2023

Symbolic Computation of Squared Amplitudes in High Energy Physics with Machine Learning

Abdulhakim Alnuqaydan
University of Kentucky, aal700@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0003-2460-6181
Digital Object Identifier: https://doi.org/10.13023/etd.2023.482

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

<div align="right">

Abdulhakim Alnuqaydan, Student

Dr. Michael Eides, Major Professor

Dr. Christopher B. Crawford, Director of Graduate Studies

</div>

Symbolic Computation of Squared Amplitudes in High Energy Physics with
Machine Learning

DISSERTATION

A dissertation submitted in partial

fulfillment of the requirements for

the degree of Doctor of Philosophy

in the College of Arts and Sciences

at the University of Kentucky

By

Abdulhakim M. Alnuqaydan

Lexington, Kentucky

Director: Dr. Michael Eides, Professor of Physics

Lexington, Kentucky

2023

ABSTRACT OF DISSERTATION

Symbolic Computation of Squared Amplitudes in High Energy Physics with
Machine Learning

The calculation of particle interaction squared amplitudes is a key step in the calculation of cross sections in high-energy physics. These complex calculations are currently performed using domain-specific symbolic algebra tools, where the computational time escalates rapidly with an increase in the number of loops and final state particles. This dissertation introduces an innovative approach: employing a transformer-based sequence-to-sequence model capable of accurately predicting squared amplitudes of Standard Model processes up to one-loop order when trained on symbolic sequence pairs. The primary objective of this work is to significantly reduce the computational time and, more importantly, develop a model that efficiently scales with the complexity of the processes. To the best of our knowledge, this model is the first that encapsulates a wide range of symbolic squared amplitude calculations and, therefore, represents a potentially significant advance in using symbolic machine learning techniques for practical scientific computations.

Abdulhakim M. Alnuqaydan

December 20, 2023

Symbolic Computation of Squared Amplitudes in High Energy Physics with

Machine Learning

By

Abdulhakim M. Alnuqaydan

<div style="text-align:right">

_____

Dr. Michael Eides

Director of Dissertation

_____

Dr. Sergei Gleyzer

Co-Director of Dissertation

_____

Dr. Christopher B. Crawford

Director of Graduate Studies

_____

December 20, 2023

Date

</div>

To my dear parents,

to my beloved wife,

and to my family,

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 Introduction

High-energy physics stands at the forefront of our quest to understand the fundamental constituents and laws of the universe. At its core, this field seeks to unravel the complex behavior of particle interactions, providing insights into the underlying principles that govern the universe at the microscopic level. Central to this field is the calculation of the cross section, a fundamental quantity that describes the probabilities of particle interactions. These calculations, often come with great complexity, have historically posed challenges to physicists. The field of high-energy physics is continually evolving, driven by advancements in theoretical models, experimental techniques, and computational tools. In recent years, the rise of artificial intelligence (AI) and machine learning (ML) has opened new avenues for addressing longstanding challenges in the field. Among these, sequence-to-sequence transformer models, inspired by the transformer architecture, have demonstrated remarkable success in natural language processing and many other domains. Their ability to capture complex dependencies in sequential data, coupled with their parallel processing capabilities, makes them an excellent candidate for symbolic calculations in high-energy physics.

In this dissertation, we embark on a pioneering journey at the intersection of high-energy physics and artificial intelligence. Its focal point is the development and application of sequence-to-sequence transformer models for symbolic calculations of squared amplitudes in particle interactions—an innovative approach proposed to revolutionize this domain. By harnessing the power of machine learning, we aim to enhance the precision, efficiency, and accessibility of these crucial calculations. During this journey, we will address the unprecedented challenge we faced during the development of this approach, and present a proof-of-concept for the symbolic calcu-

lation of the squared amplitude, a key element of cross-section calculations.

In the chapters that follow, we start delving into this interdisciplinary area by presenting an introduction to the theoretical framework of high-energy physics and squared amplitude calculations. Then, we introduce the basic concepts of machine learning and its application to high-energy physics, where we discuss related works that employ the sequence-to-sequence model in symbolic mathematics and physics. Next, we explore the architecture of transformers adjusted to the specific demands of symbolic calculations, the generation and preprocessing of data, and the empirical evaluation of our model's performance. Through a comparative analyses of the results, we illuminate the strengths and limitations of our approach, shedding light on numerous potential directions of improvement.

As we embark on this journey, we acknowledge the richness and beauty of scientific inquiry that has preceded us. We stand on the shoulders of giants—those physicists who have tirelessly striven to unravel the mysteries of the subatomic world and develop clever and sophisticated computational tools. In the spirit of scientific exploration, this dissertation seeks to forge new paths, guided by the belief that the fusion of artificial intelligence and high-energy physics holds the promise of furthering our understanding of the universe's most fundamental building blocks.

In the chapters that follow, we invite the reader to traverse this intersection of two dynamic disciplines—a convergence where artificial intelligence meets high-energy physics, and where transformative insights await discovery.

1.1 Motivation

In advancing our understanding of the universe and its elementary constituents and forces, physicists have to create computational tools that enable the testing of their theories, which they build to describe observations and experiments. These tools are crucial for testing theoretical models constructed to interpret observations and experimental data. However, the complexity and computational intensity of these tools can pose significant challenges. In high-energy physics, the calculation of squared amplitudes and cross sections is essential for validating theoretical models and measuring physical quantities. Unfortunately, these calculations often become extremely resource-intensive the greater the number of final-state particles or the greater the orders, hindering progress. Physicists frequently encounter computational bottlenecks. Manual manipulation of large symbolic expressions tends to be error prone, leading to the widespread adoption of domain-specific symbolic manipulation software tools, such as `FeynCalc` [62], `CompHep` [11], and `MARTY` [66]. It is a testament to the high mathematical sophistication and the depth of domain knowledge of the developers of these symbolic manipulation tools that such analytical calculations have been automated. Despite the success of these tools in performing these calculations, the issue of scalability, particularly in terms of managing the increase in computational time required, remains a persistent and critical challenge. Here is an example of the computational time required for squaring one amplitude of QED processes performed on a personal computer:

Process	`FeynCalc`	`MARTY`
$ee \rightarrow ee$	$\sim 1sec$	$\sim 1sec$
$ee \rightarrow ee\gamma$	$\sim 40sec$	$\sim 5sec$
$ee \rightarrow ee\gamma\gamma$	$> 10000sec$	$> 1000sec$

Table 1.1: computation time for the squaring one amplitude of QED processes performed on a personal computer

The Table. 1.1 shows how increasing the number of final state particles in the interac-

tion dramatically increases the computation time. This trend is further exacerbated when the complexity of the interaction is increased, particularly with the inclusion of loops. This indicates the need for innovative computational tools that can gracefully scale with both the increasing number of final state particles and the ascending orders of interaction. The advent of such tools might potentially contribute to accelerating advancements in high-energy theoretical calculations, and could perhaps play a role in catalyzing a paradigm shift, possibly propelling the field into new territories and areas of exploration that are not yet fully envisaged.

1.2   Why Symbolic Machine Learning?

The importance of symbolic forms of equations and physical quantities comes from the fact that a symbol form provides more information about the behavior of the physical system, underlying laws or symmetry, interrelationships among various physical quantities, which provides a clear and concise description of the physical system. The generality of the symbolic form and its applicability to multiple cases and scenarios enables physicists to make predictions about physical phenomena and explore limiting cases, asymptotic behavior, and constraints, which can provide valuable physical insights. Moreover, the symbolic form can be simplified algebraically and manipulated, which helps to choose the most convenient or insightful form of the equation for a particular problem, and makes the calculations more efficient. Moreover, symbolic expressions facilitate collaboration and communication between researchers with different backgrounds.
Machine learning has proven to be a powerful tool that helps automate many complex processes in science and technology and gives efficient ways of handling them. As symbolic computation in high energy physics gets more complicated and more time-consuming, it is worth exploring how machine learning could aid in this, since it has the ability to process and generate sequences of symbolic data. The power

- **The amplitude** ($e\,e \rightarrow e\,e\,\gamma$):

$$i\mathcal{M} = \frac{\frac{1}{2}\,ie^3\left(p_{3\rho}\gamma_\epsilon^\rho\,\gamma_{\rho\eta}\,A_j^{\rho*}(p_5)\,\mathbf{e}_{i\eta}^*(p_4)\,\mathbf{e}_{l\epsilon}^*(p_3)\,\mathbf{e}_{k\delta}(p_2)\,\mathbf{e}_{i\delta}(p_1) - \frac{1}{2}\,p_{5\sigma}\,\gamma_{\rho\epsilon}\,\gamma_\epsilon^\rho\,\gamma_{\rho\eta}\,\gamma_\epsilon^\sigma\,A_j^{\rho*}(p_5)\mathbf{e}_{i\eta}^*(p_4)\,\mathbf{e}_{l\epsilon}^*(p_3)\,\mathbf{e}_{k\delta}(p_2)\,\mathbf{e}_{i\delta}(p_1)\right)}{\left((m_e^2 - \vec{p_2}.\vec{p_4}) * \vec{p_3}.\vec{p_5}\right)}$$

- **The squared amplitude** ($e\,e \rightarrow e\,e\,\gamma$):

$$|\mathcal{M}|^2 = -\frac{e^6}{((\vec{p_3}.\vec{p_5})^2 * (m_e^2 - \vec{p_2}.\vec{p_4})^2)}(2m_e^6 + m_e^4 * (-\vec{p_1}.\vec{p_3} - \vec{p_1}.\vec{p_5} - \vec{p_2}.\vec{p_4} + 2\vec{p_3}.\vec{p_5}) + m_e^2 * (\vec{p_1}.\vec{p_2} * \vec{p_3}.\vec{p_4} + $$
$$\vec{p_1}.\vec{p_2} * \vec{p_4}.\vec{p_5} + \vec{p_1}.\vec{p_4} * \vec{p_2}.\vec{p_3} + \vec{p_1}.\vec{p_4} * \vec{p_2}.\vec{p_5} + \vec{p_1}.\vec{p_5} * \vec{p_3}.\vec{p_4} - \vec{p_2}.\vec{p_4} * \vec{p_3}.\vec{p_5}) - \vec{p_1}.\vec{p_2} * \vec{p_3}.\vec{p_5} * \vec{p_4}.\vec{p_5} - \vec{p_1}.\vec{p_4} * \vec{p_2}.\vec{p_5} * \vec{p_3}.\vec{p_5})$$

Figure 1.1: Amplitude and squared amplitude of the $ee \rightarrow ee\gamma$ scattering process.

of machine learning to learn from context, recognize patterns, incorporate domain knowledge, and its scalability makes it suitable for these tasks. Furthermore, in high-energy physics, there is an abundance of data by nature. What is more important here is that these data have many physical symmetries which makes the patterns in the data more apparent, and there are constraints imposed by physical laws that make the form of the data well-defined and in a fixed-structural form. As sequence-to-sequence models in machine learning achieved remarkable success in performing symbolic mathematics (see section 4.2), like solving calculus problems, solving differential equations, and simplifying mathematical expression, there is a compelling case for delving into their potential applications within high-energy physics calculations. This exploration could unveil new possibilities and enhance our computational capabilities in this complex field.

1.3 Why Squared Amplitude?

Although machine learning can be utilized for a variety of symbolic calculations in high-energy physics, we have opted to specifically focus on applying it to the task of squaring the amplitude of particle interactions, which is related to the likelihood of a particular interaction or process occurring. An example of amplitude and squared amplitude is shown in Fig 1.1. Some of these reasons for this choice include:

- The squared amplitude is the main ingredient of cross sections, which are among the most pivotal quantities for the high-energy physics community. Consequently, developing a tool to efficiently compute squared amplitudes holds significant potential for utility and impact within the field.

- The squared amplitude is ubiquitous across various sub-fields of high-energy physics, and making improvements in this area could have wide-ranging implications for the field.

- Computing the squared amplitude is a highly complicated process because it involves many mathematical operations, including Lorentz index contractions, color factor calculation, matrix multiplications, Dirac algebra, traces, and, in some cases, integrals. This imposes a significant challenge; hence, success in this area will be highly impactful, and techniques developed for this problem may have applications in other areas of physics and beyond.

- Calculating the squared amplitude is a task that demands substantial computational resources. This complexity escalates markedly with the addition of more particles or the inclusion of higher-order interactions. Therefore, seeking a more efficient method for conducting these calculations is important and could prove to be immensely beneficial.

- As we are employing a supervised machine learning model, which necessitates a substantial dataset of examples for training purposes, generating a sufficient numbers of squared amplitude examples, using domain-specific algebraic programs, is possible in practice.

- Tackling this complex problem encourages the development of new machine-learning techniques and computational methods and demonstrates the viability of machine learning for solving fundamental problems in high-energy physics.

1.4   Challenges

Venturing into uncharted territory by applying machine learning models to square the amplitude of particle interactions in high-energy physics presents unique challenges, given that it is a novel approach and lacks precedent. Some of these challenges include:

- The lack of pre-existing frameworks and guidelines and the need to develop a framework from scratch is a significant challenge. Likewise, deciding on the machine learning model, architecture, features, parameters, and data preprocessing requires careful consideration and experimentation.

- The inherent complexity of squaring amplitudes requires looking for innovative approaches in machine learning to manage computational demands while maintaining high accuracy.

- Symbolic mathematics presents a complete contrast to natural language tasks, where outputs can be partially correct and still hold value. In symbolic mathematics, the results are either correct or entirely incorrect. A single erroneous token in the output can render the entire answer invalid, highlighting the critical importance of accuracy in this task.

- Generating high-quality training data for the model is essential, especially given the lack of pre-existing datasets for this specific problem, which creates the need to look for the optimal way of performing this step.

- Since the expressions of the amplitude and the squared amplitude are long and complicated, it is necessary to find a way to simplify them and make them more manageable for the machine learning model. A good representation makes the underlying patterns and relationships in the data more apparent, which

facilitates easier learning by the model and leads to higher accuracy in prediction and robust performance.

- The task requires a deep understanding of both high-energy physics and machine learning, so acquiring interdisciplinary knowledge and collaboration is necessary.

As we pioneer the application of machine learning to squared particle interaction amplitudes in high-energy physics, we encounter these challenges, including the absence of established frameworks, the need for precision in symbolic mathematics, and the generation of high-quality training data. Simplifying complex expressions and fostering interdisciplinary collaboration are also critical. These challenges motivate us to explore innovative solutions and shape the future of this novel approach.

Chapter 2 The Standard Model of Particle Physics

2.1   Overview

One of the most successful achievements in the history of physical science is the Standard Model (SM) of particle physics. It is a theory that describes the fundamental constituents of matter and their interactions. It is our best theory to describe the microscopic world in a consistent and computable framework. The standard model of particle physics, developed in the 1960s and 1970s, attempts to explain all the phenomena of particle physics. It has stood for 50 years as 'the' theory of particle physics, passing numerous stringent tests. It incorporates special relativity and quantum mechanics so it is based on quantum field theory (QFT), which is the language in which we codify our knowledge about the fundamental laws of nature in a manner compatible with quantum mechanics, relativity, and locality. The standard model SM is based on symmetry principles: gauge invariance and Lorentz invariance (special relativity). The SM incorporates three of the four known fundamental forces: the strong nuclear force, the weak nuclear force, and the electromagnetic force, each of which is associated with a symmetry called a gauge symmetry. All observed microscopic phenomena can be attributed to one or more of these interactions. For instance, the strong force is responsible for keeping protons and neutrons together within atomic nuclei, the electromagnetic force is what binds electrons to atomic nuclei and atoms within molecules, and weak interactions are responsible for the nuclear β-decays of certain radioactive isotopes and the nuclear fusion processes that fuel the Sun. The fourth fundamental is gravity, but its effect on particle processes at energies we can currently explore is totally negligible.

 In the Standard Model, the elementary particles can be classified based on their spin characteristics, which is a property that describes the intrinsic angular momentum

Standard Model of Elementary Particles

Figure 2.1: Standard model of elementary particles (by MissMJ — Own work by uploader, PBS NOVA, Fermilab, Office of Science, United States Department of Energy, Particle Data Group.)

of a particle 2.1. There are two types of particles in the Standard Model: **fermions**, matter particles that have a spin of ½. They obey the Pauli exclusion principle, which means that no two identical fermions can occupy the same quantum state. The fermions in the SM are **quarks**, which come in six flavors—up, down, charm, strange, top, and bottom, and **leptons**, which come into three flavors-electron, muon, and tau, along with their corresponding neutrinos (electron neutrino, muon neutrino, and tau neutrino). Quarks are confined within larger particles called hadrons (like protons and neutrons) and are held together by the strong force. The other type of particles are the **bosons**, which have integer spin values. There are three spin-1 particles in the SM, called gauge bosons, that are force carriers: **gluon**, the force carrier for the strong force, the **W and Z**, the force carriers for the weak force, and the **photon**, the force carrier for the electromagnetic force. There is a spin-0 boson called the **Higgs boson** which is associated with the Higgs field, which gives mass

to other particles through a mechanism known as the Higgs mechanism [35].

2.2   Theoretical framework

The Standard Model is a quantum field theory where particles are treated as excited states or "quanta" of underlying fields, which permeate all of space-time. It is based on wave equations: Dirac, Maxwell, Yang-Mills, and Klein-Gordon, which all generalize Schrodinger's equation to incorporate special relativity. The dynamics and interactions arise from gauge symmetry. The Standard Model is based on the local gauge group $SU(3)_C \times SU(2)_W \times U(1)_Y$. Each factor corresponds to a fundamental force: $SU(3)_C$ for the strong force, and $SU(2)_W \times U(1)_Y$ for the electroweak force, which unifies the electromagnetic and weak forces. The behavior and interactions of the fields in the SM are described by a mathematical construct known as the Lagrangian density $\mathcal{L}$. This entity encapsulates the dynamics of all the fields and their interactions. All interactions are determined from the symmetries except their strengths (coupling constants) and masses which are determined from experimental data, like α the fine structure constant or m_e the mass of the electron. Let us talk about each part in detail:

- **Relativistic quantum mechanics:** The first attempt to incorporate special relativity into Schrodinger's equation in quantum mechanics is the Klein-Gordon equation. By starting from the expectation that the free theory should have plane wave solutions of the form $\phi \propto e^{-iEt+i\mathbf{p}.\mathbf{x}} = e^{-ip_\mu x^\mu}$ and noting that the relativistic dispersion relation $p_\mu p^\mu = m^2$ should be reproduced, we get the Klein-Gordon equation:

$$(\partial_\mu \partial^\mu + m^2)\phi = 0. \tag{2.1}$$

It is a second-order equation in space and time, manifestly Lorentz-covariant and it describes scalar (spin-0) particles. Later on, Dirac derived a relativistic

equation that is linear in time and space which has the form:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \tag{2.2}$$

where the 4 constant γ^μ are 4×4 matrices which satisfy the anti-commutation relation: $\{\gamma^\mu, \gamma_\nu\} = 2\eta_{\mu\nu}$. Since the γ^μ are 4×4 matrices, the wavefunction ψ should have 4 components. It transforms in a special way under Lorentz transformations and we call it a 4-component spinor. The Dirac equation describes spin-$1/2$ particles.

- **Lagrangian formulation:** Imagine we have a field on spacetime, which we denote generically by $\phi(x^\mu)$. Just like in classical mechanics, the action, S, is obtained by integrating the Lagrangian, L, over time. Now, we shall restrict ourselves to theories in which the Lagrangian can be obtained by integrating something called the Lagrangian density, $\mathcal{L}$, over space:

$$S = \int dt L = \int d^4 x^\mu \mathcal{L}(\phi(x), \partial^\mu \phi(x)). \tag{2.3}$$

Since we almost always deal with the Lagrangian density only, we will simply call it the Lagrangian. Given the Lagrangian, the classical (Euler-Lagrange) equations of motion are obtained by extremizing the action. Thus, consider the variation δS that results from a field variation $\delta\phi$:

$$\delta S = \int d^4 x \left(\frac{\partial \mathcal{L}}{\partial \phi} \delta\phi + \frac{\partial \mathcal{L}}{\partial(\partial^\mu \phi)} \delta(\partial^\mu \phi) \right) \tag{2.4}$$

$$= \int d^4 x \left(\frac{\partial \mathcal{L}}{\partial \phi} - \partial^\mu \frac{\partial \mathcal{L}}{\partial(\partial^\mu \phi)} \right) \delta\phi, \tag{2.5}$$

where we have integrated by part and dropped the total derivative term based on the assumption that the field vanishes at boundaries. The action is thus extremal when:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial^\mu \frac{\partial \mathcal{L}}{\partial(\partial^\mu \phi)} = 0, \tag{2.6}$$

12

which is called (Euler-Lagrange) equations. This formalism is useful for identifying symmetries of the dynamics and its consequent implications, which are encoded in Noether's theorem. Suppose that the action is invariant under some symmetry transformation of the fields, $\phi \to \phi + \delta\phi$. The fact that the action is invariant means that the Lagrangian can change at most by a total derivative, $\partial_\mu K^\mu$ (which integrates to zero in the action). Then,

$$\delta\mathcal{L} = \partial_\mu K^\mu = \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi)}\delta\partial^\mu\phi \tag{2.7}$$

$$= \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi - \partial^\mu\frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi)}\delta\phi + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi)}\delta\phi\right) \tag{2.8}$$

The first two terms on the right-hand side cancel when the equation of motion holds. Thus, classically, we have the conserved current:

$$\partial_\mu J^\mu = 0, \quad where \quad J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi)}\delta\phi - K^\mu \tag{2.9}$$

2.2.1 Quantum Electrodynamics (QED)

Quantum Electrodynamics (QED) is the quantum field theory that describes the interactions between charged particles through the exchange of photons. It is a subset of the Standard Model of particle physics and stands as one of the most precisely tested theories in physics. The QED Lagrangian can be derived by requiring that the free Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\slashed{\partial} - m)\psi, \tag{2.10}$$

where $\slashed{\partial} \equiv \gamma^\mu\partial_\mu$ and γ^μ are the Dirac matrices, be invariant under a $U(1)$ *local* transformation, $\psi(x) \to \psi'(x) = e^{ie\chi(x)}\psi$. With this transformation, the Lagrangian becomes:

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L} - e\bar{\psi}\gamma^\mu(\partial_\mu\chi)\psi. \tag{2.11}$$

If we replace the derivative ∂_μ in 2.10 with the *covariant derivative* D_μ,

$$\partial_\mu \to D_\mu = \partial_\mu + ieA_\mu, \tag{2.12}$$

where A_μ is a new field, the $e\bar{\psi}\gamma^\mu(\partial_\mu\chi)\psi$ term gets cancelled provided the new field transforms as:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu\chi. \tag{2.13}$$

Hence, the gauge-invariant Lagrangian for a spin-half fermion;

$$\mathcal{L} = \bar{\psi}(i\slashed{\partial} - m)\psi - e\bar{\psi}\gamma^\mu A_\mu\psi, \tag{2.14}$$

contains a term describing the interaction of the fermion with the new vector field A_μ, which can be identified as the photon. Adding a kinetic term for the spin-1 field $F_{\mu\nu}F^{\mu\nu}$ which is invariant under $U(1)$, we get the QED Lagrangian:

$$\mathcal{L}_{QED} = \bar{\psi}(i\slashed{\partial} - m)\psi - e\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.15}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor, and the coupling constant e is interpreted as the electric charge. Using the gauge covariant derivative $D_\mu = \partial_\mu + ieA_\mu$, the QED Lagrangian can be written as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\slashed{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.16}$$

where $\slashed{D} \equiv \gamma^\mu(\partial_\mu + ieA_\mu)$. From this Lagrangian, we can extract the QED Feynman rules, draw Feynman diagrams, and compute physical quantities that can be measured. Examples of $2 \to 2$ processes that are computed in QED are Compton scattering ($\gamma e^- \to \gamma e^-$), Møller scattering ($e^- e^- \to e^- e^-$) and Bhabha scattering ($e^- e^+ \to e^- e^+$).

QED is the most successful theory in physics in terms of the theoretical and experimental precision of its tests. Its predictions are among the most accurately tested in all of physics. It has passed the classical atomic tests, such as the Lamb shift, atomic hyperfine splittings, muonium, and positronium, see [44]. Furthermore, the theory's predictions about the anomalous magnetic moment of the electron agree with experimental measurements to an astounding level of precision. Such precise agreement

between theory and experiment reinforces QED as a fundamental and accurate description of nature.

While QED can achieved remarkable success at the tree-level, it is essential to incorporate higher-order (loop) effects in perturbation theory, because some effects only occur at the loop level. The calculation of higher-order effects is greatly complicated by divergences and the need to renormalize (while maintaining gauge invariance). The divergence we often encounter called "UV divergence" comes from integrating momenta goes to infinity, which does not converge. The technique to handle these UV divergences is called renormalization. The idea is to introduce counterterms in the Lagrangian that cancel out the infinities. The infinities are absorbed into redefinitions of the parameters of the theory, such as the electron's charge and mass. Once the counterterms are introduced, the parameters are said to be "renormalized." Then we use dimensional regularization, which involves analytically continuing the number of spacetime dimensions away from the physical four dimensions (3 space + 1 time) and performing the integrals in this d dimensions. In dimensional regularization, UV divergences manifest as poles ($\frac{1}{\epsilon}$ terms) as ϵ approaches zero. These poles are more manageable and mathematically well-defined compared to the outright infinities that one would encounter in 4-dimensional calculations. After the integrals are regulated and renormalized, the limit $\epsilon \to 0$ is taken to return to the physically relevant four-dimensional spacetime, but now without the infinities, and gauge invariance is preserved. There is another divergence called Infrared Divergence (IR) that happens when the loop momentum goes to zero, which reflects real physical phenomena, such as the emission of soft photons. These divergences are usually handled by introducing an IR cutoff or assigning the massless particle a fictitious mass, and then taking the limit as the fictitious mass vanishes.

2.2.2 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD) is a quantum field theory and the sector of the Standard Model that describes the interactions of quarks and gluons, the fundamental constituents of hadrons like protons and neutrons. QCD is a non-abelian gauge theory based on the $SU(3)_C$ group, where the subscript c refers to "color," an abstract charge-like property that quarks and gluons carry. Analogous to electric charge in QED, QCD introduces the concept of 'color' charge. Unlike electric charge which comes in positive and negative, color charge in QCD has three types: red, blue, and green. Anti-quarks carry anticolors: anti-red, anti-blue, and anti-green, the "negative" charge. The force carriers of QCD are the gluons. Unlike photons in QED, which do not carry an electric charge, gluons carry a color charge and an anticolor charge. This unique feature leads to the self-interaction of gluons, making the behavior of the strong force vastly different from the electromagnetic force. One essential feature of QCD is asymptotic freedom, which implies that the strong interaction becomes weaker as quarks come closer together (high energy) and becomes stronger as they move apart. This behavior, confirmed experimentally, is opposite to the electromagnetic force, where charges feel a stronger force as they approach each other.

To construct the QCD Lagrangian, we first need to know how non-Abelian gauge theory differs from Abelian gauge theory. In QED, the gauge group is $U(1)$, which is an Abelian group. This means that the group's generators commute with each other. The photon, which is the gauge boson of QED, does not interact with itself. Non-Abelian gauge groups, on the other hand, have generators which don't commute. As a result, the gauge bosons do interact with themselves. We call theories that are based on non-Abelian gauge group *Yang-Mills theories.*

Consider a theory with n fields represented by the potentials A_μ^a $(a = 1, 2, 3, ..., n)$. The gauge invariance occurs under the local transformation:

$$\Psi \rightarrow \Psi' = e^{-igT_a\alpha^a(x)}\Psi \tag{2.17}$$

16

The group generator of the potential A_μ^a is T_a, which obeys the commutation relation $[T_a, T_b] = i f_{abc} T_c$, where the coefficients f_{abc} are the structure constants of the group. The covariant derivative now becomes:

$$D_\mu = \partial_\mu - ig T_a A_\mu^a. \tag{2.18}$$

By analogy with the electromagnetic field tensor $F_{\mu\nu}^a$, the field tensor is defined as:

$$ig T_a F_{\mu\nu}^a = [D_\nu, D_\mu]. \tag{2.19}$$

Substituting the covariant derivative and commutator of the generators of the group is obtained:

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{bc}^a A_\mu^b A_\nu^c. \tag{2.20}$$

We see that the first two terms are similar to the field tensor in QED, but the last term reflects the fact that the generators are not part of Abelian group.

Going back to QCD, the Lagrangian is invariant under local gauge transformations of the $SU(3)_C$ group $\Psi \to \Psi' = e^{-ig_s T_a \alpha^a(x)} \Psi$. The strong interaction is represented by eight gauge fields for gluons $G_{\mu\nu}^a$ $(a = 1, 2, ..., 8)$, and each of these fields is associated with a generator $T_a = \frac{\lambda_a}{2}$, where λ_a is a Gell-Mann matrix. The covariant derivative is given by:

$$D_\mu = \partial_\mu - ig_s G_{\mu\nu}^a T_a \tag{2.21}$$

The coupling constant of the fields with gluons is g_s. The Lie algebra is defined by the commutation rule: $[T_a, T_b] = i f_{abc} T_c$ where f_{abc} is the structure constant of the $SU(3)$ group. Quarks are represented by spinor fields ψ, and since there are six different flavors of quarks (*up, down, charm, strange, top and bottom*), we can represent these fields collectively with a three-component spinor in flavor space. Thus, the QCD Lagrangian can be written as the sum of the quark kinetic term, the gluon kinetic term, and their interaction:

$$\mathcal{L}_{QCD} = \sum_f \bar{\psi}_f (i\slashed{D} - m)\psi_f - \frac{1}{4} G_{\mu\nu}^a G^{\mu\nu a}, \tag{2.22}$$

where a is a color label, taking values from 1 to 8 for $SU(3)$, f runs over quark flavors, and the field strength tensor is given by:

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^a_{bc} A^b_\mu A^c_\nu \qquad (2.23)$$

Some of the calculations in QCD are similar to the QED except for the boson (gluon) self-interaction, and color factors must be taken into account. Processes involving non-abelian vertices are extremely tedious to carry out by hand and are best handled by specialized computer algebra programs. Higher-order calculations involve all of the subtleties of gauges and ghost loops (the latter may even appear in tree-level calculations involving external gluons, which are treated in standard field theory texts.) The gluon self-interactions in QCD imply asymptotic freedom, so that one can treat the quarks and gluons as weakly coupled, and processes such as deep inelastic scattering can be calculated in perturbation theory. However, at low energies, perturbation theory no longer holds. The strong coupling and gluon self-interactions presumably lead to the confinement of quarks, gluons, and any colored states, so that only color singlet hadron states can emerge.

2.2.3 Electroweak Theory (EW)

The electroweak theory is a fundamental cornerstone of the Standard Model of particle physics. It provides a unified description of two of the four known fundamental forces in nature: electromagnetism (mediated by the photon) and the weak force (mediated by the $W^\pm$ and Z^0 bosons). This theory was developed during the mid-20th century and its key architects include Sheldon Glashow, Abdus Salam, and Steven Weinberg, all of whom were awarded the Nobel Prize in Physics in 1979 for their contributions. The electroweak interactions are described by the gauge group $SU(2)_L \otimes U(1)_Y$. The subscript L means that these transformations only act on components of the left chiral fermions and Y is the hypercharge quantum number. Electroweak interactions are represented by three non-Abelian gauge fields W^i_μ ($i = 1, 2, 3$) and Abelian gauge

field B_μ. The generators of the weak isospin group $SU(2)_L$ are $\tau_i = \sigma_i/2$ and the generators for the $U(1)_Y$ group is $y = IY/2$, where σ_i are the Pauli matrices, I the 2×2 unit matrix, and Y hypercharge quantum number. The coupling constants are g and g'. The covariant derivative is given by:

$$D_\mu = I\partial_\mu - ig\frac{\sigma_i}{2}W_\mu^i - ig'I\frac{Y}{2}B_\mu. \tag{2.24}$$

From the definition in 2.20, we can write the tensor field:

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + gf_{jk}^i W_\mu^j W_\nu^k \tag{2.25}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{2.26}$$

The Lagrangian for the free fields is:

$$\mathcal{L} = -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{2.27}$$

Performing the matrix operations in the first term of the covariant derivative 2.24, we can define:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \tag{2.28}$$

With the linear combination of the third component of the gauge field, W_μ^3 with the gauge field B_μ, we can define the Z_μ^0 boson and the photon A_μ^0 as the following:

$$A_\mu^0 = \cos\theta_W B_\mu + \sin\theta_W W_\mu^3 \tag{2.29}$$

$$Z_\mu^0 = -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3, \tag{2.30}$$

where the mixing angle, θ_W, called Weinberg angle; is defined by the ratio of the electroweak coupling constants,

$$\tan\theta_W = g'/g, \tag{2.31}$$

or equivalently, $e = g\sin\theta_W = g'\cos\theta_W$, where e is the magnitude of the electron charge. The hypercharge Y is related to the electric charge Q by the relation: $Q =$

$\frac{1}{2}(\sigma_3 + IY)$. Substituting these definitions, with the boson definitions in 2.29 and 2.30 in the covariant derivative 2.24, we obtain:

$$D_\mu = I\partial_\mu - ieQA_\mu^0 - i\frac{g}{\sqrt{2}}(\sigma^+ W_\mu^+ + \sigma^- W_\mu^-) - i\frac{g}{\cos\theta_W}(\frac{\sigma^3}{2} - Q\sin^2\theta_W)Z_\mu^0, \quad (2.32)$$

where $\sigma^\pm = \frac{1}{\sqrt{2}}(\sigma^1 \mp i\sigma^2)$. We know from experiments that the weak force violates parity, so we need to separate each fermion field into right-handed and left-handed chiral components, $\Psi = \Psi_L + \Psi_R$ using the chiral projectors $P_L = \frac{1}{2}(1 - \gamma^5)$ and $P_R = \frac{1}{2}(1 + \gamma^5)$, where $\gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4$, and $\Psi_L = P_L\Psi$ and $\Psi_R = P_R\Psi$.

Fermions, quarks and leptons, are classified into three groups called generations. A generation is formed by quarks and leptons with different charges, and generations are ordered from the lightest to the heaviest. The left component of the leptons is given by a doublet:

$$L \; doublet: \qquad l_L = \begin{pmatrix} \nu_i \\ e_i \end{pmatrix}_L, \qquad\qquad q_L = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L \qquad (2.33)$$

$$R \; singlet: \qquad u_R^i, \; d_R^i, \; e_R^i, \qquad\qquad (2.34)$$

where $(i = 1, 2, 3)$ labels the generations. Notice that there are no right-handed neutrinos because in the SM, neutrinos are massless —there is no right-chiral component. To describe the leptons and quarks interacting with the weak and electromagnetic fields, we use the Lagrangian:

$$\mathcal{L} = \bar{l}_L^i i\gamma^\mu D_{L\mu} l_L^i + \bar{l}_R^i i\gamma^\mu D_{R\mu} l_R^i. \qquad (2.35)$$

The left-handed components belong to the $SU(2)$ group while the right ones only to the $U(1)$ group. The mass term is absent in the Lagrangian because such a term is not gauge invariant. The covariant derivative $D_{L\mu}$ in this Lagrangian is given by

2.32, and $D_{R_\mu} = \partial_\mu - ieQA^0_\mu - i\frac{g}{\cos\theta_W} - Q\sin^2\theta_W Z^0_\mu$. Substituting these into the Lagrangian gives:

$$\mathcal{L}_l = i\bar{\nu}_i\gamma^\mu P_L\nu_i + i\bar{e}_i\gamma^\mu\partial_\mu e_i$$

$$- e\bar{e}_i\gamma^\mu e_i A^0_\mu + \frac{g}{2\cos\theta_W}\bar{\nu}_i\gamma^\mu P_L\nu_i Z^0_\mu$$

$$+ \frac{g}{\sqrt{2}}\bar{\nu}_i\gamma^\mu P_L e_i W^+_\mu + \frac{g}{\sqrt{2}}\bar{e}_i\gamma^\mu P_L\nu_i W^-_\mu$$

$$+ \frac{g}{2\cos\theta_W}\bar{e}_i\gamma^\mu(2\sin^2\theta_W p_R + (2\sin^2\theta_W - 1)P_L)e_i Z^0_\mu.$$

Similarly for the quarks, the Lagrangian is:

$$\mathcal{L}_q = i\bar{u}_i\gamma^\mu\partial_\mu u_i + i\bar{d}_j\gamma^\mu\partial_\mu d_j$$

$$+ \frac{2e}{3}\bar{u}_i\gamma^\mu u_i A^0_\mu - \frac{e}{3}\bar{d}_j\gamma^\mu d_j A^0_\mu$$

$$+ \frac{g}{\sqrt{2}}V_{ij}\bar{u}_i\gamma^\mu p_L d_j W^+_\mu + \frac{g}{\sqrt{2}}V^\star_{ij}\bar{d}_j\gamma^\mu p_L u_i W^-_\mu$$

$$+ \frac{g}{\cos\theta_W}\bar{u}_i\gamma^\mu(\frac{1}{2}P_L - \frac{2}{3}sin^2\theta_W)u_i Z^0_\mu$$

$$+ \frac{g}{\cos\theta_W}\bar{d}_j\gamma^\mu(-\frac{1}{2}P_L + \frac{1}{3}sin^2\theta_W)d_j Z^0_\mu.$$

To generate mass terms we use the *Higgs mechanism*, which is a crucial part of the Standard Model. This process involves introducing a scalar field called the Higgs field, a ubiquitous scalar field with a non-zero vacuum expectation value. As quantum fields interact with the Higgs field, they acquire mass proportional to the strength of this interaction. The Higgs mechanism operates through a process called *Spontaneous Symmetry Breaking (SSB)*; the Higgs field possesses a symmetric potential; however, its ground state, is not symmetric. Instead of the field resting at a zero value (which would preserve the symmetry), it settles in a non-zero value (vacuum expectation value, *vev*), thereby breaking the symmetry spontaneously. As other fields interact with this non-zero Higgs field, they acquire mass. This phenomenon of spontaneous symmetry breaking is crucial to the Higgs mechanism, allowing particles to have mass without explicitly violating gauge symmetry.
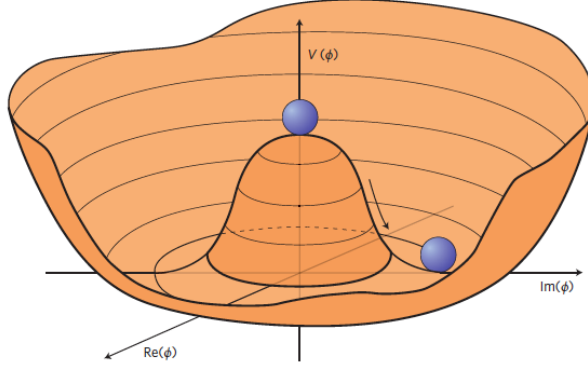
Figure 2.2: Higgs Potential (From: [30])

To implement spontaneous symmetry breaking, let us introduce a doublet of complex scalar Higgs, $\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$. The Lagrangian of this field is:

$$\mathcal{L} = (D_\mu \Phi)^\dagger D^\mu \Phi - V(\Phi), \qquad (2.36)$$

with the Higgs potential,

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \qquad (2.37)$$

For $\mu^2 < 0$, the potential has an infinite set of degenerate minima (see Fig. 2.2). The real part of the neutral field ϕ^0 can be written as $\phi^0 = \frac{1}{\sqrt{2}}(h + v + i\eta)$. After symmetry breaking, $SU(2)_L \otimes U(1)_Y \to U(1)_{em}$, the minimum of the potential must correspond to a non-zero vacuum expectation value only of the neutral scalar field Φ^0, such that,

$$\langle \Phi \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \qquad (2.38)$$

When the fields are expanded around the ground state, the kinetic part of the Lagrangian becomes:

$$\mathcal{L} = M_W^2 W_\mu^+ W^{\mu-} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu, \qquad (2.39)$$

where, $M_w = \frac{gv}{2}$, is the W boson mass; $M_Z = \frac{gv}{2\cos\theta_W}$, Z boson mass and the A_μ boson remains massless. The two free parameters of the Higgs potential μ and λ are

related to the vacuum expectation value of the Higgs field v and to the mass of the Higgs boson M_H so that,

$$v^2 = \frac{-\mu^2}{\lambda}, \qquad M_H^2 = 2\lambda v^2. \tag{2.40}$$

The leptons and quarks acquire mass through Yukawa interactions between fermion fields and the Higgs field:

$$-\mathcal{L}_Y = \bar{l}_{iL}\lambda_{ij}\Phi e_{jR} + \bar{q}_{iL}\lambda_{ij}\tilde{\Phi} u_{jR} + \bar{q}_{iL}\lambda_{ij}\Phi d_{jR} + h.c., \tag{2.41}$$

where $\tilde{\Phi} = i\tau_2\Phi^*$, and the mass term can be defined as $m_{ij} = \lambda_{ij}v/\sqrt{2}$. Therefore, the Lagrangian for the ground state ($\Phi \to \langle\Phi\rangle_0$) is:

$$
\begin{aligned}
\mathcal{L}_Y =& m_{e_i}\bar{e}_i e_i + m_{u_i}\bar{u}_i u_i + m_{d_i}\bar{d}_i d_i \\
& \frac{g}{2M_W}(m_{e_i}\bar{e}_i e_i + m_{u_i}\bar{u}_i u_i + m_{d_i}\bar{d}_i d_i)\ h^o \\
& + \frac{ig}{2M_W}(m_{e_i}\bar{e}_i\gamma^5 e_i + m_{u_i}\bar{u}_i\gamma^5 u_i + m_{d_i}\bar{d}_i\gamma^5 d_i)\ \eta^o \\
& \frac{gm_{ei}}{\sqrt{2}M_W}\bar{\nu}_i P_R e_i\phi^+ + \frac{gm_{e_i}}{\sqrt{2}M_W}\bar{e}_i P_L \nu_i\phi^- \\
& \frac{g}{\sqrt{2}M_W}\bar{u}_i(m_{d_j}P_R - m_{u_i}P_L)V_{ij}d_j\phi^+ \\
& \frac{g}{\sqrt{2}M_W}\bar{d}_i(m_{d_j}P_L - m_{u_i}P_R)V_{ij}^* u_j\phi^-.
\end{aligned}
$$

In summary, the electroweak theory is a chiral gauge theory with gauge group $SU(2)_L \otimes U(1)_Y$. This symmetry is spontaneously broken down to $U(1)_{em}$ by the Higgs mechanism which generates the gauge bosons, Higgs, and all fermion masses. The electroweak theory is extremely well tested experimentally due to two properties: observables can be reliably calculated in perturbation theory, and it can be tested in lepton-lepton collisions, which allow for very precise measurements.

2.3 Experimental Test of the Standard Model

The Standard Model (SM) exhibits remarkable agreement with a vast array of experimental results —see, for example, how cross section measurement, from the ATLAS agree with the Standard Model predictions 2.3. Precision experiments conducted over decades at high-energy physics laboratories around the world, such as CERN and Fermilab, have tested the Standard Model predictions and have firmly established its validity at energies up to the electroweak scale. Precise measurements of the properties of quarks and leptons and the force-carrying particles like the W and Z bosons, gluons, and the Higgs boson, confirm that the Standard Model's predictions to an astonishing degree of accuracy. The discovery of the Higgs boson at CERN in 2012 was a monumental milestone, filling in the last missing piece of the Standard Model and solidifying the theory's foundational status in physics.
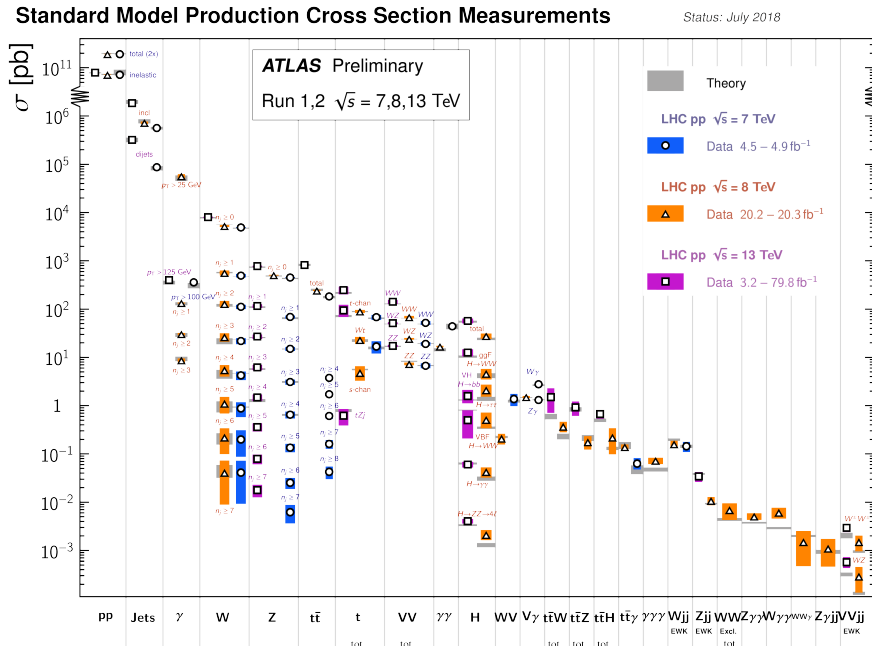
Figure 2.3: Summary of several Standard Model total production cross section measurements. (From: ATLAS Collaboration [1])

2.4 Challenges and Search for New Physics Beyond The Standard Model

Although The Standard Model of particle physics is an incredibly successful theory, it does have a number of known limitations and unanswered questions that present significant challenges highlight the need for a more comprehensive framework. The theoretical challenges of the Standard Model include:

- The Hierarchy Problem: If we consider the Standard Model as an effective field theory up to some cutoff Λ, loop corrections from the Standard Model to the Higgs mass yield a quadratic divergence. In other words, these corrections depend quadratically on the energy scale Λ^2 up to which the Standard Model is valid (potentially the Planck scale $\Lambda \sim M_{pl}$). This would naturally drive the mass of the Higgs boson to be at the Planck scale, far higher than the observed value at 125 GeV, unless there is an incredible fine-tuning of parameters to cancel out these corrections. This seems unnatural and calls for an explanation. The hierarchy problem is often seen as an indication that there must be new physics beyond the Standard Model that comes into play at energy scales between the electroweak and the Planck scales to stabilize the Higgs mass.

- The Strong CP Problem: It is related to the quantum chromodynamics (QCD) sector of the Standard Model where there is a term in the QCD Lagrangian that can violate CP symmetry, characterized by the θ parameter. If θ is nonzero, it would lead to observable effects, such as a nonzero electric dipole moment for the neutron. However, experiments have placed extremely tight constraints on the electric dipole moment of the neutron, implying that θ is very close to zero. The question then arises as to why θ is so small, essentially indistinguishable from zero. This also appears to require fine-tuning, which is generally considered to be unnatural unless there is a mechanism that explains it.

- Flavor Puzzle: The Standard Model includes three generations of quarks and

leptons, each with different masses and mixing patterns. The flavor puzzle is the question of why the particle properties have the values they have, and why the mixing angles in the PMNS and CKM matrix have the values they have? For instance, the top quark is about 36,000 times heavier than the up quark, and the tau lepton is about 3,500 times heavier than the electron. The Standard Model does not provide an explanation for this vast disparity in masses, nor does it predict the absolute values of these masses.

Likewise, the Standard Model faces a range of experimental challenges that are central to our quest to understand the fundamental particles and forces of the universe. These experimental challenges include:

- Dark Matter: The Standard Model does not account for dark matter, which makes up about 27% of the total mass in the universe. Extensions or alternatives are needed to account for these observed phenomena.

- Neutrino Masses: In the Standard Model, particles acquire mass through the Higgs mechanism, where particles interact with the Higgs field. However, this mechanism applies to Dirac masses, which require both left-handed and right-handed components of the particles. Neutrinos are only observed as left-handed particles (or right-handed antiparticles), and the Standard Model does not include right-handed neutrinos. This means that the usual Higgs mechanism cannot generate neutrino masses. One possible solution is to introduce right-handed neutrinos and incorporate a see-saw mechanism, or consider neutrinos to be Majorana particles, meaning that they are their own antiparticles. This would allow for mass generation through a different mechanism but would also have profound implications for our understanding of particle physics.

- Matter-Antimatter Asymmetry: One of the central mysteries in physics and cosmology is the abundance of matter over antimatter in the observable uni-

verse. The Standard Model does not provide a satisfactory explanation for the observed asymmetry between matter and antimatter in the universe.

- Anomalies and Discrepancies: There are a few experimental results that show possible discrepancies with Standard Model predictions, such as the anomalous magnetic moment of the muon and certain B-meson decay anomalies and the discrepancy between different measurements of the neutron lifetime. Understanding whether these are signs of new physics or due to experimental or theoretical uncertainties is a significant challenge.

These theoretical and experimental challenges highlight the limitations of the Standard Model and bring the need for new ideas, extensions, or a more fundamental theory to address these questions and deepen our understanding of the fundamental constituents and forces of nature. There are many proposed solutions, including supersymmetry and extra dimensions, which have the potential to replace the Standard model; however, they lack experimental evidence.

2.5 Squared Amplitude and Cross Section

In particle physics, the cross section measures quantitatively the probability that two particles will collide and react in a certain way under certain conditions. The cross section is a crucial quantity both for theoretical computations and experimental observations. If you know the cross section for a particular process and how many particles you're shooting at a target, you can predict how many of those particles will undergo the interaction. It acts as a bridge between theoretical predictions and experimental observations, allowing scientists to validate or challenge established models like the Standard Model. By measuring and comparing cross sections of various processes, physicists can discern the presence of new particles. Further, any deviations between measured cross sections and theoretical expectations can hint at groundbreaking phenomena beyond current understanding.

Classically, the cross section is a measure of the number of scattered particles in a given area relative to the incident particles. In quantum mechanics, we can relate this quantity, called the "differential probability dP" to the cross section $d\sigma$ by

$$d\sigma = \frac{1}{T}\frac{1}{\Phi}dP,\qquad(2.42)$$

where T is the observation time and Φ is the flux which equals magnitude of the relative velocity of the incoming particles $\vec{v}$ divided by the total volume V in which the collisions take place: $\Phi = \frac{|\vec{v}|}{V}$. If we consider the case of two colliding particles that yield n particles;

$$p_1 + p_2 \rightarrow p_j,\qquad(2.43)$$

then the **differential cross section** in the rest frame of the incoming particle is

$$d\sigma = \frac{V}{T}\frac{1}{|\vec{v_1} - \vec{v_2}|}dP.\qquad(2.44)$$

We know from quantum mechanics that the probability of going from state $|i\rangle$ to state $\langle f|$ is just the squared of the amplitude $|\langle f| S |i\rangle|^2$. So the normalized differential

probability is

$$dP = \frac{|\langle f| S |i\rangle|^2}{\langle f|f\rangle \langle i|i\rangle} d\Pi, \tag{2.45}$$

where $d\Pi$ is the differential phase space that contains all final state momenta, which is given by:

$$d\Pi = \prod_j \frac{V}{(2\pi)^3} d^3 p_j. \tag{2.46}$$

The numerator in Eq. 2.45 $\langle f| S |i\rangle$ is the invariant matrix that contains all physical information about the dynamics of the particle interaction. In particular, to go from initial state $|i\rangle$ to final state $\langle f|$ assuming both states are eigenstate of the free Hamiltonian:

$$\lim_{t_\pm \to \pm\infty} \langle f| U(t_-, t_+) |i\rangle = \langle f| S |i\rangle, \tag{2.47}$$

where S is called the S-matrix. Expanding the S-matrix

$$\langle f| S |i\rangle = \langle f| \mathbb{I} + i\mathcal{T} |i\rangle, \tag{2.48}$$

and taking the non-trivial case $\langle f| i\mathcal{T} |i\rangle$ and factoring a δ-function, which imposes 4-momentum conservation, we can define the **invariant matrix elements** $\langle i| \mathcal{M} |f\rangle$ or **the amplitude** as:

$$\langle f| i\mathcal{T} |i\rangle = (2\pi)^4 \, \delta^4(\sum_i p_i - \sum_f p_f) \, \langle i| \mathcal{M} |f\rangle. \tag{2.49}$$

Squaring this quantity, we get:

$$\begin{aligned} |\langle f| i\mathcal{T} |i\rangle|^2 &= (2\pi)^8 \delta^4(0) \, \delta^4(\sum_i p_i - \sum_f p_f) \, |\langle i| \mathcal{M} |f\rangle|^2 \\ &= (2\pi)^4 TV \delta^4(\sum_i p_i - \sum_f p_f) \, |\langle i| \mathcal{M} |f\rangle|^2, \end{aligned} \tag{2.50}$$

where we used the fact that $\delta^4(0) = \frac{TV}{(2\pi)^4}$ since the δ-function here can be regulated by the finite volume. The normalization factors in Eq. 2.45 are given by:

$$\langle i|i\rangle = (2E_1 V)(2E_2 V), \quad \langle f|f\rangle = \prod_j (2E_j V). \tag{2.51}$$

Putting everything together, **the differential cross section** of a $2 \to n$ process is:

$$d\sigma = \frac{1}{(2E_1)(2E_2)|\vec{v_1} - \vec{v_2}|}|\mathcal{M}_{if}|^2 d\Pi_{LIPS} \tag{2.52}$$

where $d\Pi_{LIPS}$ is the *Lorentz-invariant phase space*, which equals:

$$d\Pi_{LIPS} = (2\pi)^4 \delta^4 (\sum_i p_i - \sum_f p_f) \prod_j \frac{d^3 p_j}{(2\pi^3)2E_{p_j}}. \tag{2.53}$$

The only part in 2.52 that is process-dependent is the squared amplitude $|\mathcal{M}_{if}|^2$, so, calculating the cross section of an interaction requires calculating the square of the amplitude, averaging over all the spin states of the incident particles and summing over all final spin states, while calculating the kinematic term and $d\Pi_{LIPS}$ is easy and process-independent.

To determine the S-matrix elements, hence the amplitude, one begins by specifying the initial and final particle states of interest and identifying the appropriate inter-action Lagrangian, drawing all possible Feynman diagrams for the process, up to the desired perturbation order. By applying the associated Feynman rules, each diagram translates into a mathematical contribution. To isolate the pure interaction from these contributions, external particle propagators are amputated, and the result is then multiplied by the respective wave packets using the LSZ reduction formula. The summed contributions yield the matrix element for the process.

Let us take an actual example of a process and try to compute the amplitude and squared amplitude. One of the simplest processes, albeit important, is annihilation of an electron and anti-electron (positron) into a muon and anti-muon $e^+ e^- \to \mu^+ \mu^-$, as shown in Fig. 2.4. We construct the amplitude (matrix elements) of this interaction using the QED Feynman rules:

$$i\mathcal{M} = (-ie)\bar{v}_\alpha(p_2)\gamma^\mu_{\alpha\beta}u_\beta(p_1)\frac{-ig_{\mu\nu}}{k^2}(-ie)\bar{u}_\delta(p_3)\gamma^\nu_{\delta\sigma}v_\sigma(p_4), \tag{2.54}$$
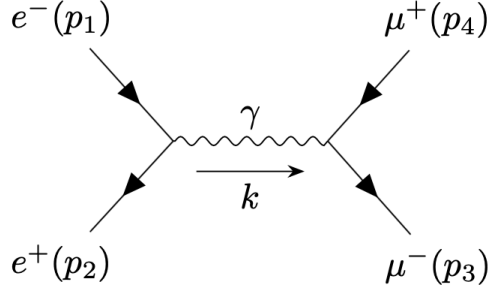
Figure 2.4: Feynman diagram of annihilation of electron and anti-electron (positron) into muon and anti-muon $e^+e^- \rightarrow \mu^+\mu^-$ process

where $k^\mu = p_1^\mu + p_2^\mu = p_3^\mu + p_4^\mu$. Each spinor comes with a spin label, i.e. $u_\alpha^{s_1}(p_1)$, but we keep them implicit. Rearranging this slightly, we get:

$$\mathcal{M} = \frac{e^2}{s}\bar{v}(p_2)\gamma^\mu u(p_1)\bar{u}(p_3)\gamma_\mu v(p_4), \tag{2.55}$$

where $s = (p_1 + p_2)^2$.

To calculate the squared amplitude $|\mathcal{M}|^2 = \mathcal{M}^\dagger \mathcal{M}$, we need to get the conjugate amplitude $\mathcal{M}^\dagger$. From the commutation relations of the gamma matrices, we can infer that $\gamma_\mu^\dagger \gamma_0 = \gamma_0 \gamma_\mu$ and $\gamma_0^\dagger = \gamma_0$; therefore:

$$(\bar{\psi}_1 \gamma^\mu \psi_2)^\dagger = (\bar{\psi}_1^\dagger \gamma_0 \gamma^\mu \psi_2)^\dagger = \bar{\psi}_2^\dagger \gamma^{\mu\dagger} \gamma_0^\dagger \psi_1 = \psi_2^\dagger \gamma_0 \gamma^\mu \psi_1 = \bar{\psi}_2 \gamma^\mu \psi_1. \tag{2.56}$$

Then, the squared amplitude becomes

$$|\mathcal{M}|^2 = \frac{e^4}{s^2}[\bar{v}(p_2)\gamma^\mu u(p_1)][\bar{u}(p_3)\gamma_\mu v(p_4)][\bar{v}(p_4)\gamma^\nu u(p_3)][\bar{u}(p_1)\gamma_\nu v(p_2)]. \tag{2.57}$$

For the unpolarized case, where the spin is not measured, we need to sum over the spin. We can do that using the trace technique, which derived form the Dirac equation as follows:

$$\sum_{s=1}^{2} u_\alpha^s(p)\bar{u}_\beta^s(p) = (\gamma^\mu p_\mu + m\mathbb{I})_{\alpha\beta} = (\not{p} + m)_{\alpha\beta} \tag{2.58}$$

$$\sum_{r=1}^{2} v_\alpha^r(p)\bar{v}_\beta^r(p) = (\gamma^\mu p_\mu - m\mathbb{I})_{\alpha\beta} = (\not{p} - m)_{\alpha\beta}. \tag{2.59}$$

31

This yields:

$$\sum_{s'}\sum_{s}[\bar{u}^{s'}(p_3)\gamma_\mu v^s(p_4)][\bar{v}^s(p_4)\gamma^\nu u^{s'}(p_3)] = (\not{p_3} + m_\mu \mathbb{I})_{\alpha\beta}\gamma^\mu_{\beta\delta}(\not{p_4} - m_\mu \mathbb{I})_{\delta\sigma}\gamma^\nu_{\sigma\alpha} \quad (2.60)$$

$$= Tr[(\not{p_3} + m_\mu)\gamma^\mu(\not{p_4} - m_\mu)\gamma^\nu]. \quad (2.61)$$

We can evaluate the traces using γ-matrix identities, such as:

$$Tr[\gamma^\mu\gamma^\nu] = 4g^{\mu\nu}, \qquad Tr[\gamma^\alpha\gamma^\mu\gamma^\beta\gamma^\nu] = 4(g^{\alpha\mu}g^{\beta\nu} - g^{\alpha\beta}g^{\mu\nu} + g^{\alpha\nu}g^{\beta\mu}). \quad (2.62)$$

Multiplying by a factor of $\frac{1}{4}$ and performing the traces, the averaged **squared amplitude** becomes:

$$< |\mathcal{M}|^2 > = \frac{8e^4}{s^2}(p_{13}p_{24} + p_{14}p_{23} + m_\mu^2 p_{12} + m_e^2 p_{34} + 2m_\mu^2 m_e^2), \quad (2.63)$$

where $p_{ij} = p_i \cdot p_j$.

This is the way we obtain the squared amplitude in such processes, it involves multiple steps. At each step, specific mathematical operations and identities are employed to advance the calculation. As we extend to encompass a greater number of final states, the amplitude naturally escalates in complexity, giving rise to a growing number of terms and expressions. Furthermore, when we incorporate higher order (loops), the complication deepens, particularly with integration associated with the loops. This highlights the escalating difficulties inherent in calculations within the domain of high-energy physics.

Chapter 3 Machine Learning

Machine learning (ML) is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform specific tasks without using explicit instructions. Instead, they rely on patterns and inferences derived from data. It has developed significantly since the 1950s, starting with Turing's Test and early algorithms like perceptrons (Rosenblatt, 1958). Despite an initial decline in interest in the 1970s, the 1980s rejuvenated the field with expert systems and backpropagation (Rumelhart, 1986). The 1990s and 2000s saw the rise of data-driven approaches and deep learning. Recent years highlight advancements like GPT [57] and BERT [28] in natural language processing and milestones like AlphaGo [63].

Machine learning is playing increasingly important roles in many aspects of modern technology, spanning from biotechnology to the design of autonomous vehicles and smart devices. Its applications are abundant and varied, including image and speech recognition, medical diagnosis, financial forecasting, content recommendation, and many others. The surge of interest and advancement in machine learning in recent years is due in part to the availability of large datasets, powerful computing resources, and refined algorithms.

There are three main types of machine learning:

1. Supervised Learning: The algorithm is trained on a labeled dataset, meaning the input data is paired with the correct output. The goal is for the model to learn a mapping from inputs to outputs and make predictions on new, unseen data.

2. Unsupervised Learning: The algorithm is provided with an unlabeled dataset

and must find patterns and relationships within the data. Common techniques include clustering and association.

3. Reinforcement Learning: The algorithm learns by interacting with an environment, receiving feedback in the form of rewards or penalties. It aims to find the best strategy, called a policy, to achieve the maximum cumulative reward over time.

In supervised deep learning, the primary goal is to train a model to approximate the function f which maps input features to outputs. Given a dataset of input features $X \in \mathbb{R}^{n \times p}$ where n is the number of samples and p is the number of features, and corresponding outputs $y \in \mathbb{R}^{n \times m}$, the objective is to find an approximation of the function $f\mathbb{R}^p \to \mathbb{R}^m$ such that:

$$f(X_i) \approx y_i, \quad \forall i \in 1, \ldots, n. \tag{3.1}$$

The model f, encapsulated within a deep neural network, is characterized by its parameters θ, making it non-linear and typically high-dimensional. The accuracy of the model's predictions, given by $f_\theta(X)$, against the true outputs y is measured using a loss function

$$\mathcal{L}(y, f_\theta(X)) = \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, f_\theta(X_i)). \tag{3.2}$$

Here, ℓ is a predefined per-sample loss (e.g., mean squared error for regression or cross-entropy for classification). To optimize the model's parameters θ, an optimization algorithm, commonly referred to as an optimizer, is employed. For many deep learning models, gradient descent or its variants are utilized. The optimizer updates the parameters θ in a direction that minimizes the loss L:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(y, f_\theta(X)). \tag{3.3}$$

In the equation above, η is the learning rate, and $\nabla_\theta \mathcal{L}$ represents the gradient of the loss with respect to the model's parameters. The training process involves it-

eratively feeding data, computing the loss, back-propagating the error through the neural network to obtain gradients. Back-propagation involves the systematic adjustment of the network's weights by assessing the impact of each weight on the overall error. Subsequently, these gradients guide the updating of the model parameters using an optimizer. This iterative process continues until reaching convergence or completing a predefined number of epochs. The goal is to have f_θ generalize well to unseen data, minimizing the discrepancy between predictions and true values. The set $\{X_i, y_i\}_{i=1,\dots,n}$ is called the *training set*. In order to test the resulting function f one usually splits the data samples into the *training set* used to learn the function and a *test set* to evaluate the performance.

Commonly used machine learning models and algorithms include:

- Linear regression: This algorithm is used to predict continuous values, based on a linear relationship between different values. For example, the technique could be used to predict house prices based on known features such as the size, location, and neighborhood of the given property.

- Logistic regression: This algorithm is used for binary classification tasks. It predicts the probability of a discrete outcome, usually between two classes, based on one or more independent variables. The output is transformed using the logistic (or sigmoid) function to fall between 0 and 1, making it suitable for probability estimation.

- Clustering: It is an unsupervised learning technique that groups similar data points based on their inherent patterns, without predefined labels. By recognizing subtle differences in data, clustering algorithms can assist data scientists in segmenting information in ways that might not be immediately obvious, revealing hidden structures or relationships within the dataset.

- Decision trees: This can be used for both regression (predicting numerical values) and classification. Decision trees use a branching sequence of linked decisions that can be represented with a tree diagram. One of the advantages of decision trees is that they are easy to audit and validate, unlike the neural network.

- Neural networks: This is the method that stands behind the machine learning revolution of the past decade. It is a computational model inspired by biological brain structures. It consists of interconnected layers of nodes or "neurons" that process and transmit information. Starting with an input layer, data passes through subsequent hidden layers and culminates in an output layer that makes predictions. During training, the network adjusts its internal parameters, known as weights w and biases, to better predict outcomes, enabling it to capture complex data relationships. A concept central to "deep learning" is the multi-layer feed-forward neural networks (FFNN), where there are l-layer fully connected neural networks, so the function $f_w(X_i)$ is parameterized as follows:

$$f_w(X) = g^{(l)}(W^{(l)}...g^{(2)}(W^{(2)}g^{(1)}(W^{(1)}X_i))), \qquad (3.4)$$

where $W^{(l)} \in \mathbb{R}^{k \times m}$ is the weight matrix of the $l-$th layer with k and m the number of neurons in the input and hidden layer, respectively, and g is called the *activation function*. Neural networks are good at recognizing patterns and play an important role in applications including natural language translation, image recognition, speech recognition, and image creation. Examples of deep learning models include: Convolutional Neural Networks (CNN) [52], Recurrent Neural Networks (RNNs) [45], Long Short-Term Memory Networks (LSTMs) [40], Restricted Boltzmann Machines (RBMs) [69], Generative Adversarial Networks (GANs) [36] and Transformers [68].

3.1 Machine Learning in Physics

The significant impact of machine learning in the industrial domain, coupled with its vast technological implementations, has piqued the interest of scientists in its capabilities for fundamental research. It is been used in biology for predicting protein structures - AlphaFold [41]. In chemistry, it is been used for molecular dynamics simulation, analytical chemistry, and computational chemistry [49]. Physics, notably, is also exploring this realm. In fact, physicists are uniquely situated to benefit from and contribute to machine learning. Many of the core concepts and techniques used in machine learning, such as Monte-Carlo methods, and variational methods, originate from physics. Moreover, "energy-based models" inspired by statistical physics form the foundation of many deep learning methods. Hence, modern physicists and astronomers have been leading proponents of harnessing "big data" for their research endeavors. For instance, experiments in LHC colliders such as CMS and ATLAS generate petabytes of data per year. In astronomy, projects such as the Sloan Digital Sky Survey (SDSS) release and analyze hundreds of terabytes of data measuring the properties of nearly a billion stars and galaxies. Researchers in these fields are increasingly incorporating recent advances in machine learning and data science, and this trend is likely to accelerate in the future [2][65]. Other areas of physics, such as quantum computation [42] and condensed matter [8] [16], are increasingly exploring the potential benefits of machine learning in their research.

High energy physics, which investigates the fundamental particles and forces of the universe, has always been at the forefront of technological innovation and data analysis. Machine learning has emerged as a critical tool in managing and making sense of this data avalanche, as it offers a suit of techniques to confront these challenges and a new perspective that motivates bold new strategies. The excitement is widespread, touching both the theoretical and applied aspects of these fields, highlighting not only the current, significant applications but also suggesting deep and transformative

changes in the future. In experimental high-energy physics, physicists use machine learning in a variety of tasks, including:

- Simulation and event generation: Simulating particle interactions is computationally expensive. Several generative machine learning methods are being explored to generate synthetic data that matches real experimental data, potentially reducing the need for exhaustive simulations [14] [59] [47] [50].

- Event and particle classification: Machine learning has been used to quickly classify and filter events, allowing physicists to focus on the most promising data. This is crucial for tasks like identifying rare decay processes or searching for new particles [58] [25] [38].

- Particles identification: The capacity of machine learning algorithms to learn from vast amounts of data and recognize intricate patterns makes them particularly suited for the challenges posed by particle identification in modern experiments [20] [37].

- Anomaly detection: Since machine learning excels at identifying patterns, it has been used to detect anomalies or rare events in the data, which could be indicative of new physics beyond the current understanding [21] [23] [17].

Machine learning is also rapidly finding new uses in theoretical high-energy physics and phenomenology, with the purpose of gaining as much insight into a physical system or process as possible, reducing the need for time, and computational resources. Examples of applying machine learning in these areas include:

- Parton model: Parton Distribution Functions (PDFs) is the probability of finding a certain parton (like a quark or gluon) with a specific momentum fraction inside a proton. These functions are essential in predicting the outcomes of

high-energy collisions involving protons, like those at the Large Hadron Collider (LHC). PDFs are derived by fitting Quantum Chromodynamics (QCD) theoretical models to experimental data, particularly from deep inelastic scattering. Recently, machine learning techniques have been introduced to aid in navigating the complex, high-dimensional parameter spaces of the fits. By leveraging machine learning, researchers can efficiently optimize these fits, potentially uncovering subtleties in the proton's internal structure and refining our understanding of particle interactions [7] [26].

- Lattice QCD: It is a computational approach to studying Quantum Chromodynamics (QCD). Lattice QCD calculations are performed by approximating the QCD path integral by a Monte Carlo sum over gauge field configurations on a discrete four-dimensional space-time. Machine learning has been integrated into Lattice QCD to handle its high dimensional data and complex optimization challenges [61] [32] [43].

- Search for new physics: Machine learning is reshaping the search for new physics models by its adeptness at analyzing complex and vast datasets. Its proficiency in pattern recognition can unveil anomalies suggesting deviations from established physics, while its capability for anomaly detection can spotlight rare events indicative of phenomena beyond the Standard Model [10] [48] [22] [31].

- Constraining Effective Field Theories (EFTs): EFTs provide a framework for describing physics at a given energy scale while integrating out higher energy details, often coming with a plethora of parameters. Machine learning assists in efficiently exploring this vast parameter space, identifying regions consistent with experimental data. By rapidly comparing EFT predictions with observational data, machine learning facilitates the optimization of these parameters and tightens constraints, aiding physicists in narrowing down plausible scenarios

and improving our understanding of fundamental interactions [12] [3] [34].

- Symbolic and numerical calculations: Machine learning offers tools to automate and optimize high-energy calculations. By recognizing patterns and making intelligent predictions, Machine learning algorithms can significantly reduce the computational overhead, making complex symbolic computations more tractable and paving the way for deeper insights into particle interactions and behaviors [4] [27] [13] [54] [64] [5].

- Symbolic regression and models discovery: Symbolic regression is a powerful tool for model discovery and equation formulation. Symbolic regression, which involves finding mathematical expressions that best fit a given dataset, allows physicists to uncover potential underlying equations or relationships in their data. Machine learning can automatically sift through vast spaces of mathematical forms, efficiently pinpointing those that best represent the observed phenomena. This capability not only aids in validating established theories but also paves the way for the potential discovery of new relationships or conservation laws within the realm of high-energy interactions [70] [15] [53] [29].

Chapter 4 Method

4.1 Sequence to Sequence Models

A Sequence-to-sequence (`seq2seq`) model is a machine learning model that is designed for tasks involving mapping an input sequence of symbols to an output sequence. It is a pivotal development that has played a significant role in Natural Language Processing (NLP), which has revolutionized various tasks including machine translation, text summarization, speech recognition, and more. The foundations of seq2seq models can be traced back to the development of recurrent neural networks (RNNs) in the 1980s [45]. RNNs, with their ability to process sequential data by maintaining hidden states, were instrumental in modeling sequences for various tasks, and in 2007, they showed great success in speech recognition tasks [33]. The breakthrough for seq2seq models came with the introduction of the Encoder-Decoder architecture in 2014 [19]. This architecture consisted of two main components: an encoder network that encodes the input sequence into a fixed-length vector representation, and a decoder network that generates the output sequence from this representation. This architecture has revolutionized machine translation with the introduction of the Long Short-Term Memory (LSTM) networks [40]. The Encoder-Decoder architecture was a significant step forward; however, its limitations when dealing with long sequences were addressed with Attention Mechanisms in 2014 [6]. Attention mechanisms allowed the decoder to focus on specific parts of the input sequence while generating the output, significantly improving the model's performance on long sequences and complex tasks. In 2017, Vaswani et al [68] introduced the *transformer* architecture, which further improved the efficiency and parallelization of seq2seq models, making them highly scalable and suitable for large-scale tasks.

4.2 Transformer Model

The transformer model was introduced in 2017 in a paper titled "Attention Is All You Need" by Ashish Vaswani et al. of the Google Brain team [68]. It represents a groundbreaking paradigm shift in the field of natural language processing and machine learning. This architectural innovation revolutionized the way we approach sequence-based tasks by eliminating the need for recurrent neural networks (RNNs) and relying instead on a novel, attention-based mechanism. The Transformer's self-attention mechanism allows it to capture complex dependencies and relationships within sequences, making it highly effective in a wide range of applications, from machine translation and text summarization to question-answering systems, language generation, biological sequence analysis and drug discovery.

As sequence-to-sequence (`seq2seq`) transformer-based model has been applied in many areas of science; the most relevant and interesting ones are the application in symbolic mathematics and physics, which represents a cutting-edge approach, leveraging deep learning to tackle complex computational challenges. These models, which have achieved remarkable success in natural language processing, are now being adapted to understand and generate mathematical expressions, manipulate symbols, and solve mathematical and physical problems. Examples of recent works in this area:

- The transformer model achieves state-of-the-art performance on various symbolic mathematics tasks, including integration, solving differential equations, and simplifying expressions. In the paper 'Deep Learning for Symbolic Mathematics' [51], the authors, Lample and Charton demonstrate that deep learning models can not only match but also surpass the performance of traditional symbolic computation systems. This model was able to predict correctly the symbolic solutions of integrals and found solutions to some problems that could

not be completed using the the standard symbolic mathematics systems within a time limit orders of magnitude longer than the model execution time. Also, pre-trained transformer models achieved excellent results [56] in symbolic calculus problems.

- Transformers models can perform linear algebra numerical calculations with high accuracy [18]. This includes matrix transposition, addition, multiplication, eigenvalues and vectors, singular value decomposition, and inversion.

- Transformer models have also been used to infer the recurrence relation of underlying sequences of numbers [24]. The authors used a transformer model to successfully find symbolic recurrence relations given the first terms of the sequence.

- Transformer models have also been used for symbolic regression, which is the task of identifying a mathematical expression that best fits a provided dataset of input and output values [67].

- In physics, the transformer model has been used to simplify polylogarithmic functions, which appears in loop calculations in high energy physics [27].

4.2.1 Transformer Architecture

The transformer model consists of two parts: *encoder* and *decoder*. The encoder maps an input sequence to a numerical vector in a d-dimension vector space in a process called *embedding*, which allows words with similar meanings to have a similar representation. Then, the model encodes the relative position of each word in the sequence, an operation referred to as *positional encoding* that has the same dimension d. The position *pos* is encoded in a sequence using a combination of *sin* and *cos* functions (alternating these functions for each dimension i in the d-dimensional
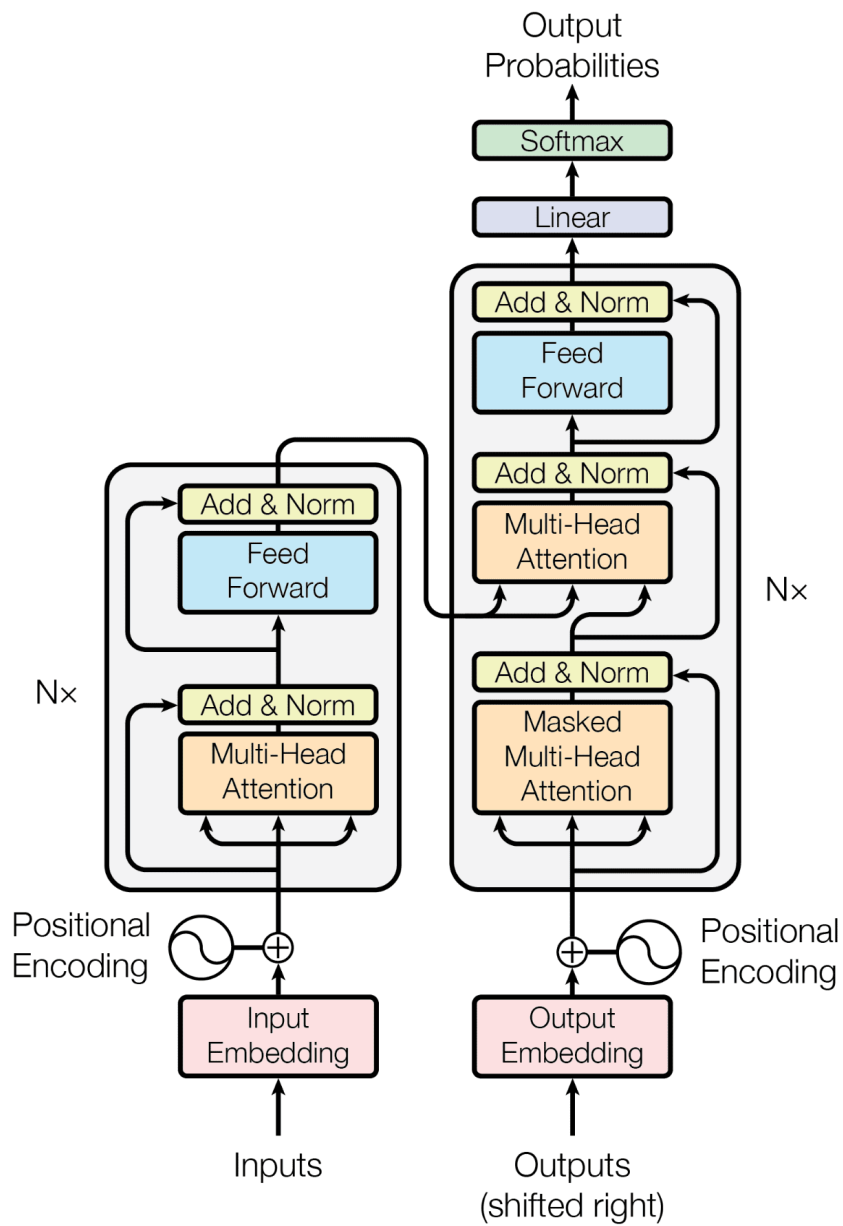
Figure 4.1: Transformer Architecture. (From: "Attention Is All You Need" by Ashish Vaswani et al. [68])

embedding vector) as follows:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right)$$

That is, each dimension of the positional encoding corresponds to a sinusoid. The wavelengths form a geometric progression from 2π to $2\pi \cdot 10000$. The positional encodings have the same dimension d as the embeddings, so that the two can be summed. Next, these vector enter the encoder block, which has two main components: the *multi-head self attention mechanism* and the *position-wise fully connected feed-forward network*. There is also a residual connection around each of the two sub-layers, followed by layer normalization. The multi-head self attention, allows each position in the input sequence to attend to all positions in the input sequence. It uses three vectors for each word: Query (Q), Key (K), and Value (V). The attention score is computed using the formula:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V, \tag{4.1}$$

where d_k is the dimension of the key vectors. The scaling factor $\sqrt{d_k}$ is used to prevent the dot product from growing too large. The model uses multiple sets of Q, K, V to capture information from different representations. Then, the results from different heads are concatenated and linearly transformed:

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O, \tag{4.2}$$

where:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V). \tag{4.3}$$

The projections are parameter matrices: $W_i^Q \in \mathbb{R}^{d \times d_K}$, $W_i^K \in \mathbb{R}^{d \times d_K}$, $W_i^V \in \mathbb{R}^{d \times d_K}$, and $W^O \in \mathbb{R}^{hd_v \times d}$. Next, a fully connected feed-forward network is applied, which consists of two linear transformations with a ReLU activation in between:

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \tag{4.4}$$

While the linear transformations are the same across different positions, they use different parameters from layer to layer.

The decoder is like the encoder, but it has an additional sub-layer to perform multi-head attention over the encoder's output, and it has a *Masked Multi-Head Self Attention Mechanism* to prevent attending to future tokens. During the training of the model, the decoder takes the output vector from the encoder, which encodes information about the input sequence to the encoder, together with the encoded target sequence one token at a time and outputs a sequence also one token at a time.

In this dissertation we describe an in-depth study of the application of transformer model to the symbolic calculation of squared amplitudes. We begin with the data preparations and the choice of hyperparameters. This is followed by the presentation of the results which followed by a discussion and possible future directions.

4.3 Data Generating

As in all supervised machine learning tasks, we need enough data to train the model. We use the symbolic computation program `MARTY` [66] to generate expressions for possible interactions in quantum electrodynamics (QED), the theory of the electromagnetic force, quantum chromodynamics (QCD), the theory of the strong forces, and the electroweak theory (EW). For our proof-of-concept, and due to resource limitations, we restrict the scope to 2-to-2 and 2-to-3 particles at tree-level processes, and 2-to-2 at loop-level processes. All interactions involving off-shell and on-shell particles, anti-particle, and gauge bosons are included. Since it is possible for different amplitudes to yield the same squared expressions, we include such amplitudes in our dataset. An example of expressions of the amplitudes $i\mathcal{M}$ and squared amplitudes $|\mathcal{M}|^2$ of a process of two incoming electrons $e(p_1)$ and $e(p_2)$ scatter into two electrons $e(p_4)$ and $e(p_5)$ and a photon $\gamma(p_3)$ are shown in Fig. 4.2

Figure 4.2: Feynman diagram, amplitude, and squared amplitude expression (without simplification) for the $ee \to ee\gamma$ scattering process produced by `MARTY`

We consider two approaches:

1. **Mapping the input (amplitude or diagram) to the entire squared amplitude expression**

$$\texttt{input} \rightarrow \texttt{output}$$

2. **Mapping the input (amplitude or diagram) to part of the squared amplitude expression**, that is, breaking the squared amplitude into n independent parts, and mapping the input to one part at a time, so the model processes each part independently. By doing this, we increase the number of data elements by a factor of n.

$$\texttt{input+"part\_1"} \rightarrow \texttt{output\_1}$$
$$\texttt{input+"part\_2"} \rightarrow \texttt{output\_2}$$
$$\texttt{input+"part\_3"} \rightarrow \texttt{output\_3}$$
$$....$$
$$\texttt{input+"part\_n"} \rightarrow \texttt{output\_n}$$

The advantage we gain from doing this segmentation is making the mapping easier as the output gets simpler and shorter, which helps improve the performance of the model. In practice, this segmentation would be beneficial if one is interested in certain orders of the squared amplitude, like when putting constraints on the observables in effective field theories (EFTs). Additionally, this way of segmenting the squared amplitude is general and it might be necessary to apply it when we go to more final states or higher order, where the squared amplitude expression becomes excessively long. However, the downside of using this approach is that the data size will be large. More details of this approach are in the next section.

```
Connection on vertex V_2: e(p_4) AntiPart OffShell e(V_2)
OffShell A(V_2) ; Connection on vertex V_1 : e(p_1) A(p_3)
OffShell e(V_1) ; Connection on vertex V_0 : e(p_2) e(p_5)
OffShell A(V_0)|
```

Figure 4.3: Feynman diagram written as a sequence for the $ee \rightarrow ee\gamma$ scattering process produced by MARTY

For each one of these approaches, there are two tasks:

1. **Mapping the amplitude to the squared amplitude**: this task is restricted to tree-level datasets only.

2. **Mapping the Feynman diagram (written as a sequence as shown in Fig 4.3) to the squared amplitude**: this task is applied to all datasets (tree-level and loop).

The reason for the second task is that the Feynman diagram sequence expression has a short and compact form compared to the amplitude which can be very long and complicated, so using the Feynman diagram as input makes the model smaller. All loop datasets are represented in Feynman diagrams only since the amplitude in loop interaction tends to be longer than what our model can take. Another notable advantage of this task is the Feynman diagram can be written by hand, if desired, without the need for a domain-specific tool to construct the amplitude.

4.4 Data Preparation

4.4.1 Simplifying squared amplitude expressions

All squared amplitude expressions are simplified with the `Python` symbolic mathematics module `SymPy` [55]. All squared amplitudes can be written in a general form consisting of a numerator and denominator. The numerator is a polynomial in the particle masses $m_{(p)}$ or in v (the vacuum expectation value) with even powers, and the coefficients are functions of masses of other particles $m_{(p)}$, momenta $p_{(j)}$, v, and the electroweak mixing angle θ_W. The denominator is also a function of masses $m_{(p)}$, momenta $p_{(j)}$, v, and the electroweak mixing angle θ_W. The squared amplitude can be written as follows:

$$Squared\ amplitude = \frac{a_0 m_{(p)}^0 + a_2 m_{(p)}^2 + a_4 m_{(p)}^4 + ... + a_n m_{(p)}^n}{D(m_{(p)}, p_{(j)}, v, \theta_W)}, \qquad (4.5)$$

or:

$$Squared\ amplitude = \frac{a_0 v^0 + a_2 v^2 + a_4 v^4 + ... + a_n v^n}{D(m_{(p)}, p_{(j)}, v, \theta_W)}, \qquad (4.6)$$

where $a_n = f(m_{(p)}, p_{(j)}, v, \theta_W)$, $j = \{1, 2, 3, 4, 5\}$, and $p = \{e, \mu, \tau, u, d, s, t, c, b, h\}$. Knowing the general form of the output makes us able to represent the squared amplitude in different ways that are suitable for machine learning, as it is important to have a short and simple representation. Therefore, for the first task (considering the whole squared amplitude), we put the factors (masses or v) first, then the coefficients, and finally the denominator, as the following:

$$Squared\ amplitude = [\{m_{(p)}^0, m_{(p)}^2, ..., m_{(p)}^n\}], [\{a_0, a_2, ..., a_n\}], [D] \qquad (4.7)$$

This way of writing the squared amplitude has several advantages:

- Structured expressions can facilitate more effective learning by making it easier for the seq2seq model to capture relevant patterns and relationships.

- Making the expression follow one structure makes them more interpretable for humans, enabling easier inspection and understanding of the data.

- In this structure, missing parentheses or operand tokens will not affect the whole expression making it mathematically meaningless. It affects only one term in the expression.

- This structure may help in scalability and extrapolation when one extrapolates to more final-state particles.

Below is an actual example of squared amplitude written in this form for the $ee \rightarrow ee\gamma$ scattering process:

```
([m_e^6, m_e^4, m_e^2, 1], [-2, 2*s_13 + s_14 + s_25 - s_34, -s_12*s_45 - s_1
3*s_25 - s_13*s_34 - s_15*s_24 + s_23*s_45 + s_24*s_35, s_13*(s_23*s_45 + s_2
4*s_35)], [s_13^2*(m_e^4 - 2*m_e^2*s_25 + s_25^2)])
```

where $s_{ij} = p_i \cdot p_j$.

For the second task (breaking the squared amplitude into parts), we split all squared amplitude into n parts based on the highest possible order of the polynomial, because it is possible to know what the highest order of the dataset by knowing the Feynman rules and the number of particles in the process, and consider the denominator as a part too. For instance, the highest order in 2-to-3 QED theory is $m_{(p)}^6$, so, we break the squared amplitude into 5 parts: a part for $m_{(p)}^0$ with its coefficient a_0, a part for $m_{(p)}^2$ with its coefficient a_2, a part for $m_{(p)}^4$ with its coefficient a_4, a part for $m_{(p)}^6$ with its coefficient a_6, and the last one is for the denominator D. If the squared amplitude has no particular order, like no $m_{(p)}^6$ for example, we put "zero".

4.4.2 Tokenization

There are many choices for making the tokenizations, and each choice has advantages and disadvantages. For the squared amplitude, we could tokenize it by symbols (like m_e or p_{12}), by terms (like $4 * m_e^2 * p_{12}.p_{34}$), or by characters. Tokenizing by symbol

or character makes the expression very long, which is not preferable, and tokenizing by term, makes the number of tokens very large and there is a chance of losing the generality. So, we choose to tokenize by each mass, product of momenta, weak mixing angle for EW, and numerical factor (for example, $4 * m_e^2 * p_{12}.p_{34} * \sin \theta_W$ is four tokens) as there are a finite number of momenta products consistent with the physical dimension (in powers of mass) and conservation laws. For squared amplitudes coming from loop interactions, there are additional symbols corresponding to the n-point function integrals that can be evaluated numerically with other tools such as `LoopTools` [39]. The amplitudes are tokenized by operators (tensors) and their indices in Fig. 4.2, while the diagrams are tokenized by particle label (name and momentum) and by vertex number as shown in Fig. 4.3. For practical computational reasons, we exclude expressions longer than 264 tokens after the simplification, which excludes 5%, 26% and 12% of all QED, QCD and EW tree-level expressions, respectively. For the loop interaction, we exclude expressions longer than 500, which excludes 56%, 51% and 24% of all QED, QCD and EW loop expressions. The data are split into three sets: training, validation and test, 70%, 15% and 15%, respectively. Figure 4.4 shows the sequence length for the dataset of the first task (amplitude-squared amplitude). We perform the tokenization, that is, the assignment of an integer to each symbol and the padding of sequences to make them of equal length using `torchtext` [60], so each sequence is then converted to a vector built from these integers.
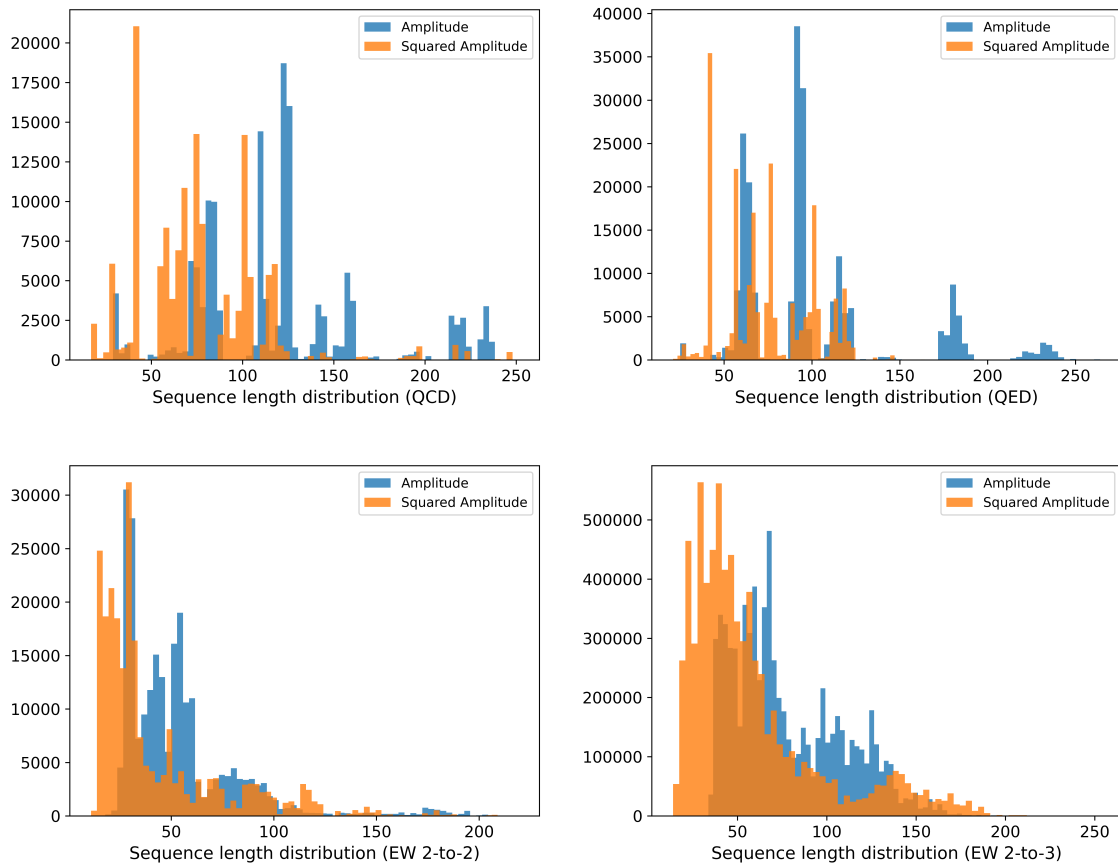
Figure 4.4: Sequence distribution of amplitudes and squared amplitudes

4.4.3 Model and Hyper-parameters:

The transformer model is implemented using `Pytorch 2.0` [60] without structural modifications. The model has $1-2$ layers and 8 attention-heads, with 512 embedding dimensions and 2048 latent dimensions. We use *cross-entropy* as the loss function, the `Adam` optimizer [46] with a learning rate of 10^{-4}. For the batch size, we used 64 for QCD and QED loop, 128 for QED, QCD and EW (2-to-2), and 512 for EW (2-to-3) and EW loop. The training was performed for 30 epochs for EW (2-to-3) due to the fact that we have a limited time for training (24 hours), 100 for QED and QCD at tree-level and loop, and 50 for EW loop. The training was performed on four NVIDIA A100 Tensor Core GPUs, which took about $2-24$ hours.

Dataset	Number of Tokens	Maximum Sequence Length
QED (amplitude) 2-to-2 & 2-to-3	input: 933 output: 1717	input: 264 output: 256
QCD (amplitude) 2-to-2 & 2-to-3	input: 2387 output: 1971	input: 251 output: 250
QED (diagram) 2-to-2 & 2-to-3	input: 90 output: 2469	input: 12 output: 196
QCD (diagram) 2-to-2 & 2-to-3	input: 68 output: 2749	input: 25 output: 255
EW (amplitude) 2-to-2	input: 922 output: 2797	input: 210 output: 195
EW (diagram) 2-to-2	input: 202 output: 3820	input: 19 output: 200
EW (amplitude) 2-to-3	input: 1449 output: 7829	input: 222 output: 253
EW (diagram) 2-to-3	input: 250 output: 7855	input: 28 output: 255
QED (diagram) Loop 2-to-2	input: 238 output: 3384	input: 16 output: 498
QCD (diagram) Loop 2-to-2	input: 152 output: 5479	input: 16 output: 362

Table 4.1: Number of tokens and sequence length for dataset in the first approach

Chapter 5 Results

5.1 Accuracy Metrics

We examined the accuracy of our model by taking a random sample of 500 amplitudes or diagrams (from the test set) that have not been used in the training of the transformer model and predicting their squared amplitudes (or part of the squared amplitudes for the second approach). We used two distinct metrics to measure the accuracy of our model:

1. **Sequence Accuracy:** The percentage of predicted symbolic expressions that *identically* match the correct expression. That is, each predicted token is correct and in the right location.

2. **Token Score:** A measure of the number of tokens (symbols) predicted correctly in the correct location in the sequence. We define a token score measure:

$$Token\ Score = \frac{n_c - n_{ex}}{n_{act}} \times 100\%,$$

where n_c is the number of tokens predicted correctly, n_{ex} is the number of extra tokens that the model predicts (if any), and n_{act} is the number of tokens in the actual sequence. This metric might initially appear impractical, as our ultimate goal is to obtain a correct expression, and a single incorrect token can render the prediction useless. However, this metric is valuable in assessing the model's proximity to the correct answer, which is useful in determining the potential benefits of further experimentation or fine-tuning. Additionally, it provides insights into the form and structure of the output, which can be particularly useful in certain scenarios.

5.2 Model Performance Results

5.2.1 First Approach

Sequence Accuracy:

The results of sequence accuracy for the entire squared amplitude are shown below in the table. 5.1:

Training Sample	process	Training Size	Sequence Accuracy
QED (amplitude)	2-to-2 & 2-to-3	213K	98.6%
QCD (amplitude)	2-to-2 & 2-to-3	205K	97.4%
EW (amplitude)	2-to-2	236K	94.8%
EW (amplitude)	2-to-3	7M	94.4%
QED (diagram)	2-to-2 & 2-to-3	244K	99.0%
QCD (diagram)	2-to-2 & 2-to-3	250K	87.7%
EW (diagram)	2-to-2	258K	93.2%
EW (diagram)	2-to-3	7M	82.3%
QED (diagram)	2-to-2 (Loop)	13K	68.9%
QCD (diagram)	2-to-2 (Loop)	5.5K	60.0%

Table 5.1: Model performance on the first approach - sequence accuracy

Token Score:

The results of the token score for the entire squared amplitude are shown below in the table. 5.2:

Training Sample	process	Training Size	Token Score
QED (amplitude)	2-to-2 & 2-to-3	213K	99.7%
QCD (amplitude)	2-to-2 & 2-to-3	205K	98.7%
EW (amplitude)	2-to-2	236K	96.1%
EW (amplitude)	2-to-3	7M	97.4%
QED (diagram)	2-to-2 & 2-to-3	244K	99.7%
QCD (diagram)	2-to-2 & 2-to-3	250K	90.0%
EW (diagram)	2-to-2	258K	95.8%
EW (diagram)	2-to-3	7M	90.9%
QED (diagram)	2-to-2 (Loop)	13K	80.0%
QCD (diagram)	2-to-2 (Loop)	5.5K	72.3%

Table 5.2: Model performance on the first approach - token accuracy.

5.2.2 Second Approach

Sequence Accuracy:

The results of sequence accuracy for each part of squared amplitude are shown in the tables below:

Training Sample	Size	Sequence Accuracy				
		Part 1	Part 2	Part 3	Part 4	Part 5
QED (amplitude) (2-to-2 & 2-to-3)	1M	100%	99.4%	99.3%	99.4%	99.9%
QED (diagram) (2-to-2 & 2-to-3)	1M	100%	99.5%	99.0%	98.9%	98.6%
QCD (amplitude) (2-to-2 & 2-to-3)	1M	100%	99.5%	98.8%	98.4%	99.4%
QCD (diagram) (2-to-2 & 2-to-3)	1M	96.0%	92.4%	88.0%	90.0%	91.0%

Table 5.3: Model performance on the second approach for QCD and QED - sequence accuracy

Training Sample	Size	Sequence Accuracy					
		Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
EW (amplitude) (2-to-2)	1.4M	99.9%	99.4%	98.4%	98.6%	94.9%	98.7%
EW (diagram) (2-to-2)	1.4M	99.9%	99.7%	99.4%	99.4%	96.7%	99.8%

Table 5.4: Model performance on the second approach for EW (2-to-2) - sequence accuracy

		Sequence Accuracy				
Training Sample	Size	Part 1	Part 2	Part 3	Part 4	Part 5
EW (amplitude) 1st generation (2-to-3)	8M	100%	100%	99.7%	99.6%	95.9%
EW (diagram) 1st generation (2-to-3)	8M	100%	100%	100%	98.8%	95.8%

Table 5.5: Model performance on the second approach for EW (2-to-3) - sequence accuracy (1)

		Sequence Accuracy		
Training Sample	Size	Part 6	Part 7	Part 8
EW (amplitude) 1st generation (2-to-3)	8M	76.8%	75.0%	99.2%
EW (diagram) 1st generation (2-to-3)	8M	76.7%	69.8%	99.0%

Table 5.6: Cont. Model performance on the second approach for EW (2-to-3) - sequence accuracy (2)

		Sequence Accuracy					
Training Sample	Size	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
QED (diagram) Loop (2-to-2)	128K	95.2%	88.5%	84.1%	78.4%	86.9%	97.1%
QCD (diagram) Loop (2-to-2)	40K	91.0%	82.0%	72.0%	71.2%	73.2%	91.5%

Table 5.7: Model performance on the second approach for QCD and QED at loop-level - sequence accuracy

Training Sample	Size	Sequence Accuracy				
		Part 1	Part 2	Part 3	Part 4	Part 5
EW (diagram) Loop (2-to-2)	1.7M	99.0%	100%	98.8%	98.4%	94.3%

Table 5.8: Model performance on the second approach for EW (2-to-2) at loop-level - sequence accuracy (1)

Training Sample	Size	Sequence Accuracy				
		Part 6	Part 7	Part 8	Part 9	Part 10
EW (diagram) Loop (2-to-2)	1.7M	95.9%	90.0%	96.5%	97.9%	85.7%

Table 5.9: Cont. Model performance on the second approach for EW (2-to-2) at loop-level - sequence accuracy (2)

Token Score:

The results of the token score for each part of the squared amplitude are shown below in the tables:

		Token Score				
Training Sample	Size	Part 1	Part 2	Part 3	Part 4	Part 5
QED (amplitude) (2-to-2 & 2-to-3)	1M	100%	99.6%	99.5%	99.5%	99.8%
QED (diagram) (2-to-2 & 2-to-3)	1M	100%	99.9%	99.6%	99.4%	98.9%
QCD (amplitude) (2-to-2 & 2-to-3)	1M	100%	99.8%	99.5%	99.3%	99.6%
QCD (diagram) (2-to-2 & 2-to-3)	1M	97.4%	95.5%	92.1%	92.4%	94.9%

Table 5.10: Model performance on the second approach for QED and QCD - token score

		Token Score					
Training Sample	Size	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
EW (amplitude) (2-to-2)	1.4M	99.9%	99.6%	97.9%	98.2%	97.8%	99.5%
EW (diagram) (2-to-2)	1.4M	99.9%	99.7%	99.4	99.4%	96.7%	99.8%

Table 5.11: Model performance on the second approach for EW (2-to-2) - token score

		Token Score				
Training Sample	Size	Part 1	Part 2	Part 3	Part 4	Part 5
EW (amplitude) 1st generation (2-to-3)	8M	100%	100%	99.9%	99.7%	97.2%
EW (diagram) 1st generation (2-to-3)	8M	100%	100%	100%	99.4%	97.4%

Table 5.12: Model performance on the second approach for EW (2-to-3) - token score (1)

		Token Score		
Training Sample	Size	Part 6	Part 7	Part 8
EW (amplitude) 1st generation (2-to-3)	8M	87.5%	86.3%	99.8%
EW (diagram) 1st generation (2-to-3)	8M	87.2%	82.4%	99.7%

Table 5.13: Cont. Model performance on the second approach for EW (2-to-3) - token score (2)

		Token Score					
Training Sample	Size	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
QED (diagram) Loop (2-to-2)	128K	96.3%	90.0%	90.0%	86.3%	93.4%	98.6%
QCD (diagram) Loop (2-to-2)	40K	93.5%	88.8%	79.8%	78.8%	81.5%	96.4%

Table 5.14: Model performance on the second approach for QED and QCD at loop-level - token score

Training Sample	Size	Token Score				
		Part 1	Part 2	Part 3	Part 4	Part 5
EW (diagram) Loop (2-to-2)	1.7M	99.8%	100%	99.2%	98.5%	96.7%

Table 5.15: Model performance on the second approach for EW at loop-level - token score (1)

Training Sample	Size	Token Score				
		Part 6	Part 7	Part 8	Part 9	Part 10
EW (diagram) Loop (2-to-2)	1.7M	97.3%	94.0	97.6%	98.8%	93.3%

Table 5.16: Cont. Model performance on the second approach for EW at loop-level - token score (2)

Table 5.1 and 5.2 summarize the performance of our models for the first approach. The model achieves a sequence accuracy **between 94.4% to 98.6%** for the first task (mapping the amplitudes) at tree-level, and **between 83.2% to 99.0%** for the second task (mapping the diagrams) at tree-level. For the token score, the model achieves **between 97.4% to 99.7%** for the first task at tree-level, and **between 90% to 99.7%** for the second task. For the loop-level dataset, the sequence accuracy is **between 60% to 68%** and a token score **between 72.3% to 80.0%**. Tables 5.3, 5.4, 5.5, and 5.6 summarize the performance of our models for the second approach (mapping to parts of the squared amplitude), the model achieves a sequence accuracy of **almost 100%** for the first three parts and the last part (the denominator) at tree-level (except the QCD diagram). The plots in Fig. 5.3 provide a comprehensive overview of the results for the second approach.
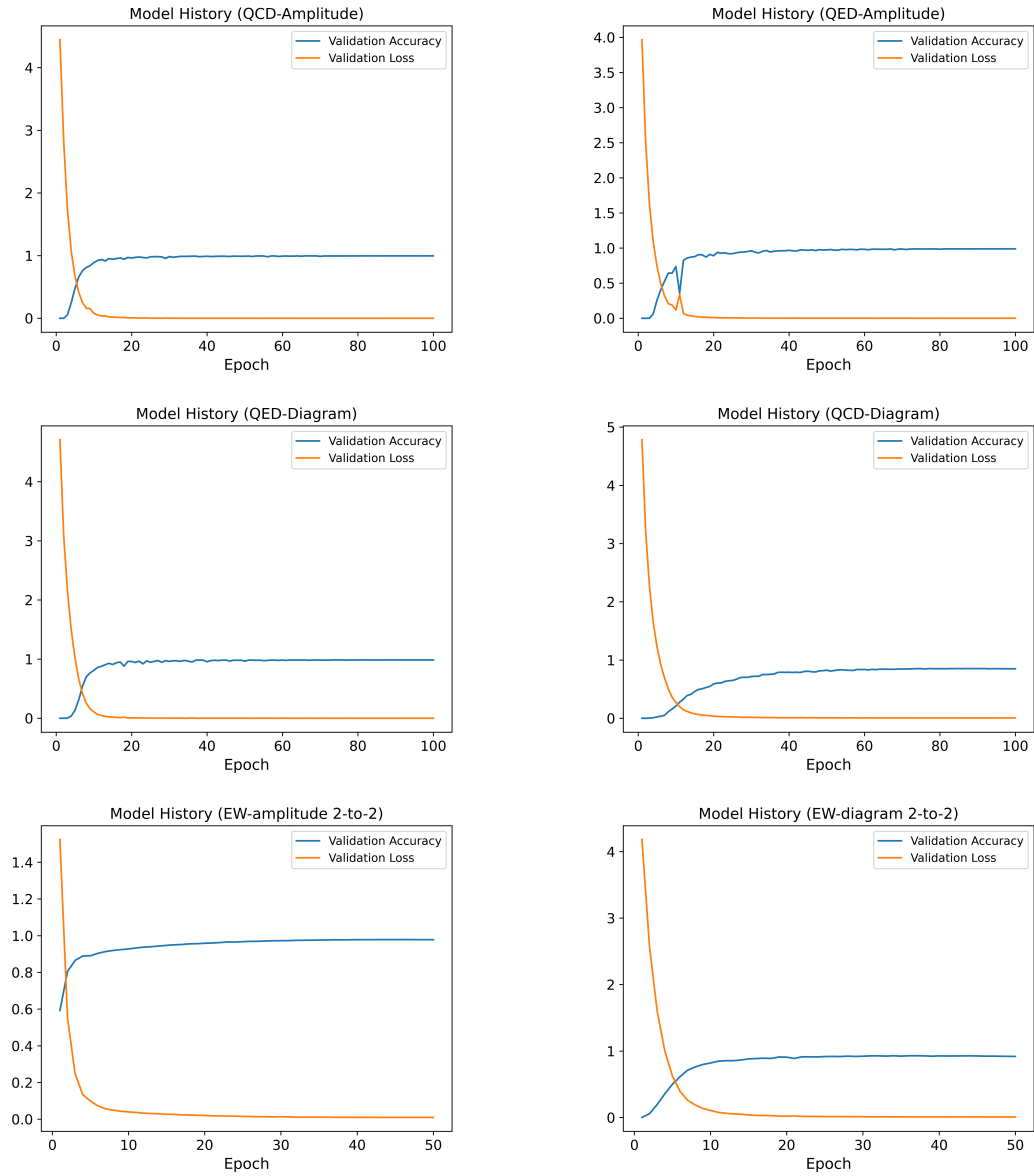
Figure 5.1: A visualization of the validation loss and accuracy during training
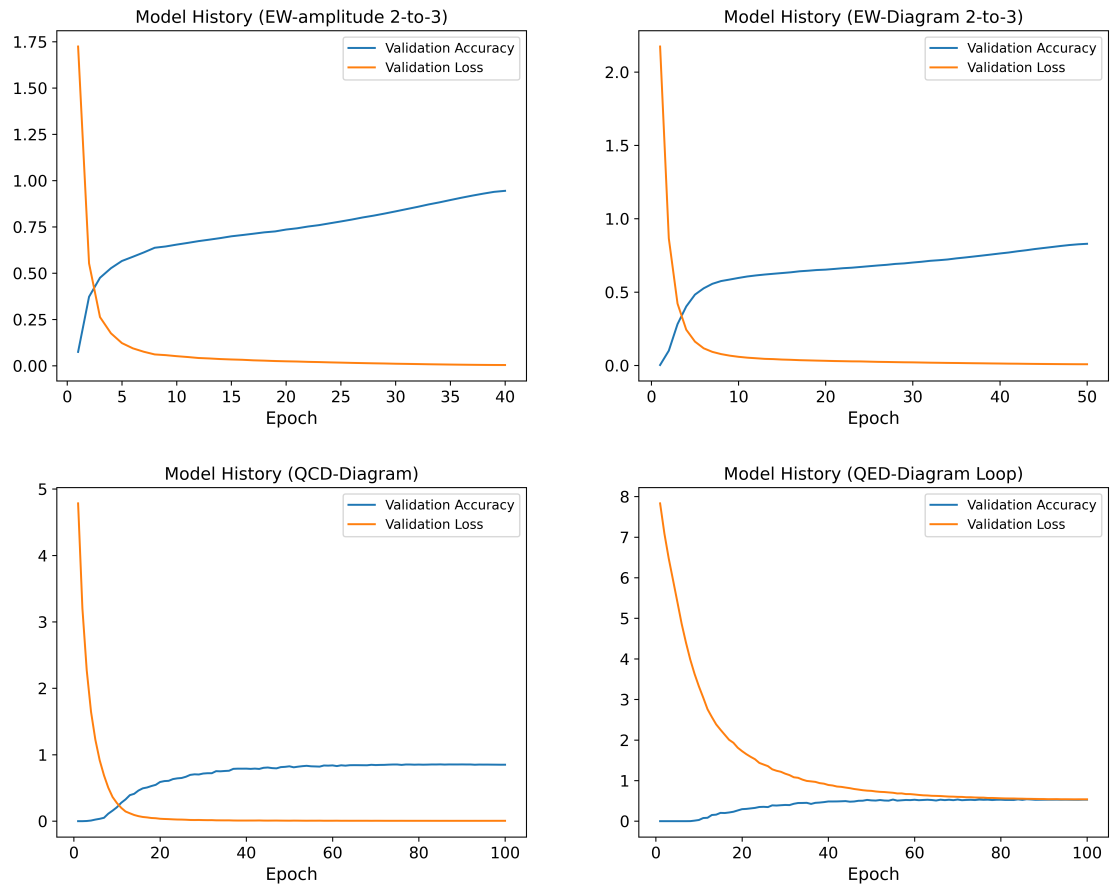
Figure 5.2: A visualization of the validation loss and accuracy during training
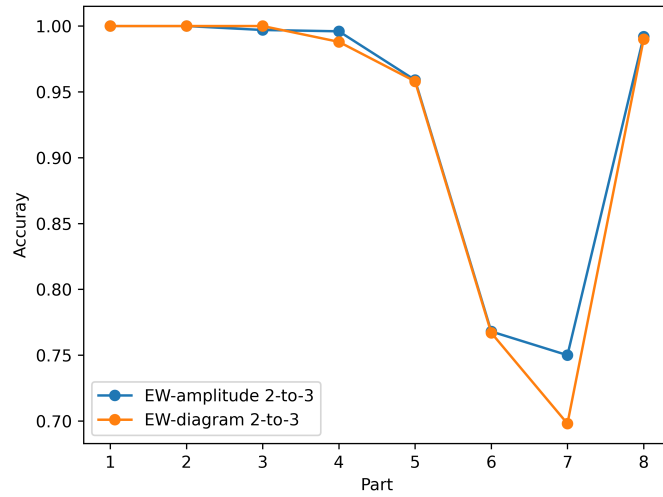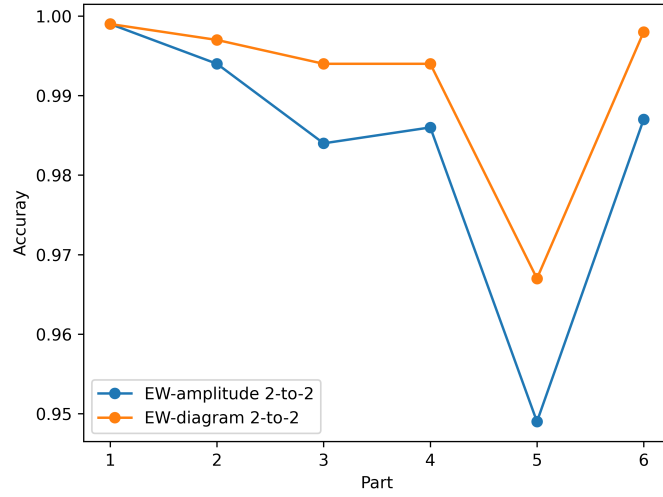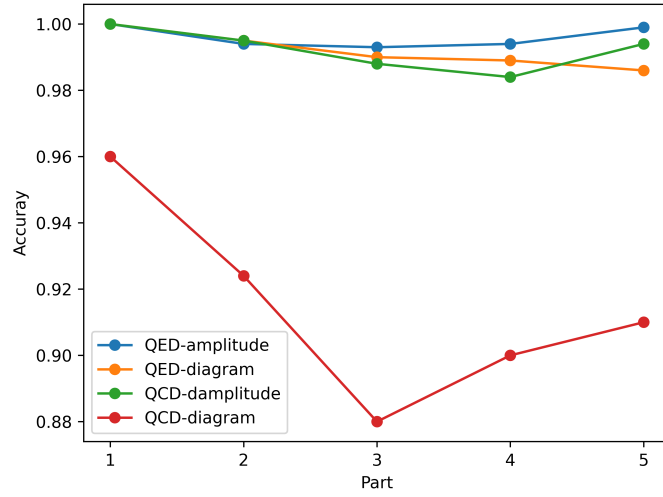
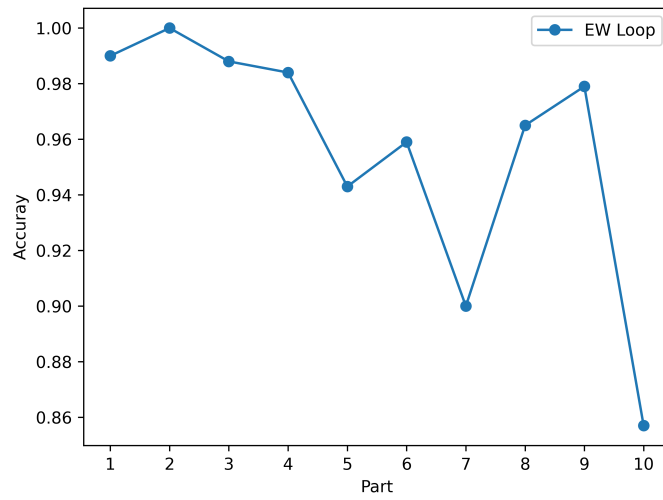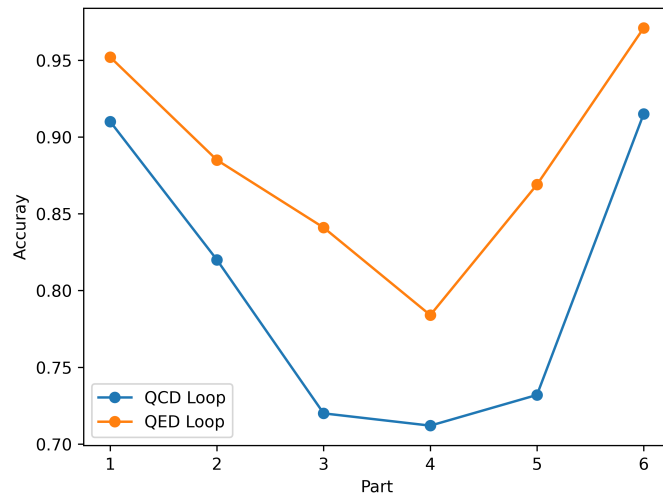Figure 5.3: Sequence accuracy results for each part in the second approach

Figure 5.4: Sequence accuracy results for each part in the second approach

Chapter 6 Discussion

6.1   Overview

The results in the previous chapter provides a detailed account of our work, setting the stage for a thorough and thoughtful discussion. It has been demonstrated in the previous chapter that the model can "do the algebra" for such a mathematical problem as complicated as squaring an amplitude, averaging and summing over the internal degrees of freedom of particles, and manipulating the whole result into a meaningful form. The time for inference, on average, is less than one second, which is similar to the time taken by `MARTY` for most of the dataset, but in some data in EW (2-to-3) and QCD, the model is **up to 6 orders of magnitudes faster** than `MARTY` for the same calculations. This proves that the scalability of our model is much better than `MARTY`. The accuracy of the sequence varies within specific ranges for different tasks in our analysis. For the first task, which involves mapping amplitudes at the tree-level, the sequence accuracy falls between 94.4% and 98.6%. In contrast, for the second task, where we map diagrams also at the tree-level, the accuracy ranges from 83.2% to 99.0%. This discrepancy highlights that the second task exhibits lower accuracy compared to the first one, possibly due to a lack of sufficient information in the input data. In particular, the difference in accuracy in quantum chromodynamics theory (QCD) can be attributed to the inherent complexity of the theory itself, which involves color factors and gluon self-interactions, elements not present in quantum electrodynamics (QED), for example. As a result, expressing the input data in the form of Feynman diagrams alone may not be adequate in QCD, primarily because the number of input tokens in Feynman diagrams is significantly smaller. A similar pattern is observed in the case of electroweak interactions (EW) for 2-to-3 processes, where accuracy is lower when compared to models that rely on amplitude

sequence information. However, in the case of QED and EW 2-to-2, the dataset can be effectively represented by Feynman diagrams, and this representation seems to capture the complexity adequately. For the loop-level dataset, the sequence accuracy ranges from 60% to 68%, and the token score falls between 72.3% and 80.0%. This lower accuracy is expected due to the limited size of the dataset and the fact that the input consists solely of Feynman diagrams, which may not fully encompass the intricacies of the mapping process.

In the second approach, which involves mapping to parts of the squared amplitude, the model achieves a sequence accuracy of nearly 100% for the first three parts and the last part (the denominator) at the tree-level, except for the QCD diagram. In the case of EW at 2-to-3 processes, the accuracy drops to 70% for some parts beyond the third part. This decrease in accuracy is anticipated because we trained the model using particles from the 1st generation exclusively, due to resource limitations. We anticipate improved performance when incorporating more particles for the 2nd and 3rd generations.

In regard to the model's ability to extrapolate, specifically, its capability to predict squared amplitude for processes involving a greater number of final state particles without comprehensive prior training, we have found some promising evidence. We trained the model on electroweak (EW) 2-to-2 data (285K), supplemented by an additional sample of 5K from EW 2-to-3 data. When testing this model on amplitudes from EW 2-to-3, it achieved a sequence accuracy of 32.4% (compared to a sequence accuracy of 2.4% when training only on those 5K). This indicates that including a small sample from data with more final state particles, along with the current data, enables the model to combine knowledge from both and more accurately predict outcomes involving more final state particles. This also implies a reduced need for extensive data generation as we go to more final-state particles. While this current result might seem low, it suggests significant potential for improvement and development. An im-

provement could be accomplished in the data representation, such as representing the squared amplitude in a format that is more suitable for extrapolations. Furthermore, adapting the model to accommodate additional terms arising from extrapolating to more final-state particles could be beneficial. This would involve altering the mass dimensions and modifying the loss function and inference process accordingly.

6.2   Remarks:

In the remarks that follow, we aim to extract key insights, address the implications of our work, and highlight potential directions for future research that arise from our study:

- We observed that the model's accuracy is influenced by three key factors: data complexity, data size, and sequence length. While quantifying data complexity directly may be challenging, we can draw insights from several indicators, including the total number of tokens, the presence of longer sequences in both input and output and the distribution of data sequence lengths. These factors, coupled with our understanding of the inherent complexity of the underlying theory, as previously mentioned, can collectively shed light on the issue of complexity. In loop dataset, all of these three factors manifest and contribute to the lower accuracy observed. As there is a strong dependence on data set size, we expect that employing a larger training dataset will likely yield improved performance. There are several ways to address the issues of data complexity including adding more details about the interaction, so the input data should encompass a comprehensive range of features, taking into consideration all complexities of interactions. Notably, considering the amplitude as an input, as opposed to the diagram, appears to have a positive impact, particularly in the case of QCD. Regarding data size and sequence length, we present empirical evidence in Table 6.1 and Table 6.2 demonstrating the positive effects of in-

71

QCD	Train on:	⅕ data	⅓ data	½ data	full data (235K)
	Sequence Accuracy:	82.0%	91.0%	94.0%	97.4%

QED	Train on:	⅕ data	⅓ data	½ data	full data (251K)
	Sequence Accuracy:	89.6%	95.8%	98.0%	98.6%

Table 6.1: Model performance on different sizes of QCD and QED dataset

QCD	Max. sequence length:	170 tokens	195 tokens	full length
	Sample size:	126K	137K	235K
	Sequence Accuracy:	99.6%	98.6%	97.4%

QED	Maximum sequence length:	170 tokens	185 tokens	full length
	Sample size:	218K	237K	251K
	Sequence Accuracy:	99.0%	98.8%	98.6%

Table 6.2: Model performance on different sequence lengths of QCD and QED dataset

creasing data size and reducing sequence length on model accuracy. However, challenges arise in datasets like QED and QCD loop datasets, where both data size and data length pose issues. The data size issue can be solved by adding more processes from theories beyond the Standard Model (BSM) which exposes the model to a greater variety of examples, potentially enhancing its ability to handle variations in data size. For addressing longer sequence lengths, specialized transformer model variants that exhibit better scalability with respect to sequence length, as demonstrated in [9], could be explored. However, we leave these considerations for future research. Additionally, fine-tuning the sequence length can be accomplished through adjustments to the tokenization process.

- The effect of representing the squared amplitude in a form that is concise, unified, and easily tokenizable is important for efficient and optimal performance. As an example, if we were to refrain from the simplification of the squared amplitude, as detailed in Section 4.4.1 and instead directly adopt it in its original

form from the source, `MARTY`, the resulting model becomes excessively large, leading to a decline in performance. Similarly, it is advisable to maintain uniformity in the representation of the amplitude, at least up to a level where it adequately encapsulates the underlying complexity.

- One of the most important issues encountered in the project is data imbalance, a common issue in many machine learning, which means that the distribution of classes or categories is not roughly equal or balanced. In other words, some classes or categories have significantly more examples (data points) than others, leading to an unequal representation of different classes in the dataset. In the context of this project, the imbalance manifests in the varying frequencies of expression forms of the squared amplitude expressions, like the number of orders illustrated in 4.7, which means some forms appear more often than others. Consequently, some expression forms occur more frequently than others. This inherent issue spans across all the theories under consideration, primarily due to the fact not every interaction is physically allowed, resulting in the prevalence of certain interactions with similar (albeit not exact) forms over others. Addressing data imbalance is a critical consideration in machine learning, and existing literature offers several strategies to mitigate this issue, including resampling, synthetic data generation, and data augmentation. However, these approaches may either be irrelevant or accompanied by disadvantages in the specific context of this project. An alternative solution that shows promise involves expanding the scope to encompass additional theories or interactions that are mathematically correct, even if they may not align with physical reality. By incorporating these mathematically valid but potentially non-physical interactions into the training dataset, we can expose the model to a broader array of examples that share similar mathematical characteristics. This approach has the potential to help mitigate the challenges posed by data imbalance and enhance the model's
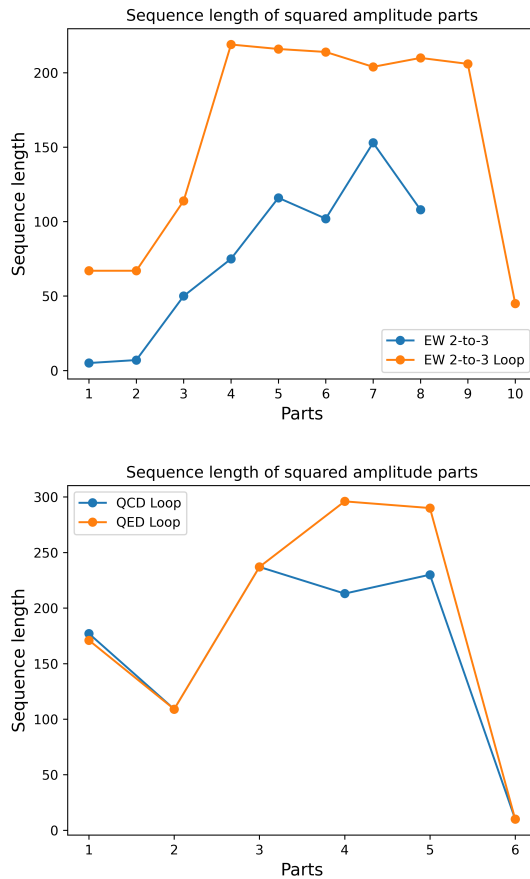
Figure 6.1: Sequence length of squared amplitude parts

performance.

- The decline in accuracy in parts beyond the third one in EW 2-to-3 and loop
  datasets can be primarily attributed to the increasing sequence length. As the
  polynomial order decreases, the sequence tends to become longer. Figure 6.1
  illustrates the maximum sequence length for each part. One potential solution
  is to customize the loss function to place greater emphasis on these specific
  parts during training.

In summary, these remarks highlight crucial factors influencing model accuracy, including data complexity, size, and sequence length. We emphasize the importance of representing the squared amplitude in a concise and uniform manner for efficiency and

optimal performance. Data imbalance poses a significant challenge, and we explore the inclusion of mathematically valid but non-physical interactions as a potential solution. Additionally, we pinpoint the accuracy decline in certain parts due to increasing sequence lengths and suggest customizing the loss function to address this issue during training. These insights inform strategies to enhance the model's performance and robustness in symbolic mathematics applications.

6.3 Future Works

Having explored the applications of seq2seq models in symbolic calculations of squared amplitude within the realm of high-energy physics, we have made significant strides in enhancing our understanding of this complex domain. While our current research has yielded valuable insights, several challenges and opportunities remain unexplored. These areas represent fertile ground for future investigations. To continue pushing the boundaries of high-energy physics and symbolic mathematics, it is imperative that we embark on a journey of further exploration. Our work has laid the foundation, and the path ahead is filled with exciting possibilities. In light of these considerations, we turn our attention to potential avenues for future research. The following sections will outline suggestions and strategies for advancing our understanding of symbolic amplitude calculations in high-energy physics. Examples of future research directions include:

Enhancing model performance: As we aim for a higher accuracy for the entire prediction of the output, there are many directions one can explore:

- Exploring transformer variants: One of the foremost limitations lies in sequence length; thus, exploring advanced transformer model variants that exhibit better scalability with respect to sequence length would be of great importance as one goes to final states with more particles or higher order the sequence length becomes overly large. By identifying or developing transformer variants that effectively handle these longer sequences, researchers can significantly enhance the model's performance and applicability.

- Including more data: The significance of data in achieving optimal performance cannot be overstated. Researchers can explore various strategies to generate additional data and examples to enrich the training dataset. This could involve using theories beyond the Standard Model, or including additional theories or

interactions that are mathematically correct, even if they may not align with physical reality.

- Incorporate physical constraints: Many physical constraints within this task have the potential to assist the model during both the training and inference phases. This can be accomplished by developing a custom loss function that directly incorporates the relevant physical principles.

- Enhancing sequence accuracy: While achieving higher accuracy for the entire output is a primary objective, it's crucial to develop methods and techniques that ensure such accuracy consistently across various tasks and expressions. Researchers can explore strategies such as multi-task learning or fine-tuning approaches that specifically target sequence accuracy improvement.

- Improving data representation: The data representation is important in the context of symbolic mathematics. Given the various ways expressions can be written, it is essential to explore and refine data preprocessing techniques. Researchers can investigate methods for standardizing and optimizing data representation to ensure consistency and effectiveness in model training and inference.

Expanding the model: As our ultimate goal is to create a model that can be practically employed for discovery, prediction, and reliability, requiring robust generalization and effective extrapolation to accommodate a broader range of final-state particles and orders, researchers have a multitude of avenues to explore, including:

- Integrating with the domain-specific programs: Combining machine learning with traditional methods and taking advantage of both holds tremendous potential. By integrating machine learning models with domain-specific programs and techniques, researchers can leverage the speed and scalability advantages

77

offered by machine learning, while still benefiting from the accuracy and precision inherent in traditional methods. A strategic approach involves a thorough analysis of the domain-specific program's algorithms to pinpoint bottlenecks within the amplitude calculation process. Machine learning can then be applied at these specific junctures, such as during simplification, expansion, or other transformations. This collaborative synergy holds the potential to propel symbolic amplitude calculations to new heights of advancement.

- Multitasking model: When calculating the squared amplitude, numerous complicated mathematical operations come into play, encompassing tasks like amplitude simplification, Lorentz index contractions, color factor calculations, matrix multiplications, Dirac algebra, and trace evaluations. It's conceivable to generate examples for each of these operations and train the model to execute them proficiently. By incorporating these auxiliary tasks alongside the primary objective of squaring the amplitude, there is potential for significant performance and robustness improvements.

- Utilizing pre-trained models: The use of pre-trained models has emerged as a state-of-the-art approach in various machine learning domains. Researchers can explore the adaptation of pre-trained models to symbolic amplitude calculations. Employing strategies such as fine-tuning and transfer learning, researchers can exploit the wealth of knowledge encapsulated within these pre-trained models. An intriguing avenue involves pre-training a multi-tasking model specifically tailored to mathematical operations in high-energy physics. This deliberate approach holds the promise of yielding substantial performance improvements.

- Addressing the uncertainty: Developing methods for error analysis and uncertainty quantification within symbolic calculations is essential for robust and

reliable results. Researchers can explore techniques such as uncertainty esti-
mation and probabilistic modeling to gain insights into the model's confidence
levels and identify potential sources of error. Ensemble methods, which com-
bine predictions from multiple models with different architectures, can also be
a valuable tool for addressing uncertainty.

These future research opportunities represent diverse avenues for advancing the field
of symbolic amplitude calculations in high-energy physics. Researchers can choose to
explore one or more of these directions based on their specific goals, resources, and
the evolving needs of the scientific community. Each avenue promises to contribute
to the ongoing progress in this fascinating and interdisciplinary field.

6.4 Conclusion

In conclusion, we show the remarkable capacity of a symbolic deep learning model to efficiently learn the mapping between particle interaction amplitudes and their squares with a high level of accuracy. This novel approach significantly outpaces traditional methods in terms of computational speed, despite the inherent complexity of the mapping process. Our findings highlight the complex interplay of factors influencing accuracy, with data size and sequence length emerging as key influencers. While acknowledging the model's limitations, the promising results obtained serve as a catalyst for future endeavors aimed at further enhancing performance.

Bibliography

[1] Atlas collaboration `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/`
     `CombinedSummaryPlots/SM/ATLAS_b_SMSummary_FiducialXsect/history.`
     `html`.

[2] K. Albertsson et al. Machine Learning in High Energy Physics Community
     White Paper. *J. Phys. Conf. Ser.*, 1085(2):022008, 2018.

[3] J. Alda, J. Guasch, and S. Penaranda. Using Machine Learning techniques in
     phenomenological studies on flavour physics. *JHEP*, 07:115, 2022.

[4] A. Alnuqaydan, S. Gleyzer, and H. Prosper. Symba: symbolic computation
     of squared amplitudes in high energy physics with machine learning. *Machine
     Learning: Science and Technology*, 4(1):015007, jan 2023.

[5] S. Badger, A. Butter, M. Luchmann, S. Pitz, and T. Plehn. Loop amplitudes
     from precision networks. *SciPost Phys. Core*, 6:034, 2023.

[6] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly
     learning to align and translate, 2016.

[7] R. D. Ball, S. Carrazza, J. Cruz-Martinez, L. D. Debbio, S. Forte, T. Giani,
     S. Iranipour, Z. Kassabov, J. I. Latorre, E. R. Nocera, R. L. Pearson, J. Rojo,
     R. Stegeman, C. Schwan, M. Ubiali, C. Voisey, and M. Wilson. An open-source
     machine learning framework for global analyses of parton distributions, 2021.

[8] E. Bedolla, L. C. Padierna, and R. Castañeda-Priego. Machine learning for
     condensed matter physics. *Journal of Physics: Condensed Matter*, 33(5):053001,
     nov 2020.

[9] I. Beltagy, M. E. Peters, and A. Cohan. Longformer: The long-document transformer, 2020.

[10] G. Bertone, M. P. Deisenroth, J. S. Kim, S. Liem, R. Ruiz de Austri, and M. Welling. Accelerating the BSM interpretation of LHC data with machine learning. *Phys. Dark Univ.*, 24:100293, 2019.

[11] E. Boos, V. Bunichev, M. Dubinin, L. Dudko, V. Ilyin, A. Kryukov, V. Edneral, V. Savrin, A. Semenov, and A. Sherstnev. CompHEP 4.4: Automatic computations from Lagrangians to events. *Nucl. Instrum. Meth. A*, 534:250–259, 2004.

[12] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez. Constraining effective field theories with machine learning. *Physical Review Letters*, 121(11), sep 2018.

[13] A. Buckley, A. Kvellestad, A. Raklev, P. Scott, J. V. Sparre, J. V. den Abeele, and I. A. Vazquez-Holm. Xsec: the cross-section evaluation code. *The European Physical Journal C*, 80(12), dec 2020.

[14] A. Butter, T. Plehn, S. Schumann, S. Badger, S. Caron, K. Cranmer, F. A. D. Bello, E. Dreyer, S. Forte, S. Ganguly, D. Gonçalves, E. Gross, T. Heimel, G. Heinrich, L. Heinrich, A. Held, S. Höche, J. N. Howard, P. Ilten, J. Isaacson, T. Janßen, S. Jones, M. Kado, M. Kagan, G. Kasieczka, F. Kling, S. Kraml, C. Krause, F. Krauss, K. Kröninger, R. K. Barman, M. Luchmann, V. Magerya, D. Maitre, B. Malaescu, F. Maltoni, T. Martini, O. Mattelaer, B. Nachman, S. Pitz, J. Rojo, M. Schwartz, D. Shih, F. Siegert, R. Stegeman, B. Stienen, J. Thaler, R. Verheyen, D. Whiteson, R. Winterhalder, and J. Zupan. Machine learning and LHC event generation. *SciPost Physics*, 14(4), apr 2023.

[15] A. Butter, T. Plehn, N. Soybelman, and J. Brehmer. Back to the Formula – LHC Edition. 9 2021.

[16] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, feb 2017.

[17] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant. Variational autoencoders for new physics mining at the large hadron collider. *Journal of High Energy Physics*, 2019(5), may 2019.

[18] F. Charton. Linear algebra with transformers, 2022.

[19] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.

[20] J. Collado, J. N. Howard, T. Faucett, T. Tong, P. Baldi, and D. Whiteson. Learning to identify electrons. *Physical Review D*, 103(11), jun 2021.

[21] J. Collins, K. Howe, and B. Nachman. Anomaly detection for resonant new physics with machine learning. *Physical Review Letters*, 121(24), dec 2018.

[22] R. T. d'Agnolo, G. Grosso, M. Pierini, A. Wulzer, and M. Zanetti. Learning new physics from an imperfect machine. *Eur. Phys. J. C*, 82(3):275, 2022.

[23] R. T. D'Agnolo and A. Wulzer. Learning new physics from a machine. *Physical Review D*, 99(1), jan 2019.

[24] S. d'Ascoli, P.-A. Kamienny, G. Lample, and F. Charton. Deep symbolic regression for recurrent sequences, 2022.

[25] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman. Jet-images — deep learning edition. *Journal of High Energy Physics*, 2016(7), jul 2016.

[26] L. D. Debbio, T. Giani, J. Karpie, K. Orginos, A. Radyushkin, and S. Zafeiropoulos. Neural-network analysis of parton distribution functions from ioffe-time pseudodistributions. *Journal of High Energy Physics*, 2021(2), feb 2021.

[27] A. Dersy, M. D. Schwartz, and X. Zhang. Simplifying Polylogarithms with Machine Learning. 6 2022.

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[29] Z. Dong, K. Kong, K. T. Matchev, and K. Matcheva. Is the machine smarter than the theorist: Deriving formulas for particle kinematics with symbolic regression. *Physical Review D*, 107(5), mar 2023.

[30] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos. A historical profile of the higgs boson. In *Advanced Series on Directions in High Energy Physics*, pages 255–274. WORLD SCIENTIFIC, aug 2016.

[31] M. Farina, Y. Nakai, and D. Shih. Searching for new physics with deep autoencoders. *Physical Review D*, 101(7), apr 2020.

[32] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh. Lattice gauge symmetry in neural networks, 2021.

[33] S. Fernández, A. Graves, and J. Schmidhuber. An application of recurrent neural networks to discriminative keyword spotting. In *Proceedings of the 17th International Conference on Artificial Neural Networks*, ICANN'07, page 220–229, Berlin, Heidelberg, 2007. Springer-Verlag.

[34] F. F. Freitas, C. K. Khosa, and V. Sanz. Exploring the standard model eft in *vh* production with machine learning. *Phys. Rev. D*, 100:035040, Aug 2019.

[35] D. Goldberg. *The standard model in a Nutshell.* Princeton University Press, 2017.

[36] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.

[37] L. K. Graczykowski, M. Jakubowska, K. R. Deja, and M. Kabus. Using machine learning for particle identification in ALICE. *JINST*, 17(07):C07016, 2022.

[38] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson. Jet flavor classification in high-energy physics with deep neural networks. *Physical Review D*, 94(11), dec 2016.

[39] T. Hahn and M. Pérez-Victoria. Automated one-loop calculations in four and d dimensions. *Computer Physics Communications*, 118(2):153–165, 1999.

[40] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[41] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[42] S. S. Kalantre, J. P. Zwolak, S. Ragole, X. Wu, N. M. Zimmerman, M. D. Stewart, and J. M. Taylor. Machine learning techniques for state recognition and auto-tuning in quantum dots. *npj Quantum Information*, 5(1), jan 2019.

[43] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan. Equivariant flow-based sampling for lattice gauge theory. *Physical Review Letters*, 125(12), sep 2020.

[44] S. G. Karshenboim. Precision physics of simple atoms: QED tests, nuclear structure and fundamental constants. *Physics Reports*, 422(1-2):1–63, dec 2005.

[45] M. Kaur and A. Mohta. A review of deep learning with recurrent neural network. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 460–465, 2019.

[46] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014.

[47] C. Krause and D. Shih. Fast and accurate simulations of calorimeter showers with normalizing flows. *Physical Review D*, 107(11), jun 2023.

[48] B. Kronheim, M. Kuchera, H. Prosper, and A. Karbo. Bayesian neural networks for fast SUSY predictions. *Physics Letters B*, 813:136041, feb 2021.

[49] D. Kuntz and A. K. Wilson. Machine learning, artificial intelligence, and chemistry: How smart algorithms are reshaping simulation and the laboratory. *Pure and Applied Chemistry*, 94(8):1019–1054, 2022.

[50] B. Käch, D. Krücker, I. Melzer-Pellmann, M. Scham, S. Schnake, and A. Verney-Provatas. Jetflow: Generating jets with conditioned and mass constrained normalising flows, 2022.

[51] G. Lample and F. Charton. Deep learning for symbolic mathematics, 2019.

[52] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019, 2022.

[53] Y. Lu, Y.-J. Wang, Y. Chen, and J.-J. Wu. Rediscovery of numerical lüscher's formula from the neural network, 2022.

[54] D. Maître and H. Truong. A factorisation-aware matrix element emulator. *Journal of High Energy Physics*, 2021(11), nov 2021.

[55] A. Meurer et al. Sympy: symbolic computing in python, Jan. 2017.

[56] K. Noorbakhsh, M. Sulaiman, M. Sharifi, K. Roy, and P. Jamshidi. Pretrained language models are symbolic mathematics solvers too!, 2023.

[57] OpenAI. Gpt-4 technical report, 2023.

[58] M. Paganini, L. de Oliveira, and B. Nachman. Survey of machine learning techniques for high energy electromagnetic shower classification. In *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA*, 2017.

[59] M. Paganini, L. de Oliveira, and B. Nachman. Accelerating science with generative adversarial networks: An application to 3d particle showers in multilayer calorimeters. *Physical Review Letters*, 120(4), jan 2018.

[60] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[61] P. E. Shanahan, D. Trewartha, and W. Detmold. Machine learning action parameters in lattice quantum chromodynamics. *Physical Review D*, 97(9), may 2018.

[62] V. Shtabovenko, R. Mertig, and F. Orellana. Feyncalc 9.3: New features and improvements. *Computer Physics Communications*, 256:107478, 2020.

[63] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[64] D. L. B. Sombillo, Y. Ikeda, T. Sato, and A. Hosaka. Unveiling the pole structure of s-matrix using deep learning, 2021.

[65] Y. Su et al. A deep learning view of the census of galaxy clusters in IllustrisTNG. *Mon. Not. Roy. Astron. Soc.*, 498(4):5620–5628, 2020.

[66] G. Uhlrich, F. Mahmoudi, and A. Arbey. – modern artificial theoretical physicist a c++ framework automating theoretical calculations beyond the standard model. *Computer Physics Communications*, 264:107928, 2021.

[67] M. Valipour, B. You, M. Panju, and A. Ghodsi. Symbolicgpt: A generative transformer model for symbolic regression, 2021.

[68] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2023.

[69] N. Zhang, S. Ding, J. Zhang, and Y. Xue. An overview on restricted boltzmann machines. *Neurocomputing*, 275:1186–1199, 2018.

[70] Z. Zhang, R. Ma, J. Hu, and Q. Wang. Approach the gell-mann-okubo formula with machine learning. *Chinese Physics Letters*, 39(11):111201, oct 2022.

Vita

<div align="center">

Abdulhakim M. Alnuqaydan

</div>

Place of Birth:

- Buraydah, Saudi Arabia

Education:

- University of Cincinnati, Cincinnati, OH, USA

  M.S. in Physics, April. 2018

- Qassim University, Saudi Arabia

  B.S. in Physics, Jun. 2012

Professional Positions:

- Teaching Assistant, Qassim University 2013–Present

Publications

- A. Alnuqaydan, S. Gleyzer, and H. Prosper. *Symba: symbolic computation of squared amplitudes in high energy physics with machine learning.* Machine Learning: Science and Technology, 4(1):015007, jan 2023.