

# **Über die Anwendung neuronaler Netze zur Messung der Top Quark Masse am LHC**

**(Using neuronal networks  
to measure the top quark mass at LHC)**

**von**

**Rosmarie Irma Wirth**

**geboren am**

**09 Dezember 1993**

**Bachelor-Arbeit im Studiengang Physik**

**Universität Hamburg**

**2018**

**Erstgutachter: Prof. Dr. Peter Schleper**

**Zweitgutachter: Dr. Fred Markus Stober**



## Zusammenfassung

In dieser Bachelorarbeit wird die Masse des top Quarks aus den Daten des CMS-Experiments von 2016 bestimmt. Zur Ereignisselektion wird ein Neuronales Netz genutzt, um die Ereignisse aus dem voll hadronischen Zerfallskanal des top Quarks zu identifizieren. Die Ergebnisse des Neuronalen Netzes werden mit der Selektion aufgrund eines kinematischen Fits verglichen.

Auf dem genutzten Datensatz identifiziert das Neuronale Netz bis zu 1,52 mal mehr richtig permutierte Topereignisse. Der Anteil falsch permutierter Topereignisse im Signal sinkt um bis zu 30,55%.

Aus dem Massenspektrum ergibt sich mit dieser Selektion eine Topmasse von  $171,98 \pm 8,43 \text{ GeV}/c^2$ .

## Abstract

This bachelor thesis measures the top quark mass by using the CMS Data of 2016. A neuronal network takes over the event selection. In this case events from the full hadronic top decay need to be identified. The selection result is compared to a selection based on a kinematic fit.

In this setup the neuronal network identifies up to 1.52 times more right permuted top events. The number of wrong permuted top events after selection decreases by up to 30.55%.

From the mass spectrum, of the selected events, the top mass is measured to be  $171.98 \pm 8.43 \text{ GeV}/c^2$ .

---



---

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
1.1. Zielsetzung und Motivation . . . . .	1
<b>2. Physikalischer Hintergrund</b>	<b>3</b>
2.1. Standardmodell . . . . .	3
2.1.1. Top Quark . . . . .	4
2.2. Das Experiment . . . . .	6
2.2.1. LHC . . . . .	6
2.3. Datenverarbeitung . . . . .	12
2.3.1. Rekonstruktion . . . . .	12
2.3.2. Kinematischer Fit . . . . .	14
2.3.3. Datensatz . . . . .	15
<b>3. Künstliche Neuronale Netze</b>	<b>17</b>
3.1. Neuron . . . . .	17
3.2. Aufbau eines künstlichen Neuronalen Netzes . . . . .	18
3.3. Training . . . . .	19
3.4. Evaluation . . . . .	20
<b>4. Training des Neuronalen Netzes</b>	<b>23</b>
4.1. Struktur . . . . .	23
4.1.1. Zielgrößen . . . . .	23
4.1.2. Simulierte Daten . . . . .	25
4.1.3. Eingabevariablen . . . . .	25
4.1.4. Relevanz der Eingabevariablen . . . . .	27

---

4.2.	Training . . . . .	30
4.2.1.	Wahl des geeignetsten Aktivierungsfunktion . . . . .	30
4.2.2.	Wahl des geeignetsten Optimierungsverfahren . . . . .	32
4.2.3.	Trainingsepochen . . . . .	34
4.2.4.	Ausgabe des Neuronalen Netzes . . . . .	36
4.2.5.	Schwellwert zur Klassifizierung der Prognosen . . . . .	37
4.3.	Analyse der Trainingsergebnisse . . . . .	40
4.3.1.	Signifikanz der Eingabeparameter . . . . .	40
<b>5.</b>	<b>Anwendung des Neuronalen Netzes</b>	<b>43</b>
5.1.	Einfluss der Topmasse des Trainingsdatensatzes auf die Topmasse der selektierten Daten . . . . .	43
5.1.1.	Ergebnisse für simulierten Daten . . . . .	44
5.1.2.	Ergebnisse für Messdaten . . . . .	45
5.2.	Vergleich der Ereignisselektion der Neuronalen Netze mit PGOF-Cut . . . . .	47
<b>6.</b>	<b>Ergebnisse</b>	<b>49</b>
	<b>Literatur</b>	<b>51</b>
	<b>A. Anhang</b>	<b>53</b>
	<b>Eidesstattliche Versicherung</b>	<b>69</b>

---

# 1. Einleitung

## 1.1. Zielsetzung und Motivation

In dieser Bachelorarbeit sollen die Daten des CMS-Detektors aus dem Jahr 2016 genutzt werden, um die Masse des Top Quarks zu bestimmen. Das Top Quark ist das schwerste Elementarteilchen des Standardmodells. Da das Higgs-Boson an die Masse koppelt ist die möglichst exakte Bestimmung der Topmasse sehr interessant. Ein großes Problem bei der Top Quark Massenbestimmung ist, dass Ereignisse, die im CMS-Detektor aufgezeichnet werden, nicht immer richtig zu den verschiedenen Prozessen zugeordnet werden können. In der aktuellen Top Quark Massenmessung werden die Ereignisse mit Hilfe der Anpassungsgüte (engl. 'Godness of Fit') bewertet. Bei dieser Selektionsmethode sind im voll hadronischen Topzerfallskanal circa 47,5% der selektierten Ereignisse tatsächlich richtig zugeordnet[11]. Ein höherer Anteil an Topereignissen im selektierten Datensatz könnte zu einer genaueren Bestimmung der Topmasse beitragen. Um die Klassifikation dieser Ereignisse zu verbessern, wird in dieser Arbeit der Einsatz eines Neuronalen Netzes zur Ereignisselektion untersucht.

Das Neuronale Netz muss zunächst erstellt und trainiert werden. Nachdem der Trainingserfolg gewährleistet ist, können die Daten aus dem CMS-Detektor eingespeist werden. Aus den selektierten Daten kann wiederum die Topmasse bestimmt werden.

---



## 2. Physikalischer Hintergrund

### 2.1. Standardmodell

Die Teilchenphysik untersucht die Grundbausteine der Materie und deren Wechselwirkungen. Elementarteilchen besitzen keine Unterstruktur und bilden in Kombination alle nicht elementaren Teilchen. Außer der Gravitation lassen sich die fundamentalen Wechselwirkungen und die Elementarteilchen im Standardmodell der Teilchenphysik zusammenfassen. Dieses elegante theoretische Konstrukt ermöglicht momentan die beste Beschreibung der experimentellen Ergebnisse.

Drei Generationen  
der Materie (Fermionen)

	I	II	III		
Masse	2,3 MeV	1,275 GeV	173,07 GeV	0	125,09 GeV
Ladung	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
Spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
Name	<b>u</b> up	<b>c</b> charm	<b>t</b> top	<b><math>\gamma</math></b> Photon	<b>H</b> Higgs Boson
Quarks	4,8 MeV $-\frac{1}{3}$ $\frac{1}{2}$ <b>d</b> down	95 MeV $-\frac{1}{3}$ $\frac{1}{2}$ <b>s</b> strange	4,18 GeV $-\frac{1}{3}$ $\frac{1}{2}$ <b>b</b> bottom	0 0 1 <b>g</b> Gluon	
	<2 eV 0 $\frac{1}{2}$ <b><math>\nu_e</math></b> Elektron- Neutrino	<0,19 MeV 0 $\frac{1}{2}$ <b><math>\nu_\mu</math></b> Myon- Neutrino	<18,2 MeV 0 $\frac{1}{2}$ <b><math>\nu_\tau</math></b> Tau- Neutrino	91,2 GeV 0 1 <b><math>Z^0</math></b> Z Boson	
	0,511 MeV -1 $\frac{1}{2}$ <b>e</b> Elektron	105,7 MeV -1 $\frac{1}{2}$ <b><math>\mu</math></b> Myon	1,777 GeV -1 $\frac{1}{2}$ <b><math>\tau</math></b> Tau	80,4 GeV $\pm 1$ 1 <b><math>W^\pm</math></b> W Boson	Eichbosonen
Leptonen					

Abbildung 2.1.: Teilchen und Eigenschaften des Standardmodells der Teilchenphysik.

In Abbildung 2.1 ist das Standardmodell der Teilchenphysik schematisch dargestellt. Dabei werden die Elementarteilchen in Materie- und Austauschteilchen unterteilt. Die Materieteilchen oder Fermionen werden in Quarks und Leptonen unterteilt, in Abbildung 2.1 in den Farben lila und grün dargestellt. Auf der rechten Seite rot und gelb werden die

Bosonen dargestellt, diese werden auch Austauscheteilchen genannt.

Alle fundamentalen Wechselwirkungen, mit Ausnahme der Gravitation, können durch Austausch eines entsprechenden Bosons beschrieben werden. Das Photon  $\gamma$  wird bei elektromagnetischen Prozessen ausgetauscht, während das Gluon  $g$  die starke Wechselwirkung vermittelt. Die schwache Wechselwirkung wird durch  $Z^0$ - oder  $W^\pm$ - Bosonen vermittelt.

Beim Higgs-Teilchen liegt eine vollkommen andere Wechselwirkung vor. Diese beruht auf spontaner Symmetriebrechung und tritt zwischen Higgs und massebehafteten Teilchen auf.

Die Fermionen mit Spin  $1/2$  werden in je drei Generationen von Leptonen und Quarks unterteilt. Die Leptonen unterliegen der schwachen Wechselwirkung und werden wiederum in nicht elektrisch geladene Neutrinos und elektrische geladenen Teilchen unterteilt. Dabei handelt es sich um Elektronen, Myonen und Tauonen, welche die elektrische Ladung  $-1$  tragen und sich durch ihre Massen unterscheiden. Wegen ihrer elektrischen Ladung können diese auch elektromagnetisch wechselwirken.

Im Gegensatz dazu unterliegen Quarks auch der starken Wechselwirkung. Durch ihre drittelzahlige elektrische Ladung wechselwirken sie auch elektromagnetisch. Sie sind ebenfalls in drei Generationen aufgeteilt und heißen up, down, charm, strange, top und bottom. Dabei nimmt das Top Quark durch seine große Masse von  $173,01 \pm 0,6 \frac{\text{GeV}}{c^2}$  [9] eine besondere Rolle, auch in Bezug auf das Higgs-Boson, ein. [8]

### 2.1.1. Top Quark

Das Top Quark ist durch seine hohe Masse sehr instabil. Mit einer Lebensdauer von  $\tau \approx 4 \cdot 10^{-25} \text{s}$  formt es im Gegensatz zu den anderen fünf Quarks keine Hadronen. Die benötigte Zeitdauer zur Bildung von Hadronen beträgt dagegen  $t \approx 10^{-23} \text{s}$ . Nur die Lebensdauern von Up, Down, Strange, Charm und Bottom Quark überschreiten diesen Wert. Wird ein Top Quark erzeugt, so zerfällt es zu schnell um experimentell direkt beobachtet werden zu können. Es können also nur die Zerfallsprodukte des Top Quarks untersucht werden.

In Abbildung 2.2 ist der Zerfallsmechanismus eines Top Quarks in ein positiv gelade-

---

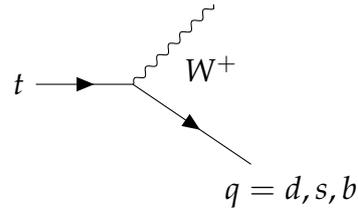


Abbildung 2.2.: Zerfall des Top Quarks in ein W Boson  $W^+$  und ein negativ geladenes Quark (Down, Strange oder Bottom).

nes W Boson und ein negativ geladenes Quark, also Down, Strange oder Bottom, dargestellt. Dabei ist der Zerfall in ein W Boson und ein Bottom Quark im Standardmodell bei weitem der häufigste, mit einer Wahrscheinlichkeit von  $0,99 \pm 0,008$ . Das  $W^+$  Boson zerfällt wiederum entweder leptonisch in ein Lepton und das dazugehörige Neutrino oder hadronisch in Quarks. Die Feynmandiagramme der Zerfälle sind in Abbildung 2.3 dargestellt. Links ist der leptonische Zerfall, wobei  $l$  die verschiedenen Leptonen  $e, \mu, \tau$  und die zugehörigen Neutrinos bezeichnet. Rechts ist der hadronische Zerfall zu sehen.

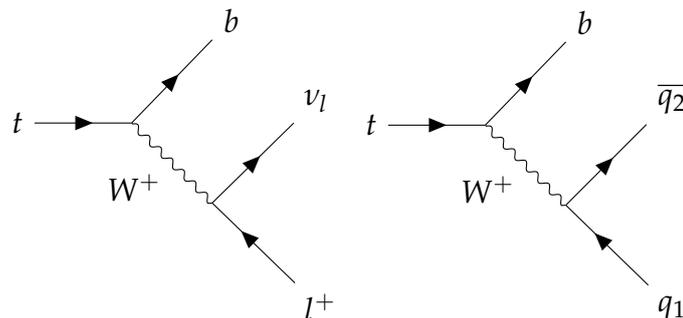


Abbildung 2.3.: Links: leptonischer Zerfall des Top Quarks ( $l = e, \mu, \tau$ ) Rechts: hadronischer Zerfallskanal in Quarks ( $q_1 \bar{q}_2 = u\bar{d}, u\bar{s}, c\bar{d}, c\bar{s}$ )

In dieser Arbeit wird der hadronische Topzerfall betrachtet. Dabei gibt es verschiedene Quarkkombinationen, die als Zerfallsprodukte des  $W^+$ -Bosons in Frage kommen:

$$W^+ \rightarrow q_1 + \bar{q}_2$$

mit

$$q_1 \bar{q}_2 = u\bar{d}, u\bar{s}, c\bar{d}, c\bar{s} \quad (2.1)$$

In dieser Analyse werden Daten des CMS-Experiment am LHC genutzt, dort werden Top

Quarks durch Proton-Proton-Kollisionen erzeugt. Bei diesen Kollisionen kann es insbesondere zur Paarerzeugung von Top Quarks kommen, welche wiederum zerfallen. Dabei wird zwischen dem dileptonischen, semileptonischen und voll hadronischen Zerfallskanal unterschieden. Im dileptonischen und voll hadronischen Kanal zerfallen  $t$  und  $\bar{t}$  ausschließlich leptonisch bzw. hadronisch, während beim semileptonischen Zerfall je ein Top Quark leptonisch und eins hadronisch zerfällt. Dabei haben der semileptonische und der voll hadronische Zerfallskanal mit je 44,4% die höchste Wahrscheinlichkeit. In dieser Arbeit wird der voll hadronische Zerfallskanal untersucht.

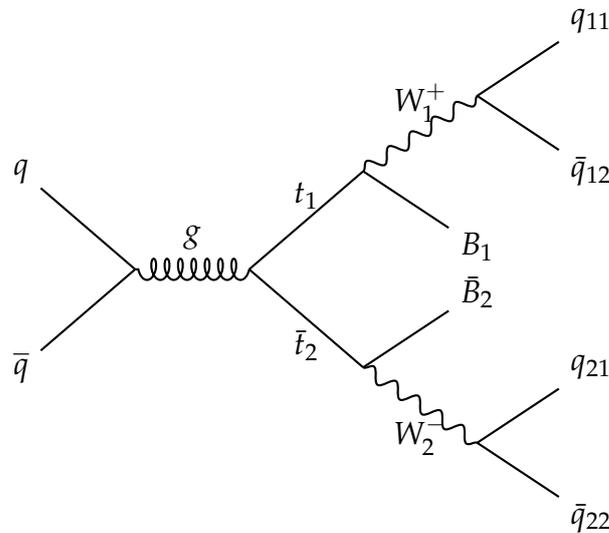


Abbildung 2.4.: Paarerzeugung von Top Quarks mit darauf folgendem Zerfall von  $t\bar{t}$ , die Zerfallsprodukte sind zur weiteren Kennzeichnung benannt.

Um im späteren Verlauf die einzelnen Zerfallsprodukte genau benennen und beschreiben zu können ist dieser Zerfall in Abbildung 2.4 dargestellt.

## 2.2. Das Experiment

### 2.2.1. LHC

Die experimentelle Untersuchung der Elementarteilchen und die Überprüfung der im Standardmodell genutzten Theorien wird zumeist durch hochenergetische Teilchenkollisionen durchgeführt.

Der Large Hadron Collider (kurz: LHC) ist der weltweit größte und leistungsfähigste Teilchenbeschleuniger, welcher am europäischen Kernforschungszentrum CERN in Genf

betrieben wird.

Dabei handelt es sich um ein unterirdisches Synchrotron mit einem Umfang von 26,7 Kilometern. Teilchen werden durch supraleitende Magnete auf einer nahezu kreisförmigen Bahn gehalten und beschleunigt. Im LHC werden zwei gegenläufige Teilchenstrahlen zur Kollision gebracht. Es können Protonen oder Bleiionen zur Kollision gebracht werden.

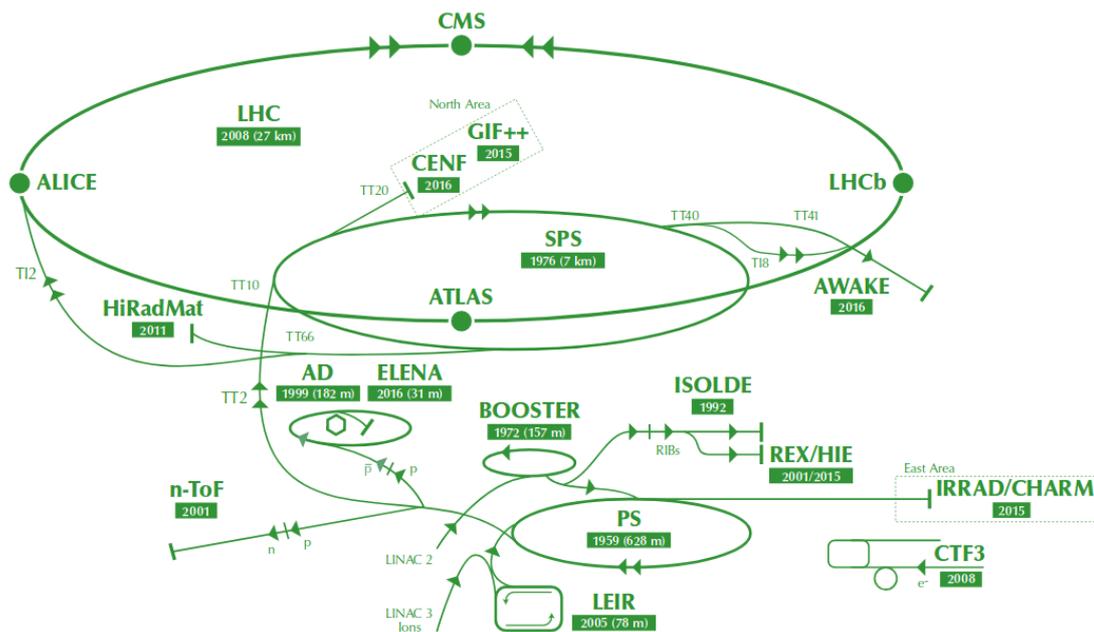


Abbildung 2.5.: Aufbau des LHC Beschleunigerkomplexes mit Vorbeschleunigern und Detektoren.

[7]

In Abbildung 2.5 ist eine Skizze des Aufbaus zu sehen. Im unteren Bereich der Graphik sind die verschiedenen Vorbeschleuniger des LHC dargestellt. Die Protonen werden in einer Protonenquelle durch Ionisation von Wasserstoff erzeugt. Diese werden mit einer kinetischen Energie von 90 keV in den Linearbeschleuniger (LINAC 2) eingeführt. Dieser besteht aus Hohlzylindern, welche durch wechselnde elektrische Ladung die Protonen über 30 Meter auf eine kinetische Energie von 50 MeV beschleunigen. Daraufhin werden die Protonen in den Proton Synchrotron Booster (BOOSTER) geleitet. Dies ist ein 157m langer Kreisbeschleuniger, welcher die kinetische Energie in 1,2 Sekunden auf 1,4 GeV und damit 91% der Lichtgeschwindigkeit hebt. Darauf folgt das Proton Synchrotron (PS) mit einem Umfang von 628 Metern mit einer Beschleunigungsleistung von 28 GeV.

Der letzte Vorbeschleuniger ist das Super Proton Synchrotron (SPS) mit einem Umfang von 6,9 Kilometern, hier werden die Protonen in 21,6 Sekunden auf 450 GeV und somit 99,9998% der Lichtgeschwindigkeit beschleunigt. Von dort werden die Protonen über zwei Verbindungswege mit gegensätzlichen Umlaufrichtungen in den LHC eingespeist. Der LHC besteht aus zwei Speicherringen, welche im Betrieb mit je 2808 Protonenpaketen aus den Vorbeschleunigern gegenläufig befüllt sind. Hier werden die Protonen innerhalb von etwa 20 Minuten auf 6,5 TeV beschleunigt. Dabei wird der 26,7km lange Ring 11000 mal pro Sekunde durchlaufen.[7] Es gibt vier Kreuzungspunkte der beiden Teilchenstrahlen, mit je einem Interaktionsvolumen von  $16\mu\text{m}$  Durchmesser und 8 cm Länge entlang der Wegachse. Pro Paket sind 115 Milliarden Protonen enthalten. In einem Kreuzungspunkt kommt es circa alle 25 ns zu einer Kollision zweier Pakete. In den Kreuzungspunkten des LHCs befinden sich die Detektoren CMS, ATLAS, ALICE und LHCb. Diese sind in Abbildung 2.5 im oberen Bereich dargestellt. Die vier Detektoren sind eigenständige Experimente mit verschiedenen Schwerpunkten. ALICE ist zum Nachweis und der Untersuchung von Quark-Gluonen-Plasmen, ATLAS und CMS dem Nachweis des Higgs-Bosons und anderer neuer Teilchen ausgelegt. Das LHCb Experiment beschäftigt sich mit b Quark Physik und CP-Verletzung.

### **CMS-Detektor**

Der CMS (Compact Muon Solenoid) Detektor ist ein 21 m langer, 18 m hoher und 15 m breiter, zylinderförmiger Aufbau, welcher am nördlichsten Kreuzungspunkt um die Strahlachse gebaut ist. Das Experiment besteht aus verschiedenen Detektoren, welche auf die Eigenschaften der zu detektierenden Teilchen angepasst sind.

In Abbildung 2.6 ist eine Darstellung des Detektors zu sehen, wobei von links oben nach rechts unten die Strahlachse verläuft. Um den blau markierten Interaktionsbereich sind zwiebelartig verschiedene Detektoren angebracht.

Bei den Protonenkollisionen entstehen unterschiedliche Zerfallsprodukte, welche alle im CMS Detektor identifiziert und gemessen werden sollen.

In Abbildung 2.7 ist ein Schnitt durch den CMS Detektor zu sehen. Die einzelnen Detektoren haben verschiedene Messeigenschaften und können so unterschiedliche Teilchen

---

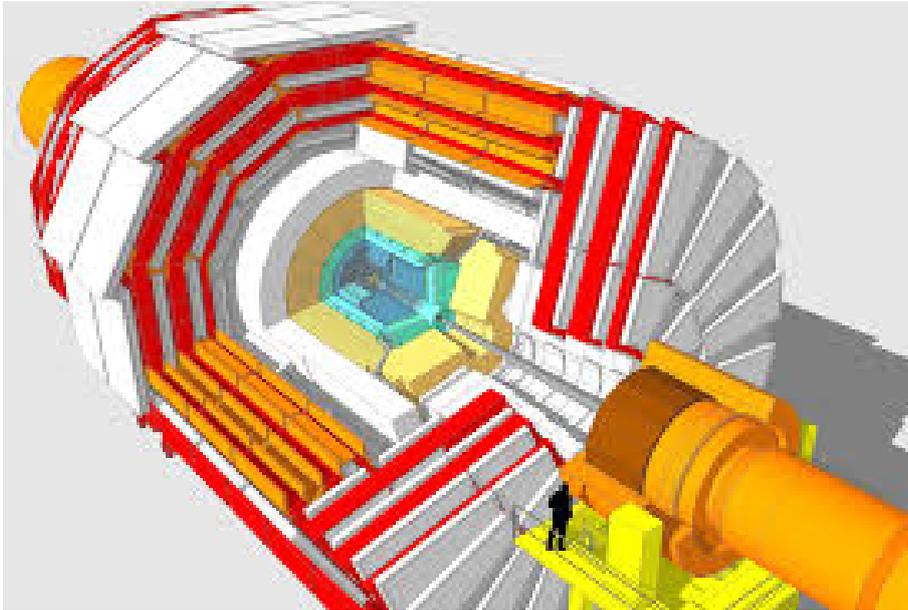


Abbildung 2.6.: Darstellung des CMS Detektors: Die Strahlachse des LHCs verläuft von links oben nach rechts unten. Um den in blau gekennzeichneten Interaktionsbereich sind zwiebelförmig die verschiedenen Detektoren angebracht.[7]

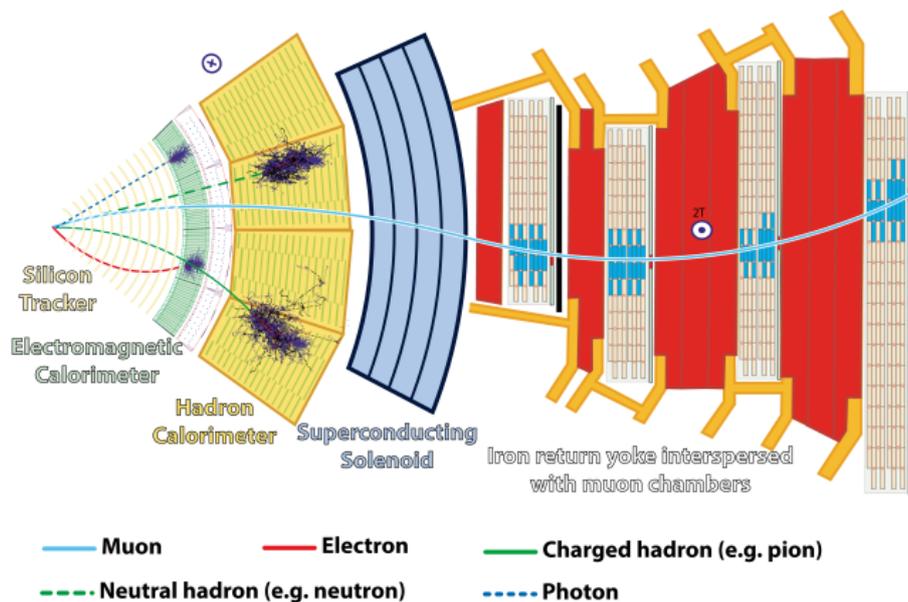


Abbildung 2.7.: Schnitt durch den CMS Detektor mit verschiedenen Teilchenbahnen und Detektorarten. [1]

detektieren. Des Weiteren sind auch verschiedene Teilchenbahnen eingezeichnet. Dabei gilt es geladene und ungeladene Hadronen und Leptonen mit verschiedensten Massen aber auch Photonen zu identifizieren.

Im Folgenden werden die Detektoren von Innen nach Außen beschrieben:

### **Spurdetektoren**

Die Spurdetektoren im CMS funktionieren nach dem Prinzip von Halbleiterdetektoren. Die Siliziumpixel sind ähnlich wie Dioden aus einem p- und einem n-dotierten Bereich zusammengesetzt. Die Zonen mit Elektronenüberfluss und -mangel sind durch eine sogenannte Sperrzone getrennt. In Sperrrichtung ist eine Gleichspannung angelegt. Trifft ein geladenes Teilchen auf einen Pixeldetektor, wird entlang der Flugbahn das Silizium ionisiert. Durch die anliegende Spannung können die Ladungsträger als Signal gemessen werden. Da die zurückzulegenden Driftstrecken sehr klein sind, beträgt die Verzögerung im Signal weniger als 10 ns. Siliziumdetektoren sind besonders geeignet um die Flugbahnen vieler geladener Teilchen aufzuzeichnen. Durch ihre schnelle Reaktionsfähigkeit und genaue Ortsauflösung können so mit Hilfe der Krümmung der Flugbahn der Impuls und die Geschwindigkeit der Teilchen bestimmt werden.

Die innerste Schicht im CMS bildet ein Siliziumpixeldetektor. Die geringe Pixelgröße des Detektors von  $100 \times 150 \mu\text{m}$  ermöglicht Ortsauflösung von bis zu  $5 \mu\text{m}$ . Da die Pixel in mehreren Schichten um den Kollisionsbereich liegen, kann ein dreidimensionales Bild der Trajektorien aller elektrisch geladenen Teilchen erstellt werden.

Daraufhin müssen entstandene Teilchen den Silizium-Streifendetektor passieren. Dieser ist in 10 Schichten bis zu einem Radius von 1,30m angeordnet. So können auch geringe Krümmungen der Teilchenbahnen erkannt werden.

### **Kalorimeter**

Die Kalorimeter geben Auskunft über die Energie der eintreffenden Teilchen. Dabei muss zwischen dem elektromagnetischen (ECAL) und dem hadronischen Kalorimeter (HCAL) unterschieden werden. In beiden Fällen soll die Energie der Teilchen im Kalorimeter vollständig deponiert und rekonstruiert werden. Hierfür sind Materialien hoher Dichte und Kernladungszahl sinnvoll.

---

---

Im **elektromagnetischen Kalorimeter ECAL** werden die Energien hochenergetischer Photonen und Elektronen gemessen. Dabei wird die Energie hauptsächlich durch Photoeffekt, Compton-Effekt und Paarbildung auf das Detektormaterial übertragen. Da bei großen Energien Paarbildung und Bremsstrahlung die dominierenden Prozesse sind, kommt es zu elektromagnetischen Schauern, welche sich lawinenartig vermehren und erst bei geringen Energien ihre Energie mittels Ionisation im Detektor deponieren. Im CMS werden  $PbWO_4$ -Kristalle als aktives Material genutzt. Diese geben die absorbierte Energie der Teilchen in Form von Licht ab. Speziell für diese Anforderungen entwickelte Photodetektoren befinden sich am Ende jedes Kristalls und verwerten die Lichtsignale. Im CMS sind fast 80 000 derartige Kristalle verbaut, welche es erlauben, die Energie von Photonen und Elektronen genau zu detektieren.

Im **hadronischen Kalorimeter HCAL** wird nun die Energie von Hadronen, also beispielsweise Protonen, Pionen und Neutronen, gemessen. Hadronen interagieren mittels starker, schwacher und elektromagnetischer Wechselwirkung. Dabei kommt es zu elastischen und inelastischen Reaktionen mit dem Detektormaterial.

Im CMS Detektor besteht das HCAL aus abwechselnden Schichten von absorbierenden Materialien und Szintillatoren. Die Szintillatoren funktionieren analog zu den Kristallen im ECAL.

### **Der Magnet**

Um die Kalorimeter liegt im CMS die supraleitende Spule, welche das Magnetfeld von 3,8 Tesla generiert. Da bewegte Ladungen im Magnetfeld eine zur Magnetfeldrichtung senkrechte Kraft erfahren, können so anhand der Flugbahnen Ladung und Impuls berechnet werden.

### **Myonspektrometer**

Myonen sind 200 mal schwerer als Elektronen und können daher mehrere Meter Eisen durchqueren ohne viel Energie zu verlieren. Deshalb werden sie in den Kalorimetern nicht absorbiert. Um auch diese Teilchen zu detektieren ist das Myonspektrometer um den Magneten herum angebracht. Es besteht aus vier Lagen Myonenkammern. Insgesamt

---

sind 1400 Myonenkammern verbaut. Aus den Ortsinformationen der Siliziumdetektoren und der Myonenkammern kann die Flugbahn der Myonen sehr präzise rekonstruiert werden. Aus dieser kann wiederum mit Hilfe der bekannten Magnetfeldkonfigurationen der Impuls bestimmt werden.

In den verschiedenen Detektoren wurden die Flugbahnen, Energien und Impulse der einzelnen Zerfallsprodukte aufgenommen. Diese müssen allerdings noch weiter verarbeitet werden bevor sie zur Analyse genutzt werden können.

## **2.3. Datenverarbeitung**

### **2.3.1. Rekonstruktion**

Bei der Rekonstruktion der Ereignisse im Detektor werden zwei wichtige Verfahren angewendet. Zum einen der vom CMS entwickelte Particle-Flow-Algorithmus und zum anderen die Jet Clustering-Algorithmus.

#### **Particle-Flow-Algorithmus**

Der Particle-Flow-Algorithmus verwendet alle Informationen der einzelnen Detektorteile eines Ereignisses, um die Teilchen genau zu identifizieren und zu vermessen.

Zuerst werden die Flugbahnen geladener Teilchen mit einem Transversalimpuls von mindestens 150 MeV rekonstruiert. Die zugehörigen Messungen werden aus dem Datensatz entfernt, um weitere Teilchenbahnen zu identifizieren. Die Energien aus den Kalorimetern werden separat zusammengefasst. Im folgenden müssen die Teilchenbahnen mit den Energiedepositionen aus den Kalorimetern kombiniert werden, sonst besteht die Gefahr von mehrfach Zählungen detektierter Teilchen. Dabei werden die Enden der Spuren der Teilchenbahnen eine Interaktionslänge in die Kalorimeter hinein verlängert und die detektierten Energien mit diesen assoziiert. Des Weiteren werden auch die Spuren im Myonenspektrometer mit Hilfe eines  $\chi^2$ -fits zu Messungen aus den Siliziumdetektoren zugeordnet. Zur Identifikation der Teilchen werden zuerst die Spuren aus dem vorherigen Schritt mit den dazugehörigen Energien als Myonen markiert. Daraufhin werden die Zuordnungen von Spuren und Energien im Kalorimeter verbessert. Dabei werden, falls mehrere

---

Energiedeposition für eine Spur in Frage kommen, nur die mit der größten Energie berücksichtigt und den Messungen aus dem HCAL eventuelle elektromagnetische Schauer zugeordnet. Wenn Spuren und Energien innerhalb der Messungenauigkeiten zusammenpassen, wird ein globaler Fit über beide durchgeführt. Alle nicht zu einer Teilchenbahn assoziierten Energien werden als ungeladene Hadronen oder Photonen rekonstruiert. Aus der Summe aller Vierer-Vektoren aller Teilchen kann nun die fehlende Transversalenergie (MET - missing transverse energy) berechnet werden, dabei wird die Z-Komponente aller Teilchen auf Null gesetzt.

Da der Particle-Flow-Algorithmus die Kombination aus Teilchenbahnen und Energiedepositionen nutzt, ist es möglich, diese auf jeweils eine der Interaktion zurück zu führen. So können Teilchen aus verschiedenen Kollisionen unterschieden werden. Damit kann der sogenannte Pileup Effekt, welcher die Verfälschung von Messungen durch Teilchen aus anderen Interaktionen bezeichnet, verringert werden.

In Abbildung 2.7 sind die Teilchenbahnen eines Myon, Elektrons, geladenen und ungeladenen Hadrons und eines Photons mit den zugehörigen detektierten Energien beispielhaft eingezeichnet. [2]

### **Jet - Rekonstruktion**

Die rekonstruierten Teilchen werden im darauf folgenden Schritt zu Jets zusammengefasst. Als Jet wird ein kollimiertes Bündel Teilchen bezeichnet. Hierfür wird der sogenannte anti-k Algorithmus verwendet. Aus den rekonstruierten Jets werden die mit  $p_T > 30$  GeV und  $|\eta| < 2.4$  selektiert. In dieser Analyse werden zwei B-Jets und vier weitere Quark-Jets für den voll hadronischen Top-Zerfall benötigt (siehe Abbildung 2.4).[10]

### **b-tagging**

Das Bottom Quark ist ein wichtiger Bestandteil des voll hadronischen Topzerfalls. Deshalb ist es notwendig Jets, welche aus einem Bottom Quark resultieren zu identifizieren. Wenn ein Meson ein Bottom Quark enthält, hat es eine recht lange Lebensdauer und kann einige Hundert Micrometer zurücklegen. Durch den verspäteten Zerfall des Bottom Quarks können diese B Quark Mesonen erkannt werden. [6]

---

Hier wird der 'Combined Secondary Vertex algorithm' [6] genutzt. Dieser nutzt unter anderem die Position und den Impuls und die Masse des zweiten Zerfallsvertex. Es sind verschiedene Arbeitspunkte definiert. In diesem Fall wird der strengste ('tight working point') genutzt. Dieser wird auch CSVv2T genannt. An diesem Arbeitspunkt hat der Algorithmus eine Effizienz von  $\approx 50\%$  und eine Falschzuordnungsrate von  $\approx 0,1\%$ , bei einem Discriminatorwert von  $>0.9535$ .

### 2.3.2. Kinematischer Fit

Beim kinematischen Fit werden die Jets eines Ereignisses mit Hilfe der Methode der kleinsten Quadrate an eine mögliche Zerfallshypothese angepasst. Dabei gibt die Anpassungsgüte an, wie gut das gemessene Ereignis der angenommenen Hypothese entspricht. In diesem Fall sollen voll hadronische Topereignisse untersucht werden. Dieses Ereignisse bestehen aus sechs Jets. Zwei davon müssen als b-Jet identifiziert sein. Daraus resultieren folgende Bedingungen für einen voll hadronischen Topzerfall:

- die Kombination aus je zwei Jets, die nicht als B markiert sind, soll die W-Bosonenmasse von 80.4 GeV ergeben[9],
- die Massen des rekonstruierten Top Quarks und anti-Top Quarks sollen gleich sein.

Während des Fits werden die Fitparameter der Jets angepasst, um das minimale  $\chi^2$  zu finden. Dabei werden alle möglichen Jetpermutationen genutzt. Für die gegebenen Bedingungen werden also zwölf Permutationen pro Ereignis getestet.

$$\chi^2 = \sum_f \sum_j \frac{(f_{j, \text{rekonstruiert}} - f_{j, \text{angepasst}})^2}{\sigma_{f_j}^2} + \sum_b d_b \quad (2.2)$$

Dabei werden durch f die Fitparameter  $E_T, \eta, \phi$  bezeichnet. Der Index j repräsentiert die Jets und b die oben genannten Bedingungen. Die Massen der Jets werden auf Null gesetzt. Die Anpassung an die genannten Bedingungen erfolgt über d:

$$d_{m_{W_i}} = m_{W_i, \text{fit}} - m_W \quad (2.3)$$

$$d_{m_t} = m_{t1, \text{fit}} - m_{t2, \text{fit}} \quad (2.4)$$

Konvergiert die Minimierung  $\chi^2$  für eine Permutation handelt es sich dabei um eine Kombination, welche die gesetzten Bedingungen erfüllt. Die Wahrscheinlichkeit des Fits  $P(\chi^2)$  wird wie folgt bestimmt:

$$P(\chi^2) = \left(2^{n/2}\Gamma\left(\frac{n}{2}\right)\right)^{-1} \int_{\chi^2}^{\infty} t^{n/2-1} e^{-t/2} dt \quad (2.5)$$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (2.6)$$

Für richtige Permutationen ist der  $\chi^2$ -Wert klein. Die  $P(\chi^2)$ -Verteilung ist so konstruiert, dass sie für korrekte Hypothesen und korrekte Annahme der Messfehler konstant ist und Werte im Bereich  $0 \leq P(\chi^2) \leq 1$  annimmt. In aktuellen Analysen wird die  $P(\chi^2)$  als Selektionsgröße genutzt. Hat ein Ereignis einen Wert  $P(\chi^2) > 0,1$  so wird dieses im weiteren als richtig permutiertes Topereignis verwendet.

### 2.3.3. Datensatz

In dieser Arbeit werden die folgenden Daten vom CMS-Detektor aus dem Jahr 2016 genutzt:

```
/JetHT/Run2016B-03Feb2017ver2-v2/MINIAOD
/JetHT/Run2016C-03Feb2017-v1/MINIAOD_allCut
/JetHT/Run2016D-03Feb2017-v1/MINIAOD_allCut
/JetHT/Run2016E-03Feb2017-v1/MINIAOD_allCut
/JetHT/Run2016F-03Feb2017-v1/MINIAOD_allCut
/JetHT/Run2016G-03Feb2017-v1/MINIAOD_allCut
/JetHT/Run2016H-03Feb2017ver2-v1/MINIAOD_allCut
/JetHT/Run2016H-03Feb2017ver3-v1/MINIAO_allCut
```

Diese Daten enthalten bereits einigen Vorselektionen.

Der sechste Jet, absteigend sortiert nach der Größe der Transversalimpulse  $p_T$ , muss  $p_T(jet_6) > 40\text{GeV}$  erfüllen. Die Summe der Transversalimpulse  $H_T := \sum_{jets} p_T$  muss  $H_T > 459\text{ GeV}$  erreichen. Dabei ist . Außerdem müssen die beiden B-Jets einen Winkelabstand von  $\Delta R_{B_1, B_2} = \sqrt{(\Delta\phi^2 + \Delta\eta^2)} > 2.0$  haben.



## 3. Künstliche Neuronale Netze

Künstliche Neuronale Netze werden in der Datenverarbeitung genutzt, um Datensätze zu abstrahieren und Muster in diesen zu erkennen. Dabei orientiert man sich am biologischem Vorbild, dem Gehirn. Künstliche Neuronale Netze bestehen ebenfalls aus Neuronen, welche durch einen Lern-Algorithmus in der Lage sind, auch komplexe Zusammenhänge aufzudecken.

Dabei wird ein Neuronales Netz zuerst mit Hilfe eines Trainingsdatensatzes auf das zu untersuchende Merkmal trainiert. Werden in ein trainiertes Netz dann Daten eingegeben wird eine Ausgabe generiert, die einer Beschreibung des Merkmals entspricht. Neuronale Netze eignen sich unter anderem zur Klassifizierung von Daten, dabei kann als Ausgabe beispielsweise die Zugehörigkeit zu verschiedenen Klassen abgefragt werden.

### 3.1. Neuron

Ein künstliches Neuronale Netz besteht aus vielen Neuronen. Ein solches Neuron ist in Abbildung 3.1 schematisch dargestellt. Die Neuronen bekommen als Eingangsinformation die Eingabeparameter oder die Ausgaben ( $x_n$ ) der Neuronen aus der vorherigen

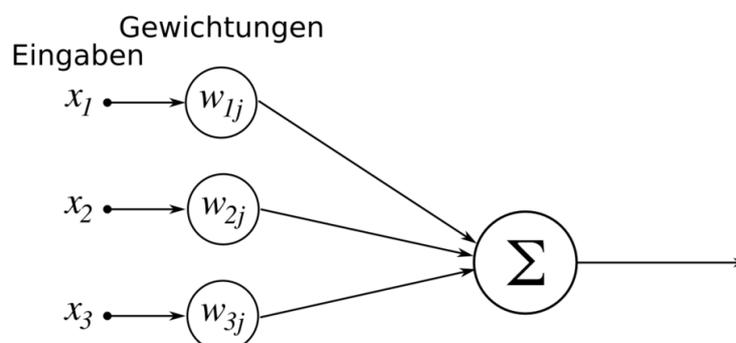


Abbildung 3.1.: Model eines Neurons.[4]

Schicht. Diese werden jeweils mit einem Faktor ( $w_n$ ) gewichtet. Durch die Anpassung der Gewichtungsfaktoren, kann das Neuronale Netz lernen. Dabei werden die Faktoren so angepasst, dass sie einem vorgegebenen Ziel möglichst nahe kommen. Das Neuron bildet aus den gewichteten Eingaben die Summe:

$$a = \sum_{i=0}^N w_i x_i + b \quad (3.1)$$

Außerdem wird dabei der sogenannte Bias  $b$  addiert. Dies ist eine weitere Stellschraube, die während des Trainings angepasst werden kann. Die entstandene Summe ist das Argument einer Aktivierungsfunktion. Das Ergebnis der Aktivierungsfunktion ist der Ausgabewert des Neurons, welches wiederum gewichtet in Neuronen der nächsten Schicht eingespeist wird.

Es gibt verschiedene Aktivierungsfunktionen, die an das Problem angepasst werden können. Da das Neuron die Summe über alle Eingangsvariablen bildet, werden meist Funktionen genutzt, welche ausschließlich auf Null und positive Werte abbilden.

## 3.2. Aufbau eines künstlichen Neuronalen Netzes

Künstliche Neuronale Netze sind in Schichten aufgebaut. Dabei ist zu unterscheiden zwischen Eingabe- und Ausgabeschicht und den inneren Schichten. Die Eingabeschicht besteht aus einem Neuron pro Eingabeparameter, während die Ausgabeschicht ebenfalls ein Neuron pro Ausgabeparameter aufweist. In Abbildung 3.2 ist ein einfacher Aufbau eines künstlichen Neuronalen Netzes mit 3 Eingabeneuronen, einer einzelnen inneren Schicht und zwei Ausgabeneuronen schematisch dargestellt. Dabei ist charakteristisch, dass die Ausgaben der einzelnen Neuronen einer Schicht als Eingabe für andere Neuronen der nächsten Schicht genutzt werden. Bei den 'fully-connected' (komplett verbundenen) Neuronalen Netzen sind alle Neuronen einer Schicht mit allen Neuronen der nächsten Schicht verbunden, wie in der Abbildung dargestellt. Im Gegensatz zu allen anderen Neuronen haben die Ausgabeneuronen keine Aktivierungsfunktion, sondern geben in der Regel lediglich die Summe aus. Somit hat das Neuronale Netz in Abbildung 3.2 trotz seiner vergleichsweise geringen Größe mit 15 Gewichtungsfaktoren ( $w_i$ ) und 5 Biaswerten ( $b$ ) schon 20 Parameter, welche während des Trainings dem Problem angepasst werden

---

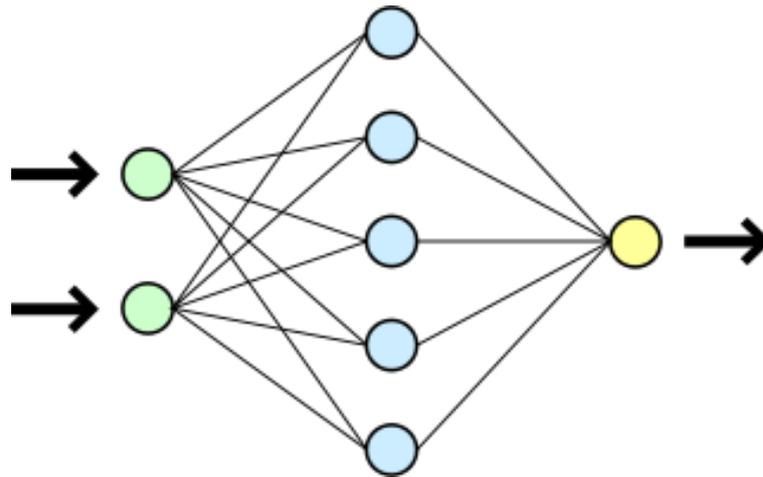


Abbildung 3.2.: Neuronales Netz mit einer Hidden Layer. [5]

können.

Je nach Komplexität des Problems ist es sinnvoll, die Anzahl an Schichten und Neuronen anzupassen. Mit wachsender Anzahl an Neuronen und Schichten nimmt auch die Komplexität des vom Neuronalen Netzes erstellten Modells zu. Dies birgt aber auch die Gefahr, dass das Neuronale Netz übertrainiert wird. In diesem Fall fehlinterpretiert das Neuronale Netz, wegen zu hoher Lernkapazität oder zu kleinem Trainingsdatensatz, unwichtige Merkmale des Trainingsdatensatzes als ausschlaggebende Eigenschaften und klassifiziert somit auf dem Trainingsdatensatz sehr gut, auf den auszuwertenden Daten allerdings nicht.

### 3.3. Training

Um ein Neuronales Netz zu trainieren wird ein Trainingsdatensatz benötigt. Dieser zeichnet sich dadurch aus, dass die gewünschten Zielwerte bekannt sind. Zielwert und Ausgabe sind die Werte der Ausgabeschicht und sind in der Regel als Vektor gegeben. Beim Training wird der Trainingsdatensatz in das Neuronale Netz eingespeist und die Ausgabe bestimmt. Die Lossfunktion bestimmt dann, wie nah die Ausgaben an den gewünschten

Werten liegen. Als Lossfunktion eignet sich beispielsweise:

$$L = \frac{1}{N} \sum_{i=1}^N \left( \begin{pmatrix} Ziel_{1,i} \\ \vdots \\ Ziel_{n,i} \end{pmatrix} - \begin{pmatrix} Ausgabe_{1,i} \\ \vdots \\ Ausgabe_{n,i} \end{pmatrix} \right)^2 \quad (3.2)$$

Dabei wird über N untersuchte Ereignisse summiert. Die Ausgabe- und Zielvektoren enthalten die Werte der n Ausgabeneuronen je Ereignis. Je nach Problemstellung kann die Lossfunktion aber auch beliebig komplex sein. Im Verlauf des Trainings wird versucht, durch Variation der Gewichte das Ergebnis der Lossfunktion zu minimieren. Um ein Minimum in der Lossfunktion zu finden gibt es verschiedene Optimierungsverfahren. Diese geben an wie die Gewichtungsfaktoren und Biaswerte angepasst werden.

### 3.4. Evaluation

Wurde nun ein künstliches Neuronales Netz erstellt und trainiert, so muss seine Güte beschrieben werden.

Schon das Training gibt Auskunft über die Vorhersagen des Neuronalen Netzes. Dafür kann ein Graph mit dem Ergebnis der Lossfunktion für die einzelnen Trainingsepochen genutzt werden, siehe Abbildung 3.3. Anhand dieser Abbildung kann der Fortschritt des Trainings abgelesen werden. Außerdem kann mit Hilfe des Vergleichsdatensatz vermieden werden, dass das Neuronale Netz wie in Kapitel 3.2 beschrieben übertrainiert wird. Das Minimum der Lossfunktion auf dem Trainings- und dem Vergleichsdatensatz beschreibt somit den Punkt des besten Trainingszustands.

Eine weitere Evaluationsmethode, welche sich besonders zum Vergleich von verschiedenen Neuronalen Netzen eignet ist die Roc-Kurve. Bei dieser Methode wird ausgewertet wie viel Prozent des Signals und wie viel Prozent des Hintergrunds richtig eingeordnet wurden. Dafür müssen zunächst die Klassen Hintergrund und Signal je nach Problem definiert werden. Wenn die Prognose des Neuronalen Netzes einen Schwellwert überschreitet, wird das Ereignis als prognostiziertes Signal definiert. Zur Erstellung der Roc-Kurve werden die Anteile für verschiedene Schwellwerte  $S \in [Prognose_{min}, Prognose_{max}]$  be-

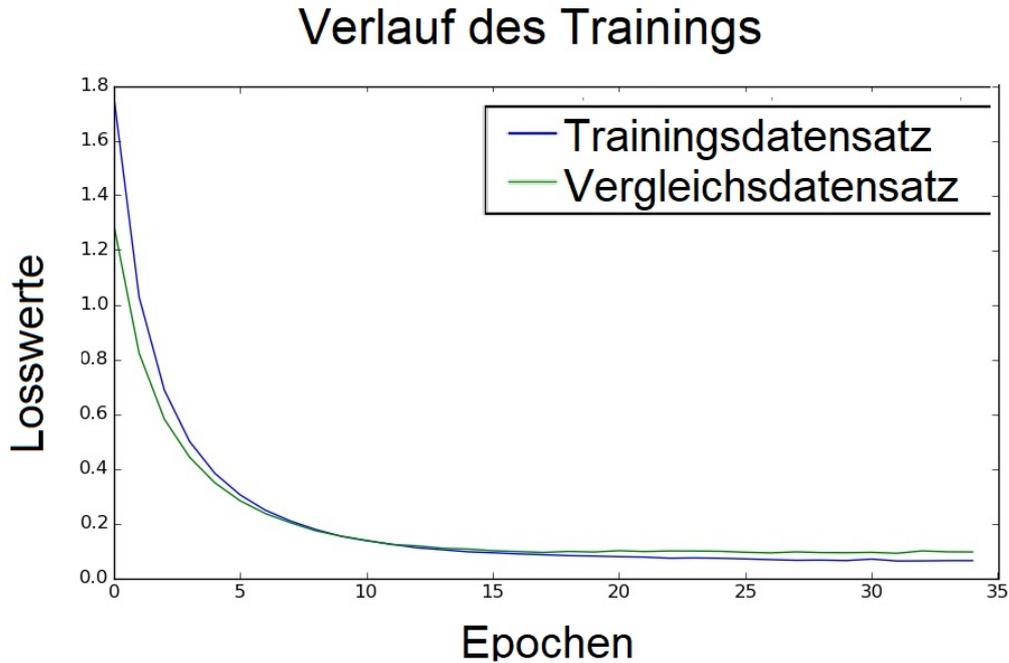


Abbildung 3.3.: Beispiel: Ergebnis der Lossfunktion nach den einzelnen Trainingsepochen, dabei sind die Ergebnisse für den Trainingsdatensatz und den Vergleichsdatensatz zu sehen. Ab der elften Epoche erreicht der Trainingsdatensatz bessere Ergebnisse als der Vergleichsdatensatz. Dies wird als Übertraining bezeichnet.

rechnet:

$$R_S = \frac{N_{Prognose>S}}{N_{wahr}} \quad (3.3)$$

Daraufhin wird wie in Abbildung 3.4 der Anteil des richtig prognostizierten Hintergrunds nach richtig prognostiziertem Signal aufgetragen. Je größer die eingeschlossene Fläche, desto klarer sind die beiden Klassen voneinander getrennt. In der Abbildung ist also das Neuronale Netz 2 besser für die Trennung zwischen Signal und Hintergrund geeignet.

Um die Güte der Roc-Kurve genauer zu klassifizieren wird der Gini-Koeffizient genutzt. Dieser beschreibt die Fläche unter der Kurve gegenüber einem 1x1-Quadrat,

$$g = 1 - \int_0^1 f(x) dx \quad (3.4)$$

wobei  $f(x)$  die Roc-Kurve beschreibt. So können auch verschiedene Klassifizierungsmethoden verglichen werden. Mit sinkendem Gini-Koeffizient steigt die Trennung zwischen Signal und Hintergrund.

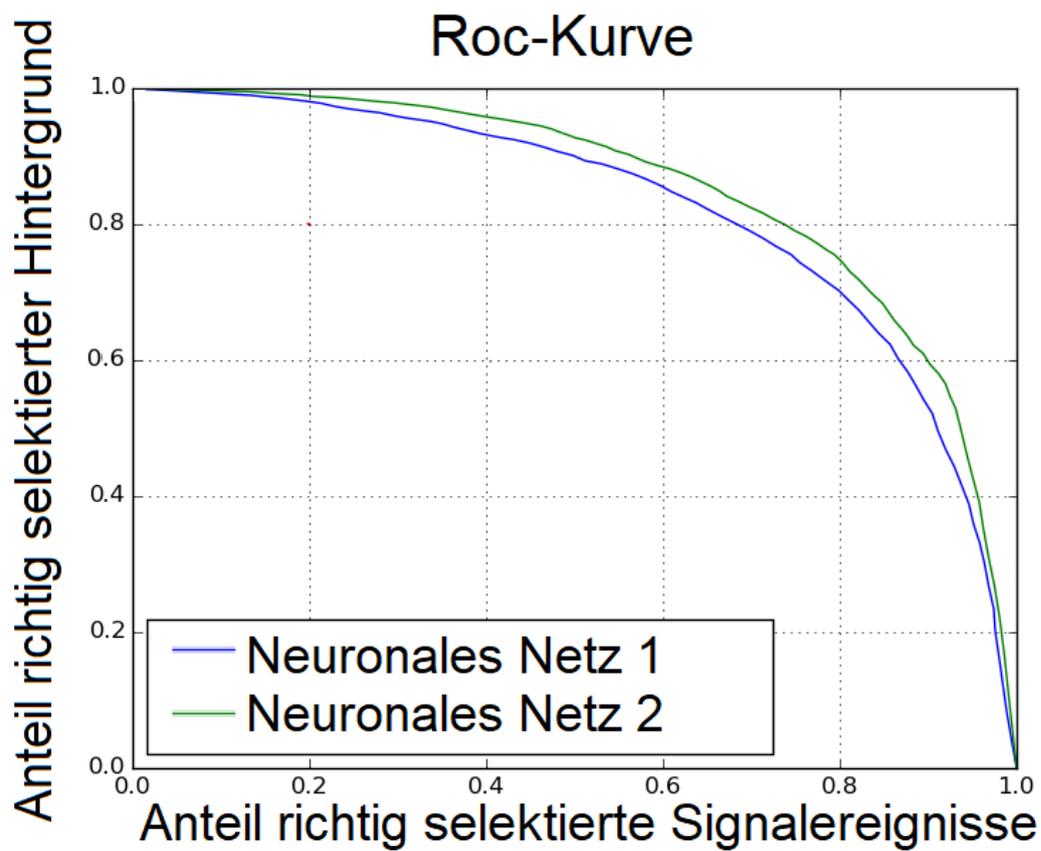


Abbildung 3.4.: Beispiel: Roc-Kurven verschiedener Neuronaler Netze. Der Anteil richtig vorhergesagter Hintergrundereignisse ist auf der Y- und die Anteil richtig vorhergesagter Signalereignisse auf der X-Achse aufgetragen.

## 4. Training des Neuronalen Netzes

Aus den Daten des CMS-Detektors soll am Ende dieser Arbeit die Masse des Top Quarks aus dem Massenspektrum bestimmt werden. Da im gegebenen Datensatz verschiedene Ereignisse enthalten sind, muss eine Methode zur Selektion geeigneter Ereignisse gefunden werden.

Um die Methoden bewerten zu können werden folgende Ziele der Selektion festgelegt:

- effiziente Identifizierung der Signalereignisse,
- Klassifizierung der Ereignistypen zur weiteren Analyse,
- möglichst genaue Bestimmung der Topmasse aus identifizierten Ereignissen.

### 4.1. Struktur

#### 4.1.1. Zielgrößen

Der CMS - Datensatz besteht einerseits aus Ereignissen hadronischer  $t\bar{t}$ -Zerfällen und andererseits aus anderen Ereignissen, welche die Rahmenbedingungen der Ereignisauswahl erfüllen. Die zweite Ereignisgruppe wird im folgenden Hintergrund Ereignisse genannt und besteht größtenteils aus QCD Ereignissen.

Des weiteren wird auf den CMS-Datensatz der kinematische Fit angewendet (siehe Abschnitt 2.3.2). Somit sind in dem Datensatz zusätzlich zu Hintergrundereignissen, auch Topereignisse mit korrekter und falscher Permutation, also Zuordnungen der Jets zu finden. Die Ereignisse mit falscher Permutation sollen nicht für die Berechnung der Topmasse genutzt werden. Außerdem soll eine Selektion zwischen richtig permutierten Topereignissen hoher und niedriger Massenabweichung  $\frac{\Delta m_t}{m_t}$  stattfinden. Somit soll in der Ausgabe die Unterteilung der gegebenen Ereignisse in diese Klassen erfolgen:

---

- korrekt permutierte Topereignisse geringer Massenabweichung  $\frac{\Delta m_t}{m_t}$ ,
- korrekt permutierte Topereignisse hoher Massenabweichung  $\frac{\Delta m_t}{m_t}$ ,
- falsche Permutationen von Topereignissen,
- andere Ereignisse, z.B. kein  $t\bar{t}$ -Ereignis, anderer Zerfallskanal, QCD.

Die Zielausgabe je Ereignis hat also vier Komponenten, welche die Zugehörigkeit zu den verschiedenen Klassen beschreiben.

Die relative Topmassenabweichung wird mit Hilfe der Fehlerfortpflanzung berechnet:

$$\frac{\Delta m_t}{m_t} = \frac{f}{m_t^2} \cdot \sqrt{(p_{T_{B1}} \cdot D_{B1})^2 + (p_{T_{q11}} \cdot D_{q11})^2 + (p_{T_{q12}} \cdot D_{q12})^2} \quad (4.1)$$

$$D_i = m_t \left( \frac{\partial m_t}{\partial p_{T_i}} \right) \quad (4.2)$$

Dabei ist  $m_t$  die Topmasse und  $f = 0.02$  der angenommene Fehler auf den transversalen Impuls[11]. Des weiteren ist  $\frac{\Delta m_t}{m_t}$  auf 30 beschränkt und normiert:

$$M_{normiert} = \frac{\min(\frac{\Delta m_t}{m_t}, 30)}{30} \quad (4.3)$$

Die Unterteilung der richtig permutierten Topereignisse gestaltet sich wie folgt:

$$K_{geringe\ Topmassenabweichung} = 1 - M_{normiert} \quad (4.4)$$

$$K_{hohe\ Topmassenabweichung} = M_{normiert} \quad (4.5)$$

Somit haben die Ereignisse geringer Massenabweichung für 4.4 ein Ergebnis nahe eins und werden stärker mit dieser Klasse identifiziert, während die Summe der beiden Ausgabenwerte für korrekte Permutation weiterhin eins bleibt.

### 4.1.2. Simulierte Daten

Für das Training und die Analyse des neuronalen Netzes ist ein Trainingsdatensatz nötig. Da für jedes Ereignis die Klassenzugehörigkeit gegeben sein muss, werden Monte-Carlo simulierte Daten genutzt. Für den Trainingsdatensatz werden für Topereignisse mit powheg und Pythia8 simulierte Daten verwendet. Die QCD Ereignisse wurden mit Madgraph und Pythia8 erzeugt.

Nach der Anwendung der Cuts (siehe 2.3), welche auf dem tatsächlichen Datensatz angewandt sind, liegen nur noch 23 900 richtig permutierte Topereignisse vor. Deshalb werden auch von den Ereignistypen "Hintergrund und falsch permutierte Topereignisse" jeweils 23 900 Ereignisse genutzt. Zwar liegt im CMS Datensatz eine völlig andere Zusammensetzung vor, das Neuronale Netz benötigt allerdings für jeden Ereignistyp genug Trainingsereignisse um diese zuverlässig zu erkennen. Außerdem soll kein Ereignistyp versehentlich bevorzugt selektiert werden. Die generierte Topmasse des Datensatzes liegt bei  $172,5 \text{ GeV}/c^2$ .

### 4.1.3. Eingabevariablen

Die genutzten Eingabevariablen sollen die Unterscheidung der beschriebenen Ereignisklassen ermöglichen. Deshalb ist es wichtig, dass diese für die verschiedenen Klassen unterschiedliche Verteilungen folgen. Es bieten sich die kinematischen Eigenschaften aller am hadronischen Topzerfall beteiligten Teilchen an.

Es werden die Impulse  $P_T, P_x, P_y, P_z$ , die Pseudorapidität  $\eta$  und Azimutwinkel  $\phi$  sowie Energie  $E$  und Masse  $M$  für alle sechs beteiligten Teilchen genutzt. Daraus ergeben sich die folgenden Eingabeparameter:

Außerdem wird die quadratische Summe aller Transversalimpulse  $ST^2$  und die Komponenten der fehlenden Transversalenergie  $E_{T,met}$  und fehlenden Transversalimpulsen  $P_{x,met}$ ,  $P_{y,met}$  genutzt. Zusätzlich wird der in der Teilchenphysik übliche relative Winkelabstand  $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  zwischen den B Quarks, W Bosonen und den jeweiligen Quark Jets genutzt. Die Winkel- und Rapiditätsdifferenzen und kombinierten Massen der B Quarkpaare und der resultierenden der W Bosonen und die im kinematischen Fit angepasste Topmasse, sowie die Fitwahrscheinlichkeit werden ebenfalls berücksichtigt. Die Summe aller Transversalimpulse  $ST^2$ , ist wie folgt definiert:

0	$B_1$	Transversalimpuls $P_T$	24	$q_{12}$	Transversalimpuls $P_T$
1	$B_1$	Pseudorapidität $\eta$	25	$q_{12}$	Pseudorapidität $\eta$
2	$B_1$	Azimutwinkel $\phi$	26	$q_{12}$	Azimutwinkel $\phi$
3	$B_1$	Masse $m$	27	$q_{12}$	Masse $m$
4	$B_1$	Impuls $P_x$	28	$q_{12}$	Impuls $P_x$
5	$B_1$	Impuls $P_y$	29	$q_{12}$	Impuls $P_y$
6	$B_1$	Impuls $P_z$	30	$q_{12}$	Impuls $P_z$
7	$B_1$	Energie $E$	31	$q_{12}$	Energie $E$
8	$B_2$	Transversalimpuls $P_T$	32	$q_{21}$	Transversalimpuls $P_T$
9	$B_2$	Pseudorapidität $\eta$	33	$q_{21}$	Pseudorapidität $\eta$
10	$B_2$	Azimutwinkel $\phi$	34	$q_{21}$	Azimutwinkel $\phi$
11	$B_2$	Masse $m$	35	$q_{21}$	Masse $m$
12	$B_2$	Impuls $P_x$	36	$q_{21}$	Impuls $P_x$
13	$B_2$	Impuls $P_y$	37	$q_{21}$	Impuls $P_y$
14	$B_2$	Impuls $P_z$	38	$q_{21}$	Impuls $P_z$
15	$B_2$	Energie $E$	39	$q_{21}$	Energie $E$
16	$q_{11}$	Transversalimpuls $P_T$	40	$q_{22}$	Transversalimpuls $P_T$
17	$q_{11}$	Pseudorapidität $\eta$	41	$q_{22}$	Pseudorapidität $\eta$
18	$q_{11}$	Azimutwinkel $\phi$	42	$q_{22}$	Azimutwinkel $\phi$
19	$q_{11}$	Masse $m$	43	$q_{22}$	Masse $m$
20	$q_{11}$	Impuls $P_x$	44	$q_{22}$	Impuls $P_x$
21	$q_{11}$	Impuls $P_y$	45	$q_{22}$	Impuls $P_y$
22	$q_{11}$	Impuls $P_z$	46	$q_{22}$	Impuls $P_x$
23	$q_{11}$	Energie $E$	47	$q_{22}$	Energie $E$

Tabelle 4.1.: Kinematische Eigenschaften aller beteiligten Teilchen.

$$ST^2 = \sum_i p_{T,i}^2 \quad (4.6)$$

mit  $i$  werden alle beteiligten Teilchen ( $B_1, B_2, q_{11}, q_{12}, q_{21}, q_{22}$ ) bezeichnet.

Die Rapiditätsdifferenz ist definiert als:

$$\Delta\eta_{i,j} = |\eta_i - \eta_j| \quad (4.7)$$

$i$  und  $j$  bezeichnen die beteiligten Teilchen, in diesem Fall  $B_1$  und  $B_2$ ,  $q_{11}$  und  $q_{12}$  oder  $q_{21}$  und  $q_{22}$ . Die Azimutwinkeldifferenz nutzt das selbe  $i$  und  $j$ :

$$\Delta\phi_{i,j} = \left| \text{modulo} \left( \frac{\phi_i - \phi_j}{2\pi} \right) \right| - \pi \quad (4.8)$$

48	$ST^2$	Summe der Transversalimpulse
49	$E_{T,met}$	fehlende Transversalenergie
50	$P_{x,met}$	fehlender Transversalimpuls
51	$P_{y,met}$	fehlender Transversalimpuls
52	$btag_1$	btag von $B_1$
53	$btag_2$	btag von $B_2$
54	$\Delta R_{B_1,B_2}$	relativer Winkelabstand $B_1, B_2$
55	$\Delta R_{W_1,W_2}$	relativer Winkelabstand $W_1, W_2$
56	$\Delta R_{q_{11},q_{12}}$	relativer Winkelabstand $q_{11}, q_{12}$
57	$\Delta R_{q_{21},q_{22}}$	relativer Winkelabstand $q_{21}, q_{22}$
58	$\Delta \eta_{B_1,B_2}$	Pseudorapiditätsdifferenz $B_1 B_2$
59	$\Delta \phi_{B_1,B_2}$	Winkeldifferenz $B_1 B_2$
60	$\Delta \alpha_{B_1,B_2}$	Raumwinkeldifferenz $B_1 B_2$
61	$\Delta \eta_{q_{11},q_{12}}$	Pseudorapiditätsdifferenz $q_{11} q_{12}$
62	$\Delta \phi_{q_{11},q_{12}}$	Winkeldifferenz $q_{11} q_{12}$
63	$\Delta \alpha_{q_{11},q_{12}}$	Raumwinkeldifferenz $q_{11} q_{12}$
64	$\Delta \eta_{q_{21},q_{22}}$	Pseudorapiditätsdifferenz $q_{21} q_{22}$
65	$\Delta \phi_{q_{21},q_{22}}$	Winkeldifferenz $q_{21}, q_{22}$
66	$\Delta \alpha_{q_{21},q_{22}}$	Raumwinkeldifferenz $q_{21} q_{22}$
67	$M_{q_{11},q_{12}}$	kombinierte Masse der Produkte $q_{21}, q_{22}$
68	$M_{q_{21},q_{22}}$	kombinierte Masse der Produkte $q_{21}, q_{22}$
68	$M_{B_1,B_2}$	kombinierte Masse der b Quarks $B_1, B_2$
70	$m_t$	gefittete Topmasse
71	$P_{\chi^2}$	Fitwahrscheinlichkeit P- GoF

Tabelle 4.2.: Zusätzliche Eingabeparameter

Dabei gibt die modulo-Funktion den Rest der Division aus. Die Raumwinkeldifferenz  $\Delta\alpha$  wird wie folgt bestimmt:

$$\Delta\alpha_{B_1 q_{11}} = \arccos\left(\frac{P_{x,B_1} \cdot P_{x,q_{11}} + P_{y,B_1} \cdot P_{y,q_{11}} + P_{z,B_1} \cdot P_{z,q_{11}}}{\sqrt{(P_{x,B_1}^2 + P_{y,B_1}^2 + P_{z,B_1}^2)} \cdot \sqrt{(P_{x,q_{11}}^2 + P_{y,q_{11}}^2 + P_{z,q_{11}}^2)}}\right) \quad (4.9)$$

#### 4.1.4. Relevanz der Eingabevariablen

Anhand der Eingabe bestimmt das Neuronale Netz die Zugehörigkeit des jeweiligen Ereignisses zu den Klassen der Ausgabe. Eingabevariablen, die keine Aussage über die Klassenzugehörigkeit liefern, müssen von diesem erkannt und niedrig gewichtet werden. Allerdings sind nicht alle Zusammenhänge trivial und direkt ersichtlich. Deshalb ist es einerseits sinnvoll keine irrelevanten, aber auch nicht zu wenige Eingabeparameter einzuspeisen.

Im Folgenden ist eine Auswahl der Eingabeparameter graphisch dargestellt. Dabei wird nach den Ereignistypen Hintergrund sowie richtig und falsch permutierte Topereignissen unterschieden. Die Darstellungen aller weiterer Eingabevariablen sind im Anhang zu finden.

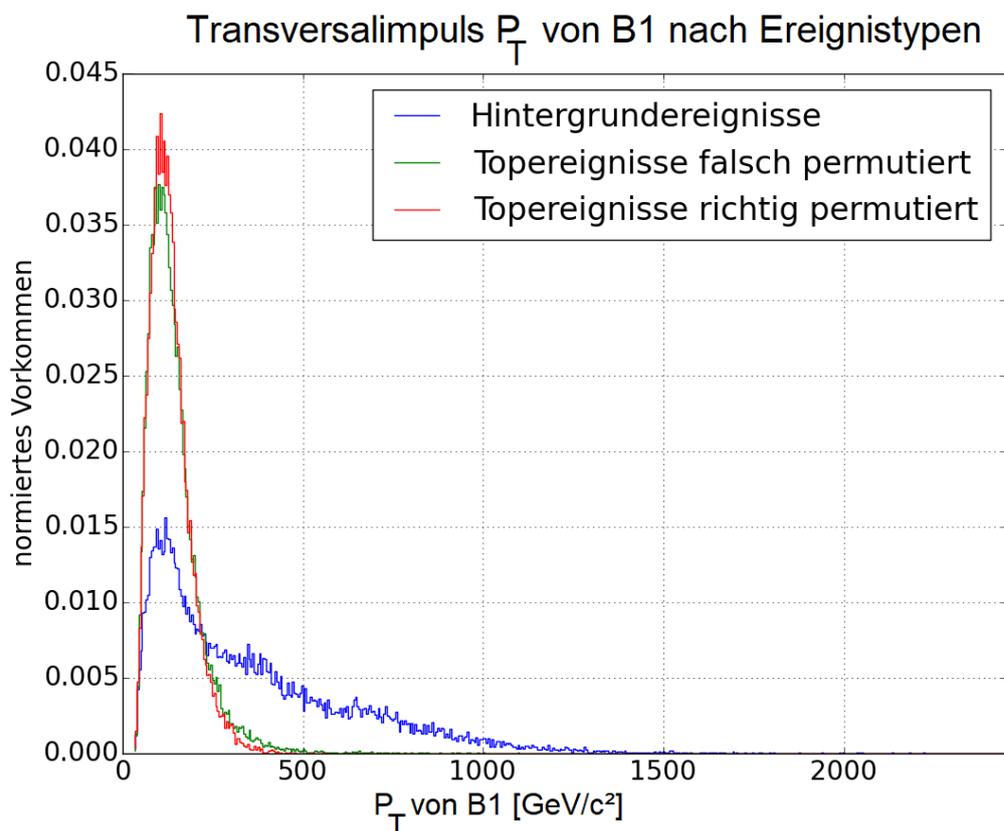


Abbildung 4.1.: Normiertes Vorkommen von Hintergrund, Richtige und Falsche Permutation von Topereignissen nach dem Transversalimpuls des  $B_1$  Quarks.

In Abbildung 4.1 ist das normierte Vorkommen von Hintergrund, richtig und falsch permutierte Topereignisse nach dem transversal Impuls von  $B_1$  aufgetragen. Es fällt auf, dass Hintergrundereignisse zu größeren transversalen Impulsen tendieren. So kann bei hohen Transversalimpulsen eine Unterscheidung zwischen Hintergrund-, richtig und falsch permutierten Ereignissen gemacht werden.

In Abbildung 4.2 ist das normierte Vorkommen der fehlenden Transversalenergie für Hintergrund, richtige und falsche Permutation von Topereignissen dargestellt.

Die Identifikation der falsch permutierten Ereignisse gestaltet sich schwieriger, da der Verlauf der Verteilungen meist sehr ähnlich zu denen für richtig permutierte oder Hinter-

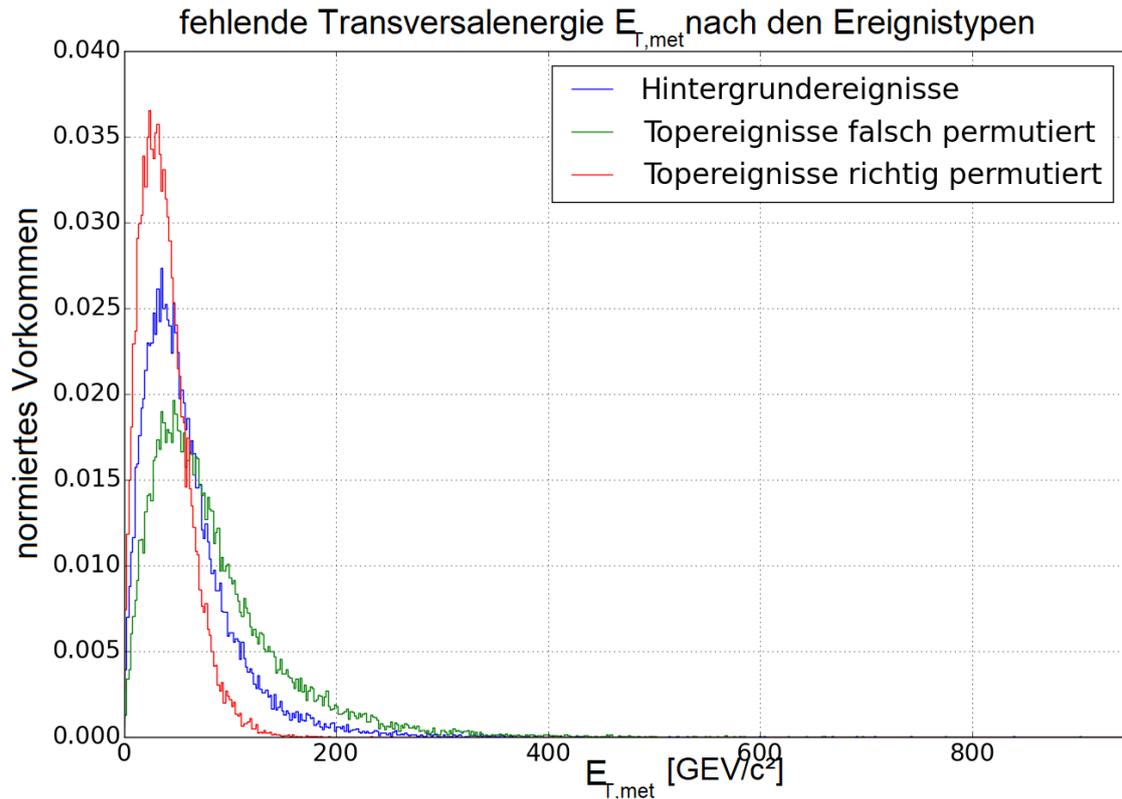


Abbildung 4.2.: Normiertes Vorkommen der fehlenden Transversalenergie  $E_{T,met}$  für die verschiedenen Klassen.

grundereignisse ist. Hierfür eignen sich allerdings die berechneten Massen der involvierten Teilchen. In Abbildung 4.3 ist das normierte Vorkommen der einzelnen Ereignistypen nach der im kinematischen Fit gefitteten Topmasse dargestellt. Nach dem Schnittpunkt der drei Verteilungen bei circa  $220 GeV/c^2$ , ist das Vorkommen von falsch permutierten und Hintergrundereignissen deutlich höher, sodass in diesem Bereich eine Unterscheidung zwischen den Klassen möglich wird. In Abbildung 4.4 ist die Rapiditätsdifferenz  $\Delta\eta_{q_{12},q_{22}}$  dargestellt. Im Bereich zwischen  $\eta > 2$  treten nahezu keine richtig permutierten Topereignisse auf. So können in diesem Bereich richtig permutierte Topereignisse fast ausgeschlossen werden.

Die verschiedenen Eingabevariablen sollen dem Neuronalen Netz die Trennung der Ereignisklassen ermöglichen. Zwar scheinen einige Parameter keine Trennung der Ereignisklassen zu ermöglichen, allerdings können diese in Kombination zu weiteren Informationen führen. Alle weiteren Graphen der Eingabevariablen sind im Anhang zu finden.

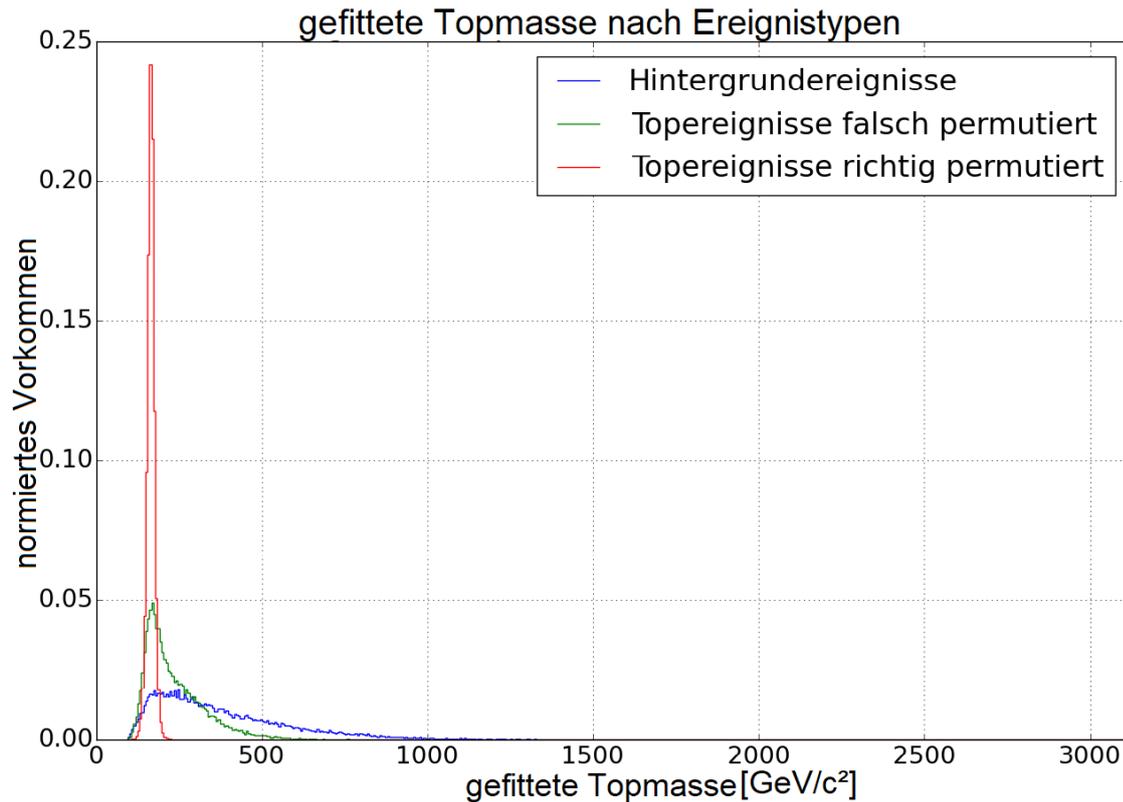


Abbildung 4.3.: Normiertes Vorkommen der gefitteten Topmasse nach den einzelnen Klassen.

## 4.2. Training

Im folgenden wird ein Neuronales Netz erstellt und verschiedene Einstellmöglichkeiten optimiert. Das Grundgerüst besteht aus 72 Eingangs- und 4 Ausgabeneuronen. Dazwischen sind 10 Schichten mit jeweils 128 Neuronen. Der Code für die Grundstruktur wurde von Torben Lange zur Verfügung gestellt.

### 4.2.1. Wahl des geeignetsten Aktivierungsfunktion

Die Aktivierungsfunktion generiert die Ausgabe der einzelnen Neuronen aus der Summe der Eingaben (siehe Kapitel 3.1).

Um die geeignetste Aktivierungsfunktion zu finden, wird ein Neuronales Netz mit verschiedenen Aktivierungsfunktionen trainiert und verglichen. Die getesteten Aktivierungsfunktionen sind ReLU, linear und sigmoid. Dabei bildet ReLU (rectified linear units) nicht auf negative Werte ab:

$$f_{relu}(x) = \max(x, 0) \quad (4.10)$$

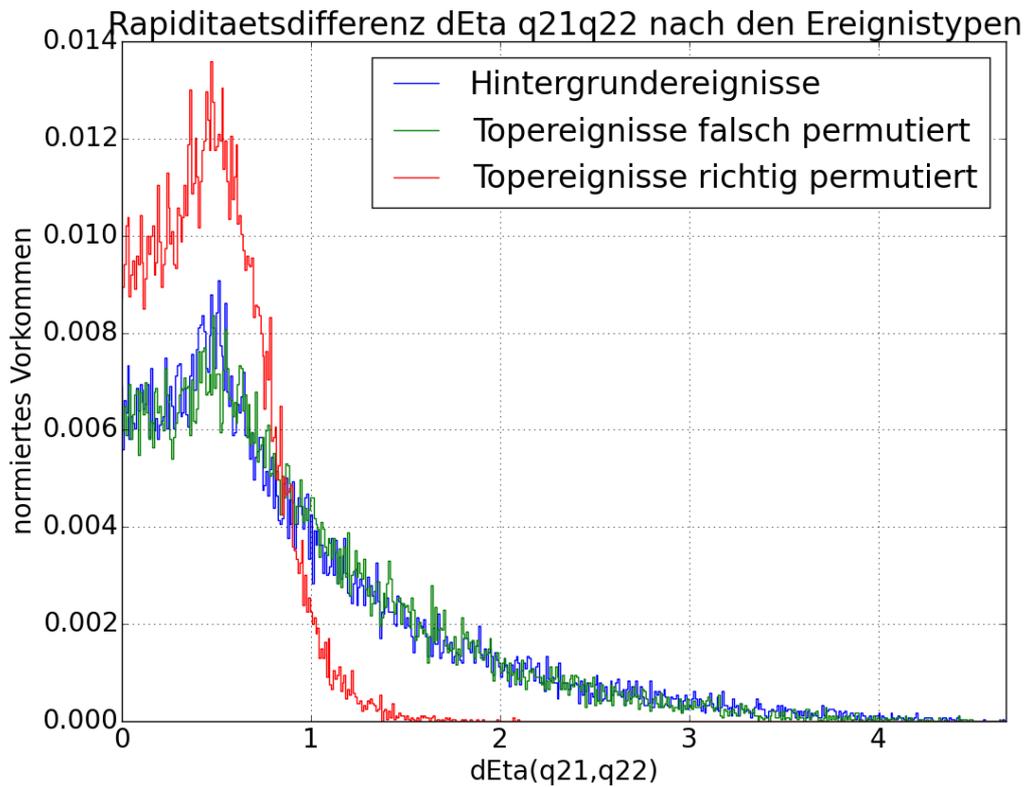


Abbildung 4.4.: Normiertes Vorkommen der Rapiditätsdifferenz  $\Delta\eta_{q_{12},q_{22}}$  zwischen den beiden Resultierenden des ersten W Bosons nach den einzelnen Klassen.

Die Sigmoidfunktion bildet ebenfalls nur auf positive Werte ab:

$$f_{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (4.11)$$

Zum Vergleich wird auch die lineare Funktion, die die Eingabe auf sich selbst abbildet ausgewertet.

Als geeignetste Aktivierungsfunktion wird diejenige mit der ausgeprägtesten Roc-Kurve, also dem geringsten Gini-Koeffizient, nach 10 Trainingsepochen angesehen. In Abbildung 4.5 sind die Roc-Kurven der verschiedenen Aktivierungsfunktionen zu sehen. Aus den Gini-Koeffizienten  $g$  kann abgelesen werden welches Neuronale Netz die beste Trennung zwischen Signal und Hintergrund vollzieht. Hier wird als Signal die Klasse richtig permutierter Topereignisse geringer Topmassenabweichung bezeichnet, die anderen Klassen werden als Hintergrund zusammengefasst.

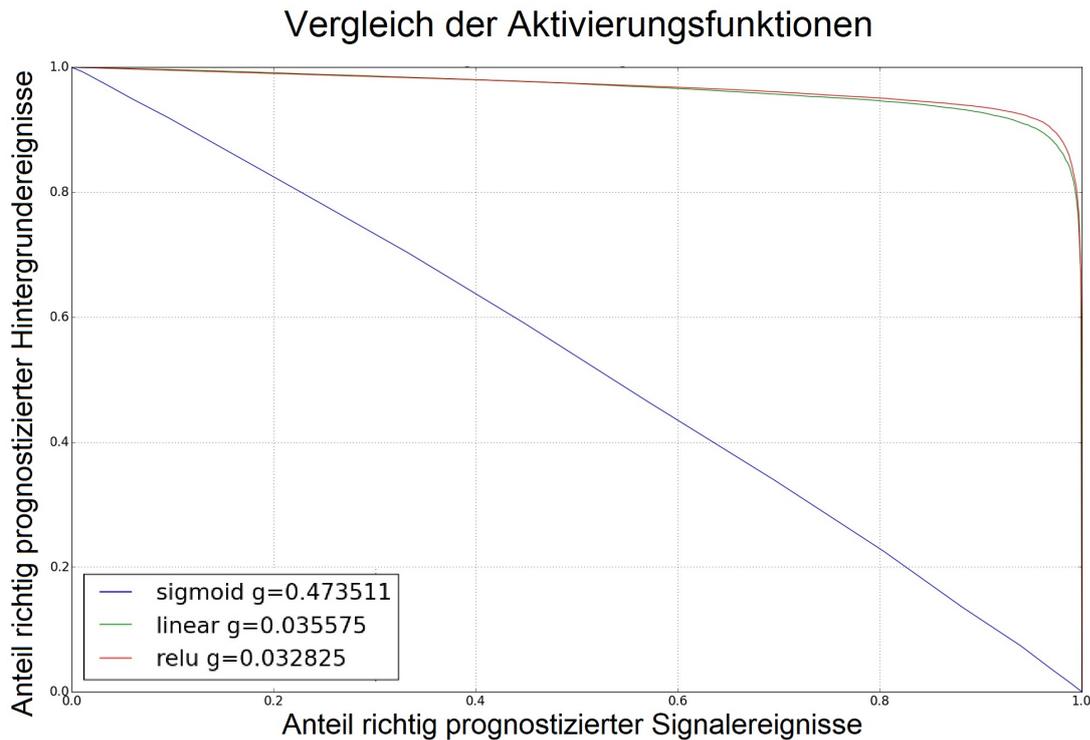


Abbildung 4.5.: Roc-Kurven und Gini-Koeffizienten  $g$  für die Aktivierungsfunktionen ReLU, sigmoid und linear, mit den richtig prognostizierten Hintergrundereignissen nach den richtig prognostizierten Signalereignissen. Hier sind richtig permutierte Topereignisse geringer Massenabweichung als Signal und die Ereignisse der übrigen Klassen als Hintergrund definiert.

Da ReLU den geringsten Gini-Koeffizient von  $g = 0.032825$  aufweist, wird diese Aktivierungsfunktion als besonders geeignet festgelegt und im folgenden genutzt.

#### 4.2.2. Wahl des geeignetsten Optimierungsverfahren

Das Optimierungsverfahren passt während des Trainings die internen Parameter des Neuronalen Netzes an, um das Ergebnis der Lossfunktion zu minimiert.

Hierfür wurden alle verfügbaren Optimierungsverfahren getestet und das geeignetste gewählt. Als geeignetstes Optimierungsverfahren wird in diesem Fall das angesehen, wel-

Aktivierungsfunktion	Gini-Koeffizient
ReLU	0.032825
linear	0.035575
sigmoid	0.473511

Tabelle 4.3.: Gini-Koeffizienten der verschiedenen Aktivierungsfunktionen.

che unter definierten Bedingungen das beste Ergebnis, also das neuronale Netz mit den besten Vorhersagen, liefert. Dabei wird das selbe Neuronale Netz mit je 10 Trainingsepochen genutzt und anhand von Roc-Kurve und Gini-Koeffizient das geeignetste festgelegt. Die ermittelten Ginikoeffizienten sind in Abbildung 4.7 abzulesen.

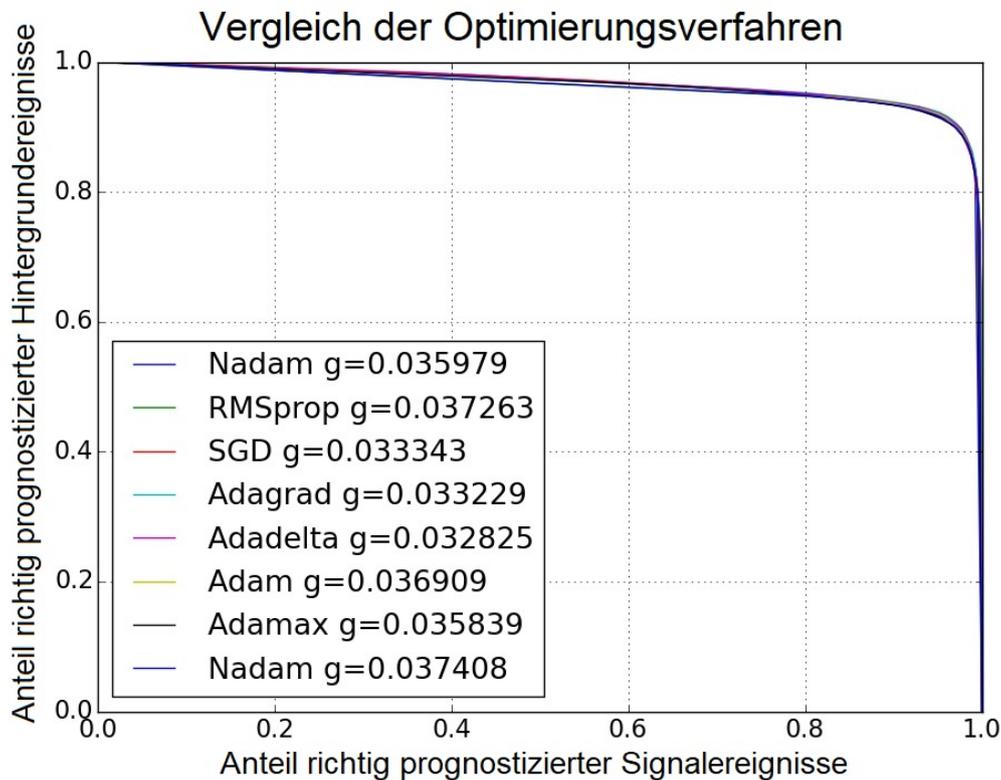


Abbildung 4.6.: Roc-Kurve für die verschiedenen Optimierungsverfahren. Den geringsten Ginifaktor liefert Adadelata mit  $g \approx 0.0328$ .

Anhand dieser Ergebnisse wird Adadelata als bester Optimierungsfunktion festgelegt. Adadelata ist im folgenden kurz beschrieben, alle weiteren sind in der Keras Dokumentation [3] beschrieben.

### Adadelata

Die Optimierungsfunktion Adadelata passt die Gewichtungsfaktoren im Lernprozess des Neuronalen Netzes an. Dabei wird die Lernrate in der Nähe eines Minimums verringert, um ein Oszillieren um das Minimum zu vermeiden. Adadelata ist eine Weiterentwicklung der Optimierungsfunktion Adagrad, dabei werden größere Änderungen an kleinen und

kleinere Änderungen an großen Gewichtungsfaktoren durchgeführt. Während Adagrad die Änderung der Gewichtungsfaktoren an Hand aller vorherigen Änderungen berechnet, nutzt Adadelta nur ein diskrete Menge an vorherigen Änderungen.[12]

### 4.2.3. Trainingsepochen

Die Länge des Trainings ist für die Leistung eines neuronalen Netzes ausschlaggebend. Bei zu vielen Trainingsepochen mit einem festen Trainingsdatensatz und einem komplexen Netz besteht die Gefahr des Übertrainings, irrelevante Charakteristika werden dabei als ausschlaggebend fehlinterpretiert. Bei zu wenigen Trainingsepochen kann das Potential des Neuronalen Netzes nicht ausgeschöpft werden.

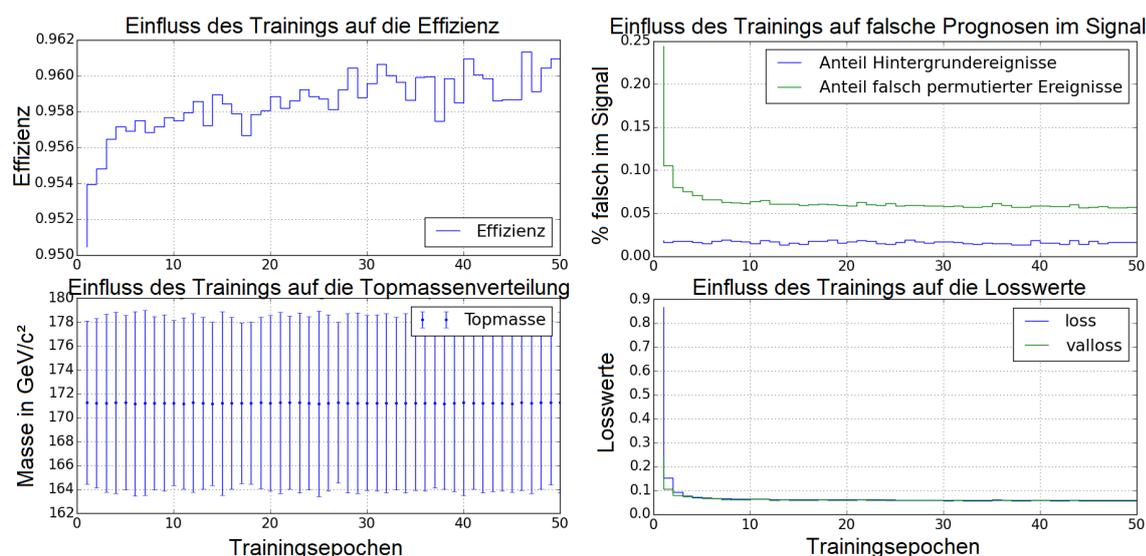


Abbildung 4.7.: Einfluss der Anzahl der Trainingsepochen auf die Effizienz der Vorhersage (oben links), die ans Voigtprofil angepasste Topmasse mit Profilbreite  $\sigma$  (unten links), die Anteile falsch als Signal klassifizierter Ereignisse (oben rechts) und die Losswerte (unten rechts).

Um beide Fehler zu vermeiden wird das Neuronale Netz über 50 Epochen trainiert, die Ergebnisse sind in Abbildung 4.8 zu sehen. Es werden nach jeder Trainingsepochen die Effizienz und die Loss- und Vergleichslosswerte ausgelesen. Außerdem wird an die Verteilung der Topmassen der prognostizierten Signalereignisse eine Voigtfunktion angelegt. Im zweiten Graphen sind die Breite  $\sigma$  und den Peak  $\mu$  dargestellt. Die Anteile falsch klassifizierter Ereignisse im Signal, im dritten Graphen. Aus diesen Größen soll die optimale

Trainingslänge abgeschätzt werden.

In diesem Training kommt es nicht zu einem Übertraining. Außerdem weisen sowohl die Topmassenverteilung als auch der Verlauf der falsch prognostizierten Ereignisse nach den ersten 10 Epochen nur noch kleine Fluktuationen auf. Die Effizienz steigt im Verlauf der 50 Epochen nur im circa einen Prozent an. Der Graph steigt die ersten 30 Epochen an und fluktuiert dann circa um  $\approx 1\%$ . Somit wird die ideale Trainingslänge auf 30 Epochen festgelegt.

Die angepasste Topmasse liegt zwischen  $171$  und  $172 \text{ GeV}/c^2$ . Die Anpassung an das Voigtprofil nach 30 Epochen ist in Abbildung 4.13 zu sehen.

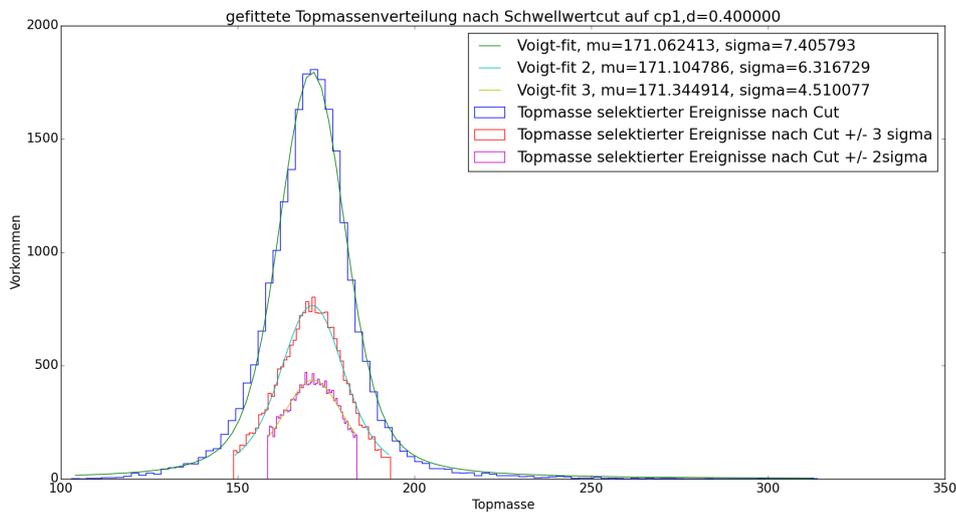


Abbildung 4.8.: Anpassung der Topmasse für prognostizierte Signalereignisse, mit einem Schwellwert von 0.4 auf die Klasse richtig permutierter Topereignisse niedriger Massenabweichung. Es wurde eine dreifache Voigtfunktionanpassung durchgeführt. Die Kurven der verschiedenen Bereiche haben unterschiedliche Höhen, da die Gesamtzahl der Bins konstant gehalten ist.

Die Voigtfunktion beschreibt die Faltung aus Gaußkurve  $G(x)$  und einer Cauchykurve  $L(x)$ :

$$V(x) = (G * L)(x) = \int G(\tau)L(x - \tau)d\tau \quad (4.12)$$

$$G(x) = \frac{e^{-x^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}} \quad (4.13)$$

$$L(x) = \frac{\gamma}{\pi(x^2 + \gamma^2)} \quad (4.14)$$

Dabei beschreibt  $\sigma$  die Breite der Gaußkurve und  $\gamma$  die Breite der Cauchyverteilung. Es wurde eine dreifache Voigtprofilanpassung durchgeführt. Dabei wurde zunächst auf die ganze Massenverteilung, daraufhin nur für den Bereich  $\mu \pm 3 \cdot \sigma$  und im folgenden für  $\mu \pm 2 \cdot \sigma$  angepasst. Die Topmassenverteilung lässt sich gut an das Voigtprofil beschreiben. Der Peak der Topmasse liegt nach der dritten Anpassung bei  $\mu = 171,34 \pm 4,51 \text{ GeV}/c^2$ . In der publizierten Topmassenstudie wird mit einem anderen Verfahren extrahiert. Aber das hier genutzte erlaubt zumindest einfache Vergleiche bezüglich der Ereignis Selektion.

#### 4.2.4. Ausgabe des Neuronalen Netzes

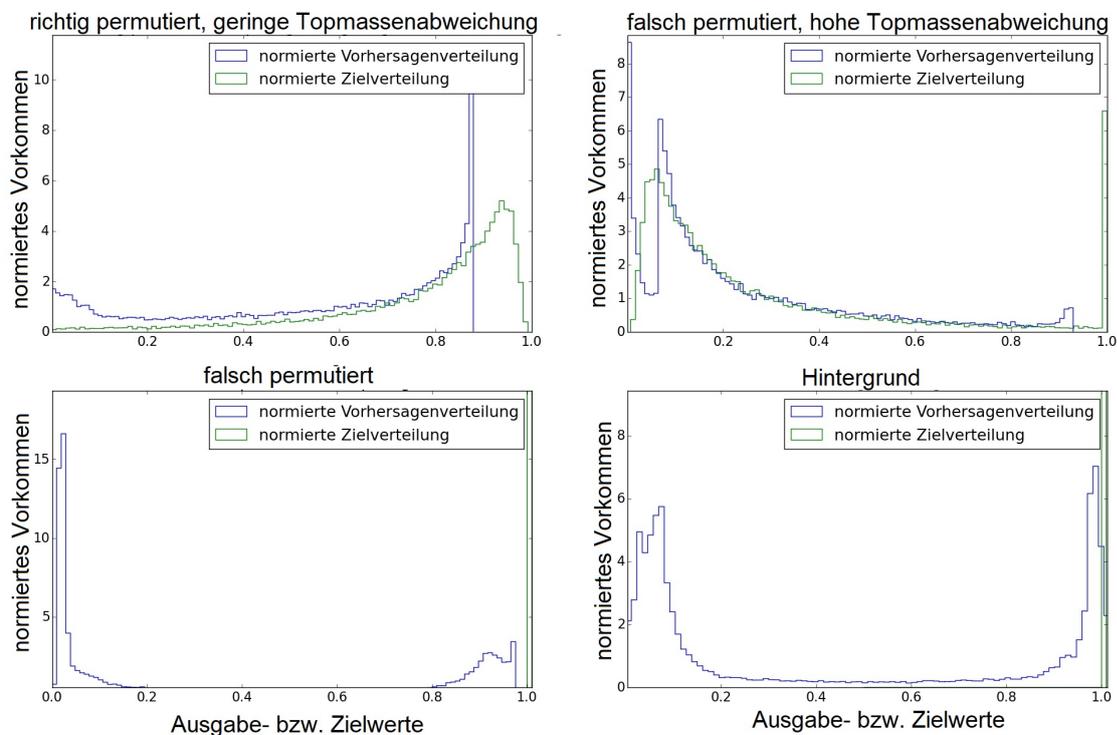


Abbildung 4.9.: Vergleich der normierten Verteilungen von Prognosen- und Zielwerten nach den einzelnen Klassen.

Um die Ausgabe des Neuronalen Netzes bewerten zu können werden die Prognosen- und Zielwerte der einzelnen Klassen untersucht. Dabei wurden die Prognosen mit dem Trainingsdatensatz erstellt. In Abbildung 4.10 sind die Verteilungen der Prognosen- und Zielwerten nach den einzelnen Klassen gezeigt. Dabei fällt auf, dass die verschiedenen Graphen für die ersten beiden Klassen wie erwartet grob der Zielfunktion folgen. Außerdem erreichen nur die Prognosen des Hintergrunds eine prognostizierte Klassenzugehör-

rigkeit von eins. Die Verteilungen für die falsch permutierten Top- und Hintergrundereignisse liegen allerdings nicht sehr gut übereinander. Die Zielwerte sind in beiden Fällen ausschließlich eins. Die Verteilungen der Prognosewerte haben jedoch beide Peaks nahe Null und nahe Eins. Zwar ist das nicht die Zielverteilung, aber kaum Ereignisse haben einen mittleren Wert als Klassenzugehörigkeit. So kann trotzdem eine Trennung der Werte vorgenommen werden. Insgesamt lassen diese Verteilungen auf eine klare Trennung der Ereignisse schließen.

#### 4.2.5. Schwellwert zur Klassifizierung der Prognosen

Das Neuronale Netz gibt eine Prognose über die Klasse eines eingespeistes Ereignis, hier ein vierdimensionaler normierter Vektor. Die vier Werte geben jeweils die prognostizierte Zugehörigkeit zu den vier Ereignisklassen an. Da am Ende richtig permutierte Topereignisse mit geringer Topmassenabweichung selektiert werden sollen, muss entschieden werden, ab welchem Schwellwert ein Ereignis zu dieser Klasse gezählt wird. Bei einem sehr hohen Schwellwert werden einige richtig permutierte Topereignisse geringer Topmassenabweichung vernachlässigt, bei einem kleinen Schwellenwert werden Ereignisse anderer Klassen mit berücksichtigt. Um den idealen Schwellwert zu finden, werden die Prognosen auf dem Trainingsdatensatz nach den Schwellwerten getrennt und die Güte der Trennung ausgewertet. Die Effizienz wird nach den Schwellwerten aufgetragen. Des weiteren ist auch der Anteil falsch prognostizierter Hintergrund und falsch permutierter Topereignisse und die angepasste Topmasse mit Breite der Verteilung der als richtig prognostizierter Ereignisse, sowie statistischer Unsicherheit zur Bewertung des Neuronalen Netzes, interessant.

In Abbildung 4.11 ist die Auswertung eines Schwellwerts  $d_{min}$ , den die Prognose für die Klasse richtig permutierter Topereignisse geringer Topmassenabweichung überschreiten muss, dargestellt.

$$P_{Klasse1} > d_{min} \quad (4.15)$$

Im oberen linken Graphen ist die Roc-Kurve dargestellt, im Vergleich ist auch der übliche Cut auf die Fitwahrscheinlichkeit zu sehen. Es fällt auf, dass die Roccurve des Neuronalen Netzes weitaus höher verläuft und somit eine bessere Trennung verspricht.

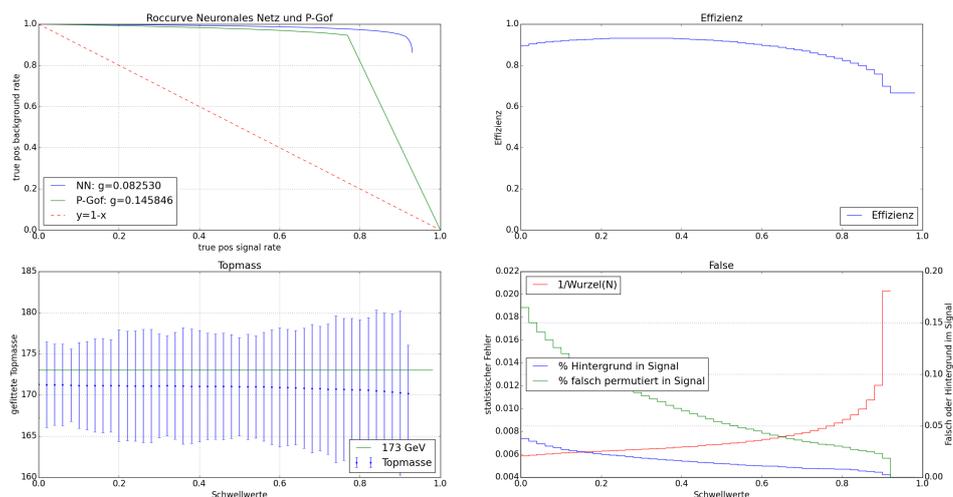


Abbildung 4.10.: Untersuchung des geeignetsten Schwellwerts der ersten Prognosekomponente anhand der Roc-Kurve, Effizienz, angepassten Topmasse und richtig und falsch prognostizierten Ereignissen, sowie statistischer Unsicherheit.

Im oberen rechten Graphen ist die Effizienz aufgetragen, der Schwellwert für die höchste Effizienz liegt zwischen 0.24 und 0.38. Im unteren linken Graphen ist die angepasste Topmasse zu sehen, die geringsten Breiten  $\sigma$  der Verteilungen treten bei einem Schwellenwert zwischen 0.1 und 0.2 auf. Bei größeren Werten werden die Verteilungen breiter. Der Peak liegt bei circa  $172\text{GeV}/c^2$  und fällt bei Schwellwerten über 0.6 ab. Der untere rechte Graph zeigt die statistische Unsicherheit und die Anteile falsch prognostizierter Ereignisse im Signal. Die Kurven verlaufen wie erwartet, mit steigendem Schwellwert sinken die falsch prognostizierten Anteile, die statistische Unsicherheit steigt allerdings. Anhand dieser wird der Schwellwert auf  $d_{min} = 0.26$  festgelegt, denn dieser liefert eine hohe Effizienz von über 90%, den Gesamtanteil falsch prognostizierter Ereignisse von unter 10%, mit einer statistischen Unsicherheit von circa 0.007%.

Auch für die anderen Ereignisklassen wird der Schwellwert festgelegt. Bei diesen allerdings nicht als minimaler Wert, sondern als maximaler. Denn die Ereignisse anderer Klassen sind im Signal unerwünscht. Die passenden Grafiken sind im Anhang zu finden. Die Schwellwerte sind in Tabelle 4.4 zu sehen.

Mit diesen Schwellwerten werden die Daten in Signal und Hintergrund getrennt. Diese Trennung liefert auf dem Trainingsdatensatz folgende Ergebnisse, Tabelle 4.5.

	Klasse	Schwellwert
1	richtig permutierte Topereignisse geringer Massenabweichung	>0.26
2	richtig permutierte Topereignisse hoher Massenabweichung	<0.6
3	falsch permutierte Topereignisse	<0.3
4	Hintergrundereignisse	<0.1

Tabelle 4.4.: Schwellwerte für die verschiedenen Ereignisklassen.

Ereignistyp	Anzahl im Datensatz	Anzahl im Signal	prozentual
richtig perm. Topereignis	23 900	20 428	85.47
falsch perm. Topereignis	23 900	953	3.99
Hintergrundereignis	23 900	291	1.22

Tabelle 4.5.: Selektionsergebnisse für die ausgewählten Schwellenwerte.

Diese Trennung selektiert besonders gut gegen Hintergrundereignisse. Außerdem fällt auf, dass nur 85,47% der richtig permutierten Topereignisse als Signal erkannt wird. Dies liegt allerdings an der Trennung zwischen richtig permutierten Topereignissen hoher und geringer Topmassenabweichung und ist somit erwünscht. In Abbildung 4.12 ist das Topmassenspektrum des Signals zu sehen. Auch hier wurde eine dreifache Anpassung an das Voigtprofil durchgeführt. Im folgenden wird  $\gamma$  auf die Zerfallsbreite des Top Quarks, von  $\Gamma = 1.41 \text{ GeV}7c^2$  festgelegt [9]. Die Funktion wurde zuerst auf dem kompletten Spektrum, dann auf  $\pm 2\sigma$  und  $\pm \sigma$  angepasst, um den Peak ideal zu beschreiben.

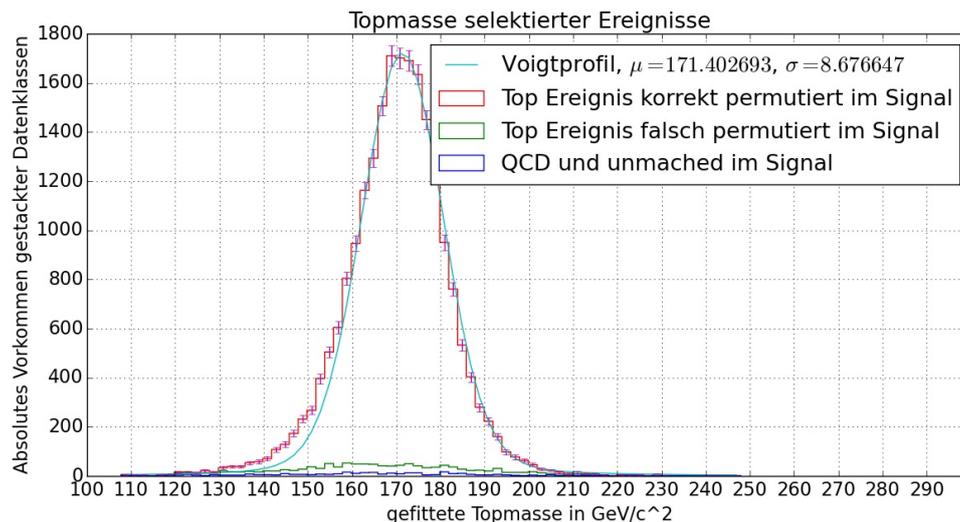


Abbildung 4.11.: Verteilung der Topmasse nach Schwellwertcut. An die Verteilung wurde ein Voigtfunktion angepasst. Außerdem sind die Anteile der anderen Klassen im Signal in grün und blau eingezeichnet.

### 4.3. Analyse der Trainingsergebnisse

Das Neuronale Netz wird durch seine Struktur im Hinblick auf das beste Ergebnis definiert und trainiert. Um den Erfolg des Trainings zu gewährleisten, wird die Leistung des Neuronalen Netzes auf Trainingsdaten überprüft.

#### 4.3.1. Signifikanz der Eingabeparameter

In Kapitel 4.1.3 und 4.1.4 werden Eingabeparameter für das Neuronale Netz festgelegt. Dabei stellt sich die Frage, welche Eingabeparameter welchen Einfluss auf die Prognose haben. Um diese Frage zu klären werden die Eingabeparameter im Bereich der Varianz verändert. Dabei wird auf alle Impulse, Massen und Energien eine Varianz von 2% angenommen. Für  $E_{T,met}$  wird eine Varianz von 4.9% und  $P_{x,met}$ ,  $P_{y,met}$  von 3.5% angenommen. Für die übrigen Parameter wurde die Breite der Verteilung der relativen Abweichung als Varianz angenommen. Diese wurden aus den Abweichungen zwischen den Rekonstruierten Jets im Monte-Carlo Datensatz bestimmt. Wird durch diese Variation des Eingabeparameters die Prognose verändert, muss dieser einen Einfluss haben. Dafür werden alle Eingabeparameter um ihre Varianz variiert und der so geänderte Datensatz in das Neuronale Netz eingespeist. In der Tabelle sind nun die Ergebnisse für den Peak der Topmassenverteilung, deren Breite, der Anteil an richtig prognostizierten Signalereignissen und die prozentuale Änderung von Signalereignissen zu sehen. Die entsprechenden Werte bei unverändertem Datensatz liegen bei:  $170.7746 \text{ GeV}/c^2$ ,  $8.1216 \text{ GeV}/c^2$ , 0.9645 und 18968.

Den größten Einfluss auf den Peak der Topmassenverteilung haben, die Topmasse  $m_t$ , die kombinierten Masse  $M_{B_1, B_2}$ ,  $M_{q_{11}, q_{12}}$ ,  $M_{q_{21}, q_{22}}$  und  $\Delta R_{W_1, W_2}$ .

	Var [%]	$\Delta$ Peak	$\Delta$ Breite	$\Delta$ Effizienz	$\Delta$ N
$P_{T, B_1}$	0.02	0.0010   0.0004	0.0029   0.00799	-0.0000   -0.0002	-0.0009   0.0011
$\eta_{B_1}$	0.022	0.0000   0.0000	0.0000   0.0000	0.0000   0.0000	0.0000   0.0000
$\phi_{B_1}$	0.011	0.0001   -0.0003	-0.0020   0.0007	0.0000   -0.0000	0.0001   -0.0001
$M_{B_1}$	0.02	0.0001   -0.0003	-0.0020   0.0007	0.0000   -0.0000	0.0001   -0.0001
$P_{x, B_1}$	0.02	-0.0002   -0.0025	0.0048   -0.0058	-0.0000   -0.0000	-0.0003   0.0004
$P_{y, B_1}$	0.02	-0.0002   -0.0003	0.0006   0.0041	-0.0001   0.0000	-0.0001   0.0001
$P_{z, B_1}$	0.02	0.0019   -0.0001	-0.0012   0.0036	-0.0000   -0.0000	-0.0003   -0.0001
$E_{B_1}$	0.02	-0.0005   -0.0077	-0.0065   0.01912	0.0001   -0.0001	0.0029   -0.0031

	Var [%]	$\Delta$ Peak	$\Delta$ Breite	$\Delta$ Effizienz	$\Delta$ N
$P_{T,B_2}$	0.02	-0.0033   0.0031	0.0066   -0.0052	0.0001   -0.0000	-0.0021   0.0018
$\eta_{B_2}$	0.022	0.0000   0.0000	0.0000   0.0000	0.0000   0.0000	0.0000   0.0000
$\phi_{B_2}$	0.011	-0.0002   0.0000	0.0026   0.0000	-0.0000   0.0000	-0.0001   0.0000
$M_{B_2}$	0.02	-0.0003   0.0000	0.0007   0.0000	-0.0000   0.0000	-0.0001   0.0000
$P_{x,B_2}$	0.02	0.0020   -0.0002	-0.0002   -0.0013	-0.0000   0.0000	-0.0002   0.0000
$P_{y,B_2}$	0.02	-0.0002   0.0000	-0.0013   0.0000	0.0000   0.0000	0.0000   0.0000
$P_{z,B_2}$	0.02	0.0021   0.0003	0.0003   -0.0038	-0.0000   0.0000	-0.0002   0.0002
$E_{B_2}$	0.02	0.0025   -0.0000	-0.0050   0.0038	0.0000   0.0000	0.0002   -0.0004
$P_{T,q_{11}}$	0.02	-0.0027   0.0014	0.0055   -0.0005	0.0001   -0.0001	-0.0015   0.0014
$\eta_{q_{11}}$	0.047	0.0005   0.0000	0.0008   0.0000	-0.0000   0.0000	-0.0001   0.0000
$\phi_{q_{11}}$	0.030	0.0022   0.0007	-0.0033   0.0078	-0.0000   -0.0000	-0.0002   -0.0001
$M_{q_{11}}$	0.02	-0.0007   -0.0009	0.0021   -0.0045	-0.0000   -0.0000	-0.0009   0.0006
$P_{x,q_{11}}$	0.02	-0.0002   -0.0010	0.0022   -0.0021	-0.0000   -0.0001	-0.0003   -0.0001
$P_{y,q_{11}}$	0.02	-0.0001   -0.0004	-0.0014   -0.0012	0.0000   0.0000	0.0000   0.0001
$P_{z,q_{11}}$	0.02	0.0010   0.0006	0.0027   0.0001	-0.0000   0.0000	-0.0004   0.0001
$E_{q_{11}}$	0.02	0.0021   -0.0027	-0.0067   0.0060	-0.0000   -0.0000	0.0016   -0.0018
$P_{T,q_{12}}$	0.02	0.0013   0.0021	0.0031   -0.0029	0.0000   -0.0003	-0.0026   0.0027
$\eta_{q_{12}}$	0.047	-0.0005   0.0000	0.0027   0.0000	-0.0000   0.0000	-0.0001   0.0000
$\phi_{q_{12}}$	0.030	0.0001   0.0001	-0.0018   0.0015	-0.0001   0.0000	0.0001   0.0002
$M_{q_{12}}$	0.02	-0.0001   -0.0008	0.0061   -0.0037	-0.0000   -0.0000	-0.0002   0.0002
$P_{x,q_{12}}$	0.02	-0.0007   -0.0006	0.0001   -0.0025	-0.0000   -0.0001	-0.0001   0.0001
$P_{y,q_{12}}$	0.02	-0.0003   0.0001	0.0007   -0.0020	-0.0000   0.0000	-0.0001   0.0001
$P_{z,q_{12}}$	0.02	0.0012   -0.0003	0.0007   0.0007	-0.0000   -0.0000	-0.0002   -0.0001
$E_{q_{12}}$	0.02	-0.0005   -0.0002	-0.0055   -0.0006	-0.0001   -0.0000	0.0003   -0.0006
$P_{T,q_{21}}$	0.02	0.0007   0.0019	0.0105   -0.0016	0.0001   -0.0003	-0.0027   0.0028
$\eta_{q_{21}}$	0.047	0.0004   -0.0002	0.0034   -0.0013	-0.0000   0.0000	-0.0001   0.0000
$\phi_{q_{21}}$	0.030	-0.0008   0.0006	-0.0018   -0.0012	-0.0000   -0.0001	-0.0001   0.0000
$M_{q_{21}}$	0.02	0.0003   -0.0006	0.0077   -0.0031	-0.0000   -0.0000	-0.0007   0.0005
$P_{x,q_{21}}$	0.02	0.0005   -0.0003	0.0068   -0.0070	-0.0000   -0.0000	-0.0003   0.0002
$P_{y,q_{21}}$	0.02	0.0000   -0.0014	0.0000   0.0033	0.0000   0.0000	0.0000   0.0000
$P_{z,q_{21}}$	0.02	0.0008   -0.0005	0.0010   -0.0011	-0.0000   0.0000	-0.0003   0.0001
$E_{q_{21}}$	0.02	-0.0004   0.0022	-0.0083   0.0060	-0.0002   0.0000	0.0012   -0.0010
$P_{T,q_{22}}$	0.02	0.0027   0.0044	0.0068   -0.0050	0.0001   -0.0002	-0.0032   0.0030
$\eta_{q_{22}}$	0.047	-0.0003   -0.0005	0.0007   0.0028	-0.0000   -0.0000	-0.0001   -0.0001
$\phi_{q_{22}}$	0.030	-0.0002   0.0005	0.0026   0.0008	-0.0000   -0.0000	-0.0001   -0.0001
$M_{q_{22}}$	0.02	-0.0002   0.0001	0.0026   -0.0020	-0.0000   -0.0001	-0.0001   0.0001
$P_{x,q_{22}}$	0.02	-0.0003   -0.0005	0.0007   -0.0027	-0.0000   0.0000	-0.0001   0.0001
$P_{y,q_{22}}$	0.02	-0.0003   0.0001	0.0026   -0.0020	-0.0001   0.0000	-0.0001   0.0001
$P_{z,q_{22}}$	0.02	0.0001   -0.0006	-0.0001   0.0009	0.0000   0.0000	0.0000   0.0000
$E_{q_{22}}$	0.02	0.0006   0.0021	-0.0059   -0.0041	-0.0000   -0.0000	0.0003   -0.0003
$ST^2$	0.02	0.0050   -0.0084	-0.0078   0.0108	0.0000   0.0000	0.0026   -0.0024
$E_{T,met}$	0.049	-0.0043   0.0007	-0.0035   0.0010	0.0004   -0.0001	-0.0002   0.0001
$P_{x,met}$	0.035	0.0005   0.0004	-0.0007   0.0012	-0.0001   -0.0000	0.0000   -0.0001
$P_{y,met}$	0.035	-0.0080   0.0115	-0.0099   0.0122	-0.0006   0.0004	0.0103   -0.0095
btag1	0.021	0.0020   0.0009	-0.0067   -0.0121	0.0002   -0.0006	-0.0082   0.0091
btag2	0.024	-0.0049   0.0027	-0.0015   -0.0177	0.0002   -0.0006	-0.0100   0.0111

	Var [%]	$\Delta$ Peak	$\Delta$ Breite	$\Delta$ Effizienz	$\Delta$ N
$\Delta R_{q_{11},q_{12}}$	0.011	-0.0013 0.0021	0.0054  0.0059	0.0000 -0.0000	-0.0007 0.0006
$\Delta R_{W_1,W_2}$	0.071	-0.0321  0.0315	0.0262   -0.0201	0.0006 -0.0005	-0.0128 0.0156
$\Delta R_{q_{11},q_{12}}$	0.0249	-0.0066  0.0077	-0.0012   -0.0089	-0.0001 -0.0000	0.0009 -0.0008
$\Delta R_{q_{21},q_{22}}$	0.025	-0.0014  0.0040	-0.0011   -0.0012	-0.0001 0.0000	0.0010 -0.0012
$\Delta \eta_{B_1,B_2}$	-0.0133	-0.0009  0.0000	0.0002  0.0022	-0.0000 -0.0000	0.0001 -0.0001
$\Delta \phi_{B_1,B_2}$	0.014	0.0025  0.0008	0.0044  0.0011	-0.0000 -0.0000	-0.0008 0.0005
$\Delta \alpha_{B_1,B_2}$	0.013	0.0010 0.0008	0.0020   -0.0032	-0.0000 -0.0000	-0.0002 0.0003
$\Delta \eta_{q_{11},q_{12}}$	0.035	-0.0020  0.0001	0.0024  0.0071	-0.0001 0.0001	0.0013 -0.0015
$\Delta \phi_{q_{11},q_{12}}$	0.031	-0.0013  0.0050	-0.0028  0.0018	-0.0001 0.0001	0.0014 -0.0015
$\Delta \alpha_{q_{11},q_{12}}$	0.042	-0.0022  0.0056	-0.0007   -0.0016	-0.0001 0.0000	0.0012 -0.0014
$\Delta \eta_{q_{21},q_{22}}$	0.035	-0.0001  0.0002	-0.0049  0.0031	-0.0000 0.0000	0.0008 -0.0012
$\Delta \phi_{q_{21},q_{22}}$	0.031	0.0004  0.0024	-0.0037  0.0062	-0.0002 0.0000	0.0011 -0.0011
$\Delta \alpha_{q_{21},q_{22}}$	0.0415	0.0019  0.0026	-0.0017  0.0075	-0.0000 0.0000	0.0009 -0.0012
$M_{q_{12},q_{22}}$	0.089	0.0600   -0.0073	-0.0618   -0.0625	-0.0039 0.00345	0.0561 -0.0498
$M_{q_{21},q_{22}}$	0.089	0.0573  0.0014	-0.0634   -0.0247	-0.0026 0.0024	0.0435 -0.0382
$M_{B_1,B_2}$	-0.088	0.0103   -0.0202	-0.0013   -0.0024	-0.0005 0.0001	0.0070 -0.0044
$m_t$	0.011	-1.6758  1.6896	-0.0965  0.0958	-0.0008 0.0005	0.0157 -0.0132
$P_{\chi^2}$	0.02	0.0002  0.0023	0.0005  0.0018	0.0001 -0.0000	-0.0017 0.0017

Tabelle 4.6.: Ergebnisse der Varianz der Eingabeparameter. Es sind die relative Verschiebung des Peaks  $\Delta$  Peak und dessen Breite  $\Delta$  Breite, sowie die relative Änderung der Effizienz und der Größe des selektierten Datensatzes  $\Delta N$  angegeben.

---

## 5. Anwendung des Neuronales Netzes

### 5.1. Einfluss der Topmasse des Trainingsdatensatzes auf die Topmasse der selektierten Daten

Da das Neuronale Netz mit simulierten Monte-Carlo Daten trainiert wird, ist es wichtig zu überprüfen, wie diese das Ergebnis beeinflussen. Im folgenden werden weitere Monte-Carlo simulierte Daten mit anderen generierten Topmassen genutzt:

```
/mtop_jsf_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8RunIISummer16Mini  
AODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_allCut/  
2017
```

```
-07-24_GC98692c161673/mtop1715_jsf1.00/
```

```
/mtop_jsf_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8RunIISummer16Mini  
AODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_allCut/  
2017
```

```
-07-24_GC98692c161673/mtop1735_jsf1.00/
```

```
/mtop_jsf_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8RunIISummer16Mini  
AODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_allCut/  
2017
```

```
-07-24_GC98692c161673/mtop1755_jsf1.00/
```

```
/mtop_jsf_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8RunIISummer16Mini  
AODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_allCut/  
2017
```

```
-07-24_GC98692c161673/mtop1785_jsf1.00/
```

Die Trainingsdatensätze bestehen, wie im bisherigen Training, zu je circa 1/3 aus den einzelnen Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse. We-

---

gen mangelnder Daten haben die Trainingsdatensätze für  $171,5\text{GeV}/c^2$ ,  $173,5\text{GeV}/c^2$  und  $178,5\text{GeV}/c^2$  nur eine Gesamtgröße von circa 18000 Ereignissen. Die Trainingsdatensatz mit der Topmasse und  $175,5\text{GeV}/c^2$  hat fast 60000 Ereignisse.

### 5.1.1. Ergebnisse für simulierten Daten

Nach dem Training wird für jedes Netz individuell der geeignetste Schwellenwert identifiziert und angewendet (analog Kapitel 4.3.5). Daraus lassen sich die Massenverteilungen bestimmen, dabei wird zwischen den tatsächlichen Klassen der selektierten Daten unterschieden. Die Verteilung wird an ein Voigtfunktion vierfach angepasst.

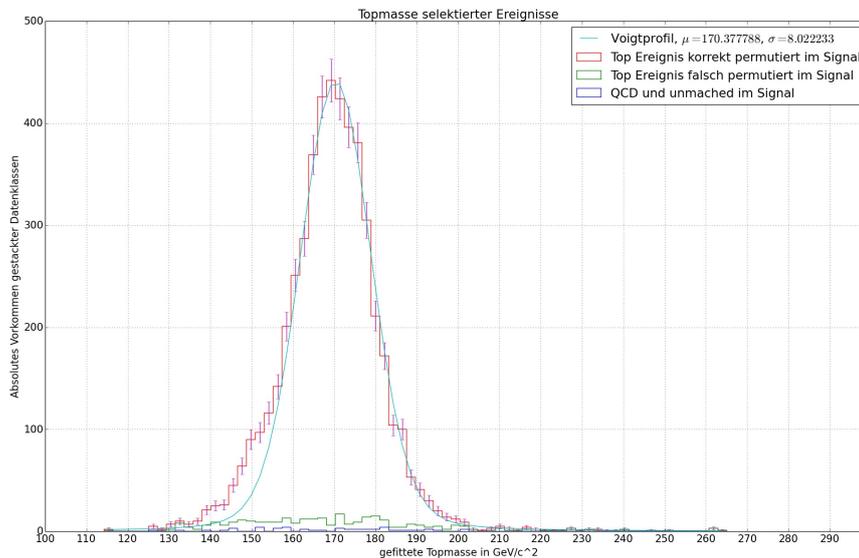


Abbildung 5.1.: Die Massenverteilung der Topmasse für ein mit der Topmasse  $171,5\text{GeV}/c^2$  trainiertes Neuronales Netz.

Die Massenverteilungen für die mit der Topmasse  $171,5\text{GeV}/c^2$  trainierten Neuronales Netz ist in Abbildung 5.1 dargestellt. Die übrigen Massenverteilungen sind im Anhang zu finden. Aus den gefitteten Peaks der Topmassenverteilungen und den generierten Topmassen des Trainingsdatensatzes ergibt sich ein linearer Zusammenhang (siehe Abbildung 5.2):

$$y = 0.980357 \cdot x + 2.2491 \frac{\text{GeV}}{c^2} \quad (5.1)$$

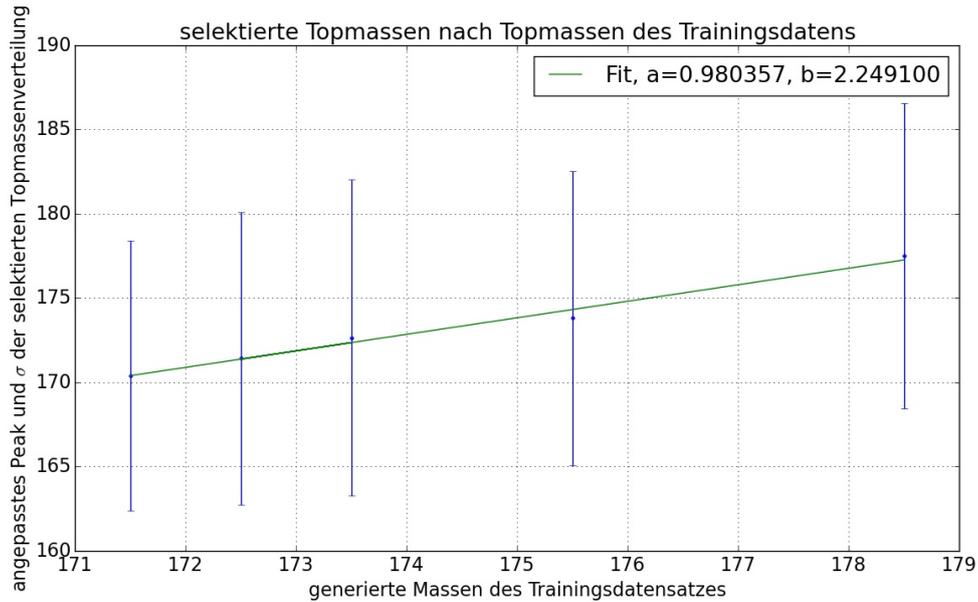


Abbildung 5.2.: Die gefitteten Topmassen der verschiedenen Neuronalen Netze nach den generierten Topmassen der Trainingsdatensätze. Der lineare Zusammenhang lässt sich durch  $y = a * x + b$  beschreiben, wobei a und b mit Hilfe eines linearen Fits angepasst wurden.

X und y sind die generierte Topmasse des Trainingsdatensatzes und der angepasste Peak der Topmassenverteilungen. Somit kann mit dieser linearen Funktion, von der generierten Topmasse des Trainingsdatensatzes auf den angepassten Peak der Massenverteilung und umgekehrt geschlossen werden.

Damit ist gezeigt, dass das Neuronale Netz keine Verschiebung der Topmassen aufgrund der generierten Topmasse des Trainingsdatensatzes aufweist.

### 5.1.2. Ergebnisse für Messdaten

In die Neuronalen Netze werden im folgenden die tatsächlichen CMS-Daten eingespeist. In diesem Fall wird allerdings eine kompliziertere Anpassung der Voigtfunktion vorgenommen. Dabei wird die Funktion auf den Bereichen  $\pm 3\sigma$ ,  $\pm 2\sigma$ ,  $\pm \sigma$  und  $\pm 0,75\sigma$  angepasst.

Hier sind nur die Ergebnisse für die generierte Topmasse von  $171,5 GeV/c^2$  und den bisher genutzten Trainingsdatensatz mit  $172,5 GeV/c^2$  dargestellt. Alle weiteren Ergebnisse sind im Anhang zu finden.

Die Topmasse des Datensatzes kann mit Hilfe der Umkehrfunktion der in Kapitel 5.1.1

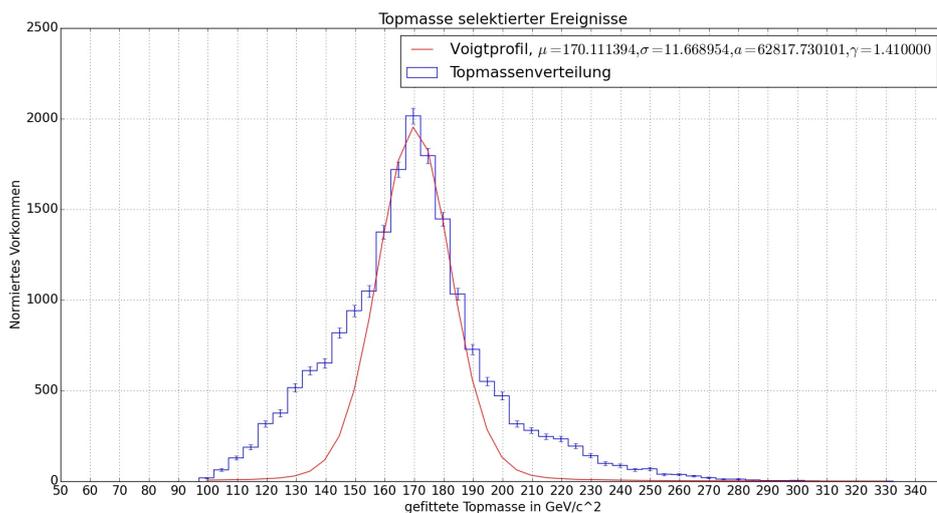


Abbildung 5.3.: Die Topmassenverteilung der selektierten Daten für ein mit der Topmasse  $171,5\text{GeV}/c^2$  trainiertes Netz, mit resultierendem Voigtprofil. Das Topmassenpeak liegt bei  $170,38\text{GeV}/c^2$  mit einer Breite von  $11,25\text{GeV}/c^2$ .

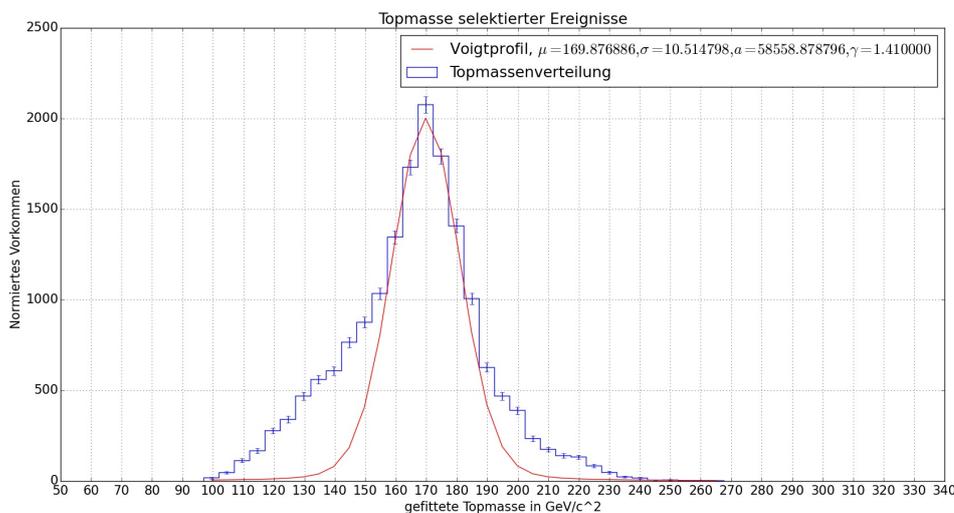


Abbildung 5.4.: Die Topmassenverteilung der selektierten Daten für ein mit der Topmasse  $172,5\text{GeV}/c^2$  trainiertes Netz, mit resultierender Voigtfunktion. Das Topmassenpeak liegt bei  $179,118\text{GeV}/c^2$  mit einer Breite von  $11,67\text{GeV}/c^2$ .

bestimmten linearen Gleichung rekonstruiert werden:

$$m_{top, \text{Datensatz}} = \frac{\mu - 2,249105 \frac{\text{GeV}}{c^2}}{0,980367} \quad (5.2)$$

Dabei ist  $\mu$  der angepasste Peak der Massenverteilung und  $m_{top, \text{Datensatz}}$  die tatsächliche Topmasse des Datensatzes. So kann analog die Breite  $\sigma$  berechnet werden. Für die mit verschiedenen Topmassen trainierten Netze ergeben sich folgende Werte:

gen. Topmasse [ $\frac{GeV}{c^2}$ ]	$\mu \pm \sigma$ [ $\frac{GeV}{c^2}$ ]	$m_{top, \text{Datensatz}} \pm \sigma$ [ $\frac{GeV}{c^2}$ ]
171,5	170,111394 ± 11,66895	171,22567 ± 9,608586
172,5	169,876886 ± 10,514798	170,986468 ± 8,431309
173,5	170,0202363 ± 11,72341	171,132691 ± 9,664138
175,5	169,938668 ± 10,90829	171,049488 ± 8,832685
178,5	170,338796 ± 11,505119	171,457633 ± 9,441473

Tabelle 5.1.: Rekonstruierte Topmassen der mit verschiedenen generierten Topmassen trainierten Neuronalen Netze.

Das neuronale Netz mit der generierten Topmasse im Trainingsdatensatz von 172,5  $GeV/c^2$  hat mit 71 700 Ereignissen den größten Datensatz fürs Training genutzt.

## 5.2. Vergleich der Ereignisselektion der Neuronalen Netze mit PGOF-Cut

Im folgenden werden die Ergebnisse der Ereignisselektion der Neuronalen Netze mit der des PGOF-Cut verglichen. In diesem Fall werden alle Ereignisse mit  $P(\chi^2) > 0,1$  dem selektierten Datensatz zugeordnet. Dann wird berechnet welcher Anteil der richtig permutierten Topereignisse im Signal zu finden ist. Außerdem werden auch die Anteile an Hintergrund- und falsch permutierten Topereignissen, welche nicht heraus selektiert wurden bestimmt. Die Ergebnisse sind in Tabelle 5.2 zu sehen:

gen. $m_{top}$ [ $GeV/c^2$ ]	N	Neuronale Netze			PGOF-Cut		
		% Top	% falsch	% QCD	% Top	% falsch	% QCD
171,5	18 000	79,78	5,30	0,9	62,41	5,97	0,78
172,5	71 700	85,47	3,99	1,22	61,71	5,66	0,83
173,5	18 000	76,01	4,02	0,8	61,92	5,32	0,8
175,5	59 842	94,07	3,91	0,78	61,91	5,63	0,69
178,5	17 747	89,02	5,53	1,11	62,10	5,63	0,85

Tabelle 5.2.: Vergleich der Anteile der Ereignistypen im selektierten Signal. Es wird zwischen der Selektion der Neuronalen Netze und des PGOF-Cuts unterschieden.

Es zeigt sich, dass die Neuronalen Netze einen weitaus größeren Teil der richtig permutierten Topereignisse erkennen als der PGOF-Cut. Außerdem sind die Anteile falsch permutierter Topereignisse im selektierten Datensatz für das Neuronale geringer also für die Cut-Methode. Nur für die Hintergrundereignisse liefert der PGOF-Cut bessere Ergebnisse als die Neuronalen Netze. Allerdings sind für beide Methoden die Anteile an falsch selektierten Hintergrundereignissen um  $\approx 1\%$ .

---

## 6. Ergebnisse

Die Ergebnisse aller Neuronalen Netze sind in der Tabelle dargestellt. Dabei bezeichnet N die Anzahl an Ereignissen im Trainingsdatensatz. Die Masse des Trainingsdatensatzes ist mit  $gen. m_{top}$  bezeichnet,  $reko. m_{top}$  ist die aus dem Massenspektrum rekonstruierte Topmasse. Außerdem sind die Anteile der verschiedenen Ereignistypen, welche die Neuronalen Netze als Signal klassifiziert haben angegeben.

$gen. m_{top}[GeV/c^2]$	N	richtig [%]	falsch [%]	QCD [%]	$reko. m_{top}[GeV/c^2]$
171,5	18 000	79,78	5,30	0,9	$171,22567 \pm 9,608586$
172,5	71 700	85,47	3,99	1,22	$170,986468 \pm 8,431309$
173,5	18 000	76,01	4,02	0,8	$171,132691 \pm 9,664138$
175,5	59 842	94,07	3,91	0,78	$171,049488 \pm 8,832685$
178,5	17 747	89,02	5,53	1,11	$171,457633 \pm 9,441473$

Tabelle 6.1.: Ergebnisse der verschiedenen Neuronalen Netze, die nach der Masse ihres Trainingsdatensatzes  $gen. m_{top}$  benannt sind. Außerdem sind die Anteile die pro Ereignisklasse als Signal identifiziert wurden angegeben. Es wird zwischen richtig oder falsch permutierten Topereignissen und Hintergrund(QCD) unterschieden. Die letzte Spalte zeigt die aus den Daten des CMS Experiments erhaltenen Topmassen.

Es ist bemerkenswert, dass die aus den Daten erhaltenen top-Massen unabhängig von der Trainingsmasse sind. Dies ist ein wichtiger Test für die Anwendbarkeit der Methode. Es fällt auf, dass die rekonstruierten Topmassen circa  $2 GeV/c^2$  von dem Literaturwert  $173,1 \pm 0,6 GeV/c^2$  [9] und etwa ein  $GeV/c^2$  von der präzisesten CMS Topmassenmessung abweicht. Dies muss in einer entsprechenden Kalibration der Methode berücksichtigt werden.

Außerdem fällt auf, dass die Neuronalen Netze eine sehr gute Selektion der richtig permutierten Topereignisse ermöglicht, bis zu 94,07% dieser wurden vom Neuronalen Netz identifiziert. Von den falsch permutierten Topereignissen werden circa 3,9 bis 5,5 % als Signal fehlinterpretiert.

Im Vergleich zu den Ergebnissen des PGOF-Cuts mit  $P(\chi^2) > 0.1$  liefern die Neuronalen Netze für die genutzten Datensätzen eine 1,23 bis 1,52 -fach bessere Selektion auf richtig permutierte Topereignisse. Außerdem werden 1,78 bis 30,55 % weniger falsch permutierte Topereignisse dem Signal zugeordnet. Die Selektion von Hintergrundereignissen kann allerdings nicht verbessert werden. Das Neuronale Netz fehlinterpretiert bis zu 1,47 mal mehr Hintergrundereignisse als Signal, als der PGOF-Cut. Trotzdem bleibt der Anteil falsch interpretierter Hintergrundereignisse auch für die neuronalen Netze unter 1,25%. Leider konnten in dieser Bachelorarbeit, wegen der strengen Cuts nur sehr kleine Datensätze für das Training der Neuronalen Netze genutzt werden. Es wäre weiterhin interessant die Anwendung Neuronaler Netze zur Ereignisselektion mit größeren Trainingsdatensätzen zu untersuchen.

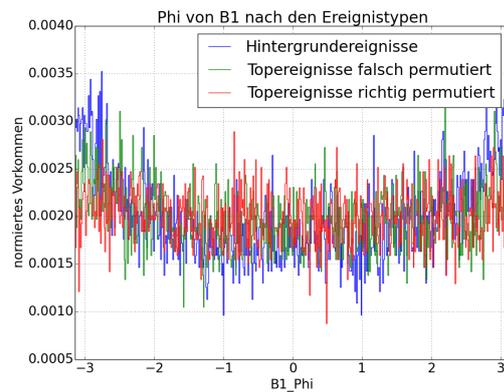
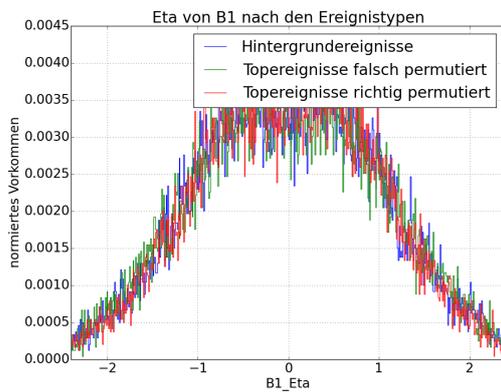
---

# Literatur

- [1] David Barney. *CMS Detector Slice*. CMS Collection. 2016. URL: `\url{https://cds.cern.ch/record/2120661}`.
  - [2] Florian Beaudette. “The CMS Particle Flow Algorithm”. In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, S. 295–304. arXiv: 1401.8155 [hep-ex]. URL: `https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf`.
  - [3] François Chollet u. a. *Keras*. `https://github.com/keras-team/keras`. 2015.
  - [4] Chrislb. *Schema eines kuenstlichen Neurons: Kuenstliches Neuron*. via Wikimedia Commons. 2005. URL: `\url{https://upload.wikimedia.org/wikipedia/commons/7/7f/ArtificialNeuronModel_deutsch.png}`.
  - [5] Mysid Dake. *A simplified view of an artifiical neural network*. via Wikimedia Commons. 2006. URL: `\url{https://upload.wikimedia.org/wikipedia/commons/3/3d/Neural_network.svg}`.
  - [6] Cristina Ferro. “B-tagging in CMS”. In: *EPJ Web Conf.* 28 (2012), S. 12055. DOI: 10.1051/epjconf/20122812055. arXiv: 1201.5292 [hep-ex].
  - [7] *LHC Guide*. 2017. URL: `{https://cds.cern.ch/record/2255762}`.
  - [8] B.R. Martin und G. Shaw. *Particle Physics*. Manchester Physics Series. Wiley, 2008. ISBN: 9780470721537. URL: `https://books.google.de/books?id=whIbrWJdEJQC`.
-

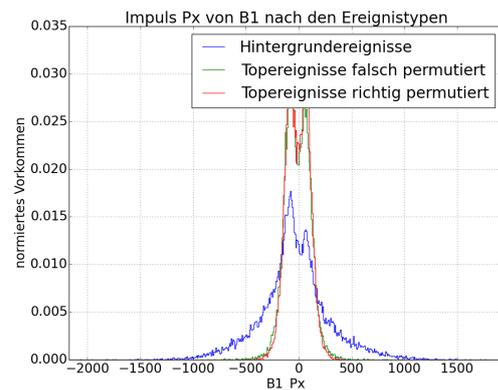
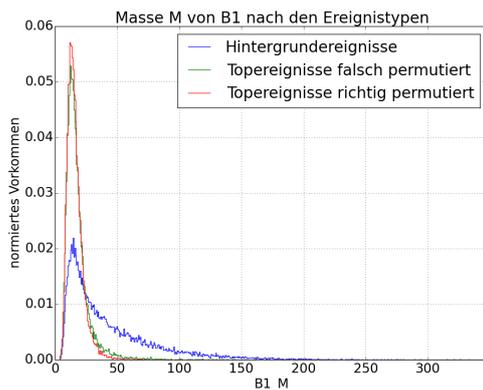
- [9] C. Patrignani u. a. “Review of Particle Physics”. In: *Chin. Phys.* C40.10 (2016), S. 100001. DOI: 10.1088/1674-1137/40/10/100001.
- [10] Gavin P. Salam. “Towards Jetography”. In: *Eur. Phys. J.* C67 (2010), S. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].
- [11] E. S. Schlieckau. “Measurement of Top-Quark Mass and Inclusive Top-Quark-Pair Production Cross Section in pp Collisions at  $s = 7/8$  TeV with CMS”. Diss. Universitaet Hamburg, 2014.
- [12] Matthew D. Zeiler. “ADADELTA: An Adaptive Learning Rate Method”. In: *CoRR* abs/1212.5701 (2012). arXiv: 1212.5701. URL: <http://arxiv.org/abs/1212.5701>.
-

# A. Anhang



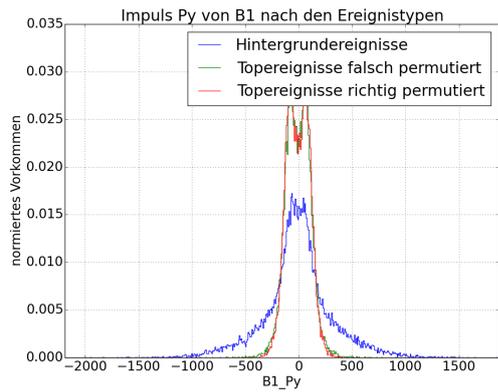
(1) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Rapidität  $\eta_{B_1}$  des ersten b Quarks.

(2) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem Azimutwinkel  $\phi_{B_1}$  des ersten b Quarks.

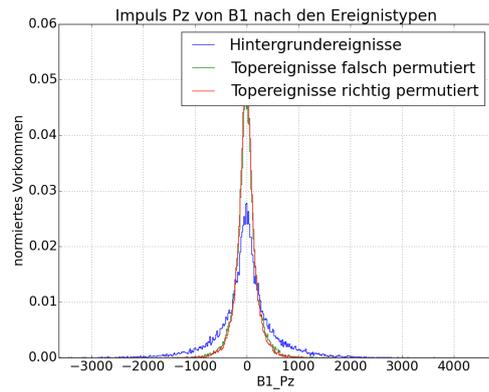


(3) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Masse  $m_{B_1}$  des ersten b Quarks.

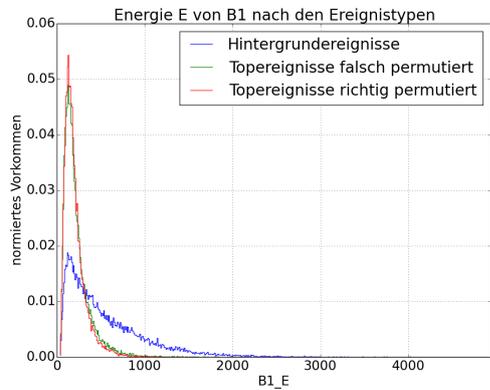
(4) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem x Impuls  $P_{x,B_1}$  des ersten b Quarks.



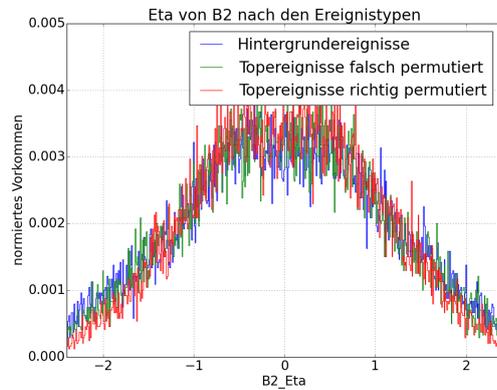
(5) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem y Impuls  $P_{y,B_1}$  des ersten b Quarks.



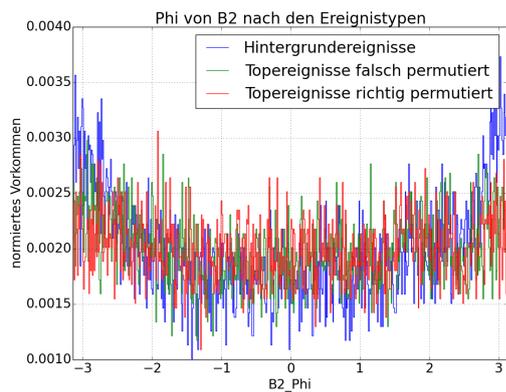
(6) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem z Impuls  $P_{z,B_1}$  des ersten b Quarks.



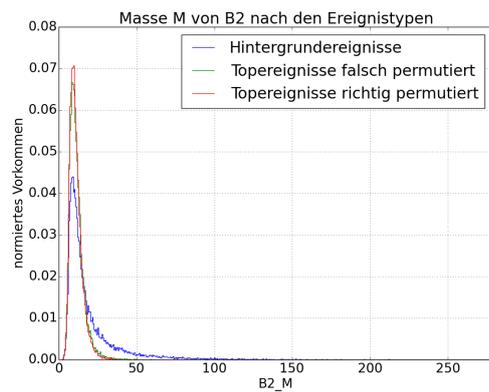
(7) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Energie  $E_{B_2}$  des ersten b Quarks.



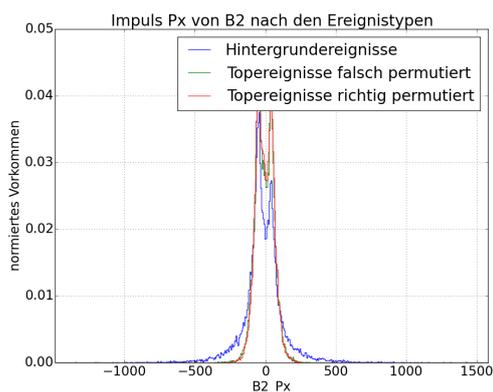
(8) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Rapidität  $\eta_{B_2}$  des zweiten b Quarks.



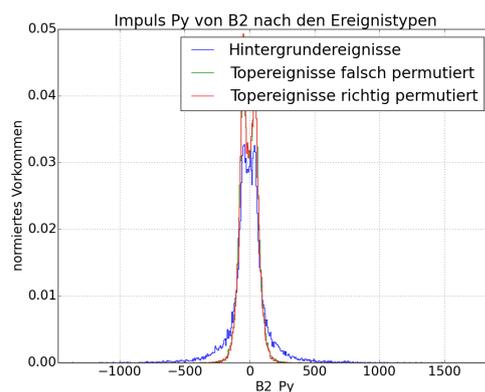
(9) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem Azimutwinkel  $\phi_{B_2}$  des zweiten b Quarks.



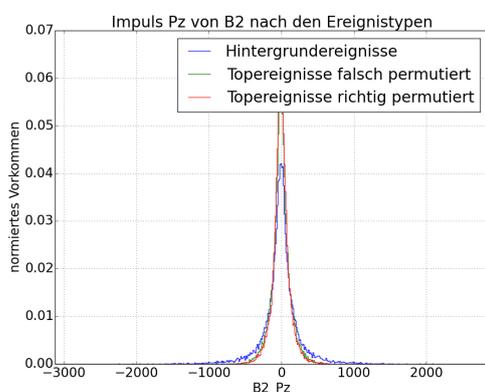
(10) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Masse  $m_{B_2}$  des zweiten b Quarks.



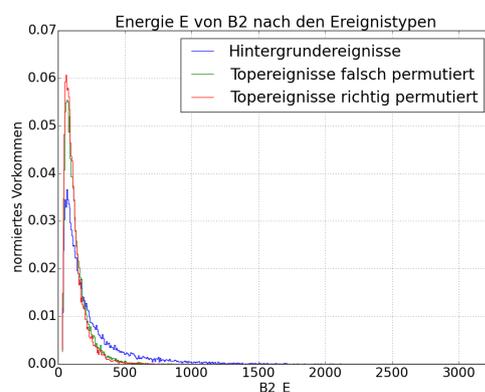
(11) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem x Impuls  $P_{x,B_2}$  des ersten b Quarks.



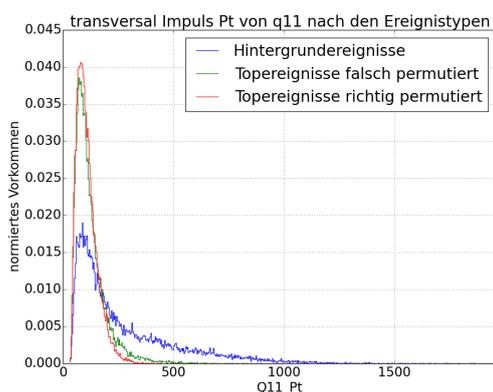
(12) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem y Impuls  $P_{y,B_2}$  des zweiten b Quarks.



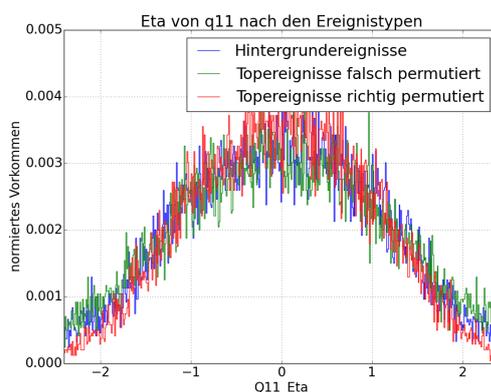
(13) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem z Impuls  $P_{z,B_2}$  des zweiten b Quarks.



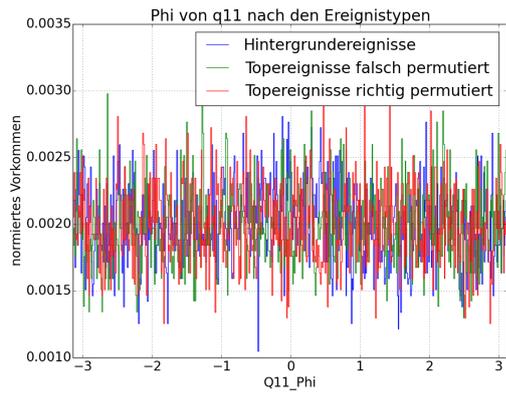
(14) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Energie  $E_{B_2}$  des zweiten b Quarks.



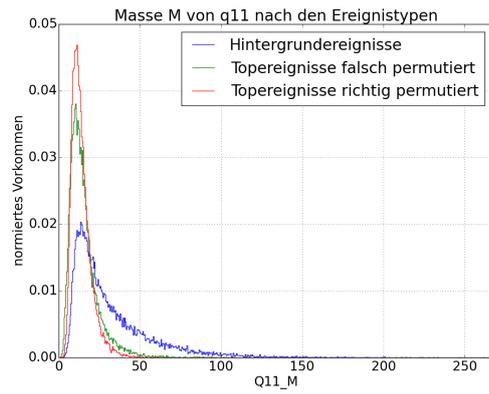
(15) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Transversalimpuls  $P_{T,q_{11}}$  des  $q_{11}$  Quarks.



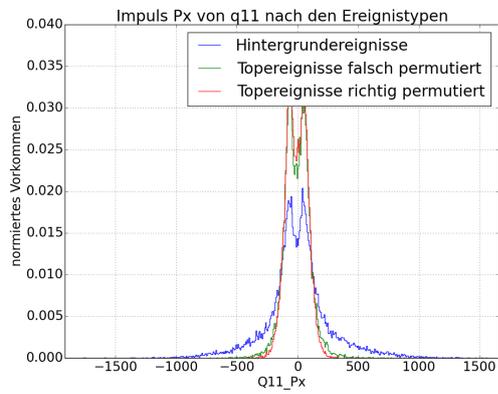
(16) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Rapidität  $\eta_{q_{11}}$  des  $q_{11}$  Quarks.



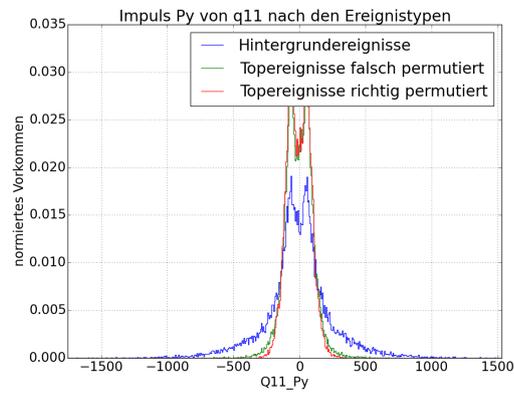
(17) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem Azimutwinkel  $\phi_{q_{11}}$  des  $q_{11}$  Quarks.



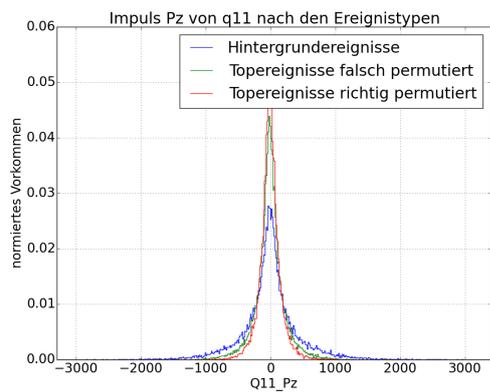
(18) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Masse  $m_{q_{11}}$  des  $q_{11}$  Quarks.



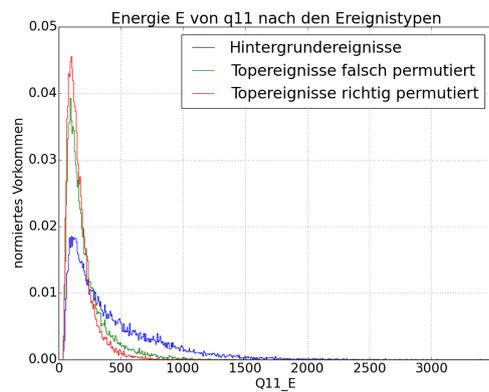
(19) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem x Impuls  $P_{x,q_{11}}$  des  $q_{11}$  Quarks.



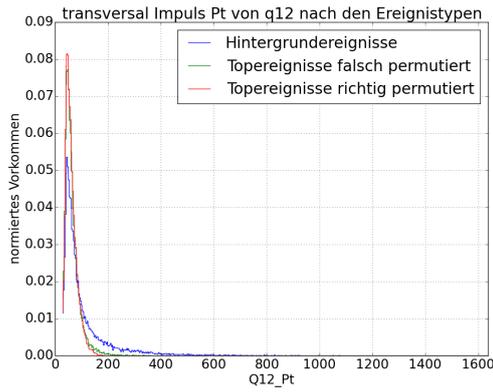
(20) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem y Impuls  $P_{y,q_{11}}$  des  $q_{11}$  Quarks.



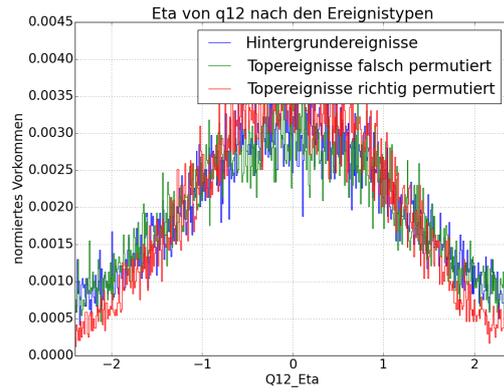
(21) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem z Impuls  $P_{z,q_{11}}$  des  $q_{11}$  Quarks.



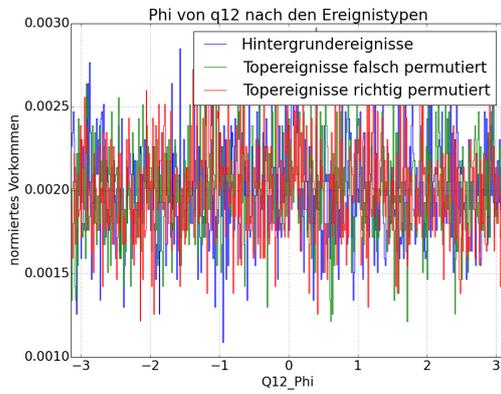
(22) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Energie  $E_{q_{11}}$  des  $q_{11}$  Quarks.



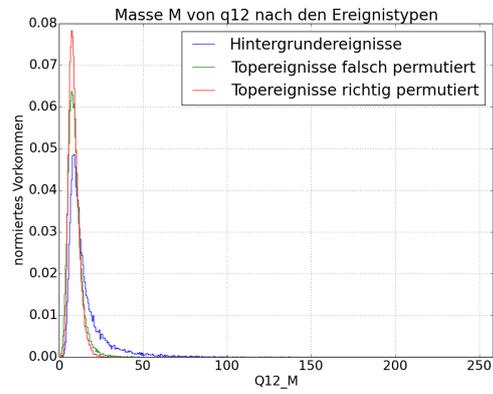
(23) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Transversalimpuls  $P_{T,q_{12}}$  des  $q_{12}$  Quarks.



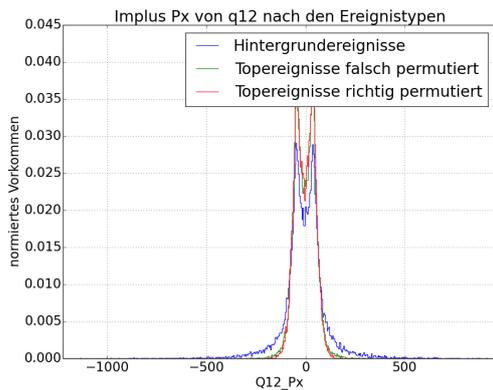
(24) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Rapidität  $\eta_{q_{12}}$  des  $q_{12}$  Quarks.



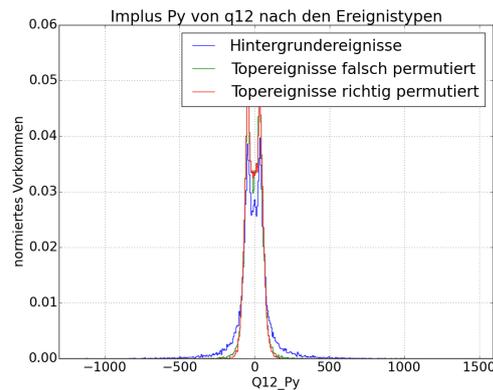
(25) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem Azimutwinkel  $\phi_{q_{12}}$  des  $q_{12}$  Quarks.



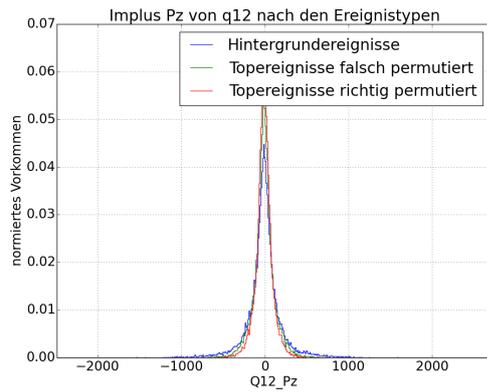
(26) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Masse  $m_{q_{12}}$  des  $q_{12}$  Quarks.



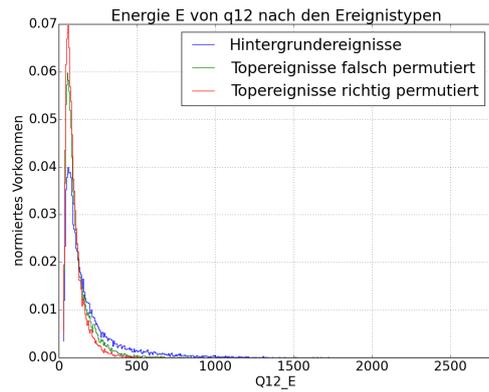
(27) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem x Impuls  $P_{x,q_{12}}$  des  $q_{12}$  Quarks.



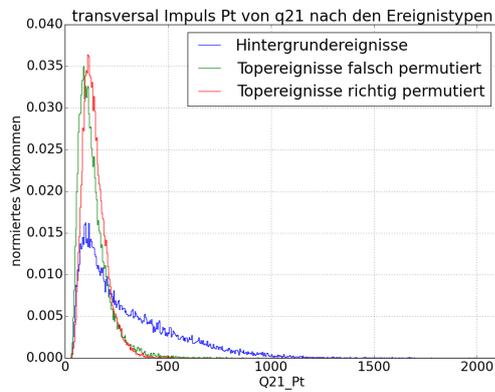
(28) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem y Impuls  $P_{y,q_{12}}$  des  $q_{12}$  Quarks.



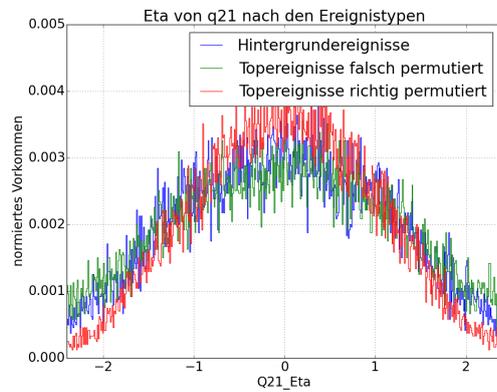
(29) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem z Impuls  $P_{z,q_{12}}$  des  $q_{12}$  Quarks.



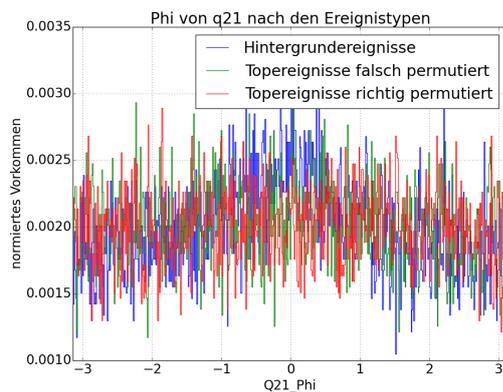
(30) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Energie  $E_{q_{12}}$  des  $q_{12}$  Quarks.



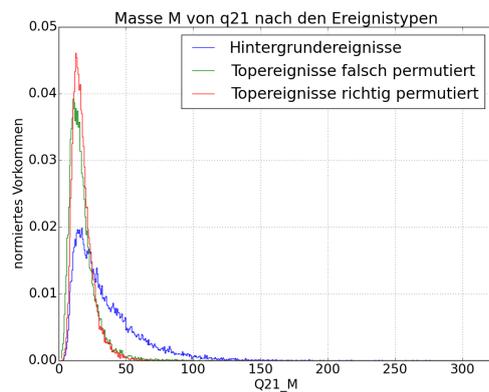
(31) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Transversalimpuls  $P_{T,q_{21}}$  des  $q_{21}$  Quarks.



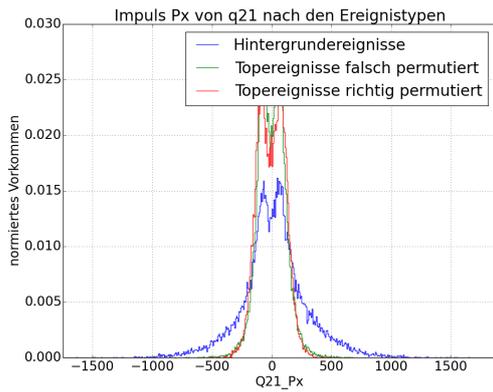
(32) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Rapidität  $\eta_{q_{21}}$  des  $q_{21}$  Quarks.



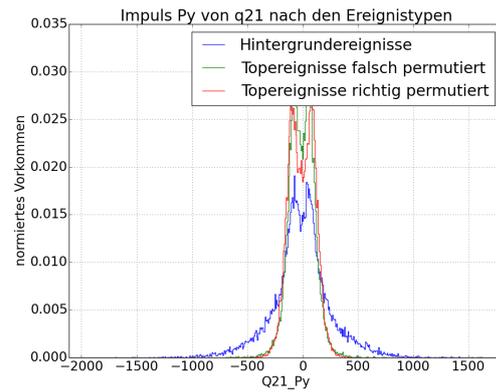
(33) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem Azimutwinkel  $\phi_{q_{21}}$  des  $q_{21}$  Quarks.



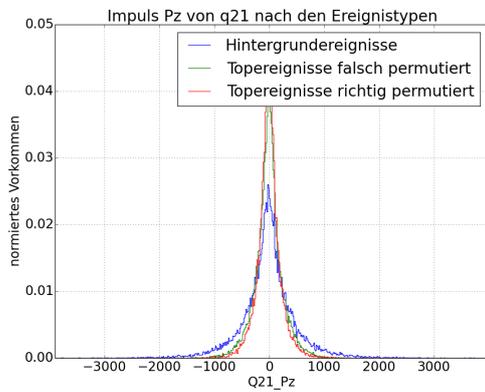
(34) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Masse  $m_{q_{21}}$  des  $q_{21}$  Quarks.



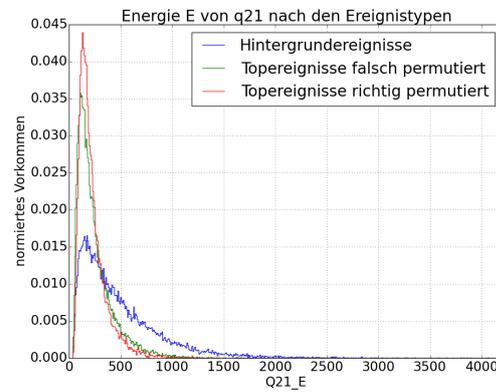
(35) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem x Impuls  $P_{x,q_{21}}$  des  $q_{21}$  Quarks.



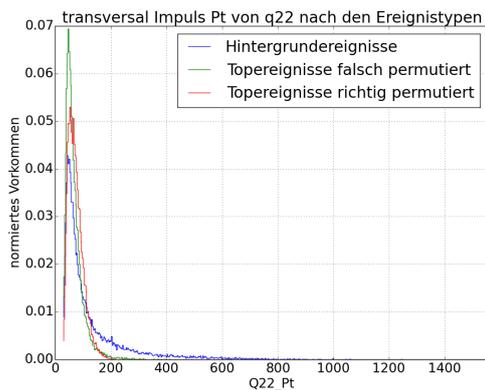
(36) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem y Impuls  $P_{y,q_{21}}$  des  $q_{21}$  Quarks.



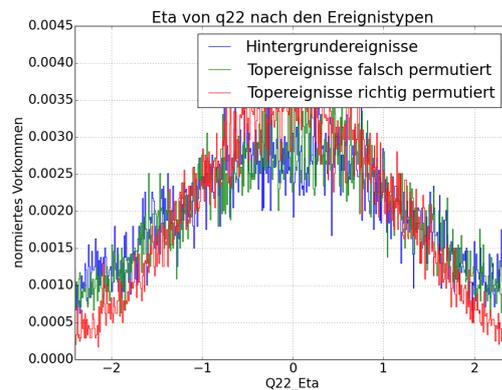
(37) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem z Impuls  $P_{z,q_{21}}$  des  $q_{21}$  Quarks.



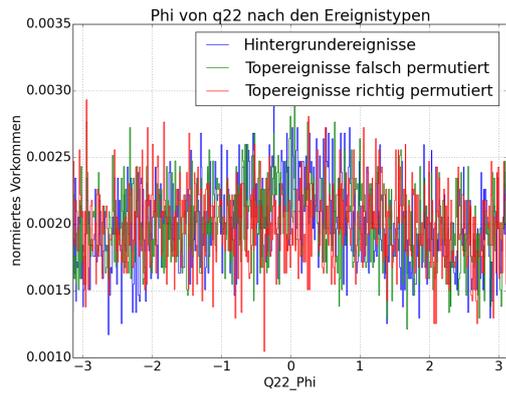
(38) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Energie  $E_{q_{21}}$  des  $q_{21}$  Quarks.



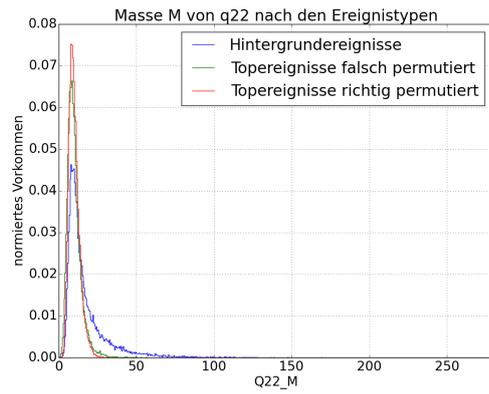
(39) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Transversalimpuls  $P_{T,q_{22}}$  des  $q_{22}$  Quarks.



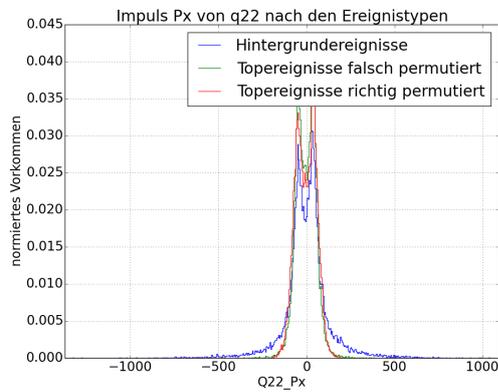
(40) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Rapidität  $\eta_{q_{22}}$  des  $q_{22}$  Quarks.



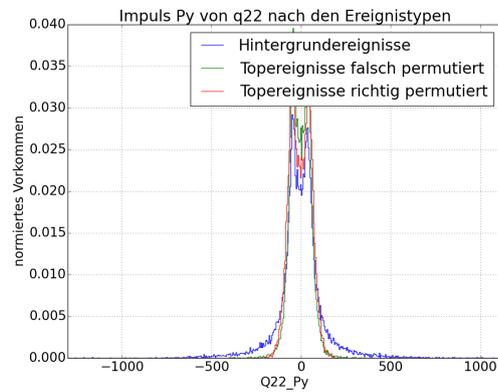
(41) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem Azimutwinkel  $\phi_{q22}$  des  $q22$  Quarks.



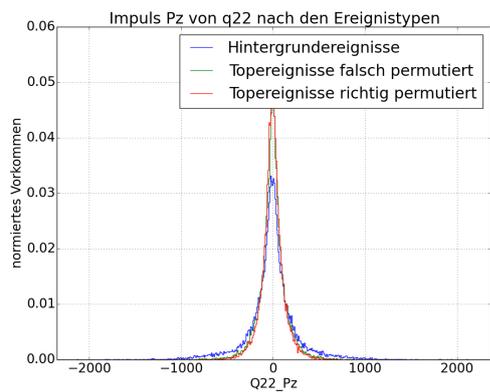
(42) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Masse  $m_{q22}$  des  $q22$  Quarks.



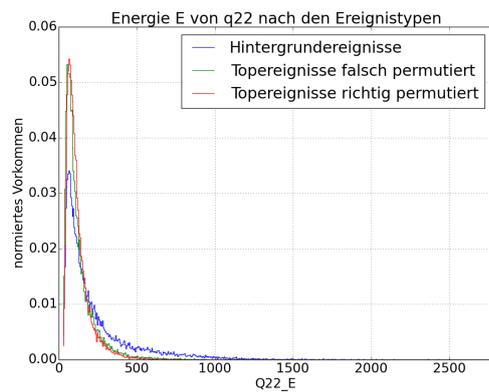
(43) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem x Impuls  $P_{x,q22}$  des  $q22$  Quarks.



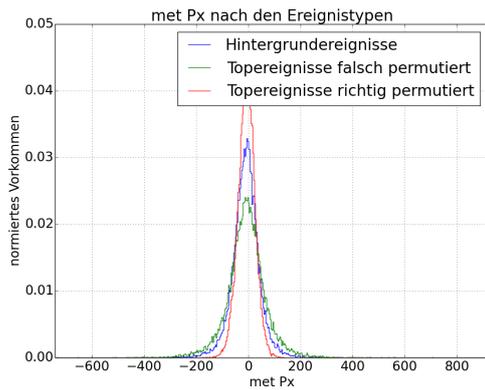
(44) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem y Impuls  $P_{y,q22}$  des  $q22$  Quarks.



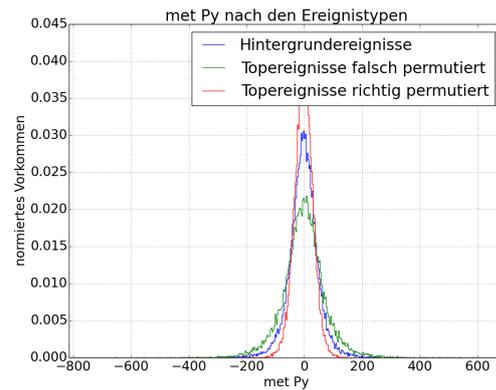
(45) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach dem z Impuls  $P_{z,q22}$  des  $q22$  Quarks.



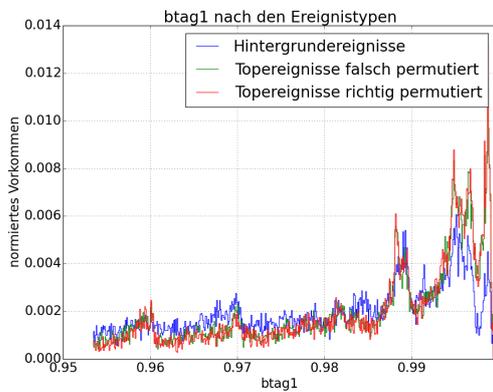
(46) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der Energie  $E_{q22}$  des  $q22$  Quarks.



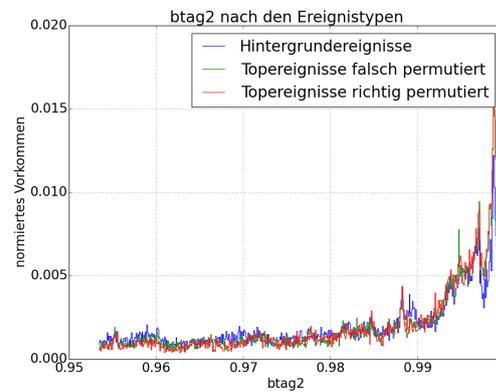
(47) Normiertes Vorkommen der Ereignistypen Hintergrundereignisse, richtig und falsch permutierte Topereignisse nach den fehlenden Transversalimpuls  $P_{x,met}$ .



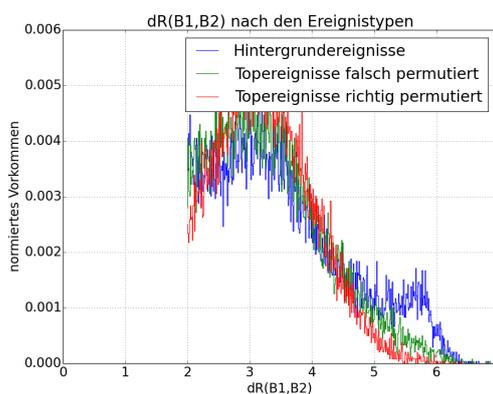
(48) Normiertes Vorkommen der Ereignistypen Hintergrundereignisse, richtig und falsch permutierte Topereignisse nach dem fehlenden Transversalimpuls  $P_{y,met}$ .



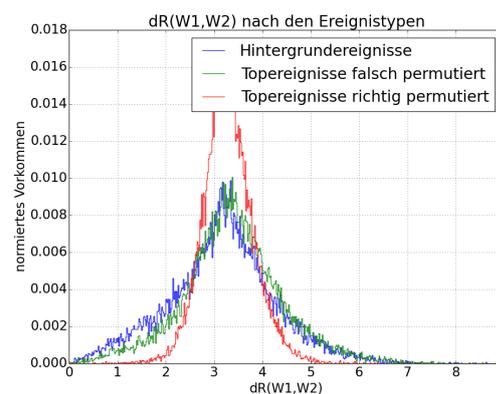
(49) Normiertes Vorkommen der Ereignistypen Hintergrundereignisse, richtig und falsch permutierte Topereignisse nach btag1.



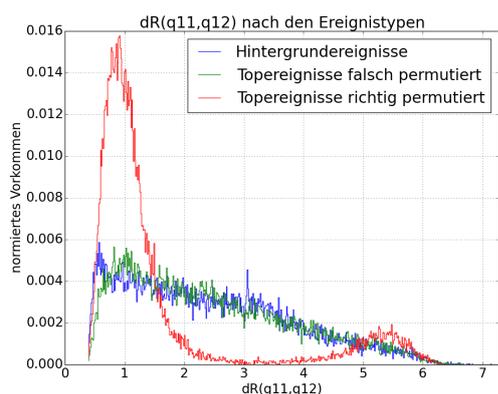
(50) Normiertes Vorkommen der Ereignistypen Hintergrundereignisse, richtig und falsch permutierte Topereignisse nach btag2.



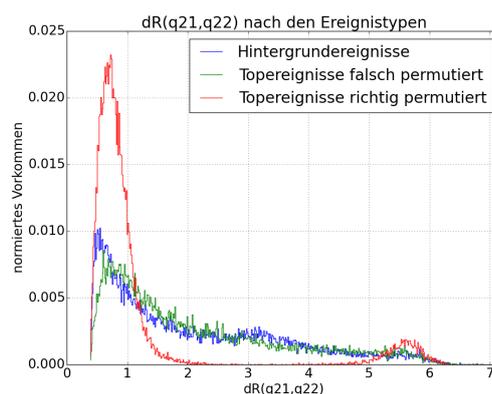
(51) Normiertes Vorkommen der Ereignistypen Hintergrundereignisse, richtig und falsch permutierte Topereignisse nach  $\Delta R$  zwischen  $B_1$  und  $B_2$ .



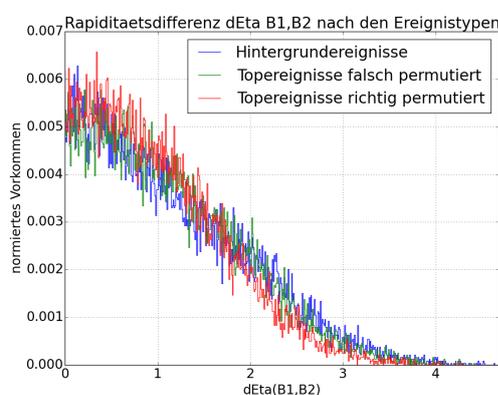
(52) Normiertes Vorkommen der Ereignistypen Hintergrundereignisse, richtig und falsch permutierte Topereignisse nach  $\Delta R$  zwischen  $W_1$  und  $W_2$ .



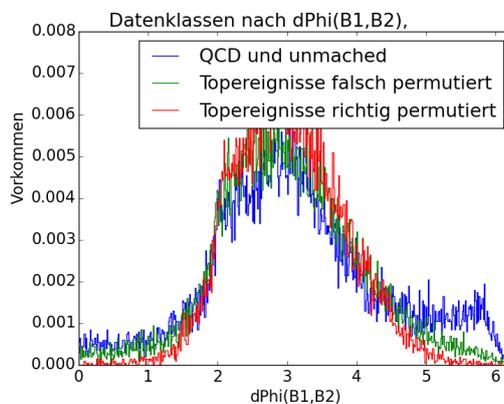
(53) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach  $\Delta R$  zwischen  $q_{11}$  und  $q_{12}$ .



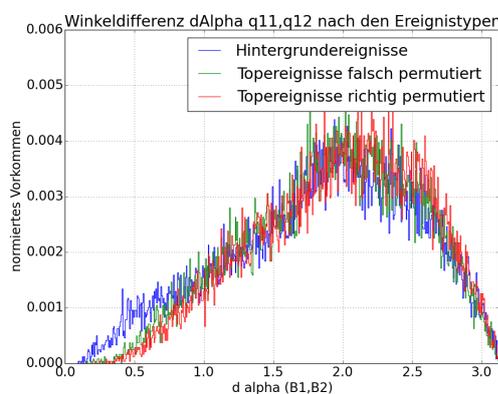
(54) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach  $\Delta R$  zwischen  $q_{21}$  und  $q_{22}$ .



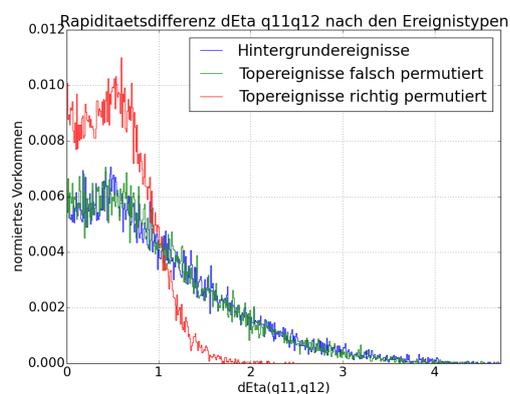
(55) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Rapiditätsdifferenz  $\Delta\eta$  zwischen  $B_1$  und  $B_2$ .



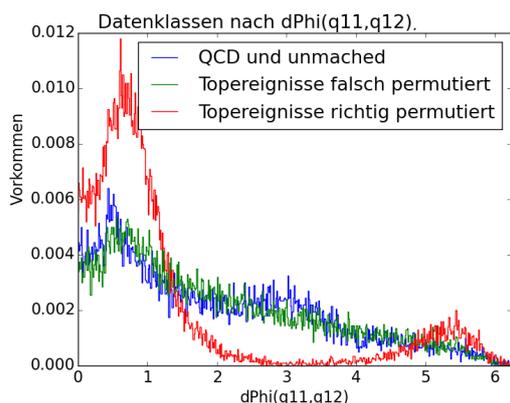
(56) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Azimutwinkeldifferenz  $\Delta\phi$  zwischen  $B_1$  und  $B_2$ .



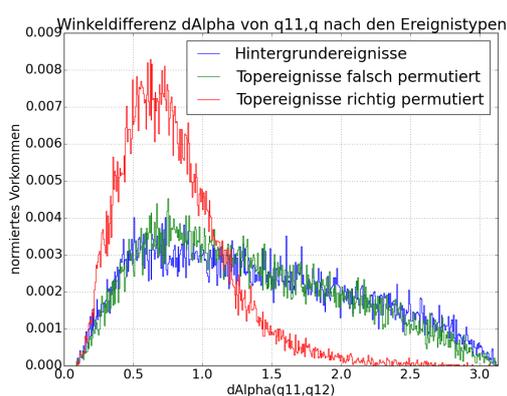
(57) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Raumwinkel  $\Delta\alpha$  zwischen  $B_1$  und  $B_2$ .



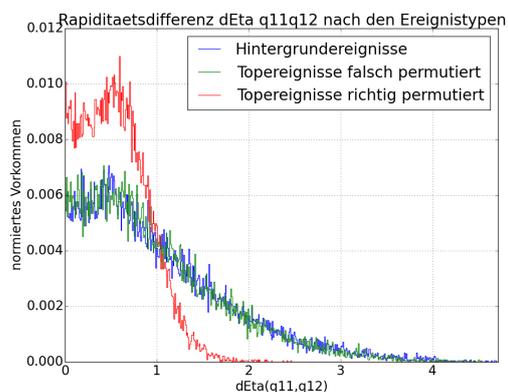
(58) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Rapiditätsdifferenz  $\Delta\eta$  zwischen  $q_{11}$  und  $q_{12}$ .



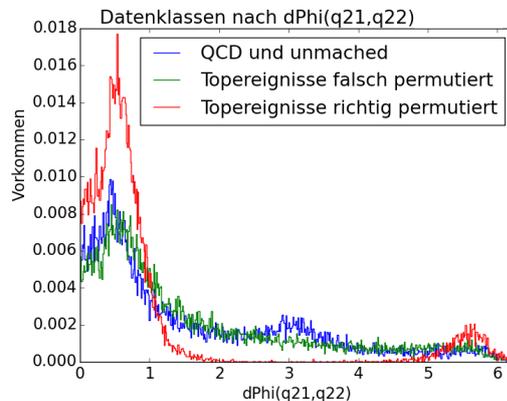
(59) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Azimutwinkeldifferenz  $\Delta\phi$  zwischen  $q_{11}$  und  $q_{12}$ .



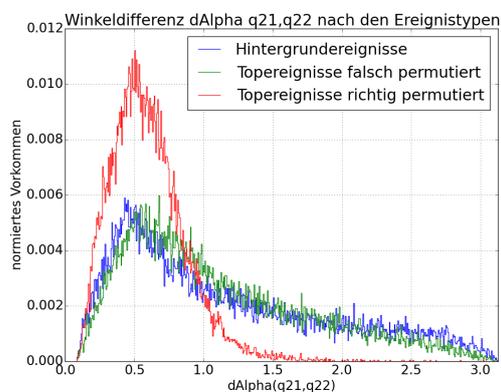
(60) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Raumwinkel  $\Delta\alpha$  zwischen  $q_{11}$  und  $q_{12}$ .



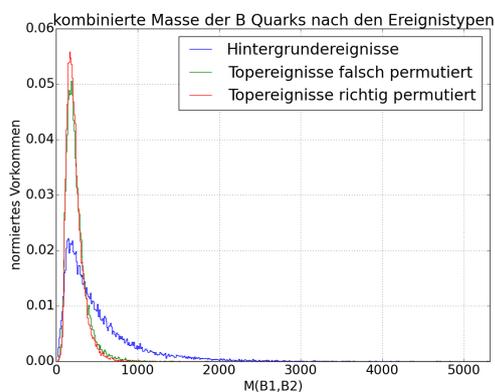
(61) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Rapiditätsdifferenz  $\Delta\eta$  zwischen  $q_{21}$  und  $q_{22}$ .



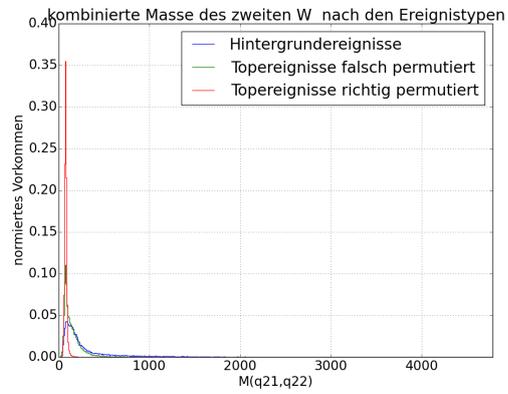
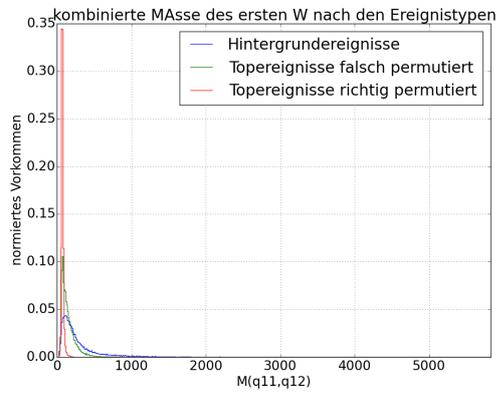
(62) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Azimutwinkeldifferenz  $\Delta\phi$  zwischen  $q_{21}$  und  $q_{22}$ .



(63) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach Raumwinkel  $\Delta\alpha$  zwischen  $q_{21}$  und  $q_{22}$ .



(64) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der kombinierten Masse von  $B_1$  und  $B_2$ .



(65) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der kombinierten Masse von  $q_{11}$  und  $q_{12}$ .

(66) Normiertes Vorkommen der Ereignistypen Hintergrund, richtig und falsch permutierte Topereignisse nach der kombinierten Masse von  $q_{21}$  und  $q_{21}$ .

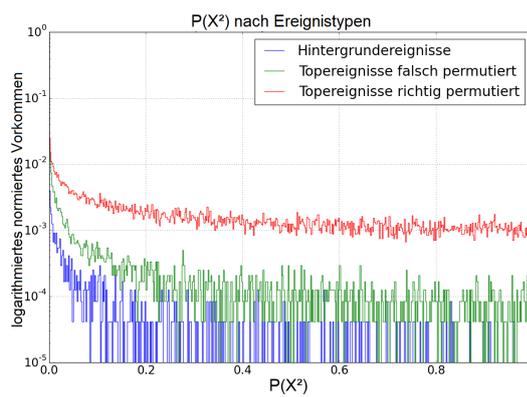


Abbildung A.1.: Normiertes Vorkommen von  $P(\chi^2)$  nach den einzelnen Klassen.

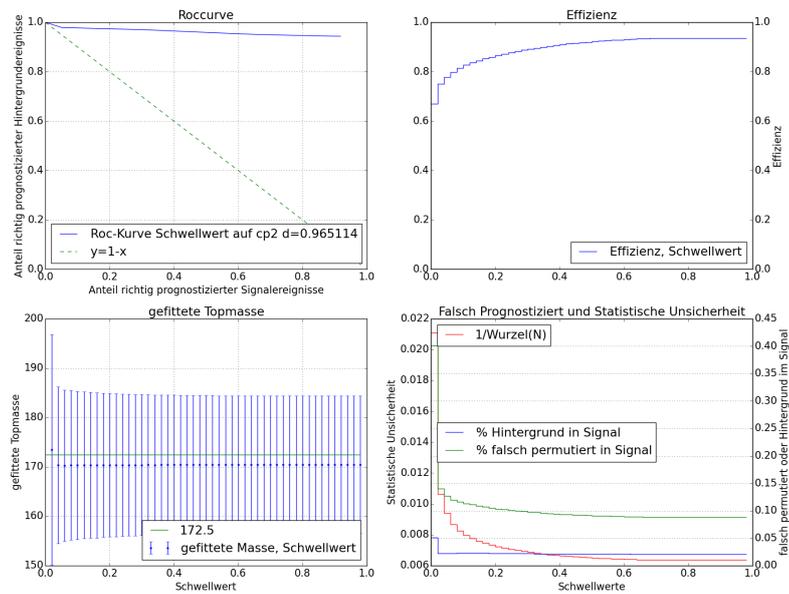


Abbildung A.2.: Untersuchung des geeignetsten Schwellwerts der zweiten Prognosekomponente anhand der Roc-Kurve, Effizienz, angepassten Topmasse und richtig und falsch prognostizierten Ereignissen, sowie statistischer Unsicherheit. Der Schwellwert wird auf 0.6 festgelegt.

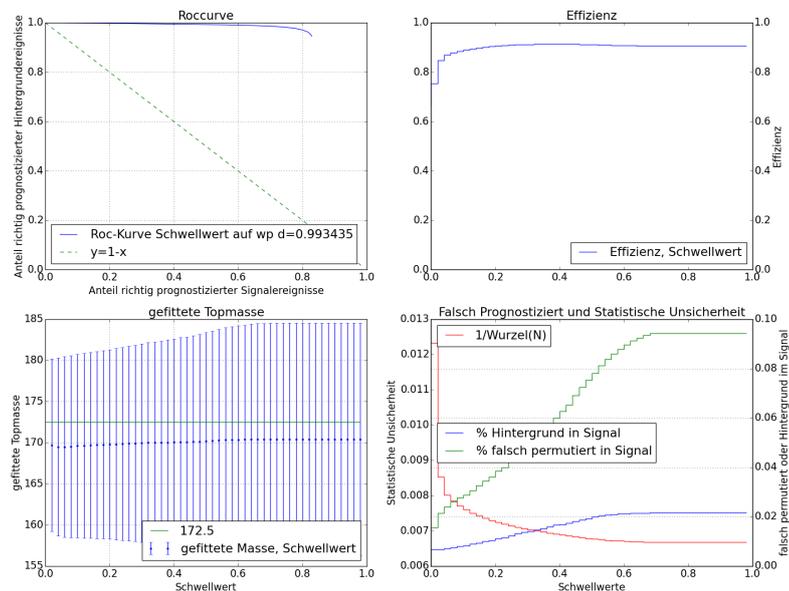


Abbildung A.3.: Untersuchung des geeignetsten Schwellwerts Prognosekomponente für falsch permutierte Topereignisse. Der Schwellwert wird auf 0.3 festgelegt.

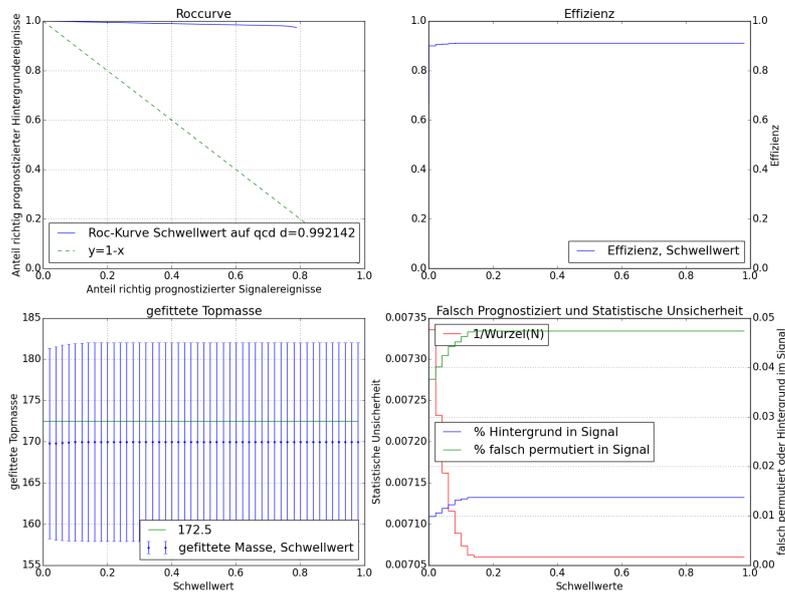
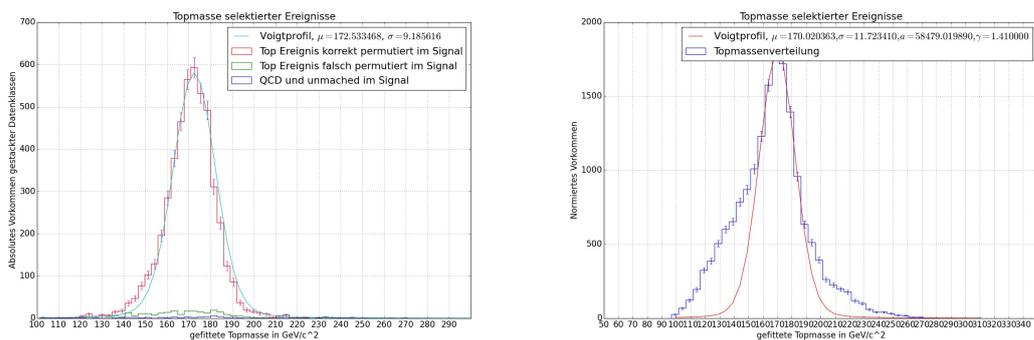
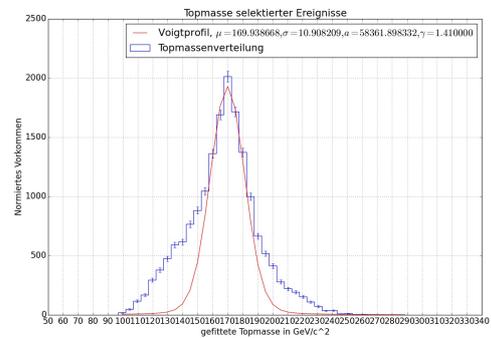
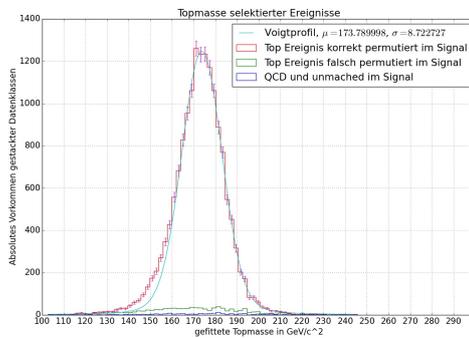


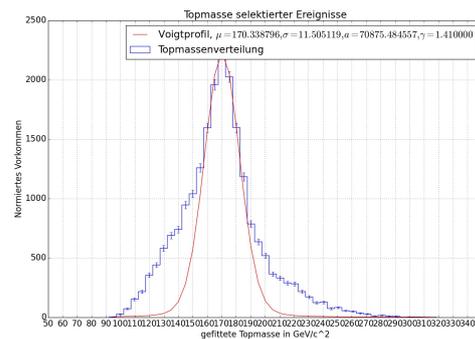
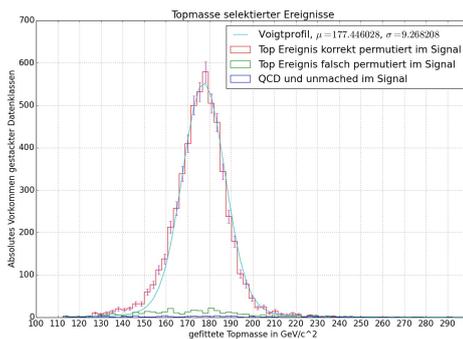
Abbildung A.4.: Untersuchung des geeignetsten Schwellwerts für die Prognosekomponente des Hintergrunds. Der Schwellwert wird auf 0.1 festgelegt.



(1) Massenspektrum mit angepasster Voigtfunktion. (2) Massenspektrum des selektierten CMS-Hier wurde der Trainingsdatensatz mit der Topmasse Datensatzes. Das selektierende Neuronale Netz  $173,5 GeV/c^2$  sowohl fürs Training als auch für das wurde mit dem simulierten Datensatz der Topmasse Massenspektrum genutzt.  $173,5 GeV/c^2$  trainiert.



(3) Massenspektrum mit angepasster Voigtfunktion. (4) Massenspektrum des selektierten CMS-Hier wurde der Trainingsdatensatz mit der Topmasse Datensatzes. Das selektierende Neuronale Netz  $175,5\text{GeV}/c^2$  sowohl fürs Training als auch für das wurde mit dem simulierten Datensatz der Topmasse Massenspektrum genutzt.  $175,5\text{GeV}/c^2$  trainiert.



(5) Massenspektrum mit angepasster Voigtfunktion. (6) Massenspektrum des selektierten CMS-Hier wurde der Trainingsdatensatz mit der Topmasse Datensatzes. Das selektierende Neuronale Netz  $178,5\text{GeV}/c^2$  sowohl fürs Training als auch für das wurde mit dem simulierten Datensatz der Topmasse Massenspektrum genutzt.  $178,5\text{GeV}/c^2$  trainiert.



# Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Fachbereiches einverstanden.

Hamburg, den \_\_\_\_\_ Unterschrift: \_\_\_\_\_