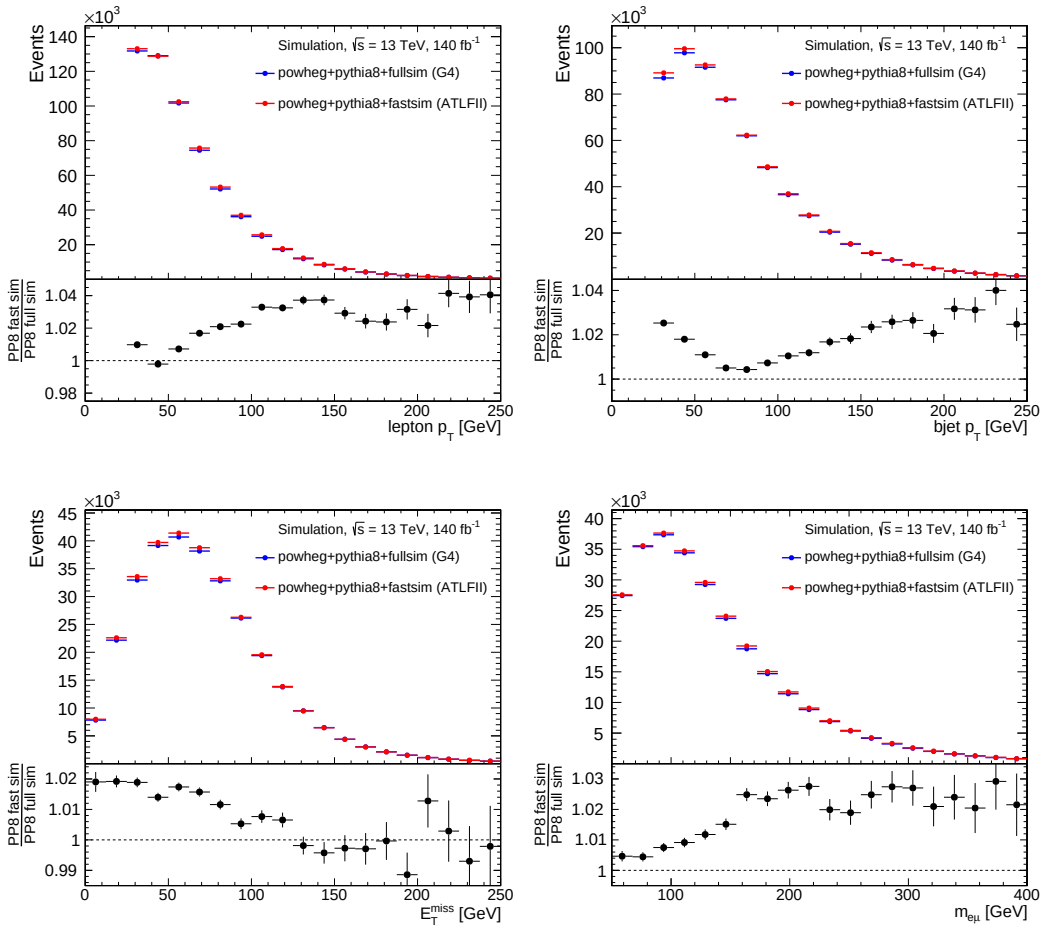

Abbildung 6.9: Vergleich von vollständiger (GEANT4, blau) und schneller (ATLASFASTII, rot) ATLAS-Simulation für den Transversalimpuls der Leptonen (oben links), den Transversalimpuls der *b*-Jets (oben rechts), die fehlende Transversalenergie (unten links) und die invariante Masse des Elektron-Myon-Paars (unten rechts). Im unteren Bereich ist dazu jeweils das Verhältnis mit statistischem Fehler gegeben.

analysiert. Der HERWIG7-Simulationssatz wird nur unter Verwendung der schnellen ATLAS-Simulation durch ATLASFASTII [112] erstellt, wodurch er folgend mit dem PYTHIA8 Simulationssatz verglichen wird, der durch die gleiche schnelle Detektorsimulation produziert wurde. So können die Detektorsimulationsunterschiede von den Generatorunterschieden entkoppelt werden. Top-Quark-Prozesse werden modelliert, indem die durch POWHEG [50–52] durchgeführte NLO Matrixelement-Berechnung des harten Streuprozesses mit Partonschauer- und Hadronosierungsgeneratoren kombiniert wird (s. Abs. 2.2.2). Die systematische Unsicherheit in der Modellierung kann durch Anpassungen[4)] interner Generatorparameter oder durch den Vergleich zweier unterschiedlicher Generatoren für Partonschauer und Hadronisierung - hier PYTHIA8 [57] und HERWIG7 [58, 59] (siehe Abs. 2.2.2 für die Unterschiede der Modellierungen) -

[4)] Hier sind beispielsweise Verbesserungen in der Theorie, wie Berechnungen höherer Ordnungen der Störungstheorie oder ein verbessertes *tuning* an experimentellen Daten gemeint.

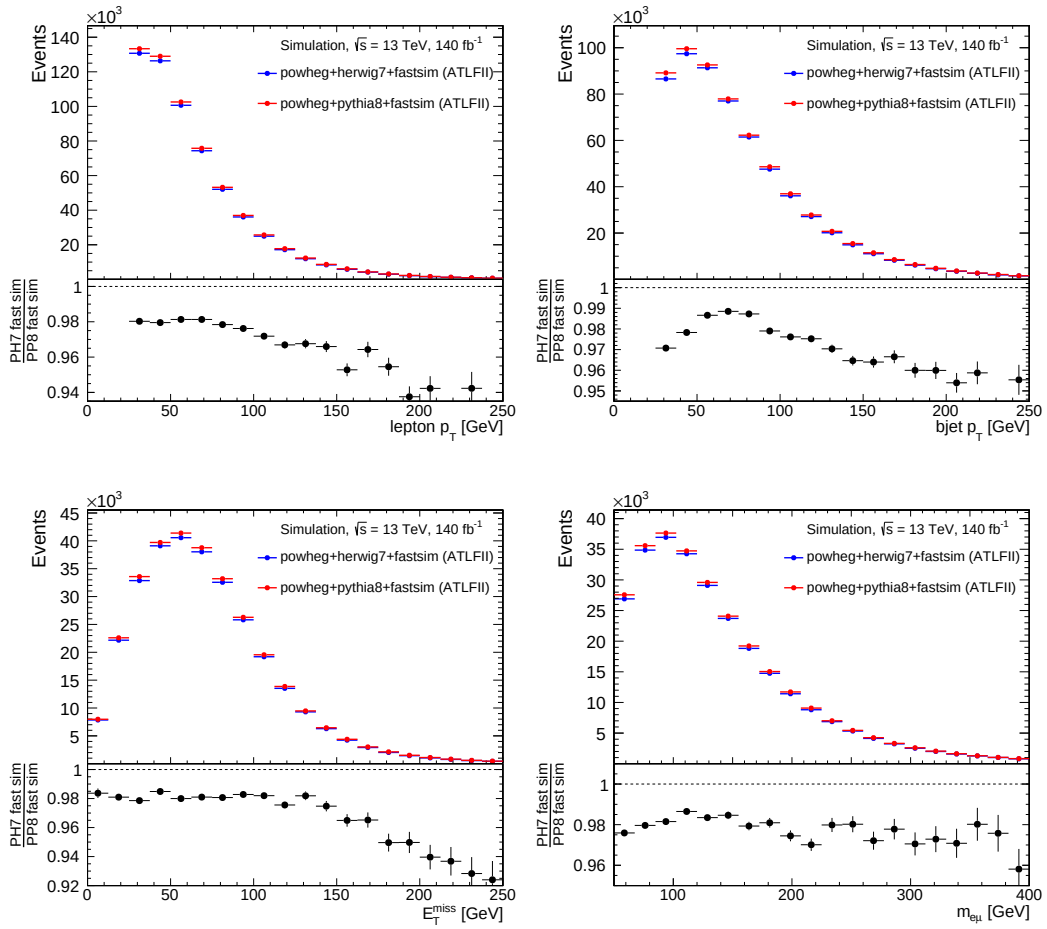

Abbildung 6.10: Vergleich unterschiedlich generierter $t\bar{t}$-Datensätze aus POWHEG+PYTHIA8 (rot) und POWHEG+HERWIG7 (blau) für den Transversalimpuls der Leptonen (oben links), den Transversalimpuls der b-Jets (oben rechts), die fehlende Transversalenergie (unten links) und die invariante Masse des Elektron-Myon-Paares (unten rechts). Im unteren Bereich ist dazu jeweils das Verhältnis inklusive statistischem Fehler gegeben.

abgeschätzt werden. Für die in Kapitel 7 beschriebene vollständige Bestimmung der Top-Yukawa Kopplung ist eine umfassende Berücksichtigung aller systematischen Unsicherheiten vorausgesetzt, wobei insbesondere die Unsicherheit der Modellierung von $t\bar{t}$-Ereignissen einen deutlichen Effekt haben kann [19, 126], weswegen sich hier auf diese konzentriert wird.

Der Unterschied der genannten Datensätze ist in Abbildung 6.10 für den Transversalimpuls der Leptonen und b-Jets, die fehlende Transversalenergie $E_{\mathrm{T}}^{\mathrm{miss}}$ und die invariante Masse $m_{e\mu}$ gezeigt. Die geringste Abweichung ist in der Verteilung von $m_{e\mu}$ mit etwa 2-4% gegeben. In den Spektren der Transversalimpulse und $E_{\mathrm{T}}^{\mathrm{miss}}$ ist ebenfalls eine geringe Abweichung von etwa 2-4% für niedrige Energiewerte erkennbar, jedoch steigt diese Abweichung mit Zunahme der Energie auf bis zu 8% für $E_{\mathrm{T}}^{\mathrm{miss}}$ an. Die zum Teil deutlichen Unterschiede weisen darauf hin, dass die Änderung des Partonschauer- und Hadronisierungsmodells einen Einfluss auf die Ausgabe des

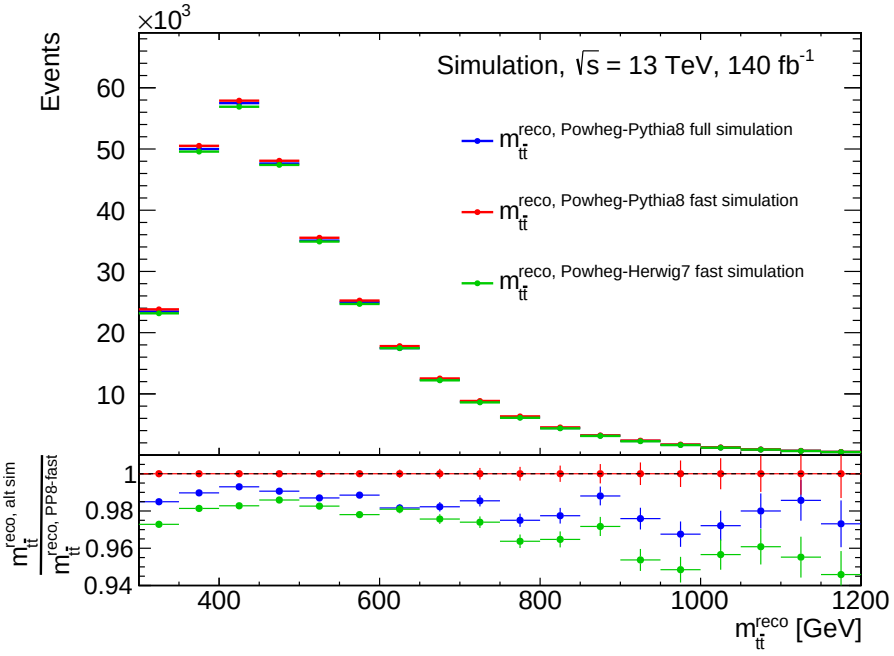

Abbildung 6.11: Vergleich der invarianten Masse $m_{t\bar{t}}^{\text{reco}}$ der Modellvorhersage für verschieden generierte $t\bar{t}$ Datensätze: POWHEG+PYTHIA8+GEANT4-Simulation (PP8-full) (blau), POWHEG+PYTHIA8+ATLASFASTII-Simulation (PH7-fast) (rot) und POWHEG+HERWIG7+ATLASFASTII-Simulation (PP8-fast) (grün). Im unteren Bereich ist jeweilige das Verhältnis zu der (PP8-fast)-Simulation gezeigt. Dazu ist der statistische Fehler gegeben.

neuronalen Netzwerks haben könnte und eine große systematische Unsicherheit in der Schätzung der Top-Yukawa Kopplung darstellt. Auf Letzteres wird in Abs. 7.2 näher eingegangen.

Einfluss auf die Massenverteilung $m_{t\bar{t}}^{\text{reco}}$ des neuronalen Netzwerks
Abbildung 6.11 betrachtet schließlich den Einfluss der unterschiedlich generierten Datensätze in Bezug auf die vom neuronalen Netzwerk ausgegebene rekonstruierte Massenverteilung $m_{t\bar{t}}^{\text{reco}}$. Dazu werden diese Datensätze nach Abschluss der Trainingsphase als Testsätze für die Modellvorhersage verwendet. Die blaue Linie repräsentiert die POWHEG+PYTHIA8+GEANT4 Simulation. Die rote Linie stellt die POWHEG+PYTHIA8+ATLASFASTII Simulation und die grüne Linie die Simulation aus POWHEG+HERWIG7+ATLASFASTII dar. Die Verhältnisse $t\bar{t}$ (PP8_{full}) / $t\bar{t}$ (PP8_{fast}) (blau) und $t\bar{t}$ (PH7_{fast}) / $t\bar{t}$ (PP8_{fast}) (grün) (in der Achsenbeschriftung zusammengefasst unter der Bezeichnung „alt sim“) sind ebenfalls dargestellt, um die Abweichungen besser aufzulösen. Allgemein weisen die drei Massenverteilungen eine übereinstimmende Form auf. Der Vergleich zwischen den verschiedenen Generatoren (PP8_{fast}) (rot) und (PH7_{fast}) (grün) zeigt eine geringe Abweichung von 2-3 % bis etwa 700 GeV und anschließend Abweichungen von bis zu rund 5%. Die blaue Verhältniskurve zeigt einen Unterschied von rund 1-2% in kleinen Werten von $m_{t\bar{t}}^{\text{reco}}$, die folgend auf etwa 3% ansteigen, womit eine hohe Übereinstimmung zwischen der vollständigen und schnellen ATLAS-Simulation gezeigt ist. Die erhöhte Unsicherheit ist auf die limitierte Statistik in den höheren Energiebereichen zurückzuführen.

7 Top-Yukawa Kopplung

Seit der Entdeckung des Higgs-Bosons in 2012 [2,3] ist die Untersuchung dessen Eigenschaften ein Hauptziel des LHC-Programms. Im Standardmodell (s. Kap. 2) wird die Masse eines Fermions über die Yukawa-Kopplung g_f (Gl. 2.9), also die Wechselwirkungsstärke mit dem Higgs-Feld, bestimmt. Das Top-Quark als schwerstes Fermion mit einer Masse $m_t = 172{,}5$ GeV [6] führt auf eine Vorhersage der Yukawa-Kopplung im SM von etwa $g_t^{\mathrm{SM}} \approx 1$[1)]. Erweiterungen des Standardmodells wie das 2-Higgs-Dublett [10] oder ein zusammengesetztes Higgs-Boson [11] modifizieren die Kopplung des Top-Quarks mit dem Higgs-Feld, was eine experimentelle Bestätigung der Vorhersage von $g_t^{\mathrm{SM}} \approx 1$ motiviert. Eine direkte Messung von g_t über die $t\bar{t}H$-Produktion [15,16] beinhaltet Annahmen über die Higgs-Kopplung zu den Zerfallsprodukten (z.B. b-Quarks). Diese Abhängigkeit kann in indirekten Messungen von g_t vermieden werden, wie es beispielsweise in der Vier-Top-Produktion $t\bar{t}t\bar{t}$-Produktion [17,18] gezeigt werden kann.

Die in dieser Arbeit motivierte Analyse der Top-Yukawa Kopplung ist ebenfalls eine indirekte Messung, die im dileptonischen Zerfallskanal von Top-Antitop-Paaren durchgeführt wird. Sie nutzt eine Untersuchung der vom neuronalen Netzwerk rekonstruierten invarianten Masse des Top-Antitop-Paares $m_{t\bar{t}}^{\mathrm{reco}}$, die im Bereich der Produktionsschwelle ($\approx 2m_t$) sensitiv auf eine virtuelle Korrektur durch den Austausch eines Higgs-Bosons ist und damit abhängig von der Stärke der Top-Yukawa Kopplung ist, die letztlich mit Hilfe eines Likelihood-Fits abgeschätzt werden kann. Folgend wird mit der Top-Yukawa Kopplung nicht g_t (s. Gl. 2.9), sondern die Top-Yukawa Kopplung im Verhältnis zum Standardmodellwert $Y_t = g_t/g_t^{\mathrm{SM}}$ gemeint.

In Abschnitt 7.1 wird zunächst ein Vergleich von kinematischen Variablen in simulierten Daten und experimentellen Daten vorgenommen. Außerdem wird die rekonstruierte Massenverteilung des neuronalen Netzwerks auf simulierten und experimentellen Daten ausgewertet und verglichen. Eine gute Übereinstimmung ist wichtig, da die letztliche Bestimmung der Top-Yukawa Kopplung aus den experimentellen Daten erfolgt. Abschnitt 7.2 analysiert den Einfluss der Top-Yukawa Kopplung auf die $m_{t\bar{t}}$-Verteilung auf dem Teilchen-Level und nach Rekonstruktion durch das neuronale Netzwerk und zeigt weiterhin einen Vergleich zu der einfach-rekonstruierbaren Verteilung m_{bbll}^{reco} der Leptonen und b-Jets. In Abschnitt 7.2.1 wird die statistische Methode erläutert, die zur Schätzung der Top-Yukawa Kopplung verwendet wird. Dazu wird in Abschnitt 7.2.2 eine erste Schätzung der Top-Yukawa Kopplung in einem Likelihood-Fit auf Asimov-Daten gegeben und näher auf den Einfluss der Unsicherheiten in Partonschauer und Hadronisierung eingegangen.

7.1 Data-MC Vergleich

Der Vergleich von experimentellen und simulierten Daten liefert einen Einblick in die Genauigkeit der theoretischen Modelle, auf denen die Simulationen beruhen. Dafür werden zunächst kinematische Variablen der Simulation bestehend aus dem Signal $t\bar{t}$ ($\mathrm{PP8_{full}}$) und Untergrund tW ($\mathrm{PP8_{full}}$) (siehe Abs. 5.1) mit den entsprechenden zwischen 2015 und 2018 aufgenommenen experimentellen Daten verglichen. In Abbildung 7.1 ist dies für den Transversalimpuls der Leptonen und b-Jets sowie für die fehlende Transversalenergie $E_{\mathrm{T}}^{\mathrm{miss}}$ und die berechnete invariante Masse des Elektron-Myon-Paars $m_{e\mu}$ dargestellt. Der Untergrundprozess tW ist stark unterdrückt und macht nur etwa 3% der Statistik aus, s. Tabelle 5.2.

[1)] Aufgrund von Strahlungskorrekturen und Skalenabhängigkeiten im Rahmen der Quantenfeldtheorie weicht der Wert von $g_t^{\mathrm{SM}} = 1$ ab. Im Folgenden wird jedoch ein Standardmodellwert von $g_t = 1$ angenommen.

Im niederenergetischen Bereich überstiegen die Daten die MC-Simulation um 1-3%. In den verwendeten Ereignisgeneratoren (s. Abs. 5.1) ist die Beschreibung vom Toponium [127] und Effekten höherer Ordnung im Bereich der Produktionsschwelle [128] nicht implementiert, sodass eine geringe Abweichung der Simulation von den Daten in diesem Bereich erwartet ist. Eine zunehmende Abweichung für hohe Energien im p_T-Spektrum der Leptonen und b-Jets ist erkennbar. Dies ist eine bekannte Eigenschaft der durch POWHEG+PYTHIA8 simulierten Daten in Messungen des $t\bar{t}$- Wirkungsquerschnitts [129, 130] und wird auf Fehlmodellierungen in der Simulation zurückgeführt. Eine Verbesserung kann durch NNLO-QCD Berechnungen und NLO elektroschwache Korrekturen erzielt werden [131], was in den hier verwendeten Generatoren nicht implementiert ist. Das Transversalimpulsspektrum der Leptonen in den simulierten Ereignissen kann durch eine Umgewichtung an die Daten angepasst werden, sodass eine perfekte Übereinstimmung vorliegt. Es wird untersucht, ob sich dies positiv auf die Übereinstimmung der Modellvorhersage zwischen Simulation und Daten auswirkt (s. Abb. 7.2). Es wird hier vorweg genommen, dass dabei jedoch nur ein vernachlässigbar kleiner Unterschied festgestellt wird und die Umgewichtung des Transversalimpulsspektrums der Leptonen für die weiteren Untersuchungen nicht unternommen wird. Die invariante Masse $m_{e\mu}$ zeigt eine gute Übereinstimmung zwischen Simulation und Daten und keine große Abweichung für hohe Energiewerte. Da diese Größe einen großen Einfluss auf die Leistung des neuronalen Netzwerks hat (s. Diskussion in Abs. 6.3.1), ist der geringe Einfluss der Umgewichtung der Transversalimpulse der Leptonen nachvollziehbar.

In dieser Arbeit sind bisher nur simulierte Datensätze für die Evaluierung des neuronalen Netzwerkes verwendet worden. Im Folgenden wird die Ausgabe des Netzwerks aus den simulierten Datensätzen $t\bar{t}$ (PP8_{full}) und tW (PP8_{full}) (siehe Abs. 5.1) mit den experimentellen Daten entsprechend einer aufgenommenen integrierten Luminosität von 140fb^{-1} verglichen. Die Anzahl der erwarteten Ereignisse nach Gl. 5.2 zeigt, dass die Gesamtereignisse aus etwa 97% $t\bar{t}$ und aus etwa 3% tW Untergrund bestehen.

Abbildung 7.2 zeigt den Vergleich der rekonstruierten invarianten Masse für die Monte-Carlo Simulationen und Daten. Der Übersicht halber wird eine logarithmische Darstellung gewählt, um den geringen Untergrundbeitrag sichtbarer zu machen. Im unteren Teil der Abbildung ist das Verhältnis genauer gezeigt. Im Bereich der Produktionsschwelle < 400 GeV überwiegt die Datenvorhersage mit einer Abweichung von etwa 5% von der MC-Simulation. Dies kann wieder auf die Vernachlässigung von Schwelleneffekten in den Simulationen zurückgeführt werden [127, 128]. Für höhere Werte in $m_{t\bar{t}}^{\text{reco}}$ wechselt sich das Verhältnis und die MC-Simulation zeigt eine leichte Unterschätzung der Ereignisse. Die Abweichung beträgt jedoch maximal etwa 5%, was auf eine gute allgemeine Übereinstimmung der Simulation mit den Daten hindeutet.

Eine vollständige Berücksichtigung aller relevanten systematischen Unsicherheiten deckt diese Abweichungen im Idealfall ab und ermöglicht eine genauere Analyse. Dabei kann sichergestellt werden, dass die zunächst gute Übereinstimmung kein „Glück“ in der Optimierung und Ausgabe des Modells ist. Dies ist für die vollständige Abschätzung der Top-Yukawa Kopplung außerdem notwendig. Zu den systematischen Unsicherheiten gehören Unsicherheiten in der Modellierung (von z.B. Partonschauer, Hadronisierung, Partondichtefunktionen, Top-Quark-Masse, ...) sowie experimentelle Unsicherheiten (z.B. Luminositätsmessung, pile-up, Auflösung und Energieskala

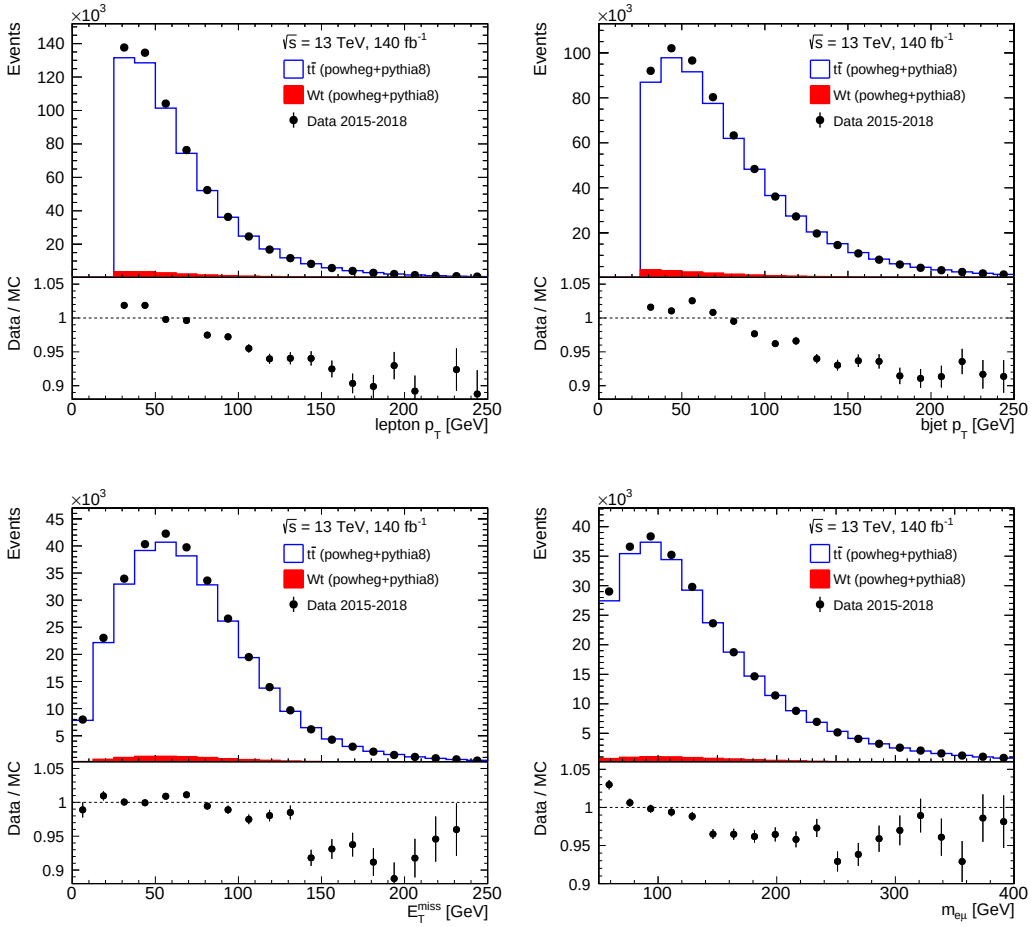

Abbildung 7.1: Vergleich von Monte-Carlo Simulation und Daten für den Transversalimpuls der Leptonen (oben links), den Transversalimpuls der b-Jets (oben rechts), die fehlende Transversalenergie (unten links) und die invariante Masse des Elektron-Myon-Paars (unten rechts). Im unteren Bereich ist dazu jeweils das Verhältnis samt statistischer Fehler gegeben. Die simulierten Daten sind aufgeteilt in den Hauptprozess $t\bar{t}$ (blau) und dem dominanten Untergrund tW (rot).

von Jets, Triggern und Identifikation von Leptonen, ...). Die in Abbildung 7.1 beobachtete Zunahme der Abweichung für höhere Energiewerte in den Transversalimpulsspektren der Leptonen und b-Jets wird bei der invarianten Masse $m_{t\bar{t}}^{\text{reco}}$ des neuronalen Netzwerks nicht beobachtet, was darauf hindeutet, dass es durch die invarianten Massen als Trainingsvariablen (s. Abs. 6.3.1) nur wenig von den Fehlmodellierungen der einzelnen Transversalimpulsspektra abhängt.

7.2 Bestimmung der Top-Yukawa Kopplung

Folgend wird das Verhältnis $Y_t = g_t/g_t^{\text{SM}}$ der Top-Yukawa Kopplung g_t zum SM-Wert g_t^{SM} definiert und untersucht. Elektroschwache Korrekturen zur $t\bar{t}$-Produktion gehen in Ordnung $\alpha_s^2\alpha_{\text{weak}}$ in den Wirkungsquerschnitt ein. Dabei haben diese nur einen kleinen Effekt auf den Gesamtwirkungsquerschnitt, weshalb sie in den Simulationen der Datensätze nicht berücksichtigt

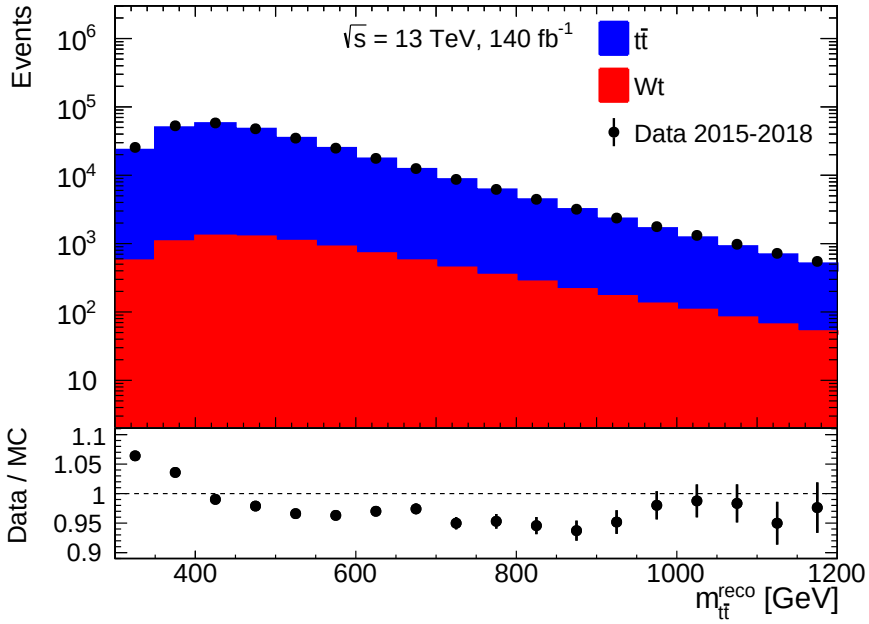

Abbildung 7.2: Vergleich der invarianten Masse $m_{t\bar{t}}^{\text{reco}}$ der Modellvorhersage für MC-Simulation und Daten. Die Simulation teilt sich in $t\bar{t}$ (blau) und Untergrund tW (rot) auf. Aufgrund des geringen Untergrunds wird eine logarithmische Darstellung gewählt. Das Verhältnis von Daten zu Simulation im unteren Bereich samt des statistischen Fehlers gezeigt.

werden. Wie in Abs. 5.2 beschrieben, ist die Korrektur durch den Austausch eines Higgs-Bosons zwischen dem produzierten $t\bar{t}$-Paar (siehe Abb. 5.2) proportional zu Y_t^2. In kinematischen Regionen, wo diese Korrekturen groß werden, kann die Form von kinematischen Verteilungen deutlich verändert werden. Physikalische Größen, die eine solche Empfindlichkeit besitzen, sind beispielsweise die invariante Masse $m_{t\bar{t}}$ des $t\bar{t}$-Paares oder der Azimuthwinkel $\cos\theta^*$ des Top-Quarks relativ zur Strahlachse.

Im Rahmen dieser Arbeit wird die rekonstruierte Masse des neuronalen Netzwerks $m_{t\bar{t}}^{\text{reco}}$ für die Untersuchung benutzt. Unter Verwendung des HATHOR v2.1-b3 Pakets [117] werden die Korrekturen als Funktion von $m_{t\bar{t}}$ und $\cos\theta^*$ implementiert (für Details siehe Abs. 5.2.2). Die dileptonischen Ereignisse werden damit sowohl auf Teilchen-Level (direkt aus der Ereignissimulation nach der Hadronisierung und vor Detektoreffekten), als auch auf Detektor-Level nach Gleichung 5.3 für verschiedene Werte von Y_t umgewichtet, wobei $Y_t = 1$ als Standardmodellvorhersage angenommen wird. Aus dem Vergleich der umgewichteten $t\bar{t}$-Datensätzen mit experimentellen Daten oder Pseudo-Daten kann letztendlich ein Wert für Y_t abgeschätzt werden. Eine Analyse in dieser Form wurde von der CMS-Kollaboration im semi-leptonischen und dileptonschen Kanal bereits durchgeführt. [19, 20]

Der Austausch eines virtuellen Higgs-Bosons zwischen dem produzierten $t\bar{t}$-Paar ist proportional zu Y_t^2. Eine Parametrisierung in Y_t^2 erlaubt damit die spätere Behandlung eines linearen Problems. In Abbildung 7.3 (oben) ist die invariante Masse $m_{t\bar{t}}^{\text{MC truth}}$ auf dem Teilchen-Level

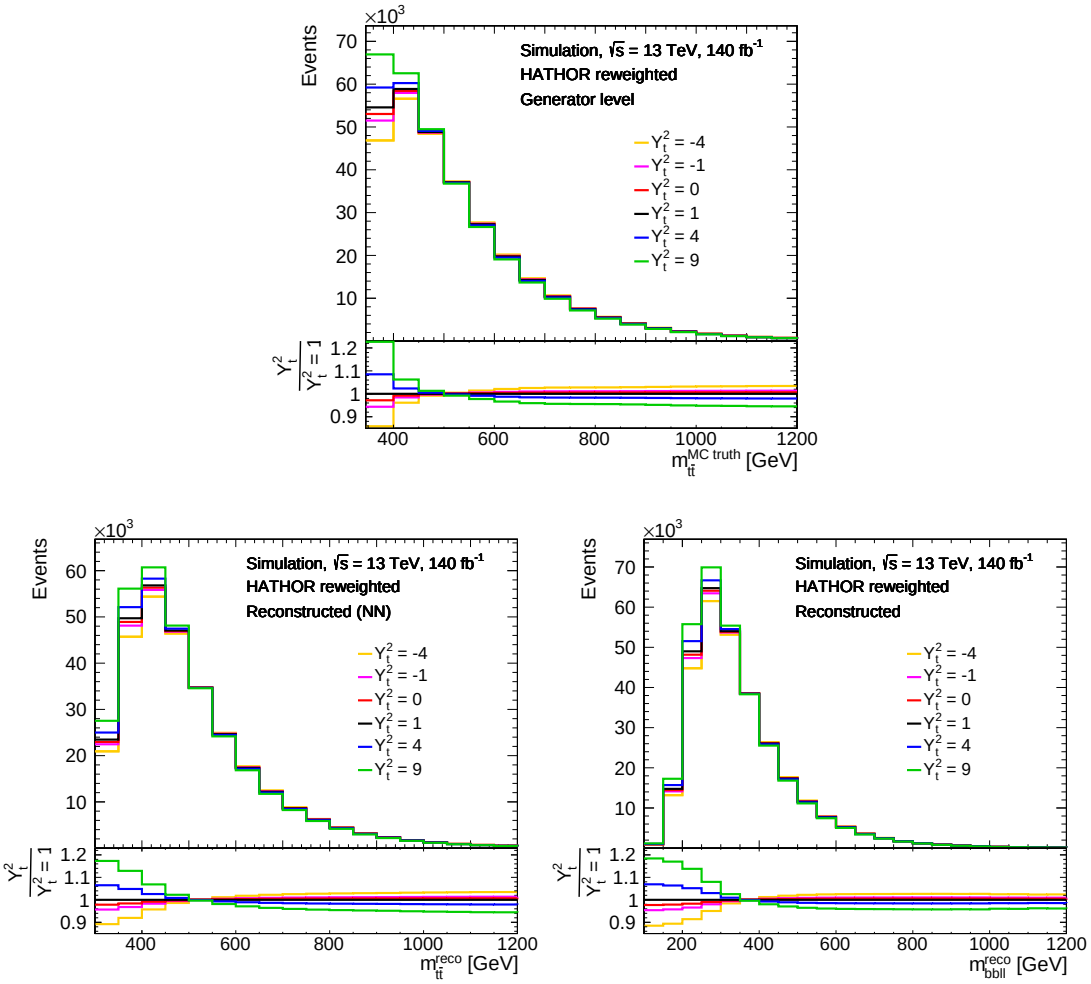

Abbildung 7.3: Abhängigkeit der sensitiven Variablen $m_{t\bar{t}}^{\text{MC truth}}$ (Teilchen-Level, oben), $m_{t\bar{t}}^{\text{reco}}$ (neuronales Netzwerk, unten links), m_{bbll}^{reco} (unten rechts) für verschiedene Werte von Y_t^2. Die Verteilungen werden nach Gleichung 5.2.2 durch HATHOR berechnete elektroschwache Korrekturen zum Wirkungsquerschnitt umgewichtet. Dazu ist jeweils das Verhältnis zum Standardmodellwert von $Y_t^2 = 1$ im unteren Bereich gegeben.

für die Werte von $Y_t^2 = (-4, -2, 0, 1, 4, 9)$ dargestellt. Insbesondere im Bereich der $t\bar{t}$-Produktionsschwelle ($\approx 2m_t$), was einer geringen relativen Geschwindigkeit zwischen dem produzierten Top-Quark und Anti-Top-Quark entspricht, ist der Wirkungsquerschnitt sensitiv auf elektroschwache Korrekturen [132]. So entspricht eine Verdopplung der Top-Yukawa-Kopplung vom SM-Wert im Bereich der Produktionsschwelle einer Änderung von etwa 9%. Außerhalb des Bereiches der Produktionsschwelle besteht keine Empfindlichkeit zum Wert von Y_t^2. Eine Kombination mit kinematischen Variablen, die eine Empfindlichkeit in höheren Energiebereichen aufweisen (z.B. der Streuwinkel $\cos\theta^*$ des Top-Quarks), kann die Genauigkeit der Messung erhöhen.

Alternativ kann für die invariante Masse des vollständigen $t\bar{t}$-Systems (hier vom neuronalen

Netzwerk gegeben) die invariante Masse m_{bbll}^{reco} der beiden Leptonen und b-Jets im Endzustand verwendet werden. Dies bietet den Vorteil, dass die Schwierigkeit der vollständigen Rekonstruktion mit zwei Neutrinos umgangen wird und trotzdem eine merkliche Empfindlichkeit auf den Wert von Y_t^2 existiert. Außerdem ist die Berechnung aus den jeweiligen Vierervektoren der einzelnen Leptonen und b-Jets simpel. In Abbildung 7.3 sind die beiden rekonstruierten kinematischen Verteilungen $m_{t\bar{t}}^{\text{reco}}$ (unten links) und m_{bbll}^{reco} (unten rechts) für verschiedene Werte von Y_t^2 dargestellt. Beim Übergang von kinematischen Verteilungen auf dem Teilchen-Level zu rekonstruierten Verteilungen des Detektor-Levels ist eine Reduktion der Empfindlichkeit zu Y_t^2 aufgrund von systematischen Unsicherheiten und experimentellen Auflösungseffekten erkennbar. So entspricht eine Verdopplung der Top-Yukawa Kopplung beider Verteilungen in Bereichen sensitiv zu Y_t^2 eine Änderung von nur noch etwa 6,5%.

7.2.1 Statistische Analyse

Im folgenden Abschnitt werden die statistischen Methoden erläutert, die zur Schätzung der Top-Yukawa Kopplung aus den kinematischen Verteilungen $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} verwendet werden. Die am häufigsten verwendete Methode zur Schätzung physikalischer Parameter in Analysen der Hochenergiephysik ist die Maximum-Likelihood-Methode [133, 134]. Dabei wird eine Likelihood-Funktion $\mathscr{L}$ maximiert, die angibt, wie wahrscheinlich die beobachteten Daten für verschiedene Werte der im Modell enthaltenen Parameter sind. Für die Schätzung von Y_t wird ein Profile-Likelihood-Fit [135] unter Berücksichtigung systematischer Unsicherheiten im TRExFitter-Framework [136] vorgenommen, welches auf der RooFit-Bibliothek [137] für die statistische Modellierung basiert. Wichtig dabei sind aufgrund des beschriebenen Einflusses der elektroschwachen Korrekturen (s. Abs. 5.2.2) vor allem Systematiken, die die Form des Wirkungsquerschnitts ändern (und nicht die Gesamtnormalisierung wie beispielsweise die Unsicherheiten der Luminosität oder des $t\bar{t}$-Wirkungsquerschnitts).

Die hier präsentierte Schätzung für die Top-Yukawa-Kopplung wird anhand der in Abbildung 7.3 gezeigten Verteilungen $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} extrahiert. Die Veränderung von Y_t^2 sorgt in bestimmten Bereichen der kinematischen Verteilungen für eine Änderung der erwarteten Ereignisse (s. Abs. 7.2). Die Verteilungen sind jeweils nach Gleichung 5.2.2 für verschiedene ganzzahlige Werte von Y_t^2 gewichtet, welche im Folgenden *Templates* genannt werden. Um eine kontinuierliche Anpassung in Y_t^2 zu gewährleisten, muss zwischen den Templates interpoliert werden. Die Korrektur durch den Austausch eines virtuellen Higgs-Bosons zwischen dem produzierten $t\bar{t}$-Paar ist proportional zu Y_t^2 (s. Abs. 5.2.2), womit Y_t^2 als der im Fit gesuchte Parameter (engl. *paramter of interest* POI) festgelegt wird. Durch die Wahl von Y_t^2 als POI produziert eine lineare Interpolation zwischen den Y_t^2-Templates genaue Resultate. Jedes Template t erhält einen Normalisierungsfaktor $w_i^t(\text{POI})$, der abhängig vom Wert für Y_t^2 ist. Die Anzahl an Ereignissen S_i im Bin i ist somit:

$$S_i = \sum_t w_i^t(\text{POI}) \cdot T_i^t, \tag{7.1}$$

wobei T_i^t die Anzahl der Ereignisse in Bin i des Templates t ist. In TRExFitter [136] wird dies durch die *template morphing* Methode implementiert. Angenommen der „wahre“ Wert für die Top-Yukawa Kopplung wäre $Y_t^2 = 1{,}5$, so würde der Fit auf ein Ergebnis konvergieren, in dem

die Templates $Y_t^2 = 1$ und $Y_t^2 = 2$ auf 0,5 mal der ursprünglichen Normalisierung angepasst werden und die restlichen Templates nicht zum Fit beitragen (Normalisierung von 0).

Die in Abbildung 7.3 gezeigten Verteilungen sind in Energieintervallen (Bins) von 50 GeV unterteilt, was genügend Statistik in jedem Bin sicherstellt. Die Wahrscheinlichkeit, dass n_i^{obs} Ereignisse in Bin i beobachtet werden, ist durch eine Poisson-Verteilung gegeben:

$$\mathcal{P}(n_i^{\text{obs}}|\mu_i(Y_t,\theta)) = \frac{\mu_i^{n_i^{\text{obs}}}(Y_t,\theta)e^{-\mu_i(Y_t,\theta)}}{n_i^{\text{obs}}!}, \tag{7.2}$$

wobei μ_i die Anzahl der erwarteten Ereignisse ist, die sich aus der Summe der Beiträge von $t\bar{t}$-Signal $s_i(\theta)$ und Untergrundprozessen $u_i(\theta)$ zusammensetzt und wie bereits erläutert abhängig von Y_t und den systematischen Unsicherheiten θ ist. Jede systematische Unsicherheit geht als sogenannter Störparameter[2)] (engl. *nuisance parameter*) in den Fit ein, die durch eine Normalverteilung $\mathcal{G}(\theta)$ um ihren Erwartungswert beschränkt werden. Die relative Wirkung der systematischen Unsicherheiten der einzelnen Templates wird als gleich groß angenommen und aus der nominalen Template $Y_t^2 = 1$ (SM-Wert) auf die anderen Templates übertragen. Da die beobachteten Ereignisse aus statistisch unabhängigen Messungen stammen, wird mit den genannten Beziehungen die Likelihood-Funktion $\mathscr{L}$ durch Multiplikation der individuellen Poisson-Verteilungen in jedem Bin i aufgestellt:

$$\mathscr{L} = \prod_{i\in\text{bins}} \mathcal{P}\left(n_i^{\text{obs}} \,\middle|\, s_i(\theta)\cdot R_i^{\text{EW}}(Y_t) + u_i(\theta)\right) \cdot \prod_{j\in\text{syst}} \mathcal{G}(\theta_j), \tag{7.3}$$

wobei $R_i^{\text{EW}}(Y_t)$ den Umgewichtungsparameter für verschiedene Werte von Y_t (s. Gl. 5.2.2) und $\mathcal{G}(\theta)$ die Normalverteilung der systematischen Unsicherheit j darstellt, die multiplikativ zur Likelihood-Funktion hinzugefügt wird [138]. Die statistische Unsicherheit in einem Bin wird im Fit als ein zusätzlicher Störparameter in Form eines Poisson-Terms in Gleichung 7.3 multipliziert und häufig als Gamma-Parameter bezeichnet [139]. Die Likelihood-Funktion kann nun bezüglich der Störparameter θ_j für jeden Wert von $\mu_i = s_i(\theta)\cdot R_i^{\text{EW}}(Y_t) + u_i(\theta)$ maximiert werden. Es wird über die Störparameter „profilliert“ und gibt eine Schätzung der Unsicherheiten ab, was den Namen *Profile Likelihood Methode* bezeichnet. Die Maximierung der „profilierten“ Likelihood-Funktion kann dann eine Schätzung für den Wert von μ_i, in diesem Fall also den besten Wert von Y_t (bzw. Y_t^2) geben. In der Praxis wird aus numerischen Gründen häufig statt der Maximierung von Gleichung 7.3 eine Minimierung der negativen Log-Likelihood-Funktion ($-\ln\mathscr{L}$) vorgenommen.

7.2.2 Durchführung eines Asimov-Fits

Eine Durchführung der beschriebenen Fit-Prozedur anhand der experimentellen Daten (aus denen der Wert für Y_t letztlich ermittelt wird) erfordert die Berücksichtigung aller relevanten systematischen Unsicherheiten, von denen am Ende des Abschnitts 7.1 einige genannt sind. Im Zeitrahmen dieser Arbeit wird nur die Systematik der Partonschauer- und Fragmentationsalgo-

[2)] Störparameter sind in der Schätzung nicht vom primären Interesse. Sie beeinflussen jedoch die Genauigkeit der Messung und müssen berücksichtigt werden.

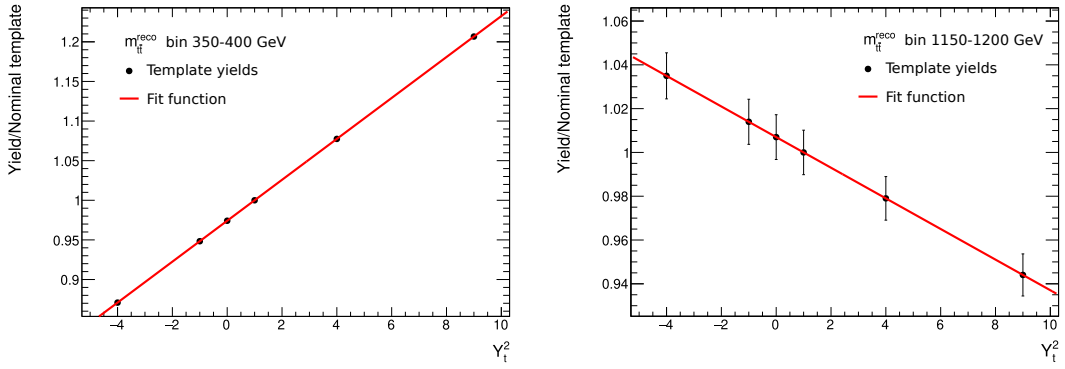

Abbildung 7.4: Die schwarzen Punkte zeigen die beobachteten Ereignisse (yields) in den $m_{t\bar{t}}^{\text{reco}}$ Bins [350-400] GeV und [1150-1200] GeV für verschiedene Werte von Y_t^2 im Verhältnis zum nominalen Wert $Y_t^2 = 1$. Die lineare Fit-Funktion ist in rot dargestellt. Statistische Unsicherheiten sind berücksichtigt.

rithmen (s. Abs. 2.2.2) untersucht. Eine nützliche Methode, um das Fit-Modell vor der Anpassung an experimentellen Daten zu prüfen, ist ein Fit an einem simulierten *Asimov*-Datensatz. Dieser wird für $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} aus dem $t\bar{t}$ (PP8$_{\text{full}}$)-Sample erstellt und ist dadurch charakterisiert, dass er die genau erwartete Anzahl an Ereignissen (umgewichtet nach Gl. 5.2.2 für den SM-Wert von $Y_t^2 = 1$) in jedem Bin enthält. Zwar wird damit keine Anpassung an tatsächlichen Daten vorgenommen und das Ergebnis für Y_t^2 durch Forderung von $Y_t^2 = 1$ „vorgegeben“, jedoch werden die Unsicherheiten genauso wie in der vollständigen statistischen Analyse berücksichtigt, wodurch die Sensitivität der Analyse eingeschätzt werden kann. Außerdem kann so ein Vergleich aus den Verteilungen $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} im Fit gezogen werden.

Zunächst wird der Likelihood-Fit anhand der Pythia8-Asimov-Daten nur unter Berücksichtigung der Unsicherheit des $t\bar{t}$-Wirkungsquerschnitts als Normierungsparameter durchgeführt. Außerdem wird auf den Untergrundbeitrag verzichtet (vgl. Gl. 7.3). Damit ergeben sich aus den genannten Verteilungen die in Tabelle 7.1 (mittlere Spalte) gezeigten Ergebnisse. Dabei zeigt die vom neuronalen Netzwerk rekonstruierte invariante Masse des $t\bar{t}$-Paares eine etwa 12% geringere Unsicherheit in Y_t^2 (damit $\approx 3{,}5\%$ in Y_t). Der Zentralwert von $Y_t^2 \approx 1$ ist bei der Anpassung an die Asimov-Daten (die der nominalen Verteilung $Y_t^2 = 1$ entsprechen) erwartet. Die Genauigkeit des Zentralwertes hängt u.a. von der gewählten Interpolation zwischen den Templates im Fit ab. In Abbildung 7.4 ist die Interpolation der Templates für Werte zwischen $Y_t^2 = -4$ und $Y_t^2 = 9$ in zwei Bins [350-400] GeV und [1150-1200] GeV für den Fit in $m_{t\bar{t}}^{\text{reco}}$ gezeigt (für m_{bbll}^{reco} sind die Ergebnisse gleichartig). Genauer ist das Verhältnis der Ereignisse (engl. *yields*) zur nominalen $Y_t^2 = 1$-Template dargestellt, wobei eine klare lineare Abhängigkeit von Y_t^2 erkennbar ist. Die umgekehrte Steigung (d.h. eine Umkehrung der Abhängigkeit von Y_t^2) ist, wie bei den Verhältniskurven in Abbildung 7.3 ab etwa 500 GeV erkennbar. Kleinere Bins (beispielsweise 25 GeV statt 50 GeV) ändern das Resultat nicht merklich.

In Abbildung 7.5 (links) ist das Resultat des Profile-Likelihood-Fits in $m_{t\bar{t}}^{\text{reco}}$ für die Werte von $Y_t^2 = 0$ bis $Y_t^2 = 3$ dargestellt. Genauer ist das Verhältnis von $-\ln \mathscr{L}(Y_t^2)$ zu dem besten Fit-Wert $-\ln \mathscr{L}(\hat{Y}_t^2)$ für die nach $Y_t^2 = 1$ produzierten Pythia8-Asimov-Daten gezeigt. Dies ist eine gängige Methode zur Visualisierung der Profile-Likelihood und wird auch Likelihood-

Tabelle 7.1: Ergebnisse des Asimov-Fits für die Verteilungen $m_{t\bar{t}}^{\text{reco}}$ (Neuronales Netzwerk) und m_{bbll}^{reco}. Gegeben ist der beste Fit-Wert für Y_t^2 an Asimov-Daten (für $Y_t^2 = 1$ generiert) produziert durch Pythia8 und Herwig7 unter sonst gleichen Fit-Bedingungen, sowie die Unsicherheit der Messung.

Verteilung	Y_t^2 Pythia8 Asimov-Fit	Y_t^2 Herwig7 Asimov-Fit
$m_{t\bar{t}}^{\text{reco}}$ (NN)	$1 \pm 0{,}197$	$0{,}51 \pm 0{,}197$
m_{bbll}^{reco}	$1 \pm 0{,}223$	$0{,}48 \pm 0{,}223$

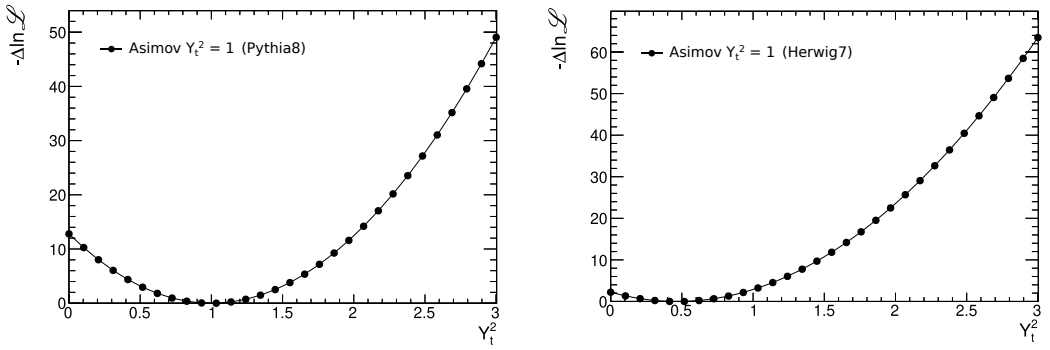

Abbildung 7.5: Das Ergebnis des Log-Likelihood-Fits an $m_{t\bar{t}}^{\text{reco}}$ der Asimov-Daten für $Y_t^2 = 1$ mit PYTHIA8 produziert (links) und HERWIG7 produziert (rechts), dargestellt als das Verhältnis von $-\ln \mathscr{L}(Y_t^2)$ zu dem besten Fit-Wert $-\ln \mathscr{L}(\hat{Y}_t^2)$. Da die Templates (s. Text) durch PYTHIA8 generiert wurden, ist ein bester Fit-Wert $Y_t^2 \neq 1$ für das HERWIG-Sample erkennbar.

Quotienten-Test genannt. Auf dessen Basis kann die statistische Signifikanz von Schätzungen von Y_t beurteilt werden, oft anhand von Konfidenzintervallen [135]. Das Minimum der Kurve entspricht dabei dem besten Fit-Wert von Y_t^2 (bei den verwendeten PYTHIA8Asimov-Daten natürlich $Y_t^2 = 1$). Die Likelihood-Kurve für m_{bbll}^{reco} ist analog berechnet und erzielt gleiche Ergebnisse.

Systematische Unsicherheit durch Partonschauer und Hadronisierung

Die oben angesprochenen Y_t^2-Templates sind durch die Generatoren POWHEG [50–52] und PYTHIA8 [57] generiert und im obigen Fit mit den entsprechend generierten Asimov-Daten verglichen. Unter ansonsten gleichen Fit-Einstellungen erfolgt nun ein Vergleich mit einem durch POWHEG und HERWIG7 [58,59] generierten Datensatz (wobei die Templates unverändert bleiben). Das Ergebnis des Likelihood-Scans ist in Abbildung 7.5 (rechts) gezeigt. Im Vergleich zum Fit mit den vorherigen POWHEG+PYTHIA8-Asimov-Daten ist nun ein von $Y_t^2 = 1$ abweichender bester Fit-Wert zu erwarten. Dies gibt Rückschlüsse auf den Einfluss der verschiedenen Generatoren auf die statistische Messung, insbesondere der systematischen Unsicherheit durch Partonschauer und Hadronisierung. Die Ergebnisse für die Verteilungen $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} sind in Tabelle 7.1 (rechte Spalte) mit den entsprechenden Unsicherheiten für Y_t dargestellt. Es ist eine Änderung des Zentralwertes von $Y_t^2 = 1$ (Asimov-Fit durch Pythia8 Samples) auf $Y_t^2 = 0.51$ bzw. $Y_t^2 = 0.48$ für $m_{t\bar{t}}^{\text{reco}}$ bzw. m_{bbll}^{reco} zu erkennen. Dies zeigt, dass die systematische Unsicherheit

durch die verschiedenen Partonschauer- und Hadronisierungsmodelle von PYTHIA8 und HERWIG7 in beiden Fits eine Abweichung von etwa 50% in Y_t^2, d.h. etwa 30% in Y_t, beträgt. Damit beeinflusst diese systematische Unsicherheit die Messung erheblich, was in der analogen Analyse im Lepton+Jets $t\bar{t}$-Zerfallskanal ebenfalls beobachtet werden kann [126].

Eine verbesserte Anpassung der durch die Generatoren PYTHIA8 und HERWIG7 simulierten Daten an experimentelle Daten kann die Genauigkeit der Messung von Y_t verbessern. Eine Rekonstruktionsmethode, die empfindlicher auf eine Änderung von Y_t ist, als die hier gezeigte Verteilung des neuronalen Netzwerks, kann ebenfalls präzisere Ergebnisse liefern. Hier ist es wichtig zu erwähnen, dass ein erweitertes neuronale Netzwerk, bei dem sich auf die Vermeidung eines Bias in $m_{t\bar{t}}$ konzentriert wird, eine Verbesserung in der Präzision liefer kann. Ein Bias, wie in Abschnitt 6.3 besprochen, würde eine zusätzliche systematische Unsicherheit liefern. Obwohl das neuronale Netzwerk eine leicht verbesserte Ungenauigkeit in dem Asimov-Fit zeigt, kann erst bei der Berücksichtigung aller systematischen Sicherheiten ein Fazit geschlossen werden, welche Methode am Ende die beste Präzision liefert. Auch erst dann wird ein Fit auf experimentellen Daten vorgenommen, der im Prinzip jedoch äquivalent mit der Fit-Routine ist, die hier bereits vorgestellt wurde. Dies sind zukünftige Untersuchungen, die über den Zeitrahmen dieser Arbeit hinausgehen.

8 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein neuronales Regressionsnetzwerk für die Rekonstruktion der invarianten Masse von Top-Anti-Top Paaren $m_{t\bar{t}}^{\text{reco}}$ im dileptonischen Zerfallskanal unter Verwendung simulierter Daten der Run-2 Datennahmeperiode am ATLAS-Experiment entwickelt. Grundsätzlich versucht das Regressionsmodell während der Trainingsphase den Betrag der Differenz zwischen der ausgegebenen Verteilung $m_{t\bar{t}}^{\text{reco}}$ und der auf Generator-Level berechneten Monte-Carlo Zielverteilung $m_{t\bar{t}}^{\text{MC truth}}$ zu minimieren. In der Trainingsphase wurden die Eigenschaften des Modells im Lernprozess untersucht. Nur ab einer bestimmten Komplexität des Netzwerks und einer damit verbundenen hohen Anzahl an internen Parametern zeigt die Verlustfunktion nach einem anfänglichen Abfall in ein lokales Minimum später im Training ein zweites Minimum, in dem das Modell verweilt. Der Übergang in das zweite Minimum bedeutet ein geringerer Verlustwert, d.h. eine geringe Differenz zwischen der Netzwerkausgabe $m_{t\bar{t}}^{\text{reco}}$ und der Zielfunktion $m_{t\bar{t}}^{\text{MC truth}}$, jedoch wird die ausgegebene Verteilung $m_{t\bar{t}}^{\text{reco}}$ während des Übergangs stark modifiziert. Anhand des relativen Fehlers ($m_{t\bar{t}}^{\text{reco}}$ - $m_{t\bar{t}}^{\text{MC truth}}$ / $m_{t\bar{t}}^{\text{MC truth}}$) in verschiedenen Massenintervallen von $m_{t\bar{t}}^{\text{MC truth}}$ wurde der Unterschied zum Zeitpunkt des ersten Minimums, während des Übergangs und im zweiten Minimum verglichen. Die Ausgabeverteilung im ersten Minimum verschiebt die Produktionsschwelle zu geringeren Energien, sodass Ereignisse unterhalb von $2m_t \approx 345$ GeV auftreten. Beim Übergang ins zweite Minimum zeigt sich ein Bias, bei dem das Modell bevorzugt Ergebnisse im Bereich der Produktionsschwelle von $m_{t\bar{t}}^{\text{MC truth}}$ generiert. Um eine möglichst Bias-freie Verteilung zu gewährleisten, wird das Training vor dem Erreichen des zweiten Minimums gestoppt. Für zukünftige Untersuchungen ist die Konstruktion einer Verlustfunktion bedeutsam, die einen entstehenden Bias verhindern kann. Außerdem können komplexere Netzwerkstrukturen, wie ein sekundäres Diskriminationsmodell implementiert werden, die einen Bias während des Training überwachen können.

Verschiedene Kombinationen von Hyperparametern des neuronalen Netzwerks wurden mit einer GridSearch des KERAS-Pakets getestet. Die Struktur wird auf drei tiefe verborgene Schichten mit jeweils 30 Neuronen und ReLU-Aktivierungsfunktion festgelegt. Die Batchgröße ($N_B = 128$), die die Anzahl der Ereignisse angibt, über die die Verlustfunktion minimiert wird, die Wahl des Optimierungsprozesses (ADAM) und dessen Lernrate ($\alpha = 0{,}0003$) sowie die Anzahl der dem Modell zur Verfügung gestellten $t\bar{t}$-Trainingsereignisse (≈4 Millionen) beeinflussen hauptsächlich Dauer des Trainings zu dem festgelegten Abbruchspunkt. Die Werte sind für eine optimale Trainingseffizienz angepasst. Verschiedene physikalische Observablen wurden für die Eingabe in das Netzwerk getestet. Die beste Kombination lieferten Paare von invarianten Massen $m_{e\mu}$, m_{eb_1}, m_{eb_2}, $m_{\mu b_1}$, $m_{\mu b_2}$ und $m_{b_1 b_2}$ der Elektronen e, Myonen μ und b-Jets im Endzustand, zusammen mit der fehlenden Transversalenergie $E_{\text{T}}^{\text{miss}}$, wobei insbesondere $m_{e\mu}$ und $E_{\text{T}}^{\text{miss}}$ den größten Einfluss auf das Training des Modells zeigten. Diese *high-level* Variablen erzielten bessere Ergebnisse, als die *low-level* Features $(p_{\text{T}},\eta,\phi,E)$ der jeweiligen Leptonen und b-Jets. Mit komplexeren Netzwerkstrukturen könnte die Bedeutung von *low-level* Variablen steigen, sodass ein Vergleich der Eingabevariablen für zukünftige Änderung an der Modellstruktur essenziell ist.

Weiterhin wurde der Einfluss unterschiedlich generierter Datensätze auf die Ausgabeverteilung $m_{t\bar{t}}^{\text{reco}}$ untersucht. Um den Einfluss verschiedener Simulationen, d.h. Modellierungen (z.B. verschiedene Partonschauer- und Hadronisierungsmodelle) des $t\bar{t}$-Prozesses zu zeigen, wurden ein POWHEG+PYTHIA8 mit einem POWHEG+HERWIG7 Datensatz verglichen. Dabei zeigte sich ebenfalls eine hohe Übereinstimmung mit Abweichungen von 2-3% bis etwa 700 GeV, die folgend bis auf rund 5% ansteigen. Auch wurde ein Vergleich von $m_{t\bar{t}}^{\text{reco}}$ zwischen Simulation und experimen-

tellen Daten, die von 2015 und 2018 entsprechend einer integrierten Luminosität von 140 fb^{-1} aufgenommen wurden, vorgenommen. Dabei wurde neben dem $t\bar{t}$-Signal auch der dominante Untegrundprozess der tW-Produktion einbezogen. Abweichungen von 5% waren im Bereich der Produktionsschwelle zu beobachten, die durch eine Implementation von Schwelleneffekten höherer Ordnung oder dem Toponium in die Simulation verringert werden können [127, 128]. Für den Rest des Spektrums waren ebenfalls Abweichungen von bis zu 5% beobachtbar, die durch eine Berücksichtigung von Effekten höherer Ordnungen in den Generatoren und einer Anpassung der internen Generatorparameter reduziert werden können. Erst unter Berücksichtigung experimenteller Unsicherheiten, wie beispielsweise bei der Luminositätsmessung, der Rekonstruktion der verwendeten Physikobjekte und der Unsicherheit sowie Modellierung des Pile-Ups, kann eine genauere Aussage zur Übereinstimmung getroffen werden. Ebenso müssen theoretische Unsicherheiten in der Modellierung, den Partonverteilungsfunktionen (PDFs) und der verwendeten Top-Quark-Masse berücksichtigt werden. Diese Untersuchungen waren im Rahmen dieser Arbeit nicht möglich und können zukünftig analysiert werden.

Die Rekonstruktion der invarianten Top-Anti-Top Masse $m_{t\bar{t}}^{\text{reco}}$ ist in dieser Arbeit durch eine Messung der Top-Yukawa Kopplung im Verhältnis zum Standardmodellwert $Y_t = g_t/g_t^{\text{SM}}$ motiviert. In der Nähe der $t\bar{t}$-Produktionsschwelle ($\approx 2m_t$) modifizieren virtuelle Korrekturen durch den Austausch eines Higgs-Bosons zwischen den produzierten Top- und Anti-Top-Quarks das differentielle Spektrum der invarianten Masse. Da die virtuelle Korrektur quadratisch in den Wirkungsquerschnitt eingeht, werden die Massenverteilungen in Y_t^2 parametrisiert, um ein lineares Problem zu behandeln. Über eine Umgewichtung der Ereignisse für verschiedene Werte von Y_t^2 wurde die invariante Masse der Top-Quarks auf dem Teilchen-Level $m_{t\bar{t}}^{\text{MC truth}}$ und Detektor-Level ($m_{t\bar{t}}^{\text{reco}}$) verglichen. Zusätzlich wurde die einfach zu berechnende Proxy-Variable m_{bbll}^{reco} als Vergleich untersucht. Eine Verdopplung von Y_t^2 im Bereich der Produktionsschwelle entspricht einer Änderung von 9% auf dem Teilchen-Level, die auf etwa 6,5% für die beiden rekonstruierten Verteilungen abnimmt. Eine letztliche Messung von Y_t durch einen in dieser Arbeit vorgestellten Maximum-Likelihood-Fit, in dem die experimentellen Daten mit dem erwarteten $t\bar{t}$-Signal für verschiedene Werte von Y_t^2 verglichen werden, ist nur unter Berücksichtigung aller systematischer Unsicherheiten möglich. Da dies den Zeitrahmen dieser Arbeit überschritt, wurde ein Fit an Asimov-Daten vorgenommen, die bereits dem erwarteten SM $t\bar{t}$-Signal $Y_t^2 = 1$ entsprechen und damit das Fit-Ergebnis „vorgeben“. Dies erlaubt trotzdem eine Bewertung verschiedener Methoden durch den Vergleich des Asimov-Fits unter gleichen Voraussetzungen und berücksichtigten Systematiken. So wurde die Empfindlichkeit der Messung für die beiden rekonstruierten Verteilungen $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} abgeschätzt. Die Fit-Ergebnisse ergaben für $m_{t\bar{t}}^{\text{reco}}$ einen Wert von $Y_t^2 = 1 \pm 0{,}197$ und für m_{bbll}^{reco} einen Wert von $Y_t^2 = 1 \pm 0{,}223$, was einer etwa 12% geringeren Unsicherheit in Y_t^2 (etwa 3,5% in Y_t) entspricht und somit eine leichte Besserung der Methode durch das neuronale Netzwerk zeigt. Doch erst durch zukünftige Messungen unter Berücksichtigung systematischer Unsicherheiten kann eine finale Aussage über die Empfindlichkeit der beiden Verteilungen getroffen werden, was letztlich auch darüber entscheidet, welche Verteilung für den endgültigen Fit an die experimentellen Daten verwendet wird. Zukünftig kann der Fit an der Verteilung $m_{t\bar{t}}^{\text{reco}}$ oder m_{bbll} durch weitere kinematische Verteilungen ergänzt werden, die in anderen Energiebereichen empfindlich auf Y_t sind, um die Präzision der Messung zu erhöhen. Dies ist beispielsweise der Streuwinkel der Top-Quarks relativ zur Strahlachse $\cos\theta^*$. Darüber hinaus ist die oben genannte potenzielle Erweiterung des neuronalen Netzwerks ent-

scheidend, um einen Bias des Modells zu vermeiden, der sonst als systematische Unsicherheit in die Messung eingeht. Zusätzlich können weitere Rekonstruktionsmethoden für $m_{t\bar{t}}^{\text{reco}}$ hinsichtlich der Präzision des Fits untersucht werden, wie alternative Deep-Learning Strukturen wie Boosted-Decision-Trees [140] oder analytische Methoden wie die Ellipsenmethode [22].

Eine systematische Unsicherheit durch verschiedene Partonschauer- und Hadronisierungsmodelle in Pythia8 und *Herwig7* generierten Asimov-Daten wurde abgeschätzt. Dabei wurde festgestellt, dass der Fit an Asimov-Daten, die mit Herwig7 erzeugt wurden, im Vergleich zu den mit Pythia8 generierten Asimov-Daten eine Abweichung von etwa 50%vom Zentralwert $Y_t^2 = 1$ für beide Verteilungen $m_{t\bar{t}}^{\text{reco}}$ und m_{bbll}^{reco} aufweist. Dies verdeutlicht den großen Einfluss dieser Modellierungsunsicherheiten auf die Messung, die durch eine verbesserte Anpassung der Generatoren Pythia8 und Herwig7 an die experimentellen Daten verringert werden kann.

Literaturverzeichnis

[1] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008.

[2] Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, September 2012.

[3] Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, September 2012.

[4] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, aug 2008.

[5] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, aug 2008.

[6] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8):083C01, 08 2020.

[7] H T Edwards. The Tevatron Energy Doubler: A Superconducting Accelerator. *Annual Review of Nuclear and Particle Science*, 35(Volume 35, 1985):605–660, 1985.

[8] CDF Collaboration. Observation of Top Quark Production in Pbar-P Collisions with the Collider Detector at Fermilab. *Physical Review Letters*, 74(14):2626–2631, April 1995.

[9] D0 Collaboration. Observation of the Top Quark. *Physical Review Letters*, 74(14):2632–2637, April 1995.

[10] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Physics Reports*, 516(1–2):1–102, July 2012.

[11] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. The minimal composite Higgs model. *Nuclear Physics B*, 719(1–2):165–187, July 2005.

[12] F. Bezrukov and M. Shaposhnikov. Why should we care about the top quark Yukawa coupling? *Journal of Experimental and Theoretical Physics*, 120(3):335–343, March 2015.

[13] Fedor Bezrukov, Javier Rubio, and Mikhail Shaposhnikov. Living beyond the edge: Higgs inflation and vacuum metastability. *Physical Review D*, 92(8), October 2015.

[14] Nello Bruscino. A gateway to new physics: direct measurement of the top Yukawa coupling to the Higgs boson, August 2017. Rheinische Friedrich-Wilhelms-Universität Bonn.

[15] The ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Physics Letters B*, 784:173–191, September 2018.

[16] The CMS Collaboration. Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(5), May 2019.

[17] The ATLAS Collaboration. Observation of four-top-quark production in the multilepton final state with the ATLAS detector. *The European Physical Journal C*, 83(6), June 2023.

[18] The CMS Collaboration. Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 847:138290, December 2023.

[19] The CMS Collaboration. Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the dilepton final state in proton-proton collisions at $\sqrt{s} = 13\ \mathrm{TeV}$. *Phys. Rev. D*, 102:092013, Nov 2020.

[20] The CMS Collaboration. Constraining the top quark Yukawa coupling from $\mathrm{t\bar{t}}$ differential cross sections in the lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13\ \mathrm{TeV}$. Technical report, CERN, Geneva, 2019.

[21] Lars Sonnenschein. Erratum: Analytical solution of $t\bar{t}$ dilepton equations [Phys. Rev. D 73, 054015 (2006)]. *Phys. Rev. D*, 78:079902, Oct 2008.

[22] Burton A. Betchart, Regina Demina, and Amnon Harel. Analytic solutions for neutrino momenta in decay of top quarks. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 736:169–178, February 2014.

[23] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep Learning and Its Application to LHC Physics. *Annual Review of Nuclear and Particle Science*, 68(1):161–181, October 2018.

[24] Burton Richter. Very high-energy electron-positron colliding beams for the study of the weak interactions. *Nucl. Instrum. Methods*, 136(1):47–60, 1976.

[25] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, New York, 2013.

[26] Christoph Berger and Gregor Herten. *Elementarteilchenphysik: von den Grundlagen zu den modernen Experimenten; 4. Auflage*. Springer Berlin, Berlin, 2023.

[27] Peter Schleper. Teilchenphysik für Fortgeschrittene. `https://www.desy.de/~schleper/lehre/TeilchenFortgeschrittene/WS_2012_13/TeilchenFortgeschrittene.pdf`, 2012. [Online; accessed 28-March-2024].

[28] Wikimedia Commons. Standard Model of Elementary Particles. `https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`, 2019. [Online; accessed 23-02-2024].

[29] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.

[30] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[31] E. Noether. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[32] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[33] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[34] Abdus Salam. Gauge Unification of Fundamental Forces. *Science*, 210(4471):723–732, 1980.

[35] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.

[36] 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.

[37] The ATLAS Collaboration. Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2017(12), dec 2017.

[38] The ATLAS Collaboration. Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector. *Journal of High Energy Physics*, 2018(6), June 2018.

[39] A. Quadt. Top quark physics at hadron colliders. *The European Physical Journal C - Particles and Fields*, 48(3):835–1000, Dec 2006.

[40] Frédéric Déliot and Petra Van Mulders. Top quark physics at the LHC. *Comptes Rendus. Physique*, 21(1):45–60, 2020.

[41] R.P. Feynman. *Photon-hadron Interactions*. Advanced Book Classics. Addison-wesley., 1972.

[42] John C. Collins, Davison E. Soper, and George Sterman. Factorization of Hard Processes in QCD, 2004.

[43] Richard Keith Ellis, William James Stirling, and Bryan R Webber. *QCD and collider physics*. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, Cambridge, 2003. Photography by S. Vascotto.

[44] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[45] Michiel Botje, Jon Butterworth, Amanda Cooper-Sarkar, Albert de Roeck, Joel Feltesse, Stefano Forte, Alexander Glazov, Joey Huston, Ronan McNulty, Torbjorn Sjostrand, and Robert Thorne. The PDF4LHC Working Group Interim Recommendations, 2011.

[46] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections. *The European Physical Journal C*, 64(4):653–680, October 2009.

[47] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C.-P. Yuan. CT10 next-to-next-to-leading order global analysis of QCD. *Physical Review D*, 89(3), February 2014.

[48] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244–289, February 2013.

[49] Andrea Giammanco. Single top quark production at the LHC. *Reviews in Physics*, 1:1–12, November 2016.

[50] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[51] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[52] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[53] Andy et al. Buckley. General-purpose event generators for LHC physics. *Physics Reports*, 504(5):145–233, July 2011.

[54] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31–145, 1983.

[55] B.R. Webber. A QCD model for jet fragmentation including soft gluon interference. *Nuclear Physics B*, 238(3):492–528, 1984.

[56] J.-C. Winter, F. Krauss, and G. Soff. A modified cluster-hadronisation model. *The European Physical Journal C - Particles and Fields*, 36(3):381–395, Aug 2004.

[57] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3, 2022.

[58] Manuel et. al Bähr. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, November 2008.

[59] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[60] Esma Mobs. The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018. 2018. General Photo.

[61] The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3(08):S08005, aug 2008.

[62] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08002, aug 2008.

[63] The ATLAS Collaboration. Public ATLAS Luminosity Results for Run-2 of the LHC. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`. Accessed: 12.06.2024.

[64] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. *The European Physical Journal C*, 83(10), October 2023.

[65] G. Avoni et al. The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. *Journal of Instrumentation*, 13(07):P07017, jul 2018.

[66] S. et al. Khalek. The ALFA Roman Pot detectors of ATLAS. *Journal of Instrumentation*, 11(11):P11013–P11013, November 2016.

[67] B. Abbott et al. Production and integration of the ATLAS Insertable B-Layer. *Journal of Instrumentation*, 13(05):T05008, may 2018.

[68] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, and E. Moyse. Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). 3 2007.

[69] The ATLAS Collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *The European Physical Journal C*, 77(10), October 2017.

[70] G Piacquadio, K Prokofiev, and A Wildauer. Primary vertex reconstruction in the ATLAS experiment at LHC. *Journal of Physics: Conference Series*, 119(3):032033, jul 2008.

[71] The ATLAS Collaboration. Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV. 2015.

[72] The ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12):P12006–P12006, December 2019.

[73] The ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data. *The European Physical Journal C*, 77(3), March 2017.

[74] Ben Bruers. Searches for dark matter and single top-quarks with the ATLAS experiment. `https://cds.cern.ch/record/2864659/files/CERN-THESIS-2022-373.pdf?version=1`. Accessed: 01.06.2024.

[75] PLImprovedWorkingPoints. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PLImprovedWPs`. Accessed: 08.06.2024.

[76] The ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 81(7), July 2021.

[77] The ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *The European Physical Journal C*, 77(7), July 2017.

[78] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, April 2008.

[79] The ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *The European Physical Journal C*, 81(8), August 2021.

[80] Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector. Technical report, CERN, Geneva, 2015. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-029.

[81] The ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *The European Physical Journal C*, 76(11), October 2016.

[82] The ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(11), November 2019.

[83] The ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(11), November 2019.

[84] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 78(11), November 2018.

[85] Iqbal Sarker. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2, 03 2021.

[86] A. Peine, C. Lütge, F. Poszler, L. Celi, Oliver Schöffski, G. Marx, and L. Martin. Künstliche Intelligenz und maschinelles Lernen in der intensivmedizinischen Forschung und klinischen Anwendung. *Anästhesiologie und Intensivmedizin*, 61:372–384, 2020. CRIS-Team Scopus Importer:2020-09-18.

[87] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.

[88] Martin Erdmann, Jonas Glombitza, Gregor Kasieczka, and Uwe Klemradt. *Deep Learning for Physics Research*. WORLD SCIENTIFIC, 2021.

[89] Ethem Alpaydin. *Maschinelles Lernen*. De Gruyter Oldenbourg, Berlin, Boston, 2022.

[90] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec 1943.

[91] F. Rosenblatt. The perceptron - A perceiving and recognizing automaton. Technical Report 85-460-1, Cornell Aeronautical Laboratory, Ithaca, New York, January 1957.

[92] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, Dec 1989.

[93] Xavier Glorot, Antoine Bordes, and Y. Bengio. Deep Sparse Rectifier Neural Networks. volume 15, 01 2010.

[94] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[95] The ATLAS Collaboration. ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset. *The European Physical Journal C*, 83(7):681, Jul 2023.

[96] The ATLAS Collaboration. Graph Neural Network Jet Flavour Tagging with the ATLAS Detector. 2022.

[97] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Graph Neural Networks: A Review of Methods and Applications. *CoRR*, abs/1812.08434, 2018.

[98] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407, 1951.

[99] J. Kiefer and J. Wolfowitz. Stochastic Estimation of the Maximum of a Regression Function. *The Annals of Mathematical Statistics*, 23(3):462 – 466, 1952.

[100] Geoffrey Hinton. Coursera Neural Networks for Machine Learning lecture 6. `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`, 2018.

[101] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, 2017.

[102] Inga Döbel, Miriam Leis, Manuel Molina Vogelsang, Dmitry Neustroev, Henning Petzka, Stefan Rüping, Angelika Voss, Martin Wegele, and Juliane Welz. Maschinelles Lernen-Kompetenzen, Anwendungen und Forschungsbedarf. *Fraunhofer IAIS, Fraunhofer IMW, Fraunhofer Zentrale. Zugriff am*, 21:2020, 2018.

[103] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[104] Enrico Guiraud, Axel Naumann, and Danilo Piparo. TDataFrame: functional chains for ROOT data analyses, January 2017.

[105] François Chollet et al. Keras. `https://keras.io`, 2015.

[106] Martín Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Software available from tensorflow.org.

[107] Charles R. Harris et al. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[108] Rene Brun et al. root-project/root: v6.18/02, June 2020.

[109] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

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

[111] CDF and D0 Collaborations. Combination of the top-quark mass measurements from the Tevatron collider. *Phys. Rev. D*, 86:092003, Nov 2012.

[112] The ATLAS Collaboration. The ATLAS Simulation Infrastructure. *The European Physical Journal C*, 70(3):823–874, Dec 2010.

[113] The ATLAS Collaboration. Operation of the ATLAS trigger system in Run 2. *Journal of Instrumentation*, 15(10):P10004–P10004, October 2020.

[114] The ATLAS Collaboration. Performance of the ATLAS muon triggers in Run 2. *Journal of Instrumentation*, 15(09):P09015–P09015, September 2020.

[115] The ATLAS Collaboration. Performance of electron and photon triggers in ATLAS during LHC Run 2. *The European Physical Journal C*, 80(1), January 2020.

[116] The ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *The European Physical Journal C*, 76(11), October 2016.

[117] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann. HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR. *Computer Physics Communications*, 182(4):1034–1046, 2011.

[118] F. Pedregosa et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[119] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), 2016.

[120] Hrushikesh Mhaskar, Qianli Liao, and Tomaso Poggio. Learning Functions: When Is Deep Better Than Shallow, 2016.

[121] Nitish Shirish Keskar and Richard Socher. Improving Generalization Performance by Switching from Adam to SGD, 2017.

[122] Difan Zou, Yuan Cao, Yuanzhi Li, and Quanquan Gu. Understanding the Generalization of Adam in Learning Neural Networks with Proper Regularization, 2021.

[123] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. Regularization for Deep Learning: A Taxonomy, 2017.

[124] P. Baldi, P. Sadowski, and D. Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, 5(1), July 2014.

[125] R. Santos, M. Nguyen, J. Webster, S. Ryu, J. Adelman, S. Chekanov, and J. Zhou. Machine learning techniques in searches for $t\bar{t}h$ in the hß$b\bar{b}$ decay channel. *Journal of Instrumentation*, 12(04):P04014–P04014, April 2017.

[126] Supriya Sinha. Private Communication, 2024. Extraction of top-Yukawa coupling from $t\bar{t}$ cross section in single leptonic final state with $\sqrt{s} = 13$ TeV data from the ATLAS detector. DESY Zeuthen ATLAS Group.

[127] Benjamin Fuks, Kaoru Hagiwara, Kai Ma, and Ya-Juan Zheng. Signatures of toponium formation in LHC run 2 data. *Phys. Rev. D*, 104(3):034023, 2021.

[128] Y. Kiyo, J. H. Kühn, S. Moch, M. Steinhauser, and P. Uwer. Top-quark pair production near threshold at LHC. *The European Physical Journal C*, 60(3):375–386, February 2009.

[129] The CMS Collaboration. Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV. *Physical Review D*, 97(11), June 2018.

[130] The ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment. *Physics Letters B*, 810:135797, 2020.

[131] Michał Czakon, David Heymes, Alexander Mitov, Davide Pagani, Ioannis Tsinikos, and Marco Zaro. Top-pair production at the LHC through NNLO QCD and NLO EW. *Journal of High Energy Physics*, 2017(10), October 2017.

[132] J. H. Kühn, A. Scharf, and P. Uwer. Weak Interactions in Top-Quark Pair Production at Hadron Colliders: An Update. *Physical Review D*, 91(1), January 2015.

[133] G. Cowan. *Statistical data analysis*. Oxford University Press, 1998.

[134] C. Amsler et al. Review of Particle Physics. *Physics Letters B*, 667(1):1–6, 2008. Review of Particle Physics.

[135] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), February 2011.

[136] TRExFitter documentation. `https://trexfitter-docs.web.cern.ch/trexfitter-docs/(2024)`. Accessed: 1.6.2024.

[137] D.Kirkby W. Verkerke. RooFit Users Manual v2.91. `https://root.cern.ch/download/doc/RooFit_Users_Manual_2.91-33.pdf`. Accessed: 05.05.2024.

[138] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. `https://arxiv.org/abs/1103.0354`, 2011.

[139] Kyle Cranmer. Practical Statistics for the LHC. `https://arxiv.org/abs/1503.07622`, 2015.

[140] Yann Coadou. *Boosted Decision Trees*, page 9–58. WORLD SCIENTIFIC, February 2022.

9 Danksagung

Zunächst möchte ich Klaus Mönig für die Möglichkeit danken, meine Masterarbeit in seiner Arbeitsgruppe durchführen zu können. Auch bedanke ich mich für die Unterstützung, die ich von ihm währenddessen erhalten habe. Des Weiteren danke ich Prof. Dr. Heiko Lacker für die Übernahme des Zweitgutachtens.

Weiterhin möchte ich mich bei der gesamten ATLAS-Forschungsgruppe in Zeuthen für die außerordentlich angenehme Atmosphäre und Unterstützung bedanken. In diesem Zusammenhang möchte ich besonders meinen beiden Betreuern Dr. Evgeniya Cheremushkina und Dr. Oliver Majersky danken, die jederzeit ein offenes Ohr für meine Fragen hatten und mir bei der Analyse tatkräftig zur Seite standen. Ein großer Dank gilt auch Dr. Thorsten Kuhl für das Korrekturlesen und die vielen Anregungen während des Schreibens der Arbeit.

Besonderer Dank gebührt meinen Eltern und meinem Bruder, die mir kontinuierlich innerhalb und außerhalb des Studiums den Rücken stützen und mir das erfolgreiche Abschließen des Studiums überhaupt ermöglicht haben. Auch danke ich ihnen für die finale Korrektur meiner großzügigen Auslegung der deutschen Sprachregeln. Ich danke meinen Freunden dafür, dass sie stets an meiner Seite standen und dass ich auch in Zukunft auf sie bauen kann. Zu guter Letzt bedanke ich mich bei meiner Freundin Maja für die wunderschöne gemeinsame Zeit seit unserem Kennenlernen und die immerwährende Unterstützung in guten sowie stressigen Zeiten.