A Study of Physics-motivated Deep learning based algorithms for Jet tagging at the LHC

A Thesis

submitted to

Indian Institute of Science Education and Research Pune

in partial fulfillment of the requirements for the

BS-MS Dual Degree Programme

by

Pradyun Hebbar

IISER PUNE

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

May, 2024

Supervisor: Stefan Kluth

Co-supervisor: Daniel Britzger

Certificate

This is to certify that this dissertation entitled '**A Study of Physics-motivated Deep learning based algorithms for Jet tagging at the LHC**' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents work carried out by Pradyun Hebbar at the Max-Planck-Institut für Physik under the supervision of Stefan Kluth, ATLAS Computing, Max-Planck-Institut für Physik during the academic year 2023-2024.
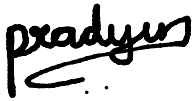
Stefan Kluth

Committee:

Stefan Kluth

Seema Sharma

Declaration

I hereby declare that the matter embodied in the report entitled "**A Study of Physics-motivated Deep learning based algorithms for Jet tagging at the LHC**" are the results of the work carried out by me at the Department of ATLAS Computing , Max-Planck-Institut für Physik in Munich , under the supervision of Stefan Kluth, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.
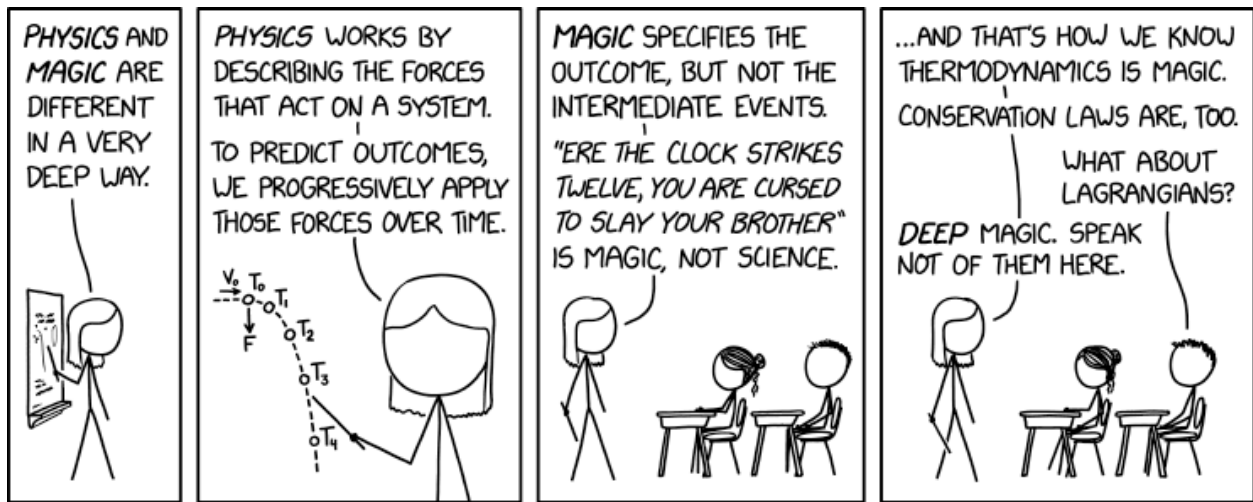
Pradyun Hebbar

IISER Roll Number – 20191114

Date – 26th March, 2024

This thesis is dedicated to the journey of self-discovery, finding what brings true fulfillment and inner peace.

Acknowledgements

I would like to express my heartfelt gratitude to my parents, Sadhana Dabake and Prasad Hebbar, whose unwavering support has been the bedrock of my journey. Their encouragement, both emotionally and financially, has enabled me to pursue my dreams in Germany. Their sacrifices and belief in my abilities have been a constant source of motivation throughout my time here.

Special thanks are due to my esteemed supervisors, Dr. Stefan Kluth and Dr. Daniel Britzger, whose guidance and expertise have shaped my research journey. Their insightful feedback, endless patience, and unwavering encouragement have been invaluable in navigating the complexities of particle physics. I am grateful for the opportunities they have provided me and for their unwavering belief in my potential.

I extend my gratitude to my dedicated group members, Josef and Johannes, for their warm welcome and camaraderie. Their collaboration, support, and willingness to share knowledge have enriched my experience in Munich. Together, we have overcome challenges, celebrated successes, and forged lasting friendships.

I am also indebted to the broader scientific community whose contributions, both direct and indirect, have propelled my research forward. From insightful discussions at conferences to valuable resources shared online, I am grateful for the collective knowledge and expertise that have shaped my understanding of particle physics.

Lastly, I would like to thank the Max Planck Institute for Physics for providing a stimulating research environment and the necessary resources for my project. The vibrant academic atmosphere and access to cutting-edge facilities have been instrumental in my academic growth.

x

Funding & Financial Support

Abstract

In this thesis, we delve into the realm of particle physics with a focus on jet tagging using deep learning algorithms. Specifically, we explore the PELICAN (Permutation Equivariant Lorentz Invariant and Covariant Aggregator Network) architecture to identify jets originating from top quarks and bottom quarks. Jet tagging is crucial for reconstructing the properties of parent particles and probing new physics phenomena beyond the Standard Model. In this thesis, we study the PELICAN architecture, verifying the claims of Lorentz symmetry and Permutation symmetry preservation in the original paper. We put PELICAN to the test in a more realistic scenario by working with the ATLAS Open Dataset and confirm PELICAN's robustness. We utilize multiple datasets to research PELICAN's performance on various input quantities to test the features that bolster PELICAN's performance. We propose incorporating 4-vector momentum data and trajectory displacement information to enhance the accuracy of jet identification. We propose novel extensions to the PELICAN architecture, including the use of spacetime displacement 4-vectors and scalar particle identification labels, to improve the tagging of heavy-flavor jets. This work not only enhances the performance of existing jet tagging algorithms but also opens new avenues for future research in the field.

Contents

Appendices   73

A ATLAS Detector   73

B Jet Representations   75

References   77

List of Tables

List of Figures

Part I

Introduction

Chapter 1

Introduction

1.1 Introduction

Particle physics is a field that delves into the fundamental constituents of matter and the forces that govern their interactions. At the forefront of this discipline lies the exploration of high-energy collisions, where particles are accelerated to extreme speeds and made to collide in controlled environments such as the Large Hadron Collider (LHC) at CERN (CERN, 2018). Through the study of these collisions, physicists aim to unravel the mysteries of the universe, from understanding the nature of dark matter and dark energy to probing the fundamental forces that shape the cosmos.

The Standard Model (Wikipedia contributors, Accessed 2024) of particle physics provides a framework for understanding the behavior of elementary particles and their interactions. It describes the electromagnetic, weak, and strong nuclear forces and classifies particles into two categories: fermions and bosons. Fermions, such as quarks and leptons, are the building blocks of matter, while bosons, including photons and W and Z bosons, mediate the fundamental forces. Despite its remarkable success in explaining a wide range of phenomena, the Standard Model has limitations and does not account for phenomena such as gravity or dark matter (Newman, 2023).

One of the crucial tasks in particle physics experiments is the identification and characterization of particles produced in high-energy collisions. Particle identification plays a

pivotal role in deciphering the underlying physics processes and discriminating between different particle types. Among the myriad particles produced in collisions, jets are of particular interest. Jets are collimated sprays of particles resulting from the fragmentation and hadronization of quarks and gluons produced in the collision. Identifying and tagging jets accurately are essential for reconstructing the properties of parent particles and probing new physics phenomena beyond the Standard Model.

In recent years, advancements in machine learning and artificial intelligence have revolutionized the field of particle physics (Kasieczka et al., 2019a). Neural networks, in particular, have emerged as powerful tools for particle identification and analysis. By training neural networks on vast datasets of collision events, physicists can extract valuable insights into the properties of particles and the dynamics of their interactions.

1.2   Motivation

This thesis focuses on identifying the source of 'collimated spray of particles' that appear in proton-proton collisions, called Jets.

Jets are the experimental signatures of quarks and gluons (partons) produced in high-energy processes such as proton-proton collisions. These partons radiate more partons as they propagate resulting in a parton shower. As quarks and gluons have a net colour charge and cannot exist freely due to colour-confinement, thus, at lower energies, they come together to form colourless hadrons, a process called hadronization. This spray of particles is called a jet. The identification of the parton that originated a jet is called jet tagging.

As jets of particles propagate through the detector, they leave signals in components such as the tracker and the electromagnetic and hadronic calorimeters. These signals are combined using jet algorithms to form a reconstructed jet. Physics analysis is usually performed on reconstructed jets.

This thesis focuses on a sub part of jet tagging called as Heavy Flavour tagging. Heavy Flavour tagging, the identification of jets originating from t, b and c quarks, is a critical component of the physics program of the ATLAS experiment (Collaboration et al., 2008) at the Large Hadron Collider (LHC). Generally, Heavy Flavour tagging is split into top tagging (for t) and flavour tagging (for b , c).
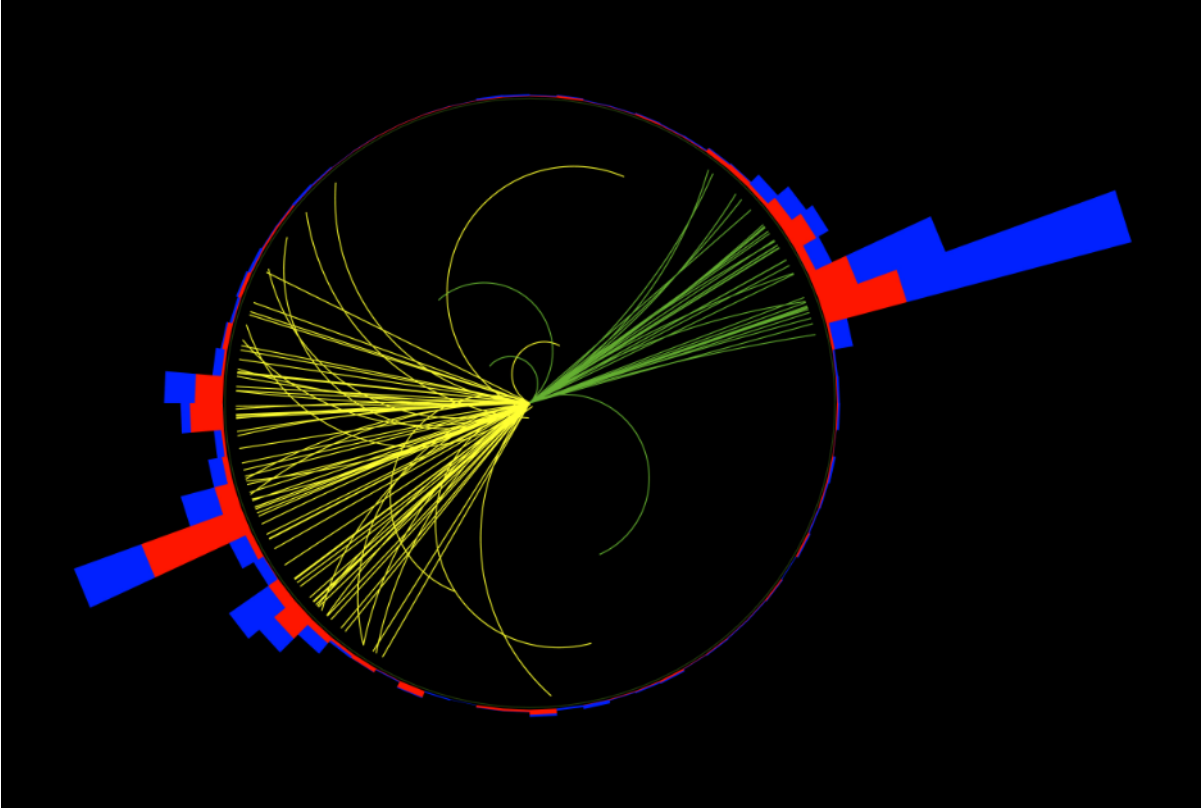
Figure 1.1: A top quark + anti-top quark event at the LHC. The quarks have hadronized into particle "jets" denoted in green and yellow. Figure has been taken from CMS Collaboration, Accessed 2024.

Flavour tagging is of particular importance for the study of the Standard Model (SM) Higgs boson and the top quark, which preferentially decay to b quarks through the channels $H \to b\vec{b}$, and $t \to bW^+$, and additionally for several Beyond Standard Model (BSM) resonances that readily decay to heavy flavour quarks (G. Aad et al., 2020). The signatures used specifically for b-tagging are -

- significant lifetime of b-hadrons, approximately 1.5 ps (10^{-12}), provides the unique signature of a secondary decay vertex which has a high mass and is significantly displaced from the primary vertex

- b hadrons may also have a tertiary decay vertex, resulting from $b \to c$ decay chains.

- the reconstructed trajectories of charged particles (henceforth simply referred to as tracks) have large impact parameters (IP) ("Graph Neural Network Jet Flavour Tagging with the ATLAS Detector" 2022)

On the other hand, for top tagging, the lifetime of the top quark ($\approx 10^{25}$s) is shorter than the hadronization time ($\approx 10^{23}$s) and much shorter than the spin decorrelation time ($\approx 10^{21}$s) (Déliot and Mulders, 2020). Therefore the top quark decays before hadronisation and its spin information is directly transferred to its decay products. Many years after its discovery, the top quark still plays a fundamental role in the program of particle physics. The top quark is the heaviest of all known elementary particles. Its mass is close to the EWSB (Electro-Weak Symmetry Breaking) scale, and it may play an important role in the understanding of the EWSB mechanism. Furthermore, the large top quark mass implies a large coupling to the Higgs boson, thus establishing a privileged link to the Higgs sector (Gallinaro, 2013). The signatures used specifically for top-tagging depend on the decay modes of the W boson-

- In the SM, the top quark decays with a branching fraction of almost 100 % to a b quark and a W+ boson $t \rightarrow bW^+$.

- The W boson will subsequently decay to either two quarks ($W \rightarrow q\bar{q}'$), which are observed as jets of particles, or to a charged lepton and a neutrino ($W \rightarrow lv$).

- In proton collisions at the LHC top quarks are dominantly produced via the strong interaction, resulting in a top quark-antiquark pair ($t\bar{t}$). But, single top (or anti-top) quark production modes via weak interaction also exist.

The technique used for top tagging and flavour tagging in the ATLAS and CMS collaborations, has been to use *High level quantity-based taggers*. These taggers are trained on a set of high level quantities which are observables (For example - Energy correlation ratios, N-subjettiness, Angularity, etc.) that can be calculated from the measured properties of the jet constituents. These quantities are hand-designed to draw out differences between signal jets and background jets *Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s}$ = 13 TeV* 2017.

In the same spirit, the natural next question is why we apply highly complex tagging algorithms to a pre-processed set of kinematic observables rather than to actual data. This question becomes especially relevant when we consider the significant conceptual and performance progress in deep learning Kasieczka et al., 2019a. Deep learning, or the use of neural networks with many hidden layers, is the tool that allows us to analyze low-level LHC data without constructing high-level observables. These are the *Constituent based, low level*
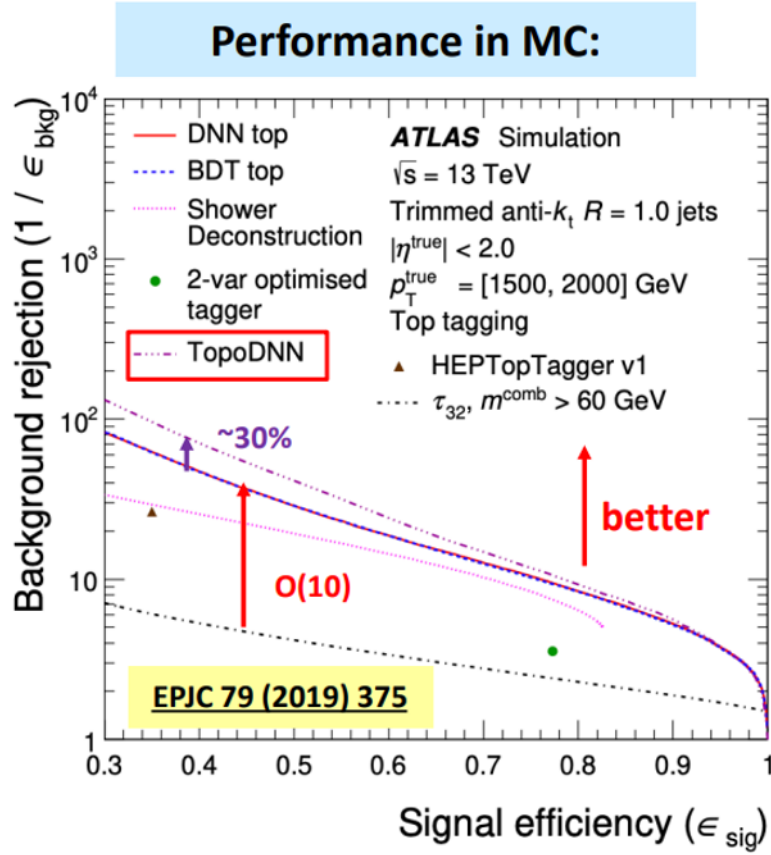
Figure 1.2: Performance of various taggers for top tagging on MC simulation. The plot shows 'High level quantity based taggers' like BDTs and DNN, in comparison to TopoDNN which is a 'Low level quantity based tagger'(Gouskos, 2020).

taggers. Studies have consistently shown that we can expect great performance improvement from low level taggers compared to the traditional taggersKasieczka et al., 2019a.

Given this background, the primary goals of my thesis are:-

- To study which deep learning based jet tagging algorithm shows state-of-the-art (SOTA) performance

- To analyze the deep learning algorithm's performance and pinpoint the features responsible for it.

- Studies have shown that introducing symmetries and physical constraints in the neural network architectures leads to a boost in the jet tagging performance Li et al., 2022. To test the performance of the deep learning algorithm on a more realistic dataset.

- To extend the same network architecture for a wider variety of use cases - flavor tagging, event tagging, 4-momenta reconstruction, etc.

In this thesis we delve into the realm of Particle Physics and Jet tagging with a focus on top tagging and bottom quark tagging using neural networks. The introductory chapter provides a foundational understanding of Particle Physics and Jet tagging, elucidating the motivations behind our focus. Following this, Chapter 2 delves into the theoretical underpinnings, exploring particle interactions, calorimeters, detectors, shower development, jet formation, measurement/reconstruction, and the pivotal process of jet tagging of particles. Moving forward, Chapter 3 introduces the application of neural networks in Particle Physics, with a particular emphasis on our chosen architecture, PELICAN (Permutation Equivariant Lorentz Invariant and Covariant Aggregator Network). Chapter 4 provides insight into the datasets utilized, offering a comprehensive understanding of their sources and characteristics. Subsequently, Chapter 5 presents the results garnered from the application of the PELICAN network to these datasets, evaluating its efficacy in top tagging (and few other particles). Finally, Chapter 6 consolidates our findings, drawing conclusions, discussing their implications, and outlining avenues for future research and refinement of our methodology. Through this structured approach, we aim to contribute to the advancement of Particle Physics and the development of Deep Learning-based algorithms for Jet tagging.

Part II

Theoretical Background

Chapter 2

Particle Interaction with Matter

To be detected, a particle must interact with the material of a detector. In this section, we explore these interactions.

A particle may interact with an atomic nucleus as its first possibility. For instance, this interaction might occur through the strong nuclear force for hadrons or via the weak interaction for neutrinos. If the particle's energy is sufficiently high, it may produce new particles, often serving as the initial step in the detection process.

Apart from these short-range interactions, a charged particle can also excite and ionize atoms along its trajectory, resulting in ionization energy losses, and emit radiation, leading to radiation energy losses. Both processes stem from the long-range electromagnetic interaction. They are important because they form the basis of most detectors for charged particles. Photons, too, are directly detected through electromagnetic interactions, and at high energies, their interactions with matter predominantly result in the production of e^+e^- pairs through the pair production process

$$\gamma \to e^+ + e^- \tag{2.1}$$

and also Compton scattering. This process must occur near a nucleus to conserve energy and momentum. All these interaction types are elaborated upon in the subsequent sections. This section relies heavily on previous literature [Leo, 1994 , Green, 2000, Virdee, 2007]

2.1   Short-range interactions with nuclei

For hadrons, the most important short-range interactions with nuclei are due to the strong nuclear force, which is important for neutral and charged particles. As an illustration, consider the simplest nucleus, a proton and an incident π^-. There are two types of possible reactions - *elastic scattering* such as

$$\pi^- + p \to \pi^- + p \tag{2.2}$$

and *inelastic scattering* such as

$$\pi^- + p \to \pi^+ + \pi^- + \pi^0 + n \tag{2.3}$$

in which the final state particles differ from those in the initial state. At high energies, many inelastic reactions are possible, most of them involving the production of several particles in the final state. The total cross-section is then given by -

$$\sigma_{total} = \sigma_{elastic} + \sum \sigma_{inelastic} \tag{2.4}$$

The probability of a hadron–nucleus interaction occurring as the hadron traverses a small thickness dx of material is given by $n\sigma_{tot}dx$, where n is the number of nuclei per unit volume in the material. Then, the mean distance traveled before an interaction occurs is given by the *collision length* -

$$l_{col} = \frac{1}{n\sigma_{tot}} \tag{2.5}$$

Neutrinos and antineutrinos can also be absorbed by nuclei, leading to weak interaction reactions of the type

$$\bar{\nu}_l + p \to l^- + X \tag{2.6}$$

where l is a lepton and X denotes any hadron or set of hadrons allowed by the conservation laws. As these proceed via the weak interaction, the cross-sections are extremely small compared to the cross-sections for strong interaction processes and thus, the *interaction lengths* are large. Such reactions are one of the basis of detecting neutrinos.

2.2 Ionization Energy Loss

Ionization energy losses are important for all charged particles, (other than electrons and positrons) at all but the highest attainable energies. When charged particles traverse through matter, they lose energy to the atomic electrons via the Coulomb interaction. The energy transferred to the electrons causes them either to be ejected from the parent atom (*ionisation*) or to be excited to a higher level (*excitation*). The theory of such energy losses was worked out by Bethe, Bloch and others in the 1930s. The Bethe-Bloch equations is -

$$\frac{dE}{dx}_{\text{ion}} = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \tag{2.7}$$

Where,

- $-\frac{dE}{dx}$: The average rate of energy loss of a charged particle per unit path length in the material.

- K: A constant.

- z: The charge of the incident particle.

- Z: The atomic number of the material.

- A: The atomic mass of the material.

- β: The velocity of the incident particle relative to the speed of light.

- γ: The Lorentz factor.

- $T_{\max}$: The maximum kinetic energy transfer in a single collision.

- I: The mean excitation energy of the material.

- m_e: The electron mass.

- c: The speed of light.

- δ: The density effect parameter.

The notable features of this formula are :-

- $\frac{dE}{dx}_{\text{ion}} \propto \frac{1}{m_e}$ shows that the energy is lost the electrons

- $\frac{dE}{dx}_{\text{ion}} \propto \frac{1}{\beta^2}$ shows that slower incident particles lose more energy

- $\frac{dE}{dx}_{\text{ion}} \propto z^2$ shows particles with more charge, lose more energy

- $\frac{dE}{dx}_{\text{minimum}}$ occurs are $\beta\gamma = 4$ (these are said to be minimum ionizing particles or MIPs)
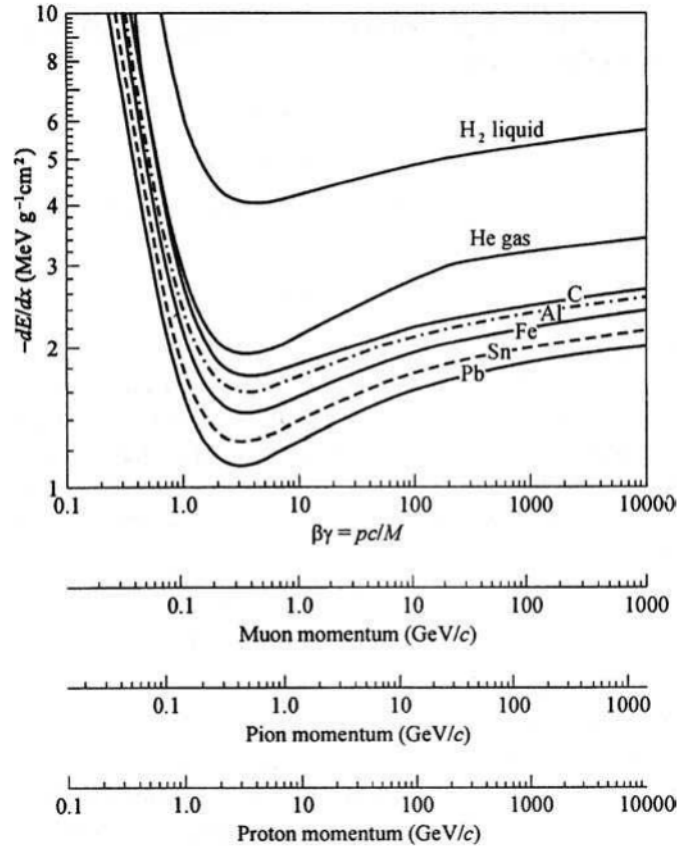
Figure 2.1: Ionization energy loss for muons, pions and protons on a variety of materials. Figure taken from (et al, 2006).

It is common practice to represent 2.7 as an *areal density*, where 2.7 is divided by mass density ρ. Examples of the behavior of 2.8 for muons, pions and protons traversing a range of materials are shown in 2.1

$$\frac{1}{\rho}\frac{dE}{dx}_{\text{ion}} \tag{2.8}$$

2.3   Radiation energy losses

When a charged particle traverses matter it can also lose energy by radiative collisions, especially with nuclei. The electric field of a nucleus will accelerate and decelerate the particles as they pass, causing them to radiate photons, and hence lose energy. This process is called *bremsstrahlung* (literally 'braking radiation' in German) and is a particularly important contribution to the energy loss for electrons and positrons. In the intense electric field of a nucleus, relativistic electrons radiate photons (bremsstrahlung), and photons can be converted into electron-positron pairs (pair creation). The dominant Feynman diagram for electron bremsstrahlung in the field of a nucleus, i.e.

$$e^- + (Z, A) \rightarrow e^- + \gamma + (Z, A) \tag{2.9}$$

is given in 2.2
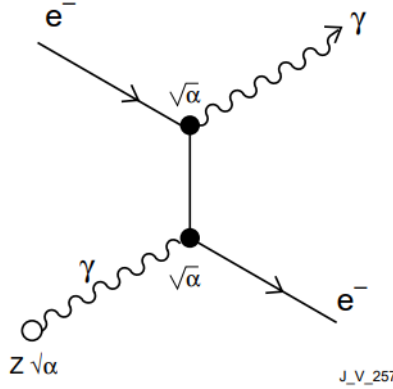
Figure 2.2: The Feynman diagram for bremstrahlung. Figure taken from (Virdee, 2007).

A detailed calculation shows that the cross-section for energy losses due to radiation is given by -

$$\rho_{\mathrm{rad}} \propto \frac{Z^2 \alpha^3}{m^2 c^4} \tag{2.10}$$

and the energy loss -

$$\frac{dE}{dx}_{rad} \propto \frac{E}{m^2} \tag{2.11}$$

From these results, we can see that radiation losses completely dominate the energy losses for electrons and positrons at high enough energies, but are much smaller than ionization losses for all particles other than electrons and positrons at all but the highest energies.

2.4   Energy loss by photons

In contrast to heavy charged particles, photons have a high probability of being absorbed or scattered through large angles by the atoms in matter. Consequently, a collimated monoenergetic beam of I photons per second traversing a layer thickness dx of matter will lose

$$dI = -\frac{I\,dx}{\lambda} \tag{2.12}$$

photons per second, which on integrating gives -

$$I(x) = I_0 e^{\frac{-x}{\lambda}} \tag{2.13}$$

where, mean free path λ is analogous to the collision length for hadronic reactions. Photons mainly lose energy via - the photoelectric effect, in which the photon is absorbed by the atom as a whole with the emission of an electron; Compton scattering, where the photon scatters from an atomic electron; and electron–positron pair production in the field of a nucleus or of an atomic electron. The corresponding cross-sections on lead are shown in 2.3
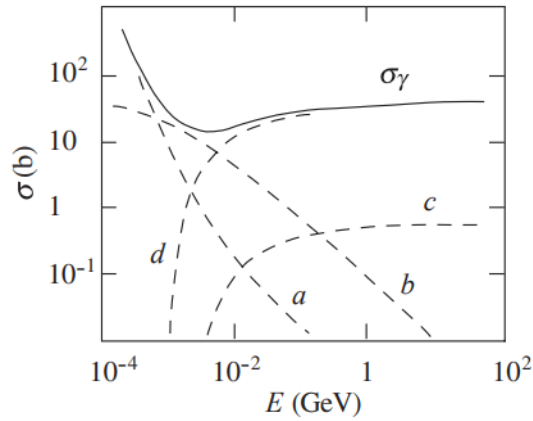
Figure 2.3: Total experimental photon cross-section on a lead atom, together with the contributions from (a) the photoelectric effect, (b) Compton scattering, (c) pair production in the field of the atomic electrons and (d) pair production in the field of the nucleus. Figure has been taken from (et al, 2006).

Chapter 3

Shower development and Jet formation

3.1 Shower Development

Now that we know how particles interact with matter, we can learn about the basic blocks in a jet - *particle showers*. A particle shower refers to a cascade of secondary particles produced when a high-energy particle interacts with matter. There are two types of showers -

- Electromagnetic Showers

- Hadronic Showers

3.1.1 Electromagnetic shower

A cascade of photons and electrons develops when a highly energetic photon or electron is incident on an absorber material. 3.1 shows a schematic of an electromagnetic shower produced by a photon. The photon pair-produces an electron-positron pair and these, in turn, produce photons via bremsstrahlung. With an increasing depth of the absorber material, the number of particles in the electromagnetic shower increases till some depth (called the shower maximum) while the mean energy carried by the particles decreases.

3.2 shows the longitudinal development of the electron shower in different materials. The multiplication of the shower particles continues until the energy carried by the particles is below
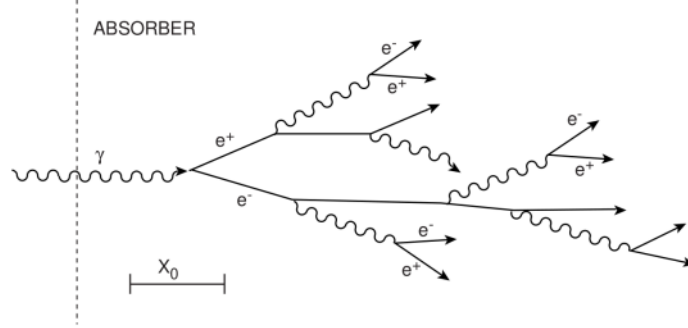
Figure 3.1: A schematic of an electromagnetic shower induced by a photon. Figure taken from (Virdee, 2007).

the critical energy (ϵ) of the material and the energy deposition increases accordingly. After the shower maximum has been reached, the energy deposition falls off exponentially. At this point, the photons lose energy mainly through the compton scattering or the photoelectric effect and the electrons(/positrons) lose energy through ionization of the absorber material. The shower maximum is deeper for the high Z material (low critical energy) as the multiplication of shower particles continues down to lower energies.
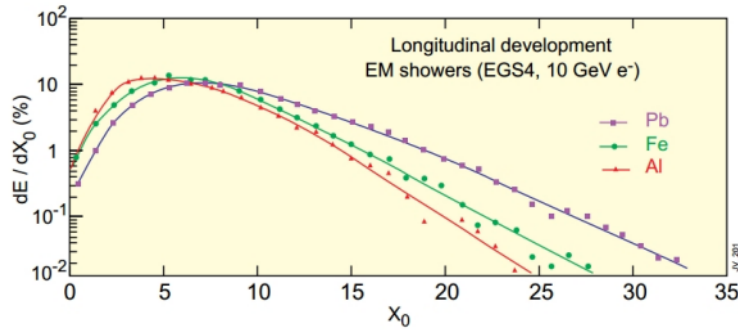
Figure 3.2: Simulation of the longitudinal development of 10 GeV electron showers in aluminum(Al), iron (Fe) and lead (Pb). Figure taken from (Virdee, 2007).

The electromagnetic shower spreads in the lateral direction due to multiple scattering of the secondary e^-, e^+ and emission of photons away from the shower axis. The lateral extent of the electromagnetic shower for different materials can be specified in terms of Moliere radius (R_M), which is the radius of the cylinder of infinite length, which contains 90 percent of the shower energy.

3.1.2 Hadronic shower

As mentioned in 2.1, a highly energetic hadron (charged or neutral) deposits its energy in a material by interacting with its nuclei via strong nuclear interactions. A charged hadron continues to lose its energy through ionization until it undergoes such a hadronic interaction. The interaction of an incoming hadron with absorber nuclei leads to multiparticle production. The secondary hadrons in turn interact with further nuclei leading to a growth in the number of particles in the cascade. Nuclei may breakup leading to spallation products. The particle multiplication continues until the energy carried by the secondary hadron goes below the threshold for producing other hadrons. The cascade contains two distinct components, namely the electromagnetic part ($\eta/\pi^0 \rightarrow \gamma\gamma$) and the hadronic part ($\pi^+, \pi^-, p, n, etc.$).
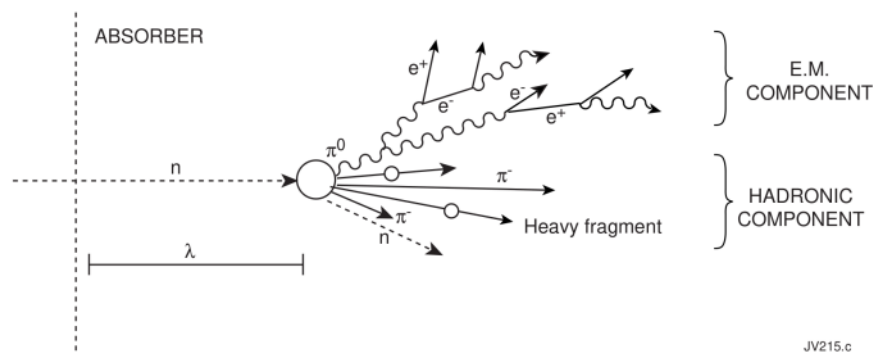
Figure 3.3: A schematic of development of a hadronic shower. Figure taken from (Virdee, 2007).

The particles emitted during the nuclear spallation process have to overcome the nuclear binding energy to escape the nuclear boundary. The energy needed for this process does not contribute to the detector signal and is termed invisible energy. Hence, in general, the energy deposited, which is available to convert to a signal, from a pion shower is less than an electron shower of the same energy.

As seen from 3.4, for hadronic showers, there is a high shower-to-shower fluctuation in energy carried by the electromagnetic component (mainly coming from π^0) for a given pion energy. Also, the depth at which the π^0 is produced varies from shower to shower. The longitudinal spread of hadronic showers is more than an electromagnetic shower of the same energy. For example, $\approx 151cm$ of iron is needed to contain 99 percent of 100 GeV π^- shower. In comparison, $\approx 44cm$ of iron is needed to contain 99 percent of 100 GeV e^- shower (Acar et al., 2023).
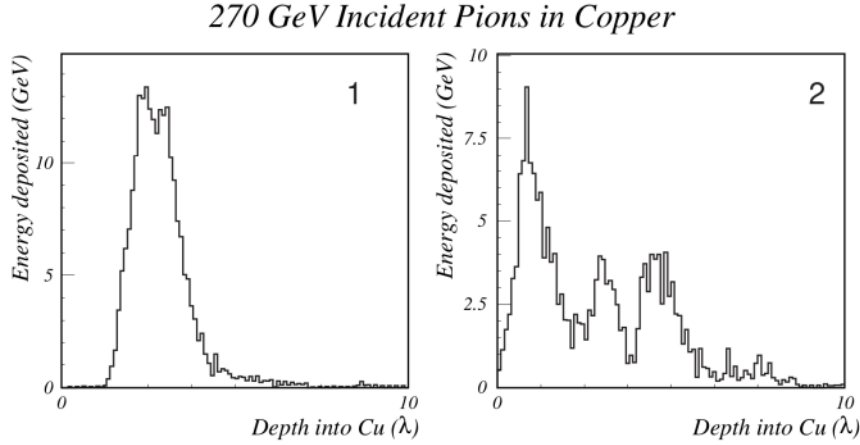
Figure 3.4: The simulation of energy deposition of two different 270 GeV pion in a block of copper. Figure taken from (Virdee, 2007).

3.2   Jet Formation

In particle physics, a jet refers to a collimated spray of particles produced from the fragmentation and hadronization of quarks and gluons. Jets typically arise from the hadronization of quarks and gluons that are produced in high-energy collisions, such as those occurring in particle accelerators like the Large Hadron Collider (LHC) at CERN.

When high-energy quarks or gluons are produced in a collision, they cannot exist freely due to the strong force confinement property of quantum chromodynamics (QCD). Instead, they undergo a process called hadronization, where they fragment into a cascade of stable hadrons (such as protons, neutrons, pions, and other mesons) before eventually being detected.

The resulting hadrons tend to be collimated, meaning they are emitted in roughly the same direction as the original quark or gluon. This collimation arises from the conservation of momentum and energy in the fragmentation process. As a result, the observed pattern of particles appears as a focused spray or jet of particles traveling in a similar direction (Marzani et al., 2019).

3.2.1   Two-jet events

Let us illustrate this with an example from Martin and Shaw, 2008.

In the centre-of-mass energy range 15–40 GeV, electron–positron annihilation into hadrons is dominated by processes like -

$$e^+ + e^- \rightarrow q + \vec{q} \tag{3.1}$$

These can be regarded as occurring in two stages: a primary electromagnetic process leading to the production of a quark–antiquark pair, followed by a strong interaction process, called fragmentation, which converts the high-energy $q\vec{q}$ pair into two jets of hadrons. These jets are emitted in opposite directions in the centre-of-mass frame in order to conserve momentum. The fragmentation process that converts the quarks into hadrons is very complicated, and the composition of the jets (i.e. the numbers and types of particles in the jet and their momenta) varies from event to event.

A typical example of such an event observed in an electron–positron colliding-beam experiment is shown in 3.5.
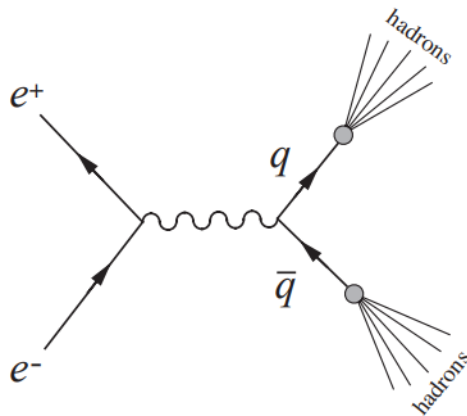
Figure 3.5: Basic mechanism of two-jet production in electron–positron annihilation. Figure taken from (Martin and Shaw, 2008).

3.3   Parton Showers

The stage between the $q\vec{q}$ pair and the hadron formation (as shown as the small bubble between the $q\vec{q}$ line and the hadrons in Figure 3.5) is called as a Parton shower.

A parton is a particle that is the constituent of hadrons. These partons, primarily quarks and gluons, interact with each other during collision processes (Nagy and Soper, 2018).

3.3.1 Quark Decays

- When a high-energy quark is involved in a collision, it can emit additional gluons as it evolves from high to low momentum.

- This emission process is governed by QCD and occurs probabilistically, with the emission probability depending on factors such as the energy and momentum transfer involved.

- The emitted gluons can subsequently split into quark-antiquark pairs, continuing the cascade of parton emissions.

3.3.2 Gluon Decays

- Gluons, being the carriers of the strong force, can self-interact and split into quark-antiquark pairs or emit additional gluons.

- Similar to quark emissions, gluon emissions occur probabilistically and depend on factors such as the energy and momentum transfer involved in the interaction.

- These emitted quarks and gluons further undergo fragmentation and hadronization processes, eventually forming collimated sprays of particles, or jets, observed in collider experiments.

The propagation of the parton shower involves the successive emissions and interactions of partons, resulting in a cascade of increasingly lower momentum partons. This cascade continues until the partons reach a low enough energy scale where non-perturbative effects become dominant, leading to the formation of stable hadrons (Ellis et al., 2003).

Chapter 4

Jet Measurement and Reconstruction

4.1 Jet Measurement

Till now we looked at the interactions of particles with matter, their development into particle showers and subsequent jet definitions. This chapter focuses on the identification of particles and their energy measurements, which is used to cluster jets in an event.

Particle Identification (ID) is a crucial aspect of most High Energy Physics experiments(the detectors used depend on the physics under study). In a typical experiment two beams of particles collide within the detectors (or a single beam collides with a fixed target). The resulting events should be reconstructed as fully as possible, where usually many particles emerge from the interaction point.

- **Tracking detectors** determine whether the particles are charged, and (in conjunction with a magnetic field) measure their momentum and the sign of their charge

- **Calorimeters** measure the energy of particles, determine whether they have electro-magnetic or hadronic interactions, and detect neutral particles

4.1.1   Measurement of Energy - Calorimeters

Hadronic calorimeters are primarily used to measure the energies of jets (Virdee, 2007). To measure the energy of a jet, it is necessary to measure the energy and position of the jet constituent particles. Calorimeters measure the energy and position of the particles that pass through them and are designed to fully contain the showers that develop from e^-, e^+, photons and hadrons. The absorber medium is where the particle shower develops and deposits most of its energy and the active medium is where the deposited energy is actually measured. Based on the active and absorber material construction, calorimeters can be of two types: homogenous and sampling calorimeters.

- In **homogeneous calorimeters**, a single medium serves both the purpose of shower development (absorber) and energy measurements (active). For example, the current CMS electromagnetic calorimeter (ECAL) is a homogenous calorimeter made up of $PbWO_4$ scintillating crystals. crystals.

- In **sampling calorimeters**, dense absorber layers are sandwiched between light active layers. Most of the energy deposition occurs in the absorber layers and the active layers take a snapshot of energy deposition at intermediate depths. These calorimeters are mainly used to measure the energy from hadronic showers, which need a larger depth for containment. For example, the current CMS hadronic calorimeter (HCAL) is a sampling calorimeter with brass as the absorber material and plastic scintillators as the active material.

4.1.2   Measurement of Momentum

In general the track of a charged particle is measured using several (N) position-sensitive detectors in the magnetic field volume. Assume that each detector measures the coordinates of the track with a precision of $\sigma(x)$. Then, the obtainable momentum resolution depends on: **L** (length of the measurement volume), **B** (magnetic field strength) and σ(position resolution).

Consider a charged particle traversing a magnetic volume of side L with velocity v, with the magnetic field pointing into the page, as shown in Figure 4.1. Assuming $\theta \approx small$,

$$\frac{L/2}{\rho} = \sin \theta/2 \approx \theta/2 \tag{4.1}$$
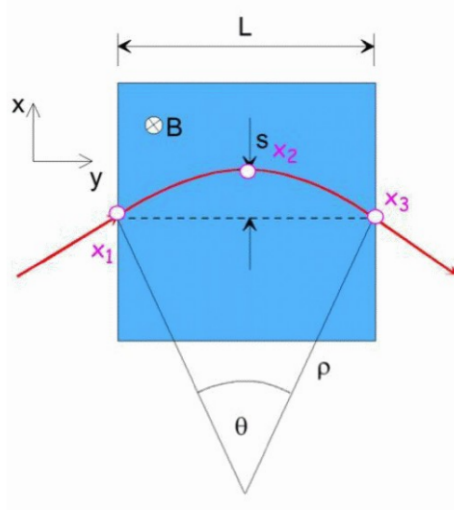
Figure 4.1: A charged particle (red) traversing a magnetic volume (blue) of side L. Figure taken from (Bracinik and Watson, 2012).

$$S = \rho \left(1 - \cos\frac{\theta}{2}\right) \approx \rho\frac{\theta^2}{8} \tag{4.2}$$

From deflection in a magnetic field,

$$\theta = \frac{L}{\rho} = \frac{0.3BL}{p_T} \Rightarrow S = \frac{0.3}{8}\frac{L^2 B}{p_T} \Rightarrow \frac{\sigma(p_T)}{p_T} = \frac{\sigma(s)}{s} \tag{4.3}$$

Say we measure sagitta from three (N=3) spacial points:

$$S = x_2 - \frac{x_1 + x_3}{2} \Rightarrow \sigma^2(s) = \frac{3}{2}\sigma^2(x) \tag{4.4}$$

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma(S)}{S} = \sqrt{\frac{3}{2}}\sigma_X \frac{8p_T}{0.3BL^2} \tag{4.5}$$

For N equidistant measurements, the momentum resolution is described by the Gluckstern formula (Bracinik and Watson, 2012):

$$\frac{\sigma(p_T)}{p_T} = \sqrt{\frac{720}{N+4}}\sigma_X \frac{8p_T}{0.3BL^2} \tag{4.6}$$

4.1.3   Particle Identification

Particle identification (PID) is crucial for jet reconstruction in particle physics experiments for several reasons[Martin and Shaw, 2008, Forty, 2017]:

- **Background Suppression:** In high-energy collisions, jets are often produced alongside a large number of other particles from the underlying event and pileup interactions. Accurate PID helps distinguish particles belonging to the jet of interest from these background particles, thereby improving the purity of the reconstructed jet sample.

- **Jet Energy Calibration:** Different types of particles deposit energy differently in the detector material. Accurate PID allows for the identification of particle types within a jet, which in turn enables better energy calibration of the reconstructed jets. By understanding the particle content of a jet, corrections can be applied to account for energy losses and ensure the accurate determination of the jet's energy.

- **Understanding Jet Substructure:** Jets may contain substructure features that provide insight into the underlying physics processes. PID allows for the identification of specific particles within a jet, such as heavy-flavor hadrons or photons, which can help discriminate between different jet substructure configurations and aid in the study of phenomena like quark/gluon discrimination, b-tagging, or Higgs boson decays to specific final states.

- **Background Rejection in Searches:** In searches for new physics or rare processes, it is essential to distinguish signal jets from background jets originating from known Standard Model processes. PID provides discriminatory power to reject background jets while preserving signal jets, thereby enhancing the sensitivity of searches for new phenomena.

Particle identification is performed using both the electromagnetic and hadronic calorimeters on the basis of lateral and longitudinal shower profiles. The demand is quite different depending on the desired physics process.

There are various techniques used for particle identification (Virdee, 2007):-

1. **Identification of Particles using Time of Flight** - At low momenta the difference in the time taken by different particles to traverse a certain distance can be used to distinguish between them.

2. **Identification of Particles using Specific Energy Loss** - The difference in the energy loss of charged particles traversing a medium at a given momentum by various particles can be used to distinguish between them.

3. **Identification of Electrons using Transition Radiation Detectors** - It is based on the principle that "Transition radiation is emitted when a charged particle moves from a medium of refractive index n1 to a medium of a different index n2." This is the concept behind the ATLAS Transition Radiation Tracker (Appendix A).

4. **Identification of Particles using Particle-deposit containment** - Electrons, characterized by their electromagnetic interactions, primarily deposit energy in the electromagnetic calorimeter (ECAL), yielding distinct clusters of energy deposits. Hadrons, on the other hand, interact via the strong force and typically deposit energy in both the ECAL and the hadronic calorimeter (HCAL). Muons, being minimally ionizing particles, traverse through the detector, leaving minimal energy deposits in the calorimeters but registering tracks in the inner tracking system and muon chambers. Neutrinos, which interact weakly with matter, often escape detection altogether, leaving behind little to no detectable signal in the detector. Their presence can be inferred by calculating the "missing energy" in an event. Quarks, being confined within hadrons due to color confinement, manifest their presence indirectly through jets.
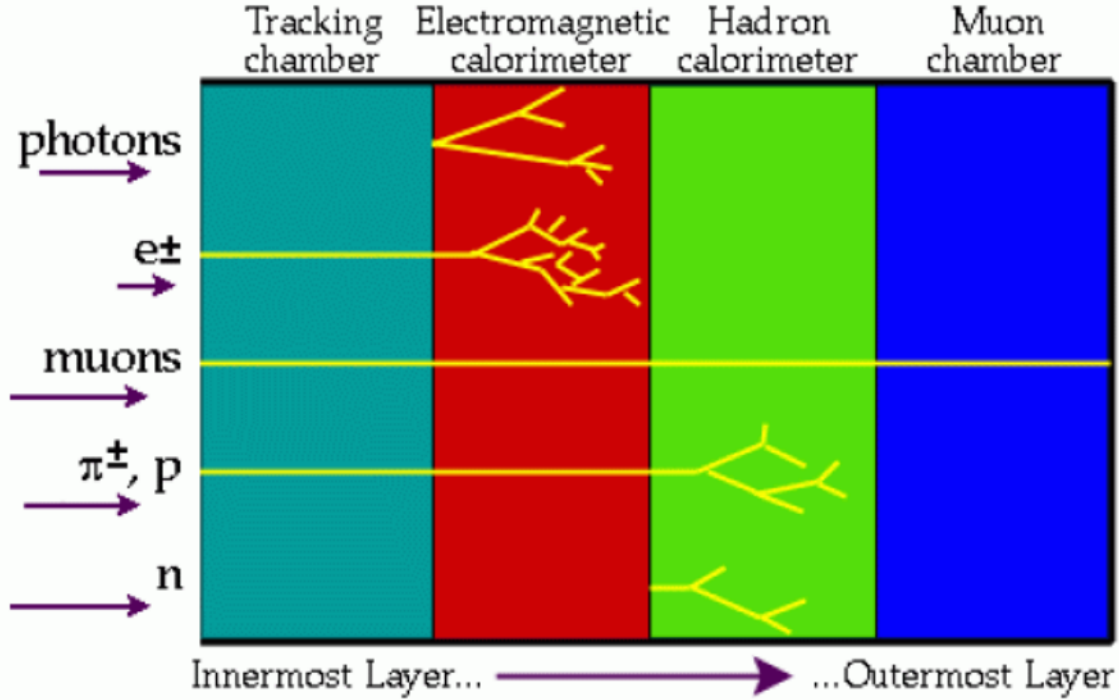
Figure 4.2: Particle identification by containing their energy deposits in specific regions of the detector. Figure taken from (Bracinik and Watson, 2012).

4.2   Jet Reconstruction

The definition of a jet as being a 'collimated sprays of particles, produced in abundance in high energy particle collisions' is oversimplified in a few aspects. Mainly, whether two particles are part of the same jet or belong to two separate jets has some degree of arbitrariness, related to what we practically mean by "collimated". The simple concept of what a jet is meant to represent is therefore not sufficient to practically identify the jets in an event. To do this, one relies on a *jet algorithm*, i.e. a well-defined procedure that tells how to reconstruct the jets from the set of hadrons in the final state of the collision (Marzani et al., 2019). There is no single, universal definition of a jet – which particles belong to a jet depend on the algorithm used to combine particles into jets (Larkoski et al., 2020). An intuitive definition for a jet algorithm consists of summing the momenta of all particles within a cone with fixed size (Sterman and Weinberg, 1977). Naive cone algorithms are not infared and collinear (IRC) safe – the requirement that the resulting jets be insensitive to arbitrarily low energy particles

and collinear splittings. IRC safety is a useful theoretical requirement for making calculations in pQCD and is also a convenient language for describing the experimental robustness to noise and detector granularity. Over the past few decades, a number of jet algorithms have been proposed. They typically fall under two big categories:

- cone algorithms

- sequential-recombination algorithms

Today, almost all studies involving jets performed at the LHC use the anti-kt algorithm (Cacciari et al., 2008), which is a modification of the kt recombination algorithm. A generalization of these algorithms leads to three classes, distinct only by the sign of the exponent of the transverse momentum $p_{T,i}$ in the inter-particle distance measure:-

$$d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta_{ij}^2}{R^2} \tag{4.7}$$

This definition is used in the results section. The original kT algorithm, with k = 1 in Equation 4.7, clusters soft and collinear particles first, the Cambridge/Aachen algorithm (CA) (Dokshitzer et al., 1997), with k = 0, prioritizes particles in the clustering solely by their angular proximity, and the anti-kT algorithm Cacciari et al., 2008, with k = 1, combines the hardest particles first. The proposal of the anti-kT algorithm is also responsible for the disappearance of cone-type algorithms in experimental studies.

Chapter 5

Jet Tagging

5.1 Jet tagging

One of the most important questions about a jet is which type of elementary particle initiates it. The technique for the identification of the elementary particle initiating a jet is called "Jet Tagging". Jets initiated by different particles exhibit different characteristics. For example, jets initiated by gluons tend to have a broader energy spread than jets initiated by quarks. High-momentum heavy particles (e.g., top quarks and W, Z, and Higgs bosons) that decay hadronically can lead to jets with distinct multi-prong structures. Therefore, the identity of the source particle can be inferred from the properties of the reconstructed jet (Qu and Gouskos, 2020). This analysis of the properties of a jet is called "Jet Substructure Analysis".

Jet substructure techniques are used for the identification of the particle origin of jets. Several substructure variables have been developed by the theoretical community that can be used along with the jet mass for jet classification. Traditionally, the term 'tagger' indicates the use of one or more of these variables to discriminate between jets coming from different types of particles (Kogler et al., 2019).

The natural next question is why we apply highly complex tagging algorithms to a pre-processed set of observables rather than to actual data. This question becomes especially relevant when we consider the significant conceptual and performance progress in machine

learning. Deep learning, or the use of neural networks with many hidden layers, is the tool which allows us to analyze low-level LHC data without constructing high-level observables. In recent years, machine learning (ML) has injected fresh blood in jet tagging. Jets are regarded as images [Almeida et al., 2015, de Oliveira et al., 2016] or as sequences [Guest et al., 2016], trees [Louppe et al., 2019], graphs [Qu and Gouskos, 2020], or sets [Komiske et al., 2019] of particles, and ML techniques, most notably deep neural networks (DNNs), are used to build new jet tagging algorithms automatically from (labeled) simulated samples or even (unlabeled) real data, leading to new insights and improvements in jet tagging. A more detailed review of Jet Representations and ML models is given in Appendix B.

Deep learning has now been applied to jets of every origin: electroweak bosons, gluons, light quarks, heavy quarks, and even BSM particles. (Larkoski et al., 2020) provides a good summary of these.

5.2   Some important particle properties

This section is meant to be a quick overview of some important particle jets that are relevant for this thesis. Some parts of this have been taken from (Kheddar et al., 2024).

5.2.1   W and Z bosons

W and Z bosons are important closely related particles described by the SM of particle physics. They play a significant role in the weak nuclear force, which is responsible for certain types of particle interactions and radioactive decay. The existence and properties of the Z boson, along with the W bosons, provided strong support for the electroweak theory and the SM as a whole. However, as with the W boson, the Standard Model has limitations and does not explain all aspects of particle physics, such as gravity, dark matter, and the hierarchy of particle masses.

Here are some key points about the W and Z bosons:

- **Charge and Variants**: The W boson comes in two varieties: the W^+ and the W^-, which carry a positive and negative electric charge, respectively. These particles are antiparticles of each other. The Z boson is a neutral elementary particle.

- **Mass and Spin**: The W bosons masses are around 80.4 GeV/c^2. The Z boson has a

relatively large mass. Its mass is around 91.2 GeV/c². Both W and Z bosons have a spin of 1, which is a measure of their intrinsic angular momentum.

- **Decay**: The W and Z bosons are unstable and have a very short lifetime. They quickly decay into other particles. For example, a W^+ boson can decay into a positron (an antielectron) and a neutrino, while a W^- boson can decay into an electron and an antineutrino. The Z can decay into various combinations of charged leptons (such as electrons and muons) and their corresponding antiparticles, as well as neutrinos and antineutrinos.

5.2.2   The Higgs boson

The Higgs boson is crucial to our understanding of how other particles acquire mass and, by extension, how the universe's structure and behavior arise. The key points about the Higgs boson are:

- **Origin of Mass:** is associated with the Higgs field, a theoretical field that permeates all of space. In the SM, particles acquire mass by interacting with the Higgs field. The more a particle interacts with this field, the greater its mass will be. This mechanism explains why some particles are heavier than others.

- **Mass and Spin:** the Higgs boson itself has a mass of around 125.1 GeV/c². It has a spin of 0, which means it has no intrinsic angular momentum.

- **Decay:** is unstable and quickly decays into other particles after its creation in high-energy collisions. The specific decay modes and products depend on the energy at which it is produced.

- **Higgs field Interaction:** is a carrier of the interaction associated with the Higgs field. When particles move through space, they interact with this field, which gives them mass. The Higgs boson itself is the quantized excitation of this field.

5.2.3   The top quark

The Top Quark is one of the fundamental particles described by the Standard Model of particle physics. It holds a special place in particle physics due to its extremely large mass

and its role in various processes involving high-energy collisions. Here are some key points about the top quark:

- **Mass:** The top quark is the heaviest known elementary particle. Its mass is approximately 173.2 $\mathrm{GeV/c^2}$, which is even heavier than an entire atom of gold.

- **Quarks and the Strong Force:** Quarks are the building blocks of protons and neutrons, which are the constituents of atomic nuclei. The top quark, like all quarks, experiences a strong nuclear force, which is responsible for holding quarks together within hadrons (particles composed of quarks).

- **Weak Decays:** Due to its high mass, the top quark is relatively short-lived and decays before it can form bound states with other quarks to create hadrons. It decays primarily through weak interaction, one of the fundamental forces described by the Standard Model.

- **Production and Detection:** The top quark is typically produced in high-energy particle collisions, such as those that occur in experiments at particle accelerators like the LHC. Due to its high mass, the top quark is often produced along with its corresponding antiquark. Researchers detect its presence indirectly by observing its decay products, which can include other quarks, leptons (such as electrons and muons), and neutrinos.

- **Role in Electroweak Symmetry Breaking:** The top quark is of particular interest in theories related to electroweak symmetry breaking, a phenomenon that explains why certain particles acquire mass. Its large mass plays a significant role in the behavior of the Higgs boson and its interactions.

Part III

Methodology

Chapter 6

Brief Overview of Neural Networks

6.1 Neural Networks

A neural network is a mathematical framework that encapsulates a functional representation. During the process of training a neural network, specific parameters of the network, such as the coefficients of its functional representation, are iteratively adjusted to make the network's output approximate a desired but often unknown function. The potential for the network to approximate any function closely depends on factors like its capacity, architecture, training duration, and the availability of relevant data[Elbrächter et al., 2021]. The objective of training is to discover a function that accurately represents the underlying problem, enabling the network to make accurate predictions for new data within the same problem domain, thus exhibiting generalization capability. However, it's important to note that achieving generalization power isn't guaranteed for all neural networks even after training (Winter, 2021).

Neural networks are composed of neurons (nodes) and weights (edges) that are arranged in consecutive layers (see Figure 6.1) such that each layer's neurons receive their inputs from the previous layer's neurons and send their outputs to the successive layer's neurons. Data is fed to the network via the first (input) layer, undergoes transformations in the intermediate (hidden) layers until it arrives at the final (output) layer that can be read out and interpreted. The amount of hidden layers separates shallow neural networks from deep networks. Deep networks can easily span tens or hundreds of hidden layers.
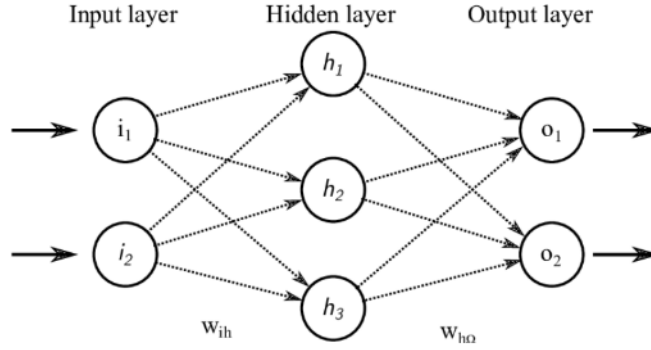
Figure 6.1: Simple neural network architecture. Figure reproduced from Chollet, 2017

Data propagating through a neural network is successively transformed by the neurons of each layer. Each neuron transforms its inputs into a single output value, called activation. In a fully connected neural network architecture (Fig 6.1) the activation of a single neuron is defined by -

$$a(\vec{x}) = \sigma(\sum_i w_{ih} x_i + b) \tag{6.1}$$

where, $\vec{x}$ represents the inputs, w_{ih} the connecting weights and b the bias term. The weights and biases are the free parameters of the network, usually initialized to random values before adjustment due to network training. σ is a non-linear function (tanh, sigmoid, ReLU, softmax, etc.) that allows the model to capture non-linear relationships in the input data.

Geometrically, a neuron activation transforms the input in three ways:

- Transform the input linearly: $\sum_i w_{ih} x_i$

- Translate to a different position by b.

- Warp by σ. This effectively maps the input into a non-linear manifold.

Successive non-linear layers allow complex transformations. These networks are referred to as deep neural networks (DNNs). Non-linearity is a decisive criterion for depth since a stack of linear hidden layers, on the other hand, can be collapsed to a single layer containing linear combinations of the former. Therefore, DNNs are much more powerful than shallow (linear) nets.

Since successive hidden layers obtain their inputs from previous layers. In this form the (forward) propagation of data through the network can be computed rather easily by basic linear algebra subprograms that are highly optimized for matrix multiplications. A graphical processing unit (GPU) can further accelerate the computation by introducing many computational cores that allow for massive parallelization. As network training is a curve fitting process involving many free parameters, care must be taken to not overfit (or underfit) the model on the training data. Overfitting happens when the parameters are tuned too much. The desired function is for example linear while the network learned a quadratic fit due to the presence of noise in the training data, i.e. the network learns a fit to random fluctuations caused by noise which do not capture the underlying problem. Underfitting is the reverse process of a linear fit to a desired quadratic function resulting from too few network parameters or insufficient training time. To prevent overfitting or underfitting, the training process should be monitored and halted at the right time. Additionally, so-called regularization techniques can be used to prevent overfitting.

The training process is guided by a so-called loss (or objective) function $\mathcal{L}$. Since neural networks can be trained for different purposes (classification, regression, reconstruction, data generation, etc.) a multitude of loss functions can be found in the literature. For a given input, the loss function takes the activations of the network's output layer and a corresponding desired output (ground truth) as input and generates a single real number to be minimized.

If the desired output (truth) has been generated in advance by experts or external sources one speaks of supervised learning. One variant of supervised learning is self-supervised learning in which the ground truth and input data are identical.

Chapter 7

Introduction to PELICAN

7.1 PELICAN

The accurate prediction of jets in particle physics experiments demands neural networks that maintain critical symmetries, such as Lorentz invariance, to ensure reliable results across different reference frames. Traditionally, achieving Lorentz invariance has involved selecting invariant observables or preserving equivariant representations in neural network architectures. However, the PELICAN architecture (Bogatskiy et al., 2022) introduces a novel approach by leveraging the complete set of Lorentz invariants, specifically pairwise dot products between input 4-momenta. By simplifying the architecture and enhancing interpretability, training, and overall performance, PELICAN addresses the complexity of jet tagging tasks effectively. To tackle the permutation structure inherent in particle data, PELICAN employs permutation-equivariant architectures, ensuring that network predictions remain invariant under permutations of input particles. This approach maintains physical consistency while modeling complex interactions between particles. By utilizing permutation-equivariant layers, PELICAN can effectively capture the intricate symmetries of the problem, enhancing its adaptability and performance in jet tagging tasks. Ultimately, by embracing fundamental symmetries like Lorentz invariance, PELICAN exemplifies the potential of physics-aware neural networks to deliver reliable and interpretable results in particle physics experiments, paving the way for advancements in understanding complex physical phenomena. (Bogatskiy et al., 2024)

7.2   PELICAN Architecture

The PELICAN architecture simplifies previous approaches in jet tagging by utilizing a complete set of Lorentz-invariants, specifically pairwise dot products between input 4-momenta. This approach enhances the architecture's interpretability, ease of training, and overall performance in jet tagging tasks.

7.2.1   Inputs and Embeddings

This component describes the preprocessing of input data, including the computation of pairwise dot products of 4-momenta and the incorporation of auxiliary beam particles to restore orientational knowledge within datasets. Embedding layers further process the dot products and scalar data, facilitating flexible processing while preserving Lorentz symmetry.

7.2.2   Permutation Equivariant Blocks

The permutation equivariant blocks transform arrays of rank 2 and play a crucial role in maintaining permutation symmetry within the architecture. These blocks consist of message and aggregation components, which apply linear aggregation functions and mix aggregators to ensure that the architecture respects permutation symmetry, which is vital for accurate jet tagging.

7.2.3   Classification and 4-Vector Regression Outputs

This component discusses how the PELICAN architecture can serve as both a classifier for jet tagging and a provider of 4-vector outputs for tasks such as momentum reconstruction. Different output layers are employed depending on the task, ensuring that the architecture can adapt to various jet tagging requirements while maintaining performance and interpretability. In the realm of particle physics experiments, particularly in tasks like jet tagging, ensuring that neural networks maintain certain symmetries is crucial for accurate and reliable predictions. One such symmetry is Lorentz invariance, which ensures that the output of the network remains unchanged under any Lorentz transformation applied to the input 4-vectors (energy-momentum vectors). This property is essential for preserving the physical consistency of predictions across different reference frames. Traditionally, enforcing Lorentz invariance in neural networks has been approached in various ways. One common method involves selecting a set of invariant observables as inputs, while another draws inspiration from convolutional

neural networks (CNNs) by preserving group-equivariant latent representations in hidden layers. These approaches aim to regulate the network's behavior and improve its performance. In this work, a different approach is taken. Instead of hand-picking invariant observables or focusing solely on equivariant representations, the complete set of Lorentz invariants on the input 4-vectors is computed. These invariants, which depend only on pairwise dot products of the 4-vectors, are used as the network input. Remarkably, it is demonstrated that these invariant dot products alone can yield state-of-the-art performance in a simpler architecture, showcasing the power of leveraging fundamental physical principles in neural network design. Furthermore, to address the challenges posed by the permutation structure of particle data, permutation-equivariant architectures are employed. These architectures ensure that the network's predictions remain invariant under permutations of the input particles, maintaining physical consistency. By utilizing permutation-equivariant layers, the network can effectively model complex interactions between particles while respecting the inherent symmetry of the problem.

Chapter 8

Evaluation Metrics

8.1 Evaluation Metrics

In the evaluation of neural network models, several metrics are commonly employed to assess their performance and effectiveness in various tasks. These metrics provide insights into different aspects of the model's behavior and can help researchers and practitioners make informed decisions about model selection and optimization. Here, we discuss the key evaluation metrics used in our analysis:

- **Accuracy:** Accuracy is a fundamental metric that measures the proportion of correctly classified instances among all instances in the dataset. It is calculated as the ratio of the number of correctly predicted samples to the total number of samples in the dataset.

- **AUC Score:** The Area Under the Receiver Operating Characteristic (ROC) Curve (AUC score) is a performance metric commonly used in binary classification tasks. It quantifies the ability of the model to distinguish between positive and negative instances across all possible thresholds. A higher AUC score indicates better discrimination ability.

- **ROC Curve:** The ROC curve is a graphical representation of the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) across different threshold values. It plots the TPR against the FPR for various threshold settings, providing insights into the model's performance at different classification thresholds.

- **True Positive Rate (TPR):** TPR, also known as sensitivity or recall, measures the proportion of positive instances that are correctly identified by the model. It is calculated as the ratio of true positives to the sum of true positives and false negatives.

- **False Positive Rate (FPR):** FPR measures the proportion of negative instances that are incorrectly classified as positive by the model. It is calculated as the ratio of false positives to the sum of false positives and true negatives.

- **Background Rejection at 0.5:** Background rejection at a specific signal efficiency point, often denoted as background rejection at 0.5, quantifies the model's ability to suppress background events while maintaining a certain level of signal efficiency. It represents the ratio of background events rejected to the total number of background events at a signal efficiency of 0.5.

These evaluation metrics provide valuable insights into the performance of neural network models and help assess their suitability for specific classification tasks. By considering multiple metrics, researchers can gain a comprehensive understanding of a model's strengths and weaknesses and make informed decisions about model optimization and deployment.

Part IV

Results & Discussion

Chapter 9

Dataset Description

9.1 Top Quark Tagging Reference Dataset

This is a dataset of Monte-Carlo simulated training and testing events for the evaluation of top quark tagging architectures. It contains a total of 1.2M training events, 400k validation events, and 400k test events.The leading 200 jet constituent four-momenta are stored, with zero-padding for jets with fewer than 200. For each event, the jet constituents are sorted by p_T, with the highest p_T one first. (Kasieczka et al., 2019b)

9.2 ATLAS Top Tagging Open Data

The ATLAS Top Tagging Open Data Set consists of jets taken from simulated collisions of protons at a center of mass energy of 13 TeV. The signal and background jets come from simulations of two different processes:

- Signal: A heavy Z boson (termed Z') with mass of 2 tera-electron-volts decaying to a top anti-top quark pair.

- Background: Jets initiated by light quarks and gluons. These particles are copious by-products of proton-proton collisions at the LHC

9.3   JetClass Dataset

The dataset includes a total of 10 types of jets. The jets in this dataset generally fall into two categories. The background jets are initiated by light quarks or gluons (q/g) and are ubiquitously produced at the LHC. The signal jets are those arising either from the top quarks (t), or from the W, Z or Higgs (H) bosons. For top quarks and Higgs bosons, it further considers their different decay modes as separate types, because the resulting jets have rather distinct characteristics and are often tagged individually.

Chapter 10

Verifying Lorentz and Permutation Symmetry

This section presents the performance evaluation of the PELICAN model across various datasets and input types. The results demonstrate the model's effectiveness in jet tagging tasks, emphasizing its adherence to fundamental symmetries and its robustness in realistic settings.

10.1 Reproduce Results from the Paper

The first step before starting any new research work is to verify the claims of the existing works. Therefore, the obvious first step was to test PELICAN's performance on the Max Planck Institute for Physics's setup and compare this to the reported values in the paper.

10.1.1 Performance on Top Quark Tagging Reference Dataset

The PELICAN model was evaluated on the Top Quark Tagging Reference Dataset, where it was trained to distinguish between hadronic top jets and QCD jets. The performance metrics, including accuracy and AUC, are summarized in Table 10.1.

Model	Accuracy	AUC
PELICAN (paper)	0.94	0.987
PELICAN (MPP)	0.94	0.987

Table 10.1: PELICAN's performance on the Top Quark Tagging Reference Dataset

These results highlight the PELICAN model's high performance and consistency when tagging top quark jets. The identical performance metrics for both the paper and MPP versions of PELICAN indicate robust implementation and training protocols.

10.2 Lorentz Symmetry in PELICAN

Symmetries in ML are known to produce less complex models which respect basic geometrical rules and arguably provide more opportunities for interpretability and explainability of the results. Even in realistic settings where the symmetries are merely approximate, symmetry-constrained architectures often outperform more general architectures in terms of pure accuracy (Li et al., 2024, but this is not always true.

PELICAN has Lorentz symmetry built inherently into the architecture (Bogatskiy et al., 2022). The Lorentz symmetry is one of the fundamental symmetries of the Standard Model of particle physics.Lorentz invariance is the mathematical encapsulation of the fact that the outcomes of physical phenomena don't depend on the inertial frame of the observer. In the context of particle accelerators, this boils down to the observation that all initial and final states of a particle interaction are the same in all inertial frames. This is formally reflected in the fact that the Standard Model of particle physics is Lorentz-invariant, and therefore any model of any physically relevant processes described by the Standard Model can be as well.

To verify that the PELICAN model respects Lorentz and permutation symmetries, experiments were conducted where the input data was transformed accordingly. The results, presented in Tables 10.2 and 10.3, show consistent performance, indicating the model's symmetry invariance.

10.3   Testing Lorentz Invariance

PELICAN incorporates Lorentz symmetry in the network architecture by making the final inputs of the network a Lorentz invariant quantity. Given a set of 4-vector inputs $p_1, p_2, ..., p_N$, it computes a complete set of Lorentz invariants on that set. As proven by H.Weyl in his book 'The Classical Groups', the space of symmetric invariants constructed out of a collection of vectors in the fundamental representation consists of functions of only the pairwise invariant dot products. In short, all totally symmetric Lorentz invariants can be written in the following form:

$$I(p_1, \ldots, p_N) = f\left(\{p_i \cdot p_j\}\right)_{i,j} \tag{10.1}$$

Hence, PELICAN's final inputs are the pairwise dot products of the input 4-vectors.

To test whether PELICAN is Lorentz symmetric in working, we look at the following:- if the inputs to the network are a collection of 4-vectors $p_1, p_2, ..., p_N$ and the output is $F(p_1, p_2, ..., p_N)$ -

$$F(\lambda p_1, \lambda p_2, ..., \lambda p_N) = F(p_1, p_2, ..., p_N) \tag{10.2}$$

where, λ is a Lorentz transformation.

10.3.1   Training

The Hyperparameter for the training are listed below:-

- Dataset = Top Quark Tagging Reference Dataset (9.1)

- GPU = NVIDIA A100-SXM4-40GB with 40960 MB of GPU memory

- Batch size = 64

- Number of epochs = 15

- Optimizer = Adam

- Activation = leaky ReLU

- Train events = 600k

- Validation events = 400k

- Test events = 400k

- Learning rate: lr-init=0.0025 to lr-final=1e-6

10.3.2 Evaluation

Once the training of the model was complete, we evaluated it on two testing samples. Sample 1 contains 400k events from the Top Quark Tagging Reference Dataset, while Sample 2 consists of Sample 1 events modified by an arbitrary Lorentz Transformation.

- For two inertial frames moving relative to their x-axis direction

$$
\Lambda = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
$$

- Let $v = 0.6c$, then, the Lorentz transformation matrix is

$$
\Lambda = \begin{pmatrix} 1.25 & -0.75 & 0 & 0 \\ -0.75 & 1.25 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
$$

Data	Accuracy	AUC
Sample 1	0.94	0.987
Sample 2 (Lorentz transformed)	0.94	0.987

Table 10.2: PELICAN's performance under Lorentz transformations

As the model performance does not change between Sample 1 and Sample 2, the hypothesis that 'PELICAN architecture is inherently Lorentz invariant due to the specialized inputs' is True.

10.4 Testing Permutation Invariance

PELICAN addresses the permutation nature of particle data by using architectures that maintain prediction consistency regardless of the order of input particles.
Experimentally, to verify that PELICAN respects permutation symmetry, the network performance should not change based on the order of input particles, i.e.

$$F(p_1, p_2, ..., p_N) = F(p_i, p_j, ..., p_k) \tag{10.3}$$

10.4.1 Training

The Hyperparameter for the training are listed below:-

- Dataset = Top Quark Tagging Reference Dataset (9.1)

- GPU = NVIDIA A100-SXM4-40GB with 40960 MB of GPU memory

- Batch size = 64

- Number of epochs = 15

- Optimizer = Adam

- Activation = leaky ReLU

- Train events = 600k

- Validation events = 400k

- Test events = 400k

- Learning rate: lr-init=0.0025 to lr-final=1e-6

10.4.2   Evaluation

The model was first trained on $\approx 600k$ events from the Top Tag Dataset. Then we evaluated it on two testing samples. Sample 1 contains 400k events from the Top Quark Tagging Reference Dataset, while Sample 2 consists of Sample 1 events that whose particle order (per event) was permuted.

Data	Accuracy	AUC
Sample 1	0.94	0.987
Sample 2 (Permuted inputs)	0.94	0.987

Table 10.3: PELICAN's performance under input permutations

The ability to maintain performance under these transformations confirms the model's invariance properties. This invariance is crucial for ensuring that the model's predictions are not biased by arbitrary changes in the data representation, thus reflecting true physical properties.

Chapter 11

Testing PELICAN's performance on a more realistic dataset

11.1 Performance on the ATLAS Top Tag Open Dataset

The PELICAN model's robustness was further tested on the more realistic ATLAS Top Tag Open Dataset. This dataset involves a challenging task of tagging top quark jets from the decay of a heavy Z boson against a background of jets initiated by light quarks and gluons. The performance metrics for this dataset are shown in Table 11.1.

Model	AUC	ACC	$1/(\epsilon_B)$ at $\epsilon_S = 0.5$	Params	Inference Time
PELICAN (no wts)	0.966	0.902	203	48327	0.85 ms
PELICAN (with wts)	0.960	0.891	150.3	48327	0.85 ms

Table 11.1: PELICAN's performance on the ATLAS Top Tag Open Dataset

PELICAN was trained on about 2 mil events which is 5 percent of the full dataset but, it still performs just as good as the other taggers reported in the paper (11.2). We believe that this observation aligns with the statement that symmetries in ML produce less complex models requiring smaller datasets without loss of performance.

Model	AUC	ACC	$\varepsilon_{bkg}^{-1}@\varepsilon_{sig} = 0.5$	$\varepsilon_{bkg}^{-1}@\varepsilon_{sig} = 0.8$	# Params	Inference Time
ResNet 50	0.885	0.803	21.4	5.13	$1,486,209$	9 ms
EFN	0.901	0.819	26.6	6.12	$1,670,451$	4 ms
hIDNN	0.938	0.863	51.5	10.5	93,151	3 ms
DNN	0.942	0.868	67.7	12.0	876,641	3 ms
PFN	0.954	0.882	108.0	15.9	689,801	4 ms
Particle Net	0.961	0.894	153.7	20.4	764,887	38 ms

Table 11.2: Best performing taggers on the ATLAS Top Tag Dataset

The performance on this dataset showcases **PELICAN's** applicability in real-world scenarios, maintaining high accuracy and AUC while demonstrating efficient inference times. This robustness is essential for practical applications in high-energy physics experiments.

Chapter 12

Comparison of PELICAN's performance using different pairwise quantities as inputs

12.1 Comparison of performance using different pairwise quantities

Currently, PELICAN is one of the best jet taggers in the literature (Bogatskiy et al., 2024). It achieves state-of-the-art performance on the benchmark with fewer learnable parameters than the previous highest-performing network. This incredible performance even at low parameters has been attributed to the: *inherent Lorentz symmetry* and *Permutation invariance to inputs*. A natural next question is whether the particular architecture of PELICAN is still performant even without Lorentz invariant inputs i.e. 'Can the performance of PELICAN be attributed to the layer of the neural network? Or does Lorentz invariance (through inputs) play a major role?' To test this, we trained PELICAN using different pairwise input quantities, non-Lorentz invariant and compared its performance.

The pairwise inputs tested are -

- **Pairwise Invariant Mass**

- **Pairwise Jet Clustering measures - kT, anti-kT, CA**

- **pT-weighted angular separation**

For the pairwise invariant mass,

$$m_{ij} = (p_i + p_j)^2 \tag{12.1}$$

For the pairwise jet clustering algorithm measures,

$$d_{ij} = \min(p_{Ti}^{2k}, p_{Tj}^{2k})\frac{\Delta_{ij}^2}{R^2} \tag{12.2}$$

where, k=1 is the kT distance measure, k=0 is the Cambridge/Aachen distance measure and k=-1 is the anti-kT distance measure.

For the pT-weighted angular separation (pAS),

$$d_{ij} = (\Delta R_{ij})^2 p_{T,i} p_{T,j} \tag{12.3}$$

12.1.1   Training

The training and testing for this hypothesis were done on both the -Top Quark Tagging Reference Dataset and the ATLAS Top Tagging Open Dataset.

The Hyperparameter for the training are listed below:-

- Dataset = Top Quark Tagging Reference Dataset (9.1) and ATLAS Top Tagging Open Dataset (9.2)

- GPU = NVIDIA A100-SXM4-40GB with 40960 MB of GPU memory

- Batch size = 64

- Number of epochs = 15

- Optimizer = Adam

- Activation = leaky ReLU

- Train events = 400k

- Validation events = 400k

- Test events = 400k

- Learning rate: lr-init=0.0025 to lr-final=1e-6

12.1.2   Evaluation

To analyze the impact of different pairwise quantities on the PELICAN model's performance, various inputs relevant to jet physics were tested. The results for different input types on the Top Quark Tagging Reference Dataset and the ATLAS Top Tag Open Dataset are summarized in Tables 12.1 and 12.2.

The evaluation of the training was done on the two dataset separately. The results are given in the tables below.

Input Type	Accuracy	AUC	$1/(\epsilon_B)$ at $\epsilon_S = 0.5$
Dot Products	0.94	0.987	402
Invariant Mass	0.94	0.987	406
CA	0.89	0.957	66
Anti-k_T	0.68	0.717	3.9
k_T	0.68	0.717	3.9
pAS	0.68	0.719	3.8

Table 12.1: PELICAN's performance with different pairwise inputs on the Top Quark Tagging Reference Dataset

Input Type	Accuracy	AUC	$1/(\epsilon_B)$ at $\epsilon_S = 0.5$
Dot Products	0.89	0.960	150.3
Invariant Mass	0.89	0.960	151
CA	0.56	0.582	2.6
Anti-k_T	0.56	0.583	2.6
k_T	0.56	0.580	2.5
pAS	0.56	0.586	2.5

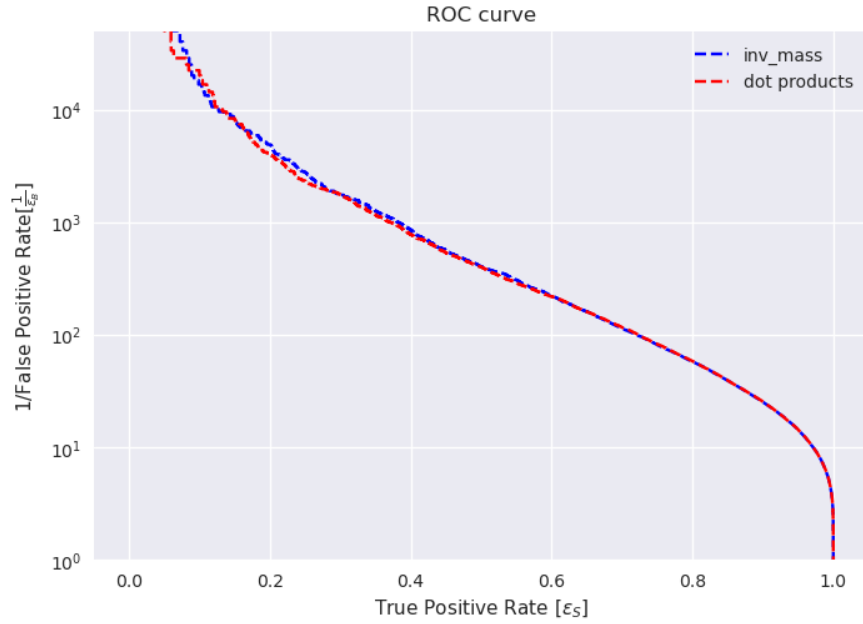**Table 12.2:** PELICAN's performance with different pairwise inputs on the ATLAS Top Tag Open Dataset
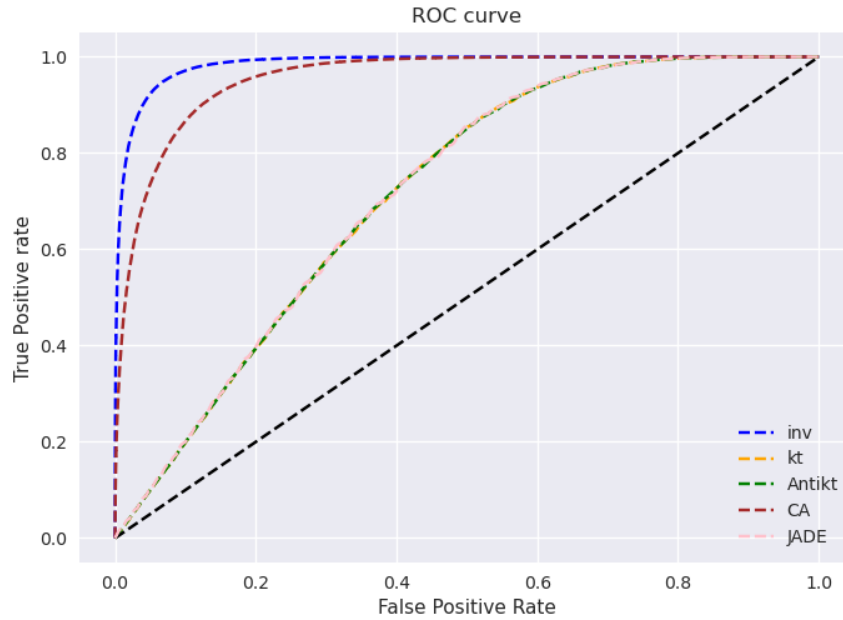
Figure 12.1: ROC Curve for PELICAN performance on Top Quark Tagging Reference Dataset
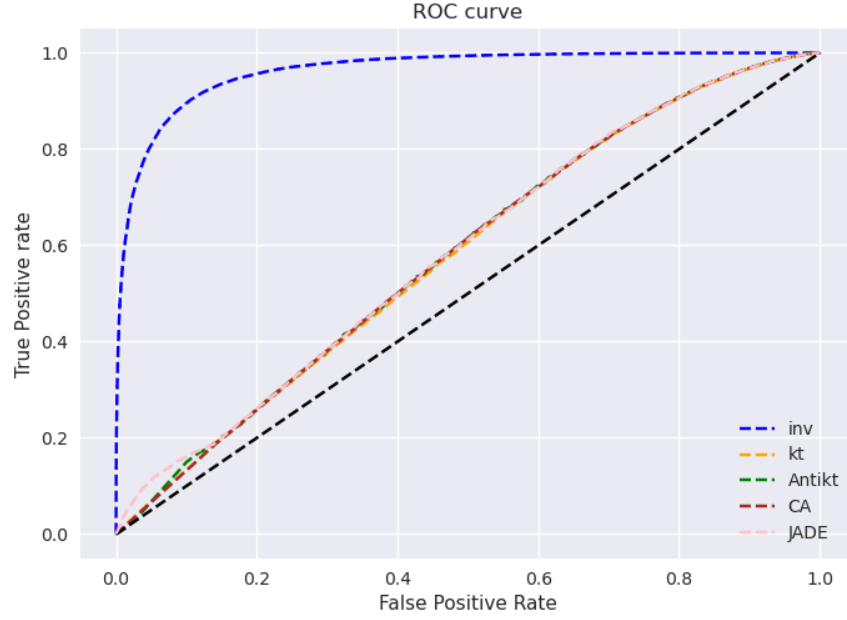
Figure 12.2: ROC Curve for PELICAN performance on ATLAS Top Tagging Open Dataset

Figure 12.3: ROC Curves for Invariant mass inputs and Regular dot product inputs coincide for Top Quark Tagging Reference Dataset
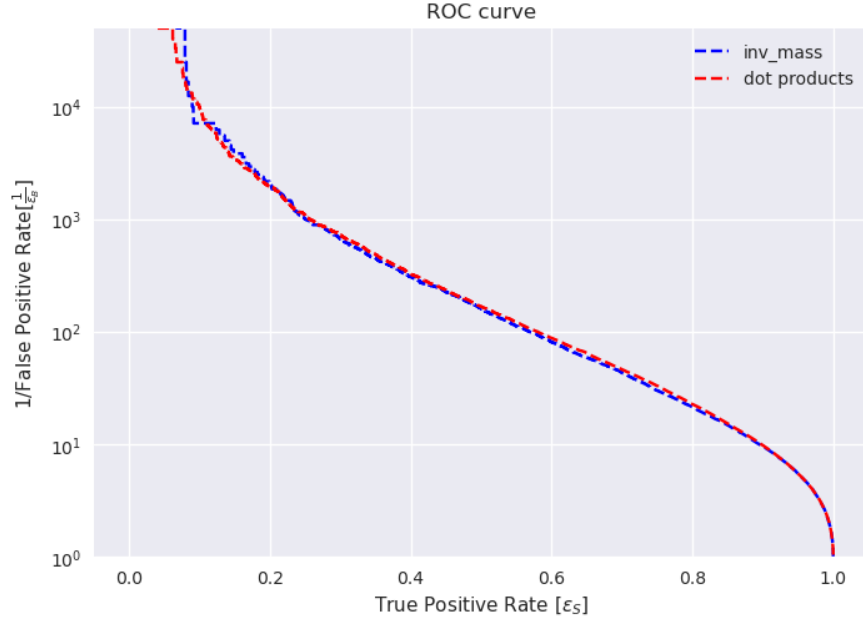
Figure 12.4: ROC Curve for Invariant mass inputs and Regular dot products coincide for ATLAS Top Tagging Open Dataset

We can see from the performance metrics and the ROC curves that PELICAN performs the best with (1) Pairwise Invariant mass or (2) Pairwise dot-products as input. There is a considerable performance difference between PELICAN having Lorentz invariant inputs like these, versus, PELICAN trained on (3) Pairwise Jet clustering measure or (4)Pairwise pT-weighted angular separation, which are not Lorentz-invariant inputs.

Further, from 12.4 the Pairwise Invariant Mass and Pairwise dot-products provide the network with the same overall information, and hence, they performance similarly. This can be seen as -

$$m_{ij} = (p_i + p_j)^2 = (E_i + E_j)^2 - \sum (p_i + p_j)^2 \tag{12.4}$$

Hence, Pairwise Invariant mass has the information of Pairwise dot-products implicit.

As a test, we also evaluated PELICAN on the kT measure with pairwise mass information implicit, but this did not perform any better than kT or antikT.

For the Pairwise pT-weighted angular separation, it can be shown that -

$$m_{ij}^2 \approx (\Delta R_{ij})^2 p_{T,i} p_{T,j} \tag{12.5}$$

But, PELICAN did not perform the best on this, which may imply that 'approximate Lorentz invariant' inputs would not improve PELICAN's performance.

These comparisons reveal that while PELICAN maintains high performance with dot products and invariant mass inputs, the model's effectiveness decreases with other input types. This indicates the importance of selecting appropriate pairwise quantities for optimal jet tagging performance.

Chapter 13

Extending PELICAN for tagging other particles

PELICAN demonstrates state-of-the-art performance for the objectives of 'Top Quark Tagging' and 'Quark-vs-gluon-initiated jet tagging'. To extend PELICAN for the task of tagging other jets such as $H \to b\vec{b}$ and $H \to c\vec{c}$, it is necessary to modify PELICAN's architecture so as to incorporate more input information during training.

The signature signals of b and c jets primarily arise from the decay products of heavy-flavor hadrons (such as B mesons for b jets and D mesons for c jets) within the jets. These signatures include:

- **Secondary Vertices**: Heavy-flavor hadrons produced in b and c jets have relatively long lifetimes compared to light-flavor hadrons. As a result, they can travel a measurable distance before decaying. The presence of secondary vertices, displaced from the primary vertex of the collision, is a distinctive signature of b and c jets.

- **Impact Parameters**: Tracks originating from the decay of heavy-flavor hadrons have large impact parameters, reflecting the displaced nature of their decay vertices. Measurement of these impact parameters within a jet provides additional discrimination power for identifying b and c jets.

- **Jet Substructure**: The substructure of b and c jets, including the distribution of

energy and momentum among their constituent particles, may exhibit characteristic patterns due to the presence of heavy-flavor hadrons. Machine learning techniques, such as jet substructure analysis, can exploit these patterns for b and c jet identification.

- **Lepton Tagging**: Some heavy-flavor hadrons decay semi-leptonically, producing charged leptons (e.g., electrons or muons) in their decay chain. The presence of isolated leptons within a jet can serve as a signature of b and c jets, particularly when combined with other discriminators.

We propose few methods to extend PELICAN to accommodate this type of extra information.

13.1   Invariant Spacetime 4-vector products

For this, we use the 9.3 JetClass dataset which is a new large and comprehensive dataset to advance deep learning for jet tagging. This dataset provides the necessary 4-vectors (E, p_x, p_y, p_z) in units of GeV for the basic functioning of the PELICAN architecture. Along side this,it includes the measured values and errors of the transverse and longitudinal impact parameters of the particle trajectories in units of mm. This trajectory displacement information is critical for tagging jets involving a bottom (b) or charm (c) quarks.

We use the transverse and longitudinal trajectory displacement information to construct a 3-dimensional trajectory displacement per particle. We these use kinematic considerations to estimate the time 't', to create the spacetime displacement 4-vector per particle per event. This set of spacetime displacement 4-vectors can be a second matrix input to PELICAN just like the usual momentum 4-vectors. The information from the two sets can be combined using a dense neural network and propagated.

By doing this, we would be able to allow PELICAN access to information about 'Displaced Vertices' which would aid in b and c jet identification.

13.2   Scalar particle identification labels

This is the simplest method we propose to extend PELICAN to tag jets of other particles. For this, in the "Input and Embed" stage of PELICAN, we add one label per particle to the scalar list -

- $L = 1$ if the particle is associated with the secondary vertex

- $L = 0$ if the particle is associated with the primary vertex

This procedure of incorporating labels is already present in the PELICAN architecture.

13.3 Performance of PELICAN on tagging other particles

13.3.1 Training

The Hyperparameter for the training are listed below:-

- Dataset = JetClass Dataset

- GPU = NVIDIA A100-SXM4-40GB with 40960 MB of GPU memory

- Batch size = 64

- Number of epochs = 15

- Optimizer = Adam

- Activation = leaky ReLU

- Train events = 800k

- Validation events = 200k

- Test events = 200k

- Learning rate: lr-init=0.0025 to lr-final=1e-6

13.3.2 Evaluation

PELICAN's performance for tagging -

- $H \rightarrow b\vec{b}$

- $H \rightarrow c\vec{c}$

- $W \to qq'$

- $Z \to q\vec{q}$

is given below.

Process	AUC	Accuracy	BgRj@$\epsilon_S = 0.5$
$H \to b\vec{b}$	0.98	0.927	403
$H \to c\vec{c}$	0.97	0.919	270
$W \to qq'$	0.96	0.904	190
$Z \to q\vec{q}$	0.96	0.900	126

Table 13.1: Performance of PELICAN for tagging other jets
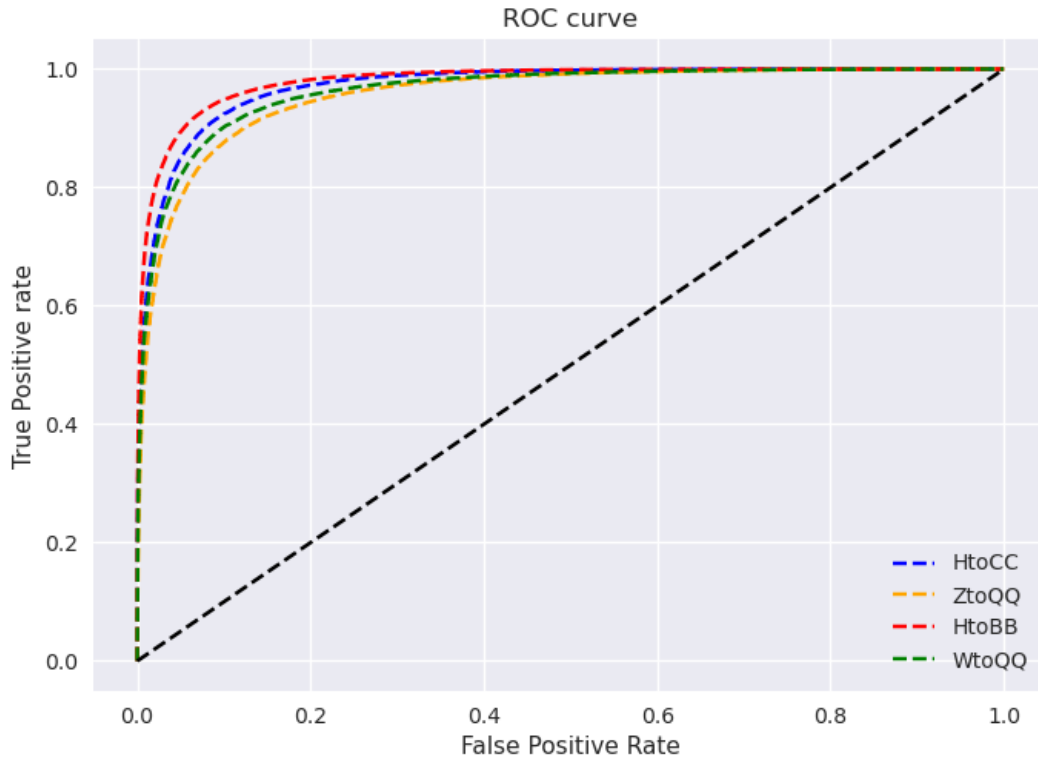
Figure 13.1: ROC Curve for PELICAN tagging Higgs, W and Z in the JetClss Dataset

From the above results, it obvious that PELICAN already performs as good as the state-of-the-art taggers. Making the improvements would improve the performance a lot.

These preliminary results show that the PELICAN model can be extended for tagging other particles, such as Higgs, W-boson, and Z-boson jets, on the JetClass Dataset. These results suggest that incorporating additional information, like impact parameters and secondary vertex information, can further enhance the model's performance while preserving its fundamental symmetries.

This extension will demonstrate the versatility of the PELICAN model, potentially suggesting that it can be adapted to a variety of jet tagging tasks. Future work will focus on refining these capabilities and integrating additional features to enhance performance further.

Appendix A

ATLAS Detector

A.1 ATLAS Detector

The ATLAS detector at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets. The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track. It is followed by the silicon microstrip tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 30 in total) above a higher energy deposit threshold corresponding to transition radiation. Reconstructed charged particles are assumed to have a charge of ± 1.

A complete overview of the ATLAS detector is provided in Ref. Collaboration et al., 2008.

Appendix B

Jet Representations

B.1   Jet Representations

Tet representations are visualizations of the complex and dynamic behavior of particle jets produced in high-energy collisions. These representations can take various forms, including images, particle clouds, and energy flow diagrams. In image representations, jets are depicted as colorful clusters of particles emerging from the collision point, providing a snapshot of the underlying physics processes. Particle clouds, on the other hand, display the spatial distribution of particles within the jet, offering insights into their energy distribution and substructure. Energy flow diagrams illustrate the flow of energy and momentum within the jet, highlighting the interactions and decays of individual particles. Each representation offers unique perspectives on jet properties and dynamics, enabling physicists to study phenomena such as jet substructure, quark-gluon discrimination, and the production of heavy-flavor particles. By leveraging advanced visualization techniques, researchers can uncover valuable insights into the fundamental particles and forces that govern the universe.

References

Aad, G., Abbott, B., Abbott, D.C., Abud, A.A., and K. Abeling, e.a., 2020. Search for new resonances in mass distributions of jet pairs using 139 fb-1 of pp collisions at \sqrt{\mathrm{s}} = 13 TeV with the ATLAS detector. *Journal of High Energy Physics 2020*.3. DOI: 10.1007/jhep03(2020)145.

Acar, B. et al., 2023. Performance of the CMS High Granularity Calorimeter prototype to charged pion beams of 20−300 GeV/c. arXiv: 2211.04740 [physics.ins-det].

Et al, W.-M.Y., 2006. Review of Particle Physics. *Journal of Physics G: Nuclear and Particle Physics 33*.1, 1. DOI: 10.1088/0954-3899/33/1/001.

Almeida, L.G., Backovic, M., Cliche, M., Lee, S.J., and Perelstein, M., 2015. Playing Tag with ANN: Boosted Top Identification with Pattern Recognition. arXiv: 1501.05968 [hep-ph].

Bogatskiy, A., Hoffman, T., Miller, D.W., and Offermann, J.T., 2022. PELICAN: Permutation Equivariant and Lorentz Invariant or Covariant Aggregator Network for Particle Physics. arXiv: 2211.00454 [hep-ph].

Bogatskiy, A., Hoffman, T., Miller, D.W., Offermann, J.T., and Liu, X., 2024. Explainable Equivariant Neural Networks for Particle Physics: PELICAN. arXiv: 2307.16506 [hep-ph].

Bracinik, J. and Watson, M., 2012. LHC Detectors. Lecture at Warwick Week.

Cacciari, M., Salam, G.P., and Soyez, G., 2008. The anti-ktjet clustering algorithm. *Journal of High Energy Physics 2008*.04, 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063.

CERN, 2018. *LHC Season 2: Facts & Figures.*

Chollet, F., 2017. *Deep Learning with Python* 1st. (USA: Manning Publications Co.).

CMS Collaboration, Accessed 2024. *Machining Jets.*

Collaboration, A., Aad, G., Abat, E., Abdallah, J., Abdelalim, A., Abdesselam, A., Abdinov, O., Abi, B., Abolins, M., Abramowicz, H., et al., 2008. The ATLAS experiment at the CERN large hadron collider.

Déliot, F. and Mulders, P.V., 2020. Top quark physics at the LHC. en. *Comptes Rendus. Physique 21*.1, 45–60. DOI: 10.5802/crphys.9.

Dokshitzer, Y.L., Leder, G.D., Moretti, S., and Webber, B.R., 1997. Better jet clustering algorithms. *JHEP 08*.

Elbrächter, D., Perekrestenko, D., Grohs, P., and Bölcskei, H., 2021. Deep Neural Network Approximation Theory. arXiv: `1901.02220 [cs.LG]`.

Ellis, R.K., Stirling, W.J., and Webber, B.R., 2003. *QCD and Collider Physics* (Cambridge University Press).

Forty, R., 2017. Particle Identification. Talk at the ICFA Instrumentation School.

Gallinaro, M., 2013. Top quark physics: A tool for discoveries. *Journal of Physics: Conference Series 447*.1, 012012. DOI: `10.1088/1742-6596/447/1/012012`.

Gouskos, L., 2020. *Machine learning techniques for jet classification ("tagging") at the LHC.*

Graph Neural Network Jet Flavour Tagging with the ATLAS Detector, 2022.

Green, D., 2000. *Physics of Particle Detectors* (Cambridge University Press). ISBN: 978-0521651515.

Guest, D., Collado, J., Baldi, P., Hsu, S.-C., Urban, G., and Whiteson, D., 2016. Jet flavor classification in high-energy physics with deep neural networks. *Physical Review D 94*.11. ISSN: 2470-0029. DOI: `10.1103/physrevd.94.112002`.

Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV, 2017. Tech. rep. Geneva: CERN.

Kasieczka, G., Plehn, T., Butter, A., Cranmer, K., Debnath, D., Dillon, B.M., Fairbairn, M., Faroughy, D.A., Fedorko, W., Gay, C., Gouskos, L., Kamenik, J.F., Komiske, P., Leiss, S., Lister, A., Macaluso, S., Metodiev, E., Moore, L., Nachman, B., Nordström, K., Pearkes, J., Qu, H., Rath, Y., Rieger, M., Shih, D., Thompson, J., and Varma, S., 2019a. The Machine Learning landscape of top taggers. *SciPost Physics 7*.1. ISSN: 2542-4653. DOI: `10.21468/scipostphys.7.1.014`.

Kasieczka, G., Plehn, T., Thompson, J., and Russel, M., 2019b. *Top Quark Tagging Reference Dataset*. Tech. rep. Version v0 (2018_03_27).

Kheddar, H., Himeur, Y., Amira, A., and Soualah, R., 2024. High-energy physics image classification: A Survey of Jet Applications. arXiv: `2403.11934 [hep-ph]`.

Kogler, R., Nachman, B., Schmidt, A., Asquith, L., Winkels, E., Campanelli, M., Delitzsch, C., Harris, P., Hinzmann, A., Kar, D., McLean, C., Pilot, J., Takahashi, Y., Tran, N., Vernieri, C., and Vos, M., 2019. Jet substructure at the Large Hadron Collider. *Reviews of Modern Physics 91*.4. ISSN: 1539-0756. DOI: `10.1103/revmodphys.91.045003`.

Komiske, P.T., Metodiev, E.M., and Thaler, J., 2019. Energy flow networks: deep sets for particle jets. *Journal of High Energy Physics 2019*.1. ISSN: 1029-8479. DOI: `10.1007/jhep01(2019)121`.

Larkoski, A.J., Moult, I., and Nachman, B., 2020. Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning. *Physics Reports 841*, 1–63. ISSN: 0370-1573. DOI: `10.1016/j.physrep.2019.11.001`.

Leo, W.R., 1994. *Techniques for Nuclear and Particle Physics Experiments* 2nd. (Springer). ISBN: 978-3-642-57821-3.

Li, C., Qu, H., Qian, S., Meng, Q., Gong, S., Zhang, J., Liu, T.-Y., and Li, Q., 2022. Does Lorentz-symmetric design boost network performance in jet physics? arXiv: 2208.07814 [hep-ph].

— 2024. Does Lorentz-symmetric design boost network performance in jet physics? *Physical Review D 109*.5. ISSN: 2470-0029. DOI: 10.1103/physrevd.109.056003.

Louppe, G., Cho, K., Becot, C., and Cranmer, K., 2019. QCD-aware recursive neural networks for jet physics. *Journal of High Energy Physics 2019*.1. ISSN: 1029-8479. DOI: 10.1007/jhep01(2019)057.

Martin, B.R. and Shaw, G., 2008. *Particle Physics* (John Wiley & Sons). ISBN: 978-0-470-03294-7.

Marzani, S., Soyez, G., and Spannowsky, M., 2019. *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-object Phenomenology* (Springer International Publishing). ISBN: 9783030157098. DOI: 10.1007/978-3-030-15709-8.

Nagy, Z. and Soper, D.E., 2018. What is a parton shower? *Physical Review D 98*.1. ISSN: 2470-0029. DOI: 10.1103/physrevd.98.014034.

Newman, D., 2023. Problems with the Standard Model of particle physics. arXiv: 2308.12295 [physics.gen-ph].

De Oliveira, L., Kagan, M., Mackey, L., Nachman, B., and Schwartzman, A., 2016. Jet-images — deep learning edition. *Journal of High Energy Physics 2016*.7. ISSN: 1029-8479. DOI: 10.1007/jhep07(2016)069.

Qu, H. and Gouskos, L., 2020. Jet tagging via particle clouds. *Physical Review D 101*.5. ISSN: 2470-0029. DOI: 10.1103/physrevd.101.056019.

Sterman, G. and Weinberg, S., 1977. Jets from Quantum Chromodynamics. *Phys. Rev. Lett. 39* (23), 1436.

Virdee, T.S., 2007. Experimental Techniques. *EP Division, CERN*.

Wikipedia contributors, Accessed 2024. *Standard Model*.

Winter, D., 2021. *Readout Chip Characterization and Convolutional Autoencoder Trigger Concept for the CMS High Granularity Calorimeter Upgrade*.