Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models

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

vorgelegt von
Peter Jeffrey McKeown

Hamburg
2024

Gutachter/innen der Dissertation:	Prof. Dr. Gregor Kasieczka Dr. Frank Gaede
Zusammensetzung der Prüfungskommission:	Prof. Dr. Daniela Pfannkuche Prof. Dr. Gregor Kasieczka Dr. Frank Gaede Prof. Dr. Marcus Brüggen Dr. Ties Behnke
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Daniela Pfannkuche
Datum der Disputation:	08.04.2024
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Markus Drescher
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

For my family

Abstract

Modern high energy physics experiments fundamentally rely on large quantities of simulated data, placing significant demands on the available computational resources. Machine learning methods based on deep generative models promise to reduce the compute time required to simulate particle showers in the calorimeter system, which constitutes the most computationally intensive part of a full detector simulation.

This work focuses on the development of a first simulation tool based on deep generative models for shower simulation in highly granular calorimeters, and subsequently studies its performance in a realistic detector geometry. In order to apply these models in a general simulation, they must provide a suitable detector response for particles incident under various angles to, and at various positions in, the detector. Crucially, the physics performance after reconstruction must remain high, which is the ultimate target of such a simulator.

We initially extend the performant Bounded Information Bottleneck Autoencoder (BIB-AE) to simulate showers from photons with varying incident energy and angle to the surface of the electromagnetic calorimeter of the International Large Detector (ILD), before studying the single particle performance of the model in terms of key calorimetric observables, both before and after reconstruction. We then further extend the model to handle an additional angle of incidence, as well as taking steps to deal with geometry irregularities in order to allow the use of the model at different positions in the calorimeter.

As a next step, we describe a generic library that enables the use of generative models with GEANT4 and DD4HEP, allowing a full integration into standard software ecosystems used in high energy physics. We outline the integration of the BIB-AE into this library, allowing a fair benchmark of the computational performance of the model. We then simulate showers at different positions with the model, in order to investigate the effects of performing simulations in an irregular calorimeter geometry.

Finally, we study the performance of the BIB-AE when used to simulate photons from neutral pion decays in the process $e^+e^- \to \tau^+\tau^-$ in terms of key physics observables. We find that while some deviations from GEANT4 occur, they are typically comparable to the Monte Carlo uncertainty, estimated from the performance differences between GEANT4 versions.

Zusammenfassung

Für moderne Experimente der Hochenergiehysik sind große Mengen simulierter Daten ein wesentlicher Bestandteil, was erhebliche Anforderungen an die verfügbaren Rechenressourcen stellt. Methoden des maschinellen Lernens, basierend auf generativen Modellen, haben das Potential die Rechenzeit zu reduzieren, welche für die Simulation von Teilchenschauern im Kalorimetersystem erforderlich ist, die den rechenintensivsten Teil einer vollständigen Detektorsimulation darstellt.

Diese Arbeit fokussiert sich auf die Entwicklung einer ersten Simulationsanwendung, die auf generativen Modellen für die Schauersimulation in hochgranularen Kalorimetern basiert, und untersucht anschließend dessen Leistung in einer realistischen Detektorgeometrie. Um diese Modelle in einer allgemeinen Simulation anwenden zu können, müssen sie eine geeignete Detektorantwort für Teilchen liefern, die unter verschiedenen Winkeln zum und an verschiedenen Positionen im Detektor einfallen. Entscheidend ist, dass die Güte physikalischer Observablen nach der Rekonstruktion hoch bleibt, was das eigentliche Ziel eines solchen Simulators ist.

Wir erweitern zunächst den leistungsstarken Bounded Information Bottleneck Autoencoder (BIB-AE), um Schauer von Photonen unterschiedlicher Einfallsenergien und -winkel zur Oberfläche des elektromagnetischen Kalorimeters des International Large Detector (ILD) zu simulieren, bevor wir die Einzelteilchensimulationsgüte des Modells im Hinblick auf entscheidende kalorimetrische Messgrößen sowohl vor als auch nach der Rekonstruktion untersuchen. Anschließend wird das Modell so erweitert, dass es einen zusätzlichen Einfallswinkel verarbeiten kann, und es werden Schritte unternommen, um geometrische Unregelmäßigkeiten zu berücksichtigen, damit das Modell an verschiedenen Positionen im Kalorimeter eingesetzt werden kann.

Als nächsten Schritt beschreiben wir eine generische Programmbibliothek, die die Verwendung von generativen Modellen mit Geant4 und DD4hep ermöglicht und eine vollständige Integration in die üblichen in der Hochenergiephysik verwendeten Softwareumgebungen erlaubt. Wir beschreiben die Integration des BIB-AE in diese Bibliothek, was einen angemessen Benchmark der Rechenleistung des Modells ermöglicht. Anschließend simulieren wir mit dem Modell Schauer an unterschiedlichen Positionen, um die Effekte der Durchführung solcher Simulationen in einer irregulären Kalorimetergeometrie zu untersuchen.

Schließlich untersuchen wir die Leistungsfähigkeit des BIB-AE, wenn es in Simulationen von Photonen aus Zerfällen neutraler Pionen im Prozess $e^+e^- \to \tau^+\tau^-$ in Bezug auf die wichtigsten physikalischen Observablen verwendet wird. Wir stellen fest, dass gewisse Abweichungen von Geant4 auftreten, diese aber typischerweise mit der Monte-Carlo-Unsicherheit vergleichbar sind, die aus den Ergebnisunterschieden zwischen den Geant4-Versionen geschätzt wird.

Contents

1 Introduction **3**

2 Particle Physics **7**
- 2.1 The Standard Model of Particle Physics 7
- 2.2 Physics Beyond the Standard Model 8
- 2.3 Future High Energy e^+e^- Colliders 9
  - 2.3.1 Collider Experiments 9
  - 2.3.2 Measurements at a Future High Energy e^+e^- Collider 11
  - 2.3.3 The International Linear Collider 12
- 2.4 Physics of the Tau Lepton 13

3 Calorimetry **15**
- 3.1 Electromagnetic Showers 15
  - 3.1.1 Electron and Positron Interactions with Matter 15
  - 3.1.2 Photon Interactions with Matter 16
  - 3.1.3 Electromagnetic Shower Development 17
- 3.2 Heavy Charged Particle Interactions with Matter 18
- 3.3 Hadronic Showers 19
- 3.4 Calorimeter Designs and Measurements 21

4 Particle Flow Reconstruction and the International Large Detector **23**
- 4.1 Particle Flow Reconstruction 23
- 4.2 The International Large Detector Concept 25
  - 4.2.1 Detector Systems 25
  - 4.2.2 Software Ecosystem 28
  - 4.2.3 Reconstruction Scheme 28

5 Machine Learning **31**
- 5.1 Neural Networks and Deep Learning 31
- 5.2 Network Architecture Components - Layers and Activation Functions 32
  - 5.2.1 Fully Connected Layers 32
  - 5.2.2 Convolutional Layers 33
  - 5.2.3 Transpose Convolutional Layers 34
  - 5.2.4 Normalisation Layers 34
  - 5.2.5 Non-linearities 35
- 5.3 Learning and Optimisation Algorithms 35
  - 5.3.1 Backpropagation 35
  - 5.3.2 Gradient Descent and Stochastic Gradient Descent 36
  - 5.3.3 Momentum and Adaptive Learning Rates 37
  - 5.3.4 ADAM: Adaptive Moment Estimation 38
- 5.4 Generative Models 38
  - 5.4.1 Variational Autoencoders 38

5.4.2 Generative Adversarial Networks 40
5.4.3 The Bounded Information Bottleneck Autoencoder 41
5.4.4 Normalising Flows 42
5.4.5 Conditioning of Generative Models 44

6 Simulation in High Energy Physics 47
6.1 Particle Transport Monte Carlo 48
6.1.1 The GEANT4 Simulation Toolkit 49
6.1.2 The DD4HEP Detector Description Toolkit 51
6.2 Classical Approaches to Fast Simulation 51
6.3 Generative Models for Fast Calorimeter Simulation 52

7 Multi-parameter Conditioning of Generative Models 55
7.1 Coordinate System Convention 55
7.2 Dataset and Reconstruction Scheme 56
7.3 Generative Model 58
7.3.1 BIB-AE Model 58
7.3.2 Latent Space Sampling 62
7.4 Results 63
7.4.1 Physics Performance 63
7.4.2 Computational Performance 75
7.5 Conclusions 75

8 A Fully Conditioned Generative Model for Electromagnetic Shower Simulation 77
8.1 Dataset 77
8.1.1 Creation of a Regular ECAL Geometry 77
8.1.2 Creation of the Dataset 80
8.2 Extensions to the Generative Model 82
8.3 Results 82
8.4 Conclusions 88

9 Integration into the International Large Detector Software Chain 91
9.1 A Library for Fast Calorimeter Simulation with Generative Models in GEANT4 and DD4HEP 91
9.2 BIB-AE Integration 95
9.2.1 Running the BIB-AE Architecture in C++ 95
9.2.2 Implementation of the Model and Geometry Interfaces 96
9.2.3 Geometry Exclusions 96
9.3 Computational Performance of the Library 97
9.4 Conclusions 99

10 Irregular Geometries 101
10.1 Investigations into the Effects of Simulation with a Regular Grid Model in an Irregular Geometry 101
10.2 Results 104
10.3 Conclusions 110

11 Physics Benchmark 113
11.1 Photons from Neutral Pions Produced in the Decay of Tau Pairs 113
11.2 Dataset Creation 113
11.3 Reconstruction Performance Comparison 117
11.4 Conclusions 123

12 Conclusions **125**

A Supplementary Material: Multi-parameter Conditioning of Generative Models 129
- A.1 Supplementary Material: Neural Network Architectures 129
- A.2 Supplementary Material: Results 134
  - A.2.1 Cell Energy Spectrum 134
  - A.2.2 Center of Gravity 137
  - A.2.3 Longitudinal Profile 138
  - A.2.4 Radial Profile 140

B Supplementary Material: A Fully Conditioned Generative Model for Electromagnetic Shower Simulation **143**
  - B.0.1 Angular Response- Theta 143
  - B.0.2 Angular Response- Phi 144
  - B.0.3 Energy Response 145
  - B.0.4 Cell Energy Spectrum 146
  - B.0.5 Number of Hits 147
  - B.0.6 Center of Gravity 148
  - B.0.7 Longitudinal Profile 149

C Supplementary Material: Irregular Geometries **151**

D Supplementary Material: Physics Benchmark **153**

Preface

The results presented in this thesis were collected at Deutsches Elektronen-Synchrotron DESY between 2020 and 2024 in collaboration with other researchers. The results presented in Chapter 7 have previously been published as:

[1] S. Diefenbacher et al. “New angles on fast calorimeter shower simulation”. Mach. Learn. Sci. Tech. 4.3 (2023), p. 035044 `10.1088/2632-2153/acefa9`

In addition, parts of the results presented in Chapter 9 have previously been included in a contribution to a conference proceedings as:

[2] P. McKeown et al. "Fast Simulation of Highly Granular Calorimeters with Generative Models: Towards a First Physics Application". PoS EPS-HEP2023 (2023), p. 568. `10.22323/1.449.0568`

The author has also been involved in the following publications directly related to the topic of this thesis during this research period:

[3] E. Buhmann et al. “Fast and Accurate Electromagnetic and Hadronic Showers from Generative Models”. EPJ Web Conf. 251 (2021), p. 03049. `10.1051/epjconf/202125103049`

[4] E. Buhmann et al. "Hadrons, better, faster, stronger". Mach. Learn. Sci. Tech. 3.2 (2022), p. 025014. `10.1088/2632-2153/ac7848`

[5] S. Bieringer et al. “Generative Models for Fast Simulation of Electromagnetic and Hadronic Showers in Highly Granular Calorimeters”. PoS ICHEP2022 (2022), `10.22323/1.414.0236`

[6] E. Buhmann et al. "CaloClouds: fast geometry-independent highly-granular calorimeter simulation". JINST 18.11 (2023), P11025. `10.1088/1748-0221/18/11/P11025`

[7] E. Buhmann et al. "CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation" (2023). arXiv: `2309.05704[physics.ins-det]`.

Chapter 1

Introduction

Decades of experimental and theoretical advances have gradually illuminated the fundamental forces and particles which underpin our universe, culminating in the Standard Model of Particle Physics (SM) [8–10]. This model represents a triumph of modern physics, and has withstood rigorous experimental testing. However, evidence of phenomena including the presence of dark matter [11–14], neutrino oscillations [15, 16] and the asymmetry between matter and anti-matter [17] indicate that the SM is not a complete description of Nature. Despite this, experiments including those operated at the Large Hadron Collider (LHC) [18] have thus far been unable to find new physics beyond the standard model.

For this reason, various future collider projects are planned or under consideration. This includes upgrades to the existing experiments for the High-Luminosity phase LHC (HL-LHC) [19], and proposed experiments at future high energy e^+e^- colliders [20–23]. A core component of the physics programmes of these experiments is making precision measurements of observables. Looking for small deviations in these measurements from the predictions made by the SM provides a means to search for new physics, and requires significant amounts of data to be collected.

In order for a comparison to be drawn between theoretical predictions and experimental observations, high energy physics experiments crucially rely on detailed and realistic *simulations*, which incorporate our best understanding of the complete physics involved. This procedure includes a *generator* which simulates the hard interaction in the initial collision and the hadronisation of any final state partons produced. Subsequently, *detector simulation* models the interactions of particles which fly through, and interact with, the detector systems. This work will focus on detector simulation.

Traditionally, particle physics experiments have relied on Monte Carlo (MC) methods to design physics-based simulation tools for use with realistic detector geometries. This approach underpins the foremost simulation toolkit used throughout particle physics, GEANT4 [25–27]. This *full* simulation approach relies on individually tracking particles through the volumes of the detector geometry in order to simulate interactions between a particle and the materials and fields present in the detector.

Full simulation is therefore resource intensive. To illustrate this, Figure 1.1 shows the fraction of the total CPU usage by the ATLAS experiment spent on various tasks in 2018. Around 40% of the total compute time was spent on detector simulation [24]. This is a major challenge for precision measurements, which are facilitated by reducing statistical uncertainties. As a result, the amount of simulated data required can be an order of magnitude larger than the amount of experimental data collected. In particular, full simulation creates a major bottleneck when used for the *calorimeter* subsystems of the detector, which rely on creating a spray of secondary particles in a *shower* in order to make a destructive measurement of the energy of a primary particle.

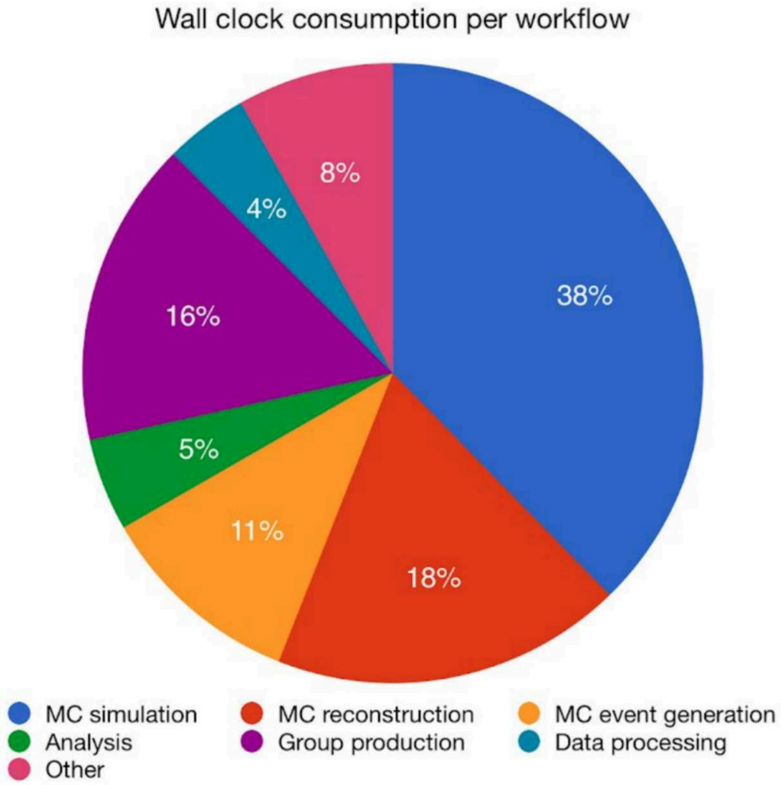


Figure 1.1: Fraction of the total CPU time spent on various activities for the ATLAS experiment in 2018. Figure from [24].

For this reason, *fast* simulation approaches have long been explored as a means of trading off some physics fidelity for simulation speed. Examples adopted for calorimeter simulation include parameterisations [31, 32] and libraries of pre-simulated showers [33, 34]. However, even with these fast simulation approaches being available, both the ATLAS and CMS experiments predict that required computing resources will be exceeded without the necessary R&D activities being conducted [29, 30]. The corresponding projections are shown in Figure 1.2.

With advances in compute accelerator hardware such as Graphics Processing Units (GPUs), coupled with developments in optimisation algorithms and model architectures, approaches based on machine learning (ML) techniques hold promise to provide fast and accurate simulation tools. In particular, *generative models* which aim to learn the distribution underlying a given set of data points, and thereby permit subsequent sampling of new data points, have been explored. Studies have investigated fast simulation applications of generative models at various levels in the simulation chain, including event generation [35–40] and modelling hadronization [41–43]. Within the detector component of simulation, the simulation of calorimeter showers with generative models has been the focus of particularly intensive investigation [4, 6, 7, 44–63].

One major challenge is that calorimeters for future experiments are planned with higher granularities in order to meet the requirements of the physics programmes. This demands the development of precise approaches to fast simulation, for which proof-of-concept studies have indicated generative models are particularly promising options [4, 45, 56, 63]. One model in particular, the *Bounded Information Bottleneck Autoencoder* (BIB-AE), was found to be especially successful in this task [4, 45, 63]. Thus far however, these studies have largely been restricted to simulating particles incident at a fixed angle to the face of the calorimeter, and at a fixed position. In order for a fast simulation tool based on generative models to be useful

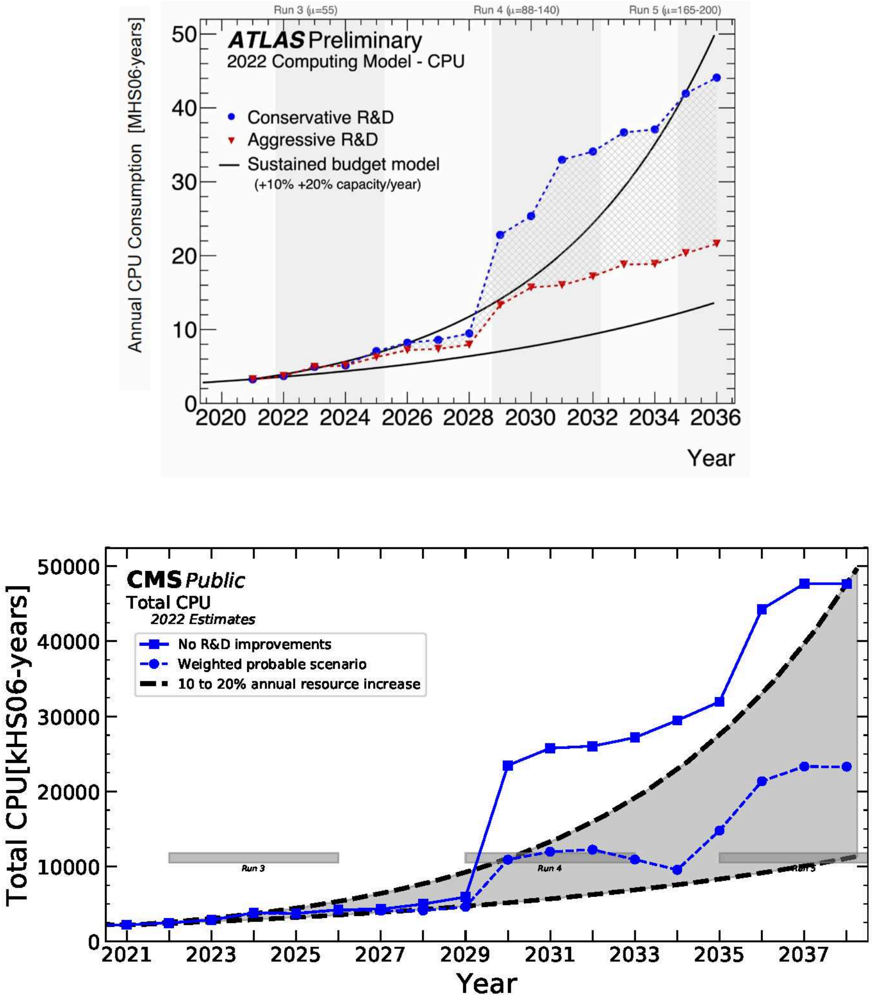


Figure 1.2: CPU consumption projections for the ATLAS (top) and CMS (bottom) experiments into the HL-LHC phase. Various R&D scenarios are shown, with the computing resource usage measured with the HEP-SPEC06 (HS06) benchmarking suite [28]. Figures adapted from [29] (top) and [30] (bottom).

for simulating showers in a full calorimeter subsystem, it must be able to provide a correct detector response for particles incident at different angles and positions. Importantly, due to the irregular nature of real detector geometries, simulating showers at different positions requires the effects of irregular geometries to be studied and the development of strategies to minimise these effects. Additionally, while it has become standard to benchmark the physics performance of a fast simulation tool on its direct *simulation* level output, downstream analysis makes use of this information after applying the full *reconstruction* chain of a given experiment. This means that the ultimate validation of a generative model based fast simulation tool must be dictated by the physics performance after reconstruction. Crucially, in order to interface with the existing reconstruction pipelines, a generative model must be integrated into the software ecosystem used by an experiment. If a seamless integration with the full simulation and reconstruction software can be achieved, benchmarking of the model can be performed on full physics processes, as opposed to single particle events.

This thesis will develop these proof-of-concept studies through to their deployment and benchmarking with production ready software. Given the strong performance of the BIB-AE model observed in previous work, this generative model will be expanded upon in this thesis. Studies will be performed using the International Large Detector (ILD) [64], a detector proposed for use at the International Linear Collider (ILC) [20] which features highly granular calorimeters.

This thesis is structured as follows. We begin by describing the theoretical background and methodological details relevant to this work. Chapter 2 provides a brief overview of the SM together with the motivations for, and fundamentals of, experiments at future high energy e^+e^- colliders. The physics of the tau lepton is also reviewed. In Chapter 3 the physics underpinning the development of calorimeter showers is described, together with the principles behind calorimeter designs and measurements. Chapter 4 introduces the particle flow reconstruction paradigm, and provides an overview of the detector systems, software ecosystem and reconstruction scheme employed for ILD. In Chapter 5, the ML techniques adopted in this thesis will be introduced, including the components used to build the network architectures, the optimisation algorithms and the generative models. Chapter 6 addresses simulation in high energy physics, with a focus on detector simulation. Particle transport Monte Carlo, which forms the basis of full simulation is introduced, before descriptions of the GEANT4 simulation toolkit and the DD4HEP detector description toolkit are provided. An overview of classical approaches to fast simulation is given, followed by a review of the current state of the field of fast calorimeter simulation using generative models.

We then turn to the development of a fast simulation tool for showers in high granularity calorimeters using generative models. In Chapter 7, the BIB-AE model will be extended by means of *conditioning* to provide an appropriate detector response for a varying incident angle to the calorimeter surface. Additionally, the single-particle reconstruction performance of the model after calorimeter clustering with a state-of-the-art particle flow reconstruction algorithm, will be investigated. In Chapter 8, the BIB-AE model will be extended to accept conditioning on an additional incident angle, and thus be fully conditioned for a general simulation of electromagnetic showers. In Chapter 9, the development of a software library which allows the integration of generative models into GEANT4 via DD4HEP will be described. Additionally, the steps taken to integrate the BIB-AE model for electromagnetic shower simulation in the barrel and endcap regions of the ILD electromagnetic calorimeter will be detailed. In Chapter 10, the implications of using the BIB-AE model to simulate showers in an irregular calorimeter geometry will be investigated, and approaches to handling the irregular geometry explored. Finally, in Chapter 11, the performance of the BIB-AE will be benchmarked in comparison to GEANT4 for the simulation of photons from π^0s produced during hadronic tau decays in the process $e^+e^- \to \tau^+\tau^-$.

Unless otherwise stated, natural units of $\hbar = c = 1$ will be used throughout this thesis.

Chapter 2

Particle Physics

Particle physics seeks to unravel the structure of our universe at the most fundamental levels. In Section 2.1 An overview will be provided of our best current understanding of elementary particles and their interactions, which is formulated in the the Standard Model of Particle Physics. While this model has withstood rigorous experimental scrutiny, it is known to not be a complete theory of Nature. Some of the motivations to look for physics beyond the Standard Model will be addressed in Section 2.2. A promising means of searching for this new physics is to build a high energy e^+e^- collider. Section 2.3.1 will outline the advantages provided by such a machine, while Section 2.3.2 will summarise some of the key measurements that could be made. In Section 2.3.3, one option for such a machine, the International Linear Collider will be described. Section 2.4 will focus on the physics of the tau lepton, which displays a rich variety of behaviours and provides experimental challenges, as well as being a key tool for physics at a future e^+e^- collider.

2.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) [8–10] describes a broad range of phenomena and has provided numerous predictions that have been precisely tested and validated by experiments. It therefore represents our best current understanding of fundamental particles and forces, and constitutes one of the most successful physical theories ever created.

The particle content of the SM is as follows, with an overview shown in Figure 2.1. It is composed of twelve spin-$\frac{1}{2}$ fermions, each with its own corresponding anti-particle with the same mass but opposite quantum numbers. Interactions can occur by means of the electromagnetic, strong and weak forces, which are mediated by four spin-1 gauge bosons. Finally, the existence of a scalar Higgs boson is necessary to allow the vector gauge bosons to acquire a mass.

The SM is a quantum field theory (QFT) which is renormalisable and provides a unified description of the three forces by means of an $\mathsf{SU}(3)_C \otimes \mathsf{SU}(2)_L \otimes \mathsf{U}(1)_Y$ local gauge symmetry. The $\mathsf{SU}(2)_L \otimes \mathsf{U}(1)_Y$ component describes interactions in the electroweak sector, which are mediated by the photon and the Z^0 and $W^\pm$ bosons. The Higgs field provides the mechanism by which *spontaneous symmetry breaking* (SSB) occurs, resulting in the breaking of electroweak symmetry to the electromagnetic subgroup [66],

$$\mathsf{SU}(2)_L \otimes \mathsf{U}(1)_Y \to \mathsf{U}(1)_{\mathrm{QED}}. \tag{2.1}$$

While the photon remains massless, this process imbues the Z^0 and $W^\pm$ bosons with a mass, as well as generating the appropriate fermion masses and mixings. It additionally gives rise to the Higgs boson [67–69].

The $\mathsf{SU}(3)_C$ component is the symmetry group which underlies the description of the strong force provided by the non-abelian gauge theory Quantum Chromodynamics (QCD)

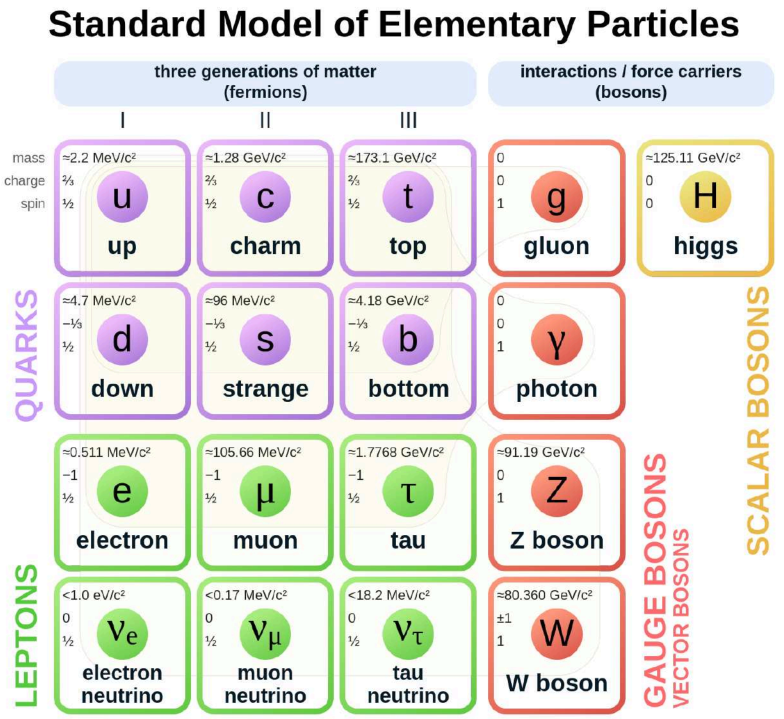

Standard Model of Elementary Particles
three generations of matter (fermions)
interactions / force carriers (bosons)
I
II
III
mass
charge
spin
≈2.2 MeV/c²
⅔
½
u
up
≈1.28 GeV/c²
⅔
½
c
charm
≈173.1 GeV/c²
⅔
½
t
top
0
0
1
g
gluon
≈125.11 GeV/c²
0
0
H
higgs
≈4.7 MeV/c²
−⅓
½
d
down
≈96 MeV/c²
−⅓
½
s
strange
≈4.18 GeV/c²
−⅓
½
b
bottom
0
0
1
γ
photon
≈0.511 MeV/c²
−1
½
e
electron
≈105.66 MeV/c²
−1
½
μ
muon
≈1.7768 GeV/c²
−1
½
τ
tau
≈91.19 GeV/c²
0
1
Z
Z boson
<1.0 eV/c²
0
½
νe
electron neutrino
<0.17 MeV/c²
0
½
νμ
muon neutrino
<18.2 MeV/c²
0
½
ντ
tau neutrino
≈80.360 GeV/c²
±1
1
W
W boson
QUARKS
LEPTONS
GAUGE BOSONS
VECTOR BOSONS
SCALAR BOSONS

Figure 2.1: Particles in the Standard Model, including their mass spin and charge. Figure from [65]

[70]. This theory describes the interactions involving quarks and gluons, as a result of them carrying colour charge. The gluons are the mediators of the strong interaction, and while they are electromagnetically neutral and massless, couple to colour charge. As a result, they are self-interacting and this gives rise to the phenomenon of *colour confinement*, meaning that no colour charged particle can exist in isolation, and must always be bound into a colour neutral state. A key consequence of this is that final state quarks and gluons produced in high energy collisions undergo *hadronisation*. This results in a collimated spray of colour-neutral hadrons called a *jet* [71].

2.2 Physics Beyond the Standard Model

While the SM has been incredibly successful in providing a theoretical description of a wide range of phenomena, it is known to be an incomplete description of nature. The most clearly lacking aspect is a description of gravity, which is considered to have a negligible strength relative to the other forces at the scales probed in high energy physics experiments.

Another example is the lack of an explanation for dark matter, which dominates the matter content of the universe. Its existence is at this point well established by numerous different astrophysical and cosmological observations. Examples include galactic rotation curves, which exhibit the signatures of unexplained excess matter outside of the galactic core [11], results from gravitational lensing [12] and the anisotropic structure of the Cosmic Microwave Background [13, 14].

A further shortcoming is that in the SM there are no right-handed neutrinos, prohibiting a coupling to the Higgs and meaning that the neutrino is massless. However, observations of neutrino oscillations [15, 16], in which neutrinos change flavour, imply that neutrino mass eigenstates exist. Hence experiments indicate that neutrinos should have a small but non-zero mass.

A final example of phenomena which the SM cannot explain is the matter-antimatter asymmetry observed in the universe, which is mostly composed of baryonic matter. No mechanism to cause such an asymmetry exists in the SM [17].

While this is by no means an exhaustive list of the problems with the SM, it serves to motivate searches for new physics beyond the standard model (BSM). These may be direct searches, which seek to directly produce and measure new particles that would be a source of new physics, or indirect searches, where precision measurements are made to look for deviations in the predictions or parameters of the SM.

2.3 Future High Energy e^+e^- Colliders

With the discovery of the Higgs boson in 2012 [72, 73], the final particle in the SM was observed. Given the shortcomings of the SM outlined in Sections 2.2, the Higgs also provides an excellent means by which to probe the SM and search for new physics. Searching for new phenomena requires both an appropriate physics environment and particle detectors suited to that environment.

2.3.1 Collider Experiments

Colliding beam experiments rely on particle accelerators to accelerate *bunches* of many particles, and bring them into collision, resulting in individual interactions called *events*. This allows such experiments to reach higher center-of-mass $\sqrt{s}$ energies than experiments in which a beam of particles is fired into a fixed target, as the full energy of the two beams is available. This gives a better scaling with the energy of the colliding beams of $\sqrt{s} = 2E_{\text{beam}}$, for two beams with equal energies E_{beam}. Since having more energy available in a collision allows the production of heavier, potentially new particles and the probing of structures on smaller scales, machines operating on the energy frontier adopt this approach.

Another important property of an accelerator is the rate at which events occur. The expected number of events for a given process per unit time is given by

$$N_{\text{proc}} = L \cdot \sigma_{\text{proc}}, \tag{2.2}$$

where σ_{proc} is the cross section of the process and L is the instantaneous luminosity provided by the accelerator, an important parameter of the machine. The total luminosity delivered over the life-cycle of an accelerator, known as the integrated luminosity, is given by

$$\mathcal{L}_{\text{int}} = \int L(t)dt. \tag{2.3}$$

Differences between Linear and Circular Colliders

Linear and circular accelerator layouts are the two key designs for particle colliders. A circular collider in the form of a synchrotron has the advantage that it is possible to circulate particles multiple times through the accelerator. This makes it possible to achieve a higher rate of collision, and thereby luminosity, as particles which do not interact in a given bunch crossing can be re-circulated in the machine. This is in general[1] not possible at a linear collider. Circular machines, however, suffer from synchrotron radiation, which arises from a charged particle following a curved trajectory in a magnetic field. The energy loss scales as

$$\Delta E \propto \frac{E^4}{m^4 R}, \tag{2.4}$$

---

[1] R&D efforts are attempting to mitigate this form of inefficiency [74]

where E is the energy of the particle, m is its mass and R is the bending radius. This means that significant energy loss occurs when trying to accelerate lighter particles, and limits the maximum energy achievable with a circular e^+e^- collider.

Motivations for a High Energy Lepton Collider

Currently the highest energy collider in operation is the Large Hadron Collider (LHC) at CERN [18]. It is a circular machine with a circumference of 26.7 km, a center-of-mass energy of up to $\sqrt{s} = 14$ TeV for proton-proton collisions and a design luminosity of $L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$. An upgrade to the LHC, the High-Luminosity LHC (HL-LHC), plans to increase the luminosity by approximately a factor of 5 [19].

A key advantage of hadron colliders such as the LHC is that they are able to reach high center-of-mass energies, as a result of the high mass of the particles that they collide. However, the physics environment at a hadron collider presents a number of challenges.

At a hadron collider, the total cross section is dominated by soft QCD processes, which cannot be described perturbatively and make precision measurements challenging. These QCD processes create many soft jets, resulting in a high detector occupancy and the production of huge backgrounds. Additionally, the high event rates necessitate the use of complex trigger systems, potentially introducing biases into the data collected and possibly missing new physics. Conversely, at a lepton collider electroweak interactions govern the production, resulting in much lower backgrounds and cleaner detector environments. In addition, the lower event rates offer the potential for triggerless operation, depending on the machine.

Finally, hadron colliders do not collide fundamental particles. Instead, hadron collisions involve not just the valence quarks, but also a sea of quarks and gluons. This means that the partons involved in the initial hard scattering carry an unknown fraction of the proton momentum, which leads to an unknown partonic centre-of-mass energy. As the initial collision energy is not known, a probabilistic description through the use of parton distribution functions (PDFs) [75] is necessary. In contrast, lepton collisions involve fundamental particles, and as such the initial state of the colliding particles is well known. This enables direct control of properties of the beam, which can be used to constraint the outgoing particle kinematics.

These factors, combined with the fact that Higgs physics has never been studied at a lepton collider, motivate a future high energy lepton collider. This makes a *Higgs factory*, a lepton collider which would produce Higgs bosons with a high rate, a priority machine in high energy physics. While numerous proposals of varying levels of maturity exist for e^+e^- Higgs factories, with both linear [21, 76–78] and circular [22, 23, 79] footprints, this work will focus on the International Linear Collider (ILC) [20, 76, 80–82], which will be described in more detail in Section 2.3.3.

Particle Detectors

Collider experiments rely on complex detector systems to identify and perform measurements of the particles produced in collision events. Modern general purpose detectors are composed of multiple sub-detector systems arranged in separate layers around the interaction point (IP). A cylindrical or polyhedral barrel region is placed around the IP with its length parallel to the beam axis, and with endcaps closing its ends in order to provide coverage down to the beam pipe. The detector is immersed in a strong axial magnetic field $\vec{B}$, which causes charged particles to follow a helical trajectory due to the Lorentz force.

The innermost subdetector is the *tracking* system, devoted to measuring charged particles while introducing as little material into the detector as possible. Around the tracking volume sits the *calorimeter* system split in an electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL), which aim to stop particles and make a destructive measurement of their energy by creating a *shower* of secondary particles. This sub-system will be the primary focus of this thesis, and will be described more in Chapter 3. Outside of these regions a dedicated system is placed for making *muon* measurements, as they typically escape the previous detector elements.

2.3.2 Measurements at a Future High Energy e^+e^- Collider

A high energy e^+e^- collider would provide an excellent means by which to perform precision measurements, as well as direct searches. In the following, we provide an overview of some of the salient measurements that can be made in the Higgs, electroweak and top sectors.

Higgs Physics

Many open questions remain about the nature of the Higgs boson, which a future e^+e^- collider would be well suited to address. Examples include whether the Higgs is truly a fundamental scalar or whether it is a composite particle, whether the Higgs is self-interacting, and whether it connects to a dark sector. Since deviations of new physics from the standard model are expected to be at most of the order of $5-10\%$, the desired precision of 1% or less makes an e^+e^- collider an ideal physics environment [83].

The golden channel at a Higgs factory is the so called Higgs-strahlung process $e^+e^- \to ZH$, which is the dominant Higgs production mode at $\sqrt{s} = 250$ GeV. By reconstructing the decay of the Z boson, which can be done very precisely using the leptonic decays $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$, the Higgs mass M_H can be computed directly from the recoil mass via

$$M^2_{recoil} = M^2_H = s - 2E_Z\sqrt{s} + M^2_Z, \tag{2.5}$$

where $\sqrt{s}$ is the center-of-mass energy, E_Z is the energy of the Z and M_Z is the reconstructed mass [84]. Since only measurements of the Z boson decay are made, this provides a very precise and model-independent means by which to study the properties of the Higgs, including its mass, decay width and couplings [85]. In particular, it allows a precise determination of the Higgs to invisible branching fraction, making it a powerful means by which to conduct dark matter searches.

Another opportunity provided by a high energy, and therefore linear, e^+e^- collider is the potential to discover and subsequently study the Higgs self-coupling. This would provide direct insights to the shape of the Higgs potential and provide a key test of the standard model. With a collision energy of 500 GeV or greater, the Higgs self-coupling can be accessed through di-Higgs production, with the dominant production modes being either di-Higgs-strahlung ($e^+e^- \to ZHH$) or WW fusion ($e^+e^- \to \nu_e\bar{\nu}_e HH$), depending on the center-of-mass energy [86–88].

Opportunities to test for CP violation in $H \to \tau^+\tau^-$ decays will be discussed in more detail in Section 2.4.

Precision Electroweak Physics

Given the challenging physics environment at a hadron collider, the sensitivity of electroweak measurements at the LHC is limited [89]. In contrast, a future e^+e^- collider would be able to probe the electroweak sector with unprecedented precision. These high precision measurements of the properties of the Z and W bosons would provide access to new physics at a scale well above the energies directly probed through higher order loop corrections [90, 91].

Such results may be obtained from dedicated runs, such as at the Z pole. Alternatively, for the ILC running at $\sqrt{s} = 250$ GeV, there is a high-cross-section for so called *radiative return* to the Z events, via the process $e^+e^- \to Z\gamma$. In this case, the photon is radiated from the incoming electron or positron as *initial state radiation* (ISR). The Z and the ISR photon are therefore produced in very forward yet opposite directions, but the Z can often still be reconstructed with the proposed detectors [83].

Top Physics

The top quark is the heaviest particle in the SM, and hence has the strongest coupling to the Higgs boson. This makes it a key probe of electroweak symmetry breaking, as well as being an important means by which to search for new physics. At an e^+e^- collider, a scan of the top quark pair production threshold can be conducted around $\sqrt{s} = 350$ GeV, whilst at higher energies the process $e^+e^- \rightarrow t\bar{t}H$ becomes accessible [83]. This will make precision measurements of the top quark properties, including its mass, width and electroweak and Yukawa couplings, possible [92, 93].

2.3.3 The International Linear Collider

The International Linear Collider (ILC) [20, 76, 80–82] is one of the foremost proposals for a Higgs factory and a high energy e^+e^- collider. The layout proposed for the ILC is shown in the schematic in Figure 2.2

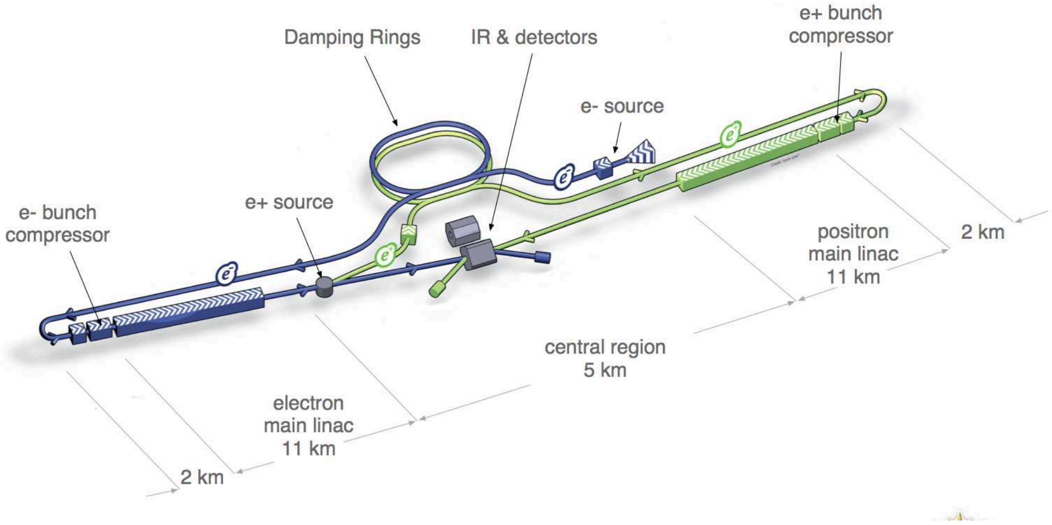


Figure 2.2: Schematic illustrating the layout of the ILC. Illustration not to scale. Figure from [20].

The baseline design for the ILC has a total length of 31 km, with two linear accelerators (linacs) based on superconducting RF cavities designed to facilitate the collision of longitudinally polarised electron and position beams. A staged approach has been proposed [94], in which a center-of-mass energy of 250 GeV would be initially selected for the machine, with the potential to upgrade the machine beyond this up to a center-of-mass energy of 1 TeV. The beams at the ILC would consist of bunch trains composed of 1312 bunches, each separated by 554 ns and containing $\sim 2 \times 10^{10}$ particles. The temporal separation between bunch trains is 200 ms, corresponding to a repetition rate of 5 Hz. At $\sqrt{s} = 250$ GeV, an instantaneous luminosity of 1.35×10^{35}cm^{-2}s^{-1} is expected. The beams have a crossing angle of 14 mrad at the collision point, and can additionally feature polarisation of up to 80% for electrons and 30% for positrons.

Two multi-purpose detector systems are envisioned for use at the ILC: the Silicon Detector (SiD) [82] and the International Large Detector (ILD) [64]. The ILD will be the focus of this thesis, and is described in more detail in Section 4.2.

2.4 Physics of the Tau Lepton

The tau lepton was discovered by the Mark I experiment conducted at the Stanford Linear Accelerator Center (SLAC) in 1975 [95]. As a third generation fermion, the tau is the heaviest known lepton with a mass of 1776.86 ± 0.12 MeV [96]. The mean lifetime of the tau is $(290.3 \pm 0.5) \times 10^{-15}$ s, meaning it can travel on the order of 10s of μm at relativistic velocities before decaying [96].

As a result of its high mass, the tau is the only lepton which decays to hadrons. The decay of the tau is mediated by the weak interaction, resulting in the production of a virtual W boson and a tau neutrino. The dominant decay modes of the tau are shown in Table 2.1. Approximately 65% of the time, the tau will undergo a decay involving hadrons, with the remainder of decays including either an electron or a muon, and usually being purely leptonic. The leptonic decay modes involve multiple neutrinos, which pass through a detector without leaving a signature and therefore result in less information being available to reconstruct the event. For this reason, the hadronic decay modes are prefered for precise measurements.

The majority of the hadronic decays of the tau occur via the intermediate resonances $\rho(770)$ or $a1(1200)$, and frequently involve one or more neutral pions. Additionally, the particles produced in the decay are typically highly boosted, and therefore more collimated than those present in QCD jets of a comparable energy [97]. This makes separating the particles in the detector, and therefore reconstructing the particles, challenging. In particular, incorrectly reconstructing the number of photons produced from any π^0s created ($\pi^0 \to \gamma\gamma$ occurs in 99% of cases [96]) means the number of π^0s produced is incorrectly reconstructed and the decay mode of the tau therefore mis-identified. For this reason, reconstruction performance for hadronic tau decays is a classic benchmark of the performance of an electromagnetic calorimeter (see Chapter 3) [97, 98].

Table 2.1: Branching ratios of the dominant decay modes of the tau lepton, separated into purely leptonic and hadronic decays and including intermediate resonances where appropriate. Decays modes of the τ^- are listed, with decay modes of the τ^+ being identical under charge conjugation and $h^\pm$ standing for $\pi^\pm$ or $K^\pm$ [96].

Category	Decay mode	Resonance	Branching ratio (%)
Leptonic	$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.82 ± 0.04
	$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.39 ± 0.04
Hadronic	$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	25.93 ± 0.09
	$\tau^- \to h^- \nu_\tau$		11.51 ± 0.05
	$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a1(1200)$	9.48 ± 0.10
	$\tau^- \to h^- h^+ h^- \nu_\tau$	$a1(1200)$	9.80 ± 0.05
	$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.76 ± 0.05

A key physics motivation for studying tau decays at a future e^+e^- collider is to investigate the CP[2] nature of the Higgs boson. In the SM, the Higgs boson is predicted to be a CP-even scalar, so the detection of any CP-odd structure in its interactions would be a clear indication of new physics. The decay $H \to \tau^+\tau^-$ is well suited as probe of the CP properties of the Higgs as it has a relatively large branching ratio of $\sim 6.3\%$ [99], and the lifetime of the tau is large enough such that it is possible for a decay vertex to be reconstructed. A precise reconstruction of the tau decay mode therefore makes it possible to use the spin of the τ lepton to investigate the CP structure of the Higgs sector [100].

---

[2] the combination of charge conjugation symmetry and parity symmetry

Chapter 3

Calorimetry

In high energy physics, the primary goal of calorimetry is to make an energy measurement of an incident particle. This is achieved by triggering an interaction between the high energy primary particle and the detector material, with a view to causing a cascade of secondary interactions and thereby a shower of secondary particles which should be contained within the volume of the calorimeter. The energy depositions of these secondary particles in the sensitive regions of the calorimeter are then recorded in order to infer the energy of the primary particle. The interaction of the particles with the calorimeter can be split into two classes- those which interact electromagnetically such as electrons, positrons and photons, and those which also interact hadronically such as pions, kaons and protons.

3.1 Electromagnetic Showers

The interaction of photons, electrons and positrons in matter is driven by a number of different processes, the relative frequency of which strongly depends on the energy scale involved.

3.1.1 Electron and Positron Interactions with Matter

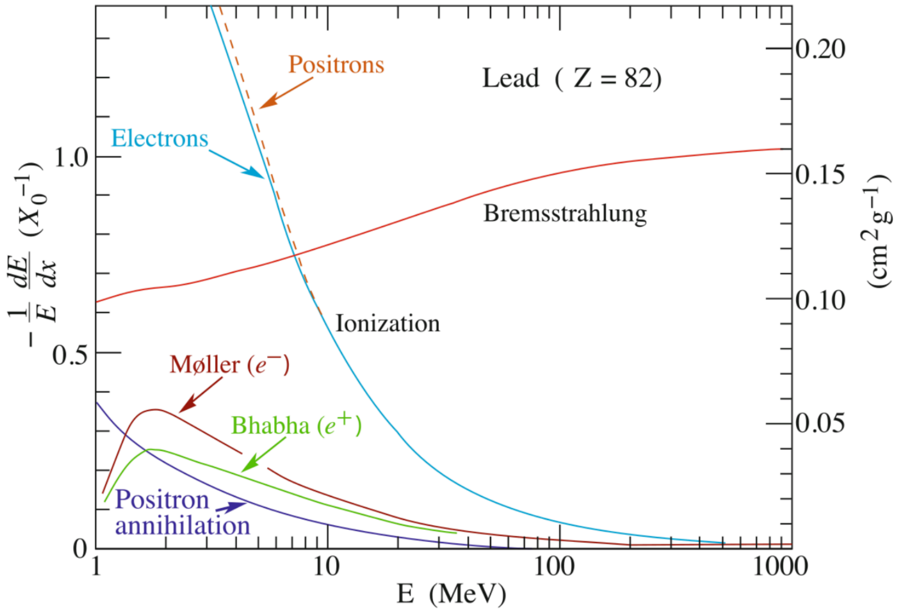


Figure 3.1: Electron (positron) energy loss per radiation length in lead as a function of particle energy for various interaction processes with matter. Figure from [96].

While the exact manner in which electrons and positrons interact with matter differs, they are predominantly driven by two processes; *ionisation* and *bremsstrahlung* [101]. *Ionisation* occurs when the electron or positron traversing the material transfers a sufficient fraction of its energy to free a bound electron, leaving behind an ionised atom. This is the dominant mode of interaction for low energy electrons or positrons. As the energy is increased above a certain threshold (about 10 MeV for lead), the relative energy lost to ionisation falls off steeply, and *bremsstrahlung* becomes the dominant interaction process. Bremsstrahlung occurs when a charge particle is deflected by the electromagnetic field of a nucleus and radiates a photon. Besides ionisation, and bremmstrahlung there are are several sub-leading processes by which electrons and positrons can interact with matter. Scattering with the electrons bound to an atom in the medium can also occur- *Møller scattering* in the case of electrons and *Bhabaha scattering* in the case of positrons. Additionally, positrons can annihilate with electrons in the medium, resulting in the production of two photons. These sub-leading processes provide small, but non-negligible contributions to the energy loss of electrons and positrons in matter, particularly for lower energies. The dependence of the energy loss for each process as a function of the particle energy in lead is shown in Figure 3.1.

3.1.2 Photon Interactions with Matter

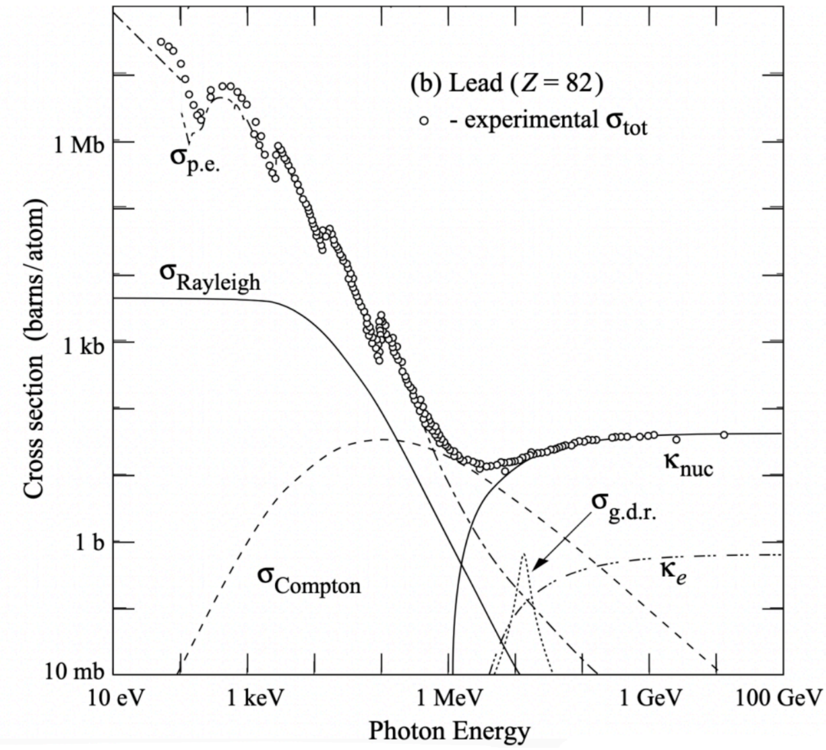


Figure 3.2: Cross sections for the interactions of photons in lead as a function of the photon energy. The total cross section obtained from experiment is shown, together with the contributions from individual interaction processes. The contributions from each process is denoted as follows: $\sigma_{p.e}$ is the contribution from the photoelectric effect, $\sigma_{Rayleigh}$ is the contribution from Rayleigh scattering, $\sigma_{Compton}$ is the contribution from Compton scattering, $\sigma_{\kappa_{nuc}}$ and σ_{κ_e} are the contributions from pair production in the electric field of a nucleus and an electron respectively, and $\sigma_{g.d.r}$ is the contribution from photonuclear interactions. Figure from [96].

Photons undergo electromagnetic interactions with matter through three main processes; *the photoelectric effect*, *Compton scattering* and *pair production* [101]. The dependence of their interaction cross-sections with matter on energy is shown for lead in Figure 3.2. *The photoelectric effect* occurs when a photon is absorbed by an atom, exciting a bound electron and freeing it from the atom. The photoelectric effect is the dominant process for lower photon energies, while the cross-section falls away for higher energies. At a material dependent threshold (around 1 MeV in lead), *Compton scattering* becomes the dominant process. This is the inelastic scattering of the photon with an electron bound to an atom (the elastic scattering process, *Rayleigh scattering*, is the sub-leading process to the photoelectric effect for lower energies, and falls off in a similar manner as the photon energy increases). The cross-section for Compton scattering rises as the photon energy increases relative to the binding energy of the atomic electron, to a peak value that is larger for materials with higher atomic numbers [102]. For energies exceeding a material dependent threshold (about 10 MeV in lead), the cross-section for Compton scattering falls off, and *pair production* becomes the dominant process. This is the production of an electron-positron pair as a result of the photon interacting with the electromagnetic field of either the atomic nucleus or an atomic electron. As such, it can only occur if the photon has an energy at least as large as two times the rest mass of the electron, equivalent to 1.022 MeV.

3.1.3 Electromagnetic Shower Development

When electrons, positrons or photons with sufficiently high energy are incident upon a dense medium, an electromagnetic cascade is triggered, with the constituent particles undergoing a mixture of the interactions described in Sections 3.1.1 and 3.1.2. For example, a primary photon may undergo pair production to produce an electron-positron pair, each of which will likely radiate a photon via bremsstrahlung. With sufficient energy, the bremsstrahlung photon can undergo pair production to produce another electron-positron pair. An electromagnetic shower evolving in this manner is shown in Figure 3.3. With this view on the development of electromagnetic showers being driven by bremsstrahlung and pair production in mind, several key parameters governing the shower become clear. The first is the *radiation length* X_0.

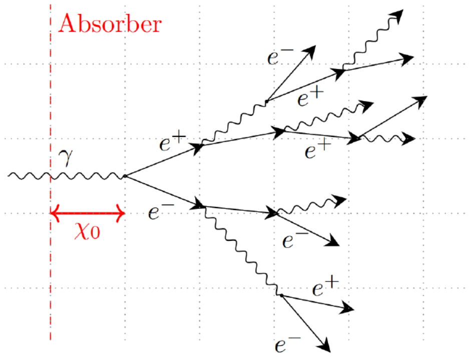


Figure 3.3: Illustration showing the development of an electromagnetic shower initiated by a photon. Figure from [103].

For electrons and positrons, X_0 characterises the length scale of bremsstrahlung interactions-being the average distance over which the particles' energy is reduced by a factor of $\frac{1}{\mathrm{e}}$ from its initial value E_0. Similarly, for photons the length scale for pair production is driven by the radiation length, with the intensity of a beam of photons being reduced to $\frac{1}{\mathrm{e}}$ of its initial value after travelling a distance of $\frac{9}{7}X_0$ [104]. This common length scale means that showers induced by electrons or positrons and those induced by photons are similar, with the major differences being that photon showers tend to start slightly further into the absorbing material, and exhibit larger fluctuations [1] [105]. The radiation length is a material dependent property, with denser materials featuring shorter radiation lengths. X_0 can be related [104] to the atomic number Z and atomic mass A by the approximation

---

[1] The fact that fluctuations are larger in photon showers than in electron or positron showers is due to

$$X_0 \approx \frac{716A}{Z(Z+1)\ln(287/\sqrt{Z})}[\text{gcm}^{-2}]. \tag{3.1}$$

The continuation of the shower is contingent on interactions that produce particles, meaning that the number of particles in the shower increases. However, with each subsequent interaction the energy is reduced. This gives rise to a second key parameter- the *critical energy* E_c. This is the energy for electrons or positrons at which the contribution to the energy loss from bremsstrahlung (which results in the production of a photon) is equal to the energy loss from ionisation (which does not produce a new particle) [105]. According to this definition [2], an empirical approximation [101] for E_c, with material dependence through the atomic number Z, is given by:

$$E_c \approx \frac{800}{Z+1.2}[\text{MeV}]. \tag{3.2}$$

This provides a means by which to characterise the approximate extent of a shower. Longitudinally, if the depth t in the material is measured in units of radiation length (i.e. $t = x/X_0$, where x is the physical depth), when the average energy per particle is less than or equal to E_c the maximum shower depth t_{max} will have been reached. For a given incident energy E_0, this can be shown [101] to be given by

$$t_{max} \approx 1.4 \ln\left(\frac{E_0}{E_c}\right). \tag{3.3}$$

The transverse extent of a shower is characterised by the *Molière radius* R_M. It does not have the same physical meaning as the interaction length X_0, but typically contains $\sim 85-90\%$ of the shower energy [105]. It is given by the approximate empirical expression [104]

$$R_M \approx \frac{21\text{MeV}}{E_c} X_0. \tag{3.4}$$

Parameterising showers in terms of their longitudinal and radial development has been used as an approach for fast simulation of calorimeter showers, and will be discussed in Section 6.2

3.2 Heavy Charged Particle Interactions with Matter

In comparison to positrons and electrons where the primary form of energy loss is radiative as a result of bremsstrahlung, for heavier charged particles such as muons and charged hadrons, energy loss is dominated by ionisation to much higher energies. This is due to the cross section for bremsstrahlung being inversely proportional to the fourth power of the mass, resulting in a strong suppression of the process for particles with higher masses. The mean energy loss per unit distance travelled due to ionisation is described by the Bethe-Bloch equation [105]

$$-\left\langle \frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{max}}{I^2} - \beta^2 - \frac{\delta}{2} \right], \tag{3.5}$$

where γ is the Lorentz factor, β is the velocity of the particle, m_e is the mass of the electron, I is the mean excitation energy of the absorbing material, Z and A are the atomic number and atomic mass respectively, z is the charge of the particle traversing the material,

---

the photon no longer existing after pair production, whereas an electron or positron can undergo subsequent bremsstrahlung interactions.

[2] There is an alternative definition of the critical energy- the energy at which the ionisation loss *per radiation length* is equal to the energy of the electron [104] [105]

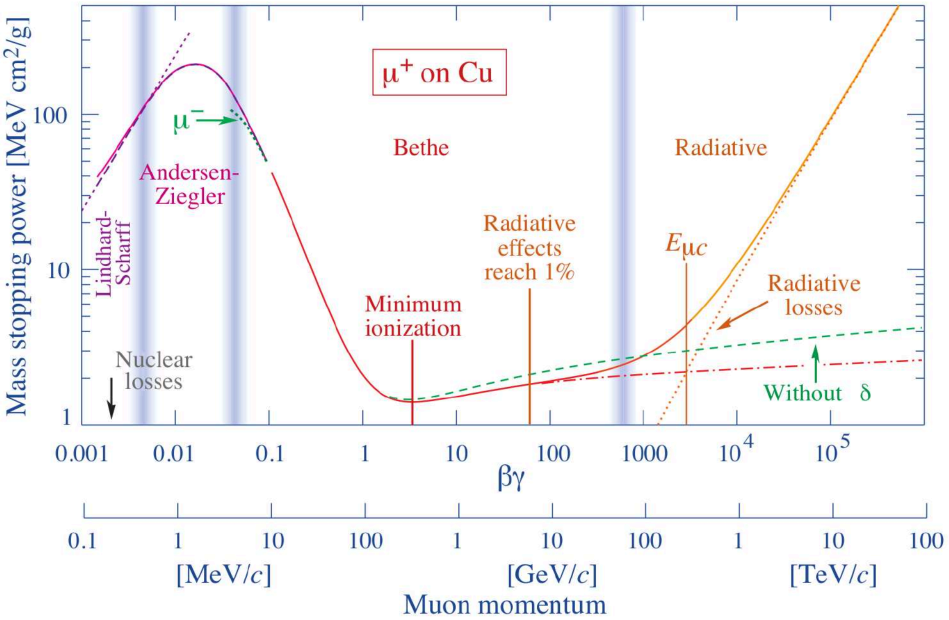


Figure 3.4: Mass stopping power $-\left\langle \frac{dE}{dx} \right\rangle$ as a function of momentum for positively charged muons in copper. Solid lines denote the total stopping power. Figure from [96].

K is the constant of proportionality and T_{max} is the maximum kinetic energy transferred per collision. The term including δ is a correction to account for the effects of polarisation in the medium through which the particle travels, and causes the logarithmic rise present at higher particle momenta to be truncated [96]. The energy loss due to ionisation for positively charged muons of various momenta in copper is shown in Figure 3.4. At $\beta\gamma \approx 3-4$, the curve reaches a a region of minimum ionisation. Given the slow rise in energy loss due to ionisation as the momentum of the muon increases, this region extends over several GeV. Charged particles with momenta corresponding to this region of minimum ionisation (the exact momentum range depends on the particle species), are known as *Minimum Ionising Particles* or *MIPs*.

As calorimeters are typically composed of relatively thin sections of material, the total energy loss in the material can fluctuate considerably. This causes a large variation in the mean energy loss, due to high energy interactions that do not occur often producing a long tail to high energies. However, the peak position of the energy loss (the most probable value) remains largely unaffected. This is because for MIPs, particles only lose a relatively small amount of energy. As such a particle can undergo many such interactions, thereby producing multiple energy depositions with values around a MIP. The stability of this value for a given detector makes it a well defined quantity, and as such an important means of calibration.

3.3 Hadronic Showers

In hadronic showers, in addition to electromagnetic interactions, interactions with the nuclei of the absorbing material involving the strong interaction can also occur. This means that hadronic showers exhibit significantly more varied and richer topologies than electromagnetic showers.

The development of a hadronic shower is illustrated in Figure 3.5. While charged hadrons also undergo ionisation interactions as described in Section 3.2, hadron showers are triggered by an initial inelastic hard scattering (hadronic interaction) with nuclei of the absorbing material. The hadronic interactions that occur in a shower can be broken into two parts — those that pertain to interactions with the nuclei, and those related to the particles in the shower. In

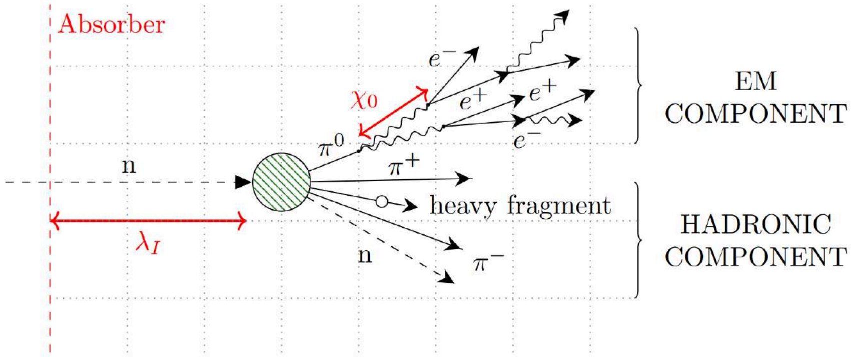


Figure 3.5: Illustration showing the development of a hadronic shower initiated by the interaction of a neutron with an atomic nucleus. The shower consists of both hadronic and electromagntic components. Figure from [103].

the first instance, a number of different hadronic processes with the nuclei of the material can occur. Examples include spallation, fission and evaporation. These processes typically result in the production of pions and eta mesons, as well as the ejection of nucleons and nuclear fragments [106]. These secondary particles and fragments can then undergo further interactions with the detector material. In the second instance, particles in the shower may decay. Neutral pions and eta mesons that are produced in the shower often undergo a decay that results in the production of two photons. For example, a π^0 decays to two photons 99% of the time [96]. This means that hadronic showers consist of a considerable *electromagnetic fraction*. While the electromagnetic fraction fluctuates from event to event, on average roughly one-third of mesons produced in the first interaction decay electromagnetically [106].

An additional feature of hadronic showers is that some fraction of the energy is lost and inherently cannot be measured. This is termed *invisible energy*. This energy can be lost through nuclear interactions, for example to binding energy, the recoil or excitation of nuclei in the material, or to neutrons that can travel large distances (on the order of cm) until they are stopped, which may not happen in the sensitive volume. Additionally, the decay of charged hadrons, may result in the production of neutrinos, which escapes the detector without leaving any signature.

The length scale which characterises the development of a hadronic shower is known as the *nuclear interaction length* λ_I. This quantity describes the mean free path between hadronic interactions, and is in general defined as

$$\lambda_I = \frac{1}{n\sigma_{had}}, \tag{3.6}$$

where σ_{had} is the cross section for the hadronic interaction and n is the number of atoms per unit volume in the target material [101]. When expressed in units of gcm^{-2}, eliminating material density effects, λ_I scales with $A^{\frac{1}{3}}$, where A is the atomic weight [105]. Since for the majority of materials, the nuclear interaction length is significantly larger than than the radiation length, hadronic showers tend to extend over much larger distances than electromagnetic showers. Additionally, because of the much greater variety of processes that can occur for a given interaction, and the corresponding large variation in the number of secondary particles that are produced, hadronic showers display much more complex topologies and exhibit significantly larger event-to-event fluctuations.

3.4 Calorimeter Designs and Measurements

The task of calorimetry is typically divided between two sub detectors; an electromagnetic calorimeter (ECAL), which is placed in front of a hadronic calorimeter (HCAL). The ECAL aims to make a measurement of particles interacting purely electromagnetically, while the HCAL aims to make a measurement of particles which also interact hadronically (in practice, since the ECAL has a non-zero hadronic interaction length (see Section 3.2), showers initiated by hadrons can also start in the ECAL). These detectors consist of an absorbing material which is typically dense in order to increase the frequency of interactions between the particles in the shower and the material, thereby maximising the containment of the shower in as small a volume as possible. In the ECAL, the exact type and amount of material is chosen as a balance between maximising the number of radiation lengths of the material (see Section 3.1.3), and minimising the number of hadronic interaction lengths (see Section 3.3).

The precision of the energy measurement provided by a calorimeter is characterised by the *resolution*. This is defined by the ratio of the uncertainty on the energy measurement σ_E to the energy of the primary particle incident on the calorimeter E_{mes}. The resolution of a calorimeter can be expressed [104] as the sum of three contributions in quadrature

$$\left(\frac{\sigma_E}{E_{mes}}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2. \tag{3.7}$$

The first term, involving a, is the so called *stochastic* term. This results from the fact that shower evolution is not a deterministic process. The $\frac{1}{\sqrt{E}}$ dependence arises from assuming Poisson statistics, and that the number of particles created N is approximately proportional to the energy of the incident particle E [105]. The second term, which involves b, is the *noise* term. It arises from electronics noise, and thus is independent of the energy of the incident particle. Finally, the *constant* term c provides a contribution that does not vary with the incident energy. It represents limitations that are inherent to the detector, such as incorrect calibrations and material inhomogeneities.

Calorimeters are broadly divided into two main categories. *Homogeneous* calorimeters are detectors with a completely sensitive volume, which also serves as the absorbing medium. The technology for these calorimeters can rely on numerous different physical processes, for example collecting ionisation charge from liquid Nobel gases, Cherenkov light from a transparent medium (e.g. lead glass) or scintillation light from materials such as BGO, CsI or PbWO4 crystals [104] [101]. An example of the latter in use is for the electromagnetic calorimeter of the CMS experiment, where lead tungstate (PbWO4) crystals are coupled to avalanche photodiodes for the detection of scintillation light [107] [108]. The key advantage of homogeneous calorimeters is that they can provide outstanding energy resolution, as the entire energy of a shower is deposited in the sensitive material. However, the cost of these detectors is typically high, and crucially for particle flow reconstruction (see Section 4.1), it is difficult to produce a highly granular and longitudinally segmented homogeneous calorimeter [104].

The second main category of calorimeter is the *sampling* calorimeter. These calorimeters consist of layers of sensitive detector material interleaved with passive absorber material. This gives sampling calorimeters a natural longitudinal segmentation. A characteristic quantity of a sampling calorimeter is the so-called *sampling fraction* f_s. In this work, we define f_s as the ratio of the *visible* energy deposited in the sensitive layers of the calorimeter to the total energy deposited in both the sensitive and passive materials of the calorimeter. This means that sampling calorimeters tend to have a worse energy resolution than homogeneous calorimeters, as the sampling fraction has to be folded into the stochastic term given in Equation 3.7. Fluctuations in whether energy is deposited in the sensitive or absorber material increase fluctuations in the visible energy of the shower, and thereby degrade the resolution. A key advantage of sampling calorimeters is that they can typically be more compact than homogeneous calorimeters, as removing the requirement for the material to be sensitive allows denser materials such as tungsten to be used. This factor, along with the fact that passive materials are less expensive, makes sampling calorimeters much more cost efficient. It is for this reason that sampling calorimeters are used almost exclusively for hadronic calorimeters in detectors for collider-based experiments, as the large volume that has to be covered prohibits the use of a homogeneous calorimeter [104]. Examples of sampling calorimeters include the Calorimeter Endcap Upgrade for the CMS experiment for the upcoming High Luminosity phase of the LHC [109] and the calorimeters developed by the CALICE collaboration for use at future Higgs Factories [110–113]. In particular, these examples of sampling calorimeters feature a highly-granular transverse segmentation of the sensors, giving them a high spatial resolution and allowing detailed information about a shower's substructure to be recorded. In the case of calorimetry at a Higgs Factory, this is essential for the particle flow approach to reconstruction (see Section 4.1), and at a hadron collider it allows for pile-up rejection. The two calorimeter designs used in this thesis will be discussed in more detail in Section 4.2.1.

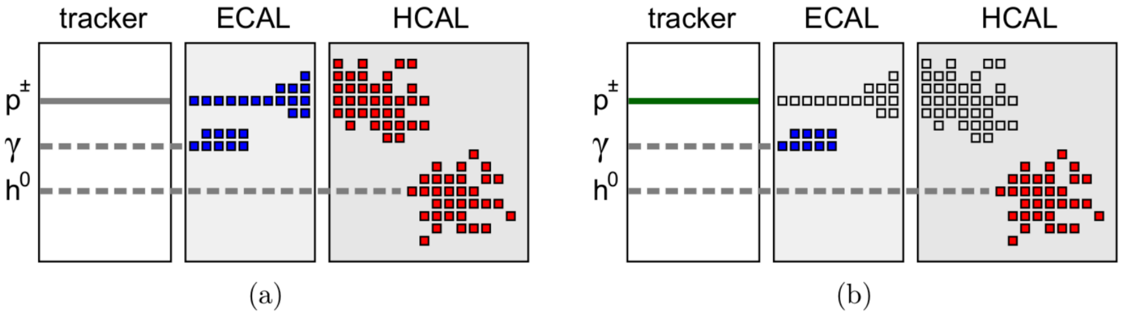
Chapter 4

Particle Flow Reconstruction and the International Large Detector

To be able to conduct the broad physics programme of a future high energy e^+e^- collider outlined in Section 2.3.2, a key requirement is to be able to distinguish hadronic decays of Z and W bosons resulting in multi-jet final states [114, 115]. This essentially means having a resolution on the invariant mass of the two bosons comparable to their widths, and necessitates a jet energy resolution of $\sigma_E/E \lesssim 3.5\%$ for $50 - 500$ GeV jets [116]. This requires a revised approach for event reconstruction, called *particle flow* reconstruction, to be adopted. This will be described in Section 4.1. The International Large Detector, which has been optimised around the particle flow approach to reconstruction, will be described in Section 4.2. This will include a description of both the subsystems included in the detector, with a focus on the calorimeter subsystems, and the software and reconstruction procedures adopted.

4.1 Particle Flow Reconstruction

Particle flow reconstruction is an alternative approach to the traditional reconstruction procedure that is required to the meet the physics goals at a future e^+e^- collider.



(a) (b)

Figure 4.1: Illustration comparing a) the traditional approach and b) the particle flow approach to calorimetry. The interaction of the particles composing a jet (charged particles (mostly hadrons) p$^\pm$, photons γ and neutral hadrons h^0) with the detector components is shown. Coloured areas show the detector regions used to perform measurements of each category of particles, while dashed lines indicate when no signature is left in the detector subsystem. Figure taken from [117].

To understand the reason for this consider the composition of a jet based on measurements at LEP [118], which is typically approximately 60% charged particles which are mostly hadrons, 30% photons and 10% neutral hadrons. In the traditional approach to reconstruction, which is shown in Figure 4.1 a), the energy measured for a jet E_{Trad} is based on a purely calorimetric measurement

$$E_{\text{Trad}} = E_{\text{ECAL}} + E_{\text{HCAL}}, \tag{4.1}$$

where E_{ECAL} and E_{HCAL} are the energy measurements made in the ECAL and HCAL respectively. In this case, the HCAL would therefore be used to measure most of the energy of the hadrons, which compose approximately 70% of the jet. Since the energy resolution of the HCAL is poor due to the complex physics governing hadronic showers (see Chapter 3), with a typical resolution of $\gtrsim 55\%/\sqrt{E(\text{GeV})}$ [116], this makes achieving the required jet energy resolution with this approach very challenging.

In the particle flow approach to reconstruction, which is illustrated in Figure 4.1 b), the goal is to perform the measurement for a given particle in the detector subsystem which provides the best resolution. This means the energy measurement of a jet E_{PFlow} is decomposed as follows

$$E_{\text{PFlow}} = E_{p^{\pm}} + E_{\gamma} + E_{h^0} \tag{4.2}$$

where $E_{p^{\pm}}$ is derived from the momentum measurement of charged particles in the tracking detector and E_{γ} and E_{h^0} are energy measurements of photons and neutral hadrons in the calorimeters. The advantage of this is that now only about 10% of the jet constituents are measured in the HCAL.

The key challenge is that the particle flow approach requires the reconstruction of the four momenta of each individual particle in an event. The energy depositions in the detector must therefore be correctly assigned to the particle which caused them. This means a correct clustering of hits in the calorimeters, including a clean separation of energy depositions originating from different particles, and the association of clusters with correctly reconstructed tracks where appropriate. This requires sophisticated clustering algorithms and calorimeters of a sufficiently high granularity to allow a separation of the showers of particles produced.

Confusion in the reconstruction procedure is a key cause of a reduced performance in terms of jet energy resolution. This confusion can take several forms, and can relate to the association of individual hits in the calorimeter to clusters, as well as the association of tracks to clusters. For instance, a *loss* in energy can occur if a neutral particle, be it a photon or a neutral hadron, is not properly separated from a nearby charged hadron in the calorimeter. This would result in the energy from the neutral particle being combined into the calorimeter cluster associated to the charged hadron, which is then discarded during the measurement in favour of the track. Conversely, a *double counting* of energy will occur if a *fragment* of the calorimeter shower caused by a charged particle is reconstructed as a separate, neutral particle, since the original charged particle will still be reconstructed from its track [116].

The current state-of-the-art Particle Flow Algorithm (PFA) is PANDORAPFA [115, 119], which consists of a complex hierarchy of pattern recognition algorithms in order to reach the required performance in terms of jet energy resolution. Of particular relevance in this thesis will be the reconstruction of photons.

In total, there are five different dedicated algorithms applied in PANDORAPFA to tackle photon reconstruction, following developments in [120, 121]. The first is the main algorithm for photon reconstruction, which clusters hits in the ECAL into a photon candidate and attempts to verify the identity of the candidate, with additional checks performed if the photon is in proximity to one or more charged particles. The next three algorithms aim to remove fragments created by energy depositions being split off of the photon when they should not be. These fragments can occur around the edges of a shower once the highly energetic core has been identified as a photon, incorrectly creating additional neutral particles. Additionally, if a photon has enough energy *leakage* will occur, meaning that the shower is not fully contained in the ECAL and continues into the HCAL, producing a fragment in the HCAL that often creates a neutral hadron. Finally, a splitting procedure is applied to separate any photons that were incorrectly merged during fragment removal. This can happen when the spatial

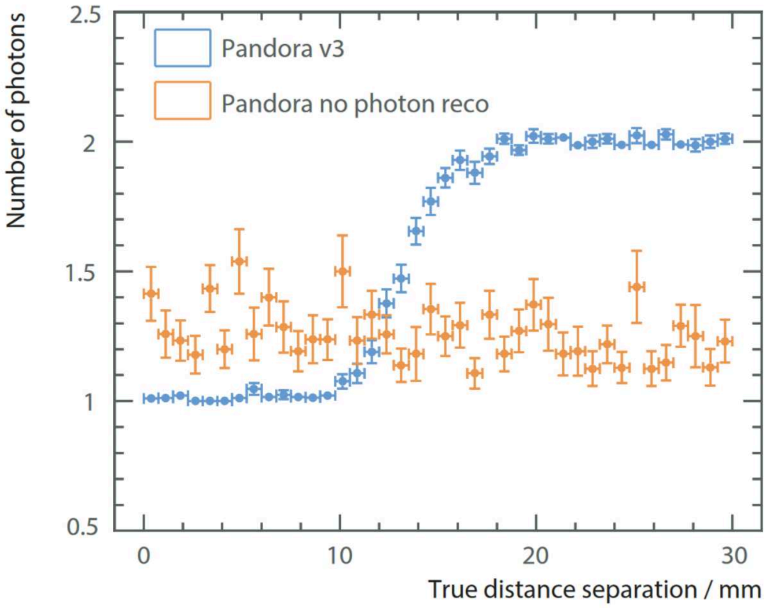


Figure 4.2: Average number of photons reconstructed as a function of the true separation between two photons. The sample was created for two photons with disparate energies of 500 GeV and 50 GeV without (orange) and with (blue) the dedicated photon reconstruction algorithms. Figure and caption taken from [120].

distance between photons is small. Figure 4.2 shows the number of reconstructed photons as a function of the separation between photons with different energies in the ILD detector for PandoraPFA applied with and without the dedicated photons reconstruction procedure. At a distance of 10 mm, it is possible to start to resolve the two photons, with a clear reconstruction of the two photons occurring at a separation of around 20 mm [120].

4.2 The International Large Detector Concept

The International Large Detector (ILD) [64, 82] is a multi-purpose detector concept proposed for operation at the ILC. The detector concept is visualised in Figure 4.3.

The detector is optimised for the particle flow approach to reconstruction, with a core goal of the design being to obtain a jet energy resolution of less the 4% for jets with energies in the range $40 - 250$ GeV [82]. This places a number of key requirements on the capabilities of the detector. Firstly, the momentum resolution must be sufficiently high, with the requirement $\sigma_p/p^2 < 5 \times 10^{-5}$ GeV^{-1}. Secondly, the material budget in front of the calorimeters has to be kept as small as possible, to reduce the effects of multiple scattering. Thirdly the detector must be highly hermetic to maximise the number of visible particles that can be detected. Finally, the calorimeters must be highly granular, to provide a high level of resolution needed for the separation of overlapping showers.

4.2.1 Detector Systems

The ILD detector model studied in this thesis has a polyhedral barrel with a total length of 13 m and a total radius of 7.8 m. The tracking and calorimeter systems are located inside a superconducting solenoid coil, which creates a 3.5 T magnetic field orientated along the beam axis. Outside of the coil sits the iron return yoke, with an integrated muon system and tail catcher calorimeter [64]. The key detector subsystems are as follows, and are illustrated in Figure 4.3 b).

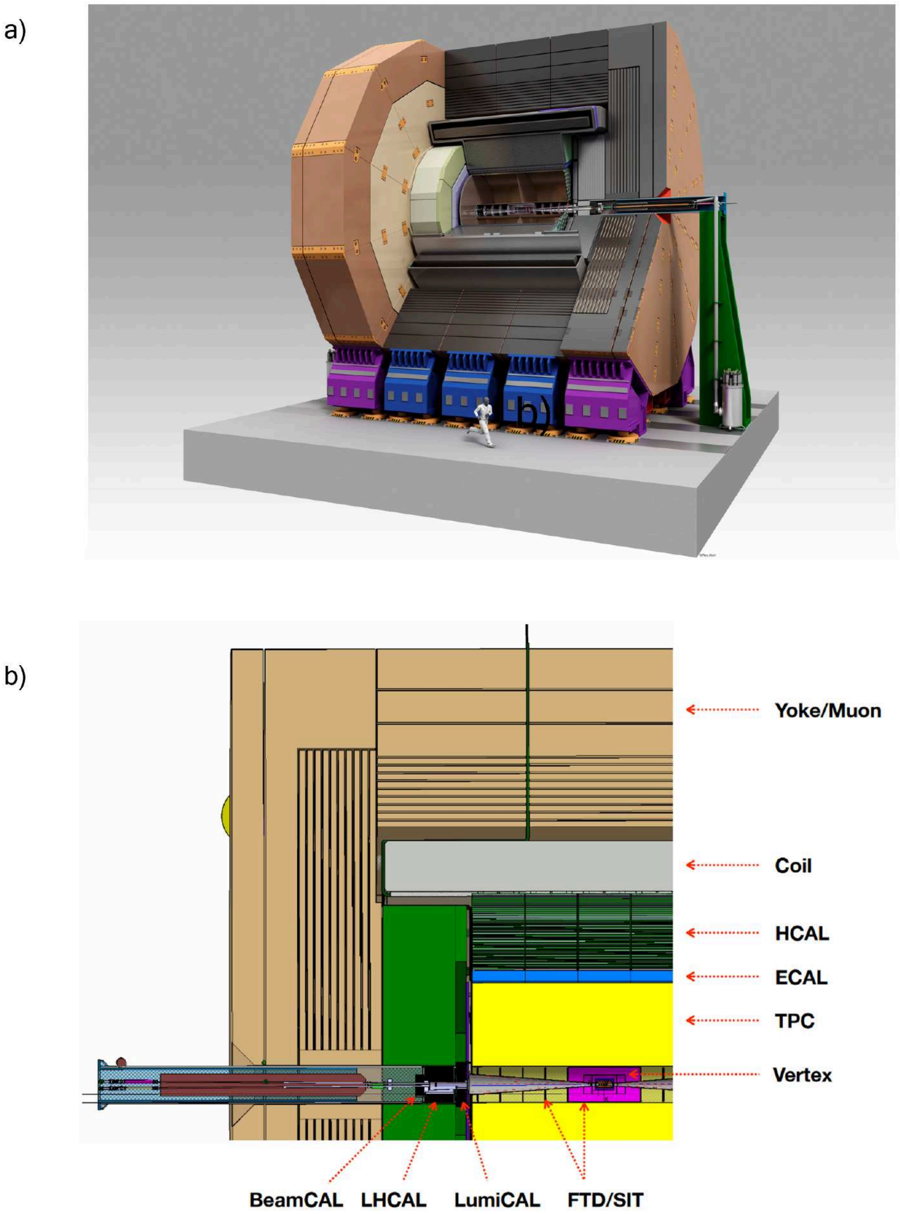


Figure 4.3: The ILD detector concept a) rendered as a three dimensional model [20] and b) in cross-section, with the key detector subsystems labelled [64].

Tracking Systems

The innermost component of the tracking systems is the vertex detector (VTX). It consists of a barrel of three double-layers of silicon pixel sensors, with the closest placed 16 mm from the interaction point. The system is optimised for a point resolution better than 3 µm, and a material thickness of less than $\sim 0.15\%\ X_0$ per layer. Additional silicon trackers are placed around the VTX detector, with two layers in the barrel (SIT) and 7 disks in the forward direction (FTD) to allow tracking of particles to low angles. Outside of this sits the main tracking device, which is a time projection chamber (TPC) with a length of about 4.6 m and a radial span from about 33 to 180 cm. Tracking is achieved by using an electric field to collect electrons produced by ionisation in the gas filling the volume, which are readout with either gas electron multiplier (GEMs) [122] or micro-mesh gaseous structure (micromegas) [123] detectors featuring pad sizes of a few mm^2. It can provide continuous tracking with up to 220 three dimensional space points, and enables particle identification through the specific energy loss dE/dx (See Chapter 3), while only providing a small contribution to the material budget. Additional silicon tracking layers are placed after the TPC and before the calorimeter systems. The full tracking system has a momentum resolution of about 2×10^{-5} GeV^{-1} [64].

Calorimeter Systems

The ILD detector concept features highly granular calorimeters, separated into an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL).

The Silicon-Tungsten Electromagnetic Calorimeter

Two proposed options for the ILD ECAL exist. Both are sampling calorimeters and use tungsten absorber layers. The first option is the SiW ECAL [110], which makes use of silicon sensors in the active layers of the calorimeter, while the second option is the Sc ECAL [111] which uses scintillator strips in the active layers.

This thesis will focus on the SiW ECAL. This design is composed of 30 active silicon layers stacked between tungsten absorber layers. The calorimeter is designed with two different sampling fractions, with the first 20 tungsten absorber layers being 2.1 mm thick, and the last 10 layers having a thickness of 4.2 mm. The total absorber thickness in the calorimeter therefore corresponds to $24\ X_0$. The active layers consist of silicon wafers which are 0.525 mm thick and feature cells of size 5×5 mm^2. The edges of wafers represent regions of additional dead material, for example as a result of readout electronics.

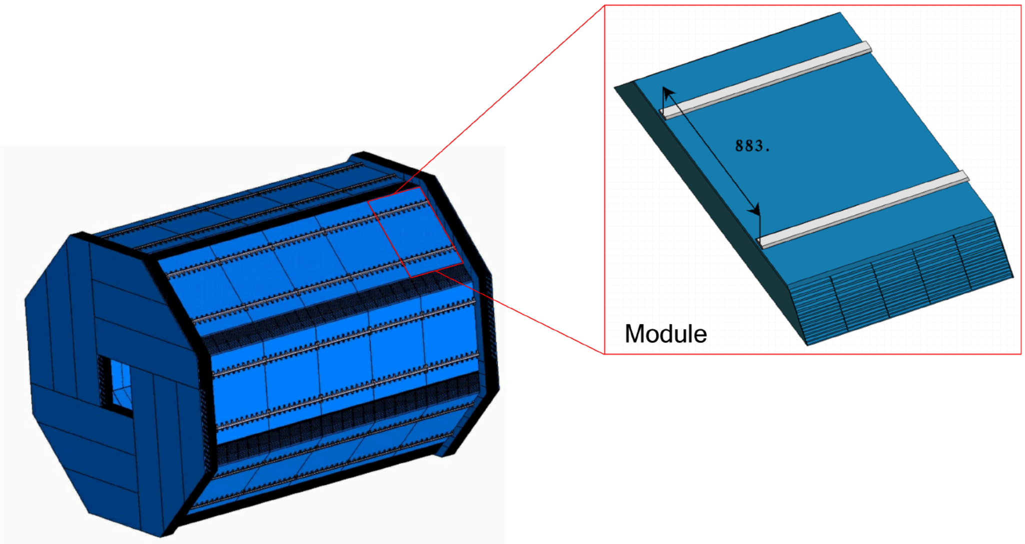


Figure 4.4: Geometrical structure of the ILD ECAL, consisting of the octagonal barrel and two endcaps. In the final construction, a plug consisting of two rings would also be inserted into the gap in the face of each endcap. Inset: one module in the barrel region consisting of a stack of every other tungsten absorber layer with a carbon-fibre supporting structure. Slabs of tungsten sandwiched between active silicon layers are then inserted into the alveoli (gaps between layers in the stack) to complete the module. Figure adapted from [124]

The basic building block of the ECAL subsystem is a *module*, which features a tungsten stack with carbon-fibre supports. The module has an alveolar structure, consisting of every other tungsten absorber layer. The structure of a module in the barrel region is shown in the inset in Figure 4.4. Into each alveolus (gap in the module structure) a *slab* is inserted, which consists of a single layer of tungsten absorber sandwiched between two active silicon layers. This gives the calorimeter an alternating structure along its depth. Due to the trapezoidal shape of the module, a staggering effect occurs between layers. This also serves to prevent a particle passing through the insensitive volumes at the edge of a wafer in consecutive layers. The ECAL subsystem is composed of an octagonal barrel, two endcaps and two rings which plug into the apertures around the beam pipe in each endcap. The octagonal barrel, assembled together with the two endcaps, is shown in Figure 4.4. These will be the principal detector regions used in this thesis. The barrel consists of 8 *staves* running along its length, with each stave being made of 5 trapezoidal modules. Each endcap consists of 12 modules divided into 4 quadrants [125]. The modules in the endcap have similar structure to those present in the barrel, but a different shape [124]. The implementation of the ECAL in the ILD detector simulation chain features a detailed geometry implemented in DD4hep detector description toolkit (see Section 4.2.2), including the aforementioned geometrical structures and a realistic material composition.

The Analogue Hadronic Calorimeter

Two proposals for the ILD hadronic calorimeter exist. The first is the Semi-Digital Hadron Calorimeter (SDHCAL) [113], and the second is the Analogue Hadron Calorimeter (AHCAL) [112]. The AHCAL will be the design used in this thesis. It consists of 48 layers of stainless steel absorbers, each with a thickness of 17.2 mm resulting in a total of $\sim 4\lambda_I$. The active layers consist of 3×3 cm^2 scintillator tiles with a thickness of 3 mm, and silicon photomultipliers which read out the tiles individually. As with the ECAL, a detailed detector model of the AHCAL is implemented in DD4hep.

Very Forward Detectors

A number of dedicated detectors are placed in the very forward region of the ILD. This includes a calorimeter with a keystone layout designed for a precise measurement of the luminosity using the pairs of electrons from Bhabha scattering (LumiCal), a calorimeter to provide additional hadronic calorimeter coverage in the forward region (LHCAL) and a system to provide calorimeter coverage down to 6 mrad, as well as bunch-by-bunch monitoring of the beam. A very similar technology to that used in the electromagnetic calorimeter will be used, with adaptions for the environment in the forward region [64].

4.2.2 Software Ecosystem

ILD makes use of the iLCSoft [126] [1] software ecosystem for the full chain of detector simulation, reconstruction and analysis. The LCIO [128] framework provides the event data model and handles data persistency. The detector description toolkit DD4hep [129] is used to create a realistically detailed detector model, in terms of the material composition, structure and readout geometry. Within DD4hep, DDG4 [130] is used to interface with Geant4 [25] (see Chapter 6 for more details on simulation), and DDRec [131] is used as the geometry interface during reconstruction. The Marlin [132] application framework is used for both reconstruction and analysis.

4.2.3 Reconstruction Scheme

The full reconstruction scheme used by ILD consists of a chain of separate steps [133]. These steps are described in the following.

---

[1] Note that key components of this software framework are also integrated into Key4hep, the common turnkey software stack under development for future colliders [127].

Digitisation and Calibration

Initially, a *digitisation* procedure is applied to the simulated hits created in the detector. This involves emulating effects stemming from the detector, such as features which are intrinsic to a given sensor, as well as those caused by the readout electronics. Additionally, for the calorimeters a *calibration* composed of two steps is applied. Firstly, the energy deposited in each cell is converted into units of MIPs (a unit system based on the most probable energy deposited by a minimum ionising particle, see Chapter 3). Secondly, the energy of the hit in MIPs is converted into a value in GeV. After the calibration procedure, the total energy deposited by a given particle in the calorimeter system corresponds to the the incident energy of the particle. An additional *gap hit correction* is applied in regions of the calorimeter with additional insensitive volumes, and hence a reduced sampling fraction which is specific to that region of the calorimeter.

Reconstruction

The first step of reconstruction is *tracking*. This involves reconstructing the trajectory of charged particles, as well as the kinematics of the trajectories. This procedure uses several different algorithms for the different regions of the tracking system [134]. The tracking is complemented with dedicated algorithms. In particular, specific algorithms are applied to reconstruct *kinks* in tracks, which may arise from bremsstrahlung or multiple scattering. In addition, specific algorithms are applied to reconstruct V^0s, which are secondary vertices which may, for example, arise form the conversion of a photon into an e^+e^- pair (*gamma conversion*).

In the second step, PandoraPFA is applied to digitised inputs, which include tracks, kinks, V^0s and calorimeter hits, as well as information from DDRec about the geometry [133]. The procedure then involves algorithms applied to cluster hits in the calorimeters, and subsequently to associate tracks and clusters where appropriate. A preference is given for splitting clusters initially rather than incorrectly merging them. Re-clustering algorithms are applied later in the chain based on topological and kinematic information [115]. An identity can then be assigned to the particle type. A dedicated procedure called *software compensation* [135] is also applied to correct for the unequal detector response between the electromagnetic and hadronic shower components. The final output is a list of reconstructed particles called *Particle Flow Objects* (PFOs), which contain important information associated with the particles including their energy, momentum and particle ID.

Finally, high-level reconstruction is applied. This involves primary and secondary vertex finding, as well as jet clustering and tagging using the *LCFIPlus* package [136]. A wide range of additional dedicated algorithms for various analysis specific applications also exist. This includes tools for particle identification (PID), which are under active development [137] and may make use of dE/dx, shower shape and time-of-flight (TOF) [138] information, as well as various event shapes. Additionally, $\gamma\gamma$-finders allow the identification of π^0 and η mesons. Candidates are identified by means of a constrained kinematic fit using the MarlinKinfit [139] package.

Chapter 5

Machine Learning

Machine Learning (ML) is a broad term used to describe a class of algorithms that can learn from data in an automated, typically iterative, fashion [140]. ML algorithms have long been explored for various tasks in High Energy Physics, including tracking, jet tagging and triggering [141–143].

However, recent advances in specialised hardware enabling significant parallelisation of computational operations (most notably with Graphics Processing Units (GPUs)) have lead to major strides in the performance achievable with ML algorithms, such that they are now routinely used.

ML algorithms can be generally split into three broad categories: supervised learning, unsupervised learning and reinforcement learning. Algorithms using reinforcement learning are beyond the scope of this work. *Supervised* learning approaches rely on the use of a label or target for each data example that guides the algorithm towards a given goal. Examples of this task include classification, where an algorithm seeks to correctly label a given example, and regression where an algorithm attempts to predict a target value. By contrast, *unsupervised* learning approaches typically aim to model the underlying probability distribution of a given dataset by learning to extract salient information contained within its structure. [140]. Generative modelling is an example of an unsupervised task, and will be the principle ML application explored in this thesis.

In Section 5.1, a brief overview of the basics of deep learning will be given, including a definition of key terms that will be used throughout this thesis. In Section 5.2, the building blocks of the architectures explored in this thesis, including network layers and activation functions, will be described. Section 5.3 will discuss the optimisation algorithms that will be employed to facilitate learning. Finally, Section 5.4 will provide an overview of the various generative models that will be important for this thesis.

5.1 Neural Networks and Deep Learning

Artificial *neural networks* (NNs) were originally inspired by networks present in biological systems, and their ability to display complex emergent computational capabilities [144]. NNs consist of a network of *nodes* (or neurons) that can interact with other nodes with a certain *weight*. It is neural networks that are the foundational components of modern *deep learning*. A network is deep if it consists of multiple layers of nodes- the first layer is called the *input* layer, while the final one is termed the *output* layer. The layers that lie in between these two are so-called *hidden* layers. The presence of depth in a network (see Section 5.2 for more detail), as well as the addition of non-linear activation functions (see Section 5.2.5) between layers, are key drivers of a networks' expressiveness.

Networks are typically not fed an entire dataset at once, but rather subsets of the data- referred to as *mini-batches* or *batches*. This helps to improve training times and efficiency.

Data is then passed through each layer of a network during *forward propagation*.

To measure the performance of the network, the output of the full network must then be evaluated by means of a task-specific *objective* or *loss function*. Once the network performance has been quantified, it can be improved by means of an *optimisation* procedure in which information is *backpropogated* through each layer consecutively, starting from the output layer. It is this updating of the connections between nodes in the network that allows a model to learn. The performance of a network is typically optimised by minimising the loss function. The optimisation algorithms that will be used in this thesis will be addressed in Section 5.3.

One forward and backward pass over a batch is called an *iteration*, while one complete use of the full training dataset is called an *epoch*.

The overall process of a model training can be controlled by *hyperparameters*. These concern anything from the structure of a model (e.g. the number of nodes in a layer, the number of layers etc.), through the manner in which the dataset is passed (e.g. the size of each batch, the number of epochs etc.), to the optimisation procedure used in the training. The training process can therefore be improved by tuning these hyperparameters.

While a network may learn to perform well on a given task with data that was used to train it, the performance of the model on this set of data does not necessarily indicate that it will *generalise* to data that was not used during training. Therefore a full data set should consist of three separate parts — a *training* dataset used to update the network parameters during training, a *validation* dataset that is not seen during training but which can be used to optimise the network performance by tuning hyperparameters and a *test* dataset for a final check of the models' generalisation ability. The test dataset crucially cannot be used in *any way* when designing a model, including training or hyperparameter choices. Once a model has been trained to a satisfactory standard, the state of the model is frozen. During *inference*, the trained model is used for the intended application by executing a forward pass to obtain the desired outputs.

5.2 Network Architecture Components - Layers and Activation Functions

We begin by describing the basic components that will be used to build networks in this thesis.

5.2.1 Fully Connected Layers

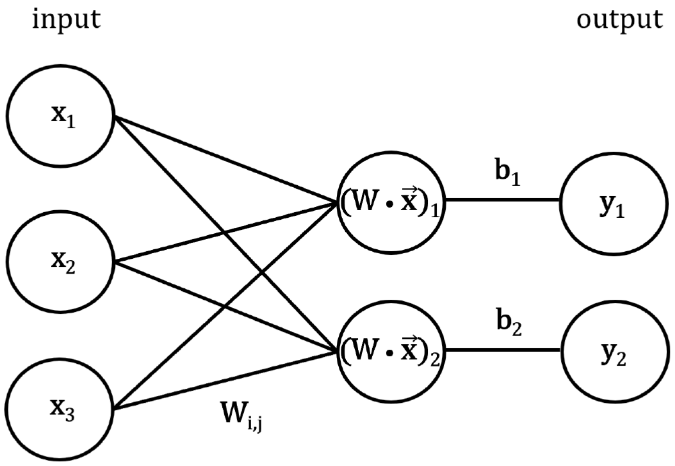


Figure 5.1: Illustration of the operation of a fully connected layer.

The simplest layer that can be used in an NN consists of a linear mapping (i.e. affine transformation) by means of a *weight* matrix $\mathbf{W}$ applied to an input vector $\vec{x}$, plus a constant *bias* vector $\vec{b}$, that produces an output y:

$$\vec{y} = \mathbf{W}\vec{x} + \vec{b}. \tag{5.1}$$

The parameters contained in $\mathbf{W}$ and $\vec{b}$ are *trainable parameters* that are learnt and updated during training. The operation of a fully connected layer is illustrated in Figure 5.1. When multiple such layers are combined, the nodes in a layer are connected to all nodes in the next layer by the respective weight given in the matrix $\mathbf{W}$. As such, these layers are often referred to as *fully connected* layers.

5.2.2 Convolutional Layers

Convolutional neural network layers are widely known to provide a performant base for building an algorithm operating on data with a regular grid-like structure (e.g. an image [145]). These layers have several advantages over fully connected layers for this type of data structure. These advantages arise from exploiting *weight sharing* and the *spatial symmetry* present in the data. The presence of weight sharing in the algorithm means that the number of trainable parameters does not depend on the number of inputs, so the number of parameters exhibits a much better scaling with the dimensions of the regular grid. This results in a more efficient algorithm. In the case of spatial symmetry, the network should ideally respond in a similar fashion to the presence of a given feature in the data independent of its position in the grid. This is known as *translational invariance* [146].

Convolutional layers operate by passing a *kernel* (or filter) over the grid. The kernel operates over a number of grid points, that is usually defined to be a subset of the overall size of the grid. This is called the *kernel size*. At each position of the kernel in the grid, the dot product is taken between each element in the input and the corresponding element of the kernel. These are then aggregated by means of a sum, producing a single scalar value for the given kernel position. Convolutional layers therefore act in a similar manner to that defined in Equation 5.1, however they use a convolution in the place of general matrix multiplication [140]. The output of a kernel once it has been passed over the grid is called a *feature map*. In practice, data often contains a vector of values at each point, which are called *channels*. A common example are the RGB pixel values in a colour image. For this reason, multiple independent convolutional filters are usually learned in parallel in a given layer.

Since only a subset of the total input is covered by the kernel, each node in a convolutional layer only receives information from a local region of the input. This is termed the *receptive field*. If multiple convolutional layers are applied one after the other, the effective receptive field on the input data gradually grows, as each kernel operation combines information from neighbouring grid points. This means that earlier layers in the network are restricted to learning local correlations, while later layers can learn correlations over larger length scales. This effect is illustrated in Figure 5.2.

The convolutional operation can be augmented to control the size of the feature map output by a layer. The *stride* defines the step size between grid positions where the kernel is applied. This means that the kernel is not applied to every position in the grid, and results in *downsampling* such that the size of the output feature map is reduced compared to the input size to the layer. If the size of the output needs to be kept the same as the input, *padding* can be applied to the data. This adds values around the edge of the grid, to allow the kernel to be applied to the very edge points of the grid. Typically, *zero padding* is used, which is the addition of zeros. Convolutional layers can be applied to regular grid-like data over varying dimensionality. In this thesis, $1\mathrm{D}$ and $3\mathrm{D}$ convolutions will be used.

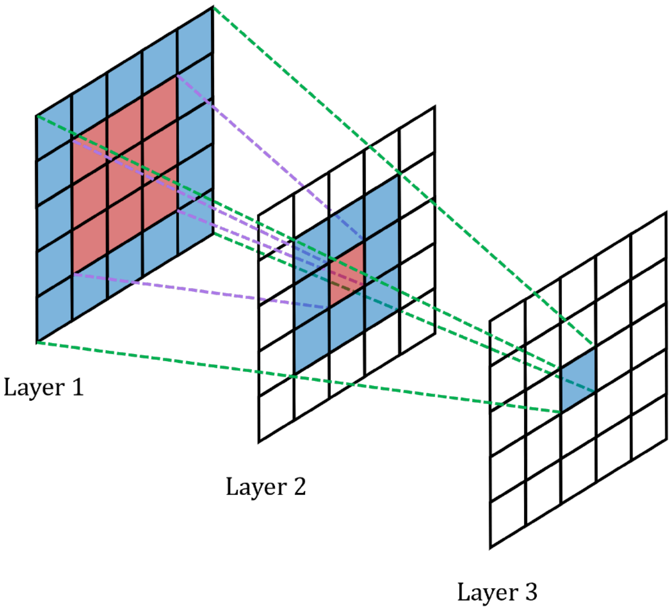


Figure 5.2: Illustration of the effect of applying multiple consecutive convolutional layers. The effective receptive field grows the deeper the layer is in the network.

5.2.3 Transpose Convolutional Layers

A transpose convolution, initially introduced in [147], can be used to perform the reverse operation to that of the standard convolution described in Section 5.2.2. While the standard convolutional operation can be applied with a stride larger than one for downsampling an image, a transpose convolution with a stride larger than one *upsamples* an image, meaning that the output of the network layer has a larger size than the input. The transpose convolution operates by using a kernel to produce an output segment that is the size of the kernel for each position in the regular grid. Multiple output segments are produced by passing the kernel across each grid position. Each of these intermediate segments are then overlaid to produce the final output [146]. This thesis will make use of 3D transpose convolutional layers.

5.2.4 Normalisation Layers

Parameters in the layers of a neural network can potentially have values that vary over several orders of magnitude. Normalisation layers were therefore introduced as a means to decouple parameter scales from the loss [148]. This amounts to a reparameterisation applied to network layers. *Batch normalisation* [149] calculates the mean μ_B and variance var_B over a batch of data x, and applies the transformation

$$x' = \gamma \frac{x - \mu_B}{\sqrt{var_B} + \epsilon} + \beta. \tag{5.2}$$

Here, γ and β are trainable parameters and ϵ is introduced to prevent division by zero [146]. During inference, the mean and variance are calculated using the population statistics, rather than batch statistics. Batch normalisation layers are typically interleaved between network layers, to maintain normalisation for the input of each layer.

An alternate approach is to use *layer normalisation* [150]. This applies the same transformation given in Equation 5.2, but replaces the mean μ_B and variance var_B, which were previously calculated over the batch, with the mean μ_I and variance var_I over the input to the layer. This removes the dependence on the statistics of a batch and does not require separate evaluation during inference.

5.2.5 Non-linearities

Thus far, the network layers discussed have applied linear mappings. However, a key part of a networks' expressiveness, and ability to model complex, nonlinear functions, is the use of *activation functions*. These are nonlinear mappings that are applied to the output of a linear layer, such as that defined in Equation 5.1, that modify the function to be inherently nonlinear. The important activation functions used in this thesis are defined as follows.

The *Rectified Linear Unit (ReLU)* [151] [152] is the most popular activation function. It maps negative values to zero, and applies the identity transformation for positive values. It is defined as

$$\mathrm{ReLU}(x) = \max(x, 0). \tag{5.3}$$

The main disadvantage of this function is that zero gradients for negative input values can produce 'dead' neurons. This means that no gradients (see Section 5.3) are backpropagated through the neuron, resulting in a reduction in the capacity of the network. The *leaky ReLU* [153] function seeks to solve this problem by introducing a slight slope for negative input values, the gradient of which is controlled by a fixed (i.e. non-trainable) parameter α. The leaky ReLU is defined as

$$\mathrm{LeakyReLU}(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0. \end{cases} \tag{5.4}$$

In some applications, the output of a network must lie between zero and one. A *sigmoid* activation can be used in these cases, which is defined as

$$\mathrm{sigmoid}(x) = \frac{1}{1 + \exp(-x)}. \tag{5.5}$$

5.3 Learning and Optimisation Algorithms

In Section 5.2, the basic components of neural networks that will be employed in this work were outlined. However, in order to enable networks to be able to learn from data, the parameters of a network must be modified depending on how the network output compares to a specific target. This requires two components:

- A target for the optimisation L — this is the loss function defined in Section 5.1. The loss of an architecture is minimised in order to improve the performance of a network. Since these functions are highly task specific, the loss functions used in this thesis will be discussed in Section 5.4.
- A method for improving the network's performance on this optimisation target. This can be achieved by descending gradients of the loss function. The process of calculating these gradients is called *backpropagation* and will be discussed in Section 5.3.1. How these gradients are actually used for learning is addressed by the *optimisation* proceedure. The basics of optimisation will be discussed in Sections 5.3.2 and 5.3.3, while the algorithm that will be employed for optimisation in this thesis will be discussed in Section 5.3.4.

5.3.1 Backpropagation

Forward propagation of information from an input x through each layer of a network produces a network output y. This can be used to compute the loss defined for the problem L. In order to filter information backward through the network, backpropagation performs an iterative procedure starting from the final output layer of the network and working back to the input layer. The aim is then to compute, in an efficient manner, the gradient of L with respect to the ith parameter of the network θ_i,

$$\nabla_{\theta,i} L = \frac{\partial L}{\partial \theta_i}. \tag{5.6}$$

The key component of backpropagation is the *chain rule* from calculus.

To illustrate this, suppose we have a single neuron performing a simple linear mapping of input x with weight w and bias b, as defined in Equation 5.1, to produce an output z_1, followed by a non-linearity σ (see Section 5.2.5), to produce an output z_2. Let the loss be some function of z_2, $L(z_2)$ [146]. The forward propagation is then defined as

$$z_1 = w \cdot x + b \tag{5.7}$$

$$z_2 = \sigma(z_1) \tag{5.8}$$

The gradient with respect to w is then given by

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial w} \tag{5.9}$$

By recursively applying the chain rule in this fashion, $\frac{\partial L}{\partial \theta_i}$ can be calculated. Having calculated the gradients, they can then be used to optimise the model.

5.3.2 Gradient Descent and Stochastic Gradient Descent

Having calculated the gradient $\nabla_{\vec{\theta}} L(\vec{\theta})$, which defines the direction of the steepest gradient in L, the simplest approach to optimisation is to move in the direction of the negative of gradient (i.e. the gradient of steepest descent). For a training dataset consisting of n examples, this means updating the parameters $\vec{\theta}$ to new values $\vec{\theta'}$ by

$$\vec{\theta'} = \vec{\theta} - \epsilon \nabla_{\vec{\theta}} L(\vec{\theta}) = \vec{\theta} - \frac{\epsilon}{n} \sum_{i=1}^{n} \nabla_{\vec{\theta}} L_i(\vec{\theta}). \tag{5.10}$$

This algorithm is therefore called *gradient descent*. The parameter ϵ is a positive scalar value called the *learning rate*, which determines the size of the step taken. The learning rate therefore has a significant impact on the performance of the optimisation. If ϵ is too large, the step might take parameters past a global minimum, whereas if ϵ is too small, the convergence of the training will be slow and there is a risk of the algorithm becoming stuck in a local minimum [140]. More sophisticated optimisation algorithms therefore typically seek to adjust the learning rate during training by using information from the previous steps in the optimisation [146]. Strategies to deal with this will be discussed in Section 5.3.3.

Gradient descent relies on calculating the gradient across the entire training dataset. Equation 5.10 implies that the computational cost therefore scales linearly with the size of the training dataset for each independent variable. This means that gradient descent does not scale well to models with a large number of trainable parameters and for large datasets. For this reason *stochastic gradient descent* (SGD) [154] has become the standard approach. Instead of using the full training dataset, a batch of m samples is drawn from the training set $\{x_1 ... x_m\}$. This means that equation 5.10 is now computed on a single sample

$$\vec{\theta'} = \vec{\theta} - \epsilon \nabla_{\vec{\theta}} L_i(\vec{\theta}). \tag{5.11}$$

This can then be used to estimate the gradient by

$$\mathbb{E}\left[\nabla L_i(\vec{\theta})\right] = \frac{1}{m} \sum_{i=1}^{m} \nabla L_i(\vec{\theta}). \tag{5.12}$$

This therefore additionally introduces a larger variance between gradients than using the complete dataset. This noise serves to help to prevent the optimisation from becoming stuck in local minima. However, it also means that the optimiser tends to bounce around a minimum. For this reason, SGD is typically used with a learning rate that is adjusted over time [140].

5.3.3 Momentum and Adaptive Learning Rates

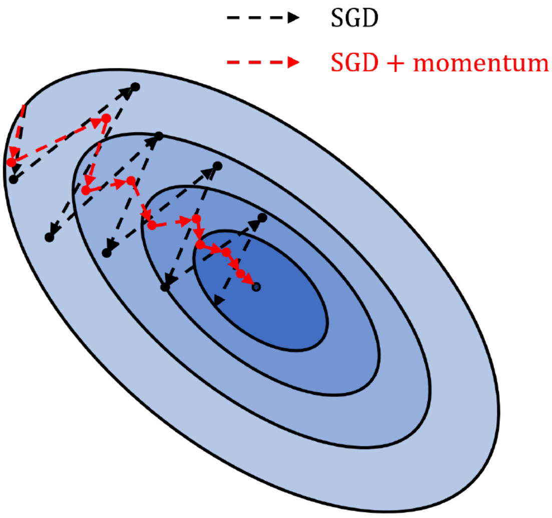


Figure 5.3: Illustration of the differences between SGD (black arrows) and SGD with momentum (red arrows). The optimisation seeks to descend a loss function, illustrated by the coloured contours. SGD tends to take large steps and bounces between the sides of the valley, whereas the addition of momentum helps to guide the optimisation to the minimum quicker and in a smoother fashion.

Several strategies have been employed to handle the potential pitfalls created by using a static learning rate. A common approach is to decrease the learning rate during training by means of *learning rate scheduling*. This can be achieved by using a learning rate *decay*, which reduces the learning rate in a manner that is pre-defined before training. Another approach is to introduce a so called *momentum* to the optimisation [155]. This makes use of previous steps in the optimisation to introduce a variable $\vec{v}$ that plays the role of the velocity at a step t. This modifies the SGD update step in Equation 5.11 to

$$\vec{v}' = \alpha\vec{v} - \epsilon\nabla L_i(\vec{\theta}), \tag{5.13}$$

$$\vec{\theta}' = \vec{\theta} + \vec{v}'. \tag{5.14}$$

The factor α is a hyperparameter that balances the relative contribution to the update from the previous step [140]. This can help to stabilise and speed up the optimisation by damping oscillations in the trajectory through the parameter space. Figure 5.3 illustrates this effect by comparing SGD to SGD with momentum.

The disadvantage of these approaches is that they depend on the specifics of the hyperparameter configuration, and hence the best schedule to use can be significantly different depending on the task, dataset and model. Therefore some algorithms introduce *adaptive learning rates*. Prominent examples include AdaGrad [156], which scales the learning rates for all model parameters inversely with the sum of the gradients of all previous steps in quadrature, and RMSProp [157], which uses an exponential moving average of gradients to scale the learning rate.

5.3.4 ADAM: Adaptive Moment Estimation

The ADAM algorithm [158] is the optimisation algorithm that will be employed in this thesis. It combines the components of RMSProp and momentum with SGD. This involves taking exponential moving averages of the first and second moments of the gradient

$$\vec{m'} = \beta_1 \vec{m} + (1 - \beta_1)\nabla L_i(\vec{\theta}) \tag{5.15}$$

$$\vec{s'} = \beta_2 \vec{s} + (1 - \beta_2)[\nabla L_i(\vec{\theta})]^2. \tag{5.16}$$

The decay of the moving averages of the two terms are controlled by β_1 and β_2 respectively. Typically an initialisation of $\vec{m_0} = \vec{s_0} = \vec{0}$ is applied, which introduces a bias that is corrected for by applying

$$\hat{\vec{m}} = \frac{\vec{m'}}{1 - \beta_1} \tag{5.17}$$

$$\hat{\vec{s}} = \frac{\vec{s'}}{1 - \beta_2}. \tag{5.18}$$

The parameters can then stepped according to

$$\vec{v'} = -\epsilon \frac{\hat{\vec{m}}}{\sqrt{\hat{\vec{s}}} + \delta} \tag{5.19}$$

$$\vec{\theta'} = \vec{\theta} + \vec{v'}, \tag{5.20}$$

where the parameter $\delta > 0$ prevents division by zero or steps that are too large. The combination of momentum and adaptive learning rates allows ADAM to converge quickly, and adapt the step size taken based on the fluctuations in gradient updates, making it rather robust to the selection of hyperparameters [140].

5.4 Generative Models

A class of machine learning model that has been developing rapidly in recent times is the so called *generative model*. These models aim to learn the underlying distribution $p(x)$ for a training dataset x, such that samples can later be drawn from the model, thereby enabling the generation of new data samples. In practice, this is achieved by the generative model learning a mapping from some form of noise onto the data space. This task is highly non-trivial and the exact components of the architecture and training procedure depend heavily on the chosen approach.

In this Section, focus will be restricted to the network architectures relevant to this thesis. In Section 5.4.1 the Variational Autoencoder will be described, followed by an overview of Generative Adversarial Networks in Section 5.4.2. In Section 5.4.3 the Bounded Information Bottleneck Autoencoder, which incorporates components from a number of different generative models, will be described. This model with be the main generative model that will be studied in this thesis. Section 5.4.4 will describe Normalising Flows, a fully invertible generative model. Finally, in Section 5.4.5, conditioning of these various models will be discussed.

5.4.1 Variational Autoencoders

The *Variational Autoencoder* (VAE) [159] [160] is a generative model that inherits directly from autoencoder architectures which are widely used for numerous unsupervised tasks [140]. An autoencoder consists of two key components — an *encoder* network that performs a mapping of input data sample $x \sim p(x)$ onto a lower dimensional latent representation z, and a *decoder* network which, attempts to reconstruct samples $\tilde{x}$ from z. The loss function minimised during the training of an autoencoder is then the *reconstruction* loss which is often chosen to be a *Mean Squared Error* (MSE), defined as

$$L_{\mathrm{MSE}} = \mathbb{E}\left[(x - \tilde{x})^2\right] = \frac{1}{N}\sum_{i=1}^{N}(x_i - \tilde{x_i})^2, \tag{5.21}$$

for an N dimensional real input.

However, if we wish to be able to later use the model for generation, the distribution of the intermediate representation $p(z)$ must be known so that samples can be drawn and passed to the decoder for generation. Using just the MSE loss in Equation 5.21 does not provide any such restriction. Therefore the loss of a general VAE[1] contains an additional term to constrain the latent representation, and is given by

$$\begin{aligned} L_{\mathrm{VAE}} &= L_{\mathrm{MSE}}(x, \tilde{x}) + \beta \cdot \mathcal{D}_{KL}[q(z|x)||p(z)] & (5.22)\\ &= L_{\mathrm{MSE}}(x, \tilde{x}) + \frac{\beta}{M}\sum_{x}\sum_{z} q(z|x)\log\left[\frac{p(z)}{q(z|x)}\right], & (5.23)\end{aligned}$$

where M is the batch size, β is a hyperparameter and $\mathcal{D}_{KL}$ denotes the Kullback-Leibler divergence (KLD), which provides a measure of the similarity between $q(z|x)$ and $p(z)$ [146].

The Kullback-Leibler divergence between probability distributions p and q is defined as

$$\mathcal{D}_{KL}(p||q) = \sum_{x} p(x)\log\left(\frac{p(x)}{q(x)}\right). \tag{5.24}$$

The encoder provides the mapping denoted by the variational posterior $q(z|x)$, which approximates the true posterior $p(z|x)$, while the conditional likelihood $p(x|z)$ represents the task of the decoder. The latent space distribution $p(z)$ is usually chosen to be an uncorrelated multivariate Gaussian. This can be achieved by producing two vectors (with size equal to that of the latent space), the first $\vec{\mu}$ describing the means of the Gaussians and the second $\vec{\sigma}$ their standard deviations.

However, in order to allow for a stochastic sampling process in the model without increasing the variance of the gradient too much during backpropagation [162], a so called *reparameterisation trick* has to be employed to obtain the vector encoding $\vec{z}$

$$\vec{z} = \vec{\mu} + \vec{\sigma} \cdot \vec{\epsilon}, \tag{5.25}$$

with the independent random variable $\epsilon \in \mathcal{N}(\mathbf{0}, \mathbb{1})$. It is therefore $\vec{z}$ that is input to the decoder.

With the choice of an uncorrelated multivariate Gaussian for $p(z)$, this means that Equations 5.22 and 5.23 reduce to

$$\begin{aligned} L_{\mathrm{VAE}} &= L_{\mathrm{MSE}}(x, \tilde{x}) + \beta \cdot \mathcal{D}_{KL}[\mathcal{N}(\mu, \sigma)||\mathcal{N}(\mathbf{0}, \mathbb{1})] & (5.26)\\ &= L_{\mathrm{MSE}}(x, \tilde{x}) + \frac{\beta}{M}\sum_{x}\frac{1}{2}(1 + \log(\sigma) - \mu^2 - \sigma^2) & (5.27)\end{aligned}$$

While the MSE component of the loss governs the reconstruction quality, the second term represents the regularisation of the latent space towards a multivariate Gaussian. This trade off, governed by β, will be explored in more detail in Section 5.4.3.

---

[1] The definition given here is technically that of the beta-VAE [161], for which the term VAE will be used synonymously.

5.4.2 Generative Adversarial Networks

Another popular class of generative model is the *Generative Adversarial Network* (GAN) [163]. This model consists of a generator network G with parameters θ which is trained in competition with a discriminator network D with parameters ϕ. The generator G tries to learn a mapping $G_\theta(z)$ from random noise variables $z \sim p(z)$ onto the target distribution of the training data $x \sim p(x)$. The discriminator D acts as a binary classifier which tries to distinguish real data x from the generator output $\tilde{x} = G_\theta(z)$, by outputting a probability $D_\phi(x)$ of a given sample being real data.

The training follows an *adversarial approach*, in which D tries to *maximise* the probability of a correct classification by means of maximising the *cross entropy*

$$L_D = \max_D \mathbb{E}_{x \sim p_{data}}[\log(D_\phi(x))] + \mathbb{E}_{z \sim p(z)}[\log(1 - D_\phi(G_\theta(z)))]. \tag{5.28}$$

The feedback from the discriminator can then be used to train the generator, which seeks to fool D by *minimising* the probability of D (with parameters ϕ fixed) performing the classification correctly,

$$L_G = \min_G \mathbb{E}_{z \sim p(z)}[\log(1 - D_\phi(G_\theta(z)))]. \tag{5.29}$$

The combination of the two losses in Equations 5.28 and 5.29, gives the loss of the GAN

$$L_{\text{GAN}} = \min_G \max_D \mathbb{E}_{x \sim p_{data}}[\log(D_\phi(x))] + \mathbb{E}_{z \sim p(z)}[\log(1 - D_\phi(G_\theta(z)))]. \tag{5.30}$$

From a game theoretic perspective, this amounts to zero sum game. This permits an optimal solution to be reached in an idealised scenario called a *Nash Equilibrium* [164]. In the case of a GAN, this corresponds to G producing samples so similar to the training data that D can no longer distinguish generated data from training data, resulting in $D_\phi = 0.5$ i.e. D randomly guessing whether a given sample is real or generated.

In reality, the inherent fluctuations induced by the adversarial nature of the training tend to make GAN trainings rather unstable, and achieving convergence can be challenging.

Wasserstein GANs

One popular modification to the traditional GAN is the *Wasserstein GAN* (WGAN) [165]. This approach adopts the *Wasserstein-1 distance* metric between the training data distribution and the generated distribution, as the target for optimisation. This metric is rooted in optimal transport theory, and is commonly referred to as the Earth Mover's Distance due to an intuitive interpretation of the metric as the minimal cost to transform one mass of earth into another [146]. The Kantorovich-Rubinstein duality [166] can be used to write the Wasserstein distance between distributions p_x and p_y in a tractable form

$$W(p_x, p_y) = \sup_{||f||_L \leq 1} \mathbb{E}_{x \sim p_x}[f(x)] - \mathbb{E}_{y \sim p_y}[f(y)], \tag{5.31}$$

where the supremum runs over all 1-Lipschitz functions. This means that rather than the binary classification task given to the discriminator in the original GAN, the adversarial network in the WGAN, called a *critic* C, now attempts to estimate the Wasserstein distance between the real data p_{data} distribution and the generated distribution p_g. Hence, the loss function for the Wasserstein GAN with Generator G can be written as

$$L_{\text{WGAN}} = \min_G \max_{C \in \mathcal{C}} \mathbb{E}_{x \sim p_{data}}[C_\phi(x)] - \mathbb{E}_{z \sim p(z)}[C_\phi(G_\theta(z))], \tag{5.32}$$

where $C \in \mathcal{C}$ denotes that the critic is required to be a 1-Lipschitz function. This constraint requires the norm of the gradients of a differentiable function to be less than or equal to one everywhere. Practically, this can be enforced by adding a so called *gradient penalty* term [167] to the loss function of the critic. This term is calculated by defining a new distribution $l \sim p_l$, which is created by random sampling uniformly along straight lines connecting generated samples $\tilde{x}$ and data samples x [146],

$$l = \epsilon x + (1 - \epsilon)\tilde{x}, \tag{5.33}$$

with $(0 \leq \epsilon \leq 1)$. The critic gradient with respect to samples from p_l can then be calculated. This results in a critic loss term for the WGAN of

$$L_C = \mathbb{E}_{z \sim p(z)}[C_\phi(G_\theta(z))] - \mathbb{E}_{x \sim p_{data}}[C_\phi(x)] + \lambda \mathbb{E}_{l \sim p_l}[(||\nabla_l C_\phi(l)||_2 - 1)^2]. \tag{5.34}$$

The third term in Equation 5.34 is the additional gradient penalty term, with the penalty coefficient λ acting as the hyperparameter for scaling the gradient penalty. This is typically set to a value of 10 [167]. In this thesis, the WGAN Critic with gradient penalty will be used.

5.4.3 The Bounded Information Bottleneck Autoencoder

The Information Bottleneck Principle

The Information Bottleneck (IB) Principle [168] was formulated to provide a theoretical means by which the information from a given random variable $X \in \mathcal{X}$, relevant for the prediction of another random variable $Y \in \mathcal{Y}$, could be extracted. It has since played an important role in approaches seeking to gain an information-theoretical understanding of Deep Learning [169].

The IB principle is defined in terms of *mutual information*, which provides a measure of the amount of information that one random variable contains about another. It is defined between random variables X and Y, jointly distributed by $p(x, y)$ and with marginals $p(x)$ and $p(y)$ respectively, as [170]

$$I(X; Y) = \mathbb{E}_{p(x,y)}\left[\log\left(\frac{p(x, y)}{p(x)p(y)}\right)\right]. \tag{5.35}$$

In the framework of the IB, a purely unsupervised (U) task in which no labels are provided can be considered to be a compression from the data space X to a lower dimensional latent space Z, which seeks to maximise the mutual information I between Z and X, while minimizing information irrelevant for the task. The problem can be formulated as a minimization of

$$L^U(\phi) = I_\phi(X; Z) - \beta I(Z; X) \tag{5.36}$$

over model parameters ϕ, with Lagrange multiplier β parameterising the trade-off between compression and retention of useful information, which forms the central part of the optimisation. This kind of problem can be seen explicitly in a generative autoencoder such as the VAE discussed in Section 5.4.1, which features an intermediate bottleneck representation.

The Bounded Information Bottleneck Autoencoder

The *Bounded Information Bottleneck Autoencoder* (BIB-AE), first proposed in [171], provides a theoretical framework that unifies many of the features present in common GAN and VAE architectures. The BIB-AE optimisation target is defined as

$$L_{\text{BIB-AE}}(\theta, \phi) = I_\phi(X; Z) - \beta I_{\theta,\phi}(Z; X), \tag{5.37}$$

where ϕ are the encoder parameters, and θ are the decoder parameters. By means of a variational decomposition [171], the BIB-AE loss can be written as

$$
\begin{aligned}
L_{\text{BIB-AE}}(\theta,\phi) =& \underbrace{\mathbb{E}_{p_{data}}\Big[\mathcal{D}_{KL}(q_\phi(z|X=x)||p_\theta(z))\Big]}_{\text{A}} - \underbrace{\mathcal{D}_{KL}(q_\phi(z)||p_\theta(z))}_{\text{B}} \\
&-\beta\Bigg[\underbrace{\mathbb{E}_{p_{data}}\Big[\mathbb{E}_{q_\phi(z|x)}[\log(p_\theta(x|z))]\Big]}_{\text{C}} - \underbrace{\mathcal{D}_{KL}(p_{data}(x)||p_\theta(x))}_{\text{D}}\Bigg],
\end{aligned} \tag{5.38}
$$

where data samples x are distributed according to $p_{data}(x)$, latent samples z are distributed according to $p(z)$ and, as in Section 5.4.1, the encoder task is represented by $q_\phi(z|x)$ and the decoder task is represented by $p_\theta(x|z)$.

Each of the four components denoted in Equation 5.38 serve a particular purpose. The first two terms (A) and (B) drive the the distribution of the encoded latent space to match the target distribution $p_\theta(z)$, thereby reducing the information contained in the latent space. The second two terms (C) and (D) ensure that the fidelity of the reconstructed samples $\tilde{x}$ is high, i.e. that they are distributed in a similar manner to the training data $x \sim p_{data}$. For this to be achieved, sufficient information must be contained in the latent space. The optimisation of the BIB-AE is therefore a balance between these two opposing objectives. The interpretation of the two components in each of the two objectives are similar. Components (B) and (D) both rely on a sampling based method, such as from adversarial feedback or a kernel-based Maximum Mean Discrepency loss (see Section 7.3.1) in order to approximate the KLD — in the first case between encoded latent space (q_ϕ) distribution and the target distribution ($p_\theta(z)$), and in the second case between the distribution reconstructed by the decoder ($p_\theta(x)$) and the data distribution ($p_{data}(x)$). In contrast, components (A) and (C) rely on an explicit computation of either the KLD in the first case or the log-likelihood in the second.

This allows for the interpretation of numerous generative models by means of these components. For example, the GAN described in Section 5.4.2 uses only component (D) in the form of the adversarial feedback from the discriminator/critic, while the VAE described in Section 5.4.1 makes use of component (A) for the KLD-based latent space constraint, and component (C) through the MSE reconstruction loss.

5.4.4 Normalising Flows

Normalising Flows (NFs) [172] [173] are a type of generative model that aim to learn a series of invertible mappings that allow the transformation of a simple distribution into a much more complicated distribution and back. This can be considered to be a special case of a VAE, with the constraint that the decoder applies exactly the inverse mapping to that of the encoder. Additionally, while a VAE usually encodes information into a lower dimensional latent space, NFs operate on a *base* distribution that has the same dimensionality as the data [146].

This base random variable $z \in Z$ has a known and tractable probability density function $p_z(z)$, while the target random variable $x \in X$ has some unknown probability density function $p_x(x)$. NFs then aim to learn a bijective mapping $X = g(Z)$ between the two, with inverse $f := g^{-1}$. Under the change of variables formula, $p_x(x)$ can then be computed from $p_z(z)$ via

$$
\begin{aligned}
p_x(x) &= p_z(f(x))\left|\det\frac{\partial f(x)}{\partial x}\right| \\
&= p_z(z)\left|\det\frac{\partial g(z)}{\partial z}\right|^{-1},
\end{aligned} \tag{5.39}
$$

where $\frac{\partial g(z)}{\partial z}$ is the Jacobian of g, J_g and $\frac{\partial f(x)}{\partial x}$ is the Jacobian of f, J_f. In practice, designing an arbitrarily complex non-linear function which is also invertible is highly non-trivial [173].

Since it is desirable for a network to be deep and hence consist of multiple layers, another important property arises from considering a set of N bijective functions $\{g_1 ... g_N\}$, each with inverse $\{f_1 ... f_N\}$. The composition g

$$g = g_N \circ g_{N-1} \circ ... \circ g_1, \tag{5.40}$$

is also bijective and has inverse

$$f = f_1 \circ ... \circ f_{N-1} \circ f_N. \tag{5.41}$$

Additionally, the determinant of the Jacobian of f can be calculated by taking the product of the individual Jacobian determinants

$$\det J_f = \prod_{i=1}^{N} \det J_{f,i}. \tag{5.42}$$

This means that multiple invertible layers can be stacked together to form an NF, which will result in a network that still obeys Equation 5.39.

An important consequence [162] of Equation 5.39 is that the logarithm of $p_x(x)$ is directly calculable for the sequence of transformations $\{g_1 ... g_N\}$ starting from the base distribution $p_z(z)$ via

$$\log p_x(x) = \log p_z(z) - \sum_{i=1}^{N} \log \left| \det J_{g_i,(i-1)} \right|. \tag{5.43}$$

This means that NFs can directly calculate the negative log-likelihood, allowing for an explicit optimisation, rather than an implicit one that is often present in other generative models e.g. GANs.

Components of Normalising Flows

Designing layers of NFs is a non-trivial task, as they must be invertible (which is not generally the case of a standard network layer), expressive enough to model a given distribution, and computationally efficient [173]. While there are a number of different approaches, this thesis will make use of the *coupling layer* [174].

A coupling layer operates by first splitting the d-dimensional input $x \in \mathbb{R}^d$ into two parts x_a and x_b. The most common method of dividing the input is to split it into two halves. A number of different transformations can then be applied. A simple example is the *affine* coupling layer. In the case of an affine coupling layer, the transformation g in the *forward* direction is,

$$y_a = x_a \tag{5.44}$$

$$y_b = \exp(s(x_a)) \odot x_b + t(x_a), \tag{5.45}$$

where s and t are arbitrary neural networks (for example a fully connected network) that apply *scaling* and *transition* respectively, and $\odot$ represents element-wise multiplication [162]. In the *backwards* direction, the inverse transformation g^{-1} can then be applied

$$x_b = (y_b - t(y_a)) \odot \exp(-s(y_a)) \tag{5.46}$$

$$x_a = y_a. \tag{5.47}$$

Note that since s and t are always applied in the same manner in both the forward and backward pass, they do not need to be invertible [146]. Multiple coupling layers can then be applied one after the other. However, since only half of the input is processed in g, a *permutation layer* is normally applied after each coupling layer. Since this permutation is applied in a

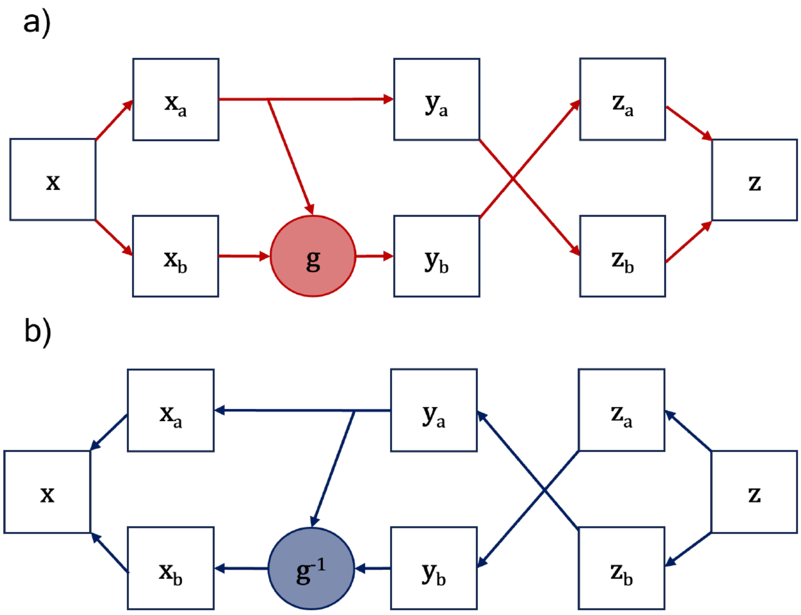


Figure 5.4: Illustration of the operation of an invertible block comprising a coupling layer and a permutation layer in a) the forward pass and b) the backward pass.

fixed fashion, it can be inverted to recover the original input. Additionally, a permutation is *volume preserving* (i.e. its Jacobian-determinant is one), so it can be used repeatedly [162]. An example of an invertible layer composed of a coupling layer and permutation layer applied in the forward and backward direction is shown in Figure 5.4

A more complicated function than an affine transformation that can be used in a coupling layer is an *spline* [175], which is a piece-wise function specified by points (knots) through which the spline passes. A *rational* spline [176] is a function f comprised of segments formed from the ratio between two polynomials. As an example, the ith segment of the spline in the dth dimension will take the functional form

$$f_d = \frac{p_a^{(i)}(x_d)}{p_b^{(i)}(x_d)}, \tag{5.48}$$

where $p_a^{(i)}$ and $p_b^{(i)}$ are order d polynomials with parameters (i.e. locations of the points and their derivatives) modelled by the output of a neural network. The spline is linear if $d = 1$, and quadratic if $d = 2$. Each segment is constructed over its respective interval, and the function is the identity mapping otherwise [173].

5.4.5 Conditioning of Generative Models

In this work, it will be essential to be able to generate data which displays particular attributes. Consider a training dataset consisting of training objects $x \sim p_{data}(x)$, whose distribution we wish to learn with a generative model G_θ, in order to be able to generate new samples $\tilde{x}$. In addition, each sample also has a vector of labels l, resulting in a set of n observations $\{(x_n, l_n)\}$. The task is then to generate random samples which all correspond to a subset of labels l_i, i.e. to use the *conditional* distribution $p_\theta(x|l)$. For this reason, providing labels to augment the generative task is called *conditioning*. Conditional generation can then be performed by passing the vector of desired labels l_i to the generator in addition to the random noise z, $G(z, l_i)$.

Adapting each of the generative models discussed so far to be conditional is fairly straightforward. In the case of the *VAE* described in Section 5.4.1, the conditioning label is provided together with the training data to the encoder, and together with the encoded latent representation to the decoder. The conditioning label is provided explicitly to the decoder in order to create a robust conditioning scheme during later sampling.

In the case of the *GAN* and *WGAN* discussed in Section 5.4.2, the generator can be passed the label by concatenation with the noise vector used for generation, while the discriminator/critic is passed the label in addition to the training data sample. In this manner, the discriminator is able to learn to correlate a given training sample with its corresponding label. This means that if the generator produces a sample which does not fit a given label, the discriminator/critic will be able to recognise the discrepancy. The addition of the label to the generator input allows the generator to make use of the adversarial feedback from the discriminator/critic by connecting it to the label.

Since the *BIB-AE* discussed in Section 5.4.3 features components from both the VAE and the GAN/WGAN, the conditioning of the model inherits directly from those models. The autoencoder at the core of the BIB-AE is conditioned in the same manner as the VAE, with both the encoder and the decoder being provided the conditioning label as additional input. The adversarial elements of the BIB-AE are also provided the conditioning labels as in the GAN/WGAN.

Finally, in the case of *NFs* described in Section 5.4.4, the base distribution is passed through a series of bijections that are conditioned on the label, in order to create the more complex transformed distribution. A given normalising flow bijection g can be made conditional by using a *hypernetwork* [177], which is conditioned on the given labels, to generate the parameters of g. By providing the conditioning labels to the transformations in this manner during training, the same conditioning scheme can be used to generate conditional samples when mapping from the base distribution to new data during inference.

Chapter 6

Simulation in High Energy Physics

Simulation in high energy physics is broken down into a chain of separate simulation steps, which can broadly be divided into *event generation* and *detector simulation*. The first step of *event generation* is the simulation of the hard scattering process between the initial state particles. This is performed by building the matrix elements and integrating across the phase space to calculate the appropriate cross sections. Since these integrals are not analytically tractable, event generators make use of MC techniques. The partons (quarks and gluons) that are produced during the process must then be evolved through a process termed *parton showering*, whereby they are successively spilt into pairs of partons. This splitting emulates the radiation of gluons or quark-antiquark pairs from a particle with colour, and continues down to a given momentum transfer scale, after which point the splitting is terminated. However, the partons in the final state of the parton shower may not exist in isolation as a result of colour confinement, and undergo a processes termed *hadronisation*. This involves non-perturbative QCD, and hence phenomenological models are used to create the final state, colourless hadrons, as well as decay short lived resonances [178]. These final state colourless hadrons appear in the detector as a collimated stream of particles in a jet.

To model the physics conditions at the ILC, ILD uses the event generator WHIZARD [179] to generate final state partons and leptons, and the event generator PYTHIA [180] to model the hadronisation process. The beam conditions are simulated with the Guinea-Pig [181] package, which provides a dedicated simulation tool for beam-beam interactions at linear colliders [64]. Note that an important part of a complete simulation scheme is also the simulation of *background* processes. At the ILC, the dominant effects arise from beam-beam interactions. This includes beamstrahlung which affects the luminosity and produces a $\gamma\gamma \rightarrow hadrons$ background, and incoherent e^+e^- pairs. These are typically overlaid to the simulated event after the detector simulation of the hard process, and will thus not be considered further in this work.

The second step in the simulation chain in high energy physics is *detector simulation*, which will be the focus of this thesis. This simulation stage takes the output of the event generation as primary particles, which are then propagated through the detector with their interactions with the materials and fields in the detector being simulated. This results in the creation of hits which represent the energy deposited in the sensitive parts of the detector. ILD uses the state-of-the-art simulation toolkit GEANT4 [25], which is interfaced through the DD4HEP [129] geometry description toolkit.

Detector simulation can be accomplished in a number of ways. *Full simulation* involves a physics-based approach relying on MC techniques to track individual particles through the detector and record their interactions. The principles behind this approach and its implementation in GEANT4 will be described in Section 6.1.1. However, given the large amount of simulated data required by high energy physics experiments and the high computational cost of full simulation, *fast simulation* techniques are now wide spread. GEANT4 therefore additionally provides an interface for fast simulation models.

Classical approaches to fast simulation will be described in Section 6.2, with a particular emphasis placed on those designed for calorimeter simulation, which constitutes the most computationally intensive part of a full simulation of a detector. Subsequently, a review of approaches to fast calorimeter simulation based on generative models will be given in Section 6.3. Particular attention will be paid to models for simulating high granularity calorimeters, which are the principle focus of this work.

6.1 Particle Transport Monte Carlo

Particle transport Monte Carlo aims to simulate the passage of individual particles through matter, from their creation until they either are destroyed or leave the detector volume. This approach breaks down the simulation of the particles' propagation into individual *steps*, with a number of potential interactions considered at each step.

Consider the general case of a particle of energy E traversing a compound material with density ρ consisting of numerous different elements, where the ith element composes a fraction f_i of the mass. The number of atoms per unit volume of that element is

$$n_i = \frac{N_A \rho f_i}{A_{m,i}}, \tag{6.1}$$

where N_A is the Avagadro constant and $A_{m,i}$ is the molar mass of the ith element. The interaction of the particle via a given process is characterised by the *mean free path* $\lambda(E)$ [182], which can be written in terms of the total cross-section of the process per atom of element i, $\sigma_i(E)$, as

$$\lambda(E) = \frac{1}{\sum_i n_i \cdot \sigma_i(E)}. \tag{6.2}$$

In order to establish the interaction point for a particle passing through a medium which consists of numerous different materials, a quantity which is material independent is required. This requirement is fulfilled by using the *number* of mean free paths n_λ [182] that a particle traverses when moving a distance x

$$n_\lambda(x) = \int_0^x \frac{dx'}{\lambda(x')}. \tag{6.3}$$

The probability of an interaction occurring within a distance x is then given by the cumulative probability

$$c(x) = 1 - p(x) = 1 - \exp(-n_\lambda(x)). \tag{6.4}$$

n_λ can then be sampled [183] by drawing a random number η from a uniform distribution between 0 and 1

$$n_\lambda = -\log(\eta). \tag{6.5}$$

The step size $s(x) = n_\lambda \cdot \lambda(x)$ is then calculated for each possible process. The shortest step size is selected, with the particle undergoing the corresponding process during the given step [182].

In practice it is not just discrete processes such as this that have to be considered — in the case of Bremsstrahlung, for example, this is because of the large number of low energy secondaries produced [102]. In this case, below a selected energy threshold E_t the interactions are treated collectively with cross-section $\sigma_{continuous}$, and do not contribute to limiting the step, only altering the particle kinematics. Therefore only the interactions with energy E resulting in secondaries above threshold are simulated individually (i.e. treated discretely), resulting in cross-section

$$\sigma_{discrete} = \int_{E_t}^{\infty} \frac{d\sigma}{dE} dE, \tag{6.6}$$

with the total cross-section σ arising from the sum of the discrete and continuous contributions [184].

Additionally, particles often propagate in an external *field*. The most common form of external field used in a detector simulation is an electromagnetic field. In this case, the change in momentum $\vec{p}$ of a relativistic charged particle moving with velocity $\vec{v}$ in a vacuum through electric $\vec{E}$ and magnetic $\vec{B}$ fields will follow the equation of motion

$$\frac{d\vec{p}}{ds} = \frac{q}{v}(\vec{E} + \vec{v} \times \vec{B}), \tag{6.7}$$

where q is the charge of the particle and ds is the differential path-length. When propagating through a medium, this motion becomes more complicated because of the elastic and inelastic forces that are present in addition to the coupling to the external field [183].

6.1.1 The GEANT4 Simulation Toolkit

The Geant4 [25] [27] (Geometry and Tracking) simulation toolkit provides a state-of-the-art C++ framework for simulating particle interactions with matter. It shall be used extensively for detector simulation throughout this thesis. Geant4 provides a large set of physics models, which makes it capable of simulating physics processes across a wide range of energies. Additionally, the framework features a uniform interface to make the addition or alteration of physics models easier with minimal modification of the existing code base. The following description provides an overview of the salient features of the toolkit that will be relevant for this thesis.

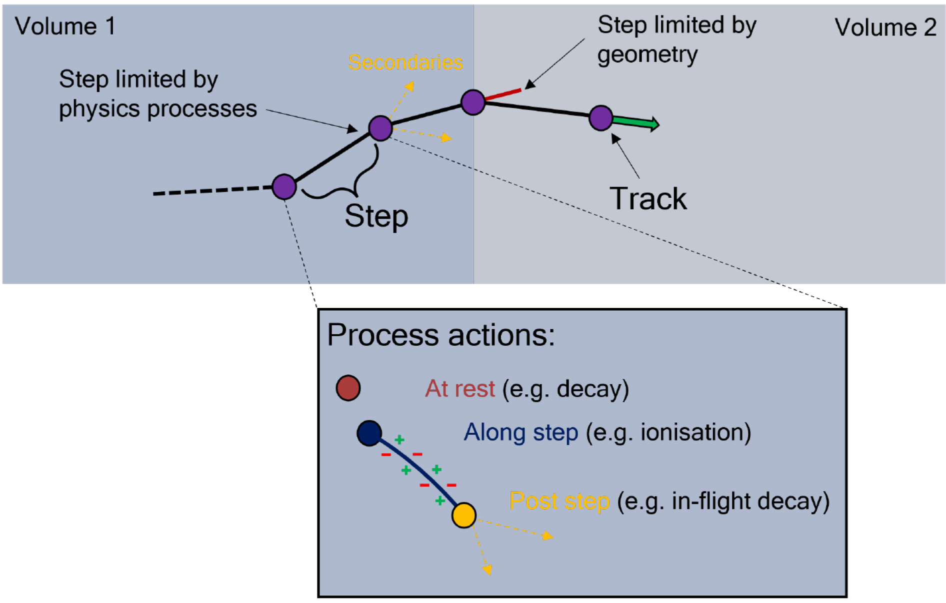


Figure 6.1: Illustration of the procedure of tracking in Geant4. A particle is moved one step at a time, with the track representing a snapshot of the particle state at each step. Steps are limited in size by a random competition between processes, to allow for the production of secondaries or if a boundary between geometry volumes is crossed. A process may be defined as a combination of three actions: at rest, along step and post step.

Simulations in Geant4 proceed in independently contained events, which prevents information being retained unnecessarily after the processing of an event is complete. The information held in an event can then be stored for downstream processing and reconstruction. It additionally provides a flexible interface to external event generators for the inputting of primary particles. This removes dependence on the specifics of a given event generator, and allows for the use of primary particles that cannot be handled with GEANT4.

The *geometry* of the detector through which particles propagate can be constructed from a hierarchy of volumes. A *logical* volume can contain other volumes, and holds information about the volume that is independent of its physical position, such as the material composition. The *physical* volume represents the placement of a logical volume within the mother volume that contains it [25]. Information about interactions is collected if they occur in a *sensitive* region, denoting an active element of the detector.

The definition of a *particle* in GEANT4 contains static information about the particle such as its name, mass, charge etc. GEANT4 typically only tracks particles with macroscopic lifetimes, with particles such as quarks and gauge bosons being handled by the generator. *Full simulation* in GEANT4 involves a detailed physics-based simulation procedure based on particle transport Monte-Carlo. At the core of this scheme is *tracking*. The manner in which tracking proceeds in GEANT4 is represented in Figure 6.1. A *track* represents a snapshot of the information relevant to a given particle for each step, such as the current position and momentum of the particle. During tracking, the track is moved in individual *steps*. The length of a step arises from a competition between *processes*, with the user being able to limit the maximum step size directly. An additional limit on the step size arises when particles from a secondary process need to be created, and also typically if a step crosses a boundary between geometry volumes. Each physics process has three allowed actions, which may be combined to formulate the process:

- *at rest*, which is used for particles at rest, such as in the case of particle decay
- *along step*, which describes a continuous interaction that occurs along the path of the particle, such as ionisation
- *post step*, which represents a point-like interaction at the end of a step, such as in-flight decay

The particles that will be used in a given simulation application and the allowed processes for each particle are defined in a *physics list*. These may include different models for interactions, with a transition between the models occurring across energy ranges. For electromagnetic physics, a standard electromagnetic package provides the complete set of processes for charged particles and photons down to 1keV. As discussed in Section 3.3, hadronic physics is significantly more complicated, and therefore difficult to fully describe with a single list. For this reason, several different hadronic physics models are available. In this thesis, the QGSP_BERT physics list will be used, which uses the Quark-Gluon String precompound model [185] for higher energies, and the Bertini Cascade model [186] for lower energy regions.

Since the full step-based simulation procedure described above involves tracking individual particles, in detectors where large numbers of particles are present in the simulation the computational cost can become significant. This is in particular true in a calorimeter, which relies on the production of many secondary particles as described in Section 3. Since the full simulation procedure also relies on the sequential production of secondaries, it cannot be easily parallelised. This motivates the use of parameterised models, which can provide a faster yet typically less accurate, emulation of the detector response. In GEANT4, a given detector region can be assigned a *parameterisation envelope*. This allows a user to terminate the full simulation of a particle which fulfils certain criteria (such as the type of particle and its energy). Information about the particle can then be passed to the relevant parameterisation model in order to emulate the response of the sub-detector in question. Once the output of the parameterisation model has been produced, it can be handed back to GEANT4 for placement back into the detector geometry. These can then be treated as if they were produced by the conventional full simulation procedure.

6.1.2 The DD4HEP Detector Description Toolkit

The DD4HEP detector description toolkit [129] provides a generic array of different tools which can be used coherently to provide a single source of information about a detector, in a manner which is sufficiently flexible to enable its use throughout the complete lifecycle of an experiment. Of particular relevance to this thesis is the DDG4 [130] component, which provides an interface to GEANT4.

This interface enables complex detector geometries to be implemented in a straightforward fashion in GEANT4, with an automatic mechanism provided for translating between geometry representations [187]. The sensitive regions of the detector can additionally be segmented in order to emulate the *readout geometry* of the detector, thereby allowing the information about physical interactions to be accumulated into *hits*. Hooks are also provided to allow the use of the fast simulation mechanism present in GEANT4.

6.2 Classical Approaches to Fast Simulation

One approach to fast simulation is to attempt to emulate the entire detector response and, at least partially, the reconstruction chain. Examples of such tools are DELPHES [188] and SGV [189], which has been used as a fast simulation tool for the ILD concept. In the case of DELPHES, a detector parameterisation is used which involves a smearing being applied to generator level quantities in order to emulate the detector response. In the case of the calorimeter component of the simulation, this smearing is used to model the energy resolution of the calorimeter, producing a single calorimeter 'object' (cluster). While SGV uses a significantly more advanced method for simulating tracking detectors based on the full covariance matrix of the track, the simulation of the calorimeters is similar. The main difference is that SGV can optionally provide a rough parameterisation of confusion in associating energy between clusters produced by charged and neutral particles (see Section 4.1). While these detector simulations are typically very fast, the fact that a reconstruction level output is produced for the calorimeters means that it cannot be used in conjunction with the real reconstruction chain of an experiment, since there are no hits produced.

An alternative approach to fast simulation is to directly target the calorimeter component of the simulation by attempting to simulate the hits in a shower. One method is to use a *shower library*. Such a library consists of a large number of showers simulated with GEANT4 for different particle types with various energies and incident positions on the calorimeter surface [33]. During the detector simulation, an appropriate shower for a given particle incident to the calorimeter is selected from the library. A slight modification to this approach was adopted by the ATLAS experiment [34], where full simulation was used for electromagnetic particles above a certain energy threshold, while a *frozen shower* from a library was used in place of full simulation for electromagnetic particles that fell below the threshold.

Another classical approach to fast calorimeter shower simulation is a *parameterisation* [31, 32]. For electromagnetic showers, the spatial distribution of energy E is broken down into three components

$$dE(\vec{r}) = Ef(t)dtf(r)drf(\phi)d\phi, \tag{6.8}$$

where the probability density functions $f(t)$, $f(r)$ and $f(\phi)$ represent the longitudinal, radial and azimuthal components respectively. The longitudinal shower depth t is represented in units of radiation length, while r denotes the radial distance perpendicular to the shower axis in units of Molière radii. The parameter ϕ describes the azimuthal angle around the shower axis, with the assumption that the energy is symmetrically distributed in ϕ. An additional distribution must be folded in to emulate the sampling fluctuations in a sampling calorimeter. An implementation of this parameterisation for calorimeter shower simulation called GFLASH is available in GEANT4 [31].

6.3 Generative Models for Fast Calorimeter Simulation

Following the seminal work of Paganini et al. [44, 53], which proposed the use of GANs for fast calorimeter simulation, a multitude of advances have been made in the application of generative models to calorimeter simulation. The vast majority of this work has focused on exploring various different generative models for this task, including GANs and their variants [57–62], VAEs and their variants [4, 45–47, 63] and Normalising Flows [48–51]. Towards the end of the work presented in this thesis, Diffusion Models [6, 7, 52, 54] have also begun to be explored. The field is now sufficiently advanced that the ATLAS experiment has deployed a GAN-based model as part of their fast simulation chain for the detector in its current form [55]. This tool involves training 300 GANs, one for each particle type and regions of the detector in bins of pseudorapidity η.

However, as described in Chapter 3, several upcoming experiments plan to use calorimeters of a significantly higher granularity. The higher granularity present in these calorimeters poses a much greater challenge for a generative model-based fast simulation tool for a number of reasons:

- Greater structure in the shower can be resolved, increasing the requirement in terms of physics fidelity of the generative model.
- The generative model must simulate the detector response across more channels. This typically means that any given network will require more parameters and therefore more computational operations.

The increased number of channels in a calorimeter subsystem also poses challenges in terms of integrating a generative model into a simulation chain, as training a large number of generative models for different regions of the detector and operating them during simulation time becomes impractical. The approach taken for integrating models into the simulation chain in this thesis will be described in Chapter 9.

Prior to the work presented in this thesis, the potential for generative models to be used as fast simulation tools for highly granular calorimeters has been explored across a range of models. Buhmann et al. [45, 63] investigated the performance of several generative models for simulating photon showers with varying energy in the highly granular ECAL present in the ILD. In particular, a BIB-AE model (see Section 5.4.3) specifically adapted to the task of calorimeter simulation was found to perform well across a range of calorimeter observables. Subsequently, Buhmann, P.M. et al. [4] applied the BIB-AE model to the significantly more challenge physics case of hadronic shower simulation in the form of charged pions with varying energy in the ILD HCAL. The improved performance of the BIB-AE in describing the cell energy spectrum of hits in the shower relative to a WGAN and a GAN (for photons in the ECAL) and a WGAN (charged pions in the HCAL) is shown in Figure 6.2. The BIB-AE has also been shown to provide an accurate description of data collected at a test beam, with various calorimetric observables being described well by a BIB-AE in comparison to two different physics lists provided by GEANT4 [190].

However, so far work on this topic has focused on the specific case of a particle incident orthogonal to, and at a fixed position on, the calorimeter surface. A notable exception is the work of Khattak et al. [56], who investigated the performance of a GAN at generating showers in an idealised highly granular ECAL for particles incident under varying polar angles uniformly distributed in the range of 60 to 120 degrees. In order to generalise these proof of concept studies to a complete detector simulation, an appropriate detector response must be simulated for particles incident to the face of the calorimeter with various energies and incident angles. A minimum complete conditioning for a generative model given a fixed particle type in this case therefore involves the incident energy and two angles.

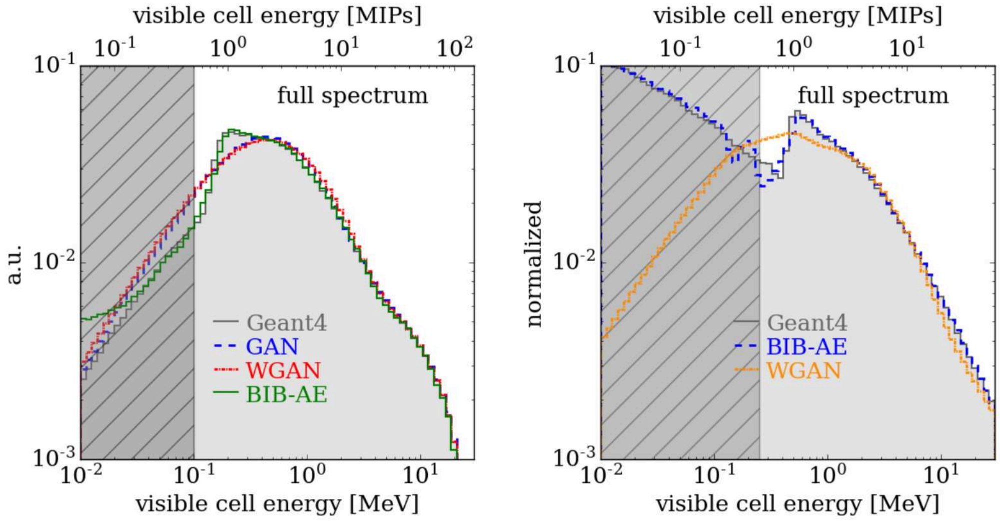


Figure 6.2: Left: Cell energy spectrum for photon showers in the ILD ECAL simulated with GEANT4 (grey, filled), a GAN (blue dashed), a WGAN (red dashed), and a BIB-AE (green dashed). Figure from [63]. Right: Cell energy spectrum for charged pion showers in the ILD HCAL simulated with GEANT4 (grey, filled), a WGAN (orange dashed) and a BIB-AE (blue dashed). Figure from [4]. The hatched region in each figure represents the region below the noise floor of the detector at 0.5 MIP. In both cases, the BIB-AE is able to better describe the steep rise in the spectrum (MIP peak) than the GAN-based networks.

Additionally, with the exception of an initial study on the effect of reconstruction on the resolution and linearity of the detector response performed by Buhmann, P.M. et al. [4], no detailed investigation has been performed into the effects of reconstruction on highly granular calorimeter showers. This is a crucial step, as the ultimate goal of a fast simulation model in replacing a full calorimeter shower simulation is to retain the physics performance after reconstruction. It is therefore essential to conduct detailed investigations into the performance of a generative model after interfacing with the full chain of reconstruction algorithms used by an experiment, both in terms of the lower level calorimeteric obeservables, as well as higher level physics observables.

Finally, in order to judge whether or not the performance of a generative model is sufficient to replace or supplement components of full simulation, it is essential to integrate generative models into the existing software chains. This crucially means also dealing with realistic detector geometries, which are often complex and bring additional challenges that have to be handled, such as insensitive volumes due to gaps in the detector or dead material and irregularities. This means that the local geometry in which a calorimeter shower occurs varies with the impact position of a particle on the calorimeter surface. In addition, it is only possible to perform a fair comparison of the computational performace of a model in comparison to GEANT4 once it has been integrated into the full simulation chain.

Only once these challenges have been overcome will a generative model be useful as a fast simulation tool, with the exact use case of the model determined in terms of which physics benchmarks are described sufficiently well.

Chapter 7

Multi-parameter Conditioning of Generative Models

The work presented in this Section has previously been published in [1] in collaboration with Sascha Diefenbacher, Engin Eren, Frank Gaede, Gregor Kasieczka, Anatolii Korol, Katja Krüger and Lennart Rustige. As such, this section includes numerous figures and tables, as well as significant amounts of text that are similar or identical to parts of that work. This author's contribution to the work included the implementation and optimisation of the BIB-AE model, the implementation and optimisation of the flow model, parts of the dataset production, the application of the reconstruction algorithm, the study of the model's performance, writing sections of the paper, coordination of the peer review process and correspondence with the journal, as well as the addressing of referee comments.

This Chapter focuses on extending the BIB-AE model for which the proof of concept application to highly granular calorimeter showers was already demonstrated, as described in Chapter 6. Previously, the performance of the BIB-AE was successfully demonstrated for the restricted case of a particle incident with varying energy, but fixed incident angle to, and position on, the calorimeter face. In this Chapter, as a first step towards creating a more general generative model for calorimeter simulation, the model will be extended to handle conditioning on an angle of incidence as well as the incident energy, with a fixed incidence position in the calorimeter. Additionally, a detailed study of calorimetric observables before and after applying PandoraPFA for single-particle reconstruction will be performed.

Section 7.2 will address the creation of the data set and the approach adopted for reconstruction. Section 7.3 will then focus on the set up of the generative model used, including the extension to the BIB-AE architecture and the necessary change in approach for latent space sampling. The model performance will then be assessed before and after reconstruction in Section 7.4, before conclusions are drawn in Section 7.5.

7.1 Coordinate System Convention

Throughout the work presented in this thesis, two key coordinate systems will be used. The first is the *global* ILD coordinate system, which will be represented by primed coordinates. The origin of the coordinate system is placed at the IP and it is orientated such that the z' axis points along the beam axis in which the initial e^+e^- collisions occur. The $y' - x'$ plane lies orthogonal to the beam axis, with the y' axis pointing vertically and the x' axis pointing horizontally. It is a right-handed coordinate system. The *polar* angle θ' represents the angle measured from the z' axis. The *azimuthal* angle ϕ' represents the angle in the $x' - y'$ plane, as measured from the x' axis.

The second coordinate system is the *local* calorimeter coordinate system. The origin is placed at the incident position of a particle to the calorimeter face. It is orientated such that the z axis points into the depth of the calorimeter, perpendicular to the orientation of the calorimeter layers. The x and y axes therefore lie paralled to the calorimeter face. The coordinate system is right-handed. The angle θ from the z axis therefore represents the direction of a particle with respect to the layers of the calorimeter, while the angle ϕ represents the orientation with respect to the cell geometry in the $x-y$ plane.

7.2 Dataset and Reconstruction Scheme

The dataset used in this study was produced using the iLCSoft [126] ecosystem (see Section 4.2.2), with full simulation performed with Geant4 [25] version 10.4 using the QGSP_BERT physics list and DD4hep [129] version 1.11. The datasets used consist of electromagnetic showers in the ILD ECAL (see Section 4.2.1). The showers are initiated by photons produced by a *particle gun* placed directly at the face of the calorimeter. In this study the position of the gun is fixed to be $(x', y', z') = (0.0\ \mathrm{mm}, 1810\ \mathrm{mm}, -50\ \mathrm{mm})$ in the global ILD coordinate system. To create the training dataset, the incident energy E was varied uniformly in the range of 10 to 100 GeV simultaneously with the global polar incident angle, which was varied uniformly from 90 to 30 degrees. The global azimuthal angle (corresponding to the angle in the $x-z$ plane in local calorimeter coordinates) was fixed to be 90 degrees. A total of $500k$ showers composed the training dataset [1]. Since a simulator based on a generative model will be used to produce showers initiated with particular energies and angles, fixed combinations of energies and angles are used to evaluate the performance of the model. A total of 9 datasets, each consisting of 1900 showers with fixed combinations of incident energies ($\{20$, 50, $90\}$ GeV) and polar angles ($\{40$, 60, $85\}$ degrees). These points were chosen to cover a broad range of the phase space, while avoiding any edge effects that might be introduced right at the boundaries of the training ranges. Another 9 independent datasets with an identical number of showers and combinations of incident energies and angles were used as validation datasets to select the best performing state of the model. A fixed calibration factor was applied to scale the hit energies in the last 10 layers of the calorimeter to account for the modification of the sampling fraction caused by the absorber layers being twice as thick in this region (see Section 4.2.1).

Importantly, since the BIB-AE model that will be used relies on the use of 3D convolutions, the data provided to the model must be in the form of a regular grid. Calorimeter shower images are created by projecting the cells in the ECAL into a regular grid of $(x, y, z) = 30 \times 60 \times 30$ in the local calorimeter coordinate system. This means that each plane of voxels for a given z position corresponds to one of the 30 calorimeter layers present in the physical ECAL. The size of grid was chosen such that on average approximately 99% of the visible energy deposited in the detector was contained for the highest incident energy and most inclined incident angle present in the data. In the grid that results from the projection, the photons are incident at a fixed cell with index $(i_x, i_y, i_z) = (15, 12, 0)$.

A key problem of this projection to a regular grid arises from the irregular nature of the physical readout geometry present in the detector. These irregularities include staggering effects between layers, and insensitive volumes within an active layer arising for example from dead material such as readout electronics or gaps between detector elements, which are necessary for the construction of the detector. The effect of the staggering between layers appears in the $z-x$ plane, as shown in Figure 7.1. These irregularities tend to cause artefacts that appear in the regular grid as empty lines of cells. These are corrected for in the datasets such that each voxel in the regular grid corresponds to exactly one sensor in the calorimeter.

---

[1] The training dataset and the workflow for creating the test/validation dataset and running reconstruction were provided by Engin Eren. The training dataset has been published and is available at [191]

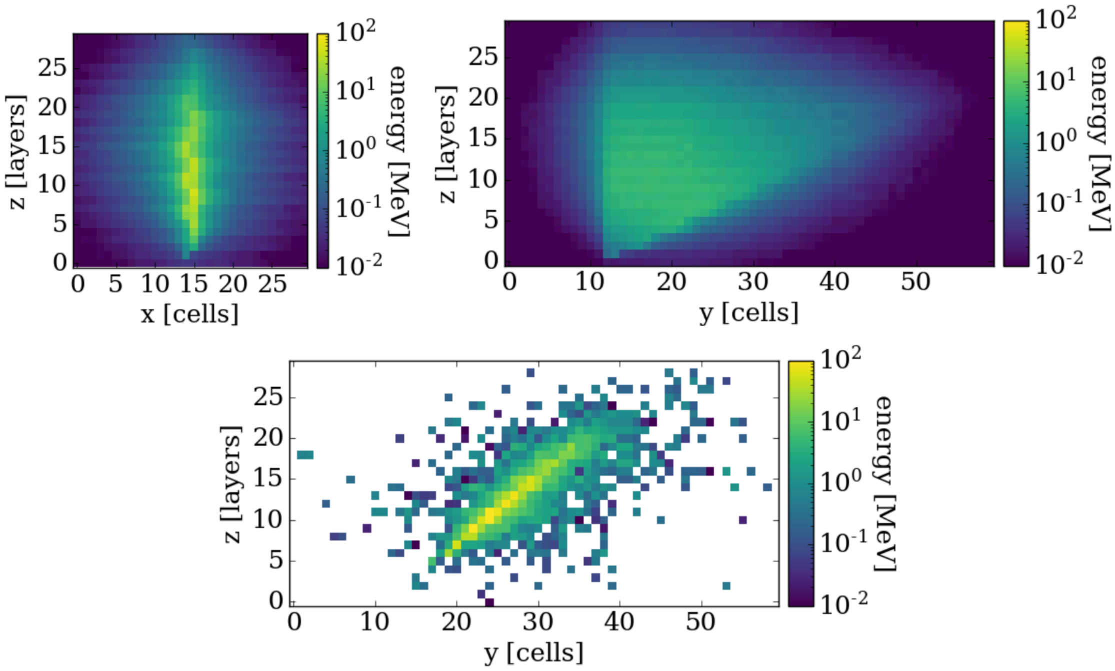


Figure 7.1: Top: Overlays of $5,000$ showers from the training dataset projected into a regular grid. Top Left: Projection of the overlay along the y axis. The staggering effect that appears between layers in the z direction is an artefact arising from projecting the showers from the irregular detector geometry into a regular grid. Top Right: Projection of the overlay along the x axis, showing the angle varied in the $z-y$ plane. Bottom: Example shower from the training dataset with incident energy 88.5 GeV and angle 45.9 degrees projected along the x axis.

In order to study the effects of reconstruction, both the fixed test samples generated with Geant4 and showers generated with the BIB-AE for identical incident energies and angles were passed through the standard reconstruction chain used by ILD, as described in Section 4.2.3. This procedure includes the application of PandoraPFA, which in the case of a neutral cluster applies calorimeter hit clustering to showers after digitisation and calibration. However, in order to be able to apply reconstruction, the hits in the shower must first be placed back into the physical geometry. This was achieved by means of a map look-up, connecting a single voxel regular grid to a single cell in the physical geometry. In order to create the map, a separate sample of negatively charged muons was created, with each cell on the calorimeter face having a muon gun positioned in front of it. With the magnetic field turned off, the muons were fired through the calorimeter perpendicular to the calorimeter face, ensuring that every cell present in the regular grid was activated. This allowed the creation of the map between the regular grid and the physical geometry [2]. For a fair comparison, both the Geant4 and BIB-AE showers were projected in the same fashion to the regular grid, and then mapped back to the detector for reconstruction. Note that the corrections for artefacts in the regular grid, and the creation of the map needed for reconstruction, crucially require the local geometry of the detector to be known. This was only possible in this study because of the fixed position in the calorimeter, and is generally not practical for an arbitrary incident position. Strategies for dealing with arbitrary incident positions will be explored in subsequent chapters.

---

[2] The map between the regular grid and the physical geometry was provided by Anatolii Korol, using a muon sample created by the author.

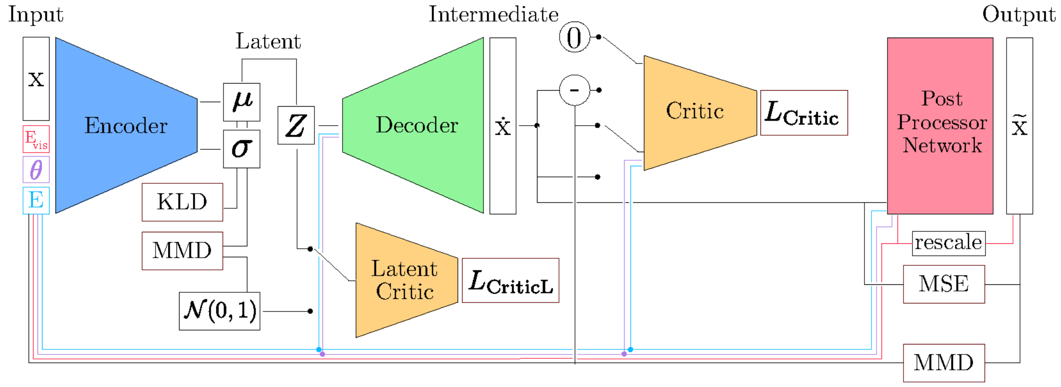


Figure 7.2: Schematic diagram of the BIB-AE architecture used during training, including each network and its corresponding loss terms. The encoder maps the input calorimeter showers to a low-dimensional latent space, which is regularised by KLD and MMD loss terms in addition to a latent critic. The decoder reconstructs shower images back from the latent space, with a dual purpose reconstruction critic simultaneously assisting latent-reconstruction and judging the quality of the output shower. The Post Processor network is trained in a second step to adjust voxels individually. The light blue and lilac lines represent an input conditioning on the energy and angle of the incident particle, while the red line represents a conditioning on the visible energy. Figure and caption from [1].

7.3 Generative Model

7.3.1 BIB-AE Model

The BIB-AE architecture used is adapted directly from the version developed in [4] by Buhmann, P.M. et al. for the extended conditioning scenario explored here. The model is implemented in PyTorch [192], with the architecture being illustrated schematically in Figure 7.2. While the theoretical foundation of the model is based on the description in Section 5.4.3, the model is adapted specifically for the task of calorimeter shower simulation. The core of the architecture is an autoencoder, consisting of an encoder network N_E to map input showers to a lower dimensional latent representation and a decoder network N_D to reconstruct showers out of the latent space. Additionally, according to the design principles outline in Section 5.4.3, a number of auxiliary networks are chosen to improve specific elements of the training or generation process.

Architecture and Loss

Two of the auxiliary components focus on the latent space. As described in Section 5.4.1, for a VAE-like architecture, the intermediate bottleneck representation must be regularised towards a known distribution, for which we make the standard choice of a standard Normal distribution $\mathcal{N}(0,1)$. The first loss component is a KLD term L_{KLD}, as described in Section 5.4.1. The second component is composed of two sampling-based methods. The first term is a Maximum Mean Discrepancy (MMD) [193] loss, which is a kernel-based method for comparing distributions. An MMD loss between samples x and x' drawn from one distribution and samples y and y' drawn from another is given by

$$MMD^2(P,Q) = \mathbb{E}[k(x,x')] + \mathbb{E}[k(y,y')] - 2\mathbb{E}[k(x,y)], \qquad (7.1)$$

where k is a positive definite kernel function. For the latent regularisation, this means comparing samples from the latent space produced by the encoder with samples from the standard Normal

$$L_{\mathrm{MMD}} = \mathrm{MMD}(N_E(x), \mathcal{N}(0,1)). \tag{7.2}$$

The second sampling-based method comprises a Wasserstein-GAN-like latent critic C_L (see Section 5.4.2), which learns to distinguish the latent space from a standard Normal distribution, resulting in a loss of

$$L_{\mathrm{Critic_L}} = \mathbb{E}[C_L(N_E(x))]. \tag{7.3}$$

Another WGAN-like critic network, the reconstruction critic C, acts on the output shower images reconstructed from the latent space by the decoder. This critic serves a dual purpose, providing feedback as to the quality of the latent-reconstructed showers directly, but also by comparing the input to the autoencoder to its output by means of a difference between the two. The loss contribution of this critic is denoted as

$$L_{\mathrm{Critic}} = \mathbb{E}[C(N_D(E(x)), x)]. \tag{7.4}$$

Following developments in [4], each of the latent and reconstruction critic networks represented in Figure 7.2, actually consist of two identical networks with independent weights. One network is trained continuously in the standard fashion, while the second has its weights and optimiser reset after each epoch. This was originally introduced in [4] to deal with the higher sparsity present in hadronic showers, where a continuously trained critic could become blind to the outer regions of a shower, and focus only on the shower core. However, we find that the approach of using dual critic networks also aids the network training with the increased sparsity present in the photon shower data in this case, due to the larger grid size used. Additionally, the minibatch discrimination [194] introduced for the BIB-AE in [4] is retained for this study. This involves additionally providing the model with information pertaining to the composition of the batch. This is achieved by producing a set of difference matricies between the output and input shower images, as well as with logarithmic scaled versions of both. This information is then incorporated into the reconstruction critic via an embedding [190].

These individual losses result in a total architectural loss for the main BIB-AE of

$$\begin{aligned} L_{\text{BIB-AE}} = \beta_{\mathrm{KLD}} \cdot L_{\mathrm{KLD}} &+ \beta_{\mathrm{MMD}} \cdot L_{\mathrm{MMD}} \\ &+ \beta_{\mathrm{Critic}_L} \cdot L_{\mathrm{Critic_L}} + \beta_{\mathrm{Critic}} \cdot L_{\mathrm{Critic}}, \end{aligned} \tag{7.5}$$

where the strength of the contribution from each term is controlled by an independent hyperparameter β_i. The values of these hyperparameters were chosen such that each loss term has a similar order of magnitude, as described in [4]. The values for each of these hyperparameters were therefore set as follows: $\beta_{\mathrm{KLD}} = 0.1$, $\beta_{\mathrm{MMD}} = 100$, $\beta_{\mathrm{Critic_L}} = 100$ and $\beta_{Critic} = 1$.

Once the main BIB-AE model has been trained the weights of this model are frozen and a separate *Post Processor* (PP) network is trained according to the developments described in [4]. This network is specifically designed to tune the model to better describe the cell energy spectrum (see Section 6.3), by applying a series of 3D convolutions with kernel size one in order to encourage the adjustment of individual voxel energies, rather than creating or removing a hit. The PP network is trained for one epoch with only an MSE loss term between the input and output showers, as defined in Equation 5.21. Subsequently, the training is continued with this MSE term, along with another 5 loss terms.

The first two are Sorted Kernel Maximum Mean Discrepancy (SK-MMD) loss terms, which were introduced in [63]. This loss computes the MMD (Equation 7.1) only on the $N_{hits} = 1,000$ most energetic hits in the shower, avoiding wasted computation on hits with very low energies. Two vectors are created from these $1,000$ hits — the first sorted by decreasing energy, and the second without any sorting applied. The MMD is then computed between corresponding vector entries lying in a window of m hits, which is passed along all $1,000$ hits with a stride s, before a final loss term is computed by means of an addition over all of the

MMD losses calculated. The values are set to $m = 150$ and $s = 25$. A Gaussian kernel k given by

$$k(x, x') = e^{\alpha(x^2+x'^2-2xx')}, \tag{7.6}$$

is used, with the scale factor α governing the size of the feature which the loss is sensitive too [190]. Two separate SK-MMD loss terms $L_{\text{SK-MMD1}}$ and $L_{\text{SK-MMD2}}$ are used with kernel factors $\alpha_1 = 40$ and $\alpha_2 = 4$ respectively. The difference in sensitivity of these two terms allows a broader range of the cell energy spectrum to be covered.

The second two Post Processor loss terms operate on the input and output shower hits sorted by energy in vectors. The first term computes an MSE between the sorted vectors, with loss contribution $L_{\text{S-MSE}}$ and the second computes a Mean Absolute Error (MAE) loss, given by

$$L_{\text{MAE}} = \mathbb{E}\,|x - \tilde{x}|, \tag{7.7}$$

which also operates on the sorted vectors, and results in a loss contribution of $L_{\text{S-MAE}}$.

A final contribution to the Post Processor loss is provided by an MSE loss computed between batches of input and output shower images, making the loss sensitive to artefacts that only appear across multiple showers. This term provides a contribution to the Post Processor loss of $L_{\text{BatchComp.}}$.

These individual losses result in a total loss for the Post Processor network of

$$\begin{aligned} L_{\text{PP}} = \beta_{\text{MSE}} \cdot L_{\text{MSE}} &+ \beta_{\text{SK-MMD1}} \cdot L_{\text{SK-MMD1}} \\ &+ \beta_{\text{SK-MMD2}} \cdot L_{\text{SK-MMD2}} + \beta_{\text{S-MSE}} \cdot L_{\text{S-MSE}} \\ &+ \beta_{\text{S-MAE}} \cdot L_{\text{S-MAE}} + \beta_{\text{BatchComp.}} \cdot L_{\text{BatchComp.}}, \end{aligned} \tag{7.8}$$

where the strength of the contribution from each term is controlled by an independent hyperparameter β_i. The values of these hyperparameters are set as follows: $\beta_{\text{MSE}} = 1$, $\beta_{\text{SK-MMD1}} = 5$, $\beta_{\text{SK-MMD2}} = 5$, $\beta_{\text{S-MSE}} = 10$, $\beta_{\text{S-MAE}} = 10$, $\beta_{\text{BatchComp.}} = 0.0001$.

Structure of the Networks

In order for the BIB-AE model to be useful as a simulation tool, it must be able to provide an appropriate detector response for given characteristics of the incident particle. Therefore, as shown in Figure 7.2, all networks composing the architecture except the latent critic are conditioned on the energy (light blue pipes) and angle (lilac pipes) of the incident particle. The Post Processor network is additionally conditioned on the visible energy sum of a given shower. How this visible energy is determined during inference time will be discussed in Section 7.3.2. The networks are built around combinations of 3D convolutions and fully connected layers. The structure of each network is as follows.

The encoder-decoder pair uses a total of eleven 3D convolutional layers, two transpose 3D convolutional layers and seven fully connected layers. Layer normalisation is applied after each convolution except for in the final layer of the decoder, and LeakyReLU activations are applied to each layer except for the final encoder and decoder layers, where linear and ReLU activations are applied respectively.

The critic network uses a total of six 3D convolutional layers, each followed by a normalisation layer, eleven fully connected layers and a further twelve 1D convolutions for the minibatch discrimination. LeakyReLU activations are applied throughout, except for in the final layer.

The latent critic is a purely fully connected architecture, consisting of four fully connected layers followed in each case by LeakyReLU activations, except in the final layer.

The Post Processor network uses a total of six 3D convolutional and three fully connected layers. No layer normalisation is used, while LeakyReLU activations are applied throughout except in the final layer. Unlike in previous implementations of Post Processing for the BIB-AE

Table 7.1: Key hyperparameters used in the training the BIB-AE model, separated into the training of the main model and the Post Processor network.

	Category	Hyperparameter	Value
Main BIB-AE		Batch Size	64
		No. Epochs	50
	Loss Terms	β_{KLD}	0.1
		β_{MMD}	100
		β_{Critic_L}	100
		β_{Critic}	1
	Initial LR	$LR_{Encoder}$	0.5×10^{-4}
		$LR_{Decoder}$	0.5×10^{-4}
		LR_{Critic_L}	0.5×10^{-4}
		LR_{Critic}	2.0×10^{-4}
Post Processing		Batch Size	32
		No. Epochs	53
	SK-MMD Loss	N_{hits}	$1,000$
		m	150
		s	25
		α_1	40
		α_2	4
	Loss Terms	β_{MSE}	1
		$\beta_{\text{SK-MMD1}}$	5
		$\beta_{\text{SK-MMD2}}$	5
		$\beta_{\text{S-MSE}}$	10
		$\beta_{\text{S-MAE}}$	10
		$\beta_{\text{BatchComp.}}$	0.0001
	Initial LR	LR_{PP}	0.5×10^{-4}

[4, 63], no layer normalisation is used between layers, as this was found to reduce the networks' ability to learn the inclination of showers.

More details on the architectures of each network are provided in Appendix A.

Training and Validation

The ADAM optimiser [158] was used for training all networks in the BIB-AE, with an exponential learning rate decay. A common initial learning rate (LR) of 0.5×10^{-4} is used for the encoder, decoder, critic and Post Processor network, while the initial learning rate for the latent critic is set to 2.0×10^{-4}. The learning rate is decayed by a factor of 0.97 after each epoch. A batch size of 64 was used during the training of the main BIB-AE, while this was reduced to 32 during the training of the Post Processor network. The BIB-AE architecture was trained for a total of 50 epochs, after which the model was frozen and the Post Processor trained for a further 53 epochs. The GEANT4 showers used for training all have a threshold applied to map hits below 1×10^{-4} MeV to zero. The key hyperparameters used in the training procedure are summarised in Table 7.1.

Due to the adversarial nature of the training, the performance of the network can fluctuate significantly between epochs. The best epoch of the training was selected by means of a fidelity scan over epochs. At each epoch, the validation data was split into two, with the first half being run through the encoder-decoder pair of the BIB-AE, and the remaining half being retained for comparison. Passing the data through the autoencoder was necessary to avoid the computational infeasibility of modelling the latent space (see Section 7.3.2) for every epoch individually, since the structure of the latent space varies during the course of training. For each epoch, seven key calorimetric observables (angular response, energy response, cell energy spectrum, number of hits above threshold, center of gravity, longitudinal profile and radial profile — see Section 7.4) were used as the criteria for selecting the best model. For each observable, a single value quantifying the network performance at that epoch was produced by means of a bin-wise area difference [3] between the GEANT4 distribution and the output of the autoencoded showers. This approach was used to determine the best performing epoch of the main BIB-AE training, with particular emphasis being placed on the quality of the energy and angular response modelling. Post Processing was then trained starting with this state of the main BIB-AE as a base, and the best epoch of Post Processing selected via a similar scan over epochs.

7.3.2 Latent Space Sampling

As described in Section 5.4.3, the optimisation of the BIB-AE represents a trade-off between information retention and compression. This means that relaxing the regularisation constraint on the latent space allows more information to be retained. However, this comes at a cost — the less regularisation is applied, the more the latent space variables will deviate from standard Normal distributions. While this more flexible latent space contains more encoded information, in order to be able to use the BIB-AE as a simulator it must be possible to draw samples from the latent representation z. Since the latent space variables no longer perfectly follow a standard Normal distribution, using latent samples drawn form this distribution degrades the performance of the network. It is for this reason that previous BIB-AE studies [4, 45] have employed Kernel Density Estimation [195] to provide an improved modelling of the latent space variables and their correlations. For a set of N data points $\{x_1, x_2, ..., x_N\}$ drawn from the latent space, the KDE is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right), \tag{7.9}$$

where $h > 0$ is a smoothing parameter called the bandwidth and K is the kernel function. Once the KDE has been fitted to the latent distribution, it can be sampled from during inference time to provide input to the decoder.

However, a problem arises with the conditioning scheme in this approach. Previously, in order to generate a sample of a given incident energy, the energy label E was concatenated with the latent sample prior to fitting the KDE. This meant that sampling from the KDE also provided a corresponding energy label for a given sample. This allowed the use of a rejection sampling technique, whereby repeated sampling was performed until a sample was found to lie within a narrow window around the desired incident energy. While this sampling technique was adequate for a single conditioning parameter, it scales poorly to the general case of multi-parameter conditioning.

For this reason, the density estimation approach adopted for latent space sampling must be improved. We therefore use a Normalising Flow (NF) model, which is well suited to this low-dimensional density estimation task, and can be conditioned in a straightforward manner as described in Section 5.4.5. The model is implemented using the PYRO [196] deep probabilistic

---

[3] This metric is calculated with a normalisation applied such that a value of one means the distributions do not overlap at all and a value of zero corresponds to perfect overlap.

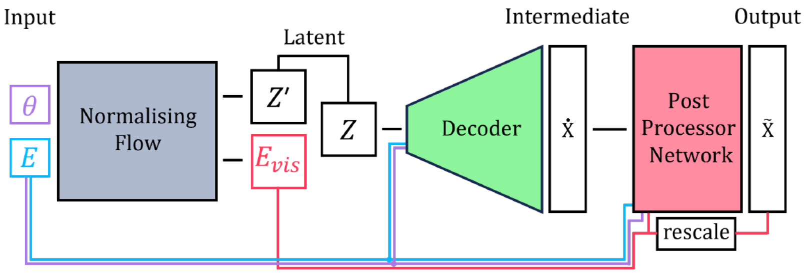


Figure 7.3: Schematic diagram of the combined BIB-AE and Normalising Flow architecture used during inference.

programming library and consists of eight blocks, each containing seven coupling layers. Six of these coupling layers are based on affine transformations [197] and one is based on element-wise rational spline bijections of linear order [175, 176]. Unlike in the previous studies where the conditioning variables were included in the density estimation, in this case we provide each layer of the NF the energy and angular conditioning labels as a two dimensional context. A pre-scaling is applied to each of the labels before being used for conditioning — the energy labels are divided by a factor of 100 and the angular labels are divided by a factor of $\frac{\pi}{2}$. In order to train the model, the 500k showers used for training are encoded with the BIB-AE model using the weights of the best performing epoch, resulting in a total of 24 latent variables that need to be modelled. In addition, the visible energy sum above the 0.5 MIP threshold of each shower in MeV divided by a factor of 10^4 was appended to the latent space. This resulted in a 25-dimensional latent space as the density estimation target. This means that the NF learns to generate a latent-encoded shower and its corresponding visible energy sum. During inference time, this means that the NF can be used to provide not only the latent variables for the decoder for a given incident energy and angle, but also the energy sum that is provided as an additional conditioning parameter for the Post Processor network. Following [48] this energy sum is then also used to perform a per-shower re-scaling of the individual voxels in the shower, such that the sum of all individual voxels above the 0.5 MIP threshold corresponds to the target energy sum.

To train the NF model, a batch size of 2048 was used, with model optimisation performed using the ADAM optimiser [158] with a learning rate of 1×10^{-4}. The model was trained for a total of 200 epochs, with the best epoch being selected from the lowest loss value. During inference, the weights of the NF model were frozen, and samples drawn in order to permit shower generation[4]. The architecture used during inference is shown in Figure 7.3.

7.4 Results

We now perform an evaluation of the performance of the BIB-AE in comparison to Geant4. In Section 7.4.1 the physics performance of the model will be studied, both before and after reconstruction. Subsequently in Section 7.4.2, the computational performance of the model will be benchmarked.

7.4.1 Physics Performance

In order to assess the physics performance of the BIB-AE, we study the ability of the model to reproduce key calorimetric observables present in the Geant4 generated showers. This comparison between BIB-AE and Geant4 involves each of the nine test points consisting of fixed combinations of incident angles and energies, as described in Section 7.2. The comparison is broken down into two parts for each observable.

---

[4] The implementation of the model and the hyperparameters used for training are available at `https://github.com/FLC-QU-hep/new_angles`

Firstly, the model performance at *simulation level* is evaluated. This involves a comparison between the direct output of the BIB-AE and GEANT4 with a calibration applied as described in Section 7.2. This provides a means to judge the ability of the BIB-AE to learn from the training data. At this level of comparison, the only additional processing applied to the showers is the mapping of hits with an energy deposition of less than 0.07875 MeV, which corresponds to an energy deposition of less than half a MIP, to zero. This emulates the processing applied in a real calorimeter, where hits with energy less than this threshold lie below the noise floor of the detector.

Secondly, a comparison is made at *reconstruction level* after processing the showers generated with both models using the reconstruction procedure described in Section 7.2. A quality criterion is placed on the data, such that only events in which one PFO is reconstructed are selected for the performance evaluation.

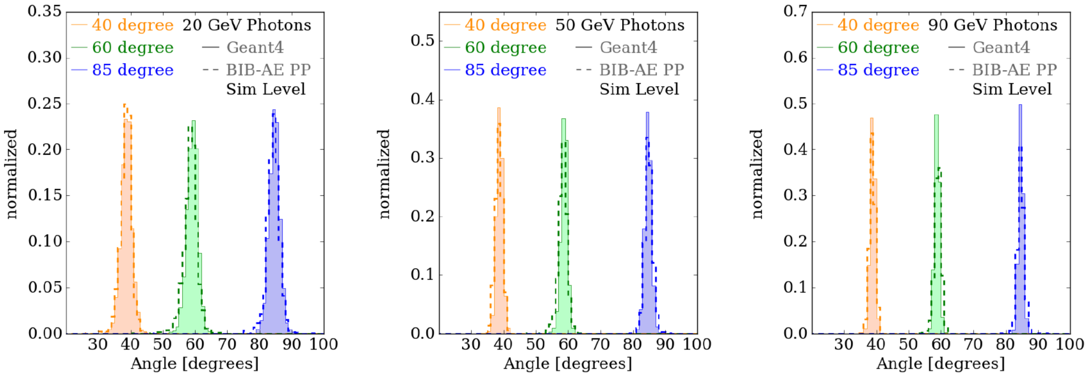


Figure 7.4: Angular response at simulation level for both GEANT4 (filled histograms), and BIB-AE generated photon showers (dashed, unfilled histograms). Distributions are shown for each incident energy of 20 GeV (left), 50 GeV (center) and 90 GeV (right), and for each incident angle of 40 degrees (orange), 60 degrees (green) and 85 degrees (blue). Figure and caption reproduced from [1].

We begin by focusing on the conditioning performance of the model, which will be studied by means of observables that are highly correlated with the angle and energy of the incident photon. Subsequently, the cell energy spectrum, number of hits, center of gravity, longitudinal profile and radial profile will be considered. For the latter observables with the exception of the number of hits, the performance of the model for each combination of incident angle and energy was assessed by calculating the Jensen-Shannon Distance (JSD) between the GEANT4 and BIB-AE distributions. The JSD between two probability distributions P and Q, with probability densities p and q respectively, is the square root of the Jensen-Shannon divergence

$$\mathrm{JSD}(P||Q) = \sqrt{\frac{\mathcal{D}_{KL}(P||M) + \mathcal{D}_{KL}(Q||M)}{2}}, \tag{7.10}$$

where $M = \frac{1}{2}(P+Q)$, and $\mathcal{D}_{KL}$ is the Kullback–Leibler divergence, defined in Equation 5.24. For brevity, only the distributions for the best and worst performing angle and energy combinations in terms of JSD score will be used to illustrate the performance for each observable. Additional test points are included in Appendix A.

In this section, a consistent colour scheme will be used to represent the various fixed incident angles of 40 degrees (orange), 60 degrees (green) and 85 degrees (blue).

Angular Response

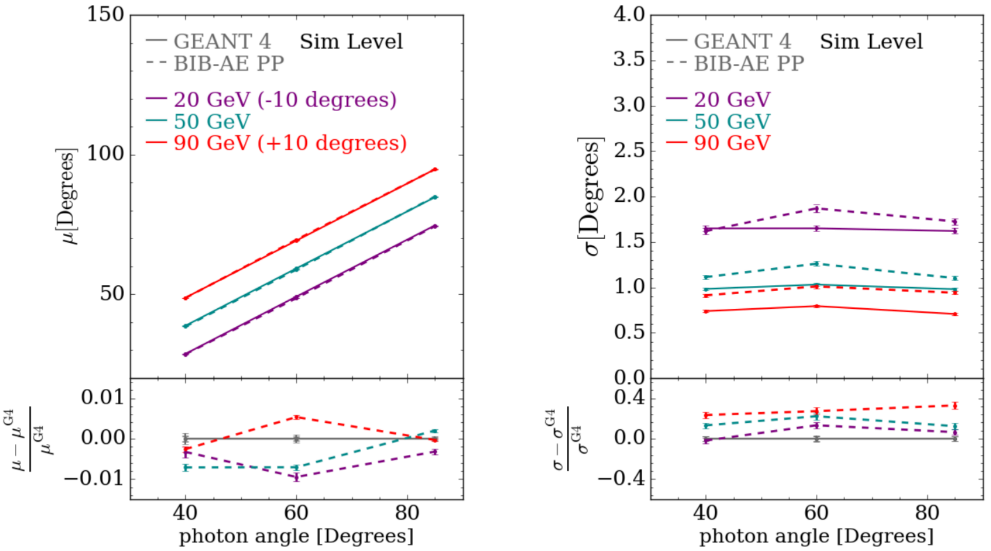


Figure 7.5: Simulation level angular linearity (left) and width (right) for both GEANT4 and BIB-AE generated showers. Curves are shown for each of the fixed incident energies of 20 GeV, 50 GeV and 90 GeV, which are coloured purple, dark cyan and red respectively. In the angular linearity plot on the left, the means for showers with energies of 20 GeV and 90 GeV are shifted by constant values of -10 degrees and $+10$ degrees respectively for visual purposes. The sub-panels in each figure show the relative deviation of the BIB-AE angular responses from their GEANT4 equivalents. The errors on each point are statistical. Figure and caption reproduced from [1].

In order to characterise the quality of the BIB-AE response with respect to the incident angle of the particle, a principle component analysis (PCA) is applied to the showers in order to obtain the principal axis of the showers. The resulting angular distributions for the BIB-AE and GEANT4 for 20 GeV (left), 50 GeV (center) and 90 GeV (right) showers are shown in Figure 7.4. In general, across the range of incident energy and angle combinations considered, the BIB-AE distributions show a high degree of agreement with the corresponding GEANT4 distributions. The clearest discrepancy appears in the form of the BIB-AE producing distributions that are somewhat broader than their GEANT4 counterparts, which becomes more pronounced as the GEANT4 distributions become sharper for photons with higher incident energies.

The angular performance at simulation level is quantified in more detail in Figure 7.5 via Gaussian fits to both the GEANT4 and BIB-AE distributions. The resulting means μ (left) and standard deviations σ (right) of the fits are plotted as functions of the incident particle angles in order to obtain the effective angular linearity and widths for each fixed incident energy of 20 GeV, 50 GeV and 90 GeV. The linearity of the response of the BIB-AE simulator and GEANT4 agree very well, with the maximum relative deviations reaching only to the 1 % level, indicating the means of the distributions being reproduced well by the BIB-AE. The effective angular resolution in the right of the figure demonstrates the aforementioned broadening of the BIB-AE distributions, with deviations in the widths of the BIB-AE distributions from their GEANT4 counterparts becoming more apparent as the incident energy is increased. For example, the angular response for 20 GeV showers at 40 degrees shows excellent agreement, whereas a deviation of almost 40% is observed for 90 GeV showers at 85 degrees.

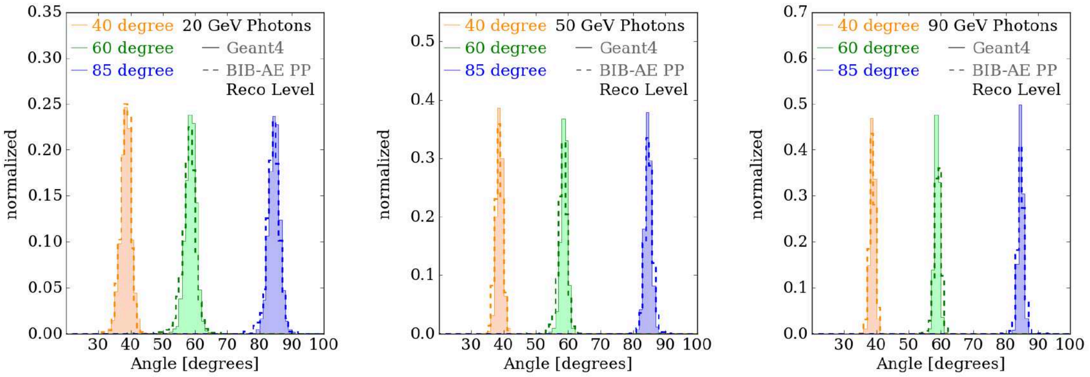


Figure 7.6: Reconstructed angular response for both GEANT4 (filled histograms), and BIB-AE generated photon showers (dashed, unfilled histograms). Distributions are shown for each incident energy of 20 GeV (left), 50 GeV (center) and 90 GeV (right), and for each incident angle of 40 degrees (orange), 60 degrees (green) and 85 degrees (blue). Figure and caption reproduced from [1].

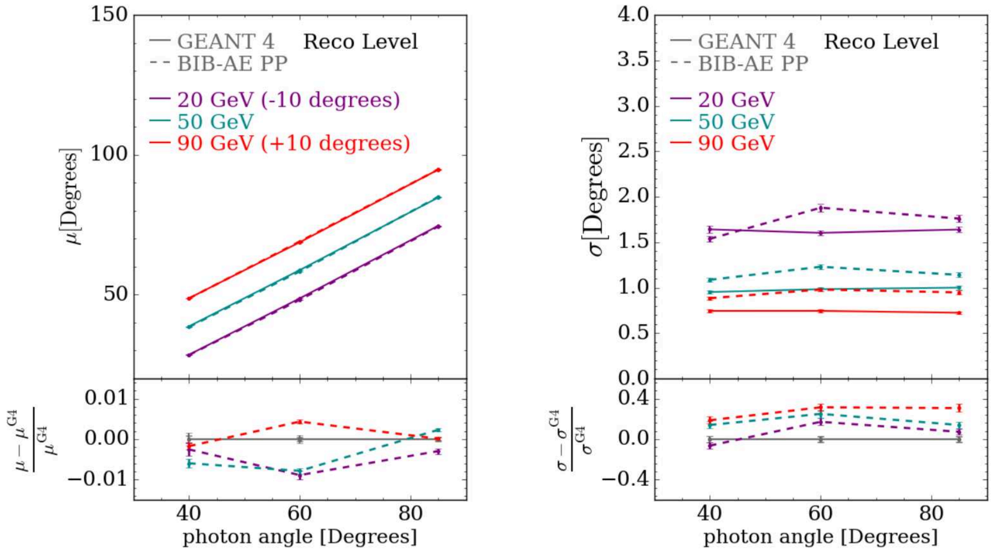


Figure 7.7: Reconstructed angular linearity (left) and width (right) for both GEANT4 and BIB-AE generated showers. Curves are shown for each of the fixed incident energies of 20 GeV, 50 GeV and 90 GeV, which are coloured purple, dark cyan and red respectively. In the angular linearity plot on the left, the means for showers with energies of 20 GeV and 90 GeV are shifted by constant values of -10 degrees and $+10$ degrees respectively for visual purposes. The sub-panels in each figure show the relative deviation of the BIB-AE angular responses from their GEANT4 equivalents. The errors on each point are statistical. Figure and caption reproduced from [1].

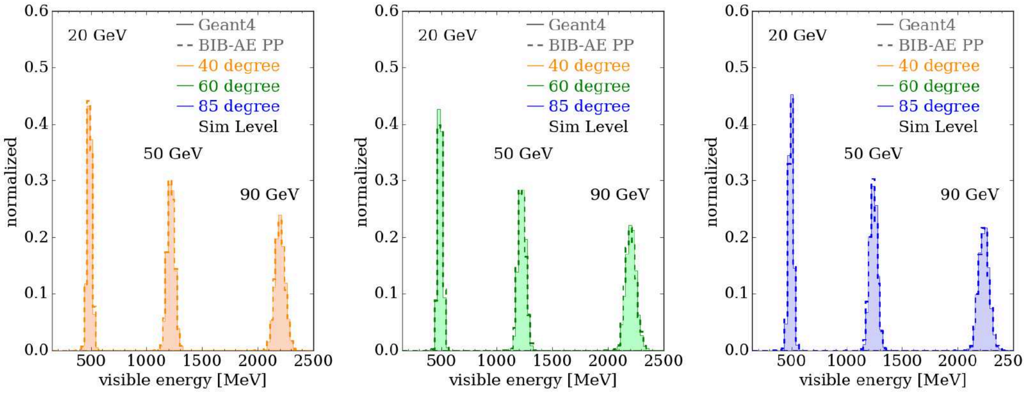


Figure 7.8: Visible energy deposited in the calorimeter at simulation level for both Geant4 and BIB-AE generated showers. The distributions are grouped according to incident photon angles of 40 degrees (left, orange), 60 degrees (center, green) and 85 degrees (right, blue). Figure and caption reproduced from [1].

The corresponding distributions after reconstruction is applied to both BIB-AE and Geant4 showers can be seen in Figure 7.6. In this case, the reconstructed angle is obtained from the intrinsic direction of the reconstructed cluster. The distributions show a very similar behaviour to that which was observed at simulation level, indicating that the reconstruction procedure has very little effect on the observed angular response. This means the broad agreement between BIB-AE and Geant4 distributions remains, with the slight mismodelling of the sharper peaks in the Geant4 distributions appearing at higher energies still being present.

The effects of reconstruction on the angular linearity and width can be seen in Figure 7.7. Since only minor changes can be seen in the means and widths of the BIB-AE with respect to Geant4, this provides further confirmation that the calorimeter hit clustering involved in the reconstruction procedure has a minimal effect on the angular performance of the BIB-AE.

Energy Response

The energy response of the BIB-AE model at simulation level is characterised by the sum of all energy depositions above the 0.5 MIP threshold which occur in the sensitive regions of the calorimeter, i.e. the total visible energy. The distributions obtained from both Geant4 and BIB-AE showers are shown in Figure 7.8. An excellent agreement can be seen between the BIB-AE distributions and their Geant4 counterparts across the full range of test points. This is a direct consequence of the per-shower re-scaling procedure made possible by the use of the NF model for latent space sampling, as described in Section 7.3.2.

A more detailed characterisation of the energy response at simulation level is provided in Figure 7.9. The mean (μ_{90}) and standard deviation (σ_{90}) are calculated using the central 90% of each of the distributions, and the linearity (left) and resolution (right) shown in the Figure as a function of the incident particle energy, each coloured according to the incident particle angle. The linearity is reproduced particularly well by the BIB-AE, especially for photons with low incident angles. The deviations from Geant4 in this case are limited to well below the 1% level. The resolution exhibits larger deviations from Geant4, but still represents a good agreement between the BIB-AE and Geant4, with the maximum relative deviations being restricted to less than 10%.

The energy response after reconstruction is shown in Figure 7.10. The distributions show the energy of the PFO reconstructed by PandoraPFA. Since this also involves a calibration procedure, this also corresponds to the incident energy of the reconstructed particle. Similarly to the simulation level results, an excellent agreement between the BIB-AE and Geant4

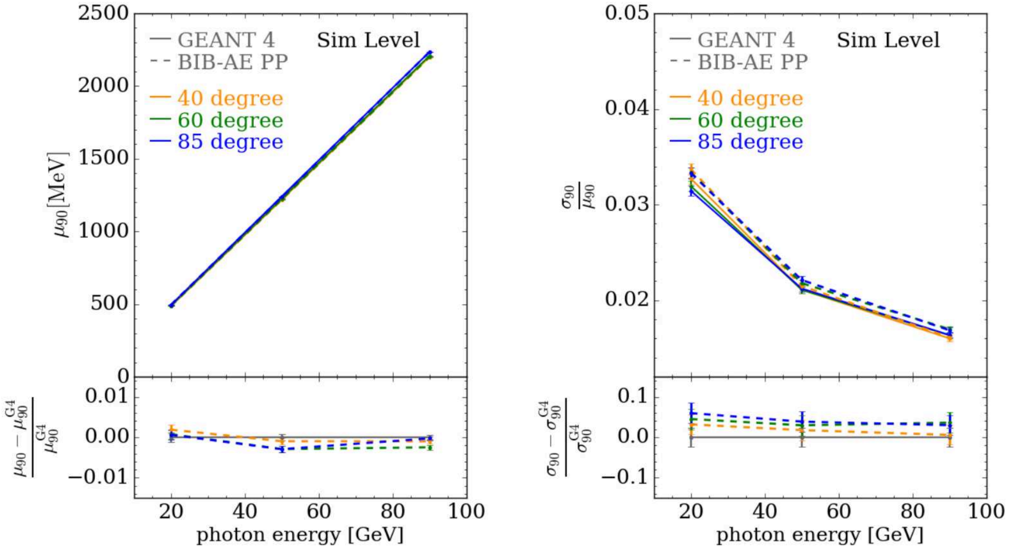


Figure 7.9: Energy linearity (left) and resolution (right) at simulation level for for both GEANT4 and BIB-AE generated showers. The curves are grouped according to the three incident angles of 40 degrees (orange), 60 degrees (green) and 85 degrees (blue). The sub-panels in each figure show the relative deviation of the BIB-AE visible energy responses from their GEANT4 equivalents. The errors on each point are statistical. Figure and caption reproduced from [1].

showers is observed. This is further supported by the PFO linearity and resolution shown in Figure 7.11. While the effect of reconstruction on the linearity is minimal, a slight degradation in the resolution is noticeable, in particular for more inclined showers. However, these deviations still remain below the 10% level across all of the test points.

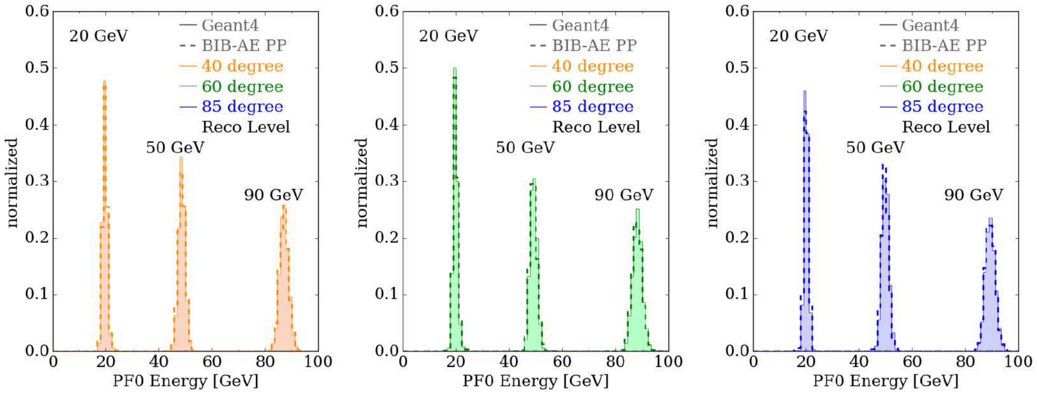


Figure 7.10: Reconstructed particle (PFO) energy for both GEANT4 and BIB-AE generated showers. The distributions are grouped according to incident photon angles of 40 degrees (left, orange), 60 degrees (center, green) and 85 degrees (right, blue). Figure and caption reproduced from [1].

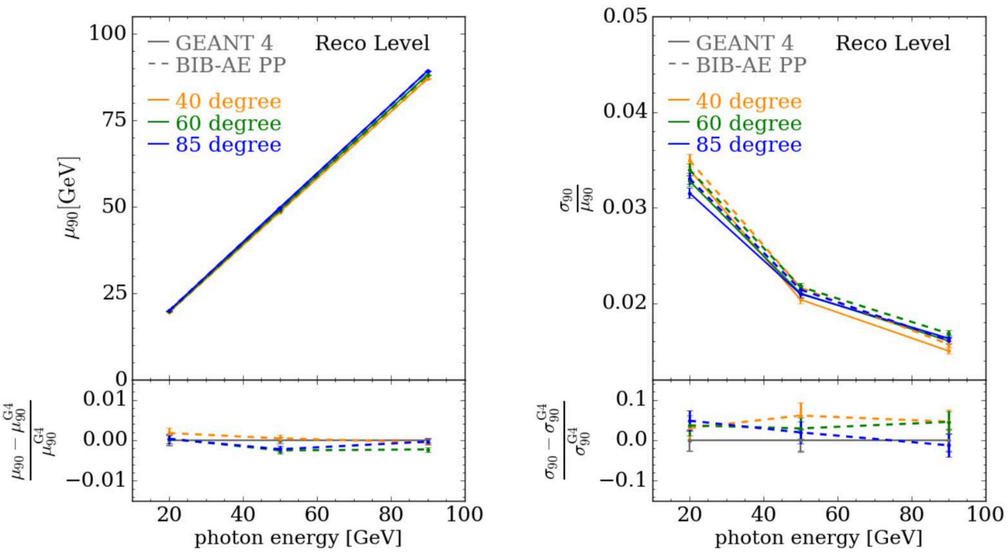


Figure 7.11: Reconstructed particle (PFO) energy linearity (left) and resolution (right) for both GEANT4 and BIB-AE generated showers. The curves are grouped according to the three incident angles of 40 degrees (orange), 60 degrees (green) and 85 degrees (blue). The sub-panels in each figure show the relative deviation of the BIB-AE visible energy responses from their GEANT4 equivalents. The errors on each point are statistical. Figure and caption reproduced from [1].

Cell Energy Spectrum

The next calorimetric observable considered is the cell energy spectrum. As described in Section 6.3, the BIB-AE was previously able to describe this distribution better than GAN-based networks for the case of single-parameter. The cell energy spectra at simulation level for the best (50 GeV, 85 degree) and worst (90 GeV, 40 degree) combinations of incident energy and angle according to the JSD between GEANT4 and BIB-AE distributions (Equation 7.10) are shown in Figure 7.12. Additional test points are included in Appendix A.2.1. The grey hatched region in the left of each figure lies below half a MIP (the threshold termed the MIP cut), and hence would be removed in a real calorimeter. For this reason, the JSD calculation used to judge the quality of modelling of the simulation cell energy spectra was only performed on the region above the MIP cut. While larger deviations from GEANT4 are seen below the MIP cut, above this threshold the model is able to provide a good description of the distributions, the shape of which now varies depending on the incident energy and angle. A key feature of these distributions is the peak that occurs at a cell energy of around 1 MIP. This feature is well described by the BIB-AE thanks to the dedicated Post Processor Network. This indicates that the previously observed ability of the BIB-AE to model this distribution well extends to more general simulation scenarios involving multiple conditioning parameters. The distributions after reconstruction for the best (50 GeV, 85 degrees) and worst (now 20 GeV, 40 degrees) combinations of incident energy and angle in terms of the JSD are shown in Figure 7.13. For these distributions, an abrupt cut-off occurs at the half MIP threshold, as lower energy hits are removed during the reconstruction procedure. The MIP peak that appeared smooth at simulation level is now smeared out because of the two MIP calibration factors that are applied during the procedure. The strong performance of the BIB-AE in describing the distribution in this region that was observed at simulation level is retained after reconstruction, even in the case of the worst performing incident energy and angle combination at 20 GeV, 40

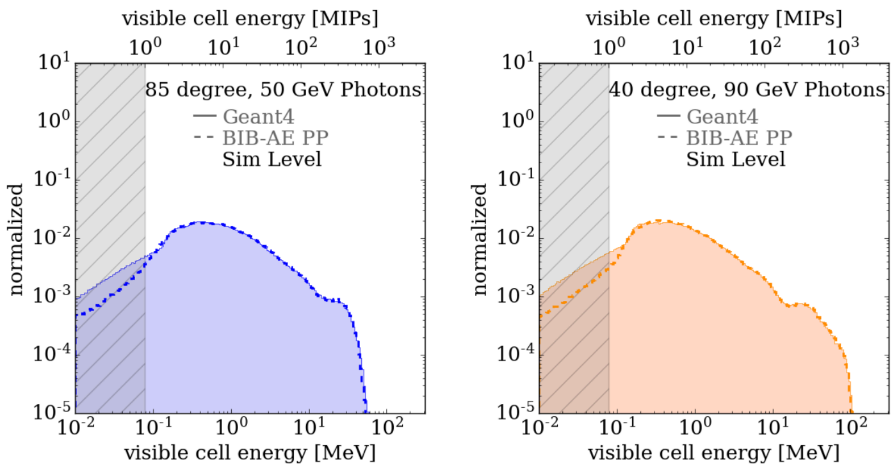


Figure 7.12: Simulation level cell energy spectra for the best (50 GeV, 85 degrees, left) and worst (90 GeV, 40 degrees, right) incident angle and energy combinations. The grey hatched area indicates the region below half a MIP. Figure and caption reproduced from [1].

degrees.

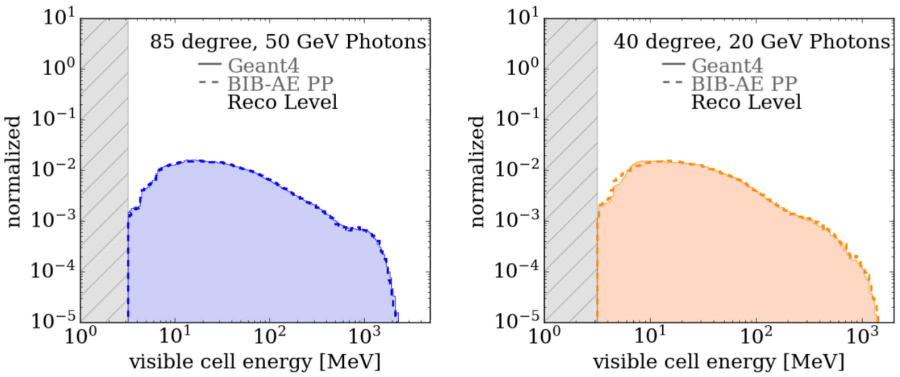


Figure 7.13: Reconstructed cell energy spectra for the best (50 GeV, 85 degrees, left) and worst (20 GeV, 40 degrees, right) incident angle and energy combinations. The grey hatched area indicates the region below half a MIP. Figure and caption reproduced from [1].

Number of Hits

Correctly describing the number of hits above the MIP cut is another regime where the Post Processor Network plays a key role. The distributions of the number of hits above threshold for each incident energy and angle combination at simulation level is shown in Figure 7.14. How well the network is able to model this distribution strongly depends on the quality of the cell energy spectrum produced around this cut, with even apparently minor discrepancies potentially having a large effect on the number of hits above the threshold. For this reason, the differences observed between the number of hits in the GEANT4 and BIB-AE showers can be attributed to the more noticeable deviations in the description of the cell energy spectra observed below the MIP cut. These discrepancies increase for showers initiated by particles with higher incident energies as a result of the larger number of hits created in the shower.

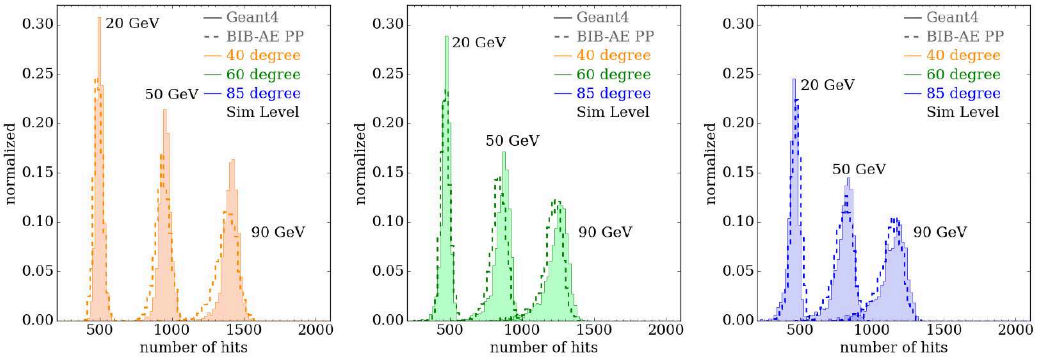


Figure 7.14: Simulation level number of hits for both Geant4 and BIB-AE generated showers. The distributions are grouped according to incident photon angles of 40 degrees (left, orange), 60 degrees (center, green) and 85 degrees (right, blue). Figure and caption reproduced from [1].

The number of hits present in the clusters after reconstruction are shown in Figure 7.15. No significant shift relative to the results obtained at simulation level is observed, indicating that the vast majority of hits above the 0.5 MIP cut are retained after reconstruction. This aligns with expectation for the electromagnetic showers explored here.

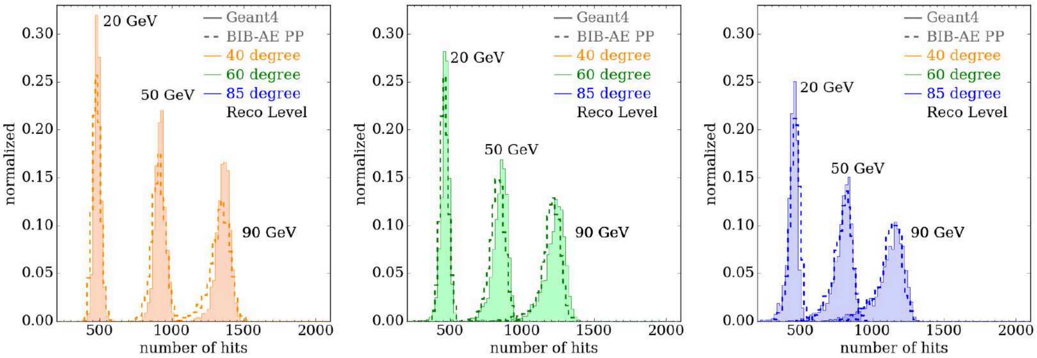


Figure 7.15: Reconstructed number of hits for both Geant4 and BIB-AE generated showers. The distributions are grouped according to incident photon angles of 40 degrees (left, orange), 60 degrees (center, green) and 85 degrees (right, blue). Figure and caption reproduced from [1].

Center of Gravity

We now turn our attention to shower shape observables, an accurate description of which is essential to prevent a large impact on downstream reconstruction. We begin by studying the performance of the model in describing the structure of showers along the depth of the calorimeter i.e. the z axis. Firstly, we consider the center of gravity of the shower along the depth of the calorimeter, which is obtained from the first moment along the z axis, and represents the position of the shower along the calorimeters' depth.

Figure 7.16 shows the center of gravity distributions along the z axis for the best (50 GeV, 85 degree) and worst (20 GeV, 85 degree) performing combinations of energy and angle in terms of JSD at simulation level, determined for the JSD between the corresponding Geant4 and BIB-AE distributions. Additional test points are included in Appendix A.2.2. Overall the BIB-AE is able to describe these distributions well, the most apparent deviation arising from

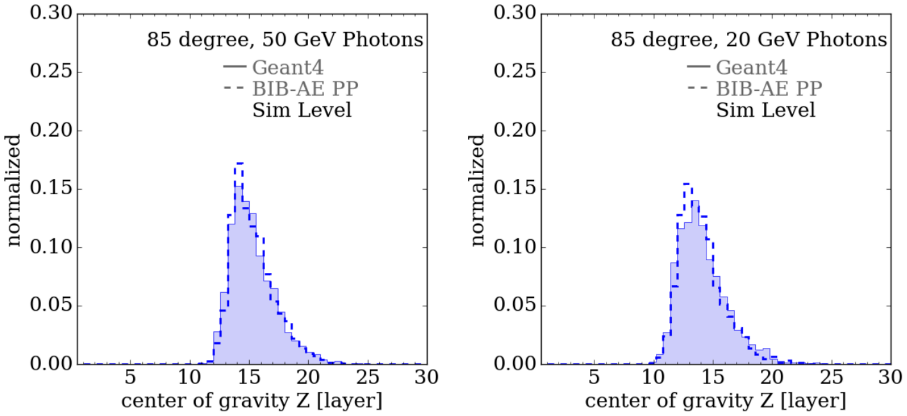


Figure 7.16: Simulation level center of gravity distributions for the best (50 GeV, 85 degrees, left) and worst (20 GeV, 85 degrees, right) incident angle and energy combinations. Figure and caption reproduced from [1].

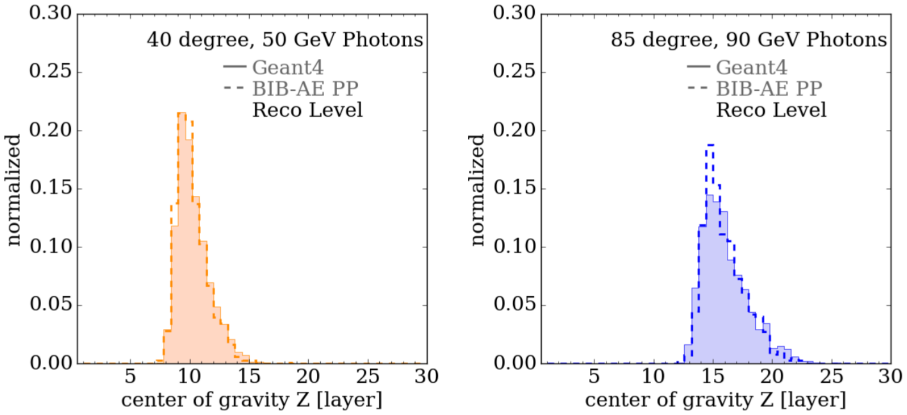


Figure 7.17: Reconstructed center of gravity distributions for the best (50 GeV, 40 degrees, left) and worst (90 GeV, 85 degrees, right) incident angle and energy combinations. Figure and caption reproduced from [1].

the BIB-AE tending to produce a slightly narrower distribution than GEANT4.

The resulting distributions for the center of gravity along the z axis after reconstruction are shown in Figure 7.17, with the best and worst performing incident energy and angle combinations now occurring at (50 GeV, 40 degrees) and (90 GeV, 85 degrees) respectively. In the best case, the distribution resulting from the BIB-AE matches the GEANT4 result very closely, while in the worst case the BIB-AE still produces a slightly narrower distribution.

Longitudinal Profile

The evolution of showers along the depth of the calorimeter is studied by means of the longitudinal profile. Throughout this thesis, this definition will be assumed when referring to the longitudinal profile, rather than the alternative definition of shower development along the principal axis. The best (90 GeV, 85 degrees) and worst (20 GeV, 40 degrees) performing combinations of incident angle in terms of JSD at simulation level are shown in Figure 7.18. Note that the apparent discontinuous step that occurs around layer 20 is a direct result of the calibration factor applied to the showers to account for the two sampling fractions present in the calorimeter (see Section 4.2). The prominence of this feature therefore increases signifi-

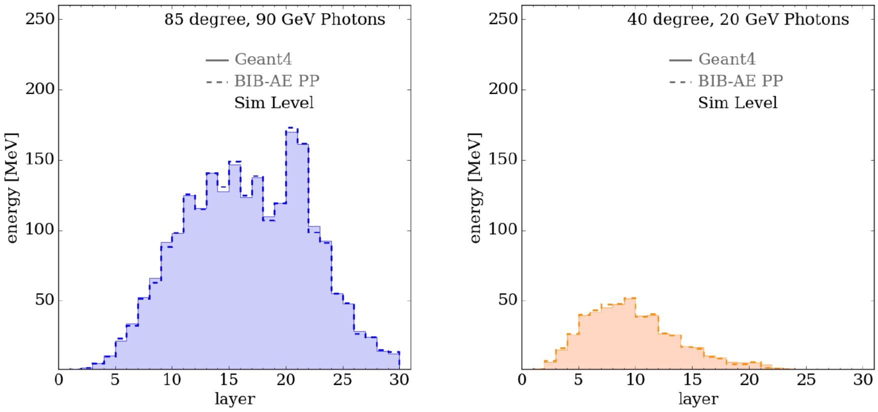


Figure 7.18: Simulation level longitudinal profiles for the best (90 GeV, 85 degrees, left) and worst (20 GeV, 40 degrees, right) incident angle and energy combinations. Figure and caption reproduced from [1].

cantly for showers with higher energy and closer to perpendicular incidence, as a result of more energy being deposited in the later layers of the calorimeter. The BIB-AE shows an impressive performance on reproducing this distribution across the range of incident energies and angle combinations studied. In particular, the BIB-AE is even able to accurately reproduce the step-like features that appear in the central layers for many combinations of incident energy and angle (as is the case for example with the 90 GeV, 85 degree showers present in the left panel of Figure 7.18). These signature steps arise from the alternating layer structure of the ECAL (see Section 4.2).

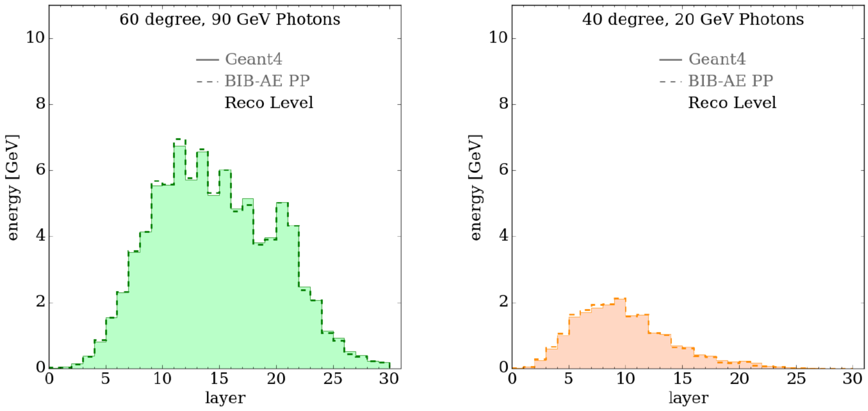


Figure 7.19: Reconstructed longitudinal profiles for the best (90 GeV, 60 degrees, left) and worst (20 GeV, 40 degrees, right) incident angle and energy combinations. Figure and caption reproduced from [1].

The longitudinal profiles for the best and worse combinations of incident angle and energy occur at (90 GeV, 60 degree) and (20 GeV, 40 degrees) after reconstruction, with these distributions being shown in Figure 7.19. The excellent performance of the BIB-AE in modelling this observable that was noted at simulation level is maintained after reconstruction across the range of test points studied.

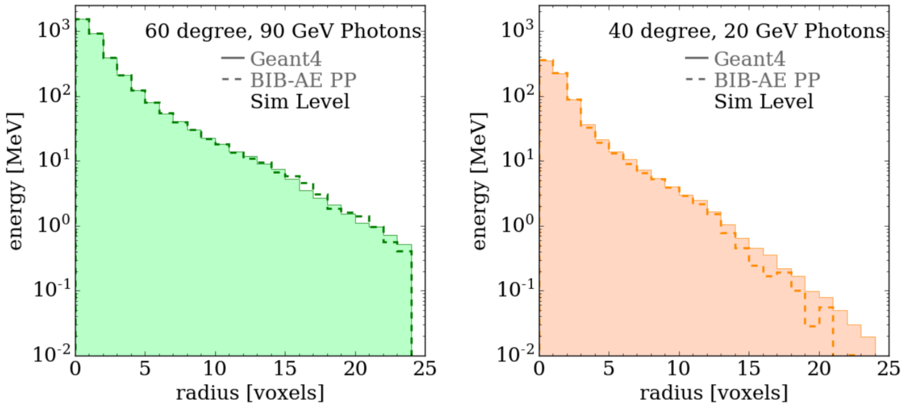


Figure 7.20: Simulation level radial profiles for the best (90 GeV, 60 degrees, left) and worst (20 GeV, 40 degrees, right) incident angle and energy combinations. Figure and caption reproduced from [1].

Radial Profile

As a final measure of shower shapes, the quality of modelling the transversal development of showers is investigated by means of the radial profile around the shower axis. Distributions for the best (90 GeV, 60 degrees) and worst (20 GeV, 40 degrees) performing combinations of incident and angle in terms of JSD are shown at simulation level in Figure 7.20. The core of the shower is well modelled across the range of test points studied, while at some test points (typically for showers with a higher inclination), deviations are observed at larger radii, with the tails of the BIB-AE distributions falling off quicker than the equivalent region of GEANT4 showers.

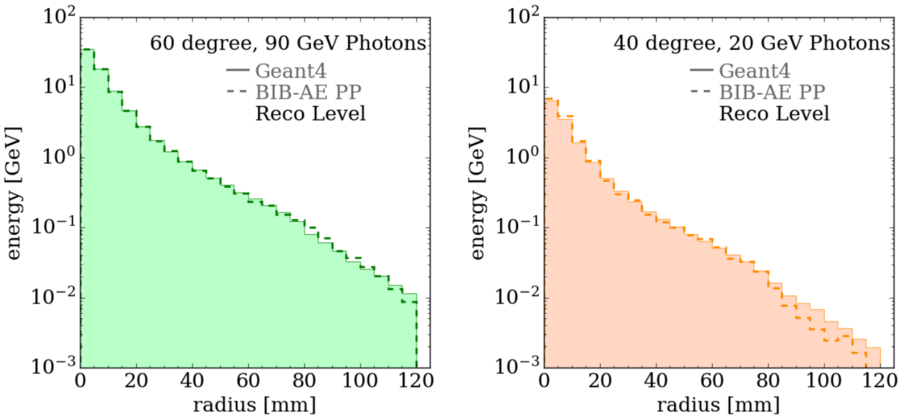


Figure 7.21: Reconstructed radial profiles for the best (90 GeV, 60 degrees, left) and worst (20 GeV and 40 degrees, right) incident angle and energy combinations. Figure and caption reproduced from [1].

The radial profiles for the best (90 GeV, 60 degrees) and worst (20 GeV, 40 degrees) combinations of incident energy and angle in terms of JSD after reconstruction are shown in Figure 7.21. The modelling of the radial profile around the core of the shower remains accurate after reconstruction, with the best performing combination of incident energy and angle also displaying a strong agreement between GEANT4 and BIB-AE at larger radii. Notably, for cases where there is a pronounced mismodelling of the profile for larger radii, reconstruction appears to reduce the level of mismatch that occurs.

7.4.2 Computational Performance

Since the ultimate usage of a generative model for fast simulation relies on being able to provide a reduction in the time required to generate showers, we finally benchmark the computational performance of the model in terms of the inference time per shower in comparison to GEANT4. Due to the sequential nature of the Monte Carlo simulation approach used in GEANT4, the computational performance of this approach is only benchmarked on CPU hardware, while the performance of the BIB-AE can be benchmarked on both CPU and GPU hardware. It should be noted that the timings for GEANT4 additionally include navigation and placement in the detector geometry and additional computational effects that cannot be replicated in the BIB-AE timings, which at this stage are performed in a pythonic software environment and in the $(30 \times 60 \times 30)$ regular grid which is produced by the BIB-AE.

Table 7.2 shows the average time to generate a shower with energy uniformly distributed in the range of $10 - 100$ GeV and angles in the range of $90 - 30$ degrees using GEANT4 and the BIB-AE. The generative model offers a significant speedup relative to GEANT4, reaching up to three orders of magnitude on a GPU.

Table 7.2: Comparison of the computational performance of the BIB-AE generator and GEANT4 on a single core of an Intel® Xeon® CPU E5-2640 v4 (CPU) and an NVIDIA® A100 with 40 GB of memory (GPU). For the BIB-AE, the best performing batch size is selected. The value shown is the mean obtained for a set of 2000 showers with uniform energy from $10 - 100$ GeV and $30 - 90$ degrees, with error arising from the standard deviation. Table and caption reproduced from [1].

Hardware	Simulator	Time / Shower [ms]	Speed-up
CPU	GEANT4	4417 ± 83	$\times 1$
	BIB-AE	362 ± 2	$\times 12$
GPU	BIB-AE	4.32 ± 0.09	$\times 1022$

7.5 Conclusions

In this Section, we have developed generative models for highly granular calorimeter simulation in two key directions. In the first instance, we have extended the BIB-AE which is one of the best performing models for the task of calorimeter simulation, to a generalised scenario which requires conditioning on both incident particle energy and angle. In a second step, we have additionally performed a detailed study into the effects on key physics observables after interfacing with the full reconstruction pipeline used in ILD.

Importantly, the model has been shown to not only provide an accurate description of key physics observables in this generalised simulation scenario, but to also maintain this performance after reconstruction is applied. The energy conditioning performance of the model was shown to be particularly strong thanks to the per-shower re-scaling that was afforded by the use of an NF to model the latent space. The angular conditioning was, while somewhat weaker in comparison, still found to provide a good description of the distributions. The ability of the BIB-AE to describe the cell energy spectrum was retained across the range of incident energies and angles, as was the description of the different shower shapes. A notable exception to this final remark are the differences that appear at larger radii for more inclined showers.

The extensions to the model, notably the modelling of the latent space with an NF have also allowed the computational performance of the model to remain strong even with the larger regular grid required to contain the showers incident under the the most inclined angles.

While this indicates a strong performance from the model in the scenario presented, in order to broaden the scope of the study, a number of key limitations must be noted. Firstly, this Section has only investigated the conditioning of the model on a single angle of incidence. In order to produce a general tool, conditioning on a second angle of incidence will be required.

Secondly, all the showers investigated were generated at a fixed point in the calorimeter. Since the calorimeter geometry is not perfectly regular, this means that at different incidence points the local geometry will appear different. Generalising to different incidence points would also make the use of the map look-up that was used in this Section to place the shower back into the detector geometry impractical, as an individual muon sample and corresponding map would have to be created for each possible incidence point.

Thirdly, while the performance of the model after reconstruction was found to be strong, only single particle reconstruction has been studied. This is in part restricted by the aforementioned map look up. In a real physics simulation, showers from different particles will overlap, and the reconstruction problem becomes significantly more challenging. Notably, when separating overlapping showers, one can imagine the radial profile of showers being crucial, meaning that the deviations observed for more inclined showers at larger radii could have significant a significant impact.

Finally, the computational benchmark presented in this Section, which is also routinely used in the literature, is not a fair comparison between GEANT4 and the BIB-AE, as the GEANT4 timings necessarily also include additional effects from running in the full simulation pipeline and in the real detector geometry. In order to provide a realistic computational benchmark, and to also investigate the more complicated reconstruction scenarios present in real physics events, generative models will have to be incorporated into the full simulation pipelines commonly used in high energy physics.

Tackling these challenges will be the focus of the remainder of this thesis.

Chapter 8

A Fully Conditioned Generative Model for Electromagnetic Shower Simulation

Having established a robust conditioning scheme for the BIB-AE in Chapter 7, we now turn our attention to the development of a fully conditioned model. This means that the model must be conditioned on two angles, as well as the energy of the incident particle. Varying another incident angle to the calorimeter face also has implications for the size of the regular grid, which must be large enough to contain maximally inclined showers of the highest incident energies. To maintain the increased simulation speed of the generative model, the grid size must be kept as small as possible. As this model will ultimately be used in a complete calorimeter subsystem, it must also be able to simulate different incident positions in the calorimeter, meaning steps must also be taken to handle the irregular geometry present in the calorimeter. This chapter will focus on minimising the effects of an irregular geometry for training a model, while the effects of using a regular grid model to simulate showers in an irregular grid will be studied in Chapter 10.

Section 8.1 will address the creation of the dataset, including the creation of a regular detector geometry and the method employed to minimise the size of the grid while retaining information about the incidence position. In Section 8.2, the extension of the BIB-AE to allow conditioning on an additional incident angle will be described, while the performance of the model will be studied in Section 8.3.

8.1 Dataset

8.1.1 Creation of a Regular ECAL Geometry

In order to create an appropriate dataset for training a fully conditioned generative model, a different approach to that taken in Chapter 7 must be adopted. Consider the physical geometry present in the ILD ECAL, a visualisation of which is shown for two layers of active material in the left of Figure 8.1. The detector geometry is irregular, and features insensitive volumes arising from readout electronics, as well as gaps in the detector and structural supports which are necessary to build a realistic physical detector, as described in Chapter 3. Insensitive volumes arising from the construction of the detector can be aligned between layers, while the position of the insensitive volumes at the edges of wafers varies between layers, with a staggering effect causing an offset between the placement of cells from one layer to the next. This was the geometry used to create the training dataset used in Chapter 7.

In that case, the irregular geometry could be projected to a regular grid because a fixed incident position to the face of the calorimeter was chosen. The procedure of removing

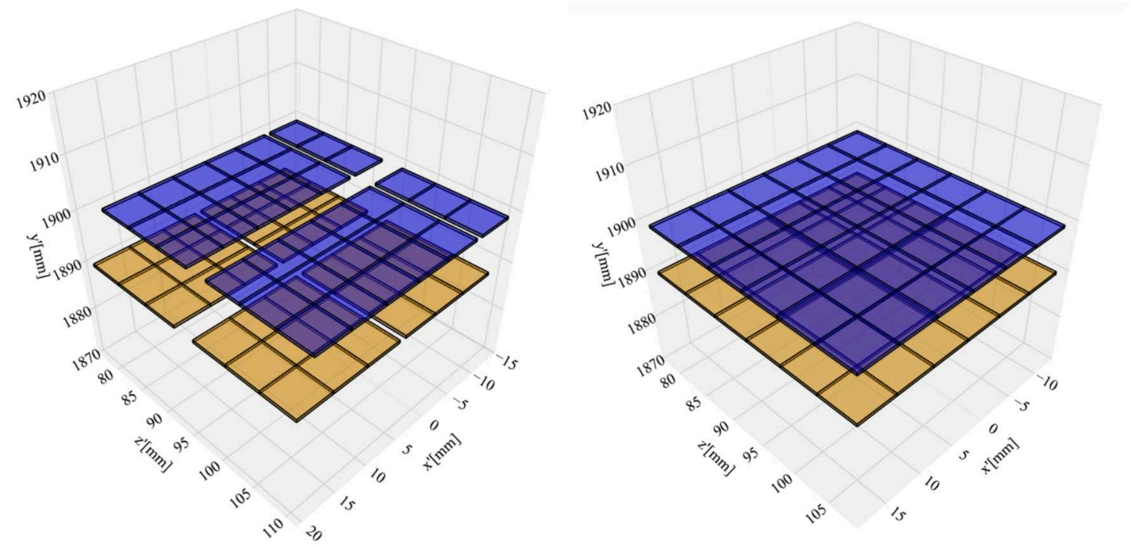


Figure 8.1: Visualisation of geometry maps for (left) the physical geometry and (right) the regular geometry for a section of two sensitive layers in the calorimeter. The physical geometry includes gaps between the cells arising from insensitive volumes such as structural supports and readout electronics, as well as a staggering effect between layers. The regular geometry consists purely of sensitive material, with the cells being perfectly aligned from one layer to the next[1].

artefcats from the projection — which appear as rows of empty cells — was therefore a constant correction. However, when further generalising the scope of the simulation to an arbitrary incident position on the face of the calorimeter, the local geometry near a given incidence point varies due to the irregular nature of the geometry. This means that a different artefact removal procedure would have to be applied for every individual position, which is clearly an approach that is practically infeasible.

Another problem arising from training the generative model on an irregular geometry is the fraction of the active layers that is sensitive. Since the local geometry changes with the incident position on the calorimeter face, the relative abundance of insensitive volumes within the local region of an active layer also varies. This means that if the model is trained on a particular position, any energy depositions that occur within insensitive volumes in the active layer are lost. As a result, using the model to simulate showers at different incident positions is not feasible, as sensitive cells and insensitive volumes are positioned differently. In order to correctly handle this when training on an irregular geometry, training would have to occur over all possible incident positions to the calorimeter face. This is again practically infeasible.

To handle these problems, a regular version of the ILD ECAL was created for use in training the model. A visualisation of the geometry of two layers in the regular ECAL is shown in Figure 8.1. The geometry is created such that only the segmentation in the sensitive layers is altered, with the detector consisting of exactly the same structure in terms of material composition and placement longitudinally (i.e. along the depth of the calorimeter). With the regular segmentation in the sensitive layers, cells are now arranged in a perfect grid meaning that there are no insensitive volumes within the active layers, so all the energy deposited in an active layer is recorded by GEANT4. This effect can be seen in Figure 8.2 (top left), from the total visible energy deposited in the calorimeter for 50 GeV photons incident perpendicularly to the face of the calorimeter. The distribution for the realistic, irregular geometry (grey, filled histogram) is compared to that of the regular geometry (blue, unfilled histogram). More energy is visible

---

[1] The scripts used to produce this visualisation were provided by Anatolii Korol

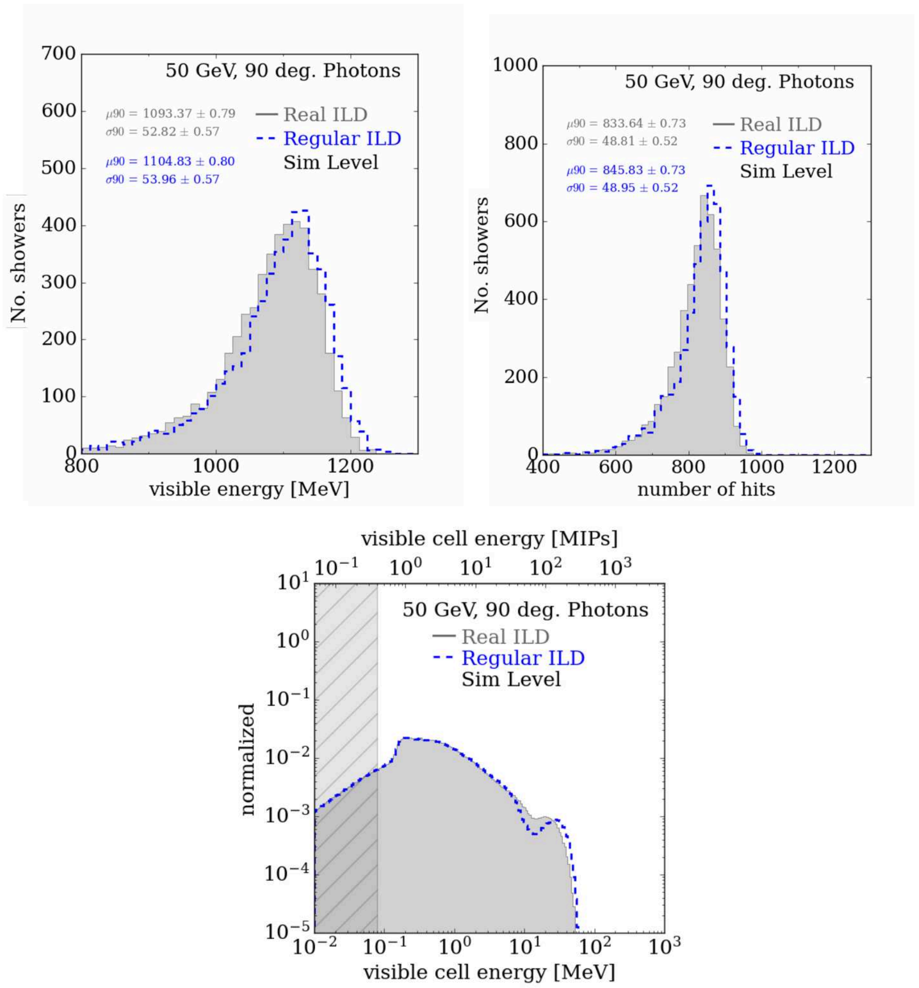


Figure 8.2: Comparison of distributions for 50 GeV photons incident orthogonal to the face of the calorimeter for the real, irregular geometry of the ILD ECAL (grey, filled distributions) and the regular ECAL geometry (blue, unfilled distributions) used to create the training data set. The distributions shown are for the total visible energy deposited in the calorimeter (top, left), the number of hits above threshold (top, right) and the cell energy spectrum (bottom). For the number of hits and the visible energy sum, the mean (μ_{90}) and standard deviation (σ_{90}) calculated on the central 90% of the distribution are shown.

in the regular detector, as the active layer is completely sensitive. The regular segmentation also means that more hits (Figure 8.2, top right) are created in the regular geometry than the irregular one. With all the energy deposited in the active layers being retained, a shower could be placed back into the physical geometry with Geant4 itself dropping energy depositions that fall into insensitive volumes within an active layer.

The staggering effect present in the irregular geometry is also removed, with cells now perfectly aligned from one layer to the next. The effect of this is visible in the cell energy spectrum produced in the detector, which is shown in Figure 8.2 (bottom). The dip (around 10 MeV for 50 GeV photons incident at 90 degrees) and subsequent peak at higher energies is significantly more pronounced in the regular geometry than in the irregular physical geometry. This is a result of the more energetic particles near the core of the shower passing through the same positions in the aligned cells in the regular geometry, whereas the staggering smears the energy depositions in the regular geometry. The lack of staggering means that no correction has to be applied for artefacts created by the projection from the irregular physical geometry into a regular grid.

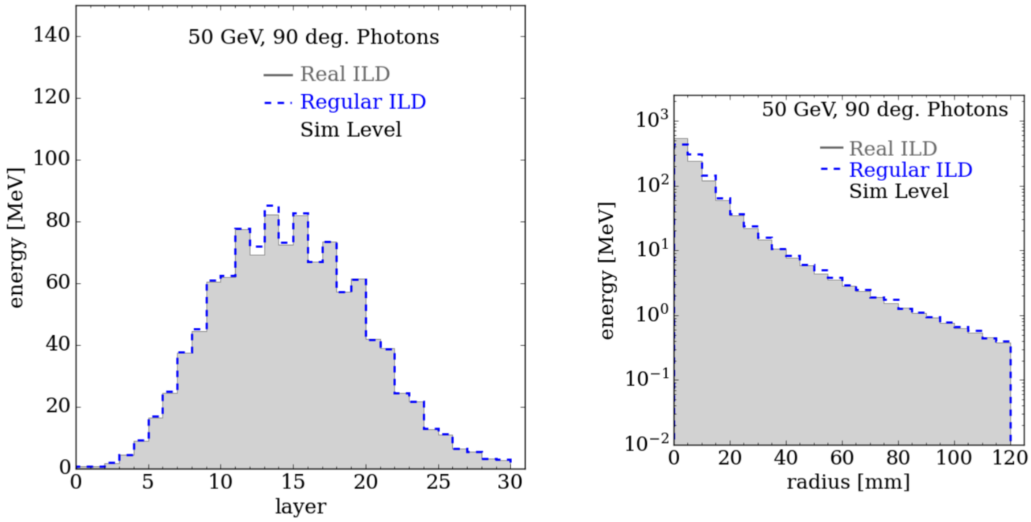


Figure 8.3: Comparison of distributions for 50 GeV photons incident orthogonal to the face of the calorimeter for the real, irregular geometry of the ILD ECAL (grey, filled distributions) and the regular ECAL geometry (blue, unfilled distributions) used to create the training data set. The distributions shown are the longitudinal profile (left) and the radial profile (right).

Finally, the effect of using the regular geometry on shower shapes is shown in Figure 8.3. In terms of the longitudinal profile (left), noticeably more energy is deposited in the regular detector for certain layers. The size of this effect in a given layer depends on the position of insensitive volume within a layer with respect to the higher energy components of a shower. For the radial profile (right), the largest effects occur closer to the shower core, where the staggering effect again plays a role in the exact distribution of energy radially.

8.1.2 Creation of the Dataset

The dataset used in this study was produced with the same software environment described in Chapter 7, using the ILCSoft [126] ecosystem (see Section 4.2.2), with full simulation performed with Geant4 [25] version 10.4, using the QGSP_BERT physics list and DD4hep [129] version 1.11. The electromagnetic showers are again initiated by photons, using a regular segmentation in the active layers of the ILD ECAL as described in Section 8.1.1.

Since two angles must now be varied, using a fixed incident position at the face of the calorimeter would require the grid size to be extended significantly in order to contain maximally inclined showers of the highest energy. As extending the grid size in this fashion would drastically increase the number of channels which the model would have to simulate, the inference time required to generate a shower with the model becomes prohibitive. The size of the grid required can be reduced by allowing the incident position on the calorimeter face to vary, however in order to be able to place a simulated shower back into the detector geometry, as will be done in Chapter 9, the incident position on the calorimeter face must be known. In order to satisfy these requirements, the following approach was adopted.

An illustration showing how the incident angle and position on the face of the calorimeter were determined is shown in Figure 8.4. The incidence position on the calorimeter face was allowed to vary with the angle of incidence. The flight direction was defined by the angle of incidence, and aligned such that the flight direction passed through a fixed point in the center of the calorimeter, at a position of $(x', y', z') = (0\ \text{mm}, 1889\ \text{mm}, -100\ \text{mm})$ in the

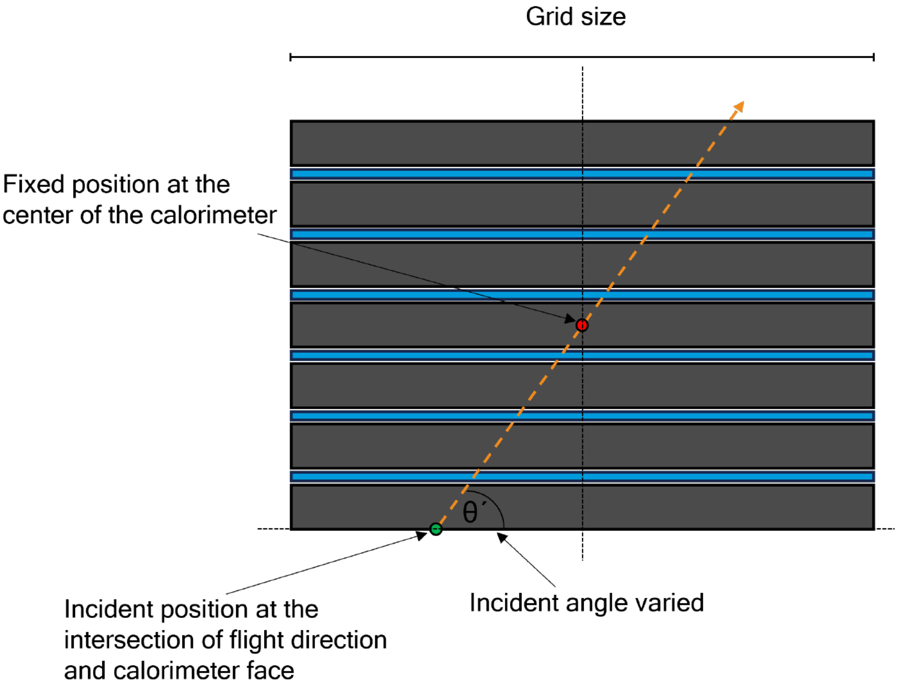

Grid size
Fixed position at the center of the calorimeter
θ´
Incident position at the intersection of flight direction and calorimeter face
Incident angle varied

Figure 8.4: Illustration of the approach adopted for varying the incident angle of the photons to the face of the calorimeter. The incident photons are always created directly at the face of the calorimeter, and the flight direction (orange dashed line) of the particle is defined by the angle of incidence. The flight direction of the incident photon always passes through the same point in the center of the calorimeter (red dot), independent of the angle of incidence. This means that the incident position on the calorimeter face (green dot) can be calculated directly from the intersection between the direction of flight and the plane of the calorimeter face. For illustrative purposes only one angle is shown, while in practice two angles were varied when creating the dataset.

ILD coordinate system, independent of the angle of incidence. The incident position on the face of the calorimeter could then be calculated by intersecting the direction of flight of the particle with the plane lying on the face of the calorimeter. This means that all photons were created directly at the face of the calorimeter. For illustrative purposes, Figure 8.4 only shows the variation of a single angle, when in practice this approach was used for two angles of incidence.

To create the training dataset, the incident energy E was varied uniformly in the range of 5 to 126 GeV, along with both the global polar incident angle θ', which was varied uniformly from 30 to 95 degrees and the global azimuthal incident angle ϕ', which was varied uniformly from 65 to 95 degrees. The range of energies and angles was chosen to cover the majority of the kinematic range for particles produced at the IP incident to the ILD ECAL for the ILC operating at a center of mass energy of 250 GeV, with an additional overshoot of the maximum expected values incorporated to prevent edge effects occurring at the boundaries of the phase space. A total of 500k showers composed the training dataset. To evaluate the performance of the model, fixed combinations of incident energies and angles were used, as in Chapter 7. However in this case, a total of 27 datasets were created, each consisting of 2000 showers with fixed combinations of incident energies ($\{10, 50, 100\}$ GeV), incident global polar angles ($\{40, 60, 90\}$ degrees) and incident global azimuthal angles ($\{70, 80, 90\}$ degrees). These points cover a broad range of the phase space, while also avoiding the very edges of the training range. Another 27 independent datasets with an identical number of showers and combinations of incident energies and angles were used as validation datasets to select the best performing state of the model. As in Chapter 7, a fixed calibration factor was applied to scale the hit energies in the last 10 layers of the calorimeter to account for the modification of the sampling fraction caused by the absorber layers being twice as thick in this region (see Section 4.2.1).

Calorimeter shower images were again created by projecting the physical cells in the active layers of the calorimeter into a regular grid of size $(x, y, z) = 30\times49\times30$. Since the dataset was created using an ECAL geometry with a regular segmentation in the active layers, no corrections for artefacts were necessary. The grid size was chosen such that on average approximately 99% of the visible energy deposited in the detector was contained for the particles incident with the highest energies and most inclined angles in the dataset.

8.2 Extensions to the Generative Model

The adaption of the BIB-AE to handle an additional conditioning parameter constitutes a straightforward extension to the model developed in Chapter 7[2]. The conditioning of the components of the model is identical, but now the second angle is appended to the vector of conditioning parameters used. The Post Processor network is again also conditioned on the visible energy sum. Due to the different range of energy and angle values in this fully conditioned scenario, a different scaling was necessary for the conditioning labels provided to the normalising flow. In this case, the incident energy labels were divided by a factor of 126, and the incident polar and azimuthal angles were divided by a factor of 1.658063. These factors correspond to the maximum values for the respective conditioning parameters. The visible energy, which was concatenated with the latent variables modelled by the NF, was divided by a factor of 3300. All models were trained in the same manner, and with the same optimiser parameters, as described in Chapter 7. The best epoch of the training was again selected by means of a scan over epochs using the validation datasets, with half of the validation data being passed through the encoder-decoder pair of the BIB-AE, and the remaining half being retained for comparison. For each epoch, eight key calorimeteric observables were considered (polar angle response, azimuthal angle response, energy response, cell energy spectrum, number of hits above threshold, center of gravity, longitudinal profile and radial profile). A bin-wise area difference was again used to quantify the performance of the network in modelling each observable at each epoch. This approach was used to determine the best epoch of BIB-AE training, after which training of the Post Processor network was started. A similar procedure was adopted to ascertain the best performing epoch of Post Processing. The BIB-AE was trained for a total of 51 epochs, after which the model was frozen and the Post Processor trained for a further 49 epochs. The NF model was trained for a total of 199 epochs, with the best epoch being selected from the lowest loss value.

8.3 Results

We now perform an evaluation of the physics performance of the fully conditioned BIB-AE model in comparison to GEANT4. This comparison will be performed at *simulation level*, and focus on the key calorimetric observables already investigated in Chapter 7 for the BIB-AE model conditioned on energy and a single angle.

---

[2] The implementation of the model and the hyperparameters used for training are available at `https://github.com/FLC-QU-hep/Two_Angle_BIBAE`

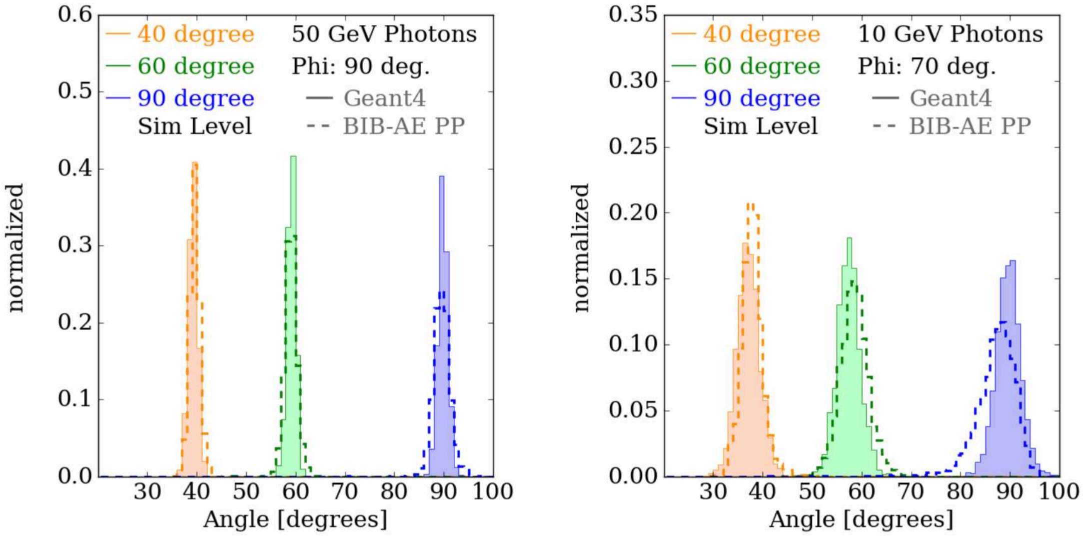


Figure 8.5: Angular response in theta at simulation level for both GEANT4 (filled histograms) and BIB-AE generated photon showers (dashed, unfilled histograms). Distributions are shown for an incident energy of 50 GeV and phi angle of 90 degrees (left) and an incident energy of 10 GeV and phi angle of 70 degrees (right). Figures are shown for each incident theta angle of 40 degrees (orange), 60 degrees (green) and 90 degrees (blue).

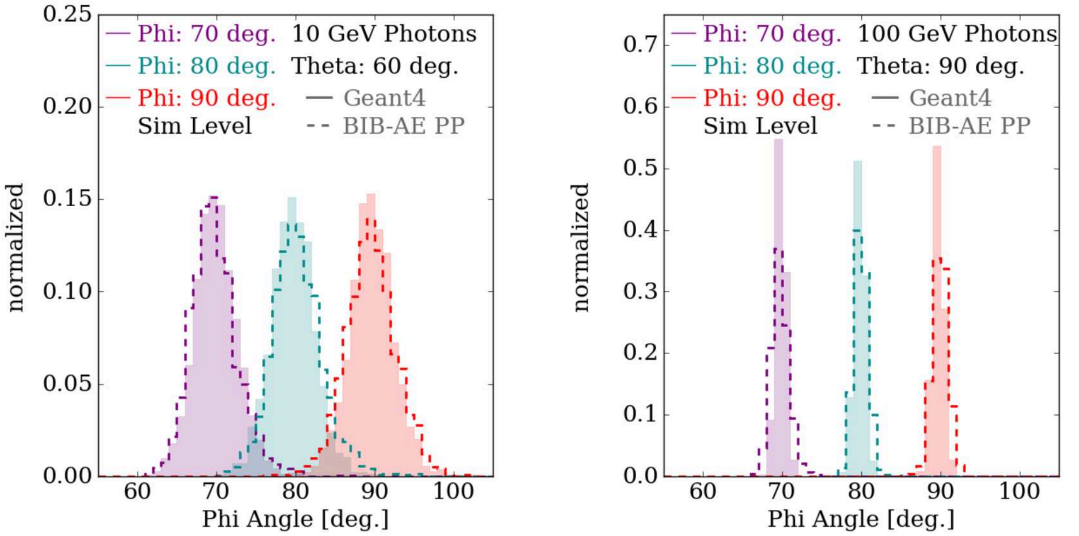


Figure 8.6: Angular response in phi at simulation level for both GEANT4 (filled histograms) and BIB-AE generated photon showers (dashed, unfilled histograms). Distributions are shown for an incident energy of 10 GeV and theta angle of 60 degrees (left) and an incident energy of 100 GeV and theta angle of 90 degrees (right). Figures are shown for each incident phi angle of 70 degrees (purple), 80 degrees (cyan) and 90 degrees (red).

The only processing applied to the showers is the mapping of hits with an energy deposition of less than 0.07875 MeV, which corresponds to an energy deposition of less than half a MIP, to zero. As stated in Chapter 7, this emulates the processing applied in a real calorimeter, where hits with energy less than this threshold lie below the noise floor of the detector. The comparison between the BIB-AE and Geant4 involves each of the 27 test points consisting of fixed combinations of incident angles and energies, as described in Section 8.1.2. Due to the large number of test points considered, only a subset of distributions for each observable will be shown. Additional test points are included in Appendix B. A comparison of the computational performance of the BIB-AE to Geant4 will be presented in Chapter 9, after the model has been integrated into the software chain, and a fair comparison can be made. In this section, a consistent colour scheme will be used to represent the various fixed incident angles. For the polar angle (theta), the test points are 40 degrees (orange), 60 degrees (green) and 90 degrees (blue). For the azimuthal angle (phi), the test points are 70 degrees (purple), 80 degrees (cyan) and 90 degrees (red).

Angular Response

The angular response is again characterised by applying a principal component analysis (PCA) to obtain the principal axis of the showers. The resulting angular distributions of the angle in theta for photons incident with 50 GeV and a phi angle of 90 degrees (left) and photons incident with 10 GeV and a phi angle of 70 degrees (right) are shown in Figure 8.5. The quality of modelling of the theta angle is quite variable — for example the angular response for the theta 40, phi 90 degree, 50 GeV showers is modelled well by the BIB-AE. However, there is noticeable mismodelling of the sharp peaks that are more pronounced for higher energy photons. Additionally, at the lower incident energies that are probed in this study, some mismodellings in the centering of the distributions are clear. This is the case for the theta 90 degree, phi 70 degree, 10 GeV showers where the BIB-AE produces not only a broader distribution, but also one that exhibit a significant bias towards lower angles. Overall, the modelling of the angle in theta is not as strong as that observed for the model conditioned on a single angle in the more constrained problem that was explored in Chapter 7. The angular distributions of the angle in phi for photons incident with 10 GeV at a theta angle of 60 degrees (left) and with 100 GeV at a theta angle of 90 degrees (right) are shown in Figure 8.6. Some distributions, such as for the 10 GeV, theta 60 degree photons shown in the left of the figure, are well modelled. However, biases and mismatches in distribution widths are observed for other test points, such as for the 100 GeV, theta 90 degree photons shown in the right of the figure.

Energy Response

The energy response of the BIB-AE model is again characterised at simulation level by the sum of all energy depositions above the 0.5 MIP threshold in the sensitive regions of the calorimeter, i.e. the visible energy. The distributions obtained for both Geant4 and the BIB-AE for a theta angle of 90 degrees and a phi angle of 90 degrees, as well as for a theta angle of 40 degrees and a phi angle of 70 degrees are shown in Figure 8.7. Across the range of energies and angles studied, a strong agreement between the distributions produced by the BIB-AE and Geant4 is observed. This is similar to what was observed in the single angle and energy conditioning scenario in Chapter 7, since the task of the NF in terms of modelling the energy sum is not significantly harder even with a more challenging conditioning scenario.

Cell Energy Spectrum

The cell energy spectrum is the next calorimetric observable studied. The simulation level cell energy spectra for the best (100 GeV, theta 60 degrees, phi 90 degrees) and worst (10 GeV,

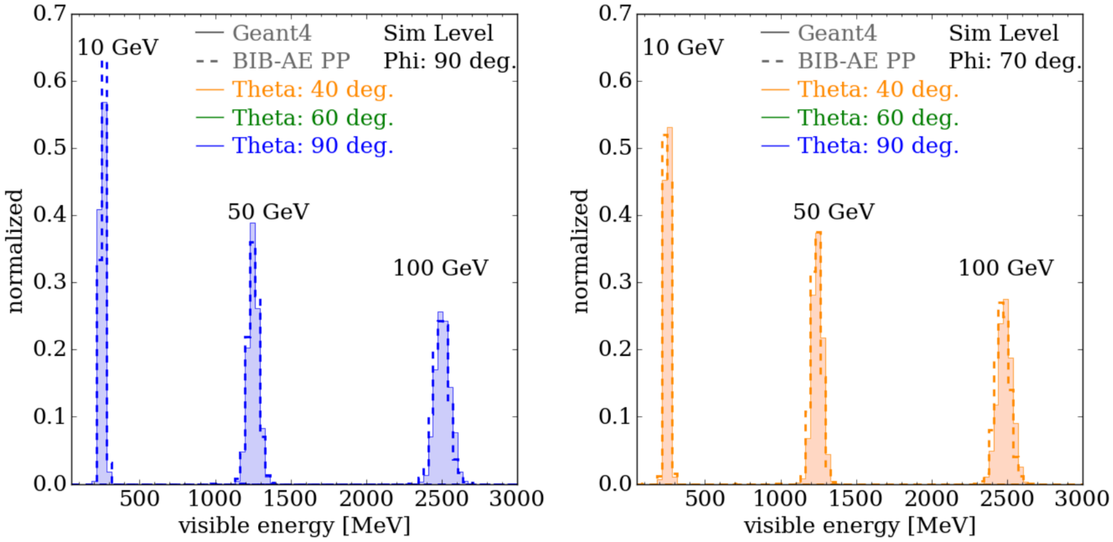


Figure 8.7: Visible energy deposited in the calorimeter at simulation level for both GEANT4 and BIB-AE generated showers. Distributions are shown for each incident energy of 10 GeV, 50 GeV and 100 GeV, for a theta angle of 90 degrees and a phi angle of 90 degrees (left) and for a theta angle of 40 and a phi angle of 70 degrees (right).

theta 40 degrees, phi 80 degrees) incident energy and angle combinations, selected according to the JSD value (Equation 7.10) between the GEANT4 and BIB-AE distributions, are shown in Figure 8.8. The grey hatched region in the left of each figure lies below the half MIP threshold, and hence would be removed in a real calorimeter. For this reason, the JSD value was only calculated for the distributions in the region above the half MIP threshold. The Post Processor Network plays a key role in how the BIB-AE models this distribution. The BIB-AE performance is worst for the low energy 10 GeV photons, which have not been studied as a test point with this model previously as this energy value lay right on the edge of the conditioning range. The modelling of the MIP peak for this energy shows clear deviations, with the BIB-AE producing a much smoother distribution in this region. However, for higher incident energies, the modelling of the spectrum is much better, including for the test points at 100 GeV which was previously at the upper boundary of the conditioning range. For these higher energies (50 and 100 GeV), the largest discrepancies appear in the high energy tail of the spectrum, where the BIB-AE often produces a spectrum that falls off too sharply compared to the GEANT4 distribution.

Number of Hits

As has previously been noted in Chapter 7, the Post Processor Network also has an important role to play in modelling the number of hits above the MIP cut. Distributions for the number of hits above threshold at simulation level for a theta angle of 60 degrees and a phi angle of 70 degrees and for a theta angle of 40 degrees and a phi angle of 80 degrees are shown for each incident energy in Figure 8.9. In the first instance, the number of hits are relatively well modelled at each of the incident energies shown, with the most noticeable deviations occurring for photons incident with energies of 50 GeV. In the latter instance, a significantly poorer modelling of the GEANT4 distributions by the BIB-AE is observed. In this case, the BIB-AE tends to produce more hits above threshold than GEANT4, with a clear relative shift in the distributions present that becomes more apparent as the incident energy of the photons is increased and more hits are present in the shower.

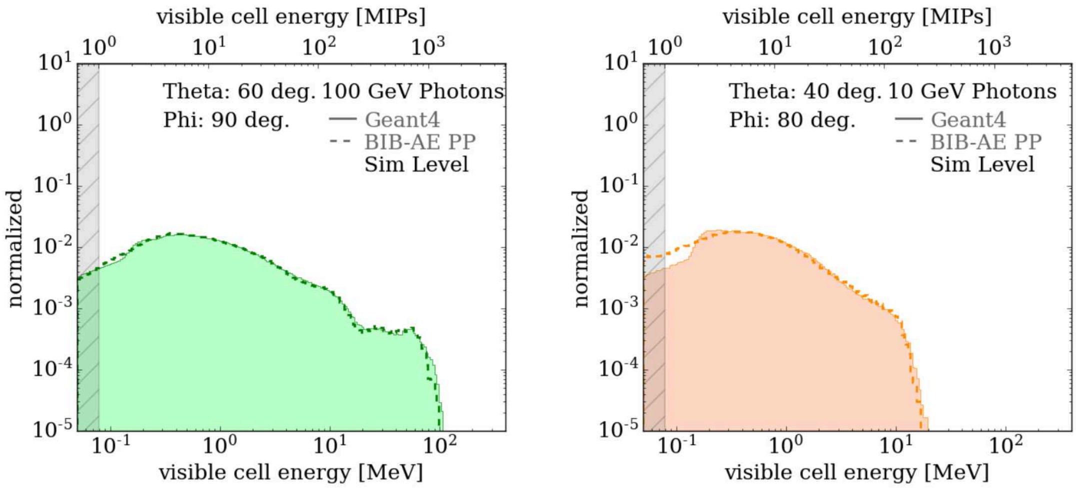


Figure 8.8: Cell energy spectra at simulation level for the best (100 GeV, theta 60 degrees, phi 90 degrees, left) and the worst (10 GeV, theta 40 degrees, phi 80 degrees, right) incident energy and angle combinations. The grey hatched area indicates the region below half a MIP.

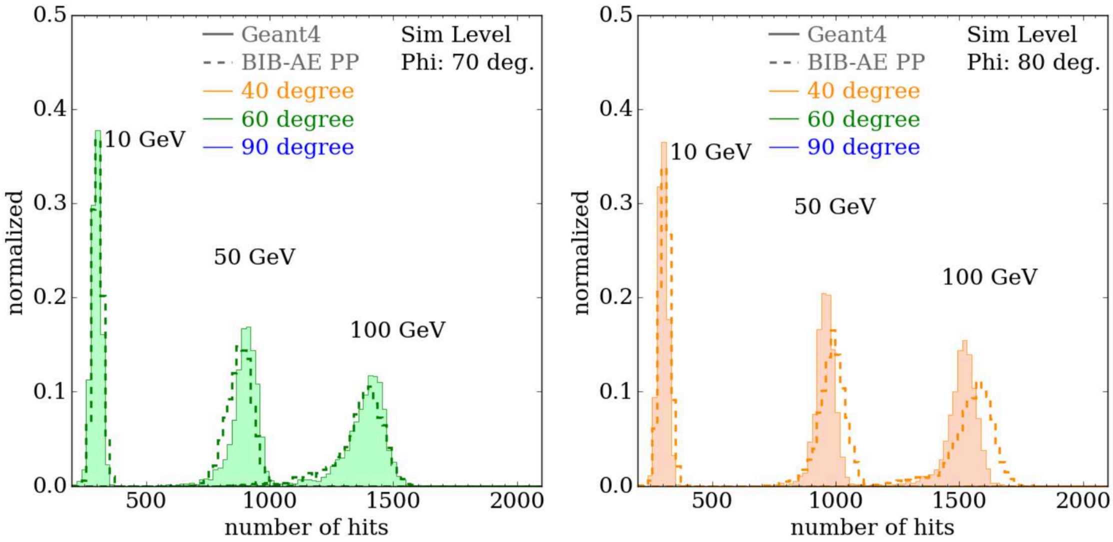


Figure 8.9: Number of hits at simulation level for both Geant4 and BIB-AE generated showers. The distributions are shown for each incident energy of 10, 50 and 100 GeV, for a theta angle of 60 degrees and a phi angle of 70 degrees (left) and for a theta angle of 40 degrees and a phi angle of 80 degrees (right).

Center of Gravity

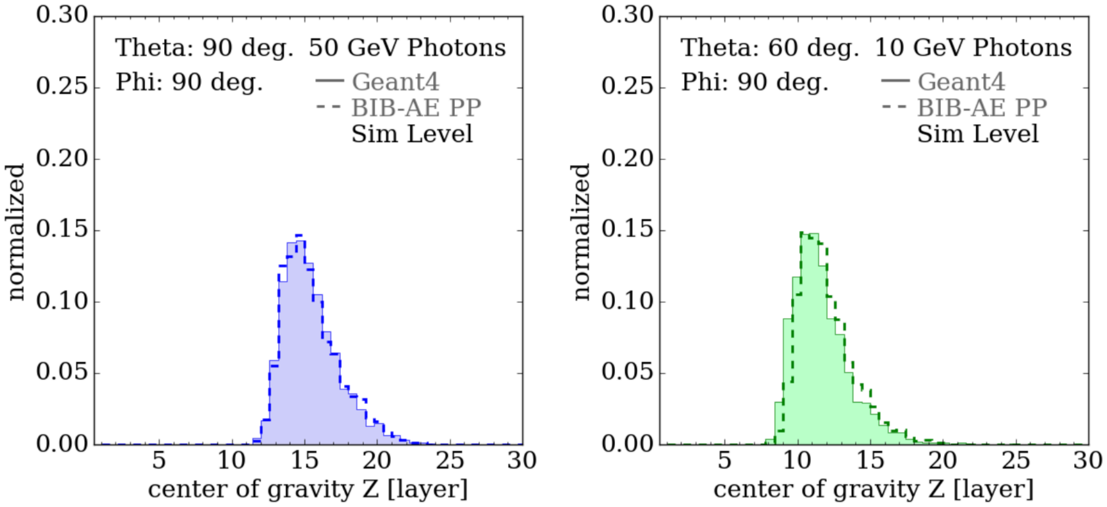


Figure 8.10: Center of gravity distributions in z at simulation level for the best (50 GeV, theta 90 degrees, phi 90 degrees, left) and worst (10 GeV, theta 60 degrees, phi 90 degrees, right) incident energy and angle combinations.

We begin our study of shower shape observables with the center of gravity of the shower along the depth of the calorimeter i.e. the z axis. This observable is obtained for each shower from the first moment along the z axis, and as such represents the position of the shower along the calorimeters' depth. Distributions of the center of gravity along the z axis are shown in Figure 8.10 for the best (50 GeV, theta 90 degrees, phi 90 degrees, left) and worst (10 GeV, theta 60 degrees, phi 90 degrees, right) incident energy and angle combinations, selected according to the JSD value (Equation 7.10) between the GEANT4 and BIB-AE distributions. In the best case, the BIB-AE produces a very good description of the GEANT4 distribution, with only minor deviations visible. In the worst case, the BIB-AE shows more noticeable deviations from the GEANT4 distribution, with the BIB-AE tending to produce the center of gravity of a shower further into the depth of the calorimeter, resulting in a distribution which is slightly shifted to higher centers of gravity relative to GEANT4.

Longitudinal Profile

The next shower shape observable studied is the longitudinal profile into the depth of the calorimeter. Distributions for the best (100 GeV, theta 60 degrees, phi 80 degrees) and worst (10 GeV, theta 40 degrees, phi 80 degrees) incident energy and angle combinations, selected according to the JSD value (Equation 7.10) between the GEANT4 and BIB-AE distributions, are shown in Figure 8.11. As in Chapter 7. The apparently discontinuous step observed around layer 20 again arises from the calibration factor applied to the showers to account for the two sampling fractions present in the calorimeter. The BIB-AE maintains an impressive level of performance in modelling this distribution, with the model still providing an accurate reproduction of the step-like features arising from the alternating structure of the calorimeter. This is true across the range of energies and angles that were investigated.

Radial Profile

Finally, the radial profile around the shower axis is shown as a means to investigate the transversal shape of showers. Distributions for the best (100 GeV, theta 60 degrees, phi 80 degrees)

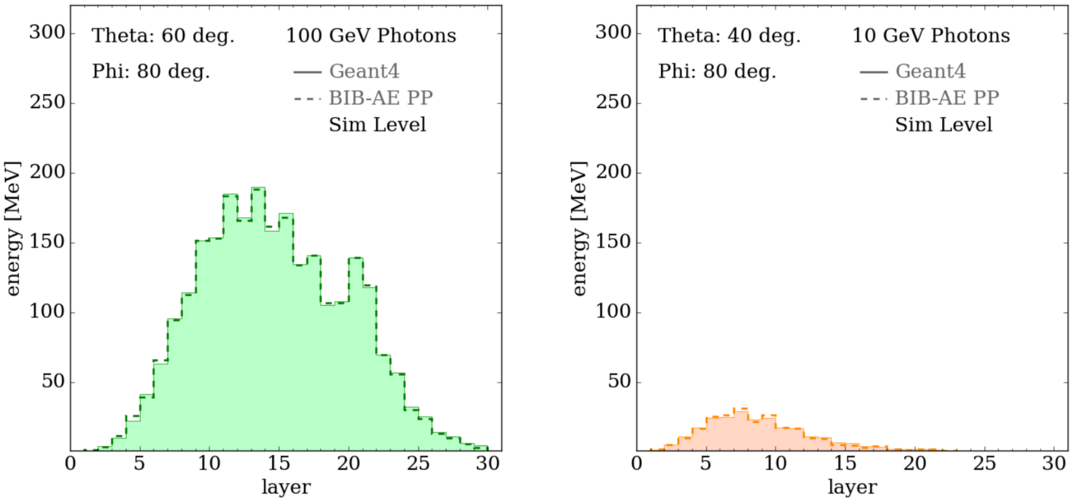


Figure 8.11: Longitudinal profiles at simulation level for the best (100 GeV, theta 60 degrees, phi 80 degrees, left) and the worst (10 GeV, theta 40 degrees, phi 80 degrees, right) incident energy and angle combinations.

and worst (10 GeV, theta 40 degrees, phi 80 degrees) incident energy and angle combinations selected according to the JSD value (Equation 7.10) between the GEANT4 and BIB-AE distributions, are shown in Figure 8.12. Generally, more noticeable deviations at larger radii appear in these distributions than were present in the single angle study in Chapter 7, and while in the best cases these are limited to the tails of the distribution, in the worst case deviations even become more apparent closer to the shower core.

8.4 Conclusions

In this chapter, we have developed a fully conditioned generative model for highly granular calorimeter simulation, extending the study performed in Chapter 7. This work has also taken steps to generalise previous studies in order to allow showers to be simulated at different positions in the calorimeter. This required the development of a geometry with a purely regular segmentation to allow all the energy deposited in the active layers of the calorimeter to be recorded. This development is important, not just for models relying on convolutions over regular grids, such as the BIB-AE explored here, but any model that would be used to produce showers which would be placed back into different positions in the real, irregular geometry. Additionally, a strategy was developed and executed to create a dataset with incident energy and two incident angles being simultaneously varied. This involved minimising the box size to retain as much of the computational speed-up as possible, while also allowing important information about the incident position of the particle to be easily computed. This is essential to be able to perform shower simulations with a generative model at different positions in a full calorimeter subsystem.

The conditioning scheme for the BIB-AE developed in Chapter 7 was extended to that required for a fully conditioned model, and the model retrained. The physics performance of the model, which has been measured here in terms of calorimetric observables for fixed combinations of incident energies and angles, was generally of a somewhat lower quality than that observed for the single angle conditioning studied in Chapter 7. Notable exceptions were the visible energy deposited in the calorimeter and the longitudinal shower profile into the depth of the calorimeter, where the BIB-AE performance remained excellent. The lower

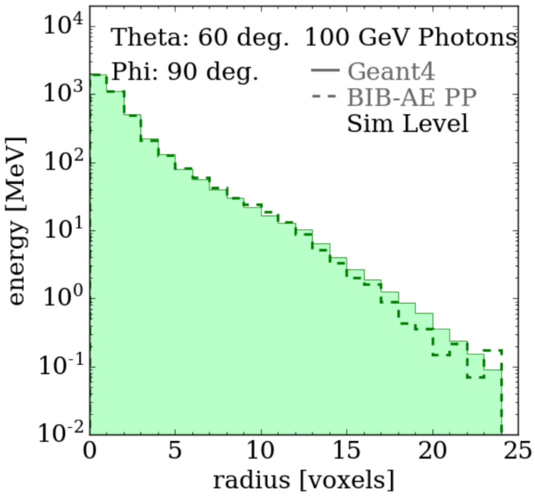


Figure 8.12: Radial profiles at simulation level for the best (100 GeV, theta 60 degrees, phi 90 degrees, left) and the worst (10 GeV, theta 40 degrees, phi 80 degrees, right) incident energy and angle combinations.

physics fidelity observed with the BIB-AE in this study could be a result of the fact that the conditioning phase space was increased significantly compared to that used in Chapter 7, but the number of training samples kept constant. This effectively means that the phase space is less densely populated by training samples, thereby making it more difficult for the BIB-AE to learn key features correctly. The number of training samples needed for training a model to a sufficient level of performance is generally an open question, and could be the focus of dedicated future studies. Additionally, a systematic re-tuning of the hyperparameters used in the BIB-AE would help tailor the model to this dataset and could therefore result in an increase in the performance of the model.

The development of the BIB-AE described here has focused on optimising the model for certain salient calorimeteric observables that we believe are representative of the simulation quality, motivated by our understanding of the physics of calorimeter showers. However, it should be reiterated that the ultimate target of such a simulator is the quality of higher level quantities after interfacing with reconstruction. This makes it very difficult to draw definitive conclusions on the quality of a model purely based on simulation level observables, as these conclusions could be heavily impacted by reconstruction and the needs of a given analysis.

Chapter 9

Integration into the International Large Detector Software Chain

The description of the software library presented in Section 9.1 has previously been published in [2], in collaboration with Erik Buhmann, Thorsten Buss, Sascha Diefenbacher, Engin Eren, Frank Gaede, Gregor Kasieczka, William Korcari, Anatolii Korol, Katja Krüger, Thomas Madlener and Lennart Rustige. As such, this section includes numerous figures, as well as significant amounts of text that are similar or identical to parts of that work. This author's contribution to the work included key involvement in the implementation of inference for models written in PyTorch *in* Geant4*, implementation of the fast simulation trigger in DDFastShowerML, conversion of the BIB-AE model into a format suitable for use in* C++*, integration of the BIB-AE model and writing sections of the paper.*

Having developed a fully conditioned generative model in Chapter 8, we will now focus on the integration of the model into the existing software chains. The ultimate target will be to use the fully conditioned BIB-AE model for simulation of electromagnetic showers (photons, positrons and electrons) in the barrel and endcap regions of the ILD ECAL, with a seamless transition with full Monte Carlo simulation in Geant4.

In order to achieve this goal, this chapter will focus on the development of a generic library that combines inference for generative models trained to simulate calorimeter showers together with DD4hep. Firstly, in Section 9.1 a software library which was developed during the course of this thesis will be described. In Section 9.2, details will be given on the integration of the BIB-AE model developed in Chapter 8 into this library, including the numerous steps taken to tackle the challenges posed by this integration. Finally in Section 9.3, the computational performance of the library will be studied, with a particular focus on the performance of the fully conditioned BIB-AE model.

9.1 A Library for Fast Calorimeter Simulation with Generative Models in Geant4 and DD4hep

The approach adopted for fast calorimeter simulation in this thesis is to use a single generative model to generate showers that can then be placed into different positions in the calorimeter. This requires a number of elements to be present in a suitable software library, including a geometry description and navigation, and a means of handing information from the full simulation suite to the fast simulation tool in an efficient manner. The software should be written in C++, which is currently used ubiquitously throughout high energy physics. Since DD4hep provides an interface to a generic fast simulation tool via the fast simulation hooks present in Geant4 (see Chapter 6), we choose to make use of these software suites to build our library. This has the major advantage of the key machinery required already being present in the tool. This includes an interface for terminating the detailed physics-based simulation provided by Geant4 via a trigger mechanism if the particle type and energy satisfy particular criteria. This automatically provides a seamless means of combining full simulation with a generative model based fast calorimeter simulation tool.

Our prototype library, called *DDFastShowerML*[1], is therefore built upon Geant4 and DD4hep. The library is designed to be generic so as to allow different types of generative models, in terms of their composition, input and output, to be used for fast calorimeter simulation in a detector geometry implemented in DD4hep. This library has to provide interfaces for several key components.

The first set of components concern the generative model. These include the implementation of the specific model that will be used, as well as an implementation for model inference with the particular software library that the model was designed for. Since different models are foreseen, the interface used for the model must be sufficiently flexible and general so as not to bind the interface to a particular kind of model in terms of the input or output that it requires. The inference interface on the other hand must allow for different inference libraries to be used, depending on what is required by the model.

Once the model inference has been run and the output obtained, it must be placed back into the detector geometry. In order to obtain an appropriate detector response, a correct hit placement must be given within the specific geometry used. Finally, since realistic geometries are not perfectly regular in order to allow for a physical construction, as we have seen in the case of the ILD ECAL, a means must also be provided to prevent the fast simulation running in specific regions of the detector. This is an essential component, as these regions of the detector will have a geometry with a significantly different structure to the region used for training. This means that the model will not be able to provide a physically meaningful detector response. Since different detectors can have drastically different structures, the interface must be sufficiently flexible to allow different regions of a detector to be excluded depending on its structure. This interface could also be used to allow different generative models to be run in different regions of the detector, meaning an appropriate model with a dedicated training could also be used to reproduce a correct response for a specific detector region. However, in this work we focus on generating the response in the bulk of the detector and simply exclude these regions, leaving a generative modelling of these more complex detector regions to future studies.

Algorithm 1 Pseudocode illustrating the order of operations for the core components of the *DDFastShowerML* library.

```
if Trigger.checkTrigger(track) == True then
    Kill full simulation of particle
    localDir = Geometry.getLocalDir(track)
    inputs = Model.prepareInputs(track, localDir)
    outputs = Inference.runInference(inputs)
    localSPs = Model.convertOutput(track, localDir, outputs)
    globalSPs = Geometry.localToGlobal(track, localSP)
    for (sp in globalSPs) do
        HitMaker.makeHit(sp, track)
    end for
else
    Full simulation of particle with Geant4
end if
```

---

[1] The code is available at `https://gitlab.desy.de/ilcsoft/ddfastshowerml`

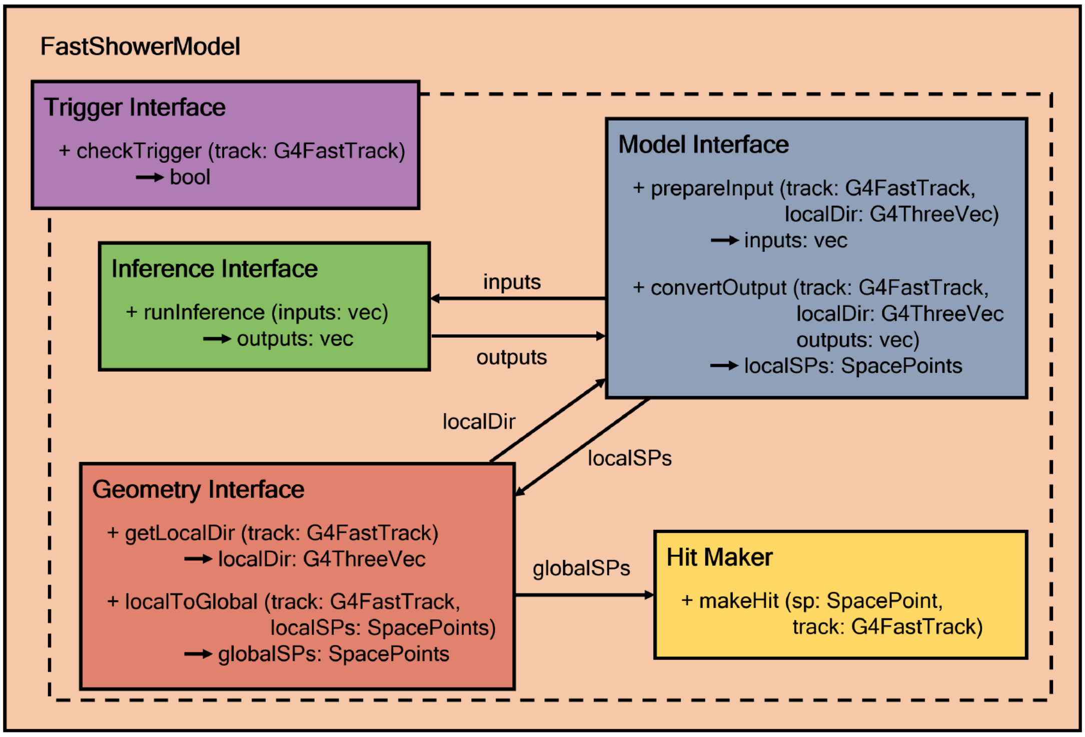

FastShowerModel
Trigger Interface
+ checkTrigger (track: G4FastTrack)
→ bool
Model Interface
+ prepareInput (track: G4FastTrack, localDir: G4ThreeVec)
→ inputs: vec
+ convertOutput (track: G4FastTrack, localDir: G4ThreeVec outputs: vec)
→ localSPs: SpacePoints
Inference Interface
+ runInference (inputs: vec)
→ outputs: vec
inputs
outputs
localDir
localSPs
Geometry Interface
+ getLocalDir (track: G4FastTrack)
→ localDir: G4ThreeVec
+ localToGlobal (track: G4FastTrack, localSPs: SpacePoints)
→ globalSPs: SpacePoints
globalSPs
Hit Maker
+ makeHit (sp: SpacePoint, track: G4FastTrack)

Figure 9.1: Class diagram for the core components of the *DDFastShowerML* library. For the purposes of illustration, slight simplifications have been made with respect to the actual implementation, which includes some memory optimisations such as re-using existing allocations from previous events.

The library addresses these requirements by using a class template at its core which consists of five key components. A template is used as it contributes no overhead at runtime, while retaining almost all flexibility. The interfaces are separated as far as possible such that they each address tasks with as little interdependence as possible. This allows them each to be replaced individually with minimal changes being required to the rest, meaning they can be readily adapted for a given application.

A class diagram showing the key components of *DDFastShowerML* is shown in Figure 9.1, with the algorithmic operation of the library being shown by the pseudocode in Algorithm **??**. These components are as follows:

Trigger

The *Trigger* interface provides a simple means to allow the definition of conditions for running the fast shower model. It sits on top of the particle type and energy triggers that are already present in the DD4HEP interface to GEANT4. For example, this interface can be used to exclude a certain fast shower model from running in particular regions of the detector geometry, and instead run full simulation. This is currently implemented to exclude the more irregular regions at the transition between sub-detector elements for the ILD detector, which will be described in more detail in Section 9.2.3.

Model

The first class is the interface for the *Model*, which contains the implementation specific to a given model. This is separated into two methods for the specific model. The first is the preparation of the input in the correct form for the model. This may include the preparation of the latent vector passed to a model, as well as the conditioning parameters that will be passed to the model to generate a specific shower. For this reason, information related to the incident particle kinematics has to be accessible. The position and energy of the incident particle can be obtained from the *FastTrack* — the Track object (see Section 6) used by GEANT4 for the particle being handled by fast simulation. In order to provide the information about the incident angle, it is necessary to introduce information from the *Geometry* component of the interface. This is given in terms of the *localDirection* (localDir) object, defined in the *Geometry* interface. The second component of the *Model* implementation is the interpretation of the model output, such that it can be converted into local space points. This interpretation is done per layer of the calorimeter in local coordinates, using the convention that the origin is placed at the entry point into the calorimeter. For this reason the *FastTrack* and *localDirection* objects are also provided to allow a uniform interpretation across different forms of model output.

Inference

The *Inference* interface provides a simple means of passing input to, and obtaining output from, the implementation of the inference library called for a specific model. The implementation for a specific library also includes the declaration of any plugin properties. Currently, *DDFastShowerML* supports the LIBTORCH [192] and ONNXRUNTIME [198] libraries for model inference. While PYTORCH models can be used with both inference libraries, we observed a better inference performance with LIBTORCH for the use case of a batch size of one on CPU hardware for the PYTORCH models that are currently in use in the library. This inference library is therefore the one that is used in this thesis.

Geometry

The role filled by the *Geometry* interface is two-fold. In the first instance, the *localDirection* is calculated to provide the flight direction of the particle in local calorimeter coordinates (see Section 7.1), which can be used in the *Model* interface in order to provide a consistent basis for model conditioning. The second role of the *Geometry* interface is to allow the placement of the space points produced as the output from the *Model* interface back into the detector geometry. This includes for example the positioning of calorimeter layers, and the conversion between global (envelope) and local calorimeter coordinates. Currently, implementations are provided for endcap and polyhedral barrel calorimeter geometries.

HitMaker

The helper class provided by GEANT4 to allow placement of energy deposits produced by a fast simulation model into the detector geometry, given that their position lies within a sensitive region of the detector. The energy deposits should be in the form of space points.

Currently, the *DDFastShowerML* library only supports single shower generation with a generative model (i.e. inference with a batch size of one) on a CPU. While there is considerable ongoing effort to make GPU hardware, and corresponding software tools, more widespread in high energy physics computing infrastructure, CPU hardware is still dominant [199]. This makes benchmarks of the computational performance of the prototype library in its current form a realistic, albeit pessimistic measure of the speed up provided by generative models for fast calorimeter simulation with current computing infrastructure. Future work should aim to extend the library to support batching of shower generation within an event, allowing the number of sequential passes through the model to be reduced, hence improving the per-event simulation time for a given model. This would allow another key development to be made, which would be the addition of GPU support. The most significant speed-ups would then be accessible via parallelisation.

The library has been tested using the KEY4HEP [200] ecosystem, with GEANT4 version 11.1 using the QGSP_BERT physics list and DD4HEP version 1.27. Note that it was necessary to update both the GEANT4 and DD4HEP versions from those used in Chapters 7 and 8, as updates to these libraries which were introduced during the course of this thesis were necessary to use generative models within these ecosystems.

9.2 BIB-AE Integration

Currently, the fully conditioned model developed in Chapter 8 has the most complete implementation in *DDFastShowerML*. A 20 GeV photon shower in the barrel of the ILD ECAL that has been simulated with the BIB-AE is shown in Figure 9.2. We will now focus on the steps taken in order to integrate this particular generative model into the library.

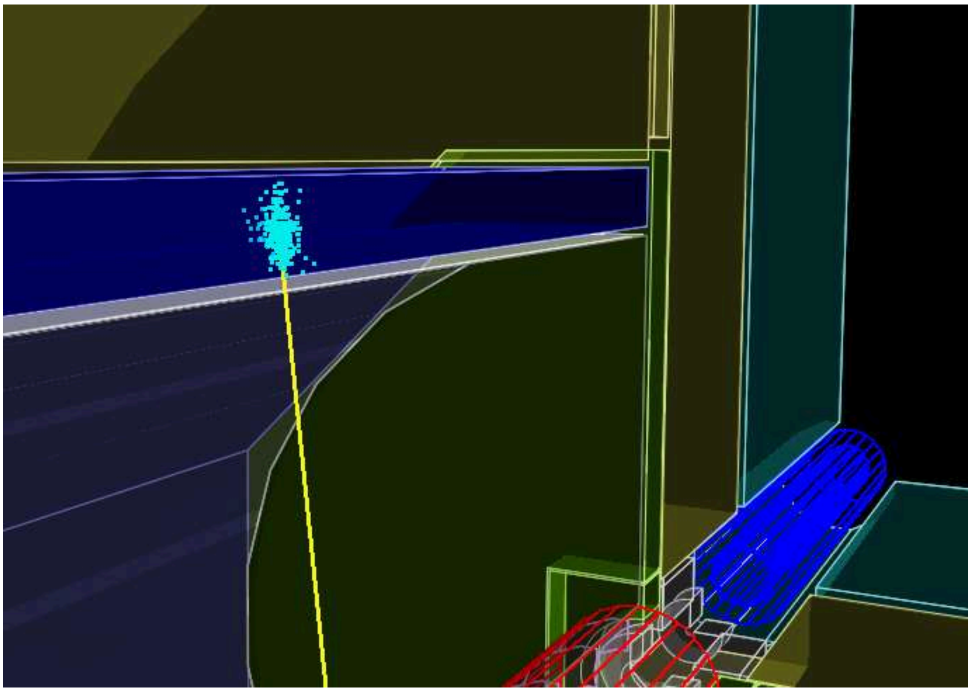


Figure 9.2: 20 GeV photon shower simulated in ILD with a BIB-AE model using the *DDFastShowerML* library. Figure from [2].

9.2.1 Running the BIB-AE Architecture in C++

Since the BIB-AE model is a multi-component model, a custom approach must be adopted for conversion into a form suitable for use in C++. The model is broken into four components: the NF for latent space sampling, the main BIB-AE network, the Post Processor network and the energy sum rescaling of the shower. Each of the components were serialised by tracing or scripting each of the individual operations that occurred in the functions that were called. For the main BIB-AE network, the Post Processor network and the rescaling procedure this was achieved using the utilities provided by TORCHSCRIPT in PYTORCH [192], which provides an intermediate representation of a PYTORCH model that can then be run in C++. For the NF model which was implemented using the PYRO [196] deep probabilistic programming library, the POUTINE effect handlers provided in PYRO were used to produce a TORCHSCRIPT model. The full BIB-AE model consisting of the four components was then converted to TORCHSCRIPT by means of a container that was used to wrap the complete sequence of operations composing the full BIB-AE. This means that when the model is called in C++, only the conditioning variables for a given incident particle need to be provided, with the output being a shower in the regular grid. This was again performed in octants in the local ϕ coordinate, and amounted to reflections in the $x-y$ plane.

9.2.2 Implementation of the Model and Geometry Interfaces

Having obtained the model in a suitable format, the implementation of the BIB-AE by means of the *Model* interface in *DDFastShowerML* reduces to the preparation of the conditioning variables required as input, and the correct conversion of the output. The angular conditioning variables have to be converted from the local calorimeter coordinate system to the global angles that were used for conditioning. As the angles used for training the model in Chapter 8 were only varied in one direction from the normal to the face of the calorimeter, transformations are applied in octants in ϕ using the *localDirection* provided by the interface, with the x and y coordinates interchanged where necessary. When the output of the model was obtained after inference, the incident position in the grid was calculated from the incident global angles to the regular grid, as described in Chapter 8. The grid containing the shower was then transformed depending on the ϕ incident angle in the local coordinate system, such that the space points produced conformed to the coordinate system employed in *DDFastShowerML*.

In the implementation of the *Geometry* interface, the exact computation of the *localDirection* three vector and the *localToGlobal* transformation depends on the detector region concerned. In the barrel, a rotation is applied such that all sectors in ϕ' are mapped to the sector which lies in the positive x' region and parallel to y'. The computations for both the *localDirection* and the *localToGlobal* are performed in this sector, before eventually being rotated back to the original sector in ϕ'. The two endcaps are treated in a common implementation. However, for the endcap in the negative z' direction, a rotation by π around the y' axis has to be applied, while no additional transformation is necessary for the endcap in the positive z' direction, as the global coordinate system aligns with the local one.

9.2.3 Geometry Exclusions

The final step required to fully integrate the BIB-AE model into *DDFastShowerML* is to prevent the model being called in certain regions of the calorimeter sub-system where the detector has a very different physical structure to that used for training the model. Figure 9.3 illustrates the issues with using the BIB-AE model, or any such model producing a regular grid output, in the regions of the ILD ECAL that present a very different physical structure to the majority of the detector. These regions lie at a transition between detector elements, and as such do not present the planar calorimeter structure that the model was trained on and that dominates the majority of the calorimeter. These regions can be excluded from the fast simulation by means of the *Trigger* interface described in Section 9.1, with an independent trigger being implemented for the barrel and endcap respectively. The regions are excluded by selecting a range of angles in global theta (θ') and global phi (ϕ').

The first type of transition is the region between the edge of the barrel and the endcap, which is shown in Figure 9.3 a). In this case, there is a gap between the barrel and the endcap, with the endcap noticeably overhanging the barrel. Additionaly, the orientation of the calorimeter layers, shown by the orange lines, changes by 90 degrees in the transition. This means it is not physically meaningful to use the BIB-AE in this region. Two ranges of angles in θ' are selected in both the barrel and endcap triggers, depending on whether the transition is in the forward (i.e $\theta' < 90$ degrees) or backward (i.e. $\theta' > 90$ degrees) endcap. In the barrel, the regions $\theta' < 40$ degrees and $\theta' > 140$ degrees are excluded from the fast simulation trigger. In the forward endcap, the region $\theta' > 33$ degrees is excluded, while in the backward endcap, the region $\theta' < 147$ degrees is excluded. This corresponds to a 7 degree window in

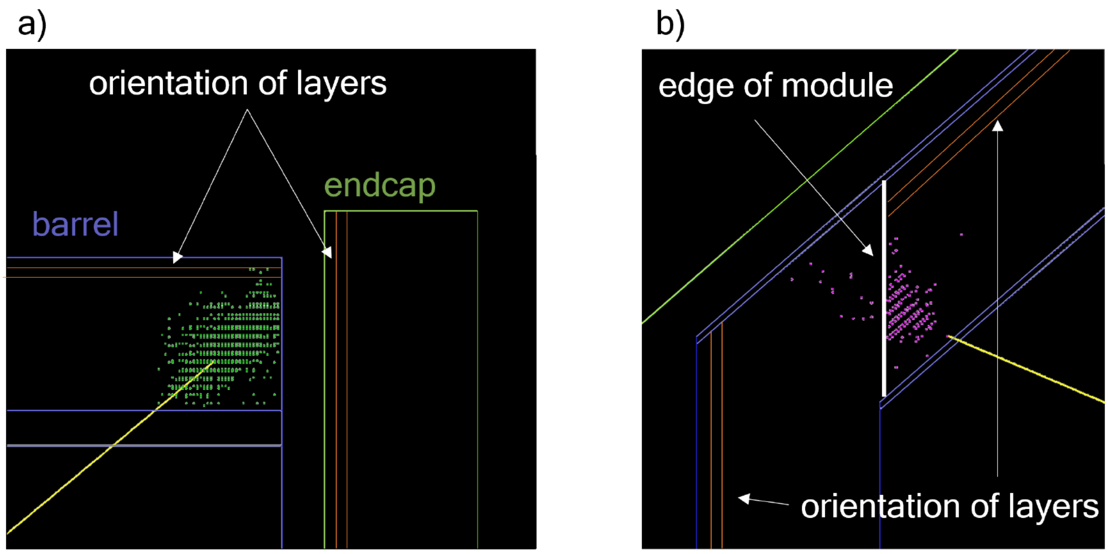


Figure 9.3: Simulation of a shower with the BIB-AE in regions of the ILD ECAL that are excluded from the fast simulation trigger, illustrating the problem with using a regular grid model in these regions. The regions of the detector that are excluded have a geometrical structure that differs significantly from that of the geometry that was used for training. The regions which are excluded are a) the transition between the barrel (purple) and endcap (green) and b) the boundary between modules (white line) in the barrel region of the ECAL. In each case, the transition involves a change in orientation of the calorimeter layers, which is indicated by the orange lines for each detector element.

θ' for each of the two endcap transitions. This exclusion range covers both the geometrical range of this transition, and the conditioning range used during the training of the BIB-AE.

The second type of transition is the intersection between modules in the octagonal barrel, which is shown in Figure 9.3 b). In this case the orientation of the calorimeter layers, shown by the orange lines, changes at the edge of the module which is indicated by the thick white line. A sharp cut-off appears for BIB-AE showers generated in this region, as the placement of hits into the detector defined in the *Geometry* interface assumes a planar orientation of calorimeter layers. The different physical structure of the calorimeter means that the shower evolution learnt by the model is also not physically meaningful in this region. These regions are excluded by means of an 8.01 degree range selected in ϕ' for each of the 8 module transitions that occur in the barrel. Due to the asymmetric nature of the intersection of the modules, these ranges are also asymmetric. The cuts applied are

$$\left(\pm\left(\frac{\pi}{8}+n\left(\frac{\pi}{4}\right)\right)-0.01\left(\frac{\pi}{180}\right)\right)<\phi'<\left(\pm\left(\frac{\pi}{8}+n\left(\frac{\pi}{4}\right)\right)+8\left(\frac{\pi}{180}\right)\right), \tag{9.1}$$

where $n = 0, 1, 2, 3$, with one cut for each corner of the barrel resulting in a total of 8 cuts. When combined together, these excluded regions account for approximately 10% of the total surface area of the calorimeter sub-system.

9.3 Computational Performance of the Library

With the BIB-AE now fully integrated into the library, it is possible to perform a fair benchmark of the computational performance of the model in comparison to GEANT4. From profiling the library, when the fast simulation model is called the simulation time is dominated by running the model inference. The combined time for the preparation of the model input, conversion of the output into space points and transformation between local and global coordinates is

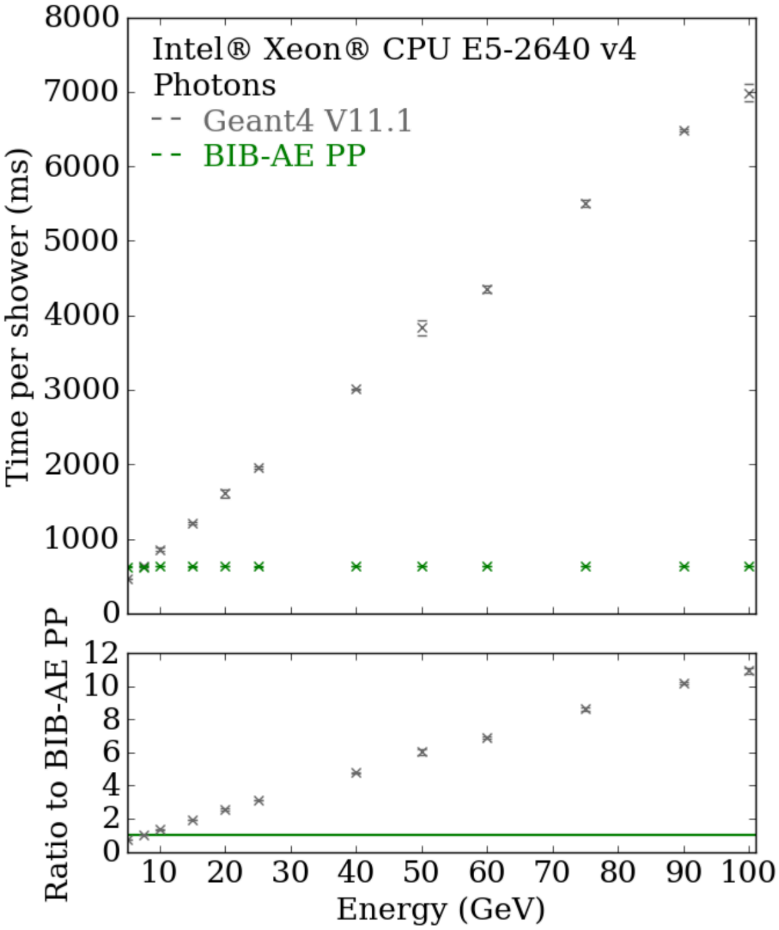


Figure 9.4: Simulation time per shower against energy of the incident particle for showers simulated with GEANT4 (grey) and with the BIB-AE (green) in *DDFastShowerML*. The sub-panel shows the ratio to the BIB-AE time, highlighting the speed-up factor relative to GEANT4. Each point is the average obtained from three sets of 2000 showers, with the error arising from the standard deviation.

typically limited to well below 0.1% of the inference time. The time spent in the hit maker is larger, but still only 1% or less of the time spent on model inference.

The time per event for photon showers against the incident energy of the photon simulated in the ILD ECAL with *DDFastShowerML* using a BIB-AE in comparison to GEANT4 is shown in Figure 9.4. All photons were fired orthogonally to the face of the calorimeter and were incident at a fixed position of $(x', y', z') = (0.0\ \mathrm{mm}, 1810\ \mathrm{mm}, -209\ \mathrm{mm})$. All simulations were performed on a single core of an Intel® Xeon® CPU E5-2640 v4. The simulation time for GEANT4, which is driven by the number of secondaries created, increases with the energy in an almost linear fashion. Conversely, since the BIB-AE model outputs a grid of a fixed size independent of the energy, and the inference time of the model constitutes almost the entirety of the simulation time per event, the simulation time of the BIB-AE is approximately flat. A speed-up by a factor of $\times 11$ over GEANT4 is achieved by the BIB-AE when simulating 100 GeV photon showers.

9.4 Conclusions

In this chapter, we have described the development of a generic C++ library that can be used to interface generative models with Geant4 and DD4hep, to allow a seamless mixing of full and fast simulation of detector geometries with a realistic structure and material composition. Additionally, the integration of the fully conditioned BIB-AE model developed in Chapter 8 has been outlined, which has also served to test all of the various features of the library. This includes the conversion of the model to a format suitable for use in C++, the implementation of the model within the software library, and the exclusion of fast simulation in certain regions of the detector geometry. It should be emphasised that although the principle subject of this chapter has been the integration of the BIB-AE model into this library, it is sufficiently flexible to generally allow different types of generative model to be implemented for use with different detector geometry descriptions in DD4hep. Additionally, while this work has focused on the integration of a model for electromagnetic shower simulation in the planar regions of the ECAL, multiple different generative models could be used together within this simulation framework. This could include, for example, for different particle types and for the regions of the detector with a different physical structure.

The computational performance of the library has been investigated on CPU hardware. Since full Monte Carlo-based simulation and fast simulation can be run in the same software environment in this library, a fair comparison has been drawn between Geant4 and the BIB-AE generative model, which was previously not possible. Importantly, it has been demonstrated that the unavoidable overheads which arise from passing information to and from the generative model and from placing the energy deposits back into the detector geometry are negligible compared to the BIB-AE inference, indicating that the simulation library developed operates in an efficient manner. The reduction in simulation time per shower when using the BIB-AE relative to Geant4 increases with the energy of the incident photon, with the BIB-AE being $\times 11$ faster than Geant4 when simulating 100 GeV photon showers. In the future, the library should be extended to allow batches of showers within an event to be generated on GPU hardware, as this is the scenario that will provide the most significant speed-ups offered by a generative model.

Chapter 10

Irregular Geometries

With the fully conditioned BIB-AE model integrated into the full simulation chain, it is now possible to easily simulate showers with the model in different regions of the geometry. This enables investigations into the effects of projecting showers simulated with a generative model in a regular grid into the realistic geometry of the ILD ECAL, which is irregular. In Chapter 8, steps were taken to create a purely regular geometry in order to create the training data for the model. This meant that no artefacts were created by having to project an irregular geometry into a regular geometry, and that all the energy deposited in the active layers of the calorimeter was retained. However, it is not clear a priori what the effects of placing showers created in a regular geometry back into an irregular geometry will be. The task is made more challenging by the constraints provided by the simulation tool, in that the structure of the irregular geometry into which the shower will be placed cannot be accessed. This means, for instance, that the cells in the regular grid cannot be placed directly onto cells in the physical geometry.

In this chapter, different means of depositing the hits belonging to a shower back into the irregular detector geometry will be investigated. In Section 10.1 two different approaches to depositing the output of a regular grid generative model into an irregular geometry will be described. Two positions in the irregular geometry with differing fractions of insensitive volumes will be investigated. In Section 10.2, the effects on key calorimetric observables for showers created by a particle of fixed incident energy and angle will be studied, before conclusions are drawn in Section 10.3.

10.1 Investigations into the Effects of Simulation with a Regular Grid Model in an Irregular Geometry

The simplest approach to depositing the output of a regular grid model is by taking the positions of the cell centers in the regular grid and placing them directly (taking into account the physical cell size and the distances between layers) into the irregular geometry. One problem that is encountered with this approach is caused by insensitive volumes in the detector geometry. If a hit lands in one of these volumes, it is not recorded by GEANT4, and this can result in significantly more energy being lost than would have been the case in a physics-based simulation of the shower in the detector geometry. One approach that will be explored to try to combat this issue is to divide the energy of a hit into cells of a size smaller than that of the physical readout cells, hereafter referred to as *sub-cells*. This approach is inspired by that adopted in [6], who used information at the level of GEANT4 steps (see Chapter 6) to learn a much higher resolution representation of showers. This helped the model to achieve a high degree of independence from the physical readout geometry. It should be noted that the model used in that case was a point cloud[1] model rather than a regular grid model, making the learning of a much higher dimensional representation of a shower feasible. In the scenario studied here, the splitting of energy in the cells of the regular grid produced by the BIB-AE cannot add any additional information to that contained in the readout geometry.

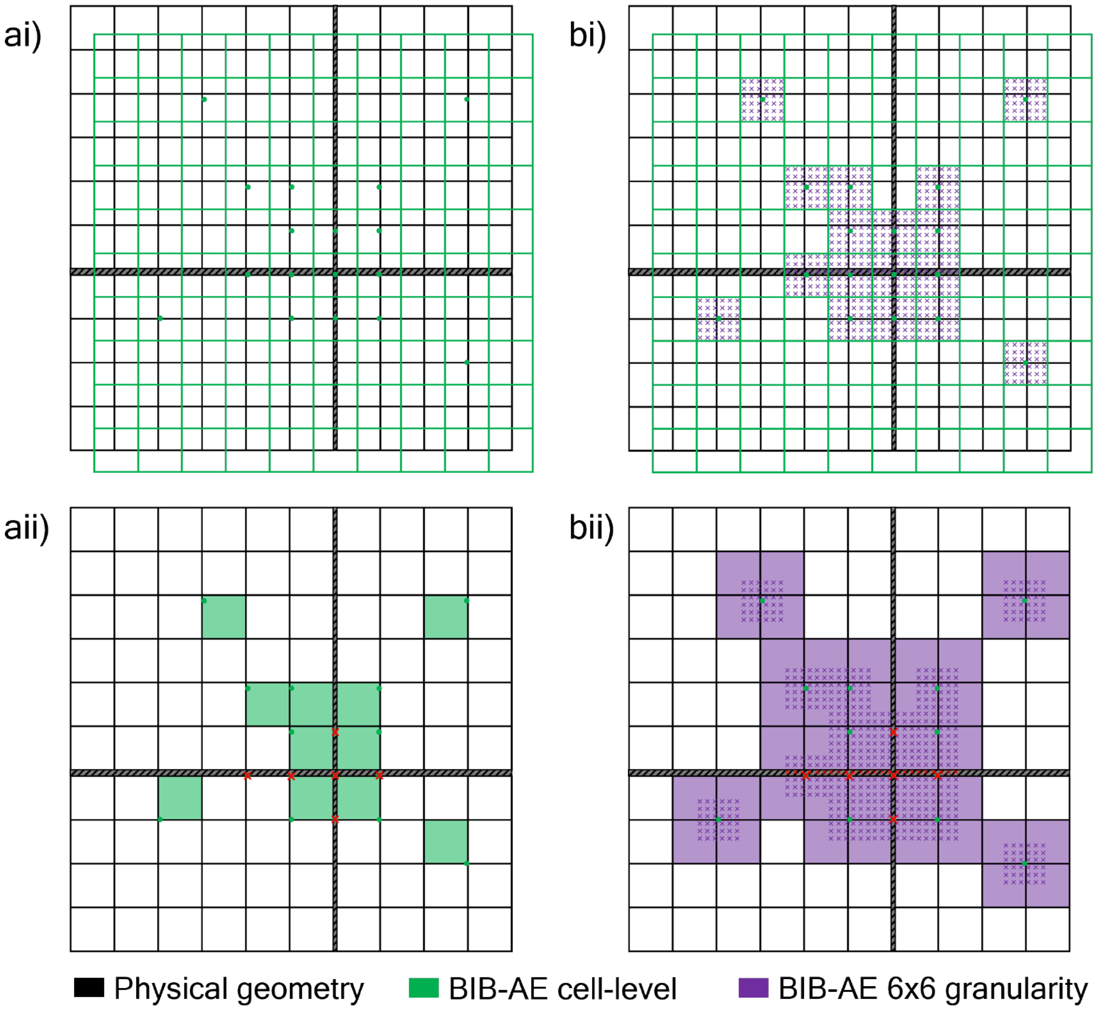


Figure 10.1: Illustration of the deposition of a single layer of a shower created in a regular grid with the BIB-AE (green) into a single layer of a mis-aligned irregular geometry (black) containing insensitive volumes (black hatched regions) of differing sizes. Two approaches to depositing the hits in the shower back into the irregular geometry are shown. ai) the entire energy of a hit in the regular grid is placed at the center of the regular grid cell. aii) if the center of the cell lands in an insensitive volume in the irregular grid, the entire energy of the hit is lost (red crosses). bi) The energy of a each hit is uniformly divided by area into sub-cells of $\times 36$ higher granularity (purple) than that present in the physical readout geometry. bii) less energy is lost when the higher granularity sub-cells land in insensitive volumes, but additional hits are also created.

Figure 10.1 illustrates the two different approaches to placing a shower back into the irregular geometry present in the physical ILD ECAL. An extreme case of mis-alignment between the regular and irregular geometries is shown to emphasise the effects. A single layer of the regular grid output of the BIB-AE model is shown in green, while a single layer of the irregular physical geometry is shown in black, with insensitive volumes of varying size shown by the black hatched regions. In Figure 10.1 ai) the regular grid output is taken directly (i.e. at cell-level) and the hit energy placed into the physical geometry at the center of the regular grid cell. Since the relative position of the cells in the regular grid to those in the irregular grid is not known, it is possible for the hit to be placed into an insensitive volume. In Figure 10.1 aii), this results in the entire energy of the hit being discarded by GEANT4 (red crosses) during simulation, even though the insensitive volume is much smaller than the size of a cell. In the alternative approach shown in Figure 10.1 bii), each cell in the regular grid output of the BIB-AE is divided into a more granular 6×6 regular grid of sub-cells (purple). The energy of the cell-level hit is divided uniformly across the more granular sub-cells according to their area (i.e. each high granularity sub-cell contains $\frac{1}{36}$th of the energy in the cell-level hit). The high granular sub-cells are then placed back into the detector geometry at the center of the sub-cell. This means that if a high granularity sub-cell now lands in dead material (red crosses), a much lower fraction of the energy of the hit is lost during simulation. However, a side effect of splitting up the energy of a hit in the regular grid in this manner appears when the regular grid is not aligned with the irregular geometry. In this case, the energy contained in a single cell in the regular grid is spilt over multiple cells in the irregular geometry.

---

[1] A point cloud is a discrete set of space points containing information associated with the points, for example their position in 3D space, the energy for a calorimeter hit etc.

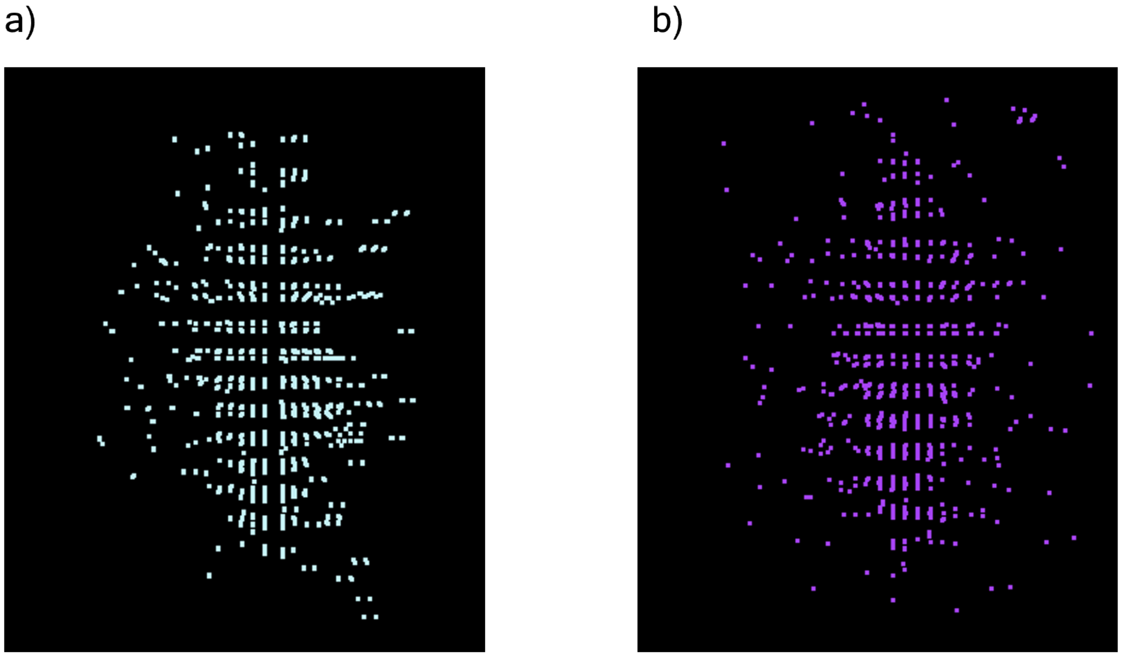


Figure 10.2: Example BIB-AE showers generated at two different positions in the irregular geometry of the ILD ECAL. A shower in position 1 is shown in a), with a large insensitive volume near the core of the shower. A shower in position 2 is shown in b), with no large insensitive volumes near the shower core.

The effect of using these two different methods for placing the BIB-AE generated showers back into the irregular ILD ECAL geometry was tested at two positions. These are shown in Figure 10.2. At *position 1*, shown in Figure 10.2 a), a large insensitive volume caused by a gap between modules in the barrel (this gap is aligned between layers, with no absorber or sensitive layers being present in this region, see Figure 8.1) lies near the core of the shower, meaning that the effects from losing hits when placing the showers into the geometry are exacerbated. Showers were created in this region by placing a photon gun right in front of the face of the ECAL at a position of $(x', y', z') = (-4.84 \times 10^{-15}\ \mathrm{mm}, 1810\ \mathrm{mm}, -100\ \mathrm{mm})$ in the global ILD coordinate system. At *position 2*, shown in Figure 10.2 b), there are no large insensitive volumes near the core of the shower. Showers were created in this region by placing a photon gun right in front of the face of the ECAL at a position of $(x', y', z') = (0.0 \text{ mm}, 1810 \text{ mm}, -209 \text{ mm})$. It should be noted that having a large insensitive volume near the shower core, as is the case at *position 1*, is a much less common scenario, as these irregularities are not as prevalent in the ILD ECAL. This makes *position 2* the more representative scenario for a shower occurring in the ILD ECAL. For each of the two positions, photons with an incident energy of 50 GeV were fired orthogonally to the face of the ECAL. As described in Chapter 9, the GEANT4 version had to be updated from version 10.4, which was used to train the fully conditioned model in Chapter 8, to version 11.1, which was necessary to create the *DDFastShowerML* library in Chapter 9. For this reason, both GEANT4 versions are used to simulate showers in both positions, in order to gauge the variations between versions and thereby give an estimate of the uncertainty on the MC simulation tool. Additionally, both methods of placing regular grid showers generated with the BIB-AE are tested at both positions. These are the cell-level placing of hits at the center of the regular grid cells, and dividing the hit energies up uniformly into a higher granular grid of sub-cells.

10.2 Results

For each of the two test positions in the ILD ECAL, 5 calorimeteric observables were investigated. These were: the visible energy sum, the number of hits, the cell energy spectrum, the longitudinal profile and the radial profile. In line with the previous studies presented in Chapters 7 and 8, a fixed calibration factor was applied to scale the hit energies in the last 10 layers of the calorimeter to account for the modification of the sampling fraction caused by the absorber layers being twice as thick in this region. Additionally, hits with an energy deposition of less than 0.07875 MeV, which corresponds to an energy deposition of less than half a MIP, are mapped to zero.

Different procedures for the splitting up of hit energies in the regular grid into a higher granular grid were explored. Splittings of both 6×6 and 8×8 were tested, however no visible increase in the amount of energy retained relative to GEANT4 was observed when increasing the granularity of the splitting. For this reason, the splitting of 6×6 was preferred. Additionally, with the 6×6 splitting adopted, the effect of only applying the splitting to hits above a threshold of 1.5 MIPs was investigated. This approach showed no major improvement in terms of the physics observables studied when compared to applying a 6×6 splitting to all cells, independent of their energy. More details on these studies can be found in Appendix C.

Position 1

We begin by studying the effect of projecting the showers produced in a regular grid by the BIB-AE into the ILD ECAL in the first position, with a large insensitive volume near the shower core.

Figure 10.3 shows the total visible energy above threshold deposited in the calorimeter (top, left), the number of hits above threshold (top, right) and the cell energy spectrum (bottom, right) for 50 GeV photon incident perpendicularly to the calorimeter face simulated with the two GEANT4 versions, and with the BIB-AE. The two different approaches for depositing the showers simulated with the BIB-AE in the regular grid into the geometry are shown in green for the cell-level approach, and purple for the high granularity sub-cell approach.

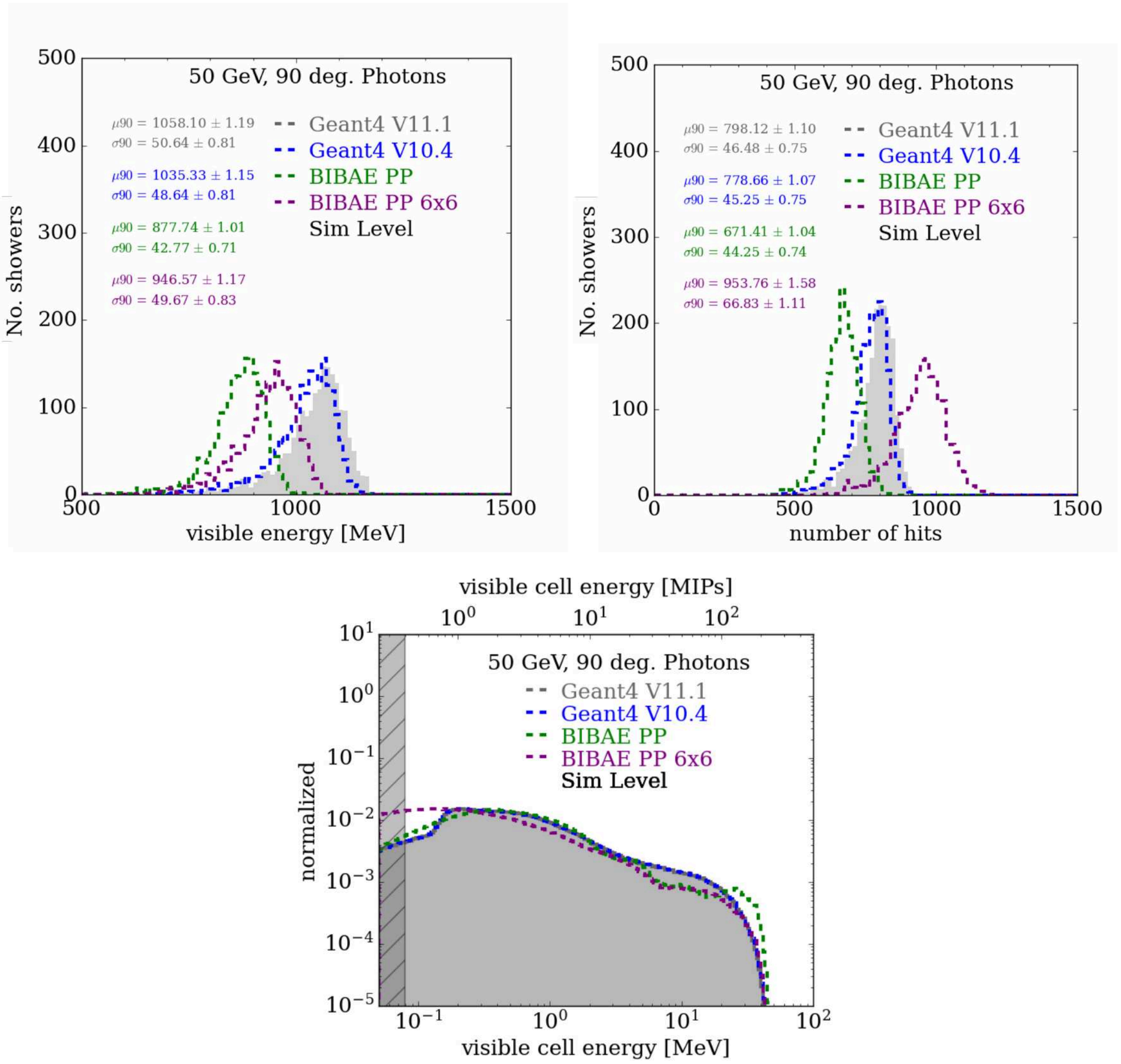


Figure 10.3: Distributions for 50 GeV photons incident orthogonal to the face of the calorimeter at *position 1*. Showers were simulated with two different Geant4 versions: 11.1 (grey, filled distributions, used in the *DDFastShowerML* library) and 10.4 (blue, unfilled distributions, used for the training of the BIB-AE model). Showers were also simulated with the fully conditioned BIB-AE model, and projected back into the irregular ILD ECAL with two different approaches. In the first approach, hits are placed at the center of the cell in the regular grid (green, unfilled distributions) and in the second the energy of a hit is split uniformly into a 6×6 grid of high granularity sub-cells (purple, unfilled distributions). The distributions shown are for the total visible energy above threshold deposited in the calorimeter (top, left), the number of hits above threshold (top, right) and the cell energy spectrum (bottom). For the number of hits and visible energy sum, the mean (μ_{90}) and standard deviation (σ_{90}) calculated on the central 90% of the distribution are shown.

Starting with the total visible energy, a clear reduction in the energy deposited in the calorimeter is observed when using the cell-level approach to placing the BIB-AE showers into the calorimeter. This is due to entire hits landing in the insensitive volume near to the core of the shower, which contains the most energetic part of the shower, resulting in a large reduction in the total energy deposited. When the output of the BIB-AE is split into the 6×6 higher granularity sub-cells, the loss of energy in the insensitive volumes of the calorimeter is significantly reduced, with a much lower shift observed in the distribution. While slightly less energy is deposited in the calorimeter with the GEANT4 version used for training (version 10.4) compared to the version used in *DDFastShowerML* (version 11.1), this does not account for the differences observed when using the BIB-AE for shower simulation. While splitting the hits in the BIB-AE shower into higher granularity sub-cells reduces the energy loss, it is still the case that significantly less energy is deposited than by the GEANT4 simulation.

Turning to the number of hits above threshold, the cell-level approach for depositing hits from the BIB-AE showers into the geometry shows a reduction in the number of hits due to them landing in the insensitive volumes. However, splitting the output of the BIB-AE into the higher granularity sub-cells creates significantly more hits than even that present in GEANT4. This is a result of the energy of the hits present in the shower in the regular grid, particularly in the regions of lower occupancy toward the edges of the shower, being split over multiple cells (see Figure 10.1).

The cell energy spectrum is influenced by a number of different factors. The peak observed for the cell-level projection (green histogram) of the BIB-AE showers in to the ILD ECAL arises from the fact that the model was trained on showers created in a perfectly regular geometry (see Figure 8.2). This feature is removed by splitting the hits in the regular grid into high granularity sub-cells (purple histogram) before depositing them into the irregular ILD ECAL geometry, after which the high energy tail of the spectrum aligns much more with the GEANT4 distributions. The opposite is true in the region of the spectrum from around 3 MeV down to about 0.3 MeV, where the cell-level distribution is much more closely aligned with the GEANT4 distributions, while significantly fewer hits created using the high granularity sub-cells have energies in this range. This a natural consequence of the energy of cells in the regular grid being divided into high granularity sub-cells, with more lower energy hits being created and the energy of existing cells in the regular grid being reduced. This effect is clear from the drastic increase in the low energy tail of the spectrum (below about 0.3 GeV) when dividing the regular grid into high granularity sub-cells, which extends below the MIP threshold. This explains why the total visible energy deposited using the high granularity splitting is still reduced with respect to the GEANT4 distributions, as hits with energies below the MIP threshold are lost.

Figure 10.4 shows the longitudinal profile (left) and radial profile (right) for the two GEANT4 versions and the two methods of depositing the regular grid BIB-AE showers back into the calorimeter. For the longitudinal profile, while both the cell-level and high granularity sub-cell approaches to depositing showers into the ILD ECAL show less energy in certain layers than the GEANT4 versions, the energy deposited using the cell-level method is systematically lower than the high granularity sub-cell approach. The relative difference between the cell-level and high granularity sub-cell approaches varies significantly, with the most pronounced drop in the cell level distribution occurring around layer 12. The strength of this effect will depend on the exact position of the edges of wafers, which create an insensitive volume, in a given layer and whether a cell-level hit lands in that volume. The exact layers in which these drops in energy occur will be highly dependent on the incident position in the calorimeter, as the effect will become more pronounced if the insensitive volumes in a layer lie closer to the core of a shower, where higher energy hits occur more often.

In the radial profile, the high granularity sub-cell approach tends to retain more energy closer to the core of the shower than the cell-level approach to projecting the BIB-AE shower in the regular grid. This is partially due to the large insensitive volume close to the shower

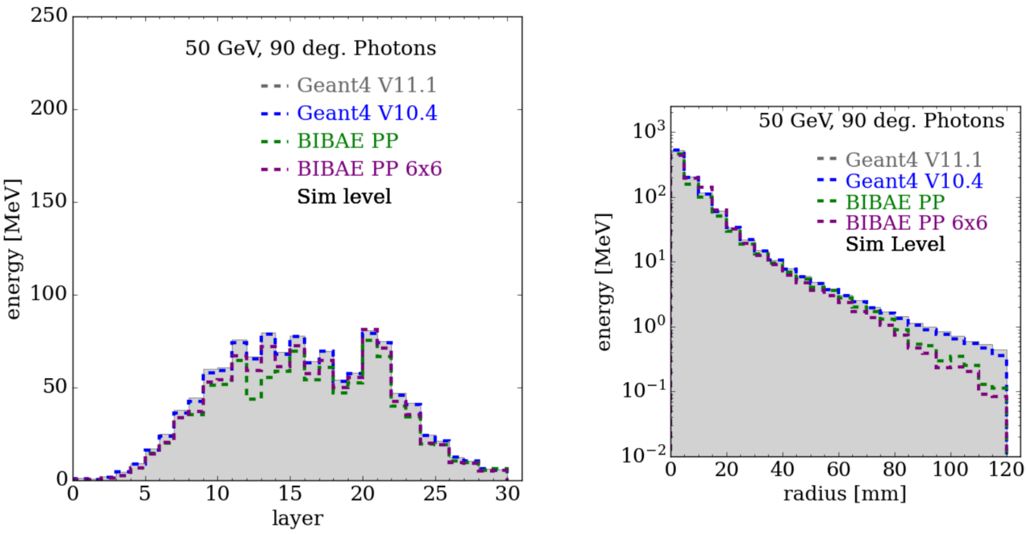


Figure 10.4: Distributions for 50 GeV photons incident orthogonal to the face of the calorimeter at *position 1*. Showers were simulated with two different GEANT4 versions: 11.1 (grey, filled distributions, used in the *DDFastShowerML* library) and 10.4 (blue, unfilled distributions, used for the training of the BIB-AE model). Showers were also simulated with the fully conditioned BIB-AE model, and projected back into the irregular ILD ECAL with two different approaches. In the first approach, hits are placed at the center of the cell in the regular grid (green, unfilled distributions) and in the second the energy of a hit is split uniformly into a 6×6 grid of high granularity sub-cells (purple, unfilled distributions). The distributions shown are for the longitudinal profile (left) and the radial profile (right).

core in *position 1*, but also because fewer isolated hits are present close to the shower core, so energy is typically only transferred between hits created in the irregular geometry of the ILD ECAL. The opposite is true at larger radii, where the energy deposited using the splitting into high granularity sub-cells falls below the cell-level distribution. This is a result of more isolated hits being present at larger radii, meaning that new hits are created, rather then energy simply being transferred between hits. Depending on the relative alignment of the regular and irregular grids, if hits are created with an energy lower than the MIP threshold they will be removed, explaining the drop in the distribution for the high granularity sub-cells. Note that the deviation between the BIB-AE and GEANT4 distributions present at larger radii is a result of the selection of a box around the shower required for the training of a regular grid model. The smallest distance from the side of the box to the center of the grid was 15 cells, corresponding to a radius of 75 mm where the difference starts to become pronounced. It should be noted that this still corresponds to more than 8 Molière radii for tungsten, meaning the vast majority of the energy in the shower is retained by the selection of the box.

Position 2

We now turn our attention to *position 2*, which does not contain any large insensitive volumes near to the core of the shower, and hence is a more representative scenario for a shower in the ILD ECAL.

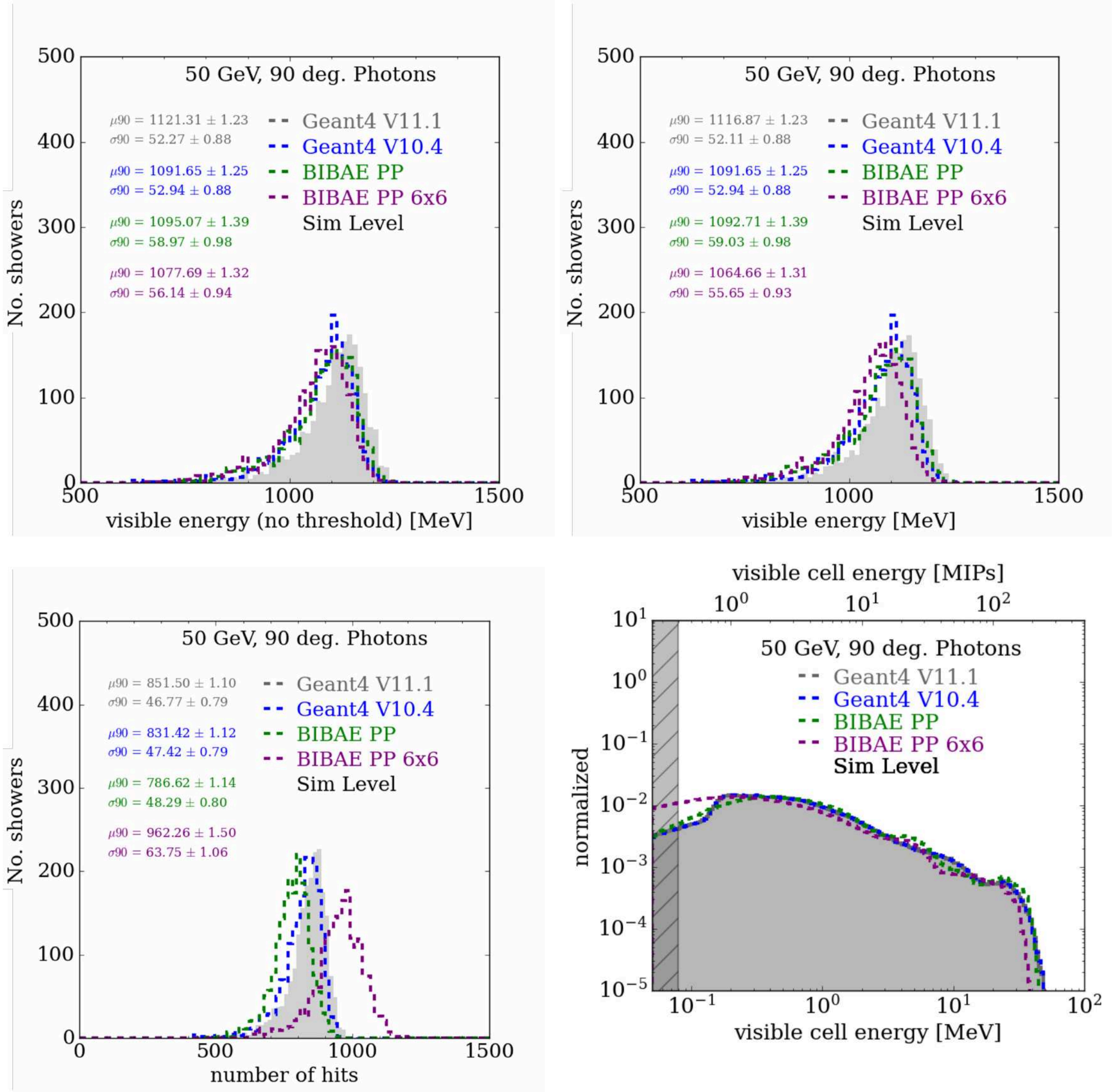


Figure 10.5: Distributions for 50 GeV photons incident orthogonal to the face of the calorimeter at *position 2*. Showers were simulated with two different GEANT4 versions: 11.1 (grey, filled distributions, used in the *DDFastShowerML* library) and 10.4 (blue, unfilled distributions, used for the training of the BIB-AE model). Showers were also simulated with the fully conditioned BIB-AE model, and projected back into the irregular ILD ECAL with two different approaches. In the first approach, hits are placed at the center of the cell in the regular grid (green, unfilled distributions) and in the second the energy of a hit is split uniformly into a 6×6 grid of high granularity sub-cells (purple, unfilled distributions). The distributions shown are for the total visible energy deposited in the calorimeter with no MIP threshold applied (top, left), the total visible energy above threshold deposited in the calorimeter (top, right), the number of hits above threshold (bottom, left) and the cell energy spectrum (bottom, right). For the number of hits and visible energy sums, the mean (μ_{90}) and standard deviation (σ_{90}) calculated on the central 90% of the distribution are shown.

Figure 10.5 shows distributions for the total visible energy without any threshold and with the MIP threshold applied (top left and right respectively), the number of hits above threshold (bottom, left) and the cell energy spectrum (bottom, right) for 50 GeV photons incident perpendicularly to the calorimeter face simulated with the two GEANT4 versions, and with the BIB-AE. The two different approaches for depositing the BIB-AE showers created in the regular grid back into the irregular geometry are shown in green for the cell-level approach, and purple for the high granularity sub-cell approach.

Starting with the visible energy before a threshold is applied, both the approaches at the cell-level and based on the splitting into high granularity sub-cells produce an energy sum that is closely aligned with the training GEANT4 version (version 10.4), with a closer agreement being observed than that seen between the GEANT4 versions themselves. The energy sum agrees well in both cases because the energy sum is well learned by the BIB-AE as a result of the combination of the NF model and the re-scaling procedure, and because this position contains significantly fewer insensitive volumes than *position 1*. This means that less energy is lost to the geometry during the placement of the shower. However, once the MIP threshold is applied, the picture changes. While the cell-level approach remains very close to the distribution from GEANT4 version 10.4, less energy is retained using the splitting of cells into high granularity sub-cells. This is because of the energy of cells being split of into low energy hits that fall below the MIP threshold.

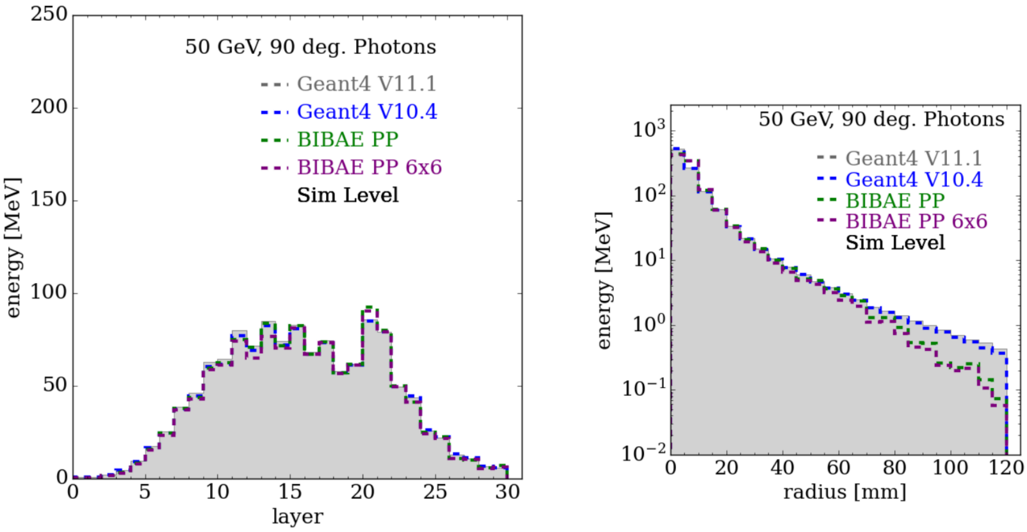


Figure 10.6: Distributions for 50 GeV photons incident orthogonal to the face of the calorimeter at *position 2*. Showers were simulated with two different GEANT4 versions: 11.1 (grey, filled distributions, used in the *DDFastShowerML* library) and 10.4 (blue, unfilled distributions, used for the training of the BIB-AE model). Showers were also simulated with the fully conditioned BIB-AE model, and projected back into the irregular ILD ECAL with two different approaches. In the first approach, hits are placed at the center of the cell in the regular grid (green, unfilled distributions) and in the second the energy of a hit is split uniformly into a 6×6 grid of high granularity sub-cells (purple, unfilled distributions). The distributions shown are for the longitudinal profile (left) and the radial profile (right).

Turning to the number of hits above threshold, the number of hits produced by the cell-level approach remains close to that from GEANT4 version 10.4, while the splitting of cells into the high granularity sub-cells creates significantly more hits. For the cell energy spectrum, a different picture compared to that seen at *position 1* is observed. The high energy tail drops off significantly from GEANT4 for the splitting into high granularity sub-cells, while this is not the case for the cell-level approach. This is a result of the energy of the highest energy hits in the shower being washed out by the splitting. The creation of lower energy hits by the splitting causes the tail to lower energies, which extends below the MIP threshold, to become significantly more pronounced.

Figure 10.6 shows the longitudinal profile (left) and radial profile (right) for the two GEANT4 versions and the two methods of depositing the BIB-AE showers created in the regular grid back into the irregular geometry. In comparison to the observations at *position 1*, the differences between the BIB-AE distributions and the GEANT4 version 10.4 distribution are noticeably smaller. In this case, the loss of hits due to the cell splitting dominates the effect of hits landing in insensitive materials, meaning that in several layers more energy is retained when using the cell-level approach, as opposed to the splitting into high granularity sub-cells. Finally, in the radial profile no major deviations are observed between the cell-level and high granularity sub-cell approaches near the core of the shower, as a result of no major insensitive volumes being present. However, at larger radii, the loss of hits after splitting due to their energies falling below the MIP threshold that was observed for *position 1* is still present.

10.3 Conclusions

In this chapter, we have investigated two different methods for placing a shower created in a regular grid with a generative model back into an irregular calorimeter geometry. Two different approaches to this shower placement were investigated. In the first, the hits in the regular grid were placed into the irregular geometry at the center of the regular grid cell, while in the second the energy of each hit in the regular grid was split uniformly into a 6×6 higher granularity regular grid of sub-cells, with the energy being deposited at the center of each sub-cell. These two approaches were studied at two different incident positions in the ILD ECAL, each with a different physical structure. In the first position, a large insensitive volume extending through the calorimeter layers was present near the shower core as a results of a gap between modules, whereas in the second position no large insensitive volumes that extended between layers were present. In the first position, an improvement was observed in terms of the total energy retained during the placement into the geometry by splitting hits into high granularity sub-cells. This was a result of significant energy loss in the insensitive volumes when placing hits at the cell-level. However, even with the splitting procedure applied, significant differences in terms of calorimeteric observables compared to full simulation with GEANT4 were still present. In the second position, the cell-level placement of regular grid showers produced with the BIB-AE performed significantly better when compared to the splitting of cells into high granularity sub-cells. This was a result of more energy being retained at cell-level due to fewer insensitive volumes being present, and the splitting of energy from hits in the regular grid to create lower energy hits which fell below the MIP threshold and were therefore discarded. Since the irregularities present in the first position are relatively rare in the ILD ECAL, the second position is much more representative of the standard case in the ILD ECAL. This means that having large insensitive volumes very close to the core of the shower is relatively rare. For this reason, in the remainder of this thesis, the cell-level approach to depositing energy will be adopted.

While this chapter has focused specifically on the BIB-AE model and the effects of the irregularities present in the ILD ECAL, the results indicate more general difficulties when using a generative model producing a regular grid to simulate a highly granular irregular calorimeter geometry. Other potential options for placing showers created in a regular grid into an irregular calorimeter geometry could be explored in the future. The most straightforward approach would be to randomly select the position within a cell in the regular grid used to deposit the hit into the irregular geometry, as opposed to simply taking the center of the cell as was done here. On average this would likely reduce the energy loss in an insensitive volume that extends between layers, such as that explored in *position 1*. This is because this approach would prevent a whole line of hits in consecutive layers from being lost. However, in the more general case of not having such a drastically different physical structure, if the regular and irregular grids are not perfectly aligned it would be possible to combined multiple hits in the regular grid into a single hit in the irreglar geometry. This would therefore likely affect the number hits and the radial profile. Fundamentally, without access to the positioning of cells and composition of the local irregular geometry, it is likely not possible to achieve a loss-less placement of a shower produced in a regular grid into an irregular geometry.

Another option would be to explore the use of a generative model that does not rely on a regular grid, such as that adopted by Buhmann, P.M. et al. in [6]. The use of such a point cloud model allows the model to learn the shower at a much higher granularity than the level of a physical readout cell, as the model is not affected by the significant increase in sparsity. While this inspired the uniform splitting of hits in the regular grid into a 6×6 higher granularity that was explored in this chapter, it is likely to provide a much better description of observables after placement into an irregular geometry. This is because the model is able to learn the physical distribution of energy within a cell, rather than the artificially uniform splitting of energy explored here.

Even with this approach, it may not be possible to achieve a sufficiently high fidelity description of showers in positions with large gaps in the geometry, and therefore a significantly different physical structure, such as *position 1*. This may therefore necessitate the use of the *Tigger* interface in *DDFastShowerML* described in Chapter 9 to exclude this region for the model trained on the planar calorimeter structure, and the use of a suitable generative model with a dedicated training in this region. Another possible solution could be to create 'gap correction' for the fast simulation in these specific regions of the calorimeter. This is already applied in the ILD reconstruction chain to account for the difference in sampling fraction present in full simulation (which is also applied during reconstruction for showers created with fast simulation, see Section 4.2.3), and an additional but similar, dedicated procedure for fast simulation would be a method of correcting the total visible energy. This procedure would therefore be specific to a particular detector geometry.

Chapter 11

Physics Benchmark

With a fully conditioned generative model capable of running seamlessly with full simulation in GEANT4, and of producing showers in the realistically detailed ILD ECAL, it is now possible to study benchmarks of the performance of the model after reconstruction with full physics processes. In Section 11.1, the physics process chosen as a first benchmark of the model will be described, while in Section 11.2 the creation of the dataset will be described. This will include the selection of samples at the generator level, as well as the approach to detector simulation and reconstruction. In Section 11.3, a study will be performed into the effects of using the BIB-AE model for electromagnetic shower simulation on reconstruction and key physics observables, before conclusions are drawn in Section 11.4.

11.1 Photons from Neutral Pions Produced in the Decay of Tau Pairs

The physics performance required from a fast simulation tool will depend heavily on the physics process under study. A natural choice of physics processes with which to benchmark the performance of a fast simulation tool for electromagnetic showers are the hadronic decay modes of the tau lepton in the process $e^+e^- \to \tau^+\tau^-$. As described in Section 2.4, such a decay commonly results in one or more boosted neutral pions, which quickly decay into photons. These factors make such a process a stringent test of the performance of a generative model used for fast simulation of photon showers, as to the high overlap of the showers produced in the calorimeters requires a high degree of simulation fidelity due to the sensitivity of the reconstruction procedure.

A study into the effect of using a generative model for photon shower simulation in this process can be broken down into two parts. In the first instance, the effects on the reconstruction of photons, which are the direct products of the model, can be studied. This can give more direct indications of which properties of the showers simulated with the model may cause differences in reconstruction in a realistic physics environment. Secondly, the effects on the reconstruction of the π^0 which produced the two photons can be investigated. As a higher level reconstructed object, the reconstructed information about these particles feeds directly into subsequent physics analyses, and as such determines whether or not a model is suitable for simulation of the given process.

11.2 Dataset Creation

In order to make a direct comparison of the photon showers simulated with the BIB-AE and with GEANT4, an identical set of events from the generator is taken, with only the detector simulation differing between samples. This removes the effects of different underlying event topologies and physics processes, which would dilute any differences in the reconstruction performance between the two simulators, making a direct comparison harder. Additionally, no background is overlaid onto events (see Chapter 6). Overlaying a background would end up increasing the detector occupancy and likely degrade the reconstruction performance, for example as a result of increased overlap between showers in the calorimeter. While this is the more realistic simulation case, it would make disentangling the effects on the reconstruction performance stemming from the simulation of photon showers in the calorimeter more challenging, and hence is not appropriate for this study.

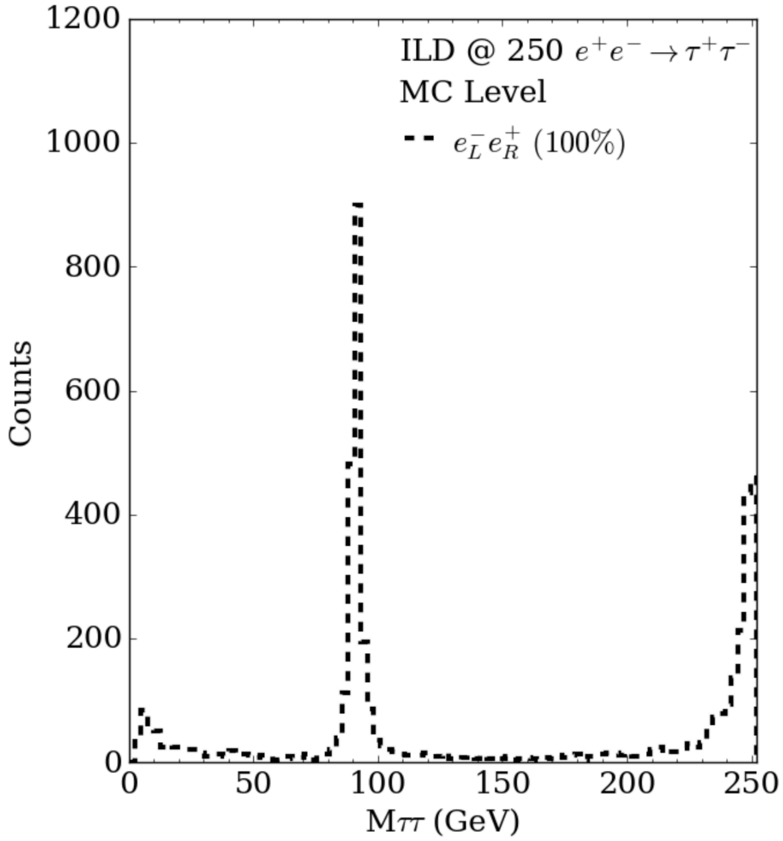


Figure 11.1: Invariant mass of the di-tau system in the process $e^+e^- \to \tau^+\tau^-$, using MC level information from the generator.

The generator samples used for this study were created by the ILD Software Working Group during the MC-2020 [201] production. The samples used were produced for an ILC physics environment running at a center-of-mass energy of 250 GeV. Event samples of the process $e^+e^- \to \tau^+\tau^-$ were generated using $\mathrm{WHIZARD}$ [179] version $2.8.5$, with a realistic ILC beam energy spectrum, and included the effects of beamstrahlung and initial state radiation. Samples were created using 100% polarised left-handed electron and right-handed positron beams $(e_L^- e_R^+)$. The tau leptons were decayed with the TAUOLA library [202]. The invariant mass of the di-tau system in the process $e^+e^- \to \tau^+\tau^-$ is shown in Figure 11.1. The pronounced peak around the Z pole (at approximately 91 GeV) is a result of radiative return events in which one or more ISR photons are produced, reducing the center of mass energy and resulting in a peak in the invariant mass around the Z mass.

A subset of events from these samples was selected for detector simulation based on the following criteria. It was required that at least one of the taus underwent a hadronic decay that resulted in the production of at least one neutral pion. Additionally, it was required that both the photons produced by at least one of the π^0s resulting from a tau decay had an energy of more than 10 GeV. This selection was made such that the photons could be of a sufficiently high energy, if they reached the calorimeter, to warrant the use of the BIB-AE to simulate the shower. As is shown in the computational benchmark in Chapter 9, for photons incident with energy less than this approximate threshold, the time to simulate a shower with GEANT4 drops below the simulation time of the BIB-AE. This means that it makes little sense to use the BIB-AE to simulate showers for low energy particles. More information on the effect of these selections on the number of π^0s can be found in Appendix D.

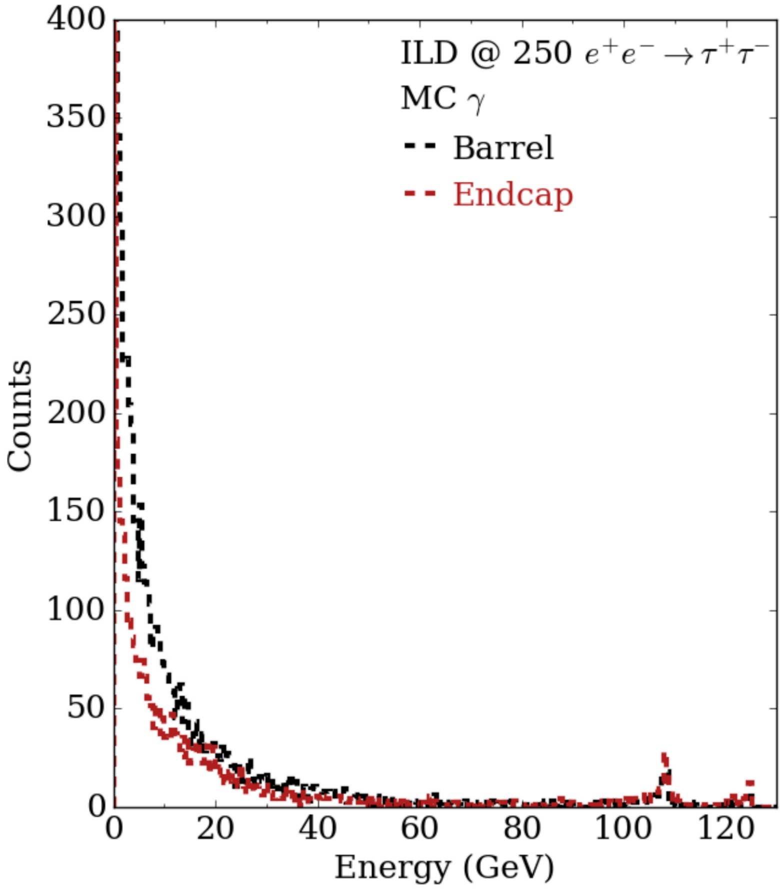


Figure 11.2: Energy spectrum for photons created the process $e^+e^- \to \tau^+\tau^-$, using MC level information from the generator. Distributions are shown for those photons which hit the barrel (black), and those which hit the endcap (red).

The energy spectrum for photons at the level of the generator in the process $e^+e^- \to \tau^+\tau^-$, separated into distributions for those photons which hit the barrel and those which hit the endcap, is shown in Figure 11.2. This includes all photons in the event which are incident on these detector subsystems, not just those produced in the π^0 decays studied, as all photons in the event are relevant to the simulation time. The majority of photons have energies well below 10 GeV. The slight peak that occurs at around 108 GeV arises as a result of ISR photons produced in radiative return events. Given the steeply falling nature of the photon energy spectrum to higher energies, together with the fact that the BIB-AE could only sensibly be used to simulate electromagnetic showers with an energy of more than 10 GeV, performing electromagnetic shower simulation with the BIB-AE for the samples selected from the process $e^+e^- \to \tau^+\tau^-$ was only a factor of approximately $\times 1.2$ faster than using pure GEANT4.

Three approaches were adopted at the level of the detector simulation, with each using identical generator level events selected by the above procedure. In the first two, all particles in the detector were simulated using full simulation with GEANT4, with one simulation performed using the Key4hep [200] ecosystem, with Geant4 version 11.1 and DD4hep version 1.27 (used with the *DDFastShowerML* library) and one simulation performed using the iLCSoft [126] ecosystem (see Section 4.2.2), with full simulation performed with Geant4 version 10.4, and DD4hep version 1.11 (which was used to train the BIB-AE). In the third sample, a mixture of fast simulation and full simulation was used. In this sample, the BIB-AE was used to simulate all electrons, positrons and photons which were incident to the endcap or barrel of the ILD ECAL with an energy of more than 10 GeV, as long as the particle additionally passed all the geometrical triggers applied during simulation discussed in Chapter 9. All other particles in this sample were simulated using Geant4 version 11.1 and DD4hep version 1.27, however for brevity this will be referred to as the BIB-AE sample from now on. In all cases, the QGSP_BERT physics list was used in Geant4. The crossing angle of the beams in the ILC was simulated by applying a Lorentz boost to generator level quantities in the x' direction.

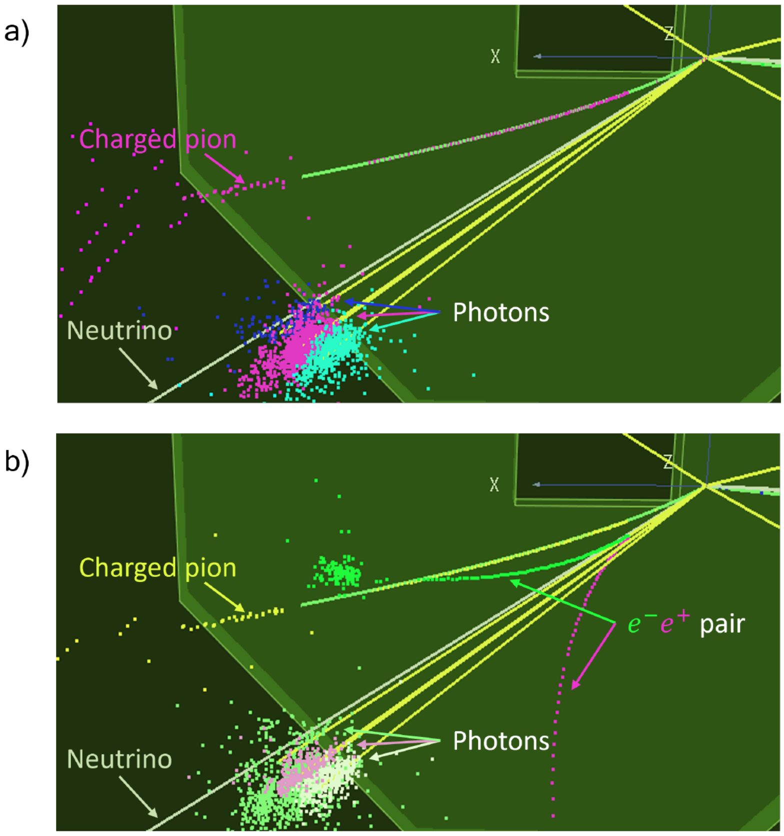


Figure 11.3: Two example events simulated fully with Geant4 where a tau lepton decays into a charged pion, two π^0s and a tau neutrino, via the process $\tau^- \to a_1(1260)^- \to \nu_\tau\pi^-\pi^0\pi^0$. Although the event is identical at the level of the generator in both cases, the outcome of reconstruction is heavily dependent on what occurs during the detector simulation. a) The two π^0s each decay into a pair of boosted photons, leading to four intermingled photon showers in the ECAL endcap. Due to the high overlap, two photon showers are merged into one, leading to only three photons being reconstructed. No π^0 is reconstructed. b) One of the photons converts into an e^+e^- pair prior to reaching the calorimeter. This gamma conversion is identified, leading to two π^0 candidates being reconstructed. In each case the particle labels are for the reconstructed particles, with the exception of the neutrino, which leaves no signature in the detector and escapes undetected.

Each of the three detector simulation approaches were repeated three times, with independent random seeds used in Geant4 in each case. This provides a means to quantify the uncertainties on reconstruction level quantities, given the high level of correlation between the three approaches that would otherwise be difficult to estimate. This correlation arises from using identical generator level input. For each of these three random seeds used for each of the three detector simulation approaches, $9,000$ $e^+e^- \to \tau^+\tau^-$ events passing the generator level selection criteria were simulated, each using an identical set of events from the generator. This resulted in 9 different datasets. All simulated events were then reconstructed using the standard reconstruction procedure adopted by ILD described in Section 4.2.3.

To understand the drastic effect that using different seeds for the detector simulation can have on the reconstruction performance for an event, consider the two different events shown in Figure 11.3. Full GEANT4 simulation is used in both cases, and both have an identical input from the generator, with a tau lepton decaying into a charged pion, two π^0s and a tau neutrino via an intermediate a_1 resonance. The particles which have been reconstructed with the ILD standard reconstruction scheme are labelled, with the exception of the neutrino, which leaves no signature in the detector and escapes undetected. Each of the two π^0s decay into two photons, with the trajectories of the generator level photons shown by the straight yellow lines. This results in four highly collimated photons being simulated by GEANT4. In a), all four photons impact upon the calorimeter, creating four overlapping photon showers. The overlapping of the showers leads to two of the photons being merged into one by the reconstruction procedure, meaning only three photons are reconstructed. The incorrect reconstruction of the photons means that no π^0s are reconstructed. In b), one of the photons converts into an e^+e^- pair prior to hitting the calorimeter. These charged particles bend in the magnetic field of the detector, and as such the showers they create in the ECAL are clearly separated from the showers created by the photons. This gamma conversion is identified by the reconstruction procedure, meaning that all four photons can be correctly reconstructed, and two π^0 candidates identified. This highlights how the different physics processes that occur during the detector simulation can directly impact whether reconstruction is successful or not.

11.3 Reconstruction Performance Comparison

We now study the reconstruction performance for each of the three detector simulation approaches, focusing on the reconstruction of the π^0s which are produced by the decays of tau leptons. We begin by investigating the reconstruction performance broadly, by considering all π^0s produced by a tau decay.

A summary of the results is shown in Table 11.1. In total, each of the 9 datasets contain 16693 π^0s created by the generator (i.e MC Truth π^0s). In the remainder of the table, the errors are derived from the repeats with three different seeds for each of the three different simulation approaches, and the value shown is the average over the seeds. In total, 8942 ± 69 π^0s are reconstructed in the sample simulated with GEANT4 V11.1, 9021 ± 119 π^0s are reconstructed in the sample simulated with GEANT4 V10.4 and 9192 ± 130 π^0s are reconstructed in the BIB-AE sample. This means that the total number of π^0s reconstructed is consistent across the different samples within errors. A different pattern emerges when looking into the quality of the reconstruction. To this end we record the number of π^0s satisfying four additional criteria. In each case, a link between MC Truth particles and reconstructed particles is used, with a weight given by the relative (energy weighted) contribution of an MC Truth particle to a reconstructed particle. In the case of multiple reconstructed particles being linked to an MC Truth particle (or vice versa), the relation with the largest weight is used. Firstly, the number of correctly reconstructed π^0s is shown (*No. Good*). While the BIB-AE sample contains the highest number of π^0s correctly reconstructed, this is consistent with the number of correctly reconstructed π^0s in the GEANT4 V10.4 sample within error. Notably, the number of correctly reconstructed π^0s in the GEANT4 V11.1 sample is significantly lower than in the GEANT4 V10.4 and BIB-AE samples. Secondly, the number of π^0s missed (*No. Missed*) is shown, meaning that there was no reconstructed π^0 linked to a given MC Truth π^0. In this case,

Table 11.1: Table showing the number of π^0s under certain conditions for simulation samples produced using GEANT4 V11.1, GEANT4 V10.4 and the BIB-AE. In each case, values are reported for the number of MC Truth π^0s (*No. True*), number of reconstructed π^0s (*No. Reco*), number of correctly reconstructed π^0s (*No. Good*), number of π^0s not reconstructed (*No. Missed*), number of π^0s reconstructed with one photon correct (*No. Confused*) and the number of π^0s reconstructed when no MC Truth π^0 was present (*No. Fake*). Values reported with errors are means over three different random seeds, with the error arising from the standard deviation.

π^0s	No. True	No. Reco	No. Good	No. Missed	No. Confused	No. Fake
GEANT4 V11.1	16693	8942 ± 69	2452 ± 33	12843 ± 27	1398 ± 33	5092 ± 80
GEANT4 V10.4	16693	9021 ± 119	2545 ± 35	12789 ± 35	1359 ± 10	5117 ± 96
BIB-AE PP	16693	9192 ± 130	2576 ± 16	12720 ± 2	1397 ± 16	5219 ± 128

the two GEANT4 versions are consistent with one another within error, while for the BIB-AE sample significantly fewer π^0s are missed by the reconstruction. Thirdly, the number of π^0s which have been reconstructed incorrectly (*No. Confused*) is shown. This means that one of the photons combined into the reconstructed π^0 was reconstructed correctly (i.e. linked to an MC Truth photon), while the other was not. In this instance, the BIB-AE and GEANT4 V11.1 agree with one another well, while the number of confused π^0s in the GEANT4 V10.4 sample is markedly lower. Finally, the number of π^0s that were reconstructed but not linked to an MC Truth π^0 (*No. Fake*) is shown in the last column. The number of fake π^0s is consistent across all samples within error.

It should be noted that the π^0 reconstruction efficiency is observed to be low, with a value of approximately 54%. Additionally, the disagreements observed between GEANT4 versions indicate that this is a major source of the variations in the reconstruction performance observed, often dominating the deviations caused by using the BIB-AE.

We now turn to a more detailed comparison of physics observables for reconstructed π^0s between the three different detector simulation approaches. As before, selections are placed on the MC Truth π^0s. It is required that each π^0 is produced by the decay of a tau lepton at the level of the MC Truth particle, that both the MC Truth photons produced by the π^0 decay have an energy above 10 GeV and that both photons pass the geometry triggers applied during simulation, which were described in Chapter 9. It was also required that both of the reconstructed photons combined into the π^0 were correctly linked to an MC Truth photon from a π^0 produced in a tau decay.

We first study the reconstructed photons. Figure 11.4 shows histograms for the number of reconstructed photons against the difference between the energy of reconstructed and corresponding MC Truth photons. In the main panel, the distributions for GEANT4 V10.4 (blue), GEANT4 V11.1 (grey) and the BIB-AE (green) are shown. Each distribution is the average of the three different random seeds, with an error arising from the standard deviation over the three seeds. In the sub-panel, the ratio of each of these three distributions to the GEANT4 V10.4 distribution is shown. The errors are propagated into the ratio by means of a sum in quadrature, with no correlation coefficient included. While deviations in the BIB-AE distribu-

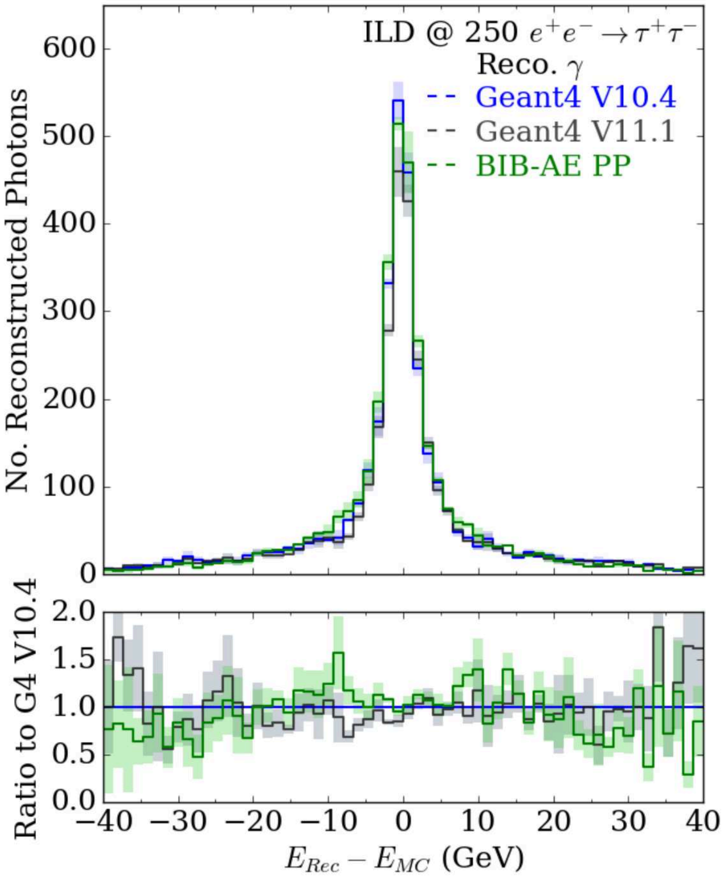


Figure 11.4: Number of photons reconstructed as a function of the difference between the energy of the reconstructed and MC photon for the $e^+e^- \to \tau^+\tau^-$ samples studied. Distributions are shown for each of the three simulation approaches; Geant4 V10.4 (blue), Geant4 V11.1 (grey) and the BIB-AE (green). Each distribution is the average over the three different random seeds used in the simulation, with an error arising from the standard deviation over the three seeds. The sub-panel shows the ratio of each distribution to Geant4 V10.4, with the error being propagated into the ratio by means of a sum in quadrature with a correlation coefficient of zero.

tion from the Geant4 V10.4 distribution are observed, these deviations are on a similar level to the deviations present in the Geant4 V11.1 distribution. Larger fluctuations are observed in the ratios for the tails of distributions, where the statistics become limited. The most noticeable deviations observed in the BIB-AE distribution occur around the ± 10 GeV difference from MC region. An excess in the number of reconstructed photons in the BIB-AE samples is observed, with the central value of the maximum deviation reaching a factor of approximately 1.5. Interestingly, the deviation at -10 GeV is almost mirrored by a reduction in the number of photons reconstructed in the Geant4 V11.1 sample around this region. These observations indicate that the differences between the BIB-AE sample and Geant4 V10.4, which was used to produce the training data, are on a similar level to the deviations between Geant4 versions, which quantifies the uncertainty in our understanding of electromagnetic shower development that is used during the MC simulation.

Figure 11.5 shows histograms for the number of reconstructed photons against the polar angle reconstructed in the global ILD coordinate system. This is also known as the *intrinsic* polar angle of the reconstructed cluster ($iTheta$), which is obtained from a principle component analysis during reconstruction. This is analogous to the angular reconstructions performed in Chapters 7 and 8, and hence gives an indication of the angular conditioning performance of

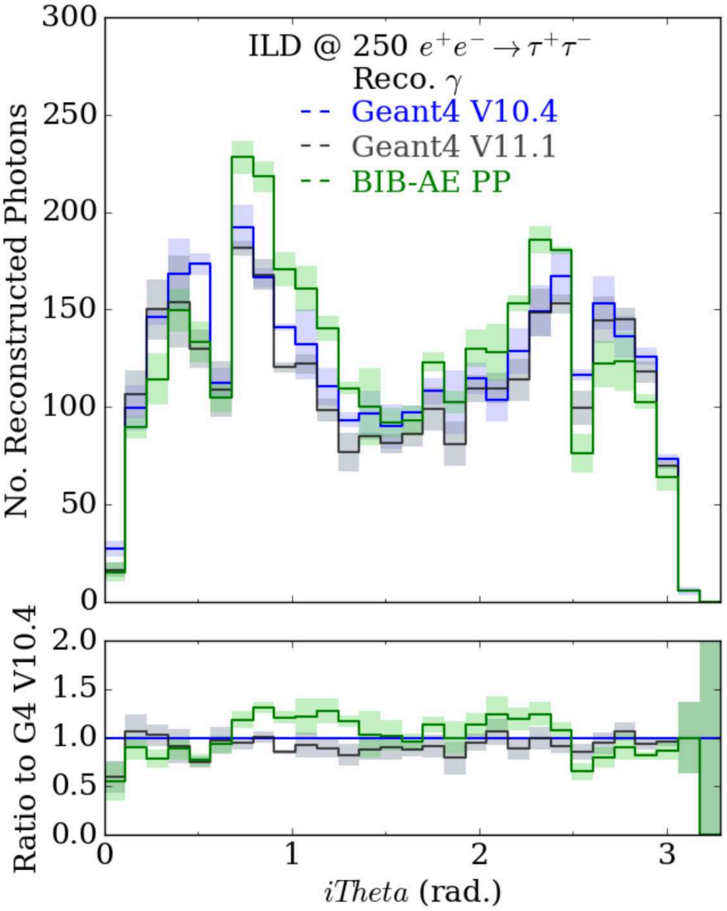


Figure 11.5: Number of photons reconstructed as a function of the intrinsic polar angle of the reconstructed photon cluster for the $e^+e^- \to \tau^+\tau^-$ samples studied. Distributions are shown for each of the three simulation approaches; GEANT4 V10.4 (blue), GEANT4 V11.1 (grey) and the BIB-AE (green). Each distribution is the average over the three different random seeds used in the simulation, with an error arising from the standard deviation over the three seeds. The sub-panel shows the ratio of each distribution to GEANT4 V10.4, with the error being propagated into the ratio by means of a sum in quadrature with a correlation coefficient of zero.

the model in this more complex physics environment. In the main panel, the distributions for GEANT4 V10.4 (blue), GEANT4 V11.1 (grey) and the BIB-AE (green) are shown. Each distribution is the average of the three different random seeds, with an error arising from the standard deviation over the three seeds. In the sub-panel, the ratio of each of these three distributions to the GEANT4 V10.4 distribution is shown. The errors are propagated into the ratio by means of a sum in quadrature, with no correlation coefficient included. The distributions reflect the structure of the detector along the direction of the global polar angle, with the pronounced dips at approximately 0.64 and 2.5 radians stemming from the cuts placed on the MC Truth particle to emulate the geometry triggers introduced in the simulation chain for the transition region between the barrel and the endcap. In this case, while a comparatively good agreement is observed between the two GEANT4 versions, the BIB-AE shows some excesses in the number of photons reconstructed towards the edges of the barrel. Since the photons travel on a straight line from the IP, as they result from the instantaneous decay of π^0s, the edges of the barrel correspond to the most inclined incident angles relative to the face of the calorimeter (i.e. in the local calorimeter coordinate system). This could therefore indicate that the performance of the model is somewhat reduced for the most inclined showers.

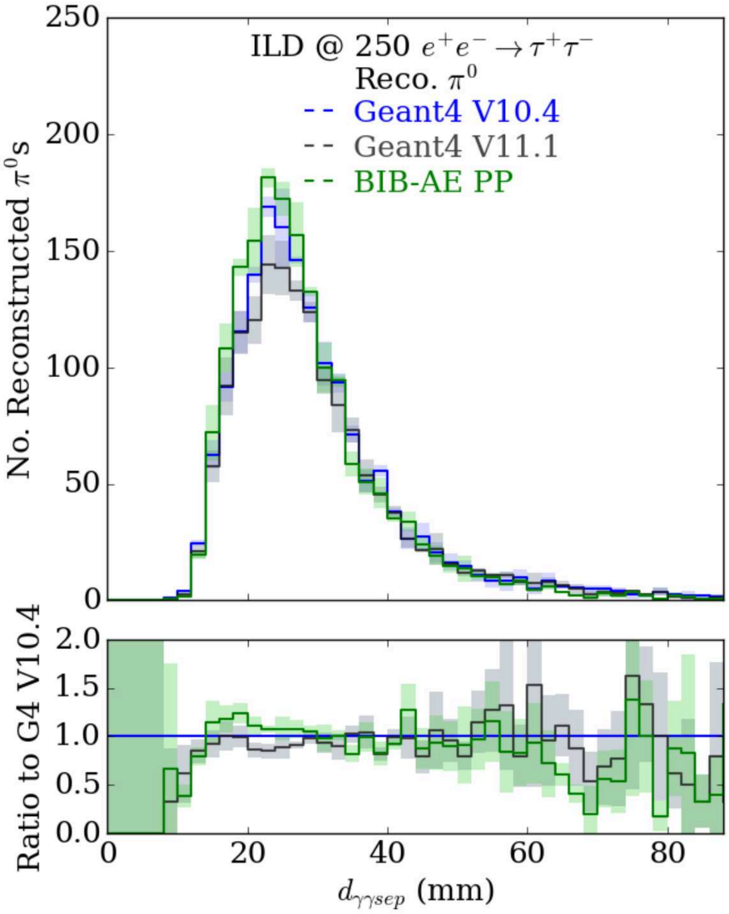


Figure 11.6: Number of π^0s reconstructed as a function of the separation between the two photon clusters combined into the π^0 for the $e^+e^- \to \tau^+\tau^-$ samples studied. Distributions are shown for each of the three simulation approaches; GEANT4 V10.4 (blue), GEANT4 V11.1 (grey) and the BIB-AE (green). Each distribution is the average over the three different random seeds used in the simulation, with an error arising from the standard deviation over the three seeds. The sub-panel shows the ratio of each distribution to GEANT4 V10.4, with the error being propagated into the ratio by means of a sum in quadrature with a correlation coefficient of zero.

We now study the reconstructed π^0s. Figure 11.6 shows histograms for the number of π^0s reconstructed against the separation between the two photon clusters combined into the reconstructed π^0s. The separation is calculated using the barycenters of the clusters. Looking for differences between the BIB-AE and GEANT4 in the reconstruction performance in terms of separating photons provides a means to verify if the modelling of the radial profile by the BIB-AE is of a sufficient quality. In the main panel, the distributions for GEANT4 V10.4 (blue), GEANT4 V11.1 (grey) and the BIB-AE (green) are shown. Each distribution is the average of the three different random seeds, with an error arising from the standard deviation over the three seeds. In the sub-panel, the ratio of each of these three distributions to the GEANT4 V10.4 distribution is shown. The errors are propagated into the ratio by means of a sum in quadrature, with no correlation coefficient included. For a separation of less than about 10 mm between the position of the photon clusters, no π^0s are reconstructed. This is because for small separations, the two photons can no longer be distinguished and are merged into a single cluster. The transition for this cluster merging occurring around 10 mm corresponds to twice the size of a cell, and is a documented performance limit of PandoraPFA [121] (see Section 4.1). As the photon separation is increased above approximately 10 mm, the two photons can be separated more easily, and the distribution rises sharply to a peak at a separation of about 25 mm. After this point, the distribution falls away into a long tail to larger radii, with a combinatorial effect coming into play for the two photons selected folded with the fact that the π^0s tend to be highly boosted, and hence large photon separations suppressed. A good agreement between the three simulation approaches is seen overall. While larger fluctuations are observed in the ratio plot for the long tail of the distribution, the clearest differences occur for the rising edge of the distribution for separations below the peak. While a slight excess in the number of reconstructed π^0s relative to Geant4 V10.4 is observed for the BIB-AE sample, this is mirrored by a slight deficit in the number of π^0s reconstructed in the Geant4 V11.1 sample relative to Geant4 V10.4. This suggests that the deviations present in the BIB-AE sample are of a similar magnitude to the differences between Geant4 versions.

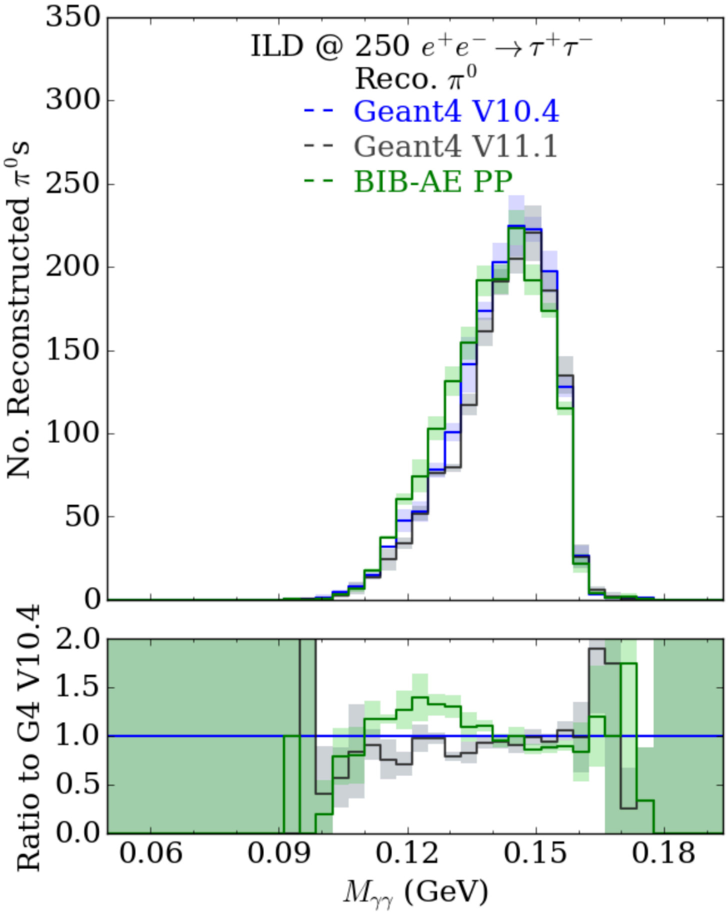


Figure 11.7: Number of π^0s reconstructed as a function of the invariant mass of the two photons combined into the π^0 for the $e^+e^- \to \tau^+\tau^-$ samples studied. Distributions are shown for each of the three simulation approaches; Geant4 V10.4 (blue), Geant4 V11.1 (grey) and the BIB-AE (green). Each distribution is the average over the three different random seeds used in the simulation, with an error arising from the standard deviation over the three seeds. The sub-panel shows the ratio of each distribution to Geant4 V10.4, with the error being propagated into the ratio by means of a sum in quadrature with a correlation coefficient of zero.

Finally, Figure 11.7 shows histograms for the number of π^0s reconstructed against the invariant mass of the two photons clustered into the π^0. The invariant mass $M_{\gamma\gamma}$ of the di-photon system was calculated using the reconstructed energy of photon i E_i and photon j E_j, as well as the opening angle between their reconstructed flight directions η, according to

$$M_{\gamma\gamma} = \sqrt{2E_iE_j(1 - \cos(\eta))}. \tag{11.1}$$

In the main panel, the distributions for Geant4 V10.4 (blue), Geant4 V11.1 (grey) and the BIB-AE (green) are shown. Each distribution is the average of the three different random seeds, with an error arising from the standard deviation over the three seeds. In the sub-panel, the ratio of each these three distributions to the Geant4 V10.4 distribution is shown. The errors are propagated into the ratio by means of a sum in quadrature, with no correlation coefficient included. Both the distribution obtained for the BIB-AE sample and for the Geant4 V11.1 sample show deviations in the rising tail below the peak of the distribution. In the case of the BIB-AE sample, significantly more π^0s are reconstructed than in the Geant4 V10.4 sample in this region, while in the case of the Geant4 V11.1 sample, fewer π^0s are reconstructed. It should be noted that the excess in the case of the BIB-AE sample is noticeably more pronounced than the deficit in the case of the Geant4 V11.1 sample. This suggests that the deviations in the BIB-AE sample are larger than the differences between Geant4 versions.

11.4 Conclusions

In this chapter, we have studied the performance of the BIB-AE model used for electromagnetic shower simulation in the process $e^+e^- \to \tau^+\tau^-$. The effect on the reconstruction performance was investigated for photons from π^0s produced in hadronic decays of the tau, with both the reconstruction of the photons directly, and the subsequent reconstruction of the π^0s being studied. This allowed the performance of the model to be studied in a much more complex and realistic physics environment than could be done previously. A comparison was made between a sample including BIB-AE shower simulation and two different Geant4 versions, showing the variation in reconstruction performance between versions and giving an indication of the MC uncertainty on Geant4. Uncertainties for each sample were also estimated by means of repeated simulations with different random seeds. These steps allowed the definition of a clearer target in terms of the simulation quality required of the BIB-AE model. In many cases, the deviations of the BIB-AE model from the Geant4 version used for training were on a similar level to the deviations observed between Geant4 versions. This was the case for the difference between the reconstructed and MC Truth energies for the photons produced by neutral pion decays and for the separation between the photon clusters combined into the reconstructed π^0. However, deviations that appear larger than the differences between Geant4 versions were observed in the intrinsic polar angle reconstructed for the most inclined showers at the edges of the barrel, and in regions of the invariant mass distribution for the di-photon system. The latter case is particularly significant, as it is a physics observable directly used in downstream analysis. It should also be noted that the simulation speed-up relative to Geant4 achieved was low, at only a factor of about $\times 1.2$. This is a result of a large majority of photons produced in the process having an energy below the threshold at which the BIB-AE could be used to simulate showers faster than Geant4. If a generative model with better energy scaling was used instead, thereby providing a speed-up relative to Geant4 for lower energy photons, this speed-up factor could increased.

Chapter 12

Conclusions

The simulation requirements of experiments in high energy physics place large burdens on the available computing infrastructure, and present a significant hurdle for the operation of future experiments. The use of full Monte Carlo based simulation for showers which develop in the calorimeter systems of detectors are the most computationally intensive components of a complete detector simulation. These challenges are particularly formidable when seeking to develop fast simulation tools for showers in high granularity calorimeters, which demand a high degree of physics fidelity from a simulator. This thesis has sought to address these issues by developing a fast simulation tool for high granularity calorimeters based on deep generative models.

To develop such a tool, it is necessary to generalise the scope of initial proof-of-concept studies which have investigated the performance of generative models for this task in restricted scenarios. This includes producing a correct detector response for particles incident with various angles to the calorimeter, and at different positions on the calorimeter surface.

To this end, in Chapter 7, the performant BIB-AE model was extended to handle photons incident with a varying angle and energy, but with a fixed position, on the ILD ECAL [64]. We found that the BIB-AE was able to model key calorimetric observables with a high degree of fidelity when compared to GEANT4, the state-of-the-art simulation tool.

We then took a further step and interfaced the simulation level output with the full reconstruction chain used by ILD, allowing the single particle reconstruction performance of the model to be studied. Importantly, it was demonstrated that the physics fidelity provided by the model remained high after reconstruction. This is a major step in the development of a generative model for fast simulation of showers in high granularity calorimeters, as the ultimate metric by which a model is judged is the physics performance after reconstruction.

However, this was not yet enough to build a complete simulation tool. Chapter 8 therefore addressed the development of a fully conditioned model. Firstly, to be able to simulate showers at different positions, it was necessary to take steps to handle the irregular nature of real calorimeter geometries. A key problem is that due to the variation in the relative fraction of insensitive volumes at different positions in the calorimeter, it is hard to place a model generated shower into different positions in the planar region of a calorimeter. Additionally, when using a generative model relying on regular grids, such as the BIB-AE, projecting the irregular detector geometry into a regular grid creates artefacts. A perfect regular version of the ILD ECAL, with purely sensitive active layers was therefore created.

Secondly, the BIB-AE was then extended to accept conditioning on a second angle, enabling a fully conditioned model to be trained for electromagnetic shower simulation. During the creation of the dataset, the size of the regular grid was minimized in order to maintain a reduced simulation time per shower. While the fidelity of the BIB-AE's modelling of certain calorimeteric observables remained high, others were less well modelled than when the model was conditioned on a single angle in Chapter 7. This could be a result of the fact that the conditioning phase space was significantly expanded, but the number of training samples were kept constant. Generally, the number of training samples needed to obtain a sufficiently performant model is an open question and could be explored in future work. Additionally, a more rigorous tuning of the hyperparameters of the BIB-AE for the sample could yield an improved performance. Nevertheless, this study constitutes the first time a generative model has been developed with a broad enough conditioning scheme to be used at different positions in the planar region of a highly granular calorimeter.

With this fully conditioned model developed, Chapter 9 addresses the important task of running the generative models in the existing software ecosystems. We have developed a generic library that is capable of seamlessly mixing full and fast simulation with generative models in realistic detector geometries. The BIB-AE model developed in Chapter 8 was then integrated into the library, allowing the various features of the library to be tested and enabling further studies with this model. For example, a fair benchmark of the computational performance of the model in comparison to GEANT4 was then possible. It was demonstrated that the per-shower simulation speed-up relative to GEANT4 increased with the incident particle energy, with the BIB-AE being $\times 11$ faster then GEANT4 for the simulation of 100 GeV photon showers.

While this study has focused on the integration of the BIB-AE for use with the ILD ECAL, it should be noted that the library is sufficiently flexible to support the use of other models including those which do not rely on a regular grid output. Additionally, the library can support the use of shower simulation in other detectors, given that they have an implementation in DD4HEP and a correspondingly trained generative model. The library therefore has the potential to serve future developments in this field by the community.

An important question that has received little attention thus far is the effect of using a generative model to simulate showers in an irregular geometry. As with the vast majority of models explored thus far in the field, the BIB-AE was trained to generate showers at the level of the readout geometry of the calorimeter. The effects of projecting showers generated with the model into different positions in an irregular detector geometry are therefore not clear apriori. Chapter 10 therefore investigated the use of the BIB-AE for shower simulation at different positions in the ILD ECAL. The effect of projecting BIB-AE generated showers into two different positions were explored. The first position had a large insensitive region running through layers near the shower core. The second position had no inter-layer large insensitive volume near the shower core, but the geometry was still irregular, with a staggering effect between layers. Inspired by broader developments that have occurred during the course of this thesis [6, 7], two approaches to projecting showers were explored. In the first case, hits in the shower were deposited at the center of the readout cells. In the second, each hit in the shower was splitting into a 6×6 higher granularity grid of sub-cells. While in the first position some effects on the calorimetric observables produced by the model showed improvement by splitting hits into higher granularity sub-cells, the model performance in the second position showed was negatively impacted. As the second position is encountered more frequently in the calorimeter, the projection of shower hits using the center of the readout cell was concluded to be preferable for the regular grid BIB-AE model used.

Chapter 11 benchmarked the performance of the BIB-AE applied to the task of simulating the photons from π^0s produced in the process $e^+e^- \to \tau^+\tau^-$. A comparison to two different GEANT4 versions provided a means to quantify the uncertainty in our understanding of the development of electromagnetic showers. This provided a well defined goal for the physics performance of the generative model. While the performance of the BIB-AE was within the deviations between Geant4 versions for several key physics observables, other observables showed some more significant discrepancies, notably in the intrinsic polar angle and invariant mass distribution of the di-photon system that was reconstructed. Due to the steep rise to low energies in the spectrum of the photons produced in this process, a speed-up of only a factor of $\times 1.2$ relative to pure full simulation was obtained. This constitutes the first benchmark of a generative model on a full physics process in a high granularity calorimeter with a realistic geometry.

Outlook

In this thesis, only the application of generative models for fast simulation of electromagnetic showers has been investigated. However, showers produced by hadrons exhibit significantly more complex topologies and larger event-to-event fluctuations, as described in Chapter 3. This makes modelling hadronic showers much more challenging, in particular for high granularity calorimeters, where it is possible to resolve their topological structure to a high degree. Being able to perform high fidelity fast simulation of hadronic showers across the calorimeter system would enable the simulation time to be reduced more significantly across a greater variety of physics processes. As described in Chapter 6, thus far the only study that attempted to simulate hadronic showers in a realistic high granularity calorimeter with a generative model was performed in [4]. This work investigated the performance of both a WGAN and a BIB-AE trained to simulate charged pion showers in the ILD AHCAL (see Chapter 4). However, a major limitation of this study was that it was only performed in the AHCAL. While this allows initial investigations into whether a model can learn the more complex structure of a hadronic shower, in a real detector showers produced by hadrons can start in the ECAL and extend into the HCAL (see Section 3). Since the geometry of these two detectors is typically significantly different, for models relying on the use of a regular grid, such as the BIB-AE studied in this thesis, this poses a significant challenge. For example, the granularity of the ECAL is typically significantly higher than that of the HCAL (for ILD, the side length of cells in the SiW-ECAL is almost an order of magnitude smaller than in the AHCAL). Another challenge is that the geometrical irregularities are typically significantly different between these calorimeter subsystems.

While geometry irregularities are particularly challenging when they differ between detector subsystems, they are still a factor that has to be considered even within a single calorimeter subsystem. As was demonstrated in the investigation performed in Chapter 10, for a regular grid based model trained using information at the level of the readout geometry of the ILD ECAL, the physics performance of the model degrades at positions in the calorimeter with additional insensitive volumes compared to the geometry used for training. This is partly due to entire hits being lost in these volumes.

One promising approach to tackling the irregular nature of a geometry is to use a model operating on a point cloud of energy depositions at a lower level than that of the physical readout geometry (i.e. clustered Geant4 steps). This approach was proposed in [6], and improved upon in [7]. While this inspired the splitting of hits explored in Chapter 10, a key difference is that by directly learning information at the level of clustered Geant4 steps, the information learnt by the model is decoupled from the readout geometry. This presents the model with the possibility to learn the distribution of energy that would usually be *within* a readout cell, and shows promise as a means for a model to achieve a high degree of geometry independence. This would allow showers to be more readily translated within the planar regions of the ILD calorimeter. A point cloud is also a more flexible form of data structure, meaning that such a model is well suited to be able to simulate showers in calorimeters with more irregular geometries, such as the Calorimeter Endcap Upgrade for the CMS experiment for the upcoming High Luminosity phase of the LHC [109]. A final advantage of a point cloud model is that it reduces the wasted computation on cells which are not hit, as these necessarily have to be simulated by a regular grid based model. This makes it a more computationally efficient solution, therefore providing the potential to achieve more pronounced speed-ups relative to GEANT4. A key challenge in applying generative models based on point clouds to fast calorimeter simulation which has yet to be addressed is the incorporation of an efficient means of local inter-point communication. This locality is naturally incorporated in a regular grid model based on convolutional operations. While this is not necessarily required for the simulation of electromagnetic showers, and indeed was not included in the studies performed in [6, 7], it will be needed to learn the detailed structure present in hadronic showers.

While a point cloud model operating on clustered GEANT4 steps provides a promising avenue to handle irregular detector geometries, it is very likely that this will not be enough to handle some of the irregularities studied in Chapter 10. In particular, positions in which the shower core lies close to a gap which extends through not just the sensitive layer, but also through the absorber layer, present a very different geometrical structure to the regular planar regions of the calorimeter on which a model would usually be trained. This could be handled by adding additional exclusion regions to the fast simulation trigger implemented in *DDFastShowerML* as described in Chapter 9. An additional dedicated model training could then be used to allow the use of fast simulation in this region of different geometrical structure.

Future work focusing on extending the functionality of *DDFastShowerML* would allow faster simulation speeds with generative models. In particular, adding support for simulation with GPUs would likely maximise the potential speed-up relative to GEANT4. This would bring further challenges, and may potentially require tailored solutions depending on the hardware used. Additionally, this would require the batching of showers to gain the most from the parallel nature of the computation. While in GEANT4 this is in principle possible within an event, inter-event batching is not and would require a major overhaul of the simulation procedure. In these regards, the addition of GPU support to the library has potential synergies with recently started efforts to create a particle transport Monte Carlo code base that can utilise GPU-based hardware acceleration [203–206].

Further extensions of *DDFastShowerML* could involve the incorporation of different detector concepts with a geometry description implemented in DD4HEP, such as those foreseen for FCC-ee [22], together with an appropriately trained model. A key future direction for the development of *DDFastShowerML* would therefore involve the integration of different generative models, trained with a conditioning scheme comparable to the one explored in this thesis. This would allow these models to be run in the full software chain used by ILD, and enable a comparison between models on benchmark physics processes.

This work has focused on studying the performance of the BIB-AE for simulating photon showers resulting from π^0 decays in the process $e^+e^- \to \tau^+\tau^-$. As this process is a particularly stringent test of the performance of a fast simulation tool for electromagnetic showers, future work could compare the performance across different generative models applied to this task. Additionally, the performance of a model used for shower simulation in different physics processes could be studied. In particular, other processes which more often result in electromagnetic showers initiated by particles with a higher energy would result in larger reductions in per-event simulation times relative to GEANT4, assuming that the particles hit a detector subsystem where the fast simulation model could be applied. For ILD, this may require training a model to simulate detector systems closer to the beam pipe, such as the LumiCal (see Chapter 4).

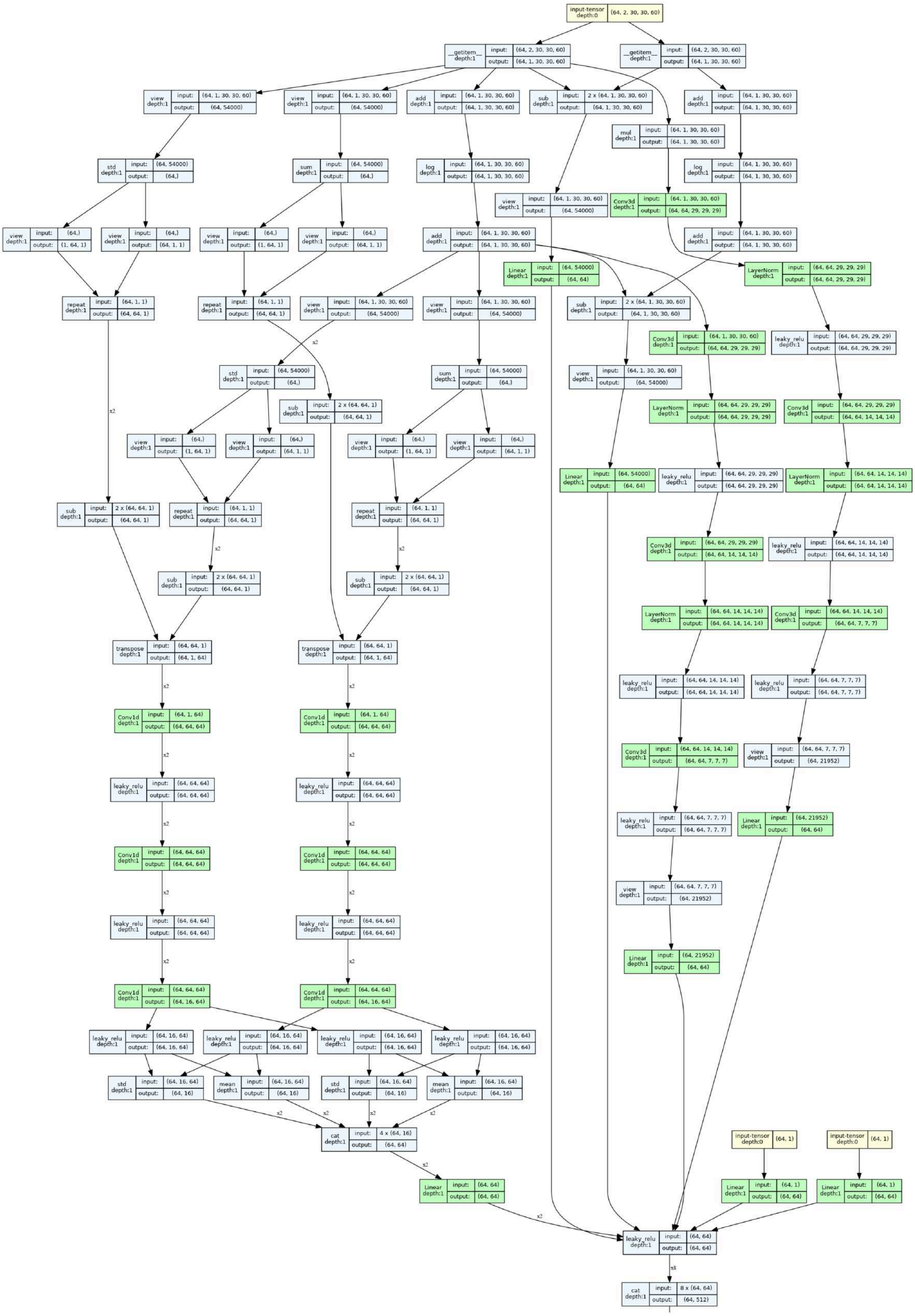
Appendix A

Supplementary Material: Multi-parameter Conditioning of Generative Models

This appendix contains additional details on the work described in Chapter 7 titled Multi-parameter Conditioning of Generative Models.

A.1 Supplementary Material: Neural Network Architectures

The BIB-AE model consists of five different networks. The architecture for the critic network is shown in Figure A.1, while the latent critic network is shown in Figure A.2. Figure A.3 shows the architecture of the encoder network, while Figure A.4 shows the architecture of the decoder network. Finally, the Post Processor network is shown in Figure A.5. All architecture diagrams were automatically generated using TorchView [207].


input-tensor depth:0 (64, 2, 30, 30, 60)
__getitem__ depth:1 input: (64, 2, 30, 30, 60) output: (64, 1, 30, 30, 60)
__getitem__ depth:1 input: (64, 2, 30, 30, 60) output: (64, 1, 30, 30, 60)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 54000)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 54000)
add depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
sub depth:1 input: 2 x (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
add depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
mul depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
std depth:1 input: (64, 54000) output: (64,)
sum depth:1 input: (64, 54000) output: (64,)
log depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
log depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 54000)
Conv3d depth:1 input: (64, 1, 30, 30, 60) output: (64, 64, 29, 29, 29)
view depth:1 input: (64,) output: (1, 64, 1)
view depth:1 input: (64,) output: (64, 1, 1)
view depth:1 input: (64,) output: (1, 64, 1)
view depth:1 input: (64,) output: (64, 1, 1)
add depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
add depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
Linear depth:1 input: (64, 54000) output: (64, 64)
LayerNorm depth:1 input: (64, 64, 29, 29, 29) output: (64, 64, 29, 29, 29)
repeat depth:1 input: (64, 1, 1) output: (64, 64, 1)
repeat depth:1 input: (64, 1, 1) output: (64, 64, 1)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 54000)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 54000)
sub depth:1 input: 2 x (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
Conv3d depth:1 input: (64, 1, 30, 30, 60) output: (64, 64, 29, 29, 29)
leaky_relu depth:1 input: (64, 64, 29, 29, 29) output: (64, 64, 29, 29, 29)
x2
std depth:1 input: (64, 54000) output: (64,)
sum depth:1 input: (64, 54000) output: (64,)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 54000)
sub depth:1 input: 2 x (64, 64, 1) output: (64, 64, 1)
LayerNorm depth:1 input: (64, 64, 29, 29, 29) output: (64, 64, 29, 29, 29)
Conv3d depth:1 input: (64, 64, 29, 29, 29) output: (64, 64, 14, 14, 14)
x2
view depth:1 input: (64,) output: (1, 64, 1)
view depth:1 input: (64,) output: (64, 1, 1)
view depth:1 input: (64,) output: (1, 64, 1)
view depth:1 input: (64,) output: (64, 1, 1)
Linear depth:1 input: (64, 54000) output: (64, 64)
leaky_relu depth:1 input: (64, 64, 29, 29, 29) output: (64, 64, 29, 29, 29)
LayerNorm depth:1 input: (64, 64, 14, 14, 14) output: (64, 64, 14, 14, 14)
sub depth:1 input: 2 x (64, 64, 1) output: (64, 64, 1)
repeat depth:1 input: (64, 1, 1) output: (64, 64, 1)
repeat depth:1 input: (64, 1, 1) output: (64, 64, 1)
Conv3d depth:1 input: (64, 64, 29, 29, 29) output: (64, 64, 14, 14, 14)
leaky_relu depth:1 input: (64, 64, 14, 14, 14) output: (64, 64, 14, 14, 14)
x2
x2
sub depth:1 input: 2 x (64, 64, 1) output: (64, 64, 1)
sub depth:1 input: 2 x (64, 64, 1) output: (64, 64, 1)
LayerNorm depth:1 input: (64, 64, 14, 14, 14) output: (64, 64, 14, 14, 14)
Conv3d depth:1 input: (64, 64, 14, 14, 14) output: (64, 64, 7, 7, 7)
transpose depth:1 input: (64, 64, 1) output: (64, 1, 64)
transpose depth:1 input: (64, 64, 1) output: (64, 1, 64)
leaky_relu depth:1 input: (64, 64, 14, 14, 14) output: (64, 64, 14, 14, 14)
leaky_relu depth:1 input: (64, 64, 7, 7, 7) output: (64, 64, 7, 7, 7)
x2
Conv1d depth:1 input: (64, 1, 64) output: (64, 64, 64)
Conv1d depth:1 input: (64, 1, 64) output: (64, 64, 64)
Conv3d depth:1 input: (64, 64, 14, 14, 14) output: (64, 64, 7, 7, 7)
view depth:1 input: (64, 64, 7, 7, 7) output: (64, 21952)
leaky_relu depth:1 input: (64, 64, 64) output: (64, 64, 64)
leaky_relu depth:1 input: (64, 64, 64) output: (64, 64, 64)
leaky_relu depth:1 input: (64, 64, 7, 7, 7) output: (64, 64, 7, 7, 7)
Linear depth:1 input: (64, 21952) output: (64, 64)
Conv1d depth:1 input: (64, 64, 64) output: (64, 64, 64)
Conv1d depth:1 input: (64, 64, 64) output: (64, 64, 64)
view depth:1 input: (64, 64, 7, 7, 7) output: (64, 21952)
leaky_relu depth:1 input: (64, 64, 64) output: (64, 64, 64)
leaky_relu depth:1 input: (64, 64, 64) output: (64, 64, 64)
Linear depth:1 input: (64, 21952) output: (64, 64)
Conv1d depth:1 input: (64, 64, 64) output: (64, 16, 64)
Conv1d depth:1 input: (64, 64, 64) output: (64, 16, 64)
leaky_relu depth:1 input: (64, 16, 64) output: (64, 16, 64)
leaky_relu depth:1 input: (64, 16, 64) output: (64, 16, 64)
leaky_relu depth:1 input: (64, 16, 64) output: (64, 16, 64)
leaky_relu depth:1 input: (64, 16, 64) output: (64, 16, 64)
std depth:1 input: (64, 16, 64) output: (64, 16)
mean depth:1 input: (64, 16, 64) output: (64, 16)
std depth:1 input: (64, 16, 64) output: (64, 16)
mean depth:1 input: (64, 16, 64) output: (64, 16)
x2
cat depth:1 input: 4 x (64, 16) output: (64, 64)
input-tensor depth:0 (64, 1)
input-tensor depth:0 (64, 1)
Linear depth:1 input: (64, 64) output: (64, 64)
Linear depth:1 input: (64, 1) output: (64, 64)
Linear depth:1 input: (64, 1) output: (64, 64)
leaky_relu depth:1 input: (64, 64) output: (64, 64)
x8
cat depth:1 input: 8 x (64, 64) output: (64, 512)

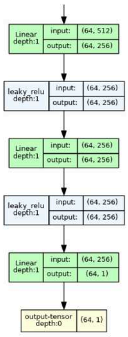

Linear depth:1
input: (64, 512)
output: (64, 256)
leaky_relu depth:1
input: (64, 256)
output: (64, 256)
Linear depth:1
input: (64, 256)
output: (64, 256)
leaky_relu depth:1
input: (64, 256)
output: (64, 256)
Linear depth:1
input: (64, 256)
output: (64, 1)
output-tensor depth:0
(64, 1)

Figure A.1: Critic architecture.

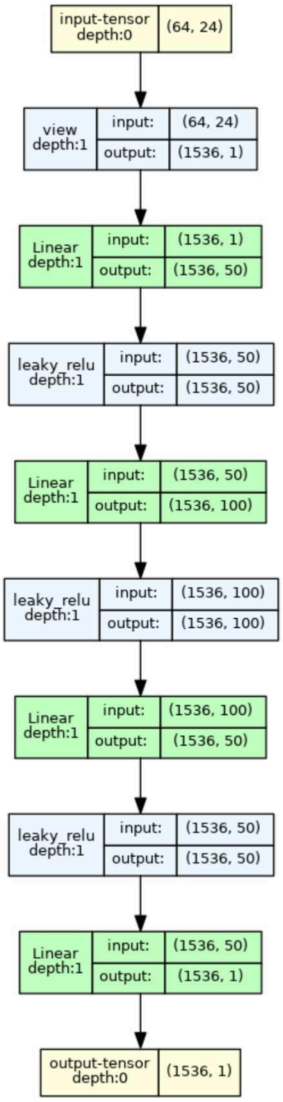

input-tensor depth:0
(64, 24)
view depth:1
input: (64, 24)
output: (1536, 1)
Linear depth:1
input: (1536, 1)
output: (1536, 50)
leaky_relu depth:1
input: (1536, 50)
output: (1536, 50)
Linear depth:1
input: (1536, 50)
output: (1536, 100)
leaky_relu depth:1
input: (1536, 100)
output: (1536, 100)
Linear depth:1
input: (1536, 100)
output: (1536, 50)
leaky_relu depth:1
input: (1536, 50)
output: (1536, 50)
Linear depth:1
input: (1536, 50)
output: (1536, 1)
output-tensor depth:0
(1536, 1)

Figure A.2: Latent critic architecture.

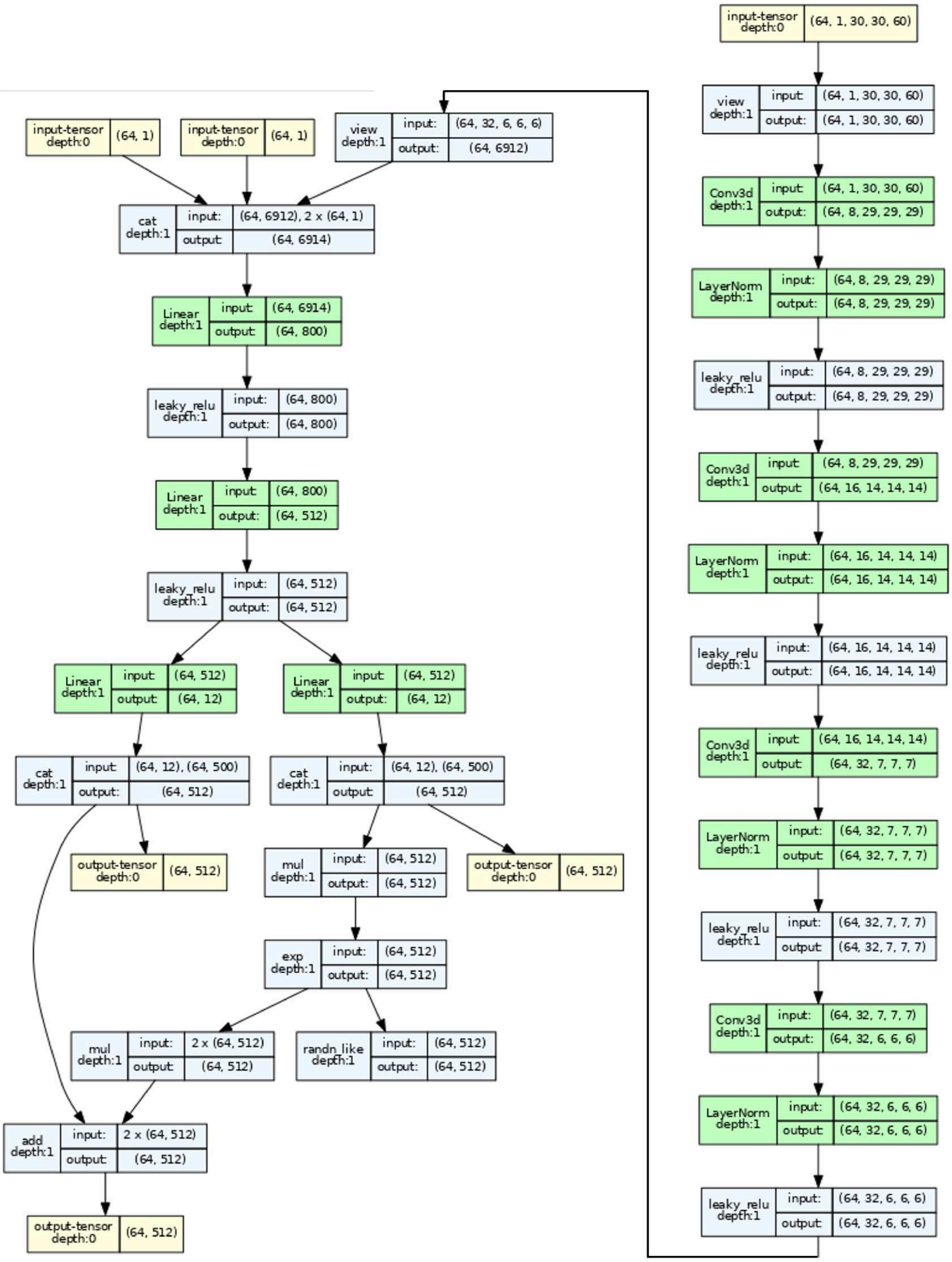

input-tensor depth:0 (64, 1, 30, 30, 60)
view depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
Conv3d depth:1 input: (64, 1, 30, 30, 60) output: (64, 8, 29, 29, 29)
LayerNorm depth:1 input: (64, 8, 29, 29, 29) output: (64, 8, 29, 29, 29)
leaky_relu depth:1 input: (64, 8, 29, 29, 29) output: (64, 8, 29, 29, 29)
Conv3d depth:1 input: (64, 8, 29, 29, 29) output: (64, 16, 14, 14, 14)
LayerNorm depth:1 input: (64, 16, 14, 14, 14) output: (64, 16, 14, 14, 14)
leaky_relu depth:1 input: (64, 16, 14, 14, 14) output: (64, 16, 14, 14, 14)
Conv3d depth:1 input: (64, 16, 14, 14, 14) output: (64, 32, 7, 7, 7)
LayerNorm depth:1 input: (64, 32, 7, 7, 7) output: (64, 32, 7, 7, 7)
leaky_relu depth:1 input: (64, 32, 7, 7, 7) output: (64, 32, 7, 7, 7)
Conv3d depth:1 input: (64, 32, 7, 7, 7) output: (64, 32, 6, 6, 6)
LayerNorm depth:1 input: (64, 32, 6, 6, 6) output: (64, 32, 6, 6, 6)
leaky_relu depth:1 input: (64, 32, 6, 6, 6) output: (64, 32, 6, 6, 6)
input-tensor depth:0 (64, 1)
input-tensor depth:0 (64, 1)
view depth:1 input: (64, 32, 6, 6, 6) output: (64, 6912)
cat depth:1 input: (64, 6912), 2 x (64, 1) output: (64, 6914)
Linear depth:1 input: (64, 6914) output: (64, 800)
leaky_relu depth:1 input: (64, 800) output: (64, 800)
Linear depth:1 input: (64, 800) output: (64, 512)
leaky_relu depth:1 input: (64, 512) output: (64, 512)
Linear depth:1 input: (64, 512) output: (64, 12)
Linear depth:1 input: (64, 512) output: (64, 12)
cat depth:1 input: (64, 12), (64, 500) output: (64, 512)
cat depth:1 input: (64, 12), (64, 500) output: (64, 512)
output-tensor depth:0 (64, 512)
mul depth:1 input: (64, 512) output: (64, 512)
output-tensor depth:0 (64, 512)
exp depth:1 input: (64, 512) output: (64, 512)
mul depth:1 input: 2 x (64, 512) output: (64, 512)
randn_like depth:1 input: (64, 512) output: (64, 512)
add depth:1 input: 2 x (64, 512) output: (64, 512)
output-tensor depth:0 (64, 512)

Figure A.3: BIB-AE encoder architecture.

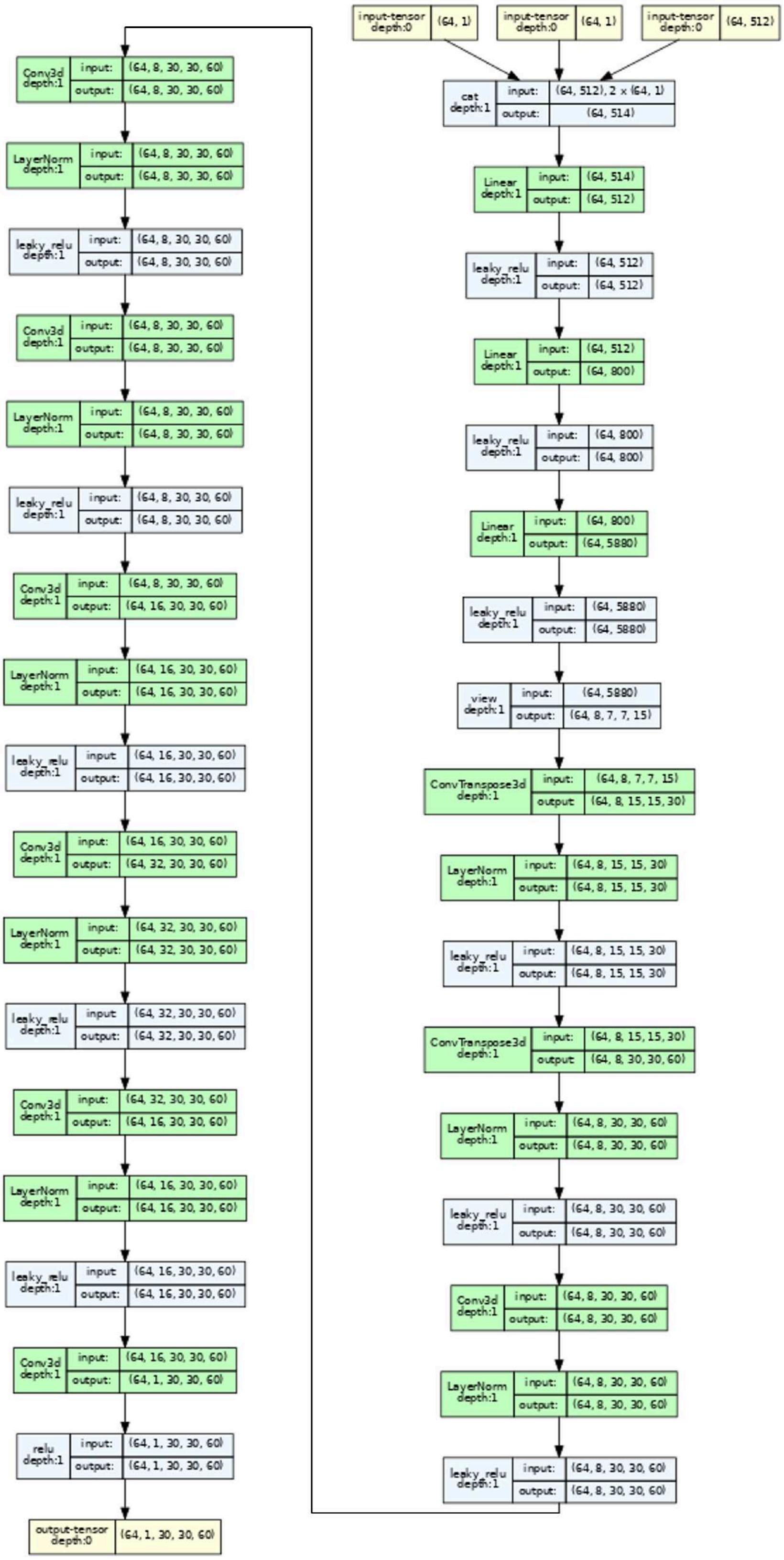

input-tensor depth:0 (64, 1)
input-tensor depth:0 (64, 1)
input-tensor depth:0 (64, 512)
cat depth:1 input: (64, 512), 2 x (64, 1) output: (64, 514)
Linear depth:1 input: (64, 514) output: (64, 512)
leaky_relu depth:1 input: (64, 512) output: (64, 512)
Linear depth:1 input: (64, 512) output: (64, 800)
leaky_relu depth:1 input: (64, 800) output: (64, 800)
Linear depth:1 input: (64, 800) output: (64, 5880)
leaky_relu depth:1 input: (64, 5880) output: (64, 5880)
view depth:1 input: (64, 5880) output: (64, 8, 7, 7, 15)
ConvTranspose3d depth:1 input: (64, 8, 7, 7, 15) output (64, 8, 15, 15, 30)
LayerNorm depth:1 input: (64, 8, 15, 15, 30) output: (64, 8, 15, 15, 30)
leaky_relu depth:1 input: (64, 8, 15, 15, 30) output: (64, 8, 15, 15, 30)
ConvTranspose3d depth:1 input: (64, 8, 15, 15, 30) output (64, 8, 30, 30, 60)
LayerNorm depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
leaky_relu depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
Conv3d depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
LayerNorm depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
leaky_relu depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
Conv3d depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
LayerNorm depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
leaky_relu depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
Conv3d depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
LayerNorm depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
leaky_relu depth:1 input: (64, 8, 30, 30, 60) output: (64, 8, 30, 30, 60)
Conv3d depth:1 input: (64, 8, 30, 30, 60) output: (64, 16, 30, 30, 60)
LayerNorm depth:1 input: (64, 16, 30, 30, 60) output: (64, 16, 30, 30, 60)
leaky_relu depth:1 input (64, 16, 30, 30, 60) output: (64, 16, 30, 30, 60)
Conv3d depth:1 input: (64, 16, 30, 30, 60) output: (64, 32, 30, 30, 60)
LayerNorm depth:1 input: (64, 32, 30, 30, 60) output: (64, 32, 30, 30, 60)
leaky_relu depth:1 input (64, 32, 30, 30, 60) output: (64, 32, 30, 30, 60)
Conv3d depth:1 input: (64, 32, 30, 30, 60) output: (64, 16, 30, 30, 60)
LayerNorm depth:1 input: (64, 16, 30, 30, 60) output: (64, 16, 30, 30, 60)
leaky_relu depth:1 input (64, 16, 30, 30, 60) output: (64, 16, 30, 30, 60)
Conv3d depth:1 input: (64, 16, 30, 30, 60) output: (64, 1, 30, 30, 60)
relu depth:1 input: (64, 1, 30, 30, 60) output: (64, 1, 30, 30, 60)
output-tensor depth:0 (64, 1, 30, 30, 60)

Figure A.4: BIB-AE decoder architecture.

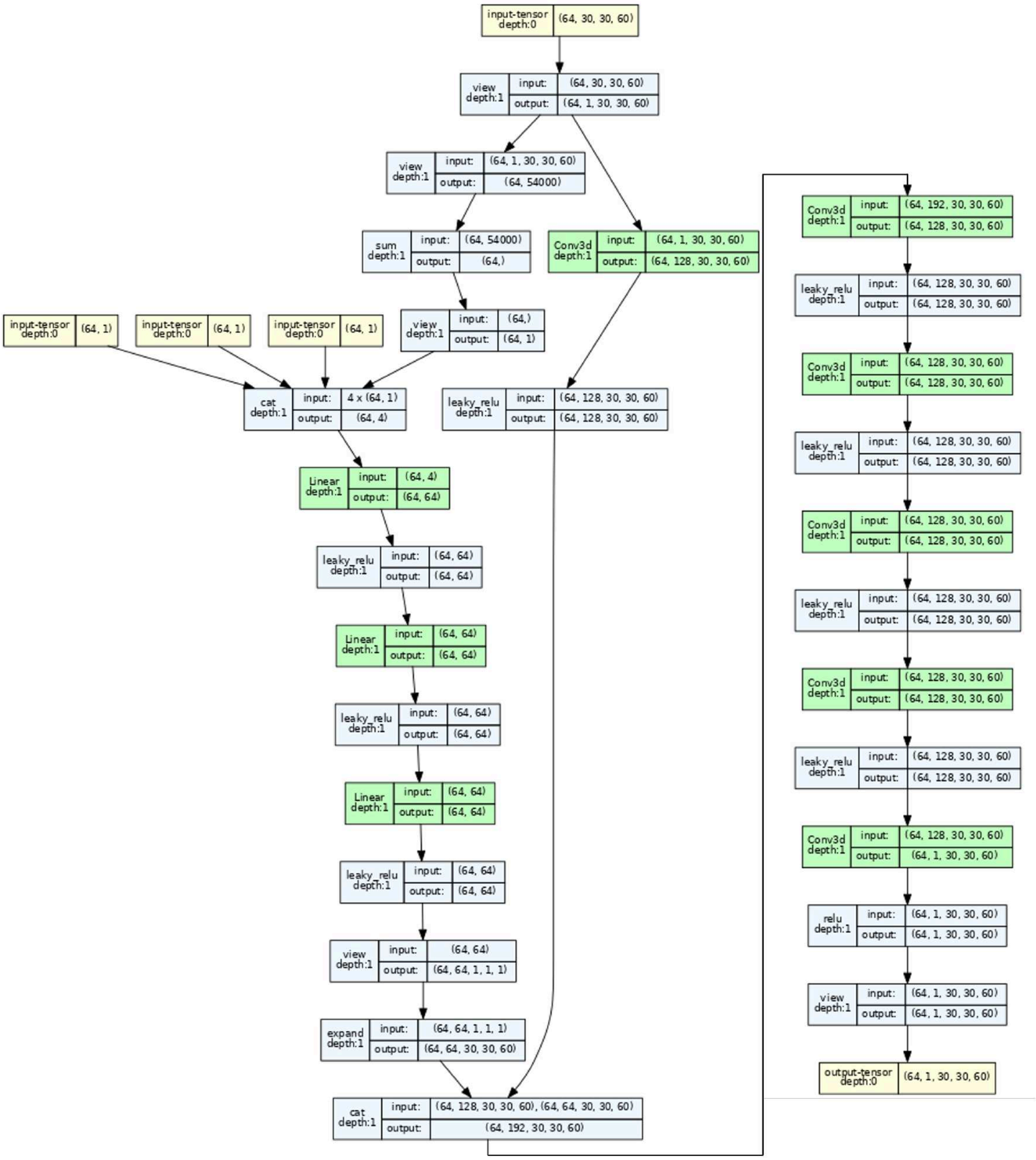

input-tensor depth:0
(64, 30, 30, 60)
view depth:1
input: (64, 30, 30, 60)
output: (64, 1, 30, 30, 60)
view depth:1
input: (64, 1, 30, 30, 60)
output: (64, 54000)
sum depth:1
input: (64, 54000)
output: (64,)
Conv3d depth:1
input: (64, 1, 30, 30, 60)
output: (64, 128, 30, 30, 60)
input-tensor depth:0
(64, 1)
input-tensor depth:0
(64, 1)
input-tensor depth:0
(64, 1)
view depth:1
input: (64,)
output: (64, 1)
cat depth:1
input: 4 x (64, 1)
output: (64, 4)
leaky_relu depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
Linear depth:1
input: (64, 4)
output: (64, 64)
leaky_relu depth:1
input: (64, 64)
output: (64, 64)
Linear depth:1
input: (64, 64)
output: (64, 64)
leaky_relu depth:1
input: (64, 64)
output: (64, 64)
Linear depth:1
input: (64, 64)
output: (64, 64)
leaky_relu depth:1
input: (64, 64)
output: (64, 64)
view depth:1
input: (64, 64)
output: (64, 64, 1, 1, 1)
expand depth:1
input: (64, 64, 1, 1, 1)
output: (64, 64, 30, 30, 60)
cat depth:1
input: (64, 128, 30, 30, 60), (64, 64, 30, 30, 60)
output: (64, 192, 30, 30, 60)
Conv3d depth:1
input: (64, 192, 30, 30, 60)
output: (64, 128, 30, 30, 60)
leaky_relu depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
Conv3d depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
leaky_relu depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
Conv3d depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
leaky_relu depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
Conv3d depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
leaky_relu depth:1
input: (64, 128, 30, 30, 60)
output: (64, 128, 30, 30, 60)
Conv3d depth:1
input: (64, 128, 30, 30, 60)
output: (64, 1, 30, 30, 60)
relu depth:1
input: (64, 1, 30, 30, 60)
output: (64, 1, 30, 30, 60)
view depth:1
input: (64, 1, 30, 30, 60)
output: (64, 1, 30, 30, 60)
output-tensor depth:0
(64, 1, 30, 30, 60)

Figure A.5: BIB-AE Post Processor architecture.

A.2 Supplementary Material: Results

This section lists tables containing the JSD values for each combination of incident angle and energy studied, along with the corresponding selected distributions at simulation and reconstruction level for the cell energy spectrum (Table A.1, Figure A.6, Figure A.7), the center of gravity (Table A.2, Figure A.8, Figure A.9) and the longitudinal (Table A.3, Figure A.10,**??**) and radial (Table A.4) profiles. In each table, the JSD values for the best and worst combinations of energy and angle which were shown in Chapter 7 are highlighted in bold.

A.2.1 Cell Energy Spectrum

Table A.1: Table showing the Jensen Shannon Distance (JSD) between the Geant4 and BIB-AE results for the simulation and reconstruction level cell energy spectrum for each combination of fixed energy and angle. For the simulation level results, the reported JSD is only calculated in the region above the MIP cut. Table and caption reproduced from [1].

		Energy (GeV)					
		SIM			REC		
		20	50	90	20	50	90
Angle (deg.)	40	0.029	0.024	**0.036**	**0.031**	0.022	0.029
	60	0.026	0.021	0.032	0.027	0.018	0.025
	85	0.028	**0.018**	0.029	0.029	**0.015**	0.021

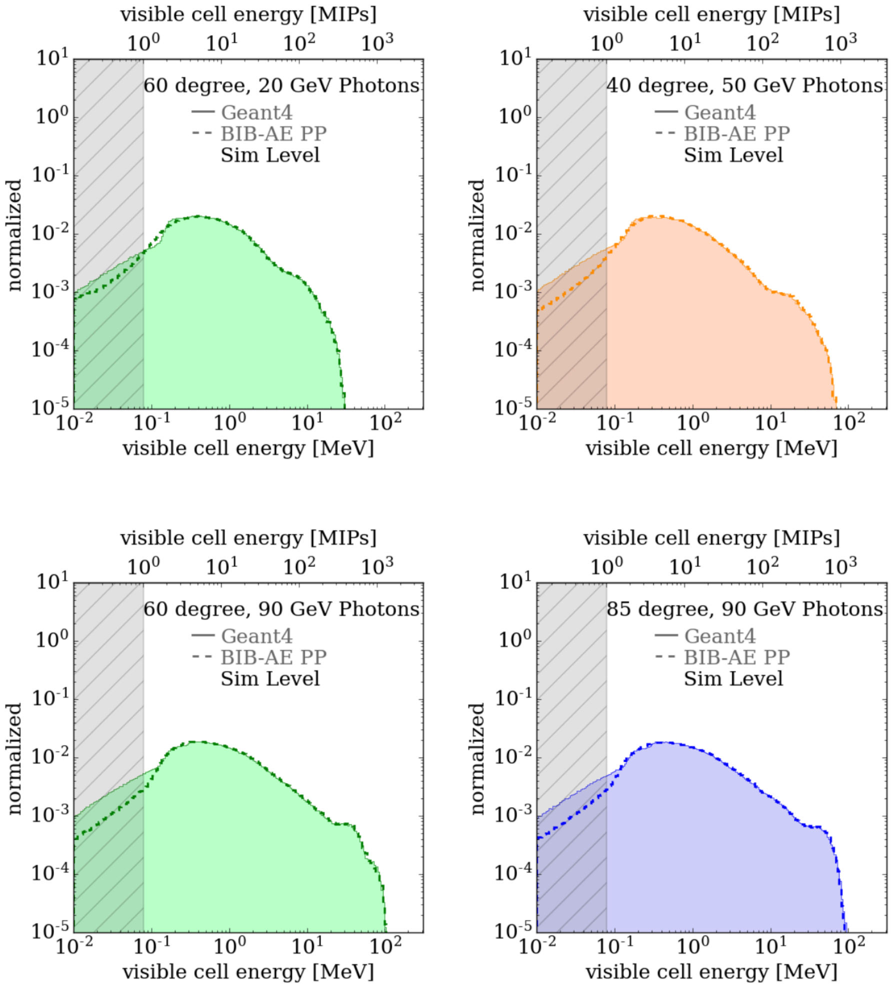


Figure A.6: Selection of additional simulation level cell energy spectra. The grey hatched area indicates the region below half a MIP.

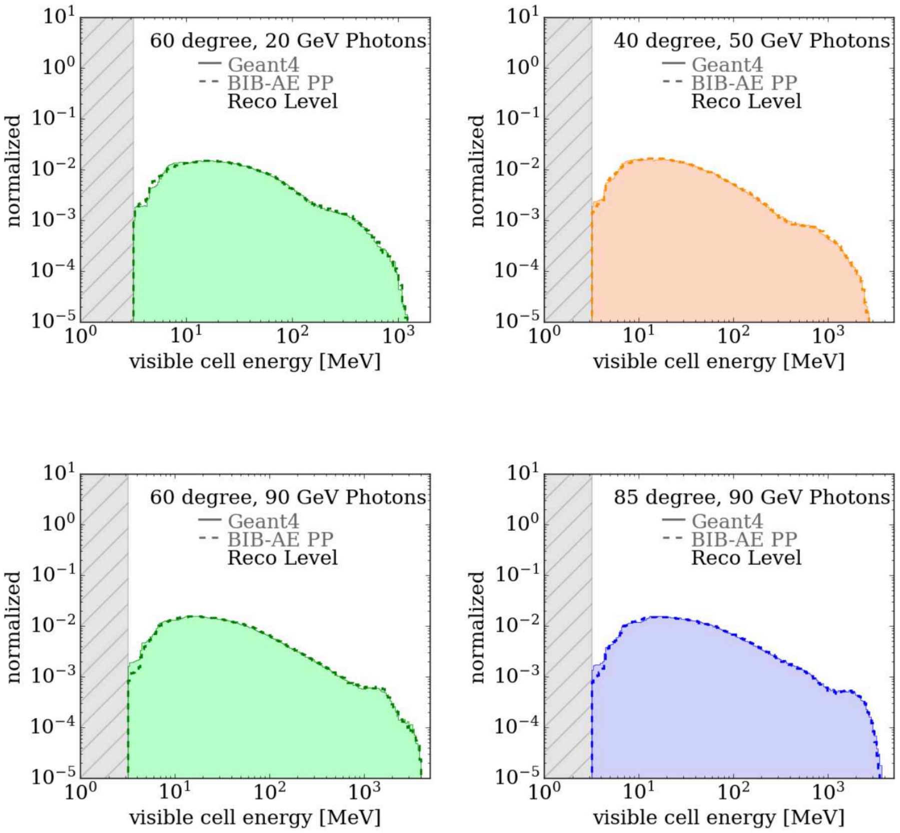


Figure A.7: Selection of additional reconstruction level cell energy spectra. The grey hatched area indicates the region below half a MIP.

A.2.2 Center of Gravity

Table A.2: Table showing the Jensen Shannon Distance (JSD) between the Geant4 and BIB-AE results for the simulation and reconstruction level center of gravity for each combination of fixed energy and angle. Table and caption reproduced from [1].

		Energy (GeV)					
		SIM			REC		
		20	50	90	20	50	90
Angle (deg.)	40	0.066	0.062	0.072	0.081	**0.057**	0.064
	60	0.062	0.059	0.058	0.067	0.064	0.070
	85	**0.076**	**0.057**	0.070	0.079	0.067	**0.091**

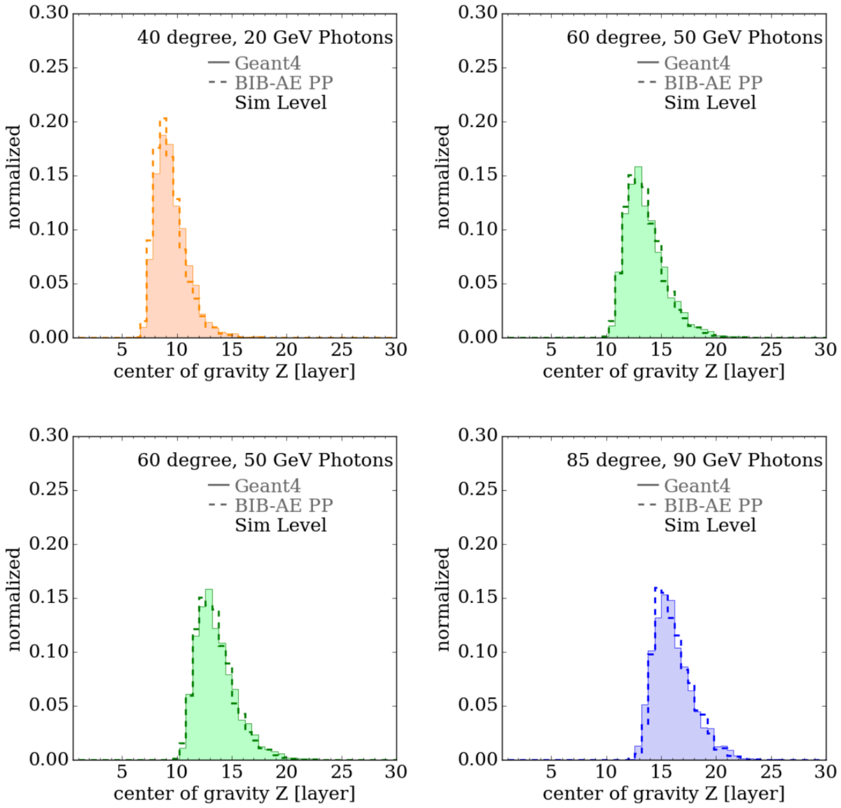


Figure A.8: Selection of additional simulation level center of gravity distributions.

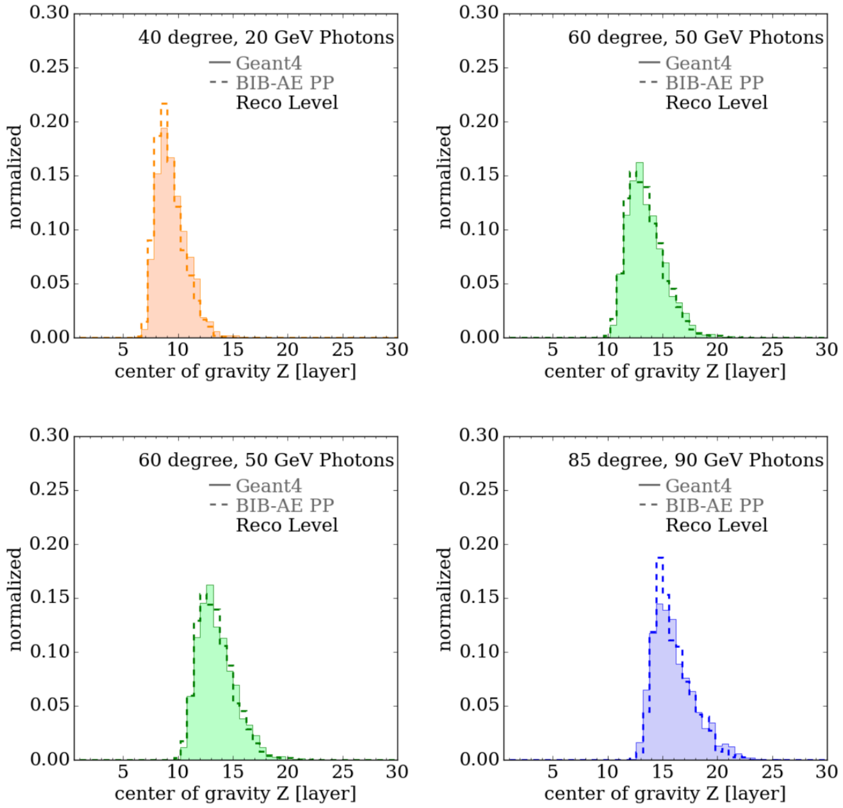


Figure A.9: Selection of additional reconstruction level center of gravity distributions.

A.2.3 Longitudinal Profile

Table A.3: Table showing the Jensen Shannon Distance (JSD) between the Geant4 and BIB-AE results for the simulation and reconstruction level longitudinal profiles for each combination of fixed energy and angle. Table and caption reproduced from [1].

		Energy (GeV)					
		SIM			REC		
		20	50	90	20	50	90
Angle (deg.)	40	**0.025**	0.021	0.023	**0.027**	0.017	0.015
	60	0.021	0.015	0.014	0.020	0.014	**0.009**
	85	0.023	0.014	**0.009**	0.023	0.015	0.011

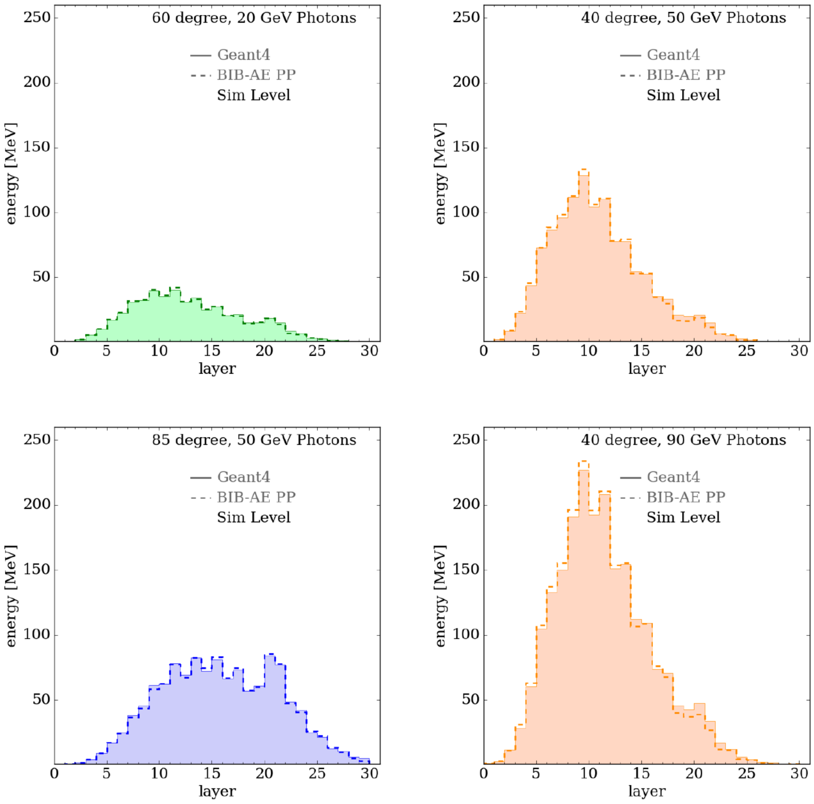


Figure A.10: Selection of additional simulation level longitudinal profiles.

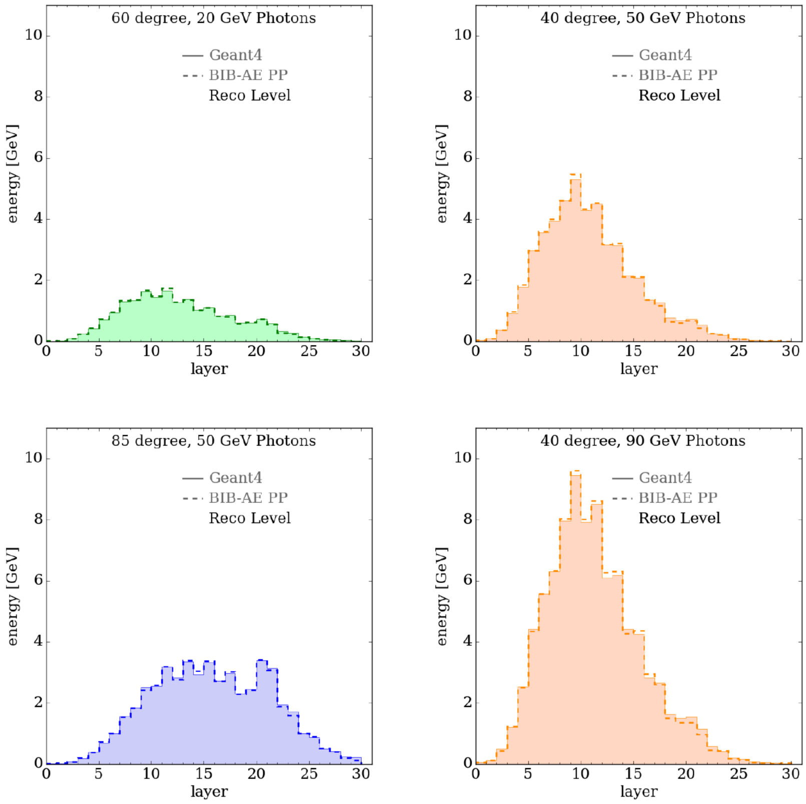


Figure A.11: Selection of additional reconstruction level longitudinal profiles.

A.2.4 Radial Profile

Table A.4: Table showing the Jensen Shannon Distance (JSD) between the Geant4 and BIB-AE results for the simulation and reconstruction level radial profiles for each combination of fixed energy and angle. Table and caption reproduced from [1].

		Energy (GeV)					
		SIM			REC		
		20	50	90	20	50	90
Angle	40	**0.019**	0.014	0.016	**0.018**	0.013	0.013
(deg.)	60	0.011	0.006	**0.006**	0.010	0.007	**0.004**
	85	0.011	0.007	0.007	0.010	0.007	0.008

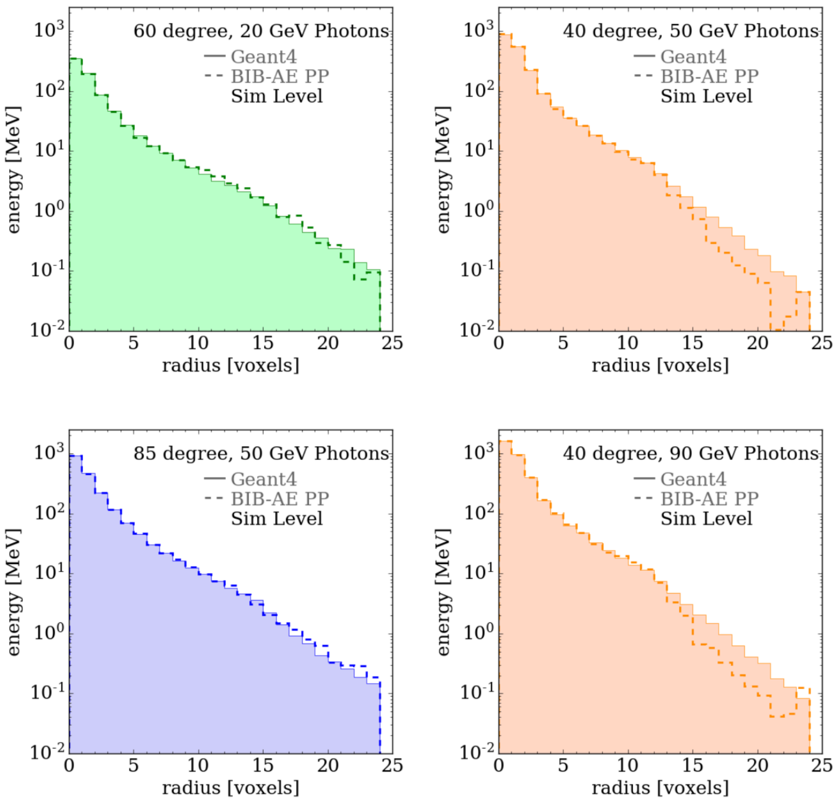


Figure A.12: Selection of additional simulation level radial profiles.

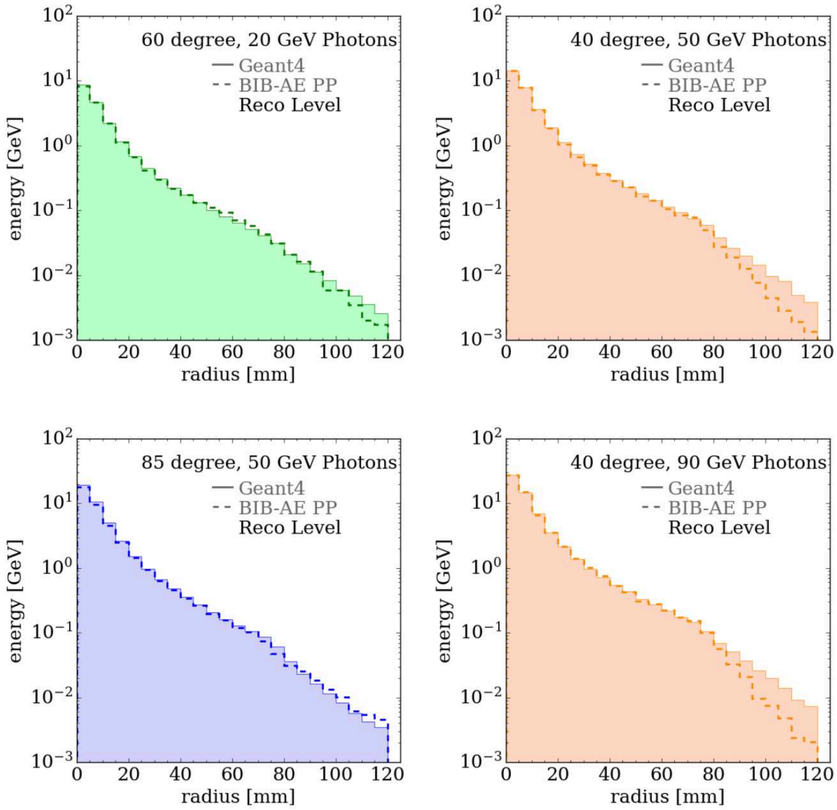


Figure A.13: Selection of additional reconstruction level radial profiles.

Appendix B

Supplementary Material: A Fully Conditioned Generative Model for Electromagnetic Shower Simulation

This appendix contains additional details on the work described in Chapter 8 titled A Fully Conditioned Generative Model for Electromagnetic Shower Simulation. A sample of the results obtained at different test points from those shown in Chapter 8 are included here, as the complete set used in the study consisted of hundreds of such distributions.

B.0.1 Angular Response- Theta

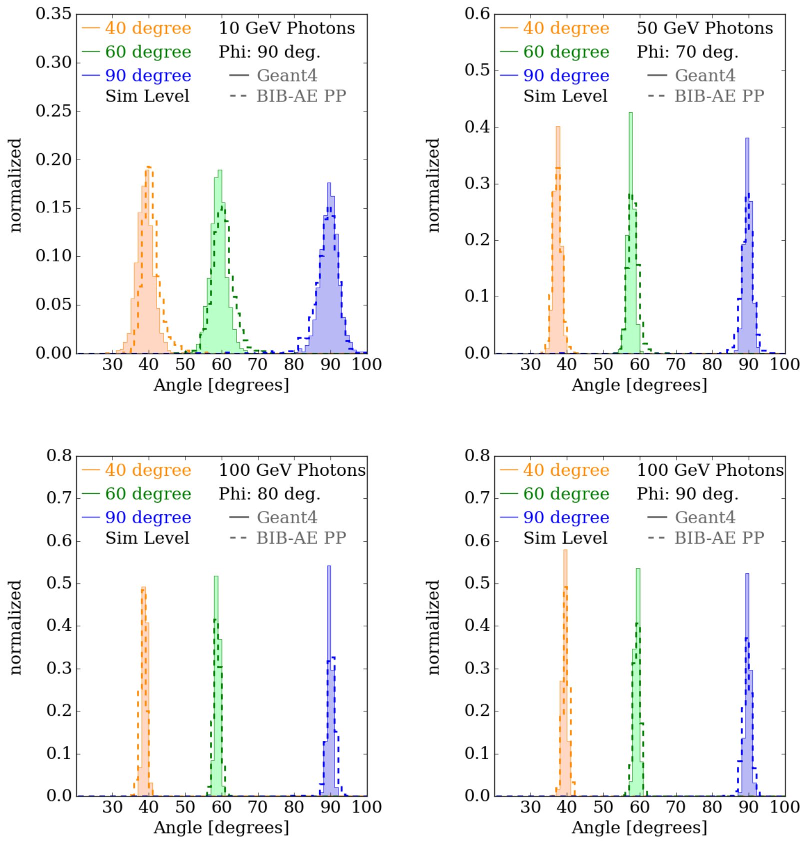


Figure B.1: Selection of additional distributions for the angular response in theta at simulation level.

B.0.2 Angular Response- Phi

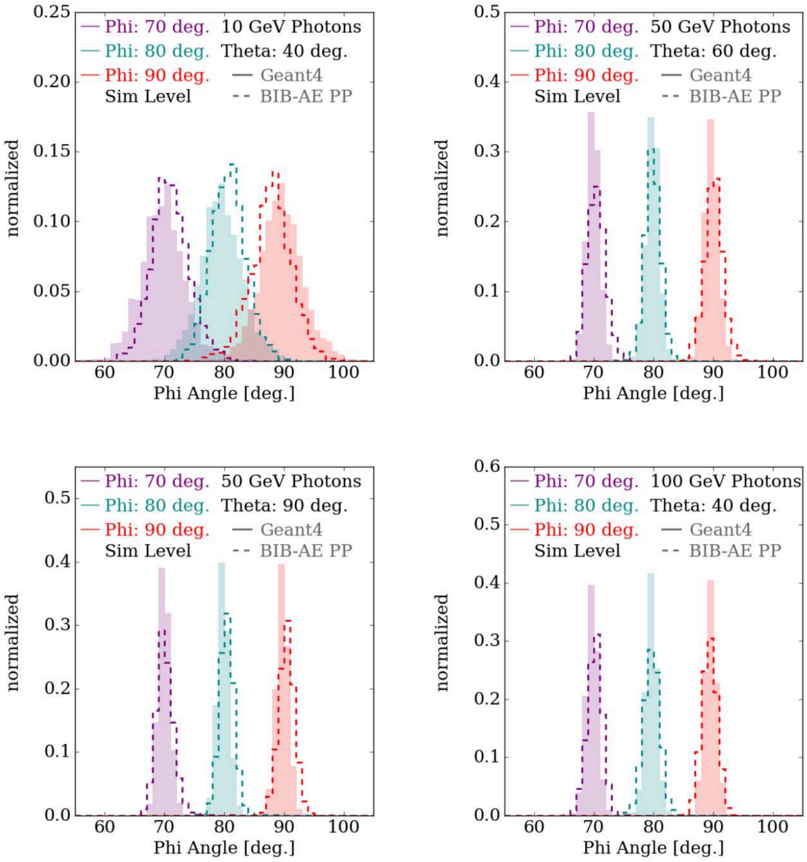


Figure B.2: Selection of additional distributions for the angular response in phi at simulation level.

B.0.3 Energy Response

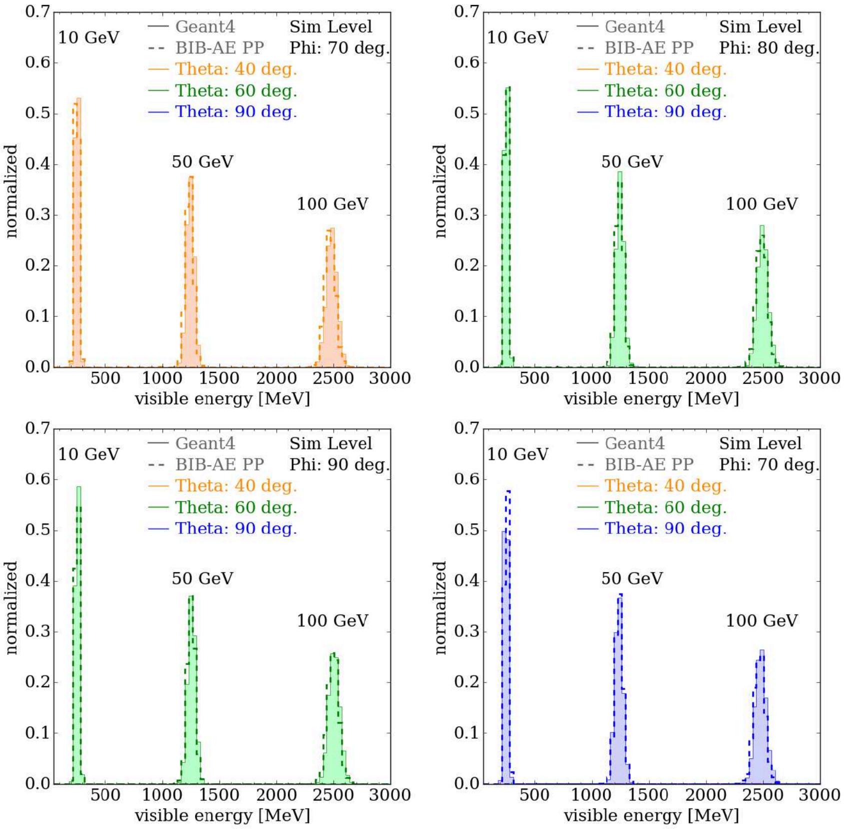


Figure B.3: Selection of additional distributions for the visible energy deposited in the calorimeter at simulation level.

B.0.4 Cell Energy Spectrum

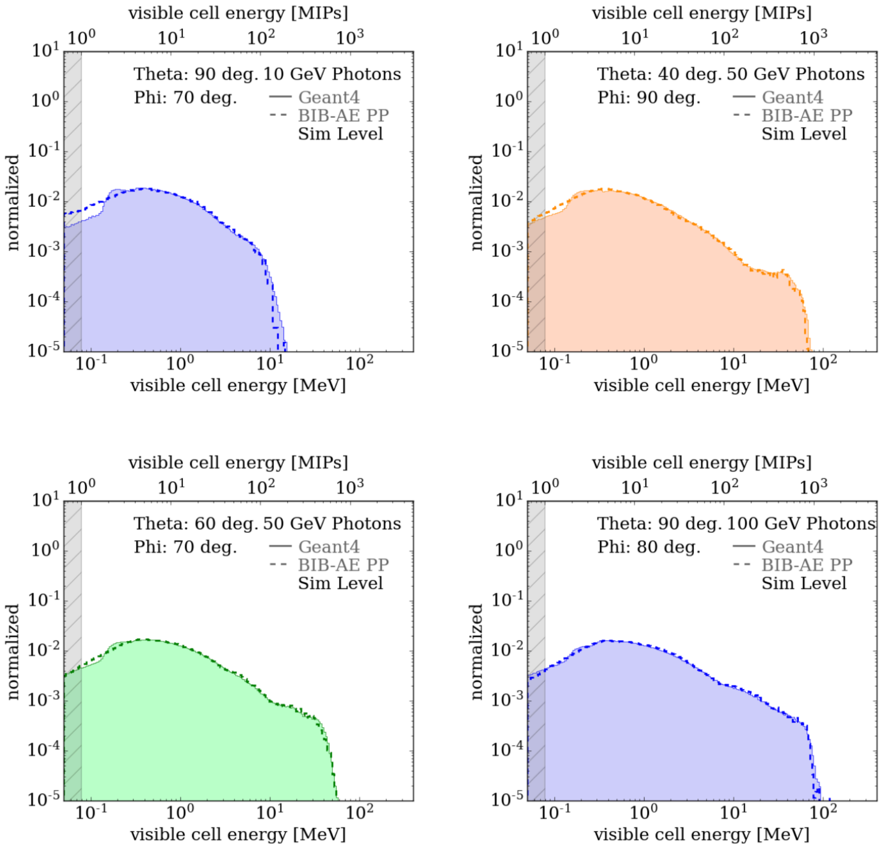


Figure B.4: Selection of additional distributions for the cell energy spectrum at simulation level.

B.0.5 Number of Hits

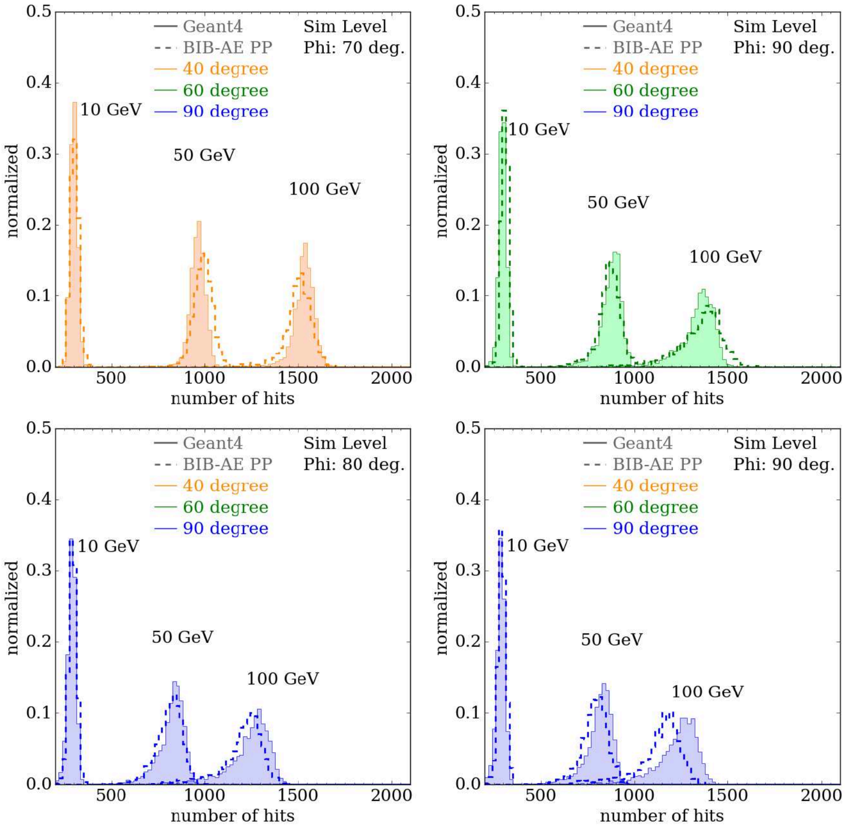


Figure B.5: Selection of additional distributions for the number of hits at simulation level.

B.0.6 Center of Gravity

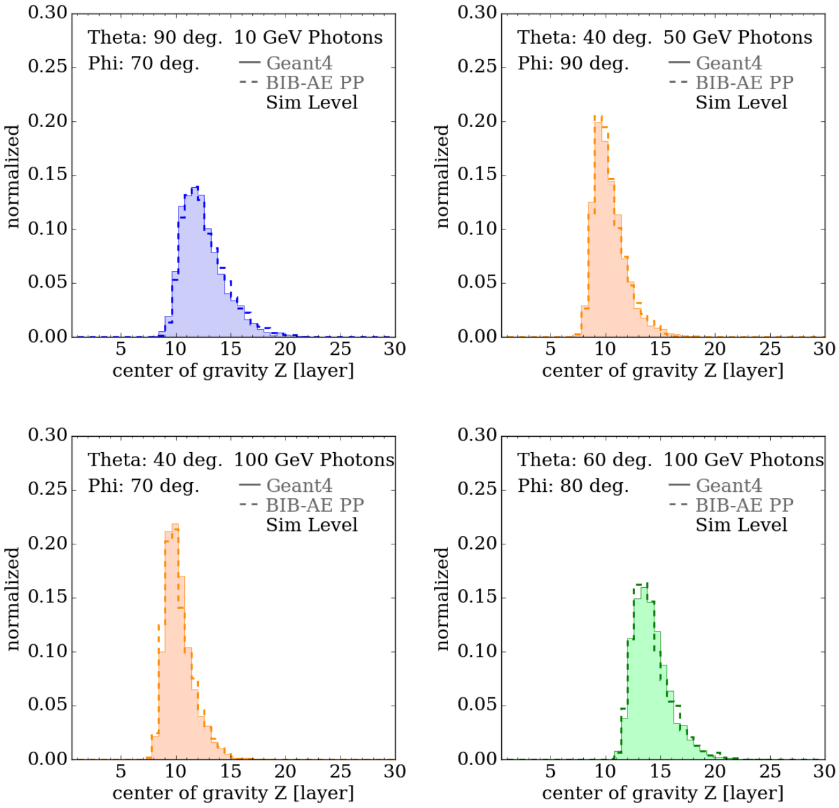


Figure B.6: Selection of additional distributions for the center of gravity in z at simulation level.

B.0.7 Longitudinal Profile

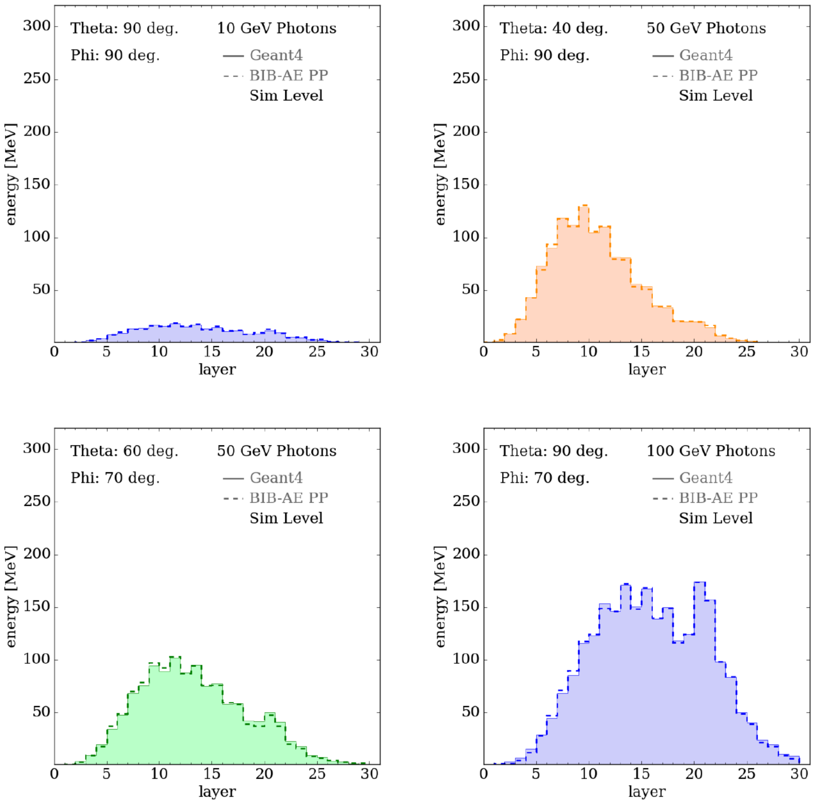


Figure B.7: Selection of additional distributions for the longitudinal profile at simulation level.

Radial Profile

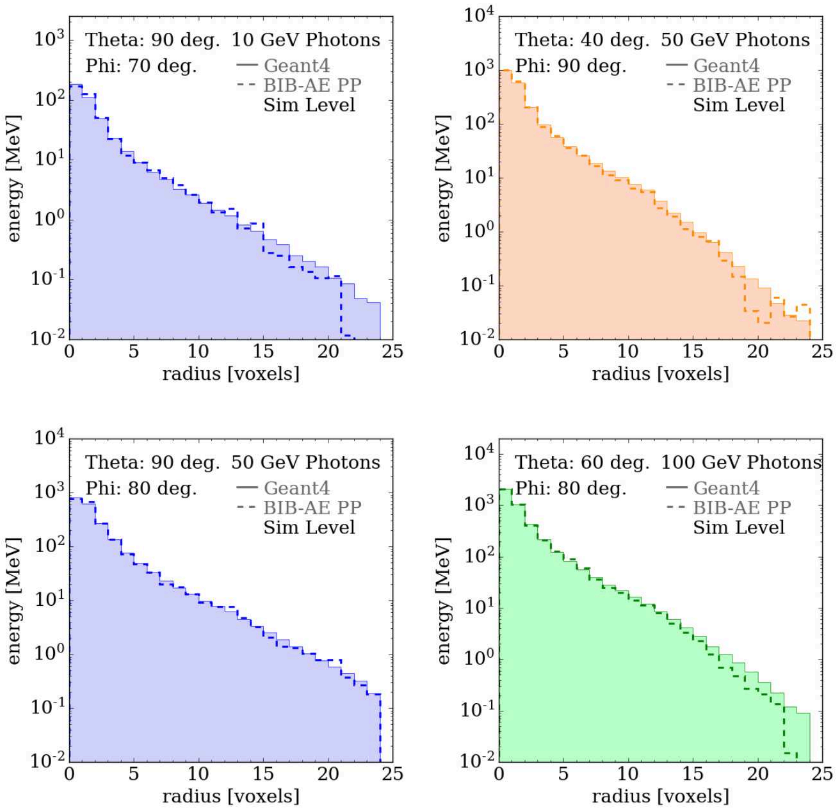


Figure B.8: Selection of additional distributions for the radial profile at simulation level.

Appendix C

Supplementary Material: Irregular Geometries

This appendix contains additional details for Chapter 10 on Irregular Geometries.

As well as the splitting of hits into 6×6 sub-cells, a splitting into 8×8 sub-cells was also investigated. The distributions for visible energy resulting from both splittings is shown in Figure C.1, before and after the MIP threshold is applied. The 6×6 splitting was therefore preferred, as no improvement was observed with the finer splitting.

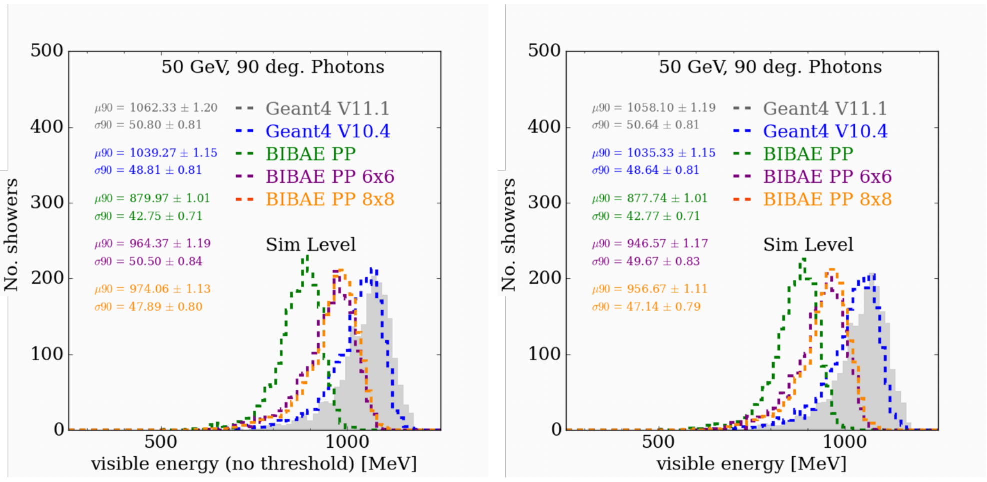


Figure C.1: Distributions for 50 GeV photons incident orthogonal to the face of the calorimeter at *position 1*. Showers were simulated with two different GEANT4 versions: 11.1 (grey, filled distributions, used in the *DDFastShowerML* library) and 10.4 (blue, unfilled distributions, used for the training of the BIB-AE model). Showers were also simulated with the fully conditioned BIB-AE model, and projected back into the irregular ILD ECAL with three different approaches. In the first approach, hits are placed at the center of the cell in the regular grid (green, unfilled distributions), in the second the energy of a hit is split uniformly into a 6×6 grid of high granularity sub-cells (purple, unfilled distributions), while in the third, the energy of a hit is split uniformly into an 8×8 grid of high granularity sub-cells (orange, unfilled distributions). The distributions shown are for the total visible energy deposited in the calorimeter with no MIP threshold applied (top, left) and the total visible energy above threshold deposited in the calorimeter (top, right). The mean (μ_{90}) and standard deviation (σ_{90}) calculated on the central 90% of the distribution are shown.

An additional approach was tested, in which only hits with an energy above 1.5 MIPs were split into 6×6 sub-cells. Figure C.2, shows the effect on the visible energy before and after applying the threshold, the number of hits and the cell energy spectrum. This approach was found to yield no major improvement.

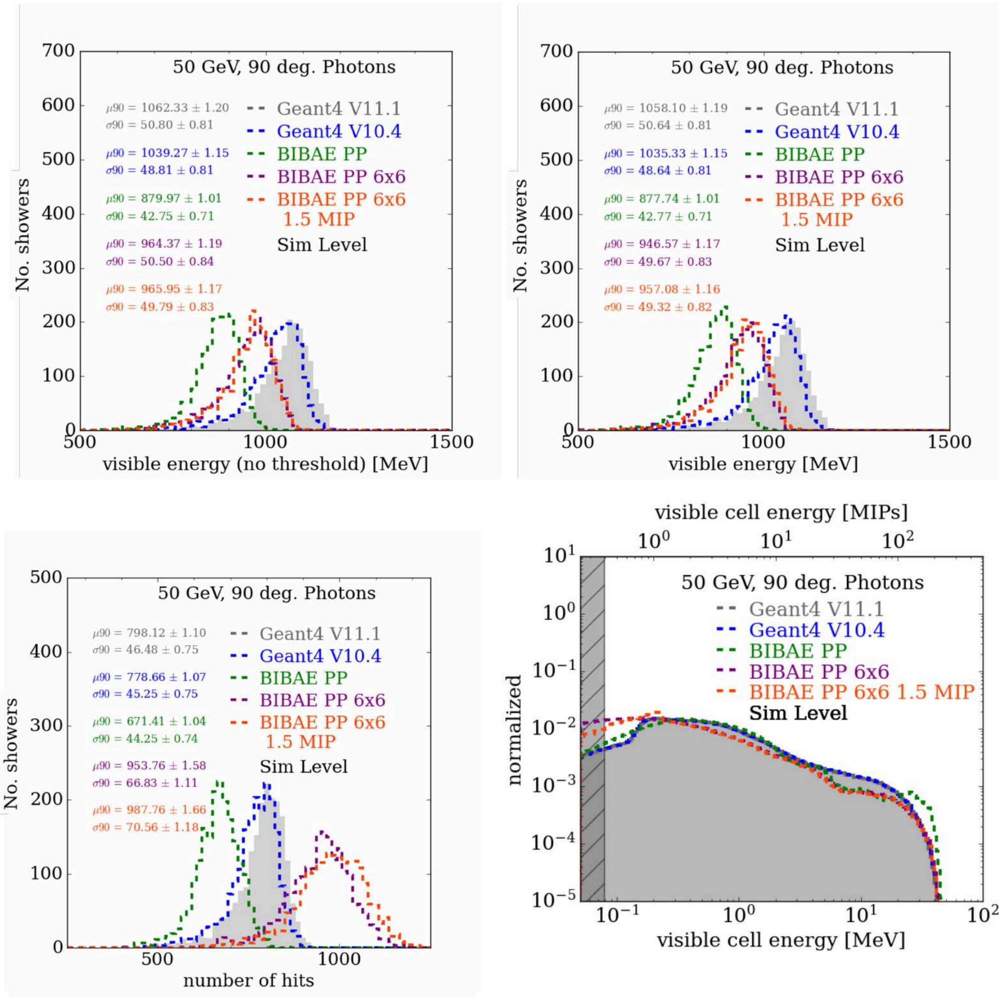


Figure C.2: Distributions for 50 GeV photons incident orthogonal to the face of the calorimeter at *position 1*. Showers were simulated with two different $\textsc{Geant4}$ versions: 11.1 (grey, filled distributions, used in the *DDFastShowerML* library) and 10.4 (blue, unfilled distributions, used for the training of the BIB-AE model). Showers were also simulated with the fully conditioned BIB-AE model, and projected back into the irregular ILD ECAL with three different approaches. In the first approach, hits are placed at the center of the cell in the regular grid (green, unfilled distributions), in the second the energy of a hit is split uniformly into a 6×6 grid of high granularity sub-cells (purple, unfilled distributions), while in the third the energy of a hit is only split into the 6×6 grid of sub-cells if the hit energy is greater than 1.5 MIPs. The distributions shown are for the total visible energy deposited in the calorimeter with no MIP threshold applied (top, left), the total visible energy above threshold deposited in the calorimeter (top, right), the number of hits above threshold (bottom, left) and the cell energy spectrum (bottom, right). For the number of hits and visible energy sums, the mean (μ_{90}) and standard deviation (σ_{90}) calculated on the central 90% of the distribution are shown.

Appendix D

Supplementary Material: Physics Benchmark

This appendix contains additional details on the work described in Chapter 11 on the physics benchmark of the simulation tool developed. Figure D.1 shows the energy spectrum of MC π^0s under various criteria, while Figure D.2 shows the distribution of the global polar angle of their momentum.

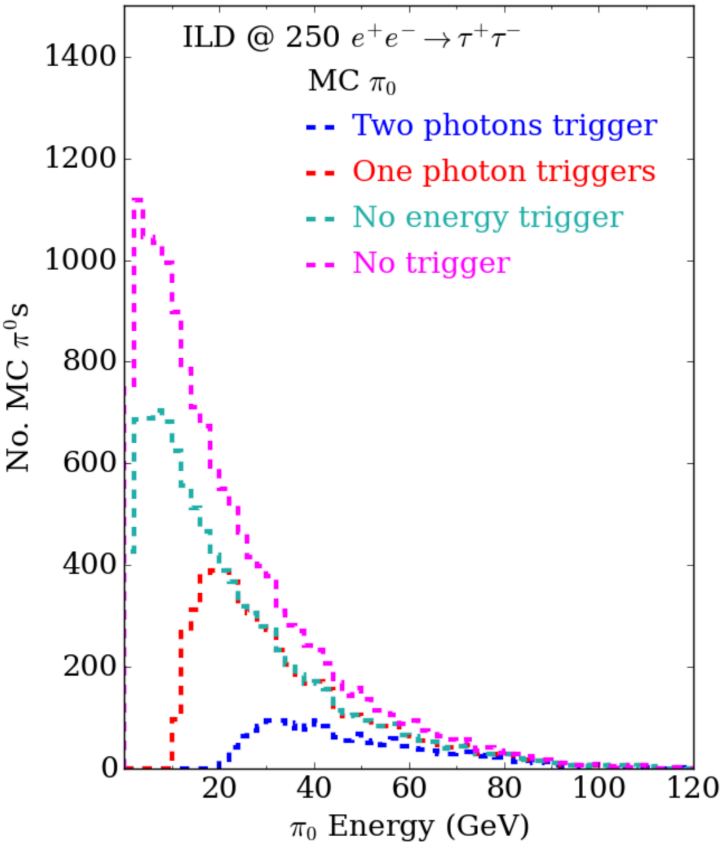


Figure D.1: Energy spectrum for π^0s created in the process $e^+e^- \rightarrow \tau^+\tau^-$, using MC level information from the generator. Distributions are shown for those π^0s which: produce two photons that pass the simulation trigger (blue), produce at least one photon that passes the simulation trigger (red), produce two photons that pass the geometry component of the simulation trigger, with no energy constraint applied (cyan). Finally, the distribution is shown for all π^0s produced, with no constraints from the simulation trigger or otherwise (lilac).

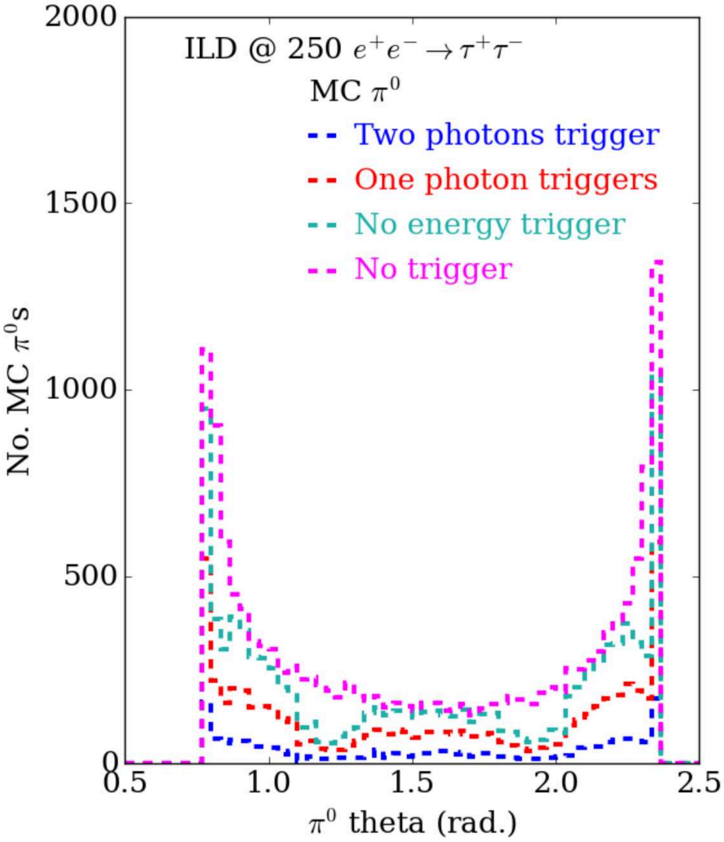


Figure D.2: Number of π^0s as a function of the global polar angle of the π^0's momentum, for the $e^+e^- \to \tau^+\tau^-$ sample studied. MC level information from the generator is used. Distributions are shown for those π^0s which: produce two photons that pass the simulation trigger (blue), produce at least one photon that passes the simulation trigger (red), produce two photons that pass the geometry component of the simulation trigger, with no energy constraint applied (cyan). Finally, the distribution is shown for all π^0s produced, with no constraints from the simulation trigger or otherwise (lilac).

Acknowledgements

The author gratefully acknowledges support by Deutsches Elektronen-Synchrotron DESY, where the work presented in this thesis was carried out. This research was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. Work presented in this thesis has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 101004761, as well as support via the KISS consortium (05D23GU4, 13D22CH5) funded by the German Federal Ministry of Education and Research BMBF in the ErUM-Data action plan.

This thesis represents the culmination of almost three and a half years of work, which would have not been possible without a huge number of people. While I may not be able to thank everyone here, I would like to call out some people/groups of people who I feel have been particularly important.

Firstly, I would like to sincerely thank Dr. Frank Gaede, who has acted as my professional supervisor at DESY during my doctorate and has been a key mentor to me throughout my academic career thus far. Similarly, I would like to thank my University Professor Prof. Dr. Gregor Kasieczka for his strong support and supervision throughout my doctorate. Finally, I would like to thank Dr. Katja Krüger, who acted as a third unofficial supervisor of sorts, providing invaluable guidance. Thank you all for your continuous support, exceptional supervision and amazing enthusiasm.

During this time, I have had the privilege to be a member of the FTX (and previously the FLC) group at DESY, as well as the joint DESY-UHH Generative working group. I am grateful to everyone from these groups who has created an amazing working environment, giving me the opportunity for great personal growth, as well as providing a friendly and sociable space. I have been particularly lucky to have had the support of a number of amazing Postdocs during my work. Thomas has no doubt borne the brunt of my (probably stupid, although he insists there is no such thing) questions, but has always provided patient assistance and useful discussions. Engin was a driving force during the first half of my work, helping me set up a solid base on which to build future work. Finally, Lennart provided a lot of important software support, as well as many interesting discussions. I have also been lucky enough to work with several incredibly gifted PhD students, including Anatolii and Bohdan, with whom I have worked closely discussing work, planning projects for students and developing ideas. A special thanks to Adriana for running a great writing club (which definitely expedited the completion of this thesis) and the lovely cake breaks.

I have also had the honour to be involved in the supervision of a number of exceptional students during my time at DESY. My thanks go to Francisca, Khrystyna and Konrad for enduring my supervision. I would also like to add Michal to this list, although I was not directly part of her supervisory team. I hope you all learnt as much as I did.

My deepest thanks go to Antoine, Daniel, Thomas, Mareike, Sascha, Engin, Lennart, Anatolii, Katja, Konrad, Uli and Jeff, who were kind enough to proof read this document and provide detailed feedback. I really appreciate the time you have given me.

I would like to conclude by thanking my friends and family. The first half of my time in Germany was particularly challenging, as I started mid way through the pandemic, and the support I have received from them has been truly humbling. A special, honourable thank you must go to Jane, without whom I would never have become a physicist, and this work would not exist. Finally my undying love goes to my parents Helen and Jeff, who have always patiently gone above and beyond the call of duty to encourage and support me in every way they can, my siblings Alexander and Catherine for brightening my world, and my dearest wife Verity for her patience, unwavering support, and letting me move away to Germany for nearly three and a half years to undertake this endeavour. This thesis is dedicated to all of you.

Bibliography

[1] S. Diefenbacher et al. "New angles on fast calorimeter shower simulation". *Mach. Learn. Sci. Tech.* 4.3 (2023), p. 035044. DOI: `10.1088/2632-2153/acefa9`. arXiv: `2303.18150 [physics.ins-det]`.

[2] P. McKeown et al. "Fast Simulation of Highly Granular Calorimeters with Generative Models: Towards a First Physics Application". *PoS* EPS-HEP2023 (2023), p. 568. DOI: `10.22323/1.449.0568`.

[3] E. Buhmann et al. "Fast and Accurate Electromagnetic and Hadronic Showers from Generative Models". *EPJ Web Conf.* 251 (2021), p. 03049. DOI: `10.1051/epjconf/202125103049`.

[4] E. Buhmann et al. "Hadrons, better, faster, stronger". *Mach. Learn. Sci. Tech.* 3.2 (2022), p. 025014. DOI: `10.1088/2632-2153/ac7848`. arXiv: `2112.09709`.

[5] S. Bieringer et al. "Generative Models for Fast Simulation of Electromagnetic and Hadronic Showers in Highly Granular Calorimeters". *PoS* ICHEP2022 (2022), p. 236. DOI: `10.22323/1.414.0236`.

[6] E. Buhmann et al. "CaloClouds: fast geometry-independent highly-granular calorimeter simulation". *JINST* 18.11 (2023), P11025. DOI: `10.1088/1748-0221/18/11/P11025`. arXiv: `2305.04847 [physics.ins-det]`.

[7] E. Buhmann et al. "CaloClouds II: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation" (2023). arXiv: `2309.05704 [physics.ins-det]`.

[8] S. L. Glashow. "Partial Symmetries of Weak Interactions". *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: `10.1016/0029-5582(61)90469-2`.

[9] S. Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: `10.1103/PhysRevLett.19.1264`.

[10] A. Salam and J. C. Ward. "Weak and electromagnetic interactions". *Nuovo Cim.* 11 (1959), pp. 568–577. DOI: `10.1007/BF02726525`.

[11] V. C. Rubin and W. K. Ford Jr. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". *Astrophysical Journal,* 159 (Feb. 1970), p. 379. DOI: `10.1086/150317`.

[12] D. Clowe et al. "A direct empirical proof of the existence of dark matter". *Astrophys. J. Lett.* 648 (2006), pp. L109–L113. DOI: `10.1086/508162`. arXiv: `astro-ph/0608407`.

[13] WMAP. "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results". *Astrophys. J. Suppl.* 180 (2009), pp. 225–245. DOI: `10.1088/0067-0049/180/2/225`. arXiv: `0803.0732`.

[14] Planck. "Planck 2015 results. XIII. Cosmological parameters". *Astron. Astrophys.* 594 (2016), A13. DOI: `10.1051/0004-6361/201525830`. arXiv: `1502.01589`.

[15] SNO. "Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory". *Phys. Rev. Lett.* 87 (2001), p. 071301. DOI: `10.1103/PhysRevLett.87.071301`. arXiv: `nucl-ex/0106015`.

[16] Super-Kamiokande. “Evidence for oscillation of atmospheric neutrinos”. *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: `10.1103/PhysRevLett.81.1562`. arXiv: `hep-ex/9807003`.

[17] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: `10.1070/PU1991v034n05ABEH002497`.

[18] “LHC Machine”. *JINST* 3 (2008). Ed. by L. Evans and P. Bryant, S08001. DOI: `10.1088/1748-0221/3/08/S08001`.

[19] I. Zurbano F. et al. “High-Luminosity Large Hadron Collider (HL-LHC): Technical design report”. 10/2020 (Dec. 2020). Ed. by I. Béjar Alonso et al. DOI: `10.23731/CYRM-2020-0010`.

[20] “The International Linear Collider Technical Design Report - Volume 1: Executive Summary” (June 2013). Ed. by T. Behnke et al. arXiv: `1306.6327 [physics.acc-ph]`.

[21] CLIC CLICdp. “The Compact Linear Collider (CLIC) - 2018 Summary Report”. 2/2018 (Dec. 2018). Ed. by P. N. Burrows et al. DOI: `10.23731/CYRM-2018-002`. arXiv: `1812.06018 [physics.acc-ph]`.

[22] FCC. “FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2”. *Eur. Phys. J. ST* 228.2 (2019), pp. 261–623. DOI: `10.1140/epjst/e2019-900045-4`.

[23] CEPC Study Group. “CEPC Technical Design Report – Accelerator” (Dec. 2023). arXiv: `2312.14363 [physics.acc-ph]`.

[24] ATLAS. “ATLAS HL-LHC Computing Conceptual Design Report” (2020).

[25] Geant4 Collaboration. “Geant4—a simulation toolkit”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/S0168-9002(03)01368-8`.

[26] J. Allison et al. “Geant4 developments and applications”. *IEEE Trans. Nucl. Sci.* 53 (2006), p. 270. DOI: `10.1109/TNS.2006.869826`.

[27] Geant4 Collaboration. “Recent developments in Geant4”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2016.06.125`.

[28] HEPiX Benchmarking Working Group. “Next generation of HEP CPU benchmarks”. *EPJ Web Conf.* 214 (2019). Ed. by A. Forti et al., p. 08011. DOI: `10.1051/epjconf/201921408011`.

[29] ATLAS. *ATLAS Software and Computing HL-LHC Roadmap*. Tech. rep. Geneva: CERN, 2022. URL: `http://cds.cern.ch/record/2802918`.

[30] CMS Offline Software and Computing. *CMS Phase-2 Computing Model: Update Document*. Tech. rep. Geneva: CERN, 2022. URL: `http://cds.cern.ch/record/2815292`.

[31] G. Grindhammer and S. Peters. “The Parameterized simulation of electromagnetic showers in homogeneous and sampling calorimeters”. *International Conference on Monte Carlo Simulation in High-Energy and Nuclear Physics - MC 93*. Feb. 1993. arXiv: `hep-ex/0001020`.

[32] G. Grindhammer, M. Rudowicz, and S. Peters. “The fast simulation of electromagnetic and hadronic showers”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 290.2 (1990), pp. 469–488. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/0168-9002(90)90566-O`.

[33] S. Glazov. “Fast Simulation of Showers in the HI SpaCal Calorimeter”. *Journal of Physics: Conference Series* 293.1 (Apr. 2011), p. 012024. DOI: `10.1088/1742-6596/293/1/012024`.

[34] E Barberio et al. “Fast simulation of electromagnetic showers in the ATLAS calorimeter: Frozen showers”. *Journal of Physics: Conference Series* 160.1 (Apr. 2009), p. 012082. DOI: `10.1088/1742-6596/160/1/012082`.

[35] A. Butter, Tilman Plehn, and Ramon Winterhalder. “How to GAN LHC Events”. *SciPost Phys.* 7.6 (2019), p. 075. DOI: `10.21468/SciPostPhys.7.6.075`. arXiv: `1907.03764 [hep-ph]`.

[36] C. Gao, Joshua Isaacson, and Claudius Krause. “i-flow: High-dimensional Integration and Sampling with Normalizing Flows”. *Mach. Learn. Sci. Tech.* 1.4 (2020), p. 045023. DOI: `10.1088/2632-2153/abab62`. arXiv: `2001.05486 [physics.comp-ph]`.

[37] E. Bothmann et al. “Exploring phase space with Neural Importance Sampling”. *SciPost Phys.* 8.4 (2020), p. 069. DOI: `10.21468/SciPostPhys.8.4.069`. arXiv: `2001.05478 [hep-ph]`.

[38] C. Gao et al. “Event Generation with Normalizing Flows”. *Phys. Rev. D* 101.7 (2020), p. 076002. DOI: `10.1103/PhysRevD.101.076002`. arXiv: `2001.10028 [hep-ph]`.

[39] T. Heimel et al. “Precision-Machine Learning for the Matrix Element Method” (Oct. 2023). arXiv: `2310.07752 [hep-ph]`.

[40] A. Butter et al. “Kicking it Off(-shell) with Direct Diffusion” (Nov. 2023). arXiv: `2311.17175 [hep-ph]`.

[41] A. Ghosh et al. “Towards a deep learning model for hadronization”. *Phys. Rev. D* 106.9 (2022), p. 096020. DOI: `10.1103/PhysRevD.106.096020`. arXiv: `2203.12660 [hep-ph]`.

[42] J. Chan et al. “Fitting a deep generative hadronization model”. *JHEP* 09 (2023), p. 084. DOI: `10.1007/JHEP09(2023)084`. arXiv: `2305.17169 [hep-ph]`.

[43] C. Bierlich et al. “Towards a data-driven model of hadronization using normalizing flows” (Nov. 2023). arXiv: `2311.09296 [hep-ph]`.

[44] M. Paganini, L. de Oliveira, and B. Nachman. “Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters”. *Phys. Rev. Lett.* 120.4 (2018), p. 042003. DOI: `10.1103/PhysRevLett.120.042003`. arXiv: `1705.02355 [hep-ex]`.

[45] E. Buhmann et al. “Decoding Photons: Physics in the Latent Space of a BIB-AE Generative Network”. *EPJ Web Conf.* 251 (2021), p. 03003. DOI: `10.1051/epjconf/202125103003`. eprint: `2102.12491`.

[46] ATLAS Collaboration. *Deep generative models for fast photon shower simulation in ATLAS*. 2022. arXiv: `2210.06204 [hep-ex]`.

[47] J. C. Cresswell et al. *CaloMan: Fast generation of calorimeter showers with density estimation on learned manifolds*. 2022. arXiv: `2211.15380 [hep-ph]`.

[48] C. Krause and D. Shih. “CaloFlow: Fast and Accurate Generation of Calorimeter Showers with Normalizing Flows” (June 2021). arXiv: `2106.05285 [physics.ins-det]`.

[49] C. Krause and D. Shih. "CaloFlow II: Even Faster and Still Accurate Generation of Calorimeter Showers with Normalizing Flows" (Oct. 2021). arXiv: `2110.11377 [physics.ins-det]`.

[50] S. Diefenbacher et al. "L2LFlows: Generating High-Fidelity 3D Calorimeter Images" (2023). arXiv: `2302.11594 [physics.ins-det]`.

[51] M. R. Buckley et al. *Inductive CaloFlow*. 2023. arXiv: `2305.11934 [physics.ins-det]`.

[52] F. T. Acosta et al. "Comparison of Point Cloud and Image-based Models for Calorimeter Fast Simulation" (July 2023). arXiv: `2307.04780 [cs.LG]`.

[53] M. Paganini, L. de Oliveira, and B. Nachman. "CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks". *Phys. Rev. D* 97.1 (2018), p. 014021. DOI: `10.1103/PhysRevD.97.014021`. arXiv: `1712.10321 [hep-ex]`.

[54] O. Amram and K. Pedro. "Denoising diffusion models with geometry adaptation for high fidelity calorimeter simulation". *Phys. Rev. D* 108.7 (2023), p. 072014. DOI: `10.1103/PhysRevD.108.072014`. arXiv: `2308.03876 [physics.ins-det]`.

[55] ATLAS Collaboration. "AtlFast3: The Next Generation of Fast Simulation in ATLAS". *Comput. Softw. Big Sci.* 6.7 (2022). DOI: `https://doi.org/10.1007/s41781-021-00079-7`.

[56] G. R. Khattak et al. "Fast Simulation of a High Granularity Calorimeter by Generative Adversarial Networks". *The European Physical Journal C* 8.386 (2022). DOI: `10.1140/epjc/s10052-022-10258-4`. arXiv: `2109.07388 [physics.ins-det]`.

[57] M. Erdmann et al. "Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks". *Comput. Softw. Big Sci.* 2.1 (2018), p. 4. DOI: `10.1007/s41781-018-0008-x`. arXiv: `1802.03325 [astro-ph.IM]`.

[58] M. Erdmann, J. Glombitza, and T. Quast. "Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network". *Comput. Softw. Big Sci.* 3.1 (2019), p. 4. DOI: `10.1007/s41781-018-0019-7`. arXiv: `1807.01954 [physics.ins-det]`.

[59] P. Musella and F. Pandolfi. "Fast and Accurate Simulation of Particle Detectors Using Generative Adversarial Networks". *Comput. Softw. Big Sci.* 2.1 (2018), p. 8. DOI: `10.1007/s41781-018-0015-y`. arXiv: `1805.00850 [hep-ex]`.

[60] The ATLAS collaboration. *Deep generative models for fast shower simulation in ATLAS*. Tech. rep. Geneva: CERN, 2018. URL: `http://cds.cern.ch/record/2630433`.

[61] F. Carminati et al. "Three dimensional Generative Adversarial Networks for fast simulation". *J. Phys. Conf. Ser.* 1085.3 (2018), p. 032016. DOI: `10.1088/1742-6596/1085/3/032016`.

[62] D. Belayneh et al. "Calorimetry with deep learning: particle simulation and reconstruction for collider physics". *The European Physical Journal C* 80.7 (July 2020). DOI: `10.1140/epjc/s10052-020-8251-9`.

[63] E. Buhmann et al. "Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed". *Comput. Softw. Big Sci.* 5.1 (2021), p. 13. DOI: `10.1007/s41781-021-00056-0`. eprint: `2005.05334`.

[64] ILD Concept Group. "International Large Detector: Interim Design Report" (Mar. 2020). arXiv: `2003.01116 [physics.ins-det]`.

[65] Cush. *Standard Model of Elementary Particles*. `https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`. Accessed: 10.01.2024. 2019.

[66] A. Pich. "The Standard Model of Electroweak Interactions". *2010 European School of High Energy Physics*. Jan. 2012, pp. 1–50. arXiv: `1201.0537 [hep-ph]`.

[67] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`.

[68] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`.

[69] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`.

[70] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the Color Octet Gluon Picture". *Phys. Lett. B* 47 (1973), pp. 365–368. DOI: `10.1016/0370-2693(73)90625-4`.

[71] A. Pich. "Aspects of quantum chromodynamics". *ICTP Summer School in Particle Physics*. June 1999, pp. 53–102. arXiv: `hep-ph/0001118`.

[72] ATLAS. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`. arXiv: `1207.7214 [hep-ex]`.

[73] CMS. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: `10.1016/j.physletb.2012.08.021`. arXiv: `1207.7235 [hep-ex]`.

[74] C. Adolphsen et al. "The Development of Energy-Recovery Linacs" (July 2022). arXiv: `2207.02095 [physics.acc-ph]`.

[75] A. D. Martin et al. "Parton distributions for the LHC". *Eur. Phys. J. C* 63 (2009), pp. 189–285. DOI: `10.1140/epjc/s10052-009-1072-5`. arXiv: `0901.0002 [hep-ph]`.

[76] "The International Linear Collider Technical Design Report - Volume 3.II: Accelerator Baseline Design" (June 2013). Ed. by C. Adolphsen et al. arXiv: `1306.6328 [physics.acc-ph]`.

[77] M. Bai et al. "C^3: A "Cool" Route to the Higgs Boson and Beyond". *Snowmass 2021*. Oct. 2021. arXiv: `2110.15800 [hep-ex]`.

[78] B. Foster, R. D'Arcy, and C. A. Lindstrom. "A hybrid, asymmetric, linear Higgs factory based on plasma-wakefield and radio-frequency acceleration". *New J. Phys.* 25.9 (2023), p. 093037. DOI: `10.1088/1367-2630/acf395`. arXiv: `2303.10150`.

[79] V. N. Litvinenko, T. Roser, and M. Chamizo-Llatas. "High-energy high-luminosity e^+e^- collider using energy-recovery linacs". *Phys. Lett. B* 804 (2020), p. 135394. DOI: `10.1016/j.physletb.2020.135394`. arXiv: `1909.04437 [physics.acc-ph]`.

[80] ILC. "The International Linear Collider Technical Design Report - Volume 2: Physics" (June 2013). Ed. by H. Baer et al. arXiv: `1306.6352 [hep-ph]`.

[81] "The International Linear Collider Technical Design Report - Volume 3.I: Accelerator \& in the Technical Design Phase" (June 2013). Ed. by C. Adolphsen et al. arXiv: `1306.6353 [physics.acc-ph]`.

[82] H. Abramowicz et al. "The International Linear Collider Technical Design Report - Volume 4: Detectors" (June 2013). Ed. by T. Behnke et al. arXiv: `1306.6329 [physics.ins-det]`.

[83] ILC International Development Team. "The International Linear Collider: Report to Snowmass 2021" (Mar. 2022). arXiv: `2203.07622 [physics.acc-ph]`.

[84] P. Garcia-Abia and W. Lohmann. "Measurement of the Higgs cross-section and mass with linear colliders". *Eur. Phys. J. direct* 2.1 (2000). Ed. by T. Behnke et al., p. 2. DOI: `10.1007/s1010500c0002`. arXiv: `hep-ex/9908065`.

[85] J. Yan et al. "Measurement of the Higgs boson mass and $e^+e^- \to ZH$ cross section using $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ at the ILC". *Phys. Rev. D* 94.11 (2016). [Erratum: Phys.Rev.D 103, 099903 (2021)], p. 113002. DOI: `10.1103/PhysRevD.94.113002`. arXiv: `1604.07524 [hep-ex]`.

[86] C. F. Dürig. "Measuring the Higgs Self-coupling at the International Linear Collider". PhD thesis. Hamburg U., 2016. DOI: `10.3204/PUBDB-2016-04283`.

[87] CLICdp. "Double Higgs boson production and Higgs self-coupling extraction at CLIC". *Eur. Phys. J. C* 80.11 (2020), p. 1010. DOI: `10.1140/epjc/s10052-020-08567-7`. arXiv: `1901.05897 [hep-ex]`.

[88] J. M. Torndal et al. "Higgs self-coupling measurement at future e^+e^- colliders". *2023 European Physical Society Conference on High Energy Physics*. Nov. 2023. arXiv: `2311.16774 [hep-ex]`.

[89] R. Covarelli, M. Pellen, and M. Zaro. "Vector-Boson scattering at the LHC: Unraveling the electroweak sector". *Int. J. Mod. Phys. A* 36.16 (2021), p. 2130009. DOI: `10.1142/S0217751X2130009X`. arXiv: `2102.10991 [hep-ph]`.

[90] J. Fan, M. Reece, and L.-T. Wang. "Possible Futures of Electroweak Precision: ILC, FCC-ee, and CEPC". *JHEP* 09 (2015), p. 196. DOI: `10.1007/JHEP09(2015)196`. arXiv: `1411.1054 [hep-ph]`.

[91] J. De Blas et al. "On the future of Higgs, electroweak and diboson measurements at lepton colliders". *JHEP* 12 (2019), p. 117. DOI: `10.1007/JHEP12(2019)117`. arXiv: `1907.04311 [hep-ph]`.

[92] D. Asner et al. "Top quark precision physics at the International Linear Collider". *Snowmass 2013: Snowmass on the Mississippi*. July 2013. arXiv: `1307.8265 [hep-ex]`.

[93] CLICdp. "Top-Quark Physics at the CLIC Electron-Positron Linear Collider". *JHEP* 11 (2019), p. 003. DOI: `10.1007/JHEP11(2019)003`. arXiv: `1807.02441 [hep-ex]`.

[94] Linear Collider. "The International Linear Collider Machine Staging Report 2017" (Nov. 2017). arXiv: `1711.00568 [physics.acc-ph]`.

[95] M. L. Perl et al. "Evidence for Anomalous Lepton Production in e+ - e- Annihilation". *Phys. Rev. Lett.* 35 (1975), pp. 1489–1492. DOI: `10.1103/PhysRevLett.35.1489`.

[96] Particle Data Group. "Review of Particle Physics". *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. ISSN: 2050-3911. DOI: `10.1093/ptep/ptac097`. eprint: `https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf`.

[97] T. H. Tran et al. "Reconstruction and classification of tau lepton decays with ILD". *Eur. Phys. J. C* 76.8 (2016), p. 468. DOI: `10.1140/epjc/s10052-016-4315-2`. arXiv: `1510.05224 [physics.ins-det]`.

[98] D. Jeans. "Tau lepton reconstruction at collider experiments using impact parameters". *Nucl. Instrum. Meth. A* 810 (2016), pp. 51–58. DOI: `10.1016/j.nima.2015.11.030`. arXiv: `1507.01700 [hep-ex]`.

[99] J. de Blas et al. "Higgs Boson Studies at Future Particle Colliders". *JHEP* 01 (2020), p. 139. DOI: `10.1007/JHEP01(2020)139`. arXiv: `1905.03764 [hep-ph]`.

[100] D. Jeans and G. W. Wilson. "Measuring the CP state of tau lepton pairs from Higgs decay at the ILC". *Phys. Rev. D* 98.1 (2018), p. 013007. DOI: `10.1103/PhysRevD.98.013007`. arXiv: `1804.01241 [hep-ex]`.

[101] L. Cerrito. *Radiation and Detectors*. Springer Cham, May 2017. ISBN: 978-3-319-53181-6. DOI: `10.1007/978-3-319-53181-6`.

[102] C. W. Fabjan and H. Schopper, eds. *Particle Physics Reference Library: Volume 2: Detectors for Particles and Radiation*. Cham: Springer, 2020. ISBN: 978-3-030-35317-9, 978-3-030-35318-6. DOI: `10.1007/978-3-030-35318-6`.

[103] Y. Israeli. "Energy Reconstruction in Highly Granular Calorimeters for Future Electron-Positron Colliders". PhD thesis. Munich, Tech. U., 2019.

[104] C. W. Fabjan and F. Gianotti. "Calorimetry for particle physics". *Rev. Mod. Phys.* 75 (4 Oct. 2003), pp. 1243–1286. DOI: `10.1103/RevModPhys.75.1243`.

[105] M. Livan and R. Wigmans. *Calorimetry for Collider Physics, an Introduction*. Springer Cham, July 2019. ISBN: 978-3-030-23653-3. DOI: `10.1007/978-3-030-23653-3`.

[106] R. Wigmans. *Calorimetry: Energy Measurement in Particle Physics*. 2nd ed. Oxford University Press, Sept. 2017. ISBN: 9780198786351. DOI: `10.1093/oso/9780198786351.001.0001`.

[107] CMS Collaboration. *The CMS electromagnetic calorimeter project : Technical Design Report*. Tech. rep. CERN-LHCC-97-033. CMS-TDR-4. Geneva: CERN, 1997.

[108] CMS Collaboration. "Performance and operation of the CMS electromagnetic calorimeter". *Journal of Instrumentation* 5.03 (2010), T03010. DOI: `10.1088/1748-0221/5/03/T03010`.

[109] CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Tech. rep. CERN-LHCC-2017-023. CMS-TDR-019. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2293646`.

[110] CALICE. "Design and Electronics Commissioning of the Physics Prototype of a Si-W Electromagnetic Calorimeter for the International Linear Collider". *JINST* 3 (2008), P08001. DOI: `10.1088/1748-0221/3/08/P08001`. arXiv: `0805.4833`.

[111] CALICE. "Performance of the first prototype of the CALICE scintillator strip electromagnetic calorimeter". *Nucl. Instrum. Meth. A* 763 (2014), pp. 278–289. DOI: `10.1016/j.nima.2014.06.039`. arXiv: `1311.3761 [physics.ins-det]`.

[112] CALICE. "Construction and Commissioning of the CALICE Analog Hadron Calorimeter Prototype". *JINST* 5 (2010), P05004. DOI: `10.1088/1748-0221/5/05/P05004`. arXiv: `1003.2662 [physics.ins-det]`.

[113] CALICE. "Description and stability of a RPC-based calorimeter in electromagnetic and hadronic shower environments". *JINST* 18.03 (2023), P03035. DOI: `10.1088/1748-0221/18/03/P03035`. arXiv: `2207.06291 [physics.ins-det]`.

[114] J. C. Brient. "Improving the jet reconstruction with the particle flow method: An introduction". *11th International Conference on Calorimetry in High-Energy Physics (Calor 2004)*. Mar. 2004, pp. 445–451.

[115] M. A. Thomson. "Particle Flow Calorimetry and the PandoraPFA Algorithm". *Nucl. Instrum. Meth. A* 611 (2009), pp. 25–40. DOI: `10.1016/j.nima.2009.09.009`. arXiv: `0907.3577 [physics.ins-det]`.

[116] J. S. Marshall and M. A. Thomson. “Pandora Particle Flow Algorithm”. *International Conference on Calorimetry for the High Energy Frontier*. 2013, pp. 305–315. arXiv: `1308.4537 [physics.ins-det]`.

[117] N. Feege. “Low-energetic hadron interactions in a highly granular calorimeter”. PhD thesis. Hamburg U., 2011.

[118] I. G. Knowles and G. D. Lafferty. “Hadronization in Z^0 decay”. *J. Phys. G* 23 (1997), pp. 731–789. DOI: `10.1088/0954-3899/23/7/003`. arXiv: `hep-ph/9705217`.

[119] J. S. Marshall and M. A. Thomson. “The Pandora Software Development Kit for Pattern Recognition”. *Eur. Phys. J. C* 75.9 (2015), p. 439. DOI: `10.1140/epjc/s10052-015-3659-3`. arXiv: `1506.05348 [physics.data-an]`.

[120] B. Xu. “Detectors and Physics at a Future Linear Collider”. PhD thesis. Cambridge U., Dec. 2017. DOI: `10.17863/CAM.16867`.

[121] B. Xu. “Improvement of photon reconstruction in PandoraPFA”. *International Workshop on Future Linear Colliders*. Feb. 2016. arXiv: `1603.00013 [physics.ins-det]`.

[122] F. Sauli. “GEM: A new concept for electron amplification in gas detectors”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 386.2 (1997), pp. 531–534. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/S0168-9002(96)01172-2`.

[123] Y. Giomataris et al. “MICROMEGAS: a high-granularity position-sensitive gaseous detector for high particle-flux environments”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 376.1 (1996), pp. 29–35. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/0168-9002(96)00175-1`.

[124] ILD SiW ECAL Collaboration. “Technical instrumentation R&D for ILD SiW ECAL large scale device”. *JINST* 13.03 (2018), p. C03047. DOI: `10.1088/1748-0221/13/03/C03047`. arXiv: `1712.05680 [physics.ins-det]`.

[125] R. Pöschl and H. Videau. *SiEcal Interface Control Document*. `https://edmsdirect.desy.de/item/D00000001162465,A,1,1`. Accessed: 12.12.2023. 2017.

[126] *iLCSoft Project Page*. `https://github.com/iLCSoft`. Accessed: 22.11.2023. 2016.

[127] Key4hep. “Key4hep: Progress Report on Integrations”. *26th International Conference on Computing in High Energy & Nuclear Physics*. Dec. 2023. arXiv: `2312.08152 [hep-ex]`.

[128] F. Gaede et al. “LCIO: A Persistency framework for linear collider simulation studies”. *eConf* C0303241 (2003), TUKT001. arXiv: `physics/0306114`.

[129] M. Frank et al. “DD4hep: A Detector Description Toolkit for High Energy Physics Experiments”. *J. Phys. Conf. Ser.* 513 (2014), p. 022010. DOI: `10.1088/1742-6596/513/2/022010`.

[130] M. Frank et al. “DDG4 A Simulation Framework based on the DD4hep Detector Description Toolkit”. *J. Phys. Conf. Ser.* 664.7 (2015), p. 072017. DOI: `10.1088/1742-6596/664/7/072017`.

[131] ILD CLICdp. “DD4Hep based event reconstruction”. *J. Phys. Conf. Ser.* 898.4 (2017). Ed. by Richard Mount and Craig Tull, p. 042017. DOI: `10.1088/1742-6596/898/4/042017`.

[132] F. Gaede. “Marlin and LCCD: Software tools for the ILC”. *Nucl. Instrum. Meth. A* 559 (2006). Ed. by J. Blumlein et al., pp. 177–180. DOI: `10.1016/j.nima.2005.11.138`.

[133] P. Bambade et al. “The International Linear Collider: A Global Project” (Mar. 2019). arXiv: `1903.01629 [hep-ex]`.

[134] F. Gaede et al. "Track reconstruction at the ILC: the ILD tracking software". *J. Phys. Conf. Ser.* 513 (2014). Ed. by D. L. Groep and D. Bonacorsi, p. 022011. DOI: `10.1088/1742-6596/513/2/022011`.

[135] H. L. Tran et al. "Software compensation in Particle Flow reconstruction". *Eur. Phys. J. C* 77.10 (2017), p. 698. DOI: `10.1140/epjc/s10052-017-5298-3`. arXiv: `1705.10363 [physics.ins-det]`.

[136] T. Suehara and T. Tanabe. "LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies". *Nucl. Instrum. Meth. A* 808 (2016), pp. 109–116. DOI: `10.1016/j.nima.2015.11.054`. arXiv: `1506.08371 [physics.ins-det]`.

[137] U. Einhaus et al. "Implementation, performance and physics impact of particle identification at Higgs factories". *2023 European Physical Society Conference on High Energy Physics*. Nov. 2023. arXiv: `2311.09102 [hep-ex]`.

[138] B. Dudar et al. "Development of time-of-flight particle identification for future Higgs factories". *2023 European Physical Society Conference on High Energy Physics*. Nov. 2023. arXiv: `2311.04720 [hep-ex]`.

[139] B. List, J. List, and DESY. "MarlinKinfit: An Object–Oriented Kinematic Fitting Package". *LC Notes* (2009). DOI: `10.3204/PHPPUBDB-10294`.

[140] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[141] B. H. Denby. "Neural Networks and Cellular Automata in Experimental High-energy Physics". *Comput. Phys. Commun.* 49 (1988), pp. 429–448. DOI: `10.1016/0010-4655(88)90004-5`.

[142] L. Lonnblad, C. Peterson, and T. Rognvaldsson. "Finding Gluon Jets With a Neural Trigger". *Phys. Rev. Lett.* 65 (1990), pp. 1321–1324. DOI: `10.1103/PhysRevLett.65.1321`.

[143] F. Gaede. "Exclusive production of Phi mesons in e p scattering at the H1 experiment at HERA". PhD thesis. Universität Kiel, 1998.

[144] J. J. Hopfield. "Neural networks and physical systems with emergent collective computational abilities ". *Proc. Natl. Acad. Sci. U S A.* 79.8 (1982), pp. 2554–8. DOI: `10.1073/pnas.79.8.2554`.

[145] Y. Lecun et al. "Gradient-based learning applied to document recognition". *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: `10.1109/5.726791`.

[146] M. Erdmann et al. *Deep Learning for Physics Research*. World Scientific, 2021. DOI: `10.1142/12294`. eprint: `https://www.worldscientific.com/doi/pdf/10.1142/12294`.

[147] M. D. Zeiler et al. "Deconvolutional networks". *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 2528–2535. DOI: `10.1109/CVPR.2010.5539957`.

[148] K. Lyu, Z. Li, and S. Arora. "Understanding the Generalization Benefit of Normalization Layers: Sharpness Reduction". *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022, pp. 34689–34708. URL: `https://proceedings.neurips.cc/paper_files/paper/2022/file/dffd1c523512e557f4e75e8309049213-Paper-Conference.pdf`.

[149] S. Ioffe and C. Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. ICML'15. Lille, France: JMLR.org, 2015, pp. 448–456.

[150] J. Lei Ba, J. R. Kiros, and G. E Hinton. "Layer normalization". *arXiv preprint arXiv:1607.06450* (2016).

[151] K. Fukushima. "Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements". *IEEE Transactions on Systems Science and Cybernetics* 5.4 (1969), pp. 322–333. DOI: `10.1109/TSSC.1969.300225`.

[152] R. H. Hahnloser et al. "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit". *Nature* 405.6789 (2000), pp. 947–51. DOI: `10.1038/35016072`.

[153] A. L Maas, A. Y Hannun, and A. Y. Ng. "Rectifier nonlinearities improve neural network acoustic models". *Proc. ICML*. Vol. 30. 2013.

[154] L. Bottou. "Online Algorithms and Stochastic Approximations". *Online Learning and Neural Networks*. Ed. by D. Saad. revised, oct 2012. Cambridge, UK: Cambridge University Press, 1998. URL: `http://leon.bottou.org/papers/bottou-98x`.

[155] B. Polyak. "Some methods of speeding up the convergence of iteration methods". *Ussr Computational Mathematics and Mathematical Physics* 4 (1964), pp. 1–17. URL: `https://api.semanticscholar.org/CorpusID:120243018`.

[156] J. Duchi, E. Hazan, and Y. Singer. "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". *J. Mach. Learn. Res.* 12 (July 2011), pp. 2121–2159. ISSN: 1532-4435.

[157] G. Hinton. *Neural Networks for Machine Learning, Lecture 6e*. `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`. 2012. URL: `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`.

[158] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Y. Bengio and Y. LeCun. 2015. DOI: `10.48550/arXiv.1412.6980`. eprint: `1412.6980`.

[159] D. P. Kingma and M. Welling. *Auto-Encoding Variational Bayes*. 2022. arXiv: `1312.6114 [stat.ML]`.

[160] D. J. Rezende, S. Mohamed, and D. Wierstra. *Stochastic Backpropagation and Approximate Inference in Deep Generative Models*. 2014. arXiv: `1401.4082 [stat.ML]`.

[161] I. Higgins et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". *International Conference on Learning Representations*. 2017. URL: `https://openreview.net/forum?id=Sy2fzU9gl`.

[162] J. M. Tomczak. *Deep Generative Modeling*. Springer Cham, 2022. DOI: `10.1007/978-3-030-93158-2`.

[163] I. Goodfellow et al. "Generative Adversarial Networks". *Commun. ACM* 63.11 (Oct. 2020), pp. 139–144. ISSN: 0001-0782. DOI: `10.1145/3422622`.

[164] J. F. Nash. "Equilibrium Points in N-Person Games." *Proc. Natl. Acad. Sci. U S A.* 36 1 (1950), pp. 48–9. URL: `https://api.semanticscholar.org/CorpusID:38771718`.

[165] M. Arjovsky, S. Chintala, and L. Bottou. "Wasserstein GAN" (2017). arXiv: `1701.07875 [stat.ML]`.

[166] D.A. Edwards. "On the Kantorovich–Rubinstein theorem". *Expositiones Mathematicae* 29.4 (2011), pp. 387–398. ISSN: 0723-0869. DOI: `https://doi.org/10.1016/j.exmath.2011.06.005`.

[167] I. Gulrajani et al. "Improved Training of Wasserstein GANs". *Advances in Neural Information Processing Systems 30*. 2017, pp. 5767–5777. arXiv: `1704.00028 [cs.LG]`. URL: `http://papers.nips.cc/paper/7159-improved-training-of-wasserstein-gans.pdf`.

[168] N. Tishby, F. C. Pereira, and W. Bialek. "The information bottleneck method" (2000). arXiv: `0004057 [physics.data-an]`.

[169] N. Tishby and N. Zaslavsky. "Deep learning and the information bottleneck principle" (2015). arXiv: `1503.02406 [cs.LG]`.

[170] D. J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge: Cambridge Univ. Pr., 2003, 628 pages. ISBN: 0521642981.

[171] S. Voloshynovskiy et al. "Information Bottleneck Through Variational Glasses" (2019). arXiv: `1912.00830 [cs.CV]`.

[172] D. J. Rezende and S. Mohamed. "Variational Inference with Normalizing Flows" (2015). arXiv: `1505.05770 [stat.ML]`.

[173] I. Kobyzev, S. J.D. Prince, and M. A. Brubaker. "Normalizing Flows: An Introduction and Review of Current Methods". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.11 (2021), pp. 3964–3979. DOI: `10.1109/TPAMI.2020.2992934`.

[174] L. Dinh, D. Krueger, and Y. Bengio. *NICE: Non-linear Independent Components Estimation*. 2015. arXiv: `1410.8516 [cs.LG]`.

[175] C. Durkan et al. *Neural Spline Flows*. 2019. arXiv: `1906.04032 [stat.ML]`.

[176] H. M. Dolatabadi, S. Erfani, and C. Leckie. *Invertible Generative Modeling using Linear Rational Splines*. 2020. arXiv: `2001.05168 [stat.ML]`.

[177] D. Ha, A. Dai, and Q. V. Le. *HyperNetworks*. 2016. arXiv: `1609.09106 [cs.LG]`.

[178] M. Bachtis. "Event Simulation". *Heavy Neutral Particle Decays to Tau Pairs: Detected with CMS in Proton Collisions at $\sqrt{s} = 7TeV$*. Cham: Springer International Publishing, 2014, pp. 47–54. ISBN: 978-3-319-03257-3. DOI: `10.1007/978-3-319-03257-3_4`.

[179] W. Kilian, T. Ohl, and J. Reuter. "WHIZARD: Simulating Multi-Particle Processes at LHC and ILC". *Eur. Phys. J. C* 71 (2011), p. 1742. DOI: `10.1140/epjc/s10052-011-1742-y`. arXiv: `0708.4233 [hep-ph]`.

[180] T. Sjostrand, S. Mrenna, and P. Z. Skands. "PYTHIA 6.4 Physics and Manual". *JHEP* 05 (2006), p. 026. DOI: `10.1088/1126-6708/2006/05/026`. arXiv: `hep-ph/0603175`.

[181] D. Schulte. "Beam-beam simulations with GUINEA-PIG" (1999). URL: `https://cds.cern.ch/record/382453`.

[182] Geant4 Collaboration. *Geant4 Physics Reference Manual, Release 11.1, Rev. 7.1*. 2023. URL: `https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsReferenceManual/fo/PhysicsReferenceManual.pdf`.

[183] A. F. Bielajew and B. Boulevard. "Fundamentals of the Monte Carlo method for neutral and charged particle transport". *Monte Carlo Methods Lecture Series, University of Michigan*. 2016. URL: `http://websites.umich.edu/~nersa590/MCbook.pdf`.

[184] S. M. Seltzer. "An Overview of ETRAN Monte Carlo Methods". *Monte Carlo Transport of Electrons and Photons*. Ed. by T. M. Jenkins, W. R. Nelson, and A. Rindi. Boston, MA: Springer US, 1988, pp. 153–181. ISBN: 978-1-4613-1059-4. DOI: `10.1007/978-1-4613-1059-4_7`.

[185] G. Folger and J. P. Wellisch. "String parton models in GEANT4". *eConf* C0303241 (2003), MOMT007. arXiv: `nucl-th/0306007`.

[186] D.H. Wright and M.H. Kelsey. "The Geant4 Bertini Cascade". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 804 (2015), pp. 175–188. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2015.09.058`.

[187] M. Frank. *DDG4: A Simulation Toolkit for High Energy Physics Experiments using Geant4 and the DD4hep Geometry Description, version 1.0.* 2013. URL: `https://dd4hep.web.cern.ch/dd4hep/usermanuals/DDG4Manual/DDG4Manual.pdf`.

[188] DELPHES 3. "DELPHES 3, A modular framework for fast simulation of a generic collider experiment". *JHEP* 02 (2014), p. 057. DOI: `10.1007/JHEP02(2014)057`. arXiv: `1307.6346 [hep-ex]`.

[189] M. Berggren. "SGV 3.0 - a fast detector simulation". *International Workshop on Future Linear Colliders (LCWS11).* Mar. 2012. arXiv: `1203.0217 [physics.ins-det]`.

[190] S. D. Diefenbacher. "Topics in Generative Modeling of Particle Physics Data". PhD thesis. Universität Hamburg, 2022.

[191] S. Diefenbacher et al. *Photon Showers in a High Granularity Calorimeter with Varying Incident Energy and Angle.* Zenodo, Apr. 2023. DOI: `10.5281/zenodo.7786846`.

[192] A. Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". *Advances in Neural Information Processing Systems 32* (2019). Ed. by H. Wallach et al. Advances in Neural Information Processing Systems 32 pp. 8024–8035, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[193] A. Gretton et al. "A Kernel Method for the Two-Sample Problem". *CoRR* (2008). arXiv: `0805.2368 [cs.LG]`.

[194] T. Salimans et al. "Improved Techniques for Training GANs" (2016). arXiv: `1606.03498 [cs.LG]`.

[195] E. Parzen. "On Estimation of a Probability Density Function and Mode". English. *The Annals of Mathematical Statistics* 33.3 (1962). The Annals of Mathematical Statistics **33**, pp. 1065, pp. 1065–1076. URL: `http://www.jstor.org/stable/2237880`.

[196] E. Bingham et al. "Pyro: Deep Universal Probabilistic Programming". *J. Mach. Learn. Res.* 20 (2019), 28:1–28:6. URL: `http://jmlr.org/papers/v20/18-403.html`.

[197] L. Dinh, J. Sohl-Dickstein, and S. Bengio. "Density estimation using Real NVP" (2016). arXiv: `1605.08803 [cs.LG]`.

[198] ONNX Runtime developers. *ONNX Runtime*. `https://onnxruntime.ai/`. Accessed: 30.12.2023. 2021.

[199] A. Boehnlein et al. *HL-LHC Software and Computing Review Panel, 2nd Report.* Tech. rep. CERN-LHCC-2022-007 ; LHCC-G-183. Geneva: CERN, Mar. 2022. URL: `http://cds.cern.ch/record/2803119/`.

[200] V. Völkl et al. "The Key4hep turnkey software stack". *PoS* ICHEP2022 (2022), p. 234. DOI: `10.22323/1.414.0234`.

[201] H. Ono and A. Miyamoto. "Status of ILD new 250 GeV common MC sample production". *International Workshop on Future Linear Colliders.* May 2021. arXiv: `2105.06040 [physics.acc-ph]`.

[202] S. Jadach, J. H. Kuhn, and Z. Was. "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons". *Comput. Phys. Commun.* 64 (1990), pp. 275–299. DOI: `10.1016/0010-4655(91)90038-M`.

[203] S. R. Johnson et al. “Novel features and GPU performance analysis for EM particle transport in the Celeritas code”. *EPJ Web Conf.* 251 (2021), p. 03030. DOI: `10.1051/epjconf/202125103030`.

[204] S. C. Tognini et al. “$Celeritas$: GPU-accelerated particle transport for detector simulation in High Energy Physics experiments”. *Snowmass 2021*. Mar. 2022. arXiv: `2203.09467 [physics.data-an]`.

[205] G. Amadio et al. “Offloading electromagnetic shower transport to GPUs”. *J. Phys. Conf. Ser.* 2438.1 (2023), p. 012055. DOI: `10.1088/1742-6596/2438/1/012055`. arXiv: `2209.15445 [hep-ex]`.

[206] B. M. Gruber, G. Amadio, and S. Hageböck. “Challenges and opportunities integrating LLAMA into AdePT” (Feb. 2023). arXiv: `2302.08252 [hep-ex]`.

[207] TorchView developers. *TorchView Project Page*. `https://github.com/mert-kurttutan/torchview/releases`. Accessed: 02.02.2024. 2022.

Declaration on oath

I hereby declare in lieu of oath that I have written this dissertation myself and that I have not used any auxiliary materials or sources other than those indicated.

Hamburg, 12.02.2024

P. M

Signature of the doctoral student