

Technische Universität Dresden
Bereich Mathematik und Naturwissenschaften
Fakultät Physik
Institut für Kern- und Teilchenphysik

Wissenschaftliche Arbeit im Fach Physik

zur Erlangung des Hochschulgrades

1. Staatsexamen

im Studiengang

Lehramt an Gymnasien

Detektorspezifische Energierekonstruktion mit neuronalen Netzen für die ATLAS Flüssig-Argon-Kalorimeter

Gutachter: Prof. Dr. Arno Straessner
Dr. Frank Siegert

Eingereicht von: Alexander Keyn

Geboren am: 06.01.1993

Dresden, 06.10.2022

Abstrakt

Nach dem Phase-II Upgrade des LHC wird die Luminosität erwartungsgemäß um das 3- bis 5-fache steigen. Die dadurch erhöhte Anzahl an Teilchenkollisionen erhöht die Ansprüche an die Technik und die Messelektronik, auch im LAr-Kalorimeter des ATLAS-Detektors. Derzeitige Forschungen untersuchen, ob zur Energierekonstruktion der bisher eingesetzte Optimalfilter mit Maximums Finder durch künstliche neuronale Netze ersetzt werden kann. In der Diskussion sind die beiden vielversprechenden 3Conv- und 4Conv-Netze, die bisher nur auf einer Zelle des Middle-Layer des LAr-Kalorimeters trainiert wurden. In der vorliegenden Arbeit wurden die beiden Netze ohne und mit Training auf fünf weitere Zellen des LAr-Kalorimeters angewendet, wobei die zugehörigen Datensätze mit AREUS simuliert wurden. Dabei zeigte sich, dass das Laden der Parameter in die Netze vor dem Training die Trainingsdauer verkürzte und die Energierekonstruktion im Vergleich zur Auswertung ohne Training deutlich verbesserte.

Inhaltsverzeichnis

Abkürzungsverzeichnis	2
1. Einleitung	3
2. Physikalische und theoretische Grundlagen	4
2.1. Der Large Hadron Collider (LHC)	4
2.2. Der ATLAS-Detektor am LHC	5
2.3. Das Flüssig-Argon-Kalorimeter	8
2.4. Maschinelles Lernen	11
2.4.1. Künstliche neuronale Netze	11
2.4.2. Training von neuronalen Netzen	14
3. Detektorspezifische Energierekonstruktion im LAr-Kalorimeter	15
3.1. Nutzung von Faltungsnetzen für die Energierekonstruktion	15
3.2. Betrachtete Detektorregionen	18
3.3. Performanz der CNN's ohne Training auf den Zellen	19
3.4. Performanz der CNN's mit Training auf den Zellen	22
3.4.1. Einfluss der Anzahl an Epochen und der Lernrate	23
3.4.2. Einfluss der Energieschwelle auf die Pulserkennung	24
3.4.3. Einfluss statistischer Unsicherheiten	26
4. Fazit und Ausblick	28
5. Anhang	29
Literaturverzeichnis	36
Abbildungsverzeichnis	38

Abkürzungsverzeichnis

ALICE	Eigenname, Akronym für <i>A Large Ion Collider Experiment</i>
AREUS	ATLAS Readout Electronics Upgrade Simulation
ATLAS	Eigenname, Akronym für <i>A Toroidal LHC Apparatus</i>
BC	Paket-Kollision (engl.: <i>bunch crossing</i>)
BEE	<i>Back-End-Elektronik</i>
CERN	europäische Organisation für Kernforschung
CMS	Eigenname, Akronym für <i>Compact Myon Solenoid</i>
CNN	Faltungsnetz (engl.: <i>convolutional neural network</i>)
EMB	<i>electromagnetic barrel</i>
EMEC	<i>electromagnetic endcaps</i>
F05, F10	Zellen aus dem Front-Layer mit einer Pseudorapidität von 0,5015625 bzw. 1,0015625
FEE	<i>Front-End-Elektronik</i>
HL-LHC	<i>High Luminosity LHC</i>
LHC	<i>Large Hadron Collider</i>
LHCb	Large Hadron Collider beauty
M05, M10, M20	Zellen aus dem Middle-Layer mit einer Pseudorapidität von 0,5125, 1,0125 bzw. 2,0125
NN	künstliches neuronales Netz (engl.: <i>artificial neural network</i>)
OF	Optimalfilter
OFMax	Optimalfilter mit Maximumserkennung (engl.: <i>Optimal Filter with Maximum Finder</i>)
PS	Zelle aus dem Presampler
3Conv-Netz	CNN bestehend aus zwei Schichten für das <i>Pulse Tagging</i> und einer Schicht für die <i>Energy Reconstruction</i>
4Conv-Netz	CNN bestehend aus zwei Schichten für das <i>Pulse Tagging</i> und zwei Schichten für die <i>Energy Reconstruction</i>

1. Einleitung

Derzeit wird untersucht, ob künstliche neuronale Netze nach dem High-Luminosity Upgrade des Large Hadron Collider zur Energierekonstruktion im Flüssig-Argon-Kalorimeter des ATLAS-Detektors geeignet sind. In bisherigen Forschungen wurden dazu zwei künstliche neuronale Faltungsnetze mit drei bzw. vier Schichten entwickelt, die im Folgenden 3Conv- und 4Conv-Netz genannt werden und die bisher nur auf einer Zelle aus dem Middle-Layer des Kalorimeters trainiert wurden. Diese Arbeit untersucht, ob die beiden Netze auch auf anderen Zellen des Kalorimeters eingesetzt werden können. Dazu wurden fünf weitere Zellen ausgewählt, die für die verschiedenen Regionen des Kalorimeters repräsentativ sind. Die beiden Netze wurden dann sowohl ohne als auch mit Training auf den Datensätzen zu diesen Zellen ausgewertet. Für die Trainingsphase wurde untersucht, welchen Einfluss die Anzahl an Epochen, die Lernrate und die Energieschwelle der Pulserkennung auf die Ergebnisse haben.

2. Physikalische und theoretische Grundlagen

2.1. Der Large Hadron Collider (LHC)

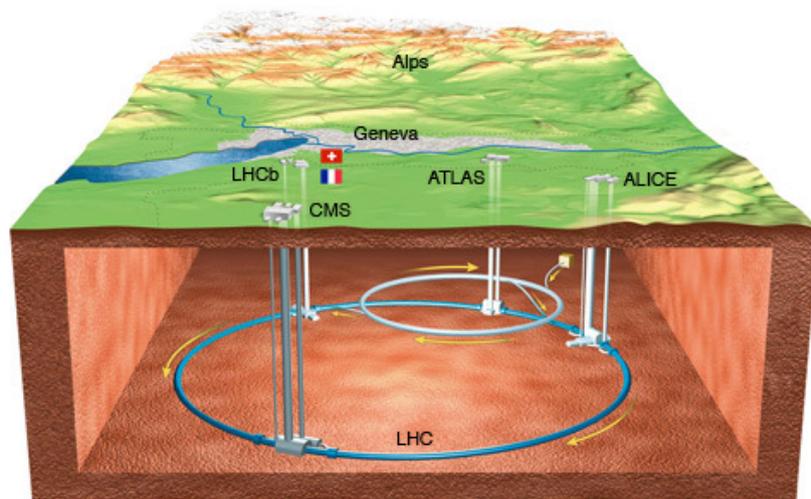


Abb. 1: Schematische Gesamtansicht des LHC. [1]

Der Large Hadron Collider in der Nähe von Genf in der Schweiz ist ein kreisförmiger Teilchenbeschleuniger mit einem Umfang von rund 27 km (siehe Abb. 1), womit er der bisher weltweit größte Teilchenbeschleuniger ist. In ihm werden in einem ultrahohem Vakuum zwei Strahlen aus Protonen oder Atomkernen, die in Teilchenpakete (engl.: *bunches*) eingeteilt sind, getrennt in entgegengesetzter Richtung auf nahezu Lichtgeschwindigkeit beschleunigt, bevor sie dann in einem der vier Detektoren entlang des Rings aufeinander treffen. Dazu sind im LHC ca. 1200 Dipolmagnete zur Beschleunigung und 400 Quadrupolmagnete zur Fokussierung der Teilchenpakete verbaut. Dabei werden supraleitende Elektromagnete benutzt, weshalb diese mit Helium auf konstante $-271,3\text{ °C}$ herabgekühlt werden müssen. Aktuell kann bei den Paket-Kollisionen (engl.: *bunch crossings*; Abk.: BC) im LHC eine Schwerpunktsenergie von 13,6 TeV erreicht werden. Dabei treffen die Pakete mit einer Frequenz von ca. 40 MHz aufeinander, weshalb im Detektor im Mittel alle 25 ns eine Kollision stattfindet. Durch diese Kollisionen können sehr massereiche Teilchen erzeugt werden, die nach einer begrenzten Lebensdauer in leichtere Teilchen zerfallen, deren Impuls und Energie in den Detektoren gemessen werden. Für die Vermessungen der Teilchenschauer werden am LHC die vier Detektoren ALICE, ATLAS, CMS und LHCb benutzt. Aus den erhaltenen Messdaten können Rückschlüsse auf die Art der erzeugten Teilchen und deren Eigenschaften gezogen werden sowie bekannte Phänomene oder neue Theorien untersucht werden. Dazu gehört u. a. die Untersuchung von Dunkler Materie, kleinen Schwarzen Löcher, Supersymmetrie und das Higgs-Boson. [2]

Die Experimente am LHC wurden seit seiner Inbetriebnahme 2010 in mehrere Phasen (engl.: *runs*) unterteilt. Zwischen jeder Phase fanden Verbesserungen am LHC statt, um die Messgenauigkeiten und -möglichkeiten zu erhöhen. Ab 2026 soll das LHC zum High Luminosity LHC (kurz: HL-LHC) aufgerüstet werden, um die Luminosität im Vergleich zu Run 3 um das 3- bis 5-fache zu erhöhen. Damit steigen die Anzahl an Kollisionen in den Detektoren deutlich an, womit auch die absolute Häufigkeit von seltenen Ereignissen pro Strahlenkreuzung, wie der Erzeugung von Higgs-Bosonen, ansteigt. Diese Verbesserungen gehen mit höheren Ansprüchen an das Vakuum, die Messgenauigkeiten, die Kühlung und den Schutz von Maschinen und Menschen einher. Eine Übersicht der Betriebsphasen des LHC bzw. HL-LHC sind in Abb. 2 dargestellt. [3]



Abb. 2: Übersicht der Betriebsphasen des LHC bzw. HL-LHC. [3]

2.2. Der ATLAS-Detektor am LHC

Der ATLAS-Detektor hat die Form eines 46 m langen Zylinders mit einem Durchmesser von 25 m und einem Gesamtgewicht von 7000 t. Er wird u. a. zur Suche nach schweren Teilchen oder nach Teilchen, die durch die Supersymmetrie vorhergesagt werden, benutzt. So wurde in ihm 2012 die Existenz des Higgs-Boson nachgewiesen [4, 5]. Sein Aufbau kann in die vier Bereiche Magnetsystem, Inneren Detektor, Kalorimeter und Myon-Spektrometer unterteilt werden, deren Systembestandteile in Abb. 3 dargestellt sind.

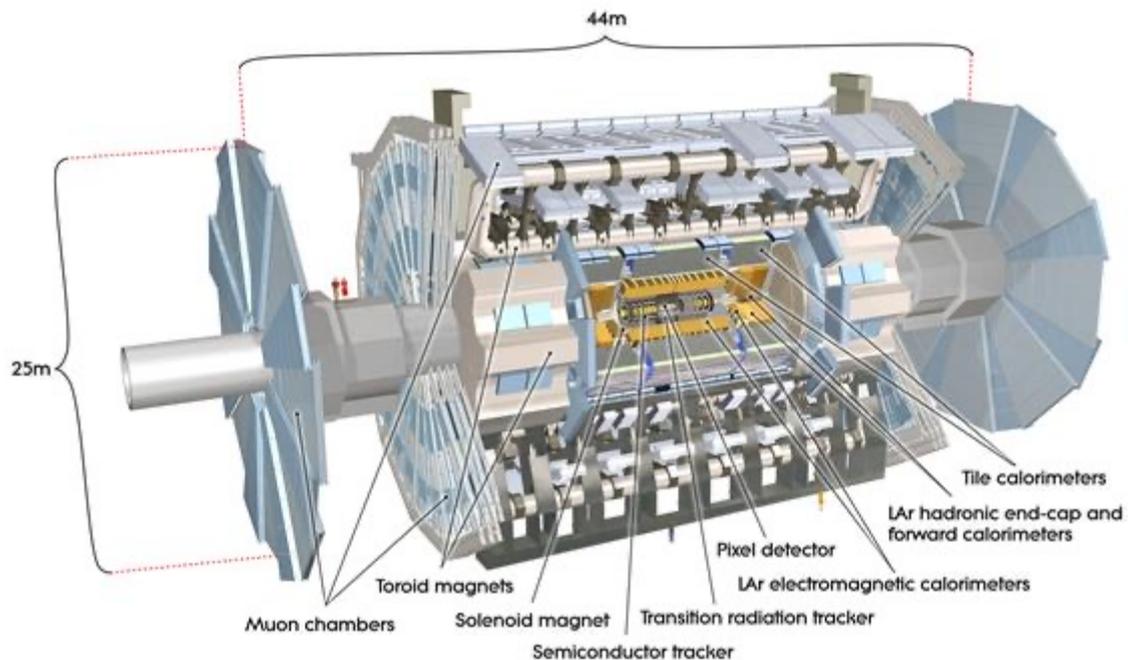


Abb. 3: Schematischer Aufbau des ATLAS-Detektors [6]

Die vier Bereiche und deren Systeme sind dabei so konzipiert, dass sie einen möglichst großen Raumwinkel abdecken zur Detektion der Zerfallsprodukte, die bei der Kollision in der Mitte des Detektors entstehen. So müssen die Elementarteilchen auf ihrem Weg vom Kollisionsursprung nach außen alle Bereiche passieren, bis sie absorbiert werden oder den Detektor verlassen, und dabei mit den verschiedenen Materialien in den Detektoren wechselwirken.

Dabei passieren sie als erstes den Inneren Detektor, der aus den drei Hauptkomponenten *Pixel Detector*, *Semiconductor Tracker* und *Transition Radiation Tracker* besteht. Aus der im *Pixel Detector* deponierten Energie werden der Ursprung der Kollision und der Impuls der Teilchen berechnet. Im *Semiconductor Tracker* wird die Flugbahn der geladenen Teilchen verfolgt und danach werden im *Transition Radiation Tracker* neben der Flugbahn auch die Ladung der Teilchen detektiert. Bei dem letzten Detektor wird ausgenutzt, dass die geladenen Elementarteilchen ein Gas ionisieren, wodurch ein messbarer elektrischer Stromfluss entsteht. [7]

Nach dem inneren Teil des ATLAS-Detektors folgt das Kalorimeter, das aus zwei Systemen besteht: das innen liegende Flüssig-Argon-Kalorimeter (engl.: *Liquid Argon Calorimeter*, Abk.: LAr-Kalorimeter) und das außen liegende Hadronen-Kalorimeter (engl.: *Tile Hadronic Calorimeter*). Auf das Flüssig-Argon-Kalorimeter wird im Abschnitt 2.3 ausführlich eingegangen. In ihm werden die elektromagnetischen Elementarteilchen

vollständig absorbiert. Die hadronischen Elementarteilchen werden erst im Hadronen-Kalorimeter vollständig absorbiert, das aus sich abwechselnden Schichten von Stahl- und szintillierenden Plastikkacheln besteht. Die Elementarteilchen deponieren ihre gesamte Energie in den Stahlkacheln und erzeugen dadurch hadronische Schauer, die in dem Kalorimeter ein Szintillationslicht erzeugen, das dann in einen messbaren elektrischen Strom umgewandelt wird, dessen Intensität direkt proportional zur deponierten Energie ist. Mit Hilfe der Kachelstruktur kann daraus zusätzlich die Flugrichtung der Elementarteilchen bestimmt werden. [8]

Der ATLAS-Detektor schließt mit dem Myon-Spektrometer ab, welches aus vier Systemen besteht: *Thin Gap Chambers*, *Resistive Plate Chambers*, *Monitored Drift Tubes* und *Cathode Stripe Chambers*. In ihnen werden die Myonen und deren Impuls gemessen, da diese aufgrund ihrer hohen Masse fast unverändert die vorherigen Schichten des ATLAS-Detektors durchdringen. [9]

Das Magnetsystem, der vierte Bereich des ATLAS-Detektor, besteht aus drei Systemen: *Central Solenoid Magnet*, *Barrel Toroid* und *End-Cap Toroids*. Der *Central Solenoid Magnet* umschließt den Inneren Detektor und besteht aus supraleitenden Niob-Titan-Leitungen, mit denen er ein Magnetfeld zur Ablenkung der geladenen Elementarteilchen erzeugt. Das *Barrel Toroid* und die zwei *End-Caps Toroids* sind ebenso aus supraleitenden Leitungen aufgebaut und erzeugen ein Magnetfeld zur Ablenkung von Myonen, wobei die *End-Caps Toroids* vor allem die Teilchen ablenkt, deren Flugbahn nahe der Teilchenstrahlröhre verläuft. Durch die Ablenkung der Teilchen in den erzeugten Magnetfeldern können aufgrund der Lorentz-Kraft Rückschlüsse auf die Ladung der Elementarteilchen

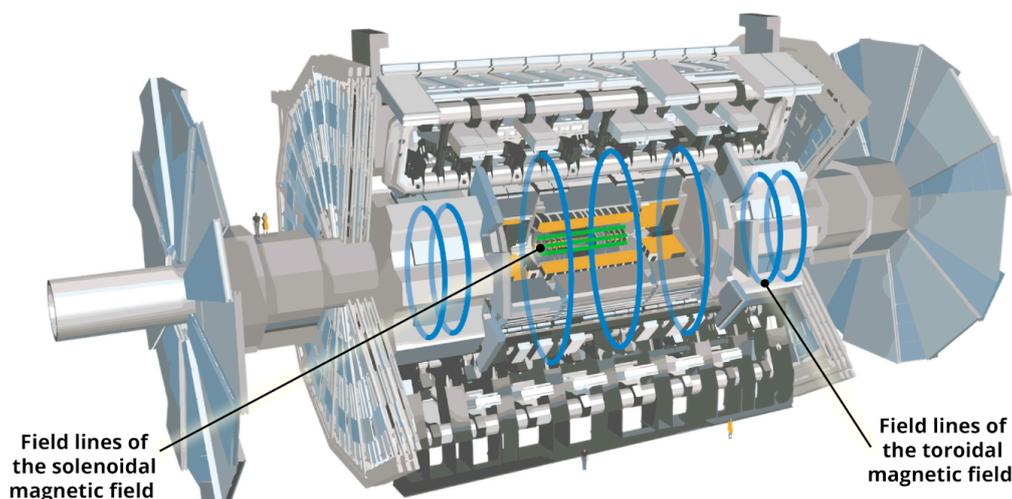


Abb. 4: Verlauf der durch das Magnetsystem erzeugten Feldlinien im ATLAS-Detektor [10]

und deren Impuls gezogen werden. Die erzeugten Feldlinien des Magnetsystems sind in Abb. 4 veranschaulicht. [10]

2.3. Das Flüssig-Argon-Kalorimeter

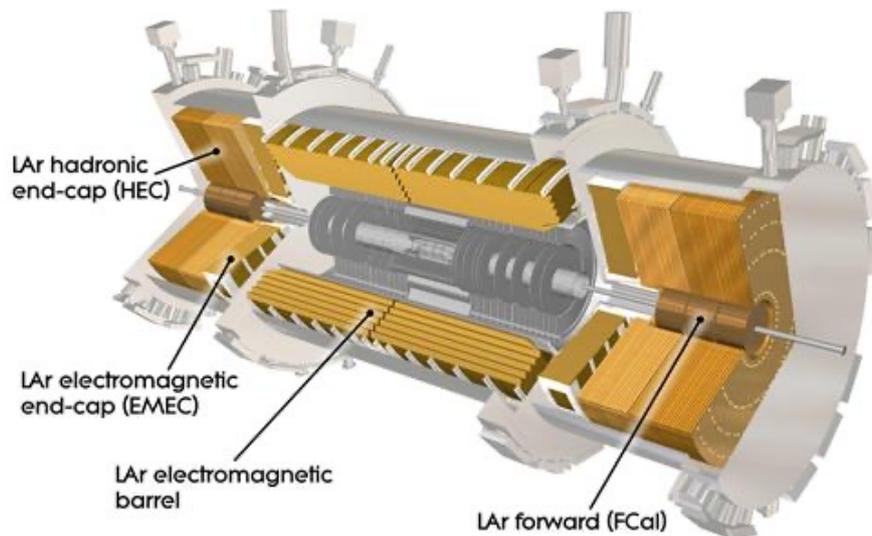


Abb. 5: Schematischer Aufbau des LAr-Kalorimeters im ATLAS-Detektor [13]

Das LAr-Kalorimeter ist ein elektromagnetisches Kalorimeter zur Detektion von elektrisch geladenen Teilchen. Das Ziel ist, aus deren Energiedepositionen die Flugbahn, die Ladung und den Impuls zu berechnen, wobei leichtere Teilchen, wie Photonen oder Elektronen, komplett absorbiert werden. Dazu werden verschiedene Detektoren eingesetzt, wie in Abb. 5 zu sehen ist. Die relevanten Detektoren für die betrachteten Untersuchungen sind das *LAr Electromagnetic Barrel* (Abk.: EMB) und die *LAr Electromagnetic End-Caps* (Abk.: EMEC).

Das EMB besteht aus zwei Zylinderringen, die zusammen mit dem *Solenoid Magnet* von einem Kryostat gekühlt werden. Das EMEC besteht aus zwei Rädern (engl.: *wheels*), die senkrecht zur Strahlenachse verlaufen und je eine Seite des EMB abdecken. Beide Räder bestehen ihrerseits aus einem inneren und einem äußerem Rad (engl.: *inner/outer wheel*, Abk.: IW/OW). Das EMEC wird zusammen mit dem *LAr hadronic end-cap* in einem eigenen Kryostat gekühlt.

Die elektromagnetischen Kalorimeter bestehen aus akkordeonförmig gebogenen Schichten von Kupfer-Kapton-Elektroden zwischen Blei-Absorptionsplatten, die durch wabenförmige Abstandshalter in Position gehalten werden. Die Zwischenräume sind mit

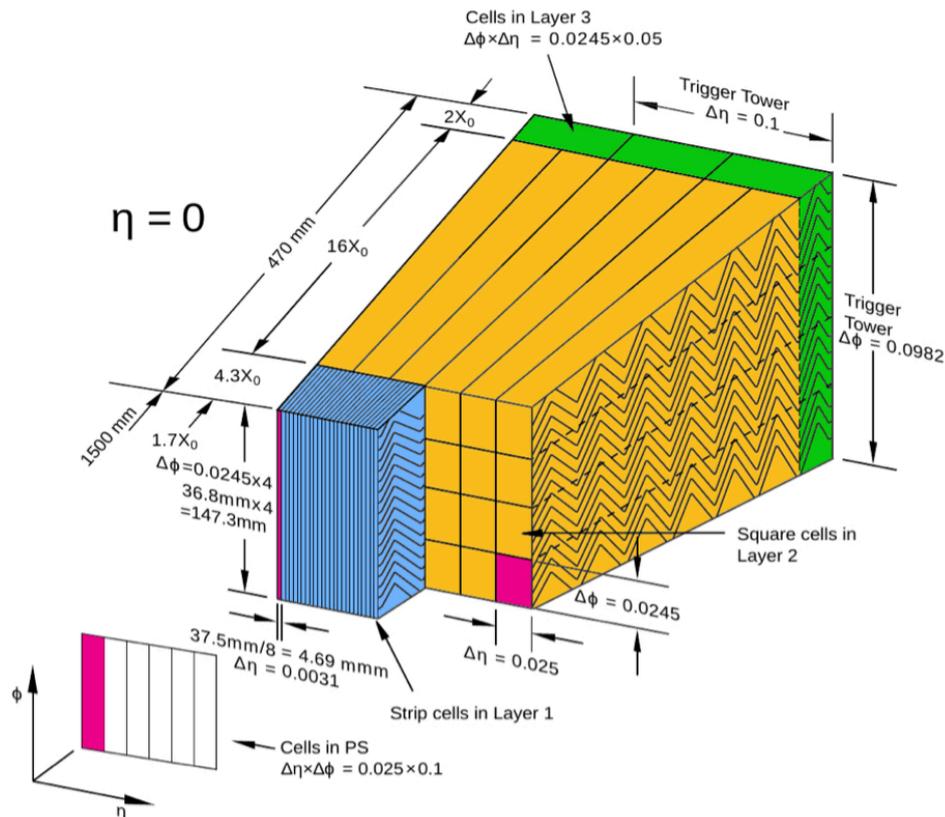


Abb. 6: Schematische Darstellung eines Moduls des Elektromagnetischen Kalorimeters, unterteilt in Schichten (eingefärbt) und Zellen. Schichten: Weiß: *presampler*, Blau: *front layer*, Gelb: *middle layer*, Grün: *back layer*. [14]

flüssigem Argon gefüllt, das für das Kalorimeter namensgebend ist und konstant auf einer Temperatur von -184 °C gehalten wird. Jedes Modul des EMB und des EMEC-OW besteht aus drei solcher Schichten von Zellen mit unterschiedlicher Granularität, während die Module des EMEC-IW nur aus zwei solchen Schichten besteht. [11, S. 8 f.; 12]

Die bei der Kollision entstandenen Elementarteilchen wechselwirken mit den Bleiplatten, wodurch sie an Energie verlieren und dabei Teilchenschauer erzeugen. Dabei ist die Größe und Energie des Teilchenschauers proportional zum Energieverlust des Elementarteilchens. Auf dem Weg zwischen den Bleiplatten ionisieren die elektrisch geladenen Elementarteilchen das flüssige Argon. Da zwischen den Absorptionsplatten und den Elektroden eine Hochspannung von 2 kV anliegt, entsteht auf diese Weise ein messbarer Ionisationsstrom mit einem Dreieckspuls, der ca. 450 ns bis zum vollständigen Abklingen andauert. Unter Berücksichtigung der Lage der ionisierten Zellen kann daraus der Ursprung, die Flugbahn und die Energie des Elementarteilchens berechnet werden.

Die für diese Berechnungen notwendige Elektronik besteht aus einer *Front-End*-Elektronik (Abk.: FEE) und einer *Back-End*-Elektronik (Abk.: BEE). Erstere ist direkt am Detektor platziert, während letztere außerhalb, in einem eigenen Raum ist.

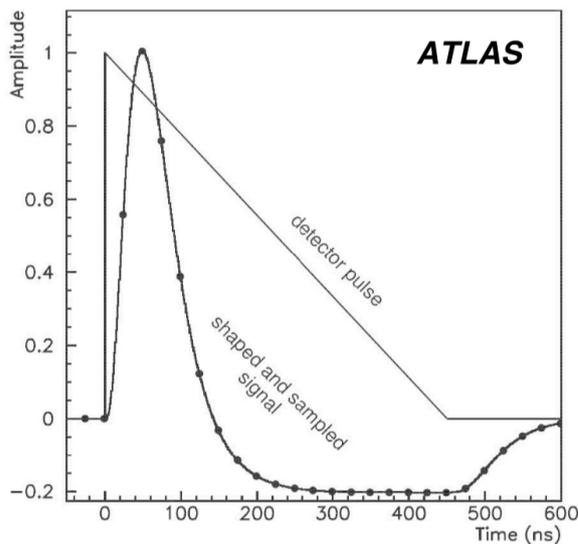


Abb. 7: Dreieckspuls des Ionisationsstromes im LAr-Kalorimeter und digitalisierte Pulsform. Die Punkte markieren die mögliche Lage der digitalisierten Messpunkte. [11]

Die FEE ist dafür zuständig, die gemessenen Ionisationsströme zu einem analogen Puls zu formen und diesen zu digitalisieren. Die erzeugte Pulsform (siehe Abb. 7) ist so gewählt, dass der Puls innerhalb von ca. 15 ns bis zum Peak ansteigt, schnell in den negativen Bereich abfällt und dann langsam auf Null steigt. Auf Grund dieses Verlaufes ist das Integral über den kompletten Pulsverlauf gleich Null. Dadurch wird die Überlagerung von nur positiven Impulsen vermieden und der Mittelwert aller überlagerten Pulse konstant gehalten. Diese Pulsüberlagerungen entstehen, da die Kollisionen alle 25 ns stattfinden, aber die erzeugten Ionisationsströme in den Detektorzellen bis zu 450 ns andauern, wodurch sich die Amplituden der einzelnen Pulse addieren. Dieser Effekt wird kurz als Pileup bezeichnet. Zur Reduzierung der produzierten Datenmengen erfolgt eine Digitalisierung der Signale mit einer Abtastrate, die der Kollisionsfrequenz des LHC entspricht. Die gewählte Pulsform gewährleistet auch den Informationsgehalt über die maximale Amplitude, weil die Digitalisierungsrate das Abtasten mehrerer Punkte des positiven Signalanteils erlaubt. Je höher diese Rate ist, desto besser lässt sich der positive Signalanteil digitalisieren, was auch eine bessere Energierekonstruktion erlaubt. Die digitalisierte Datensequenz wird zur weiteren Verarbeitung an die BEE gesendet. [11, S. 8 ff.]

In der BEE wird aktuell ein Optimalfilter (Abk.: OF) zur Energie- und Zeitrekonstruktion verwendet, der für beide Rekonstruktionen eine Linearkombination aus jeweils bis zu fünf aufeinanderfolgenden Werten der Datensequenz nutzt [11, S. 151]. Diese Werte werden dann durch Koeffizienten so gewichtet, dass je nach Bereich des Kalorimeters alle Störungen, wie Pileup oder elektronisches Rauschen, minimiert werden und die deponierte Energie anhand der Amplitude bestmöglich erkannt wird. Kurz gefasst werden die Gleichungen (1) und (2) benutzt:

$$A[n] = \sum_{i=0}^4 a_i \cdot w[n - i] \quad (1)$$

$$(A\tau)[n] = \sum_{i=0}^4 b_i \cdot w[n - i] \quad (2)$$

mit: Amplitude A des Pulses; Zeitpunkt τ des Ereignisses; Koeffizienten a_i der Energie- bzw. b_i der Zeitrekonstruktion; Datenpunkt n aus der Datensequenz w . [15, S. 474]

Zur Erkennung der Peaks in der Datensequenz wird zusätzlich zum OF der *Maximum Finder* benutzt. Dieser nicht-lineare Algorithmus wird nur im Trigger eingesetzt, da er mit einer Latenz von nur einem BC sehr schnell arbeitet. Ein Datenpunkt wird von dem Algorithmus dann als Peak erkannt, wenn er positiv und größer als sein vorheriger und sein nachfolgender Datenpunkt ist, wie in Gleichung (3) erkennbar ist [16, S. 56]:

$$y[n] = \begin{cases} x[n - 1] & \text{falls } x[n - 2] < x[n - 1] > x[n] \text{ und } x[n - 1] > 0, \\ 0 & \text{sonst.} \end{cases} \quad (3).$$

2.4.Maschinelles Lernen

Maschinelles Lernen ist ein Oberbegriff für Computer, die mit Algorithmen statistische Modelle zu Trainingsdaten aufbauen und diese gegenüber Testdaten optimieren. Anhand dieser Modelle lernen sie Muster bzw. Gesetzmäßigkeiten zu erkennen und wenden diese auf neue, unbekannte Daten an, um diese zu beurteilen. Sie ahmen im Prinzip den Lernprozess eines Menschen nach, mit dem Unterschied, dass ihr Lernen auf ein sehr eng gefasstes Gebiet begrenzt ist. Mögliche Anwendungen findet das maschinelle Lernen bei der Erkennung von Handschriften, Sprachen, Texten und Kreditkartenbetrug. Eine Methode des maschinellen Lernens ist das Deep Learning, bei dem künstliche neuronale Netze zur Informationsverarbeitung eingesetzt werden, die zwischen einer Eingabe- und Ausgabeschicht mehrere unsichtbare Schichten haben.

2.4.1.Künstliche neuronale Netze

Künstliche neuronale Netze (engl.: *artificial neural network*, Abk.: NN) sind nach dem Vorbild des menschlichen Gehirns geschaffen. In einem Graphen werden die Neuronen

als Knoten und deren Verknüpfung durch Kanten symbolisiert werden. Typischerweise werden die Kanten durch Pfeile dargestellt, deren Richtung angibt, welche Eingabewerte ein Knoten benutzt und von welchen Knoten sein Ausgabewert benutzt wird. In den Knoten wird durch eine mathematische Funktion aus den Eingabewerten ein Ausgabewert berechnet. Dieser Ausgabewert kann in anderen Knoten zu weiteren Berechnung herangezogen werden. So können in einem NN mehrere Schichten (engl.: *layers*) entstehen.

Um in jeder Schichten eine flexible Verarbeitung der Eingabewerte in den Knoten zu gewährleisten, werden nicht-lineare Aktivierungsfunktionen benutzt. Zwei verbreitete Aktivierungsfunktionen sind die rektifizierte lineare Einheit (engl.: *rectified linear unit*, Abk.: ReLU; Gl. 4) und die Sigmoidfunktion (Gl. 5).

$$g_{\text{ReLU}}(x) = \begin{cases} x & \text{falls } x \geq 0, \\ 0 & \text{sonst.} \end{cases} \quad (4)$$

$$g_{\text{Sigmoid}}(x) = \frac{1}{1 + \exp(-x)} \quad (5)$$

Die ReLU-Funktion ist einfach zu implementieren, da sie den Eingabewert übernimmt, wenn dieser positiv ist, und sonst den Wert Null verwendet. Die Sigmoid-Funktion eignet sich für NN die binäre Entscheidungen treffen sollen, wie z. B. bei der Klassifizierung von Daten, da ihre Funktionswerte im Intervall zwischen Null und Eins liegen. [16]

Hier die mathematisch formale Zusammenfassung eines NN, wie sie auch in Abb. 8 beispielhaft aufgegriffen wird:

Ein NN ist eine Funktion

$$f_{\text{NN}}: \mathbb{R}^M \rightarrow \mathbb{R}^N \quad (6)$$

mit M Eingabe- und N Ausgabewerten, die eine Verkettung von D -Schichten mit den Schichtindexen $i = 1, \dots, D$ und den schichtabhängigen Parametern $A_i \in \mathbb{R}^{m_i \times n_i}$ (Gewichtsmatrix) und $b_i \in \mathbb{R}^{n_i}$ (Bias) ist, wobei die Schichten wiederum durch mehrdimensionale Funktionen

$$f_i(x): \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}, x \mapsto g_i(A_i x + b_i) \quad (7)$$

beschrieben werden, die durch die Aktivierungsfunktion g_i in den Knoten einer Schicht gesteuert werden. Aus den Gl. 6 und 7 folgt die Gl. 8.

$$\begin{aligned} f_{\text{NN}}(x) &= (f_D \circ f_{D-1} \circ \dots \circ f_2 \circ f_1)(x) \\ &= f_D (f_{D-1} (\dots (f_2 (f_1(x)))))) \end{aligned} \quad (8)$$

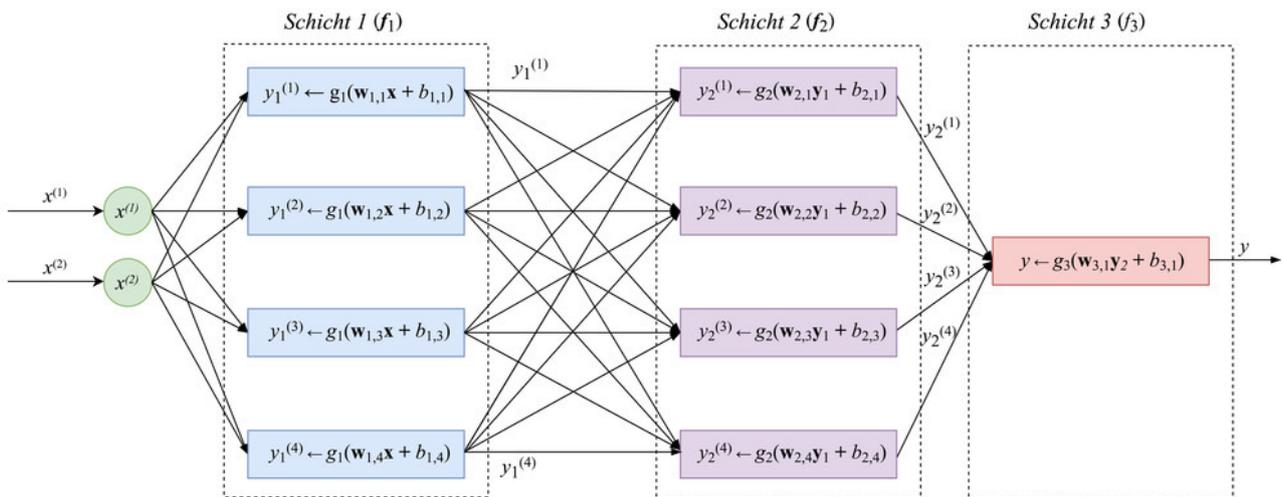


Abb. 8: Darstellung eines künstlichen NN mit drei Schichten, einem zwei-dimensionalen Eingangsvektor und einem skalaren Ausgangsvektor. [17, Abb. 6.1]

Zwei spezielle Typen von NN sind rekurrente NN (Abk.: RNN) und Faltungsnetze (engl.: *Convolutional NN*, Abk.: CNN). Bei RNN's werden Datensequenzen in einzelne Abschnitte unterteilt, die nacheinander eingelesen und verarbeitet werden. Das Besondere bei diesen Netzen ist, dass Zwischenergebnisse aus einem vorangegangenen Abschnitt in die Berechnungen des aktuellen Abschnitts einfließen. So bilden sich an den Knoten Schleifen, die eine Art Gedächtnis darstellen und durch Aktivierungsfunktionen gesteuert werden, damit nur bestimmte Informationen behalten werden. Damit sind RNN's für kurze Datensequenzen gut geeignet, da die Muster für jeden Abschnitt individuell erkannt und genutzt werden.

Im Gegensatz zum RNN werden CNN's für eine kontinuierliche Verarbeitung der Datensequenzen eingesetzt. Dabei werden kleine Bereiche der Eingabewerte auf einmal eingelesen und verarbeitet, solange bis die komplette Sequenz abgearbeitet wurde. Die Größe des eingelesenen Bereichs wird durch den Kern (engl.: *kernel*) festgelegt. Ein Vorteil dieses Vorgehens ist, dass die Anzahl an Rechenoperationen sinkt, weil die

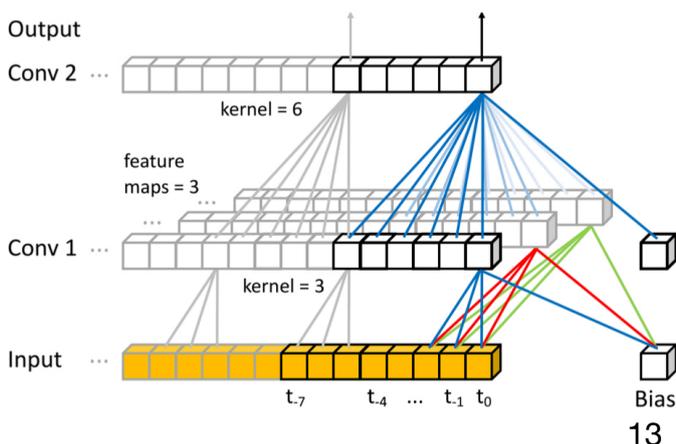


Abb. 9: Darstellung eines CNN mit zwei Schichten: Conv 1 (*kernel* = 3, *feature maps* = 3) und Conv 2 (*kernel* = 6, *feature maps* = 1). [14, Abb. 11]

Dimensionen der Eingabewerte, der Gewichtsmatrix und des Bias (siehe Gl. 6) reduziert werden. Ein weiterer Vorteil ist, dass innerhalb einer Schicht der Kern gleichzeitig durch verschiedene Funktionen im CNN verarbeitet werden kann, um so verschiedene Informationen gleichzeitig zu erhalten. Dieses Vorgehen wird durch den Begriff *feature maps* ausgedrückt. Diese Informationen fließen später in weiteren Rechenschritten des CNN wieder ein. [16; 17, Kap. 6.2]

2.4.2. Training von neuronalen Netzen

Der Lernprozess des NN besteht in der Verarbeitung der Eingabedaten, der Bewertung der erhaltenen Rechenergebnisse und der selbstständigen Veränderung des Bearbeitungsprozesses aufgrund der Bewertung, damit die nachfolgenden Bewertungen besser werden. Die Kategorisierung des Lernprozesses erfolgt in über-, teilüber- und unüberwachtes sowie bestärkendes Lernen. Im Rahmen dieser Arbeit wurde nur überwachtes Lernen angewendet, weshalb dieses im Folgenden erläutert wird.

Bei einem überwachten Lernprozess werden für das Training Testdatenpaare $(x_i | y_i)$ aus den Eingangsdaten und den zugehörigen, bekannten Ergebnissen vorgegeben. Mit dem CNN werden aus und zu den Eingangsdaten die Trainingsdaten $f_{\text{NN}}(x)$ berechnet, wobei die Parameter in Gl. 6 für die Berechnungen vorgegeben oder zufällig initialisiert werden. Anschließend erfolgt ein Vergleich der Testdatenpaare mit den Trainingsdatenpaaren durch eine Verlustfunktion (engl.: *loss function*). Ihr Ergebnis gibt an, wie groß die Abweichung zwischen den erhaltenen und den gewünschten Ergebnissen ist. Für die Verlustfunktion können beispielsweise die Euklidische Norm oder die mittlere quadratische Abweichung (engl.: *mean square error*, Abk.: MSE) verwendet werden, aber auch eigens für das Problem speziell angepasste Funktionen sind möglich. Je größer der Verlust ist, desto mehr werden die Parameter des CNN verändert. Dabei soll der Verlust langfristig minimiert werden. Damit das Auffinden des Minimums nicht zu viele Ressourcen verbraucht werden Abbruchkriterien eingeführt. Mögliche Kriterien sind bspw. das Unterschreiten eines Zielwertes der Verlustfunktion oder das Überschreiten einer Anzahl an Trainingsdurchläufen, welche auch Epochen (engl.: *epochs*) genannt werden. Eine Epoche ist gekennzeichnet durch das einmalige Verwenden aller Test- und Trainingsdaten, einschließlich deren Vergleich und der abschließenden Parameteränderungen. [17, Kap. 1.2 und 1.3; 18, Kap. 5.1]

3. Detektorspezifische Energierekonstruktion im LAr-Kalorimeter

In vorangegangenen Arbeiten wurde der Einsatz von NN's auf einer beispielhaften Zelle des LAr-Kalorimeters entwickelt. Im nun folgenden Teil dieser Arbeit wird der Einsatz auf verschiedenen Zellen untersucht, um festzustellen, ob die bisherigen Erkenntnisse nur zellengebunden sind oder sich verallgemeinern lassen. Dabei wird angenommen, dass die CNN's auf jeder Zelle unterschiedlich gut funktionieren. Dies ist auf die Lage und den Aufbau der Zellen zurückzuführen, wodurch in jeder Zelle u. a. die angelegten Spannungen, die erzeugten Ionisations- und Driftströme und auch die Messungenauigkeiten voneinander verschieden sind. Andererseits ist die Annahme auch aus der Sicht legitimiert, dass CNN's sich bei ihrem Training stark auf Besonderheiten der untersuchten Zelle und des zum Training bereitgestellten Datensatzes spezialisieren können.

Für diese Untersuchungen wird in diesem Kapitel zuerst die Struktur der eingesetzten CNN's erläutert, die bisher für die Energierekonstruktionen benutzt wurden. Anschließend werden die Zellen vorgestellt, die als Beispiel für alle Zellen der untersuchten Regionen des LAr-Kalorimeter stehen. Da technisch bedingt die Pulsform in jeder dieser Zellen unterschiedlich ist, wird auf diese als nächstes eingegangen, bevor sich die Arbeit den Ergebnissen des Trainings der CNN's auf den jeweiligen Regionen zuwendet. Der Schwerpunkt liegt dabei auf der Betrachtung der zweidimensionalen Histogramme, welche die Energierekonstruktion über der Zeit zwischen zwei Teilchenpaketen der CNN's mit einer Farbskala abbilden. Punkte in der Farbe gelb haben die meisten Events, wohingegen die in lila die wenigsten haben.

3.1.Nutzung von Faltungsnetzen für die Energierekonstruktion

Auf Basis von Simulationen wurde gezeigt, dass die Verwendung von NN's für die Energierekonstruktion am ATLAS-Detektor langfristig in Betracht zu ziehen ist (siehe [19]). Dies liegt u. a. daran, dass die untersuchten NN's gegenüber dem derzeit eingesetzten OF die Signale besser vom Hintergrundrauschen trennen, die Energien beim Pileup mit besserer Auflösung rekonstruieren und bei den ermittelten transversalen Energien einen kleineren Bias haben. Dabei wurden, neben anderen Netzen, CNN's mit drei bzw. vier

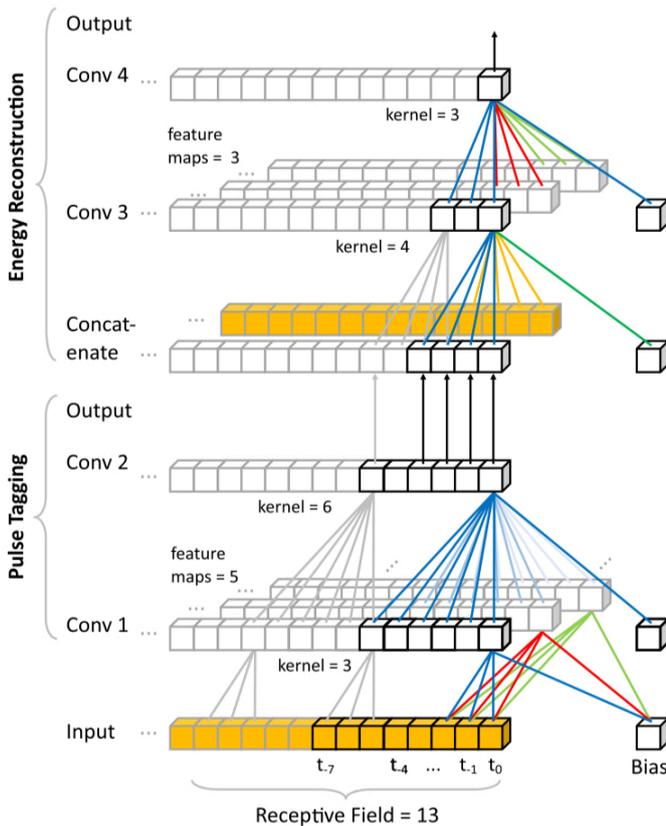


Abb. 10: Struktur des 4Conv-Netzes bestehend aus je zwei Schichten für die Pulserkennung und die Energierekonstruktion. Die Datenverarbeitung erfolgt von unten nach oben. Für die Energierekonstruktion werden sowohl die Eingangsdaten als auch die Ergebnisse aus der Pulserkennung verwendet, weshalb ein *Concatenate-Layer* eingefügt ist. [19, Fig. 2]

Schichten verwendet (Abk.: 3Conv- bzw. 4Conv-Netz) [19]. Diese werden im Folgenden vorgestellt, da diese beiden Netze die bisher besten Ergebnisse lieferten und die nachfolgenden getroffenen Erkenntnisse auf diesen beiden Netzen aufbauen.

Das 3Conv- und das 4Conv-Netz benutzen für die Pulserkennung (engl.: *pulse tagging*) die gleiche Netzstruktur, während die Netzstruktur für die Energierekonstruktion (engl.: *energy reconstruction*) bei Ersterem aus einer Schicht und bei Letzterem aus zwei Schichten besteht. Die Struktur des 4Conv-Netzes ist in Abb. 10 dargestellt. Beide Netze arbeiten nach dem gleichen Prinzip. Zuerst wird aus den Eingangsdaten der Zeitpunkt des Entstehens des Ionisationspuls in der Zelle berechnet. Dafür ist die zweischichtige Pulserkennung da, die bei beiden Netzen identisch ist und die Sigmoidfunktion (Gl. 5) zur Aktivierung benutzt. Anschließend werden die Ergebnisse der Pulserkennung mit den Eingangsdaten im *Concatenate-Layer* verknüpft, bevor daraus in ein bzw. zwei Schichten die transversalen Energien rekonstruiert werden. Für diese Berechnungen wird die ReLU-Funktion (Gl. 4) zur Aktivierung benutzt. Als Ergebnis der Berechnungen geben beide Netze nur die rekonstruierte transversale Energie zu einem Zeitpunkt aus.

Die Kennwerte der beiden Netze, also die Anzahl an Schichten und *feature maps*, die Kerngröße, die Anzahl an Parametern u. a. sind in Tab. 1 gelistet. Aus dieser geht hervor,

Tab. 1: Konfigurationen des 3Conv- und 4Conv-Netzes, die die gleiche Struktur zur Pulserkennung und zur Energierekonstruktion eine Struktur aus einer bzw. zwei Schichten haben. [19]

	3-conv			4-conv			
	Tagging		Energy Recon- struction	Tagging		Energy Recon- struction	
Layer index	1	2	3	1	2	3	4
Kernel Size	3	6	21	3	6	4	3
Dilation Rate	1	1	1	1	1	1	1
Feature Maps	5	1	1	5	1	3	1
Activation Function	Sigmoid		ReLU	Sigmoid		ReLU	
Number of Parameters	51		43	51		37	
MAC units	45		42	45		33	
Receptive Field	28			13			

dass aus den gewählten Strukturen für das 3Conv-Netz 94 Parameter und für das 4Conv-Netz 88 Parameter trainiert werden müssen. Die Parameter bilden zusammen mit den *Multiplier-Accumulator Units* (Abk.: MAC-Units) einen Indikator für den Ressourcenverbrauch bei den zur Berechnung eingesetzten *Field Programmable Gateway Arrays* (Abk.: FPGA). Die FPGA ermöglichen eine parallele Verarbeitung der am ATLAS-Detektor erwarteten Datenmengen und bieten den Vorteil, dass auch nach ihrer Produktion ihre Schaltungsstruktur am Einsatzort den Gegebenheiten angepasst werden kann. Der Wert des in der Tabelle aufgeführten *Receptive Field* gibt dabei an, wieviele digitalisierte Eingangsdaten für die Berechnung eines Ausgabewerts berücksichtigt werden. [19]

Die Struktur der Netze erfordert es, dass erst die Parameter der Pulserkennung und danach die der Energierekonstruktion trainiert werden. Für das überwachte Training der Netze wird die *ATLAS Readout Electronics Upgrade Simulation* (Abk. AREUS) verwendet. Diese ermöglicht die Simulation des kompletten Datenverarbeitungsprozesses von der einzelnen Zelle bis hin zur Elektronik. Simuliert werden u. a. die Signalumwandlung von analog zu digital, die bitgenaue Signalverarbeitung, Pileup und Hintergrundrauschen. Die Simulationen mit unterschiedlichen Signalbedingungen können für unterschiedliche Trainings verwendet werden, womit auch die Möglichkeit besteht, diverse Bewertungsmethoden und veränderte NN's auf dem gleichen Datensatz zu testen. [20]

Neben den Parametern gibt es auch noch Hyperparameter, wie z. B. die Lernrate (engl.: *learning rate*). Mit den Hyperparametern wird das Trainingsverhalten des NN beeinflusst. Im Spezifischen reguliert die Lernrate die Änderung der Trainingsparameter nach dem Durchlauf einer Epoche. Dazu wird der Fehlergradient berechnet, der anschließend mit der Lernrate multipliziert wird, um die Größe der Parameteränderungen zu erhalten. Kleine Lernraten führen somit zu kleinen Änderungen der Trainingsparameter und damit zu mehr Epochen, um ein Minimum der multidimensionalen Verlustfunktion zu erhalten. Bei großen Lernraten besteht die Gefahr, dass die Parameter nie so angepasst werden, dass ein Minimum erreicht werden kann. [21]

3.2. Betrachtete Detektorregionen

Tab. 2: Übersicht über die untersuchten Detektorregionen

Region	Kürzel	Pseudorapidität η	Azimutwinkel ϕ	Rauschen 3σ in GeV
EMB Front Layer	F05	0,5015625	0,05	0,03894
EMB Front Layer	F10	1,0015625	0,05	0,029184
EMB Middle Layer	M05	0,5125	0,0125	0,234
EMB Middle Layer	M10	1,0125	0,0125	0,2064
EMB Presampler	PS	0,0125	0,05	0,14307
Outer EMEC Middle Layer	M20	2,0125	0,0125	0,159

Mit AREUS wurden Datensätze für sechs Zellen des LAr-Kalorimeters aus unterschiedlichen Detektorregionen erzeugt. Die Auswahl der Zellen erfolgte nach der Geometrie und der Lage im LAr-Detektor. Die Lage der Zellen wird spezifiziert durch die Pseudorapidität und den Azimutwinkel. Dazu wird ein rechtshändiges Koordinatensystem mit dem Ursprung im Zentrum des Detektors, dem nominellen Kollisionspunkt, gewählt. Die x-Achse zeigt zum Zentrum des LHC-Rings, die y-Achse in Richtung der Erdoberfläche und die z-Achse in Richtung des Teilchenstrahls. In der transversalen Ebene werden Zylinderkoordinaten (r, ϕ) benutzt und der Azimutwinkel wird um die z-Achse herum gemessen. Die Pseudorapidität ist abhängig vom Polarwinkel und berechnet sich nach Gl. 9.

$$\eta = -\ln \tan(\theta/2) \quad (9)$$

Anhand dieser Definition werden mit dem in Abschnitt 2.3 vorgestellten Aufbau des LAr-Kalorimeters Teilchen innerhalb einer Pseudorapidität von $|\eta| < 4,9$ detektiert [11, S. 7]. Die Lage der untersuchten Zellen, deren Hintergrundrauschen und die im Folgenden für die Zellen verwendeten Abkürzungen sind Tab. 2 zu entnehmen. Für die Auswertung der Datensätze wurden nur Energien bis max. 5 GeV berücksichtigt, während bei den Datensätzen zu den Trainings wahre Energien bis max. 30 GeV simuliert wurden.

3.3. Performanz der CNN's ohne Training auf den Zellen

In den bisherigen Untersuchungen wurden die CNN auf der Zelle M05 trainiert [19]. Die so trainierten 3Conv- und 4Conv-Netze wurden als Erstes ohne ein weiteres Training zur Evaluation auf die anderen fünf Zellen angewendet. Dabei wurde für jede Zelle der jeweilige Wert des Hintergrundrauschens (siehe Tab. 2) als Energieschwelle zur Puls-erkennung und Energierekonstruktion angewendet.

Aus der Breite der 98%-Intervallen in Abb. 11 leitet sich ab, dass die noch untrainierten CNN's mit den bisherigen Parametern am besten auf den Zellen M05, M10 und M20 arbeiten. Dies wird am identischen Aufbau der Zellen liegen, welche alle aus dem Middle-Layer sind. Tendenziell lässt sich beim 3Conv-Netz eine bessere Energierekonstruktion bei einer größeren Pseudorapidität ableiten. Da die Zellen des Middle-Layer alle ungefähr in der gleichen relativen Eindringtiefe der Schauer liegen, kann die Pseudorapidität nicht der finale Grund sein. Ein Vergleich des Rauschens zwischen den drei Zellen zeigt, dass dieses mit steigender Rapidität abnimmt, so dass ein direkter Zusammenhang zwischen dem elektrischen Rauschen und der Energierekonstruktion vorliegen könnte. Die breitesten Verteilungen haben die Zellen F05, F10 und PS, wobei hier mit steigender Rapidität ein stärkeres Rauschen in den Zellen und eine tendenziell schlechte Energierekonstruktion der CNN's vorliegt. Bei dem 4Conv-Netz lassen sich diese Aussagen nur bedingt aus den Intervallbreiten ableiten. Dies könnte u. a. daran liegen, dass es die Energien insgesamt schlechter als das 3Conv-Netz rekonstruiert, wie es auch im Zusammenhang mit der gröberen Skaleneinteilung in Abb. 11 zu sehen ist.

Bei den Intervallen der Mittelwerte unter Einbezug der Standardabweichung sind größere Unterschiede zu erkennen als bei den Medianen mit 98%-Intervall. Dabei lassen sich die Zellen wieder in die Gruppe M05, M10, M20 und die Gruppe F05, F10, PS aufteilen. Die geringeren Abweichungen zwischen den Mittelwert- und den Median-Intervallen bei der

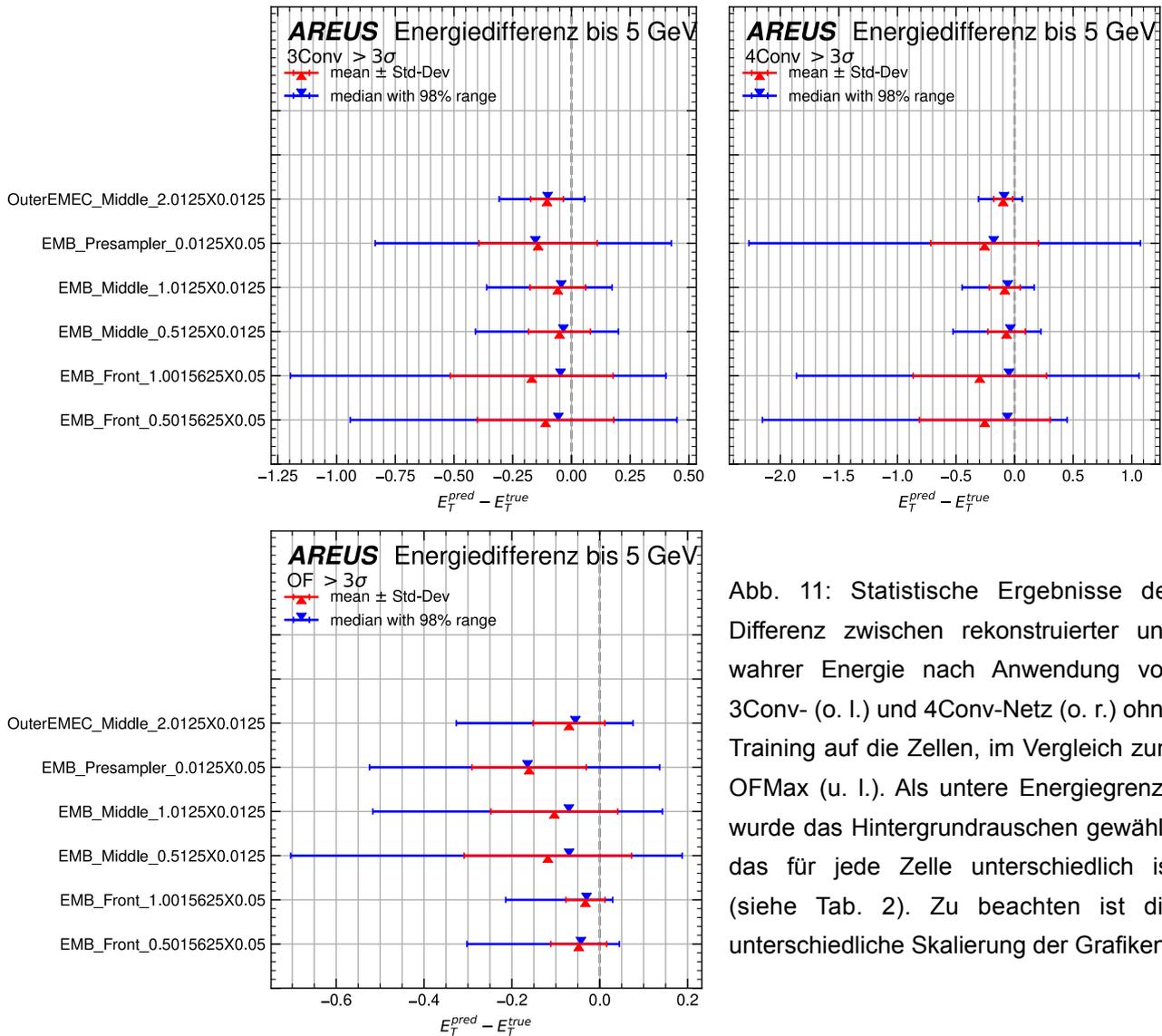


Abb. 11: Statistische Ergebnisse der Differenz zwischen rekonstruierter und wahrer Energie nach Anwendung von 3Conv- (o. l.) und 4Conv-Netz (o. r.) ohne Training auf die Zellen, im Vergleich zum OFMax (u. l.). Als untere Energiegrenze wurde das Hintergrundrauschen gewählt, das für jede Zelle unterschiedlich ist (siehe Tab. 2). Zu beachten ist die unterschiedliche Skalierung der Grafiken.

ersten Gruppe zeigen, dass diese die Energien besser rekonstruiert als die zweite Gruppe. Außerdem liegen auch deren Mittelwerte allesamt näher bei Null als bei der zweiten Gruppe. Dies wird durch die Abb. 12 bestätigt, in der die Verteilung der Differenzen zwischen der wahren Energie, die größer 1 GeV sein soll, und deren dazugehörigen rekonstruierten Energien dargestellt ist. Das 3Conv-Netz hat eine bessere Performanz als das 4Conv-Netz, weil es schmalere und höhere Verteilungen hat, allerdings sind beide CNN bei den Zellen F05, F10 und PS deutlich schlechter als der OFMax, der besonders gute Ergebnisse auf den Zellen F05 und F10 liefert.

Es wurde versucht die erhaltenen Unterschiede auf die Pulsform zurückzuführen, die in jeder Zelle verschieden ist. Allerdings ist kein allgemeiner Zusammenhang zwischen der Statistik der Energierekonstruktionen in Abb. 11 und der auf die deponierte Energie

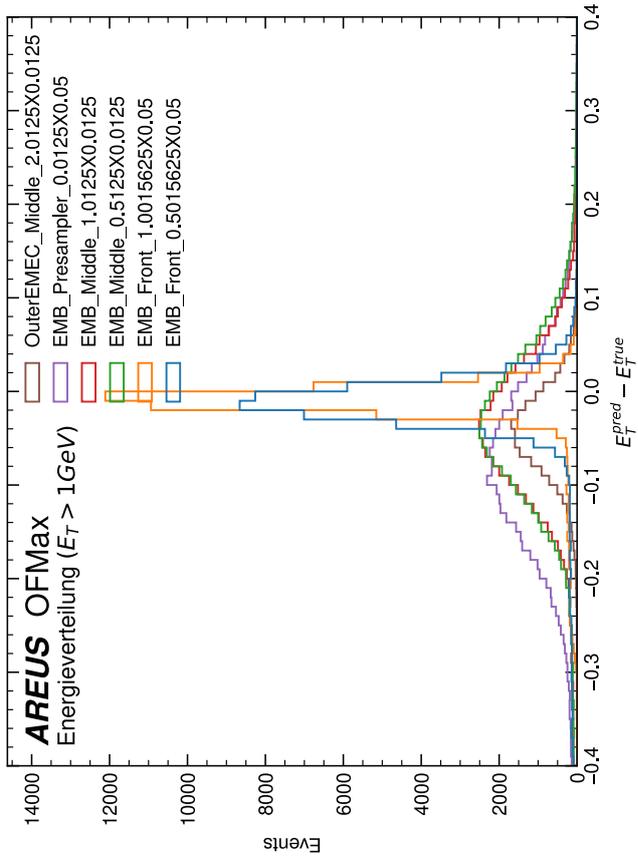
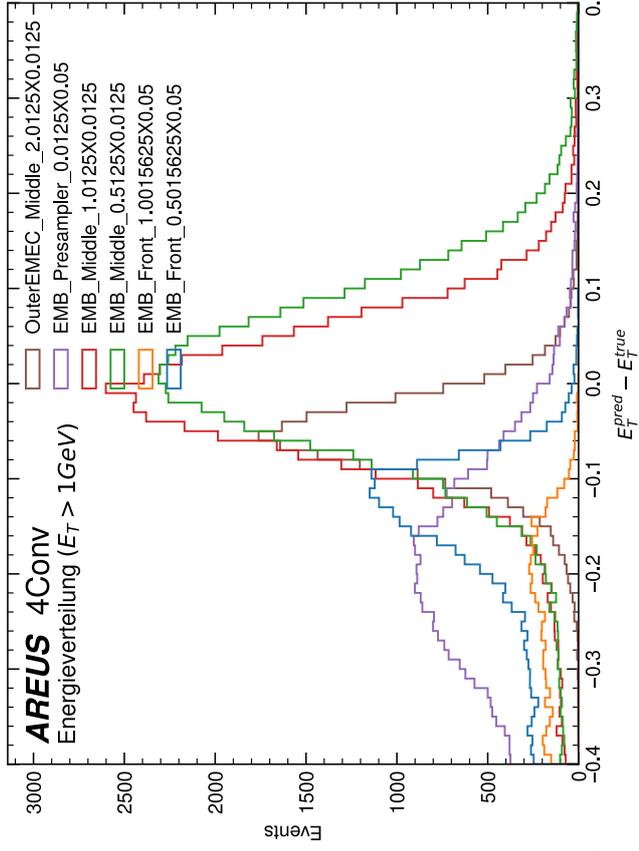
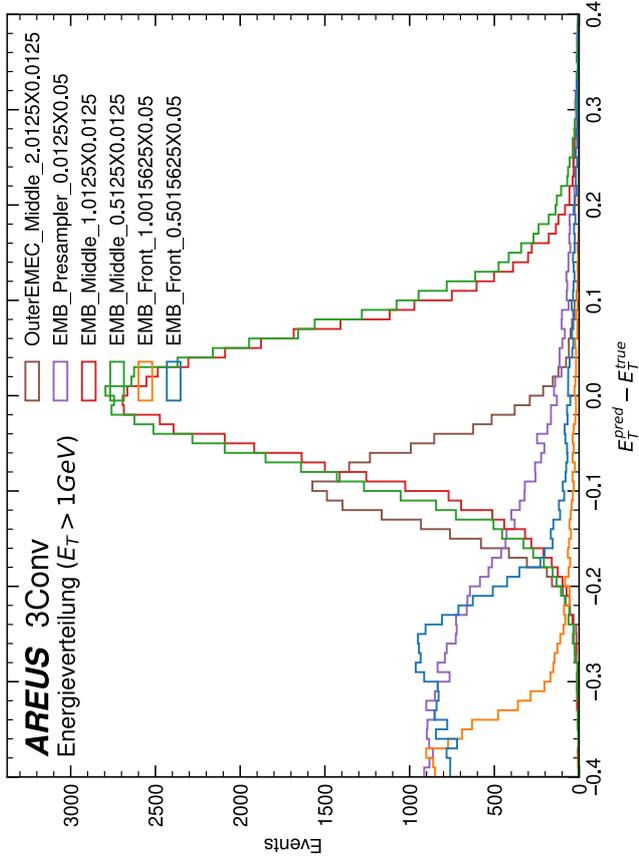


Abb. 12: Verteilung der Energiedifferenzen in GeV nach Anwendung von 3Conv- (o. l.) und 4Conv-Netz (o. r.) ohne Training auf die Zellen, im Vergleich zum OFMax (u. l.). Es sind nur die Differenzen dargestellt, bei denen die wahre Energie größer 1 GeV ist.

normierten Pulsform aus Abb. 13 zu erkennen. Aus den Pulsformen lässt sich die bereits festgestellte starke Ähnlichkeit zwischen den Zellen M05 und M10 nochmal bestätigen, da deren Verlauf fast identisch ist. Die Pulsformen der anderen Zellen sind zueinander stark

ähnlich, wobei die Unterschiede im Bereich des Peaks auf die Breite der Bins zurückzuführen sind. Dass die Ergebnisse der Netze bei der Zelle M20 von den Zellen F05, F10 und PS abweicht liegt an den baulich bedingten Unterschieden.

Aus der Anwendung der CNN's ohne Training auf die Zellen lässt sich zusammenfassend ableiten, dass beide CNN auf den Zellen M05, M10 und M20 zueinander relativ ähnliche Ergebnisse erzeugen, ebenso wie auf den Zellen F05, F10 und PS. Trotzdem müssen die Netze auf jeder Zelle einzeln trainiert werden, insbesondere auf den letztgenannten drei Zellen, da sich alle Zellen in den Pulsformen und ihrer jeweiligen Lage in Bezug auf den Kollisionspunkt unterscheiden.

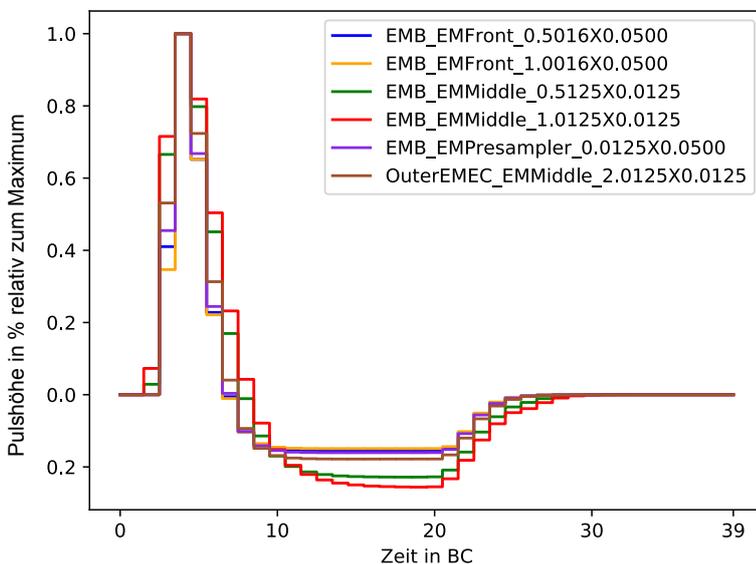


Abb. 13: Pulsformen der sechs Zellen, normiert auf den Anteil zum Maximum.

3.4. Performanz der CNN's mit Training auf den Zellen

Im nachfolgenden Ansatz wurden die CNN's auf allen Datensätzen neu trainiert, außer auf dem der Zelle M05, da diese bereits in anderen Arbeiten ausführlich untersucht wurde. Dabei wurden vor den Trainings die Parameter für die Zelle M05 in die CNN's geladen, bevor diese auf den anderen Regionen neu trainiert wurden. Anschließend wurde der Einfluss der Anzahl an Epochen, der Lernrate und der Energieschwelle untersucht. Der Ausgangspunkt für diese Untersuchungen war, dass bei den Zellen F05, F10 und PS für die anfänglich erstellten 2D-Histogrammen die Zeitintervalle zwischen zwei Pulsen insgesamt zu klein waren, obwohl der mittlere Abstand zwischen den Signalen 25 BC betragen sollte. (siehe Abb. 14).

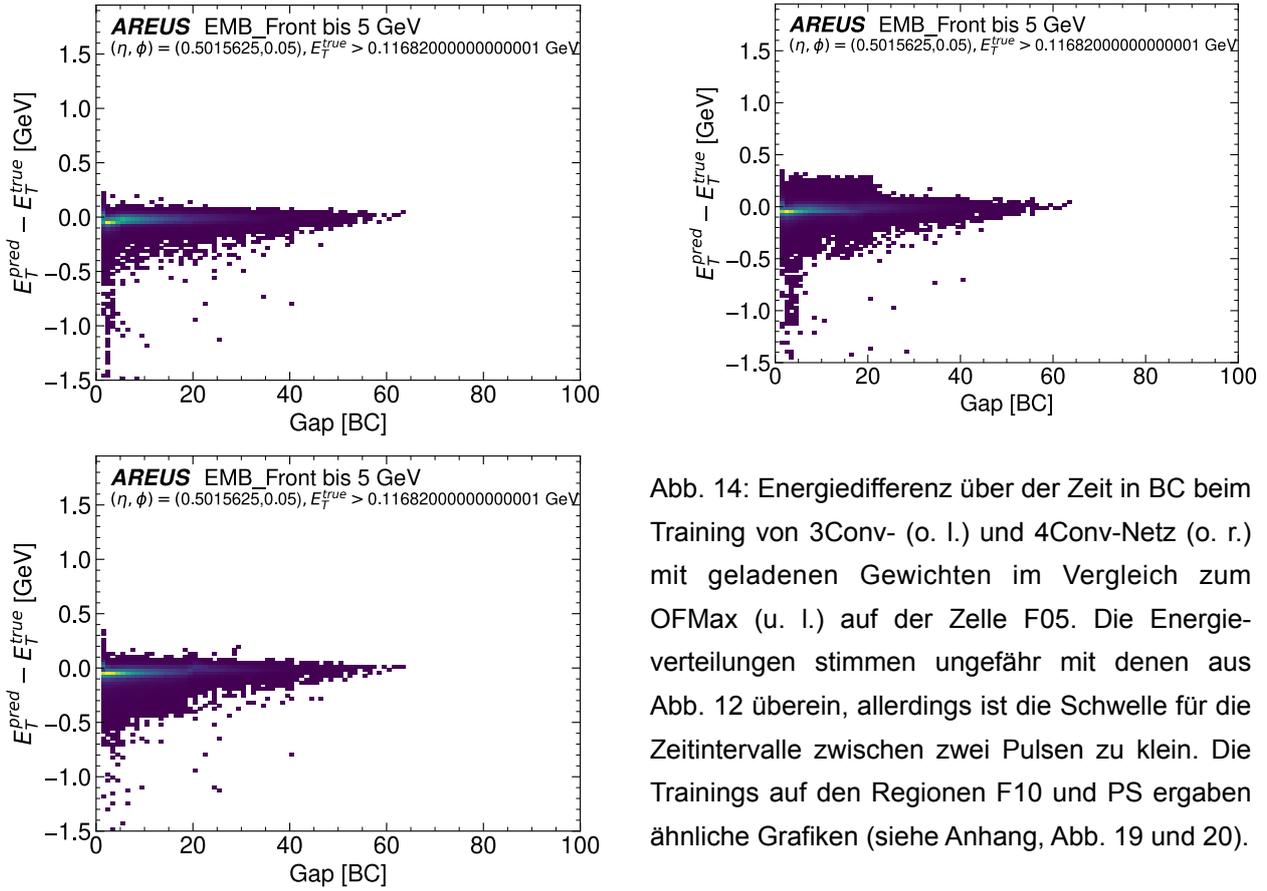


Abb. 14: Energiedifferenz über der Zeit in BC beim Training von 3Conv- (o. l.) und 4Conv-Netz (o. r.) mit geladenen Gewichten im Vergleich zum OFMax (u. l.) auf der Zelle F05. Die Energieverteilungen stimmen ungefähr mit denen aus Abb. 12 überein, allerdings ist die Schwelle für die Zeitintervalle zwischen zwei Pulsen zu klein. Die Trainings auf den Regionen F10 und PS ergaben ähnliche Grafiken (siehe Anhang, Abb. 19 und 20).

3.4.1. Einfluss der Anzahl an Epochen und der Lernrate

Bei den Trainings wurden die Ergebnisse von 200, 300 bzw. 400 Epochen untersucht. Dazu wurde sich die Verlustrate angesehen, die Werte zwischen Null und Eins annehmen kann und auf vier Nachkommastellen genau angegeben wurde. Bei den ersten 100 Epochen gab es geringfügige Änderung im Bereich von 10^{-3} bis 10^{-4} und bei den nächsten 100 Epochen waren diese Änderungen kleiner als 10^{-4} . Daher wurde bei allen weiteren Trainings die Anzahl von Epochen auf 200 gesetzt, da hier das Verhältnis zwischen dem Rechenaufwand, der bei steigender Epochenanzahl steigt, und der Anzahl an Epochen den größten Nutzen brachte.

Bei den Lernraten wurden die Werte $5 \cdot 10^{-3}$, $5 \cdot 10^{-4}$, $5 \cdot 10^{-5}$ bis $5 \cdot 10^{-6}$ untersucht. Die Anzahl von 200 Epochen wurde beibehalten. Für die Zellen F05, F10 und PS ergab sich dabei keine Verbesserung bzgl. der Zeitintervalle zwischen zwei Pulsen, weil die Signalabstände für die meisten Energien immer noch unter 25 BC lagen (siehe Anhang, Abb. 21 und 22).

Für die Zellen M10 und M20 konnten kleine Veränderungen beobachtet werden (Abb. 15 bzw. Anhang, Abb. 21 und 22). Diese Veränderungen waren bei beiden CNN von gleicher Art. Bei der Zelle M10 liefert eine Lernrate von $5 \cdot 10^{-4}$ die besten Ergebnisse, weil sich hier die meisten Energien um den angezielten Punkt bei 25 BC und 0 GeV Energiedifferenz konzentrieren, weshalb diese Lernrate bei den Trainings beibehalten wurde (siehe Abb. 15). Für die anderen Zellen wird angenommen, dass die Lernrate von $5 \cdot 10^{-4}$ zusammen mit 200 Epochen ebenso die besten Resultate erzielt, weil die Lernrate und die Epochen Hyperparameter sind, die von der Netzstruktur und deren Parameter nicht beeinflusst wird.

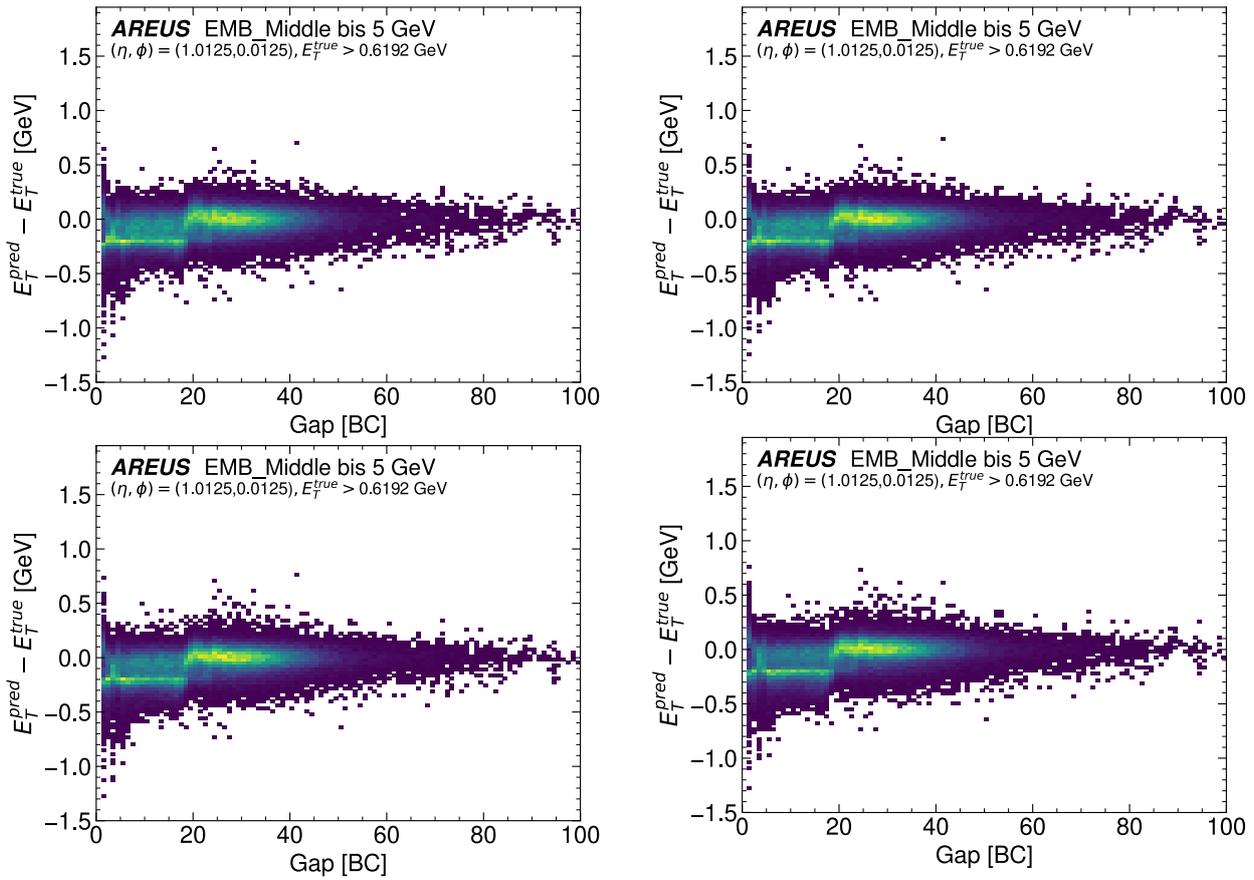


Abb. 15: Energiedifferenz über der Zeit in BC auf der Zelle M10 nach dem Training mit geladenen Gewichten auf dem 4Conv-Netz zu den Lernraten $5 \cdot 10^{-3}$ (o. l.), $5 \cdot 10^{-4}$ (o. r.), $5 \cdot 10^{-5}$ (u. l.) und $5 \cdot 10^{-6}$ (u. r.). Zu betrachten sind die Helligkeitsunterschiede in den gelben Regionen.

3.4.2. Einfluss der Energieschwelle auf die Pulserkennung

Für die Zelle M05, wird der Wert von drei Sigma, der hier 240 MeV beträgt, als Energieschwelle für das Rauschen sowohl bei der Pulserkennung als auch bei der Energie-rekonstruktion angewendet. Dieser Ansatz wurde sowohl bei den Trainings der anderen Zellen als auch beim Erstellen der 2D-Histogramme übernommen, wobei sie auf das

Rauschen in der jeweiligen Zelle (Tab. 2, S. 18) angepasst wurde. Dies führte jedoch bei den Zellen F05, F10 und PS zu Widersprüchen in den 2D-Histogrammen, weil die meisten Energien nicht um den erwarteten Punkt von 25 BC und 0 GeV Energiedifferenz lagen. Da aber die Energieverteilungen vorläufig akzeptabel waren und nur die Darstellung bzgl. der Zeit nicht funktionierte, wurde für letztere für die Darstellung die Energieschwelle auf 0,5 bzw. 1,0 GeV hochgesetzt und die Diagramme neu gezeichnet. Auch bei den Zellen M10 und M20 wurden die Diagramme erneuert, um einen optimalen Schwellenwert auf den Bereich um drei Sigma, um 0,5 GeV oder um 1,0 GeV einzugrenzen.

Wie in Abb. 16 beispielhaft für die Zelle F10 gezeigt ist, werden die Pulse aufgrund der höheren Energieschwelle besser den Zeiten in BC zugeordnet. Wird die Schwelle zu hoch gesetzt, sinkt die Anzahl an dargestellten rekonstruierten Energien, wie an den Skalen zu erkennen ist. Dies trifft auf allen Zellen zu (siehe Anhang, Abb. 23). Deshalb ist der Schwellenwert von drei Sigma für die Zellen M10 und M20 besser als die höheren

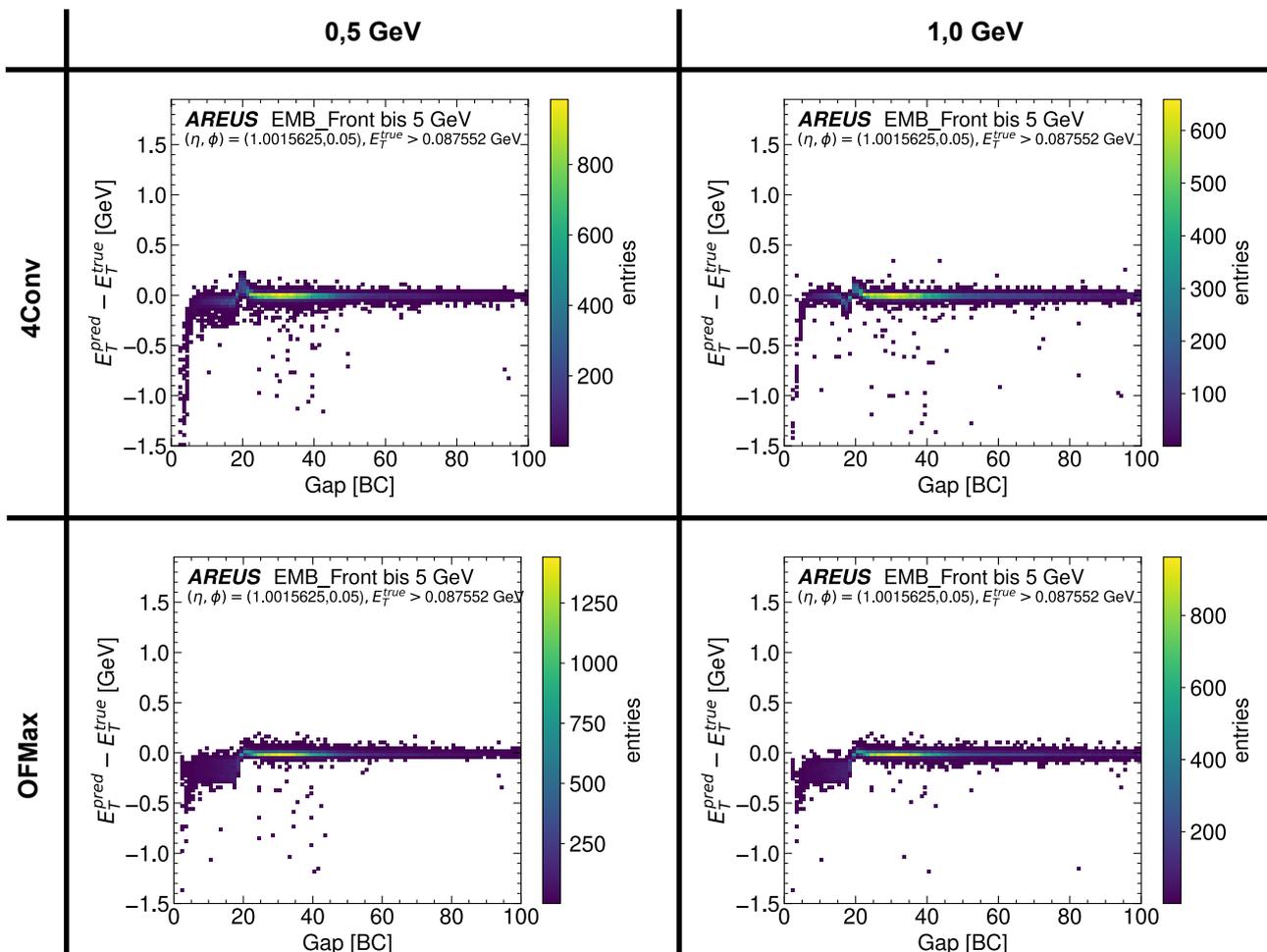


Abb. 16: Energiedifferenz über der Zeit auf der Zelle F10 nach dem Training mit geladenen Gewichten mit den Energieschwellen 0,5 GeV und 1,0 GeV für die Pulserkennung auf dem 4Conv-Netz im Vergleich zum OFMax

Schwellenwerte. An den Skalen ist ebenso zu sehen, dass der OFMax weiterhin die meisten Energien rekonstruiert. Im Vergleich zu dem 4Conv-Netz liegen diese weiter vom Ziel von 0 GeV für die Energie-differenz entfernt. Bei allen CNN und dem OFMax sind im Bereich um 20 BC Unregel-mäßigkeiten zu erkennen, die aufgrund des elektrischen Rauschens entstehen. Das Energien bei Zeiten über 60 BC rekonstruiert werden liegt am Pileup.

Aus den relativen Energieverteilungen wird deutlich, dass beide CNN auf den Zellen M10, M20 und PS vergleichbare oder sogar bessere Ergebnisse liefern als der OFMax. Insbesondere ist das 3Conv-Netz besser als das 4Conv-Netz, auch bei den Zellen, bei denen beide Netze schlechter als der OFMax sind. Dies zeigt sich daran, dass das 3Conv- gegenüber dem 4Conv-Netz höhere, schmalere Peaks und kürzere Flanken hat, wie es beispielhaft in Abb. 17 gezeigt wird (weitere Grafiken im Anhang, Abb. 22).

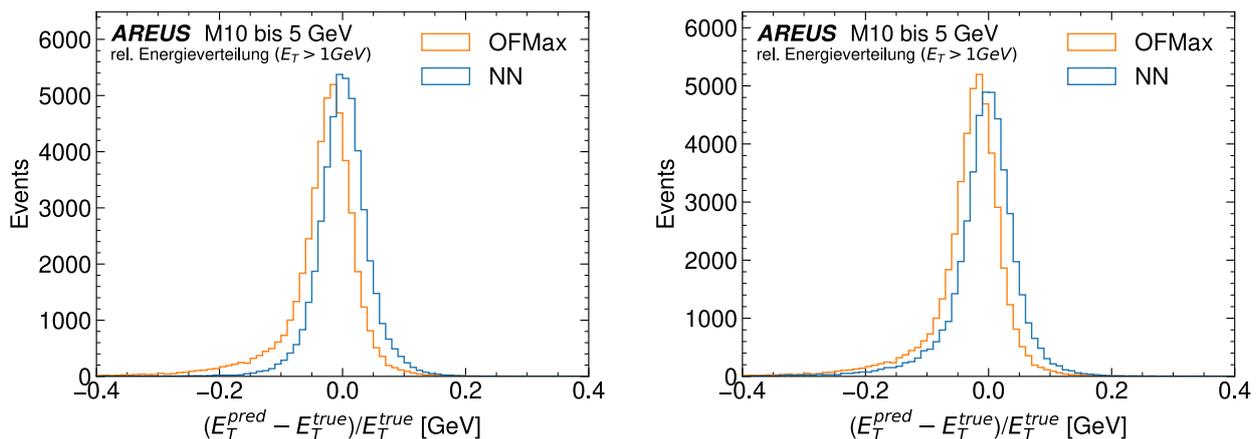


Abb. 17: Energiedifferenz normiert auf die wahre Energie der Zelle M10 nach dem Training mit geladenen Gewichten vom 3Conv- (links) und 4Conv-Netz (rechts) im direkten Vergleich zum OFMax.

3.4.3. Einfluss statistischer Unsicherheiten

Anhand der Zelle M10 wurde bei den Trainings beispielhaft getestet, ob zwei Trainings bei den gleichen Startbedingungen zu verschiedenen Parametern bzw. Ergebnissen kommen können. Für die beiden Trainings wurden das 4Conv-Netz, 200 Epochen, eine Lernrate von $5 \cdot 10^{-4}$ und eine Energieschwelle von 1,0 GeV für die Pulserkennung gewählt. In den in Abb. 18 dargestellten Vergleich sind mehrere Unterschiede auffindbar, vor allem in den Randbereichen der Verteilung, wobei einer beispielhaft hervorgehoben wurde. Aus dem Vorhandensein von Unterschieden lässt sich der Einfluss von statistischen Unsicherheiten

auf die Parameter während der Trainings nachweisen. Auch wenn die in Abb. 18 dargestellten Ergebnisse sehr ähnlich sind, kann trotzdem schätzungsweise ein Training von 100 durch eine besonders gute Verteilung hervorstechen. Einerseits liegt das darin begründet, dass die Verlustfunktion multidimensional ist und damit das Ermitteln eines lokalen Minimums viel einfacher ist als das eines globalen Minimums. Andererseits ist zu berücksichtigen, dass bei den numerischen Rechnungen die Ergebnisse nicht immer exakt dargestellt werden können, wodurch es zu Rundungsfehlern kommen kann, die sich bei vielen Rechenoperationen aufsummieren können. Der Grund dafür ist die begrenzte Anzahl an Bits für die Mantisse einer jeden Zahl.

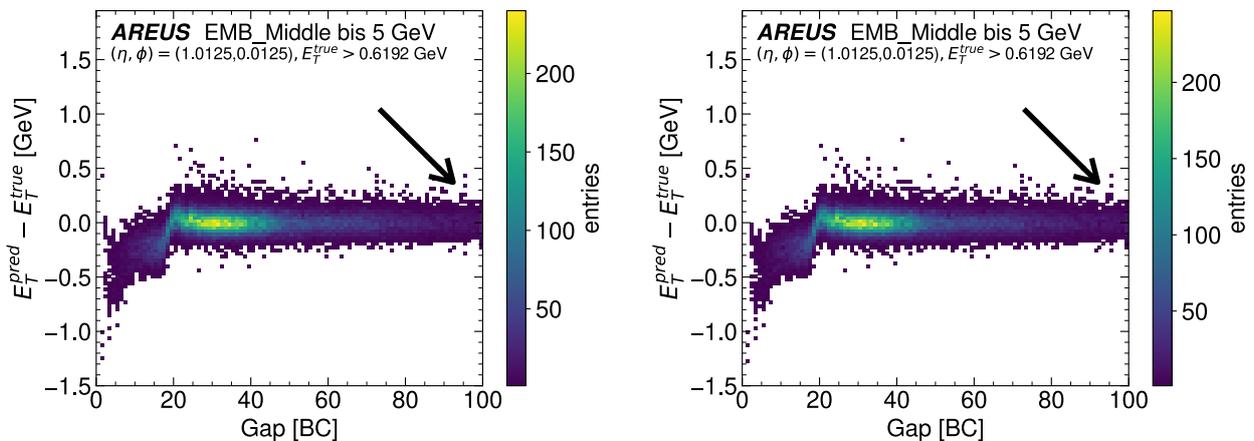


Abb. 18: Energiedifferenz über der Zeit in BC der Zelle M10 nach zwei Trainings auf dem 4Conv-Netz mit gleichen Ausgangsbedingungen zum Test auf statistische Unsicherheiten. Mit den Pfeilen wird ein Unterschied beispielhaft hervorgehoben.

4. Fazit und Ausblick

Die Anwendung des 3Conv- und 4Conv-Netzes liefert auch auf anderen Zellen des LAr-Kalorimeters brauchbare Ergebnisse, wenn die Parameter aus einer anderen Region vorher geladen und dann neu trainiert werden. Durch dieses Vorgehen werden die Trainingszeiten verkürzt, da bereits mit Startwerten begonnen wird, die potentiell nahe eines Minimums der neuen Verlustfunktion liegen. Die Ergebnisse haben auch gezeigt, dass die CNN's auf den Middle-Layern die Energien genauso gut oder besser rekonstruieren als der OFMax. Auf den Front-Layern sind die Ergebnisse der CNN's bisher schlechter als beim OFMax, sollten aber durch weitere Trainings besser werden können. Ein weiterer Vorteil dieser Methode ist, dass die bisherigen Netzwerkstrukturen beibehalten werden können.

Für zukünftige Anwendungen der CNN's ist zu erwarten, dass es für jede Zelle besser ist die Parameter aus einer anderen, baulich ähnlichen Zelle zu laden, weil dann die Unterschiede zwischen den Startwerten und den neu trainierten Werten geringer sind und das Training schneller gegen eine optimale Lösung konvergiert. Allerdings wurden bisher nur Daten bis 30 GeV für die Trainings betrachtet. In weiteren Arbeiten sollten in den Trainings die Einflüsse von höheren Energien und statistischen Unsicherheiten auf die Ergebnisse eingehend untersucht werden.

5. Anhang

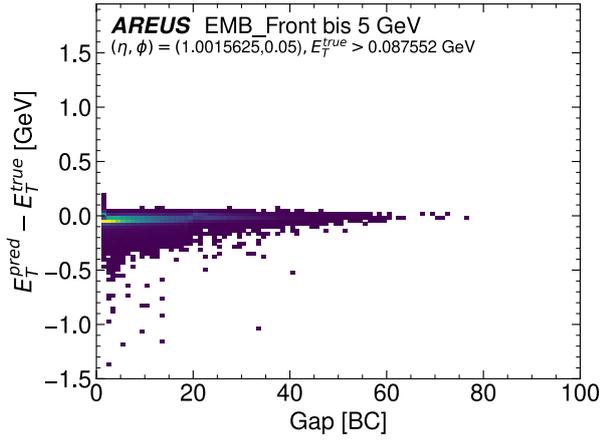
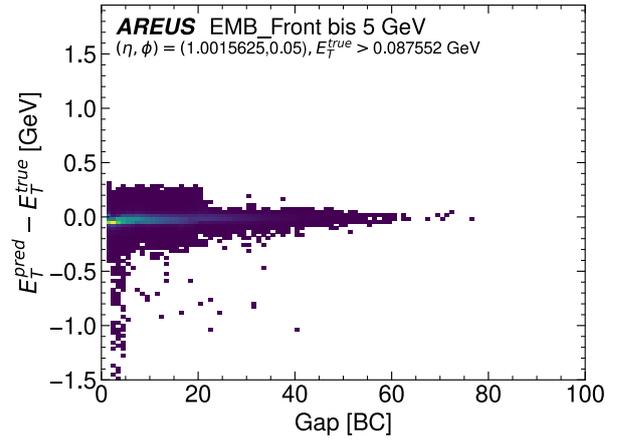
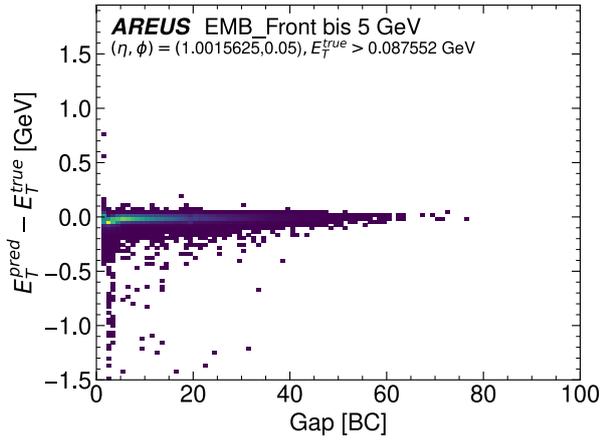


Abb. 19: Energiedifferenz über der Zeit in BC beim Training von 3Conv- (o. l.) und 4Conv-Netz (o. r.) mit geladenen Gewichten im Vergleich zum OFMax (u. l.) auf der Zelle F10.
 (Anhang zu Abschnitt 3.4, Abb. 14)

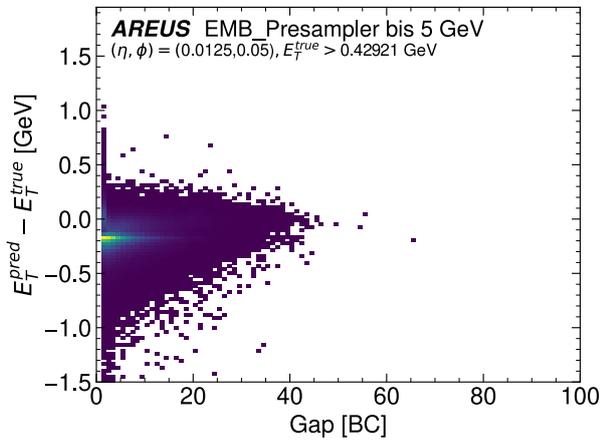
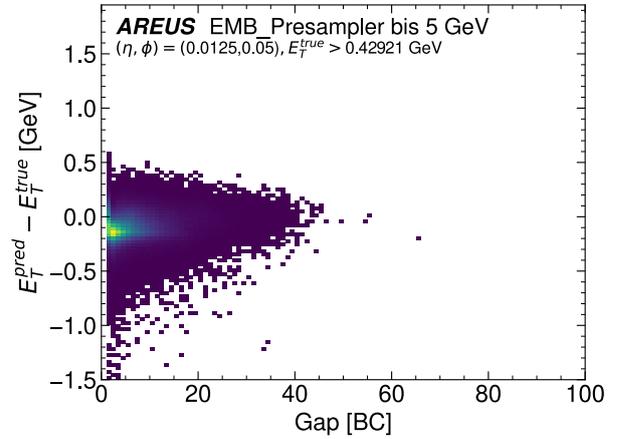
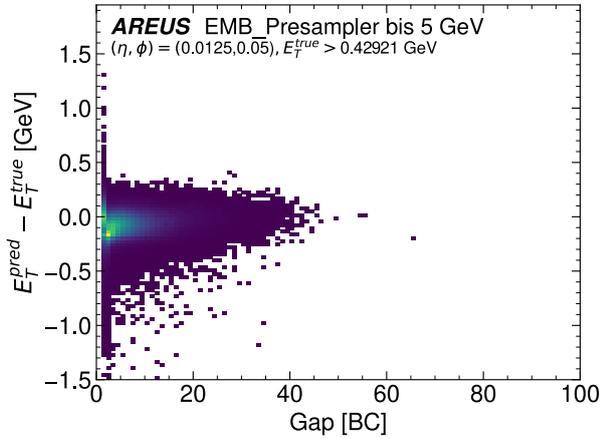


Abb. 20: Energiedifferenz über der Zeit in BC beim Training von 3Conv- (o. l.) und 4Conv-Netz (o. r.) mit geladenen Gewichten im Vergleich zum OFMax (u. l.) auf der Zelle PS.
 (Anhang zu Abschnitt 3.4, Abb. 14)

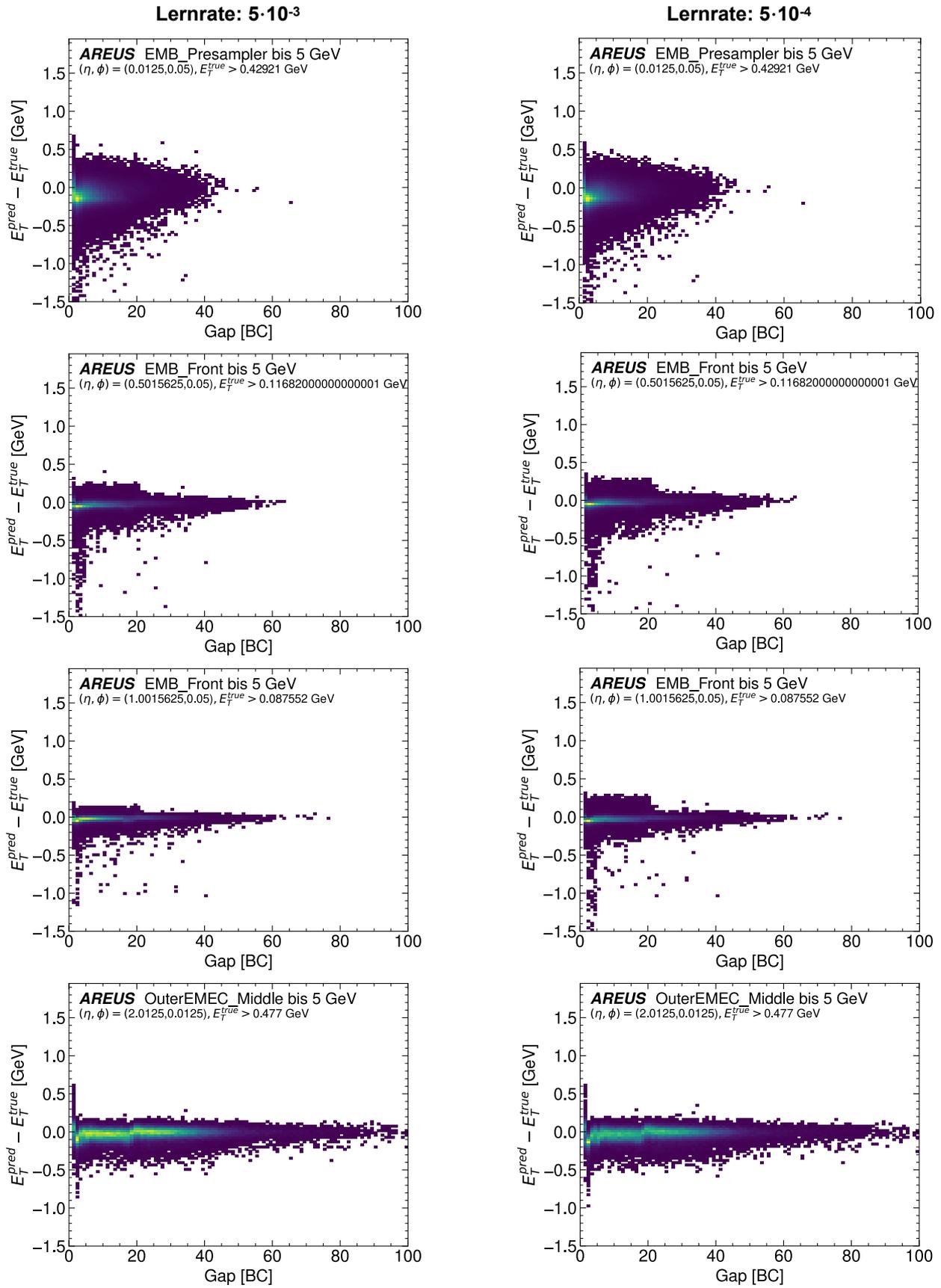


Abb. 21: Energiedifferenz über der Zeit in BC auf den Zellen PS, F05, F10 und M20 (v. o. n. u.) nach dem Training mit geladenen Gewichten auf dem 4Conv-Netz zu den Lernraten $5 \cdot 10^{-3}$ (links) und $5 \cdot 10^{-4}$ (rechts). Das 3Conv-Netz ergibt ähnliche Grafiken. (Anhang zu Abschnitt 3.4.1, Abb. 15)

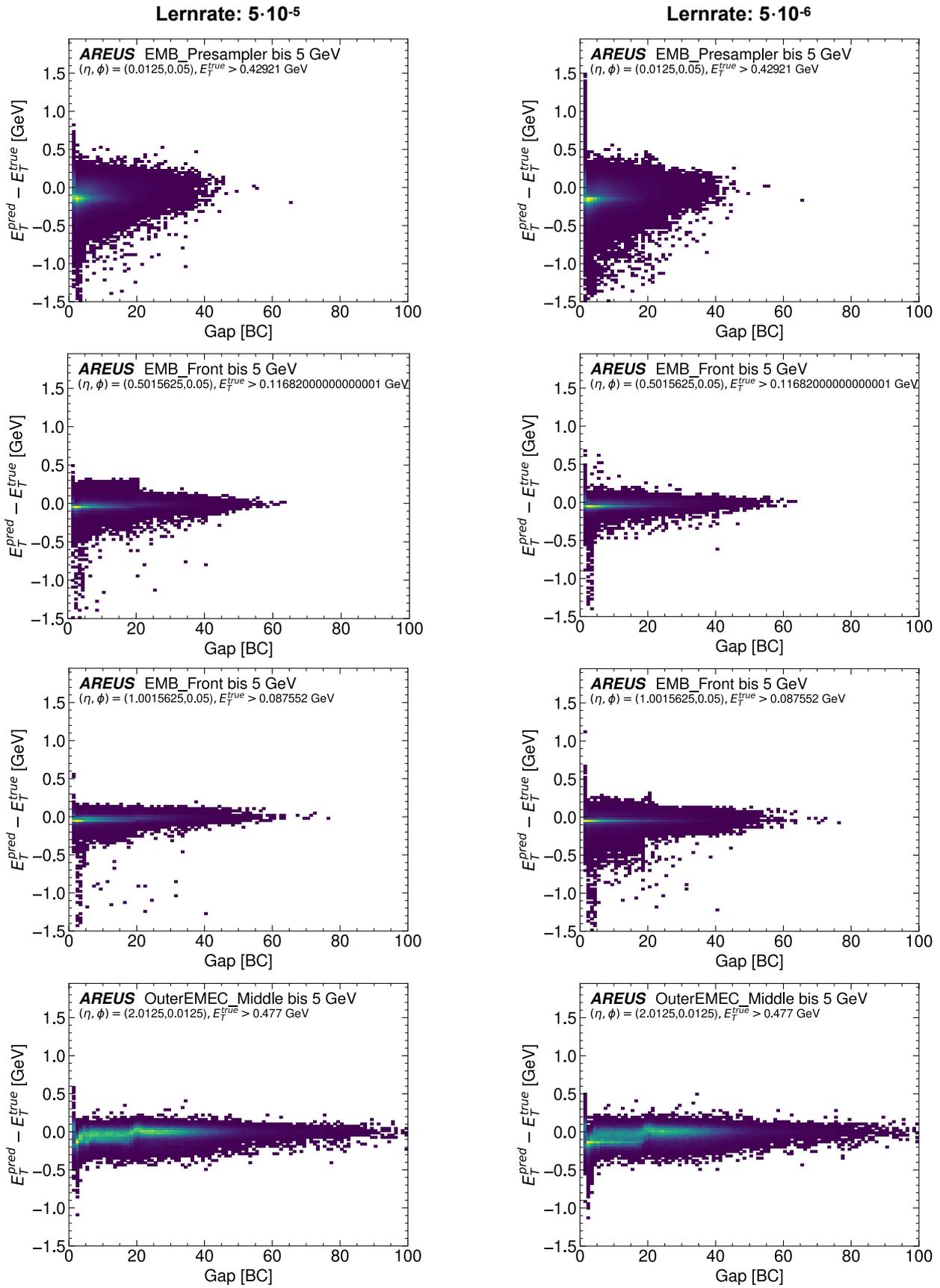


Abb. 22: Energiedifferenz über der Zeit in BC auf den Zellen PS, F05, F10 und M20 (v. o. n. u.) nach dem Training mit geladenen Gewichten auf dem 4Conv-Netz zu den Lernraten $5 \cdot 10^{-4}$ (links) und $5 \cdot 10^{-6}$ (rechts). Das 3Conv-Netz ergibt ähnliche Grafiken. (Anhang zu Abschnitt 3.4.1, Abb. 15)

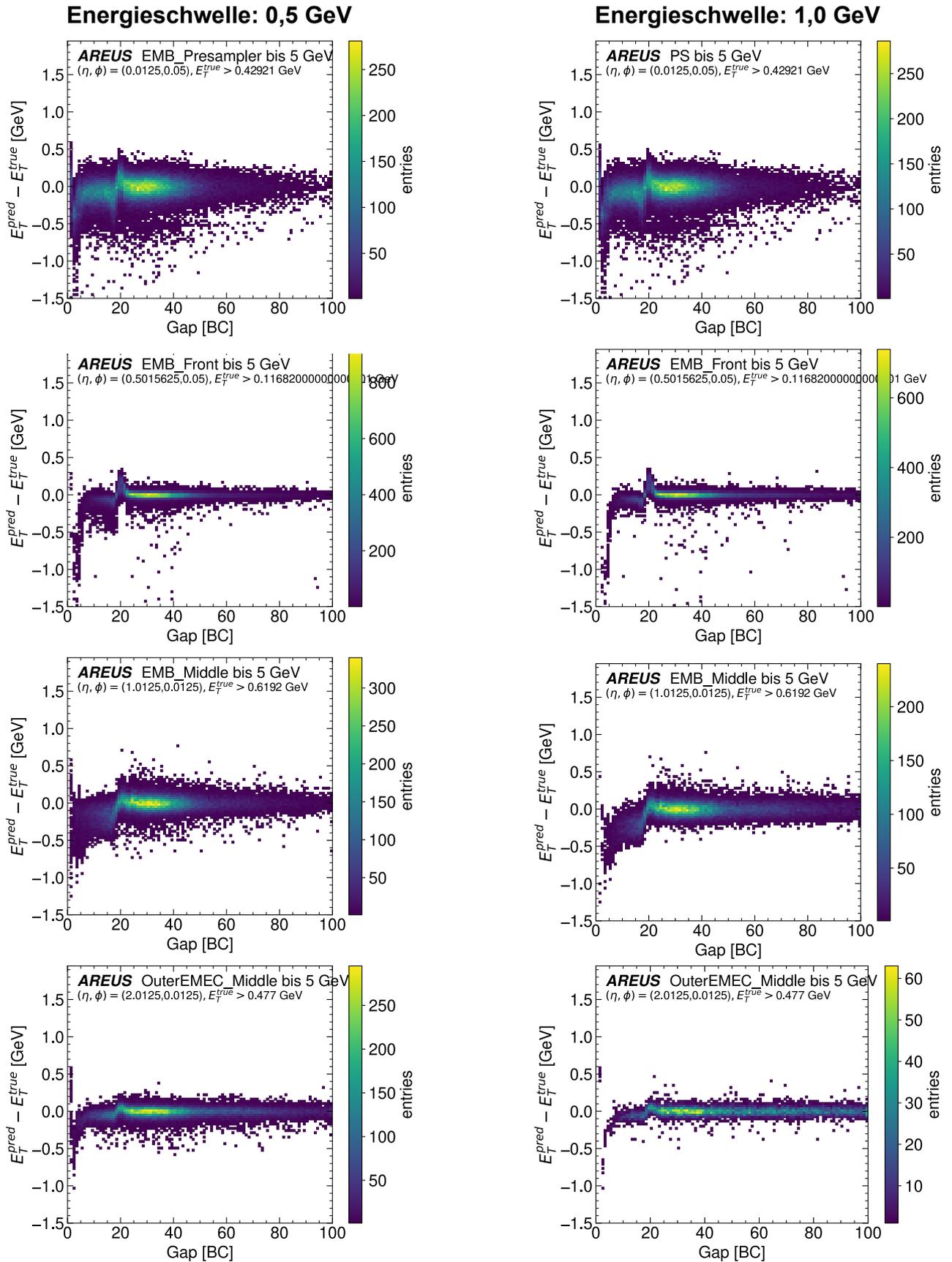


Abb. 23: Energiedifferenz über der Zeit auf den Zellen PS, F05, M10, M20 (v. o. n. u.) nach dem Training auf dem 4Conv-Netz mit geladenen Gewichten und den Energieschwellen 0,5 GeV und 1,0 GeV für die Pulserkennung. (Anhang zu Abschnitt 3.4.2, Abb. 16)

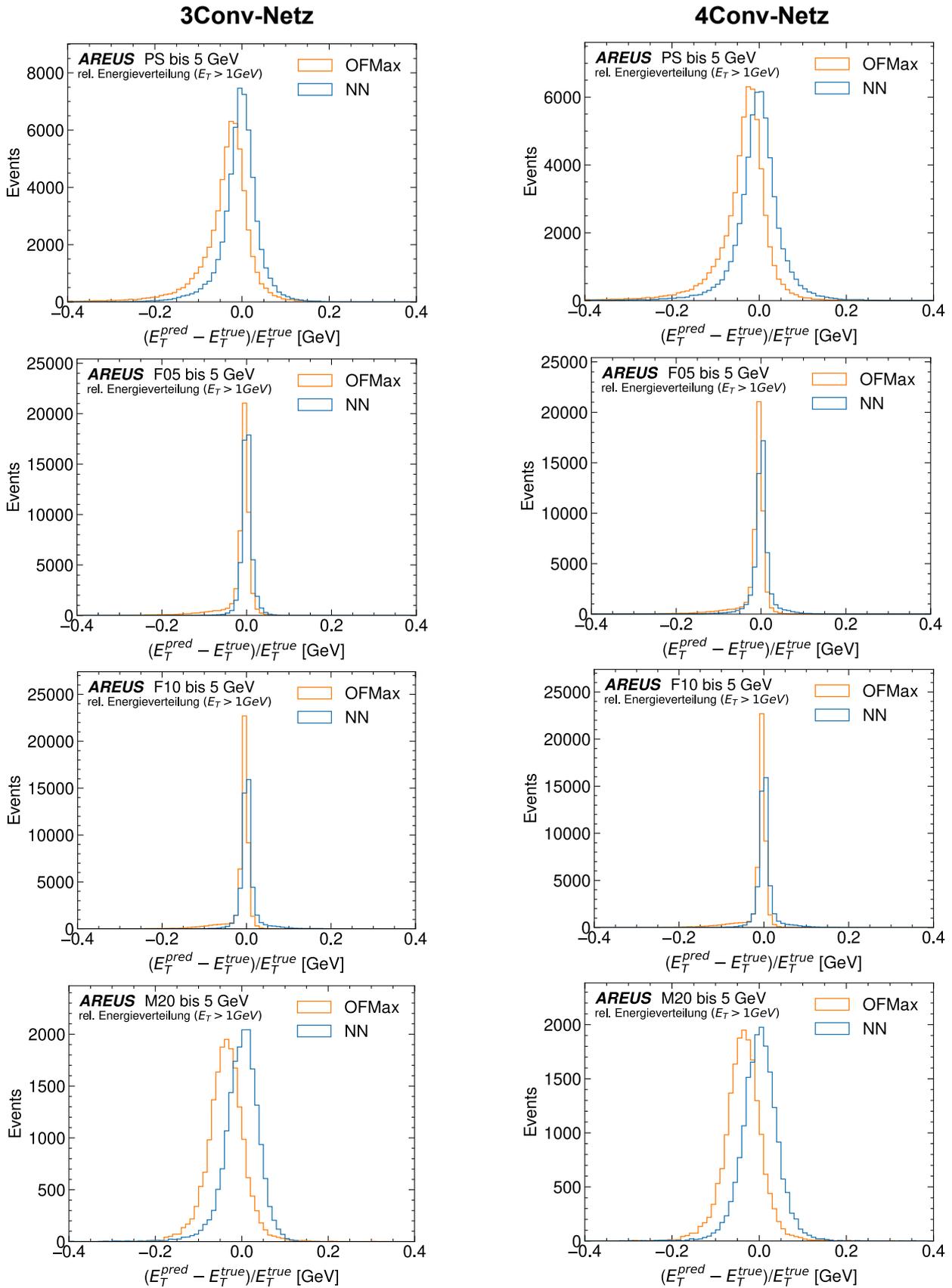


Abb. 24: Energiedifferenz normiert auf die wahre Energie der Zellen PS, F05, F10 und M20 (v. o. n. u.) nach dem Training mit geladenen Gewichten vom 3Conv- (links) und 4Conv-Netz (rechts) im direkten Vergleich zum OFMax. (Anhang zu Abschnitt 3.4.2, Abb. 17)

Literaturverzeichnis

- [1] CERN. The LHCb Experiment. URL: <https://lhcb-outreach.web.cern.ch/detector/>.
- [2] CERN. The Large Hadron Collider. URL: <https://home.web.cern.ch/science/accelerators/large-hadron-collider>.
- [3] CERN. The HL-LHC project. URL: <https://hilumilhc.web.cern.ch/content/hl-lhc-project>.
- [4] CERN. ATLAS. URL: <https://home.web.cern.ch/science/experiments/atlas>.
- [5] CERN. How did we discover the Higgs boson? URL: <https://home.cern/science/physics/higgs-boson/how>.
- [6] Pequeno, Joao: Computer generated image of the whole ATLAS detector. URL: <https://cds.cern.ch/record/1095924>.
- [7] ATLAS Experiment. The Inner Detector. URL: <https://atlas.cern/Discover/Detector/Inner-Detector>.
- [8] ATLAS Experiment. Calorimeter. URL: <https://atlas.cern/Discover/Detector/Calorimeter>.
- [9] ATLAS Experiment. Muon Spectrometer. URL: <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- [10] ATLAS Experiment. Magnet System. URL: <https://atlas.cern/Discover/Detector/Magnet-System>.
- [11] CERN. Geneva. The LHC experiments Committee. ATLAS Liquid Argon Calorimeter Phase-II Upgrade : Technical Design Report. Techn. Ber. CERN-LHCC2017-018 ; ATLAS-TDR-027. Geneva: CERN, Sep. 2017. DOI: 10.23731/CYRM2020-0010. URL: <http://cds.cern.ch/record/2285582>.
- [12] ATLAS Experiment. Calorimeter. URL: <https://atlas.cern/Discover/Detector/Calorimeter>.
- [13] Pequeno, Joao: Computer generated image of the ATLAS Liquid Argon. URL: <https://cds.cern.ch/record/1095928>.
- [14] Elst, Vincent: „Optimierung der Energiemessung im Flüssig-Argon-Kalorimeter des ATLAS-Detektors am LHC mit künstlichen neuronalen Netzen“. Technische Universität Dresden, Institut für Kern- und Teilchenphysik: 04.10.2021. URL: https://iktp.tu-dresden.de/IKTP/pub/21/StExArbeit_VincentElst.pdf.
- [15] Cleland, W. E.; Stern, E. G.: “Signal processing considerations for liquid ionization calorimeters in a high rate environment”. In: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 338.2 (1994), S. 467–497. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(94\)91332-3](https://doi.org/10.1016/0168-9002(94)91332-3). URL: <https://www.sciencedirect.com/science/article/pii/0168900294913323>.
- [16] Keras. Layer activation functions. URL: <https://keras.io/api/layers/activations/>.
- [17] Andriy Burkov. Machine Learning kompakt. 1. Auflage. Frechen: mitp Verlag, Juli 2019. ISBN: 9783958459977. URL: https://learning.oreilly.com/library/view/machine-learning-kompakt/9783958459977/Text/buchmlch6.xhtml#sigil_toc_id_67.

- [18] Ian Goodfellow, Yoshua Bengio und Aaron Courville. Deep Learning. MIT Press, 2016. URL: <http://www.deeplearningbook.org>.
- [19] Aad, Georges; Berthold, Anne-Sophie; Calvet, Thomas; u. a.: Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS LAr Calorimeters. Techn. Ber. Geneva: CERN, 2021. URL: <https://cds.cern.ch/record/2775033>.
- [20] Madysa, Nico: AREUS - A Software Framework for ATLAS Readout Electronics Upgrade Simulation. Techn. Ber. Geneva: CERN, 2018. DOI: 10.1051/epjconf/201921402006. URL: <https://cds.cern.ch/record/2645881>.
- [21] Brownlee, Jason (25.01.2019): Understand the Impact of Learning Rate on Neural Network Performance. URL: <https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/>.

Abbildungsverzeichnis

Nr.	Bezeichnung	Seite
1	Schematische Gesamtansicht des LHC	4
2	Übersicht der Betriebsphasen des LHC bzw. HL-LHC	5
3	Schematischer Aufbau des ATLAS-Detektors	6
4	Verlauf der durch das Magnetsystem erzeugten Feldlinien im ATLAS-Detektor	7
5	Schematischer Aufbau des LAr-Kalorimeters im ATLAS-Detektor	8
6	Schema eines Moduls des EM-Kalorimeters	9
7	Dreieckspuls des Ionisationsstromes im LAr-Kalorimeter	10
8	Darstellung eines künstlichen NN mit drei Schichten	13
9	Darstellung eines CNN mit zwei Schichten	13
10	Struktur des 4Conv-Netzes	16
11	Statistische Ergebnisse der CNN's ohne Training	20
12	Energieverteilung der CNN's ohne Training	21
13	Pulsformen der Zellen	22
14	2D-Histogramme der Zelle F05 nach dem Training mit geladenen Gewichten und Energieschwelle von 3 Sigma	23
15	Gegenüberstellung der Lernraten für die Zelle M10	24
16	Gegenüberstellung der Energieschwellen für die Pulserkennung für die Zelle F10	25
17	Relative Energieverteilung des 3Conv-Netzes auf der Zelle M10	26
18	2D-Histogramme der Zelle M10 nach zwei Trainings auf dem 4Conv-Netz mit gleichen Ausgangsbedingungen	27
19	2D-Histogramme der Zellen F10 nach dem Training mit geladenen Gewichten und Energieschwelle von 3 Sigma	30
20	2D-Histogramme der Zellen PS nach dem Training mit geladenen Gewichten und Energieschwelle von 3 Sigma	30
21	Gegenüberstellung der Lernraten $5 \cdot 10^{-3}$ und $5 \cdot 10^{-4}$ für die Zellen PS, F05, F10 und M20	31
22	Gegenüberstellung der Lernraten $5 \cdot 10^{-5}$ und $5 \cdot 10^{-6}$ für die Zellen PS, F05, F10 und M20	32
23	Gegenüberstellung der Energieschwellen für die Pulserkennung des 4Conv-Netzes für die Zellen PS, F05, M10 und M20	33

Nr.	Bezeichnung	Seite
24	Gegenüberstellung der relativen Energieverteilungen für die Zellen PS, F05, F10 und M20	34

Selbstständigkeitserklärung

Hiermit versichere ich, Alexander Keyn, die vorliegende Arbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben sowie alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, durch die Angabe der Quellen als Entlehnung kenntlich gemacht zu haben.

Dresden, 06.10.2022