

# **Erkennung von hadronischen Tau-Leptonzerfällen mit dem verbesserten ATLAS-Kalorimetertrigger**

---

## **Bachelor-Arbeit**

zur Erlangung des akademischen Grades

## **Bachelor of Science**

vorgelegt von

**Christian Wiel**

**geboren am 18.7.1993 in Dresden**

der

Technischen Universität Dresden

Institut für Kern- und Teilchenphysik  
Fachrichtung Physik

Fakultät Mathematik und Naturwissenschaften

2016

Eingereicht am 22. Februar 2016

1. Gutachter: Prof. Dr. Arno Straessner
2. Gutachter: Prof. Dr. Michael Kobel

## Zusammenfassung

Im Rahmen des geplanten Phase-I-Upgrades im Jahr 2018 soll der Kalorimeter-Trigger des ATLAS-Detektors verbessert werden. Danach werden dem Trigger für seine Entscheidung die Daten des Flüssig-Argon-Kalorimeters mit höherer räumlicher Auflösung zur Verfügung stehen. Anhand der Ergebnisse einer Simulation dieser Daten werden in dieser Arbeit Schauerprofilvariablen auf ihre Fähigkeit hadronische Tau-Leptonzerfälle zu identifizieren getestet. Dafür werden insbesondere auch die Informationen des Tile-Kalorimeters einbezogen. Für die Berechnung der Schauerprofilvariablen kommen einfache Rechenoperationen zum Einsatz, die in der Trigger-Elektronik umgesetzt werden können.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
<b>2. Das ATLAS-Experiment</b>	<b>2</b>
2.1. Der Large Hadron Collider . . . . .	2
2.2. Der ATLAS-Detektor . . . . .	2
2.2.1. Das Flüssig-Argon-Kalorimeter . . . . .	3
2.2.2. Das Tile-Kalorimeter . . . . .	4
2.2.3. Phase-I-Upgrade . . . . .	5
2.3. Tau-Leptonen . . . . .	6
<b>3. Erkennung hadronischer Tauzerfälle</b>	<b>7</b>
3.1. Simulationsdatensätze . . . . .	7
3.1.1. Klassifizierung der Anti- $k_t(4)$ -truth-Jets . . . . .	8
3.2. Schauerprofilvariablen . . . . .	10
3.3. Methoden zur Trennung von Signal und Untergrund . . . . .	19
3.3.1. Einfacher Schnitt in einer Variable . . . . .	19
3.3.2. Likelihood-Methode . . . . .	19
3.3.3. Boosted Decision Trees . . . . .	20
3.4. Auswertung . . . . .	22
3.4.1. Anwendung einfacher Schnitte in einer Variable . . . . .	22
3.4.2. Anwendung von Boosted Decision Trees und Likelihood-Methode . . . . .	23
3.5. Zusammenfassung und Ausblick . . . . .	27
<b>A. Anhang</b>	<b>29</b>
<b>Abkürzungsverzeichnis</b>	<b>31</b>
<b>Abbildungsverzeichnis</b>	<b>32</b>
<b>Tabellenverzeichnis</b>	<b>33</b>
<b>Literaturverzeichnis</b>	<b>34</b>

# 1. Einleitung

Das Ziel des geplanten Ausbaus des Large Hadron Collider (LHC) im Jahr 2018 ist es, die instantane Luminosität auf  $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  zu verdoppeln, um auch extrem seltene physikalische Ereignisse nachweisen zu können. Dabei wird auch das Triggersystem des Kalorimeters des ATLAS-Detektors verbessert, indem die Signale des Flüssig-Argon-Kalorimeters mit höherer Energie- und Ortsauflösung ausgelesen werden. Aus Daten einer Simulation dieser Signale sollen in dieser Arbeit Schauerprofilvariablen berechnet werden, anhand derer hadronisch zerfallende Tau-Leptonen von QCD-Jets oder Elektronen unterschieden werden können. Zur Berechnung der Schauerprofilvariablen sollen einfache Rechenoperationen wie Addition und Multiplikation verwendet werden, da diese in die Trigger-Elektronik implementiert werden können.

In Kapitel 2 wird das ATLAS-Experiment kurz vorgestellt. Dabei wird auf den LHC, den ATLAS-Detektor, dessen Kalorimeter, das Phase-I-Upgrade des Kalorimeter-Triggers und Tau-Leptonen Bezug genommen.

In Kapitel 3 werden zunächst in Abschnitt 3.1 die Simulationsdatensätze und deren Erzeugung beschrieben. Anschließend werden in Abschnitt 3.2 die einzelnen Schauerprofilvariablen vorgestellt, deren Eignung durch hadronische Tauzerfälle verursachte Jets von anderen Untergrundjets zu unterscheiden in Abschnitt 3.4 untersucht wird. Dafür wird die Nachweiseffizienz eindimensionaler Schnitte (Abschnitt 3.4.1) und multivariater Methoden (Abschnitt 3.4.2) ermittelt.

## 2. Das ATLAS-Experiment

### 2.1. Der Large Hadron Collider

Der Large Hadron Collider (LHC) [1] ist ein Teilchenbeschleuniger der Europäischen Organisation für Kernforschung (CERN) in der Nähe von Genf. Er wurde im 26,7 km langen Tunnelring des Large Electron-Positron Collider (LEP) nach dessen Stilllegung im Jahr 2000 installiert. Der LHC dient zur Beschleunigung und Kollision von Protonen oder geladenen Ionen. Nachdem die Teilchen verschiedene Vorbeschleuniger durchlaufen haben, werden sie gegenläufig in die zwei separaten ringförmigen Strahlröhren des LHC gespeist und dort weiter beschleunigt. Dabei werden sie durch magnetische Felder von bis zu 8 T, welche von supraleitenden Niob-Titan-Magneten, die mit suprafluidem Helium auf unter 2 K gekühlt werden, auf ihrer Bahn gehalten [2].

Bei der Kollision an einem der vier Kollisionspunkte besitzen zwei aufeinander treffende Protonen eine Schwerpunktsenergie von bis zu  $\sqrt{s} = 14$  TeV. An den Kollisionspunkten befinden sich die vier großen Experimente ATLAS [3], CMS, ALICE und LHCb.

### 2.2. Der ATLAS-Detektor

In Abbildung 2.1 ist der Aufbau des ATLAS-Detektors dargestellt. Der ATLAS-Detektor ist spiegelsymmetrisch um die Ebene, die transversal zum Strahl im Kollisionspunkt liegt, aufgebaut und besteht aus mehreren um die Strahlachse angeordneten aufeinanderfolgenden zylindrischen Schichten.

Dem Kollisionspunkt am nächsten liegt der innere Detektor (ID). Mit diesem kann unter anderem der Impuls geladener Teilchen aus ihrer Ladung und dem Krümmungsradius ihrer Spur im Magnetfeld von 2 T, welches den ID durchdringt, bestimmt werden.

Die nächste Detektorschicht bildet das Kalorimeter, welches sich in das Flüssig-Argon-(LAr)-Kalorimeter<sup>1</sup> [4] und das Tile-Kalorimeter unterteilt. Es dient der Bestimmung der Energie von Teilchen. Der Aufbau und die Funktionsweise des LAr- und Tile-Kalorimeters wird in Abschnitt 2.2.1 und 2.2.2 genauer beschrieben.

Die Kalorimeter sind vom Myonenspektrometer umgeben, mit dem sich die Impulse

---

<sup>1</sup>Englisch: Liquid Argon

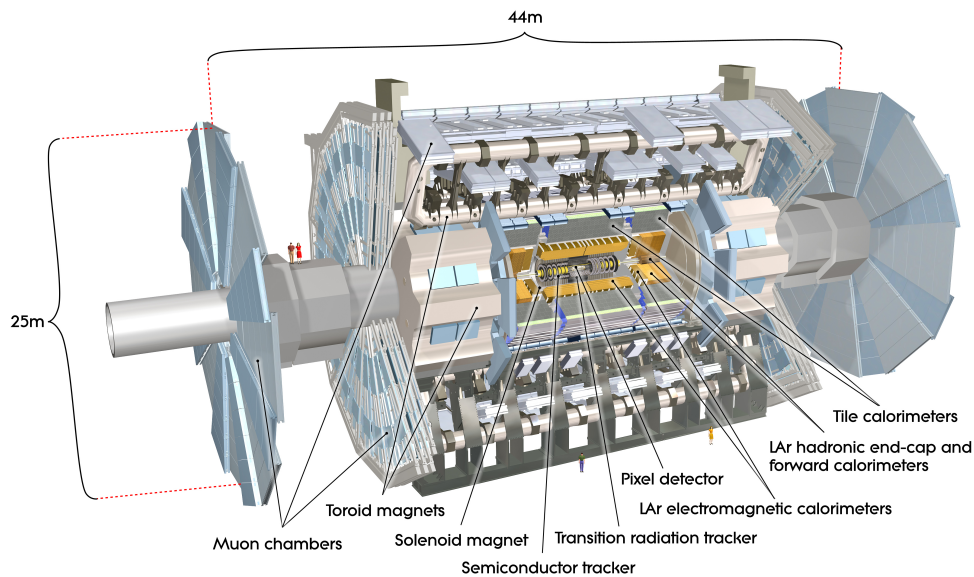


Abbildung 2.1.: Aufbau des ATLAS-Detektors [5].

von Myonen bestimmen lassen.

Zur räumlichen Beschreibung werden die Koordinaten Pseudorapidität  $\eta$  und Azimutalwinkel  $\phi$  eingeführt. Wird ein kartesisches Koordinatensystem definiert, wobei der Koordinatenursprung im Kollisionspunkt liegt und die z-Achse in Strahlrichtung, die x-Achse in Richtung des Mittelpunkts des LHCs und die y-Achse nach oben zeigt, ergibt sich  $\phi$  als Winkel gemessen von der x-Achse und  $\eta$  nach Formel 2.1, wobei  $\theta$  der Polarwinkel gemessen von der z-Achse ist.

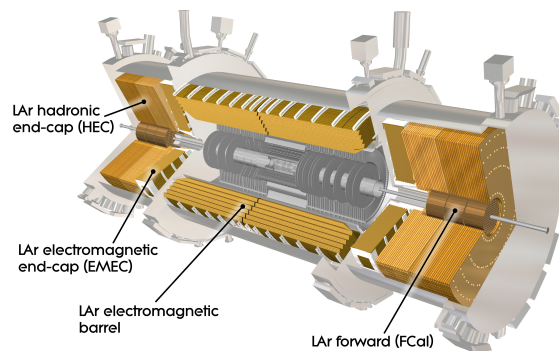
$$\eta = -\ln \tan(\theta/2) \quad (2.1)$$

Pseudorapidität  $\eta = 0$  beschreibt so zum Beispiel eine Ebene transversal zum Strahl im Kollisionspunkt ( $\theta = \pi/2$ ). Auf diesen Koordinaten wird außerdem der folgende Abstand  $\Delta R$  definiert:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (2.2)$$

### 2.2.1. Das Flüssig-Argon-Kalorimeter

Das LAr-Kalorimeter (siehe Abbildung 2.2) erfasst den Bereich  $|\eta| < 4,9$ . Es besteht aus dem elektromagnetischen Barrel-Kalorimeter (EMB) ( $|\eta| < 1,475$ ) im Zentrum und je einem elektromagnetischen ( $1,375 < |\eta| < 3,2$ ) und hadronischen ( $1,5 < |\eta| < 3,2$ )



**Abbildung 2.2.:** Aufbau des LAr-Kalorimeters [6].

Endkappenkalorimeter in Vorwärts- und Rückwärtsrichtung. Die Vorwärtskalorimeter ( $3,1 < |\eta| < 4,9$ ) befinden sich zwischen Strahl und Endkappenkalorimeter.

Es handelt sich bei den Kalorimetern um Sampling-Kalorimeter. Das bedeutet, dass sie abwechselnd aus Absorber- und Nachweisschicht aufgebaut sind (siehe Abbildung 2.3). Die hochenergetischen Teilchen wechselwirken mit der Absorberschicht und verursachen Schauer von Sekundärteilchen, die das aktive Medium in der Nachweisschicht ionisieren. Die Ionen und Elektronen driften zu den Elektroden, welche ein elektrisches Feld im aktiven Medium erzeugen, und verursachen einen Strom, aus welchem die Energie des Teilchens rekonstruiert wird.

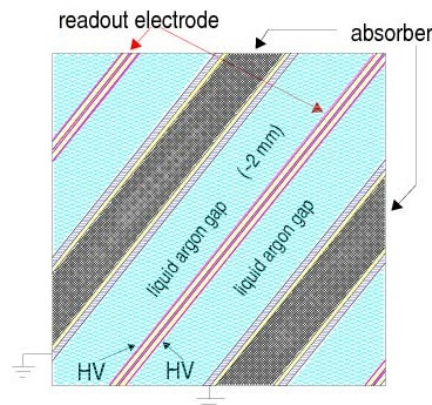
Als aktives Medium kommt flüssiges Argon zum Einsatz. Das Absorbermaterial und die Anordnung der Schichten ist für die Kalorimeter unterschiedlich. Beim EMB wird dafür beispielsweise Blei verwendet.

Das EMB und das elektromagnetische Endkappenkalorimeter bestehen aus vier Schichten, die von innen nach außen wie folgt benannt werden: Presampler, Front Layer, Middle Layer und Back Layer.

### 2.2.2. Das Tile-Kalorimeter

Das Tile-Kalorimeter liegt außerhalb des LAr-Kalorimeters im Bereich  $|\eta| < 1,7$ . Es handelt sich bei diesem ebenfalls um ein Samplingkalorimeter, wobei Stahl das Absorbermedium ist und Szintillator-Platten das aktive Medium sind. Es unterteilt sich in ein zentrales ( $|\eta| < 1,0$ ) und zwei erweiterte Barrel-Kalorimeter ( $1,0 < |\eta| < 1,7$ ). Die Energien werden für den Trigger im Bereich einzelner sogenannter Tile Tower (TT) aufsummiert. Ein TT überdeckt einen  $\Delta\eta \times \Delta\phi$ -Bereich von  $0,1 \times 0,1$ .





**Abbildung 2.3.:** Aufbau der Schichten der Kalorimeter in der Vergrößerung [7].

### 2.2.3. Phase-I-Upgrade

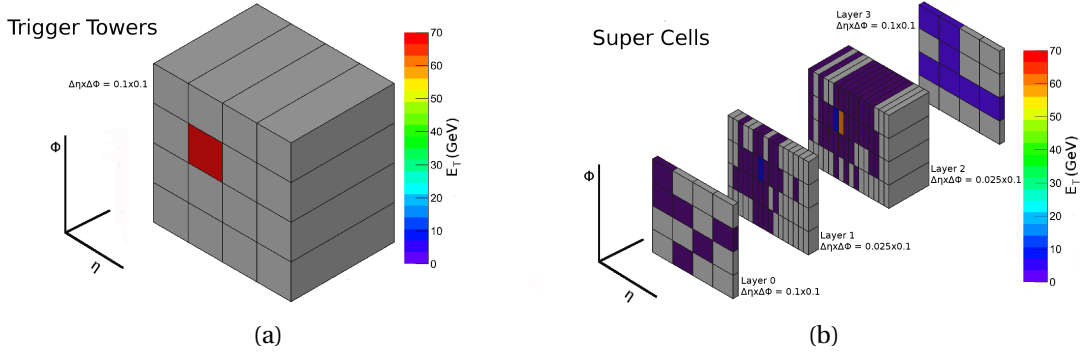
Während des aktuellen Laufs des LHC von 2015-2017 kommt es mit einem zeitlichen Abstand von 25 ns zu Bunchcrossings<sup>2</sup> (BCs) bei denen Bündel von bis zu  $10^{11}$  Protonen aufeinandertreffen. Diese hohe Frequenz ist wichtig, um auch für seltene Ereignisse eine ausreichende Statistik zu erhalten.

Neben exzellenter Auflösung von Energie und Spur der Teilchen ist es deshalb erforderlich eine schnelle Auswahl physikalisch interessanter Ereignisse zu treffen, insbesondere da einige Ereignisse um Größenordnungen seltener vorkommen als andere. Dies ist die Aufgabe des Triggers. In seiner aktuellen Form verwendet dieser nur die Informationen sogenannter Trigger Tower, welche die Energien aller Zellen in einem  $\eta \times \phi$ -Bereich von  $0,1 \times 0,1$  aufsummieren.

Ein Ziel des Phase-I-Upgrades [8] des LAr-Kalorimeters ist es, Informationen mit höherer Granularität und Auflösung sowie longitudinale Schauerinformation für den Level-1 Trigger zugänglich zu machen, indem anstelle der Trigger Tower Superzellen (SCs) verwendet werden. Damit soll dessen Effizienz zum Auslösen auf Elektronen, Photonen und Tau-Leptonen erhöht und gleichzeitig die Untergrundunterdrückung verbessert werden. Dies ist wichtig, da die instantane Luminosität<sup>3</sup> mit dem geplanten Upgrade 2018 auf  $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  verdoppelt werden soll. Ein Vergleich der Granularität der Trigger Tower und SCs ist in Abbildung 2.4 dargestellt. Die SCs besitzen im Front- und Middle Layer eine Ausdehnung von  $\Delta\eta \times \Delta\phi = 0,025 \times 0,1$  und im Presampler und Back Layer eine Ausdehnung von  $\Delta\eta \times \Delta\phi = 0,1 \times 0,1$ .

<sup>2</sup>Deutsch: Strahlkreuzung

<sup>3</sup>Maß für die Häufigkeit von Ereignissen im Verhältnis zum Wirkungsquerschnitt



**Abbildung 2.4.:** Informationen über Energiedeposition eines Elektrons mit 70 GeV transversaler Energie, die dem Trigger nach aktuellem Stand (a) und nach geplanter Phase-I-Upgrade (b) bereitstehen [8].

## 2.3. Tau-Leptonen

Tau-Leptonen sind Elektronen ähnliche Elementarteilchen. Sie besitzen die gleiche Ladung wie diese, allerdings eine fast 3500 mal größere Masse von  $(1776,82 \pm 0,16)$  MeV [9]. Wegen ihrer kurzen Lebenszeit von  $(290,6 \pm 1,0) \cdot 10^{-15}$  s [9] werden sie durch ihre Zerfallsprodukte nachgewiesen. Tau-Leptonen können aufgrund ihrer hohen Masse sowohl leptonisch (zu ca. 35%) als auch hadronisch (zu ca. 65%) zerfallen. Der hadronische Zerfall findet überwiegend zu Pionen statt. Die Häufigkeiten der wichtigsten Zerfallskanäle sind in Tabelle 2.1 zusammengefasst.

Die Erkennung von Tau-Leptonen in den Ereignissen am ATLAS-Detektor ist wichtig, um z.B. Higgs-Bosonen, die über  $H \rightarrow \tau^- \tau^+$  zerfallen, nachzuweisen. Außerdem werden von Theorien, die das Standardmodell erweitern, weitere Teilchen vorhergesagt, die in Tau-Leptonen zerfallen. Zum Beispiel werden für das MSSM (minimales supersymmetrisches Standardmodell) noch weitere Higgs-Bosonen erwartet [10].

**Tabelle 2.1.:** Häufigkeit einiger ausgewählter Zerfallskanäle des Tau-Leptons (für  $\tau^+$  gelten die ladungskonjugierten Zerfallskanäle) [9].

<b>leptonische Zerfälle</b>	$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$	$(17,41 \pm 0,04)\%$
	$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$	$(17,83 \pm 0,04)\%$
<b>hadronische Zerfälle</b>	$\tau^- \rightarrow \pi^- \pi^0 \nu_\tau$	$(25,52 \pm 0,09)\%$
	$\tau^- \rightarrow \pi^- \nu_\tau$	$(10,83 \pm 0,06)\%$
	$\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$	$(9,80 \pm 0,07)\%$
	$\tau^- \rightarrow \pi^- 2\pi^0 \nu_\tau$	$(9,30 \pm 0,11)\%$
	$\tau^- \rightarrow \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	$(4,62 \pm 0,06)\%$

## 3. Erkennung hadronischer Tauzerfälle

Das Ziel dieses Kapitels ist es, Tau-Jets, welche aus den hadronischen Zerfallsprodukten des Tau-Leptons bestehen, von anderen hadronischen Untergrundjets oder Elektronen zu unterscheiden. Der hadronische Untergrund, der bei den Proton-Proton-Kollisionen entsteht, wird hauptsächlich durch das Hadronisieren von Quarks aufgrund von Effekten der Quantenchromodynamik (QCD) verursacht. Aus diesem Grund werden diese Jets im Folgenden auch als QCD-Jets bezeichnet. Im Gegensatz zu Tau-Jets bestehen QCD-Jets meist aus einer größeren Anzahl von Teilchen und verursachen einen breiteren Schauer im Kalorimeter.

### 3.1. Simulationsdatensätze

Um mögliche Variablen zu finden, anhand derer es dem Trigger möglich wäre hadronische Tau-Zerfälle vom sonstigen Jetuntergrund zu unterscheiden, wurden Simulationsdatensätze, welche die jeweiligen Ereignisse enthalten, für die Arbeit bereitgestellt.

Die Dateien bestehen aus Informationen über die enthaltenen Teilchen, deren Energie und Flugrichtung sowie deren Zerfallsprodukte und in welcher SC des EMB oder welchem TT des Tile-Kalorimeter welche Energie deponiert wurde. Zur deponierten Energie jeder Zelle ist auch die rekonstruierte Energie enthalten. Weiterhin ist die zugehörige ID des BCs beinhaltet.

Die deponierten Energien wurden mittels des Anti- $k_t$ -Algorithmus [11] mit Parameter  $R = 0,4$  zu Jets zusammengefasst, deren transversaler Impuls  $p_T$  und Richtung  $(\eta, \phi)$  in den Daten gespeichert ist. Diese Jets werden im Folgenden als Anti- $k_t(4)$ -truth-Jets bezeichnet.

Die rekonstruierte Energie ist für die SCs im EMB und TTs im Tile-Kalorimeter im Bereich  $|\eta| < 1,3$  vorhanden.

Die Datensätze werden in verschiedenen Schritten produziert. Ein Monte Carlo Eventgenerator erzeugt die Teilcheninformation. Anschließend wird mittels GEANT4 [12] und einem detaillierten Modell des ATLAS-Detektors die in ihm deponierte Energie simuliert.

Aus dieser ermittelt AREUS<sup>1</sup> [13] eine rekonstruierte Energie für jede SC. Diese Information soll dem Trigger nach dem Phase-I-Upgrade des LAr-Kalorimeters bereitstehen.

Zur Untersuchung der Tau-Leptonen wurden zwei Dateien bereitgestellt. Eine enthält  $Z \rightarrow \tau\tau$ -Ereignisse<sup>2</sup> bei 8 TeV Schwerpunktsenergie der Protonen, die andere<sup>3</sup>  $H \rightarrow \tau\tau$  bei 14 TeV. Diese Wahl wurde getroffen, um eine weite Energieverteilung der Taus zu erhalten.

Für die Analyse des QCD-Jetuntergrundes wurde eine Datei<sup>4</sup> bestehend aus Dijets genutzt, die bereits in [13] verwendet wurde. Für diese wurden Proton-Proton-Kollisionen ausgewählt, bei denen der resultierende Jet mit dem höchsten transversalen Impuls im Bereich  $20 \text{ GeV} < p_T < 80 \text{ GeV}$  liegt.

Zum Vergleich mit Elektronen wird eine weitere Datei<sup>5</sup> hinzugezogen, für die als Event-generator lediglich eine Partikelkanone verwendet wurde, welche Elektronen mit einer uniformen  $p_T$ -Verteilung im Bereich  $20 \text{ GeV} < p_T < 50 \text{ GeV}$  erzeugt. Die  $p_T$ -Verteilung der Anti- $k_t(4)$ -truth-Jets der Tau- und Dijet-Dateien ist in Abbildung 3.1 dargestellt.

### 3.1.1. Klassifizierung der Anti- $k_t(4)$ -truth-Jets

In Abschnitt 3.4 sollen Tau-Jets von Untergrundjets unterschieden werden. Dafür müssen zunächst die jeweiligen Anti- $k_t(4)$ -truth-Jets der Dateien ausgewählt werden.

Aus der Dijet und der Elektronen-Datei werden alle Anti- $k_t(4)$ -truth-Jets mit  $p_T > 20 \text{ GeV}$  gewählt.

Für die Jets der Tau-Datei wird neben der Energieauswahl von  $p_T > 20 \text{ GeV}$  noch festgestellt, ob der Jet durch ein hadronisch zerfallendes Tau-Lepton verursacht wurde. Dazu werden alle Taus auf ihre Zerfallsprodukte untersucht. Tritt dabei kein Lepton auf, werden im Radius von  $\Delta R < 0,2$  um die Koordinaten des Taus alle Anti- $k_t(4)$ -truth-Jets als von hadronischen Tauzerfällen verursacht klassifiziert. Die Verteilung von  $\Delta R$  ist in Abbildung 3.2 (a) dargestellt.

Um eine Orientierung über die Streuung der Zerfallsprodukte zu erhalten, wurde die Verteilung des Abstands  $\Delta R_{\pi,\tau}$  zwischen Pion und verursachendem Tau in Abbildung 3.2 (b) dargestellt.

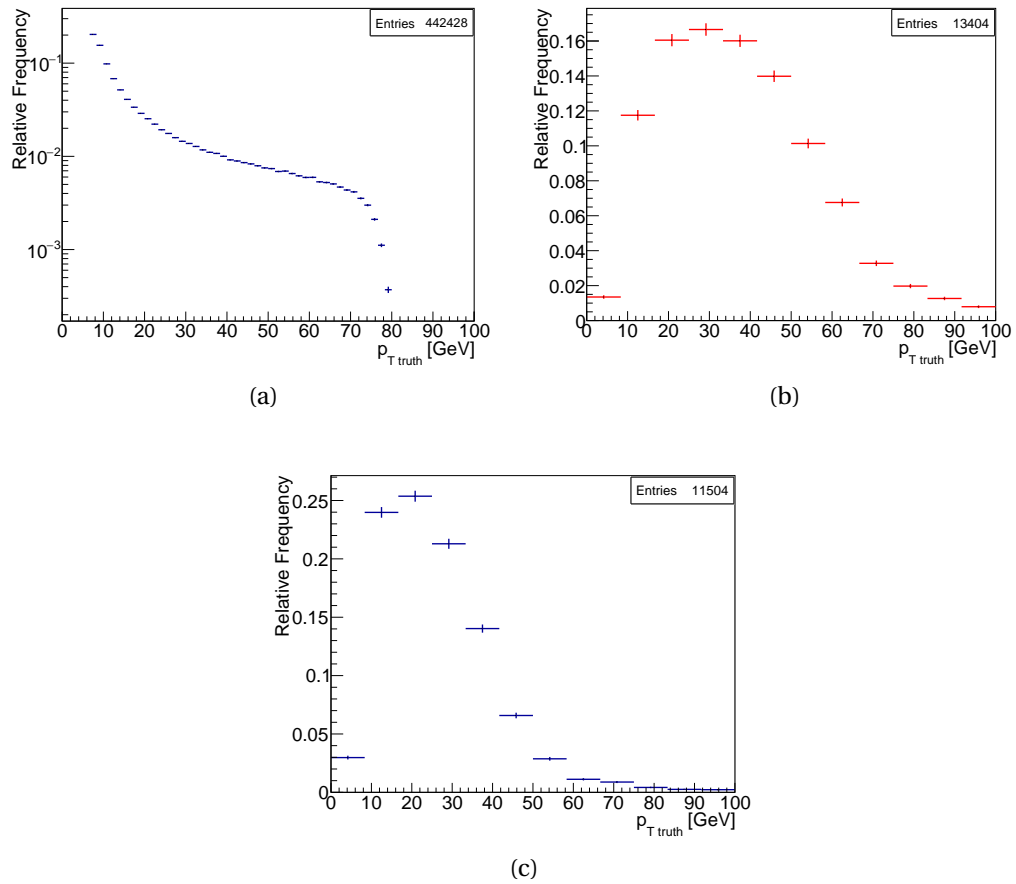
<sup>1</sup>ATLAS Readout Electronics Upgrade Simulation

<sup>2</sup>user.olga.mc12\_8TeV.147772.Sherpa\_CT10\_Ztautau.merge.HITS.e1434\_s1499\_s1504\_01328035

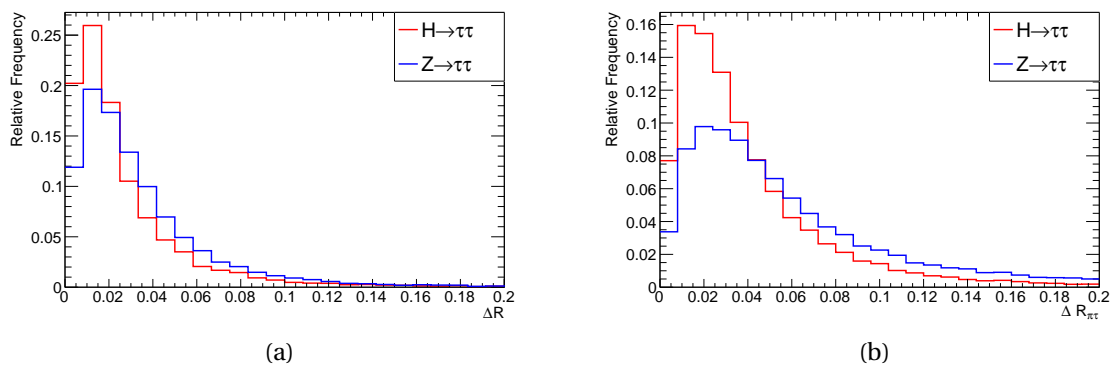
<sup>3</sup>user.olga.mc12\_14TeV.161577.PowHegPythia8\_AU2CT10\_ggH125\_tautauhh.merge.HITS.e1904\_s1682\_s1691\_01304813

<sup>4</sup>user.jgrohs.mc12\_14TeV.147911.Pythia8\_AU2CT10\_jetjet\_JZ1W.merge.AREUS.e1996\_s1729\_s1720.Mar19.13h

<sup>5</sup>user.jgrohs.SingleElectrons\_EMB\_20to50GeV.AREUS.Nov27.14h



**Abbildung 3.1.:** Verteilung der transversalen Impulse  $p_{T, \text{truth}}$  der Anti- $k_t(4)$ -truth-Jets für (a) Dijets (b)  $H \rightarrow \tau\tau$  und (c)  $Z \rightarrow \tau\tau$ .



**Abbildung 3.2.:** Verteilung der Abstände zwischen hadronisch zerfallendem Tau-Lepton und (a) Anti- $k_t(4)$ -truth-Jet, (b) Pionen aus Zerfall.

## 3.2. Schauerprofilvariablen

Es werden sowohl für QCD-Jets als auch für hadronisch zerfallende Tau-Leptonen Variablen berechnet, die deren Schauer im Kalorimeter beschreiben. Anhand dieser Schauerprofilvariablen soll in Abschnitt 3.4 eine Unterscheidung zwischen Taus und QCD-Jets stattfinden. Da es das Ziel ist Taus zu erkennen und dabei möglichst viele QCD-Jets zu verwerfen, werden Taus als Signal und QCD-Jets als Untergrund bezeichnet.

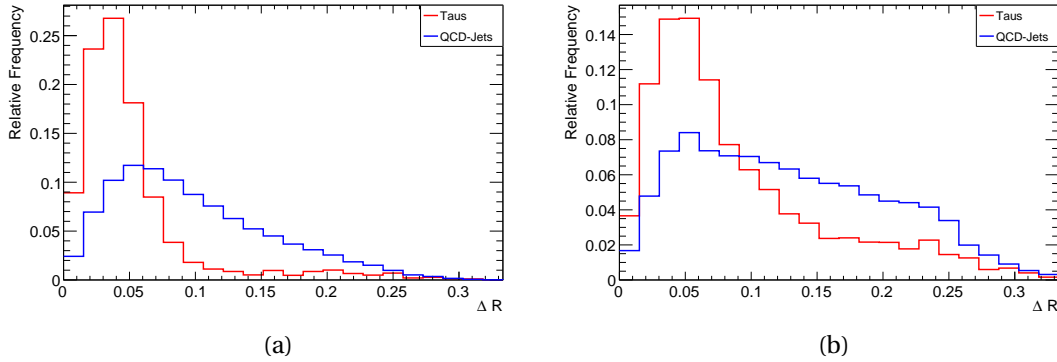
Als Ausgangspunkt der Berechnung werden die in Abschnitt 3.1.1 klassifizierten Anti- $k_t(4)$ -truth-Jets der Tau- und Dijetdateien verwendet. In einem  $\Delta\eta \times \Delta\phi$ -Bereich von  $0,4 \times 0,4$  um diesen Ausgangspunkt wird im Middle-Layer des EMB die SC und im Tile-Kalorimeter der TT mit maximaler Energiedeposition gesucht. Überschreitet deren Energie die Schwelle von 1 GeV und sind sie das lokale Maximum der sie umgebenden SCs bzw. TTs, werden die im nachfolgendem beschriebenen Schauerprofilvariablen für sie berechnet. Die Bezeichnung SC bzw. TT mit maximaler Energiedeposition impliziert im Folgenden immer, dass die genannten Bedingungen erfüllt sind.

Der Abstand  $\Delta R$  zwischen Anti- $k_t(4)$ -truth-Jets und SC mit maximaler Energiedeposition für QCD- und Tau-Jets ist in Abbildung 3.3 dargestellt. Insgesamt wurde für 7843 der Tau-Jets, 55862 der QCD-Jets und 98008 der Elektronen eine SC und oder ein TT mit maximaler Energiedeposition und  $E_{SC(TT)} > 1\text{GeV}$  gefunden. Die relativen Häufigkeiten für verschiedene Kombinationen von diesen sind in Tabelle 3.1 zusammengefasst. Das Maximum der Verteilungen des Abstands zwischen Anti- $k_t(4)$ -truth-Jet zu SC bzw. TT mit maximaler Energiedeposition liegt für Taus bei geringeren Werten.

Aufgrund der unterschiedlichen Granularität des Tile-Kalorimeters und des elektromagnetischen Kalorimeters, werden unterschiedliche Variablen für die SC und den TT mit maximaler Energiedeposition definiert.

**Tabelle 3.1.:** Relative Häufigkeit mit der jeweilige Kombination aus SC und TT mit maximaler Energieposition und  $E_{SC(TT)} > 1\text{GeV}$  auftritt.

	Tau-Jets	QCD-Jets
Nur Tile-Kalorimeter	37%	13%
Nur Middle Layer des EMB	20%	26%
Middle Layer und Tile-Kalorimeter ( $\Delta R \leq 0.1$ )	23%	19%
Middle Layer und Tile-Kalorimeter ( $\Delta R > 0.1$ )	19%	40%
Nur Presampler, Front- oder Backlayer des EMB	1%	3%



**Abbildung 3.3.:** Verteilung des Abstands  $\Delta R$  zwischen Koordinaten des Anti- $k_t(4)$ -truth-Jets und SC mit maximaler Energiedeposition im Middle Layer des EMB (a) bzw. TT mit maximaler Energiedeposition im Tile-Kalorimeter (b).

### Schauerprofilvariable $r_{EM}$

Der energiegewichtete Radius  $r_{EM}$  beschreibt wie stark der Schauer im Middle Layer des EMB kollimiert ist.

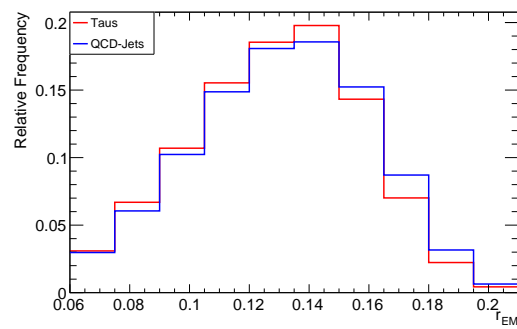
Er wurde auch in [14] verwendet und berechnet sich nach Formel 3.1, wobei die Summe über alle SCs im Radius  $\Delta R < 0,3$  um die SC mit maximaler Energiedeposition läuft:

$$r_{EM} = \frac{\sum_i E_i \cdot \Delta R_i}{\sum_i E_i}. \quad (3.1)$$

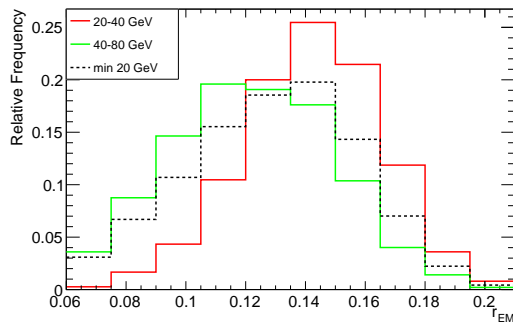
Die Verteilungen der relativen Häufigkeiten der Variable sind in Abbildung 3.4 dargestellt. Der Unterschied zwischen den Verteilungen von QCD-Jets und Taus ist gering. Die Histogramme für Jets oder Taus für unterschiedliche Bereiche der Impulse der Anti- $k_t(4)$ -truth-Jets (Abb. 3.4 (b),(c)) weisen deutlichere Unterschiede auf, allerdings entwickeln sich die Verteilungen für QCD-Jets und Taus sehr ähnlich, sodass trotzdem keine große Abweichung zwischen diesen erkennbar ist.

Die Verteilungen von  $r_{EM}$  haben sich in [14] für QCD-Jets und Taus stärker unterschieden. Das Maximum der Verteilung für QCD-Jets lag bei etwas höheren und das der Verteilung der Taus bei etwas niedrigeren Werten als hier.

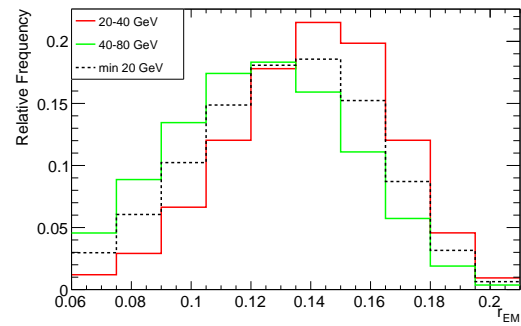
Der Grund dafür sind Unterschiede in den gewählten Simulationsdatensätzen. Zum einen besitzen diese unterschiedliche Energieverteilungen. Während in [14] die Mittelwerte der  $p_T$ -Verteilungen der Anti- $k_t(4)$ -truth-Jets bei über 60 GeV lagen, liegen sie hier im Bereich  $(42 \pm 1)$  GeV. Auch für den Energiebereich  $40 \text{ GeV} < p_T < 80 \text{ GeV}$  liegen die Mittelwerte hier nur im Bereich  $(54 \pm 2)$  GeV. Zum anderen wurden in [14] als Untergrund die hadronischen Jets von Topquark-Antitopquark-Paar-Produktionen gewählt. Weiterhin wurde mit deponierten Energien und nicht wie hier mit rekonstruierten Energien gearbeitet.



(a)



(b)



(c)

**Abbildung 3.4.:** Verteilung der relativen Häufigkeiten von  $r_{EM}$  für alle Energien für QCD-Jets und Taus (a) und verschiedene Energiebereiche für Taus (b) und QCD-Jets (c).



### Schauerprofilvariable $f_{\text{core}}$

Eine weitere als  $f_{\text{core}}$  [8] bezeichnete Variable beschreibt, ähnlich wie  $r_{\text{EM}}$ , die Form des Schauers im Middle Layer des EMB. Sie ist das Verhältnis der Energiesumme eines kleinen  $\Delta\eta \times \Delta\phi$  -Bereichs zu der eines großen Bereichs um die SC mit maximaler Energiedeposition:

$$f_{\text{core}} = \frac{E_{\Delta\eta \times \Delta\phi=0,175 \times 0,1}}{E_{\Delta\eta \times \Delta\phi=0,275 \times 0,3}}. \quad (3.2)$$

Die Verteilungen für diese Variable sind in Abbildung 3.5 dargestellt. Der Wert von  $f_{\text{core}}$  ist für Taus im Durchschnitt höher als für QCD-Jets, das bedeutet, dass die Energie bei Tau-Jets stärker um die SC mit maximaler Energiedeposition kollimiert ist. Dies entspricht den Erwartungen, da Taus nur in geringe Anzahlen von Tochterteilchen zerfallen und die Energie somit auf einen kleineren Bereich verteilt ist als bei QCD-Jets.

Weiterhin ist in Abbildung 3.5 (c) und (d) zu sehen, dass der Unterschied der Verteilungen für QCD-Jets und Taus für höhere Energien sogar wächst. Die Energie der Anti- $k_t(4)$ -truth-Jets steht dem Trigger jedoch nicht zur Verfügung, da diese zum einen nicht auf den von AREUS rekonstruierten sondern den deponierten Energien beruht und zum anderen Jet-Cluster-Algorithmen wie der Anti- $k_t$  Algorithmus aufwändig sind und deshalb auf Trigger-Level nicht zum Einsatz kommen.

### Schauerprofilvariable $\rho_{\text{oi}}$

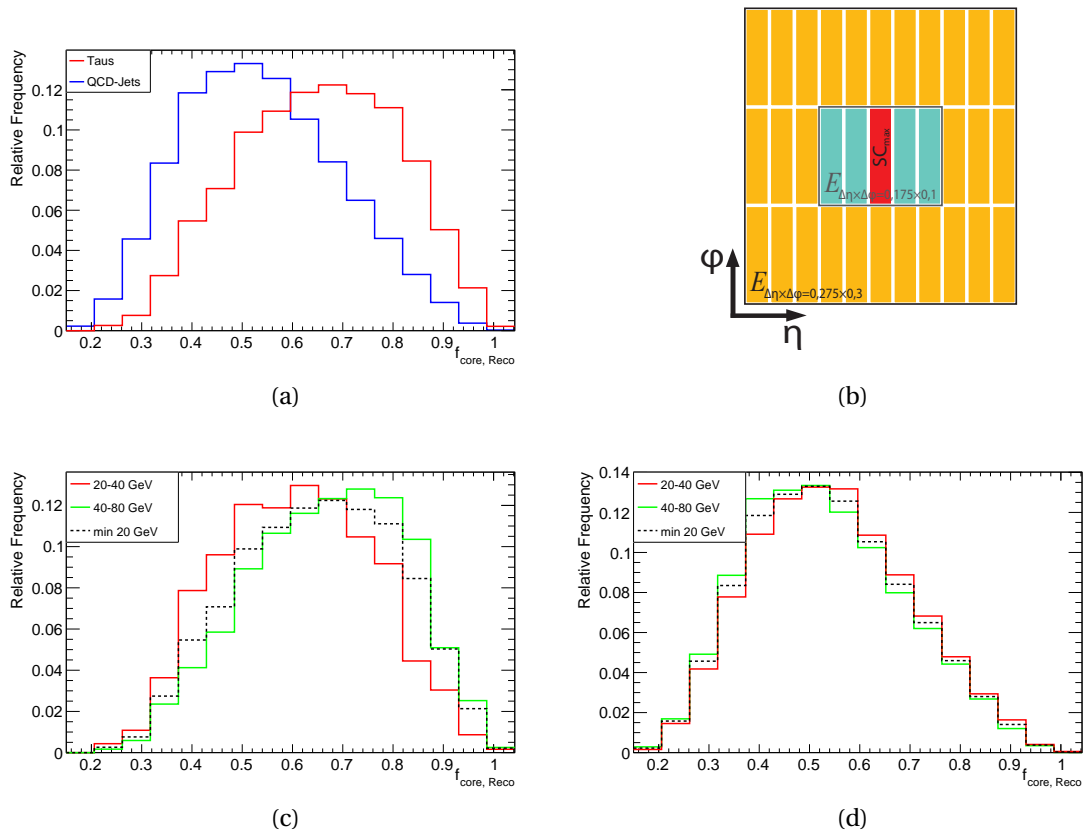
Die Variable  $\rho_{\text{oi}}$  wurde in [14] eingeführt und ist ebenfalls eine Variable für den Schauer im Middle Layer des EMB.

Sie entspricht dem Verhältnis zwischen  $E_{\text{außen}}$  und  $E_{\text{innen}}$ :

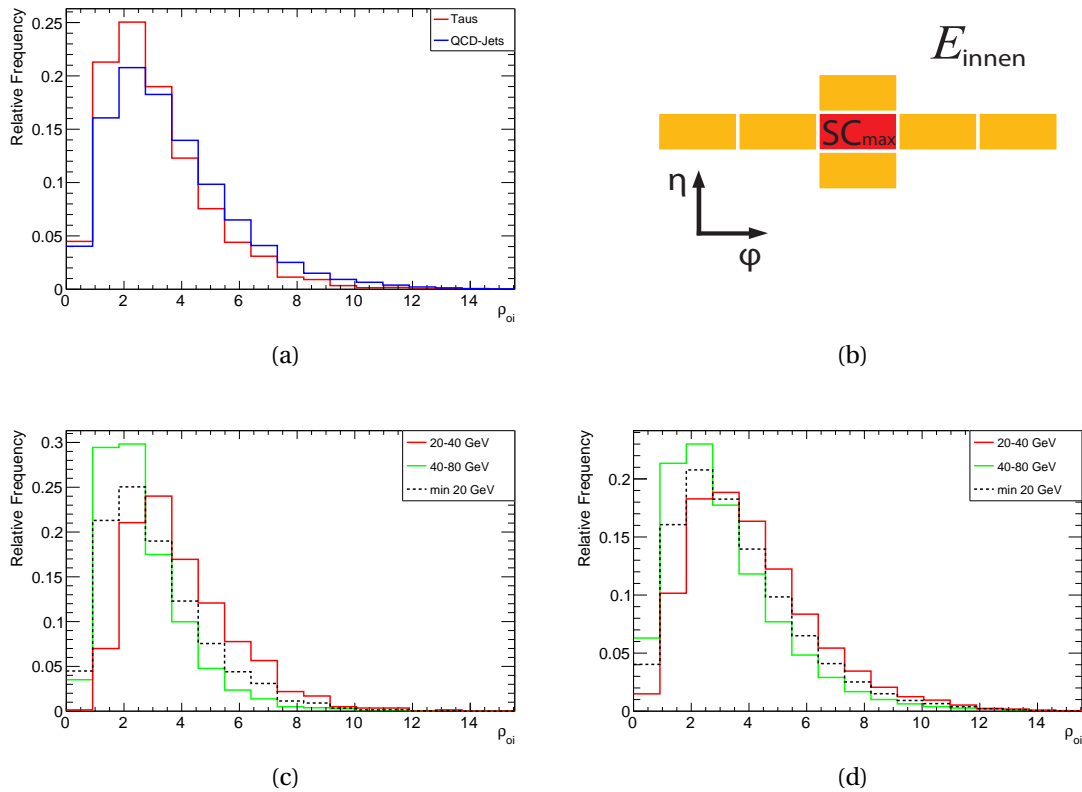
$$\rho_{\text{oi}} = \frac{E_{\Delta\eta \times \Delta\phi=0,7 \times 0,7} - E_{\text{innen}}}{E_{\text{innen}}} = \frac{E_{\text{außen}}}{E_{\text{innen}}}. \quad (3.3)$$

Die Energie  $E_{\text{innen}}$  ist die Summe der Energien der SC mit maximaler Energiedeposition, der Energie der je zwei nächsten SCs in positiver und negativer  $\phi$ -Richtung zu dieser und der beiden benachbarten SCs in  $\eta$ -Richtung zu dieser (siehe Abbildung 3.6 (b)).  $E_{\text{außen}}$  ist die Energie in einem  $\Delta\eta \times \Delta\phi$  -Bereich von  $0,7 \times 0,7$  um die SC mit maximaler Energiedeposition ohne  $E_{\text{innen}}$ . Ähnlich wie  $f_{\text{core}}$  basiert sie darauf, dass die Energie bei Taus zu einem größeren Teil in den zentralen SCs vorliegt als bei QCD-Jets. Der Wert von  $\rho_{\text{oi}}$  ist je kleiner, desto stärker die Energie auf die inneren SCs konzentriert ist. Aus den Abbildungen 3.6 ist zu entnehmen, dass diese Konzentration bei Taus im Durchschnitt größer ist. Außerdem verschieben sich die Verteilungen bei höheren  $p_{\text{T, truth}}$  für Taus und QCD-Jets zu geringeren Werten.

Die Form der Verteilungen entspricht denen in [14], allerdings war die Verteilung für



**Abbildung 3.5.:** Verteilung der relativen Häufigkeiten von  $f_{\text{core}}$  für alle Energien für QCD-Jets und Taus (a) und verschiedene Energiebereiche für Taus (c) und QCD-Jets (d). Darstellung der SCs über die summiert wird (b).



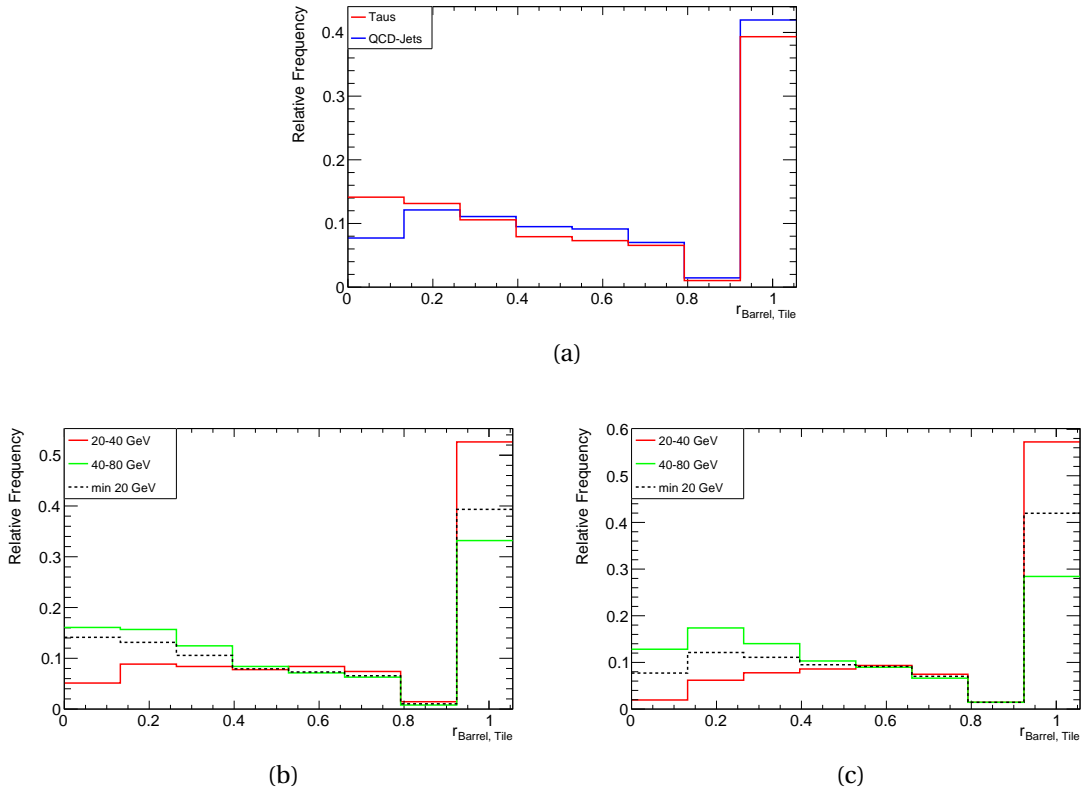
**Abbildung 3.6.:** Verteilung der relativen Häufigkeiten von  $\rho_{oi}$  für alle Energien für QCD-Jets und Taus (a) und verschiedene Energiebereiche für Taus (c) und QCD-Jets (d). Darstellung der SCs über die für  $E_{\text{innen}}$  summiert wird (b).

QCD-Jets wesentlich breiter und der Peak bei ca.  $\rho_{oi} = 1,5$  nur bei Taus vorhanden. Die Gründe für die Unterschiede sind wie bei  $r_{EM}$  mit den Unterschieden der Simulationsdatensätzen zu erklären.

### Schauerprofilvariable $r_{\text{Barrel, Tile}}$

Im Gegensatz zu den bisherigen Variablen werden für  $r_{\text{Barrel, Tile}}$  sowohl LAr-Kalorimeter-Informationen als auch Tile-Kalorimeter-Informationen verwendet, indem das Verhältnis der Energien der beiden Schichten gebildet wird. Es wird die longitudinale Form der Schauer über verschiedene Detektorschichten genutzt.

Die Variable  $r_{\text{Barrel, Tile}}$  wird nach Gleichung 3.4 berechnet, wobei  $E_{\text{Barrel}, \Delta\eta \times \Delta\phi = 0,1 \times 0,1}$  die Energie aller SCs im EMB in einem  $\Delta\eta \times \Delta\phi$ -Bereich von  $0,1 \times 0,1$  um die SC mit maximaler Energiedeposition im Middle Layer ist. Der Wert von  $E_{\text{Tile}, 3 \times 3}$  entspricht der



**Abbildung 3.7.:** Verteilung der relativen Häufigkeiten von  $r_{\text{Barrel, Tile}}$  für alle Energien für QCD-Jets und Taus (a) und verschiedene Energiebereiche für Taus (b) und QCD-Jets (c).

Energie eines  $3 \times 3$  Feldes von TTs, wobei der zentrale TT derjenige mit minimalem  $\Delta R$  zur SC mit maximaler Energiedeposition des Middle Layer ist.

$$r_{\text{Barrel, Tile}} = \frac{E_{\text{Barrel}, \Delta\eta \times \Delta\phi = 0,1 \times 0,1}}{E_{\text{Barrel}, \Delta\eta \times \Delta\phi = 0,1 \times 0,1} + E_{\text{Tile}, 3 \times 3}} \quad (3.4)$$

Die Verteilungen in Abbildung 3.7 zeigen, dass für QCD-Jets und Taus in ca. 40 % der Fälle keine Energie im Tile hinter der SC mit maximaler Energiedeposition des Middle Layers deponiert wird. Dies ist hauptsächlich auf die geringe Energie der QCD- und Tau-Jets zurückzuführen, da sich dieser Anteil für  $40 \text{ GeV} < p_{T, \text{truth}} < 80 \text{ GeV}$  auf ca. 30 % reduziert. Bei gleicher Energie deponieren die Taus mit höherer Wahrscheinlichkeit mehr Energie im Tile-Kalorimeter als QCD-Jets. Dies kann damit zusammenhängen, dass das Tau nur in wenige Teilchen zerfällt, die somit trotz gleicher Gesamtenergie eine höhere Durchschnittsenergie besitzen.

## Schauerprofilvariable $r_{\text{Tile 3, Barrel 3}}$

Im Unterschied zu den bisherigen Variablen wird  $r_{\text{Tile 3, Barrel 3}}$  berechnet, nachdem ein TT mit maximaler Energiedeposition gefunden wurde. Auch für  $r_{\text{Tile 3, Barrel 3}}$  werden Informationen aus LAr- und Tile-Kalorimeter kombiniert, indem ein Verhältnis gebildet wird.

Die Variable ergibt sich aus der Summe der Energien der SCs aller Layer des EMB im  $\Delta\eta \times \Delta\phi$ -Bereich von  $0,3 \times 0,3$  um die Koordinaten des TT mit maximaler Energiedeposition  $E_{\text{Barrel}\Delta\eta \times \Delta\phi=0,3 \times 0,3}$  und aus der Energie  $E_{\text{Tile},3 \times 3}$  des  $3 \times 3$ -Feldes von TTs um den TT mit maximaler Energiedeposition:

$$r_{\text{Tile 3, Barrel 3}} = \frac{E_{\text{Tile},3 \times 3}}{E_{\text{Tile},3 \times 3} + E_{\text{Barrel}\Delta\eta \times \Delta\phi=0,3 \times 0,3}}. \quad (3.5)$$

Anhand der Verteilungen in Abbildungen 3.8 wird deutlich, dass die Tau-Jets, wenn sie das Tile-Kalorimeter treffen, dort einen größeren Anteil ihrer Energie deponieren als der QCD-Jetuntergrund. Dies kann darauf zurückgeführt werden, dass die einzelnen Pionen, die als Zerfallsprodukte des Taus entstehen, eine hohe Energie besitzen und somit erst im Tile-Kalorimeter Energie deponieren. Insbesondere die geladenen  $\pi^\pm$ -Teilchen erzeugen mit hoher Wahrscheinlichkeit erst im Tile-Kalorimeter einen Jet. Dies zeigt sich auch daran, dass für ca. 37% der 7843 Tau-Zerfälle keine SC mit maximaler Energiedeposition im EMB gefunden wird. Dafür sind überwiegend die Zerfallskanäle  $\tau^- \rightarrow \pi^- \nu_\tau$  und  $\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$ , die ca. 32% der 65% hadronischen Tauzerfälle ausmachen, verantwortlich, da bei diesen kein  $\pi^0$  entsteht, welches nahezu sofort in zwei Photonen zerfällt, die ihre Energie mit hoher Wahrscheinlichkeit im EMB deponieren.

## Weitere Variablen

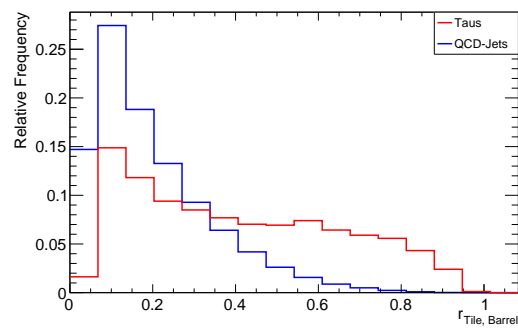
Um später in Abschnitt 3.4.2 mehrere Variablen für die Unterscheidung von Tau-Leptonen und Untergrund einzubeziehen, werden noch einige weitere, ähnlich definierte Variablen eingeführt.

Es werden 28 weitere  $f_{\text{core}}$ -Variablen mit unterschiedlichen großen und kleinen Bereichen definiert, die Tabelle A.1 im Anhang entnommen werden können.

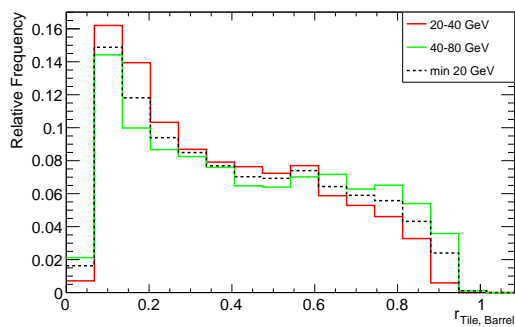
Die Variable  $r_{\text{Tile 3, Barrel 1}}$  wird analog zu  $r_{\text{Tile 3, Barrel 3}}$  definiert mit der Ausnahme, dass statt eines  $0,3 \times 0,3$ -Bereiches im EMB ein  $0,1 \times 0,1$ -Bereich gewählt wird.

Für das Tile-Kalorimeter werden die Formfaktoren  $f_{\text{Tile}, 3, 1}$ ,  $f_{\text{Tile}, 5, 3}$  und  $f_{\text{Tile}, 5, 1}$  eingeführt, die sich nach folgender Formel berechnen:

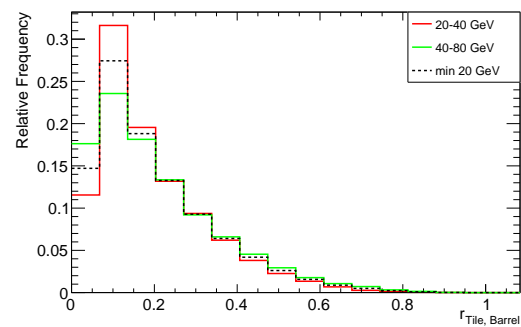
$$f_{\text{Tile}, a, b} = \frac{E_{\text{Tile}, b \times b}}{E_{\text{Tile}, a \times a}}. \quad (3.6)$$



(a)



(b)



(c)

**Abbildung 3.8.:** Verteilung der relativen Häufigkeiten von  $r_{\text{Tile } 3, \text{ Barrel } 3}$  für alle Energien für QCD-Jets und Taus (a) und verschiedene Energiebereiche für Taus (b) und QCD-Jets (c).

## 3.3. Methoden zur Trennung von Signal und Untergrund

Im Folgenden sollen Methoden vorgestellt werden, anhand derer die Schauerprofilvariablen zum Trennen von Signal und Untergrund verwendet werden können.

### 3.3.1. Einfacher Schnitt in einer Variable

Um die Trennkraft einer einzelnen Variablen festzustellen, wird ein einfacher Schnitt an der gleichen Stelle innerhalb der Verteilungen der Variable von Signal und Untergrund durchgeführt, wobei alle Ereignisse auf der einen Seite des Schnitts als Signal und auf der anderen Seite als Untergrund klassifiziert werden.

Anschließend wird die Nachweiseffizienz  $\epsilon$  für Untergrund und Signal nach Formel 3.7 berechnet, wobei  $N_{\text{cut}}$  die Anzahl der Ereignisse von Untergrund bzw. Signal sind, die durch den Schnitt als Signal klassifiziert wurden und  $N_{\text{all}}$  die jeweilige Gesamtanzahl von Signal- bzw. Untergrundereignissen ist.

$$\epsilon = \frac{N_{\text{cut}}}{N_{\text{all}}} \quad (3.7)$$

Bei einer perfekten Unterscheidung wäre die Nachweiseffizienz des Signals  $\epsilon_s = 1$  und die Nachweiseffizienz des Untergrunds  $\epsilon_b = 0$ , da somit jedes Signalereignis auch als Signal und jedes Untergrundereignis auch als Untergrund klassifiziert worden wäre. Überschneiden sich jedoch die Verteilungen der Variablen für Untergrund und Signal kann keine perfekte Unterscheidung getroffen werden. Es muss somit immer ein Kompromiss zwischen  $\epsilon_s$  und  $\epsilon_b$  getroffen werden, welcher von der Stelle an der geschnitten wird abhängt.

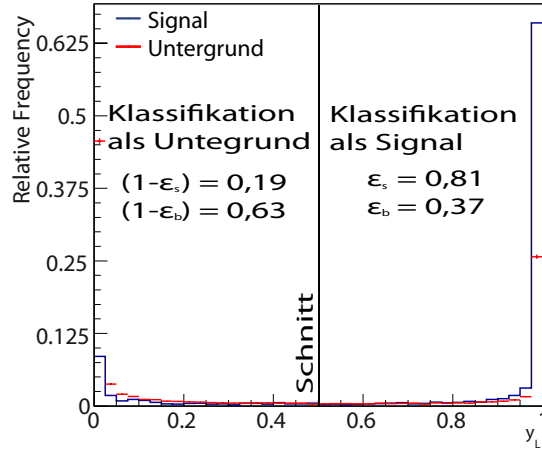
Daher werden  $\epsilon_s$  und  $\epsilon_b$  für viele mögliche Schnittstellen innerhalb der Verteilung der Schauervariablen berechnet und anschließend die Untergrundunterdrückung  $(1 - \epsilon_b)$  in Abhängigkeit von der Signaleffizienz  $\epsilon_s$  dargestellt.

### 3.3.2. Likelihood-Methode

Im Gegensatz zu einem einfachen Schnitt werden bei der Likelihood-Methode mehrere Variablen in die Entscheidung, ob es sich bei einem Ereignis um Untergrund oder Signal handelt, mit einbezogen.

Die Likelihood-Methode wird über das TMVA-Modul [15], welches in die Analyseumgebung ROOT [16] implementiert ist, angewendet. Es teilt die Daten in zwei Teile. Der erste dient zum Training der Methode, der zweite zum Testen dieser.

Die Verteilungen der Variablen  $x_k$  von Untergrund und Signal des ersten Teils werden



**Abbildung 3.9.:** Es ist ein Beispiel für die Verteilung der Relativen Häufigkeit der Likelihood für Signal und Untergrund dargestellt. Die resultierenden Nachweiseffizienzen für einen Schnitt bei  $y_L = 0,5$  sind eingetragen.

von TMVA geglättet und die resultierenden Wahrscheinlichkeitsdichtefunktionen (WDF)  $p_{S(B),k}(x_k)$  gespeichert [15]. Für den zweiten Teil der Daten berechnet TMVA anhand der WDFs für jedes Ereignis  $i$  die Likelihood  $y_L(i)$  nach Formel 3.8 [15].

$$y_L(i) = \frac{L_S(i)}{L_S(i) + L_B(i)} \quad (3.8)$$

Wobei sich  $L_{S(B)}$  nach Formel 3.9 [15] berechnen.

$$L_{S(B)}(i) = \prod_{k=1}^{n_{\text{var}}} p_{S(B),k}(x_k(i)) \quad (3.9)$$

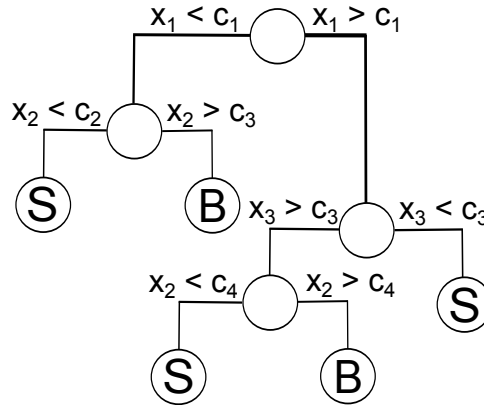
Es ergeben sich für Signal und Untergrund je eine Verteilung der Likelihoods  $f(y_L)$ . In diesen wird ein einfacher Schnitt (Abschnitt 3.3.1) durchgeführt, um die Nachweiseffizienz  $\epsilon_s$  und  $\epsilon_b$  zu bestimmen. In Abbildung 3.9 ist ein Beispiel für die resultierende Verteilung der relativen Häufigkeit der Likelihoods  $y_L$  von Signal und Untergrund dargestellt.

### 3.3.3. Boosted Decision Trees

Für die Anwendung von Boosted Decision Trees (BDTs) wird ebenfalls die Implementierung durch TMVA verwendet. Auch bei BDTs werden mehrere Variablen in die Entscheidung einbezogen. Die erste Hälfte der Daten von Untergrund und Signal dient zum Training der Methode und die zweite Hälfte zum Testen.

Ein einzelner Entscheidungsbaum versucht die Daten in Signal und Untergrund aufzuteilen, indem er an mehreren aufeinanderfolgenden Gabelungen (Nodes) Schnitte in den





**Abbildung 3.10.:** Es ist ein Beispiel für einen Entscheidungsbaum dargestellt. Der Ablauf der Entscheidung erfolgt von oben nach unten. Am Ende der Entscheidungswege werden die Ereignisse entweder als Signal (S) oder Untergrund (B) klassifiziert.

Klassifikationsvariablen durchführt. Jeder so entstandene Weg im Entscheidungsbaum führt entweder zum Ergebnis Signal oder Untergrund (siehe Abbildung 3.10). Für die BDT-Methode werden viele verschiedene Entscheidungsbäume trainiert, indem die Ereignisse anhand der Missklassifikation vorheriger Entscheidungsbäume gewichtet werden.

Ob ein Ereignis der zweiten Hälfte der Daten als Signal oder Untergrund klassifiziert wird, entscheidet eine gewichtete Abstimmung aller so trainierter Entscheidungsbäume. TMVA berechnet für die BDTs für die verwendeten Variablen die variable importance<sup>6</sup>. Diese ergibt sich als Summe der quadrierten Separationszunahmen an den Nodes einer Variable. Die Separation ist wie in Formel 3.10 [15] definiert. Dabei sind  $\hat{y}_S(x)$  und  $\hat{y}_B(x)$  die WDFs des Signals bzw. Untergrunds einer Variable.

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(x) - \hat{y}_B(x))^2}{\hat{y}_S(x) + \hat{y}_B(x)} dx \quad (3.10)$$

<sup>6</sup>Deutsch: Wichtigkeit der Variable

## 3.4. Auswertung

Im Folgenden sollen die in Abschnitt 3.3 vorgestellten Methoden angewendet werden, um die Trennkraft der einzelnen Variablen oder von verschiedenen Teilmengen der Variablen zu untersuchen.

### 3.4.1. Anwendung einfacher Schnitte in einer Variable

Es werden einfache Schnitte auf die Verteilungen der relativen Häufigkeiten der Schau-erprofilvariablen der Taus und Jets angewendet. In Abbildung 3.11 ist als Ergebnis die Kurve der Untergrundunterdrückung zur Signaleffizienz für verschiedene Variablen dargestellt.

Je höher die Kurve einer Variable liegt, desto besser ist ein Schnitt in dieser in der Lage bei gleicher Signaleffizienz den Untergrund zu unterdrücken. Für diese Methode stellen sich somit  $f_{\text{core}}$  und  $r_{\text{Tile 3, Barrel 3}}$  als beste Variablen heraus. Die anderen Variablen besitzen eine Kurve, die nah an der Diagonalen liegt, welche für zwei identische Verteilungen für Signal und Untergrund stehen würde.

In [8] wurde für  $f_{\text{core}}$  ein besseres Ergebnis erzielt, obwohl Taus mit geringerer Energie verwendet worden sind und somit ein schlechteres Ergebnis erwarten werden würde. Für Tau-Leptonen mit  $p_T = 25 \text{ GeV}$  wurde bei einer Signaleffizienz von  $\epsilon_s = 80\%$  eine Untergrundunterdrückung von  $(1 - \epsilon_b) = 60\%$  erreicht, während hier nur ein  $(1 - \epsilon_b) = 50\%$  erreicht wird. Dies kann zum einen mit einer anderen Wahl des Untergrunds zusammenhängen und zum anderen damit, dass hier keine Tau-Leptonen mit festem  $p_T$  gewählt werden.

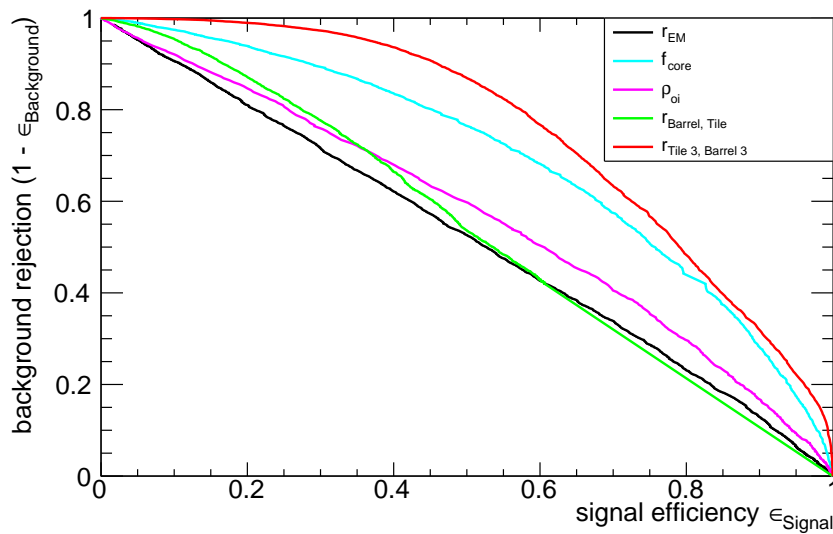
Für  $r_{\text{EM}}$  wird hier im Gegensatz zu [8] fasst keine Trennung von Taus und Untergrund festgestellt. Dies kann neben den für  $f_{\text{core}}$  genannten Gründen daran liegen, dass das hier gewählte  $r_{\text{EM}}$  aus [14] etwas anders definiert ist. Während in [8] die Summe über alle SCs im Bereich  $\Delta\eta \times \Delta\phi = 0,4 \times 0,4$  um die SC mit maximaler Energiedeposition lief, läuft sie hier über alle im Radius  $\Delta R < 0,3$ .

In Abbildung 3.12 ist dargestellt, welche inverse Untergrundeffizienz  $\epsilon_b^{-1}$  die eingeführten Variablen ermöglichen, wenn als Signal die Tau-Datei und als Untergrund die Elektron-Datei verwendet werden. Zum Vergleich ist für  $r_{\text{Tile 3, Barrel 3}}$  die Kurve für QCD-Jets als Untergrund eingetragen.

Die Stufen in den Kurven resultieren aus der geringen Anzahl von Elektronen die durch den Schnitt als Signal eingestuft werden. Übersteigen die Kurven den dargestellten Wertebereich bedeutet dies, dass alle Elektronen richtig klassifiziert werden.

Die Variablen  $r_{\text{EM}}$  und  $\rho_{\text{oi}}$  sind sehr gut geeignet um Tau-Jets von Elektronen zu unterscheiden. Für eine Signaleffizienz von  $\epsilon_s = 80\%$  wird eine inverse Untergrundeffizienz von  $\epsilon_b^{-1} > 10^3$  erreicht.

Die Kurve von  $r_{\text{Tile 3, Barrel 3}}$  liegt fasst über den gesamten Bereich der Signaleffizienz um



**Abbildung 3.11.:** Vergleich der Untergrundunterdrückung in Abhängigkeit von der Signaleffizienz für verschiedene Variablen für Taus als Signal.

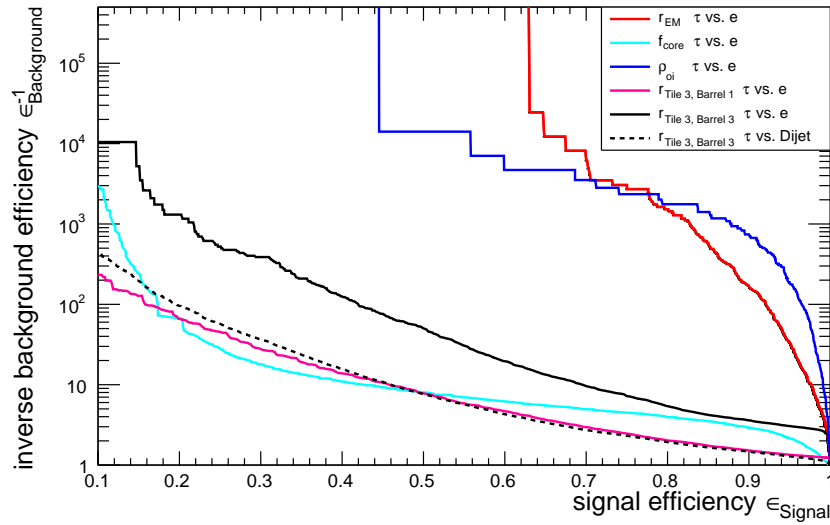
einen Faktor kleiner als  $10^{-3}$  niedriger als die von  $r_{EM}$  und  $\rho_{oi}$ .

Für die Unterscheidung von Tau-Jets zu Elektronen kann mit einfachen Schnitten ein gutes Ergebnis erzielt werden. Es wird deutlich, dass für die Unterscheidung von Tau-Jets zu QCD-Jetuntergrund andere Methoden in Betracht gezogen werden sollten.

### 3.4.2. Anwendung von Boosted Decision Trees und Likelihood-Methode

Um eine bessere Trennung von Taus und QCD-Jetuntergrund zu erreichen und festzustellen welche Möglichkeiten das Zusammenspiel mehrerer Variablen bietet, wird die BDT- und die Likelihoodmethode angewendet.

Bisher wurden alle Variablen gesondert betrachtet. Es kommt jedoch häufig vor, dass ein Tau-Jet nur im Tile oder nur im EMB eine SC mit maximaler Energiedeposition verursacht, die die Schwelle von 1 GeV übersteigt (siehe Tabelle 3.1). Das bedeutet, dass nicht für jedes Tau alle Variablen berechnet werden können. Um jedoch nicht alle Taus zu verwerfen, die nicht überall Energie deponieren und da es möglicherweise auch eine gute Klassifikation sein kann, welche Variable verfügbar ist, wird eine Variable, sollte sie nicht berechnet werden können, auf einen festen Wert außerhalb der möglichen Werte der Verteilung gesetzt.



**Abbildung 3.12.:** Vergleich der inversen Untergrundeﬃzienz in Abhangigkeit von der Signaleﬃzienz fur verschiedene Variablen fur Taus gegen Elektronen bzw. gegen Dijets.

Zunachst wird die BDT-Methode<sup>7</sup> unter Verwendung aller Variablen aus Abschnitt 3.2 durchgefuhrt. Das Ranking nach der variable importance (siehe Abschnitt 3.3.3) ist in Tabelle A.2 zusammengefasst. Die Variablen  $r_{\text{Tile 3, Barrel 3}}$  und  $f_{\text{core}}$ , die auch in Abschnitt 3.4.1 fur einfache Schnitte die besten Resultate erzielt haben, sind auch fur die BDT-Methode die wichtigsten.

Die Variable  $f_{\text{core}}$  ist wichtiger als die anderen zusatzlichen, gleich definierten Variablen fur die blo die Energie in Bereichen anderer Groe summiert wird. Dies entspricht der Erwartung, da in [8] die Groe der Bereiche von  $f_{\text{core}}$  fur die Tauerkennung optimiert wurden. Um festzustellen, welchen Gewinn die weiteren  $f_{\text{core}}$ -Variablen bringen, wird das Training ohne diese wiederholt. Abbildung 3.13 (a) kann entnommen werden, dass die weiteren  $f_{\text{core}}$ -Variablen die inverse Untergrundeﬃzienz um den Faktor 10 verbessern. Zum weiteren Vergleich wird das Training noch fur die 12 besten Variablen im Ranking wiederholt. Das Ergebnis ist nur geringfugig schlechter als das fur alle 37 Variablen.

Insgesamt ermoglicht diese Methode eine wesentlich bessere Klassifikation als die einfachen Schnitte in einer Variablen. Es ist moglich bei  $\epsilon_{\text{tau}}$  von 90 % trotzdem ca. 90 % aller QCD-Jets zu verwerfen. Unter der Annahme, dass QCD-Jets  $10^3$  mal haufiger als Taus auftauchen, waren ca. 99 % der als Tau klassifizierten Events trotzdem auf QCD-Jets zuruckzufuhren.

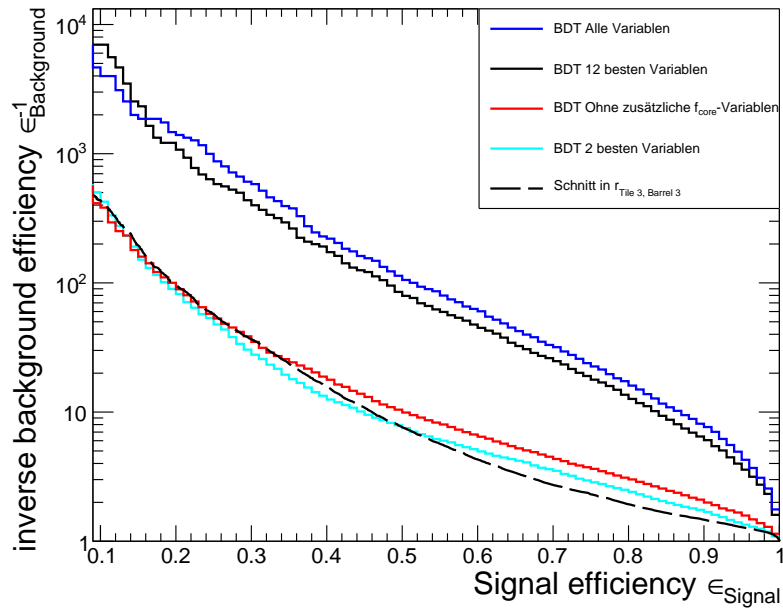
Das Ergebnis fur den Schnitt in der Variable  $r_{\text{Tile 3, Barrel 3}}$  in Abbildung 3.13 (a) erscheint im Vergleich zu den anderen Methoden zunachst besser als erwartet. Allerdings werden

<sup>7</sup>NTrees=850, nEventsMin=150, MaxDepth=3, BoostType=AdaBoost, AdaBoostBeta=0.5, SeparationType=GiniIndex, nCuts=20, PruneMethod=NoPruning

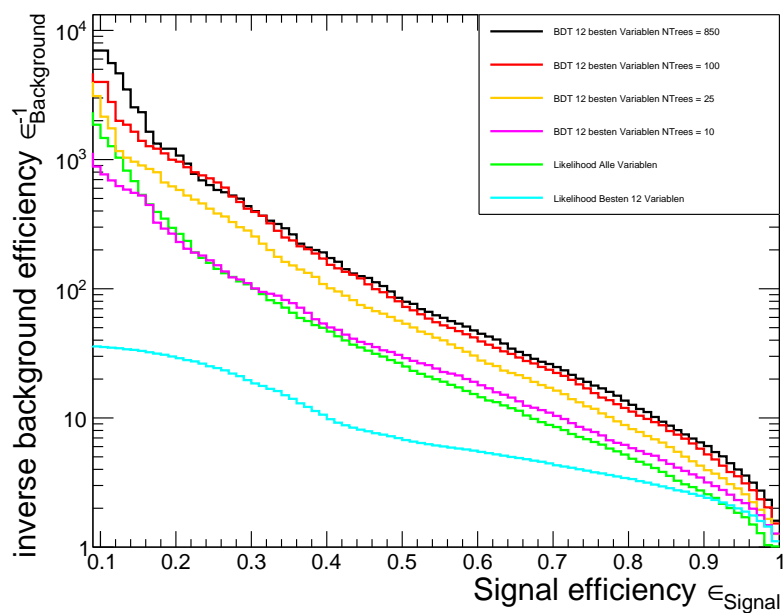
für die Nachweiseffizienzen beim Schnitt nur die Ereignisse einbezogen für die die jeweilige Variable auch berechnet werden kann. Bei den multivariaten Analysen werden hingegen alle Ereignisse einbezogen bei denen mindestens eine Variable gesetzt ist.  $r_{\text{Tile 3, Barrel 3}}$  kann z.B. nur in ca. 80 % der Fälle berechnet werden.

Als Alternative wird das Training mit der Likelihood-Methode durchgeführt und die Ergebnisse mit denen der BDT-Methode für verschiedene Anzahlen der Entscheidungsbäume verglichen (Abbildung 3.13 (b)). Die BDT-Methode erzielt bessere Ergebnisse als die Likelihood-Methode. Die BDT-Methode mit 12 Variablen und 10 Entscheidungsbäumen führt nahezu zum gleichen Ergebnis wie die Likelihood-Methode mit allen 37 Variablen. Die Anzahl von Entscheidungsbäumen lässt sich für die BDT-Methode fast ohne Leistungseinbußen von 850 auf 100 reduzieren.

Ein BDT mit 25 Entscheidungsbäumen und somit insgesamt nur 184 Nodes lieferte ein Ergebnis welches zwischen dem von Likelihood-Methode und BDT-Methode mit 850 Entscheidungsbäumen liegt.



(a)



(b)

**Abbildung 3.13.:** Vergleich der Kurven der inversen Untergrundeffizienz für Klassifikationen mit verschiedenen Variablenmengen und Methoden.

## 3.5. Zusammenfassung und Ausblick

Zunächst wurden in Abschnitt 3.2 aus einer Simulation der Informationen höherer Granularität, die dem Triggersystem des ATLAS-Detektors nach dem Phase-I-Upgrade bereitstehen, Schauerprofilvariablen berechnet. Die wichtigste Ursache für Unterschiede in den Verteilungen der relativen Häufigkeiten der Variablen von Taus und QCD-Jetuntergrund ist, dass der hadronische Zerfall des Tau-Leptons hauptsächlich zu geringen Anzahlen von Tochterpartikeln in einem kleinen  $\Delta R$ -Kegel stattfindet.

Die Schauerprofilvariablen wurden in Abschnitt 3.4 auf ihre Qualität hadronisch zerfallende Tau-Leptonen gegen QCD-Jetuntergrund zu identifizieren untersucht, indem sie für einfache Schnitte und für BDT- und Likelihood-Methodentraining verwendet wurden. Dabei haben sich  $f_{\text{core}}$  und  $r_{\text{Tile 3, Barrel 3}}$  als die besten Variablen erwiesen.

Die im BDT-"variable importance"-Ranking wichtigste Variable  $r_{\text{Tile 3, Barrel 3}}$  wurde berechnet, indem zum ersten Mal die mittels AREUS simulierten Tile-Kalorimeter- und LAr-Kalorimeter-Informationen zusammen in einer Schauerprofilvariable verwendet wurden. Dies zeigt, dass es sinnvoll ist für die Triggerentscheidung ein Zusammenspiel aus Tile- und LAr-Kalorimeter in Betracht zu ziehen.

Für eine Signaleffizienz von  $\epsilon_s = 80\%$  erreicht ein Schnitt in  $r_{\text{Tile 3, Barrel 3}}$  eine Untergrundunterdrückung von  $(1 - \epsilon_b) = 50\%$ , ein Schnitt in  $f_{\text{core}}$  eine von  $40\%$ , die Likelihood-Methode eine von  $80\%$  und die BDT-Methode  $90\%$ .

Für die BDT-Methode haben sich noch weitere  $f_{\text{core}}$ -ähnliche Variablen mit einem kleinen Innenbereich von  $\Delta\eta \times \Delta\phi = 0.075 \times 0.1$  und verschiedenen großen Außenbereichen als nützlich erwiesen.

Für die Unterscheidung von Tau-Leptonen und Elektronen wurde mit einfachen Schnitten in den Variablen  $r_{\text{EM}}$  und  $\rho_{\text{oi}}$  ein gutes Ergebnis erzielt. Für eine Nachweiseffizienz der Taus von  $\epsilon_s = 80\%$  können  $99,9\%$  der Elektronen unterdrückt werden.

Das Triggersystem wird auf Field Programmable Gate Arrays (FPGAs) implementiert. Bei FPGAs handelt es sich um integrierte Schaltkreise in die logische Schaltungen programmiert werden können. In einem nächsten Schritt sollte deshalb die Berechnung der Schauerprofilvariablen mit einem FPGA untersucht werden. Dies sollte prinzipiell möglich sein, da zur Berechnung hauptsächlich einfache Rechenoperationen zum Einsatz gekommen sind.

Die multivariaten Methoden dienen vorrangig dazu, festzustellen welche Möglichkeiten das Zusammenspiel der Variablen prinzipiell bietet, da es möglicherweise sehr aufwändig ist eine BDT-oder Likelihood-Methode auf einem FPGA zu programmieren. Eine Machbarkeitsstudie über eine FPGA-Implementierung dieser Methoden wäre aufgrund der besseren Ergebnisse allerdings interessant.

Um eine Triggerrate zu berechnen, wäre es wichtig einzubeziehen mit welcher Häufigkeit und welchem Impuls Signal und Untergrund tatsächlich auftreten. Um dafür

Dateien zu erzeugen, die Tau-Leptonen einer beliebigen  $p_T$ -Verteilung unabhängig vom physikalischen Entstehungsprozess enthält, wäre eine Tau-Partikelkanone nützlich. Weiterhin wäre dafür nötig einen Algorithmus zu definieren, der unabhängig von Anti- $k_t$ -truth-Jets ist und nur auf den Informationen des Triggers beruht. Beispielsweise könnte ein Ausgangspunkt für die Berechnung der Variablen eine TT oder eine SC sein, welche die maximale Energie aller SCs und TTs in einem Bereich  $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$  besitzt. Der Abstand von dieser zu anderen SCs bzw. TTs, die einen gewisse Energieschwelle überschreiten, könnte ein weiterer Klassifikator sein.

Eine weitere Optimierung der Bereiche, in denen die Energien für die Variablen summiert werden, könnte sich als nützlich erweisen. Für den Bereich, in dem im EMB für die beste Variable  $r_{\text{Tile 3, Barrel 3}}$  die Energien summiert werden, wird zum Beispiel die höhere Granularität des LAr-Kalorimeters noch nicht genutzt.



# A. Anhang

**Tabelle A.1.:** Weitere Variablen  $f_{\text{core}} = \frac{E_{\text{kleiner Bereich}}}{E_{\text{großer Bereich}}}$ .

Bezeichnung	großer $\Delta\eta \times \Delta\phi$ -Bereich	kleiner $\Delta\eta \times \Delta\phi$ -Bereich
$f_{\text{core}, 5, 3}$	$0,125 \times 0,1$	$0,075 \times 0,1$
$f_{\text{core}, 7, 3}$	$0,175 \times 0,1$	$0,075 \times 0,1$
$f_{\text{core}, 7, 5}$	$0,175 \times 0,1$	$0,125 \times 0,1$
$f_{\text{core}, 9, 3}$	$0,225 \times 0,3$	$0,075 \times 0,1$
$f_{\text{core}, 9, 5}$	$0,225 \times 0,3$	$0,125 \times 0,1$
$f_{\text{core}, 9, 7}$	$0,225 \times 0,3$	$0,175 \times 0,1$
$f_{\text{core}, 11, 3}$	$0,275 \times 0,3$	$0,075 \times 0,1$
$f_{\text{core}, 11, 5}$	$0,275 \times 0,3$	$0,125 \times 0,1$
$f_{\text{core}, 11, 7}$	$0,275 \times 0,3$	$0,175 \times 0,1$
$f_{\text{core}, 11, 9}$	$0,275 \times 0,3$	$0,225 \times 0,3$
$f_{\text{core}, 13, 3}$	$0,325 \times 0,5$	$0,075 \times 0,1$
$f_{\text{core}, 13, 5}$	$0,325 \times 0,5$	$0,125 \times 0,1$
$f_{\text{core}, 13, 7}$	$0,325 \times 0,5$	$0,175 \times 0,1$
$f_{\text{core}, 13, 9}$	$0,325 \times 0,5$	$0,225 \times 0,3$
$f_{\text{core}, 13, 11}$	$0,325 \times 0,5$	$0,275 \times 0,3$
$f_{\text{core}, 15, 3}$	$0,375 \times 0,5$	$0,075 \times 0,1$
$f_{\text{core}, 15, 5}$	$0,375 \times 0,5$	$0,125 \times 0,1$
$f_{\text{core}, 15, 7}$	$0,375 \times 0,5$	$0,175 \times 0,1$
$f_{\text{core}, 15, 9}$	$0,375 \times 0,5$	$0,225 \times 0,3$
$f_{\text{core}, 15, 11}$	$0,375 \times 0,5$	$0,275 \times 0,3$
$f_{\text{core}, 15, 13}$	$0,375 \times 0,5$	$0,325 \times 0,5$
$f_{\text{core}, 17, 1}$	$0,425 \times 0,7$	$0,075 \times 0,1$
$f_{\text{core}, 17, 3}$	$0,425 \times 0,7$	$0,125 \times 0,1$
$f_{\text{core}, 17, 5}$	$0,425 \times 0,7$	$0,175 \times 0,1$
$f_{\text{core}, 17, 7}$	$0,425 \times 0,7$	$0,225 \times 0,3$
$f_{\text{core}, 17, 9}$	$0,425 \times 0,7$	$0,275 \times 0,3$
$f_{\text{core}, 17, 11}$	$0,425 \times 0,7$	$0,325 \times 0,5$
$f_{\text{core}, 17, 13}$	$0,425 \times 0,7$	$0,375 \times 0,5$

**Tabelle A.2.:** Variable importance bei BDT-Training mit allen Variablen.

Rang	Variable	Variable importance
1	$r_{\text{Tile 3, Barrel 3}}$	$6,253 \cdot 10^{-2}$
2	$f_{\text{core}}$	$5,058 \cdot 10^{-2}$
3	$f_{\text{core, 13, 3}}$	$4,255 \cdot 10^{-2}$
4	$f_{\text{core, 9, 3}}$	$3,890 \cdot 10^{-2}$
5	$f_{\text{core, 15, 3}}$	$3,855 \cdot 10^{-2}$
6	$f_{\text{core, 17, 3}}$	$3,802 \cdot 10^{-2}$
7	$f_{\text{core, 11, 3}}$	$3,760 \cdot 10^{-2}$
8	$f_{\text{core, 7, 3}}$	$3,288 \cdot 10^{-2}$
9	$f_{\text{Tile, 3, 1}}$	$3,138 \cdot 10^{-2}$
10	$r_{\text{Barrel, Tile}}$	$3,100 \cdot 10^{-2}$
11	$f_{\text{core, 5, 3}}$	$2,911 \cdot 10^{-2}$
12	$r_{\text{Tile 3, Barrel 1}}$	$2,892 \cdot 10^{-2}$
13	$\rho_{\text{oi}}$	$2,763 \cdot 10^{-2}$
14	$f_{\text{core, 17, 9}}$	$2,592 \cdot 10^{-2}$
15	$f_{\text{core, 17, 15}}$	$2,586 \cdot 10^{-2}$
16	$f_{\text{core, 9, 5}}$	$2,530 \cdot 10^{-2}$
17	$f_{\text{core, 13, 7}}$	$2,467 \cdot 10^{-2}$
18	$f_{\text{core, 15, 9}}$	$2,432 \cdot 10^{-2}$
19	$f_{\text{core, 17, 7}}$	$2,427 \cdot 10^{-2}$
20	$f_{\text{core, 15, 13}}$	$2,407 \cdot 10^{-2}$
21	$f_{\text{core, 13, 5}}$	$2,298 \cdot 10^{-2}$
22	$f_{\text{core, 17, 13}}$	$2,297 \cdot 10^{-2}$
23	$f_{\text{core, 13, 9}}$	$2,286 \cdot 10^{-2}$
24	$f_{\text{core, 15, 7}}$	$2,153 \cdot 10^{-2}$
25	$f_{\text{Tile, 5, 3}}$	$2,119 \cdot 10^{-2}$
26	$f_{\text{core, 9, 7}}$	$2,112 \cdot 10^{-2}$
27	$f_{\text{core, 11, 9}}$	$2,042 \cdot 10^{-2}$
28	$f_{\text{core, 13, 11}}$	$2,041 \cdot 10^{-2}$
29	$f_{\text{core, 17, 5}}$	$1,964 \cdot 10^{-2}$
30	$f_{\text{core, 17, 11}}$	$1,956 \cdot 10^{-2}$
31	$f_{\text{core, 15, 11}}$	$1,947 \cdot 10^{-2}$
32	$r_{\text{EM}}$	$1,885 \cdot 10^{-2}$
33	$f_{\text{Tile, 5, 1}}$	$1,847 \cdot 10^{-2}$
34	$f_{\text{core, 7, 5}}$	$1,814 \cdot 10^{-2}$
35	$f_{\text{core, 15, 5}}$	$1,765 \cdot 10^{-2}$
36	$f_{\text{core, 11, 7}}$	$1,563 \cdot 10^{-2}$
37	$f_{\text{core, 11, 5}}$	$1,532 \cdot 10^{-2}$

# Abkürzungsverzeichnis

BC	Bunchcrossing
BDTs	Boosted Decision Trees
CERN	Europäische Organisation für Kernforschung
EMB	Elektromagnetisches Barrel-Kalorimeter
FPGA	Field Programmable Gate Array
ID	innerer Detektor
LAr-Kalorimeter	Flüssig-Argon Kalorimeter
LEP	Large Electron-Positron Collider
LHC	Large Hadron Collider
SC	Superzelle
TT	Tiletower
WDF	Wahrscheinlichkeitsdichtefunktion

# Abbildungsverzeichnis

2.1. Aufbau des ATLAS-Detektors [5]. . . . .	3
2.2. Aufbau des LAr-Kalorimeters [6]. . . . .	4
2.3. Aufbau der Schichten der Kalorimeter in der Vergrößerung [7]. . . . .	5
2.4. Vergleich von Superzellen und Triggertower . . . . .	6
3.1. $p_{T, \text{truth}}$ -Verteilungen der Dateien . . . . .	9
3.2. $\Delta R$ -Verteilungen . . . . .	9
3.3. Abstand Jet Heißeste Superzelle . . . . .	11
3.4. Schauerprofilvariable $r_{EM}$ . . . . .	12
3.5. Schauerprofilvariable $f_{core}$ . . . . .	14
3.6. Schauerprofilvariable $\rho_{oi}$ . . . . .	15
3.7. Schauerprofilvariable $r_{Barrel, Tile}$ . . . . .	16
3.8. Schauerprofilvariable $r_{Tile\ 3, Barrel\ 3}$ . . . . .	18
3.9. Beispiel für Likelihood Verteilung und Entscheidungsbaum . . . . .	20
3.10. Beispiel für Entscheidungsbaum . . . . .	21
3.11. Vergleich der Untergrundunterdrückung für Taus als Signal . . . . .	23
3.12. Vergleich der inversen Untergrundeﬃzienz für Taus und Elektronen als Signal . . . . .	24
3.13. Vergleich der Untergrundunterdrückung für verschiedene Methoden und Variablenmengen . . . . .	26

# Tabellenverzeichnis

2.1. Zerfallskanäle des Tau-Leptons . . . . .	6
3.1. Relative Häufigkeit mit der jeweilige Kombination aus SC und TT mit maximaler Energieposition und $E_{SC(TT)} > 1\text{GeV}$ auftritt. . . . .	10
A.1. weitere $f_{\text{core}}$ -Variablen . . . . .	29
A.2. Variable importance bei BDT-Training mit allen Variablen. . . . .	30

# Literaturverzeichnis

- [1] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.
- [2] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report Vol.1: The LHC Main Ring*, .
- [3] ATLAS, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [4] ATLAS LArG, H. Wilkens, *The ATLAS liquid argon calorimeter: An overview*, J. Phys. Conf. Ser. **160** (2009) 012043.
- [5] J. Pequeno, *Computer generated image of the whole ATLAS detector*, Mar, 2008.
- [6] J. Pequeno, *Computer generated image of the ATLAS Liquid Argon*, Mar, 2008.
- [7] ATLAS Liquid Argon Calorimeter, C. Gabaldon, *Drift time measurement in the ATLAS liquid argon electromagnetic calorimeter using cosmic muons*, J. Phys. Conf. Ser. **293** (2011) 012050.
- [8] M. C. Aleksa, W. P. Cleland, Y. T. Enari, M. V. Fincke-Keeler, L. C. Hervas, F. B. Lanni, S. O. Majewski, C. V. Marino, and I. L. Wingerter-Seez, *ATLAS Liquid Argon Calorimeter Phase-I Upgrade Technical Design Report*, Tech. Rep. CERN-LHCC-2013-017. ATLAS-TDR-022, CERN, Geneva, Sep, 2013.  
<http://cds.cern.ch/record/1602230>. Final version presented to December 2013 LHCC.
- [9] Particle Data Group, K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.
- [10] N. D. Christensen, T. Han, and S. Su, *MSSM Higgs Bosons at The LHC*, Phys. Rev. **D85** (2012) 115018, arXiv:1203.3207 [hep-ph].
- [11] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].
- [12] GEANT4, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

- [13] J. P. Grohs, *Simulation of the future Phase-1 Trigger Readout Upgrade of the Liquid-Argon Calorimeter of the ATLAS Detector at the LHC*. PhD Thesis, Technische Universität Dresden, 2015.
- [14] R. Wolff, *Erkennung von Tau-Leptonen mit dem verbesserten Trigger des ATLAS-Flüssigargon-Kalorimeters*. Bachelor Arbeit, TU Dresden, 2013.
- [15] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.
- [16] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, in *AIHENP'96 Workshop, Lausanne*, vol. 389, pp. 81–86. 1996.

# Danksagung

Ich bedanke mich bei allen, die mich bei dieser Arbeit unterstützt haben. Prof. Dr. Arno Straessner danke ich für das interessante Thema, das ich bearbeiten durfte. Maximilian Hils, Dr. Olga Novgorodova und Dr. Wolfgang Mader danke ich für das aufmerksame und kritische Lesen und Kommentieren meiner Arbeit. Beim gesamten Elektronik-Upgrade Team bedanke ich mich für die angenehme Arbeitsatmosphäre und die guten Ratschläge in den wöchentlichen Treffen, die mir immer weiter geholfen haben.



# Versicherung

Hiermit versichere ich, die vorliegende Arbeit selbständig, ohne fremde Hilfe und ohne Benutzung anderer als der von mir angegebenen Quellen angefertigt zu haben. Alle aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche gekennzeichnet. Die Arbeit wurde noch keiner Prüfungsbehörde in gleicher oder ähnlicher Form vorgelegt.

Dresden, 22. Februar 2016

Christian Wiel