

Studien zur Verbesserung der Sensitivität für Tests der CP-Invarianz in der Higgs-Boson-Produktion via VBF am ATLAS-Experiment durch Verwendung neuronaler Netzwerke zur Rekonstruktion des Higgs-Boson-Viererimpulses

Alexandra Spitzer



BACHELORARBEIT

**Studien zur Verbesserung der
Sensitivität für Tests der CP-Invarianz in
der Higgs-Boson-Produktion via VBF
am ATLAS-Experiment durch
Verwendung neuronaler Netzwerke zur
Rekonstruktion des
Higgs-Boson-Viererimpulses**

vorgelegt von
Alexandra Spitzer

Betreuer
Prof. Dr. Markus Schumacher

14. Januar 2022



Albert-Ludwigs-Universität Freiburg

Datum: 14. Januar 2022
Autor: Alexandra Spitzer
Physikalisches Institut
Albert-Ludwigs-Universität Freiburg

Erklärung

Hiermit versichere ich, die eingereichte Bachelorarbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt zu haben. Wörtlich oder inhaltlich verwendete Quellen wurden entsprechend den anerkannten Regeln wissenschaftlichen Arbeitens (lege artis) zitiert. Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht anderweitig als Bachelorarbeit eingereicht wurde.

Ort, Datum: Freiburg, 14. Januar 2022

Unterschrift:

Zusammenfassung

In der vorliegenden Arbeit wird eine CP-Studie vorgestellt, die neuronale Netzwerke einsetzt, um den Viererimpuls des Higgs-Bosons, produziert in der Vektorbosonfusion, im $H \rightarrow \tau_e \tau_\mu$ Zerfall zu rekonstruieren und damit die Sensitivität für CP-Tests zu verbessern. Zur Untersuchung der Kopplungsstruktur des Higgs-Teilchens an elektroschwache Vektorbosonen werden simulierte Ereignisse von Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV verwendet und der Mittelwert der CP-ungeraden Optimalen Observable ermittelt. Für die Berechnung der Optimalen Observable wird der Viererimpuls des Higgs-Bosons benötigt. Die Auflösung dieser ist daher durch die Rekonstruktionsgüte des Higgs-Boson-Viererimpulses begrenzt. Zur Rekonstruktion der Viererimpulse werden verschiedene Regressionsnetzwerke mit optimierten Hyperparametern trainiert. Die sich daraus ergebenden Viererimpulse werden mit den Viererimpulsen, die mit der Methode des *Missing Mass Calculators* bestimmt wurden, verglichen. In der Auflösung der Komponenten des Higgs-Boson-Viererimpulses konnten signifikante Verbesserungen erreicht werden. Auch in der Auflösung der Optimalen Observable konnte eine Verbesserung von 11% erreicht werden. Die Sensitivität der Optimalen Observablen für CP-Verletzung konnte jedoch nicht deutlich verbessert werden.

Abstract

A CP study using feedforward neural networks to reconstruct the four-momentum of the Higgs boson produced via vector-boson fusion with the $H \rightarrow \tau_e \tau_\mu$ decay channel is presented. In order to study the coupling structure of the Higgs particle to electroweak gauge bosons, simulated events of proton proton collisions at a center of mass energy of $\sqrt{s} = 13$ TeV are used and the mean value of the CP-odd Optimal Observable is investigated. For the calculation of the Optimal Observable the Higgs boson's four-momentum is required. The resolution of the Optimal Observable is therefore limited by the reconstruction quality of the Higgs boson's four-momentum. Various regression networks are trained with optimized hyperparameters to reconstruct the four-momentum of the Higgs boson. The results are compared with the four-momenta determined by the method of the Missing Mass Calculator. Significant improvements could be achieved concerning the resolution of the components of the Higgs boson's four-momentum. Also, an improvement of 11% was achieved in the resolution of the Optimal Observable. However, the sensitivity of the Optimal Observables to CP violation could not be increased significantly.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 1 |
| 2 | Theoretische Grundlagen | 3 |
| 2.1 | Standardmodell der Teilchenphysik | 3 |
| 2.2 | Brout-Englert-Higgs-Mechanismus | 6 |
| 2.3 | Das Higgs-Boson | 8 |
| 2.3.1 | Produktion und Zerfall des Higgs-Bosons | 8 |
| 2.4 | Diskrete Symmetrien | 10 |
| 2.4.1 | Parität | 10 |
| 2.4.2 | Ladungskonjugation | 11 |
| 2.4.3 | CP-Symmetrie | 11 |
| 2.5 | Parametrisierung von CP-Verletzung und CP-sensitive Observablen | 11 |
| 3 | Das ATLAS Experiment | 14 |
| 3.1 | Der ATLAS Detektor | 14 |
| 3.1.1 | Koordinatensystem | 14 |
| 3.1.2 | Detektor-Komponenten | 15 |
| 4 | Neuronale Netze | 19 |
| 4.1 | Allgemeiner Aufbau | 19 |
| 4.2 | Trainingsprozess | 21 |
| 4.3 | Minimierung der Verlustfunktion mit Adam-Optimierer | 23 |
| 4.4 | Early-Stopping-Algorithmus | 23 |
| 4.5 | Anpassung der Hyperparameter mit OPTUNA | 24 |
| 5 | Der analysierte Datensatz | 25 |
| 5.1 | Merkmale des Datensatzes | 25 |
| 5.2 | Skalierung der Ereignisse | 27 |
| 6 | Selektion der Signalereignisse | 28 |
| 6.1 | Objektselektion | 28 |
| 6.2 | Vorselektion | 28 |
| 6.3 | Finale Selektion | 30 |
| 7 | Rekonstruktion des Higgs-Boson-Viererimpulses mittels Regressionsnetzwerken | 32 |
| 7.1 | Vergleich von vier verschiedenen neuronalen Netzen | 32 |
| 7.1.1 | Neuronales Netz 1 - kartesische Koordinaten | 32 |
| 7.1.2 | Neuronales Netz 2 - kartesische Koordinaten mit Jets | 33 |
| 7.1.3 | Neuronales Netz 3 - Polarkoordinaten | 34 |
| 7.1.4 | Neuronales Netz 4 - Polarkoordinaten mit Jets | 35 |
| 7.1.5 | Ergebnisse der OPTUNA-Optimierungen | 36 |
| 7.1.6 | Gemeinsamkeiten der vier neuronalen Netze im Trainings- und Vorhersageprozess | 39 |
| 7.1.7 | Vergleich der Rekonstruktionsgüte für den Higgs-Boson-Viererimpuls | 41 |
| 7.2 | Berechnung der Optimalen Observable | 45 |

| | | |
|----------|--|-----------|
| 7.2.1 | Beiträge der verschiedenen Objekte zur Auflösung der Optimalen Observable | 45 |
| 7.2.2 | Auflösungsvermögen der Optimalen Observable für die Rekonstruktion mit neuronalen Netzen | 47 |
| 8 | Bestimmung der Sensitivität von \tilde{d} | 49 |
| 8.1 | Simulation von Verteilungen der Optimalen Observable für $\tilde{d} \neq 0$ | 49 |
| 8.2 | Abhängigkeit der Optimalen Observablen von \tilde{d} | 49 |
| 8.2.1 | Eichkurven | 51 |
| 9 | Zusammenfassung | 57 |
| A | Visualisierungsabbildungen OPTUNA | 59 |
| A.1 | Neuronales Netz 2 | 59 |
| A.2 | Neuronales Netz 3 | 60 |
| A.3 | Neuronales Netz 4 | 61 |
| B | Verteilungen der Eingangsgrößen | 62 |
| B.1 | Neuronales Netz 1 | 62 |
| B.2 | Neuronales Netz 2 | 64 |
| B.3 | Neuronales Netz 3 | 66 |
| B.4 | Neuronales Netz 4 | 67 |
| C | Abbildungen zum Auflösungsvermögen des Higgs-Boson-Viererimpulses | 70 |
| C.1 | Neuronales Netz 1 | 70 |
| C.2 | Neuronales Netz 2 | 71 |
| C.3 | Neuronales Netz 3 | 73 |
| C.4 | Neuronales Netz 4 | 74 |
| | Literatur | 75 |
| | Danksagung | 80 |

1 Einleitung

Woraus besteht Materie und wodurch wird sie zusammen gehalten? Diese Frage stellen sich Forscher schon seit vielen Jahren. Heute fasst das sogenannte „Standardmodell der Elementarteilchen“ den jetzigen Wissensstand aller Elementarteilchen zusammen. Das Standardmodell beinhaltet die sogenannten Fermionen, die über die Grundkräfte durch den Austausch sogenannter Eichbosonen, wechselwirken [1] [2] [3]. Um in das Standardmodell Massenterme für die Fermionen und Eichbosonen einführen zu können ohne dabei die Forderung nach lokaler Eichinvarianz der Theorien zur Beschreibung der fundamentalen Wechselwirkungen zu verletzen, entwickelten Peter Higgs [4] [5], François Englert und Robert Brout [6], sowie Gerald Guralnik, Carl R. Hagen and T. W. B. Kibble [7] 1964 den Brout-Englert-Higgs-Mechanismus. Voraussetzung hierfür ist ein skalares Feld mit einem Potential, durch dessen nichtverschwindenden Vakuumerwartungswert Massenterme in das Standardmodell eingeführt werden können. Die vorhergesagte Existenz eines neuen Bosons durch die Anregung dieses skalaren Feldes, konnte durch den Nachweis des Higgs-Bosons im Sommer 2012 am *Large Hadron Collider* (LHC) bestätigt werden [8] [9]. Anhand von Daten aus Proton-Proton-Kollisionen, die am LHC aufgenommen werden, werden die Eigenschaften des Higgs-Bosons, wie zum Beispiel dessen CP-Natur, intensiv untersucht. Da die Verletzung der CP-Invarianz eine der drei Sacharow-Bedingungen ist, die zur Erklärung der beobachteten Baryonenasymmetrie in unserem Universum erforderlich sind, wird im Higgs-Sektor nach neuen Quellen von CP-Verletzung gesucht. In dieser Arbeit werden hierfür unter anderem Methoden des maschinellen Lernens bzw. neuronale Netze verwendet. Die Bedeutung von diesen hat in den letzten Jahren mit der zunehmenden Verfügbarkeit von Computerressourcen stark zugenommen. Mit neuronalen Netzen können komplexe und nichtlineare Funktionen modelliert werden [10].

Das Ziel dieser Bachelorarbeit ist es, die Rekonstruktion des Viererimpulses des Higgs-Bosons im $H \rightarrow \tau_e \tau_\mu$ Zerfallskanal unter der Annahme von einer integrierten Luminosität $\mathcal{L} = 139 \text{ fb}^{-1}$ in pp-Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13 \text{ TeV}$ zu verbessern. Hierfür werden vier neuronale Netze mit unterschiedlichen Sätzen von Eingangsvariablen und Zielvariablen festgelegt. Zwar bestehen die Zielvariablen stets aus den Komponenten des Higgs-Bosons auf Wahrheitsniveau, jedoch werden diese bei den ersten neuronalen Netzen in kartesischen und bei den anderen beiden Netzen in Polarkoordinaten definiert. Nach einer Hyperparameter-Optimierung aller Netze mittels OPTUNA [11], werden alle Netze trainiert und für die Vorhersage des Higgs-Boson Viererimpulses verwendet. Die Qualität der Rekonstruktion wird mittels Auflösungsverteilungen bewertet und mit der Leistung der aktuell verwendeten Rekonstruktionsmethode - der sogenannte *Missing Mass Calculator* (MMC) [12] - verglichen. Zusätzlich wird zum Vergleich auch die Methode der Vektoraddition der rekonstruierten Transversalimpulse der sichtbaren und unsichtbaren Zerfallsprodukte des Higgs-Bosons herangezogen. Diese Methode kann jedoch nur für die Rekonstruktion des Transversalimpulses und der ϕ -Komponente des Higgs-Bosons genutzt werden. Die mit den neuronalen Netzen vorhergesagten Viererimpulse des Higgs-Bosons werden dann verwendet, um die CP-ungerade Optimalen Observablen der vier Netze zu berechnen. Die Auflösung dieser wird anschließend mit der des MMCs verglichen. Zur Analyse der Sensitivität von CP-Tests werden sogenannte Eichkurven herangezogen, die den Mittelwert der Optimalen Observable in Abhängigkeit von \tilde{d} darstellen. Aus diesen Eichkurven kann dann auf *Tilded* geschlossen werden.

Die vorliegende Arbeit ist wie folgt aufgebaut: Das folgende Kapitel gibt einen Überblick

über die theoretischen Grundlagen. Hier wird unter anderem in die Theorie des Standardmodells, des Brout-Englert-Higgs-Mechanismus und der CP-sensitiven Observablen eingeführt. In Kapitel 3 wird der ATLAS-Detektor mit seinem Koordinatensystem und Subdetektoren vorgestellt. Kapitel 4 befasst sich mit der Theorie der in dieser Arbeit verwendeten neuronalen Netze, wobei hier auch auf die Optimierungs-Software OPTUNA eingegangen wird. Anschließend erfolgt in Kapitel 5 eine Beschreibung der Merkmale des für diese Arbeit verwendeten Datensatzes, während in Abschnitt 6 das Vorgehen zur Skalierung der Ereignisse auf $\mathcal{L} = 139 \text{ fb}^{-1}$ beschrieben wird. Im siebten Kapitel werden die Selektionskriterien, die auf den Datensatz angewendet wurden, vorgestellt. In Kapitel 8 folgen die Studien zu den neuronalen Netzen, wobei hier die Ergebnisse von vier Netzen miteinander verglichen und zur Berechnung der Optimalen Observable verwendet werden. Daraufhin erfolgt in Kapitel 9 die Bestimmung der Sensitivität des CP-Tests durch Berechnung der erwarteten Unsicherheit auf den Parameter \tilde{d} , der die Stärke der CP-Verletzung im HVV -Vertex beschreibt. Schließlich findet sich in Kapitel 10 eine Zusammenfassung der Arbeit.

2 Theoretische Grundlagen

Dieses Kapitel soll einen Überblick über die Theorie bezüglich der in dieser Arbeit zugrundeliegenden Physik geben. Zuerst wird das Standardmodell der Teilchenphysik anschaulich erläutert. In dem nächsten Unterkapitel wird dann der Brout-Englert-Higgs-Mechanismus beleuchtet. Auch das Higgs-Boson und seine Produktionsprozesse am LHC und Zerfallskanäle werden im Rahmen dieses Kapitel vorgestellt. Außerdem erfolgt eine kurze Einführung in die diskreten Symmetrien der Paritäts (\mathcal{P})- und Ladungskonjugation (\mathcal{C}). Der letzte Abschnitt befasst sich schließlich mit der CP-sensitiven Optimalen Observable und wie diese genutzt werden kann, um CP-verletzende des Higgs-Bosons an elektroschwache Vektorbosonen zu untersuchen.

2.1 Standardmodell der Teilchenphysik

Das Standardmodell [13] (SM) der Teilchenphysik bildet das Grundgerüst unseres heutigen Verständnis der Struktur der Materie. Auf Basis der vereinheitlichten Theorie der elektroschwachen Wechselwirkung und der Quantenchromodynamik werden im SM drei der vier fundamentalen Wechselwirkungen mit sehr hoher Präzision beschrieben: die elektromagnetische, die schwache und die starke Wechselwirkung. Lediglich die Gravitation ist nicht im Standardmodell enthalten. Diese ist jedoch im Vergleich zu den drei anderen Wechselwirkungen im Rahmen der Hochenergiepartikelphysik vernachlässigbar klein.

Abbildung 1 zeigt eine schematische Übersicht des Standardmodells. Es besteht aus 12 Fermionen. Das sind Elementarteilchen mit einem Spin von $s = \frac{1}{2}$, zu denen die Quarks und die Leptonen gehören. Quarks tragen eine Farbladung und unterliegen der starken, der schwachen, sowie der elektromagnetischen Wechselwirkung, während die Leptonen der schwachen Wechselwirkung und auch der elektromagnetischen Wechselwirkung, sofern sie elektrisch geladen sind, unterliegen. Jedes Fermion hat ein zugehöriges Antiteilchen mit invertierten additiven Quantenzahlen.

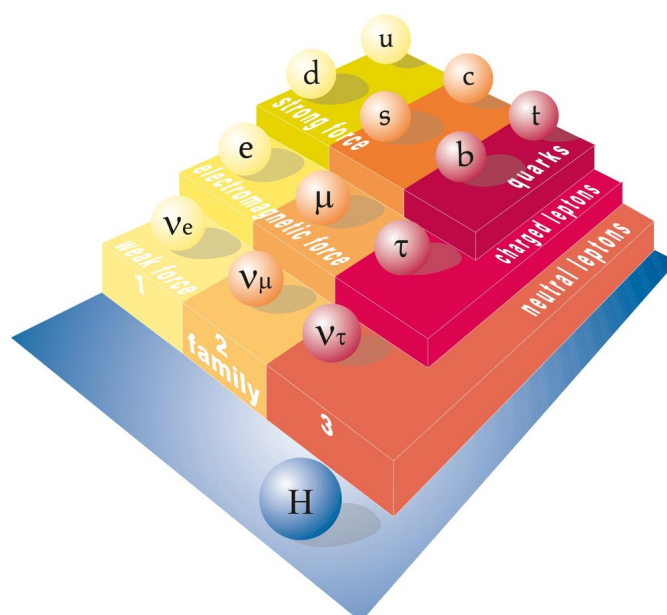


Abbildung 1: Standardmodell der Teilchenphysik [14]; Es besteht aus Leptonen, Quarks, Eichbosonen und dem Higgs-Boson.

Die drei Familien der Leptonen und Quarks

Die Quarks werden zusammen mit den Leptonen in drei 'Familien' eingeteilt. Die Leptonen der ersten Familie umfassen das Elektron e und das zugehörige Elektron-Neutrino ν_e , während das schwerere Myon μ und das Myon-Neutrino ν_μ die Leptonen der zweiten Familie bilden. Letzlich gehört zu der dritten Familie das Tau-Lepton τ und das Tau-Neutrino ν_τ . Das Elektron, Myon und Tau-Lepton tragen eine elektrische Ladung von $Q = -1e$, wohingegen die Neutrinos elektrisch ungeladen ($Q = 0$) sind und im SM als masselos angenommen werden.

Die Quarks der ersten Familie sind das Up-Quark u mit einer elektrischen Ladung von $Q = \frac{2}{3}e$ und das Down-Quark d mit $Q = -\frac{1}{3}e$. Das Charm-Quark c trägt wie das Up-Quark eine elektrische Ladung von $Q = \frac{2}{3}e$ und bildet zusammen mit dem Strange-Quark s , welches mit $Q = -\frac{1}{3}e$ elektrisch geladen ist, die Quarks der zweiten Familie. Zu der dritten Familie gehört das Top-Quark t mit $Q = \frac{2}{3}e$ und das Bottom-Quark b mit $Q = -\frac{1}{3}e$.

Quarks kommen nur in Bindungszuständen, den sogenannten Hadronen, vor. Zu diesen gehören die Mesonen, die aus einem Quark-Antiquark-Paar gebildet werden, und die Baryonen, welche sich aus drei Quarks zusammensetzen. Der Spin der einzelnen Quarks koppelt im Falle der Mesonen zu einem ganzzahligen Gesamtspin, während Baryonen einen halbzahligen Gesamtspin haben [13].

Eichbosonen

Zu den Elementarteilchen gehören außerdem die Vektorbosonen, auch Eichbosonen genannt. Sie haben im Gegensatz zu den Fermionen einen ganzzahligen Spin. Alle weiteren wichtigen Eigenschaften der Vektorbosonen sind in Tabelle 1 aufgeführt. Die Existenz der Eichbosonen basiert auf dem Eichprinzip, das besagt, dass die Invarianz der Dirac-Gleichung gegenüber lokalen Phasentransformationen der Wellenfunktion nur gegeben ist, wenn äußere Felder existieren, an welche die geladenen Teilchen koppeln. Diese externen Felder entsprechen im Rahmen der elektromagnetischen Wechselwirkung dem Photon A , im Rahmen der schwachen Wechselwirkung den W^\pm -Bosonen und dem Z^0 -Boson, sowie im Rahmen der starken Wechselwirkung den acht Gluonen g [15]. Über die Eichbosonen werden also die fundamentalen Kräfte vermittelt. Letztere sollen im folgenden Abschnitt genauer betrachtet werden.

| Eichboson | Wechselwirkung | Masse M [GeV] | Ladung Q [e] |
|------------|--------------------|--------------------|----------------|
| Photon A | elektromagnetische | 0 | 0 |
| W^\pm | schwache | $80,379 \pm 0,012$ | ± 1 |
| Z^0 | schwache | $91,188 \pm 0,002$ | 0 |
| Gluon g | starke | 0 | 0 |

Tabelle 1: Die vier Eichbosonen und die drei fundamentalen Kräfte, die sie vermitteln. Außerdem werden die Masse M und die elektromagnetische Ladung Q des jeweiligen Bosons angegeben [16].

Wechselwirkungen

Die Wechselwirkungen des Standardmodells lassen sich durch Quantenfeldtheorien beschreiben. Um die verschiedenen Kraftfelder zu beschreiben, werden die entsprechenden Lagrange-Dichten herangezogen, die ihr klassisches Analogon in der Lagrange-Funktion $L(q_i, \dot{q}_i)$ finden. Während diese in der klassischen Mechanik von den verallgemeinerten Koordinaten q_i und Geschwindigkeiten \dot{q}_i abhängen, wird die Lagrange-Dichte $\mathcal{L}(\phi(x), \partial_\mu \phi(x))$ in Abhängigkeit des Feldes $\phi(x)$ und seines Vierergradienten $\partial_\mu \phi(x)$ definiert. Zu den Lagrange-Gleichungen der Feldtheorie gelangt man, indem man $\delta S = 0$ fordert, wobei

$$S = \int_{t_1}^{t_2} dt \int d^3x \mathcal{L}(\phi(x), \partial_\mu \phi(x)) \quad (1)$$

die Wirkung beschreibt [15].

Die Theorie der Quantenelektrodynamik (kurz: QED) beschreibt die elektromagnetische Wechselwirkung zwischen Elektronen, Positronen oder anderen, elektrisch geladenen Teilchen und Photonen. Die zugehörige U(1)-Symmetriegruppe ergibt sich aus der Tatsache, dass man die absolute Phase der Wellenfunktion z.B. eines Elektrons nicht messen kann. Eine Konsequenz der Invarianz unter der globalen U(1)-Transformation ist im Fall der QED, dass die elektrische Ladung erhalten ist. Aus der Forderung nach lokaler Eichinvarianz folgt die Kopplung der elektrischen Ladung an das Photon. Aus dieser Symmetrie resultiert außerdem die Masselosigkeit des Photons [17].

Um hingegen die starke Wechselwirkung zwischen Quarks und Gluonen zu beschreiben, wird die Quantenchromodynamik herangezogen. Diese kann mittels einer SU(3)-Gruppe beschrieben werden, deren globale Symmetrie eine neue Erhaltungsgröße liefert: die Farbladung C . Jedes Quark trägt demnach eine Farbladung, die entweder rot (r), grün (g) oder blau (b) ist. Die Wechselwirkung bleibt ungeändert bei Vertauschung der Farben. Die Felder, die bei lokalen SU(3)_C-Transformationen eingeführt werden, koppeln an die Farbladung der Quarks, tragen selber eine Farbladung und können die Farbe der Quarks ändern. Die acht dazugehörigen, masselosen Feldquanten heißen Gluonen [15], [17], [18].

Die schwache Wechselwirkung tritt zwischen Leptonen, Antileptonen, Quarks, sowie Antiquarks auf und wird im Rahmen der elektroschwachen Vereinigung zusammen mit der elektromagnetischen Wechselwirkung behandelt. Wie bei den anderen Wechselwirkungen, wählt man auch hier eine Eichgruppe, nämlich die U(1)_Y × SU(2)_{I_W}-Gruppe, so dass die Lagrange-Dichte lokal invariant bleibt. I_W bezieht sich hier auf den schwachen Isospin, an welchen die schwache Wechselwirkung koppelt, während Y die Hyperladung beschreibt. Diese ersetzt die elektrische Ladung Q aus der QED und erfüllt zusammen mit der dritten Komponente des Isospins I_{W₃} die Gell-Mann-Nishijama-Formel $Q = I_{W_3} + \frac{Y}{2}$. Es ist sinnvoll, den schwachen Isospin als neue Quantenzahl einzuführen, da man die Teilchen, die durch Emission schwacher Eichbosonen ineinander übergehen können, in Multipletts eines schwachen Isospins zusammenzufassen kann. Die linkshändigen Fermionen mit Isospin $I = \frac{1}{2}$ koppeln an die W-Feldquanten, während man bei den rechtshändigen, geladenen Leptonen und Quarks, die Singulets mit Isospin $I = 0$ bilden, keine Kopplung beobachtet. Das Z⁰-Boson hingegen koppelt an beide Multipletts.

Das Problem der elektroschwachen Vereinheitlichung ist, dass die Forderung nach Invarianz der Lagrange-Dichte unter lokalen Eichtransformationen masselose Feldquanten

erfordert. Jedoch trifft dies nur auf das Photon A zu, wohingegen die Eichbosonen der schwachen Wechselwirkung vergleichsweise hohe Massen haben, wie es in Tabelle 1 aufgeführt ist. Dieses Problem kann gelöst werden, indem man ein zusätzliches Skalarfeld über den Brout-Englert-Higgs-Mechanismus einführt [15], [17], [18], [13]. Dieser wird im folgenden Kapitel näher erläutert.

2.2 Brout-Englert-Higgs-Mechanismus

Um Massenterme in die elektroschwache Wechselwirkung einzuführen, ohne die Invarianz unter lokaler $U(1)_Y \times SU(2)_{I_W}$ -Eichtransformation zu verletzen, ist es notwendig ein skalares Feld ϕ mit schwacher Hyperladung $Y = 1$ in Form eines komplexen Dupletts mit Isospin $I_W = \frac{1}{2}$ zu definieren [19]:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (2)$$

Hierbei beziehen sich die Indizes $+$ und 0 auf die elektromagnetische Ladung in Einheiten der Elementarladung e . Die Lagrange-Dichte dieses Feldes ([18], [20]), unter Abwesenheit anderer Felder, lässt sich schreiben als

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \quad (3)$$

wobei $D_\mu \phi$ die kovariante Ableitung beschreibt, die hier eingeführt wird, um eine global invariante in eine lokal invariante Lagrange-Dichte umzuwandeln. Der Ausgangspunkt des Higgs-Mechanismus ist das in Gleichung 3 vorkommende Higgs-Potential $V(\phi)$. Der erste Term dieses Potentials

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad \lambda > 0 \quad (4)$$

wird auch als Massenterm mit dem Massenparameter μ bezeichnet, während der zweite Term den (Selbst)kopplungsparameter λ des Higgsfeldes beschreibt.

Das Minimum bzw. der quantenmechanische Grundzustand von $V(\phi)$ lässt sich mit der üblichen Methode berechnen und liefert eine triviale Lösung bei $\phi_{\min,1} = 0$ für $\mu^2 < 0$ und eine nichttriviale Lösung bei $\phi_{\min,2} = \sqrt{\frac{\mu^2}{2\lambda}}$, wobei $\mu^2 > 0$. Dies liefert einen Kreis entarteter Minima mit Radius $\sqrt{\frac{\mu^2}{2\lambda}}$. Zur Veranschaulichung des Higgspotentials und seinen Minima sei auf Abbildung 2 verwiesen.

Eine gebräuchliche Wahl des Minimums [18] lautet:

$$\phi_{\min} = \begin{pmatrix} 0 \\ \sqrt{\frac{\mu^2}{2\lambda}} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad (5)$$

Hierbei ist v der Vakuumerwartungswert, welcher spontan aus dem Kreis der entarteten Minima mit Radius $\frac{\mu^2}{2\lambda}$ ausgewählt wird. Durch die Festlegung auf einen Grundzustand wird die Symmetrie der $U(1)_Y \times SU(2)_{I_W}$ -Gruppe spontan zur $U(1)_Q$ gebrochen.

Da Berechnungen in der Quantenfeldtheorie keine exakten Lösungen liefern, wird oft eine Reihenentwicklung um das Minimum durchgeführt. Folglich wird die unitäre Eichung

$$\phi_{\text{uni}} = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \quad (6)$$

betrachtet. Hierbei ist h das Higgs-Boson, das Fluktuationen um den Vakuumerwartungswert v beschreibt. Setzt man dieses ϕ_{uni} in die Lagrange-Dichte ein, so erhält man nach einigen Umformungen und Reparametrisierungen, die hier nicht weiter betrachtet werden sollen, Massen [21] für die geladenen W -Bosonen:

$$m_{W^\pm} = \frac{gv}{\sqrt{2}} \quad (7)$$

Die Massen für das Photon A und das Z^0 -Boson [21] ergeben sich aus zwei Linearkombinationen in der Lagrange-Dichte, die mit dem schwachen Mischungswinkel [2] θ_W parametrisiert werden:

$$\theta_W = \arctan \frac{g'}{g} \quad (8)$$

Die Stärke der Kopplung an den schwachen Isospin I_W wird hierbei durch die Kopplungskonstante g beschrieben, wohingegen g' die Kopplung an die schwache Hyperladung Y repräsentiert. Das Photon bleibt masselos, während das neutrale Z^0 -Boson eine Masse von

$$m_{Z^0} = \frac{m_{W^\pm}}{\cos \theta_W} \quad (9)$$

erhält. Die Masse des Higgs-Bosons ist in der Theorie nicht vorhersagbar, und muss folglich experimentell bestimmt werden. Der Vakuumerwartungswert [18] des Higgs-Potentials v hingegen lässt sich über

$$v = 246 \text{ GeV} = (\sqrt{2}G_F)^{-\frac{1}{2}} \quad (10)$$

berechnen, wobei G_F die Fermi-Konstante bezeichnet. Auch die Massen der Fermionen m_f [21] lassen sich durch spontane Symmetriebrechung generieren, indem man der Lagrange-Dichte einen Kopplungsterm mit einer neuen Kopplungsstärke λ_f hinzufügt:

$$m_f = \frac{\lambda_f v}{\sqrt{2}} \quad (11)$$

λ_f wird als Yukawa-Kopplung bezeichnet, welche die Kopplung des Higgs-Feldes an Fermionen beschreibt. Die Stärke dieser Kopplung ist proportional zur Masse der Fermionen.

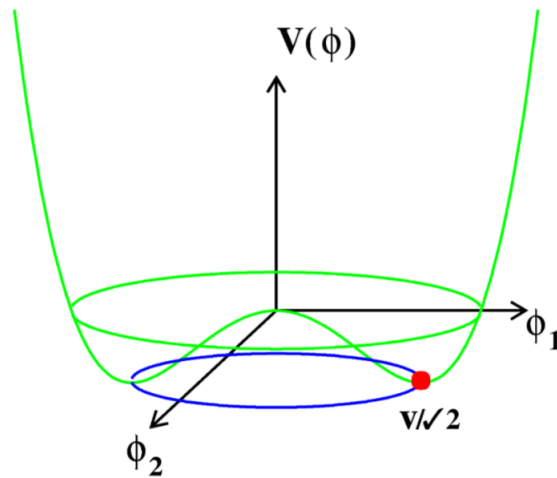


Abbildung 2: Das Higgs-Potential in Abhängigkeit von zwei Feldkomponenten ϕ_1 und ϕ_2 . Bei $\phi_1 = \phi_2 = 0$ existiert ein nicht stabiler Zustand. Der energetisch günstigste Zustand liegt bei einem Wert ungleich Null [22].

2.3 Das Higgs-Boson

Der Brout-Englert-Higgs-Mechanismus setzt die Existenz des Higgs-Bosons im Standardmodell voraus. Dieses entspricht der elementaren Anregung des Higgs-Feldes, ist elektrisch ungeladen und ist das einzige Elementarteilchen mit Spin 0. Außerdem ist es CP-gerade: $J^{CP} = 0^+$. Da das Higgs-Teilchen nach einer sehr kurzen Zeit zerfällt, ist es nicht direkt im Detektor nachweisbar und muss folglich über seine Zerfallsprodukte rekonstruiert werden. Der experimentelle Nachweis des Higgs-Bosons und die Bestimmung seiner Masse von $M_H = (126,0 \pm 0,4(stat.) \pm 0,4(sys.))$ GeV gelang im Juli 2012 am LHC mit Daten bei einer integrierten Luminosität von etwa $4,8 \text{ fb}^{-1}$ bei einer Schwerpunktsenergie von $\sqrt{s} = 7$ TeV, sowie $5,8 \text{ fb}^{-1}$ bei $\sqrt{s} = 8$ TeV [23]. Der aktuell beste Wert der Higgs-Masse liegt bei $M_H = (125,25 \pm 0,17)$ GeV [16] [24].

2.3.1 Produktion und Zerfall des Higgs-Bosons

Produktionsprozesse

Die Higgs-Boson-Produktion in Proton-Proton-Kollisionen wird dominiert von der Gluonfusion, welche links in Abbildung 3 zu sehen ist. Hierbei fällt auf, dass die Gluonen nicht direkt an das Higgs-Boson koppeln, sondern indirekt über eine Top- oder Bottom-Quark-Schleife. Obwohl dies ein Prozess höherer Ordnung ist, kommt die Gluonfusion am häufigsten vor. Das liegt an der vergleichsweise starken Kopplung des Top-, und auch des Bottom-Quarks an das Higgs-Boson aufgrund ihrer hohen Massen [13].

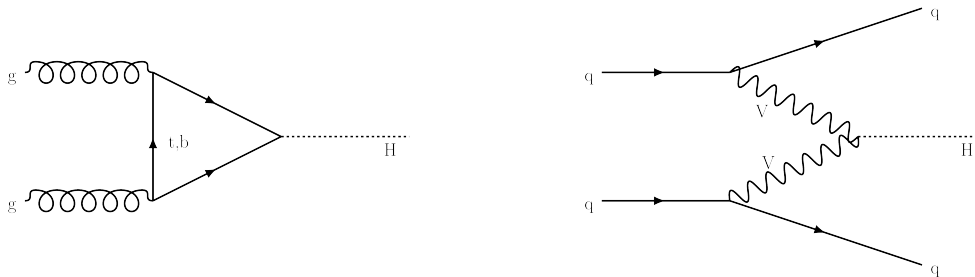


Abbildung 3: Die zwei häufigsten Produktionsprozesse des Higgs-Boson in Proton-Proton-Kollisionen. *Links*: Feynman-Diagramme der Gluonfusion über eine Quarkschleife. *Rechts*: Feynman-Diagramm der Vektorbosonfusion, wobei V für die Vektorbosonen W^\pm und Z^0 stehen.

Ein Feynman-Diagramm des zweit häufigsten Produktionsprozess ist rechts in Abbildung 3 dargestellt. Da hier die von den Quarks q (oder Antiquarks) abgestrahlten Vektorbosonen W^+ und W^- oder zwei Z^0 in ein Higgs-Teilchen fusionieren, wird dieser Prozess Vektorbosonfusion (VBF) genannt [13]. Wie in Abbildung 4 zu sehen ist, ist der Wirkungsquerschnitt σ des Vektorbosonfusions-Prozesses um eine etwa Größenordnung kleiner als der der Gluon-Gluon-Fusion.

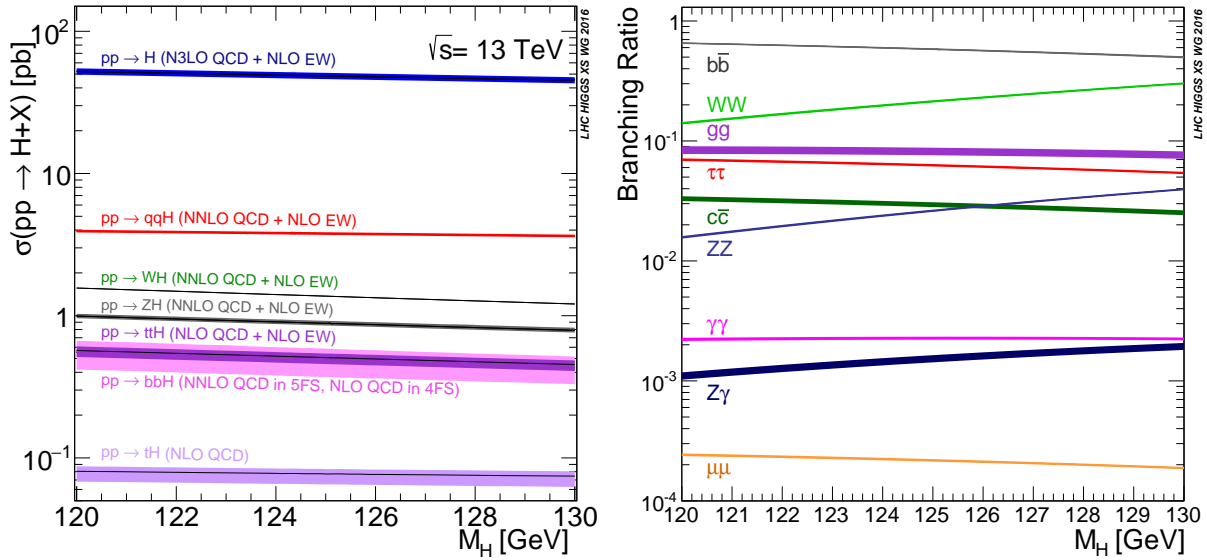


Abbildung 4: *Links*: Wirkungsquerschnitte σ für die Produktion von Higgs-Bosonen bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV in einem Massenbereich des Higgs-Teilchens von 120 bis 130 GeV. Der dominante Gluon-Gluon-Fusionsprozess ($pp \rightarrow H$) ist im Vergleich zum VBF-Prozess ($pp \rightarrow qqH$) um etwa einen Faktor 10 größer; *Rechts*: Verzweigungsverhältnisse des Higgs-Bosons in Abhängigkeit von seiner Masse M_H ebenfalls im Bereich von 120 bis 130 GeV für die wichtigsten Zerfallsprozesse [25].

Zerfallsprozesse

Die Zerfallsprodukte des Higgs-Bosons sind Fermionen- oder Bosonen-Paare. Ein Zerfall, bei dem masselose Teilchen beteiligt sind, wie z.B. bei $H \rightarrow gg$, ist über Schleifen von z.B. Top-Quarks möglich. Die Verzweigungsverhältnisse der häufigsten Zerfallsprozesse in Abhängigkeit der Higgs-Boson-Masse ist in Abbildung 4 dargestellt. Der Higgs-Zerfall in ein Bottom- und Antibottom-Quark mit einem Verzweigungsverhältnis und relativer Unsicherheit von $5,82 \times 10^{-1} \pm_{1,2\%}^{1,3\%}$ bei einer Higgs-Masse von $M_H = 125$ GeV [26] ist der dominante Prozess.

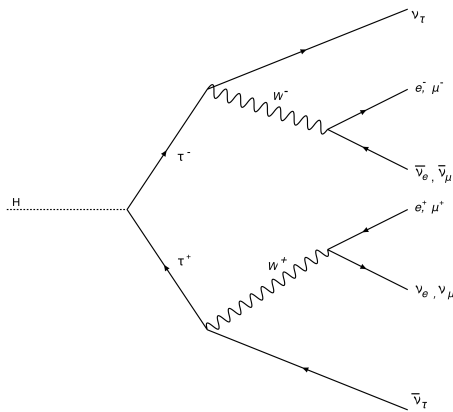


Abbildung 5: Feynman-Diagramm des leptonen Zerfalls $H \rightarrow \tau\tau$

Der in dieser Arbeit analysierte Zerfall in zwei τ -Leptonen ist mit einem Verzweigungsverhältnis und relativer Unsicherheit von $6,27 \times 10^{-2} \pm 1,6\%$ [26] weniger wahrscheinlich. Dennoch ist dies der häufigste leptonen Zerfall. Wie das Higgs-Boson, ist auch das τ -Lepton instabil und zerfällt mit einer mittleren Lebensdauer von $(290,3 \pm 0,5) \times 10^{-25}$ s [16] und kann somit nur über seine Zerfallsprodukte nachgewiesen werden. Ein τ -Lepton kann hadronisch oder leptonisch zerfallen. Ein System, das aus zwei τ -Leptonen besteht - wie der $H \rightarrow \tau\tau$ Zerfallskanal - hat somit drei verschiedene Zerfallsmöglichkeiten. Zum Einen gibt es den dileptonischen

Zerfall ($\tau_{\text{lep}}\tau_{\text{lep}}$), dessen Feynman-Diagramm in Abbildung 5 dargestellt ist. Zum Anderen können die Teilchen rein hadronisch ($\tau_{\text{had}}\tau_{\text{had}}$) zerfallen. Schließlich gibt es noch den semileptonischen Zerfall ($\tau_{\text{lep}}\tau_{\text{had}}$), bei dem ein τ -Lepton leptonisch und das andere hadronisch zerfällt. Die Wahrscheinlichkeit für den rein hadronischen Zerfall ($H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$) liegt bei ca. 42%, während der semileptonische Zerfall ($H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$) eine Wahrscheinlichkeit von 46% und der rein leptonische Zerfall einer Wahrscheinlichkeit von 12% hat. Dies kann mittels folgender Tabelle nachvollzogen werden.

| Zerfallskanal | $\tau^- \rightarrow$ | Verweigungsverhältnis [%] |
|---------------|--------------------------------|---------------------------|
| leptonisch | $e^- \bar{\nu}_e \nu_\tau$ | $17,82 \pm 0,04$ |
| | $\mu^- \bar{\nu}_\mu \nu_\tau$ | $17,30 \pm 0,04$ |

Tabelle 2: Verzweigungsverhältnisse des τ -Leptons im leptonischen Zerfallskanal. Das τ -Teilchen kann hierbei entweder in ein Elektron e oder ein Myon μ zerfallen [16].

Im Rahmen dieser Analyse werden nur Ereignisse aus dem VBF-Kanal und vollständig leptonische Zerfälle des τ -Lepton-Paares berücksichtigt. Hierbei werden nur $H \rightarrow \tau_e \tau_\mu$ Zerfälle betrachtet.

2.4 Diskrete Symmetrien

Symmetrien sind Operationen, welche die Eigenschaften eines physikalischen Systems unverändert lassen. Hierbei gibt es kontinuierliche Symmetrien, die eine unendliche Anzahl an Symmetrieeoperationen besitzen und diskrete Symmetrien, bei denen die Anzahl der Symmetrieeoperationen begrenzt ist [18].

In diesem Kapitel werden zwei diskrete Symmetrien - die Parität \mathcal{P} und die Ladungskonjugation \mathcal{C} - näher betrachtet.

2.4.1 Parität

Die Parität, welche im 20. Jahrhundert von Paul Wigner eingeführt wurde, ist in der schwachen Wechselwirkung maximal verletzt. Dies zeigte 1956 das Wu-Experiment [27], bei dem der β -Zerfall von Cobalt untersucht wurde. Somit wurde im Rahmen der schwachen Wechselwirkung die Vermutung, dass ein gespiegelter physikalischer Prozess aus Symmetriegründen genauso existiert wie sein Ursursungsprozess, widerlegt.

Wendet man den Paritätsoperator \mathcal{P} auf einen Zustand $|\psi(\vec{r})\rangle$ mit Ortsvektor \vec{r} an, so wird der gesamte Raum an den drei Ebenen, die von den Einheitsvektoren aufgespannt werden, am Ursprung des Koordinatensystems gespiegelt:

$$\mathcal{P} |\psi(\vec{r})\rangle = |\psi(\vec{r})\rangle^1 \quad (12)$$

Da ein zweimaliges Anwenden des Paritätsoperators $\mathcal{P}^2 |\psi(\vec{r})\rangle = |\psi(\vec{r})\rangle$ die Wellenfunktion in sich selbst überführt, kommen als Eigenwerte nur ± 1 in Frage:

$$\mathcal{P} |\psi(\vec{r})\rangle = \pm |\psi(\vec{r})\rangle \quad (13)$$

¹Die in Gleichung 12, 13 und 14 verwendete Notation heißt Dirac-Notation und wurde im Rahmen der Quantenmechanik eingeführt. Hierbei wird ein quantenmechanischer Zustand mit einem Ket-Vektor $|\psi(\vec{r})\rangle$ dargestellt, wobei $|\psi(-\vec{r})\rangle$ sein Spiegelbild beschreibt.

So nennt man Eigenfunktionen $|\psi(\vec{r})\rangle$ \mathcal{P} -gerade, wenn ihr Eigenwert $+1$ beträgt und \mathcal{P} -ungerade, wenn ihr Eigenwert -1 beträgt.

2.4.2 Ladungskonjugation

Eine weitere Symmetrietransformation beschreibt die Ladungskonjugation [17]. Hierbei kommt es zu einem Austausch von Teilchen mit ihren Antiteilchen und somit zu einer Spiegelung des Vorzeichens der Ladung und allen anderen additiven Quantenzahlen, wohingegen die Masse, Impuls, Energie und Spin des Teilchens unverändert bleibt. Die Eigenwerte des Ladungskonjugationsoperators betragen ± 1 , da das doppelte Anwenden des \mathcal{C} -Operators auf ein Teilchen wieder das Teilchen ergibt:

$$\mathcal{C}^2 |\psi(\vec{r})\rangle = |\psi(\vec{r})\rangle \quad (14)$$

Die Eigenzustände des \mathcal{C} -Operators sind Teilchen, die gleichzeitig ihr eigenes Antiteilchen sind, und somit bezüglich aller additiven Quantenzahlen neutral sind. Dazu gehören z.B. das Photon A und das neutrale Pion π^0 .

Wie die Parität, ist auch die Ladungskonjugation in der elektromagnetischen und starken Wechselwirkung erhalten, jedoch nicht in der schwachen Wechselwirkung [28]. Dies wird beispielsweise durch das Verhalten von Neutrinos unter Ladungskonjugation bestätigt. Hier wird ein linkshändiges Neutrino in ein linkshändiges Antineutrino transformiert. Letzteres wird experimentell jedoch nicht beobachtet. Antineutrinos kommen in der Natur nur rechtshändig vor.

2.4.3 CP-Symmetrie

Führt man die \mathcal{C} - und \mathcal{P} -Transformation hintereinander aus, so spricht man von einer \mathcal{CP} -Transformation. Dabei spielt die Reihenfolge der Anwendung der Operatoren auf einen Zustand keine Rolle. In vielen zuvor \mathcal{C} - oder \mathcal{P} -verletzenden Prozessen der schwachen Wechselwirkung führt die Hintereinanderschaltung der beiden Operatoren zu einer Aufhebung der Symmetriebrechung. So geht ein linkshändiges Neutrino unter \mathcal{CP} -Transformation in ein rechtshändiges Antineutrino über.

Allerdings wurde 1964 von Christenson, Cronin, Fitch und Turlay [29] erstmals eine \mathcal{CP} -Verletzung beim Zerfall langlebiger K -Mesonen beobachtet. Diese zerfallen nicht nur in drei geladene Pionen, sondern zu einem geringen Anteil auch in zwei geladene Pionen. Im Standardmodell wird die \mathcal{CP} -Verletzung in der schwachen Wechselwirkung mithilfe der Cabibbo-Kobayashi-Maskawa-Matrix (CKM-Matrix) erklärt [21].

2.5 Parametrisierung von CP-Verletzung und CP-sensitive Observablen

In einem modellunabhängigen Ansatz kann die effektive $U(1)_Y \times SU(2)_{I_W}$ invariante Lagrangedichte für \mathcal{CP} -verletzende HVV-Wechselwirkungen ($V = W^\pm, Z, A$) parametrisiert werden, indem die Lagrangedichte des Standardmodells mit \mathcal{CP} -ungeraden Operatoren der Massendimension sechs erweitert wird, die das Higgs-Feld und elektroschwache Eichfelder beinhalten [30] [31] [32]. Vernachlässigt man den Beitrag von \mathcal{CP} -verletzenden geladenen Dreifach-Eichboson-Kopplungen lässt sich die effektive Lagrangedichte schreiben

als

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{f_{\tilde{B}B}}{\Lambda^2} \mathcal{O}_{\tilde{B}B} + \frac{f_{\tilde{W}W}}{\Lambda^2} \mathcal{O}_{\tilde{W}W}, \quad (15)$$

wobei $f_{\tilde{B}B}$ und $f_{\tilde{W}W}$ dimensionslose Wilson-Koeffizienten, $\mathcal{O}_{\tilde{B}B}$ und $\mathcal{O}_{\tilde{W}W}$ die Operatoren, sowie Λ die Skala der neuen Physik beschreiben [32] [33].

Unter der Annahme, dass alle Wechselwirkungen zwischen dem Higgs-Boson und anderen SM-Teilchen nach den Vorhersagen des Standardmodells stattfinden, kann die effektive Lagrangedichte nach der elektroschwachen Symmetriebrechung in der unitären Eichung aufgestellt werden. In der Massenbasis des Higgs-Bosons H , des Photons A und der schwachen Eichbosonen Z und W^\pm wird diese geschrieben als

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \tilde{g}_{HAA} H \tilde{A}_{\mu\nu} A^{\mu\nu} + \tilde{g}_{HAZ} H \tilde{A}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HZZ} H \tilde{Z}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HWW} H \tilde{W}_{\mu\nu}^+ W^{-\mu\nu} \quad (16)$$

mit $V^{\mu\nu}$ und $\tilde{V}^{\mu\nu} = \epsilon^{\mu\nu\rho\sigma} V_{\rho\sigma}$ ($V = W^\pm, Z, A$). Die Koeffizienten \tilde{g}_{HVV} beschreiben die Kopplungsparameter der CP-ungeraden Operatoren. Aufgrund der Forderung nach $U(1)_Y \times SU(2)_{I_W}$ Invarianz sind nur zwei der vier Kopplungsparameter unabhängig voneinander. Diese hängen von den zwei Parametern aus Gleichung 15 wie folgt ab:

$$\tilde{d} = -\frac{m_W^2}{\Lambda^2} f_{\tilde{W}W} \quad \tilde{d}_B = -\frac{m_W^2}{\Lambda^2} \frac{\sin^2 \theta_W}{\cos^2 \theta_W} f_{\tilde{B}B} \quad (17)$$

Basierend auf der Annahme, dass $\tilde{d} = \tilde{d}_B$ gilt, reduziert sich die Anzahl der Parameter zu

$$\tilde{g}_{HAA} = \tilde{g}_{HZZ} = \frac{1}{2} \tilde{g}_{HWW} = \frac{g}{2m_W} \tilde{d} \quad \tilde{g}_{HAZ} = 0, \quad (18)$$

wobei g die Kopplungskonstante der schwachen Wechselwirkung darstellt. Setzt man diese Relationen in Gleichung 16 ein, so ergibt sich eine Abhängigkeit der effektiven Lagrangedichte von einem einzigen Parameter \tilde{d} . Somit ist es möglich, die Stärke der CP-Verletzung bei der VBF-Higgs-Bosonen-Produktion durch einen Parameter - \tilde{d} - zu beschreiben [31] [32] [33]. Das entsprechende Matrixelement

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \tilde{d} \cdot \mathcal{M}_{\text{CP-odd}}, \quad (19)$$

das proportional zu dem differentiellen Wirkungsquerschnitt ist, beinhaltet im ersten Term die CP-geraden Anteile der Standardmodell-Vorhersage, wohingegen der zweite Term die Beiträge der CP-ungeraden Matrixelemente beinhaltet. Die CP-ungeraden Matrixelemente, die aus den in 15 eingeführten Operatoren hervorgehen, werden hierbei in Abhängigkeit des Parameters \tilde{d} beschrieben. Durch Quadrieren des Matrixelements

$$|\mathcal{M}|^2 = |\mathcal{M}_{\text{SM}}|^2 + \tilde{d} \cdot 2\text{Re}(\mathcal{M}_{\text{SM}}^* \mathcal{M}_{\text{CP-odd}}) + \tilde{d}^2 \cdot |\mathcal{M}_{\text{CP-odd}}|^2 \quad (20)$$

ergeben sich zwei CP-gerade Terme, nämlich $|\mathcal{M}_{\text{SM}}|^2$ und $|\mathcal{M}_{\text{CP-odd}}|^2$, die folglich nicht für eine mögliche CP-Verletzung verantwortlich sind. Der mittlere Term $2\text{Re}(\mathcal{M}_{\text{SM}}^* \mathcal{M}_{\text{CP-odd}})$ hingegen ist CP-ungerade und stellt somit eine mögliche Quelle von CP-Verletzung dar. Bildet man den Quotienten von diesem Interferenzterm und dem quadrierten Matrixelement des Standardmodells gelangt man zur Definition der Optimalen Observablen:

$$\mathcal{OO} = \frac{2\text{Re}(\mathcal{M}_{\text{SM}}^* \mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\text{SM}}|^2} \quad [31][32][33] \quad (21)$$

Der Endzustand des VBF-Produktionsprozesses, bestehend aus dem Higgs-Boson und zwei Quarks, kann durch sieben Phasenraumvariablen beschrieben werden, wenn man die Masse des Higgs-Bosons festlegt, die Massen der Jets vernachlässigt und Impulserhaltung in der transversalen Ebene annimmt. Das Konzept der Optimalen Observablen beruht nun darauf, dass sie die gesamte Phasenrauminformation in einer einzigen Beobachtungsgröße zusammenfasst und gleichzeitig die höchste Empfindlichkeit für CP-Verletzungen bei kleinen Abweichungen vom Standardmodell aufweist.

Um eine Suche nach CP-Verletzung vorzunehmen, kann der Mittelwert der Optimalen Observable gemessen werden. Hierbei wird ausgenutzt, dass die Optimale Observable eine CP-ungerade Observable ist und daher bei Anwenden des CP-Operators auf \mathcal{OO} folgendes gilt:

$$\mathcal{CP} \mathcal{OO} = -\mathcal{OO} \quad (22)$$

Bei CP-Invarianz muss die Verteilung der Optimalen Observablen \mathcal{OO} symmetrisch sein und einen Mittelwert von $\langle \mathcal{OO} \rangle = 0$ haben. Würde man hingegen einen von null verschiedenen Mittelwert beobachten $\langle \mathcal{OO} \rangle \neq 0$, wäre dies ein klares Indiz für CP-Verletzung [34] [35] [36] [37].

Durch Integrieren über den Wirkungsquerschnitt ergibt sich die erwartete Abhängigkeit des Mittelwerts der Optimalen Observable von \tilde{d} :

$$\langle \mathcal{OO} \rangle = \frac{\int \mathcal{OO} d\sigma_{\text{SM}} + \tilde{d} \int \mathcal{OO} d\sigma_{\text{CP-odd}} + \tilde{d}^2 \int \mathcal{OO} d\sigma_{\text{CP-even}}}{\int d\sigma_{\text{SM}} + \tilde{d} \int d\sigma_{\text{CP-odd}} + \tilde{d}^2 \int d\sigma_{\text{CP-even}}} \quad (23)$$

Im Zähler verschwinden die Terme, in denen die CP-ungerade Optimale Observable über die CP-geraden Phasenräume ($d\sigma_{\text{SM}}, d\sigma_{\text{CP-even}}$) integriert werden. Im Nenner hingegen verschwindet das Integral $\int d\sigma_{\text{CP-odd}}$. Daraus ergibt sich für den Mittelwert der Optimalen Observable folgende Gleichung:

$$\langle \mathcal{OO} \rangle = \frac{\tilde{d} \int \mathcal{OO} d\sigma_{\text{CP-odd}}}{\int d\sigma_{\text{SM}} + \tilde{d}^2 \int d\sigma_{\text{CP-even}}} \quad (24)$$

Für kleine Werte von \tilde{d} gilt in erster Näherung somit eine lineare Abhängigkeit der Optimalen Observable von \tilde{d} ($\langle \mathcal{OO} \rangle = \text{konst.} \cdot \tilde{d}$). Dieses lineare Verhalten bei kleinen Werten von \tilde{d} wird im Laufe der Analyse genutzt, um die Sensitivität auf \tilde{d} zu untersuchen.

3 Das ATLAS Experiment

Am Europäischen Labor für Teilchenphysik CERN befindet sich mit einem Umfang von etwa 27 km, der größte und energiereichste Teilchenbeschleuniger der Welt - der *Large Hadron Collider* (LHC). In der Datennahmeperiode 2 (*Run 2*) zwischen 2015 und 2018 wurden hier unter anderem zwei entgegengesetzt zirkulierende Protonstrahlen zur Kollision gebracht. Dabei umfasste jeder Strahl 2500 Pakete, welche auf eine Energie von 6,5 TeV beschleunigt wurden. Die Spitzenluminosität betrug $2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ [38] und die integrierte Gesamtluminosität, die vom ATLAS Detektor geliefert wurde, belief sich 2018 auf 63.4fb^{-1} [39]. Im Oktober 2022 soll die dritte Datennahmeperiode (*Run 3*) gestartet werden [40], bei dem die Energie ursprünglich geplant weiter auf knapp 14 TeV bei einer Luminosität von $2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ [38] erhöht werden soll.

Am LHC sind vier Haupt-Experimente an den Kollisionspunkten installiert: ALICE, CMS, LHCb und ATLAS. Die in dieser Arbeit durchgeführten Analysen und Studien beziehen sich auf den ATLAS-Detektor, welcher im folgenden Abschnitt vorgestellt wird.

3.1 Der ATLAS Detektor

Der ATLAS-Detektor ist ein Mehrzweck-Teilchendetektor mit einer vorwärts-rückwärts-symmetrischen zylindrischen Geometrie. Er ist 25 Meter hoch, 44 Meter lang und hat ein Gewicht von etwa 7000 Tonnen [41]. Sein Querschnitt ist in Abbildung 7 zu sehen.

3.1.1 Koordinatensystem

ATLAS verwendet ein rechtshändiges Koordinatensystem [41], wobei der Ursprung am Kollisionspunkt der zwei gegenläufigen Teilchenstrahlen liegt. Wie in Abbildung 6 dargestellt, zeigt die z -Achse entlang des Strahlrohrs, während die x -Achse vom Kollisionspunkt zum Zentrum des LHC-Rings zeigt. Die y -Achse weist nach oben. Die Transversalebene ist senkrecht zum Strahl gerichtet und entspricht der x - y -Ebene. Der Azimuthwinkel ϕ beschreibt den Winkel zur x -Achse in der transversalen Ebene. Der Polarwinkel θ gibt den Winkel bezüglich der z -Achse an. Überlicherweise wird der Polarwinkel in die Pseudorapidität

$$\eta = -\ln(\tan \theta/2) \quad (25)$$

transformiert. Für Teilchen mit Energie E , deren Masse bezüglich ihres Impulses vernachlässigbar klein ist ($m \ll p$), stimmt die Pseudorapidität η mit der Rapidität

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (26)$$

überein, wobei p_z die Impulskomponente entlang der Strahlachse angibt. Entlang dieser Achse ist der Abstand zweier Teilchen bezüglich ihrer Pseudorapidität

$$\Delta y = y_2 - y_1 \quad (27)$$

invariant unter Lorentz-Boost.

Eine weitere nützliche Variable ist ΔR , welche über

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (28)$$

definiert ist und den Winkelabstand zweier Objekte angibt. Da für $m \ll p$ der Abstand bezüglich der Pseudorapidität $\Delta\eta = \eta_2 - \eta_1$ zweier Teilchen lorentzinvariant ist, gilt dasselbe somit auch für ΔR .

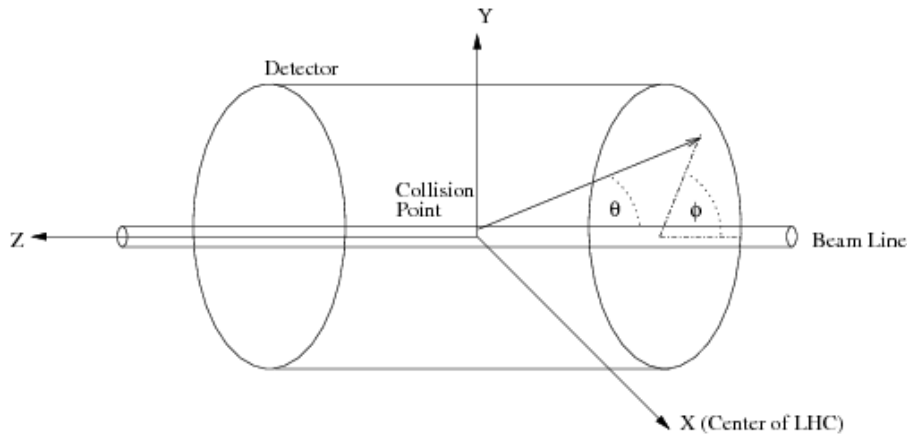


Abbildung 6: Das rechtshändige Koordinatensystem des ATLAS-Detektors mit seiner 2π -Symmetrie. In der Abbildung ist der Azimutwinkel ϕ , der Polarwinkel θ , der Kollisionsspunkt am Ursprung des Koordinatensystems, sowie die Richtung des Teilchenstrahls zu sehen [42].

3.1.2 Detektor-Komponenten

Das **Magnet-System** [41] des ATLAS-Detektors setzt sich aus einem supraleitendem Solenoid, sowie drei supraleitenden Toroiden zusammen. Zwei der drei Toroiden bilden die Endkappen des Magnetsystems, während der zentrale, zylinderförmige Toroid das Zentrum des Experiments umgibt. Jeder Toroidmagnet enthält acht Spulen, wobei die Endkappen-Spulen um $22,5^\circ$ gegenüber dem zentralen Toroid gedreht sind, um einen radialen Überlapp zu erreichen. Mit diesem Toroidspulensystem wird das Magnetfeld für das Myonspektrometer bereitgestellt. Im zentralen Bereich wird hier ein Magnetfeld von 1 T und in den Endkappenregionen ein Magnetfeld von 0,5 T erreicht. Das Solenoid hingegen wird dafür genutzt, ein 2 Tesla starkes, homogenes Magnetfeld parallel zur Strahlachse zu erzeugen, in welches der **innere Spurdetektor** eingebettet ist. Dieser enthält drei verschiedene Subsysteme, welche die Trajektorie, den Impuls und die Ladung von elektrisch geladenen Teilchen innerhalb eines Pseudorapiditätsbereichs von $|\eta| < 2,5$ misst. In der Datennahmepériode 2 wurden Daten mit einer Effizienz von mehr als 99% aufgezeichnet. Die innerste Komponente bildet der Silizium-Pixel-Detektor. Die erste Pixellage, die sogenannte 'b-layer', befindet sich in 33,25 mm Entfernung vom Strahlrohr und wurde erst für die zweite Datennahmepériode eingebaut, um die Präzision und Robustheit der Spurrekonstruktion zu verbessern [43]. Die Pixelgröße beträgt hier $50 \times 250 \mu\text{m}^2$, während die Pixel der drei darauffolgenden Lagen eine Größe von $50 \times 400 \mu\text{m}^2$ haben [41]. Um den Pixeldetektor herum befindet sich ein Silizium-Streifendetektor. Seine Detektormodulen und integrierten Auslestreifen sind auf vier zylindrische Lagen, sowie auf zwei Endkappen mit je neun ebenen Scheiben, verteilt. Mit diesem Detektor wird eine Auflösung von $17 \mu\text{m}$ in der Transversalebene und $580 \mu\text{m}$ in Strahlrichtung erreicht [41]. Die äußerste Komponente des inneren Detektors, der Übergangsstrahlungsdetektor, besteht aus 350848 gasgefüllten (70% Ar, 27% CO₂ und 3% O₂ [44]) Driftröhren, die je einen Durchmesser von 4 mm haben [43]. Im Inneren der Röhren befindet sich ein vergoldeter Wolfram-Draht. Durchqueren geladene Teilchen die Röhrenwände, so geben sie

elektromagnetische Strahlung im Röntgenbereich (Übergangsstrahlung) ab, welche das Gasgemisch in den Röhren ionisiert [41]. Für die Teilchenidentifikation werden die Energiedepositionen der Übergangsstrahlung im Detektor genutzt. Elektronen bzw. Positronen deponieren typischerweise eine Energie von 8-10 keV, während minimal-ionisierende Teilchen wie Pionen etwa 2 keV deponieren [30].

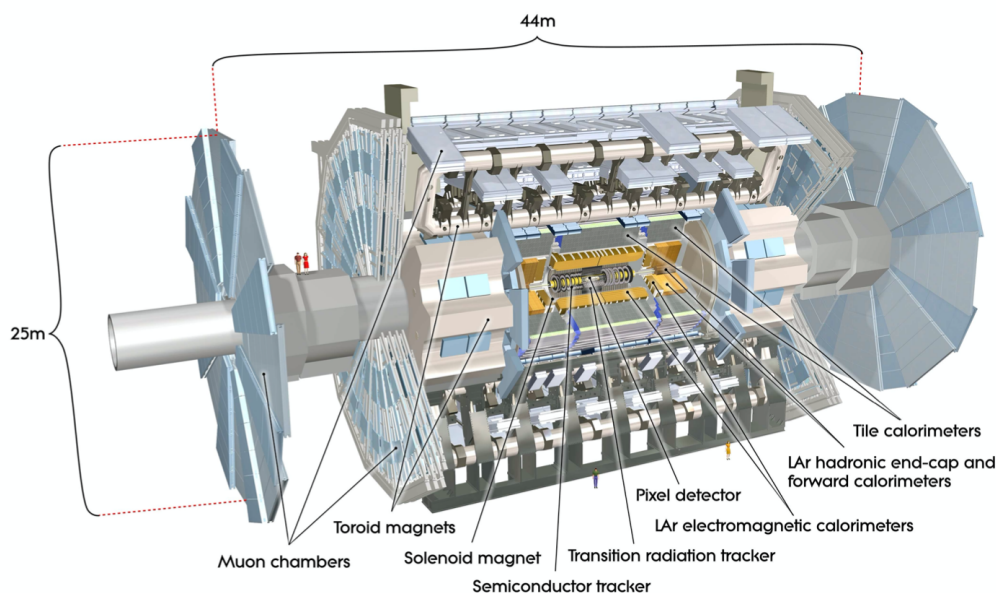


Abbildung 7: Querschnitt des ATLAS-Detektors mit seinen verschiedenen Komponenten, welche im Text näher beschrieben werden. Der ATLAS-Detektor ist zylinderförmig gebaut, wobei er 25 m hoch und 44 m lang ist [41].

Um den inneren Detektor herum befindet sich das **elektromagnetische Kalorimeter**. Es ist für eine hochauflösende Messung der Energie von Elektronen und Photonen mittels elektromagnetischen Schauern ausgelegt und besteht aus einem zylindrischen Körper, sowie zwei Endkappenteilen, die einen Pseudorapiditätsbereich von $|\eta| < 1,475$ bzw. $1,375 < |\eta| < 3,2$ abdecken. Er ist in abwechselnde Schichten von Bleiabsorbern, Ausleseelektroden und flüssigem Argon als Sensormaterial unterteilt, wobei die Elektroden und Absorberplatten die Form einer Ziehharmonika haben. Dadurch wird über den gesamten ϕ -Bereich eine einheitliche Auflösung erreicht [41]. Hadronen geben nur ein Teil ihrer Energie im elektromagnetischen Kalorimeter ab. Sie dringen in das **Hadronenkalorimeter** ein. Dieses ist aus einem zentralen Barrel² und zwei verlängerten Barrels aufgebaut, wobei die zentrale Komponente einen Bereich von $|\eta| < 1,0$ bzw. $0,8 < |\eta| < 1,7$ abdeckt, während die Endkappen den Bereich $1,5 < |\eta| < 3,2$ abdecken. Im zentralen Bereich werden zur Detektion von Hadronen Stahlabsorber und Plastiksintillatoren als aktives Material verwendet. Die Endkappen des hadronischen Kalorimeters befinden sich hinter denen des elektromagnetischen Kalorimeters. Als Absorbermaterial werden hier Kupferplatten verwendet, die senkrecht zum Strahlrohr angeordnet sind. Zwischen diesen und einigen Elektroden befindet sich flüssiges Argon als Sensormaterial.

Das Vorwärtskalorimeter (FCal), welches sich zwischen der Strahlachse und dem hadroni-

²Ein Barrel ist ein zylinderförmiges, hohles Gebilde.

schen Endkappenkalorimeter befindet, deckt den Bereich $3,1 < \eta < 4,9$ ab. Es besteht aus drei Modulen, wobei das erste Modul Kupfer als Absorbermaterial verwendet, um elektromagnetische Schauer zu detektieren. Die restlichen Module sind für die Messung der hadronischen Schauer zuständig und verwenden Wolfram als Absorbermaterial. In den Modulen sind konzentrische Stäbe und Rohre parallel zur Strahlachse angeordnet. In den Lücken zwischen den Stäben und Rohren befindet sich flüßiges Argon als Sensormaterial [41].

Schließlich bildet das **Myonspektrometer** die äußerste Komponente des ATLAS-Detektors. Hier werden die Spuren der Myonen, die die inneren Teile des Detektors relativ ungehindert durchlaufen, detektiert. Beim Passieren der Myonen durch diesen Teil des Detektors werden die Bahnen dieser durch ein starkes torodiales Magnetfeld gekrümmt, wodurch eine Messung des Myonenimpulses ermöglicht wird. Über den größten Teil des η -Bereiche wird diese Präzisionsmessung der Spurkoordinaten mit überwachten Driftrohren, welche eine räumliche Auflösung von $80 \mu\text{m}$ pro Röhre und $35 \mu\text{m}$ pro Kammer entlang der Strahlachse haben, durchgeführt. Bei großen Pseudorapititäten hingegen werden Kathodenstreifen-Kammern, die eine höhere Granularität besitzen, verwendet. Diese haben eine hohe Ratenkapazität, sowie eine hohe Raum- und Zeitauflösung. So wird mit den Kathodenstreifen-Kammern eine Auflösung von $40 \mu\text{m}$ pro Kammer in Richtung der Biegung erreicht, wobei quer zur Richtung der Biegung eine Auflösung von ca. 5 mm pro Kammer in geliefert werden kann. Für $|\eta| < 1,4$ wird das Magnetfeld durch den zentralen Toroid erzeugt. Hier werden zur Messung der Spuren Kammern eingesetzt, die in drei zylindrischen Lagen in Abständen von etwa 5 m , $7,5 \text{ m}$ und 10 m um die Strahlachse angeordnet sind. Zwei kleinere Endkappenmagnete sorgen für das Magnetfeld im Bereich von $1,6 < |\eta| < 2,7$ und im Übergangsbereich ($1,4 < |\eta| < 1,6$) wird die magnetische Ablenkung durch eine Kombination aus Barrel- und Endkappenfeldern gewährleistet [41].

Da die Kollisionsrate beim LHC 40 MHz beträgt, und nur ca. 1000 Ereignisse gemessen werden können, ist ein zweistufiges **Triggersystem** [45] notwendig. Dieses ist in der Lage aus der enormen Menge an Daten, die Ereignisse herauszufiltern, die interessant sind und ist somit für die Entscheidung verantwortlich, ob Daten einer bestimmten Kollision aufgezeichnet werden, oder nicht. Die erste Instanz des Triggersystems bildet der L1-Trigger [45]. Er ist hardwarebasiert und selektiert Myonen, Elektronen, Photonen, Jets und hadronisch zerfallenden τ -Leptonen, die eine hohen Transversalimpuls haben. Aber auch Ereignisse mit großer Gesamtenergiedeposition werden erfasst. Der L1-Trigger akzeptiert Ereignisse innerhalb einer durchschnittlichen Entscheidungszeit von $2,5 \mu\text{s}$ und reduziert die Detektorausleserate auf 100 kHz . Die sogenannten interessanten Regionen, welche die η - und ϕ -Koordinaten der ausgewählten Ereignisse enthalten, werden dann vom L1-Trigger zum *High Level Trigger* (HLT) weitergeleitet. Dieser ist im Gegensatz zum L1-Trigger software-basiert und verringert die Triggerrate von 100 kHz auf 1 kHz bei einer durchschnittlichen Entscheidungszeit von ca. 200 ms . Hierbei nutzt der HLT bei voller Granularität alle verfügbaren Detektordaten innerhalb der interessanten Regionen und verwendet Algorithmen, um bestimmte Teilchen wie Elektronen oder Jets zu identifizieren, sowie um ihre Eigenschaften zu bestimmen oder globale Merkmale des Ereignisses zu erfassen. Die Auswahl und Verarbeitung der interessanten Ereignisse erfolgen mit Offline-Analyseverfahren, basierend auf Standard-ATLAS-Ereignisrekonstruktions- und -Analyseanwendungen [41]. Zusammenfassend sind in Tabelle 3 die angestrebte Auflösung und die η -Abdeckung der einzelnen Detektorkomponenten dargestellt.

| Detektorkomponente | Geforderte Auflösung | η -Abdeckung | |
|------------------------------------|---|----------------------|----------------------|
| | | Messung | Trigger |
| Innerer Detektor | $\sigma_{p_T}/p_T = 0,05\%p_T \oplus 1\%$ | $\pm 2,5$ | |
| Elektromagnetisches Kalorimeter | $\sigma_E/E = 10\%/\sqrt{E} \oplus 0,7\%$ | $\pm 3,2$ | $\pm 2,5$ |
| Hadronisches Kalorimeter | | | |
| - zentral und Endkappen | $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ | $\pm 3,2$ | $\pm 3,2$ |
| - vorwärts | $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ | $3,1 < \eta < 4,9$ | $3,1 < \eta < 4,9$ |
| Myonspektrometer | $\sigma_{p_T}/p_T = 10\%$ bei $p_T = 1\text{TeV}$ | $\pm 2,7$ | $\pm 2,4$ |

Tabelle 3: Übersicht der angestrebten Auflösung und der η -Abdeckung der einzelnen ATLAS-Detektorkomponenten. Die verwendete \oplus -Notation ist folgendermassen zu verstehen: $a \oplus b = \sqrt{a^2 + b^2}$. Wenn nicht anders gekennzeichnet, sind alle Energie- und Impulswerte in GeV angegeben [41].

4 Neuronale Netze

Künstliche, neuronale Netze sind nichtlineare, statistische Modelle, die zur Informationsverarbeitung und Datenanalyse dienen. Sie sind eine der am weitesten verbreiteten Anwendungen des maschinellen Lernens und haben besonders in den letzten zehn Jahren an großer Bedeutung gewonnen. Die zentrale Idee besteht darin, den komplexen, nicht-linearen Zusammenhang zwischen Eingangs- und Ausgangsgrößen zu modellieren. Es gibt verschiedene Arten von neuronalen Netzen, die sowohl für Klassifizierungs-, als auch für Regressionsprobleme, verwendet werden [10]. Im Rahmen dieser Arbeit werden tiefe *Feed-Forward*-Regressionsnetze verwendet, weswegen sich die kommenden Unterkapitel auf diese Art beziehen. Da in dieser Analyse, die von dem neuronalen Netz vorhergesagten Werte mit den bereits bekannten, wahren Zielwerten verglichen werden und dieser Vergleich zur Anpassung der Netzparameter herangezogen wird, wird diese Form des Trainings 'überwachtes Lernen' genannt. Es wird erklärt, wie ein neuronales Netz aufgebaut ist, wie das Netz aus Daten lernt und wie die Verlustfunktion mit dem Adam-Optimierungs-Algorithmus, der für die Analyse in dieser Arbeit verwendet wurde, minimiert wird. Außerdem erfolgt im letzten Unterkapitel eine kurze Vorstellung eines frei zugänglichen Softwarepakets namens OPTUNA [11], welches in dieser Arbeit für die Hyperparameter³-Optimierung verwendet wurde.

4.1 Allgemeiner Aufbau

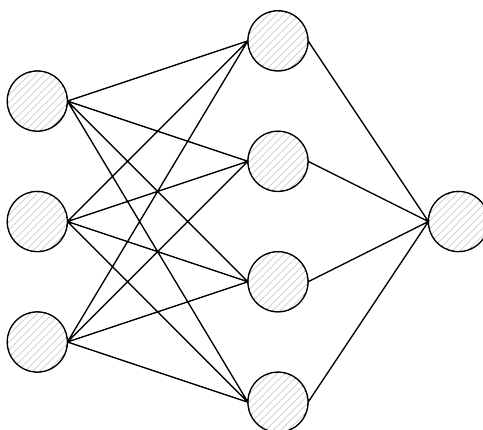


Abbildung 8: Beispielhafter Aufbau eines einfachen neuronalen Netzes: Es besteht aus der Eingangs-Schicht mit drei Knoten, einer verborgenen Schicht mit vier Knoten und der Ausgabeschicht mit einem Knoten.

Ein neuronales Netz ist aus Neuronen bzw. Knoten, den Grundbausteinen, aufgebaut. Diese Neuronen sind in Schichten angeordnet, wobei es drei unterschiedliche Arten von Schichten gibt. Zum Einen gibt es die Eingangs-Schicht, deren Knoten die Werte der Eingangs-Variablen aufnehmen. Die zweite Art von Schicht wird als verborgene Schicht bezeichnet. Jeder Knoten dieser Schicht ist mit jedem Knoten der ersten Schicht verbunden. Sofern es in einem neuronalen Netz mehr als eine verborgene Schicht gibt, spricht

³Ein Hyperparameter ist ein Parameter eines neuronalen Netzes, der nicht im Trainingsprozess bestimmt wird, sondern bereits vor Beginn des Trainingsprozess manuell festgelegt werden muss. Die Wahl der Hyperparameter in einem Netz können die Leistung des Netzes erheblich beeinflussen [46].

man von einem tiefen Netz. Die Knoten der letzten Schicht, die sogenannte Ausgangsschicht, enthalten die Werte, die vom neuronalen Netz für einen bestimmten Satz von Eingabewerten vorhergesagt wurden. Möchte man nur ein Merkmal vorhersagen, so enthält die Ausgangsschicht entsprechend nur einen Knoten. Eine weitere wichtige Eigenschaft eines *Feed-Forward*-Netzes ist, dass nur die Knoten aus aufeinanderfolgenden Schichten miteinander verbunden sind. Hierbei ist jede Verbindung i zwischen den Neuronen mit einem Gewicht w_i assoziiert, sowie jeder Knoten mit einer Bias b behaftet. Ausgeschlossen davon sind jedoch die Neuronen der Eingabe-Schicht.

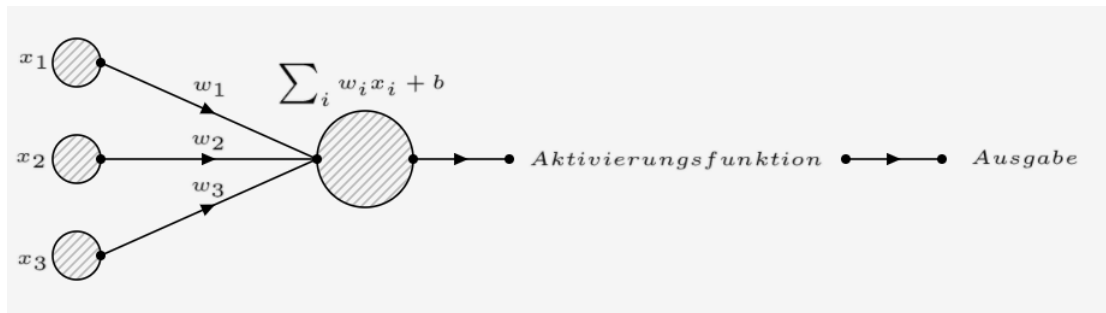


Abbildung 9: Graphische Darstellung eines Knotens in einem neuronalen Netzwerk

Wie in Abbildung 9 dargestellt, repräsentiert ein Knoten einer verborgenen oder Ausgangsschicht eine mathematische Funktion. Hier werden die in den Knoten eingehenden Gewichte w_i mit den entsprechenden Eingabe-Werten x_i multipliziert und die verschiedenen Ergebnisse für jede Verbindung i aufsummiert. Anschließend wird noch der, für einen Knoten charakteristische, Bias b zu der Summe hinzu addiert. Dieses Zwischenergebnis wird dann in eine sogenannte Aktivierungsfunktion σ eingesetzt, was die endgültige Ausgabe eines Knoten $j + 1$ liefert:

$$(\text{Ausgabe})_i^{j+1} = \sigma\left(\sum_i w_i x_i^j + b_i^{j+1}\right) \quad (29)$$

Durch die Aktivierungsfunktionen werden Nichtlinearitäten in das Modell eingeführt und eine unterschiedliche Wahl der nicht-linearen Funktionen führt zu unterschiedlichen Rechen- und Trainingseigenschaften der Neuronen. In Abbildung sind mögliche Aktivierungsfunktionen abgebildet. Die am häufigsten verwendete Funktion

$$\sigma_{\text{ReLU}}(x) = \max(0, x) \quad (30)$$

ist die ReLU-Funktion (engl.: *rectified linear unit*). Für Werte größer als Null ist die Funktion linear mit einer konstanten positiven Steigung, während sie für Werte kleiner oder gleich Null kontinuierlich bei Null liegt. Da die Ausgabe der Knoten in der letzten Schicht eines Regressionsnetzwerkes jede reelle Zahl annehmen können soll, wird hier üblicherweise die Identität als Aktivierungsfunktion gewählt [10].

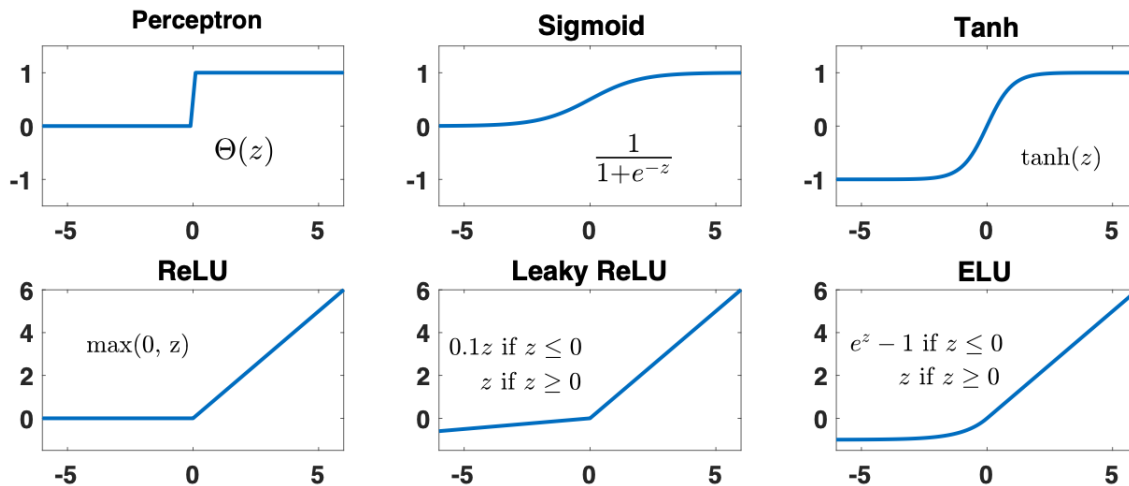


Abbildung 10: Ein Überblick möglicher Aktivierungsfunktionen. Am häufigsten verwendet wird die ReLU-Funktion. Knoten der Eingabe-Schicht haben keine Aktivierungsfunktionen [10].

4.2 Trainingsprozess

Bevor ein neuronales Netz trainiert wird, wird der gegebene Datensatz üblicherweise zunächst in ein Trainings- und Testdatensatz aufgeteilt. Der Trainingsdatensatz beinhaltet die Daten, die zum Trainieren des Modells und somit zur Bestimmung der Netzwerkparameter verwendet werden, wohingegen der Testdatensatz dafür eingesetzt wird, die Leistung eines bereits trainierten Modells zu bewerten [10].

Um ein neuronales Netzwerk zu trainieren, muss jedoch zuallererst eine Verlustfunktion definiert werden. In Regressionsproblemen wird häufig die mittlere quadratische Abweichung

$$\text{MSE}(\vec{w}) = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i(\vec{w}))^2 \quad (31)$$

als Verlustfunktion gewählt. Hierbei beschreibt n die Anzahl der Datenpunkte (bzw. Ereignisse), y_i den i -ten Zielwert und \hat{y}_i den vom neuronalen Netz vorhergesagten Wert. Der Vektor \vec{w} beinhaltet die Parameter des Netzes - die Gewichte und die Biases. Wenn ein Netzwerk mehr als eine Ausgabe, also eine m -dimensionale Ausgaben liefern soll, werden in obiger Gleichung y und \hat{y} durch m -dimensionale Vektoren \vec{y} und $\vec{\hat{y}}$ ersetzt und vor der Summe wird noch ein Faktor $\frac{1}{m}$ hinzugefügt.

Ein beliebiges Netz kann nun mit der Wahl der Verlustfunktion trainiert werden. Hierfür wird die Methode des Gradientenabstiegs angewendet, um die besten Gewichte und Biases zu finden und somit die Verlustfunktion zu minimieren.

Für die Berechnung der Gradienten ist im Falle eines neuronalen Netzes ein spezieller Algorithmus, namens Fehler-Rückpropagierung (engl.: *Error Backpropagation*) notwendig. Dieser zielt auf die Berechnung zweier Gradienten, $\frac{\partial f}{\partial b_j^l}$ und $\frac{\partial f}{\partial w_{jk}^l}$ ab und basiert auf vier fundamentalen Gleichungen. Um diese zu verstehen, gilt es ein Netzwerk mit $l = 1, \dots, L$ Schichten zu betrachten. Die Aktivierung a_j^l mit der Aktivierungsfunktion σ und dem

Bias b_j^l des j -ten Neurons in der l -ten Schicht ist

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) = \sigma(z_j^l) \quad (32)$$

mit

$$z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l. \quad (33)$$

w_{jk}^l ist das Gewicht, das mit der Verbindung zwischen dem k -ten Neuron in der $l-1$ -ten Schicht zu dem j -ten Neuron in der l -ten Schicht assoziiert ist. Die erste der vier fundamentalen Gleichungen

$$\Delta_j^L = \frac{\partial f}{\partial z_j^L} = \sum_k \frac{\partial f}{\partial a_k^L} \frac{\partial a_k^L}{\partial z_j^L} = \frac{\partial f}{\partial a_j^L} \frac{\partial a_j^L}{\partial z_j^L} = \frac{\partial f}{\partial a_j^L} f'(z_j^L) \quad (34)$$

ist für die Berechnung des Fehlers des j -ten Knoten in der Ausgabe-Schicht L zuständig. Dieser ist als Änderung der Verlustfunktion f in Abhängigkeit der gewichteten Eingabe z_j^L definiert. Die Summe über die k Neuronen in der Ausgabe-Schicht L fällt hierbei im dritten Schritt von obiger Gleichung weg, da die Aktivierung a_k^L nur von z_j^L abhängt, wenn $k = j$ ist.

Die Berechnung des Fehlers Δ_j^l des j -ten Knoten in der l -ten Schicht in Bezug auf den Fehler eines Knoten in der nächsten $l+1$ -ten Schicht erfolgt mit der zweiten fundamentalen Gleichung:

$$\Delta_j^l = \frac{\partial f}{\partial z_j^l} = \sum_k \frac{\partial f}{\partial z_k^{l+1}} \frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \frac{\partial z_k^{l+1}}{\partial z_j^l} \Delta_k^{l+1} \quad (35)$$

Hierbei ist z_k^{l+1} wie in Gleichung 33 definiert (setze $l = l+1$), wodurch die Ableitung $\frac{\partial z_k^{l+1}}{\partial z_j^l} = w_{jk}^{l+1} \sigma'(z_j^l)$ ergibt. Folglich kann man die zweite fundamentale Gleichung schreiben als:

$$\Delta_j^l = \sum_k w_{jk}^{l+1} \Delta_k^{l+1} \sigma'(z_j^l) \quad (36)$$

An dieser Gleichung wird zum ersten Mal deutlich, warum der Algorithmus 'Fehler-Rückpropagation' heißt. Denn hier wird der Fehler des k -ten Knoten in der $l+1$ -ten Schicht rückwärts durch das Netzwerk gegeben und ergibt so den Fehler der Ausgabe der l -ten Schicht.

Die dritte fundamentale Gleichung bezieht sich auf die Änderungsrate der Verlustfunktion f in Bezug auf eine Bias b_j^l im Netz:

$$\frac{\partial f}{\partial b_j^l} = \Delta_j^l \quad (37)$$

Die Änderungsrate der Verlustfunktion bezüglich eines beliebigen Gewichtes w_{jk}^l im Netz lässt sich durch die vierte und letzte elementare Gleichung berechnen:

$$\frac{\partial f}{\partial w_{jk}^l} = \frac{\partial f}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{jk}^l} = \Delta_j^l a_k^{l-1} \quad (38)$$

Diese vier Gleichungen werden im Algorithmus in folgender Reihenfolge angewendet. Zuerst werden die Aktivierungen aller Knoten der Eingabe-Schicht gemäß Gleichung

32 berechnet. Dann wird ebenfalls gemäß Gleichung 32 die Aktivierungen der weiteren Schichten berechnet, wobei hier vorwärts von Schicht $l - 1$ zu Schicht l gearbeitet wird. Daraufhin wird der Fehler der Ausgabe Δ^L mithilfe von Gleichung 34 bestimmt. Nun kommt es zum vierten Schritt, dem Rückpropagieren der Fehler. Für $l = L, L - 1, L - 2, \dots$ werden nun die Fehler Δ^l berechnet (siehe Gleichung 36). Im letzten Schritt wird dann der Gradient mittels Gleichung 37 und 38 ermittelt [10].

4.3 Minimierung der Verlustfunktion mit Adam-Optimierer

In der Praxis wird der Rückpropagations-Algorithmus üblicherweise mit einem Lernalgorithmus kombiniert. Dabei werden Algorithmen, die auf der Methode des Gradientenabstiegs basieren, am häufigsten verwendet. Beim Algorithmus des sogenannten 'Mini-Batch Gradientenabstiegs' wird der gesamte Trainingsdatensatz in n 'Mini-Batches', also in n Untermengen der Gesamtdaten, eingeteilt. Hierfür wird zunächst ein Trainingsdatensatz ausgewählt. Dieser ist ein Datensatz von Beispielen, der während des Lernprozesses verwendet wird und dazu dient, die Parameter (z. B. Gewichte) eines Klassifizierers anzupassen

Nach Durchlaufen einer Untermenge, werden dann die Parameter \vec{w} des Modells, die die Verlustfunktion f minimieren, entgegen der Richtung des Gradienten der Verlustfunktion $\nabla_{\vec{w}} f(\vec{w})$, aktualisiert

$$\vec{w} = \vec{w} - \eta \cdot \nabla_{\vec{w}} f(\vec{w}) \quad (39)$$

und der Prozess wird für die darauffolgenden Batches wiederholt. Hierbei ist η die Lernrate. Diese bestimmt die Schrittgröße, mit welcher das Netz in Richtung des lokalen Minimums der Verlustfunktion geht und somit, wie stark das Netz die Gewichtung einzelner Knoten adaptiert. Problematisch hierbei ist hauptsächlich eine optimale Wahl der Lernrate. Ist sie zu groß, wird das lokale Minimum übersprungen und es kommt zu einer Oszillation um das Minimum herum. Eine zu kleine Lernrate hingegen benötigt sehr viel Rechenzeit. Daher gibt es einige an dieses Problem angepasste Algorithmen.

Im Rahmen dieser Arbeit wird zur Minimierung der Verlustfunktion ausschließlich der Adam-Optimierer [47] verwendet. Die grundlegende Idee hierbei ist es, adaptive Lernraten für jeden Parameter zu berechnen, indem ein exponentiell abklingender Mittelwert vergangener Gradienten und vergangener quadrierten Gradienten gespeichert wird [48]. Ein wichtiger Hyperparameter des Algorithmus ist die Batch-Größe. Sind alle Batches durchlaufen, so hat der Lernalgorithmus den gesamten Trainingsdatensatz durchgearbeitet, was als Epoche bezeichnet wird. Die Anzahl der Epochen ist ein weiterer Hyperparameter und ist in der Regel oft im Größenbereich von Hunderten. Nach Durchlaufen aller Epochen ist der Trainingsprozess abgeschlossen [49].

4.4 Early-Stopping-Algorithmus

Die Epochen-Anzahl ist ein Hyperparameter, der die Leistung des Netzes erheblich beeinflussen kann. Zu wenige Epochen können dazu führen, dass das Netz komplexere Strukturen nicht lernen kann und zu viele Epochen können zu Übertraining führen. Übertraining beschreibt ein unerwünschtes Verhalten eines maschinellen Lernalgorithmus. Hierbei erzielt das Modell eine signifikant höhere Leistung auf dem Trainingsdatensatz als auf dem Testdatensatz, wodurch das Modell somit nicht übertragbar und anwendbar auf neue, unbekannte Daten ist [10]. Um die richtige Wahl der Epochenanzahl zu erleichtern, wird

der *Early-Stopping*-Algorithmus beim Trainieren eines neuronalen Netzwerkes verwendet. Dieser stoppt den Trainingsprozess, wenn keine Verbesserung mehr in der Leistung des Modells bezüglich des Testdatensatzes beobachtet wird. Es gibt zwei wichtige Hyperparameter, `min_delta` und `patience`, die vor dem Training festgelegt werden können. Der `min_delta` Parameter legt fest, wie groß eine absolute Änderung der Verlustfunktion mindestens sein muss, damit diese Änderung als Verbesserung eingestuft wird. Liegt der Wert der Änderung unter dem Wert von `min_delta`, so wird das Training abgebrochen, sofern der `patience`-Parameter auf Null gesetzt ist. Dieser beschreibt die Anzahl der Epochen, die weiter trainiert werden, obwohl der Algorithmus keine Verbesserung mehr beobachten kann. Der *Early-Stopping*-Algorithmus ist in der frei zugänglichen Keras [50]-Bibliothek implementiert. Dieser Bibliothek liegt die 'Tensorflow'[51]-Bibliothek zugrunde, die Implementierungen von Netzwerk-Architektur, Verlustfunktionen, Optimierern und mehr enthält.

4.5 Anpassung der Hyperparameter mit OPTUNA

Die Wahl des optimalen Aufbaus eines neuronalen Netzes ist in vielen Fällen problemspezifisch und erfordert umfangreiche numerische Analysen. Die Leistung des Netzes wird nicht nur durch die Anzahl der versteckten Schichten, sondern auch durch die Anzahl der Knoten in jeder Schicht beeinflusst.

OPTUNA [11], [52] bietet die Möglichkeit nach den optimalen Hyperparametern automatisiert zu suchen. Es wird eine Zielfunktion definiert, die minimiert oder maximiert werden soll und verschiedene Architekturen des zu optimierenden neuronalen Netzes enthält. Sie spannt also den Parameter-Suchraum auf, welcher dynamisch konstruiert werden kann, indem Vorschläge für verschiedene Parameter in die Zielfunktion eingebaut werden können. So gibt es für alle Arten von Hyperparametern angepasste Optionen. Beispielsweise können Fließkomma-Parameter mit `trials.suggest_uniform`, ganze Zahlen mit `trials.suggest_int` und kategoriale Parameter mit `trials.suggest_categorical` vorgeschlagen werden. Eine einzelne Bewertung der Zielfunktion wird Versuch genannt, während der gesamte Optimierungsprozess als Studie bezeichnet wird. Bei OPTUNA basiert die Optimierung auf einem Bayesianischen Ansatz [53]. Dies bedeutet, dass im Gegensatz zur Zufalls- oder Rastersuche vergangene Bewertungsergebnisse aufgezeichnet werden, die dann zur Bildung eines probabilistischen Modells - ein sogenanntes Surrogats - verwendet werden. Zweck des Surrogats ist es, die Zielfunktion zu 'ersetzen', da das Surrogat deutlich einfacher als die Zielfunktion zu optimieren ist. Grundsätzlich funktioniert der Optimierungsprozess folgendermaßen: Zuerst wird ein Surrogat erstellt, woraufhin Hyperparameter ermittelt werden, die für die Surrogatfunktion am besten geeignet sind. Anschließend erfolgt eine Bewertung der wahren Zielfunktion mit diesen Hyperparametern, worauf eine Aktualisierung des Surrogatmodells mit den neuen Ergebnissen erfolgt. Der Prozess der Hyperparameter-Ermittlung bis hin zur Aktualisierung des Surrogatmodells wird solange wiederholt, bis die maximale Anzahl an Versuchen oder die maximale Zeit erreicht ist.

5 Der analysierte Datensatz

In diesem Kapitel wird der in dieser Arbeit analysierte Datensatz vorgestellt. Hierbei werden auf die wichtigsten Merkmale des Datensatzes und auf die Skalierung der Ereignisse auf eine integrierte Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ eingegangen.

5.1 Merkmale des Datensatzes

Im Rahmen dieser Arbeit wird mit Monte-Carlo-simulierten Ereignissen gearbeitet. Daher steht, im Gegensatz zu echten Daten, die im ATLAS-Detektor gemessen wurden, für jedes Ereignis der wahre Wert der Zielvariablen zu Verfügung. Von der Generation der Ereignisse bis hin zu dem bereitgestellten Datensatz, wurden bereits Vorselektionen angewendet. Die Details dazu sind in Kapitel 6 erläutert. Für den Datensatz wurden nur Ereignisse des Signalprozesses (VBF $H \rightarrow \tau\tau \rightarrow e\mu + 4\nu$) generiert und simuliert.

Die Daten wurden in Form eines 'Pandas DataFrame' [54] zu Verfügung gestellt. Hierbei handelt es sich um eine zweidimensionale Tabelle, wobei eine Zeile im DataFrame ein Ereignis darstellt und eine Spalte mit den Werten einer Variable des Datensatzes besetzt ist. Im gesamten Datensatz stehen 115 Variablen zu Verfügung. Die Anzahl der Zeilen bzw. Ereignisse hängt von den angewandten Selektionskriterien ab, die in Kapitel 6 diskutiert werden.

Zu den wichtigsten Merkmalen des Datensatzes gehören die Komponenten des Viererimpulses des Higgs-Bosons auf Wahrheitsniveau $p_{H,\text{truth}}^\mu$. Diese sind in Polarkoordinaten (p_T, η, ϕ, M) gegeben, wobei eine Masse von 125 GeV generiert wurde. Des Weiteren sind die Viererimpulse der zwei Partonen p_0 und p_1 auf Parton-Niveau der VBF-Produktion gegeben. Diese Partonen hadronisieren zu Jets j_0 und j_1 , deren Viererimpulse auf Wahrheits- und Rekonstruktionsniveau bezüglich des Hadron-Niveaus ebenfalls im Datensatz enthalten sind. Auch die Zerfallsprodukte des Higgs-Bosons auf Wahrheits-, sowie auf Rekonstruktionsniveau, sind gegeben. Dazu zählen die Viererimpulse der zwei Leptonen ℓ_0 und ℓ_1 , sowie die Viererimpulse der Neutrinos, die aber aufgrund ihrer geringen Wechselwirkung nicht direkt im Detektor nachgewiesen werden können. Jedoch gilt unter der Annahme, dass der Transversalimpuls der Partonen in den Kollisions-Protonen im Anfangszustand Null ist, in der Transversalebene des Detektors Impulserhaltung. So kann der Transversalimpuls der Neutrinos berechnet werden, indem die Transversalimpulse aller nachgewiesenen Teilchen vektoriell aufsummiert werden. Das Ergebnis ist von Null verschieden, sodass dieser fehlende Transversalimpuls E_T^{miss} in einem rekonstruierten Ereignis der Summe der vier Neutrinos zugeordnet wird:

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2} \quad (40)$$

Unter diesen Bedingungen lässt sich der Transversalimpuls des Higgs-Bosons im Fall von $H \rightarrow \tau\tau$ durch einfache vektorielle Aufaddierung der Zerfallsteilchen der τ -Leptonen mittels

$$\vec{p}_{T,\text{add}}(H) = \vec{p}_T(\ell_0) + \vec{p}_T(\ell_1) + \vec{p}_T(E_T^{\text{miss}}) \quad (41)$$

rekonstruieren. Hierbei ist $\vec{p}_T(\ell_0)$ der Transversalimpuls des führenden Leptons und $\vec{p}_T(\ell_1)$ der Transversalimpuls des zweitführenden Leptons. Da die fehlende Transversalenergie nur in der x - y -Ebene gemessen wird, gibt es keine η - und M -Komponente des rekonstruierten Viererimpulses. Die Methode des einfachen Aufaddierens der Zerfallsteilchen der τ -Leptonen liefert also keinen vollständigen Viererimpuls des Higgs-Teilchens.

Diese fehlende Information wird durch den sogenannten *Missing Mass Calculator* (MMC) [12] aufgefüllt. Die mit dem MMC rekonstruierten Viererimpulse des Higgs-Bosons sind ebenfalls im Datensatz enthalten. Der Algorithmus des Missing Mass Calculator wird verwendet, um die Kinematik von Ereignissen im $\tau\tau$ -Zerfall zu rekonstruieren. Hierfür benötigt der Algorithmus im rein leptonischen Zerfallskanal Informationen von acht Variablen: die Dreierimpulse und die invarianten Massen der zwei τ -Neutrinos. Zur Berechnung dieser steht ein Gleichungssystem, das aus nur vier Gleichungen besteht, zu Verfügung. Folglich hat das Gleichungssystem keine exakte Lösung. Jedoch sind die verschiedenen Lösungen unter Berücksichtigung der Ereigniskinematik nicht gleich wahrscheinlich. Über die bekannten Wahrscheinlichkeitsdichtefunktionen der τ -Zerfallskinetiken wird dann die wahrscheinlichste Lösung bestimmt. Die Qualität der durch den MMC rekonstruierten Higgs-Boson-Viererimpulse wird im Laufe dieser Arbeit mit Higgs-Boson-Viererimpulsen verglichen, die durch den Einsatz maschinellen Lernens rekonstruiert wurden.

5.2 Skalierung der Ereignisse

Die Datennahmeperiode 2 (*Run 2*) lieferte einen Datensatz von Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 13 TeV, welcher einer integrierten Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ entspricht. Um die in dieser Arbeit verwendeten Monte-Carlo-simulierten Ereignisse in Relation zu der gesammelten Datenmenge des Run 2 zu setzen, müssen die simulierten Ereignisse skaliert werden, sodass man die Anzahl der erwarteten Ereignissen $N_{\text{exp}}^{\mathcal{L}}$ in 139 fb^{-1} erhält. Hierfür muss die Anzahl der simulierten Ereignisse N_{sim} mit einem Faktor gewichtet werden. Zu diesem Faktor gehört die inverse Anzahl aller Ereignisse N_{all} , die generiert wurden. Hinzu kommen die integrierte Luminosität \mathcal{L} , der Wirkungsquerschnitt des Signalprozesses x_{sec} , der K -Faktor des Generator-Filters und die zugehörige Filter-Effizienz ϵ_{filter} :

$$N_{\text{exp}}^{\mathcal{L}} = \frac{N_{\text{sim}}}{N_{\text{all}}} \cdot \mathcal{L} \cdot x_{\text{sec}} \cdot K \cdot \epsilon_{\text{filter}} \quad (42)$$

Hierbei beinhaltet die Variable x_{sec} den Wirkungsquerschnitt des VBF-Produktionsprozesses σ_{VBF} , das Verzweigungsverhältnis des $H \rightarrow \tau\tau$ Zerfalls und das Verzweigungsverhältnis des $H \rightarrow e\mu 4\nu$ Zerfalls für eine Higgs-Boson-Masse von $m_{\text{H}} = 125,09 \text{ GeV}$.

Die Werte, die für die Berechnung des Gewichtungsfaktors verwendet wurden, sind in Tabelle 4 zusammengefasst. Die Anzahl an simulierten Ereignissen N_{sim} ändert sich bei Anwenden der in Kapitel 6 beschriebenen Selektion und wird hier daher nicht explizit angegeben.

| | |
|---|----------|
| Anzahl Ereignisse vor Generator-Filter N_{all} | 12093522 |
| Luminosität $\mathcal{L} [\text{fb}^{-1}]$ | 139 |
| Wirkungsquerschnitt des Signalprozesses $x_{\text{sec}} [\text{fb}^{-1}]$ | 236,41 |
| K -Faktor | 1 |
| Filter-Effizienz $\epsilon_{\text{filter}} [\%]$ | 5,7926 |

Tabelle 4: Übersicht der Werte, die man für die Skalierung der Monte-Carlo-Ereignisse auf eine integrierte Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ benötigt. Der Wert für den Wirkungsquerschnitt und die $H \rightarrow \tau\tau$ Verzweigungsverhältnisse sind aus [25] entnommen.

6 Selektion der Signalereignisse

Die in dieser Analyse verwendeten Monte-Carlo-simulierten Ereignisse des Signalprozesses wurden mit den POWHEG[55][56][57][58]- und PYTHIA8[59]-Generatoren erzeugt. In diesem Kapitel werden die einzelnen Schritte zur Selektion dieser Ereignisse erläutert. Mit der Objektselektion werden bestimmte Anforderungen an die Objekt-Kandidaten gestellt, die im Signalprozess enthalten sind. Mit den definierten Objekten werden dann in der Vorselektion weitere Forderungen definiert, welche Ereignisse für eine weitere Betrachtung in der Analyse erfüllen müssen. Schließlich erfolgen in der finalen Selektion aus technischen Gründen weitere Schnitte, sowie ein Abgleich der Jets auf Rekonstruktions- und Wahrheitsniveau.

6.1 Objektselektion

Für die Objektselektion werden unterschiedliche Anforderungen an die rekonstruierten Objekte gestellt, um die Wahrscheinlichkeit den richtigen Teilchenkandidaten auszuwählen zu erhöhen. Hierfür wird gefordert, dass der Transversalimpuls des aus dem Tau-Zerfall stammenden Elektrons größer als 15 GeV und der des Myons größer als 10 GeV sein soll. Der Betrag der η -Koordinate eines Myons oder Elektrons soll außerdem maximal 2,5 betragen. Des Weiteren wird gefordert, dass der Transversalimpuls des führenden Jets, welcher der Jet mit dem höchsten Transversalimpuls ist, größer als 40 GeV und der des zweitführenden Jets, welcher der Jet mit dem zweithöchsten Transversalimpuls ist, größer als 20 GeV sein ist. Bezüglich der η -Koordinate des führenden und zweitführenden Jets ist gefordert, dass der Betrag dieser maximal 4,5 beträgt. Der Transversalimpuls der fehlenden Transversalenergie soll außerdem > 20 GeV sein. In Tabelle 5 sind die Schnitte zur Objektselektion zusammengefasst.

Mittels des *particle-flow*-Algorithmus [60] werden Jetkandidaten, sogenannte *particle-flow*-Objekte, gebildet, die dann mittels des Anti- k_T -Algorithmus [61] mit einem Abstandsparameter von $R = 0,4$ zu Jets rekonstruiert werden.

| | p_T -Schnitt | $ \eta $ |
|--|----------------|----------|
| Elektronen e | > 15 GeV | $< 2,5$ |
| Myonen μ | > 10 GeV | $< 2,5$ |
| Führender Jet j_1 | > 40 GeV | $< 4,5$ |
| Zweitführender Jet j_2 | > 20 GeV | $< 4,5$ |
| Fehlende Transversalenergie E_T^{miss} | > 20 GeV | - |

Tabelle 5: Selektionsschnitte auf den Transversalimpuls p_T und die Pseudorapidität η für die Elektronen, Myonen, Jets und der fehlenden Transversalenergie E_T^{miss} . Für E_T^{miss} gibt es keinen Schnitt auf η , da E_T^{miss} nur in der Transversalebene (x - y -Ebene) gemessen wird und es somit keine η -Komponente von E_T^{miss} gibt.

6.2 Vorselektion

Um den Datensatz zu reduzieren und Speicherkapazitäten zu entlasten, wurden Vorselektions-Kriterien auf den Datensatz angewendet. In folgender Tabelle 6 ist eine Übersicht der verschiedenen Schnitte bzw. Kriterien und die jeweils zugehörige Anzahl

der simulierten Ereignisse N_{sim} , sowie die Anzahl der Ereignisse, die man bei einer Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ erwartet N_{exp} , dargestellt.

| Name des Schnitts | N_{sim} | N_{exp} bei 139 fb^{-1} |
|------------------------------------|------------------|--|
| All events | 1240754 | 195,29 |
| Any Trigger | 1240754 | 195,29 |
| Veto hadronic decaying Taus | 1231972 | 193,91 |
| NLep = 2 | 1231972 | 193,91 |
| Different flavour | 616405 | 97,02 |
| Leading lep ID cut | 614661 | 96,75 |
| Sublead lep ID cut | 610549 | 96,10 |
| Leading lep iso cut | 610549 | 96,10 |
| Sublead lep iso cut | 610549 | 96,10 |
| Trigger matching cut | 610549 | 96,10 |
| Veto SS evt | 610549 | 96,10 |
| MET cut | 542335 | 85,36 |
| Veto 0-jet events | 542335 | 85,36 |
| Leading jet $p_T > 40 \text{ GeV}$ | 498493 | 78,46 |
| $\Delta R(e, \mu) < 2,5$ | 375727 | 59,14 |
| $\Delta \eta(e, \mu) < 1,5$ | 342886 | 53,97 |
| Final selection | 276074 | 43,45 |

Tabelle 6: Die Tabelle zeigt die Schnitte der Vorselektion; Für jeden Schnitt wird die Anzahl der Monte-Carlo-simulierten Ereignisse N_{sim} nach Anwenden des Schnittes, sowie die Anzahl der erwarteten Ereignisse N_{exp} bei einer Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$, angegeben.

Mit dem Schnitt **Any Trigger** werden diverse Kombinationen von Ein-Lepton-Triggern und Zwei-Lepton-Triggern gefordert. Da in dieser Arbeit der dileptonische τ -Zerfall betrachtet wird, soll es keine hadronisch zerfallenden τ -Leptonen geben, was durch den Schnitt **Veto hadronic decaying Taus** sichergestellt wird. **NLep** hingegen gibt die Anzahl an leptonisch zerfallenen Tau-Leptonen an, die, dem Signalprozess entsprechend, gleich zwei ist. Weiterhin wird mit **Different flavour** verlangt, dass die τ -Leptonen in zwei Leptonen verschiedener Art zerfallen. So wird der Hauptuntergrund von $Z \rightarrow ee, Z \rightarrow \mu\mu$, der ein τ -Zerfallskanal mit gleicher Leptonen-Art zufolge hätte, unterdrückt. Der **Leading lep ID cut** und **Sublead lep ID cut** legt fest, unter welchen Qualitätskriterien das führende und das zweitführende Lepton als solche akzeptiert werden. Das führende Lepton ist das Lepton mit dem höchsten Transversalimpuls p_T , von dem mindestens das lockere (*loose*) [62] [63] Identifikationskriterium gefordert wird. Das zweitführende Lepton hingegen ist das Lepton mit dem zweithöchsten Transversalimpuls p_T und muss ebenfalls mindestens ein lockeres Identifikationskriterium erfüllen. Um zu verhindern, dass die Leptonen mit z.B. Jets verwechselt werden, werden die Isolations-Schnitte **Leading lep iso cut** und **Sublead lep iso cut** angewendet. Für die Elektronen e und Myonen μ werden unterschiedliche Qualitätskriterien gefordert, die in folgender Tabelle eingesehen werden können.

| | Qualitätskriterium |
|---------------------|--|
| Leading lep ID cut | <i>loose</i> |
| Sublead lep ID cut | |
| Leading lep iso cut | für e : <i>fc loose</i> |
| Sublead lep iso cut | für μ : <i>fc tight track only</i> |

Tabelle 7: Übersicht der für die Leptonen-Selektion geforderten Qualitätskriterien bezüglich der Identifikation [62] [63] und Isolation.

Der **Trigger matching cut** stellt sicher, dass die Zerfalls-Leptonen mit dem größten Transversalimpuls den Trigger aktiviert haben. Außerdem gibt es ein Veto auf Ereignisse, die zwei τ -Leptonen mit gleichem Vorzeichen bezüglich der elektrischen Ladung enthalten (**Veto SS evt**). Da das Higgs-Teilchen elektrisch neutral ist, müssen die τ -Leptonen unter Berücksichtigung der Ladungserhaltung entgegengesetzt geladen sein. Der Schnitt **MET cut** auf die in Kapitel 5.1 eingeführte fehlende Transversalenergie wurde bereits in der Objektselektion aufgeführt und liegt bei $E_T^{miss} > 20\text{GeV}$. Ebenfalls werden durch den Schnitt **Veto 0-jet events** Ereignisse verworfen, bei denen kein Jet existiert, wobei der führende Jet einen Transversalimpuls von über 40 GeV haben muss. Es werden nur Ereignisse akzeptiert, für die der Winkelabstand ΔR (siehe Gleichung 28) zwischen dem Elektron e und Myon μ kleiner als 2,5 ist und für die der Abstand bezüglich der η -Koordinate, wie sie in Gleichung 25 definiert ist, maximal 1,5 beträgt. Für weitere Details sei auf die Publikation [31] verwiesen.

6.3 Finale Selektion

Zusätzlich zu den Anforderungen, die bereits an den ursprünglichen Datensatz gestellt wurden, ist es sinnvoll, weitere Schnitte vorzunehmen, um eine bestimmte Daten-Qualität zu gewährleisten. Aus technischen Gründen ist es notwendig, den Schnitt aus der Objektselektion, dass mindestens zwei Jets in jedem Ereignis existieren, zu wiederholen. In Tabelle 8 ist zu sehen, dass ca. 47000 der simulierten Ereignisse bei der Forderung $p_T(j_1) > 20\text{ GeV}$ wegfallen. Das liegt daran, dass bei diesen ca. 47000 Ereignissen kein zweiter Jet j_1 gefunden werden konnte, der diese p_T -Schwelle übersteigen konnte. Auch die Forderung, dass mindestens zwei Leptonen existieren, wobei das (zweit)führende Lepton im Fall des Elektron einen Transversalimpuls von größer als 15 GeV haben muss und im Fall des Myons einen Transversalimpuls von größer als 10 GeV haben muss, wird erneut gestellt. Hier blieb die Anzahl an Ereignissen unverändert.

Zusätzlich zu den gerade beschriebenen Schnitten erfolgt ein Abgleich der rekonstruierten Jets zu den Jets auf Wahrheitsniveau. Hierfür wird verlangt, dass der Winkelabstand ΔR zwischen dem führenden Jet auf Hadron-Rekonstruktions-Niveau j_0 und dem führenden Jet auf Hadron-Wahrheits-Niveau $j_{0,\text{truth}} < 0,4$ ist. Dasselbe gilt für die zweitführenden Jets j_0 und $j_{0,\text{truth}}$. In Tabelle 8 sind die Schnitte der finalen Selektion zusammengefasst. Mit den $N_{\text{sim}} = 200623$ Signal-Ereignissen, die am Ende der finalen Selektion vorhanden sind, werden die neuronalen Netze trainiert.

| Schnitte | N_{sim} | N_{exp} in 139 fb^{-1} |
|---|------------------|---|
| kein zusätzlicher Schnitt | 276067 | 43,45 |
| $p_{\text{T}}(j_0) > 40 \text{ GeV}$ | 276067 | 43,45 |
| $p_{\text{T}}(j_1) > 20 \text{ GeV}$ | 228917 | 36,03 |
| $p_{\text{T}}(\ell_0) > 15 \text{ GeV}$ für e bzw. $> 10 \text{ GeV}$ für μ | 228917 | 36,03 |
| $p_{\text{T}}(\ell_1) > 15 \text{ GeV}$ für e bzw. $> 10 \text{ GeV}$ für μ | 228917 | 36,03 |
| $\Delta R(j_0, j_{0,\text{truth}}) < 0,4$ | 228163 | 35,91 |
| $\Delta R(j_1, j_{1,\text{truth}}) < 0,4$ | 200623 | 31,58 |

Tabelle 8: Zusammenfassung der zusätzlich angewandten Schnitte auf die Ereignisse; N_{sim} beschreibt die Anzahl der simulierten Ereignisse nach Anwenden der Schnitte und N_{exp} beschreibt die Anzahl der bei $\mathcal{L} = 139 \text{ fb}^{-1}$ erwarteten Ereignisse.

7 Rekonstruktion des Higgs-Boson-Viererimpulses mittels Regressionsnetzwerken

Für die Berechnung der CP-ungeraden Optimalen Observablen werden als Eingangsgrößen, neben den Viererimpulsen der aus der VBF-Produktion stammenden Jets, der Viererimpuls des Higgs-Bosons benötigt. Folglich ist die Auflösung der Optimalen Observablen durch die Rekonstruktionsgüte des Higgs-Boson-Viererimpulses begrenzt, was im Unterkapitel 7.2 gezeigt wird. Um die Sensitivität für CP-Verletzung der Optimalen Observablen zu erhöhen, ist also eine präzise Rekonstruktion des Higgs-Boson-Viererimpulses wünschenswert. In diesem Kapitel wird untersucht, inwieweit der Einsatz von Methoden des maschinellen Lernens, die Rekonstruktionsqualität des Viererimpulses verbessert. Hierfür wurden im Rahmen dieser Arbeit neuronale Regressions-Netzwerke verwendet. Um die Rekonstruktionsgüte dieser Methode beurteilen zu können, werden dessen Ergebnisse mit denen von anderen Rekonstruktion-Methoden verglichen.

Im ersten Abschnitt dieses Kapitels werden vier verschiedene Netze mit unterschiedlichen Eingangs-Größen und Strukturen miteinander verglichen. Hierfür wird auch auf die Ergebnisse der OPTUNA-Optimierungen näher eingegangen. Im darauffolgenden Abschnitt erfolgt ein Vergleich der Auflösungen der Higgs-Boson-Viererimpulse, welche mit den neuronalen Netzen, dem MMC und der Methode der vektoriellen Aufaddierung der Zerfallsprodukte des Di-Tau-Systems rekonstruiert wurden. Schließlich wird in den letzten beiden Unterkapitel die Optimale Observable der verschiedenen Rekonstruktionsmethoden berechnet und deren Auflösung miteinander verglichen.

7.1 Vergleich von vier verschiedenen neuronalen Netzen

Im Rahmen dieser Arbeit wurden vier verschiedene neuronale Netze trainiert und für die Vorhersage des Viererimpulses des Higgs-Teilchens verwendet. Inwiefern sich die verschiedenen Modelle unterscheiden wird im folgenden Kapitel im Detail erläutert.

7.1.1 Neuronales Netz 1 - kartesische Koordinaten

In die Eingabe-Schicht des ersten Netzwerkes NN_1 gehen Werte von zehn Variablen ein. Zu diesen gehören die kartesischen Viererimpulse der beiden Leptonen, sowie die p_x - und p_y -Komponente der fehlenden Transversalenergie E_T^{miss} . Da bei Proton-Proton-Kollisionen der longitudinale Impuls der wechselwirkenden Partonen nicht bekannt ist, gibt es keine p_z -Komponente von E_T^{miss} . Als Ziel-Variablen werden die Komponenten des Viererimpulses des wahren Higgs-Teilchens in kartesischen Koordinaten angegeben. Da dieser Viererimpuls im ursprünglichen Datensatz für jedes Ereignis in Polarkoordinaten (p_T, η, ϕ, M) gegeben ist, wurde der Viererimpuls in kartesische Koordinaten transformiert. Dasselbe gilt für alle Eingangs-Variablen. Zusammenfassend sind in Tabelle 9 die Ein- und Ausgabe-Variablen dargestellt.

| Eingangs-Variablen | Beschreibung | |
|---------------------------|-----------------------|----------------------------------|
| $p_x(\ell_0)$ | x -Impulskomponente | des führenden Leptons |
| $p_y(\ell_0)$ | y -Impulskomponente | |
| $p_z(\ell_0)$ | z -Impulskomponente | |
| $E(\ell_0)$ | Energie | |
| $p_x(\ell_1)$ | x -Impulskomponente | des zweitführenden Leptons |
| $p_y(\ell_1)$ | y -Impulskomponente | |
| $p_z(\ell_1)$ | z -Impulskomponente | |
| $E(\ell_1)$ | Energie | |
| $p_x(E_T^{miss})$ | x -Impulskomponente | der fehlenden Transversalenergie |
| $p_y(E_T^{miss})$ | y -Impulskomponente | |
| Ziel-Variablen | | |
| $p_x(H, \text{truth})$ | x -Impulskomponente | des wahren Higgs Teilchens |
| $p_y(H, \text{truth})$ | y -Impulskomponente | |
| $p_z(H, \text{truth})$ | z -Impulskomponente | |
| $E(H, \text{truth})$ | Energie | |

Tabelle 9: Übersicht der Eingangs- und Zielvariablen des ersten neuronalen Netzes NN_1 (Kartesische Koordinaten). Die Verteilungen der Eingangs-Variablen sind im Anhang B.1 in Abbildung 22 bis 24 dargestellt.

7.1.2 Neuronales Netz 2 - kartesische Koordinaten mit Jets

Das zweite neuronale Netz enthält die gleichen Eingangs-Variablen wie das erste neuronale Netz, jedoch werden zusätzlich zu den Viererimpulsen der beiden Leptonen, sowie der p_x - und p_y -Komponente der fehlenden Transversalenergie E_T^{miss} noch die Viererimpulse der beiden aus der Vektorbosonfusion stammenden Jets hinzugefügt. Das Netz hat also acht Eingangs-Variablen mehr, als das Netz zuvor. Dies ist auch in Tabelle 10 einzusehen. Die Ziel-Variablen sind, wie bereits bei dem ersten Netz, die kartesischen Komponenten des Viererimpulses des wahren Higgs-Teilchens.

| Eingangs-Variablen | Beschreibung | |
|--|--|----------------------------------|
| $p_x(\ell_0)$ $p_y(\ell_0)$ $p_z(\ell_0)$ $E(\ell_0)$ | x -Impulskomponente y -Impulskomponente z -Impulskomponente Energie | des führenden Leptons |
| $p_x(\ell_1)$ $p_y(\ell_1)$ $p_z(\ell_1)$ $E(\ell_1)$ | x -Impulskomponente y -Impulskomponente z -Impulskomponente Energie | des zweitführenden Leptons |
| $p_x(j_0)$ $p_y(j_0)$ $p_z(j_0)$ $E(j_0)$ | x -Impulskomponente y -Impulskomponente z -Impulskomponente Energie | des führenden Jets |
| $p_x(j_1)$ $p_y(j_1)$ $p_z(j_1)$ $E(j_1)$ | x -Impulskomponente y -Impulskomponente z -Impulskomponente Energie | des zweitführenden Jets |
| $p_x(E_T^{miss})$ $p_y(E_T^{miss})$ | x -Impulskomponente y -Impulskomponente | der fehlenden Transversalenergie |
| Ziel-Variablen | | |
| $p_x(H, \text{truth})$ $p_y(H, \text{truth})$ $p_z(H, \text{truth})$ $E(H, \text{truth})$ | x -Impulskomponente y -Impulskomponente z -Impulskomponente Energie | des wahren Higgs Teilchens |

Tabelle 10: Übersicht der Eingangs- und Zielvariablen des zweiten neuronalen Netzes NN_2 (Kartesische Koordinaten mit Jets). Die dazugehörigen Input-Verteilungen sind im Anhang B.2 in Abbildung 22 bis 26 gezeigt.

7.1.3 Neuronales Netz 3 - Polarkoordinaten

Während das erste und das zweite Netz auf kartesischen Koordinaten basieren, soll das dritte Netz NN_3 in Polarkoordinaten trainiert werden. Die 2π -Symmetrie der ϕ -Koordinate des ATLAS-Detektors kann ausgenutzt werden, indem das gesamte Koordinatensystem so gedreht wird, dass $\phi_{\text{rot}}(E_T^{miss}) = 0$. Das hat den Vorteil, dass die Variable der fehlenden Transversalenergie $\phi(E_T^{miss})$ als Eingangs-Variable im Training entfällt. So verläuft der Trainingsprozess schneller ab. Außerdem sind die relevanten ϕ -Winkel durch die Drehung zu einem gemeinsamen Nullpunkt hin, um die Null verteilt, was dem Netz das Lernen der Daten erleichtert. Die genauen Schritte der Koordinatensystem-Drehung von ϕ zu ϕ_{rot} kann mit folgenden Formel nachvollzogen werden:

$$\phi_{\text{rot}}(E_T^{miss}) = \phi(E_T^{miss}) - \phi(E_T^{miss}) = 0 \quad (43)$$

$$\phi_{\text{rot}}(\ell_i) = \phi(\ell_i) - \phi(E_T^{miss}) \quad (44)$$

Hierbei beschreibt ℓ_i die das i -te Lepton.

Des Weiteren galt es, das 'Sprungproblem' der ϕ -Koordinate zu beachten, das aus der Doppeldeutigkeit der, physikalisch gesehen ähnlichen, Werte von knapp 2π und 0 folgt. Um dieses Problem zu unterbinden, wurden der $\sin \phi_{\text{rot}}$ und der $\cos \phi_{\text{rot}}$ anstelle des

rotierten Winkels ϕ_{rot} selbst in den Eingangs-Variablen verwendet. In den Ziel-Variablen hingegen ist dies nicht der Fall, da so eine lückenlose Rücktransformation des ϕ -Winkels möglich ist. In folgender Tabelle sind die Eingangs- und Ausgangsvariablen des Netzes zusammengefasst.

| Eingangs-Variablen | Beschreibung | |
|---|---|----------------------------|
| $p_T(\ell_0)$ $\eta(\ell_0)$ $\sin(\phi_{\text{rot}}(\ell_0))$ $\cos(\phi_{\text{rot}}(\ell_0))$ | fehlender Transversalimpuls Pseudorapidity Sinus des rotierten ϕ -Winkels Cosinus des rotierten ϕ -Winkels | des führenden Leptons |
| $p_T(\ell_1)$ $\eta(\ell_1)$ $\sin(\phi_{\text{rot}}(\ell_1))$ $\cos(\phi_{\text{rot}}(\ell_1))$ | fehlender Transversalimpuls Pseudorapidity Sinus des rotierten ϕ -Winkels Cosinus des rotierten ϕ -Winkels | des zweitführenden Leptons |
| $p_T(E_T^{\text{miss}})$ | Transversalimpuls der fehlenden Transversalenergie | |
| Ziel-Variablen | | |
| $p_T(H, \text{truth})$ $\eta(H, \text{truth})$ $\phi_{\text{rot}}(H, \text{truth})$ $M(H, \text{truth})$ | fehlender Transversalimpuls Pseudorapidity Rotierter Azimuth-Winkel Masse | des wahren Higgs Teilchens |

Tabelle 11: Übersicht der Eingangs- und Zielvariablen des dritten neuronalen Netzes NN_3 (Polarkoordinaten). Die jeweiligen Verteilungen der Eingangs-Variablen sind im Anhang B.3 in Abbildung 27 bis 29 dargestellt.

7.1.4 Neuronales Netz 4 - Polarkoordinaten mit Jets

Für das letzte Netz, das im Rahmen dieser Arbeit vorgestellt wird, sind die Eingangs- und Ausgabevariablen, wie im vorherigen Netz, in Polarkoordinaten gegeben. Es hat die gleichen Eingangs-Variablen wie das dritte Netz und zusätzlich die entsprechenden Koordinaten der Jets. Auch hier wird eine Rotation des Koordinatensystems bezüglich den ϕ -Koordinaten der Eingangs- und Ausgangsgrößen gemäß Formeln 43, 44 vorgenommen, wobei für die Rotation der i Jets j_i folgende Formel gilt:

$$\phi_{\text{rot}}(j_i) = \phi(j_i) - \phi(E_T^{\text{miss}}) \quad (45)$$

Auch hier wurde zusätzlich der $\sin \phi_{\text{rot}}$ und der $\cos \phi_{\text{rot}}$ angewendet. Dies ist in Tabelle 12 dargestellt.

| Eingangs-Variablen | Beschreibung | |
|---|--|----------------------------|
| $p_T(\ell_0)$ $\eta(\ell_0)$ $\sin(\phi_{\text{rot}}(\ell_0))$ $\cos(\phi_{\text{rot}}(\ell_0))$ | fehlender Transversalimpuls Pseudorapidität Sinus des rotierten ϕ -Winkels Cosinus des rotierten ϕ -Winkels | des führenden Leptons |
| $p_T(\ell_1)$ $\eta(\ell_1)$ $\sin(\phi_{\text{rot}}(\ell_1))$ $\cos(\phi_{\text{rot}}(\ell_1))$ | fehlender Transversalimpuls Pseudorapidität Sinus des rotierten ϕ -Winkels Cosinus des rotierten ϕ -Winkels | des zweitführenden Leptons |
| $p_T(j_0)$ $\eta(j_0)$ $\sin(\phi_{\text{rot}}(j_0))$ $\cos(\phi_{\text{rot}}(j_0))$ | fehlender Transversalimpuls Pseudorapidität Sinus des rotierten ϕ -Winkels Cosinus des rotierten ϕ -Winkels | des führenden Jets |
| $p_T(j_1)$ $\eta(j_1)$ $\sin(\phi_{\text{rot}}(j_1))$ $\cos(\phi_{\text{rot}}(j_1))$ | fehlender Transversalimpuls Pseudorapidität Sinus des rotierten ϕ -Winkels Cosinus des rotierten ϕ -Winkels | des zweitführenden Jets |
| $p_T(E_T^{\text{miss}})$ | Transversalimpuls der fehlenden Transversalenergie | |
| Ziel-Variablen | | |
| $p_T(H, \text{truth})$ $\eta(H, \text{truth})$ $\phi_{\text{rot}}(H, \text{truth})$ $M(H, \text{truth})$ | fehlender Transversalimpuls Pseudorapidität Rotierter Azimuth-Winkel Masse | des wahren Higgs Teilchens |

Tabelle 12: Übersicht der Eingangs- und Zielvariablen des vierten neuronalen Netzes NN_4 (Polarkoordinaten mit Jets). Die entsprechenden Eingangs-Verteilungen sind im Anhang B.4 in Abbildung 27 bis 31 zu finden.

7.1.5 Ergebnisse der OPTUNA-Optimierungen

Mit der Festlegung der Ein- und Ausgabe-Variablen der vier verschiedenen neuronalen Netze können nun die Hyperparameter dieser optimiert werden. Hier soll zunächst das Vorgehen bezüglich der Optimierungsprozesse mit OPTUNA beschrieben werden. Für jedes der vier Modelle wurde eine OPTUNA-Optimierung vollzogen, wobei ein Optimierungsprozess 100 Versuche enthielt. Es wurden also für jedes der vier Systeme 100 verschiedene Netze trainiert und bewertet. Aus einer groben Abschätzung folgt, dass dies ca. einem Drittel der möglichen Parameterkombinationen entspricht, die in der Zielfunktion der OPTUNA-Optimierung festgelegt wurden. Daher erscheint die Wahl von 100 Versuchen auch in Hinblick auf die Rechenkapazität des verwendeten Laptops als sinnvoll.

Bei jeder Optimierung wurden 70% der gesamten Daten zum Trainieren und die restlichen 30% zum Validieren verwendet. Der Parameterraum, welcher die Struktur des Netzes bestimmt, wurde für jedes der vier OPTUNA-Systeme folgendermaßen gewählt: Ein Netz kann aus einer bis fünf versteckten Schichten aufgebaut sein und jede dieser Schichten kann entweder 20, 50 oder 100 Knoten besitzen. Die Aktivierungsfunktion der versteckten Schichten ist in allen Fällen die ReLU-Funktion, wie sie in Gleichung 30 definiert ist. Die Aktivierungsfunktion der Ausgabe-Schicht entspricht stets der Identität. Als Lernalgorithmus wurde der Adam-Optimierer eingesetzt, dessen Verlust-

funktion, sowie Metrik in allen vier Systemen der mittleren quadratischen Abweichung (siehe Gleichung 31) entspricht. Auch die möglichen Kombinationen der Lernrate wurden in allen Systemen gleich gewählt, wobei hier Zahlen von $1 \cdot 10^{-5}$ bis $1 \cdot 10^{-1}$ auf logarithmischer Skala möglich sind. Des Weiteren wurde in der Zielfunktionen der OPTUNA-Optimierungen festgelegt, dass die Batch-Größe im Trainingsprozess entweder 32, 64, 96 oder 128 betragen kann. Die möglichen Werte der Hyperparameter sind in Tabelle 13 zusammengefasst. Hinzu kommt, dass für jeden Versuch der vier Optimierungsprozesse 500 Epochen angesetzt waren, die jedoch aufgrund des Early-Stopping Algorithmus nicht voll ausgeschöpft wurden. Für die Abschätzung des Parameters `min_delta` des Early-Stopping Algorithmus (siehe Abschnitt 4.4) wurden vor den OPTUNA-Optimierungen für jedes der vier Systeme einfache Netze trainiert, um abschätzen zu können, in welchem Größenbereich die Werte der Verlustfunktion auf dem Validierungsset zu erwarten sind. Der Parameter wurde dann so ausgewählt, dass er ca. 3% der zu erwartenden mittleren quadratischen Abweichung entspricht. Dieser Wert wurde recht klein gewählt, um auch kleine Verbesserungen am Ende des Trainingsprozesses zu berücksichtigen. Für Netzwerk 1 und 2 beträgt `min_delta` = 20. Bei Netzwerk 3 hingegen wurde der Parameter auf 0,85 und bei Netzwerk 4 auf 1,16 gesetzt. Der `patience` Parameter wurde in allen Fällen auf 20 Epochen gesetzt, um ein zu frühzeitiges Abbrechen des Trainings zu verhindern.

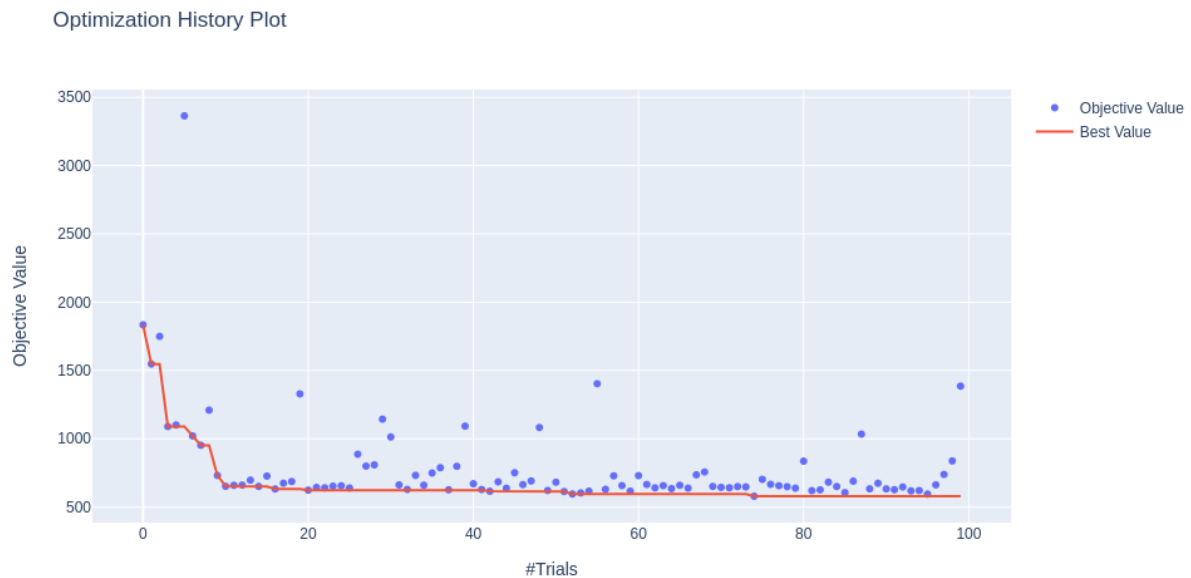
| Hyperparameter | Mögliche Werte |
|----------------------|---------------------------|
| Anzahl der Schichten | 1, 2, 3, 4, oder 5 |
| Knoten pro Schicht | 20, 50, oder 100 |
| Lernrate | auf logarithmischer Skala |
| Batch-Größe | 32, 64, 96, 128 |

Tabelle 13: Übersicht über den gewählten Parameter-Suchraum mit den Hyperparametern und ihren möglichen Werten

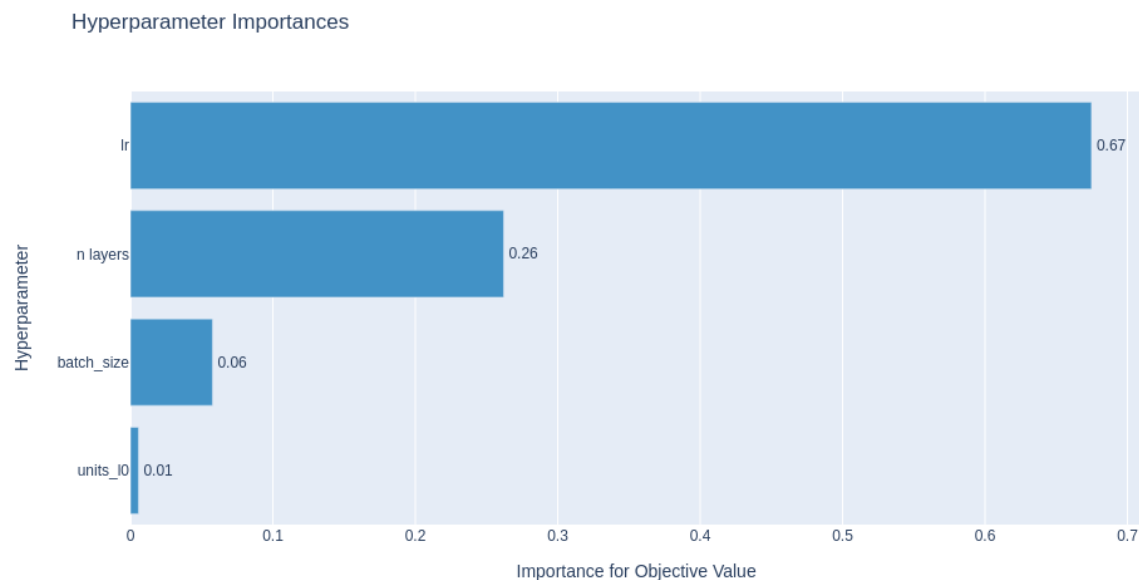
Um einen besseren Einblick in die Funktionsweise von OPTUNA zu bekommen, werden in unten stehender Abbildung zwei Graphen dargestellt, die den Optimierungsprozess des ersten neuronalen Netzes visualisieren. Zum Einen ist ein Graph, der den Trainings- und Studienverlauf der Optimierung darstellt, aufgeführt. Die x -Achse gibt hierbei die Anzahl der Versuche pro Optimierungsprozess an und die y -Achse gibt den Wert der Zielfunktion und somit den Wert des mittleren quadratischen Fehlers an. Die einzelnen Bewertungen eines Versuchs sind als blaue Punkte dargestellt und die rote Linie verbindet die Bewertungen, die am besten abgeschnitten haben. Wie man sieht, ist der mittlere quadratische Fehler der ersten Bewertungen noch vergleichsweise hoch und nach ca. zehn Epochen können nur noch kleine Verbesserungen in der Bewertung der Zielfunktion erzielt werden. Dies spricht dafür, dass die gewählte Anzahl von 100 Versuchen geeignet ist und man eventuell sogar noch eine etwas kleinere Gesamtanzahl der Versuche hätte wählen können.

Zum Anderen wird im zweiten Graph die Bedeutung der einzelnen Hyperparameter für die Optimierung der Zielfunktion abgebildet. In diesem Fall hat die Lernrate (`lr`) die größte Auswirkung auf die Gesamtleistung des Algorithmus, gefolgt von einer weitaus geringeren Auswirkung der Anzahl an Schichten des Netzes (`n_layers`). Die Batch-Größe (`batch_size`) ist für die Bewertung der Zielfunktion am dritt-wichtigsten, während die

Anzahl der Knoten in der ersten Schicht (`units_10`) des Netzes mit einem Wert von 0,01 kaum beiträgt. Die restlichen Hyperparameter, die optimiert wurden, sind in der Abbildung nicht dargestellt, da sie eine noch geringere Bedeutung als `units_10` für die Zielfunktion haben.



(a) Verlauf des OPTUNA-Optimierungsprozesses; `#Trials` gibt die Anzahl der Versuche an und *Objective Value* beschreibt den Wert der Zielfunktion, welche in diesem Fall dem mittleren quadratischen Fehlers entspricht.



(b) Prozentuale Wichtigkeit (engl.: *Importance*) der einzelnen Hyperparameter für die Zielfunktion (engl. *Objective Function*); Die Beschreibung der Parameter in der Abbildung erfolgt im beiliegenden Text.

Abbildung 11: Abbildungen zur Visualisierung des OPTUNA-Optimierungsprozesses für das erste neuronale Netz. Die entsprechenden Abbildungen der Netze NN_2 bis NN_4 sind im Anhang A.1 - A.3 in Abbildung 19 bis 21 dargestellt.

Die Ergebnisse der vier verschiedenen Optimierungsprozesse mit OPTUNA sind in Tabelle 14 dargestellt. Hierbei liegt die Anzahl der versteckten Schichten des ersten neuronalen Netzes bei fünf, wobei die erste, die zweite, die dritte und die fünfte Schicht 100 Knoten haben und die vierte Schicht 50 Knoten hat. Die Lernrate des Adam-Optimierers beträgt 0,0010 und die Batchgröße 64.

Das zweite Netz NN_2 enthält vier versteckte Schichten, wobei die ersten drei Schichten 100 Knoten haben und die letzte Schicht 20 Knoten hat. Die Lernrate des Adam-Optimierers von 0,0013 ist die größte von allen hier beschriebenen Netzen. Die Batch-Größe liegt bei 96.

Beim dritten Netz NN_3 befinden sich zwischen der Eingangsschicht und der Ausgangsschicht vier versteckte Schichten, wobei Schicht 1 und 3 eine Knotenzahl von 100 haben, Schicht 2 eine Knotenzahl von 20 und Schicht 4 eine Knotenzahl von 50 hat. Der Adam-Optimierer arbeitet hier am besten mit einer Lernrate von 0,0007. Die Batch-Größe liegt bei 96.

Das vierte Netz setzt sich aus einer Eingangsschicht mit 17 Knoten und einer Ausgangsschicht mit 4 Knoten zusammen. Dazwischen liegen fünf versteckte Schichten, wobei die erste und letzte Schicht 100 Knoten besitzt, während die zweite und dritte 20 Knoten, und die vierte 50 Knoten aufweist. Die Lernrate des Adam-Optimierers liegt bei 0,0009 und die Batch-Größe beträgt 96.

| | | NN_1 | NN_2 | NN_3 | NN_4 |
|----------------------|-----------|--------|--------|--------|--------|
| Anzahl der Schichten | | 5 | 4 | 4 | 5 |
| Knoten pro Schicht | Schicht 1 | 100 | 100 | 100 | 100 |
| | Schicht 2 | 100 | 100 | 20 | 20 |
| | Schicht 3 | 100 | 20 | 100 | 20 |
| | Schicht 4 | 50 | 100 | 50 | 50 |
| | Schicht 5 | 100 | | | 100 |
| Lernrate | | 0,0010 | 0,0013 | 0,0007 | 0,0009 |
| Batch-Größe | | 64 | 96 | 32 | 96 |

Tabelle 14: Ergebnisse der OPTUNA-Optimierungen für die neuronalen Netze NN_1 , NN_2 , NN_3 und NN_4

7.1.6 Gemeinsamkeiten der vier neuronalen Netze im Trainings- und Vorhersageprozess

Der nächste Schritt in der Analyse ist es, die vier verschiedenen Modelle zu trainieren. Bei diesem Prozess gibt es einige Gemeinsamkeiten.

Alle vier neuronale Netze wurden darauf trainiert, den Viererimpuls des Higgs-Bosons zu rekonstruieren. Die Vorgehensweise für die Vorbereitung des Trainings der verschiedenen Netze und die anschließende Vorhersage war für alle Modelle identisch und wird hier nur einmal im Detail erläutert.

Nachdem die Eingangs- und Ausgangs-Variablen definiert, sowie die optimalen Hyperparameter der vier verschiedenen Netze bestimmt wurden, wurde der gesamte Datensatz in einen Trainings- und Test-Set aufgeteilt, wobei in das Trainings-Set, wie schon bei den OPTUNA-Optimierungen, 70% und in das Test-Set 30% der Daten eingingen. Danach wurden die Daten mit Sklearns Standardscaler [64] so skaliert, dass die Verteilungen der

Eingangs-Werte einen Mittelwert von Null und eine Varianz von Eins haben. So wird verhindert, dass es verschiedene Skalengrößen in den Eingangs-Variablen gibt, wodurch das Netz die Daten besser und effizienter lernen kann. In dem darauf folgenden Schritt werden die vier neuronalen Netze mit den Hyperparametern, die gemäß den OPTUNA-Optimierungen optimal sind, definiert. Die restlichen Hyperparameter der Modelle gleichen denen, die bereits für die OPTUNA-Optimierungen gewählt wurden: Jeder versteckten Schicht liegt die ReLU-Aktivierungsfunktion zugrunde und die Verlustfunktion, sowie Metrik des Adam-Optimierers entspricht stets der mittleren quadratischen Abweichung. Im Falle der ersten beiden neuronalen Netze lautet die explizite Verlustfunktion folglich

$$\text{MSE}_{\text{NN}_1, \text{NN}_2}(\vec{w}) = \frac{1}{n} \sum_i^n (p_x^i - \hat{p}_x^i(\vec{w}))^2 + (p_y^i - \hat{p}_y^i(\vec{w}))^2 + (p_z^i - \hat{p}_z^i(\vec{w}))^2 + (E^i - \hat{E}^i(\vec{w}))^2, \quad (46)$$

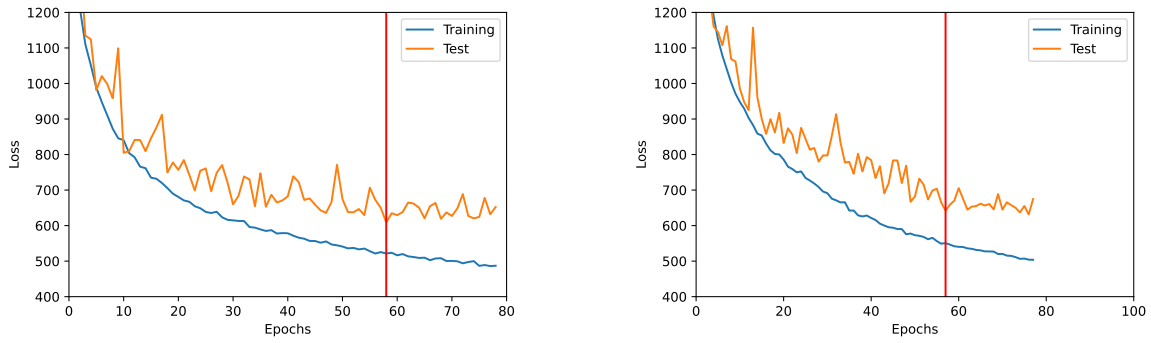
während die Verlustfunktion des dritten und vierten Netzes folgendermaßen definiert ist:

$$\text{MSE}_{\text{NN}_3, \text{NN}_4}(\vec{w}) = \frac{1}{n} \sum_i^n (p_T^i - \hat{p}_T^i(\vec{w}))^2 + (\eta^i - \hat{\eta}^i(\vec{w}))^2 + (\phi^i - \hat{\phi}^i(\vec{w}))^2 + (M^i - \hat{M}^i(\vec{w}))^2, \quad (47)$$

Hierbei bezieht sich i auf das i -te Ereignis und n beschreibt die Gesamtzahl an Ereignissen. Alle mit dem Hut gekennzeichneten Variablen stehen für die Vorhersagen des Netzwerkes mit den Netzwerkparametern \vec{w} , während die Variablen ohne Hut die wahren Zielvariablen beschreiben.

Für jedes Netz wird die exponentielle Abklingrate des Adam-Lernalgorithmus für den Mittelwert vergangener Gradienten auf $\beta_1 = 0,9$ und die exponentielle Abklingrate für die Varianz vergangener Gradienten auf $\beta_2 = 0,999$ gesetzt, was den Standardwerten entspricht. Für jedes Netz werden außerdem die Hyperparameter des Early-Stopping Algorithmus festgelegt, wobei auch hier die gleichen Kriterien wie schon in den OPTUNA-Optimierungen galten. Der `patience`-Parameter wurde bei allen Netzen auf 20 Epochen gesetzt. Für das erste und zweite Netz beträgt `min delta` = 20. Beim dritten Netz hingegen wurde der `min delta` Parameter auf 0,85 gesetzt. Im Falle des vierten Netzes liegt der `min delta`-Parameter bei 1,16. Daraufhin wurden die Modelle erst kompiliert und dann trainiert.

Ist der Trainingsprozess abgeschlossen, so wurden die trainierten Modelle genutzt, um den Viererimpuls des Higgs-Teilchens für jedes Ereignis vorherzusagen. Die Vorhersage wird hierbei über den gesamten Eingangs-Datensatz durchgeführt, um die volle Statistik für die weitere Analyse nutzen zu können. Es wurde überprüft, ob es einen Einfluss hat, dass das Modell 70% dieser Daten schon vom Trainieren kennt und es stellte sich heraus, dass dieser Effekt vernachlässigbar klein ist.



(a) Trainingskurve des ersten neuronalen Netzes

(b) Trainingskurve des zweiten neuronalen Netzes

Abbildung 12: Trainingskurven der ersten beiden Netzwerke, welche die Anzahl der Epochen (engl.: *Epochs*) auf der x -Achse und den Wert der Verlustfunktion (engl.: *Loss*) (in diesem Fall der mittlere quadratische Fehler) auf der y -Achse zeigen. Die rote Linie markiert den Punkt, an dem die besten Netzparameter gefunden wurden.

In Abbildung 12 sind beispielhaft die Trainingskurven der ersten beiden neuronalen Netze gezeigt. Hier wird der Zusammenhang zwischen der Anzahl an trainierten Epochen und dem Wert der Verlustfunktion auf dem Trainings- und Testdatensatz dargestellt. Die rote Linie markiert den Punkt, an dem die besten Netzparameter gefunden wurden, jedoch wird der Trainingsprozess hier nicht sofort gestoppt, da der *patience*-Parameter des Early-Stopping-Algorithmus auf 20 Epochen eingestellt wurde. Dadurch wird ein zu frühes Stoppen des Trainingsprozesses verhindert.

7.1.7 Vergleich der Rekonstruktionsgüte für den Higgs-Boson-Viererimpuls

Jedes der vier trainierten Modelle wurde dafür genutzt, den Viererimpuls des Higgs-Teilchens für jedes Ereignis des Input-Datensatzes vorherzusagen. Die ersten beiden Netze liefern den Viererimpuls in kartesischen Koordinaten $(p_x(H), p_y(H), p_z(H), E(H))$, während die anderen beiden Netze die Polar-Komponenten des Higgs-Viererimpulses $(p_T(H), \eta(H), \phi_{\text{rot}}(H), M(H))$ zurückgeben. Der Winkel ϕ_{rot} wird anschließend zurück rotiert, sodass man $\phi(H)$ erhält.

In diesem Kapitel erfolgt ein Vergleich der vorhergesagten Viererimpulse p_{H,NN_i}^μ der i Modelle, indem die Auflösungen der rekonstruierten Viererimpulse mit den wahren Viererimpulsen des Higgs-Bosons $p_{H,\text{truth}}^\mu$ als Referenzwert berechnet werden. Zunächst erfolgt die Berechnung der Auflösung (Res) für die Viererimpulse, die mittels der vier verschiedenen Modelle vorhergesagt wurden:

$$\text{Res}(p_{H,NN_i}^\mu)_j = (p_{H,\text{truth}}^\mu)_j - (p_{H,NN_i}^\mu)_j \quad (48)$$

Hierbei bezieht sich j auf eine Komponente des Viererimpulses. Da es das Ziel ist, die Ergebnisse des *Missing Mass Calculators* mit Methoden maschinellen Lernens zu verbessern, wird zum Vergleich auch die Auflösung des MMC berechnet:

$$\text{Res}(p_{H,MMC}^\mu)_j = (p_{H,\text{truth}}^\mu)_j - (p_{H,MMC}^\mu)_j \quad (49)$$

Zusätzlich dazu, werden die Auflösungen der p_T - und ϕ -Komponenten berechnet, die durch vektorielle Aufaddierung gemäß Gleichung 41 rekonstruiert wurden.

Es werden also zwei Rekonstruktionsmethoden, die den vollständigen Viererimpuls des Higgs-Bosons liefern, miteinander verglichen - die Methode des maschinellen Lernens und die Methode des Missing Mass Calculators. Die berechneten Werte für die Auflösungen wurden für alle vier Modelle in Histogramme gefüllt, wobei im Rahmen dieses Kapitels beispielhaft die p_T -, η -, ϕ -, M - und E -Auflösungsabbildungen des ersten Netzes in Abbildung 13 aufgeführt werden. Die restlichen Auflösungsabbildungen der vier neuronalen Netze sind im Anhang C in Abbildung 32 bis 36 dargestellt. Aus diesen Histogrammen wurde dann für jede Rekonstruktionsmethode der RMS (engl: *Root Mean Square*)

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (50)$$

ermittelt. N beschreibt die Anzahl der Datenpunkte x_i , während \bar{x} den Mittelwert einer Verteilung angibt. Die berechneten Werte für den RMS bezüglich der Auflösungen in (p_T, η, ϕ, M, E) sind in Tabelle 15 dargestellt. Hierfür wurden im Vorfeld die vorhergesagten Viererimpulse $(p_x(H), p_y(H), p_z(H), E(H))$ der ersten beiden kartesischen Modelle in Polarkoordinaten $(p_T(H), \eta(H), \phi(H), M(H))$ transformiert. Dies gilt auch für die Auflösungsabbildungen, wobei die Abbildungen der ursprünglichen kartesischen Komponenten der Viererimpulse im Anhang in Abbildung 32 gezeigt sind.

| Variable | RMS | | | | | |
|-------------|-------|------|-----------------|-----------------|-----------------|-----------------|
| | MMC | add | NN ₁ | NN ₂ | NN ₃ | NN ₄ |
| p_T [GeV] | 20,0 | 20,2 | 14,7 | 14,9 | 11,7 | 11,7 |
| η | 0,110 | - | 0,098 | 0,100 | 0,278 | 0,385 |
| ϕ | 0,15 | 0,13 | 0,12 | 0,12 | 0,19 | 0,19 |
| M [GeV] | 20,1 | - | 11,0 | 14,9 | 0,8 | 0,9 |
| E [GeV] | 49 | - | 26 | 27 | 57 | 65 |

Tabelle 15: Übersicht der aus den Auflösungsabbildungen folgenden Werte für den RMS in den Variablen p_T, η, ϕ, M, E ; Der RMS des MMC und der RMS der durch die Methode der vektoriellen Aufaddierung der Zerfallsteilchen des Higgs-Bosons (RMS(add)) stehen im direkten Vergleich zu den RMS-Werten, welche die Auflösungsabbildungen der neuronalen Netzen NN_i geliefert haben.

In Tabelle 15 ist zu erkennen, dass die neuronalen Netze 1 und 2 bezüglich allen Variablen signifikante Verbesserungen der Auflösung im Vergleich zum MMC und der Additions-Methode erzielen konnten. Mit dem dritten und dem vierten neuronalen Netzwerk konnten keine Verbesserungen bezüglich der η -, ϕ - und E -Komponente erzielt werden, aber eine signifikante Verbesserung von 95% bei der Massenrekonstruktion im Vergleich zum MMC.

Die deutlich bessere Auflösung in der Masse M der letzten beiden Netze lässt sich dadurch erklären, dass diese Netze auf die M -Koordinate des wahren Higgs-Teilchens trainiert wurden, während die ersten beiden Netze auf die E -Koordinate des wahren Higgs-Teilchens trainiert wurden. Hinzu kommt, dass die Massen-Vorhersage der letzten beiden Netze besonders einfach war, da die Ereignisse für Higgs-Massen von 125 GeV simuliert wurden. Die E -Koordinate, die von den ersten beiden Netzen vorhergesagt wurde, hat einen deutlich größeren Wertebereich, wodurch diese Variable für die Netze schwieriger zu lernen war.

Die untenstehende Abbildung 13 zeigt die zu der Tabelle zugehörigen Auflösungsabbildungen des ersten neuronalen Netzes. Hierbei fällt auf, dass die Auflösung der durch die Rekonstruktionsmethode der vektoriellen Addition der Higgs-Zerfallsteilchen nur in der p_T - und η -Koordinate dargestellt ist. Auch in der Tabelle ist dies wiedergespiegelt. Der Grund hierfür ist, dass für die eben genannte Rekonstruktionsmethode die fehlende Transversalenergie E_T^{miss} verwendet wird, die ausschließlich in der x - y -Ebene gemessen wird, wodurch die η -Komponente, sowie die M - und E -Komponente der fehlenden Transversalenergie nicht wohldefiniert sind. Daher bietet diese Rekonstruktionsmethode keine faire Vergleichsbasis für den *Missing Mass Calculator* und die neuronalen Netzwerke.

In der Auflösungsabbildung 13 ist zu erkennen, dass die p_T -Auflösung des neuronalen Netzes um 26,67% besser als die vom MMC, sowie um 27,28% besser als die vom einfachen Aufaddieren der Higgs-Boson-Zerfallsteilchen ist. Auch in der ϕ -Komponente lässt sich eine prozentuale Verbesserung von 17,76% gegenüber dem MMC und eine Verbesserung von 4,01% gegenüber der Additions-Methode beobachten. In der η -Komponente liegt die prozentuale Verbesserung des neuronalen Netzes im Vergleich zum MMC bei 10,85% und in der M -Komponente bei 45,32%. Die größte Verbesserung lässt sich bei der Energie E feststellen, die bei 46,49% liegt. In allen Variablen ließen sich also signifikante Verbesserungen in der Auflösung erreichen.

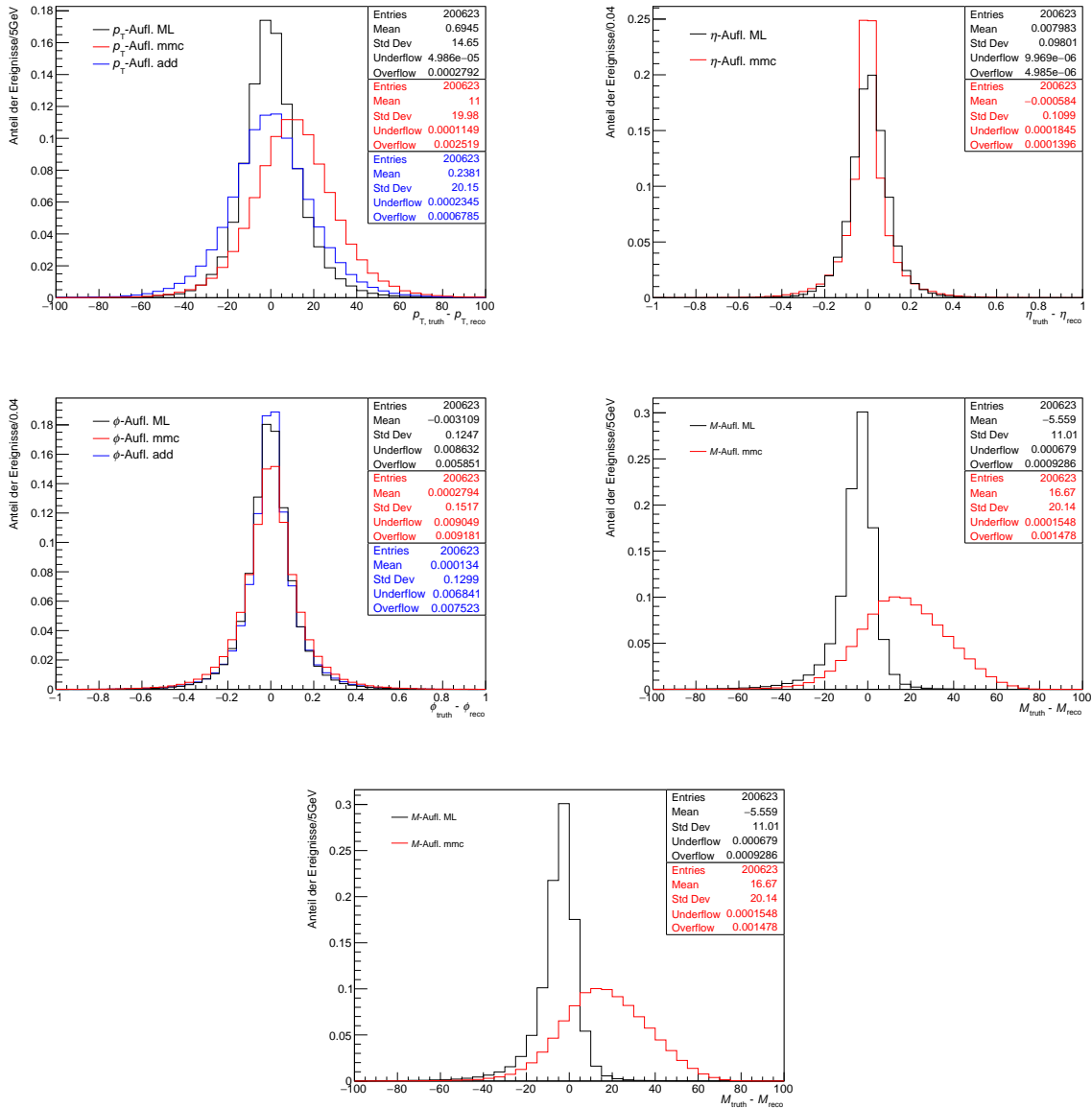


Abbildung 13: Auf Eins normierte Auflösungsabbildungen des ersten neuronalen Netzes (ML), des *Missing Mass Calculators* (MMC) und der Rekonstruktions-Methode des vektoriellen Aufaddierens der Higgs-Zerfallsteilchen (add); Gezeigt ist jeweils die Auflösung bezüglich des Transversalimpulses p_T , der Pseudorapidität η , dem Azimutwinkel ϕ , der Masse M , sowie der Energie E des rekonstruierten Higgs-Bosons. Die Statistikboxen sind in englischer Sprache dargestellt, wobei *Entries* die Anzahl der Einträge im Histogramm, *Mean* den Mittelwert der Verteilung und *Std Dev* die Standardabweichung bzw. den RMS der Verteilung angibt. Die Anzahl der Werte, die kleiner als die Untergrenze des in der Verteilungen vorliegenden Wertebereichs sind und daher nicht abgebildet werden, wird durch *Underflow* wiedergegeben, wobei *Overflow* die Anzahl der Werte angibt, die im positiven Bereich über den gezeichneten Wertebereich hinausgehen. Die Bedeutung der Einträge in den Statistikboxen ist in allen dieselbe und wird im Folgenden nicht mehr erläutert.

7.2 Berechnung der Optimalen Observable

Die Optimale Observable \mathcal{OO} wurde für jedes Ereignis mittels dem Python-Modul `vbfcprw` [65] [66] [67] berechnet. Zur Berechnung der Optimalen Observable benötigt man den Viererimpuls des Higgs-Bosons p_H^μ und die Viererimpulse der beiden Partonen $p_{p_0}^\mu$ und $p_{p_1}^\mu$ bzw. der Jets $p_{j_0}^\mu$ und $p_{j_1}^\mu$, die aus der Vektor-Boson-Fusion stammen.

7.2.1 Beiträge der verschiedenen Objekte zur Auflösung der Optimalen Observable

In Abbildung 14 ist dargestellt, welchen Einfluss die verschiedenen Eingangs-Größen, die zur Berechnung der \mathcal{OO} benötigt werden, auf die Auflösung der Optimalen Observablen haben. Der Referenzwert der Auflösungsabbildung ist die Optimale Observable, die mit den wahren Higgs-Viererimpulsen und den Viererimpulsen der Partonen, die die volle Wahrheitsinformation tragen, berechnet wurde. Nun können die Eingangs-Größen zur Berechnung der Optimalen Observable von Variablen auf Wahrheitsniveau zu Variablen auf Rekonstruktionsniveau variiert werden. Auf Seiten der Jets stehen die Viererimpulse der Partonen, der wahren Jets, und der rekonstruierten Jets zu Verfügung, während für den Higgs-Boson-Viererimpuls der wahre Viererimpuls und der mit dem MMC rekonstruierte Viererimpuls verwendet werden kann. Die möglichen Kombinationen von Eingangsgrößen für die Berechnung der Optimalen Observable sind in folgender Liste zusammengefasst:

- Higgs-Boson-Viererimpuls aus MMC $p_{H,MMC}^\mu$ und Viererimpulse der Partonen auf Wahrheitsniveau $p_{parton_0}^\mu$ und $p_{parton_1}^\mu$
- Higgs-Boson-Viererimpuls aus MMC $p_{H,MMC}^\mu$ und Viererimpulse der Jets $p_{j_0,truth}^\mu$ und $p_{j_1,truth}^\mu$ auf Wahrheitsniveau
- Higgs-Boson-Viererimpuls aus MMC $p_{H,MMC}^\mu$ und Viererimpulse der Jets $p_{j_0}^\mu$ und $p_{j_1}^\mu$ auf Rekonstruktionsniveau
- Higgs-Boson-Viererimpuls auf Wahrheitsniveau $p_{H,truth}^\mu$ und Viererimpulse der Jets $p_{j_0,truth}^\mu$ und $p_{j_1,truth}^\mu$ auf Wahrheitsniveau
- Higgs-Boson-Viererimpuls auf Wahrheitsniveau $p_{H,truth}^\mu$ und Viererimpulse der Jets $p_{j_0}^\mu$ und $p_{j_1}^\mu$ auf Rekonstruktionsniveau

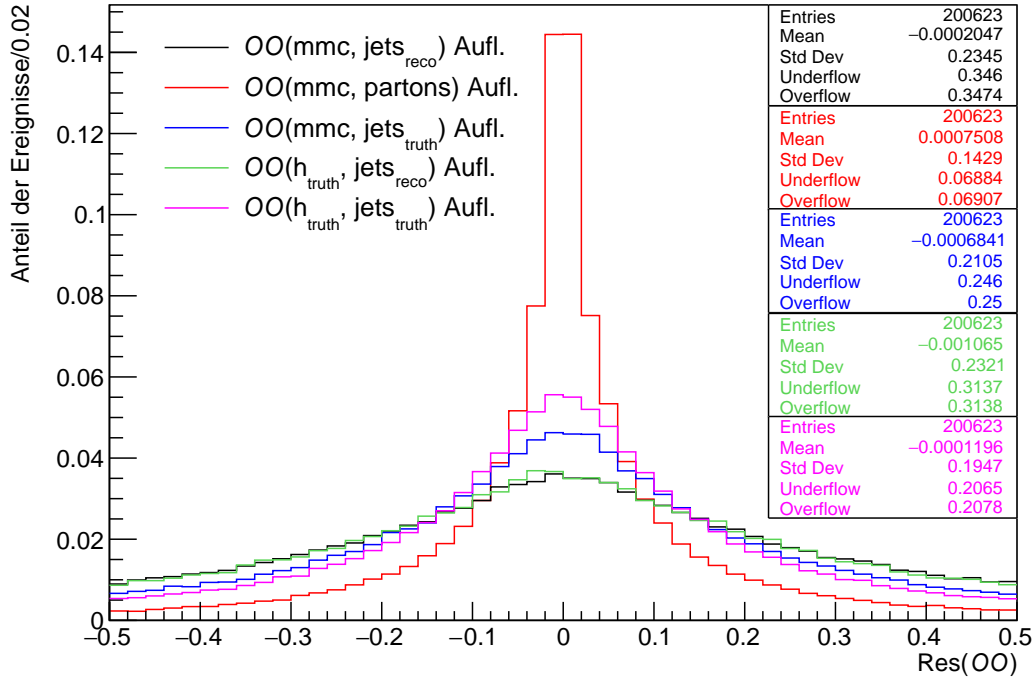


Abbildung 14: Auflösung der Optimalen Observable für verschiedene Eingangs-Größen auf Wahrheits- und Rekonstruktionsniveau. Für die Abbildung wurde als Referenzwert die Optimalen Observable verwendet, die mit den Viererimpulsen der beiden Partonen und den wahren Higgs-Boson-Impulsen berechnet wurde.

Die aus der Auflösungsabbildung 14 folgenden Werte für den RMS für die verschiedenen Kombinationen der Eingangsgrößen zur Berechnung der Optimalen Observable sind in Tabelle 16 gezeigt. Die beste Auflösung hat die Optimalen Observable, die mit dem Viererimpuls des MMC und den Viererimpulsen der VBF-Partonen berechnet wurde. Die Auflösung sinkt um ca. 36% bei der Optimalen Observable, die mit dem wahren Viererimpuls des Higgs-Bosons und den wahren Jets berechnet wurde. Dies ist ein signifikanter Verlust, was deutlich macht, dass die Partonen, die zu Jets hadronisieren, einen erheblichen Einfluss auf die Auflösung der Optimalen Observable haben. Dieser starke Verlust in der Auflösung könnte möglicherweise an einer fehlerhaften Parton-Jets-Assoziation bei einigen Ereignissen liegen. Bei einer Berechnung der Optimalen Observable mit dem Viererimpuls des MMC und den wahren Viererimpulsen der Jets lässt sich eine Verschlechterung in der Auflösung im Vergleich zu Letzterem von ca. 8% beobachten. Hier wird deutlich, dass die Auflösung in der Optimalen Observable durch die Rekonstruktionsgüte des Higgs-Boson-Viererimpulses limitiert ist, weswegen eine präzise Rekonstruktion der Viererimpulse von großer Bedeutung ist - nicht nur, um die Auflösung in der Optimalen Observable zu verbessern, sondern um somit auch die Sensitivität von CP-Tests zu erhöhen. Verwendet man für die \mathcal{OO} -Berechnung die Viererimpulse der rekonstruierten Jets und den Viererimpuls des wahren Higgs-Bosons, so folgt eine weitere Verschlechterung der Auflösung um etwa 10% im Vergleich zu der Optimalen Observable, die mit dem Viererimpuls des MMC und den wahren Viererimpulsen der Jets berechnet wurde. Die Rekonstruktionsqualität der Jets hat also auch einen erheblichen Einfluss auf die \mathcal{OO} -Auflösung. Die schlechteste Auflösung hat die Optimalen Observable,

die mit den Viererimpulsen des MMC und den Viererimpulsen der rekonstruierten Jets berechnet wurde, was den Erwartungen entspricht, da beide Berechnungs-Komponenten auf Rekonstruktions-Niveau sind.

| Variable | RMS | | | | |
|----------------|--------------------------------------|---------------------------------------|--------------------------------------|---|--|
| | $p_{H,MMC}^\mu$ & $p_{parton_i}^\mu$ | $p_{H,MMC}^\mu$ & $p_{j_i,truth}^\mu$ | $p_{H,MMC}^\mu$ & $p_{j_i,reco}^\mu$ | $p_{H,truth}^\mu$ & $p_{j_i,truth}^\mu$ | $p_{H,truth}^\mu$ & $p_{j_i,reco}^\mu$ |
| \mathcal{OO} | 0,1429 | 0,2105 | 0,2345 | 0,1947 | 0,2321 |

Tabelle 16: Übersicht der aus der Auflösungsabbildung folgenden Werte für den RMS in der Optimalen Observable \mathcal{OO} für die verschiedenen Kombinationen der Eingangsgrößen zur Berechnung der Optimalen Observable

7.2.2 Auflösungsvermögen der Optimalen Observable für die Rekonstruktion mit neuronalen Netzen

Nun wird untersucht, inwieweit die Viererimpulse, die mit den neuronalen Netzen vorhergesagt wurden, die Auflösung in der Optimalen Observable verbessern können. Es werden fünf verschiedene Methoden zur Rekonstruktion des Viererimpulse des Higgs-Bosons miteinander verglichen. Zum Einen stehen zur Berechnung der Optimalen Observablen die von den vier neuronalen Netzwerken vorhergesagten Viererimpulse (p_{H,NN_1}^μ , p_{H,NN_2}^μ , p_{H,NN_3}^μ , p_{H,NN_4}^μ) des Higgs-Teilchens zu Verfügung und zum Anderen kann die Optimale Observable mit dem Viererimpuls des MMC $p_{H,MMC}^\mu$, der in Abschnitt 5.1 beschrieben ist, berechnet werden. Da es sich bei den Regressionsnetzen, sowie bei dem *Missing Mass Calculator* um Rekonstruktionsmethoden handelt, werden zur Berechnung der \mathcal{OO} die Jets auf Rekonstruktionsniveau verwendet. Als Referenzwert wird die Optimale Observable mit dem Viererimpuls des wahren Higgsteilchens $p_{H,truth}^\mu$ und den Viererimpulsen der wahren Jets $p_{j_0,truth}^\mu$ und $p_{j_1,truth}^\mu$ verwendet.

Schließlich können mit den sechs Viererimpulsen des Higgs-Bosons sechs verschiedene Optimale Observablen berechnet werden: die optimalen Observablen, welche die neuronalen Netze geliefert haben (\mathcal{OO}_{NN_1} , \mathcal{OO}_{NN_2} , \mathcal{OO}_{NN_3} , \mathcal{OO}_{NN_4}), die optimale Observable des *Missing Mass Calculators* (\mathcal{OO}_{MMC}) und die wahre optimale Observable (\mathcal{OO}_{truth}). Für jede dieser Optimalen Observablen wird gefordert, dass $-12 < \mathcal{OO} < 12$ erfüllt wird, um die Effekte von Ausreißerwerten zu verringern.

Mit den nun ermittelten Optimalen Observablen, kann die Auflösung mit der wahren Observable \mathcal{OO}_{truth} als Referenzwert berechnet werden

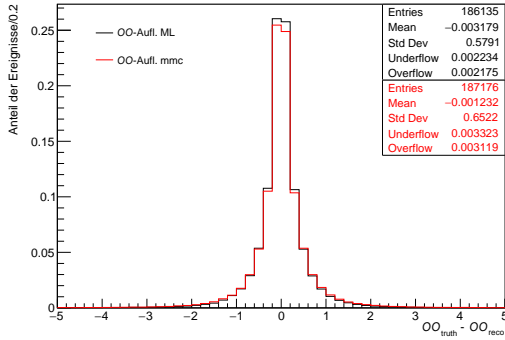
$$\text{RES}(\mathcal{OO}_{reco}) = \mathcal{OO}_{truth} - \mathcal{OO}_{reco}, \quad (51)$$

wobei sich der Index *reco* auf eine der Rekonstruktionsmethoden (NN_1 , NN_2 , NN_3 , NN_4 , oder MMC) bezieht. Wie im vorherigen Kapitel werden auch hier Auflösungsabbildungen erstellt, die in Abbildung 15 dargestellt sind. Aus diesen ergeben sich die RMS-Werte. Die Ergebnisse sind in Tabelle 17 gezeigt:

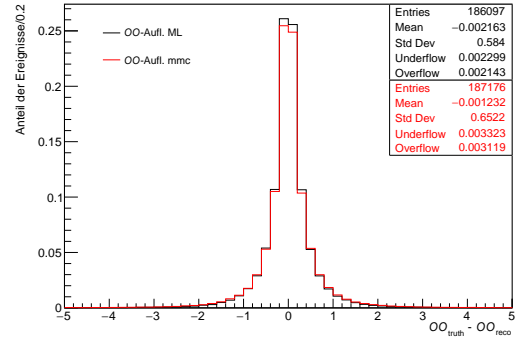
| Variable | RMS | | | | |
|----------------|-------|-----------------|-----------------|-----------------|-----------------|
| | MMC | NN ₁ | NN ₂ | NN ₃ | NN ₄ |
| \mathcal{OO} | 0,652 | 0,579 | 0,584 | 0,623 | 0,673 |

Tabelle 17: Übersicht der RMS-Werte der Optimalen Observablen \mathcal{OO} für die verschiedenen Rekonstruktionsmodelle

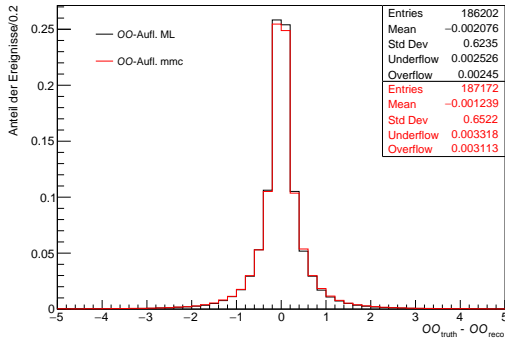
Die größte Verbesserung in der Auflösung der Optimalen Observablen konnte mit dem ersten neuronalen Netz NN_1 erreicht werden. Hier liegt prozentuale Verbesserung im Vergleich zum MMC bei 11,20%. Auch das zweite neuronale Netz NN_2 konnte mit 10,43% eine Verbesserung von über zehn Prozent liefern. Das dritte neuronale Netz NN_3 konnte die Auflösung im Vergleich zum MMC um 4,45% verbessern, während das vierte neuronale Netz NN_4 keine Verbesserung lieferte. Die Gründe für die Unterschiede in der Auflösungsverbesserung der vier Netzwerke konnte im Rahmen dieser Arbeit nicht mehr untersucht werden.



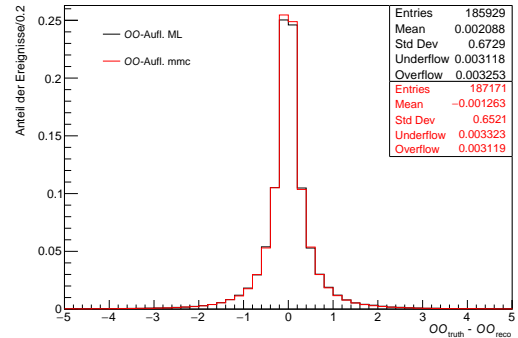
(a) Erreichte OO -Auflösung mit dem ersten neuronalen Netz (kartesische Koordinaten) im Vergleich zum MMC



(b) Erreichte OO -Auflösung mit dem zweiten neuronalen Netz (kartesische Koordinaten mit Jets) im Vergleich zum MMC



(c) Erreichte OO -Auflösung mit dem dritten neuronalen Netz (Polarkoordinaten) im Vergleich zum MMC



(d) Erreichte OO -Auflösung mit dem vierten neuronalen Netz (Polarkoordinaten mit Jets) im Vergleich zum MMC

Abbildung 15: Normierte Histogramme, welche die Auflösung in der Optimalen Observable OO mit der wahren Optimalen Observablen als Referenzwert zeigen. Es werden jeweils die beiden Rekonstruktionsmethoden (reco) des maschinellen Lernens (ML, schwarz) und des Missing Mass Calculators (mmc, rot) verglichen.

8 Bestimmung der Sensitivität von \tilde{d}

Im Standardmodell ($\tilde{d} = 0$) wird das Higgs-Boson als ein neutrales, skalares Teilchen mit gerader Parität eingeführt. Jedoch könnten neue Quellen von CP-Verletzung im Higgs-Sektor existieren. Eine mögliche Quelle stellen anomale Kopplungen des Higgs-Bosons an elektroschwache Vektorbosonen HVV dar, die in diesem Kapitel untersucht werden. Hierfür wird die CP-ungerade Optimale Observable verwendet, deren Mittelwert zu verschiedenen \tilde{d} berechnet wird, wobei \tilde{d} die Stärke der CP-Verletzung beschreibt (siehe Kapitel 2.5). Durch Umgewichten der Signalereignisse können verschiedene Szenarien mit $\tilde{d} \neq 0$ simuliert werden. Hierfür wird der Zusammenhang zwischen den Mittelwerten der Optimalen Observable und den verschiedenen \tilde{d} -Werten in sogenannten Eichkurven dargestellt. Aus diesen Eichkurven kann die Sensitivität für die Messung von \tilde{d} ermittelt werden.

Die folgende CP-Analyse gliedert sich wie folgt:

- a) Simulation von \mathcal{OO} -Verteilungen für $\tilde{d} \neq 0$ durch Umgewichtung der Ereignisse
- b) Bestimmung der Eichkurven, d.h. Bestimmung des Zusammenhangs zwischen dem Mittelwert der Optimalen Observable $\langle \mathcal{OO} \rangle$ und \tilde{d}
- c) Bestimmung der bei einer integrierten Luminosität von 139 fb^{-1} erwarteten Fehler auf \tilde{d}

8.1 Simulation von Verteilungen der Optimalen Observable für $\tilde{d} \neq 0$

Die durch die Monte-Carlo-Generatoren simulierten Ereignisse basieren auf dem Standardmodell mit $\mathcal{CP} = 1$ für das Higgs-Boson und $\tilde{d} = 0$, wodurch eine Untersuchung der CP-Eigenschaften des Higgs-Bosons nicht möglich ist. Um die Sensitivität auf eine mögliche CP-Verletzung untersuchen zu können, müssen die Ereignisse folglich für \tilde{d} -Werte ungleich Null umgewichtet werden. Hierzu wird für jedes Ereignis ein Gewicht

$$\begin{aligned}
 w(\tilde{d}) &= \frac{|\mathcal{M}|^2}{|\mathcal{M}_{\text{SM}}|^2} \\
 &= 1 + \tilde{d} \cdot \frac{2\text{Re}(\mathcal{M}_{\text{SM}}^* \mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\text{SM}}|^2} + \tilde{d}^2 \cdot \frac{|\mathcal{M}_{\text{CP-odd}}|^2}{|\mathcal{M}_{\text{SM}}|^2} \\
 &= 1 + \tilde{d} \cdot w_{\text{lin}} + \tilde{d}^2 \cdot w_{\text{quad}}
 \end{aligned} \tag{52}$$

berechnet, wobei w_{lin} und w_{quad} die linearen bzw. die quadratischen Anteile der Gewichte sind, die in dem für die Analyse verwendeten Datensatz enthalten sind und ebenfalls mit dem `vbfprw`-Modul berechnet werden können. Für die quadratischen Anteile der Gewichte wurde $w_{\text{quad}} < 400$ gefordert, um eine zu große Gewichtung einzelner Ereignisse zu verhindern bzw. um den Effekt von Ausreißern zu minimieren.

8.2 Abhängigkeit der Optimalen Observablen von \tilde{d}

Im Folgenden wird die Abhängigkeit der Verteilungen der Optimalen Observablen von \tilde{d} untersucht. Abbildung 16 zeigt die Verteilungen der Optimalen Observablen bei drei

verschiedenen \tilde{d} -Werten. Zum Einen wird ein negativer Wert $\tilde{d} = -0,10$ und zum Anderen ein positiver Wert $\tilde{d} = 0,05$ im Gegensatz zu der Standardmodell-Vorhersage $\tilde{d} = 0$ betrachtet. Wie aus der Abbildung zu entnehmen ist, ist die Verteilung für $\tilde{d} = 0$ symmetrisch um die Null verteilt. Für den negativen \tilde{d} -Wert wird die Verteilung asymmetrischer und verschiebt sich ins Negative. Auch bei dem positiven \tilde{d} -Wert kann eine Asymmetrie in der Verteilung beobachtet werden und die Verteilung ist nun ins Positive verschoben. Ein \tilde{d} -Wert von ungleich Null beeinflusst also die Verteilung der Optimalen Observablen, womit diese auf CP-Verletzung sensitiv ist. Die Verteilungen der Optimalen Observable für verschiedene \tilde{d} -Werte hängt nicht nur vom Vorzeichen, sondern auch vom Betrag von \tilde{d} ab. Dies ist ebenfalls in Abbildung 16 zu erkennen.

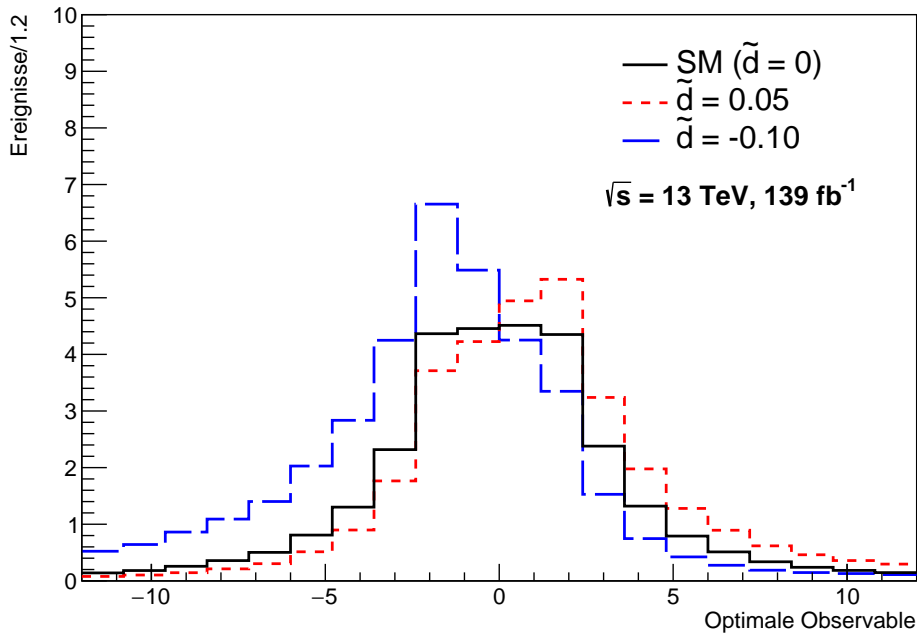


Abbildung 16: Die hier gezeigten Verteilungen der Optimalen Observable für drei verschiedene \tilde{d} -Werte sind auf eine integrierte Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

Um nun die Abhängigkeit der Optimalen Observable \mathcal{OO} von \tilde{d} zu untersuchen, werden die Eichkurven zur Analyse herangezogen. Diese bilden den gewichteten Mittelwert der \mathcal{OO} -Verteilung in Abhängigkeit von \tilde{d} ab. Der Mittelwert wird mit den Gewichten w_i gemäß Gleichung 52 gewichtet, um Szenarien mit $\tilde{d} \neq 0$ untersuchen zu können. Im Falle des Standardmodells ($\tilde{d} = 0$) betragen die Gewichte für jedes Ereignis $w(0) = 1$. Der gewichtete Mittelwert von n Ereignissen wird über

$$\langle \mathcal{OO} \rangle = \frac{\sum_{i=1}^n w_i \cdot (\mathcal{OO})_i}{\sum_{i=1}^n w_i} \quad (53)$$

und dessen Standardabweichung über

$$s_{(\mathcal{OO})} = \frac{1}{\sqrt{n-1}} \sqrt{\frac{\sum_{i=1}^n w_i \cdot (\mathcal{OO}^2)_i}{\sum_{i=1}^n w_i} - \langle \mathcal{OO} \rangle^2} \quad (54)$$

berechnet. Die Standardabweichung des gewichteten Mittelwerts $s_{\langle \mathcal{O} \rangle, \text{exp}}$, die man bei einer Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ erwarten würde, berechnet sich ebenfalls über Formel 54, jedoch werden die Gewichte w_i entsprechend skaliert.

8.2.1 Eichkurven

Die berechneten Mittelwerte der Optimalen Observablen können nun in Abhängigkeit von \tilde{d} aufgetragen werden. In Abbildung 17 ist ein Beispiel einer Eichkurve gezeigt. Da die Umgewichtung der Ereignisse zu einer Asymmetrie in der Verteilung führt, ist der Mittelwert für $\tilde{d} \neq 0$, von Null verschieden. Außerdem erkennt man für kleine $|\tilde{d}|$ -Werte ≈ 0 einen linearen Verlauf der Eichkurve, da hier die in \tilde{d} quadratischen Beiträge im Matrixelement vernachlässigbar klein sind. Für große $|\tilde{d}|$ -Werte nähert sich der Mittelwert asymptotisch der Null. Dies entspricht auch den theoretischen Erwartungen.

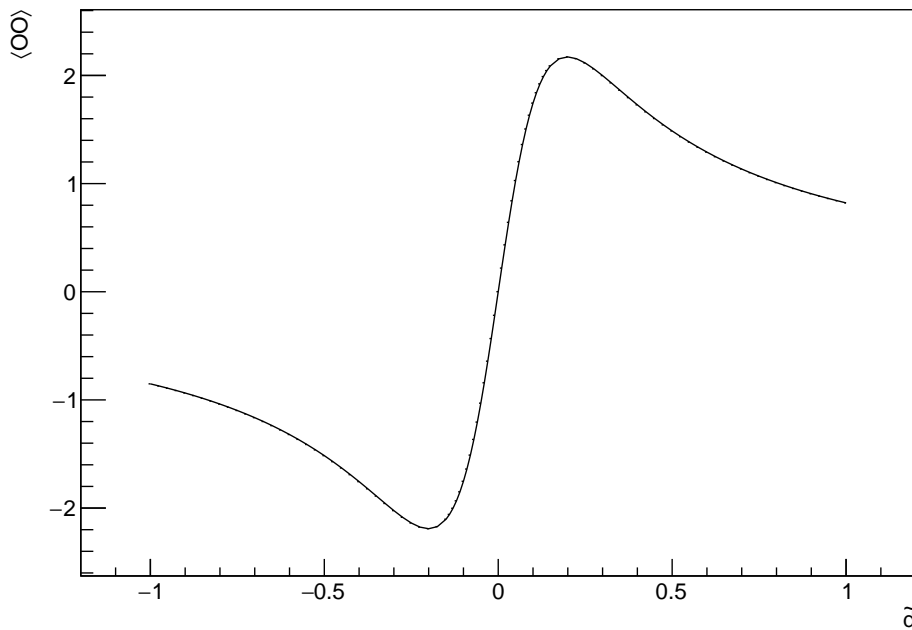


Abbildung 17: Ein Beispiel einer Eichkurve; Hierfür wurden die mit dem ersten neuronalen Netzwerk NN_1 rekonstruierten Higgs-Boson-Viererimpulse verwendet.

Mit den sechs verschiedenen Optimalen Observablen ($\mathcal{O}_{\text{NN}_1}, \mathcal{O}_{\text{NN}_2}, \mathcal{O}_{\text{NN}_3}, \mathcal{O}_{\text{NN}_4}, \mathcal{O}_{\text{MMC}}, \mathcal{O}_{\text{truth}}$) wurden für die weitere Analyse sechs verschiedene Eichkurven untersucht. Für einen Sensitivitätsvergleich zwischen den einzelnen Eichkurven wurden lineare Fits in einem Bereich von $\tilde{d} = [-0,04, 0,04]$ angefertigt. Die sechs verschiedenen Eichkurven mit den linearen Fits sind in Abbildung 18 gezeigt. Hierbei spricht eine große Steigung a und ein vergleichsweise kleiner Fehler auf die Mittelwerte der Observablen $s_{\langle \mathcal{O} \rangle}$ für eine hohe Sensitivität. Der Fehler auf \tilde{d} ergibt sich aus dem erwarteten linearen Verlauf der Eichkurve für kleine \tilde{d} -Werte

$$\tilde{d} = \frac{\langle \mathcal{O} \rangle}{a} \quad (55)$$

unter der Annahme, dass der Mittelwert der Optimalen Observable bei $\tilde{d} = 0$ Null beträgt bzw. der y -Achsenabschnitt gleich Null ist. Diese Annahme wurde auch bestätigt. Denn

wie in Tabelle 18 ersichtlich ist, ist der Mittelwert der Optimalen Observable bei $\tilde{d} = 0$ für alle Modelle mit Null verträglich.

Für den Fehler auf \tilde{d} gilt folgende Formel:

$$s_{\tilde{d}} = \frac{s_{\langle \mathcal{O} \mathcal{O} \rangle}}{a} \quad (56)$$

Hierbei wird die Unsicherheit in a vernachlässigt. Der bei einer integrierten Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ erwartete Fehler auf \tilde{d} berechnet sich schließlich über

$$s_{\tilde{d},\text{exp}} = \frac{s_{\langle \mathcal{O} \mathcal{O} \rangle, \text{exp}}}{a}, \quad (57)$$

wobei das Vorgehen zur Berechnung von $s_{\langle \mathcal{O} \mathcal{O} \rangle, \text{exp}}$ in Kapitel 8.2 aufgezeigt ist.

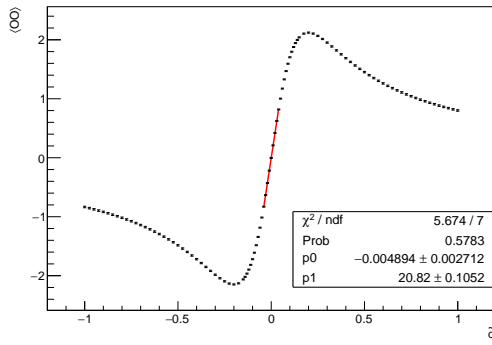
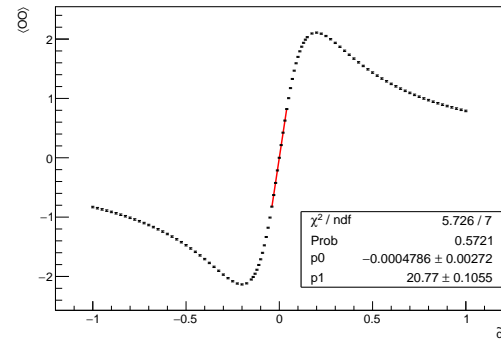
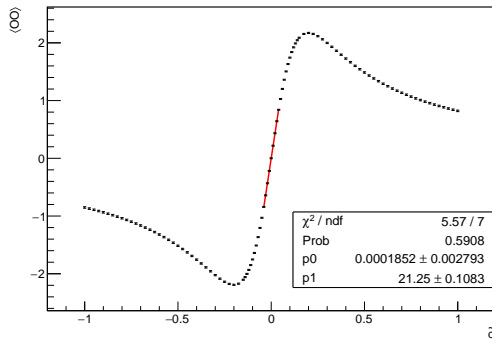
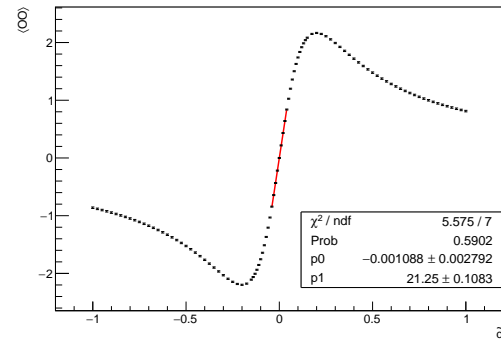
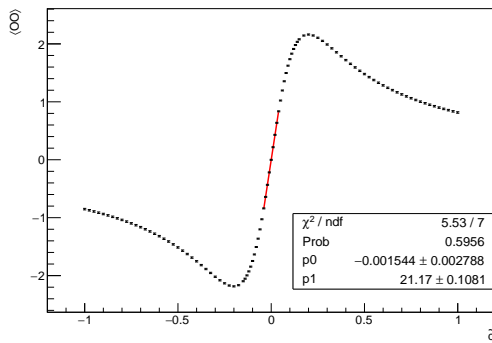
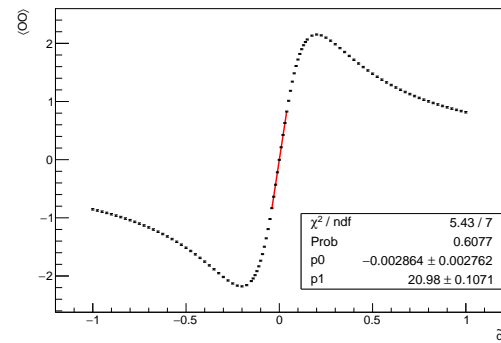
(a) Linearer Fit an die Eichkurve der wahren Optimalen Observable $\langle \mathcal{O}_{\text{truth}} \rangle$ (b) Linearer Fit an die Eichkurve der Optimalen Observablen des MMC $\langle \mathcal{O}_{\text{MMC}} \rangle$ (c) Linearer Fit an die Eichkurve der Optimalen Observable des ersten neuronalen Netzes $\langle \mathcal{O}_{\text{NN}_1} \rangle$ (d) Linearer Fit an die Eichkurve der Optimalen Observable der zweiten neuronalen Netzes $\langle \mathcal{O}_{\text{NN}_2} \rangle$ (e) Linearer Fit an die Eichkurve der Optimalen Observable des dritten neuronalen Netzes $\langle \mathcal{O}_{\text{NN}_3} \rangle$ (f) Linearer Fit an die Eichkurve der Optimalen Observable des vierten neuronalen Netzes $\langle \mathcal{O}_{\text{NN}_4} \rangle$

Abbildung 18: Eichkurven mit linearen Fits im Bereich von $\tilde{d} = [-0,04, 0,04]$ (in rot markiert); Die eingezeichneten Fehlerbalken bilden $s_{\tilde{d}}$ ab. Der χ^2/ndf -Wert in den Statistikboxen gibt den χ^2 -Wert im Verhältnis zu der Anzahl an Freiheitsgraden (ndf) wieder. prob gibt die χ^2 -Wahrscheinlichkeit an. p_0 und p_1 geben die Werte der Fitparameter mit den jeweiligen Unsicherheiten wieder, wobei p_0 den y -Achsenabschnitt und p_1 die Steigung der Gerade angibt.

Die Ergebnisse aller linearen Fits für die Steigung a und den zugehörigen Fehler s_a , sowie die Güte des Fits, die mit der χ^2 -Wahrscheinlichkeit (in Tabelle: prob) angegeben wird, sind in Tabelle 18 zusammengefasst. Die Mittelwerte der sechs verschiedenen Optimalen Observablen ($\langle \mathcal{O} \rangle_{\text{NN}_1}, \langle \mathcal{O} \rangle_{\text{NN}_2}, \langle \mathcal{O} \rangle_{\text{NN}_3}, \langle \mathcal{O} \rangle_{\text{NN}_4}, \langle \mathcal{O} \rangle_{\text{MMC}}, \langle \mathcal{O} \rangle_{\text{truth}}$) für $\tilde{d} = 0$ und ihre Fehler $s_{\langle \mathcal{O} \rangle}$ sind dort ebenfalls zusammengefasst. Außerdem wird hier auch der bei einer Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ erwartete Fehler des Mittelwerts der Optimalen Observablen $s_{\langle \mathcal{O} \rangle, \text{exp}}$ angegeben. Hinzu kommt der Fehler $s_{\tilde{d}}$ auf \tilde{d} und der bei einer Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ erwartete Fehler $s_{\tilde{d}, \text{exp}}$ gemäß Gleichung 57. Des Weiteren ist die Anzahl der erwarteten Ereignisse N_{exp} in der Tabelle enthalten.

| | $\langle \mathcal{O} \rangle$ | $s_{\langle \mathcal{O} \rangle}$ | $s_{\tilde{d}}$ | a | s_a | prob | $s_{\langle \mathcal{O} \rangle, \text{exp}}$ | $s_{\tilde{d}, \text{exp}}$ | N_{exp} |
|-----------------|-------------------------------|-----------------------------------|-----------------|-------|-------|------|---|-----------------------------|------------------|
| truth | -0,004533 | 0,0081 | 0,000390 | 20,82 | 0,105 | 0,58 | 0,65 | 0,0311 | 29,37 |
| MMC | -0,000002 | 0,0081 | 0,000392 | 20,77 | 0,106 | 0,57 | 0,65 | 0,0313 | 29,46 |
| NN ₁ | 0,000602 | 0,0084 | 0,000394 | 21,25 | 0,108 | 0,59 | 0,67 | 0,0314 | 29,30 |
| NN ₂ | -0,000519 | 0,0084 | 0,000394 | 21,25 | 0,108 | 0,59 | 0,67 | 0,0314 | 29,29 |
| NN ₃ | -0,001125 | 0,0084 | 0,000395 | 21,17 | 0,108 | 0,59 | 0,67 | 0,0314 | 29,31 |
| NN ₄ | -0,003015 | 0,0084 | 0,000395 | 21,17 | 0,108 | 0,60 | 0,67 | 0,0315 | 29,26 |

Tabelle 18: In dieser Tabelle stehen die Ergebnisse der Eichkurven für $\tilde{d} = 0$, die durch die neuronalen Netze NN_i erzielt wurden, im direkten Vergleich zu den Ergebnissen des Missing Mass Calculators (MMC) und zu den Ergebnissen auf Wahrheitsniveau (truth). Die Beschreibung der einzelnen hier aufgezeigten Variablen erfolgt im beiliegenden Text.

Die χ^2 -Wahrscheinlichkeiten aller Fits liegen in einem ähnlichen Größenbereich von rund 60%, was dafür spricht, dass das lineare Modell die zugrunde liegenden Daten gut beschreiben kann. Des Weiteren ist in der Tabelle zu erkennen, dass der Mittelwert der Optimalen Observable $\langle \mathcal{O} \rangle$ für alle Modelle innerhalb einer Standardabweichung mit Null verträglich ist und die Ergebnisse somit mit der SM-Vorhersage von $\tilde{d} = 0$ übereinstimmen, so dass kein Anzeichen einer CP-Verletzung zu beobachten ist. Ebenfalls ist anzumerken, dass die Anzahl an erwarteten Ereignissen N_{exp} von Modell zu Modell leicht unterschiedlich ist. Dies liegt daran, dass die Werte der Optimalen Observable in jedem Modell individuell auf $|\mathcal{O}_{\text{reco}}| < 12$ beschränkt wurden und somit für jedes Modell unterschiedlich viele Ereignisse in diesem Wertebereich liegen.

Eine große Steigung a und ein vergleichsweise kleiner Fehler auf den Mittelwert der Optimalen Observable $s_{\langle \mathcal{O} \rangle}$ sprechen für eine hohe Sensitivität, da dies zu einem kleinen Fehler auf \tilde{d} führt. Die Steigungen der linearen Fits, die aus den Daten der neuronalen Netze folgen, sind leicht größer als die Steigungen, die aus den Daten des MMCs und den Daten auf Wahrheitsniveau folgen. Gleichzeitig ist jedoch bei den neuronalen Netzwerken der Fehler auf den Mittelwert der Optimalen Observablen $s_{\langle \mathcal{O} \rangle}$ um etwa 3,7% größer im Vergleich zum MMC und dem *truth*-Modell. Dies spiegelt sich auch in den bei einer integrierten Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ erwarteten Fehlern auf die Mittelwerte der Optimalen Observablen $s_{\langle \mathcal{O} \rangle, \text{exp}}$ wider.

Um nun die Sensitivität für die Messung von \tilde{d} der fünf Rekonstruktionsmethoden mit dem *truth*-Modell zu vergleichen, wird der Fehler $s_{\tilde{d}}$ bzw. der erwartete Fehler $s_{\tilde{d}, \text{exp}}$ auf \tilde{d} betrachtet. Hier ist zu erkennen, dass der erwartete Fehler auf \tilde{d} im Falle des MMCs im Vergleich zum Modell auf Wahrheitsniveau um 0,6% größer ist. Die Eichkurven der ersten

drei neuronalen Netze hingegen liefern im Vergleich zum *truth*-Modell einen um 1,0% größeren erwarteten Fehler auf \tilde{d} , während die Eichkurve des vierten neuronalen Netzes in einem um 1,3% größeren erwarteten Fehler auf \tilde{d} resultiert. Vergleicht man die fünf Rekonstruktionsmethoden untereinander, so lässt sich feststellen, dass die Eichkurve des MMCs im Gegensatz zu denen der ersten drei neuronalen Netze einen um etwa 0,3% kleineren erwarteten Fehler $s_{\tilde{d},\text{exp}}$ liefert. Die Eichkurve des vierten neuronalen Netzes führt zu einem um etwa 0,6% größeren erwarteten Fehler auf \tilde{d} im Vergleich zum MMC.

Trotz der größeren Steigungen, die die Eichkurven der neuronalen Netze aufweisen, konnte somit mit den neuronalen Netzen keine Erhöhung der Sensitivität in \tilde{d} erreicht werden, da im Falle aller neuronalen Netze keine Verringerung des Fehlers $s_{\tilde{d}}$ bzw. des erwarteten Fehlers $s_{\tilde{d},\text{exp}}$ auf \tilde{d} im Vergleich zum MMC beobachtbar. Die Optimalen Observablen der vier neuronalen Netze sind somit nicht sensitiver auf eine mögliche CP-Verletzung als die des MMCs.

Die leichte Verschlechterung der Sensitivität in \tilde{d} der Optimalen Observablen der ersten beiden neuronalen Netze ist bemerkenswert, denn wie in Kapitel 7.2.2 dargestellt, konnte eine signifikante Verbesserung in der Auflösung der Optimalen Observable von rund 10% mit den ersten beiden neuronalen Netzen erreicht werden. Es gilt zu untersuchen, welche Gründe es dafür geben könnte, dass sich die Verbesserung in der Auflösung der Optimalen Observable nicht in einer Verbesserung der Sensitivität in \tilde{d} widerspiegelt. Hierfür wurden die in Kapitel 7.2.2 gezeigten Auflösungsabbildungen genauer untersucht. Der Referenzwert jeder Auflösungsabbildung ist die Optimale Observable, die mit den Higgs-Viererimpulsen auf Wahrheitsniveau, sowie den Viererimpulsen der Jets - ebenfalls auf Wahrheitsniveau - berechnet wurde. In Tabelle 19 ist für jede Rekonstruktions-Methode der RMS-Wert der $\mathcal{OO}_{\text{reco}}$ -Verteilung und der RMS-Wert der Referenz-Verteilung auf Wahrheitsniveau dargestellt. Der Wert des RMS der Referenz-Verteilung ändert sich für jedes Rekonstruktions-Modell (MMC und die neuronalen Netze), da die Werte der Optimalen Observable in jedem Modell individuell auf $|\mathcal{OO}_{\text{reco}}| < 12$ beschränkt wurden und alle Ereignisse verworfen wurden, für die die Optimale Observable nicht in diesem Wertebereich liegt. Dies hat zur Folge, dass für jedes Rekonstruktions-Modell unterschiedliche Werte der Optimalen Observable auf Wahrheitsniveau bzw. unterschiedliche Ereignisse verworfen wurden.

Des Weiteren ist eine Auffälligkeit im Falle des MMC zu erwähnen: der RMS-Wert der Referenz-Verteilung auf Wahrheitsniveau ist größer als der RMS-Wert der Optimalen-Observablen-Verteilung des MMCs selbst. Dies bedeutet, dass die Streuung der Werte in der Verteilung auf Wahrheitsniveau größer ist, als die in der Verteilung auf Rekonstruktionsniveau, was zunächst keinen Sinn ergibt. Der Grund hierfür ist, dass die Optimalen Observablen der Rekonstruktions-Modelle auf $|\mathcal{OO}_{\text{reco}}| < 12$ geschnitten wurde. Die Werte der Optimalen Observablen auf Wahrheitsniveau, die hier als Referenz dienen, wurden jedoch nicht eingeschränkt, wodurch Werte von $|\mathcal{OO}_{\text{truth}}| > 12$ in den RMS miteingehen. Würde man für die Optimale Observable auf Wahrheitsniveau ebenfalls $|\mathcal{OO}_{\text{truth}}| < 12$ fordern, wäre der RMS-Wert dieser Verteilung wieder kleiner als der RMS-Wert der \mathcal{OO} -Verteilung des MMCs.

| | RMS($\mathcal{O}\mathcal{O}$) | RMS($\mathcal{O}\mathcal{O}_{\text{ref}}$) |
|-----------------|---------------------------------|--|
| MMC | 3,52 | 3,72 |
| NN ₁ | 3,61 | 3,58 |
| NN ₂ | 3,61 | 3,58 |
| NN ₃ | 3,60 | 3,61 |
| NN ₄ | 3,61 | 3,62 |

Tabelle 19: Die Tabelle zeigt für jede Rekonstruktions-Methode den RMS-Wert der $\mathcal{O}\mathcal{O}_{\text{reco}}$ -Verteilungen (reco = (MMC, NN_i)) und den RMS-Wert der Referenz-Verteilung $\mathcal{O}\mathcal{O}_{\text{ref}}$ auf Wahrheitsniveau. In die Berechnung von $\mathcal{O}\mathcal{O}_{\text{ref}}$ gingen die wahren Viererimpulse der zwei Jets auf Teilchenniveau und der wahre Viererimpuls des Higgs-Bosons ein.

Zukünftige Studien und eine Variation der Signalselektion sollten ergeben, wie die verbesserte Rekonstruktion des Higgs-Boson-Viererimpulses mit neuronalen Netzwerken auch optimal für die Messung von \tilde{d} genutzt werden könnte.

9 Zusammenfassung

Ziel dieser Arbeit war es, die Sensitivität von CP-Tests im VBF $H \rightarrow \tau_e \tau_\mu$ Kanal unter der Annahme von einer integrierten Luminosität von $\mathcal{L} = 139 \text{ fb}^{-1}$ von pp-Kollisionen bei $\sqrt{s} = 13 \text{ TeV}$ am ATLAS-Experiment zu verbessern. Hierzu sollten Methoden des maschinellen Lernens für die Rekonstruktion des Higgs-Boson-Viererimpulses angewendet werden.

In der vorliegenden Arbeit wurde der Higgs-Viererimpuls mit neuronalen Netzen rekonstruiert. Hierbei wurden vier verschiedene Sätze von Eingangs-Variablen definiert. Der Satz des ersten neuronalen Netzes NN_1 enthielt die kartesischen Komponenten des Viererimpulses der zwei Leptonen, sowie die p_x - und p_z -Komponente der fehlenden Transversalenergie. Beim zweiten neuronalen Netz NN_2 kommen zu den eben genannten Eingangs-Variablen noch die kartesischen Komponenten des Viererimpuls der beiden Jets hinzu. Die Ziel-Variablen der ersten beiden Netze waren die kartesischen Komponenten des Higgs-Boson-Viererimpulses. Das dritte neuronale Netz NN_3 enthält die gleichen Eingangs-Variablen wie das erste Netz, jedoch sind diese hier nicht in kartesischen Koordinaten, sondern in Polarkoordinaten gegeben. Außerdem sind hier die entsprechenden ϕ -Winkel zu einem gemeinsamen Nullpunkt hin rotiert, die dann in Sinus- und Cosinus-Funktionen eingesetzt werden. Der Satz der Eingangs-Variablen des vierten Netzes NN_4 gleicht letzterem, wobei hier noch die Viererimpulse der beiden Jets - ebenfalls in Polarkoordinaten - berücksichtigt werden. Für die ϕ -Winkel gilt gleiches wie beim dritten neuronalen Netz. Die Ziel-Variablen der letzten beiden Netze waren die polaren Komponenten des Higgs-Viererimpulses.

Für jedes Modell NN_i wurden einige Hyperparameter, nämlich die Anzahl der Schichten, die Knoten pro Schicht, die Lernrate und die Batch-Größe, mittels OPTUNA optimiert. Mit den Informationen der vier OPTUNA-Optimierungen wurden die unterschiedlichen Netze trainiert. Die vier trainierten Netze wurden dann genutzt, um die Higgs-Viererimpulse vorherzusagen, wobei die Qualität dieser mittels Auflösungsabbildungen mit der Qualität der Viererimpulse, die durch den Missing Mass Calculator rekonstruiert wurden, verglichen wurde. Hierbei war für die ersten beiden neuronalen Netze eine signifikante Verbesserung der Auflösung in allen Komponenten des Higgs-Viererimpulses zu beobachten. Mit dem ersten neuronalen Netz konnte in p_T eine prozentuale Verbesserung gegenüber dem MMC von ca. 27% , in η eine Verbesserung von ca. 11%, in ϕ eine Verbesserung von ca. 18% und in M eine Verbesserung von etwa 45% erreicht werden.

Mit den rekonstruierten Higgs-Viererimpulsen der vier verschiedenen Modelle wurde dann für jedes Modell individuell die Optimale Observable, die auf $|\mathcal{OO}| < 12$ geschnitten wurde, berechnet. Daraufhin wurde die Auflösung der Optimalen Observable der vier Netze mit der Auflösung der Optimalen Observable des MMC verglichen, wobei mit dem ersten Netz eine Verbesserung von etwa 11% und mit dem zweiten Netz eine Verbesserung von etwa 10% erzielt werden konnte.

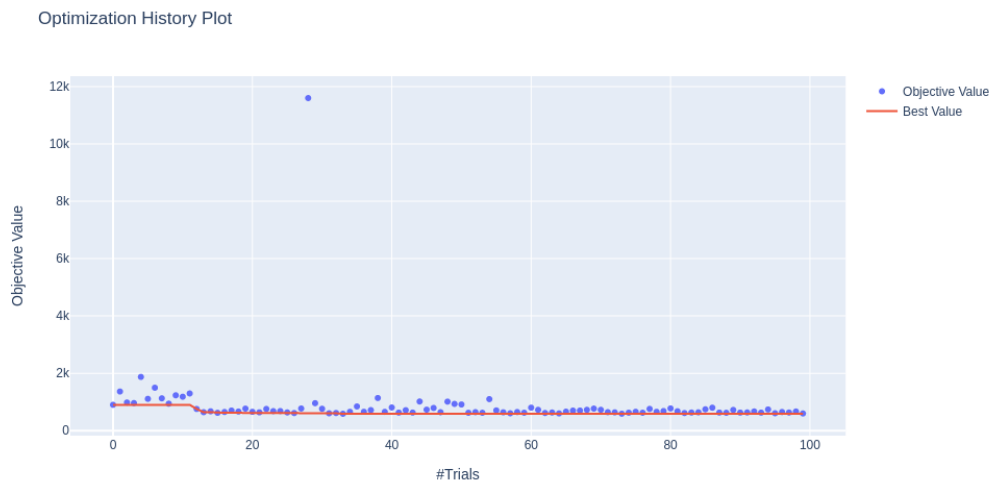
Mittels Eichkurven wurde dann der Zusammenhang zwischen den Mittelwerten der Verteilungen der CP-ungeraden Optimalen Observable $\langle \mathcal{OO} \rangle$ und \tilde{d} bestimmt. Hierfür wurde die Optimale Observable auf Wahrheitsniveau, sowie die Optimale Observable der vier verschiedenen neuronalen Modelle und des MMC betrachtet. Die Eichkurven für kleine \tilde{d} zeigen einen linearen Verlauf und für jede Eichkurve wurde ein linearer Fit im Bereich von $\tilde{d} = [-0,04, 0,04]$ durchgeführt. Mit den Informationen der Eichkurven und mit den

Ergebnissen der linearen Fits konnte dann die Sensitivität in der Messung von \tilde{d} , der die Stärke der CP-Verletzung in HVV -Kopplungen parametrisiert, analysiert werden. Hierbei wurde festgestellt, dass die Optimalen Observablen der vier neuronalen Netze im Vergleich zum MMC gleich sensitiv sind und somit keine Verbesserung in der Sensitivität bewirkt werden konnte.

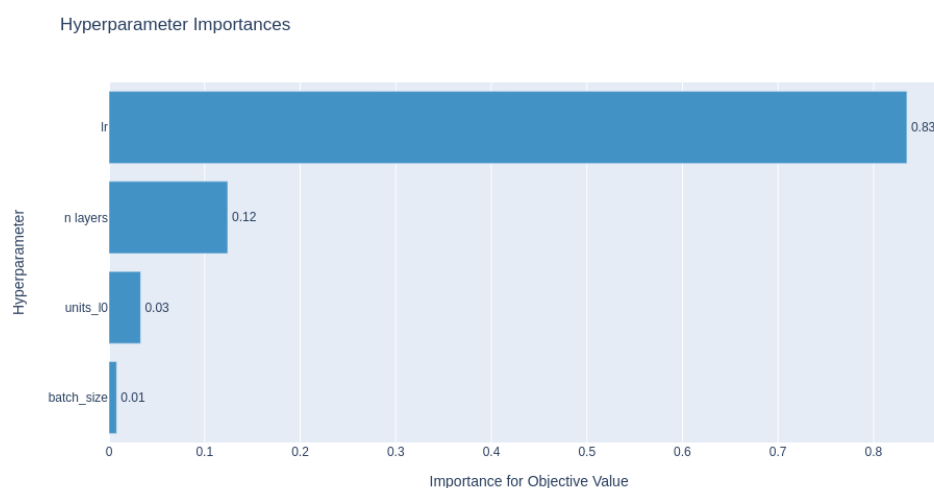
Trotz der positiven Ergebnisse bezüglich der verbesserten Rekonstruktionsgüte des Higgs-Boson-Viererimpulses und der verbesserten Auflösung in der Optimalen Observable, gilt es noch im Detail zu verstehen, warum sich diese positiven Ergebnisse nicht in einer Verbesserung in der Sensitivität widerspiegeln. Außerdem wäre es von Interesse, eine einzelne Zielvariable der letzten beiden Netze NN_3, NN_4 zu verändern, nämlich statt der Masse M , die Energie E als Zielvariable festzulegen. Dies würde möglicherweise zu einer besseren Gesamt-Leistung der Netze 3 und 4 führen und die große Diskrepanz in der Massenauflösung zwischen Netz 3 und 4 und Netz 1 und 2 auflösen. Des Weiteren wäre es interessant, die Effekte einer besseren Parton-Jet-Assoziation zu untersuchen, da die Optimale Observable, die mit den Viererimpulsen der Partonen berechnet wird, sehr sensitiv ist.

A Visualisierungsabbildungen OPTUNA

A.1 Neuronales Netz 2



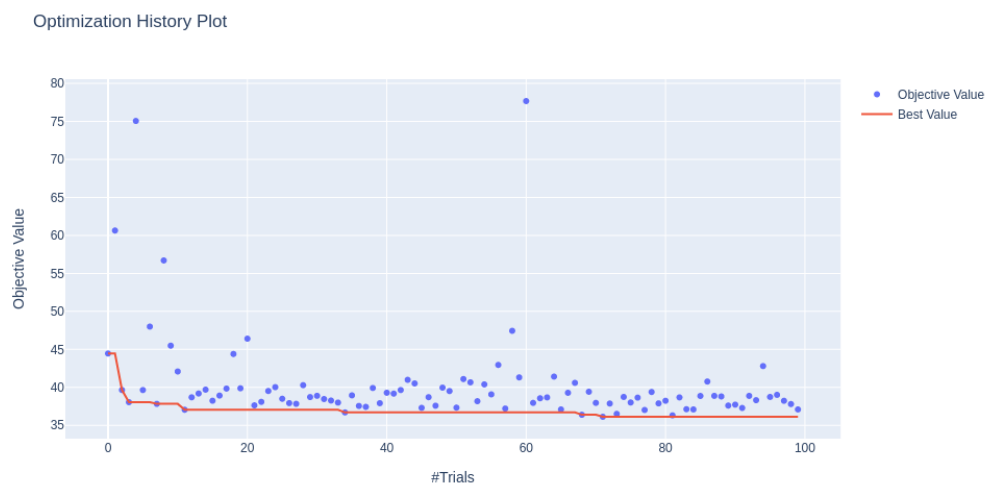
(a) Verlauf des OPTUNA-Optimierungsprozesses; $\# Trials$ gibt die Anzahl der Versuche an und $Objective Value$ beschreibt den Wert der Zielfunktion, welche in diesem Fall dem mittleren quadratischen Fehlers entspricht. Mit den blauen Punkten wird die Bewertung der Zielfunktion für jeden einzelnen Versuch gekennzeichnet. Die rote Linie verbindet die Bewertungen, die im Verlauf des Prozesses den niedrigsten Wert des mittleren quadratischen Fehlers geliefert haben. Die Bedeutung der Legende ist für die folgenden Abbildungen identisch und wird deshalb hier nur einmal erläutert. Man erkennt an der Abbildung, dass 100 Versuche für den Optimierungsprozess ausreichend waren, da schon nach wenigen Versuchen keine merkliche Verbesserung mehr erreicht werden konnte.



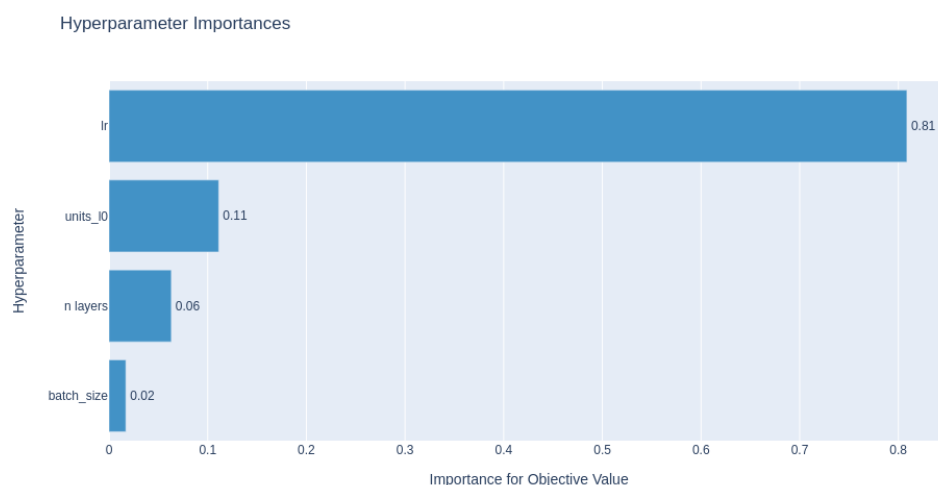
(b) Prozentuale Wichtigkeit (engl.: *Importance*) der einzelnen Hyperparameter für die Zielfunktion (engl. *Objective Function*; lr steht für die Lernrate, welche die größte Bedeutung für die Zielfunktion hat, gefolgt von der Anzahl der Schichten des Netzes $n layers$. $units l_0$ steht für die Anzahl der Knoten in der ersten Schicht des Netzes und $batch size$ gibt die Größe eines Batches an.

Abbildung 19: Abbildungen zur Visualisierung des OPTUNA-Optimierungsprozesses für das zweite neuronale Netz.

A.2 Neuronales Netz 3



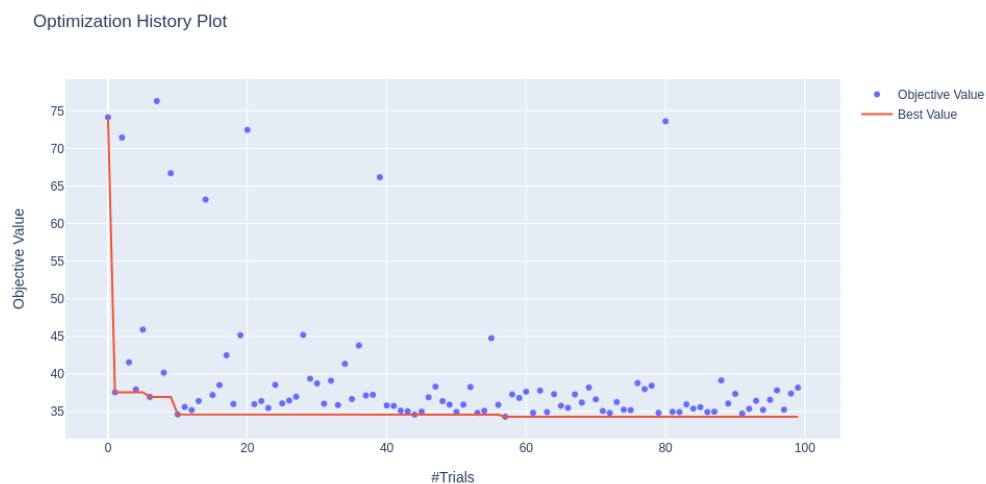
(a) Verlauf des OPTUNA-Optimierungsprozesses; $\# Trials$ gibt die Anzahl der Versuche an und *Objective Value* beschreibt den Wert der Zielfunktion (MSE). Auch hier sieht man, dass 100 Versuche ausreichend waren.



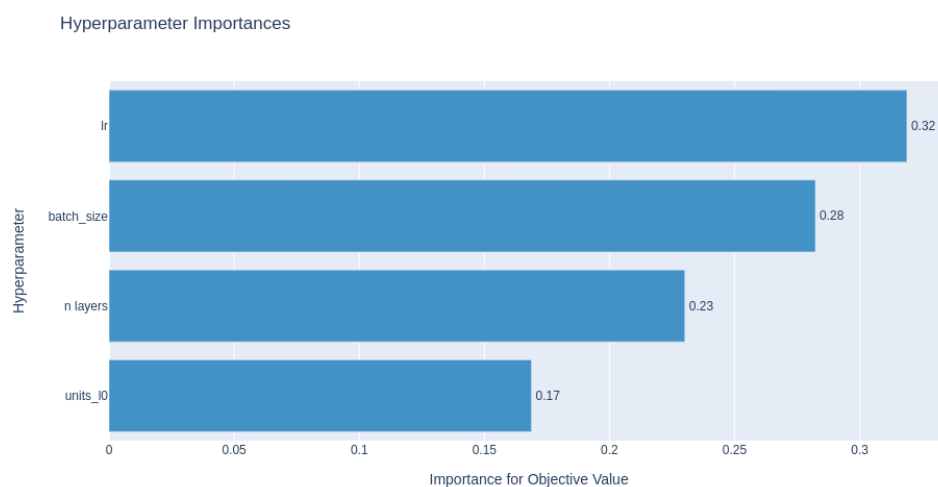
(b) Prozentuale Wichtigkeit (engl.: *Importance*) der einzelnen Hyperparameter für die Zielfunktion (engl. *Objective Function*). Den dominierenden Beitrag liefert hier die Lernrate lr . Die Anzahl der Knoten in der ersten Schicht des Netzes $units_{l0}$, die Gesamtanzahl der Schichten des Netzes $n\ layers$ und die Batch-Größe ($batch\ size$) haben relativ dazu eine deutlich kleinere Bedeutung.

Abbildung 20: Abbildungen zur Visualisierung des OPTUNA-Optimierungsprozesses für das dritte neuronale Netz.

A.3 Neuronales Netz 4



(a) Verlauf des OPTUNA-Optimierungsprozesses; $\# Trials$ gibt die Anzahl der Versuche an und $Objective Value$ beschreibt den Wert der Zielfunktion (MSE). Auch hier kann nach ca. 10 Versuchen keine signifikante Verbesserung mehr in den Bewertungen der Zielfunktion beobachtet werden.



(b) Prozentuale Wichtigkeit (engl.: *Importance*) der einzelnen Hyperparameter für die Zielfunktion (engl. *Objective Function*). Die Bedeutung der einzelnen Hyperparameter ist hier im Vergleich zu den oberen Abbildungen ausgeglichener.

Abbildung 21: Abbildungen zur Visualisierung des OPTUNA-Optimierungsprozesses für das vierte neuronale Netz.

B Verteilungen der Eingangsgrößen

B.1 Neuronales Netz 1

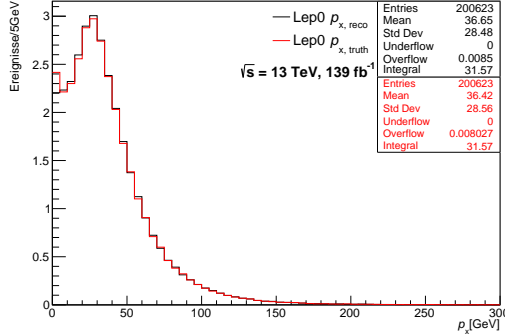
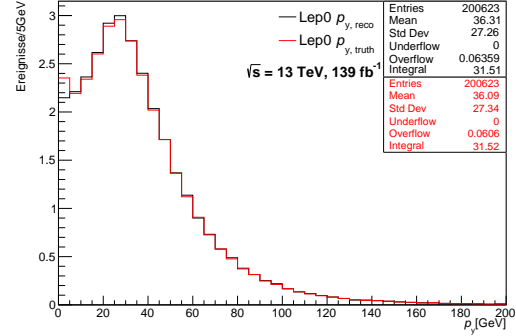
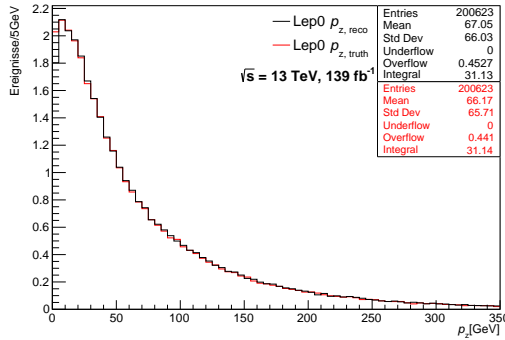
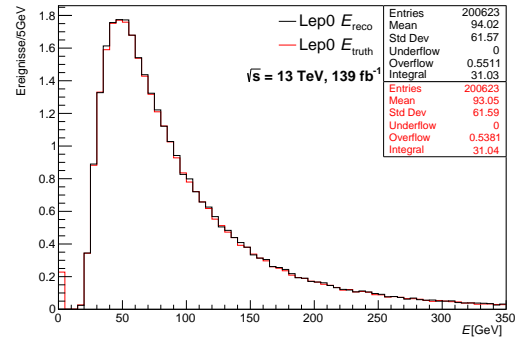
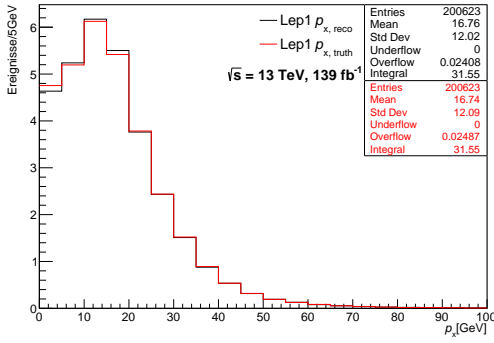
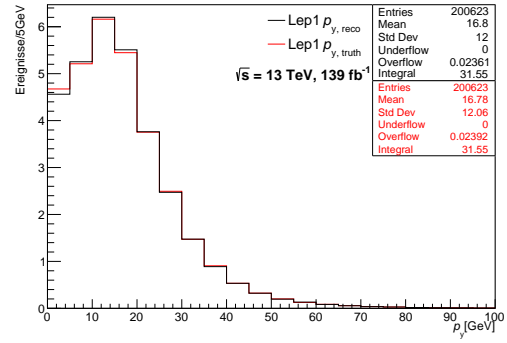
(a) Verteilung des Impulses in x -Richtung(b) Verteilung des Impulses in y -Richtung(c) Verteilung des Impulses in z -Richtung(d) Verteilung der Energie E ; Der Peak bei Null kommt durch falsch rekonstruierte Ereignisse zustande, die auf Wahrheitsniveau nicht gefunden werden können und daher der Null zugeordnet werden.

Abbildung 22: Alle hier gezeigten Verteilungen der Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des führenden Leptons (in Abb.: Lep0) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

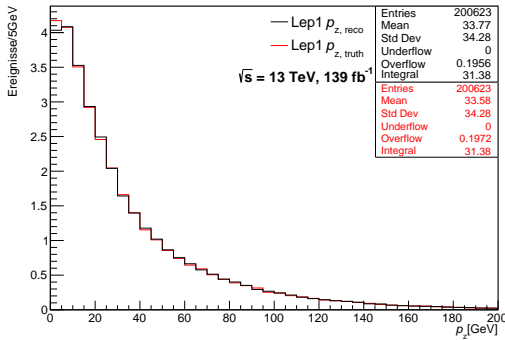
Die Statistikboxen sind in englischer Sprache dargestellt, wobei *Entries* die Anzahl der Einträge im Histogramm, *Mean* den Mittelwert der Verteilung und *Std Dev* die Standardabweichung bzw. den RMS der Verteilung angibt. Die Anzahl der Werte, die kleiner als die Untergrenze des vorliegenden Wertebereichs sind und daher nicht abgebildet werden, wird durch *Underflow* wiedergegeben, wobei *Overflow* die Anzahl der Werte angibt, die im positiven Bereich über den gezeichneten Wertebereich hinausgehen. Der *Integral*-Wert liefert das Integral der Bin-Inhalte innerhalb des gezeichneten Wertebereichs und gibt somit die Anzahl der erwarteten Ereignisse bei $\mathcal{L} = 139 \text{ fb}^{-1}$ an. Die Bedeutung der Einträge in den Statistikboxen ist in allen dieselbe und wird im Folgenden nicht mehr erläutert.



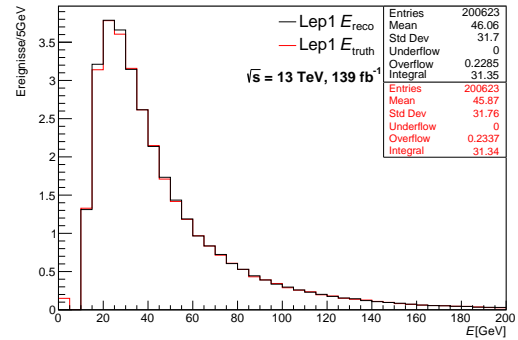
(a) Verteilung des Impulses in x -Richtung



(b) Verteilung des Impulses in y -Richtung

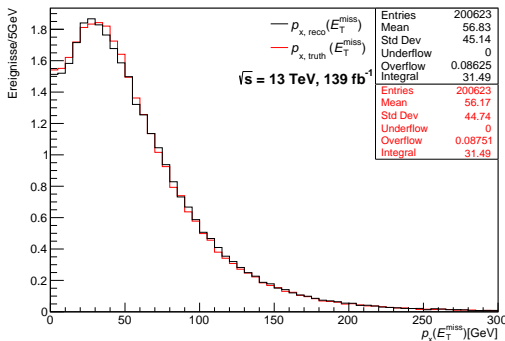


(c) Verteilung des Impulses in z -Richtung

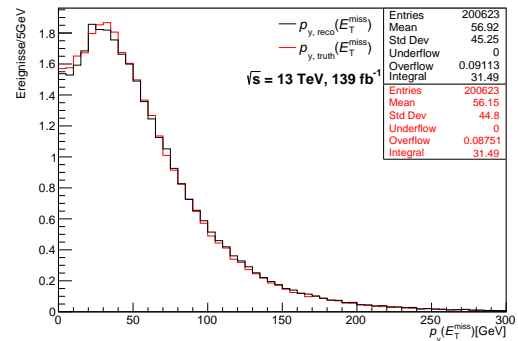


(d) Verteilung der Energie E ; Der Peak bei Null kommt durch falsch rekonstruierte Ereignisse zustande, die auf Wahrheitsniveau nicht gefunden werden können und daher der Null zugeordnet werden.

Abbildung 23: Alle hier gezeigten Verteilungen der Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des zweitführenden Leptons (in Abb.: Lep1) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.



(a) Verteilung des Impulses in x -Richtung



(b) Verteilung des Impulses in y -Richtung

Abbildung 24: Die hier gezeigten Verteilungen der Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) der fehlenden Transversalenergie E_T^{miss} sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

B.2 Neuronales Netz 2

Die Verteilungen der ersten Eingangs-Variablen des neuronalen Netzes 2 sind dieselben wie die des ersten neuronalen Netzes und können im obigen Unterkapitel eingesehen werden. Hinzu kommen die Verteilungen der Jets, die in den zwei folgenden Abbildungen dargestellt sind.

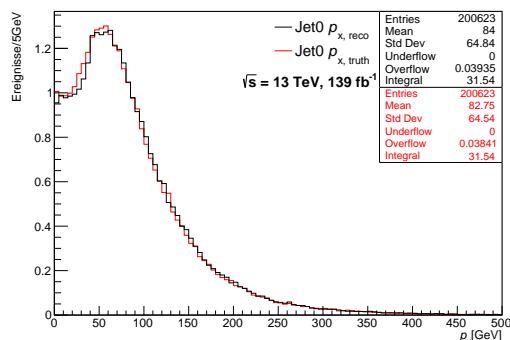
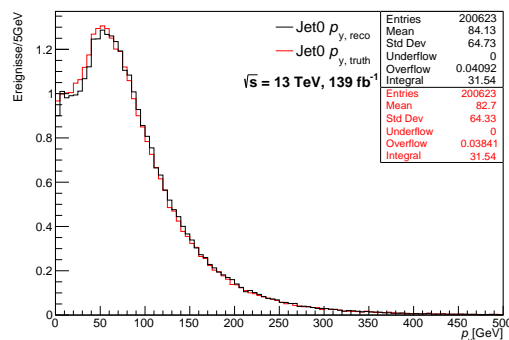
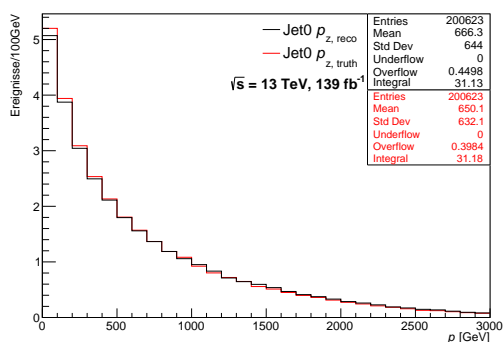
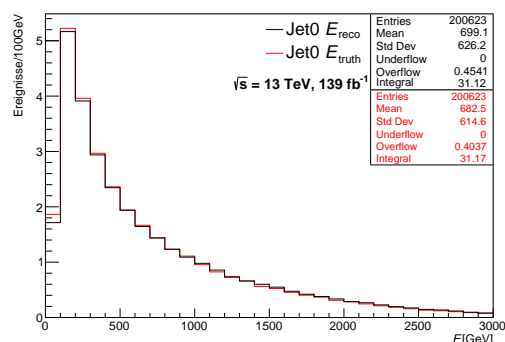
(a) Verteilung des Impulses in x -Richtung(b) Verteilung des Impulses in y -Richtung(c) Verteilung des Impulses in z -Richtung(d) Verteilung der Energie E

Abbildung 25: Alle hier gezeigten Verteilungen der Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des führenden Jets (in Abb.: Jet0) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

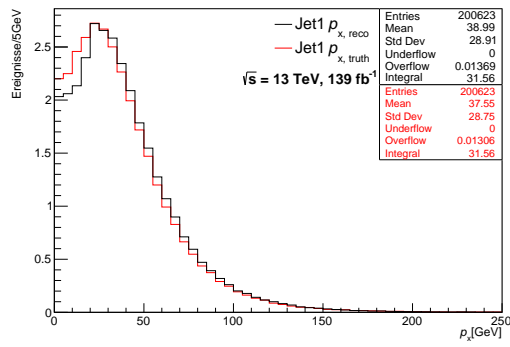
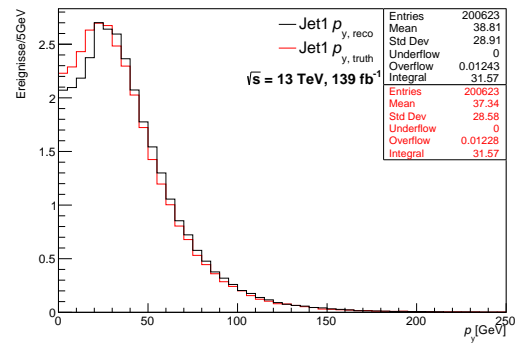
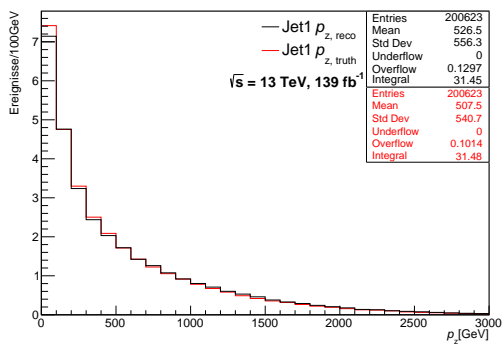
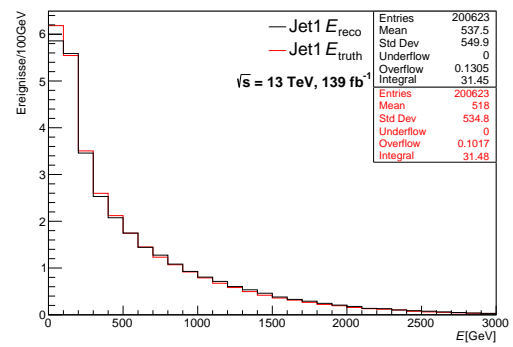
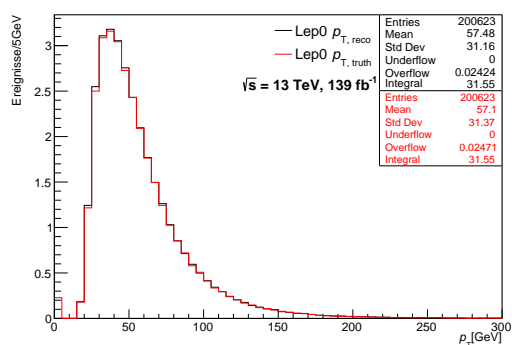
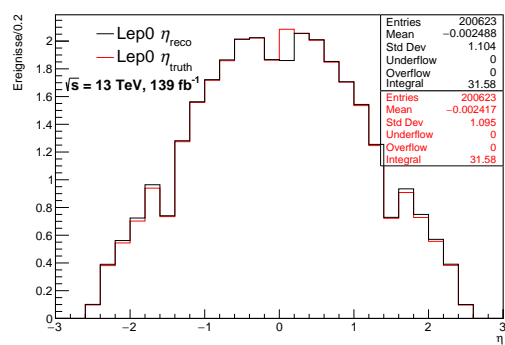
(a) Verteilung des Impulses in x -Richtung(b) Verteilung des Impulses in y -Richtung(c) Verteilung des Impulses in z -Richtung(d) Verteilung der Energie E

Abbildung 26: Alle hier gezeigten Verteilungen der Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des zweitführenden Jets (in Abb.: Jet1) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

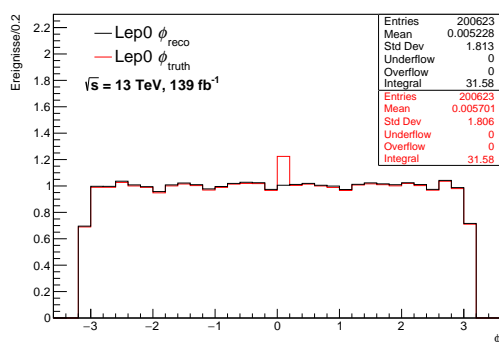
B.3 Neuronales Netz 3



(a) Verteilung des Transversalimpulses p_T ; Der Peak bei Null kommt durch falsch rekonstruierte Ereignisse zustande, die auf Wahrheitsniveau nicht gefunden werden können und daher der Null zugeordnet werden.



(b) Verteilung der Pseudorapidity η



(c) Verteilung des Azimuthwinkels ϕ

Abbildung 27: Die hier gezeigten Verteilungen der ersten drei Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des führenden Leptons (in Abb.: Lep0) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

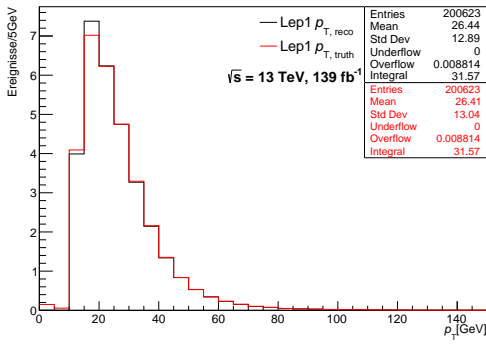
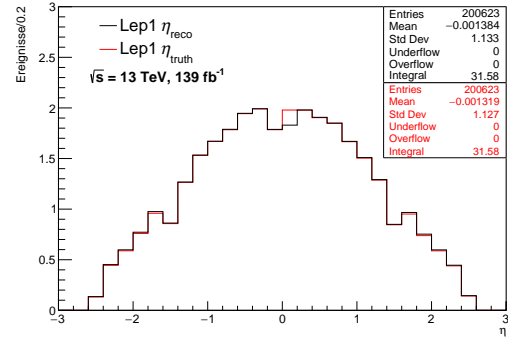
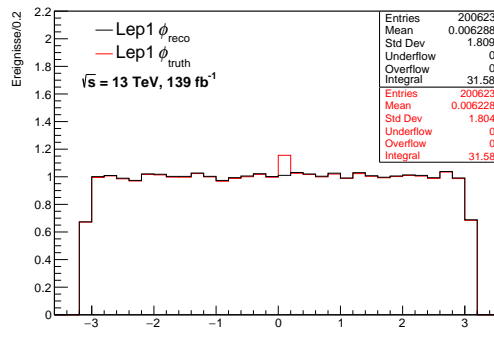
(a) Verteilung des Transversalimpulses p_T (b) Verteilung der Pseudorapidity η (c) Verteilung des Azimuthwinkels ϕ

Abbildung 28: Die hier gezeigten Verteilungen der ersten drei Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des zweitführenden Leptons (in Abb.: Lep1) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

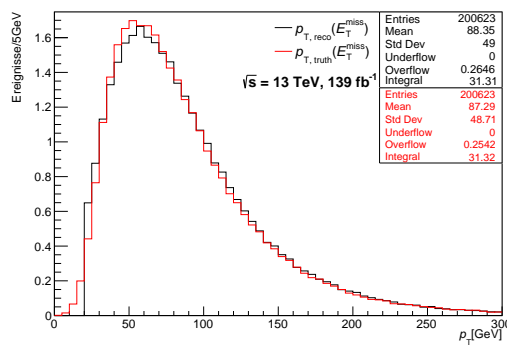


Abbildung 29: Die hier gezeigte Verteilung des Transversalimpulses p_T auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) der fehlenden Transversalenergie E_T^{miss} ist auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

B.4 Neuronales Netz 4

Die Verteilungen der ersten Eingangs-Variablen des neuronalen Netzes 4 sind dieselben wie die des dritten neuronalen Netzes und können im obigen Unterkapitel eingesehen

werden. Hinzu kommen die Verteilungen der Jets, die in den zwei folgenden Abbildungen dargestellt sind.

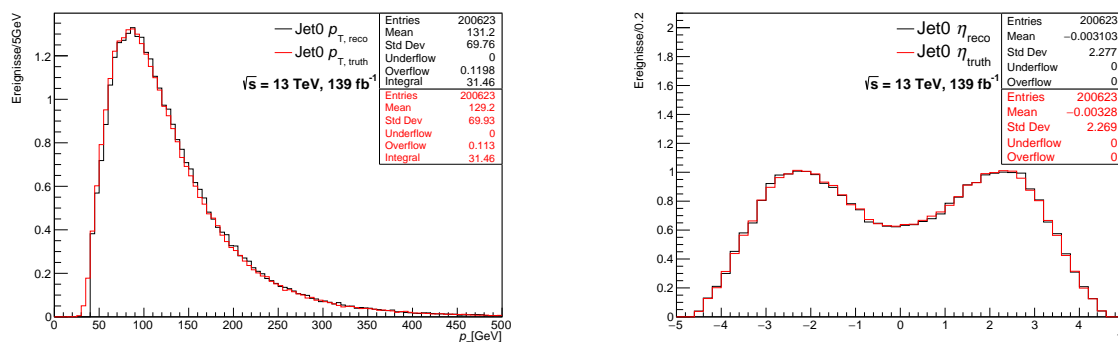
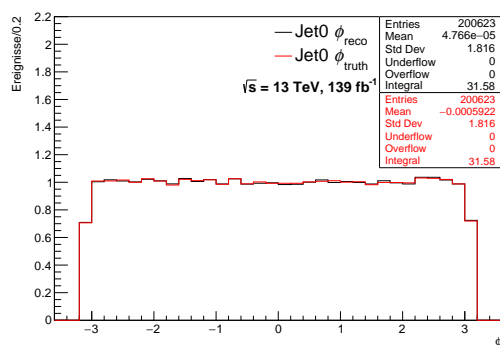
(a) Verteilung des Transversalimpulses p_T (b) Verteilung der Pseudorapidität η (c) Verteilung des Azimuthwinkels ϕ

Abbildung 30: Die hier gezeigten Verteilungen der ersten drei Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des führenden Jets (in Abb.: Jet0) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

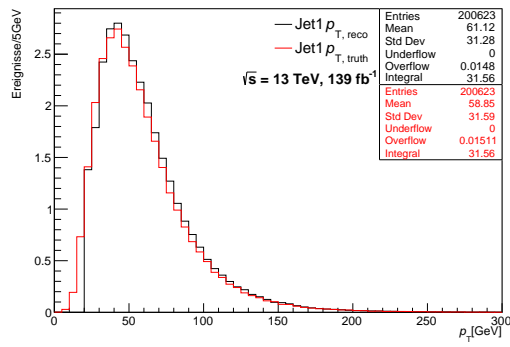
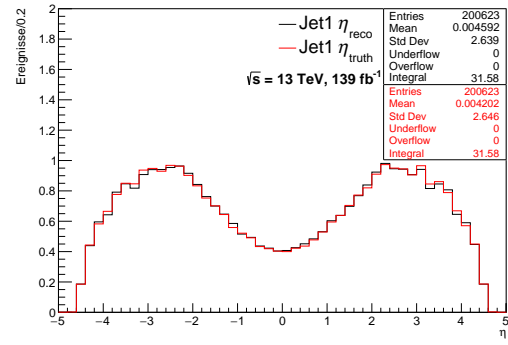
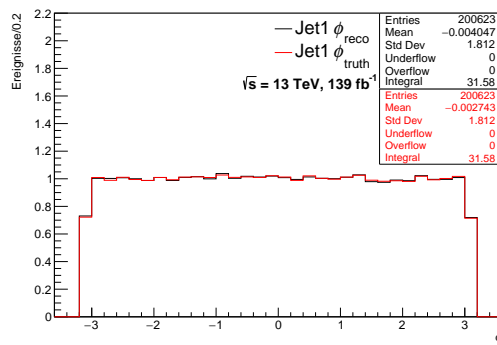
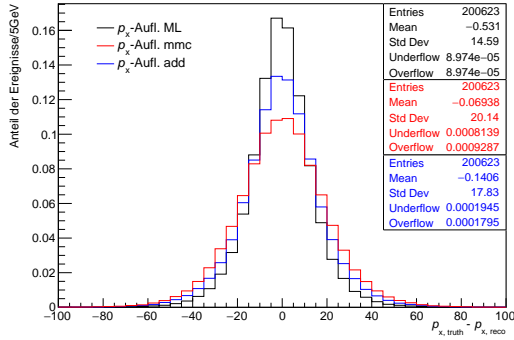
(a) Verteilung des Transversalimpulses p_T (b) Verteilung der Pseudorapidity η (c) Verteilung des Azimuthwinkels ϕ

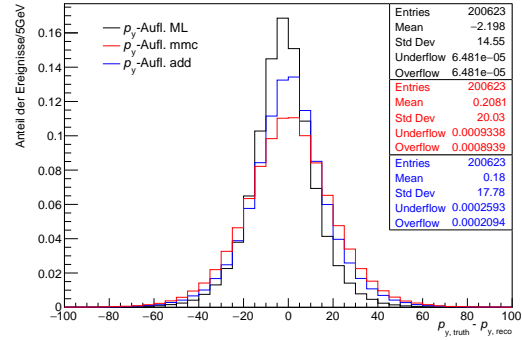
Abbildung 31: Die hier gezeigten Verteilungen der ersten drei Komponenten des Viererimpulses auf Wahrheitsniveau (*truth*) und Rekonstruktionsniveau (*reco*) des zweitführenden Jets (in Abb.: Jet1) sind auf $\mathcal{L} = 139 \text{ fb}^{-1}$ skaliert.

C Abbildungen zum Auflösungsvermögen des Higgs-Boson-Viererimpulses

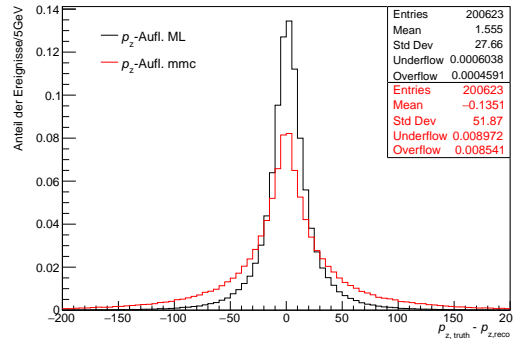
C.1 Neuronales Netz 1



(a) Auflösung bezüglich des Impulses in x -Richtung p_x



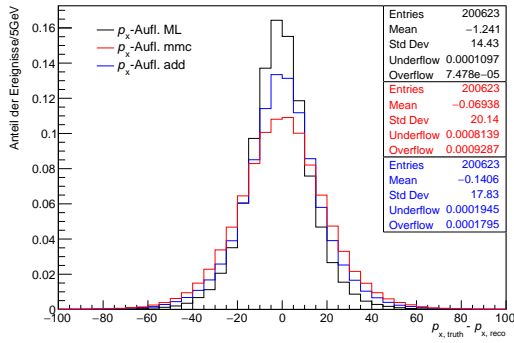
(b) Auflösung bezüglich des Impulses in y -Richtung p_y



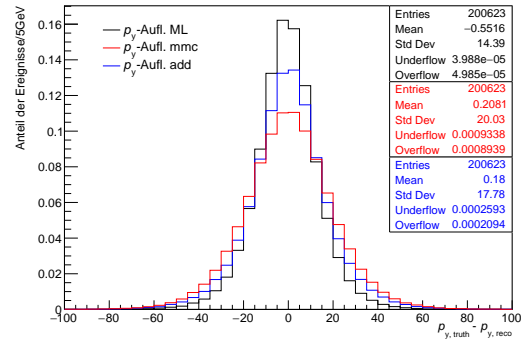
(c) Auflösung bezüglich des Impulses in z -Richtung p_z

Abbildung 32: Die restlichen auf Eins normierten Auflösungsabbildungen des ersten neuronalen Netzes (ML), des Missing Mass Calculators (MMC) und der Rekonstruktionsmethode des vektoriellen Aufaddierens der Higgs-Zerfallsteilchen (add); Gezeigt ist jeweils die Auflösung der p_x , p_y und p_z des rekonstruierten Higgs-Bosons. Die Statistikboxen sind in englischer Sprache dargestellt, wobei *Entries* die Anzahl der Einträge im Histogramm, *Mean* den Mittelwert der Verteilung und *Std Dev* die Standardabweichung bzw. den RMS der Verteilung angibt. Die Anzahl der Werte, die kleiner als die Untergrenze des in der Verteilungen vorliegenden Wertebereichs sind und daher nicht abgebildet werden, wird durch *Underflow* wiedergegeben, wobei *Overflow* die Anzahl der Werte angibt, die im positiven Bereich über den gezeichneten Wertebereich hinausgehen. Die Bedeutung der Einträge in den Statistikboxen ist in allen dieselbe und wird im Folgenden nicht mehr erläutert.

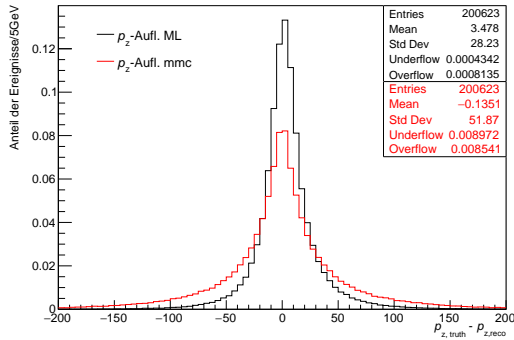
C.2 Neuronales Netz 2



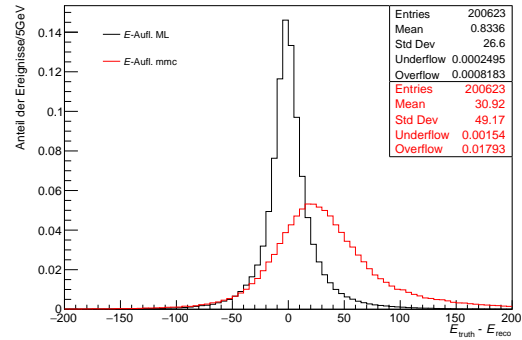
(a) Auflösung bezüglich des Impulses in x -Richtung p_x



(b) Auflösung bezüglich des Impulses in y -Richtung p_y



(c) Auflösung bezüglich des Impulses in z -Richtung p_z



(d) Energie-Auflösung

Abbildung 33: Die auf Eins normierten Auflösungsabbildungen des zweiten neuronalen Netzes (ML), des Missing Mass Calculators (MMC) und der Rekonstruktions-Methode des vektoriellen Aufaddierens der Higgs-Zerfallsteilchen (add); Gezeigt ist jeweils die Auflösung in p_x , p_y , p_z und der Energie E des rekonstruierten Higgs-Bosons.

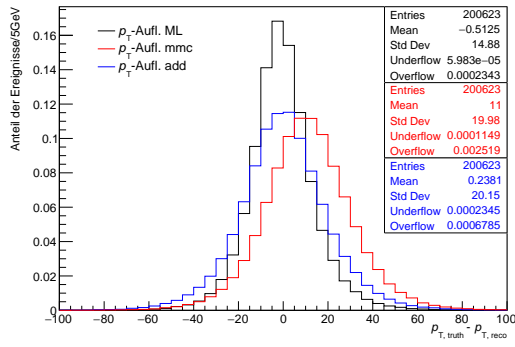
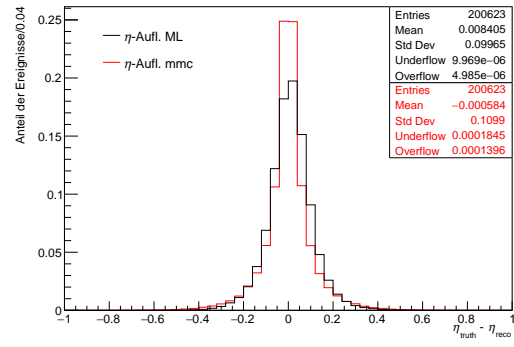
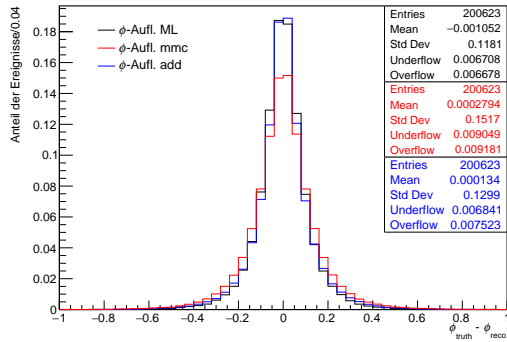
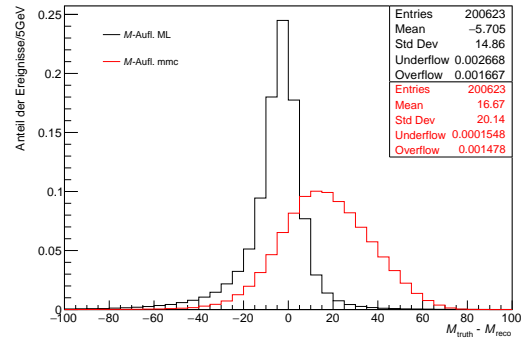
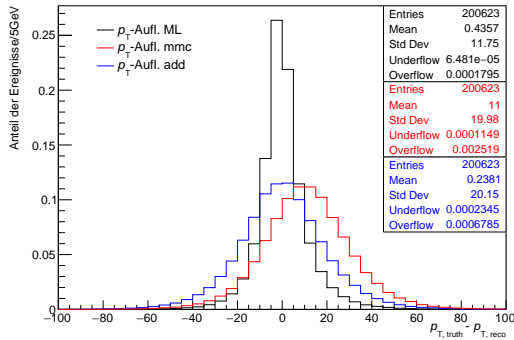
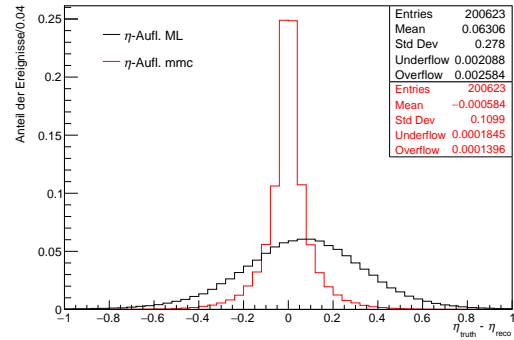
(a) Auflösung bezüglich des Transversalimpulses p_T (b) Auflösung bezüglich der Pseudorapidity η (c) Auflösung bezüglich dem Azimuthwinkel ϕ (d) Auflösung bezüglich der Masse M

Abbildung 34: Auf Eins normierte Auflösungsabbildungen des zweiten neuronalen Netzes (ML), des Missing Mass Calculators (MMC) und der Rekonstruktions-Methode des vektoriellen Aufaddierens der Higgs-Zerfallsteilchen (add); Gezeigt ist jeweils die Auflösung in p_T , η , ϕ und M des rekonstruierten Higgs-Bosons.

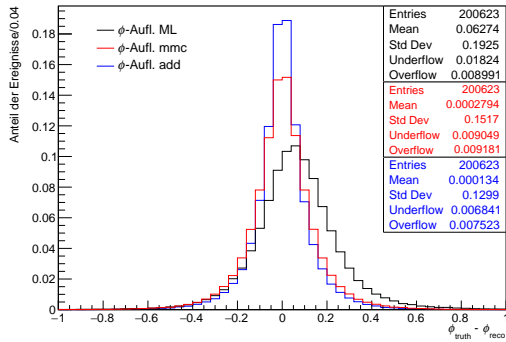
C.3 Neuronales Netz 3



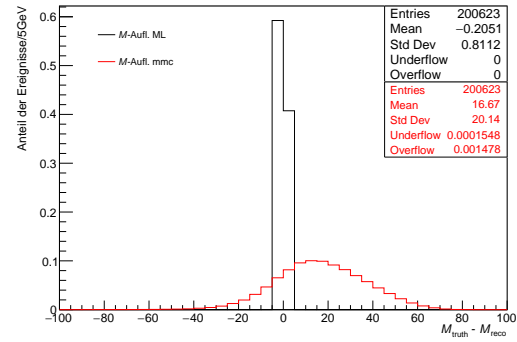
(a) Auflösung bezüglich des Transversalimpulses p_T



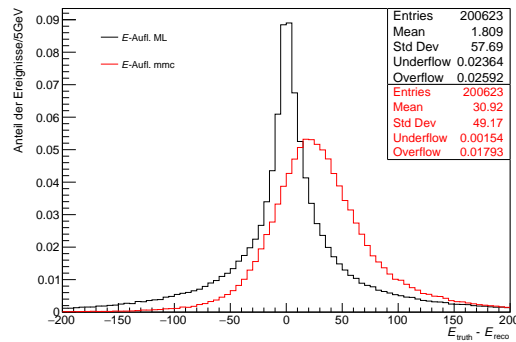
(b) Auflösung bezüglich der Pseudorapidität η



(c) Auflösung bezüglich dem Azimutwinkel ϕ



(d) Auflösung bezüglich der Masse M



(e) Auflösung bezüglich der Energie E

Abbildung 35: Auf Eins normierte Auflösungsabbildungen des dritten neuronalen Netzes (ML), des Missing Mass Calculators (MMC) und der Rekonstruktions-Methode des vektoriellen Aufaddierens der Higgs-Zerfallsteilchen (add); Gezeigt ist jeweils die Auflösung in p_T , η , ϕ , M , und E des rekonstruierten Higgs-Bosons.

C.4 Neuronales Netz 4

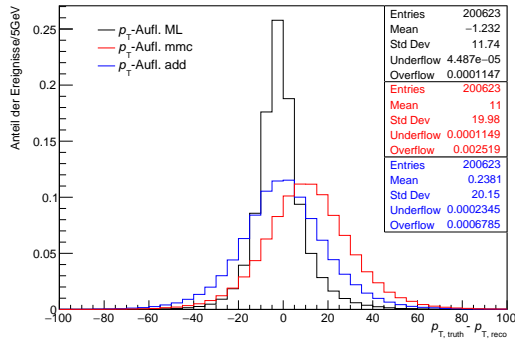
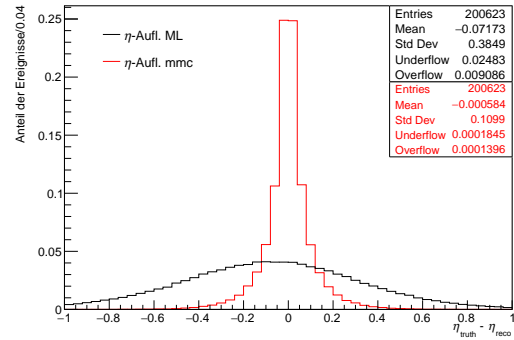
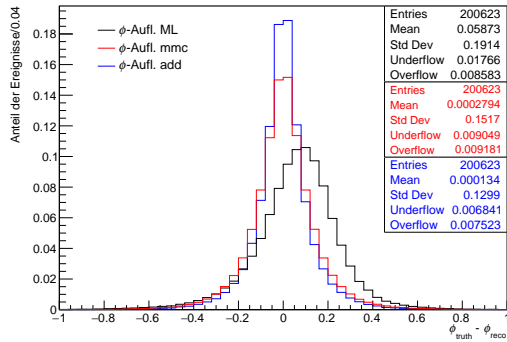
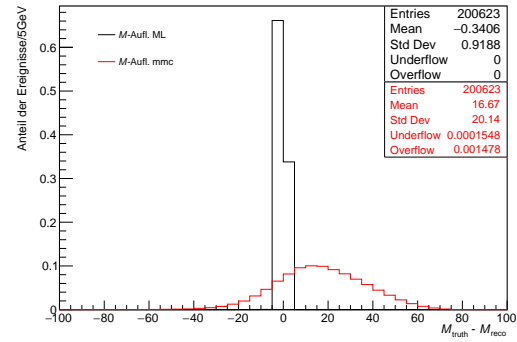
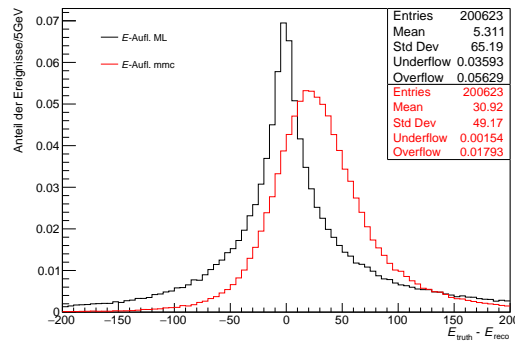
(a) Auflösung bezüglich des Transversalimpulses p_T (b) Auflösung bezüglich der Pseudorapidität η (c) Auflösung bezüglich dem Azimuthwinkel ϕ (d) Auflösung bezüglich der Masse M (e) Auflösung bezüglich der Energie E

Abbildung 36: Auf Eins normierte Auflösungsabbildungen des vierten neuronalen Netzes (ML), des Missing Mass Calculators (MMC) und der Rekonstruktions-Methode des vektoriellen Aufaddierens der Higgs-Zerfallsteilchen (add); Gezeigt ist jeweils die Auflösung bezüglich des Transversalimpulses p_T , der Pseudorapidität η , dem Azimuthwinkel ϕ , der Masse M , sowie der Energie E des rekonstruierten Higgs-Bosons.

Literatur

- [1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579–588, [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.
- [3] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367–377.
- [4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.
- [5] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.
- [6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.
- [7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.
- [8] A. Collaboration, *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** (2012) 1–29.
- [9] C. Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30–61, <http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [10] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, *A high-bias, low-variance introduction to Machine Learning for physicists*, Physics Reports **810** (2019) 1–124.
- [11] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, *Optuna: A Next-generation Hyperparameter Optimization Framework*, in *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019.
- [12] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A new mass reconstruction technique for resonances decaying to*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481–489, <http://dx.doi.org/10.1016/j.nima.2011.07.009>.
- [13] B. Povh et al., *Teilchen und Kerne*. Springer Spektrum, ISBN: 9783642378225, Deutschland, 2014.
- [14] DESY. https://www.desy.de/e409/e116959/e119238/media/6006/2002-02-11_S_M_-Grafik_stufig_eng_thumbnail.jpg, 2002.
- [15] P. Schmüser, *Feynmangraphen und Eichtheorien für Experimentalphysiker*. Springer-Verlag Berlin Heidelberg, ISBN: 9783540584865, Berlin Heidelberg, 1995.
- [16] Particle Data Group Collaboration, P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01.

- [17] O. Philippen, *Quantenfeldtheorie und das Standardmodell der Teilchenphysik*. Springer Spektrum, ISBN: 9783662578193, Deutschland, 2018.
- [18] J. Schwichtenberg, *Durch Symmetrie die moderne Physik verstehen*. Springer Spektrum, ISBN: 9783662538111, Deutschland, 2017.
- [19] A. G. Caren Hagner, *Teilchenphysik für Fortgeschrittene (+ Vorlesung Masse) - Der Higgs-Mechanismus*, <https://www.desy.de/~geiser/Lehre/SS09/higgs09.pdf>, 2009.
- [20] D. J. Griffiths, *Introduction to elementary particles*. Physics textbook, ISBN: 9780471603863. Wiley, New York, NY, 1987.
- [21] J. Bleck-Neuhaus, *Elementare Teilchen*. Springer Spektrum, ISBN: 9783642325786, Deutschland, 2013.
- [22] M. Schumacher, *Suche nach neutralen Higgsbosonen mit dem OPAL-Detektor bei LEP2*. PhD thesis, Universität Bonn, Bonn-IR-99-19, 1999.
- [23] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1–29, [arXiv:1207.7214](https://arxiv.org/abs/1207.7214) [hep-ex].
- [24] C. Collaboration, *A measurement of the Higgs boson mass in the diphoton decay channel*, *Physics Letters B* **805** (2020) 135425.
- [25] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs, CERN-2017-002-M. CERN, Geneva, Oct, 2016. <https://cds.cern.ch/record/2227475>.
- [26] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev. D* **98** (2018) 030001.
- [27] 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** (1957) 1413–1415.
- [28] W. Demtröder, *Experimentalphysik 4*. Springer Spektrum, ISBN: 9783642214769, Berlin, Heidelberg, 2014.
- [29] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, *Phys. Rev. Lett.* **13** (1964) 138–140.
- [30] V. A. MITSOU, *THE ATLAS TRANSITION RADIATION TRACKER*, *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications* (2004), http://dx.doi.org/10.1142/9789812702708_0073.
- [31] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \rightarrow \tau\tau$ channel in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$ with the ATLAS detector*, *Physics Letters B* **805** (2020) 135426, <https://dx.doi.org/10.1016/j.physletb.2020.135426>.
- [32] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel*

- with the ATLAS detector, *The European Physical Journal C* **76** (2016), <http://dx.doi.org/10.1140/epjc/s10052-016-4499-5>.
- [33] A. Lösle, *Measurement of the Higgs-Boson Production Cross-Section and Test of CP Invariance in Vector-Boson Fusion Production with $H \rightarrow \tau^+\tau^-$ Decays in the Fully-Leptonic Final State with the ATLAS Detector at the LHC*. PhD thesis, Freiburg U., 2020.
- [34] M. Diehl and O. Nachtmann, *Optimal observables for measuring three-gauge-boson couplings in $e^+e^- \rightarrow W^+W^-$* , pp. , 301–304. 1996. [arXiv:hep-ph/9603207](https://arxiv.org/abs/hep-ph/9603207) [[hep-ph](#)].
- [35] M. Diehl, O. Nachtmann, and F. Nagel, *Triple gauge couplings in polarised $e^-e^+ \rightarrow W^-W^+$ and their measurement using optimal observables*, *The European Physical Journal C* **27** (2003) 375–397.
- [36] D. Atwood, S. Bar-Shalom, G. Eilam, and A. Soni, *CP violation in top physics*, *Physics Reports* **347** (2001) 1–222.
- [37] D. Atwood and A. Soni, *Analysis for magnetic moment and electric dipole moment form-factors of the top quark via $e^+e^- \rightarrow t\bar{t}$* , *Phys. Rev. D* **45** (1992) 2405–2413.
- [38] J. Boyd, *LHC Run-2 and Future Prospects*, tech. rep., Jan, 2020. [arXiv:2001.04370](https://arxiv.org/abs/2001.04370). proceedings from lecture given at CERN-JINR ESHEP 2019.
- [39] ATLAS Collaboration Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, tech. rep., CERN, Geneva, Jun, 2019. <https://cds.cern.ch/record/2677054>.
- [40] LHC commission, *Longer term LHC schedule*, 2021. <https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm>. [Online; accessed 23-December-2021].
- [41] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [42] M. Schott and M. Dunford, *Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV. Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* **74** (2014) 60 p, [arXiv:1405.1160](https://arxiv.org/abs/1405.1160), <https://cds.cern.ch/record/1699952>.
- [43] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, *Eur. Phys. J. C* **80** (2020) 1194, [arXiv:2007.07624](https://arxiv.org/abs/2007.07624) [[hep-ex](#)].
- [44] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, *Journal of Instrumentation* **12** (2017) P05002–P05002, <http://dx.doi.org/10.1088/1748-0221/12/05/P05002>.
- [45] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, *Journal of Instrumentation* **15** (2020) P10004–P10004, <http://dx.doi.org/10.1088/1748-0221/15/10/P10004>.
- [46] J. Brownlee, *What is the Difference Between a Parameter and a Hyperparameter?*, 2017. <https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/>.

- [47] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego*. 2015. <http://arxiv.org/abs/1412.6980>.
- [48] S. Ruder, *An overview of gradient descent optimization algorithms*, [arXiv:1609.04747](https://arxiv.org/abs/1609.04747) [cs.LG].
- [49] J. Brownlee, *Difference Between a Batch and an Epoch in a Neural Network*, 2018. <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>.
- [50] F. Chollet et al., *Keras*, <https://keras.io>, 2015.
- [51] M. A. et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. <https://www.tensorflow.org/>. Software available from tensorflow.org.
- [52] J. Bergstra, D. Yamins, and D. Cox, *Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures*, in *Proceedings of the 30th International Conference on Machine Learning*, pp. , 115–123. 2013. <https://proceedings.mlr.press/v28/bergstra13.html>.
- [53] Will Koehrsen, *A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning*, 2018. <https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-> [Online; accessed 20-December-2021].
- [54] Wes McKinney, *Data Structures for Statistical Computing in Python*, pp. , 56 – 61. 2010.
- [55] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [arXiv:hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [56] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092](https://arxiv.org/abs/hep-ph/0709.2092) [hep-ph].
- [57] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [arXiv:1002.2581](https://arxiv.org/abs/1002.2581) [hep-ph].
- [58] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, *JHEP* **02** (2010) 037, [arXiv:0911.5299](https://arxiv.org/abs/0911.5299) [hep-ph].
- [59] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Computer Physics Communications* **191** (2015) 159–177, <http://dx.doi.org/10.1016/j.cpc.2015.01.024>.
- [60] A. Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *The European Physical Journal C* (2017), <https://doi.org/10.1140/epjc/s10052-017-5031-2>.
- [61] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-ktjet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063–063.

- [62] A. Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C* **79** (2019).
- [63] A. Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C* **76** (2016).
- [64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825–2830.
- [65] Dr. Michael Böhler, *vbfcprw*, Private communication, 2021.
<https://github.com/frboehler/vbfcprw>. [Online; accessed 21-December-2021].
- [66] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, *Computer Physics Communications* **195** (2015) 161–171.
- [67] C.-P. Yuan, *CTEQ parton distribution functions: CT10*, 2012.
https://ct.hepforge.org/PDFs/ct10_2010.html. [Online; accessed 21-December-2021].

Danksagung

Der Prozess des Erarbeitens meiner Bachelorarbeit war für mich ein sehr positives Erlebnis. Ich habe in einer kurzen Zeit von drei Monaten sehr viel gelernt und hatte große Freude daran, in das Thema meiner Bachelorarbeit einzutauchen und einen Eindruck davon zu bekommen, wie die Forschung in der Teilchenphysik aussieht.

An erster Stelle möchte ich mich bei Herrn Prof. Dr. Markus Schumacher bedanken - nicht nur für die Bereitstellung dieses sehr interessanten Themas, sondern auch für die durchweg gute und intensive Betreuung mit vielzähligen, lehrreichen Diskussionen über die Physik.

Besonderer Dank gilt ebenso meinem Betreuer Dr. Michael Böhler, der mich stets unterstützt und ermutigt hat und mich im Verständnis dieses Forschungsfeldes sehr voran gebracht hat.

Ein großer Dank geht auch an Dr. Valerie Lang, die immer ein offenes Ohr hatte und mich besonders bei der Wahl der Sprache dieser Bachelorarbeit unterstützt hat.

Danke auch an Fabian Becherer, mit dem ich ein Büro geteilt habe und der jederzeit für kleinere und größere Fragen zur Hilfe bereit stand und ein treuer Mensa-Kumpane war. Bezüglich Mensa möchte ich mich auch bei Dr. Dirk Sammel für die lustigen und interessanten Gespräche bedanken.

Der gesamten Arbeitsgruppe möchte ich dafür danken, dass ich von Beginn an als neues Team-Mitglied freundlich aufgenommen wurde.