Application of Generative Neural Networks to High Fidelity Simulation of Particle Detectors in High Energy Physics

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

vorgelegt von
Anatolii Korol

Hamburg
2026

Gutachter/innen der Dissertation: Prof. Dr. Gregor Kasieczka
Dr. Frank Gaede

Zusammensetzung der Prüfungskommission: Prof. Dr. Sven-Olaf Moch
Prof. Dr. Gregor Kasieczka
Dr. Frank Gaede
Prof. Dr. Luisa Lucie-Smith
Dr. Ties Behnke

Vorsitzende/r der Prüfungskommission: Prof. Dr. Sven-Olaf Moch

Datum der Disputation: 13.03.2026

Vorsitzender des Fach-Promotionsausschusses PHYSIK: Prof. Dr. Johannes Haller

Leiter des Fachbereichs PHYSIK: Prof. Dr. Markus Drescher

Dekan der Fakultät MIN: Prof. Dr.-Ing. Norbert Ritter

Zusammenfassung

Die Monte Carlo Simulation von Teilchenwechselwirkungen bildet die Grundlage für nahezu alle Aspekte moderner Kollisionsexperimente, wobei die Rechenkosten bei steigender Luminosität und zunehmender Detektorkomplexität die verfügbaren Ressourcen überschreiten werden. Die Simulation von Kalorimeterschauern macht den größten Anteil dieser Kosten aus, insbesondere für die hochgranularen Designs zukünftiger Anlagen. Tiefe generative Modelle des maschinellen Lernens bieten einen vielversprechenden Ansatz zur Lösung dieses Problems, indem sie aus vergleichsweise kleinen Monte Carlo Datensätzen lernen und neue Schauer effizienter erzeugen.

Diese Arbeit stellt CaloClouds vor, ein generatives Modell zur schnellen Simulation elektromagnetischer Schauer in hochgranularen Abtastkalorimetern, entwickelt und evaluiert anhand des elektromagnetischen Silizium-Wolfram Kalorimeters des International Large Detector (ILD). Im Gegensatz zu früheren Ansätzen basierend auf festen Voxelgittern beschreibt CaloClouds Schauer als Punktwolken – Mengen energiegewichteter Koordinaten mit kontinuierlichen räumlichen Positionen. Diese Formulierung minimiert Projektionsartefakte beim Abbilden auf irreguläre Detektorgeometrien und ermöglicht die Nutzung eines einzigen Modells für verschiedene Detektorpositionen mit gleicher longitudinaler Struktur.

Das Modell kombiniert zwei komplementäre Komponenten: ShowerFlow, einen Normalizing Flow für globale Eigenschaften wie Schichtenergien und Anzahl der Energiedepositionen, und ein Diffusionsmodell zur Erzeugung einzelner Energiedepositionen. Eine Vorverarbeitungspipeline wandelt Geant4-Simulationsschritte in hochauflösende Punktwolken um und liefert Trainingsdaten mit höherer Granularität als die physikalischen Auslesezellen.

Um die Beiträge der Datenrepräsentation von der Modellleistung zu trennen, führt diese Arbeit *optimale Schauer-Generatoren* ein, konstruiert durch Projektion von Geant4-Schritten auf reguläre Gitter unterschiedlicher Auflösung ohne generatives Modell. Vergleichstests zeigen, dass CaloClouds eine Genauigkeit nahe der des feinsten optimalen Generators erreicht, was darauf hindeutet, dass das Modell sein maximales Potenzial für die gegebene Datenrepräsentation ausschöpft. Gleichzeitig zeigen Vergleiche mit grobgranularen Generatoren grundlegende Einschränkungen von Festgitter-Darstellungen unabhängig von der Modellarchitektur.

Die vollständige Integration in die ILD-Softwarekette über die DDML-Bibliothek ermöglicht die Auswertung mit Standard-Rekonstruktionswerkzeugen einschließlich Pandora Particle Flow. Physikalische Benchmarks umfassen Ein-Photon-Observablen, Di-Photon-Separation und π^0-Rekonstruktion in $e^+e^- \to \tau^+\tau^-$-Ereignissen, wobei der Großteil der Ergebnisse eine erstklassige Übereinstimmung mit Geant4 zeigt. Auf CPU-Hardware ist CaloClouds mehr als 120-fach schneller als die vollständige Geant4-Simulation, während auf GPU-Hardware Beschleunigungen um mehrere Tausend möglich sind.

Insgesamt etablieren diese Ergebnisse CaloClouds als effektive Lösung, die Geschwindigkeit und Genauigkeit für schnelle Kalorimetersimulation ausbalanciert und eine Grundlage für die Simulationsanforderungen künftiger Kollisionsexperimente bildet.

Abstract

Monte Carlo simulation of particle interactions underpins nearly every aspect of modern collider experiments, yet its computational cost will exceed available resources as luminosities and detector complexities increase. Calorimeter shower simulation accounts for the dominant share of this cost, particularly for the highly granular designs planned for future facilities. Deep generative Machine Learning models offer a promising path toward addressing this bottleneck by learning from relatively small amounts of Monte Carlo data and then sampling new showers more efficiently.

This work introduces CaloClouds, a generative model for fast electromagnetic shower simulation in highly granular sampling calorimeters, developed and evaluated using the silicon-tungsten electromagnetic calorimeter of the International Large Detector (ILD) as a case study. In contrast to previous approaches built on fixed voxel grids, CaloClouds represents showers as point clouds – sets of energy-weighted coordinates with continuous spatial positions. This formulation minimizes the projection artefacts that arise when mapping grid-based outputs onto irregular detector readout geometries and enables the use of a single model across different detector positions with the same longitudinal structure.

The model combines two complementary components: ShowerFlow, a normalising flow that captures global properties such as per-layer energies and number of energy depositions, and a diffusion model that generates individual energy depositions. A dedicated preprocessing pipeline converts Geant4 simulation steps into high-resolution point clouds, providing training data at finer granularity than the physical readout cells.

To disentangle the contributions of data representation from model performance, this work introduces *optimal shower generators* constructed by projecting Geant4 steps onto virtual grids of varying resolution without generative modelling. Benchmarks against these references show that CaloClouds achieves fidelity close to the finest-grained optimal generator, indicating that the model reaches its maximum potential given the data representation. Simultaneously, comparisons with the coarse-grained optimal shower generators reveal fundamental limitations of fixed-grid representations independent of the generative model architecture.

Full integration into the ILD software chain via the DDML library enables evaluation with standard reconstruction tools, including Pandora particle flow. Physics benchmarks span single-photon observables, di-photon separation, and π^0 reconstruction in $e^+e^- \to \tau^+\tau^-$ events, with the majority demonstrating state-of-the-art agreement with Geant4. On CPU hardware, CaloClouds achieves more than 120-fold acceleration relative to full Geant4 simulation, while on GPU hardware, speedups of several thousand-fold are achievable.

Taken together, these results establish CaloClouds as an effective solution that balances speed and accuracy for fast calorimeter simulation, providing a foundation for meeting the simulation demands of next-generation collider experiments.

Contents

Preface

The results presented in this thesis were obtained between 2022 and 2026 at Deutsches Elektronen-Synchrotron DESY in Hamburg, in collaboration with colleagues and fellow researchers. The findings described in this thesis have been published in the following works:

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

[2] Erik Buhmann et al. "CaloClouds II: ultra-fast geometry-independent highly-granular calorimeter simulation". In: *JINST* 19.04 (2024), P04020. DOI: 10.1088/1748-0221/19/04/P04020. arXiv: 2309.05704 `[physics.ins-det]`

[3] Thorsten Buss et al. "CaloClouds3: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation". In: (Nov. 2025). arXiv: 2511.01460 `[physics.ins-det]`

[4] Thorsten Buss et al. "A First Full Physics Benchmark for Highly Granular Calorimeter Surrogates". In: (Nov. 2025). arXiv: 2511.17293 `[hep-ex]`

Additionally, the author has been involved in the following publications that are directly related to the topic of this thesis:

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

[6] Thorsten Buss et al. "CaloHadronic: a diffusion model for the generation of hadronic showers". In: (June 2025). arXiv: 2506.21720 `[physics.ins-det]`

[7] Oz Amram et al. "CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation". In: (Oct. 2024). Ed. by Claudius Krause et al. arXiv: 2410.21611 `[physics.ins-det]`

[8] Erik Buhmann et al. "Fast Simulation of Highly Granular Calorimeters with Generative Models: Towards a First Physics Application". In: *PoS* EPS-HEP2023 (2024), p. 568. DOI: 10.22323/1.449.0568

1 Introduction

The main goal of particle physics is to understand the fundamental building blocks and forces of our universe. So far, the most comprehensive description is provided by the Standard Model [9–11]. It describes three of the four fundamental forces of nature – strong, weak, and electromagnetic, excluding gravitational – each with its own mediator. According to the Standard Model, matter consists of twelve different fermionic particles: six quarks and six leptons.

The Standard Model has been remarkably successful in describing many known phenomena in particle physics and in providing experimental confirmation of its predictions. With the discovery of the Higgs boson [12, 13] by the ATLAS and CMS experiments at the Large Hadron Collider (LHC), the predicted set of Standard Model particles was completed. However, the Standard Model cannot be considered a complete theory, as it fails to describe several observed phenomena, such as the existence of dark matter and dark energy, gravity, the asymmetry between matter and antimatter in the universe, and the hierarchy problem. These limitations motivate searches for physics beyond the Standard Model. To this end, multiple high-luminosity experiments are being planned. Among them is the upgrade of the LHC to the High-Luminosity phase (HL-LHC) [14], as well as different experiments at future e^+e^- colliders, such as International Linear Collider (ILC) [15], the Compact Linear Collider (CLiC) [16] and the Future Circular Collider (FCC) [17]. The new experiments focus on high-precision measurements aimed at detecting even the smallest deviations from the Standard Model. Such precision is achieved, among other factors, through the large amount of data collected by the experiments.

A common way to compare theoretical predictions with experimental results is through detailed simulations of the facility and the physical processes within it, based on the latest theoretical knowledge. Additionally, Monte Carlo (MC) methods are crucial for developing and testing analysis tools and algorithms that are applied to experimental data. These simulations include both the particle collisions and a realistic model of the detector itself, along with the interactions of the collision products with its components. Traditional simulation in particle physics is performed with MC methods. Accurate detector geometries are most often simulated using the GEANT4 framework [18–20].

A full MC simulation of processes in a particle physics experiment must account for all particle interactions with matter along their trajectories through the detector facility. This is a highly CPU-intensive and extremely time-consuming task. As illustrated in Fig. 1.1, the ATLAS experiment in 2018 devoted more than one third of its total CPU time to MC simulations.

Moreover, future experiments designed for new physics searches will operate at much higher luminosities and produce significantly larger numbers of particles to be simulated. Furthermore, the design of detectors for high-luminosity physics is being adapted to provide finer granularity and improved stability under higher particle fluxes. These changes affect all of the detector parts. Therefore, the current approach to experimental simulations must be reconsidered to ensure that the studies remain feasible and efficient. To address these challenges, fast simulation techniques have become widely adopted across the field. These methods trade some level of physical accuracy for substantial computational speedup [21–24]. However, most existing fast simulation tools generate calorimeter responses at the reconstruction level rather than at the hit level, making them incompatible with the actual reconstruction software chains used by experiments. This incompatibility becomes especially critical with the emergence of sophisticated Particle Flow algorithms, which were designed to exploit the fine granularity of modern calorimeters and fundamentally rely on access to individual hit information.

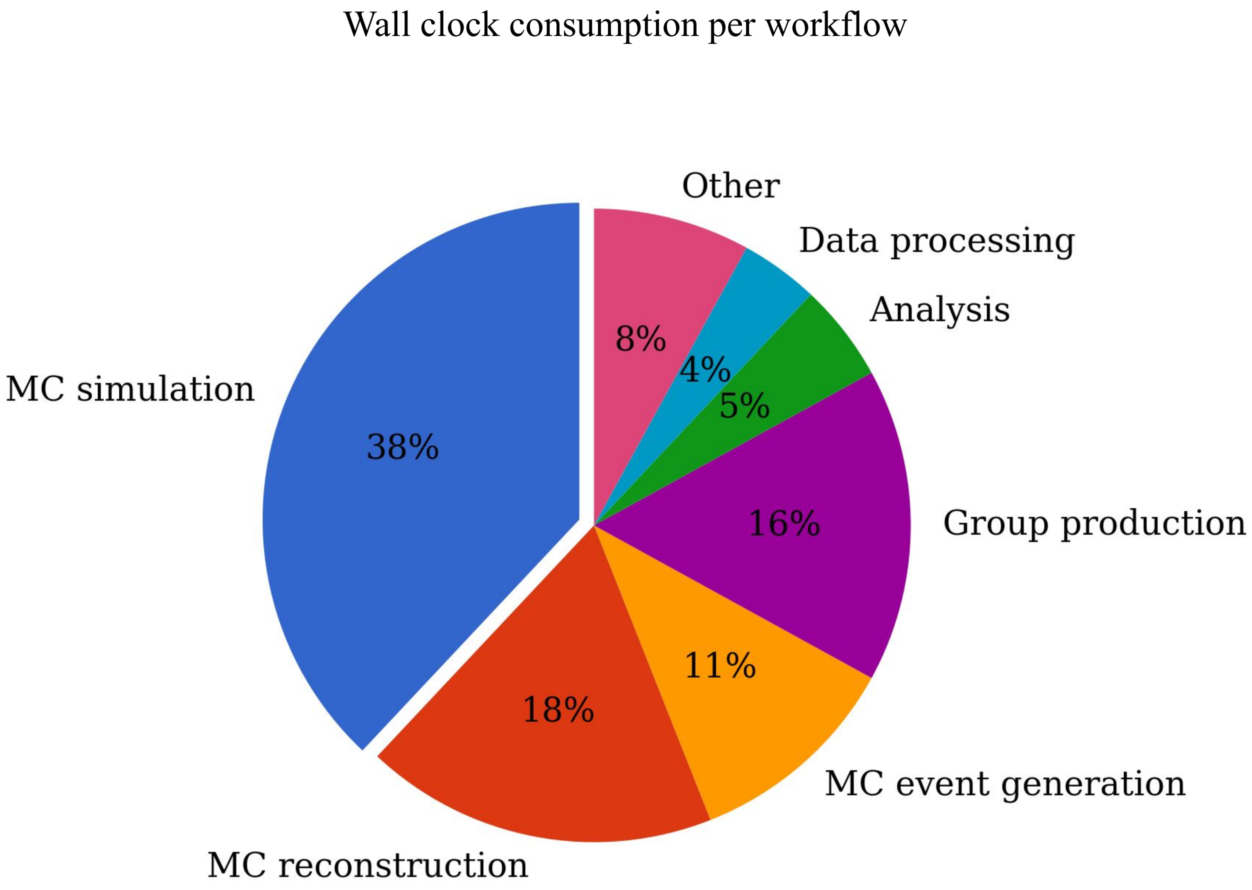


Figure 1.1: Distribution of CPU time across activities for the ATLAS experiment in 2018. Figure adapted from [25].

Moreover, despite their speed advantages, current fast simulation approaches still demand substantial computing resources that may prove infeasible for the needs of upcoming high-luminosity experiments. As illustrated in Fig. 1.2, the ATLAS and CMS collaborations predict excessive CPU usage for the HL-LHC phase even when employing fast simulation approaches.

Consequently, new approaches to simulating future physics experiments must be considered. Recent advances in generative Machine Learning and artificial intelligence, along with the increasing computational power of graphical processing units (GPUs), offer a wide range of possibilities across many scientific domains, including particle physics. Generative Machine Learning models can be trained to learn the underlying theoretical distributions represented by data samples and to generate new data following these patterns. Fast simulation using generative models has been explored for various processes, such as event generation and hadronization [26–33].

Simulation of the calorimeter showers has been a particularly interesting investigation area. The challenges associated with simulating high granularity and intense particle showers have

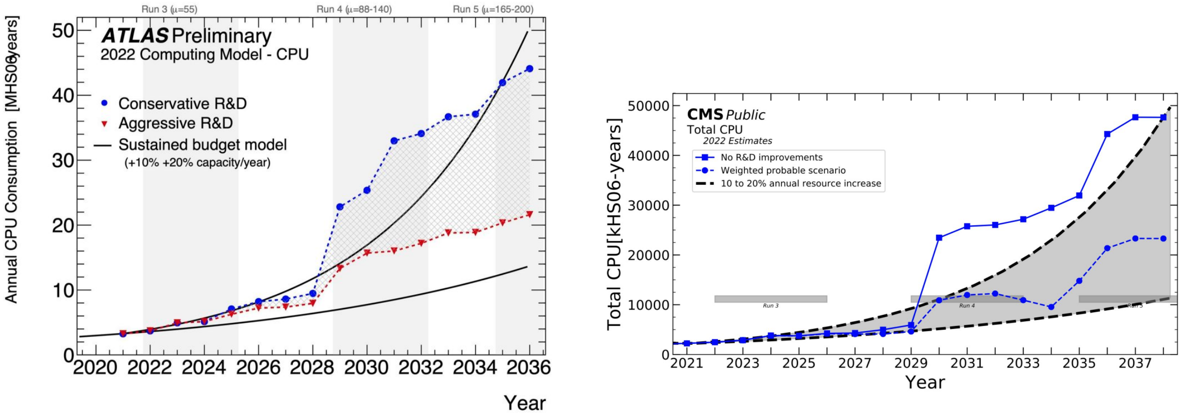


Figure 1.2: Projected CPU usage for the ATLAS (left) and CMS (right) experiments during the HL-LHC phase. Different R&D scenarios are illustrated, with computing resources evaluated using the HEP-SPEC06 (HS06) benchmarking suite [34]. The figures from [35] (left) and [36] (right).

been addressed in several studies using generative models [37–40]. The model that has proven to be most successful for reproducing detailed calorimeter responses under such conditions is the Bounded Information Bottleneck Autoencoder (BIB-AE) [37, 38, 40]. However, although previous studies have integrated BIB-AE into the realistic simulation chain, the BIB-AE model has primarily relied on a regular grid representation of the calorimeter. This approximation introduces non-negligible deviations in the reconstructed shower properties due to the simplification of the actual detector geometry [41].

The majority of previous efforts in this field have also been limited to models relying on fixed, regular detector readout geometries [7]. This thesis introduces the CaloClouds model, a geometry-independent diffusion model designed to overcome these limitations by generating calorimeter showers as point clouds for the electromagnetic calorimeter of the International Large Detector (ILD) [42].

The remainder of this thesis is structured as follows. Chapter 2 gives an overview of the principles of the Standard Model and outlines the motivation for exploring physics beyond it. Chapter 3 presents insights into the design of calorimeters used in current and future experiments, as well as the development and propagation of particle showers within them. In Chapter 5, the main principles of simulation in High Energy Physics are introduced and discussed. Chapter 6 introduces the fundamentals of Machine Learning techniques, which serve as the foundation for the simulation methods central to this work. It covers topics ranging from basic mathematical concepts, such as gradient descent, to advanced architectures like generative neural networks. Chapter 7 provides detailed descriptions of the CaloClouds model – developed to generate electromagnetic showers in highly granular calorimeters, with high speed and fidelity. This chapter describes the scope of the simulation studies conducted in this thesis, the model architecture, data processing and data representations required for training. The integration of this model into the ILD software framework, making it part of the overall simulation chain, is described in Chapter 8. Finally, Chapter 9 presents the results of the thesis, including a comprehensive evaluation of the model, and advantages of the proposed model, particularly in terms of its speed-accuracy trade-off, are highlighted. All major achievements of this work are summarized in Chapter 10, which also outlines possible future developments and applications.

2 Particle Physics

From the earliest times, humankind has been driven by curiosity about the structure of the world around it. This curiosity has been accompanied by the desire to understand the phenomena and forms of matter that occur both on Earth and beyond. Comprehension of the structure of matter is fundamental for understanding the structure of the Universe itself.

So far, our understanding of the elementary building blocks of matter has evolved from rather simplistic and inaccurate theories – such as that proposed by Anaximenes of Miletus, who believed that the world is composed of fire, water, earth, and air – to highly detailed descriptions that reach down to the elementary particles known today (see Fig. 2.1). The fundamental structure of matter in the Universe is the subject of particle physics, and the most successful theoretical framework describing the basic constituents of nature so far the Standard Model of Particle Physics (SM).

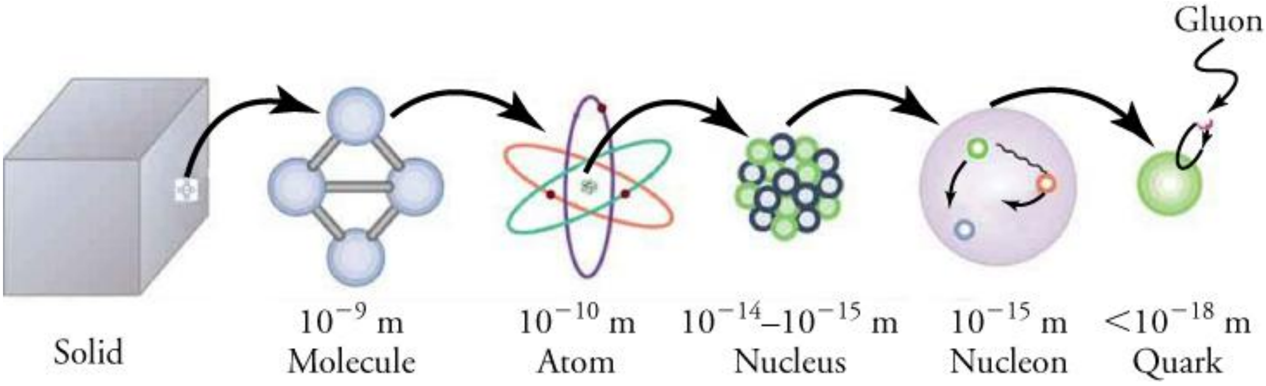


Figure 2.1: Simplified view on a structure of matter and building blocks of nature, starting from a solid body and ending with quarks and gluons. Figure from [43].

2.1 Remark on the Units

The scales encountered in particle physics are extremely small, which makes the use of standard S.I. units impractical. For this reason, alternative units are commonly used. These units are summarized in the table below:

Measure	Particle physics units	Standard S.I. units
Energy	1 eV (electron-Volts)	1.602×10^{-19} J
Area	1 barn	10^{-28} m^2

Additionally, it is common in particle physics to work in *natural units*, in which fundamental constants are set equal to unity, $\hbar = c = 1$.

Given the above, and using Einstein's energy-momentum relation, the natural units for mass and momentum in particle physics are also the electronvolt (eV).

2.2 The Standard Model of Particle Physics

The Standard Model [9–11] successfully describes numerous phenomena in particle physics. Its predictions have been confirmed by a wide range of experimental evidence.

According to the Standard Model, the fundamental constituents of matter are leptons and quarks. The interactions between them are mediated by field- or gauge bosons[1]. The set of particles in the Standard Model is shown in Fig. 2.2.

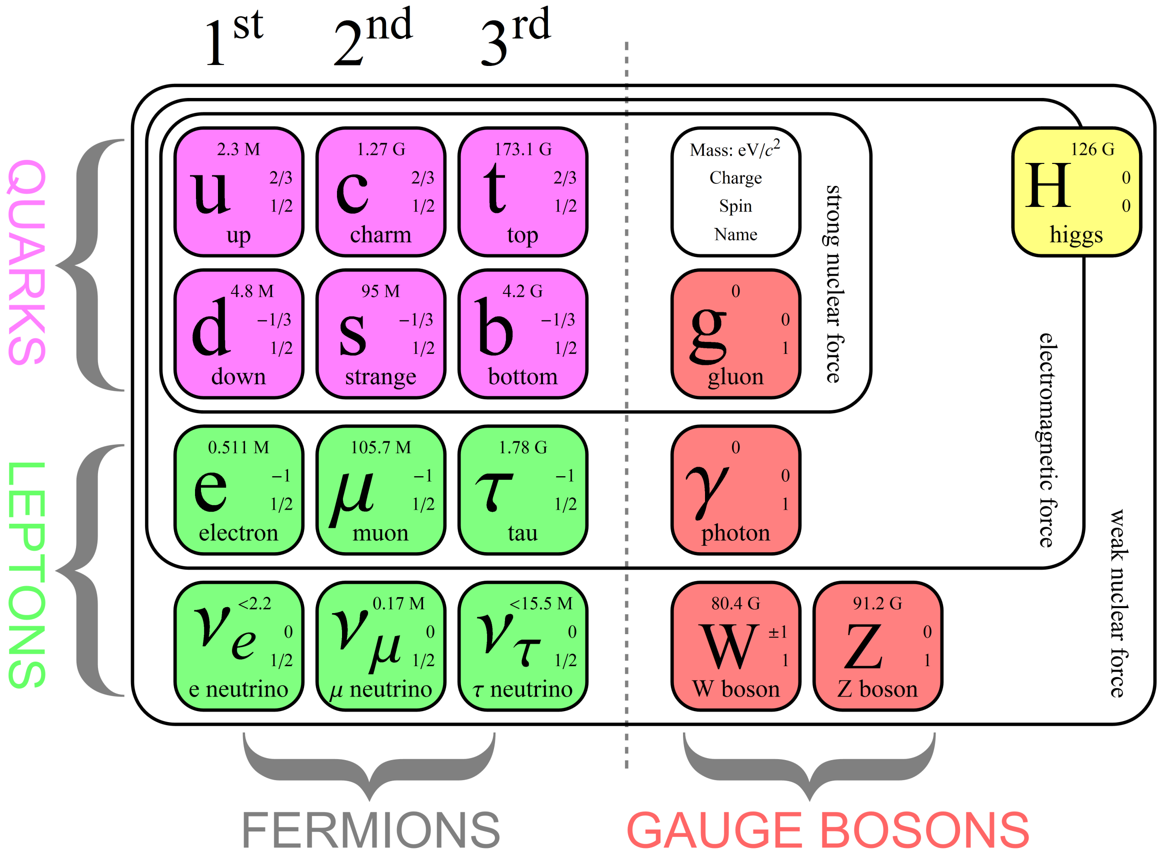


Figure 2.2: Standard Model set of elementary particles and gauge bosons with their main features. Figure from [44].

The Standard Model is a quantum field theory (QFT) that describes three of the four fundamental forces in nature: the electromagnetic, weak, and strong interactions.

The quantum field theory describing the electromagnetic interaction is Quantum Electrodynamics (QED). All particles with non-zero electric charge participate in electromagnetic interactions, which are mediated by the exchange of photons (γ).

The weak interaction is named for its comparatively low strength: it is approximately $10^4 - 10^5$ times weaker than the electromagnetic interaction and $10^6 - 10^7$ times weaker than the strong interaction [45]. The charge associated with the weak interaction is called *weak isospin*. The weak processes occur through two types of interactions: neutral-current interactions, mediated by the Z^0, and charged-current interactions, mediated by the $W^{\pm}$.

The unification of electromagnetic and weak interactions was described by Glashow, Salam, and Weinberg, forming the electroweak theory [46]. In the Standard Model, this is described

---

[1] A *boson* is any particle, which has an integer spin and which is characterised by Bose-Einstein statistics.

by an $SU(2)_L \otimes U(1)_Y$ gauge symmetry. The $SU(2)_L$ component involves three gauge bosons, W_1, W_2, and W_3. The Z boson arises as a combination of W_3 and the $U(1)_Y$ boson B, coupling to both left- and right-handed fermions, while $W^{\pm}$ mediate charged current interactions, and couple only to left-handed fermions.

Gauge bosons are massless under exact symmetry, but in nature $W^{\pm}$ and Z are massive due to spontaneous symmetry breaking via the Higgs mechanism:

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

The Higgs field gives masses to the $W^{\pm}$, Z bosons, and fermions, while the photon remains massless. The quantum of the Higgs field is the Higgs boson [47–49].

The $SU(3)_C$ component of the Standard Model describes the strong interaction through Quantum Chromodynamics (QCD) [50], a non-abelian gauge theory of quarks and gluons. All particles with non-zero color charge can participate in the strong interaction. Quarks and gluons carry colour charge, and gluons, being massless but colour-charged, can interact both with quarks and with each other. Gluon self-interaction requires a postulation of a phenomenon called *color confinement*, in which coloured particles cannot exist in isolation and must form colour-neutral states. This means that coloured particles cannot propagate freely. Instead, they are always bound into colour-neutral states.

Because confinement prevents free quarks and gluons from being formed, those produced in high-energy collisions undergo hadronisation, forming narrow sprays of colour-neutral hadrons known as *jets* [51].

All of the phenomena, described above, have been experimentally confirmed. The Large Hadron Collider (LHC) at CERN has played a critical role in improving high energy physics by producing particle interactions at previously unreachable energy levels. Discoveries such as the Higgs boson [12, 13] have validated key aspects of the Standard Model, yet also raised new questions about the fundamental structure of matter.

2.2.1 Limitations of the Standard Model

Although the Standard Model is highly successful in describing numerous phenomena in particle physics and is supported by a vast amount of experimental evidence, it still has important limitations. For example, it doesn't take into consideration dark energy, dark matter, or gravity. It also does not explain why there are exactly three generations of particles (see Fig. 2.2) or why there is an asymmetry between matter and antimatter in the Universe.

One important aspect that is not addressed by the Standard Model is the mass of neutrinos. Neutrino oscillations, which were experimentally established by the Super-Kamiokande experiment [52] and the Sudbury Neutrino Observatory [53], demonstrate that neutrinos change flavor as they propagate through space. Such behavior requires neutrinos to have non-zero mass, which is not possible in the Standard Model without postulating new fields and symmetries that haven't been observed.

The Standard Model also faces the *hierarchy problem*, which arises in determining the mass of the Higgs boson. Without fine-tuned quantum corrections, the Higgs mass would naturally increase toward the Planck scale of $\sim 10^{18}$GeV. Supersymmetry [54] could resolve this issue by canceling these corrections, but it currently lacks experimental evidence.

These gaps suggest that there is physics beyond the Standard Model, which has sparked a lot of research and investigation into pottential extensions such as supersymmetry and string theory.

2.3 Future Collider Experiments

Taking all the limitations of the Standard Model into account, it is clear that the Higgs boson offers important opportunities for exploring and searching for new physics. The experimental facilities capable of producing and studying particles such as the Higgs boson are particle colliders and the detectors built at their interaction points. This section provides an overview of collider experiments, their key characteristics, and the physics opportunities they offer, with particular focus on future electron-positron Higgs factories.

2.3.1 Collider Experiments

In the collider experiments, groups of charged particles or ions (*bunches*) are accelerated to high energies and brought into collision, producing a variety of physical processes and creating new particles. Two main types of collisions can be realized in such experiments:

- **Fixed-target experiments**: a bunch of accelerated particles collides with a stationary target, typically a layer of matter.
- **Collider experiments**: two bunches of accelerated particles are brought into collision with each other.

In this thesis, the focus is on collider experiments, in which a higher center-of-mass energy can be achieved, thus more and havier particles can be produced as the result. For collisions of two particle beams with equal energies, the center-of-mass energy is given by

$$\sqrt{s} = 2E_{\text{beam}}, \tag{2.2}$$

where E_{beam} is the energy of one of the beams immediately before the collision.

It is also very important how frequently a collider produces collisions. In particular, the study of rare processes requires large event statistics, making a high collision rate essential. The quantity that characterizes the collision rate in a collider is the *luminosity.* It is defined such that the number of events of a specific physical process produced per unit time is given by

$$N_{\text{proc}} = L \cdot \sigma_{\text{proc}}, \tag{2.3}$$

where L is the instaneous luminosity and σ_{proc} is the cross section of the process. The total luminosity accumulated over time is expressed as

$$\mathcal{L}_{\text{int}} = \int L(t), dt. \tag{2.4}$$

The integrated luminosity is typically evaluated over the operational lifetime of the collider.

There are two typical designs of collider experiments: **circular** and **linear** colliders. Both designs have their advantages and disadvantages. Circular synchrotron colliders are more commonly built, as they allow particles to circulate through the accelerator multiple times, which greatly increases integrated luminosity. However, each time an accelerated charged particle follows a curved trajectory, it loses energy due to synchrotron radiation. The energy loss caused by synchrotron radiation can be expressed as

$$\Delta E \propto \frac{E^4}{m^4 R}, \tag{2.5}$$

where E is the particle energy, m is its mass, and R is the radius of its trajectory.

Equation 2.5 shows that synchrotron energy losses scale with the fourth power of energy and inversely with the fourth power of mass. For hadrons, which are much heavier than leptons, these losses remain manageable even at very high energies, making circular hadron colliders practical at the TeV scale. For electron-positron colliders, many successful circular machines have operated at center-of-mass energies up to around 200 GeV, including LEP at CERN and the currently operating SuperKEKB in Japan. However, reaching significantly higher energies with circular lepton colliders becomes increasingly challenging due to the E^4 scaling of synchrotron radiation losses. For this reason, proposals for future lepton colliders span both circular designs, such as FCC-ee, optimized for precision measurements at moderate energies, and linear designs, such as ILC or CLIC, targeting higher energy reach.

2.3.2 Importance of Experiments with Lepton Collisions

There have been, and continue to be, several large circular collider experiments that have achieved breakthrough results and enabled the discovery of many particles and physical phenomena. Currently, the largest and highest-energy collider in operation is the Large Hadron Collider (LHC) at CERN [55]. It has a circumference of 26.7 km, reaches a center-of-mass energy for proton–proton collisions of up to $\sqrt{s} = 14\,\mathrm{TeV}$, and was designed to operate at a luminosity of $L = 10^{34}\,\mathrm{cm^{-2}s^{-1}}$. The instantaneous luminosity is planned to be increased by a factor of five through the High-Luminosity upgrade of the LHC (HL-LHC) [56], which would increase the total integrated luminosity by a factor of ten.

Such large hadronic collider experiments can reach very high center-of-mass energies, but at the same time they produce a large number of soft hadronic jets. This leads to high detector occupancies and makes precision measurements particularly challenging.

In contrast, lepton-lepton collisions exhibit low background levels, as electroweak interactions do not produce large numbers of jets. Consequently, the event rates remain comparatively low, enabling data collection with minimal or even no triggers. Another major advantage of lepton collisions is that leptons are fundamental particles, so they interact directly with one another. In contrast, hadron collisions involve interactions between partons, whose momenta at the time of the interaction are not precisely known. For colliding leptons, however, the centre-of-mass energy of collisions is precisely defined event-by-event.

These characteristics make lepton colliders essential facilities for precision measurements in high-energy physics. This work uses MC samples produced with the International Linear Collider (ILC) [15, 57–59] in mind. However, the results of this study are applicable to any future collider project. In particular, this work can be extremely useful for the e^+e^- Future Circular Collider project (FCC-ee) [17] studies, as the detector designs at these colliders share many similar features.

2.3.3 International Linear Collider

Several concepts for future high-energy electron-positron colliders serving as dedicated Higgs boson production facilities are under consideration worldwide. This work focuses on the International Linear Collider [15, 57–59], a proposed linear collider design. The proposed layout of the ILC is shown in Fig. 2.3.

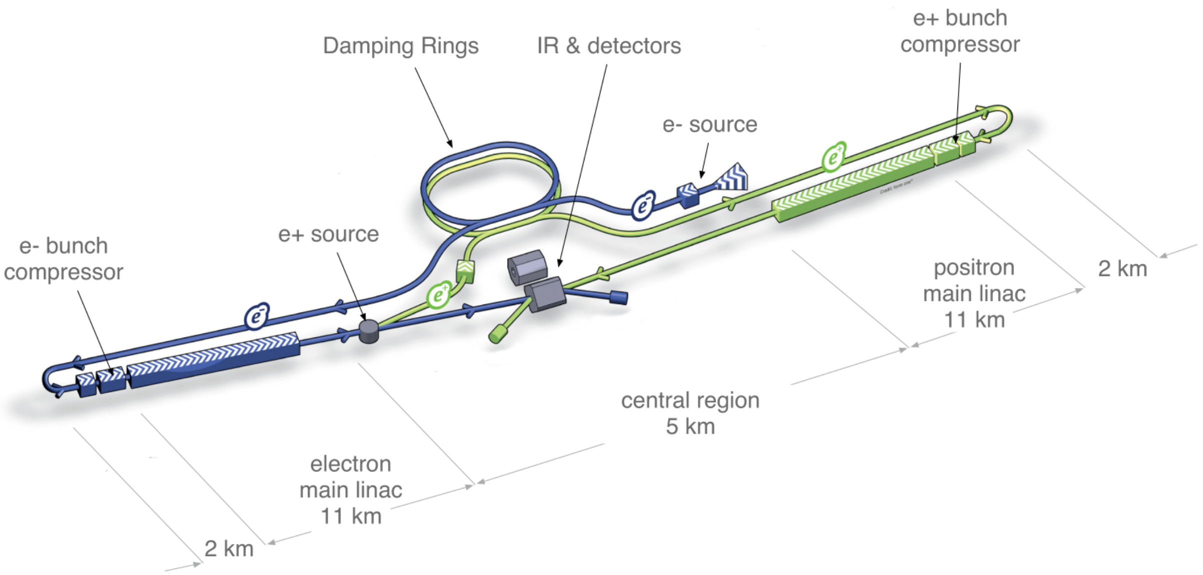


Figure 2.3: Schematic layout of the ILC, indicating all the major subsystems. Figure from [15].

The ILC will consisnt of two linear accelerators of a total length of 31 km, which bring longitudinally polarized electron and positron beams. The International Linear Collider is designed to be realised in a staged approach [60]. The initial stage foresees operation at a center-of-mass energy of $\sqrt{s} = 250\,\text{GeV}$. The accelerator design allows for subsequent energy upgrades, extending the achievable center-of-mass energy to values of up to 1 TeV.

The electron and positron beams are delivered in the form of bunch trains. Each train comprises 1312 bunches, with a temporal spacing of 554 ns between successive bunches. Each bunch contains approximately 2×10^{10} particles. The time interval between consecutive bunch trains is 200 ms, corresponding to a repetition rate of 5 Hz. At a center-of-mass energy of $\sqrt{s} = 250\,\text{GeV}$, the expected instantaneous luminosity is $1.35 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$.

At the interaction point, the beams collide with a crossing angle of 14 mrad. The collider design further includes the option of longitudinal beam polarisation, with electron polarisation of up to 80% and positron polarisation of up to 30%.

Two multipurpose detectors are proposed to operate at ILC: the Silicon Detector (SiD) [59] and the International Large Detector (ILD) [42]. Both detector concepts are designed to provide excellent performance for precision measurements, with a particular emphasis on efficient tracking, high-resolution calorimetry, and precise particle flow reconstruction. While sharing common physics goals and overall performance requirements, the two detectors differ in their technological choices and designs. The SiD has compact detector layout and an all-silicon tracking system, whereas the ILD features a larger detector volume and a gaseous time projection chamber as its central tracking device. These complementary approaches allow for cross-validation of measurements and systematic studies. This thesis will focus on the ILD.

2.4 Physics at Lepton Collider Experiments

Lepton collider experiments offer unique opportunities for precision measurements. This section provides an overview of selected processes that can be studied at such facilities.

2.4.1 Electroweak Physics

In contrast to hadron colliders, e^+e^- collider experiments provide a highly suitable environment for precision studies in electroweak physics, owing to the absence of large QCD backgrounds and the precise knowledge of the initial-state kinematics. In particular, high-precision measurements of the properties of the W and Z bosons become possible, enabling sensitivity to energy scales significantly beyond those directly accessible, through their impact on higher-order loop corrections [61, 62].

Another important process occurs at a center-of-mass energy of $\sqrt{s} = 250\,\text{GeV}$, where events with radiative return to the Z resonance are produced with a sizable rate. These events arise from the reaction $e^+e^- \to Z\gamma$, in which a photon is emitted by the incoming electron or positron as initial-state radiation. The emitted photon and the resulting Z boson are predominantly produced in opposite directions and at small polar angles with respect to the beam axis. Despite this forward direction of the emission, the Z boson can still be efficiently reconstructed using the detector concepts foreseen for the ILC [63].

2.4.2 Top Physics

Precision measurements of top-quark properties may also be performed at lepton colliders. For example, the process $e^+e^- \to t\bar{t}H$ provides direct access to the top-quark Yukawa coupling and allows for high-precision determinations of the top-quark mass as well as its electroweak couplings [64, 65]. The top quark, as the heaviest known elementary particle and the one with the strongest coupling to the Higgs boson, constitutes a powerful probe for searches for physics beyond the Standard Model.

2.4.3 Higgs Physics

Despite the discovery of the Higgs boson, many fundamental questions concerning its nature remain unanswered. In particular, it is still unclear whether the Higgs field corresponds to an elementary scalar particle or arises from a more complex underlying structure, whether it exhibits self-interactions consistent with the Standard Model, and whether it plays a role in physics beyond the visible sector, such as interactions with dark matter. Addressing these questions requires measurements of Higgs properties with a precision at the percent level or better, as potential effects of new physics are generally expected to induce deviations of only a few percent from the Standard Model predictions. In this context, a future e^+e^- collider offers a very well-controlled experimental environment for precision Higgs studies.

At center-of-mass energies around $\sqrt{s} = 250\,\text{GeV}$, Higgs production is dominated by the Higgs-strahlung process, $e^+e^- \to ZH$. By reconstructing the Z boson through its clean leptonic decay modes, $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$, the presence of the Higgs boson can be inferred independently of its decay products. The Higgs mass can then be extracted from the recoil using the relation below:

$$M^2_{\text{recoil}} = s - 2E_Z\sqrt{s} + M^2_Z, \tag{2.6}$$

where $\sqrt{s}$ is the center-of-mass energy, E_Z is the measured energy of the Z boson, and M_Z its reconstructed mass [66]. Since this method relies solely on the reconstruction of the Z boson, it provides a model-independent determination of Higgs properties, including its mass, total decay width, and absolute couplings [67]. In particular, it enables a precise measurement of the Higgs branching fraction into invisible final states, offering a sensitive probe of possible Higgs interactions with dark sector particles.

At higher center-of-mass energies, a linear e^+e^- collider further allows access to Higgs self-interactions, which are directly related to the shape of the Higgs potential. For collision energies of 500 GeV and above, the Higgs self-coupling can be studied through double Higgs production processes. Depending on the energy regime, the dominant contributions arise either from double Higgs-strahlung, $e^+e^- \to ZHH$, or from vector-boson fusion, $e^+e^- \to \nu_e\bar{\nu}_e HH$. Measurements of these processes provide a crucial test of the Standard Model and are sensitive to deviations induced by new physics [68–70].

Detailed studies of Higgs decay modes offer additional opportunities to probe the structure of the Higgs sector. Among these, the decay $H \to \tau^+\tau^-$ is of special interest, as it allows for a direct investigation of the CP structure of the Higgs–lepton interaction. The tau physics and the prospects for such studies will be discussed in detail in the following section.

2.4.4 Tau Leptons Physics

In the Standard Model, the Higgs boson is a CP-even scalar, so any CP-odd effects in its interactions would indicate new physics. The decay $H \to \tau^+\tau^-$ is well suited to probe the Higgs CP properties, due to its sizable branching ratio of $\sim 6.3\%$ [71] and the tau‘s sufficiently long lifetime, which allows the decay vertex to be reconstructed. Precise reconstruction of tau decay modes is essential for using tau spin correlations to study the Higgs CP structure [72]. Further details on tau reconstruction are provided in this section.

As shown in Fig. 2.2, the tau lepton is the heaviest lepton, with a mass of 1776.86 ± 0.12 MeV and a mean lifetime of $(290 \pm 0.5) \times 10^{-15}$ s [73]. Due to its large mass, the tau lepton is the only lepton capable of decaying into hadrons. The main decay modes of the tau leptons are listed in the Table 2.1. According to Table 2.1, the hadronic decay modes are dominant, accounting for approximately 65% of the total branching ratio. In contrast, the leptonic decay modes produce neutrinos, which cannot be detected directly. For this reason, the hadronic decay modes of the tau lepton are preferred in experimental studies.

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

Table 2.1: Dominant decay modes of tau leptons, including intermediate resonances, and their branching ratios. τ^- decays are listed, while τ^+ decay modes are identical under charge conjugation. $h^\pm$ is standing for $\pi^\pm$ or $K^\pm$. Table taken from [73].

Nevertheless, hadronic tau decays pose significant reconstruction challenges. Most proceed via intermediate resonances $\rho(770)$ or $a_1(1200)$ and frequently produce neutral pions. The decay products are highly boosted and more collimated than particles in QCD jets of comparable energy [74], which complicates their separation and accurate reconstruction. Miscounting photons from π^0 decays ($\pi^0 \to \gamma\gamma$, 99% of cases [73]) can result in incorrect π^0 multiplicities and misidentification of the tau decay mode. Thus, hadronic tau reconstruction serves as a standard benchmark for electromagnetic calorimeter performance [72, 74].

3 Calorimetry

Calorimeters are essential components of all high-energy physics experiments. Their primary purpose is to measure the energy of particles - mostly secondary particles produced in the initial interaction - that deposit their energy through interactions with the detector material. By analyzing the resulting particle showers, the energy of the original particles can be inferred. Depending on the particle type, energy deposition in calorimeters occurs via electromagnetic or hadronic processes, and the characteristics of the electromagnetic interaction can further depend on the mass of the particle. This chapter provides an overview of the fundamental principles of calorimetry and their application in particle physics experiments.

3.1 Electromagnetic Interaction of Light Particles with Matter

In this section, the electromagnetic interaction of common light particles, such as electrons, positrons, and photons will be described.

3.1.1 Light Charged Particles Interaction with Matter

The interaction modes of light ionizing charged particles, such as electrons and positrons, are illustrated in Fig. 3.1. As shown in the figure, the dominant mechanisms by which electrons and positrons interact with matter are as follows:

- **Ionization**: this is the dominant interaction mode. Ionisation occurs when an electron or positron transfers part of its kinetic energy to an atomic electron, resulting in the ejection of the latter and the formation of an ionised atom. As the particle energy increases beyond a material-dependent threshold (approximately 10 MeV in lead), the fractional energy loss due to ionisation decreases significantly. At this point *Bremsstrahlung* becomes significant.
- **Bremsstrahlung**: it arises when a charged particle is accelerated in the electromagnetic field of a nucleus of the matter, leading to the emission of a photon.
- **Electron scattering**: interactions with atomic electrons can result in elastic scattering. This process is referred to as *Møller scattering* for electrons and *Bhabha scattering* for positrons.
- **Positron annihilation**: positrons may undergo annihilation with electrons in the surrounding material, producing a pair of photons in the final state.

3.1.2 Photon Interaction with Matter

Photons also have different electromagnetic interaction processes with matter (see Fig. 3.2). The dominant interaction processes with matter are as follows:

- **Photoelectric Effect:** occurs when a photon is absorbed by an atom, transferring enough energy to free a bound electron. This process dominates at low photon energies, but its cross-section decreases rapidly as the photon energy increases.

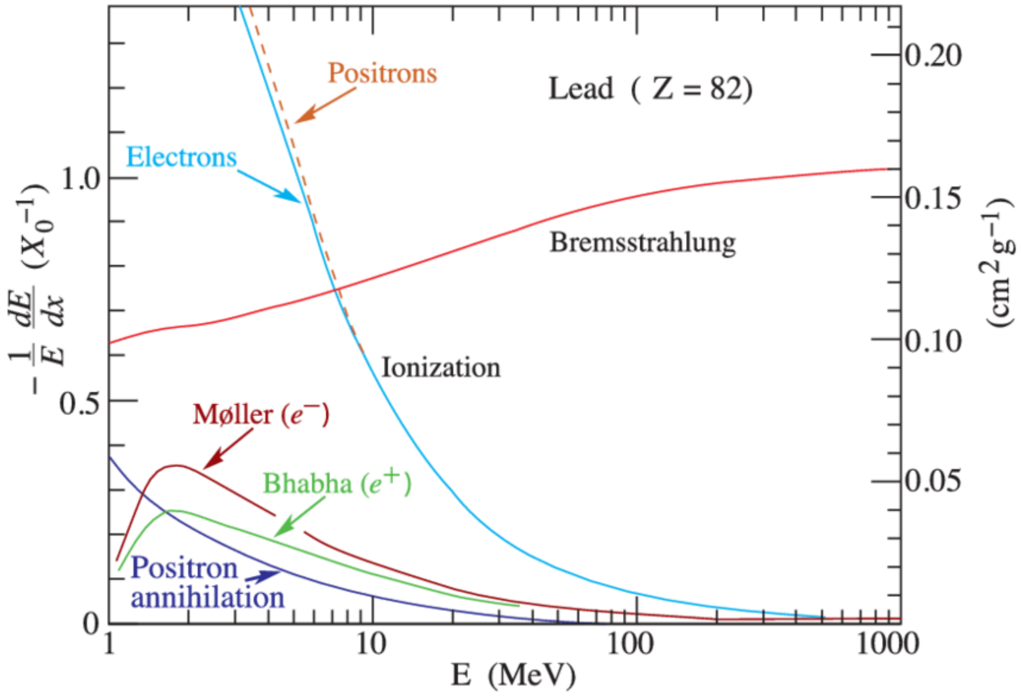


Figure 3.1: Fractional energy loss per radiation length in lead as a function of electron or positron energy. Figure from [73].

- **Compton Scattering:** becomes the primary interaction mechanism at intermediate photon energies (around 1 MeV for lead). It is an inelastic scattering process in which the photon collides with a bound electron. The cross-section scales approximately with the number of atomic electrons (atomic number Z), since each electron can participate in the scattering process.

- **Rayleigh scattering:** an elastic scattering process, is a subdominant contribution at lower energies and its cross-section falls rapidly with increasing photon energy.

- **Pair Production:** this process dominates at high photon energies (above roughly 10 MeV in lead). In this process, the photon is converted into an electron-positron pair in the electromagnetic field of a nucleus of the matter or an electron. Pair production can only occur if the photon energy exceeds 1.022 MeV, which corresponds to the sum of electron and positron rest mass.

3.1.3 Electromagnetic Showers

When high-energy electrons, positrons, or photons enter a dense material, they can initiate a chain reaction of particle interactions known as an electromagnetic cascade. An electron can emit bremsstrahlung photons, which, if sufficiently energetic, can undergo pair production, creating a new electron and positron. These particles can then emit additional bremsstrahlung photons, and the process continues, forming a growing avalanche of particles. This results in an electromagnetic shower, an example of which is shown in Fig. 3.3.

For electrons and positrons, the characteristic length scale of bremsstrahlung interactions, X_0, as shown in Fig. 3.3, can be approximated as follows:

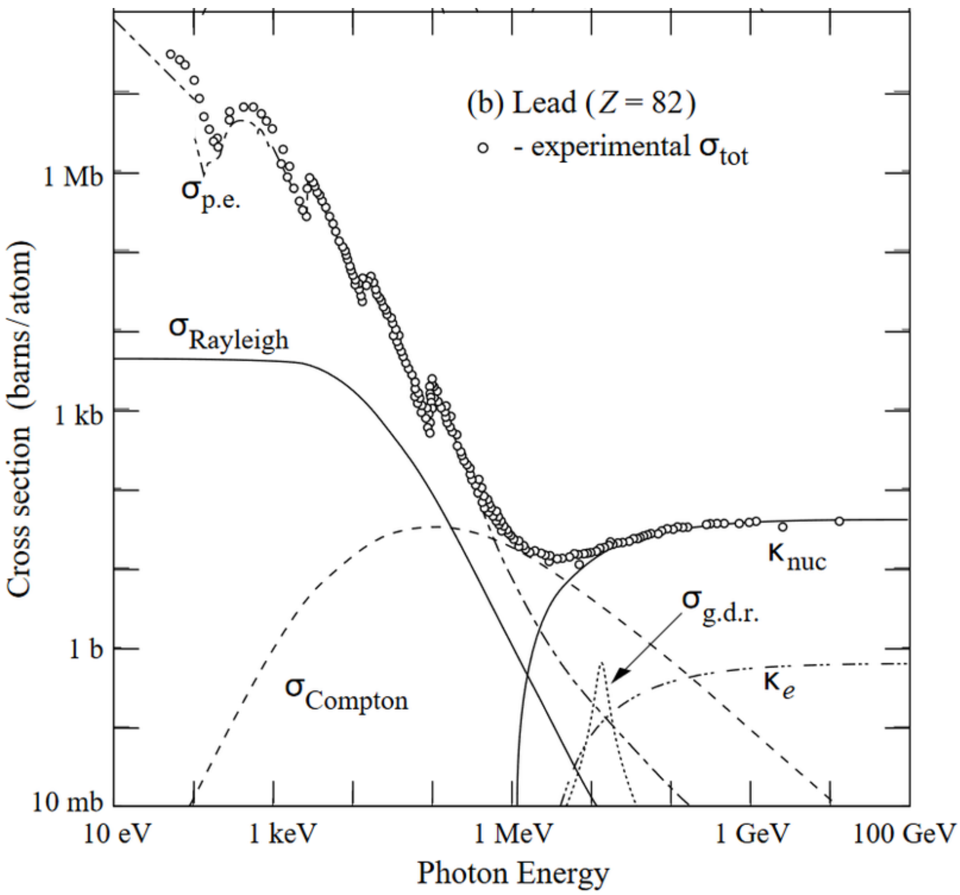


Figure 3.2: Cross sections of photon interactions with lead as function of photon energy. The total cross section obtained from experiment is shown, together with the contributions from individual interaction processes. The contributions from the individual processes are denoted as follows: $\sigma_{\text{p.e.}}$ for the photoelectric effect, σ_{Rayleigh} for Rayleigh scattering, σ_{Compton} for Compton scattering, κ_{nuc} and κ_e for pair production in the electric field of a nucleus and an electron, respectively, and $\sigma_{\text{g.d.r.}}$ for photonuclear interactions. Figure from [73].

$$X_0 \approx \frac{716\,A}{Z(Z+1)\ln(287/\sqrt{Z})}\ [\mathrm{g\,cm^{-2}}], \tag{3.1}$$

where Z is the atomic number and A is the atomic mass of the material in which the shower evolves [76]. This formula is valid for single-element materials and does not account for composite materials.

Physically, X_0 represents the average distance over which an electron or positron loses a fraction $1/e$ of its energy through bremsstrahlung. For photons, a similar characteristic length scale governs pair production, with the photon intensity reduced to $1/e$ of its initial value after traveling a distance of $\frac{9}{7}X_0$ [76].

This common scale means that showers produced by electrons, positrons, and photons are similar in extent. The main differences are that photon-initiated showers tend to start slightly deeper in the material, as the first pair-production interaction occurs after a stochastic conversion distance, while an electron continues propagating after a bremsstrahlung emission and can emit another photon [77]. Denser materials have shorter radiation lengths, resulting in more compact showers.

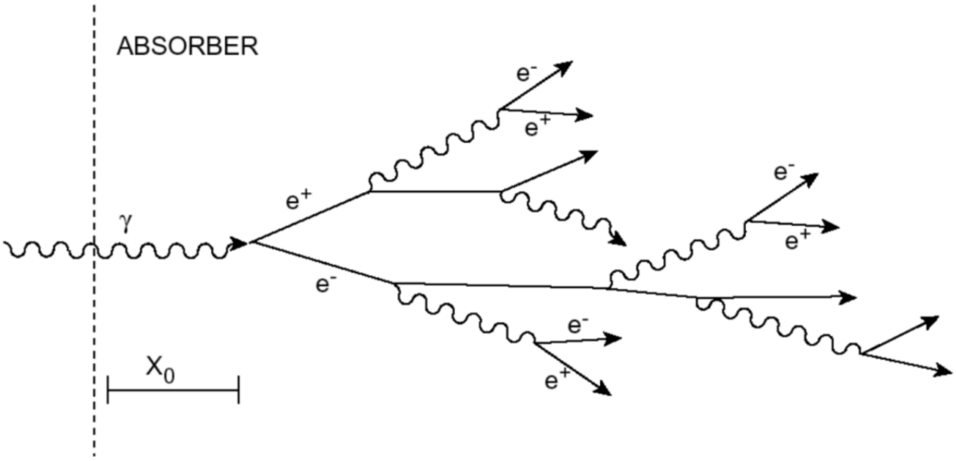


Figure 3.3: Illustration of the propagation process of an electromagnetic shower initiated by a photon. Figure from [75].

As an electromagnetic shower develops, particle-producing interactions increase the number of particles and at the same time their energy progressively reduces. A key parameter describing this behaviour is the critical energy E_c, defined as the energy at which the energy loss due to bremsstrahlung equals that due to ionisation for electrons or positrons [77]. An empirical approximation for E_c, is given by [72]:

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

where Z is the atomic number of the matter in which the shower is propagating.

The longitudinal development of the shower can be characterised using the depth t, measured in units of radiation length ($t = x/X_0$). The maximum shower depth is reached when the average particle energy falls below E_c. For an incident particle energy E_0, the position of the shower maximum is approximately [78]:

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

The transverse size of the shower is described by the Molière radius R_M, an approximate measure that corresponds to ~90% containment of the shower energy [77]. It is given by the empirical relation [76]:

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

3.2 Heavy Charged Particles Interaction with Matter

Due to their larger masses, heavy charged particles have other dominant interaction modes than light charged leptons and photons. Bremsstrahlung is proportional to the inverse fourth power of the mass of the particle, thus it becomes non-prominent for such particles as muons and charged hadrons. In contrast, energy loss through ionisation constitutes the primary interaction mechanism for these particles as they traverse matter. The mean energy loss per unit path length (or a mass stopping power) due to ionisation is described by the Bethe-Bloch equation [77]:

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

where γ and β represent the Lorentz factor and the velocity of the particle, respectively, m_e is the electron mass, and I is the mean excitation energy of the material. The quantities Z and A are the atomic number and atomic mass of the absorber, through which the charged particle is traveling, while z is the charge of the incident particle. The constant K sets the overall scale of the energy loss, and $T_{\max}$ is the maximum kinetic energy that can be transferred to an electron in a single collision. The term involving δ accounts for medium polarisation effects, which reduce the logarithmic increase of the ionisation energy loss at high particle momenta [73].

The ionisation energy loss, as given by equation 3.5, for positively charged muons traversing copper is illustrated in Fig. 3.4. The energy loss reaches a minimum at values of $\beta\gamma \approx$ 3-4, defining the region of minimum ionisation. Owing to the slow increase of ionisation losses with increasing momentum beyond this point, this minimum extends over a broad momentum range of several GeV. Charged particles with momenta in this region are referred to as Minimum Ionising Particles (MIPs), although the exact momentum range depends on the particle species.

MIPs are important for the calorimeter calibration, as they provide a stable reference for the energy loss. Although the overall shape of the energy-loss distribution may vary, the location of its peak, corresponding to the most probable energy loss, remains stable. For MIPs, the energy transferred in a single interaction is small, allowing the particle to experience many interactions as it traverses the detector material. This results in a large number of energy depositions clustered around the characteristic MIP value. This characteristic energy loss is highly reproducible for a given detector and material, thus it constitutes a well-defined reference point.

3.2.1 Hadronic Showers

Hadronic showers are phenomena that are analogous to electromagnetic showers (see Sec. 3.1.3) but typically exhibit more complex and variable topologies. In addition to electromagnetic interactions, hadrons also undergo strong interactions, which play a significant role in the development of the shower (see Fig. 3.5).

The development of a hadronic shower involves several distinct interaction mechanisms and energy-loss processes, which together lead to its complex and highly variable structure. The main contributions can be summarised as follows.

- First, charged hadrons traversing matter continuously lose energy through **ionisation**, in the same manner as described for other charged particles. While this process is always present, it typically represents only a small fraction of the total energy deposition in a hadronic shower.

- Second, the shower is initiated by a **hard inelastic interaction** between the incoming hadron and the nuclei of the absorber material. Such hadronic interactions give rise to a variety of nuclear processes, including spallation, fission, and nuclear evaporation. These reactions can produce secondary hadrons, most notably pions and η mesons, as well as liberated nucleons and heavier nuclear fragments [80]. All of these products may

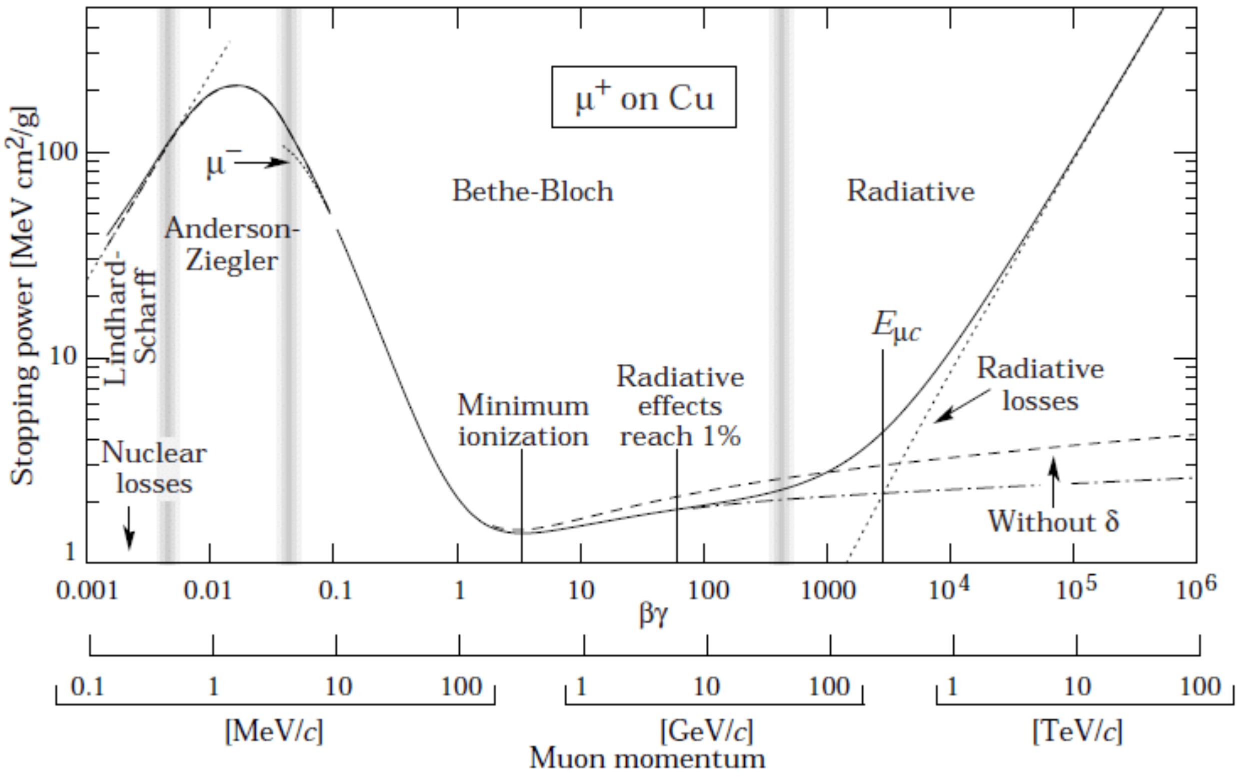


Figure 3.4: Mass stopping power for the positively charged muons in copper as a function of their momenta. Figure from [79].

subsequently undergo further interactions with the surrounding material, contributing to the continued development of the shower.

- Third, unstable particles produced in the shower can **decay**. In particular, neutral pions and η mesons predominantly decay into two photons [73], thereby feeding energy into an electromagnetic sub-shower. As a result, hadronic showers contain a significant **electromagnetic component**. Although the size of this component fluctuates from event to event, on average roughly one third of the mesons created in the initial interaction decay electromagnetically [80].

- Finally, a portion of the shower energy is not directly observable and is referred to as **invisible energy**. This includes energy consumed in nuclear binding and excitation, energy transferred to recoiling nuclei, and energy carried by neutrons that may travel considerable distances before being captured, potentially outside the instrumented detector volume. In addition, decays of charged hadrons can produce neutrinos, which escape the detector without depositing measurable energy.

The quantity λ_I in Fig. 3.5 is called the *nuclear interaction length* and is the mean free path between hadronic interactions. It is expressed as follows:

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

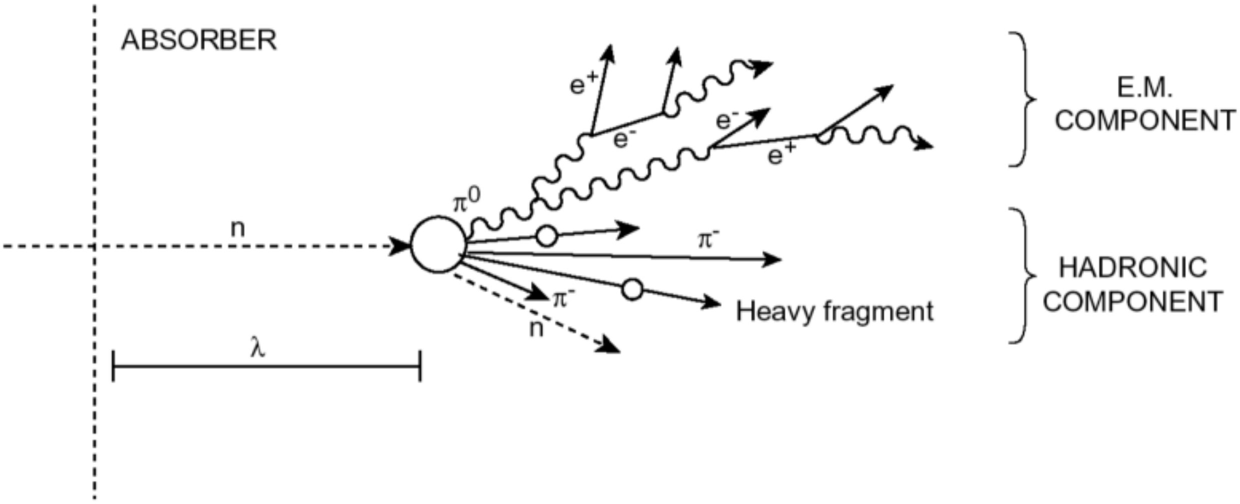


Figure 3.5: Illustration of the propagation process of a hadronic shower initiated by a neutron. The shower consists of hadronic and electromagnetic components. Figure from [75]

where σ_{had} represents the cross section of hadronic interactions, while n denotes the number of atoms per unit volume in the material with which the particle is interacting [78]. By expressing the interaction length λ_I in units of $\mathrm{g\,cm^{-2}}$, thus removing the effects of the material's density, it becomes clear that λ_I approximately scales as $A^{\frac{1}{3}}$ [77].

In conclusion, for most materials, the nuclear interaction length is significantly larger than the radiation length (see Sec. 3.1.3). Moreover, hadronic showers exhibit much more variable topologies due to the greater number of possible interactions. As a result, hadronic showers propagate over longer distances and produce a more diverse set of final-state particles compared to electromagnetic showers.

3.3 Typical Calorimeter Designs

As discussed in the preceding sections, electromagnetic and hadronic showers differ in their nature, topology, and characteristics. Consequently, their detection and measurement are generally treated separately and divided between two calorimeter parts: *electromagnetic calorimeter* (ECAL) and *hadronic calorimeter* (HCAL).

The ECAL has a finite hadronic interaction length (see Section 3.2.1), so some hadrons can begin their showers there as well. Both ECAL and HCAL detectors are built with dense absorbing materials to encourage frequent particle interactions, ensuring that showers are contained within a compact volume. For the ECAL, the choice and amount of material is carefully balanced: it must provide enough radiation lengths to effectively measure electromagnetic showers while keeping the hadronic interaction length low (see Sec. 3.2.1).

One of the main characteristics of calorimeters is their energy resolution. It is expressed as follows [76]:

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

Here, σ_E is the uncertainty on the energy measurement in the calorimeter and E_{mes} is the measured energy of the primary particle.

Equation 3.7 has multiple components. The first component, which contains a, is the stochastic term. The $\frac{1}{\sqrt{E}}$ dependence comes from assuming Poisson statistics, where the number of particles N produced is roughly proportional to the energy E of the incoming particle [77]. The second component, containing b, is the noise term. This arises from electronic noise and does not depend on the particle's energy. Finally, the constant c contributes a fixed amount, independent of energy. It accounts for intrinsic detector limitations, such as calibration errors or material inconsistencies.

Calorimeter designs can be divided into two main types:

- **Homogeneous calorimeters:** The entire detector volume is sensitive and also acts as the absorber. It can be, for example, a noble gas (producing Cherenkov light in the interaction with the incident particles) or a scintillator material. They provide excellent energy resolution, but are expensive and difficult to make highly granular or segmented [76].

- **Sampling calorimeters:** The detector is built from alternating layers of sensitive and passive material. This means that such detectors are by nature longitudinally segmented. A characteristic quantity of a sampling calorimeter is the so-called sampling fraction f_s, which is the ratio of the visible energy deposited in the sensitive layers to the total energy deposited in both the sensitive and passive materials of the calorimeter. Sampling calorimeters are more compact and cost-efficient, but have lower energy resolution due to fluctuations in energy deposition [76].

3.4 Particle Flow Reconstruction in Calorimeters

To study the physical processes occurring after a collision, accurate particle reconstruction algorithms must be employed to identify both the particles within a jet and the primary particle that produced them. The particle flow reconstruction algorithm is a method used in modern calorimetry to improve the measurement of particle energies. It combines information from tracking detectors and calorimeters to reconstruct individual particles within a jet, aiming to achieve the best possible energy resolution. By identifying charged and neutral components separately, particle flow minimizes the impact of detector limitations and provides a more accurate picture of the event.

Traditionally, particle energies are reconstructed solely from calorimeter measurements, which include the ECAL and HCAL components (see Sec. 3.3). The reconstructed energy can therefore be expressed as follows:

$$E_{rec} = E_{ECAL} + E_{HCAL}, \tag{3.8}$$

where E_{ECAL} denotes the energy measured in the ECAL and E_{HCAL} the energy measured in the HCAL. In the case of a lepton collider, such as LEP [81], jets are composed predominantly of hadrons, with approximately 60% charged particles (mostly hadrons), 30% photons, and 10% neutral hadrons. As a result, the dominant contribution to the jet energy measurement originates from the HCAL. Since hadronic showers are governed by complex physical processes, the HCAL has a relatively poor energy resolution, typically of the order of $\gtrsim 55\%/\sqrt{E(\mathrm{GeV})}$ [82]. As hadrons account for roughly 70% of the jet energy, this significantly limits the achievable jet energy resolution when relying solely on the calorimeter measurements.

In contrast, the *particle flow reconstruction algorithm* (PFA) takes advantage of the detector systems with better resolution to measure energies of the jets (see Fig. 3.6). The energy in this case can be represented as follows:

$$E_{PF} = E_{p^\pm} + E_\gamma + E_{h^0}. \tag{3.9}$$

Here, the quantity $E_{p^\pm}$ is obtained from the momentum measurement of charged particles in the tracking detector, while E_γ and E_{h^0} correspond to the calorimetric energy measurements of photons and neutral hadrons, respectively. This approach significantly reduces the reliance on the HCAL, as only about 10% of the jet constituents are measured there. However, there are also challenges in this approach. The main challenge of the particle flow algorithm is the reconstruction of the four-momenta of all individual particles in an event, which requires assigning each detector energy deposition to its originating particle. This demands accurate clustering of calorimeter hits, clear separation of overlapping showers and correct association of clusters with reconstructed tracks, made possible only by sophisticated algorithms and highly granular calorimeters. Reconstruction confusion is a major source of degraded jet energy resolution and can arise from incorrect hit-to-cluster or track-to-cluster associations. For example, energy is lost if a neutral particle is merged with a nearby charged hadron and its energy is discarded in favour of the track measurement, while energy is double counted if part of a charged particle shower is misidentified as a separate neutral particle [82].

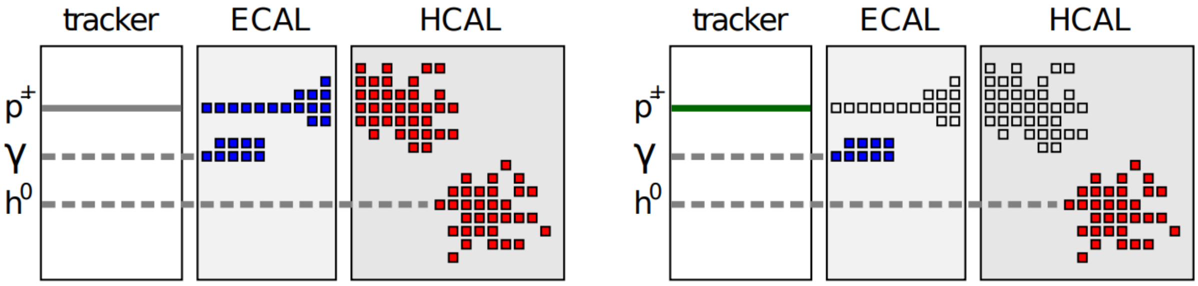


Figure 3.6: Schematic comparison between conventional calorimetry (left) and particle flow calorimetry (right). In the conventional approach, particle energies are determined solely from their energy deposits in the calorimeter, whereas in particle flow reconstruction the momenta of charged particles are measured using the curvature of their trajectories in the tracking system. Figure from [83].

The current state-of-the-art particle flow algorithm is PandoraPFA [85, 86], which is a comprehensive framework for reconstructing all particle types in an event, including charged hadrons, photons, and neutral hadrons. Particularly relevant to this thesis are the photon reconstruction components, consisting of five main algorithms. The primary photon reconstruction algorithm clusters ECAL hits into photon candidates and validates their identity, applying additional checks when photons are found near charged particles. Three fragment removal algorithms then operate in sequence to eliminate spurious neutral fragments. The first targets fragments created at the edges of electromagnetic showers after the energetic core has been identified as a photon. The second addresses fragments arising from energy leakage, where high-energy photon showers extend beyond the ECAL into the HCAL and are incorrectly reconstructed as neutral hadrons. The third further suppresses residual photon-related fragments that result from incomplete or incorrect shower containment. Finally, a photon splitting algorithm separates photons that were incorrectly merged during fragment removal,

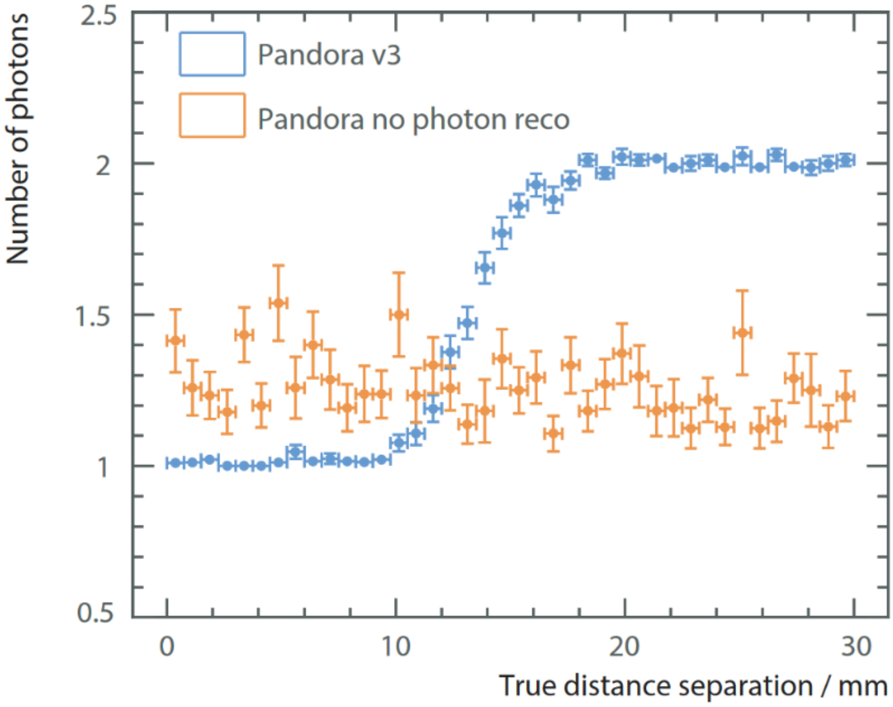


Figure 3.7: Average number of reconstructed photons as a function of the true separation between two photons. The sample consists of photon pairs with significantly different energies, 500 GeV and 50 GeV, shown without (orange) and with (blue) the dedicated photon reconstruction algorithms. Figure from [84].

which can occur when photons are spatially close to one another.

The impact of these dedicated photon reconstruction algorithms is illustrated in Fig. 3.7, which shows the average number of reconstructed photons as a function of the true separation between two incident photons with energies of 500 GeV and 50 GeV. Without the photon reconstruction algorithms, Pandora fails to correctly identify both photons across the full separation range, consistently reconstructing around 1.2–1.4 photons regardless of their actual separation. In contrast, when the photon reconstruction algorithms are enabled, Pandora correctly identifies both photons once their separation exceeds approximately 12 mm, achieving a reconstruction efficiency close to 2 photons for separations above 15 mm. This demonstrates the critical importance of sophisticated photon reconstruction for resolving closely spaced electromagnetic showers.

4 The International Large Detector Concept

The International Large Detector (ILD) [42, 59] is a proposed general-purpose detector designed for the International Linear Collider (ILC). ILD aims to provide precise measurements of particle interactions, leveraging advanced tracking systems, highly granular calorimetry, and a powerful magnetic field to achieve exceptional momentum and energy resolutions. The three dimentional model of ILD is shown in Fig. 4.1. The length of the ILD is 13 m and the total radius is 7.8 m.

Its design is optimized for particle flow reconstruction algorithms (see Sec. 3.4). The main goal of the design was to reach less than a 4% jet energy resolution for the energy ranges of the jets of 40 – 250 GeV, motivated by the separation between Z and W hadronic decay channels.

This chapter presents a description of the ILD detector systems and its software framework.

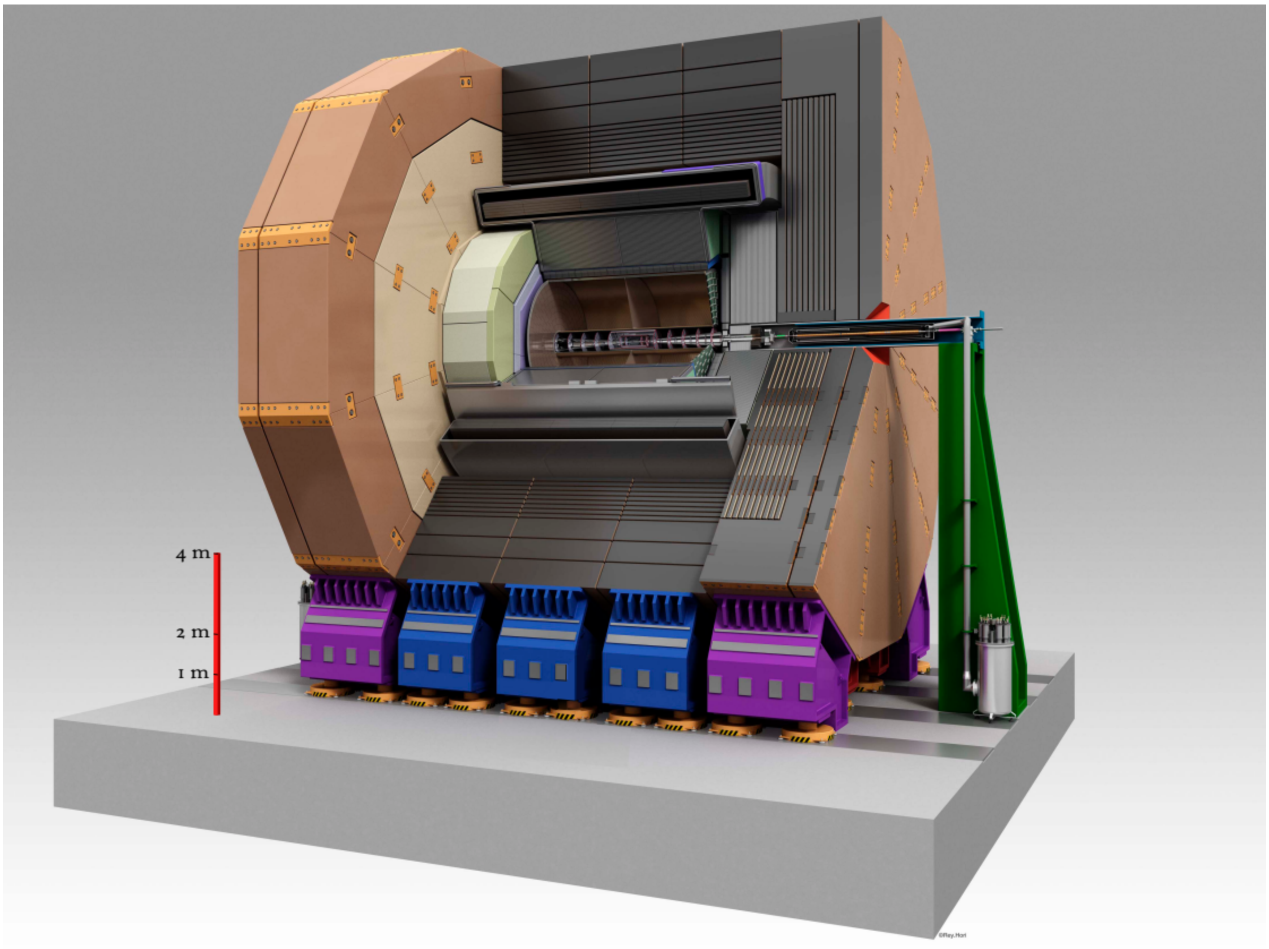


Figure 4.1: Three dimentional model of the International Large Detector. Figure from [87].

4.1 Detector Systems

The main detector components of the ILD are shown in Fig. 4.2. This section provides an overview of each of them.

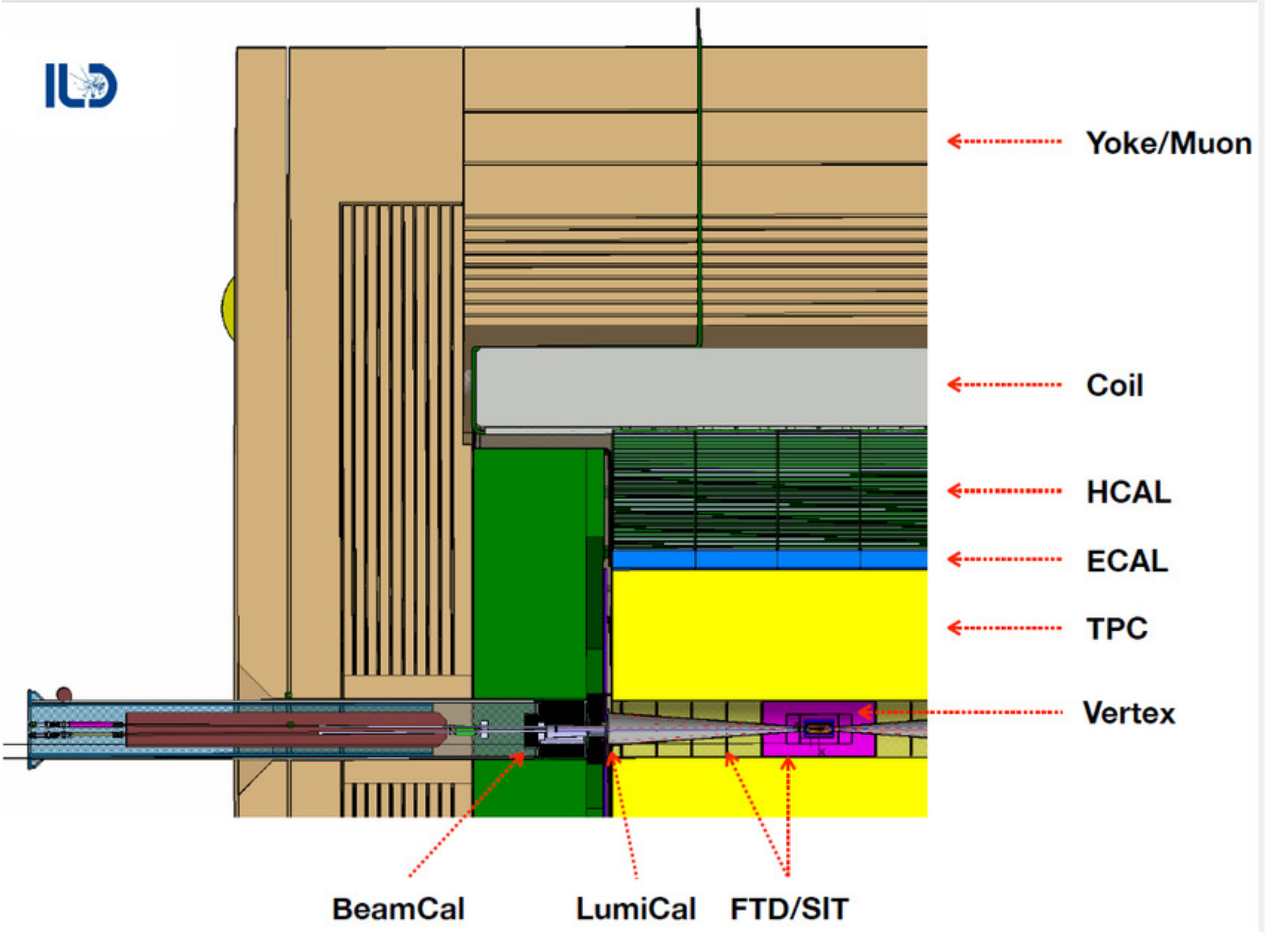


Figure 4.2: Single quadrant view of the ILD detector with all of the main detector systems highlighted. Figure from [88].

4.1.1 Tracking Detector

The part of the ILD, closest to the collision point, is the tracking system. It consists of the following main components (see Fig. 4.2):

- **Vertex Detector (VTX):** The innermost component, composed of a barrel of three double-layers of silicon pixel sensors. The first layer is located 16 mm from the interaction point. The detector is optimized for a point resolution better than 3 μm and a material thickness of less than $\sim 0.15\% \, X_0$ per layer.
- **Silicon Inner Tracker (SIT) and Forward Tracking Detector (FTD):** Additional silicon layers surrounding the VTX. The SIT has two barrel layers, while the FTD consists of seven forward disks, extending tracking coverage to low polar angles.
- **Time Projection Chamber (TPC):** The main tracking facility surrounding the silicon trackers. It has a length of about 4.6 m and a radial span from approximately 33 cm to

180 cm and if filled with a low-mass, fast-drifting gas mixture. Ionization electrons produced in the gas volume are collected using an electric field and read out with either Gas Electron Multipliers (GEMs) [89] or Micro-Mesh Gaseous Structures (Micromegas) [90] with pad sizes of a few mm^2. The TPC provides continuous tracking with up to 220 three-dimensional space points per track and enables particle identification via specific energy loss (dE/dx) measurements, while keeping the material budget low. Additional silicon layers placed after the TPC and before the calorimeter system, improving momentum resolution and track linking.

The full tracking system achieves a momentum resolution of approximately
2×10^{-5} GeV^{-1} [42].

4.1.2 Calorimeter

The tracking system is followed by the calorimeter, which consists of an ECAL and HCAL (see sec. 3.3).

Electromagnetic Calorimeter

The electromagnetic calorimeter foreseen for the ILD detector is realized as a sampling calorimeter employing tungsten as the absorber material. It currently has two alternative design concepts.

One concept, commonly referred to as the SiW ECAL [91], utilizes silicon-based sensors as the active detection medium. The second concept, known as the Sc ECAL [92], relies on scintillator strips to provide the active response.

This thesis focuses on the SiW ECAL implementation. This configuration is a sequence of thirty silicon sensor layers interleaved with tungsten absorber plates. To optimize performance, two absorber thicknesses are employed: the first twenty tungsten layers have a thickness of 2.1 mm, while the remaining ten layers are twice as thick, at 4.2 mm. When combined, the full absorber stack corresponds to a total depth of 24 radiation lengths. The sensitive elements consist of silicon wafers with a thickness of 0.525 mm, segmented into square cells measuring 5×5 mm^2. Inevitably, non-sensitive regions arise at wafer boundaries, primarily due to mechanical constraints and the presence of readout electronics.

From a mechanical perspective, the fundamental unit of the ECAL is the module (see Fig. 4.3). Each module contains a tungsten absorber assembly supported by carbon-fibre structures. An alveolar architecture is adopted, in which alternating tungsten plates form a set of cavities. Within each cavity, referred to as an alveolus, a slab is inserted that comprises a single tungsten absorber layer positioned between two silicon sensor layers. This arrangement leads to a repetitive absorber–sensor pattern along the longitudinal direction of the calorimeter. Owing to the trapezoidal geometry of the modules, successive layers are laterally offset, producing a staggered configuration. This feature reduces the likelihood that particles traverse inactive edge regions in multiple consecutive layers.

The complete ECAL system is organized into several large-scale components: an octagonal barrel section, two endcaps, and two additional ring structures that fill the regions surrounding the beam pipe openings in the endcaps (see Fig. 4.4).

The barrel and endcaps together constitute the primary detector regions considered in this study. The barrel section is formed from eight longitudinal staves, each of which is assembled from five trapezoidal modules. The endcaps are composed of twelve modules arranged into four quadrants. Although the endcap modules share the same fundamental construction

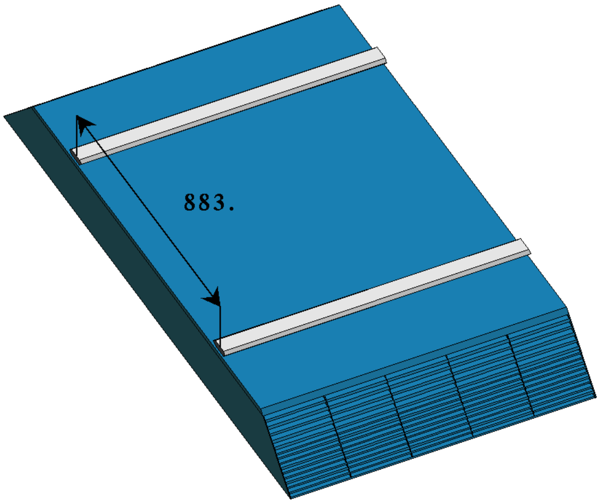


Figure 4.3: One ILD ECAL module, consisting of a carbon-fibre-supported stack of alternating tungsten absorber layers, with tungsten slabs between silicon sensor layers inserted into the intervening gaps. Figure from [93].

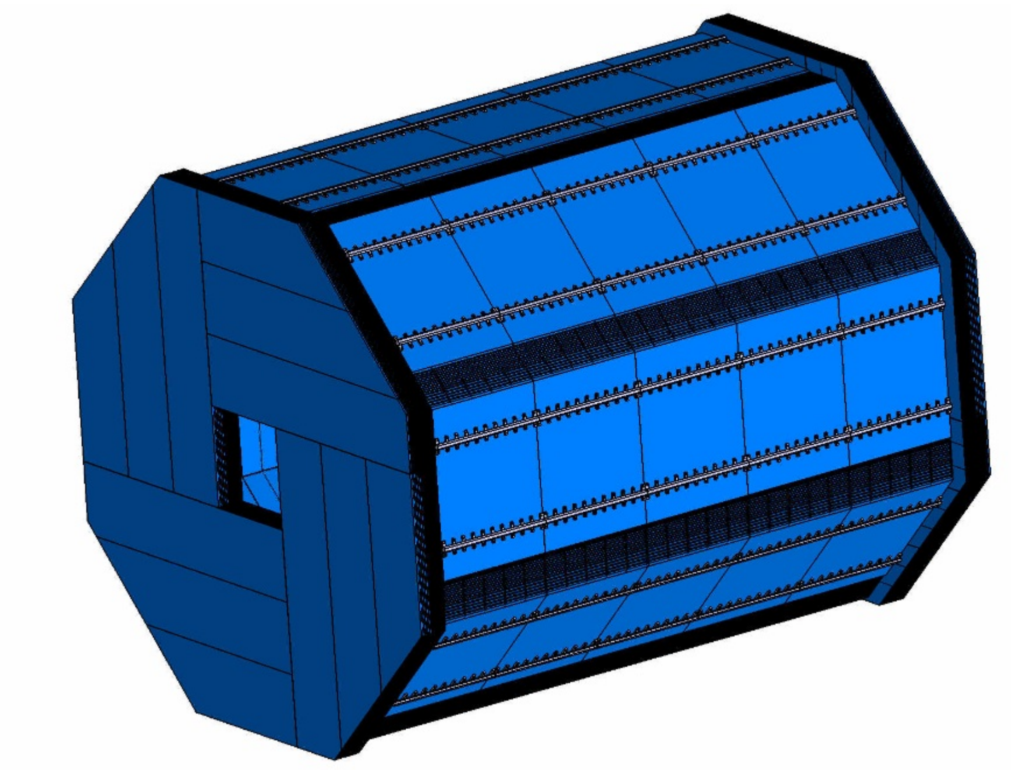


Figure 4.4: Geometrical model of the octagonal ECAL barrel and two endcaps. Figure from [93].

principles as those in the barrel, their geometrical shapes differ to accommodate the detector layout.

Within the ILD simulation framework, the ECAL is modeled using a highly detailed geometrical description implemented through the DD4HEP detector description toolkit (will be described further in this chapter). This implementation incorporates the full modular structure as well as a realistic representation of the materials present, ensuring an accurate simulation of the detector response.

Hadronic Calorimeter

Two alternative concepts have been proposed for the ILD hadronic calorimeter: a semi-digital option (SDHCAL) [94] and an analogue option (AHCAL) [95]. This thesis focuses on the analogue hadron calorimeter. It is composed of 48 layers of stainless-steel absorber plates, each with a thickness of 17.2 mm, giving a total depth of approximately $5\,\lambda_\mathrm{I}$. The active layers consist of scintillator tiles of size 3×3 cm^2 and a thickness of 3 mm, with each tile read out individually by a silicon photomultiplier.

A detailed detector description, including geometry and material composition, is implemented in the DD4HEP simulation framework.

Very Forward Detectors

A set of specialized detectors is to be installed in the very forward region of the ILD to ensure precise measurements in areas not fully covered by the central detector systems. This includes the LumiCal, a calorimeter with a keystone geometry optimized for high-precision luminosity measurements via the detection of Bhabha scattering electron-positron pairs. Complementing this, the LHCAL provides additional hadronic calorimeter coverage in the forward region, while a dedicated forward calorimeter system extends the calorimetric acceptance down to polar angles as small as 6 mrad and enables bunch-by-bunch monitoring of the beam conditions. These forward detectors leverage technologies closely related to those employed in the electromagnetic calorimeter, with specific adaptations to withstand the unique radiation environment and high occupancy rates characteristic of the forward region. This combination of precision instrumentation ensures both accurate luminosity determination and comprehensive calorimetric coverage in the forward direction [42].

4.1.3 Solenoid Magnet

The tracking and calorimeter systems are enclosed within a *superconducting solenoid coil*, which generates a magnetic field of 3.5 T oriented along the beam axis. This strong, uniform magnetic field bends the trajectories of charged particles, allowing precise momentum measurements in the central tracking detectors. Outside the solenoid, the iron return yoke completes the magnetic circuit and supports additional detector systems.

4.2 Software Ecosystem

The software ecosystem of the ILD is based on the ILCSOFT [96]. It is a comprehensive software framework developed for the simulation, reconstruction, and analysis of events at the ILC and similar future lepton colliders. It provides a common environment for all aspects of detector development, enabling realistic performance studies and comparison of different detector concepts.

The main software frameworks used in ILD studies include:

- LCIO [97]: Provides the event data model and manages data storage.
- DD4HEP [98]: A toolkit for building a realistic detector model, including its materials, structure, and readout geometry. DDG4 [99] is integrated within DD4HEP, this module connects the detector model to GEANT4 for detailed particle simulation. DDREC [100] is also a part of DD4HEP, it provides the interface between the detector geometry and reconstruction algorithms.
- MARLIN [101]: A general application framework used for both reconstruction and physics analysis of simulated events.

4.2.1 Digitisation and Calibration

Simulated detector hits are first converted into digitised signals, emulating the response of the sensors and readout electronics. For the calorimeters, a two-step calibration is applied:

1. Conversion of the deposited energy in each cell into MIP units (see the definition in Sec. 3.2), based on the most probable energy deposited by a minimum ionizing particle.
2. Conversion of energy in MIPs to absolute energy units (GeV), so that the measured energy corresponds to the particle's incident energy.

Additional corrections are applied in regions of the calorimeter with gaps or insensitive volumes, compensating for reduced sampling fractions.

4.2.2 Reconstruction

The reconstruction stage combines tracking and particle flow to build a complete picture of the event:

- **Tracking:** Charged particle trajectories are reconstructed using specialized algorithms, including dedicated procedures for kinks (from bremsstrahlung or scattering) and secondary vertices (V0s), such as photon conversions to e^+e^- pairs.
- **Particle Flow Reconstruction:** For this, the PANDORAPFA algorithm is employed. Tracks, calorimeter hits, kinks, and V0s are clustered and associated, with particle identities assigned. The clusters are rather splitted initially, then wrongly merged. Re-clustering is applied on the later steps [85]. Software compensation [102] corrects for differences in response between electromagnetic and hadronic showers. The output is organised as a collection of *Particle Flow Objects* (PFOs), which include energy, momentum, and particle type.
- **High-Level Reconstruction:** Includes primary and secondary vertex finding, jet clustering and tagging (LCFIPLUS [103]), particle identification (PID) [104] using dE/dx, shower shapes, or time-of-flight (TOF) information [105], and reconstruction of neutral mesons (π^0, η) via $\gamma\gamma$-finders and constrained kinematic fits (MARLINKINFIT [106]).

5 Simulation and Computational Approaches in High-Energy Physics

Simulation plays a central role in modern high-energy physics, serving as the primary interface between theoretical predictions and experimental observations. Detailed simulations of particle collisions, detector responses, and reconstruction algorithms are essential for interpreting experimental data, optimizing detector designs, and searching for new physics phenomena. However, the computational demands of traditional simulation methods have become increasingly challenging, particularly as experiments move toward higher luminosities and more complex detector systems. This chapter provides an overview of simulation methodologies in high-energy physics, with emphasis on the computational challenges that motivate the development of new fast simulation techniques.

5.1 Monte Carlo Simulations in High-Energy Physics

The term ”Monte Carlo” refers to computational algorithms that rely on repeated random sampling to obtain numerical results. In particle physics, Monte Carlo methods are used to simulate stochastic processes – from the fundamental quantum mechanical nature of particle interactions to the statistical fluctuations in detector responses.

The Monte Carlo approach to particle simulation follows a basic procedure. It starsts with a given initial conditions (particle type, energy, direction), tracks the particle through its trajectory by sampling interaction points according to cross-sections and mean free paths, at each interaction point sample the type of interaction and its outcome (energy transfer, secondary particle production, etc.), repeat for all secondary particles until all particles fall below energy thresholds or exit the detector volume, and finally record relevant quantities such as energy depositions and their positions.

By repeating this procedure many times, the simulation builds up a statistical representation of the underlying physical processes. The accuracy of Monte Carlo simulation improves with the square root of the number of samples, following the standard statistical scaling where the uncertainty scales as $1/\sqrt{N}$ with N being the number of simulated events.

Simulations in high-energy physics serves multiple essential purposes across the entire lifecycle of an experiment. During the design phase, simulations enable the optimization of detector geometries and the evaluation of different technological choices. Simulations are also used to develop and validate reconstruction algorithms, estimate selection efficiencies, and model backgrounds. Finally, for physics analysis, simulations provide the Monte Carlo samples necessary to compare theoretical predictions with experimental measurements and to extract physics results.

The simulation chain for a typical collider experiment consists of several stages. In event generation, the hard scattering process and subsequent parton shower and hadronization are simulated using specialized Monte Carlo event generators, commonly implemented in tools such as PYTHIA [107] or WHIZARD [108]. Detector simulation then propagates particles through the detector geometry and simulates their interactions with detector materials and fields using GEANT4 [18, 20]. This is followed by digitization, which converts energy depositions into detector hits accounting for electronic noise and inefficiencies, and finally reconstruction applies pattern recognition and clustering algorithms to reconstruct particles from the detector hits.

Among these stages, detector simulation typically dominates the computational cost, particularly for calorimeter systems where electromagnetic and hadronic showers generate large numbers of secondary particles.

5.2 Geant4: The Standard Tool for Detector Simulation

The GEANT4 [18, 20] (GEometry ANd Tracking, version 4) is the most widely used toolkit for simulating the passage of particles through matter in high-energy physics. Originally developed for high-energy physics applications, it has since been adopted in nuclear physics, accelerator physics, and medical physics. GEANT4 is implemented in C++ and provides a comprehensive framework for detailed geometric modeling of complex detector systems, particle tracking through arbitrary geometries with automatic handling of boundary crossings, a comprehensive library of physics processes covering electromagnetic, hadronic, and optical interactions, flexible output and analysis interfaces, and support for parallel processing and distributed computing.

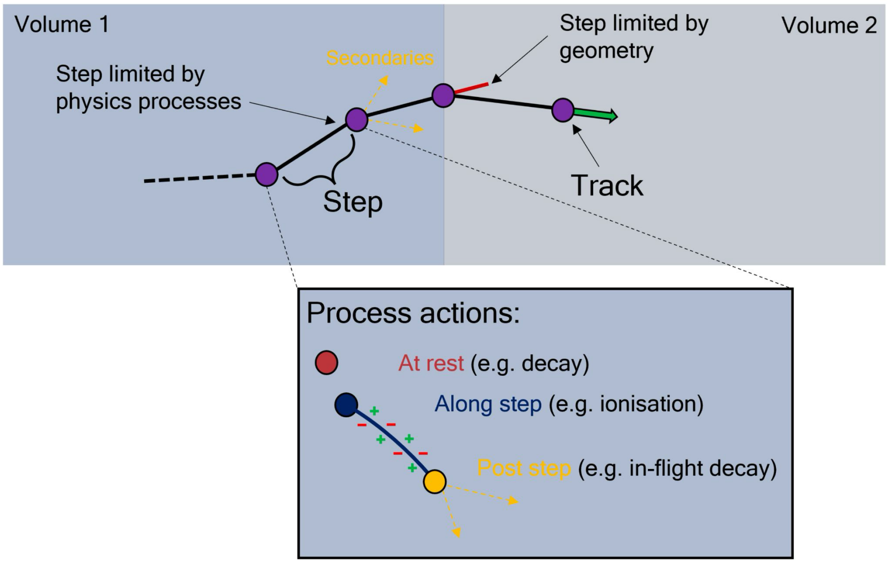


Figure 5.1: Illustration of the procedure of tracking in GEANT4. A particle is moved one step at a time, with the track representing a snapshot of the particle state at each step. Steps are limited in size by a random competition between processes, to allow for the production of secondaries or if a boundary between geometry volumes is crossed. A process may be defined as a combination of three actions: at rest, along step and post step. Figure from [41].

The fundamental unit of simulation in GEANT4 is the "step" – the smallest segment of a particle's path through material, as illustrated in Fig. 5.1. Each step represents a portion of the particle trajectory during which the particle's state (position, energy, direction) is updated based on physical processes. During each step, GEANT4 updates the particle's position along its trajectory, checks for boundary crossings between different materials or detector volumes, simulates continuous energy loss processes (ionization, multiple scattering), samples discrete interactions (bremsstrahlung, pair production, nuclear interactions), creates and tracks secondary particles produced in interactions, and records energy deposited in sensitive detector regions.

The step size is determined dynamically based on multiple factors: the material properties, the particle energy, the proximity to volume boundaries, and physics process characteristics. A single high-energy electron traversing a calorimeter might undergo hundreds or thousands of steps before being absorbed or falling below the tracking threshold. For a single 50 GeV photon shower in the ILD electromagnetic calorimeter, approximately 20,000 individual GEANT4 steps are recorded on average in the sensitive detector layers, which constitute only a few percent of the total detector material (see Chapter 7).

5.3 Computational Challenges of Full Simulation

The computational demands of full Monte Carlo simulation have grown to represent a major challenge for high-energy physics experiments. As shown in Chapter 1, the ATLAS experiment devoted more than one third of its total CPU resources to Monte Carlo simulation during 2018 [25]. This fraction is expected to increase substantially for future high-luminosity experiments.

Several factors contribute to the computational intensity of full detector simulation. A single high-energy particle entering a calorimeter can initiate a cascade producing thousands of secondary particles. Each of these particles must be tracked individually through potentially hundreds of steps. As discussed in Section 3.1.3, electromagnetic showers develop through the interplay of bremsstrahlung and pair production processes, continuing to multiply until particle energies fall below the critical energy where ionization losses begin to dominate over radiative processes.

The accuracy of GEANT4 simulation arises from its detailed step-by-step tracking of every particle. At each step, the simulation must query the geometry to determine the current material, calculate all possible interaction processes and their probabilities, sample interaction outcomes using random number generators, update particle states and create secondary particles, and check for volume boundaries and tracking cuts. This granular approach is essential for physical accuracy but becomes computationally expensive when multiplied by large shower multiplicities. Moreover, the full simulation procedure relies on the sequential production of secondaries, as each interaction must be simulated before its daughter particles can be tracked, making it difficult to parallelize efficiently within a single shower.

Hadronic showers are significantly more complex than electromagnetic showers, involving a much broader range of physics processes including strong interactions producing multiple secondary hadrons, neutral and charged pion production with $\pi^0 \to \gamma\gamma$ decays feeding electromagnetic sub-showers, nuclear reactions, long-lived particles requiring tracking over extended distances, and neutron transport with complex energy-dependent cross-sections. The stochastic nature of hadronic interactions leads to large shower-to-shower fluctuations, requiring detailed modeling of each process.

Modern and future calorimeters feature increasingly fine segmentation to enable advanced reconstruction techniques such as Particle Flow algorithms. The ILD electromagnetic calorimeter uses 5×5 mm^2 cells (see Chapter 4), the CMS High Granularity Calorimeter employs cells as small as 0.52 cm^2 [109]. Fine granularity increases computational cost through more frequent volume boundary crossings requiring geometry queries, larger numbers of hits to be created and processed, increased memory requirements for storing hit collections, and more complex reconstruction requiring evaluation of spatial relationships between many hits.

Future high-luminosity experiments further compound these computational challenges. The HL-LHC [14] aims to collect 3000 fb^{-1} of integrated luminosity, roughly an order of magnitude more data than the original LHC design. Higher luminosity at hadron colliders leads to increased pileup, with multiple interactions occurring in the same or nearby bunch crossings, dramatically increasing the particle multiplicity per recorded event. Proposed future lepton colliders such as the ILC [15], the FCC-ee [17], and CLIC [16] will produce even larger event samples while operating in cleaner environments with minimal pileup. However, the computational burden remains substantial due to the large number of events required for precision measurements and the complexity of highly granular detector designs. Additionally, precision measurements require simulated samples significantly larger than the collected data to minimize statistical uncertainties.

As illustrated in Fig. 1.2 in Chapter 1, both ATLAS and CMS project that their computing needs for the HL-LHC era will substantially exceed available resources [35, 36].

5.4 Fast Simulation Approaches

To address these computational challenges, the high-energy physics community has developed various fast simulation techniques that trade some level of physical detail for substantial speedup.

Parametric fast simulation techniques model calorimeter responses using parameterized functions rather than detailed particle tracking. These approaches work by characterizing average shower shapes through fits to full simulation, parameterizing observables (longitudinal and lateral profiles, energy resolution, etc.) as functions of particle type, energy, and incident angle, and sampling from these parameterizations when fast simulation is invoked. Examples include FastCaloSim [110] used by ATLAS, fast shower parameterizations [21, 22] based on analytical functions describing shower development, and frozen showers [24] which store and reuse pre-simulated showers.

Parametric approaches can achieve much faster inference time compared to full simulation. However, they have significant limitations that make them increasingly unsuitable for future experiments. Most produce calorimeter responses at the cluster level rather than individual hits, making them incompatible with reconstruction algorithms that rely on hit-level information. Modern reconstruction approaches, particularly Particle Flow algorithms [111], fundamentally rely on the fine granularity of calorimeters to separate energy depositions from different particles in dense environments, requiring access to individual calorimeter hits with their precise spatial positions and timing information. The move toward highly granular calorimeters with millions of readout channels, such as the ILD ECAL with approximately 80 million individual cells, amplifies this fundamental incompatibility. Additionally, being based on average shower profiles and simplified analytical functions, parametric methods cannot fully reproduce the shower-to-shower variations observed in real particle showers.

5.5 Machine Learning for Fast Simulation

Recent advances in deep generative modeling offer a new approach for fast simulation that addresses many limitations of traditional approaches. Machine learning models can learn complex, high-dimensional distributions directly from training data and generate new samples more efficiently.

The application of generative models to calorimeter simulation has been explored through various architectures including Generative Adversarial Networks (GANs) [40, 112–118], Variational Autoencoders (VAEs) [5, 40, 115, 119], Normalizing Flows [120–124], and Diffusion Models [1–3, 6, 125–127].

The key advantages of generative ML models for calorimeter simulation include direct learning from data without requiring manual parameterization, hit-level generation compatible with actual reconstruction chains, ability to model complex multi-dimensional correlations, fast inference particularly when executed on GPUs, and flexibility to adapt to different detector geometries with fine-tuning. However, machine learning approaches also face challenges including the need for large training samples of full simulation as ground truth, extensive benchmarking requirements to validate physics accuracy, careful integration into experiment software frameworks, and ongoing research into understanding and quantifying systematic uncertainties from ML modeling.

Most generative models for calorimeter simulation rely on fixed-grid (voxelized) representations, mapping detector hits onto regular 3D tensors for training with standard deep learning architectures. However, this approach encounters fundamental limitations. The high sparsity of shower data in highly granular calorimeters creates large, mostly empty tensors that are computationally inefficient to process, and the fixed-grid mapping entangles shower physics with position-specific geometrical features (such as gaps between cells and staggering patterns), significantly limiting the model's ability to generalize to different impact positions without retraining. Chapter 6 introduces the machine learning concepts and techniques underlying modern generative models, followed by Chapter 7 which presents CaloClouds, a point cloud-based diffusion model that addresses these limitations by representing showers as sets of energy-weighted points in continuous space, enabling both computational efficiency and geometry independence in the transverse plane.

6 Machine Learning

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that focuses on the development of algorithms capable of learning from examples to solve problems, without being explicitly programmed. The ability to learn from examples is the key feature of ML algorithms, which makes them very powerful in solving problems that are difficult to describe with a formal set of rules [128].

Driven by recent advancements in specialized hardware that enable extensive parallelization of computational operations, machine learning has revolutionized high energy physics, and computer science in general, allowing efficient analysis of large datasets generated by experiments like those at LHC. In this rapidly evolving field, machine learning is used for a wide variety of tasks ranging from real-time analysis [129, 130] and anomaly detection [131–134] to performing particle tagging [135–139] and reconstruction [140, 141] using machine learning models.

The present work explores the use of generative ML to accelerate particle shower simulations in high-granularity calorimeters, addressing the computational bottlenecks outlined earlier in Section 5, for future experiments in high energy physics.

This chapter is organized into several sections to provide a basic understanding of machine learning principles and an overview of different generative model architectures. Section 6.1 introduces the foundational concepts and basics of machine learning, describing its general principles and techniques. Section 6.2 introduces generative machine learning, with subsections dedicated to specific architectures: Generative Adversarial Networks 6.2.1, Variational Autoencoders 6.2.2, Normalizing Flows 6.2.3, and Diffusion Models 6.2.4, outlining benefits and use cases of each approach.

6.1 Basics of Machine Learning

Machine learning as a field is very broad, with a wide range of algorithms and techniques. Although it is typically divided into three main categories: supervised learning, unsupervised learning and reinforcement learning, each solving different types of problems [142].

Supervised learning refers to machine learning algorithms that are trained using labeled datasets, where each input example is associated with a corresponding output. The goal is to learn a function that accurately maps inputs to outputs for new, previously unseen inputs. A simple example of supervised learning could be a particle identification task, where the input data is the detector response from a particle interaction, and the output is the type of particle responsible for that interaction.

Unsupervised learning on the other hand is a type of ML where the algorithm is trained to learn the underlying patterns or distributions in data without the use of labeled examples. An example of unsupervised learning could be an anomaly detection task, where the goal is to identify the data points that do not fit the data distribution on which the algorithm has been trained.

Reinforcement learning is a type of ML where the algorithm learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. An example of reinforcement learning could be a beam control system, where the goal is to maximize the luminosity of the beam by adjusting the beam parameters.

Despite differences among these three subfields, they all share a common training technique involving iterative gradient-based optimization procedure aimed at minimizing or maximizing a **loss function** – sometimes called a **cost function** or **criterion** [128]. A loss function quantifies how accurately a model fits the training data. This self-optimization process is precisely what classifies these algorithms as machine learning. Even a simple linear regression can be considered a form of machine learning if its solution is found through an automated iterative optimization, e.g. **Gradient Descent**.

6.1.1 Gradient Descent

The most common optimization technique used in machine learning is gradient descent. It is used in the vast majority of ML algorithms except for rule-based learning algorithms such as classical Decision Trees (DT) or K-Nearest Neighbors (KNN). Gradient descent minimizes the loss function by iteratively adjusting the model parameters. This procedure involves calculating the gradient, which provides the direction of steepest ascent for the loss function, and then moving in the opposite direction to minimize the error [128].

To describe gradient descent, consider a simple linear regression model, where the goal is to find a linear function that predicts a target value based on the input, previously unseen feature value. Suppose that a training set is given that contains input features x and output targets y. Then the model can be described as:

$$f_{w,b}(x) = wx + b, \tag{6.1}$$

where w and b are parameters of the model, also referred to as weights, that are adjusted during the optimization process. In order to find the best w and b parameters, the loss function needs to be defined. The most common loss function for linear regression is Mean Squared Error (MSE) and it can be defined as follows:

$$L_{w,b} = \frac{1}{m}\sum_{i=1}^{m}(f_{w,b}(x_i) - y_i)^2, \tag{6.2}$$

where y_i is the target value, $f_{w,b}(x_i)$ is the model's prediction, sometimes denoted as $\hat{y}_i$, and m is the number of examples in the training set.

Mathematically, the gradient of the loss function with respect to the parameters can be defined as:

$$\nabla_{w,b} L_{w,b} = \left[\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b}\right], \tag{6.3}$$

For the linear regression loss function using MSE, these partial derivatives are computed as:

$$\frac{\partial L}{\partial w} = \frac{2}{m}\sum_{i=1}^{m}(f_{w,b}(x_i) - y_i)x_i, \tag{6.4}$$

$$\frac{\partial L}{\partial b} = \frac{2}{m}\sum_{i=1}^{m}(f_{w,b}(x_i) - y_i), \tag{6.5}$$

The parameters are then updated in each iteration as follows:

$$w := w - \alpha \frac{\partial L}{\partial w}, \tag{6.6}$$

$$b := b - \alpha \frac{\partial L}{\partial b}, \tag{6.7}$$

where α is the learning rate, a hyperparameter that controls the step size at each iteration. A suitable learning rate is crucial, too large may cause overshooting of the minimum, and too small might lead to excessively slow convergence. Gradient descent continues iteratively updating the parameters until convergence is achieved or a predefined number of iterations is completed.

6.1.2 Stochastic Gradient Descent

Gradient descent is fundamental for training most machine learning models. However, a common challenge is the need for large training sets to achieve good generalization, which simultaneously increases computational demands. The computational complexity of the 6.3 operation is $O(m)$, meaning that as the training set expands to millions of examples, even a single gradient step can become computationally impractical.

The core idea behind stochastic gradient descent (SGD) is to approximate the gradient using a small subset of the data. At each iteration of the algorithm, a **minibatch** $B = \{x_1, x_2, ..., x_m\}$ is sampled randomly from the full training set. The minibatch size m is typically a small fixed number – often between 1 and a few hundred, and crucially it remains constant even as the overall training set size grows. As a result, it becomes possible to effectively train on datasets with millions of examples by computing gradient updates using only around a hundred examples at each step [128].

Beyond this fundamental method, there are advanced optimization techniques that improve convergence properties and stability of the training.

The momentum method [143] accelerates the training, especially when gradients are small, noisy, or when the optimization surface has high curvature. It works by keeping track of a weighted average of previous gradients, encouraging the updates to continue moving in the same direction.

Adaptive Gradient Algorithm (Adagrad) [144] dynamically adapts the learning rate for each parameter, scaling inversely to the accumulated squared gradients. Adagrad is particularly beneficial in sparse data scenarios but tends to excessively reduce learning rates over time.

Root Mean Square Propagation (RMSProp) [145] is a modification of Adagrad, RMSProp employs an exponentially decaying moving average to counteract Adagrad's aggressive learning rate decay, providing stable training across different parameter scales.

Adaptive Moment Estimation (Adam) [146] combines the advantages of momentum-based methods and RMSProp, maintaining an exponentially decaying moving average of past gradients and squared gradients, resulting in adaptive and efficient parameter updates. Due to its robustness and minimal need for hyperparameter tuning, Adam has been utilized in this work.

6.1.3 Artificial Neural Networks and Deep Learning

While linear models, such as those described in Section 6.1.1, can be effective for many simple tasks, they have limitations. Most famously, they are unable to learn the XOR function, where

$f_{w,b}([0,1]) = 1$ and $f_{w,b}([1,0]) = 1$, but $f_{w,b}([0,0]) = 0$ and $f_{w,b}([1,1]) = 0$. To address these limitations, artificial neural networks (ANNs) were developed. These models are inspired by the structure and function of biological neurons in the brain, and designed to approximate arbitrary functions via learned transformations.

The fundamental computational unit in a neural network is the artificial neuron, also known as a perceptron. A neuron receives a set of input features $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, applies a weighted sum, adds a bias, and then transforms this value using a non-linear function called an **activation function**:

$$y = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{6.8}$$

where $\mathbf{w} = \{w_1, ..., w_n\}$ are set of learnable weights, b is a learnable bias term, and $\sigma(\cdot)$ is the activation function.

The introduction of non-linearity σ is very important. Without it, no matter how many layers are stacked, the composition of linear functions is still a linear function, so the entire model would still represent the equivalent of a single linear transformation. Non-linear activations such as those presented in Fig. 6.1, ReLU (Rectified Linear Unit), tanh, and sigmoid, allow neural networks to introduce non-linear mappings from inputs to outputs.

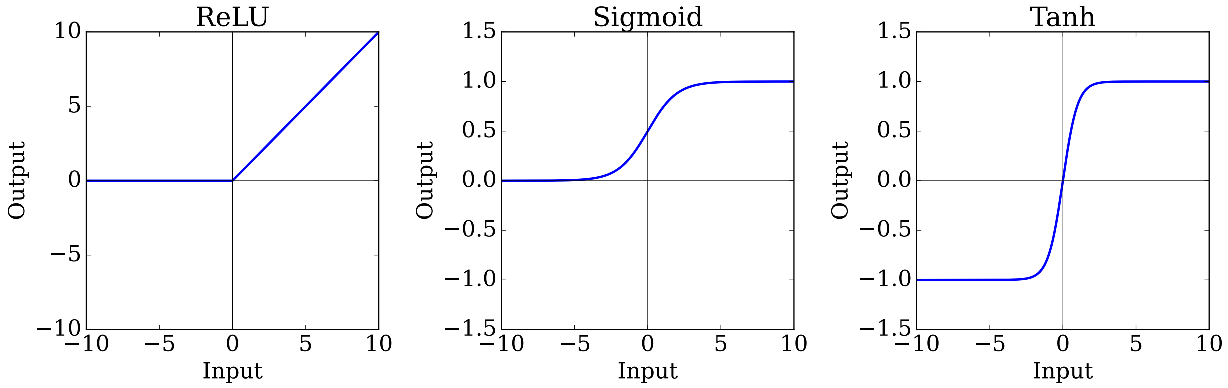


Figure 6.1: Illustration of the most common activation functions. The ReLU function (left) is defined as $\sigma(x) = \max(0, x)$, the sigmoid function (center) is defined as $\sigma(x) = \frac{1}{1+e^{-x}}$, and the tanh function (right) is defined as $\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

However, a single neuron is still limited in its expressiveness. To model more complex functions, neurons are connected in layers, where the output of one layer serves as input to the next. This is usually called a **feedforward neural network** or **multi-layer perceptron** (MLP); it consists of an input layer, multiple hidden layers, and an output layer. When such networks consist of many hidden layers, they are often referred to as **deep neural networks** (DNNs). A simple example of a feedforward neural network is shown in Fig. 6.2.

Over the years, many architectural components and design innovations have been developed to improve performance and address specific tasks, such as convolutional layers for image processing [147], recurrent layers that are used for sequential data processing [148], or attention mechanisms for natural language processing [149]. Additionally, many different techniques have been developed to increase training speed and stability: normalization techniques, residual connections, and many others [128].

While a comprehensive review of all these components is beyond the scope of this work, it is important to recognize that the underlying principle remains the same: a neural network is a

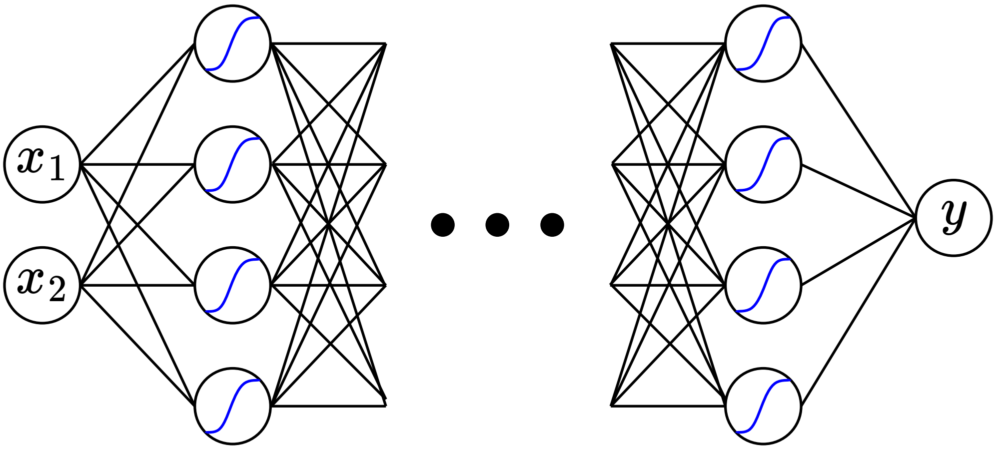


Figure 6.2: Illustration of a simple feedforward neural network. The input layer consists of 2 features x_1 and x_2. Each line represents a connection between two neurons, with the weights w_i associated with each connection. Each neuron aggregates the inputs from the previous layer via summation and applies a non-linear activation function $\sigma(\cdot)$. The final neuron y usually does not apply any activation function, and represents the output of the model.

sequence of transformations applied to the input, with non-linearities applied between them, trained to minimize the cost function. As the field continues to evolve, new architectures and training methods emerge, yet this key idea remains the foundation upon which deep learning is built.

6.1.4 Backpropagation

In order to train a neural network with Gradient Descent 6.1.1, the gradients of the loss function with respect to the model parameters need to be computed. The standard method for training is **backpropagation**, an algorithm that efficiently computes gradients of the loss function with respect to each parameter in the network.

Backpropagation is a special case of the chain rule from calculus. Since a neural network is essentially a nested composition of functions, the derivative of the loss L with respect to any parameter w (e.g., a weight in one of the layers) can be expressed as:

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w}, \tag{6.9}$$

where y is an intermediate output affected by w. By applying the chain rule recursively from the output layer back to the input layer, the gradients for all parameters in the network can be computed.

6.2 Generative Machine Learning

Generative machine learning involves the use of generative models that learn to produce new data samples that resemble the training data. To be able to draw a new sample, models have to learn the underlying distribution of the training data. Practically, this is achieved by training a model to perform a mapping from a simple prior distribution (e.g., Gaussian or uniform) to the data space, such that new data samples can be generated by sampling from the prior distribution and passing the sampled noise through the trained model, as can be seen in the diagram in Fig. 6.3. In other words, a generative model is a function that transforms a noise into some structure.

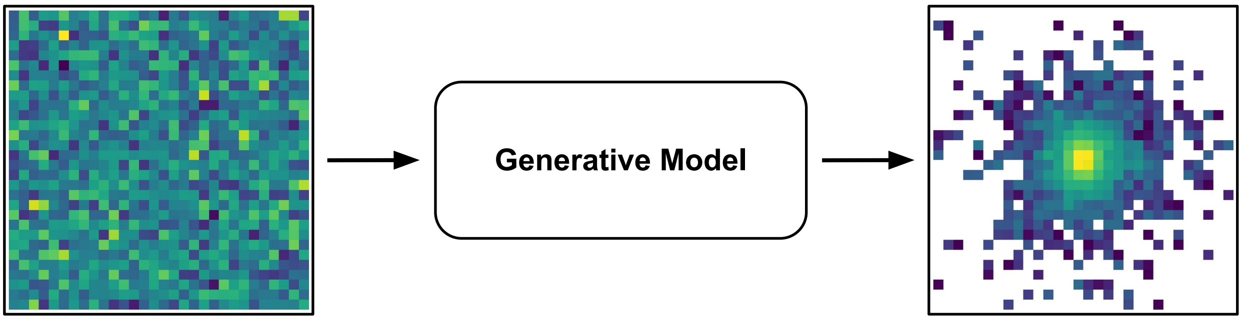


Figure 6.3: Illustration of a generative model mapping. A sample drawn from a simple Gaussian prior distribution (left) is transformed by the generative model into a structured output representing a simulated calorimeter shower (right). The shower was generated using the CaloClouds III model.

There are many different generative models that have been developed over the years, each achieving this transformation in a different way, offering different benefits and drawbacks. In the following sections, the most common types of generative models are discussed. While the generative modeling field is rapidly evolving, and newer models generally differ from the formalisms presented here, the key conceptual ideas remain the same.

6.2.1 Generative Adversarial Networks

The most iconic generative models, Generative Adversarial Networks (GANs), were introduced by Goodfellow et al. in Ref. [150]; they offer an elegant approach in training a generative model by framing the problem as a game between two neural networks, a generator G and a discriminator D. The generator generates samples from a random noise vector z drawn from a simple prior distribution, while the discriminator attempts to distinguish between real samples x from the training data and generated samples $\tilde{x}$ produced by the generator. The discriminator is trained to maximize the probability of correctly classifying real and generated samples, while the generator is trained to minimize the probability of the discriminator correctly classifying the generated samples. This can be represented by the following loss function:

$$L_{D,G} = \min_G \max_D \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))], \tag{6.10}$$

where $p_{data}(x)$ is the distribution of the training data and $p_z(z)$ is the distribution of the noise vector z.

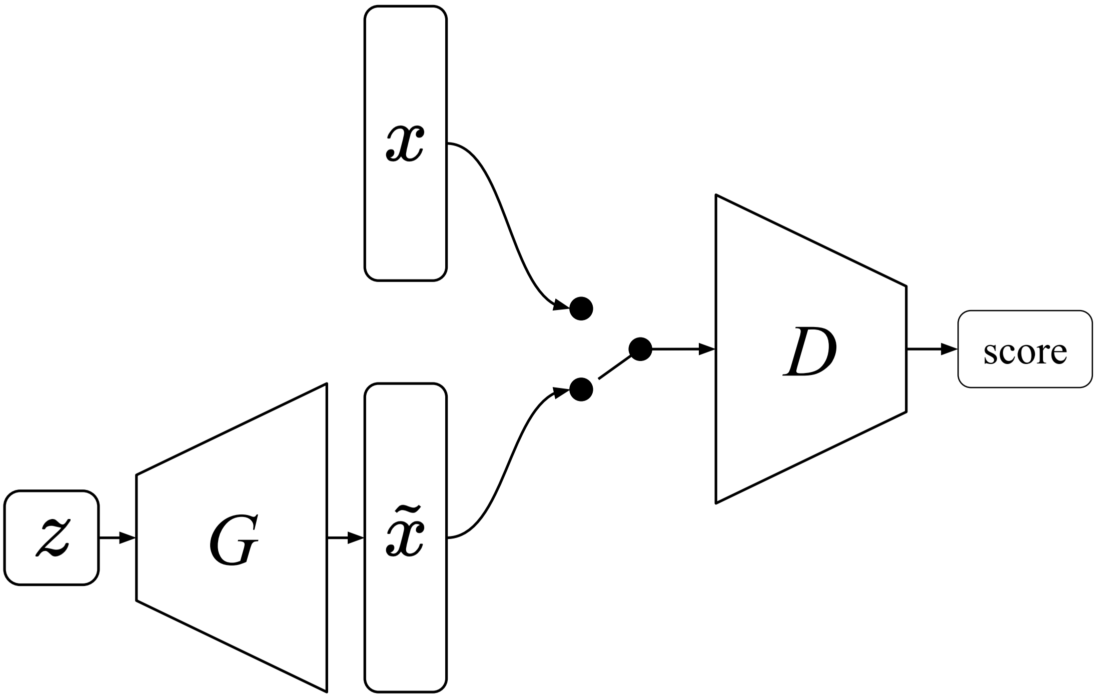


Figure 6.4: Overview of the GAN architecture. The generator G generates samples from a random noise vector z, while the discriminator D attempts to distinguish between real samples x and generated samples $\tilde{x}$. The generator and discriminator are trained in an adversarial manner, with the generator weights optimized to minimize the probability of the discriminator correctly classifying the generated samples, and the discriminator weights optimized to maximize the probability of correctly classifying betwen real and generated samples.

A simple diagram of the GAN architecture is shown in Fig. 6.4. Theoretically, in this adversarial min-max game, the generator and discriminator are trained until they reach a Nash Equilibrium, where the generator produces samples that are indistinguishable from the real samples and the discriminator is unable to classify the samples correctly. In reality, however, this is very difficult to achieve, the training process is very unstable and often leads to mode collapse, when the generator produces a limited variations of samples, they may look close to real samples but do not cover the full distribution of the training data.

6.2.2 Variational Autoencoders

Variational Autoencoders (VAEs) [151] are based on and derived from the autoencoder architecture, which is a type of neural network with the primary goal of learning a compressed representation of the input data. An autoencoder consists of two main components, an encoder E and a decoder D. The encoder maps the input data x to a lower-dimensional latent space z, while the decoder reconstructs the latent representation back to the original data space $\tilde{x}$.

The VAE extends the autoencoder architecture by introducing a probabilistic approach to the latent space representation. This is achieved by parameterizing the encoder as a neural network that outputs the mean μ and variance σ^2 of the latent distribution, which is usually set to be an uncorrelated multidimensional Gaussian distribution. The overall structure of the VAE architecture is shown on Fig. 6.5.

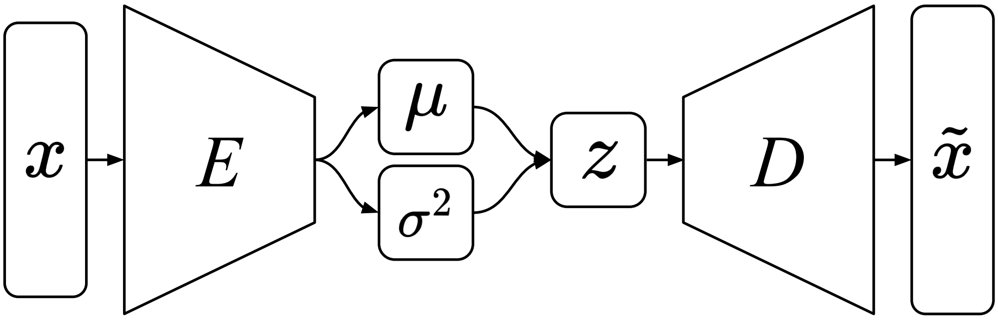


Figure 6.5: Overview of the VAE architecture. The encoder E maps the input data x to a lower-dimensional latent space z, while the decoder D reconstructs the latent representation back to the original data space $\tilde{x}$. The latent space is parameterized as a Gaussian distribution with mean μ and variance σ^2, which is used to sample new latent vectors.

By putting a constraint on the latent space to be a known distribution, new latent vectors can be easily sampled and new samples can be generated by passing them through the decoder. This constraint is usually enforced by minimizing the Kullback-Leibler divergence (KLD) between the approximated posterior distribution $q(z|x)$ and the prior distribution $p(z)$:

$$D_{KL}(q(z|x)||p(z)) = \sum_x \sum_z q(z|x) \log \frac{q(z|x)}{p(z)}, \tag{6.11}$$

Combined with the reconstruction loss, which is usually set to be the MSE between the input and reconstructed data, the total VAE loss function can be defined as:

$$L_{\text{VAE}} = \mathbb{E}[(x - \tilde{x})^2] + \mathbb{E}[D_{KL}(q(z|x)||p(z))], \tag{6.12}$$

In order to be able to perform backpropagation through the stochastic sampling process, a reparameterization trick is used, introduced in Ref.[151], that expresses the latent space z as:

$$z = \mu + \sigma \odot \epsilon, \tag{6.13}$$

where ϵ is a random noise vector drawn from a standard normal distribution, and $\odot$ is the element-wise multiplication.

With the prior being a standard normal distribution, the total loss function can be simplified to:

$$L_{\text{VAE}} = \mathbb{E}[(x - \tilde{x})^2] + \beta \cdot \mathbb{E}[\frac{1}{2}(1 + \log(\sigma^2) - \mu^2 - \sigma^2)], \tag{6.14}$$

where β is a hyperparameter that controls the strength of the KLD term.

Overall, VAEs offer a more stable approach in training generative models in comparison to GANs, as they do not rely on the adversarial training process and use reconstruction loss to guide the training. However, VAEs are often criticized for producing “blurry” samples, as the reconstruction loss tends to average out the details in the generated samples. This requires additional loss terms in the loss function, such as GAN-like adversarial loss [152], to improve the quality of the generated samples.

6.2.3 Normalizing Flows

Normalizing Flows (NFs) [153] are a class of generative models that models complex data distributions by applying a sequence of invertible and differentiable transformations to a simple base distribution. As for GANs and VAEs the base distribution typically set to be a standard multidimensional Gaussian distribution. Such transformation results in a flexible generative model that can both generate new samples and compute exact likelihoods, which distinguishes normalizing flows from GANs and VAEs.

Given a data sample x, normalizing flows learn to perform a mapping from the data space to the latent space z by applying a series of invertible transformations f_k, referred to as the normalizing direction:

$$z = f_K \circ f_{K-1} \circ \cdots \circ f_1(x), \tag{6.15}$$

where f_k are the invertible and differentiable transformations, and K is the number of transformations. The inverse mapping, also known as **generative direction** or **sampling direction**, reconstructs a data point from the latent space z:

$$x = g_1 \circ g_2 \circ \cdots \circ g_K(z), \tag{6.16}$$

where $g_k := f_k^{-1}$ are the inverse transformations. A simple diagram of the normalizing flow architecture is shown in Fig. 6.6.

This invertibility of the flow allows for both sampling and exact computation of the likelihood of the latent space z for a given data sample x:

$$\log p_x(x) = \log p_z(z) - \sum_{k=1}^{K} \log \left| \det \left(\frac{\partial f_k}{\partial h_{k-1}} \right) \right|, \tag{6.17}$$

where $\frac{\partial f_k}{\partial h_{k-1}}$ is the Jacobian of the transformation f_k, $h_0 = x$ and $h_k = f_k(h_{k-1})$ for $k = 1, \ldots, K$, and $p_Z(z)$ is the base (prior) distribution.

Computing the Jacobian determinant for a general transformation is computationally expensive; the complexity of this operation is $O(n^3)$, where n is the number of dimensions in the data space. To be practical, the transformations f_k must be carefully designed to ensure that their Jacobian determinants are tractable, i.e., can be computed efficiently.

Many different types of transformations have been proposed in the literature, such as affine coupling layers [154–156], autoregressive flows [157], or neural ODEs [158]. The present work makes use of coupling layers to construct the normalizing flow model for modeling the per-layer quantities of the electromagnetic shower.

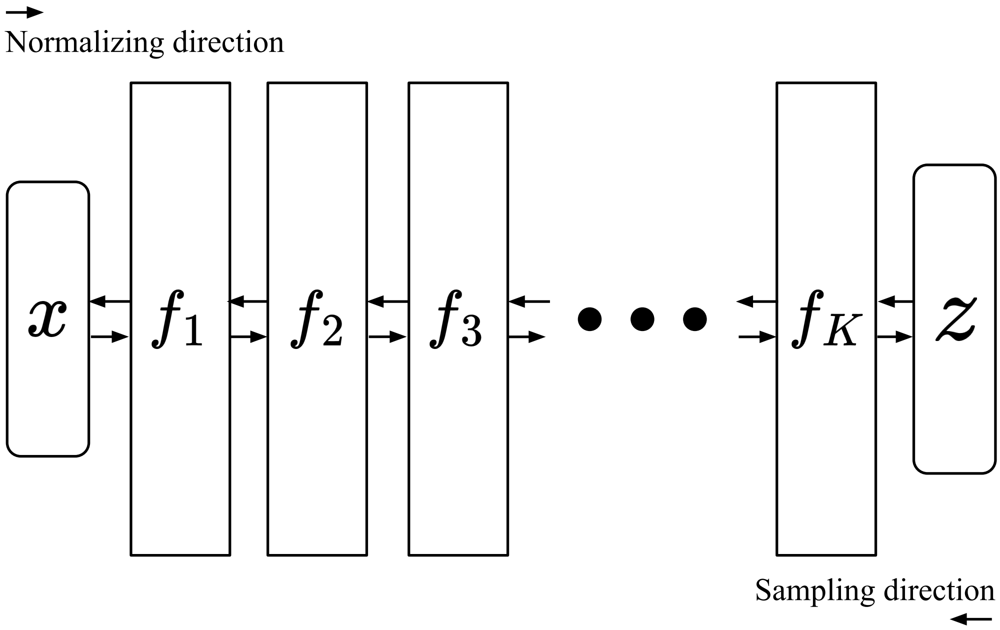


Figure 6.6: Schematic view of the normalizing flow model. The data sample x is transformed into the latent space z by applying a series of invertible transformations f_k during training (normalizing direction). The inverse mapping reconstructs a data point from the latent space z by applying the inverse transformations f_k^{-1} (sampling direction).

6.2.4 Diffusion Models

Diffusion models are a class of generative models that learn to reverse a gradual perturbation process applied to the data. The key idea involves defining a forward process that transforms structured data into noise through a sequence of steps, followed by training a model to invert this process and reconstruct the original data from noise. Once trained, the model can generate new data samples by sampling noise from a known distribution (e.g., Gaussian) and applying the learned reverse process step by step. A simple diagram of the diffusion model architecture is shown in Fig. 6.7.

Since the introduction of the diffusion model setup in Ref. [159], two main mathematical frameworks have evolved in this field: the denoising diffusion probabilistic model (DDPM) formalism [160] and the score-based formulation [161].

The denoising diffusion formalism, originally proposed in the DDPM framework [160], takes a discrete-time perspective. The forward process adds Gaussian noise to the data over T time steps according to:

$$x_t = x_{t-1}\sqrt{1-\beta_t} + \epsilon\sqrt{\beta_t}, \tag{6.18}$$

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t \mid x_{t-1}\sqrt{1-\beta_t}, \beta_t I), \tag{6.19}$$

where $\epsilon \sim \mathcal{N}(0, I)$ and β_t is a noise schedule that controls the amount of noise added at each step. After a sufficiently large number of steps, the data sample is almost indistinguishable from pure Gaussian noise. The reverse process is also assumed to follow a Gaussian distribution:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1} \mid \mu_\theta(x_t, t), \beta_t I), \tag{6.20}$$

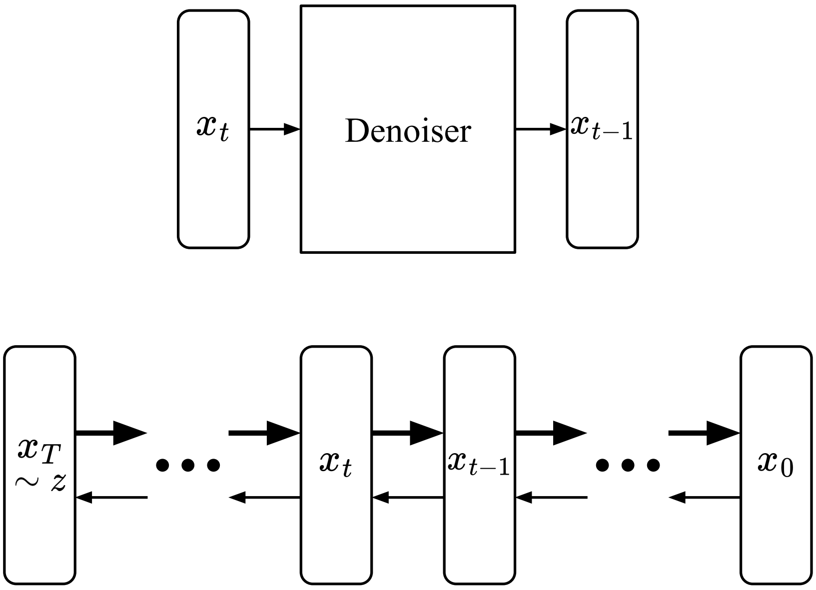


Figure 6.7: Schematic view of the diffusion model. The forward process gradually transforms data sample x_0 into nearly Gaussian noise X_T, while parameterized by the neural network "Denoiser" reverses process by reconstructing the original data from noise.

where the mean $\mu_\theta(x_t, t)$ is predicted by a neural network.

In the score-based approach, the forward process is modeled as a stochastic differential equation (SDE) of the form:

$$dx = f(x, t)dt + g(t)dW, \tag{6.21}$$

where $f(x, t)$ and $g(t)$ are predefined drift and diffusion functions, respectively, and dW represents a Wiener process (Brownian motion), with time $t \in [0, 1]$. The reverse-time SDE, used for generation, is given by:

$$dx = \left[f(x, t) - g(t)^2 \nabla_x \log p_t(x)\right] dt + g(t), dW, \tag{6.22}$$

where $\nabla_x \log p_t(x)$ is the score function, representing the gradient of the log probability density of the data at time t. The score function is approximated using a neural network, which is trained to predict the score at each time step.

Although the score-based and DDPM approaches differ in formalism and implementation, they have been shown to be mathematically equivalent under specific conditions, specifically when using variance-preserving drift and diffusion functions [161]. This equivalence means that a model trained under one formalism is also optimal under the other.

Diffusion models have gained significant attention in recent years due to their ability to generate high-quality samples with stable training across various domains. However, the iterative nature of the generation process makes them computationally expensive and requires additional techniques to speed up the sampling process, such as consistency distillation [162]. Nevertheless, they offer a flexible framework that allows to balance computational resources against generation quality.

7 CaloClouds

The next three chapters present the main achievements of this work: the development and evaluation of a point cloud generative model for fast calorimeter simulation. This chapter discusses the purpose and applications of the CaloClouds model, how it works, its architecture design choices, and the improvements made through versions 1 to 3. Chapter 8 describes the integration of generative models into the ILD software chain, and Chapter 9 presents a comprehensive performance evaluation.

The results presented in the following three chapters have been obtained in collaboration with Erik Buhmann, Thorsten Buss, Henry Day-Hall, Sascha Diefenbacher, Engin Eren, Frank Gaede, Gregor Kasieczka, William Korcari, Katja Krüger, Thomas Madlener, Peter McKeown, Martina Mozzanica, and Lorenzo Valente, and have been previously published in [1–4, 7]. Accordingly, the following three chapters include numerous figures, as well as substantial portions of text that are similar or identical to parts of those publications.

The author's contributions to the CaloClouds models are substantial and span conception, implementation, and evaluation. Specifically, the author was responsible for designing and implementing the data pipelines, including the creation, preprocessing, and postprocessing of datasets for all studies in [1–4], except for the one in [7]. The author designed and implemented the overall CaloClouds framework and carried out the training and evaluation of the original model in [1]. In [2], the author implemented and trained the ShowerFlow module of CaloClouds2. In CaloClouds3 [3], the author led the training and optimization of the diffusion model. Finally, the author performed detailed analysis and benchmarking of the models in [4].

7.1 Purpose and Applications of the CaloClouds Model

The primary goal of this work is the development and evaluation of a generative machine learning model to accelerate electromagnetic shower simulation in the ILD detector, a next-generation detector originally designed for operation at the ILC (see Chapter 4 for a detailed description of the ILD detector). The ILD detector serves as one of the primary options for an e^+e^- Higgs factory, and is optimized for the Particle Flow approach to event reconstruction (see Sec. 3.4). This approach places stringent requirements on detector hermeticity, calorimeter granularity, and the minimization of material budget ahead of the calorimeters.

This work primarily focuses on the silicon-tungsten (Si-W) electromagnetic calorimeter (ECAL) of the ILD, whose detailed specifications are provided in Chapter 4.

The CaloClouds model is a point cloud generative model developed to generate electromagnetic showers in highly granular calorimeters, with high speed and fidelity. Unlike the vast majority of generative models in this field, which rely on fixed-grid representations [7], the CaloClouds model represents showers as sets of energy-weighted points in space. This point cloud representation offers significantly better information density than voxel-based approaches, which is particularly advantageous for highly granular calorimeters where the sparsity of shower data in a fine grid creates challenges for learning. By avoiding the fixed grid structure, the model can capture shower physics at high resolution without the computational overhead of processing large, mostly empty tensors. Furthermore, the continuous spatial representation within each layer enables geometry independence in the transverse plane, allowing generated showers to be projected onto arbitrary detector positions without retraining.

The primary motivation behind CaloClouds is to alleviate the computational bottleneck of traditional full detector simulations, such as those performed with Geant4, which dominate computing budgets in current and future HEP experiments. With ever-increasing luminosity and detector complexity at experiments like the HL-LHC, ILC, or FCC, the need for scalable and adaptable fast simulation solutions has become more urgent. CaloClouds addresses this need by enabling orders-of-magnitude faster generation of calorimeter showers while preserving essential physical features.

A key feature of this model is the data preprocessing scheme that enables the use of a point cloud representation for energy depositions, as discussed in Section 7.2. This geometry-independent approach decouples the electromagnetic shower from the fixed readout layer layouts of the detector, enabling the model to be applied to any position with the same material structure. The same modeling approach can also be applied to other highly granular calorimeter systems, such as CMS-HGCAL [109], without requiring significant changes to the model architecture, though retraining or finetuning on the target detector's material structure would be necessary.

While the initial development focused on electromagnetic showers for the ECAL of the ILD detector concept, the CaloClouds paradigm has been extended in subsequent studies to simulate hadronic showers in a combined ECAL-HCAL system [6], and for different detector designs [163, 164]. This demonstrates the model's ability to adapt to various calorimeter geometries, readout schemes, and physics scenarios with minimal modifications.

All simulations are performed within the Key4hep[165] software ecosystem. Full simulation is conducted using Geant4 [166] version 11.2.2, employing the QGSP_BERT physics list, and DD4hep [167] version 1.30. A detailed and realistic model of the ILD detector geometry (version: `ILD_l5_o1_v02`), as implemented in DD4hep, is utilized for both full and fast simulation workflows.

7.2 CaloClouds Data Representation

The performance of machine learning models relies heavily on the data representation they are trained on [128]. The right data representation can significantly improve a model's performance, or in some cases, even turn an impossible problem into a trivial task. A simple example is illustrated in Figure 7.1: consider classifying points in a 2D space by drawing a straight line to separate two classes. In Cartesian coordinates (left panel), perfect separation is impossible, achieving at best approximately 65% accuracy. However, representing the same data in polar coordinates (right panel) makes the task trivial – a vertical line achieves perfect separation.

Similarly, the excellent performance of the CaloClouds model can be attributed to its data representation strategy. This section describes how calorimeter shower data is represented and preprocessed to enable geometry-independent fast simulation.

Throughout this section, shower data will be described in a local Cartesian coordinate system centred on the photon's impact point, where x and y span the transverse plane (parallel to the calorimeter layers) and z points along the depth of the calorimeter in the direction of shower propagation.

The Challenge: Detector Geometry and Shower Representation

The ILD ECAL geometry presents a challenge for machine learning approaches. While the overall material composition and longitudinal layer structure are consistent across the detec-

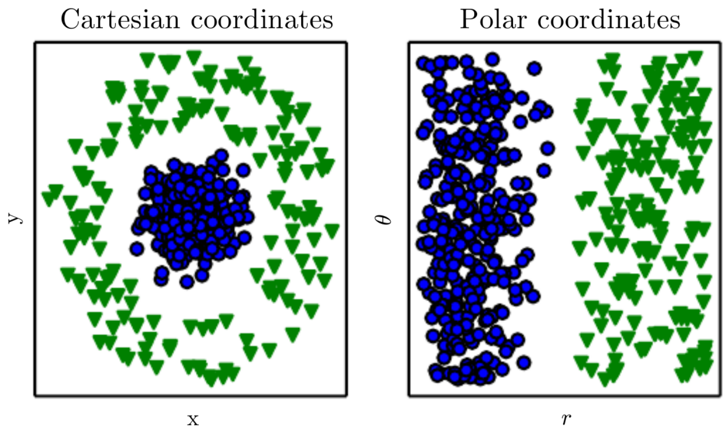


Figure 7.1: Effect of data representation on model performance. The left panel shows a dataset represented in Cartesian coordinates, where it is impossible to separate the two classes with a straight line. The right panel shows the same dataset represented in polar coordinates, where a perfect separation can be achieved with a vertical line. Figure from [128].

tor, the readout structure features irregularities: gaps between sensors, silicon wafer edges, mechanical supports, and a staggering pattern between layers. Figure 7.2 illustrates this with four 50 GeV photons fired from the interaction point – despite varying impact positions, the shower shapes appear similar due to the consistent material structure, yet the readout irregularities vary with position.

Previous studies typically mapped energy depositions from the irregular detector geometry onto a fixed, regular grid representation, such as a 3D tensor (Figure 7.3), to enable training with standard deep learning architectures like 3D convolutional neural networks [40, 41]. However, this one-to-one mapping from detector cells to a regular grid introduces a fundamental problem: the shower data becomes entangled with position-specific geometrical features.

The staggering effect, illustrated in Figure 7.4, exemplifies this issue. Models trained on such representations inadvertently learn not only electromagnetic shower physics, but also detector-specific artefacts tied to the training location. This creates a fundamental limitation: the model becomes biased to specific incidence positions and loses generalisability across the calorimeter. While conditioning the model on exact shower positions could in principle restore spatial generalisation, this approach is computationally infeasible due to the enormous number of possible incidence positions and orientations.

The CaloClouds Solution: Regularised Geometry and Point Clouds

To overcome these limitations, CaloClouds employs a two-part strategy: a regularised readout geometry combined with a high-resolution point cloud representation.

The first step involves simulating showers in a regularised version of the ILD ECAL geometry, as illustrated in Figure 7.5. By eliminating insensitive regions and staggering offsets (while preserving layer thicknesses), the model learns only the physics of shower development, decoupled from detector-specific readout irregularities. This approach was also adopted in [41] for training the Bounded Information Bottleneck Autoencoder (BIB-AE) model [40]. However, as demonstrated in [41], while this regularised approach decouples showers from position-

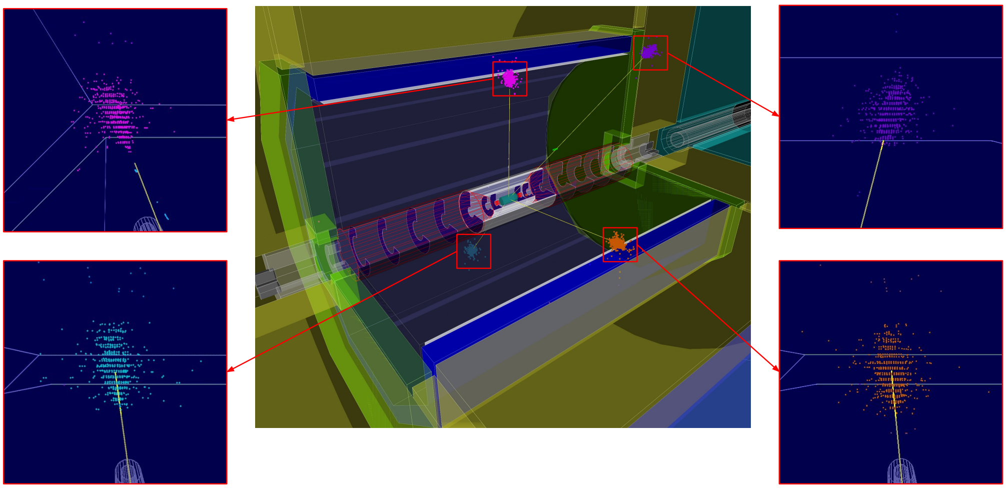
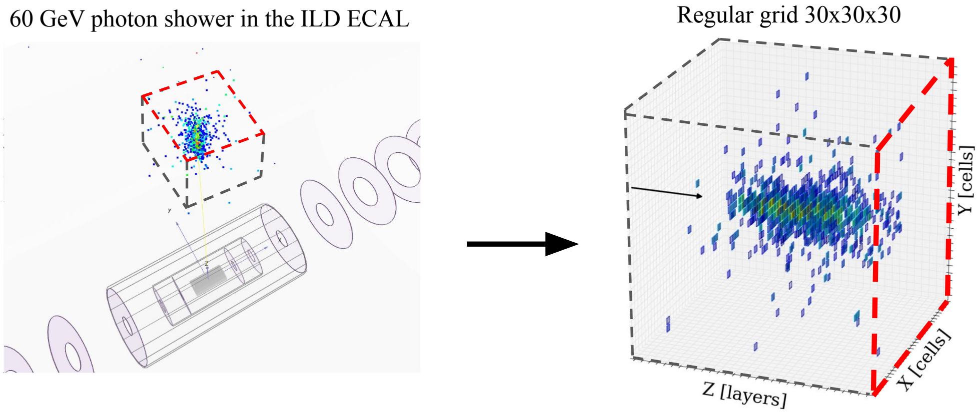


Figure 7.2: ILD event display with four 50 GeV photons fired from the interaction point in random directions. Each inset highlights the localised electromagnetic shower in the ECAL for a different photon. Despite varying impact positions across the detector volume, the shower shapes appear similar to each other due to the consistent material structure and longitudinal layering. This regularity justifies simplifying the training dataset to a single location with variable energy and angle, significantly reducing the dimensionality of the training phase space.


60 GeV photon shower in the ILD ECAL

Regular grid 30x30x30

Figure 7.3: Mapping of calorimeter showers from irregular detector geometry to a regular grid representation. The left panel shows the ILD event display with a 60 GeV photon shower in the physical geometry of the ECAL. The right panel shows the same shower mapped onto a regular 30 × 30 × 30 grid representation, where each voxel corresponds to a cell in the detector. Figure adapted from [40].

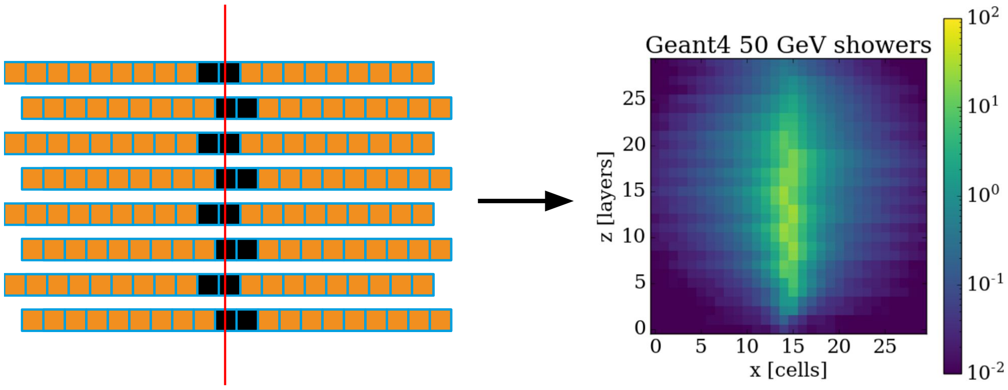


Figure 7.4: Staggering effect in the ILD ECAL geometry. The left panel shows a simplified schematic view of the ECAL layer structure in the physical geometry, where the sensitive layers (in orange) are staggered in the x direction. The red line illustrates the trajectory of a particle traversing the calorimeter, and the black cells indicate the activated cells. The right panel shows the projection of 2,000 photon showers with 50 GeV energy onto the regular grid. The staggering effect is clearly visible in the x direction as a periodic pattern in the core of the shower. The right panel is taken from [40].

specific geometry artefacts, having the same spatial resolution in the regular and physical geometries is insufficient to avoid projection artefacts when mapping generated showers to arbitrary detector locations. These artefacts arise from the discrete nature of the voxel-based representation, where energy depositions are confined to fixed grid cells that do not perfectly align with the target geometry after projection.

The CaloClouds approach takes this strategy a step further by creating a high-resolution point cloud representation from the raw Geant4 simulation steps, as illustrated in Figure 7.6. During full simulation with Geant4, a very large number of individual energy depositions (on average 20,000 per shower for photon showers at 50 GeV) from secondary particles traversing the sensors is created in the sensitive materials. These Geant4 steps represent the ultimate resolution of the simulation – a true point cloud of energy depositions with continuous spatial coordinates, completely independent of any readout geometry layout. By training on this high-resolution point cloud and projecting to the physical geometry only at inference time, CaloClouds avoids the projection artefacts inherent to voxel-based approaches.

Preprocessing Pipeline

While the raw Geant4 steps could in principle be used directly as input to the point cloud model, their number is prohibitively large. Therefore, a pre-clustering procedure is applied to reduce the data to a manageable size while preserving spatial structure.

The preprocessing pipeline operates as follows. First, the ILD detector's azimuthal symmetry is exploited: the calorimeter is divided into 8 identical octants in ϕ, allowing training and evaluation to be performed on a single octant. The Geant4 steps are then pre-clustered by

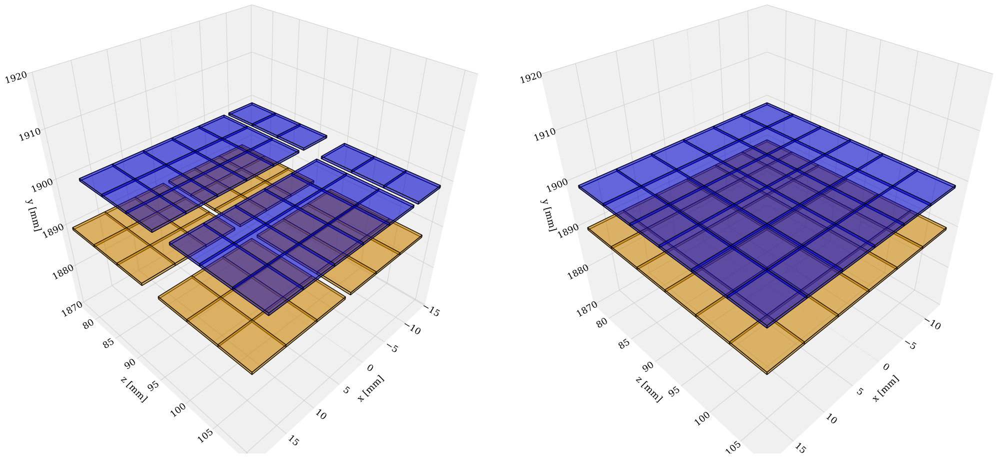


Figure 7.5: Visualisation of geometry maps for (left) the physical geometry and (right) the regular geometry for a section of two sensitive layers in the calorimeter. The physical geometry includes gaps between the cells arising from insensitive volumes such as structural supports and readout electronics, 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.

binning the energy deposits in each layer into a grid with higher granularity than the physical cell size. In CaloClouds2, a 6×6 grid per detector cell (36 times higher granularity) was used, with points placed at the centre of each bin. CaloClouds3 adopts a slightly coarser 5×5 grid per cell, as this was found to have minimal impact on kinematics after projection while improving computational efficiency. Additionally, CaloClouds3 places points at the location of the highest-energy step within each bin rather than at the bin centre, better preserving the spatial structure of the shower. In each event, the binning grid is given a uniform random offset of up to half a cell to reduce discretisation effects. This creates a set of up to $\sim 6,000$ points per shower.

The preprocessing pipeline defines a bounding box around the showers: ±200 mm in x and y for CaloClouds2, extended to ±250 mm (100×100 cells) in CaloClouds3. The centre of each layer is shifted in the x and y directions to remove the shower's tilt, as illustrated in Figure 7.7. The shift is calculated from the layer depth and incident particle momentum, aligning the shower core along the z-axis. This preprocessing helps the diffusion model learn the details of backscattering rather than focusing on the gross shape, which is known from the incident particle direction.

Since there are no recorded Geant4 steps in the absorber layers, the data is transformed such that the absorber regions are removed and the active layers become contiguous. The z positions within the active layers are uniformly smeared to produce data that is smooth in z, whereas in the original ILD simulation, energy is deposited predominantly at the centre of each layer for technical reasons. Finally, the cluster positions are normalised such that the bounding box boundaries correspond to positions of -1 and $+1$ respectively, improving numerical stability during training.

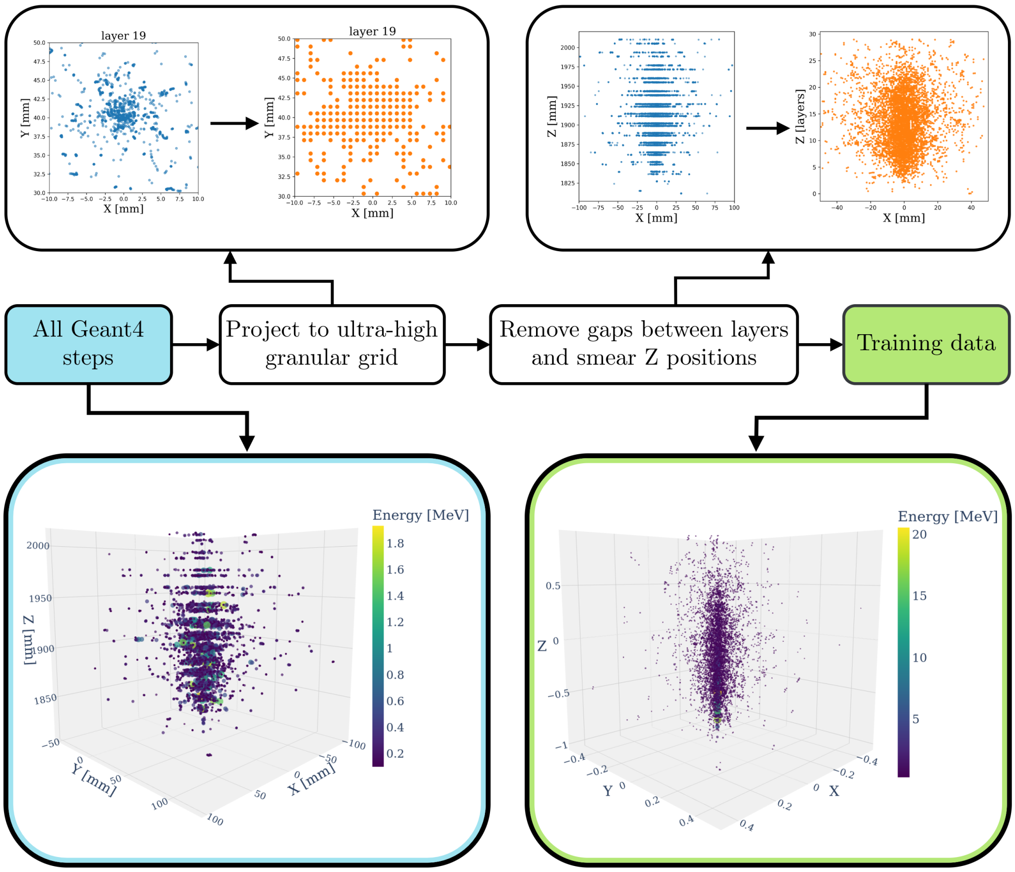


Figure 7.6: Data preprocessing pipeline for CaloClouds. The raw Geant4 simulation steps provide fine-grained energy depositions along particle trajectories, with on average 20,000 steps per shower. These steps are first projected onto a regular grid with higher granularity than the calorimeter cells, reducing the number of points while preserving spatial structure. The gaps between active layers are then removed and the z positions are smeared to produce smooth distributions. The resulting high-resolution point cloud representation with normalised coordinates enables flexible projection to arbitrary detector geometries without reconstruction artefacts.

7.3 Model Architecture

Having established how calorimeter showers are represented as high-resolution point clouds, this section describes the generative model architecture used to generate them. All three CaloClouds model versions share a common three-step generation procedure, illustrated in Figure 7.8: (1) a normalising flow generates global shower properties, (2) a diffusion model generates the energy depositions, and (3) a calibration step combines these outputs into a physically consistent shower.

Shower Flow: Global Property Generation

The first component, referred to as ShowerFlow (shown in purple in Figure 7.8), is a normalising flow model that generates global shower properties conditioned on the incident particle. In CaloClouds3, the conditioning variables are the normalised incident energy E and the normalised momentum direction $(p_x, p_y, p_z)/|p|$. While versions 1 and 2 used only the incident energy as conditioning. The decision to use Cartesian coordinates for the

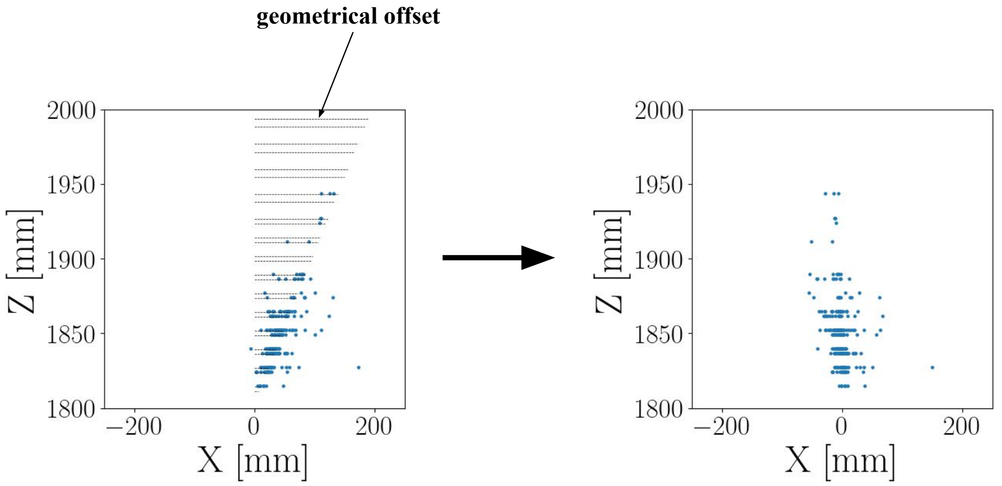


Figure 7.7: Removal of shower tilt in the preprocessing pipeline. The left panel shows a shower from an angled incident photon, where the shower core is displaced in x as a function of depth z. The right panel shows the same shower after applying layer-by-layer geometrical offsets to remove the tilt, aligning the shower core along the z-axis. This transformation allows the model to learn shower physics independently of the incident angle.

direction, rather than a pair of angles, was driven by a desire to avoid degeneracy in the input space. Consider a local polar angle measuring the deviation of the incident photon from one that travels at 90° to the surface of the layers. When this local polar angle is 0, there are an infinite number of azimuthal angles specifying the same direction, which is challenging for interpolation. Some minor instability is observed at $(0, 0, 1)$, as this point is singular. Occasionally, this leads to unphysical events from ShowerFlow (with many negative energies or numbers of points), but it is rare enough (around 0.1% of samples) to be handled by rejection sampling.

The features predicted by ShowerFlow have evolved across model versions to remove redundancies and improve the stability of the training:

- In CaloClouds, ShowerFlow predicted the number of points per layer and the total shower energy.
- In CaloClouds2, ShowerFlow additionally predicted the energy per layer and the centre of gravity (CoG) of the shower in x, y, and z.
- In CaloClouds3, ShowerFlow is limited to predicting only the number of points per layer and the energy per layer, with the CoG prediction removed (see below).

The CaloClouds3 ShowerFlow architecture interleaves 12 affine couplings and 2 spline couplings, substantially reduced from the 60 affine and 10 spline couplings used in CaloClouds2. This simplification was found to be sufficient to capture the complexity of the distribution while being faster at inference time and more stable to train. The output values are scaled by fixed factors (7864 for point counts, 3.4 for energies) to rescale the majority of each distribution to lie between 0 and 1. The model is trained using the Adam optimiser with a learning rate of 0.0002 and batch size of 62.

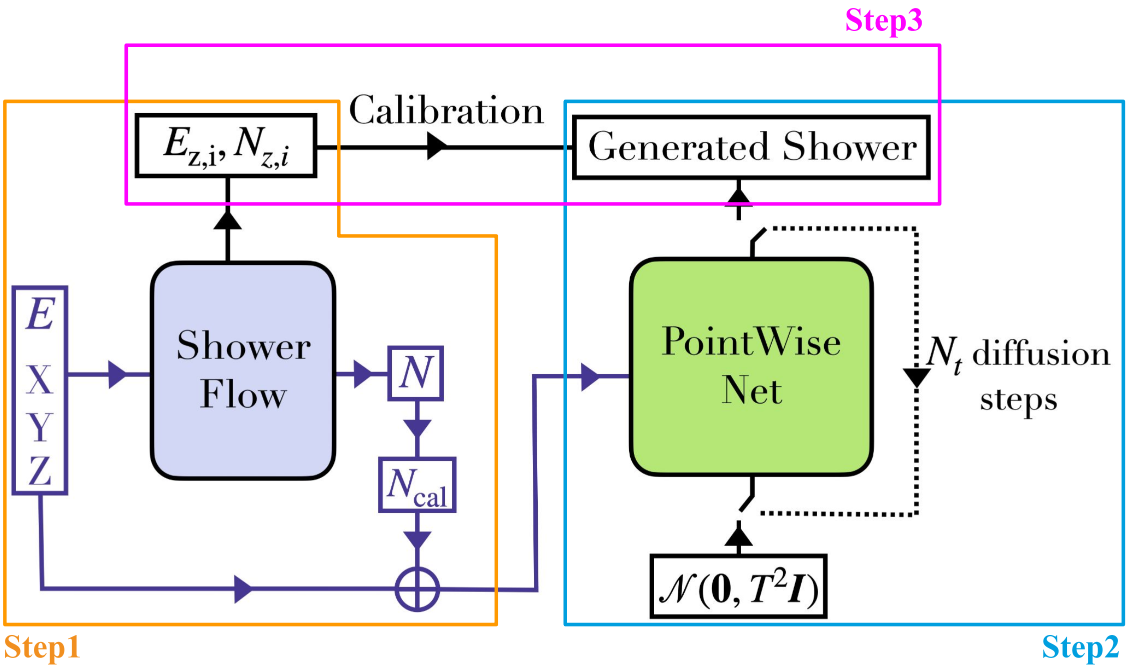


Figure 7.8: CaloClouds3 model architecture overview. Step 1 (orange): ShowerFlow, a normalising flow, generates global shower properties – the number of points $N_{z,i}$ and energy $E_{z,i}$ per layer – conditioned on the incident particle energy E and direction (X, Y, Z). Step 2 (cyan): a diffusion model generates $N_{\text{cal}} = N \times scalefactor$ individual points from Gaussian noise $\mathcal{N}(\mathbf{0}, T^2\mathbf{I})$ in N_t denoising steps. Step 3 (magenta): calibration assigns points to layers according to ShowerFlow predictions and scales energies to match the predicted layer energies, producing the final generated shower.

Point-Wise Diffusion Model

The second step employs a point-wise diffusion model (shown in green in Figure 7.8) that generates the actual point cloud of energy depositions. A key simplification in the CaloClouds approach is the assumption that all points in the shower are independent and identically distributed (i.i.d.). While electromagnetic showers do exhibit correlations between hits – for example, from secondary particles scattering at wide angles – treating points as i.i.d. dramatically simplifies the model architecture and reduces inference time.

The diffusion model generates points by iteratively denoising samples drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0}, T^2\mathbf{I})$. In CaloClouds2 and CaloClouds3, the diffusion model has been distilled to a single denoising step, enabling rapid generation. Unlike CaloClouds2, the CaloClouds3 diffusion model is no longer conditioned on the total number of points; removing this conditioning was found to have no impact on physics performance. The conditioning is therefore the same as for ShowerFlow: the incident particle energy and direction.

Each generated point consists of spatial coordinates (x, y, z) and an energy value expressed on a logarithmic scale, ensuring the energy spectrum covers all real numbers.

Calibration and Final Assembly

The third step (shown in magenta in Figure 7.8) combines the outputs of ShowerFlow and the diffusion model into a physically consistent shower through a calibration procedure.

Starting from the first layer the photon reaches and working outwards:

1. The number of points predicted by ShowerFlow for each layer are selected from the pool of points generated by the diffusion model. The selected points are assigned z positions at the centre of their respective layer.

2. The total energy in each layer is rescaled to match the energy predicted by ShowerFlow.

Two fixed scale factors are applied to correct for residual biases in the occupancy and total energy of the shower. This is simplified from CaloClouds2, where occupancy was corrected with a polynomial fit.

Removal of Centre of Gravity Calibration

In CaloClouds2, ShowerFlow predicted values for the CoG perpendicular to the shower axis, and all points were shifted to match during calibration. While this achieved good agreement with the Geant4 CoG distribution, the approach is unphysical. The tails of the true CoG distribution arise not from shifting the bulk of points, but from correlated, off-centre clusters of hits produced by wide-angle secondary particles. Since the diffusion model draws points i.i.d., these correlated structures cannot be replicated.

In CaloClouds3, the CoG calibration was therefore removed, and the x and y centre of gravity naturally yielded by the diffusion model is left unmodified.

8 Integration into the International Large Detector Software Chain

This chapter describes the integration of generative models for fast calorimeter simulation into the ILD software chain. The work builds on collaborative efforts within the FTX-SFT group at DESY and was initially introduced as DDFastShowerML in [41], later renamed to DDML and published in [4].

Peter McKeown initially developed the DDML library in collaboration with Thomas Madlener and Frank Gaede, which forms the foundation of this integration framework. Figures in this chapter are from [4]. The author contributed by scripting the CaloClouds model for deployment and implementing adaptations to DDML for integrating the CaloClouds model into the ILD simulation chain. Thorsten Buss developed, trained, and integrated the L2LFlows model (Chapter 9) into DDML.

8.1 The DDML Library

To study generative model performance in realistic physics simulations with event reconstruction, the models must integrate with the experiment's standard software ecosystems, typically implemented in C++. This section describes the DDML library, a generic framework for integrating generative models for fast calorimeter simulation into full simulation applications.

8.1.1 Overview and Motivation

The DDML[1] library [168] was developed as part of the Key4hep software stack to integrate generative models for fast calorimeter simulation into full simulation applications. It follows the Geant4 Par04 example for running machine learning inference in fast simulation models [169]. Built on top of Geant4 and DD4hep toolkits, the library provides access to extensive functionality. This includes an interface to a Geant4 trigger mechanism for terminating physics-based full simulation in favor of an alternative simulation approach. A trigger is associated with a specific geometrical detector region and activates when an impinging particle satisfies certain criteria (particle type, energy, etc.). This enables seamless incorporation of generative model-based fast calorimeter simulation into full simulation applications.

8.1.2 Design Principles

The DDML library supports a generic approach to fast simulation with generative models through three key requirements:

1. Support different types of generative models regarding structure, required inputs, and generated outputs.
2. Enable different engines for model inference.
3. Allow different detector geometries implemented in DD4hep.

---

[1] https://github.com/key4hep/DDML

Two conventions are adopted for the library. First, a local right-handed coordinate system is defined with the origin at the calorimeter entry point, the z' axis orthogonal to the calorimeter face pointing inward, and the x' axis aligned with the magnetic field direction. This enables consistent handling of model output regardless of where a particle strikes the detector. Second, model outputs are interpreted as local space points in this coordinate system, providing a generic means of interpreting outputs independent of architectural details.

8.1.3 Architecture

The library is split into multiple interfaces via a class template, decoupling the three aspects detailed above for easier extension and maintenance. The interfaces are outlined below, with a class diagram shown in Figure 8.1 and the order of operations in Algorithm 1.

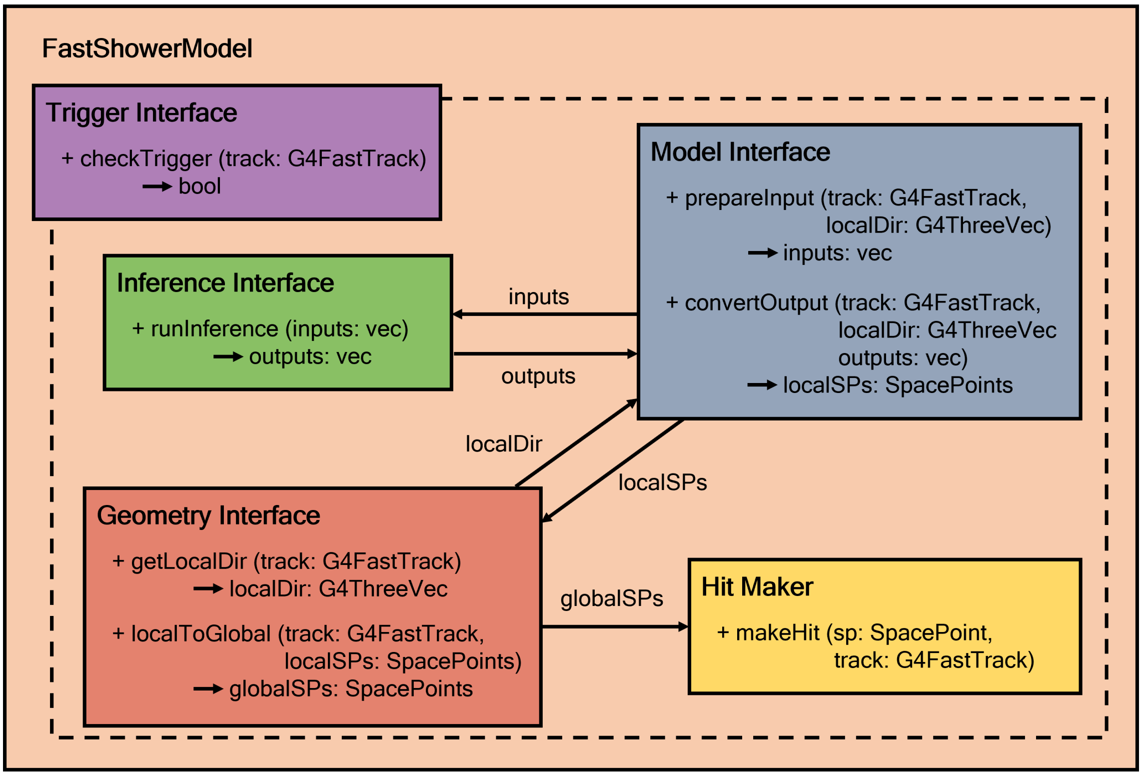


Figure 8.1: Class diagram illustrating the core components of the DDML library. See the text for a detailed description of the interfaces. Figure from [41].

Trigger Interface

The Trigger Interface extends the trigger on particle type and energy existing in DD4HEP and GEANT4. It enables excluding a fast simulation model from running in certain detector regions (*checkTrigger*). This provides a simple means of handling calorimeter regions with irregular structure, where full simulation or a separate generative model can run instead.

Model Interface

The Model Interface provides model-specific implementations with two roles. First, it prepares the input (*prepareInputs*) in the format expected by the model, with a *localDir* object available

Algorithm 1 Pseudocode illustrating the order of operations for the core components of the DDML 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
```

to provide local direction information at the calorimeter face for use as conditioning input. Second, it interprets the model output for conversion into local space points (*convertOutput*).

Inference Interface

The Inference Interface provides a simple means of calling the inference library for a model (*runInference*). Both LIBTORCH [170] and ONNXRUNTIME [171] inference libraries are currently supported. Additionally, functionality exists for loading pre-simulated shower libraries from HDF5 files, intended for model prototyping and representation investigations.

Geometry Interface

The Geometry Interface performs two roles. First, it computes the local direction (*getLocalDir*), providing the Model Interface a consistent means of model conditioning. Second, it places the local space points from the Model Interface into the detector geometry (*localToGlobal*), including conversion from local calorimeter coordinates to global envelope coordinates and placement of hits onto sensitive detector elements. It must be implemented per geometry, with disk endcap, polyhedral, and cylindrical barrel calorimeter geometries already supported.

HitMaker Interface

A GEANT4 helper class that enables placement of energy deposits from the fast simulation model, provided their position lies within a sensitive detector element (*makeHit*).

The DDML library currently supports only single shower generation with a generative model (batch size of one) on a CPU, which remains the dominant hardware in high energy physics computing infrastructure [172]. Future support for batched shower generation and GPU acceleration is planned, which would enable the most significant simulation speed-ups relative to GEANT4 through parallelization.

8.1.4 DDML Implementation for the ILD Detector

The DDML implementation for the ILD detector used in this study includes both the CaloClouds3 and L2LFlows models described in Chapter 7. The models were converted to a C++-compatible format, with architecture components serialized through a combination of tracing and scripting individual operations. This was achieved primarily using TorchScript utilities in PyTorch [170], with the ShowerFlow component of the CaloClouds3 architecture utilizing POUTINE effect handlers from the PYRO [173] deep probabilistic programming library in which it was implemented.

This study focuses on the ILD detector's barrel region. To leverage the detector's symmetry, showers at different positions are generated using models trained on showers in a single location, as described in Chapter 7.2. This is possible because the regularized calorimeter used for the training dataset has no dead material within active layers and no gaps, making the model a valid simulator for most of the detector. In particularly irregular regions, full simulation runs instead using the trigger interface from Section 8.1.

Two types of regions are excluded. First, the barrel-endcap transition region contains a gap and a change in calorimeter layer orientation. Therefore, barrel edges at $\theta < 40$ degrees and $\theta > 140$ degrees in the global ILD coordinate system are excluded from fast simulation. Second, the intersection between staves of the octagonal barrel exhibits asymmetrical calorimeter layer orientation changes. These 8 transition regions are excluded by 8 cuts in the global ILD coordinate system, each consisting of an 8.01 degree range in φ.

This implementation demonstrates DDML library's flexibility in handling realistic detector geometries while seamlessly integrating generative models into the full simulation chain.

9 Results and Benchmarks

This chapter presents a comprehensive evaluation of the generative models developed for electromagnetic shower simulation in the ILD detector. The assessment follows a systematic approach, progressing from idealized reference scenarios to increasingly complex and realistic physics applications. The results have been previously published in [4].

The L2LFlows model included in the benchmarks was developed, trained, and integrated into the ILD simulation pipeline by Thorsten Buss. The DDML library and integration framework were developed by Peter McKeown in a collaboration with Thomas Madlener and Frank Gaede, with contributions from the author for specific features required for the benchmarks. The plots and studies presented in this chapter were performed by the author.

The evaluation framework is designed to address two fundamental questions: first, how well do generative models reproduce the physics of electromagnetic shower development, and second, what are the intrinsic limitations imposed by different data representations independent of model performance. To answer these questions, this work introduces a novel benchmarking methodology based on *optimal shower generators* that isolate the effects of data preprocessing strategies from generative model capabilities.

The benchmarking strategy proceeds through several levels of complexity:

- **Optimal generators** (Section 9.2): Establish theoretical performance limits for different data representations by using idealized generators derived directly from Geant4 simulation steps.

- **Single particle performance** (Section 9.3): Evaluate fundamental shower characteristics including energy profiles, resolution, linearity, and angular reconstruction for individual electromagnetic showers.

- **Multi-particle scenarios** (Section 9.4): Test performance with overlapping showers using controlled di-photon benchmarks to probe reconstruction capabilities under realistic conditions.

- **Full physics applications** (Section 9.5): Assess model fidelity in complete physics processes using tau lepton decays, representing the ultimate test of practical applicability.

- **Computational performance** (Section 9.6): Quantify the speed advantages of generative models compared to traditional Geant4 simulation across different energy ranges.

- **Systematics from simulation methodology** (Section 9.7): Study geometric biases when applying models trained on idealized to realistic geometries.

- **Observable comparison summary** (Section 9.8): Provide a quantitative overview of model agreement with Geant4 across all benchmark observables using JS divergence and L_1 metrics.

Throughout this evaluation, the performance of two primary generative models is compared: CaloClouds3, a point cloud-based diffusion model that operates with high resolution point cloud shower representation, and L2LFlows, a grid-based normalizing flow model that uses fixed voxelized representations. The BiBAE [41, 174], the previous state-of-the-art generative model developed and integrated for the ILD detector, is not included in the physics performance benchmarks due to differences in the simulation setup and Geant4 version used in previous studies compared to the one employed here. The performance of Optimum (x1) serves as a proxy to the upper bound performance of BiBAE, since Optimum (x1) is essentially the training data for the BiBAE model. However, timing results for the BiBAE are included in Section 9.6. These models represent different approaches in data preprocessing strategies and provide complementary insights into the trade-offs between flexibility, accuracy, and computational efficiency.

The results presented demonstrate that generative models can achieve high fidelity electromagnetic shower simulation in realistic physics scenarios while providing substantial computational speedups over traditional methods. The analysis also reveals important limitations, particularly regarding the impact of data representation choices on the achievable performance.

9.1 Benchmarking Methodology

As described in Chapter 8, the DDML library enables the integration of different generative models into the ILD simulation chain using the DD4hep toolkit [175]. This allows particle showers to be generated within the standard ILD software chain and reconstructed using particle flow reconstruction with PandoraPFA [111], enabling realistic physics benchmarking of generative models. All benchmarking samples used in this study have gone through the complete software chain including event reconstruction.

Although studying single-shower observables is important to gauge the performance of a model, as is now standard in the literature, in real physics events showers from multiple particles may overlap. This significantly increases the complexity of evaluating the performance of a model, as the breadth of the phase space makes disentangling the interplay of overlapping showers with reconstruction algorithms a challenging task. For this reason, a step-by-step approach is taken, ultimately building towards benchmarking the models in a full physics setting.

The evaluation begins by studying the performance of the models in terms of single particle observables. Next, the simplest scenario for a multi-particle test is examined – two photons fired into the face of the ECAL. This is a standard benchmark which has also been used for the development of reconstruction algorithms, such as PandoraPFA [176]. This approach provides an isolated and controllable means of probing generative model performance, as well as allowing connections to be drawn to the performance on single particle observables.

Finally, the performance of generative models for the simulation of photon showers in a full physics process is studied. As the performance required of a fast simulation tool will depend heavily on the physics process for which the tool is used, a physics process is desired that provides a stringent test of a fast simulation tool for electromagnetic showers. To this end, hadronic decay modes of the tau lepton in the process $e^+e^- \rightarrow \tau^+\tau^-$ are chosen.

As discussed in Chapter 2, the tau lepton is of interest for many precision studies planned for future e^+e^- collider experiments, particularly for probing the Higgs sector and its CP structure [71, 177–180]. Hadronic decay modes of the tau, which account for approximately 65% of tau decays, frequently involve one or more neutral pions produced via intermediate resonances such as $\rho(770)$ or $a_1(1200)$ (see Table 2.1 in Chapter 2). Since each π^0 almost always decays to two photons, correctly reconstructing the number of π^0s produced is essential for determining the tau decay mode and spin state. The high boost and collimation of the decay products makes this challenging, and the presence of numerous overlapping photon showers in such events makes hadronic tau decays a standard benchmark of electromagnetic calorimeter performance [74, 181]. These events are therefore ideal for exposing any flaws in the performance of a fast simulation tool for electromagnetic calorimeters. Correctly reconstructing these π^0s involves not only distinguishing the number of photons from overlapping showers (the performance of which can be linked to the aforementioned isolated di-photon benchmark), but also correctly inferring the kinematics from shower-level observables (which can be linked to the aforementioned single particle benchmark).

9.2 Optimal Shower Generators

To better understand the effects and potential limitations of projecting showers into a realistic detector geometry, three different shower representations are considered, each with a corresponding *optimal shower generator*. Each of these optimal generators is derived from simulations run with GEANT4 on the regular ILD ECAL introduced in section 7.2, from which all individual energy depositions within the sensitive layers, so called GEANT4 steps, are extracted. This approach enables quantification of the effect of projecting a regular grid with a given granularity onto the actual detector readout geometry, and allows isolation of the effects of the data representation from the performance of a given generative model. The three representations are denoted as $\mathcal{R}_{\times 1}$, $\mathcal{R}_{\times 9}$, and $\mathcal{R}_{(\text{steps})}$.

The first representation $\mathcal{R}_{\times 1}$ features a granularity identical to that of the ILD ECAL (5×5 mm^2), and has a corresponding optimal generator OPTIMUM (X1) shown in Figure 9.1 (left). Here, each simulated step from GEANT4 is projected onto a virtual regular grid, matching exactly the realistic detector cell sizes. Thus, OPTIMUM (X1), provides the ideal reference for the achievable performance for any model trained on the readout geometry of the physical detector.

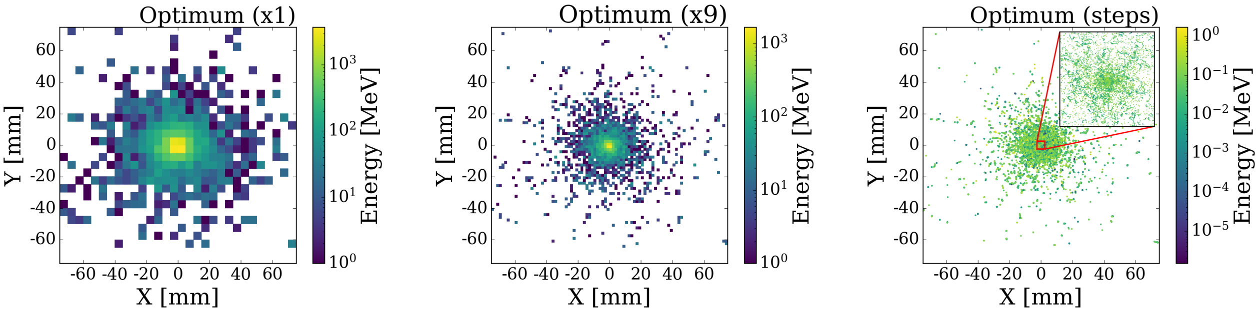


Figure 9.1: Visualization of the same 90 GeV electromagnetic shower in a lateral projection of the ILD ECAL, represented using the three optimal generators: OPTIMUM (X1) (left), OPTIMUM (X9) (center), and OPTIMUM (STEPS) (right).

The second representation $\mathcal{R}_{\times 9}$ increases the lateral granularity by a factor of three in each dimension, resulting in nine times more cells (1.67×1.67 mm^2) per layer compared to $\mathcal{R}_{\times 1}$. This representation has an optimal generator OPTIMUM (X9), shown in Figure 9.1 (center). Such an increased granularity allows for a finer spatial resolution of showers, reducing projection-related effects. It serves as a reference for understanding the impact of increasing the granularity of the data representation on the fidelity of the projected showers.

Finally, representation $\mathcal{R}_{(\text{steps})}$, with corresponding generator OPTIMUM (steps) shown in Figure 9.1 (right), provides the most detailed reference scenario by avoiding any spatial projection onto a predefined grid altogether. Instead, it directly utilizes the ultimate resolution – GEANT4 simulation steps – as they occur within the sensitive material. This results in the highest achievable spatial resolution, completely free from projection artifacts, and represents the ultimate benchmark for evaluating the accuracy and potential information loss associated with any spatial discretization scheme.

By leveraging these optimal generators – OPTIMUM (X1), OPTIMUM (X9), and OPTIMUM (steps), insight is gained into the intrinsic limits imposed by projection artifacts, independent of generative model performance. Moreover, they provide invaluable baselines against which generative models can be assessed.

9.3 Single Particle Performance

A detailed validation of individual particle showers is crucial to ensure that these models correctly capture essential physics characteristics before being used in more complex, multi-particle scenarios or physics analyses.

An evaluation of key calorimetric observables for electromagnetic showers generated by single photons using the generative models described previously is presented. Radial and longitudinal shower profiles, energy resolution, linearity, and the intrinsic shower angle reconstruction are studied. These observables are usually chosen in the literature because they encapsulate the essential features of electromagnetic shower development, and significantly influence the performance of particle identification and reconstruction algorithms.

The generative models are further benchmarked against the optimal generators introduced in Section 9.2. These optimal generators represent idealized performance scenarios that isolate and quantify the intrinsic limitations arising from spatial discretization effects and detector irregularities, independent of the generative model itself. By comparing the performance of the generative models against both these optimal generators and the standard GEANT4 simulation, a clear distinction can be made between artifacts arising from the data representation and the intrinsic capabilities of the approaches to generative modeling explored. This structured approach enables a detailed assessment of generative model fidelity and highlights areas requiring further improvement.

Dataset

The training dataset used in the study consisted of ~ 3 million samples, initiated by photons with incident energies uniformly distributed in the range of 1–126 GeV, created at a position of $[X = 0,\ Y = 1804.7\text{ mm},\ Z = -50\text{ mm}]$ at the front face of the ECAL. The incident angles were varied within a cone of up to 60° in θ' (relative to the normal to the calorimeter layers), with an azimuthal angle φ' uniformly distributed in $[0°, 360°)$, in the local frame of reference where the Z axis is aligned with the normal to the calorimeter layers. This configuration enabled uniform sampling over the space of possible incident directions on the calorimeter face.

In order to validate the generalization capability of the trained models, independent test samples were generated at multiple positions, uniformly distributed over a single ECAL barrel segment. Test samples were produced at fixed photon energies between 10 and 100 GeV in 10 GeV increments. For each test energy, a sample consisted of 3,000 photon showers with positions selected randomly on the front surface of the ECAL. The incident directions were chosen such that they mimicked particles originating from the interaction point (IP), across the angular ranges $43° < \theta < 137°$ and $79° < \varphi < 109°$ in the ILD global coordinate system, thereby effectively covering one complete stave of the barrel. This configuration ensures full coverage of the region where models are expected to operate, exposing them to a wide variety of incident directions and local geometries of the sensitive layers – thereby enabling the evaluation of model performance under realistic conditions.

Results

Shower Profiles

The performance evaluation of the generative models begins by examining their ability to reproduce the characteristic EM shower shapes, assessed through comparisons of radial and longitudinal energy profiles, which are key observables reflecting the spatial energy deposition pattern within the calorimeter.

The following observables are computed at the reconstruction level, after the generative models have been fully integrated into the simulation pipeline. This ensures that any potential geometric or systematic effects introduced by the integration framework are accounted for in the performance evaluation.

The radial energy profile, shown in Figure 9.2 (left), illustrates the mean deposited energy as a function of orthogonal distance from the axis aligned with the direction of the incident particle to the center of the cell. The longitudinal energy profile, displayed in Figure 9.2 (right), represents the average energy deposited per calorimeter layer along the depth of the detector.

Both models reproduce the longitudinal profile with good accuracy. The CaloClouds3 model achieves the closest agreement with Geant4, with deviations typically within a few percent. The L2LFlows model performs similarly well, although it exhibits deviations of up to 15% near the start and end of the calorimeter. Most of these deviations are due to the finite simulation volume required for fixed grid models.

At first inspection, the radial energy profile appears to only be reproduced well by the optimal generators – Optimum (x1), Optimum (x9), and Optimum (steps), while the CaloClouds3 and L2LFlows models have notable discrepancies at larger radial distances. However, it is important to note that the radial profile of electromagnetic showers is inherently steep and narrowly peaked, with the energy density rapidly decreasing with distance from the shower axis. In fact, more than 90% of the total energy of the shower is contained within a radius of 30 mm, indicating that deviations at large radii have a limited impact on the overall shower description.

To better assess the fidelity of the models in the region that dominates the shower energy density, Figure 9.3 zooms into the first 30 mm from the shower axis. Note that the dip in the first bin arises from the fact that the binning here is applied at a distance less than the width of a cell, whereas the hits in the shower are necessarily at the center of a cell after reconstruction. This is the most relevant part of the shower which plays a crucial role in the separation of overlapping showers during particle reconstruction.

In this region, the CaloClouds3 and L2LFlows models show good agreement with Geant4, with relative deviations generally remaining below 10%.

By contrast, Optimum (x1) underestimates the energy density by up to 20% in the innermost bins. This demonstrates the intrinsic limitation imposed by the $\mathcal{R}_{\times 1}$ representation, where the coarser lateral granularity fails to resolve the sharply peaked energy profile near the shower axis. The lack of detail in this region can significantly impact the reconstruction of particle showers with a large degree of overlap. This highlights a significant disadvantage of generative models trained on fixed grids using the true detector granularity.

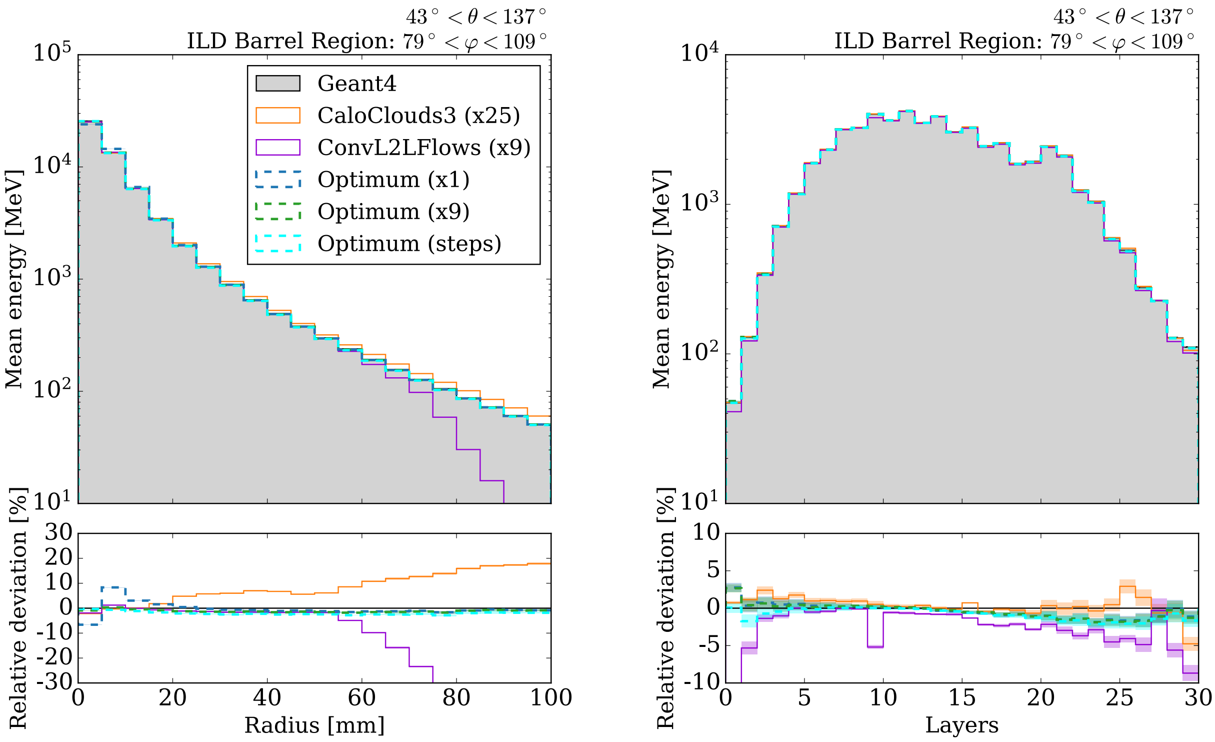


Figure 9.2: Radial (left) and longitudinal (right) energy profiles of electromagnetic showers, computed at the simulation level after integration of the generative models. The radial profile shows the mean reconstructed energy as a function of distance from the shower axis, while the longitudinal profile shows the mean reconstructed energy per calorimeter layer. These observables provide a detailed characterization of the transverse and longitudinal shower structure and are critical benchmarks for assessing how well generative models replicate Geant4 showers in realistic detector geometry settings. The color coding corresponds to the different generative models: CaloClouds3 (orange), L2LFlows (violet), Optimum (x1) (blue), Optimum (x9) (green), and Optimum (steps) (cyan). The Geant4 reference is shown in the light gray filled histogram. The color coding is consistent across all figures in this section. The shaded bands indicate statistical uncertainties; lower panels show relative deviations with respect to the Geant4 baseline.

Energy Resolution and Linearity

The energy resolution and linearity of the generative models are now investigated, two critical performance metrics for calorimeter simulation. The energy resolution quantifies the model's ability to accurately reproduce the fluctuations in the deposited energy, directly affecting the precision with which particle energies can be measured. Linearity assesses how accurately the reconstructed energy scales with the true particle energy, essential for ensuring unbiased energy measurements across a large range of incident particle energies.

The energy resolution is evaluated by measuring the relative width of the reconstructed energy distribution for photons at various fixed energies, defined as $\frac{\sigma_{90}}{\mu_{90}}$, where σ_{90} and μ_{90} are standard deviation and mean of the central 90% of the distribution, shown as a function of the incident photon energy in Figure 9.4 (left). Figure 9.4 (right) displays the linearity, expressed as the mean reconstructed energy divided by the true incident energy.

Similar to the radial profile results, the resolution plot shows the same trend. Among the

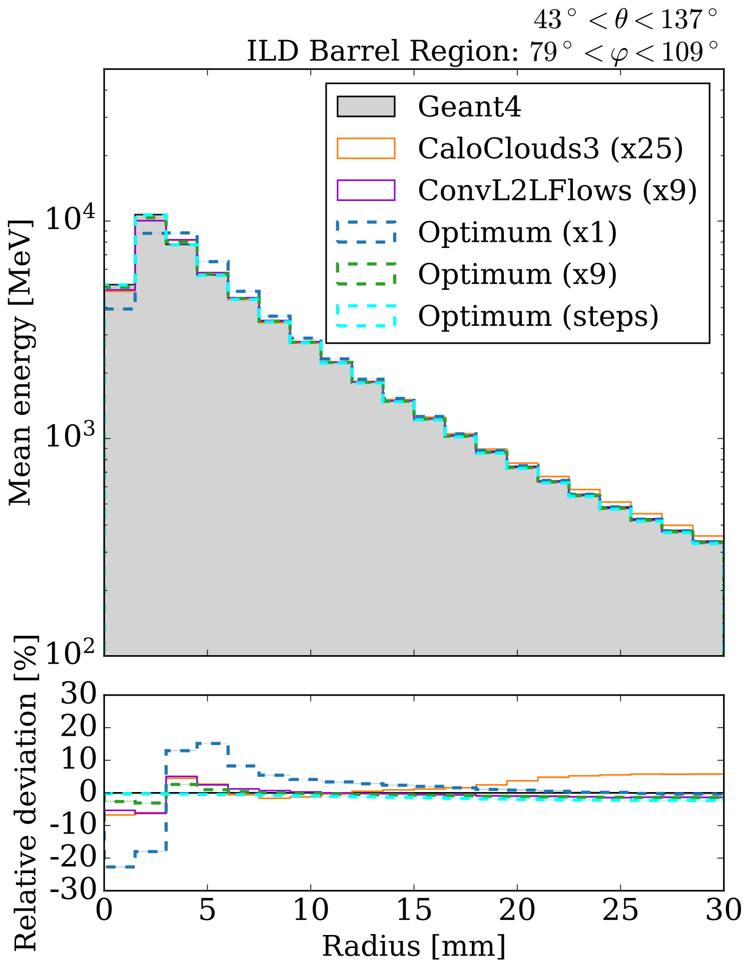


Figure 9.3: Radial energy profile of the showers, zoomed in to the first 30 mm from the shower axis. The shaded error bands correspond to the statistical uncertainty in each bin. The lower subplot shows the relative deviation of the radial energy profile with respect to the Geant4 reference. The color coding is consistent with Figure 9.2.

optimal generators, Optimum (x1) performs the worst, showing significant deviation from the Geant4 baseline. This demonstrates that the coarse $\mathcal{R}_{\times 1}$ representation lacks sufficient granularity to accurately capture the intrinsic energy fluctuations of electromagnetic showers. Despite being derived from full Geant4 simulation, this representation inherently limits the achievable fidelity due to the loss of fine spatial information.

With increased granularity, Optimum (x9) shows improvement, more closely tracking the Geant4 resolution. Finally, Optimum (steps), which directly uses the individual Geant4 steps without any spatial discretization, comes closest to reproducing the full simulation, representing the maximum achievable performance.

Importantly, the two generative models follow the trends of their respective representations. The CaloClouds3 model, trained on de-quantized $\mathcal{R}_{\times 25}$, almost reaches the performance of Optimum (steps). L2LFlows trained on the regular $\mathcal{R}_{\times 9}$ data representation, performs comparably to Optimum (x9), although the deviations from the optimal representation are larger than for CaloClouds3.

These findings demonstrate that both models have successfully learned from their respective training data representations, achieving performance that closely matches the optimal

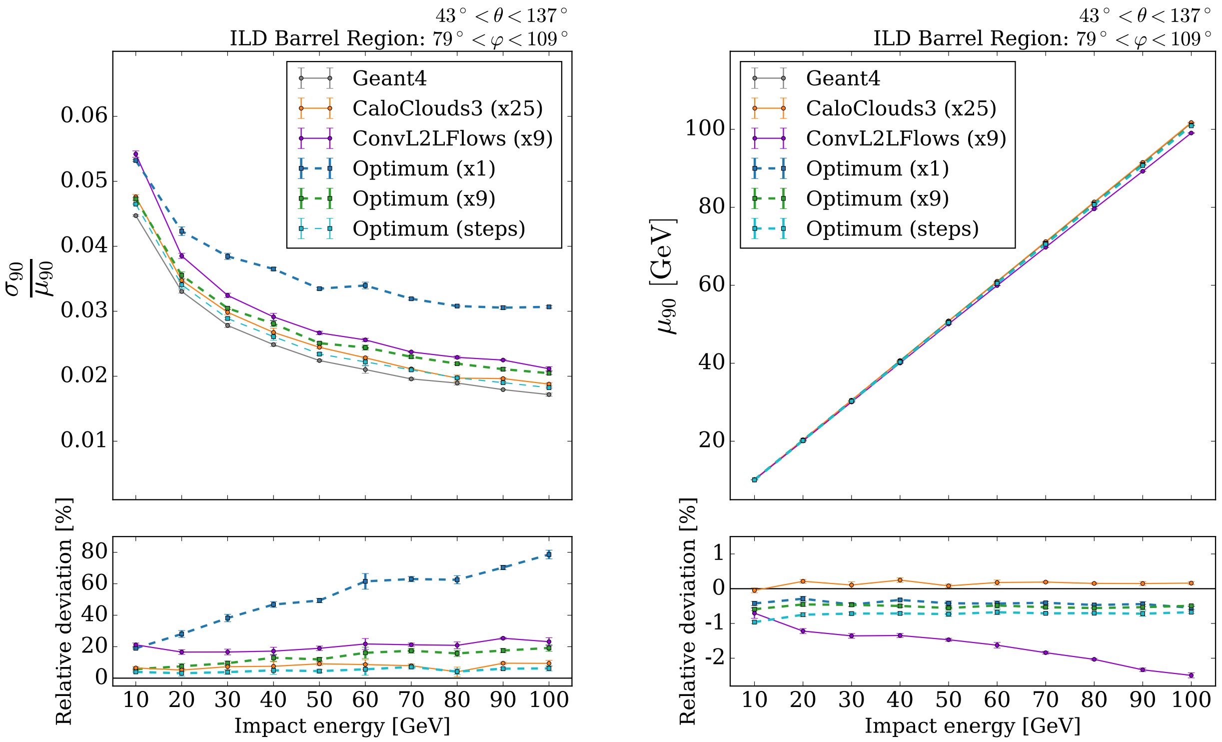


Figure 9.4: Energy resolution (left) and linearity (right) of reconstructed photon showers in the ILD ECAL. The resolution is defined as the relative width σ_{90}/μ_{90} of the central 90% interval of the reconstructed energy distribution. The linearity is given by the mean μ_{90} of this central interval as a function of the incident photon energy.

generators derived from the same representation.

The linearity is reproduced well by all models being within a $\sim 3\%$ relative deviation from Geant4, over the entire energy range tested – from 10 to 100 GeV with 10 GeV steps. Notably, the CaloClouds3 model exhibits the best agreement, showing negligible deviation from Geant4. This is a result of a simple scaling applied during integration, where each generated shower is rescaled by a constant factor determined from the difference in average visible energy at 50 GeV between the model and Geant4.

In principle, similar scaling could be applied to all models. However, for L2LFlows, and any grid-based model in general, such calibration is more challenging due to its restricted generation region, where an increasing fraction the shower's energy leaks out with increasing energy of the incident particle. As a result, the discrepancy between generated and true visible shower energy becomes more pronounced at higher energies, limiting the effectiveness of global rescaling.

This highlights a fundamental trade-off faced by grid-based models – they must balance performance against computational efficiency. Increasing the size of the generation volume can improve accuracy by reducing energy leakage, but it also significantly increases memory usage and inference time. As a result, achieving high fidelity with grid-based models requires careful tuning of the generation volume to remain computationally feasible while minimizing physical artifacts.

Intrinsic Angle Reconstruction

In a similar fashion to previous studies using the BiBAE model [5, 41], the previous state-of-the-art generative model applied to the ILD detector, the angular response of the generative models is evaluated by comparing the reconstructed intrinsic angles of showers simulated with Geant4 to those generated by the models. A principal component analysis (PCA) is applied to all reconstructed hits of each shower to determine its principal axis. The resulting angular distributions of the azimuthal and polar angles, denoted as *iPhi* and *iTheta*, respectively, are presented as the difference between the reconstructed and true angles of the incoming particle direction. These distributions are shown in Figure 9.5.

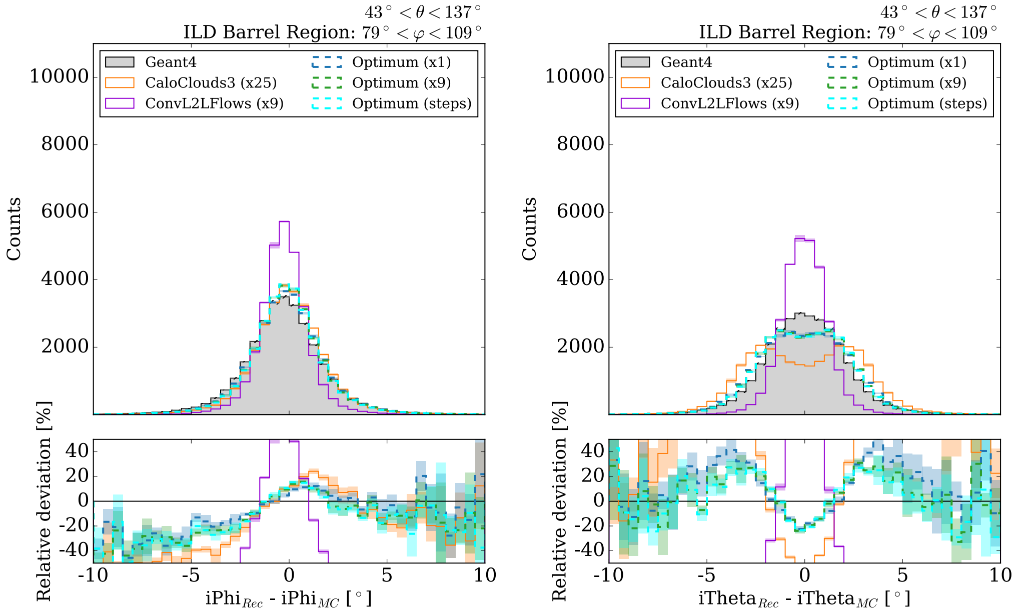


Figure 9.5: Distributions of the differences between reconstructed and true intrinsic angles for showers in the ILD barrel region ($43° < \theta < 137°$, $79° < \varphi < 109°$), comparing Geant4 to various generative models. Left: azimuthal angle (*iPhi*). Right: polar angle (*iTheta*). The top panel shows the angle residual distributions, while the bottom panel presents the relative deviation with respect to Geant4. A PCA is applied to all reconstructed hits to extract the principal axis of the shower.

As shown in Figure 9.5, both Geant4 and the generative models produce angular distributions that are centered around zero with comparable widths for both *iPhi* and *iTheta*, indicating that the models qualitatively reproduce the angular response observed in detailed simulation.

The CaloClouds3 and optimal shower generators, however, show a tendency to overestimate the polar angle, resulting in a double-peaked structure in the *iTheta* distribution. This indicates that this double-peak effect is likely related to the methodology of placing showers generated in the regularized detector into the real detector readout (see Section 9.7). This effect is more pronounced for CaloClouds3, and is likely compounded by the fact that CaloClouds3 slightly overestimates the energy for larger radii (i.e. further from the shower axis).

L2LFlows on the other hand shows a noticeably sharper peak in both the polar and azimuthal angle distributions. This observation, combined with the fact that L2LFlows generates showers within a tightly constrained spatial region [4, 123], suggests that a simple

PCA applied to all hits in the shower may be a suboptimal procedure for reconstructing the intrinsic shower angle. This is likely due to the high sensitivity of PCA to hits located far from the shower core, which can disproportionately influence the estimated principal axis.

To address this, an energy-based hit selection is applied, retaining only the top 4% most energetic hits before running PCA. This suppresses noise from peripheral hits and improves the stability of the reconstructed direction, meaning that this approach results in an improved angular reconstruction algorithm. Quantitative studies with more details on this reconstruction improvement were done in [3]. The results are shown in Figure 9.6.

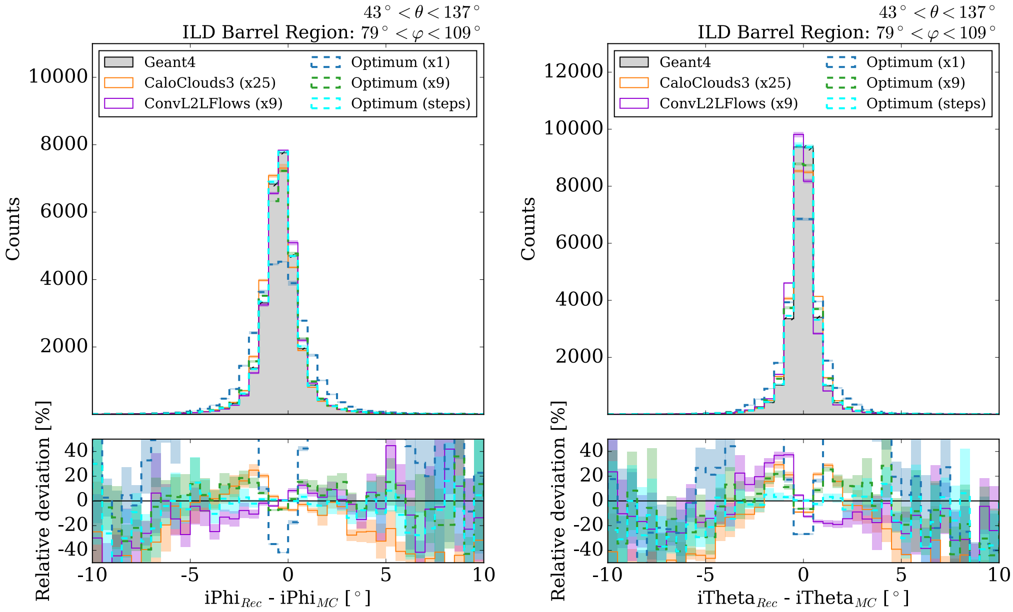


Figure 9.6: Distributions of the differences between reconstructed and true intrinsic angles for showers in the ILD barrel region ($43^\circ < \theta < 137^\circ$, $79^\circ < \varphi < 109^\circ$), comparing Geant4 to various generative models. Left: azimuthal angle (*iPhi*). Right: polar angle (*iTheta*). The top panel shows the angle residual distributions, while the bottom panel presents the relative deviation with respect to Geant4. Here, only the top 4% most energetic hits are used for PCA-based extraction of the shower axis, resulting in significantly improved angular resolution compared to using all hits.

All optimum shower generators, including CaloClouds3 and L2LFlows, now demonstrate angular resolutions that closely match the Geant4 reference. It is notable that the distribution produced by the Optimum (x1) shower generator shows the largest discrepancy, producing a slightly wider distribution than Geant4. The improved reconstruction shows better alignment with Geant4, and the double-peak structure previously observed in the polar angle distribution disappears entirely.

9.4 Di-Photon Separation

Having studied the single particle performance of the models, a multi-particle benchmark is now introduced. While studying single-shower observables, as is now standard in the literature, is important to gauge the performance of a model, in real physics events showers from multiple particles may overlap. This significantly increases the complexity of evaluating the performance of a model, as the breadth of the phase space makes disentangling the interplay of overlapping showers with reconstruction algorithms a challenging task.

For this reason a step-by-step approach is taken, ultimately building towards benchmarking the model in a full physics setting. The evaluation begins by studying the simplest scenario for a multi-particle test – two photons fired into the face of the ECAL. This is a standard benchmark which has also been used for the development of reconstruction algorithms, such as PandoraPFA [182]. This approach provides an isolated and controllable means of probing generative model performance, as well as allowing connections to be drawn to the single particle observables studied in Section 9.3.

Di-Photon Dataset

The Di-Photon benchmark dataset consists of 15,000 samples for each of three photon energy configurations: 5 GeV, 20 GeV, and 100 GeV, with each sample containing two photons of equal energy. The positions of the photons are randomly sampled within a small patch of the ILD ECAL barrel, such that their separation ranges from 0 mm to a maximum of 90 mm. The photons are produced directly at the face of the ILD ECAL barrel, with their direction of flight being orientated such that they appear to have been produced at the IP. This configuration of flight direction is chosen in this study because PandoraPFA assumes by default that photons are produced at the IP, and uses this information when performing particle identification.

The photon positions are randomly sampled to expose the di-photon system to different local geometries of the sensitive layers, while ensuring that the separation between the two photons varies uniformly from 0 to 90 mm. The photons impact the upper barrel module, within a narrow angular range of $\theta \in [83°, 87°]$ and $\phi \in [88°, 92°]$, corresponding to a small localized patch of the detector. To ensure that all photon pairs remain fully contained within the fast simulation trigger region (see Section 8.1) and avoid contamination from fully simulated showers, the center of this patch is placed well inside the boundaries of the trigger region. These samples are created for Geant4, each of the Optimum generators described in Section 9.2, and for both the L2LFlows and CaloClouds3 models. In each sample, both showers are produced with the respective generator.

Di-Photon Results

The di-photon benchmark allows the performance of simulators to be studied in a scenario where multiple showers overlap, emphasizing shower characteristics that are not directly probed by studying single shower observables, while maintaining a controlled environment that prevents any contamination which may be present in a realistic physics process. To evaluate this, the number of reconstructed photons is plotted against the separation between the two photons in Figure 9.7 and 9.8, for symmetric di-photon energies of 5 GeV (left), 20 GeV (middle) and 100 GeV (right). The performance is shown for Geant4, CaloClouds3, L2LFlows, Optimum (x1), and Optimum (x9). Note that Optimum (steps) is excluded to aid visibility in the plots and to reduce computational cost, as it performed on a similar level to Geant4. The error bars represent the binomial error in each case. The red line present in each plot represents two photons being reconstructed on average, which is the optimal case for this reconstruction scenario.

For all energies, at separations of less than approximately 6 mm, only a single photon is reconstructed on average. This corresponds to slightly more than one cell's worth of separation, meaning that the two individual shower cores cannot be resolved, with all hits being clustered into a single photon. In addition, it is easier for two photons to be resolved as their incident energy is increased. For photon pairs with incident energies of 100 GeV, the average number of reconstructed photons quickly rises to two for separations of only 10–20 mm, while for incident photon pairs with energies of 5 GeV the average number of reconstructed photons rises much more slowly, and only reaches an average of two for separations of around 60–65 mm. This is because incident photons with a higher energy have a higher energy core, which is also more densely populated than for lower energy photon showers, making them easier to distinguish [183].

Due to the relative simplicity of separating higher energy photons, both models and the optimal generators agree well with Geant4 in terms of the distribution of the average number of reconstructed photons for photon pairs with incident energies of 100 GeV, shown in Figure 9.8. Relative deviations in this instance appear around the level of a few percent, with the largest differences arising for the L2LFlows model.

At lower photon pair energies of 5 GeV and 20 GeV, shown in Figure 9.7, more significant deviations occur. Here, the Optimum (x1) generator shows significant discrepancies across a large range of separations, reaching relative deviations around the 10–15% level. Both the CaloClouds3 and the L2LFlows models show more contained relative deviations, typically less than 5%, with the most noticeable exception being for the CaloClouds3 model at 20 GeV and separations of around 10-20 mm where the relative deviation slightly exceeds this level.

These results indicate that training directly on the detector readout causes major discrepancies in the reconstruction performance. This is likely due to artifacts created when placing hits back into the detector geometry. The radial profile is of particular importance when separating such showers, as the reconstruction is especially sensitive to how the profiles of the two showers interfere. The effects observed around the core of the radial profile for photon showers produced with Optimum (x1) in Figure 9.3 are therefore considered a major factor in the poor performance of this generator. The differences observed for L2LFlows and CaloClouds3 are linked to the deviations in the radial profile observed in Figure 9.2 (left), although these are less influential as they occur further out from the shower core. The steep fall in the radial profile of the L2LFlows model, predicted by the constrained cut required for a regular grid model, seems to have little influence on the reconstruction performance in this test, as at large separations it is easy to distinguish the showers from their core alone.

Even with a perfect generative model trained on $\mathcal{R}_{\times 1}$, we cannot outperform Optimum (x1), as it represents the ideal case scenario for the $\mathcal{R}_{\times 1}$ data representation. This highlights the intrinsic limitations imposed by the $\mathcal{R}_{\times 1}$ granularity, where deviations in photon reconstruction efficiency of up to 15% can occur when two showers are in close proximity. These deviations primarily arise from the generator's mismodeling of the radial profile of electromagnetic showers – a feature that is crucial for separating overlapping photons during reconstruction. As shown in Section 9.3, the Optimum (x1) generator performs worst in modeling the radial profile.

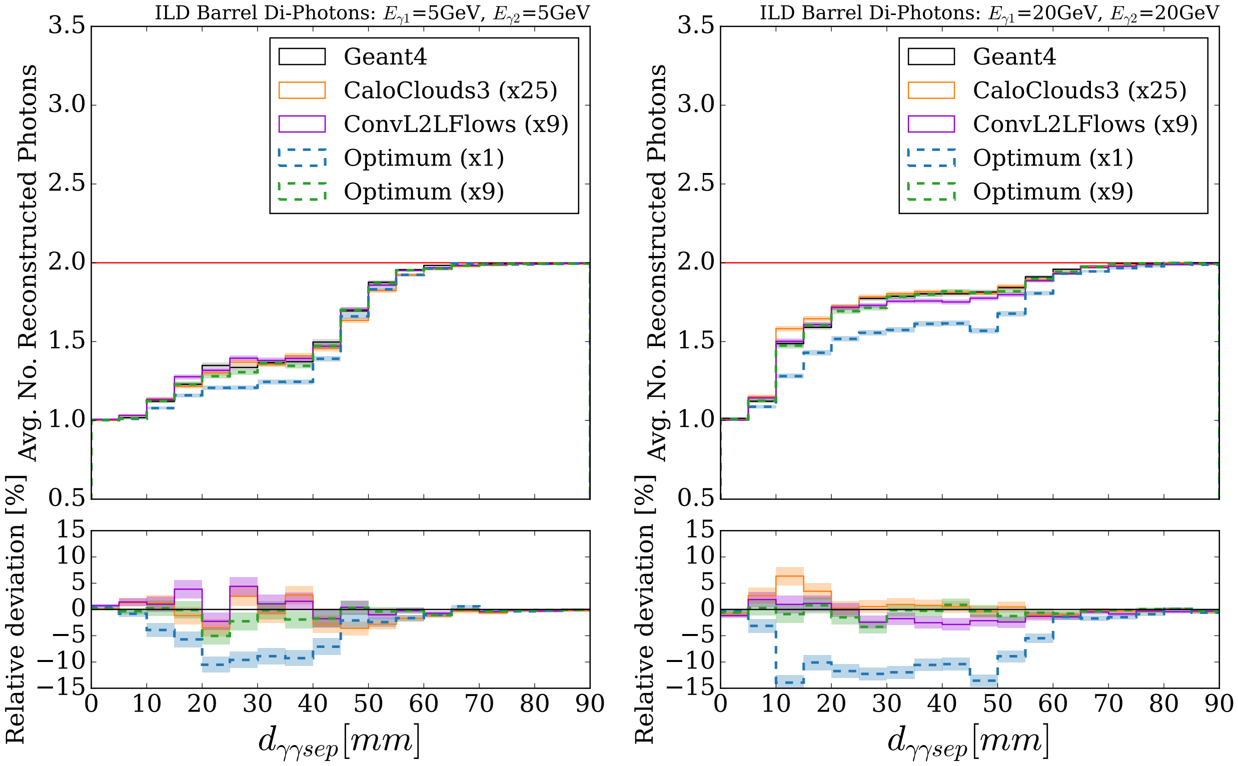


Figure 9.7: Performance of photon reconstruction for di-photon events at 5 GeV (left) and 20 GeV (right) photon energies, evaluated as a function of transverse shower separation $d_{\gamma\gamma\text{sep}}$. The red horizontal line indicates the ideal case of perfect reconstruction of two photons. The shaded error bands correspond to the statistical uncertainty in each bin. The lower subplots show the relative deviation of the number of reconstructed photons with respect to the Geant4 reference.

In contrast, increasing the spatial resolution, as done in the $\mathcal{R}_{\times 9}$ representation leads to significantly better agreement with Geant4, as can be seen by the Optimum (x9) results. Notably, both CaloClouds3 and L2LFlows are able to learn from this higher-resolution data, achieving a high level of agreement with Geant4.

The 100 GeV case, shown in Figure 9.8, demonstrates excellent agreement with the Geant4 reference across all generators. The high-energy photons produce showers with a very dense energy core, which increases the contrast between nearby energy deposits and allows individual photons to be identified even at small separations. All models, including CaloClouds3, L2LFlows, Optimum (x1), and Optimum (x9), track the Geant4 baseline closely over the full separation range, with relative deviations generally within 3%. However, the Jensen-Shannon (JS) divergence, which quantifies differences between two distributions, suggests that Optimum (x1) performs the worst among the models. A full comparison across all energies is provided in Table 9.1.

9.5 Full Physics Benchmark

The performance of generative models for the simulation of photon showers in a full physics process is now studied. As the performance required of a fast simulation tool will depend

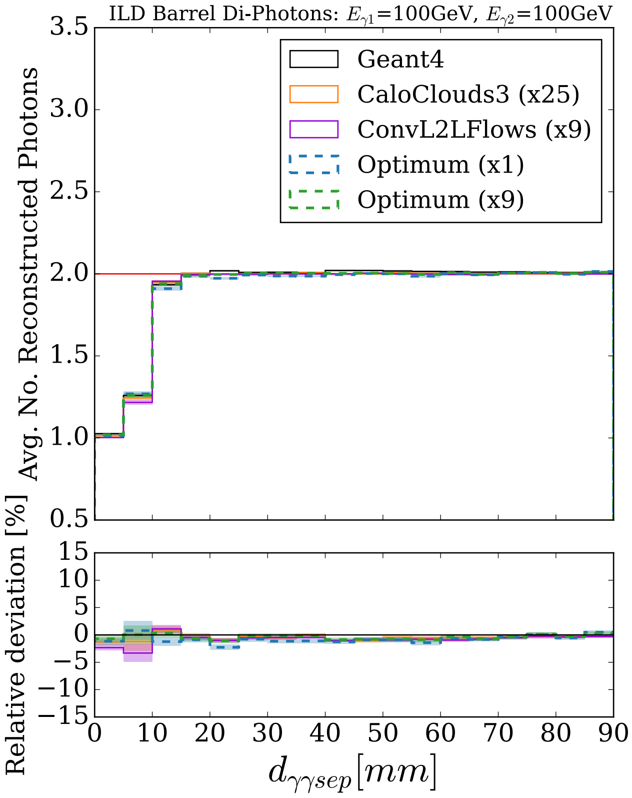


Figure 9.8: Performance of photon reconstruction for di-photon events at 100 GeV photon energy, evaluated as a function of transverse shower separation $d_{\gamma\gamma\text{sep}}$. The red horizontal line indicates the ideal case of perfect reconstruction of two photons. The shaded error bands correspond to the statistical uncertainty in each bin. The lower subplot shows the relative deviation of the number of reconstructed photons with respect to the GEANT4 reference.

heavily on the physics process for which the tool is used, a physics process is chosen that provides a stringent test of a fast simulation tool for electromagnetic showers. To this end, hadronic decay modes of the tau lepton in the process $e^+e^- \to \tau^+\tau^-$ are selected. The dataset used is described in Section 9.5. The physics results are then studied in Section 9.5.

Dataset

The dataset used in this study consists of samples of the process $e^+e^- \to \tau^+\tau^-$ in an ILC running scenario at a center-of-mass energy of 250 GeV. MC Generator samples, provided by the ILD Software Working Group in the MC-2020 production [184], were created with WHIZARD [108] version 2.8.5. A realistic ILC beam energy spectrum and crossing angle, as well as the effects of bremsstrahlung and initial state radiation were included. All samples contained beams of 100% polarised left-handed electrons and right-handed positrons ($e_L^- e_R^+$). The decay of the tau leptons in the samples were simulated with the TAUOLA library [185].

In order to enable a direct comparison between the various shower representations and models described in Sections 9.2 and generative models, the same set of MC generator inputs were used in all cases. This means that for each case, only the detector simulation differs, removing any differences in underlying event topologies or physics processes. In addition, no background is overlaid onto events. These two choices reduce event dilution that would only serve to obscure the performance of the calorimeter shower simulators.

Several selection criteria were put on the events, to further enhance the sensitivity of the analysis to the performance of the calorimeter shower simulators. It was required that all events contained at least one π^0 produced in a tau decay, with the π^0 then decaying into two photons. Both of the photons were required to have an energy above 5 GeV and to satisfy the geometry region trigger described in Section 8.1.4. Detector simulation was then run for each of the various shower representations and models described in Sections 9.2 and generative models. The software configuration employed, used the DDML implementation described in Section 8.1.4. All photons with an energy above 5 GeV incident on the electromagnetic calorimeter which passed the geometry region trigger were then simulated with the appropriate simulator.

For each calorimeter shower simulator, samples were created with three different random seeds for the detector simulation. This provides a means of estimating the uncertainties on post-reconstruction physics observables that would otherwise be difficult to estimate, given the high level of correlation arising from using identical generator level input. In total this meant that samples, consisting of 3 random seeds each containing 6791 events, were generated for GEANT4, each of the optimal generators OPTIMUM (X1), OPTIMUM (X9), OPTIMUM (STEPS) and for both the CALOCLOUDS3 and L2LFLOWS models. The standard ILD reconstruction chain was then applied to each sample.

Results

The post-reconstruction performance of the various simulation approaches for the full physics benchmark based on π^0s produced in hadronic decays of the tau lepton is presented. These results are split into an investigation into the global performance of the π^0 reconstruction in the process $e^+e^- \to \tau^+\tau^-$ in Section 9.5, followed by a study of the modeling of key π^0 physics observables in Section 9.5.

Global Reconstruction Performance

The evaluation begins by studying the overall quality of the reconstruction of π^0s produced in the tau decays for each shower simulator in comparison to GEANT4. For this evaluation, it is necessary to create a relation between Monte Carlo (MC) Truth particles and reconstructed particles. The weight of each relation is determined by the energy weighted contribution of each MC Truth particle to a reconstructed particle. If multiple MC Truth particles have a relation to a given reconstructed particle (or vice versa), the one with the highest weight is taken.

These relations are used to define four different categories of reconstruction quality into which a given reconstructed π^0 may fall. Firstly, nGood is defined as the number of correctly reconstructed π^0s. Secondly, nFake is defined as the number of π^0s reconstructed without being linked to an MC Truth π^0. Thirdly, nConfused is defined as the number of π^0s reconstructed where only one of the constituent photons was reconstructed correctly. Finally nMissed represents the number of MC Truth π^0s for which there was no reconstructed π^0. The total number of π^0s reconstructed, nRecoPi0, is also calculated. nFake is constrained on a per-event basis by the equation nFake = nRecoPi0 − nGood − nConfused.

The results for each of the optimum generators OPTIMUM (X1), OPTIMUM (X9), OPTIMUM (STEPS), together with the L2LFLOWS and CALOCLOUDS3 models, are shown in Figure 9.9. In each case, the results are plotted in comparison to the GEANT4 result. For each category, the results show the average over the 3 different GEANT4 random seeds, with the error calculated from the standard deviation across seeds.

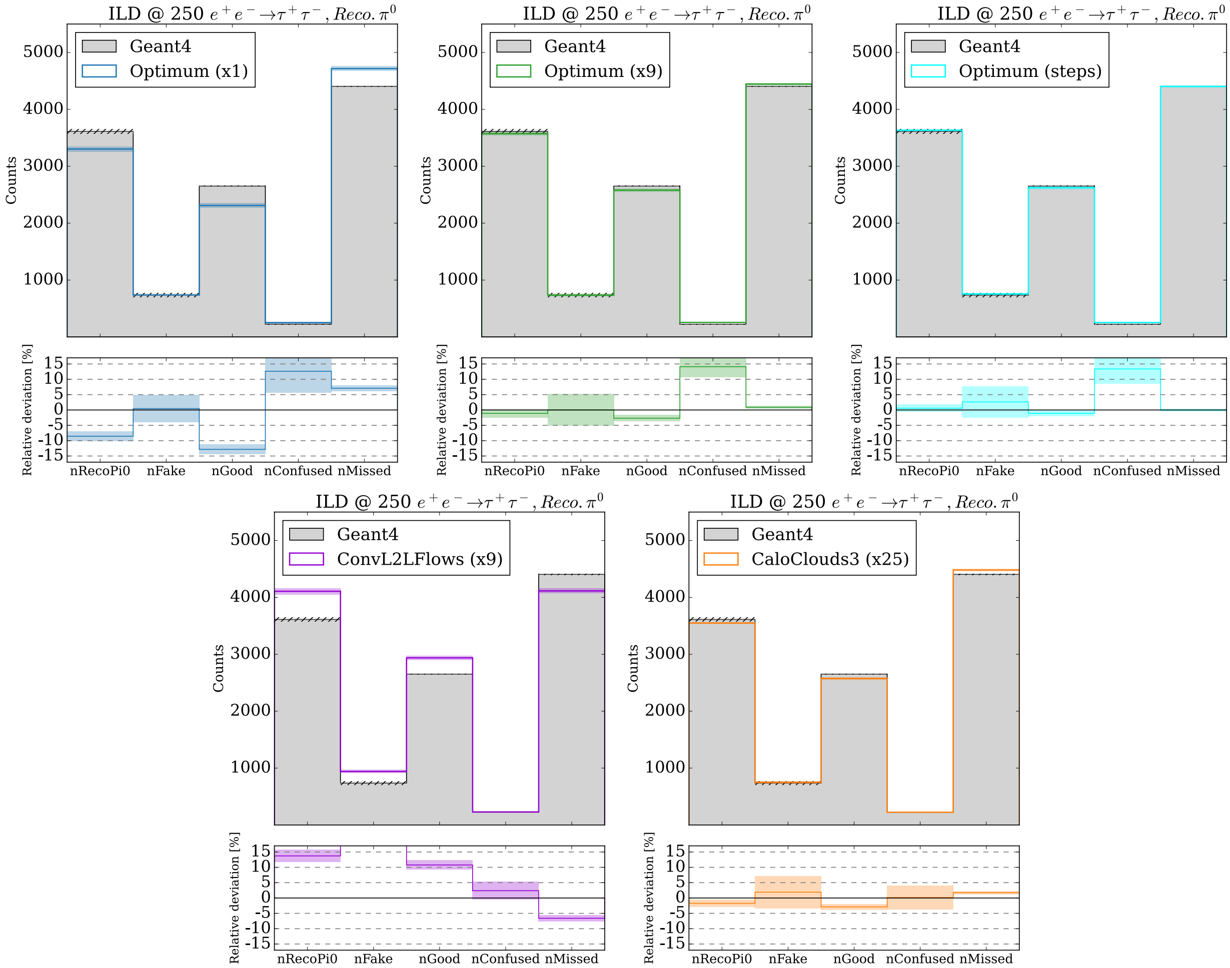


Figure 9.9: Overall reconstruction quality of π^0s produced by tau decays in the process $e^+e^- \to \tau^+\tau^-$ for the Optimum (x1) (blue), Optimum (x9) (green), Optimum (steps) (cyan) generators, and the L2LFlows (violet) and CaloClouds3 (orange) models in comparison to Geant4 (gray). The quality of the reconstruction is characterized by five different categories: nRecoPi0, the total number of π^0s reconstructed; nFake, the number of π^0s reconstructed without a corresponding π^0 in the MC Truth; nGood, the number of correctly reconstructed π^0s; nConfused, the number of π^0s reconstructed only partially correctly; and nMissed, the number of MC Truth-level π^0s that were not reconstructed. The errors on each category are derived from the standard deviation across three different random seeds used for the detector simulation. The lower panel represents the relative deviation from Geant4 in each category.

For the optimum generators, the cell-level generator Optimum (x1) shows the clearest mismatches. Large deviations from the Geant4 samples are observed in all categories except nFake. This includes on average almost 10% fewer π^0s being reconstructed in total. On average more than 10% fewer π^0s are correctly reconstructed, while many more π^0s are reconstructed incorrectly (nConfused is on average more than 10% greater, while nMissed is on average more than 5% larger) with respect to Geant4. It should be noted that while nConfused exhibits a high relative deviation, nConfused itself is low meaning that the absolute deviation is small.

Increasing the granularity of the resolution drastically decreases the deviations with respect to the GEANT4 samples, with the OPTIMUM (X9) generator showing much closer agreement. However, deviations are still clearly visible, with on average fewer nGood π^0s relative to GEANT4 and a similar excess in nConfused as was observed for OPTIMUM (X1).

At the step-level resolution in OPTIMUM (STEPS), the average for each category is consistent with GEANT4 within error except for nConfused. Again, the deviation of nConfused relative to GEANT4 is of a similar magnitude to that found for the OPTIMUM (X1) and OPTIMUM (X9) representations. These results demonstrate that increasing the granularity of the representation past that of the detector readout results in improved physics performance for this use case.

Turning to the performance of the generative models, L2LFLOWS shows major differences to GEANT4 in a large number of categories, including relative deviations in excess of 10% for nRecoPi0 and nGood, and in excess of 15% for nFake. The deviation with respect to GEANT4 for nMissed exceeds the 5% level, while nConfused shows the closest agreement with a relative deviation of only a few percent. These deviations are consistent with the significant discrepancies observed for single particle observables produced with the L2LFLOWS model in Section 9.3.

By contrast, the CALOCLOUDS3 model shows much closer agreement with the GEANT4 baseline, with all categories showing relative deviations only at the level of a few percent. While at first glance it may be surprising that CALOCLOUDS3 is able to achieve better performance than OPTIMUM (STEPS), for example in the nConfused observable, it should be noted that an extra calibration was applied to the CALOCLOUDS3 model as described in Section 9.3.

Physics Observable Performance

The performance of the individual simulators in terms of key π^0 physics observables is now studied. In order to perform a fair comparison, all π^0s selected are required to have been correctly reconstructed (i.e., those that fell into the nGood category described in Section 9.5). As before, the results show the average over 3 different GEANT4 random seeds, with the error calculated from the standard deviation across seeds. It should be noted that some of these distributions are already constrained by the selection of π^0s that fall into the nGood category. This is due to quality criteria imposed by the π^0 reconstruction procedure, which includes performing a constrained kinematic fit [106].

Firstly, the invariant mass of the reconstructed π^0, $M_{\pi^0\mathrm{Rec}}$, is calculated, which is equivalent to the invariant mass of the di-photon system $M_{\gamma\gamma}$ given by

$$M_{\gamma\gamma} = \sqrt{2E_iE_j(1-\cos(\eta))}, \tag{9.1}$$

where E_i is the reconstructed energy of photon i, E_j is the reconstructed energy of photon j, and η is the opening angle between their reconstructed flight directions. The invariant mass distributions for GEANT4, each of the optimum generators, and both the L2LFLOWS and CALOCLOUDS3 models are shown in Figure 9.10 (left). All models show broad agreement within the stated uncertainties around the bulk of the distribution. Larger relative deviations appear for all optimum generators and both models in the tails of the distribution, although the increasing errors on the ratio make the exact discrepancy less clear.

The next observable studied is the difference between the reconstructed energy of the π^0 and the energy of the corresponding MC particle, $E_{\pi^0\mathrm{Rec}} - E_{\pi^0\mathrm{MC}}$, which is shown in Figure 9.10 (right). It can be seen that in comparison to GEANT4 the reconstructed energy tends to be slightly biased towards lower energies than that of the MC particle, although this deviation

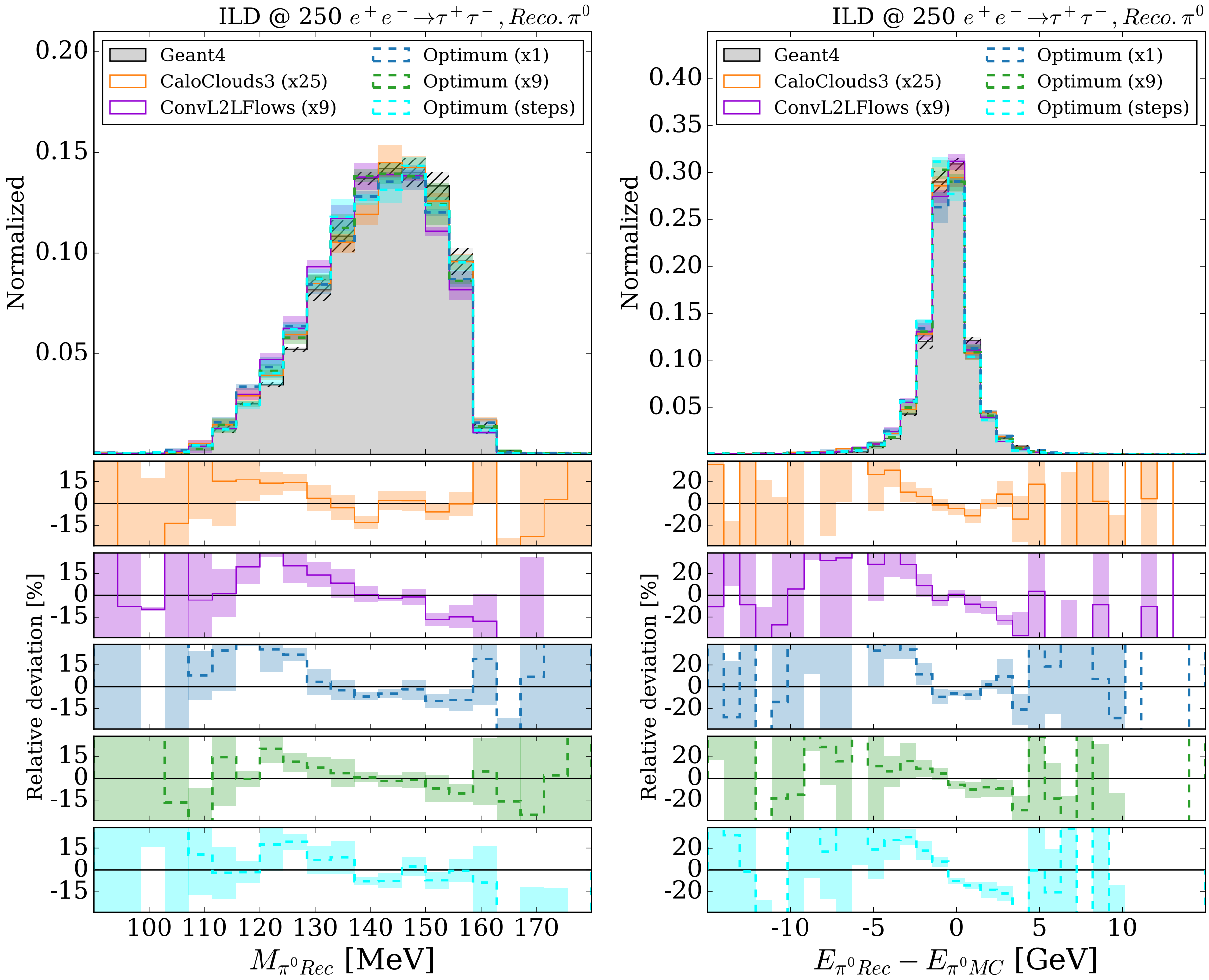


Figure 9.10: Key physics observables for π^0s produced by tau decays in the process $e^+e^- \to \tau^+\tau^-$: the reconstructed π^0 invariant mass $M_{\pi^0 Rec}$ distribution (left), and the difference between the reconstructed energy of the π^0 $E_{\pi^0 Rec}$ and the corresponding MC Truth π^0 $E_{\pi^0 MC}$ (right). Each of the optimal shower generators OPTIMUM (x1) (blue), OPTIMUM (x9) (green) and OPTIMUM (STEPS) (cyan), as well as the two generative models CALOCLOUDS3 (orange) and L2LFLOWS (violet) are shown, with the GEANT4 reference shown in gray. In both cases, the errors are derived from the standard deviation across three different random seeds used for the detector simulation. The lower panels in each case show the relative deviation from GEANT4.

is heavily suppressed by the large magnitude of the uncertainties. This effect appears to be strongest for OPTIMUM (STEPS) and L2LFLOWS, which correlates with the shifts in linearity for single photons observed for these approaches in Section 9.3.

Finally, the reconstructed angular differences for the π^0 in both the θ, $\theta_{\pi^0\mathrm{Rec}} - \theta_{\pi^0\mathrm{MC}}$, and ϕ, $\phi_{\pi^0\mathrm{Rec}} - \phi_{\pi^0\mathrm{MC}}$, directions are shown in Figure 9.11. For the distribution in θ, the clearest deviation occurs for the OPTIMUM (x1) representation at the level of the detector readout, which produces a noticeably broader distribution with relative deviations quickly exceeding the 30% level away from the core of the distribution. Clear mismodelings from the L2LFLOWS model are also present in both the θ and ϕ distributions, with relative deviations around the 20% level around the core of the distributions. The other models and generators show good

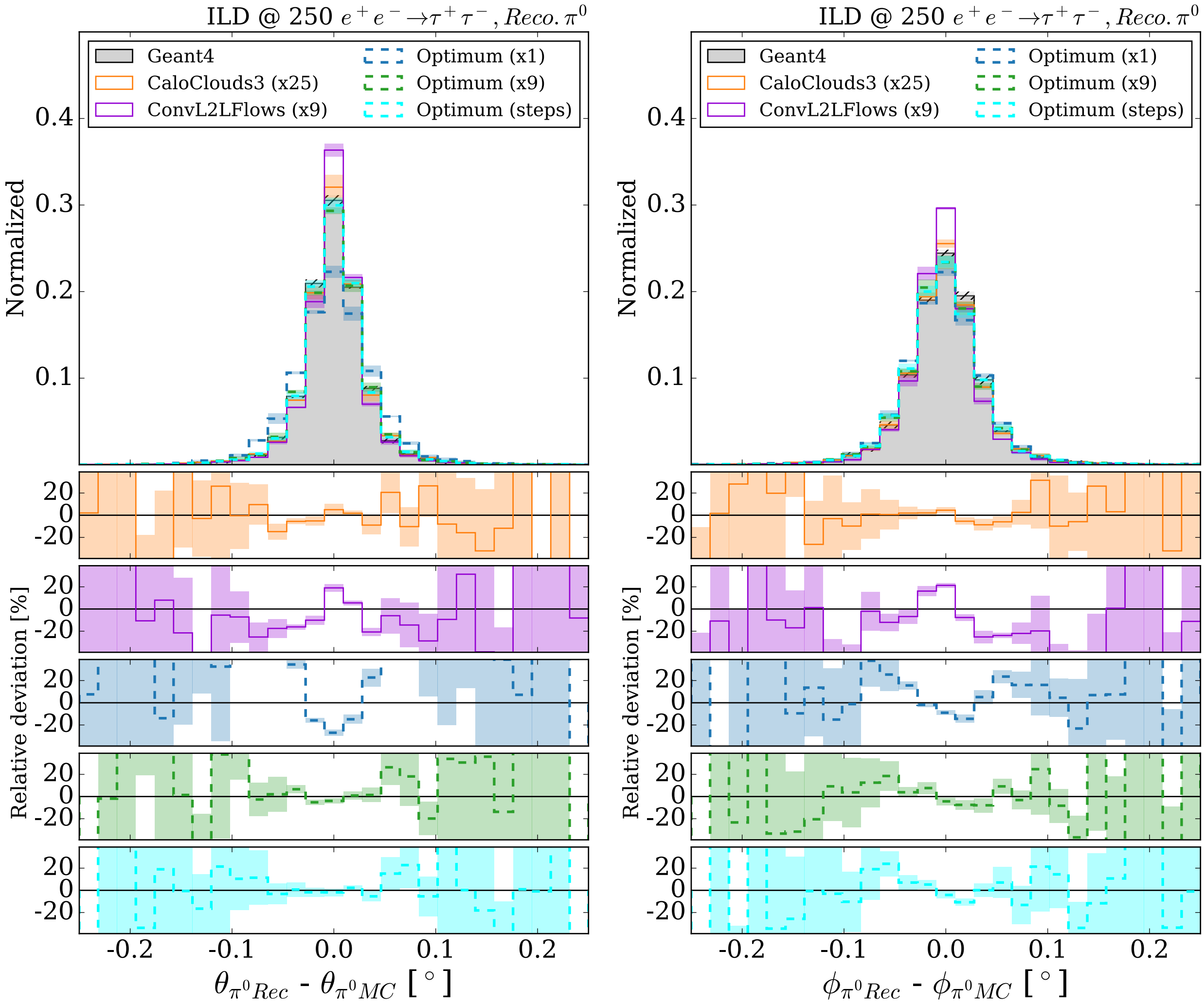


Figure 9.11: Reconstructed angular deviations for π^0s produced by tau decays in the process $e^+e^- \to \tau^+\tau^-$: the difference between the reconstructed $\theta_{\pi^0\mathrm{Rec}}$ and MC Truth $\theta_{\pi^0\mathrm{MC}}$ direction in θ (left), and the difference between the reconstructed $\phi_{\pi^0\mathrm{Rec}}$ and MC Truth $\phi_{\pi^0\mathrm{MC}}$ direction in ϕ (right). Each of the optimal generators OPTIMUM (X1) (blue), OPTIMUM (X9) (green) and OPTIMUM (STEPS) (cyan), as well as the two generative models CALOCLOUDS3 (orange) and L2LFLOWS (violet) are shown, with the GEANT4 reference shown in gray. In both cases, the errors are derived from the standard deviation across three different random seeds used for the detector simulation. The lower panels in each case show the relative deviation from GEANT4.

agreement with GEANT4 across both angular distributions, with deviations typically within the uncertainties.

9.6 Computational Performance

To quantify the speed advantage of the fast simulators in the full ILD software chain, the wall-clock time per shower on a single CPU core of an AMD EPYC™ 7402 is measured. The measurements for GEANT4 and for all generative models were taken on the same machine and software setup. As a result of the integration of the models using the DDML library described in Chapter 8, all timing measurements are directly comparable to those of Geant4, thanks to the use of an identical software configuration, including overheads such as hit placement in the detector geometry. This enables us to perform a fair and realistic timing benchmark, which is not possible without model integration.

Figure 9.12 (left, top) reports the time per shower as a function of photon energy 10–100 GeV. The Geant4 baseline lies at the few second level and grows linearly with energy, reflecting the increasing number of interaction steps. CaloClouds3 shows a similar trend, but with a much smaller slope. By contrast, the grid-based models, L2LFlows and BiBAE, show no energy dependence and remain flat across the full range.

Complementary to the time per shower as a function of photon energy, Figure 9.12 (right) shows the average single core throughput over 10–100 GeV, normalized to the Geant4 baseline (×1). Single-shower simulation speed-up factors of ×129.29 for CaloClouds3, ×11.66 for BiBAE, and ×1.20 for L2LFlows are observed across this range of incident photon energies. The striking gain for CaloClouds3 reflects its lightweight point-wise inference.

Overall, the scaling behavior is favorable for grid-based models as their inference time is constant with respect to shower energy because the computational cost is fixed by the constant voxelized volume. As a consequence, the relative speed-up over Geant4 grows with energy. At the low-energy end of the spectrum, where Geant4 showers contain fewer steps, the absolute latency gap narrows and the grid models can approach Geant4 in wall-clock time. Toward higher energies, where Geant4 scales approximately linearly with the number of interaction steps, the flat cost of L2LFlows (and other grid models such as the BiBAE) becomes increasingly advantageous.

By contrast, the point-cloud based CaloClouds3 shows a mild energy dependence, with its simulation time rising with energy as the number of generated points grows with incident energy. However, the slope remains well below that of Geant4, yielding a more uniform though less asymptotically dramatic speed-up across the full range. For completeness, it should be noted that model initialization (loading weights, caching etc.) is excluded from the timings shown, as including it only affects small sample sizes and does not change the observed scaling trends.

In addition to inference speed, the memory footprint and model size are important practical factors for deployment within large-scale simulation workflows. The compiled CaloClouds3 model has a total weight size of approximately 27 MB, corresponding to a memory footprint of 198 MB during inference. By contrast, the grid-based L2LFlows model is substantially larger, with a compiled weight size of 2.5 GB and a peak memory footprint of about 4.4 GB. This reflects the higher parameter count and larger activation maps inherent to convolutional grid-based architectures.

While the scaling behavior is favorable for grid-based models – their inference time is essentially constant with shower energy because the compute is fixed by the voxelized volume – they also exhibit larger physics performance deviations at higher energies due to leakage from the bounded generation volume. As energy increases, a growing fraction of the shower can reach the box boundary and leak out, degrading containment and biasing observables (see Section 9.3 and the single particle profiles in Section 9.3). Mitigations such as expanding the generation volume reduce leakage but increase memory footprint and latency, reducing part of the speed advantage.

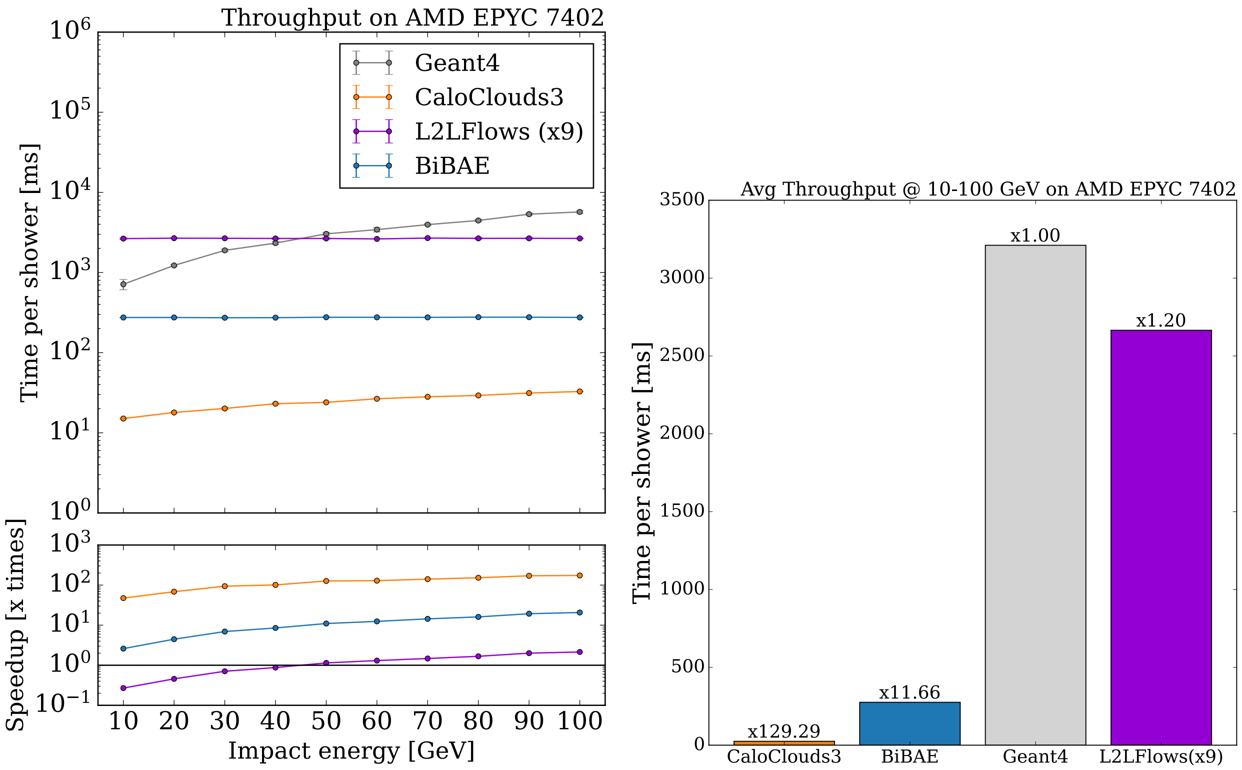


Figure 9.12: Single particle timing on a single core of an AMD EPYC™ 7402 CPU within the full ILD chain. Left: wall-clock time per shower (top) and speed-up vs. Geant4 (bottom) for photons at energy range from 10 to 100 GeV with 10 GeV steps. Right: average single core throughput over 10–100 GeV, normalized to Geant4.

9.7 Systematics from Simulation Methodology

When applying generative models trained at a single reference location to other positions within the detector, systematic biases can arise from differences between the idealized training geometry and the realistic detector structure. This section examines these effects by studying the spatial distribution of deposited energy across the ILD ECAL barrel.

Figure 9.13 shows the average reconstructed energy pattern for photon showers, visualized as YZ projections of the mean energy per voxel across 90,000 events. The vertical direction Y corresponds to the calorimeter layers, while Z indicates the position within the upper ECAL barrel in mm. All optimal shower generators and generative models reproduce the overall longitudinal and transverse structure observed in the Geant4 reference, including layer-dependent modulations and geometrical features of the detector segmentation. No significant deviations are visible at this scale, demonstrating that all models are properly integrated into the detector geometry and capture the overall event topology with high fidelity.

All methods exhibit a characteristic concave curvature in the reconstructed energy pattern, a direct consequence of the angular distribution of incident photons. Photons entering at non-normal angles traverse a larger effective thickness of absorber material, causing earlier shower development and thus higher energy deposits in shallower calorimeter layers (lower Y indices). This leads to the observed bending of the average energy profile across the stave.

In addition, a secondary increase in energy is visible in the last layers ($Y > 20$) across all models. This is attributed to the thicker absorber layers implemented in the final third of the ILD ECAL stack, which result in a small but systematic increase in energy deposition per layer in this region.

In contrast, the relative per-voxel energy difference maps shown in Figure 9.14 reveal small but systematic biases. To the right of each panel, zoomed-in views highlight regions near module boundaries and absorber gaps. Because all surrogate models were trained on idealized geometries without inactive regions, they tend to overestimate the early energy deposition. The showers in the surrogate models start slightly earlier since the absorber material is present in every layer homogeneously in the idealized geometry. In the realistic ILD geometry, however, the presence of structural gaps in the absorber delays the shower onset, allowing the energy to penetrate deeper. This results in a characteristic underestimation of energy in the later layers for all surrogates when compared to the full Geant4 simulation.

These effects originate from the interplay between the idealized local training setup and the complex global detector structure. They represent geometry-dependent systematic biases that persist even for the most accurate surrogate, Optimum (steps), and define a practical upper limit on achievable agreement when a single conditional model is reused across the barrel.

Because the affected regions are spatially confined, extending the approach with localized trainings or lightweight region-specific calibrations offers a straightforward path to further reduce the remaining discrepancies.

9.8 Observable Comparison Summary

Table 9.1 provides a concise overview of the agreement between the evaluated surrogate models and the Geant4 reference across all benchmark observables. The comparison employs the Jensen-Shannon (JS) divergence and the mean absolute error (L_1) as quantitative metrics, both measuring the deviation of a model's reconstructed distributions from those obtained with Geant4. Lower values correspond to better agreement.

Uncertainties are estimated from repeated evaluations using identical event samples simulated with different random seeds, for both Geant4 and the surrogate models.

Among the reference optimal shower generators, Optimum (x1)– trained directly at detector readout granularity – shows systematically larger deviations, indicating that coarse spatial representations limit achievable fidelity. Progressively finer geometrical resolution in Optimum (x9) and Optimum (steps) significantly improves the match, providing an effective upper bound on achievable performance for surrogates.

9.9 Conclusions

This chapter presented the first comprehensive evaluation of generative calorimeter surrogates under full physics conditions, incorporating realistic detector geometry, particle flow reconstruction, and diverse physics processes. The integration of generative models into the DD4hep software ecosystem through the DDML library enables production-ready benchmarking and provides access to reconstruction-level physics observables, addressing a critical gap in the validation of fast simulation approaches.

New reconstruction-level benchmarks have been introduced for the evaluation of generative models designed for electromagnetic shower simulation in highly granular calorimeters. These include a first multi-particle benchmark involving di-photon separations and a full physics benchmark based on hadronic decays of the tau lepton in the process $e^+e^- \to \tau^+\tau^-$. While

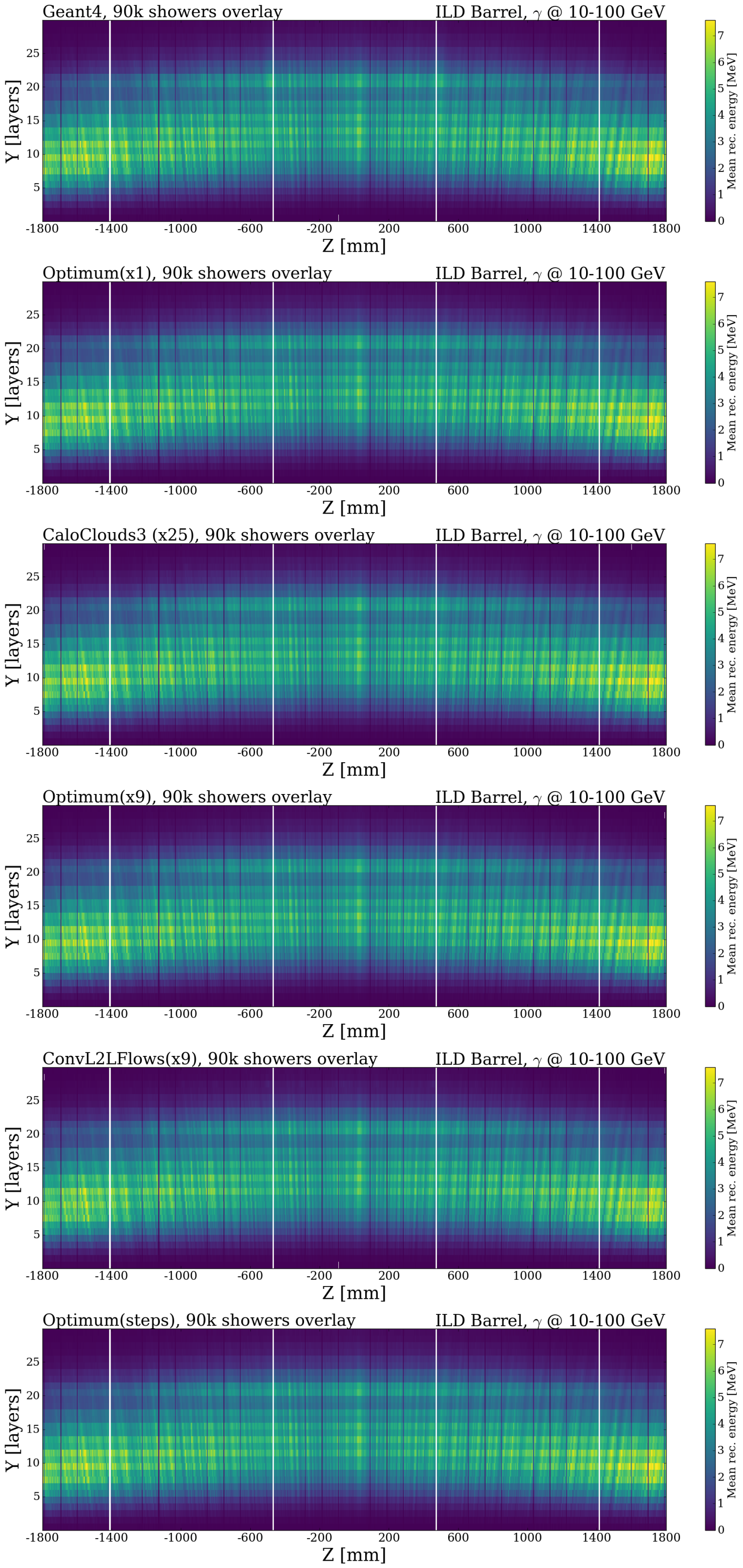


Figure 9.13: Average reconstructed energy of 90,000 photon showers incident on the upper barrel segment of the ILD ECAL, projected onto the YZ plane. Each panel shows one generator or model (name of the model in the title of each panel), with Y denoting the calorimeter depth (layer index) and Z denoting the global detector coordinate in mm. The color scale corresponds to the mean deposited energy per voxel.

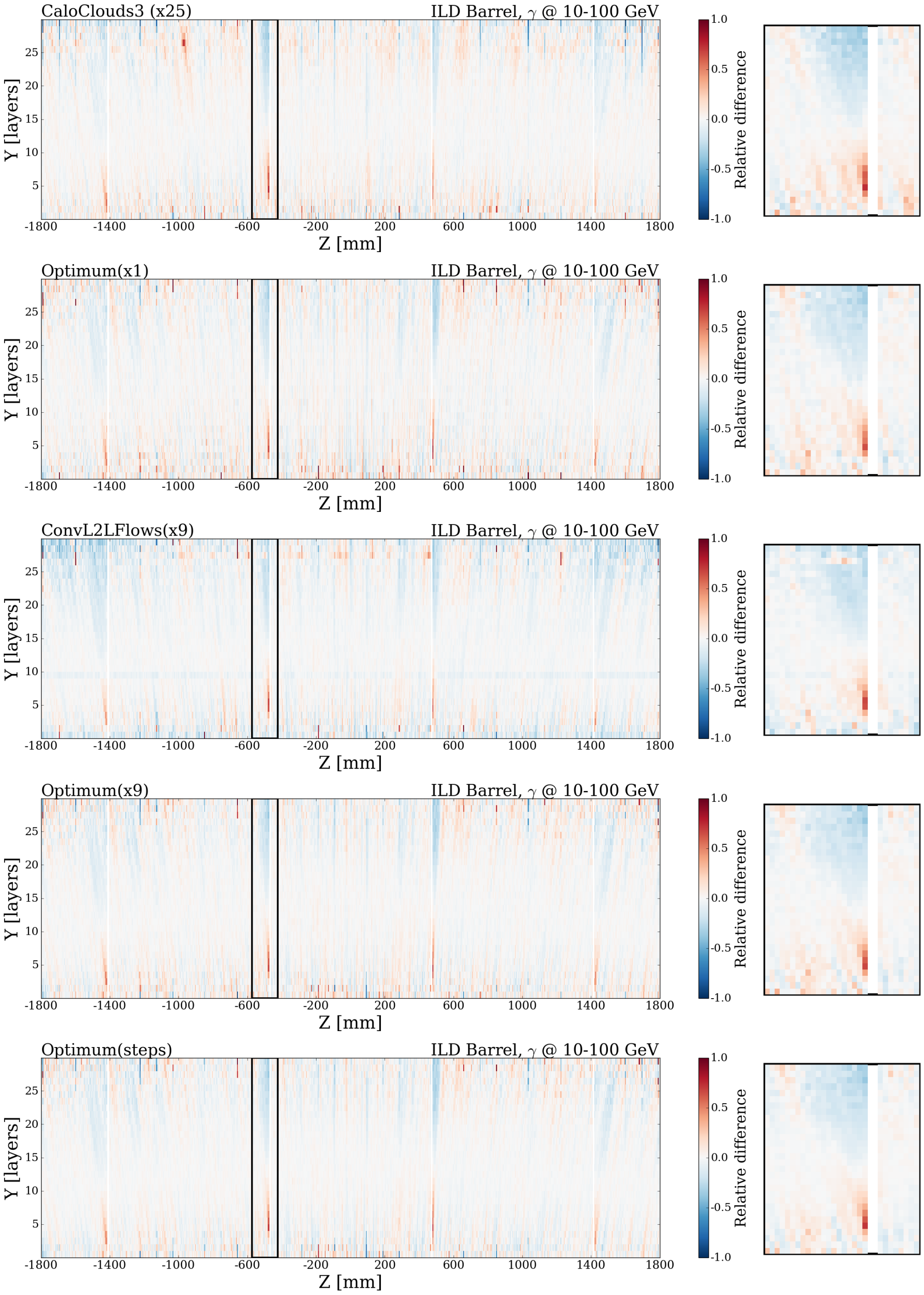


Figure 9.14: Relative per-voxel deviation in mean reconstructed energy in the ILD ECAL with respect to Geant4 for each surrogate model (name of the model in the title of each panel). The accompanying zoomed-in panels (on the right of each panel) show the regions near calorimeter module gaps. The consistent overestimation at shower onset and underestimation at larger depths arise from the absence of absorber gaps in the idealized training geometry. These localized geometry-dependent effects set a practical limit on surrogate accuracy when transferring models trained at a single reference position to the full ILD geometry.

Metric	Optimum (x1)	Optimum (x9)	Optimum (steps)	L2LFlows (x9)	CaloClouds3 (x25)
Single shower observables					
$\mathrm{JS}^{E_{\mathrm{radial(100)}}}(\times 10^{-4})$	4.71 ± 0.04	**0.04 ± 0.01**	0.06 ± 0.01	*8.37 ± 0.03*	1.11 ± 0.02
$\mathrm{JS}^{E_{\mathrm{radial(30)}}}(\times 10^{-4})$	*23.33 ± 0.16*	0.46 ± 0.02	**0.04 ± 0.01**	1.92 ± 0.06	2.24 ± 0.06
$\mathrm{JS}^{E_{\mathrm{long}}}(\times 10^{-4})$	**0.049 ± 0.013**	0.050 ± 0.013	**0.049 ± 0.013**	*0.313 ± 0.023*	0.050 ± 0.011
$\mathrm{JS}^{iPhi_{\mathrm{res(100\%)}}}(\times 10^{-4})$	**13.84 ± 1.58**	22.72 ± 2.07	24.68 ± 2.02	*342.18 ± 7.96*	48.64 ± 3.62
$\mathrm{JS}^{iPhi_{\mathrm{res(4\%)}}}(\times 10^{-4})$	*299.91 ± 7.08*	9.33 ± 1.20	**1.37 ± 0.53**	6.58 ± 1.20	16.85 ± 1.93
$\mathrm{JS}^{iTheta_{\mathrm{res(100\%)}}}(\times 10^{-4})$	44.40 ± 3.15	37.72 ± 2.71	**37.52 ± 2.82**	*553.91 ± 10.26*	374.08 ± 8.83
$\mathrm{JS}^{iTheta_{\mathrm{res(4\%)}}}(\times 10^{-4})$	*195.27 ± 5.54*	13.23 ± 1.32	**2.02 ± 0.66**	39.12 ± 2.49	32.10 ± 2.16
$\mathrm{L_1}^{\frac{\sigma_{90}}{\mu_{90}}}$	*114.33 ± 5.43*	29.84 ± 5.13	**11.63 ± 5.08**	49.31 ± 5.35	17.98 ± 5.26
$\mathrm{L_1}^{\mu_{90}}$	0.25 ± 0.04	0.29 ± 0.03	0.40 ± 0.03	*1.06 ± 0.04*	**0.09 ± 0.04**
Multi-particle observables					
$\mathrm{JS}^{\gamma\gamma_{rec.}}(\times 10^{-6})$@5GeV	*182.69 ± 27.87*	**18.09 ± 10.28**	–	28.52 ± 13.56	36.30 ± 14.21
$\mathrm{JS}^{\gamma\gamma_{rec.}}(\times 10^{-6})$@20GeV	*380.74 ± 32.14*	**10.54 ± 6.04**	–	19.56 ± 8.86	34.21 ± 12.79
$\mathrm{JS}^{\gamma\gamma_{rec.}}(\times 10^{-6})$@100GeV	5.57 ± 2.94	**2.17 ± 1.24**	–	*7.28 ± 4.48*	2.95 ± 1.77
$\mathrm{JS}^{M_{\pi^0}}(\times 10^{-4})$	23.12 ± 7.38	**13.91 ± 6.89**	25.83 ± 8.58	*32.06 ± 8.74*	22.08 ± 7.40
$\mathrm{JS}^{E_{\pi^0}}(\times 10^{-4})$	31.40 ± 9.93	**19.62 ± 6.74**	*35.88 ± 8.02*	27.24 ± 8.22	26.87 ± 8.82
$\mathrm{JS}^{\theta_{\pi^0 res}}(\times 10^{-4})$	*143.62 ± 16.41*	10.97 ± 4.94	**7.04 ± 4.27**	34.97 ± 7.63	10.69 ± 4.90
$\mathrm{JS}^{\phi_{\pi^0 res}}(\times 10^{-4})$	25.07 ± 6.98	13.90 ± 6.05	16.11 ± 6.38	*49.74 ± 10.14*	**8.68 ± 4.91**
$\mathrm{JS}^{\pi^0_{rec}}(\times 10^{-4})$	9.90 ± 1.66	0.73 ± 0.28	**0.49 ± 0.33**	*12.75 ± 2.3*	0.51 ± 0.28

Table 9.1: Quantitative comparison of models' performance relative to Geant4. Metrics are JS divergence or L_1 distance as indicated, lower values correspond to better agreement. Bold entries denote the best agreement with Geant4 per observable, italics denote the worst.

previous studies have focused primarily on single-particle shower shapes and energy distributions, the tests presented in this chapter assess model performance in the context of complete event reconstruction, examining whether generative surrogates preserve the correlations and subtle detector effects essential for realistic physics analysis.

The comparison of shower representations at different granularities reveals that building datasets directly at the level of detector readout granularity results in significant deviations in key physics observables when using realistic detector geometries. These deviations are visible across all levels of post-reconstruction observables – from single particle observables through to higher level π^0 observables in the tau decay process. As the quality of shower representation fundamentally limits the maximum achievable performance of a generative model trained on a given dataset, optimized representations that operate on shower information at a finer level than the detector readout are essential. This remains true beyond highly granular calorimeters, as demonstrated by recent work which optimized a voxelized representation for photons in the barrel region of the calorimeter of the ATLAS experiment, showing significant improvement over the current fast simulation tool [186].

Two distinct generative models with different data representations were evaluated in this study. L2LFlows is designed to operate on a fixed grid structure, while CaloClouds3 employs a point cloud representation. Due to the limitations of training directly on detector readout, both models are trained on finer granularity shower representations. However, this requirement has greater impact on the L2LFlows model, which must use a restricted granularity due to limitations imposed by the regular grid structure. As a result, the L2LFlows model shows poor simulation throughput relative to Geant4 and exhibits several deviations in observables, including higher-level physics quantities. A significant factor in the larger deviations observed for L2LFlows results from the constrained bounding box necessitated by the use of a regular grid, which is the major reason this approach exhibited consistently larger deviations than the corresponding Optimum (x9) generator with its far less restrictive bounding box.

By contrast, CaloClouds3 operates on a more granular and therefore more accurate shower representation while achieving more than two orders of magnitude faster simulation throughput than Geant4 for single showers with energies between $10-100$ GeV on a single CPU core. This demonstrates that point cloud representations offer not only a more efficient solution than regular grids but also enable a superior speed-accuracy trade-off for highly granular calorimeter simulation in realistic applications. The results confirm that CaloClouds3 achieves excellent agreement with Geant4 across all benchmark observables, with its performance approaching the theoretical upper bound established by optimal shower generators.

While some deviations are present in CaloClouds3 observables, particularly for single particle showers, these are suppressed in subsequent reconstruction for the physics cases studied. This emphasizes that while studying model performance at the level of single showers provides useful model insight, model performance must ultimately be judged on post-reconstruction level physics observables. The relative importance of certain shower features depends on the target downstream analysis – for example, the deviations observed in intrinsic cluster angles may be more relevant for certain searches for physics beyond the Standard Model [187]. It should be noted that the extent to which these deviations appear for CaloClouds3 depends on the approach used to determine the intrinsic angle (see [3] for more details).

Some deviations remain even for the Optimum (steps) generator, setting an upper limit for CaloClouds3 performance. These deviations are caused by shower placement into isolated regions in the geometry which feature greater irregularities, arising from the idealized training geometry lacking absorber gaps and detailed module structure. While reconstructed energy in these regions could be corrected by dedicated calibrations, improving the modeling of shower structure would require the addition of dedicated model trainings in those specific detector regions.

The DDML library introduced as part of this work provides essential functionality for integrating fast simulation models back into the software ecosystems used in particle physics experiments, enabling both placement into actual detector geometry and access to reconstruction-level observables. Being designed to be generic, the library facilitates the straightforward addition of new ML models and detector geometries. While this work focused on the ILD detector, other detectors proposed for future colliders such as FCC-ee are already supported, including the CLD detector [188] and the ALLEGRO detector [189].

The benchmark establishes that generative machine learning models, particularly point cloud-based approaches, can serve as reliable alternatives to traditional Monte Carlo simulation for electromagnetic calorimetry in highly granular detectors. Future work could investigate the case of electron or positron showers, where charged particle tracking and track-cluster associations may add increased importance to particular shower observables. To achieve significantly faster simulation throughput at the level of full physics events, addressing the challenge of highly granular hadron shower simulation will be necessary. Recent work [6] has demonstrated that the use of a diffusion-transformer mechanism in point cloud models can provide accurate modeling of hadron showers, including for single particle observables at the post-reconstruction level. However, further work is still required to develop a model which can achieve significantly faster simulation throughput, as well as to develop benchmarks similar to the ones shown here targeted at hadronic showers.

10 Conclusions

This thesis addresses the computational challenge of simulating calorimeter showers in highly granular detectors for high energy physics experiments. As experiments move toward higher luminosities and more complex detector designs, traditional Monte Carlo simulation methods based on Geant4 require substantial computational resources. The work presented here demonstrates how deep generative machine learning models can provide faster simulation while maintaining sufficient fidelity for physics measurements.

10.1 Summary of Results

This work presents CaloClouds, a novel generative model for fast electromagnetic shower simulation in highly granular calorimeters, developed and evaluated using the silicon-tungsten electromagnetic calorimeter of the International Large Detector [42] as a case study. The model represents calorimeter showers as point clouds (sets of energy-weighted coordinates with continuous spatial positions) rather than fixed voxel grids. This approach addresses limitations of previous grid-based methods and allows for geometry-independent simulation within regions of consistent material structure.

The CaloClouds model comprises two components, ShowerFlow and a diffusion model. ShowerFlow is a normalizing flow that generates global shower properties such as per-layer energies and energy deposition counts, and a diffusion model generates individual energy depositions.

A crucial contribution of this work is the introduction of *optimal shower generators*, idealized references constructed by projecting Geant4 simulation steps onto virtual grids of varying resolution without generative modeling. These benchmarks allow for separation between the limitations imposed by data representation choices and those arising from generative model performance. Comparisons with these references show that CaloClouds3 achieves fidelity close to the finest-grained optimal generator (Optimum (steps)), suggesting that the model performs near the limits imposed by the data representation. Comparisons with coarse-grained optimal generators reveal limitations of fixed-grid representations that are independent of model architecture.

The model has been integrated into the ILD simulation and reconstruction software chain through the DDML library [168], enabling evaluation with standard reconstruction tools including Pandora particle flow algorithms [111]. Physics benchmarks spanning single-photon observables, di-photon separation, and π^0 reconstruction in $\tau^+\tau^-$ events show state-of-the-art agreement with Geant4 for most observables. These benchmarks include both shower-level metrics and reconstructed particle properties.

On CPU hardware, CaloClouds3 achieves more than 120-fold speedup relative to full Geant4 simulation, while on GPU hardware, speedups of several thousand-fold are achievable [190]. These speedups, combined with the demonstrated physics fidelity, indicate that CaloClouds provides a practical approach for electromagnetic shower simulation. The point cloud representation offers several advantages. It provides better information density than voxel-based approaches, which is particularly beneficial for highly granular calorimeters where shower data is sparse in fine grids. By avoiding fixed grid structures, the model can capture shower physics at high resolution without the computational overhead of processing large, mostly empty tensors. The point cloud approach and the point-wise formulation of the diffusion model also result in a compact model with a memory footprint of ~200 MB during inference, substantially smaller than grid-based approaches which typically require several gigabytes for the same detector. Additionally, a single trained model can be applied to different detector positions with the same longitudinal structure, reducing the need for position-specific retraining that would be required by grid-based approaches.

10.2 Outlook

The results presented in this thesis have broader implications for fast simulation in high energy physics.

Computational efficiency for future experiments: For the High-Luminosity LHC [14] and proposed future colliders such as the ILC [15], FCC [17], and CLIC [16], where simulation needs are projected to exceed available computing resources, generative models represent one possible approach. The speedup achieved could allow experiments to generate larger Monte Carlo samples while remaining within computing budgets.

Geometry independence and adaptability: By decoupling shower generation from specific detector readout geometries, the point cloud approach enables a single trained model to be applied across different positions in a detector with consistent material structure. The continuous spatial representation also reduces projection artifacts that arise when mapping fixed-grid outputs onto irregular detector geometries. The successful adaptation to the CMS High-Granularity Calorimeter [164] demonstrates the model's flexibility across different detector designs.

Validation methodology: The optimal shower generators introduced in this work provide a benchmarking methodology that can be applied to other fast simulation challenges, including hadronic shower modeling.

10.3 Future Directions

While this thesis establishes CaloClouds as an effective solution for fast electromagnetic shower simulation, several promising directions remain for future development.

Extension to hadronic showers: The point cloud approach developed for electromagnetic showers is being extended to hadronic shower simulation in combined ECAL-HCAL systems. Preliminary work with collaborators has shown initial results [6], and in very recent work [190], a new model sharing the same point cloud representation has been developed that can generate different particle types with a single model in combined ECAL-HCAL systems, achieving exceptional agreement with Geant4, though further development is required to improve inference speed. Hadronic showers present additional challenges due to their more complex physics, larger spatial extent, and greater variability. Continued development of hadronic shower models will be needed for more comprehensive fast simulation coverage.

Multi-detector deployment: While this work focused on the ILD detector, the geometry-independent nature of the point cloud approach makes it well-suited for adaptation to other detector concepts. The model has been successfully adopted for the CMS-HGCAL [191], demonstrating transferability of this approach across different detector designs. Broader deployment to additional calorimeter technologies would further validate the generality of this approach, though practical deployment will require further validation and optimization.

Foundation models and transfer learning: As more data becomes available from multiple experiments and detector designs, it may be possible to develop foundation models pre-trained on diverse calorimeter data that can be fine-tuned for specific detector configurations. Such an approach could reduce data requirements for training models on new detectors.

10.4 Concluding Remarks

This thesis demonstrates that deep generative machine learning models, when combined with appropriate data representations, can provide fast and accurate simulation of electromagnetic showers in highly granular calorimeters. The CaloClouds model achieves more than 120-fold speedup relative to Geant4 while maintaining physics fidelity across a comprehensive set of benchmarks, from single-photon observables to π^0 reconstruction in $e^+e^- \to \tau^+\tau^-$ events.

The point cloud representation addresses key limitations of grid-based approaches, offering better information density, reduced memory footprint, and geometry independence. The evaluation framework based on optimal shower generators provides a methodology for separating data representation effects from model performance, applicable to future fast simulation development.

As high energy physics moves toward more complex detectors and higher luminosities, fast simulation approaches will be essential to address computational constraints. The results presented here demonstrate that machine learning techniques can contribute meaningfully to this challenge.

Acknowledgements

I would like to begin by acknowledging the support from Deutsches Elektronen-Synchrotron DESY, where this research was conducted. The computational work was made possible through access to the Maxwell computational resources at DESY, Hamburg, Germany. Financial support was provided via the KISS consortium (05D23GU4, 13D22CH5) funded by the German Federal Ministry of Research, Technology and Space (BMFTR) in the ErUM-Data action plan. I am also deeply grateful for the support received from the Helmholtz Initiative and Networking Fund's initiative for refugees as a refugee of the war in Ukraine.

Completing this thesis over the past three and a half years would not have been possible without the contributions and support of many individuals. While I cannot mention everyone, I would like to express my gratitude to those who have been especially important to this journey.

I am profoundly grateful to my supervisors Dr. Frank Gaede and Prof. Dr. Gregor Kasieczka for their guidance, mentorship, and unwavering support throughout my doctoral studies. Their insights and encouragement have shaped both this work and my development as a researcher in essential ways. I also wish to thank Dr. Katja Krüger for generously sharing her expertise and patiently explaining the intricacies of calorimetry.

I would also like to express my deep gratitude to Dr. Oleg Bezshyyko, my undergraduate supervisor, who introduced me to the world of research and laid the foundation for my academic journey. His early guidance helped form my approach to scientific problems.

My time as a member of the FTX group and the DESY-UHH Generative working group has been incredibly enriching. These groups have provided not only a stimulating research environment but also a warm and collaborative community. I have had the pleasure of working alongside talented PhD students and postdoctoral researchers, including Peter, Thomas, Henry, Konrad, Thorsten, Erik, Sascha, Engin, William, Martina, Lorenzo, and Bohdan. I would also like to thank all other members of the FTX and DESY-UHH Generative working groups for creating such a welcoming and productive atmosphere. Thank you all for the insightful discussions and collaborative spirit.

I have also had the pleasure of supervising Konrad during his studies. I hope you learned as much from our work together as I did.

I would like to extend special thanks to Peter, Thomas, Henry, Konrad, Taras, Hanna, Bohdan, and Mykyta for dedicating their time to carefully proofread this thesis. Your feedback has been incredibly helpful.

Finally, my heartfelt gratitude goes to my family for their constant love and support. To my mother and father, thank you for always being there, for your unconditional love and protection. You have given me a home that is always a safe place to return to. To my brothers Max and Jenya, and to my sisters Toma and Masha, thank you for all the moments we share, for your friendship, and for going through everything together. Special recognition goes to Jenya, who first sparked my interest in physics and inspired me to pursue this path. To my girlfriend Mariia, thank you for your patience, understanding, and encouragement through the challenges of this journey. This achievement would not have been possible without all of you.

The writing of this thesis took place amid circumstances of profound difficulty, touching the lives of my loved ones, colleagues, and countless others around me. I consider myself immensely fortunate to have been surrounded and shielded by such extraordinary people. In turbulent times, the ability to feel safe and at peace becomes an invaluable blessing. This

sense of safety and peace is more than comfort, it is an essential condition for scientific pursuit, which requires focus and resilience in the face of uncertainty. However, this sense of security does not exist in isolation, it is made possible by the dedication and sacrifice of innumerable individuals, most of whom remain unknown to me. My debt to them is beyond expression, yet I offer my deepest gratitude, thank you.

Bibliography

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

[2] Erik Buhmann et al. "CaloClouds II: ultra-fast geometry-independent highly-granular calorimeter simulation". In: *JINST* 19.04 (2024), P04020. DOI: 10.1088/1748-0221/19/04/P04020. arXiv: 2309.05704 `[physics.ins-det]`.

[3] Thorsten Buss et al. "CaloClouds3: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation". In: (Nov. 2025). arXiv: 2511.01460 `[physics.ins-det]`.

[4] Thorsten Buss et al. "A First Full Physics Benchmark for Highly Granular Calorimeter Surrogates". In: (Nov. 2025). arXiv: 2511.17293 `[hep-ex]`.

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

[6] Thorsten Buss et al. "CaloHadronic: a diffusion model for the generation of hadronic showers". In: (June 2025). arXiv: 2506.21720 `[physics.ins-det]`.

[7] Oz Amram et al. "CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation". In: (Oct. 2024). Ed. by Claudius Krause et al. arXiv: 2410.21611 `[physics.ins-det]`.

[8] Erik Buhmann et al. "Fast Simulation of Highly Granular Calorimeters with Generative Models: Towards a First Physics Application". In: *PoS* EPS-HEP2023 (2024), p. 568. DOI: 10.22323/1.449.0568.

[9] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: 10.1016/0029-5582(61)90469-2.

[10] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[11] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[12] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys.Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 `[hep-ex]`.

[13] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys.Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 `[hep-ex]`.

[14] O. Aberle et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: 10.23731/CYRM-2020-0010.

[15] "The International Linear Collider Technical Design Report - Volume 1: Executive Summary". In: (June 2013). Ed. by Ties Behnke et al. arXiv: 1306.6327 `[physics.acc-ph]`.

[16] T. K. Charles et al. "The Compact Linear Collider (CLIC) - 2018 Summary Report". In: *CERN Yellow Rep. Monogr.* 2 (2018). Ed. by P. N. Burrows et al., pp. 1–112. DOI: 10.23731/CYRM-2018-002. arXiv: 1812.06018 `[physics.acc-ph]`.

[17] Michael Benedikt et al. *FCC-ee: The Lepton Collider*. Tech. rep. 2. Geneva: CERN, 2019. DOI: 10.1140/epjst/e2019-900045-4.

[18] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *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: 10.1016/S0168-9002(03)01368-8.

[19] J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278. DOI: 10.1109/TNS.2006.869826.

[20] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.06.125.

[21] G. Grindhammer and S. Peters. *The Parameterized Simulation of Electromagnetic Showers in Homogeneous and Sampling Calorimeters*. 2000. arXiv: hep-ex/0001020 `[hep-ex]`.

[22] G. Grindhammer, M. Rudowicz, and S. Peters. "The fast simulation of electromagnetic and hadronic showers". In: *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: 10.1016/0168-9002(90)90566-O.

[23] Sasha Glazov. "Fast Simulation of Showers in the HI SpaCal Calorimeter". In: *Journal of Physics: Conference Series* 293.1 (Apr. 2011), p. 012024. DOI: 10.1088/1742-6596/293/1/012024.

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

[25] ATLAS Collaboration; Project Managers: James Catmore, Alessandro Di Girolamo. *ATLAS HL-LHC Computing Conceptual Design Report*. Tech. rep. CERN Document Server record 2729668; submitted 3 September 2020; CERN, 2020.

[26] Anja Butter, Tilman Plehn, and Ramon Winterhalder. "How to GAN LHC events". In: *SciPost Phys.* 7 (2019), p. 075. DOI: 10.21468/SciPostPhys.7.6.075. URL: https://scipost.org/10.21468/SciPostPhys.7.6.075.

[27] Christina Gao, Joshua Isaacson, and Claudius Krause. "<tt>i- flow</tt>: High-dimensional integration and sampling with normalizing flows". In: *Machine Learning: Science and Technology* 1.4 (Nov. 2020), p. 045023. ISSN: 2632-2153. DOI: 10.1088/2632-2153/abab62.

[28] Enrico Bothmann et al. "Exploring phase space with Neural Importance Sampling". In: *SciPost Phys.* 8 (2020), p. 069. DOI: 10.21468/SciPostPhys.8.4.069.

[29] Christina Gao et al. "Event generation with normalizing flows". In: *Phys. Rev. D* 101 (7 Apr. 2020), p. 076002. DOI: 10.1103/PhysRevD.101.076002.

[30] Theo Heimel et al. "Precision-machine learning for the matrix element method". In: *SciPost Physics* 17.5 (Nov. 2024). ISSN: 2542-4653. DOI: 10.21468/scipostphys.17.5.129.

[31] Anja Butter et al. *Kicking it Off(-shell) with Direct Diffusion*. 2024. arXiv: 2311.17175 `[hep-ph]`.

[32] Aishik Ghosh et al. "Towards a deep learning model for hadronization". In: *Physical Review D* 106.9 (Nov. 2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.106.096020.

[33] Christian Bierlich et al. "Towards a data-driven model of hadronization using normalizing flows". In: *SciPost Physics* 17.2 (Aug. 2024). ISSN: 2542-4653. DOI: 10.21468/scipostphys.17.2.045.

[34] Domenico Giordano et al. "Next Generation of HEP CPU Benchmarks". In: *23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018)*. Vol. 214. EPJ Web of Conferences. Published online 17 September 2019. 2019, p. 08011. DOI: 10.1051/epjconf/201921408011.

[35] ATLAS Collaboration. *ATLAS Software and Computing HL-LHC Roadmap*. Tech. rep. CERN-LHCC-2022-005, LHCC-G-182. Submitted March 3, 2022; last modified August 23, 2023. CERN, 2022.

[36] CMS Offline Software and Computing. *CMS Phase-2 Computing Model: Update Document*. Tech. rep. CMS-NOTE-2022-008, CERN-CMS-NOTE-2022-008. Published 07 Jul 2022; 45 pages. CERN, July 2022.

[37] Erik Buhmann et al. "Hadrons, better, faster, stronger". In: *Machine Learning: Science and Technology* 3.2 (July 2022), p. 025014. DOI: 10.1088/2632-2153/ac7848.

[38] Erik Buhmann et al. "Decoding Photons: Physics in the Latent Space of a BIB-AE Generative Network". In: *EPJ Web of Conferences* 251 (2021). Ed. by C. Biscarat et al., p. 03003. ISSN: 2100-014X. DOI: 10.1051/epjconf/202125103003.

[39] Gul Rukh Khattak et al. *Fast Simulation of a High Granularity Calorimeter by Generative Adversarial Networks*. 2021. arXiv: 2109.07388 `[physics.ins-det]`.

[40] Erik Buhmann et al. "Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed". In: *Computing and Software for Big Science* 5.1 (May 2021). ISSN: 2510-2044. DOI: 10.1007/s41781-021-00056-0.

[41] P. McKeown. "Development and Performance of a Fast Simulation Tool for Showers in High Granularity Calorimeters based on Deep Generative Models". PhD thesis. Deutsches Elektronen-Synchrotron DESY, Hamburg, 2024, p. 186. DOI: 10.3204/PUBDB-2024-01825.

[42] Halina Abramowicz et al. "International Large Detector: Interim Design Report". In: (Mar. 2020). arXiv: 2003.01116 `[physics.ins-det]`.

[43] Texas Gateway. *Introduction: Structure of Matter*. https://texasgateway.org/resource/introduction-26#Figure_23_00_Structure. Accessed: 2025-11-11.

[44] Prof. Dr. Nicola Serra and University of Zurich the Physik-Institut. *Standard Model*. https://www.physik.uzh.ch/groups/serra/StandardModel.html. Last modified April 25, 2015. Retrieved from UZH Physik-Institut outreach pages. 2015.

[45] James William Rohlf. *Modern Physics from A to Z*. New York: John Wiley and Sons, 1994. ISBN: 978-0-471-57270-1.

[46] S. M. Bilenky and J. Hosek. "Glashow–Weinberg–Salam Theory of Electroweak Interactions and the Neutral Currents". In: *Physics Reports* 90 (1982), pp. 73–157.

[47] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[48] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Sept. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[49] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[50] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet gluon picture". In: *Physics Letters B* 47.4 (1973), pp. 365–368. ISSN: 0370-2693. DOI: 10.1016/0370-2693(73)90625-4.

[51] A. Pich. *Aspects of Quantum Chromodynamics*. 2000. arXiv: hep-ph/0001118 `[hep-ph]`.

[52] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[53] Q. R. Ahmad et al. "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 89 (1 June 2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301.

[54] STEPHEN P. MARTIN. "A SUPERSYMMETRY PRIMER". In: *Perspectives on Supersymmetry*, pp. 1–98. DOI: 10.1142/9789812839657_0001.

[55] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001. URL: https://doi.org/10.1088/1748-0221/3/08/S08001.

[56] I. Béjar Alonso et al., eds. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report*. Vol. 10. Accessed: 2025-xx-xx. CERN Yellow Reports: Monographs, 2020. DOI: 10.23731/CYRM-2020-0010.

[57] Howard Baer et al. *The International Linear Collider Technical Design Report - Volume 2: Physics*. 2013. arXiv: 1306.6352 `[hep-ph]`.

[58] Chris Adolphsen et al. *The International Linear Collider Technical Design Report - Volume 3.II: Accelerator Baseline Design*. 2013. arXiv: 1306.6328 `[physics.acc-ph]`.

[59] Ties Behnke et al. *The International Linear Collider Technical Design Report - Volume 4: Detectors*. 2013. arXiv: 1306.6329 `[physics.ins-det]`.

[60] Lyn Evans and Shinichiro Michizono. *The International Linear Collider Machine Staging Report 2017*. 2017. arXiv: 1711.00568 `[physics.acc-ph]`.

[61] JiJi Fan, Matthew Reece, and Lian-Tao Wang. *Possible Futures of Electroweak Precision: ILC, FCC-ee, and CEPC*. 2015. arXiv: 1411.1054 `[hep-ph]`.

[62] Jorge de Blas et al. "On the future of Higgs, electroweak and diboson measurements at lepton colliders". In: *Journal of High Energy Physics* 2019.12 (Dec. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep12(2019)117.

[63] Alexander Aryshev et al. *The International Linear Collider: Report to Snowmass 2021*. 2023. arXiv: 2203.07622 `[physics.acc-ph]`.

[64] D. Asner et al. *Top quark precision physics at the International Linear Collider*. 2014. arXiv: 1307.8265 `[hep-ex]`.

[65] H. Abramowicz et al. “Top-quark physics at the CLIC electron-positron linear collider”. In: *Journal of High Energy Physics* 2019.11 (Nov. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep11(2019)003.

[66] Pablo Garcia-Abia and Wolfgang Lohmann. *Measurement of the Higgs Cross Section and Mass with Linear Colliders*. 1999. arXiv: hep-ex/9908065 `[hep-ex]`.

[67] 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”. In: *Phys. Rev. D* 94 (11 Dec. 2016), p. 113002. DOI: 10.1103/PhysRevD.94.113002.

[68] Claude Fabienne Duerig. “Measuring the Higgs Self-coupling at the International Linear Collider”. Dissertation, Universität Hamburg, 2016. Dissertation. Hamburg: Universität Hamburg, 2016, p. 246. DOI: 10.3204/PUBDB-2016-04283.

[69] Philipp Roloff et al. “Double Higgs boson production and Higgs self-coupling extraction at CLIC”. In: *The European Physical Journal C* 80.11 (Nov. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-08567-7.

[70] Julie Munch Torndal et al. *Higgs self-coupling measurement at future e^+e^- colliders*. 2023. arXiv: 2311.16774 `[hep-ex]`.

[71] J. de Blas et al. “Higgs Boson Studies at Future Particle Colliders”. In: *JHEP* 01 (2020), p. 139. DOI: 10.1007/JHEP01(2020)139. arXiv: 1905.03764 `[hep-ph]`.

[72] D. Jeans and G. W. Wilson. “Measuring the CP state of tau lepton pairs from Higgs decay at the ILC”. In: *Phys. Rev. D* 98.1 (2018), p. 013007. DOI: 10.1103/PhysRevD.98.013007. arXiv: 1804.01241 `[hep-ex]`.

[73] Particle Data Group et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097.

[74] Trong Hieu Tran et al. “Reconstruction and classification of tau lepton decays with ILD”. In: *Eur. Phys. J. C* 76.8 (2016), p. 468. DOI: 10.1140/epjc/s10052-016-4315-2. arXiv: 1510.05224 `[physics.ins-det]`.

[75] Yasmine Israeli. *Basic Concepts of Calorimetry.* IMPRS Colloquium, Max Planck Institute for Physics, December 2016. PDF available at https://indico.mpp.mpg.de/event/4688/contributions/10913/attachments/8561/9481/Israeli_IMPRScolloquium.pdf. 2016.

[76] Christian W. Fabjan and Fabiola Gianotti. “Calorimetry for particle physics”. In: *Rev. Mod. Phys.* 75 (4 Oct. 2003), pp. 1243–1286. DOI: 10.1103/RevModPhys.75.1243.

[77] Michele Livan and Richard Wigmans. *Calorimetry for Collider Physics, an Introduction*. UNITEXT for Physics. Springer Cham, 2019. ISBN: 978-3-030-23653-3. DOI: 10.1007/978-3-030-23653-3.

[78] Lucio Cerrito. *Radiation and Detectors: Introduction to the Physics of Radiation and Detection Devices*. Graduate Texts in Physics. Cham, Switzerland: Springer, 2017. ISBN: 978-3-319-53181-6. DOI: 10.1007/978-3-319-53181-6.

[79] Vanja Morisbak and Farid Ould-Saada. “Search for New Physics with ATLAS at LHC - Z’ dilepton resonance at high mass”. In: ().

[80] Richard Wigmans. *Calorimetry: Energy Measurement in Particle Physics*. Oxford University Press, Sept. 2017. ISBN: 9780198786351. DOI: 10.1093/oso/9780198786351.001.0001.

[81] I G Knowles and G D Lafferty. “Hadronization in decay”. In: *Journal of Physics G: Nuclear and Particle Physics* 23.7 (July 1997), p. 731. DOI: 10.1088/0954-3899/23/7/003.

[82] J. S. Marshall and M. A. Thomson. *The Pandora Particle Flow Algorithm*. 2013. arXiv: 1308.4537 `[physics.ins-det]`.

[83] Nils Feege. “Low-energetic hadron interactions in a highly granular calorimeter”. PhD thesis. Hamburg U., 2011.

[84] Boruo Xu. “Detectors and Physics at a Future Linear Collider”. PhD thesis, Department of Physics, University of Cambridge. PhD thesis. University of Cambridge, 2017. DOI: 10.17863/CAM.16867.

[85] M.A. Thomson. “Particle flow calorimetry and the PandoraPFA algorithm”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 611.1 (2009), pp. 25–40. ISSN: 0168-9002. DOI: 10.1016/j.nima.2009.09.009.

[86] J. S. Marshall and M. A. Thomson. “The Pandora software development kit for pattern recognition”. In: *The European Physical Journal C* 75.9 (Sept. 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3659-3.

[87] Ties Behnke. *The ILD Detector at the ILC: Input to the European Particle Physics Strategy Update*. Tech. rep. Contribution 3295752 at Indico event 765096. Input to the European Particle Physics Strategy Update 2018–2020, 2018. URL: https://indico.cern.ch/event/765096/contributions/3295752/.

[88] H. Abramowicz et al. *The ILD Detector: A Versatile Detector for an Electron-Positron Collider at Energies up to 1 TeV*. June 2025. DOI: 10.48550/arXiv.2506.06030.

[89] F. Sauli. “GEM: A new concept for electron amplification in gas detectors”. In: *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: 10.1016/S0168-9002(96)01172-2.

[90] Y. Giomataris et al. “MICROMEGAS: a high-granularity position-sensitive gaseous detector for high particle-flux environments”. In: *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: 10.1016/0168-9002(96)00175-1.

[91] J. Repond et al. “Design and Electronics Commissioning of the Physics Prototype of a Si-W Electromagnetic Calorimeter for the International Linear Collider”. In: *JINST* 3 (2008), P08001. DOI: 10.1088/1748-0221/3/08/P08001. arXiv: 0805.4833.

[92] K. Francis et al. “Performance of the first prototype of the CALICE scintillator strip electromagnetic calorimeter”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 763 (Nov. 2014), pp. 278–289. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.06.039.

[93] V. Balagura. “Technical instrumentation R&D for ILD SiW ECAL large scale device”. In: *Journal of Instrumentation* 13.03 (Mar. 2018), pp. C03047–C03047. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/03/c03047.

[94] J. Apostolakis et al. “Description and stability of a RPC-based calorimeter in electromagnetic and hadronic shower environments”. In: *Journal of Instrumentation* 18.03 (Mar. 2023), P03035. ISSN: 1748-0221. DOI: 10.1088/1748-0221/18/03/p03035.

[95] The CALICE collaboration et al. “Construction and commissioning of the CALICE analog hadron calorimeter prototype”. In: *Journal of Instrumentation* 5.05 (May 2010), P05004–P05004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/5/05/p05004.

[96] iLCSoft. *iLCSoft: Linear Collider Software*. https://github.com/iLCSoft. GitHub repository, accessed 24 January 2026.

[97] Frank Gaede et al. *LCIO - A persistency framework for linear collider simulation studies*. 2003. arXiv: physics/0306114 `[physics.data-an]`.

[98] M Frank et al. “DD4hep: A Detector Description Toolkit for High Energy Physics Experiments”. In: *Journal of Physics: Conference Series* 513.2 (June 2014), p. 022010. DOI: 10.1088/1742-6596/513/2/022010.

[99] M. Frank et al. “DDG4 A Simulation Framework based on the DD4hep Detector Description Toolkit”. In: *Journal of Physics: Conference Series* 664.7 (Dec. 2015), p. 072017. DOI: 10.1088/1742-6596/664/7/072017.

[100] A Sailer et al. “DD4Hep Based Event Reconstruction”. In: *Journal of Physics: Conference Series* 898.4 (Oct. 2017), p. 042017. DOI: 10.1088/1742-6596/898/4/042017.

[101] F. Gaede. “Marlin and LCCD—Software tools for the ILC”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (2006). Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research, pp. 177–180. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.11.138.

[102] Huong Lan Tran et al. “Software compensation in particle flow reconstruction”. In: *The European Physical Journal C* 77.10 (Oct. 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5298-3.

[103] Taikan Suehara and Tomohiko Tanabe. “LCFIPlus: A framework for jet analysis in linear collider studies”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 808 (Feb. 2016), pp. 109–116. ISSN: 0168-9002. DOI: 10.1016/j.nima.2015.11.054.

[104] Ulrich Einhaus et al. *Implementation, performance and physics impact of particle identification at Higgs factories*. 2023. arXiv: 2311.09102 `[hep-ex]`.

[105] Bohdan Dudar et al. *Development of time-of-flight particle identification for future Higgs factories*. 2023. arXiv: 2311.04720 `[hep-ex]`.

[106] B. List, J. List, and DESY. “MarlinKinfit: An Object–Oriented Kinematic Fitting Package”. In: *LC Notes* (2009). DOI: 10.3204/PHPPUBDB-10294. URL: https://bib-pubdb1.desy.de/record/88030.

[107] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 Physics and Manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 `[hep-ph]`.

[108] W. Kilian, T. Ohl, and J. Reuter. “WHIZARD: Simulating Multi-Particle Processes at LHC and ILC”. In: *Eur. Phys. J. C* 71 (2011), p. 1742. DOI: 10.1140/epjc/s10052-011-1742-y. arXiv: 0708.4233 `[hep-ph]`.

[109] “The Phase-2 Upgrade of the CMS Endcap Calorimeter”. In: (2017). DOI: 10.17181/CERN.IV8M.1JY2.

[110] Takashi Yamanaka. “The ATLAS calorimeter simulation FastCaloSim”. In: *Journal of Physics: Conference Series*. 2011. DOI: 10.1088/1742-6596/331/3/032053.

[111] J. S. Marshall and M. A. Thomson. “The Pandora Software Development Kit for Pattern Recognition”. In: *Eur. Phys. J. C* 75.9 (2015), p. 439. DOI: 10.1140/epjc/s10052-015-3659-3. arXiv: 1506.05348 `[physics.data-an]`.

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

[113] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. “CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks”. In: *Phys. Rev. D* 97.1 (2018), p. 014021. DOI: 10.1103/PhysRevD.97.014021. arXiv: 1712.10321 `[hep-ex]`.

[114] Martin Erdmann, Jonas Glombitza, and Thorben Quast. “Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network”. In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 4. DOI: 10.1007/s41781-018-0019-7. arXiv: 1807.01954 `[physics.ins-det]`.

[115] Erik Buhmann et al. “Hadrons, better, faster, stronger”. In: *Mach. Learn. Sci. Tech.* 3.2 (2022), p. 025014. DOI: 10.1088/2632-2153/ac7848. arXiv: 2112.09709 `[physics.ins-det]`.

[116] Baran Hashemi et al. “Ultra-high-granularity detector simulation with intra-event aware generative adversarial network and self-supervised relational reasoning”. In: *Nature Commun.* 15.1 (2024). [Erratum: Nature Commun. 115, 5825 (2024)], p. 4916. DOI: 10.1038/s41467-024-49104-4. arXiv: 2303.08046 `[physics.ins-det]`.

[117] Matej Srebre et al. “Generation of Belle II Pixel Detector Background Data with a GAN”. In: *EPJ Web Conf.* 245 (2020). Ed. by C. Doglioni et al., p. 02010. DOI: 10.1051/epjconf/202024502010.

[118] Benno Käch. “Generative Modelling in High-Energy Physics”. PhD thesis. Hamburg U., 2024. DOI: 10.3204/PUBDB-2024-05063.

[119] Baran Hashemi. “Deep Generative Models for Ultra-High Granularity Particle Physics Detector Simulation: A Voyage From Emulation to Extrapolation”. PhD thesis. LMU, 2023. DOI: 10.5282/edoc.34137. arXiv: 2403.13825 `[physics.ins-det]`.

[120] Claudius Krause and David Shih. “Fast and accurate simulations of calorimeter showers with normalizing flows”. In: *Phys. Rev. D* 107.11 (2023), p. 113003. DOI: 10.1103/PhysRevD.107.113003. arXiv: 2106.05285 `[physics.ins-det]`.

[121] Claudius Krause and David Shih. “Accelerating accurate simulations of calorimeter showers with normalizing flows and probability density distillation”. In: *Phys. Rev. D* 107.11 (2023), p. 113004. DOI: 10.1103/PhysRevD.107.113004. arXiv: 2110.11377 `[physics.ins-det]`.

[122] Sascha Diefenbacher et al. “L2LFlows: generating high-fidelity 3D calorimeter images”. In: *JINST* 18.10 (2023), P10017. DOI: 10.1088/1748-0221/18/10/P10017. arXiv: 2302.11594 `[physics.ins-det]`.

[123] Thorsten Buss et al. “Convolutional L2LFlows: generating accurate showers in highly granular calorimeters using convolutional normalizing flows”. In: *JINST* 19.09 (2024), P09003. DOI: 10.1088/1748-0221/19/09/P09003. arXiv: 2405.20407 `[physics.ins-det]`.

[124] Allison Xu et al. "Generative machine learning for detector response modeling with a conditional normalizing flow". In: *JINST* 19.02 (2024), P02003. DOI: 10.1088/1748-0221/19/02/P02003. arXiv: 2303.10148 `[hep-ex]`.

[125] Vinicius Mikuni and Benjamin Nachman. "Score-based generative models for calorimeter shower simulation". In: *Phys. Rev. D* 106.9 (2022), p. 092009. DOI: 10.1103/PhysRevD.106.092009. arXiv: 2206.11898 `[hep-ph]`.

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

[127] Vinicius Mikuni and Benjamin Nachman. "CaloScore v2: single-shot calorimeter shower simulation with diffusion models". In: *JINST* 19.02 (2024), P02001. DOI: 10.1088/1748-0221/19/02/P02001. arXiv: 2308.03847 `[hep-ph]`.

[128] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[129] Tae Min Hong et al. "Nanosecond machine learning event classification with boosted decision trees in FPGA for high energy physics". In: *JINST* 16.08 (2021), P08016. DOI: 10.1088/1748-0221/16/08/P08016. arXiv: 2104.03408 `[hep-ex]`.

[130] Nemer Chiedde. "Machine learning for real-time processing of ATLAS liquid argon calorimeter signals with FPGAs". In: *JINST* 17.04 (2022), p. C04010. DOI: 10.1088/1748-0221/17/04/C04010. arXiv: 2111.08590 `[physics.ins-det]`.

[131] Gregor Kasieczka et al. "The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics". In: *Rept. Prog. Phys.* 84.12 (2021), p. 124201. DOI: 10.1088/1361-6633/ac36b9. arXiv: 2101.08320 `[hep-ph]`.

[132] Anna Hallin et al. "Classifying anomalies through outer density estimation". In: *Phys. Rev. D* 106.5 (2022), p. 055006. DOI: 10.1103/PhysRevD.106.055006. arXiv: 2109.00546 `[hep-ph]`.

[133] Claudius Krause et al. "Anomaly detection with flow-based fast calorimeter simulators". In: *Phys. Rev. D* 110.3 (2024), p. 035036. DOI: 10.1103/PhysRevD.110.035036. arXiv: 2312.11618 `[hep-ph]`.

[134] Debajyoti Sengupta et al. "CURTAINs flows for flows: Constructing unobserved regions with maximum likelihood estimation". In: *SciPost Phys.* 17.2 (2024), p. 046. DOI: 10.21468/SciPostPhys.17.2.046. arXiv: 2305.04646 `[hep-ph]`.

[135] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. "Energy Flow Networks: Deep Sets for Particle Jets". In: *JHEP* 01 (2019), p. 121. DOI: 10.1007/JHEP01(2019)121. arXiv: 1810.05165 `[hep-ph]`.

[136] Vinicius Mikuni and Benjamin Nachman. "OmniLearn: A Method to Simultaneously Facilitate All Jet Physics Tasks". In: (Apr. 2024). arXiv: 2404.16091 `[hep-ph]`.

[137] Shiqi Gong et al. "An efficient Lorentz equivariant graph neural network for jet tagging". In: *JHEP* 07 (2022), p. 030. DOI: 10.1007/JHEP07(2022)030. arXiv: 2201.08187 `[hep-ph]`.

[138] Huilin Qu, Congqiao Li, and Sitian Qian. "Particle Transformer for Jet Tagging". In: (Feb. 2022). arXiv: 2202.03772 `[hep-ph]`.

[139] Alexander Bogatskiy et al. “PELICAN: Permutation Equivariant and Lorentz Invariant or Covariant Aggregator Network for Particle Physics”. In: (Nov. 2022). arXiv: 2211.00454 `[hep-ph]`.

[140] Joosep Pata et al. “Improved particle-flow event reconstruction with scalable neural networks for current and future particle detectors”. In: *Commun. Phys.* 7.1 (2024), p. 124. DOI: 10.1038/s42005-024-01599-5. arXiv: 2309.06782 `[physics.data-an]`.

[141] Laurits Tani et al. “A unified machine learning approach for reconstructing hadronically decaying tau leptons”. In: *Comput. Phys. Commun.* 307 (2025), p. 109399. DOI: 10.1016/j.cpc.2024.109399. arXiv: 2407.06788 `[hep-ex]`.

[142] Christopher M. Bishop and Hugh Bishop. *Deep Learning: Foundations and Concepts*. Springer, 2024. ISBN: 978-3-031-45467-7. DOI: 10.1007/978-3-031-45468-4.

[143] Boris T. Polyak. “Some methods of speeding up the convergence of iteration methods”. In: *USSR Comput. Math. Math. Phys.* 4.5 (1964), pp. 1–17. DOI: 10.1016/0041-5553(64)90137-5. URL: https://www.semanticscholar.org/paper/Some-methods-of-speeding-up-the-convergence-of-Polyak/4b53e3f719ff983eef867c6d8deac5dbe38aecb4.

[144] John Duchi, Elad Hazan, and Yoram Singer. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”. In: *J. Mach. Learn. Res.* 12 (2011), pp. 2121–2159. URL: https://www.jmlr.org/papers/volume12/duchi11a/duchi11a.pdf.

[145] Geoffrey Hinton. *Neural Networks for Machine Learning - Lecture 6a: Overview of mini-batch gradient descent*. Lecture slides from Coursera course. 2012. URL: https://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf.

[146] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: (2014). arXiv: 1412.6980 `[cs.LG]`.

[147] Yann LeCun et al. “Gradient-Based Learning Applied to Document Recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791. URL: https://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf.

[148] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735. URL: https://www.bioinf.jku.at/publications/older/2604.pdf.

[149] Ashish Vaswani et al. “Attention Is All You Need”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017, pp. 5998–6008. DOI: 10.48550/arXiv.1706.03762.

[150] Ian J. Goodfellow et al. “Generative Adversarial Networks”. In: (June 2014). DOI: 10.48550/arXiv.1406.2661. arXiv: 1406.2661 `[stat.ML]`.

[151] Diederik P. Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: (Dec. 2013). DOI: 10.48550/arXiv.1312.6114. arXiv: 1312.6114 `[stat.ML]`.

[152] Anders Boesen Lindbo Larsen et al. “Autoencoding beyond pixels using a learned similarity metric”. In: (Dec. 2015). DOI: 10.48550/arXiv.1512.09300. arXiv: 1512.09300 `[cs.LG]`.

[153] Danilo Jimenez Rezende and Shakir Mohamed. “Variational Inference with Normalizing Flows”. In: (2015). Presented at the 32nd International Conference on Machine Learning (ICML 2015). arXiv: 1505.05770 `[stat.ML]`.

[154] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. “Density estimation using Real NVP”. In: (May 2016). Accepted at ICLR 2017. DOI: 10.48550/arXiv.1605.08803. arXiv: 1605.08803 `[cs.LG]`.

[155] Conor Durkan et al. "Neural Spline Flows". In: *arXiv preprint arXiv:1906.04032* (2019). Published at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada. DOI: 10.48550/arXiv.1906.04032. arXiv: 1906.04032 `[stat.ML]`.

[156] Hadi M. Dolatabadi, Sarah Erfani, and Christopher Leckie. "Invertible Generative Modeling using Linear Rational Splines". In: *arXiv preprint arXiv:2001.05168* (2020). Accepted at the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Palermo, Sicily, Italy. DOI: 10.48550/arXiv.2001.05168. arXiv: 2001.05168 `[stat.ML]`.

[157] Diederik P. Kingma et al. "Improving Variational Inference with Inverse Autoregressive Flow". In: (June 2016). DOI: 10.48550/arXiv.1606.04934. arXiv: 1606.04934 `[cs.LG]`.

[158] Ricky T. Q. Chen et al. "Neural Ordinary Differential Equations". In: (June 2018). DOI: 10.48550/arXiv.1806.07366. arXiv: 1806.07366 `[cs.LG]`.

[159] Jascha Sohl-Dickstein et al. "Deep Unsupervised Learning using Nonequilibrium Thermodynamics". In: *arXiv preprint arXiv:1503.03585* (2015). DOI: 10.48550/arXiv.1503.03585. arXiv: 1503.03585 `[cs.LG]`.

[160] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising Diffusion Probabilistic Models". In: (June 2020). DOI: 10.48550/arXiv.2006.11239. arXiv: 2006.11239 `[cs.LG]`.

[161] Yang Song et al. "Score-Based Generative Modeling through Stochastic Differential Equations". In: (Nov. 2020). DOI: 10.48550/arXiv.2011.13456. arXiv: 2011.13456 `[cs.LG]`.

[162] Yang Song et al. "Consistency Models". In: (Mar. 2023). Presented at ICML 2023. DOI: 10.48550/arXiv.2303.01469. arXiv: 2303.01469 `[cs.LG]`.

[163] Frank Gaede, Gregor Kasieczka, and Lorenzo Valente. "Cross-Geometry Transfer Learning in Fast Electromagnetic Shower Simulation". In: (Nov. 2025). arXiv: 2512.00187 `[physics.ins-det]`.

[164] *It's about time: a Point Cloud Generative Model for the CMS High Granularity Calorimeter*. CMS Performance Note CMS-DP-2025-016. Available on the CMS information server. CERN, 2025.

[165] V. Völkl et al. "The Key4hep turnkey software stack". In: *PoS* ICHEP2022 (2022), p. 234. DOI: 10.22323/1.414.0234.

[166] Geant4 Collaboration. "Geant4—a simulation toolkit". In: *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: 10.1016/S0168-9002(03)01368-8.

[167] M. Frank et al. "DD4hep: A Detector Description Toolkit for High Energy Physics Experiments". In: *J. Phys. Conf. Ser.* 513 (2014), p. 022010. DOI: 10.1088/1742-6596/513/2/022010.

[168] Frank Gaede et al. *key4hep/DDML: v0.2.0*. Version v0.2.0. Oct. 2025. DOI: 10.5281/zenodo.17475367.

[169] J. Allison et al. *Par04 Example*. https://github.com/Geant4/geant4/tree/master/examples/extended/parameterisations/Par04. Accessed: 06.08.2025. 2025.

[170] A. Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *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.

[171] ONNX Runtime developers. *ONNX Runtime*. https://onnxruntime.ai/. Accessed: 30.12.2023. 2021.

[172] A Boehnlein et al. *HL-LHC Software and Computing Review Panel, 2nd Report*. Tech. rep. Geneva: CERN, 2022. URL: http://cds.cern.ch/record/2803119.

[173] E. Bingham et al. “Pyro: Deep Universal Probabilistic Programming”. In: *J. Mach. Learn. Res.* 20 (2019), 28:1–28:6. URL: http://jmlr.org/papers/v20/18-403.html.

[174] Sascha Daniel Diefenbacher. “Topics in Generative Modeling of Particle Physics Data”. PhD thesis. Hamburg U., 2022.

[175] Markus Frank et al. “DD4hep: A Detector Description Toolkit for High Energy Physics Experiments”. In: *J. Phys. Conf. Ser.* 513 (2014). Ed. by D. L. Groep and D. Bonacorsi, p. 022010. DOI: 10.1088/1742-6596/513/2/022010.

[176] Boruo Xu. “Improvement of photon reconstruction in PandoraPFA”. In: *International Workshop on Future Linear Colliders*. Feb. 2016. arXiv: 1603.00013 `[physics.ins-det]`.

[177] Mogens Dam. “The τ challenges at FCC-ee”. In: *Eur. Phys. J. Plus* 136.9 (2021), p. 963. DOI: 10.1140/epjp/s13360-021-01894-y.

[178] Lorenzo Calibbi, Xabier Marcano, and Joydeep Roy. “Z lepton flavour violation as a probe for new physics at future e^+e^- colliders”. In: *Eur. Phys. J. C* 81.12 (2021), p. 1054. DOI: 10.1140/epjc/s10052-021-09777-3. arXiv: 2107.10273 `[hep-ph]`.

[179] Sahabub Jahedi and Abhik Sarkar. “Exploring optimal sensitivity of lepton flavor violating effective couplings at the e+e- colliders”. In: *Phys. Rev. D* 110.9 (2024), p. 095021. DOI: 10.1103/PhysRevD.110.095021. arXiv: 2408.00190 `[hep-ph]`.

[180] A. Gutiérrez-Rodríguez, C. Pérez-Mayorga, and A. González-Sánchez. “Sensitivity Estimates on the Electromagnetic Dipole Moments of the τ-Lepton at Future e^+e^- Linear Colliders”. In: *Int. J. Theor. Phys.* 61.5 (2022), p. 132. DOI: 10.1007/s10773-022-05119-5.

[181] Daniel Jeans. “Tau lepton reconstruction at collider experiments using impact parameters”. In: *Nucl. Instrum. Meth. A* 810 (2016), pp. 51–58. DOI: 10.1016/j.nima.2015.11.030. arXiv: 1507.01700 `[hep-ex]`.

[182] Boruo Xu. “Improvement of photon reconstruction in PandoraPFA”. In: *International Workshop on Future Linear Colliders*. Feb. 2016. arXiv: 1603.00013 `[physics.ins-det]`.

[183] Boruo Xu. “Detectors and Physics at a Future Linear Collider”. PhD thesis. Cambridge U., Dec. 2017. DOI: 10.17863/CAM.16867.

[184] H. Ono and A. Miyamoto. “Status of ILD new 250 GeV common MC sample production”. In: *International Workshop on Future Linear Colliders*. May 2021. arXiv: 2105.06040 `[physics.acc-ph]`.

[185] S. Jadach, J. H. Kuhn, and Z. Was. "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons". In: *Comput. Phys. Commun.* 64 (1990), pp. 275–299. DOI: 10.1016/0010-4655(91)90038-M.

[186] *Photon showers in the ATLAS fast calorimeter simulation: A voxelized dataset with minimized information loss and improved ML models*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2942061.

[187] Lawrence Lee et al. "Collider Searches for Long-Lived Particles Beyond the Standard Model". In: *Prog. Part. Nucl. Phys.* 106 (2019). [Erratum: Prog.Part.Nucl.Phys. 122, 103912 (2022)], pp. 210–255. DOI: 10.1016/j.ppnp.2019.02.006. arXiv: 1810.12602 `[hep-ph]`.

[188] N. Bacchetta et al. *CLD – A Detector Concept for the FCC-ee*. Tech. rep. Nov. 2019. arXiv: 1911.12230 `[physics.ins-det]`.

[189] Michaela Mlynarikova. "Design and performance of the calorimeter system for ALLEGRO FCC-ee detector concept". In: *EPJ Web Conf.* 320 (2025), p. 00022. DOI: 10.1051/epjconf/202532000022.

[190] Thorsten Buss et al. "AllShowers: One model for all calorimeter showers". In: (Jan. 2026). arXiv: 2601.11716 `[physics.ins-det]`.

[191] CMS Collaboration. *It's about time: a Point Cloud Generative Model for the CMS High Granularity Calorimeter*. Tech. rep. CMS-DP-2025-016, CERN-CMS-DP-2025-016. CMS Detector Performance Summary. CERN, May 2025. URL: https://cds.cern.ch/record/2932517.

[192] OpenAI. *ChatGPT*. https://chat.openai.com. Large language model. 2024.

[193] Anthropic. *Claude*. https://claude.ai. Large language model. 2024.

[194] GitHub. *GitHub Copilot*. https://github.com/features/copilot. AI pair programmer. 2024.

Declaration on oath

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.

If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI. The following gAI tools were used: ChatGPT [192], Claude Code [193], and GitHub Copilot [194].

04.02.2026

Date

Signature of doctoral candidate