



**TECHNISCHE  
UNIVERSITÄT  
DRESDEN**



INSTITUT FÜR  
KERN- UND  
TEILCHENPHYSIK

---

# **Optimierung von künstlichen Neuronalen Netzen für die Energierückgewinnung des Flüssigargonkalorimeters am ATLAS-Detektor**

Bachelor-Arbeit  
zur Erlangung des Hochschulgrades  
Bachelor of Science  
im Bachelor-Studiengang Physik

vorgelegt von

**GENNARO ROMANELLI**  
geboren am 01.12.1994 in Neuhaus am Rennweg

Institut für Kern- und Teilchenphysik  
Fakultät Physik  
Bereich Mathematik und Naturwissenschaften  
Technische Universität Dresden  
2023

Eingereicht am 23. Oktober 2023

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



## **Abstract**

The Large Hadron Collider is set to be upgraded from 2026 to 2028 for the high luminosity run, leading to an increase in the average number of beam intersections and events. This increase in events also affects background and pile-up. To improve energy reconstruction, the readout electronics of the liquid argon calorimeters will be improved, and research is being conducted on using neural networks implemented on field programmable gate arrays for optimal measurement. With this upgrade, the ATLAS trigger system will be able to record and process successive signals. However, the reconstruction of electronic signals from overlapping events has been challenging. This scientific work aims to analyze optimization methods, such as adjusting the architecture or training data set, to better reconstruct these directly successive events.

## **Zusammenfassung**

Im Zuge der Erweiterung des Large Hadron Colliders von 2026 bis 2028 für den High-Luminosity-Run wird sich die Luminosität und somit auch die mittlere Strahlkreuzungszahl um ein Vielfaches erhöhen. Infolgedessen kommt es zu einer Steigerung aller Ereignisse. Das betrifft auch die störenden. Eine verbesserte Performanz der Energierekonstruktion ist im Laufe des zeitgleich stattfindenden Ausbaus der Ausleseelektronik der Flüssigargonkalorimeter geplant. Für eine optimierte Messung wird der Einsatz von auf Field Programmable Gate Arrays implementierten Neuronalen Netzwerken erforscht. Mit dem Upgrade erhält das ATLAS-Triggersystem die Möglichkeit, direkt aufeinanderfolgende Signale aufzuzeichnen und für die Datenverarbeitung verfügbar zu machen. Da die elektronischen Signale solcher Ereignisse überlappen, ist ihre Rekonstruktion bisher häufig falsch. Diese wissenschaftliche Arbeit analysiert Optimierungsmethoden, wie das Anpassen der Architektur oder des Trainingsdatensatzes von Neuronalen Netzwerken, um solche direkt aufeinanderfolgenden Ereignisse rekonstruieren zu können.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theoretischer Hintergrund</b>	<b>2</b>
2.1	Das CERN . . . . .	2
2.2	ATLAS - Aktuelles Design . . . . .	3
2.2.1	Flüssigargonkalorimeter . . . . .	6
2.2.2	High-Luminosity-LHC . . . . .	8
2.3	Neuronale Netzwerke als Alternative . . . . .	9
2.3.1	Training und Evaluation Neuronaler Netzwerke . . . . .	9
2.3.2	Verlustfunktion und Optimierung Neuronaler Netze . . . . .	11
2.3.3	Convolutional Neural Networks . . . . .	14
2.3.4	AREUS-Simulation . . . . .	15
2.4	Hyperparametersuche . . . . .	17
<b>3</b>	<b>One-Gap-Überlappung</b>	<b>19</b>
3.1	Einfluss der Rekonstruktionslatenz . . . . .	22
3.1.1	Performanz der Rekonstruktion von One-Gap-Überlappungen . . . . .	24
3.1.2	Allgemeine Performanz der Energierekonstruktion . . . . .	27
3.2	Variation von Rekonstruktionslatenz und Kernelgröße . . . . .	28
3.3	Hyperparameter und erweiterter Datensatz . . . . .	30
3.3.1	Performanz der Rekonstruktion von One-Gap-Überlappungen . . . . .	33
3.3.2	Allgemeine Performanz der Energierekonstruktion . . . . .	35

3.3.3	Effektivität der Hyperbandsuche und des erweiterten Datensatzes .	40
<b>4</b>	<b>Zusammenfassung und Ausblick</b>	<b>42</b>
<b>5</b>	<b>Anhang</b>	<b>44</b>
<b>6</b>	<b>Literatur</b>	<b>51</b>

# Abkürzungsverzeichnis

**ADC** Analog-to-Digital-Converter

**ALICE** A Large Ion Collider Experiment

**ANN** Artificial Neural Networks

**AREUS** ATLAS Readout Electronic Upgrade Simulation

**ATLAS** A Toroidal LHC Aparatus

**BC** Bunch Crossing

**BSM** Beyond Standard Model

**CERN** Europäische Organisation für Kernforschung

**CMS** Compact Muon Solenoid

**CNN** Convolutional Neural Network

**DR** Dilationsrate

**EMB** Electromagnetic Barrel

**EMEC** Electromagnetic End Cap

**FCal** Forward Calorimeter

**FoV** Field of View

**FPGA** Field Programmable Gate Array

**HEC** Hadronic End Cap

**HL** High Luminosity

**LAr** Flüssigargon

**LEP** Large-Electron-Positron-Beschleuniger

**LHC** Large Hadron Collider

**LHCb** Large Hadron Collider beauty

**MAE** mittlere absolute Abweichung

**MSE** mittlere quadratische Abweichung

**OF** Optimalfilter

**ReLU** Rectified Linear Unit

**RL** Rekonstruktionslatenz

**ROC** Receiving Operator Curve

**SCT** Halbleitertracker

**TRT** Übergangstracker



# 1 Einleitung

In den 1960er Jahren sagten die Physiker Brout, Englert, Higgs, Guralnik, Hagen und Kibble die Existenz des Higgs-Bosons voraus [1], [2], [3]. Diese Vorhersage wurde durch die Experimente von A Toroidal LHC Aparatus (ATLAS) und Compact Muon Solenoid (CMS) im Jahr 2012 unabhängig bestätigt [4]. Higgs und Englert erhielten 2013 den Physik-Nobelpreis für ihre theoretischen Vorhersagen. Ab diesem Zeitpunkt waren alle erwarteten und prognostizierten Elementarteilchen des Standardmodells der Teilchenphysik entdeckt - eine bedeutende Errungenschaft.

Die Physik Beyond the Standard Model (BSM), das ist Physik über das Standardmodell hinaus, setzt hier an. Viele offene Fragen und unerklärte Effekte bleiben bestehen. Gibt es nur ein Higgs-Boson oder mehrere? Was verbirgt sich hinter dunkler Materie? Existieren unentdeckte langlebige Teilchen? Welche Vorhersagen treffen neue theoretische BSM-Modelle? Supersymmetrische Theorien und ihre Fähigkeit, die bekannten Probleme des Standardmodells zu lösen, sind Gegenstand zahlreicher offener Fragestellungen [5], [6].

Um Antworten darauf zu finden, soll der Large Hadron Collider von 2026 bis 2028 für die sogenannten High-Luminosity-Runs und auch ATLAS im Phase-II-Upgrade überarbeitet werden [7], [8], [9]. Dabei soll die Design-Luminosität versiebenfacht und die Zahl der Teilchenkollisionen erhöht werden. Diese Vervielfachung führt zu einem höheren Pile-up und Untergrundsignal. Teil des ATLAS-Detektors sind die Flüssigargonkalorimeter. Es stellt sich die Frage, ob der dort für die Energierekonstruktion genutzte Optimalfilter durch alternative Rekonstruktionsalgorithmen ersetzt werden kann. Ein derzeit erforschter Ersatz nutzt Neuronale Netzwerke. Diese sollen auf die Voraussage der in den Zellen des Flüssigargonkalorimeters deponierten Teilchenenergien optimiert werden. Durch das Upgrade erhält das ATLAS-Triggersystem die Fähigkeit, direkt aufeinanderfolgende Ereignisse aufzuzeichnen. Weder der Optimalfilter, noch die Neuronale Netze können solche Teilchenmessungen bisher rekonstruieren. Deshalb werden in dieser wissenschaftlichen Arbeit mehrere Methoden zur Realisierung dieses Aspekts untersucht.

## 2 Theoretischer Hintergrund

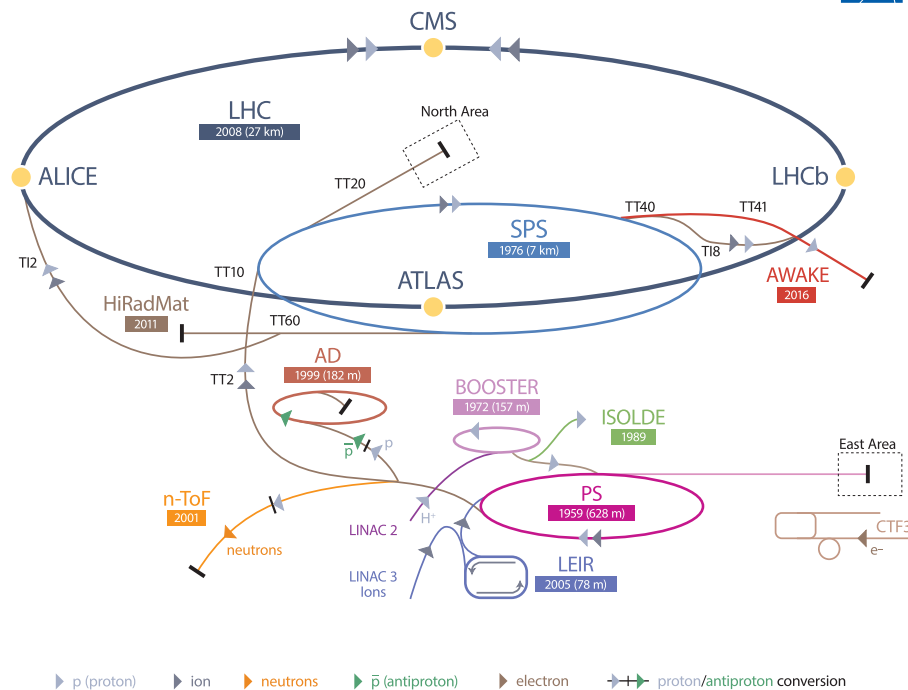
### 2.1 Das CERN

Bereits in den frühen 1980er Jahren planten Teilchenphysiker den Nachfolger des Large-Electron-Positron-Beschleuniger (LEP) - den sogenannten Large Hadron Collider (LHC) [10]. Die Energie am LEP kann durch Energieverluste der Synchrotronstrahlung nicht weiter erhöht werden. Eine Konstruktion für höhere Energien sind Hadronenbeschleuniger. Für die Verifikation des Standardmodells, zur Suche des Higgs-Bosons und dem Überprüfen weiterer physikalischer Phänomene gab man dem Vorhaben 1994 statt [11]. Nach vielen Jahren der Planung, Konstruktion und Kollaboration konnte der LHC zum ersten Mal am 10. September 2008 Protonen in den Beschleunigerring schicken.

Am 20. November 2009 wurde der LHC nach einem Defekt wieder in Betrieb genommen. Seitdem werden im LHC Hadronen (z.B. Protonen) oder Schwerionen innerhalb zweier Strahlröhren entgegengesetzt beschleunigt und an mehreren Punkten des Ringes zur Kollision gebracht. Vier Interaktionspunkte befinden sich bei den Hauptexperimenten A Large Ion Collider Experiment (ALICE), Large Hadron Collider beauty (LHCb), CMS und ATLAS. Die Europäische Organisation für Kernforschung (CERN) führt weitere Untersuchungen und Kollaborationen am gesamten Campus aus (siehe Abbildung 2.1). Mit knapp 27 km Umfang ist der LHC der größte Teilchenbeschleuniger der Welt. Um die hohen Schwerpunktsenergien der Teilchenstrahlen von bis zu 13.6 TeV zu erreichen, müssen die Protonen über mehrere Vorbeschleuniger (Booster, PS, SPS) angetrieben werden. Im LHC angekommen halten Dipolmagneten die nahezu lichtschnellen Teilchen auf der Bahn, Quadrupolmagnete fokussieren den Strahl. Das Zusammentreffen der Teilchen wird Bunch Crossing (BC) genannt. Diese treten alle 25 Nanosekunden (Strahlkreuzungszahl von 40 MHz) auf. Jede dieser Strahlkreuzungen verursachen in Run 1 und Run 2 laut Design im Mittel 20, derzeit sogar 60 Proton-Proton-Ereignisse [12].



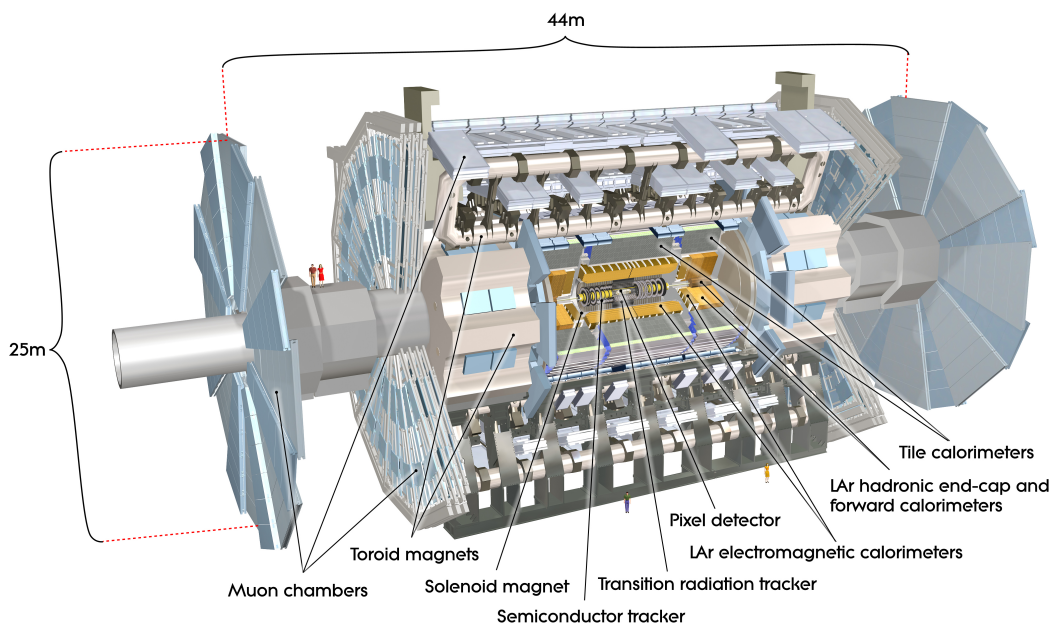
## CERN's Accelerator Complex



**Abbildung 2.1:** Übersicht der Experimente am LHC inklusive aller Vorbeschleuniger [13].

## 2.2 ATLAS - Aktuelles Design

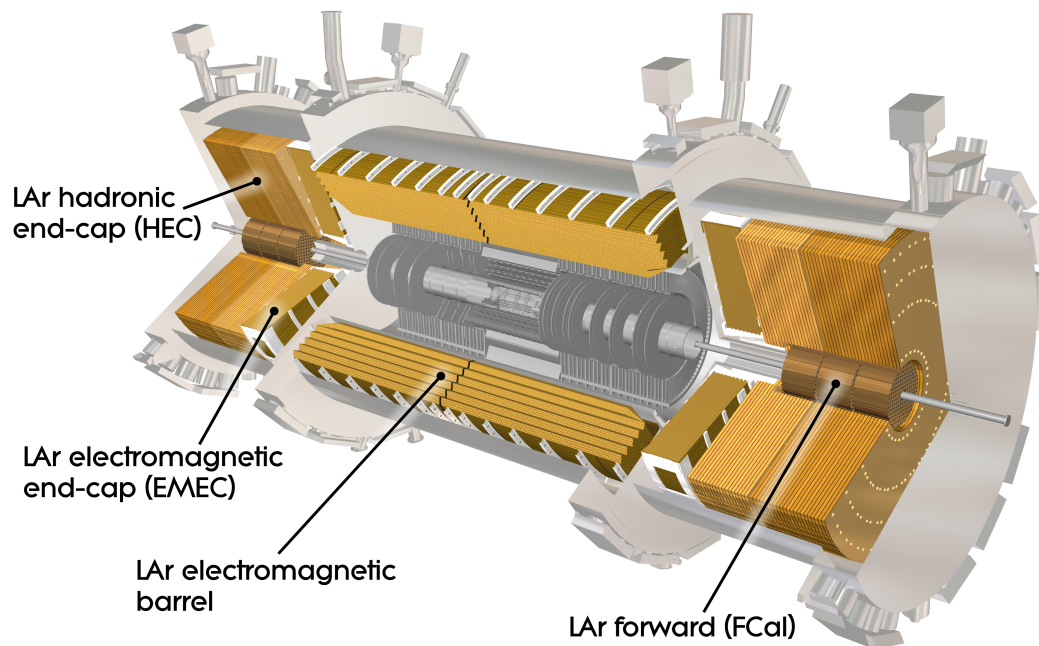
ATLAS ist eines der vier Hauptexperimente am LHC. Wie auch am CMS ist es ein General-Purpose-Detektor. Der Detektor hat einen zylinderförmigen Aufbau. Das Gewicht von ATLAS liegt bei 7000 Tonnen. Das physikalische Ziel ist die Rekonstruktion von Higgs-Bosonen aus Zerfallsprodukten, die Erstellung von Beyond Standard Model (BSM)-Analysen, die Bestätigung und Überprüfung des Standardmodells der Teilchenphysik und die Analyse weiterer theoretischer Modelle. Dies wird durch Messung der in Kollisionen erzeugten sekundären Teilchen gewährleistet. Der Ursprungsort wird aus dem Schnittpunkt der Teilchenspuren extrapoliert. Es werden Impuls und Energie aufgenommen und aus diesen Geschwindigkeit und Masse bestimmt. Die Ladung wird aus der Krümmung der Teilchenspuren innerhalb eines externen Magnetfeldes bestimmt. Im Zwiebelprinzip angeordnet, besitzt der Detektor verschiedene Komponenten, die unterschiedliche Aufgaben übernehmen. In mehreren Schichten werden die vorher genannten Informationen ausgelesen. Daraus erfolgt eine Rekonstruktion von Teilchen und Jets.



**Abbildung 2.2:** Computergenerierte Grafik des ATLAS-Detektors im Zwiebelschalenprinzip [14].

ATLAS ist 46 Meter lang, 26 Meter hoch und breit. Im Zentrum befindet sich der Interaktionspunkt der Bunches. Die ersten Detektorelemente sind Pixeldetektor, dann Halbleitertracker (SCT) und Übergangstracker (TRT). Darauf folgen das elektromagnetische Kalorimeter, das hadronische Kalorimeter und zum Schluss das Myonenspektrometer. Im Inneren gibt es einen Solenoidmagneten und für das Myonenspektrometer ein Toroidmagnetsystem. Im inneren Detektor (Pixeldetektor, SCT, TRT) wird der Teilchenimpuls anhand ihrer Flugbahn im Magnetfeld bestimmt. Elektromagnetische Teilchen, wie Photonen und Elektronen, werden im elektromagnetischen Kalorimeter gestoppt und ihre Energie gemessen. Hadronen werden nur gebremst und kommen erst im hadronischen Kalorimeter zum Stillstand. Myonen passieren alle bisherigen Schichten und werden im letzten Teil, dem Myonenspektrometer, gemessen. Neutrinos sind die einzigen Teilchen im Standardmodell, die nicht mit dem Detektor interagieren und ihn passieren. Diese und theoretische exotische Teilchen können nur über die fehlende transversale Energie  $E_T^{\text{miss}}$  und die Abschirmung des Detektors rekonstruiert werden.

ATLAS ist radialsymmetrisch um den Strahl und spiegelsymmetrisch in der  $xy$ -Ebene aufgebaut. Nicht jeder Punkt des Detektors kann Teilchen messen. Um diese Messabdeckung akkurat und eindeutig aufzuzeichnen, wird ein spezielles Koordinatensystem für zylinderförmige Detektoren genutzt. In Abbildung 5.1 wird dieses Koordinatensystem dargestellt.



**Abbildung 2.3:** Aufbau des Flüssigargonkalorimeters des ATLAS-Detektors zur Messung elektromagnetischer und hadronischer Teilchen [15].

Es wird die Pseudorapidität  $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$  mit dem Polarwinkel  $\theta$  (in  $yz$ -Ebene) und der Azimuthalwinkel  $\phi$  (in  $xy$ -Ebene) genutzt [16]. Das Verhältnis der Pseudorapidität zum Polarwinkel wird in Abbildung 5.2 gezeigt. Wie bereits erwähnt werden am ATLAS Protonen und Schwerionen kollidiert. Da es sich dabei nicht um Elementarteilchen handelt, kann nur mit Partonen, den einzelnen Bestandteilen dieser Teilchen, gerechnet werden. Das führt dazu, dass die genaue Energie der einzelnen Partonen nicht bekannt ist. Die Energie in Strahlrichtung kann für Berechnungen und Analysen nicht genutzt werden. Aus diesem Grund nutzt man die sogenannte transversale Energie  $E_T$ . In transversaler Richtung werden die Teilchen nicht beschleunigt. Sie haben vor der Kollision eine Gesamtenergie und einen Gesamtimpuls von null. Die Energie- und Impulserhaltung wird hier genutzt. Wenn sich die Gesamtenergie der Zerfallsprodukte nicht zu null aufsummiert, dann müssen weitere Zerfallsprodukte entstanden sein, welche nicht vom Detektor gemessen werden können. Diese von null abweichende Energie nennt man auch fehlende Energie  $E_T^{\text{miss}}$ . Eine genaue Messung dieser ist wichtig.

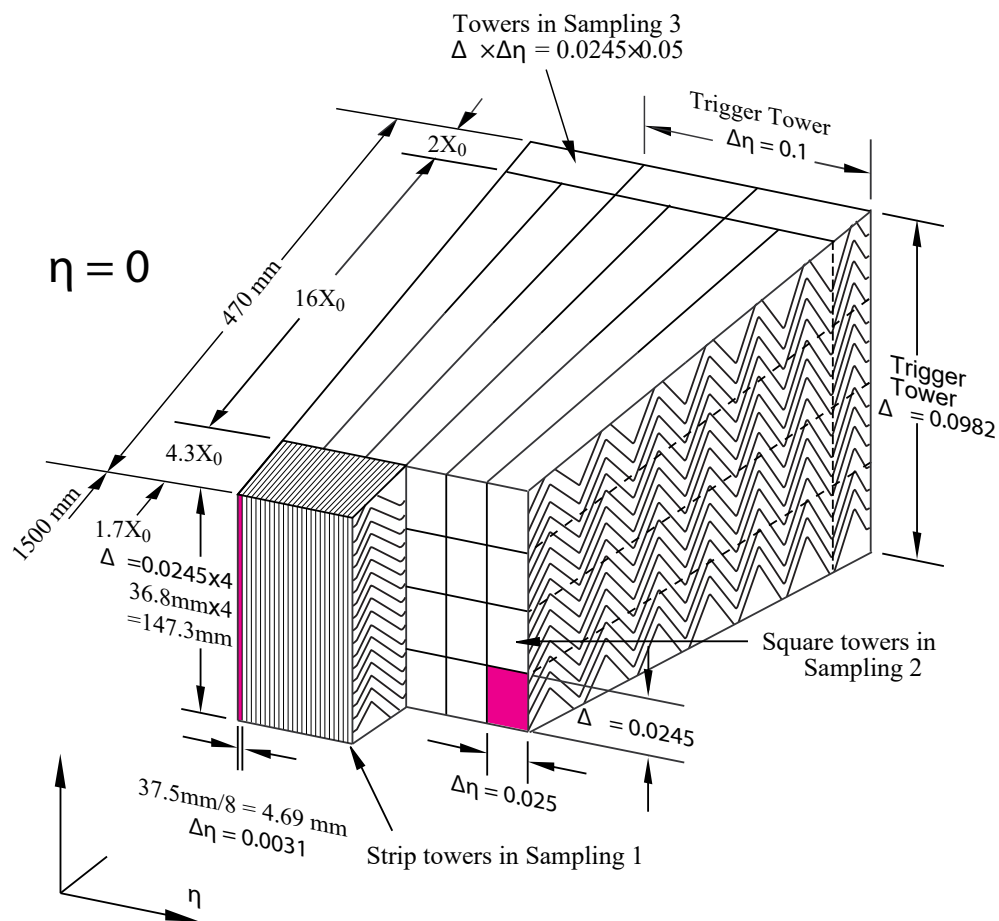
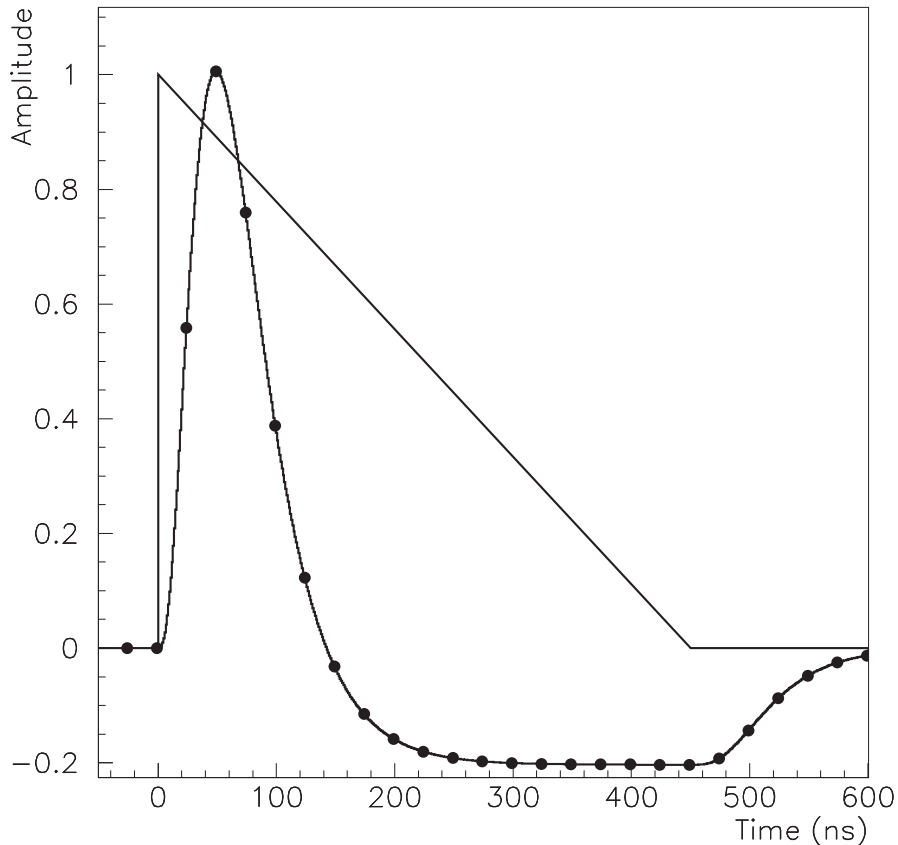


Abbildung 2.4: Akkordeonkonstruktion der Flüssigargonkalorimeter [17].

### 2.2.1 Flüssigargonkalorimeter

An den inneren Detektor grenzen das elektromagnetische Flüssigargonkalorimeter und auch das hadronische Kalorimeter. Es besteht aus ungefähr 182.000 Zellen. Die Pseudorapidität des elektromagnetischen Kalorimeters befindet sich im Bereich  $|\eta| < 3.2$ , das Forward Calorimeter (FCal) liegt bei  $3.0 < |\eta| < 4.9$ . Unterteilt wird es in das Electromagnetic Barrel (EMB), das Electromagnetic End Cap (EMEC), das Hadronic End Cap (HEC) und das FCal (siehe Abbildung 2.3). Die bereits erwähnten Bestandteile der Kalorimeter benutzen Flüssigargon (LAr) als aktives Medium. Bei der Durchquerung dieser Substanz werden Argon-Atome ionisiert. Die dabei entstehenden Elektronen driften im elektrischen Feld zu den Ausleselektroden und können gemessen werden.



**Abbildung 2.5:** Elektronische Signalform einer LAr-Zelle [17].

Die Kalorimetersubstruktur ist akkordeonförmig (siehe Abbildung 2.4). In EMB und EMEC wechselt sich Absorbermaterial und das aktive Medium nacheinander ab (siehe Abbildung 2.4). Stahl dient der Festigkeit des Detektors. Trifft ein Teilchen ein, wird ein Dreieckspuls generiert. Dieser wird dann durch einen analogen Shaper (CRRC<sup>2</sup>-Filter) zu einem Bipolarpuls (siehe Abbildung 2.5) transformiert. Danach wird das Analogsignal mittels Analog-to-Digital-Converter (ADC) digitalisiert [17]. Bei mehreren aufeinanderfolgenden Teilchen überlagern sich Dreieckspulse so, dass eine genaue Energierekonstruktion schwierig ist.

Eine präzise Energierekonstruktion der Flüssigargonkalorimeter ist für wichtige Ereignisse von Relevanz. Elektroschwache Interaktionen sind für besondere Zerfälle, wie z. B. des Higgs-Bosons oder der W-Bosonen, relevant. Auch Kaonen und Pionen zerfallen in elektrisch geladene Teilchen. Eine bessere Rekonstruktion ermöglicht eine höhere Messgenauigkeit, unter anderem für Standardmodellparameter und auch für die Messung theoretischer exotischer Teilchen über  $E_T^{\text{miss}}$ .

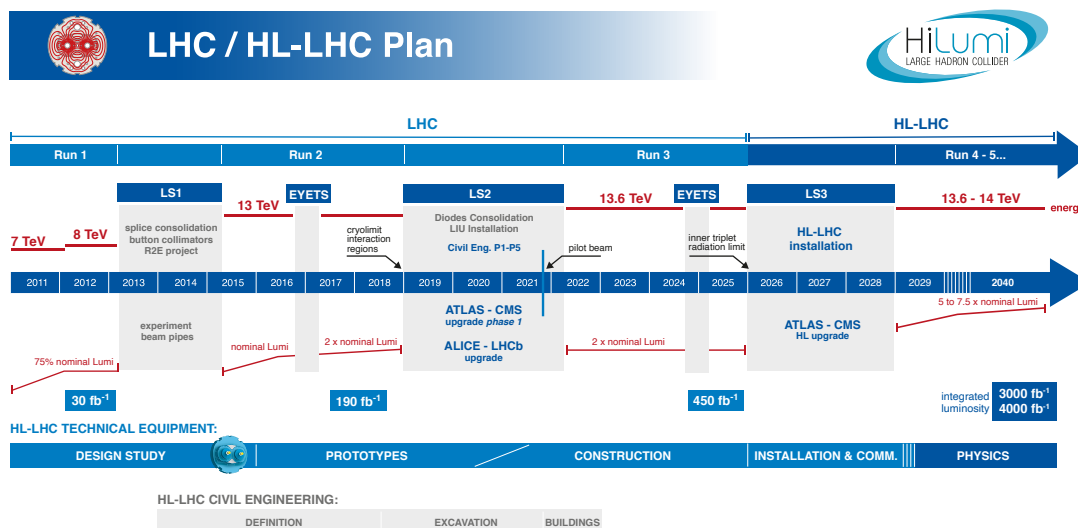


Abbildung 2.6: Zeitplan der LHC-Upgrades und -Erweiterungen [7].

## 2.2.2 High-Luminosity-LHC

Seit den ersten Kollisionen am LHC wurde mit dem Ausbau des Beschleunigers die Schwerpunktsenergie von  $\sqrt{s} = 7$  TeV bis zu  $\sqrt{s} = 13.6$  TeV im Jahre 2022 erhöht. Luminosität beschreibt die Anzahl der Teilchenkollisionen pro Fläche und Zeit. Die designte mittlere Zahl an Strahlkreuzungen entspricht  $\mu = 20$  je BC. Diese soll mit der Hochluminositätsenerweiterung zwischen 2026 und 2028 auf über  $\mu = 140$  erhöht werden (siehe Abbildung 2.6). Das sind sieben Mal mehr initiale Ereignisse, die bis zu den Messzellen noch exponentiell zunehmen und dort gemessen werden. Dadurch kommt es häufiger zu Ereignissen, die die Auslese und Rekonstruktion bewältigen müssen. Auch die Erzeugung niedrig-energetischer Teilchen wie Pionen und damit auch das Pile-up nehmen zu. Pile-Up ist die Überlagerung der Pulse durch zeitlich aufeinanderfolgende Energiepositionen.

Erwähnenswert ist die bei den Strahlkreuzungen entstehende enorme Anzahl an Daten. Von diesen müssen die wichtigen und interessanten intelligent aufgezeichnet werden. Das entscheidet das sogenannte Triggersystem. Das ist eine Herausforderung für ATLAS und die im High Luminosity (HL)-Run auftretenden Luminositäten. Es bestehen nur begrenzte Kapazitäten für das Echtzeit-Triggern. Der derzeitige Standardalgorithmus zur Energierekonstruktion ist der Optimalfilter (OF) [18]. Als Alternative hierfür könnte in Zukunft auf Field Programmable Gate Arrays (FPGAs) implementierte Artificial Neural Networks (ANN)s dienen [9].

Der OF wird definiert als

$$y_t = \sum_{n=0}^{M-1} x_{t-n} \cdot a_n \quad (2.1)$$

wobei  $y_t$  die Ausgabe des OFs für BC  $t$  ist, vom ADC sind  $\vec{x}$  die Eingabedaten,  $a_n$  die Koeffizienten des OFs und  $M$  die Filtertiefe. Generell ist der Filter die beste mathematische lineare Operation für die Energieauflösung, allerdings nur bei isolierten, nicht überlagernden Signalen.

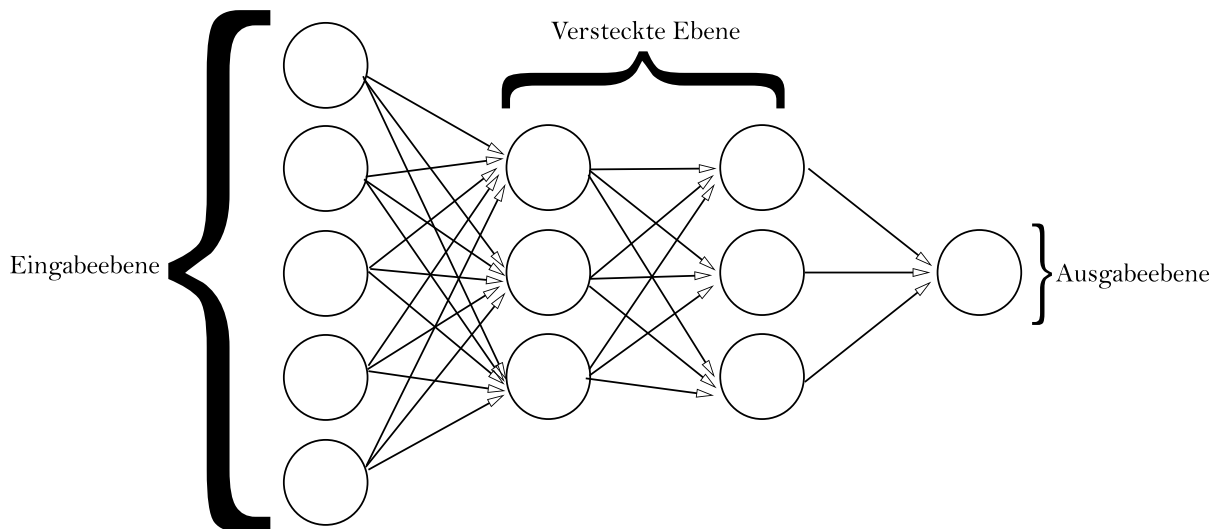
## 2.3 Neuronale Netzwerke als Alternative

Neuronale Netzwerke sind Teil des Forschungsschwerpunktes der künstlichen Intelligenz. Es wurde sich hierbei an biologischen Neuronen orientiert. Neuronale Knoten bekommen Eingaben und geben Informationen wieder aus. Neuronen werden in verschiedenen Schichten, auch Layer oder Ebenen genannt, zum Lösen spezifischer Probleme eingesetzt. Heutzutage werden solche Netzwerke für eine Vielzahl von Aufgaben verwendet, so z. B. Computer Vision, Spracherkennung, Sprachverarbeitung und viele mehr.

Künstliche Neuronale Netzwerke werden auch ANN genannt. Simple Architekturen eines ANN bestehen aus einem Eingabelayer, einem Verarbeitungslayer und dem Ausgabebelayer. Beliebige komplizierte Varianten können konstruiert werden. Das folgende Unterkapitel erläutert die Trainings Neuronaler Netzwerke ([19], Kapitel 1).

### 2.3.1 Training und Evaluation Neuronaler Netzwerke

Neuronale Netzwerke können als Funktion dargestellt werden. Dabei ist  $x$  der Eingabewert des Netzes,  $y^*$  der Zielwert,  $y$  der Ausgabewert und die Funktion ist abhängig vom Parameter  $\theta$ . Diese Annäherung kann mit  $y = f^*(x, \theta)$  abgebildet werden. Aus der Tiefe der Neuronalen Netze entspringt auch die Bezeichnung *Deep Learning*. Eine häufig gewählte Darstellungsform Neuronaler Netze ist ein Multilayer-Perceptron, welches in Abbildung 2.7 gezeigt wird. Es ist die einfachste Architektur eines ANNs. Hierbei handelt es sich um ein Neuronales Netzwerk mit vier Ebenen: einer Eingabeebene, zwei versteckten Ebenen und einer Ausgabebene. Netze können allgemein beliebig komplizierter gemacht werden.



**Abbildung 2.7:** Neuronales Netzwerk mit vier Ebenen: Einer Eingabeebene, zwei versteckten Ebenen und einer Ausgabeebene.

Das Multilayer-Perceptron besteht aus vielen einzelnen Neuronen. Ein solches ist in Abbildung 5.3 zu sehen. Bei diesem werden die Eingabedaten  $x_1, x_2, x_3$  mit den Gewichten  $w_1, w_2, w_3$  multipliziert, mit einem Bias  $b$  addiert und die Summe  $\sum_{i=1}^n x_i w_i + b$  darüber gebildet. Am Ende kann eine Aktivierungsfunktion für die Ausgabe des Neurons für bessere Trainingserfolge sorgen. Zur Spezialisierung der Netze wird ein Trainingsdatensatz verwendet, um Gewichte anzupassen. Jeder Eingabewert eines Netzwerkes wird mit einem Gewicht transformiert. Ein Gewicht hat dementsprechend einen Einfluss auf die Stärke der in das Netz eingespeisten Datenwerte. Für ein Netz wird ein sogenanntes *Training* durchgeführt und darauf folgend eine *Evaluation*.

Eine Form des Trainings wird *Supervised Learning* genannt. Die Netze bekommen einen Trainingsdatensatz, der Eingabewerte und dazugehörige Zielwerte enthält. Als Beispiel soll ein Netz die Fellfarbe eines Hundes erkennen. Es wird mit verschiedenen Hundebildern mit bekannter und gekennzeichnete Fellfarbe trainiert. Je mehr Daten in das Netz eingespeist werden, desto mehr kann es lernen. Beim *Unsupervised Learning* hingegen muss das Netzwerk Zusammenhänge eigenständig erkennen. Ein bereits trainiertes Netz kann auf unbekannte Daten angewandt werden. Dabei wird der Trainingserfolg des Netzes evaluiert. Wenn der Erfolg ausbleibt, können die nachfolgenden Szenarien ursächlich sein.

Beim **Overfitting** ist die Ausgabe der ANNs für die Trainingsdaten gut. Bei der Anwendung des trainierten Netzwerkes auf einen unbekanntem Datensatz entsteht aber ein ungenügendes Ergebnis. Das kann passieren, wenn der Trainingsdatensatz zu spezifisch



oder das Netz zu groß ist. Das Netzwerk hat sich auf den Trainingsdatensatz spezialisiert und beherrscht es nicht, diesen zu generalisieren. Overfitting kann auch auftreten, wenn die Variabilität des Trainingsdatensatzes zu gering ist. Ein Beispiel für Overfitting ist ein Netzwerk, das Bilder mit Autos erkennen soll. Jetzt wird das Netzwerk nur mit vielen Bildern von Autos in einer Garage trainiert. Dem Netz wird es dann nicht möglich sein, Autos in anderen Umgebungen, wie z. B. auf einer Landstraße, zu erkennen.

Ein weiteres Szenario für Misserfolg beim Training ist das **Underfitting**. Dabei kann das trainierte Netz zwischen Eingangs- und Ausgabedaten keinen sinnvollen Zusammenhang erkennen. Es wird auch davon gesprochen, dass ein Netzwerk nicht konvergiert. Netzwerke können auch ohne Underfitting nicht konvergieren. Diese können in einem lokalen Minimum oder Sattelpunkt stecken bleiben. Auch dann geben Netze keine sinnvollen Ergebnisse aus und konvergieren nicht ([19], Kapitel 5; [20]).

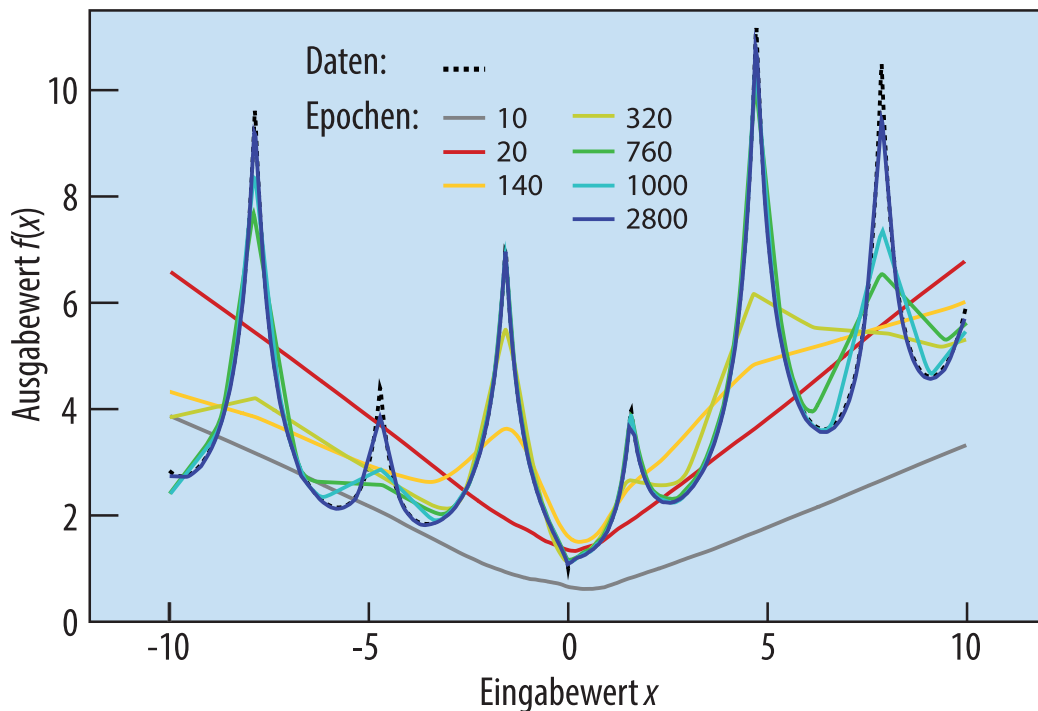
Im Gegensatz zum Overfitting, bei dem das Netz auf den Trainingsdatensatz spezialisiert wird, kann das Netz beim Underfitting für keinen Datensatz guten Ausgaben generieren. Die sogenannte Verlustfunktion wird beim Underfitting groß.

### 2.3.2 Verlustfunktion und Optimierung Neuronaler Netze

Ein Maß dafür, wie gut ein Netzwerk lernt, ist die *Verlustfunktion*, manchmal auch Loss- oder Kostenfunktion genannt. Ihre Minimierung ist das angestrebte Ziel des Trainings Neuronaler Netze. Häufig wird die mittlere quadratische Abweichung (MSE) als Verlustfunktion genutzt. Für die allgemeine Optimierung der Energierekonstruktion hat sich allerdings die mittlere absolute Abweichung (MAE) als vorteilhaft herausgestellt. Die Verlustfunktion  $C_{\text{MSE}}(w, b)$  erzeugt breitere Energiedifferenzen im Vergleich zur MAE. Erklärt wird das dadurch, dass stark abweichende Werte stärker gewichtet werden als kleine. Das vernachlässigt die Optimierung bereits geringer Werte [21]. Die MAE wird definiert als

$$C_{\text{MAE}}(w, b) \equiv \frac{1}{n} \sum_x |(y(x) - y^*)| \quad (2.2)$$

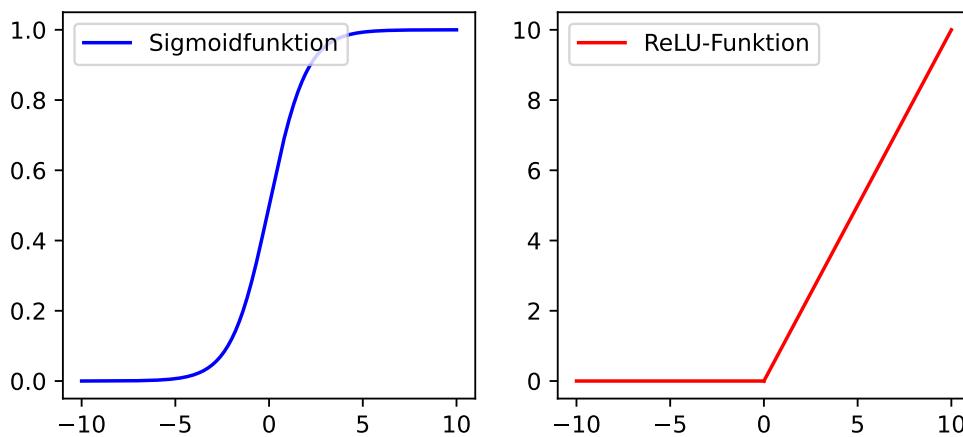
Das Netz verbessert sich im Vergleich zu einer vorherigen Epoche durch die Minimierung der gewählten Kostenfunktion. Ein Training durchläuft mehrere Epochen. Nach jeder Epoche werden die Gewichte der Netzwerkknoten evaluiert. Der Prozess dieser Annäherung an einen Datensatz ist dargestellt in Abbildung 2.8. Dafür könnte der analytische Gradient der Verlustfunktion ermittelt werden. Er ist eine Vektorgröße und würde die Ableitung an jedem Punkt ermitteln. Effizienter ist der numerische *Gradient-Descent*-Algorithmus, aber es gibt auch weitere Verfahren.



**Abbildung 2.8:** Grafische Darstellung der graduellen Optimierung eines Neuronales Netzwerkes mit steigender Epochenzahl [22].

Der Algorithmus ist in Abbildung 5.4 dargestellt. Mit jedem weiteren Iterationsschritt werden die Gewichte so angepasst, dass sie dem Minimum näherkommen. Das wird so oft wiederholt, bis man ein Minimum oder eine Abbruchbedingung erreicht hat. Eine Abbruchbedingung kann zum Beispiel das Unterschreiten eines vorher festgelegten Verlustwertes sein. Beim Gradient-Descent-Algorithmus kann es passieren, dass die Gewichte bei einem lokalen Minimum oder einem Sattelpunkt hängen bleiben und das globale Minimum der Verlustfunktion nicht erreicht wird. Das zu vermeiden gewährleisten Optimierungsalgorithmen, welche die Lernraten dynamisch anpassen [23]. Ein solcher Optimierungsalgorithmus ist Adam, welcher bei den Trainings zur Energierekonstruktion genutzt wird.

Da die von einem Netz zu nähernden Funktionen in der Regel nicht linear sind, werden sogenannte *Aktivierungsfunktionen* genutzt. Damit kann ein Neuronales Netz grundsätzlich jede Funktion nähern und komplexere Beziehungen und Eigenschaften in den Daten abbilden. Ohne Aktivierungsfunktion wären die einzigen Operationen lineare zwischen Gewichten und Eingabedaten.



**Abbildung 2.9:** Zwei häufig genutzten Aktivierungsfunktionen: Die Sigmoidfunktion und die ReLU-Funktion.

Häufig werden die in Abbildung 2.9 dargestellten Funktionen genutzt. Die Sigmoidfunktion ist im Grunde eine geglättete Stufenfunktion und definiert als

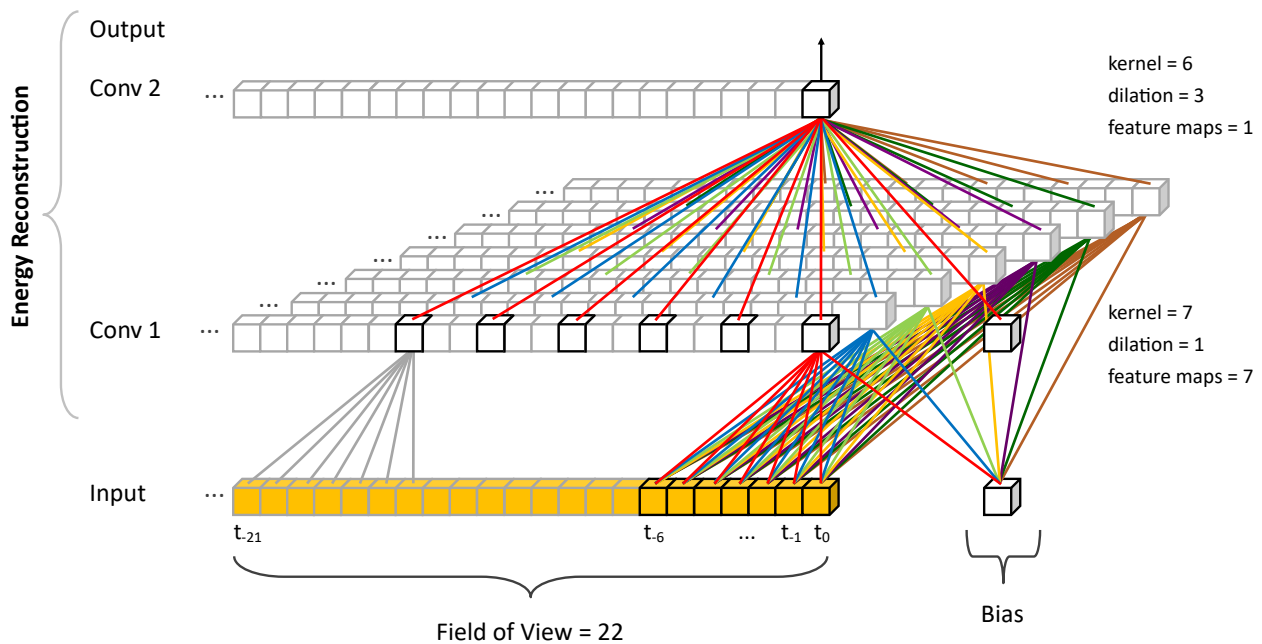
$$S(x) = \frac{1}{1 + \exp(-x)} \quad (2.3)$$

und die Rectified Linear Unit (ReLU)-Funktion

$$\text{ReLU}(x) = \max(0, x) \quad (2.4)$$

welche für Werte kleiner null einen Ausgabewert von null zurückgibt und linear ist für Werte, die größer als null sind.

Das Netzwerk durchläuft Schritt für Schritt jede Ebene, wird mit den Gewichten der Knotenpunkte linear transformiert und durch eine Aktivierungsfunktion geschickt. In der letzten Schicht wird die Ausgabe des Netzes mit dem Zielwert verglichen, um den Verlust zu bestimmen. Das Netz kann mit einer hierauf folgenden Rückwärtspropagation die Parameter einzelner Netzwerkknoten so verändern, dass sich die Ausgabe des Netzes bereits nach geringer Trainingszeit verbessert. Wenn ein Knotenpunkt einer Ebene eine starke Erhöhung der Verlustfunktion verursacht, wird diese Abweichung an die nächsten Knoten weitergeben. Die Rückwärtspropagation geht deshalb von der letzten Ebene (Ausgabebene) rückwärts und optimiert die Parameter, um die Verlustfunktion zu minimieren.



**Abbildung 2.10:** Grafische Darstellung des derzeit für die Energierekonstruktion verwendeten Convolutional Neural Network (CNN)s mit zwei Ebenen [24].

### 2.3.3 Convolutional Neural Networks

Besondere ANNs sind Neuronale Faltungsnetzwerke (CNN) ([19], Kapitel 9). Sie nutzen eine Faltung in mindestens einer Ebene, welche definiert wird als

$$y = w \otimes x + b \quad (2.5)$$

mit der Eingabe  $x$ , der Ausgabe  $y$ , den Parametern des Kernels  $w$  und den Parametern des Bias  $b$ , einer systematischen Verschiebung des Ausgabewertes.

Abbildung 2.10 visualisiert die Architektur eines CNNs. Das Sichtfeld (Field of View (FoV)) bestimmt, wie viele Daten in die Berechnung eines einzelnen Ausgabewertes eingehen. Mit der Dilation kann eben dieses vergrößert werden. Die Parameterzahl bleibt dabei klein, das Sichtfeld wird aber vergrößert.

Das in Abbildung 2.10 dargestellte Netz besteht aus zwei Conv1D-Ebenen zur Energierekonstruktion. Grundsätzlich ist ein Vorteil der CNNs, dass sie einfacher als dichte, voll-vernetzte ANNs sind. Die Architektur des zur Zeit präferierten Netzes für die Energierekonstruktion ist in der folgenden Tabelle dargestellt [25].

	Filtergröße	Kernelgröße	Aktivierungsfunktion	Dilationsrate (DR)
Ebene 1	7	7	ReLU	1
Ebene 2	1	6	ReLU	3

**Tabelle 2.1:** Architektur des derzeit präferierten Standardnetzes mit zwei Ebenen.

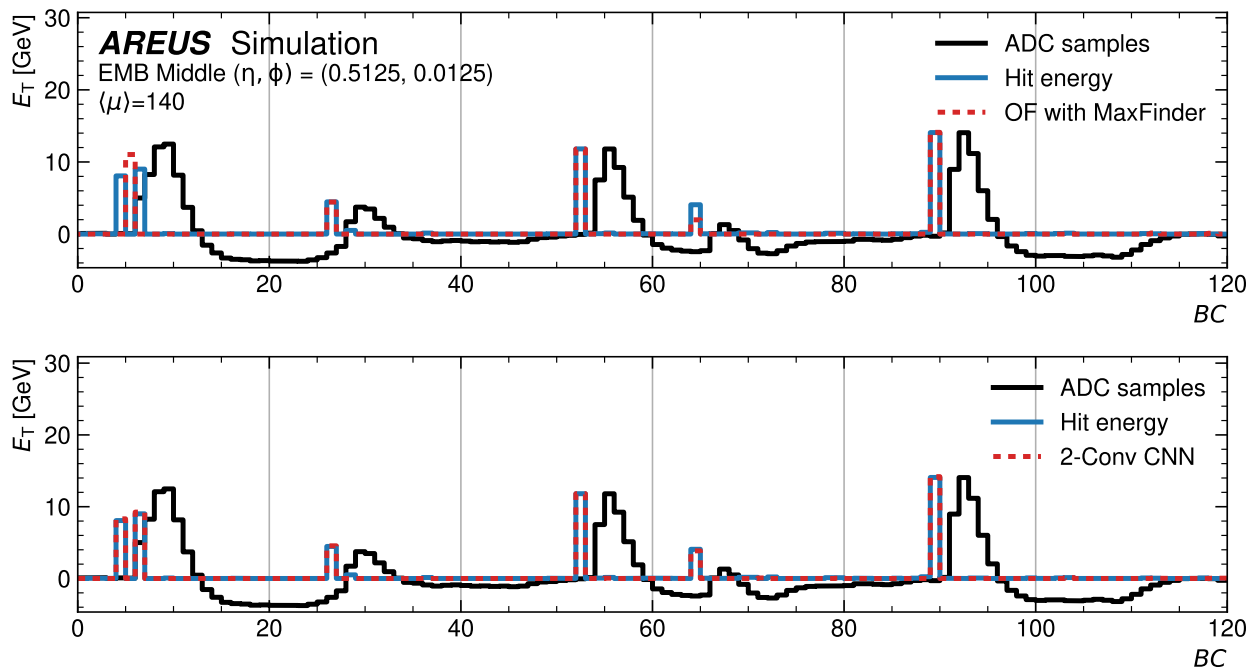
### 2.3.4 AREUS-Simulation

Im oberen Teil der Abbildung 2.11 wird die Energierekonstruktion mit dem derzeit im Kalorimeter genutzten OF mit Maximum Finder dargestellt. Wenn aus drei Samples das mittlere ein Maximum ist, gibt der Maximum Finder dieses aus.

Ansonsten wird eine Null zurückgegeben. Die Ausgabe des OF wird an der simulierten Signalenergie ausgerichtet, um beide besser vergleichbar zu machen. In der unteren Grafik wird die Rekonstruktion eines optimierten CNNs im Vergleich zum oberen mit OF aufgezeichnet.

Die dargestellte Beispielsequenz wurde mit der Simulationssoftware ATLAS Readout Electronic Upgrade Simulation (AREUS) generiert [26]. Erst werden simulierte Energien erzeugt (siehe Abbildung 2.11 in blau), dann werden diese als Puls simuliert. Vorhandenes Pile-up wird mit dem simulierten Puls überlagert und digitalisiert zu ADC-Samples. Mit dieser Methode werden die Datensätze zum Training erzeugt. Sie werden für eine Zelle der mittleren Lage des EMB (EMB-Middle) mit den Parametern  $(\eta, \phi) = (0.5125, 0.0125)$  simuliert. Diese enthalten Pile-up, sowie künstlich hinzugefügte Signal-Energien. Trainiert werden die Netze mit Hilfe eines kombinierten Datensatzes. Die einzelnen Datensätze sind wie folgt aufgebaut:

Es gibt einen Datensatz mit konstantem Abstand von 45 BCs zwischen zwei Signalen, welcher Constant-Gap genannt wird. Ein weiterer Datensatz mit zufallsverteilten (gaußverteilten) Abständen zwischen zwei Signalen heißt Random-Gap. Der Erwartungswert (Mittelwert) der Gaußverteilung der Abstände liegt bei  $\mu = 30$  und seine Standardabweichung bei  $\sigma = 10$ . Für beide Datensätze, Random-Gap und Constant-Gap, gibt es eine Variante mit niedriger Energie und eine mit hoher Energie. Die hohe Energie ist gleichverteilt bis zu 80 % der Gesamtenergie. Das entspricht zirka 150 GeV. Die niedrige Energie ist ebenfalls gleichverteilt bis zu einer Energie von zirka 20 GeV. Die hier simulierte Energie ist nicht die transversale Energie  $E_T$ , sondern die tatsächliche Teilchenenergie. Neben den vier bereits eingeführten Datensätzen gibt es noch einen Datensatz, der lediglich Pile-up enthält. Dementsprechend befinden sich keine größeren Energiedepositionen darin.



**Abbildung 2.11:** Grafische Darstellung der AREUS-Simulation von Teilchenenergien mit ADC [26]. Im oberen Plot werden Signale mit OFMax rekonstruiert, im unteren Plot mit dem Standardnetzwerk [24].

Alle Datensätze haben eine Ausdehnung der Bunches von fünf Zentimetern. Es gibt noch einen weiteren Random-Gap-Datensatz, der eine Bunchausdehnung von 50 cm besitzt. Damit befinden sich in den zusammengesetzten Daten diese sechs Teildatensätze. Mit drei unterschiedlichen Random-Gap-Datensätzen sind sie überproportional repräsentiert. Jeder der genannten sechs Datensätze wird auf eine Länge von 10000 BCs generiert. Von diesem Sample mit 10000 BCs gibt es 2200 Stück je Datensatz. Für das Training haben die Teildatensätze jeweils  $1000 \cdot 10000$  BCs an Samples. Der Rest wird für die Validierung verwendet. Es gibt einen weiteren Datensatz für eine separate Evaluation der Trainings. Dieser hat eine Gleichverteilung der Abstände der Signale zwischen 0 und 70 BCs und heißt Uniform-Gap. Der Datensatz besitzt  $1200 \cdot 10000$  BCs an Samples zum Evaluieren und hat eine Bunchlänge von fünf Zentimetern.

Im Laufe der Arbeit wurde ein siebter Trainingsdatensatz mit dem Namen Custom-Gap simuliert. In diesem folgen zwei Signale direkt aufeinander. Danach sind 38 BCs nur mit Pile-up gefüllt. Insgesamt ist das eine Länge von 40 BCs. Für das Training wurden wieder  $1000 \cdot 10000$  BCs an Samples generiert, zum Validieren  $1200 \cdot 10000$  BCs. Der Datensatz ohne Custom-Gap-Datensatz wird von hier an *Standarddatensatz*, der mit Custom-Gap *erweiterter Datensatz* genannt.

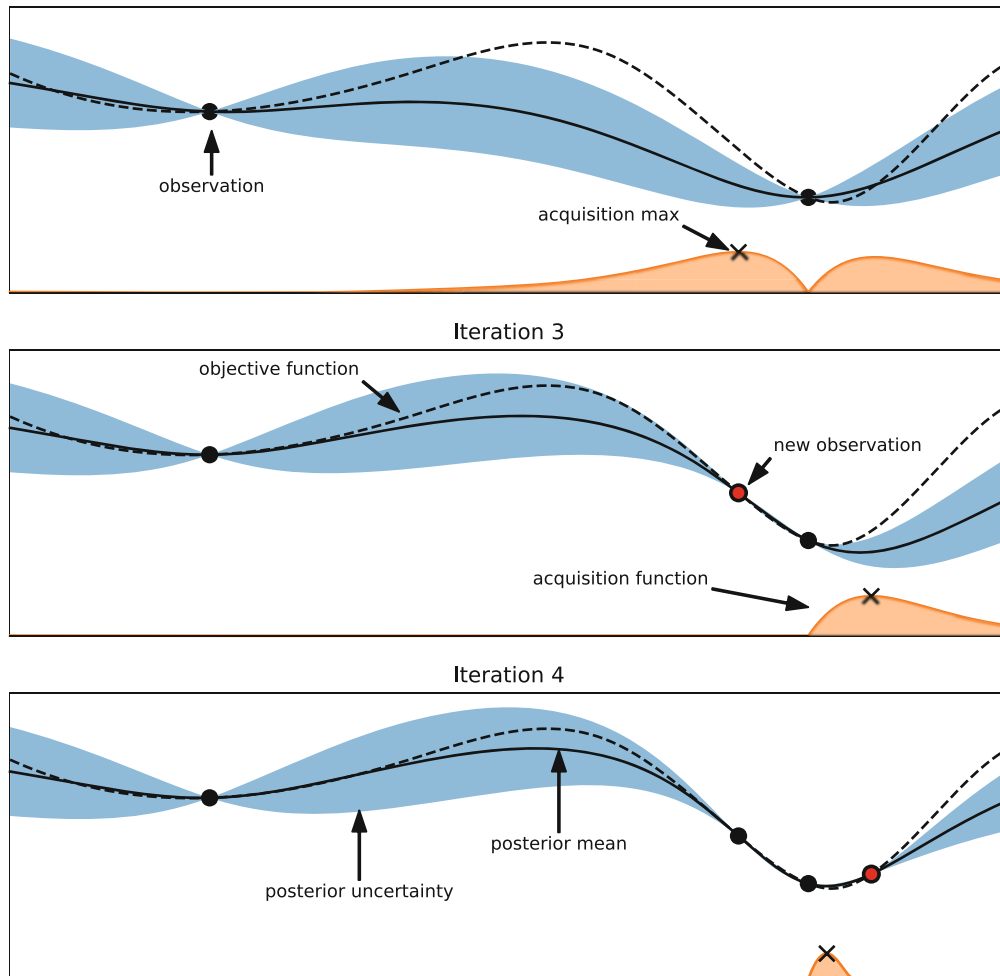
Ein zu optimierendes Verhalten des Standardnetzwerkes ist das bereits erwähnte Überlagern zweier aufeinanderfolgender Ereignisse. Sobald Ereignisse mit großen Abständen voneinander in einer Zelle gemessen werden, kann das Standardnetz und auch der OF die Energie mit geringer Abweichung zum tatsächlichen Wert voraussagen. Je näher sich die Energiedepositionen in einer Zelle allerdings kommen, umso fehleranfälliger wird die Rekonstruktion. Die Länge des Signalpulses hat eine Größe von zirka 25 BC. Wenn zwei hochenergetische Energiedepositionen innerhalb dieses Zeitintervalls auftreten, überlagern sich die Pulse. Dies kann zur Vergrößerung oder Verkleinerung der Amplitude des zweiten Pulses führen. Ersteres tritt auf, wenn sich die beiden Maxima überlagern. Zweiteres tritt dann auf, wenn Maxima und Unterschwinger des Pulses aufeinandertreffen. Die Verbesserung der Performanz im Vergleich zum OF kann das derzeitige Standardnetz bereits gewährleisten.

## 2.4 Hyperparametersuche

Neben den im Training Neuronaler Netze zu optimierenden Gewichten gibt es sogenannte *Hyperparameter*. Diese haben einen besonderen Einfluss auf die Genauigkeit, Effizienz und Rechenzeit eines Netzwerkes. Hyperparameter werden vor dem Training festgelegt. Beispiele für Hyperparameter sind Ebenenzahl, Filter- und Kernelgröße, Aktivierungsfunktionen und Verlustfunktionen des Netzes.

Es stellt sich die Frage, wie optimale Hyperparameter ausgewählt werden können. Die Rechenzeit und Effizienz ist dabei mit einzubeziehen. Für diese Aufgabe gibt es Algorithmen zur Hyperparameteroptimierung. Einfache Implementierungen sind z. B. die *Gittersuche* wie auch die *Zufallssuche*. Die Gittersuche iteriert jede Kombination der Hyperparameter und wählt die beste aus. Hier werden alle Parameterkombinationen durchlaufen und geprüft. Das kostet hohe Rechenzeit und kann unter Umständen auch optimale Hyperparameterwerte verfehlen. Bei der Zufallssuche wird wahllos über alle Kombinationen iteriert. Das kann eine effizientere Methode darstellen [27].

Mit der *Bayes-Optimierung* wird im Gegensatz zur Zufallssuche nicht mehr beliebig gesucht. Dieser Algorithmus sagt aus vorherigen Iterationen Wahrscheinlichkeiten für das nächste mögliche Parameteroptimum voraus. Es wird ein Regressionsmodell für zwei Iterationspunkte erstellt. Aus diesem Modell wird der nächste Iterationspunkt mit dem Maximum der Akquisitionsfunktion geschätzt (siehe Abbildung 2.12). Das kann die Rechenzeit im Vergleich zur Zufallssuche verringern und erhält dabei die Effizienz [28].



**Abbildung 2.12:** Visualisierung der Funktionsweise der Bayes-Optimierung [28].

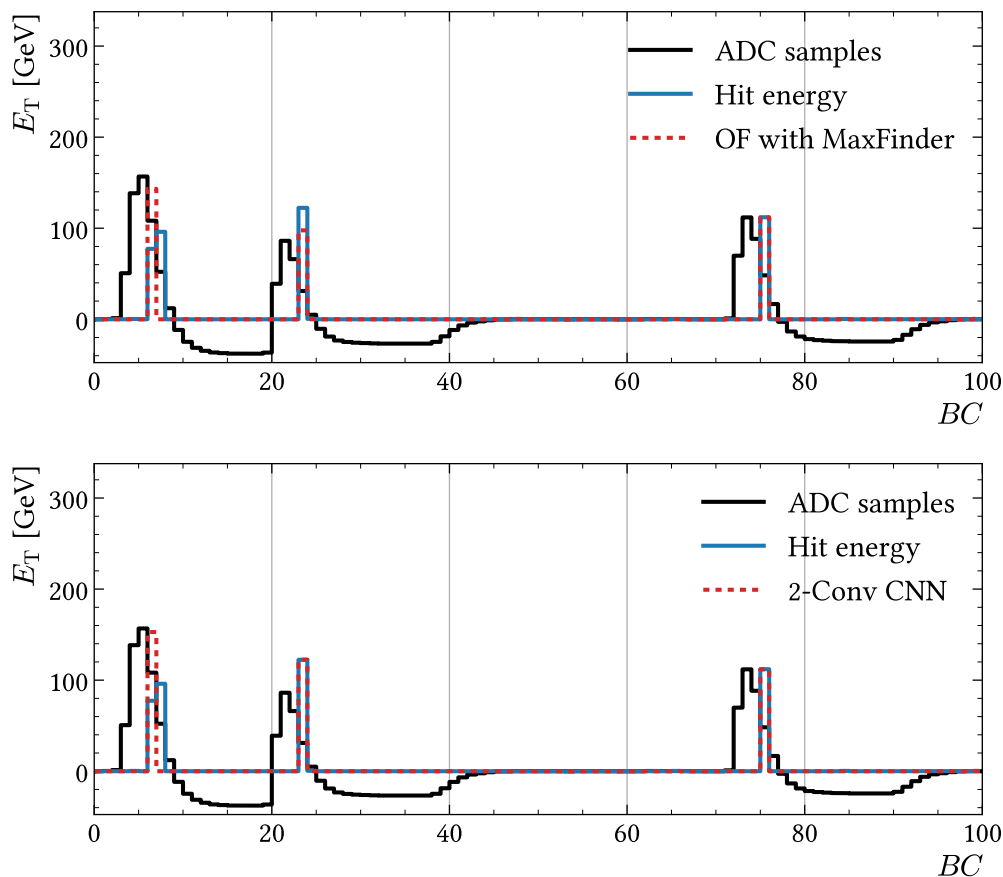
Für diese wissenschaftliche Arbeit wurde der sogenannte *Hyperbandalgorithmus* verwendet. Er nutzt die Zufallssuche mit stetiger Halbierung, *Successive Halving* genannt. Dabei verwendet er den Parameter Rechenbudget  $B$ . Abhängig von diesem wird der Algorithmus mehrfach iteriert. Für eine Iteration werden die Starthyperparameter zufällig gewählt und mit ihnen einige Epochen trainiert. Danach wird die Hälfte der Netze mit schlechterer Genauigkeit verworfen. Nachdem durch alle Iterationen durchlaufen wurde, kann das beste Modell ermittelt werden [29]. Die Implementierung dieses Algorithmus für Hyperparameter der CNNs für die Energierekonstruktion wurde in der Arbeit von P. Welle analysiert [30]. Eine Erkenntnis seiner Analyse ist die Einsparung der Rechenleistung zum Optimieren der Netze im Vergleich zu anderen Algorithmen. Die Modellstruktur wurde in eine Hyperbandsuche implementiert. Im Verlauf dieser Arbeit wird die Architektur der Netze zur Optimierung der One-Gap-Überlappung noch eine Rolle spielen.



### 3 One-Gap-Überlappung

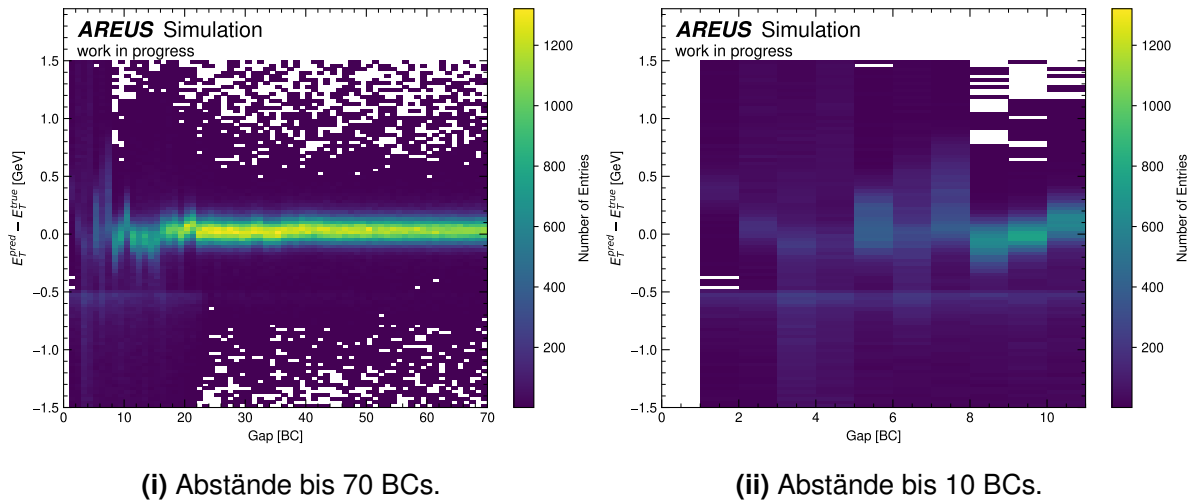
Obwohl die Energierekonstruktionsqualität mit CNNs im Vergleich zum OF bereits wesentlich verbessert werden konnte, bleibt die Energierekonstruktion bei der Überlagerung zweier aufeinanderfolgender Signale auch für die Faltungsnetzwerke eine Herausforderung. Die Abweichung der vorausgesagten Energie  $E_T^{\text{pred}}$  zur tatsächlichen  $E_T^{\text{true}}$ , die ab jetzt  $E_T^{\text{dif}}$  genannt wird, ist bei direkt aufeinanderfolgenden Signalen am größten. Das ist der Fall, wenn nach einer Energiedeposition in einer Zelle ein weiteres Teilchen im nächsten BC in der gleichen Zelle deponiert. Eine Verbesserung der Rekonstruktion aufeinanderfolgender Signale ist wünschenswert, da mit Run 4 das Triggersystem die Aufzeichnung solcher Szenarien ermöglicht. Bisher lag zwischen getriggerten Ereignissen eine Zeitspanne von fünf BCs. Die genaue Ermittlung der Energie  $E_T^{\text{miss}}$  bei allen möglichen Abständen zweier Signale soll im Idealfall möglich sein, da insbesondere Ereignisse mit hoher  $E_T^{\text{miss}}$  BSM-Physik bedeuten können. Die Auseinandersetzung mit dem Szenario der One-Gap-Überlappung ist ein Fokus dieser Arbeit. Im derzeitigen Run sind solche Ereignisse selten. Durch die Erhöhung der Luminosität im HL-Run werden diese häufiger auftreten.

Da der Effekt der falschen Energiezuweisung am stärksten bei der Überlagerung direkt aufeinanderfolgender Signale auftritt, befasst sich die Arbeit mit der Optimierung dieser Ereignisse. Im weiteren Verlauf wird von One-Gap-Überlappungen gesprochen. Dabei handelt es sich um zwei unabhängige Ereignisse in unmittelbar aufeinanderfolgenden BCs. Die Lücke zwischen zwei Ereignissen ist deshalb ein BC breit. Ist diese zwei BCs weit, folgt die Bezeichnung Two-Gap-Überlappung und so fort. Es wird zu Analysezwecken zwischen der wahren und der vorausgesagten (Ausgabe des CNNs) Energie des ersten Signals  $E_T^{(1)\text{true}}$  und  $E_T^{(1)\text{pred}}$  und des zweiten Signals  $E_T^{(2)\text{true}}$  und  $E_T^{(2)\text{pred}}$  unterschieden. Hierbei werden die Einflüsse verschiedener Maßnahmen untersucht und deren Effektivität sowie Genauigkeitsveränderungen anderer Parameter und der Energierekonstruktion überprüft. Daraus resultieren die Fragestellungen, ob ein Neuronales Netzwerk auf überlagerte Messungen optimiert werden und genaue Voraussagen erzeugen kann. Welche Folgen hat eine solche Optimierung und kann sie ohne Performanceeinbußen in anderen Bereichen funktionieren?



**Abbildung 3.1:** Grafische Darstellung von AREUS-simulierten One-Gap-Ereignissen mit ADC. Im oberen Plot werden Signale mit OFMax rekonstruiert, im unteren Plot mit dem Standardnetzwerk [31].

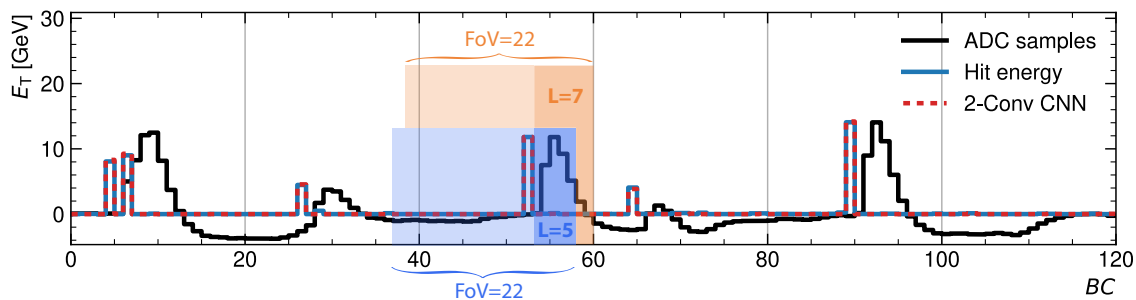
Eine Darstellung eines One-Gap-Ereignisses mit schlechter Rekonstruktion ist in Abbildung 3.1 zu sehen. Im Bereich von 0 bis 10 BCs treffen zwei aufeinanderfolgende Signalenergien  $E_T^{\text{true}}$  (blau) ein. Die ADC-Ausgabe dieser überlagerten Ereignisse besitzt im Vergleich zu den zwei weiteren isolierten Signalen (bei 24 und 76 BCs (blau)) eine größere Breite und eine schneller fallende Flanke. Weiterhin wird das erste Signal überschätzt und das zweite Signal auf eine Energie  $E_T^{\text{pred}} = 0$  GeV (rot) gesetzt und somit unterschätzt. Das Standardnetzwerk rekonstruiert One-Gap-Überlappungen derzeit häufig mit dieser Tendenz. Der umgekehrte Fall trifft jedoch auch ein. In Abbildung 5.5 ist der Fall dargestellt, indem die Energie des ersten Signals  $E_T^{(1)\text{true}}$  klein ist und die des zweiten Signals  $E_T^{(2)\text{true}}$  deutlich größer. Hier wird die vorausgesagte Energie des ersten Signals  $E_T^{(1)\text{pred}} = 0$  GeV und die zweite  $E_T^{(2)\text{true}}$  überschätzt. Die CNNs haben zur Zeit die Tendenz, bei einem bestimmten Verhältnis der wahren Signalenergien eines der Signale auf 0 GeV zu setzen und das andere Signal zu überschätzen. Häufiger kommt die Überschätzung des ersten Signals vor.



**Abbildung 3.2:** Häufigkeiten von  $E_T^{\text{dif}}$  über den Abständen zwischen zwei Signalen. Das Standardnetz erzeugte die Energie  $E_T^{\text{pred}}$ .

Die Qualität der Energierekonstruktion wird mit Hilfe der Abbildung 3.2 dargestellt. Die Grafik zeigt die Differenz zwischen vorhergesagter und wahrer deponierter Energie in Abhängigkeit der Lücke zweier aufeinanderfolgender Signale. Als Signal zählen hierbei Energiedepositionen mit  $E_T^{\text{true}} > 0.5$  GeV. Die One-Gap-Überlappungen befinden sich bei einem Abstand von eins. Der linke Teil der Abbildung 3.2 (i) zeigt eine höhere Dichte von Ereignissen um den Nullpunkt von  $E_T^{\text{dif}}$ . Das Signal hat eine Länge von 25 BCs. Ab Abständen über 22 BCs ist die Verteilung um den Nullpunkt zentriert. Es gibt Ereignisse, die höhere Energieabweichungen haben, aber deren Häufigkeit ist geringer. Bei Abständen kleiner als 22 kann man die immer schlechter werdende Energierekonstruktion beobachten. Das ist der Bereich, indem sich zwei Signale überlappen. Das Verschwinden der Ereignishäufung um den Nullpunkt von  $E_T^{\text{dif}}$  (gelb-grünes Band) bei kleinen Abständen, besonders bei One-Gap-Ereignissen, ist der zu optimierende Effekt.

Im rechten Teil der Abbildung 3.2 (ii) ist eine feiner aufgelöste Achse für die Abstände visualisiert. Hier kann die Energierekonstruktion kleiner Abstände zwischen zwei Ereignissen besser beobachtet werden. Alle ANNs zeigen eine Häufung an Ereignissen bei  $E_T^{\text{dif}} = -0.5$  GeV. Diese entsteht durch einen Energiecut von 0.5 GeV, um das Pile-up aus der Grafik zu entfernen. Da jedes Signal Pile-Up enthält, würden ansonsten alle Signale eine Lücke von einem BC haben. In der Abbildung ist  $E_T^{\text{dif}}$  weit gestreut.  $E_T^{\text{pred}}$  wird dementsprechend nicht gut rekonstruiert. Im Optimalfall sollte das Netzwerk eine Verteilung um  $E_T^{\text{dif}} = 0$  für alle Abstände zwischen zwei Signalen rekonstruieren.



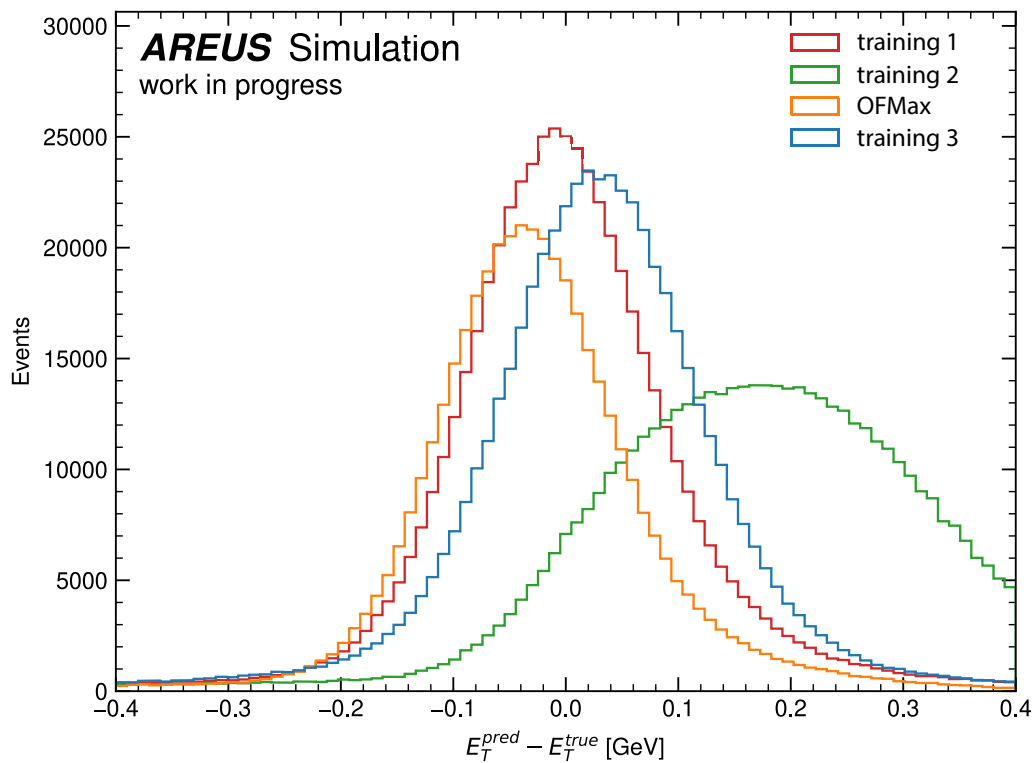
**Abbildung 3.3:** Energierekonstruktion eines 2Conv-Netzwerkes mit Veranschaulichung der Rekonstruktionslatenz (L) [24].

### 3.1 Einfluss der Rekonstruktionslatenz

Eine Überlegung zur Optimierung ist die Tatsache, dass ein überlagerter bipolarer Puls breiter und höher ist als der Puls eines einzelnen Signals. Daher wird als erster Ansatz das FoV verschoben und die Rekonstruktionslatenz (RL) erhöht. Bei Anwendung der modifizierten Methode mit erhöhtem FoV und RL wird eine Erkennung des breiteren Pulses erwartet, sodass die Qualität der Signalverarbeitung verbessert wird. Im Normalfall nutzt das CNN fünf BCs um das Maximum des Signals. Wenn die RL um ein oder zwei BCs erhöht wird, wird die abfallende Flanke des Signals erfasst, damit das Netz den Zusammenhang zwischen zwei Pulsen erkennt. Zusätzlich soll die richtige Energie den jeweiligen Signalen zugeordnet werden.

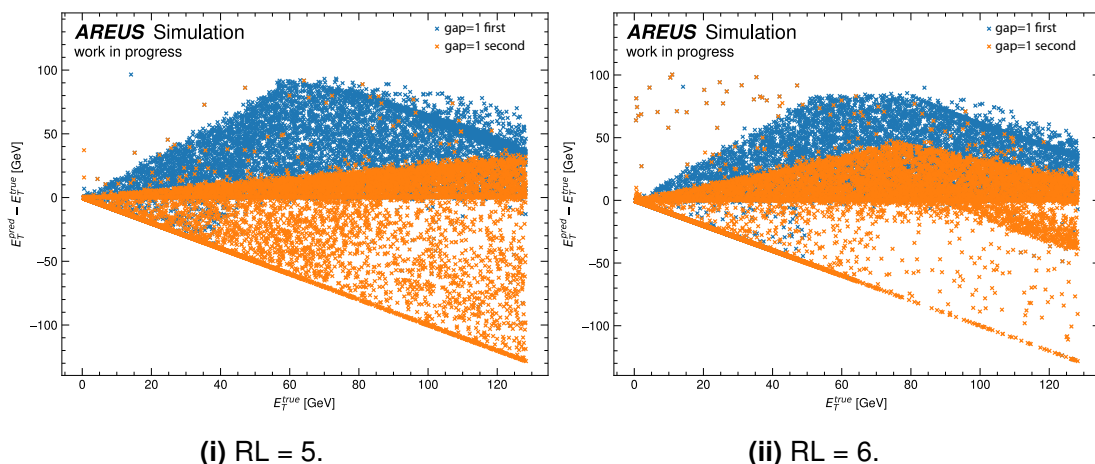
In Abbildung 3.3 wird die RL dargestellt. Der blaue Bereich (hellblau inklusive dunkelblau) ist die Konfiguration des Standardnetzes mit FoV von 22 und der RL von fünf (dunkelblau). Der orange Bereich stellt ein Netz dar, bei welchem die RL um zwei BC erhöht wurde. Dabei ist die Größe der RL sieben BCs (dunkelorange) und das FoV noch immer 22 BCs. Das FoV wurde hier um zwei BCs verschoben. Die Latenz erhöht sich um 50 Nanosekunden. Das derzeitige Triggersystem hat eine solche Latenz nicht implementiert. Durch die Vergrößerung dieser Parameter wird ein breiterer Bereich vom Netz zeitgleich verarbeitet. Damit wird bei One-Gap-Überlappungen ein größerer Teil der Signale berücksichtigt. Breitere Signale sollten besser erkannt werden.

Mit diesen Überlegungen ist die RL implementiert. Für alle Ansätze der Ausarbeitung sind jeweils zwanzig Trainings mit 100 Epochen angefertigt. Konkret sind hier die besten Modelle für die Architektur des Standardnetzes mit der Optimierung für eine RL von fünf, sechs und sieben angefertigt.



**Abbildung 3.4:** Ereignisanzahl über  $E_T^{\text{dif}}$  für die drei besten Trainingsmodelle des Standardnetzes mit RL von fünf.

Zusätzlich wurden Grafiken äquivalent zu Abbildung 3.2 und auch Histogramme analysiert, wie das in Abbildung 3.4 dargestellte. Bei diesem ist sofort erkennbar, dass z. B. beim Modell des Trainings 2 (grün) Underfitting auftritt. Die Verteilung ist zum Nullpunkt von  $E_T^{\text{dif}}$  stark verschoben. Weiterhin ist die Standardabweichung, also die Breite der Verteilung, größer als die der anderen dargestellten Netze. Ein optimales Netzwerk für die Energierekonstruktion hätte eine um den Nullpunkt verteilte Häufung mit maximaler Ereignisanzahl und würde einer Delta-Distribution gleichen. Realistischer wäre eine schmale Gaußkurve mit niedriger Standardabweichung und großer Ereignisanzahl bei  $E_T^{\text{dif}} = 0$ . Im besten Fall hat ein Netzwerk eine geringe Standardabweichung und einen Mittelwert, welcher sich um den Nullpunkt zentriert. Mit Hilfe des Histogramms kann schnell die Performanz der Energierekonstruktion verschiedener CNNs und des OFs miteinander verglichen werden. Dies wird für die Evaluation der Performanzparameter genutzt. Im Laufe der Ausarbeitung wurde der Verlauf der Verlustfunktion MAE und des Validierungsverlustes jedes einzelnen Trainingsmodells ermittelt. Daraus erfolgte die Auswahl des besten Modells. Da die Optimierung der Trainings über die Minimierung der Verlustfunktion stattfand, erscheint das eine objektive und schnelle Wahl zur Bestimmung bester Trainings zu sein.



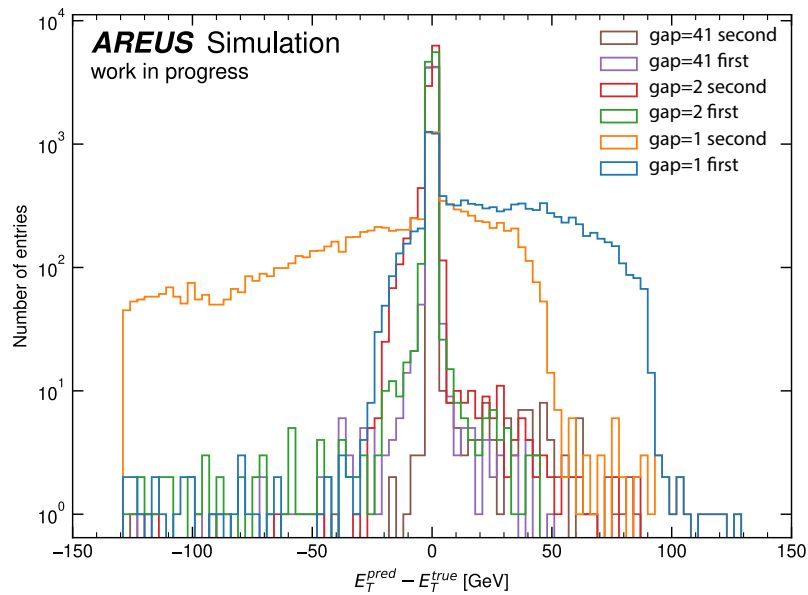
**Abbildung 3.5:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{true}}$  über  $E_T^{\text{dif}}$  für das erste und zweite Signal mit Standardnetz und unterschiedlicher RL.

### 3.1.1 Performanz der Rekonstruktion von One-Gap-Überlappungen

Mithilfe des in Abschnitt 2.3.4 erläuterten Uniform-Gap-Datensatzes wurden zur Analyse der Rekonstruktion von One-Gap-Überlappungen Evaluationsanalysen mit dem Standardnetz mit der RL fünf, sechs und sieben durchgeführt. Die Ergebnisse wurden für alle Ereignisse mit den Lücken von einem BC maskiert. Eine Analysedarstellung ist in Abbildung 3.5 für die RL von fünf und sechs dargestellt. Aufgetragen werden hierbei  $E_T^{\text{dif}}$  über  $E_T^{\text{true}}$  jeweils für das erste und zweite Signal für One-Gap-Überlappungen. Mehrere Effekte und komplexe Strukturen fallen in der Rekonstruktion auf.

Im Optimalfall sollte das CNN  $E_T^{\text{pred}}$  genau rekonstruieren, also nahe Werte um  $E_T^{\text{true}}$  schätzen. Dann würde sich das erste Signal (blau) um den Nullpunkt von  $E_T^{\text{dif}}$  streuen. Gleiches würde mit dem zweiten, auf das erste folgende Signal, in orange dargestellt, stattfinden. Das bedeutet, dass das Netzwerk beiden Signalen die richtige Energie zuweist. Im folgenden wird die derzeitige Energierekonstruktion für One-Gap-Überlappungen mit dem Standardnetz erörtert.

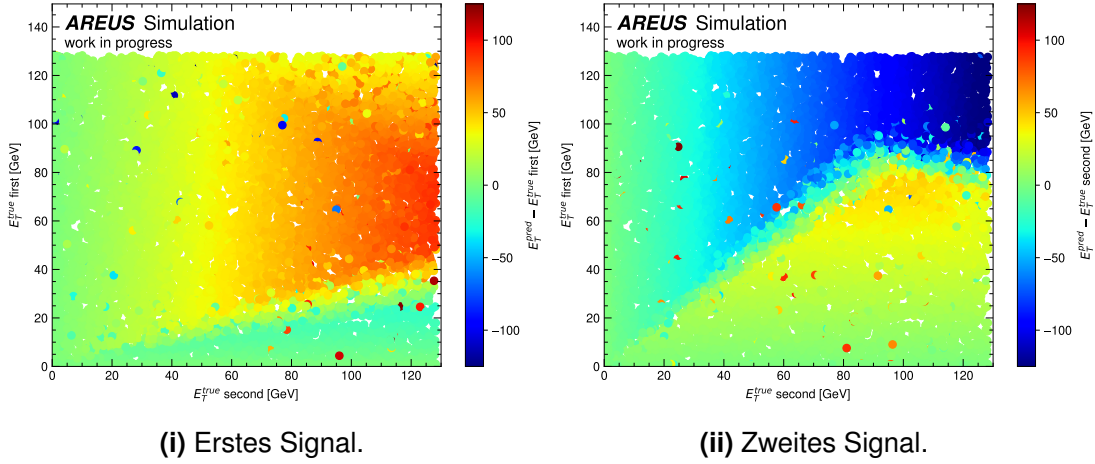
Das zweite Signal (orange) hat eine Ereignishäufung auf einer Diagonalen mit negativem Anstieg. Das sind die Ereignisse, welche durch das Netzwerk mit  $E_T^{\text{pred}}$  von Null rekonstruiert werden. Die Einträge in der Diagonale mit positivem Anstieg sind überschätzte Energien  $E_T^{\text{pred}}$ . Das Netzwerk hat die Tendenz, das erste Signal zu überschätzen und das zweite Signal zu unterschätzen. Es müsste folglich lernen, diese Tendenz zu korrigieren. Dabei sollte es zu keiner Verschlechterung der Performanz des Netzes kommen.



**Abbildung 3.6:** Ereignisanzahl über  $E_T^{\text{dif}}$  in logarithmischer Skala für das erste und zweite Signal für One-Gap-, Two-Gap- und 41-Gap-Überlappungen mit Standardnetz mit RL sechs.

In Abbildung 3.6 ist  $E_T^{\text{dif}}$  in logarithmischer Skala dargestellt. Dabei wird unterschieden zwischen dem ersten und zweiten Signal und für beide jeweils One-Gap-, Two-Gap- und 41-Gap-Überlappungen. Mit dieser Darstellung wurde die Performanz der Energierekonstruktion in unterschiedlichen Teilbereichen analysiert. Die Diskrepanz der wahren Energie  $E_T^{\text{true}}$  des ersten und des zweiten Signals kann so verglichen werden. Dabei erzeugt die Energierekonstruktion häufig Asymmetrien zwischen beiden Signalen. Die Tendenz der Überschätzung des ersten Signals (blau) und der Unterschätzung des zweiten Signals (orange) ist im Training mit dem Standarddatensatz erkennbar.

Das mit dem Standarddatensatz trainierte Standardnetz mit RL von sechs kann z. B. die 41-Gap-Überlappungen für das erste (lila) und zweite Signal (braun) gut rekonstruieren. 41-Gap-Überlappungen sind die Ereignisse, bei denen zwei Signale unabhängig voneinander gemessen werden und sich nicht mehr überlagern. Das Netzwerk rekonstruiert dementsprechend isoliert auftretende Signale gut. Für Two-Gap-Überlappungen kann jedoch eine Verbreiterung der Verteilung beobachtet werden. Die One-Gap-Überlappungen haben viele Ereignisse mit großen Werten von  $E_T^{\text{dif}}$ . Im Optimalfall sollte eine Maßnahme, wie beispielsweise die Veränderung der Netzwerkarchitektur oder des Trainingsdatensatzes, Ereignisse der One-Gap-Überlappungen nahe zum Nullpunkt von  $E_T^{\text{dif}}$  rücken. Wünschenswert ist, dass die Two-Gap- und 41-Gap-Überlappungen hinsichtlich ihrer Verteilungen mindestens gleich bleiben.

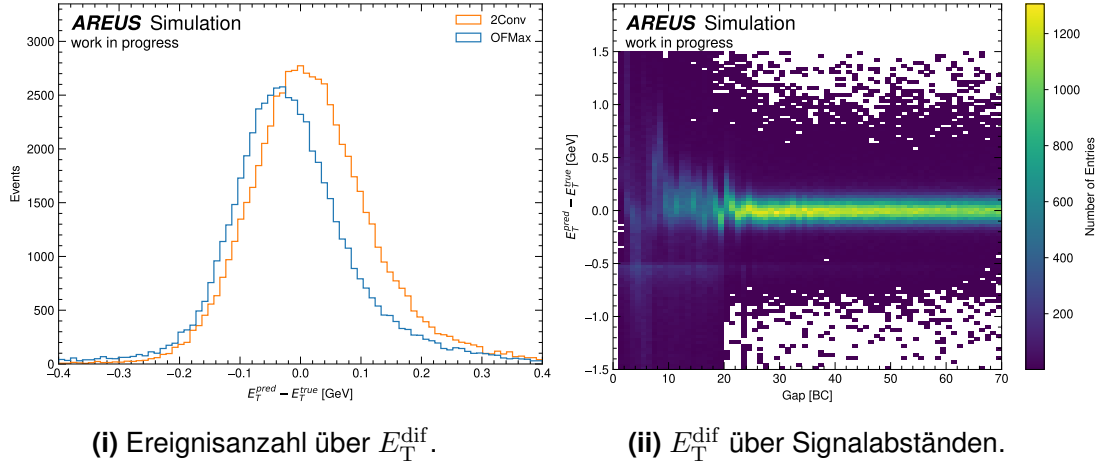


**Abbildung 3.7:**  $E_T^{\text{dif}}$  in Abhängigkeit von  $E_T^{(1)\text{true}}$  und  $E_T^{(2)\text{true}}$  für alle One-Gap-Überlappungen mit einer RL von sechs.

Abbildung 3.7 visualisiert die Energiedifferenz  $E_T^{\text{dif}}$  des ersten Signals (i) und des zweiten Signals (ii) in Abhängigkeit der wahren Energie  $E_T^{\text{true}}$  des ersten und des zweiten Signals. Die Trainingsmodelle wurden mit einer RL von sechs trainiert.

Die Darstellung aus Abbildung 3.7 ermöglicht eine bessere Erkennung der Übergänge zwischen besser und schlechter rekonstruierten Bereichen im Vergleich zur Abbildung 3.6. Eine optimale Rekonstruktion würde für die Darstellungen für das erste und zweite Signal lediglich grüne Einträge, also jene nahe dem Nullpunkt von  $E_T^{\text{dif}}$ , erzeugen. Deshalb sollte die Grafik nahezu einfarbig sein. Die Grafiken für das erste und zweite Signal des mit dem Standarddatensatz trainierten Standardnetzes zeigen jedoch starke Abweichungen von diesem Idealbild. Abbildung 3.7 (i) stellt die One-Gap-Überlappungen mit der Abweichung  $E_T^{(1)\text{dif}}$  des ersten Signals dar. Es gibt darin zwei getrennte Bereiche. Von einer wahren Energie  $E_T^{\text{true}}$  beider Signale von 0 GeV ist ein Grenzübergang bis zu zirka  $E_T^{(1)\text{true}} \approx 47$  GeV und  $E_T^{(2)\text{true}} \approx 130$  GeV erkennbar. Wenn beispielsweise das erste Signal eine Energie von  $E_T^{(1)\text{true}}$  von Null hat und  $E_T^{(2)\text{true}}$  eine beliebige Energie, dann kann das Netz gut rekonstruieren. Gleiches gilt für den umgekehrten Fall, indem  $E_T^{(2)\text{true}}$  gleich Null ist, da einzelne Signale von den CNNs erkannt werden. Ist aber zum Beispiel die Energie des ersten Signals  $E_T^{(1)\text{true}} \approx 80$  GeV und das zweite Signal hat eine annähernd gleich hohe oder höhere Energie, dann weist das Netz in der Regel dem Signal mit höherer Energie die gesamte Energie zu. Der Grenzübergang entsteht durch eben diesen Effekt. Gleichzeitig besteht noch die Tendenz der Überschätzung des ersten Signals und der Unterschätzung des zweiten Signals, welche in Abbildung 3.6 eindeutiger zu erkennen ist.



(i) Ereignisanzahl über  $E_T^{\text{dif}}$ .(ii)  $E_T^{\text{dif}}$  über Signalabständen.**Abbildung 3.8:** Analysedarstellungen für die Evaluation mit einer RL von sechs.

Der Grenzübergang wird im Gegensatz zur linken Grafik in Abbildung 3.7 (i) der Abweichung  $E_T^{(2)\text{dif}}$  des zweiten Signals nicht unterbrochen. Es gibt einen Punkt bei  $E_T^{(2)\text{true}} \approx 95$  GeV und  $E_T^{(1)\text{true}} \approx 87$  GeV, ab dem der Grenzübergang nicht durchgehend ist und wieder sinkt. Das liegt an der Übersättigung beider Signale, welche insgesamt einen Wert von zirka 182 GeV überschreiten. Durch die Saturationsdetektion würde in den Low-Gain gewechselt werden und dieses Absinken damit verschwinden, sodass der Effekt in der Realität nicht sichtbar ist. Der Grenzbereich würde in der echten Messung durch diesen Wechsel weiter ansteigen. In dieser Darstellung tritt der Effekt nur für das zweite Signal auf.

### 3.1.2 Allgemeine Performanz der Energierekonstruktion

Die Veränderungen mit der Erhöhung der RL auf sechs und sieben sind in Abbildung 5.6 dargestellt. Für RL sechs hat das erste Signal marginal niedrigere Höchstwerte für  $E_T^{\text{dif}}$ . Die Dichte der Ereignisse im negativen Bereich von  $E_T^{\text{dif}}$  für das zweite Signal hat sich etwas ausgedünnt. Diese Ereignisse liegen im zweiten Bereich mit positiven  $E_T^{\text{dif}}$ , welcher sich verbreitert hat. Hier kann keine signifikante Verbesserung festgestellt werden. RL sieben hingegen hat zumindest im zweiten Signal eine direkt sichtbare Veränderung. Ein dichter Teil der Ereignisse im negativen Bereich von  $E_T^{\text{dif}}$  ist auf höhere Werte verschoben. Diese Werte sind allerdings nicht genau und nur marginal besser. Bei Überprüfung anderer Modelle mit RL sechs und sieben konnte auch festgestellt werden, dass solche Rekonstruktionen zufällig auftreten. Ob diese mit der Veränderung der RL zusammenhängen, kann nicht geschlussfolgert werden. Es könnte sich dabei lediglich um stochastische Effekte des Trainings der Neuronalen Netze handeln.

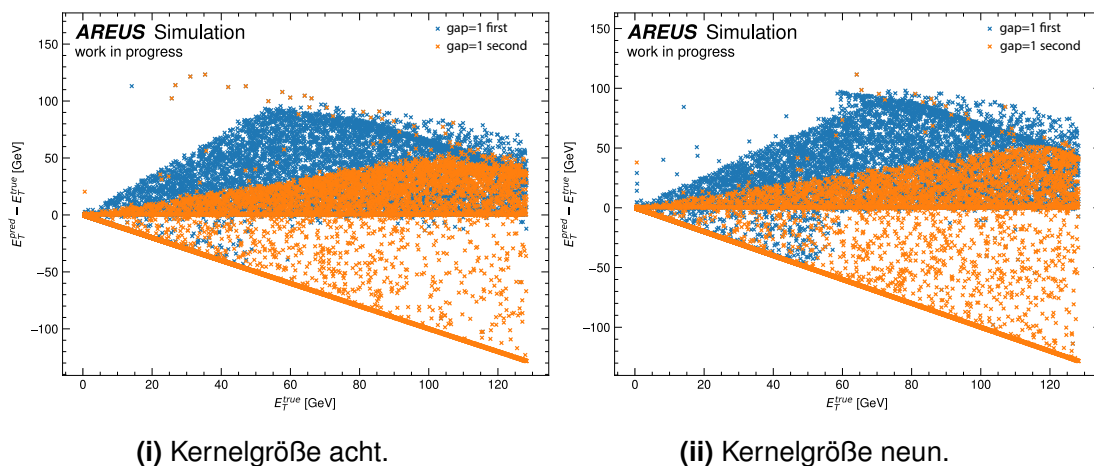
In den beiden Abbildungen 3.8 (i) und (ii) kann die allgemeine Performanz der Energierekonstruktion des CNN evaluiert werden. Die im Abschnitt (i) dargestellte Verteilung hat nur marginale Veränderungen zum OF und rekonstruiert die Energie gut. In Abschnitt (ii) der Abbildung verschlechtert sich die Performanz der Energierekonstruktion im Bereich von Abständen zwischen zwei Signalen kleiner als 22 BCs. Im Vergleich zum Standardnetz gibt es aber keine signifikanten Verbesserungen oder Verschlechterungen. Die One-Gap-Überlappungen können mit der Veränderung der RL allein nicht besser rekonstruiert werden.

Zusammenfassend kann gesagt werden, dass die Neuronalen Netzwerke die breiteren Signale von One-Gap-Überlappungen durch ein Verändern der RL nicht erlernt haben. Auch die Reproduzierbarkeit marginaler Verbesserungen in der Performanz ist nicht gegeben. Viele der trainierten Modelle hatten minimale bis keine genaueren Energierekonstruktionen und im schlechtesten Falle sogar Einbußen in anderen wünschenswerten Eigenschaften der Rekonstruktion. Da der ursprüngliche Ansatz für die Verbesserung der Genauigkeit von One-Gap-Überlappungen keine signifikanten Ergebnisse erzielte, wurde der Ansatz der RL überarbeitet und angepasst.

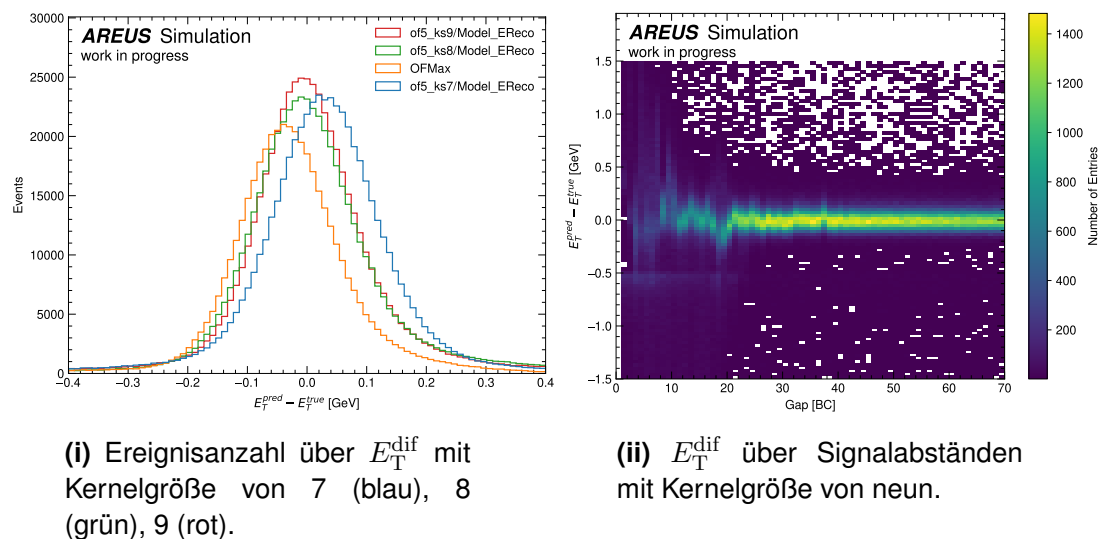
## 3.2 Variation von Rekonstruktionslatenz und Kernelgröße

Mit dem Erhöhen der RL bei standardmäßiger Kernelgröße von sieben wurde eine verbesserte Erkennung des Signalmaximums durch das CNN erwartet. Das Erhöhen der RL um bis zu zwei BCs führte allerdings zu keiner signifikanten Rekonstruktionsverbesserung. Eine Erhöhung der Kernelgröße geht mit einer Erhöhung des FoV einher. Das würde zu einer besseren Vergleichbarkeit mit der Rekonstruktion des Standardnetzes führen. Eine höhere Kernelgröße bedeutet jedoch eine ebenfalls erhöhte Parameteranzahl des CNNs.

Die nächste Überlegung nach der Vergrößerung des Inputs um bis zu zwei Zeiteinheiten ist die zusätzliche Erhöhung der Kernelgröße des Standardnetzwerkes. Um die Veränderung der Kernelgröße und der RL systematisch zu analysieren, wird jede Kombination der Kernelgröße und RL miteinander verglichen, die modifizierten Netzwerke trainiert und evaluiert. Die gewählten Kernelgröße sind sieben, acht und neun und die RL wie im vorherigen Ansatz fünf, sechs und sieben.



**Abbildung 3.9:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{dif}}$  über  $E_T^{\text{true}}$  für das erste und zweite Signal mit Standardnetz mit RL von fünf und angepasster Kernelgröße.



**Abbildung 3.10:** Analysedarstellungen für die Evaluation mit einer RL von fünf.

Der bisherige Standard hat eine RL von fünf und eine Kernelgröße von sieben. Es entstehen dementsprechend sechs weitere Variationen der Netzwerke. ANNs mit Kernelgröße sieben und jeweils mit RL fünf, sechs und sieben existieren bereits und einige Ergebnisse sind in den Abbildungen 3.5 und 5.6 dargestellt. Für Kernelgröße acht und neun wurden weitere Netze trainiert.

Teile der neuen Ergebnisse sind in den Abbildungen 3.9, 5.7 und 5.8 dargestellt. Auch hier konnte keine signifikante Verbesserung der Genauigkeit in der Energierekonstruktion der One-Gap-Überlappung erzielt werden. Bei der Analyse der Grafiken erkennt man leichte Variationen. Verbesserte Rekonstruktionen wurden aber nicht erreicht.

Auch die logarithmische Darstellung in Abbildung 5.9 zeigt keine Verbesserung. In Abbildung 5.10 (i) wird die Übersättigungsdetektion diesmal im ersten Signal erreicht. Ansonsten ist keine signifikante Veränderung eingetreten.

In Abbildung 3.10 (i) sind alle Veränderungen der Kernelgröße in einem Histogramm dargestellt. Die grobe Verteilung hat sich nicht maßgeblich geändert. Auch die in Abbildung 3.10 (ii) dargestellte Anzahl der Einträge hat sich im Bereich der Abstände kleiner als 22 BCs nicht verändert. Vergleichbare Analysedarstellungen zu den in Abschnitt 3.2 angerissenen für eine Kernelgröße von neun sind im Anhang in den Abbildungen 5.9 und 5.10 hinterlegt. Minimale Veränderungen jener Darstellungen könnten auf statistische Varianz zurückzuführen sein.

### 3.3 Hyperparameter und erweiterter Datensatz

Das reine Verändern der Architektur mittels Anpassung des FoV und der RL führte nicht zum gewünschten Erfolg. Deshalb wurde als nächster Ansatz der Trainingsdatensatz verändert. Zum vorherigen Datensatz wurde ein neuer hinzugefügt. Dieser ist ausschließlich mit One-Gap-Ereignissen gefüllt, wie in in Abschnitt 2.3.4 beschrieben. Zusätzlich wurde die Parameterzahl der CNNs auf bis zu 400 erhöht. Daraufhin wurde eine Hyperparametersuche für Netze mit jeweils zwei und drei Ebenen mit RL fünf, sechs und sieben und den maximalen 400 Netzwerkparametern durchgeführt. Verwendung fand der in Abschnitt 2.4 erläuterte Hyperbandalgorithmus.

	Filtergröße	Kernelgröße	Aktivierungsfunktion	DR
Ebene 1	33	7	ReLU	1
Ebene 2	1	4	ReLU	1

**Tabelle 3.1:** Beste Architektur mit zwei Ebenen für RL fünf.

	Filtergröße	Kernelgröße	Aktivierungsfunktion	DR
Ebene 1	28	7	ReLU	1
Ebene 2	1	6	ReLU	1

**Tabelle 3.2:** Beste Architektur mit zwei Ebenen für RL sechs.

	Filtergröße	Kernelgröße	Aktivierungsfunktion	DR
Ebene 1	26	12	ReLU	1
Ebene 2	1	2	ReLU	1

**Tabelle 3.3:** Beste Architektur mit zwei Ebenen für RL sieben.

	Filtergröße	Kernelgröße	Aktivierungsfunktion	DR
Ebene 1	6	12	ReLU	1
Ebene 2	8	5	ReLU	1
Ebene 3	1	9	ReLU	1

**Tabelle 3.4:** Beste Architektur mit drei Ebenen für RL fünf.

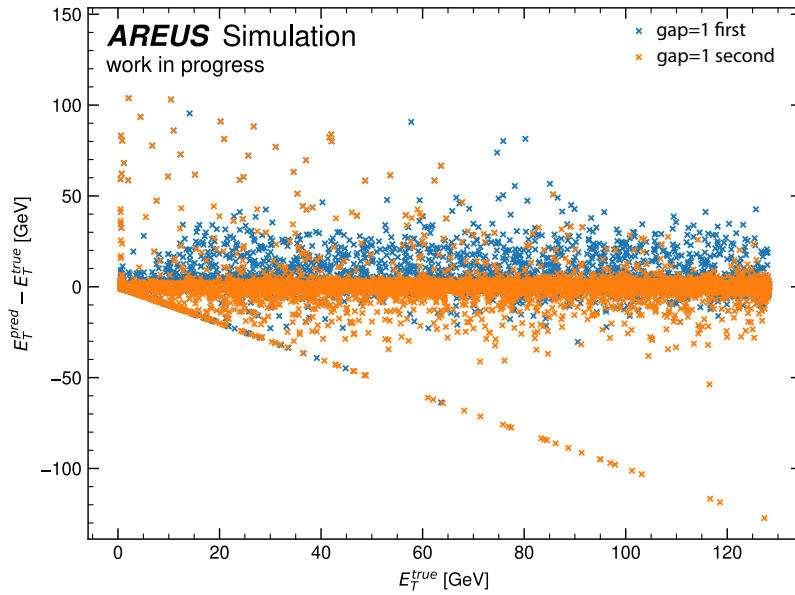
	Filtergröße	Kernelgröße	Aktivierungsfunktion	DR
Ebene 1	5	14	ReLU	1
Ebene 2	19	2	ReLU	1
Ebene 3	1	6	ReLU	1

**Tabelle 3.5:** Beste Architektur mit drei Ebenen für RL sechs.

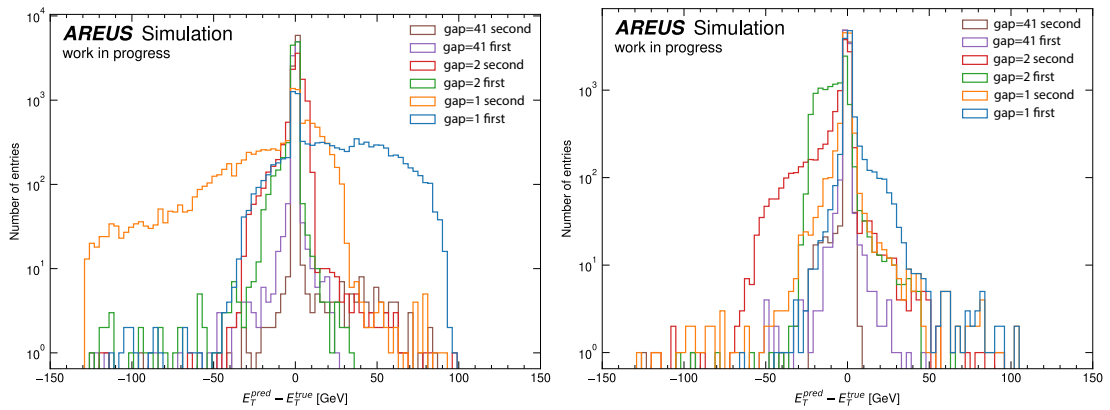
	Filtergröße	Kernelgröße	Aktivierungsfunktion	DR
Ebene 1	14	17	ReLU	1
Ebene 2	4	2	ReLU	1
Ebene 3	1	7	ReLU	1

**Tabelle 3.6:** Beste Architektur mit drei Ebenen für RL sieben.

Für die sechs Netzwerkstrukturen aus der Hyperbandsuche und für das Standardnetz mit RL fünf, sechs und sieben wurden wieder 20 Trainings mit 100 Epochen durchgeführt. Die Trainings wurden einmal mit Standarddatensatz und einmal mit erweitertem Datensatz angefertigt. Im Vergleich zu vorherigen Ansätzen konnten eindeutige Veränderungen in der Rekonstruktion der Neuronalen Netze festgestellt werden.



**Abbildung 3.11:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{dif}}$  über  $E_T^{\text{true}}$  für das erste und zweite Signal mit dem Standardnetz mit RL fünf trainiert mit erweitertem Datensatz.



**(i)** Training mit Standarddatensatz.

**(ii)** Training mit erweitertem Datensatz.

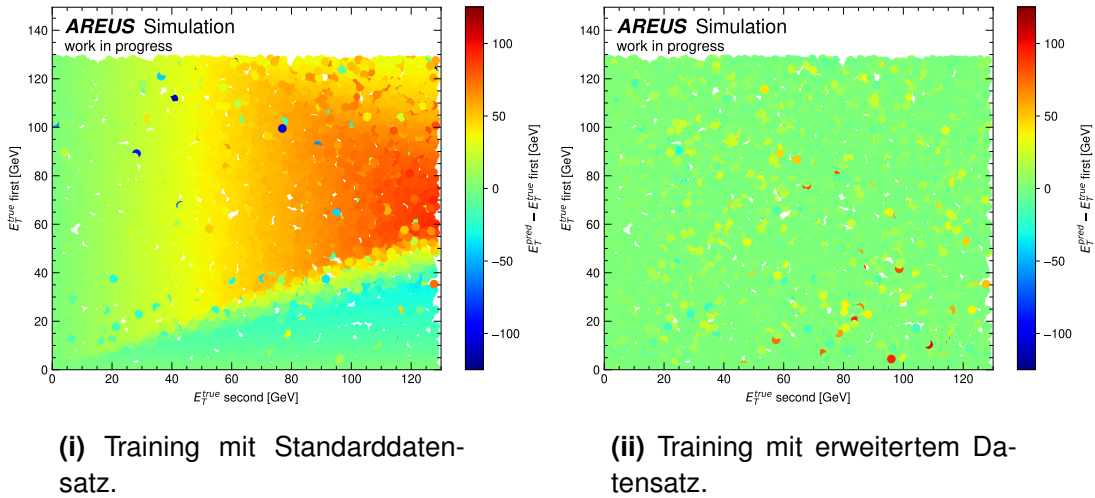
**Abbildung 3.12:** Ereignisanzahl über  $E_T^{\text{dif}}$  in logarithmischer Skala für das erste und zweite Signal für One-Gap-, Two-Gap- und 41-Gap-Überlappungen mit Standardnetz mit RL fünf.

### 3.3.1 Performanz der Rekonstruktion von One-Gap-Überlappungen

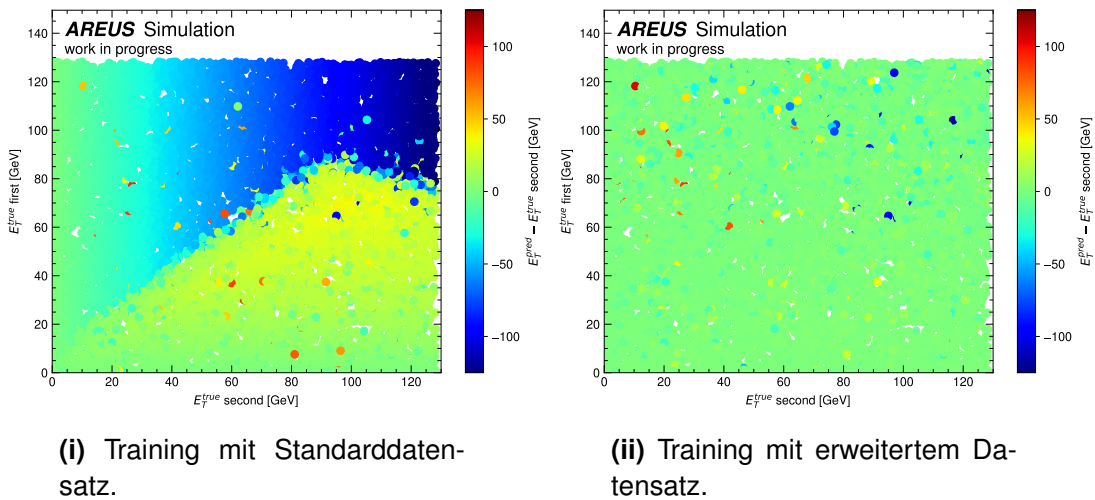
Der größte Einfluss konnte mit dem erweiterten Datensatz erzeugt werden. Der Vergleich von Abbildung 3.11 mit Abbildung 3.5 zeigt eine enorme Verbesserung der Energierekonstruktion. Ein großer Anteil der Ereignisse befindet sich um den Nullpunkt von  $E_T^{\text{dif}}$ . Ausreißer sind noch vorhanden. Falsche Energien für das zweite Signal werden bei einer wahren Energie  $E_T^{\text{true}} = 0$  erzeugt. Kein Training mit dem Standarddatensatz hat eine bessere Rekonstruktion der One-Gap-Überlappungen erreichen können. In dieser visuellen Darstellung kam die Architektur mit zwei Ebenen mit RL sechs, welche mit dem erweiterten Datensatz trainiert wurde, dem Optimalfall am nächsten (siehe Abbildung 5.11).

Das Training mit dem erweiterten Datensatz hat bei allen Architekturen den größten Einfluss für eine gute Rekonstruktion der One-Gap-Überlappungen. Weder die modifizierte Architektur durch den Hyperbandalgorithmus oder die RL hatten für die Performanz dieser Darstellung einen merklichen Einfluss. Diese marginalen Veränderungen könnten lediglich durch statistische Varianz entstanden sein.

Bei der Analyse des Standardnetzes in Abbildung 3.12, welches jeweils mit dem Standarddatensatz und erweitertem Datensatz trainiert wurde, sind spezifische Veränderungen auffällig. Beim Training mit dem erweiterten Datensatz rücken die One-Gap-Ereignisse des ersten und zweiten Signals deutlich näher zum Nullpunkt. Das deutet darauf hin, dass das Hinzufügen mit dem ausschließlich mit One-Gap-Überlappungen (und Pile-up) gefüllten Datensatz die Rekonstruktion eben jener Ereignisse deutlich verbessert. Die Two-Gap-Überlappungen (rot und grün) vom Training mit dem erweiterten Datensatz sind im Vergleich zum Training mit dem Standarddatensatz allerdings weiter vom Nullpunkt von  $E_T^{\text{dif}}$  weggerückt. Das bedeutet, dass das Netz die One-Gap-Überlappungen gut rekonstruieren kann, aber an der Performanz bei den Two-Gap-Überlappungen eingebüßt hat. Auch die 41-Gap-Überlappungen werden mit dem erweiterten Datensatz schlechter rekonstruiert. Die beste Rekonstruktion erreichte für diese Darstellung die Architektur mit drei Ebenen mit RL sechs, trainiert mit dem erweiterten Datensatz (siehe Abbildung 5.12). Im Vergleich zum mit dem erweiterten Datensatz trainierten Standardnetz ist die Rekonstruktion der Überlappungen verbessert. Allerdings ist die Rekonstruktion des ersten Signals der Two-Gap-Überlappung schlechter, als die mit dem Standarddatensatz trainierten Standardnetz.



**Abbildung 3.13:**  $E_T^{(1) \text{ dif}}$  in Abhängigkeit von  $E_T^{(1) \text{ true}}$  und  $E_T^{(2) \text{ true}}$  für alle One-Gap-Überlappungen.



**Abbildung 3.14:**  $E_T^{(2) \text{ dif}}$  in Abhängigkeit von  $E_T^{(1) \text{ true}}$  und  $E_T^{(2) \text{ true}}$  für alle One-Gap-Überlappungen.

Die linken Grafiken sind Trainings mit dem Standarddatensatz, die rechten mit dem erweiterten Datensatz. Beide wurden mit dem Standardnetz mit einer RL von fünf generiert. Im Gegensatz zur Abbildung 3.12 kann in dieser Darstellung die Energierekonstruktion in Abhängigkeit der wahren Energie  $E_T^{\text{true}}$  vom ersten und zweiten Signal analysiert werden.



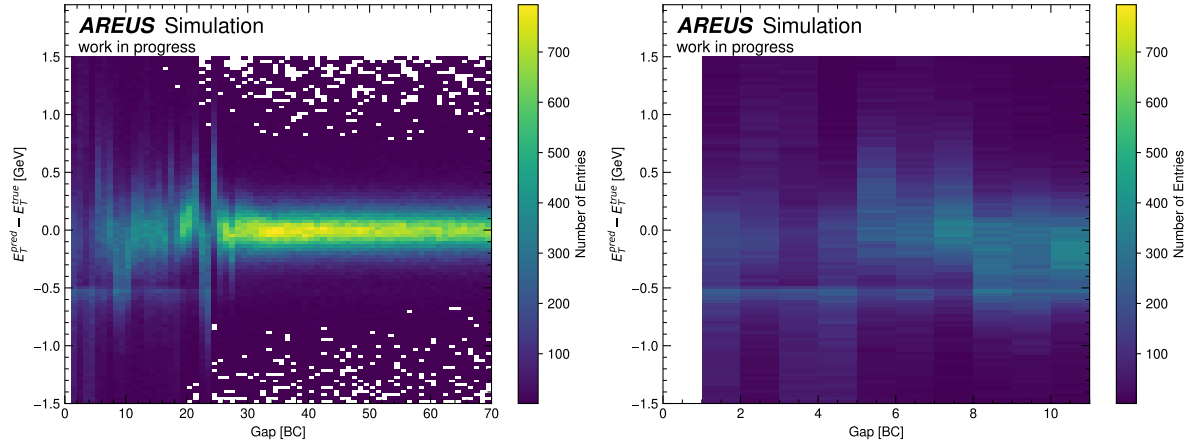
Die rechte Seite der beiden Abbildungen 3.13 (ii) und 3.14 (ii) zeigt mit dem erweiterten Datensatz generierte Trainings. Im Vergleich zum Training mit dem Standarddatensatz ist eine verbesserte Rekonstruktion direkt erkennbar. Beide Signale sind nahezu einfarbig grün gefärbt. Das lässt auf eine gute Rekonstruktion der One-Gap-Überlappungen schließen. Es sind wenige Ereignisse in beiden Grafiken erkennbar, die schlecht rekonstruiert wurden und abweichen. Werden die Ergebnisse nun mit den besten Rekonstruktionen für die One-Gap-Überlappungen in Abbildung 5.13 verglichen, ist eine fehlerfreie Rekonstruktion zu erkennen. In der Grafik ist, abgesehen von wenigen Ausnahmen,  $E_T^{\text{dif}}$  annähernd Null. Somit ist die Energierekonstruktion wesentlich verbessert.

Eine wichtige Bemerkung bezüglich dieser Darstellung ist, dass hier nur die Rekonstruktion der One-Gap-Überlappungen analysiert und evaluiert werden kann. Die Trainings mit erweitertem Datensatz geben eine homogene Verteilung von  $E_T^{\text{dif}}$  um den Nullpunkt wieder. Für eine allgemeine Aussage der Performanz des CNNs ist das nicht ausreichend. Die Performanz aller anderen Abstände und die Evaluation der Energiebereiche muss ebenso ausgewertet werden. Diese Darstellung kann somit nicht ausschließlich genutzt werden. Weitere Darstellungen sind zwingend notwendig.

### 3.3.2 Allgemeine Performanz der Energierekonstruktion

Die beste Architektur zur Energierekonstruktion der One-Gap-Überlappungen mit drei Ebenen, RL sechs und mit erweitertem Datensatz trainiert, ist in Abbildung 3.15 dargestellt. Diese Grafik kann mit Abbildung 3.2 des Standardnetzes, welcher mit dem Standarddatensatz trainiert wurde, verglichen werden.

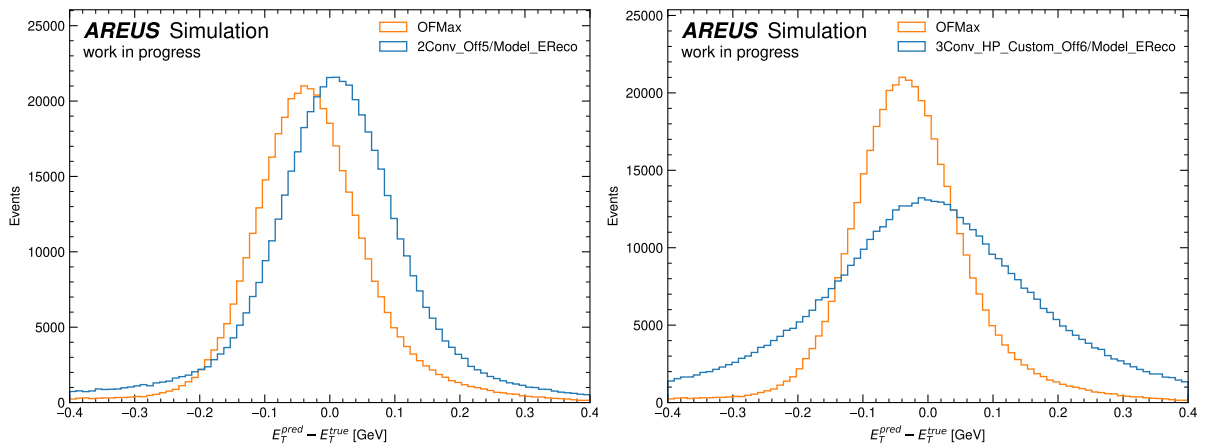
Auffällig ist bei den Darstellungen (i) mit Abständen zwischen zwei Signalen bis zu 70 BCs, dass die Breite der Verteilung um den Nullpunkt von  $E_T^{\text{dif}}$  bei Abständen größer als 22 BCs beim besten Training deutlich verbreitert ist. Gleichzeitig sind die Ereignisse bei Abständen von einem BC, was in diesem Fall der One-Gap-Überlappung entspricht, in dem auf die One-Gap-Überlappung optimierten Training deutlich dichter um  $E_T^{\text{dif}} = 0$ . Das Ziel der besseren Rekonstruktion für One-Gap-Überlappungen konnte somit erreicht werden. Die breitere Ausprägung der Ereignisse um den Nullpunkt von  $E_T^{\text{dif}}$  für alle anderen Abstände lässt allerdings darauf schließen, dass die Rekonstruktion im Vergleich zum Standardnetz schlechter geworden ist. Dieser Effekt kann für alle mit dem erweiterten Datensatz trainierten Netzwerke, unabhängig von der gewählten Architektur, beobachtet werden.



(i) Abstände bis 70 BCs.

(ii) Abstände bis 10 BCs.

**Abbildung 3.15:** Häufigkeiten von  $E_T^{\text{dif}}$  über den Abständen zwischen zwei Signalen. Die Grafik wurde mit Netzwerkarchitektur mit drei Ebenen und RL sechs erstellt, der für die One-Gap-Überlappungen besten Architektur.



(i) Standardnetz mit RL fünf trainiert mit Standarddatensatz (blau).

(ii) Bestes Netzwerk für One-Gap-Überlappungen mit drei Ebenen und RL sechs trainiert mit erweitertem Datensatz (blau).

**Abbildung 3.16:** Ereignisanzahl über  $E_T^{\text{dif}}$  für verschiedene Netzwerkarchitekturen.

In Abbildung 3.16 (i) ist in blau das mit dem Standarddatensatz trainierte Standardnetz mit RL fünf dargestellt. Abbildung 3.16 (ii) visualisiert in blau die für die One-Gap-Rekonstruktion optimierte Architektur mit drei Ebenen, RL sechs und erweitertem Datensatz. Der Unterschied beider blauer Verteilungen ist nicht zu übersehen. Die optimierte Architektur hat ein viel geringeres Maximum. Die Verteilungen werden für viele Trainings mit dem erweiterten Datensatz stark asymmetrisch. Der Median weicht deutlich vom Mittelwert der Verteilung ab. Das kann für alle mit dem erweiterten Datensatz trainierten Netzwerke beobachtet werden.

Grundsätzlich lassen die Beobachtungen der beiden beschriebenen Grafiken darauf schließen, dass die Trainings mit dem erweiterten Datensatz allgemein eine Verschlechterung der Energierekonstruktion, außer bei den One-Gap-Überlappungen, zur Folge haben.

In Abbildung 3.17 ist ein sogenannter Sternplot zu sehen. Er bietet sich an, um unterschiedliche Performanzparameter in einer Abbildung darzustellen und somit Parameter verschiedener Netzwerke miteinander vergleichen zu können. In der Darstellung ist das Training des Standardnetzes mit RL fünf und Standarddatensatz in orange, mit erweitertem Datensatz in blau, dargestellt. Ist die Wertung ( $S_C$ ) eines Performanzparameters kleiner als null, dann ist das CNN in diesem schlechter als der OF. Der Wert wird innerhalb des roten Bereiches markiert. Ist die Wertung gleich Null, dann sind OF und CNN gleich gut in dem Parameter. Der Wert des Netzes befindet sich somit genau auf dem Rand des roten Bereiches des OF. Wenn die Wertung größer ist als Null, dann ist das CNN in diesem Parameter besser und der Wert wird außerhalb des roten Bereiches eingezeichnet.

Parameter, die minimiert werden sollen, haben eine Wertung nach folgender Gleichung

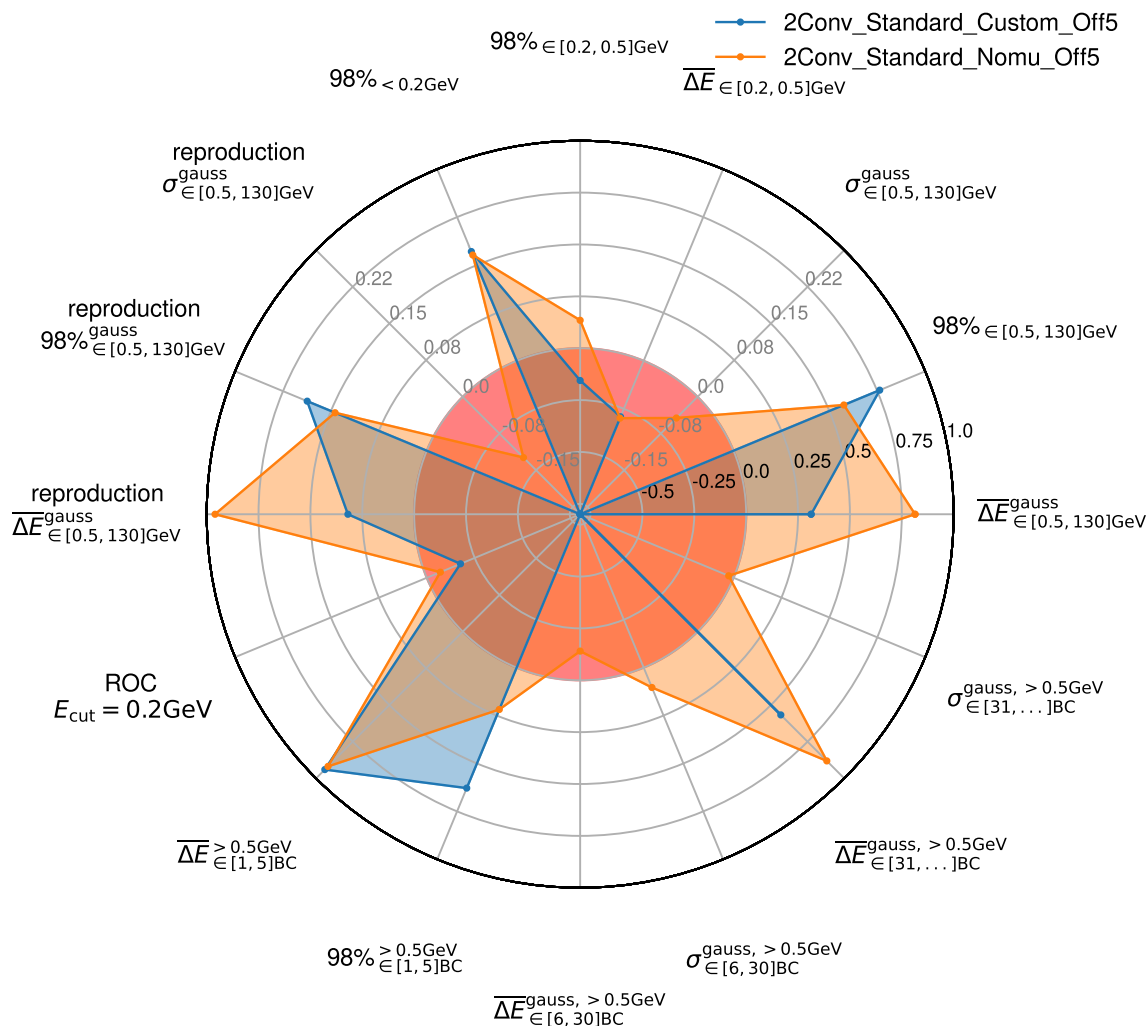
$$S_C = 1 - \frac{X_{ANN}}{X_{OF}} \quad (3.1)$$

und Parameter, die maximiert werden sollen, haben eine Wertung nach

$$S_C = \frac{X_{ANN} - X_{OF}}{1 - X_{OF}} \quad (3.2)$$

mit den Parametern  $X_{ANN}$  des ANNs und denen des OF  $X_{OF}$ .

Grundsätzlich sind alle Parameter, bis auf die Receiving Operator Curve (ROC) mit Cutoff bei 0.20115 GeV zu minimieren. Diese ROC ist ein Maß für die Signaleffizienz und Backgroundunterdrückung. Die Reproduzierbarkeitsparameter errechnen sich aus allen trainierten und konvergierten Modellen einer Architektur und sind ein Maß dafür, wie gut diese im Bezug auf den Parameter im Mittel rekonstruieren. Aus diesen kann dementsprechend geschlussfolgert werden, inwiefern die Performanzparameter  $\Delta E$ ,  $\sigma$ , 98 %<sub>[0.5,130]GeV</sub> des besten Netzes reproduzierbar sind.



**Abbildung 3.17:** Sternplot für Standardnetz mit RL fünf. In orange das Training mit Standarddatensatz, in blau das Training mit erweitertem Datensatz. Die Datenpunkte des roten Kreis stellen den OF dar und haben den Wert Null. Der äußerste Punkt des Kreises ist der Wert eins. Ist der Parameter eines Netzes gleich gut wie der des OF, bekommt er den Wert Null.

Der Sternplot gibt einen Überblick über die Performanz und Reproduzierbarkeit eines CNN im Vergleich zum OF. Die Darstellung enthält keine spezifischen Merkmale, wie z. B. die Rekonstruktion der One-Gap-Überlappungen. Eine Verbesserung oder Verschlechterung dieser Eigenschaft eines Netzes kann hier nicht präzise überprüft werden. Der Sternplot allein ist deswegen kein allgemeingültiges Maß zur Analyse der Rekonstruktionsqualität eines Netzes, sondern stellt wesentliche Eigenschaften des Netzes dar und macht sie vergleichbar. Bei der Verbesserung der One-Gap-Rekonstruktion soll es zu keiner schlechteren Rekonstruktion für andere BC-Abstände zwischen zwei Signalen kommen. Wünschenswert wären deshalb Parameterwerte des blauen Trainings, die mindestens genau so groß sind, wie die der orangen Architektur in Abbildung 3.17.

Für den Energiebereich von 0.5 bis 130 GeV gibt es drei Parameter. In dieser Größenordnung befinden sich die Teilchenenergien, die als interessant eingestuft werden. Dieses Intervall deckt dementsprechend alle Ereignisse, unabhängig ob eine Überlappung aufgetreten ist oder nicht und somit die allgemeine Rekonstruktionsgenauigkeit des Netzes ab. In dem Bereich wird  $E^{\text{dif}}$  als Gaußkurve mit dem Mittelwert  $\overline{\Delta E}^{\text{gauss}}$  und der Standardabweichung  $\sigma^{\text{gauss}}$  angepasst. Zusätzlich gibt es den 98 %-Bereich, der neben der Standardabweichung ein weiteres Maß für die Breite der Verteilung darstellt, aber ein größeres Intervall der Verteilung mit einbezieht.

Für die Abstandsbereiche zwischen zwei Signalen von [6, 30] BCs und [31, . . .] BCs gibt es jeweils die auf eine Gaußkurve angepasste Standardabweichung  $\sigma$  und den Mittelwert  $\overline{\Delta E}$ . Für Abstände größer als 31 BCs rekonstruiert das CNN nicht überlappende Signale. Im Abstandsbereich von sechs bis dreißig BCs gibt es sich überlappende Ereignisse. Dieser wird bereits vom Standardnetz gut rekonstruiert. Im Abstandsbereich der One-Gap-Überlappungen bis zu Abständen von fünf BCs ist die Varianz der wahren zur vorausgesagten Energie zu groß, um einen Gaußschen Mittelwert und Standardabweichung zu ermitteln. Es wird hier der arithmetische Mittelwert von  $\overline{\Delta E}^{>0.5\text{GeV}}$  und auch der 98 % $^{>0.5\text{GeV}}$ -Bereich verglichen. Das bedeutet, dass eine Asymmetrie das Verteilungsmaximum nicht wiedergibt. Zusätzlich muss erwähnt werden, dass der Sternplot ein Vergleich von OF und dem jeweiligen Trainingsmodell ist und deshalb keine absoluten Werte ausgibt. Wenn der OF zum Beispiel etwas sehr schlecht für einen Parameter rekonstruiert, dann werden sich Werte des Netzes im Sternplot einer Eins annähern. Für ein Netz ist es einfach, einen schlechten Parameter des OF zu übertrumpfen. Ist ein Parameter des OF allerdings sehr gut, dann wird der Wert des Netzes eine Eins nur schwer erreichen, da das Netz nicht viel besser werden kann.

Im Bereich von [1, 5] BCs ist die Energierekonstruktion überlappender Signale in bisherigen Untersuchungen am schlechtesten. Der Abstand von einem BC entspricht hier genau den One-Gap-Überlappungen. So spezifisch wie die logarithmische Abbildung 3.12 ist der Sternplot für die One- und Two-Gap-Überlappungen jedoch nicht, kann aber für einen Überblick hilfreich sein.

In den folgenden Absätzen wird zunächst das mit Standarddatensatz trainierte Standardnetz mit RL fünf in Abbildung 3.17 in orange analysiert, um es mit dem erweiterten Datensatz trainierten Standardnetz in blau zu vergleichen. Im Bereich von 0.5 bis 130 GeV sind der 98 %-Bereich und der Mittelwert  $\overline{\Delta E}^{\text{gauss}}$  gut in der Rekonstruktion. Die Gaußsche Standardabweichung ist für die Trainings mit erweitertem Datensatz (blau)

jedoch deutlich schlechter. Für den Bereich von  $[6, 30]$  BCs ist die Rekonstruktion mit Standarddatensatz schlechter, mit erweitertem Datensatz sogar um ein Vielfaches. Für die Abstände ab 31 BCs ist der Gaußsche Mittelwert  $\overline{\Delta E}^{\text{gauss}}$  gut, aber die Standardabweichung ist für den Standarddatensatz schlecht und für den erweiterten Datensatz noch schlechter. Die einzige, allerdings im Sternplot nicht gut sichtbare Verbesserung ist die der One-Gap-Überlappungen.

Trainings mit beiden Datensätzen erzeugten schlechtere ROC und eine daraus folgende niedrigere Signaleffizienz und Backgroundunterdrückung. Die Reproduzierbarkeit ist für Standarddatensatz und erweiterten Datensatz für den Energiemittelwert gut, für die Standardabweichung insbesondere mit dem erweiterten Datensatz schlecht. Auch im besten Training für die One-Gap-Überlappungen in Abbildung 5.14 kann eine deutliche Verschlechterung der allgemeinen Performanz beobachtet werden. Der Abstandsbereich von einem bis fünf BCs ist jedoch minimal besser.

### 3.3.3 Effektivität der Hyperbandsuche und des erweiterten Datensatzes

Die Analyse aller Methoden zur Verbesserung der One-Gap-Rekonstruktion wurde mithilfe verschiedener Darstellungen angefertigt und kann nur mit der Betrachtung all dieser möglichst objektiv stattfinden. Lediglich eine Darstellung heranzuziehen, z. B. den Sternplot, ist nicht aussagekräftig genug. Eine Evaluation der allgemeinen Performanz wurde mithilfe der Darstellungen aus der Abbildung 3.8 und dem Sternplot getroffen. Umfangreichere Analysen für die One-Gap-Überlappungen konnten mit den Abbildungen 3.5, 3.6 und 3.7 erstellt werden. Nur mit dem gemeinsamen Einsetzen dieser Darstellungen kann eine allumfassende Auswertung der Maßnahmen betrachtet werden.

Beste Architektur für die One-Gap-Überlappung war die mit der Hyperbandsuche generierte Architektur mit drei Ebenen und einer RL von sechs. Die RL hatte allgemein nur einen geringen Einfluss. Die besten Ergebnisse entstanden mit veränderter RL in Verbindung mit dem Training mit erweitertem Datensatz. Das kann möglicherweise durch statistische Varianz verursacht werden. Ebenfalls unklar ist die Effektivität der Veränderung der Architektur mithilfe der Hyperbandsuche der Trainings auf den erweiterten Datensatz. Das Standardnetz ist aus 99 Parametern zusammengesetzt. Bei den Architekturen des Hyperbandalgorithmus wurden 400 Parameter genutzt. Allein die Parametererhöhung könnte eine Verbesserung der Energierekonstruktion mit sich bringen. Die Optimierung auf diesen Datensatz verbessert die Energierekonstruktion nur für One-

Gap-Überlappungen, für alle anderen Abstände tritt eine Verschlechterung auf. Die dafür aufkommende Rechenzeit kann effizienter für andere Betrachtungen genutzt werden.

Das Training mit dem erweiterten Datensatz hat neben der Verbesserung der One-Gap-Rekonstruktion eine Verschlechterung der Rekonstruktion aller anderen Abstände zur Folge. Dieser Effekt ist wahrscheinlich auf ein Overfitting zurückzuführen. Der Custom-Gap-Datensatz beansprucht ein Siebtel aller Daten. Die anderen Daten bestehen aus drei Random-Gap-Datensätzen, zwei Constant-Gap-Datensatz und einem Pile-up-Datensatz. Die Varianz der restlichen Daten ist deutlich größer. Die mit dem erweiterten Datensatz trainierten Netze erlernen demzufolge die One-Gap-Ereignisse zuzuordnen und werden zu spezifisch darauf trainiert. Das könnte eine Erklärung der Verschlechterung der anderen Parameter sein. Die Gewichtung des Custom-Gap-Datensatzes spielt folglich eine tragende Rolle für weitere Optimierungsschritte.

## 4 Zusammenfassung und Ausblick

Im Zuge des High-Luminosity-Upgrades des Large Hadron Colliders stellt sich die Frage, ob durch den Einsatz von Neuronalen Netzen eine Erhöhung der Messgenauigkeit erreicht werden kann. Eine momentan bevorzugte Netzwerkarchitektur ist ein Faltungsnetzwerk mit zwei Ebenen. Die bisher trainierten Netzwerke sind nicht in der Lage, sogenannte One-Gap-Überlagerungen zu rekonstruieren. Der Schwerpunkt der wissenschaftlichen Arbeit lag darauf, die Energierekonstruktion von direkt aufeinanderfolgenden Signalen zu ermöglichen. Dabei war es auch wichtig, die allgemeine Rekonstruktionsqualität zu erhalten. Dafür wurde zunächst die Netzwerkarchitektur angepasst und die Effektivität dieser Maßnahmen analysiert. Die Architektur wurde über das Einführen der Rekonstruktionslatenz, der Variation der Rekonstruktionslatenz und der Kernelgröße und letztendlich einer Variation der Rekonstruktionslatenz und durch den Hyperbandalgorithmus verbessert und getestet. Die durchgeführten Maßnahmen hatten lediglich einen geringen Einfluss auf die Energierekonstruktion.

Das Training mit einem erweiterten Datensatz hat sich als vielversprechendste Methode erwiesen. Der zusätzlich angereicherte Datensatz enthält One-Gap-Überlappungen und Pile-up. Infolgedessen können die Netzwerke mit allen geprüften Architekturen eine deutliche Verbesserung in der Energierekonstruktion für One-Gap-Überlappungen erzielen, jedoch auf Kosten der Rekonstruktionsqualität im Allgemeinen.

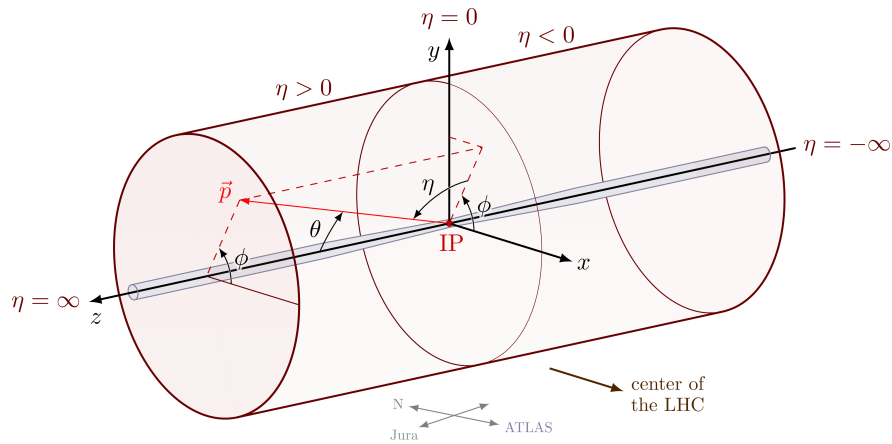
Die One-Gap-Überlappung kann unter Umständen besser rekonstruiert werden. Das Ziel der mindestens gleichbleibenden Energierekonstruktion für den Rest der Rekonstruktion konnte jedoch nicht gewährleistet werden. Dies könnte durch ein Overfitting der Daten des Custom-Gap-Datensatzes bedingt sein. Die Gewichtung des Trainingsdatensatzes gilt es in fortführenden Studien zu optimieren. Ein einfacher Ansatz wäre, den Custom-Gap-Datensatz weniger zu gewichten. Da die Two-Gap-Überlappungen aufgrund des Trainings mit erweitertem Datensatz schlechter rekonstruiert wurden, ist eine Anpassung des neu hinzugefügten Datensatzes eine weitere Möglichkeit.



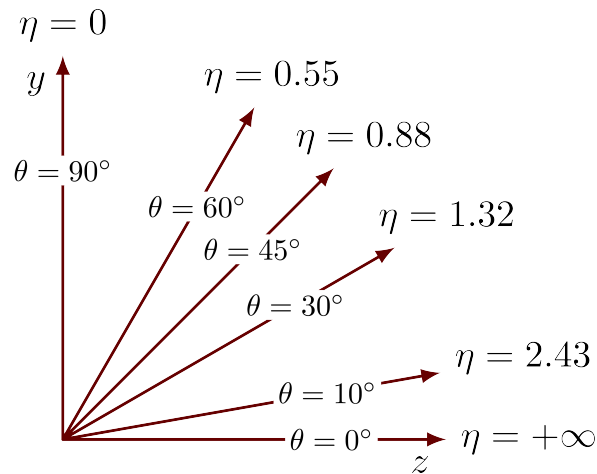
Unter Umständen kann der Custom-Gap-Datensatz aus kleineren Teildatensätzen mit Lücken zwischen zwei Ereignissen von ein bis fünf Bunch Crossings aufgebaut sein, um diese am schlechtesten rekonstruierten Abstandsbereiche zwischen zwei Signalen besser zu rekonstruieren und gleichzeitig kein Overfitting zu verursachen. Die Optimierung der Mischung der Datensätze mit One-Gap-Überlappungen und anderen Ereignissen kann mithilfe der Hyperbandsuche automatisch implementiert werden. Nach dieser Anpassung der Energierekonstruktion für alle Abstandsbereiche ist es wünschenswert, den Einfluss der Rekonstruktionslatenz erneut zu prüfen. Das beste Trainingsmodell für die One-Gap-Überlagerung verwendete eine erhöhte Rekonstruktionslatenz.

Wenn zukünftige Analysen ergeben, dass eine Verbesserung der One-Gap-Überlappung zu einer Verschlechterung der restlichen Energierekonstruktion führt, stellen sich folgende Fragen: Welche Ereignisse sind bedeutend? Welche Gewichtung haben die One-Gap-Ereignisse im Vergleich zu anderen Ereignissen? Ist die Optimierung der One-Gap-Überlappung wichtiger, als die restliche Rekonstruktionsfähigkeit? Da Ereignisse mit One-Gap-Abstand eher selten auftreten werden und Ereignisse mit größeren Abständen den Großteil aller physikalischen Messungen ausmachen, ist bei diesem Umstand die ungenaue Rekonstruktion der One-Gap-Überlappungen eine hinzunehmende Folge der Energierekonstruktion. Obwohl eine gute Rekonstruktion dieser Teilchenenergien wichtig ist, würde es dadurch zu einer Verschlechterung des Netzwerkes auf Kosten der Optimierung weniger Ereignisse kommen.

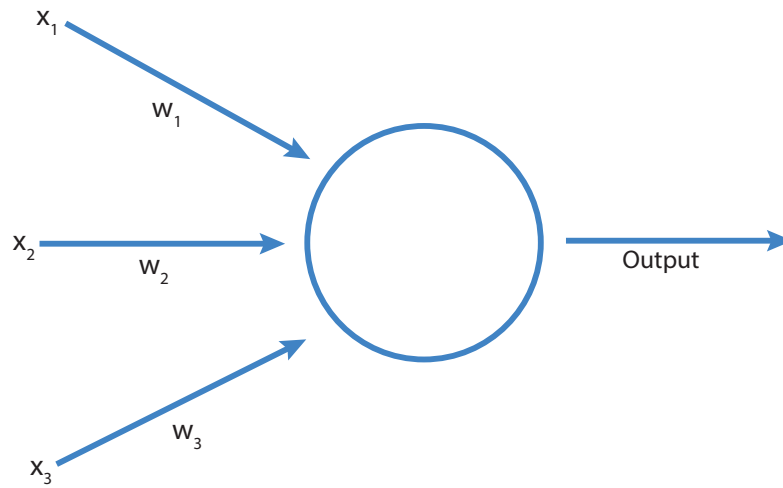
# 5 Anhang



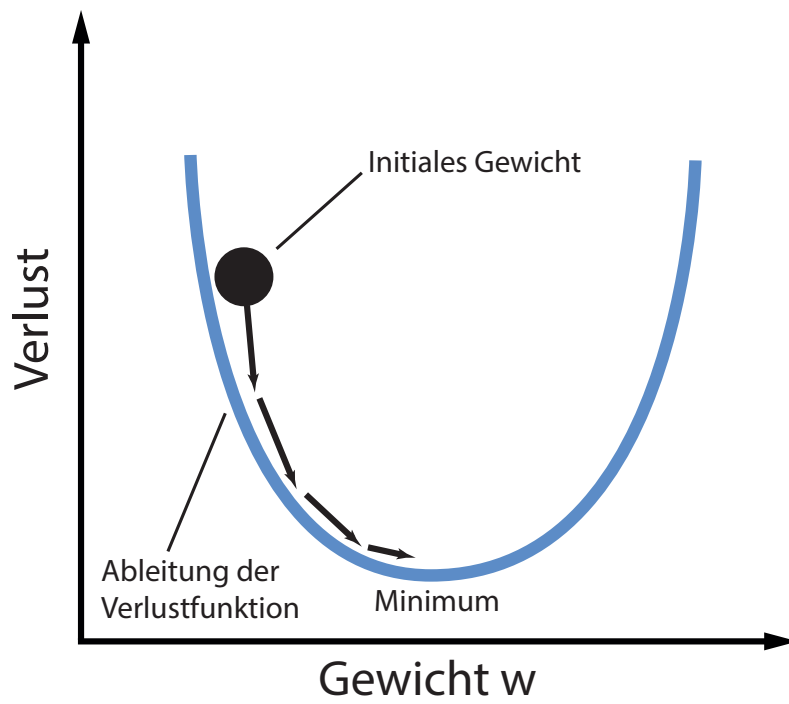
**Abbildung 5.1:** Darstellung des üblichen Koordinatensystems eines zylinderförmigen Detektors am Beispiel des CMS [32].



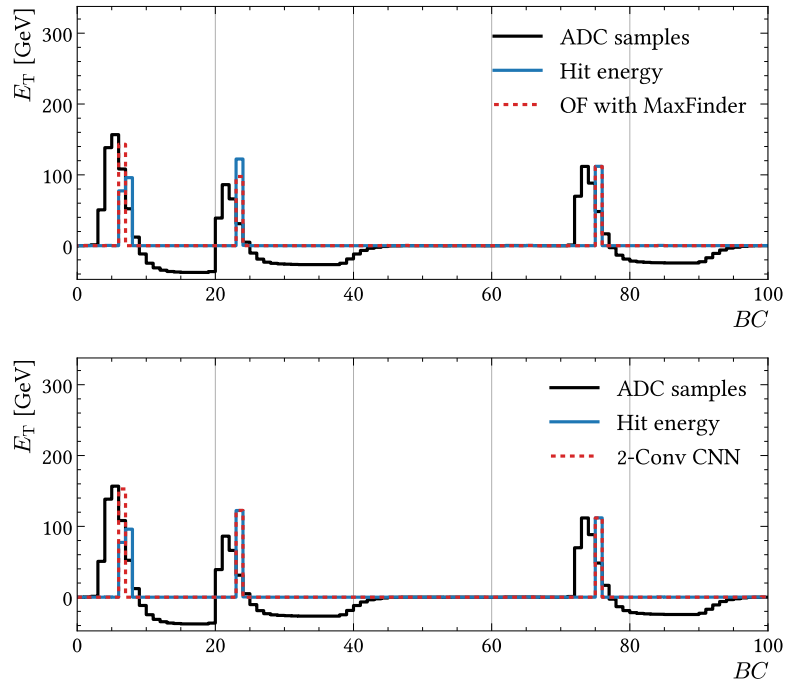
**Abbildung 5.2:** Darstellung der Pseudorapidität [32].



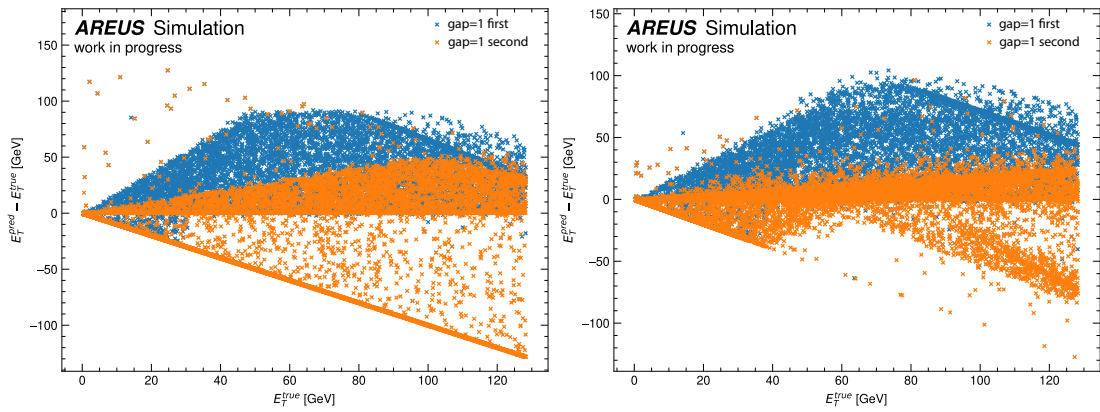
**Abbildung 5.3:** Grafische Darstellung der Funktionsweise eines in Neuronalen Netzen verwendeten Neurons.



**Abbildung 5.4:** Darstellung des Gradient-Descent-Algorithmus.



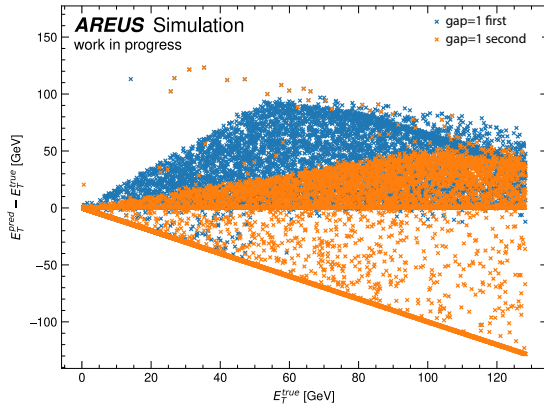
**Abbildung 5.5:** Grafische Darstellung von AREUS-simulierten One-Gap-Ereignissen mit ADC. Im oberen Plot werden Signale mit OFMax rekonstruiert, im unteren Plot mit dem Standardnetzwerk [31].



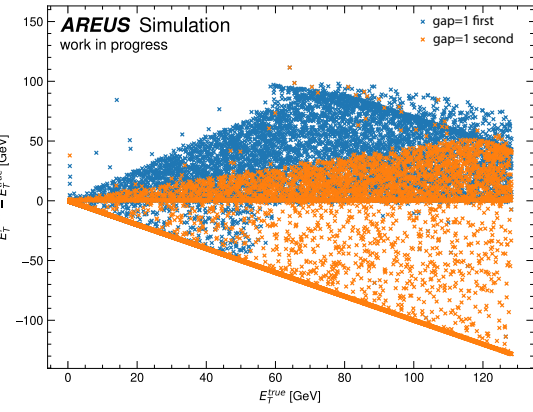
(i) RL = 6, Kernelgröße = 7.

(ii) RL = 7, Kernelgröße = 7.

**Abbildung 5.6:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{true}}$  über  $E_T^{\text{dif}}$  für das erste und zweite Signal mit dem Standardnetzwerk mit Kernelgröße sieben.

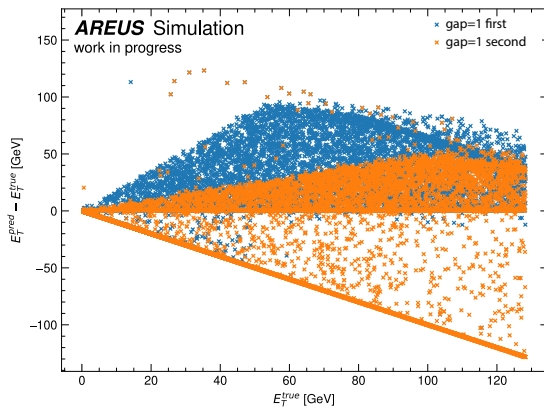


(i) RL = 6, Kernelgröße = 8.

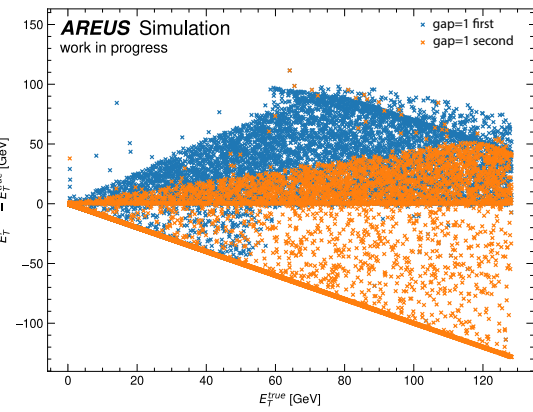


(ii) RL = 6, Kernelgröße = 9.

**Abbildung 5.7:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{true}}$  über  $E_T^{\text{dif}}$  für das erste und zweite Signal mit modifizierter RL und Kernelgröße.

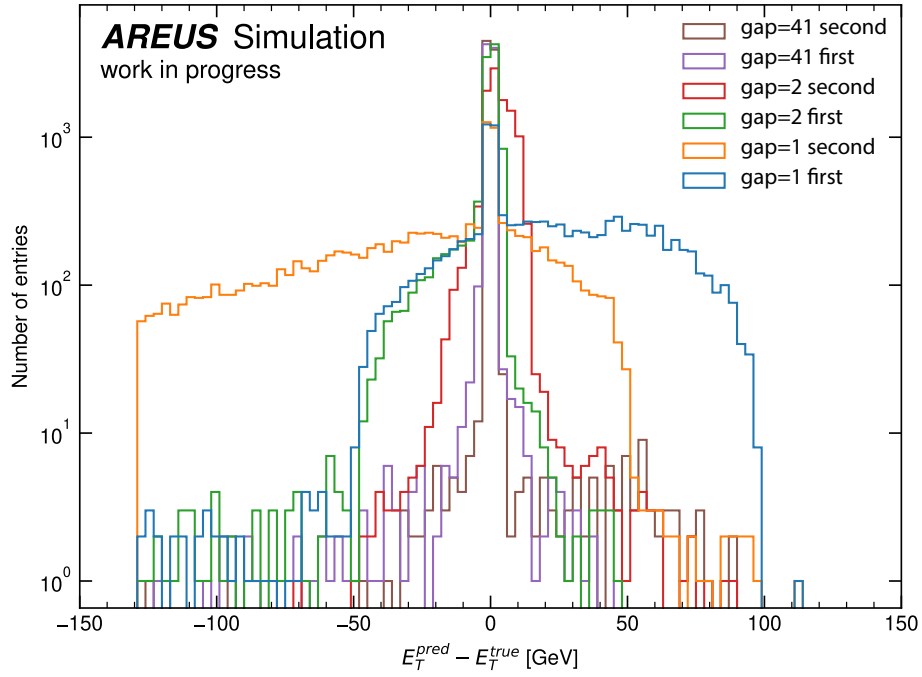


(i) RL = 7, Kernelgröße = 8.

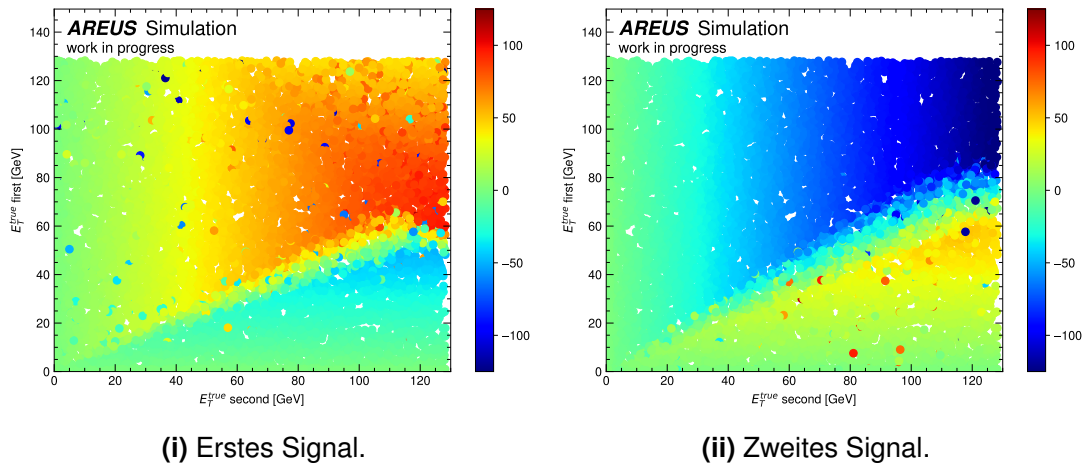


(ii) RL = 7, Kernelgröße = 9.

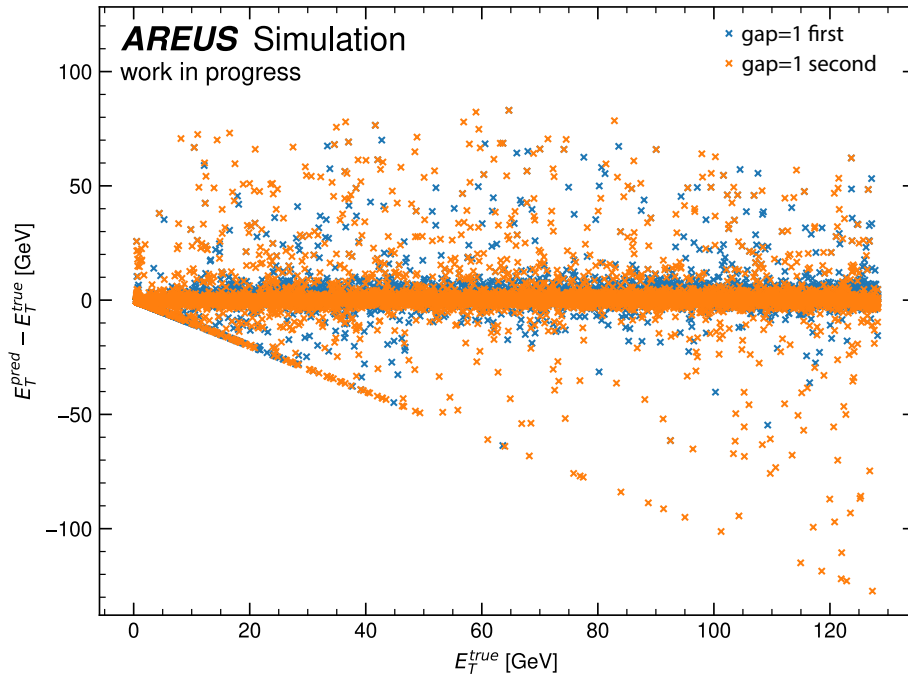
**Abbildung 5.8:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{true}}$  über  $E_T^{\text{dif}}$  für das erste und zweite Signal mit modifizierter RL und Kernelgröße.



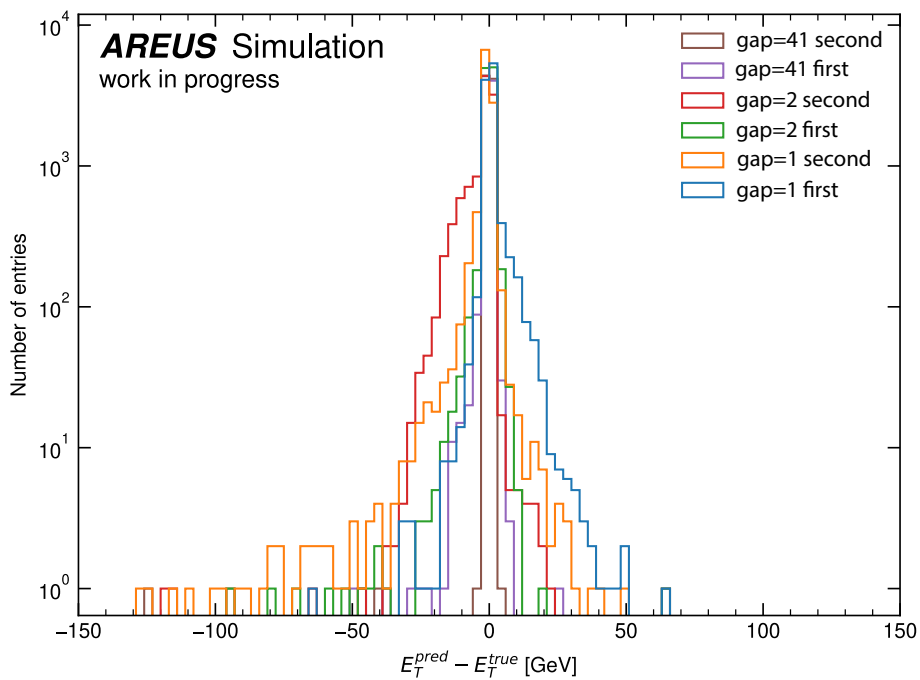
**Abbildung 5.9:** Ereignisanzahl über  $E_T^{dif}$  in logarithmischer Skala für das erste und zweite Signal für One-Gap-, Two-Gap- und 41-Gap-Überlappungen mit CNN mit RL sechs und Kernelgröße neun.



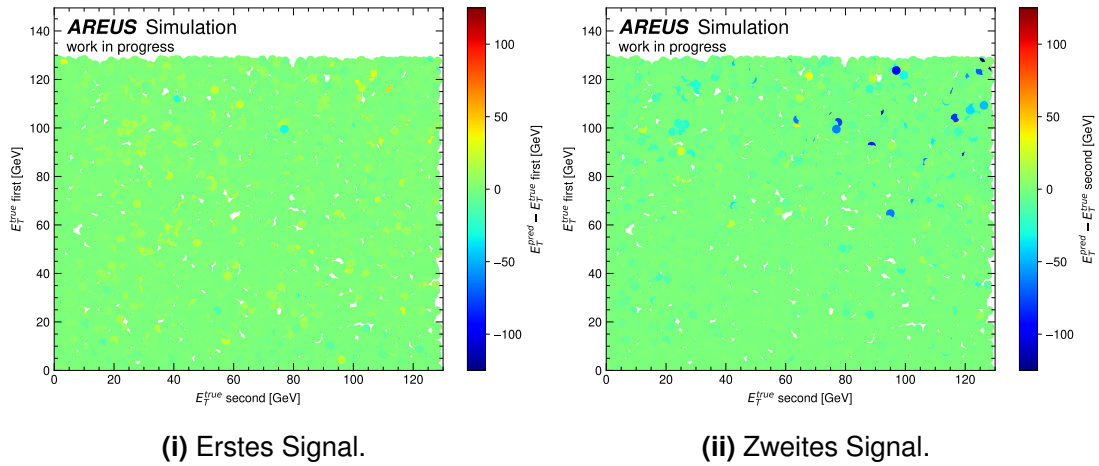
**Abbildung 5.10:**  $E_T^{dif}$  in Abhängigkeit von  $E_T^{(1)\ true}$  und  $E_T^{(2)\ true}$  für alle One-Gap-Überlappungen mit einer RL von fünf und Kernelgröße neun und Standarddatensatz.



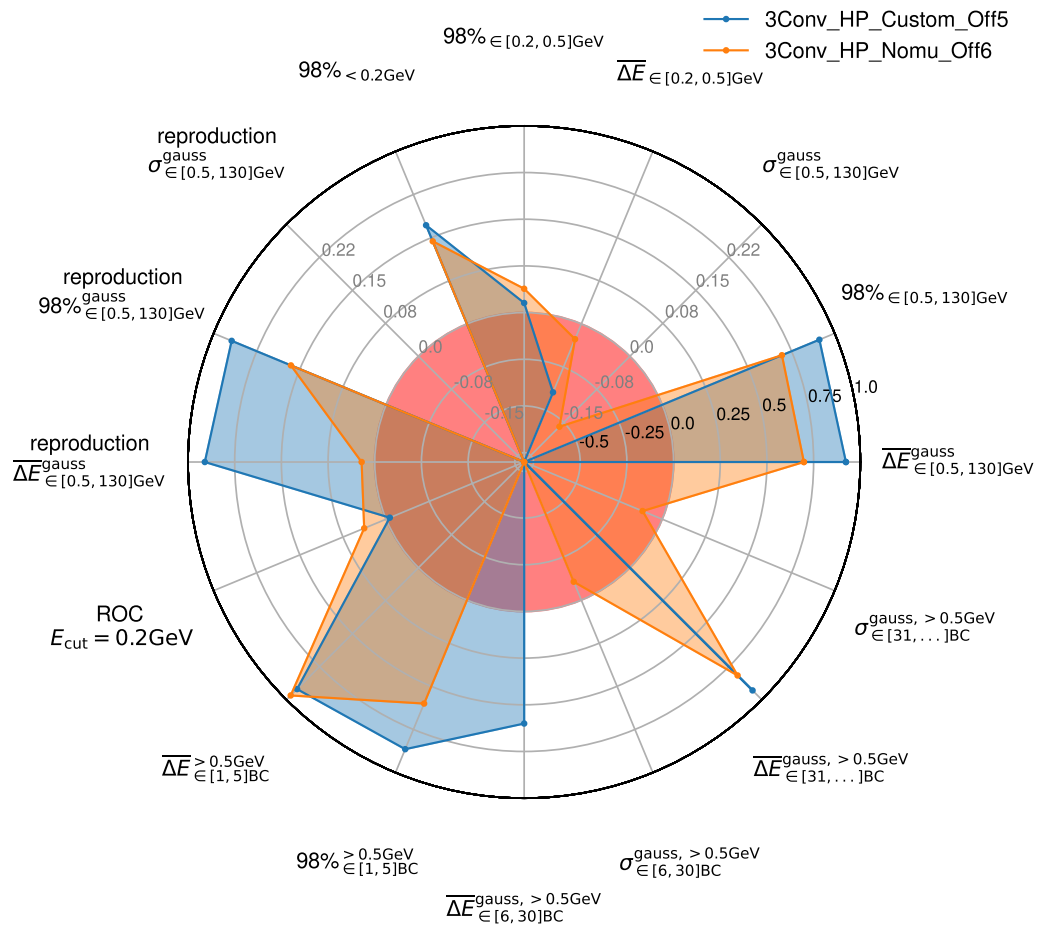
**Abbildung 5.11:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{true}$  über  $E_T^{dif}$  für das erste und zweite Signal mit zwei Ebenen mit RL sechs trainiert mit erweitertem Datensatz.



**Abbildung 5.12:** Ereignisanzahl über  $E_T^{dif}$  in logarithmischer Skala für das erste und zweite Signal für One-Gap-, Two-Gap- und 41-Gap-Überlappungen mit CNN mit drei Ebenen und RL sechs.



**Abbildung 5.13:** Rekonstruktion der One-Gap-Überlappungen mit  $E_T^{\text{true}}$  über  $E_T^{\text{dif}}$  für das erste und zweite Signal mit Netzwerkarchitektur mit drei Ebenen und RL sechs trainiert mit erweitertem Datensatz.



**Abbildung 5.14:** Sternplot für Architektur mit drei Ebenen und RL sechs. In orange das Training mit Standarddatensatz, in blau das Training mit erweitertem Datensatz. Die Datenpunkte des roten Kreis stellen den OF dar und haben den Wert Null. Der äußerste Punkt des Kreises ist der Wert eins. Ist der Parameter eines Netzes gleich gut wie der des OF, bekommt er den Wert Null.



## 6 Literatur

- [1] F. Englert und R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.*, Jg. 13, S. 321–323, 9 Aug. 1964. DOI: 10.1103/PhysRevLett.13.321. Adresse: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [2] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.*, Jg. 13, S. 508–509, 16 Okt. 1964. DOI: 10.1103/PhysRevLett.13.508. Adresse: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [3] G. S. Guralnik, C. R. Hagen und T. W. B. Kibble, “Global Conservation Laws and Massless Particles,” *Phys. Rev. Lett.*, Jg. 13, S. 585–587, 20 Nov. 1964. DOI: 10.1103/PhysRevLett.13.585. Adresse: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [4] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, Jg. 716, Nr. 1, S. 1–29, 2012, ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. Adresse: <https://www.sciencedirect.com/science/article/pii/S037026931200857X>.
- [5] CMS Collaboration, *Search for the decay of a heavy Higgs boson H into two lighter Higgs bosons h and h\_S in the h(h\_S(bb)) final state at 13TeV*, Geneva, 2021. Adresse: <http://cds.cern.ch/record/2758271>.
- [6] K. Whalen, “BSM searches at the LHC,” 2019. Adresse: <https://cds.cern.ch/record/2654369>.
- [7] *LS3 schedule change | High Luminosity LHC Project*, 6.09.2023. Adresse: <https://hilumilhc.web.cern.ch/article/ls3-schedule-change>.
- [8] T. Affolder, “ATLAS Phase II Upgrade,” 2022. Adresse: <https://cds.cern.ch/record/2799535>.
- [9] ATLAS Collaboration, “ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report,” CERN, Geneva, Techn. Ber., 2017. DOI: 10.17181/CERN.6QIO.YGH0. Adresse: <https://cds.cern.ch/record/2285582>.

- [10] O. S. Brüning, P. Collier, P. Lebrun u. a., *LHC Design Report* (CERN Yellow Reports: Monographs). Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. Adresse: <https://cds.cern.ch/record/782076>.
- [11] CERN, *LHC construction approved*, 16.12.1994. Adresse: <https://timeline.web.cern.ch/lhc-construction-approved>.
- [12] A. Lopes und M. L. Perrey, “FAQ-LHC The guide,” 2022. Adresse: <https://cds.cern.ch/record/2809109>.
- [13] Marcastel, Fabienne CERN Graphic Design service, *CERN's Accelerator Complex*, 2013. Adresse: <https://cds.cern.ch/record/1621586>.
- [14] Joao Pequeno, *Computer generated image of the whole ATLAS detector*, 2008. Adresse: <https://cds.cern.ch/record/1095924>.
- [15] Joao Pequeno, *Computer generated image of the ATLAS Liquid Argon*, 2008. Adresse: <https://cds.cern.ch/record/1095928>.
- [16] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1* (Technical design report. ATLAS). Geneva: CERN, 1999. Adresse: <https://cds.cern.ch/record/391176>.
- [17] *ATLAS liquid-argon calorimeter: Technical Design Report, S. 5, 8* (Technical design report. ATLAS). Geneva: CERN, 1996. DOI: 10.17181/CERN.FWRW.F00Q. Adresse: <https://cds.cern.ch/record/331061>.
- [18] W. Cleland und E. Stern, “Signal processing considerations for liquid ionization calorimeters in a high rate environment,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Jg. 338, Nr. 2, S. 467–497, 1994, ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(94\)91332-3](https://doi.org/10.1016/0168-9002(94)91332-3). Adresse: <https://www.sciencedirect.com/science/article/pii/0168900294913323>.
- [19] I. Goodfellow, Y. Bengio und A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [20] AWS Amazon Team, *Machine Learning und KI: Was ist Overfitting?* 18.09.2023. Adresse: <https://aws.amazon.com/de/what-is/overfitting/>.
- [21] C. Gutsche, *The Influence of Varying Pulse Shapes on the Energy Reconstruction of ATLAS Liquid Argon Calorimeter Signals using Convolutional Neural Networks*, 6.10.2022.
- [22] M. Erdmann, “Deep Learning - Erkenntnisgewinn durch moderne datengetriebene Methoden,” *Physik Journal*, Jg. 4, S. 31–36, 2020. Adresse: <https://pro-physik.de/zeitschriften/physik-journal/2020-4/>.

- [23] M. Nielsen, *Neural Networks and Deep Learning Online Book, Chapter 1*, 18.09.2023. Adresse: <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [24] ATLAS LAr Calorimeter Group, *LArCaloPublicResultsUpgrade < AtlasPublic < TWiki*, 10.09.2023. Adresse: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResultsUpgrade>.
- [25] Personal communication with Anne-Sophie Berthold, *Architektur eines optimierten Netzwerkes für die LAr-Energierückrekonstruktion*, 21.09.2023.
- [26] P. Horn, "AREUS - a software framework for the ATLAS Readout Electronics Upgrade Simulation," 2018. Adresse: <https://cds.cern.ch/record/2314235>.
- [27] J. Bergstra und Y. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, Jg. 13, Nr. 10, S. 281–305, 2012. Adresse: <http://jmlr.org/papers/v13/bergstra12a.html>.
- [28] M. Feuerer und F. Hutter, "Hyperparameter Optimization," in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff und J. Vanschoren, Hrsg. Cham: Springer International Publishing, 2019, S. 10, ISBN: 978-3-030-05318-5. DOI: 10.1007/978-3-030-05318-5\_1. Adresse: [https://doi.org/10.1007/978-3-030-05318-5\\_1](https://doi.org/10.1007/978-3-030-05318-5_1).
- [29] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh und A. Talwalkar, "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization," *Journal of Machine Learning Research*, Jg. 18, Nr. 185, S. 1–52, 2018. Adresse: <http://jmlr.org/papers/v18/16-558.html>.
- [30] P. Welle, *Anwendung von künstlichen neuronalen Netzen auf die Energierückrekonstruktion der ATLAS-Flüssigargon-Kalorimeter*, 27.06.2022.
- [31] zur Verfügung gestellt von Anne-Sophie Berthold, 20.10.2023.
- [32] I. Neutelings, *TikZ.net - CMS coordinate system*, 24.09.2023. Adresse: [https://tikz.net/axis3d\\_cms/](https://tikz.net/axis3d_cms/).



# Danksagung

Ich möchte mich bei Prof. Dr. Arno Straessner für die Teilhabe in der Detektor-Arbeitsgruppe des IKTP und deren Meetings, wie auch der Möglichkeit der Bearbeitung dieses Themas bedanken. Die Auseinandersetzung mit Neuronalen Netzen, der Detektorphysik und die Option, mir diesbezüglich Wissen anzueignen, hat mir viel Spaß bereitet. Ganz besonderer Dank gilt Anne-Sophie Berthold, Alexander Lettau, Nick Fritzsche, Johann Voigt, Markus Helbig, Philipp Welle und Rainer Hentges für die stete Bemühung, mir all meine Fragen zu beantworten und mir bei technischen und inhaltlichen Fragen und Problemen beiseite zu stehen und mich zu unterstützen.

Auch für die Unterstützung meine Mutter möchte ich an dieser Stelle bedanken.



## **Erklärung**

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am Institut für Kern- und Teilchenphysik ohne unzulässige Hilfe Dritter verfasst und alle Quellen als solche gekennzeichnet habe.

Gennaro Romanelli  
Dresden, Oktober 2023