

# **Showers Separation in Five Dimensions using Machine Learning Techniques**

**Dissertation zur Erlangung des Doktorgrades an der Fakultät für  
Mathematik, Informatik und Naturwissenschaften  
Fachbereich Physik der Universität Hamburg**

vorgelegt von  
Jack Christopher Hutchinson Rolph

Hamburg, 2023





<b>Gutachter/innen der Dissertation:</b>	Prof. Dr. Erika Garutti Jun.-Prof. Dr. Gregor Kasieczka
<b>Zusammensetzung der Prüfungskommission:</b>	Prof. Dr. Robin Santra Prof. Dr. Erika Garutti Prof. Dr. Bernd Kniehl Jun.-Prof. Dr. Gregor Kasieczka Dr. Katja Krüger
<b>Vorsitzende/r der Prüfungskommission:</b>	Prof. Dr. Robin Santra
<b>Datum der Disputation:</b>	07.09.2023
<b>Vorsitzender des Fach-Promotionsausschusses PHYSIK:</b>	Prof. Dr. Günter H. W. Sigl
<b>Leiter des Fachbereichs PHYSIK:</b>	Prof. Dr. Wolfgang J. Parak
<b>Dekan der Fakultät MIN:</b>	Prof. Dr.-Ing. Norbert Ritter

### **Colophon**

This document was typeset with the help of *KOMA-Script* and  $\LaTeX$  using the *kaobook* class.

The source code of this book is available at:

<https://github.com/fmarotta/kaobook>



Mr. Tweedy:  
*What is it?*

Mrs. Tweedy:  
*It's a pie machine, you idiot. Chickens go in, pies come out.*

Mr. Tweedy:  
*Ooh, what kind of pies?*

Mrs. Tweedy:  
*Apple.*

Mr. Tweedy:  
*My favourite!*

Dialogue, Lord, P., Park, N. (2000). *Chicken Run*. DreamWorks  
Distribution.



# Zusammenfassung/Abstract

**Zusammenfassung (Deutsch):** Diese Arbeit präsentiert Methoden und Werkzeuge zur Kalibrierung und Betrieb des CALICE Analogen Hadronenkalorimeters (AHCAL), eines hochgranularen Stahl-Szintillator-Kalorimeters, das für die Verwendung in der Teilchenflußrekonstruktion (Particle Flow) in einem zukünftigen Präzisions-Leptonen-Kollisions-Experiment entwickelt wurde und eine hervorragende Energieauflösung für Jets erfordert. Diese Kalorimetrie-Methode basiert auf hochgranularen Kalorimetern, exzellenter Detektor-Kalibrierung und anspruchsvollen Clustering-Algorithmen, um Energie deposition (energy deposits) von verschiedenen Teilchen zu unterscheiden. Das AHCAL verfügt über etwa 22.000 Auslesekanäle, die Silizium-Photomultiplier (SiPMs) zur Detektion von Szintillationslicht verwenden, und ist einzigartig in seiner Fähigkeit, sowohl Energie als auch einen Zeitstempel mit einer zeitlichen Auflösung von bis zu 100 Pikosekunden zu messen.

In dieser Arbeit werden Software-Werkzeuge und Algorithmen entwickelt, um das AHCAL und die SiPMs zu kalibrieren und zu betreiben. Zunächst wird ein flexibles Monte-Carlo-Programm namens *LightSimTastic* vorgestellt, das die Antwort der SiPMs im linearen Regime simuliert, in dem Sättigungseffekte vernachlässigt werden können. Die Eingabewerte für das Programm sind die mittlere Anzahl und die zeitliche Verteilung der Geiger-Entladungen (Geiger discharge) von Photonen sowie die Dunkelzählrate (dark count rate). Anschließend wird ein weiteres Software-Werkzeug zur Charakterisierung der SiPM-Spektren namens *PeakOTron* eingeführt. Dieses Programm fittet die gesamten Ladungsspektren (charge spectrum) an, einschließlich der Intervalle zwischen den Photoelektronen-Spitzen (photoelectron peaks), was es ermöglicht, neben der mittleren Anzahl detektierter Photonen auch Verstärkung (gain), Verstärkungsverteilung (gain spread), prompten Übersprechen (prompt cross-talk), Grundlinie (pedestal) und elektronisches Rauschen (electronics noise), die Dunkelzählrate sowie die Wahrscheinlichkeit und Zeitkonstante von Nachentladungen aus den Ladungsspektren zu bestimmen. Die Startwerte der Anpassungsparameter werden aus den Ladungsspektren extrahiert, und das Programm liefert eine gute Beschreibung sowohl von Simulations- als auch von experimentellen Daten. Drittens wird ein neuronales Netzwerkmodell (neural network model) zur Softwarekompensation entwickelt, das für das AHCAL verwendet wird. Dabei werden räumliche und zeitliche Ereignisinformationen des AHCAL und Energieinformationen verwendet, um die Empfindlichkeit für die Entwicklung von Schauern und den Neutronenanteil des Hadronenschauers zu verbessern. Die Methode erzeugte eine lineare Detektorantwort (detector response) bei der Kompensation von sowohl simulierten als auch experimentellen Hadronenschauer-Daten. Sie übertraf eine veröffentlichte Kon-

trollmethode in Bezug auf die Auflösung für jede untersuchte Teilchenenergie. Schließlich werden neuronale Netzwerkmodelle zur Schauererkennung angewendet, um geladene und synthetische neutrale Hadronenschauer zu trennen. Das AHCAL erweist sich als ein äußerst effektiver Teilchenfluß-Kalorimeter, bei dem mehr als 90 % der Ereignisse in der Kalorimeterauflösung für die meisten Kombinationen von Teilchenenergien rekonstruiert werden können. Dies verbessert sich signifikant in den anspruchsvollsten Fällen unter Verwendung von Zeitinformationen als zusätzliche Clustering-Dimension.

**Abstract (English):** This thesis presents methods and tools for the calibration and operation for the CALICE Analogue Hadronic Calorimeter (AHCAL), a highly-granular, steel-scintillator calorimeter designed for use in Particle Flow in a future precision lepton collider experiment requiring excellent jet energy resolution. This calorimetry method relies on highly-granular calorimeters, excellent detector calibration and sophisticated clustering algorithms to resolve energy deposits from different particles. The AHCAL has around 22,000 readout channels, utilising silicon photomultipliers (SiPMs) to read scintillation light, and is unique for its capacity to measure both energy and a timestamp with up to 100 ps timing resolution.

In this thesis, software tools and algorithms are developed to calibrate and operate the AHCAL and SiPMs. Firstly, a flexible Monte Carlo program called `LightSimTastic` is presented which simulates the response of SiPMs in the linear regime in which saturation effects can be ignored. Inputs to the program are the mean number and time distribution of Geiger discharges from photons, and the dark-count rate. Then, another software tool for the characterisation of SiPM spectra called `PeakOTron` is introduced. This program fits the entire charge spectra, including the intervals in-between the photoelectron peaks, which allows determining, in addition to the mean number of detected photons, gain, gain spread, prompt cross-talk, pedestal, and electronics noise, the dark-count rate as well as the probability and time constant of after-pulses from charge spectra. The starting values of the fit parameters are extracted from the charge spectra, and the program provides a good description of both simulation and experimental data. Thirdly, a neural network model for software compensation developed for the AHCAL is presented, using spatial and temporal event information from the AHCAL and energy information, which is expected to improve sensitivity to shower development and the neutron fraction of the hadron shower. The method produced a linear detector response in compensating both simulation and experimental hadron shower data. It outperformed a published control method in terms of resolution for every particle energy studied. Lastly, neural network models for shower separation are applied to separating charged and synthetic neutral hadron shower events using the AHCAL detector. The AHCAL is demonstrated to be a highly effective Particle Flow Calorimeter, with  $> 90\%$  of events being reconstructed in the calorimeter resolution for most particle energy combinations that improve significantly in the most challenging cases using timing information as an additional clustering dimension.

**High-Learned Mardlin' 'Bout What I Dun (Norfolk):** This big ol' barney what I 'rit is 'bout ' what I larned' 'bout a big ol' whatsit called a A-hach-CAL for gorpini' at tiny party-cools when yoo hull'em at wun'nother lyke cars on the Acl Straight. As' got luds'a tiny little doodahs called silicon pho-to-multi-plyers or S-eye-PMs for seein' the lyte whass' cumini' orf bits of old party-cools; there's more'o them doodahs than idjuts and rum'muns on Prince O'Wales Rud on Fryday Nyght, or even Portman Rud on a Match Day. Yoo need lots of 'em 'cos yoo myte strain yer eyes if yoo look too hard; an 'as got the tyme as well, 'case you fergit.

In this ol' barney, I fiddled 'bout with the old compooter so 'at myte make the A-hach-CAL work 'bit bettugh. I made a thin' called LightSimtastic, whass a tool for sim-yew-latin' the charge and current orf a S-eye-PM. I also made a thin' called PeakOTron, whass'a tool for figurin' out wass gun'ann in that old S-eye-PM from the charge. Thass' got luds of bits'un bobs what it spoos out, lyke the gain, the gain sprehd, the prompt cross-blarin', pedestal, dark count rayte, and even the probability of arfter-pulses and arfter-pulses tyme-constants. Blow me down, 'at 'ent too bad compared to the ol' sim-yew-lation and data. Thass' also got a bit gunn'an 'bout a noo-ral network was trained for figurin' out what the old A-hach-CAL shuda measured when it dint to improve the resulooshon, and 'as a bit speshul cos 'at dun't get stuck in the old energy range lyke other meffurds. 'At uses tyme and spatial infermation, an even does a bit bettugh than the standard meffurd, which 'ent too bad, all thins' considered. Fynally, a buncha noo-ral networks for figurin' out whass'a nootral and whass'a charged shower in the old' A-hach-CAL were trained. 'At prove the A-hach-CAL 'ent too bad at figurin' out whass'wat; more 'an 90% of the events, are within old resulooshon for most combinashuns of party-cool energies, and get qyte a bit bettugh for rum'muns using tyme infermation.




# Eidesstattliche Versicherung/Declaration on Oath

**Deutsch:** Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

**English:** I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

**Norfolk:** I swear on my lyfe an 'afore Delia Smith that I 'rit this big 'ol barney on my tod, 'ent stul nuthin' and dint git help from no-h-where other than what I 'rit in the barney.

Unterschrift:  \_\_\_\_\_

Jack Christopher Hutchinson Rolph

Hamburg, September 21, 2023




# Erklärung über die Identische Inhalte in elektronischer und gedruckter Form/Declaration of Identical Content in Electronic and Printed Format

**Deutsch:** Ich versichere, dass dieses gebundene Exemplar der Dissertation und das in elektronischer Form eingereichte Dissertationsexemplar (über den Docata-Upload) und das bei der Fakultät (zuständiges Studienbüro bzw. Promotionsbüro Physik) zur Archivierung eingereichte gedrucktegebundene Exemplar der Dissertationsschrift identisch sind.

**English:** I, the undersigned, declare that this bound copy of the dissertation and the dissertation submitted in electronic form (via the Docata upload) and the printed bound copy of the dissertation submitted to the faculty (responsible Academic Office or the Doctoral Office Physics) for archiving are identical.

**Norfolk:** I, who syne here, swear on my lyfe that the binded copy of the barney what 'rit and the barney what I send orf on the old compooter (usin' the Docata whatsit) and the barney what I 'rit and printed out and give to the faculty (the responsible High-Learnin' Office or the Doctoral Office Physics) for sticking somewhere where the Sun dunt shyne are 'sactly the same.

Unterschrift: 

Jack Christopher Hutchinson Rolph

Hamburg, September 21, 2023



# Contents

Zusammenfassung/Abstract	vii
Eidesstattliche Versicherung/Declaration on Oath	xi
Erklärung über die Identische Inhalte in elektronischer und gedruckter Form/Declaration of Identical Content in Electronic and Printed Format	xiii
Contents	xv
<b>1 Introduction</b>	<b>1</b>
<b>2 Background and Motivation</b>	<b>5</b>
2.1 The Standard Model . . . . .	5
2.1.1 The Higgs Boson . . . . .	7
2.1.2 Top Quark . . . . .	8
2.2 Linear Lepton Colliders . . . . .	8
2.2.1 Overview of Lepton Colliders . . . . .	8
2.2.2 The Importance of Jets at Future Lepton Colliders . . . . .	10
2.2.3 Example Measurements at Future Lepton Colliders . . . . .	13
2.3 Interaction of Particles With Matter . . . . .	15
2.3.1 Ionisation . . . . .	15
2.3.2 Energy Loss Mechanisms of Electrons and Positrons . . . . .	16
2.3.3 Energy Loss Mechanisms of Photons . . . . .	16
2.3.4 Electromagnetic Showers . . . . .	16
2.3.5 Energy Loss Mechanisms of Hadrons . . . . .	18
2.3.6 Hadronic Showers . . . . .	20
2.3.7 Simulations of Hadron Showers . . . . .	22
2.4 Calorimeters . . . . .	24
2.4.1 Design of Calorimeters . . . . .	25
2.4.2 Energy Resolution of Sampling Calorimeters . . . . .	26
2.4.3 Improving Hadronic Energy Resolution . . . . .	28
2.4.4 Linearity of Response . . . . .	30
2.4.5 Particle Flow . . . . .	30
2.4.5.1 Confusion . . . . .	33
2.5 Apparatus and Experimental Methods . . . . .	34
2.5.1 CALICE AHCAL . . . . .	34
2.5.1.1 Design of the AHCAL Prototype . . . . .	34
2.5.1.2 Event Information . . . . .	37
2.5.1.3 2018 SPS June Testbeam . . . . .	39
2.5.1.4 Software and Simulation Tools for AHCAL . . . . .	40
2.5.1.5 Five-Dimensional Calorimetry . . . . .	41
2.5.2 Silicon Photomultipliers . . . . .	42
2.5.2.1 Semiconductors . . . . .	43
2.5.2.2 Silicon and Doping . . . . .	43
2.5.2.3 p-n Junctions . . . . .	44
2.5.2.4 The Design of Silicon Photomultipliers . . . . .	45
2.5.3 Characterisation of SiPMs . . . . .	47
2.5.3.1 SiPM Performance Parameters and Noise Sources . . . . .	47

2.5.3.2	Characterisation of SiPMs from Charge Spectra . . . . .	49
2.5.4	Machine Learning . . . . .	51
2.5.4.1	Design of Neural Networks . . . . .	52
2.5.4.2	Activation Functions . . . . .	56
2.5.4.3	Loss Functions and Optimisation Algorithms . . . . .	59
2.5.4.4	Graph Neural Networks . . . . .	60
<b>3</b>	<b>LightSimtastic: A Tool For SiPM Simulation</b>	<b>63</b>
3.1	Overview Of Study . . . . .	63
3.2	Method And Tools . . . . .	64
3.2.1	Detector Response Mode and Free Parameters . . . . .	64
3.2.1.1	Geiger Array . . . . .	65
3.2.1.2	Time And Integration Gate ( $t_0, t_{\text{gate}}, t_{\text{start}}$ ) . . . . .	65
3.2.1.3	Gain Fluctuations ( $\sigma_1$ ) . . . . .	66
3.2.1.4	Number, Times and Amplitudes of Photon-Induced Geiger Discharges ( $N_{\gamma G}, n_{\gamma G}, A_{i\gamma G}, t_{i\gamma G}$ ) . . . . .	67
3.2.1.5	Number, Times and Amplitudes of Dark Counts ( $DCR, n_{\text{dark}}, A_{i\text{dark}}, t_{i\text{dark}}$ ) . . . . .	67
3.2.1.6	Number, Times and Amplitudes of Prompt Cross-Talk ( $p_{\text{pXT}}, n_{\text{pXT}}, A_{i\text{pXT}}, t_{i\text{pXT}}$ ) . . . . .	67
3.2.1.7	Number, Times and Amplitudes of Delayed Cross-Talk ( $p_{\text{dXT}}, n_{\text{dXT}}, A_{i\text{dXT}}, t_{i\text{dXT}}$ ) . . . . .	68
3.2.1.8	Number, Times and Amplitudes of After-pulses ( $p_{\text{Ap}}, n_{\text{Ap}}, A_{i\text{Ap}}, t_{i\text{Ap}}$ ) . . . . .	68
3.2.1.9	Pulse Shape Model . . . . .	69
3.2.1.10	Charge of The Event . . . . .	69
3.2.1.11	Transient of The Event, Electronics Response Function and Width ( $\sigma_I, \sigma_R, R(t)$ ) . . . . .	70
3.3	Results . . . . .	70
3.4	Conclusion . . . . .	73
<b>4</b>	<b>Peak0Tron: A Tool For SiPM Characterisation</b>	<b>75</b>
4.1	Overview of Study . . . . .	75
4.2	Theory . . . . .	76
4.2.1	Detector Response Model And Free Parameters . . . . .	77
4.2.2	Generalised Poisson Distribution . . . . .	77
4.2.3	Model for After-pulses . . . . .	79
4.2.4	Model for Photon-Induced Discharges . . . . .	80
4.2.5	Dark Count Model . . . . .	81
4.3	Methods And Tools . . . . .	83
4.3.1	Model Input Parameters . . . . .	83
4.3.2	Data Preparation . . . . .	83
4.3.3	Effective Gain using the Fourier Transform ( $G_{\text{FFT}}^*$ ) . . . . .	83
4.3.4	Peak Finding . . . . .	84
4.3.4.1	Initial Estimation of Peak Positions . . . . .	84
4.3.4.2	Pedestal Estimation . . . . .	84
4.3.4.3	Improved Peak Position Estimate . . . . .	85
4.3.5	Determination of the Input Parameters . . . . .	87
4.3.5.1	Pedestal Position and Width, Gain Spread ( $Q_0, \sigma_0, \sigma_1$ ) . . . . .	87
4.3.5.2	Estimates of $\mu$ and $\lambda$ . . . . .	87
4.3.5.3	Dark Count Rate Estimate ( $DCR$ ) . . . . .	88

4.3.5.4	After-pulse Parameters ( $p_{Ap}, \tau_{Ap}$ ) . . . . .	88
4.3.6	Implementation of the Fit . . . . .	89
4.4	Results . . . . .	90
4.4.1	Validation of Peak0Tron with Simulated Spectra . . . . .	90
4.4.1.1	Overall Performance Of Fit . . . . .	91
4.4.1.2	Systematic Biases and Statistical Uncertainties Of Fit . . . . .	98
4.4.1.3	CPU Time for the Fit and the Prefit . . . . .	98
4.4.2	Fits to Experimental Data . . . . .	102
4.4.2.1	SiPMs and Setup . . . . .	102
4.4.2.2	Peak0Tron Fits . . . . .	104
4.4.3	Requirements and Limitations of Fits with Peak0Tron . . . . .	112
4.5	Conclusion . . . . .	112
<b>5</b>	<b>5D Software Compensation with Machine Learning</b>	<b>115</b>
5.1	Overview of Study . . . . .	115
5.1.1	Biasing of Software Compensation Models . . . . .	116
5.1.2	Proposed Method to Reduce Bias in Software Compensation Models	119
5.2	Methods and Tools . . . . .	121
5.2.1	Neural Network SC Method . . . . .	121
5.2.2	Control SC Method . . . . .	122
5.2.3	Datasets and Training . . . . .	123
5.2.3.1	Datasets . . . . .	123
5.2.3.2	Training . . . . .	139
5.3	Results . . . . .	140
5.3.1	Simulation Model . . . . .	140
5.3.1.1	Response Distributions in Simulation . . . . .	140
5.3.1.2	Resolution and Linearity of Response in Simulation . . . . .	140
5.3.1.3	Correlations with Spatial and Temporal Information in Simulation . . . . .	143
5.3.1.4	Robust Statistics of Response Distributions in Simulation	145
5.3.2	Data Model . . . . .	148
5.3.2.1	Response Distributions in Data . . . . .	148
5.3.2.2	Resolution and Linearity of Response in Data . . . . .	148
5.3.2.3	Correlations with Spatial and Temporal Information in Data	151
5.3.2.4	Robust Statistics of Response Distributions in Data . . . . .	154
5.3.2.5	Comparison of Simulation Model to Data Model . . . . .	157
5.4	Conclusion . . . . .	162
<b>6</b>	<b>5D Shower Separation with Machine Learning</b>	<b>165</b>
6.1	Overview of Study . . . . .	165
6.2	Methods and Tools . . . . .	167
6.2.1	Network Architectures . . . . .	167
6.2.1.1	Overview . . . . .	167
6.2.1.2	Summary of Models . . . . .	167
6.2.2	Raw Datasets . . . . .	175
6.2.3	Data Synthesis Techniques . . . . .	177
6.2.3.1	Generation of Synthetic Neutral Events . . . . .	177
6.2.3.2	Shower Displacement . . . . .	189
6.2.3.3	Synthetic Neutral and Charged Hadron Shower Events . . . . .	194
6.2.4	Synthetic Datasets and Training . . . . .	195
6.2.4.1	Two-Shower Charged-Synthetic Neutral Datasets . . . . .	195

6.2.4.2	Training . . . . .	204
6.3	Results . . . . .	205
6.3.1	Simulation Models . . . . .	207
6.3.1.1	Example Event Displays . . . . .	207
6.3.1.2	Linearity of Response . . . . .	211
6.3.1.3	Spread of Confusion Energy . . . . .	211
6.3.1.4	Fraction of Events Reconstructed Within Calorimeter Res- olution vs. Charged/Neutral Particle Combination . . . . .	215
6.3.1.5	Fraction of Events Reconstructed Within Calorimeter Res- olution vs. Lateral Shower Distance . . . . .	220
6.3.1.6	Fraction of Events Reconstructed Within Calorimeter Res- olution vs. Longitudinal Shower Distance . . . . .	221
6.3.1.7	Skewness of Confusion Energy . . . . .	225
6.3.1.8	Examples Event Displays of Failure Modes . . . . .	229
6.3.1.9	Software Compensation of Separated Neutral Hadron Showers . . . . .	232
6.3.2	Data Model . . . . .	237
6.3.2.1	Example Event Displays . . . . .	237
6.3.2.2	Linearity Of Response . . . . .	241
6.3.2.3	Spread of Confusion Energy . . . . .	241
6.3.2.4	Fraction of Events Reconstructed Within Calorimeter Res- olution vs. Charged/Neutral Particle Combination . . . . .	241
6.3.2.5	Fraction of Events Reconstructed Within Calorimeter Res- olution vs. Lateral Shower Distance . . . . .	245
6.3.2.6	Fraction of Events Reconstructed Within Calorimeter Res- olution vs. Longitudinal Shower Distance . . . . .	245
6.3.2.7	Skewness of Confusion Energy . . . . .	245
6.3.2.8	Software Compensation of Separated Neutral Hadron Showers . . . . .	247
6.4	Conclusion . . . . .	252
<b>7</b>	<b>Conclusion</b>	<b>255</b>
7.1	Summary of Research . . . . .	255
7.2	Future Prospects . . . . .	257
7.3	Closing Remark . . . . .	259
<b>8</b>	<b>Appendix</b>	<b>261</b>
8.1	Peak0Tron: A Python Tool For SiPM Characterisation . . . . .	261
8.1.0.1	Geiger-Discharge Probability for After-pulses . . . . .	261
8.2	Software Compensation for AHCAL using Graph Neural Networks . . . . .	264
8.3	Shower Separation in Five Dimensions Using Machine Learning . . . . .	276
	<b>Bibliography</b>	<b>297</b>



High-energy particle colliders are used to study the products of collisions to explore the fundamental building blocks of matter and interactions. Lepton colliders can provide precision measurements of quantities of scientific interest, such as cross-sections, coupling constants and the masses of the particles of the Standard Model and beyond. In particular, since the discovery of the Higgs Boson in 2012 [1], linear lepton colliders such as the International Linear Collider (ILC) and Compact Linear Collider (CLIC) have been favoured by the physics community for achieving unprecedented precision in the study of the properties of the Higgs boson and other fundamental particles, with centre-of-mass-energies in the range 250 GeV to several TeV.

Both ILC and CLIC require exceptional jet energy resolution in the order of around 3% for jet energies in the range 50-200 GeV for precision Higgs measurements. This is expected to be achieved using Particle Flow (PF), which exploits  $4\pi$ -hermetic tracking of charged particles, highly granular calorimeters and sophisticated clustering algorithms to associate charged tracks to energy deposits in the calorimeters.

The Analogue Hadronic Calorimeter (AHCAL) is a steel-scintillator (Fe-Sc) highly granular hadron calorimeter that utilises silicon photomultipliers (SiPMs) as detectors and is designed for Particle Flow. It has around 22,000 readout channels and can measure energy and a timestamp with up to 100 picoseconds timing resolution. The AHCAL is designed to be highly granular, so it can accurately measure the energy deposited by individual particles and the substructure of hadron shower events. This allows for sophisticated clustering techniques to separate and measure the energies of individual particles within the jet, which is necessary to achieve the required jet energy resolution. Each readout channel is coupled to a silicon photomultiplier (SiPM), a photo-detector with single-photon resolution, is insensitive to magnetic fields, have excellent timing resolution and has high photon-detection efficiency, making them ideal photosensors for AHCAL.

All detector and event reconstruction aspects must be optimised to achieve the challenging jet energy resolution requirements achievable using PF. This thesis touches on several fundamental ingredients of PF reconstruction.

Firstly, at the detector level, the energy measurement achieved by the AHCAL calorimeter must be as accurate as possible. To do this, single-channel calibration is required for the SiPMs of the AHCAL. This highly complex task requires characterising all the device's properties, including noise effects such as thermally-induced 'dark' counts, cross-talk between individual pixels and delayed pulses due to de-trapping effects in damaged silicon. These effects are typically not accounted for in standard techniques for characterising a SiPM. This shortcoming can induce significant biases and uncertainties during the calibration of the

detector. Furthermore, the SiPMs of the AHCAL are too numerous for channel-by-channel supervision, meaning the AHCAL calibration should be automated. This motivates research into more sophisticated models to simulate the behaviour of and automatically characterise SiPMs using models that include the effect of detector noise.

Secondly, jet energy resolution in PF is found to be limited most by the capacity for energy deposits in the calorimeters to be correctly associated with individual particles rather than the resolution of the calorimeters for jets with energy greater than around 50 GeV. This concept is called '*confusion*'. Confusion depends primarily on the performance of the sophisticated pattern recognition algorithms used to reconstruct the event. The high granularity of the AHCAL detector enables the device to capture a rich spatiotemporal measurement of particle showers induced by interacting hadrons. In PF, one of the main challenges will be the effective separation of a charged from a neutral particle shower. This fact motivates research into algorithms to optimally exploit the available substructure of hadron shower events to minimise confusion using the AHCAL calorimeter. In addition, the added benefit of timing information is unknown and studying its influence is important in verifying its importance to PF reconstruction.

Thirdly, the resolution of the calorimeters is the most important contributing factor to jet energy resolution for jet energies below 50 GeV. Above this limit, calorimeter resolution plays a secondary but essential role in reducing confusion. More accurate measurements of the energy of a hadron shower can improve the matching of tracks to energy depositions by charged particles, an essential ingredient for PF reconstruction. This again motivates research into algorithms that can help to improve the detector resolution with AHCAL. The influence of time information on the reconstruction is also relatively unknown.

These topics are the focus of the research performed in this thesis, which is structured as follows:

- ▶ In Chapter 2, the physics and motivation for a linear collider, the AHCAL, silicon photomultipliers and machine learning techniques are introduced, as well as concepts critical to the thesis such as key SiPM characterisation parameters and the motivation for calibration tools for SiPMs, as well as the concepts of software compensation and confusion in clustering algorithms designed for Particle Flow are introduced;
- ▶ In Chapter 3, a flexible Monte Carlo program called `LightSintastic` for the simulation of transients and charge spectra for SiPMs is introduced, explained and validated;
- ▶ In Chapter 4, a program for the characterisation of SiPM charge spectra called `PeakOTron` is introduced, which includes noise effects such as after-pulses, dark counts and cross-talk in the detector response model used to describe the spectra. The detector response model is explained, and the pre-fit routine is described to initialise the parameters. The model is then evaluated on simulation from `LightSintastic` and on experimental data;

- ▶ In Chapter 5, a neural network model is devised for software compensation of the AHCAL. The choice of model is explained and compared to the performance of a standard CALICE control method. The models are trained on simulation and experimental data, and the linearity of response, resolution and comparisons between features of the models trained on simulation and data, as well as with and without 100 ps timing information are presented and studied;
- ▶ In Chapter 6, three neural network models from the literature are applied to shower separation of a charged and synthetic-neutral hadron shower observed with AHCAL by reconstruction of the energy fractions belonging to each hadron shower. Considerations are made to assess the quality of synthetic neutral hadron showers produced from charged hadron showers measured with AHCAL, and the appropriate inter-shower distance distributions are assessed. The results of the models are then compared with each other when trained on simulation and data and with and without timing information, and the properties of the shower separation algorithms presented and studied;
- ▶ Finally, In Chapter 7, the conclusions of each study chapter are summarised.

Each study chapter includes its own independent overview and conclusion, motivating and summarising each study in depth. Furthermore, the contributions by the author to each chapter is highlighted in a box at the start of the text.



# Background and Motivation

# 2

This section acts to motivate the research of this thesis. Firstly, the state of modern particle physics is discussed. Then the case is made for an electron-positron linear collider to extend the current understanding of particle physics. Next, an overview of the energy deposition processes of particles interacting with matter is provided. Next, an introduction to calorimeters, their design and calorimeter energy resolution is provided. Then, the concept of Particle Flow is motivated. Finally, the apparatus and methods studied in this thesis and summaries of the motivation for the studies performed are discussed.

## 2.1 The Standard Model

The Standard Model (SM) is the theory classifying all known fundamental particles and three of the four forces of nature (electromagnetism and the weak and strong forces). The Standard Model describes seventeen known fundamental particles, not including antiparticles. These are split into two groups and four subgroups based on their observed properties.

The main two groups of particles are *fermions* and *bosons*. Fermions are particles with a half-integer spin, while bosons have an integer spin. Fermions also obey the Pauli Exclusion principle, whereas bosons do not.

### Pauli Exclusion Principle

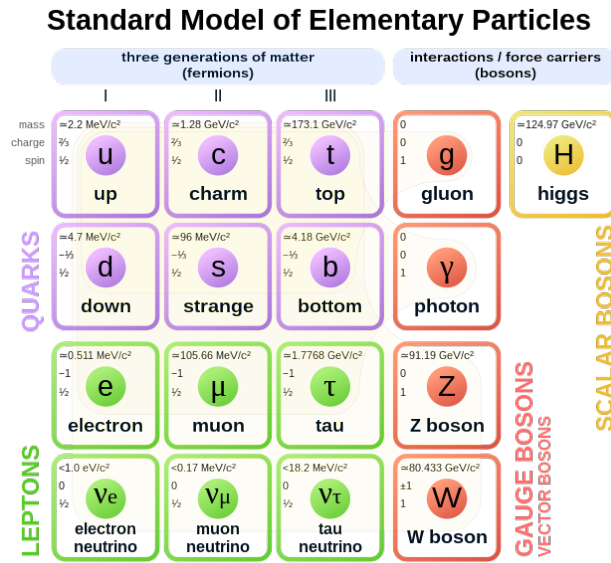
"Two particles with non-integer spin cannot occupy the same quantum state."

There are two other groups of fermions: the first class of fermions are *quarks*. These elementary spin-half particles interact through the strong nuclear force, electromagnetic force, weak nuclear force, and gravity. There are six flavours of quark: the *up*, *charm* and *top* ( $u$ ,  $c$ ,  $t$ ) quarks have an electric charge of  $+\frac{2}{3}q_0$  and the *down*, *strange* and *bottom* ( $d$ ,  $s$ ,  $b$ ) quark have an electric charge of  $-\frac{1}{3}q_0$ , where  $q_0$  is the elementary charge. Quarks carry an additional charge, known as 'colour'. There are six labels for colour charges: red, green, blue, and their 'anti-colours'. Quarks in nature always form bound states in the form 'colourless' triplets or pairs (*baryons* and *mesons*, respectively), and some forms of exotic matter<sup>1</sup>, constituting most of the visible matter in the universe.

The second class of fermions are *leptons*. These elementary spin-half particles interact through electromagnetism, weak nuclear force, and gravity. They are distinguished from quarks because they do not interact via the strong nuclear force. There are three flavours of lepton, each with a charged and neutral variant. The *electron*, *muon* and *tau lepton* ( $e^-$ ,  $\mu^-$ ,  $\tau^-$ ) are the charged leptons. These share the same charge and spin yet differ

2.1 The Standard Model . . . . .	5
The Higgs Boson . . . . .	7
Top Quark . . . . .	8
2.2 Linear Lepton Colliders . . . . .	8
Overview of Lepton Colliders . . . . .	8
The Importance of Jets at Future Lepton Colliders . . . . .	10
Example Measurements at Future Lepton Colliders . . . . .	13
2.3 Interaction of Particles With Matter . . . . .	15
Ionisation . . . . .	15
Energy Loss Mechanisms of Electrons and Positrons . . . . .	16
Energy Loss Mechanisms of Photons . . . . .	16
Electromagnetic Showers . . . . .	16
Energy Loss Mechanisms of Hadrons . . . . .	18
Hadronic Showers . . . . .	20
Simulations of Hadron Showers . . . . .	22
2.4 Calorimeters . . . . .	24
Design of Calorimeters . . . . .	25
Energy Resolution of Sampling Calorimeters . . . . .	26
Improving Hadronic Energy Resolution . . . . .	28
Linearity of Response . . . . .	30
Particle Flow . . . . .	30
2.5 Apparatus and Experimental Methods . . . . .	34
CALICE AHCAL . . . . .	34
Silicon Photomultipliers . . . . .	42
Characterisation of SiPMs . . . . .	47
Machine Learning . . . . .	51

1: An exotic form of matter known as 'pentaquarks' have been observed by the LHCb Collaboration in 2015 [3], which is a bound state consisting of five quarks. These particles have thus far only been observed under experimental conditions. They are not currently known to occur naturally.



**Figure 2.1:** The Standard Model of particle physics. Purple, green, red and yellow lines indicate quarks, leptons, gauge bosons and scalar bosons, respectively. Masses are shown in units of  $\text{GeV}/c^2$  (see Section 2.2.1). Taken from [2].

in invariant mass. They interact through electromagnetism, the weak nuclear force and gravity. The *electron*, *muon* and *tau neutrinos* ( $\nu_e, \nu_\mu, \nu_\tau$ ) are the neutral leptons. These particles have a neutral charge and are the lightest particles of the Standard Model. They interact through gravity and the weak nuclear force.

Every fermion has a partner known as an '*antiparticle*'. Antiparticles of quarks and charged leptons have the opposite charge to their partner. An antiparticle relevant to this thesis is the '*positron*' ( $e^+$ ), which has the same mass and opposite charge to the electron ( $e^-$ ). Neutrinos also have antiparticles called '*antineutrinos*', with neutral charges.

Bosons mediate the interactions of nature: when fermions interact, it is through the exchange of bosons. Bosons are grouped in two classes: *gauge bosons*, with unit spin, and *scalar bosons*, with zero spin. The gauge bosons are as follows: photons and gluons ( $\gamma, g$ ) are massless bosons responsible for mediating the electromagnetic and strong forces, respectively. The charged  $W^\pm$  and neutral Z bosons mediate the weak force. The Higgs boson is a unique scalar boson discussed in Section 2.1.1.

The Standard Model has been demonstrated to be a highly robust and predictive description of the observable universe since its inception in the mid-part of the 20<sup>th</sup> century [4]. However, it is not without its current limitations. For example, the Standard Model cannot explain several observed natural phenomena. Three examples are the hierarchy of the strengths of fundamental forces, the nature and properties of dark matter and dark energy, which are expected to account for the bulk of the mass-energy of the universe and the abundance of matter compared to antimatter (matter composed of antiparticles) in the universe ('*baryogenesis*'). Concepts such as these are described as '*beyond Standard Model*' (BSM) physics.

Modern experimental particle physics seeks to validate the Standard Model to discover unknown physical phenomena that might explain

such discrepancies. This can be achieved by studying the properties of existing particles and the presence of new particles and interactions.

### 2.1.1 The Higgs Boson

The Higgs is a scalar boson with a neutral charge. As of 2020, the most accurate measurement of the Higgs mass has been measured to be  $m_H = 125.38 \pm 0.14 \text{ GeV}/c^2$  by the CMS experiment, utilising a combination of the quad-leptonic ( $H \rightarrow ZZ \rightarrow 4\ell$ ) and diphoton ('golden',  $H \rightarrow \gamma\gamma$ ) decay channels [5].

A longstanding problem with the Standard Model was the paradox that introducing particle masses violated the local gauge invariance of quantum electrodynamics (QED). QED is one of the most experimentally accurate physical theories in existence and describes the interactions of particles through electromagnetism. In theory, gauge invariance for bosons was achieved at the time under the assumption that they were massless. When the  $W$  and  $Z$  bosons were discovered in 1983 [6], they were observed to be among the heaviest particles of the Standard Model, with rest masses of around  $80 \text{ GeV}/c^2$  and  $91 \text{ GeV}/c^2$ , respectively. This mass could not be explained by the existing theory for QED at the time.

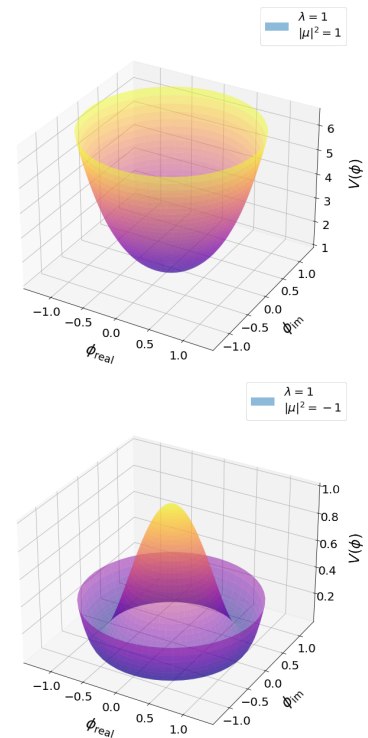
The massive properties of the  $W$  and  $Z$  bosons were eventually explained by adding a self-interacting complex scalar field to the QED Lagrangian. The potential permits two solutions: one with a minimum value corresponding to the vacuum state, which describes massless bosons such as photons, and another with a non-zero vacuum expectation value of the potential. Configurations of the potential are depicted in Figure 2.2. This non-zero vacuum state permits the scalar potential to acquire mass through a process known as 'spontaneous symmetry breaking' [7].

Since a particle consistent with the Standard Model Higgs boson was jointly discovered by the ATLAS and CMS experiments at the LHC in 2012, [1], a primary focus of contemporary particle physics is to make precise, thorough measurements of its properties.

To establish a comprehensive understanding of the Higgs boson's properties and its connection to the symmetry-breaking mechanism, three crucial experimental studies must be conducted [8]:

- ▶ the mass, lifetime and width, spin and CP (*charge conjugation parity*) quantum numbers <sup>2</sup> of the particle must be measured with high precision;
- ▶ The couplings of the particle to the weak vector bosons, leptons and quarks must be found to increase linearly with their mass;
- ▶ The particle's self-coupling must be measured since this quantity determines the aforementioned non-zero vacuum expectation value of the Higgs field.

Confirmation that the particle behaves consistently with a Standard Model Higgs boson ultimately validates the current understanding of the symmetry-breaking mechanism in the context of the Standard Model and would complete the framework of the theory as it is currently understood. For instance, it can help to explain the mass hierarchy of the fermions of



**Figure 2.2:** Higgs 'Mexican Hat Potential',  $V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4$ , where  $\phi = 1/\sqrt{2}(\phi_1 + i\phi_2)$  is a complex scalar field,  $\mu$  is a symmetry-breaking term and  $\lambda$  is a self-interaction term. Top: symmetry breaking term  $\mu^2 > 0$ . Bottom: symmetry breaking term  $\mu^2 < 0$ . [7]

2: CP (charge conjugation parity) quantum numbers refer to the eigenvalue of the C (charge) and P (parity) quantum operators. CP-symmetry refers to the notion that a particle behaves the same way if the signs of the charge and the momentum are flipped. The Standard Model predicts that the Higgs boson is 'CP-even', meaning that it obeys this symmetry.



the Standard Model through a process called '*fermionic mass generation*', whereby the strength of the coupling of the Higgs field helps to determine their mass [7]. Therefore, it is currently of great interest to the particle physics community to investigate such properties of the Higgs boson.

### 2.1.2 Top Quark

The top quark is a fermion has a charge of  $\frac{2}{3}q_0$ . It was discovered in 1995 by the CDF and D0 experiments at Tevatron, located at Fermilab in the USA [9, 10] and completed the full set of quarks in the Standard Model. It is the heaviest particle of the Standard Model. As of 2023, the CMS experiment has measured to top quark mass using the channel  $t\bar{t} \rightarrow W^-bW^+\bar{b}$  to quarks, leptons and neutrinos (see Section 2.2.1) to be  $171.77 \pm 0.37 \text{ GeV}/c^2$  [11].

The top quark coupling to the Higgs is the largest expected among fermions. It, therefore, plays a role in the determination of the higher-order processes that involve the Higgs boson<sup>3</sup> [12]. Subsequently, simultaneous measurements of the top quark and Higgs masses allow a comparison of the measured values with the predictions of the Standard Model, which would provide evidence for the validity of the existing theory. Furthermore, the higher-order processes involving the top quark influence the shape of the Higgs potential. This can have implications for the long-term stability of the Higgs field and the properties of the particles to which it couples (*vacuum-stability*). Therefore, precision measurements of the properties of the top quark are also of great interest to the scientific community.

3: Higher-order processes refer to processes that involve more than the minimum number of interaction vertices in a Feynman diagram (a 'leading order diagram'). These processes can be understood as more complicated interactions and are less probable than leading order processes. They can nonetheless influence calculations of Standard Model observables, such as cross-sections and particle masses.

## 2.2 Linear Lepton Colliders

### 2.2.1 Overview of Lepton Colliders

A *particle collider* is a device designed to collide particles at high energy, typically of the order gigaelectronvolts<sup>4</sup> (GeV) to teraelectronvolts (TeV) at the modern energy frontier to study their constituents and the fundamental interactions between the particles of nature.

Currently, perhaps the most famous particle collider in the world is the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland, with a centre-of-mass energy of  $\sqrt{s} = 14 \text{ TeV}$ . The LHC is designed to collide protons or heavy ions and was responsible for the Higgs discovery discussed in the preceding section. Here,  $s$  is a Mandelstam variable which encodes the energy, momentum and angles of a scattering process in a Lorentz-invariant fashion, and  $\sqrt{s}$  is the centre-of-mass energy of the collision between two relativistic particles.

This thesis focuses on lepton colliders, unlike hadron colliders such as the LHC. Lepton colliders are generally designed with the goal of precision measurements in particle physics for two main reasons:

- ▶ leptons are fundamental, pointlike, non-composite particles. Each collided particle has a precisely defined initial state and momentum from collision to collision, which aids in event reconstruction and

4: Particle energies are typically measured in electron-volts (eV). An electron-volt is the energy required to move an elementary charge through one volt of electric potential.



results in collisions where all four-momentum contributes to the hard interaction. By contrast, hadrons are composite (composed of *partons*, or quarks and gluons, collectively), with momentum split with an unknown fraction between not only the ‘valence’ quarks which give the hadron its properties but also the ‘quark-gluon sea’ that is predicted to exist inside hadrons at highly-relativistic momenta due to the Heisenberg Uncertainty Principle;

- ▶ leptons may only interact by electromagnetism, the weak force and gravity, which drastically reduces the number of possible interactions that occur during a collision compared to hadrons, the latter of which may also interact via the strong force. Lepton colliders, therefore, benefit from significantly less background than if hadrons were used;

### Heisenberg Uncertainty Principle

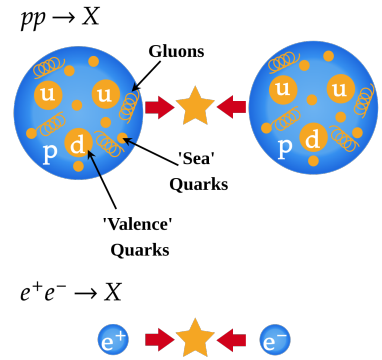
*The product of the uncertainties on the position and momentum (equivalently, energy and lifetime) of a particle is never less than  $\frac{\hbar}{2}$ .*

Two designs of particle colliders exist: circular and linear. The former accelerates particles continuously before collision in a ring; the latter accelerates them only once. While circular colliders have been used with leptons in the past (e.g. Large Electron-Positron Collider, or LEP, with  $\sqrt{s} = 209$  GeV, originally located at CERN in Geneva, Switzerland), the linear collider design is typically used for leptons. This is because of a phenomenon known as ‘*synchrotron radiation*’, which causes photons to be radiated from accelerating charged particles.

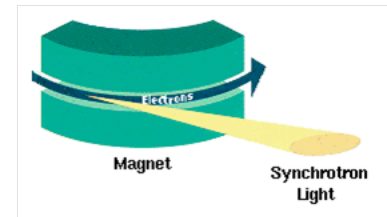
This process affects transverse acceleration more than longitudinal acceleration and arises due to special relativity. The radiated power from this source of radiation and scales as  $P \propto \frac{\gamma^4}{R}$ ,  $\gamma = \frac{E_{\text{particle}}}{m_{\text{particle}}}$ , where  $\gamma$  is the Lorentz factor,  $E_{\text{particle}}$  is the total energy of the particle, and  $m_{\text{particle}}$  is the rest mass of the particle and  $R$  is the radius of the synchrotron. For this reason, electrons radiate photons in this way a factor of  $\left(\frac{m_p}{m_e}\right)^4 = 1.13 \times 10^{13}$  faster than protons in a synchrotron, where  $m_p$  and  $m_e$  are the proton and electron rest masses. One solution to this problem is to increase the radius of the circular collider, but it is often prohibitively costly to build ever-larger instrumentation for this purpose. Since linear colliders only accelerate the particle in the longitudinal direction, synchrotron radiation is greatly reduced in this design<sup>5</sup>. In addition, one of the advantages of a linear collider is its ability to operate, with minimal modification, at any energy within a wide range that its technology makes available [14], thereby allowing a broad physics program.

$e^+e^-$  colliders are therefore the experiments of choice for precision measurements of the Higgs boson, the electroweak sector and potentially for BSM physics searches. An example of such an accelerator that has already been built is the Stanford Linear Accelerator ( $\sqrt{s} \approx 91$  GeV, located at SLAC in California, USA), which was used to discover the charm quark in 1974 [16] and the  $\tau$  lepton in 1975 [17].

It should be noted that linear colliders are not the only available option for a future precision collider. An alternative experiment that has been

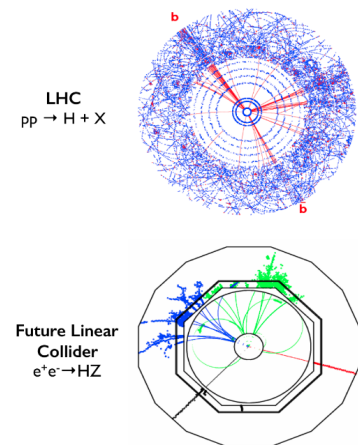


**Figure 2.3:** Diagram illustrating the difference between relativistic collisions of hadrons (protons) and leptons (electron/positron). The substructure of the proton is indicated by the orange circles (quarks) and gluons.



**Figure 2.4:** Illustration of the tangential emission of synchrotron radiation from accelerated electrons. Taken from [13].

5: Synchrotron radiation is not always an undesirable effect, since it may be used as a high-intensity X-ray source i.e. PETRA III at DESY, Hamburg, Germany.



**Figure 2.5:** Illustration of event displays illustrating the ‘cleaner’ events observed in the detector array with a future linear collider compared to a hadron collider. Top: LHC  $pp \rightarrow H + X$  process. Bottom: FLC  $e^+e^- \rightarrow HZ$ . Taken from [15].

proposed is the Future Circular Collider (FCC-ee) circular collider, which can study the  $Z$  and  $W^\pm$  bosons, Higgs boson and top quark and would have a  $\sqrt{s}$  of between 88 and 365 GeV [18].

This thesis nonetheless focuses on future linear  $e^+e^-$  colliders. Summaries of two proposed future  $e^+e^-$  linear collider programs are briefly summarised for reference.

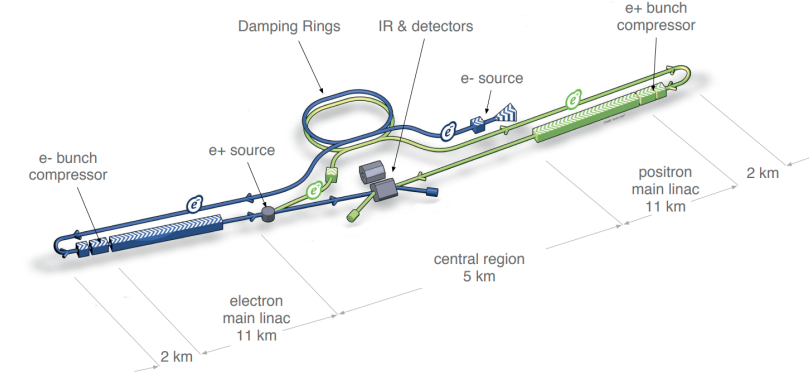
**The International Linear Collider (ILC)** The ILC is a proposed high-luminosity linear  $e^+e^-$  collider experiment based on 1.3 GHz superconducting radio-frequency (SCRF) accelerating technology [14]. The accelerator would be  $\sim 31$  km in length. It would initially have a  $\sqrt{s} = 250$  GeV, and would be initially designed as a ‘*Higgs factory*’ (a collider that produces a high frequency of Higgs bosons). The physics program could be extended in an upgrade program to achieve  $\sqrt{s} = 500$  GeV-1 TeV. It is also noted that ILC will produce a bunch crossing rate of around 5 Hz. A schematic layout of the proposed collider is shown in Figure 2.6a.

**The Compact Linear Collider (CLIC)** The CLIC is also a proposed high-luminosity linear collider experiment based on a 12 GHz conventional radio-frequency (RF) source [19]. The accelerator would be 11.4 km in length. The accelerator physics program would be extended in three stages:  $\sqrt{s} = 380$  GeV,  $\sqrt{s} = 1$  TeV, and  $\sqrt{s} = 3$  TeV. A schematic layout of the proposed collider is shown in Figure 2.6b.

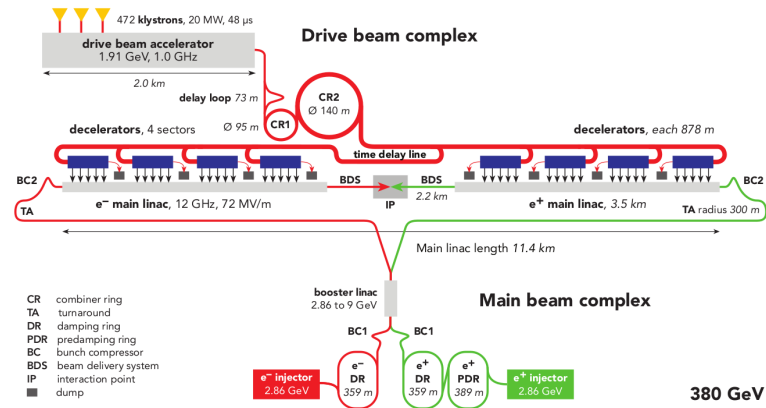
## 2.2.2 The Importance of Jets at Future Lepton Colliders

As previously mentioned, Higgs boson precision measurements are a consistent priority for the physics programs of future lepton collider experiments. Additionally, the electroweak sector and top quark precision measurements are also of interest.

The Higgs and top quark production cross-sections available to a lepton collider experiment operating in the centre-of-mass energy range achievable by ILC and CLIC, as well as the Higgs decay branching ratios for the low-mass range of the Higgs invariant mass, are shown in Figure 2.7. Two of the major Higgs production modes that may be studied at ILC and CLIC, ‘*Higgstrahlung*’ ( $e^+e^- \rightarrow ZH$ ) and ‘*vector boson fusion*’ ( $e^+e^- \rightarrow H\nu_e\bar{\nu}_e$ ), and the main top production mode ( $e^+e^- \rightarrow t\bar{t}$ ), are shown as Feynman diagrams in Figure 2.8. The branching ratios to hadrons of the  $W$  and  $Z$  bosons to hadrons are shown in Table 2.1.



(a) Schematic layout of the ILC, indicating all the major subsystems (not to scale). Taken from [14].



(b) Schematic layout of the CLIC, indicating all the major subsystems (not to scale). Taken from [19].

Figure 2.6: Schematic layouts of ILC and CLIC, shown in Figures 2.6a and 2.6b, respectively.

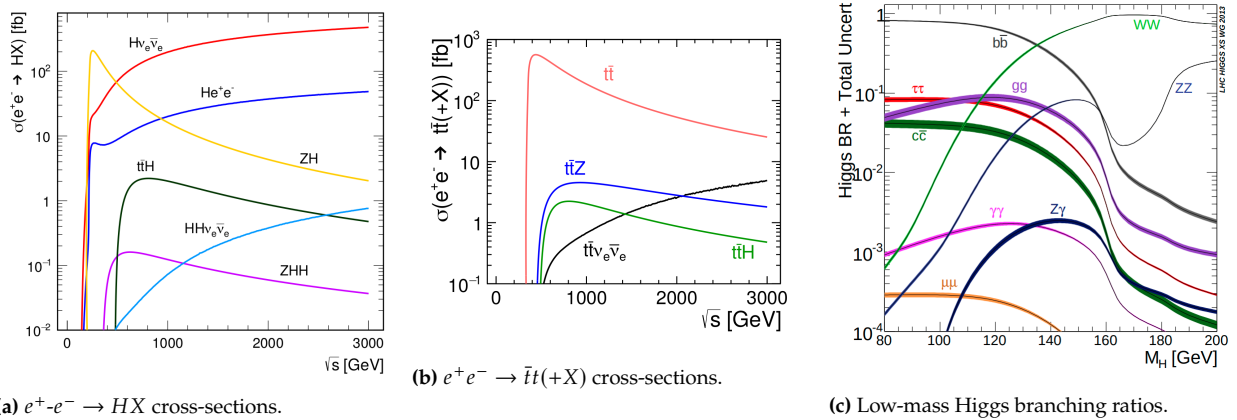
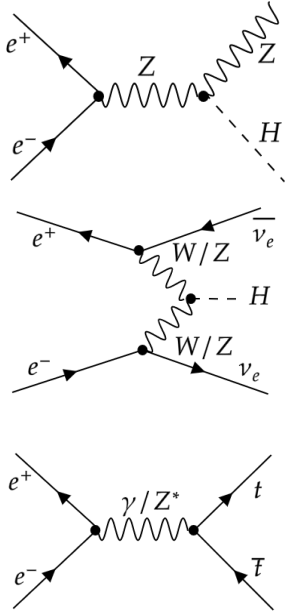


Figure 2.7: Figures 2.7a and 2.7b show the cross-sections of  $e^+e^- \rightarrow HX$  and  $e^+e^- \rightarrow t\bar{t}(+X)$  in the  $\sqrt{s} = 0 - 3$  TeV range. Both taken from [19]. Figure 2.7c shows the Higgs branching ratios for the low mass range on the Higgs. Taken from [20].

Particle	$\frac{\Gamma_{\text{Hadron}}}{\Gamma}$
$W^\pm$	$67.41 \pm 0.27$
$Z$	$69.911 \pm 0.056$

**Table 2.1:** Hadronic decay fractions produced during the decays of the  $W$  and  $Z$  weak vector bosons. Values obtained from [21].



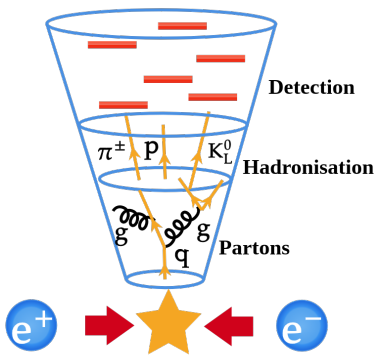
**Figure 2.8:** Feynman diagrams of the two main Higgs production interactions and the main top quark production interaction at linear colliders such as ILC and CLIC. The top diagram shows Higgstrahlung (yellow line of Figure 2.7a). The middle diagram shows Vector Boson Fusion. The bottom diagram shows direct  $t\bar{t}$  production (red line of Figure 2.7b). Higher order processes of this interaction can also produce a Higgs (dark green line of Figure 2.7a)

The information in these figures and tables highlight an important observation about the interactions that are intended to be studied with a future linear collider experiment:

- ▶ Higgs bosons decay predominantly into  $b\bar{b}$  pairs, or weak vector boson pairs, as demonstrated by the black, green and blue lines of Figure 2.7a. The  $W$  and  $Z$  bosons also decay frequently to hadron final states, shown in Table 2.1;
- ▶ Higgstrahlung produces a  $Z$  boson, shown as the yellow line of Figure 2.7a. This fact is critical for Higgs mass measurements, which is discussed fully in Section 2.2.3.
- ▶ the primary production mode for top quarks is via direct production of  $t\bar{t}$  quark pairs. However, the top quark is not observed directly, instead decaying via  $t\bar{t} \rightarrow W^+bW^-\bar{b}$ , which itself decays with a branching ratio of 45.7% into final states involving neutrinos and quarks only and with 54.1% into final states of charged leptons, neutrinos and quarks [21]. Notably, the top quark production can radiate a  $Z$  boson or a Higgs boson as a higher-order process, as shown by the blue and green lines of Figure 2.7b.

Quarks are therefore produced in many of the most probable interactions involving the Higgs and in top quark production. As previously mentioned, quarks have a colour charge and cannot exist in isolation. Instead, the quarks produced from inelastic interactions of high-energy particles form bound states in a process known as *hadronisation*. At the highly relativistic energies of linear collider experiments, the interactions typically manifest as a final state in the form of highly collimated beams of particles known as ‘jets’.

Figure 2.10 illustrates the formation of a jet. At the extremely high energy densities observed at the interaction point of  $e^+e^- \rightarrow X' \rightarrow q\bar{q}X$ , quarks may be viewed as asymptotically free and emit soft gluon radiation as they propagate through space, which may themselves decay into  $q\bar{q}$  pairs. Almost instantaneously following this stage, the quarks and gluons hadronise. The final-state particles that constitute the jet are charged and neutral hadrons, charged leptons and neutrinos. Except for neutrinos, the energy of these particles may be measured with detectors placed around the interaction point of the linear collider, from which events are reconstructed (see Section 2.4.1).



**Figure 2.9:** A diagram illustrating the formation of a jet from a quark produced in  $e^+e^- \rightarrow X' \rightarrow q\bar{q}X$ .

Jet energy resolution describes how well the detector can reconstruct the jet’s energy, which is the sum of the total energy carried by each particle depositing energy in the calorimeter, correctly attributed to the jet. A state-of-the-art jet energy resolution is required for both ILC and CLIC since it is essential for the unambiguous identification of many decay channels and enhances precision measurements by reducing the integrated luminosity required for many measurements by processes with final states involving quarks [22].

An example observable where jet energy resolution is relevant is the invariant mass<sup>6</sup> of the particle(s) that initiated the jets. For instance, the invariant mass of jets from  $W$  and  $Z$  bosons allows unambiguous distinction between processes involving them (see Section 2.4.5).

Therefore, the detector apparatus used in future linear lepton collider experiments designed for precision measurements of Higgs properties must be tailored specifically to have state-of-the-art jet energy resolution.

### 2.2.3 Example Measurements at Future Lepton Colliders

To illustrate the importance of jet energy resolution to future linear collider experiments, two complimentary examples of experiments pertinent to modern particle physics are outlined where excellent jet energy resolution at future lepton colliders plays a role in precision Higgs measurements, using the ILC as an example:

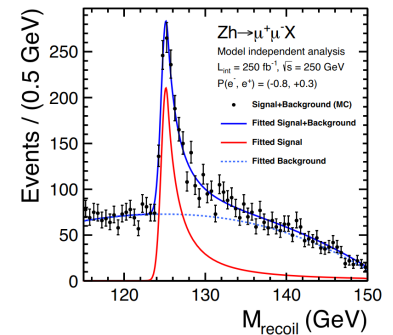
**Higgs Recoil Mass Measurement** Perhaps one of the most critical Higgs measurements that can be made at a lepton collider is a model-independent measurement of the Higgs mass and the Higgstrahlung cross-section. As mentioned in Section 2.2.1, the initial conditions (i.e. the centre of mass energy and momentum) of a lepton collision are precisely defined, which is not the case for a hadron collision. Using the Higgstrahlung process ( $e^+e^- \rightarrow ZH$ ), knowledge of the centre-of-mass energy of the collision and the momenta of the decay products of the  $Z$  can be used to reconstruct the Higgs mass indirectly via the relationship  $M_R^2 = (p_{CM} - p_{Z \rightarrow X})^2$ , where the  $M_R$  is the recoil mass,  $p_{CM}$  is the centre-of-mass energy and  $p_{Z \rightarrow X}$  is the sum of the momenta of the decay products ( $c = 1$  in this definition). The critical advantage of this method at a lepton collider is that any invisible decays of the Higgs can be included in the mass measurement, which is impossible at a hadron collider for the reasons presented.

The recoil mass study is typically presented with  $p_{Z \rightarrow X}$  as the leptonic decay channel,  $Z \rightarrow \mu^+\mu^-$  and  $Z \rightarrow e^+e^-$ . This is because (anti)muons and electrons/positrons can be identified precisely from their energy signatures, have well-defined kinematic properties and have comparatively low backgrounds compared to other decays.

There would be two important experimental outcomes using this method. Firstly, with an integrated luminosity of  $250 \text{ fb}^{-1}$  at ILC, uncertainty on the Higgs mass of at least  $30 \text{ MeV}$  can be achieved, [23] which is a factor of 5 smaller than the current best measurement by CMS [5]. Additionally, the same recoil mass study at ILC would be an independent validation of this study.

However, the branching ratio of the  $Z$  to leptons is small ( $\text{BR}(Z \rightarrow \mu^+\mu^-) \simeq \text{BR}(Z \rightarrow e^+e^-) \simeq 3.4\%$  [21]). By contrast, the branching ratio of the hadronic decays of the  $Z$  is more than an order of magnitude larger ( $\text{BR}(Z \rightarrow q\bar{q}) \simeq 70\%$ , see Table 2.1). Achieving state-of-the-art jet energy resolution is crucial to utilise this channel effectively. Combining the lepton and hadron channels could improve the precision of the mass and the cross-section, or the required integrated luminosity to achieve the same precision could be reduced [23].

6: Invariant mass describes a characteristic of particles that is conserved under Lorentz transformation (i.e. all reference frames). It is reconstructed at lepton collider experiments from the square root of the difference between the total energy and three-momentum of all particles that were involved in an interaction, which may be comprised of several jets for the interactions involving decays to quarks.



**Figure 2.10:** Recoil mass distribution and model-independent analysis of the process  $HZ \rightarrow \mu^+\mu^-X$  in simulation. The solid red line, dashed and solid blue line indicate a fit to signal, background and a combined fit, respectively. Taken from [8].

### Higgs Couplings to Standard Model Particles and Higgs Full Width

According to the Standard Model, the Higgs couplings to weak vector bosons, charged leptons, and quarks are proportional to their mass (i.e.  $g_{HAA} \propto M_A$ , where  $g_{HAA}$  is the coupling, and  $M_A$  is the mass of the decay product). The couplings are important verifications of the Standard Model, which expect 'generation universality', or the idea that the Higgs ought to couple to the different generations of particles equally. Deviations from this expectation could indicate BSM physics [8].

The square of the Higgs coupling is proportional to the partial width of the process (i.e.  $g_{HAA}^2 \propto \Gamma(H \rightarrow AA) = \Gamma_H \cdot \text{BR}(H \rightarrow AA)$ , where  $\Gamma(H \rightarrow AA)$  and  $\text{BR}(H \rightarrow AA)$  are the partial width and branching ratio of  $H \rightarrow AA$ ,  $\Gamma_H$  is the full width of the Higgs and  $A$  is a particle to which the Higgs couples. It is then the determination of  $\Gamma_H$  with high precision and in a model-independent way that is the most major stepping stone to determining the absolute scale of the couplings to Standard Model particles.

At ILC, Higgstrahlung at  $\sqrt{s} = 250$  GeV, the product of the cross-section and branching ratio of  $H \rightarrow AA$ ,  $\sigma_{ZH} \cdot \Gamma(H \rightarrow AA)$ , the cross-section,  $\sigma_{ZH}$  and the partial width for  $H \rightarrow AA$ ,  $\Gamma(H \rightarrow AA)$  can all be obtained with high precision. For instance, with an integrated luminosity of  $250 \text{ fb}^{-1}$ , the the Higgstrahlung cross-section,  $\sigma_{ZH}$  can also be measured, with a resolution of  $\Delta\sigma_{ZH}/\sigma_{ZH} = 2.6\%$  [8] using the well-known branching ratios of the  $Z$ .

From this combined information, the branching ratios can be isolated, and the full width reconstructed from  $\Gamma_H = \Gamma(H \rightarrow AA)/\text{BR}(H \rightarrow AA)$ . For the reconstruction of  $\Gamma_H$ ,  $A = Z, W^\pm$  are chosen. With the addition of the bottom quark decay of the  $W$ -boson fusion process,  $\sigma_{v\bar{v}H} \times \text{BR}(H \rightarrow b\bar{b})$ ,  $\Gamma(H \rightarrow W^+W^-)$  can be reconstructed at  $\sqrt{s} = 500$  GeV, at which this process becomes much more probable. Combining measurements at  $\sqrt{s} = 250$  GeV with  $250 \text{ fb}^{-1}$  integrated luminosity and at  $\sqrt{s} = 500$  GeV with  $500 \text{ fb}^{-1}$  integrated luminosity, a model-independent measurement of  $\Gamma_H$  with resolution  $\Delta\Gamma_H/\Gamma_H = 5\%$  can be achieved with ILC [24].

Presently, the most accurate measurement of  $\Gamma_H$  was made in 2023 by the ATLAS experiment utilising the  $H \rightarrow ZZ \rightarrow 4\ell$  and  $H \rightarrow ZZ \rightarrow 2\ell 2\nu$  channels with  $\ell = e, \mu$ . Under the assumption that no new particles enter the production of the virtual Higgs boson, a value of  $\Gamma_H = 4.5_{-2.5}^{+3.3} \text{ MeV}$  was obtained [25]. This corresponds to a resolution on the full width of more than 50%. A measurement from the ILC using the method described therefore improves the resolution on the  $\Gamma_H$  by a factor of 10 and again would serve as an accurate and model-independent cross-check to this study.

A straightforward need for exceptional jet energy resolution in this measurement is the need to distinguish the dijet events from  $H \rightarrow ZZ \rightarrow q\bar{q}q\bar{q}$  and  $H \rightarrow W^+W^- \rightarrow q\bar{q}q\bar{q}$  in this procedure. This can be achieved if the invariant mass distributions of the  $W^\pm$  and  $Z$  jets can be cleanly separated during analysis. More details on this topic are given in Section 2.4.5.



## 2.3 Interaction of Particles With Matter

It is necessary to review the physical processes that govern energy loss in matter to understand the design of energy detectors intended to be used in future lepton colliders such as ILC and CLIC. A summary of the main energy depositing processes and associated phenomena is introduced in this section.

### 2.3.1 Ionisation

When traversing matter, charged hadrons and leptons may *ionise* the detector medium. Ionisation means the liberation of electrons from an atomic nucleus, resulting in a charged remnant called an *ion*. Ionisation is particularly relevant to the energy loss of 'heavy' charged particles (i.e. with an invariant mass larger than the electron, most notably muons, for which it is the main energy loss mechanism in matter)<sup>7</sup> The average energy loss of charged particles due to ionisation is governed by the '*Bethe-Bloch Equation*', shown in Equation 2.1 and in Figure 2.11:

$$\left\langle \frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left( \frac{1}{2} \log \left( \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} \right) - \beta^2 - \frac{\delta}{2} \right) \quad (2.1)$$

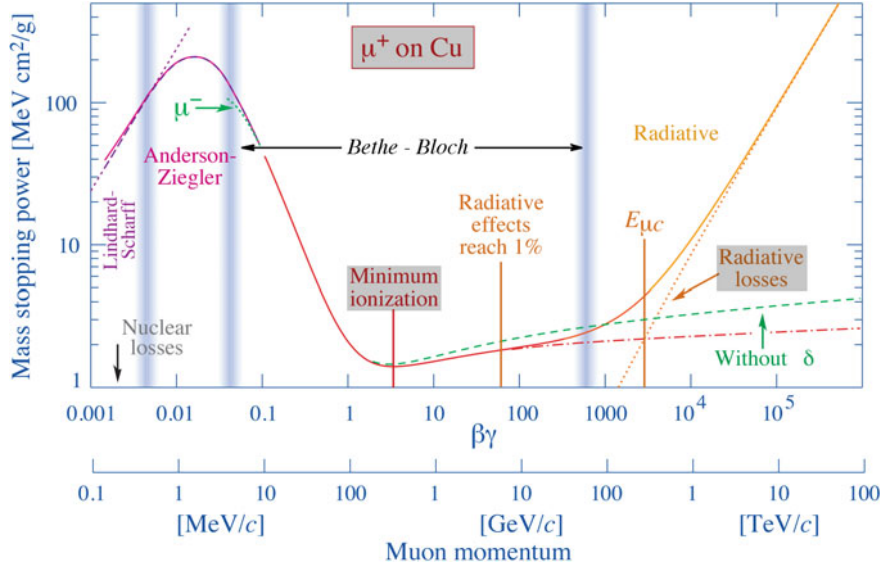
Where  $Z$  and  $A$  are the atomic and mass number of the traversed matter,  $I$  is the mean excitation energy of the traversed matter,  $z$  is the number of elementary charges carried by the traversing charged particle,  $W_{\max}$  is the maximum transferable energy by the particle,  $K = 4\pi N_A r_e^2 m_e c^2$ , where  $N_A$  is Avogadro's number and  $r_e$  is the radius of the electron,  $c$  is the speed of light,  $\beta$  is Lorentz beta, and  $\delta$  is a correction term describing the density effect<sup>8</sup>.

While the Bethe-Bloch Equation describes the average energy deposition of a massive charged particle, the distribution of ionising energy loss of a single particle is described in thin sheets of matter by the Landau distribution [27]. This is a highly skewed distribution. Therefore, the distribution has a most probable value (MPV) lower than the average. The reason for the skew is that significant fluctuations in the ionisation energy known as '*straggling*' may occur. An extreme example of straggling is when an ionising charged particle transfers its kinetic energy to an atomic electron entirely.

Figure 2.11 shows that the Bethe-Bloch equation has a local minimum, highlighted in red as '*minimum ionisation*'. Particles depositing this energy are known as '*minimum ionising particles*' (MIPs). The closest particle in nature to a minimum ionising particle is a muon. Muons are, nonetheless, therefore often used as a '*standard candle*' for calibrating the energy response of calorimeters at different energies, as the MPV of their deposited energy remains relatively constant with momentum in the range of energies 1-100 GeV/c. However, it should be noted that they are only true minimum ionising particles at the lower end of this momentum range.

7: While all charged particles, experience radiative losses it is suppressed by  $m_{\text{particle}}^{-4}$ , as with synchrotron radiation. Muons have a significantly greater mass than the electron ( $\frac{m_\mu}{m_e} = 208.8$ ). Therefore, radiative effects are suppressed for this particle for energies relevant to modern collider physics.

8: The density effect arises due to the relativistic flattening and lengthening of the electric field of the traversing particle, which increases the range over which the particle can interact with matter, and a competing effect due to polarisation of the absorber medium [26]. The consequence of this effect is shown by comparison of the red and green lines of Figure 2.11.



**Figure 2.11:** A diagram illustrating the behaviour of the Bethe-Bloch equation with momentum,  $\beta\gamma$ , for an anti-muon ( $\mu^+$ ) interacting with Cu absorber. Taken from [28].

### 2.3.2 Energy Loss Mechanisms of Electrons and Positrons

Two main energy regimes exist for the energy loss of electrons and positrons. Above energies of 100 MeV, an energy loss process known as *bremstrahlung*<sup>9</sup> dominates. This process is induced by the interaction of the particle with the electromagnetic fields of atomic nuclei. It, therefore, depends on the proton number of the absorber material ( $Z$ ) and falls off as  $E_{\text{particle}}^{-1}$  [28]. Below 10 MeV, ionisation, with smaller contributions from scattering processes such as Møller<sup>10</sup> and Bahabha<sup>11</sup> scattering dominate (see Figure 2.12a). The crossover point between ionisation and *bremstrahlung* is known as the ‘critical energy’ and is parameterised for solid and liquid matter according to Equation 2.2 [28]:

$$\varepsilon_c = \frac{610}{(Z + 1.24)} [\text{MeV}] \quad (2.2)$$

### 2.3.3 Energy Loss Mechanisms of Photons

Photons interact with matter in a markedly different way than electrons or positrons. In particular, the photon is neutral and cannot ionise the detector medium. Photons with energy  $> 1$  MeV typically lose energy by pair-production<sup>12</sup>. Notably, this process results in the entire energy of the photon being transferred to the pair: the photon no longer exists after these interactions. Below 1 MeV, the photoelectric effect<sup>13</sup> with contributions from photon-scattering processes such as Compton scattering<sup>14</sup> and Rayleigh scattering (although this is not an energy loss mechanism)<sup>15</sup> dominate [28] (see Figure 2.12b).

### 2.3.4 Electromagnetic Showers

Electrons and photons of GeV-scale energy typically induce cascades of particles called ‘*electromagnetic showers*’. While photons produced by

9: *Bremstrahlung* (‘deceleration radiation’) is the emission of photons as a result of a charged particle changing direction in an electromagnetic field. This effect arises due to special relativity. Synchrotron radiation is a type of *bremstrahlung*.

10: Møller scattering describes the elastic scattering process  $e^-e^- \rightarrow e^-e^-$ .

11: Bahabha scattering describes the elastic scattering process  $e^+e^- \rightarrow e^+e^-$ .

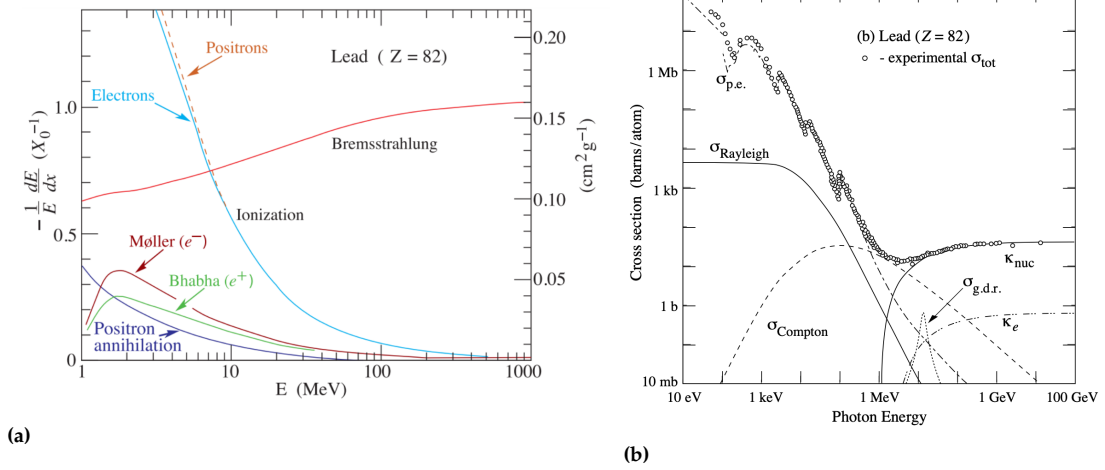
12: Pair-production describes the creation of an electron-positron pair by the interaction of a photon with the electromagnetic field of an atomic nucleus,  $\gamma \rightarrow e^+e^-$ . The photon must have at least enough energy to produce the pair at rest (i.e.  $E_{\text{particle}} \geq 2m_e$ ,  $m_e = 0.511$  MeV).

13: The photoelectric effect describes the excitation and subsequent emission of an electron from an atom by an incident photon.

14: Compton scattering describes the elastic scattering of a photon with a stationary electron, resulting in energy transfer to the electron and consequently a decrease in the photon’s frequency ( $e^-\gamma \rightarrow e^-\gamma$ ).

15: Rayleigh scattering describes photon scattering from a particle much smaller than the wavelength of the light. Photons scattered in this way do not lose energy and affect only the spatial distribution of the energy [28]. This scattering process is why the sky appears blue.





**Figure 2.12:** Figure 2.12a shows energy loss of  $e^\pm$  as a function of particle energy in Pb, by interaction process. Figure 2.12b shows cross-sections for photon interactions in Pb. Cross-sections are defined as follows:  $\sigma_{p.e.}$  indicates photoelectric effect;  $\sigma_{Compton/Rayleigh}$  indicates Compton/Rayleigh scattering;  $\sigma_{g.d.r.}$  is the cross section for ‘giant dipole resonance’, a type of collective excitation where atomic protons move one way and neutrons move another.  $\kappa_{nuc/e}$  indicates pair-production from the electromagnetic field of an atomic nucleus/electron [29]. Both taken from [28].

bremsstrahlung are typically ‘soft’ and undergo Compton scattering or are absorbed by the photoelectric effect, photons of energy above 5 – 10 MeV go on to induce pair-production in the matter they are traversing. The electrons and positrons produced in this way experience bremsstrahlung, thereby radiating more photons. The result is a ‘cascade’ of photons and  $e^+e^-$  pairs that proceed until the energy of the electron (positron) reaches  $\epsilon_c$  of Equation 2.2<sup>16</sup>.

The longitudinal development of electromagnetic showers is described in terms of a quantity called ‘radiation length’ ( $X_0$ )<sup>17</sup>.  $X_0$  is typically parameterised in terms of the atomic number ( $Z$ ) and mass number ( $A$ ) of the absorbing matter, according to Equation 2.3 [28]:

$$X_0 = \frac{716.4 A}{Z(Z+1) \log\left(\frac{287}{\sqrt{Z}}\right)} \left[ \frac{\text{g}}{\text{cm}^2} \right] \quad (2.3)$$

Additionally, the lateral development of a hadron shower can be defined in terms of a ‘Moliere radius’, given by Equation 2.4 [28]:

$$\rho_M = E_s \frac{X_0}{\epsilon_c} \quad (2.4)$$

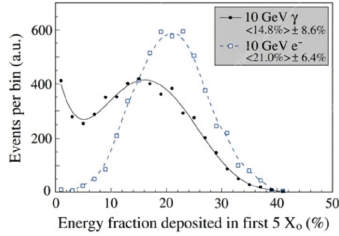
where  $\rho_M$  is the Moliere radius,  $X_0$  is radiation length,  $\epsilon_c$  is the critical energy and  $E_s = 21.2 \text{ MeV}$ . The Moliere radius describes the lateral distance from the axis along which the electromagnetic shower develops (‘shower axis’) in which a hadron shower deposits around 85-90% of its energy.

Electrons and positrons lose energy continuously as soon as they enter matter, while photons interact only after  $\frac{9}{7}X_0$  on average. This behaviour affects the development of electron and photon-induced electromagnetic showers:

16: At  $E_{\text{particle}} \leq \epsilon_c$ , the  $e^+e^-$  pairs are more likely to undergo ionisation or scattering. The positrons may also annihilate with atomic electrons (see Figure 2.12a).

17: The radiation length is defined as the distance after which an incident electron inducing an electromagnetic shower loses a fraction of  $e^{-1}$  (36.8%) of its original energy via bremsstrahlung in matter, or 7/9 of the mean free path of a photon before undergoing pair production in matter.

- ▶ Photon-induced showers deposit energy deeper within matter than electron/positron induced showers;
- ▶ The energy fluctuations of photon showers are typically greater than for electron/positron-induced showers due to variations in the starting position of the photon shower.



**Figure 2.13:** Distributions of energy fraction deposited in the first  $5 X_0$  in Pb of 10 GeV electron and photon showers (taken from [30]).

An example of the differing distributions of energy loss of photons and electrons are shown in Figure 2.13. An illustration of the development of an electron and photon-induced electromagnetic shower is shown in Figure 2.14.

### 2.3.5 Energy Loss Mechanisms of Hadrons

Hadrons interact by the strong force, as well as electromagnetism, the weak force and gravity. Hadrons may induce other energy loss mechanisms and behave differently to electrons, positrons or photons.

A neutral hadron may only interact with matter after first encountering an atomic nucleus. The distance it travels depends on the density of the matter and the inelastic nuclear cross-section with that matter. On average, a neutral hadron only interacts only after one *nuclear interaction length* ( $\lambda_I$ ). By contrast, a charged hadron will typically ionise the detector medium before such a collision occurs.  $\lambda_I$  is always longer than the radiation length<sup>18</sup> (i.e.  $\frac{\lambda_I}{X_0} = 9.5$  in steel [28], see Figure 2.15).  $\lambda_I$  also depends on the incident hadron (e.g. neutrons and pions have different  $\lambda_I$ ). The nuclear interaction length is given by [31]:

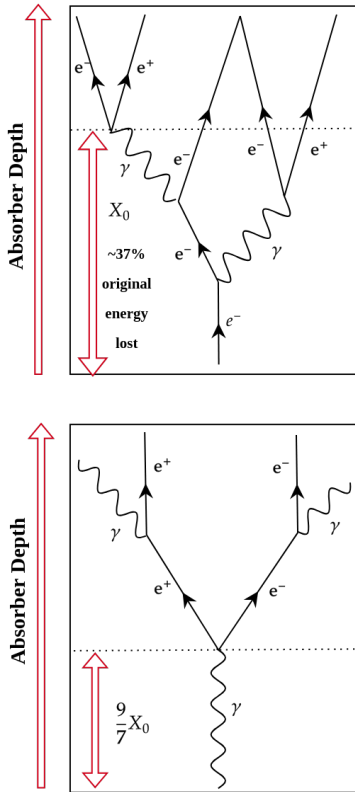
$$\lambda_I = \frac{A}{N_A \cdot \sigma_I} \left[ \frac{\text{g}}{\text{cm}^2} \right] \approx 35 \frac{A^{1/3}}{\rho} [\text{cm}] \quad (2.5)$$

where  $N_A$  is Avogadro’s constant,  $A$  is the mass number of the absorber,  $\sigma_I$  is its nuclear inelastic cross-section, and  $\rho$  is the specific density of the traversed matter.

Hadrons interact with matter mainly by the following processes:

**Nuclear Spallation** This process refers to the ejection of particles in an atomic nucleus (nucleons) and light hadrons (most notably  $\pi^\pm$ ,  $\pi^0$  and  $\eta$  mesons, see Section 2.3.6) in a fast nuclear cascade resulting from quasi-free collisions of an incident hadron with the nucleons. The particles ejected from this interaction are called the ‘*intra-nuclear cascade*’ and may go on to induce further secondary or tertiary spallation events in other atomic nuclei.

**Nuclear Evaporation** Excited nuclei from a collision with a hadron may ‘*evaporate*’, radiating neutrons, gamma rays (photons) and  $\alpha$  particles (He nuclei) as they de-excite.



**Figure 2.14:** Illustration of the development of electron and photon-induced electromagnetic showers. Top: electron-induced EM shower. Bottom: photon-induced EM shower.

18: For this reason, ECALs are always placed before HCALs in the detector arrays of collider experiments.

**Nuclear Fission** Given sufficient energy transfer, the nucleus may undergo fission. This process may release gamma rays (photons) and neutrons.

**Invisible Energy in Hadronic Interactions** Two concepts relevant to this thesis arise from consideration of these processes:

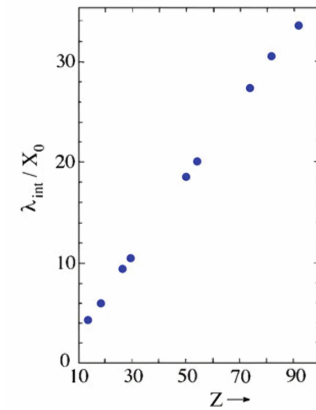
- ▶ the liberation of nucleons from atoms results in energy loss in the form of '*nuclear binding energy*'<sup>19</sup>. This energy cannot be observed and is called '*invisible energy*'. To a lesser extent, invisible energy can also be in the form of particles produced in the intra-nuclear cascade (e.g. neutrinos produced by  $\pi^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu}_\mu)$ ). The fraction of invisible energy is also highly stochastic from event to event;
- ▶ neutrons are abundantly produced in these processes, and their overall number is directly proportional to the proportion of invisible energy lost [28]. Neutrons may deposit energy indirectly in matter, and are typically *delayed* by comparison to the energy deposited by electromagnetic processes;

Two important examples of neutron energy deposition processes are:

**Neutron Elastic Scattering** Neutrons from nuclear evaporation of energy in the eV to MeV range may elastically scatter from atomic nuclei with high probability. The smaller the mass of the atomic nuclei, the greater the average fraction of energy lost by the neutron in the material because the average fraction of transferred energy in this scattering process scales with  $(A+1)^{-1}$ . Hydrogenous matter<sup>20</sup> is, therefore, a better neutron moderator than heavier matter [28]. The energy transferred to the nucleus may be re-emitted as a photon by the recoiling atomic nucleus. Energy depositions from this process were observed an exponential distribution with a time constant of 7.7 ns from the initial interaction in a steel calorimeter prototype [32].

**Neutron Capture** Neutrons of energy in the eV range may be captured by an atomic nucleus, which is more frequent than for charged particles because they are not electrostatically repelled. Notably, the binding energy lost by the nucleus that expelled the captured neutron came from is given to another nucleus, which may de-excite through the emission of observable gamma rays (high-energy  $\gamma$ ). This process is much slower due to the time of flight of the neutron undergoing moderation in the matter. Energy depositions from this process were observed to have an exponential distribution with a time constant of 76 ns from the initial interaction in a steel calorimeter prototype [32].

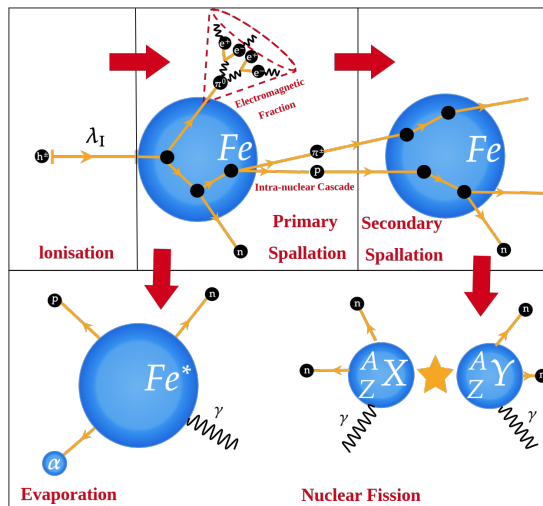
A diagram illustrating the multi-stage energy loss mechanisms of hadrons is shown in Figure 2.16. A plot showing the results of the timing study performed on mixed positive hadron beams of 60 GeV performed in [32] is shown in Figure 2.17.



**Figure 2.15:** Ratio of radiation length to nuclear interaction length as a function of absorber proton number. Note that it is always greater than one. Taken from [28].

19: Due to mass-energy equivalence, the mass of bound nucleons of hadrons is less than their individual masses due to the potential energy (nuclear binding energy) that holds them together. The mass-energy required to hold a nucleus together is known as the '*mass defect*'.

20: Hydrogenous matter is matter containing a large fraction of hydrogen. For example, water is used as a neutron moderator for fission reactors because it is hydrogenous.



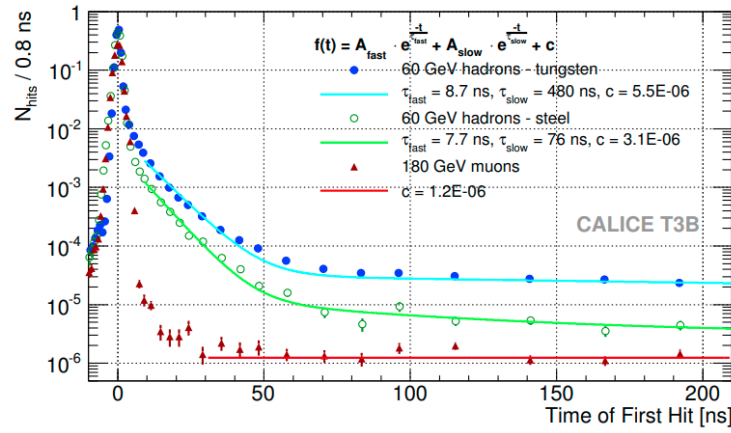
**Figure 2.16:** A diagram illustrating the multi-stage energy loss mechanisms of a charged hadron interacting with Fe nuclei. The red arrows indicate the order of the processes.

### 2.3.6 Hadronic Showers

Similarly to electrons, positrons and photons, hadrons may induce a particle cascade. However, as discussed, the physical processes that affect the hadron shower differ considerably from electromagnetic showers. A hadronic shower consists of energy deposited in two fractions:

**Electromagnetic (EM) Fraction** In the spallation process, pions of all charges are produced in approximately equal numbers on average. However, the neutral pion,  $\pi^0$  decays nearly instantaneously and almost certainly into two photons ( $\pi^0 \rightarrow \gamma\gamma$ ), while the decay of  $\pi^\pm$  to  $e^\pm$  is suppressed<sup>21</sup>, meaning they have a much longer lifetime and are therefore much more likely to go on to induce further nuclear reactions. Photons may only interact with other matter via the processes shown in Figure 2.12b. The  $\eta$  meson may also be produced in spallation, and typically decays by processes that include photons and  $\pi^0$  mesons (see Table 2.2). If above the critical energy, these photons initiate an electromagnetic shower (i.e. when a  $\pi^0$  is produced, none of its energy may be used to induce further nuclear reactions). The proportion of the hadron’s energy deposited in this way is called the ‘*electromagnetic fraction*’.

21: The energetically favourable weak decay of  $\pi^\pm \rightarrow e^\pm + \nu_e(\bar{\nu}_e)$  is heavily suppressed due to a concept called ‘*helicity conservation*’. Helicity is the particle’s spin, projected onto the direction of its motion. ‘Right-handed’ means that the direction of spin is the same as the direction of its motion; ‘left-handed’ means it is the opposite. Experimentally, antineutrinos always spin ‘right-handed’, while electrons prefer to spin ‘left-handed’, with ‘right-handed’ helicity suppressed by a factor of Lorentz  $\beta$ . However, by angular-momentum conservation, this decay requires the electron to be ‘right-handed’ and is therefore suppressed. The muon decay  $\pi^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu}_\mu)$  is preferential due to the much higher mass of the muon permitting a higher likelihood of its ‘right-handedness’ [33]. Therefore,  $\pi^\pm$  decays generally do not produce electrons required for an electromagnetic shower to initiate and have a longer lifetime than  $\pi^0$ .



**Figure 2.17:** Time of first hit distribution of muon data with steel absorbers and hadron data with steel and tungsten absorbers in a time range of -10 ns to 200 ns. Taken from [32].

The fraction of energy deposited in this way depends upon the production of  $\pi^0$  and  $\eta$  mesons. As the energy of an incident particle increases, so does the number of particles in the intra-nuclear cascade increase on average (due to more generations of nuclear spallations being energetically possible). Since the  $\pi^0$  and  $\eta$  mesons are produced in greater numbers and cannot go on to induce further nuclear reactions, the electromagnetic fraction, therefore, increases with incident particle energy. Phenomenologically, the average EM fraction scales as a power law with particle energy:

$$\langle f_{\text{em}} \rangle = 1 - \left( \frac{E}{E_0} \right)^{k-1} \quad (2.6)$$

where  $E_0$  is an absorber-dependent constant related to the average multiplicity in hadronic interactions (0.7-1.3 GeV in Cu up to Pb) for  $\pi^\pm$  showers, where  $k \simeq 0.82$  [28].

**Hadronic (HAD) Fraction** The remaining energy that is not deposited by the electromagnetic fraction is called the *hadronic fraction*, which is deposited predominantly via nuclear interactions. This fraction contains the highly stochastic proportion of ‘invisible energy’.

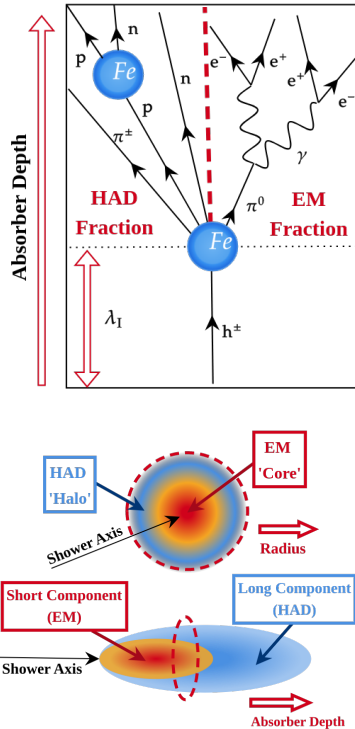
**Development of Hadron Showers** The development of a hadron shower is briefly mentioned, as the following concepts are used in this thesis.

A hadron shower develops, on average, around a ‘shower axis’, which depends on the axis of incidence of the particle onto the matter it traverses. This parameter is defined as the axis of motion of the impinging hadron on the matter. The probability of the shower initiating in matter depends follows an exponential distribution characterised by the nuclear interaction length [28].

An average hadron shower develops laterally with an energy-dense ‘core’, and an energy-sparse ‘halo’, which comprises most of the EM and HAD fraction, respectively. Hadron showers are also expected to develop

Process	$\Gamma_{\text{E}}^{\text{H}}$ (%)	$\tau_{\text{mother}}$ (ns)
$\pi^+ \rightarrow \mu^+ \nu_{\mu}$	99.98770	26.0
$/\pi^- \rightarrow \mu^- \bar{\nu}_{\mu}$	$\pm 0.00004$	$\pm 0.005$
$\pi^0 \rightarrow \gamma\gamma$	98.823	$8.52 \times 10^{-8}$
	$\pm 0.034$	$\pm 1.8 \times 10^{-9}$
$\eta \rightarrow 2\gamma$	$39.41 \pm 0.20$	$5.0 \times 10^{-9}$
$\eta \rightarrow 3\pi^0$	$32.68 \pm 0.23$	$\pm 3.0 \times 10^{-10}$
$\eta \rightarrow \pi^+ \pi^- \pi^0$	$22.92 \pm 0.28$	
$\eta \rightarrow \pi^+ \pi^- \gamma$	$4.22 \pm 0.08$	
$\eta \rightarrow \pi^0 2\gamma$	$2.56 \pm 0.22$	

**Table 2.2:** Highest probability decays involving  $\pi^\pm$ ,  $\pi^0$  and  $\eta$  mesons, their branching ratios and mean lifetimes of the mother particle in ns. Values obtained from [21].



**Figure 2.18:** Illustrations of the development of a hadronic shower in Fe. Top: development of a hadron shower in terms of processes. The left side shows the development of the hadronic fraction; the right side shows the electromagnetic fraction. The interaction vertex is replaced with a Fe nucleus where the multiple vertices of nuclear cascades are expected. Bottom: Illustration of the average development of a hadron shower, as a function of radius and absorber depth. The colour axis red to blue indicates high to low energy density.

22: It is well-known from the famous ‘double-slit experiment’ that particles can behave like waves [35]. It is relevant because this wavelength dictates the interaction scale of a massive particle.

23: A Fermi Gas is a state of matter composed of non-interacting, quasi-free fermions in a potential well. Unlike a Bose gas, which is composed of bosons, the Pauli Exclusion Principle prevents the occupation of energy levels by fermions with identical quantum numbers [39].

isotropically as a function of azimuth angle to the shower axis, on average, due to multiple scattering. However, individual showers often develop asymmetrically on an event-by-event basis. An average hadron shower develops longitudinally with a ‘short’ EM component and a ‘long’ HAD component. The lateral and longitudinal development of hadron showers is uncorrelated, on average. A charged hadron will ionise the detector medium before a hadron shower initiates, whereas a neutral shower will not.

A study on the average hadron shower spatial energy density in a highly granular calorimeter is available in [34]. A diagram illustrating the development of a hadron shower, by contrast to Figure 2.14, is shown in Figure 2.18.

### 2.3.7 Simulations of Hadron Showers

Hadrons interacting with matter involve significantly more physical processes than electrons or photons, owing to the possibility of strong interactions. Moreover, the interaction scale changes with particle momentum, as expected from the de Broglie wavelength relationship  $\lambda = \hbar/p^{22}$ . Simulations of hadrons interacting with matter must consider this fact.

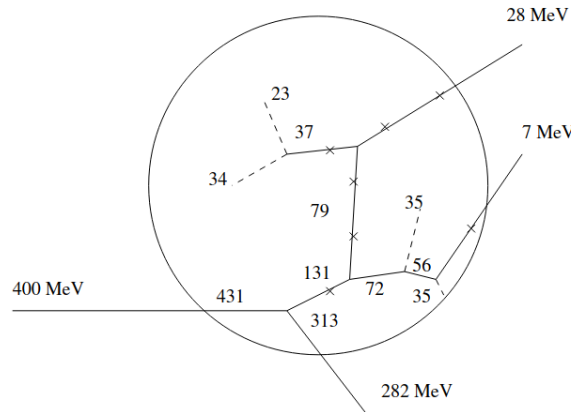
For this reason, hadron shower simulations require several different regimes depending on the energy of particles in the hadron shower and corresponding physical processes. The components of the model used in the hadron shower simulations of this thesis relevant to GeV-scale hadron interactions, in addition to the simpler to describe electromagnetic processes (e.g. ionisation, bremsstrahlung, pair-production, Compton scattering, which are included as standard in standard hadron shower simulation physics lists [36]) are therefore summarised in brief.

**Bertini Cascade Model** The Bertini Cascade Model is a model of intra-nuclear cascades, as previously defined. This model is applicable in the typical energy range  $200 \text{ MeV} < E_{\text{particle}} < 3 \text{ GeV}$  [38]. In this range, hadrons can be treated as colliding with a ‘Fermi gas’<sup>23</sup> of nucleons in the nucleus. A cascade initiates due to the incident particle interacting with the nucleons of a target nucleus, either elastically or inelastically, and producing secondary particles. These particles can produce further secondaries until one of two things happens: the secondaries are absorbed by the nucleus, which results in delayed nuclear evaporation or fragmentation, or they escape the nucleus.

At around 200 MeV, de Broglie wavelength of the interacting hadrons are of the same order as the mean free path in dense matter. Below this limit, atomic nuclei are more likely to experience excitation and evaporate. Above the 3 GeV limit, several effects mean the cascade model is no longer valid as the Lorentz contraction of the nucleus and increasingly forward scattering angles, the ‘nucleon gas’ model is no longer appropriate. This limit nonetheless can and has been extended significantly beyond the 3 GeV limit to a maximum of around 10 GeV.

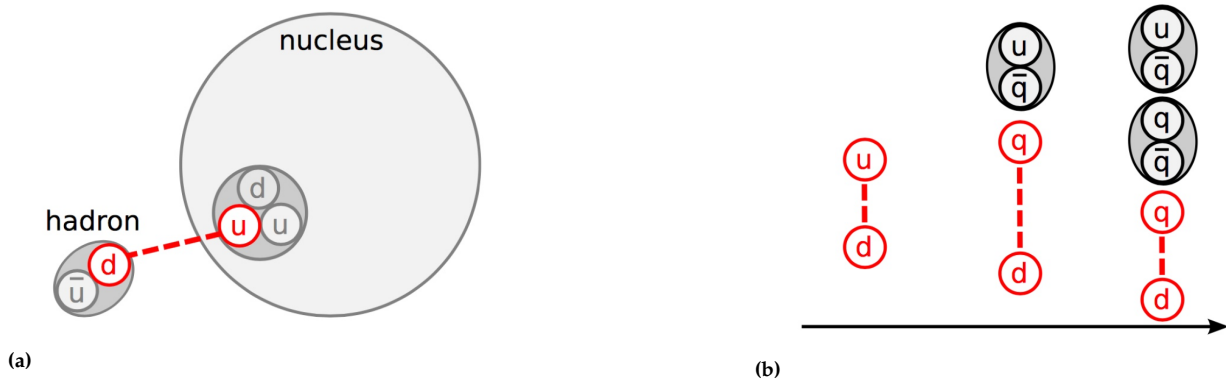
The Bertini Cascade Model has been implemented in the Geant4 simulation package and consists of five main steps [37]. First, the incident hadron’s position on the nucleus’s surface is uniformly selected. Then,





**Figure 2.19:** Diagram illustrating the Bertini Cascade Model. The energies of an incident 400 MeV nucleon and escaping secondaries are shown as solid lines exiting the bounds of the circle, which is the nucleus. Crosses indicate where interactions could have taken place but did not because of the Pauli Exclusion Principle. The lines leading nowhere indicate particles that exited the nucleus. Taken from [37].

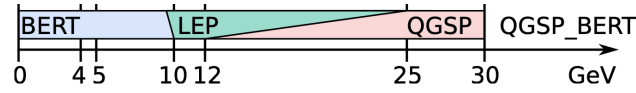
the free particle-particle and region-dependent densities are used to select a path for that hadron. The properties of the products of the next interaction, such as momentum and particle type, are found. The process then proceeds for secondaries so long as the Pauli Exclusion Principle is satisfied and  $E_{\text{particle}} > 2 \text{ MeV}$ . A diagram illustrating an intra-nuclear cascade proceeding according to the Bertini Cascade Model is shown in Figure 2.19.



**Figure 2.20:** Diagram illustrating the Quark-Gluon String Model. Figure 2.20a illustrates the string formation process. Figure 2.20b illustrates the hadronisation process by successive formation of mesons and new strings.

**Quark-Gluon String Model** As previously mentioned, at energies over 10 GeV, the cascade model is no longer an appropriate description of interactions of hadrons with nuclei. At momenta above 20 GeV, the de Broglie wavelength of the hadron probes the partonic substructure of the interacting particles. String-parton models are frequently used to describe hadron interactions at this momentum scale. One such model is the Quark-Gluon String Model, also implemented in Geant4 [40].

This model calculates the hadron-nucleon collision probabilities, determining the number of participating nucleons in the interaction. Then, a ‘string’ is formed corresponding to the parton densities and coupling pairs of quarks by colour. The interaction is treated as a form of pomeron



**Figure 2.21:** The energy ranges of the Bertini Cascade Model (BERT), Low Energy Parameterisation Model (LEP) and Quark-Gluon String precompound Model (QGSP) energy regions. Modified from [43].

24: Pomerons are a mathematical construct used in 'Regge Theory', which is used in studying processes involving high-energy scattering amplitudes, such as those taking place during a hadron shower. Pomerons in this context can be understood as a colourless, composite gluon object, used in the theory as a mediator of the strong force (and modelled as a pair of colour triplet strings in Geant4) [40, 41].

exchange<sup>24</sup>. The string is defined according to the four-momenta of its constituent quarks. Hadrons are formed by successive splitting into a hadron and a new string until the energy falls below a certain threshold. In summary, this process 'fragments' the nucleons of the target nucleus, producing secondary particles and treated as intra-nuclear cascade particles as required. Figure 2.20 illustrates this model's string forming and fragmentation process.

**Empirical Parameterisation Models** In transition energy ranges where neither of the former two models are appropriate descriptions of nature, empirical models are used to describe which hadrons will be produced [42].

**QGSP\_BERT\_HP** The QGSP\_BERT\_HP physics list is used throughout this thesis to generate simulated hadron shower events. In this model, three parameterisations are used: the Bertini Cascade (BERT) up to 10 GeV, an empirical parameterisation called the 'Low Energy Parameterisation' (LEP) from 10-25 GeV, and the Quark-Gluon String pre-compound Model (QGSP) onward in energy. The choice of model is randomly selected in the transition regions, which are demonstrated in Figure 2.21, such that smooth transitions between models is achieved [44]. The HP stands for 'high-precision', which is used to improve the accuracy of modelling neutrons with less than 20 MeV energy and is critical for the studies performed in this thesis due to assessing the potential importance of timing information to Particle Flow clustering algorithms and the contribution of neutrons to indirect energy depositions that are correlated with the proportion of energy carried in the HAD fraction (see Section 2.5.1.5).

## 2.4 Calorimeters

'Calorimetry' means 'measurement of heat'. It was first used in thermodynamics, the branch of physics that relates energy, heat, temperature and work [28]. In contemporary physics, calorimetry generally refers to 'measurement of energy'. In experimental particle physics, calorimetry refers to using devices to measure the energy of final-state particles. These devices are called 'calorimeters'. A myriad of designs exists for calorimeters. Calorimeters may be broadly described as an instrument composed of a volume of matter with which particles may interact, with sensors to measure the products of interactions with that matter.

Calorimeters play a critical role in reconstructing events at all collider experiments. There are typically two main classes of calorimeters used in general collider detector apparatus:



- ▶ *Electromagnetic Calorimeters (ECALs)*: designed to measure the energy of electrons, positrons and photons ( $e^\pm, \gamma$ );
- ▶ *Hadronic Calorimeters (HCALs)* designed to measure the energy of hadrons ( $p, n, \pi^\pm, K_L^0, K^\pm$  etc.)

Some common particles do not typically deposit their energy in calorimeters and are measured with different detectors, or not at all<sup>25</sup>. However, most final-state particles produced in high-energy particle collisions will interact with these detectors, providing a measurable signal corresponding to the particle energy.

25: Two examples are (anti)muons ( $\mu^\pm$ ), which do not deposit energy through any other means than ionisation of the detector medium up to momenta of around 1 TeV. The momentum of these particles is therefore typically measured with special instrumentation called '*muon spectrometers*'. Another example is (anti)neutrinos ( $\bar{\nu}, \nu$ ), which are extremely unlikely to interact at all with any detector apparatus.

### 2.4.1 Design of Calorimeters

The basic design of calorimetric devices for high-energy particles is reviewed in this section. In particular, attention is drawn to the benefits and limitations of sampling calorimeters, around which this thesis is based.

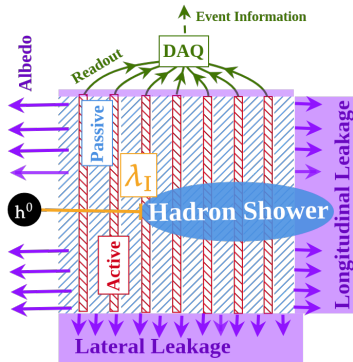
Calorimeters come in two main designs:

**Homogenous Calorimeters** Homogenous calorimeters consist entirely of '*active material*'. The active material is material which is used to induce a measurable signal. Typical examples of such material are scintillating crystals (e.g. lead tungstate,  $\text{PbWO}_4$ ) or certain types of glass (e.g. lead glass) An example of a homogenous calorimeter was the OPAL barrel/endcap calorimeter at LEP, consisting of  $\sim 10,500$   $10 \times 10 \times 37 \text{ cm}^3$  blocks of lead glass [45].

**Sampling Calorimeters** Sampling calorimeters are interleaved layers of '*passive material*' and active material. Passive material is required to induce interactions with the incident hadron, producing secondary particles which can be measured in the active material. The passive material is typically composed of dense matter (e.g. Fe, Cu, Pb, U, alloys like steel), so the hadron deposits as much energy as possible in the matter.

Active materials typically include a '*scintillator*' which produces a photon signal Scintillators are florescent materials that produce a measurable signal from the secondary particles. The active material in calorimeters can be made from a wide range of materials (e.g. plastic, organic crystals/liquids, and noble elements such as Ar). Fluorescence in these materials is achieved through processes such as the photoelectric effect. Scintillators may also include wavelength-shifting material to produce a frequency suited to read-out. However, the signal from scintillation light is typically too weak to read out without amplification. Devices called *photomultipliers* are employed to convert photons from energy-depositing processes into a macroscopic signal. A design of photomultiplier used in modern calorimeters is discussed in depth in Section 2.5.2. The active material in sampling calorimeters may often also be designed with special considerations in mind, such as the use of hydrogenous matter (e.g. plastic) as a neutron moderator (see Section 2.3.6).

26:  $\eta$  (pseudorapidity, defined  $\eta = \ln\left(\tan\left(\frac{\theta}{2}\right)\right)$ , where  $\theta$  is polar angle) and  $\phi$  (azimuth angle), both defined concerning the beam axis, is a coordinate system approximately invariant to longitudinal Lorentz-boosts invariant co-ordinates used in collider experiments to reconstruct the four-momenta of particles produced in collider events. The approximation is valid for particles as long as Lorentz  $\beta \approx 1$ .



**Figure 2.22:** Illustration of the typical design of a hadron sampling calorimeter. The calorimeter consists of interleaved layers of passive (shaded blue) material and active (shaded red) material, from which the calorimeter signal is read out from photomultipliers by a ‘data acquisition’ (DAQ) hardware-software interface that produces event information for further processing. The purple regions indicate leakage may occur for sufficiently deep or broad hadron showers. Purple arrows indicate albedo (‘backscattering’), which may occur in the first few layers of the calorimeter.

The active material of sampling calorimeters may also be sub-divided to improve sensitivity to the spatial distribution of energy inside a calorimeter (‘segmentation’ or ‘granularity’), which is particularly important for the spatial resolution of energy deposited by independent particles and is essential for accurate reconstruction of collision events. Sampling calorimeters are also often much cheaper to construct than homogenous calorimeters and are therefore used ubiquitously as detectors for collider experiments. In addition, the amount of passive and active material can be modified in a sampling calorimeter. All HCALs are therefore sampling calorimeters because their response to hadrons requires ‘compensation’ (see Section 2.4.3).

One example of an HCAL is the CMS-HB+/- barrel HCAL, a Fe/CuZn-Sc calorimeter. [46]. The HB+/- absorber consists of a 40 mm-thick front steel plate, 8×50.5 mm, and 6 × 56.5 mm thick brass plates in order of outwards placement, with around 70,000 individual 3.7/9 mm-thick scintillator tiles. The detector apparatus has a total thickness of around  $5.8 \lambda_I$  at  $90^\circ$ , and a granularity corresponding to  $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$  chosen to balance the complexity of adding additional readout channels with spatial resolution<sup>26</sup>, though this varies with pseudorapidity.

Another example of an HCAL is the ATLAS Fe-Sc TileCal barrel HCAL [47, 48]. The TileCal is composed of alternating layers of 14.1 mm-thick iron plates interleaved with 3 mm-thick plastic scintillator, read out to photo-multiplier tubes (PMTs) by wavelength-shifting fibres. In total, TileCal has approximately 5000 cells and 10,000 readout channels. The total detector extends for around  $7.4 \lambda_I$  at  $90^\circ$ , and has a granularity corresponding to  $(\Delta\eta, \Delta\phi) = (0.1, 0.1)$  and  $(\Delta\eta, \Delta\phi) = (0.1, 0.2)$  in the 3 outermost longitudinal layers.

## 2.4.2 Energy Resolution of Sampling Calorimeters

Energy resolution describes how much the calorimeter signal fluctuates when measuring a particle of known energy. The energy resolution is the most crucial figure of merit for a calorimeter. The energy resolution is typically measured as a unitless quantity, defined as the ratio of the spread of the measured energy to the known particle energy, or, more frequently, the mean reconstructed energy by the calorimeter ( $\frac{\sigma_E}{E}$ ). When referring to calorimeters, the measured energy is also referred to as ‘calorimeter response’ to the particle, which can differ depending on the type of particle.

The main causes for imperfect resolution for sampling calorimeters are summarised:

**Sampling Fluctuations** Since only a portion of the material is active in a sampling calorimeter, only a fraction of the total energy deposited is measured by the calorimeter. This is called the ‘sampling fraction’, and is given by Equation 2.7:

$$f_{\text{samp}} = \frac{E_{\text{MIP}}^{\text{active}}}{E_{\text{MIP}}^{\text{active}} + E_{\text{MIP}}^{\text{passive}}} \quad (2.7)$$

Where  $E_{\text{MIP}}^{\text{active}}$  and  $E_{\text{MIP}}^{\text{passive}}$  is the total energy deposited by a MIP in the active and passive material and respectively because many different energy depositing processes occur with stochastically varying numbers of particles and fluctuating energies deposited by each of those particles, all sampling calorimeters suffer from ‘*sampling fluctuations*’. The sampling fluctuations depend on the sampling fraction, the number of particles contributing to the measurement, and the number of sensors in the active region of the calorimeter. The number of particles in a shower is expected to be Poisson distributed. The number of particles produced is directly proportional to the energy of the event, and therefore the resolution scales as  $\frac{\sigma}{E} \propto E_{\text{particle}}^{-\frac{1}{2}}$ . Sampling fluctuations correspond to the upper limit on the resolution for ECALs [28].

**Stochastic Fluctuations** HCALs, by contrast to ECALs, suffer an additional penalty to resolution in the form of the fraction of invisible energy contributing to each measurement. On average, the proportion of invisible energy in a hadron shower is around 30–40 % of the HAD fraction of shower energy. However, this varies and can be up to 60 % [28]. The proportion of observable energy from each sensitivity is denoted  $e$  and  $h$  respectively, where  $\frac{e}{h} > 1$  in general. Ways to account for this source of uncertainty are discussed in Section 2.4.3. Stochastic fluctuations, as with sampling fluctuations, scales as  $\frac{\sigma}{E} \propto E_{\text{particle}}^{-\frac{1}{2}}$ , for the same reasons. Stochastic fluctuations are the main source of uncertainty in measuring energy for HCALs.

**Detector Calibration Quality** The calibration quality of the readout of individual calorimeter detectors may be sub-optimal. For instance, the inter-calibration of the readout channels contributes to fluctuations in response. This is because hadron showers develop differently in the calorimeter volume from event to event. Different calibrations of the individual sensors therefore result in different responses to hadrons of the same energy. This contribution to resolution is expected to be directly proportional to particle energy. Detector calibration is a topic of this thesis and is discussed in the context of this thesis in Section 2.5.2.

**Electronics Noise** Electronics noise may also affect resolution. For instance, the photomultiplier readout measures signals in a certain window (‘*gate length*’). Without a discharge, the measured charges fluctuate (see Section 2.5.2 for more details). The fluctuations are of a certain, fixed energy. The resolution, therefore, scales as the reciprocal of the energy of the particle  $\frac{\sigma}{E} \propto E_{\text{particle}}^{-1}$ .

**Leakage** Calorimeters are designed typically of a depth and breadth sufficient to contain a hadron shower, which depends on the nuclear interaction length of the absorber. However, in some instances, energy may escape the calorimeter either laterally or, more frequently, longitudinally, meaning it cannot be measured. Additionally, energy can back-scatter from the first few layers of passive material the calorimeter, which constitutes a form of leakage (‘*albedo*’). As more generations of particles are produced in higher energy showers, they experience higher proportions

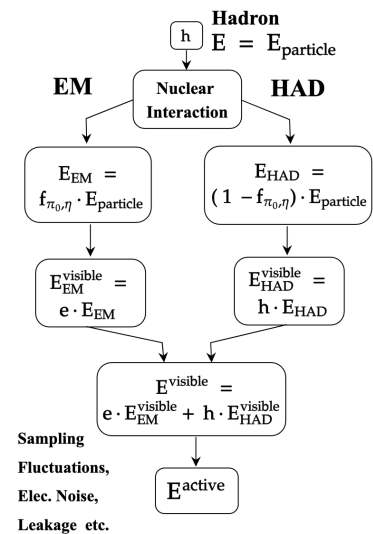


Figure 2.23: Flow chart indicating the relationship between hadron energy and the energy measured by a hadron calorimeter.

of energy loss in this way, called 'leakage'. Special calorimeters called 'tail catchers' are used to tag longitudinal leakage.

**Calorimeter Resolution** Assuming negligible electronics noise and leakage, calorimeter resolution may be described in terms of Equation 2.8:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E_{\text{particle}}}} \oplus b \quad (2.8)$$

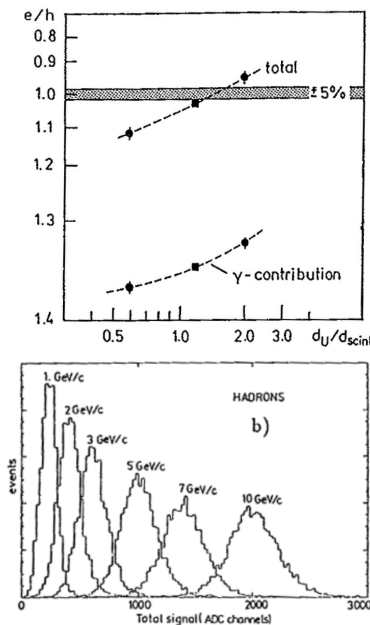
Where  $a$  describes the contribution of combined stochastic and sampling fluctuations,  $b$  describes the quality of detector calibration, and  $\oplus$  indicates addition in quadrature.

### 2.4.3 Improving Hadronic Energy Resolution

As a result of invisible energy loss in hadron showers, the EM response and HAD response tend not to be intrinsically equal in a sampling hadron calorimeter (i.e. typically,  $\frac{e}{h} > 1$  for hadron calorimeters). However, there are several ways in which the invisible energy of the hadronic fraction may be accounted for, which are discussed in brief:

**Compensating Calorimeters** A 'compensating' calorimeter is designed to equalise  $e$  and  $h$  by careful design. A compensating calorimeter requires several key design features [28, 31]:

- ▶ *A sampling calorimeter design:* compensation cannot be achieved in a homogeneous calorimeter because the response to  $e$  and  $h$  cannot be tuned (i.e. it is composed entirely of the same active material, which cannot be modified);
- ▶ *Hydrogenous active material:* the response to neutrons by the calorimeter can be used to enhance  $h$ . Hydrogenous active materials such as plastic are, therefore, frequently used in compensating calorimeters to enhance the HAD sampling fraction;
- ▶ *A high-Z passive material:* compensation is achieved by hardware by attenuating  $e$  and enhancing  $h$ . A high-Z passive material (relative to active material) reduces the sampling fraction for electrons and photons because the cross-sections for the processes shown in Figure 2.12 scale with powers of the atomic number of the absorber<sup>27</sup>. This property means more EM energy is deposited in the passive (and therefore less in the active) material, thereby attenuating the EM sampling fraction;
- ▶ *Tuned sampling fractions:* the sampling calorimeter requires careful tuning of the thickness ratio of passive and active material to balance attenuation of the EM sampling fraction from the high-Z absorber and the enhancement from the hydrogenous active material such that  $\frac{e}{h} = 1$  is achieved.



**Figure 2.24:** Top: The  $\frac{e}{h}$  of the ZEUS calorimeter as a function of the thickness ratio of U passive material to active plastic Sc material. Bottom: ZEUS Forward Calorimeter (FCAL) response to hadrons of different momenta. Both taken from [31].

27: Specifically, some example EM process cross-sections are proportional to atomic number [31]:

- ionisation  $\propto Z$
- pair production  $\propto Z^2$
- bremsstrahlung  $\propto Z^2$
- photoelectric effect  $\propto Z^4 - Z^5$
- Compton scattering  $\propto Z$

An example of a compensating calorimeter was the U-Sc ZEUS calorimeter, originally located at DESY, Hamburg, which achieved the best hadronic energy resolution of any calorimeter in history thus far ( $a \approx 35\%/\sqrt{E_{\text{particle}}}$ ) [31]. This calorimeter was designed to exploit uranium's high- $Z$  and naturally fissile properties to produce a greater fraction of neutrons and attenuate the EM signal and used a doped polystyrene-based plastic scintillator (SCSN-38) to increase  $h$ . The thickness of the absorber of active and passive materials was chosen carefully to achieve compensation (see Figure 2.24).

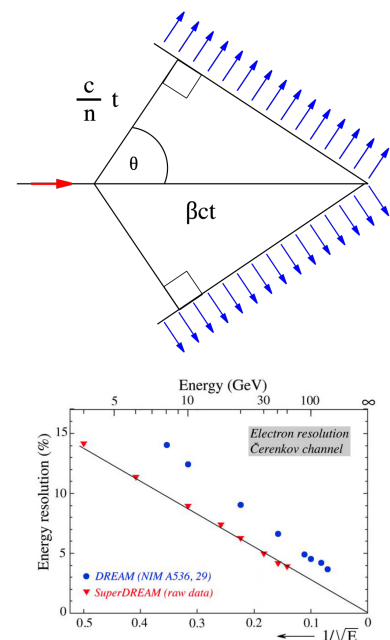
**Dual-Readout Calorimeters** When a relativistic charged particle traverses through dielectric<sup>28</sup> material, in which its refractive index is greater than that of vacuum ( $n > 1$ ), it emits Cerenkov radiation. This phenomenon occurs due to polarisation and subsequent de-excitation of the traversed dielectric matter by the charged particle. This phenomenon is analogous to the 'sonic boom' that occurs when an object surpasses the speed of sound in air.

Dual-readout detectors exploit this phenomenon to measure the EM fraction of a shower on an event-by-event basis [52]. This is achieved by simultaneously measuring energy in two ways: by Cerenkov light produced by charged particles and separately by scintillator light. The average responses from both readout methods are calibrated to be equal for electrons. However, when measuring a hadron shower, the Cerenkov signal is only sensitive to the EM fraction. At the same time, the scintillator is sensitive to signals from both the EM and HAD fractions of the event. If the HAD fraction is non-zero, then the scintillator and Cerenkov signals will differ. By exploiting the ratio of the two readout signals and their respective  $\frac{c}{h}$  ratios are independent of the hadron shower energy, the EM fraction of the event can be reconstructed. Then, from this information, the equivalent signal that would have been produced if the HAD fraction had been zero (i.e. the response expected from an EM shower produced by an electron of the same energy as the interacting hadron) can be reconstructed.

A contemporary example of this detector design is the RD50 SuperDREAM Cu-Fibre Calorimeter. This calorimeter prototype consists of layers of Cu containing  $2 \times 20,148$  Sc-Cerenkov fibres, read out by 8 PMTs [50].

**Software Compensation** If the design requirements of the calorimeter prohibit a compensating design (e.g. the thicknesses of the passive and active material cannot be modified within the constraints placed on the instrumentation), then the event information can be used to compensate the hadron shower. In particular, the energy density (e.g. the dense EM 'core' vs the sparse HAD 'halo') and temporal development of a hadron shower (e.g. delayed energy depositions caused by neutron interactions) may be used to infer information about  $e$  and  $h$  on an event-by-event basis. Software compensation is one of the topics of this thesis and is discussed in depth in Chapter 5.

28: Dielectric material is material that can be polarised in an electric field.



**Figure 2.25:** Top: illustration of the 'shock cone' of Cerenkov radiation produced by a relativistic charged particle entering from a medium with a lower refractive index to a medium with a higher refractive index. Here,  $n$  is the material's refractive index,  $c$  is the speed of light,  $t$  is time, and  $\beta c$  is the particle's velocity. Taken from [49] Bottom: electron resolution of the SuperDREAM and DREAM Dual-Readout calorimeter prototype [50, 51]. The fit shown yielded a resolution of  $13.9\%/\sqrt{E} \oplus 0.5\%$  in response to electrons. Hadron resolution is not available.

### 2.4.4 Linearity of Response

In an ideal case, a detector should produce a linear response. This means that the signal produced by the calorimeter increases linearly with the energy of the incident particle (i.e. the signal is proportional to the particle energy [28]). Several causes of calorimeter non-linearity are discussed in brief:

**Ionisation Quenching** The light yield for scintillators reduces as the energy loss of a heavy charged particle increases due to an increased likelihood that ionised electrons and ions recombine into atoms once again. This results in a non-linear light yield that reduces the energy loss of the charged particle [28]. Birk's Law gives the formula describing the light yield of a particle traversing a scintillator, shown in Equation 2.9 [53]:

$$\frac{dL}{dx} = S \frac{\frac{dE}{dx}}{1 + kB \frac{dE}{dx}} \quad (2.9)$$

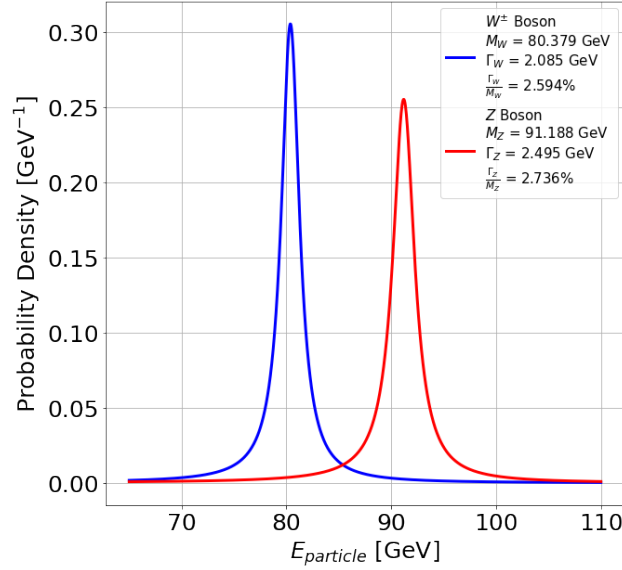
where  $L$  is light yield,  $S$  is scintillator efficiency,  $\frac{dE}{dx}$  is the energy loss of the ionising particle,  $k$  is the quenching probability and  $B$  is a constant relating local density of ionized molecules along the path of the particle to the specific energy loss. This quantity varies from scintillator to scintillator. Since  $kB$  acts as an overall scaling factor, it is known as 'Birk's constant'.

**Photodetector Saturation** Silicon-based photomultipliers may detect only one photon at once and are thus typically built as arrays of individual cells. Under a certain illumination intensity, the photomultiplier units are all active at once and cannot detect more photons. Therefore, as the intensity of illumination increases on the sensor, its response also becomes non-linear and saturates at a certain value. This nonlinearity can, to some extent, be corrected in the calibration of the detector by applying an inverse mapping of the nonlinear response function, known as a 'saturation curve' [54].

### 2.4.5 Particle Flow

Concerning Section 2.2.2, many physical interactions of scientific interest involving the Higgs will result in final states involving multiple jets of particles. Invariant mass reconstruction from two or more jets is required at future linear colliders for particle identification (PID) and event reconstruction. One typical benchmark for detectors used in future linear colliders is an adequate resolution to enable separation of the  $W^\pm$  and  $Z$  boson invariant masses from their hadronic decays (final-state jets). The best possible resolution of the invariant masses depends on the decay width of these particles and their invariant mass. Figure 2.26 shows the distribution of the energy of the  $W^\pm$  and  $Z$  bosons and illustrates that an invariant mass resolution of the same  $\sigma_{\text{particle}}/E_{\text{particle}} \simeq \Gamma_Z/E_Z \simeq \Gamma_{W^\pm}/E_{W^\pm} \simeq 2.7\%$ , which yields a  $3.6\sigma$  separation on the peaks of the Breit-Wigner distribution for the hadronic decays of the  $W$  and  $Z$  boson. For the typical jet energies at ILC, where the typical jet energy,  $E_j$ , is in





**Figure 2.26:** Plot showing the Breit-Wigner distribution for the  $W^\pm$  and  $Z$  boson, where  $M_{\text{particle}}$  is invariant mass and  $\Gamma_{\text{particle}}$  is decay width. The red and blue lines indicate the  $Z$  and  $W^\pm$  bosons respectively, with values taken from [21] and shown in the legend.

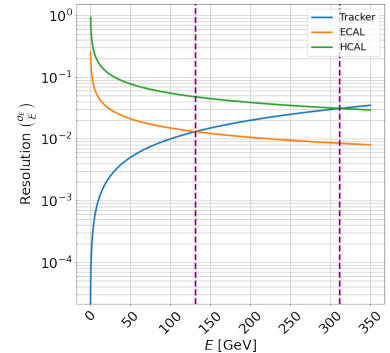
the range 150–350 GeV for  $\sqrt{s}$  in the range 0.5–1 TeV, results in a required jet energy resolution of  $\frac{\sigma_j}{E_j} \lesssim 30\%/\sqrt{E_j}$ , which is around 3% in the range of jet energies from 50–200 GeV [55].

Studies on jets from short-lived particles have shown that they are composed primarily of charged hadrons and leptons, with further contributions from photons neutral hadrons and a small fraction of neutrinos, which cannot be detected (see Table 2.3).

Traditional calorimetry measures the energy of all hadrons with the HCAL. However, because of stochastic fluctuations due to invisible energy as discussed in Section 2.3.6, HCALs always have the worst energy resolution of any detector used in collider experiments. This means most of the jet energy measured with the HCAL will result in sub-optimal jet energy resolutions.

An alternative approach is to exploit that a jet is expected to be composed primarily of charged particles. The momentum of a relativistic particle can be used to approximate the energy of a particle in cases where Lorentz  $\gamma \simeq 1$ . Momentum may be measured by using specialised detectors (*'trackers'*) to reconstruct its path in a strong magnetic field ( $B \sim \mathcal{O}(\text{T})$ ) using conservation of angular momentum and the Lorentz force. The measured track may then be associated with energy depositions in the ECAL or HCAL, and this energy is excluded from the energy depositions in the calorimeters. The energy of the photons and neutral hadrons may then be measured with the ECAL and HCAL, respectively. This choice means that around 60% of the total jet energy can be measured with a resolution many orders of magnitude better than relying on the calorimeters only. If the event's four-momenta of each observable particle are well reconstructed, the *'missing energy'* from neutrinos may also be inferred.

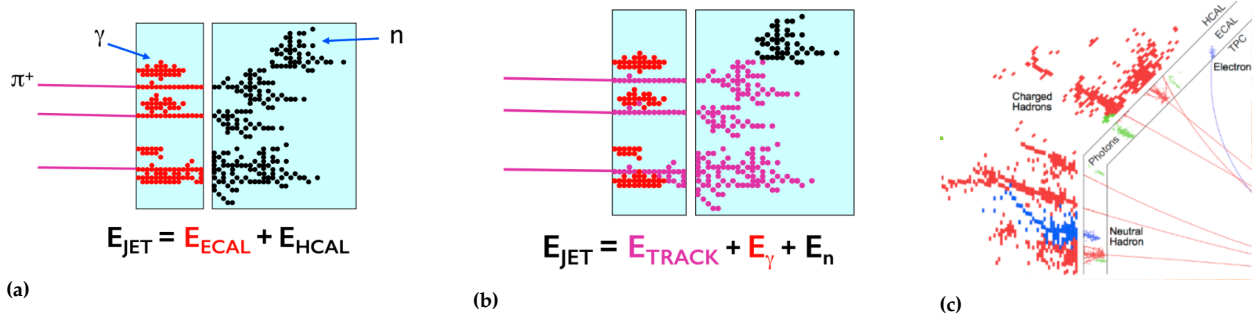
This type of calorimetry is called *'Particle Flow'* (PF). PF was first developed as *'energy flow'* for the ALEPH experiment at LEP 2, which achieved a



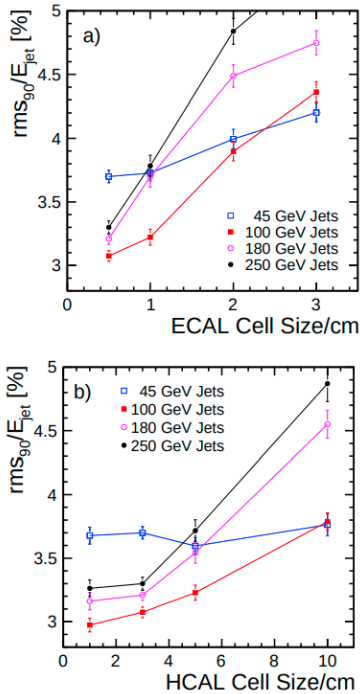
**Figure 2.27:** Comparison of particle resolutions from the Tracker, ECAL and HCAL shown in Table 2.3. The tracker is superior for measuring electrons/positrons of energies below 131 GeV and charged hadrons with energy below around 311 GeV, indicated by the dashed purple lines.

Particle Type	Detector	Energy Fraction	Energy Resolution	Jet Energy Resolution
$h^\pm/l^\pm$	Tracker	$\sim 0.6E_j$	$10^{-4}E_{X^\pm}$	$< 3.6 \times 10^{-5}E_j$
$\gamma$	ECAL	$\sim 0.3E_j$	$0.15/\sqrt{E_\gamma}$	$0.08/\sqrt{E_j}$
$h^0$	HCAL	$\sim 0.1E_j$	$0.55/\sqrt{E_{h^0}}$	$0.17/\sqrt{E_j}$
$\nu_l(\bar{\nu}_l)$	N/A	negligible	N/A	N/A

**Table 2.3:** Table describing Particle Flow. Charged hadrons/leptons ( $h^\pm/l^\pm$ ) are measured using their momentum by tracking their trajectory in a magnetic field, which improves the jet energy resolution compared to only using the calorimeters. The remaining photons ( $\gamma$ ) and neutral hadrons ( $h^0$ ) are measured using the ECAL and HCAL respectively. This means that only around 10% of the jet's total energy is measured using the HCAL for a given jet, thereby improving the jet energy resolution over traditional calorimetry. Taken from [55].



**Figure 2.28:** Figure 2.28a demonstrates the method utilised by traditional calorimetry to measure the energy of jets. Figure 2.28b demonstrates the PF method, by comparison. Figure 2.28c demonstrates the Pandora PFA algorithm, as applied to an ILC jet using the expected PF-based *International Linear Detector* (ILD). Each colour represents a different particle type, reconstructed with the Pandora PFA algorithm. All taken from [15].



**Figure 2.29:** The jet energy resolution as a function of calorimeter sensor size, motivating the existence of highly granular calorimeters. Results obtained for  $Z \rightarrow q\bar{q}$  induced jets. Top: ECAL. Bottom: HCAL. Taken from [57].

jet energy resolution corresponding to  $\sigma_{E_j} = 65\% \sqrt{E_j}$  [56]. However, this is well above the required jet energy resolution of ILC.

PF has since been developed to meet the standards for future linear collider experiments such as ILC and CLIC. Both experiments intend to use PF as the cornerstone of their detector design to achieve state-of-the-art jet energy resolution.

PF requires reconstruction of the to be achieved, several detector design requirements must be adhered to:

- ▶ *High Performance, Hermetic Tracking of Charged Particles:* to make accurate momentum measurements, a full  $4\pi$  tracking system with an excellent resolution is required such that the momenta of all charged particles produced in an inelastic collision are reconstructed accurately [58];
- ▶ *Highly Granular Calorimeters:* the energy depositions from individual particles must be resolvable to allocate energy depositions to charged tracks correctly. This has been found in simulation to be achievable with highly segmented detector arrays, which improve the jet energy resolution for both ECAL and HCAL with higher granularity detectors. To achieve the jet energy resolutions required for ILC, The cell size for ECAL and HCAL transverse granularity ought to be at least  $5 \times 5 \text{ mm}^2$  and  $3 \times 3 \text{ cm}^2$ , respectively. [57];
- ▶ *Energy Clustering Algorithms:* to associate the energy depositions from highly granular calorimeters to tracks and to determine which depositions came from neutral particles and which came from charged particles, sophisticated clustering algorithms are required.



An algorithm designed to perform such as *Pandora Particle Flow Algorithm* (Pandora PFA). Pandora PFA has been demonstrated to be able to achieve resolutions of  $< 3.8\%$  for jet energies in the range 40-420 GeV using simulations of jets observed with the International Linear Detector (ILD) proposed to be used at the ILC [55].

### 2.4.5.1 Confusion

Unlike traditional calorimetry, PF performance is not limited by ECAL/H-CAL resolution performance except at energies below 40 GeV. Instead, the resolution is dominated by the effectiveness of correctly allocating energy between showers.

Therefore, the main difference between traditional calorimetry and PF is that the former requires detectors with good energy resolution, and the latter requires detectors capable of good pattern recognition. The degradation of jet energy resolution in PF due to misallocation of energy is known as ‘*confusion*’. There are three main types of confusion, shown in Figure 2.32:

- ▶ *Photon Resolution*: photonic energy depositions in ECALs may be incorrectly allocated to the charged track, thereby ‘losing’ the energy from the particle to the charged track momentum, shown in Figure 2.32a;
- ▶ *Neutral Resolution*: energy from the neutral particle may be incorrectly allocated to charged shower, thereby ‘losing’ the energy of the hadron shower to the charged track momentum, shown in Figure 2.32b;
- ▶ *failures in Reconstructing Shower Fragments*: Fragments of showers may be reconstructed as separate particles, thereby double counting the energy depositions, shown in Figure 2.32c.

Therefore, sophisticated pattern recognition techniques for energy depositions are critical to PF since its performance dictates the jet energy resolution at the energies of future linear colliders. This fact motivates research into pattern recognition techniques to resolve charged and neutral energy depositions and their benefits and limitations (see Figure 2.30.).

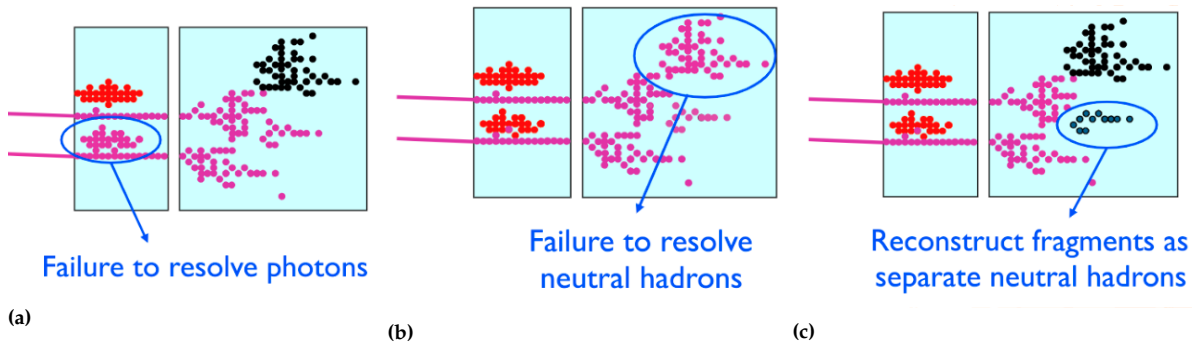


Figure 2.32: Types of confusion experienced by PandoraPFA. All taken from [15]

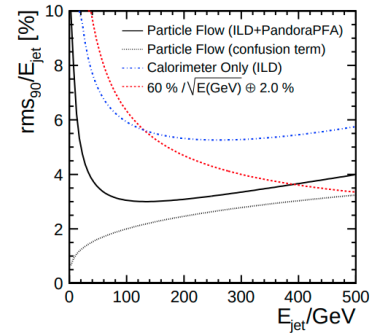


Figure 2.30: The jet energy resolution obtained from PF (using Pandora PFA and a simulation of the International Linear Detector (ILD), the proposed detector for ILC). The dotted line indicates the proportion of confusion. The dot-dashed curve shows jet energy resolution obtained the detector simulation alone. The dashed curve shows the jet energy resolution measured with a traditional calorimetric approach, using the hadron calorimeter. Taken from [55].

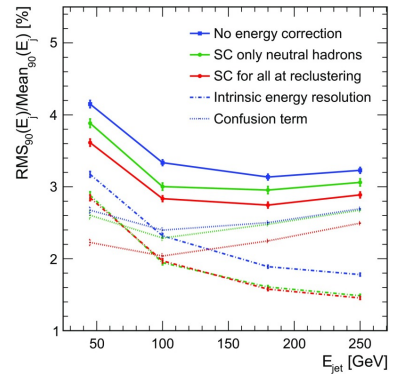


Figure 2.31: Jet energy resolution for the ILD detector for ILC with the minimum optimal HCAL cell size without energy correction (blue), when applying SC for neutral clusters only (green) and when applying SC for all clusters (red). The long-dashed lines show the energy resolution, the dotted lines present the confusion term as in Figure 2.30. Taken from [15].

A different topic discussed in this thesis is software compensation (SC) for a highly granular calorimeter. Highly granular calorimeters will typically not be compensating [55]. As shown in Figure 2.31, this technique has been shown to improve the performance of Pandora PFA, which allows superior measurements of neutral shower energy and association of energy deposits to charged tracks by correcting for the invisible energy that may contribute to a hadron shower event [59]. This observation motivates research into methods to exploit best the rich spatial energy-density information available from highly granular calorimeters for compensation.

It is also noted that no usage of timing information for compensation or shower separation for highly granular calorimeters is a generally new topic. This also motivates research into the possible gains in performance using this additional information.

## 2.5 Apparatus and Experimental Methods

This section introduces and discusses the apparatus and methods used in this thesis.

### 2.5.1 CALICE AHCAL

In this section, the CALICE AHCAL highly-granular calorimeter prototype is introduced, and its design and features are described. Special attention is drawn to the capacity of the detector to measure the timestamp of energy depositions.

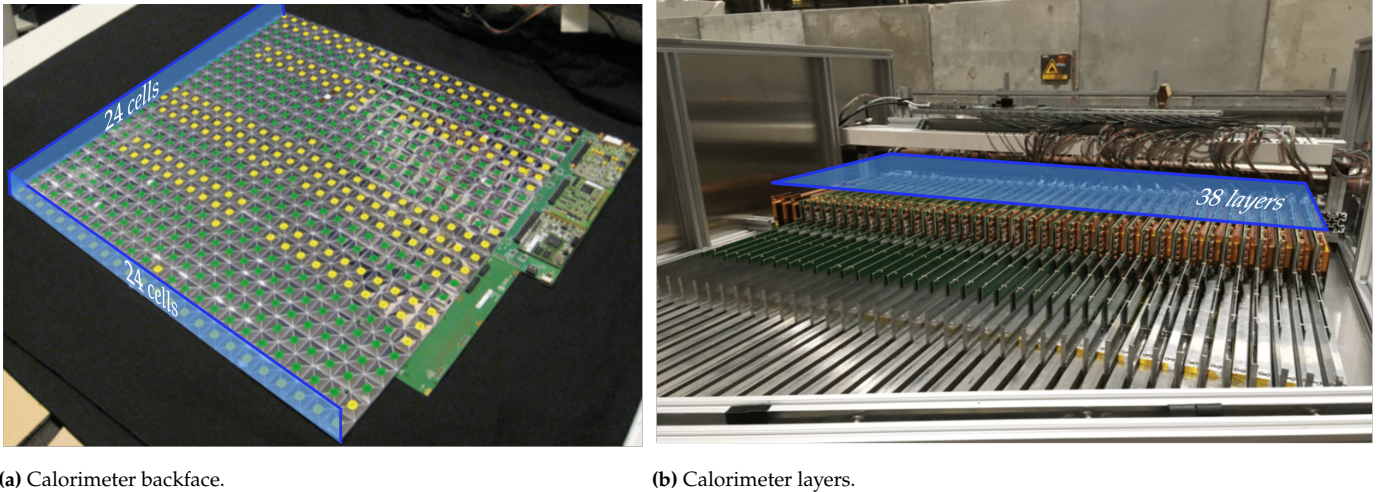
#### 2.5.1.1 Design of the AHCAL Prototype

The *CALICE Analogue Hadronic Calorimeter* (AHCAL) is a Fe-Sc highly granular calorimeter prototype built by the CALICE Collaboration, designed for PF in future linear collider experiments such as ILC or CLIC [60]. It is a sampling hadron calorimeter, consisting of steel plates of the thickness of 16 mm as passive material and plastic scintillator tiles of  $30 \times 30 \times 3 \text{ mm}^3$  volume, which places it at the upper limit on granularity required for effective PF as discussed in Section 2.4.5. The photon signal is read from the scintillator by optical-range photodetectors called ‘*silicon photomultipliers*’, or SiPMs. The physics of the SiPMs used in the AHCAL are introduced separately and in detail in Section 2.5.2.

The detector consists of  $24 \times 24 \times 38$  ( $\sim 22,000$ ) individual SiPM-on-tile readout channels interleaved between layers of steel absorber. The AHCAL has a depth of around  $4.5 \lambda_I$ . The detector is designed to fit inside the diameter of the solenoid of a future linear detector. The AHCAL’s depth is insufficient for an ILC hadron calorimeter optimized for  $\sqrt{s} = 250 \text{ GeV}$ , as it only has fewer than the 48 layers required for  $6 \lambda_I$  longitudinal shower containment. This means that longitudinal leakage affects the performance of the AHCAL [60]. It is also noted that the AHCAL is a non-compensating calorimeter due to constraints on absorber width. A previous prototype for the AHCAL was measured to have an energy-dependent  $\frac{e}{\pi}$  ratio of between 1.25 and 1.4 in simulation

	Technological Prototype (2018)
Readout	SiPM on scintillator tile
Absorber Layers	38
Number Of Sensors	21,888
Absorber thickness	17.2 mm
Layer thickness	26.1 mm
Tile thickness	3 mm
Cell Area	$30 \times 30 \text{ mm}^2$
$X_0$	27.3 mm
$\lambda_I$	237.1 mm
Moliere Radius	24.9 mm
Total Depth	$\sim 4.5 \lambda_I$ $\sim 0.11 \lambda_I/\text{layer}$ $\sim 40.8 X_0$ $\sim 1.05 X_0/\text{layer}$

**Table 2.4:** Table of design components and values of the CALICE AHCAL Prototype. Modified from [34].



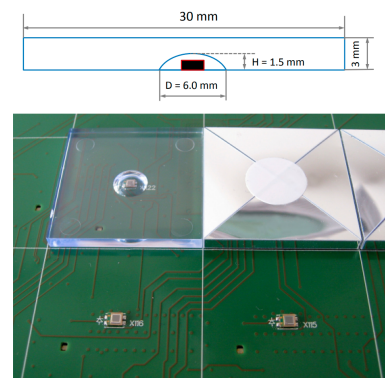
**Figure 2.33:** Pictures of the CALICE AHCAL, depicting the calorimeter at test beam. The cells are shown in Figure 2.33a, wrapped in foil to improve photon sensitivity. A stack of thirty-eight layers is shown in Figure 2.33b.

and experimental data [61]. The high granularity of the calorimeter is driven by the need to resolve the energies of hadrons in jets initiating showers in the detector [60]. Both of these topics are studied in this thesis. Images of the CALICE AHCAL are shown in Figure 2.33

A summary of the important individual components and design features and aspects of the readout of the AHCAL detector are reviewed as follows:

**Scintillator Tiles and SiPM Readout** The scintillator tiles used in AHCAL are composed of a polystyrene-based material doped with POPOP wavelength shifter, thereby giving secondary scintillation corresponding to a peak spectral wavelength in the high optical frequency range (410 nm, violet light) [60]. The tiles have a small hemispherical dimple of 1.5 mm in height and 6 mm in diameter embedded within the scintillator tile, where the SiPM readout is located. This dimple provides optimal placement of the SiPM and focuses the scintillator light. Each tile is individually wrapped in highly reflective foil (3M Enhanced Specular Reflector [62]) to improve the sensitivity of the tile via total internal reflection. The foil is cut and scored for folding by laser and automatically folded and wrapped around the tile using a robotic method [63], which allows scaling to the large number of sensors required for the AHCAL. Furthermore, the scintillator tiles are glued directly to a printed circuit board (PCB) substrate upon which the readout circuitry is also placed (see Figure 2.34. The sensitivity of the SiPM to photons using this design was found to be uniform, which is of adequate yield for a single MIP for most of the scintillator volume (97.1 % of the scintillator volume yielded within 10 % response of the mean 20.6 MIP photoelectron response to  $^{90}\text{Sr}$   $\beta$  electrons [64]).

The AHCAL detector utilises the Hamamatsu S13360-1325PE MPPC as photodetectors for the SiPM-on-tile design. This device has an active area of  $1.3 \times 1.3 \text{ mm}^2$  and has 2668 pixels per device. It is used in the AHCAL due to its sharp response, fast recovery time, its comparatively low rates of noise. During the 2018 testbeam campaign, the SiPMs of the AHCAL



**Figure 2.34:** Top: Design schematic of the plastic scintillator tile used in the AHCAL. The black square indicates the SiPM. Taken from [64]. Bottom: Photograph of the scintillator tile, with and without foil. Taken from [63].



were operated at 5 V over-voltage (for a definition of this quantity, see Section 2.5.2) [65]. A detailed discussion on silicon photomultipliers is provided in Section 2.5.2. Further details and references on the specific manufacturer values of this detector are provided in Section 4.4.2, as part of the research presented in that section.

Each combined ‘SiPM-On-Tile’ read-out unit is hereafter referred to as a ‘calorimeter cell’, and an active cell during an event as a ‘hit’.

**SPIROC2E Readout Chip** The SPIROC2E (SiPM Integrated Read Out Chip) is a dedicated front-end electronics unit used for readout from the SiPMs of the AHCAL, manufactured by Omega [66]. It is an application-specific circuit (ASIC) capable of auto-triggering and can read out from 1 to 2000 p.e. with a timing resolution of 100 ps from 36 independent channels.

An analogue memory array of 16 for each channel allows for storing the time and charge measurement of each SiPM in the readout. The SPIROC2E can operate at under 25  $\mu\text{W}$ , alleviating the need for additional cooling. The chip can operate in two modes: a high-gain and a low-gain mode, which modify the dynamic range of operation. The voltage from the SiPM is shaped and converted to digital information by a 12-bit Wilkinson Analogue-to-Digital Converter (ADC) unit<sup>29</sup>, which allows an analogue readout of the signal from the SiPMs to which it is connected, which may later be converted into calibrated MIP units.

The SPIROC2E Time-to-Digital Converter (TDC) unit allows the digitisation of the timestamp of an event. The TDC operates with two ramped voltages that operate during one clock cycle, one for even bunch crossings and one for odd, with a length given by the bunch crossing length of the ILC design (200 ns), during ‘ILC mode’ and 4000 ns during ‘test beam mode’. The out-of-phase ramped voltages are cycled to reduce the dead time between clock cycles. The voltage from the ramp is stored in a register of 16 memory cells, from which the time is reconstructed. The timestamp is extracted by converting the voltage to a TDC count similarly to the energy signal from the SiPM. The conversion from TDC units to nanoseconds is achieved by calibration of each SPIROC2E chip their channels using an external clock. A linear relationship is expected between the external clock and the measured TDC value. The slope of this line is common to all channels on a chip, while the offset is channel-dependent. Both are extracted for each chip and channel via a fit. The hit time in nanoseconds is then extracted during an event using this slope and offset, relative to the time of a coincidence between two trigger scintillators placed in front of the AHCAL, is used to extract the time in nanoseconds (see [68, 69] for further details on this process). The underlying principle of the calibration of TDC is nonetheless demonstrated in Figure 2.37.

**HCAL Base Unit** The HCAL Base Unit (HBU) is an electronics unit that supports all the components required to output the event information from the layer to the DAQ system of the AHCAL. A single HBU has dimensions  $36 \times 36 \text{ cm}^2$ , and supports four SPIROC2E chips, reading out from 144 individual scintillator-SiPM cells. The HBU has a power unit (POWER), a calibration (CALIB) and a detector interface (DIF), shown in Figure 2.40.

29: A Wilkinson ADC unit operates by comparing a charging capacitor to an input voltage. The capacitor is then discharged, and the time it recovers is proportional to the signal’s amplitude. During discharge, a high-frequency clock is sampled and stored in a register. The number of pulses from the clock during this time is proportional to the voltage, thereby digitising the signal [67].

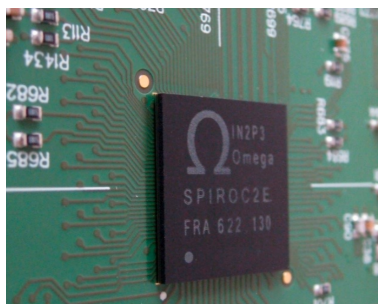


Figure 2.35: Photograph of the SPIROC2E Readout Chip. Taken from [70].

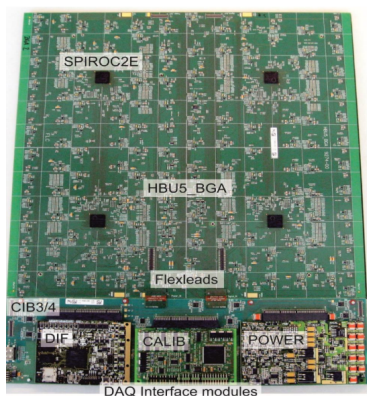


Figure 2.36: Photograph image of the AHCAL HBU. Taken from [70].

The power unit controls the power supplied to the components and may allow the calorimeter system to operate in 'power pulsing mode'. Power-pulsing mode at ILC refers to 'pulsed' bunch trains that arrive at the interaction point of the detector. During an event, data acquisition is turned on; the power is turned off between trains. This alleviates the need for additional cooling and reduces power consumption. The HBU, therefore, exploits the relatively long bunch crossing timescale expected of future linear colliders such as ILC to reduce power draw such that the need for cooling is reduced (see [71]).

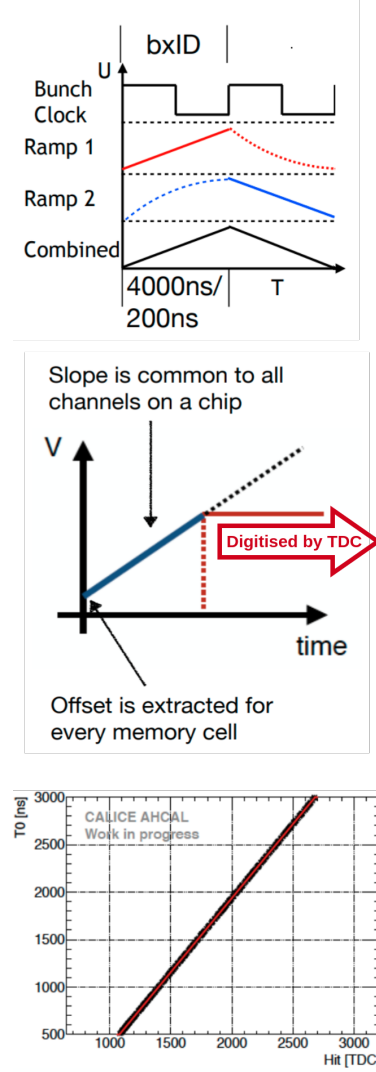
Highly relevant to this thesis is that the HBU contains an internal LED calibration system for the SiPMs. The LED Calibration unit uses UV-LEDs that are placed beneath each Scintillator-SiPM tile. UV-LEDs are used because the SiPMs are required only to be triggered only once and therefore requires fast pulses of  $\sim 10$  ns during calibration, at an intensity that may vary between no SiPM signal and the saturation of the SiPM [72]. The reasons for the need for automated calibration of the SiPM signals are discussed in depth in Section 2.5.2, and are a significant focus of this thesis. The voltages supplied to the LED calibration for each LED are controlled by the calibration unit.

Finally, the detector interface mediates the interface between the aforementioned control units and chips via *field-programmable gate array* (FPGA) units. It provides the output of the detector system to the AHCAL DAQ.

### 2.5.1.2 Event Information

As observed by the AHCAL, a hadron shower event consists of many active cells. An active cell is denoted as a 'hit'. Each 'hit' has a coordinate within the matrix of calorimeter cells. The lateral spatial position of an active cell is defined relative to the front face of the calorimeter is denoted  $I_{\text{hit}}$  and  $J_{\text{hit}}$ , both with values between and including 1 and 24 cell units. The longitudinal spatial position (depth in layers) is denoted  $K_{\text{hit}}$ . These quantities are measured in cell units and defined as values between and including 1 and 38 cell units, corresponding to the calorimeter layers. The energy of an active cell is denoted  $E_{\text{hit}}$ , measured in Analogue-to-Digital counts by the SiPM cells, calibrated to the energy deposited by a minimum ionising particle in one cell (MIP) [73].  $E_{\text{hit}}$  takes a value between a noise threshold at 0.5 MIP and the SiPM saturation energy. The TDC also produces the time of a hit,  $t_{\text{hit}}$  is bounded between the time at which the energy deposited in a given cell crosses a pre-defined threshold, normalised throughout this thesis 0 ns, smeared by the resolution, and the chosen gate length for the measurement of an event. Additionally, the incident position of a charged particle in lateral coordinates may be measured by a tracker before interaction, as would be the case in a Particle Flow reconstruction, which is denoted as a vector  $[I_{\text{track}}, J_{\text{track}}]$ . Furthermore, the number of active cells in the event is denoted  $N_{\text{hits}}$ . The total reconstructed energy in a hadron shower observed by AHCAL is denoted  $E_{\text{sum}}$  and described by Equation 2.10.

$$E_{\text{sum}} = \sum_{i=0}^{N_{\text{hits}}} E_{\text{hit},i} \quad (2.10)$$



**Figure 2.37:** Diagrams illustrating the method extracting the time in nanoseconds from a hit. The top figure shows the relationship between the two ramps and a clock cycle. Middle shows the relationship between voltage and time. The bottom figure shows the slope extraction and offset for a chip and channel from an external clock ( $T_0$ ). Adapted from [68].

The 'centre-of-gravity' is defined as the energy-weighted mean position of the hadron shower and is used as described by Equation 2.11:

$$[\text{CoG}_I, \text{CoG}_J, \text{CoG}_K] = \sum_{i=0}^{N_{\text{hits}}} \frac{E_{\text{hit},i} \cdot [I_{\text{hit},i}, J_{\text{hit},i}, K_{\text{hit},i}]}{E_{\text{hit},i}} \quad (2.11)$$

In this thesis, calorimeter cells are always used as a unit of distance. In this unit of measurement, the calorimeter origin point is treated as the leftmost and bottommost cell in the first calorimeter layer relative to the beam.

Additionally, studies are sometimes performed with transformed coordinate systems. The additional variables are outlined as follows:

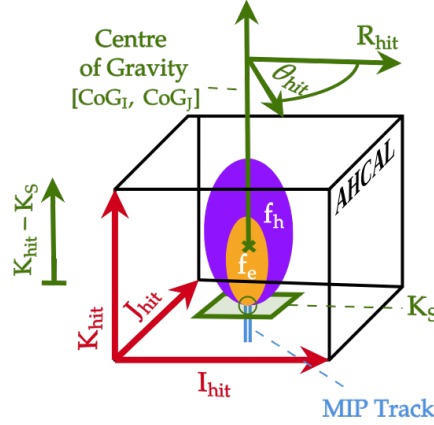
**Hit Radius and Azimuthal Angle** The hit radius defines the distance, in calorimeter cells, from the lateral centre-of-gravity. The azimuthal angle describes a corresponding angle of the hit in radians. The relationships of these quantities are shown in Equations 2.12-2.13:

$$R_{\text{hit}} = \sqrt{(I_{\text{hit}} - \text{CoG}_I)^2 + (J_{\text{hit}} - \text{CoG}_J)^2} \quad (2.12)$$

$$\theta_{\text{hit}} = \arctan2(J_{\text{hit}} - \text{CoG}_J, I_{\text{hit}} - \text{CoG}_I) \quad (2.13)$$

Where  $\arctan2$  is the 2-argument arctangent that implicitly includes the sign of the angle [74]. The reasons for using these variables are twofold. Firstly, this coordinate transformation makes hadron showers, on average, translation invariant, thereby making algorithms designed with hadron shower data implicitly robust to varying positions of impinging particles on the calorimeter front face. Secondly, this coordinate system more adequately describes the lateral development of hadron showers in matter, on average (see Section 2.3.6).

**Shower Start-Normalised Hit Depth** The shower-starting position of a single hadron shower is denoted  $K_S$ . An algorithm achieves the measurement of  $K_S$ . The algorithm studies the depth at which the energy deposited by a hadron shower event contained within a radius of 10 cm from the lateral centre-of-gravity (i.e.  $R_{\text{hit}}$ ) starts to increase laterally relative the energy contained within a radius of 2 cm. It is compared to the ratio of one (i.e. the result expected for muons, which typically deposit their energy via ionisation, and therefore in a highly localised manner compared to the granularity of the detector). The algorithm uses a moving average window of three layers to smooth any fluctuations (e.g. energy deposits from the production of  $\delta$ -electrons from ionisation). If the ratio becomes less than one, then the hadron shower must have initiated and yields the value of  $K_S$  in layers (see [75] and [76] for more details). The position of a hit in a hadron shower relative to this position is denoted  $K_{\text{hit}} - K_S$ . This quantity is used for the same reasons as  $R_{\text{hit}}$  and  $\theta_{\text{hit}}$ , since the hadron shower only develops after  $K_S$ .



**Figure 2.38:** Diagram illustrating the relationship of  $R_{hit}$ ,  $\theta_{hit}$  and  $K_{hit} - K_S$  to the standard co-ordinates provided by the AHCAL readout for shower event initiated by a charged hadron in AHCAL.  $f_h$  and  $f_e$ , coloured purple and orange, indicate the expected hadronic and electromagnetic fractions of the event, respectively.

**Rescaled Energy and Time** The detector primarily measures individual cell energy deposits from ionisation. It is well-known that ionisation energy loss is Landau-distributed [28]. As discussed in 2.3.6, the temporal distribution of the hadron shower in steel had been observed to follow a bi-exponential structure: a fast component due to evaporation of MeV-scale neutrons and a slow component due to nuclear  $\gamma$  from neutron capture and  $\gamma$  from recoil protons induced by neutron elastic scattering [32]. For these reasons, the distributions of  $E_{hit}$  and  $t_{hit}$  are highly skewed. Consequently, it is often beneficial to transform the  $E_{hit}$  and  $t_{hit}$  co-ordinates. A logarithm may be applied to the positive hit energy ( $E_{hit} \rightarrow \log E_{hit}$ ). The temporal component may have negative values due to smearing, which is a stochastic effect and depends on the scale of the time resolution. Therefore,  $t_{hit}$  is transformed with an inverse hyperbolic sine ( $t_{hit} \rightarrow \operatorname{arcsinh} t_{hit}$ ). This transformation performs similarly to the logarithm for positive values while permitting the presence of a small proportion of negative values.

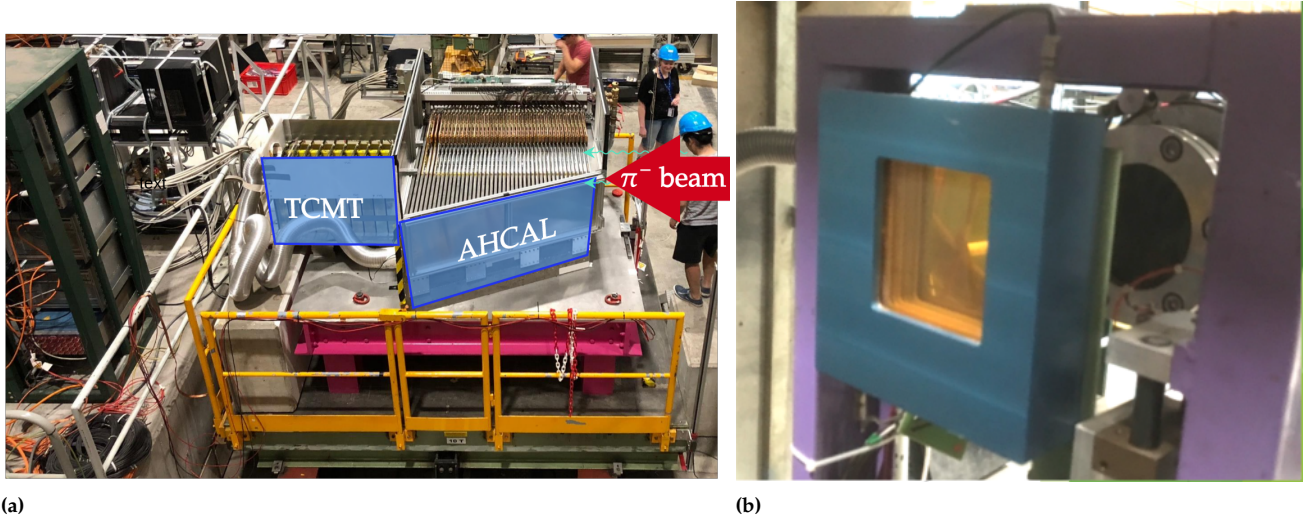
### 2.5.1.3 2018 SPS June Testbeam

In May, June and October 2018, CALICE performed an experiment at the ‘Super Proton Synchrotron’ (SPS) at CERN to study the AHCAL response to muons, electrons and hadrons ( $\pi^-$  mesons). Several  $10^5$  events were measured for each particle, for particle energies in the range 10-200 GeV, using the full technological prototype for the AHCAL detector.

The testbeam campaign was performed for technical and scientific research [77]. Technical studies were focused on a demonstration of the effectiveness ‘SiPM-on-tile’ technology and the reliability of the performance of the detector in power-pulsing mode. Physics studies were focused on the study of the intrinsic resolution and linearity of the AHCAL and energy development profiles[34], its shower separation capabilities using Pandora PFA [78] and particle identification capabilities of the detector [79].

The June 2018 testbeam is studied in this thesis. This testbeam added several pieces of additional instrumentation to the AHCAL. A single HBU





**Figure 2.39:** Pictures of the CALICE AHCAL at 2018 Testbeam. Figure 2.39a depicts the calorimeter setup at the 2018 June SPS testbeam, indicating the beam position and instrumentation. Taken from [77]. Figure 2.39b depicts a wire chamber used for track reconstruction at the 2018 June testbeam. Taken from [81].

was used as a ‘pre-shower’ detector, mainly for triggering. In addition, The hadronic calorimeter was complemented by a steel-scintillator Tail Catcher/Muon Tracker (TCMT) detector, composed of 320 extruded scintillator strips of  $5 \times 0.5 \text{ cm}^2$  area packaged in  $16 \times 1 \text{ m}$  square planes interleaved between steel plates. The TCMT corresponds to a total additional depth of  $5.8 \lambda_I$  to the AHCAL detector [80]. The TCMT was used in this analysis to tag leakage from the AHCAL. In addition, the incident track was reconstructed using four *delay wire-chambers* (DWC) of  $10 \times 10 \text{ cm}^2$  size, which allowed reconstruction of the position (a pseudo-‘track’) of the charged particle impinging onto the calorimeter [81]. The June 2018 Testbeam data is utilised in this thesis for validation studies. Both the AHCAL and TCMT are utilised in this thesis where applicable.

It should be noted that the last calorimeter layer of the AHCAL during this testbeam utilised a ganged layer of  $6 \times 6 \text{ cm}^2$  tiles during this testbeam (‘*Tokyo layer*’). Energy depositions in this layer are omitted in this thesis due to the need for isotropy of the cell size distribution in the calorimeter for the studies of shower separation performed.

Pictures of the CALICE AHCAL at the June Testbeam are shown in Figure 2.39. Further experimental details can be found in [60].

#### 2.5.1.4 Software and Simulation Tools for AHCAL

In this thesis, 2018 June Testbeam data and simulations of this experiment were used to study high-level algorithms for hadron shower clustering and software compensation relied on both simulation, detector descriptions, digitisation of simulation and event reconstruction. Several software modules were used for this purpose, which is introduced in brief.



**iLCSoft** iLCSoft [82] is the underlying software framework to simulate future linear collider instrumentation. iLCSoft v02-00-01 is used in this thesis. It consists of several specialised submodules which focus on different aspects of the detector simulation and data reconstruction for instrumentation:

- ▶ Simulation of particle interactions is provided by the ubiquitous Geant4 simulation package [83]. In this thesis,  $\pi^-$  and  $K_L^0$  hadron showers are simulated interacting with the AHCAL apparatus from a 'beam gun' at a distance of 50 m from the detector front face. The simulation used the QGSP\_BERT\_HP physics list, using Geant4 v.10.3p2.
- ▶ Full descriptions of AHCAL detector geometry, detector materials, alignment, readout and calibration are provided for simulation using the DD4hep v.01-07-01 package [84];
- ▶ Event reconstruction is handled with the 'Modular Analysis and Reconstruction for the Linear Collider' (Marlin) framework. Marlin provides the option to apply custom event reconstruction and analysis modules to the parity between the methods used to reconstruct data and simulation [85];

**CALICESoft** CALICESoft [86] consists of specialised algorithms and modules for use with Marlin to perform digitisation of simulation, reconstruction and analysis of both data and simulation taken at the various testbeam campaigns. CALICESoft v.04-15 includes the specific simulation, digitisation and reconstruction modules used for June 2018 simulation and data in this thesis. Comparisons between simulation and experimental data from the 2018 SPS June Testbeam are studied in depth in [34] and are presented independently in this thesis in Chapter 5.

### 2.5.1.5 Five-Dimensional Calorimetry

The AHCAL is a unique detector in that it is capable of recording energy depositions with up to 100 ps time resolution. This makes the calorimeter five-dimensional, rather than the typical four-dimensional calorimeters required by PF. Sub-nanosecond timing resolution is expected to improve PF algorithms, which are presently discussed.

**Timing in Software Compensation Algorithms** As previously mentioned in Section Studies of time development of hadron showers in the AHCAL have revealed that a steel calorimeter such as AHCAL is sensitive to three distributions of energy deposits from hadron showers: the electromagnetic fraction, which is instantaneous, and the energy depositions from the intra-nuclear cascade initiated by elastic neutron scattering, and those from neutron capture. Timing information is expected to provide improved event-by-event compensation capability, as the fraction of energy deposited in the later epochs of the hadron shower development is highly correlated to the missing energy fraction. Further discussion and research on this subject are presented in Chapter 5.

Parameter	Value
Particle Type	$\pi^- / K_L^0$ Beam
Physics List	QGSP_BERT_HP
Beam Gun, $x$	-15 mm
Beam Gun, $y$	-15 mm
Beam Gun, $z$	-50 m
Beam Width $x$	27.1 mm
Beam Width $y$	25.8 mm

**Table 2.5:** Table of Geant4 simulation parameters used for hadron shower simulations in AHCAL in this thesis.

The extent to which timing information can optimally improve both aspects of PF is unknown. It, therefore, motivates the work in this thesis to assess the possible improvements in shower separation performance that may be achievable with the AHCAL calorimeter operating at 100 ps timing resolution.

**Timing in Shower Separation Algorithms** An obvious use for timing information in hadron shower separation is separating showers by time-of-flight (i.e. hadrons that shower deeper in the calorimeter have later energy depositions compared to those that showered in the first few layers). However, on an event-by-event basis, the temporal development of a hadron shower is expected to vary significantly due to stochastic fluctuations in the hadron shower. An appropriate algorithm may utilise this information to cluster hadron showers not only in space but also in time, utilising differences between the energy-temporal profiles of the two showers. It is unclear how useful timing information will improve shower separation capability compared to without it. Further discussion and research are presented in Chapter 6.

As a caveat in this thesis, it should be noted that timing information is not currently available in June 2018 SPS Testbeam data to the quality required for use in machine learning algorithms. This is a consequence of chip occupancy effects, which result in artificial hit times arising from the challenge of determining a timing offset in these cases. This effect results in non-physical negative times and artificially many hits in the critical transition period between the fast EM component of the hadron shower and the slow HAD component (first 10 ns)<sup>30</sup>.

30: When a SPIROC2E chip experiences high memory-cell occupancy, the spread of the distribution of hit times becomes so variable that it is effectively impossible to calculate an offset for the timing information (see Figure 2.37). This may extend to 'negative times' because the timing depends on the parity of the bunch crossing and the timing information being encoded as a voltage rather than an incremental value relative to an external clock. This effect worsens in significance as the occupancy increases, such as when a shower is centred on a single ASIC rather than spread out across multiple ASICs, which affects the critical region of times between the fast EM component of the hadron shower and the slower HAD component and reduces for later, lower occupancy energy depositions deposited by neutron capture [87].

## 2.5.2 Silicon Photomultipliers

Silicon photomultipliers (SiPMs), '*multi-pixel photon counters*' (MPPC), or *Geiger Avalanche Photodiode* (G-APD) when referring to a single pixel, are semiconductor devices that are used to amplify the signal of photons in the near-infrared to the ultraviolet range of frequencies to be detected. This design of photodetector is advantageous because it offers single photon resolution, high detection efficiency, unaffected by magnetic fields, and sub-nanosecond time resolution making them ideal candidates for calorimeter readout. However, SiPMs are versatile energy detectors used in a variety of other fields: two further examples include in medicine, for combining positron emission tomography with magnetic resonance imaging, which capitalises on the invariance to magnetic fields, as infrared detection and ranging (LIDAR) sensors for self-driving cars, which utilise the high efficiency and fast timing resolution of these detectors.

In this section, the concept of a semiconductor is introduced. Silicon, a semiconducting element, is then introduced, and the special properties that arise from doping the material with other elements are discussed. The concept of a p-n junction is then introduced and described. The design of the SiPM device itself is then described in brief. The signal of SiPMs under illumination by light are then discussed, as well as sources of correlated noise, and related to the AHCAL detector. The need to characterise SiPM charge spectra is then discussed in the context of the AHCAL, which motivates part of the studies of this thesis.

### 2.5.2.1 Semiconductors

At the atomic level, from the Pauli Exclusion Principle, fermions such as electrons cannot occupy the same quantum state around an atomic nucleus. This concept gives rise to the '*atomic band structure*': electrons occupy discrete energy levels around atomic nuclei. More than one electron may sit on the same band if their quantum numbers allow it (e.g. a He nucleus permits two electrons of antiparallel spin to sit on the same ground state energy level).

Electrons in an atom's band structure may be considered discrete, single-particle states. Assuming thermal equilibrium, a hypothetical state where there is a 50% probability of occupation by an electron is called the '*Fermi level*' of the material. The shell of electrons closest to the Fermi level is called the '*valence band*'. The first unfilled state above the Fermi level is called the '*conduction band*'. The difference between the valence and conduction bands is called the '*band gap*'. The Fermi level of the material sits between these two bands.

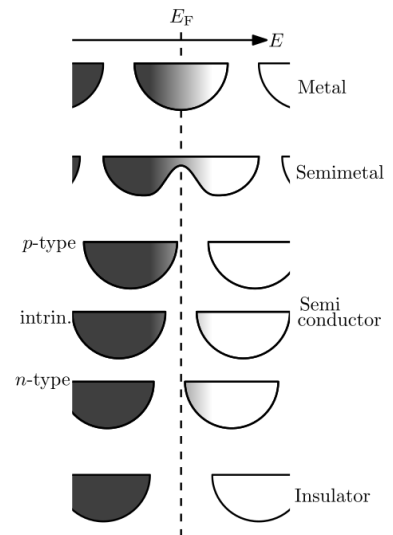
Conductive materials (e.g. metals like Fe or Cu) have overlapping conduction and valence bands. Therefore electrons may readily flow to and from the valence band to the conduction band. Conduction is negligibly likely under any condition for insulators (e.g., plastic, glass, quartz). Materials called '*semiconductors*' have an electrical conductivity that falls somewhere between a conductor and an insulator. This means that there is some finite probability for semiconductors, at ordinary temperatures and pressures, to conduct. Put another way; the bands are close enough to the Fermi level in these materials so that they may be thermally occupied with electrons, or the absence thereof ('*holes*').

Semiconductors are used ubiquitously in modern electronics because they can be made to conduct when certain conditions are met (e.g. in transistors). In the context of particle physics detectors, this usually means interacting with a particle, such as a photon or ionising radiation, thereby producing a signal only when the particle is observed.

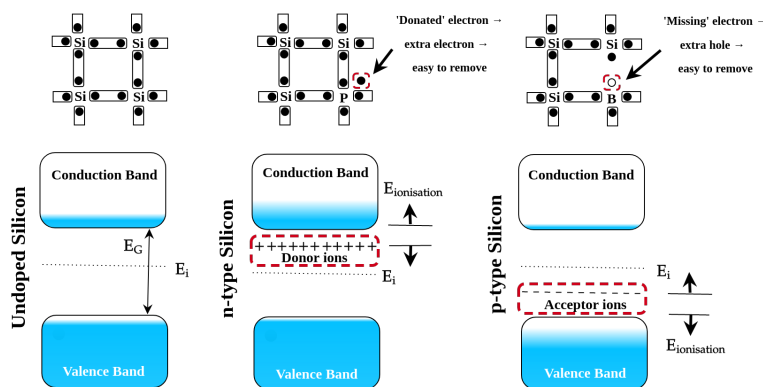
### 2.5.2.2 Silicon and Doping

Silicon (Si,  $Z=14$ ) is a semiconducting element. In nature, pure silicon forms a crystal. A crystal is any matter that forms regular structures at the microscopic level that repeats throughout ('*crystal lattice*'). Silicon crystal is composed of single atoms, which means the crystal structure is very stable and has well-known conductivity and electrical properties. Silicon is plentiful in the Earth's crust in the form of silicates and silica (silicon dioxide) and is cheap to extract. Silicon has four valence electrons and a band gap energy of 1.12 eV.

Silicon, on its own, is a poor conductor and insulator. However, it is a particularly special element that can be doped with other elements to change its electrical properties by making it more conductive. The crystal structure of silicon may be modified with two main classes of dopants:



**Figure 2.40:** Diagram illustrating the difference between conductors, semiconductors and insulators. The width indicates the energy level, the height is the density of states available, and the dashed line indicates the Fermi level. Black through white shading indicates the probability an electron inhabits the state (black: all states filled, white: no states filled), Taken from [88].



**Figure 2.41:** Diagram illustrating the difference between undoped silicon, p-type silicon and n-type silicon. The upper diagram illustrates the crystal structure of the silicon, with black circles indicating electrons and white circles indicating holes.  $E_G$  indicates the band gap.  $E_i$  is the intrinsic Fermi level (i.e. as would be the case if unmodified by the dopants). The + and - symbols indicate the charge of the ions.  $E_{ionisation}$  indicates the ionisation energy required for transitions between states. The blue-to-white shading indicates the probability of containing a filled electron state. Adapted from [89].

**Donors** Donor atoms modify the Fermi level of the silicon so that it is close to the conduction band of the crystal and introduce a new energy level above the valence band and below the Fermi level. Electrons at this level may be easily excited into the conduction band. In summary, this silicon makes it more probable to have an electron in the conduction band than in undoped silicon and, therefore, ‘donates’ electrons. The resulting material is called ‘*n-type silicon*’. Donors have five valence electrons, meaning Group V elements (e.g. P, As, Sb, Bi) are used as donors in silicon. Of them, phosphorus (P) is most commonly used as a dopant for silicon.

**Acceptors** Acceptor atoms modify the Fermi level of the silicon so that it is close to the valence band and introduce a new energy level above the Fermi level and below the conduction band. Electrons at this level may be excited there from the valence band, meaning a hole is left behind. In summary, this type of silicon makes it more probable that an electron is not in the valence band of the undoped silicon and, therefore, capable of ‘accepting’ electrons there. The resulting material is called ‘*p-type silicon*’. Acceptors have three valence electrons, meaning Group III elements (e.g. B, Al, Ga, In) are used as acceptors in silicon. Of them, Boron (B) is most commonly used as a dopant for silicon.

Different concentrations of doping in silicon may also be used in the same device. The relative doping concentrations are indicated by + and - superscripts. Ordinary doped silicon is of a concentration of  $1 \times 10^{10} \text{ cm}^{-3}$  dopant and can be many orders of magnitude higher depending on the application.

### 2.5.2.3 p-n Junctions

Doped silicon has greatly improved conductivity compared to undoped silicon. However, for use in particle detector experiments, the silicon must be sensitive to the influence of very small numbers of charge carriers produced by interactions of light or charged particles with the crystal.

For instance, a single photon will produce a single electron-hole pair if it interacts with a silicon atom via the photoelectric effect, which is 'a drop in the ocean' of free charge carriers in silicon provided by even the lightest of doping concentrations, thereby making the event undetectable. This can be resolved by introducing a 'depletion region' in the silicon where the number of free charge carriers is significantly reduced.

Constructing a 'p-n junction' may create a depletion region. A p-n junction is a basic diode: it allows current to flow in one direction ('forward bias') but not the other ('reverse bias').

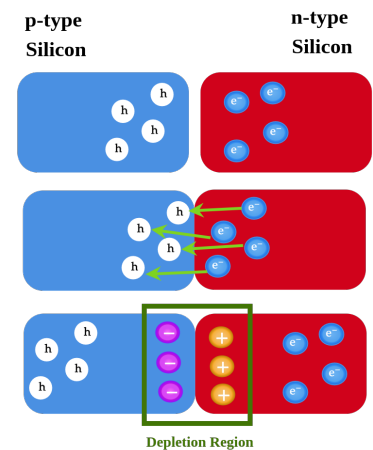
When placed together, holes diffuse across the junction from the p-type to the n-type silicon and recombine with the free electrons, and vice versa. This leaves negative ions in the p-type silicon and positive ions in the n-type silicon. The high electric field of these ions at the junction prevents further diffusion. This creates a region where the number of free charge carriers is depleted compared to the rest of the junction. This is called the depletion region. It should be noted that, under applied reverse bias, more free charge carriers are pulled from the silicon; therefore, the width of the depleted region subsequently expands.

#### 2.5.2.4 The Design of Silicon Photomultipliers

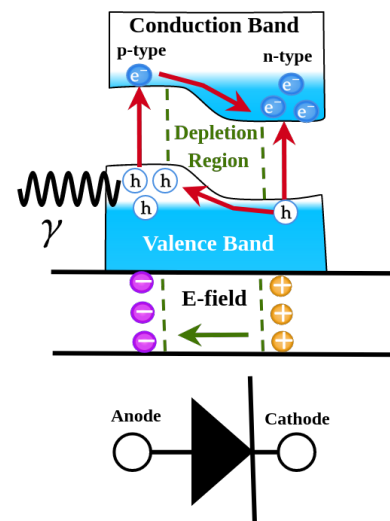
A SiPM consists of matrices of *single-photon avalanche diodes* (SPADs). The typical size of the pixel is  $10 \times 10 \mu\text{m}^2$  to  $100 \times 100 \mu\text{m}^2$ . The matrices consist of hundreds to several tens of thousands of individual pixels, each of which may measure the presence of single photons.

The light-detection sensor of a SiPM pixel may be viewed as a p-n junction operated under reverse bias and with an applied voltage across the device. In this configuration, minimal current flows through the p-n junction when in a quiescent state. However, when a photon induces the photoelectric effect in the depletion region or a valence electron is thermally excited in the depletion region, it generates an additional electron-hole pair, known as 'excess charge carriers'. The excess charge carriers drift towards the regions of the majority concentrations (electrons to cathode, holes to anode). This is because the charge carriers must travel opposite the direction of the electric field in the depletion region by moving from the high to the low potential region of the p-n junction.

While current flows in this configuration, it is much more likely that free charge carriers will recombine with ions than produce a measurable signal. If, however, the electric field of the depletion region is sufficiently large (typically  $\sim 300 \text{ kV cm}^{-1} / 30 \text{ V } \mu\text{m}^{-1}$ ), the electrons will accelerate and create more charge carriers by 'impact ionisation'. This leads to charge carrier multiplication. If the applied voltage is sufficiently large, a self-sustaining micro-plasma tunnel of around  $10 \mu\text{m}$  diameter, typically involving around  $10^5$ - $10^6$  avalanche electrons per photoelectron called a 'Geiger avalanche' [91]. This pixel state is called 'breakdown'. Breakdown may only occur at voltages greater than the 'breakdown voltage',  $V_{\text{bd}}$ . The voltage requirement arises because the electric field in the depletion region must be sufficiently strong that charge carrier multiplication can occur (i.e. the kinetic energy of charge carriers must therefore become greater than the 1.12 eV gap between the silicon valence and conduction band within the mean free path of the charge carrier in



**Figure 2.42:** Diagram illustrating the diffusion of electrons from the n-type to the p-type junction, which is indicated by blue and red regions, respectively. Purple and orange circles indicate ions formed by electrons and holes recombining with silicon atoms. Blue and white circles indicate electrons and holes. The green rectangle indicates the depletion region, where no further electron-hole recombination can occur.



**Figure 2.43:** Diagram illustrating the breaking of equilibrium at the p-n junction by excitation by a photon with a energy larger than the silicon band gap. Symbols as in Figure 2.41. Adapted from [90].

silicon). If this condition is met, at least one electron-hole pair will be produced per ionising impact. Therefore, SiPMs are operated at a voltage above  $V_{bd}$ , called the bias voltage,  $V_b$  [92]. The operation voltage of a SiPM is typically quoted as an 'over-voltage', which is defined  $V_{OV} = V_b - V_{bd}$ .

Next, when the diode breaks down, current flows through the diode, thereby producing a measurable signal from a photon. However, a Geiger avalanche is self-sustaining and will continue indefinitely if nothing is done. Therefore, a 'quenching resistor' component acts to 'quench' the Geiger avalanche and restore the diode to its quiescent state. The voltage at which the avalanche may not sustain itself is called the 'off-voltage' ( $V_{off}$ ), and another discharge may be measured by the device. It should be noted that in the study of [93], the off-voltage and the breakdown voltage have been measured to be different. This result is expected because the energy required to initiate a Geiger avalanche is greater than the energy required to sustain it [94]. Since there remains ambiguity as to the exact difference between the off-voltage and breakdown voltage, over-voltage can also be quoted relative to the off-voltage,  $V_{OV} = V_b - V_{off}$ .

In the absence of noise effects, the signal from a SiPM device under illumination by light is a pulse of current caused by a Geiger avalanche initiated by a photon. Each SPAD pixel can independently produce a signal from a photon under illumination. The number of photons the device measures during an event is proportional to the total charge measured within an integration window. The number of photons incident on the SiPM array can therefore be measured.

The pulse is bi-exponentially distributed, depending upon two time constants [91]:

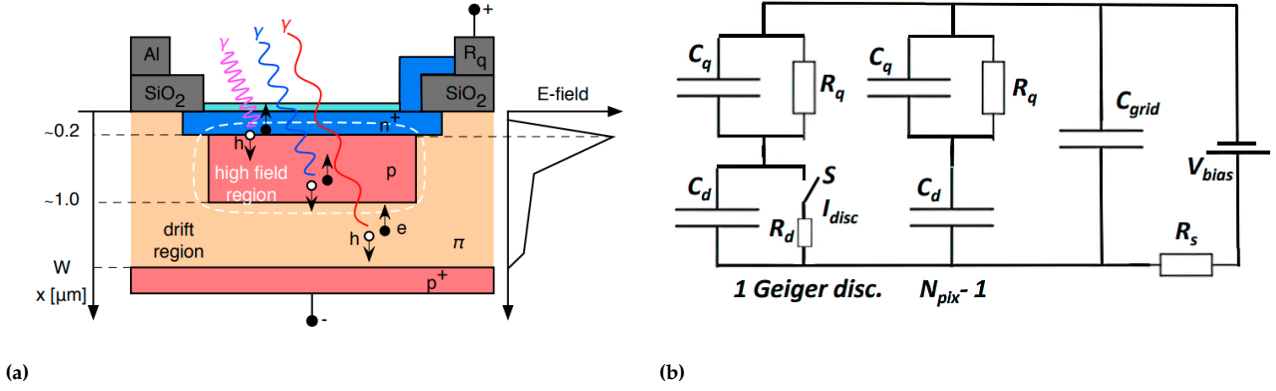
- ▶ *Fast-component time constant,  $\tau_f$* : the fast component of the pulse is caused by parasitic capacitance<sup>31</sup> to the silicon bulk,  $\tau_f = R_s \cdot C_q$ , where  $R_s$  is a shunt resistance that converts the current signal into a voltage and  $C_q$  is parasitic capacitance. This time constant is around 1-5 ns.
- ▶ *Slow-component time constant,  $\tau_s/\tau$* : the slow component of the pulse is caused by the recharging of the pixel through the quenching resistor, ( $\tau_s = R_q \cdot (C_d + C_q)$ ), where  $R_q$  is quenching resistance, and  $C_q$  and  $C_d$  are the parasitic and diode capacitance, respectively. This time constant is around 10-40 ns.

A cross-section of a SiPM pixel is shown in Figure 2.44a, showing a circuit diagram of a SiPM in Figure 2.44b.

The leading producer of SiPMs in the world contemporary with this thesis is Hamamatsu [92].

31: Parasitic capacitance is produced by the interaction of the electric fields of two close-by components in a circuit.





**Figure 2.44:** Figure 2.44a shows a cross-section of a typical SiPM pixel under illumination.  $n$  and  $p$  indicate  $p$  and  $n$ -type silicon, and  $\pi$  is the 'drift region'. The rotated graph on the right-hand side indicates the electric field in the device. The high electric field region (the depletion region) indicates the junction between the  $p$  and  $n$  type material. It indicates where amplification of the excess charge carriers is produced via the photoelectric effect. The depth of the device is indicated on the left-hand side (Property of the University of Hamburg). Figure 2.44b shows an example circuit diagram illustrating the design of a SiPM pixel array.  $V_{bias}$  ( $V_b$ ) represents a reverse-bias voltage applied to the SiPM, to increase the depletion region of the diode and thereby increase the amplification of the device.  $R_d$  and  $C_d$  and represent the resistance and capacitance of each pixel, respectively.  $I_{disc}$  represents the discharge current caused by a Geiger Discharge.  $S$  represents a 'switch' that switches on during a discharge and off once a certain off-voltage ( $V_{off}$ ) is reached.  $R_s$  is a shunt resistance that converts the current signal into a voltage.  $R_q$  and  $C_q$  represent a quenching resistance and a parasitic capacitance to the silicon bulk. Taken from [91].

## 2.5.3 Characterisation of SiPMs

The main variables of interest characterising the behaviour of a SiPM under illumination and noise sources that influence the pulses are now introduced [95–97]. Next, the concept of a charge spectrum is introduced, and an example experiment for measuring them is described. How they can be used to characterise SiPMs is explained. The relevance of SiPM characterisation to the AHCAL calibration is then discussed.

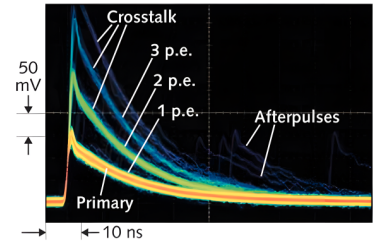
### 2.5.3.1 SiPM Performance Parameters and Noise Sources

Parameters for describing the behaviour of SiPMs are introduced and discussed in the following section.

**Gain** The total amplification achieved by the SiPM sensor measured in charge is defined as the product of the overvoltage and the combined capacitance of the diode and parasitic capacitance of the quenching resistor:

$$G = (C_d + C_q) \cdot V_{OV} \quad (2.14)$$

It should be noted that only a fraction of the total gain is reconstructed if the pulse is integrated within a gate. The gain measured, in this case, is called an 'effective gain',  $G^* = f_{gate} \cdot G$ , where  $G^*$  in this thesis is defined as the effective gain in this thesis and  $f_{gate}$  is the fraction of total integrated charge of a SiPM pulse allowed by an integration gate length [91]. The gain may also be quoted in avalanche electrons per photoelectron in units of the elementary charge. It is also clear that gain increases linearly with over-voltage. Therefore, the off-voltage of the SiPM may be extracted from experiment. It also converts the measured charge to a number of photoelectrons that can be counted.



**Figure 2.45:** A representative two-dimensional histogram representing the voltage over a 50  $\Omega$  resistor produced by a SiPM over time in nanoseconds of a Hamamatsu SiPM under low-intensity illumination by pulsed light, after a factor of 10 linear charge amplification. The bi-exponential pulse shape of the discharges of the SiPM is visible. The number of photoelectrons and discharges from correlated noise sources, such as cross-talk and afterpulsing are indicated. Property of the University of Hamburg.

32: The fill factor of a SiPM describes the ratio of the active area to the total area of the pixel.

33: Quantum efficiency is the joint probability that a photon traverses the anti-reflective coating of the pixel and produces an electron-hole pair that reaches the depletion region. It depends on the frequency of light.

34: The avalanche-triggering probability describes the likelihood that a photon-induced electron-hole pair induces a Geiger avalanche. It depends on over-voltage, the pixels' temperature and the light frequency. The frequency dependence arises because higher frequencies of light have a lower attenuation depth in silicon and, therefore may not travel deep enough into the depletion region to be fully amplified. The thermal dependence arises because producing a free charge carrier in silicon requires a photon and a phonon from the silicon lattice. Therefore, as the temperature decreases, so too does the likelihood that a Geiger avalanche occurs.

35: 'Interstitial' means an atom that is not part of the regular structure of the crystal.

**Pedestal Position** The *pedestal position* (or simply '*pedestal*') is the mean charge measured from the SiPM device when no photons are incident on the device. It converts the charge to some photoelectrons that can be counted.

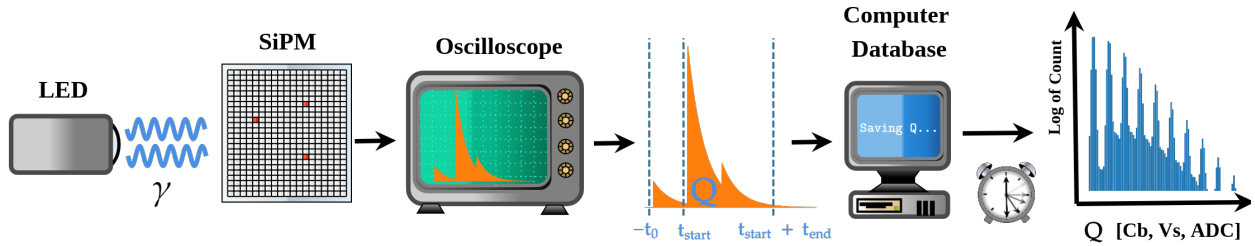
**Photon Detection Efficiency** The '*photon detection efficiency*' (PDE) of a SiPM describes the total probability an incident photon is to induce a Geiger avalanche and produce a measurable signal. The PDE is equal to the product of the '*fill factor*' of the device<sup>32</sup>, the '*quantum efficiency*'<sup>33</sup> and the '*avalanche-triggering probability*'<sup>34</sup> [95]. As the light intensity on a SiPM increases, more pixels will fire, and the PDE will saturate. The SiPM will no longer produce additional discharges, resulting in the non-linearity of the SiPM response under illumination (see Section 2.4.4 and Section 2.5.3.2). Furthermore, the PDE of a SiPM will change depending on the temperature of the SiPM. Consequently, the PDE is a critical figure-of-merit of a SiPM's performance. More details on measurements of the PDE are discussed in Appendix Section 8.1.0.1.

**Electronics Noise and Gain Fluctuations** The integrated charge from a SiPM around the pedestal value is normally distributed due to electronics noise (denoted  $\sigma_0$ ). As the number of firing pixels increases, so too do gain fluctuations within one pixel and across multiple pixels (denoted  $\sigma_1$ ) contribute to the noise. Gain fluctuations arise primarily from pixel-to-pixel fluctuations of  $V_{\text{off}}$ .  $V_{\text{off}}$  fluctuates because the quenching of Geiger discharges is a stochastic process. Differences in the capacitance of pixels in a SiPM also contribute to the gain fluctuations [96].

**Prompt and Delayed Cross-talk** Cross-talk is a type of discharge where charge carriers from a primary Geiger discharge in a pixel contribute to the signal measured by other pixels and are produced as a result of secondary photons from a Geiger avalanche producing an electron-hole pair in another pixel. Suppose the photon passes into the depleted region of another pixel directly. In that case, it is called '*prompt cross-talk*', and a coincident pulse of the same amplitude as the primary Geiger discharge pulse. If the secondary photon enters the non-depleted region of another pixel, the charge carriers will diffuse to the depleted region of that cell; it is called '*delayed cross-talk*'. It will produce another delayed pulse of the same amplitude as the primary Geiger discharge. The influence of cross-talk is typically defined as an '*excess noise factor*'.

**After-pulsing** Charge carriers can be captured by '*defects*' in the silicon crystal (e.g. '*deep defects*' such as interstitial<sup>35</sup> carbon and oxygen atoms in the crystal, known as a CiOi defect, or vacancy-related traps in the silicon) which introduce undesired energy levels in the silicon band gap. Charge carriers may be captured by such a defect and released later. This is called '*trapping*'. The charge carriers may be released after some time and contribute a delayed signal to the same pixel with diminished amplitude depending on the recharge state of the SiPM. Afterpulses can only occur if a primary Geiger discharge first takes place in the pixel. Delayed cross-talk in the same cell as the primary discharge is analogous





**Figure 2.46:** Example experiment illustrating how SiPM spectra may be measured. A SiPM is illuminated with light from an LED or laser. Geiger avalanches are produced in the SiPM pixels, producing voltage pulses. Next, the signal from the SiPM is amplified and measured with an oscilloscope. The charge is then integrated within a window known as a gate ( $t_{\text{gate}}$ ) from a time slightly before the start of the pulse ( $t_{\text{start}}$ ), and may include a period before the gate to measure dark counts ( $t_0$ ). Finally, the charge ( $Q$ ) is recorded in a computer database. The distribution of a charge spectrum is produced from many discharges, which may be used to characterise the SiPM under test.

to an afterpulse. The probability with which a discharge produces a single afterpulse is called '*after-pulse probability*'.

**Dark Counts** Thermal excitations of electron-hole pairs can also initiate discharges, called 'dark counts', which are uniformly distributed in time. This effect is temperature-dependent and may increase with radiation damage to the SiPM. The contribution of dark counts to a SiPM signal is measured as a '*dark count rate*' (*DCR*), measured as a frequency typically in the order of kHz. After irradiation, it may increase to MHz.

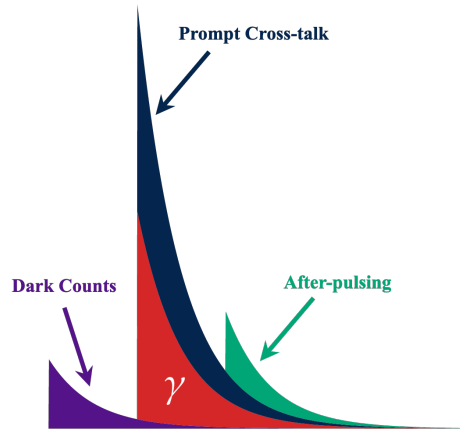
### 2.5.3.2 Characterisation of SiPMs from Charge Spectra

One method of extracting SiPM parameters is to measure the device's response under illumination by low-intensity LED or laser. The voltage pulses from the SiPM may be measured and recorded using an oscilloscope. The current produced by the SiPM may be integrated within a certain gate length, yielding a charge. The result is a distribution of charges, known as a '*charge spectrum*'. A diagram illustrating an experiment for measuring a charge spectrum is presented in Figure 2.46. This diagram also shows an SiPM charge spectrum that is influenced by correlated noise, dark counts and after-pulses. A breakdown of the contribution of several common sources of correlated noise to a single photon pulse is shown in Figure 2.47.

As mentioned in Section 2.5.1.1, there are  $\sim 22000$  channels in AHCAL, and therefore as many SiPMs. The characterisation of each SiPM must be performed as part of the detector calibration procedure. Detector calibration here means the accurate conversion of the discharge amplitudes measured from the pulses of SiPMs to an appropriate physical energy unit, as well as monitoring changes in the response of the detector to temperature<sup>36</sup> and, eventually, degradation in the performance of the devices with radiation damage. As discussed in Section 2.4.2, detector calibration plays a role in detector resolution and therefore must be performed as accurately as possible.

Section 2.5.1.2 introduced  $E_{\text{hit}}$ , the calibrated energy measured by an AHCAL cell. This must be obtained from the signals measured by the SiPMs. Specifically, several steps are required to produce a calibrated

36: This is due to the  $V_{\text{off}}$ -temperature dependence, discussed in Section 2.5.2.4. Therefore, the bias voltage of SiPMs must be increased on the SiPMs as the temperature increases. The average temperature of the detector will increase during operation, so the detector's performance must be monitored.



**Figure 2.47:** Illustration of several possible sources of noise that might influence a particular event. The red pulse indicates a discharge from a photon incident on the device. Noise sources, such as dark counts, cross-talk and after-pulsing, are shown in purple, indigo and green.

37: Each aspect of AHCAL calibration was studied in depth in the references. Specific details on the methods and results obtained during these studies are available.

and equalized response from all calorimeter cells. Three main steps are required <sup>37</sup> :

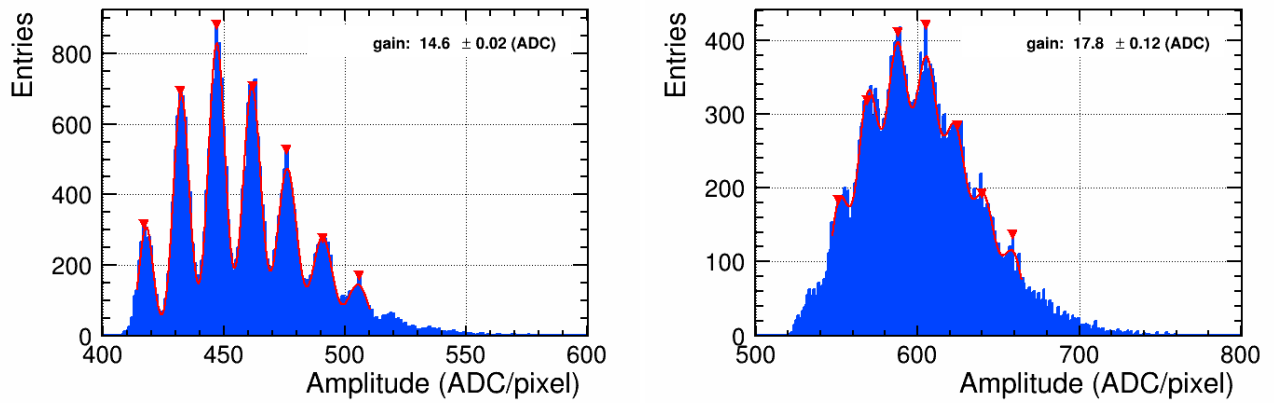
- ▶ *MIP Calibration*,  $C_{\text{MIP}}$ : a measurement of each channel to a muon from a testbeam (i.e. a MIP-like particle) to determine the MPV, in ADC units from the expected Landau energy distribution from ionisation observed from the device [78];
- ▶ *Saturation Correction Function*  $f_{\text{sat}}^{-1}$ : a measurement of a function that linearises the response of a SiPM as the intensity of illumination increases on the device, which is related to the PDE of the SiPM. This correction factor is necessary to linearise the response of the SiPMs under intense illumination that may be experienced in high-energy-density events (i.e. the EM fraction of a hadron shower)[34, 54];
- ▶ *High-Gain/Low-Gain Intercalibration Factor*,  $C_{\text{IC}}$ : a measurement of a factor which accounts for discontinuous ‘switching’ between the high-gain mode of the readout (used for low-intensity signals, to improve the photon resolution, for up to around 100 fired pixels) and low-gain mode of the readout (used for high-intensity signals, to reduce power draw) [98].

$E_{\text{hit}}$  is therefore obtained using Equation 2.15a

$$E_{\text{hit}} = f_{\text{sat}}^{-1} (K \cdot C_{\text{IC}}) \cdot \frac{(Q - Q_0) \cdot C_{\text{IC}}}{C_{\text{MIP}}} [\text{MIP}] \quad (2.15a)$$

$$K = \frac{(Q - Q_0)}{G_*} \quad (2.15b)$$

where  $Q$  is the ADC value measured by the SiPM, and  $K$  is defined as the pedestal and gain-normalised charge measured by the SiPM, in the number of photoelectrons (p.e),  $Q_0$  and  $G_*$  are the SiPM pedestal and effective gain.



(a) Low dark count rate.

(b) High dark count rate.

**Figure 2.48:** Histogram illustrating a gain measurement of a Hamamatsu S13360-1325PE MPPC SiPM of an AHCAL cell, taken during an LED run in external trigger mode, measured in ‘analogue to digital’ (ADC) charge units, indicated by the blue filled region. The distribution has been fitted using a ‘multi-Gaus’ distribution, indicated by a red line. The peaks are indicated by downwards facing red arrows. Figure 2.48a shows a fit of this model to a spectrum with a low dark count rate. Figure 2.48b shows the same model applied to a spectrum with a high dark count rate. Both taken from [34].

It is clear from this equation that all named quantities need to be extracted for each channel separately. Mass calibration of the SiPMs is therefore necessary.

SiPMs may be characterised using the distribution of discharges measured from the device under illumination. These distributions are called ‘charge spectra’ and contain a wealth of information on the performance of the SiPM device and the sources of noise it experiences under experimental conditions. It is used to extract the pedestal and gain.

Firstly, in Figure 2.48a, not all the peaks have been fitted, and the individual normal distributions exhibit deviations from the experimental peak values. The effect is even more relevant in Figure 2.48b; the effect of dark counts, which correspond to the underlying distribution that reduces the prominence of the peaks, has not been included in the model. This means that the normal distributions alone produce a poor agreement with experimental data. This disagreement motivates a tool to fit the entire charge spectrum, thereby obtaining more accurate values for the gain and pedestal, which propagate to better detector calibration and, therefore, better AHCAL energy resolution. Furthermore, such a tool would implicitly fit the pedestal position that may readily be extracted from these spectra, thereby making a separate measurement unnecessary. In addition, noise effects such as after-pulses, cross-talk and dark count play a role in the saturation correction function, and it is interesting to study these effects en masse.

## 2.5.4 Machine Learning

Machine learning describes a subset of artificial intelligence (AI) involving the study of methods of using empirical data to make predictive models of physical reality.

In many cases, computer algorithms written by a human being often perform effectively enough to achieve an intended result. Pandora PFA is

38: Boosted decision trees are classifiers based on sets of sequential decisions applied to a set of input data. The 'gradient boosting' refers to the combination of several weak classifiers which together form a strong classifier [99] [100].

39: A graphics processing unit (GPU) is a computer component that is designed for rendering images and was used almost exclusively for producing 3D graphics for video games up until the mid-2000s. Compared to a 'central processing unit' (CPU), present in every computer, a GPU has many small processors, which greatly accelerate tasks that can be divided into smaller subtasks that run in parallel. The time taken for the tensor operations required in machine learning can be greatly sped up using this parallelism, making training deep learning models possible. The leading manufacturers of GPUs in the world today are NVidia [101] and AMD [102]

40: A time-projection chamber is a homogeneous particle detector that enables a full two or three-dimensional reconstruction of a particle trajectory or a decay.

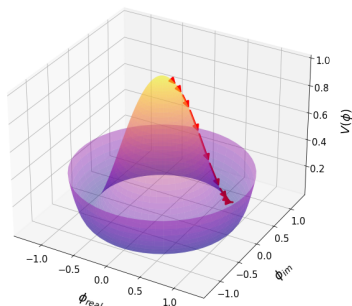
an example of such an algorithm. However, as the number of dimensions or complexity of a problem increases, so too do the number of conditions, caveats and nuances required to achieve an intended goal, and the effectiveness of an algorithm typically reduces.

Another solution is to use empirical data to infer a model based on trial and error ('learning'), which can be rendered as a mathematical problem to be solved. A simple form of machine learning used ubiquitously in particle physics is 'boosted decision trees'<sup>38</sup>, which have been implemented for particle identification (PID) using event information from the AHCAL detector, and is used for analysis in this thesis [79]. Developments in the performance and price of the 'graphics processing unit' (GPU)<sup>39</sup>, 'deep learning' has become a reality and influenced the modern field of AI. Deep learning, though nebulous in definition, refers to models with thousands up to billions of tunable parameters in layers of operations. The deepest machine learning model in the world as of 2023 is the GPT-3 natural language model, which has 175 billion free parameters [103, 104]. Concerning particle physics, such models allow exploitation of the entire parameter space available from the experiment to perform some desired algorithmic result [105].

Examples of applications of machine learning to the challenges faced by the physics community at large are numerous as of the writing of this thesis. Three examples are given: the tagging of  $b$ -jets in the ATLAS detector, of particular interest for top quark reconstruction [106]; background selection for the decays of two-body jet decays using correlated jet-substructure variables in the ATLAS detector [107] and image recognition for time projection chambers for the reconstruction of neutrino interactions that occurred in the LAr *time-projection chambers*<sup>40</sup> of the proposed DUNE experiment, providing the neutrino flavour and decay products of interactions of neutrinos [108] (see [109] for additional literature on applications of machine learning to particle physics).

In particular reference to the AHCAL detector, machine learning the generation of fast Monte Carlo simulations of showers in highly granular calorimeters has been performed in [110].

#### 2.5.4.1 Design of Neural Networks



**Figure 2.49:** Example of gradient descent minimisation. The gradient of the function is evaluated in small steps, and the original position is updated, step by step, based on that value, indicated by the red arrows, until a minimisation criterion is achieved.

Machine learning comprises five components: a set of input features,  $x \in \mathbb{X} = \mathbb{R}^N$ ; a set of target values,  $y \in \mathbb{Y} = \mathbb{R}^M$ ; a model with tunable free parameters,  $f : \mathbb{R}^N \rightarrow \mathbb{R}^M$ ; a loss function, that characterises the performance of the function  $\mathcal{L}[f] : \mathbb{X}^{\mathbb{Y}} \rightarrow \mathbb{R}$  and a strategy for optimisation [105]. Here,  $\mathbb{R}$  refers to the real set,  $N$  and  $M$  are the dimensions of the set,  $\mathbb{X}$  and  $\mathbb{Y}$  are the sets of input and targets,  $x$  and  $y$  are samples from each set, respectively. In this thesis,  $\hat{y}$  is defined as the predicted value of  $y$  from the model  $f$ ,  $\hat{y} = f(x; \theta)$ , where  $\theta$  is defined as all the free parameters that make up the model.

A machine learning model uses 'gradient descent' to achieve the local minimum of the cost function. At its most basic level, this is achieved through making small updates in the values of the parameters, given the gradient of the loss function:

$$\theta \rightarrow \theta - \eta \cdot \Delta \mathcal{L} \quad (2.16)$$

where  $\eta$  is a small step (*'learning rate'*) that defines the step by which the gradient should be updated. The principle of gradient descent is demonstrated by minimising the Higgs Potential by gradient descent shown in Figure 2.49.

"A *'neural network'* is a machine learning model designed to mimic brain processes for learning functions. It consists of sets of *'neurons'* that apply sequential biased weightings to input data. These biases and weightings are optimised to achieve the desired output. By connecting multiple neurons, the network gains the capacity to generate more complex outputs. Neural networks are employed in two studies within this thesis. A brief summary of the learning process of a neural network is provided henceforth.

There are two steps in the training of neural network models: a *'feed-forward'* step and a *'backpropagation'* step [111]. These are explained as follows in the context of the simplest possible neural network, consisting of single neurons. It is then shown the method is extensible to neural networks consisting of many neurons per layer.

**Feed-Forward Step** The simplest possible *'fully connected'* neural network can be summarised as a series of layers of single neurons,  $z^{(l)}$ , that multiply a previous input, the input  $x$  or the output from a previous layer,  $a^{(l-1)}$  by weight  $w^{(l)}$ , and adds a bias of  $b^{(l)}$ , where the superscripts ( $l$ ) is the 'layer' of the neuron, of  $L$  total layers, as shown in Equation 2.20. Here, *'fully connected'* means that the outputs of all layers are connected to all the inputs of the next layer, which in the simplest case is only to one other neuron. In order that nonlinear relationships can be learned between the input and the output,  $z^{(l)}$  is passed through a non-linear *'activation function'*,  $\sigma$ , to produce the output for the next layer (see Section 2.5.4.2). This is shown in Equation 2.17b

$$z^{(l)} = w^{(l)} \cdot a^{(l-1)} + b^{(l)} \quad (2.17a)$$

$$a^{(l)} = \sigma(z^{(l)}) \quad (2.17b)$$

This process of propagating inputs through different layers by applying weights and activations continues until reaching the output layer. Then, the loss function,  $\mathcal{L}$ , is applied to the output of the neural network,  $a^{(L)}$ , by comparison to the desired output,  $y$  (i.e.  $\mathcal{L}(a^{(L)}, y)$ ). This function condenses the output of the model into a single scalar value, which describes the overall performance of the network.

This description summarises the feed-forward step of the neural network, which is visualised in Figure 2.50a.

**Backpropagation Step** The network must update all the weights and biases in the network based on the value of the loss. The backpropagation step achieves this goal.

If we consider the outputs of the last output of the network at layer  $L$ ,  $a^L$ , then we may calculate the partial derivative of the loss with respect to the output of the network,  $\frac{\partial \mathcal{L}}{\partial a^{(L)}}$ , which may be calculated numerically. By the chain rule, the derivative for the weight, the bias and the activation input in the previous neuron may also be calculated using a product of partial derivatives if the functions used to produce the output are continuously differentiable.

For instance, the weight of the neuron of the output may be calculated according to Equation 2.18:

$$\frac{\partial \mathcal{L}}{\partial w^{(L)}} = \left( \frac{\partial z^{(L)}}{\partial w^{(L)}} \right) \cdot \left( \frac{\partial a^{(L)}}{\partial z^{(L)}} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial a^{(L)}} \right) = a^{(L-1)} \cdot \sigma' \left( z^{(L)} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial a^{(L)}} \right) \quad (2.18)$$

where all symbols have their previous meanings,  $\sigma'$  is the derivative of the activation function with respect to  $z^{(L)}$ .

The weight can then be updated using a *learning rate*,  $\eta$ , that defines a small step for the weight in that layer to move based on the value of the gradient. While more advanced schemes exist for updating the weights, the basic idea is presented in Equation 2.19:

$$w^{(L)} \rightarrow w^{(L)} - \eta \cdot \left( \frac{\partial \mathcal{L}}{\partial w^{(L)}} \right) \quad (2.19)$$

The bias can be updated similarly.

The partial derivative of the neuron in the penultimate layer,  $\frac{\partial \mathcal{L}}{\partial a^{(L-1)}}$  may also be calculated using the information in the last layer, according to Equation 2.20:

$$\frac{\partial \mathcal{L}}{\partial a^{(L-1)}} = \left( \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \right) \cdot \left( \frac{\partial a^{(L)}}{\partial z^{(L)}} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial a^{(L)}} \right) = w^{(L)} \cdot \sigma' \left( z^{(L)} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial a^{(L)}} \right) \quad (2.20)$$

Since the same rules apply to the previous layer, the derivative in Equation 2.20 can be used to calculate the derivative of the weights and biases of the penultimate layer, which can be used in the layer before, and so on. This recursive operation 'backpropagates' the gradient from the loss backwards, all the way back to the first layer of the network.

The critical step of backpropagation allows the calculation of the updates of all the weights and biases of the network until the first layer of the neural network is reached.

This description summarises backpropagation, the most important feature of how a neural network learns.

**Generalisation to Multiple Neurons Per Layer** In practice, as in the human brain, a neural network will have more than one neuron and receive input from all previous layers. In this case, the neuron  $z_k^{(l)}$  refers to the  $k^{\text{th}}$  neuron in a layer. However, most of the training method for the single neuron case remains the same.

In this case,  $z_k^{(l)}$  becomes a sum of a vector of weights applied as a dot product to all the inputs of the previous layer. There are as many weights in a neuron as inputs into the neuron with a single bias, as shown in Equation 2.21a.  $z_k^{(l)}$  is once again passed through the activation function to produce an activation for the neuron,  $a_k^{(l)}$ , as shown in Equation 2.21b:

$$z_k^{(l)} = \sum_{j=0}^{N_{l-1}-1} w_{jk}^{(l)} \cdot a_j^{(l-1)} + b_k^{(l)} \quad (2.21a)$$

$$a_k^{(l)} = \sigma(z_k^{(l)}) \quad (2.21b)$$

where the additional index  $j$  corresponds to the weight index of one of the activations from the previous layer, and  $N_{l-1}$  is the number of neurons in the previous layer. It should be noted that, for a weighting vector less than the number of input activation values, then the operation becomes a cross-correlation rather than a dot product, which is referred to as '*convolution*' in machine learning and deep learning communities.

The feed-forward step proceeds otherwise in the same way as the single neuron case. The feed-forward part of the multi-neuron case is shown in Figure 2.51a.

In backpropagation, once again starting from the output layer  $L$ , a derivative of each weight in the previous layer neuron must be calculated, which is performed in the same way as for the single neuron case, but for each weight in the neuron:

$$\frac{\partial \mathcal{L}}{\partial w_{jk}^{(l)}} = \left( \frac{\partial z_j^{(L)}}{\partial w_{jk}^{(l)}} \right) \cdot \left( \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial a_j^{(L)}} \right) \quad (2.22)$$

Once again, there is only one bias per neuron and the derivative proceeds as in the one-dimensional case.

However, for the derivative of the activation value of the neuron in the penultimate layer,  $\frac{\partial \mathcal{L}}{\partial a_k^{(L-1)}}$ , there are multiple neurons in layer  $L$  that received input from that neuron. However, the derivatives may simply be added together across the layer  $L$ . Put another way, the total gradient for a neuron in  $L - 1$  combines the gradients of the neurons that used their activation value in the last layer,  $L$ . This is shown in Equation 2.23 and visualised in Figure 2.51:

$$\frac{\partial \mathcal{L}}{\partial a_k^{(L-1)}} = \sum_{k=0}^{N_L-1} \left( \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \right) \cdot \left( \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial a_j^{(L)}} \right) \quad (2.23)$$

Once again, all the weights and biases of previous layers can be calculated through backpropagation. The backpropagation step of the neural network is shown in Figure 2.51b.

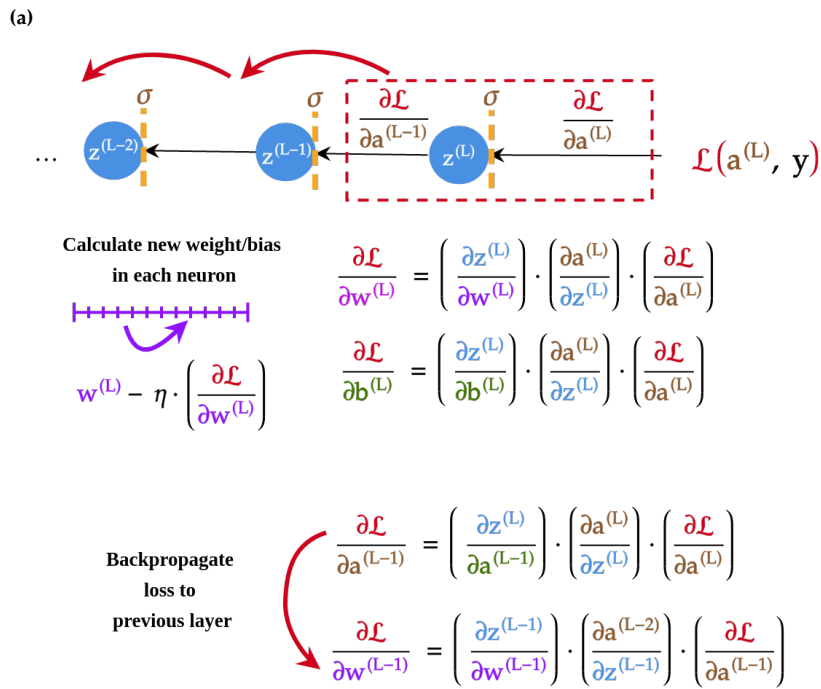
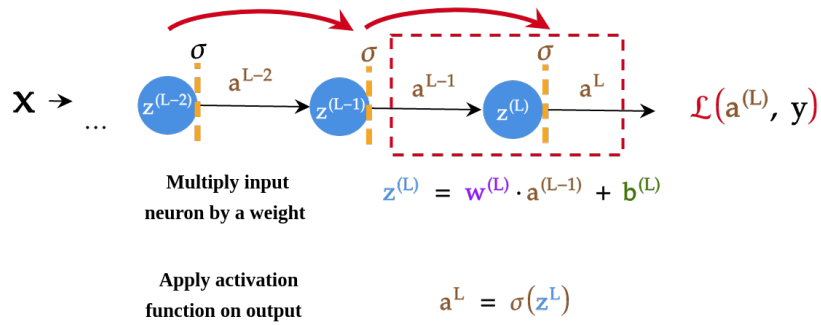
In practice, special software libraries are used for highly optimised automatic differentiation. The two main libraries are TensorFlow [112] and Torch [113].

This section summarises the basic principle of the operation of neural networks.

#### **2.5.4.2 Activation Functions**

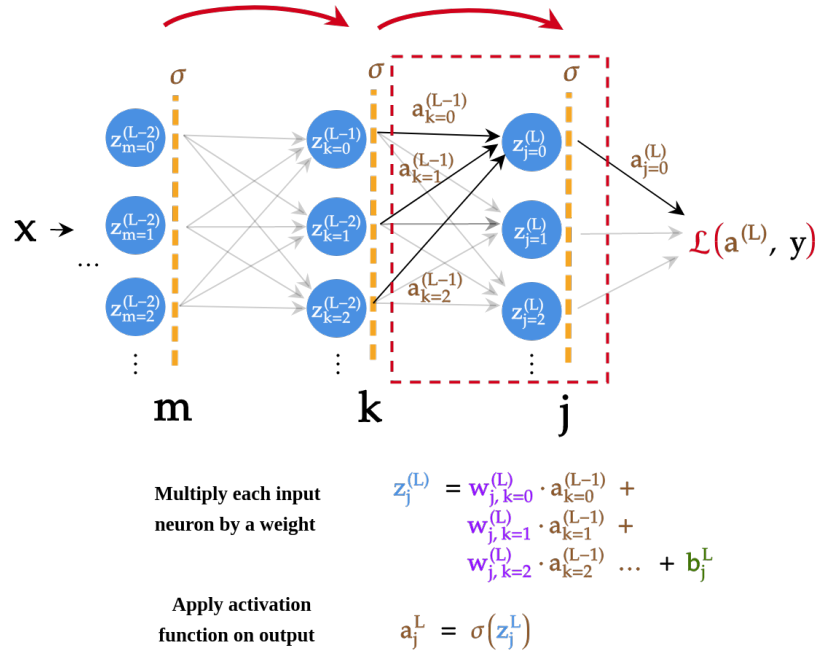
Activation functions are nonlinear operations applied to the outputs of neural networks. These enable more complex output from neural networks by mapping the neuron output to a non-linear space. Non-linear activation functions applied to neural networks have been shown to allow a neural network model to act as a universal function approximator [114].



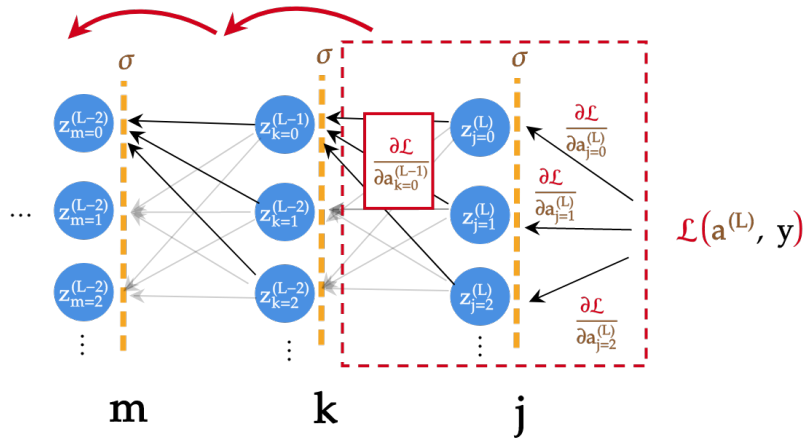


(b)

**Figure 2.50:** Figure 2.50a and Figure 2.50b show the feed-forward and backpropagation steps of a single-neuron neural network. Colour coding is provided for each of the elements of the process: the neuron output ( $z^{(l)}$ ), the weights ( $w^{(l)}$ ), the biases ( $b^{(l)}$ ), the activation function and values ( $\sigma, a^{(l)}$ ) and the loss  $\mathcal{L}$ , are coloured in blue, purple, green, brown and red respectively. The orange lines indicate the application of  $\sigma$  to  $z^{(l)}$ . Black arrows indicate the 'direction' of the flow of information. Red arrows specifically indicate the backpropagation of the loss.  $\eta$  is the learning rate. Relevant partial derivatives are shown with colour coding to illustrate how the chain rule is used to determine the derivatives for the weights, bias and for activation in the previous layer.



(a)



(b)

**Figure 2.51:** Figure 2.51a and Figure 2.51b show the feed-forward and backpropagation steps of a single-neuron neural network. The notation and colour coding is as in Figure 2.50. Additionally,  $j, k$  and  $m$  indicate the indices for neurons in layers  $L, L - 1$  and  $L - 2$ . The feed-forward and backpropagation steps show the specific activation and derivatives contributing to the layer's first neuron.

In this thesis, three activation functions are used, and are highlighted in brief:

**Rectified Linear Unit** The rectified linear unit (ReLU) is defined as the output of a neuron if greater than zero, and zero otherwise:

$$\sigma_{\text{ReLU}}(z_i^{(l)}) = \max(z_i^{(l)}, 0) \quad (2.24)$$

ReLU is used in this thesis as the output activation for AHCAL cell energy, which must always be a positive number.

**Leaky Rectified Linear Unit** The Leaky ReLU tackles the problem of 'vanishing gradients'<sup>41</sup> in deep neural networks by giving negative values a small gradient,  $\epsilon$ , to negative values. In this way, a gradient is defined for all the possible outputs of a neural network except where the neuron outputs 0:

$$\sigma_{\text{LReLU}}(z_i^{(l)}) = \begin{cases} z_i^{(l)} & \text{if } z_i^{(l)} > 0 \\ \epsilon \cdot z_i^{(l)} & \text{otherwise} \end{cases} \quad (2.25)$$

Leaky ReLU activations are used in this thesis's deep neural network models to reduce the effect of vanishing gradients.

**Softmax** The Softmax function takes the output of multiple neurons,  $\mathbf{z}$ , and maps them to a probability for each input neuron:

$$\sigma_{\text{Softmax}}(\mathbf{z}_i^l) = \frac{e^{z_i^{(l)}}}{\sum_{j=0}^{\dim(\mathbf{z})-1} e^{z_j^{(l)}}} \quad \text{for } i = 0, \dots, \dim(\mathbf{z}) - 1$$

$$\text{and } \mathbf{z} = (z_0^{(l)}, \dots, z_{\dim(\mathbf{z})-1}^{(l)}) \in \mathbb{R}^{\dim(\mathbf{z})} \quad (2.26)$$

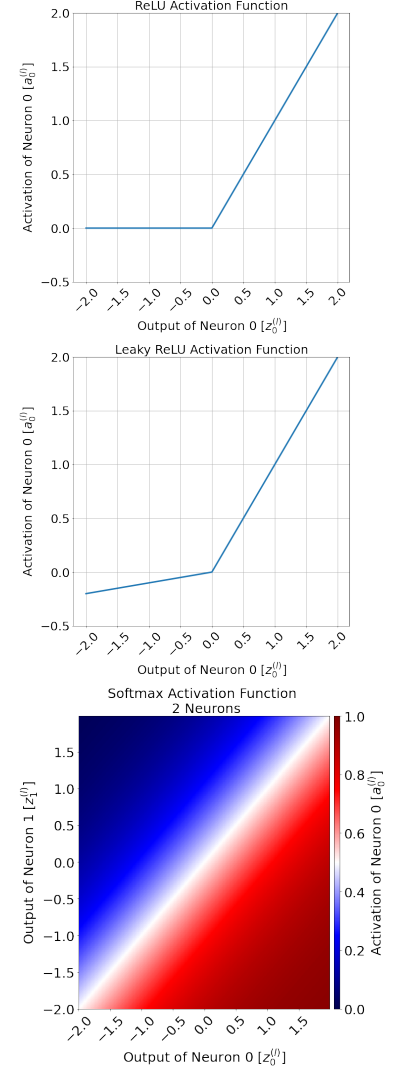
Softmax is used in this thesis as the output activation for fractions of energy in a single AHCAL cell. If two hadron showers are observed simultaneously by the detector, a model to separate them may use a Softmax output layer to produce a fraction for each belonging to each hadron shower.

Plots showing each activation function are shown in Figure 2.52.

### 2.5.4.3 Loss Functions and Optimisation Algorithms

A 'loss function' ( $\mathcal{L}$ ) is an objective function that defines a 'cost' describing the performance of the neural network function  $f$ . The gradient updates are performed relative to the gradient of this function with respect to a desired output,  $y$ . Loss functions in this thesis are tailored specifically for the tasks to be performed and are introduced in the relevant sections.

To accelerate the training of neural networks, optimisation algorithms are often employed in machine learning algorithms. The stochastic



**Figure 2.52:** Activation functions used in this thesis. Top: ReLU. Middle: Leaky ReLU, with a gradient of 0.1 for negative values. Bottom: Softmax over two neurons. The activation of the first neuron is shown on the colour axis as a function shown as a function of the output of the two neurons. The activation of the second neuron is the dual of this function.

41: The vanishing gradient problem is a nuisance effect experienced by deep neural networks. This effect occurs when, during the backpropagation step, the partial derivative of the weights in a neuron becomes so small that no change in the weight occurs. This may result in the failure of the network to update the weights due to 'dead neurons' in the network [115].

optimisation technique employed in this thesis is called ‘*adaptive moment estimation*’ (ADAM) [116]. ADAM employs a more advanced approach compared to the simpler method described in Equation 2.16 for updating the free parameters’ learning rates in the model. It achieves this by estimating the first and second moments of the gradients and computing individual adaptive learning rates accordingly. This is done through the use of two hyperparameters,  $\beta_1, \beta_2 \in [0, 1)$  and control the exponential decay rates of the moving averages for the gradient and its square. All parameters are optimised according to their own individual learning rate in this scheme.

Machine learning algorithms require an approximately consistent dynamic range between the input variables in order for the loss function. This nearly always motivates the parameter space to be chosen carefully.

For AHCAL data, the cells and layers of the calorimeter are already recorded in the unit distance, so the dynamic range of the input is acceptable for machine learning. The energy and time are of a much greater dynamic range, and so are re-scaled according to Section 2.5.1.2.

#### 2.5.4.4 Graph Neural Networks

Graph neural networks are particular classes of neural networks that operate on mathematical constructs called ‘*graphs*’. Graphs are used to describe pairwise relationships between points in space. A graph consists of two components: ‘*vertices*’, corresponding to a measured value in space, and ‘*edges*’, the vectors between vertices. Graph neural networks (GNNs) are designed to exploit data as graphs [117].

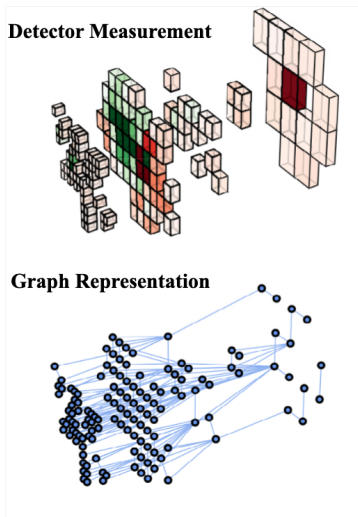
An example of a graph representation of a hadron shower as observed with a highly granular calorimeter array is shown in Figure 2.53.

Graphs are advantageous representations of experimental data for several reasons:

- ▶ *Effective Representation of Complex Data*: Graphs implicitly capture complex relationships between elements of a set of data.
- ▶ *No Grid-Based Structure*: Neural networks typically rely on grid-like structures. A graph, by contrast, places no constraints on the dimensions of the data used to define the graph;
- ▶ *Handling of Sparsity*: a graph may represent sparse data without the requirement that the data be projected to a grid without an artificial ordering scheme;

A summary of the basic structure of a graph neural network is explained in brief.

A graph typically consists of vertices, denoted  $V$ , which could be energy deposits in a calorimeter, and  $\mathcal{E}$ , the edges between them (i.e. spatial distances). A basic graph convolution operation shares act both transform the graph and share inform network consists of two main components: an ‘*update*’ function, defined  $\phi$ , and an ‘*pooling*’ function, defined  $\rho$  [118]. The update function most commonly takes the form of a fully connected neural network (FCN), also known as a ‘*multi-layer perceptron*.’ The pooling



**Figure 2.53:** Top: a representation of a hadron shower, measured by highly granular calorimeters. Each voxel represents an energy deposit in a set of calorimeters. Bottom: the same hadron shower, rendered as a graph. Adapted from [118]

operation involves one or more operations that condense information about the graph, such as maximum, sum, mean, variance, skew, an FCN, or combinations thereof. A graph neural network operates by pooling information about  $V$  and  $\mathcal{E}$  with  $\rho$  and using that information as the input into  $\phi$  to update the graph. This concept is known as ‘*message passing*’, because pooled information about vertices is used to inform the update of edges, and vice versa. Pooled information (‘*messages*’) about  $V$  can be used to update  $\mathcal{E}$  and vice versa in any particular configuration. In this way, information is diffused throughout a small graph region. Through successive graph ‘convolutions,’ nodes can be updated with information about the entire graph.

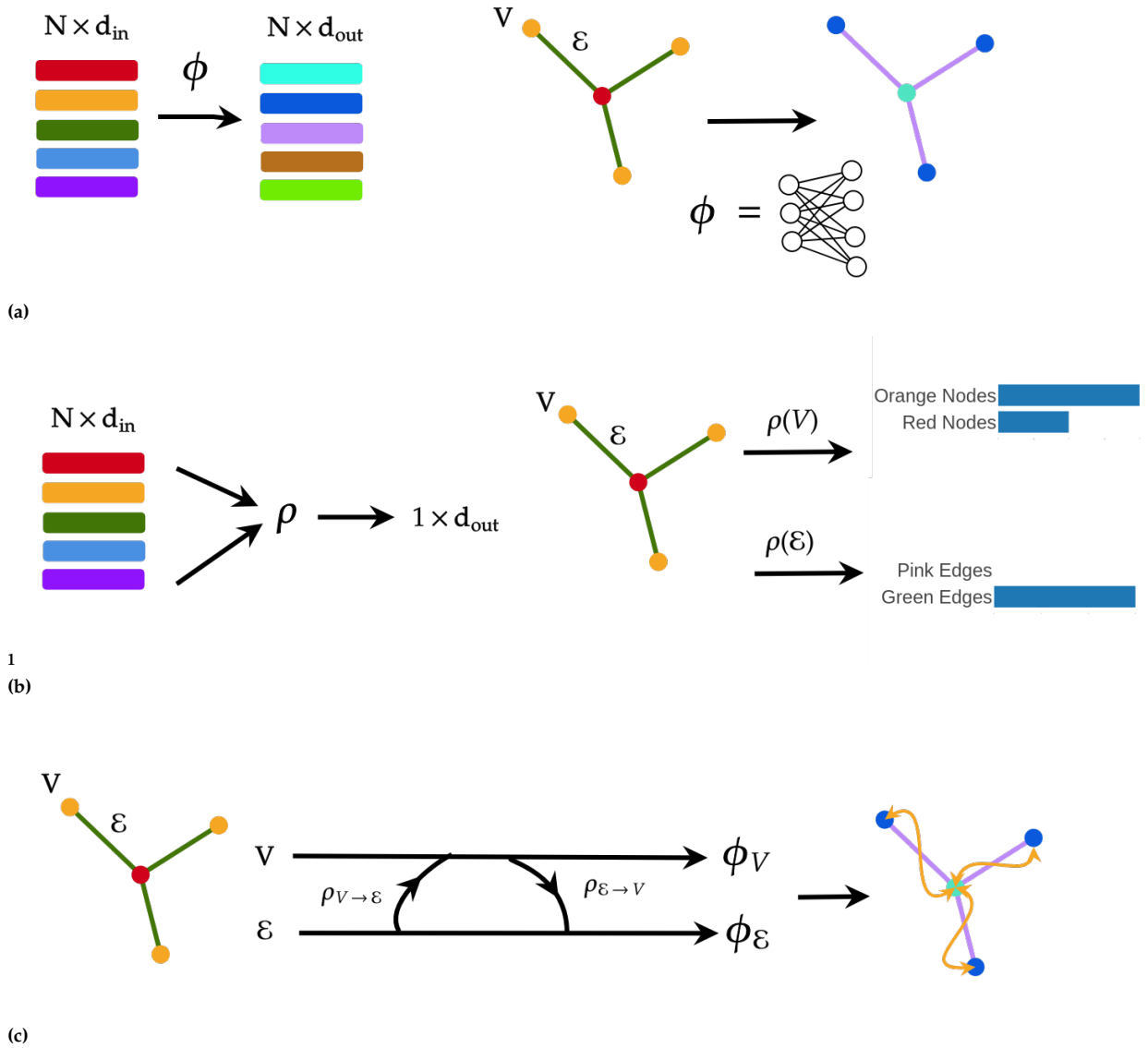
The concept of an update function, pooling function and graph convolution are summarised in Figure 2.54.

One application of GNNs to high energy physics is in the form of event reconstruction, specifically in clustering and classification tasks [117]. Highly relevant for this thesis is that graph neural networks have been shown to outperform traditional grid-based networks at distinguishing the energy deposits of two individual hadron showers observed by the same highly-granular calorimeter [120], similarly to Pandora PFA. This study demonstrated the benefit of graph neural networks applied to the task of reducing confusion for a four-dimensional, highly granular hadron calorimeter.

Referencing the discussions of Section 2.5.1.5, the effectiveness of a five-dimensional (i.e. temporal) clustering has yet to be demonstrated. The success of graph neural networks in this study motivates applying these models to five-dimensional calorimetry by clustering in time. A graph-neural network trained with time will illustrate the benefits of using temporal information for clustering and how this manifests.

Furthermore, the same principles of graph neural networks can be used to study the local energy density of hadron showers, exploiting the efficient representation of graphs to perform software compensation and detector calibration, which local-neighbourhood graphs may achieve. The message-passing properties of GNNs may also be used for software compensation studies.

These observations motivate research into the use of graph neural networks for both shower separation and software compensation.



**Figure 2.54:** Diagrams illustrating the components of a graph neural network. Figure 2.54a shows the ‘update’ function  $\phi$ .  $\phi$  acts on a graph, composed of vertices and edges, denoted by  $V$  and  $\mathcal{E}$ , respectively, indicated by points connected with lines, with a certain fixed size, indicated by the coloured bars, and applies the same function to the elements of the set to achieve an updated representation of the input. Figure 2.54b shows a pooling function,  $\rho$ , which performs an aggregation operation on the graph about its properties. In the example shown,  $\rho$  counts the colours of the vertices (nodes) and edges, indicated by the bar charts. Figure 2.54c indicates a graph convolution operation. The black lines indicate the operations used to modify the graph. Message passing is demonstrated by  $\rho_{V \rightarrow \mathcal{E}}$  and  $\rho_{\mathcal{E} \rightarrow V}$ , which act on the vertices (edges) and inform the update of the edges (vertices) by pooling information about them with  $\rho[\mathcal{E}]$  ( $\rho_V$ ) and using it to update the vertices (edges) with  $\phi_V$  ( $\phi_{\mathcal{E}}$ ). The result is a modified graph, where the orange lines indicate the diffusion of information throughout the graph. All diagrams were modified from [118] and [119].

# LightSimtastic: A Tool For SiPM Simulation

# 3

## Contributions by Author

*The author implemented a program for simulating SiPM charge spectra in Python and transients under the supervision of and developed by Prof. Dr. Robert Klanner. This chapter is an abridged version of the paper of [96], to which the author is a co-author.*

## Original Abstract

A Monte Carlo program is presented which simulates the response of SiPMs in the linear regime in which saturation effects can be ignored. Inputs to the program are the mean number and the time distribution of Geiger discharges from photons, as well as the dark-count rate. For every primary Geiger discharge, correlated Geiger discharges due to prompt and delayed cross-talk and after-pulses are simulated, and a table, called Geiger Array, of the amplitudes and times of all Geiger discharges in a specified time window is generated.

A number of different physics-based models and statistical treatments for the simulation of correlated Geiger discharges can be selected. The Geiger Arrays for many events together with different options for the pulse shapes of single Geiger discharges can be used to simulate charge spectra, as measured by a current-integrating charge-to-digital converter. In addition, current transients convolved with electronics response functions, as recorded by a digital oscilloscope, and time-difference distributions can be simulated. The program can be used to compare simulation results with different assumptions to experimental data, and thus find out which model is the most appropriate for a given SiPM, optimise the operating conditions and readout for a given application, and test programs for extracting SiPM parameters from experimental data. A Python version of the program is available on request.

3.1 Overview Of Study . . . . .	63
3.2 Method And Tools . . . . .	64
Detector Response Mode and Free Parameters . . . . .	64
3.3 Results . . . . .	70
3.4 Conclusion . . . . .	73

## 3.1 Overview Of Study

Silicon Photomultipliers (SiPMs) are arrays of single-photon avalanche diodes (SPADs) operated above the breakdown voltage. Their single-photon detection capability and their high photon-detection efficiency (PDE) have led to many applications in industry and in science from astrophysics over high-energy physics to nuclear medical imaging.

SiPMs can operate in one of two states:

- ▶ *Low-occupancy*: where the number of Geiger discharges is small compared to the number of pixels of the SiPM within an integration window similar to that of the pulse-shape of the SiPM;

- *High-occupancy*: where the number of Geiger discharges reaches or possibly exceeds the number of pixels of the SiPM.

The former case is studied in this chapter. Specifically, in this regime, it is assumed that the response of the SiPM is to good approximation linear (at least one fired pixel per photon initiating a Geiger avalanche) and that correlations between Geiger discharges from noise such as dark counts and after-pulses are negligible.

A simulation program for SiPMs could be used for several applications. Relevant to this thesis, the program is used to verify the Peak0Tron characterisation program of Chapter 4 and to evaluate the effect of different light sources, parameterisations for afterpulse and cross-talk and readout on the response of SiPMs to light.

Several simulation tools exist for low-occupancy SiPMs. A Monte Carlo program simulating the multiplication process responsible for Geiger discharges is presented in [121]. Programs that simulate the transients' shape for different options for the readout electronics are discussed in [122, 123] and associated references. Monte Carlo programs focused on the light readout from scintillators with SiPMs are studied in [124–126], the latter of which has been implemented in Geant4 [127]. The study of [128] focuses on optimising the time resolution of SiPMs for positron emission tomography (PET) scanners.

However, these models do not offer the possibility to modify the pulse shapes of light and the SiPM discharge, the contribution of correlated noise sources such as dark counts or afterpulses, or the readout mode or electronic response function. Therefore, a simple, flexible simulation tool called *LightSimtastic* was developed to perform simulations of SiPM charge spectra and transients.

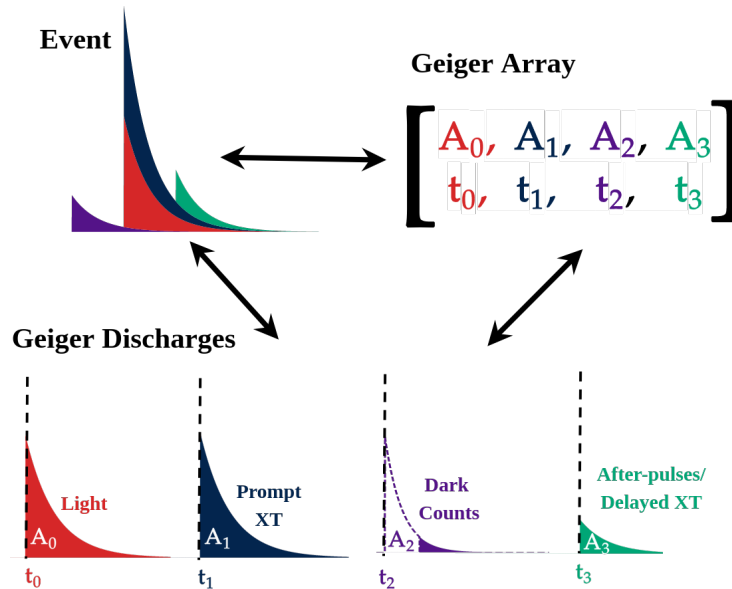
This study first introduces and describes the free parameters and methods of the program, including the concept of the '*Geiger Array*'. Then, example charge spectra and transients under different conditions are provided.

## 3.2 Method And Tools

### 3.2.1 Detector Response Mode and Free Parameters

The detector response model used in *LightSimtastic* is superficially similar to the one used in *Peak0Tron*, which is discussed in depth in Chapter 4. A critical difference is that the detector response model of each Geiger discharge and its time are individually simulated, and the possibility to modify the types of distribution used for the Geiger discharge light, cross-talk and afterpulse distributions. By contrast, in *Peak0Tron*, the model is presented as a probability density function. A general overview is provided in this section relevant to the operation of the program, and references to the more detailed discussion of Section 4.2.1 is provided as necessary.





**Figure 3.1:** Diagram illustrating the concept of the Geiger Array in LightSintastic. An 'event' consists of some (or no) Geiger discharges. The Geiger Array describes the amplitude and time of a discharge. From this information, and parameters provided to the model, transients and total discharge amplitudes can be reconstructed.

### 3.2.1.1 Geiger Array

LightSintastic is built around the concept of a 'Geiger Array'. The LightSintastic program simulates events, which correspond to a period during which the signal from the SiPM is read out (the 'gate'). For each event, the Geiger Array holds both the amplitude of a discharge and the time the discharge took place, both simulated by the program. Each discharge has a user-selected pulse shape assigned to it. The Geiger array can therefore be used to generate both charge spectra and transients as observed on an oscilloscope.

Free parameters and probability distributions of the LightSintastic detector response model are shown in Table 3.1. Table 3.1a shows the free parameters of the detector response model, Table 3.1 shows the probability distributions available to the user. The physics motivation for each element is discussed in brief.

A diagram illustrating the concept of the Geiger Array is shown in Figure 3.1

### 3.2.1.2 Time And Integration Gate ( $t_0, t_{\text{gate}}, t_{\text{start}}$ )

In LightSintastic, the time interval  $-t_0 \leq t \leq t_{\text{start}} + t_{\text{gate}}$  is modelled by the program, where  $t$  is the time of the Geiger discharge,  $-t_0$  is the minimum time at which Geiger discharges occur in the simulation,  $t_{\text{start}}$  is the start of an integration gate with length  $t_{\text{gate}}$ .  $n_t$  is the bins used for transient simulations.

Group	Symbol	Description	Default
pulse shape	$\tau$	time constant of slow component	20 ns
	$\tau_f$	time constant of fast component	1.5 ns
	$r_f$	fraction fast component	0.2
gate and time parameters	$t_{\text{start}}$	gate start	-5 ns
	$t_{\text{gate}}$	gate length	100 ns
	$-t_0$	start time simulation	-100 ns
	$n_t$	number time bins of transient	1000
primary Geiger discharges light	$N_{\gamma G}$	mean number of discharges	3
	$n_{\gamma G}$	actual number of discharges	Table 3.1b
	$A_{i\gamma G}$	amplitudes of discharges	local
	$t_{i\gamma G}$	times of discharges	Table 3.1b
primary Geiger discharges dark counts	$D\bar{C}R$	dark-count rate	500kHz
	$n_{\text{dark}}$	actual number of dark counts	Table 3.1b
	$A_{i\text{dark}}$	amplitudes of discharges	local
	$t_{i\text{dark}}$	times of discharges	Table 3.1b
prompt cross-talk	$p_{pXT}$	probability	0.1
	$n_{pXT}$	actual number of discharges	Table 3.1b
	$A_{ipXT}$	amplitudes of discharges	local
	$t_{ipXT}$	times of discharges	Table 3.1b
delayed cross-talk	$p_{dXT}$	probability	0.1
	$\tau_{dXT}$	time constant	25 ns
	$n_{dXT}$	actual number of discharges	Table 3.1b
	$A_{idXT}$	amplitudes of discharges	local
after-pulses	$t_{idXT}$	times of discharges	Table 3.1b
	$p_{AP}$	probability	0.15
	$\tau_{AP}$	time constant	7.5 ns
	$\tau_{\text{rec}}$	Geiger probability recovery time	20 ns
noise and electronics response function	$n_{AP}$	actual number of discharges	Table 3.1b
	$A_{iAP}$	amplitudes of discharges	Table 3.1b
	$t_{iAP}$	times of discharges	Table 3.1b
	$\sigma_1$	gain fluctuations	0.05
Geiger array	$\sigma_0$	electronics noise Q-measurement	0.075
	$\sigma_I$	transient current noise	0.05
	$R(t)$	electronics response function	Gauss
	$\sigma_R$	rms width of Gaussian $R(t)$	1.5 ns
Geiger array	$n_G$	total number of Geiger discharges	output
	$A_{iG}$	amplitudes of Geiger discharges	output
	$t_{iG}$	times of Geiger discharges	output

(a)

Group	Symbol	Options	Default
Geiger discharge light	$n_{\gamma G}$	Poisson, Gauss, $\delta$ -function	Poisson
	$t_{i\gamma G}$	Poisson, Gauss, 2 Exp., Double exp.	Gauss
prompt cross-talk	$n_{pXT}$	Binomial, Poisson, Borel	Binomial
delayed cross-talk	$n_{dXT}$	Binomial, Poisson, Borel	Borel
	$t_{idXT}$	Exponential	Exponential
after-pulses	$n_{dXT}$	Binomial, Poisson	Binomial
	$A_{iAP}$	Exponentials; see text	Exponential
	$t_{iAP}$	Exponential	Exponential

(b)

**Table 3.1:** Table 3.1a shows parameters and variables used in the simulation program. Default gives the values of input parameters which are implemented as default in the Python program, which the user can modify. References to Table 3.1b are provided where appropriate. Local and output indicate that the variables are local or output results, respectively. The units used in the program are in s for time and Hz for rates. The amplitudes,  $A$ , are normalised to 1. The  $\sigma$  parameters are also normalised to 1, meaning they represent a fraction of the gain of the SiPM. Table 3.1b shows the available probability distributions to describe Geiger discharges and default values. Further discussion on key SiPM parameters can be found in Section 2.5.2.

### 3.2.1.3 Gain Fluctuations ( $\sigma_1$ )

The gain of a SiPM can fluctuate. The main reason is that Geiger avalanches are statistical processes, and fluctuations are expected from discharge to discharge. Other reasons for gain fluctuations can be pixel-wise capacitance variations or electric fields inside a single pixel. The reasons are discussed in Section 2.5.2. To simulate this property of Geiger discharges, their amplitudes are multiplied with a Gaussian-distributed random number with width  $\sigma_1$ . This parameter is also discussed in Section 4.2.1.

### 3.2.1.4 Number, Times and Amplitudes of Photon-Induced Geiger Discharges ( $N_{\gamma G}$ , $n_{\gamma G}$ , $A_{i\gamma G}$ , $t_{i\gamma G}$ )

Photons incident on the SiPM has a probability and time distribution depending on their source.

- ▶ *Number of Geiger Discharges:*  $N_{\gamma G}$  is the mean number of Geiger discharges induced by photons from incident light on the SiPM. The number of discharges in an event  $n_{\gamma G}$  can be either constant or sampled from a Poisson or Gaussian distribution. The Poisson distribution describes pulsed light sources like LED or laser. The Gaussian distribution is more appropriate for light from a scintillator;
- ▶ *Times of Geiger Discharges:*  $t_{i\gamma G}$  are sampled for each discharge  $i_{\gamma G} \in [0, n_{\gamma G}]$  from one of three distributions: a Gaussian which describes light from LED or laser, a bi-exponential distribution  $(e^{-\lambda_1 t} - e^{-\lambda_2 t}) \cdot \lambda_1 / (\lambda_2 - \lambda_1)$  which describes the distributions of scintillator light read out from wavelength-shifters;
- ▶ *Amplitudes of Geiger Discharges:* amplitudes  $A_{i\gamma G}$  for each discharge  $i_{\gamma G} \in [0, n_{\gamma G}]$  are normalised to one unit of gain, smeared by  $\sigma_1$ .

### 3.2.1.5 Number, Times and Amplitudes of Dark Counts ( $DCR$ , $n_{\text{dark}}$ , $A_{i\text{dark}}$ , $t_{i\text{dark}}$ )

Dark counts are, in general, discharges by thermally induced charge carriers. More discussion is given in Section 2.5.2.

- ▶ *Number of Geiger Discharges:* the number of dark counts is modelled as a Poisson distribution with rate parameter  $\mu_{\text{dark}} = DCR \cdot (t_0 + t_{\text{start}} + t_{\text{gate}})$ . In the limit  $t_{\text{start}} \rightarrow 0$ , the definition provided in Section 4.2.1 is recovered;
- ▶ *Times of Geiger Discharges:* dark counts are uniformly distributed in time and therefore, for each dark count  $i_{\text{dark}} \in [0, n_{\text{dark}}]$ , the uniform distribution is sampled from  $-t_0$  to  $t_{\text{start}} + t_{\text{gate}}$ .
- ▶ *Amplitudes of Geiger Discharges:* similarly as for light-induced discharges, Amplitudes  $A_{i\text{dark}}$  for each discharge  $i_{\text{dark}} \in [0, n_{\text{dark}}]$  are normalised to one unit of gain, smeared by  $\sigma_1$ .

### 3.2.1.6 Number, Times and Amplitudes of Prompt Cross-Talk ( $p_{\text{pXT}}$ , $n_{\text{pXT}}$ , $A_{i\text{pXT}}$ , $t_{i\text{pXT}}$ )

Prompt cross-talk occurs by photons produced in the Geiger avalanche in the depletion region of another pixel (i.e. from bremsstrahlung by accelerated charge carriers in the high electric field region of the pixel) and are observed in another pixel (see Section 2.5.2).

- ▶ *Number of Geiger Discharges:* to sample  $n_{\text{pXT}}$ , three probability distributions are implemented: Binomial, Poisson and Borel characterised by  $p_{\text{pXT}}$ .<sup>1</sup>

1: The Borel distribution, which describes a branching process, whereby the number of new discharges caused by ongoing discharges is distributed as a Poisson distribution with a mean number of discharges,  $\lambda \in [0, 1]$  [129]. The probability distribution of the combined discharges from light and prompt cross-talk is called the *Generalised Poisson* distribution and is introduced in Section 4.2.2.

- ▶ *Times of Geiger Discharges*: Geiger discharges are, by definition, prompt because the photons travel at light speed ( $c = 3 \times 10^5 \mu\text{m ns}^{-1}$ , and the pixel size is of the order  $30 \mu\text{m}$ ). Therefore, it can be considered to occur at the same time as the primary Geiger discharge.
- ▶ *Amplitudes of Geiger Discharges*: similarly as for light-induced discharges, Amplitudes  $A_{i_{\text{pXT}}}$  for each discharge  $i_{\text{pXT}} \in [0, n_{\text{pXT}}]$  are normalised to one unit of gain, smeared by  $\sigma_1$ .

### 3.2.1.7 Number, Times and Amplitudes of Delayed Cross-Talk ( $p_{\text{dXT}}$ , $n_{\text{dXT}}$ , $A_{i_{\text{dXT}}}$ , $t_{i_{\text{dXT}}}$ )

Delayed cross-talk refers to photons produced in the Geiger avalanche in the non-depleted region of another pixel. By contrast to prompt cross-talk, the charge carriers must first drift to the depletion region before a Geiger discharge is initiated, meaning that it is delayed compared to the primary Geiger discharge. See Section 2.5.2 for details). Delayed cross-talk and after-pulses, discussed in the next subsection, are indistinguishable if they occur in the same pixel.

- ▶ *Number of Geiger Discharges*: to sample  $n_{\text{dXT}}$ , as for prompt cross-talk, three probability distributions are implemented: Binomial, Poisson and Borel, characterised by  $p_{\text{dXT}}$ .
- ▶ *Times of Geiger Discharges*: to model the diffusion of the charge carriers through the non-depleted region, for each  $i_{\text{dXT}} \in [0, n_{\text{dXT}}]$ , and an additional random number is added to the time of the original discharge of  $e^{-t/\tau_{\text{dXT}}}/\tau_{\text{dXT}} \cdot \Theta(t)$ , where  $\Theta(t)$  is the Heaviside step function <sup>2</sup> ;
- ▶ *Amplitudes of Geiger Discharges*: similarly as for light-induced discharges, Amplitudes  $A_{i_{\text{dXT}}}$  for each discharge  $i_{\text{dXT}} \in [0, n_{\text{dXT}}]$  are normalised to one unit of gain, smeared by  $\sigma_1$ .

2: This is not an accurate description of the diffusion process, and is not considered accurate in the original publication [96].

### 3.2.1.8 Number, Times and Amplitudes of After-pulses ( $p_{\text{Ap}}$ , $n_{\text{Ap}}$ , $A_{i_{\text{Ap}}}$ , $t_{i_{\text{Ap}}}$ )

After-pulses are discharges caused by charge carriers being trapped by defects in the silicon crystal, which contribute additional energy levels in the silicon band gap. The charge carriers are then released, potentially resulting in another delayed discharge, with a probability depending on how much the pixel has recovered (see Section 2.5.2). Additional discussion of the afterpulse model is provided in Section 4.2.3.

- ▶ *Number of Geiger Discharges*: the probability of an afterpulse taking place is time-dependent because it depends on the recharge state of the SiPM. Physically, an afterpulse is much less likely to occur in a pixel if it has not recovered sufficiently from the primary Geiger avalanche. The afterpulse probability is sampled from  $p_{\text{Ap}} \cdot (1 - e^{-t/\tau_{\text{rec}}}) \cdot \Theta(t)$ , where  $\Theta(t)$  is the Heaviside step function. In other words,  $n_{\text{Ap}}$  is conditioned on the distribution of  $t_{i_{\text{Ap}}}$ . Experimental justification for the parameterisation  $(1 - e^{-t/\tau_{\text{rec}}})$ ,

which describes the Geiger discharge probability as the pixel recovers, is provided in Appendix Section 8.1.0.1. Equation 4.11 gives the total afterpulse probability density in the gate.

- ▶ *Times of Geiger Discharges*: to model the de-trapping of charge carriers, for each  $i_{Ap} \in [0, n_{Ap}]$ . A random number is added to the time of the original discharge drawn from an exponential distribution  $e^{-t/\tau_{Ap}}/\tau_{Ap} \cdot \Theta(t)$ , where  $\Theta(t)$  is again the Heaviside step function.
- ▶ *Amplitudes of Geiger Discharges*: the amplitude of an afterpulse is based on the recovery of the voltage over the pixel from  $V_b$  to  $V_{off}$ , which occurs with a time constant  $\tau^3$ . The amplitude is given by  $(1 - e^{-t/\tau}) \cdot \Theta(t)$ , and are smeared by  $\sigma_1$ .

Each of the discharges and times in Section 3.2.1.4-3.2.1.8 are added to the Geiger Array. All the discharges occurring during the event are included in the Geiger Array.

### 3.2.1.9 Pulse Shape Model

Each pulse is modelled the sum of two exponentials, given by Equation 3.1:

$$I_1(t) = \left( \frac{1 - r_f}{\tau} \cdot e^{-t/\tau_s} + \frac{r_f}{\tau_f} \cdot e^{-t/\tau_f} \right) \cdot \Theta(t) \quad (3.1)$$

where  $\tau$  and  $\tau_f$  are the time constants of the slow and fast component of the SiPM pulse, and  $r_f$  is the fraction of the pulse integral belonging to the fast component.  $\Theta(t)$  is again the Heaviside step function. The pulse shape is normalised to one.

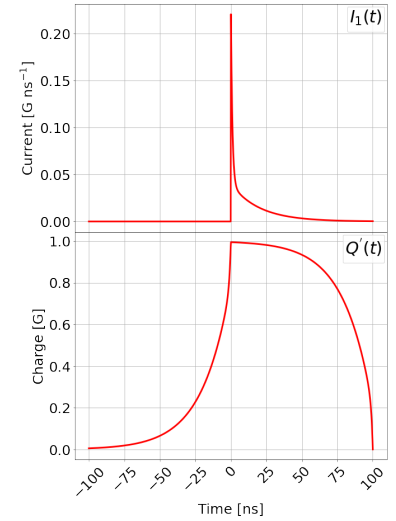
### 3.2.1.10 Charge of The Event

The charge for each event is constructed using Equation 3.2

$$Q' = \sum_{iG=1}^{n_G} A_{iG} \cdot \int_{t_{start}}^{t_{start} + t_{gate}} I_1(t - t_{iG}) dt \quad (3.2)$$

where  $Q'$  is the charge,  $n_G$  are the total number of Geiger discharges in the event ( $n_G = n_{\gamma G} + n_{dark} + n_{pXT} + n_{dXT} + n_{Ap}$ ),  $A_{iG}$  and  $t_{iG}$  are the amplitude and time of each Geiger discharge, and  $I_1(t)$  is from Equation 3.1. A random variable from a Gaussian centred at 0 with width of  $\sigma_0$  is added to simulate electronic noise. The distribution of  $Q'$  for many events simulates a typical charge spectrum measured from experimental data.

3: This 'recovery' is not the same as the probability of the Geiger discharge, which has a time constant  $\tau$  and is typically shorter than  $\tau$  (see Appendix Section 8.1.0.1)



**Figure 3.2:** Plots of pulse shape and integrated charge as a function of time, of Equation 3.1 and 3.2 using the input values are shown in Table 4.3, except for the fast component of the pulse, which is set to a  $r_f = 0.2$  and  $\tau_f = 11$  ns and a gate length of  $t_{gate} = 100$  ns, for demonstration purposes. Note the bottom plot ( $Q'$ ) is shifted in time compared to the upper plot ( $I_1$ ).

### 3.2.1.11 Transient of The Event, Electronics Response Function and Width ( $\sigma_I, \sigma_R, R(t)$ )

Transients may similarly be simulated according to Equation 3.3

$$I'(t) = \sum_{iG=1}^{n_G} A_{iG} \cdot I_1(t - t_{iG}) \quad (3.3)$$

with the same notation as Equation 3.2. The distribution is then convolved with the electronics response function  $R(t)$ , which is a Gaussian with width  $\sigma_R$  by default.

The entire program is developed to be reliant on no more than four external common Python libraries: `numpy` [130], `scipy` [131], `pandas` [132] and `matplotlib` [133].

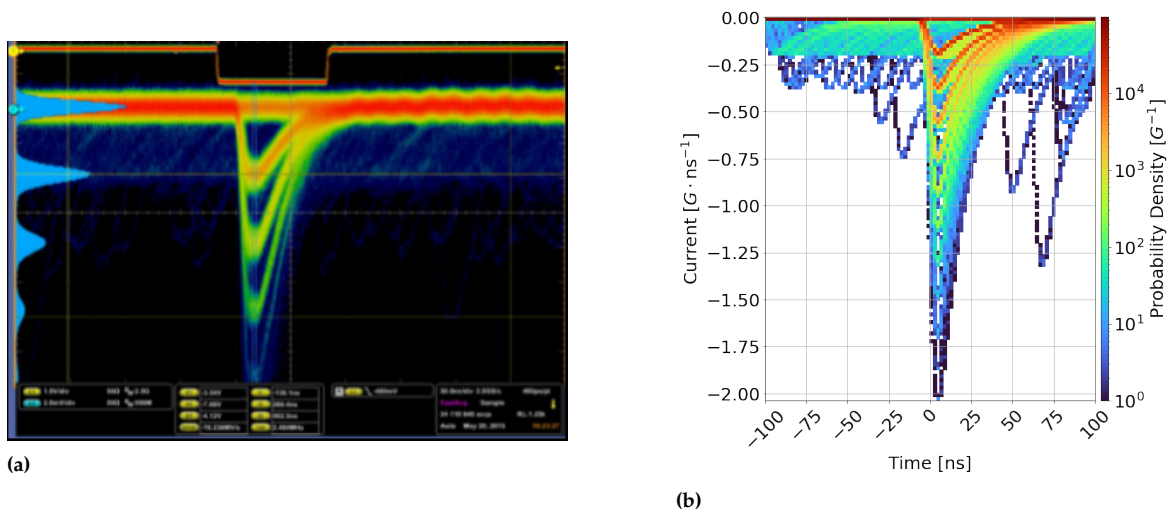
## 3.3 Results

A qualitative comparison between transients from the CAEN educational kit operated at 31 V [134] and the transients produced by `LightSimtastic` are shown in Figure 3.3. Scans through the input parameters of Table 3.1a, which are later used in Chapter 4, are shown in Figure 3.4.

#### Summary Comment to Figures 3.3 and 3.4

The `LightSimtastic` program can produce transients similar to those observed in experimental data, producing a wide range of charge spectra depending on the input parameters.

### Comparison of Transients Of LightSimgastic With Transient of CAEN Kit

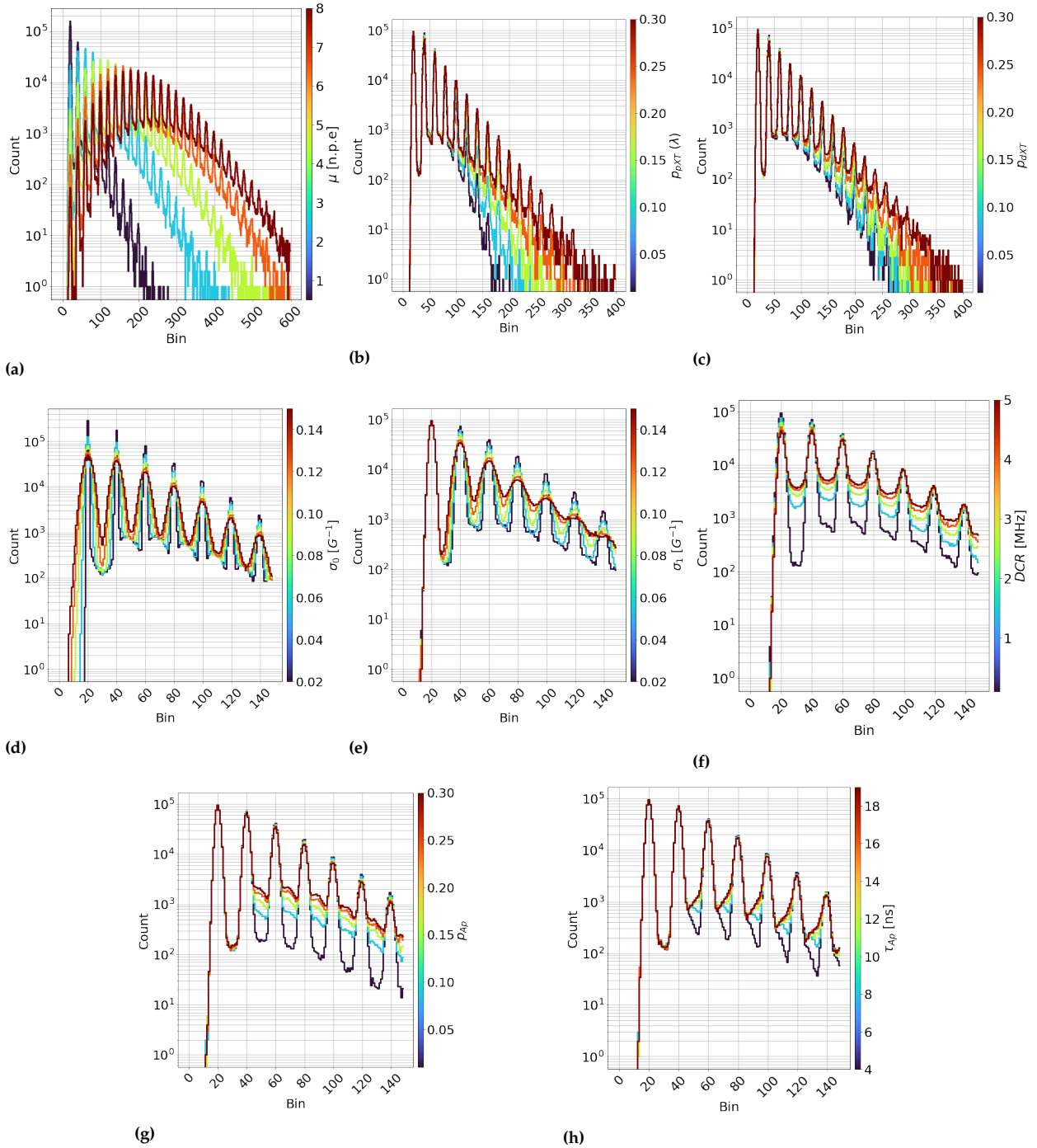


**Figure 3.3:** Qualitative comparison of the transients from experimental data and `LightSimgastic` simulation. Figure 3.3a shows experimental data recorded with the CAEN setup [134], which has a bandwidth of 125 MHz operated at 31 V. Figure 3.3b shows the same pulses, produced under baseline conditions outlined in Table 4.3, albeit with the  $\sigma_R$  of 3 ns. Figure 3.3b shows the probability density of the currents obtained for  $1 \times 10^4$  events.

#### Comments:

- ▶ Each band of each histogram corresponds to an additional Geiger discharge and is well distinguished.
- ▶ Afterpulses and dark counts are visible in both figures.

## Scans of Simulated Charge Spectra of LightSimtastic As a Function Of The Parameters Of Table 3.1a



**Figure 3.4:** Scans of simulated charge spectra produced with LightSimtastic. Figures 3.4a-3.4h show scans of the parameters relative to the baseline shown in Table 4.3 for  $\mu$ ,  $p_{pXT}$  ( $\lambda$ ),  $p_{dXT}$  (using  $\tau_{dXT} = 25$  ns and scanned over the same range as  $p_{pXT}$ ),  $\sigma_0$ ,  $\sigma_1$ ,  $DCR$ ,  $p_{Ap}$  and  $\tau_{Ap}$ , in that order. Each simulation has  $1 \times 10^6$  events and is presented in units of bin, with bin width =  $0.05G$ , where  $G$  is gain. The colour axis indicates the value of the changed variable, with red to blue indicating the low to high values of that variable. The  $x$ -scales are chosen to illustrate the influence of the parameter best.

**Comments:**

- Overall, the simulation tool is demonstrated to produce a wide range of possible SiPM charge spectra shapes.



### 3.4 Conclusion

A Monte Carlo simulation program for SiPM transients and charge spectra has been implemented in Python. The flexible program allows flexible modelling of charge spectra based on various SiPM parameters and probability distributions. The program can be used to compare experimental data with simulation, evaluate the influence on SiPM response under different operating conditions and readout schemes and the performance of methods to characterise SiPMs, as discussed in Chapter 4.



# Peak0Tron: A Tool For SiPM Characterisation

# 4

## Contributions By The Author

The detector response model for SiPMs of [97] was implemented as a Python class for automated SiPM characterisation. A pre-fitting algorithm to produce estimates for the fit and the implementation of the afterpulse model of [135] into the detector response model was developed under the supervision of Prof. Dr. Robert Klanner. The author performed all validation on simulation and experimental data. This chapter is presented from [136], for which the author of the thesis is also the main author, with minor modifications and with a section describing the theory of the model, which is based on the Appendix of that paper, which the co-authors of that paper edited.

## Original Abstract

A Python program has been developed which fits a published detector-response model to SiPM charge spectra to characterise SiPMs. Spectra for SiPMs illuminated by low-intensity pulsed light with a Poisson-distributed number of photons and a time spread of order nanoseconds or less can be analysed. The entire charge spectra, including the intervals in-between the photoelectron peaks, are fitted, which allows determining, in addition to the mean number of detected photons, gain, gain spread, prompt cross-talk, pedestal, and electronics noise, the dark-count rate as well as the probability and time constant of after-pulses. The starting values of the fit parameters are extracted from the charge spectra.

The program performance has been evaluated using simulated charge spectra with the different SiPM parameters varied in a wide range. By analysing 100 simulated spectra for every parameter set, the biases and statistical uncertainties of the individual parameters have been determined. It is found that the parameters are precisely determined and that the entire spectra are well described, in most cases with a  $\chi^2/\text{NDF}$  close to 1. In addition, measured spectra for two types of SiPMs for a wide range of over-voltages have been analysed. The program achieves mostly a good description of the spectra, and the parameters determined agree with the values from the producers and expectations.

The program can be used for detailed analyses of single spectra, but, as it is compatible with the native Python multiprocessing module, also for the automatic characterisation of large samples of SiPMs.

4.1 Overview of Study . . . . .	75
4.2 Theory . . . . .	76
Detector Response Model And Free Parameters . . . . .	77
Generalised Poisson Distribution . . . . .	77
Model for After-pulses . . . . .	79
Model for Photon-Induced Discharges . . . . .	80
Dark Count Model . . . . .	81
4.3 Methods And Tools . . . . .	83
Model Input Parameters . . . . .	83
Data Preparation . . . . .	83
Effective Gain using the Fourier Transform ( $C_{\text{FFT}}^*$ ) . . . . .	83
Peak Finding . . . . .	84
Determination of the Input Parameters . . . . .	87
Implementation of the Fit . . . . .	89
4.4 Results . . . . .	90
Validation of Peak0Tron with Simulated Spectra . . . . .	90
Fits to Experimental Data . . . . .	102
Requirements and Limitations of Fits with Peak0Tron . . . . .	112
4.5 Conclusion . . . . .	112

## 4.1 Overview of Study

Silicon Photomultipliers (SiPMs) are arrays of single-photon avalanche diodes (SPADs) operated above the breakdown voltage. Their single-

photon detection capability and high photon-detection efficiency (PDE) have led to many applications in industry and science, from astrophysics over high-energy physics to nuclear medical imaging.

Different methods have been developed to characterise SiPMs [91, 95, 122]. This paper describes a software tool to determine the values of parameters that may be extracted from charge spectra, namely the number of detected photons, gain, gain spread, prompt cross-talk, after-pulsing, dark count rate, and electronics noise.

Spectra with and without illuminating the SiPM can be analysed. They are obtained by integrating the SiPM current during a gate. For the light source, it is assumed that a Poisson distribution can describe the number of photons and that their time spread is short compared to the gate length. The entire charge spectrum is fitted by the detector response model (DRM), proposed in [97], that describes the spectrum, including the regions in-between the peaks, accounting for prompt cross-talk, after-pulsing and dark counts. However, delayed cross-talk, as described in [96], is not implemented. This method is in contrast to the standard methods of analysing charge spectra, which is to fit the peaks corresponding to 0, 1, 2, . . . discharges by individual Gaussian distributions to extract the gain, the gain spread, the distribution of the number of discharges, and the electronics noise [137, 138]. Since the latter approach does not include pulses from dark counts and after-pulses, their influence on the values of the measured SiPM parameters is not apparent.

A generally-available Python module has been developed to provide a robust and user-friendly way to fit the detector response model of [97] to characterise SiPMs. In [97, 139] the detector response model was used to analyse SiPM charge spectra.

In Section 4.3 a modification of the after-pulse probability of [97] to account for the recharging of the SiPM, introduced in [96], is presented. It results in an improved description of the effects of after-pulses. Given that the fit has ten free parameters, determining their initial values, presented in Section 4.3.1, is an essential part of the software tool.

Details of the fit are discussed in Section 4.3.6. The validation of the program for a wide range of SiPM parameters is presented in Section 4.4.1, using SiPM spectra generated by the simulation program of [96]. Finally, in Section 4.4.2, the program is used to analyse experimental data from two SiPMs operated at room temperature for a wide range of over-voltages.

In this chapter, a 'Geiger discharge' indicates a discharge of the SiPM caused by the device entering a breakdown for any reason. A 'primary Geiger Discharge' initiates a macroscopic cascade in the SiPM. This can be, for instance, a discharge caused by a photon incident on the device or a dark count. By contrast, other types of discharge are dependent on the primary discharge having first occurred.

## 4.2 Theory

The program described in this chapter is a Python implementation of the SiPM detector response model for photons and dark counts of [97], with

improved treatment of after-pulses and the simultaneous treatment of the signals from photons and dark counts.

This section thoroughly describes the model and its physical motivation for later reference in the chapter.

### 4.2.1 Detector Response Model And Free Parameters

The detector response model used in this paper is defined in [97]. However, a notable change is the treatment of after-pulses, which includes the influence of the recharging of the SiPM on the after-pulse probability. The definitions of the free and fixed parameters of the model are first given in Table 4.1a and Table 4.1b.

The probability density function (p.d.f.) implemented in PeakOTron to describe charge spectra of SiPMs in response to low-intensity light and dark counts is summarised in Eq. 4.1. It has nine free parameters,  $\theta$ , explained in Table 4.1a. The remaining fixed parameters are shown in Table 4.1b.

$$f_{\text{DRM}}(K; \theta) = f_{\gamma}(K; \theta) * f_{\text{dark}}(K; \theta), \quad (4.1)$$

where  $f_{\gamma}(K; \theta)$  and  $f_{\text{dark}}(K; \theta)$  are the photon and the dark-count induced p.d.fs, respectively,  $K = (Q - Q_0)/G^*$  is the charge in the number of photo-electrons scale (p.e.), with  $Q$  the measured charge,  $G^*$  the effective gain,  $Q_0$  the pedestal, which is the mean measured charge of the 0 p.e. peak, and  $*$  the convolution operator<sup>1</sup>.

In this chapter, a distinction is made between the effective gain,  $G^*$ , which is the integral of the SiPM current pulse from a single primary Geiger discharge over the gate of length  $t_{\text{gate}}$ . The gain,  $G$ , the integral for  $t_{\text{gate}} \rightarrow \infty$ , which is used in the simulation program of [96]. Here,  $t_{\text{gate}}$  is the length of the integration range, or 'gate length', during which an SiPM pulse is integrated.

The model is presented using the variable  $K$ , the measured charge,  $Q$ , with the pedestal,  $Q_0$ , subtracted and divided by the gain,  $G^*$ . Thus, for the variable  $K$ , the mean of the pedestal peak is at 0, and the mean of the 1 p.e. peak is at 1. The probability density as a function of  $Q$  is obtained by dividing the probability density in the variable  $K$  by  $G^*$ .

The probability distributions  $f_{\gamma}(K; \theta)$  and  $f_{\text{dark}}(K; \theta)$ , as well as the treatment of the after-pulse model, are detailed later in this section.

### 4.2.2 Generalised Poisson Distribution

A common model for the discrete probability of a SiPM discharge induced by a single photon, with the possibility to produce cross-talk, is the 'Generalised Poisson Distribution' (GP) [141, 142]. The GP distribution is a discrete probability distribution and is the convolution of a Poisson distribution and the Borel distribution. The Borel distribution describes a Poisson-distributed number of successive discharges (prompt cross-talk),

1: A critical point of note is that the distribution of the sum of two independent random variables is given by their convolution:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \quad (4.2)$$

where  $Z = X + Y$ , and  $X$  and  $Y$  are independent random variables. This fact is used implicitly henceforth. Proof may be found in [140].

Parameter	Definition	Range
$\mu$	Mean Number of Primary Geiger Discharges from Photons	$10^{-10}$ to $\infty$
$\lambda$	GP-Branching Parameter	$10^{-10}$ to $1 - 10^{-10}$
$G^*$	Effective Gain	1 Bin to $\infty$
$Q_0$	Pedestal Position	$-\infty$ to $+\infty$
$\sigma_0$	Pedestal Width	0.1 Bin to $\infty$
$\sigma_1$	Gain Spread	0.1 Bin to $\infty$
DCR	Dark Count Rate	1 Hz to $\infty$
$p_{Ap}$	After-pulse Probability	$10^{-10}$ to $1 - 10^{-10}$
$\tau_{Ap}$	After-pulse Time Constant	3 ns to $t_{gate}/2$
$A_{sc}$	Scale Factor	$N_{events} \pm 3 \cdot \sqrt{N_{events}}$

(a)

Parameter	Definition	Default
$\tau$	Slow Time Constant SiPM Pulse	20 ns
$t_0$	Time Before Gate for Dark Counts	100 ns
$\tau_{rec}$	Recovery Time of SiPM	$0.65 \cdot \tau$
$t_{gate}$	Length of Integration Gate	100 ns
$i_{\gamma}^{max}$	Max. No. of Photon Primary Discharges	–
$i_{dark}^{max}$	Max. No. of Dark Primary Discharges	6
$\lambda_{red, Ped}^d, n_{\sigma}^d, n_{\sigma}^u$	Parameters for Non-Gaussian Pedestals	2, 2, 2
$N_{Peak}$	Min. No. of Events in Peaks	100
bin0	First Bin for Fit	0
bin_method	Binning Method	Knuth's Rule
prefit_only	Run Prefit Only	false

(b)

**Table 4.1:** Tables 4.1a and 4.1b show the free and fixed parameters of the Peak0Tron fits and their default values, which the user can change. The maximum number of primary discharges from photons,  $i_{\gamma}^{max}$  is obtained from the charge spectrum.

with a mean number of  $\lambda$  successors. The total number of discharges is therefore the sum of the original Poisson process initiating the discharge (e.g. from packets of laser light) and the Borel-distributed successive discharges, yielding the GP distribution. This distribution has been found to yield a good description of SiPM discharge probability [143, 144].

The GP distribution is given by Equation 4.3.

$$GP_{k,\mu,\lambda} = \frac{\mu \cdot (\mu + k \cdot \lambda)^{k-1} \cdot e^{-(\mu+k\cdot\lambda)}}{k!} \quad (4.3)$$

where  $k$  are the total number of Geiger discharges,  $\mu$  is the mean number of Geiger discharges,  $\lambda$  is the GP-branching parameter describing prompt cross-talk probability, defined in the range  $\lambda \in [0, 1]$ .

It is noted for later reference that the first raw moment and the second and third central moments of the GP-distribution ( $m_1, m_2, m_3$ ), and the corresponding data moments ( $M_1, M_2, M_3$ ), are given in terms of the GP parameters  $\mu$  and  $\lambda$  by [142, 144]:

$$\begin{aligned} m_1 &= \mu/(1-\lambda), & M_1 &= G^* \cdot m_1 + Q_0, \\ m_2 &= \mu/(1-\lambda)^3, & M_2 &= (G^*)^2 \cdot m_2, \\ m_3 &= \mu \cdot (1+2\lambda)/(1-\lambda)^5, & M_3 &= (G^*)^3 \cdot m_3. \end{aligned} \quad (4.4)$$

From these equations,  $\lambda$  is calculated using:

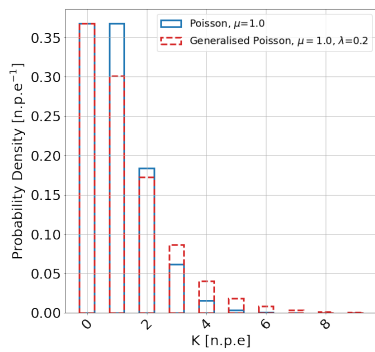
$$(1+2\lambda) = \frac{(M_1 - Q_0) \cdot M_3}{M_2^2}, \quad (4.5)$$

and  $\mu$  and  $G^*$  from:

$$\mu \cdot (1-\lambda) = \frac{(M_1 - Q_0)^2}{M_2}, \quad (4.6)$$

$$\frac{G^*}{(1-\lambda)^2} = \frac{M_2}{(M_1 - Q_0)}. \quad (4.7)$$

Therefore,  $\mu, \lambda$  and  $G^*$  can be independently estimated using the moments of the GP-distribution.



**Figure 4.1:** Probability mass functions of the Poisson and Generalised Poisson distributions in blue and red dashed lines, respectively. The input values are shown in Table 4.3.

### 4.2.3 Model for After-pulses

In contrast to what was expected, in [97], the additional charge of a single after-pulse had to be modelled by an exponential distribution to describe the measured spectra. The expected charge distribution derived in Appendix A of [97] did not describe experimental data. In the model, an after-pulse time dependence  $e^{-t_{Ap}/\tau_{Ap}}$  and a signal reduction by a factor  $1 - e^{-t_{Ap}/\tau}$  was assumed. The time between the after-pulse and the primary Geiger discharge is  $t_{Ap}$ , the after-pulse time constant is  $\tau_{Ap}$ , and the voltage-recovery time constant is  $\tau$ .

The model did not take into account the reduction of the Geiger-discharge probability during the recharging of the pixel, which in PeakOTron is parameterized by  $1 - e^{-t_{Ap}/\tau_{rec}}$ , with the recovery-time constant  $\tau_{rec}$ . The contribution of a single after-pulse to the charge spectrum is described by Equation 4.8:

$$p_{Ap} \cdot f_{Ap}(t_{Ap}; \tau_{rec}, \tau_{Ap}, t_{gate}), \quad (4.8)$$

where:

$$f_{Ap}(t_{Ap}) = \begin{cases} (1 - e^{-t_{Ap}/\tau_{rec}}) \cdot e^{-t_{Ap}/\tau_{Ap}} / \text{Norm} & 0 < t_{Ap} < t_{gate} \\ 0 & \text{otherwise.} \end{cases} \quad (4.9)$$

In Equation 4.9,  $f_{Ap}$  is the after-pulse probability density function (p.d.f.),  $p_{Ap}$  the probability of a single after-pulse for a single primary Geiger discharge, and Norm the normalisation factor. Justification for the model is provided in Appendix Section 8.1.0.1.

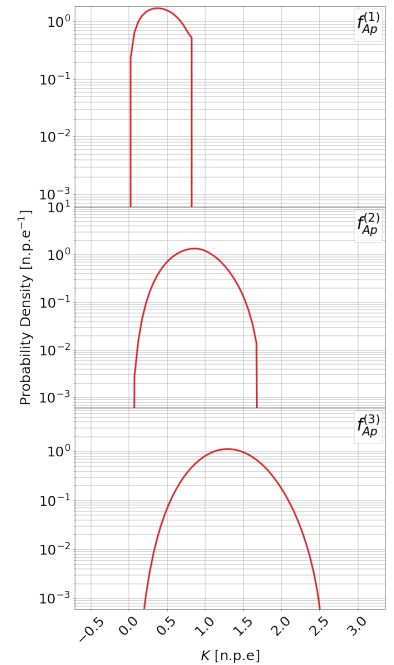
The motivation for this parameterisation is now discussed. Compared to a photon-induced signal at the time  $t = 0$ , the signal from an after-pulse at  $0 \leq t \leq t_{gate}$  is reduced by the factor  $(1 - e^{-t/\tau}) \cdot \int_0^{t_{gate}} e^{-t'/\tau} dt'$ . The first term describes the decrease of the signal due to the recharging of the pixel, and the second term is the fraction of the signal integrated by the gate.

The after-pulse-time probability density is modelled in Equation 4.9. The first term,  $1 - e^{-t/\tau_{rec}}$ , describes the decrease in Geiger-discharge probability due to the recharging of the pixel, and the second term,  $e^{-t/\tau_{Ap}}$ , the time distribution of charge carriers de-trapped from states in the silicon band gap. Both terms are parametrisations, which are only approximate. The second term assumes de-trapping from a single state only and no electric-field dependence of  $\tau_{Ap}$ .

For a single Geiger discharge at  $t = 0$  and a gate of length  $t_{gate}$  starting at  $t = 0$ , Norm of Equation 4.9 is given by Equation 4.10:

$$\text{Norm}(\tau_{Ap}, \tau_{rec}, t_{gate}) = \frac{\tau_{Ap} \left( \tau_{Ap} - e^{-\frac{t_{gate}}{\tau_{Ap}}} \left( \tau_{Ap} + \tau_{rec} \left( 1 - e^{-\frac{t_{gate}}{\tau_{rec}}} \right) \right) \right)}{(\tau_{Ap} + \tau_{rec})}. \quad (4.10)$$

The treatment of after-pulses is then the same as in Appendix A of [97]. As charge spectra are fitted, changing the after-pulse-time variable  $t_{Ap}$



**Figure 4.2:** Probability density functions of after-pulses, shown as red lines as a function of  $K$ . The top figure shows the 1<sup>st</sup> order distribution defined in Equation 4.11. The middle and lower plot show the 2<sup>nd</sup> and 3<sup>rd</sup> order autoconvolutions. The input values are shown in Table 4.3.



to charge  $K$  is required. The relationship between  $K$  and  $t_{\text{Ap}}$  has two branching solutions, one for  $0 < t_{\text{Ap}} \leq t_{\text{gate}}/2$  and one for  $t_{\text{gate}}/2 < t_{\text{Ap}} \leq t_{\text{gate}}$ . The probability density is calculated as the sum of the two branches.

$$f_{\text{Ap}}(K; \tau, \tau_{\text{Ap}}, \tau_{\text{rec}}, t_{\text{gate}}) = \left| \frac{dK}{dt}(t_{\text{Ap}}(K; \tau, t_{\text{gate}}); \tau, t_{\text{gate}}) \right|^{-1} \cdot (f_{\text{Ap}}(t_{\text{Ap}}(K; \tau, t_{\text{gate}}), \tau_{\text{Ap}}, \tau_{\text{rec}}) + f_{\text{Ap}}(t_{\text{gate}} - t_{\text{Ap}}(K; \tau, t_{\text{gate}}), \tau_{\text{Ap}}, \tau_{\text{rec}})), \quad (4.11)$$

where:

$$\left| \frac{dK}{dt}(t_{\text{Ap}}; \tau, t_{\text{gate}}) \right| = \frac{2 \left| \sinh((t_{\text{gate}}/2 - t_{\text{Ap}})/\tau) \right| e^{-t_{\text{gate}}/2\tau}}{\tau}, \quad (4.12)$$

and

$$t_{\text{Ap}}(K; \tau, t_{\text{gate}}) = \frac{t_{\text{gate}}}{2} - \tau \operatorname{arcosh} \left( \frac{(1-K) e^{t_{\text{gate}}/2\tau} + e^{-t_{\text{gate}}/2\tau}}{2} \right), \quad (4.13)$$

with  $f_{\text{Ap}}(K)$  defined in the range  $0 \leq K \leq (1 - e^{-t_{\text{gate}}/2\tau})^2$ . The derivations of Equations 4.12 and 4.13 for Geiger discharges induced by photons at  $t = 0$  can be found in [97].

#### 4.2.4 Model for Photon-Induced Discharges

The treatment of the photon-induced charge spectrum is the same as in [97], except that the after-pulse distribution is replaced by Equation 4.11. The probability density distribution is shown in Equation 4.14.

$$f_{\gamma}(K; \theta) = \text{GP}_{0,\mu,\lambda} \cdot \mathcal{N}(K; 0, \sigma_0/G^*) + \sum_{i=1}^{i_{\gamma}^{\max}} \text{GP}_{i,\mu,\lambda} \cdot \mathcal{N}(K; i, \sigma(i; \sigma_0/G^*, \sigma_1/G^*)) * \left( \text{Bi}_{0,i,p_{\text{Ap}}} \cdot \delta(K) + \text{Bi}_{1,i,p_{\text{Ap}}} \cdot f_{\text{Ap}}^{(1)}(K; \tau, \tau_{\text{Ap}}, \tau_{\text{rec}}, t_{\text{gate}}) + \sum_{j=2}^i \text{Bi}_{j,i,p_{\text{Ap}}} \cdot f_{\text{Ap}}^{(j)}(K; \tau, \tau_{\text{Ap}}, \tau_{\text{rec}}, t_{\text{gate}}) \right), \quad (4.14)$$

where  $\mathcal{N}$ , GP, Bi and  $\delta$  represent the normal, Generalised Poisson, Binomial and Dirac delta distributions, \* the convolution operator,  $\theta$  the parameters of Table 4.1a,  $f_{\text{Ap}}^{(i)}(K; \theta)$  the  $i - 1^{\text{st}}$  auto-convolution of  $f_{\text{Ap}}(K; \theta)$  (i.e.  $f_{\text{Ap}}^{(1)}(K; \theta) = f_{\text{Ap}}(K; \theta)$ ,  $f_{\text{Ap}}^{(2)}(K; \theta) = f_{\text{Ap}}(K; \theta) * f_{\text{Ap}}(K; \theta)$ , etc.), and  $\sigma(i; \sigma_0, \sigma_1) = \sqrt{\sigma_0^2 + i \cdot \sigma_1^2}$ .

Here, the convolved Gaussian distributions describe the combined effect of the electronic noise from the experimental setup ( $\sigma_0$ ) and the 'gain smearing' ( $\sigma_1$ ). The latter component arises from fluctuations in the off-voltage of the SiPM due to the stochastic nature of Geiger avalanches or pixel-to-pixel fluctuations in the capacitance or electric field [93, 96] and is introduced in Section 2.5.2. These processes are uncorrelated and therefore added in quadrature.

Delayed cross-talk is not implemented.

#### 4.2.5 Dark Count Model

The probability density distribution for a single dark pulse in the time interval  $-t_0 < t_{\text{dark}} < t_{\text{gate}}$  is given by Equation 4.15:

$$f_{\text{dark}}^{(1)}(K; \tau, t_0, t_{\text{gate}}) = \begin{cases} \frac{\tau}{t_0 + t_{\text{gate}}} \cdot \left( \frac{1}{K} + \frac{1}{1-K} \right) & \text{for } K_{\text{dark}}^{\min} \leq K \leq K_{\text{dark}}^{\max}, \\ \frac{\tau}{t_0 + t_{\text{gate}}} \cdot \left( \frac{1}{1-K} \right), & \text{for } 0 < K \leq K_{\text{dark}}^{\min}, \\ 0 & \text{otherwise,} \end{cases} \quad (4.15)$$

where  $K_{\text{dark}}^{\max} = (1 - e^{-t_{\text{gate}}/\tau})$  and  $K_{\text{dark}}^{\min} = e^{-t_0/\tau} (1 - e^{-t_{\text{gate}}/\tau})$ . Note that  $t_0$  is defined to be positive.

The probability density distributions for more than one primary Geiger discharge from dark counts for  $-t_0 < t < t_{\text{gate}}$ , are obtained by auto-convolutions of  $f_{\text{dark}}^{(1)}$ . Prompt cross-talk distributions are stretched single dark-count distributions:

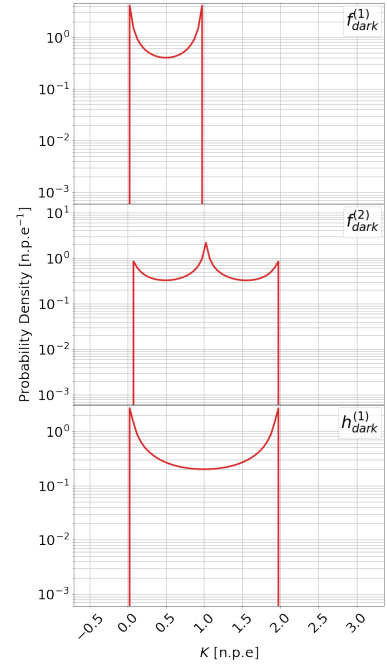
$$h_{i,\text{dark}}(K; \tau, t_0, t_{\text{gate}}) = \frac{f_{\text{dark}}^{(1)}(K/(i+1); \tau, t_0, t_{\text{gate}})}{i+1}, \quad (4.16)$$

where  $i$  is the number of cross-talk discharges.

The program calculates the charge distributions to arbitrary numbers of dark counts. The first four terms are given in Table 4.2, modified from [97]. The sum of all terms yields  $f_{\text{dark}}$ . The number of primary discharges are assumed to be Poisson distributed (P), characterised by the mean  $\mu_{\text{dark}} = \text{DCR} \cdot (t_0 + t_{\text{gate}})$ . The prompt cross-talk discharges are Borel distributed (B), characterised by the probability  $\lambda$ .

After-pulses and delayed cross-talk are not implemented in the present dark count model.

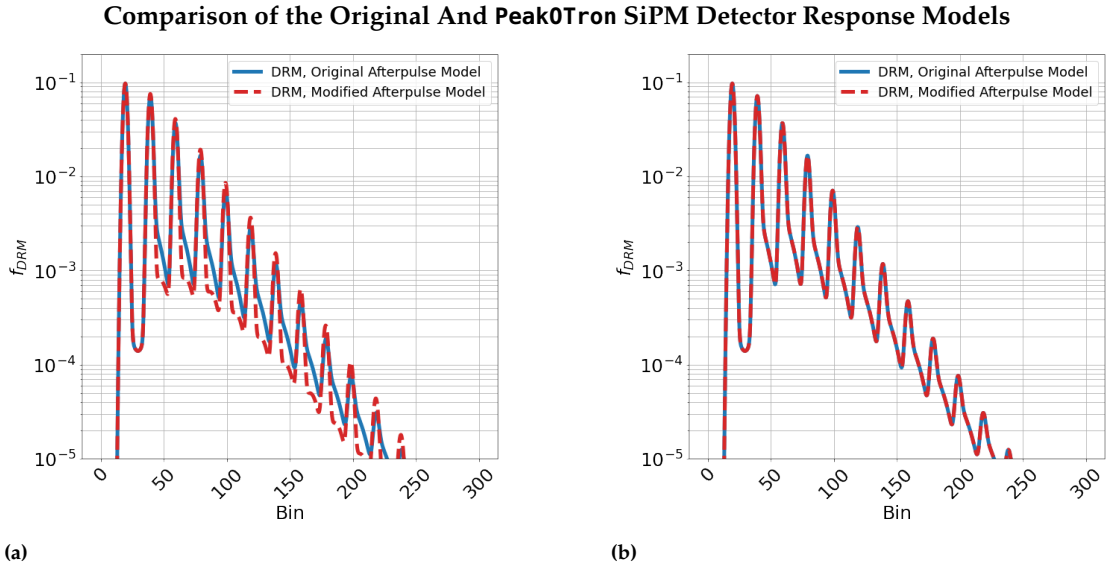
A comparison of the modified model and the original model of [97] are shown in Figure 4.4a and 4.4b.



**Figure 4.3:** Probability density functions of dark counts, shown as red lines as a function of  $K$ . The top figure shows the 1<sup>st</sup> order distribution defined in Equation 4.15. The middle figure shows the 2<sup>nd</sup> order auto-convolution. The bottom figure shows the 1<sup>st</sup>-order 'stretched' probability density functions defined in Equation 4.16. The input values are shown in Table 4.3.

No. of discharges	Comb.	Primary Geiger Probability	Cross Talk Probability	Distribution
0	1	$P_{0,\mu_{\text{dark}}}$	N/A	$\delta$
1	1	$P_{1,\mu_{\text{dark}}}$	$B_{0,\lambda}$	$f_{\text{dark}}^{(1)}$
2	1	$P_{1,\mu_{\text{dark}}}$	$B_{1,\lambda}$	$h_{1,\text{dark}}$
2	1	$P_{2,\mu_{\text{dark}}}$	$(B_{0,\lambda})^2$	$f_{\text{dark}}^{(2)}$
3	1	$P_{1,\mu_{\text{dark}}}$	$B_{2,\lambda}$	$h_{2,\text{dark}}$
3	2	$P_{2,\mu_{\text{dark}}}$	$B_{0,\lambda} \cdot B_{1,\lambda}$	$f_{\text{dark}}^{(1)} * h_{1,\text{dark}}$
3	1	$P_{3,\mu_{\text{dark}}}$	$(B_{0,\lambda})^3$	$f_{\text{dark}}^{(3)}$
4	1	$P_{1,\mu_{\text{dark}}}$	$B_{3,\lambda}$	$h_{3,\text{dark}}$
4	2	$P_{2,\mu_{\text{dark}}}$	$B_{0,\lambda} \cdot B_{2,\lambda}$	$f_{\text{dark}}^{(1)} * h_{2,\text{dark}}$
4	1	$P_{2,\mu_{\text{dark}}}$	$(B_{1,\lambda})^2$	$h_{1,\text{dark}}^{(2)}$
4	3	$P_{3,\mu_{\text{dark}}}$	$(B_{0,\lambda})^2 \cdot B_{1,\lambda}$	$f_{\text{dark}}^{(2)} * h_{1,\text{dark}}$
4	1	$P_{4,\mu_{\text{dark}}}$	$(B_{0,\lambda})^4$	$f_{\text{dark}}^{(4)}$

**Table 4.2:** The first four terms of the model of [97] for discharges from dark counts, with a modified notation. P stands for the Poisson- and B for the Borel-probability distribution. Note, that [97] has a typographical error in the zero-discharge column, which is corrected here.



**Figure 4.4:** Comparison of the charge spectra of the detector response model of [97] (blue continuous line) with the model of this paper (red dashed line). The spectra are generated using the program of [96] with the parameters of Table 4.3 for Figure 4.4a, and the same parameters with  $\tau_{\text{rec}} \rightarrow 0$  for Figure 4.4b.

**Comments:**

- ▶ the modified model reduces the probability of afterpulse discharges as the SiPM recharges. This results in the reduced probability of the region between the peaks, indicated by the dashed red line in Figure 4.4a;
- ▶ the original distribution of [97] can be recovered with the limit  $\tau_{\text{rec}} \rightarrow 0$ , or that the SiPM recovers instantaneously.

## 4.3 Methods And Tools

The design of the software is presented hereafter.

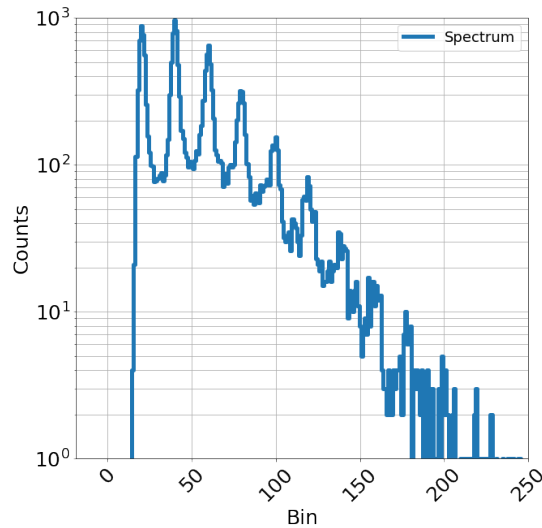
### 4.3.1 Model Input Parameters

PeakOTron requires charge spectra as input data. At first, the data is prepared as a histogram in Section 4.3.2, and then initial estimates for the effective gain in Section 4.3.3, pedestal and peak positions in Section 4.3.4, and of the dark-count rate in Section 4.3.5 are made. These estimates are used to determine the input parameters for the fit.

### 4.3.2 Data Preparation

Users can provide charge spectra in arbitrary units (C, V s, ADC, ...) either as histograms or lists of charge values, accepted in a standard numpy array format [130]. If a list of charges is provided, the program supports manual or automatic binning using one of the three methods (Scott's rule, Freedman-Diaconis Rule, Knuth's Rule) [145–148]. Figure 4.5 shows an example of a SiPM spectrum simulated using the program of [96]. The prefitting, i.e. the determination of the input parameters for the fit, and the fit itself, is demonstrated using this histogram.

Example Simulated Charge Spectrum To Illustrate The Prefitting Method Of PeakOTron

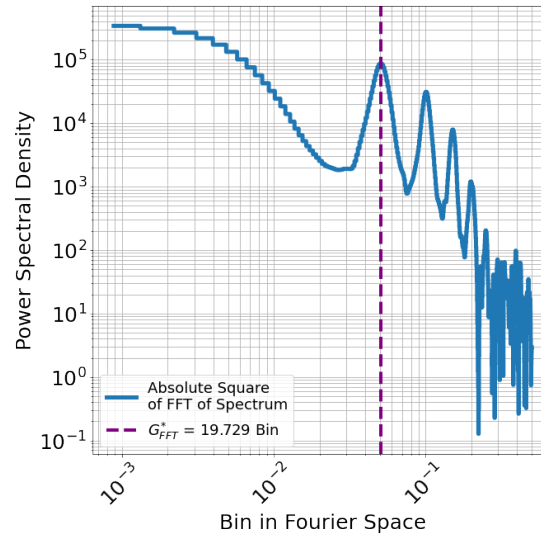


**Figure 4.5:** Exemplary charge spectrum containing  $2 \times 10^4$  events, which have been simulated using the program of [96] with the baseline values of Table 4.3 except for  $DCR = 5$  MHz. For the bin width  $0.05 \cdot G$  has been chosen, where  $G$  is the total charge of a single Geiger discharge. As the assumed gate width  $t_{\text{gate}} = 100$  ns, the effective gain  $G^* = 19.865$  Bin.

### 4.3.3 Effective Gain using the Fourier Transform ( $G_{\text{FFT}}^*$ )

In this step, an estimate for the effective gain,  $G_{\text{FFT}}^*$ , is made. First, the frequency domain representation of the histogram is calculated using the numpy Fast Fourier Transform [130]. Next, the power spectral density is calculated by taking the absolute square of the frequency-domain

### Power Spectral Density of Simulated Charge Spectrum Used To Estimate Used To Estimate Effective Gain in Peak0Tron



**Figure 4.6:** Power spectral density of the Fourier-transformed charge spectrum shown in Figure 4.5. The dashed vertical line indicates the gain frequency extracted using a spline fit. Its reciprocal is the estimate of the effective gain,  $G_{\text{FFT}}^*$ , shown in the legend in Bin units, which can be compared to  $G^* = 19.865$  Bin for the simulation.

histogram. Then, a spline fit is made to the power spectral density [149], and the position of the lowest peak is determined. The reciprocal of this value is an estimate of the effective gain,  $G_{\text{FFT}}^*$ .

An example of the method, as applied to the example histogram, is shown in Figure 4.6.

### 4.3.4 Peak Finding

#### 4.3.4.1 Initial Estimation of Peak Positions

A cubic spline fit to the entire charge spectrum is made, and the position of the highest peak is defined as the reference peak position,  $Q_{\text{max}}$ . The remaining peak positions in the spectrum are obtained from  $Q_{\text{max}} \pm i \cdot G_{\text{FFT}}^*$  for positive integers  $i$ .

This peak-finding method is chosen because it does not require events in the peak to estimate its position. This can occur for the pedestal peak if the mean number of Geiger discharges is high, and thus the probability for pedestal events is low.

#### 4.3.4.2 Pedestal Estimation

Assuming that the first three moments of the charge distribution can be approximately described using the moments of a Generalized Poisson (GP) distribution, the pedestal position can be estimated [141, 144].

As illustrated previously in Section 4.2.2, the gain  $G^*$  is related to the pedestal,  $Q_0$ , the first raw moment,  $M_1$ , and the second and third central moments,  $M_2$  and  $M_3$ , of the charge spectrum by:

$$G^*(Q_0, M_1, M_2, M_3) = \left( \frac{M_2}{(M_1 - Q_0)} \right) \cdot (1 - \lambda(Q_0; M_1, M_2, M_3))^2 \quad (4.17a)$$

$$\mu(Q_0, M_1, M_2, M_3) = \left( \frac{(M_1 - Q_0)^2}{M_2 \cdot (1 - \lambda(Q_0, M_1, M_2, M_3))} \right) \quad (4.17b)$$

$$\lambda(Q_0, M_1, M_2, M_3) = \frac{1}{2} \left( \frac{(M_1 - Q_0) \cdot M_3}{M_2^2} - 1 \right) \quad (4.17c)$$

The parameters of the GP distribution are  $\mu$  and  $\lambda$ , with  $\lambda$  the branching parameter and  $\mu$  the mean value for  $\lambda = 0$ .

$G^*$ ,  $\mu$  and  $\lambda$  may be estimated from a charge spectrum using these relationships [141, 144]. The relationship of Equation 4.17b is introduced here for subsequent use. An outline of proof for these relations has been provided in Section 4.2.2.

The pedestal is estimated by minimizing the square of the difference between  $G^*$ , calculated from the charge spectrum using Equation 4.17a, and  $G_{\text{FFT}}^*$ , the gain extracted from the power spectral density:

$$Q_0^{\text{est}} = \arg \min_{Q_0 \leq Q_{\text{max}}} \left( (G^*(Q_0; M_1, M_2, M_3) - G_{\text{FFT}}^*)^2 \right). \quad (4.18)$$

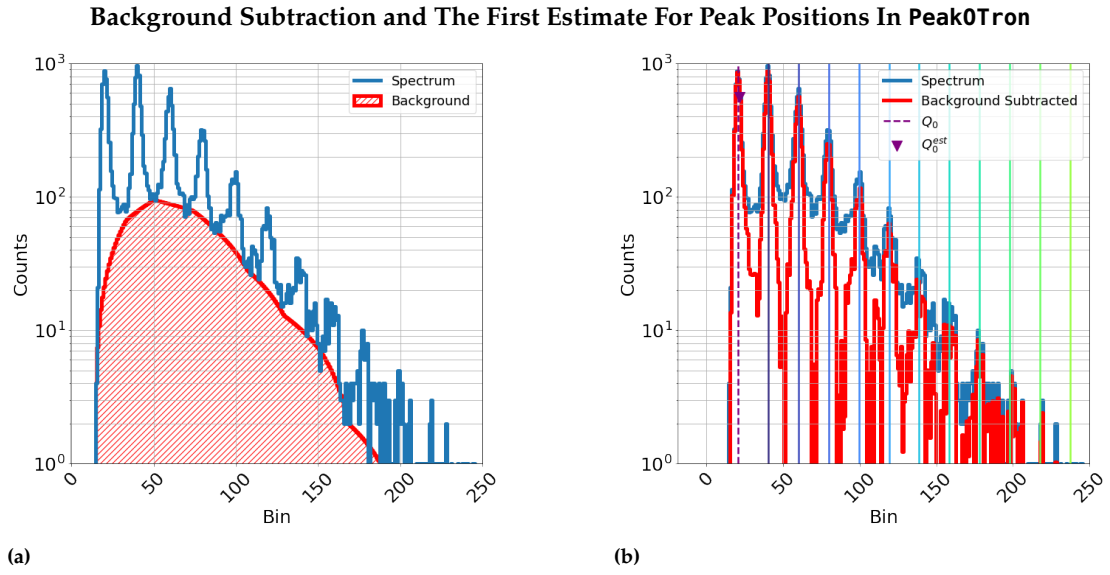
The function  $\arg \min$  gives the value of  $Q_0$  which minimizes the expression in parentheses. The specified limit for  $Q_0$  assures that the pedestal value,  $Q_0$ , is less or equal to  $Q_{\text{max}}$ , the peak with the maximum number of counts of the spectrum. Finally, the candidate peak from the set obtained in Section 4.3.4 nearest to  $Q_0^{\text{est}}$  is selected as pedestal. The peaks in the set with values less than  $Q_0^{\text{est}}$  are removed.

#### 4.3.4.3 Improved Peak Position Estimate

The peaks sit on a background from dark counts and delayed correlated pulses. If the background has a finite slope, the peak position is shifted. To improve the estimated peak positions a background is subtracted.

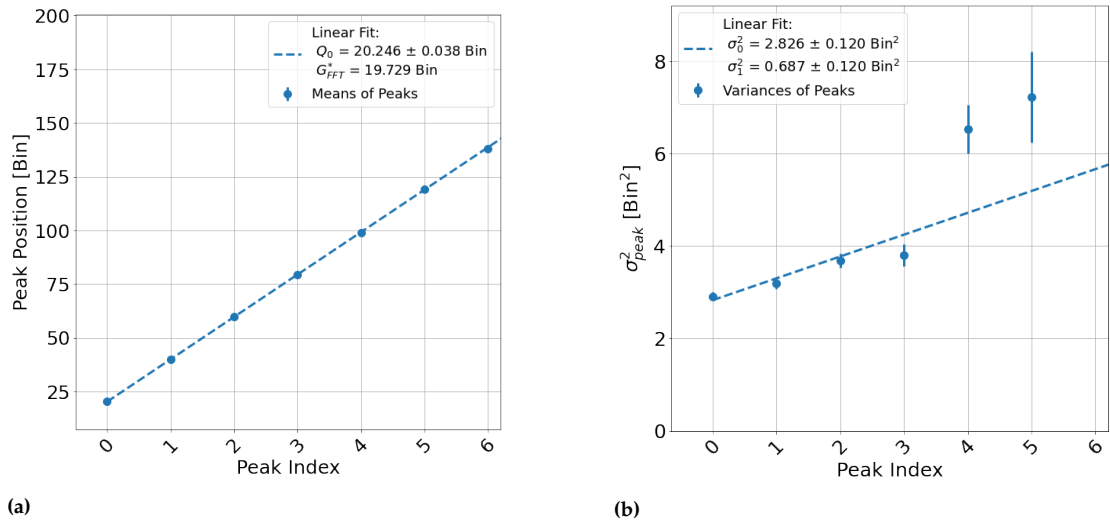
The background is estimated by a cubic spline fit to the minima of the spectrum in-between the peaks, which requires that the peaks are resolved. An example of the estimated background is shown in Figure 4.7a. The bin contents are set to zero if the background subtraction results in negative numbers. Figure 4.7b shows the background-subtracted spectrum.

The estimates of the peak positions are improved by determining an improved  $Q_{\text{max}}$  from the background-subtracted spectrum and by applying the methods described in Section 4.3.4.1.



**Figure 4.7:** Figure 4.7a shows the original and the estimated background as a blue line and a red-shaded area. Figure 4.7b shows the background-subtracted charge spectra as a red line with the peak positions indicated by vertical lines, estimated after the background subtraction. The inverted triangle indicates the position of the estimated pedestal,  $Q_0^{\text{est}}$ , and the dashed vertical purple line indicates the nearest peak position. The coloured lines indicate subsequent peaks.

### Linear Fits To Mean And Variance Of Peak Positions in Peak0Tron For The Prefitting of The $Q_0$ , $\sigma_0$ and $\sigma_1$ Free Parameters of the Detector Response Model



**Figure 4.8:** Straight-line fits to the means and variances extracted as described in Section 4.3.5.1. Figure 4.8a shows the fit to the means, with the slope fixed to  $G_{\text{FFT}}^*$ , is shown as dashed line. The intercept determines the prefit value for the pedestal position,  $Q_0$ . Its value is given in the insert. Figure 4.8b shows the fit (dashed line) to the variances of the peaks. The intercept determines the prefit value for  $\sigma_0^2$ , and the slope, the prefit value for  $\sigma_1^2$ . The values are given in the insert.



### 4.3.5 Determination of the Input Parameters

This section discusses the determination of the input parameters for the fits to the charge spectra, using the background-subtracted spectrum and the initial estimates of the effective gain, pedestal and peak positions.

#### 4.3.5.1 Pedestal Position and Width, Gain Spread ( $Q_0, \sigma_0, \sigma_1$ )

The pedestal is re-estimated in this step. Ranges of  $\pm G_{\text{FFT}}^*/2$  from each estimated peak position are selected from the background-subtracted spectrum, with the requirement that more than  $N_{\text{Peak}}$  events are observed in that range. The default value of  $N_{\text{Peak}} = 100$  may be changed by the user.

First, the mean,  $m$ , and the standard deviation,  $\sigma$ , of the spectrum in the range of the pedestal peak are calculated. If  $\sigma < G_{\text{FFT}}/4$ , the sub-range  $m \pm 2 \cdot \sigma$  is selected, and a Gaussian fit is performed to the background-subtracted spectrum in this sub-range. Then, a new sub-range is selected using the  $m$  and  $\sigma$  from the Gaussian fit. This fitting procedure is repeated for a maximum of ten iterations or until  $m$  and  $\sigma$  have changed by less than 1% of the bin width from the preceding iteration. Once one of the criteria is fulfilled, the  $m$  and  $\sigma$  from the last iteration are recorded.

The iterative fit procedure described for the pedestal is then repeated for each subsequent peak. This procedure results in a mean and a standard deviation for each peak. If there are fewer than three peaks with at least  $N_{\text{Peak}}$  events, then the means and standard deviations in the ranges  $\pm G_{\text{FFT}}^*/2$  from the three peaks which contain most events in that ranges are used instead. Once the described procedure has been completed, straight-line fits are performed to the means and standard deviations from the iterative procedure.

First, a straight-line fit to the mean peak positions versus peak number with the slope fixed to  $G_{\text{FFT}}^*$  is performed. The intercept is the final estimate for  $Q_0$ . Next, a straight-line fit to the variances,  $\sigma^2$ , versus peak number is performed. The intercept and slope are used to obtain the final estimates of  $\sigma_0$  and  $\sigma_1$ . Both fits are performed with MIGRAD, using the Huber Loss cost function, which reduces the influence of outliers. The Huber Loss is a combination of a quadratic and a linear cost function that attributes a lower weight to outliers than the purely quadratic cost function used for the  $\chi^2$  goodness-of-fit <sup>2</sup>.

In Fig. 4.8a the straight-line fit to the means for estimating  $Q_0$ , and in Fig. 4.8b the straight-line fit to the variances for estimating  $\sigma_0$  and  $\sigma_1$ , are shown.

#### 4.3.5.2 Estimates of $\mu$ and $\lambda$

The mean number of photon-induced primary Geiger discharges,  $\mu$ , and the prompt cross-talk probability,  $\lambda$ , are calculated from Eq. 4.17b and Eq. 4.17c, respectively, with the moments calculated from the original charge spectrum shown in Figure 4.5. The number of photoelectron peaks in the spectrum to be fitted is  $i_{\gamma}^{\text{max}} = \text{floor}((Q_{\text{up}} - Q_0) / G_{\text{FFT}}^*)$ , where

2: In order to reduce the sensitivity to outliers, Huber Loss is used as the cost function of the straight-line fits in Section 4.3.5.1. Huber Loss is defined in Equation 4.19:

$$L_{\delta}(z) = \begin{cases} \frac{1}{2}z^2 & \text{for } |z| \leq \delta \\ \delta (|z| - \frac{1}{2}\delta) & \text{otherwise,} \end{cases} \quad (4.19)$$

where  $z = (\hat{y} - y) / \sigma_y$  with the value of the fit function,  $\hat{y}$ , the measured value,  $y$ , and  $\sigma_y$ , the uncertainty of  $y$ .  $\delta$  is the threshold parameter, which is typically set to 1.345, to ensure at least a 95% statistical efficiency compared to a least squares fit, if the true distribution is a Gaussian [150].

$Q_{\text{up}}$  is the maximum charge of the spectrum and  $\text{floor}(x)$  gives the largest integer  $\leq x$ .

#### 4.3.5.3 Dark Count Rate Estimate (DCR)

The starting values of  $DCR$  for the fit are calculated using Equation 4.20a and Equation 4.20b:

$$DCR = DCR' \cdot e^{DCR' \cdot \tau} \quad (4.20a)$$

$$DCR' = \frac{dN_{\text{dark}}/dK(K = 0.5)}{4 \cdot \tau \cdot N_{0.5}}, \quad (4.20b)$$

where  $N_{0.5}$  is the number of entries in the spectrum up to  $K = 0.5$  and  $dN/dK(K = 0.5)$  is estimated from the counts of the histogram in the range  $0.45 \leq K \leq 0.55$ .

As discussed in [97], in the absence of illumination, the spectrum as a function of  $K$  of a single dark count randomly distributed in time is  $dN/dK = DCR \cdot \tau \cdot (1/K + 1/(1 - K))$ . From this follows that for a total of  $N_{\text{dark}}$  events,  $dN/dK(K = 0.5) = 4 \cdot DCR \cdot \tau \cdot N_{\text{dark}}$ . In [97], it is also shown that the mean number of dark counts with  $K > 0.5$  is  $\mu_{\text{dark}} = t_{\text{gate}} \cdot DCR$ . If  $\mu_{\text{dark}} \ll 1$ , one can replace  $N_{\text{dark}}$  by  $N_{0.5}$ , giving  $dN/dK(K = 0.5) \approx 4 \cdot DCR \cdot t_{\text{gate}} \cdot N_{0.5}$ . If  $\mu_{\text{dark}}$  increases and approaches 1, the approximation  $N_{0.5} \approx N_{\text{dark}}$  worsens, and in addition,  $N_{0.5}$  is reduced by the Poisson probability of no dark count in the time interval  $t_{\text{gate}}$ ,  $P(0; \mu_{\text{dark}}) = e^{-t_{\text{gate}} \cdot DCR}$ . At the same time, the probability of more than one dark count producing a significant signal increases, which further weakens above arguments. Using the simulation program of [96], it was found that replacing  $e^{-\mu_{\text{dark}}}$  by  $e^{-\tau \cdot DCR}$ , yields better initial values for  $DCR$ . An example of the  $K$ -ranges used to estimate  $DCR$  is shown in Figure 4.9.

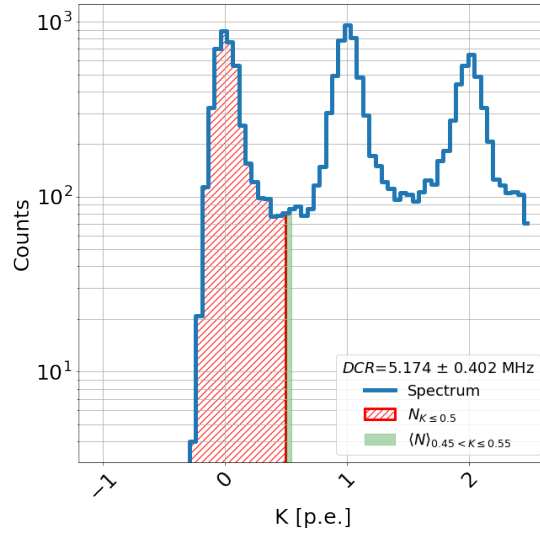
The maximum number of peaks from dark counts in the fit is  $i_{\text{dark}}^{\text{max}}$ . Its default value is 6. The user may modify the value above or equal to a minimum of 4.

#### 4.3.5.4 After-pulse Parameters ( $p_{\text{Ap}}$ , $\tau_{\text{Ap}}$ )

The after-pulse parameters cannot be readily extracted from the spectrum without performing the fit. Therefore, the ad-hoc initial values,  $p_{\text{Ap}} = 0.1$  for the after-pulse probability, and  $\tau_{\text{Ap}} = 5 \text{ ns}$  for the after-pulse time constant, are used.

To take into account the physics constraints  $0 \leq p_{\text{Ap}} < 1$  and  $\tau_{\text{Ap}} \geq 0$ , the parameter limits shown in Table 4.1a are applied in the fit. These choices can be changed by the user.

### Illustration of the Method Used To Prefit The Dark Count Rate In Peak0Tron



**Figure 4.9:** Charge spectrum of Figure 4.5 for  $K \leq 2.5$  p.e. The estimate of  $N_{\text{dark}}$  is shown in red-line shading, and the region for determining  $dN_{\text{dark}}/dK(K = 0.5)$  in solid green shading. The estimate of  $DCR$  using Equation 4.20a is given in the insert. The spectrum was simulated with  $DCR = 5$  MHz.

#### 4.3.6 Implementation of the Fit

After determining the input parameters of the model, the spectra are fitted with the binned maximum-likelihood method using MIGRAD implemented in `iminuit`, a Python interface to the MINUIT2 C++ package [151]. The logarithmic likelihood function used is:

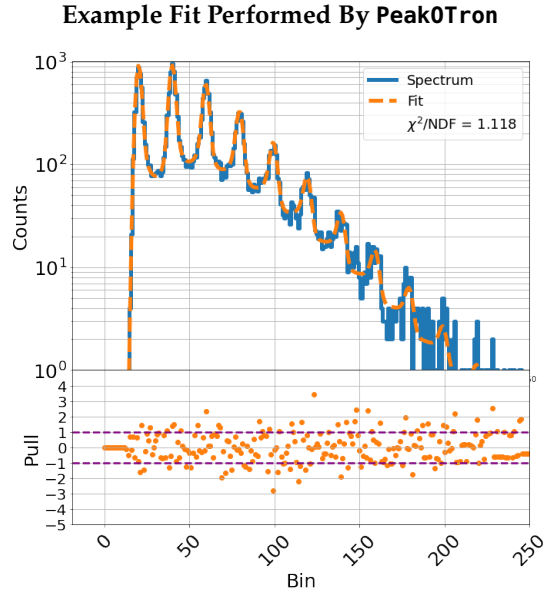
$$\mathcal{L}_{BL}(Q, N; \theta) = - \sum_{b \in \text{bins}} \left( N_b \cdot \ln \left( \frac{\hat{N}_b(Q; \theta)}{N_b} \right) + (N_b - \hat{N}_b(Q; \theta)) \right). \quad (4.21)$$

The bin index is  $b$ ,  $N$  denotes the histogram,  $N_b$  are the counts in bin  $b$ , and  $\hat{N}_b$  are the counts in bin  $b$  predicted by the model.

$\hat{N}_b$  is obtained from  $A_{\text{sc}} \cdot f_{\text{DRM}}(Q; \theta) \cdot \Delta Q$ , where the scaling factor  $A_{\text{sc}} \approx N_{\text{events}}$  is a free parameter,  $Q$  the measured charge,  $\Delta Q$  the bin width, and  $f_{\text{DRM}}$  the p.d.f. of the detector response model. The last term in parentheses of Equation 4.21 results in a pure parabolic behavior for each term at the minimum. In addition to the ten free parameters of the fit, Peak0Tron also uses a number of fixed parameters, which are given in Table 4.1b together with their default values, which can be changed by the user.

The Peak0Tron fit result for the spectrum of Figure 4.5 is shown in Figure 4.10, together with the pulls, the difference of fitted and measured number of counts divided by the estimated statistical uncertainty. For the uncertainty the square root of the fitted number of events, which can be less than one, has been assumed. It can be seen that the model provides a description of the simulated spectrum within its statistical uncertainty.

Frequently, measured spectra show non-Gaussian tails below the pedestal peak. Examples are given in Section 4.4.2. To deal with this problem,  $\chi_{\text{red}}^2$ , the  $\chi^2$  up to a charge of  $Q_0 + n_\sigma \cdot \sigma_0$  divided by the corresponding



**Figure 4.10:** Fit to the spectrum of Figure 4.5 using Peak0Tron. The blue continuous line shows the spectrum and the dashed orange line the fit result. The subfigure at the bottom shows the pulls, the difference of the counts of the spectrum minus the fit results, divided by the statistical uncertainty of the data. The pulls and the  $\chi^2/\text{NDF}$ , given in the insert, allow judging the quality of the fit.

number of bins for the Peak0Tron fit to the entire spectrum, is calculated. If  $\chi_{\text{red}}^2 > \chi_{\text{red, Ped}}^2$ , the spectrum starting at  $Q_0 - 4 \cdot \sigma_0$  is fitted. If also in this case  $\chi_{\text{red}}^2 > \chi_{\text{red, Ped}}^2$ , the fit is repeated for charge values exceeding  $Q_0 - 3.5 \cdot \sigma_0$ . This procedure is iterated in  $0.5 \cdot \sigma_0$  steps until either  $\chi_{\text{red}}^2 \leq \chi_{\text{red, Ped}}^2$  or the limit  $Q_0 - n_{\sigma}^d \cdot \sigma_0$  is reached. The default values of  $n_{\sigma}^d$ ,  $n_{\sigma}^u$  and  $\chi_{\text{red, Ped}}^2$ , which can be changed by the user, are given in Table 4.1b.

Peak0Tron is compatible with the native Python multiprocessing module [152]. Thus, it is recommended that fits of many SiPM spectra are performed in parallel. In addition, Peak0Tron fit objects can be directly stored on disk, and recovered for later analyses [153].

## 4.4 Results

In this section, the performance of the Peak0Tron program is evaluated using simulation and experimental data.

### 4.4.1 Validation of Peak0Tron with Simulated Spectra

The performance of Peak0Tron was validated using spectra simulated with the program from [96], which is also discussed in Chapter 3. First, baseline parameters were selected with values typical for SiPMs. Each parameter was scanned in a wide range of values while keeping the other parameters fixed to the baseline values. Table 4.3 shows the baseline values and the scan ranges for each parameter. For every parameter set 100 simulations, each with  $2 \times 10^4$  events, were made.

The simulation program produces for every event charge values in units of n.p.e., which were scaled by the effective gain,  $G^*$ , and shifted by the

pedestal,  $Q_0$ , with the values given in Table 4.3. The charge values were binned into a histogram with the bin widths shown in Table 4.3, and then fitted with Peak0Tron.

The following models were used for the simulations:

- ▶ The SiPM pulse from photons was modeled by an exponential starting at  $t = 0$  with the time constant  $\tau$  and the area n.p.e. The charge was obtained by integrating the SiPM pulse from  $t = 0$  to  $t = t_{\text{gate}}$ . As shown in Figure 4.24, SiPM pulses typically have two components: a slow component due to the recharging of the pixel and a fast one arising from a capacitance parallel to the quenching resistor [91]. Like in [97], no contribution from a fast component was simulated;
- ▶ Primary photon-induced SiPM pulses were generated with Poisson-distributed n.p.e. values with a mean of  $\mu$  n.p.e;
- ▶ Primary dark-count induced SiPM pulses were generated with a charge of one n.p.e. Their number was modeled by a Poisson distribution with the mean  $\mu_{\text{dark}} = DCR \cdot (t_{\text{gate}} + t_0)$ , and their times were uniformly generated in the time interval  $-t_0 < t < t_{\text{gate}}$ .
- ▶ Prompt cross-talk, which causes discharges at the same time as the primary Geiger discharges, was generated with a Borel distribution [154] with the branching parameter  $\lambda$ ;
- ▶ After-pulses for primary and prompt-cross-talk discharges were generated with the time distribution given by Equation 4.9 and an amplitude proportional to  $(1 - e^{-t_{\text{Ap}}/\tau})$ ;
- ▶ Delayed cross-talk, which was not modelled in [97], was not simulated.

For each of these simulations the spectrum was fitted with Peak0Tron. More information on the simulation is presented in Chapter 3. Fitting 100 simulated spectra for every parameter set allows estimating the systematic bias and the statistical uncertainty of the fitted parameters from the mean and RMS spread of the distribution of the differences between fitted and simulated parameter values, respectively.

#### 4.4.1.1 Overall Performance Of Fit

The results of the fits are presented in Figures 4.11-4.18, which show the fitted and prefit parameter values, their biases and their statistical uncertainties. In addition, for the extreme values of the parameter scan range, simulated and fitted spectra together with the pulls are presented. The results for the bias and the statistical uncertainty of each parameter scan for the values from the fit and from the prefit are summarised in Table 4.4. It should also be noted that for most parameters the fit improves the bias and statistical uncertainty of the prefit values.

Parameter	Baseline	Scan Range	Scaling
$Q_0$	20.0 Bin	–	constant
$G$	20.0 Bin	–	constant
$G^*$	19.865 Bin	–	constant
$\mu$	1	0.5 – 8	linear
$\lambda$	0.2	0.01 – 0.3	linear
$\sigma_0$	0.075 G (1.5 Bin)	(0.02 – 0.15) G (0.4 – 3) Bin	linear
$\sigma_1$	0.02 G (0.4 Bin)	(0.02 – 0.15) G (0.4–3) Bin	linear
$DCR$	100 kHz	100 kHz – 5 MHz	linear
$p_{AP}$	0.0272	0.0027 – 0.0818	linear
$\tau_{AP}$	7.5 ns	(4.0 – 19.0) ns	linear
$\tau$	20 ns	–	constant
$\tau_{rec}$	20 ns	–	constant
$t_0$	100 ns	–	constant
$t_{gate}$	100 ns	–	constant
$r_{fast}$	0	–	constant
bin width	0.05 G	(0.01 – 0.25) G	linear
$N_{events}$	$2 \times 10^4$ events	$(10^3 - 5 \times 10^5)$ events	linear

**Table 4.3:** Summary of the baseline values and scan ranges of the parameters used in the simulations for the validation of Peak0Tron.  $G^*/G = (1/\tau) \cdot \int_0^{t_{gate}} e^{-t/\tau} dt$  is the fraction of the SiPM signal integrated during the gate.

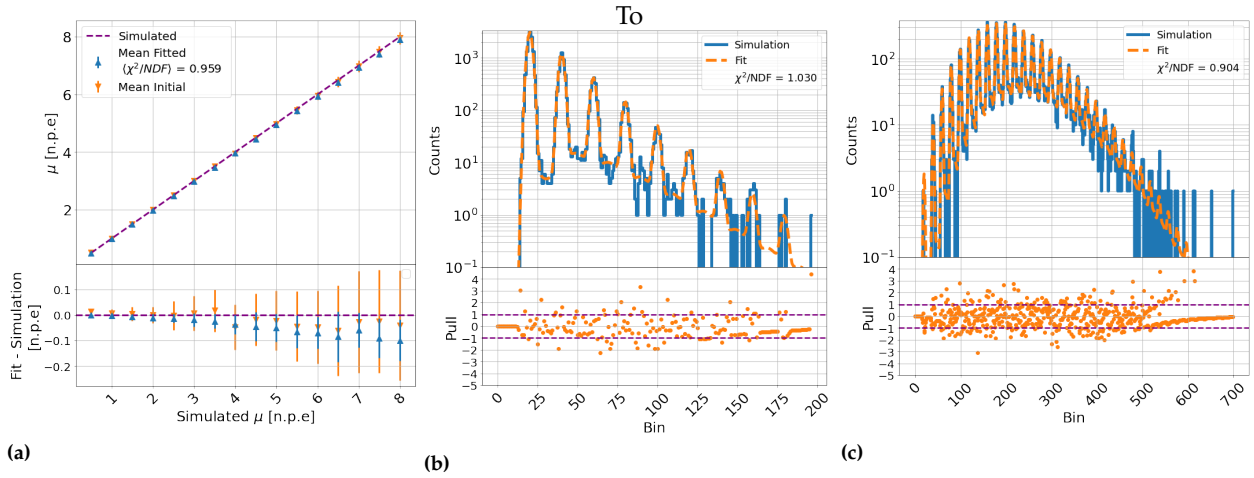
Parameter	unit	Bias	Stat. Uncertainty	Bias	Stat. Uncertainty
$Q_0$	Bin	-0.0017	0.0250	-0.124	0.185
$G$	Bin	0.0017	0.0186	0.073	0.092
$\mu$	–	-0.0319	0.0385	-0.020	0.120
$\lambda$	–	0.0075	0.0057	-0.005	0.015
$\sigma_0$	Bin	0.0307	0.0282	0.043	0.190
$\sigma_1$	Bin	-0.0108	0.0356	-0.207	0.353
$p_{AP}$	–	0.0009	0.0023	–	–
$\tau_{AP}$	ns	-0.2681	0.9618	–	–
$DCR$	MHz	-0.0587	0.1546	-0.130	0.298

**Table 4.4:** Biases and statistical uncertainties of the fitted parameters for the scans of Table 4.3.

#### Summary Of Figures 4.11-4.18 and Table 4.4

From Figures 4.11-4.18 it is concluded that the simulated spectra are well described by the fit with values of  $\chi^2/NDF$  close to one and no regions with significant differences between fit and simulation. The figures and Table 4.4 show that, for the parameters which are varied in the scan the biases are small: below 0.1 p.e. for  $\mu$ , below 0.002 Bin for  $G^*$ , below 0.002 Bin for  $Q_0$ , below 0.01 for  $\lambda$ , about 0.03 Bin for  $\sigma_0$ , about 0.01 Bin for  $\sigma_1$ , about 0.001 for  $p_{AP}$ , and below 0.5 ns for  $\tau_{AP}$ . Typically the biases are smaller than the statistical uncertainties.

**$\mu$  Scan of Fits With Peak0Tron To Simulation In The Range  $\mu = 0.5 - 8$  p.e.**

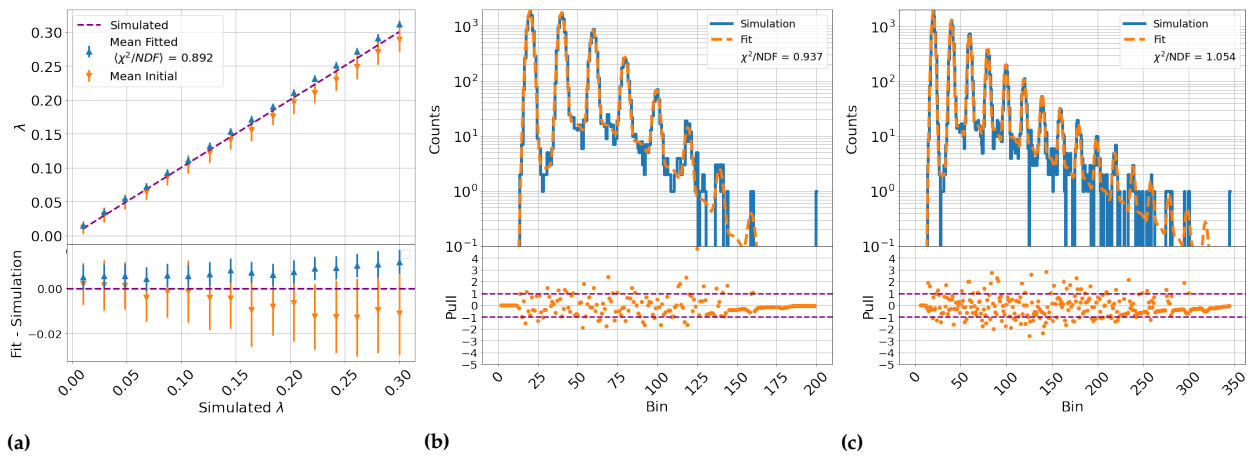


**Figure 4.11:** Comparison of the fit and the prefit values to the simulated values for the scan  $\mu = 0.5 - 8$  p.e.. Figure 4.11a shows mean fitted and prefit values and in the subplot the mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.11b and 4.11c show simulated charge spectrum and fit results, and below the pulls for  $\mu = 0.5$  p.e. and  $\mu = 8$  p.e., respectively.

**Comments:**

- Minor systematic biases are observed as  $\mu \approx 8$  in Figure 4.11a. This is a consequence of  $\mu$  being more challenging to measure as proportionally fewer events are spread across a greater number of Geiger discharge peaks. It is, however, demonstrated in Table 4.4, this effect is minor overall.

**$\lambda$  Scan of Fits With Peak0Tron Simulation In The Range  $\lambda = 0.01 - 0.3$**



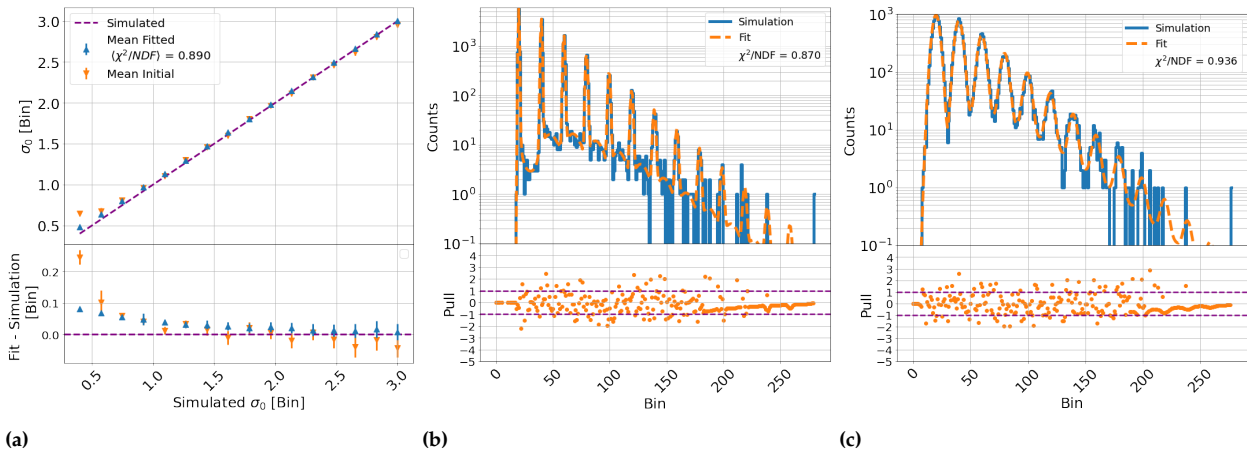
**Figure 4.12:** Comparison of the fit and the prefit values to the simulated values for the scan  $\lambda = 0.01$  to 0.3. Figure 4.12a shows mean fitted and prefit values and in the subplot the mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.12b and 4.12c show simulated charge spectrum and fit results, and below the pulls for  $\lambda = 0.01$  and  $\lambda = 0.3$ , respectively.

**Comments:**

- An overall bias is observed for both prefit and fit. The reason is unknown. However, it is shown to reduce with the number of events in the sample in Figure 4.19, and is therefore likely a consequence of limited statistics.



$\sigma_0$  Scan of Fits With Peak0Tron To Simulation In The Range  $\sigma_0 = 0.02 - 0.15$  G

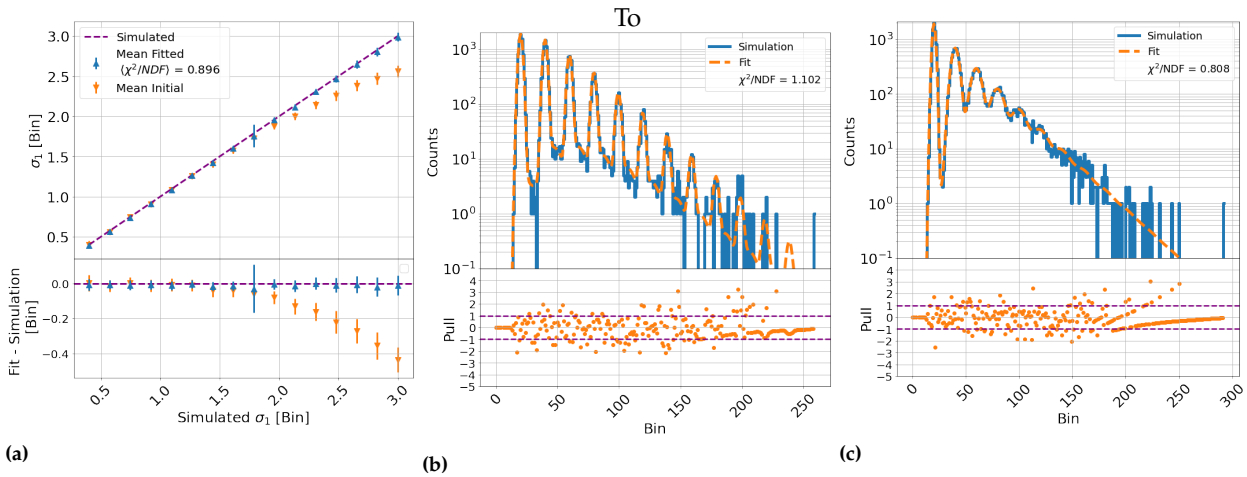


**Figure 4.13:** Comparison of the fit and the prefit values to the simulated values for the scan  $\sigma_0 = 0.02 - 0.15$  G. Figure 4.13a shows fitted and prefit values, and below, mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.13b and 4.13c show simulated charge spectrum and fit results, and below the pulls for  $\sigma_0 = 0.02$  G and  $\sigma_0 = 0.3$  G, respectively.

**Comments:**

- ▶ As shown in Figure 4.13a, the bias of  $\sigma_0$  increases if  $\sigma_0$  is smaller than the bin width. This could be cured if in the fit the integral over the bins of the fit function is used instead of its value at the bin centre.

$\sigma_1$  Scan of Fits With Peak0Tron To Simulation In The Range  $\sigma_1 = 0.02 - 0.15$  G

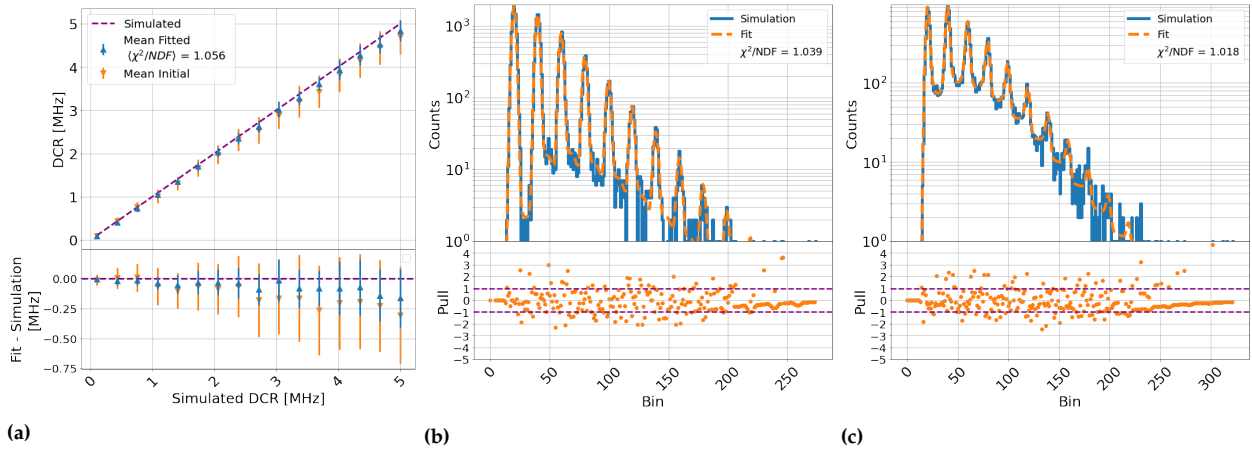


**Figure 4.14:** Comparison of the fit and the prefit values to the simulated values for the scan  $\sigma_1 = 0.02 - 0.15$  G. Figure 4.14a shows fitted and prefit values, and below, mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.14b and 4.14c show simulated charge spectrum and fit results, and below the pulls for  $\sigma_1 = 0.02$  G and  $\sigma_1 = 0.3$  G, respectively.

**Comments:**

- ▶ Figure 4.14a illustrates that the bias of  $\sigma_1$  increases for the prefit, which illustrates the increased challenge of estimating the gain spread as the width of the peaks increase with independent Gaussian fits. By contrast, the model can describe  $\sigma_1$ .

### DCR Scan of Fits With Peak0Tron To Simulation In The Range $DCR = 0.1 - 5$ MHz

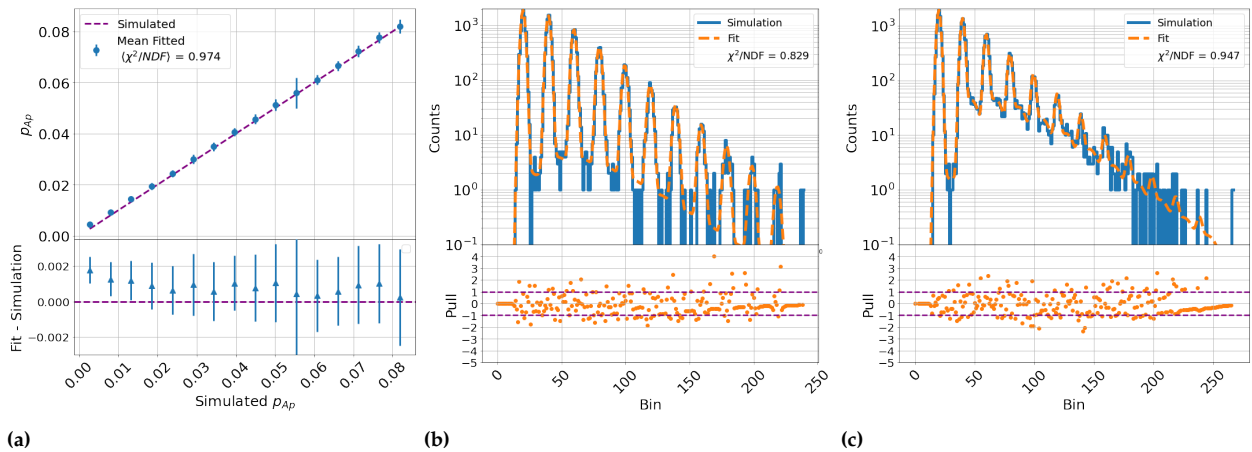


**Figure 4.15:** Comparison of the fit and the prefit values to the simulated values for the scan  $DCR = 0.1 - 5$  MHz. Figure 4.15a shows fitted and prefit values, and below, mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.15b and 4.15c show simulated charge spectrum and fit results, and below the pulls for  $DCR = 0.1$  MHz and  $DCR = 5$  MHz, respectively.

#### Comments:

- Figure 4.15a illustrates that the bias and error of the fitted  $DCR$  increases for the prefit with simulated  $DCR$ . By contrast, the model is capable of describing  $DCR$  with smaller errors and bias than the prefit.

### $p_{Ap}$ Scan of Fits With Peak0Tron To Simulation In The Range $p_{Ap} = 0.0027 - 0.0818$

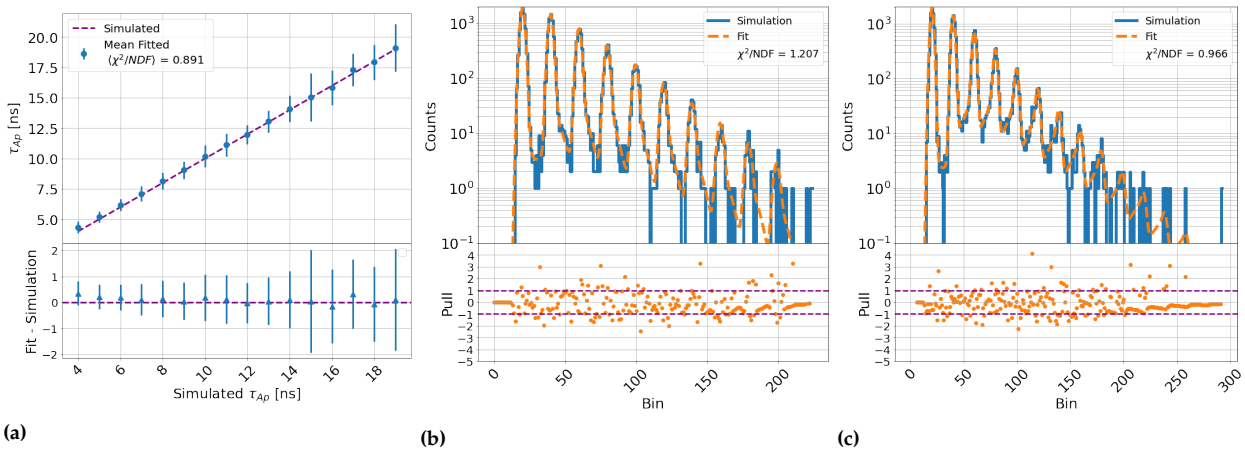


**Figure 4.16:** Comparison of the fit and the prefit values to the simulated values for the scan  $p_{Ap} = 0.0027 - 0.0818$ . Figure 4.16a shows fitted and prefit values, and below, mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.16b and 4.16c show simulated charge spectrum and fit results, and below the pulls for  $p_{Ap} = 0.0027$  and  $p_{Ap} = 0.0818$ , respectively.

#### Comments:

- Figures 4.16b-Figure 4.16c indicate that the model is able to describe the regions between the peaks caused by afterpulses.

$\tau_{Ap}$  Scan of Fits With Peak0Tron To Simulation In The Range  $\tau_{Ap} = 4 - 19$  ns

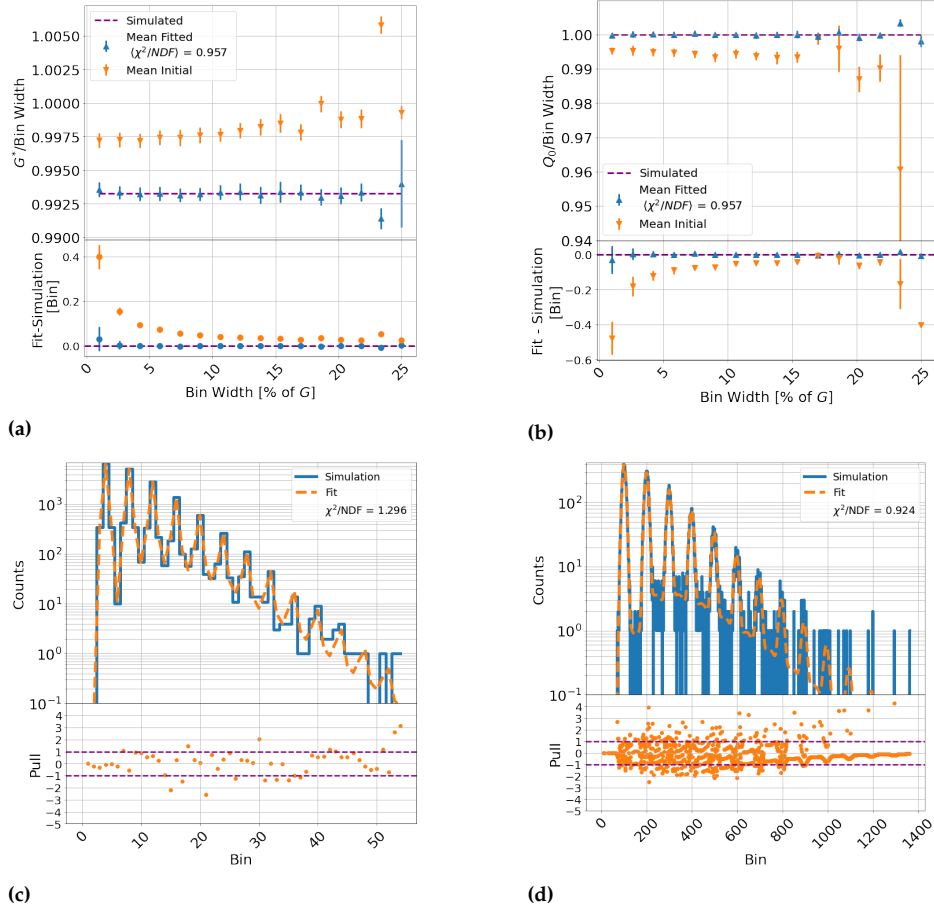


**Figure 4.17:** Comparison of the fit and the prefit values to the simulated values for the scan  $\tau_{Ap} = 4 - 19$  ns. Figure 4.17a shows fitted and prefit values, and below, mean difference and spread of the fitted/prefit values minus the simulated values vs. the simulated values. Figures 4.17b and 4.17c show simulated charge spectrum and fit results, and below the pulls for  $p_{Ap} = 0.0027$  and  $p_{Ap} = 0.0818$ , respectively.

**Comments:**

- ▶ Figure 4.17a shows that the afterpulse time constant can be reconstructed with around 1 ns precision.
- ▶ Figures 4.17b-Figure 4.17c indicate that the model is able to describe the regions between the peaks caused by afterpulses.

### Bin Width Scan of Fits With Peak0Tron To Simulation In The Range 0.01 – 0.25 G



**Figure 4.18:** Comparison of the fit and the prefit values to the simulated values for the scan bin width = 0.01 – 0.25 G. Figure 4.18a shows ratios of the mean fitted and prefit values for  $G^*$  to the simulated  $G$  values in bin widths, and below, the mean differences and spreads of the fitted/prefit values minus the simulated values. Figure 4.18b shows ratios of the mean fitted and prefit values for  $Q_0$  to the simulated values, and below, the mean differences and spreads of the fitted/prefit values minus the simulated values. Figure 4.18c and Figure 4.18d show simulated charge spectrum and fit results, and below the pulls for the bin width for 0.01 G and 0.25 G, respectively.

#### Comments:

- Figure 4.18a It can be seen that for a bin width of 0.01 G, the  $G^*$  bias is less than 0.05 bins, which corresponds to a relative bias of  $5 \times 10^{-4}$ .
- Figure 4.18b show that, independent of the bin width, the fits determine  $Q_0$  with an accuracy of a small fraction of the bin width;
- Figures 4.18c-4.18d indicate that the model can describe distributions with a variety of bin widths.

#### 4.4.1.2 Systematic Biases and Statistical Uncertainties Of Fit

So far, only the uncertainties of the parameters scanned have been presented. However, changing one parameter in the simulation may influence the uncertainties of other parameters. Figures 4.20-4.22 show the biases and statistical uncertainties on  $Q_0$ ,  $G^*$ ,  $\mu$ ,  $\lambda$ , and  $DCR$  for the scans of bin width,  $\mu$ , and  $DCR$ . Also, it is expected that the number of events used,  $N_{\text{events}}$ , will increase the accuracy of the fit tool. Figure 4.19 shows the dependence of the statistical uncertainty of the fitted parameters on the number of entries in the spectrum, for the baseline-parameter set.

##### Summary Comment To Figures 4.19-4.22

With few exceptions, the biases and systematic uncertainties are small. The increased error on  $DCR$  and  $\lambda$  in the  $\mu$  and  $DCR$  scans can be attributed to having proportionally fewer events from which these parameters may be estimated as they increase (i.e. fewer events per photoelectron peak). Also, it is noted that the accuracy of the parameters increases with the number of events as expected.

#### 4.4.1.3 CPU Time for the Fit and the Prefit

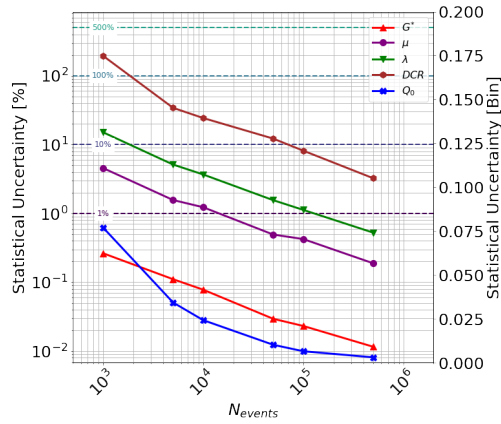
Figure 4.23 shows the mean CPU time and its spread for 100 fits to spectra simulated with the baseline parameters using an Intel® Xeon® E5-2698 v4 CPU operating at 2.2 GHz for scans of  $\mu$ ,  $DCR$  and the number of events.

##### Summary Comment to Figure 4.23

The prefit time never exceeded 0.5 s. The mean overhead for the prefit is 0.15 s, and for the fit 16 s. The fit time increases exponentially with  $\mu$  and linearly with  $DCR$ .

This section, therefore, demonstrates that PeakOTron can fit and precisely describe the simulated SiPM spectra over a wide range of parameter values and reconstruct the parameters with high accuracy, and can be used to fit spectra in a reasonable timescale.

Statistical Uncertainty of Fitted Parameters in Peak0Tron vs. Number of Events

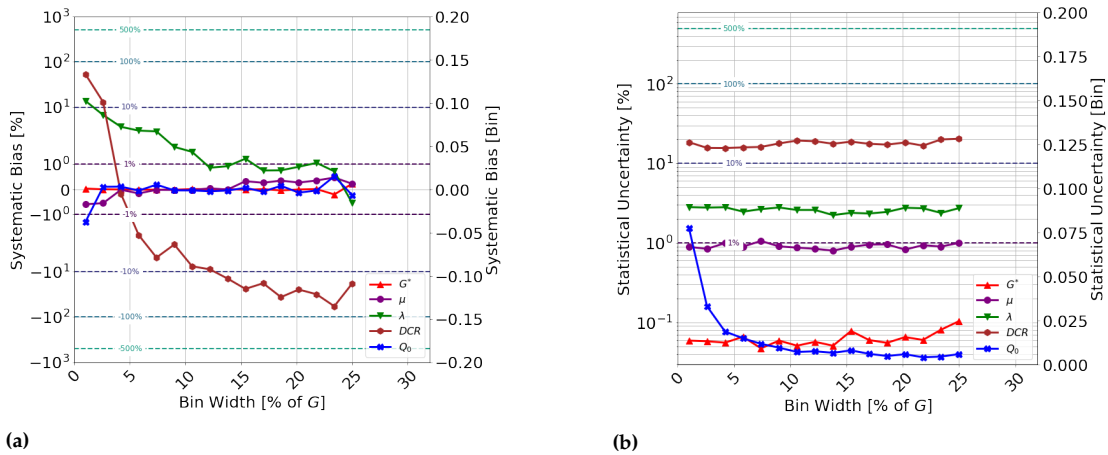


**Figure 4.19:** Statistical uncertainty of fitted parameters as a function of  $N_{\text{events}}$ , the number of entries in the spectrum, for the baseline-parameter set. For  $Q_0$ , the uncertainties are shown in bin widths, shown on the scale on the right, and for  $\mu$ ,  $G^*$ ,  $\lambda$  and  $DCR$ , as a percentage of their values, shown on the scale on the left.

Comments:

- As expected, the variances of parameters follow approximately a  $1/\sqrt{N_{\text{events}}}$  dependence.

Systematic Bias and Statistical Uncertainty of Fitted Parameters in Peak0Tron vs. Bin Width

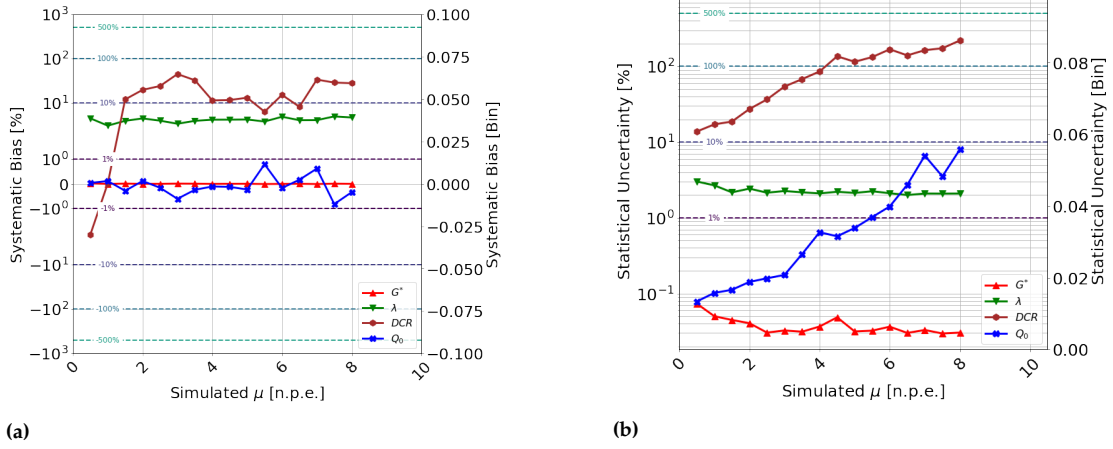


**Figure 4.20:** Figure 4.20a and Figure 4.20b show the bias and statistical uncertainty of  $Q_0$ ,  $G^*$ ,  $\mu$ ,  $\lambda$ , and  $DCR$  for the scans of the bin width. For  $Q_0$ , the uncertainties are shown in bin widths, shown on the scale on the right, and for  $G^*$ ,  $\mu$ ,  $\lambda$  and  $DCR$ , as a percentage of their values, shown on the scale on the left.

Comments:

- Figure 4.20a shows that the bin width has a significant influence on the determination of  $\lambda$  and  $DCR$ , but hardly affects  $\mu$ ,  $G^*$  and  $Q_0$ .

**Systematic Bias and Statistical Uncertainty of Fitted Parameters in Peak0Tron vs.  $\mu$**

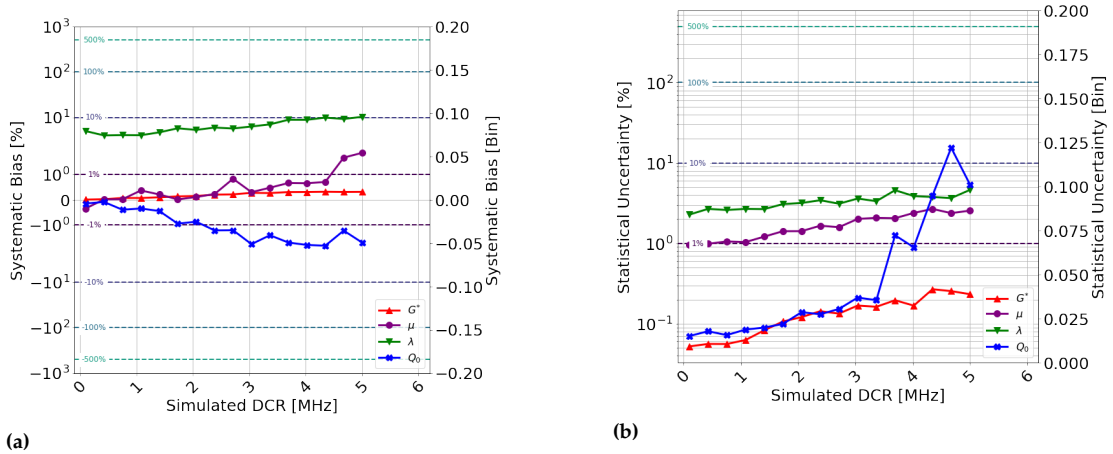


**Figure 4.21:** Figure 4.21a and Figure 4.21b show the bias and statistical uncertainty of  $Q_0$ ,  $G^*$ ,  $\mu$ ,  $\lambda$ , and  $DCR$  for the scans of  $\mu$ . For  $Q_0$ , the uncertainties are shown in bin widths, shown on the scale on the right and for  $G^*$ ,  $\mu$ ,  $\lambda$  and  $DCR$ , as a percentage of their values, shown on the scale on the left.

**Comments:**

- From Figure 4.21b it is concluded that a change in  $\mu$  influences significantly the determination of  $DCR$ , but hardly of  $Q_0$ ,  $G^*$ ,  $\mu$ , and  $\lambda$ .

**Systematic Bias and Statistical Uncertainty of Fitted Parameters in Peak0Tron vs.  $DCR$**



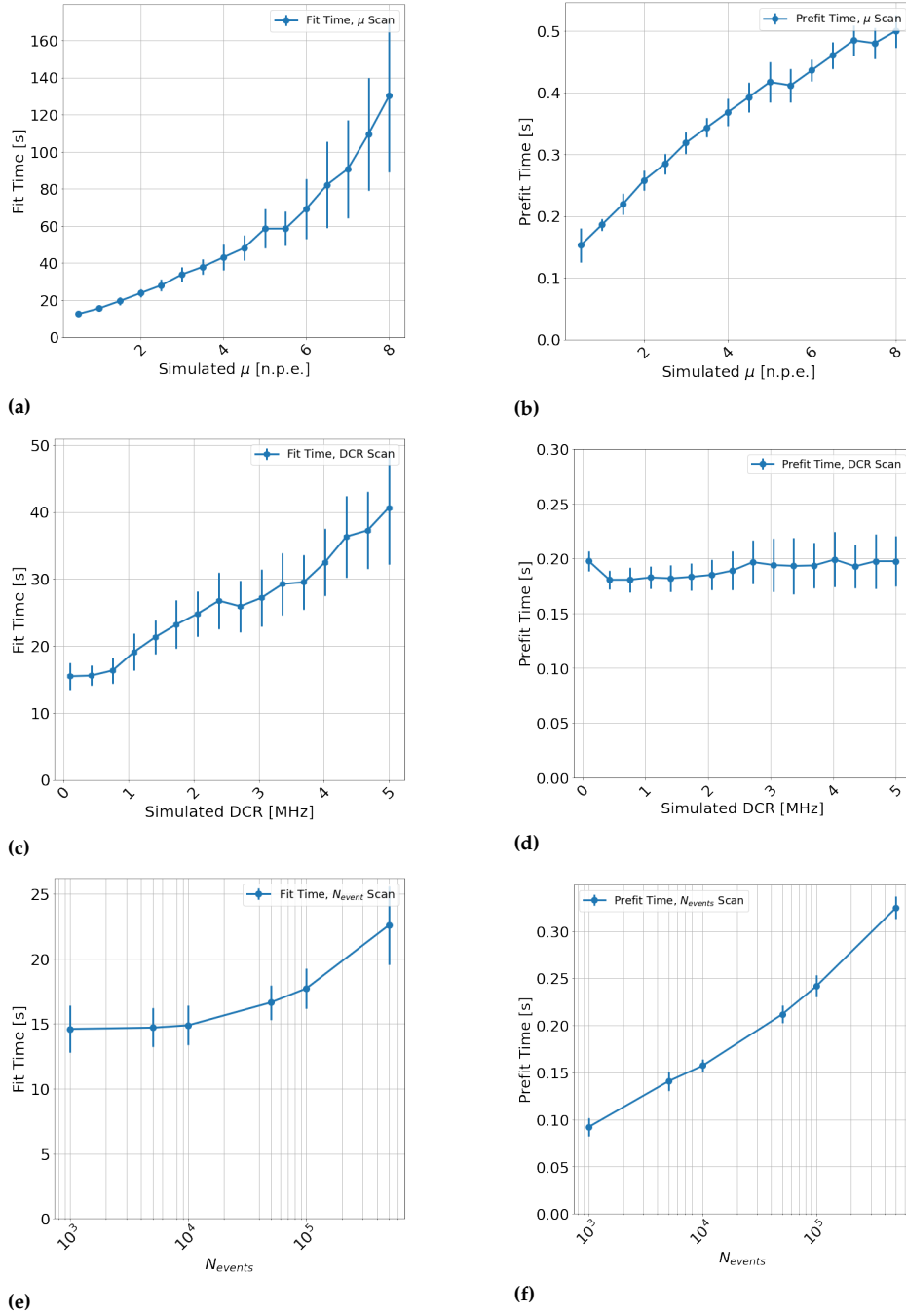
**Figure 4.22:** Figure 4.22a and Figure 4.22b show the bias and statistical uncertainty of  $Q_0$ ,  $G^*$ ,  $\mu$ ,  $\lambda$ , and  $DCR$  for the scans of  $\mu$ . For  $Q_0$ , the uncertainties are shown in bin widths, shown on the scale on the right, and for  $G^*$ ,  $\mu$ ,  $\lambda$  and  $DCR$ , as a percentage of their values, shown on the scale on the left.

**Comments:**

- From Figure 4.22a and Figure 4.22b and shows that the biases and statistical uncertainties remain small when increasing  $DCR$ .



## Fit and Prefit Time in Peak0Tron For Different Parameters



**Figure 4.23:** Figure 4.23a-4.23b, Figure 4.23c-4.23d and Figure 4.23e-4.23f show the mean fit times and their spread for the 100 simulations of  $2 \times 10^4$  events each. The mean fit times and their spread shown as points and error bars for the scans of  $\mu$ , DCR, and of  $N_{\text{events}}$  on each row, respectively. The left column shows the fit time and the right column shows the prefit time.

## Comments:

- The mean CPU time per fit increases approximately exponentially with  $\mu$  and linearly with  $\mu_{\text{dark}}$  (approximately according to  $\langle t_{\text{fit}} \rangle \propto e^{0.31 \cdot \mu} \cdot (0.1 + \mu_{\text{dark}})$ ).
- The prefit time increases linearly with  $\mu$ , logarithmically with the number of events, and is approximately independent of  $\mu_{\text{dark}}$ .
- As expected for a binned log-likelihood fit, the fit time increases only slowly with the number of events.

SiPM	Area [mm <sup>2</sup> ]	Pixel size [μm]	Pixels	PDE [%]	G [q <sub>0</sub> ]	DCR [kHz mm <sup>-2</sup> ]	V <sub>off</sub> [V]
PM1125NS-SB0	1.2 × 1.2	25	2304	25	1.5 × 10 <sup>6</sup>	typ.: 210	27.3
S13360-1325PE	1.3 × 1.3	25	2668	30	0.7 × 10 <sup>6</sup>	typ.: 70, max.: 210	51.1

**Table 4.5:** Manufacturers' specifications of the Ketek SiPM PM1125NS-SB0 [155] and the Hamamatsu MPPC S13360-1325PE [156]. Area refers to the photo-sensitive area. The values for PDE, G (in units of elementary charges,  $q_0$ ), DCR and V<sub>off</sub> refer to a temperature of 25 °C and an over-voltage of 5 V. They are typical values that may differ among SiPMs. The photon-detection efficiency (PDE) refers to a wavelength of 430 nm for the Ketek and to 450 nm for the Hamamatsu SiPM.

SiPM	$r_f$	$\tau_f$ [ns]	$\tau$ [ns]	V <sub>off</sub> [V]
PM1125NS-SB0	0.04 ± 0.01	0.92 ± 0.05	34.0 ± 0.8	27.17 ± 0.01
S13360-1325PE	0.24 ± 0.01	1.62 ± 0.02	22.0 ± 0.6	51.57 ± 0.01

**Table 4.6:** SiPM pulse-shape parameters determined from fits to the transients. The measured transients are shown as continuous blue lines in Figure 4.24. The function  $A \cdot \left( (1 - r_f) \cdot e^{-t/\tau} + r_f \cdot e^{-t/\tau_f} \right)$  is fitted to the data. Here,  $\tau$  and  $\tau_f$  are the slow and fast time constants, respectively, and  $r_f$  is the fractional contribution of the fast component. The voltages at which the Geiger discharge stops, V<sub>off</sub>, are obtained from the fits of the effective gain vs voltage shown in Figure 4.27.

## 4.4.2 Fits to Experimental Data

In this section, the PeakOTron fit program is applied to charge spectra taken from the

### 4.4.2.1 SiPMs and Setup

Measured spectra from two SiPMs have been analysed: a Hamamatsu MPPC S13360-1325PE [156] and a Ketek SiPM PM1125NS-SBO [155]. Both have a pixel size of 25 μm. Their properties are summarised in Table 4.5.

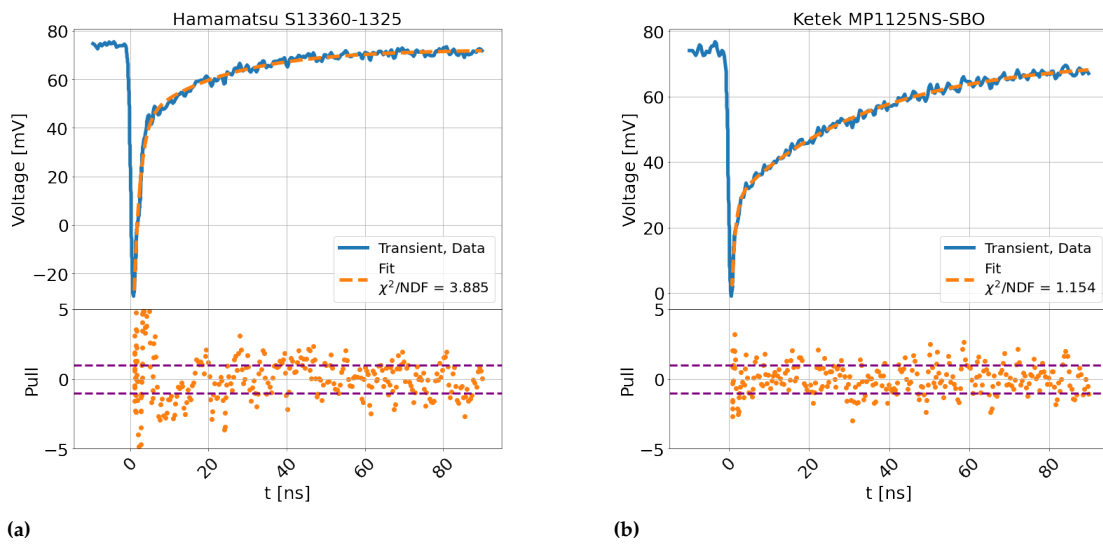
Charge measurements were performed with the SiPM educational kit from CAEN [134, 157]. It consists of a power supply and amplification unit (PSAU). The SiPMs are soldered to custom printed circuit boards that can be plugged into the PSAU. The PSAU consists of an AC-coupled amplifier, a leading-edge discriminator and a coincidence logic. After amplification, the pulses are digitised by a DT5720A CAEN Desktop Digitiser, with a sampling frequency of 250 MS/s.

An LED driver powers an LED, which emits light of approximately 400 nm wavelength with a sub-nanosecond rise time and a 5 ns decay time. The light is transported to the SiPM by an optical fibre. The CAEN kit with the SiPM is located in a light-tight Al housing, which also serves as electric shielding.

Example transients of the two SiPMs from single Geiger discharges are displayed in Figure 4.24. They show a fast and a slow time component, and the sum of two exponentials can fit them. The fit results are summarized in Table 4.6.

For obtaining the charge spectra, the transients are integrated during a gate with the width  $t_{\text{gate}} = 104$  ns, starting 4 ns before the start of the signal from the light pulse. Figure 4.25 shows charge spectra for low-intensity illumination for a range of bias voltages for both SiPMs.

### Typical Waveforms of Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

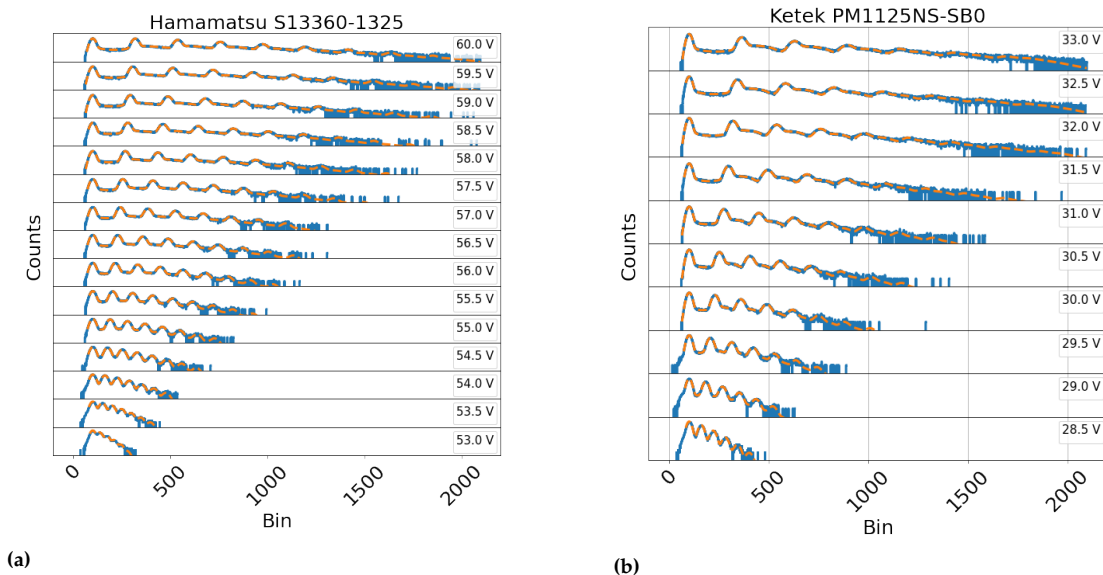


**Figure 4.24:** Figures 4.24a and 4.24b show typical waveforms (voltage vs time) of the Hamamatsu MPPC S13360-1325PE and the Ketek SiPM PM1125NS-SBO for single Geiger discharges are shown as blue continuous lines, respectively. A dashed square corresponds to 20 ns in the horizontal and 20 mV in the vertical direction. The sum of two exponentials fits the transients. The assumed measurement uncertainties are 1 mV. The fit results are shown as dashed orange lines, and the fitted parameters are reported in Table 4.6. Pulls are shown on the subplot.

#### Comments:

- the pulls on each figure shown indicate good agreement with the double-exponential structure of the SiPM pulse.

### Fitted Charge Spectra vs. Over-voltage For Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO



**Figure 4.25:** Figure 4.25a and 4.25b show measured charge spectra in logarithmic scale of the Hamamatsu MPPC S13360-1325PE and of the KETEK SiPM PM1125NS-SBO, respectively, illuminated with low-intensity light, for increasing bias voltages. The results of the Peak0Tron fits are shown as orange dashed lines.

#### Comments:

- Generally, good agreement is observed for all studied spectra, as indicated by the agreement of the dashed orange line with the blue histogram.

SiPM	$\mu_0$ [p.e.]	$V_0$ [V]	$V_{bd} - V_{off}$ [V]	$\tau$ [ns]	$V_{off}$ [V]
PM1125NS-SB0	$0.87 \pm 0.01$	$2.22 \pm 0.03$	$0.13 \pm 0.01$	$34.0 \pm 0.8$	$27.15 \pm 0.01$
Si3360-1325PE	$1.37 \pm 0.01$	$2.91 \pm 0.03$	$0.31 \pm 0.01$	$22.0 \pm 0.6$	$51.58 \pm 0.01$

**Table 4.7:** Values of the parameters from the fits of Equation 4.22 to the data of Figure 4.28. The values for  $\tau$  and of  $V_{off}$  are taken from Table 4.6.

#### 4.4.2.2 Peak0Tron Fits

The results of the Peak0Tron fits to the measured spectra presented in Figure 4.25, are shown in the Figures 4.27 to 4.36, as a function of over-voltage. The over-voltage is the difference of the bias voltage and  $V_{off}$ , where  $V_{off}$ , which is the voltage at which the Geiger discharge stops. This quantity is obtained from the intercept of a straight-line fit of  $G^*$  as a function of bias voltage presented in Figure 4.27. The breakdown voltage is the voltage at which  $V_{bd}$ , is extracted from a fit to the mean number of photon-induced primary Geiger discharges,  $\mu$ , as a function of over-voltage. This quantity is expected to saturate with increasing overvoltage, as expected from the voltage-dependence of  $PDE$ . The voltage dependence of  $\mu$  Equation 4.22.

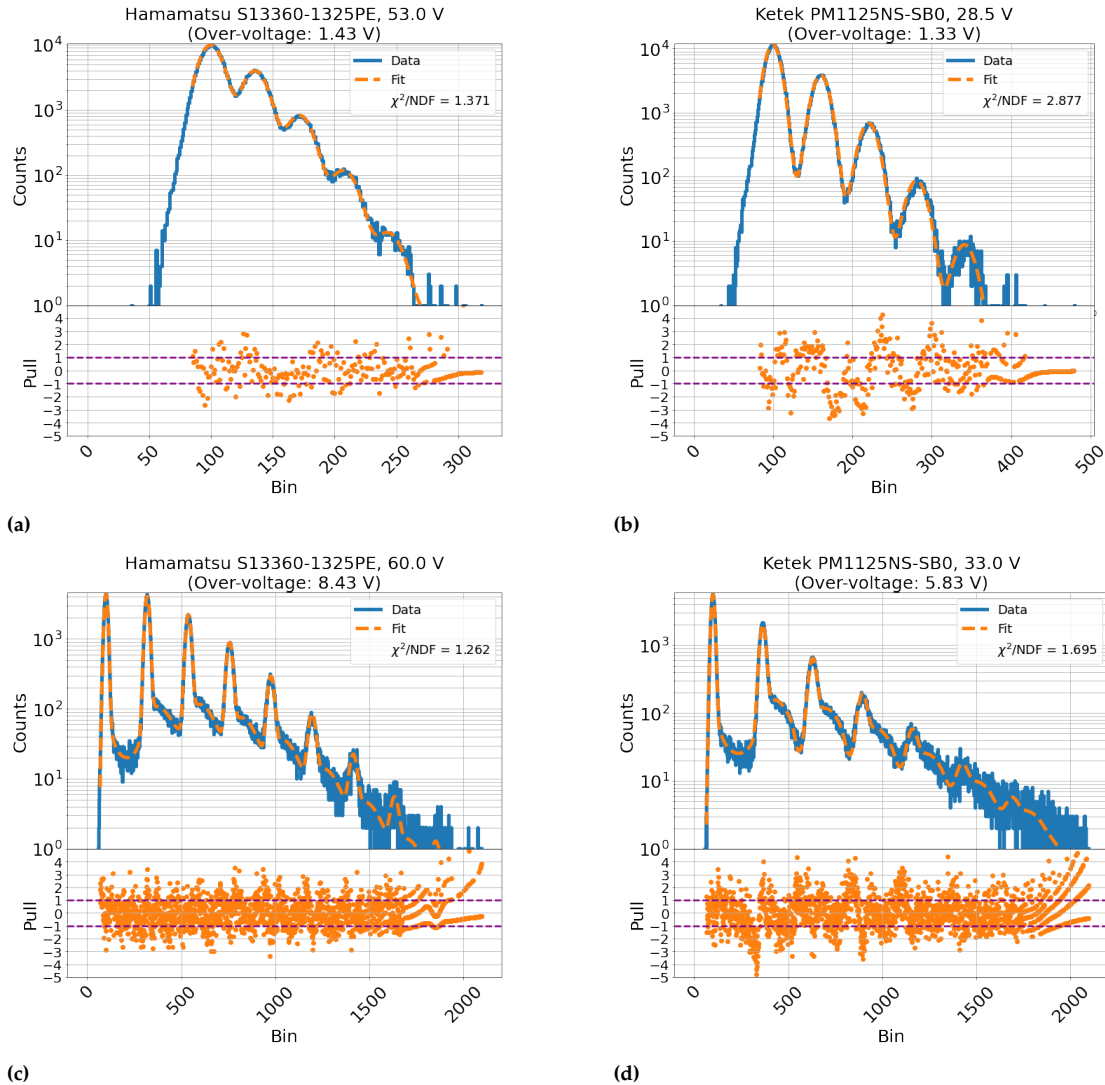
$$\mu(V) = \mu_0 \cdot \left(1 - e^{-\max(V-V_{bd}, 0)/V_0}\right), \quad (4.22)$$

where  $\mu_0$  is the  $\mu$ -saturation value,  $V_{bd}$  the breakdown voltage and  $V_0$  a parameter which characterises the voltage dependence.

#### Summary Comment To Figures 4.27-4.36

The parameters extracted with Peak0Tron from the Hamamatsu MPPC Si3360-1325PE and Ketek SiPM PM1125NS-SBO under test are in agreement with expectations from the manufacturer as outlined in Table 4.5 and with expected voltage dependences, including the dark count rate and afterpulse probability and time constant. The  $\chi^2/NDF$  indicates generally good agreement with the SiPM spectra.

### Charge Spectra of the Hamamatsu MPPC S13360-1325PE and the Ketek SiPM PM1125NS-SBO SiPMs, Illuminated With Low-Intensity Light For The Lowest and Highest Voltages Of The Voltage Scans

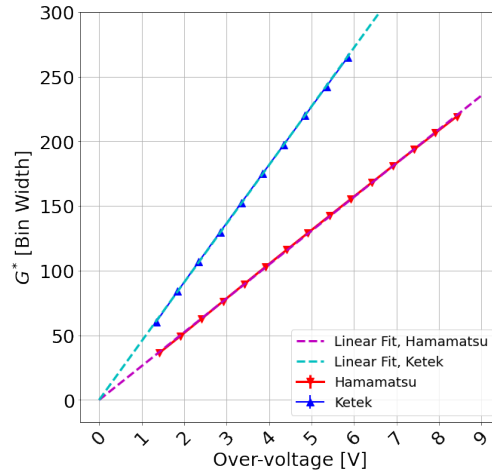


**Figure 4.26:** Measured (blue continuous lines) and fitted (orange dashed lines) charge spectra of the Hamamatsu MPPC S13360-1325PE and the Ketek SiPM PM1125NS-SBO SiPMs, illuminated with low-intensity light for the lowest and highest voltages of the voltage scans. In the lower subfigures, the pulls, the differences measured minus fitted divided by the square root of the expected number of counts, are displayed. Figures 4.26a and 4.26c show the results for the Hamamatsu MPPC operated at 53 V and 60 V, respectively, and Figures 4.26b and 4.26d for the Ketek SiPM, operated at 28.5 V and 33 V, respectively. Note that for low over-voltages the spectra are only fitted above  $Q_0 - 2 \cdot \sigma_0$ .

#### Comments:

- It is noted that the spectra at low over-voltages show non-Gaussian tails for charge values below the pedestal  $Q_0$ . Using the iterative procedure described in Section 4.3.6, the fit is only performed for charge values  $Q \geq Q_0 - n \cdot \sigma_0$ . A value of  $n = 2$  is found for low over-voltages, and  $n$  increases to 4 at high over-voltages.
- Else, as in Figure 4.25.

### $G^*$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

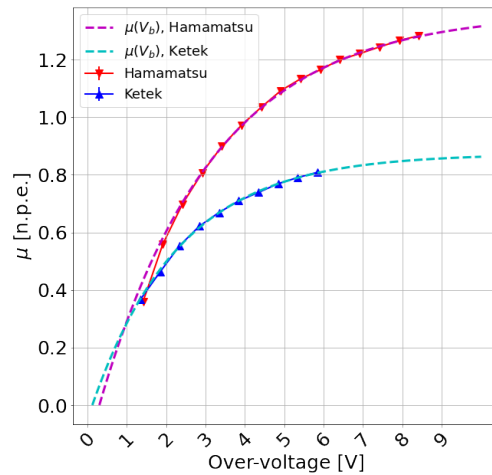


**Figure 4.27:**  $G^*$  as a function of over-voltage for the Ketek SiPM and the Hamamatsu MPPC, shown as blue and red triangles, respectively. The over-voltages are the differences of the bias voltages and  $V_{\text{off}}$ , where  $V_{\text{off}}$  is obtained from the intercepts of the straight-line fits to  $G^*$  as a function of bias voltage. The dashed lines cyan and magenta lines represent the straight-line fits shown as a function of over-voltage.

#### Comments:

- ▶ The gain is expected to increase linearly with over-voltage, as shown in Equation 2.14. Therefore,  $V_{\text{off}}$  can be extracted from the offset of a linear fit.
- ▶ The measured  $V_{\text{off}}$  values obtained this way, which are reported in Table 4.6, agree with the values from the producers.

### $\mu$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO



**Figure 4.28:** The mean number of photon-induced primary discharges,  $\mu$ , in units of n.p.e. as a function of over-voltage for the Ketek SiPM and the Hamamatsu MPPC, shown as blue and red triangles, respectively. The fits using Equation 4.22 are shown by solid lines, and the extrapolations by dashed lines. The fit parameters are given in Table 4.7.

#### Comments:

- ▶ As expected from the voltage dependence of the photon detection efficiency,  $\mu$  increases rapidly at low over-voltages and then flattens. The cyan and red lines indicate good agreement is observed with the model of Equation 4.22, corresponding to RMS deviations between the fit and the  $\mu$  values from Peak0Tron of around  $5 \times 10^{-3}$  p.e.
- ▶ Table 4.7 shows the values of the parameters determined by the fit. It is noted that  $V_{\text{bd}} > V_{\text{off}}$ , i.e. the breakdown voltage is larger than the voltage at which the Geiger discharge is quenched. Similar observations have been reported in [93]. A difference is expected because a Geiger avalanche takes more energy to initiate than sustain.

### Reduced $\chi^2$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

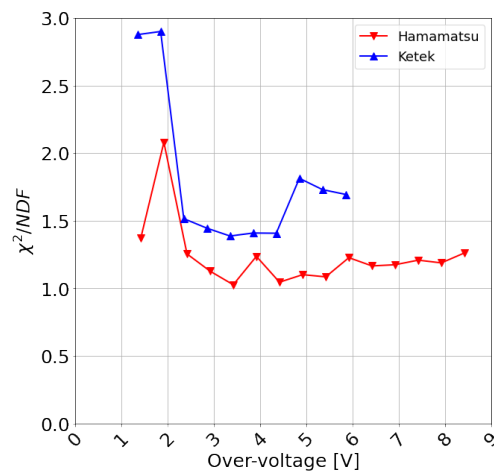


Figure 4.29:  $\chi^2/\text{NDF}$  as a function of over-voltage.

#### Comments:

- ▶ The degradation of the fit at low over-voltages is a result of non-Gaussian pedestals, as indicated by Figure 4.26. Possible causes of non-Gaussian tails are low frequency (multiple of 50 Hz) noise or dark pulses preceding the gate at times significantly earlier than  $-\tau$ , for which the AC-coupling causes a negative baseline shift at the time of the gate.
- ▶ It is seen that for the Ketek SiPM the  $\chi^2/\text{NDF}$  increases to about 1.8 for over-voltages exceeding 4.5 V. The reason for this worsening of the fit quality is not understood.

### $\lambda$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

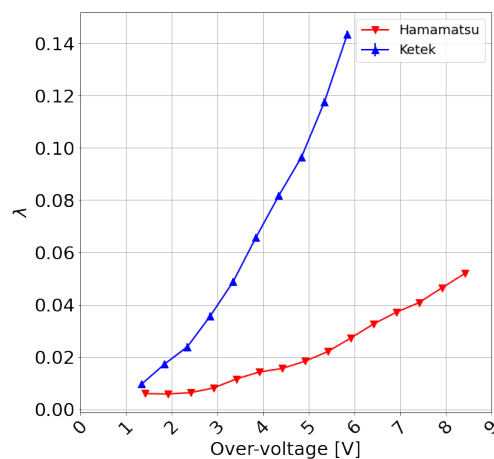


Figure 4.30:  $\lambda$  as a function of over-voltage, as in Figure 4.28.

#### Comments:

- ▶ The prompt-cross-talk parameter  $\lambda$  is observed to increase with over-voltage.
- ▶ It is noted that for the Ketek SiPM, the value of  $\lambda$  is larger than for the Hamamatsu MPPC.

### $Q_0$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

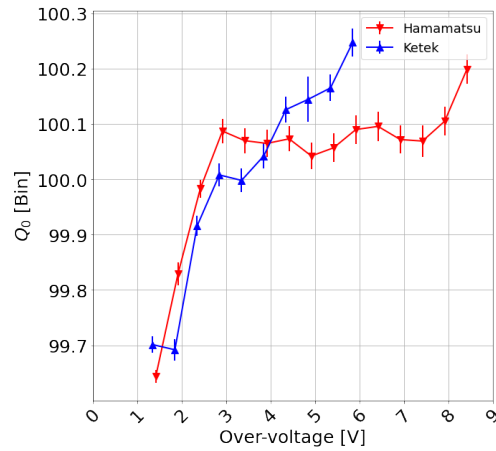


Figure 4.31:  $Q_0$  as a function of over-voltage, as in Figure 4.28.

#### Comments:

- A small increase of less than one bin is observed with over-voltage, which is ascribed to the non-Gaussian tails of the pedestal peak.

### $\sigma_0$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

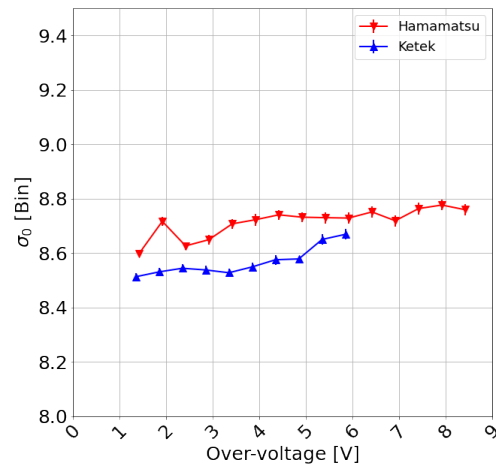


Figure 4.32:  $\sigma_0$  as a function of over-voltage, as in Figure 4.28.

#### Comments:

- A small increase of less than one bin is observed with over-voltage, which is ascribed to the non-Gaussian tails of the pedestal peak.
- The value of  $\sigma_0$ , which is about 8.6 bins for both SiPMs, is ascribed to the electronics noise of the setup.



### $\sigma_1/G^*$ As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

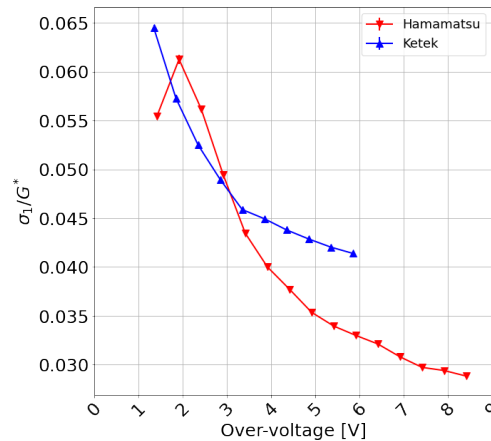


Figure 4.33: The relative gain spread,  $\sigma_1/G^*$  as a function of over-voltage, as in Figure 4.28.

#### Comments:

- It is observed that the relative gain spread decreases with over-voltage for both SiPMs, but more so for the Hamamatsu SiPM.
- As the width of the  $k^{\text{th}}$  photoelectron peak is  $\sigma_{k\text{pe}} = \sqrt{\sigma_0^2 + k \cdot \sigma_1^2}$ , the decrease of  $\sigma_1/G^*$  means that the ability to separate n.p.e. peaks improves significantly with over-voltage. This observation can also be deduced from Figure 4.26.

### DCR As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

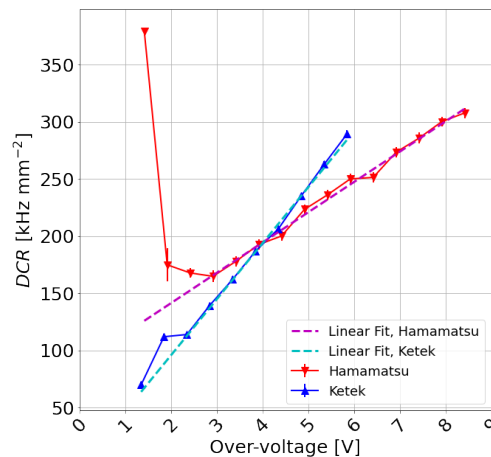


Figure 4.34: DCR as a function of over-voltage, as in Figure 4.28.

#### Comments:

- The value of DCR, increases linearly from 160 kHz mm<sup>-2</sup> at an over-voltage of 2.9 V to 310 kHz mm<sup>-2</sup> at 8.4 V for the Hamamatsu MPPC, and from 140 kHz mm<sup>-2</sup> at 2.8 V to 310 kHz mm<sup>-2</sup> at 5.8 V for the Ketek SiPM. Thus, DCR per unit area of the Ketek SiPM increases faster with over-voltage than of the Hamamatsu SiPM.
- The values obtained for the DCR at 5 V approximately agree with the manufacturers' values given in Table 4.5.
- At low over-voltages, the determination of DCR is problematic: Its value is mainly derived from the spectrum at the minimum between the pedestal and the one photoelectron peak. If the two peaks overlap, as is the case in Figure 4.26a, the contribution of dark counts to the spectrum cannot be determined reliably. This is apparently is the case for the Hamamatsu MPPC at low over-voltage, where an unphysically high DCR value is seen.

$p_{Ap}$  As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

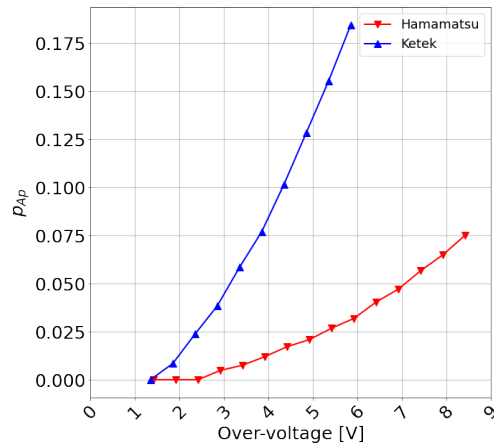


Figure 4.35:  $p_{Ap}$  as a function of over-voltage, as in Figure 4.28.

**Comments:**

- ▶ As expected, the probability of after-pulses,  $p_{Ap}$ , increases with over-voltage. The reason is that the number of charge carriers trapped by states in the Si band-gap is proportional to the number of charge carriers in the avalanche and thus to the gain.
- ▶ The non-linear dependence of  $p_{Ap}$  reflects the fact that the spatial distribution of the trapped charge carriers is approximately uniform, whereas the Geiger-discharge probability depends on position.
- ▶ For the Ketek SiPM,  $p_{Ap}$  is between 4 and 18% in the over-voltage range studied, which is significantly higher than for the Hamamatsu MPPC, where it is between 1 and 7.5% in the wider over-voltage range.

$\tau_{Ap}$  As a Function Of Overvoltage for The Hamamatsu MPPC S13360-1325PE and Ketek SiPM PM1125NS-SBO

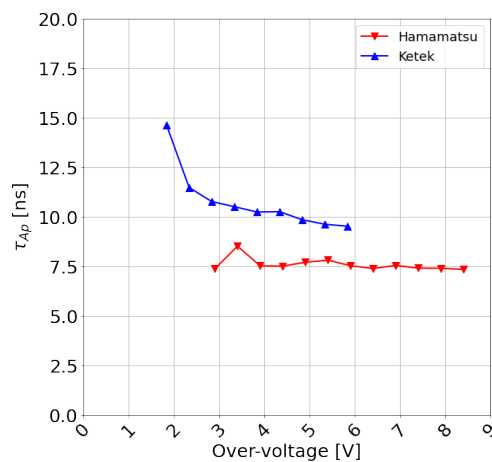


Figure 4.36:  $\tau_{Ap}$  as a function of over-voltage, as in Figure 4.28.

**Comments:**

- ▶ The time constants for after-pulse candidates,  $\tau_{Ap}$ , for both SiPMs have only a minor over-voltage dependence and are quite similar for both SiPMs, about 10 ns for the Ketek SiPM, and 7.5 ns for the Hamamatsu MPPC.
- ▶ It is noted that only few determinations of  $\tau_{Ap}$  are reported in the literature, and most of them do not account for the reduction of Geiger-discharge probability because of the recharging of the pixels. These analyses use the time differences between Geiger discharges and not the charge spectra.
- ▶ In [158], a fast trap with  $\tau_{Ap} = 15$  ns and a slow trap with  $\tau_{Ap} = 82$  ns are reported. [159] finds that the fast trap, with  $\tau_{Ap} \approx 10$  ns, is 2.5 times more effective at trapping charge carriers than the slow trap with  $\tau_{Ap} \approx 100$  ns. Qualitatively, the results from Peak0Tron agree with these findings.

### 4.4.3 Requirements and Limitations of Fits with Peak0Tron

As a caveat, the model makes several assumptions, and the SiPMs and the charge spectra must meet several requirements to successfully determine the SiPM parameters with Peak0Tron.

- ▶ The program assumes the model described in [97]. For SiPMs for which these assumptions are not valid, the parameters determined by the program may have significant systematic biases.
- ▶ The model assumes that a single exponential can describe the SiPM pulse shape. For SiPMs with a fast in addition to the slow component, the contribution of the fast component should be  $\lesssim 25\%$ . This is the case for practically all SiPMs.
- ▶ The peaks of different n.p.e. values have to be clearly resolved. This may not be true for high electronics noise or high DCR from radiation damage or ambient light. In addition, the bin width should be smaller than a quarter of the peak separation, and the determination of  $G^*$ ,  $\sigma_0$  and  $\sigma_1$  becomes unreliable for bin widths larger than  $\sigma_0/2$ .
- ▶ Threshold cuts, which remove a part of the pedestal peak, can result in poor fits and biased results for the gain, the gain spread, the electronics noise, and the pedestal position.
- ▶ The maximum number of dark counts for the time interval  $-t_0$  to  $t_{\text{gate}}$  is set to  $i_{\text{dark}}^{\text{max}} = 6$ . If the probability of more than 6 dark counts in this time interval is significant, this number has to be increased at the cost of additional CPU time.
- ▶ The determination of the after-pulse parameters is sensitive to additional correlated noise that affects the inter-peak regions, such as delayed cross-talk, which is not modelled in Peak0Tron.

## 4.5 Conclusion

A generally-available Python program, called Peak0Tron, is presented, which uses an improved version of the detector response model of [97] to fit SiPM charge spectra. Different to other programs, entire spectra, including the regions in between the photoelectron peaks, are fitted. This allows determining from charge spectra, in addition to gain, the mean number of photon-induced primary Geiger discharges, prompt cross-talk probability, pedestal, electronics noise and gain variations, the dark-count rate, the after-pulse probability and the after-pulse time constant. The initial values of the parameters for the fit are obtained from the charge spectra.

Using charge spectra simulated with the program of [96], it is shown that for a wide range of parameter values, Peak0Tron provides a good description of the spectra and achieves a precise determination of the parameters. Analysing for every parameter set 100 spectra, each with

$2 \times 10^4$  simulated events, bias and statistical uncertainty of the parameters are obtained. Finally, it is shown, that experimental charge spectra of two types of SiPMs measured over a wide range of over-voltages are well described by the model using the parameters obtained with Peak0Tron. The voltage dependencies of the parameters agree with expectations.



# 5D Software Compensation with Machine Learning

# 5

## Contributions By The Author

*The author fully devised the proposed SC model discussed in this section. The author implemented and analysed all subsequent software compensation methods described henceforth.*

## 5.1 Overview of Study

The response of hadronic calorimeters may be described in terms of two components: an electromagnetic component (i.e.  $\pi^0/\eta \rightarrow \gamma\gamma$ , nuclear  $\gamma$ ), and a hadronic component, which contains the remainder of energy depositing processes. The calorimeter response is split into an EM response ( $e$ ), and a HAD response ( $h$ ). Part of the hadronic fraction cannot be detected and is called 'invisible energy' (e.g. neutrinos, nuclear binding energy losses). This fraction also experiences significant stochastic fluctuations from event to event.

A calorimeter for which  $\frac{e}{h} \neq 1$  is called a non-compensating calorimeter. The highly granular sampling calorimeters designed to be used in PF are typically non-compensating because PF calorimeters require a high degree of longitudinal segmentation, which imposes limitations on the structure of the calorimeter.

Compensation describes a method to equalise  $e$  and  $h$ , typically by attenuating  $e$  and enhancing  $h$ . Therefore, software compensation (SC) algorithms are employed for this purpose and operate by estimating the EM fraction of a shower using information measured in each event. Notably, spatial and temporal readout information available from highly granular calorimeter may be used for SC:

- ▶ A hadron shower in a calorimeter exhibits an EM-dominated, energy-dense 'core' that propagates over a short longitudinal and lateral range and a HAD-dominated, diffuse energy-sparse 'halo', which propagates over a wider range [28]. A highly granular calorimeter may be able to resolve these two components and therefore exploit spatial energy density for SC;
- ▶ The number of neutrons produced in nuclear interactions is proportional, on average, to the invisible energy of the hadron shower. Energy deposits from neutrons can be measured indirectly predominantly from  $\gamma$  produced by recoil protons from elastic neutron scattering in hydrogenous active material such as plastic scintillator and nuclear  $\gamma$  from neutron capture. Energy deposits induced by neutrons are delayed by an exponential distribution with a time constant of around 78 ns in steel, which is distinguishable from the instantaneous EM component of the shower [32]. A time-sensitive

5.1 Overview of Study . . . . .	115
Biasing of Software Compensation Models . . . . .	116
Proposed Method to Reduce Bias in Software Compensation Models . . . . .	119
5.2 Methods and Tools . . . . .	121
Neural Network SC Method . . . . .	121
Control SC Method . . . . .	122
Datasets and Training . . . . .	123
5.3 Results . . . . .	140
Simulation Model . . . . .	140
Data Model . . . . .	148
5.4 Conclusion . . . . .	162

hadron calorimeter may therefore exploit temporal information for SC.

Artificial neural network models have already been demonstrated to effectively exploit the spatial development of hadron showers to improve SC. For example, a study performed in [160] demonstrated that a deep neural network was found to improve the response of a highly-granular hadron calorimeter system from  $48\%/\sqrt{E_{\text{particle}}} \oplus 2.2\%$  to  $37\%/\sqrt{E_{\text{particle}}} \oplus 1\%$  using simulation. However, a similar study performed for AHCAL in [161] demonstrated the incapacity of a similar model to interpolate or extrapolate compensation from the limited hadron shower data typically available for such studies. This effect means the model cannot be effectively used in an experimental setting, thereby motivating the development of a model that can exploit the spatial and temporal information from AHCAL and simultaneously remain unbiased to the training particle energies.

In this section, a neural network was designed to perform SC on simulated  $\pi^-$  hadron showers observed with the AHCAL calorimeter, using the spatial and temporal energy density from the event rather than just energy. This information was expected to reduce the effect of stochastic fluctuations by improving sensitivity to the shower development and the neutron fraction of the event. In addition, the neural network was carefully designed to reduce the effect of energy biasing. Finally, the neural network was compared to the standard CALICE software compensation method used as a control. The results were then compared.

In the following analysis, ‘*simulation*’ and ‘*data*’ are defined as event information produced by Geant4 simulation or during the CALICE 2018 Testbeam campaign. Furthermore,  $E$  and  $\sigma_E$  is the mean and standard deviation of the response of the calorimeter to a hadron of  $E_{\text{particle}}$ ,  $a$  describes the combined sampling and stochastic fluctuations experienced by the calorimeter,  $b$  the quality of detector calibration as defined in Equation 2.8.

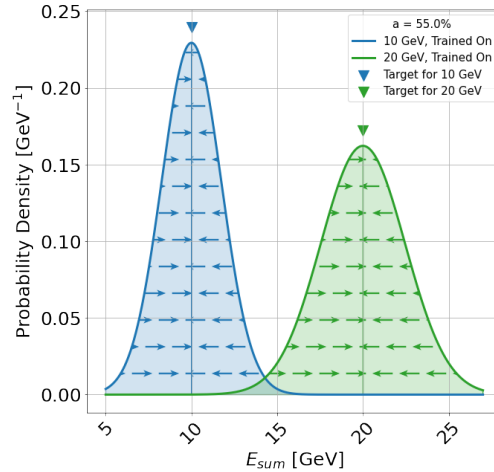
The response of the calorimeter to a hadron shower event in AHCAL is defined as the sum of the energy density deposited in each of the active cells in the event ( $E_{\text{sum}} = \sum_{i=0}^{\text{event}} E_{\text{hit}}$ ). Compensated active cell energy densities are denoted  $\widehat{E}_{\text{hit}}$ , and their sum is the compensated response of the calorimeter,  $\widehat{E}_{\text{sum}}$ . All other symbols are defined in Section 2.5.1.2.

### 5.1.1 Biasing of Software Compensation Models

Software compensation models are typically trained indirectly since the EM/HAD fraction is unknown in a hadron shower event a priori. The resolution of a hadron calorimeter is described in AHCAL according to Equation 2.8. This equation is valid under the assumption of a normally-distributed response (i.e. full shower containment, negligible electronics noise).

Since reductions in the calorimeter response width,  $\sigma_E$  imply compensation due to a smaller stochastic resolution term,  $a$ ,  $\chi^2$  minimisation of the calorimeter response to the known particle energy are commonly used to optimise SC algorithms [59, 80]. However, over-training can occur if the model is exposed to the mean responses of the calorimeter using





**Figure 5.1:** Illustration of the method of most software compensation algorithms. The theoretical calorimeter response distributions of a perfectly calibrated non-compensating hadron calorimeter ( $a = 55\%$ ,  $b = 0\%$ ) for 10 GeV and 20 GeV hadron showers are shown in shaded blue and green, respectively. The coloured arrows indicate the direction of the gradient defined by a mean square loss in the reconstructed energy for each sample.

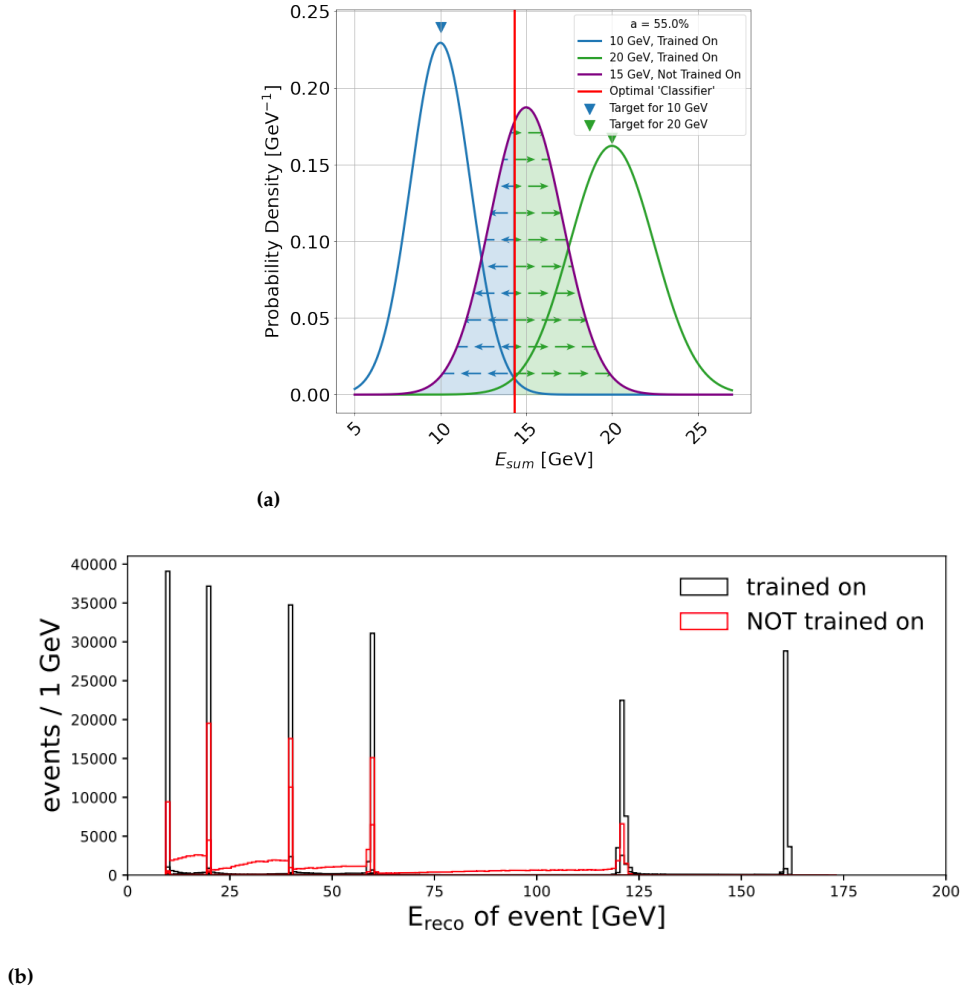
training samples with coarsely binned particle energies that are typically available for this purpose. In particular, two failure modes have been observed for software compensation models applied to AHCAL:

**Interpolation Failure** It was observed in [161] that training using coarsely binned particle momenta in the training set resulted in failures of the model to reconstruct the responses of hadron showers it had not been trained to compensate. A likely hypothesis is that the model learned to ‘classify’ the hadron shower events by learning the mean responses of the training distribution.

For example, suppose an SC model has learned the mean response of the training distributions at 10 GeV and 20 GeV. In that case, a likely result might be that the SC model infers that a 15 GeV sample is either a 10 GeV sample with a large EM fraction or a 20 GeV sample with a large HAD fraction, thereby ‘classifying’ the 15 GeV sample according to the data it was trained on. A visual example of this is Figure 5.2a. The results of [161] pertinent to this hypothesis are presented in Figure 5.2b, which shows that the SC network trained in that study was unable to compensate hadron showers with energies in the ranges between particle energies. A visual explanation of this hypothesis and the results of [161] illustrating this behaviour are presented in Figure 5.3a and Figure 5.3b, respectively.

**Extrapolation Failure** Even when the binning of the momentum of the training sample was increased in [161], biasing to the edges of the training range was observed. The reason for this can also be explained by the algorithm learning the mean particle energies, specifically because there is no penalty to overcompensating the training sample’s uppermost/lowermost bin edges.

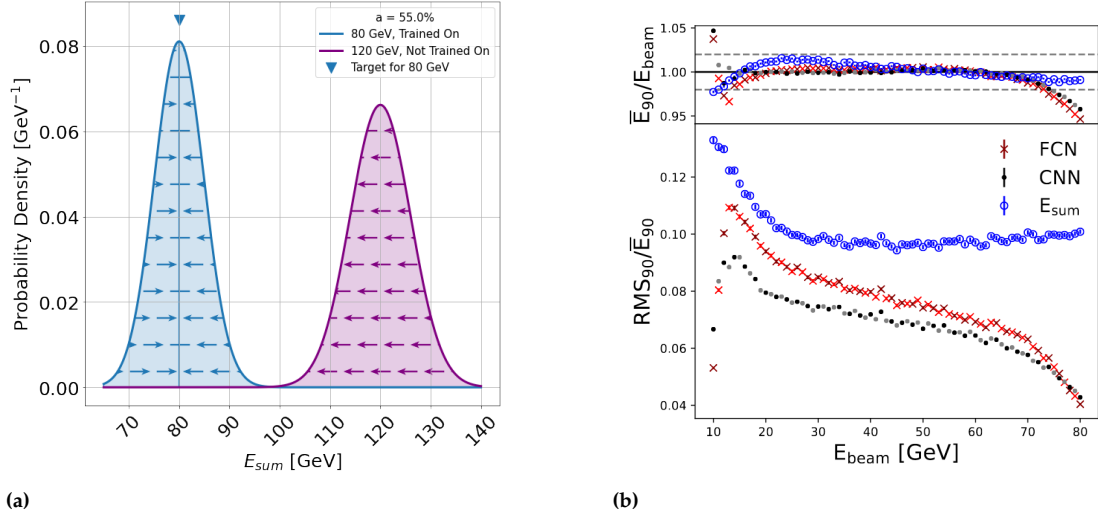
For example, suppose the SC algorithm is trained up to a maximum particle energy of 80 GeV and is allowed to learn the mean response of the 80 GeV. Since no penalty is associated with overcompensating events



**Figure 5.2:** Figure 5.2a shows a hypothesis for a source of interpolation failure in SC models. The purple shaded distribution indicates the response distribution of a standard hadron calorimeter with stochastic term  $a = 55\%$  to 15 GeV hadron showers, to which an SC algorithm has never been exposed. The blue and green lines indicate the same distributions as in Figure 5.1, to which an SC algorithm has been exposed. The best linear discriminant is indicated by the red line, with a possible solution to minimising  $\chi^2$  deviation to the particle energy indicated by the directions of the blue and green arrows beneath the purple curve. Figure 5.2b illustrates the compensated energy responses obtained using a convolutional neural network (CNN) applied to compensation of AHCAL hadron shower data in [161]. The black and red lines indicate the compensated response distributions of the algorithm presented there, applied to a validation sample of  $\pi^-$  hadron shower events used for training and not used for training, respectively. The black lines indicate the correctly compensated samples. The red lines show that the model cannot interpolate compensation between those training energies, meaning it cannot be used experimentally.

with a higher/lower response than the mean response than the training sample's uppermost/lowermost bin edges, the model will likely learn to overcompensate these training energy bins. For instance, according to this hypothesis, a 120 GeV sample trained using this model will always be weighted down in energy.

In summary, biasing effects are often observed in data-driven SC models due to biasing of the model to features of the training dataset. These limitations would not be a problem if unlimited training data were available. However, it is costly in time and resources to generate hadron shower simulations of arbitrary particle energies and to collect experimental data at a fine granularity and ever higher energies. This motivates a method to exploit the available dataset without biasing.



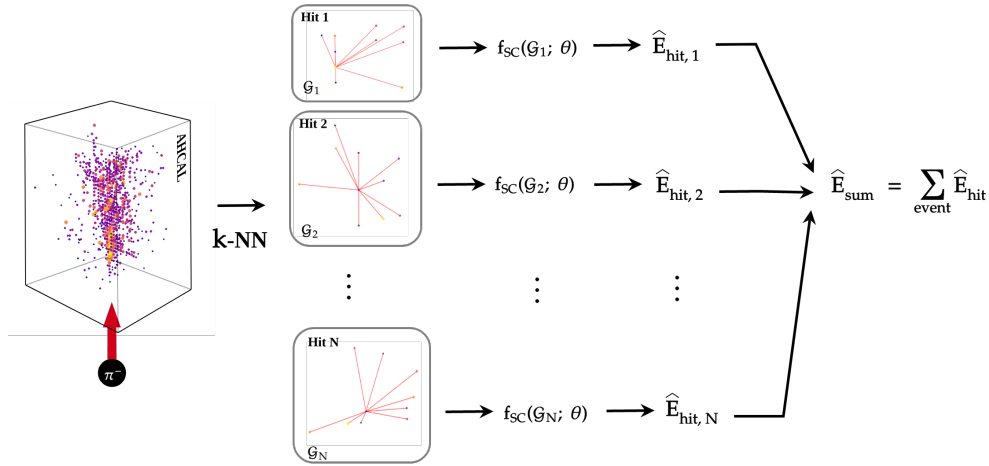
**Figure 5.3:** Figure 5.3a presents a hypothesis for a source of extrapolation failure in SC models. The shaded blue distribution represents a response distribution of a standard hadron calorimeter with stochastic term  $a = 55\%$  to 80 GeV hadron showers, to which an SC algorithm has been exposed. The purple shaded distribution indicates the response distribution of the same calorimeter to 120 GeV hadron showers, to which the same SC algorithm has not been exposed. The blue and purple arrows indicate the expected behaviour of the algorithm for the 80 GeV and 120 GeV distributions, respectively. Figure 5.3b illustrates the result of a fully connected neural network (FCN) and convolutional neural network (CNN) applied to the compensation of AHCAL hadron shower simulation. At the edges of the validation sample, severe biasing of the model is observed. A rapid decrease in resolution may be observed in the bottom plot below 15 GeV and above 60 GeV, indicating unphysical resolution, and the non-linear response produced by the algorithm below 15 GeV and above 60 GeV. This result means the model has biased to the training range of particle momenta and cannot be used experimentally.

### 5.1.2 Proposed Method to Reduce Bias in Software Compensation Models

For biasing to be reduced, constraints must be introduced to the model to prevent the SC model from learning the properties of the training calorimeter response distributions while allowing the model to attenuate or enhance active cell energies.

The simplest way to achieve this goal is to explicitly blind the model to the response distribution entirely by exploiting the highly-granular properties of the AHCAL calorimeter. This model would estimate a cell's EM/HAD fraction based on the energy density of the surrounding cells for every active cell in an event, independently of the overall energy distribution observed by the calorimeter. Knowledge of the overall distribution of energy of a hadron shower, and therefore the overall response of the calorimeter to the hadron, is no longer necessary for the SC algorithm to succeed.

One way of achieving this is applying  $k$ -nearest neighbours ( $k$ -NN) clustering to the active cells of AHCAL during an event and using a graph neural network. This method finds the nearest  $k$  neighbours to each active cell, representing the local energy density surrounding that cell. The hadron shower could then be 'split' into a set of clusters. In a machine learning context, each batch element is not an entire hadron shower event, which would be typical in most methods, but instead a cluster surrounding each cell. Each cluster is associated back to its hadron shower after processing by the model. The SC model would then determine an independent attenuation or enhancement of the active cell energy based on the energy distribution of that cluster, independently of all other clusters. The model, therefore, acts to classify each hit in an



**Figure 5.4:** Illustration of a method for software compensation by which biasing may be reduced. First, a hadron shower of  $N$  measured with AHCAL, indicated by the event display on the left, is decomposed into a series of  $k$ -NN cluster graphs, indicated by the vertices and red lines, indicating edges between them, denoted  $\mathcal{G}$ . At this stage, each active cell is now represented as a local neighbourhood graph,  $\mathcal{G}_i$ , where  $i$  is the index of the active cell. In this diagram, the 9 nearest neighbours are shown for illustration. Next, for each cluster, a SC model,  $f_{SC}(\mathcal{G}_i; \theta)$ , is applied to each graph, where  $\theta$  is the model's free parameters, producing an attenuated or enhanced calorimeter response to the kernel cell of  $\mathcal{G}_i$ ,  $\hat{E}_{hit,i}$ . The sum of the individually weighted active cells is then the compensated calorimeter response,  $\hat{E}_{sum}$ .

event as part of the EM/HAD fraction based on its local energy density and attenuates or enhances the cell energy accordingly. In summary, the model is designed to achieve compensation by cell classification rather than the overall response of the calorimeter or event properties. Furthermore, this model cannot learn undesirable features, like the number of active cells or reconstructed energy of the hadron shower by design. The fundamental idea is illustrated in Figure 5.4.

It is noted that a similar technique was applied to overcome bias in [162]. In this study, the goal was to produce photorealistic images from video game footage. In a first attempt, artefacts were initially found due to global differences in feature distributions between target domains (e.g. amount of vegetation). The study significantly alleviated the effect by using small patches of images from both domains to train the model. In the case of the presented study, it is argued that software compensation is analogously biased and can also benefit from the same subdivision as in that study.

From a physics perspective, it may be argued that the total reconstructed hadron shower energy is a poor indicator of the distribution of energy-depositing processes caused by stochastic fluctuations. Instead, the EM and HAD fractions may be inferred entirely from local energy density in a highly granular calorimeter, governed by the particular energy-depositing processes surrounding a particular active cell of the hadron shower.

## 5.2 Methods and Tools

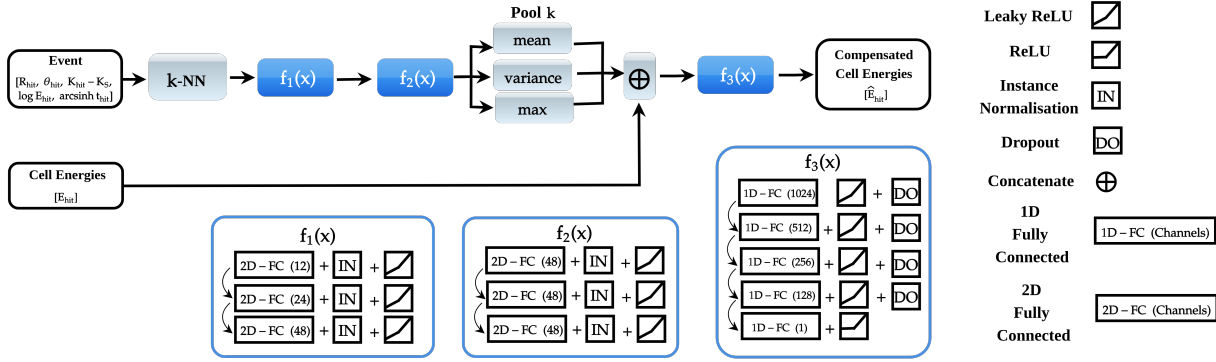
The methods for studying software compensation for AHCAL are presented in this section. First, an implementation of a neural network model based on the proposal discussed in Section 5.1.2 and additional considerations for its application are discussed. Then, a standard control method is introduced, and its implementation is described.

### 5.2.1 Neural Network SC Method

The neural network architecture was designed based on a single Edge-Conv operator of the Dynamic Graph Convolutional Neural Network (DGCNN) as described in [163]. The transformed coordinate system described in Section 2.5.1.2 was used as the input.

- ▶ *Input*: The neural network is provided with two inputs. The first is the hadron shower event in natural/transformed coordinates  $(R_{\text{hit}}, \theta_{\text{hit}}, K_{\text{hit}} - K_S, \log E_{\text{hit}}, \text{arcsinh } t_{\text{hit}})$ , where  $\text{arcsinh } t_{\text{hit}}$  is optional. The second is the original cell energy,  $E_{\text{hit}}$ , which is used to inform the neural network of the output scale of the compensated energy. Definitions for these variables are given in Section 2.5.1.2;
- ▶ *k-NN clustering*: The neural network clusters the input according to the  $k$ -nearest neighbours to each active cell in the event. Their positions and vectors relative to that cell are calculated, including a self-loop;
- ▶ *Up-scaling*: Each cluster is up-scaled through a module consisting of three sequential 2D fully-connected convolutional layers 12, 24, and 48 channels, each using leaky ReLU activation and instance normalisation;
- ▶ *Processing*: Each cluster is passed through a deep processing layer consisting of 3 sequential 2D fully-connected layers of 48 channels, each using leaky ReLU activation and instance normalisation;
- ▶ *Aggregation*: The maximum, mean and variance of the cluster dimension  $k$  is used as activation values for the cluster. These are concatenated with the cell energies of the event for each active cell;
- ▶ *Down-scaling*: The final layers of the network are five dense layers, with 1024, 512, 256 and 128 channels and leaky ReLU activation, with an output layer with ReLU activation such that the final output is positive. All dense layers, excluding the final layer, include dropout with probability  $p_{\text{dropout}}$ . The neural network's final output is a single value for each active cell, which is the compensated hit energy,  $\hat{E}_{\text{hit}}$ . The sum of these outputs yields the total compensated response,  $\hat{E}_{\text{sum}} = \sum^{\text{event}} \hat{E}_{\text{hit}}$ , where  $\hat{E}_{\text{hit}} \in [0, \infty]$ .

A diagram representing the proposed neural network architecture is shown in Figure 5.5.



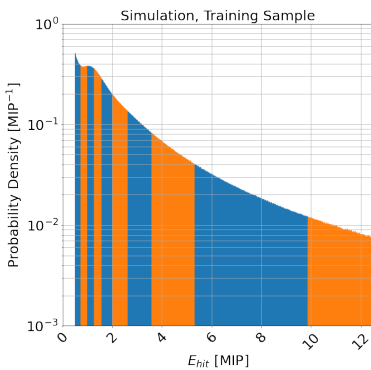
**Figure 5.5:** Flowchart describing the proposed neural network for software compensation studied in this chapter. The black, blue and grey boxes indicate inputs and outputs, convolutional operations and general operations, respectively. Additional operations are specified on the right of the figure.

## 5.2.2 Control SC Method

The neural network is compared to the standard CALICE software compensation method called '*local software compensation*', abbreviated hereafter as the control method and based on the method of [164], is described as follows.

The  $E_{\text{hit}}$  distribution is binned in deciles (i.e. a 10% probability for a given  $E_{\text{hit}}$  to be found in any one of the bins). For each bin, an appropriate function is used for weighting. A function approximator in the form of a second-order Chebyshev polynomial of the first kind,  $\omega_b$ , is defined as a function of the total calorimeter response,  $E_{\text{sum}}$ , scaled using a factor,  $S$ , such that  $E_{\text{sum}}/S \in [0, 1]$  for the typical range of hadron shower energies of AHCAL ( $S = 150$  GeV).  $\omega_b$  has three free parameters,  $\alpha_b$ ,  $\beta_b$  and  $\gamma_b$ , shown in Eq. 5.1:

$$\omega_b(E_{\text{sum}}; S, \alpha_b, \beta_b, \gamma_b) = \alpha_b + \beta_b \cdot \left( \frac{E_{\text{sum}}}{S} \right) + \gamma_b \cdot \left( 2 \left( \frac{E_{\text{sum}}}{S} \right)^2 - 1 \right) \quad (5.1)$$



**Figure 5.6:** Distribution of  $E_{\text{hit}}$  from the training dataset distribution shown in Table 5.1. Each alternating blue and orange coloured area indicates a decile bin.

For each bin, the corresponding weight is calculated. Finally, the energy of each active cell within the ranges defined by bin  $b$  is scaled by  $\omega_b$ .

$$\hat{E}_{\text{sum}} = \sum_b^{\text{bins}} \omega_b \cdot E_{\text{sum},b} \quad (5.2)$$

The idea underlying this method is that higher hit energy bins attenuate the energy, as these are more likely to belong to an EM fraction and enhance the energy of low energy bins, which are more likely to belong to the HAD fraction. A figure showing the decile binning used in this study is shown in Figure 5.6.

## 5.2.3 Datasets and Training

### 5.2.3.1 Datasets

The simulation and data used for the study presented in this chapter are introduced, analysed and studied.

Both the neural network model defined in Section 5.2.1 and the control model defined in Section 5.2.2 were trained and validated using simulated  $\pi^-$  hadron shower events observed with the AHCAL detector. The simulation of the particle showers was achieved using Geant4 [83], with a full detector simulation developed using DD4hep [84]. Additional effects, such as digitisation of the analogue signal and reconstruction of the detector variables, were achieved using CALICESoft [86]. The physics list used was QGSB\_BERT\_HP. The simulation was based on the June CALICE testbeam study taken at the Super Proton Synchrotron at CERN in 2018 [60]. Data used in the study was taken from the same experiment, where reconstruction of the detector variables were also achieved using CALICESoft.

**Selection Criteria** The sample was subject to the following selection criteria. Data and simulation were subject to the same cuts except where specified:

The following selection criteria were applied:

- ▶ events were required to be identified using the standard CALICE particle identification algorithm [79] as being a single particle and having less than a 0.5% probability of being a muon to exclude non-showering, 'punch-through' pions;
- ▶ the 38<sup>th</sup> layer of the AHCAL is ganged and requires special treatment beyond the scope of this paper. Therefore, energy deposits were considered up to the 38<sup>th</sup> layer of the calorimeter;
- ▶ events with a correctly reconstructed track position (i.e. a track position with a corresponding position inside the  $24 \times 24$  cell AHCAL front-face) and a shower starting layer within layers 1-4 of the AHCAL calorimeter were selected. This choice was made to reduce the effect of longitudinal and lateral leakage on the experiment. These cuts are supplemented, for the measurement of detector resolution and linearity only, by an additional cut using the TCMT detector. This criterion requires the TCMT to measure a total deposited energy of less than 25 MIP ( $E_{\text{sum}}^{\text{TCMT}} < 25 \text{ MIP}/670 \text{ MeV}$ ).

The training and validation dataset of simulation consisted of simulated showers induced by  $\pi^-$  hadrons with  $E_{\text{particle}}$  in the range 10-80 GeV, in increasing steps of 10 GeV. By contrast, the test sample contained showers induced by  $\pi^-$  hadrons with  $E_{\text{particle}}$  in the range 10-120 GeV, in increasing steps of 5 GeV. The finer granularity tests the hypothesis that the neural network is unbiased to the particular particle energies used for training. Energies higher than the training range are included to test the generalisation capacity of each compensation method. By contrast, the training, testing and validation sample of data was performed across all available samples, to assess the performance of the methods when

Type Sample	June 2018 SPS Testbeam Data				Simulation			
	Test	Test + TCMT Cut	Training	Validation	Test	Test + TCMT Cut	Training	Validation
$E_{\text{particle}}$ [GeV]								
10	6472	6460	51773	6472	20826	20759	18719	2080
15	-	-	-	-	21969	21685	-	-
20	9439	9233	75512	9439	23425	22808	21428	2381
25	-	-	-	-	25193	24124	-	-
30	-	-	-	-	24031	22491	21901	2434
35	-	-	-	-	24154	22065	-	-
40	10384	9378	83064	10383	24195	21513	23552	2617
45	-	-	-	-	23122	19981	-	-
50	-	-	-	-	27337	22889	24737	2749
55	-	-	-	-	19636	16009	-	-
60	13223	10684	105782	13223	22503	17728	24479	2720
65	-	-	-	-	25584	19374	-	-
70	-	-	-	-	18951	13889	24864	2763
75	-	-	-	-	15827	11204	-	-
80	11666	8298	93325	11666	22272	15165	25308	2813
85	-	-	-	-	22577	14875	-	-
90	-	-	-	-	26210	16618	-	-
95	-	-	-	-	20605	12475	-	-
100	-	-	-	-	17706	10385	-	-
105	-	-	-	-	17410	9873	-	-
110	-	-	-	-	16885	9161	-	-
115	-	-	-	-	18706	9820	-	-
120	10713	5829	85701	10713	18192	9239	-	-
Total Events	61897	49882	495157	61896	497316	395131	184988	20557

**Table 5.1:** Table of events used for training SC models after all cuts except the TCMT cut (shown separately), split into simulation and data and by the testing, training and validation samples and by data and simulation. Hyphens indicate 0 events.

applied to a realistic training dataset. The event count of the final training, testing and validation datasets are summarised in Table 5.1.

Comparisons of the main event-level, cell-level and shower profile distributions of the training sample for the entire samples of 10 GeV, 40 GeV and 80 GeV common between the simulation and data training sets are presented in Figures 5.7-5.21. Event level distributions are shown in Figures 5.7-5.11. Cell-level distributions are shown in Figures 5.12-5.18. Shower profile distributions are shown in Figures 5.19-5.21.

Additionally, Equation 2.8 cannot be applied to a calorimeter under the influence of leakage, as the model assumes full shower containment. For this reason, when evaluating the resolution of the calorimeter before and after compensation has been applied, an additional cut using the *tail catcher/muon tracker detector* (TCMT) that was included in the June 2018 testbeam and discussed in Section 2.5.1.1 and may be used as an independent detector to tag leakage events. Analysis of the proposed TCMT cut is presented in Figure 5.22 and Figure 5.23.

#### Summary Comment to Figures 5.7-5.21

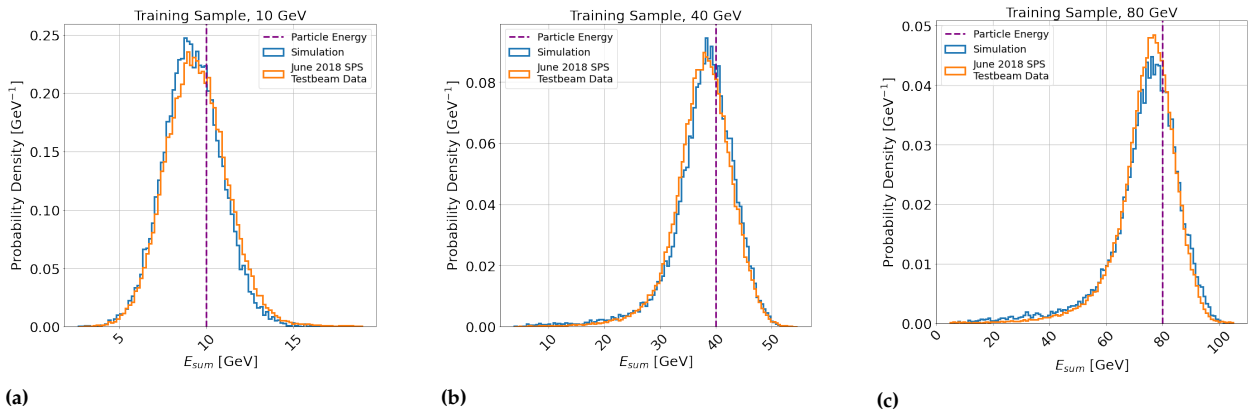
Simulation and data have similar distributions in most event-level and cell-level distributions. However, several significant discrepancies relevant to energy reconstruction are observed. The simulation produces more active cells than data and has a higher energy density core than in data. This motivates the study of the effectiveness of training on simulation and data independently and studies of the differences between the models.



**Summary Comment to Figure 5.22 and Figure 5.23**

The proposed TCMT cut is demonstrated to cause a significant reduction in the proportion of leaked events, evidenced by the normally-distributed response at each studied energy. The energy distributions after the cut are also in good agreement between the simulation and data after the cut has been applied, which indicates that the cut can be used in both.

## Event Level: Distributions of Reconstructed Shower Energy

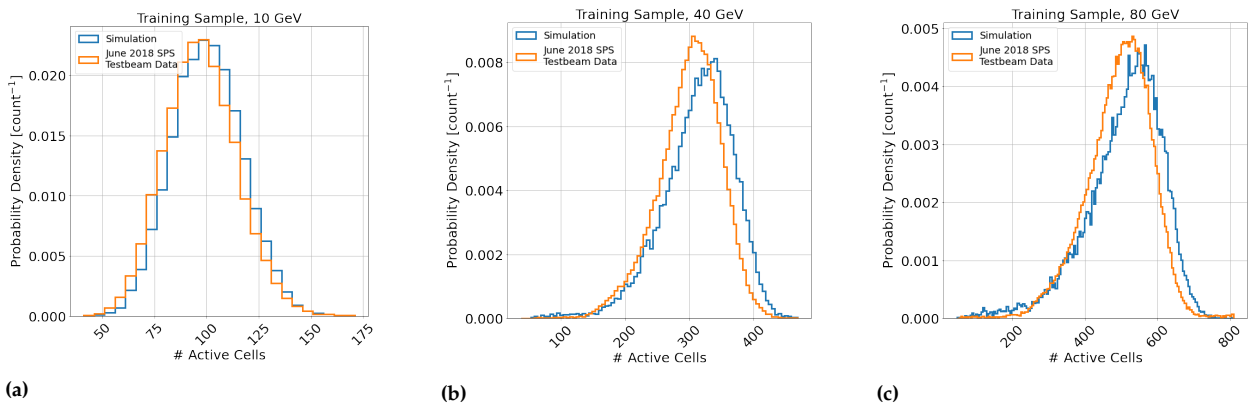


**Figure 5.7:** Distributions of the reconstructed shower energy ( $E_{\text{sum}}$ ) for the training dataset for software compensation, for 10, 40 and 80 GeV samples, shown in the left, middle and right columns, respectively. The blue histograms indicate simulation, and the orange histograms indicate June 2018 SPS Testbeam data.

## Comments:

- ▶ Figures 5.7a-5.7c indicate qualitatively good agreement between the reconstructed calorimeter response,  $E_{\text{sum}}$ , in simulation and data. The absence of a peak at low shower energy demonstrates the removal of punch-through pions by the applied cuts.
- ▶ Although a stringent shower start criterion is applied, a left-skewed power-law leakage tail is visible in Figure 5.7b and 5.7c. This effect demonstrates the necessity of an additional cut during validation to reduce the effect of leakage.

## Event Level: Distributions of Number of Active Cells

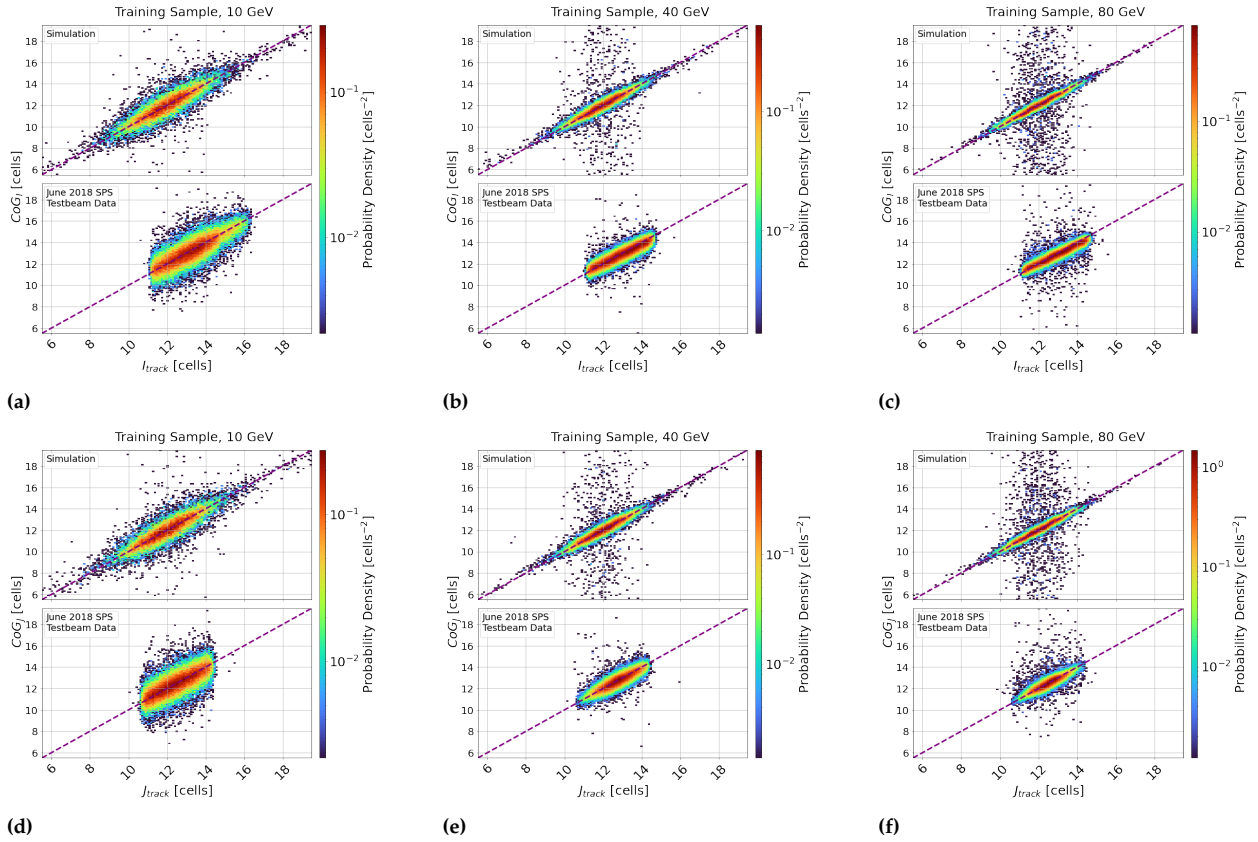


**Figure 5.8:** Distributions of the number of active cells per hadron showers for software compensation. Else, as in Figure 5.7.

## Comments:

- ▶ Figures 5.8a-5.8c indicate that more cells are active during  $\pi^-$  hadron events in simulation than in data on average.
- ▶ Because of the good agreement of Figures 5.7a-5.7c, this means the average energy density of the hadron showers must be greater on average in simulation than in data (i.e. more energy deposited per active cell, on average).
- ▶ MIP-tracks are expected to have no more than about 38 active cells in the event. The absence of a peak in Figures 5.8a-5.8c indicates the removal of punch-through pions by cuts.
- ▶ As in Figure 5.7, leakage is observed in the left-skewed tail of the distribution.

### Event Level: Joint Distributions of Lateral Centres-of-Gravity And Track Position

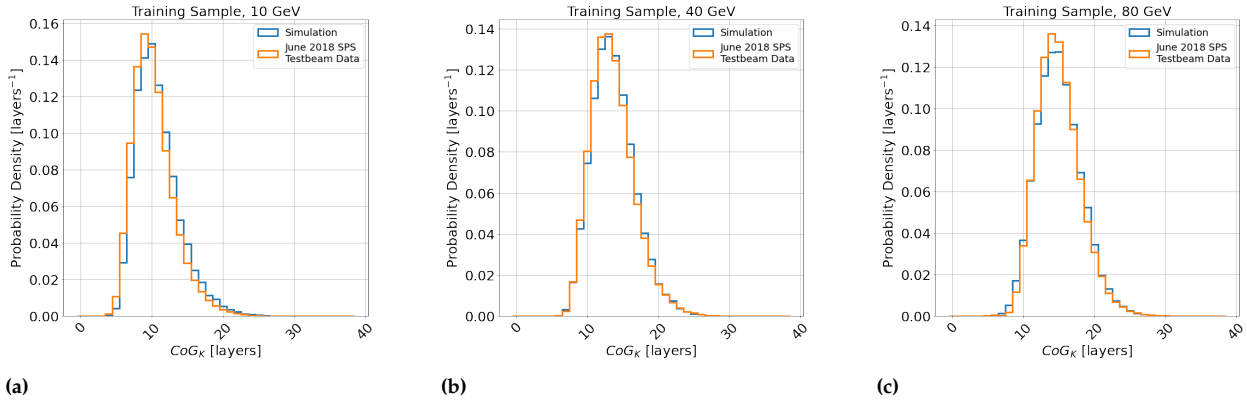


**Figure 5.9:** Joint distributions of the lateral centre-of-gravity and track position for the training dataset for software compensation, for 10, 40 and 80 GeV samples, shown in the left, middle and right columns, respectively. Figures 5.9a-5.9c show the distribution of  $\text{CoG}_I$ ,  $I_{\text{track}}$ , and Figures 5.9d-5.9f show the distribution of  $\text{CoG}_J$ ,  $J_{\text{track}}$ . The upper and lower subplots indicate the distribution of simulation and data, respectively. The colour axis indicates probability density. Else, as in Figure 5.7.

#### Comments:

- ▶ As a cross-check, the track position is expected to be highly correlated with the centre-of-gravity of the hadron shower. If the track information is correctly reconstructed, then the distribution should be similar to a 1:1 correspondence, illustrated by the purple line in the figures.
- ▶ Figures 5.9a-5.9c and Figures 5.9d-5.9f indicate that track positions and lateral centres-of-gravity of the hadron showers in the dataset are close to the central co-ordinates of the calorimeter.
- ▶ The agreement of each distribution with the purple line indicates the track is highly correlated with the centre of gravity, with a 1:1 correspondence on average.
- ▶ The correlation between the lateral centres-of-gravity and the track position are observed to increase as the particle energy increases. This may be explained because of a larger EM-fraction leads to a more energy-dense core, on average, as the particle energy increases. If the core carries a greater fraction of the total shower energy on average and varies less in space, it is expected that the centre of gravity also varies less compared to the known particle position.
- ▶ It is noted by comparing the upper and lower subfigures that there are differences between the distribution of tracks and centres of gravity between simulation and data. The reason for this is that the delay wire chamber constrains the position of tracks in data, while in simulation, there is no constraint on the track's position.

**Event Level: Distributions of Longitudinal Centres-of-Gravity**

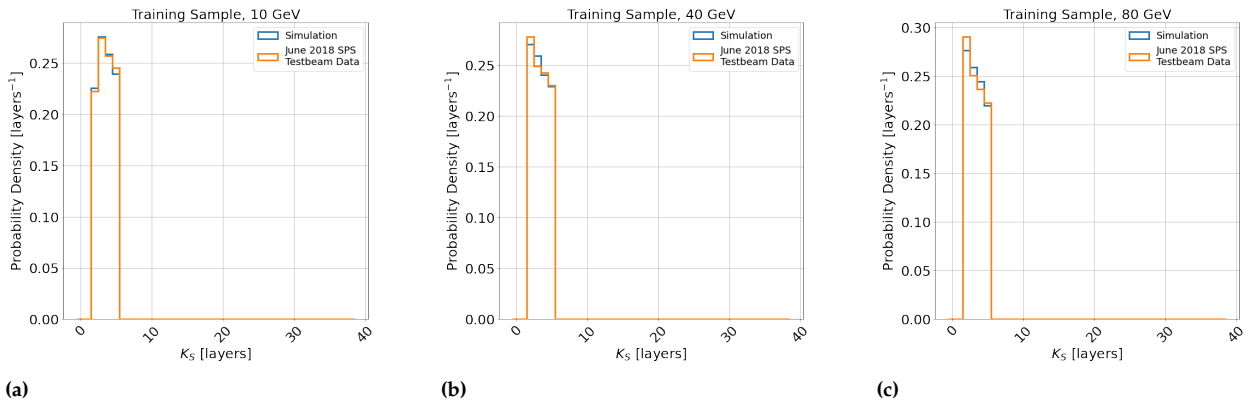


**Figure 5.10:** Distributions of the shower starting layer ( $K_S$ ) for the training dataset for software compensation. As in Figure 5.7.

**Comments:**

- Figures 5.10a - 5.10c indicate qualitatively good agreement between the longitudinal centre-of-gravity between simulation and data.

**Event Level: Distributions of Shower Start Layer**

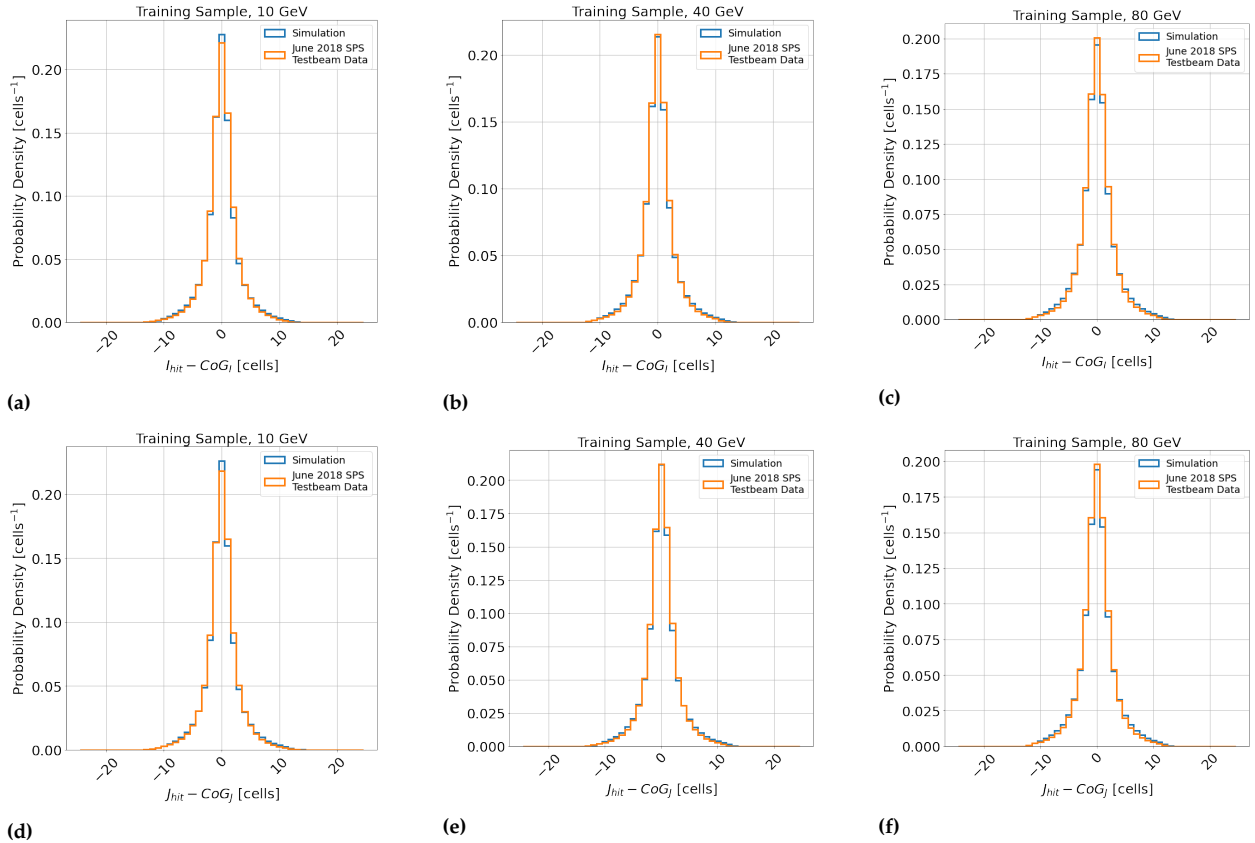


**Figure 5.11:** Distributions of the shower starting layer ( $K_S$ ) for the training dataset for software compensation. As in Figure 5.7.

**Comments:**

- Figures 5.11a-5.11c indicate that the selection criteria have functioned as intended, and the shower start occurs within the first four layers of the calorimeter.
- Qualitatively good agreement is observed between the distributions of simulation and data.

### Cell Level: Distributions of Lateral Active Cell Positions Relative to Centre-Of-Gravity

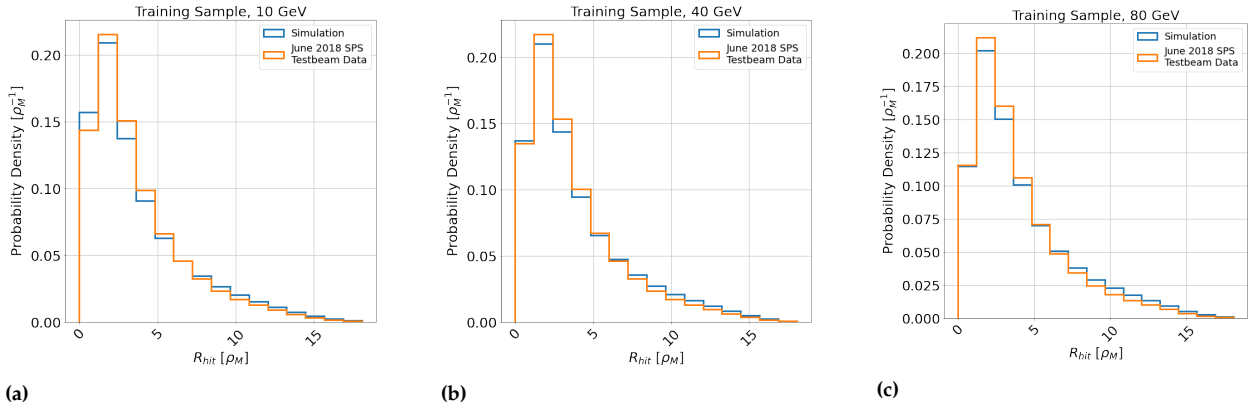


**Figure 5.12:** Distributions of the lateral cell indices for the training dataset for software compensation. Figures 5.12a-5.12c and Figures 5.12d-5.12f show the distributions of  $I_{hit}$  and  $J_{hit}$  minus their corresponding centres-of-gravity,  $CoG_I$  and  $CoG_J$ , respectively, presented in units of cells. Figures 5.13a-5.13c show the hit radius distributions as defined in 2.5.1.1. Else, as in Figure 5.7.

#### Comments:

- Figures 5.12a-5.12c, Figures 5.12d-5.12f indicate that simulation produces slightly more active cells in the halo of the hadron shower than in data (see Figure 5.13).
- Notwithstanding the previous remark, qualitatively good agreement is observed between the distributions of simulation and data.

### Cell Level: Distributions of Hit Radius

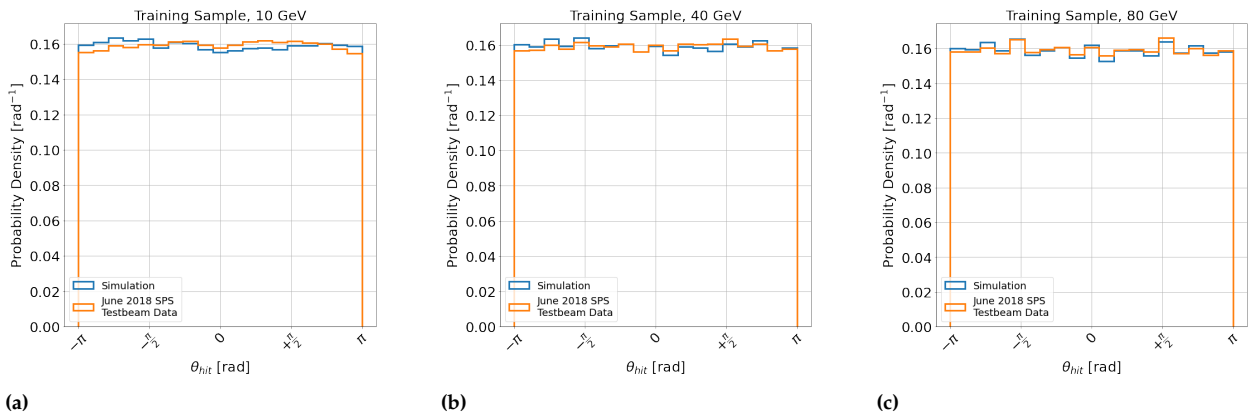


**Figure 5.13:** Distributions of hit radius ( $R_{hit}$ ) as defined in Section 2.5.1.1 for the training dataset for software compensation. Else, as in Figure 5.7.

**Comments:**

- ▶ Figures 5.13a-5.13c indicate that simulation produces slightly more active cells in the halo of the hadron shower ( $R_{hit} > 1 \rho_M$ ) than in data.
- ▶ Notwithstanding the previous remark, qualitatively, good agreement is observed between the distributions of simulation and data.

### Cell Level: Distributions of Hit Azimuthal Angle

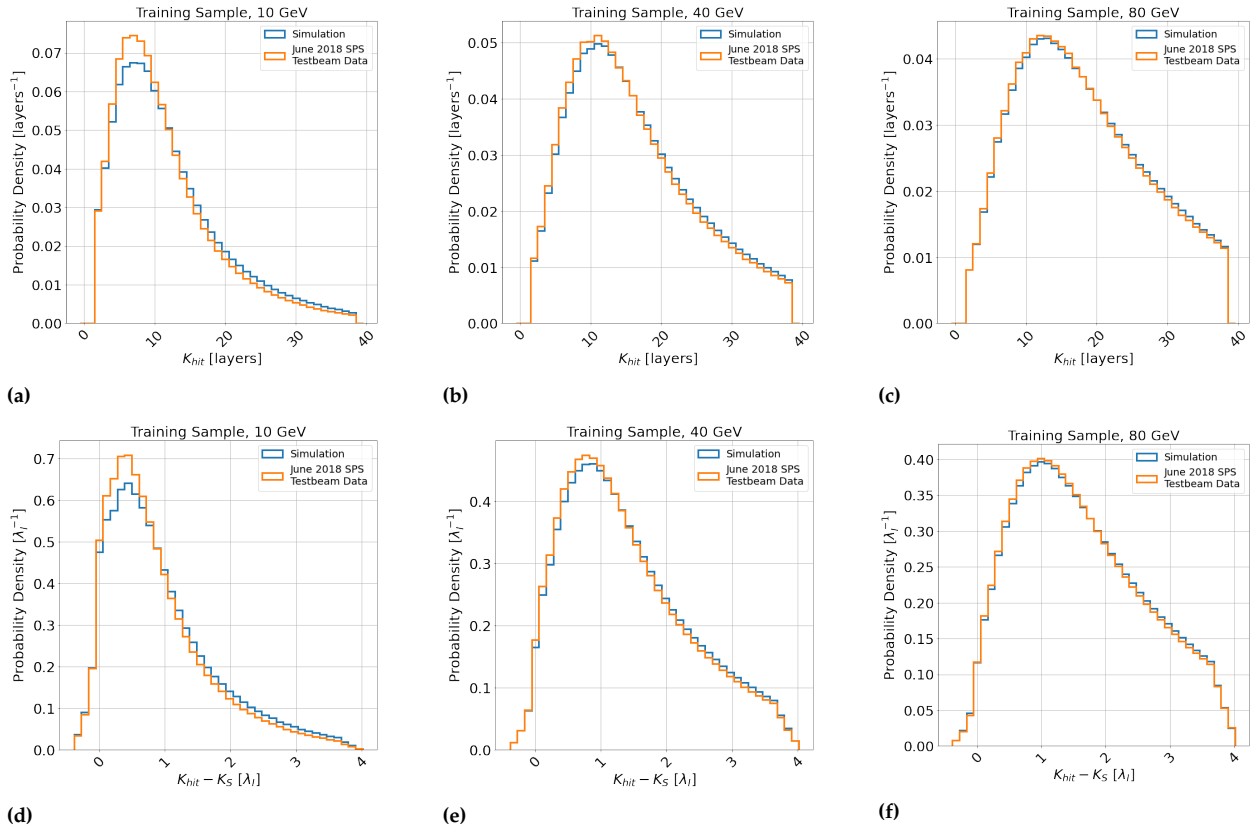


**Figure 5.14:** Distributions of hit azimuthal angle ( $\theta_{hit}$ ) as defined in Section 2.5.1.1 for the training dataset for software compensation. Else, as in Figure 5.7.

**Comments:**

- ▶ Figures 5.14a-5.14c indicate that simulation and data result in isotropically distributed azimuth angles with relative to the centre-of-gravity. This is expected from multiple scattering expected of hadron showers.
- ▶ Qualitatively good agreement is observed between the distributions of simulation and data.

### Cell Level: Distributions of Reconstructed and Shower-Start Normalised Longitudinal Layer Position

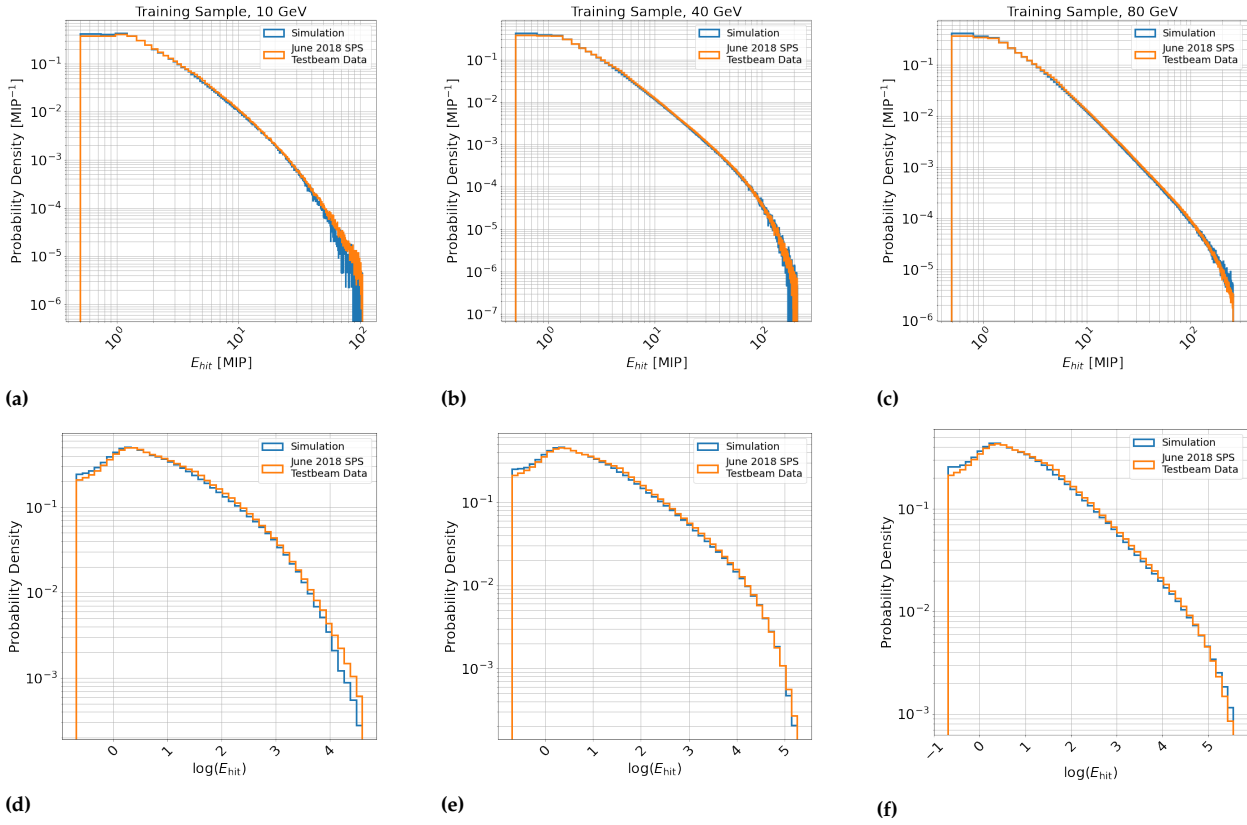


**Figure 5.15:** Distributions of longitudinal cell-level indices for the training dataset for software compensation, respectively. Figures 5.15a-5.15c and Figures 5.15d-5.15f show the distributions of  $K_{\text{hit}}$  and  $K_{\text{hit}} - K_S$ , respectively. Else, as in Figure 5.7.

#### Comments:

- Figures 5.15a-5.15c and Figure 5.15d-5.15f indicate qualitatively good agreement between the distributions of active cells in the longitudinal spatial dimensions of the calorimeter for the 40 GeV and 80 GeV samples.
- For the 10 GeV sample, a more significant proportion of the hits in the sample are observed later in the longitudinal development of the hadron shower. This observation indicates that more cells are active in data near the longitudinal centre-of-gravity for 10 GeV  $\pi^-$  hadrons than for simulation.
- For the 40 GeV and 80 GeV samples, the distribution is truncated at the end of the calorimeter, which is not the case for the 10 GeV sample. This observation is due to the effect of longitudinal leakage in the AHCAL calorimeter.

**Cell Level: Distributions of Active Cell Energy and its Logarithm**



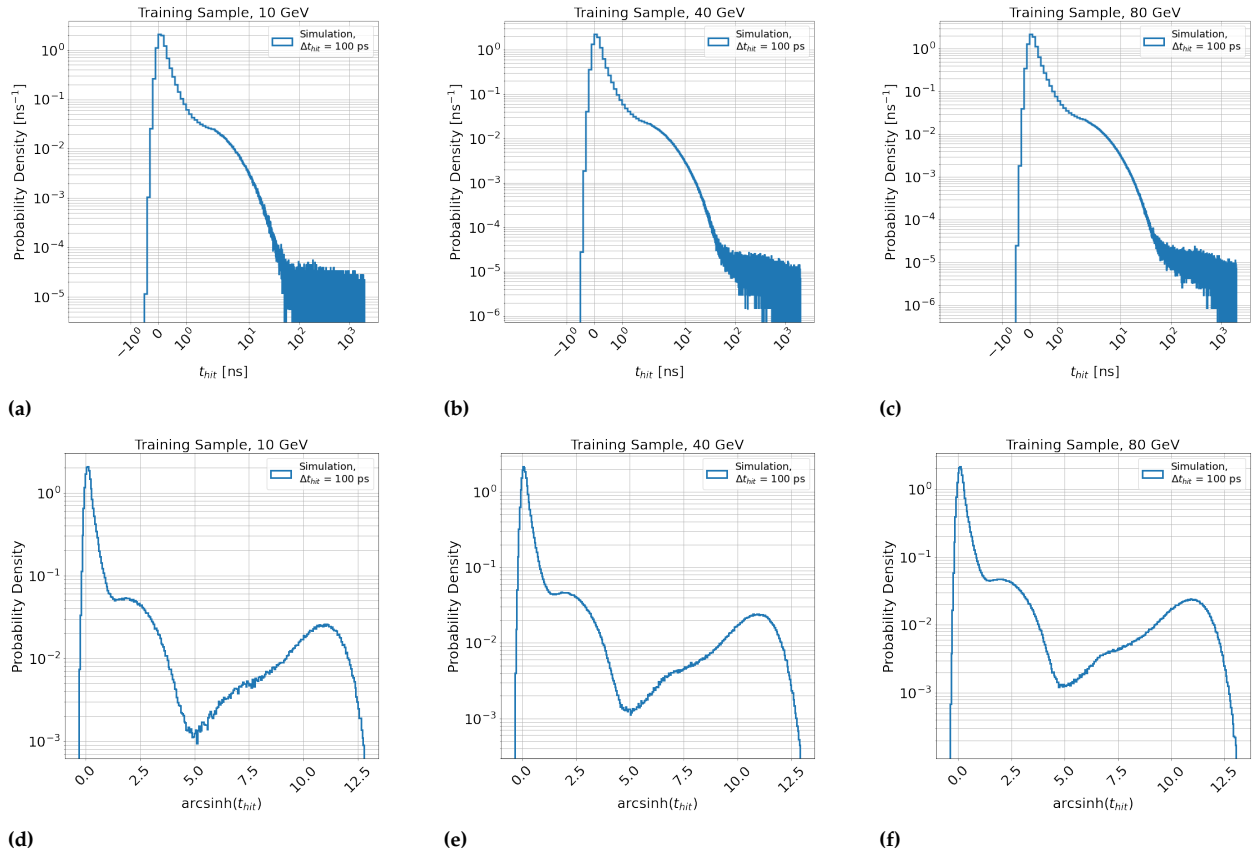
**Figure 5.16:** Distributions of active cell energy for the training dataset for software compensation, respectively. Figures 5.16a-5.16c and Figures 5.16d-5.16f show the distributions of  $E_{hit}$  and its logarithm, respectively. Else, as in Figure 5.7.

**Comments:**

- ▶ Figures 5.16a- 5.16c and Figures 5.16d-5.16f indicate that, while simulation and data generally agree, the simulation over-predicts the number of active cells with  $E_{hit} < 1$  MIP. Figure 5.21 presents a more substantive analysis of this discrepancy.
- ▶ Figures 5.16d-5.16f demonstrate the effect of applying the logarithm to the hit energy distribution. The distribution is demonstrated to reduce the variable’s variance, indicated by the greatly reduced scale on the  $x$ -axis.



### Cell Level: Distributions of Active Cell Timestamp and its Hyperbolic Sine in Simulation with a Resolution of 100 ps

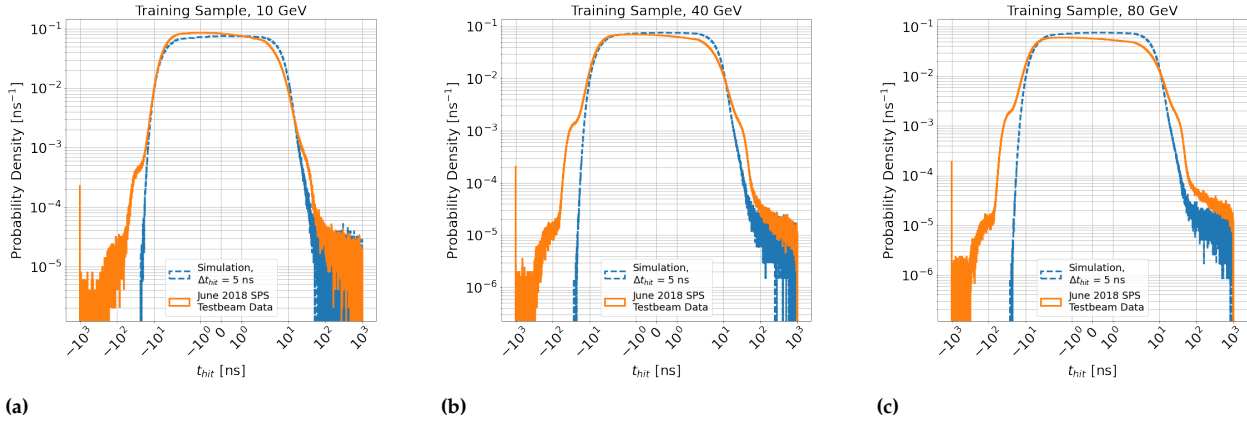


**Figure 5.17:** Distributions of the cell timestamp for the training dataset for software compensation in simulation, with 100 ps time resolution. Figures 5.17a-5.17c and Figures 5.17e-5.17f show the distributions of the active cell timestamp in simulation smeared by of 100 ns and its hyperbolic sine, respectively.

#### Comments:

- Figures 5.17c-5.17c indicate the bi-exponential structure of time (smeared by Gaussian noise) expected from the fast and slow component of the time distribution illustrated in Figure 2.17 of Section 2.3.6.
- Figures 5.17d-5.17f indicate the effect of applying a hyperbolic sine to the hit time distribution. The distribution is demonstrated to reduce the variable's variance, indicated by the greatly reduced scale on the  $x$ -axis. Additionally, the rising probability density observed in the range  $5 < \text{arcsinh}(t_{hit}) < 12.5$  ( $74 \text{ ns} < t_{hit} < 0.1342 \text{ ms}$ ) is due to an exponentially distributed variable being modified by a logarithm-like function and is not a physical effect.

**Cell Level: Distributions of Active Cell Timestamp with a Resolution of 5 ps**

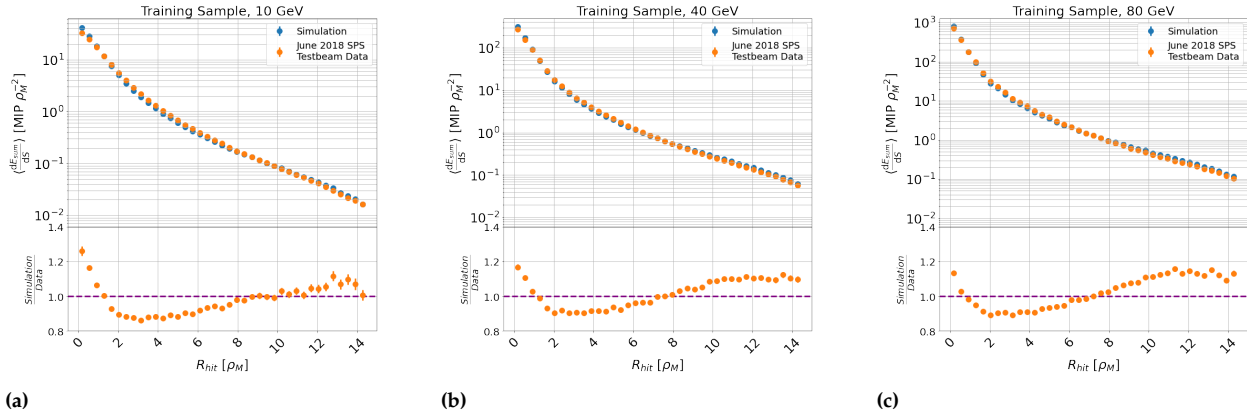


**Figure 5.18:** Distributions of the cell timestamp for the training dataset for software compensation in simulation, with 5 ns time resolution. Figures 5.18a-5.18c show the distributions of the active cell timestamp in simulation smeared by of 5 ns.

**Comments:**

- Figures 5.18a-5.18c indicate that there are significant deviations in the time distribution in simulation and data, as demonstrated by the shoulders in the data (orange) in the range  $10 \text{ ns} < t_{\text{hit}} < 100 \text{ ns}$  and  $-100 \text{ ns} < t_{\text{hit}} < -10 \text{ ns}$  and at that do not appear in simulation (blue). These effects arise from the chip occupancy effects discussed in Section 2.5.1.1. Therefore, hit timestamp information is not used in training or evaluating data.

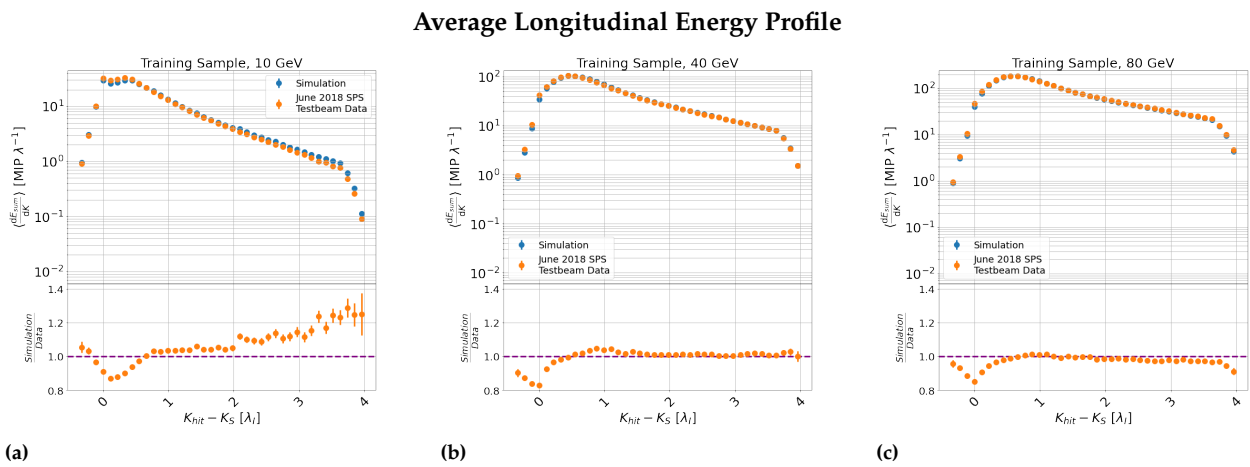
**Average Radial Energy Profile**



**Figure 5.19:** Figures 5.19a-5.19c show the average radial shower energy profile distributions per unit radial surface area of a circle with radius  $R$  (denoted  $S$ ) Blue and orange dots indicate simulation and data, respectively.

**Comments:**

- Figures 5.19a-5.19c describe mean energy deposited by a hadron shower per unit area of a circle (i.e. a thin ring) around the centre-of-gravity ( $S = 2\pi R_{\text{hit}} \cdot dR_{\text{hit}}$ , where  $dR_{\text{hit}}$  is a bin width). A circle is used because, as demonstrated by Figure 5.14, energy is deposited by hadron showers isotropically in azimuthal angle due to multiple scattering.
- It is observed that simulation produces excess energy deposited in the core and far from the hadron shower core, independently of hadron shower energy. These results are corroborated by the analysis of [34], which observed similar trends.

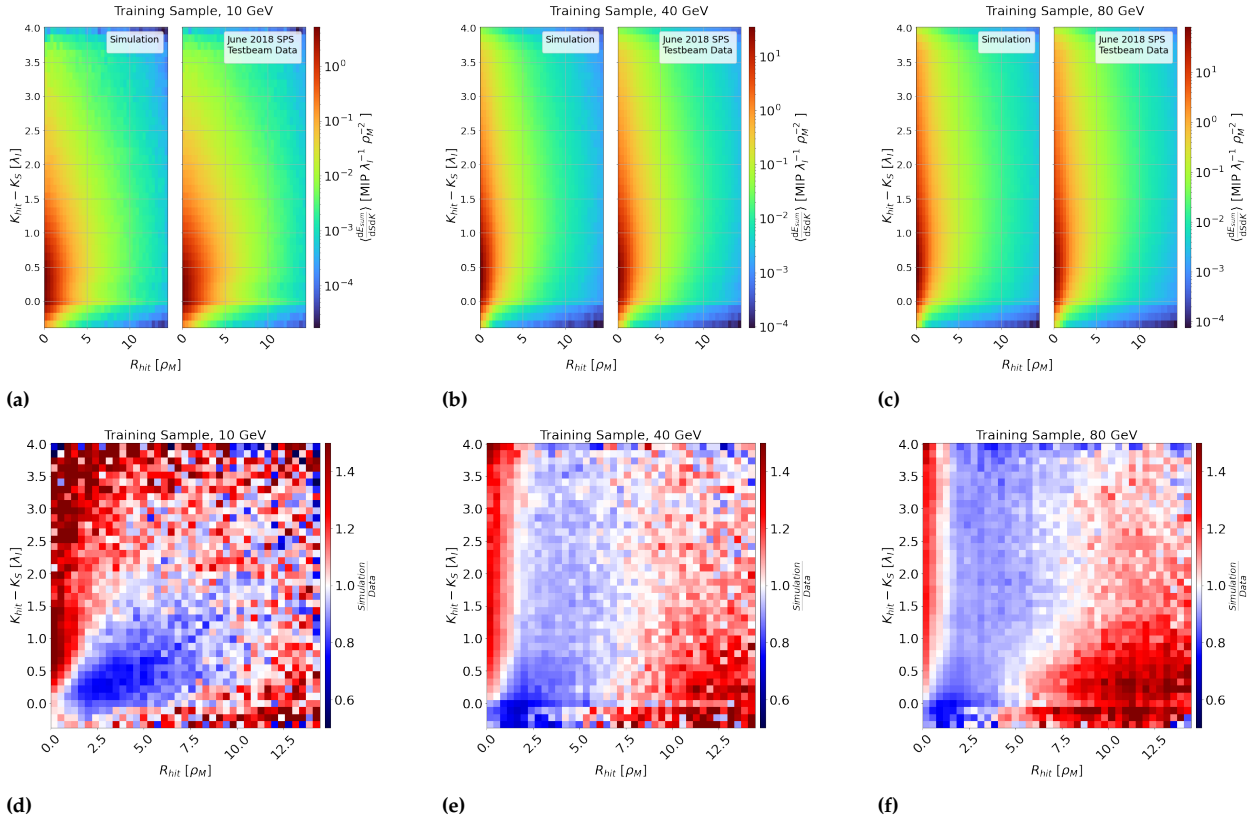


**Figure 5.20:** Figures 5.19a-5.19c show the average longitudinal shower energy profile distributions per layer of the calorimeter, relative to shower starting position ( $K_S$ ). Blue and orange dots indicate simulation and data, respectively.

**Comments:**

- ▶ Figures 5.19a-5.19c describe mean energy deposited by a hadron shower per layer of the calorimeter, relative to shower starting position ( $K_S$ ).
- ▶ It is observed that simulation produces excess energy deposited close to the shower start, independently of hadron shower energy, and is otherwise similar. These results are corroborated by the analysis of [34], which observed similar trends.

### Average Joint Radial-Longitudinal Energy Profile and Ratio of Simulation to Data

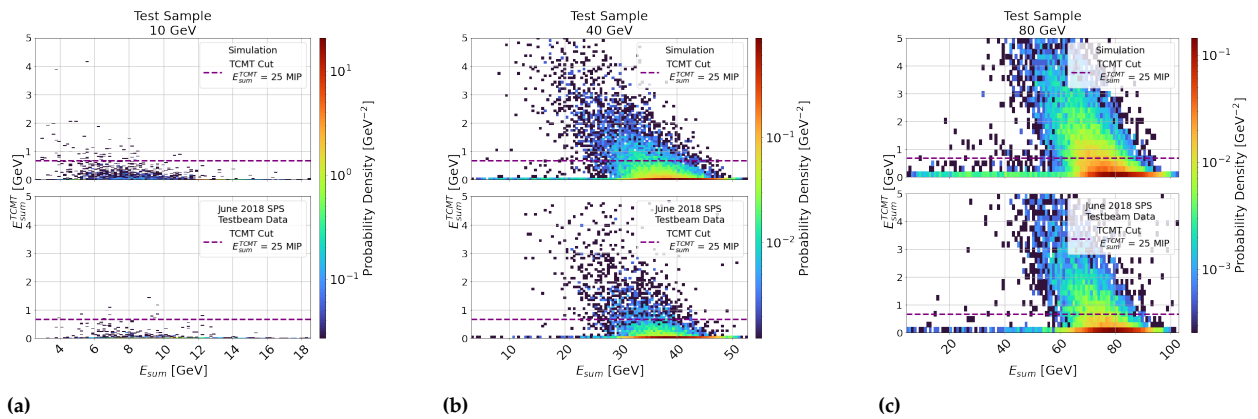


**Figure 5.21:** Figures 5.21a-5.21c show the joint average radial shower energy profile distributions per unit radial surface area of a circle with radius  $R$  (denoted  $S$ ), per layer of the calorimeter relative to the shower start ( $K_S$ ). The colour axis indicates probability density. Figures 5.21d-5.21f show the ratio of simulation and data shown in Figures 5.21a-5.21c, where the colour axis indicates the ratio.

#### Comments:

- ▶ Figures 5.21a-5.21c indicate that simulation has a slightly narrower energy profile than data. This observation agrees with the hypothesis that the core is more energy dense in simulation than in data.
- ▶ Figures 5.21d-5.21f indicate an excess of energy density in the hadron shower core ( $R_{hit} < 1 \rho_M$  and  $K - K_S > 0$ ) and in the region where  $K - K_S > 1 \lambda_I$  and  $R_{hit} > 7.5 \rho_M$ . It is in excess by around 20% for all studied samples in these regions.
- ▶ A minor dearth of energy density is observed in the region  $1 < R_{hit} < 1 \rho_M < 7.5$  and  $K - K_S > 1 \lambda_I$ , also by around 20%.

### Joint Distribution of the Reconstructed Shower Energy Measured by AHCAL and the TCMT

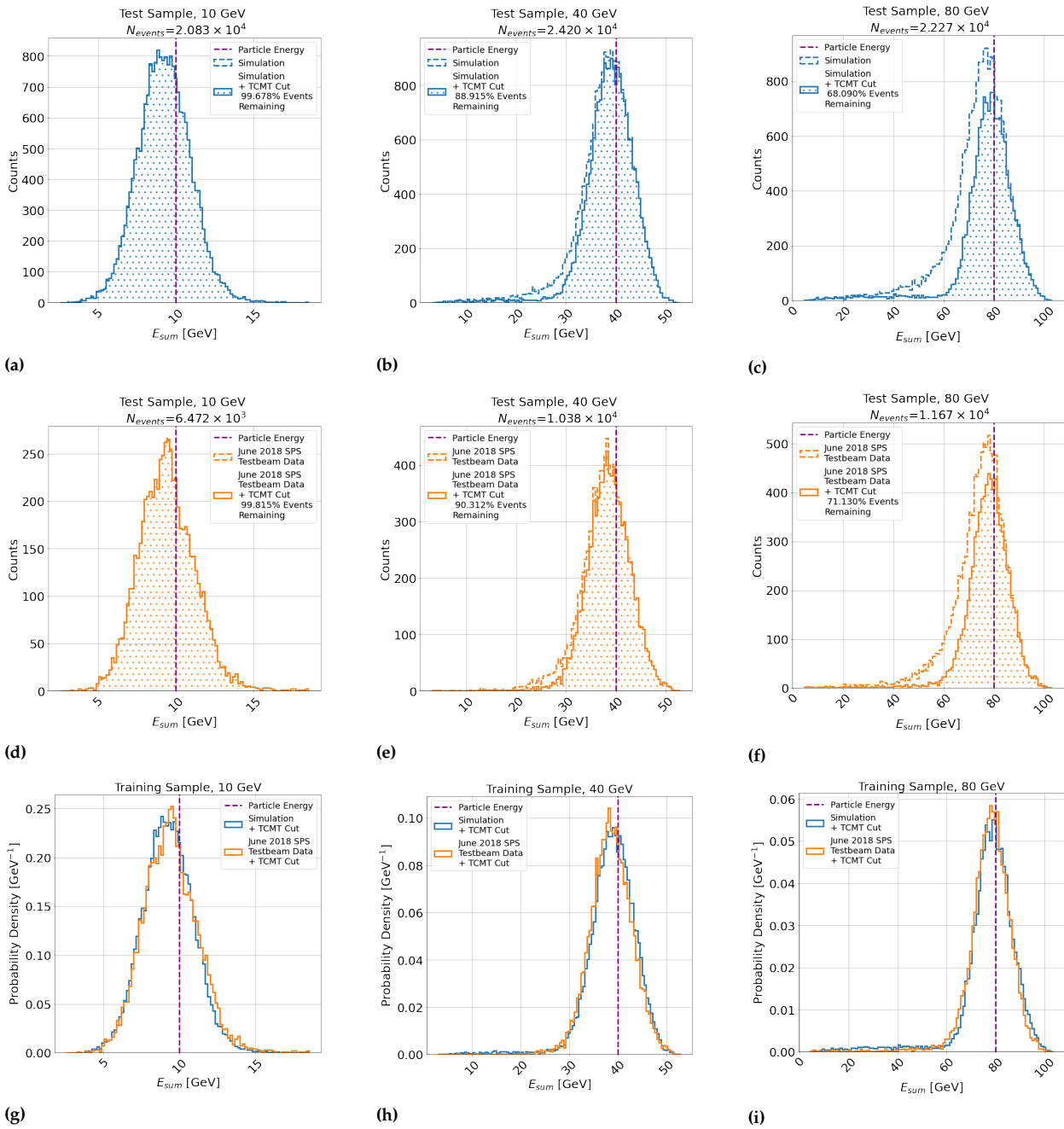


**Figure 5.22:** Joint distributions of the total reconstructed energy measured by the AHCAL ( $E_{\text{sum}}$ ) and the TCMT ( $E_{\text{sum}}^{\text{TCMT}}$ ) for the testing sample. The subplots indicate simulation and data, respectively. The purple dashed line indicates the applied cut. The colour axis indicates probability density.

#### Comments:

- ▶ Figure 5.22a indicates that 10 GeV hadron showers are, in essence, already well contained within the calorimeter geometry in both simulation and data. This observation is indicated by the low vertical spread of the distribution, which means that the TCMT typically measure a signal for a 10 GeV hadron shower.
- ▶ Figure 5.22b-5.22c indicates that 40 GeV and 80 GeV hadron showers are not well contained within the calorimeter. A weak correlation is observed between the TCMT and AHCAL reconstructed energy, indicating that the signal in the TCMT increases as the AHCAL signal decreases, which that the TCMT tags the calorimeter leakage.
- ▶ The cut, indicated by the purple dashed line, shows that it retains events which are comparatively well contained in the calorimeter. More stringent cuts were not found to significantly influence the  $E_{\text{sum}}$  distribution, yet greatly reduced the numbers of available statistics.

### Reconstructed Energy Distributions Before and After the TCMT Cut is Applied



**Figure 5.23:** Distributions of reconstructed energy before and after the TCMT cut is applied. Figures 5.23a-5.23a and Figures 5.23a-5.23d show the distributions of simulation, in blue and orange, respectively. The dashed line and the solid line filled with dots indicate the distribution before and after the cut has been applied. Figure 5.23g-5.23h show a comparison of the simulation and data. Else, as in Figure 5.7.

**Comments:**

- ▶ The 10 GeV samples shown in Figure 5.23a and Figure 5.23d indicate that the cut has little to no influence on the distribution of these samples since the showers are already well contained. By contrast, Figures 5.23b- 5.23c and Figures 5.23e- 5.23f indicate that the cut significantly reduces the leakage tail that appears at high energies.
- ▶ Figures 5.23g-5.23g indicate that the cut results in similar distributions in simulation and data, respectively. This means that the cut performs similarly in both samples.

### 5.2.3.2 Training

For simulation, two independent neural networks based on the model defined in Section 5.2.1 were trained on the training dataset: one without timing information and one with timing information. For data, a single neural network was trained without timing information. The proposed compensation networks were developed in PyTorch [113] and trained using the PyTorch Lightning research framework [165] on an NVidia V100 GPU. The ADAM optimiser was used to improve the convergence rate for ten epochs. The hyperparameters used for training are shown in Table 5.2, selected based on the results of a scan using Optuna hyperparameter optimisation framework [166], and shown in Table 5.2.

The control method was also trained using the training dataset for simulation and data, using the MIGRAD algorithm of the Minuit minimisation program [167]. Weights were initialised such that the compensation algorithm acted as the identity operator ( $\alpha_b = 1, \beta_b = 0, \gamma_b = 0$ ).

The loss was chosen to be the  $\chi^2$  goodness-of-fit of the compensated energy to the known particle energy of the hadron shower:

$$\mathcal{L}(\widehat{E}_{\text{sum}}; E_{\text{particle}}) = \frac{(\widehat{E}_{\text{sum}} - E_{\text{particle}})^2}{E_{\text{particle}} \cdot (1 \text{ GeV})} \quad (5.3)$$

The denominator in the loss arises from the uncertainty on the Poisson-distributed sampling quanta measured by the calorimeter,  $\sigma_E = \sqrt{E_{\text{particle}}}$ . The dummy constant of 1 GeV in the denominator is formally included to make the loss unitless and merely acts to scale the loss. The mean loss was used for both implementations to optimise the control and network methods. For the network methods, the epoch with the smallest mean loss of the validation sample was chosen for further study. The control method was minimised with the MIGRAD algorithm until the mean training loss reached convergence.

Parameter	Value
Learning Rate	$9 \times 10^{-5}$
Batch Size	32
$\beta_1$	0.9
$\beta_2$	0.999
$p_{\text{dropout}}$	0.15
$k$	20

**Table 5.2:** Table of hyperparameters used to train the neural network. In this table,  $\beta_1$  and  $\beta_2$  are the ADAM momentum parameters,  $p_{\text{dropout}}$  is the dropout probability, and  $k$  is the number of nearest-neighbours per cluster. The parameters were informed by a hyperparameter scan using Optuna [166].

## 5.3 Results

In this section, analysis is performed to assess the performance and properties of the models described in Section 5.2.1 and Section 5.2.2, trained on the datasets described in Section 5.2.3.1 and trained as described in Section 5.2.3.2.

### 5.3.1 Simulation Model

This section discusses the model trained and applied to the simulation test set in Table 5.1.

#### 5.3.1.1 Response Distributions in Simulation

A subsample of the normalised energy response distributions for each particle energy for hadron showers in the test sample with the TCMT cut applied is shown in Figure 5.24. A full set of figures for each available particle energy are available in Appendix Figure 8.4. The uncompensated sample and each sample after compensation are shown in each plot. The Freedman-Diaconis rule was applied to each sample to determine the bin width [147].

#### 5.3.1.2 Resolution and Linearity of Response in Simulation

The energy response distributions for each particle energy in the testing dataset, with the TCMT, cut applied, were estimated by fitting the calorimeter response distributions shown in Figure 8.4 with a normal distribution in the range of  $\pm 2$  standard deviations from their mean. The location and scale parameters of the fit,  $\mu$  and  $\sigma$ , were used to estimate  $E$  and  $\sigma_E$  of Equation 2.8 and used to study resolution ( $\sigma/\mu$  vs.  $E_{\text{particle}}$ ) and linearity of response ( $\mu/E_{\text{particle}}$  vs.  $E_{\text{particle}}$ ). The fitted values are shown in Appendix Table 8.3.

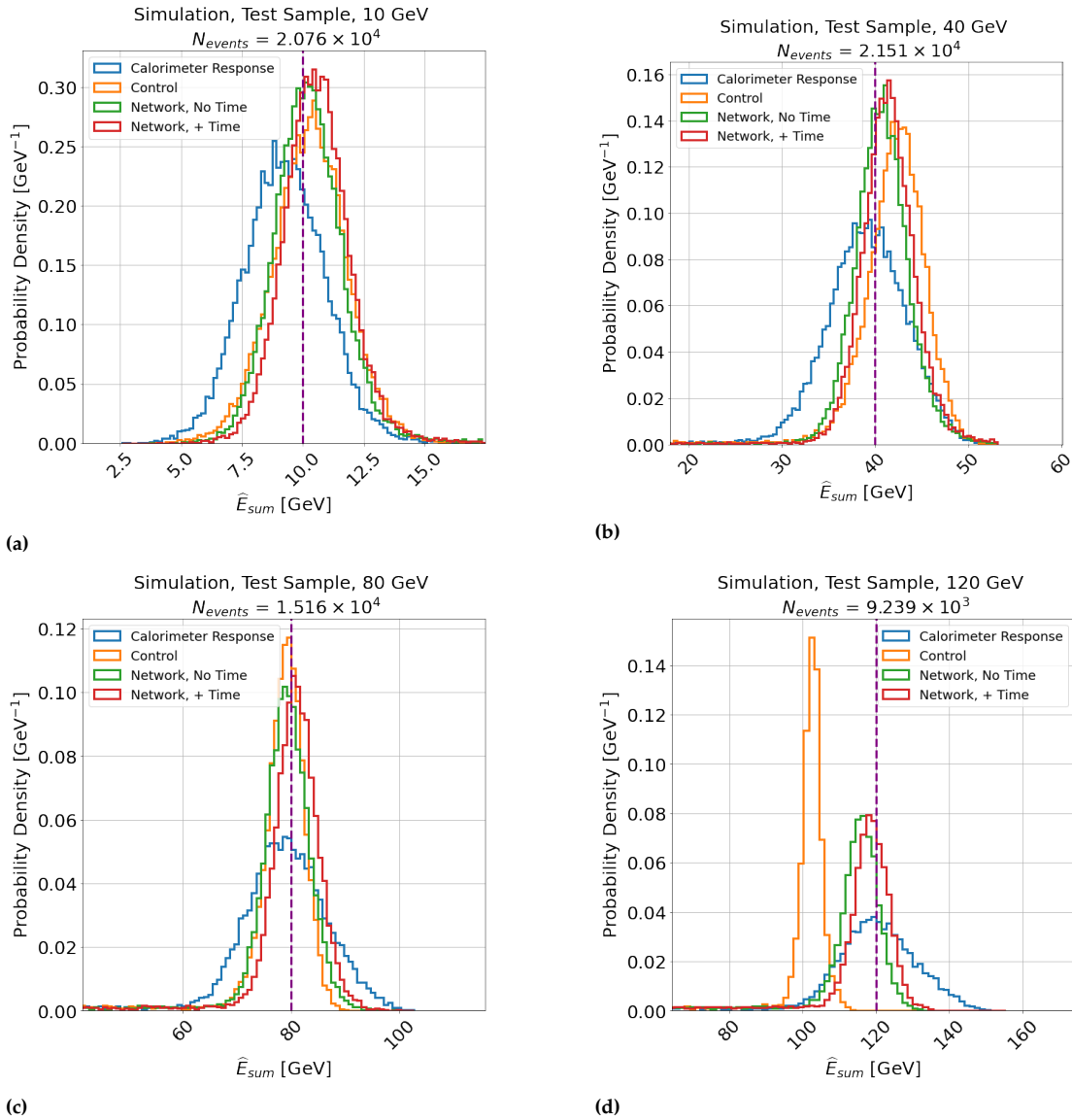
The location and scale parameters of the fit,  $\mu$  and  $\sigma$ , were used to estimate  $E$  and  $\sigma_E$  of Equation 2.8 and used to study resolution ( $\sigma/\mu$  vs.  $E_{\text{particle}}$ ) and linearity of response ( $\mu/E_{\text{particle}}$  vs.  $E_{\text{particle}}$ ). The results are shown in Figure 5.25. The fit values for Equation 2.8 are shown in Table 5.3.

#### Summary Comment to Figure 5.25 and Table 5.3

The neural network SC methods applied to simulation are demonstrated to perform compensation beyond the edge of the training range. In contrast, the control is observed to bias to the training range. The machine learning methods outperform the control method in the linearity of response and resolution, resulting in an improvement of the intrinsic stochastic resolution term of 9.3% and 12.2%, outperforming the control method in both cases. The neural networks also reduce the constant resolution term by 5%, indicating the neural networks perform some detector calibration and SC.



### Reconstructed Energy Distributions for SC Models Trained On and Applied To Simulation, with the TCMT Cut Applied

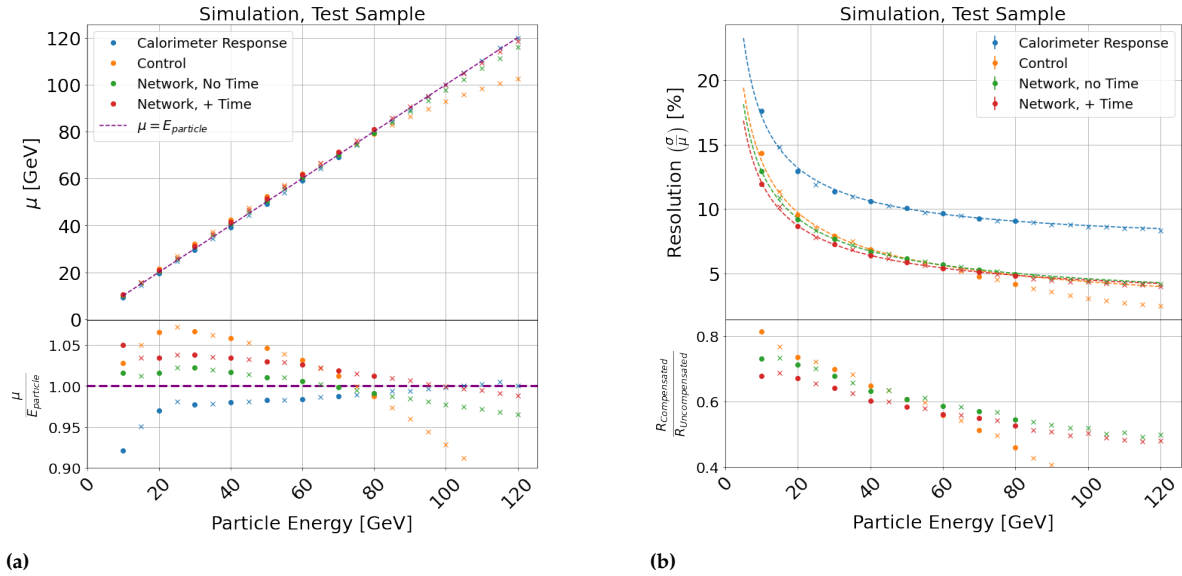


**Figure 5.24:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the test dataset of Table 5.1, with the TCMT cut applied. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods, respectively.  $E_{\text{particle}}$  is indicated as a dashed purple line. In this sample, the 10 GeV, 40 GeV, 80 GeV and 120 GeV samples are shown.

#### Comments:

- Figures 5.24a-5.24b shows that the neural network methods outperform the control method in the range of samples at particle energies of 10 GeV and 40 GeV, indicated by the lower spread of the response than for the control method.
- The inclusion of timing information results in superior energy resolution than excluding it. The likely reason for this is an enhancement of the HAD fraction is expected to play a more significant role in compensation at this energy scale due to a larger HAD fraction on average than at higher energies in the training sample. This occurs because the EM fraction increases on average with  $E_{\text{particle}}$  [28], indicated by the lower variance of the red distributions compared to the green.
- The linearity of samples between the training energies demonstrates that the neural network methods produce a more linear response than the control method and are, therefore, able to interpolate SC to samples between training energies.
- Figure 5.24c indicates that the control method outperforms the neural network methods in the range of samples for this sample. However, by examination of Figure 5.24d, it becomes apparent that this result is due to the control method biasing to the upper edge of the training sample. This statement is justified by the artificial attenuation of the response by the control method, resulting in a highly non-linear compensated response. In contrast, the neural network methods preserve the linearity of response beyond the training range. Therefore, it is demonstrated that the neural network model can extrapolate the compensation to higher particle energies without further training.

### Linearity and Resolution Fits for SC Models Trained On and Applied To Simulation, with the TCMT Cut Applied



**Figure 5.25:** Figure 5.25a and Figure 5.25b show AHCAL linearity of response and resolution to simulation using all methods under test, respectively. Blue indicates intrinsic calorimeter response, while orange, green and red indicate the control, network without and network with time compensation methods, respectively. Circle and cross markers indicate energies for both training and testing, and testing only. Figure 5.25a shows the fitted  $\mu$  to  $E_{\text{particle}}$ , where the purple dashed line indicates  $\mu = E_{\text{particle}}$ . The bottom subplot shows the ratio of  $\mu$  to  $E_{\text{particle}}$ . Figure 5.25b shows the fitted  $\sigma/\mu$ , where the dashed lines indicate fits of Equation 2.8. The bottom subplot indicates the ratio of the resolution of each compensation method to the intrinsic calorimeter response.

#### Comments:

- Figure 5.25a indicates the neural network methods offer improved linearity of response compared to the control, which overestimates the hadron shower energy by up to 5% compared to 2-3% for the network methods for most of the training range of particle energies. Moreover, the network and control methods are demonstrated to interpolate within the training range. However, for particle energies greater than 80 GeV, the control method fails to reconstruct the particle energy entirely. The divergence of the orange points from the dashed purple line demonstrates this observation.
- Figure 5.25b demonstrates that for  $E_{\text{particle}}$  up to around 60 GeV, the neural network methods produce superior compensation, indicated by the smaller value of the compensated response to the intrinsic response. However, beyond this range, the resolution produced by the control method diverges from the model of Equation 2.8. For this reason, the fit to this method was only performed for  $E_{\text{particle}}$  in the range 10-60 GeV. By contrast, the uncompensated and network methods showed good agreement with the expectations of Equation 2.8 and were fitted over the entire range.

	a [%]	b [%]	$\chi^2/NDF$
Calorimeter Response	$49.516 \pm 0.401$	$7.147 \pm 0.067$	4.575
Control	$43.387 \pm 0.119$	$0.010 \pm 2.873$	14.333
Network, no Time	$40.236 \pm 0.217$	$2.158 \pm 0.087$	0.857
Network, + Time	$37.275 \pm 0.208$	$2.448 \pm 0.070$	1.440

**Table 5.3:** Table of fitted parameters of Equation 2.8 to the training range of energies in the simulation shown as dashed lines in Figure 5.25b, except for the control method, which was fitted up to 60 GeV due to the effect of energy biasing.

#### Comments:

- The uncompensated stochastic resolution for simulated  $\pi^-$  hadron showers in AHCAL is in agreement within 1-2% with the  $a = 51.7 \pm 0.97\%$  obtained in the study of [34].
- The neural network solutions improve the calorimeter's stochastic resolution,  $a$ , by around 9.3% (around 3% better than control method) and 12.2% (around 6% better than control method) compared to the uncompensated resolution, using spatial information and both spatial and temporal information, respectively. It is also noted that the neural networks also produce a smaller constant term,  $b$ , by around 5%, indicating that the neural network is performing detector calibration. The control method produces a constant term with a comparatively large error compared to the neural networks SC and uncompensated response.
- By contrast to the control, the neural network methods produce a much lower reduced  $\chi^2$  consistent with 1, indicating superior agreement with expectations of the model in Equation 2.8.

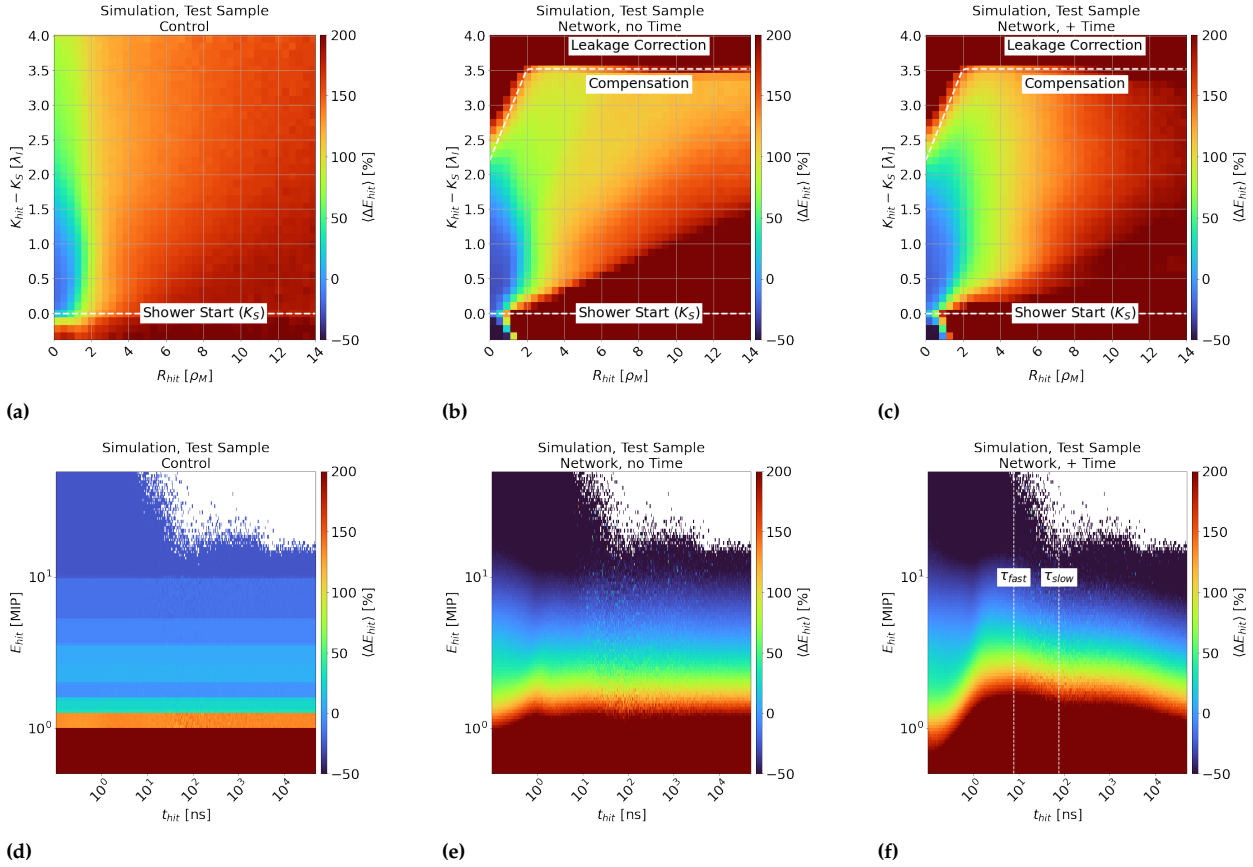
### 5.3.1.3 Correlations with Spatial and Temporal Information in Simulation

The results for the spatial correlations ( $R_{\text{hit}}, K_{\text{hit}} - K_S$ ) and energy-temporal correlations ( $t_{\text{hit}}, E_{\text{hit}}$ ) are shown in Figure 5.26 the top row (Figures 5.26a-5.26c) and bottom row (Figures 5.26d-5.26f), respectively, to the test sample without the tail-catcher cut applied, presented as the percentage change in energy as a result of the compensation algorithm having been applied.

#### Summary Comment to Figure 5.26

The neural network SC methods are demonstrated to learn a more sophisticated weighting technique that more effectively describes the expected temporal and spatial development of the hadron shower than the control. This improvement manifests as a superior sensitivity to the shower shape, leakage correction and MIP-track subtraction. The inclusion of time also results in a continuous weighting compared to the control and exhibits time dependence following the expectations of bi-exponential time distribution. These additional effects corroborate with the results of Section 5.3.1.2, where improved detector calibration was observed.

### Average Percentage Change of Original AHCAL Active Cell Energy for SC Models Trained On and Applied To Simulation, without the TCMT Cut Applied



**Figure 5.26:** Average percentage change in active cell energy in simulation ( $E_{hit}$ ) as a result of compensation as a function of  $R_{hit}$  and  $K_{hit} - K_S$  (top row, presented in units of Moliere radius from the lateral centre-of-gravity,  $\rho_M = 24.9$  mm, and nuclear interaction length from the shower start,  $\lambda_I = 237.1$  mm, respectively), and  $E_{hit}$  and  $t_{hit}$  (bottom row, presented in units of MIP and ns, respectively). Each column indicates the control, network method without and with timing information in that order. The colour axis indicates the percentage change, where blue regions indicate where the energy has been attenuated, and green through red shows where the energy has been enhanced. White space indicates no data available. Regions of interest are labelled accordingly for reference. The values of  $\tau_{slow}$  and  $\tau_{fast}$  were taken from [32].

#### Comments:

- Figure 5.26a demonstrates that the control method shows only a weak dependence on lateral and longitudinal development of the shower, with attenuation occurring only within  $R_{hit} \lesssim 1\rho_M$  (the EM fraction) and enhancement beyond, with minor variation, as expected. By contrast, the neural network methods attenuate and enhance the active cell energy with much stronger spatial dependence, indicated by the broadening of the weighting profile with longitudinal shower development.
- Two additional effects are observed for the network methods, shown in Figure 5.26b and Figure 5.26c. The first is a tendency to enhance  $E_{hit}$  in the region above the white dashed line. The second is to attenuate  $E_{hit}$  where  $R_{hit} \lesssim 1\rho_M$  (close to the lateral shower core) and  $K_S < 0$  (before the shower start). These effects are not present in Fig. 5.26a and must therefore be a consequence of including spatial information in the models. These results suggest the network models have learned leakage correction and to remove the energy deposited by minimum ionisation of the  $\pi^-$  particle before showering, demonstrating an improved capacity of the proposed model to learn the physical properties of hadron showers and detector geometry compared to the control method.
- Figures 5.26d- 5.26f demonstrate that all methods are observed to attenuate active cell energies above 5 MIP and enhance below that threshold, which is expected of all SC algorithms. The binned structure of the weighting of the control method is visible in Figure 5.26d. By contrast, the neural network methods in Figures 5.26e and 5.26f indicate a continuous weighting function has been learned. Furthermore, Figure 5.26f indicates that the model with timing information enhances the threshold for energy deposited in the order of several ns to several tens of ns, A reduction in the threshold is observed after around 100 ns. These observations are consistent with the timescales of the two main neutron energy-depositing processes discussed in Section 2.5.1.1. Comparison of Figure 5.26f and Figures 5.26d and 5.26e indicate that this effect must be due to the inclusion of timing information since no such effect is observed in the control or method without timing information.

#### 5.3.1.4 Robust Statistics of Response Distributions in Simulation

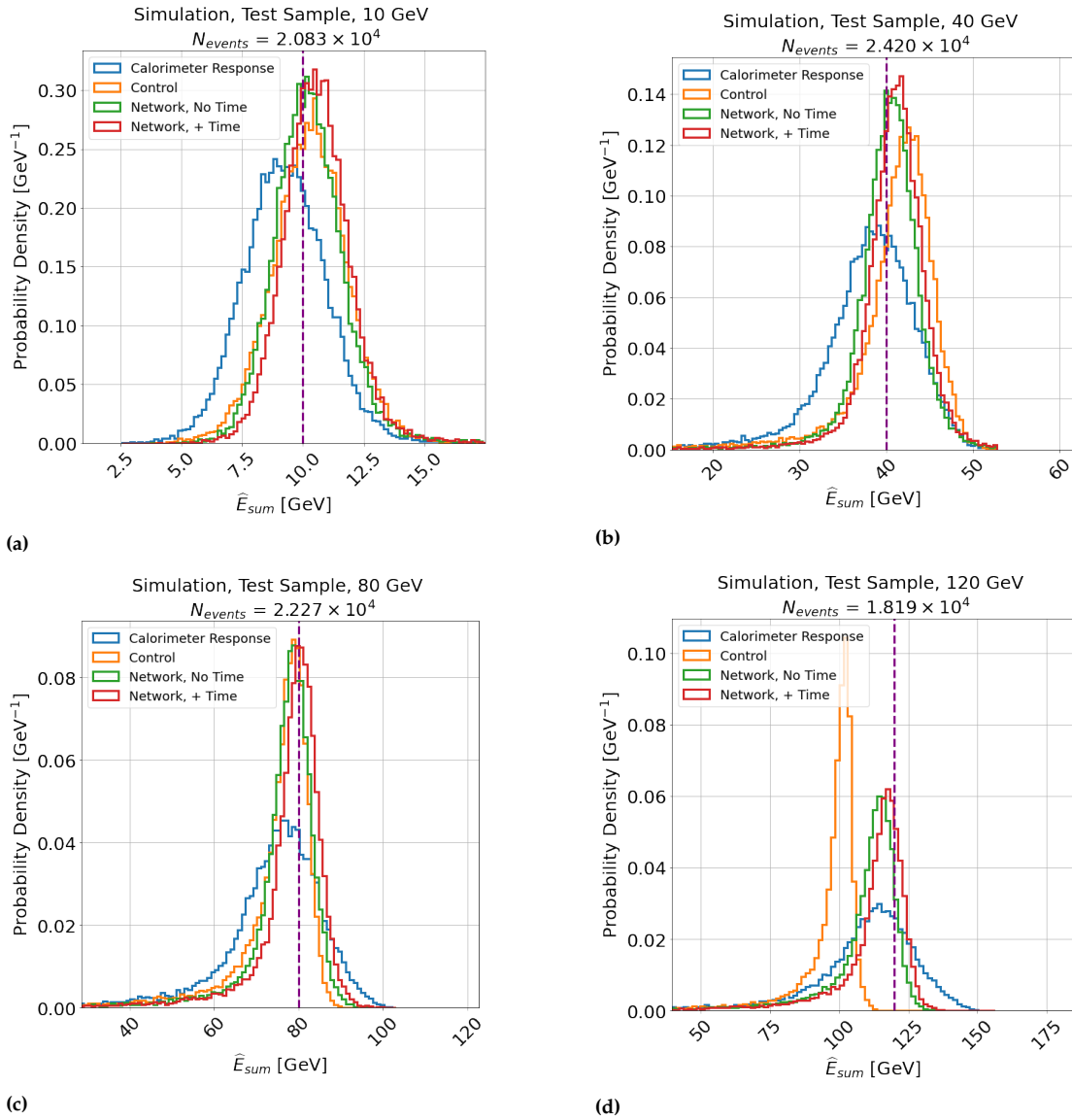
A study of the consequence of the compensation methods was also performed on simulation without applying the TCMT cut (i.e. the same selection criteria as the training data). This was performed to assess the influence of leakage correction on the energy spectra of the test sample. Figure 5.27 shows the same distributions as in Figure 5.24, without the TCMT cut applied. A full set of distributions is available in Appendix Section 8.4.

The consequences of the compensation methods on the resolution distribution with leakage are presented in terms of robust, non-parametric statistics of the compensated distributions. Robust statistics were chosen over standard moments (mean, standard deviation, standard skewness), as outliers may heavily influence these statistics and lead to incorrect conclusions about the distribution, which is particularly relevant to energy spectra with a high proportion of leakage. By contrast, robust statistics require a significantly greater number of statistics to be outliers (e.g. arbitrarily small/large observations) before yielding an incorrect (e.g. arbitrarily small/large) result. The distributions' centrality, spread and skewness were studied with the median, median absolute deviation [168] (MAD), and medcouple [169]. The median describes the value that separates the lower from the upper half of the ordered values of a distribution. The MAD is defined as the median of absolute differences between the distribution and its median. The medcouple is a scaled median difference of the left and right half of a distribution relative to its median. The medcouple is defined between -1 (entirely left-tailed) and 1 (entirely right-tailed). A medcouple of 0 indicates left-right-tailed symmetry. The results are shown in Figure 5.28.

##### Summary Comment for Figure 5.27 and Figure 5.28

The SC models result in better linearity of response and skewness and smaller variance than the control. This indicates that the network methods have learned not only SC but leakage correction, indicating that the model has learned the local properties of the detector.

**Reconstructed Energy Distributions for SC Models Trained On and Applied To Simulation, without the TCMT Cut Applied**

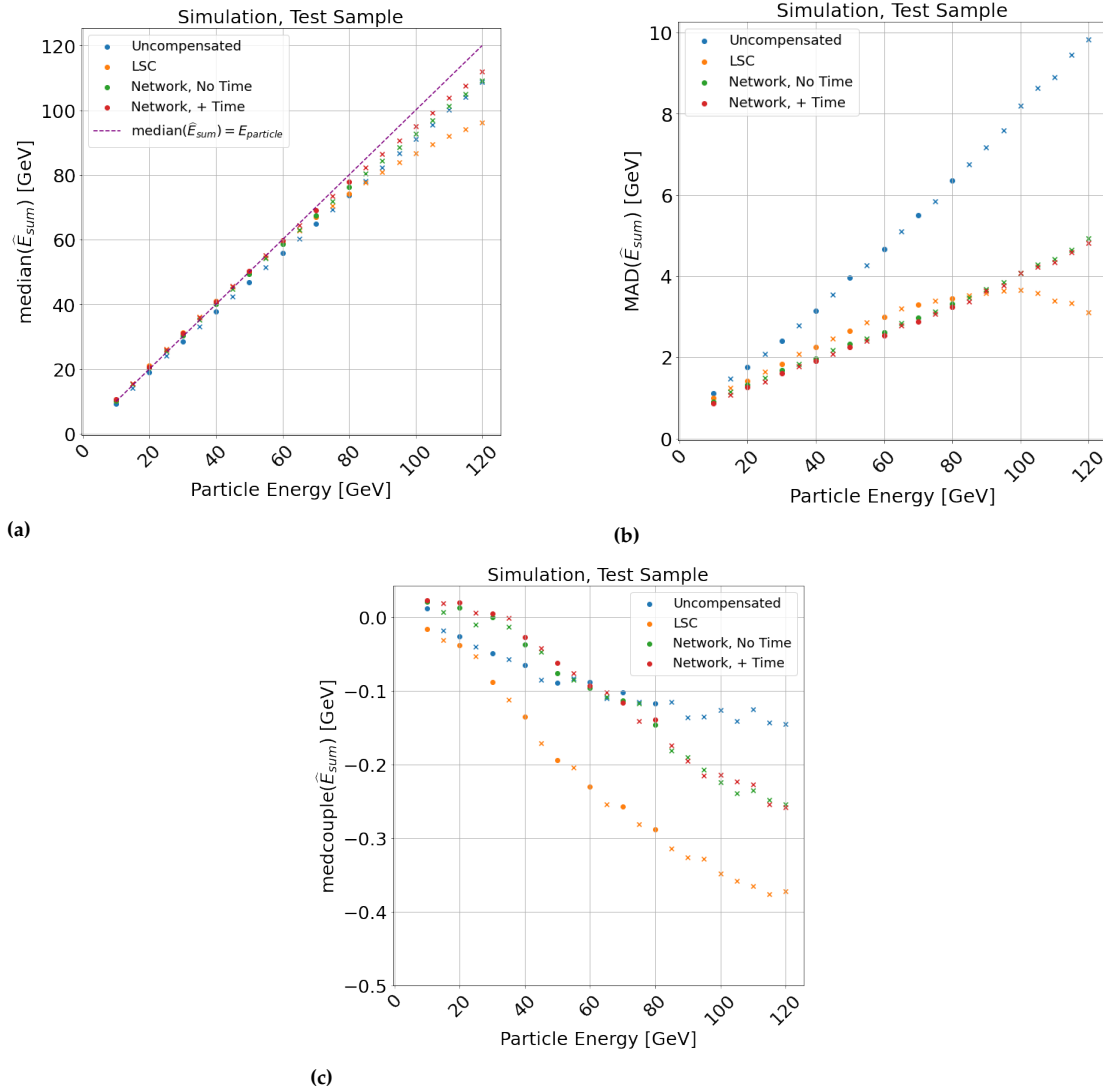


**Figure 5.27:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the test dataset of Table 5.1, without the TCMT cut applied. Else, as in Figure 5.24.

**Comments:**

- ▶ The distributions shown indicate that the application of the neural network methods results in a reduction of leakage, shown as the reduced tail in the green and red distributions compared to the blue.
- ▶ Else, as in Figure 5.24.

### Median, MAD and Medcouple of the Reconstructed Energy Distributions for SC Models Trained On and Applied To Simulation, without the TCMT Cut Applied



**Figure 5.28:** Figure 5.28a, Figure 5.28b and Figure 5.28c show the median, median absolute deviation (MAD) and medcouple of the response distributions of each model applied to the test sample of simulation, without TCMT cut. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods. Circle and cross markers indicate energies used for both training and testing and testing only. Figure 5.28a shows the fitted  $\mu = E_{particle}$ , where the purple dashed line indicates  $\mu = E_{particle}$ .

#### Comments:

- Figure 5.28a demonstrates the same conclusions as made for Figure 5.25a. The shift observed beyond the training range in Figure 5.28a compared to Figure 5.25a is a consequence of the leakage biasing the distribution and demonstrates the necessity of performing a leakage cut to assess the resolution of the methods.
- Figure 5.28b shows that the spread of the neural network methods is always the same or smaller than the control method, with un-physical resolutions above the 80 GeV training sample for the control method, again due to energy biasing for this method. The improvement in resolution due to time is less pronounced than shown in Figure 8.4 and Figure 5.25b. This is a consequence of the MAD remaining sensitive to leakage and once again demonstrates the necessity of performing a leakage cut.
- Figure 5.28c shows that, for the training range, the left-tailed skewness is reduced (i.e. closer to 0) over the uncompensated distribution for the neural network methods, whereas it is increased significantly using the control method. Above the training range, both methods produce a more heavily left-tailed distribution than the uncompensated sample. This may be explained by the compensation of high-energy hadron showers reducing the proportion of events above the median compared to below. Nonetheless, the network methods produce a less left-skewed response distribution than the control method. This plot demonstrates the positive effect of the leakage compensation learned by the neural network on the calorimeter response.

## 5.3.2 Data Model

In this section, the models trained on the training set and applied to the test set of data in Table 5.1 are discussed. It is also compared to the network model trained on the training set of simulation. This is done to assess the differences between simulation and data and the effectiveness of a model trained on simulation and applied to data.

### 5.3.2.1 Response Distributions in Data

As in Section 5.3.1, the normalised response distributions of data are shown in Figure 5.29. A full response distribution set is available in Appendix Figure 8.5.

#### Summary Comment to Figure 5.29

As in Section 5.3.2.1, the network methods produce a superior resolution than the control method for experimental data, indicated by the smaller variance of the responses using the neural network than the control method for most of the training dataset.

### 5.3.2.2 Resolution and Linearity of Response in Data

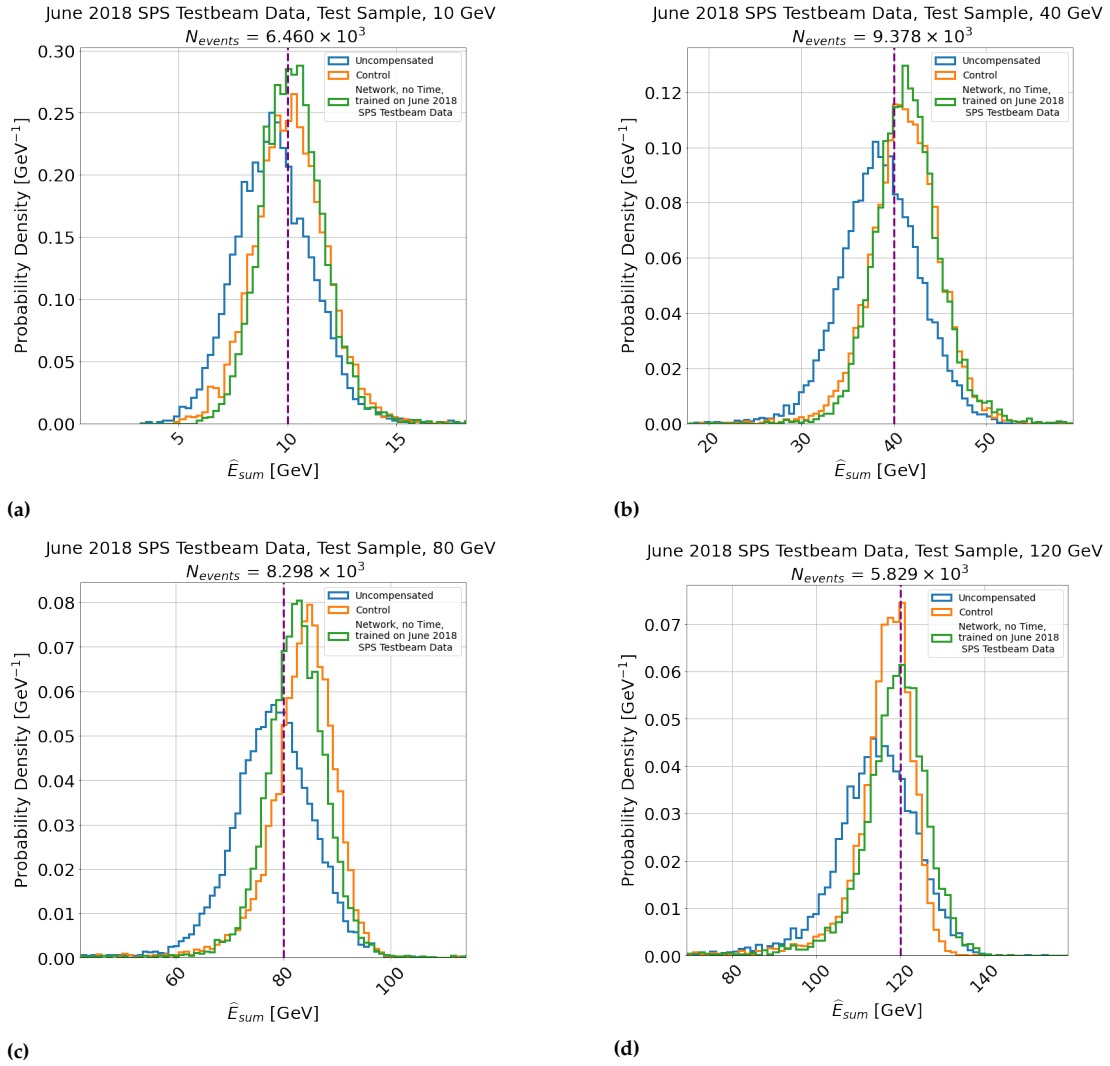
As in Section 5.3.2.1, the study resolution ( $\sigma/\mu$  vs.  $E_{\text{particle}}$ ) and linearity of response ( $\mu/E_{\text{particle}}$  vs.  $E_{\text{particle}}$ ) is shown in Figure 5.30. The fitted values are shown in Appendix Table 8.4. Fit results of Equation 2.8 are shown in Table 5.3.

#### Summary Comment to Figure 5.25 and Table 5.3

As in Section 5.3.2.1, the machine learning methods outperform the control method in resolution, resulting in an improvement of the intrinsic stochastic resolution term of 9.3 % and 12.2 %, outperforming the control method in both cases. The neural networks also reduce the constant resolution term by 2 %, indicating the neural networks perform some detector calibration and SC. A slightly superior linearity of response overall is observed compared to the control method and less than for simulation.



### Reconstructed Energy Distributions for SC Models Trained On and Applied To 2018 June SPS Testbeam Data, with the TCMT Cut Applied

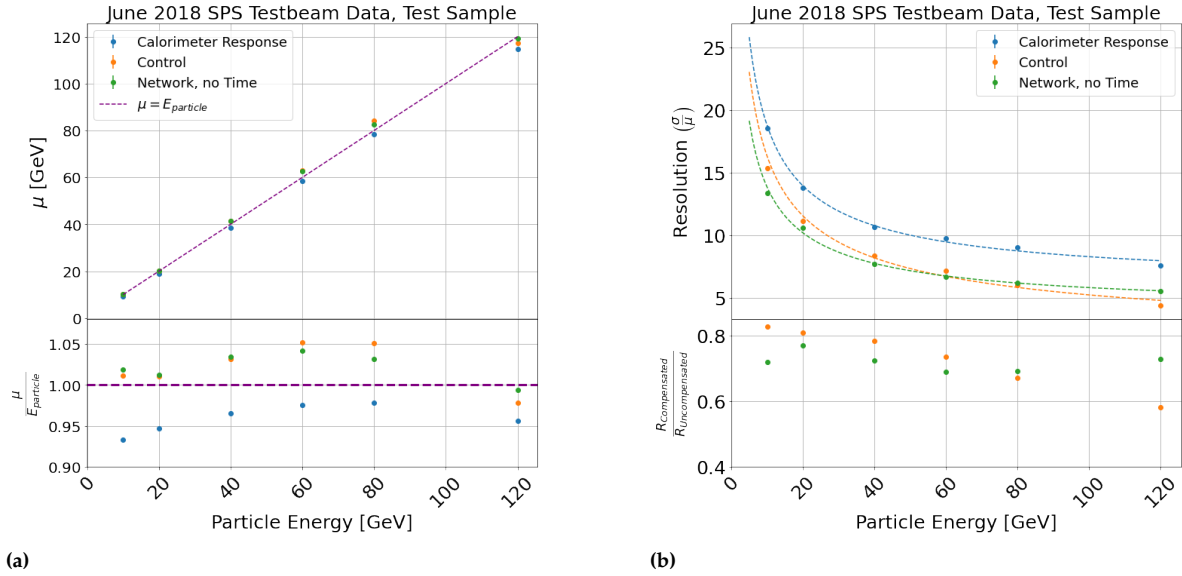


**Figure 5.29:** Example normalised histograms showing the data calorimeter response before and after compensation applied to the test dataset of Table 5.1, with the TCMT cut applied. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods, respectively.  $E_{\text{particle}}$  is indicated as a dashed purple line. In this sample, the 10 GeV, 40 GeV, 80 GeV and 120 GeV samples are shown.

#### Comments:

- As in Figure 5.24, the neural network method produces superior resolution than the control up to the 120 GeV sample shown in Figure 5.29d, for which the control method is superior. This again indicates that the control method also has a bias to the training range in data.

### Linearity and Resolution Fits for SC Models Trained On and Applied To 2018 June SPS Testbeam Data, with the TCMT Cut Applied



**Figure 5.30:** Figure 5.30a and Figure 5.30b show AHCAL linearity of response and resolution to data using all methods under test, respectively. Blue indicates intrinsic calorimeter response, while orange and green indicate the control, network without time compensation methods, respectively. Circle markers indicate energies used for both training and testing. Figure 5.30a shows the fitted  $\mu$  to  $E_{\text{particle}}$ , where the purple dashed line indicates  $\mu = E_{\text{particle}}$ . The bottom subplot shows the ratio of  $\mu$  to  $E_{\text{particle}}$ . Figure 5.30b shows the fitted  $\sigma/\mu$ , where the dashed lines indicate fits of Equation 2.8. The bottom subplot indicates the ratio of the resolution of each compensation method to the intrinsic response.

#### Comments:

- ▶ Figure 5.30a indicates that, from 10-40 GeV, the neural network and control methods provide similar linearity of response. In the 40 GeV-120 GeV range, the neural network method demonstrates a slightly superior linearity of response by no more than 1%. Notably, the improvement in linearity of response in simulation, shown in 5.25a, is more significant than in data. Nonetheless, all methods are observed to reconstruct the linearity of response within 5% of  $E_{\text{particle}}$  and overestimate the hadron shower energy as in Figure 5.30a.
- ▶ Figure 5.30b indicates that the neural network outperforms the control method significantly in data up to 80 GeV. This is reflected in Table 5.4.

	a [%]	b [%]	$\chi^2/NDF$
Calorimeter Response	$56.124 \pm 0.684$	$6.117 \pm 0.133$	10.063
Control	$51.501 \pm 0.416$	$1.017 \pm 0.325$	38.919
Network, No Time	$41.859 \pm 0.491$	$4.064 \pm 0.083$	6.477

**Table 5.4:** Table of fitted parameters of Equation 2.8 to the training range of energies in simulation shown as dashed lines in Figure 5.30b. All methods were fitted to the full energy range of the test sample.

#### Comments:

- ▶ The uncompensated stochastic resolution for data  $\pi^-$  hadron showers in AHCAL is in agreement within 1-2% with the stochastic resolution term  $a = 57.70 \pm 1.06\%$  obtained in [34].
- ▶ The neural network is seen to improve the stochastic resolution term,  $a$ , by around 9.6% compared to the control method and yields a similar stochastic resolution to simulation, within 1-2% of the same neural network model trained on and applied to simulation in Table 5.3, which suggests the neural network can be trained effectively with experimental data.
- ▶ It should be noted as a caveat that the control method produces a smaller constant  $b$  term than the neural network by around 3%, which is most likely biased since the reduced  $\chi^2$  of the fit is significantly worse for the control than the neural network methods.
- ▶ As in Section 5.3.1.2, the neural network model provides a slightly better reduced  $\chi^2$  than uncompensated data, whereas the control method produces a significantly worse agreement to uncompensated data.

### 5.3.2.3 Correlations with Spatial and Temporal Information in Data

As in Section 5.3.1.3, the spatial and energy-temporal correlations are presented in Figure 5.32.

As an additional measurement, the relationship between  $E_{\text{hit}}$  and the average  $\widehat{E}_{\text{hit}}$  (i.e. the relationship between the input active cell energy and the compensated cell energy) was performed.

The curve was fitted for all active cells and for selected MIP-track hits<sup>1</sup> and is shown in Figure 5.33 and Figure 5.33a, respectively.

This curve relating  $E_{\text{hit}}$  and the average  $\widehat{E}_{\text{hit}}$  was highly non-linear, and therefore local minima, local maxima of its curvature<sup>2</sup> and the crossing points between the curve and  $\langle \widehat{E}_{\text{hit}} \rangle = E_{\text{hit}}$ , indicating where attenuation and enhancement take place, were measured from a spline fit to the measured compensated energy value. Regions are also compared to the study of energy ranges presented in [171], which proposes that calorimeter energy deposits can be characterised according to three bins: EM-like ( $E_{\text{hit}} > 3.5$  MIP), HAD-like ( $1.7 \text{ MIP} \leq E_{\text{hit}} \leq 3.5$  MIP) and track-like ( $E_{\text{hit}} < 1.7$  MIP).

Curvature is defined according to Equation 5.4:

$$\kappa(x) = \frac{|y''(x)|}{(1 + y'^2(x))^{3/2}} \quad (5.4)$$

where  $\kappa(x)$  is curvature,  $x$  and  $y(x)$  are a line's independent and dependent variables, and primes indicate the order of the derivative.

#### Summary Comment to Figure 5.32

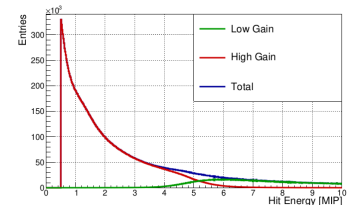
As in Section 5.3.1.1, the machine learning methods learn a more sophisticated weighting technique than the control. A different weighting is learned for simulation than data, which reflects the differences in energy density between simulation and data observed in Figure 5.21.

#### Summary Comment to Figure 5.33

Regions of interest in the relationship between  $E_{\text{hit}}$  and the average  $\widehat{E}_{\text{hit}}$  are found to be in agreement with the values predicted for different physics regimes by [171]. The threshold between attenuation and enhancement, on average, strongly agrees with the threshold for the gain-switching mode of AHCAL. The neural network learned these features without prior information than hadron shower events. The track weighting is found to weight down MIP-like hits with energy in the range 0.8-1.5 MIP, which is in agreement with the results of Figure 5.32.

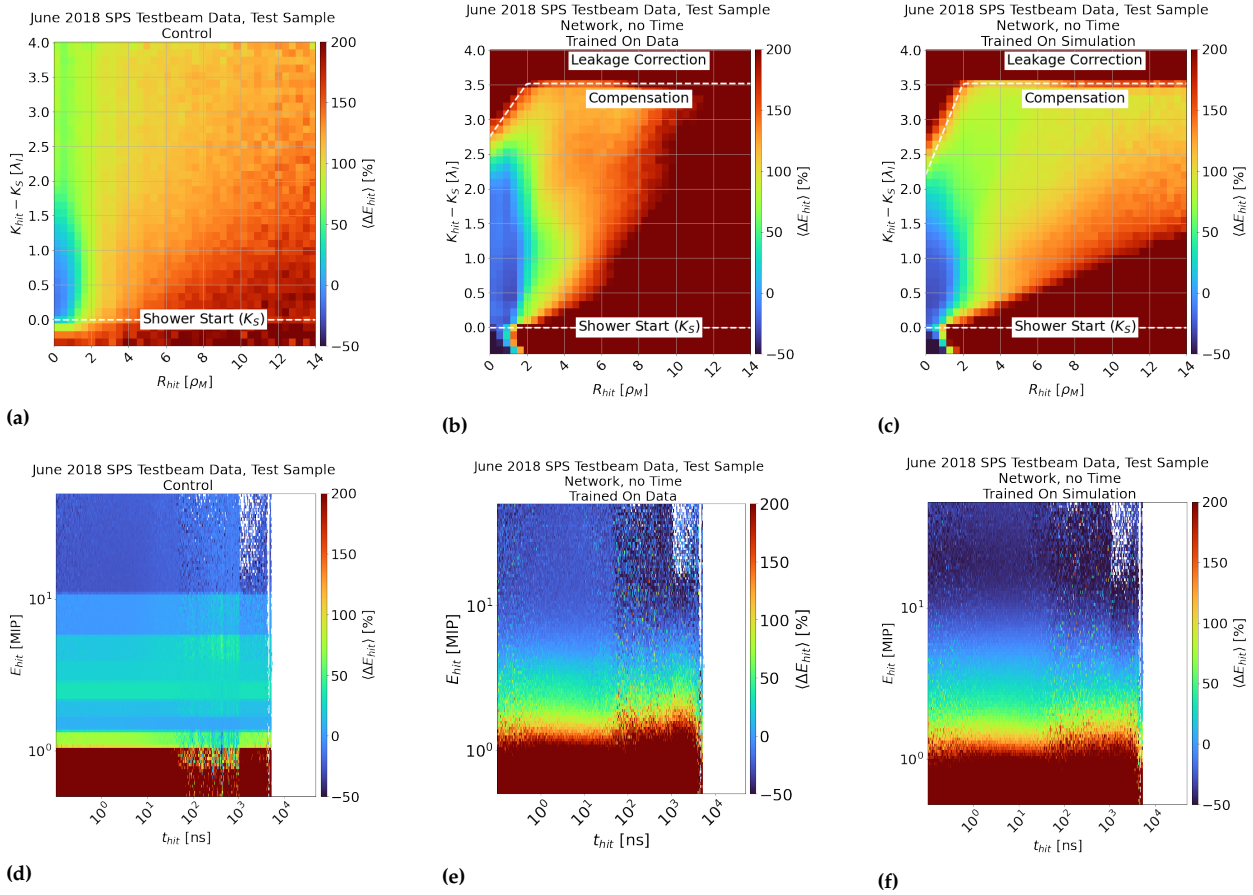
1: MIP-track hits are selected according to a special cut which is described in full in Section 6.2.3.1 in Chapter 6. The cut consists of hits fulfilling the following criteria:  $R_{\text{hit}} < 60$  mm,  $E_{\text{hit}} < 3$  MIP and  $K_{\text{hit}} - K_S \leq -2$

2: 'Knee' and 'elbow' points of a curve are subjective and depend on the choice of coordinate system [170]. Additionally, 'bumps' in the curve are not local minima because they may be on an increasing slope. Local maxima in curvature are therefore used to indicate where the line changes direction most prominently in a mathematically straightforward manner.



**Figure 5.31:** Reconstructed global hit energy spectrum of 100 GeV electron showers for the June 2018 testbeam indicating hit amplitudes measured in high gain (red line) or low gain (blue line) mode. The sum of red and blue curves is indicated by the green curve, which shows a smooth transition around 5 MIP. Taken from [78].

**Average Percentage Change of Original AHCAL Active Cell Energy for SC Models Trained On and Applied To 2018 SPS June Testbeam Data, without the TCMT Cut Applied**

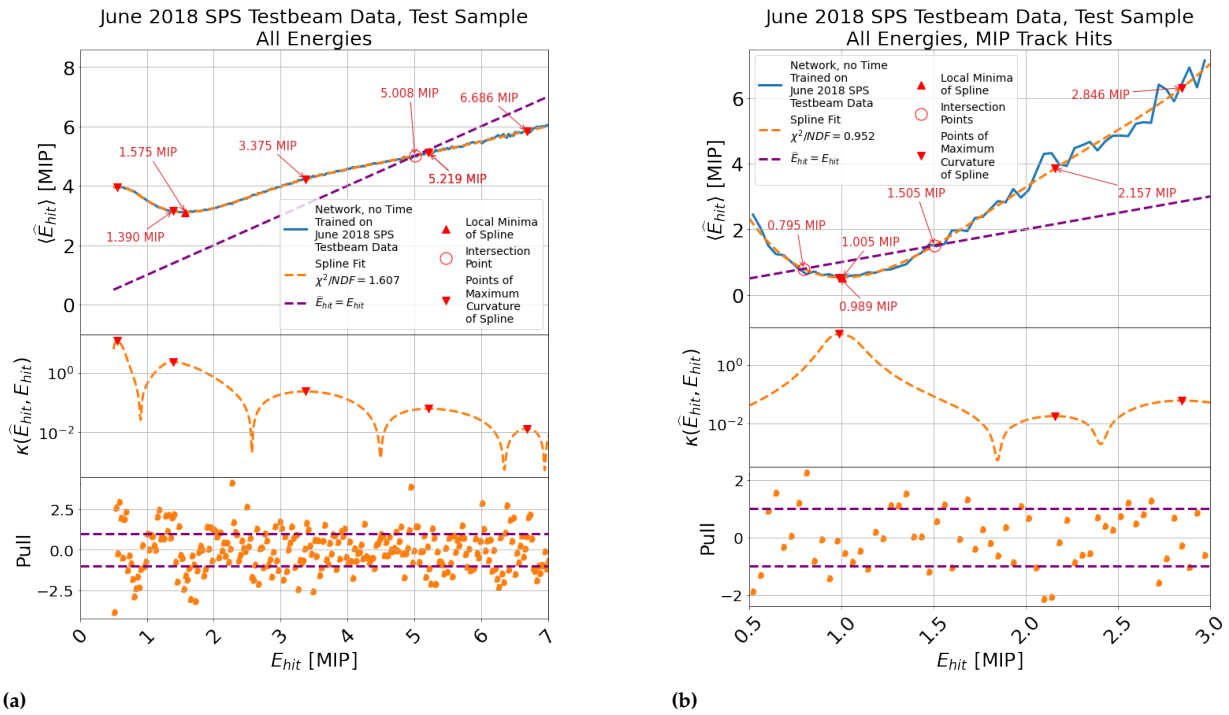


**Figure 5.32:** Average percentage change in active cell energy in data. Each column indicates the control and network methods without timing information in that order, with the middle column trained on data and the rightmost column trained on simulation. Else, as in Figure 5.26.

**Comments:**

- ▶ The control method shown in Figure 5.32a indicates weaker spatial correlations of the compensation than Figure 5.32b.
- ▶ Figure 5.32b demonstrates the model trained on data also learns a different spatial weighting than the control method trained on data 5.32a. Most notably, the core and halo, indicated by the blue region where  $(R_{hit} < 1 \rho_M \text{ and } K - K_S > 0)$  and in the red region where  $R_{hit} < 7.5 \rho_M$ . This is due to the different energy densities observed between simulation and data shown in Figure 5.21. Further analysis is provided in the comments to Figure 5.33 and Figure 5.36a.
- ▶ Figure 5.32d and Figure 5.32e both indicate that the model learns to attenuate active cell energies above and enhance active cell energies below  $E_{hit} = 5$  MIP, and do not show strong correlations with the time measured by AHCAL. This agrees with Section 5.3.1.3 and Figure 5.26 that when timing information is not included, correlations with time are not observed.

### Average Compensated Active Cell Energy Compared to the Original Active Cell Energy using the SC Model Applied to 2018 June SPS Testbeam Data, without the TCMT Cut Applied



**Figure 5.33:** Relationships between average compensated hit energy ( $\langle \hat{E}_{hit} \rangle$ ) and the original calorimeter active cell energy ( $E_{hit}$ ). The top subplot of Figure 5.33a and 5.33b show the original curve, shown in blue, and a spline fit, shown as a dashed orange line, to the network model output, without timing information, trained on June 2018 SPS Testbeam data for all active cells and MIP-track cells, respectively. The red upwards-facing triangles, red downwards-facing triangles and red empty circle markers indicate local minima, local maxima in curvature, and crossing points with  $\langle \hat{E}_{hit} \rangle = E_{hit}$ , therefore indicating where enhancement and attenuation are observed. The middle subplot shows the curvature of the spline on a logarithmic scale. The bottom subplot shows the pulls of the spline to data.

#### Comments:

- ▶ Good agreement is observed between the spline fit and the average compensated energy, with a reduced  $\chi^2$  of 1.6 in Figure 5.33 and 0.95 in Figure 5.33b.
- ▶ Figure 5.33 indicates a local minimum in the compensation curve found at 1.6 MIP, and local maxima of curvature (excluding the trivial point at 0.5 MIP) are found at 1.4 MIP, 3.4 MIP, 5.2 MIP and 6.7 MIP. It is noted that the range between the local minimum and the next local maximum of curvature (1.6 MIP and 3.4 MIP) is similar to the range suggested in [171] (1.7 MIP and 3.5 MIP, and differ both by 0.1 MIP (2.68 MeV).
- ▶ It is also noteworthy that there is a single crossing point for the line  $\langle \hat{E}_{hit} \rangle = E_{hit}$  at 5 MIP. This value is the energy at which the SIPMs of AHCAL switch from low-gain to high-gain mode, which is illustrated in Figure 5.31. The model learns to attenuate active cell energies above this value and enhance them below, indicating it has learned properties of the detector apparatus with no prior information.
- ▶ Figure 5.33b indicates a local minimum at 1 MIP, points of local maxima of curvature at 0.99 MIP, 2.2 MIP and 2.8 MIP. Two crossing points are observed for the line  $\langle \hat{E}_{hit} \rangle = E_{hit}$ , at 0.8 MIP and 1.5 MIP. The main observation is that the model attenuates the energy between these limits, most strongly at 1 MIP, and the model enhances it beyond them. The result is consistent with the MIP-subtraction weighting shown in Figure 5.32a.

#### 5.3.2.4 Robust Statistics of Response Distributions in Data

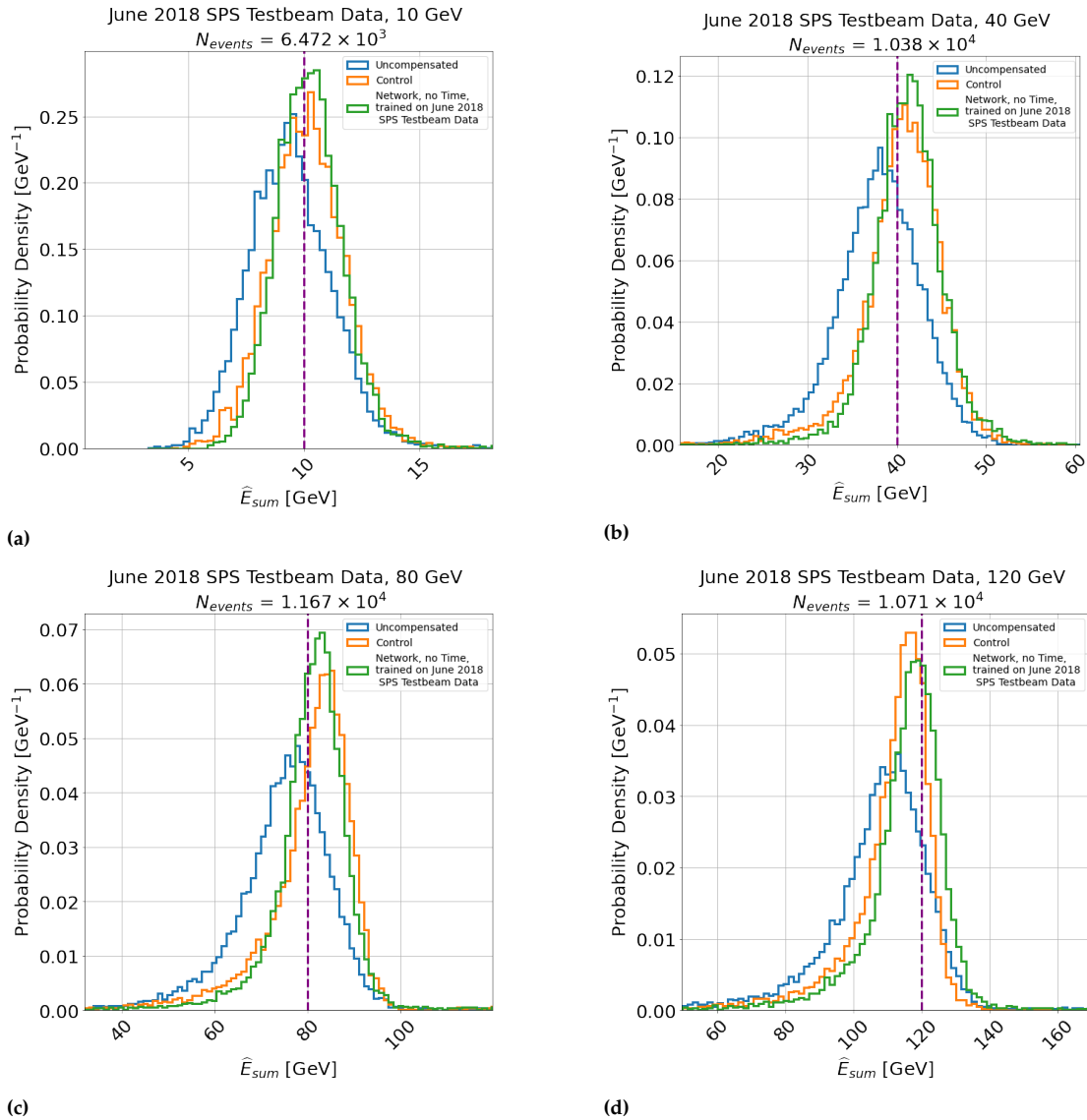
As in Section 5.3.1.4, sample distributions without the TCMT cut are presented for data in Figure 5.34, with a full set of figures available in Appendix Figure 8.6.

Robust statistics were applied to the calorimeter response distributions without the TCMT cut, to assess the effect of each model on leakage. The results are shown in Figure 5.35.

#### Summary Comment for Figure 5.27 and Figure 5.28

As in Section 5.3.1.4, the SC models resulting in better linearity of response and skewness and smaller variance than the control. This indicates that the network methods have learned not only SC but leakage correction, indicating that the model has learned local properties of the detector.

### Reconstructed Energy Distributions for SC Models Trained On and Applied To 2018 June SPS Testbeam Data, without the TCMT Cut Applied

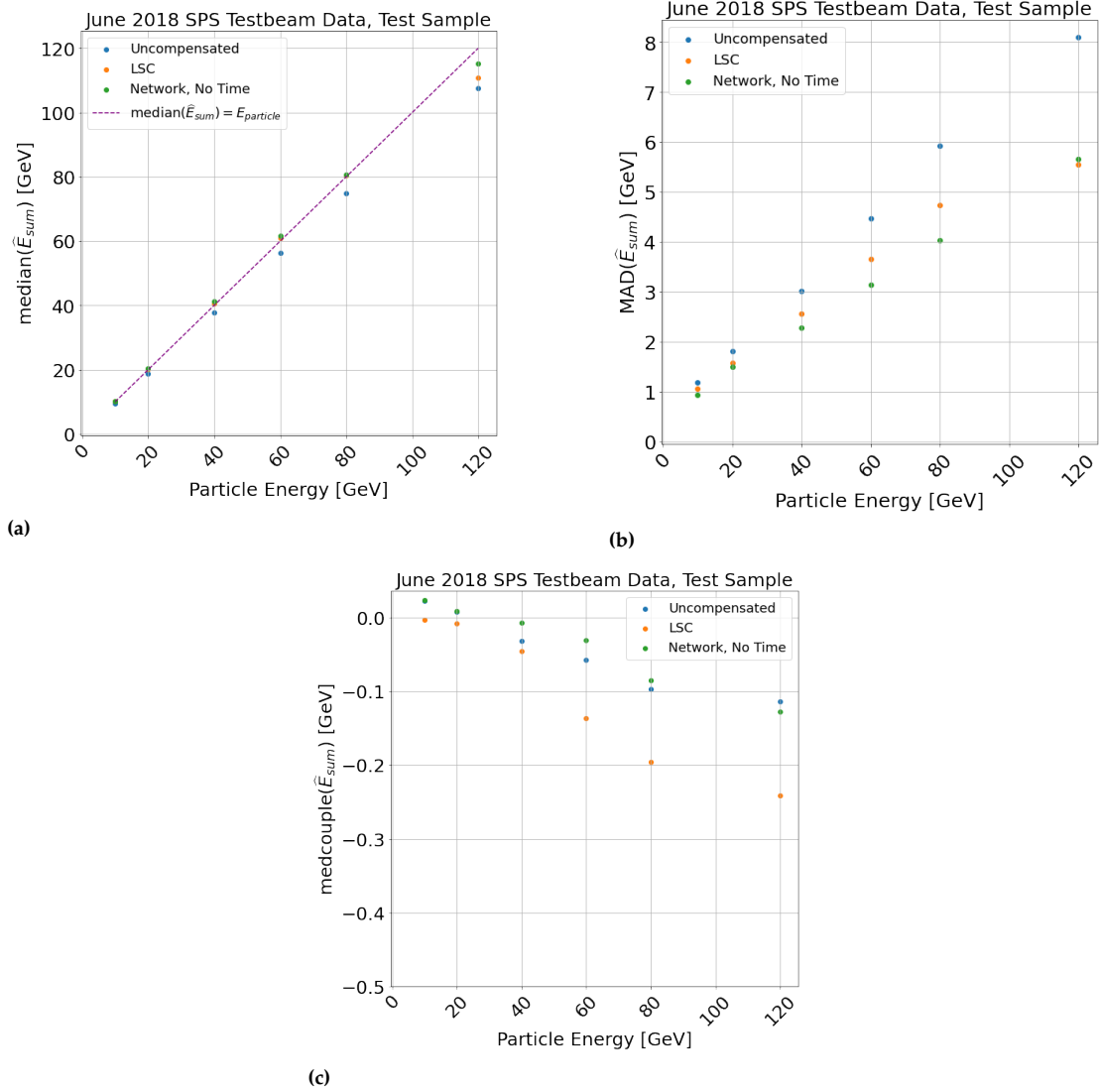


**Figure 5.34:** Example normalised histograms showing the data calorimeter response before and after compensation applied to the test dataset of Table 5.1, without the TCMT cut applied. Else, as in Figure 5.29.

#### Comments:

- As in Figure 5.24.

**Median, MAD and Medcouple of the Reconstructed Energy Distributions for SC Models Trained On and Applied To 2018 June SPS Testbeam Data, without the TCMT Cut Applied**



**Figure 5.35:** Figure 5.35a, Figure 5.35b and Figure 5.35c show the median, median absolute deviation (MAD) and medcouple of the response distributions of each model applied to the test sample of data, without TCMT cut. Else, as in Figure 5.28

**Comments:**

- As in Figure 5.28.



### 5.3.2.5 Comparison of Simulation Model to Data Model

The differences between models trained on simulation and data are now discussed.

The mean compensated energy ( $\widehat{E}_{\text{hit}}$ ) is presented as a function of the original compensated energy ( $E_{\text{hit}}$ ) for the model trained on simulation and data in Figures 5.33 and 5.33b for the test sample in data, and are compared using the hit classification regimes of [171]. Figure 5.33 shows the overall average compensated energy, while 5.33b shows the MIP-track hits as discussed in Section 5.3.2.3.

Additionally, the response distributions of the model trained on simulation and data are shown in Figure 5.37. Finally, the influence on the mean response  $E$  and  $\sigma_E$  of the response distributions by using a model trained on simulation and a model trained on data are studied using the location and scale parameter using the Gaussian fits applied to measure resolution. These are shown in Figure 5.38.

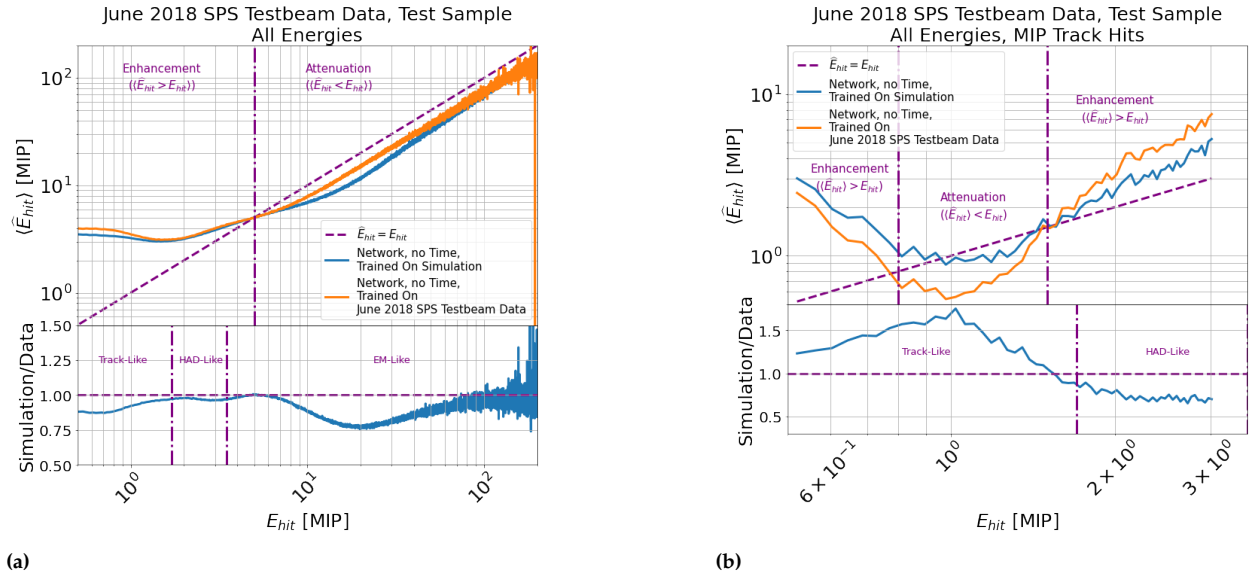
#### Summary Comment for Figure 5.36

Simulated hadron shower events were observed in Figure 5.21 to be more energy dense than hadron showers in data, on average. The average modified active cell energy from the compensation algorithm trained on data is observed to always be greater, on average, than for simulation. These observations are in agreement. This result suggests that the compensated response trained on simulation will produce a systematically lower mean response when applied to data.

#### Summary Comment for Figure 5.37

While the network compensated response distributions are essentially the same for 10 GeV hadron showers, as the shower energy increases, the mean of the response is less for the network trained on simulation than on data. The most probable reason for this is that the SC network trained on simulation over-attenuates the hadron shower core because it has been trained on hadron showers which are more energy dense on average.

**Average Compensated Active Cell Energy Compared to the Original Active Cell Energy using the Network SC Model Trained on Simulation and the Network SC Model Trained on 2018 June SPS Testbeam Data, Both Applied to Data, without the TCMT Cut Applied**

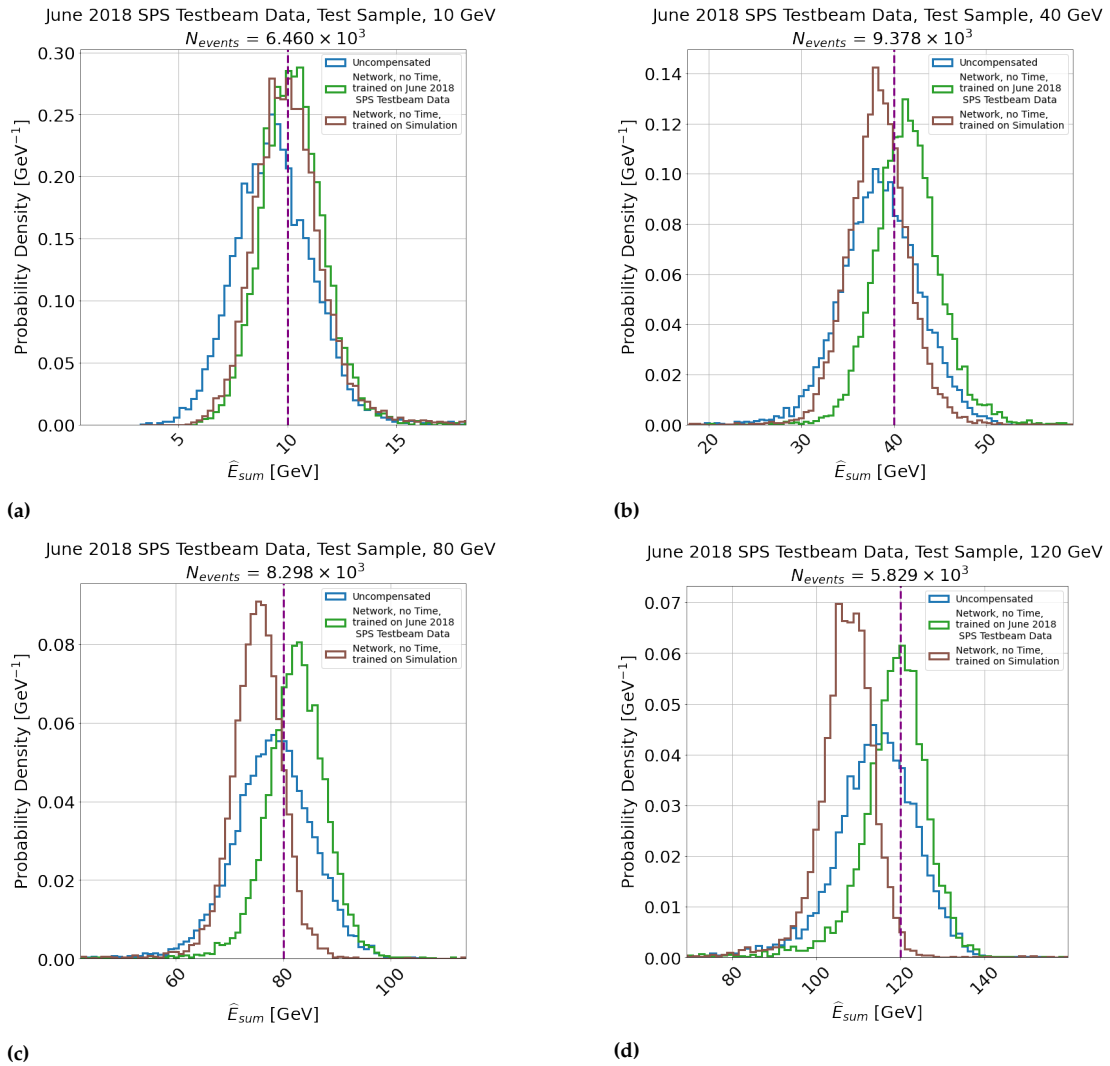


**Figure 5.36:** Relationships between average compensated hit energy ( $\langle \hat{E}_{hit} \rangle$ ) and the original calorimeter active cell energy ( $E_{hit}$ ). As in Figure 5.33, except that the network model output, without timing information, trained on simulation is shown as the blue curve, and the same model trained on data is shown as the orange curve. The vertical purple dash-dotted lines estimate the crossing point between attenuation and enhancement of the cell energy. The bottom subplot indicates the ratio of simulation to data. The regions for EM-like, HAD-like and track-like energy depositions from [171] are indicated.

**Comments:**

- ▶ Figures 5.21a-5.21c indicate that simulation has an excess of energy density both in the hadron shower core and in the hadron shower halo in comparison to data. Compensation attenuates the EM-core's energy density and enhances the HAD-halo's energy density. It should be expected then that the SC model trained on simulation should attenuate the EM-like active cell energies of the hadron shower more strongly and enhance the HAD-like halo more weakly in the model trained in simulation than for the model trained in data (i.e.  $\hat{E}_{hit}$  should be greater overall for data than simulation because both the core and halo of data are observed to be less energy-dense overall than for simulation).
- ▶ Figure 5.36a shows that the model trained on data, indicated by the orange line, always produces a larger mean compensated response than simulation, indicated by the blue line in the subplot falling below the dashed purple line. This agrees with the expectations of Figure 5.21a-5.21c.
- ▶ Figure 5.36b indicates the opposite behaviour for most of the distribution. This again agrees with the expectations of Figure 5.21a-5.21c, which indicates the energy density in the MIP region close to the core and before the shower start has more energy in data than in simulation, indicated by the blue region in those plots.

### Reconstructed Energy Distributions for SC Models Trained On Simulation and 2018 June SPS Testbeam Data and Applied To Data, with the TCMT Cut Applied



**Figure 5.37:** Example normalised histograms showing the simulated calorimeter response after compensation applied to the data test dataset of Table 5.1. Blue lines indicate intrinsic calorimeter response, while green and brown lines indicate the network without time compensation methods, trained on simulation and data, respectively.  $E_{particle}$  is indicated as a dashed purple line.

#### Comment:

- Comparing the brown curve to the green curve indicates that at 10 GeV, the response for the network trained with simulation is around the same as the one trained on data. However, as the particle energy increases, the mean response of the brown distribution decreases compared to the known particle energy. This result indicates that the model underestimates the compensated energy. This result agrees with expectations from Figure 5.36a and therefore the hypothesis that the model trained on simulation over-attenuates the energy of the hadron shower core.

The influence on the mean response,  $E$ , and the response width,  $\sigma_E$ , was then studied to determine whether or not the effect was a result of a systematic discrepancy of the model, as indicated in Figure 5.36, or an effect of biasing. For this reason, the  $\mu$  and  $\sigma$  of the data distributions were fitted with a line and Equation 5.5, based on Equation 2.8:

$$\sigma_E = a\sqrt{E} \oplus bE \quad (5.5)$$

The variables were treated separately such that the expectations of the width of each distribution can be studied independently of the particular energy scaling of each model (i.e.  $\sigma$  and  $\mu$  should independently agree with expectations). The agreement with expectations was then assessed using the difference of each model to the fit value at each particle energy.  $\mu$  and  $\sigma$  were measured the same way as described in Section 5.3.1.2.

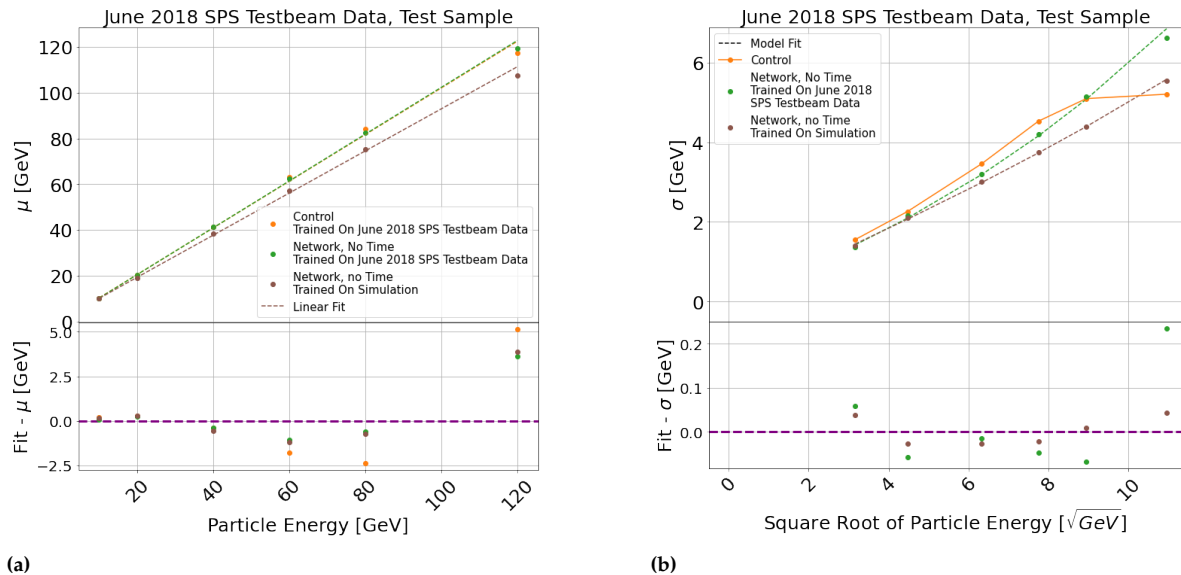
Under the hypothesis that biasing does not occur, the mean calorimeter response should remain linear, and the width should follow the expectations of Poisson-distributed stochastic fluctuations. In other words, divergences indicate bias from the physics expectations for compensation. For instance, this could manifest as saturation of either value.

The results are shown in Figure 5.38.

#### Summary Comment for Figure 5.37

Good agreement is observed with expectations of the relationship between  $\mu$ ,  $\sigma$  and the particle energy with both neural networks trained on simulation and data, respectively. There is, therefore, no evidence to support the hypothesis that the decrease in the mean response due to training the SC model with simulation compared to data is due to biasing; by contrast, the  $\sigma$  of the control method trained on data saturates with particle energy and is not in agreement with Equation 5.5, indicating bias to the training range.

### Relationship of $\mu$ and $\sigma$ of Reconstructed Energy Distributions of SC Models Trained On Simulation and 2018 June SPS Testbeam Data and Applied To Data with Particle Energy, with the TCMT Cut Applied



**Figure 5.38:** The top subplot of Figure 5.38a and Figure 5.38b show the  $\mu$  and  $\sigma$  for the network model trained with data and the network model trained with simulation. Orange, green and brown indicate the control method trained on data for reference, the network model without time trained on data and the network model without time trained on simulation, respectively. Dashed lines show fits. A linear fit is presented in Figure 5.38a as a function of particle energy, while a fit of Equation 5.5 is shown in Figure 5.38b as a function of the square root of the particle energy. The bottom subplot shows the difference between the compensated energy and the best-fit value. In Figure 5.38b, the solid orange line indicates the line between points for the control method, which could not be fitted with Equation 5.5 due to bias.

#### Comments:

- ▶ Figure 5.38a shows that, as expected, the mean response is lower on average for the SC model trained on simulation than the SC model trained on data. The reasons for this can be explained by the simulation being more energy-dense on average, as shown in Figure 5.21 and corroborated by Figure 5.36a.
- ▶ Figure 5.38a indicates that the  $\mu$  of all models under test agree well with the linearity of response. The neural network SC model trained on simulation agrees with the linear fit as the same model trained on data, indicated by the same differences observed for the brown and green points in the subplot. Better overall agreement is observed with the neural network methods than with the control method with the hypothesis of linearity, once again indicated by the greater differences shown by the orange points in the subplot.
- ▶ Figure 5.38b indicates that the  $\sigma$  of the neural network SC model trained on simulation is also found to agree with the expectations of Equation 5.5 as the same model trained on data, indicated by the similar agreement of the brown and green points in the subplot. By contrast, the control method could not be fitted with Equation 5.5, and is observed to saturate at particle energies greater than 80 GeV. This is a clear indication of unphysical bias to the training range.
- ▶ It is notable that the brown curve of Figure 5.38b sits below the green curve, indicating a smaller width. However, it is ambiguous from this plot as to whether this is due to superior compensation or the lower mean response predicted by the SC model.

## 5.4 Conclusion

A neural network method for performing software compensation was devised, trained, and tested on simulation for the AHCAL calorimeter. The model was designed to function using a local energy density estimate to overcome biasing effects on particle energies.

The neural network model was trained with and without timing information and is compared to a control method after accounting for the effect of leakage compensation learned by the networks. They yielded superior overall compensation and linearity of response to the control method when trained on simulation, resulting in calorimeter resolutions of  $40.2\%/\sqrt{E_{\text{particle}}} \oplus 2.2\%$  and  $37.2\%/\sqrt{E_{\text{particle}}} \oplus 2.4\%$ . This corresponded to an improvement of stochastic resolution by 9.3% and 12.2%, respectively. The constant resolution term was also found to reduce by around 5%, indicating that the model was capable of detector calibration. Both methods obtained a linearity of response within around 2-3% of the particle energy. The network without timing information trained on CALICE 2018 SPS testbeam data achieved a comparable resolution of  $41.9\%/\sqrt{E_{\text{particle}}} \oplus 4.0\%$ . This corresponded to an improvement in stochastic resolution term by 9.6% and of constant resolution by around 2%, indicating that the model can be trained with limited experimental data to a similar level as simulation. Additionally, the control method was observed to bias to the training range of particle energies. In contrast, the neural network method was demonstrated to both interpolate and extrapolate compensation to energies not used for training.

The basic method of SC learned by the networks was found to be in agreement with expectations: the attenuation of high-energy (EM) deposits and the enhancement of low-energy deposits. However, the network method was found to apply SC differently depending on the stage of the shower development, both in space and in time, the latter of which was found to agree with expectations of a bi-exponential time distribution for energy deposits in a steel-scintillator calorimeter expected from [32]. Additionally, the relationship between mean compensated active and original cell energy indicated similar energy regions for different energy depositions proposed in [171], without prior knowledge, and that therefore the SC model is capable of learning energy regimes similar to track-like, hadronic and electromagnetic energy depositions directly from experimental data. The model was also observed to attenuate active cell energies above 5 MIP and enhance them below that threshold. This result agrees with the hypothesis that the model has learned the gain-switching mode of the SiPM readout of the AHCAL, again without prior knowledge. Moreover, the model was observed to be capable of learning MIP-track subtraction and leakage correction in both simulation and data. Combined, these results strongly suggest that both the physical properties of hadron showers and the AHCAL detector can be learned by the network and contribute to superior hadron shower resolution.

The model was observed to produce different behaviour if trained on simulation and data, which could be explained by comparing these differences to the observations made on the difference in energy density between them. It was noted that the model trained on simulation could be applied to data, but resulted in a proportionally lower average response. This effect was observed not to arise from energy biasing and was most

likely due to the differences in energy density between hadron showers in simulation and data.

In summary, this study, therefore, indicates that superior resolution can be obtained in highly granular calorimeters using spatio-temporal event information and neural networks and that careful model design can be used to overcome the limitations of previous data-driven compensation techniques by reducing energy biasing.





# 5D Shower Separation with Machine Learning

# 6

## Contributions By The Author

*The author first processed, simulated, and analysed both simulation and experimental data presented in this chapter. Specifically, a method of producing synthetic charged-neutral hadron shower events from charged hadron shower events was thoroughly examined and validated. Appropriate shower separation networks were selected from literature and rewritten in PyTorch to support the inclusion of timing and track information and to follow a similar data-flow to Pandora PFA. The author then analysed the results, focusing on assessing the improvements in clustering achievable by a temporally sensitive calorimeter.*

6.1 Overview of Study . . . . .	165
6.2 Methods and Tools . . . . .	167
Network Architectures . . . . .	167
Raw Datasets . . . . .	175
Data Synthesis Techniques . . . . .	177
Synthetic Datasets and Training . . . . .	195
6.3 Results . . . . .	205
Simulation Models . . . . .	207
Data Model . . . . .	237
6.4 Conclusion . . . . .	252

## 6.1 Overview of Study

For future linear colliders to achieve the challenging jet energy resolutions required of around 3 % in the range of jet energies from 50-200 GeV, highly granular calorimeters must be exploited by accurate energy clustering algorithms as part of PF. Pandora PFA is an example of a clustering algorithm for resolving the energy deposits of particles.

It has been demonstrated in [55] that the main contributing factor to jet energy resolution for jet energies greater than 50 GeV using Pandora PFA is ‘confusion’ between the energy deposits of particles, as discussed in Section 2.4.5. Therefore, it is of scientific interest to reduce confusion between particles by developing methods to cluster hadron shower energy deposits more accurately by optimally exploiting the calorimeter geometry and the rich energy density information available in highly granular calorimeters.

Machine learning models provide a powerful tool for developing a bespoke shower separation algorithm using event information. As discussed in Section 2.4.5, PF relies on highly granular calorimeters and sophisticated clustering and pattern recognition algorithms. However, traditional convolutional neural networks (CNNs) assume a dense 2D or 3D ‘energy image’ matrix of voxels as an input. Typically, hadron shower events activate no more than a small fraction of the total number of sensors in the detector (e.g. the most probable number of sensors activated during a simulated 80 GeV hadron shower comprises only 2 % of the total sensors in the AHCAL calorimeter), resulting in relatively inefficient models for highly granular calorimeters due to the sparsity of relevant information in the input. It is also noted that a 5D input (e.g. the inclusion of timing information) results in even more sparsity due to the curse of dimensionality<sup>1</sup> Furthermore, these models assume a regular grid of sensors, which is problematic for clustering for detectors where the geometry is more complex, such as the transition region between barrel or endcap calorimeters.

1: The ‘curse of dimensionality’ is a phenomenon in high-dimensional analysis of data, where the sparsity of data in the measured dimensions typically increases because the increased volume of the space that a measurement may occupy [172].

2: The study of [120] includes a calorimeter of differing sensor sizes. In principle, this also applies to the AHCAL, with the ganged Tokyo layer in the 39<sup>th</sup> layer. However, due to the requirements that the performance of the algorithm be tested on experimental data, a method of producing synthetic hadron shower events is required and discussed in Section 6.2.3.3, which requires displacing the event in the calorimeter space. In this case, the energy deposited in the ganged layer becomes ambiguous and, therefore, cannot be considered in this study.

A study performed in [120] demonstrated improved performance of graph neural networks over CNNs applied to shower separation in calorimetry. The model presented in this study gains by being designed to learn an optimal representation of detector geometry, exploiting a graph neural network. The study, however, involved a calorimeter simulation with a greatly lower average sensor density than AHCAL (a total of 895 sensors, compared to the 21,888 sensors of AHCAL, in a calorimeter of approximately half its volume). Thus, it is unknown how such models perform with the increased granularity of AHCAL. This study also includes only the basic 4D calorimeter information: the total energy deposited in each cell, the cell position and layer number<sup>2</sup>. In a future  $e^+e^-$  collider experiment utilising PF, two additional sources of information are available: tracking information for charged particles (the track position and energy of the impinging particle in a magnetic field) and, for AHCAL, timing information. Both of these sources of information are expected to influence clustering performance:

- ▶ Track information provides an advantage to shower separation because a strong correlation exists between the centre-of-gravity of a charged hadron shower and the position of the charged track, which can be used to infer with high accuracy the shower axis of the charged hadron shower;
- ▶ It is unknown whether or not timing information available from AHCAL benefits the task of shower separation and by how much. An example of how timing information might aid event reconstruction is that the temporal distribution of late-showering hadrons will be delayed by the time of flight in the calorimeter. Timing information is expected to provide a richer description of the hadron shower. For example, sub-showers ought to be correlated in time relative to the hadron shower core, providing an additional degree of freedom that can be used to determine whether a fragment was part of a specific shower, even in cases where spatial resolution is insufficient to do so.

This study evaluated the performance of three published neural network models for hadron shower separation in the AHCAL highly granular calorimeter between a charged and synthetic neutral hadron shower produced using data synthesis techniques. The performance of the algorithms is tested using both simulation and experimental data. Furthermore, the software compensation algorithm described in Chapter 5 is applied to the separated shower events in simulation, and the results are studied.

As an important caveat, this study focuses on the effectiveness of clustering energy deposits from a charged and neutral hadron shower with the AHCAL detector. It neither assesses the effectiveness of the track-cluster association required to 'label' individual energy deposits as either charged or neutral nor the effectiveness of determining the number of simultaneous hadron showers in an event. Results should be interpreted with these caveats in mind. This discussion is considered in more depth in Section 6.2.4.2.

In this chapter, superscripts  $Q$  and  $N$  indicate variables associated with charged and neutral hadrons or showers. Furthermore, the same notation

as in Chapter 5 is used for predicted shower energy,  $\widehat{E}_{\text{sum}}$ . For instance,  $E_{\text{sum}}^N$  and  $\widehat{E}_{\text{sum}}^N$  denote the neutral energy measured by the calorimeter and the value reconstructed by the neural networks. Additionally, the predicted and true fractions of energy belonging to an event in a particular active cell of the calorimeter are denoted  $f_{\text{hit}}$  and  $\widehat{f}_{\text{hit}}$ , respectively.

## 6.2 Methods and Tools

In this section, the methods and tools used to experiment are introduced. First, the raw data from the simulation and CALICE 2018 Testbeam are introduced. Then, data synthesis techniques are introduced to produce the two-shower events.

### 6.2.1 Network Architectures

#### 6.2.1.1 Overview

The AHCAL is a five-dimensional calorimeter. As a form of data, the hadron shower event may be described adequately as a *'point cloud'*, or a set of points with Cartesian coordinates in space, time and energy.

In the study of this chapter, the aim is to produce a fraction of energy for each active cell belonging to one of two hadron showers, one charged and one neutral. This task is analogous to *'semantic segmentation'*, where each point is labelled with a class. Therefore, machine learning models with a demonstrable capacity to segment complex data with an underlying structure based on the geometrical distribution of the points, such as 3D models or scenes, are appropriate to apply to shower separation.

Therefore, three published network models suited to the task of hadron shower separation in highly granular calorimeters were studied in this analysis: PointNet [173], Dynamic Graph Convolutional Neural Network (DGCNN) [163], and GravNet [120]. Of these networks, only GravNet is specifically designed for the study of this chapter.

In this chapter, the same notation is used as in Chapter 5.

#### 6.2.1.2 Summary of Models

This section provides a summary of the models used to assess the performance of hadron shower separation and summaries of the specific implementations and modifications to the models to support the inclusion of timing and track information. First, modifications to vanilla models<sup>3</sup> are discussed. Then, specific implementations of each model are reviewed in brief.

3: 'Vanilla software' is a common term for software that has not been expressly customised [174]. It refers to the most common flavour of ice-cream [175].

**Modifications to Vanilla Models** Changes based on the requirements of the studied models are reasoned and described. These changes are relatively minor and discussed as follows.

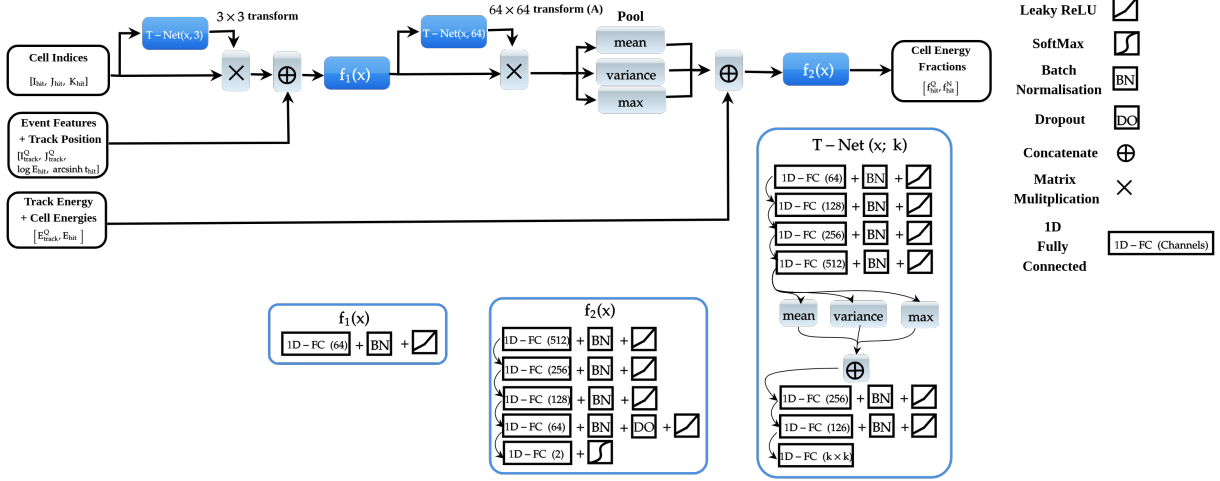
Firstly, AHCAL offers timing and track information from the delay wire chamber (see Section 2.5.1.1) which was not used in the shower separation study of [120]. Therefore, it is ambiguous how this information should be added to the network and in what order. For this reason, the information flow into each network is staggered to reduce the complexity of the high-dimensional problem and encourage the models to learn inspired by the successful template of Pandora PFA. The steps relevant to hadron shower clustering are highlighted, and the staggering is explained in the context of each relevant step of the algorithm [55]:

- ▶ After treating tracks for ‘kinks’ and potential decays of neutral particles, Pandora PFA first selects isolated active sensors, excludes them from the initial clustering stage, and aggregates information related to the cell geometry and hit distribution in the calorimeter sensor array. This step requires only spatial information. Therefore, the first operation of all networks uses only the cell indices of the active cells of an AHCAL event to first build a spatial representation of a hadron shower;
- ▶ Pandora PFA then applies a cone-based clustering algorithm starting from the track projection onto the calorimeter’s front face, from the innermost to the outermost layer. Then, topological clustering based on templates of energy depositing processes (i.e. pattern recognition, such as MIP tracks, which have been demonstrated in Chapter 5 to be identified by neural networks, illustrated in Figures 5.26a and 5.26a). For this reason, the shower ‘feature’ information (active cell energy/time stamp information, track position) is fed into the network later than the spatial information, where the main bulk of clustering and pattern recognition takes place.
- ▶ After topological clustering, a statistical clustering step is performed to inform re-clustering by aggregating the energy of a cluster ( $E_{\text{sum}}^Q$ ) and associating it to a charged track energy ( $E_{\text{track}}^Q$ ). This step is performed to improve performance for hadron showers from jets with  $E_j > 50 \text{ GeV}$ , where it is expected that there will be significant confusion between hadron showers. Therefore, in the final stages of each model, track energy and unmodified hit energy information are added such that the algorithm can correct any clustering mistakes using this information.

4: Aggregation functions summarise global properties of a data set. The maximum is typically used.

5: Variance is used instead of standard deviation for two reasons. Firstly, the square root function adds no additional information to the aggregation and is thus unnecessary. Secondly, in the special case that the variance is close to 0, which can happen if the input tensor is padded with a constant (i.e. the nearest neighbours to the padded value are also the padded value), then the gradient of the standard deviation is not defined in PyTorch.

Secondly, the study of [120] observed that the choice of aggregation functions<sup>4</sup> played a critical role in the convergence of the model. In particular, aggregation functions that included the properties of the entire set of vertices (i.e. active cells in the event) improved convergence. For this reason, in addition to the maximum (used in the basic implementations of all studied networks), the mean and the variance were also included<sup>5</sup>. Since this choice leads to three times as many channels, additional fully connected layers were added in certain cases to downsample this additional information where necessary.



**Figure 6.1:** Diagram illustrating the PointNet implementation in this study. As in Figure 5.5. Additionally, matrix multiplication is indicated as a  $\times$  symbol, batch normalisation is denoted 'BN' and the Softmax activation is indicated in the legend.

Lastly, each model was designed to have around  $2 \times 10^6$  weights, though this varied slightly from model to model depending upon the particular hyperparameters. This is a factor of 20 more free parameters than in [120]. However, around 20 times the number of sensors are in AHCAL as was used in that study (see Section 6.1). This choice preserves the proportion of around 90-100 weights per sensor used in the reference.

Other differences specific to each network are mentioned explicitly in the following network summaries.

**PointNet** This neural network architecture is designed for learning on generic point clouds [173]. The main feature of this model is invariance to the order of the input, which is achieved through the multiplication of a transformation matrix and the application of symmetric aggregation functions<sup>6</sup> to the input points. This is chosen in preference to ordering the high-dimensional points, as, in general, there does not exist a stable ordering of the set of points in the presence of point perturbations, meaning sorting is not a viable option for unordered points. By contrast, multiplying the input by an affine<sup>7</sup> transformation matrix determined from a multi-layer perceptron and activating the output, a general function can be approximated. The model also supports higher-dimensional features (e.g. energy, time, track position), necessary to describe a hadron shower event fully.

PointNet is a robust baseline neural network for the segmentation of 3D point clouds and is included in the study for that reason. Additionally, the state-of-the-art shower separation typically employs graph neural networks, and the application of PointNet allows examination of the advantages of more sophisticated neural network models. Application of this network model has been found in a physics setting, for instance, as a geometrically-sensitive partial differential equation solver for fluid dynamics [177]. It should be noted that this method is not, however, a graph neural network since no graph is constructed before or during the application of the model.

6: Here, symmetric refers to a specific type of operation that does not depend on the order of the input. For this study, the maximum, mean and variance are symmetric functions. For example, the maximum, mean and variance of the vector  $[1, 2, 3]$  and  $[3, 2, 1]$  are 3, 2 and 1, respectively. Put another way, the value of the aggregation function is independent of the permutations of the input vector.

7: An 'affine' transformation means that the transformation preserves co-linearity of points (i.e. a minimum of three points on the same line before transformation are also on a line after the transformation), the parallelism (i.e. lines that are parallel before the transformation remain that way after the transformation), convexity (i.e. the extreme points before transformation are also the extreme points after transformation) and the ratios of the lengths of parallel line segments. An affine transformation has a linear algebra representation of  $y(x) = A \cdot x + b$ , where  $y$  is some transformed vector,  $x$  is an input vector,  $A$  is a transformation matrix with shape  $\dim(x) \times \dim(x)$ , and  $b$  is a translation vector with shape  $\dim(x)$ . Rotations and mirroring are examples of affine transformations [176].



The specific implementation used in this study was based on an implementation of PointNet in PyTorch [178] and is illustrated in Figure 6.1. It is described as follows:

- ▶ *3D Transform*: The active cell indices of the event (i.e.  $(I_{\text{hit}}, J_{\text{hit}}, K_{\text{hit}})$ ) are passed through a ‘transformation network’ (T-Net), which consists of two parts: an upsampling module of four sequential 1D fully-connected layers with 64, 128, 256 and 512 channels, which is activated upon using the mean, variance and maximum over the channels, and a downsampling module of a further three sequential 1D fully-connected layers of 256, 128 and 9 ( $3 \times 3$ ) channels, with batch normalisation and leaky ReLU activation for all layers except the last layer. The output is a matrix, initialised as the identity matrix, multiplied by the input of this step. More details of the PointNet Encoder can be found in [173].
- ▶ *Upscaling and Concatenation With Features*: The additional ‘features’ of the hadron shower relevant to clustering ( $\log E_{\text{hit}}, \text{arcsinh } t_{\text{hit}}, I_{\text{track}}, J_{\text{track}}$ ) are concatenated to the output and upsampled by a 1D fully-connected layers with 64 channels;
- ▶ *Feature Transform*: Similarly to the 3D Transform, the output is passed through a second T-Net of the same structure as the 3D Transform with a  $64 \times 64$  layers. The output is a second matrix, also initialised as the identity matrix and defined  $A$  in this section, and multiplied by the input of this step as is the case for the *3D Transform* step.
- ▶ *Aggregation*: The mean, variance and maximum over the points are used to activate over the points and the remaining active cell energy and track energy information  $[E_{\text{hit}}, E_{\text{particle}}^Q]$  are concatenated to the output.
- ▶ *Downscaling and Output*: The final module of the network consists of four sequential 1D fully-connected layers with 512, 256, 128 and 64 channels, with batch normalisation and leaky ReLU activation, and dropout in the last layer of the sequence. The final layer has two outputs, one for  $\hat{f}_{\text{hit}}^Q$  and one for  $\hat{f}_{\text{hit}}^N$ , with Softmax activation. The result is a set of reconstructed energy fractions for each active cell in the event.

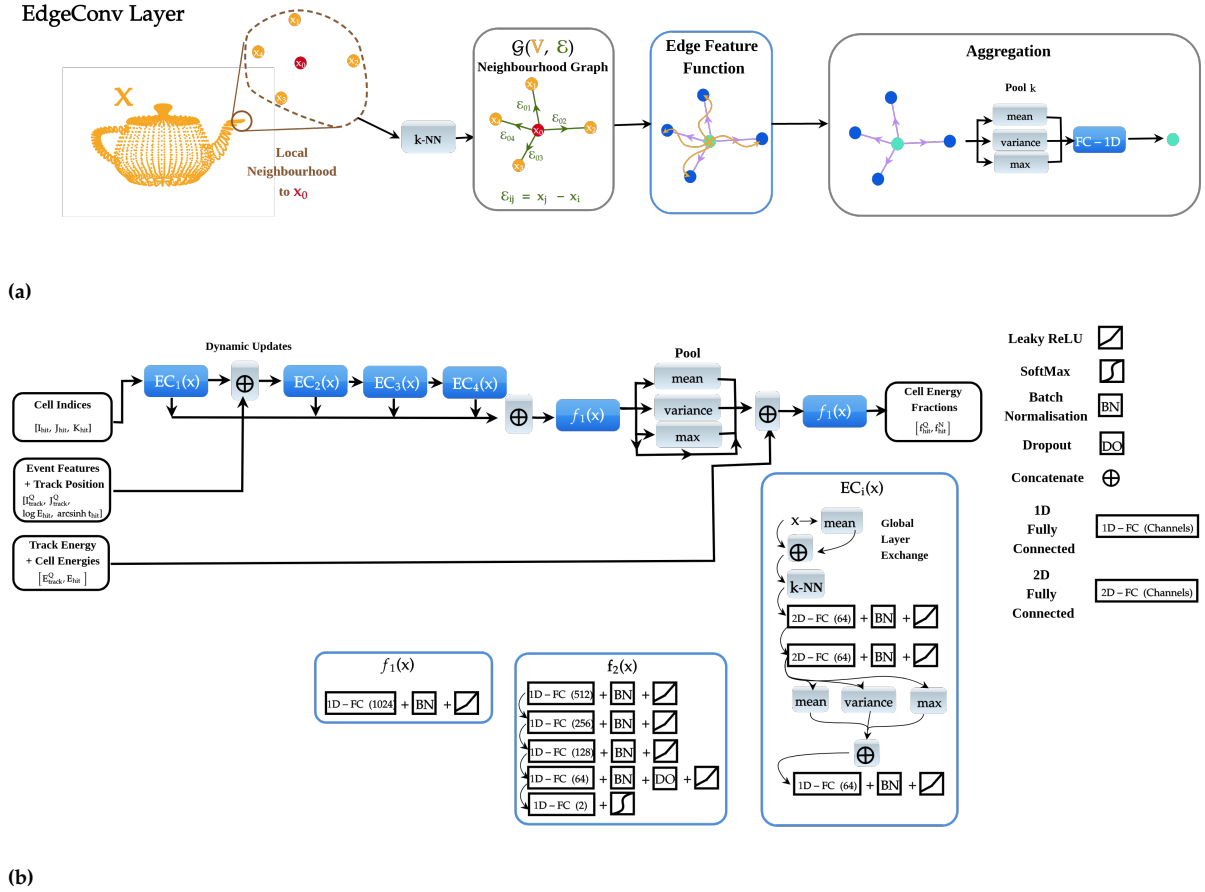


**Figure 6.2:** Qualitative performance of PointNet on 3D scene segmentation using the S3DIS dataset [179]. The top plot shows the input point cloud, representing an office scene in 3D. The bottom plot shows the PointNet model, trained to classify each point as one of 13 categories. Taken from [173].

Importantly, the  $A$  matrix is constrained to be close to an orthogonal matrix by including a regularisation condition in the loss as in Equation 6.1 [173]:

$$\mathcal{L}_{\text{reg}}(A) = \|I - AA^T\|^2 \quad (6.1)$$

where  $I$  is the identity matrix, superscript  $T$  indicates the matrix transpose operation. This is because orthogonal matrices produce a flip or rotation and preserve the lengths of vectors and angles between vectors, meaning the transformation is affine as required by the model [176].

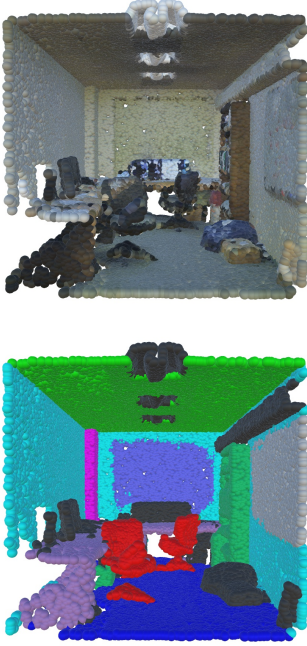


**Figure 6.3:** Figure 6.3a illustrates the EdgeConv operator. The orange Utah teapot [180] indicates some ‘point cloud’, or distribution of points,  $x$ , with some underlying distribution, to be operated upon. The orange dots indicate vertices,  $V$ , of a local-neighbourhood  $k$ -NN graph around the central red dot. The green arrows indicate the vectors (edges) between the central red dot and its neighbours,  $\mathcal{E}$ . The orange arrows indicate ‘message-passing’ between the vertices and the edges as discussed in Figure 2.54c, which modifies the graph as indicated by the colour inversion. Figure 6.3b shows a diagram illustrating the DGCNN implementation in this study. Else, as in Figure 6.1

**DGCNN** This graph neural network model is also designed for learning on generic point clouds [163]. DGCNN was designed as an improvement on PointNet. Specifically, PointNet operates on vertices (i.e. calorimeter energy deposits) individually in the T-Net and does not encode information about the model’s local energy density in space and time. By contrast, DGCNN encodes local energy density information by calculating the  $k$ -nearest neighbours ( $k$ -NN) denoted ‘vertex’ features ( $\mathcal{V}$ ) and the vectors between these nearest neighbours in the calorimeter space or ‘edge’ features, ( $\mathcal{E}$ ). A self-loop is included. A special operation called ‘EdgeConv’ is then applied, which consists of convolutional layers with activation followed by a symmetric pooling operation. The ‘dynamic’ aspect of the network is that finding neighbours and applying convolutions and pooling are repeated sequentially, resulting in a ‘dynamically updated’ graph. In the context of PF, the model can learn properties of the calorimeter readout at a local and calorimeter-wide level by diffusing local information (i.e. the energy density of the  $k$ -NN) across the graph by successive re-clustering. The model also supports higher-dimensional features (e.g. energy, time, track position), necessary for a full description of a hadron shower event.

Application of this network model has also been found in a physics setting. In addition to the hadron shower separation task of [120], the

8: Top-quark tagging means identifying a jet based on whether or not it originated from a top quark. Scientific interest in the top quark is discussed in Section 2.1.1.



**Figure 6.4:** Qualitative performance of DGCNN on 3D scene segmentation using the S3DIS dataset [179]. The top plot shows the input point cloud, representing an office scene in 3D. The bottom plot shows the DGCNN model, trained to classify each point as one of 13 categories. Taken from [163].

9: Shortcut, or ‘skip’ connections are typically used in deep neural networks that enable deeper layers in the neural network to learn to do nothing at all, which can improve performance when the complexity of the problem is simpler than the model architecture allows for. This concept is known as the ‘degradation problem’ [184]. They may also be used to capture information at varying scales, which may improve performance by enabling comparison between those scales [163].

template of DGCNN and EdgeConv has been demonstrated to be used for jet-tagging. In [181], state-of-the-art top-quark tagging<sup>8</sup> performance using jets reconstructed via Particle Flow in ATLAS [182] with a specially modified version of the same model.

The specific implementation used in this study was based on the implementation of [183]. EdgeConv and the DGCNN model used are illustrated in Figure 6.3a and 6.3b, respectively.

- ▶ **EdgeConv:** The EdgeConv operation is briefly explained. This function is designed to perform a convolution on a graph. The idea is that the graph’s origin is like the central pixel of an image, and the neighbours to that pixel are like a surrounding ‘patch’ of the image [163]. The operation consists of three stages, illustrated in Figure 6.3a. Firstly, a cluster of points is found using a  $k$ -NN operation, and the vectors between the vertices (edges) are found. Then, the vertices and edges are jointly operated on by an ‘edge feature function’, which is, in this case, a series of standard fully-connected convolutional layers. Then, the graph is aggregated. Additional aggregation functions are used in this study, and therefore an additional fully-connected convolutional layer is used to condense the aggregates to a single value. More information on this operation can be found in [163];
- ▶ **Clustering:** Four EdgeConv operators are defined. In each operation, the input is concatenated with its mean. Then, the  $k$ -NN clustering is applied, and the data is passed through two fully-connected 2D layers with 64 channels. The mean, variance and max are calculated over the clusters and passed through a 1D fully-connected layer with 64 channels with batch normalisation and leaky ReLU activation for all layers. The first EdgeConv operator uses only active cell indices of the event,  $(I_{\text{hit}}, J_{\text{hit}}, K_{\text{hit}})$ . The additional ‘features’ of the hadron shower that are relevant to clustering  $(\log E_{\text{hit}}, \text{arcsinh } t_{\text{hit}}, I_{\text{track}}, J_{\text{track}})$  are concatenated to the output are later concatenated to the output of the first EdgeConv operator before the second operator is applied. At each stage, the output from each EdgeConv operator is recorded and later concatenated into a single tensor. These are called ‘*shortcut connections*’<sup>9</sup>.
- ▶ **Embedding:** One shared 1D fully-connected layer with 1024 channels with batch normalisation and leaky ReLU activation condenses the features learned at the clustering stage;
- ▶ **Aggregation:** The mean, variance and maximum over the points are used to activate over the points and the remaining active cell energy and track energy information  $(E_{\text{hit}}, E_{\text{particle}}^Q)$  are concatenated to the output.
- ▶ **Downscaling and Output:** The final module of the network consists of four sequential 1D fully-connected layers with 512, 256, 128 and 64 channels, with batch normalisation and leaky ReLU activation, and dropout in the last layer of the sequence. The final layer has two outputs, one for  $\hat{f}_{\text{hit}}^Q$  and one for  $\hat{f}_{\text{hit}}^N$ , with Softmax activation. The result is a set of reconstructed energy fractions for each active cell in the event.

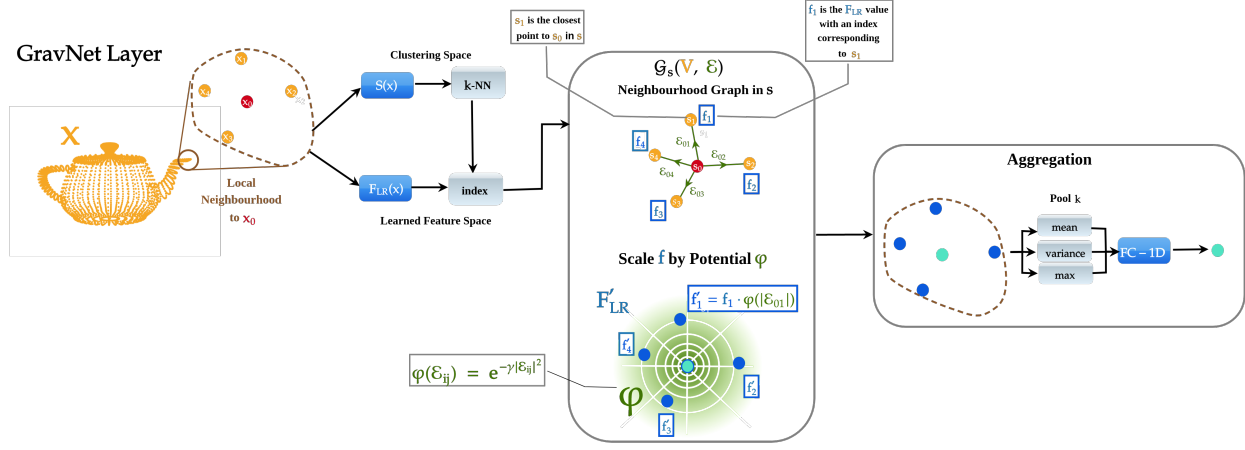


**GravNet** This graph neural network model exploits dynamic graph updates similarly to DGCNN. It is specifically designed for reconstructing hadron showers in highly granular calorimeters.

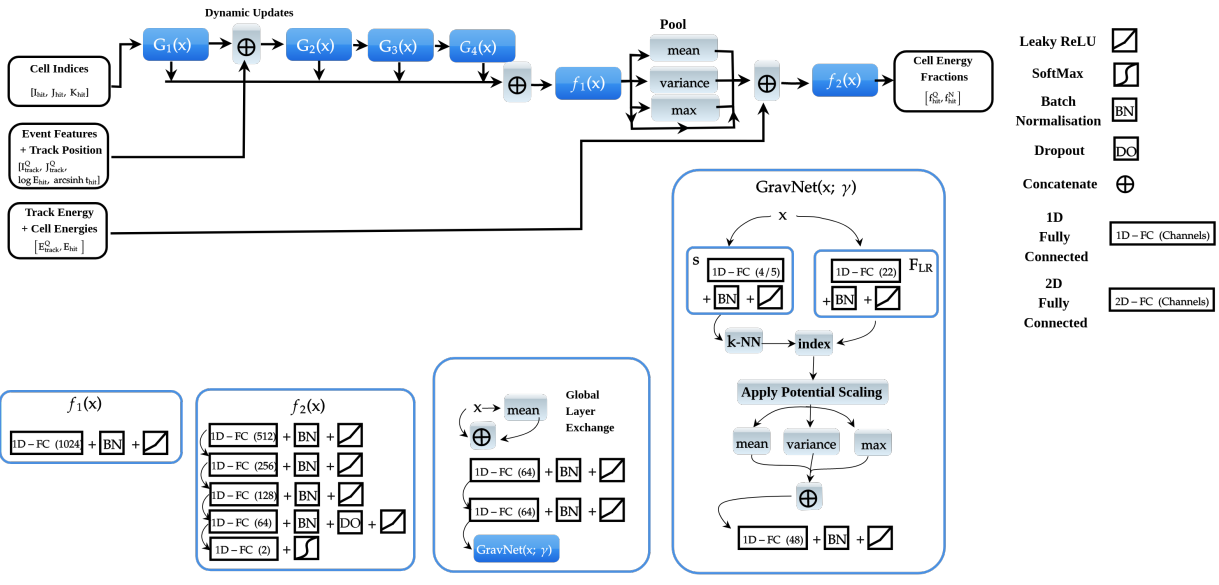
DGCNN is powerful but computationally intensive due to dynamically updating the graph at each stage during clustering in high-dimensional space. GravNet overcomes this limitation by learning two embeddings: a low-dimensional clustering space and a high-dimensional feature space. The distance between the points in the clustering space is used to assign attention to the features employing a ‘potential well’, from which GravNet gets its name. The learned features are scaled by the potential, which indirectly performs message-passing between the edges and the vertices. This method is advantageous in computation because the dimensions of the  $k$ -NN calculation are greatly reduced. Additionally, GravNet offers some control over the strength of the potential well by allowing its width to be controlled by a hyperparameter,  $\gamma$ , which was not explored in the reference<sup>10</sup>.

GravNet was designed for the application studied in this chapter. It is demonstrated in [120] to produce a similar performance to DGCNN when applied to the task of hadron shower separation in a segmented tungsten calorimeter.

10: In the original implementation of [185], it was noted that the  $\gamma$  factor was not modified. It is noted that a similar procedure is applied in a dimensionality reduction technique called *kernel principal component analysis*, and for similar reasons, in ‘*support vector machines*’, another machine learning classification method. Both methods call the ‘potential well’ a ‘*Gaussian radial basis function*’, which acts to map a lower dimensional space to a higher one with more dimensions by combining the former in non-linear combinations [186]. In the quoted applications, performance is nearly always improved if  $\gamma$  is treated as a hyperparameter to be tuned. For this reason,  $\gamma$  is tuned in this study.



(a)



(b)

**Figure 6.5:** Figure 6.5a illustrates the GravNet operator. The orange Utah teapot [180] indicates some ‘point cloud’, or distribution of points,  $x$ , with some underlying distribution, to be operated upon. The  $S$  and  $F_{LR}$  operations indicate the clustering and feature space representations learned by the neural network. The green arrows and  $\mathcal{E}$  correspond to vectors between the red point and its orange neighbours, its norm indicating a distance in a potential well given by  $\varphi$ , indicated by the shaded green region.  $f$  and  $f'$  indicate unscaled and potential-scaled features. Else as in Figure 6.3a. Figure 6.3b shows a diagram illustrating the DGCNN implementation in this study. Else, as in Figure 6.1.

The GravNet model was based on the implementation of [185]. GravNet and specifically the GravNet layer that substitutes the EdgeConv are illustrated in Figure 6.5a and Figure 6.5b, respectively. The model has the same underlying structure as the DGCNN model, except the GravNet layer and the structure of the clustering operator replacing EdgeConv.

- *GravNet Layer*: The GravNet layer operation is briefly explained. This function projects points to a low-dimensional clustering space,  $s$ , and high-dimensional learned feature space,  $F_{LR}$ , respectively. This is achieved with two individual 1D fully-connected layers with 4 (5 if time is included) and 22 channels, respectively. Then, the  $k$ -NN cluster is found in the  $s$ -space, similar to DGCNN. At this stage, the Euclidean distances of the neighbours to the origin of the graph (including a self-loop, as in DGCNN) are calculated for each cluster, denoted  $|\mathcal{E}_{ij}|$ , where  $\mathcal{E}_{ij}$  is the vector from the origin of the graph at vertex  $i$  to a neighbour at vertex  $j$ <sup>11</sup>. The elements of the  $F_{LR}$  space are then scaled by a Gaussian potential,  $\varphi(|\mathcal{E}_{ij}|) = e^{-\gamma|\mathcal{E}_{ij}|^2}$ , where  $\gamma$  is a hyperparameter to be tuned. For example, if  $f$  is a vector of features in  $F_{LR}$ ,  $f'_j = \varphi(|\mathcal{E}_{ij}|) \cdot f_j$ . The result is aggregated across the cluster using the maximum, mean and variance. Then, a 1D fully-connected layer with 48 channels, batch normalisation and leaky ReLU activation as an output layer to condense the aggregates to a single value. More information on this operation can be found in [120].
- *Clustering*: As for DGCNN, except that before each GravNet layer, the mean is concatenated to the input and passed through two 1D fully-connected layers with batch normalisation and leaky ReLU activation with 64 channels.

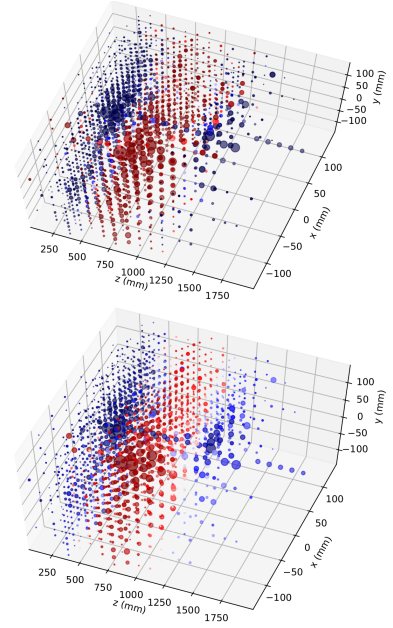
The remaining steps are as for DGCNN.

## 6.2.2 Raw Datasets

The raw simulation and data used for the study presented in this chapter are introduced, analysed and studied.

The studied models for shower separation discussed in Section 6.2.1 were trained and validated using simulated  $\pi^-$  and  $K_L^0$  hadron shower events observed with the AHCAL detector. The simulation of the particle showers was achieved using Geant4 [83], with a full detector simulation developed using DD4hep [84]. Additional effects, such as digitisation of the analogue signal and reconstruction of the detector variables, were achieved using CALICESoft [86]. The physics list used was QGSB\_BERT\_HP. The simulation was based on the June CALICE testbeam study taken at the Super Proton Synchrotron at CERN in 2018 [60]. Data used in the study was taken from the same experiment, where the detector variables were reconstructed using CALICESoft.

**Selection Criteria** The sample was subject to the following selection criteria. Data and simulation were subject to the same cuts as in Chapter 5 except where specified:



**Figure 6.6:** Qualitative performance of GravNet on hadron shower separation in a 4D calorimeter composed of tungsten for two charged pions with an energy of around 50 GeV. The colours indicate the energy fraction belonging to each hadron shower. Top: Ground Truth Event. Bottom: Reconstructed Event. Taken from [120].

<sup>11</sup>: In reality, the  $k$ -NN clustering step ranks distances based on their ‘top  $k$  negative square distance’, using the equation  $-|\mathcal{E}_{ij}|^2 = -|x_i|^2 - |x_j|^2 + 2\langle x_i, x_j \rangle$ , where  $x_i$  and  $x_j$  are two vertices to be compared, and  $\langle x_i, x_j \rangle$  is the inner product between  $x_i$  and  $x_j$  and  $|\mathcal{E}_{ij}|^2$  is the square Euclidean distance between  $x_i$  and  $x_j$ . This is mentioned because taking the square root only to square the result again would be redundant.

- ▶ events were required to be identified using the standard CALICE particle identification algorithm [79] as being a single particle and having less than a 0.5% probability of being a muon to exclude non-showering, ‘punch-through’ pions;
- ▶ the 38<sup>th</sup> layer of the AHCAL is ganged and requires special treatment beyond the scope of this paper. Therefore, energy deposits were considered up to the 38<sup>th</sup> layer of the calorimeter;
- ▶ events with a correctly reconstructed track position (i.e. a track position with a corresponding position inside the  $24 \times 24$  cell AHCAL front-face).
- ▶ events were required to have at least 50 active cells after the MIP-track cut discussed in the following section. This criterion reduces the influence of partially-showering punch-through pions, which may initiate a small cascade and continue through the calorimeter. Details of this cut are discussed in Section 6.2.3.

The event count of the final training, testing and validation datasets are summarised in Table 6.1.

Comparisons of the main event-level, cell-level and shower profile distributions of the training sample for the entire samples of 10 GeV, 40 GeV and 80 GeV common between the simulation and data training sets are presented in Appendix Figures 8.7-8.21, which have the same comments as Figures 5.7-5.21.

Comments to these plots are the same as in Chapter 5, and the figures are shown for consistency between analyses.

Hadron Type Purpose Particle Energy [GeV]	$K_L^0$	$\pi^-$			Simulation		
	Simulation Analysis	June 2018 Testing	SPS Testbeam Training	Data Validation	Simulation Testing	Training	Validation
5	-	-	-	-	40685	36966	4108
10	-	21396	171166	21396	68812	62333	6926
15	-	-	-	-	75224	67379	7487
20	-	32221	257762	32220	77759	70866	7874
25	-	-	-	-	81379	73023	8114
30	-	-	-	-	80971	74180	8243
35	-	-	-	-	78646	75619	8403
40	78146	34428	275424	34428	77055	76142	8461
45	-	-	-	-	73620	76994	8555
50	-	-	-	-	86014	77430	8604
55	-	-	-	-	63218	77881	8654
60	-	44600	356799	44600	72306	78550	8728
65	-	-	-	-	82256	78779	8754
70	-	-	-	-	59806	79042	8783
75	-	-	-	-	49390	79417	8825
80	-	37790	302315	37789	69106	79713	8858
85	-	-	-	-	70091	79848	8872
90	-	-	-	-	80773	79918	8880
95	-	-	-	-	62647	79614	8846
100	-	-	-	-	54433	79292	8811
105	-	-	-	-	53377	77300	8589
110	-	-	-	-	52632	77853	8651
115	-	-	-	-	58750	75091	8344
120	-	34881	279044	34880	56274	74835	8316
Total Events	78146	205316	1642510	205313	1625224	1788065	198686

**Table 6.1:** Table of events used for training shower separation models and performing analysis models after all cuts, split into simulation and data and by the testing, training and validation samples and by data and simulation. An additional sample of 40 GeV  $K_L^0$  hadrons simulated under the same conditions as the  $\pi^-$  hadrons is included for analysis. Hyphens indicate 0 events.

### 6.2.3 Data Synthesis Techniques

Data synthesis describes the creation of artificial examples of a source of information used to test mathematical models and to train artificial neural networks [187]. Monte Carlo simulation is ubiquitously used in particle physics for these purposes. For instance, the study in [55], which demonstrates the performance of Pandora PFA and its capacity to achieve the jet energy resolution required for future linear colliders, is based on simulation.

An example of data synthesis is Monte Carlo simulation, which can generate two-particle events, as presented in the study of [120]. However, a technique of producing synthetic neutral hadron showers and artificial two-shower events is required to validate shower separation methods on experimental data and simulation is required so that the performance of the networks can be evaluated from being trained on both simulation and data fairly. Specifically, data has shown to have non-negligible differences in energy density to simulation (see the comments to Figures 5.19-5.21 in Section 5.2.3.1). Additionally, there is no available neutral hadron shower data for AHCAL because these particles cannot be guided to the calorimeter by magnetic chicane. Therefore, the only way to assess the performance of shower separation techniques on experimental data is to synthesise showers produced by neutral hadrons from charged ones.

This section discusses the methods of data synthesis and studies validating these methods.

#### 6.2.3.1 Generation of Synthetic Neutral Events

On average, no differences in the physical processes involved in a hadron interacting with dense matter are expected between charged and neutral hadrons, except for minimum ionisation before a hard interaction. This assumption may be exploited to produce synthetic neutral hadrons.

An ad-hoc cut developed in [78] was applied to select MIP-tracks from hadron showers observed with AHCAL. This cut operates under the assumption that the MIP-track is expected to leave a rectilinear, highly-localised energy signature in the calorimeter, parallel to the axis of motion of the particle through the matter, before the shower starting layer.

The cut, therefore, consists of three conditions:

- ▶  $R_{\text{hit}} < 60 \text{ mm}$

If a hadron shower occurs, the MIP-track rectilinear energy signature is expected to occur close to the hadron shower core. This cut also allows for the possibility that the MIP-track may be shared between multiple 'towers' of cells if the particle is centred on the edges of the AHCAL cells. This is particularly relevant if there is a slight angle at which the particle enters the calorimeter;

- ▶  $E_{\text{hit}} < 3 \text{ MIP}$

The energy from ionising heavy particles is expected to be Landau-distributed. The choice of 3 MIP balances the efficiency of selecting

MIP-like active cells while reducing the number of EM-like active cells removed by the cut in the EM-core of the shower;

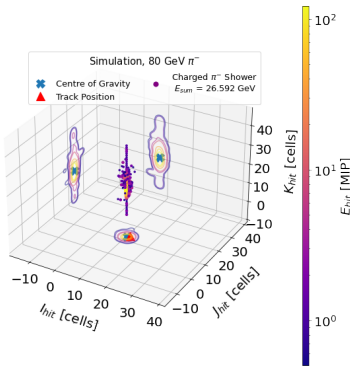
- $K \leq K_S - 2$  layers

The MIP-track is expected to occur before the first hard interaction by the hadron with matter. A study of the shower start algorithm in [78] indicated that there is a 90 % efficiency at finding the shower start within  $\pm 2$  layers. For this reason, the cut is chosen slightly before the selected shower start to reduce the likelihood of selecting hits from the EM-core.

As in [78], a synthetic neutral hadron can be produced by subtracting energy from the event that satisfies the selection criteria.

This assumption has clear limitations:

- *'Lost' Ionisation Energy*: the subtracted energy of the active cells identified in a hadron shower event would not have been deposited after the hard interaction in a true neutral shower. This means that energy that would have contributed to further generations of particles in the fake neutral hadron event is 'lost' in the pseudo-neutral event;
- *Partially-Showering Charged Hadrons*: There is a non-negligible probability that a  $\pi^-$  hadron will 'punch through' the AHCAL, ionising the passive medium only. In contrast, a neutral hadron of the same energy will not interact with the detector. A  $\pi^-$  hadron will shower within the AHCAL detector with cumulative probability  $P_{\text{shower}} = 1 - e^{-4.2} = 98.5\%$ , where  $4.2 \lambda_I$  is the longitudinal depth of AHCAL, and therefore the probability that punch-through pions are observed with AHCAL is considerable. While it has been demonstrated that these 'muon-like' events can be excluded with high-efficiency [79], a complication is the possibility of producing pseudo-neutral hadron showers from partially-showering charged hadrons, which can occur if there is a 'glancing blow' with an atomic nucleus. Attempting to generate synthetic neutral hadron showers from such 'partially showering' pions, the event will still contain the MIP-track of the 'punch through' hadron after the shower is initiated, which would not be detected for a neutral particle of the same energy. An example event display is provided in Figure 6.7;
- *Ad-hoc Cut*: Based on physical expectations, the cut presented is ad-hoc and has not been optimised.



**Figure 6.7:** An example 'partially showering' 80 GeV  $\pi^-$  hadron shower as observed with AHCAL. Each axis represents the spatial coordinates of the calorimeter, and the purple points indicate  $E_{\text{hit}}$  in a logarithmic scale. The lines on each calorimeter face indicate the 'energy shadow' of the hadron shower to indicate its profile. This example illustrates an example of an event where a hadron only partially showered and continued through the calorimeter, depositing only a fraction of its energy. Removal of the MIP-track of this event would not result in a convincing neutral event.

12: The offset of 3 layer presented in Figure arises because no energy is expected to be removed by the cut before  $K_S = 4$  (the third layer of AHCAL). Thus,  $K_S = 3$  layer is subtracted from the  $x$ -axis with the expectation that the curve crosses the  $y$ -axis at near zero.

It is then necessary to attempt to quantify the quality of the cut at producing pseudo-neutral hadrons, which is studied in this section.

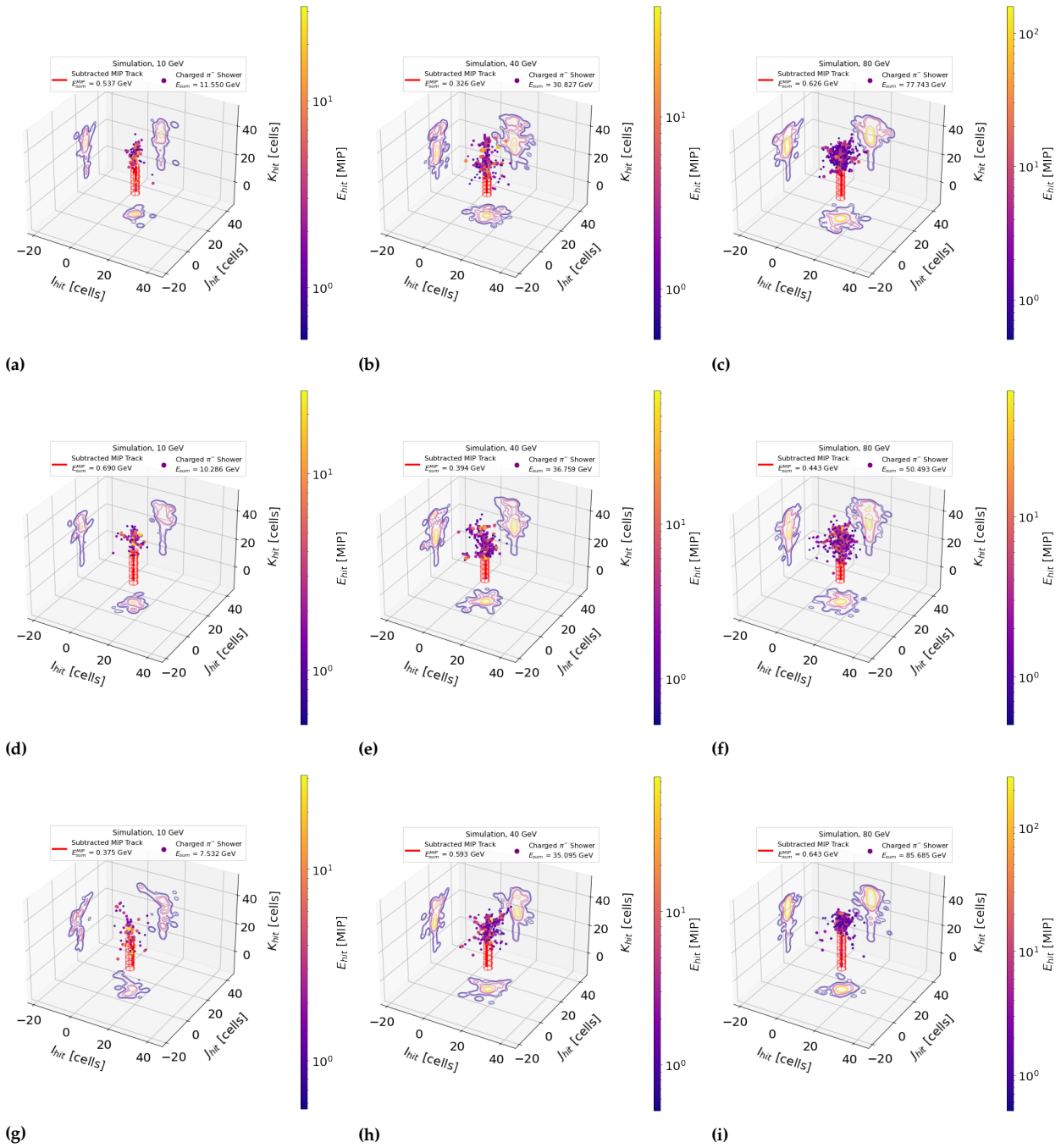
Firstly, the expectations from physics for the cut applied were analysed. Example cases of MIP-track selection according to the described cut is shown for simulation in Figure 6.8, which can be used to inspect the behaviour of the cut visually. Quantitative analyses were then performed to assess the performance of the MIP cut. The average total calorimeter response to the MIP-track according to the applied cut is studied as a function of the shower starting layer,  $K_S$ , presented with an offset

of 3 layer subtracted to aid in interpretation<sup>12</sup>, shown in Figure 6.9. The results were compared to expectations on ionising energy loss by relativistic heavy charged particles in the momentum range 10-100 GeV. The distributions of the active cell energy,  $E_{\text{hit}}$ , were measured and fitted with a Landau-Gaussian convolution distribution [188] in Figure 6.10, to measure the likelihood that the deposited energy was caused by ionisation.

#### **Summary Comment to Figures 6.8-6.10**

The relationship between the energy deposited and the number of active cells per layer agrees with physical expectations of ionisation energy loss by charged, heavy particles in Fe at the energy scale of the hadrons studied. The active cells selected by the MIP-track cut agree with a Landau-Gaussian distribution for the energy range between 0.5-2 MIP. Since the same bias is applied to simulation as in data and therefore does not affect the comparison of the results from simulation to data.

## Event Displays of Simulated Pion Showers observed with AHCAL Illustrating the MIP-Track Cut



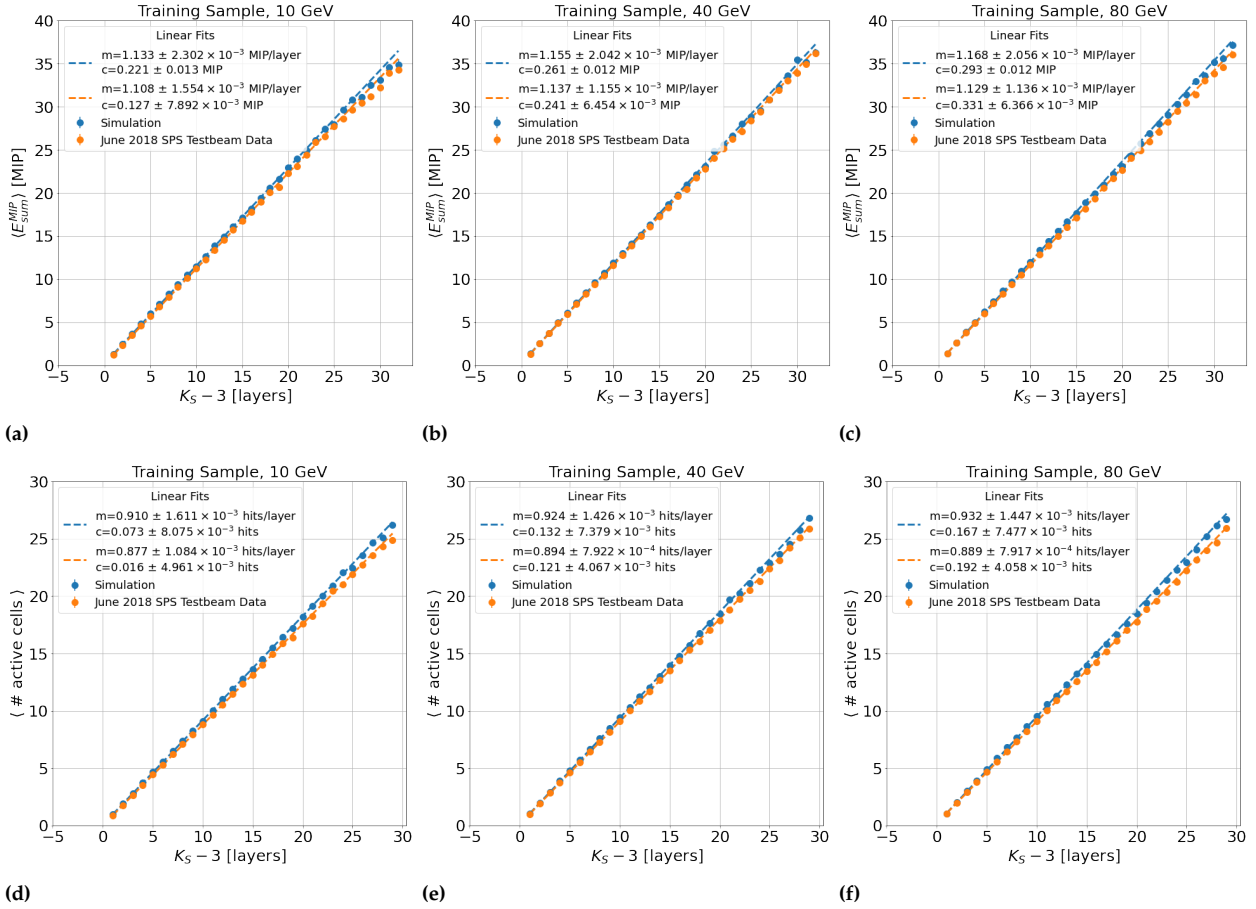
**Figure 6.8:** Event displays of  $\pi^-$  hadron showers demonstrating the MIP cut applied to simulated  $\pi^-$  hadrons of the training sample. Selected MIP-track hits are shown as red circles. The red cylinder indicates the cut region in space. The energy criterion cannot be shown. The leftmost, middle and rightmost column indicate examples of 10 GeV, 40 GeV and 80 GeV  $\pi^-$  hadron showers, respectively.

**Comments:**

- Inspection of the examples shown indicates that the cut is qualitatively effective at selecting the MIP-track. This is indicated by the expectation of the rectilinear MIP-track energy signature, before the shower start.



### Average Reconstructed Hadron Shower Energy and Number of Active Cells for Selected MIP-Tracks in Events used for Shower Separation Training

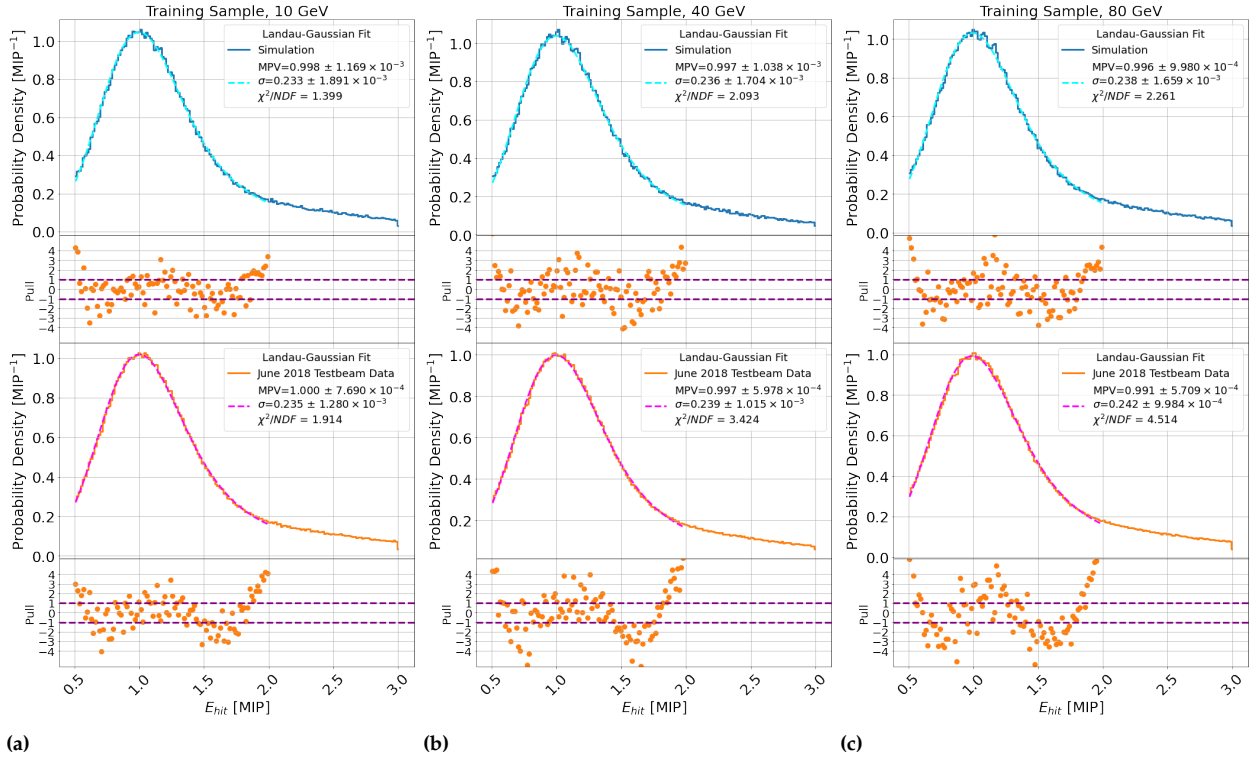


**Figure 6.9:** Figures 6.9a-6.9c show average total response to energy classified as a MIP-track in Section 6.2.3, as a function of the shower starting layer. The y-axis indicates the average energy deposited in an event in MIP. Figures 6.9d-6.9f show the average number of cells classified as a MIP-track in Section 6.2.3. The x-axis for both sets of figures indicates the shower starting depth in layer (starting from  $K_S = 3$  layer). The blue and orange circle markers indicate simulation and data, respectively. Dashed lines indicate linear fits, with values shown in the legend.  $m$  indicates the additional energy deposited per layer/ additional hits per layer traversed of AHCAL in MIP/layer or the number of hits/layer,  $c$  indicates the offset in MIP or hits, which should pass through 0. Results from the 10 GeV, 40 GeV and 80 GeV training samples are shown in the leftmost, middle and rightmost columns, respectively.

#### Comments:

- In Figures 6.9a-6.9c and Figures 6.9d-6.9f for both simulation and data, a clear linear relationship between the shower starting position and the average energy subtracted by the MIP-track cut in a given event is observed. This means that the change in the cumulative energy and number of active cells selected by the cut are approximately constant with the shower starting layer. A constant energy loss per layer at the relativistic momenta of the impinging particle is expected from the Bethe-Bloch formula shown in Equation 2.1, which indicates the cut is selecting ionising energy deposits.
- Figures 6.9a-6.9c indicate that the cumulative average energy loss selected by the cut is directly proportional to the shower starting position in both simulation and data. The average increase in energy deposited with the shower starting layer is observed to be slightly larger than 1 MIP/layer, which is expected due to the average of the Landau distribution describing ionising energy loss being larger than the most probable value (MPV) of the distribution (see Section 2.3.1 and Figure 6.10).
- Figures 6.9d-6.9f show that slightly less than 1 hit/layer is selected by the cut is observed, which agrees with a similar comparison performed in [78]. Furthermore, the constant terms for all fits are in agreement with the expectation of passing through the origin (i.e. no active cells are selected at or below  $K_S = 3$  layer). For Figures 6.9a-6.9c, the fit constant is less than the AHCAL noise threshold at 0.5 MIP and for Figures 6.9d-6.9f, it is well below 1 hit.
- The fitted slopes between simulation and data for average hit energy and average number of active cells are greater by only 1% and 3%, respectively, indicating the cut has a similar effect on both samples. The fitted intercept values are all close to zero active cells, indicating that no energy is selected by the cut for  $K_S \leq 3$ , as expected.

### Fits of Landau-Gauss Distribution to the Distributions of Active Cell Energy Selected by the MIP-Track Cut



**Figure 6.10:** Active cell energy ( $E_{hit}$ ) of hits passing the MIP-track selection criteria, for 10 GeV, 40 GeV and 80 GeV events from the training samples of simulation and data. The normalised histograms are shown in blue and orange for simulation and data, with fits of the Landau-Gaussian function up to 2 MIP shown as cyan and magenta dashed lines, respectively. Subplots indicate fit pulls. The MPV and  $\sigma$  of the Landau-Gaus distribution are shown in the legend.

#### Comments:

- ▶ For  $E_{hit} < 1.8$  MIP, good agreement is observed between both simulation and data with a convolved Landau-Gaus distribution. In addition, the MPV of each distribution is around 1 MIP. The MIP scale corresponds to the most probable value MPV of the ionisation energy deposited by 40 GeV muons from the May and June 2018 SPS Testbeam data, using a total of  $1.4 \times 10^7$  events [78].
- ▶ While the Landau-Gaussian convolution generally agrees with simulation and data where  $E_{hit} < 1.8$  MIP, clear deviations are observed beyond this range, as indicated by the deviation observed in the residuals at around  $1.8 < E_{hit} \leq 2$  MIP and beyond. This result indicates that the cut includes additional depositions not caused by ionisation. These cells are most likely part of the hadronic shower start, which can be included in the cut due to the  $\pm 2$  layer uncertainty on the shower starting position.

Hadron Type Purpose Particle Energy [GeV]	$K_L^0$ Simulation			$\pi^-$ Simulation		
	Testing	Training	Validation	Testing	Training	Validation
40	11732	54800	11614	11411	53201	11530

**Table 6.2:** Subsample of Table 6.1, used for training the gradient-boosted decision tree to assess the performance of the synthetic neutral hadron showers produced via the MIP-track cut.

The studies of Figures 6.8-6.10 provide evidence that the cut selects primarily ionising energy deposits of interacting charged hadrons, and therefore operates as intended. These studies, however, offer no conclusive statement as to the quality of the synthetic pseudo-neutral events if the MIP track is removed from the event.

For this reason, a separate study was performed to quantify this property. The properties of simulated  $K_L^0$  hadron showers were compared to the  $\pi^-$  hadron showers of the training dataset. A simple and illustrative method was devised, which may describe the method's performance in a single summary statistic for 40 GeV hadrons in simulation.

Firstly, the training sample of 40 GeV  $\pi^-$  and the entire  $K_L^0$  sample from simulation was selected from the available simulation shown in Table 6.1 and was itself split into a training, validation and test sample. The events are shown in Table 6.2.

Next, the standard CALICE AHCAL Particle Identification (PID) classifier is introduced. This is a gradient-boosted decision tree classifier implemented in the LightGBM framework [79, 86, 189]. This classifier has been demonstrated to utilise event-level information to accurately classify hadrons, electrons and muon-like events observed with AHCAL. The model uses the following thirteen event-level variables:  $E_{\text{sum}}$ , the total number of active cells in the event, the average hit radius ( $\langle R_{\text{hit}} \rangle$ ),  $\text{CoG}_K$ ; the fraction of  $E_{\text{sum}}$  deposited in the first 22 AHCAL Layers, the fraction of  $E_{\text{sum}}$  deposited after the shower start layer  $K_S$ , the fraction of  $E_{\text{sum}}$  in the 'core' of hadron shower (defined as satisfying  $R_{\text{hit}} < 1$  cell,  $\geq 2$  adjacent active cells and  $> 0$  cells active in the same layer)", the 'track' fraction of  $E_{\text{sum}}$ <sup>13</sup> (defined as satisfying  $\geq 2$  adjacent active cells and 0 cells active in the same layer), the 'detached' fraction of  $E_{\text{sum}}$  (defined as having 0 adjacent active cells)", the shower starting layer ( $K_S$ ), the number of active cells in the event occurring after  $K_S$ , the number of active cells contributing to the 'track fraction' of  $E_{\text{sum}}$  and finally the number of active cells contributing to the event in the last 4 AHCAL layers.

The gradient-boosted decision tree, using the same conditions as was used to train the standard CALICE PID, was re-optimised to distinguish the training and validation set of simulated  $K_L^0$  from simulated  $\pi^-$  hadron showers in AHCAL. The new classifier was then applied to the testing dataset, without applying the MIP-track cut on the event. Then, the classifier was applied again to the same testing dataset, but this time applying the MIP-track cut to the  $\pi^-$  events. The classifier's performance in both cases was then measured using the *area-under-curve* (AUC) of the classifier's *receiver operating characteristic* (ROC) curve. The ROC curve measures the relationship between the true positive rate and the false positive rate (the ratio of showers correctly/incorrectly classified as  $K_L^0$  to all showers correctly/incorrectly classified as  $K_L^0$ ).

13: Note that this definition is not the same as the MIP-track cut described. This cut includes rectilinearly-distributed energy deposits beyond the shower starting layer, with no requirement that they close to the shower core.

Hyperparameter	Value
Objective	SoftMax
Metric	Multi-Log Loss
Classes	2
Metric Frequency	1
Leaves	10
Max Depth	10
Min Child Samples	10
Learning Rate	0.1
Feature Fraction	0.9
Bagging Fraction	0.8
Bagging Frequency	5

**Table 6.3:** Table of hyperparameters used for the gradient-boosted decision tree used to analyse the effectiveness of the MIP-track cut. Further information about these parameters can be found in [190].

AUC is a measure of inter-class separability. An AUC above 0.5 means the classifier is better than a random guess. An AUC of 0.5 means the classifier is no better than a random guess. Therefore, if the applied MIP-track cut results in synthetic pseudo-neutral hadron showers more similar to real neutral particles in the test sample, it then follows that the AUC of the classifier should reduce from above 0.5 to near 0.5 after the cut is applied. The performance of the cut at producing synthetic pseudo-neutral events can be measured as the difference between the AUC after the cut and the random classifier, before and after the cut is applied. Evaluating cut performance in this way is advantageous as a classifier optimally condenses the distributions of potentially correlated shower variables into a single quantity with clear metrics for evaluation.

The importance of each variable to the model is shown as a bar chart in Figure 6.11. The five most important PID distributions to the classifier are shown in Figures 6.12a-6.12e, with a full set of distributions available in Appendix Figure 8.22. The ROC curves of the model applied to the testing sample of Table 6.1 as described is shown in Figure 6.13.

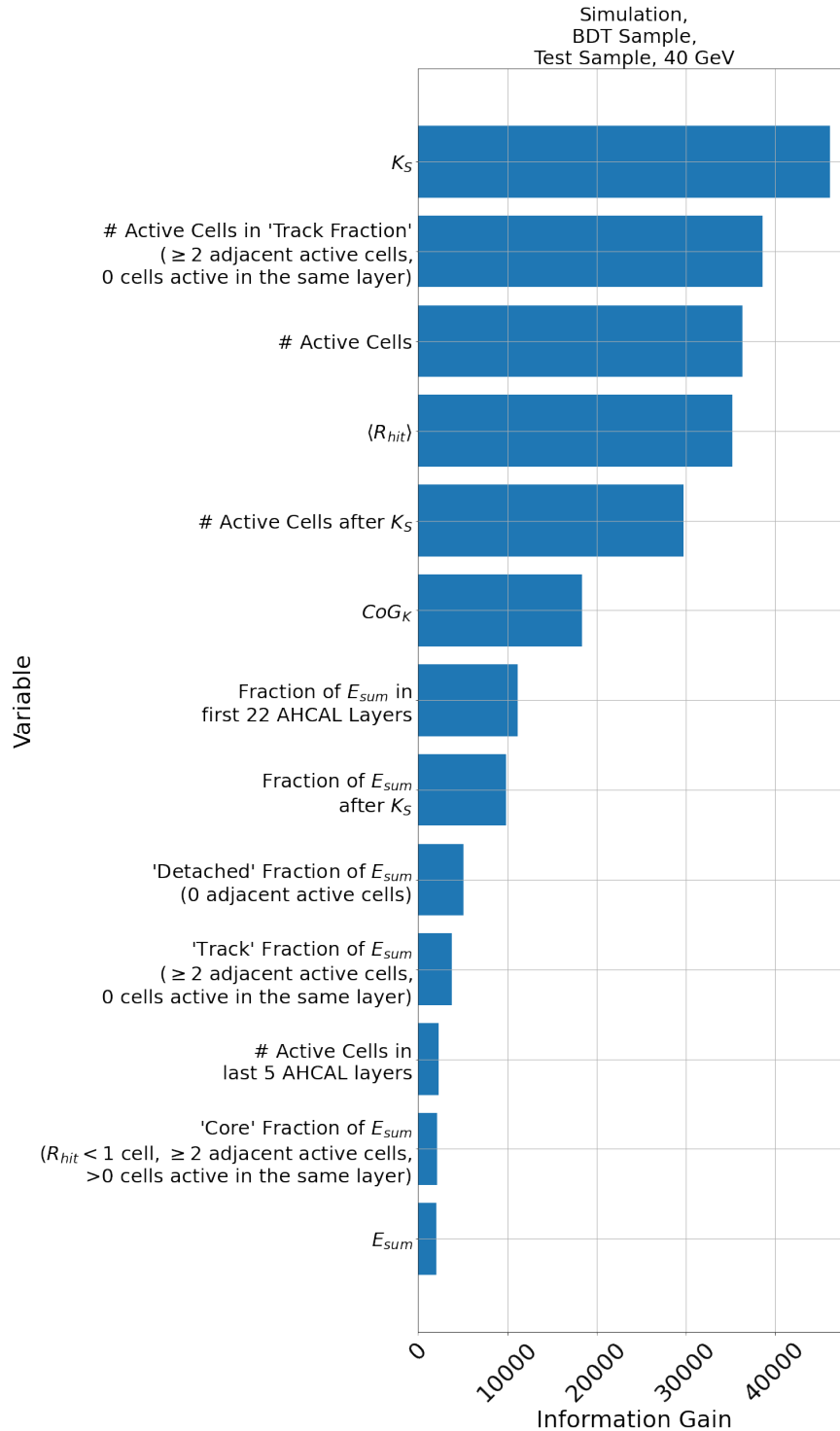
Furthermore, as a cross-check, the comparisons of the differential energy distributions of the synthetic neutral produced with a  $\pi^-$  hadron shower by applying the MIP-track cut and the  $K_L^0$  hadron samples, in the same manner as was performed in Figures 8.19a-8.21f. The results are shown in Figure 6.14.

#### Summary Comment to Figures 6.11-6.14

The AUC of the ROC curve of the trained  $\pi^- / K_L^0$  classifier drops from 0.85 to 0.517 if the MIP-cut is applied to the  $\pi^-$  events in the testing dataset. This means it is highly effective at producing convincing synthetic neutral 40 GeV hadron showers at the event level. These showers also have an energy density distribution that deviates by no more than a 10 % deviation compared to  $K_L^0$  hadrons.

This analysis shows that the synthetic neutral hadrons developed using the method described in this section may be treated as acceptable replacements for true neutral hadron shower events, on average.

### Information Gain of Boosted Decision Tree Classifier by PID Variable

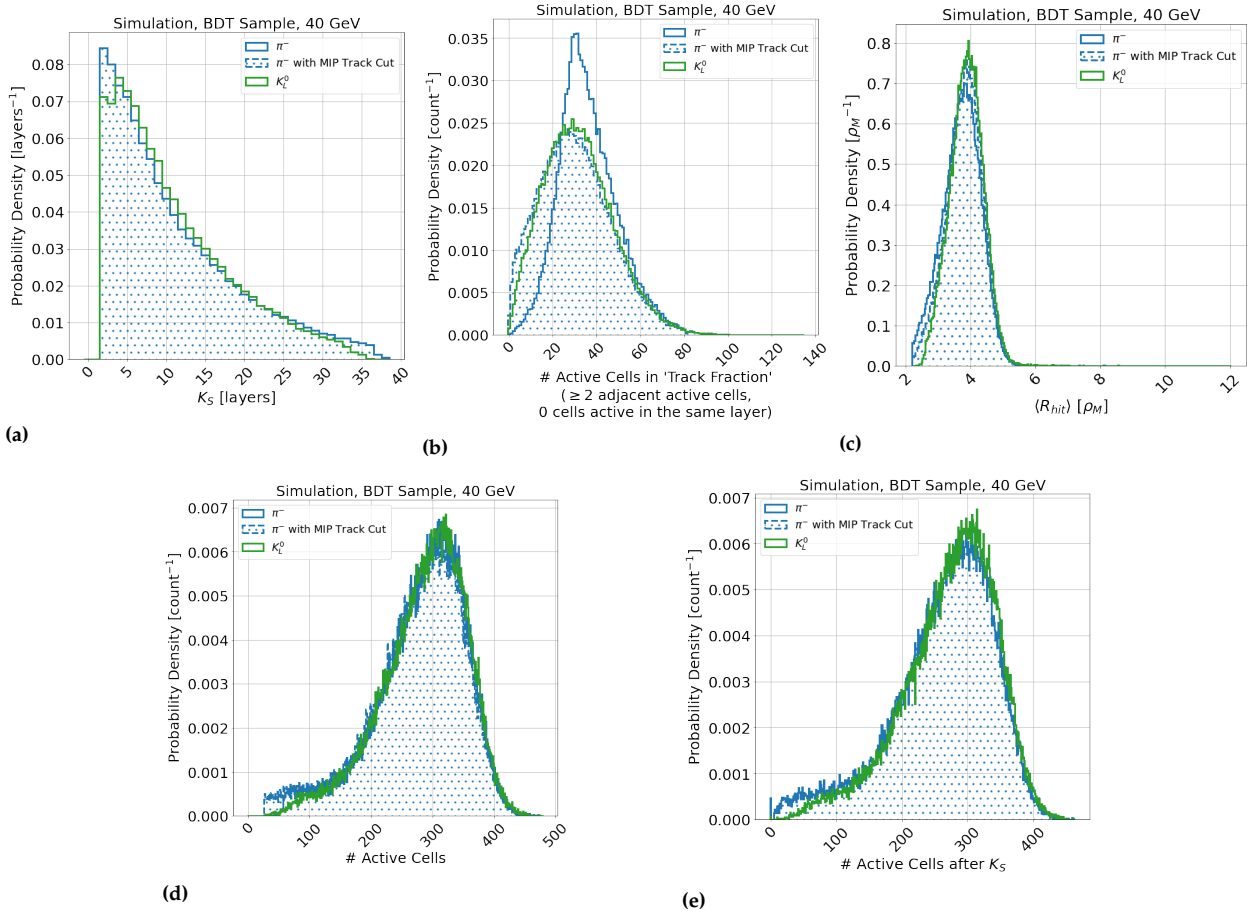


**Figure 6.11:** Information gain (difference in entropy after a split in the tree compared to before) as blue bars for different variables presented in Figure 8.22.

**Comments:**

- Figure 6.11 indicates that the five most significant variables for the classifier model are, in descending order of importance:  $K_S$ , the number of active cells in the track fraction,  $\langle R_{hit} \rangle$ , the number of active cells in the event, and the number of cells after the shower start, and that other variables are comparatively less relevant to determining the distinction between  $K_L^0$  and  $\pi^-$  hadron showers.

**Distribution of Top 5 PID Variables by Information Gain for Comparison of 40 GeV  $\pi^-$ /Synthetic Neutral and  $K_L^0$  Simulated Hadron Showers in AHCAL**

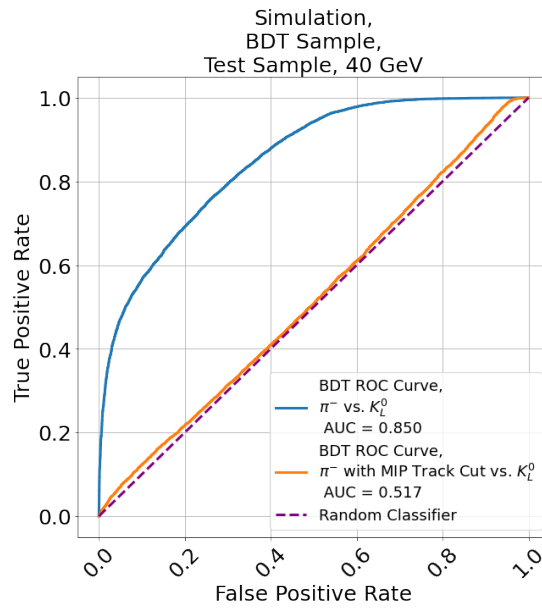


**Figure 6.12:** Top 5 PID variables calculated according to [79, 86], compared for the simulated  $\pi^-$  and  $K_L^0$ , from Table 6.2 in terms of information gain of the BDT, as shown in Figure 6.11. The green line indicates  $K_L^0$  hadron showers, while the blue and blue dashed lines indicate  $\pi^-$  hadron showers, with and without the applied MIP-track cut.

**Comments:**

- In Figures 6.12a-6.12e, it may be seen that the solid blue line differs from the green line in these figures, illustrating that differences are observed between  $K_L^0$  and  $\pi^-$  hadrons. Furthermore, the variables with the most information gain in the model are related to the expectations for variables associated with the MIP-track, and therefore it is suggestive the model is using this information to distinguish the charged from the neutral hadron shower.

### Receiver Operating Curve (ROC) for the BDT $\pi^-/K_L^0$ Classifier, Applied Before and After the Application of The MIP-Track Cut

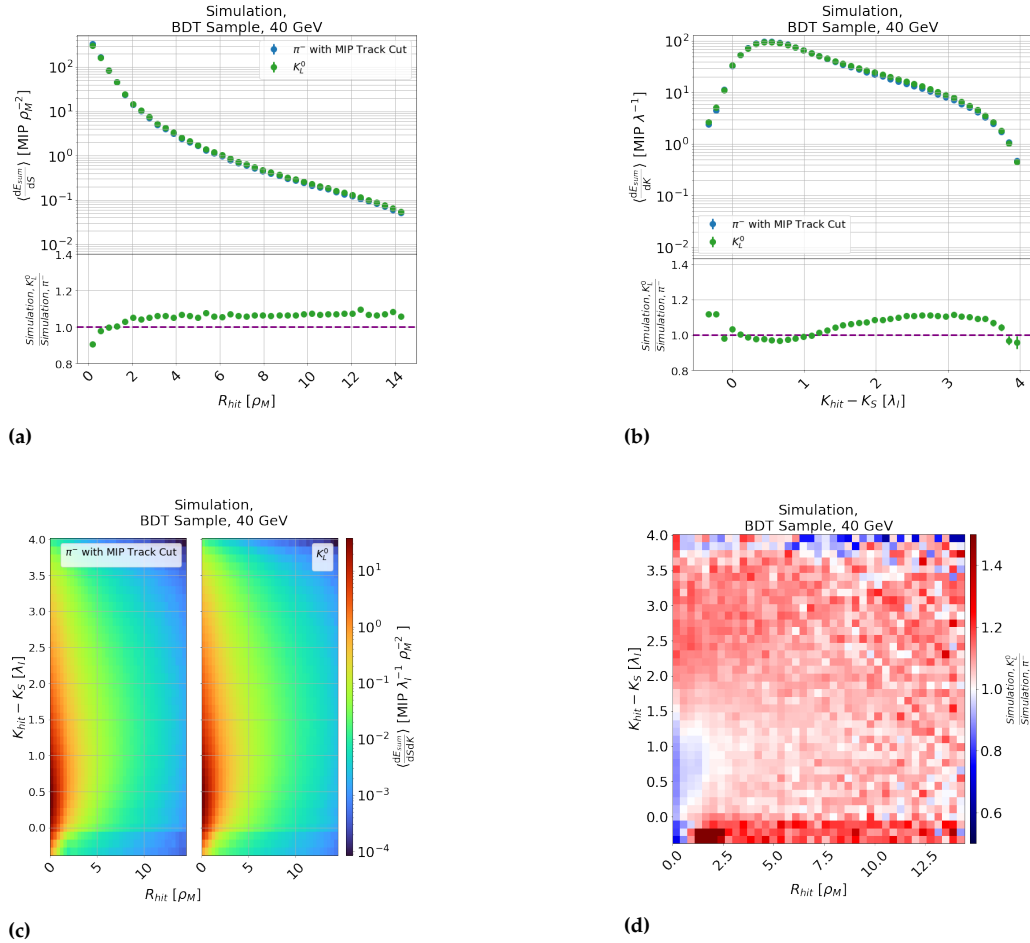


**Figure 6.13:** ROC curves from the trained classifier applied to the test sample from the BDT sample of Table 6.2. The blue and orange lines indicate the performance of the model applied to the same testing sample of the BDT sample of  $K_L^0$  and  $\pi^-$  hadron shower event variables, without and with the MIP-track cut applied. The purple dashed line indicates the expected curve for a random classifier.

#### Comments:

- Figure 6.13 shows that the classifier performs with an AUC of 0.85 for  $\pi^-$  and  $K_L^0$  showers in the BDT sample test sample before the MIP-track cut, and an AUC of 0.517 to the same sample after the MIP-track cut was applied. It may be concluded that event level information can be used to distinguish between 40 GeV  $K_L^0$  and  $\pi^-$  hadron showers significantly more frequently than a random classifier, and that the removal of the energy of deposited by the MIP-track results in a classifier only marginally better than a random classifier. This means that the synthetic neutral hadron showers produced from  $\pi^-$  events and  $K_L^0$  showers have similar event level distributions, and the removal of the MIP-track obtained the similarity.

Radial, Longitudinal And Joint Radial-Longitudinal Energy Profiles to Compare  $\pi^-$  and  $K_L^0$  Hadron Showers



**Figure 6.14:** Figures 6.14a and 6.14b show the differential shower profile distributions in radial surface area ( $S$ ) and longitudinal depth ( $K$ ), presented on the  $x$ -axis in units of Moliere radius and on the  $y$ -axis in units of nuclear interaction lengths, for the neutral events produced from simulated  $\pi^-$  hadrons showers with the MIP-track cut applied and simulated  $K_L^0$  hadron showers samples at 40 GeV, respectively. The subplots indicate the ratio of the simulated  $K_L^0$  to the neutral. Figures 6.14c indicate the joint (2D) differential energy distribution in  $S$  and  $K$ , presented in the same units. Figure 8.21d shows the ratio between the simulated  $K_L^0$  to the neutral of Figure 6.14d, presented in the same units on the  $x$ -axis and  $y$ -axis, and the ratio presented on the colour axis.

**Comments:**

- ▶ Figures 6.14a and 6.14b indicate a slight excess of energy density overall for the  $K_L^0$  by comparison to the synthetic neutral, except the region within  $1 \rho_M$  from the core and within the  $0-1 \lambda_I$  from the shower starting layer, which shows a deficit. Nonetheless, both distributions agree within 10%, demonstrated in the subplots. Similar trends were observed in a similar analysis performed in [78].
- ▶ Figures 6.14c and 6.14d indicates the same overall trends as Figures 6.14a and 6.14b. The blue region before the shower start where  $R_{hit} < 1 \rho_M$  indicates that the MIP-track cut leaves an excess of energy density, indicating the cut is suboptimal.



### 6.2.3.2 Shower Displacement

Confusion between two hadron shower events is expected to be correlated to many observables (e.g. particle energy, shower starting position, energy density). However, an obvious caveat of synthetically producing charged-neutral hadron shower events is that the lateral distance between the particles initiating the hadron shower, which plays a critical role in the level of intrinsic confusion, is left as a free parameter.

The study of [120] uses a uniform distribution of impinging particle positions on the calorimeter. One critical observation is that the average distance between uniformly selected incidence positions of the hadrons upon a circle/square sampling region scales linearly with the radius/side length. However, the study of [78] indicates an approximately S-shaped, and therefore nonlinear, dependence of confusion energy on inter-shower distances. Evaluation of the appropriate value of this quantity is therefore critical when considering the quotation of performance of the neural networks, which, unlike algorithms such as Pandora PFA, bias to the dataset they trained upon. This also means both hadron showers must be displaced in the synthesised dataset, rather than just one, to prevent the networks from biasing to incorrect assumptions (e.g. the algorithm may learn to expect all charged showers to be observed in the centre of the calorimeter, which is an unfair bias).

A study ensures that the average distance between the charged and neutral shower is neither too small nor too large. In the former case, the algorithms will be biased to mostly unresolvable events. In the latter case, the shower separation task becomes trivial on average (i.e. separation of energy deposits with a plane in space through the calorimeter cells), which would mean the influence of the additional benefits of a temporal calorimeter cannot be effectively studied. The study produces a physically-motivated relationship between the radial distance from the hadron shower and the average cumulative integrated energy at that distance is obtained which may be used to calculate an appropriate average separation distance based on the average distribution of energy of a hadron shower in AHCAL, such that the model has a balance between well-separated and overlapping hadron shower events.

Firstly, as in Figures 8.19a-8.19c, the differential energy loss by a hadron shower per unit area of a circle (i.e. a thin ring) around the centre of gravity ( $dS = 2\pi R_{\text{hit}} dR_{\text{hit}}$ , where  $dR_{\text{hit}}$  is a bin width), using the entire training sample of Table 6.1, was determined. In this study, the bin width was chosen using the Freedman-Diaconis bin rule [147], resulting in much finer binning than shown in these figures. A spline was then fitted to the distribution. The spline was then integrated over the surface area, according to Equation 6.2:

$$\langle E_{\text{sum}}(R_{\text{hit}}) \rangle = \int_0^{R_{\text{hit}}} \left\langle \frac{dE_{\text{sum}}}{dS} \right\rangle \cdot 2\pi R_{\text{hit}} dR_{\text{hit}} \quad (6.2)$$

The cumulative average energy deposited in AHCAL as a function of  $R_{\text{hit}}$  can be obtained for each particle energy of the training sample by numerical integration. The cumulative integral was then normalised to a percentage of  $\langle E_{\text{sum}} \rangle$ , for each particle energy. These curves were evaluated at the of  $\langle E_{\text{sum}} \rangle$  to find the corresponding  $R_{\text{hit}}$ , corresponding

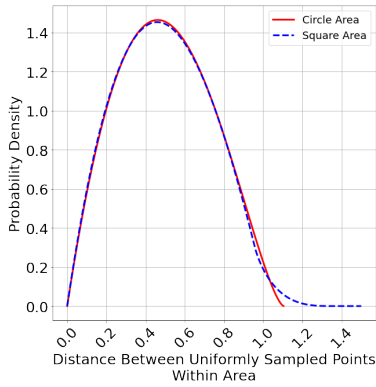
to the radius of a circle with its origin at the centre-of-gravity, within which that percentage of the total calorimeter response for that energy would be measured, on average. This quantity determines the relationship between the particle energy and how far away in radial distance they should be separated.

The distance corresponding to 80 % of  $\langle E_{\text{sum}} \rangle$  was chosen to characterise the synthetic dataset. It was found that the measured relationship between  $R_{\text{hit}}$  and  $\rho_M$  could be fitted within 1-2 % of the value using an ad-hoc relationship shown in Equation 6.3:

$$R_{\text{hit}}^{80\%}(E_{\text{particle}}) = a_R + b_R \cdot \log E_{\text{particle}} + c_R \cdot E_{\text{particle}} \quad (6.3)$$

The distribution of synthetic shower separation distances can then be calculated. The choice was made to distribute the events randomly and uniformly in a circle of radius  $R_{\text{circ}}$  so that the neural network learns no anisotropic biases in the distribution of the centres-of-gravity of the hadron shower. The radius is chosen such that 80 % of  $\langle E_{\text{sum}} \rangle$  would be integrated, on average.

The AHCAL calorimeter front-face is square ( $24 \times 24$  cells). Thus, it may seem natural to sample distances to separate hadron showers within a square area. However, uniform random sampling in a square area produces a more heavily right-tailed distribution of separation distances (i.e. skewed to greater distances) than for the same separation distance in a circle [191]. A more symmetric distribution of inter-shower distances about the mean will result in a fairer assessment of the shower separation models because the models will experience less bias to already well-separated events. The distributions are plotted for reference in Figure 6.15



**Figure 6.15:** Comparison of the distribution of distances between points sampled uniformly with the same mean distance of 0.5 in a circle area, shown in red, and in a square area, shown in blue. Remade from [191].

The relationship between the radius of a circle and the average distance at which two uniformly distributed points within the circle is given by [192]:

$$R_{\text{sep}} = \frac{128}{45\pi} R_{\text{circ}} \quad (6.4)$$

Finally, the  $R_{\text{circ}}$  required to achieve 80 % of  $\langle E_{\text{sum}} \rangle$  is given by Equation 6.5:

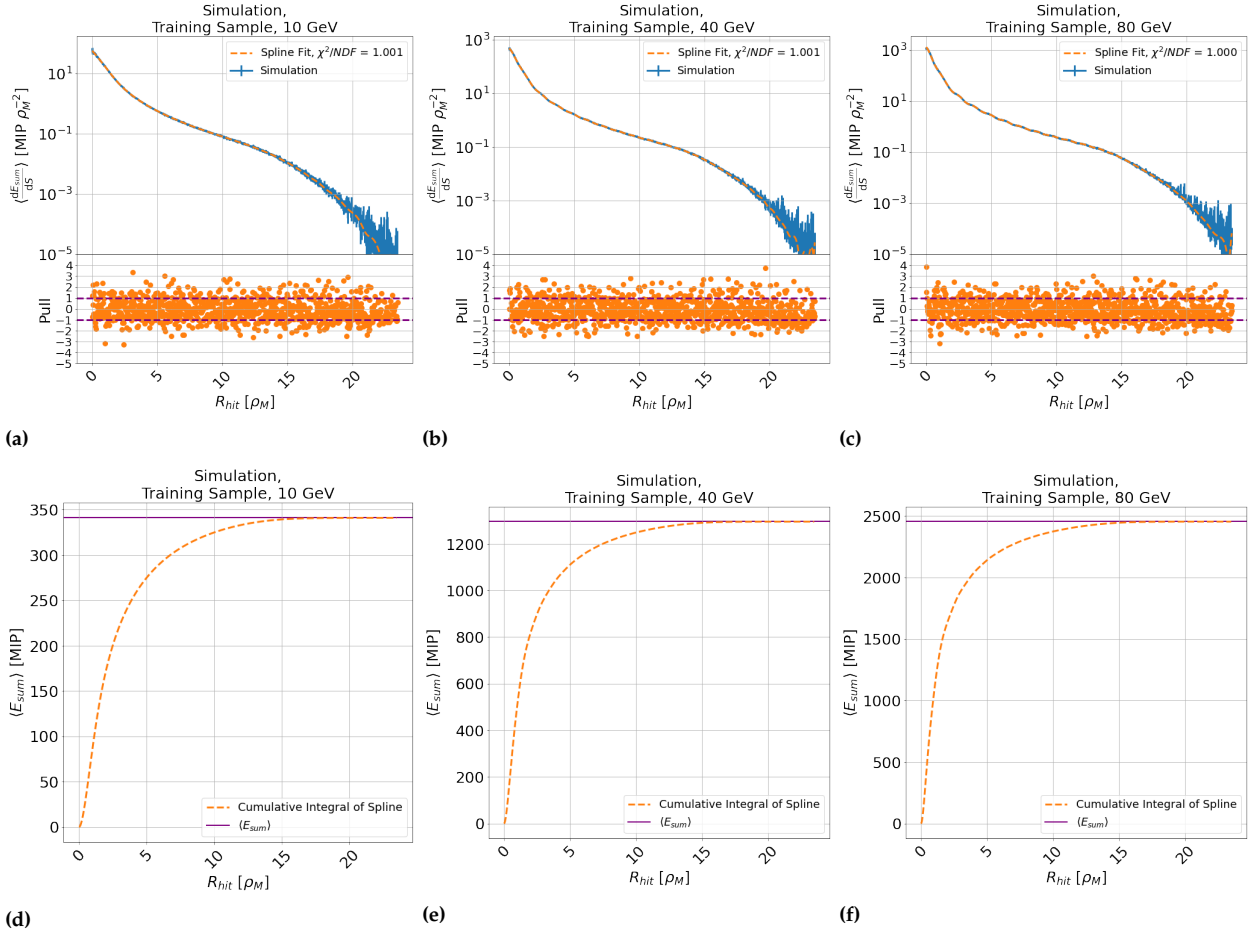
$$R_{\text{circ}}(E_{\text{particle}}^Q, E_{\text{particle}}^N) = \frac{45\pi}{128} \left( R_{\text{hit}}^{80\%}(E_{\text{particle}}^Q) + R_{\text{hit}}^{80\%}(E_{\text{particle}}^N) \right) \quad (6.5)$$

The method described was applied to the training sample of Table 6.1. The differential energy distributions and their cumulative integrals are shown for the 10 GeV, 40 GeV and 80 GeV samples in Figure 6.16. The distances corresponding to 60 %, 70 %, 80 % and 90 % of  $\langle E_{\text{sum}} \rangle$ , as well as a fit of Equation 6.3 to the 80 % curve, are shown in Figure 6.17.

**Summary Comment to Figure 6.16 and Figure 6.17**

The applied spline fits are found to give excellent agreement to the differential energy distributions. The distance at which 80 % of the reconstructed energy is measured from the shower core, on average, is found to follow an approximately logarithmic relationship according to Equation 6.3 in agreement within 1-2 %.

### Radial Energy Profile and Radial Cumulative Integral of the Events Used for Shower Separation Training

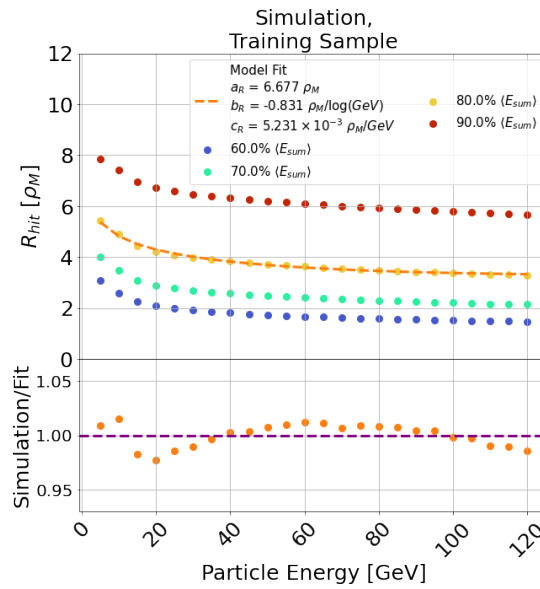


**Figure 6.16:** Figures 6.16a-6.16c show differential energy loss per unit surface area as a function of  $R_{hit}$  for the training sample of  $\pi^-$  simulation in Table 6.1, as shown in Figures 8.19a-8.19c, using finer binning. The blue points show the simulation, and the orange dashed line indicates a spline fit to the points. The pulls are shown in the bottom subplot. Figures 6.16d-6.16f indicate the cumulative integral of the spline fit, according to Equation 6.2. The solid purple line indicates the expected mean energy loss from the training sample, illustrating that the model saturates at that value.

#### Comments:

- ▶ The spline fits shown in Figures 6.16a-6.16c indicate excellent agreement to simulation, with a reduced  $\chi^2 \approx 1$ . This observation is also illustrated in the pulls of the fit.
- ▶ Figures 6.16d-6.16f indicate that the cumulative integral of Figures 6.16a-6.16c saturate at the expected mean value for the hadron shower energy. This means that the integration has been performed correctly.

**Relationship Between Radial Distance and Integrated Energy, for 60 %, 70 %, 80 % and 90 % Percentiles of Average Reconstructed Energy for Events used in Shower Separation Training**



**Figure 6.17:** Extracted distances at which proportions of the average  $\pi^-$  hadron shower energy in simulation have been integrated. The blue, teal, orange, and red circle markers indicate the 60 %, 70 %, 80 % and 90 % percentiles of  $\langle E_{\text{sum}} \rangle$ , presented as a function of particle energy. The dashed orange line indicates an ad-hoc fit to the 80 % percentile.

**Comments:**

- ▶ Two effects can be observed in Figure 6.17. The first is that the distance required to separate different percentages of energy changes non-linearly, indicated by the non-equidistant vertical spacing between the different coloured points. This is expected from the expectation of bi-exponential energy distribution as a function of the differential area from the centre-of-gravity [34]. The other effect is that the radius required to integrate a certain percentage of the hadron shower decreases with the particle energy. This is also expected since the EM-fraction of the hadron shower increases with particle energy, thus becoming more energy dense (see Section 2.3.6).
- ▶ The distance corresponding to 80 % of  $\langle E_{\text{sum}} \rangle$  was chosen to characterise the synthetic dataset. The subplot of the figure indicates that the measured relationship between  $R_{\text{hit}}$  and  $\rho_M$  could be fitted within 1-2 % of the value using an ad-hoc relationship, shown in Equation 6.3. For 80 % of  $\langle E_{\text{sum}} \rangle$ ,  $a_R = 6.677 \rho_M$ ,  $b_R = -0.831 \rho_M / \log \text{GeV}$  and  $c_R = 5.231 \times 10^{-3} \rho_M / \text{GeV}$ .

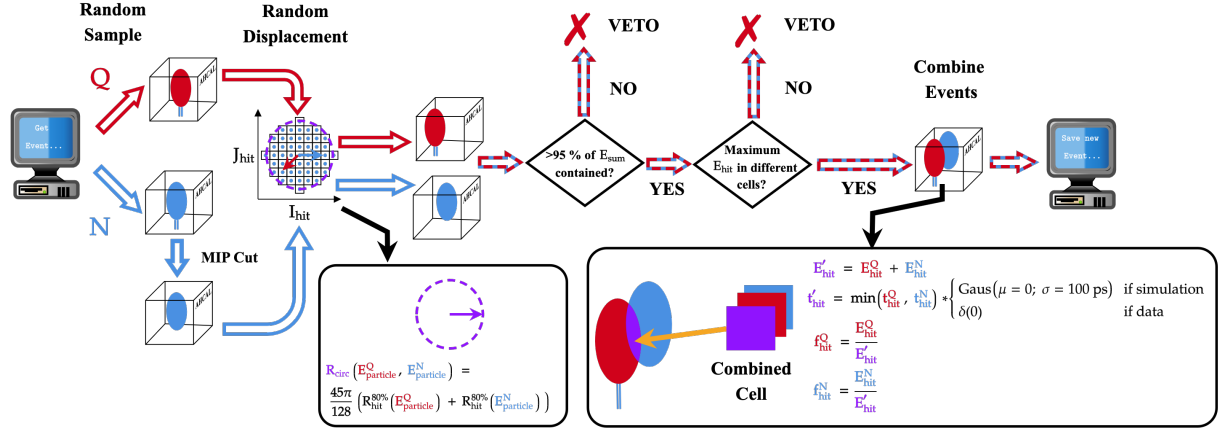
### 6.2.3.3 Synthetic Neutral and Charged Hadron Shower Events

With the results of Section 6.2.3.1 and Section 6.2.3.2, an algorithm to combine a sample of single  $\pi^-$  hadron showers into an event with a charged hadron and a synthetic neutral hadron shower is introduced.

The algorithm is defined as follows:

1. Two  $\pi^-$  hadron shower events are selected by a weighted random subsample of the main sample. One is designated the charged candidate,  $Q$ ; the other the synthetic neutral candidate,  $N$ . The weight is selected such that it is no more likely to combine a certain combination of particle energies than any other;
2. For the neutral shower only, the MIP-track cut described in Section 6.2.3.1 is applied to  $N$ .
3. Four integers,  $\Delta I^Q$ ,  $\Delta I^N$ ,  $\Delta J^Q$  and  $\Delta J^N$ , are defined as distances in cells by which to displace both  $Q$  and  $N$  in  $I$  and  $J$  in calorimeter space. These quantities are sampled within a circle of  $R_{\text{circ}}$  and centred at  $I_{\text{hit}} = 0$  cell,  $J_{\text{hit}} = 0$  cell. All of the active cells and track positions of  $Q$  and  $N$  are then shifted by the two sampled integers. For example, all the hits in  $Q$  are displaced by  $\Delta I^Q$  and  $\Delta J^Q$ , (i.e.  $I_{\text{hit}}^Q \rightarrow I_{\text{hit}}^Q + \Delta I^Q$ ,  $J_{\text{hit}}^Q \rightarrow J_{\text{hit}}^Q + \Delta J^Q$ ), and the track position is also shifted by the same integers (i.e.  $I_{\text{track}}^Q \rightarrow I_{\text{track}}^Q + \Delta I^Q$ ,  $J_{\text{track}}^Q \rightarrow J_{\text{track}}^Q + \Delta J^Q$ ). The modified 'track position' of  $N$  is only used in Figure 6.23 and Figure 6.24 in this section;
4. Any active cells from both events displaced outside the calorimeter are cut from the event. If there is more than 95% of each event's energy still contained within the calorimeter and if the maximum energy cells are not shared for consistency with the study of [120], then the algorithm proceeds. At this stage, event-level variables such as the centre-of-gravity are recalculated based on the MIP-track cut and removed active cells due to the containment criteria. If the criteria are not satisfied, the event is vetoed, and the integers for displacement are sampled again until the criteria are satisfied. The fraction of energy remaining in the event is recorded for analysis (see Figure 6.21);
5.  $Q$  and  $N$  are merged by combining the energy of the events. For active cells in  $Q$  and  $N$  sharing energy:
  - ▶ the active cell energies from  $Q$  and  $N$  are summed, such that  $E_{\text{hit}} = E_{\text{hit}}^Q + E_{\text{hit}}^N$ ;
  - ▶ the minimum  $t_{\text{hit}}$  of  $Q$  and  $N$  is chosen to characterise the active cell, such that  $t_{\text{hit}} = \min(t_{\text{hit}}^Q, t_{\text{hit}}^N)$ . In the simulation, the true  $t_{\text{hit}}$  is known, and therefore a random Gaussian smearing of the hit time of 100 ps is applied after the minimum is selected. Since the  $t_{\text{hit}}$  in data is not studied and already includes the resolution of the AHCAL detector, no additional smearing is applied;
  - ▶ The fractions of energy belonging to each hadron shower,  $f_{\text{hit}}^Q$  and  $f_{\text{hit}}^N$  are calculated from the new  $E_{\text{hit}}$ , such that  $f_{\text{hit}}^Q = E_{\text{hit}}^Q / E_{\text{hit}}$  and  $f_{\text{hit}}^N = E_{\text{hit}}^N / E_{\text{hit}}$ ;
6. The event is then stored for later use.

The result of the algorithm is an charged-synthetic neutral AHCAL two-hadron shower event that may be applied to both simulation and experimental data to train and validate machine learning algorithms for shower separation. A diagram of the algorithm is shown in Figure 6.18.



**Figure 6.18:** Illustration of the shower-combination algorithm used to produce synthetic charged-neutral hadron shower events from a sample of pure  $\pi^-$  hadron showers observed with AHCAL. The red and blue arrows indicate the paths taken by the charged and synthetic neutral hadron shower. The squared circle indicates the distribution of available cells by which the event can be displaced from its entry position.

## 6.2.4 Synthetic Datasets and Training

### 6.2.4.1 Two-Shower Charged-Synthetic Neutral Datasets

The synthetic datasets used in the analysis in this section were produced using the method described in Section 6.2.3.3. Showers were combined in the training, validation and testing samples using events exclusively from the corresponding datasets shown in Table 6.1, such that the network performance could be evaluated without the possibility that the models had been exposed to the same events during the training phase as during the in-training validation or testing phases.

For simulation,  $7.2 \times 10^5$  and  $8 \times 10^4$  synthetic charged-neutral hadron shower events were produced for training and validating the neural networks during the training phase, respectively, while  $8 \times 10^5$  events were produced to test the models. It is reiterated that each combined sample contained events purely from the corresponding source samples outlined in Table 6.1.

The same number of events for training and validation samples were chosen for data. However, owing to a smaller sample of test events in the source, as shown in Table 6.1, a smaller sample of  $2 \times 10^5$  events was used to test the models using data.

Appendix Table 8.6 and Appendix Table 8.7 break down the synthetic datasets used for the study for simulation and data, respectively. The weighted sampling is illustrated by the equal proportions of each possible combination of hadron shower energies.

A summary analysis of applying the algorithm of Section 6.2.3.3 was performed. Figure 6.19 and Figure 6.20 show example event displays of the synthetic training sample in simulation before and after the shower

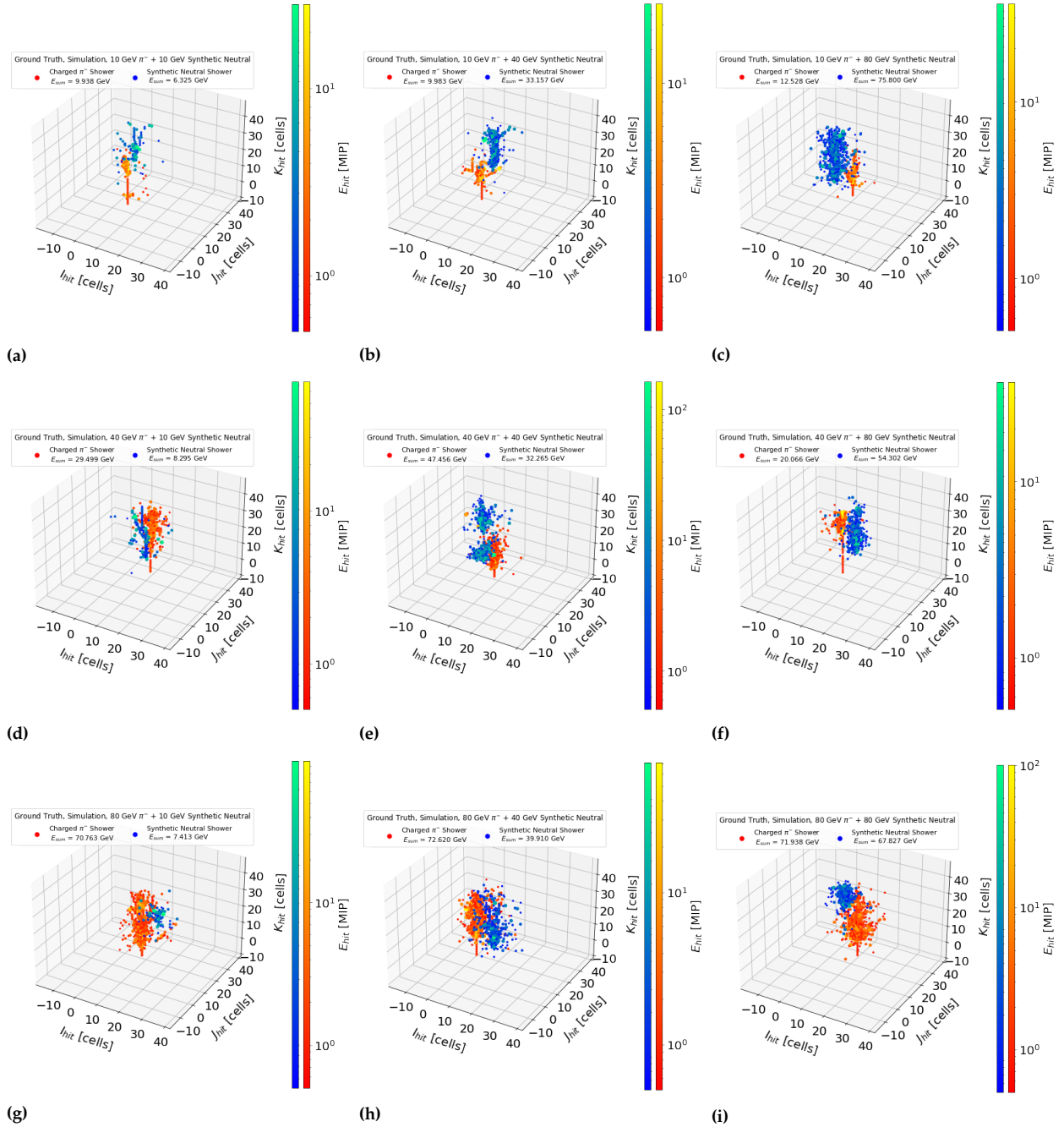
combination has been applied. Figure 6.21 show the distributions of the fractions of the  $E_{\text{sum}}$  remaining in the calorimeter after the shower combination algorithm has been applied, and the distribution of distances between the charged and synthetic neutral hadron showers after combination. Figure 6.23 and Figure 6.24 shows the joint distribution between the track position and the lateral centre-of-gravity after combination. Figure 6.25 show the joint distribution of the centres-of-gravity before and after the combination.

**Summary Comment to Figures 6.19-6.25 and Appendix Tables 8.6-8.6**

Equal proportions of each shower combination are obtained by weighted sampling. Good agreement is observed with respect to expectations on the distribution of showers in the calorimeter and the fraction of energy remaining after the synthesis algorithm was applied.



## Event Displays of Ground Truth Charged and Synthetic Neutral Hadron Showers in Simulation Observed With AHCAL



**Figure 6.19:** Example event displays of combined synthetic neutral and charged hadron showers. Events are picked from the list in Table 6.1, for combinations of 10 GeV, 40 GeV and 80 GeV hadron shower events. The red and blue points indicate an unmodified simulated charged  $\pi^-$  hadron shower and a synthetic neutral hadron shower produced with the MIP-track cut of Section 6.2.3.1, displaced from their initial position by random uniform displacement according to the method of Section 6.2.3.2. Figures 6.20a-6.20i show event displays of the same events, combined according to the algorithm of Section 6.2.3.3.

### Comments:

- The event displays qualitatively illustrate that the charged and synthetic neutral hadron showers are distributed in a wide range of configurations and that the MIP-track has been subtracted for the neutral hadron shower.

### Event Displays of Merged Charged and Synthetic Neutral Hadron Showers in Simulation Observed With AHCAL

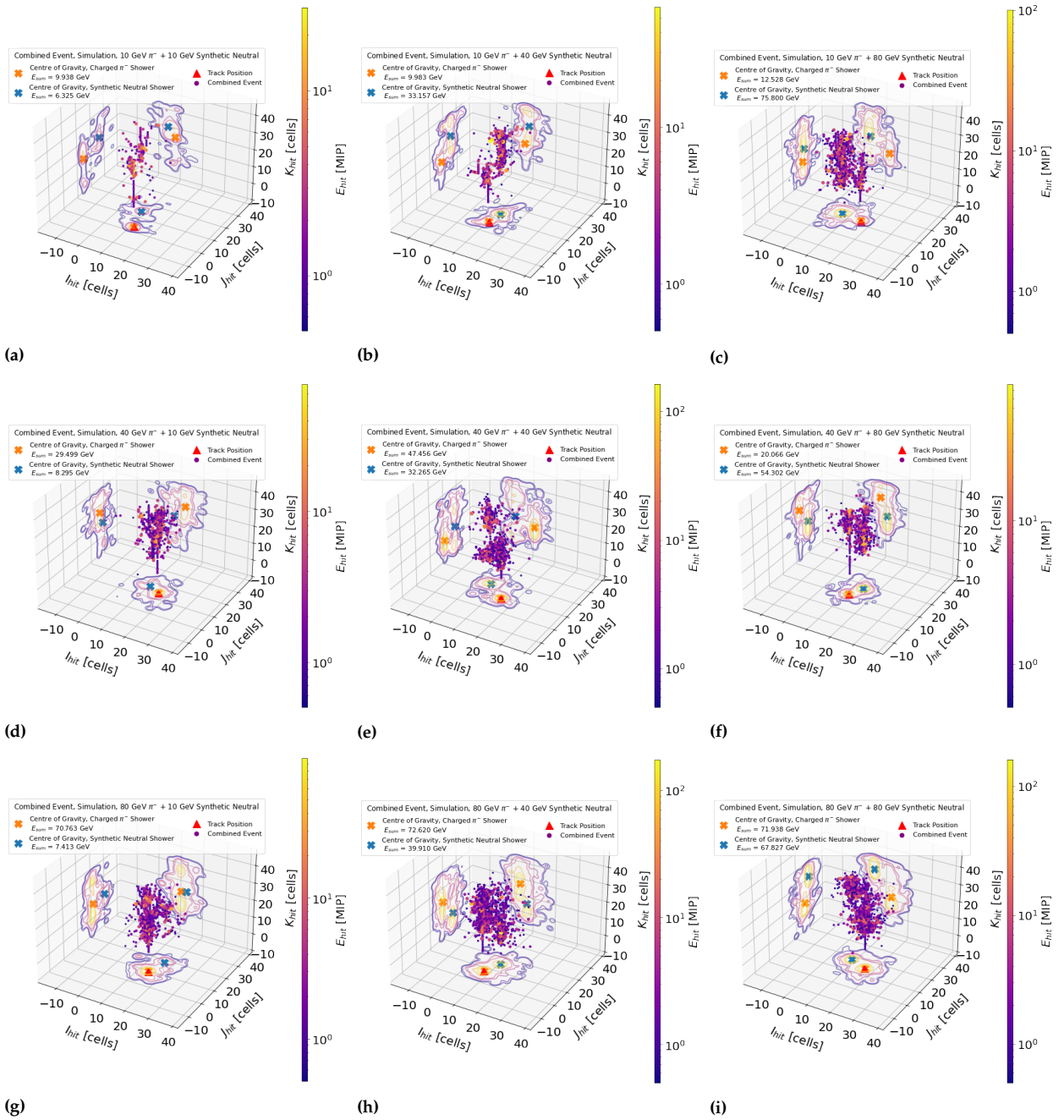
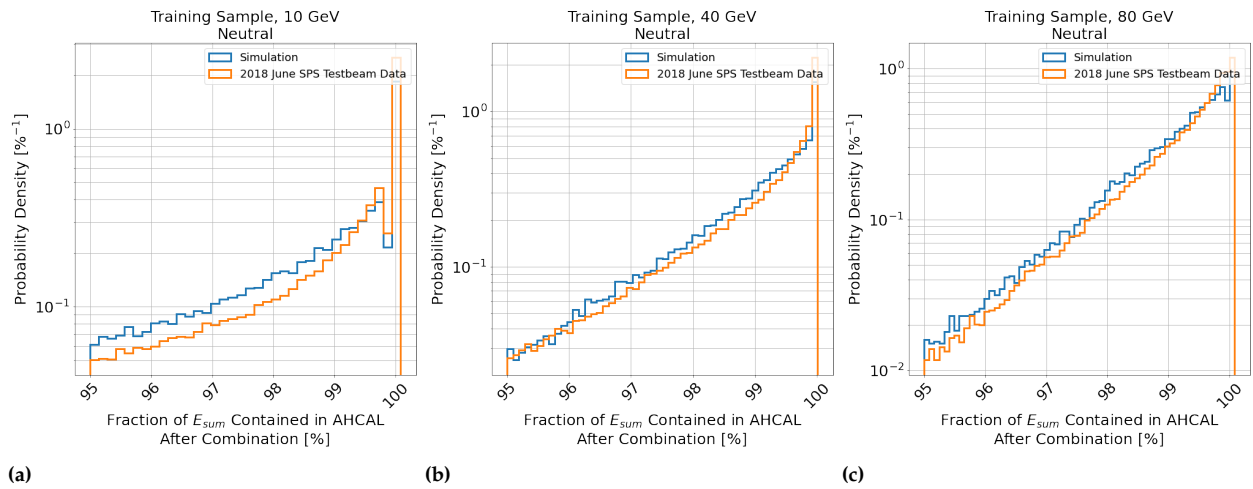


Figure 6.20: Example event displays of combined synthetic neutral and charged hadron showers. Events are the same hadron shower events as shown in Figure 6.19.

**Comments:**

- Figures 6.20a-6.20i demonstrate that the algorithm of Section 6.2.3.3 and that the combined events are qualitatively close enough together that not all of the hadron showers are trivially separable, as indicated by the overlapping ‘energy shadows’ projected on the sides of the plot. The red triangle also demonstrates that the track has been displaced since it remains highly correlated with the centre of gravity of the charged shower.

### Fraction of Reconstructed Synthetic Neutral Energy Contained within the AHCAL After Application of the Shower Combination Algorithm

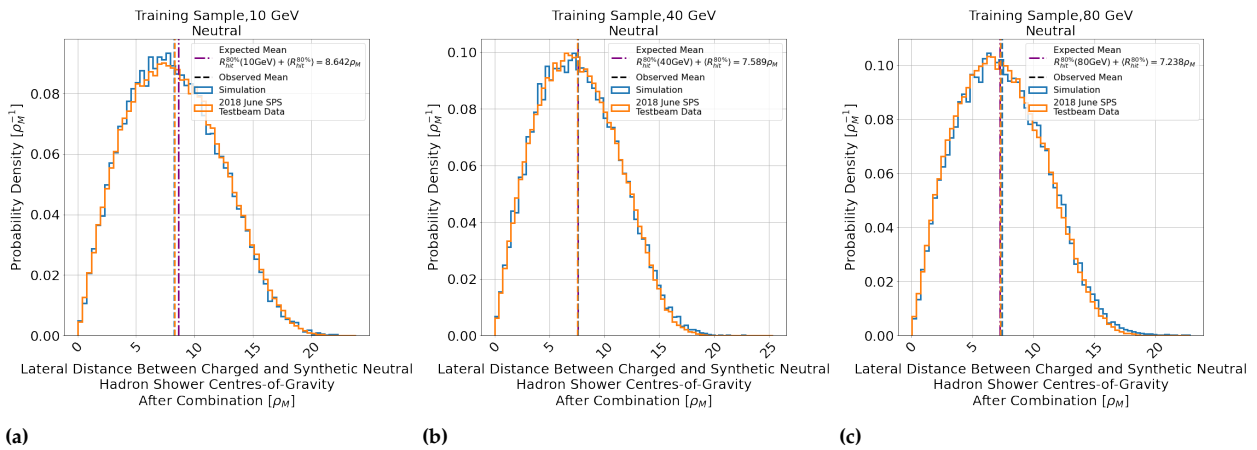


**Figure 6.21:** Distribution of the fraction of energy remaining in the AHCAL after the shower combination algorithm has been applied (i.e. the percentage of energy that would have been observed in the AHCAL if the incidence position had been shifted by  $\Delta I^N$  and  $\Delta J^N$  compared to if it had never been displaced) for 10 GeV, 40 GeV and 80 GeV synthetic neutral hadron showers, after the shower combination algorithm of Section 6.2.3.3 has been applied. Blue and orange lines indicate simulation and data, respectively.

#### Comments:

- Figures 6.21a-6.21c illustrate that the fraction of energy contained by the calorimeter after the shower combination algorithm for the neutral hadron shower is in the range 95-100%.
- It may be observed by comparing the axes of Figure 6.21a and 6.21c that, as the energy of the particle increases, so too does the proportion of events with energy missing due to being observed outside the calorimeter decrease, for both simulation and data. This is expected from the greater influence of the HAD halo for 10 GeV hadron showers than 80 GeV hadron showers, meaning that energy is more likely to be deposited outside the calorimeter when the event has a sparser halo.
- In the figures, the distribution of contained energy is found to follow an approximately exponential distribution, which is consistent between data and simulation. It is noted that slight differences were observed between data and simulation in Figure 6.21a, though these differ. The reasons for this effect are unknown but are likely related to the greater energy density of simulation compared to data in the HAD halo as discussed in Section 5.2.3.1 and shown in the red regions at  $R_{hit} > 2\rho_M$  in Appendix Figure 8.21. This observation does not significantly affect the distribution at 40 GeV and 80 GeV.

### Inter-shower Distance Distributions Between the Charged and Synthetic Neutral Hadron Showers After Application of the Shower Combination Algorithm

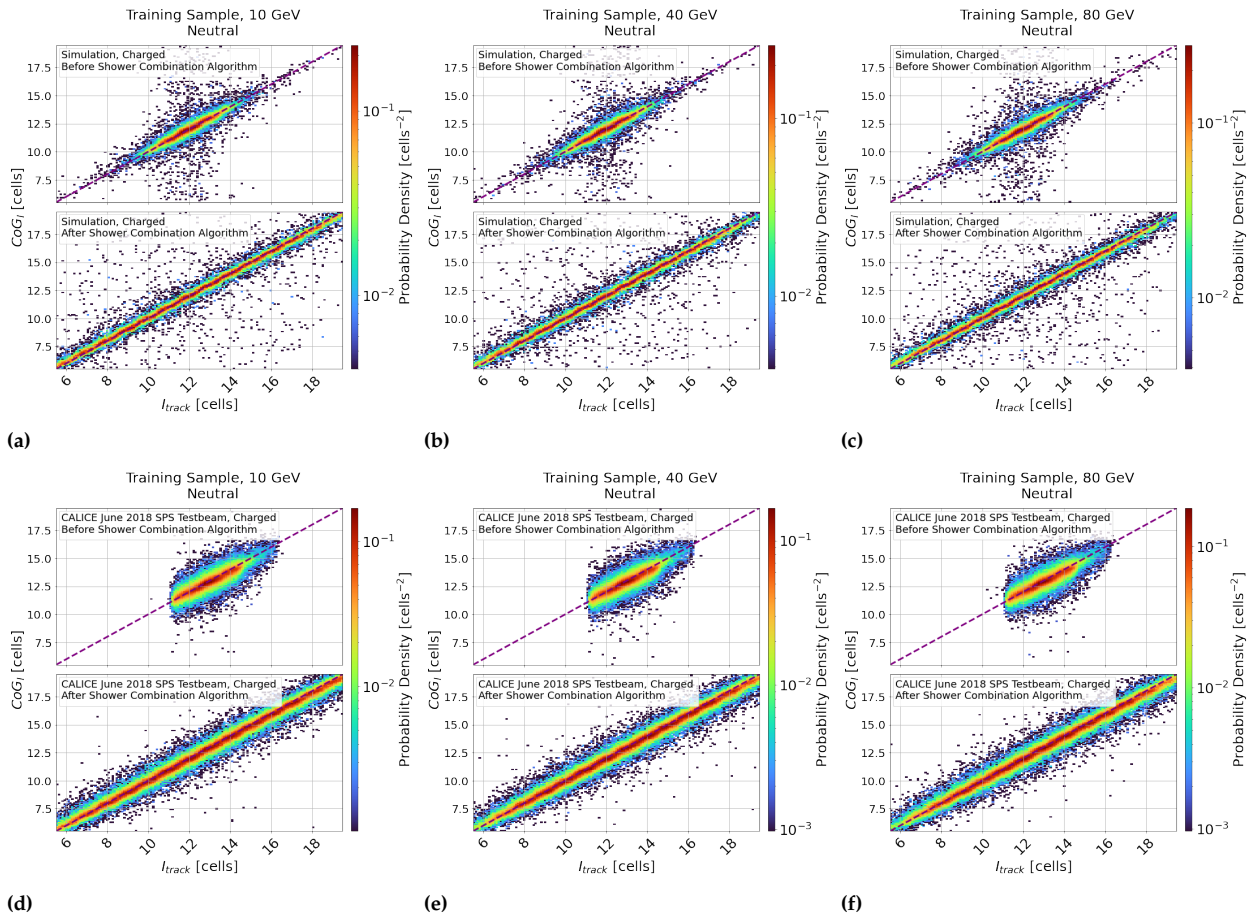


**Figure 6.22:** Distributions of the distance between the lateral centres-of-gravity (CoG<sub>I</sub>, CoG<sub>J</sub>) after the shower combination algorithm has been applied after the shower combination algorithm has been performed, for 10 GeV, 40 GeV and 80 GeV synthetic neutral hadron showers. Blue and orange lines indicate simulation and data, respectively. The expected mean distance between hadron showers from Equation 6.3 is shown as a dash-dotted purple line, which is the sum of the distance at the particular particle energy and the average of the curve shown in Figure 6.17, sampled at all possible charged particle energies shown in Table 6.1. The mean of the simulation and data distributions are shown as blue and orange lines respectively, for comparison.

#### Comments:

- ▶ Figures 6.22a-6.22c indicate excellent agreement of the mean separation distance for each sample of synthetic neutral particle energy with the expectations of Equation 6.3. This indicates that the distance distribution agrees with the expectations of the centres of gravity being separated in a circle. Additionally, the shape of the distribution also agrees with the example illustrated by the blue curve in Figure 6.15. A fit is not performed here, since the circular distribution is demonstrated visually in Figure 6.25. The distance between showers is around  $7.5\rho_M$  overall.
- ▶ Simulation and data produce the same distributions of shower distances, which is expected. A broad distribution of distances between hadron showers is observed for both samples. The lower and upper quartile of each distribution is around  $4.7\rho_M$  and  $10.5\rho_M$ , respectively, indicating a wide range of possible shower configurations. It is also noted that the total probability of a distance falling above and below the mean is approximately equal. Therefore, it may be argued that a fair dataset for training shower separation models has been produced.

### Joint Distributions Lateral Centres-of-Gravity And Track Position in $I$ of Synthetic Neutral After Application of the Shower Combination Algorithm

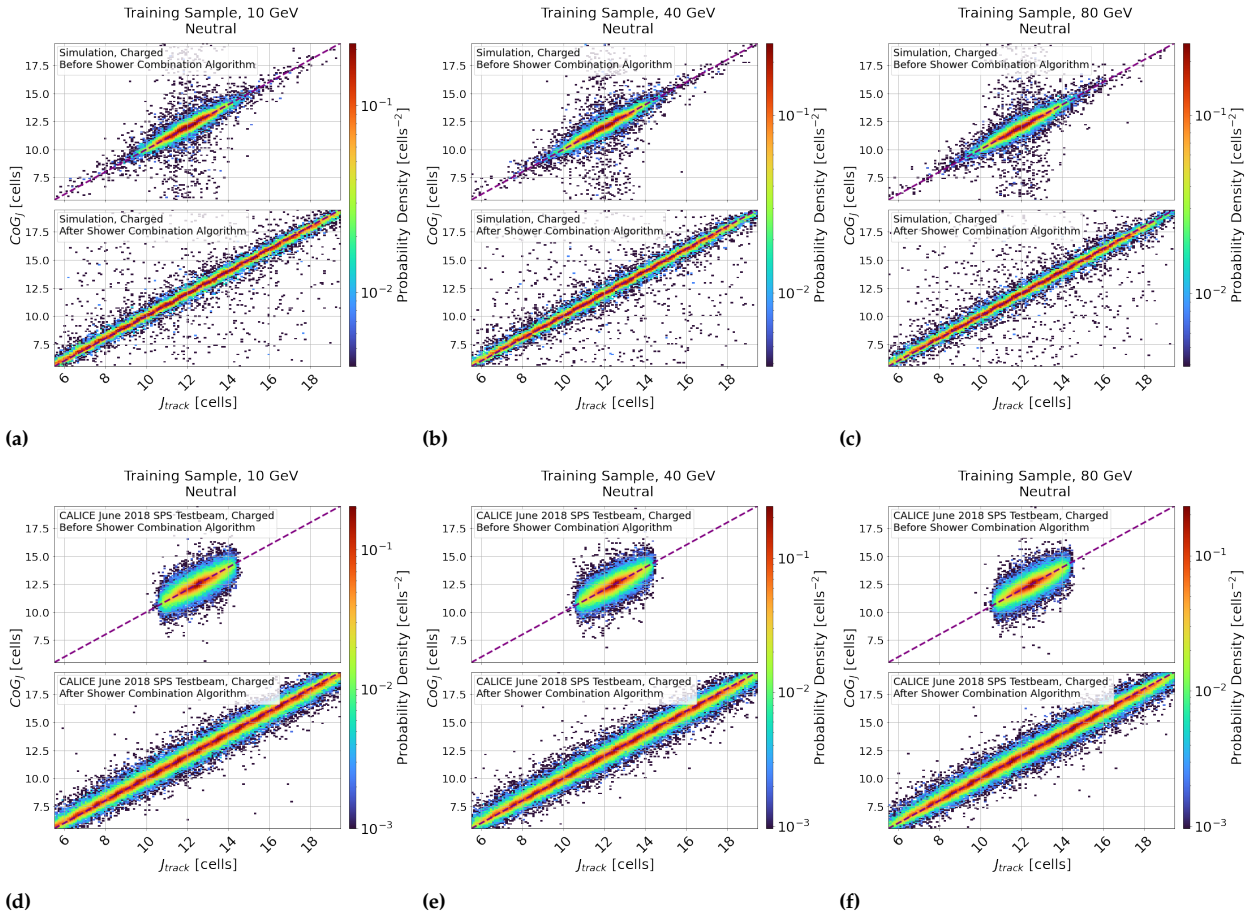


**Figure 6.23:** Figures 6.23a - 6.23c and Figures 6.23d - 6.23f show the joint distributions between  $\text{CoG}_I$  and  $I_{\text{track}}$ , before and after the shower combination has been performed. The top and the bottom subplot of each figure indicate the distribution before and after the combination, respectively. The purple dashed line indicates  $\text{CoG}_I = I_{\text{track}}$  where appropriate. The colour axis indicates probability density.

#### Comments:

- Figures 6.23a-6.23c and Figures 6.23d-6.23f indicate that the track position remains highly correlated with the centre-of-gravity after the shower separation algorithm was applied.

**Joint Distributions Lateral Centres-of-Gravity And Track Position in  $J$  of Synthetic Neutral After Application of the Shower Combination Algorithm**

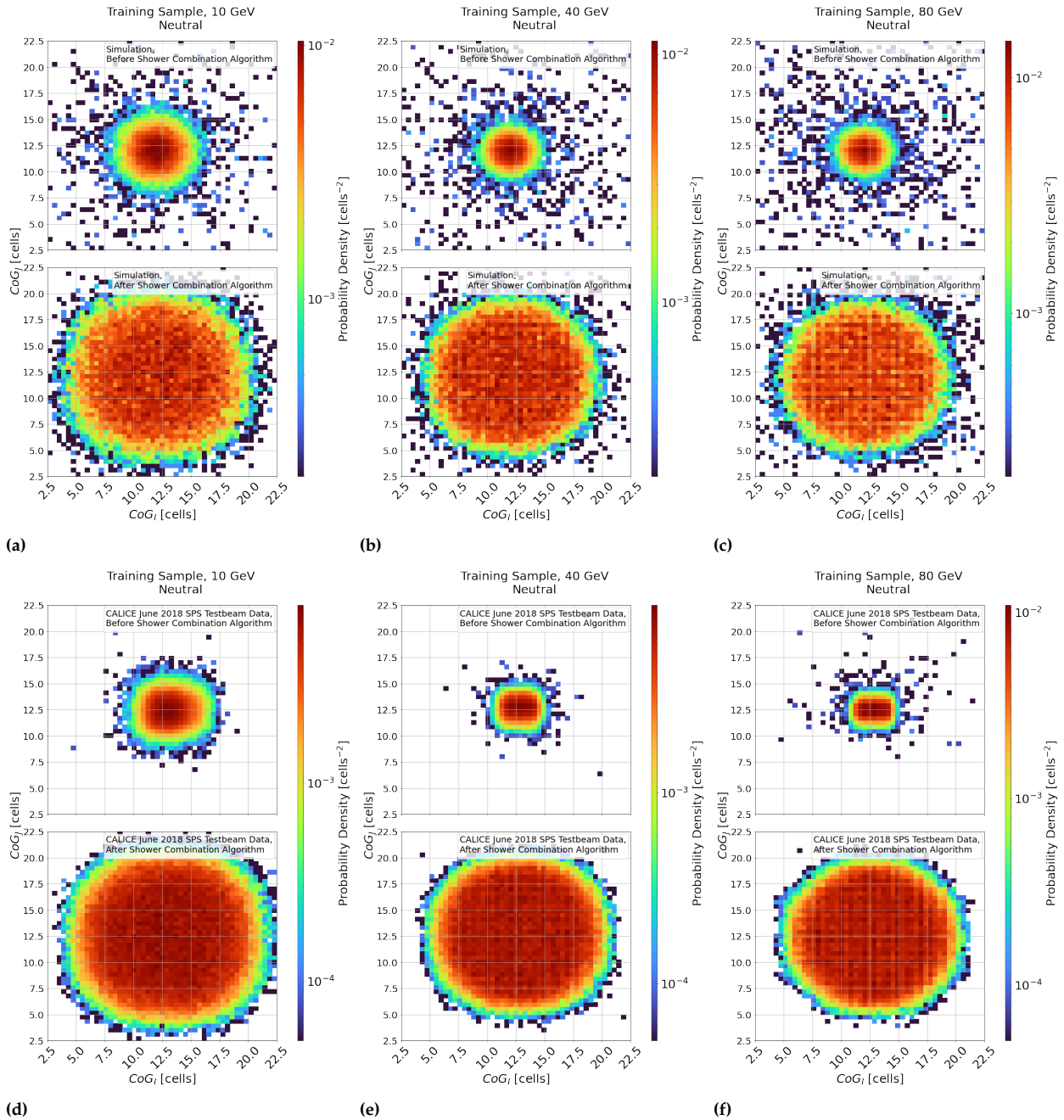


**Figure 6.24:** Figures 6.24a - 6.24c and Figures 6.24d - 6.24f show the joint distributions between  $CoG_J$  and  $J_{track}$ , before and after the shower combination has been performed. The top and the bottom subplot of each figure indicate the distribution before and after the combination, respectively. The purple dashed line indicates  $CoG_J = J_{track}$  where appropriate. The colour axis indicates probability density.

**Comments:**

- As in Figure 6.23.

### Joint Distributions Of Centres-of-Gravity in $I$ and $J$ of Synthetic Neutral After Application of the Shower Combination Algorithm



**Figure 6.25:** Figures 6.25a - 6.25c and Figures 6.25d - 6.25f show the joint distributions between  $CoG_I$  and  $CoG_J$ , before and after the shower combination has been performed, for simulation and data respectively. The top and the bottom subplot of each figure indicate the distribution before and after combination, respectively. The colour axis indicates probability density.

#### Comments:

- In Figure 6.25, it is demonstrated that the centre-of-gravity distribution of the synthetic hadron shower is broadened by the displacement of the hadron shower, by convolving the intrinsic distribution with a circle function. This indicates a broader range of possible positions for the hadron shower after the combination.



Parameter	PointNet, no Time	PointNet, + Time	GravNet, no Time	GravNet, + Time	DGCNN, no Time	DGCNN, + Time
Learning Rate	$2.567 \times 10^{-4}$	$5.681 \times 10^{-5}$	$2.012 \times 10^{-4}$	$5.169 \times 10^{-4}$	$1.252 \times 10^{-5}$	$1.660 \times 10^{-4}$
$p_{\text{dropout}}$	0.332	0.259	0.268	0.469	0.167	0.164
$\gamma$	-	-	8.137	12.815	-	-
$k$	-	-	16	24	15	18
$\beta_1$	0.9	0.9	0.9	0.9	0.9	0.9
$\beta_2$	0.99	0.99	0.99	0.99	0.99	0.99

**Table 6.4:** Table of hyperparameters used to train the neural network. In this table,  $\beta_1$  and  $\beta_2$  are the ADAM momentum parameters,  $p_{\text{dropout}}$  is the dropout probability, and  $k$  is the number of nearest-neighbours per cluster. Hyphens indicate hyperparameters that do not apply to the model. The parameters were informed by a hyperparameter scan using Optuna [166].

#### 6.2.4.2 Training

For simulation, two independent neural networks based on the model defined in Section 6.2.1 were trained on the training dataset: one without timing information and one with timing information for each model under test. For data, a single neural network with the best performance in the simulation was trained without timing information. The proposed compensation networks were developed in PyTorch [113] and trained using the PyTorch Lightning research framework [165] on an NVidia V100 GPU. The ADAM optimiser was used to improve the convergence rate for ten epochs. The hyperparameters used for training are shown in Table 6.4, selected based on the results of a parameter scan using Optuna hyperparameter optimisation framework [166], and shown in Table 6.4. It is noted that the  $\gamma$  parameter of GravNet was also varied as a hyperparameter.

The loss was chosen to be the same as in the study of [120]. This study applied a square-root energy-weighted mean square loss during training to encourage the models to correctly cluster the most energy-dense parts of the event.

It was noted in [120] that a spurious source of confusion was the ‘*shower-swapping*’. This was an effect where the neural network correctly clustered the showers but was compared to the wrong permutation of hadron showers during training and evaluation. This effect arises because the event comprises the sum of  $Q$  and  $N$  with a single track position and track energy in the example studied. The network has no inherent information to determine the correct order to output.

Assigning a fixed order to the output fractions without any meaningful basis from which the network can learn would be arbitrary when the study’s goal is to evaluate the energy clustering performance of the model, which specifically refers to the AHCAL detector’s ability to reconstruct the energy deposits accurately by characterising the shower structure and clustering them into distinct showers, not to label them as charged or neutral. The same argument is given by the authors of the study of [120]: ‘*since the labelling is irrelevant in a clustering problem, this behaviour (‘shower swapping’) is not a real inefficiency of the (shower clustering) algorithm*’.

In [120], while it is mentioned that the effect was accounted for in the model evaluation, it is ambiguous whether or not any attempt was made to resolve the problem during training. Therefore, adopting a permutation-invariant training regime to address this issue is a more suitable solution for the experiment performed in this chapter than only accounting for the effect after the fact. In this approach, the neural



network learns to predict the energy fractions without being constrained by a specific output ordering scheme.

Firstly, the loss function finds the permutation of the output that best matches the predicted and true energy fractions. It is presented in a way applicable for future studies involving more than two hadron showers of any permutation of  $Q$  or  $N$ , shown in Equation 6.6:

$$\mathcal{L}(\widehat{f}_{\text{hit}}; f_{\text{hit}}, E_{\text{hit}}) = \min_{H' \in \text{Sym}(H)} \sum_{i=0}^{\dim(H)} \frac{\sum_{\text{event}} \sqrt{E_{\text{hit}} \cdot f_{\text{hit}}^{H_i}} \cdot (\widehat{f}_{\text{hit}}^{H'_i} - f_{\text{hit}}^{H_i})^2}{\sum_{\text{event}} \sqrt{E_{\text{hit}} \cdot f_{\text{hit}}^{H_i}}} \quad (6.6)$$

where  $H$  denotes the set of possible hadrons in the event, in order  $H = \{Q, N\}$ , where  $Q$  and  $N$  are charged and neutral hadrons inducing showers in the AHCAL,  $H'$  denotes an element in the set of permutations of  $H$ ,  $\text{Sym}(H)$ , where  $\text{Sym}$  is the group of all permutations of the set  $H$ .

Secondly, during the evaluation, the combination of showers with the lowest loss was used to assign the predicted energy fractions to the ground truth showers.

It is nonetheless noted that two unavoidable biases arise in training in this analysis:

- ▶ *The Model Assumes A Charged and A Neutral Hadron Shower:* the shower separation studies only the case of a simultaneous charged and a neutral hadron shower event. The model is not designed to support arbitrary numbers of hadron showers;
- ▶ *The Model Cannot Explicitly Label An Output:* following the argumentation provided in this section and the study of [120], the bias of ambiguity of the output order must also be accepted in this study if 'shower-swapping' is to be eliminated as a source of spurious confusion.

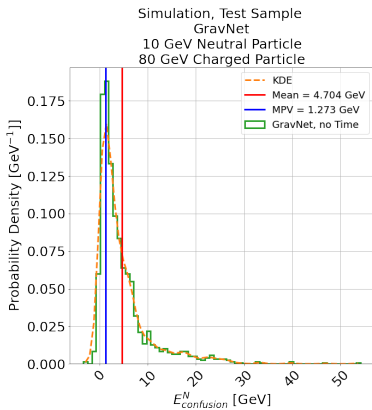
In summary, while the permutation-invariant training regime and the consideration of 'shower-swapping' address important challenges in the study, they also introduce certain limitations. It is important to recognize these limitations and interpret the results within the context of the specific event scenario and the inherent ambiguities of the clustering problem.

Future studies could explore more complex scenarios and models (e.g. a 'shower labelling' module to perform cluster-track association after shower separation) to further enhance the understanding of machine learning performance in a full Particle Flow analysis.

## 6.3 Results

In this section, the shower separation capabilities of each neutral network presented in Section 6.2.1 are presented and analysed.

14: RMS<sub>90</sub> is an observable used by the PF community to quantify the spread of the response of the calorimeter, oftentimes to quantify resolution. It is formally defined as the minimum standard deviation in the set of all possible central 90% percentile ranges allowed by the data in question (i.e. calculate the standard deviation in a sliding window, covering all percentile ranges between 0-90% to 10-100%, and select the range that produces the lowest standard deviation). It is used in preference to the standard deviation in PF due to the prevalence of extreme outliers when studying highly skewed distributions, or distributions with extreme outliers, which calls for greater robustness than the statistic allows. It is mentioned in [28] that: ‘.one should then not use the term ‘energy resolution’ for the results obtained in this way, and compare results obtained in terms of RMS<sub>90</sub> with genuine energy resolutions. This misleading practice is followed by the proponents of Particle Flow Analysis’. Arguably, a more robust measurement for spread achieving the same result is the median absolute deviation (MAD), for the same reasons as Chapter 5. RMS<sub>90</sub> is nonetheless presented for the consistency with other analyses.



**Figure 6.26:** Illustration of extracting the MPV from a highly skewed confusion energy distribution. The green line indicates the histogram of the confusion energy distribution, for GravNet, without time. The orange dashed line is the fitted KDE spline. The blue and red lines indicate the MPV and the mean, respectively.

The neutral hadron shower,  $N$ , is the reference shower. Confusion energy is therefore defined according to the difference between the true and predicted reconstructed calorimeter response and the Equation 6.7:

$$E_{\text{confusion}}^N(\widehat{E}_{\text{sum}}^N; E_{\text{sum}}^N) = \widehat{E}_{\text{sum}}^N - E_{\text{sum}}^N \quad (6.7)$$

The mean, MPV, RMS<sub>90</sub><sup>14</sup> MAD and medcouple of the confusion energy distributions are presented in this section to study the properties of the models up to the third statistical moment and therefore provide a broad overview of the reconstruction performance.

The MPV of confusion energy may be very different from the mean if its distribution is very heavily skewed, and allows additional discussion if there is a difference between the two. Therefore, this study is measured using a non-parametric density estimation technique called a *kernel density estimate* (KDE). The particular method used in this study convolves a Gaussian around the confusion energy measured in the event, with width given by a ‘bandwidth’ term,  $h$ , normalised to the number of events, as shown in Equation 6.8:

$$\widehat{f}_{\text{KDE}}(E; E_{\text{confusion}}^N, h) = \frac{1}{N_{\text{event}}} \sum_{\text{event} \in \text{Events}} \mathcal{N}\left(\frac{E - E_{\text{confusion}}^N}{h}\right) \quad (6.8)$$

where  $\widehat{f}_{\text{KDE}}$  is a continuous density estimate of the confusion distribution. The bandwidth parameter  $h$  is selected using Silverman’s Rule and the KDEpy [193] package and interpolated using a spline. The MPV is extracted by finding the root of the spline with the maximum probability density, similar to the method for initial gain estimation from SiPM charge spectra presented in Chapter 4. An example of the application of the method and the difference between MPV and mean is presented in Figure 6.26.

Finally, the energy reconstruction quality as measured by the spread (RMS<sub>90</sub>, MAD), while indicative, is not necessarily representative of the performance of the shower separation without considering its resolution. The precision of the energy of a particle cannot be greater than the width of the calorimeter response distribution caused by stochastic fluctuations at a particular particle energy. An alternative measurement of performance is therefore required.

$f_{\text{rec}}$  is therefore defined as the total number of events with confusion energy less than the threshold criterion, which is the calorimeter resolution at the particle energy of the neutral,  $\sigma_E$  as defined in Equation 5.5, divided by the total number of events in the sample. This formula provides a quantitative measure of how well the models reconstruct the energy of neutral hadrons separated from charged ones compared to the calorimeter’s energy resolution, shown in Equation 6.9:

$$f_{\text{rec}} = \frac{N_{|E_{\text{confusion}}^N| < \sigma_E}}{N_{\text{events}}} \quad (6.9)$$

$|E_{\text{confusion}}^N| < \sigma_E$  is satisfied, and  $N_{\text{events}}$  are the total number of events in a studied subsample of the test dataset.

$f_{\text{rec}}$  is used as the main figure of merit in for clustering performance in this study.

### 6.3.1 Simulation Models

In this section, the shower separation models described in Section 6.2.1 trained on simulation are evaluated.

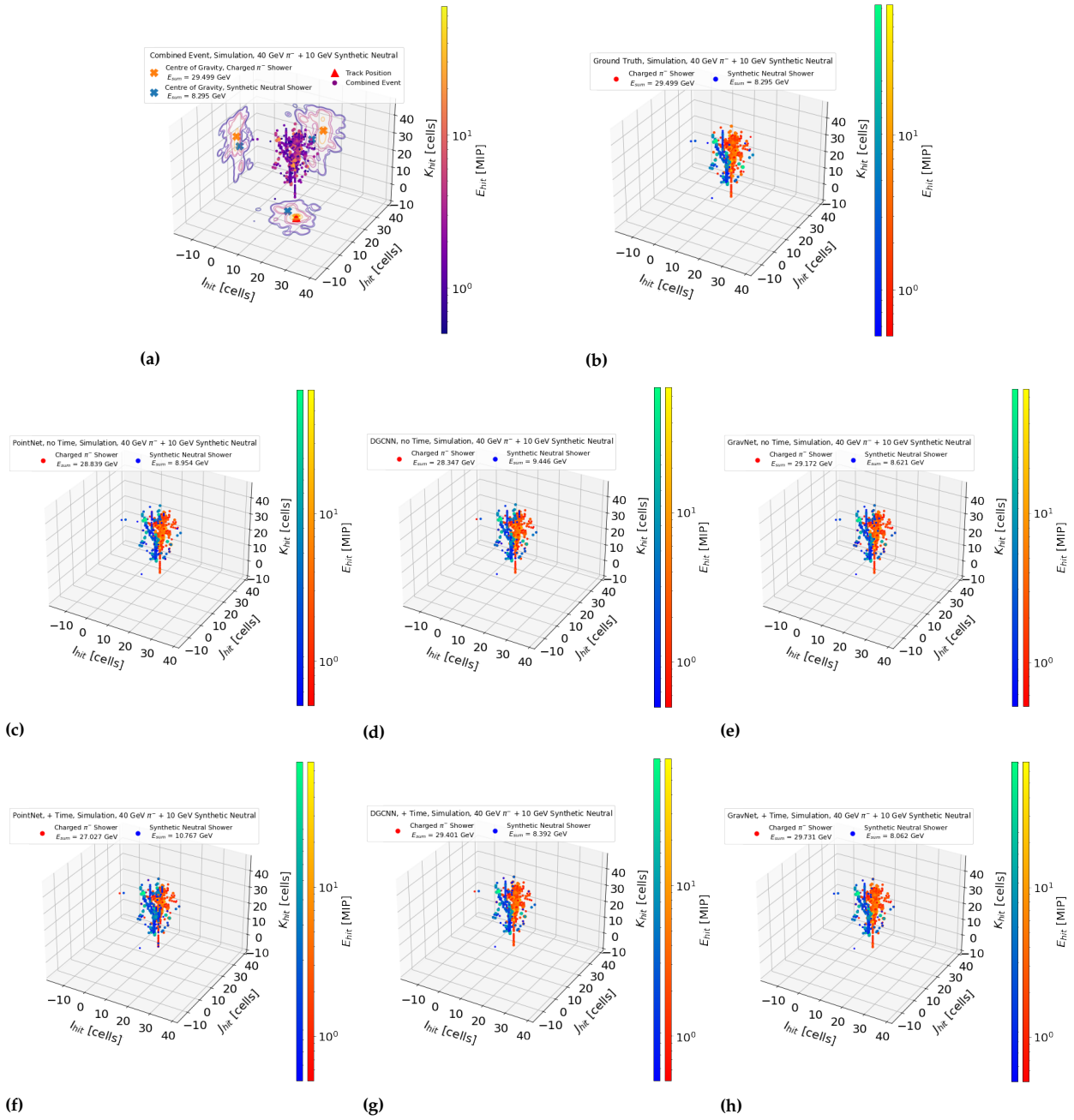
#### 6.3.1.1 Example Event Displays

Example event displays of three events, separating a 10 GeV, 40 GeV and 80 GeV synthetic neutral hadron shower from a 40 GeV charged hadron shower are shown in Figures 6.27-6.29.

#### Summary Comment to Figures 6.27-6.29

The models are demonstrated to qualitatively reconstruct shower events as expected. The showers are demonstrated to be compared in the same permutation as the original event, indicating that the evaluation can be performed with minimal contribution from 'shower-swapping' confusion.

**Example Event Displays of a 10 GeV Synthetic Neutral and a 40 GeV Charged Hadron Shower, After the Application of Each Shower Separation Model Under Test, Applied To The Test Sample of Simulation**

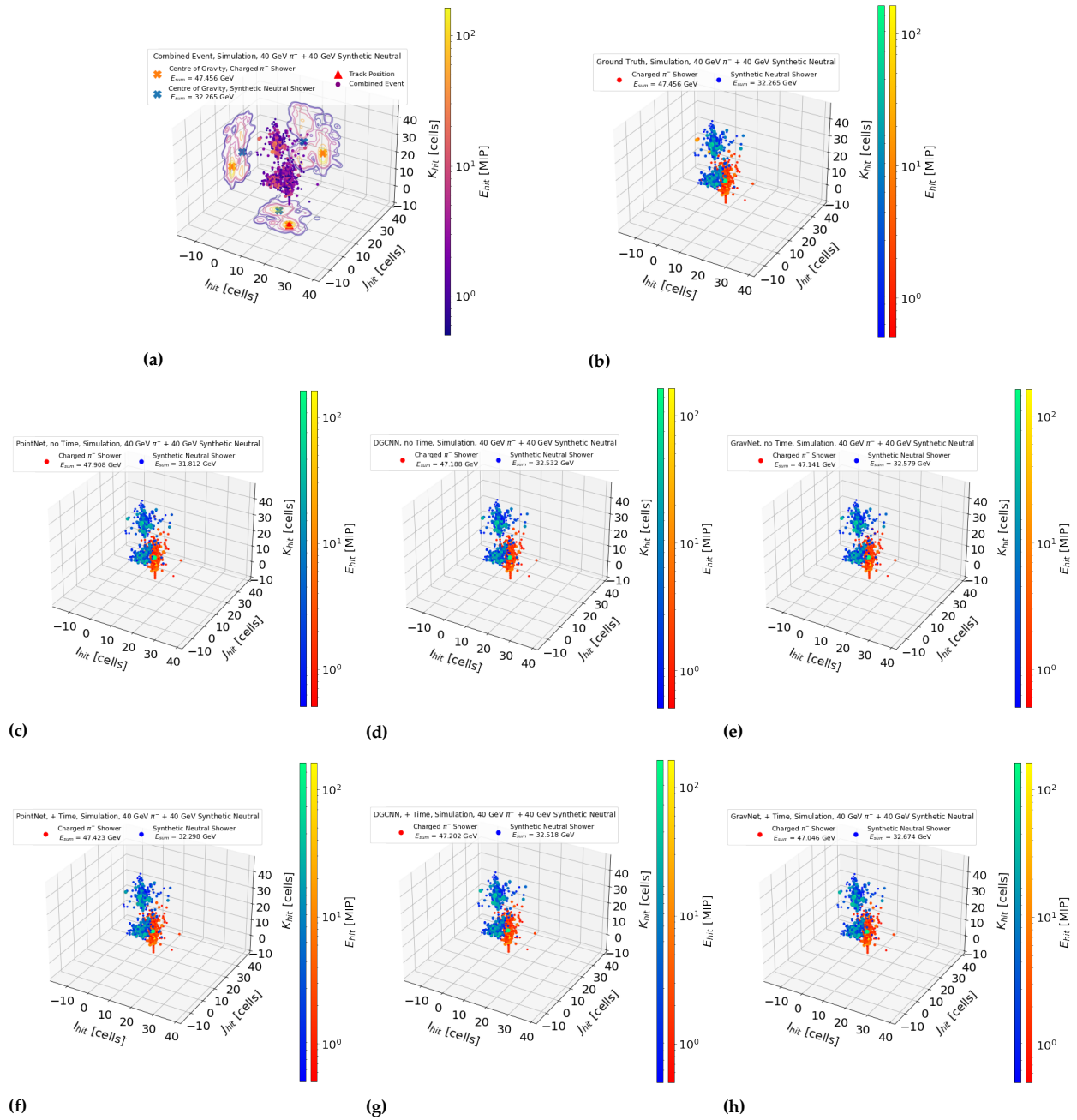


**Figure 6.27:** Example event displays illustrating the application of the trained algorithms to a test charged and synthetic neutral hadron shower event. Figure 6.27a shows the initial combined event, as in Figure 6.20d. Figure 6.27b shows the original event to be reconstructed, as in Figure 6.19d. Figures 6.27c-6.27e and Figures 6.27f-6.27h indicate the reconstructed event obtained by applying PointNet, DGCNN and GravNet models as described in Section 6.2.1 (left to right) for the models trained without (middle row) and with timing information (bottom row). Else, as in Figure 6.19 and Figure 6.20.

**Comments:**

- Figure 6.19. Figures 6.27c-6.27e and Figures 6.27f-6.27h with Figure 6.19 indicate the models under test are qualitatively able to resolve the chosen hadron shower event, indicated by their agreement.
- The agreement of the colours indicates that the best output channel permutation has been selected for study, meaning that the contribution of confusion can be studied with no possibility of ‘shower swapping’.

### Example Event Displays of a 40 GeV Synthetic Neutral and a 40 GeV Charged Hadron Shower, After the Application of Each Shower Separation Model Under Test, Applied To The Test Sample of Simulation

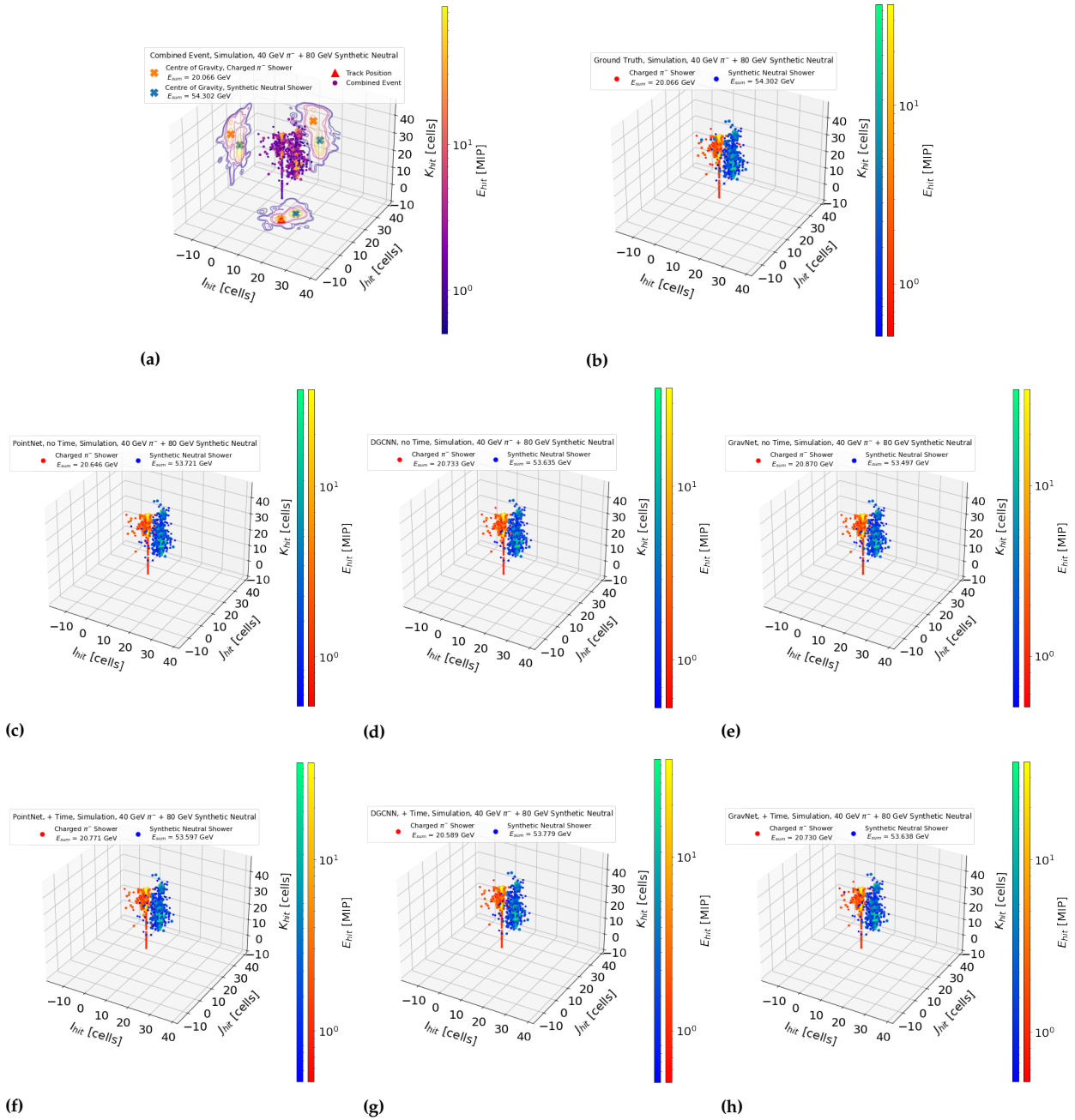


**Figure 6.28:** Example event displays illustrating the application of the trained algorithms to a test charged and synthetic neutral hadron shower event. Figure 6.28a shows the initial combined event, as in Figure 6.20e. Figure 6.28b shows the original event to be reconstructed, as in Figure 6.19e. Figures 6.28c-6.28e and Figures 6.28f-6.28h indicate the reconstructed event obtained by applying PointNet, DGCNN and GravNet models as described in Section 6.2.1 (left to right) for the models trained without (middle row) and with timing information (bottom row). Else, as in Figure 6.19 and Figure 6.20.

#### Comments:

- As in Figure 6.28.

**Example Event Displays of a 80 GeV Synthetic Neutral and a 40 GeV Charged Hadron Shower, After the Application of Each Shower Separation Model Under Test, Applied To The Test Sample of Simulation**



**Figure 6.29:** Example event displays illustrating the application of the trained algorithms to a test charged and synthetic neutral hadron shower event. Figure 6.29a shows the initial combined event, as in Figure 6.20f. Figure 6.29b shows the original event to be reconstructed, as in Figure 6.19f. Figures 6.29c-6.29e and Figures 6.29f-6.29h indicate the reconstructed event obtained by applying PointNet, DGCNN and GravNet models as described in Section 6.2.1 (left to right) for the models trained without (middle row) and with timing information (bottom row). Else, as in Figure 6.19 and Figure 6.20.

**Comments:**

- As in Figure 6.28.



### 6.3.1.2 Linearity of Response

The distributions of the reconstructed hadron shower energy are presented as a function of the original shower energy for all implemented shower separation models under test and are shown in Figure 6.31. The linearity of response for the reconstructed neutral is measured using the mean and the MPV, as defined, of the confusion energy distribution as defined in Equation 6.7 shown in Figure 6.31.

#### Summary Comment to Figure 6.31

All neutral networks frequently reconstruct neutral hadron shower events accurately, which is indicated by the MPV of the distribution being well within 1-1.5 GeV for the test dataset. However, a significantly more biased mean indicates energy-dependent skewness, which was also observed in [120]. Further analysis is required to understand this result and is discussed in Section 6.3.1.7. A clear improvement in reconstruction quality is observed from the inclusion of timing information in the model for DGCNN and GravNet.

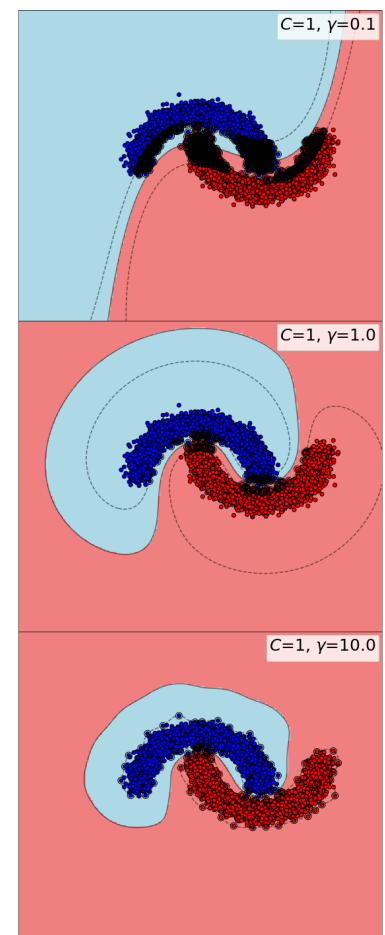
### 6.3.1.3 Spread of Confusion Energy

The complete distributions of the reconstructed hadron shower energy are presented as a function of the original shower energy for all implemented shower separation models under test and are shown in Figure 6.31, with the  $RMS_{90}$  and MAD of each distribution shown.

#### Summary Comment to Figure 6.32

For the graph neural networks (DGCNN and GravNet), a significant improvement in resolution due to the inclusion of timing information was observed. For the best-performing neural network (GravNet) this corresponded to a reduction of the MAD by around 23%. By contrast, PointNet did not observe an improvement due to timing information.

The failure of PointNet compared to the considerable success of DGCNN and GravNet to exploit the timing information available from AHCAL indicates a physical result. It is mentioned in [163] that: *'Instead of working on individual points like PointNet.. we exploit local geometric structures by constructing a local neighbourhood graph and applying convolution-like operations, on the edges connecting neighbouring pairs of points, in the spirit of graph neural networks.'* Therefore, the use of 'local geometry' (i.e. 'local' energy density) summarises the fundamental difference between the PointNet and DGCNN and GravNet models. It may therefore be tentatively concluded that the reason for the improvement of DGCNN and GravNet compared to PointNet is that timing information improves 'local' energy density clustering. Corroborating the above statement, it is noted from Table 6.4 that the optimised  $\gamma$  parameter for the GravNet network is much larger with timing information. As previously mentioned,  $\gamma$  indicates the 'strength' of the potential in GravNet, which is related to the significance of local energy density in the model. Therefore, including timing information may directly increase the 'locality' of the GravNet



**Figure 6.30:** Support Vector Machine classifier using a 'radial basis function' similar to GravNet, for different values of the  $\gamma$  'potential strength' parameter with a constant regularisation parameter,  $C$ . The red and blue points indicate data to be classified, and the blue and red regions indicate the decision regions, the black line indicates the boundary and the dashed lines indicate the  $\pm\sigma$  margin. The effect of increasing gamma is a more 'local' classifier, which is indicated by the shrinking of the decision boundary around the blue points. Modified from [194].

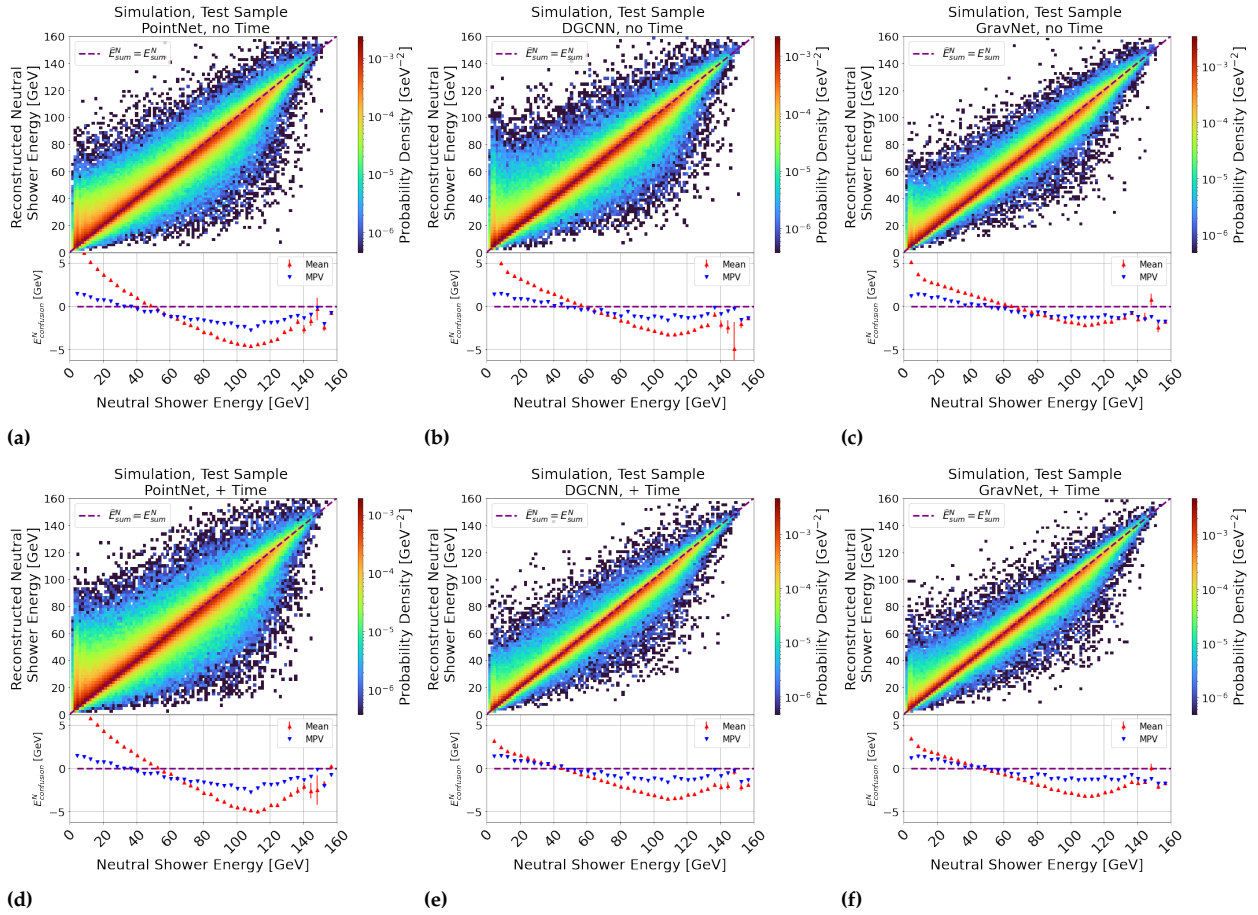
model operating under optimal conditions. This is demonstrated visually in Figure 6.30 using a support vector machine, a type of classifier that uses a 'radial basis function' kernel similar to the 'potential' function used in GravNet, illustrates the increasing 'locality' of the classifier.

The hypothesis is presently challenging to verify without a standard aggregate measure of the amount of 'substructure' or 'locality' in a hadron shower. The locality would refer to the number of locally varying energy deposits, such as sub-showers, tracks, decaying particles, etc. Further research into a 'N-subshoweriness' observable, analogous to 'N-subjettiness'<sup>15</sup>, which would be beneficial for this purpose. Preliminary studies have investigated event reconstruction methods involving neural networks and cluster counting using unsupervised clustering techniques [196]. However, these techniques have yet to be fully verified and remain conjecture.

15: 'N-subjettiness' is an observable used in jet reconstruction to determine the agreement of a jet with having multiple 'cores' or 'prongs' associated with the jet, which are associated with electroweak boson decays (i.e.  $W^\pm/Z \rightarrow q\bar{q}$ ) or top-quark decays. By contrast, QCD jets are expected to only have one core. This quantity has applications in jet-tagging [195].



### Joint Distribution between Predicted and True Reconstructed Neutral Shower Response of the Test Sample of Simulation for Each Shower Separation Model Under Test

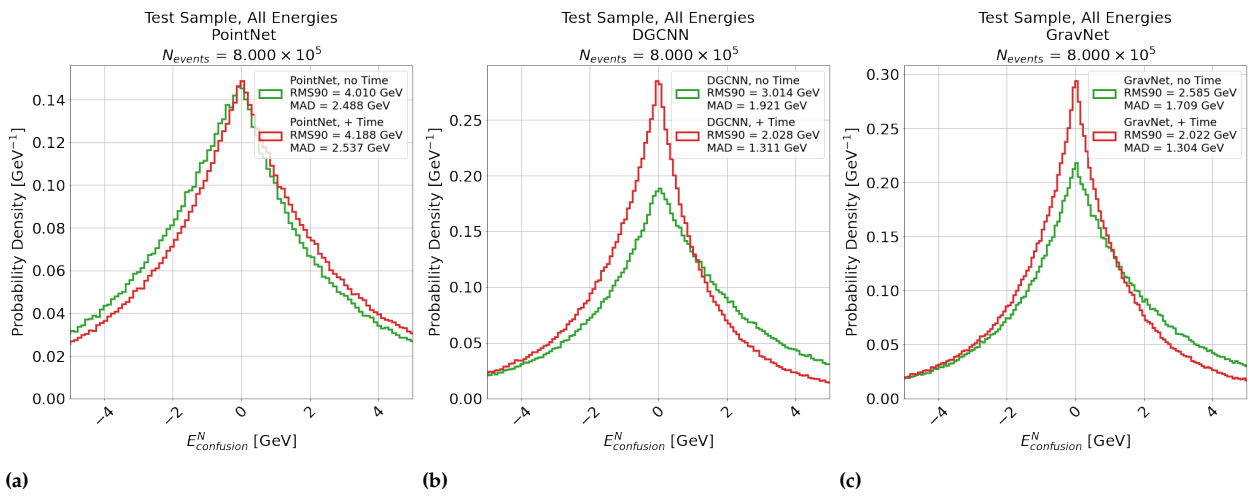


**Figure 6.31:** Figures 6.31a -6.31b (top row) Figures 6.31d -6.31e show the joint distributions of the predicted and true reconstructed neutral shower response for each shower separation model under test, without and with timing information, respectively. The colour axis indicates probability density. The purple dashed line indicates perfect reconstruction. The bottom subplot shows the mean and MPV on the  $y$ -axis at each bin along the  $x$ -axis.

#### Comments:

- The MPV and the red region indicated on the histograms indicate, in general, that the models learn to frequently reconstruct the neutral shower with nearly the same energy as the original. The MPV differs by no more than 1.5 GeV, and typically varies much less than that.
- The mean and green region in the figures indicate a bias. This means that, for all models, the shower energy of neutrals tends to be overestimated below 60 GeV and underestimated above that value. The bias is observed for all models under test and is indicated by the larger slope of the mean compared to the MPV of each distribution.
- Comparison of the top row with the bottom row indicates that DGCNN and GravNet show significant improvements in the reconstruction of the neutral hadron shower for DGCNN and GravNet with the inclusion of 100 ps time resolution, indicated by the narrowing of the distributions around the purple line. By contrast, PointNet shows no improvement.

### Distributions of Confusion Energy for the Entire Test Sample of Simulation, With and Without Timing Information, For Each Shower Separation Model Under Test



**Figure 6.32:** Figures 6.32a -6.32c show the distributions of the neutral confusion energy for each shower separation model under test. The green and red lines indicate the same models, without and with timing information, respectively. RMS<sub>90</sub> and MAD are shown in the legend for each model.

#### Comments:

- Comparison of Figure 6.32a and Figures 6.32b-6.32c indicate that the PointNet model produces similar performance for reconstructing hadron showers whether or not timing information is included. By contrast, the DGCNN and GravNet models show significantly better resolution overall than PointNet, and also show significant improvements using timing information.
- The best-performing model both with and without timing information was GravNet by a small margin. GravNet was found to achieve a better result without timing information and produced overall a similar result to DGCNN with timing information. This corresponded to a 23 % reduction in the MAD of the distribution.

### 6.3.1.4 Fraction of Events Reconstructed Within Calorimeter Resolution vs. Charged/Neutral Particle Combination

The fraction of events reconstructed within the calorimeter resolution as described in Equation 6.9 as a function of the combination of charged and neutral particle energy are shown in Figure 6.34. Sample confusion energy distributions are shown for the best-performing network, GravNet, in Figure 6.35, and for PointNet and DGCNN in Appendix Figures 8.23 and Figure 8.24. The resolution was obtained from the study of Chapter 5 presented in Table 5.3 of  $R_{\text{sim}} = 49.5\% / \sqrt{E_{\text{particle}}} \oplus 7.2\%$ .

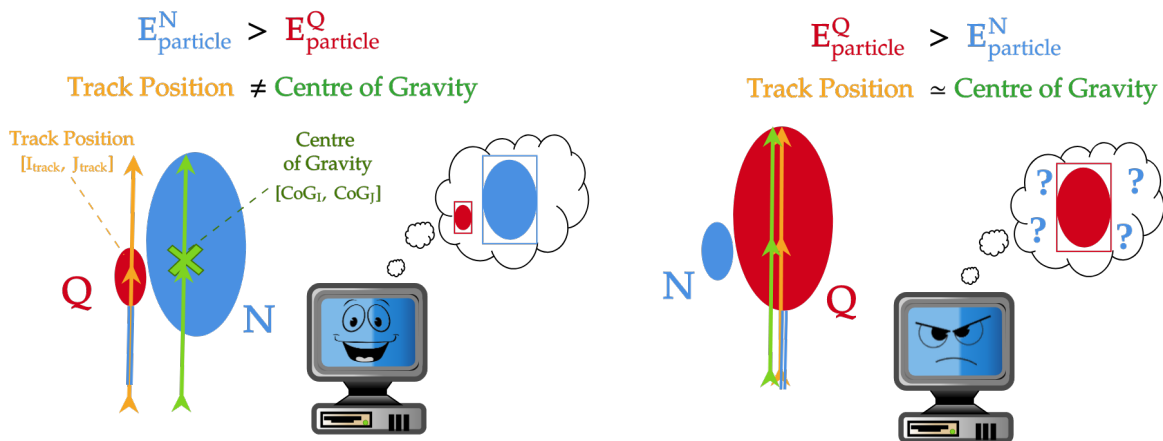
#### Summary Comment to Figure 6.34

Above 80-90 % of events are reconstructed within the calorimeter resolution where the neutral hadron energy is greater than the charged hadron energy. The performance degrades for the opposite case. Timing information is observed to result in a 15-20 % improvement in the number of events reconstructed within the resolution for the DGCNN and GravNet models, with no improvement for the PointNet model in the case where the charged shower energy is greater than the neutral. No strong improvement was observed in the opposite case.

A plausible hypothesis for this observation is that the track information is employed by the models. If the neutral hadron has more energy than the charged hadron, the centre-of-gravity (dominated by the neutral hadron) is sufficiently displaced from the track position that both axes can be used for clustering. In the opposite case, the centre-of-gravity (dominated by the charged hadron) is similar to the track position, and therefore less information is gained as to the position of the other hadron shower in the opposite case. The concept is illustrated in Figure 6.33.

An ideal test for this hypothesis would involve training the models with a combination of double neutral shower pairs and charged-neutral pairs, and evaluating the differences in performance of the two possible event configurations. However, this study is beyond the scope of the study and is left to future research.

A weaker experiment is nonetheless performed to probe the hypothesis. A subsample of 10 GeV synthetic neutrals, separated at a distance greater than  $12 \rho_M$  for charged showers of any other energy in the sample was selected. At this distance, no less than around 75-80 % of 10 GeV hadron showers are correctly reconstructed within the calorimeter resolution (see Figure 6.37g) regardless of their energy, by comparison with Figure 6.17. The distance of the centre-of-gravity of the entire two-shower event to the charged track position was calculated. It is expected that the distance between the track and the lateral centre-of-gravity ought to decrease as the charged shower energy increases. Then, the mean absolute confusion energy achieved by GravNet, with and without time, as a percentage of the true shower energy, was calculated as a function of both distance and charged shower energy. Suppose the confusion increases as the distance between the centre-of-gravity and the track position decreases. In that case, it may be concluded that these variables are correlated with one



**Figure 6.33:** Illustration of the hypothesis made for the asymmetry in reconstruction performance depending on the charged-neutral energy difference. Charged and neutral hadron showers are indicated as red and blue ovals, and the orange and green lines indicate the centre-of-gravity and charged track of the event. The diagram on the left indicates the case where the neutral hadron shower has more energy than the charged. In this case, the track is sufficiently displaced from the centre of gravity that both may be used as clustering axes. By contrast, the diagram on the right indicates the case where the charged shower has more energy than the neutral shower. The track and centre-of-gravity are sufficiently similar that little advantage to separation comes with its inclusion by comparison.

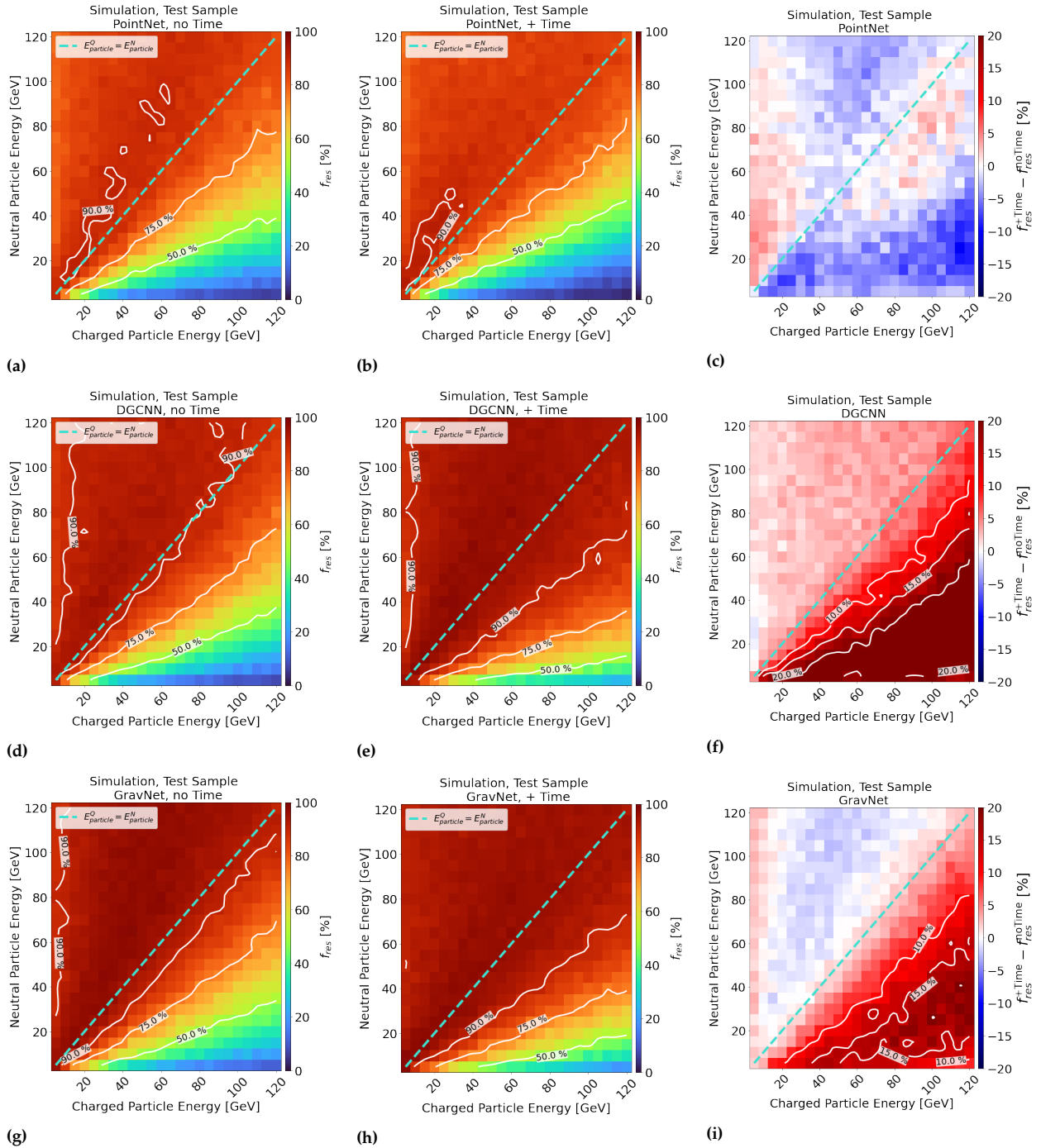
another and is indicative of the hypothesis made. The correlations are shown in Figure 6.36.

#### Summary Comment To Figure 6.36

It may be concluded that the charged shower energy, the distance between the track position and centre-of-gravity of the two-shower event and the proportion of confusion energy are strongly correlated for 10 GeV neutrals separated at lateral distances greater than  $10 \rho_M$ , which is in agreement with the observations of Figure 6.34. The improvement due to timing information also equalises as the charged shower energy decreases (distance between track and centre-of-gravity increases), which is also in agreement with the results of Figure 6.34. These observations cannot be explained by overlapping showers due to the constraint on the shower distance.

Therefore, it may be concluded that the hypothesis remains compatible with observations, even when the possibility of overlapping showers is excluded. Nonetheless, further study as suggested is required to fully interpret the results obtained in this section.

### Matrices of the Fraction of Events Reconstructed Within Calorimeter Resolution for Each Shower Separation Model Under Test Applied To The Test Sample of Simulation

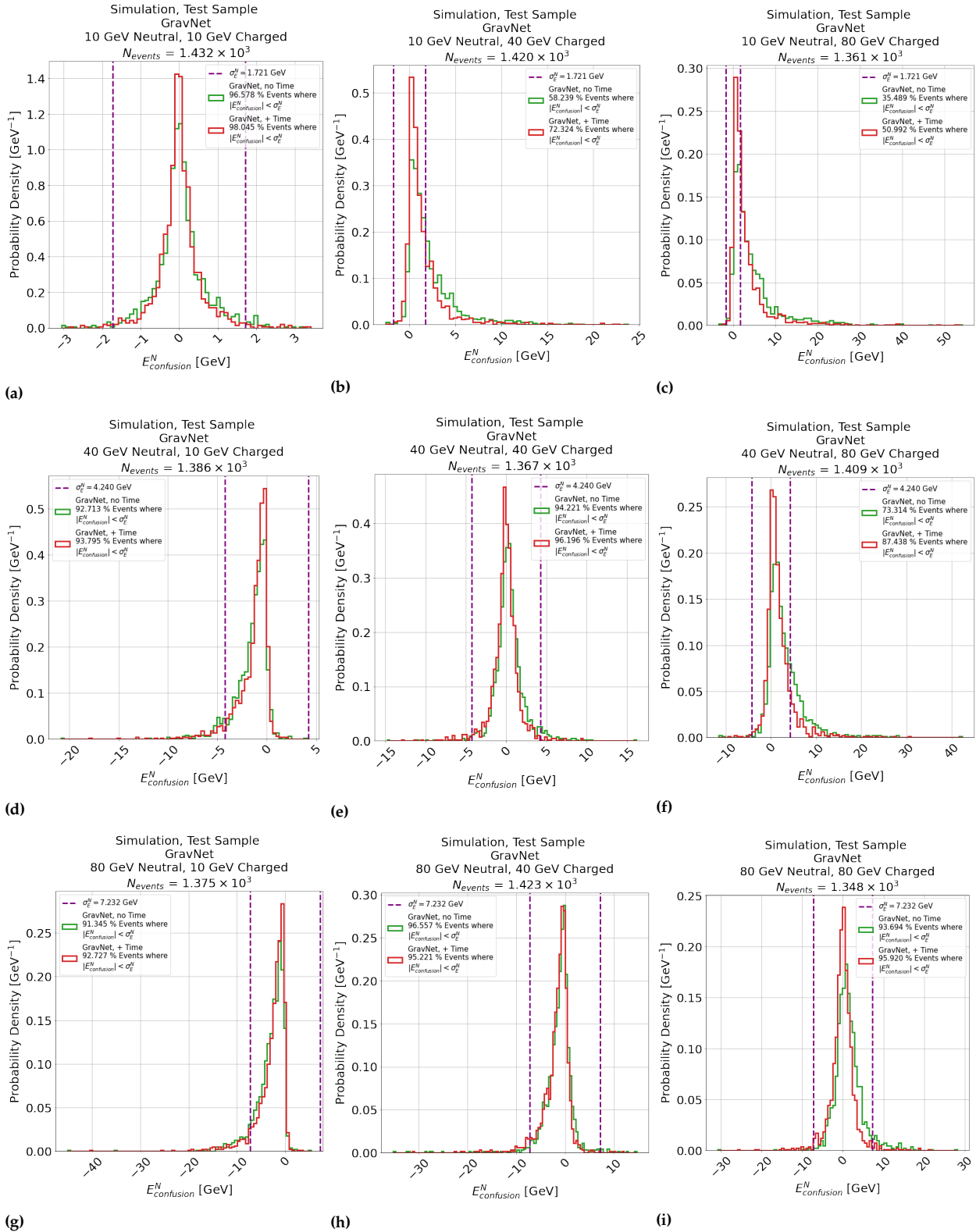


**Figure 6.34:** Figures 6.34a, 6.34d 6.34g and Figures 6.34b, 6.34e 6.34h show the matrices of the fraction of events with energy reconstructed within the calorimeter resolution as a function of the charged and neutral particle energy, for PointNet, DGCNN and GravNet, respectively, where red to blue indicates a higher to lower percentage of events. The turquoise dashed line indicates  $E_{particle}^Q = E_{particle}^N$ , while white lines indicate contours. Figures 6.34c, 6.34f 6.34i indicate the ratios of the fractions for models trained with and without timing information.

#### Comments:

- The leftmost column (Figures 6.34a, 6.34d and 6.34g) and middle column (Figures 6.34b, 6.34e and 6.34h) indicate that there is an asymmetry in shower reconstruction performance depending on whether or not the charged particle energy is greater than the neutral particle energy. Where the neutral hadron had more energy than the charged hadron, above 80 % (90 % for DGCNN and GravNet) of events are reconstructed within the calorimeter resolution. By contrast, performance degrades in the opposite case.
- The rightmost column (Figures 6.34c, 6.34f and 6.34i) indicates that for DGCNN and GravNet, improvements of up to an additional 15-20 % of events by the use of timing information is observed for the case where the charged hadron has more energy than the neutral hadron, indicated by the red region below the equality line, while in the opposite case, no strong improvement is observed.

### Distributions of Confusion Energy for Possible Combinations of Charged and Neutral Hadron Particle Energies at 10 GeV, 40 GeV and 80 GeV For All Shower Separation Models Under Test, Applied To The Test Sample of Simulation

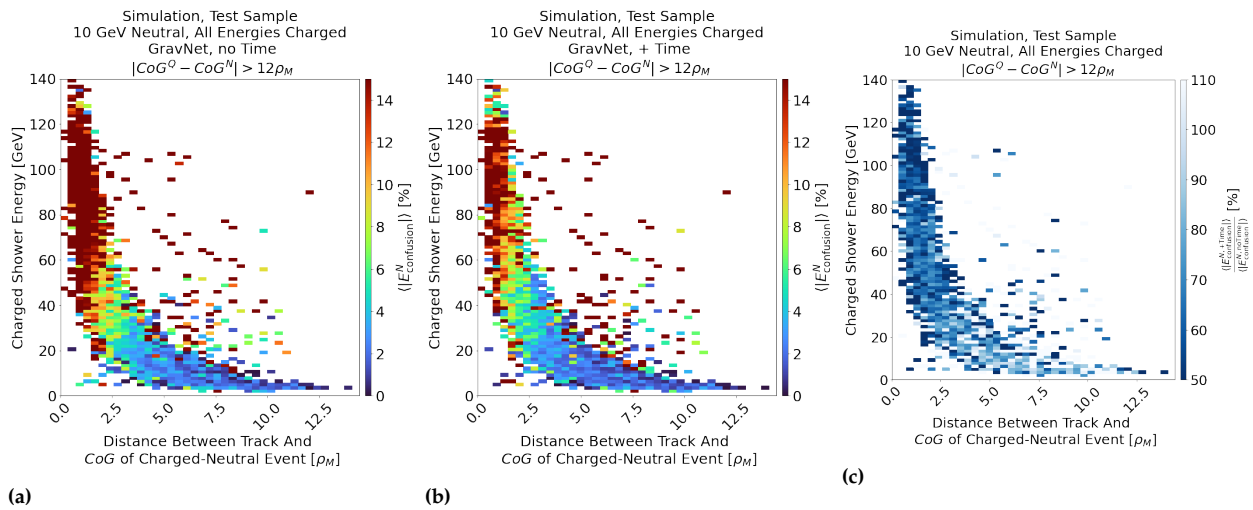


**Figure 6.35:** Distributions of confusion energy for the GravNet network applied to the test sample. The green and red lines indicate the models trained without and with timing information, respectively. The purple dashed lines indicate the resolution of the AHCAL calorimeter in simulation.

#### Comments:

- ▶ A correlation between left and right-tailed skewness in confusion energy is observed as the disparity between hadron shower energies increase. This is discussed in Section 6.3.1.7.
- ▶ In all cases, an improvement in energy reconstruction due to the inclusion of timing information is observed, indicated by the greater proportion of events falling within the purple dashed lines.

**Correlation Between Charged Shower Energy, Distance of Track to Charged and Neutral Shower Centre of Gravity and Mean Confusion Energy from GravNet, With and Without Timing Information, for 10 GeV Neutral Showers Displaced From Charged Showers, Both from the Test Sample of Simulation, By Distances Of Greater Than  $12\rho_M$**



**Figure 6.36:** Figures 6.36a-6.36b show the joint correlation between charged shower energy and the distance between the position of the charged track and centre-of-gravity of the entire event on the  $y$ -axis and  $x$ -axis, respectively. The colour axis shows mean absolute percentage of confusion energy in percent of the shower energy for the GravNet model. Events separated by more than  $12\rho_M$  are shown. Figure 6.36a shows the model trained without timing information and Figure 6.36b with timing information. Figure 6.36c shows the ratio of the model's performance with and without timing information, indicated in percent on the colour axis. Blue to white on this plot means superior to no improvement in confusion.

**Comments:**

- Figures 6.36a-6.36b the banana-shaped distribution indicates the expected behaviour that, as the charged shower carries a higher proportion of the energy of the combined event, the distance between the centre of gravity of the whole event and the track position decreases, and vice versa.
- The gradient from red to blue in these plots indicates that as the charged shower energy decreases and the distance between the centre-of-gravity and track increases, the model's performance also improves. It may be concluded that model performance, the distance of track to centre-of-gravity and the proportion of confusion energy are all correlated and in agreement with the results of Figure 6.34.
- Figure 6.36c indicates that each model's improvement in confusion energy equalises as the shower energy decreases. This means that the relevance of timing information is also correlated to the charged shower energy and the distance between centre-of-gravity and track, and is also in agreement with the observations of Figure 6.34.



### 6.3.1.5 Fraction of Events Reconstructed Within Calorimeter Resolution vs. Lateral Shower Distance

As in Section 6.3.1.4, the fraction of reconstructed events within the calorimeter resolution was determined as a function of lateral shower distance for each model under test. The results are shown in Figure 6.37. The number of correctly reconstructed events is expected to saturate with separation distance.

#### Summary Comment to Figure 6.37

As the lateral distance between hadron showers increases, the fraction of reconstructed events also increases and saturates after around 10-15  $\rho_M$  for all studied energies. Including timing information increased  $f_{\text{rec}}$  up to 10% for the DGCNN and GravNet networks. There is limited evidence to support the hypothesis that the improvement of the performance of the GravNet and DGCNN models with timing is correlated with the lateral shower distance.

A similar study measuring the performance of Pandora PFA as a function of the inter-shower distance between charged-neutral shower pairs was performed in [78]. Specifically, 10 GeV synthetic neutrals showers were combined with 10 GeV and 30 GeV charged hadron showers produced from  $\pi^-$  hadrons simulated using QGSB\_BERT\_HP in the same way as described in this chapter. Good agreement is observed comparing the energy profile distributions of the samples used in both studies is presented in Figure 8.28. In [78], the definition of  $f_{\text{rec}}$  as defined in this thesis, in Equation 6.9, uses  $1.5\sigma_E$  instead of  $1\sigma_E$  for the performance<sup>16</sup>. Therefore,  $f_{\text{rec}}^{1.5\sigma} = N_{|E_{\text{confusion}}^N| < 1.5\sigma_E} / N_{\text{events}}$  is defined for this measurement with the same variables defined as in Equation 6.9.

A comparison is presented as a clustering algorithm's lateral shower separation capabilities are crucial for assessing its performance. As a caveat, Pandora PFA does not assume the presence of two showers in the event, whereas this is true for the trained neural networks. This means that an unavoidable bias is introduced. Therefore, directly comparing the methods is impossible without a full dedicated study. Nonetheless, a comparison is indicative of potential improvements over Pandora PFA using machine learning. The results are shown in Figure 6.38.

16: It was confirmed with the author of [78] that the definition of the quantity called  $f_{\text{rec}}$  in this thesis is equivalent to the quantity  $P_{2\sigma}$ , so named to indicate coverage of the central  $2\sigma_E$  region (i.e.  $\pm 1\sigma_E$ ) of the confusion energy distribution.  $P_{3\sigma}$ , therefore, indicates the central  $3\sigma_E$  region i.e.  $\pm 1.5\sigma_E$ .



**Summary Comment to Figure 6.38**

The models tested, DGCNN and GravNet, significantly outperformed Pandora PFA in shower separation, according to the study of [76]. For 10 GeV charged showers and 10 GeV synthetic neutral hadron showers, DGCNN and GravNet achieved up to a 25 % increase in reconstructed events within the calorimeter resolution, regardless of timing information. PointNet performed better than Pandora PFA but not as well as the other models. For 30 GeV charged showers and 10 GeV synthetic neutral hadron showers, GravNet and DGCNN showed even greater improvements, reaching up to 30 % without timing information and 35 % to 45 % with timing information. PointNet performed similarly to Pandora PFA in this case.

These results indicate a promising improvement compared to the standard method of shower separation applied in Particle Flow, notwithstanding the specified caveats.

### 6.3.1.6 Fraction of Events Reconstructed Within Calorimeter Resolution vs. Longitudinal Shower Distance

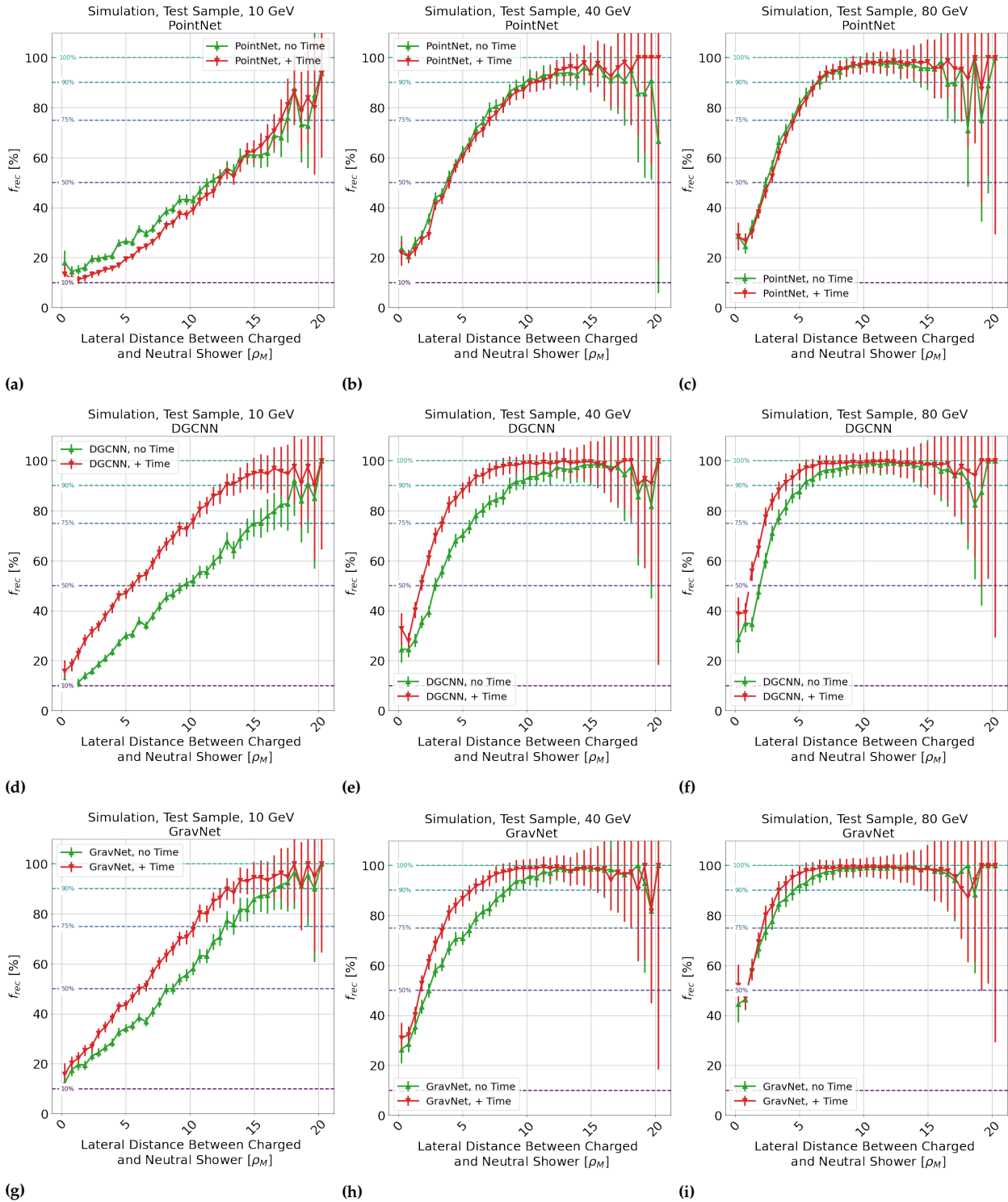
As in Section 6.3.1.4, the fraction of reconstructed events within the calorimeter resolution was determined as a function of longitudinal shower distance for showers within  $5 \rho_M$  of one another, for each model under test, measured as the difference between the shower starting positions in nuclear interaction lengths for ‘closeby’ hadron showers. This study allows analysis of whether or not the performance increase due to timing information is correlated with the difference in the shower starting positions. This would imply that clustering in time improves the ability to resolve longitudinally displaced energy deposits. The results are shown in Figure 6.39.

**Summary Comment to Figure 6.39**

The performance of the models increases as the difference between shower starts increases, with less of an effect the greater the particle energy is. There is limited evidence to support the hypothesis that the improvement of the performance of the GravNet and DGCNN models due to time is correlated with longitudinal distance.

Therefore, the improvement due to timing information for DGCNN and GravNet affects all hadron showers, with no strong dependence on the particular distribution of the hadron showers in the event. This is compatible with the hypothesis that the use of timing information to hadron shower clustering is a ‘local’ phenomenon that requires more complex, locally-defined observables to understand fully.

**Mean Fraction of Events Reconstructed Within Calorimeter Resolution vs. Lateral Shower Distance for 10 GeV, 40 GeV and 80 GeV Synthetic Neutrals Separated From 5-120 GeV Charged Showers for Each Shower Separation Model Under Test Applied To The Test Sample of Simulation**

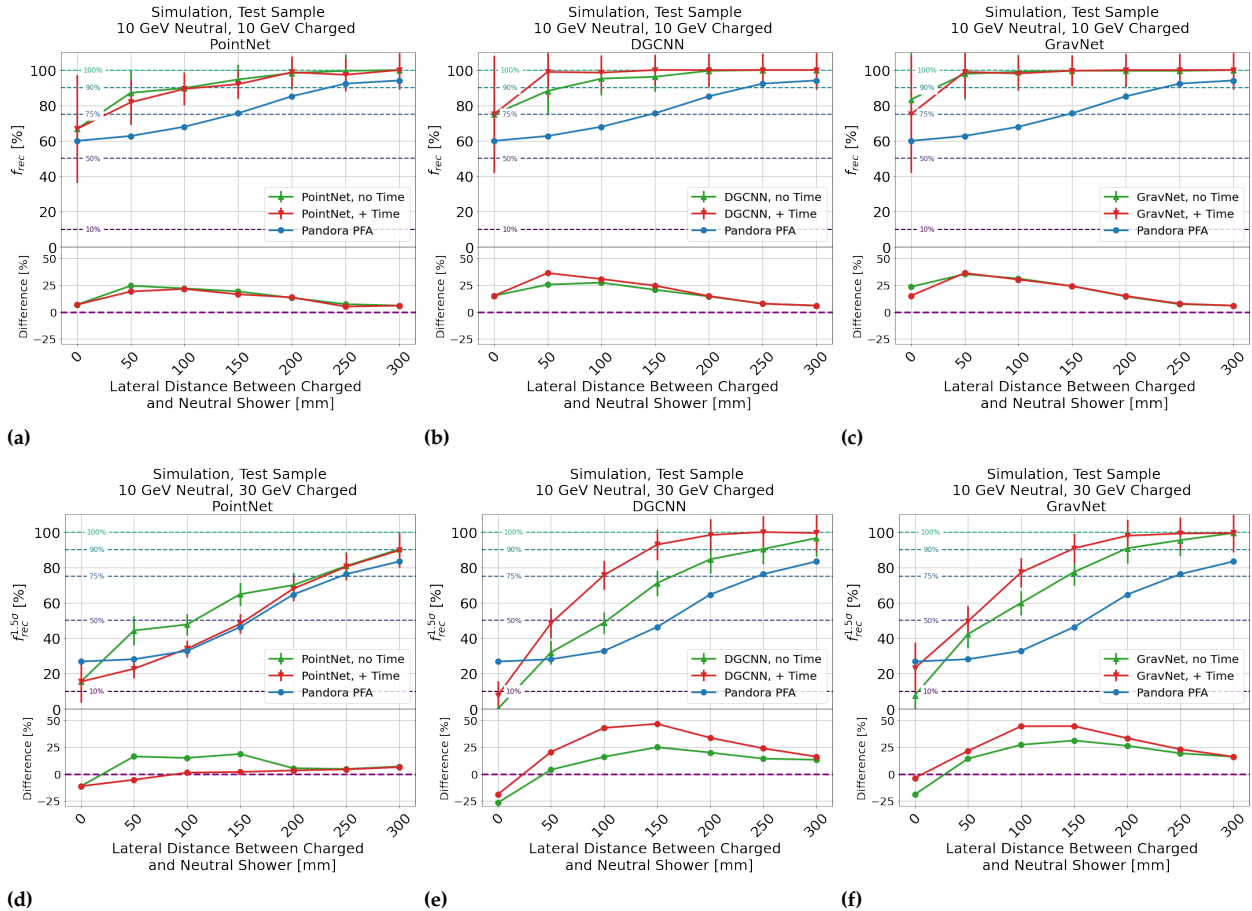


**Figure 6.37:** Figures 6.37a-6.37c, Figures 6.37d-6.37f and Figures 6.37g-6.37i show the mean fraction of events reconstructed within the calorimeter resolution as a function of lateral shower distance for PointNet, DGCNN and GravNet, respectively. The green and red points show the models trained without and with timing information, respectively.

**Comments:**

- For all models, as the energy of the neutral particle increases, the reconstructed fraction also increases and saturates after around 10-15  $\rho_M$ . The distance required decreases with energy, which is expected due to the increased energy density of hadron showers with particle energy due to a greater EM-fraction.
- For the DGCNN and GravNet models shown in Figures 6.37d-6.37f and Figures 6.37g-6.37i, significant improvements by up to 10% are observed with the inclusion of timing information, indicated by the red line compared to the green line and is in agreement with the results of Figure 6.34. It is noted that this improvement typically increases with separation distance before equalising at around 10-15  $\rho_M$  separation distance. However, no strong correlation is observed between the improvement and shower distance.

### Mean Fraction of Events Reconstructed Within Calorimeter Resolution vs Lateral Shower Distance for 10 GeV Synthetic Neutral Showers Separated from 10 GeV and 30 GeV Charged Showers, For Each Shower Separation Model Under Test Applied To The Test Sample of Simulation, Compared to the Study of [78]

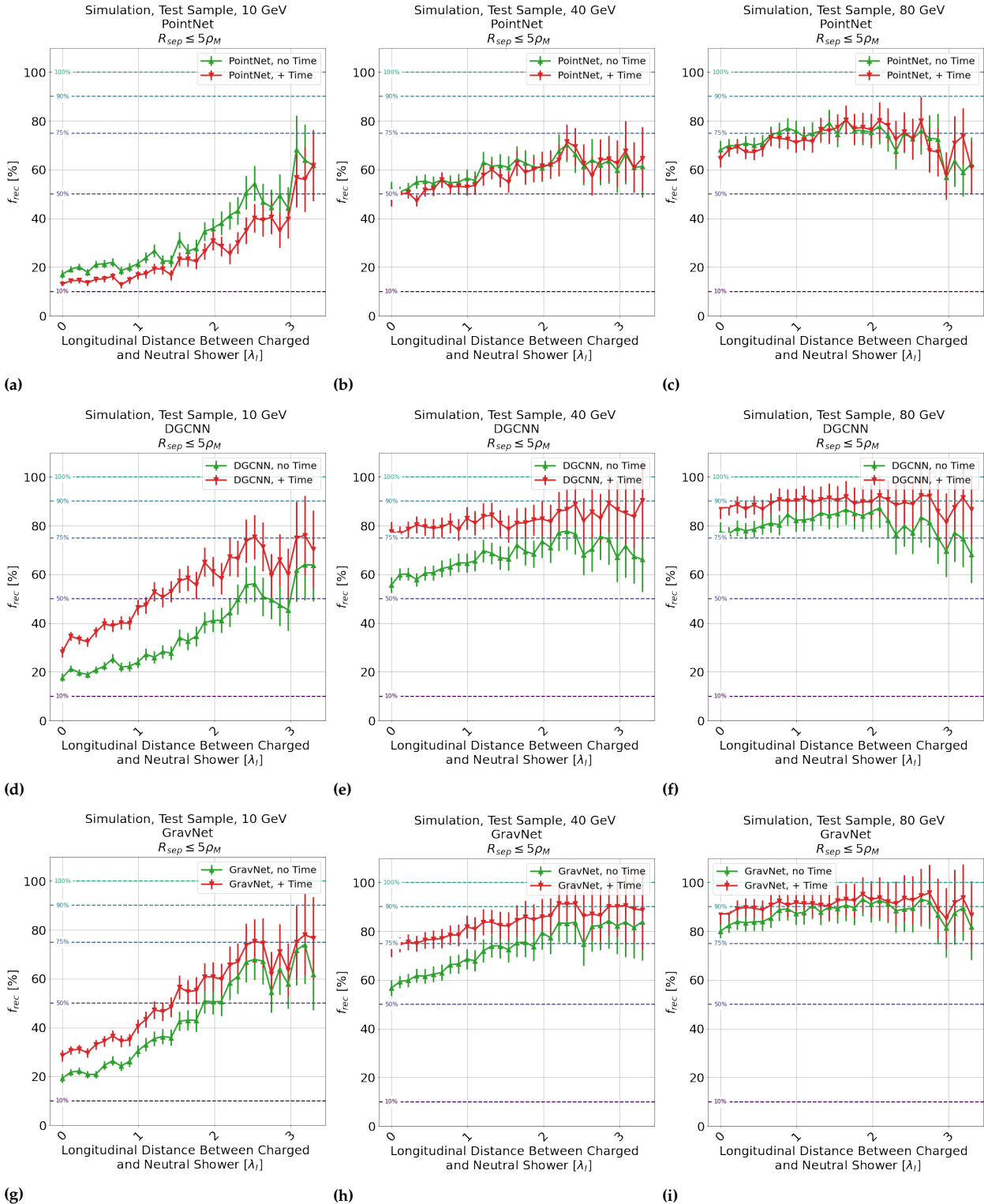


**Figure 6.38:** Figures 6.38a-6.38c and Figures 6.38d-6.38f show the mean fraction of events reconstructed within the calorimeter resolution as a function of lateral shower distance for PointNet, DGCNN and GravNet, for 10 GeV synthetic neutral showers separated from 10 GeV and 30 GeV Charged Showers, respectively. The green and red points show the models trained without and with timing information, respectively. The blue markers indicate the performance of Pandora PFA as applied in the study of Figure 9.14 of [78], and were extracted from these plots.

#### Comments:

- ▶ Except in the case of PointNet in Figure 6.38d, all networks show a significant improvement in performance compared to Pandora PFA.
- ▶ For 10 GeV synthetic neutral showers separated from 10 GeV charged showers, GravNet and DGCNN reconstruct nearly all of the events within the calorimeter resolution at inter-shower distances greater than 50 mm, corresponding to a maximum improvement of an additional 25% of events reconstructed within the calorimeter resolution, with or without timing information.
- ▶ For 10 GeV synthetic neutral showers separated from 30 GeV charged showers, GravNet and DGCNN show a significant improvement over Pandora PFA inter-shower distances greater 50 mm, corresponding to a maximum improvement of an additional 25-30% of events reconstructed within the calorimeter resolution without timing information, and an additional 35-45% of events reconstructed within the resolution with timing information.
- ▶ It is notable that the reconstruction efficiency of Pandora PFA ‘flattens out’ by comparison to the methods under test at distances below 50 mm since a more reasonable expectation would be continuous degradation in the performance as a function of distance, as is observed for the models under test. The reason for this effect is unknown.
- ▶ It is interesting to note that the performance of PointNet with timing information in Figure 6.38d is essentially the same as Pandora PFA for most distances under test, and is more consistent with the ‘s’-shape of the performance in this case than GravNet or DGCNN. However, the reasons why are again unknown.

**Mean Fraction of Events Reconstructed Within Calorimeter Resolution vs. Longitudinal Shower Distance for Showers Within  $5\rho_M$  Of One Another for 10 GeV, 40 GeV and 80 GeV Synthetic Neutrals Separated From 5-120 GeV Charged Showers for Each Shower Separation Model Under Test Applied To The Test Sample of Simulation**



**Figure 6.39:** Figures 6.39a-6.39c, Figures 6.39d-6.39f and Figures 6.39g-6.39i show the mean fraction of events reconstructed within the calorimeter resolution as a function of longitudinal shower distance (i.e. the shower starting position of both showers) for PointNet, DGCNN and GravNet, respectively. The green and red points show the models trained without and with timing information, respectively.

**Comments:**

- For all models, an improvement in the shower separation capacity of hadron showers with differing shower starting positions is observed for 10 GeV hadron showers and is less prevalent for other particle energies.
- In Figures 6.39d-6.39f and Figures 6.39g-6.39i, significant improvements of up to additional 10% due to timing information is observed. This improvement, however, is largely independent of the difference in the shower starting positions within the error.

### 6.3.1.7 Skewness of Confusion Energy

Biases in confusion energy are observed in the study, which is indicated in Figure 6.35, Figure 8.23 and Figure 8.24. These biases may be quantified by the skewness of the confusion energy distributions such that they may be studied.

To quantify the skewness, the medcouple as defined in Chapter 5 was calculated for all possible shower energy combinations for each combination of charged and neutral particle energies as in Figure 6.40. It is noted that essentially the same conclusions may be obtained with standard skewness, and therefore this is not shown.

#### Summary Comment to Figure 6.40

All methods of shower separation tend to produce an 'altruistic' shower separation technique, where the higher energy shower 'donates' energy to the lower energy one. This is indicated by the tendency for the neutral shower energy to be overestimated where the charged particle energy is greater than the neutral particle energy, and vice versa.

The study of [120] observed similar skewness in confusion when separating two charged hadrons with implementations of GravNet and DGCNN. In the study, the effect was attributed to the expectation value of the loss function being above/below the limits of  $f_{\text{hit}} \in [0, 1]$ , resulting in bias that results in this effect. No verification of this argument is provided in the reference and is therefore speculative.

This claim can be assessed by comparison to the results of Pandora PFA, which has no loss function and is instead a designed algorithm. A study to the one performed in this chapter was performed in [78], instead validating the performance of Pandora PFA in shower separation, using simulation and 2018 June Testbeam data obtained under the same conditions as presented in this chapter. Specific differences in the analysis chain may be found in the reference. Under the hypothesis made in [120], if the effect of skewness is entirely a result of the loss function, then the Pandora PFA method ought not to be skewed when separating hadron showers with a disparity between particle energies.

A comparison between distributions is shown between studies in Figure 6.41

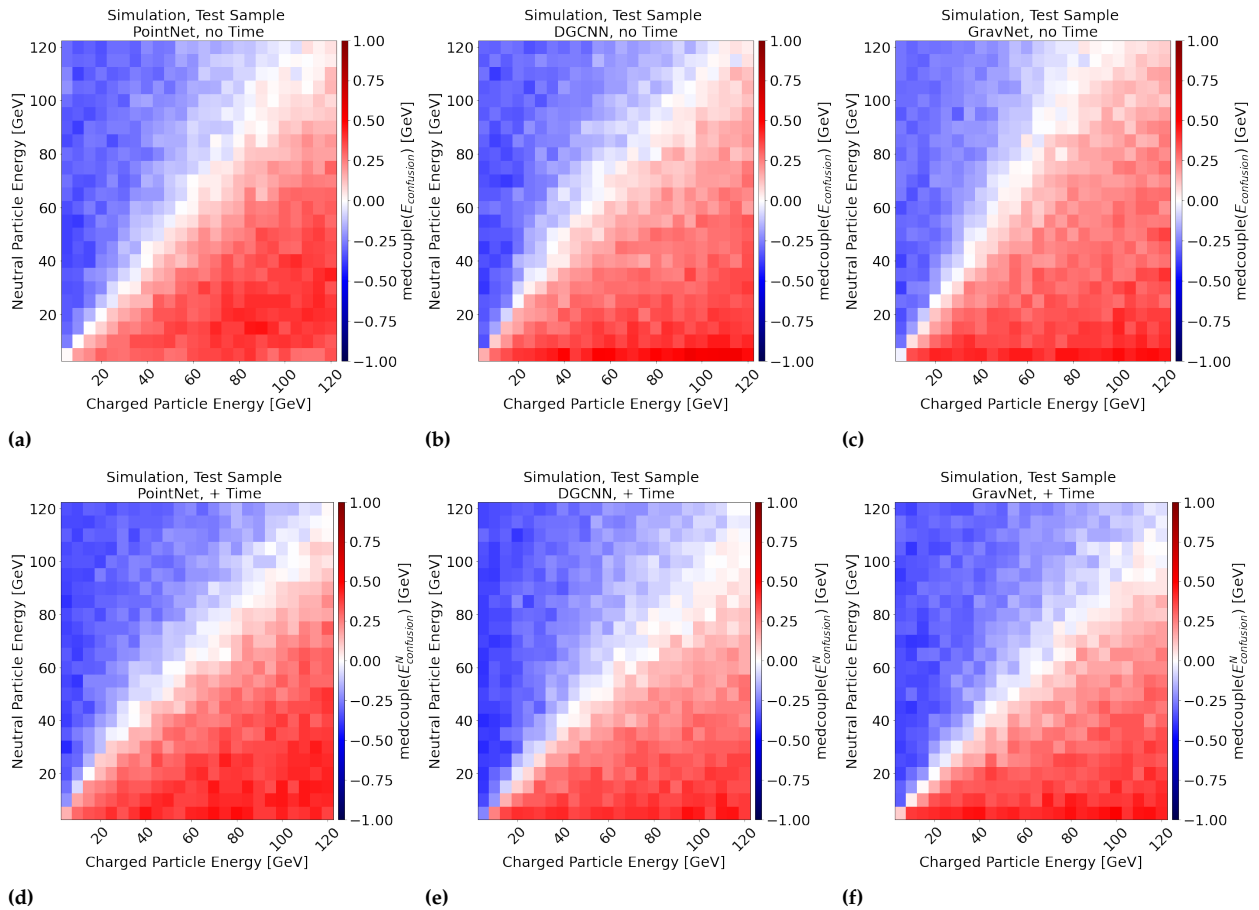
#### Summary Comment to Figure 6.41

It may be concluded that both distributions produced by Pandora PFA and GravNet, without timing information, applied to shower separation of a 10 GeV neutral and a 30 GeV charged hadron shower, result in similar distributions with right-tailed skewness. The hypothesis made in [120] cannot explain the agreement between studies.

Another plausible hypothesis is that distributing energy from a high-energy to a low-energy hadron shower during clustering is the optimal strategy for reconstructing an event with a significant disparity between them. It is noted that Pandora PFA preferentially splits up true clusters

instead of aggregating them in the initial clustering phase, by design [55], which states that: *"by design the initial clustering stage errs on the side of splitting up true clusters rather than merging energy deposits from more than one particle into a single cluster."* This strategy is logical: the most energy-dense region of a shower event is more probable to be a shower. It can be used as a clustering prior alongside the track position. If true, then the neural networks have learned the same technique autonomously. However, this hypothesis is also speculative and requires further analysis beyond the studies performed to confirm.

### Matrices of Medcouple of Confusion Energy Distributions for Each Shower Separation Model Under Test, Applied to The Test Sample of Simulation

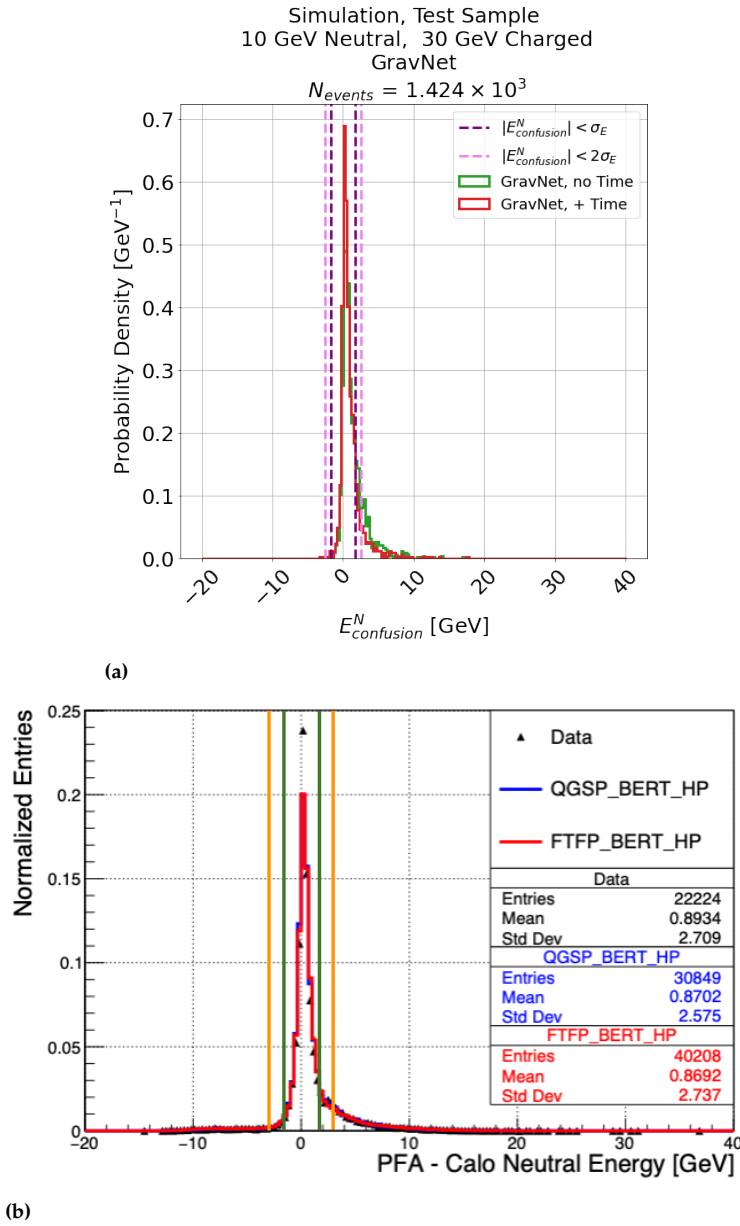


**Figure 6.40:** Figures 6.40a-6.40c and Figures 6.40d-6.40f show the medcouple as a function of the charged and neutral particle energy, for PointNet, DGCNN and GravNet, respectively. The colour axis indicates the medcouple. Red to blue indicates right-to-left-tailed skewness, whereas white indicates no skewness. Red indicates that the confusion energy tends towards the overestimation of the shower energy, while blue tends towards the underestimation of the shower energy.

#### Comments:

- In all shower separation models, with or without timing information, a clear observation is that the neutral shower energy tends to be more frequently underestimated where the neutral shower energy is greater than the charged shower energy, and vice versa. In the case where they are approximately equal, negligible skewness is observed. It may be concluded that the shower with the higher of the two energies typically 'donates' energy to the lower energy shower.

### Comparison of the Confusion Energy Distributions Comparing 10 GeV Neutral from 30 GeV Charged Hadron Showers By GravNet and Pandora PFA



**Figure 6.41:** Figure 6.41a shows the confusion energy distribution produced by GravNet, without timing information, for the test sample of simulation, separating a 10 GeV neutral hadron shower from a 30 GeV charged hadron shower. The purple and lilac dashed lines indicate the  $1\sigma$  and  $1.5\sigma$  resolution bands for simulation. Figure 6.41b shows the confusion energy distribution of the same event configuration reconstructed at a distance of 275–325 mm (approximately 11–13  $\rho_M$ , i.e. well-separated) applied to an analogous shower separation task performed in [78] using Pandora PFA, showing June 2018 Testbeam Data, and two sources of the simulation produced using the same conditions as this the study presented in this Chapter. The green and orange lines indicate the  $1\sigma$  and  $1.5\sigma$  resolution bands for CALICE June 2018 Testbeam Data. Taken from [78]. Note that the original nomenclature for the limits in [78] errantly presented the limits as  $2\sigma$  and  $3\sigma$  in that study.

**Comments:**

- The distributions of confusion energy, including the skewness observed in the GravNet model under test, show excellent agreement with the study of [78], indicating that similar skewness is also observed in Pandora PFA.



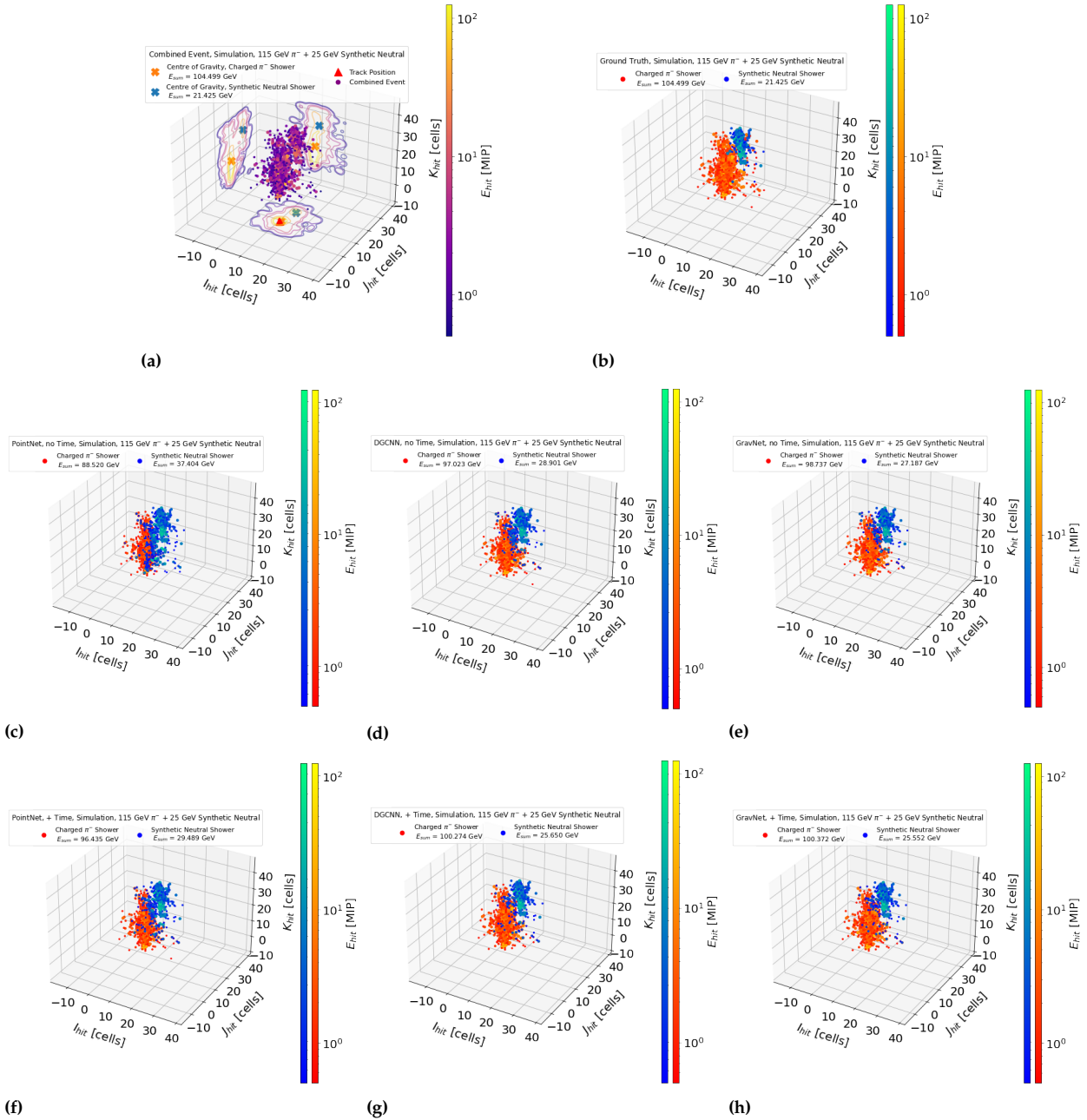
### 6.3.1.8 Examples Event Displays of Failure Modes

Two main examples of failure modes (i.e. where the confusion energy was greater than  $3\sigma_E$ ) were observed: the case where the neutral particle energy is greater and the case where the charged and neutral hadron showers are observed with close centres of gravity. Both modes are expected from the analysis in this section and are presented in Figure 6.42 and Figure 6.42.

#### Summary Comment To Figure 6.42 and Figure 6.43

Confusion energy is expected to affect the HAD-halo in events with  $E_{\text{particle}}^Q \geq E_{\text{particle}}^N$ . It is noted timing may aid in the reassociation of clusters that may otherwise be difficult to separate purely from the spatial dependence of the energy. In the case of two completely overlapping showers, with similar shower starting layers and impinging positions, bias is observed towards the shower with the lowest energy.

**Example Event Displays of a 25 GeV Synthetic Neutral Hadron Shower and a 115 GeV Charged Shower, After the Application of Each Shower Separation Model Under Test, With  $|E_{\text{confusion}}^N| > 3\sigma_E$**

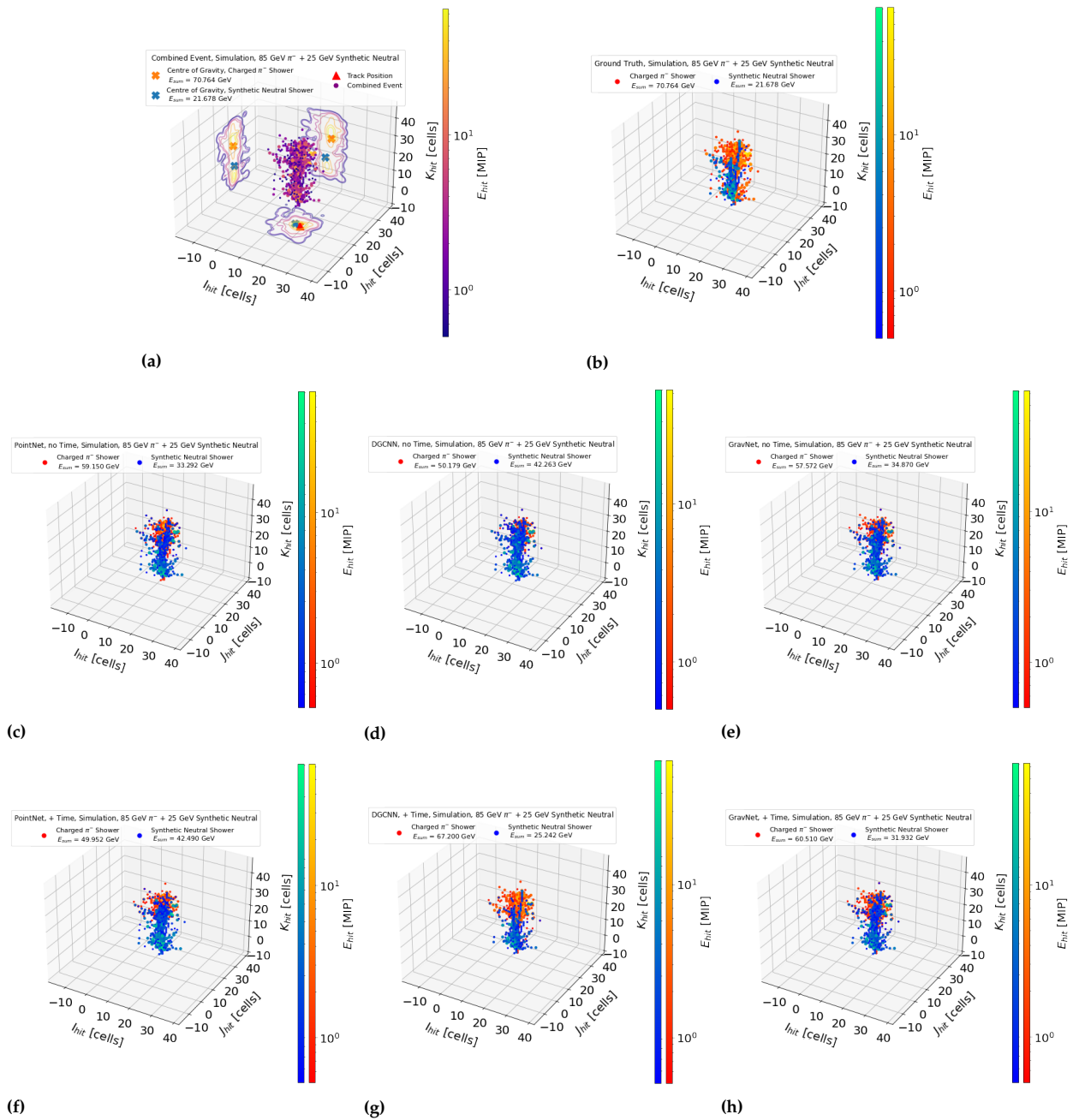


**Figure 6.42:** Example event displays illustrating the application of the trained algorithms to test charged and synthetic neutral hadron shower event. This case gives more than a  $3\sigma_E$  confusion energy in the observed confusion case where  $E_{\text{particle}}^Q > E_{\text{particle}}^N$ . Else, as in Figure 6.19 and Figure 6.20.

**Comments:**

- ▶ It is noted that, in general, most confusion energy is in the HAD-halo region of the hadron shower, in the form of energy that belonged to the charged hadron shower in red to the neutral shower in blue. This is observed for all models studied in this example and is therefore expected to degrade the performance of SC models in the HAD-halo.
- ▶ Comparison of Figures 6.42d and 6.42g and Figures 6.42e and 6.42h indicates that the addition of timing information results in a better association of the confusion energy at the ‘boundary’ between the two showers. This could indicate that confusion of neutral fragments is reduced using timing information compared to without.

Example Event Displays of a 85 GeV Synthetic Neutral Hadron Shower and a 25 GeV Charged Shower, After the Application of Each Shower Separation Model Under Test, With  $|E_{\text{confusion}}^N| > 3\sigma_E$



**Figure 6.43:** Example event displays illustrating the application of the trained algorithms to test charged and synthetic neutral hadron events. This case gives more than a  $3\sigma_E$  confusion energy due to a shower configuration where one is on top of the another. Else, as in Figure 6.19 and Figure 6.20.

#### Comments:

- In this event, both hadron showers were observed in a configuration where one shower is on top of the another with similar shower starting positions. It is expected to be broadly unresolvable. This configuration thus grants some insight into the default behaviour of the algorithms.
- All models, with or without time, is observed to allocate more energy to the neutral shower than the charged one than in the ground truth where no clear separation of showers is achieved, indicating a learned bias in the reconstruction.
- Additionally, comparison of Figures 6.43d and 6.43g to Figures 6.43e and ?? indicate that the DGCNN and GravNet models result in superior clustering using timing information.

### 6.3.1.9 Software Compensation of Separated Neutral Hadron Showers

In this section, the performance of the software compensation models developed in Chapter 5 is applied to simulation for the best performing shower separation model, GravNet, using the same two-standard deviation Gaussian fitting method as described in Section 5.3 of Chapter 5.

In addition to the cuts specified in Section 6.2.2, three additional cuts were applied to the testing dataset:

- ▶ events were selected as having reconstructed energy within  $1\sigma_E$  for the particle energy under test;
- ▶ the same tail-catcher cut is applied as in Chapter 5 of  $E_{\text{sum}}^{\text{N,TCMT}} < 25\text{MIP}$  is applied to reduce leakage to neutral hadron showers reconstructed within the resolution of the AHCAL. It is assumed that the tail catcher energy is known perfectly since it is used for tagging leakage in this study and does not contribute to the evaluation of the clustering performance;
- ▶ a 0.5MIP active cell energy threshold is applied to the minimum energy of the separated neutral event. This is required because the SoftMax function cannot output a hard 1 or 0 for a fraction, often resulting in negligibly small energy deposits below the noise threshold for AHCAL. This is problematic because the SC model has no exposure to active cells with energies below 0.5MIP and was observed to enhance the energy of these spurious confusion hits very strongly as if they were part of the HAD fraction. This problem can be cured by applying a threshold cell energy of that value to the separated neutral event, or training the model with energy fraction information as proposed later in this section.

For GravNet, a total of 82% and 87% of the entire test sample were reconstructed within the calorimeter resolution. Therefore this study limits the amount of confusion experienced by the SC model. However, the SC model trained in Chapter 5 has no exposure to events with any amount of confusion energy. The influence of the confusion is studied by comparing the performance of the SC model on the whole dataset and separately for the subset  $E_{\text{particle}}^{\text{N}} \geq E_{\text{particle}}^{\text{Q}}$ , which is demonstrated in Figure 6.34g and 6.34h to be the best case scenario for shower separation. It is expected that SC will perform optimally for the subset and degrade as confusion energy is introduced.

The uncompensated and compensated reconstructed neutral energy distributions are shown for all events in Figure 6.44 and for  $E_{\text{particle}}^{\text{N}} \geq E_{\text{particle}}^{\text{Q}}$  in Figure 6.45. The linearity and resolution for both cases are shown in Figure 6.46.

### Summary Comment to Figures 6.44-6.46

The SC network trained in Chapter 5 improves resolution for separated neutral showers in simulation, indicating that the technique can be applied to separated hadron showers. Bias is served, which can be attributed to confusion by comparison of the results with for cases where  $E_{\text{particle}}^N > E_{\text{particle}}^Q$ .

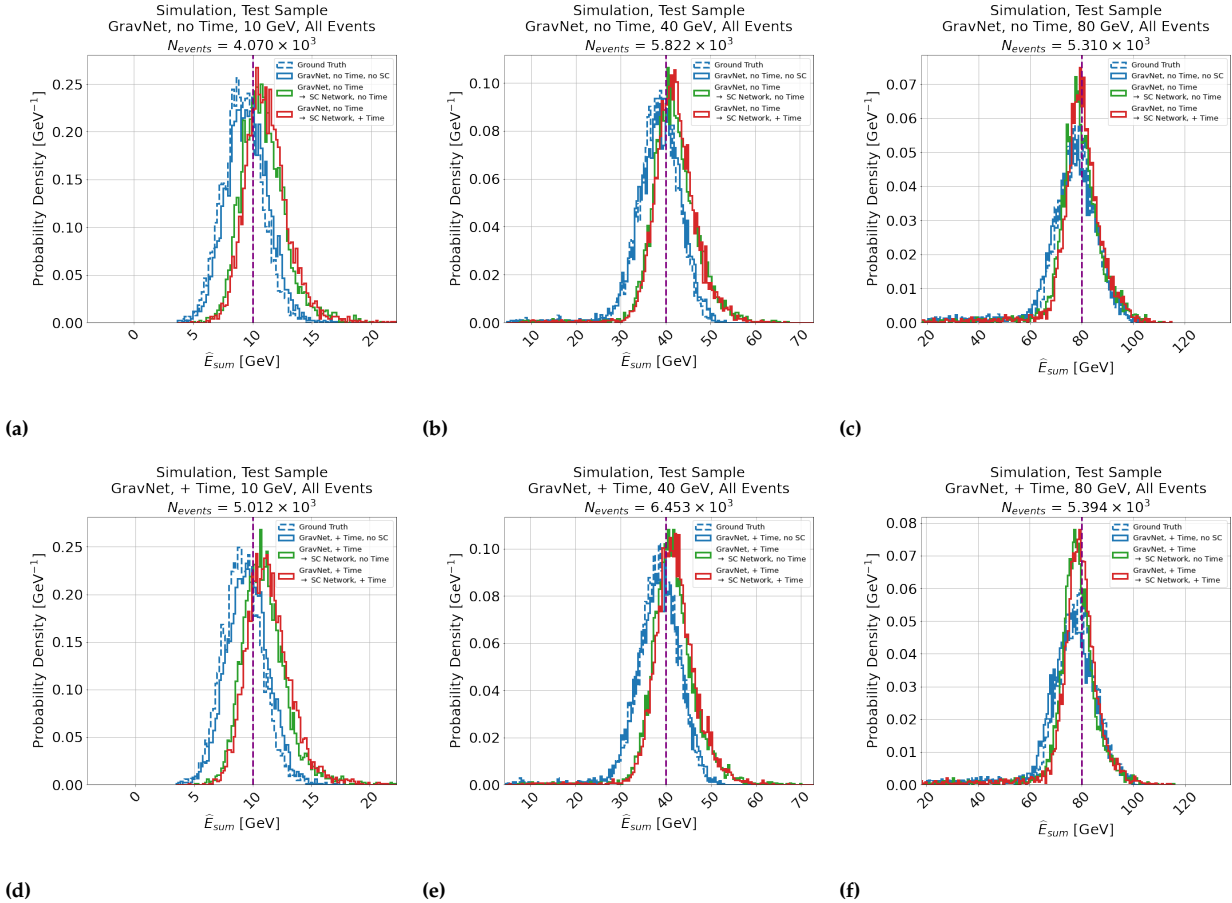
A limitation to training the SC model without including confusion and then applying it to showers with confusion energy is observed bias in the reconstruction.

As a potential solution to the caveats presented in this section, a practical application of the SC model will likely have to be made aware of the predicted energy fractions from the shower separation model such that the SC model can adapt to the concept of confusion energy. The models could be trained in tandem as part of the same network: one to separate the energy deposits and another to perform energy reconstruction.

This could be achieved by first determining fractions of each hadron shower using a shower separation network like GravNet then, indexing the energy fractions of neighbours as part of the  $k$ -NN operation applied in the SC network as explained in Section 5.2.1. This would allow the network to identify clusters that are likely to have come from confusion and develop a more sophisticated weighting technique to account for this possibility. This would involve minor modifications to the existing models and could be achieved by training the SC model as a submodule of the shower separation network.

Nonetheless, the resolution is improved in both cases studied, and therefore the SC network presented in Chapter 5 indicates a promising new technique for PF software compensation.

**Synthetic Neutral Reconstructed Energy Distributions Produced After Shower Separation By GravNet On Events Reconstructed Within The Calorimeter Resolution, Without and With Timing Information, After Separation, Without and With SC, Without And With SC Timing Information, with TCMT Cut Applied, For All Charged Hadron Shower Energies, Applied To 10 GeV, 40 GeV and 80 GeV Neutral Hadron Particle Energies In The Test Sample of Simulation**

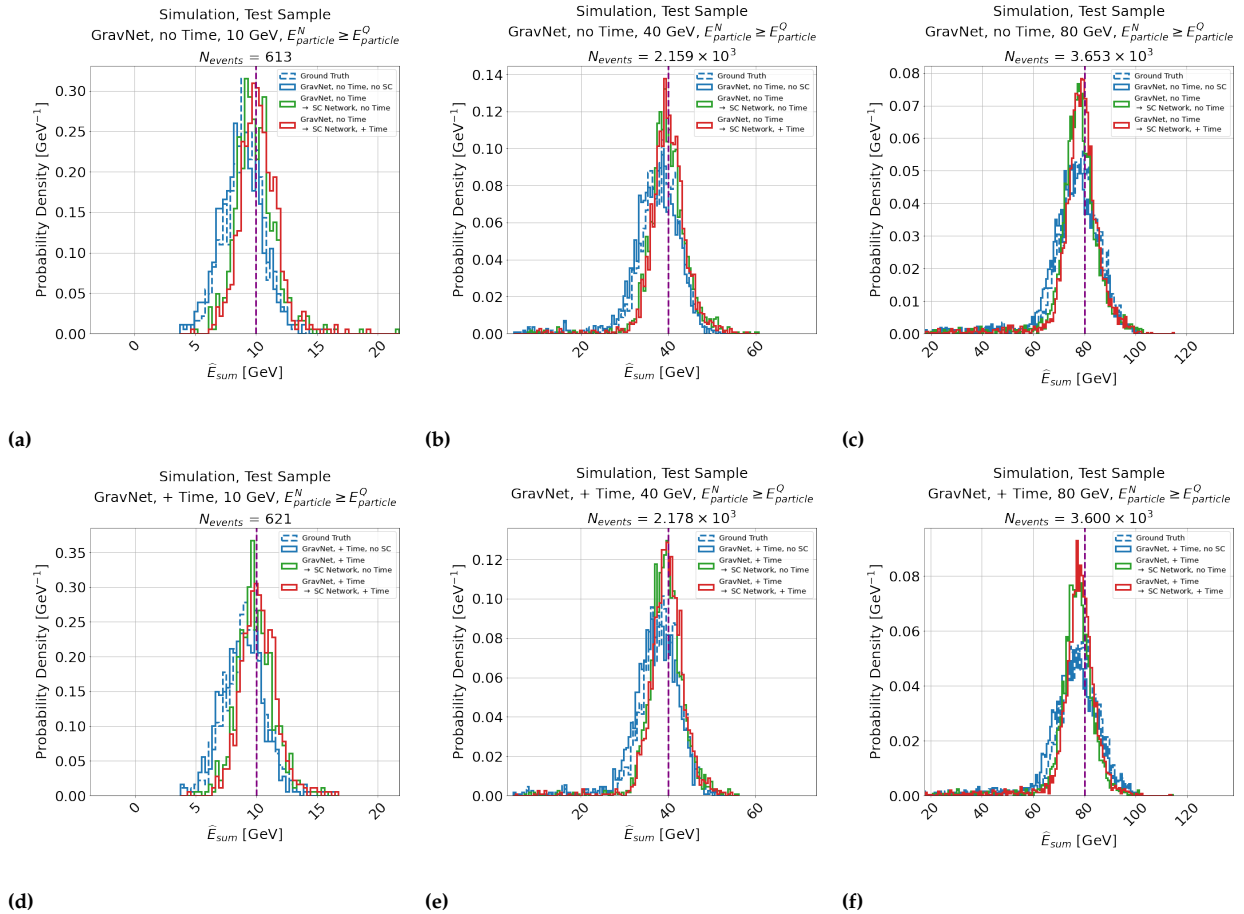


**Figure 6.44:** Reconstructed neutral hadron shower energy distributions for the GravNet network applied to the test sample of simulation, with no other selection criteria applied. The blue dashed line indicates the ground truth (perfectly known) reconstructed energy distribution. The green and red lines indicate the SC trained and applied without and with timing information, respectively, applied after separation. The purple dashed line indicates the particle energy.

**Comments:**

- Comparison of the blue dashed and blue lines illustrate that the GravNet model typically produces a reconstructed energy distribution close to the original shower energy distribution, thereby indicating that GravNet is an effective shower separation algorithm.
- It is observed that the SC networks tend to produce distributions with a smaller width, indicating that compensation is being performed.
- Figure 6.44a and Figure 6.44d indicate a systematic bias in the mean reconstructed shower energy. This bias was not observed in Chapter 5, and therefore can only be a consequence of the confusion energy.
- Overall, the distributions observed appear qualitatively Gaussian.

**Synthetic Neutral Reconstructed Energy Distributions Produced After Shower Separation By GravNet On Events Reconstructed Within The Calorimeter Resolution, Without and With Timing Information, After Separation, Without and With SC, Without And With SC Timing Information, with TCMT Cut Applied, For  $E_{particle}^N > E_{particle}^Q$ , Applied To 10 GeV, 40 GeV and 80 GeV Neutral Hadron Particle Energies In The Test Sample of Simulation**

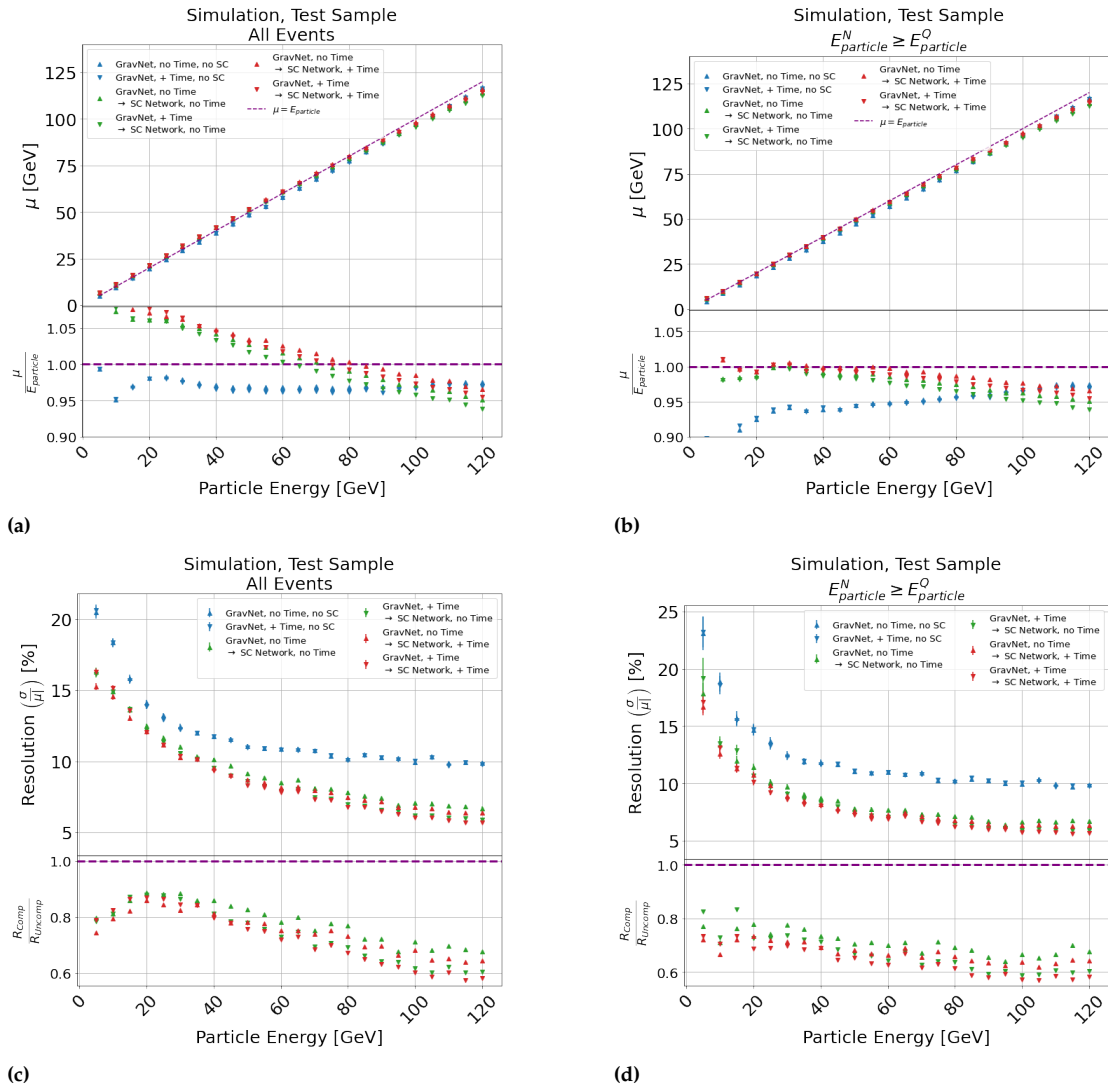


**Figure 6.45:** Reconstructed neutral hadron shower energy distributions for the GravNet network applied to the test sample of simulation, with no other selection criteria applied. The blue dashed line indicates the ground truth (perfectly known) reconstructed energy distribution. The green and red lines indicate the SC trained and applied without and with timing information, respectively, applied after separation. The purple dashed line indicates the particle energy.

**Comments:**

- Comparison of Figure 6.44 with this figure indicates that, overall, smaller bias and proportionally greater compensation are  $E_{particle}^N > E_{particle}^Q$  than when all events are included. This is expected from the prior observation that the reconstruction worsens depending on the particular combination of charged and neutral hadron.
- Else, as in Figure 6.44.

**Linearity And Resolution of Separated Neutral Hadron Showers Obtained From GravNet Model With and Without SC, Applied To The Test Sample Of Simulation, Studied For All Events and Events Where  $E_{particle}^N < E_{particle}^Q$**



**Figure 6.46:** Figures 6.46a and 6.46b and Figures 6.46c and 6.46d show the linearity and resolution of the neutral showers after separation by GravNet, using all events and for the case  $E_{particle}^N < E_{particle}^Q$ . The up and down arrows indicate the use of timing information in the shower separation model. The blue, green and red colours indicate the intrinsic calorimeter responses, and SC networks with and without timing information, after separation. Else, as in Figures 5.25a and 5.25b.

**Comments:**

- Figures 6.46c and 6.46d indicate an improvement in resolution is observed by application of SC for both studied cases, as observed in Figures 6.44 and 6.45.
- Comparison of Figures 6.46a and 6.46b and Figures 6.46c and 6.46d indicate that, if applied to the entire dataset, a significantly greater bias (no more than 6-7%) and slightly worse resolution is produced than in the case where  $E_{particle}^N > E_{particle}^Q$ , which produces a linear response and a superior overall resolution. In particular, confusion energy affects the events in the energy range 10-40 GeV, which in particular experience overestimation of the shower energy. Small amounts of confusion energy could explain this in the HAD halo, which is energy-enhanced and produces a greater overall SC response.
- The effect could be explained by the inclusion of low-energy hits from another shower, which are typically in the HAD-halo and most likely to be confused between showers (e.g. in Figure 6.42, confusion energy is wrongly attributed to the neutral hadron shower. Most of it could be confused with the HAD halo.). The study of Chapter 5 indicates that these energy deposits are likely to be enhanced, resulting in a general tendency to overestimate the response of the event. This would also explain the downwards-sloping bias observed in Figure 6.46a compared to Figure 6.46b.
- The inclusion of timing information produces a marginally superior resolution, as indicated by the red arrows being lower overall than the green arrows. improvement in resolution in Figures 6.46c and 6.46d. The biases are observed Figures 6.46a and 6.46b are found to be relatively similar with time as without.



## 6.3.2 Data Model

In this section, the best-performing network from the previous section, GravNet, as described in Section 6.2.1 was trained on data, without timing information, and evaluated. The corresponding network trained on simulation was then evaluated using the same events to determine if significant differences were observed between the performance on simulation and data and if the shower separation can be trained on a limited data sample.

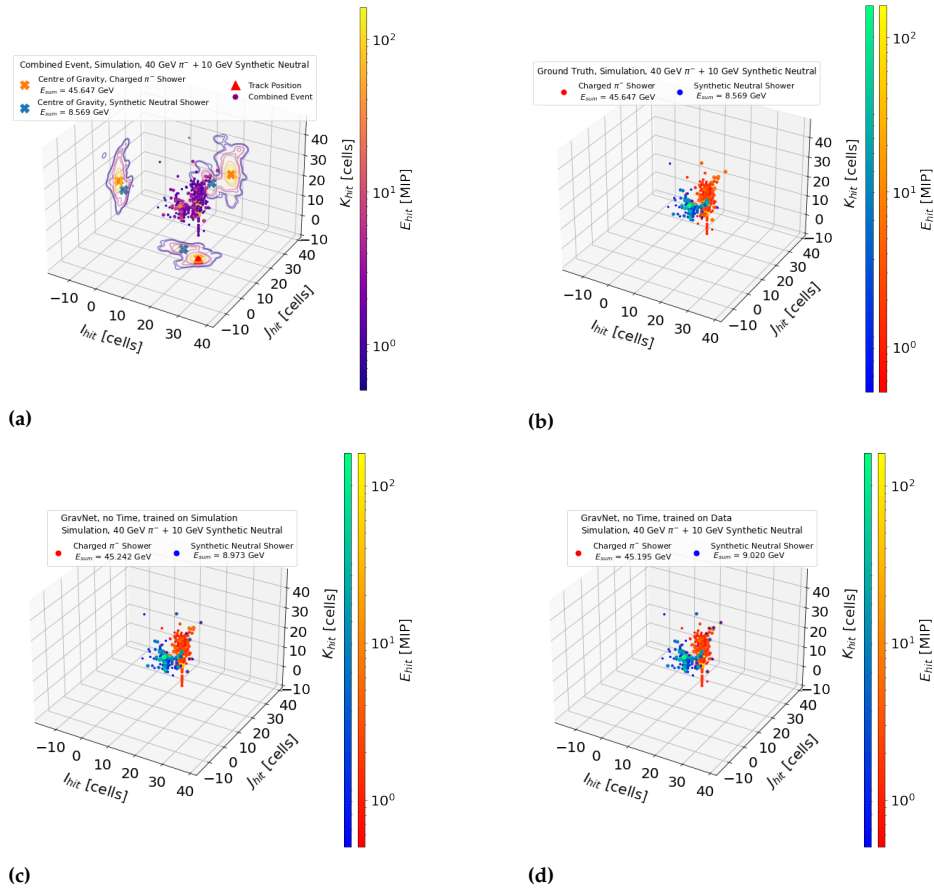
### 6.3.2.1 Example Event Displays

As in Section 6.3.1.1, example event displays are shown for 10 GeV, 40 GeV and 80 GeV neutral hadron showers separated from a 40 GeV charged hadron shower in Figures 6.47-6.49. The models trained on simulation and data are compared in these example events.

#### Summary Comment to Figures 6.47-6.49

The shower separation models show qualitatively similar performance when applied to shower separation of data showers as simulation showers, evidenced by the similar performance of the model trained on simulation as the model trained on data.

**Example Event Displays of a 10 GeV Synthetic Neutral and a 40 GeV Charged Hadron Shower, After the Application of The GravNet Shower Separation Model Trained On Simulation And Data, Applied To Testing Dataset in Data**

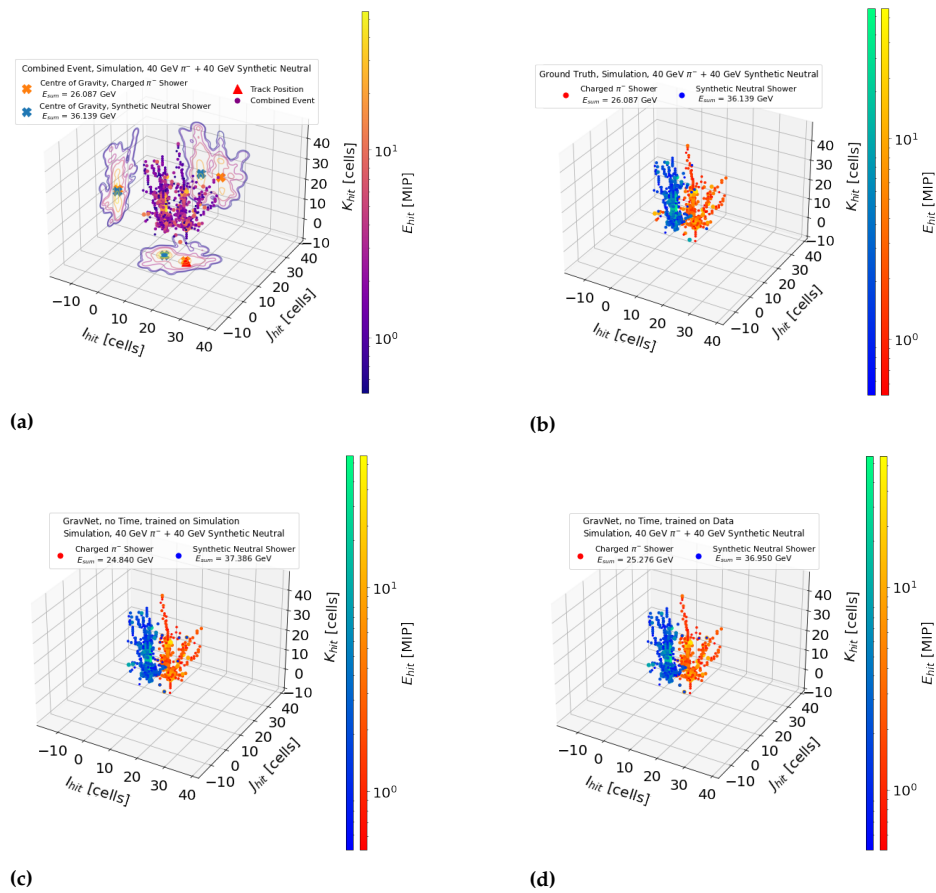


**Figure 6.47:** Example event displays illustrating the application of the trained algorithms to a test charged and synthetic neutral hadron shower event. Figure 6.47a shows the initial combined event, as in Figure 6.20. Figure 6.47b shows the original event to be reconstructed, as in Figure 6.19. Figures 6.47c and 6.47d show the GravNet models as described in Section 6.2.1 trained on simulation and data, respectively.

**Comments:**

- ▶ As for Figures 6.27-6.29, the GravNet model qualitatively separates hadron showers in good agreement with the ground truth.
- ▶ Figures 6.47c and 6.47d indicate that the model trained on simulation and data perform similarly overall.

**Example Event Displays of a 40 GeV Synthetic Neutral and a 40 GeV Charged Hadron Shower, After the Application of The GravNet Shower Separation Model Trained On Simulation And Data, Applied To Testing Dataset in Data**

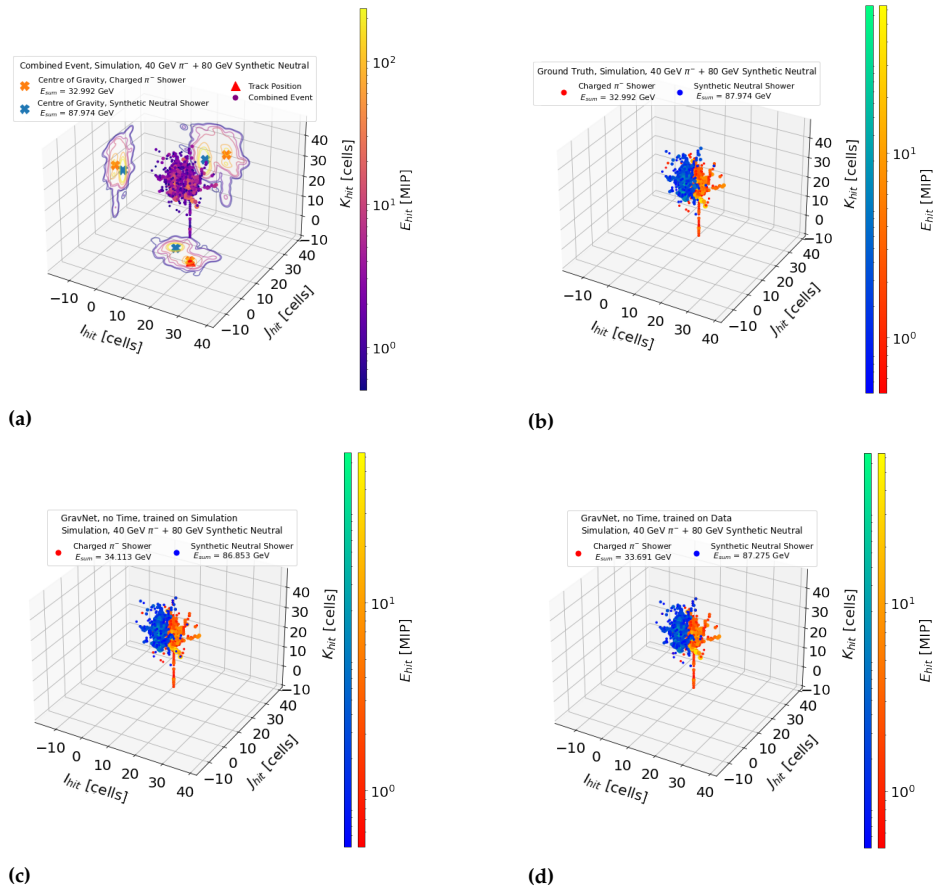


**Figure 6.48:** Example event displays illustrating the application of the trained algorithms to a test charged and synthetic neutral hadron shower event. Figure 6.48a shows the initial combined event, as in Figure 6.20. Figure 6.48b shows the original event to be reconstructed, as in Figure 6.19. Figures 6.48c and 6.48d show the GravNet models as described in Section 6.2.1 trained on simulation and data, respectively.

**Comments:**

- As in Figure 6.47.

**Example Event Displays of a 80 GeV Synthetic Neutral and a 40 GeV Charged Hadron Shower, After the Application of The GravNet Shower Separation Model Trained On Simulation And Data, Applied To Testing Dataset in Data**



**Figure 6.49:** Example event displays illustrating the application of the trained algorithms to a test charged and synthetic neutral hadron shower event. Figure 6.49a shows the initial combined event, as in Figure 6.20. Figure 6.49b shows the original event to be reconstructed, as in Figure 6.19. Figures 6.49c and 6.49d show the GravNet models as described in Section 6.2.1 trained on simulation and data, respectively.

**Comments:**

- As in Figure 6.47.

### 6.3.2.2 Linearity Of Response

As in 6.3.1.3, the distributions of the reconstructed hadron shower energy are presented as a function of the original shower energy for all for the GravNet model, trained on simulation and data. They are shown in Figure 6.50. The linearity of response for the reconstructed neutral is once again measured using the mean and the MPV.

#### Summary Comment to Figure 6.50

Bias is once again observed in the energy reconstruction, indicating this effect is not a consequence of using simulation to train the models. Performance is similar between networks trained on simulation and data.

### 6.3.2.3 Spread of Confusion Energy

As in Section 6.3.1.3, the complete distributions of the reconstructed hadron shower energy are presented as a function of the original shower energy for the GravNet model, trained on simulation and data. They are shown in Figure 6.51, with the  $RMS_{90}$  and MAD of each distribution shown.

#### Summary Comment to Figure 6.51

Both networks trained on simulation and data achieve similar performance, with minor differences in the order of 100 MeV in MAD and  $RMS_{90}$ . This indicates that shower separation networks are unaffected by the particular average energy density of the hadron showers and can be applied in experimental data.

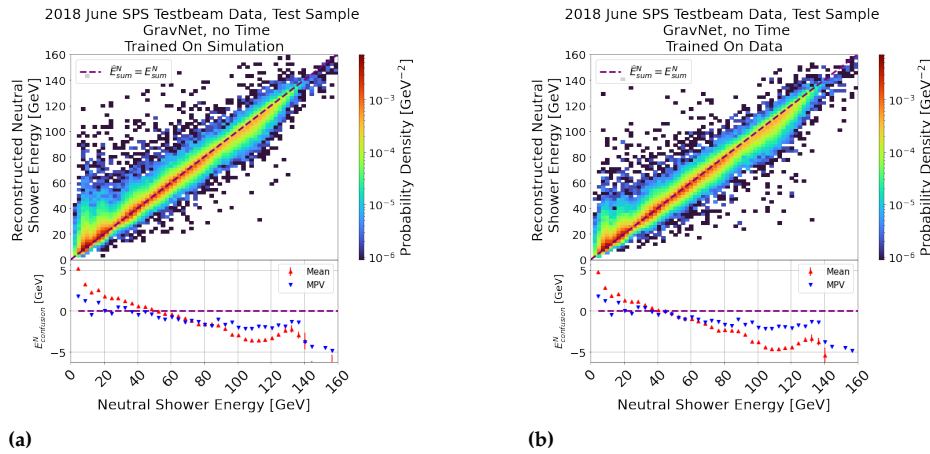
### 6.3.2.4 Fraction of Events Reconstructed Within Calorimeter Resolution vs. Charged/Neutral Particle Combination

As in Section 6.3.1.4, the fraction of events reconstructed within the calorimeter resolution as described in Equation 6.9 as a function of the combination of charged and neutral particle energy are shown in Figure 6.52. The resolution was obtained from the study of Chapter 5 presented in Table 5.4 of  $R_{sim} = 56.1\%/\sqrt{E_{particle}} \oplus 6.1\%$ .

#### Summary Comment to Figure 6.52

As in Section 6.3.1.4, above 80-90 % of events are reconstructed within the calorimeter resolution where the neutral hadron energy is greater than the charged hadron energy. The performance degrades for the opposite case. There is no more than a 5 % difference between the results obtained from training with simulation and data.

### Joint Distribution between Predicted and True Reconstructed Neutral Shower Response of the Test Sample of Data For GravNet Trained on Simulation and Data

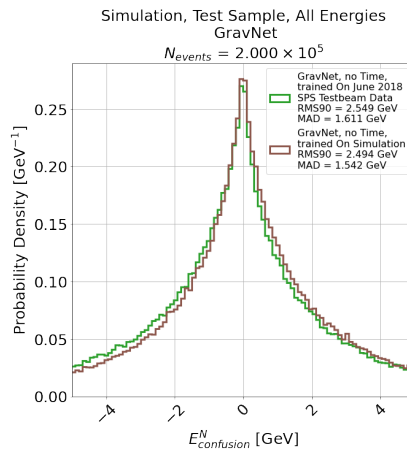


**Figure 6.50:** Distributions of the neutral confusion energy for each shower separation model under test. The green and brown lines indicate the GravNet models, trained on RMS<sub>90</sub> and MAD are shown in the legend for each model.

**Comments:**

- ▶ Comparison of Figure 6.50a and Figure 6.50b indicate that the two models perform with about the same overall performance at separating hadron showers in data.
- ▶ Similar biasing is observed when applying the models to data, as in Figure 6.31.

### Distributions of Confusion Energy for the Entire Test Sample of Data, With and Without Timing Information, For GravNet Trained Without and With Timing Information

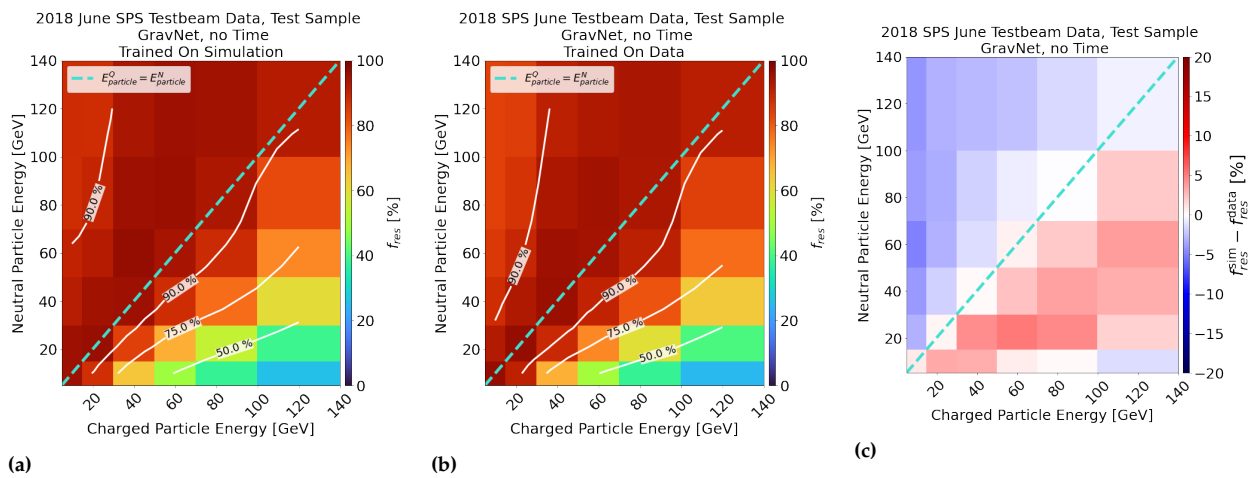


**Figure 6.51:** Distributions of the neutral confusion energy for each shower separation model under test. The green and brown lines indicate the GravNet models, trained on RMS<sub>90</sub> and MAD are shown in the legend for each model.

**Comments:**

- ▶ Comparison of the brown and green curves indicate that the two models perform with about the same overall performance at separating hadron showers in data. A slightly larger RMS<sub>90</sub> and MAD are observed for the model trained with simulation than the one trained with data, though the width depends on the energy combinations chosen for study.
- ▶ This result implies that the differences in energy density observed in Section 5.2.3.1 and Section 6.2.2 do not significantly affect the performance of the reconstruction or that shower separation models.
- ▶ It is noted that the SC network of Chapter 5 observed a much greater influence of the energy density on the output of the model. However, this is expected because the shower separation algorithm presented does not involve reconstructing the energy but instead the fraction of the shower energy belonging to each shower in each cell.

### Matrices of the Fraction of Events Reconstructed Within Calorimeter Resolution for Each Shower Separation Model Under Test Applied To The Test Sample of Data

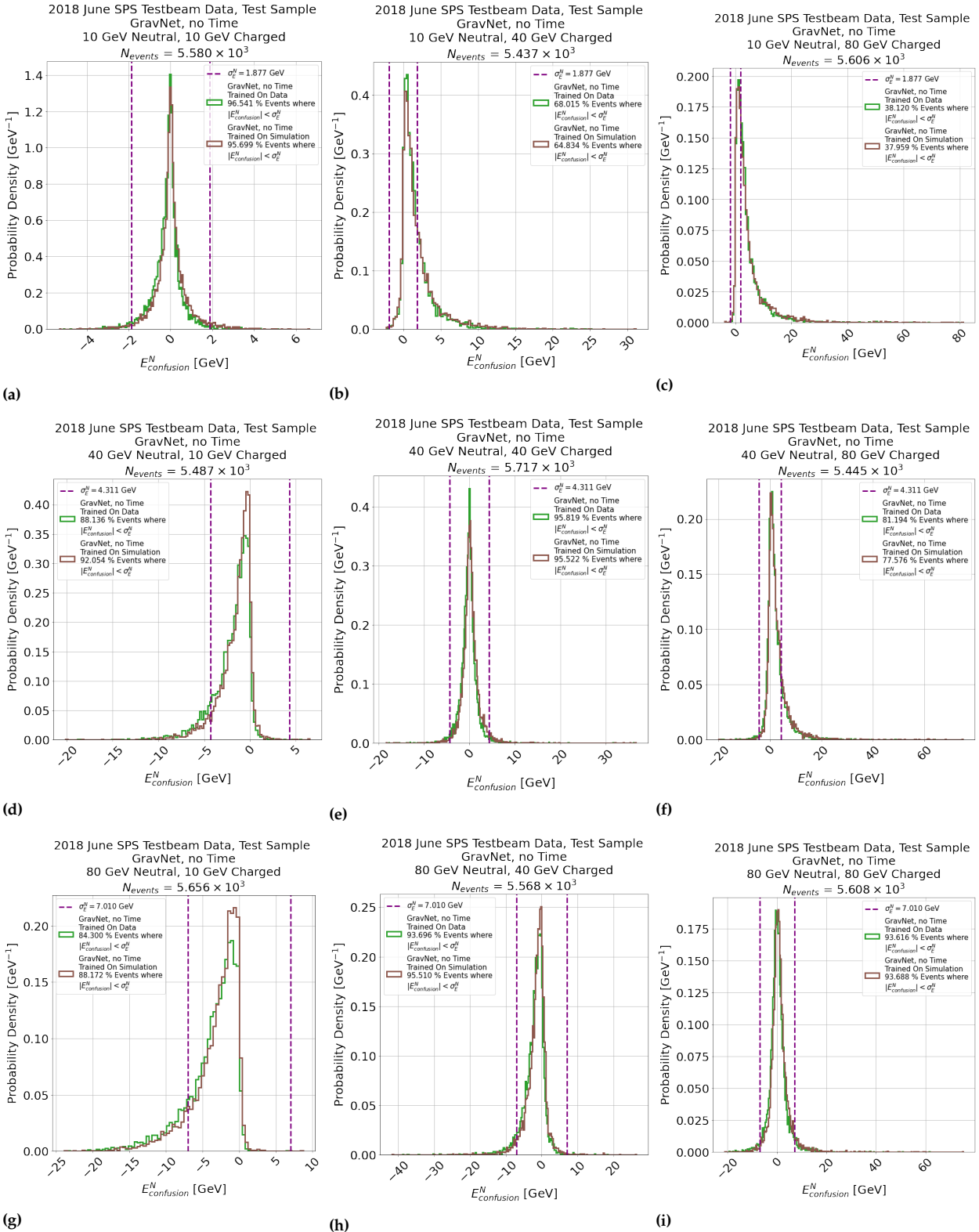


**Figure 6.52:** Figures 6.52a and 6.52b show the matrices of the fraction of events with energy reconstructed within the calorimeter resolution as a function of the charged and neutral particle energy for GravNet, trained with and without timing information, respectively, where red to blue indicates a higher to lower percentage of events. The turquoise dashed line indicates  $E_{particle}^Q = E_{particle}^N$ , while white lines indicate contours. Figure 6.52c indicates the ratios of the fractions.

#### Comments:

- Figures 6.52a-6.52b indicate the same asymmetry as observed in Figure 6.34. This result indicates that the effect is not necessarily from using simulation or data to train the neural networks.
- Figure 6.52c indicates that the fraction of events reconstructed within the resolution of the AHCAL by the GravNet network trained on simulation and data varies by no more than around 5%. This result means that the performance of the neural networks is not strongly related to the choice to use simulation or data for training the models.

### Distributions of Confusion Energy for Possible Combinations of Charged and Neutral Hadron Particle Energies at 10 GeV, 40 GeV and 80 GeV For GravNet Trained On Simulation And Data, Applied To The Test Sample Of Data



**Figure 6.53:** Distributions of confusion energy for the GravNet network applied to the test sample of data. The green and brown lines indicate the models trained on data and simulation, respectively. The purple dashed lines indicate the resolution of the AHCAL calorimeter in simulation.

**Comments:**

- ▶ No significant differences between the model trained on simulation and on data were observed.
- ▶ Else, as in Figure 6.35.



### 6.3.2.5 Fraction of Events Reconstructed Within Calorimeter Resolution vs. Lateral Shower Distance

As in Section 6.3.1.5, the fraction of reconstructed events within the calorimeter resolution was determined as a function of lateral shower distance for the GravNet model trained on simulation and data. The results are shown in Figure 6.54.

#### Summary Comment to Figure 6.54

As the lateral distance between hadron showers increases, so too does the fraction of reconstructed events increase and saturate after around 10-15  $\rho_M$ . No significant differences between the model trained on simulation and on data were observed.

### 6.3.2.6 Fraction of Events Reconstructed Within Calorimeter Resolution vs. Longitudinal Shower Distance

As in Section 6.3.1.6, the fraction of reconstructed events within the calorimeter resolution was determined as a function of longitudinal shower distance for the GravNet model trained on simulation and data. The results are shown in Figure 6.55.

#### Summary Comment to Figure 6.55

The performance of the models increases as the difference between shower starts increases, with less of an effect the greater the particle energy is. No significant differences between the model trained on simulation and on data were observed.

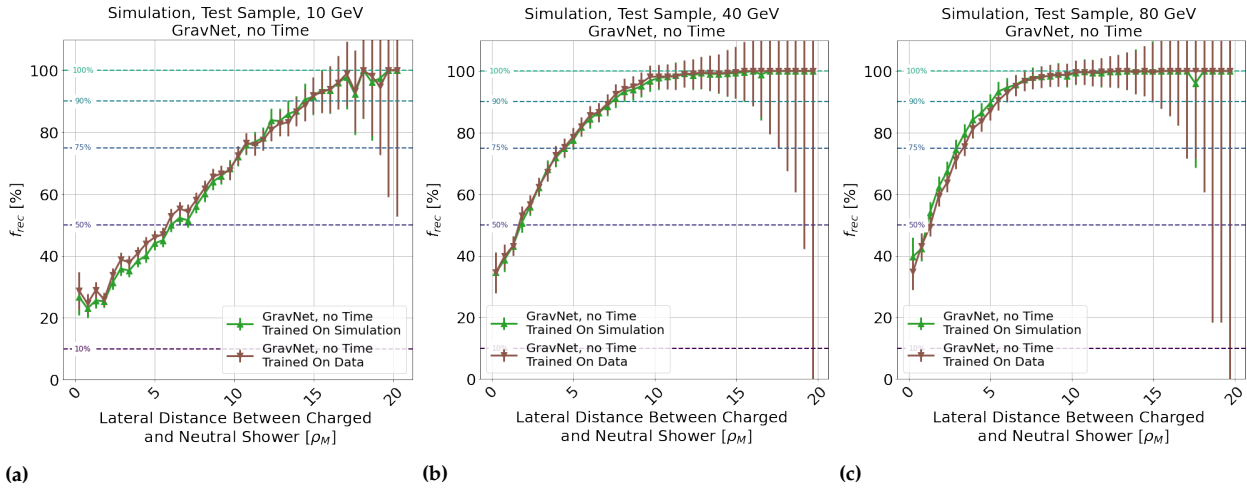
### 6.3.2.7 Skewness of Confusion Energy

As in Section 6.3.1.7, the medcouple as defined in Chapter 5 was calculated for all possible shower energy combinations for each combination of charged and neutral particle energies as in Figure 6.40. The results are shown in Figure 6.56

#### Summary Comment to Figure 6.56

All methods of shower separation tend to produce an 'altruistic' shower separation technique, where the higher energy shower 'donates' energy to the lower energy one. No significant differences between the model trained on simulation and on data were observed.

**Mean Fraction of Events Reconstructed Within Calorimeter Resolution vs. Lateral Shower Distance for 10 GeV, 40 GeV and 80 GeV Synthetic Neutrals Separated From 10-120 GeV Charged Showers for GravNet Trained On Simulation and Data, Applied To The Test Sample of Data**

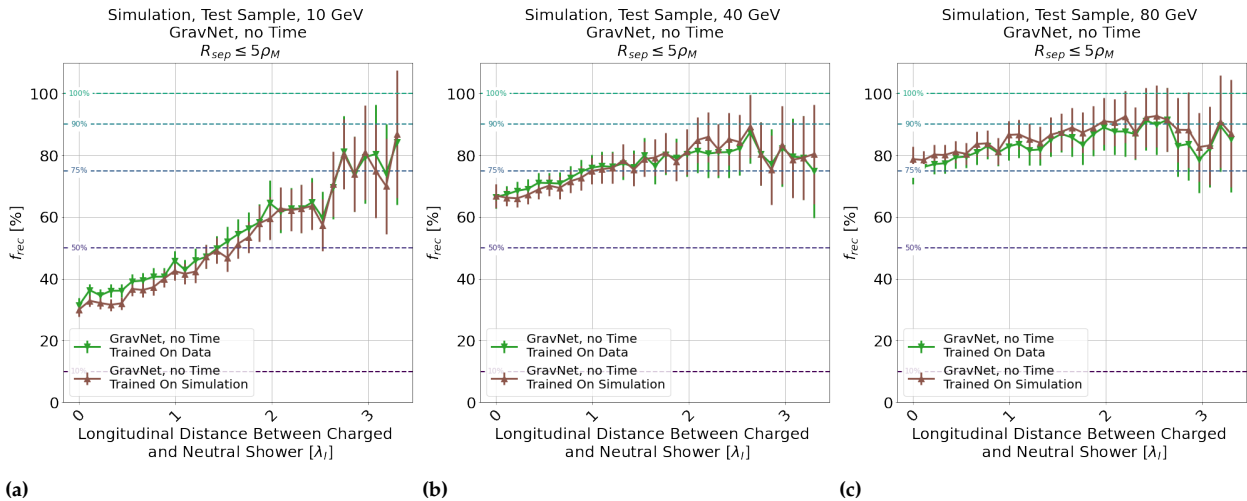


**Figure 6.54:** Figures 6.54a-6.54c, show the mean fraction of events reconstructed within the calorimeter resolution as a function of lateral shower distance for GravNet. The green and brown points show the models trained on simulation and data, respectively.

**Comments:**

- ▶ No significant differences between the model trained on simulation and on data were observed.
- ▶ Else, as in Figure 6.37.

**Mean Fraction of Events Reconstructed Within Calorimeter Resolution vs. Longitudinal Shower Distance for 10 GeV, 40 GeV and 80 GeV Synthetic Neutrals Separated From 10-120 GeV Charged Showers for GravNet Trained On Simulation and Data, Applied To The Test Sample of Data**

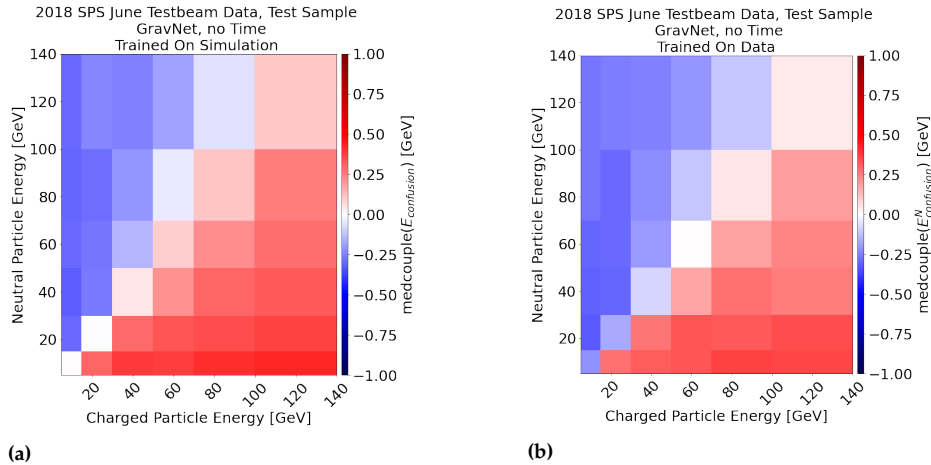


**Figure 6.55:** Figures 6.55a-6.55c, show the mean fraction of events reconstructed within the calorimeter resolution as a function of longitudinal shower distance (i.e. the shower starting position of both showers) for GravNet. The green and brown points show the models trained on simulation and data, respectively.

**Comments:**

- ▶ No significant differences between the model trained on simulation and on data were observed.
- ▶ Else, as in Figure 6.39.

### Matrices of Medcouple of Confusion Energy Distributions for Each Shower Separation Model Under Test, Applied to The Test Sample of Simulation



**Figure 6.56:** Figures 6.56a and 6.56b show the medcouple as a function of the charged and neutral particle energy, for PointNet, DGCNN and GravNet, respectively. The colour axis indicates the medcouple. Red to blue indicates right-to-left-tailed skewness, whereas white indicates no skewness. Red indicates that the confusion energy tends towards overestimation of the shower energy, while blue tends towards underestimation of the shower energy.

#### Comments:

- ▶ No significant differences between the model trained on simulation and on data were observed.
- ▶ Else, as in Figure 6.40.

#### 6.3.2.8 Software Compensation of Separated Neutral Hadron Showers

As in Section 6.3.1.9, SC is applied to the separated neutral shower in data, using the same method and cuts as described in that section. Around 81 % of events were reconstructed within the calorimeter resolution for both the model trained on simulation and data and therefore SC applies to the majority of events studied.

The uncompensated and compensated reconstructed neutral energy distributions are shown for all events in Figure 6.57 and for  $E_{\text{particle}}^N \geq E_{\text{particle}}^Q$  in Figure 6.58. The linearity and resolution for both cases are shown in Figure 6.59.

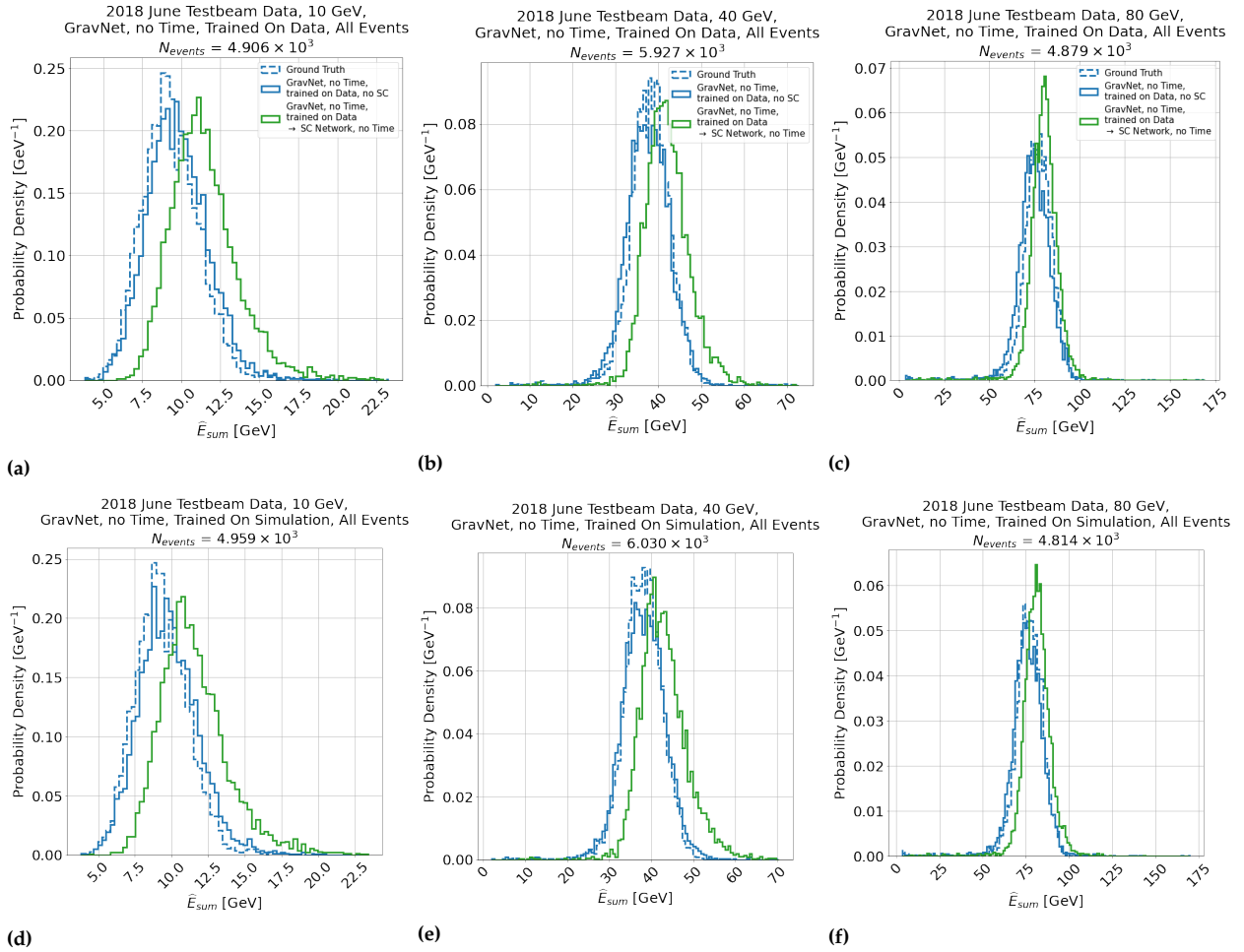
#### Summary Comment to Figures 6.57-6.59

The SC network trained in Chapter 5 improves resolution for separated neutral showers in data, indicating that the technique can be applied to separated hadron showers, and experiences lower bias for the  $E_{\text{particle}}^N > E_{\text{particle}}^Q$  sample than the sample where all events are used for the same reasons as discussed in Section 6.3.1.9. No significant differences between the model trained on simulation and on data were observed.

This result illustrates that both the shower separation and software compensation networks trained in and this chapter and Chapter 5 are applicable to experimental data. This means that both models can be

applied in an experimental setting, which therefore validates the work performed in both of those chapters.

**Neutral Reconstructed Energy Distributions Produced After Shower Separation By GravNet On Events Reconstructed Within The Calorimeter Resolution, Trained On Simulation And Data, After Separation, With and Without SC, with TCMT Cut Applied, For All Charged Hadron Shower Energies, Applied To 10 GeV, 40 GeV and 80 GeV Neutral Hadron Particle Energies In The Test Sample of Data**

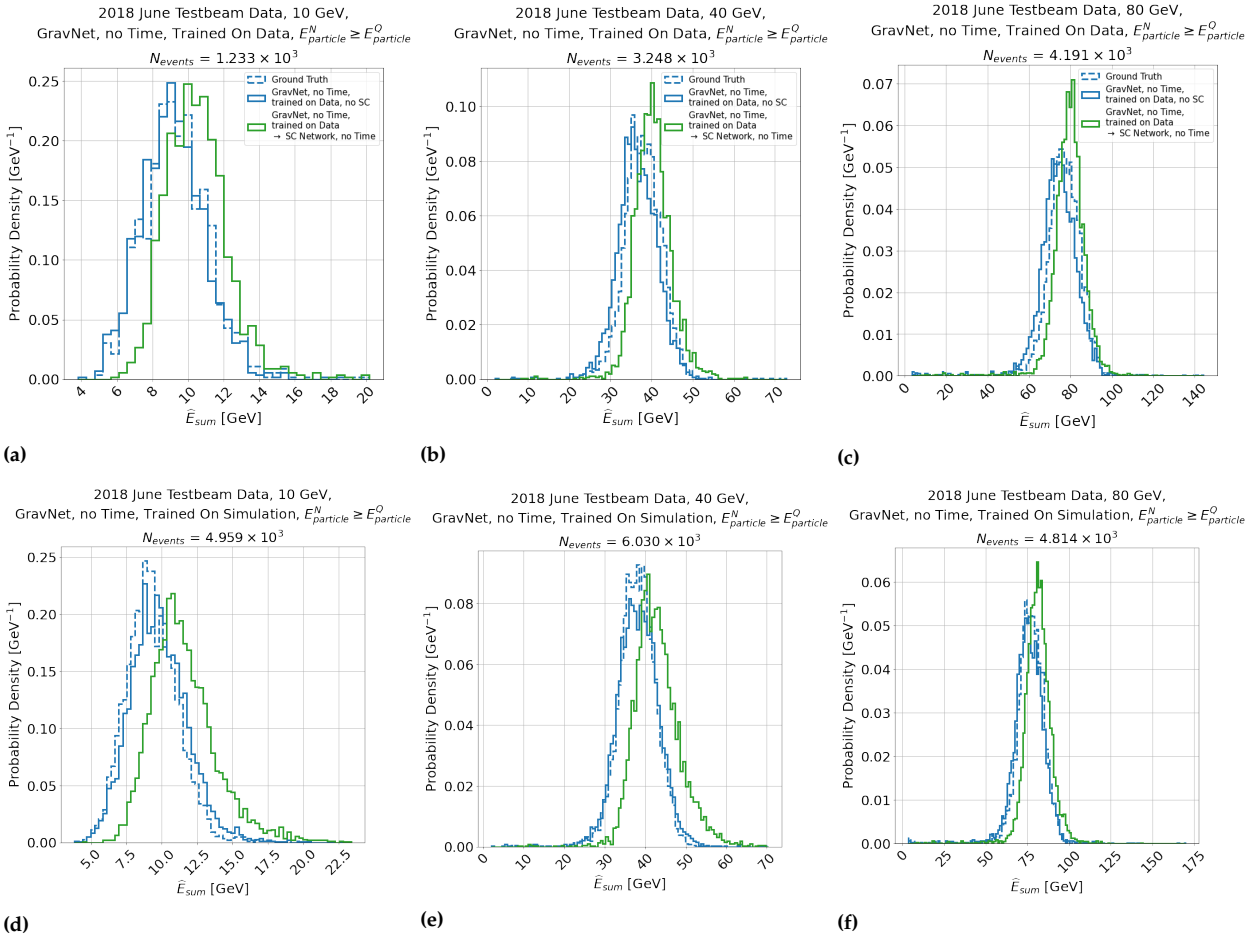


**Figure 6.57:** Reconstructed neutral hadron shower energy distributions for the GravNet network applied to the test sample of data, with no other selection criteria applied. The blue dashed line indicates the ground truth (perfectly known) reconstructed energy distribution. The green lines indicate the SC trained and applied without timing information on data, applied after separation. The top and bottom rows show the shower separation network trained on data and simulation. The purple dashed line indicates the particle energy.

**Comments:**

- ▶ No significant differences are observed in the response obtained for the shower separation model trained on simulation and data, both in qualitatively good agreement with the ground truth, indicated by the agreement of the dashed and solid blue lines.
- ▶ Comparison of the solid blue and green lines indicate that compensation is being applied, indicated by the reduced width of the distribution.
- ▶ Else, as in Figure 6.44.

**Neutral Reconstructed Energy Distributions Produced After Shower Separation By GravNet On Events Reconstructed Within The Calorimeter Resolution, Trained On Simulation And Data, After Separation, With and Without SC, with TCMT Cut Applied, For For  $E_{particle}^N < E_{particle}^Q$ , Applied To 10 GeV, 40 GeV and 80 GeV Neutral Hadron Particle Energies In The Test Sample of Data**



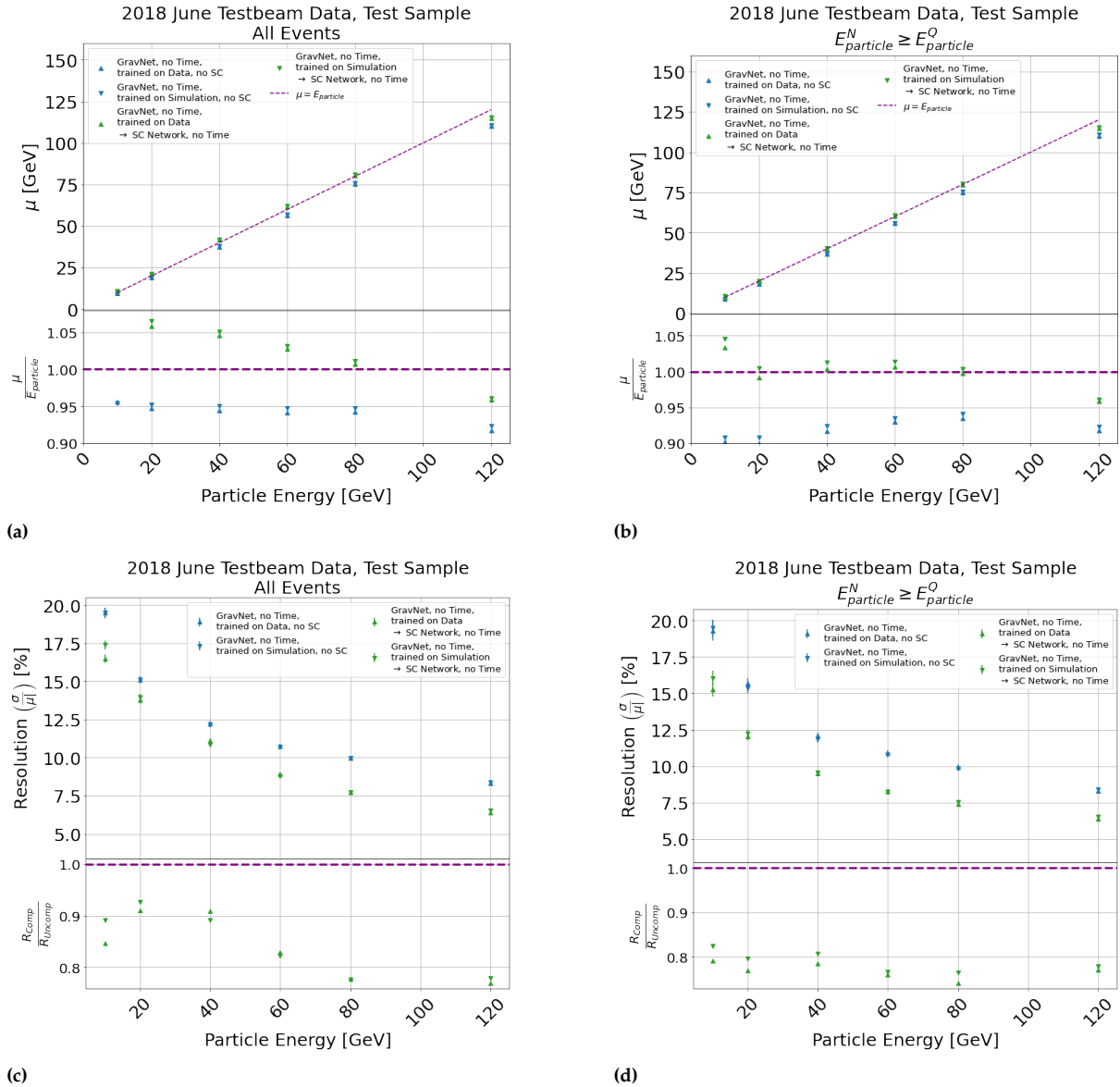
**Figure 6.58:** Reconstructed neutral hadron shower energy distributions for the GravNet network applied to the test sample of data, with no other selection criteria applied. The blue dashed line indicates the ground truth (perfectly known) reconstructed energy distribution. The green lines indicate the SC trained and applied without timing information on data, applied after separation. The top and bottom rows show the shower separation network trained on data and simulation. The purple dashed line indicates the particle energy.

**Comments:**

- Comparison of Figure 6.57 with this figure indicates that, overall, smaller bias and proportionally greater compensation are  $E_{particle}^N > E_{particle}^Q$  than when all events are included, which is the same conclusion as for Figure 6.45.
- Else, as in Figure 6.45.

### Linearity And Resolution of Separated Neutral Hadron Showers Obtained From GravNet Model With and Without SC, Applied To The Test Sample Of Data, Studied For All Events and Events Where

$$E_{\text{particle}}^N < E_{\text{particle}}^Q$$



**Figure 6.59:** Figures 6.59a and 6.59b and Figures 6.59c and 6.59d show the linearity and resolution of the neutral showers after separation by GravNet, using all events and for the case  $E_{\text{particle}}^N < E_{\text{particle}}^Q$ . The up and down arrows indicate the use of timing information in the shower separation model. The blue, green and red colours indicate the intrinsic calorimeter responses, and SC networks with and without timing information, after separation. Else, as in Figures 5.30a and 5.30b.

#### Comments:

- Figures 6.59c and 6.59d indicate that reconstruction of the event trained with or without timing information has a negligible effect on the resolution;
- Figure 6.59d indicates a maximum of a 20% improvement in the resolution is observed for the  $E_{\text{particle}}^N < E_{\text{particle}}^Q$  case.
- Else, as in Figure 6.46.

## 6.4 Conclusion

Three neural networks designed for the classification of point clouds were applied to the task of separating a charged and neutral hadron shower with the AHCAL highly-granular calorimeter in order to evaluate the shower separation performance of the calorimeter. Firstly, a method for producing synthetic neutral events by removal of the MIP-track was evaluated and found to produce events nearly indistinguishable from  $K_L^0$  hadrons. Secondly, a choice of the average lateral distance between hadron showers was considered, and a baseline distance at which 80 % of a hadron shower of a particular energy would be integrated was chosen as a baseline average distance. Synthetic two-shower events well contained within the AHCAL were then produced using a shower combination algorithm. The properties of the clustering by the neural network were evaluated, including the influence of timing information in the neural networks, both for simulation and data.

Firstly, it was observed that PointNet did not observe an improvement in resolution using timing information, whereas DGCNN and GravNet were capable of doing so. For the best-performing neural network (GravNet) this corresponded to a reduction of the MAD by around 23 %. By contrast, PointNet did not observe an improvement due to timing information. This result was tentatively attributed to the improved sensitivity of GravNet and DGCNN to 'local energy density' compared to PointNet, which does not exploit this information by design. This result was also compatible with the improvements observed with lateral and longitudinal shower separation distances, which suggested a minor to no correlation between the improvement due to timing information and these observables. This result tentatively suggests that the benefit of using timing information in shower separation algorithms is more sophisticated than was first anticipated. Specifically, the increased optimal  $\gamma$  hyperparameter of the GravNet model indicates that the decision boundary achieved captures more of the local structure, or substructure of the hadron shower using timing information than without. However, this result is speculation and requires further analysis beyond the scope of this study to fully verify.

Secondly, all models exhibited asymmetry in performance depending on whether or not the charged shower had more energy than the neutral one. For  $E_{\text{particle}}^N \geq E_{\text{particle}}^Q$ , 80-90 % of the events were reconstructed within the calorimeter resolution. For  $E_{\text{particle}}^Q > E_{\text{particle}}^N$ , model performance was observed to degrade. This can be attributed to the better performance achievable with a disparity between track position and the centre-of-gravity of the most energetic shower. This is rarely the case where the charged shower has more energy than the neutral. This result was compatible with observed trends between the confusion energy, the charged shower energy and the difference between the charged track and the centre of gravity. Additionally, timing information was found to explicitly increase the number of events reconstructed correctly of the case  $E_{\text{particle}}^Q > E_{\text{particle}}^N$  by 15-20 %, making a strong case for the temporal sensitivity of AHCAL for the use of shower separation, and also suggests that the model exploits track information during clustering. This result may be inferred with certainty that timing information provides a better



description of the hadron shower development than without it.

Thirdly, a comparison was made to a previous shower separation study using AHCAL and Pandora PFA in [78] of the fraction of events reconstructed within 1.5 calorimeter resolutions was evaluated as a function of inter-shower distance in this chapter and from that study. The results indicate significant improvements in the performance of the tested models compared to Pandora PFA. For example, the number of events reconstructed within the AHCAL resolution increased by up to an additional 25% for 10 GeV charged showers separated from 10 GeV synthetic neutral hadron showers, and by a maximum of 35-45% for 30 GeV charged showers separated from 10 GeV synthetic neutral showers. Although the models are trained with the assumption of the presence of both a charged and synthetic neutral hadron shower in the event, these findings tentatively suggest that the clustering capacity of GravNet and DGCNN, applied to Particle Flow shower separation, outperforms Pandora PFA in the task of two-shower separation. Further investigation could involve integrating the neural networks studied in this chapter as a module within Pandora PFA and evaluating the reconstruction performance with and without the module.

Fourthly, all models exhibited skewness in the confusion energy distributions, which suggested that all models independently learned an 'altruistic' energy allocation model, which preferentially allocated more confusion energy to the lower of the two energy showers than the one with more. It was noted that the same distribution is observed for the same particle energies and types in a complementary study using Pandora PFA, tentatively suggesting the models autonomously learned a similar strategy as humans have designed.

Fifthly, the SC network of Chapter 5 was applied to the separated neutral showers. Improvements in the resolution were observed at all energies, but biasing was observed due to the inclusion of confusion energy. This is a consequence of the strong influence confusion energy can have on the energy density of hadron showers with comparatively low energy ( $< 40$  GeV). It is suggested that this could be resolved by including energy fractions in the SC models.

Finally, the studies on simulation were repeated for 2018 June SPS Test-beam data, and the simulation and data-trained models were compared. No significant difference was observed between the model trained on simulation and the model trained on data regarding performance and properties. SC was also applied to data and achieved similar performance as on simulation.

In summary, this study suggests that the AHCAL calorimeter is a highly effective Particle Flow Calorimeter, enhanced by the temporal sensitivity of the detector, and the models presented in Chapter 5 and this chapter can be applied to experimental data.



## 7.1 Summary of Research

7.1 Summary of Research . . . . .	255
7.2 Future Prospects . . . . .	257
7.3 Closing Remark . . . . .	259

Future lepton collider experiments are required to achieve a jet energy resolution of around 3% in the jet energy range from 50-200 GeV to for precision particle physics measurements. Particle Flow (PF) is expected to attain this challenging requirement by utilising excellent tracking, highly granular calorimeters and sophisticated clustering algorithms to associate charged tracks with energy deposits in the calorimeters.

The CALICE AHCAL Fe-Sc prototype is a highly-granular calorimeter prototype designed for PF, with around 22,000 individual SiPM-on-tile readout channels per  $m^2$ , and can measure the energy density of hadron showers in both space and time.

All detector and event reconstruction aspects of the AHCAL must be optimised to achieve these challenging jet energy resolution requirements, which was the primary motivation for this thesis.

The research topics and conclusions of each chapter are therefore summarised henceforth:

- High-precision calorimeter calibration requires single-channel calibration parameters. Therefore, two generally-available software tools, `LightSimTastic` and `Peak0Tron`, were developed to simulate and automatically characterise the charge spectra of SiPMs for this purpose.

In Chapter 3, `LightSimTastic` was demonstrated to provide a highly flexible simulation tool for SiPMs that can be used to study the influence of different physics models and detector properties on the device's response.

In Chapter 4, `Peak0Tron` was demonstrated to accurately measure critical SiPM parameters, such as the gain and pedestal position, and non-standard parameters, like after-pulse probability, after-pulse time-constant and dark count rate, from charge spectra. These tools have a straightforward application in calibrating and monitoring the individual channels of the AHCAL calorimeter. For instance, `Peak0Tron` can be used to measure and account for effects such as decreasing breakdown voltage with increases in temperature during the running of AHCAL. However, notably, the tools developed can also be used in applications beyond the AHCAL and therefore represents a helpful addition to the field of SiPM characterisation.

- The HCAL resolution plays a role in jet energy resolution for jet energies below 50 GeV and for the improved association of energy deposits from charged particles to charged tracks. However, the AHCAL is a non-compensating calorimeter. This means that

the effects of stochastic fluctuations in the energy measurement must be reduced by specially designed software compensation (SC) algorithms that equalise the electromagnetic and hadronic response of the calorimeter.

In Chapter 5, a neural network method was developed to overcome the limitations of energy biasing in software compensation models. This was achieved by blinding the model to the overall energy of the hadron shower and instead focusing the attention of the network explicitly on local distributions of energy in the AHCAL during a hadron shower. This is in contrast to many standard SC methods, which typically result in bias to the training range due to overtraining due to limited particle energies.

The application of the SC network resulted in calorimeter resolutions for simulation of  $40.2\%/\sqrt{E_{\text{particle}}}\oplus 2.2\%$  and  $37.2\%/\sqrt{E_{\text{particle}}}\oplus 2.4\%$ , without and with timing information. The network without timing information trained on CALICE 2018 SPS testbeam data also achieved a comparable resolution of  $41.9\%/\sqrt{E_{\text{particle}}}\oplus 4.0\%$ . In both cases, the neural network outperformed the control method.

Explicit dependencies of the compensation on the development of the hadron shower in space and time were observed, in line with expectations from physics. The model also highlighted energy regimes in agreement with the expectations from literature, with no prior information than experimental data, and highlighted differences between the energy densities of simulation and data. This method directly improves the hadronic energy resolution and, therefore, the clustering capability of the AHCAL. In addition, the improved performance with timing information motivates the temporal sensitivity of AHCAL.

- The jet energy resolution in PF for jet energies above 50 GeV is influenced most heavily by confusion between energy deposits. Of particular interest is reducing the confusion charged and neutral hadron showers. Sophisticated clustering techniques are required to utilise the five-dimensional sensitivity of the AHCAL for this purpose. Machine learning has been demonstrated to accurately cluster energy deposits of hadron showers in highly granular calorimeters. However, existing studies have not explicitly included critical information available to AHCAL, such as track information for the charged shower and timing information.

In Chapter 6, neural network methods from the literature were applied to the hadron shower separation of a single charged and single synthetic neutral hadron shower to evaluate the additional clustering performance that can be achieved using temporal clustering information. The experiment was carefully evaluated, accounting for data synthesis techniques such as inter-shower distance and the validity of generating synthetic hadron shower events from charged ones by topological cuts. The neural networks were then trained to attribute a fraction of energy to each hadron shower in the event.

The neural networks demonstrated that AHCAL is a highly effective Particle Flow Calorimeter and that upwards of 90% of events

can be reconstructed within the AHCAL calorimeter resolution in simulation and data in most cases where the neutral shower has more energy than the charged shower. Timing information was highly beneficial for cases where the charged shower had more energy than the neutral shower, increasing the number of events within the resolution by 10-15 %. Additionally, while the study's limitations play a role in the interpretation of the result, the performance of the machine learning models was demonstrated to outperform Pandora PFA at the task of shower separation.

The importance of timing information was found to be more complex than first anticipated, indicated by the capacity of graph neural networks to exploit timing information compared to a point-based model without access to local-energy density information, which showed no substantial improvement with timing information. Bias was observed, where the model allocated energy preferentially from the higher to the lower-energy shower. This was noted as the case for Pandora PFA in an adjacent study. It indicates that the neural networks learned to preferentially 'split up' energy deposits rather than merge them. The energy separation capabilities of the neural network were also superior regarding the lateral resolution to this external study.

The SC model of Chapter 5 was successfully applied after shower separation, indicating a promising new possibility of training both energy reconstruction and shower separation algorithms in tandem. The models trained on simulation performed the same as those trained on data, meaning that the shower separation algorithms can be applied in an experimental setting. This study demonstrates the performance of the AHCAL detector and again motivates temporal sensitivity as an observable.

These results are interpretable within the caveats of only applying to the case of a charged shower separated from a synthetic neutral one and that the model performs no track-labelling.

## 7.2 Future Prospects

Finally, a selection of prospects for further study is presented which could use this work as a basis.

**AHCAL Calibration with Peak0Tron** A topic of future research is the application of Peak0Tron to LED calibration spectra taken from AHCAL. This study was already performed in [34], but did not include the effects of detector noise in the model. It would therefore be of interest to determine the influence of SiPM noise effects and changes in the values obtained for the gain and pedestal position. It would also prove a vital validation study for the method presented in this thesis.

**End-to-End Machine-Learning-based Particle Flow** Two modules of Particle Flow have been assessed in this thesis: software compensation and shower separation. Both have been proven to be successful in their intended purpose. However, there are clear limitations to the study. A true PFA would ultimately result in ‘Particle Flow Objects’ (PFOs), high-level objects condensing the combined information about an event from the trackers and calorimeter. The study in this thesis has focused only on limited subsets of this problem under specific conditions. Additionally, the influence of detector noise on the effectiveness of training these models has not been assessed.

Contemporary developments have proposed a method of Particle Flow reconstruction called ‘*object condensation*’ [197]. The fundamental idea behind this is similar to standard Particle Flow algorithms: to condense the properties of a hadron shower into a single object. The advantage of this method is that the number of showers is not fixed and that an arbitrary number of additional observables that are typically submodules of a PFA would be condensed into a single function. Another advantage is that it implicitly includes the possibility of identifying detector noise.

Based on this method and using a simplified version of the CMS Highly Granular Calorimeter (HGCal) proposed for High-Luminosity LHC (HL-LHC), the proposed successor of the LHC as a calorimeter, the study of [198] developed an end-to-end PFA utilising the GravNet model with an object condensation approach for PF. The method demonstrated successful single  $e^-$ ,  $\gamma$  and  $\pi^+$  energy reconstruction, and even transverse momenta of reconstructed  $q\bar{q} \rightarrow t\bar{t}$  induced-jets. Notably, timing information, software compensation, PID and tracking information were not included in the study.

A clear prospect is to implement this method for AHCAL. Many of the basic ingredients (e.g. fake neutrals, shower combination techniques, the GravNet model) have already been established in Chapter 6 and can be readily repurposed to produce a more complex event topology and implement the missing ingredients into this study. However, one challenge would be the application of software compensation in this framework. As has been demonstrated in Chapter 5, energy reconstruction is highly sensitive to over-training in the single-particle case, and it is unclear how to resolve the opposed approaches of blinding the model to the reconstructed energy for learning compensation weights to produce an unbiased, generalised SC algorithm, while simultaneously producing object-like representations of particles in an event using object condensation. A possible solution is to train the SC algorithm to operate on multiple hadron showers (i.e. before separation) and use it as an input into the object condensation network. This may be possible due to the locally-defined nature of the model.

**Generative Design for Particle Flow** The effect of detector noise on confusion is not trivial. In particular, it is unclear if it influences the ideal AHCAL calorimeter design (i.e. cell size, absorber thicknesses). A remarkable approach to this problem is to utilise a fully end-to-end machine-learning-based PFA to optimise these quantities as ‘hyperparameters’. This would allow a complete study into the optimal detector design to achieve PF.

The procedure of optimising a device's characteristics or design using on an optimisation problem is called '*generative design*'. Generative design has been successfully used in physics cases to optimise instrumentation. For instance, [199] used generative design to optimise the shape of the magnetic coils of stellarators for nuclear fusion research. This study is mentioned as it is analogous in its complexity and stringent design requirements to those of a PF calorimeter.

Such an undertaking would require accurate simulations of jets and particle interactions with matter, models of detector noise, carefully defined design constraints and a PFA that can be entirely optimised using arbitrary detector geometry and configurations. An end-to-end ML-based PFA capable of identifying detector noise that could be trained to perform best with fundamentally different detector designs could be one ingredient for the potential of generative design to be applied to collider experiments and confirm if the current specifications of AHCAL remain optimal. However, this is speculation and would require considerable effort, yet it remains an intriguing prospect.

### 7.3 Closing Remark

This thesis has presented promising developments in traditional and artificial intelligence-enhanced detector calibration techniques for the highly granular AHCAL hadron calorimeter, from the sub-detector level to high-level reconstruction algorithms. It is hoped that the methods and techniques developed and evaluated in this thesis reach their full potential and can be used to achieve state-of-the-art jet-energy resolution in future lepton collider experiments.

Software developed in this thesis is available from <https://gitlab.desy.de/jack.rolph>.





## 8.1 Peak0Tron: A Python Tool For SiPM Characterisation

### 8.1.0.1 Geiger-Discharge Probability for After-pulses

In Equation 4.9, the time dependence of the Geiger-breakdown probability during the recharging of the pixel for a primary Geiger discharge at  $t = 0$  is parameterised by Equation 8.1:

$$p_{\text{Geiger}}(t) = (1 - e^{-t/\tau_{\text{rec}}}) \quad (8.1)$$

where  $\tau_{\text{rec}}$  is the recovery time constant. This appendix discusses the motivation for this parameterisation and how  $\tau_{\text{rec}}$  can be estimated from data. For a primary discharge at  $t = 0$ , the time dependence of the voltage over the pixel is given by Equation 8.2:

$$V_d(t) = V_{\text{off}} + (V_b - V_{\text{off}}) \cdot (1 - e^{-t/\tau}), \quad (8.2)$$

where  $V_b$  is the bias voltage,  $V_{\text{off}}$  the voltage at which the Geiger discharge stops, and  $\tau$  the time constant of the slow component of the SiPM pulse.

In Section 4.4.2,  $\mu(V_b)$  has been determined for the Hamamatsu MPPC and the Ketek SiPM and fitted using Equation 4.22. The values of the parameters  $\mu_0$  and  $V_0$  from the fit, and of  $V_{\text{off}}$ , are given in Table 4.7.

Using  $\mu(V_b)$  in Equation 4.22 and  $V_d(t)$  in Equation 8.2, the Geiger-discharge probability at time  $t$  relative to the saturation value for the bias voltage  $V_b$  is estimated by Equation 8.3:

$$p_{\text{Geiger}}^{\text{rel}}(t) \approx \frac{\mu(V_d(t))}{\mu(V_b)}. \quad (8.3)$$

It is noted that the spatial distribution of the charge carriers which produce after-pulses is very different from the distribution of the charge carriers from light with a wavelength of 400 nm (violet light). As the Geiger-breakdown probability depends on the position at which a charge carrier is generated, one may doubt the validity of Equation 8.3 for after-pulses. However, if the shape of the voltage dependence of the Geiger-breakdown probability is approximately independent of position, this approach is valid.

Figure 8.1 shows  $p_{\text{Geiger}}^{\text{rel}}$  for several  $V_b$  values using Equation 8.3 with the parameters of Table 4.7, as continuous lines for the Ketek and Hamamatsu SiPMs, respectively.  $p_{\text{Geiger}}^{\text{rel}}$  only approximately follow  $(1 - e^{-t/\tau_{\text{rec}}})$  dependencies, which are shown by the dashed lines. The value

8.1 Peak0Tron: A Python Tool For SiPM Characterisation . . . . .	261
8.2 Software Compensation for AH-CAL using Graph Neural Networks . . . . .	264
8.3 Shower Separation in Five Dimensions Using Machine Learning . . . . .	276

of  $\tau_{\text{rec}}(V_b)$  is obtained by demanding that  $p_{\text{Geiger}} = p_{\text{Geiger}}^{\text{rel}}$  at  $t = 0$  and  $t = \tau$ , which gives Equation 8.4:

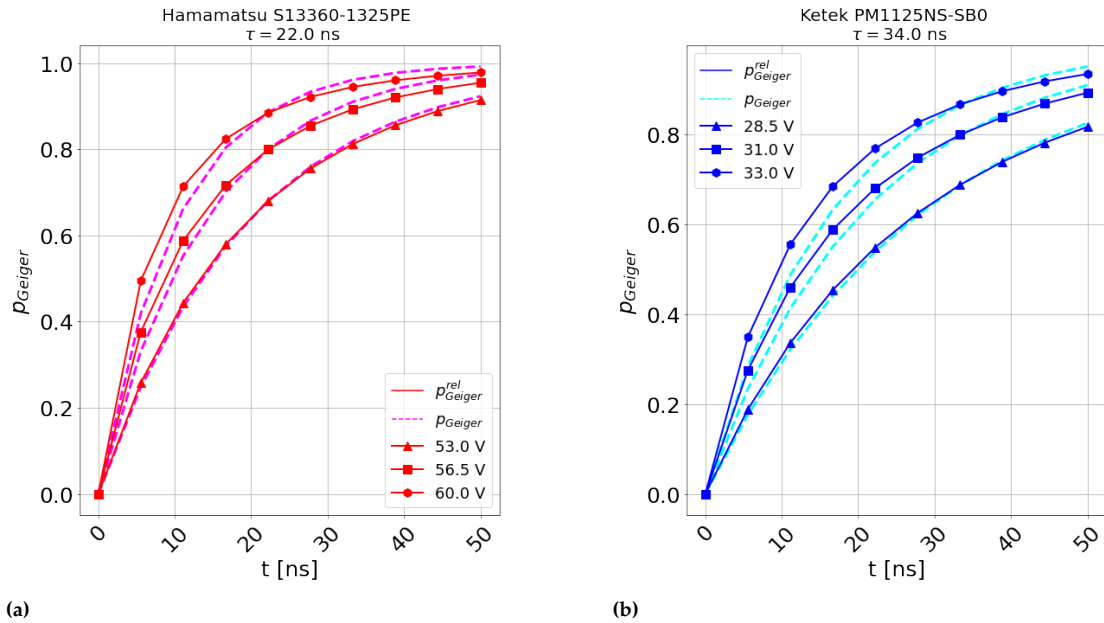
$$\tau_{\text{rec}}(V_b) = \frac{\tau}{\ln\left(e^{\frac{V_b - V_{\text{bd}}}{V_0}} - 1\right) - \ln\left(e^{\frac{V_b - V_{\text{bd}}}{V_0 e}} - 1\right)}. \quad (8.4)$$

Figure 8.2 shows  $f_{\text{rec}}$ , the ratio  $\tau_{\text{rec}}/\tau$ , for the Hamamatsu MPPC and the Ketek SiPM as a function of over-voltage using the parameters of Table 4.7.

It can be seen that  $\tau_{\text{rec}}$  decreases with over-voltage, and that at a given over-voltage,  $\tau_{\text{rec}}$  for the Ketek SiPM, which has  $\tau = 34$  ns, is 5 to 10 % lower than for the Hamamatsu MPPC with  $\tau = 22$  ns.

The values of  $\tau_{\text{rec}} = f_{\text{rec}} \cdot \tau$  shown in Figure 8.2 were used for the fits presented in Section 4.3.6. If  $f_{\text{rec}}$  is not known, Peak0Ttron will use the value 0.65. If  $f_{\text{rec}}$  is known, the user can set its value. It is concluded that the parameterisation of Equation 8.1 provides an appropriate description of the decrease of the Geiger-breakdown probability for after-pulses.

### Geiger Breakdown Probability As a Function Of Time Measured Using Peak0Tron For The Hamamatsu and Ketek SiPMs Under Test

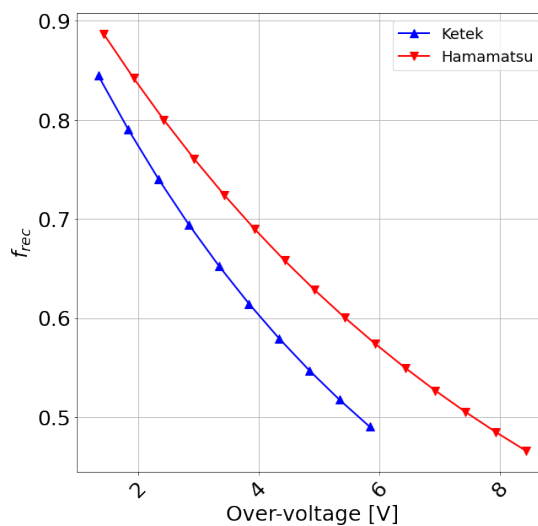


**Figure 8.1:** Relative Geiger-breakdown probability as a function of the time of the after-pulse after the primary Geiger discharge for different bias voltages,  $V_b$ . The continuous lines are the values determined using Equation 8.1 and the dashed lines the parametrisation  $1 - e^{-t/\tau_{rec}}$  with the  $\tau_{rec}$  values of Figure 8.2, in Figure 8.1a the Hamamatsu MPPC at  $V_b = 53\text{ V}$ ,  $56.5\text{ V}$  and  $60\text{ V}$ , and in Figure 8.1b for the Ketek SiPM at  $V_b = 28.5\text{ V}$ ,  $31\text{ V}$ , and  $33\text{ V}$ .

**Comments:**

- Good agreement is observed between the approximation made by Equation 8.1 and the measured Geiger discharge probability by Peak0Tron.

### Measured Ratio of $\tau_{rec}$ to $\tau$ As a Function Of Over-voltage For The Hamamatsu and Ketek SiPMs Under Test



**Figure 8.2:** The over-voltage dependence of  $f_{rec} = \tau_{rec}/\tau$  for the Hamamatsu MPPC ( $\tau = 22\text{ ns}$ ) and the Ketek SiPM ( $\tau = 34\text{ ns}$ ) determined, as described in the text.

**Comments:**

- The measured ratio shows that as over-voltage increases, so too does the recovery time constant  $\tau_{rec}$  decrease.

## **8.2 Software Compensation for AHCAL using Graph Neural Networks**

$E_{\text{particle}}$ [GeV]	Run Number	June 2018 SPS Testbeam Data			Total Events
		Testing	Training	Validation	
10	61265	1768	14595	1827	18190
	61267	1607	12382	1534	15523
	61268	1577	12593	1564	15734
20	61313	1520	12203	1547	15270
	61269	1904	15094	1742	18740
	61270	1813	15078	1901	18792
	61271	2047	16372	2043	20462
	61272	1796	14595	1867	18258
40	61273	1879	14373	1886	18138
	61274	2590	20666	2643	25899
	61275	2683	22146	2738	27567
	61276	2797	21626	2596	27019
	61277	1235	10278	1321	12834
60	61278	1079	8348	1085	10512
	61259	2005	16578	2060	20643
	61260	2810	22985	2834	28629
	61262	2396	19111	2447	23954
	61263	2099	16159	2038	20296
80	61394	1893	15237	1919	19049
	61395	2020	15712	1925	19657
	61279	2315	18356	2239	22910
	61280	2216	18233	2273	22722
	61281	2407	19440	2412	24259
120	61282	2473	19453	2429	24355
	61283	2255	17843	2313	22411
	61284	2812	22129	2776	27717
	61287	2589	21256	2622	26467
	61288	2713	21086	2650	26449
	61289	2599	21230	2665	26494
Total Events		61897	495157	61896	618950

**Table 8.1:** Table of events used from CALICE June 2018 SPS Testbeam data, split by particle energy and run number, used for training, validation and testing of SC models.

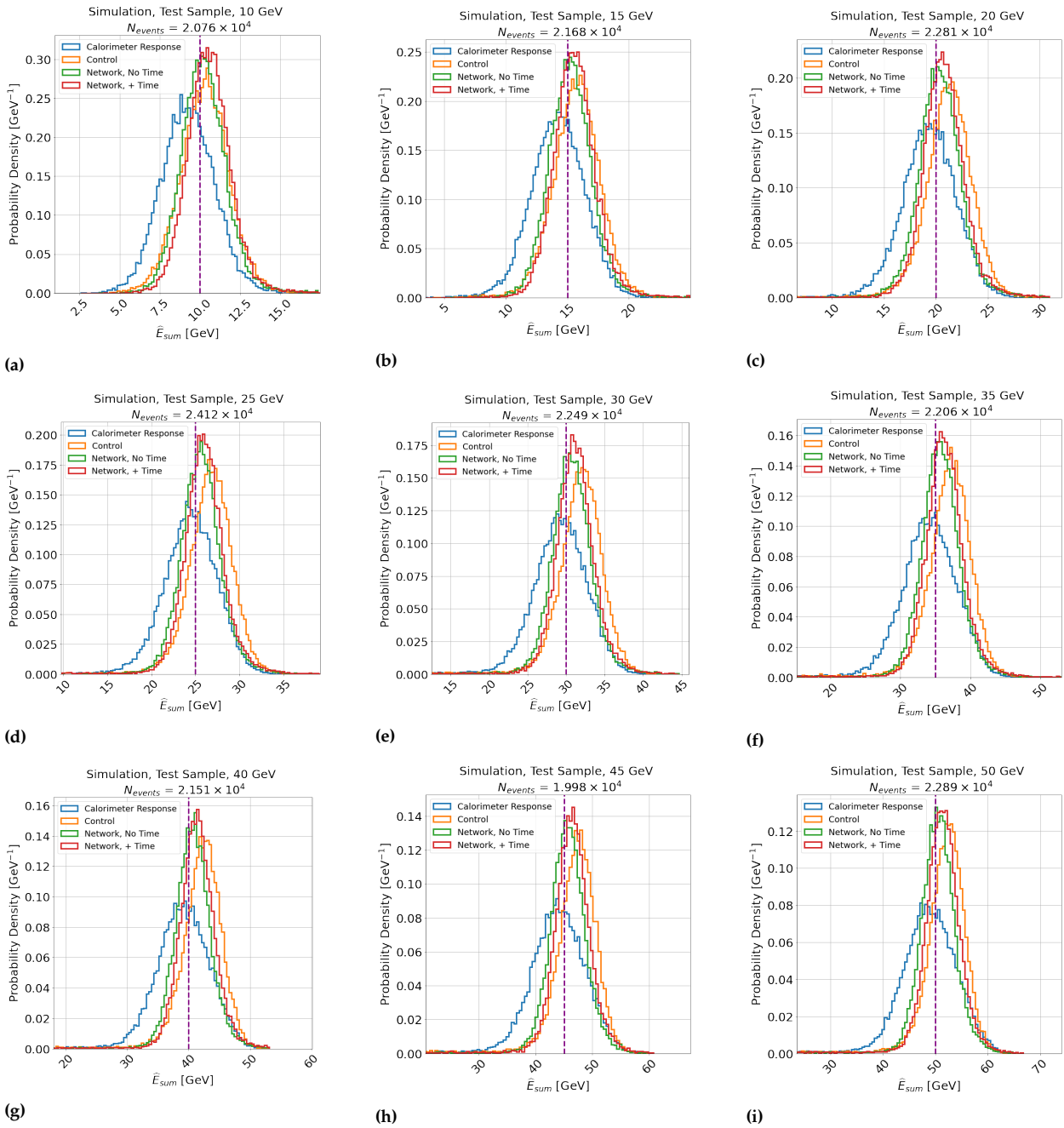
Bin Range [MIP]	Simulation			Bin Range [MIP]	June 2018 SPS Testbeam Data		
	$\alpha_b$	$\beta_b$	$\gamma_b$		$\alpha_b$	$\beta_b$	$\gamma_b$
0.500 - 0.735	-1.120	13.000	-5.499	0.500 - 0.770	-9.252	17.803	-16.047
0.735 - 1.002	-0.372	5.134	-4.727	0.770 - 1.059	-8.529	14.738	-11.256
1.002 - 1.272	-0.350	2.523	-3.054	1.059 - 1.351	-0.046	1.902	-1.797
1.272 - 1.585	-0.460	1.999	-1.587	1.351 - 1.698	1.283	-0.379	0.057
1.585 - 2.013	-0.290	1.634	-1.012	1.698 - 2.179	2.104	-1.325	0.720
2.013 - 2.631	0.022	1.275	-0.949	2.179 - 2.875	1.025	0.384	-0.235
2.631 - 3.584	0.218	1.138	-0.632	2.875 - 3.957	1.271	0.001	0.104
3.584 - 5.328	0.387	0.658	-0.457	3.957 - 5.847	1.325	0.013	0.304
5.328 - 9.881	0.579	0.259	-0.281	5.847 - 10.930	0.814	0.426	0.021
9.881 - $\infty$	0.788	-0.057	0.043	10.930 - $\infty$	0.148	1.032	-0.243

(a)

(b)

**Table 8.2:** Bin ranges and weights obtained for the control method described in Section 5.2.2. Table 8.2a shows the values obtained for simulation. Table 8.2b shows the values obtained for CALICE June 2018 SPS Testbeam data.

### Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To Simulation with TCMT Cut

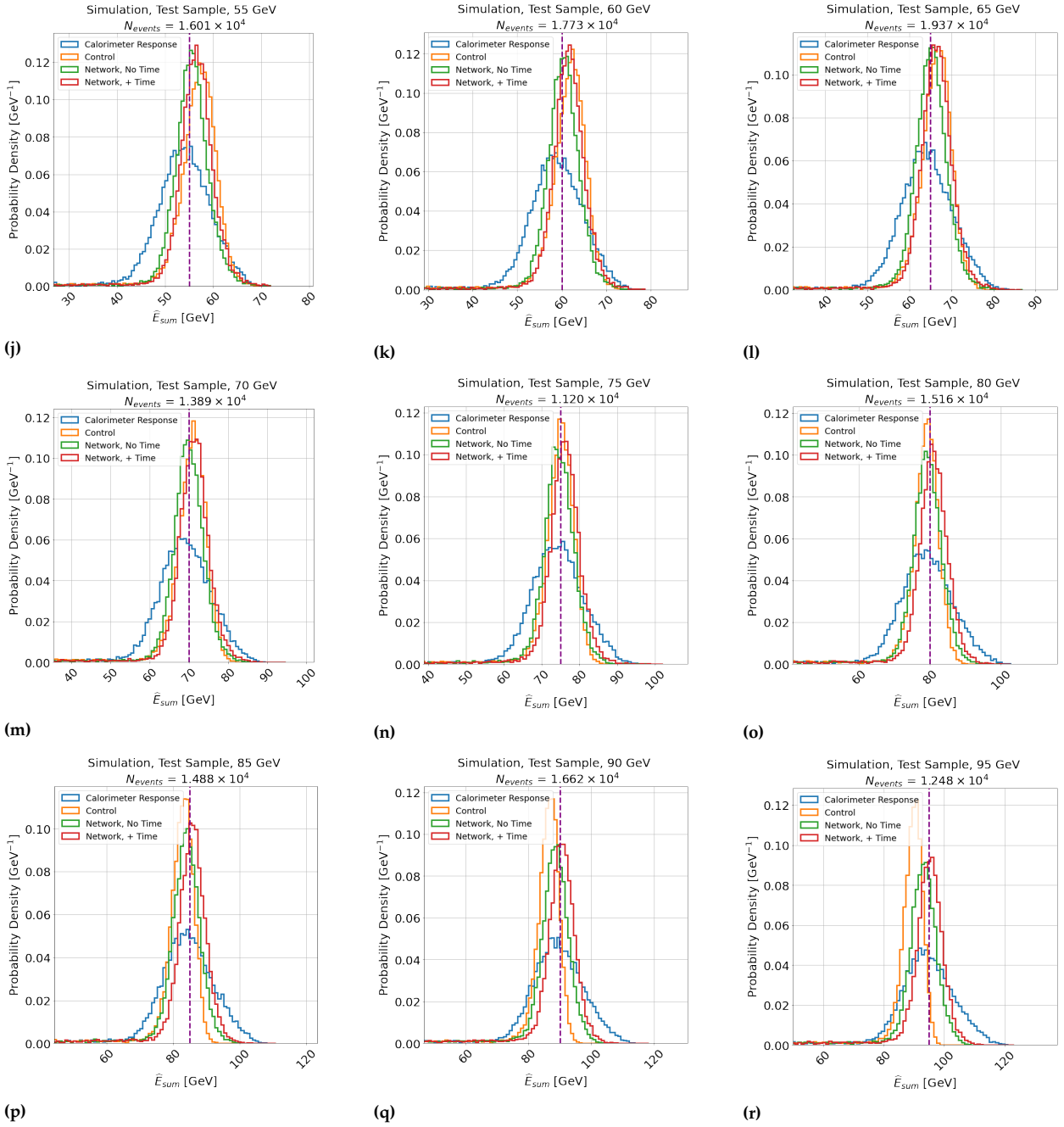


**Figure 8.3:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the simulation test dataset of Table 5.1, with the TCMT cut applied. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods, respectively.  $E_{particle}$  is indicated as a dashed purple line.

**Comments:**

- As in Figure 5.24.

**Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To Simulation with TCMT Cut**

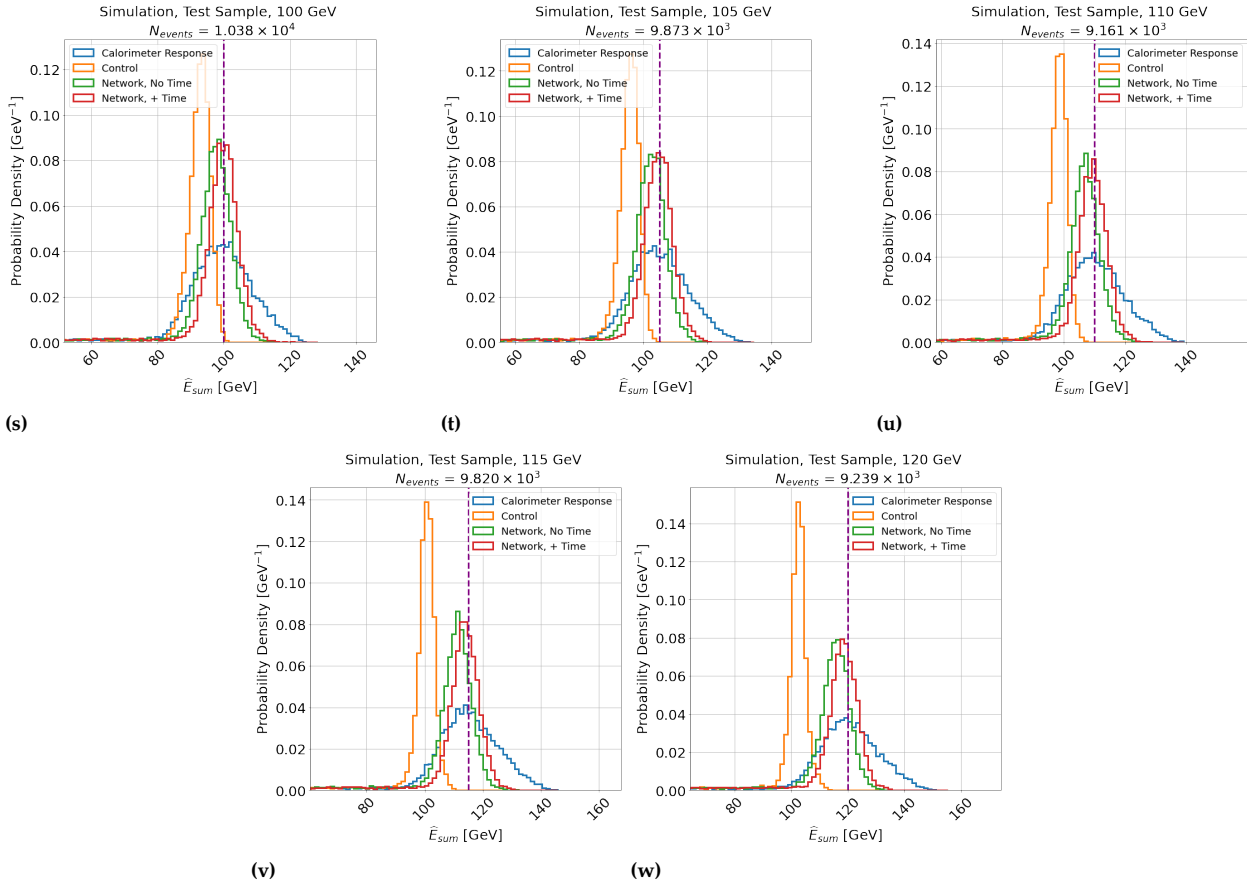


**Figure 8.3 continued:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the test dataset of Table 5.1, with the TCMT cut applied. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods, respectively.  $E_{particle}$  is indicated as a dashed purple line.

**Comments:**

- As in Figure 5.24.

**Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To Simulation with TCMT Cut**



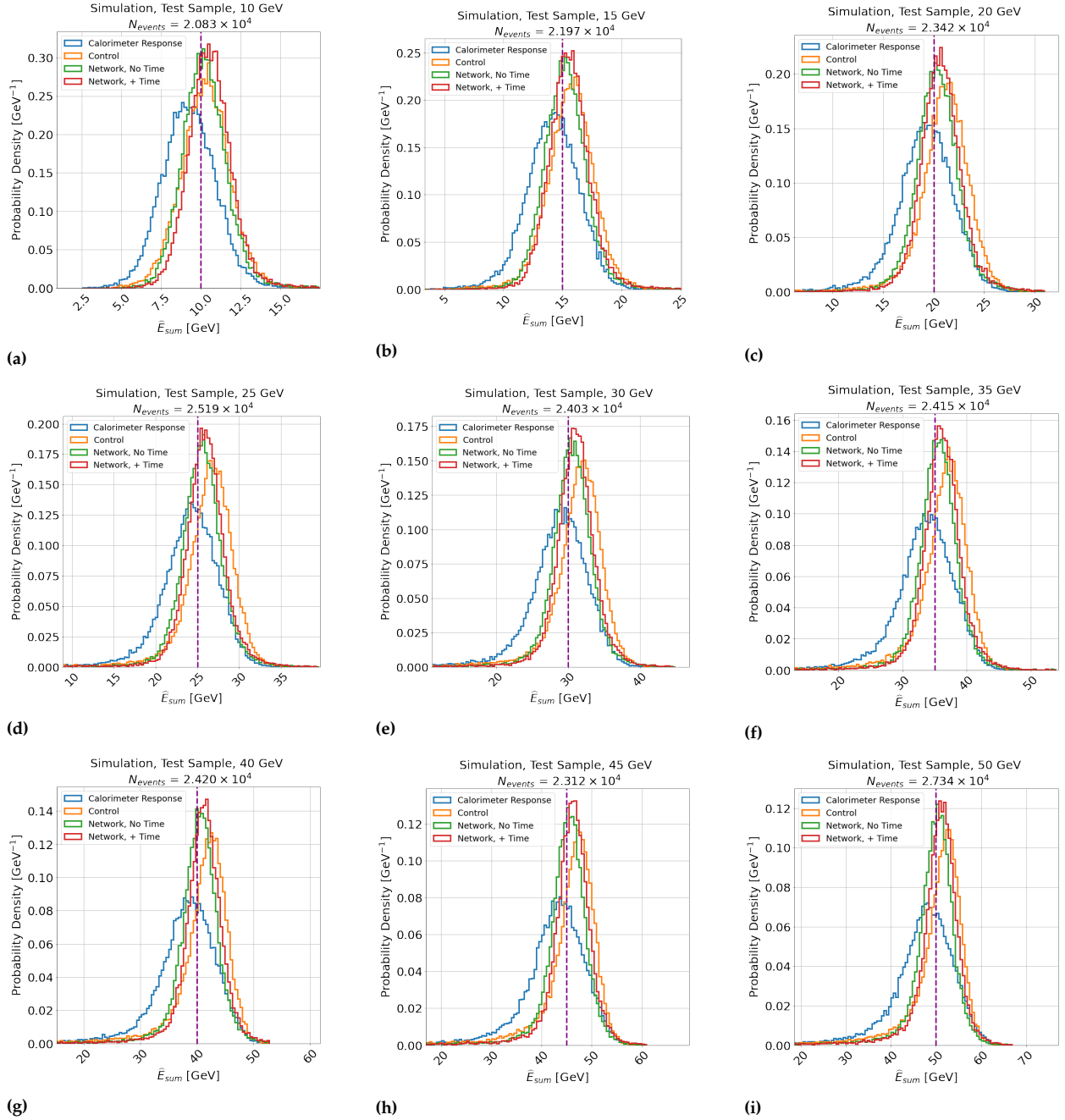
**Figure 8.3 continued:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the test dataset of Table 5.1, with the TCMT cut applied. Blue lines indicate intrinsic calorimeter response, while orange, green and red lines indicate the control, network without and network with time compensation methods, respectively.  $E_{particle}$  is indicated as a dashed purple line.

**Comments:**

- As in Figure 5.24.



**Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To Simulation, without the TCMT Cut Applied**

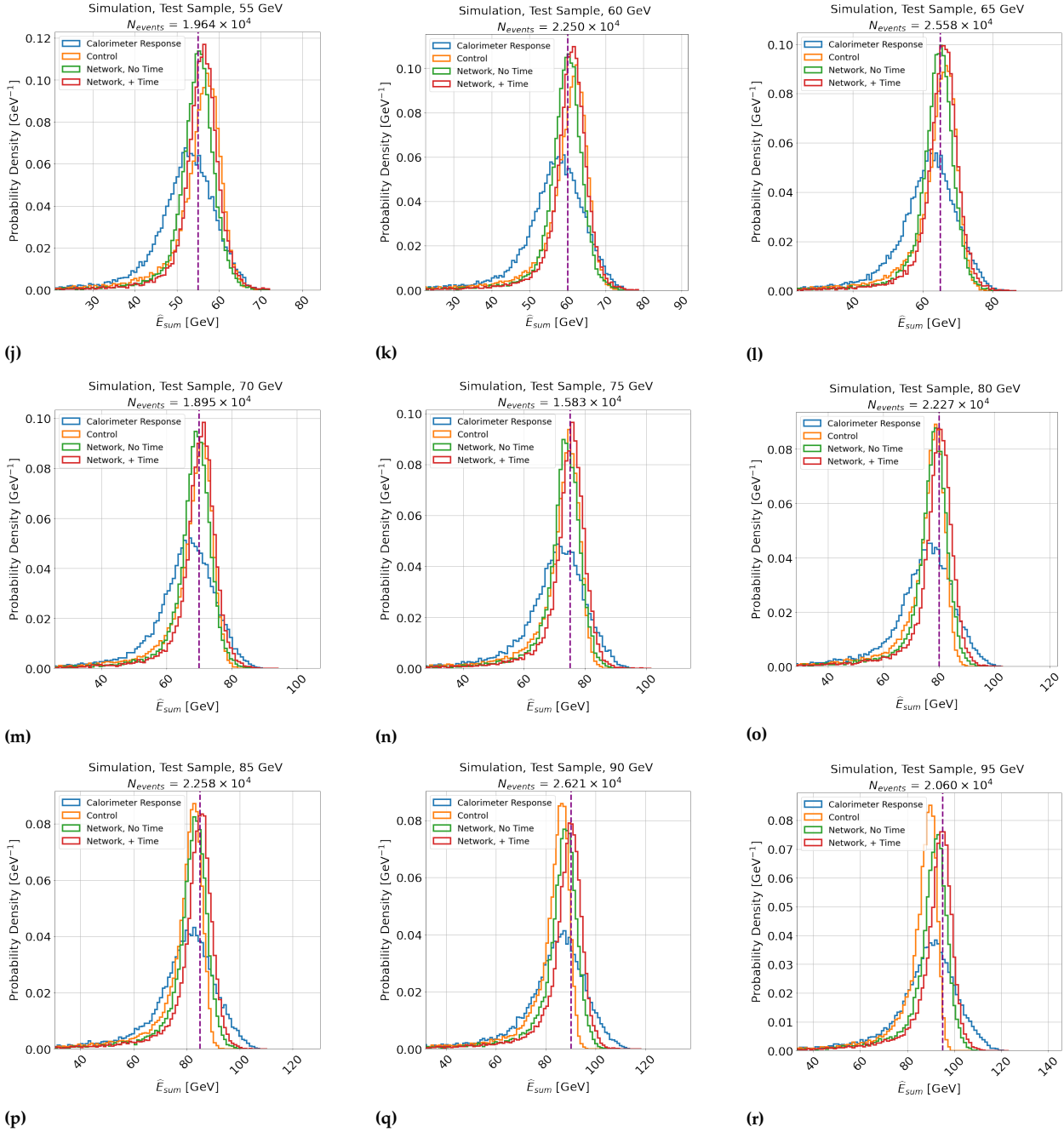


**Figure 8.4:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the simulation test dataset of Table 5.1, without the TCMT cut applied. Else, as in Figure 8.4.

**Comments:**

- As in Figure 5.27.

**Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To Simulation, without the TCMT Cut Applied**

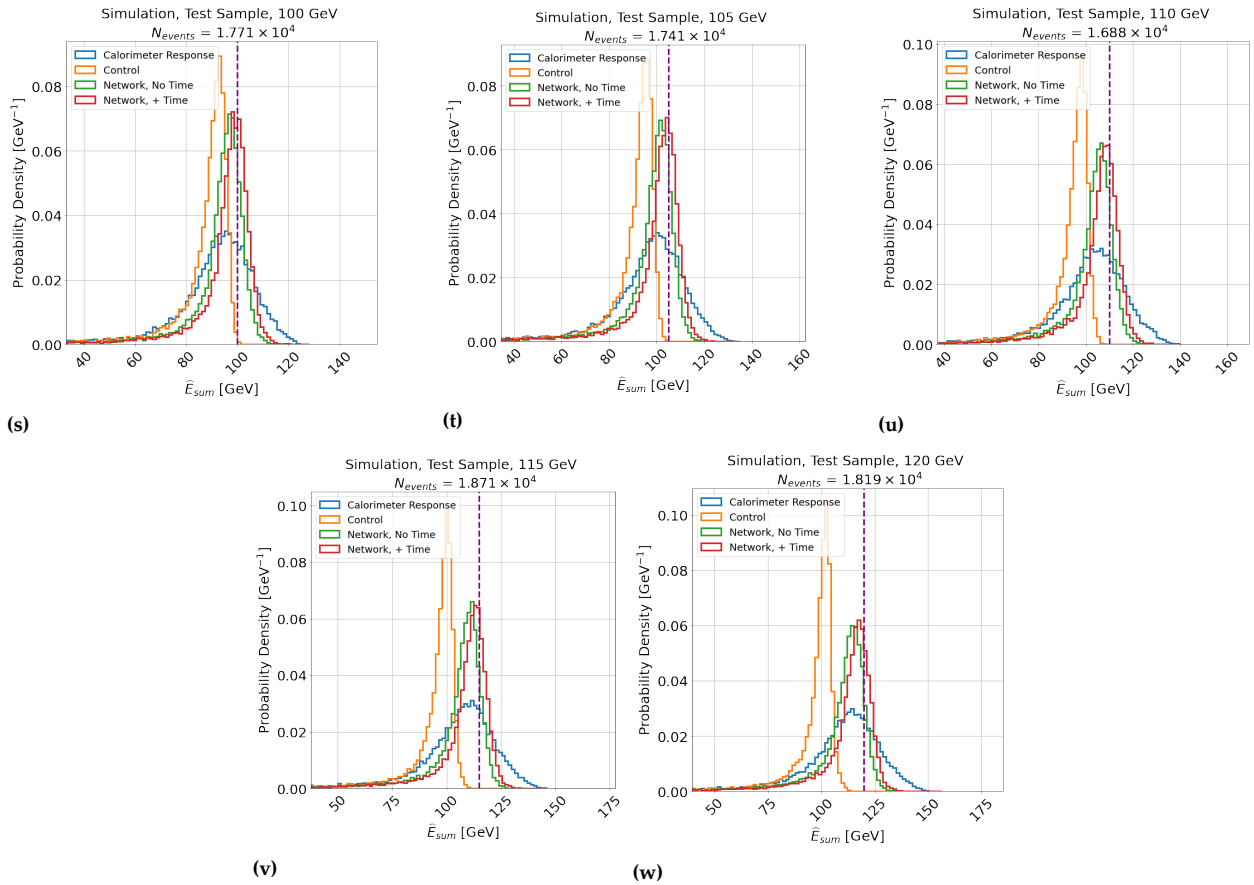


**Figure 8.4 continued:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the simulation test dataset of Table 5.1, without the TCMT cut applied. Else, as in Figure 8.4.

**Comments:**

- ▶ As in Figure 5.27.

**Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To Simulation without the TCMT Cut Applied**

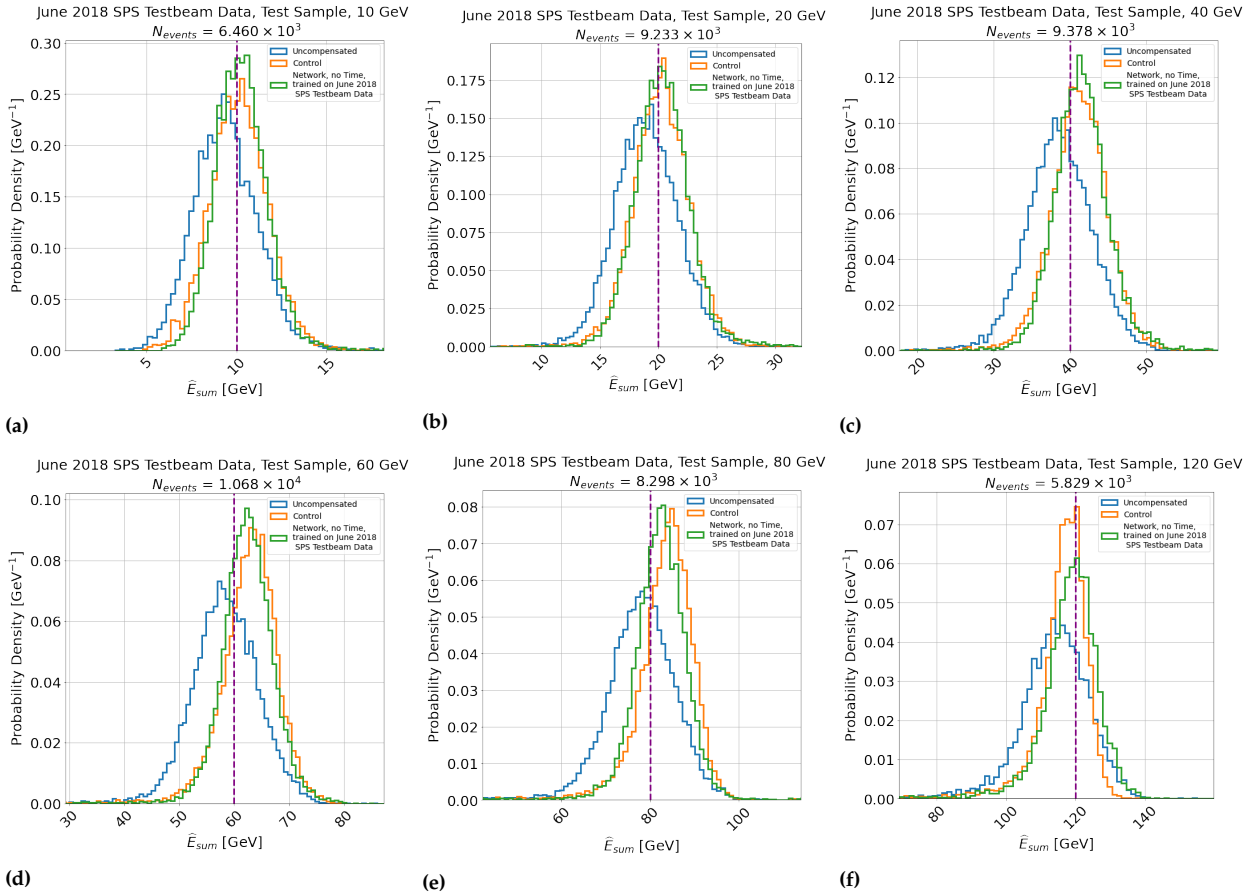


**Figure 8.4 continued:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the simulation test dataset of Table 5.1, without the TCMT cut applied. Else, as in Figure 8.4.

**Comments:**

- As in Figure 5.27.

### Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To 2018 June Testbeam Data with the TCMT Cut Applied

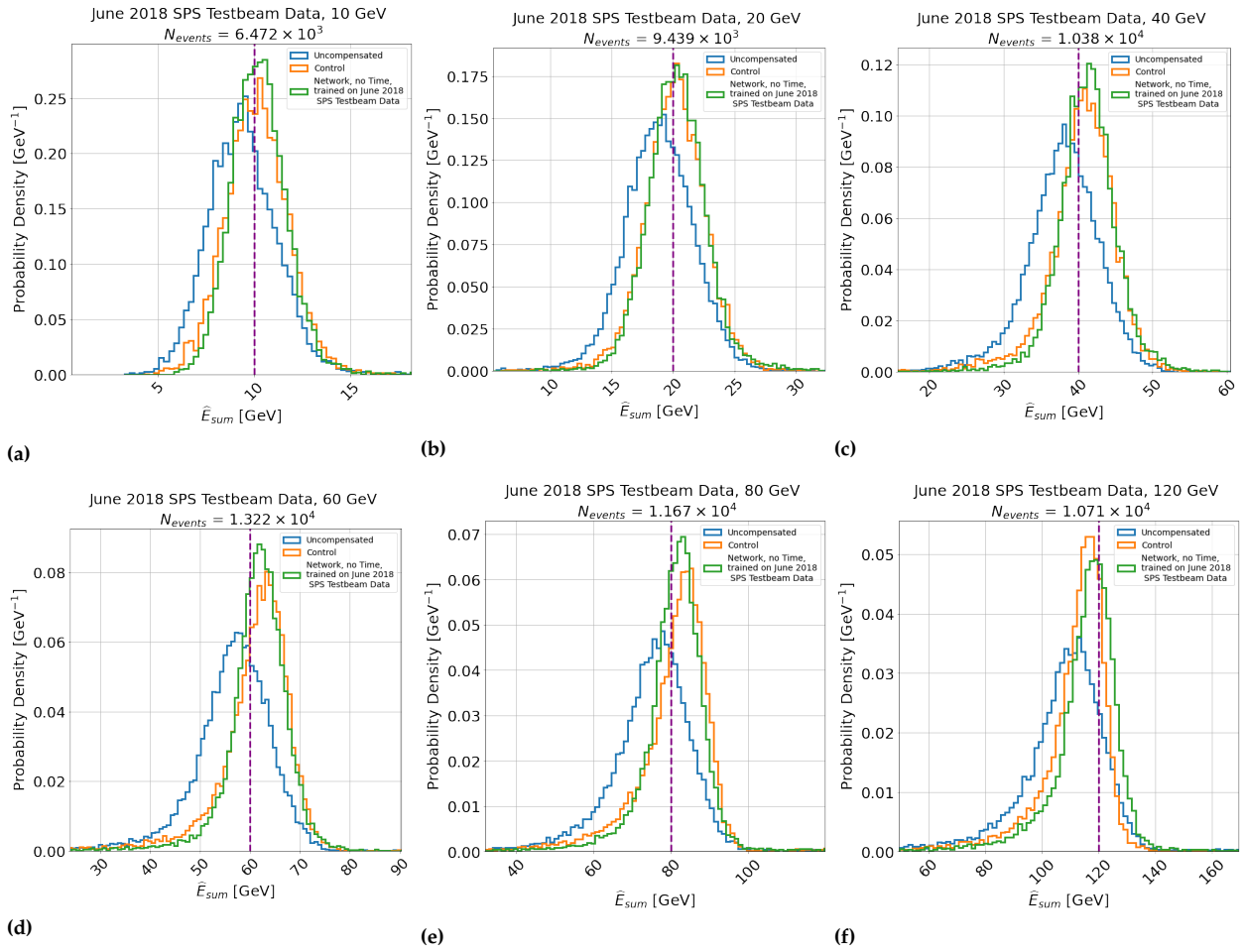


**Figure 8.5:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the test dataset of data of Table 5.1, with the TCMT cut applied. Blue lines indicate intrinsic calorimeter response, while orange and green lines indicate the control and network without time compensation methods.  $E_{particle}$  is indicated as a dashed purple line.

**Comments:**

- As in Figure 5.29.

**Reconstructed Energy Distributions, Before and After SC, Trained On and Applied To 2018 June Testbeam Data without the TCMT Cut Applied**



**Figure 8.6:** Example normalised histograms showing the simulated calorimeter response before and after compensation applied to the test dataset of data of Table 5.1, without the TCMT cut applied. Else, as in Figure 8.5.

**Comments:**

- As in Figure 5.34.

$E_{\text{particle}}$ [GeV]	Simulation $\mu$				Simulation $d\mu$			
	CR	CTRL	NN,-Time	NN,+Time	CR	CNTRL	NN,-Time	NN,+ Time
10	9.214	10.281	10.157	10.502	0.013	0.012	0.010	0.009
15	14.257	15.754	15.186	15.512	0.016	0.014	0.012	0.012
20	19.401	21.307	20.326	20.692	0.019	0.015	0.013	0.013
25	24.516	26.796	25.569	25.954	0.021	0.016	0.015	0.014
30	29.330	31.990	30.661	31.136	0.024	0.018	0.017	0.016
35	34.230	37.160	35.688	36.247	0.027	0.020	0.018	0.018
40	39.199	42.311	40.656	41.368	0.030	0.021	0.019	0.019
45	44.148	47.374	45.651	46.468	0.034	0.023	0.022	0.021
50	49.119	52.316	50.526	51.466	0.034	0.022	0.021	0.021
55	54.066	57.157	55.555	56.583	0.043	0.028	0.027	0.026
60	59.033	61.864	60.352	61.535	0.045	0.027	0.026	0.026
65	64.099	66.412	65.148	66.438	0.045	0.026	0.026	0.026
70	69.145	70.838	69.905	71.315	0.057	0.031	0.032	0.032
75	74.178	74.941	74.578	76.158	0.066	0.035	0.037	0.037
80	79.235	78.972	79.270	80.992	0.061	0.029	0.033	0.032
85	84.506	82.781	83.912	85.779	0.064	0.029	0.034	0.033
90	89.475	86.410	88.615	90.592	0.063	0.027	0.033	0.032
95	94.663	89.725	93.157	95.238	0.078	0.031	0.039	0.039
100	99.804	92.822	97.723	99.969	0.088	0.034	0.044	0.044
105	104.983	95.722	102.337	104.685	0.095	0.032	0.047	0.047
110	110.224	98.338	106.931	109.416	0.103	0.031	0.050	0.049
115	115.567	100.585	111.316	113.949	0.106	0.028	0.050	0.049
120	120.000	102.513	115.874	118.618	0.101	0.028	0.053	0.052

(a)

$E_{\text{particle}}$ [GeV]	Simulation $\sigma$				Simulation $d\sigma$			
	CR	CTRL	NN,-Time	NN,+Time	CR	CNTRL	NN,-Time	NN,+ Time
10	1.625	1.474	1.311	1.255	0.012	0.011	0.009	0.008
15	2.112	1.792	1.651	1.579	0.015	0.012	0.010	0.010
20	2.512	2.027	1.873	1.795	0.018	0.013	0.011	0.011
25	2.905	2.294	2.128	2.016	0.019	0.013	0.013	0.012
30	3.320	2.527	2.351	2.258	0.021	0.015	0.014	0.013
35	3.754	2.779	2.576	2.487	0.023	0.016	0.015	0.015
40	4.151	2.905	2.721	2.642	0.025	0.017	0.016	0.015
45	4.518	3.081	2.961	2.858	0.027	0.018	0.018	0.018
50	4.946	3.192	3.092	3.023	0.027	0.018	0.017	0.017
55	5.253	3.316	3.296	3.182	0.033	0.022	0.023	0.022
60	5.701	3.335	3.416	3.328	0.034	0.021	0.022	0.022
65	6.059	3.406	3.598	3.510	0.034	0.021	0.022	0.022
70	6.406	3.367	3.695	3.628	0.044	0.023	0.027	0.027
75	6.736	3.376	3.840	3.751	0.051	0.027	0.032	0.032
80	7.174	3.297	3.917	3.862	0.046	0.022	0.027	0.027
85	7.540	3.167	4.021	3.925	0.048	0.021	0.029	0.029
90	7.843	3.078	4.111	4.042	0.045	0.020	0.028	0.028
95	8.297	2.943	4.237	4.141	0.059	0.021	0.033	0.034
100	8.553	2.803	4.354	4.318	0.062	0.023	0.036	0.037
105	9.049	2.741	4.425	4.422	0.071	0.023	0.039	0.039
110	9.341	2.664	4.588	4.488	0.078	0.024	0.044	0.041
115	9.832	2.603	4.666	4.643	0.081	0.026	0.040	0.041
120	10.000	2.493	4.813	4.742	0.027	0.028	0.045	0.044

(b)

$E_{\text{particle}}$ [GeV]	Reduced $\chi^2$			
	CR	CTRL	NN,-Time	NN,+Time
10	1.372	1.295	1.271	0.888
15	0.713	1.585	1.504	1.442
20	1.323	1.568	2.049	2.076
25	1.192	2.083	2.304	2.774
30	1.241	2.114	1.563	2.433
35	1.445	2.924	2.554	2.925
40	1.059	4.017	2.351	1.930
45	1.507	4.508	2.317	3.282
50	1.155	5.076	3.070	3.439
55	1.533	4.386	3.182	3.192
60	1.552	4.851	2.798	3.629
65	2.093	5.700	3.268	3.897
70	1.848	4.928	3.018	3.406
75	1.842	4.609	2.920	3.279
80	2.637	5.149	3.208	3.371
85	3.641	6.326	3.737	4.290
90	3.908	7.194	4.236	4.416
95	3.823	7.087	3.968	4.226
100	4.303	6.677	3.269	3.695
105	3.658	6.090	4.212	3.961
110	4.316	4.584	4.485	4.469
115	5.032	4.759	4.460	4.318
120	4.872	5.095	4.749	4.555

(c)

**Table 8.3:** Table of  $\mu$  and  $\sigma$  from the Gaussian fits performed on the SC models trained on simulation in Section 5.3.1. Tables 8.3a, 8.3b and 8.3c show the  $\mu$ ,  $\sigma$  and their errors, and reduced  $\chi^2$ , as a function of particle energy for each studied method applied to the testing dataset. CR, CTRL, NN,-Time and NN,+Time are abbreviations of: 'intrinsic calorimeter response', 'control method', 'neural network, without time' and with 'neural network, with time', respectively.

June 2018 SPS Testbeam Data								
$E_{\text{particle}}$ [GeV]	$\mu$				$d\mu$			
	CR	CTRL (Data)	NN <sub>r</sub> -Time (Data)	NN <sub>r</sub> -Time (Sim)	CR	CNTRL (Data)	NN <sub>r</sub> -Time (Data)	NN <sub>r</sub> -Time (Sim)
10	9.334	10.113	10.190	9.949	0.025	0.022	0.018	0.019
20	18.947	20.213	20.244	19.048	0.031	0.026	0.025	0.024
40	38.623	41.278	41.355	38.301	0.047	0.038	0.036	0.033
60	58.501	63.090	62.503	57.331	0.060	0.047	0.043	0.038
80	78.246	84.081	82.514	75.260	0.081	0.062	0.058	0.049
120	114.690	117.347	119.229	107.497	0.121	0.080	0.092	0.077

(a)

June 2018 SPS Testbeam Data								
$E_{\text{particle}}$ [GeV]	$\sigma$				$d\sigma$			
	CR	CTRL (Data)	NN <sub>r</sub> -Time (Data)	NN <sub>r</sub> -Time (Sim)	CR	CNTRL (Data)	NN <sub>r</sub> -Time (Data)	NN <sub>r</sub> -Time (Sim)
10	1.733	1.554	1.363	1.407	0.023	0.019	0.016	0.016
20	2.614	2.260	2.151	2.094	0.029	0.023	0.022	0.021
40	4.124	3.456	3.198	3.012	0.042	0.033	0.031	0.029
60	5.702	4.526	4.202	3.750	0.050	0.039	0.036	0.031
80	7.058	5.093	5.144	4.381	0.065	0.046	0.048	0.040
120	8.734	5.203	6.611	5.532	0.096	0.062	0.077	0.059

(b)

June 2018 SPS Testbeam Data				
$E_{\text{particle}}$ [GeV]	Reduced $\chi^2$			
	CR	CTRL (Data)	NN <sub>r</sub> -Time (Data)	NN <sub>r</sub> -Time (Sim)
10	1.310	1.009	1.103	1.659
20	0.976	0.736	0.413	1.163
40	0.685	1.211	1.753	1.119
60	0.907	3.111	1.314	1.358
80	0.922	4.382	1.907	1.881
120	1.968	6.185	2.974	2.332

(c)

**Table 8.4:** Table of  $\mu$  and  $\sigma$  from the Gaussian fits performed on the SC models trained on data in Section 5.3.2. Tables 8.4a, 8.4b and 8.4c show the  $\mu$ ,  $\sigma$  and their errors, and reduced  $\chi^2$ , as a function of particle energy for each studied method applied to the testing dataset. Else, as in Table 8.3.

### **8.3 Shower Separation in Five Dimensions Using Machine Learning**



$E_{\text{particle}}$	Run Number	June 2018 SPS Testbeam Data			Total Events
		Testing	Training	Validation	
10	61265	6035	48303	5852	60190
	61267	5065	40849	5197	51111
	61268	5228	41604	5230	52062
20	61313	5068	40410	5117	50595
	61269	6356	50808	6337	63501
	61270	6597	51576	6433	64606
	61271	7109	56050	7011	70170
	61272	6113	49783	6136	62032
40	61273	6046	49545	6303	61894
	61274	8575	69057	8446	86078
	61275	9171	72997	9038	91206
	61276	8886	71835	9176	89897
	61277	4249	33473	4135	41857
	61278	3547	28062	3633	35242
	61259	7013	55785	7103	69901
60	61260	9616	77097	9587	96300
	61262	8037	64325	8039	80401
	61263	7004	55398	6812	69214
	61394	6277	51200	6460	63937
	61395	6653	52994	6599	66246
	61279	7349	59005	7392	73746
80	61280	7334	58841	7267	73442
	61281	7963	62809	7839	78611
	61282	7939	63362	7896	79197
	61283	7205	58298	7395	72898
	61284	9111	72223	9058	90392
	61287	8602	69014	8527	86143
120	61288	8499	68864	8476	85839
	61289	8669	68943	8819	86431
Total Events		205316	1642510	205313	2053139

**Table 8.5:** Table of events used from CALICE June 2018 SPS Testbeam data, split by particle energy and run number, used for training, validation and testing of SC models.

Event Level: Distributions of Reconstructed Shower Energy

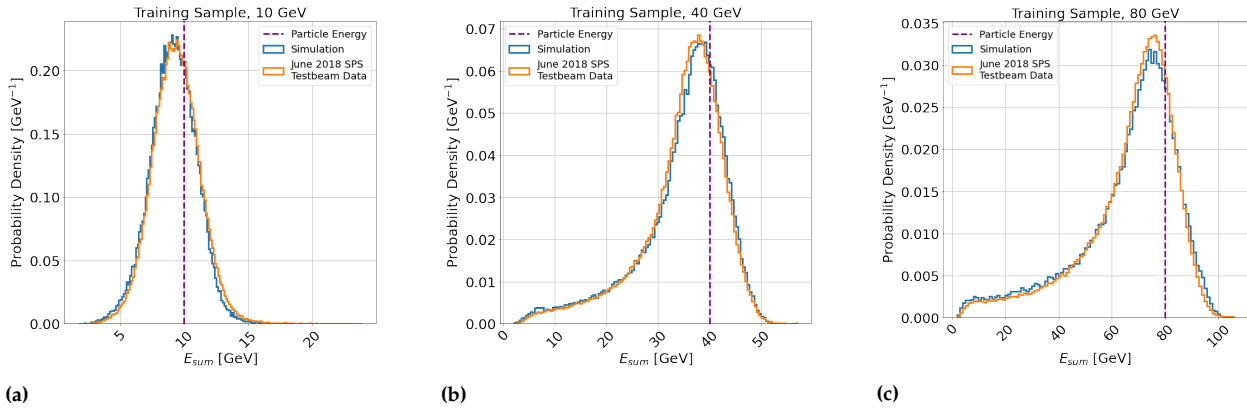


Figure 8.7: Distributions of the reconstructed shower energy ( $E_{sum}$ ) for the training dataset for shower separation, for 10, 40 and 80 GeV samples, shown in the left, middle and right columns, respectively. The blue histograms indicate simulation, and the orange histograms indicate June 2018 SPS Testbeam data.

Comments:

- As in Figure 5.7.

Event Level: Distributions of Number of Active Cells

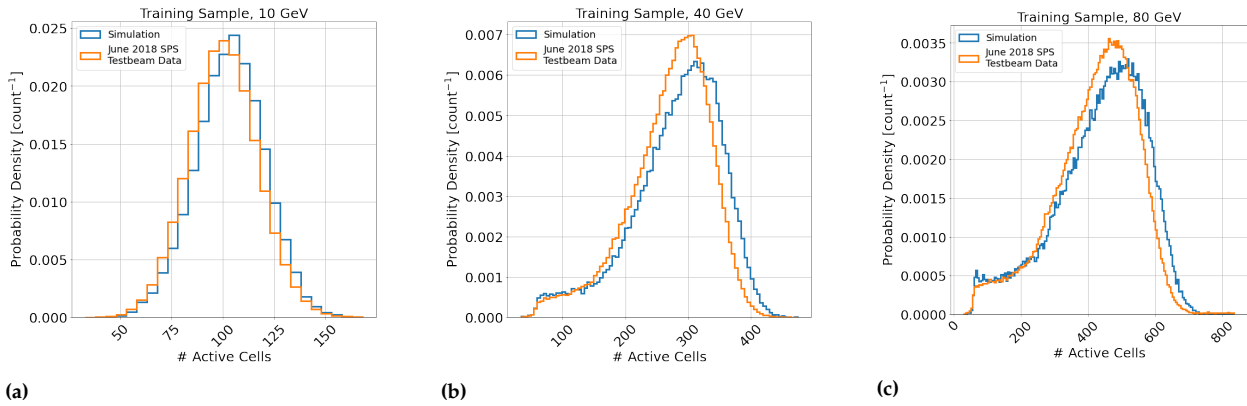
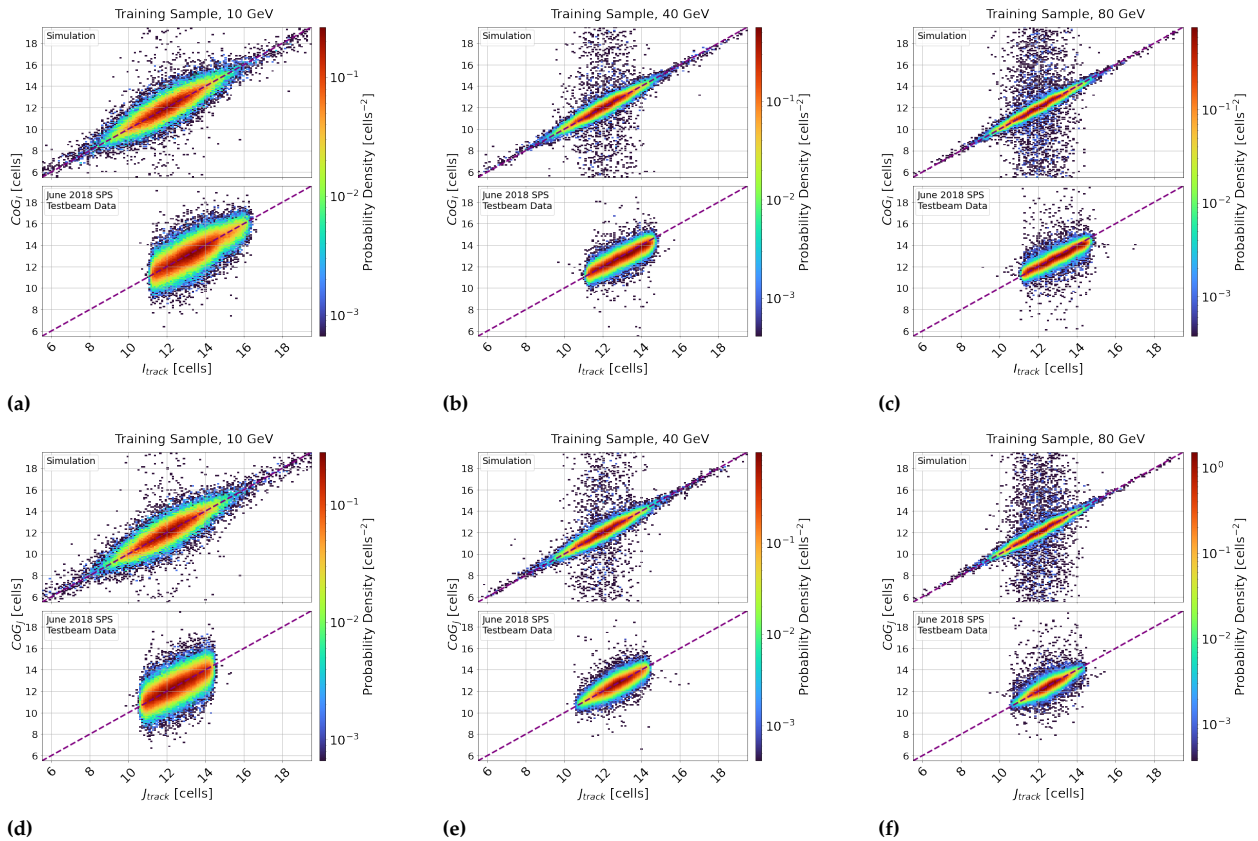


Figure 8.8: Distributions of the number of active cells per hadron showers for shower separation. As in Figure 8.7.

Comments:

- As in Figure 5.8.

Event Level: Joint Distributions of Lateral Centres-of-Gravity And Track Position

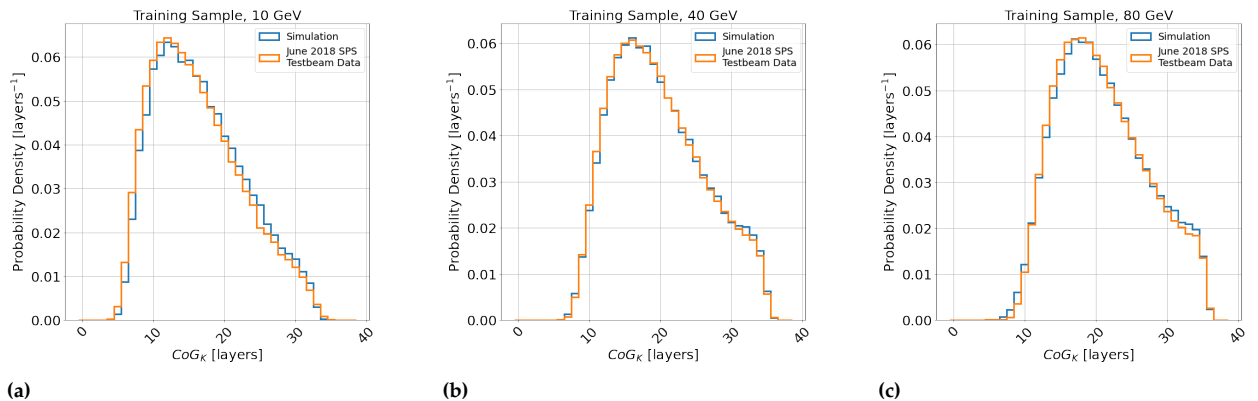


**Figure 8.9:** Joint distributions of the lateral centre-of-gravity and track position for the training dataset for shower separation, for 10, 40 and 80 GeV samples, shown in the left, middle and right columns, respectively. Figures 8.9a-8.9c show the distribution of  $CoG_I$ ,  $I_{track}$ , and Figures 8.9d-8.9f show the distribution of  $CoG_J$ ,  $J_{track}$ . The upper and lower subplots indicate the distribution of simulation and data, respectively. The colour axis indicates probability density. Else, as in Figure 8.7

Comments:

- As in Figure 5.9.

Event Level: Distributions of Longitudinal Centres-of-Gravity

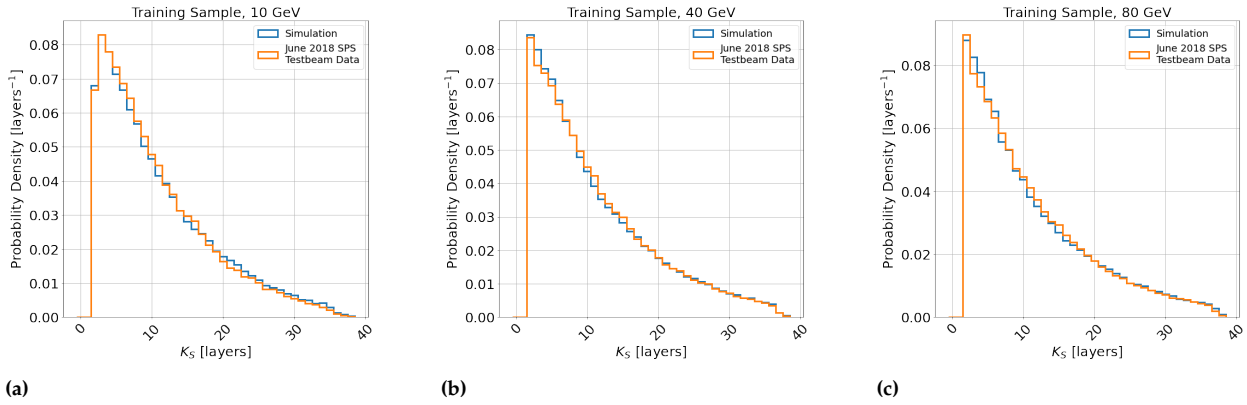


**Figure 8.10:** Distributions of the shower starting layer ( $K_S$ ) for the training dataset for shower separation. As in Figure 8.7.

Comments:

- As in Figure 5.10

### Event Level: Distributions of Shower Start Layer

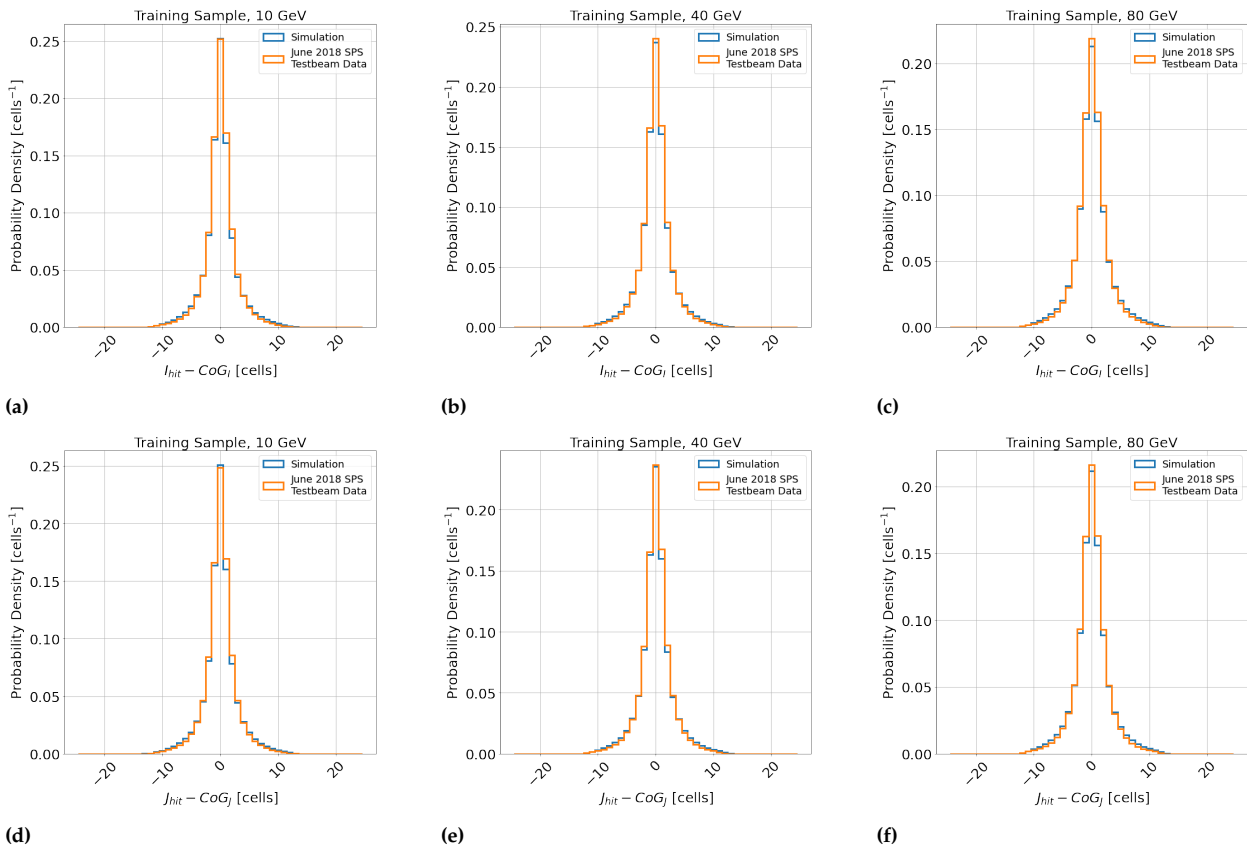


**Figure 8.11:** Distributions of the shower starting layer ( $K_S$ ) for the training dataset for shower separation. As in Figure 8.7.

**Comments:**

► As in Figure 5.11.

### Cell Level: Distributions of Lateral Active Cell Positions Relative to Centre-Of-Gravity



**Figure 8.12:** Distributions of the lateral cell indices for the training dataset for shower separation. Figures 8.12a-8.12c and Figures 8.12d-8.12f show the distributions of  $I_{hit}$  and  $J_{hit}$  minus their corresponding centres-of-gravity,  $CoG_I$  and  $CoG_J$ , respectively, presented in units of cells. Figures 8.13a-8.13c show the hit radius distributions as defined in 2.5.1.1. Else, as in Figure 8.7.

**Comments:**

► As in Figure 5.12.

Cell Level: Distributions of Hit Radius

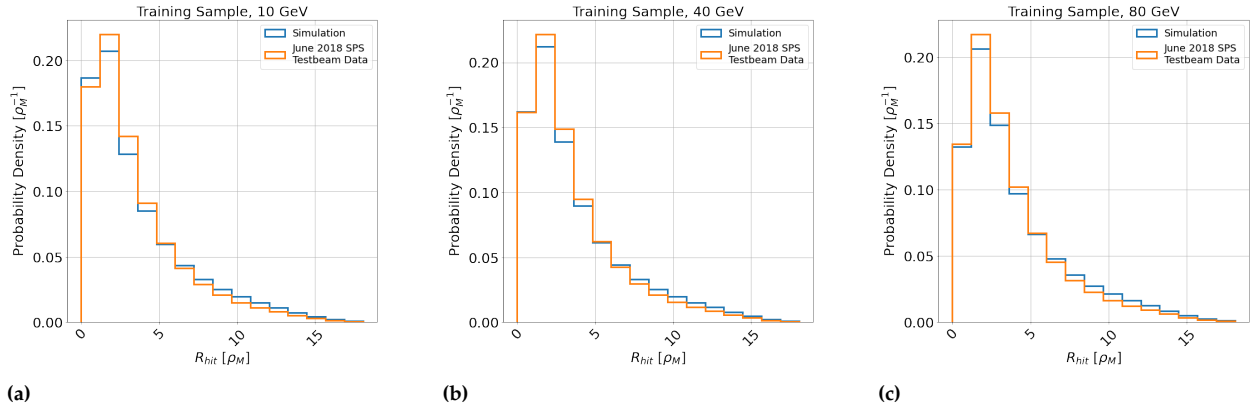


Figure 8.13: Distributions of hit radius ( $R_{hit}$ ) as defined in Section 2.5.1.1 for the training dataset for shower separation. Else, as in Figure 8.7.

Comments:

► As in Figure 5.13.

Cell Level: Distributions of Hit Azimuthal Angle

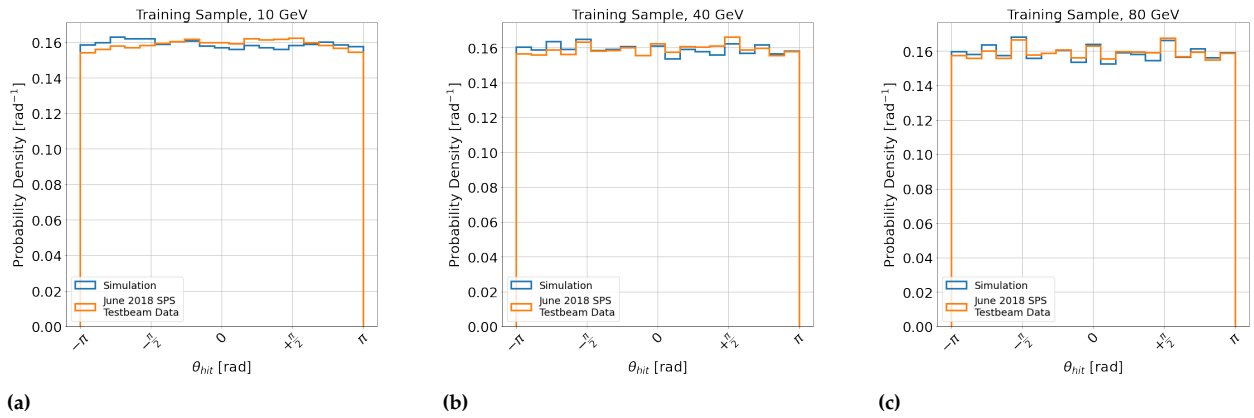
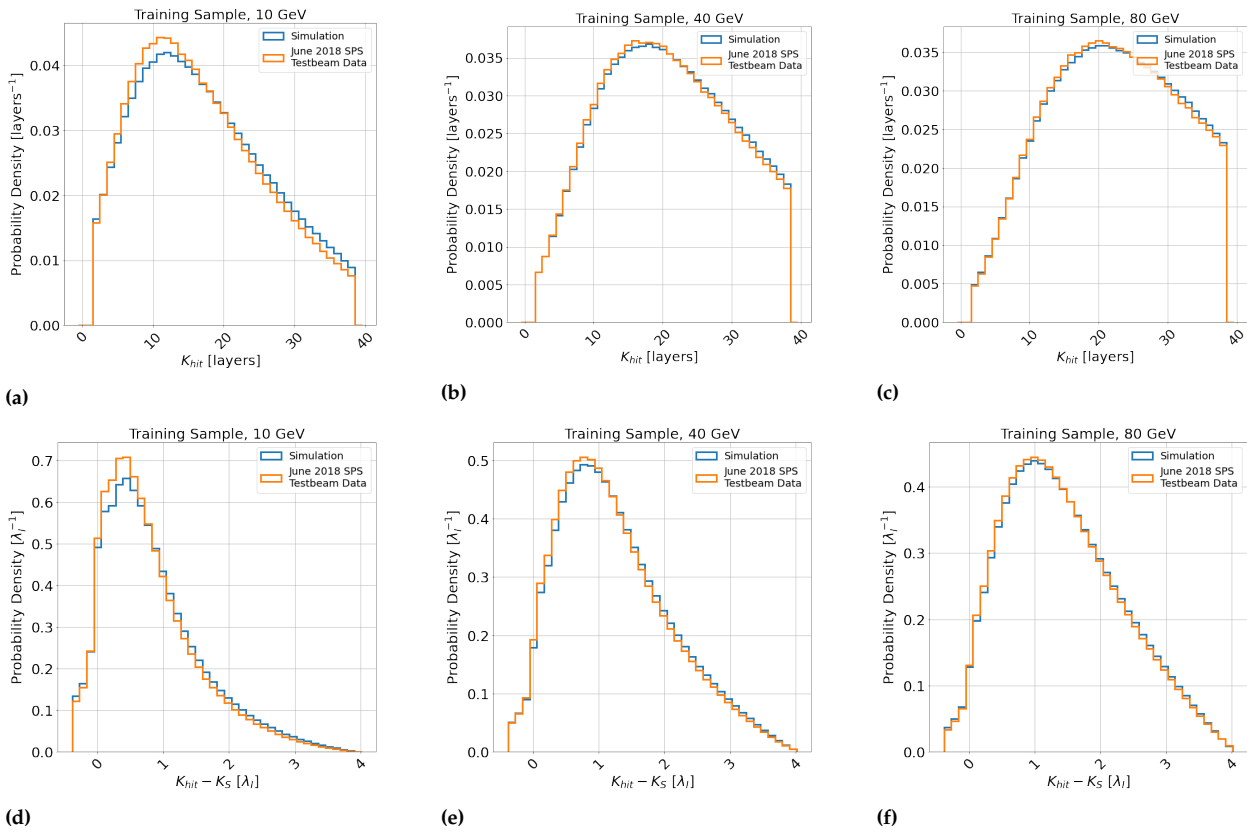


Figure 8.14: Distributions of hit azimuthal angle ( $\theta_{hit}$ ) as defined in Section 2.5.1.1 for the training dataset for shower separation. Else, as in Figure 8.7.

Comments:

► As in Figure 5.14.

**Cell Level: Distributions of Reconstructed and Shower-Start Normalised Longitudinal Layer Position**

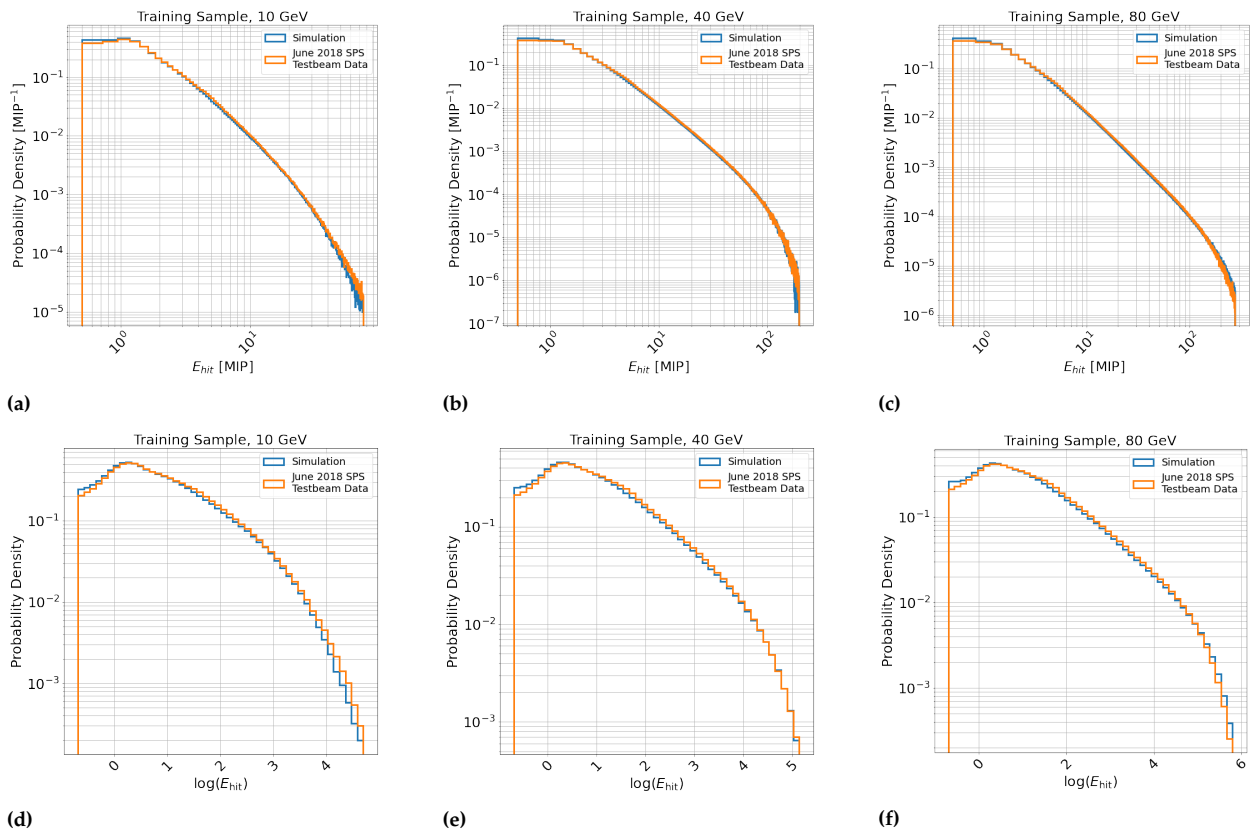


**Figure 8.15:** Distributions of longitudinal cell-level indices for the training dataset for shower separation, respectively. Figures 8.15a-8.15c and Figures 8.15d-8.15f show the distributions of  $K_{hit}$  and  $K_{hit} - K_S$ , respectively. Else, as in Figure 8.7.

**Comments:**

- As in Figure 5.15.

**Cell Level: Distributions of Active Cell Energy and its Logarithm**

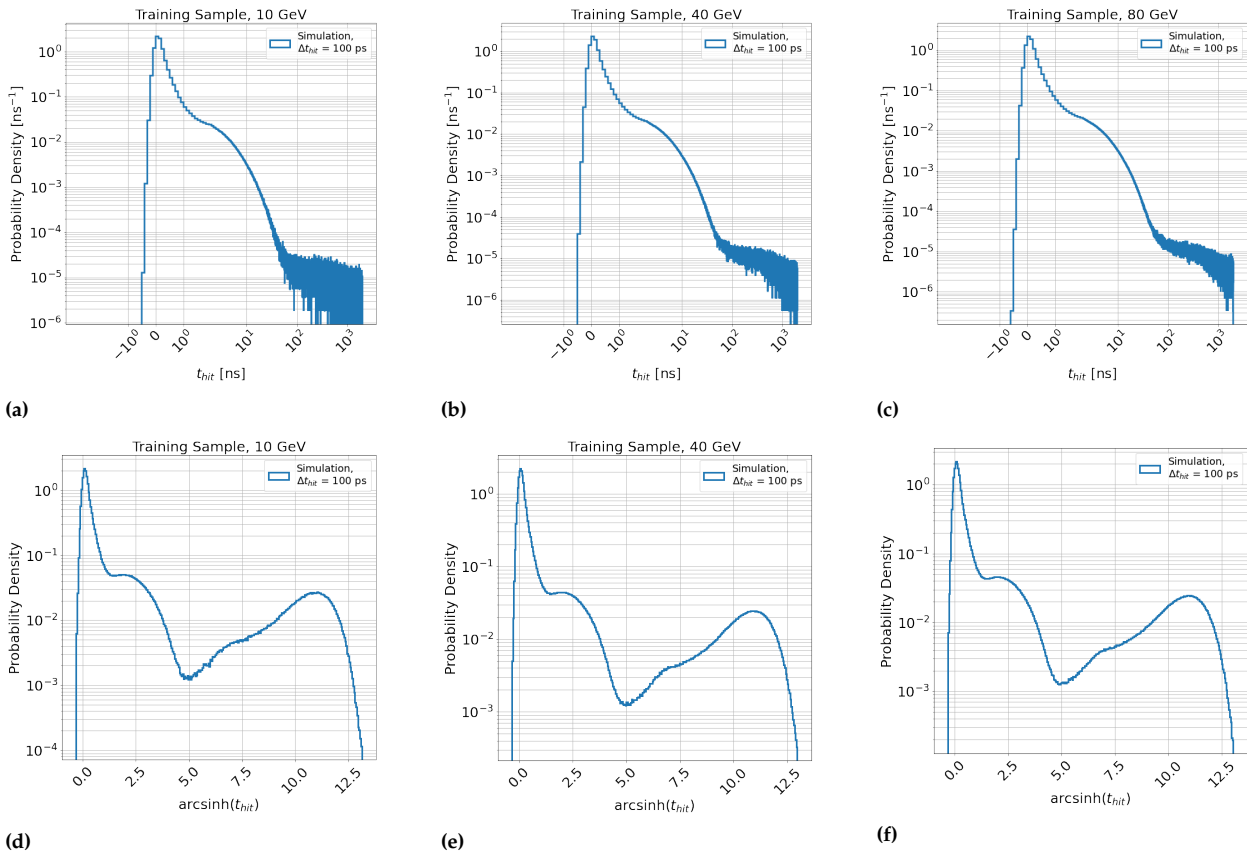


**Figure 8.16:** Distributions of active cell energy for the training dataset for shower separation, respectively. Figures 8.16a-8.16c and Figures 8.16d-8.16f show the distributions of  $E_{hit}$  and its logarithm, respectively. Else, as in Figure 8.7.

**Comments:**

- As in Figure 5.16.

**Cell Level: Distributions of Active Cell Timestamp and its Hyperbolic Sine in Simulation with a Resolution of 100 ps**



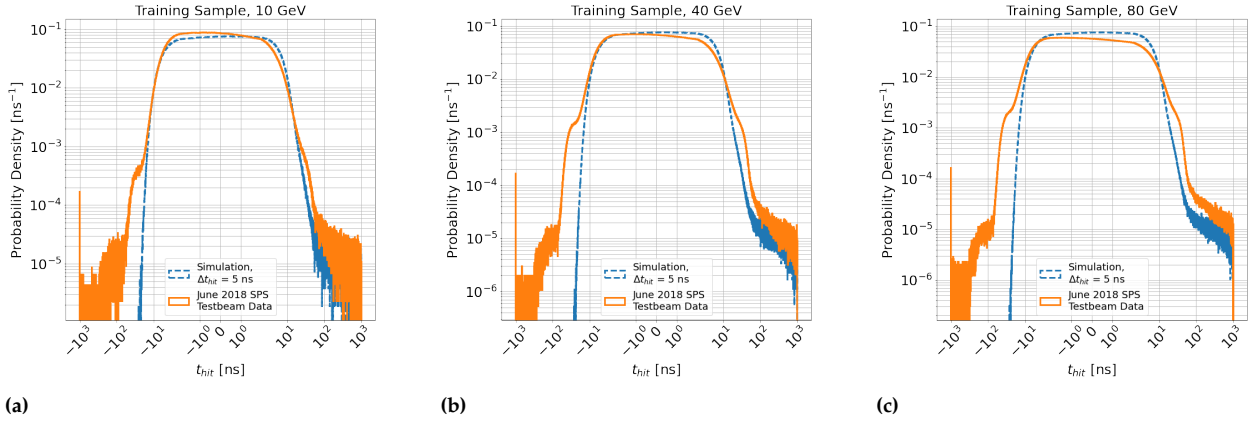
**Figure 8.17:** Distributions of the cell timestamp for the training dataset for shower separation in simulation, with 100 ps time resolution. Figures 8.17a-8.17c and Figures 8.17e-8.17f show the distributions of the active cell timestamp in simulation smeared by of 100 ns and its hyperbolic sine, respectively.

**Comments:**

- As in Figure 5.17.



**Cell Level: Distributions of Active Cell Timestamp with a Resolution of 5 ps**

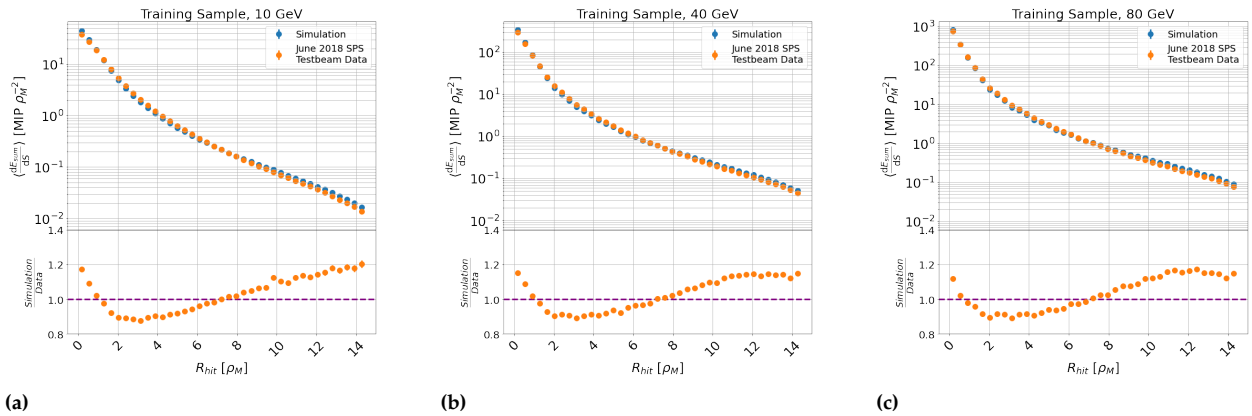


**Figure 8.18:** Distributions of the cell timestamp for the training dataset for shower separation in simulation, with 5 ns time resolution. Figures 8.18a-8.18c show the distributions of the active cell timestamp in simulation smeared by of 5 ns.

**Comments:**

- As in Figure 5.18.

**Average Radial Energy Profile**

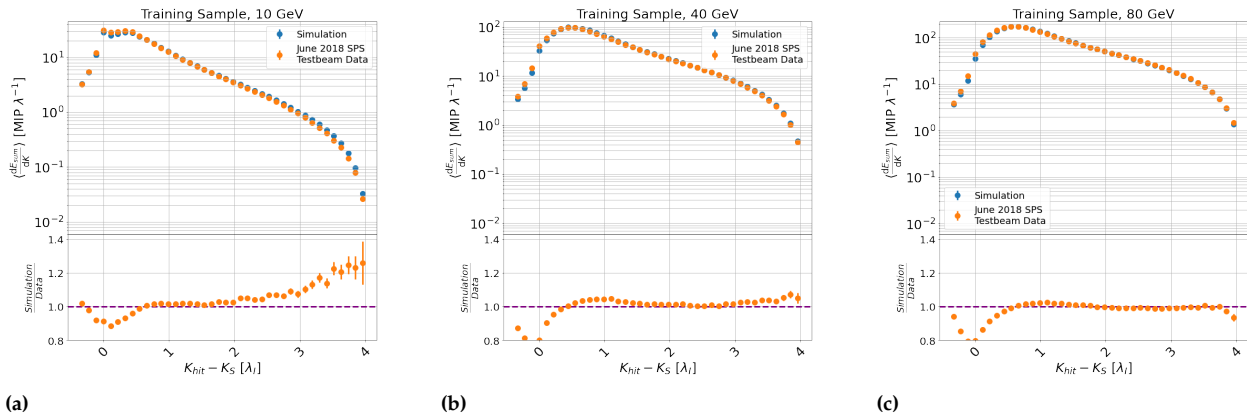


**Figure 8.19:** Figures 8.19a-8.19c show the average radial shower energy profile distributions per unit radial surface area of a circle with radius  $R$  (denoted  $S$ ) Blue and orange dots indicate simulation and data, respectively.

**Comments:**

- As in Figure 5.19.

### Average Longitudinal Energy Profile

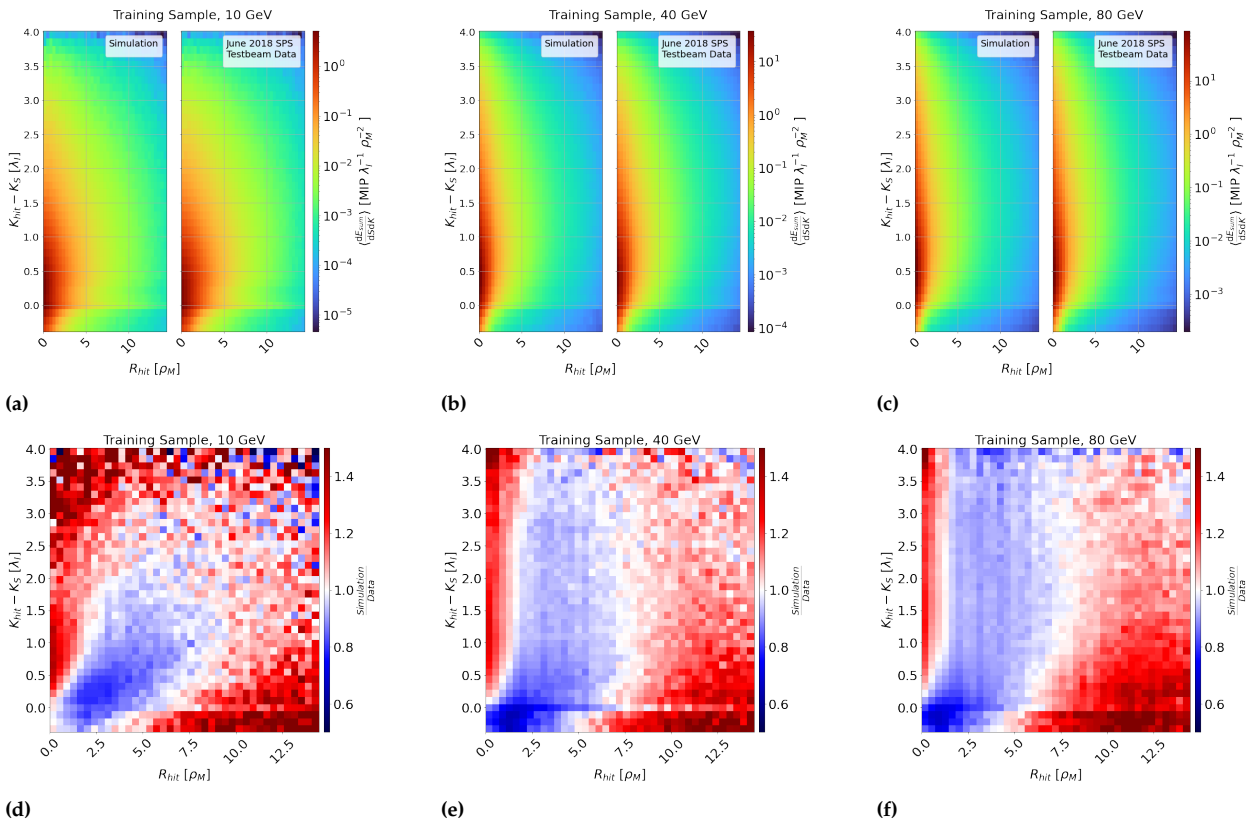


**Figure 8.20:** Figures 8.19a-8.19c show the average longitudinal shower energy profile distributions per layer of the calorimeter, relative to shower starting position ( $K_S$ ). Blue and orange dots indicate simulation and data, respectively.

**Comments:**

► As in Figure 5.20.

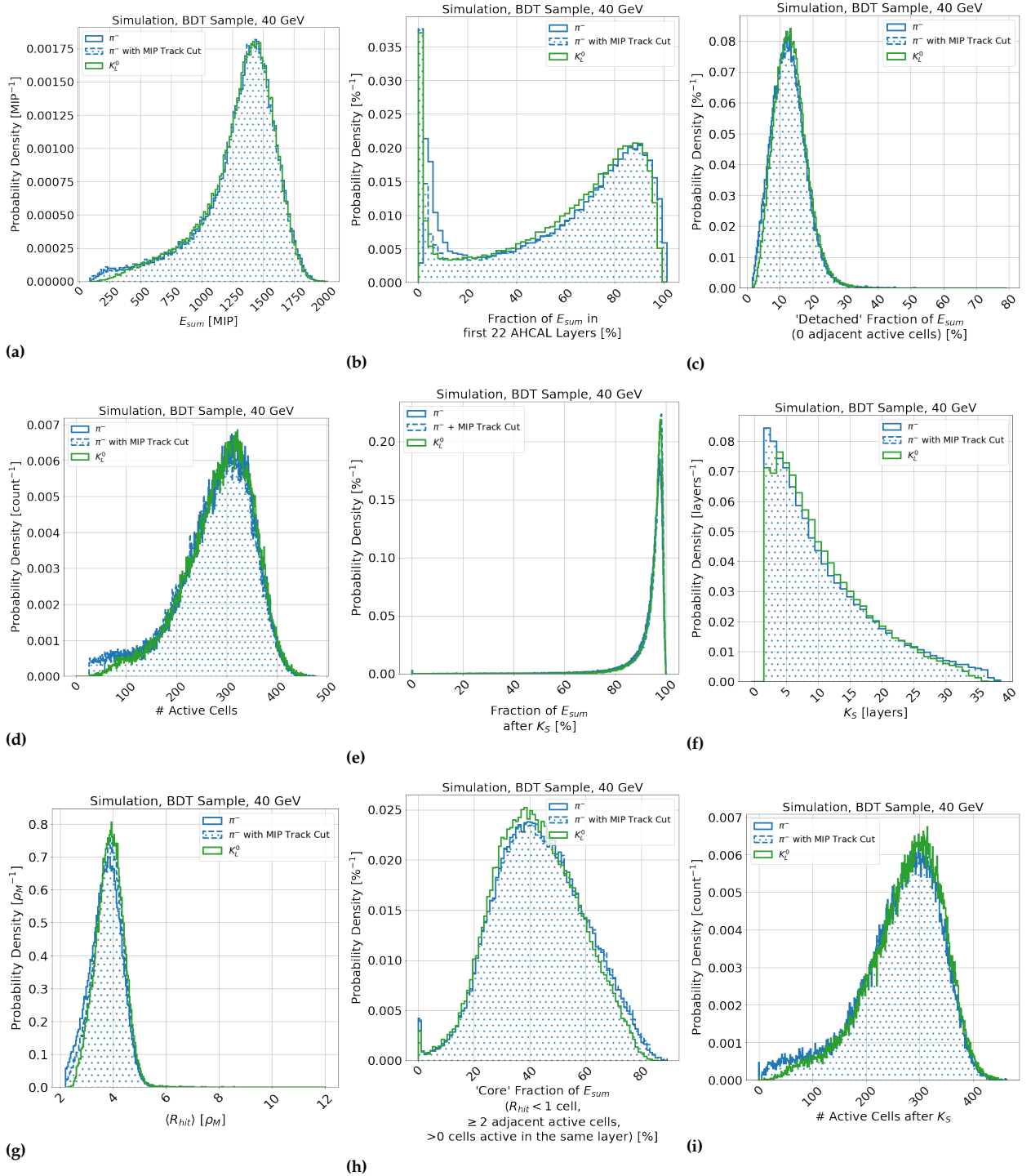
### Average Joint Radial-Longitudinal Energy Profile and Ratio of Simulation to Data



**Figure 8.21:** Figures 8.21a-8.21c show the joint average radial shower energy profile distributions per unit radial surface area of a circle with radius  $R$  (denoted  $S$ ), per layer of the calorimeter relative to the shower start ( $K_S$ ). The colour axis indicates probability density. Figures 8.21d-8.21f show the ratio of simulation and data shown in Figures 8.21a-8.21c, where the colour axis indicates the ratio.

**Comments:**

► As in Figure 5.21.

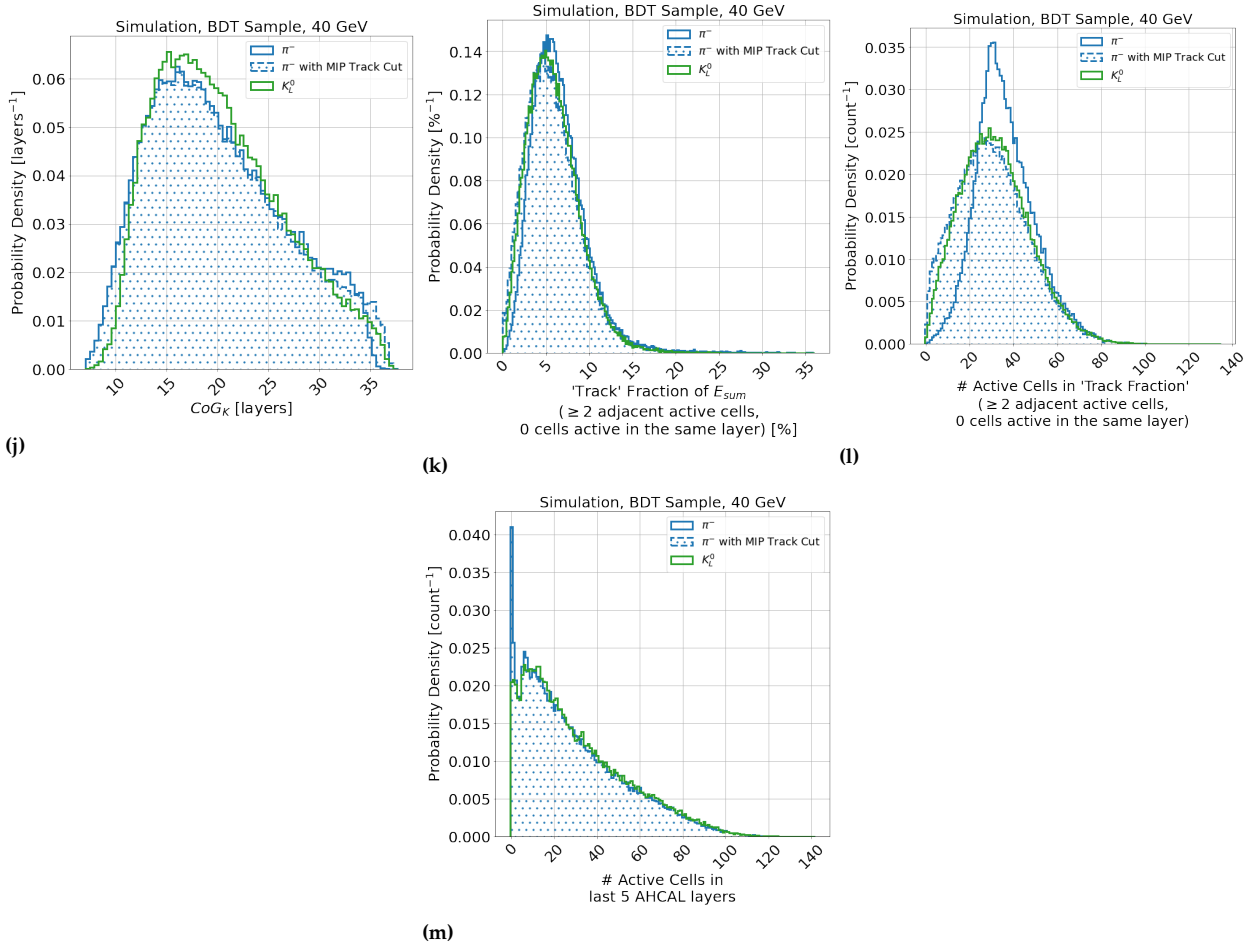
**PID Variables for Comparison of 40 GeV  $\pi^-$ /Synthetic Neutral and  $K_L^0$  Simulated Hadron Showers in AHCAL**


**Figure 8.22:** Particle ID variables calculated according to [79, 86], compared for the simulated  $\pi^-$  and  $K_L^0$ , from Table 6.2. The green line indicates  $K_L^0$  hadron showers, while the blue and blue dashed lines indicate  $\pi^-$  hadron showers, with and without the applied MIP-track cut.

**Comments:**

- As in Figure 6.12.

**PID Variables for Comparison of 40 GeV  $\pi^-$ /Synthetic Neutral and  $K_L^0$  Simulated Hadron Showers in AHCAL**



**Figure 8.22 continued:** Particle ID variables calculated according to [79, 86], compared for the simulated  $\pi^-$  and  $K_L^0$ , from Table 6.2. The green line indicates  $K_L^0$  hadron showers, while the blue and blue dashed lines indicate  $\pi^-$  hadron showers, with and without the applied MIP-track cut.

**Comments:**

- As in Figure 6.12.

		Simulation, Training Sample																				Total Events				
$E_{particle}^N$ [GeV]	$E_{particle}^Q$ [GeV]	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100		105	110	115	120
5		1184	1191	1239	1304	1217	1162	1238	1234	1208	1237	1251	1270	1247	1298	1312	1252	1230	1298	1172	1282	1199	1285	1265	1183	29758
10		1252	1283	1210	1212	1257	1280	1253	1209	1218	1280	1271	1288	1255	1279	1264	1243	1273	1240	1257	1314	1220	1195	1276	1283	30112
15		1192	1273	1220	1273	1281	1177	1267	1286	1270	1291	1146	1264	1269	1206	1190	1149	1320	1289	1241	1258	1250	1243	1277	1214	29846
20		1219	1274	1251	1158	1213	1232	1318	1243	1279	1241	1198	1308	1291	1265	1272	1298	1319	1214	1250	1309	1237	1242	1262	1258	30151
25		1230	1190	1219	1260	1275	1238	1167	1300	1212	1295	1226	1262	1242	1255	1233	1284	1269	1214	1312	1294	1296	1200	1250	1253	29976
30		1237	1265	1287	1298	1152	1288	1214	1295	1260	1232	1218	1297	1165	1193	1339	1230	1223	1232	1263	1247	1289	1248	1240	1208	29920
35		1207	1234	1230	1318	1207	1272	1279	1235	1309	1225	1156	1191	1295	1230	1218	1224	1324	1266	1229	1208	1296	1271	1258	1261	29953
40		1255	1250	1220	1270	1278	1245	1230	1271	1213	1255	1217	1243	1277	1252	1233	1222	1254	1253	1232	1257	1329	1271	1249	1235	30011
45		1278	1332	1262	1254	1309	1217	1306	1235	1268	1271	1358	1331	1249	1265	1246	1210	1242	1266	1221	1248	1254	1262	1220	30408	
50		1206	1273	1311	1194	1265	1289	1257	1220	1204	1321	1281	1192	1224	1233	1191	1227	1273	1241	1309	1305	1201	1247	1237	1282	29983
55		1237	1306	1302	1286	1221	1255	1247	1255	1294	1317	1233	1257	1171	1252	1248	1246	1273	1240	1177	1265	1334	1250	1264	1252	30082
60		1228	1216	1249	1246	1187	1251	1291	1308	1163	1243	1253	1288	1237	1236	1223	1219	1241	1257	1260	1279	1204	1240	1214	1234	29767
65		1240	1225	1224	1243	1205	1268	1264	1217	1244	1238	1322	1294	1248	1166	1238	1273	1311	1214	1225	1184	1188	1241	1224	1239	29735
70		1280	1240	1185	1289	1239	1237	1230	1289	1249	1325	1263	1203	1263	1179	1243	1213	1245	1243	1253	1223	1204	1277	1256	1231	29859
75		1212	1229	1195	1226	1297	1222	1256	1293	1246	1221	1241	1193	1220	1243	1216	1277	1309	1244	1288	1297	1256	1273	1215	1185	29854
80		1318	1287	1240	1297	1314	1197	1305	1226	1248	1293	1238	1228	1271	1283	1253	1240	1283	1229	1241	1329	1278	1261	1333	1244	30336
85		1169	1244	1256	1312	1277	1266	1242	1264	1261	1327	1289	1298	1284	1248	1221	1239	1268	1228	1265	1265	1283	1249	1305	1240	30300
90		1280	1242	1289	1309	1224	1249	1266	1260	1250	1209	1296	1183	1284	1298	1226	1256	1253	1253	1193	1237	1215	1346	1217	1261	30096
95		1265	1252	1230	1241	1283	1335	1231	1273	1269	1248	1247	1319	1213	1214	1260	1261	1231	1209	1308	1211	1222	1237	1258	1236	30053
100		1216	1200	1240	1210	1221	1212	1294	1261	1233	1279	1229	1230	1242	1222	1274	1234	1226	1254	1329	1219	1286	1205	1330	1264	29910
105		1283	1253	1280	1285	1270	1250	1242	1283	1259	1245	1297	1185	1253	1244	1237	1250	1229	1233	1311	1218	1276	1305	1220	1274	30151
110		1289	1301	1218	1264	1200	1152	1258	1277	1286	1259	1225	1311	1216	1284	1246	1192	1323	1285	1240	1232	1262	1246	1229	1270	30064
115		1204	1259	1232	1252	1236	1265	1229	1207	1300	1257	1207	1193	1260	1252	1223	1197	1247	1267	1244	1274	1222	1213	1221	1192	29653
120		1277	1281	1230	1223	1264	1279	1232	1270	1233	1198	1291	1226	1179	1246	1195	1298	1221	1278	1247	1246	1340	1229	1274	1265	30022
Total Events		29758	30100	29819	30224	29892	29838	30116	30211	29976	30307	29953	30054	29855	29843	29801	29734	30387	29947	30077	30201	30040	30058	30025	29784	720000

(a)

		Simulation, Validation Sample																				Total Events				
$E_{particle}^N$ [GeV]	$E_{particle}^Q$ [GeV]	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100		105	110	115	120
5		147	134	144	125	146	138	158	130	145	134	144	125	154	136	148	115	122	141	134	138	126	153	160	128	3325
10		143	151	144	137	143	151	178	156	128	143	144	148	141	122	142	150	161	151	124	147	162	121	144	125	3456
15		140	138	115	130	137	142	139	125	151	145	133	134	148	150	145	149	108	153	152	165	133	124	129	143	3328
20		154	154	149	146	158	139	115	159	129	135	131	134	132	126	153	133	141	130	157	139	135	147	153	119	3368
25		155	124	119	140	128	138	126	146	132	120	147	130	148	139	127	143	126	156	123	148	141	130	126	123	3235
30		156	133	136	146	129	122	142	134	145	156	134	142	136	148	139	110	147	136	133	159	148	142	167	139	3379
35		132	136	122	158	142	115	154	120	133	137	140	125	164	130	138	142	115	127	148	133	130	128	143	141	3253
40		130	137	127	148	130	151	148	142	152	151	138	147	164	138	131	134	156	115	132	122	153	123	156	152	3377
45		139	140	122	133	136	142	131	131	134	149	135	154	132	128	132	137	140	137	109	133	126	163	132	120	3235
50		138	130	127	156	127	155	133	133	134	135	141	139	131	151	152	143	134	165	136	168	135	144	144	116	3367
55		131	142	136	139	156	147	149	128	136	143	128	142	130	145	139	146	158	130	129	154	135	133	144	144	3364
60		151	155	144	117	129	127	146	153	137	138	159	144	134	149	138	131	153	133	144	137	149	125	129	138	3360
65		143	136	133	122	138	131	141	139	130	144	129	159	143	118	155	134	139	148	153	143	132	128	170	157	3365
70		148	152	137	141	144	116	127	154	146	135	131	138	119	126	129	135	146	149	140	139	130	118	131	131	3262
75		163	124	148	140	157	121	129	124	142	123	163	149	128	125	136	119	146	132	136	131	135	147	116	133	3267
80		142	132	156	147	128	139	126	136	133	122	124	139	121	142	123	136	149	139	123	112	132	137	140	137	3215
85		130	120	146	134	137	147	161	137	147	127	148	135	140	137	133	147	146	139	120	115	141	146	161	154	3348
90		125	131	126	146	117	135	148	141	145	111	133	159	142	143	149	153	132	150	158	133	131	129	131	131	3299
95		139	144	171	136	148	136	154	126	125	159	148	132	142	138	137	166	149	114	132	143	134	139	125	144	3381
100		139	126	170	151	150	140	142	152	150	140	137	141	141	123	125	144	148	133	148	130	134	132	118	149	3363
105		129	151	152	153	129	145	146	128	161	130	138	145	124	154	118	141	153	124	117	144	127	139	130	124	3302
110		142	127	119	131	152	137	130	126	142	169	142	148	153	145	117	137	131	143	143	132	132	119	144	126	3287
115		148	146	156	132	164	165	153	140	149	160	127	117	136	156	152	123	135	165	159	147	134	144	140	139	3487
120		131	163	136	135	124	129	169	127	126	134	122	148	141	134	158	134	127	151	146	141	131	158	148	164	3377
Total Events		3395	3326	3335	3343	3349	3308	3445	3287	3352	3340	3316	3374	3344	3303	3316	3302	3362	3361	3296	3353	3266	3269	3381	3277	80000

(b)

		Simulation, Testing Sample																				Total Events
$E_{particle}^N$ [GeV]	$E_{particle}^Q$																					

June 2018 SPS Data, Training Sample							
$E_{\text{particle}}^N$ [GeV]	10	20	40	60	80	120	Total Events
$E_{\text{particle}}^Q$ [GeV]							
10	20074	20081	19744	19953	20035	19941	119828
20	19898	19821	20229	19745	19765	20026	119484
40	19776	19984	20024	19990	20040	20008	119822
60	20038	20099	19741	20117	19784	20270	120049
80	20120	20008	20170	20083	20070	20110	120561
120	19940	19983	20045	20198	20085	20005	120256
Total Events	119846	119976	119953	120086	119779	120360	720000

(a)

June 2018 SPS Data, Validation Sample							
$E_{\text{particle}}^N$ [GeV]	10	20	40	60	80	120	Total Events
$E_{\text{particle}}^Q$ [GeV]							
10	2257	2274	2211	2179	2162	2285	13368
20	2209	2177	2156	2247	2205	2300	13294
40	2290	2279	2163	2221	2246	2256	13455
60	2260	2202	2187	2258	2179	2226	13312
80	2172	2203	2193	2254	2199	2236	13257
120	2183	2264	2213	2244	2223	2187	13314
Total Events	13371	13399	13123	13403	13214	13490	80000

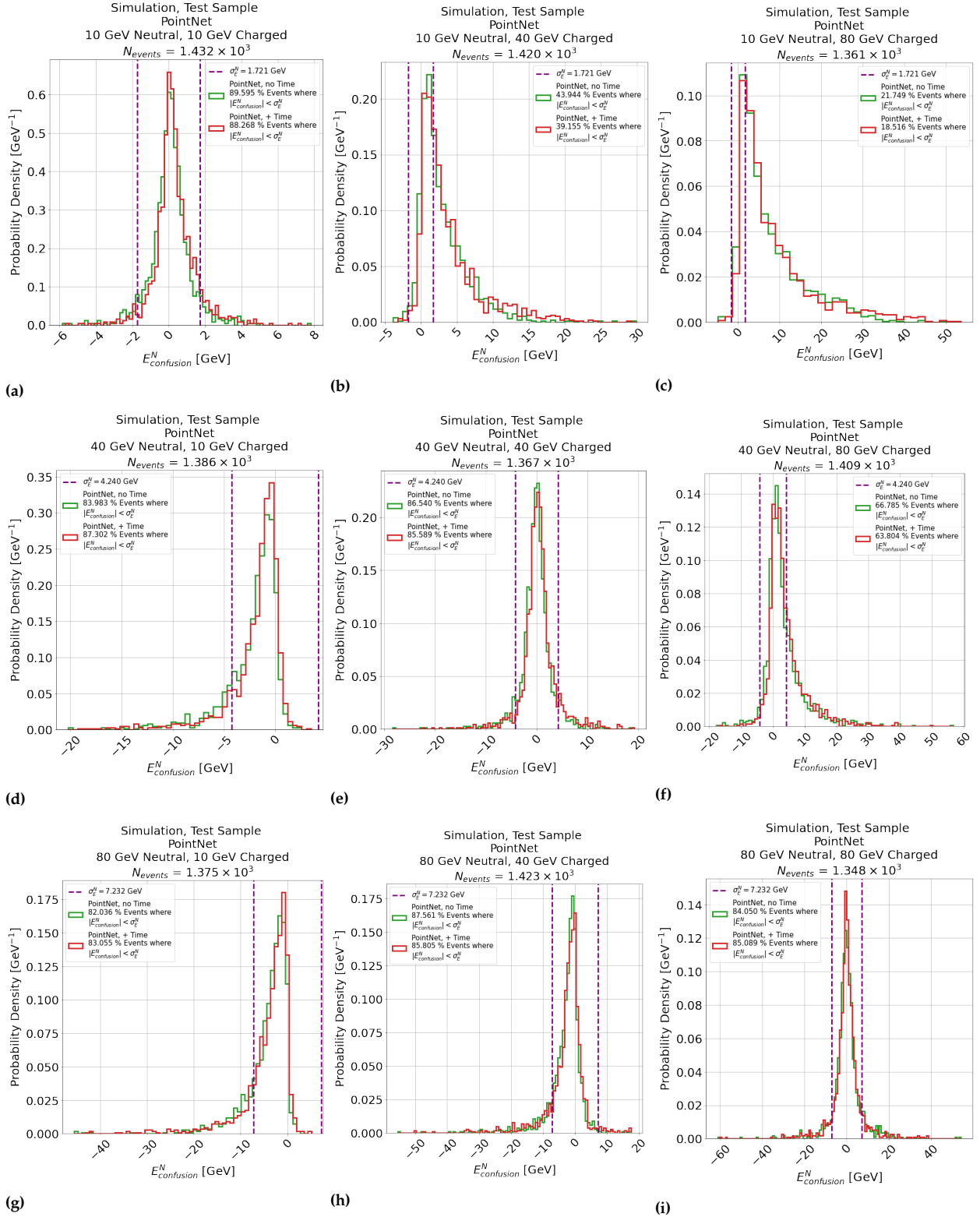
(b)

June 2018 SPS Data, Testing Sample							
$E_{\text{particle}}^N$ [GeV]	10	20	40	60	80	120	Total Events
$E_{\text{particle}}^Q$ [GeV]							
10	5683	5575	5607	5505	5467	5460	33297
20	5719	5466	5506	5536	5652	5540	33419
40	5532	5568	5608	5500	5619	5556	33383
60	5658	5662	5436	5469	5544	5546	33315
80	5548	5647	5650	5630	5534	5463	33472
120	5430	5561	5486	5410	5551	5676	33114
Total Events	33570	33479	33293	33050	33367	33241	200000

(c)

**Table 8.7:** Table 8.7a, 8.7b and 8.7c show the event tables for the training, validation and testing samples of data, combined from the source events shown in Table 6.1. Each table is split by each possible combination of charged and neutral hadron shower particle energy, indicated by the  $E_{\text{particle}}^Q$  and  $E_{\text{particle}}^N$  combinations, respectively.

### Distributions of Confusion Energy for Possible Combinations of Charged and Neutral Hadron Particle Energies at 10 GeV, 40 GeV and 80 GeV using PointNet

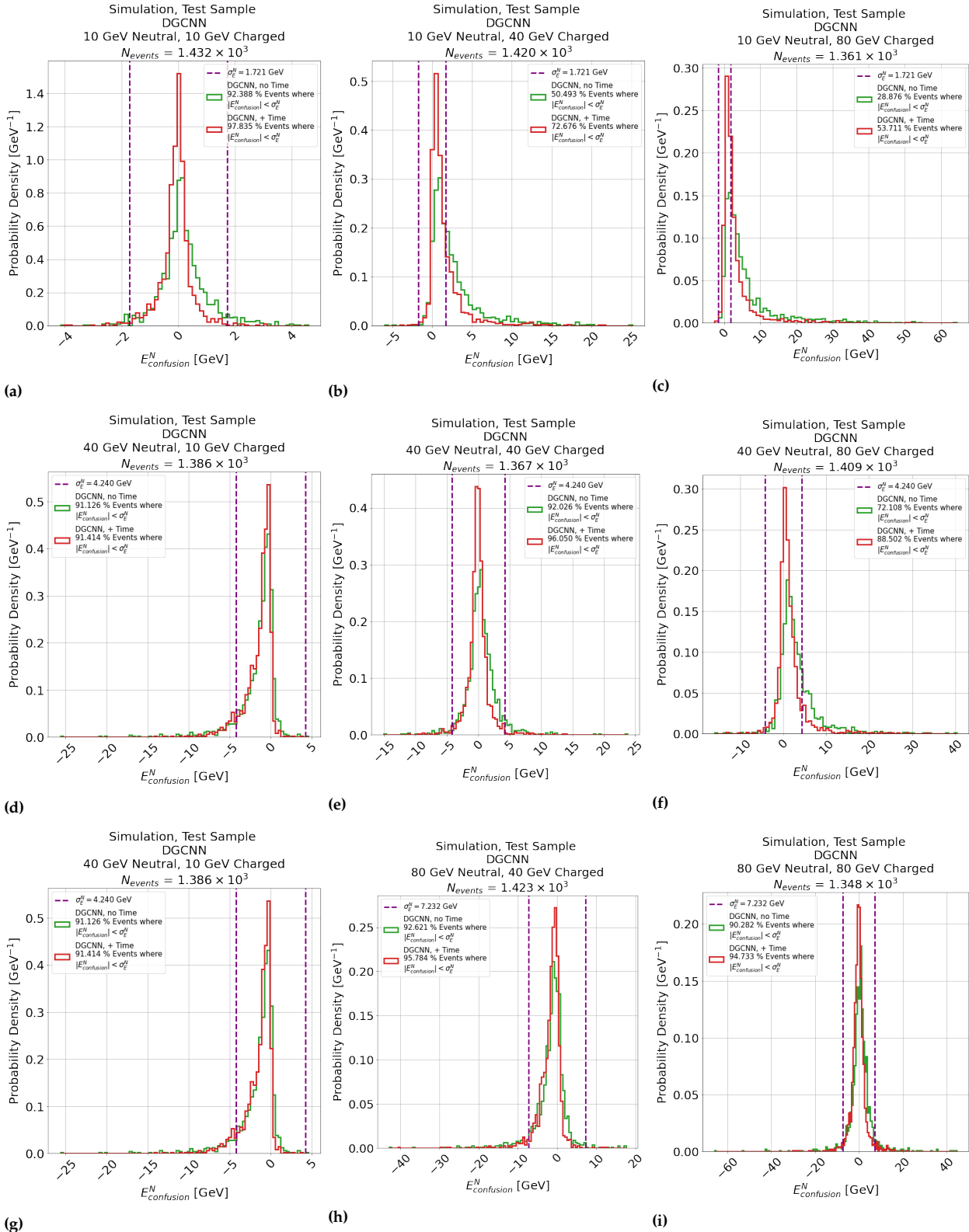


**Figure 8.23:** Distributions of confusion energy for the PointNet network applied to the test sample. The green and red lines indicate the models trained without and with timing information, respectively. The purple dashed lines indicate the resolution of the AHCAL calorimeter in simulation, as determined in Chapter 5.

#### Comments:

- As in Figure 6.35, except that no overall improvement is observed using timing information.

### Distributions of Confusion Energy for Possible Combinations of Charged and Neutral Hadron Particle Energies at 10 GeV, 40 GeV and 80 GeV using DGCNN



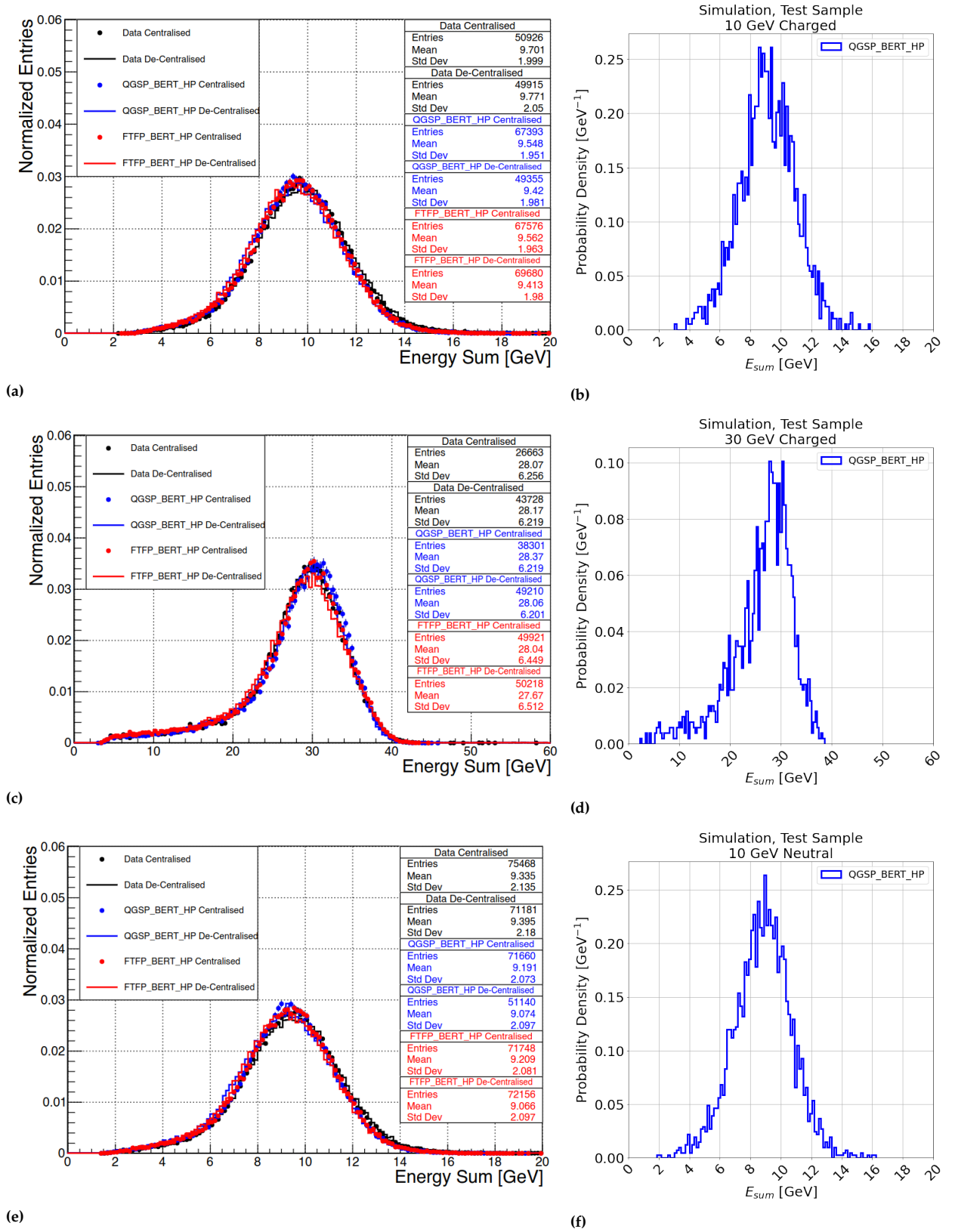
**Figure 8.24:** Distributions of confusion energy for the DGCNN network applied to the test sample. The green and red lines indicate the models trained without and with timing information, respectively. The purple dashed lines indicate the resolution of the AHCAL calorimeter in simulation, as determined in Chapter 5.

**Comments:**

► As in Figure 6.35.



### Comparison Of Reconstructed Energy Distribution Used In Shower Separation in The Pandora PFA Two-Particle Shower Separation Study of [78] and In Chapter 6

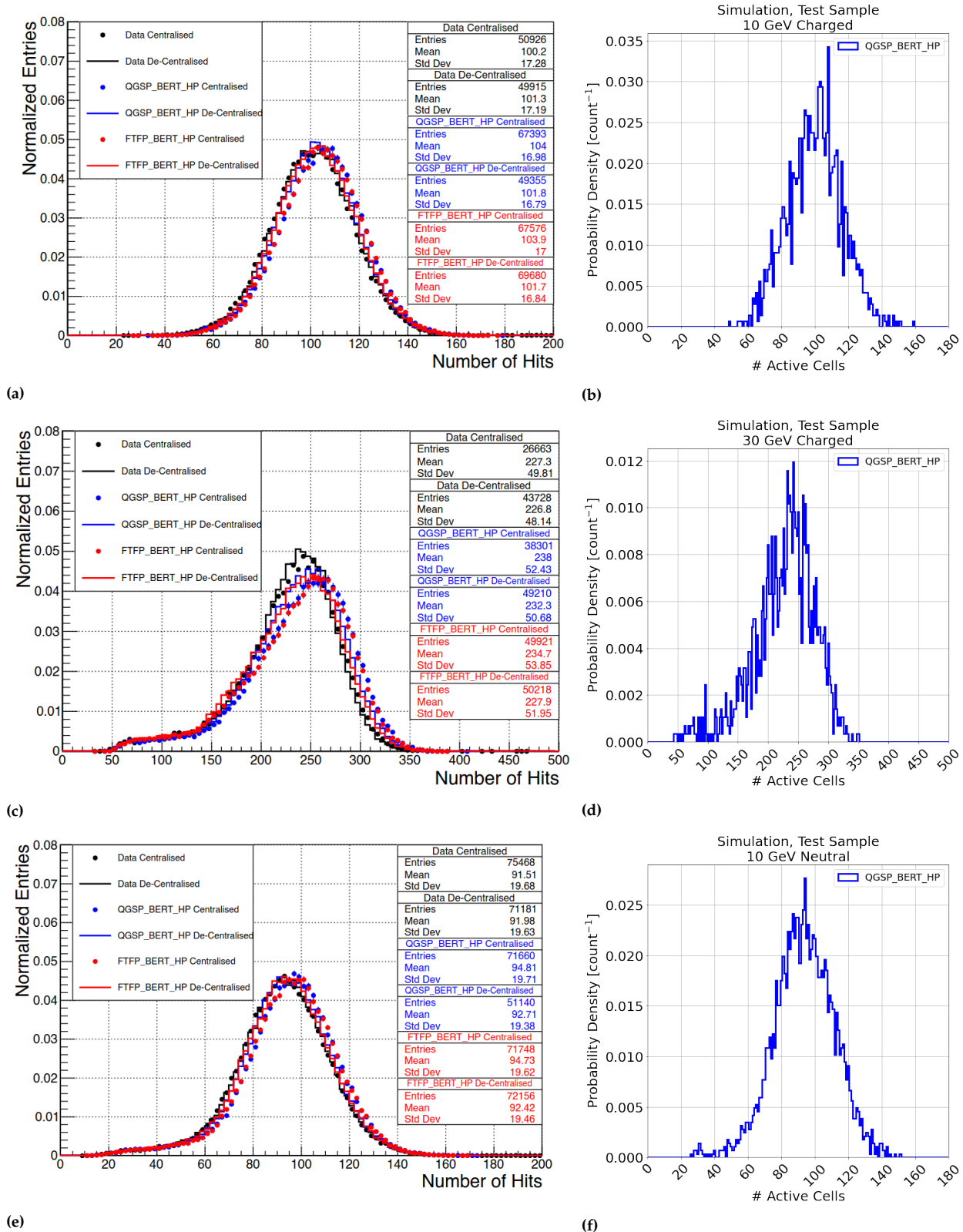


**Figure 8.25:** Distributions of reconstructed energy comparing the simulated testing dataset of 8.6 to the distributions of the study of [78]. The left and right columns shows distributions from [76] and from Table 8.6. Each row shows the distributions for  $E_{\text{particle}}^Q = 10 \text{ GeV}$ ,  $E_{\text{particle}}^Q = 30 \text{ GeV}$  and  $E_{\text{particle}}^N = 10 \text{ GeV}$ . Figures 8.25b, 8.25d and 8.26f are selected from the samples where the combination of charged (Q) and neutral (N) shower is 10 GeV Q + 10 GeV N, 30 GeV Q + 10 GeV N, and 10 GeV Q + (10 GeV N or 30 GeV N) respectively. In both figures, blue indicates simulation with the QGSP\_BERT\_HP physics list.

#### Comments:

- Good agreement is observed between the reconstructed energy distribution of the two studies.

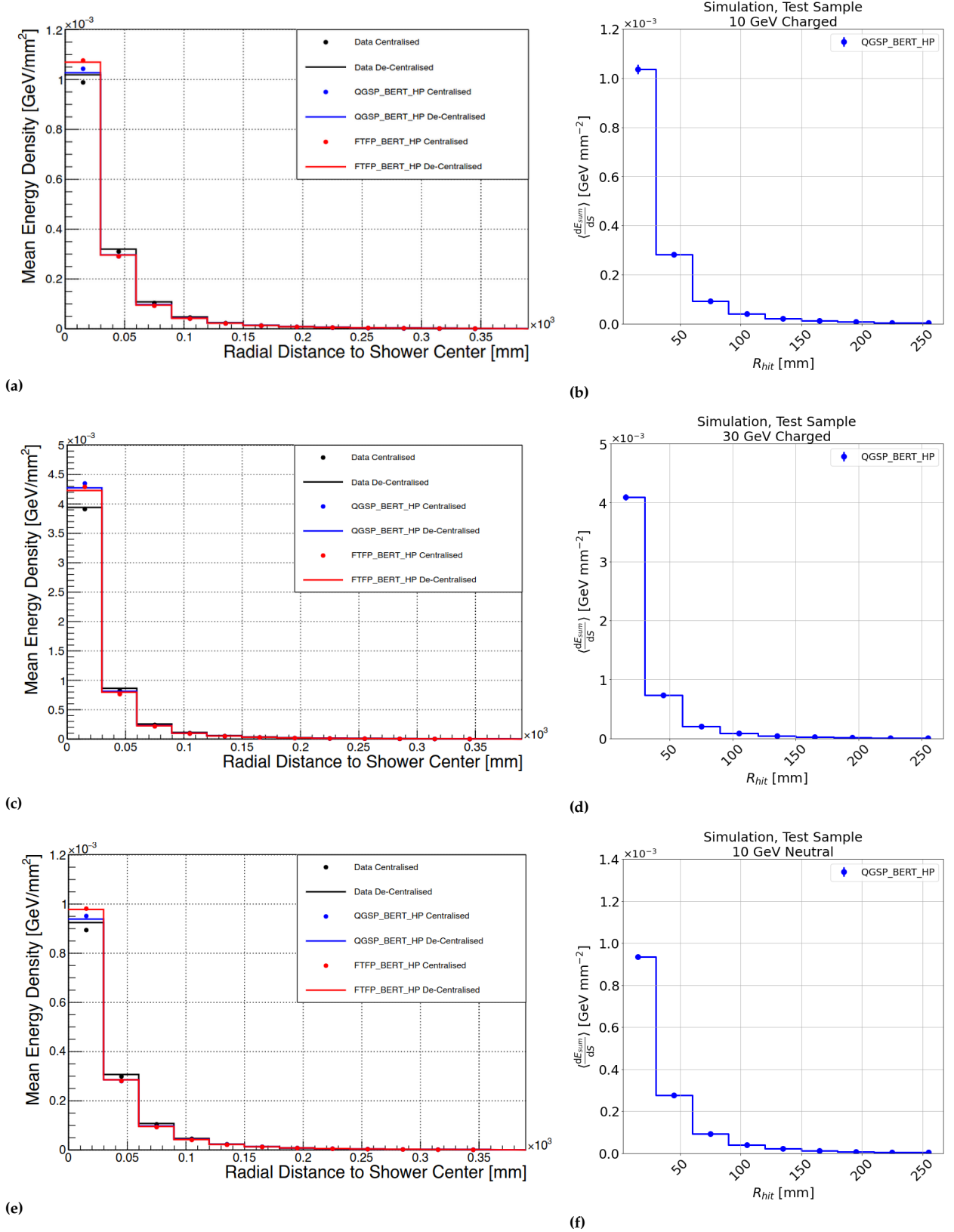
### Comparison Of Number of Active Cells Distributions Used In Shower Separation in The Pandora PFA Two-Particle Shower Separation Study of [78] and In Chapter 6



**Figure 8.26:** Distributions of number of active cells comparing the simulated testing dataset of 8.6 to the distributions of the study of [78]. The left and right columns shows distributions from [76] and from Table 8.6. Each row shows the distributions for  $E_{particle}^Q = 10\text{ GeV}$ ,  $E_{particle}^Q = 30\text{ GeV}$  and  $E_{particle}^N = 10\text{ GeV}$ . Figures 8.25b, 8.25d and 8.26f are selected from the samples where the combination of charged (Q) and neutral (N) shower is 10 GeV Q + 10 GeV N, 30 GeV Q + 10 GeV N, and 10 GeV Q + (10 GeV N or 30 GeV N) respectively. In both figures, blue indicates simulation with the QGSP\_BERT\_HP physics list.

**Comments:**

- Good agreement is observed between the number of active cells distribution of the two studies.

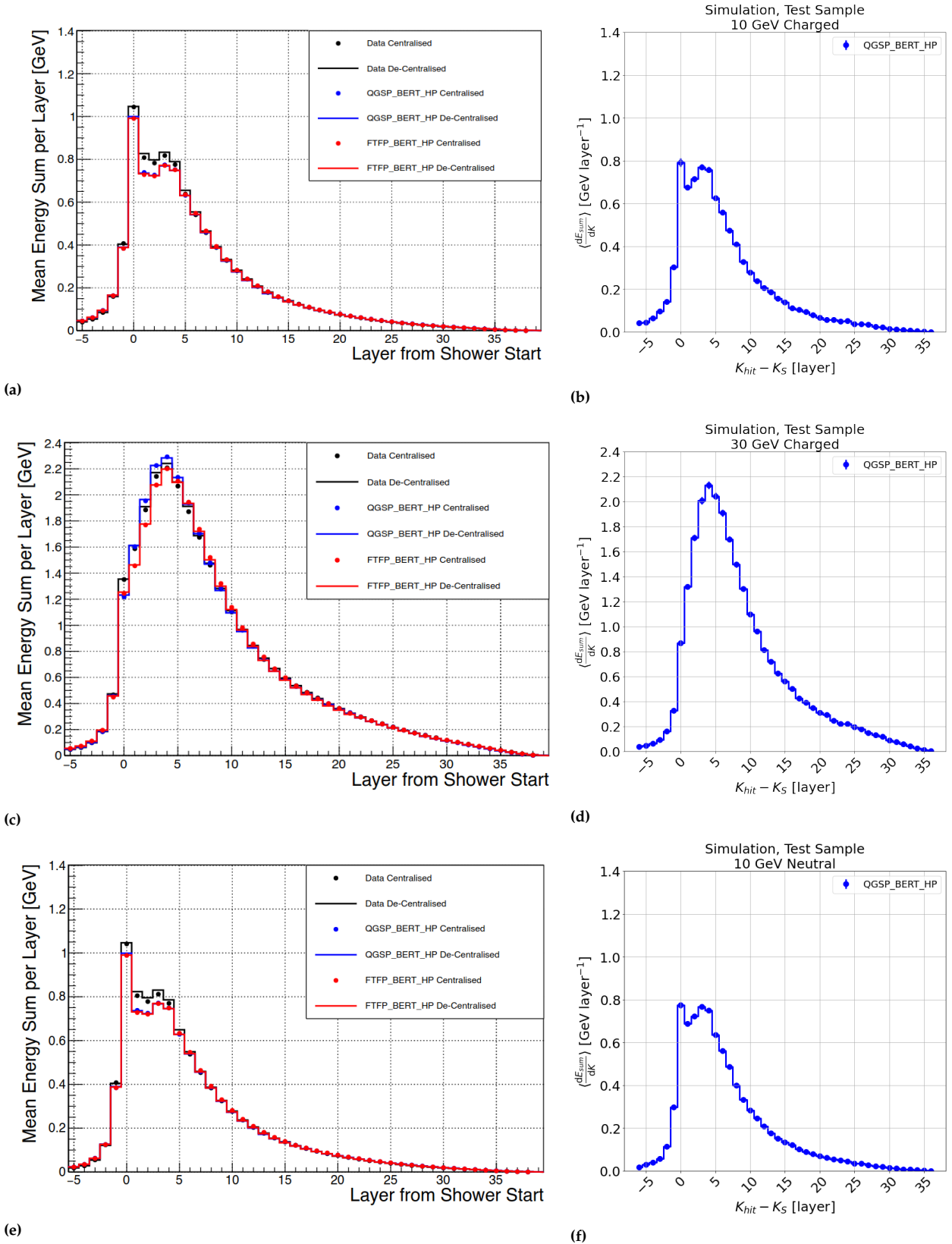
**Comparison Of Radial Energy Profiles Used In Shower Separation in The Pandora PFA Two-Particle Shower Separation Study of [78] and In Chapter 6**


**Figure 8.27:** Radial Energy Profiles comparing the simulated testing dataset of 8.6 to the distributions of the study of [78]. The left and right columns shows distributions from [76] and from Table 8.6. Each row shows the distributions for  $E_{\text{particle}}^Q = 10 \text{ GeV}$ ,  $E_{\text{particle}}^Q = 30 \text{ GeV}$  and  $E_{\text{particle}}^N = 10 \text{ GeV}$ . Figures 8.25b, 8.25d and 8.26f are selected from the samples where the combination of charged (Q) and neutral (N) shower is 10 GeV Q + 10 GeV N, 30 GeV Q + 10 GeV N, and 10 GeV Q + (10 GeV N or 30 GeV N) respectively. In both figures, blue indicates simulation with the QGSP\_BERT\_HP physics list.

**Comments:**

- Good agreement is observed between the radial energy profiles distribution of the two studies.

### Comparison Of Longitudinal Energy Profiles Used In Shower Separation in The Pandora PFA Two-Particle Shower Separation Study of [78] and In Chapter 6



**Figure 8.28:** Longitudinal Energy Profiles comparing the simulated testing dataset of 8.6 to the distributions of the study of [78]. The left and right columns shows distributions from [76] and from Table 8.6. Each row shows the distributions for  $E_{particle}^Q = 10 \text{ GeV}$ ,  $E_{particle}^Q = 30 \text{ GeV}$  and  $E_{particle}^N = 10 \text{ GeV}$ . Figures 8.25b, 8.25d and 8.26f are selected from the samples where the combination of charged (Q) and neutral (N) shower is 10 GeV Q + 10 GeV N, 30 GeV Q + 10 GeV N, and 10 GeV Q + (10 GeV N or 30 GeV N) respectively. In both figures, blue indicates simulation with the QGSP\_BERT\_HP physics list.

**Comments:**

- Good agreement is observed between the radial energy profiles distribution of the two studies, with the exception of the bin at  $K_{hit} = K_S$  comparing Figures 8.28a and 8.28b and Figures 8.28e 8.28f. The reason for this discrepancy is unclear.

# Bibliography

References are presented in order of citation.

- [1] The ATLAS Collaboration. 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC'. In: *Phys. Lett. B.* 716.1 (Sept. 2012), pp. 1–29. doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). (Visited on 06/28/2022) (cited on pages 1, 7).
- [2] *Standard Model*. In: *Wikipedia*. Page Version ID: 1157506439. May 29, 2023. (Visited on 05/29/2023) (cited on page 6).
- [3] The LHCb Collaboration. 'Observation of  $J/\psi p$  Resonances Consistent with Pentaquark States in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  Decays'. In: *Phys. Rev. Lett.* 115.7 (Aug. 12, 2015). Publisher: American Physical Society, p. 072001. doi: [10.1103/PhysRevLett.115.072001](https://doi.org/10.1103/PhysRevLett.115.072001). (Visited on 07/01/2022) (cited on page 5).
- [4] R. Oerter. *The theory of almost everything: the Standard Model, the unsung triumph of modern physics*. New York: Pi Press, 2006. 327 pp. (cited on page 6).
- [5] The CMS Collaboration. 'A measurement of the Higgs boson mass in the diphoton decay channel'. In: *Phys. Lett. B.* 805 (June 2020), p. 135425. doi: [10.1016/j.physletb.2020.135425](https://doi.org/10.1016/j.physletb.2020.135425). (Visited on 05/22/2023) (cited on pages 7, 13).
- [6] The UA1 Collaboration. 'Experimental observation of isolated large transverse energy electrons with associated missing energy at  $s = 540$  GeV'. In: *Phys. Lett. B.* 122.1 (Feb. 24, 1983), pp. 103–116. doi: [10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2). (Visited on 06/28/2022) (cited on page 7).
- [7] S. Dawson. *Introduction to the physics of Higgs bosons*. CERN Document Server. Nov. 17, 1994. URL: <https://cds.cern.ch/record/272356> (visited on 05/27/2021) (cited on pages 7, 8).
- [8] G. Moortgat-Pick et al. 'Physics at the  $e^+ - e^-$  linear collider'. In: *Eur. Phys. J. C.* 75.8 (Aug. 18, 2015). doi: [10.1140/epjc/s10052-015-3511-9](https://doi.org/10.1140/epjc/s10052-015-3511-9). (Visited on 05/19/2023) (cited on pages 7, 13, 14).
- [9] F. Abe et al. 'Observation of top quark production in  $\bar{p}p$  collisions'. In: *Phys. Rev. Lett.* 74 (1995). \_eprint: hep-ex/9503002, pp. 2626–2631. doi: [10.1103/PhysRevLett.74.2626](https://doi.org/10.1103/PhysRevLett.74.2626) (cited on page 8).
- [10] S. Abachi et al. 'Observation of the top quark'. In: *Phys. Rev. Lett.* 74 (1995). Number: arXiv:hep-ex/9503003, pp. 2632–2637. doi: [10.1103/PhysRevLett.74.2632](https://doi.org/10.1103/PhysRevLett.74.2632). (Visited on 05/23/2023) (cited on page 8).
- [11] The CMS Collaboration. 'Measurement of the top quark mass using a profile likelihood approach with the lepton+jets final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV'. In: arXiv:2302.01967 (Feb. 3, 2023). doi: [10.48550/arXiv.2302.01967](https://doi.org/10.48550/arXiv.2302.01967). (Visited on 05/23/2023) (cited on page 8).
- [12] G. Degrassi et al. 'Higgs mass and vacuum stability in the Standard Model at NNLO'. In: *J. High Energ. Phys.* 2012.8 (Aug. 2012), p. 98. doi: [10.1007/JHEP08\(2012\)098](https://doi.org/10.1007/JHEP08(2012)098). (Visited on 05/23/2023) (cited on page 8).
- [13] *Synchrotron radiation*. In: *Wikipedia*. Page Version ID: 1095267776. June 27, 2022. (Visited on 06/28/2022) (cited on page 9).
- [14] T. Behnke et al. 'The International Linear Collider Technical Design Report - Volume 1: Executive Summary'. In: (June 26, 2013). Number: arXiv:1306.6327. (Visited on 06/28/2022) (cited on pages 9–11).
- [15] J. Marshall. *PandoraPFA*. <https://github.com/PandoraPFA/>. original-date: 2015-09-30T15:28:17Z. Apr. 5, 2022. (Visited on 07/15/2022) (cited on pages 9, 32, 33).
- [16] J.-E. Augustin et al. 'Discovery of a Narrow Resonance in  $e^+ - e^-$  Annihilation'. In: *Phys. Rev. Lett.* 33 (Dec. 2, 1974), pp. 1406–1408. doi: [10.1103/PhysRevLett.33.1406](https://doi.org/10.1103/PhysRevLett.33.1406). (Visited on 05/23/2023) (cited on page 9).



- [17] M. Perl et al. 'Evidence for Anomalous Lepton Production in  $e^+ - e^-$  Annihilation'. In: *Phys. Rev. Lett.* 35 (Dec. 1, 1975). Publisher: American Physical Society, pp. 1489–1492. doi: [10.1103/PhysRevLett.35.1489](https://doi.org/10.1103/PhysRevLett.35.1489). (Visited on 05/23/2023) (cited on page 9).
- [18] A. Abada et al. 'FCC-ee: The Lepton Collider'. In: *Eur. Phys. J. Spec. Top.* 228.2 (June 1, 2019), pp. 261–623. doi: [10.1140/epjst/e2019-900045-4](https://doi.org/10.1140/epjst/e2019-900045-4). (Visited on 04/05/2023) (cited on page 10).
- [19] P. N. Burrows et al. 'CERN Yellow Reports: The Compact Linear  $e^+ - e^-$  Collider (CLIC): 2018 Summary Report'. In: (Jan. 1, 1970). Artwork Size: 58.39 MB Publisher: CERN, 58.39 MB. doi: [10.23731/CYRM-2018-002](https://doi.org/10.23731/CYRM-2018-002). (Visited on 06/28/2022) (cited on pages 10, 11).
- [20] The LHC Higgs Cross Section Working Group. 'Handbook of LHC Higgs Cross Sections: 3. Higgs Properties'. In: (2013). doi: [10.5170/CERN-2013-004](https://doi.org/10.5170/CERN-2013-004). (Visited on 06/29/2022) (cited on page 11).
- [21] M. Tanabashi, K. Hagiwara, K. Hikasa, et al. 'Review of Particle Physics'. In: *Phys. Rev. D* 98.3 (Aug. 2018). Publisher: American Physical Society, p. 030001. doi: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001) (cited on pages 12, 13, 21, 31).
- [22] A. Djouadi et al. 'International Linear Collider Reference Design Report Volume 2'. In: arXiv:0709.1893 (Sept. 12, 2007). (Visited on 07/04/2022) (cited on page 12).
- [23] K. Fujii et al. 'Physics Case for the International Linear Collider'. In: *arXiv* arXiv:1506.05992 (June 26, 2015). doi: [10.48550/arXiv.1506.05992](https://doi.org/10.48550/arXiv.1506.05992). (Visited on 05/23/2023) (cited on page 13).
- [24] C. Dürig et al. 'Model Independent Determination of  $HWW$  coupling and Higgs total width at ILC'. In: *arXiv* arXiv:1403.7734 (Apr. 1, 2014). (Visited on 05/23/2023) (cited on page 14).
- [25] The ATLAS Collaboration. 'Evidence of off-shell Higgs boson production from ZZ leptonic decay channels and constraints on its total width with the ATLAS detector'. In: *arXiv* arXiv:2304.01532 (Apr. 4, 2023). doi: [10.48550/arXiv.2304.01532](https://doi.org/10.48550/arXiv.2304.01532). (Visited on 05/23/2023) (cited on page 14).
- [26] D. Groom and S. Klein. 'Passage of Particles Through Matter'. In: *Eur. Phys. J. C* 15.1 (Mar. 2000), pp. 163–173. doi: [10.1007/BF02683419](https://doi.org/10.1007/BF02683419). (Visited on 03/05/2021) (cited on page 15).
- [27] L. Landau. 'On the energy loss of fast particles by ionization'. In: *J. Phys.* 8 (1944), p. 201. (Visited on 05/28/2021) (cited on page 15).
- [28] M. Livan and R. Wigmans. *Calorimetry for Collider Physics, an Introduction*. UNITEXT for Physics. Springer International Publishing, 2019. (Visited on 03/10/2021) (cited on pages 16–19, 21, 24, 27, 28, 30, 39, 115, 141, 206).
- [29] B. Berman and S. Fultz. 'Measurements of the giant dipole resonance with monoenergetic photons'. In: *Rev. Mod. Phys.*, v. 47, no. 3, pp. 713–761 (July 1, 1975). Institution: Lawrence Livermore Laboratory, University of California, Livermore, California 94550. doi: [10.1103/RevModPhys.47.713](https://doi.org/10.1103/RevModPhys.47.713). (Visited on 06/30/2022) (cited on page 17).
- [30] R. Wigmans and M. Zeyrek. 'On the differences between calorimetric detection of electrons and photons'. In: *Nucl. Instrum. Meth. A* 485 (2002), pp. 385–398. doi: [10.1016/S0168-9002\(01\)02141-6](https://doi.org/10.1016/S0168-9002(01)02141-6) (cited on page 18).
- [31] J. Krüger. 'The Uranium Scintillator Calorimeter for the ZEUS Detector at the Electron - Proton Collider HERA: The Heart of ZEUS'. PhD thesis. Universität Hamburg, 1992 (cited on pages 18, 28, 29).
- [32] The CALICE Collaboration. 'The Time Structure of Hadronic Showers in highly granular Calorimeters with Tungsten and Steel Absorbers'. In: *J. Inst.* 9.7 (July 18, 2014), P07022–P07022. doi: [10.1088/1748-0221/9/07/P07022](https://doi.org/10.1088/1748-0221/9/07/P07022). (Visited on 04/22/2022) (cited on pages 19, 21, 39, 115, 144, 162).
- [33] M. Thomson. 'Handout 9: The Weak Interaction and V-A'. 2009 (cited on page 20).
- [34] O. Pinto. 'Shower Shapes in a Highly Granular SiPM-on-Tile Analog Hadron Calorimeter'. Accepted: 2022-11-03T12:50:01Z. doctoralThesis. Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky, 2022. (Visited on 11/29/2022) (cited on pages 22, 34, 39, 41, 50, 51, 134, 135, 142, 150, 193, 257).
- [35] *Matter wave*. In: *Wikipedia*. Page Version ID: 1141800035. Feb. 26, 2023. (Visited on 03/17/2023) (cited on page 22).

- [36] The Geant4 Electromagnetic Physics Working Group. 'Progress in Geant4 Electromagnetic Physics Modelling and Validation'. In: *J. Phys.: Conf. Ser.* 664.7 (). Publisher: IOP Publishing, p. 072021. doi: [10.1088/1742-6596/664/7/072021](https://doi.org/10.1088/1742-6596/664/7/072021). (Visited on 03/17/2023) (cited on page 22).
- [37] A. Heikkinen, N. Stepanov, and J. Wellisch. 'Bertini intra-nuclear cascade implementation in Geant4'. In: arXiv:nucl-th/0306008 (June 2, 2003). doi: [10.48550/arXiv.nucl-th/0306008](https://doi.org/10.48550/arXiv.nucl-th/0306008). (Visited on 03/17/2023) (cited on pages 22, 23).
- [38] D. Wright and M. Kelsey. 'The Geant4 Bertini Cascade'. In: *Nucl.Instrum.Meth.A* 804 (Dec. 21, 2015), pp. 175–188. doi: [10.1016/j.nima.2015.09.058](https://doi.org/10.1016/j.nima.2015.09.058). (Visited on 03/17/2023) (cited on page 22).
- [39] *Fermi gas*. In: *Wikipedia*. Feb. 11, 2023. (Visited on 03/17/2023) (cited on page 22).
- [40] G. Folger and J. Wellisch. 'String Parton Models in Geant4'. In: arXiv:nucl-th/0306007 (June 2, 2003). doi: [10.48550/arXiv.nucl-th/0306007](https://doi.org/10.48550/arXiv.nucl-th/0306007). (Visited on 03/17/2023) (cited on pages 23, 24).
- [41] *Regge theory*. In: *Wikipedia*. Page Version ID: 1100147204. July 24, 2022. (Visited on 03/17/2023) (cited on page 24).
- [42] The Geant4 Collaboration. 'Geant4 Physics Reference Manual'. In: (Dec. 9, 2022) (cited on page 24).
- [43] O. Hartbrich. *Scintillator Calorimeters for a Future Linear Collider Experiment*. Artwork Size: pages 182 Publication Title: Bergische Universität Wuppertal Volume: Diss. Deutsches Elektronen-Synchrotron, DESY, Hamburg, 2016, pages 182. doi: [10.3204/PUBDB-2016-02800](https://doi.org/10.3204/PUBDB-2016-02800). (Visited on 03/17/2023) (cited on page 24).
- [44] J. Apostolakis et al. 'Geant4 Physics Lists for HEP'. In: *2008 IEEE Nuclear Science Symposium Conference Record*. 2008 IEEE Nuclear Science Symposium and Medical Imaging conference (2008 NSS/MIC). Dresden, Germany: IEEE, Oct. 2008, pp. 833–836. doi: [10.1109/NSSMIC.2008.4774655](https://doi.org/10.1109/NSSMIC.2008.4774655). (Visited on 03/17/2023) (cited on page 24).
- [45] The OPAL Collaboration. 'The OPAL detector at LEP'. In: *Nucl.Instrum.Meth.A* 305.2 (July 20, 1991), pp. 275–319. doi: [10.1016/0168-9002\(91\)90547-4](https://doi.org/10.1016/0168-9002(91)90547-4). (Visited on 07/05/2022) (cited on page 25).
- [46] The CMS Collaboration. 'The CMS experiment at the CERN LHC'. In: *J. Inst.* 3.8 (Aug. 14, 2008), S08004–S08004. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004). (Visited on 04/11/2023) (cited on page 26).
- [47] A. Henriques. 'The ATLAS tile calorimeter'. In: *2015 4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*. 2015 4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA). Apr. 2015, pp. 1–7. doi: [10.1109/ANIMMA.2015.7465554](https://doi.org/10.1109/ANIMMA.2015.7465554) (cited on page 26).
- [48] T. Davidek. 'Performance and Calibration of the ATLAS Tile Calorimeter'. In: *Instruments* 6.3 (Sept. 2022). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 25. doi: [10.3390/instruments6030025](https://doi.org/10.3390/instruments6030025). (Visited on 04/27/2023) (cited on page 26).
- [49] *Cherenkov radiation*. In: *Wikipedia*. Page Version ID: 1092672832. June 11, 2022. (Visited on 07/12/2022) (cited on page 29).
- [50] R. Wigmans. 'The new RD52 (DREAM) fiber calorimeter'. In: *J. Phys.: Conf. Ser.* 404 (Dec. 2012). Publisher: IOP Publishing, p. 012068. doi: [10.1088/1742-6596/404/1/012068](https://doi.org/10.1088/1742-6596/404/1/012068). (Visited on 07/14/2022) (cited on page 29).
- [51] G. Gaudio, R. Wigmans, and J. Hauptman. 'RD52 Status Report 2014 - Dual-Readout Calorimetry for High-Quality Energy Measurements'. In: (2014). Place: Geneva. (Visited on 05/12/2023) (cited on page 29).
- [52] S. Lee, M. Livan, and R. Wigmans. 'Dual-Readout Calorimetry'. In: *Reviews of Modern Physics* 90.2 (Apr. 26, 2018), p. 025002. doi: [10.1103/RevModPhys.90.025002](https://doi.org/10.1103/RevModPhys.90.025002). (Visited on 04/28/2023) (cited on page 29).
- [53] J. Birks. 'Scintillations from Organic Crystals: Specific Fluorescence and Relative Response to Different Radiations'. In: *Proceedings of the Physical Society. Section A* 64.10 (Oct. 1, 1951), pp. 874–877. doi: [10.1088/0370-1298/64/10/303](https://doi.org/10.1088/0370-1298/64/10/303). (Visited on 03/17/2023) (cited on page 30).

- [54] S. Krause. 'Studies of the Response of Silicon Photomultipliers and Testbeam Data Analysis of a Highly Granular Analog Hadron Calorimeter Prototype'. PhD thesis. Johannes Gutenberg-Universität Mainz, 2020. doi: [10.25358/OPENSCIENCE-1642](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-65358-OPENSCIENCE-1642). (Visited on 07/28/2022) (cited on pages 30, 50).
- [55] M. Thomson. 'Particle Flow Calorimetry and the PandoraPFA Algorithm'. In: *Nucl.Instrum.Meth.A* 611.1 (Nov. 2009), pp. 25–40. doi: [10.1016/j.nima.2009.09.009](https://doi.org/10.1016/j.nima.2009.09.009). (Visited on 07/14/2022) (cited on pages 31–34, 165, 168, 177, 226).
- [56] The ALEPH Collaboration. 'Performance of the ALEPH detector at LEP'. In: *Nucl.Instrum.Meth.A* 360 (1995), pp. 481–506. doi: [10.1016/0168-9002\(95\)00138-7](https://doi.org/10.1016/0168-9002(95)00138-7). (Visited on 07/15/2022) (cited on page 32).
- [57] M. Thompson. 'Advanced Particle Flow'. AIDA-2020 First annual meeting. (Visited on 07/15/2022) (cited on page 32).
- [58] F. Gaede et al. 'Track reconstruction at the ILC: the ILD tracking software'. In: 513.2 (June 2014). Publisher: IOP Publishing, p. 022011. doi: [10.1088/1742-6596/513/2/022011](https://doi.org/10.1088/1742-6596/513/2/022011). (Visited on 07/15/2022) (cited on page 32).
- [59] H. Tran et al. 'Software compensation in particle flow reconstruction'. In: *Eur Phys J C Part Fields* 77.10 (2017), p. 698. doi: [10.1140/epjc/s10052-017-5298-3](https://doi.org/10.1140/epjc/s10052-017-5298-3). (Visited on 07/18/2022) (cited on pages 34, 116).
- [60] The CALICE Collaboration. 'Design, Construction and Commissioning of a Technological Prototype of a Highly Granular SiPM-on-tile Scintillator-Steel Hadronic Calorimeter'. In: *arXiv* arXiv:2209.15327 (Sept. 30, 2022). doi: [10.48550/arXiv.2209.15327](https://doi.org/10.48550/arXiv.2209.15327). (Visited on 05/30/2023) (cited on pages 34, 35, 40, 123, 175).
- [61] M. Chadeeva. 'Extraction of h/e and calorimeter response from fits to the longitudinal shower profiles in the CALICE Sc-Fe AHCAL'. In: (Oct. 2014) (cited on page 35).
- [62] *3M™ Enhanced Specular Reflector (ESR) - Application Guide*. URL: <https://multimedia.3m.com/mws/media/13892480/application-guide-for-esr.pdf> (cited on page 35).
- [63] The CALICE Collaboration. 'A highly granular SiPM-on-tile calorimeter prototype'. In: 1162 (Jan. 2019), p. 012012. doi: [10.1088/1742-6596/1162/1/012012](https://doi.org/10.1088/1742-6596/1162/1/012012). (Visited on 07/18/2022) (cited on page 35).
- [64] Y. Liu et al. 'A Design of Scintillator Tiles Read Out by Surface-Mounted SiPMs for a Future Hadron Calorimeter'. In: arXiv:1512.05900 (Dec. 18, 2015). (Visited on 07/18/2022) (cited on page 35).
- [65] M. Robles-Manzano. 'Beam Tests of the CALICE AHCAL'. 8th Beam Telescopes and Test Beams Workshop. Jan. 29, 2020. (Visited on 08/01/2022) (cited on page 36).
- [66] 'Omega SPIROC2 Datasheet'. In: (), p. 48 (cited on page 36).
- [67] G. Knoll. *Radiation detection and measurement*. 3rd ed. New York: Wiley, 2000. 802 pp. (cited on page 36).
- [68] L. Emberger. 'AHCAL Time Calibration'. CALICE Collaboration Meeting at CERN. Oct. 1, 2019. (Visited on 07/28/2022) (cited on pages 36, 37).
- [69] The CALICE Collaboration, ed. *Time Analysis of the Partially Equipped CALICE Analog Hadronic Calorimeter Technological Prototype with Tungsten Absorber*. 2019. 16 pp. (cited on page 36).
- [70] A. Elkhali. 'Analog Hadronic Calorimeter for a Future Linear Collider'. Wissenschaftliche Abschlussarbeiten » Dissertation. Universität Wuppertal, Fakultät für Mathematik und Naturwissenschaften » Physik » Dissertationen, Nov. 5, 2020. doi: [10.25926/2jx6-f298](https://doi.org/10.25926/2jx6-f298). (Visited on 07/19/2022) (cited on page 36).
- [71] C. Graf. 'Power Pulsing Studies of an Extended Analogue Hadron Calorimeter Layer'. Bachelorarbeit. Universität Heidelberg, 2019 (cited on page 37).
- [72] I. Polak and J. Kvasnicka. 'Fast calibration UV LED system for CALICE scintillator based tile hadron calorimeter'. In: *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC). ISSN: 1082-3654. Oct. 2012, pp. 1240–1243. doi: [10.1109/NSSMIC.2012.6551304](https://doi.org/10.1109/NSSMIC.2012.6551304) (cited on page 37).



- [73] The CALICE Collaboration, ILD concept group. *Calibration of the Scintillator Hadron Calorimeter of ILD*. CALICE Analysis Note CAN-018, CALICE-CAN-2009-003. 2009 (cited on page 37).
- [74] *atan2*. In: *Wikipedia*. Page Version ID: 1100166236. July 24, 2022. (Visited on 07/26/2022) (cited on page 38).
- [75] B. Acar et al. 'Performance of the CMS High Granularity Calorimeter prototype to charged pion beams of 20 – 300 GeV/c'. In: (2023) (cited on page 38).
- [76] D. Heuchel. 'Shower Start Finding and Pandora PFA Studies - AHCAL 2018 Prototype'. CALICE Collaboration Meeting at CERN. CERN Main Auditorium, Oct. 2, 2019. (Visited on 07/28/2022) (cited on pages 38, 221, 293–296).
- [77] K. Krüger. 'The CALICE AHCAL - a highly granular SiPM-on-tile hadron calorimeter prototype'. CHEF2019 - Calorimetry for The High Energy Frontier. Centennial Hall, Kyushu University, Japan, Nov. 25, 2019. (Visited on 07/25/2022) (cited on pages 39, 40).
- [78] D. Heuchel. 'Particle Flow Studies with Highly Granular Calorimeter Data'. Place: Heidelberg. Dissertation. 2022. doi: [10.11588/heidok.00031794](https://doi.org/10.11588/heidok.00031794). (Visited on 07/25/2022) (cited on pages 39, 50, 151, 177, 178, 181, 182, 188, 189, 220, 223, 225, 228, 253, 293–296).
- [79] V. Bocharnikov. 'Particle identification methods for the CALICE highly granular SiPM-on tile calorimeter'. In: *Verhandlungen der Deutschen Physikalischen Gesellschaft (Aache2019issue 2019)*. Place: Germany INIS Reference Number: 51001777, p. 1 (cited on pages 39, 52, 123, 176, 178, 183, 186, 287, 288).
- [80] The CALICE Collaboration. *Software Compensation for Hadronic Showers in the CALICE AHCAL and Tail Catcher with Cluster-based Methods*. CALICE Analysis Note CAN-021, CALICE-CAN-2010-003. CALICE, 2010 (cited on pages 40, 116).
- [81] L. Liu. 'Beam tracking in AHCAL prototype test beam'. CALICE Collaboration Meeting at Shanghai. Shanghai Jiao Tong University, Sept. 20, 2018. (Visited on 07/25/2022) (cited on page 40).
- [82] *iLCSoft*. <https://github.com/iLCSoft>. (Visited on 07/26/2022) (cited on page 41).
- [83] S. Agostinelli et al. 'Geant4—a simulation toolkit'. In: *Nucl.Instrum.Meth.A* 506.3 (July 1, 2003), pp. 250–303. doi: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). (Visited on 06/02/2021) (cited on pages 41, 123, 175).
- [84] M. Petrič et al. 'Detector Simulations with DD4hep'. In: *J. Phys.: Conf. Ser.* 898 (Oct. 2017), p. 042015. doi: [10.1088/1742-6596/898/4/042015](https://doi.org/10.1088/1742-6596/898/4/042015). (Visited on 06/02/2021) (cited on pages 41, 123, 175).
- [85] F. Gaede. 'Marlin and LCCD—Software tools for the ILC'. In: *Nucl.Instrum.Meth.A*. Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research 559.1 (Apr. 1, 2006), pp. 177–180. doi: [10.1016/j.nima.2005.11.138](https://doi.org/10.1016/j.nima.2005.11.138). (Visited on 07/25/2022) (cited on page 41).
- [86] The CALICE Collaboration. *CALICESoft*. <https://stash.desy.de/projects/CALICE>. (Visited on 07/27/2022) (cited on pages 41, 123, 175, 183, 186, 287, 288).
- [87] L. Emberger. 'Precision Timing in Highly Granular Calorimeters and Applications in Long Baseline Neutrino and Lepton Collider Experiments'. PhD thesis. Munich, Tech. U., 2022 (cited on page 42).
- [88] *Fermi level*. In: *Wikipedia*. Page Version ID: 1088309216. May 17, 2022. (Visited on 07/20/2022) (cited on page 43).
- [89] A. S. Grove. *Physics and technology of semiconductor devices*. New York, NY: Wiley, 1967. 366 pp. (cited on page 44).
- [90] *A technical guide to silicon photomultipliers (MPPC) - Section 1 | Hamamatsu Photonics*. URL: <https://hub.hamamatsu.com/us/en/technical-notes/mppc-sipms/a-technical-guide-to-silicon-photomultipliers-MPPC-Section-1.html> (visited on 07/21/2022) (cited on page 45).
- [91] R. Klanner. 'Characterisation of SiPMs'. In: *Nucl.Instrum.Meth.A* 926 (May 2019), pp. 36–56. doi: [10.1016/j.nima.2018.11.083](https://doi.org/10.1016/j.nima.2018.11.083). (Visited on 03/07/2023) (cited on pages 45–47, 76, 91).
- [92] *Hamamatsu Photonics*. URL: <https://www.hamamatsu.com/jp/en.html> (visited on 08/02/2022) (cited on page 46).

- [93] V. Chmill et al. ‘Study of the breakdown voltage of SiPMs’. In: *Nucl.Instrum.Meth.A* 845 (Feb. 2017), pp. 56–59. doi: [10.1016/j.nima.2016.04.047](https://doi.org/10.1016/j.nima.2016.04.047). (Visited on 03/07/2023) (cited on pages 46, 81, 106).
- [94] O. Marinov et al. ‘Theory of microplasma fluctuations and noise in silicon diode in avalanche breakdown’. In: *J. Appl. Phys.* 101.6 (Mar. 30, 2007). Publisher: American Institute of Physics AIP, p. 064515. doi: [10.1063/1.2654973](https://doi.org/10.1063/1.2654973). (Visited on 10/07/2022) (cited on page 46).
- [95] C. Piemonte and A. Gola. ‘Overview on the main parameters and technology of modern Silicon Photomultipliers’. In: *Nucl.Instrum.Meth.A* 926 (May 2019), pp. 2–15. doi: [10.1016/j.nima.2018.11.119](https://doi.org/10.1016/j.nima.2018.11.119). (Visited on 03/07/2023) (cited on pages 47, 48, 76).
- [96] E. Garutti, R. Klanner, J. Rolph, et al. ‘A computer program to simulate the response of SiPMs’. In: arXiv:2006.11150 (June 19, 2020). (Visited on 07/27/2022) (cited on pages 47, 48, 63, 68, 76, 77, 81–83, 88, 90, 112).
- [97] V. Chmill et al. ‘On the characterisation of SiPMs from pulse-height spectra’. In: *Nucl.Instrum.Meth.A* 854 (May 2017), pp. 70–81. doi: [10.1016/j.nima.2017.02.049](https://doi.org/10.1016/j.nima.2017.02.049). (Visited on 07/07/2022) (cited on pages 47, 75–77, 79–82, 88, 91, 112).
- [98] A. Provenza, K. Krüger, and E. Garutti. ‘Calibration and Analysis of Data taken with the Technological Prototype of the Analog Hadron Calorimeter (AHCAL) for a Detector at the International Linear Collider’. Number: PUBDB-2019-00522. PhD thesis. Verlag Deutsches Elektronen-Synchrotron, 2018. (Visited on 07/29/2022) (cited on page 50).
- [99] *Decision tree learning*. In: *Wikipedia*. Page Version ID: 1098190172. July 14, 2022. (Visited on 08/01/2022) (cited on page 52).
- [100] *Gradient boosting*. In: *Wikipedia*. July 21, 2022. (Visited on 08/01/2022) (cited on page 52).
- [101] *NVIDIA*. URL: <https://www.nvidia.com/> (visited on 08/02/2022) (cited on page 52).
- [102] *AMD*. URL: <https://www.amd.com/> (visited on 08/02/2022) (cited on page 52).
- [103] T. Brown et al. ‘Language Models are Few-Shot Learners’. In: (2020) (cited on page 52).
- [104] *GPT-3*. In: *Wikipedia*. Page Version ID: 1101347177. July 30, 2022. (Visited on 08/02/2022) (cited on page 52).
- [105] G. Karagiorgi et al. ‘Machine Learning in the Search for New Fundamental Physics’. In: arXiv:2112.03769 (Dec. 7, 2021). (Visited on 08/01/2022) (cited on page 52).
- [106] The ATLAS Collaboration. ‘ATLAS  $b$ -jet identification performance and efficiency measurement with  $t\bar{t}$  events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV’. In: *Eur. Phys. J. C* 79.11 (), p. 970. doi: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8). (Visited on 08/01/2022) (cited on page 52).
- [107] The ATLAS Collaboration. *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*. Tech. rep. Geneva: CERN, 2018 (cited on page 52).
- [108] The DUNE Collaboration. ‘Neutrino interaction classification with a convolutional neural network in the DUNE far detector’. In: *Phys. Rev. D* 102.9 (Nov. 9, 2020). Publisher: American Physical Society, p. 092003. doi: [10.1103/PhysRevD.102.092003](https://doi.org/10.1103/PhysRevD.102.092003). (Visited on 08/01/2022) (cited on page 52).
- [109] *A Living Review of Machine Learning for Particle Physics*. HEPML-LivingReview. URL: <https://iml-wg.github.io/HEPML-LivingReview/> (visited on 08/02/2022) (cited on page 52).
- [110] E. Buhmann et al. ‘Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed’. In: *Comput Softw Big Sci* 5.1 (May 26, 2021), p. 13. doi: [10.1007/s41781-021-00056-0](https://doi.org/10.1007/s41781-021-00056-0). (Visited on 03/27/2023) (cited on page 52).
- [111] G. Sanderson. *Backpropagation calculus | Chapter 4, Deep learning*. Nov. 3, 2017. (Visited on 08/02/2022) (cited on page 53).
- [112] M. Abadi et al. ‘TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems’. In: arXiv:1603.04467 (Mar. 16, 2016). doi: [10.48550/arXiv.1603.04467](https://doi.org/10.48550/arXiv.1603.04467). (Visited on 03/27/2023) (cited on page 56).
- [113] A. Paszke et al. ‘Automatic differentiation in PyTorch’. In: (2017) (cited on pages 56, 139, 204).

- [114] G. Cybenko. 'Approximation by superpositions of a sigmoidal function'. In: *Math. Control Signal Systems* 2.4 (Dec. 1, 1989), pp. 303–314. doi: [10.1007/BF02551274](https://doi.org/10.1007/BF02551274). (Visited on 08/02/2022) (cited on page 56).
- [115] S. Basodi et al. 'Gradient amplification: An efficient way to train deep neural networks'. In: *Big Data Mining and Analytics* 3.3 (Sept. 2020). Conference Name: Big Data Mining and Analytics, pp. 196–207. doi: [10.26599/BDMA.2020.9020004](https://doi.org/10.26599/BDMA.2020.9020004) (cited on page 59).
- [116] D. Kingma and J. Ba. 'Adam: A Method for Stochastic Optimization'. In: arXiv:1412.6980 (Jan. 29, 2017). (Visited on 05/03/2023) (cited on page 60).
- [117] S. Thais et al. 'Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges'. In: arXiv:2203.12852 (Mar. 25, 2022). doi: [10.48550/arXiv.2203.12852](https://doi.org/10.48550/arXiv.2203.12852). (Visited on 08/02/2022) (cited on pages 60, 61).
- [118] J. Shlomi, P. Battaglia, and J-R. Vlimant. 'Graph neural networks in particle physics'. In: *Mach. Learn.: Sci. Technol.* 2.2 (Jan. 8, 2021), p. 021001. doi: [10.1088/2632-2153/abbf9a](https://doi.org/10.1088/2632-2153/abbf9a). (Visited on 08/06/2022) (cited on pages 60, 62).
- [119] B. Sanchez-Lengeling et al. 'A Gentle Introduction to Graph Neural Networks'. In: *Distill* 6.9 (Sept. 2, 2021), e33. doi: [10.23915/distill.00033](https://doi.org/10.23915/distill.00033). (Visited on 10/28/2022) (cited on page 62).
- [120] S. Qasim et al. 'Learning representations of irregular particle-detector geometry with distance-weighted graph networks'. In: *Eur. Phys. J. C* 79.7 (July 18, 2019), p. 608. doi: [10.1140/epjc/s10052-019-7113-9](https://doi.org/10.1140/epjc/s10052-019-7113-9). (Visited on 08/08/2022) (cited on pages 61, 166–169, 171, 173, 175, 177, 189, 194, 204, 205, 211, 225).
- [121] A. Spinelli and A. Lacaita. 'Physics and numerical simulation of single photon avalanche diodes'. In: *IEEE Transactions on Electron Devices* 44.11 (Nov. 1997). Conference Name: IEEE Transactions on Electron Devices, pp. 1931–1943. doi: [10.1109/16.641363](https://doi.org/10.1109/16.641363) (cited on page 64).
- [122] A. Fabio and S. Gundacker. 'Understanding and simulating SiPMs'. In: *Nucl.Instrum.Meth.A* 926 (May 2019), pp. 16–35. doi: [10.1016/j.nima.2018.11.118](https://doi.org/10.1016/j.nima.2018.11.118). (Visited on 03/07/2023) (cited on pages 64, 76).
- [123] P. Calò et al. 'SiPM readout electronics'. In: *Nucl.Instrum.Meth.A. Silicon Photomultipliers: Technology, Characterisation and Applications* 926 (May 11, 2019), pp. 57–68. doi: [10.1016/j.nima.2018.09.030](https://doi.org/10.1016/j.nima.2018.09.030). (Visited on 03/15/2023) (cited on page 64).
- [124] J. Pulko et al. 'A Monte-Carlo model of a SiPM coupled to a scintillating crystal'. In: *J. Inst.* 7 (Feb. 1, 2012). ADS Bibcode: 2012JInst...7.2009P, p. 2009. doi: [10.1088/1748-0221/7/02/P02009](https://doi.org/10.1088/1748-0221/7/02/P02009). (Visited on 03/15/2023) (cited on page 64).
- [125] K. Jha et al. 'Simulating Silicon Photomultiplier Response to Scintillation Light'. In: *IEEE transactions on nuclear science* 30.1 (Feb. 2013), pp. 336–351. doi: [10.1109/TNS.2012.2234135](https://doi.org/10.1109/TNS.2012.2234135) (cited on page 64).
- [126] T. Niggemann et al. 'G4SiPM: A novel silicon photomultiplier simulation package for Geant4'. In: *Nucl.Instrum.Meth.A* 787 (July 2015), pp. 344–347. doi: [10.1016/j.nima.2015.01.067](https://doi.org/10.1016/j.nima.2015.01.067). (Visited on 03/15/2023) (cited on page 64).
- [127] E. Dietz-Laursonn et al. 'GODDeSS: a Geant4 extension for easy modelling of optical detector components'. In: *J. Inst.* 12.4 (Apr. 26, 2017), P04026–P04026. doi: [10.1088/1748-0221/12/04/P04026](https://doi.org/10.1088/1748-0221/12/04/P04026). (Visited on 03/15/2023) (cited on page 64).
- [128] S. Gundacker et al. 'Time of flight positron emission tomography towards 100ps resolution with L(Y)SO: an experimental and theoretical analysis'. In: *J. Inst.* 8.7 (July 29, 2013), P07014–P07014. doi: [10.1088/1748-0221/8/07/P07014](https://doi.org/10.1088/1748-0221/8/07/P07014). (Visited on 03/15/2023) (cited on page 64).
- [129] *Borel distribution*. In: *Wikipedia*. Page Version ID: 992690075. Dec. 6, 2020. (Visited on 03/16/2023) (cited on page 67).
- [130] C. Harris et al. 'Array programming with NumPy'. In: *Nature* 585.7825 (Sept. 17, 2020), pp. 357–362. doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). (Visited on 03/07/2023) (cited on pages 70, 83).
- [131] P. Virtanen et al. 'SciPy 1.0: fundamental algorithms for scientific computing in Python'. In: *Nat Methods* 17.3 (Mar. 2, 2020), pp. 261–272. doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2). (Visited on 03/16/2023) (cited on page 70).

- [132] The Pandas Development Team. *pandas-dev/pandas: Pandas*. <https://zenodo.org/record/3509134>. Version v2.0.0rc1. Mar. 16, 2023. doi: [10.5281/ZENODO.3509134](https://doi.org/10.5281/ZENODO.3509134). (Visited on 03/16/2023) (cited on page 70).
- [133] J. Hunter. 'Matplotlib: A 2D Graphics Environment'. In: *Comput. Sci. Eng.* 9.3 (2007), pp. 90–95. doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55). (Visited on 03/16/2023) (cited on page 70).
- [134] V. Arosio et al. 'An educational kit based on a modular Silicon Photomultiplier system'. In: *2013 3rd International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications (ANIMMA)*. 2013 3rd International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications (ANIMMA). Marseille, France: IEEE, June 2013, pp. 1–7. doi: [10.1109/ANIMMA.2013.6728000](https://doi.org/10.1109/ANIMMA.2013.6728000). (Visited on 03/07/2023) (cited on pages 70, 71, 102).
- [135] E. Garutti et al. 'Simulation of the response of SiPMs; Part I: Without saturation effects'. In: *Nucl.Instrum.Meth.A* 1019 (Dec. 2021), p. 165853. doi: [10.1016/j.nima.2021.165853](https://doi.org/10.1016/j.nima.2021.165853). (Visited on 07/07/2022) (cited on page 75).
- [136] J. Rolph et al. 'PeakOTron: A Python Module for Fitting Charge Spectra of Silicon Photomultipliers'. In: arXiv:2301.11833 (Jan. 27, 2023). doi: [10.48550/arXiv.2301.11833](https://doi.org/10.48550/arXiv.2301.11833). (Visited on 03/08/2023) (cited on page 75).
- [137] P. Eckert et al. 'Characterisation studies of silicon photomultipliers'. In: *Nucl.Instrum.Meth.A* 620.2 (Aug. 2010), pp. 217–226. doi: [10.1016/j.nima.2010.03.169](https://doi.org/10.1016/j.nima.2010.03.169). (Visited on 03/07/2023) (cited on page 76).
- [138] V. Arosio et al. 'A robust and semi-automatic procedure for Silicon Photomultipliers characterisation'. In: *J. Inst.* 12.3 (Mar. 8, 2017), pp. C03030–C03030. doi: [10.1088/1748-0221/12/03/C03030](https://doi.org/10.1088/1748-0221/12/03/C03030). (Visited on 03/07/2023) (cited on page 76).
- [139] M. Zvolnsky. 'Simulation, Image Reconstruction and SiPM Characterisation for a Novel Endoscopic Positron Emission Tomography Detector'. Artwork Size: pages 219 Pages: pages 219 Publication Title: Dissertation. 2017. doi: [10.3204/PUBDB-2017-13685](https://doi.org/10.3204/PUBDB-2017-13685). (Visited on 03/07/2023) (cited on page 76).
- [140] *Convolution of probability distributions*. In: *Wikipedia*. Page Version ID: 1123552785. Nov. 24, 2022. (Visited on 03/07/2023) (cited on page 77).
- [141] P. Consul and G. Jain. 'A Generalization of the Poisson Distribution'. In: *Technometrics* 15.4 (Nov. 1973), pp. 791–799. doi: [10.1080/00401706.1973.10489112](https://doi.org/10.1080/00401706.1973.10489112). (Visited on 03/07/2023) (cited on pages 77, 84, 85).
- [142] P. Consul and F. Famoye. 'Generalized Poisson Distribution'. In: *Lagrangian Probability Distributions*. Boston: Birkhäuser-Verlag, 2006, pp. 165–190. doi: [10.1007/0-8176-4477-6\\_9](https://doi.org/10.1007/0-8176-4477-6_9). (Visited on 03/07/2023) (cited on pages 77, 78).
- [143] S. Vinogradov. 'Analytical models of probability distribution and excess noise factor of solid state photomultiplier signals with crosstalk'. In: *Nucl.Instrum.Meth.A*. New Developments in Photodetection NDIP11 695 (Dec. 11, 2012), pp. 247–251. doi: [10.1016/j.nima.2011.11.086](https://doi.org/10.1016/j.nima.2011.11.086). (Visited on 03/07/2023) (cited on page 78).
- [144] S. Vinogradov. 'Skewness-based characterization of silicon photomultipliers'. In: *Eur. Phys. J. C* 82.5 (May 2022), p. 490. doi: [10.1140/epjc/s10052-022-10444-4](https://doi.org/10.1140/epjc/s10052-022-10444-4). (Visited on 03/07/2023) (cited on pages 78, 84, 85).
- [145] The Astropy Collaboration. 'Astropy: A community Python package for astronomy'. In: *Astronomy & Astrophysics* 558 (Oct. 2013), A33. doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068). (Visited on 03/07/2023) (cited on page 83).
- [146] D. Scott. 'On optimal and data-based histograms'. In: *Biometrika* 66.3 (1979), pp. 605–610. doi: [10.1093/biomet/66.3.605](https://doi.org/10.1093/biomet/66.3.605). (Visited on 03/07/2023) (cited on page 83).
- [147] D. Freedman and P. Diaconis. 'On the histogram as a density estimator: L2 theory'. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57.4 (Dec. 1, 1981), pp. 453–476. doi: [10.1007/BF01025868](https://doi.org/10.1007/BF01025868). (Visited on 11/27/2022) (cited on pages 83, 140, 189).



- [148] K. Knuth. ‘Optimal Data-Based Binning for Histograms’. In: *arXiv:physics/0605197* (Sept. 16, 2013). (Visited on 06/03/2021) (cited on page 83).
- [149] *scipy.interpolate.UnivariateSpline* — *SciPy v1.10.1 Manual*. URL: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.UnivariateSpline.html> (visited on 03/07/2023) (cited on page 84).
- [150] P. Huber. ‘Robust Statistics’. In: *Int. Encycl. Stat. Sci.* Ed. by L. Miodrag. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1248–1251. DOI: 10.1007/978-3-642-04898-2\_594 (cited on page 87).
- [151] H. Dembinski, O. Piti, and C. Deil. *scikit-hep/iminuit: v2.8.4*. <https://zenodo.org/record/5561211>. Version v2.8.4. Oct. 11, 2021. DOI: 10.5281/ZENODO.5561211. (Visited on 03/07/2023) (cited on page 89).
- [152] *multiprocessing* — *Process-based parallelism*. Python documentation. URL: <https://docs.python.org/3/library/multiprocessing.html> (visited on 03/07/2023) (cited on page 90).
- [153] *joblib/joblib*. <https://github.com/joblib/joblib>. original-date: 2010-05-07T06:48:26Z. Mar. 7, 2023. (Visited on 03/07/2023) (cited on page 90).
- [154] E. Borel. ‘Sur l’emploi du théorème de Bernoulli pour faciliter le calcul d’une infinité de coefficients. Application au problème de l’attente à un guichet’. In: *C. R. Acad. Sci.* 214 (1942), pp. 452–456. (Cited on page 91).
- [155] *SiPM – Silicon Photomultiplier PM1125-EB / PM1150-EB*. URL: <https://www.ketek.net/wp-content/uploads/2017/01/KETEK-PM1125-EB-PM1150-EB-Datasheet.pdf> (cited on page 102).
- [156] *MPPC (Multi-Pixel Photon Counter) S13360 series* (cited on page 102).
- [157] *SP5600E - Educational Photon Kit*. CAEN - Tools for Discovery. URL: <https://www.caen.it/products/sp5600e/> (visited on 03/07/2023) (cited on page 102).
- [158] Y. Du and F. Retière. ‘After-pulsing and cross-talk in multi-pixel photon counters’. In: *Nucl. Instrum. Meth. A* 596.3 (Nov. 2008), pp. 396–401. DOI: 10.1016/j.nima.2008.08.130. (Visited on 03/07/2023) (cited on page 111).
- [159] G. Kawata et al. ‘Probability Distribution of After-pulsing in Passive-quenched Single-photon Avalanche Diodes’. In: *IEEE Transactions on Nuclear Science* (2017), pp. 1–1. DOI: 10.1109/TNS.2017.2717463. (Visited on 03/07/2023) (cited on page 111).
- [160] M. Aleksa et al. ‘Calorimeters for the FCC-hh’. In: arXiv:1912.09962 (Dec. 20, 2019). DOI: 10.48550/arXiv.1912.09962. (Visited on 10/13/2022) (cited on page 116).
- [161] E. Buhmann. ‘Deep Learning based Energy Reconstruction for the CALICE AHCAL’. Master’s Thesis. University of Hamburg, July 2019 (cited on pages 116–118).
- [162] S. Richter, H. AlHaija, and V. Koltun. *Enhancing Photorealism Enhancement*. May 10, 2021. DOI: 10.48550/arXiv.2105.04619. URL: <http://arxiv.org/abs/2105.04619> (visited on 02/14/2023) (cited on page 120).
- [163] Y. Wang et al. ‘Dynamic Graph CNN for Learning on Point Clouds’. In: arXiv:1801.07829 (June 11, 2019). DOI: 10.48550/arXiv.1801.07829. (Visited on 01/09/2023) (cited on pages 121, 167, 171, 172, 211).
- [164] The CALICE collaboration. ‘Hadronic energy resolution of a combined high granularity scintillator calorimeter system’. In: *J. Inst.* 13.12 (Dec. 2018), P12022. DOI: 10.1088/1748-0221/13/12/P12022. (Visited on 11/25/2022) (cited on page 122).
- [165] W. Falcon and The PyTorch Lightning Team. *PyTorch Lightning*. Version 1.4. Mar. 2019. DOI: 10.5281/zenodo.3828935. (Visited on 11/25/2022) (cited on pages 139, 204).
- [166] T. Akiba et al. ‘Optuna: A Next-generation Hyperparameter Optimization Framework’. In: arXiv:1907.10902 (July 25, 2019). DOI: 10.48550/arXiv.1907.10902. (Visited on 11/25/2022) (cited on pages 139, 204).
- [167] Scikit-HEP Project. *scikit-hep/iminuit*. <https://github.com/scikit-hep/iminuit>. original-date: 2012-12-18T18:45:17Z. Nov. 22, 2022. (Visited on 11/25/2022) (cited on page 139).

- [168] P. Rousseeuw and C. Croux. 'Alternatives to the Median Absolute Deviation'. In: *J. Am. Stat. Assoc.* 88.424 (1993). Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 1273–1283. doi: [10.2307/2291267](https://doi.org/10.2307/2291267). (Visited on 03/30/2022) (cited on page 145).
- [169] G. Brys, M. Hubert, and A. Struyf. 'A Robust Measure of Skewness'. In: *J. Comput. Graph. Stat.* 13.4 (2004), pp. 996–1017. doi: [10.1198/106186004X12632](https://doi.org/10.1198/106186004X12632) (cited on page 145).
- [170] A. Worthing and J. Geffner. *Treatment of Experimental Data*. Google-Books-ID: F7jPAAAAMAAJ. J. Wiley & Sons, Incorporated, 1944. 360 pp. (cited on page 151).
- [171] V. Morgunov and A. Raspereza. 'Novel 3D Clustering Algorithm and Two Particle Separation with Tile HCAL'. In: arXiv:physics/0412108 (Dec. 17, 2004). doi: [10.48550/arXiv.physics/0412108](https://doi.org/10.48550/arXiv.physics/0412108). (Visited on 11/29/2022) (cited on pages 151, 153, 157, 158, 162).
- [172] N. Altman and M. Krzywinski. 'The curse(s) of dimensionality'. In: *Nature Methods* 15.6 (June 1, 2018). Number: 6 Publisher: Nature Publishing Group, pp. 399–400. doi: [10.1038/s41592-018-0019-x](https://doi.org/10.1038/s41592-018-0019-x). (Visited on 01/01/2023) (cited on page 165).
- [173] C. Qi et al. 'PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation'. In: arXiv:1612.00593 (Apr. 10, 2017). doi: [10.48550/arXiv.1612.00593](https://doi.org/10.48550/arXiv.1612.00593). (Visited on 01/09/2023) (cited on pages 167, 169, 170).
- [174] *Vanilla software*. In: *Wikipedia*. Page Version ID: 1102240600. Aug. 4, 2022. (Visited on 01/10/2023) (cited on page 167).
- [175] K. Hilden, E. Hutchinson, and L. Currie. *Iced: 180 Very Cool Concoctions*. Google-Books-ID: 02YG2aTL89MC. Allen & Unwin, 2006. 430 pp. (cited on page 167).
- [176] *Affine transformation*. In: *Wikipedia*. Page Version ID: 1125144549. Dec. 2, 2022. (Visited on 01/09/2023) (cited on pages 169, 170).
- [177] A. Kashefi and T. Mukerji. 'Physics-informed PointNet: A deep learning solver for steady-state incompressible flows and thermal fields on multiple sets of irregular geometries'. In: *J. Comput. Phys.* 468 (Nov. 2022), p. 111510. doi: [10.1016/j.jcp.2022.111510](https://doi.org/10.1016/j.jcp.2022.111510). (Visited on 01/09/2023) (cited on page 169).
- [178] X. Yan. *Pytorch Implementation of PointNet and PointNet++*. [https://github.com/yanx27/Pointnet\\_Pointnet2\\_pytorch](https://github.com/yanx27/Pointnet_Pointnet2_pytorch). original-date: 2019-03-04T14:24:30Z. Jan. 9, 2023. (Visited on 01/09/2023) (cited on page 170).
- [179] I. Armeni et al. 'Joint 2D-3D-Semantic Data for Indoor Scene Understanding'. In: arXiv:1702.01105 (Apr. 5, 2017). doi: [10.48550/arXiv.1702.01105](https://doi.org/10.48550/arXiv.1702.01105). (Visited on 01/31/2023) (cited on pages 170, 172).
- [180] *Utah teapot*. In: *Wikipedia*. Page Version ID: 1136234480. Jan. 29, 2023. (Visited on 01/30/2023) (cited on pages 171, 174).
- [181] H. Qu and L. Gouskos. 'ParticleNet: Jet Tagging via Particle Clouds'. In: *Phys. Rev. D* 101.5 (Mar. 26, 2020), p. 056019. doi: [10.1103/PhysRevD.101.056019](https://doi.org/10.1103/PhysRevD.101.056019). (Visited on 01/31/2023) (cited on page 172).
- [182] G. Kasieczka et al. 'Top Quark Tagging Reference Dataset'. In: (Mar. 22, 2019). Type: dataset. doi: [10.5281/zenodo.2603256](https://doi.org/10.5281/zenodo.2603256). (Visited on 01/31/2023) (cited on page 172).
- [183] Y. Wang. *Dynamic Graph CNN for Learning on Point Clouds*. <https://github.com/WangYueFt/dgcnn>. original-date: 2018-02-03T05:49:29Z. Jan. 7, 2023. (Visited on 01/09/2023) (cited on page 172).
- [184] K. He et al. 'Deep Residual Learning for Image Recognition'. In: arXiv:1512.03385 (Dec. 10, 2015). (Visited on 01/30/2023) (cited on page 172).
- [185] J. Kieseler. *caloGraphNN*. <https://github.com/jkiesele/calographnn>. original-date: 2019-01-23T10:12:06Z. Oct. 2, 2022. (Visited on 01/02/2023) (cited on pages 173, 175).
- [186] B. Schölkopf, A. Smola, and K-R. Müller. 'Kernel principal component analysis'. In: *Artificial Neural Networks — ICANN'97*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 1997, pp. 583–588. doi: [10.1007/BFb0020217](https://doi.org/10.1007/BFb0020217) (cited on page 173).
- [187] S. Nikolenko. 'Synthetic Data for Deep Learning'. In: arXiv:1909.11512 (Sept. 25, 2019). doi: [10.48550/arXiv.1909.11512](https://doi.org/10.48550/arXiv.1909.11512). (Visited on 12/06/2022) (cited on page 177).

- [188] *pylandau*. <https://github.com/SiLab-Bonn/pylandau>. original-date: 2015-04-28T15:27:19Z. May 25, 2022. (Visited on 12/08/2022) (cited on page 179).
- [189] G. Ke et al. 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree'. In: *Adv Neural Inf Process Syst*. Vol. 30. Curran Associates, Inc., 2017. (Visited on 12/09/2022) (cited on page 183).
- [190] *lightgbm.LGBMClassifier* — *LightGBM 3.3.3.99 documentation*. URL: <https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html> (visited on 12/10/2022) (cited on page 183).
- [191] J. Cohen and D. Courgeau. 'Modeling distances between humans using Taylor's law and geometric probability'. In: *Math. Popul. Stud.* 24 (May 23, 2017), pp. 197–218. DOI: [10.1080/08898480.2017.1289049](https://doi.org/10.1080/08898480.2017.1289049) (cited on page 190).
- [192] B. Burgstaller and F. Pillichshammer. 'The Average Distances Between Two Points'. In: *Bull. Aust. Math. Soc.* 80.3 (Dec. 2009), pp. 353–359. DOI: [10.1017/S0004972709000707](https://doi.org/10.1017/S0004972709000707). (Visited on 07/23/2020) (cited on page 190).
- [193] T. Odland. *KDEPy*. <https://github.com/tommyod/KDEPy>. original-date: 2018-01-19T15:27:57Z. Feb. 15, 2023. (Visited on 02/17/2023) (cited on page 206).
- [194] C. Domínguez. *Effect of Hyperparameters and Kernels on SVMs*. URL: <https://kaggle.com/code/carlosdg/effect-of-hyperparameters-and-kernels-on-svms> (visited on 03/03/2023) (cited on page 211).
- [195] D. Napoletano and G. Soyez. 'Computing  $N$ -subjettiness for boosted jets'. In: *J. High Energ. Phys.* 2018.12 (Dec. 2018), p. 31. DOI: [10.1007/JHEP12\(2018\)031](https://doi.org/10.1007/JHEP12(2018)031). (Visited on 02/20/2023) (cited on page 212).
- [196] V. Bocharnikov. 'Hadronic shower substructure reconstruction with graph neural networks'. 14th Annual Meeting of the Helmholtz Alliance "Physics at the Terascale". Nov. 23, 2021. (Visited on 02/20/2023) (cited on page 212).
- [197] J. Kieseler. 'Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data'. In: *Eur. Phys. J. C* 80.9 (Sept. 2020), p. 886. DOI: [10.1140/epjc/s10052-020-08461-2](https://doi.org/10.1140/epjc/s10052-020-08461-2). (Visited on 05/24/2023) (cited on page 258).
- [198] S. Qasim et al. 'End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks'. In: *Eur. Phys. J. C* 82.8 (Aug. 29, 2022), p. 753. DOI: [10.1140/epjc/s10052-022-10665-7](https://doi.org/10.1140/epjc/s10052-022-10665-7). (Visited on 05/24/2023) (cited on page 258).
- [199] M. Landreman. 'An improved current potential method for fast computation of stellarator coil shapes'. In: *Nucl. Fusion* 57.4 (Feb. 2017). Publisher: IOP Publishing, p. 046003. DOI: [10.1088/1741-4326/aa57d4](https://doi.org/10.1088/1741-4326/aa57d4). (Visited on 05/26/2023) (cited on page 259).







