UNIVERSITY OF CALIFORNIA SAN DIEGO

Understanding the High Energy Higgs Sector with the CMS Experiment and Artificial Intelligence

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Physics

by

Raghav Kansal

Committee in charge:

Javier Duarte, Chair
Maurizio Pierini
Hao Su
Zhuowen Tu
Frank Wuerthwein

2024

The Dissertation of Raghav Kansal is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To my family.

EPIGRAPH

तम आसीत्तमसा गूहळमग्रे प्रकेतं सलिलं सर्वाऽइदम् ।
तुच्छ्येनाभ्वपिहितं यदासीत्तपसस्तन्महिनाजायतैकम् ॥३ ॥

- नासदीय सूक्त

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My interest in physics was ignited in the 10th grade by the discovery of the Higgs boson and Stephen Hawking's *The Universe in a Nutshell*. These past five years, dedicated to understanding the Higgs and the mysteries of our early universe, have been a fulfillment of dreams then born and will be a part of my life I cherish forever.

I have many people to thank for allowing my childhood passion to flourish into this dissertation. I start first and foremost with my parents and little Kli for their love and support at every step. I thank as well my entire extended family for making me feel at home around the world, from Delhi to Pondicherry and California to Texas. I am especially grateful to all my grandparents, whose wisdom, selflessness, and memory inspire me always.

I would not have survived this PhD — particularly the two long years of lockdown and two long quarters of E&M, without my amazing friends, old and new, in San Diego: Aneesh, Biswa, Chris, Davide, Dro, Elliot, Gerald, Hulk, and Varun (to name a few). I thank as well my fellow CMS students, Farouk and Yanxi, with whom I moved across continents and conducted an ancillary PhD in ping pong. Along with them, I thank my friends in Geneva and Chicago, including Fifi, Jay, Priyanka, and the LPC crew, for making my two years working at CERN and Fermilab so memorable. Most of all, I thank Praniti, for her sweetness and support throughout.

Like the universe, my journey in high energy physics (HEP) began with a bang: the CERN Openlab summer student program. It was a breathtaking experience, and I am grateful to Cliff, Frank, and my supervisor Maurizio for their support then and throughout my career since. Maurizio, in particular, introduced me to the (Nobel-prize winning!) potential of AI in HEP, and I have been hooked ever since.

More importantly, he introduced me to Javier right as we were both, perhaps serendipitously, joining UCSD in the Fall of 2019. Javier has been the most brilliant, kind, and supportive advisor I could have asked for, and I thank him for teaching me a lot more than just physics. Along with him, I thank many awesome postdocs and scientists, including Cristina, Daniel, Nhan, Petar, and Si for their mentorship through the years.

Parts I—III are primarily original work for this dissertation, discussing the standard model, the CMS experiment and the LHC, AI and ML, and statistics, and building on several references listed therein. On the topic of statistics, I thank Javier and Nick Smith for countless discussions (as well as their much-needed help with analysis software and the CMS combine tool)!

Part IV presents novel methods for producing and validating fast simulations of the CMS detector using AI. I thank Maurizio for introducing me to this topic as a summer student and his guidance since, and Javier for supporting this work, and all the research directions it bloomed, since the beginning of my PhD. I also thank my fellow students on the topic, Mary and Breno; Nadya for her support during my time at CERN; Kevin Pedro for lending his expertise on CMS simulations; and the IRIS-HEP institute and the Fermilab LPC for supporting this work through the IRIS-HEP fellowship and the LPC AI fellowship and graduate scholarship, respectively.

Part V describes searches for high energy Higgs-boson pair production in the $b\bar{b}VV$ channel using data collected by the CMS experiment during Run 2 of the LHC. I thank Cristina for her hands-on guidance on both the physics and technical aspects from the start and her patience as I refactored our codebase every week. I thank as well Javier, Petar, Si, and the rest of our boosted double-Higgs working group, as well as Nhan and the DASZLE team, for their advice and support. I thank finally Nick, Lindsey, and all the Coffea and Scikit-HEP developers for building a wonderful and supportive Pythonic HEP ecosystem.

Part VI on the JetNet library and Lorentz-equivariant ML represents a collection of work [61–63] on which I mentored some amazing students at UC San Diego and more. I thank them all for choosing me as their mentor, and Javier for encouraging and supporting us graduate students in engaging in so many rewarding mentorship opportunities.

Chapter 7 is, in part, a reprint of the materials as they appear in R. Kansal. "Symmetry Group Equivariant Neural Networks," (2020); and NeurIPS, 2021, R. Kansal; J. Duarte; H. Su; B. Orzari; T. Tomei; M. Pierini; M. Touranakou; J.-R.

Vlimant; and D. Gunopulos. Particle Cloud Generation with Message Passing Generative Adversarial Networks. The dissertation author was the primary investigator and author of these papers.

Part IV is, in part, a reprint of the materials as they appear in the NeurIPS ML4PS Workshop, 2020, R. Kansal; J. Duarte; B. Orzari; T. Tomei; M. Pierini; M. Touranakou; J.-R. Vlimant; and D. Gunopulos. Graph generative adversarial networks for sparse data generation in high energy physics; NeurIPS, 2021, R. Kansal; J. Duarte; H. Su; B. Orzari; T. Tomei; M. Pierini; M. Touranakou; J.-R. Vlimant; and D. Gunopulos. Particle Cloud Generation with Message Passing Generative Adversarial Networks; and Phys. Rev. D, 2023, R. Kansal; A. Li; J. Duarte; N. Chernyavskaya; M. Pierini; B. Orzari; and T. Tomei; Evaluating generative models in high energy physics; and the NeurIPS ML4PS Workshop, 2024, A. Li; V. Krishnamohan; R. Kansal; J. Duarte; R. Sen; S. Tsan; and Z. Zhang; Induced generative adversarial particle transformers. The dissertation author was the primary investigator and (co-)author of these papers.

Chapters 4.3.2 and 4.3.4 and Part V, in part, are currently being prepared for the publication of the material by the CMS collaboration. The dissertation author was the primary investigator and author of these papers.

Part VI is, in part, a reprint of the materials as they appear in JOSS, 2023, R. Kansal; C. Pareja; Z. Hao; and J. Duarte; JetNet: A Python package for accessing open datasets and benchmarking machine learning methods in high energy physics; and Eur. Phys. J. C, 2023, Z. Hao; R. Kansal; J. Duarte; and

N. Chernyavskaya; Lorentz group equivariant autoencoders. The dissertation author was the primary investigator and (co-)author of these papers.

VITA

2019	Bachelor of Science in Physics and Computer Engineering University of California San Diego *Summa cum laude*
2019	CERN Openlab Summer Student
2019	IRIS-HEP Fellow
2019-2024	Graduate Student Researcher University of California San Diego
2021-2022	Artificial Intelligence Fellow Fermilab LHC Physics Center
2023-2024	Graduate Scholar Fermilab LHC Physics Center
2024	Doctor of Philosophy in Physics University of California San Diego
2024-	AI/Schmidt Postdoctoral Scholar Research Associate California Institute of Technology and Fermilab

PUBLICATIONS

Note: as a member of the CMS collaboration, I have been an author on all CMS papers since 2019. The following includes only the CMS publications to which I made significant contributions during my PhD.

1. CMS Collaboration, "Search for Nonresonant Pair Production of Highly Energetic Higgs Bosons Decaying to Bottom Quarks and Vector Bosons", in prep, CMS-HIG-23-012 (2023).

2. CMS Collaboration, “Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the two b quarks and four light quarks final state”, in prep, CMS-B2G-23-007 (2023).

3. A. Li*, V. Krishnamohan*, R. Kansal, J. Duarte, R. Sen, S. Tsan, and Z. Zhang, “Induced generative adversarial particle transformers”, NeurIPS ML4PS Workshop (2023), arXiv:2312.04757.

4. R. Kansal, C. Pareja, Z. Hao, and J. Duarte, “JetNet: A Python package for accessing open datasets and benchmarking machine learning methods in high energy physics”, **JOSS** 8, 5789 (2023).

5. Z. Hao, R. Kansal, J. Duarte, and N. Chernyavskaya, “Lorentz group equivariant autoencoders”, **Eur. Phys. J. C** 83, 485 (2023), arXiv:2212.07347.

6. R. Kansal, A. Li, J. Duarte, N. Chernyavskaya, M. Pierini, B. Orzari, and T. Tomei, “Evaluating generative models in high energy physics”, **Phys. Rev. D** 107, 076017 (2023), arXiv:2211.10295.

7. CMS Collaboration, “Search for Nonresonant Pair Production of Highly Energetic Higgs Bosons Decaying to Bottom Quarks”, **Phys. Rev. Lett.** 131, 041803 (2023), arXiv:2205.06667.

8. R. Kansal, J. Duarte, H. Su, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, “Particle Cloud Generation with Message Passing Generative Adversarial Networks”, NeurIPS (2021), arXiv:2106.11535.

9. F. Mokhtar, R. Kansal, and J. Duarte, “Do graph neural networks learn traditional jet substructure?”, NeurIPS ML4PS Workshop (2022), arXiv:2211.09912.

10. M. Touranakou, N. Chernyavskaya, J. Duarte, D. Gunopulos, R. Kansal, B. Orzari, M. Pierini, T. Tomei, and J.-R. Vlimant, “Particle-based fast jet simulation at the LHC with variational autoencoders”, **Machine Learning: Science and Technology** 3, 035003 (2022), arXiv:2203.00520.

11. A. Apresyan, D. Diaz, J. Duarte, S. Ganguly, R. Kansal, N. Lu, C. M. Suarez, S. Mukherjee, C. Peña, B. Sheldon, and S. Xie, “Improving Di-Higgs Sensi-

tivity at Future Colliders in Hadronic Final States with Machine Learning", Contribution to Snowmass Summer Study (2022), arXiv:2203.07353.

12. Y. Chen et al., "A FAIR and AI-ready Higgs boson decay dataset", **Sci. Data** 9, 31 (2021), arXiv:2108.02214.

13. F. Mokhtar, R. Kansal, D. Diaz, J. Duarte, J. Pata, M. Pierini, and J.-R. Vlimant, "Explaining machine-learned particle-flow reconstruction", NeurIPS ML4PS Workshop (2021), arXiv:2111.12840.

14. S. Tsan, R. Kansal, A. Aportela, D. Diaz, J. Duarte, S. Krishna, F. Mokhtar, J.-R. Vlimant, and M. Pierini, "Particle graph autoencoders and differentiable, learned energy mover's distance", NeurIPS ML4PS Workshop (2021), arXiv:2111.12849.

15. B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J. Duarte, R. Kansal, J.-R. Vlimant, and D. Gunopulos, "Sparse Data Generation for Particle-Based Simulation of Hadronic Jets in the LHC", ICML LXAI Workshop (2021), arXiv:2109.15197.

16. R. Kansal, J. Duarte, B. Orzari, T. Tomei, M. Pierini, M. Touranakou, J.-R. Vlimant, and D. Gunopulos, "Graph generative adversarial networks for sparse data generation in high energy physics", NeurIPS ML4PS Workshop (2020), arXiv:2012.00173.

ABSTRACT OF THE DISSERTATION

Understanding the High Energy Higgs Sector with the CMS Experiment and Artificial Intelligence

by

Raghav Kansal

Doctor of Philosophy in Physics

University of California San Diego, 2024

Javier Duarte, Chair

This dissertation describes efforts towards understanding the Higgs boson at the highest energies humanly accessible, using the CMS experiment at the Large Hadron Collider and advances in artificial intelligence (AI) and machine

learning (ML). We present searches for resonant and nonresonant Higgs-boson (H) pair production in the all-hadronic two beauty-quark and two vector boson (V) final state, using a novel strategy to measure the quartic HHVV coupling and search for new Higgs-like bosons. By targeting highly Lorentz-boosted Higgs pairs, we probe effects of potential new physics in the high energy Higgs sector, which could hold answers to fundamental mysteries of nature such as baryon asymmetry.

To enable these and future searches, we introduce as well significant developments in AI/ML, including in the identification of boosted $\mathrm{H} \rightarrow \mathrm{VV}$ decays with deep transformer networks and advances in AI-accelerated fast simulations of the CMS detector. The latter notably includes the development of the first, highly performant generative models for point-cloud data in high energy physics, which have the potential to improve CMS' computational efficiency by up to three orders of magnitude. We also highlight novel solutions to the important and challenging problems of calibrating and validating these ML techniques. Finally, we present new approaches to search for new physics in a model-agnostic manner, using physics-informed ML methods equivariant to Lorentz transformations.

The quartic HHVV coupling is observed (expected) to be constrained to $[-0.04, 2.05]$ ($[0.05, 1.98]$) at the 95% confidence level relative to the standard model prediction, representing the second-most sensitive measurement of this coupling by CMS to date. Exclusion limits on the production cross section of new heavy resonances decaying to two Higgs-like bosons are expected to be as low as 0.3 fb for high resonance masses.

Introduction

The story so far: In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move.

- Douglas Adams, The Restaurant at the End of the Universe

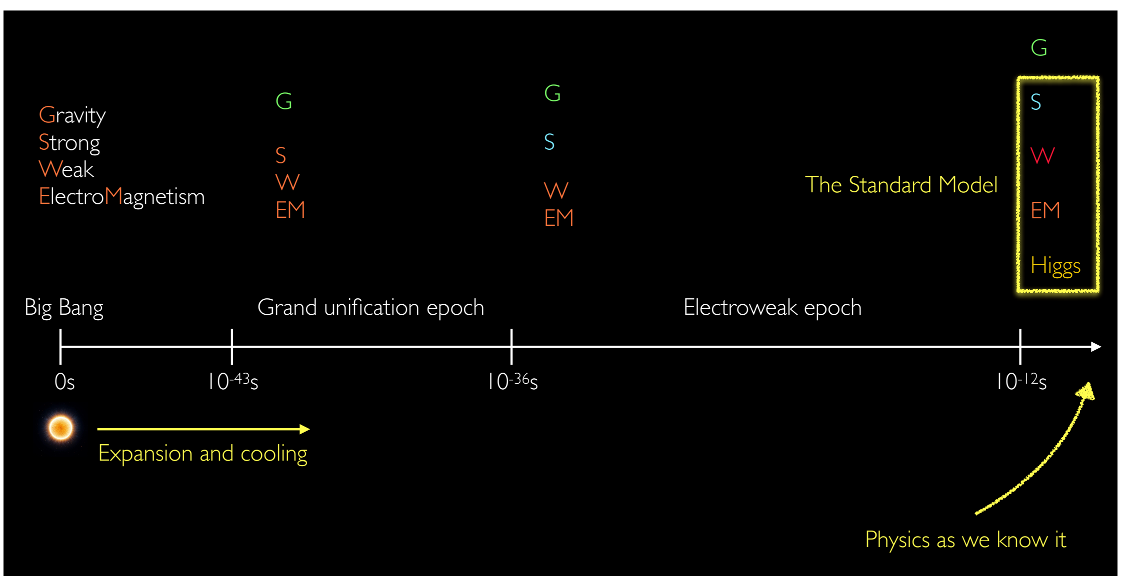


Figure 1. The timeline and evolution of forces in the early universe.

The universe started with a bang. A massive burst of energy, temperature, and pressure, with all four fundamental forces — electromagnetism, nuclear weak, nuclear strong, and gravity — as one. Immediately after, the universe expanded and cooled, and after about 10^{-43} seconds, gravity parted ways. 10^{-36}s later, the strong force separated as well and finally, by around 10^{-12}s, so did the weak and electromagnetic forces,"turning on" the Higgs field in the process and leaving us with the fundamental forces and laws of physics as we know them today (Figure 1).

Electromagnetism, the nuclear weak and strong forces, the Higgs field, and all known elementary particles can be elegantly described by the standard model (SM) of particle physics. Over the last 60 years, it has proven a monumentally successful theory, both explaining and predicting physical phenomena up to energies produced naturally only within a nanosecond of the Big Bang. These include the prediction of the Higgs boson 50 years before its discovery, explanations for radioactive decay and the binding of atomic nuclei, and the unification of the electromagnetic and weak forces. However, despite its triumphs, there remain fundamental mysteries that the SM cannot explain.

The most glaring of these is its reconciliation, or lack thereof, with gravity, for which a quantum, SM-compatible theory has proven elusive. There is also abundant cosmological evidence of "dark" matter and energy, constituting 95% of the universe and yet finding no justification from the SM. Other subtle mysteries include the inconsistency between the matter-antimatter *a*symmetry we observe and the symmetry the SM predicts, the mechanism for neutrino masses, and the origin of flavor.

The work in this dissertation is motivated by the strong possibility of many of the answers being tied to the Higgs boson. It is our newest discovered and least understood elementary particle, and the unique nature of the Higgs field and its interactions leaves open vast potential for intriguing new physics in this sector. As electroweak symmetry breaking, i.e., the separation of the weak and electromagnetic forces, is intimately connected to a phase transition of the Higgs field, many theories naturally link this transition with the breaking of the matter-antimatter [64] and flavor symmetries [65] as well. The Higgs boson may also be the connection between the SM and the dark sector [66], while the "Higgs-Saw" [67] mechanism is a promising explanation for dark energy.

Predictions of these theories include new, rare, Higgs-like particles and/or minute deviations to the interactions of the Higgs boson from the SM. However, as many of the phenomena therein would have occurred during the electroweak epoch or earlier (see Figure 1), these effects would manifest only at the highest energies, comparable to that of < 1ps after the Big Bang. This dissertation presents two complementary efforts to probe such effects, by (1) searching for new, highly energetic Higgs bosons, and (2) measuring Higgs interactions uniquely sensitive to new, high energy physics.

We do so using the Large Hadron Collider (LHC) at CERN. The LHC accelerates and collides extremely high-speed protons, producing energies comparable to the early universe just 10ps after the Big Bang. We observe these collisions with the Compact Muon Solenoid (CMS) experiment, one of four massive detectors at the LHC, and one of the two that discovered the Higgs boson in 2012. Crucially, we emphasize that, with the exponentially increasing rate of collisions and data at the LHC, the CMS experiment is entering an era of unprecedented potential for scientific discovery.

To fully realize this, however, and maximize the impact of our new data, significant computational innovation is required. To this end, we also present in this dissertation several novel AI techniques to identify high energy Higgs bosons, accelerate simulations of the CMS detector, and complement traditional data analysis techniques with model-agnostic searches for new physics. Particular emphasis is placed on the development of physics-informed machine learning (ML) algorithms, which uniquely leverage biases of high energy physics (HEP) data to improve their performance and robustness. Namely, we introduce the first generative models for *point-cloud* data in HEP, which respect the sparsity and high granularity of detector data, and the first anomaly detection models equivariant to Lorentz transformations.

We also describe significant efforts towards *validating* such AI techniques, which is critical for them to ultimately have an impact in the field. Specifically, we apply a novel method for calibrating ML algorithms targeting Higgs to vector boson decays, which has proven effective not only for the analyses presented in this dissertation but for the broader CMS physics program as well. We additionally present several studies and new statistical techniques for evaluating fast simulations. The combination of these and our new AI models has the potential to revolutionize the computing paradigm in CMS, improving the computational efficiency of our simulations by up to three orders of magnitude, and ensuring trust in their modeling of the underlying physics.

This dissertation is organized as follows. Part I introduces the theoretical basis for this dissertation, starting with the mathematical framework behind symmetries in physics (Chapter 2) and of quantum field theory (Chapter 3) before detailing the SM of particle physics (Chapter 4). Part II then describes the experimental apparatus used in this dissertation: the LHC (Chapter 5) and the CMS experiment (Chapter 6). Part III concludes the background material with an introduction to ML in HEP (Chapter 7), as well as the data analysis and statistical framework used in this dissertation (Chapter 8).

Parts IV—VI comprise the novel contributions of this dissertation. Part IV presents new methods for producing and validating fast simulations of the CMS detector using ML, which will be critical to maximizing the scientific output of the LHC in the coming decade. These methods leverage advancements in generative modeling to develop novel, physics-informed simulation techniques that are orders of magnitude faster than traditional methods. We also discuss new techniques for robust evaluation of such fast simulation techniques, and the outlook for their use in CMS.

Part V then presents two novel searches to understand the high energy

Higgs sector of the SM, targeting the production of Lorentz-boosted Higgs boson pairs, which decay into two beauty quarks and two vector bosons. Such searches are critical to understanding the properties of the Higgs boson and searching for the effects of new physics at very high energies. We discuss the analysis techniques used in these searches, particularly the use of deep transformer networks to identify Higgs-boson decays to two vector bosons for the first time, and competitive constraints achieved on new physics models and the two-Higgs-two-vector-boson coupling.

Finally, Part VI outlines the development of new software to facilitate research in ML and HEP and ML techniques that respect the symmetries of the high energy collisions that we study. Namely, we introduce the JetNet Python package, which has proven impactful in this field, and a novel ML algorithm for searching for new physics while remaining robust to Lorentz-transformations of our data.

Part I

Theoretical Background

Chapter 1

Introduction to the Standard Model

God used beautiful mathematics in creating the world. — Paul Dirac [68]

The standard model (SM) of particle physics is perhaps the greatest scientific theory of all time. It is a mathematical representation of three fundamental forces, all known elementary particles, and their collective interactions. In a broader sense, it is also the culmination of centuries of iterative, syncretic experimental results and theoretical advances, from Newton's laws of motion up to the discovery of the Higgs boson. That such a wide array of seemingly idiosyncratic physical phenomena and theories — electricity, magnetism, radioactive decays, quantum mechanics, special relativity, the structure of the atom, the binding of the nucleus, the behavior of elementary particles, and more — can all be encapsulated at their most primordial level into a single theory exemplifies the beauty of the SM.

This beauty is perhaps most apparent when viewing the SM through the lens of *symmetries*. Symmetries provide an elegant way to precisely describe the extremely complex physics mentioned above. Indeed, superficially, the SM can be viewed simply as a classification of elementary particles and their interactions according to their behavior under different symmetries of the universe and its mathematical description.

This is illustrated in Figure 1.1, listing the SM particles and their properties. They are first divided into two classes, fermions and bosons, based on how they behave under Lorentz transformations — a fundamental physical symmetry of nature. This simple distinction has profound implications: fermions constitute matter, i.e. what all the "stuff" in the universe is made out of, while bosons are the particles responsible for forces and their interactions. Specifically, the photon mediates electromagnetism, the $W^{\pm}$ and Z bosons the weak force, and gluons the strong force. There is also the Higgs boson, which is special: it does not mediate a force in the classical sense, but its interactions with elementary particles are what imbues them with mass.

Each force is intimately tied to a symmetry in the SM, and particles are further distinguished by their behavior under these symmetries — or, equivalently, how they are affected by this force. Fermions are divided by those interacting (quarks) and not interacting (leptons) with the nuclear strong force, while each of their rows in Figure 1.1 further separates them by different "charges" under the weak force. Additionally, each particle's mass and electric charge represent

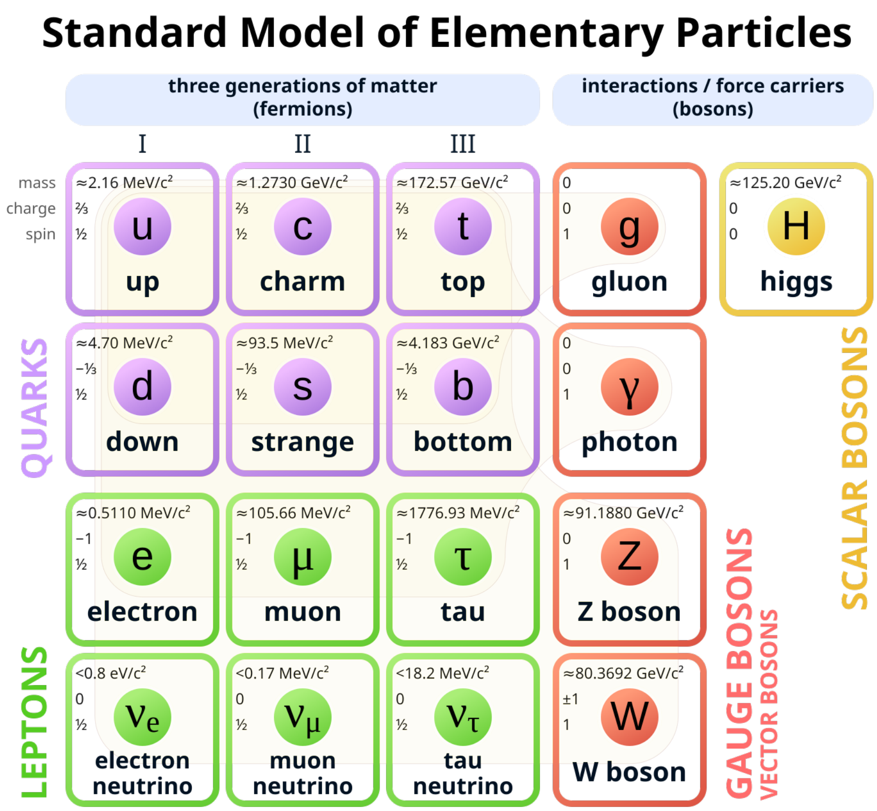


Figure 1.1. Particles and their classifications in the SM, reproduced from Ref. [1].

the strength of its interaction with the Higgs and electromagnetic fields, respectively. Finally, we can see a mysterious *almost*-symmetry: there are three copies, or "flavors" or "generations", of each fermion, which are entirely identical but for their masses (e.g. the electron, muon, and tau family of particles). Such a structure may suggest the presence of new, yet-to-be-discovered forces tied to this symmetry.

The goal of Part I is to make this picture more precise, and lay the theoretical foundation for the work discussed in this dissertation. The mathematical frameworks needed to do so are called group theory and quantum field theory (QFT), and are the subjects of Chapters 2 and 3, respectively. Equipped with these tools, we then describe the SM in Chapter 4, including the interactions discussed above and, of most relevance to the subject of this dissertation, the phenomenon of jets, the Higgs sector, and Higgs boson pair production within and beyond the SM.

These chapters build off of several great resources, including:

- David Tong's extremely useful and insightful lecture notes on QFT [69], gauge theories [70], and the standard model [71];
- John McGreevy's great course on symmetry in physics [72] (which I had the pleasure of attending in the Fall of 2020);
- Frederic Schuller's precise lectures on the geometric anatomy of theoretical physics [73];
- Tony Zee's *Group Theory in a Nutshell for Physicists* [74] and *Quantum Field Theory in a Nutshell* [75];
- Peskin and Schroeder's classic *An Introduction to Quantum Field Theory* [76];
- Gavin Salam's lectures on *Elements of QCD for hadron colliders* [21];
- and Hong Liu [77] and Ricardo Matheus' [78] clear, recorded lectures on QFT.

Chapter 2

Symmetries in physics

Perfectly balanced, as all things should be. — Thanos

Symmetry is a powerful and beautiful way to understand nature. Intuitively, a symmetry is a transformation that leaves an object unchanged. For example, a plain square has a four-fold rotational symmetry: it looks identical rotated once, twice, thrice, or four times by 90°.

Similarly, in physics, a symmetry is a transformation that leaves the laws of physics unchanged. Electromagnetism, for example, is invariant to translations in space or time: electric charges and currents should behave the same in San Diego 5 years ago as in Geneva today. Understanding such symmetries, and accounting for them in our mathematical formulation, has been a guiding principle in the development of the SM over the 20th century, and is one in understanding it as well.

In recent years, symmetries have also guided the development of machine learning algorithms in becoming more powerful and efficient. A particular focus is placed in this dissertation on such *equivariant* algorithms, which respect the symmetries and *inductive biases* of our high energy physics data. This chapter lays the foundation for these ideas, which we discuss in more detail in Chapter 7 and contribute to in Chapter **??**.

In this chapter, we first introduce the framework for describing symmetries, group theory, in Section 2.1. We then describe Lie algebras for continuous symmetries, and derive representations for the algebra corresponding to 3D rotations, in Section 2.2. We conclude in Section 2.3 with a discussion of the Lorentz and Poincaré groups, comprising the fundamental symmetries of spacetime, whose irreducible representations are what we call particles.

2.1 Group theory

The mathematical formalism for describing symmetries is called *group theory*.

Definition 2.1.1. The fundamental object in group theory is a *group*, defined as a pair $(G, \bullet)$, where G is a set and $\bullet : G \times G \to G$ is the group operation, which together satisfies the following properties:

i) Associativity: $\forall a, b, c \in G : (a \bullet b) \bullet c = a \bullet (b \bullet c)$.
ii) Identity element: $\exists e \in G : \forall a \in G : a \bullet e = e \bullet a = a$.

iii) Inverse element: $\forall a \in G : \exists a^{-1} \in G : a \bullet a^{-1} = a^{-1} \bullet a = e$.

Definition 2.1.2. Note from Definition 2.1.1 that the group operation is not necessarily commutative ($\forall a, b \in G : a \bullet b = b \bullet a$). If this condition does hold, the group is called an *abelian* group.

Example 2.1.1. To formalize the four-fold rotation symmetry of a square discussed above, we can define the group $\mathbb{Z}_4$ as $(\{0, 1, 2, 3\}, +_4)$, where $+_4$ is addition modulo 4, and the elements of the set can represent rotations by $0°$, $90°$, $180°$, and $270°$, respectively. One can check that $\mathbb{Z}_4$ satisfies all the properties of an abelian group.

Group representations

To make the abstract mathematical structure of the group more concrete, we next consider *representations* of groups.

Definition 2.1.3. A group representation R, of dimension d, is a mapping of the group elements to $d \times d$ matrices $D(g)$ in some d-dimensional vector space V, such that the group operation is preserved: $D(g_1)D(g_2) = D(g_1 \bullet g_2)$. Necessarily, this means that $D(e) = \mathbb{1}$, the identity matrix of V. Representations of a group are not unique, and arbitrarily many new represenations can be constructed simply by taking tensor sums and products, denoted by the $\oplus$ and $\otimes$ symbols respectively, of existing ones.

Definition 2.1.4. An *irreducible representation* (irrep) is one with no non-trivial invariant subspaces, i.e., it cannot be decomposed into the tensor sums of smaller-dimensional representations.[1]

Example 2.1.2. The group $\mathbb{Z}_4$ from Example 2.1.1 can be represented simply as scalar complex numbers ($V = \mathbb{C}$):

$$\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & e^{i\frac{\pi}{2}} & e^{i\pi} & e^{i\frac{3\pi}{2}} \end{array} \tag{2.1.1}$$

One can check this satisfies the conditions of Definition 2.1.3, and since it is 1-dimensional, it is also irreducible.

Definition 2.1.5. Every group has a $|G|$-dimensional *regular representation* R^{reg}, where $|G|$ is the number elements of the group, called the *order* of the group. The vector space $V = \text{span}\{|g\rangle \,|\, g \in G\}$, and the representation is defined such that

$$D^{\text{reg}}(g)\,|h\rangle = |gh\rangle\,. \tag{2.1.2}$$

Example 2.1.3. For our $\mathbb{Z}_4$ group, we can use the set of four basis vectors $\{|0\rangle = \mathbf{e}_0, |1\rangle = \mathbf{e}_1, |2\rangle = \mathbf{e}_2, |3\rangle = \mathbf{e}_3\}$ in $\mathbb{R}^4$, and derive the matrices $D^{\text{reg}}(g)$ such that

[1] Technically, certain pathological reducible representations of non-compact groups also cannot be decomposed into irreps, so "non-decomposability" is a necessary but insufficient condition for irreps.

they transform $|g\rangle$ according to the respective group operations:

$$
\begin{aligned}
D^{\text{reg}}(0) &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad D^{\text{reg}}(1) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \\
D^{\text{reg}}(2) &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad D^{\text{reg}}(3) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.
\end{aligned}
\tag{2.1.3}
$$

The regular representation has some fun properties, such as its reducibility into irreps with each irrep appearing as many times in the decomposition as its dimension. For us, it will mostly serve as a useful way to think about the *adjoint* representation we will encounter below.

Continuous symmetries

Symmetries can be *discrete*, as above, as well as continuous.

Example 2.1.4. A circle has a continuous 2D rotational symmetry; rotations by any angle θ leave it invariant. This corresponds to the *special orthogonal* group in 2-dimensions SO(2).

Definition 2.1.6. More generally, the orthogonal group in n dimensions, $\mathrm{O}(n)$, is defined as the group of orthogonal, or "distance-preserving", $n \times n$ matrices M, s.t. $MM^T = \mathbb{1}$. The *special orthogonal* group $\mathrm{SO}(n)$ is the subgroup of $n \times n$ orthogonal matrices with determinant 1, essentially retaining only rotations while removing reflections.

As their definition suggests, the $\mathrm{SO}(n)$ group elements have a natural representation as the $n \times n$ rotation matrices. For SO(2), these are of the form:

$$M(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{2.1.4}$$

where $\theta \in [0, 2\pi)$ is the angle of rotation. These $n \times n$ matrix representations are called the *fundamental* or *defining* representations of $\mathrm{SO}(n)$.

Definition 2.1.7. SO(2) is *isomorphic* — meaning identical to in terms of its group-theoretic properties — to the *unitary* group U(1). The *unitary* group $\mathrm{U}(n)$ is the group of $n \times n$ unitary matrices, i.e., those satisfying $M^\dagger M = MM^\dagger = \mathbb{1}$, where $M^\dagger$ is the conjugate transpose, or Hermitian conjugate (h.c.) of M. The special unitary group $\mathrm{SU}(n)$, again is the subgroup of $n \times n$ unitary matrices with determinant 1. As we will soon see, these groups effectively define the structure of the SM.

U(1) has the simple 1D fundamental representation:

$$M(\theta) = e^{i\theta}, \tag{2.1.5}$$

i.e., all complex numbers of unit magnitude.

Definition 2.1.8. An infinite group is *compact* if a group-invariant sum or integral over the group elements is finite. U(1) is compact, as

$$\int_0^{2\pi} d\theta = 2\pi \tag{2.1.6}$$

is finite. Indeed, all SO(n) and SU(n) groups are compact.

Examples of important non-compact groups include the group of translations in n dimensions and the Lorentz group, which we will discuss in detail in Section 2.3.

2.2 Lie algebras

We next introduce the concepts of Lie groups and Lie algebras, which are highly useful in understanding the structure and representations of continuous groups.

Definition 2.2.1. A *Lie group* is a group that is also a differentiable manifold, or "smooth". Virtually all continuous groups we consider in physics are Lie groups. What this means is that we can think of the operation of any arbitrary group element as equivalent to N successive infinitesimal operations of the form

$$g(\varepsilon_A) = \mathbb{1} + i\varepsilon_A T_A, \tag{2.2.1}$$

where ε_A are infinitesimal and indexing the continuous group parameters, e.g. rotation angles for SO(n), T_A are called the *generators* of the group, and we are using Einstein notation, implicitly summing over the index A. Thus, for a general element $g(\theta_A)$, where $\theta_A = N\varepsilon_A$ as defined above, we have

$$g(\theta_A) = \left(\mathbb{1} + i\frac{\theta_A}{N}T_A\right)^N \xrightarrow{N\to\infty} e^{i\theta_A T_A}. \tag{2.2.2}$$

This is somewhat analogous to Taylor expansion in calculus, except for Lie groups only the first order / derivative term is necessary to capture the group behavior.[2]

Definition 2.2.2. The *Lie algebra*, $\mathfrak{g}$, of a group is defined by the set of commutation relations between its generators:[3]

$$[T_A, T_B] = i f_{ABC} T_C, \tag{2.2.3}$$

where $[T_A, T_B] = T_A T_B - T_B T_A$ is the commutator of T_A and T_B, and f_{ABC} are called the *structure constants* of $\mathfrak{g}$. As $[T_A, T_B] = -[T_B, T_A]$, the structure constants must be totally antisymmetric in the swapping of their indices.

[2]This is because, based on the Campbell-Baker-Hausdorff [79] formula, higher order terms in the expansion of exponential form of g in Eq. 2.2.2 involve only commutators of the generators.

[3]An *algebra* $(V, \bullet)$ is a vector space V with a bilinear operation $\bullet : V \times V \to V$. Examples include the cross product of vectors and matrix multiplication of square matrices. The Lie algebra is the special case where $\bullet$ is the commutator.

Example 2.2.1. For the U(1) group, we can see directly from Eq. 2.1.5 that the sole generator of the group is $T = 1$. This has the rather uninteresting Lie algebra $\mathfrak{u}(1)$ of $[1, 1] = 0$, stemming from the fact that the group is abelian. Next, we look at the more interesting SO(3) and SU(2) groups, where the power of Lie algebras shines.

Fundamental and adjoint representations of the $\mathfrak{so}(3)$ and $\mathfrak{su}(2)$ algebras

We now introduce two important representations of Lie algebras, using the SO(3) and SU(2) groups as examples — both because of their importance in physics, and as their derivation introduces a number of useful concepts for the following sections. SO(3) and SU(2) are very closely related: SU(2) is a *double cover* of SO(3), which means that every rotation in SO(3) can be mapped to two elements of SU(2). Importantly, however, they are locally isomorphic near the identity, meaning they have the same Lie algebra.

We can derive the generators of SO(3) by using the properties of the special orthogonal group ($R^T R = \mathbb{1}$). From Eq. 2.2.1, we have

$$\begin{aligned} &R(\varepsilon) \cong \mathbb{1} + i\varepsilon T \\ &R^T R = \mathbb{1} + i\varepsilon(T^T + T) + \mathcal{O}(\varepsilon^2) \stackrel{!}{=} \mathbb{1} \\ \Rightarrow\ &T^T = -T. \end{aligned} \tag{2.2.4}$$

Thus T are antisymmetric matrices, of which for $N = 3$ dimensions there are three linearly independent ones:

$$J_x = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_y = i\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad J_z = i\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{2.2.5}$$

labeled as x, y, z as they represent rotations around the respective axes. The factor of i ensures the reality of the infinitesimal rotations in Eq. 2.2.4 and also that the generators are Hermitian.[4] These provide us with the fundamental representation of $\mathfrak{so}(3)$, and should be familiar as the angular momentum operators in quantum mechanics (QM). By exponentiating these, as in Eq. 2.2.2, we obtain the fundamental representation of the SO(3) group: $R(\theta) = e^{i\theta_i J_i}$.

To find the fundamental representation of $\mathfrak{su}(2)$, we can follow the same procedure as above, using the unitarity constraint $R^\dagger R = \mathbb{1}$ for $N = 2$ dimensional complex matrices, which yields:

$$\begin{gathered} T_1 = \frac{1}{2}\sigma_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad T_2 = \frac{1}{2}\sigma_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ T_3 = \frac{1}{2}\sigma_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \end{gathered} \tag{2.2.6}$$

[4]Note that the conventions around this factor are inconsistent in the literature and likely, despite our best efforts, will be inconsistent in this chapter as well.

where σ_i are the Pauli matrices — the angular momentum operators for the spin of spin-1/2 particles in QM. Either set of generators yield the following Lie algebra of both groups:

$$[T_A, T_B] = i\epsilon_{ABC}T_C, \tag{2.2.7}$$

where the structure constants f_{ABC} of the algebra are simply ϵ_{ABC}, the totally antisymmetric Levi-Civita tensor.

Structure constants themselves furnish the following representation of the corresponding Lie algebra:

$$[T_A]_{BC} = -if_{ABC}. \tag{2.2.8}$$

This can be confirmed by plugging this representation into the commutator in Eq. 2.2.3 and using the Jacobi identity [80]. As B, C index the number of generators, we see that this representation has a dimension equal to the number of generators of the Lie algebra, and it is called its *adjoint* representation. It is analogous to the regular representation (Definition 2.1.5) for a Lie algebra, with the underlying vector space spanned by the generators $V = \text{span}\{|T_A\rangle\}$ and the requirement that $D(T_A)|T_B\rangle = if_{ABC}|T_C\rangle$.

Definition 2.2.3. The dimension of a Lie group is defined as the number of generators of the group. Thus, it is the same as the dimension of the adjoint representation.

As it turns out, for $\mathfrak{so}(3)$ and $\mathfrak{su}(2)$, the adjoint representation $[T_A]_{BC} = -i\varepsilon_{ABC}$ is simply the fundamental representation of $\mathfrak{so}(3)$. More generally, the dimensions of the fundamental and adjoint representations of SO(n) and SU(n) are given in Table 2.1. The significance of these representations, as we will see, is that the force carriers (i.e., gauge bosons) of the SM live in the adjoint representation of their associated gauge group, while the matter particles live in either their fundamental or trivial representations.

Table 2.1. Dimensions of the fundamental and adjoint representations of the SO(n) and SU(n) groups.

Group	**dim(Fundamental)**	**dim(Adjoint)**
SO(n)	n	$n(n-1)/2$
SU(n)	n	$n^2 - 1$

General representations

So far we have discussed two representations of the $\mathfrak{so}(3)$ and $\mathfrak{su}(2)$ algebras. The general representations can be derived in much the same way as finding the eigenstates of the angular momentum operator in QM. We first choose a basis in which one of the generators, conventionally J_z, is diagonal, and label eigenvectors of J_z as $|m\rangle$ with eigenvalue m:

$$J_z |m\rangle = m |m\rangle . \tag{2.2.9}$$

These eigenvectors, by definition, form a basis for the representations of the generators, so counting them tells us the dimensions of allowed representation. To do so, we define the “raising” and “lowering” operators $J_\pm = J_x \pm iJ_y$, with commutation relations

$$[J_z, J_\pm] = \pm J_\pm, \quad [J_+, J_-] = 2J_z. \tag{2.2.10}$$

These are named so because

$$J_z J_\pm |m\rangle = [J_\pm Jz \pm J_\pm] |m\rangle = (m \pm 1) J_\pm |m\rangle , \tag{2.2.11}$$

i.e., $J_\pm |m\rangle$ are eigenvectors of J_z with eigenvalues $m \pm 1$, implying

$$J_\pm |m\rangle = c^\pm_{m\pm 1} |m \pm 1\rangle , \tag{2.2.12}$$

where $c_{m\pm 1}$ are normalization constants. Now if we assume that the representation is finite-dimensional and label the highest-weight state $|j\rangle$ — such that $J_+ |j\rangle = 0 \Leftrightarrow c^+_{j+1} = 0$ — we can iteratively lower the state and solve for the normalization constants until we reach the lowest-weight state. By doing so we find that $c^-_{-j-1} = 0 \Rightarrow$ the lowest weight state is in fact $|-j\rangle$.[5]

[5]See, for example, Chapter IV.2 in Zee [74] for a more detailed derivation.

Thus, we conclude the algebra allows $2j+1$-dimensional representations spanned by $\{\left|-j\right\rangle, \left|-j+1\right\rangle, \ldots, \left|j-1\right\rangle, \left|j\right\rangle\}$, with $j \in \mathbb{Z}^{\geq 0}/2$ (non-negative integers and half-integers only). Each possible j indexes a different representation of the group, and any eigenstate can thus be labeled by $|j, m\rangle$. We have already seen the $j = 1/2$ and $j = 1$ representations explicitly in Eqs. 2.2.6 and 2.2.5, respectively, while the $j = 0$ is simply the trivial representation of the group ($D(g) = 1 \ \forall g \in G$).

Definition 2.2.4. More generally, irreducible representations of a group are labeled by eigenvalues of the *Casimir invariants*, or Casimirs, of the group. Casimirs are operators that commute with all generators of the group. For $\mathfrak{so}(3)$ and $\mathfrak{su}(2)$, there is only one Casimir,

$$J^2 = J_x^2 + J_y^2 + J_z^2. \tag{2.2.13}$$

This is the total angular momentum operator, which we know from QM commutes with all the J_is and for any eigenstate $|j, m\rangle$ has eigenvalue $j(j+1)$:

$$J^2 \left|j, m\right\rangle = j(j+1) \left|j, m\right\rangle. \tag{2.2.14}$$

As expected, since the Casimir commutes with all the generators, its eigenvalues depend only on the irrep j. We have also seen that individual states can be further labeled using the eigenvalues of a set of maximally commuting operators, in this case $\{J^2, J_z\}$.

These representations can directly be used to derive those corresponding to the SU(2) and SO(3) group except that, surprisingly, the latter does not admit the half-integer irreps; essentially, SU(2) has double the irreps because it is the double cover of SO(3). Overall, the irreps of $\mathfrak{su}(2)$ and $\mathfrak{su}(3)$ are quite significant in physics, with direct applications to classical and quantum mechanics, and, moreover, they will also serve as the building blocks for the representations of the Lorentz and Poincaré groups in the next section.

2.3 Particles are irreps of the Poincaré group

The Poincaré group comprises all the physical symmetries of "flat" spacetime (i.e, without gravity), i.e. all the transformations which leave the laws of physics invariant. These include Lorentz transformations (boosts and rotations) and spacetime translations.

Particles can be defined as a "set of states which mix only among themselves under Poincaré transformations" (Schwartz [81] Ch. 8.1), leaving attributes like their mass and spin invariant. Elementary particles are those for which there is no smaller subset of states that also have this property. Thus, they correspond exactly to irreducible representations of the Poincaré group! That the physical and seemingly nebulous concept of a particle can be so precisely defined and characterized by a mathematical analysis of the symmetries of spacetime is one of the most beautiful results of fundamental physics.

In this section, we describe the irreps of the Poincaré group, starting first with the Lorentz group alone.

The (proper, orthochronous) Lorentz group

We know from special relativity that "flat" spacetime (i.e., without gravity) is described by 4D Minkowski space $\mathbb{R}^{1,3}$. This is a real vector space equipped with the metric $\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$, which defines distances, or inner products $\langle \cdot, \cdot \rangle$, between 4-vectors $x_\mu = (x_0, x_1, x_2, x_3)$ as:

$$\langle x, y \rangle \equiv x_\mu y^\mu \equiv \eta_{\mu\nu} x^\mu y^\nu = x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3. \tag{2.3.1}$$

Definition 2.3.1. The *Lorentz group* is the group of all matrices M orthogonal under the Minkowski metric $M^T \eta M = \eta$, and is called $\mathrm{O}(1,3)$. This is the analog in flat spacetime to distance-preserving transformations in Euclidean space (e.g., $\mathrm{O}(3)$).

Definition 2.3.2. The *proper, orthochronous* Lorentz group $\mathrm{SO}^+(1,3)$ is the subgroup of $\mathrm{O}(1,3)$ matrices continuously connected to the identity. Physically, these are the transformations that preserve the orientation of space and direction of time, and are typically what we refer to as *Lorentz transformations*. The two transformations of $\mathrm{O}(1,3)$ not included in $\mathrm{SO}^+(1,3)$ are parity $P = \mathrm{diag}(1, -1, -1, -1)$ and time reversal $T = \mathrm{diag}(-1, 1, 1, 1)$ (shown in the 4-vector representation), which flip the sign of spatial and temporal components of 4-vectors, respectively. Surprisingly, these are not symmetries of nature — they are violated by the weak interaction! Generally, in this chapter, when we talk about the Lorentz group or Lorentz invariance, we are referring only to the proper, orthochronous Lorentz group.

Generators of the Lorentz group

Lorentz transformations Λ are generated by six antisymmetric matrices, three for boosts (K_i) and three for rotations (J_i). In the 4-vector representation, these are:

$$K_x = -i\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ K_y = -i\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ K_z = -i\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$
$$J_x = i\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix},\ J_y = i\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix},\ J_z = i\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.3.2}$$

Lorentz transformations can thus be represented as

$$\Lambda(\theta, \beta) = e^{i(\theta_i J_i + \beta_i K_i)}, \tag{2.3.3}$$

where θ and β are the rotation and boost parameters, respectively, An important property of the Lorentz group is that it is not compact. This is related to the fact that the generators for boosts K_i in the representation above are not Hermitian, which means the corresponding group elements $e^{i\beta_i K_i}$ are not unitary. In fact, there are no finite-dimensional unitary representations of the Lorentz group [82]. Unitarity of operators is an important condition for the invariance of physical properties under transformations in QM, and the consequences of this for the SM will be discussed in Chapter **??**.

Lie algebra of the Lorentz group

From Eq. 2.3.2, we can derive the commutation relations of the generators and, hence, the Lie algebra:

$$
\begin{aligned}
[K_i, K_j] &= -i\epsilon_{ijk}J_k, \\
[J_i, J_j] &= i\epsilon_{ijk}J_k, \\
[J_i, K_j] &= i\epsilon_{ijk}K_k.
\end{aligned}
\tag{2.3.4}
$$

Moreover, if we define the operators

$$
J_i^+ = \frac{1}{2}(J_i + iK_i), \quad J_i^- = \frac{1}{2}(J_i - iK_i), \tag{2.3.5}
$$

we find that $\mathfrak{so}(1,3)$ contains two mutually commuting $\mathfrak{su}(2)$ subalgebras:

$$\begin{aligned}
[J_i^+, J_j^+] &= i\epsilon_{ijk}J_k^+, \\
[J_i^-, J_j^-] &= i\epsilon_{ijk}J_k^-, \\
[J_i^+, J_j^-] &= 0.
\end{aligned} \tag{2.3.6}$$

This implies the irreps of $\mathfrak{so}(1,3)$ are simply two copies of the irreps of $\mathfrak{su}(2)$ from Section 2.2, indexed as (j_1, j_2) with $j_1, j_2 \in \mathbb{Z}^{\geq 0}/2$ and dimension $(2j_1+1)(2j_2+1)$.

With this, we can easily obtain the generators J_i, K_i for the smallest few irreps:

$$(0,0): \qquad J_i^+ = J_i^- = 0 \qquad \Rightarrow \qquad J_i = K_i = 0. \tag{2.3.7}$$

$$(1/2,0): \qquad J_i^+ = \frac{1}{2}\sigma_i,\ J_i^- = 0 \qquad \Rightarrow \qquad J_i = \frac{1}{2}\sigma_i, K_i = -\frac{i}{2}\sigma_i. \tag{2.3.8}$$

$$(0,1/2): \qquad J_i^+ = 0,\ J_i^- = \frac{1}{2}\sigma_i \qquad \Rightarrow \qquad J_i = \frac{1}{2}\sigma_i, K_i = \frac{i}{2}\sigma_i. \tag{2.3.9}$$

$$\vdots$$

The $(1/2,1/2)$ irrep is actually our familiar 4-vector representation, but it is more involved to recover the generators in the same form as Eq. 2.3.2.[6]

[6]See e.g. Ref. [83].

Representations of the Lorentz group

It turns out the above four irreps of the Lorentz group are all we need for the SM. Their nomenclature and corresponding elementary particle fields are listed in Table 2.2. Notably, fermions are classified as those with half-integer total spin $j = j_1 + j_2$, and bosons with integer j. Their radically different behavior is a consequence of the Spin-Statistics theorem [84] (a notoriously difficult theorem to prove [85]), which states that half-integer spin particles obey Fermi-Dirac statistics and integer spin particles Bose-Einstein statistics.

All known fermionic particle fields live in the $(1/2, 0) \oplus (0, 1/2)$, or *Dirac spinor*, representation. The $(1/2, 0)$ and $(0, 1/2)$ representations are called the *left-* and *right-handed Weyl spinors* respectively, where the handedness refers to the direction of their spin angular momentum relative to their momentum. Physically, this means there is a left-handed and right-handed copy of each fermion, and they have to be packaged together in a Dirac spinor to have masses without violating parity, as we discuss in Chapter **??**. We will also see that left- and right-handed representations can be equivalently thought of as particles and antiparticles.

The $(1/2, 0) \oplus (0, 1/2)$ representation technically also includes real *Majorana spinors* as a subspace, which can represent neutral fermions. The only candidate for these in the SM are right-handed neutrinos and, in fact, the existence of such Majorana neutrinos could potentially explain the curiously small *left-handed* neutrino masses through a process called the *seesaw mechanism* [86, 87]. To date, however, no experimental evidence for these, such as neutrinoless double beta decay [88] or same-sign charged dilepton decays [89], has been observed.

On another technical note, the Lorentz group, similar to SO(3), does not itself admit half-integer, fermionic representations. Thus, the true spacetime symmetry group is actually the double cover of $SO(1,3)$, $Spin(1,3)$! Indeed, there are many subtleties to the Lorentz group, some of which will be revisited in the context of Lorentz-group equivariant neural networks in Chapter **??**. To conclude, however, it is worth emphasizing again the remarkable physical insight these seemingly abstract group-theoretic concepts deliver. We are able to classify a fundamental dichotomy of particle physics — bosons versus fermions, and their completely different behavior — simply by their representation under the Lorentz (or, rather, the $Spin(1,3)$) group!

Table 2.2. Representations of the Lorentz group and their associated particle fields in the SM.

Representation (j_1, j_2)	**Name**	**Elementary Fields**
$(0,0)$	Scalar	Higgs boson
$(1/2, 0)$	Left-handed Weyl spinor	—
$(0, 1/2)$	Right-handed Weyl spinor	—
$(1/2, 0) \oplus (0, 1/2)$	Dirac spinor	All fermions
$(1/2, 1/2)$	Vector	$g, \gamma, W^{\pm}$, and Z gauge bosons

Lie algebra of the Poincaré group

The Poincaré group is Lorentz transformations plus spacetime translations. Just as angular momentum generates rotations, translations are generated by the momentum operator P_μ. P_μ and the Lorentz generators J_i and K_i together comprise the generators of the Poincaré group, and its algebra is thus the Lorentz algebra (Eq. 2.3.4) plus the commutation relations with the P_μs:[7]

$$\begin{aligned}
[P_\mu, P_\nu] &= 0, \\
[J_i, P_0] &= 0, \\
[J_i, P_j] &= i\epsilon_{ijk} P_k, \\
[K_i, P_0] &= -iP_i, \\
[K_i, P_j] &= i\eta_{ij} P_0.
\end{aligned} \tag{2.3.10}$$

As is conventional, the Greek indices run over all four spacetime dimensions, while the Latin indices only the three spatial.

The Poincaré algebra can be expressed more compactly by first combining

[7]See Appendix **??** for a derivation.

the Lorentz generators into the antisymmetric tensor $M_{\mu\nu}$:

$$M_{\mu\nu} = \begin{pmatrix} 0 & K_x & K_y & K_z \\ -K_x & 0 & J_z & -J_y \\ -K_y & -J_z & 0 & J_x \\ -K_z & J_y & -J_x & 0 \end{pmatrix} \quad \Rightarrow \Lambda(\omega) = e^{\frac{i}{2}\omega^{\mu\nu}M_{\mu\nu}}, \tag{2.3.11}$$

with $\omega_{\mu\nu}$ another antisymmetric tensor containing the six rotation and boost parameters. The algebra can be then written as:

$$\begin{aligned} [M_{\mu\nu}, M_{\rho\sigma}] &= i(\eta_{\nu\rho}M_{\mu\sigma} - \eta_{\mu\rho}M_{\nu\sigma} - \eta_{\nu\sigma}M_{\mu\rho} + \eta_{\mu\sigma}M_{\nu\rho}), \\ [M_{\mu\nu}, P_\rho] &= i(\eta_{\nu\rho}P_\mu - \eta_{\mu\rho}P_\nu), \\ [P_\mu, P_\nu] &= 0. \end{aligned} \tag{2.3.12}$$

Irreps of the Poincaré group

As we saw from Section 2.2, we can derive the irreps of an algebra using its Casimir invariants (Definition 2.2.4). Each set of their eigenvalues uniquely labels an irrep, while each basis state within the irreps is indexed by eigenvalues of a maximal set of commuting operators (e.g., $\{J^2, J_z\} \to |j, m\rangle$ for $\mathfrak{so}(3)$). Note that in the following, we simply provide a sketch of the derivations and point to, for example, Zee GT [74] Chapter VII.2 and Tong SM [71] Chapter 1.1.2 for more detailed proofs and discussion.

The Casimirs of the Poincaré algebra are the operators

$$P^2 = P_\mu P^\mu \text{ and } W^2 = W_\mu W^\mu, \tag{2.3.13}$$

where

$$W_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} P^\sigma \tag{2.3.14}$$

is the *Pauli-Lubanski vector*, the relativistic analog of the angular momentum operator J_i. Furthermore, P_μ commutes with both, and we can label its eigenstates as $|p\rangle$,

$$P_\mu |p\rangle = p_\mu |p\rangle \,, \tag{2.3.15}$$

which represent single-particle states with 4-momentum p_μ. These are therefore eigenstates of P^2 as well, with eigenvalues m^2, the squared mass of the particle:

$$P^2 |p\rangle = p_\mu p^\mu |p\rangle = m^2 |p\rangle \,. \tag{2.3.16}$$

Thus, we see that the mass of a particle, m, is one label of the irreps, with states therein indexed by p_μ. The other label, the particle spin j, is based on the eigenvalue of W^2:

$$W^2 |p, j\rangle \propto j(j+1) |p, j\rangle \,. \tag{2.3.17}$$

The easiest way to see this is to, for a given $m > 0$, pick a single eigenstate $|p^*\rangle$. The simplest is the rest frame $p^*_\mu = (m, 0, 0, 0)$. The subgroup of Poincaré transformations which leave $|p^*\rangle$ invariant is called its *little group*. In this case, it comprises all 3D rotations — i.e., SO(3). Indeed, if we look at the Pauli-Lubanski vector acting on $|p^*\rangle$,

$$W_\mu |p^*\rangle = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} p^{*\sigma} |p^*\rangle \quad \Rightarrow W_0 = 0, W_i = -mJ_i, \tag{2.3.18}$$

we simply recover the generators J_i of $\mathfrak{so}(3)$.[8] Therefore,

$$W^2 |p^*, j\rangle = m^2 J^2 |p^*, j\rangle = m^2 j(j+1) |p^*, j\rangle\,, \tag{2.3.19}$$

using the eigenvalues of J^2 from Eq. 2.2.14. Although we chose here to look at a specific state $|p^*\rangle$, this can be shown to hold for all states $|p\rangle$ in the irrep.[9] Thus, we see that irreps of the Poincaré group, and, hence, particles, are characterized by their mass m and spin j.

Massive versus massless particles

Continuing with the massive, $m > 0$, particle case, we know as well from Section 2.2 that the eigenstates within the $|p, j\rangle$ irreps are further labeled by their spin

[8] This also motivates why W_μ can be thought of as relativistic angular momentum.

[9] By choosing an eigenstate $|p^*, j\rangle$ of P^2 and looking for transformations which leave it invariant, we "induced" a subgroup, SO(3), and used its representation theory to derive the irreps of the Poincaré group. Such a representation is hence called an *induced representation*.

along a particular axis: $J_z \left| p, j, m_j \right\rangle = m_j \left| p, j, m_j \right\rangle$, with $m_j \in \{-j, -j+1, \ldots, j-1, j\}$. Thus, massive particles $\left| m, j \right\rangle$ exist in $2j+1$ spin states $\otimes$ an infinite number of momentum states $\left| p \right\rangle$, $p_\mu p^\mu = m^2$.

However, this is not the case for *massless* particles, which have a different little group. Recall that we can never boost into the rest frame of a massless particle to define the simple $\left| p^* \right\rangle$ we did above. Instead, let us consider the next-best state $\left| p' \right\rangle$, $p'_\mu = (E, 0, 0, E)$. Its little group turns out to be E(2), the Euclidean group in 2D, whose representations and implications for massless particles are considerably more involved. However, the upshot is that it as well has irreps characterized by spin j (and mass $m = 0$), but with only two *helicity* eigenstates therein.

As it turns out, physically, the mass of particles is based on the strength of their interactions (or lack thereof) with the Higgs boson, the particle at the center of this dissertation. We will discuss the mechanisms for these and all fundamental interactions in the next chapter, and see that symmetries are crucial in understanding them.

Chapter 3

Quantum field theory

Quantum mechanics describes nature as absurd from the point of view of common sense. And yet it fully agrees with experiment. So I hope you can accept nature as She is — absurd. — Richard Feynman

The standard model is a quantum field theory (QFT). It describes the universe as a collection of fields associated with the various elementary particles. At each point in spacetime, there is a random probability for these fields to interact and create or destroy their respective particles.

This means we have an electron field, a photon field, a Higgs field, etc. spread across the universe, and all electrons, photons, and Higgs bosons are identical *quantum excitations* of these. The interactions of the electron and photon fields, for example, are what we experience as electromagnetism.

As Feynman says, this may all sound absurd. Fields are highly unintuitive, "unphysical" concepts. It can be hard to imagine that particles, matter, and indeed all of us, are simply a collection of quanta probabilistically popping out and dropping back into an abstract cosmic sea.

Not only that, historically, QFT often appeared intractable and even nonsensical, yielding results such as negative energy and infinite mass particles. Its development underwent multiple periods of stagnation and ardent opposition, including by Richard Feynman who suggested in 1945 that field theory be abandoned altogether [90] before changing his mind and making seminal contributions to quantum electrodynamics.

Yet, through the collective efforts of generations of physicists, QFT can now explain nearly every observed phenomenon in particle physics, up to the highest experimental energies. Moreover, it has made some of the most staggering and precise predictions in the history of physics, all of which proved to be in complete agreement with experiment. These range from the calculation of the electron's magnetic moment up to 12 significant digits, to the prediction of the Higgs boson 50 years before its discovery. Its unprecedented experimental success is why we believe "it is the language in which the laws of Nature are written" (Tong SM [71]).

In this chapter, we first introduce classical and quantum field theory for free particles in Section 3.1, before discussing their interactions and connection to physical observables in Section 3.2. We then detail *gauge theories* and the beautiful connection between symmetries, QFT, and the forces of nature, in Section 3.3.

We conclude with a description of the Higgs mechanism, which is of particular relevance to this dissertation, in Section 3.4. Further background on classical mechanics and mathematical details of quantization, interactions, and spinors can be found in Appendix **??**.

3.1 Free scalar field theory

Historically, field theory was in part an attempt to develop *local* theories rather than those, such as Newtonian gravity, implying *action-at-a-distance*.[1] The idea is to associate each point in space and time with a value or set of values $\phi_a(\mathbf{x}, t)$, called fields. As long as these fields interact only at the same point in spacetime or, at most, with their immediate neighbors (via their derivatives), the theory is guaranteed to be local. Classic examples include the vector-valued electric and magnetic fields $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{B}(\mathbf{x}, t)$. The behavior of the fields is encapsulated by the *Lagrangian* of the system.

In this section, we briefly recap the Lagrangian formulation of classical field theory (Section 3.1.1) and Noether's theorem connecting symmetries to conserved quantities (Section 3.1.2). We then present the quantized form of the free scalar field, and the interpretation of particles as excitations of these fields, in Section 3.1.3. Finally, we conclude in Section 3.1.4 with a discussion of particle *propagators*, which are the connection between these abstract fields and

[1] See Weinberg's notes on a history of QFT [90] for a nice summary of its historical development.

the physical observables we can measure in experiments. Further background on classical Lagrangian and Hamiltonian mechanics can be found in Appendix **??**, and on quantization in Appendix **??**.

3.1.1 Classical field theory

The Lagrangian of a classical field $\phi(\mathbf{x}, t)$ is given as a function of the field and its derivatives:

$$L(t) = \int d^3x \; \mathcal{L}(\partial_\mu \phi, \phi), \tag{3.1.1}$$

where $\mathcal{L}$ is the *Lagrangian density*. The action is the integral of L over time, or $\mathcal{L}$ over spacetime:

$$S = \int L \; dt = \int d^4x \; \mathcal{L}. \tag{3.1.2}$$

The equations of motion (EOMs) of the field is derived from the *principle of stationary action*, which states that the true path is an extremum of S, yielding the *Euler-Lagrange* (E-L) equations:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{3.1.3}$$

The Klein-Gordon equation

The Lagrangian for a *free, scalar* relativistic field $\phi(\mathbf{x}, t)$ is:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2, \tag{3.1.4}$$

The E-L equation for this Lagrangian is called the *Klein-Gordon equation*:

$$\partial_\mu\partial^\mu\phi + m^2\phi \equiv (\Box + m^2)\phi = 0, \tag{3.1.5}$$

where $\Box \equiv \partial_\mu\partial^\mu = \partial_t^2 - \nabla^2$ is the *d'Alembertian* operator.

The Klein-Gordon equation is essentially the relativistic generalization of the Schrödinger equation. Just as the Schrödinger equation quantizes the non-relativistic EOM $E = p^2/2m$, the Klein-Gordon equation converts the relativistic EOM for a free particle

$$E^2 = p^2c^2 + m^2c^4 \tag{3.1.6}$$

into quantum operator form, with $E \longrightarrow \hat{E} = i\hbar\partial_t$ and $p \longrightarrow \hat{p} = -i\hbar\nabla$:

$$\begin{aligned}\hat{E}^2 &= \hat{p}^2c^2 + m^2c^4 \\ -\hbar^2\partial_t^2\phi &= -\hbar^2c^2\nabla^2\phi + m^2c^4\phi \\ \Rightarrow\ (\partial_t^2 - c^2\nabla^2 &+ \frac{m^2c^4}{\hbar^2})\psi = 0.\end{aligned} \tag{3.1.7}$$

Natural units

It is conventional in high energy physics to use *natural units*:

$$\hbar = c = 1. \tag{3.1.8}$$

Besides being notationally convenient, this enables all dimensionful physical quantities to be described by the same scale — conventionally, in terms of energy, e.g. in units of electronvolts (eV). For example:

- Mass: $E = mc^2 \longrightarrow m = E$,
- Compton wavelength: $\lambda = \hbar/mc \longrightarrow \lambda = 1/E$,
- Momentum: $p = \hbar/\lambda \longrightarrow p = E$.

We define each quantity to have a dimension in terms of energy, i.e. energy, mass, and momentum all have dimension $[E] = [m] = [p] = 1$, while length has dimension $[\lambda] = -1$. Thus, in natural units Eqs. 3.1.5 and 3.1.7 are identical.

3.1.2 Symmetries and Noether's theorem

Noether's theorem states an important consequence of continuous symmetries of a system: they are associated with a physical conserved currents. For example, translational and rotational invariance of the potential energy imply conservation of momentum and angular momentum, respectively.

More precisely, if a continuous transformation on the field

$$\phi(x) \longrightarrow \phi'(x) = \phi(x) + \epsilon\Delta\phi(x) \tag{3.1.9}$$

is a symmetry, leaving the EOMs invariant, it can be shown to imply the existence of a conserved current j^μ:

$$\partial_\mu j^\mu = 0, \quad j^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\Delta\phi - \mathcal{J}^\mu, \tag{3.1.10}$$

and conserved charge Q

$$Q = \int_{\text{allspace}} d^3x\ j^0. \tag{3.1.11}$$

Example: translation symmetry

Consider a translation-invariant theory, such as for the free scalar field (Eq. 3.1.4). A spacetime translation $x^\mu \longrightarrow x^\mu - a^\mu$ leads to the transformation $\phi(x) \longrightarrow \phi(x + a) \simeq \phi(x) + a^\mu\partial_\mu\phi(x)$, yielding the conserved current:

$$(j^\mu)_\nu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - \delta^\mu_\nu\mathcal{L}, \tag{3.1.12}$$

which we call the energy-momentum tensor T^μ_ν. The associated conserved quantities (or "charges") are the total energy and momentum of the field configuration:

$$E = \int d^3x\, T^{00}, \quad P^i = \int d^3x\, T^{0i}. \tag{3.1.13}$$

For our free scalar field, this turns out to be:

$$\begin{aligned} E &= \int d^3x\, \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}m^2\phi^2, \\ P^i &= \int d^3x\, \dot{\phi}\, \partial^i\phi. \end{aligned} \tag{3.1.14}$$

The interpretation of these as energy and momenta is described further in Appendix **??**.

Example: a U(1) internal symmetry

Symmetries such as translational and rotational invariance are spacetime, or external, symmetries. An *internal symmetry* is a transformation that acts only on the fields, at each point of spacetime. A simple example is the complex scalar field $\psi(x)$, for which we can write down the free Lagrangian:

$$\mathcal{L} = \partial_\mu\psi^*\partial^\mu\psi - m^2\psi^*\psi. \tag{3.1.15}$$

This Lagrangian possesses an internal U(1) symmetry: it is invariant under $\psi(x) \to e^{i\alpha}\psi(x)$ for any constant α. Noether's theorem tells us this too has the important physical consequences of a conserved current and charge:

$$j^\mu = i(\psi^* \partial^\mu \psi - \psi \partial^\mu \psi^*), \quad Q = \int d^3x \; i(\psi^* \partial^0 \psi - \psi \partial^0 \psi^*). \tag{3.1.16}$$

Once quantized, we will see this exactly corresponds to the conservation of electric charge!

In fact, we say that a field that transforms as so under a global U(1) rotation

$$\begin{aligned} \psi(x) &\to e^{iq\alpha}\psi(x), \\ \psi^*(x) &\to e^{-iq\alpha}\psi^*(x), \end{aligned} \tag{3.1.17}$$

is *charged* under the U(1) symmetry, with charge q (and its complex conjugate with charge $-q$).

3.1.3 Quantization

Details of quantization can be found in Appendix **??**; briefly, the quantized solution to the Klein-Gordon equation for a real field is akin to a superposition of plane waves:

$$\hat{\phi}(\mathbf{x}, t) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} (\hat{a}_\mathbf{p} \, e^{ip\cdot x} + \hat{a}^\dagger_\mathbf{p} \, e^{-ip\cdot x}) \tag{3.1.18}$$

where $p \cdot x = p_\mu x^\mu$ is the 4D spacetime inner product and $p_\mu = (\omega_p = \sqrt{|\mathbf{p}|^2 + m^2}, \mathbf{p})$. The quantum operators $\hat{a}_\mathbf{p}$ and $\hat{a}^\dagger_\mathbf{p}$ are annihilation and creation operators, respectively, for a particle with momentum $\mathbf{p}$ and mass m, just as for states in a quantum harmonic oscillator.[2]

The solutions for a complex scalar field are (see Appendix **??**):

$$\begin{aligned}
\hat{\psi}(\mathbf{x}, t) &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} (\hat{b}_\mathbf{p}\, e^{ip\cdot x} + \hat{c}^\dagger_\mathbf{p}\, e^{-ip\cdot x}), \\
\hat{\psi}^\dagger(\mathbf{x}, t) &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_p}} (\hat{b}^\dagger_\mathbf{p}\, e^{-ip\cdot x} + \hat{c}_\mathbf{p}\, e^{ip\cdot x}),
\end{aligned} \tag{3.1.19}$$

where we now have two sets of creation and annihilation operators, $\hat{b}_\mathbf{p}$ and $\hat{b}^\dagger_\mathbf{p}$, and $\hat{c}_\mathbf{p}$ and $\hat{c}^\dagger_\mathbf{p}$, again for particles with momentum $\mathbf{p}$ and mass m but *opposite charges*, respectively, under the U(1) symmetry described above. These are interpreted as particles and antiparticles.

Particle eigenstates

The Hilbert space of the free scalar field theory comprises the vacuum — 0-particle — state $|0\rangle$ and the states — i.e., particles — created by the creation operators

[2]Indeed, one can view particles simply has excited states of a continuous set of QHOs for different masses and momenta. This is made more precise in Appendix **??**.

acting on it:

$$\begin{aligned}
\hat{a}_{\mathbf{p}} \left|0\right\rangle = 0 \;\forall \mathbf{p} \quad &\Rightarrow \quad H\left|0\right\rangle = 0, \\
\left|\mathbf{p}\right\rangle \propto \hat{a}^\dagger_{\mathbf{p}} \left|0\right\rangle \quad &\Rightarrow \quad H\left|p\right\rangle = \omega_{\mathbf{p}} \left|p\right\rangle. \\
\left|\mathbf{p}_1, \mathbf{p}_2\right\rangle \propto \hat{a}^\dagger_{\mathbf{p}_1}\hat{a}^\dagger_{\mathbf{p}_2} \left|0\right\rangle \quad &\Rightarrow \quad H\left|\mathbf{p}_1, \mathbf{p}_2\right\rangle = (\omega_{\mathbf{p}_1} + \omega_{\mathbf{p}_2})\left|\mathbf{p}_1, \mathbf{p}_2\right\rangle \\
&\vdots \\
\left|\mathbf{p}_1, \ldots, \mathbf{p}_n\right\rangle \propto \hat{a}^\dagger_{\mathbf{p}_1} \ldots \hat{a}^\dagger_{\mathbf{p}_n} \left|0\right\rangle \quad &\Rightarrow \quad H\left|\mathbf{p}_1, \ldots, \mathbf{p}_n\right\rangle = \left(\sum_{i=1}^{n} \omega_{\mathbf{p}_i}\right)\left|\mathbf{p}_1, \ldots, \mathbf{p}_n\right\rangle,
\end{aligned} \tag{3.1.20}$$

where H is the Hamiltonian operator of the theory (see Appendix **??**), and $\left|\mathbf{p}\right\rangle$ is a state with momentum $\mathbf{p}$ and energy $\omega_p = \sqrt{|\mathbf{p}|^2 + m^2}$: i.e., a single particle of mass m. This is essentially the sum of the Hilbert spaces of an infinite number of QHOs, across all momenta, and is called the *Fock space*.

The momentum eigenstates are normalized such that their inner products are Lorentz scalars:

$$\left|\mathbf{p}\right\rangle = \sqrt{2E_{\mathbf{p}}}\, \hat{a}^\dagger_{\mathbf{p}} \left|0\right\rangle \quad \Rightarrow \quad \left\langle \mathbf{q}|\mathbf{p}\right\rangle = 2E_{\mathbf{p}}\; \delta^3(\mathbf{q} - \mathbf{p}). \tag{3.1.21}$$

For a scalar field, the creation operators commute amongst themselves, which means the states $\left|\mathbf{p}_1, \ldots, \mathbf{p}_n\right\rangle$ are symmetric under exchange of particles, and thus describe *bosons*.[3]

[3]Technically, the individual momentum eigenstates are not "physical", as they are not normalizable. Instead, particles exist in the form of a wavepacket with some spread in momenta $\varphi(\mathbf{p})$, which we typically assume to be smaller than our detector resolution can thus ignore (see Appendix **??**.)

Interpretation of the field operators

The action of the fields themselves on the vacuum is:

$$\hat{\phi}(\mathbf{x})\,|0\rangle = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2E_{\mathbf{p}}}e^{ip\cdot x}\,|\mathbf{p}\rangle\,. \tag{3.1.22}$$

This is very similar to the Fourier transform of the position eigenstate $|\mathbf{x}\rangle$ in nonrelativistic QM, except with an integral measure that is now Lorentz-invariant due to our normalizations above. Thus, we can roughly interpret $\phi(\mathbf{x})$ as an operator which creates a particle at position $\mathbf{x}$. However, we will see next that $\phi(\mathbf{x})\,|0\rangle \equiv |\mathbf{x}\rangle$, unlike in QM, is not *exactly* localized in position (although it's pretty close).

3.1.4 Propagators and Green functions

We now discuss briefly the concept of *propagators* in QFT, primarily because of their importance in relating quantum fields to physical observables of the theory, like scattering amplitudes (Section 3.2.2), but also because of some interesting insights they offer regarding our quantized particle states. The propagator is the amplitude for the associated particle at a spacetime point y to be found at x:

$$D(x-y) \equiv \langle 0|\phi(x)\phi(y)|0\rangle = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2E_{\mathbf{p}}}e^{ip\cdot(x-y)}. \tag{3.1.23}$$

This is also called the *two-point correlation function* between x and y.

Interestingly, it can be shown that, for a particle with mass $m > 0$, for space-like separated points, e.g. $x_0 = y_0, |\mathbf{x} - \mathbf{y}| \equiv r$,

$$D(r) \sim e^{-mr}, \tag{3.1.24}$$

i.e., it is not 0![4] However, it exponentially decays at rate of $1/m$, or the Compton wavelength. This tells us that there is a fundamental physical limit in relativistic QM to which a particle can be localized in space (or, at least, to which we can measure its position, related to the uncertainty principle).

Green functions

The propagator is closely related to the Green function $\Delta(x)$ of the Klein-Gordon equation, which is the solution (or response) to a delta function source:

$$(\Box + m^2)\Delta(x) = \delta^4(x). \tag{3.1.25}$$

The Green function $\Delta(x - y)$ effectively describes the effect on the field at x due to a localized source at y; hence, the connection to the two-point correlation

[4]Mathematically, this stems from the $1/2E_\mathbf{p}$ factor in the integral required for Lorentz invariance. Note, however, that this does not violate causality, since the commutator $[\phi(x), \phi(y)] = D(x - y) - D(y - x) = 0$ for spacelike separated points, meaning physically they cannot affect each other. For a complex field, $[\psi(x), \psi^*(y)] = 0$ has the interesting interpretation of a particle's amplitude for $x \rightarrow y$ being canceled by its antiparticle's amplitude for $y \rightarrow x$. Or, inversely, this tells us that causality necessitates the existence of antiparticles (Peskin and Schroeder [76] Chapter 2.4).

function, or propagator, above.

The form of $\Delta(x)$ can be found by Fourier transforming this equation to be:

$$\Delta(x) = \int \frac{d^4p}{(2\pi)^4} \frac{i}{p^2 - m^2} e^{-ip\cdot x}, \tag{3.1.26}$$

This has a pole on the real line at $p^2 = m^2 \Leftrightarrow E = \pm\sqrt{\mathbf{p}^2 + m^2}$, which means there is an ambiguity in defining the contour integral.

The choice of contour leads to four different Green functions, each with a different physical interpretation. The one choice we make in QFT is the *Feynman prescription*, often defined as

$$\Delta_F(x) = \int \frac{d^4p}{(2\pi)^4} \frac{i}{p^2 - m^2 + i\varepsilon} e^{-ip\cdot x}, \tag{3.1.27}$$

where the $i\varepsilon$ term resolves the ambiguity by shifting the poles infinitesimally above and below the real line, $\Delta_F(x)$ is called the *Feynman propagator*. It is related to our normal propagator above by

$$\Delta_F(x-y) = \begin{cases} \langle 0|\phi(x)\phi(y)|0\rangle = D(x-y) & \text{if } x^0 > y^0 \\ \langle 0|\phi(y)\phi(x)|0\rangle = D(y-x) & \text{if } x^0 < y^0 \end{cases} \equiv \langle 0|T\phi(x)\phi(y)|0\rangle\,, \tag{3.1.28}$$

where we call T the time-ordering operator.

3.2 Interactions

Like the silicon chips of more recent years, the Feynman diagram was bringing computation to the masses. — Julian Schwinger

We next make the field theory more interesting by adding interactions. We will continue with our scalar fields, first discussing the types of interactions we consider and the important concept of *renormalizability* in Section 3.2.1. We then focus on *weakly coupled* theories, where we can treat the interactions as small perturbations, as described in Section 3.2.2, and then discuss how to calculate the probability of interactions occurring using Feynman diagrams in Section 3.2.3. Finally, we outline how to translate these probabilities into the physical quantities we measure, namely decay rates and cross sections, in Section 3.2.4.

3.2.1 Interactions in the Lagrangian

Before diving into the calculations, it is useful to get an idea of the types of interactions that are "relevant" in a QFT using dimensional analysis. Consider the following generic Lagrangian for a single real scalar field:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + \sum_{n=3}^{\infty}\frac{\lambda_n}{n!}\phi^n. \tag{3.2.1}$$

The ϕ^n terms are what are new, representing interactions, and λ_n are called their *coupling constants*, determining their respective strengths. Broadly speaking, we only know how to make meaningful analytic calculations for interactions which we can treat as small perturbations to the free Lagrangian; indeed, there is much we do not understand about *strongly-coupled* theories such as QCD.

How do we decide whether an interaction is “small”? It certainly depends on the coupling constant, but λ is not necessarily dimensionless. The Lagrangian has energy (or mass) dimension 1 (using natural units, see Section 3.1.1), so

$$[\mathcal{L}] = 4, [m] = 1 \Rightarrow [\phi] = 1 \Rightarrow [\partial_\mu] = 1, [\lambda_n] = 4 - n. \tag{3.2.2}$$

We need λ to be small *relative* to different things, depending on its dimension. In fact, we use its dimension (or, equivalently, that of the interaction term) to categorize different interactions.

Relevant, marginal, and irrelevant interactions

$[\lambda_3] = 1$: This means λ_3 must be small compared to some energy E, which is typically the energy scale of our experiment or process of interest. Such an interaction therefore becomes a larger perturbation at lower energies, and smaller at high energies. These terms are called *relevant* because they affect the physics that we usually deal with.

$[\lambda_4] = 0$**:** These are called *marginal* interactions, which are small if $\lambda_4 \ll 1$.

$[\lambda_n] < 0, n > 4$**:** These interactions are small at low energies and large at high energies. Because of this, we typically do not need to consider them in a QFT; hence, they are called *irrelevant*.

Thus, in a sense, QFT is quite simple — we need only consider relevant and marginal interactions! In this case, $\lambda_3\phi^3$ and $\lambda_4\phi^4$. The same dimensional analysis also shows why we do not consider terms with more than two derivatives.

When we do want to explore the effects of irrelevant interactions, we can parametrize them as generic operators in the Lagrangian which are suppressed by powers of $(E/\Lambda)^{n-4}$, where Λ is the energy scale at which we expect these interactions to become relevant. This is (one of) the ideas behind *effective field theory* (EFT) [91, 92].

Renormalizability

The types of interactions present in a theory also determine its *renormalizability*. Calculations in QFT are inherently plagued by infinities, one of which we encountered as the zero-point energy of the quantized free scalar field (Section **??**). A general method for handling *ultraviolet* (UV) infinities — those which arise from integrating over momenta up to $|\mathbf{p}| \to \infty$ — is to impose a cut-off energy scale Λ on these integrals.

By doing so, we are essentially admitting, rightfully so, that we do not know what is going on arbitrarily high energies; hence, we do not expect our theory to be valid beyond Λ. We then, after performing the integrals, can take the limit $\Lambda \to \infty$ and hope and pray our result is independent of Λ. This is a simplified picture of *renormalization*.

However, the strength of irrelevant interactions only grows with energy, so $\Lambda \to \infty$ will lead to a divergence. Hence, we call theories with irrelevant interactions *non-renormalizable*. The SM is a renormalizable QFT and thus, as for our simple scalar field theory, its possible interactions are helpfully constrained. Most likely, it is simply an EFT of a higher energy theory, with the nonrenormalizable terms heavily suppressed by the scale of new physics!

3.2.2 S-matrix elements

As discussed above, we will focus on interactions in weakly-coupled theories, where they can be treated as small perturbations to the free Lagrangian. The quantized interaction terms comprise different combinations of creation and annihilation operators, corresponding to existing particles interacting, getting destroyed, and/or creating new ones. Broadly, we call these *scattering* processes, and the amplitude of these occurring is called the *S-matrix* element $\langle f|S|i\rangle$ between the initial and final particles states $|i\rangle$ and $|f\rangle$. The operator S, for scattering, is called the S-matrix.

Note that so far we have only been discussing the abstract notion of fields in the Lagrangian. We have highlighted many connections and interpretations relating fields to physical particles, but they are not the same; *fields are not particles*.[5] The S-matrix elements between particles are the physical quantities we measure: they are the basic *observables* of QFT.

Formally, fields and particles are related through the LSZ reduction formula [93], which expresses S-matrix elements in terms of the Green functions of the field (Section 3.1.4). The formula states that the S-matrix element between n incoming and m outgoing asymptotically free, on-shell particles is the residue of the $n + m$ particle pole of the associated fields' Green functions.[6]

This is a very powerful result in QFT. In this section, we heuristically explain its practical consequence, which is that the S-matrix element can be calculated using the time-ordered product of the interacting fields, up to different orders in the interaction coupling constant. In the following section, we then present the even more practical method of calculating such time ordered products using Feynman diagrams.

[5]This point is well emphasized in Aneesh Manohar's notes on EFT [91].

[6]Useful discussions of this can be found in Peskin and Schroeder [76] Chapter 7 and Schwartz [81] Chapter 6.

Scalar Yukawa Lagrangian

We will use *scalar Yukawa theory* as an example, which couples together our real and complex scalar fields, ϕ and ψ:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + \partial_\mu\psi^\dagger\partial^\mu\psi - M^2\psi^\dagger\psi - g\phi\psi^\dagger\psi. \tag{3.2.3}$$

The interaction term $g\phi\psi^\dagger\psi$ is called a *Yukawa interaction*, and the weak coupling condition is $g \ll m, M$.

A similar theory was originally developed by Hideki Yukawa to model the strong nuclear force between nucleons (ψ) via a hypothesized meson (ϕ) [94]. Indeed, such a meson was discovered a decade later via cosmic rays, and is called the pion [95]. Nobel Prizes were awarded for both the prediction and discovery. The difference in our theory is the scalar rather than fermionic nucleon, for simplicity; we will still, however, be able to reproduce the iconic physical feature of the theory: the Yukawa potential.

Under the weak coupling condition, we can treat the interaction term as a perturbation to the free Lagrangian and use perturbation theory and the *interaction picture* of QM to calculate the S-matrix elements for processes at any order in g (see Appendix **??**). For example, diagrams of first-order processes such as meson decay ($\phi \to \psi\psi$) and nucleon-antinucleon annihilation ($\psi\psi \to \phi$) are shown in Figure 3.1.

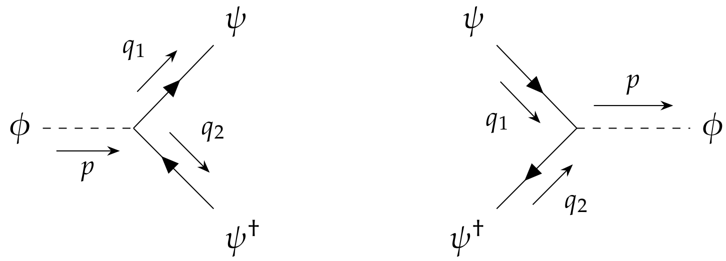


Figure 3.1. Feynman diagrams for meson decay (left) and nucleon-antinucleon annihilation (right).

Explicit calculation yields the S-matrix element for both processes (Appendix **??**):

$$\langle f|S|i\rangle^{(1)} = -ig(2\pi)^4\delta^{(4)}(p - q_1 - q_2). \tag{3.2.4}$$

The delta function ensures momentum conservation, and is in fact a general feature of all S-matrix elements. We typically define

$$\langle f|S - \mathbb{1}|i\rangle \equiv i(2\pi)^4\delta^{(4)}(\Sigma\, p)\mathcal{M}, \tag{3.2.5}$$

where $\mathcal{M}$ is called the *matrix element* of the process, and is the nontrivial component we must compute. For our first-order processes, the matrix element is simply $\mathcal{M} = -g$. However, explicit calculations quickly become intractable at higher orders; instead, we present a simpler alternative in the next section.

3.2.3 Feynman diagrams

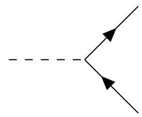
Feynman diagrams are intuitive and powerful tools for calculating S-matrix elements. We have already seen examples for our first-order meson decay and nucleon-antinucleon annihilation processes in Figure 3.1. They encode a lot of information (some of which is redundant, shown only for these first diagrams for clarity) and, as we will see, directly give us the matrix elements of the processes. Feynman diagrams for higher-order processes can be constructed by adding more vertices and *internal lines* connecting them. Details and some conventions used in this dissertation are given in Appendix **??**.

Feynman rules for scalar Yukawa theory

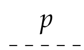
To read off the matrix element from a Feynman diagram, we take the product of factors associated to each element of the diagram, according to the *Feynman rules* of the theory. These rules are ultimately derived from and encode all our information about the underlying Lagrangian. They can be written in either position or momentum space; since we are working with momentum eigenstates, we will use the latter.

Definition 3.2.1. For our scalar Yukawa theory, the Feynman rules for calculating $i\mathcal{M}$ are:[7]

[7] These are derived nicely in Peskin and Schroeder [76] Chapter 4.7, albeit with fermionic electrons instead of our scalar "nucleons".

1. Vertices:  $= -ig$

2. Internal lines (propagators)

   Mesons:  $= \dfrac{i}{p^2 - m^2 + i\varepsilon}$ Nucleons:  $= \dfrac{i}{q^2 - M^2 + i\varepsilon}$

3. Impose momentum conservation at each vertex.
4. Integrate over the momentum k flowing through each loop $\int {}^{d^4k}/_{(2\pi)^4}$.

Note that the factors associated with internal lines are exactly the Feynman propagators from Section 3.1.4, which is in line with their interpretation as the amplitude for a particle to propagate from one point to another. For internal lines, the convention is for momentum to flow in the same direction as the particle flow, even for antiparticles. We see immediately that these rules reproduce the matrix element $\mathcal{M} = -g$ for our first-order processes, as expected. We discuss loops briefly at the end of this section; however, we focus primarily on *tree-level diagrams*, those without loops.

Nucleon-antinucleon scattering

One interesting higher-order example is nucleon-antinucleon scattering $\psi\psi^\dagger \to \psi\psi^\dagger$. At lowest order, we have the diagrams shown in Figure 3.2.

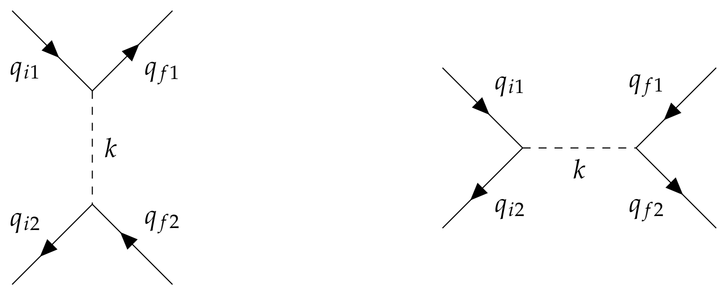


Figure 3.2. The two lowest order nucleon-antinucleon scattering diagrams.

The first two Feynman rules result in the same matrix element (Eq. **??**) for both. Imposing momentum conservation we find:

$$i\mathcal{M} = i(\mathcal{M}_{\text{left}} + \mathcal{M}_{\text{right}}) = (-ig)^2\left[\frac{1}{(q_{f1} - q_{i1})^2 - m^2} + \frac{1}{(q_{i1} + q_{i2})^2 - m^2}\right]. \tag{3.2.6}$$

Virtual particles

Note that by momentum conservation, the exchange meson does not have mass m, as $k^2 \neq m^2$. We say that this meson is a *virtual particle* and is *off-shell* (referring to the "mass shell" in k at $k^2 = m^2$). This may appear dangerously unphysical; however, we are saved by the fact that such off-shell particles always appear internally in the diagram and thus can never be observed. In a sense, they can be viewed simply as a mathematical convenience in QFT; no one knows their correct physical interpretation, if any.[8]

[8]To quote Hong Liu, "In physics, when we don't understand something, we give it a name and then claim we understand it." [77].

Mandelstam variables

Because these types of 2-by-2 scattering processes are so common in particle physics, they have standard names, based on the momenta in the denominator of the matrix element.

Definition 3.2.2. For incoming particle momenta p_{i1} and p_{i2} and outgoing momenta p_{f1} and p_{f2}, the *Mandelstam variables* are defined as:

$$\begin{aligned} s &= (p_{i1} + p_{i2})^2 = (p_{f1} + p_{f2})^2, \\ t &= (p_{i1} - p_{f1})^2 = (p_{i2} - p_{f2})^2, \\ u &= (p_{i1} - p_{f2})^2 = (p_{i2} - p_{f1})^2. \end{aligned} \tag{3.2.7}$$

We can see that the matrix elements for nucleon-antinucelon scattering (Eq. 3.2.6) can be rewritten in terms of t and s as:

$$\begin{aligned} i\mathcal{M}_{\text{left}} &= (-ig)^2 \cdot \frac{1}{t - m^2}, \\ i\mathcal{M}_{\text{right}} &= (-ig)^2 \cdot \frac{1}{s - m^2}. \end{aligned} \tag{3.2.8}$$

Hence, they are referred to as t-channel and s-channel diagrams, respectively. An example of a u-channel diagram appears for nucleon-nucleon scattering in Figure **??**. Intuitively, s is the total energy in the COM frame squared, while t and u are a measure of how much momentum is exchanged between the scattered particles (see Appendix **??**).

Resonances

Note an important point about s−channel diagrams: the amplitude diverges as $s \to m^2$.[9] Or, in other words, as the COM energy approaches the mass of the exchanged particle (as long as $m > 2M$).

This divergence is interpreted as a *resonance* in the cross section (see below) of the scattering process as a function of $\sqrt{s}$, and allows us to discover new particles. Figure 3.3 shows a great example for $e^+e^- \to$ hadron scattering by a series of HEP experiments with a magnificent peak at 96 GeV, the Z boson mass.

The classical limit and the Yukawa potential

It is important to check our QFT recovers classical physics in the appropriate limit. It will also be useful to translate the somewhat abstract idea of amplitudes to the familiar concepts of forces and potentials. We will do so by considering the nonrelativistic limit ($|\mathbf{p}| \ll M$) of our above amplitudes and using the Born approximation relating the scattering amplitude between two particles to the

[9] We are saved from this potential infinity by a factor to be added to the denominator due to meson decay (Tong SM [71] Chapter 3.5).

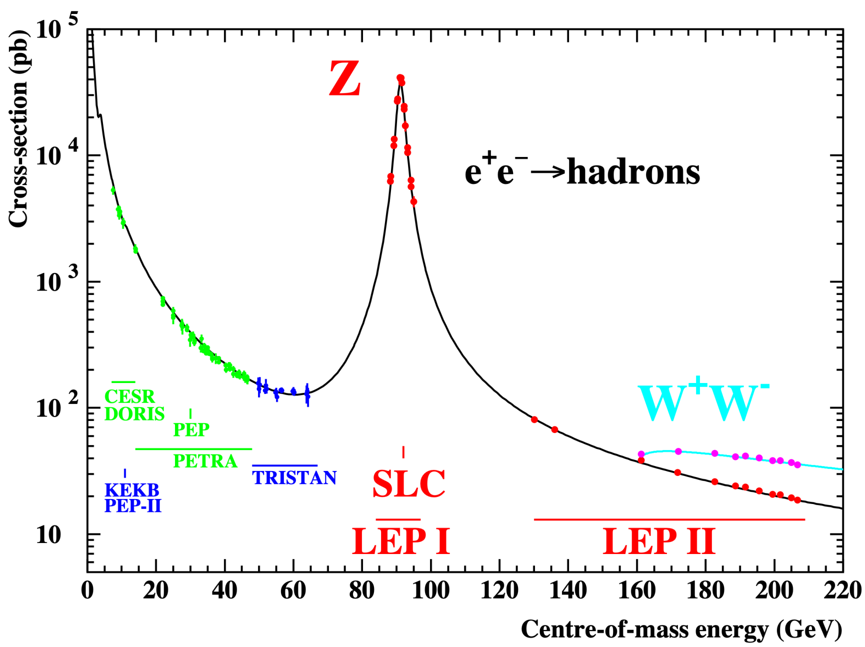


Figure 3.3. Cross section for $e^+e^- \to$ hadron scattering as a function of $\sqrt{s}$ with a clear resonance at the Z boson mass, reproduced from Ref. [2].

potential between them $U(\mathbf{r})$:

$$\mathcal{M} = \langle \mathbf{p}_f | U(\mathbf{r}) | \mathbf{p}_i \rangle = -i \int U(\mathbf{r}) e^{i(\mathbf{p}_f - \mathbf{p}_i)\cdot \mathbf{r}} d^3 r, \tag{3.2.9}$$

where $\mathbf{r}$ is the displacement between the particles.

First, let us consider what this potential would be classically. The static Klein-Gordon equation for a delta-function source:

$$(\nabla^2 - m^2)\phi(\mathbf{r}) = \delta^3(\mathbf{r}), \tag{3.2.10}$$

can be found via the Fourier transform to be:

$$\phi(\mathbf{r}) = \frac{e^{-mr}}{4\pi r}. \tag{3.2.11}$$

We can interpret this to be the profile of ϕ around a nucleon (the delta function source), and thus conversely the potential felt by another nucleon via the meson and the Yukawa interaction, under the assumption $M \gg m$. This is entirely analogous to gauge potential A_0 in electrostatics generated by a δ-function source acting as the electric potential for a test charge.

Going back to our amplitude for nucleon-antinucleon scattering, the s-channel diagram vanishes in the nonrelativistic limit (which essentially means it does not have a simple classical interpretation), while the t-channel diagram actually stays the same:

$$i\mathcal{M} = -(-ig)^2 \cdot \frac{1}{\left|\mathbf{p}_f - \mathbf{p}_i\right|^2 - m^2}. \tag{3.2.12}$$

Plugging this into the LHS of Eq. 3.2.9 and inverting the RHS integral gives us:

$$U(\mathbf{r}) = -\frac{g^2}{4M^2} \cdot \frac{e^{-mr}}{4\pi r}. \tag{3.2.13}$$

This is exactly the classical potential we found in Eq. 3.2.11! It is weighted by the coupling constant g and M to get the correct dimensions, and with a minus sign telling us potential is attractive.

Thus, we are able to reproduce Newtonian forces from the nonrelativisic limit of QFT. We also have the new interpretation of forces as simply manifestations of interactions in the Lagrangian, occurring through the exchange of virtual particles.

This potential is called the *Yukawa potential*, describing a force mediated by a massive boson. As expected, in the limit $m \to 0$, we recover the familiar $1/r$ Coulomb potential, which is mediated by the massless photon. We can check that we obtain the same potential for nucleon-nucleon scattering and, more generally, that all forces mediated by scalars are attractive. In fact, this is true for spin-2 particles as well, which is why gravity is universally attractive! On the other hand, forces mediated by spin-1 particles, such as EM, can be either attractive or repulsive, with the charges of the particles involved determining the sign of each diagram. See e.g. Zee QFT [75] Chapter I.5 for a useful discussion.

Fourth-order diagrams and loops

So far, we have only considered *tree-level* diagrams, the simplest to calculate. This is in contrast to diagrams with *loops*, which can occur at higher order in perturbation theory. For example, at fourth-order we can have diagrams like those in Figure 3.4 for nucleon scattering.

Such diagrams contribute integrals over the loop momentum k to the matrix element, which can notoriously diverge. To deal with this requires a process called *renormalization*, which, briefly, involves defining a cut-off energy scale Λ for these integrals, beyond which we claim the theory is invalid. Experimentally, the main consequence is that physical parameters like the mass of particles and coupling constants in fact depend on the energy scale at which they are measured!

3.2.4 Decay rates and cross sections

In this section, we translate our S-matrix elements to physical observables: cross sections and decay rates.

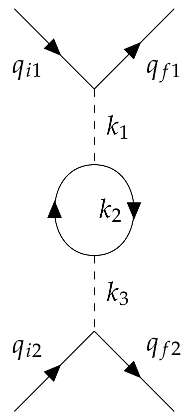


Figure 3.4. An example of a higher-order scattering diagram with a "loop".

Cross section

Classically for a scattering experiment, the number of particles scattered N is related to the cross sectional area σ as:

$$N = \sigma T \Phi, \tag{3.2.14}$$

where T is the total time and Φ is the flux of incoming particles (number of incoming particles per unit area and unit time). In QM, we define the cross section σ similarly, but in terms of the probability of scattering P instead of N:

$$\sigma = \frac{P}{\Phi T}. \tag{3.2.15}$$

This is a more abstract quantity in QM, but it still has units of area. The number of scattering events N is related to σ by a factor we call the *luminosity* L:

$$N = \sigma L. \tag{3.2.16}$$

Here, we simply consider this the definition of luminosity, but for a collider, for example, it can be derived from the properties of the input particle beams (as will be discussed in Part II). Often, we are interested in the *differential cross section* $d\sigma$ with respect to kinematic variables like the solid angle Ω or energy, so we write:

$$d\sigma = \frac{dP}{\Phi T}. \tag{3.2.17}$$

As in QM, this probability P is proportional to the square of the amplitude $|\langle f|S|i\rangle|^2$:

$$dP = \frac{\left|\langle f|S|i\rangle\right|^2}{\langle f|f\rangle \langle i|i\rangle} d\Pi, \tag{3.2.18}$$

where $\langle f|f\rangle$ and $\langle i|i\rangle$ are the normalization factors for the final and initial states (they are not equal to 1 as discussed in Section 3.1.4), and $d\Pi$ is the differential region of final state momenta.

For the case of two incoming particles (which is what is most relevant for this dissertation), we can put all of this together to obtain the relation between differential cross section and the matrix element $\mathcal{M}$:

$$d\sigma = \frac{1}{(2E_1)(2E_2)\,|\mathbf{v}_1 - \mathbf{v}_2|}\,|\mathcal{M}|^2\,d\Pi_{\text{LIPS}}, \tag{3.2.19}$$

where E_1 and E_2 are the energies of the incoming particles, $\mathbf{v}_1$ and $\mathbf{v}_2$ are their velocities, and $d\Pi_{\text{LIPS}}$ is called the Lorentz-invariant phase space of the final state momenta:

$$d\Pi_{\text{LIPS}} = (2\pi)^4\delta^{(4)}(\Sigma p)\prod_{\text{final states } j}\frac{d^3p_j}{(2\pi)^3}\frac{1}{2E_j} \tag{3.2.20}$$

For the case of $2 \to 2$ scattering, in the COM frame, this simplifies considerably:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{CM}} = \frac{1}{64\pi^2 E_{\text{CM}}^2}\frac{|\mathbf{p}_f|}{|\mathbf{p}_i|}\,|\mathcal{M}|^2\,\theta(E_{\text{CM}} - m_3 - m_4), \tag{3.2.21}$$

and even more so when the all four masses are equal:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{CM}} = \frac{1}{64\pi^2 E_{\text{CM}}^2}\,|\mathcal{M}|^2\,. \tag{3.2.22}$$

Decay rate

The other type of process we are interested in are decays. The decay rate Γ is simply the probability of decay per unit time:

$$\Gamma = \frac{P}{T}. \tag{3.2.23}$$

Using our expression for P from above and simplifying, we find:

$$d\Gamma = \frac{1}{2m} |\mathcal{M}|^2 \, d\Pi_{\text{LIPS}}, \tag{3.2.24}$$

in the rest frame of the decaying particle, where m is its mass. If multiple decays of the same particle are possible, we sum over the final states in the phase space integral. The total Γ is then called the *width* of the particle, and $1/\Gamma \equiv \tau$ is its half-life.

For our simple meson decay $\phi \to \psi^\dagger \psi$, we have at tree level:

$$d\Gamma = \frac{g^2}{2m} d\Pi_{\text{LIPS}} \quad \Rightarrow \quad \Gamma = \frac{g^2}{32\pi m} \left(1 - \frac{4M^2}{m^2}\right)^{1/2}, \tag{3.2.25}$$

where we performed the integral over $d\Pi_{\text{LIPS}}$ (see Ref. [96] 4.2). This is in fact not too far off the expression for the decay width of the Higgs boson to fermions. What we are missing of course is that fermions are spin-1/2 particles, and we need to sum over their spin states. Fermions are described by spinor field theory, detailed in Appendix **??**.

3.3 Gauge theories

Nature seems to take advantage of the simple mathematical representations of the symmetry laws. When one pauses to consider the elegance and the beautiful perfection of the mathematical reasoning involved and contrast it with the complex and far-reaching physical consequences, a deep sense of respect for the power of the symmetry laws never fails to develop. — C. N. Yang

So far, we have discussed spin-0 scalar bosons (and the spin-$\frac{1}{2}$ fermions in Appendix **??**); the last set of SM particles are the spin-1 *gauge bosons*. These are the particles which mediate all three fundamental forces in the SM: electromagnetism, the weak force, and the strong force. Fortunately, compared to spinors, they live in the simpler and familiar vector representation of the Lorentz group.

On the other hand, they are intrinsically tied to a unique type of internal, *local*, symmetry in QFT: *gauge symmetry*. Unlike, say, Lorentz or spacetime translation invariance, this is not a fundamental physical symmetry of nature, and is not associated with any consrervation law. Instead, it simply describes a redundancy in our mathematical formulation of the gauge theory, stemming from the fact that the vector fields used to describe the gauge bosons have more degrees of freedom (DoFs) than the physical particles themselves. The DoFs are thereby reduced by identifying fields related by a gauge symmetry transformation to be the same physical state, known as the principle of *gauge invariance*. This is entirely analogous to requiring that a change of coordinate system not affect the physics. A deeper discussion of the motivations behind gauge invariance can be found in Appendix **??**.

In this section, we first introduce the simplest gauge boson, the photon, and its associated U(1) gauge symmetry in Section 3.3.1. Coupling this to matter and quantizing the theory gives us QED, the relativistic quantum theory of electromagnetism (Section 3.3.2). We then generalize this to, and quantize, non-abelian gauge theories, known as Yang-Mills theories, in Sections 3.3.3 and 3.3.3, respectively. We conclude with a discussion of renormalization and the *running* of coupling constants in Section 3.3.4.

3.3.1 Maxwell Theory

Gauge symmetries are a generalization of internal global symmetries, such as the U(1) symmetry from Section 3.1.2, to a *local* symmetry, where the symmetry transformation can be a function of spacetime. We are most familiar with this concept from classical E&M, in which Maxwell's laws are invariant under transformations of the 4-vector potential $A_\mu = (\phi, \mathbf{A})$ of the form:

$$A_\mu(x) \to A_\mu(x) + \frac{1}{e}\partial_\mu \alpha(x), \tag{3.3.1}$$

for an arbitrary function $\alpha(x)$, where e is a conventional constant that we will soon interpret as the coupling constant of the theory.

Recall that A_μ is related to the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$, by:

$$\mathbf{E} = -\nabla\phi - \partial_t \mathbf{A}, \qquad \mathbf{B} = \nabla \times \mathbf{A}, \tag{3.3.2}$$

and the Maxwell equations can be derived from the Lagrangian:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{3.3.3}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{3.3.4}$$

is the *field strength* tensor. One can confirm that (1) $F_{\mu\nu}$ and, hence, the Lagrangian is invariant under the gauge transformation in Eq. 3.3.1, and (2) the resulting E-L EOMs are exactly the homogeneous Maxwell equations. Thus, classical E&M was our earliest and simplest gauge theory, although the significance and generalization of gauge invariance only became clear with the advent of QFT.

Gauge invariance significantly restricts the possible terms in the Lagrangian (and thus considerably simplifies the theory). Notably, mass terms like $m^2 A_\mu^2$ *violate* gauge invariance, which is why gauge bosons are necessarily massless, without something special like the Higgs mechanism (Section 3.4). As discussed in Appendix **??**, gauge invariance also ensures the renormalizability of the theory and reduces the DoFs of A_μ such that, once quantized, we can identify it as the photonic field.

Interactions with scalars

The U(1) nature of the gauge transformation becomes more apparent when we try to couple the photon to other particles. Note that our Lagrangian above contains terms of the form $(\partial_\mu A_\nu)^2$ so A_μ (and indeed all spin-1 fields) have dimension 1.

Let us consider a scalar field ϕ: we can write renormalizable, scalar terms like $A_\mu^2\phi^2$ and $A_\mu\phi\partial_\mu\phi$; however, they do not look gauge invariant. To make them so, we must require that ϕ *also* transforms under the same gauge transformation in a way that compensates the change in A_μ.

The simplest way is to take ϕ to be a *complex* scalar field and "promote" its inherent global U(1) symmetry to a local one:

$$\phi(x) \to e^{iQ_\phi \alpha(x)}\phi(x), \tag{3.3.5}$$

where we say Q_ϕ represents the charge of ϕ under the U(1) symmetry.[10] We can

[10]Note that such a transformation is not possible with a *real* field, which necessarily has 0 charge and does not couple with the photon.

then define the *covariant derivative* acting on ϕ as:[11]

$$D_\mu \phi = (\partial_\mu - ieQ_\phi A_\mu)\phi, \tag{3.3.6}$$

where e is the same coupling constant from Eq. 3.3.1.

One can check that $D_\mu \phi$ transforms under the gauge transformation as:

$$D_\mu \phi \longrightarrow e^{iQ_\phi \alpha(x)} D_\mu \phi, \tag{3.3.7}$$

meaning $(D_\mu \phi)^\dagger D^\mu \phi$ provides us with a gauge invariant interaction term for the Lagrangian. Thus, we have a gauge invariant *scalar* QED Lagrangian:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (D_\mu \phi)^\dagger D^\mu \phi - m^2 \left|\phi\right|^2 . \tag{3.3.8}$$

Note that the commutator of the covariant derivative is in fact not a derivative at all, but proportional to the field strength tensor:

$$[D_\mu, D_\nu]\phi = ([\partial_\mu, \partial_\nu] - ie[\partial_\mu, A_\nu] + ie[\partial_\nu, A_\mu])\phi = -ieF_{\mu\nu}\phi. \tag{3.3.9}$$

Thus, we can define $F_{\mu\nu} \equiv \frac{i}{e}[D_\mu, D_\nu]$, which will prove useful for non-abelian gauge symmetries later in this section.

[11]As discussed above, this is the same concept as the covariant derivative in GR, with the gauge field A_μ acting as a connection on a U(1) fiber bundle analogously to the Levi-Civita connection between tangent bundles. Essentially, it encodes the change in the local phase of ϕ across spacetime (see Peskin and Shroeder [76] Chapter 15.1 for a nice derivation of this).

Generally, we choose the normalization $Q_e = -1$ for the electron field, so e becomes our familiar elementary charge (in natural units) and $\alpha \equiv e^2/4\pi \approx 1/137$ is the famous dimensionless fine structure constant.[12]

Interactions with spinors

The case for spinors is not so different. The definition of the covariant derivative remains the same, so combining the "covariant" Dirac Lagrangian with the free photonic yields the QED Lagrangian:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi. \tag{3.3.10}$$

This is in fact the most general possible Lorentz-invariant, renormalizable, P-symmetric Lagrangian for a spinor field with a U(1) gauge symmetry, and can thus be derived from the requirement of gauge invariance alone (as done in e.g. Peskin and Shroeder [76] Chapter 15.1). This is a general feature of the SM: every possible term permitted by gauge invariance and the usual physical requirements of Lorentz invariance etc. is included in the Lagrangian (with one possible exception that forms the basis for the strong CP problem [97, 98]).

[12]Technically, this value varies with our energy scale, as we will discuss in Section 3.3.4, and $1/137$ is its asymptotic value at low energies.

Expanding out the Lagrangian, we have:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - e\bar{\psi}\gamma^\mu\psi A_\mu, \tag{3.3.11}$$

where we see this interaction term is simply $-ej^\mu A_\mu$ with $j^\mu = \bar{\psi}\gamma^\mu\psi$ the conserved current associated with the *global* U(1) symmetry we found in Section **??**. One can check that the E-L EOMs for A_μ now correspond to the *inhomogeneous* Maxwell equations with a source term $J_\mu \equiv -ej_\mu$:

$$\partial_\mu F_{\mu\nu} = J_\nu, \tag{3.3.12}$$

reproducing our beloved E&M from this field theory formulation!

3.3.2 Quantum electrodynamics

The quantized version of the above is what we call *quantum electrodynamics* (QED): the QFT of electromagnetic interactions. It has proven an extraordinarily successful theory, serving as a model for the remainder of the SM as well as theories for condensed matter phenomena.

The exact path to quantizing A_μ depends on the choice of gauge. We will forego those details and simply use physical intuition — namely, that the photon has only two physical, transverse polarizations — to motivate the result:

$$A_\mu(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} \sum_{\lambda=1}^{2} \left(\epsilon_\mu^\lambda(p) \hat{a}_p^\lambda e^{-ip\cdot x} + \epsilon_\mu^{\lambda *}(p) \hat{a}_p^{\lambda\dagger} e^{ip\cdot x} \right), \tag{3.3.13}$$

where $\epsilon_\mu^\lambda(p)$ are the two transverse polarization basis vectors and a_p^λ and $a_p^{\lambda\dagger}$ are the photon annihilation and creation operators.

The photon propagator depends as well on the choice of gauge. Expanding the homogeneous photon EOM, Eq. 3.3.12, gives:

$$\partial_\mu \partial^\mu A_\nu - \partial_\nu \partial_\mu A^\mu = J_\nu, \tag{3.3.14}$$

which in momentum space becomes:

$$(-p^2 \eta_{\mu\nu} + p_\mu p_\nu) A_\mu = J_\nu. \tag{3.3.15}$$

Recall that the propagator is the inverse of the operator on the LHS for a delta-function source; however, due to the redundant DoFs of A_μ, this is not directly invertible without first fixing a gauge.

The cleanest way to do so is to add a "Lagrange multiplier" term representing the gauge fixing condition to the Lagrangian. The most common choice is the *Lorenz gauge*, $\partial_\mu A^\mu = 0$, which makes Lorentz-invariance manifest and to enforce which we can include the term $-\frac{1}{2\xi}(\partial_\mu A^\mu)^2$. One can confirm that the EOM for ξ is exactly the Lorenz gauge condition. Inverting the new EOM for A_μ

gives us the (Feynman) photon propagator:

$$\Delta_{\mu\nu}(p) = \frac{-i}{p^2 + i\epsilon}\left[\eta_{\mu\nu} + (1-\xi)\frac{p_\mu p_\nu}{p^2}\right]. \tag{3.3.16}$$

This is called the R_ξ gauge and different values of ξ correspond to different propagators, each with their own advantages and disadvantages for calculations. In QED, we typically take $\xi = 1$, called the Feynman-'t Hooft gauge, for simplicity:

$$\Delta_{\mu\nu}(p) = \frac{-i\eta_{\mu\nu}}{p^2 + i\epsilon}. \tag{3.3.17}$$

Definition 3.3.1. With this, we can write down the Feynman rules for QED, with spinor (α, β) and 4-vector (μ, ν) indices labeled explicitly for clarity:

These Feynman rules can be applied to simple tree-level processes similarly to Yukawa theory (see Sections 3.2.3 and **??**). These include several important processes such as electron-electron scattering $e^-e^- \to e^-e^-$ via a virtual photon, Compton scattering $\gamma e^- \to \gamma e^-$, electron-positron annihilation $e^+e^- \to \gamma\gamma$, and electron-positron (or Bhabha) scattering $e^+e^- \to e^+e^-$. The former (and its variations with other charged particles) is what we generally experience as electromagnetism, and can recover the Coloumb potential in the non-relativistic limit.

3.3.3 Yang-Mills Theory

Following the remarkable success of QED and GR, a generalization of such gauge theories to *non-abelian* symmetries was proposed by Chen Ning Yang and Robert Mills in 1953 [99], today referred to as *Yang-Mills* theories. These theories picked up steam in the 1960s when the concept of spontaneous symmetry breaking was developed to give mass to the gauge bosons (Section 3.4) and it was realized that both the weak and strong interactions can be described by SU(2) and SU(3) Yang-Mills theories, respectively. They are hence a cornerstone of the SM, and we will now briefly outline their construction, generalizing the U(1) gauge symmetry from the previous section.

Non-abelian gauge transformations

In Yang-Mills theory, we allow non-gauge fields to transform locally under *any* Lie group G, in an arbitrary representation R of the group (generally, in the SM, R is either the fundamental or trivial representation). This means the fields ψ are actually vectors of $\dim(R)$ (on top of their usual spinor or 4-vector indices etc.), and transform as:

$$\psi(x) \to \psi'(x) = e^{i\alpha^a(x)T^a_R}\psi(x) \equiv V(x)\psi(x), \tag{3.3.18}$$

where T_R^a are the generators of G in the representation R and $V(x) = e^{i\alpha^a(x)T_R^a}$ is the gauge transformation. To construct a G-invariant Lagrangian, we again need to define a covariant derivative with gauge fields A_μ^a connecting the local transformations of ψ across spacetime:

$$D_\mu\psi = (\partial_\mu - igA_\mu^a T_R^a)\psi. \tag{3.3.19}$$

Observe that we must have as many gauge fields as there are group generators to counter all possible gauge transformations $V(x)$; i.e., there are $\dim(G)$ A_μs, living in the *adjoint* representation of G (see Chapter 2.2). The gauge field is often represented more conveniently as a "Lie-algebra-valued" field (i.e., as an object in the Lie algebra):

$$A_\mu \equiv A_\mu^a T^a. \tag{3.3.20}$$

We can derive how A_μ transforms by requiring the covariant derivative to transform identically to ψ (the same as in the abelian case):[13]

$$D_\mu\psi \to D'_\mu\psi = (\partial_\mu - igA'_\mu)V\psi \stackrel{!}{=} VD_\mu\psi(x), \tag{3.3.21}$$

where g is the coupling constant. One can check this is satisfied for the trans-

[13]For a detailed derivation see e.g. Ricardo Matheus' QFT Lectures [78] Part 34.

formed gauge field:

$$A'_\mu = V A_\mu V^{-1} - \frac{i}{g}(\partial_\mu V)V^{-1}. \tag{3.3.22}$$

For infinitesimal gauge transformations $V \simeq 1 + i\alpha^a T^a_R$, this can be written in terms of the components as:

$$A'^a_\mu T^a = A^a_\mu T^a + \frac{1}{g}\partial_\mu \alpha^a T^a + iA^a_\mu \alpha^b [T^b, T^a] = A^a_\mu T^a + \frac{1}{g}\partial_\mu \alpha^a T^a - f^{abc} A^a_\mu \alpha^b T^c, \tag{3.3.23}$$

where f^{abc} are the structure constants of the Lie algebra of G. The second term represents the gauge transformation, same as in the abelian case, while the third term is new and is the transformation property for a field in the adjoint representation.

The field strength tensor

The final piece we need for the Lagrangian is a gauge-invariant kinetic term for the gauge fields, generalizing the electromagnetic field strength tensor $F_{\mu\nu}$. We can construct this, as in the abelian case, using the commutator of covariant derivatives:

$$F_{\mu\nu} \equiv \frac{i}{g}[D_\mu, D_\nu] = (\partial_\mu A^a_\nu - \partial_\nu A^a_\mu) - ig[A_\mu, A_\nu]. \tag{3.3.24}$$

Again, this reduces to the E&M tensor for an abelian symmetry, where the commutator term is 0. In the non-abelian case, the commutator term adds new *self-interaction* terms to the gauge fields. One can check that $F_{\mu\nu}$ transforms as:

$$F_{\mu\nu} \to V F_{\mu\nu} V^{-1}, \tag{3.3.25}$$

or, infinitesimally, in terms of components as:

$$F^a_{\mu\nu} T^a \to F^a_{\mu\nu} T^a + f^{abc} F^b_{\mu\nu} \alpha^c T^a, \tag{3.3.26}$$

which we can recognize as the transformation of a field in the adjoint representation (Eq. 3.3.23 without the gauge transformation term).

Clearly, for non-abelian theories, the field-strength tensor alone, or even $F_{\mu\nu}F^{\mu\nu}$, is no longer gauge-invariant; however, its *trace* is:

$$\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] \to \mathrm{Tr}\left[V F_{\mu\nu} V^{-1} V F^{\mu\nu} V^{-1}\right] = \mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] \tag{3.3.27}$$

using the cyclic property of the trace, providing us with a gauge-invariant kinetic term for the gauge fields. In terms of components, this is:

$$\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] = F^a_{\mu\nu} F^{a\mu\nu}\, \mathrm{Tr}\left[T^a T^a\right] \tag{3.3.28}$$

The value of $\mathrm{Tr}\left[T^a T^a\right]$ is a normalization constant that is conventionally chosen to be $\frac{1}{2}$ for the fundamental representation. Expanding out $(F^a_{\mu\nu})^2$ gives us cubic and quartic self-interaction terms for the gauge fields.

The Yang-Mills Lagrangian

Combining all of the above, we have the gauge-invariant Yang-Mills Lagrangian:

$$\mathcal{L} = -\frac{1}{2}\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] + \bar{\psi}(i\not{D} - m)\psi, \tag{3.3.29}$$

or, in explicit, component form:

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \bar{\psi}_i[\delta_{ij}(i\not{\partial}_\mu - m) + g\not{A}^a T^a_{ij}]\psi_j, \tag{3.3.30}$$

where the indices i and j are running over the fermion fields in the representation R. Note again that a mass term $m^2 A^a_\mu A^{a\mu}$ would violate gauge invariance without the Higgs mechanism.

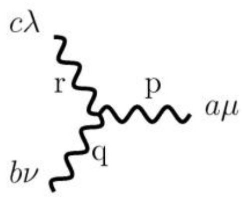
Interestingly, despite the extra self-interaction terms, there remains only one free parameter in the theory: the coupling constant g. This is why the SM, despite its apparent complexity, has so few free parameters, particularly in the "gauge sector" (the majority of free parameters are related to couplings in the *Higgs* sector). It is also worth emphasizing that the primary difference *physically* between abelian and non-abelian gauge theories is that the gauge bosons are charged under the gauge group in the latter (and, hence, self-interact).

Quantum Yang-Mills Theory

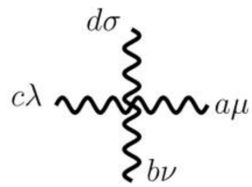
The form of the quantized gauge fields in Yang-Mills are similar to the U(1) case, except now with the extra adjoint representation indices. The process of quantization and deriving the propagator, however, is considerably more involved for non-abelian theories. The core idea of adding an R_ξ gauge-fixing term to the Lagrangian is similar, but due to the gauge fields' non-trivial transformation property, the proper treatment necessitates the introduction of imaginary internal particles called *Faddeev-Popov ghosts* to cancel gauge-dependent terms. Somewhat similar to virtual particles, these ghosts are purely mathematical artifacts required to maintain gauge- and Lorentz-invariance of the quantized theory. The full details of this process can be found in e.g. Peskin and Shroeder [76] Chapter 16; the upshot is simply some extra Feynman rules involving ghost particles in the theory.

The new Feynman rules for *non-abelian Yang-Mills* theories are shown in Figure 3.5. The gauge bosons are conventionally referred to as "gluons" but these rules are general. Note the cubic and quartic gauge boson vertices, as well as the ghost particle (c) diagrams, unique to non-abelian theories. The phenomenology of Yang-Mills theories in the SM will be discussed in the next chapter.

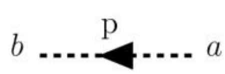
gluon propagator: $$D^{ab}_{\mu\nu}(p) = \frac{-i\delta^{ab}}{p^2+i0}\left[\eta_{\mu\nu} - \frac{(1-\xi)p_\mu p_\nu}{p^2+i0}\right]$$



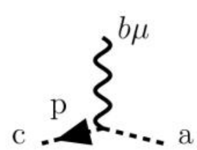
3–gluon vertex: $$\Gamma^{abc}_{\mu\nu\lambda}(p,q,r) = -gf^{abc}[(p-q)_\lambda\eta_{\mu\nu} + (q-r)_\mu\eta_{\nu\lambda} + (r-p)_\nu\eta_{\mu\lambda}]$$



4–gluon vertex: $$\Gamma^{abcd}_{\mu\nu\lambda\sigma} = -ig^2 f^{abe}f^{cde}(\eta_{\mu\lambda}\eta_{\nu\sigma} - \eta_{\mu\sigma}\eta_{\nu\lambda}) -ig^2 f^{ace}f^{bde}(\eta_{\mu\nu}\eta_{\sigma\lambda} - \eta_{\mu\sigma}\eta_{\nu\lambda}) -ig^2 f^{ade}f^{bce}(\eta_{\mu\nu}\eta_{\sigma\lambda} - \eta_{\mu\lambda}\eta_{\nu\sigma})$$



ghost propagator: $$C^{ab}(p) = \frac{i\delta^{ab}}{p^2+i0}$$



$\bar{c}cg$ – vertex: $$\Gamma^{abc}(p) = gf^{abc}p_\mu$$

Figure 3.5. Feynman rules unique to non-abelian Yang-Mills theories, reproduced from Ref. [3].

3.3.4 Running couplings and asymptotic freedom

As discussed briefly in Section 3.2, in order to handle divergences from higher order "loop" diagrams in perturbation theory, a class of mathematical techniques called *renormalization* is employed. A perhaps surprising physical consequence of this is that parameters of the theory are dependent on the energy scale at which they are probed. Their dependence is described the *renormalization group equations* or *flow*.

The renormalization group is an extremely deep subject with applications in many areas of physics. The most relevant result for us is the *running* of the coupling constants in gauge theories — i.e., the strength of the corresponding forces as a function of the energy scale. This is shown for the relevant U(1), SU(2), and SU(3) gauge symmetries of the SM in Figure 3.6.

We see firstly that the electromagnetic interaction strength increases with energy scale. Physically, this is understood through the *vacuum polarization* via virtual electron-positron pair creation, which "screen" the electric charges of real particles more effectively at longer distances, thereby weakening the force.[14]

A notable, Nobel-prize winning, 1973 result of Frank Wilczek, David Gross, and David Politzer, however, was an inverse dependence on energy for

[14]Interestingly, QED has a Landau pole: a finite value of the energy scale for which the interaction strength is infinite. However, this value is so high (10^{286} GeV) as to have no practical consequence, and likely points to the breakdown of perturbation theory, that is used to derive the running coupling, at such a scale.

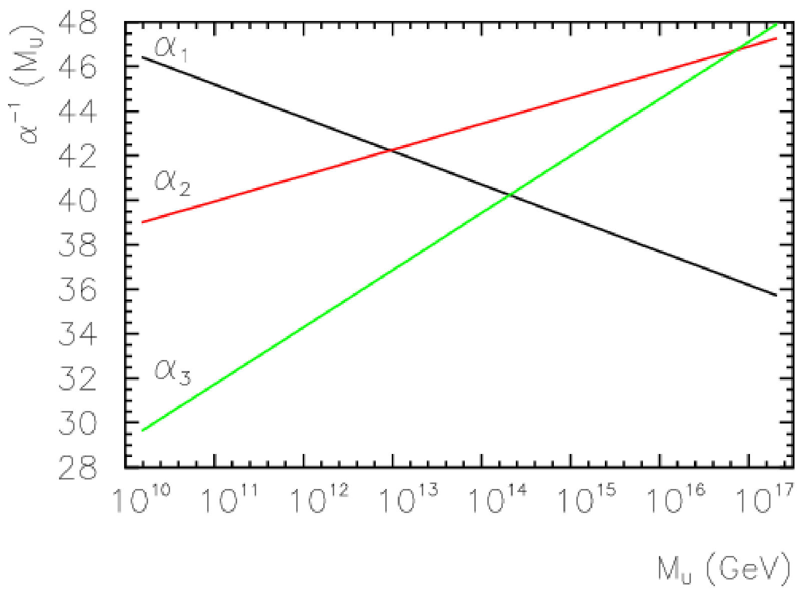


Figure 3.6. The running of the inverse strength of the SM coupling constants, with the strong coupling constant (SU(3)) in green, weak (SU(2)) in red, and electromagnetic (U(1)) in black, reproduced from Ref. [4].

non-abelian gauge theories [100, 101], as shown for the SU(2) and SU(3) couplings in Figure 3.6.[15] This phenomenon is called *asymptotic freedom*, as in the high energy limit the theory is effectively one of free particles. It is a notable feature of the strong force, as will be discussed in Chapter 4.1.

[15]Technically, this depends on the gauge group and the number of fermions in the theory; for both the weak and strong forces, this number is sufficiently small (see e.g. Peskin and Shroeder [76] Chapter 16).

3.4 The ABEGHHK (Higgs) mechanism

As highlighted in the previous section, gauge bosons in pure Yang-Mills theories are massless. This is in conflict, however, with the short observed range of the weak force, implying massive mediatory bosons. To resolve this, a series of work in the early 1960s by Anderson, Brout, Englert, Guralnik, Hagen, Higgs, and Kibble (ABEGHHK) yielded a mechanism to give mass to the gauge bosons without violating gauge invariance [102–105], based on the concept of *spontaneous symmetry breaking* developed by Nambu [106, 107] and others.

By 1970, Glashow, Salam, Weinberg and others were able to use this mechanism to formulate a combined theory of weak and electromagnetic interactions, known as "electroweak" or Weinberg-Salam theory [108–110]. Electroweak unification has been one of the most significant breakthroughs in theoretical physics with several Nobel prizes cumulatively awarded for these developments.

In this section we outline the ABEGHHK mechanism — commonly (but reductively) referred to as the "Higgs mechanism" — first for an abelian gauge theory in Section 3.4.1 and then for non-abelian gauge theories 3.4.2 like the SM.

3.4.1 The abelian Higgs mechanism

The Higgs mechanism is based on the idea of spontaneous symmetry breaking (SSB), where the ground states of a physical system violate the overall symmetry. The classic example is the so-called "sombrero" potential for a complex scalar field ϕ:

$$V(\phi) = -\frac{\lambda}{2}(|\phi|^2 - v^2)^2, \tag{3.4.1}$$

for constants λ and v, shown in Figure 3.7. The potential has is symmetric under a U(1) transformation of $\phi \to e^{i\alpha}\phi$, but any specific ground state of $|\phi| = v$ will break this symmetry, as shown in the figure. SSB is a crucial concept in physics, with several applications in condensed matter and particle physics, including chiral symmetry breaking in QCD (see e.g. Tong SM [71] Chapter 3.2).

The Higgs mechanism is an application of SSB to *gauge* symmetries. The interpretation here of SSB a bit fiddly since, as emphasized above, gauge symmetries are not physical and cannot be spontaneously broken;[16] what actually breaks is the corresponding *global* symmetry, as we outline below.

Consider our QED Lagrangian for a complex scalar field ϕ with the above

[16]This is an implication of Elitzur's theorem [111].

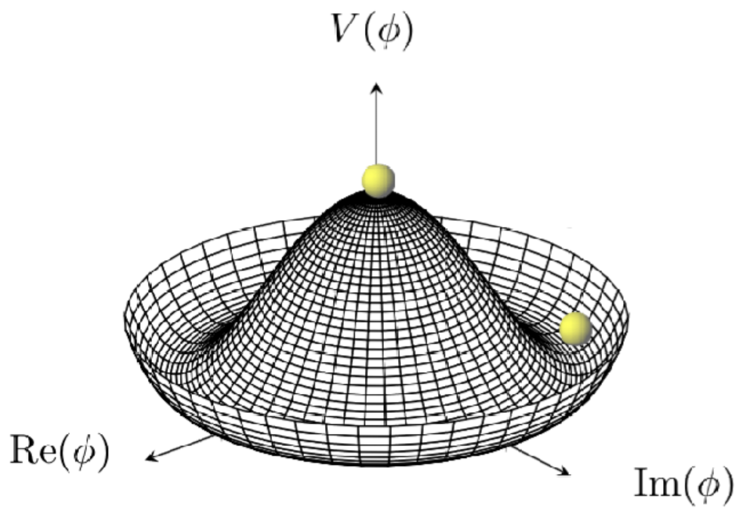


Figure 3.7. The "sombrero" potential for the Higgs field, reproduced from Ref. [5]. An initial state and a ground state breaking the U(1) symmetry are represented by the green balls at the top and bottom of the potential, respectively.

potential:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger D^\mu\phi + \frac{\lambda}{2}(|\phi|^2 - v^2)^2. \tag{3.4.2}$$

As before, this Lagrangian possesses a U(1) gauge symmetry; however, this symmetry is "broken" by a particular ground state $\phi = ve^{i\delta}$ (we can take $\delta = 0$ WLOG). The fluctuations around the ground state can be parametrized as:

$$\phi(x) = (v + \sigma(x))e^{i\theta(x)}, \tag{3.4.3}$$

where σ and θ are two real fields. Plugging this into the Lagrangian gives us:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \partial_\mu\sigma\partial^\mu\sigma + (v+\sigma)^2(\partial_\mu\theta - eA_\mu)(\partial^\mu\theta - eA^\mu) - \lambda(2v^2\sigma^2 + 2v\sigma^3 + \frac{\sigma^4}{4}). \tag{3.4.4}$$

We see first that σ can be interpreted as a normal scalar quantum field, with a quadratic mass term with $m_\sigma^2 = 2\lambda v^2$. The θ term is a bit more unusual;[17] it only appears in the combination $\partial_\mu\theta - eA_\mu$. Hence, we can simply redefine the gauge field as $A'_\mu \equiv A_\mu + \frac{1}{e}\partial_\mu\theta$, allowing it to "absorb" this DoF. Note that this takes the form of a gauge transformation of A_μ and thus does not affect the field strength tensor $F_{\mu\nu}$. The resulting Lagrangian is then:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \partial_\mu\sigma\partial^\mu\sigma + e^2(v+\sigma)^2A'_\mu A'^\mu - \lambda(2v^2\sigma^2 + 2v\sigma^3 + \frac{\sigma^4}{4}), \tag{3.4.5}$$

where we now have a mass term for the "gauge boson", $m_A^2 = 2e^2v^2$!

3.4.2 The non-abelian Higgs mechanism

There is an analogous mechanism for a non-abelian gauge symmetry, as in the SM. One crucial difference is that the symmetry may only partially break from the gauge group G to a subgroup H (for example from SU(2) to a U(1)). In this case, the gauge bosons corresponding to the generators of G's broken symmetries

[17] In a non-gauge-theory, the θ field would be considered a massless "Goldstone boson" resulting from the spontaneously breakdown of the symmetry.

acquire mass as above, while the generators of H remain massless Goldstone bosons; as we will see in Chapter 4.2, in the SM these correspond to the massive $W^{\pm}$ / Z bosons and the massless photon, respectively. See e.g. Tong SM [71] Chapter 2.3.3 for an example.

Chapter 4

The Standard Model of Particle Physics

Before I came here, I was confused about this subject. Having listened to your lecture, I am still confused. But on a higher level. — Enrico Fermi[1]

We are now ready to describe the standard model (SM)! It is a renormalizable, quantum Yang-Mills theory, and is illustrated nicely in Figure 4.1. Before electroweak symmetry breaking (EWSB) — a form of spontaneous symmetry breaking (SSB) — it possessed the gauge symmetry $\mathrm{SU(3)_C \times SU(2)_L \times U(1)_Y}$, with C, L, and Y standing for color, left, and hypercharge, respectively. These three groups correspond to the strong, weak, and electromagnetic forces, with eight, three, and one generators or gauge bosons, respectively. The relative strengths of each interaction, as well as gravity's, are shown in Table 4.1, based on the equivalent of the fine structure constant of each force, $\alpha_f = g_f/4\pi$, where g_f is the

[1] Also an accurate representation of my understanding of the SM before and after my PhD.

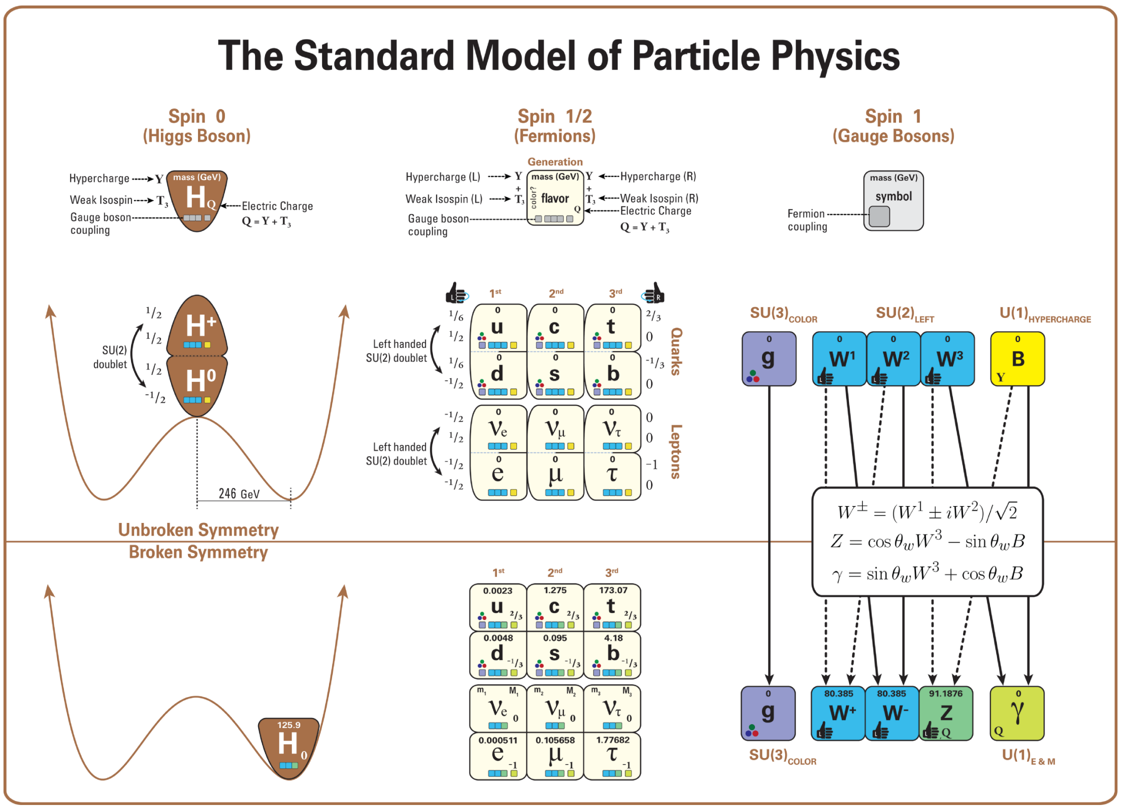


Figure 4.1. A graphical summary of the SM, reproduced from Ref. [6].

respective coupling constant.

The SM contains six fermions charged and uncharged under the $SU(3)_C$ symmetry each, called the "quarks" and "leptons", respectively. The left-handed fermions live as pairs in $SU(2)_L$ doublets, while the right-handed fermions in singlets. The six types of fermions are referred to as different "flavors", grouped into three generations as in Figure 4.1.

The SM also contains a complex scalar $SU(2)_Y$-doublet called the Higgs field, which is associated with EWSB. As shown in Figure 4.1, it initially is at the center of a "sombrero" potential because of which, before EWSB, the gauge bosons, fermions, and the Higgs field are all massless.

EWSB is hypothesized to have occurred during the electroweak epoch (see Figure 1), where the $SU(2)_L \times U(1)_Y$ global symmetry broke to the $U(1)_{EM}$ of QED. Through this, the Higgs field obtained a non-zero vacuum expectation value (VEV), imbuing all the fermions, three of the gauge bosons, and the Higgs boson with mass — a process referred to as the Higgs mechanism. The outcome is the state of the universe and physics as we know it.

Of course, as outlined in the introduction, this picture does not explain myriad phenomena in fundamental physics, including dark matter, dark energy, baryon asymmetry, and neutrino masses. This is why it is crucial to test the SM as rigorously and in as broad a phase space as possible, in order to identify any cracks that may point to new physics.

Table 4.1. Approximate magnitude of the strengths of the four fundamental forces at an energy scale of around 100 MeV.

Force	**Strength**
Electromagnetic	$\alpha_{\mathrm{EM}} \approx \frac{1}{137}$
Weak	$\alpha_W \approx \frac{1}{30}$
Strong	$\alpha_s \approx 1$
Gravity	$\alpha_G \approx 10^{-38}$

In this chapter, we will briefly walk through different areas of the SM. Having discussed QED in Chapter 3.3, we begin with the remaining two fundamental interactions: quantum chromodynamics (Section 4.1); and weak interactions and electroweak unification (Section 4.2). Finally, in Section 4.3, we will discuss the Higgs sector and pair production of the Higgs boson, which is the focus of this dissertation.

4.1 Quantum chromodynamics

Quantum chromodynamics (QCD) is a quantum Yang-Mills field theory describing the strong force, with the gauge group SU(3). SU(3) has eight generators and, hence, eight gauge bosons (G_μ) called gluons. The only other elementary particles which interact with the strong force — i.e., which don't live in the trivial representation of SU(3) — are the quarks. They live in the three-dimensional fundamental representation and thus possess three extra DoFs beyond vanilla spinors, which we call their "color" (hence, quantum *chromo*dynamics). The three orthogonal eigenstates in this representation are colloquially referred to as labeled red, green, and blue, and mathematically the quark fields (q_α) labeled with extra color indices $i = 1, 2, 3$.

Putting this together, the QCD Lagrangian, with all the indices labeled explicitly is:

$$\mathcal{L} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \sum_{f=1}^{6} \bar{q}_{\alpha f i} [\delta_{ij} (i\partial\!\!\!/_{\alpha\beta} - m\delta_{\alpha\beta}) + g_s G\!\!\!/^a_{\alpha\beta} t^a_{ij}] q_{\beta f j} \tag{4.1.1}$$

where g_s is the strong coupling constant, $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu$ is the gluon field strength tensor, f^{abc} are the structure constants of SU(3), t^a are the generators of SU(3) in the fundamental representation, the sum over f is running over the six flavors, and the indices a and i, j label the eight gluons and the three colors of quarks, respectively. The six flavors of quarks have different masses and charges, as shown in Figure 4.2.

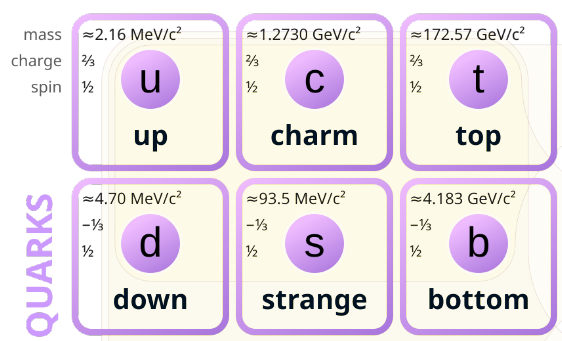


Figure 4.2. The quarks in the SM, reproduced from Ref. [1].

QCD is an extremely rich and complex theory due to its non-abelian gauge symmetry, the six different flavors of quarks, and the unique strength and running of its coupling, shown in Figure 4.3. Observe its property of weak coupling and asymptotic freedom at high energies, versus the extremely high $\mathcal{O}(1)$ value of α_s at low energies leading to the phenomenon of *confinement*. Note also that α_s appears to diverge in Figure 4.3 at around 200 MeV, a sign of perturbation theory breaking down. This 200 MeV limit is considered the characteristic energy scale of QCD, Λ_{QCD}.[2]

The $\mathcal{O}(1)$ coupling strength means that the standard perturbative techniques we have discussed are not applicable at our usual energy scales; instead, we must rely on nonperturbative techniques such as numerical simulations of QCD on a discretized spacetime lattice (see e.g. Schwartz [81] Chapter 25). Because of this, QCD is one of the least understood and most exciting areas of study in modern physics.

4.1.1 Asymptotic freedom and confinement

As discussed above, a key phenomenological characteristic of the strong force is asymptotic freedom, wherein at high energies quarks and gluons behave as free particles. This also means that perturbative techniques can be applied at high energies; indeed, we can derive an analogous $1/r^2$ "Coulomb force", based on

[2] The phenomenon of an energy scale arising from a dimensionless coupling constant is known as *dimensional transmutation* (see e.g. Tong SM [71] Chapter 3).

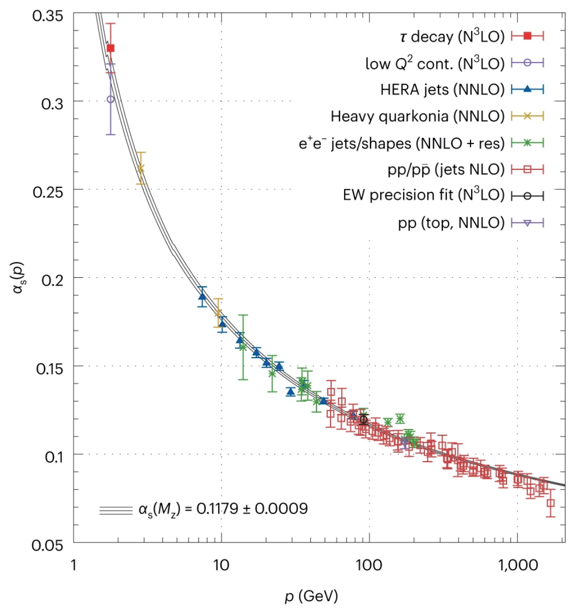


Figure 4.3. The theoretically predicted running of the strong coupling $\alpha_s = g_s^2/4\pi$ as a function of the energy scale along with experimental measurements, reproduced from Ref. [7].

tree-level quark-quark scattering amplitudes, for quarks at very short distances. This force turns out to always be attractive between quarks and antiquarks, as well as between two or even three quarks in different color states: the "aim" of the force appears to always be to form color-neutral bound states. These are called *mesons* for the case of an antiquark and quark pair, and *baryons* for three quarks.

At longer distances we enter the strong-coupling and nonperturbative regime, in which the dynamics are harder to understand. However, through lattice QCD simulations, we are able to see the emergence of a "flux tube" pulling quarks together as they are pushed apart, as shown in Figure 4.4. This phenomenon is referred to as *confinement*, and it means we can never observe free quarks or gluons outside high-energy colliders. Both the long- and short-distance behavior of the strong force conspire to always confine quarks in color-neutral hadrons. The scale of confinement is naturally set by $\Lambda_{\text{QCD}} \approx 200\,\text{MeV}$, which is hence roughly the radius of the proton and other hadrons (1fm in SI).

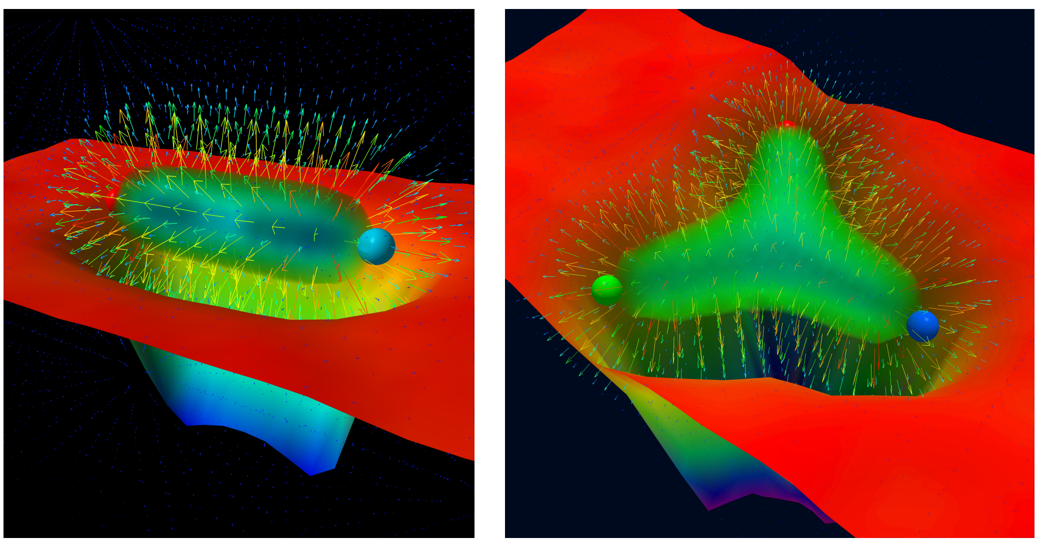


Figure 4.4. "Flux tubes" between a quark and anti-pair inside a meson (left) and three quarks in a proton (right), reproduced from Refs. [8, 9].

4.1.2 Quarks and the eightfold way

Since their discovery in the 1920s and 30s, the proton and neutron and were believed to be elementary particles along with the electron and photon. In fact, due to confinement, the first experimental evidence of quarks was not found until the 1960s. However, already in 1932, the remarkably similar masses of the two nucleons surprised physicists and led Heisenberg, Wigner, and others to hypothesize an underlying SU(2) symmetry between them (later named *isospin*) [112, 113]. The intrigue only increased in the next decades, during which new cosmic ray, cyclotron, and bubble chamber experiments discovered a veritable "zoo" of hadrons, exemplified by a 1964 table of particles in Figure 4.5. While all appeared elementary, several had surprisingly similar properties such as mass and spin, and could also be grouped into invariant subspaces of the isospin group.

In 1961, Murray Gell-Mann and Yuval Ne'eman independently realized that the new hadrons could elegantly fit into representations of a larger symmetry group, SU(3) [114, 115]. Gell-Mann and George Zweig in 1964 then independently showed that this could be explained physically by hadrons being composed of combinations of three fundamental particles, named the "up", "down", and "strange quarks", with the former two carrying isospin up and down, respectively [116, 117]. Gell-Mann named this model the "eightfold way" (since $\dim(SU(3)) = 8$) and was awarded the Nobel Prize in 1969 for this work.

Examples of baryons (three-quark hadrons) in the octet and decuplet (di-

Table of Elementary Particles (Compiled by W. H. Barkas and A. H. Rosenfeld). Errors Are Not Shown

Family	Particle	Spin	Mass (Mev.)	Mean life (second)
Photon	γ	1	0	stable
Leptons	$\nu(\bar{\nu})$	$\frac{1}{2}$	$<2 \times 10^{-4}$	stable
	$e^-(e^+)$	$\frac{1}{2}$	0.511	stable
	$\mu^-(\mu^+)$	$\frac{1}{2}$	105.7	2.26×10^{-6}
Mesons	$\pi^+(\pi^-)$	0	139.6	2.6×10^{-8}
	π°	0	135.0	2×10^{-16}
	$K^+(K^-)$	0	494	1.2×10^{-8}
	$K^\circ(\bar{K}^\circ)$	0	498	K_1°: 1.0×10^{-10}
	$(\lvert m(K_1^\circ) - m(K_2^\circ)\rvert \sim 5 \times 10^{-12})$			K_2°: $\sim 7 \times 10^{-8}$
Baryons	$p(\bar{p})$	$\frac{1}{2}$	938.2	stable
	$n(\bar{n})$	$\frac{1}{2}$	939.5	1.0×10^3
	$\Lambda^\circ(\Lambda^\circ)$	$\frac{1}{2}$	1115.5	2.5×10^{-10}
	$\Sigma^+(\overline{\Sigma^+})$	$\frac{1}{2}$	1189	0.8×10^{-10}
	$\Sigma^-(\overline{\Sigma^-})$	$\frac{1}{2}$	1197	1.6×10^{-10}
	$\Sigma^\circ(\overline{\Sigma^\circ})$	$\frac{1}{2}$	1193	theory $\sim 10^{-19}$
	$\Xi^-(\overline{\Xi^-})$	?	1318	1.3×10^{-10}
	$\Xi^\circ(\overline{\Xi^\circ})$	?	~ 1312	$\sim 2 \times 10^{-10}$

Figure 4.5. A table of what were considered to be elementary particles in 1964, reproduced from Ref. [10].

mension 8 and 10, respectively) representations of SU(3) are shown in Figure 4.6, sorted by their isospin along the "z" axis (I_3 = # of up quarks - # of down quarks) and strangeness (S = # of strange quarks). Note that this SU(3) symmetry is only approximate; it is broken by the different masses of the quarks. However, their significantly smaller masses compared to Λ_{QCD} mean it remains a useful symmetry for categorizing hadrons. On the other hand, broader "symmetries" such as SU(4) through SU(6) including the heavier charm, bottom and top quarks are broken so heavily by their higher masses that they are not helpful for characterizing the heavier hadrons.

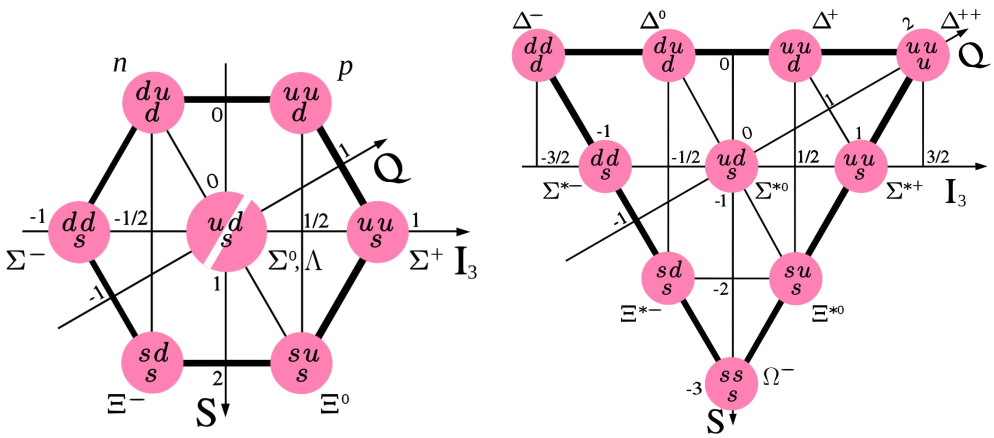


Figure 4.6. Baryons in the octet (left) and decuplet (right) representations of SU(3), reproduced from Ref. [11].

This fourth "charm" quark was notably predicted by Sheldon Glashow, John Iliopoulos, and Luciano Maiani in 1970 to explain the observed suppression of Z-boson-mediated flavor-changing neutral currents [118] (and also to match the number of known leptons at the time). This, and the quark model as a whole, was famously validated by the discovery of a 3.1 GeV charm-anti-charm bound state, named the J/ψ meson, simultaneously by Burton Richter's team at the Stanford Linear Accelerator Center (SLAC) and Samuel Ting's team at Brookhaven National Laboratory in 1974 [119, 120], both of whom received the Nobel Prize in 1976.

A year before this, Makoto Kobayashi and Toshihide Maskawa had proposed the existence of a *third* generation of quarks to explain the observed CP-violation in weak interactions [121]. This proposal gained more traction after the J/ψ discovery, as well as the discovery of a third-generation lepton, the τ, by Martin Lewis Perl's team in electron-positron collisions at SLAC between 1973 and 1977 [122].

In the end, both third generation quarks were discovered at the Fermi National Accelerator Laboratory (Fermilab): first the bottom quark in 1977 by Leon Lederman's team on the E288 experiment [123]; and then, much later, the top quark in 1995 by the CDF and DØ experiments at the Tevatron [124, 125]. The bottom quark was discovered indirectly, as with the charm quark, through the observation of a bottom quark-antiquark bound state called bottomium, or the Υ meson, in proton-nucleon collisions.

The top quark, on the other hand, is highly unique because of its high 173 GeV mass, and it decays too quickly to form bound states. Hence, it is the only quark to have been observed "directly", through its decays to a W boson and a bottom quark. It is the heaviest known elementary particle, which is why its discovery required the 1 TeV center-of-mass energy proton-antiproton collisions of the Tevatron. The unique nature of the heavy quarks leads to a rich phenomenology at high energy colliders such as the LHC, particularly in the context of the *jets* they form (Section 4.1.4).

4.1.3 The parton model

Some physicists, including Gell-Mann himself, initially believed quarks not to be real particles but simply mathematical conveniences to describe hadrons. It was only through *deep inelastic scattering* (DIS) experiments in the 1960s and 70s at SLAC — in which high energy electrons were shot at protons (in the form of hydrogen) to probe their inner structure — that it was confirmed that protons are indeed not point-like particles.

To explain this behavior, Richard Feynman and others proposed the *parton model* of the proton (and other hadrons). In this, protons are composed of point-like particles called *partons* that are what actually interact with the electrons in DIS, as illustrated in Figure 4.7. Though initially partons were abstract entities, we now identify them as the quarks and gluons of QCD. At the energies required for DIS (and modern hadron colliders), the "partonic" cross-section of electron-parton scattering (or parton-parton scattering) ($\hat{\sigma}$) can be calculated using standard perturbation theory and Feynman diagrams.

To then derive the total "hadronic" electron-proton cross-section, we must integrate over all possible electron-parton interactions, weighted by the probability of finding a parton carrying a fraction x of the proton's momentum at an energy scale Q^2. This is described by the *parton distribution functions* (PDFs) $f_i(x, Q^2)$, where i represents the type of parton. PDFs cannot be calculated perturbatively and must be determined from experimental data. Examples for the proton at $Q^2 = 10\,\text{GeV}$ are shown in Figure 4.8; observe that the up and down quarks — called the *valence* or "real" quarks — dominate at high x, while at lower x there are gluons as well as other *sea* (i.e., virtual) quarks.

The overall hadronic cross-section for DIS is thus:

$$\sigma_{eh} = \sum_i \int_0^1 \mathrm{d}x\, f_i(x, Q^2)\hat{\sigma}_{ei}(Q^2, \mu_r), \tag{4.1.2}$$

where μ_r is the scale used for renormalization when calculating the partonic cross-sections. The separation of the perturbative and nonperturbative parts of the cross-section is called *factorization*, and the fact that this is possible is proved in the *factorization theorem* [126].

As also illustrated in Figure 4.7, high energy hadron-hadron collisions such as those at the LHC involve a similar, but more complicated, interaction. The corresponding cross-section involves integrating over two partons' momenta (one each from the two colliding hadrons):

$$\sigma_{hh} = \sum_{i,j} \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 f_i(x_1, Q^2) f_j(x_2, Q^2)\hat{\sigma}_{ij}(Q^2, \mu_r). \tag{4.1.3}$$

This is known as the "master formula" for cross-sections at the LHC.[3] PDFs are generally measured via DIS at electron-proton colliders, and are then crucial inputs to the above equation for hadron colliders. There is also hope of deriving these through lattice QCD simulations.

[3]See lectures by Torsten Pfoh [127] and Joey Huston [128] for useful pedagogical discussions.

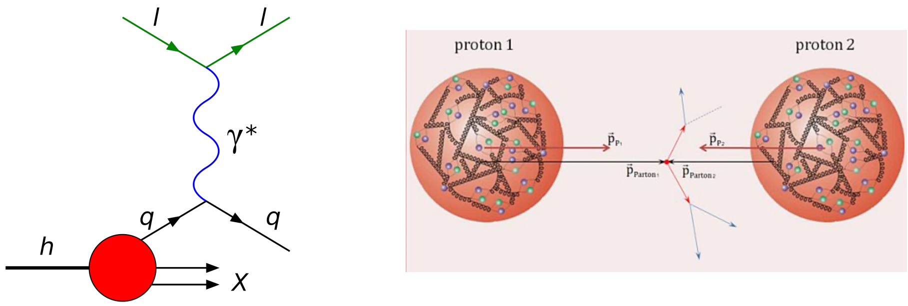


Figure 4.7. Feynman diagram for deep inelastic scattering, reproduced from Ref. [12] (left) and an illustrative example of proton-proton collisions reproduced from Ref. [13] (right).

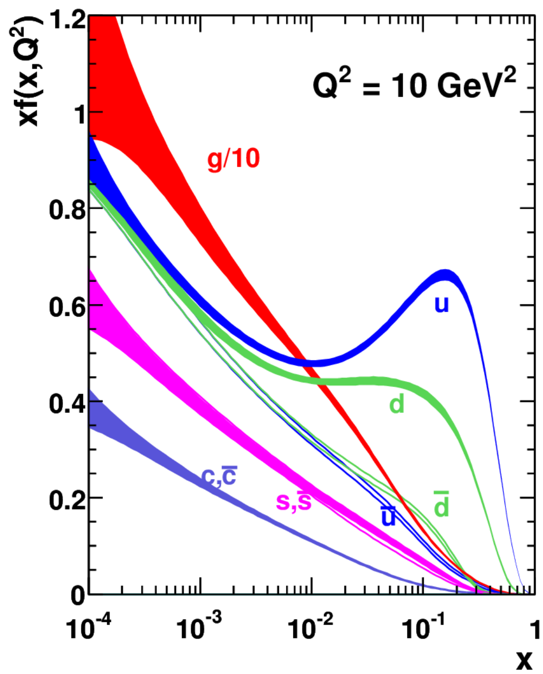


Figure 4.8. PDFs for the proton at $Q^2 = 10\,\text{GeV}$, reproduced from Ref. [14].

The partonic cross section

The partonic cross-section $\hat{\sigma}(Q^2, \mu_r)$ is an important theoretical input for measurements at high energy colliders. The dependence on μ_r is perhaps surprising; however, it represents the fact that $\hat{\sigma}$ is calculated perturbatively: the μ_r dependence only appears in the highest order term of the expansion. Indeed, this scale dependence would disappear at infinite order in perturbation theory. While it may seem a nuisance, in fact, it provides a convenient handle to estimate the *uncertainties* on our theoretical predictions by simply varying μ_r and μ_f.[4]

One important feature to keep in mind regarding the perturbative calculations for hadron colliders is that the leading order (LO) predictions are often a factor of $\gtrsim 2$ off the higher order next-to-LO (NLO) and next-to-NNLO (NNLO) calculations. This is exemplified in the predictions for Z boson production at the LHC, shown in Figure 4.9. The reason for this, despite α_s being reasonably small (≈ 0.1) at the scale for this process $m_Z \simeq 90\,\text{GeV}$, is simply that the $\mathcal{O}(\alpha_s)$ corrections have large coefficients [21]. This is why measurements at the LHC relying on LO simulations often multiply the cross-section with an NLO / LO "K-factor".

Practically, matrix elements are first calculated as a function of the input and output "hard particle" momenta, after which event generator programs such as MadGraph [129] use Monte Carlo (MC) methods to sample events appro-

[4]See Ref. [21] 4.1 for further discussion.

priately from the overall phase space. NLO and NNLO calculations are more complicated and often involve weighting events negatively to represent subtractions at higher orders [130].

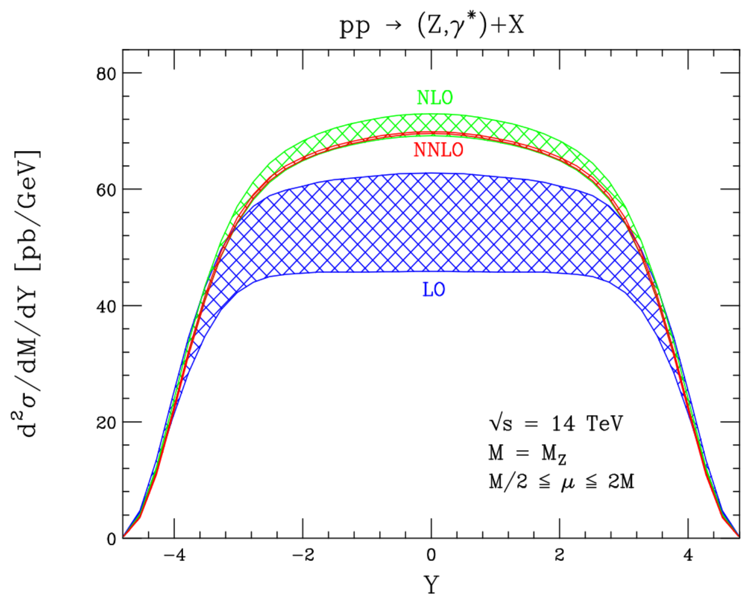


Figure 4.9. LO, NLO, and NNLO predictions and uncertainties for pp to Z boson production, differential in rapidity Y at the LHC, reproduced from Ref. [15].

Parton evolution

Each parton has a certain probability of radiating another quark or gluon, with a fraction of the original parton's momentum, z. These are called parton splitting functions, $P_{ij}(z)$, depicted in Figure 4.10, and can be calculated perturbatively in

QCD (see e.g. Ref. [21]). They are then further convolved with PDFs to derive their evolution with the energy scale:

$$\frac{\mathrm{d}f_i(x,Q^2)}{\mathrm{d}Q^2} = \frac{1}{Q^2}\sum_j \int_x^1 \frac{\mathrm{d}z}{z} f_j(x/z, Q^2) P_{ji}(z). \tag{4.1.4}$$

Equations 4.1.4 are called the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations, after five physicists who developed them in the 1970s, and are analogous to the renormalization group flows of coupling constants. The dependence of the PDFs on the energy scale has been confirmed in DIS experiments, which are then also used to fit the parameters of the PDFs, as shown in Figure 4.11.

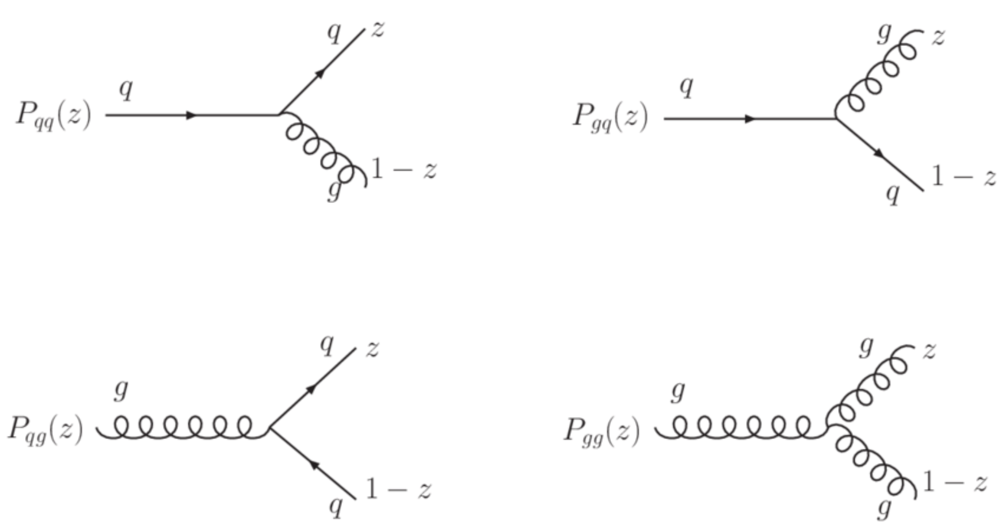


Figure 4.10. The splitting functions for quarks and gluons, reproduced from Ref. [16].

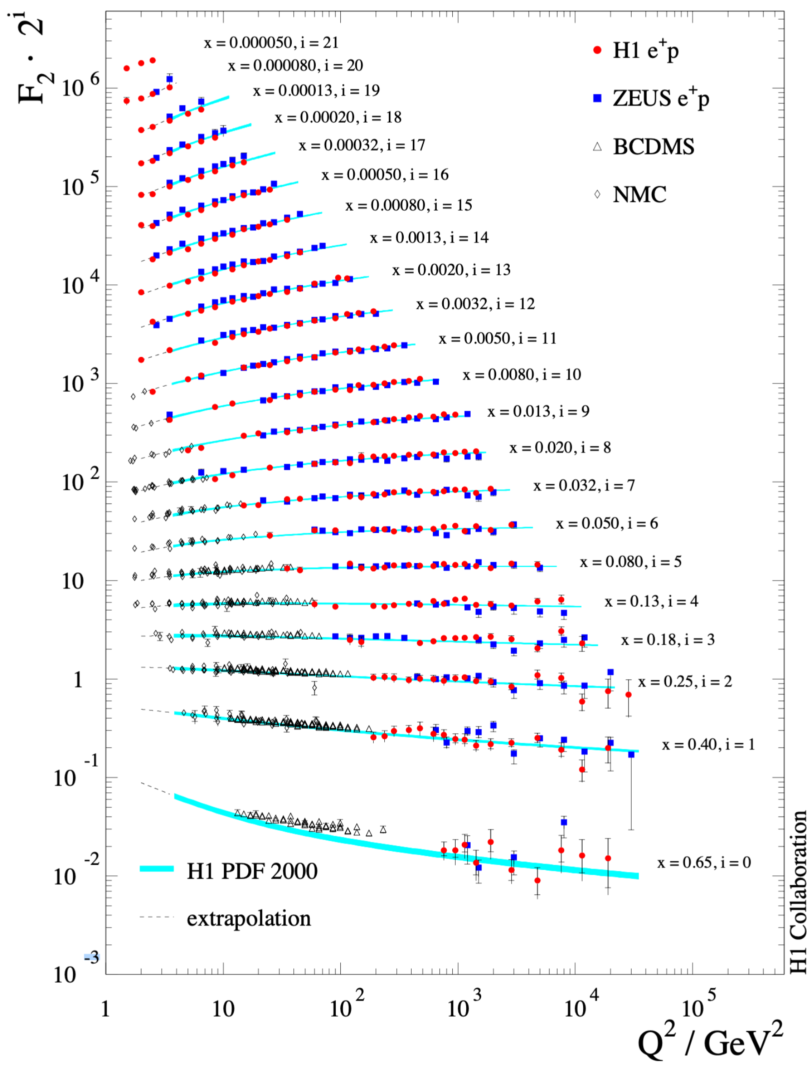


Figure 4.11. PDF measurements at different energy scales Q^2 and momentum fraction x by the H1 collaboration in DIS experiments, reproduced from Ref. [17].

4.1.4 Jets

As one may infer from the DGLAP equations (Eq. 4.1.4), when high energy partons are produced at a collider, they will probabilistically radiate further and further partons — called *parton showering* — until they approach the confinement scale and start forming bound hadrons — called *hadronization*. For sufficiently high energy initial partons, the resulting hadrons will appear as a collimated spray of particles in the detector, called a *jet* (Figures 4.12 and 4.13).

Since quarks and gluons are never observed in isolation, their production can only be inferred by understanding the jets they form. Moreover, at a hadron collider, the high-energy hadrons continuously radiate partons *before* and *after* the collision as well, with the resulting jets referred to as *initial* and *final state radiation* (ISR and FSR), respectively. Such jets are by far the most prevalent outputs of collisions at the LHC and, hence, represent a significant background in many measurements and searches, particularly those searching for hadronic final states.

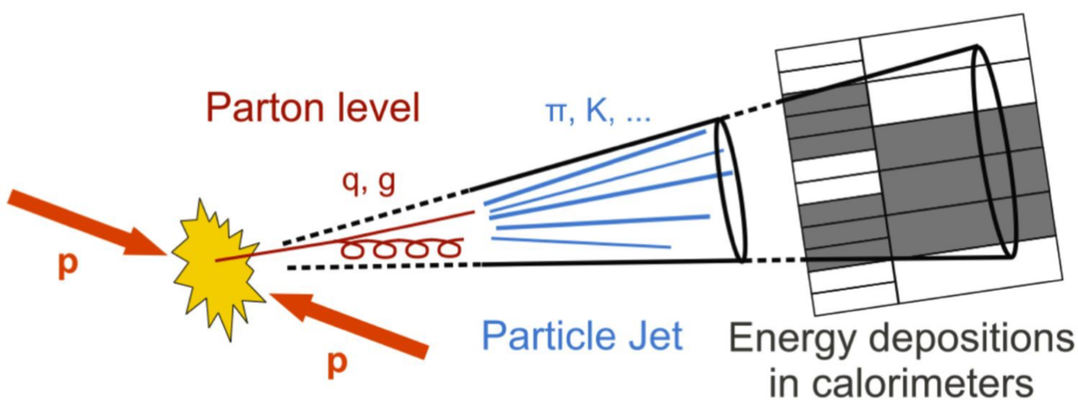


Figure 4.12. A cartoon of a jet, reproduced from Ref. [18].

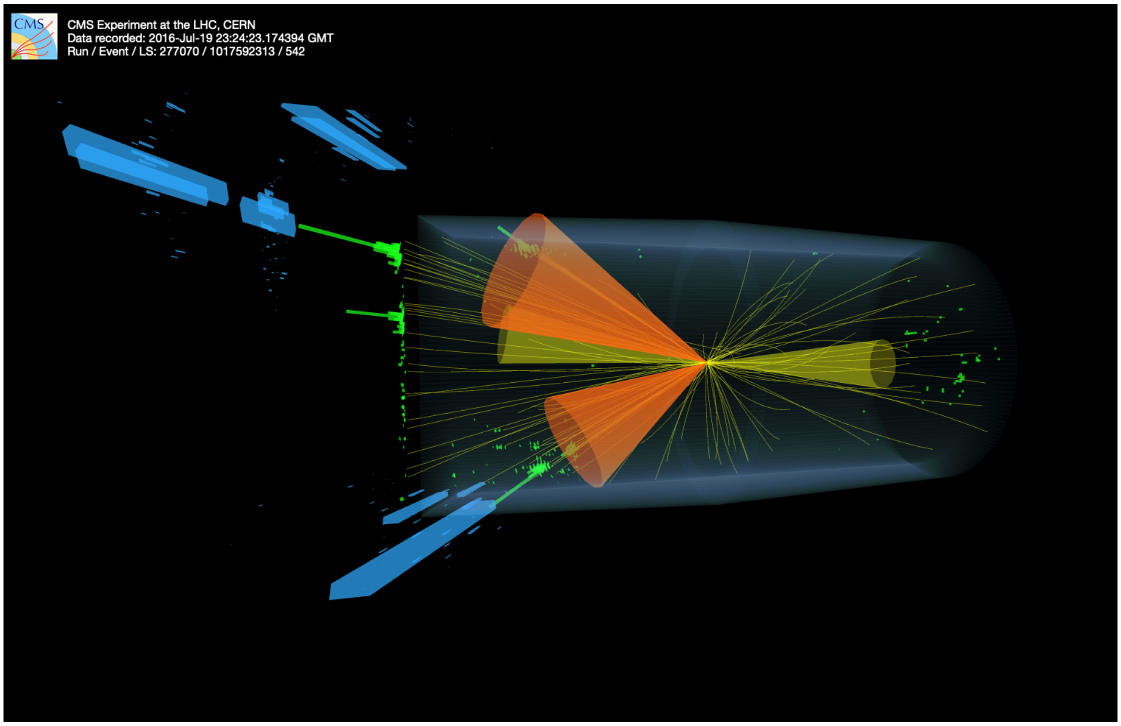


Figure 4.13. An example of real jets in an event collected by CMS and identified in the search described in Chapter **??** [19, 20]. An interactive version of this event display is available at https://cms3d.web.cern.ch/HIG-23-012/.

Parton showering

Jets can be understood and modeled by factorizing the dynamics. As above, the parton scattering cross section (referred to as the *hard process* and calculated perturbatively) is separated from the PDFs (measured from data) and their evolution (DGLAP equations). This evolution is what produces the showering, and is modeled by numerically iterating through Q^2 (or, equivalently, through time) and randomly emitting new partons according to the splitting functions via MC sampling.

There are several subtleties involved in this process which numerical parton shower generators, such as PYTHIA [131], HERWIG [132], and SHERPA [133] must account for. First, the probability of gluon emission diverges in the soft — i.e., low gluon energy — and collinear — small gluon angle with the parent parton — limits. Physically, this can be interpreted as the limit of our experimental resolution: at a certain point we cannot resolve two close-by or detect arbitrarily soft particles.

These are known as the *infrared* and *collinear* (IRC) divergences, respectively, and are typically regulated by introducing cut-off energies and angles for emissions (below which we can reasonably argue that perturbation theory is anyway invalid). These divergences also mean that when analyzing jets in experimental data, care must be taken in defining observables to be *IRC-safe*, meaning that jet clustering algorithms and physical properties derived therein should not be sensitive to arbitrarily soft or collinear emissions.

Another issue is that a naive combination of the hard matrix element and subsequent parton shower calculations may lead to double-counting of emissions, as illustrated in Figure 4.14. This necessitates a careful "matching procedure", such as the most common MLM scheme [134], which defines cut-off energy and angular scales to separate the matrix element and parton shower phase spaces. Other considerations include preserving unitarity, color coherence and color flow, and differences between ISR and FSR (see e.g. Refs. [135, 136]).

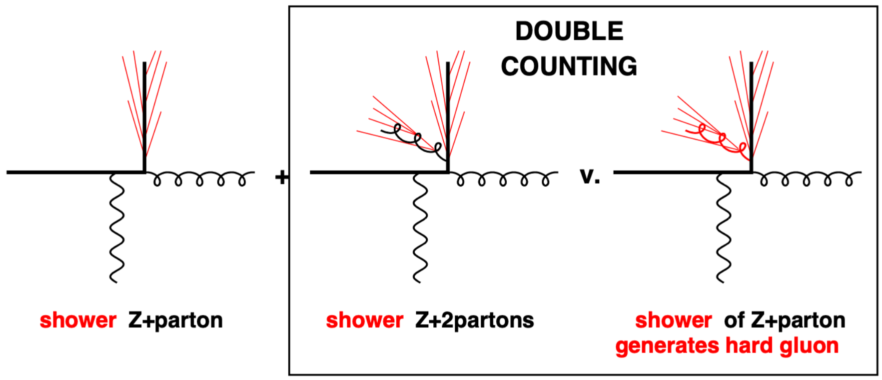


Figure 4.14. An illustration of double-counting when combining matrix element predictions (in black) with parton showering algorithms (in red) for Z+parton and Z+2-parton events, reproduced from Ref. [21].

Hadronization

The final element of the factorized process is hadronization, once the parton shower approaches the confinement scale. This is a completely nonperturbative process and, hence, like PDFs, we must rely on numerical simulations and experimental measurements.

Lattice QCD simulations, such as those shown in Figure 4.4, indicate that in the low energy limit, the effective potential between quarks increases linearly with distance, resembling string tension:

$$V(r) = \sigma r, \tag{4.1.5}$$

where σ is the string tension coefficient. In fact, this analogy can be extended further: above a certain energy, the string appears to "snap", in the sense that it becomes possible and energetically more favorable to produce a quark-antiquark pair.

This analogy the basis of the *Lund string model* of hadronization [**?**], illustrated in Figure 4.15. The strong force between the final state partons is modeled as a series of strings stretched between them that probabilistically break into new partons. Other models are based on clustering partons into color-neutral combinations [132].

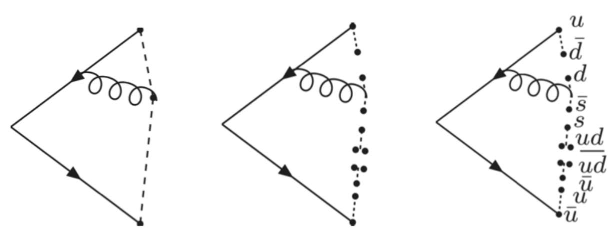


Figure 4.15. An illustration of the Lund string model of hadronization, reproduced from Ref. [22].

4.2 Electroweak interactions

The weak interaction is the last of the three fundamental forces we discuss in the SM. Apart from its relatively weak coupling constant (Table 4.1), it is unique in several ways: (1) it couples only to left-chiral fermions, thereby violating parity (P) and charge conjugation (C); (2) it is the only force with massive gauge bosons, resulting in short-range interactions; and (3) it is the only force that "sees" and can change the flavors of the fermions. Hence, it is responsible for radioactive decays and the instability of all hadrons and leptons bar the proton and electron. Its couplings to the different flavors also lead to CP-violation, as we discuss in Section 4.2.4.

4.2.1 Weak interactions

The first theory of weak interactions was Enrico Fermi's 1933 theory of beta decay [137]: the decay of the neutron to a proton, $n \to p + e^- + \bar{\nu}_e$, through a four-fermion interaction (Figure 4.16, left). Fermi was inspired by Dirac's nascent theory of QED, and using similar perturbative techniques, his theory proved successful in describing weak decays. The same principle was also applied to other weak decays, such as muon decay (Figure 4.16, right) and pion decay.

As it turned out, the four-fermion interaction is of mass dimension 6 and not renormalizable (see Chapter 3.2), leading to the scattering cross-section diverging at high energies. This is, of course, because these interactions are in fact mediated by the massive weak $W^\pm$ and Z gauge bosons, which become relevant around their mass scale of $\mathcal{O}(100\,\mathrm{GeV})$. We now understand the Fermi theory as an effective field theory (EFT) valid for energies much lower than $100\,\mathrm{GeV}$, wherein the W and Z boson DoFs can be integrated out and nonrenormalizable interactions are allowed — they are just suppressed by factors of $({}^{1}/{M_W})^2$. This suppression is why the weak interaction is so weak, with a coupling constant of $\approx \mathcal{O}(10^{-6})$ at the mass scale of the proton.

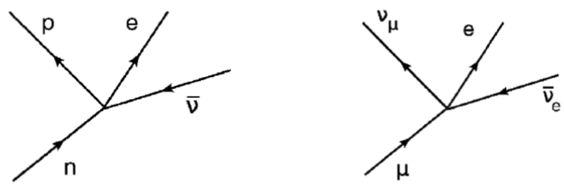


Figure 4.16. Feynman diagrams for beta decay (left) and muon decay (right) in Fermi's theory.

The weak interaction is described by an SU(2) Yang-Mills theory, and is sometimes referred to as quantum flavordynamics (QFD) because of its deep connection to flavor, as we will discuss. However, "vanilla" Yang-Mills theories cannot accommodate massive gauge bosons; hence, it was only after the development of the ABEGHHK (Higgs) mechanism in the 1960s that this description gained traction. Specifically, Sheldon Glashow, Abdus Salam, Steven Weinberg and others showed that the spontaneous breaking of an SU(2) × U(1) symmetry to U(1) could not only yield massive weak gauge bosons, but also naturally incorporate QED with a massless photon [108–110].

This combined *electroweak* theory has been experimentally confirmed in many stages: first with the discovery of *neutral currents* involving neutrinos with the Gargamelle bubble chamber at CERN in 1973 [138], the first evidence for the Z boson (the only neutral boson that couples to neutrinos); then with the direct discovery of the W and Z bosons at the Super Proton Synchrotron (SPS) in 1983 [139–142], as well as precision measurements of electroweak parameters such as the W and Z masses with the Large Electron-Positron Collider (LEP) in the 1990s; and finally with the discovery of the Higgs boson at the LHC in 2012 [59, 143], a particle predicted by the ABEGHHK mechanism, and the ongoing measurements of its properties.

4.2.2 Before electroweak symmetry breaking

Electroweak interactions are associated with the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge symmetry, which is "spontaneously broken" to $\mathrm{U}(1)_{\mathrm{EM}}$ through the ABEGHHK mechanism during electroweak symmetry breaking (EWSB).[5] We label the three gauge bosons of $\mathrm{SU}(2)_L$ as W^1, W^2, W^3 and of $\mathrm{U}(1)_Y$ as B, with coupling constants g and g', respectively.

The fermions in the SM can be categorized by their representations, or charges, under the three gauge symmetries before EWSB, as in Table 4.2. The bold numbers indicate the dimension of the representation under the respective symmetry group, while the regular numbers are the charges under the $\mathrm{U}(1)_Y$ group — referred to as their *hypercharge*, Y.

The weak interactions specifically are associated with the $\mathrm{SU}(2)_L$ gauge

[5] As discussed in Chapter 3.4, technically, the gauge symmetry cannot be broken — what breaks is the associated global symmetry.

Table 4.2. The representations and charges of fermionic and scalar fields in the SM under the $\mathrm{U}(1)_Y$, $\mathrm{SU}(2)_L$, and $\mathrm{SU}(3)_C$ gauge symmetries. The right-handed neutrino is included here for completeness but has not been experimentally confirmed.

	$\mathrm{U}(1)_Y$	$\mathrm{SU}(2)_L$	$\mathrm{SU}(3)_C$
Q_L	$+1/6$	$\mathbf{2}$	$\mathbf{3}$
L_L	$-1/2$	$\mathbf{2}$	$\mathbf{1}$
u_R	$+2/3$	$\mathbf{1}$	$\mathbf{3}$
d_R	$-1/3$	$\mathbf{1}$	$\mathbf{3}$
e_R	-1	$\mathbf{1}$	$\mathbf{1}$
ν_R?	0	$\mathbf{1}$	$\mathbf{1}$
H	$+1/2$	$\mathbf{2}$	$\mathbf{1}$

symmetry, and their key characteristic is that they violate parity: they only couple to left-handed fermions and right-handed antifermions (hence, the subscript L). Specifically, the left-handed quarks (u_L and d_L) and leptons (ν_L and e_L) reside in $\mathrm{SU}(2)_L$ doublets:

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \tag{4.2.1}$$

while right-handed fermions live in the trivial representation. The L and R subscripts indicate left- and right-chiral Weyl spinors, respectively. Note that there are actually three generations of fermions, which we index as u^i for $i = 1, 2, 3$; however, before EWSB, there is no distinction between them as they are all massless. Often, we will omit this index when the properties across generations are identical.

The SM also contains a scalar, SU(2)-doublet "Higgs" field H, which is listed in Table 4.2 as well. Its dynamics are governed by the Lagrangian:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu H)^\dagger (D^\mu H) - \lambda (H^\dagger H - \frac{v^2}{2})^2, \tag{4.2.2}$$

where v is a constant and

$$D_\mu H = \left[\partial_\mu - igW_\mu - \frac{i}{2}g' B_\mu\right] H \tag{4.2.3}$$

is the covariant derivative of H. The potential resembles the sombrero potential from Figure 3.7, but with a 2D rather than 1D complex field.

The Higgs field is able to couple to the fermions without violating gauge symmetry through Yukawa interactions. The most general possible Yukawa terms are 3×3 matrices across the three generations.

$$\mathcal{L}_{\text{Yukawa}} = -y^d_{ij} \bar{Q}^i_L H d^j_R - y^u_{ij} \bar{Q}^i_L \tilde{H} u^j_R - y^e_{ij} \bar{L}^i_L H e^j_R - y^\nu_{ij} \bar{L}^i_L \tilde{H} \nu^j_R + \text{h.c.}, \tag{4.2.4}$$

where $\tilde{H} = i\sigma_2 H^\dagger$ is the charge-conjugated Higgs field, i and j index the three generations of fermions, y^d_{ij} are the Yukawa coupling constant matrices, and h.c. denotes the Hermitian conjugate of all the preceding terms.

The Higgs field contracts with the fermionic left-handed SU(2)-doublets to produce an SU(2)-singlet, while the quark fields contract with each other to form

SU(3)-singlets. One can also check that through our clever choice of H vs. $\tilde{H}$, the total hypercharge of each term is 0. Thus, each Yukawa term is independently gauge invariant.

The overall electroweak Lagrangian is:

$$\begin{aligned}\mathcal{L}_{\text{EW}} = -\frac{1}{4}W^a_{\mu\nu}W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ + \bar{Q}_L i\not{D} Q_L + \bar{L}_L i\not{D} L_L + \bar{u}_R i\not{D} u_R + \bar{d}_R i\not{D} d_R + \bar{e}_R i\not{D} e_R \\ + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}},\end{aligned} \tag{4.2.5}$$

Note that without EWSB, not only are all the gauge bosons of the theory massless, but so are the fermions: the usual fermionic mass terms of the form $mu_L u_R$ violate the $\text{SU}(2)_L$ gauge symmetry. As we will see, the ABEGHHK mechanism is what generates masses for all the fermions, through the Higgs Yukawa couplings, as well as the three weak gauge bosons.

4.2.3 Electroweak symmetry breaking

EWSB occurs when the Higgs field spontaneously breaks the global $\text{SU}(2)_L \times \text{U}(1)_Y$ symmetry by moving to a ground state of the potential. Without loss of generality, we can choose this ground state to be:

$$\langle H \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{4.2.6}$$

where $\langle H \rangle$ is the vacuum expectation value (VEV) of the Higgs field. As before, we can parametrize the fluctuations around this ground state as:

$$H = e^{i\xi^A(x)T^A} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{4.2.7}$$

where h is a real scalar field, T^A are three generators of the broken $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry, and ξ^A are the corresponding Goldstone bosons.

The ABEGHHK mechanism for gauge theories effectively involves the original gauge bosons absorbing these Goldstone bosons, thereby acquiring mass (see Chapter 3.4). This can be equivalently thought of as simply a convenient choice of gauge in which $\xi^A(x) = 0$. In the end, after some algebra, the gauge + Higgs sector of the electroweak Lagrangian after EWSB looks like:

$$\begin{aligned}\mathcal{L}_{\mathrm{GH}} = -\frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \frac{1}{2} \partial_\mu h \partial^\mu h - \underbrace{\lambda h^2 \left(v + \frac{h}{2} \right)^2}_{\equiv V(h)} \\ + \frac{1}{8} (v + h)^2 \left[g^2 (W^1_\mu)^2 + g^2 (W^2_\mu)^2 + (g W^3_\mu - g' B_\mu)^2 \right],\end{aligned} \tag{4.2.8}$$

where $V(h)$ is the new Higgs potential.

We now have mass terms for three gauge, as well as the Higgs, bosons. As always, we are free to define the gauge boson fields as we wish, and it turns out the most convenient choice is:

$$
\begin{aligned}
W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2), \\
Z_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(g W_\mu^3 - g' B_\mu), \\
A_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(g' W_\mu^3 + g B_\mu).
\end{aligned}
\tag{4.2.9}
$$

It is conventional to define the *Weinberg* or *weak mixing angle* θ_W:

$$
\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}, \quad \sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}}, \tag{4.2.10}
$$

to simplify the forms of Z and A above:

$$
\begin{aligned}
Z_\mu &= W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W, \\
A_\mu &= W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W.
\end{aligned}
\tag{4.2.11}
$$

Experimentally, we have determined the free parameters of this theory to be:

$$
v \approx 250\,\mathrm{GeV}, \quad \lambda \approx 0.35, \quad g \approx 0.64, \quad g' \approx 0.34 \quad \Rightarrow \quad \sin^2\theta_W \approx 0.223, \tag{4.2.12}
$$

at an energy scale of the Z boson mass.

The Lagrangian can hence be written as:

$$\begin{aligned}\mathcal{L}_{\mathrm{GH}} = -\frac{1}{4}W^{+}_{\mu\nu}W^{-\mu\nu} - \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + m_W^2 W^{+}_{\mu}W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu \\ + \frac{2m_W^2}{v}hW^{+}_{\mu}W^{-\mu} + \frac{m_Z^2}{v}hZ_\mu Z^\mu + \frac{m_W^2}{v^2}h^2 W^{+}_{\mu}W^{-\mu} + \frac{m_Z^2}{4v^2}h^2 Z_\mu Z^\mu \\ + \frac{1}{2}\partial_\mu h \partial^\mu h - m_h^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4,\end{aligned} \quad (4.2.13)$$

where

$$m_W = \frac{1}{2}vg \approx 80\,\mathrm{GeV}, \quad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \approx 91\,\mathrm{GeV}, \quad m_h = \sqrt{2\lambda}v \approx 125\,\mathrm{GeV}. \quad (4.2.14)$$

Note that in addition to the gauge-boson self-interactions, the Higgs field also has a trilinear ($\lambda v h^3$) and quartic ($1/4\lambda h^4$) self-interaction terms. While some manner of EWSB has been confirmed experimentally through the discovery of the Higgs boson, its full nature, and the full form of the Higgs potential, can only be determined through measurements of these terms. The trilinear self-coupling, in particular, can be accessed at the LHC through pair production of the Higgs boson, as we will discuss in Section 4.3. Higgs pair production also allows exclusive access to the quartic $hhVV$ couplings, where V is a weak gauge boson.

The photon and re-emergence of Dirac spinors

A_μ corresponds to the one unbroken symmetry of the original $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group. In terms of the original generators — three T^As for $\mathrm{SU}(2)_L$ and one Y for $\mathrm{U}(1)_Y$ — A_μ corresponds to linear combination $Q = T^3 + Y$. This is the massless photon field, with the gauge group $\mathrm{U}(1)_{\mathrm{EM}}$.

The eigenvalue of Q for each fermion corresponds to their electric charge under this group. For the $\mathrm{SU}(2)_L$-singlet fields, T_3 has no value and hence $Q = Y$, while for the doublets, T_3 has eigenvalues $\pm 1/2$ for the upper and lower components, respectively: e.g. for u_L, $Q = 1/2 + 1/6 = 2/3$ and for e_L, $Q = -1/2 - 1/2 = 1$. In the end, we see that the left- and right-chiral fermion field pairs have the same charge under this remaining unbroken symmetry, so they can again form Dirac spinors:

$$\psi_u = \begin{pmatrix} u_L \\ u_R \end{pmatrix}, \quad \psi_d = \begin{pmatrix} d_L \\ d_R \end{pmatrix}, \quad \psi_e = \begin{pmatrix} e_L \\ e_R \end{pmatrix}, \quad \psi_\nu = \begin{pmatrix} \nu_L \\ \nu_R \end{pmatrix}, \tag{4.2.15}$$

times three for each generation.

Each Weyl spinor pair interacts identically with the photon and gluons but the W and Z bosons continue to couple only to the left-handed components.

For example the W^+–fermion coupling is:

$$\mathcal{L}_{W^+f} = -\frac{g}{\sqrt{2}} W_\mu^+ \underbrace{(\bar{u}_L \bar{\sigma}^\mu d_L + \bar{\nu}_L \bar{\sigma}^\mu e_L)}_{\equiv J_\mu^+}, \tag{4.2.16}$$

where J_μ^+ is called the weak charged current. In terms of Dirac spinors, this can be written using the projection operator (Eq. **??**):

$$\mathcal{L}_{W^+f} = -\frac{g}{\sqrt{8}} W_\mu^+ \left(\bar{\psi}_u \gamma^\mu (1-\gamma_5) \psi_d + \bar{\psi}_\nu \gamma^\mu (1-\gamma_5) \psi_e \right). \tag{4.2.17}$$

Recall that $\bar{\psi}\gamma^\mu\psi$ is a Lorentz vector, while $\bar{\psi}\gamma^\mu\gamma_5\psi$ is a pseudo- or axial-vector, so the weak current is effectively an axial vector subtracted from a vector. Indeed, historically, the weak interaction was referred to as "V-A" theory.

A hierarchy problem

Generally, if we encounter a new energy scale in nature, we are either able to connect it in some way to an existing scale or a (broken) symmetry of the theory. For example, the mass of the proton $\sim 1\,\text{GeV}$ is based on the QCD confinement scale $\Lambda_{\text{QCD}} \sim 200\,\text{MeV}$, which in turn is related to the Planck scale $\Lambda_{\text{Planck}} \sim 10^{19}\,\text{GeV}$ through dimensional transmutation. The mass of the pion on the other hand, $\approx 140\,\text{MeV}$, is a consequence of chiral symmetry breaking in QCD. Physicists such as Dirac and Gell-Mann have in fact proposed these criteria as a principle of "naturalness" for physical theories [144, 145].

There are many examples of mysterious energy scales appearing experimentally, which were either later rationalized or in fact even used to correctly predict new physics, such as the prediction of the charm quark mass based on the small mass difference between the K_L^0 and K_S^0 mesons. The electroweak energy scale of $\approx 100\,\text{GeV}$ is one such example which has yet to be explained.

A common way of expressing the problem is based on the Higgs mass: if we believe the SM to be an EFT valid up to some energy scale Λ, and if we have *a priori* no other energy scale to which to tie the Higgs mass, then we expect higher order corrections to its bare mass to be of order Λ. For example, if there is no new physics up to the Planck scale, then we are left with a bare mass and correction both of order 10^{19} GeV. The fact that the actual mass is 125 GeV implies a cancellation between the two, or *finetuning*, at a $^{125}/_{10^{19}} \approx 10^{-15}\%$ level. This is considered highly "unnatural" and is called a *hierarchy problem*, perhaps hinting at new physics.

One possibility is that Λ is in fact on the order of the Higgs mass, and we are simply yet to find the new degrees of freedom at this $\mathcal{O}(100\,\text{GeV})$ scale. Or, if we accept a higher level of finetuning, e.g. at the 10% or 1% levels, Λ can be pushed up further to 1–10 TeV. Effectively, we can invert the hierarchy problem into setting a bound on new physics!

Another solution is the existence an underlying (approximate) symmetry of nature "protecting" the Higgs mass from higher order corrections, similar to the chiral symmetry for the pion mass. The most promising candidate is *supersymmetry*, postulating an additional global symmetry between bosons and fermions [146]. Both these solutions possibly hint at an extended scalar sector, with either another Higgs doublet or new scalar singlets, for example, through the minimal supersymmetric extension of the SM (MSSM) [147] or based on two-real-scalar-singlet models [148]. This is one motivation behind the search for new Higgs bosons described in this dissertation. A more detailed, pedagogical discussion of naturalness can be found in e.g. Nathaniel Craig's IAS lectures [149].

4.2.4 Fermion masses and flavor

After EWSB, we have

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+h \end{pmatrix}, \quad \tilde{H} = \frac{1}{\sqrt{2}}\begin{pmatrix} v+h \\ 0 \end{pmatrix}, \tag{4.2.18}$$

which yields the following Yukawa Lagrangian:

$$\mathcal{L}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}}(v+h)\left[y^d_{ij}\bar{d}^i_L d^j_R + y^u_{ij}\bar{u}^i_L u^j_R + y^e_{ij}\bar{e}^i_L e^j_R + y^\nu_{ij}\bar{\nu}^i_L \nu^j_R\right] + \text{h.c.}. \tag{4.2.19}$$

Again, we are free to redefine fields and can choose a basis for all the fermion fields in which the Yukawa matrices are diagonal:

$$u^i_L \to (V^u)^i_j u^j_L, \quad d^i_L \to (V^d)^i_j d^j_L, \quad u^i_R \to (U^u)^i_j u^j_R, \quad d^i_R \to (U^d)^i_j d^j_R, \tag{4.2.20}$$

such that

$$V^{\dagger u} Y^u U^u = \mathrm{diag}(m_u, m_c, m_t), \quad V^{\dagger d} Y^d U^d = \mathrm{diag}(m_d, m_s, m_b), \qquad (4.2.21)$$

and same for the leptons (though again with the caveat that the right-handed neutrino, ν_R, has not been experimentally confirmed).

This is the *mass eigenstate* basis, where each generation and type of fermion has terms of the form:

$$\mathcal{L}_{\mathrm{Yukawa}} = -\frac{v}{\sqrt{2}} y^u \bar{u}_L u_R - \frac{h}{\sqrt{2}} y^u u_L u_R + \mathrm{h.c.} + \ldots, \qquad (4.2.22)$$

i.e., a mass term $m_X = \frac{v y^X}{\sqrt{2}}$ and a Yukawa interaction term with the Higgs field of strength y^X. The same Yukawa constant determines both the mass of the particle and its coupling to the Higgs field; indeed, in Appendix **??**, we show how this is used to determine the Higgs to fermion decay rates!

Like v and λ, the Yukawa couplings are all free parameters of the SM which can only be determined experimentally.

The CKM and PMNS matrices

Observe that the transformation we needed to diagonalize the Yukawa matrices in Eq. 4.2.20 violates the $\mathrm{SU}(2)_L$ symmetry, by transforming the up and down components of Q_L and L_L independently. This is actually okay because the $SU(2)_L$ symmetry was already broken through EWSB; however, crucially, this means the weak eigenstates (also called the *flavor eigenstates*), which the W bosons couple together, are not the same as the mass eigenstates!

Explicitly, any term which couples the up or down components of the SU(2)-doublet only with their respective right-handed components, such as the kinetic terms $\partial_\mu u_L \partial^\mu u_L$ or the electromagnetic interaction $A_\mu \bar{\psi}_u \gamma^\mu \psi_L$, is invariant under the transformation in Eq. 4.2.20. It is only the weak charged currents (Eq. 4.2.16) in the SM which mix the two. Including the three generations, the positive current is:

$$J^+_\mu = \sum_i \bar{u}^i_L \bar{\sigma}^\mu d^i_L + \sum_i \bar{\nu}^i_L \bar{\sigma}^\mu e^i_L \tag{4.2.23}$$

in the flavor eigenstate basis (the negative current is simply the h.c.). But if we attempt to transform to the mass basis via Eq. 4.2.20:

$$J^+_\mu = \sum_i \bar{u}^i_L \bar{\sigma}^\mu \underbrace{[(V^u)^\dagger V^d]_{ij}}_{\equiv V_{\text{CKM}}} d^j_L + \sum_i \bar{\nu}^i_L \bar{\sigma}^\mu \underbrace{[(V^\nu)^\dagger V^e]_{ij}}_{\equiv V_{\text{PMNS}}} e^i_L, \tag{4.2.24}$$

we are left with these two matrices, called the *Cabibbo-Kobayashi-Maskawa* (CKM) [121] and *Pontecorvo-Maki-Nakagawa-Sakata* (PMNS) [150] matrices mixing the quark and lepton flavors, respectively. The upshot is that the W bosons can not only mix the components of (the former) $SU(2)_L$ doublets, but also different mass eigenstates! The Z and photon currents do not have this property, which is why we say there are no *flavor-changing neutral currents* (FCNCs) in the SM, at least at tree level.

The magnitude of this mixing by the Ws is determined by the CKM and PMNS matrices. They are 3×3 unitary matrices that, after accounting for the various constraints imposed by the fermion masses, unitarity, etc., have 4 free parameters each. They are most often parametrized as three real Euler angles and one complex phase, which again have to be determined experimentally. Importantly, this complex phase means the weak interaction violates CP-symmetry as well!

To see why, we can go back to the Yukawa interactions, this time with the hermitian conjugates included:

$$\mathcal{L}_{\text{Yukawa}} \sim y^d_{ij}\bar{d}^i_L d^j_R + y^{d*}_{ij}\bar{d}^i_R d^j_L + \ldots \tag{4.2.25}$$

One can check that the two field terms $\bar{d}^i_L d^j_R$ and $\bar{d}^i_R d^j_L$ are CP-conjugates of each other, which means invariance under CP requires $y^d_{ij} = y^{d*}_{ij}$, i.e. the Yukawa matrix to be real. The complex phases in the CKM and PMNS matrices therefore lead to CP-violation in the SM. Interestingly, for fewer than three generations, the two matrices cannot be imaginary. Kobayashi and Maskawa discovered this in 1973, after the observation of CP-violation in 1964, and predicted a third generation of quarks to explain CP-violation in the SM, for which they were awarded the Nobel prize in 2008.

Flavor is perhaps the least understood area of the SM: Why are their exactly three generations each of quarks and leptons? Why their particular hierarchy of masses? Is it a coincidence that nature chose the exact minimum number of generations needed to allow for CP-violation? All of these mysteries point to the strong possibility of new physics in the flavor sector.

4.3 The Higgs sector

The Higgs boson, being the only scalar in the SM and uncharged under the $\mathrm{U}(1)_{\mathrm{EM}}$ and $\mathrm{SU}(3)_C$ symmetries, may appear to be the simplest particle in the theory. However, these same properties also mean that the Higgs sector is not as strongly constrained by gauge invariance, renormalizability, etc. as the gauge and fermionic sectors. Indeed, the Higgs sector contains the majority of the free parameters of the SM: the Yukawa couplings (12 masses of the fermions + 8 more parameters from the CKM and PMNS matrices), the Higgs VEV, and the Higgs mass (or, equivalently, λ in the Higgs potential). Without it, the SM would only have three free parameters: the three forces' coupling constants!

This is why a significant motivation for the next decades of the LHC, as well as future "Higgs factory" colliders, is to precisely characterize the Higgs sector. In this section, we first describe how this is possible at the LHC and discuss recent experimental constraints. We then motivate measurements of Higgs pair production, both in the SM and through BSM decays of heavy resonances, which are the focus of this dissertation and a key target of the current and upcoming LHC physics program.

4.3.1 Higgs boson production and measurements at the LHC

Higgs bosons are produced at the LHC through a variety of parton-parton interactions, as shown in Figure 4.17. Because of their high mass, they have a lifetime of roughly $\mathcal{O}(10^{-22})$s and decay immediately into two vector bosons or two fermions at tree-level, with further decays possible through loops. The decay probability depends on the strength of the respective interactions, which we see from Section 4.2 are proportional to the mass or the mass squared for fermions and vector bosons, respectively, though with the probability lowered for decays that are not kinematically accessible (i.e., when the total mass of the decay products is greater than the Higgs').

This is illustrated in Figure 4.18, which shows the branching fractions (BFs) of the Higgs boson as a function of its mass. Generally, we see the higher the mass of the decay product, the higher the BF; however, as the Higgs mass decreases, first the $\mathrm{t\bar{t}}$ and later the W and Z boson decays become kinematically inaccessible, leading to decreasing decay probabilities.

The Higgs boson was initially observed by the CMS and ATLAS experi-

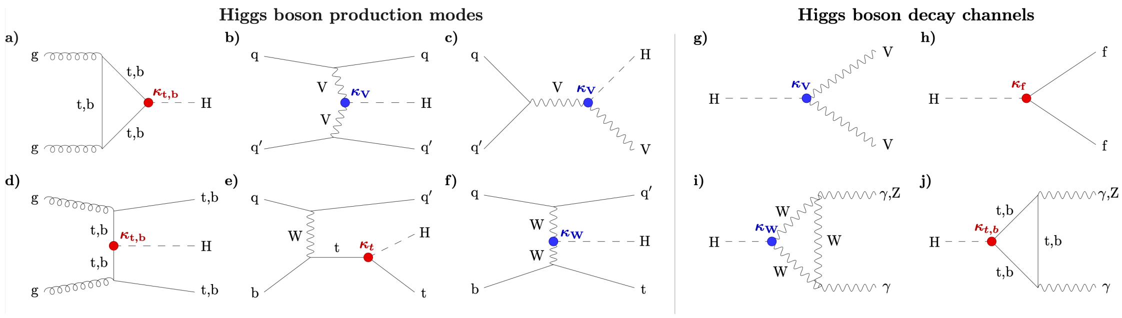


Figure 4.17. Single Higgs boson production modes and decay channels at the LHC, reproduced from Ref. [23].

ments in 2012 through a combination of several decay channels. Since then, the two experiments have been making steady progress in the precise measurements of the various Higgs properties. For example, Figure 4.19 shows the overall constraints on the Higgs to fermion and vector boson couplings and the Higgs mass by the CMS experiment. Constraints are based on the κ-framework [151], where κ_X scales the Higgs-X coupling strength with $\kappa_X = 1$ corresponding to the SM prediction. Changes to the coupling strength due to new physics are thus generically captured by deviations from $\kappa_X = 1$.

4.3.2 Higgs pair production in the SM

Two couplings of the Higgs boson which have not been well-constrained are the trilinear Higgs self-coupling (HHH), with coupling modifier κ_λ, and the Higgs quartic coupling to vector bosons (HHVV), with modifier κ_{2V}. As discussed in Section 4.2.3 and illustrated in Figure 4.20, measuring the Higgs self-coupling in particular is necessary to fully characterize the Higgs potential, deviations

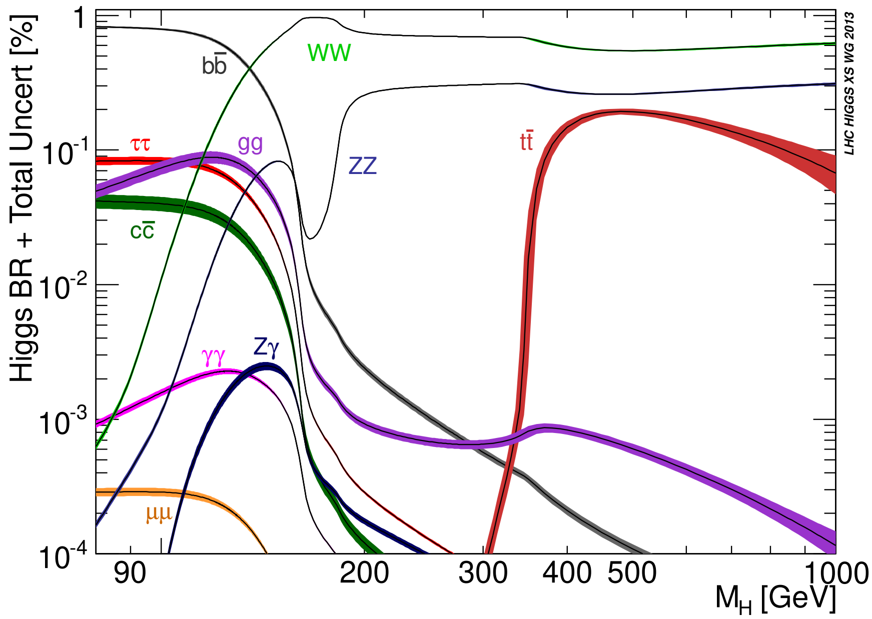


Figure 4.18. Higgs branching fractions predicted in the SM as a function of m_H (reproduced from Refs. [24, 25]).

to which could hint at BSM explanations to mysteries such as baryon asymmetry [152]. As we describe below, both couplings can be probed exclusively through Higgs pair production (HH), which is why it is a key physics target for the upcoming high-luminosity era of the LHC.

HH production in the SM occurs dominantly through gluon fusion (ggF), with a small production cross section $\sigma_{\mathrm{ggF}} = 31.05^{+2.2\%}_{-5.0\%} \pm 3\%(\mathrm{PDF} + \alpha_s)^{+4\%}_{-18\%}(m_{\mathrm{t}})$ fb [153, 154] at a center of mass energy of 13 TeV and $m_{\mathrm{H}} = 125$ GeV, and subdominantly through vector boson fusion (VBF), with a smaller production cross section $\sigma_{\mathrm{VBF}} = 1.726^{+0.03\%}_{-0.04\%} \pm 2.1\%(\mathrm{PDF}+\alpha_s)$ fb [25]. At leading order, the ggF production mode has contributions from diagrams that involve the trilinear HHH

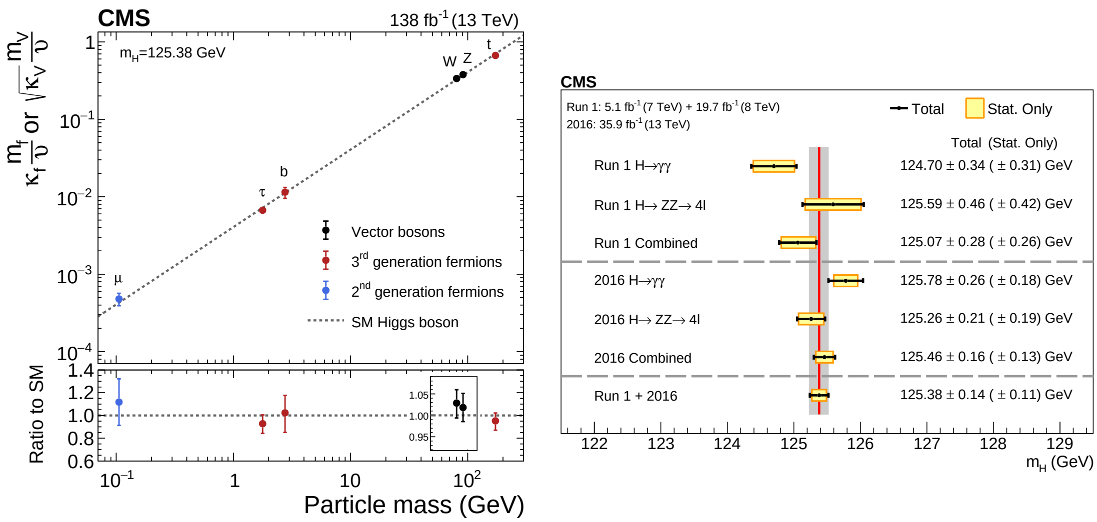


Figure 4.19. Constraints on Higgs to fermion and vector boson couplings, reproduced from Ref. [23] (left) and measurements of the Higgs mass, reproduced from Ref. [26] (right) by the CMS experiment.

Higgs self-coupling and the emission of two Higgs bosons through a top quark loop, while the VBF production mode has contributions from three diagrams involving the trilinear HHH, HVV, and quartic HHVV couplings (Figures 4.21 and 4.22). It also features the distinct final state signature of two, typically forward, jets in addition to the two Higgs bosons.

The production cross section and kinematic properties of the HH system are altered if values of the Higgs self-coupling, the top Yukawa coupling, and/or the quartic HHVV coupling are modified due to beyond the SM (BSM) effects. Notably, at the energy scale of the LHC, the leading contribution to the VBF production amplitude is the scattering of longitudinal vector bosons, which scales as $\sim m_{\mathrm{HH}}^2(\kappa_{2\mathrm{V}} - \kappa_{\mathrm{V}}^2)$ [155], where, as above, κ_λ, $\kappa_{2\mathrm{V}}$, and κ_{V} are defined to be multiplicative modifiers of the HHH, HHVV, and HVV couplings from their

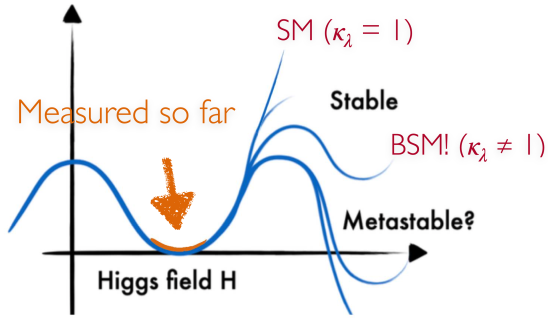


Figure 4.20. Cartoon of the Higgs potential in the SM and potential deviations due to BSM physics.

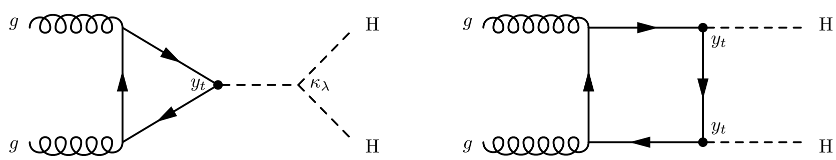


Figure 4.21. Leading-order diagrams for nonresonant HH production via gluon gluon fusion.

SM values, respectively.

In the SM, with $\kappa_{2V} = \kappa_V = 1$, VBF production is suppressed since the left-most $(HVV)^2$ and right-most HHVV VBF diagrams in Figure 4.22 cancel; however, BSM deviations to HHVV can spoil the cancelation, significantly enhancing this mode. This departure from the SM could be more visible at high energies, as illustrated in Figure 4.23, which shows the increase and shift towards higher m_{HH} of the differential VBF HH production cross section for enhanced and reduced κ_{2V} values. Thus, measuring high-m_{HH} nonresonant VBF HH production, with

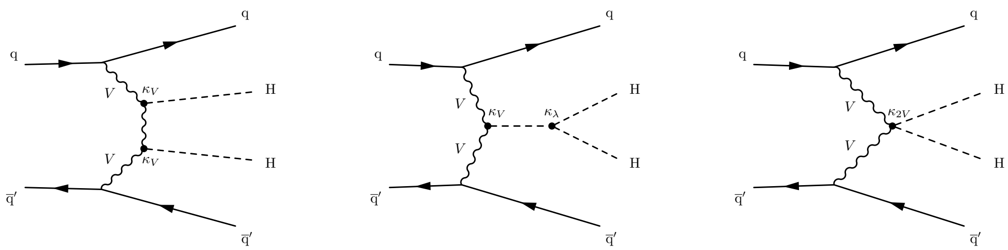


Figure 4.22. Leading-order diagrams for nonresonant HH production via vector boson fusion. In this chapter, we refer to the left-most VBF diagram as the $(\mathrm{HVV})^2$ and the right-most as the HHVV diagram.

both Higgs bosons highly Lorentz-boosted, is a powerful probe of the HHVV coupling.

This is evidenced by the current $\kappa_{2\mathrm{V}}$ constraint in CMS being dominated by the search for boosted HH in the $\mathrm{b\bar{b}b\bar{b}}$ channel, with an observed (expected) 95% confidence level (CL) constraint of $[0.6, 1.4]$ ($[0.7, 1.4]$), excluding $\kappa_{2\mathrm{V}} = 0$ for the first time [156]. This is followed by CMS searches in the resolved $\mathrm{b\bar{b}b\bar{b}}$ [28] and $\mathrm{b\bar{b}}\tau\tau$ [29] channels, with constraints of $[-0.1, 2.2]$ ($[-0.4, 2.5]$) and $[-0.4, 2.6]$ ($[-0.6, 2.8]$), respectively. Similarly, the strongest $\kappa_{2\mathrm{V}}$ constraint from the ATLAS experiment is from the boosted $\mathrm{b\bar{b}b\bar{b}}$ search [157], with an observed (expected) 95% CL constraint of $[0.55, 1.49]$ ($[0.3, 1.7]$).

The success of searches in the boosted $\mathrm{b\bar{b}b\bar{b}}$ channel motivates further exploration of high-m_{HH} HH production. This dissertation presents the first search in the all-hadronic $\mathrm{b\bar{b}VV}$ channel, where one Higgs boson decays to $\mathrm{b\bar{b}}$ while the other to WW or ZZ, where $\mathrm{W} \to \mathrm{q\bar{q}}$ and $\mathrm{Z} \to \mathrm{q\bar{q}}$. The branching fractions for the $\mathrm{b\bar{b}}$ and all-hadronic VV decays are 0.58 and 0.11 respectively, for

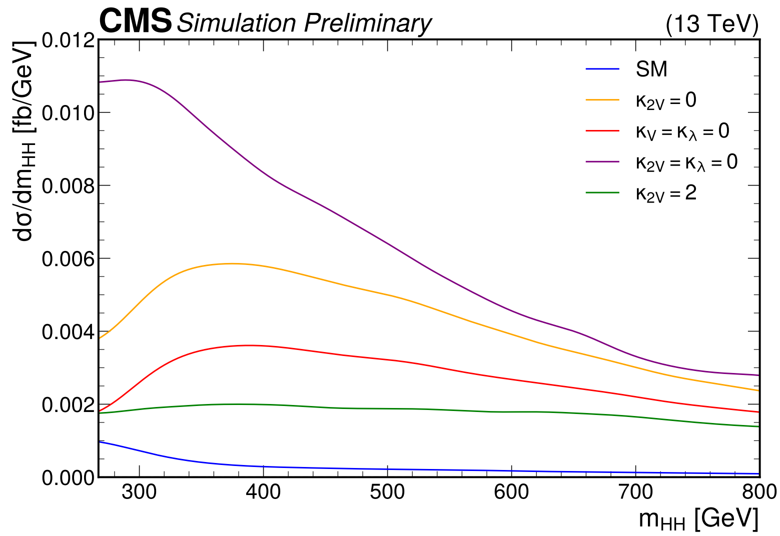


Figure 4.23. Differential cross section at 13 TeV center of mass for VBF HH production as a function of the invariant mass of the HH system (m_{HH}) for different diagrams and couplings.

a total branching fraction $\mathcal{B}(\mathrm{HH} \to \mathrm{b\bar{b}}(\mathrm{VV} \to 4\mathrm{q})) = 2 \cdot 0.58 \cdot 0.11 = 0.13$, which is the second largest behind $\mathrm{b\bar{b}b\bar{b}}$. The analysis primarily aims to constrain $\kappa_{2\mathrm{V}}$ and also sets an exclusion limit on the inclusive HH production cross-section. It is not expected to be sensitive to κ_λ because of the focus on the high-m_{HH} regime.

Another benefit of the high-m_{HH} regime is the significantly reduced QCD multijet background, which otherwise makes such all-hadronic searches extremely challenging. Because of the two Higgs bosons' high Lorentz-boosts, this regime also features the unique experimental signature of the $\mathrm{b\bar{b}}$ and $\mathrm{VV} \to 4\mathrm{q}$ decays each being reconstructed as single wide-radius jets. Such merged $\mathrm{H} \to \mathrm{b\bar{b}}$ jets have been identified to great effect in CMS using deep neural networks (DNNs) [156, 158], but attaining similar signal versus background discrimination for H → VV jets remains an open challenge. To this end, we introduce a new attention-based DNN, referred to as the global particle transformer (GloParT) to not only enable this search but open new possibilities for searches in boosted-VV channels as well (Chapter **??**).

4.3.3 Experimental status of HH measurements with CMS

H1 \ H2	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

Figure 4.24. HH decays and their respective branching fractions (reproduced from Ref. [27]).

The decays and branching fractions (BFs) of the Higgs boson pairs are shown in Figure 4.24. Three of these final states have emerged as experimental "golden channels" — the channels expected to yield the highest signal-to-background-events ratio for SM HH production:

- $b\bar{b}b\bar{b}$: This channel has the highest BF (34%) and, despite the large QCD multijet background due to the all-hadronic final state, it benefits from unique signatures of heavy-flavor b-jets, such as the presence of secondary vertices and displaced tracks due to the long lifetimes of b-hadrons. Both the resolved [28] and boosted [156] Run 2 CMS analyses (Figure 4.25) have been highly effective, with the latter benefitting from the high BF of this decay mode, the exponential reduction of the QCD multijet background in the boosted regime, and significant recent advances in $b\bar{b}$-jet classification and reconstruction (as will be discussed in Chapter **??**).

- $b\bar{b}\tau\tau$: This has an intermediate BF of 7% but relatively lower background of primarily Drell-Yan (Z/γ^*), top quark pair production ($t\bar{t}$), and QCD multijet events (Figure 4.26, reproduced from Ref. [29]). It benefits from similar deep learning techniques for b-jet tagging, and targets all-hadronic ($\tau_h\tau_h$) and semi-leptonic ($\tau_h\tau_e$ or $\tau_h\tau_\mu$) τ-lepton decays using a variety of traditional and ML techniques.

- $b\bar{b}\gamma\gamma$: Despite the small BF (0.3%) of this channel, the $H \to \gamma\gamma$ decay provides a clean experimental signature with a sharply peaking resonance over a small background of QCD multijet + γ events (Figure 4.27, from Ref. [30]).

More recently, the $b\bar{b}WW$ channel has been explored in the douple-lepton (ll$\nu\nu$) and single-lepton (lνqq) WW final states [31], which have a large combined BF of 13.4%. The former features a clean experimental signature of two opposite-sign leptons but a small BF of 2.6%, while the latter has the higher BF of 10.8% but larger top quark background as well (Figure 4.28). Because of this, the two channels have similar sensitivities to the HH cross section.

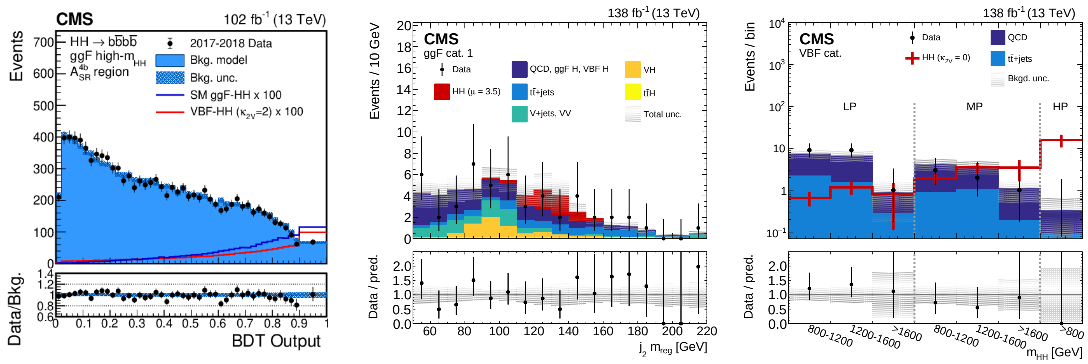


Figure 4.25. Distribution of events in the high-m_{HH} ggF category of the Run 2 CMS HH $\to$ b$\bar{b}$b$\bar{b}$ resolved analysis [28] as a function of the BDT discriminant (left), and of the Run 2 boosted analysis' [156] most sensitive ggF category, as a function of the second-highest tagged b$\bar{b}$-jet's mass (middle), and VBF categories, as a function of m_{HH} (right).

The limits set on the HH cross section by each channel, and their combinations, are shown in Figure 4.29, and as a function of κ_λ and κ_{2V} in Figure 4.30. The three "golden channels" each offer roughly similar sensitivities to the cross section and κ_λ limits; however, the constraint on κ_{2V} is dominated by the boosted b$\bar{b}$b$\bar{b}$ channel, because of the enhancement of boosted HH production at BSM κ_{2V} deviations, as discussed in Section 4.3.2. Its observed (expected) 95% confi-

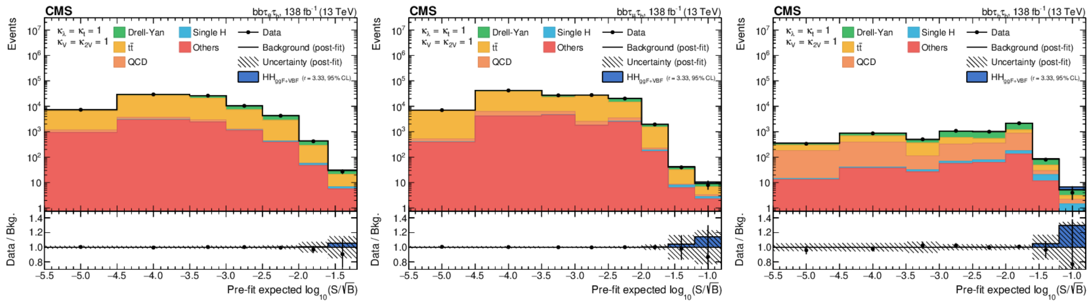


Figure 4.26. Combination of bins of all postfit distributions of the Run 2 CMS $HH \to b\bar{b}\tau\tau$ analysis [29], ordered according to the expected signal-to-square-root-background ratio, separately for the $\tau_h\tau_e$ (left), the $\tau_h\tau_\mu$ (center), and $\tau_h\tau_h$ (right) channels.

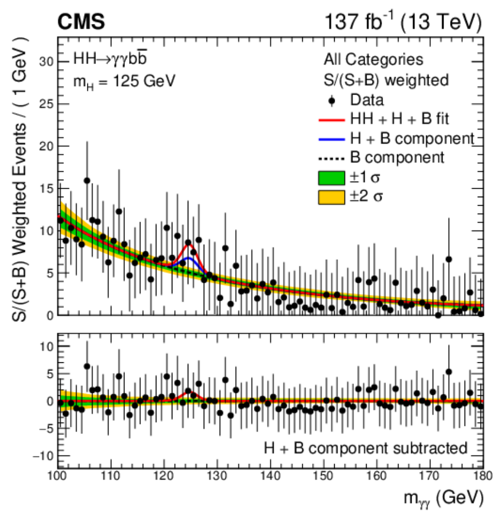


Figure 4.27. Invariant two-photon mass distribution of the Run 2 CMS $HH \to b\bar{b}\gamma\gamma$ analysis [30], combined for all signal categories, weighted by S/(S+B), where S (B) is the number of signal (background) events extracted from the signal-plus-background fit.

dence level (CL) constraint is $[0.6, 1.4]$ ($[0.7, 1.4]$). This is followed by the resolved $b\bar{b}b\bar{b}$ [28] and $b\bar{b}\tau\tau$ [29] channels, with constraints of $[-0.1, 2.2]$ ($[-0.4, 2.5]$) and $[-0.4, 2.6]$ ($[-0.6, 2.8]$), respectively. Similarly, the strongest κ_{2V} constraint from the ATLAS experiment is from the recent boosted $b\bar{b}b\bar{b}$ search [157], with an

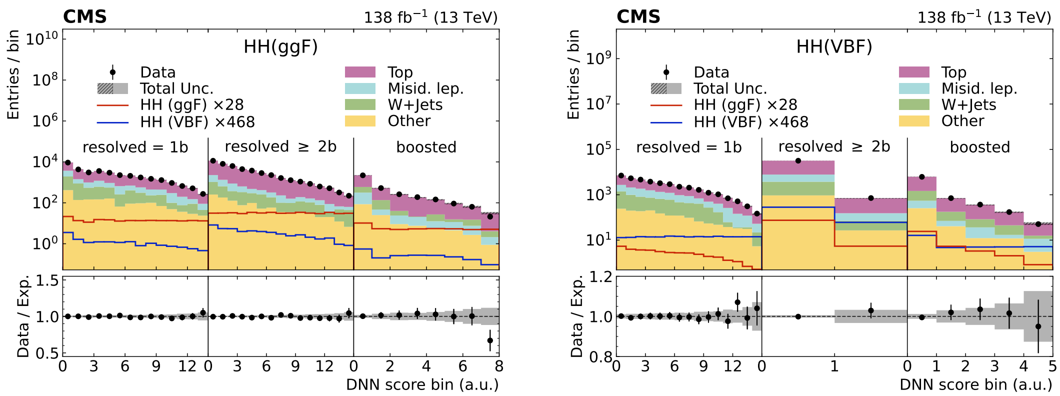


Figure 4.28. Distribution of events in the resolved 1b, resolved ≥ 2b, and boosted signal categories of the Run 2 CMS semi-leptonic HH → $b\bar{b}WW$ analysis [31] as a function of their DNN discriminant in the single-lepton (left) and double-lepton (right) final states.

observed (expected) 95% CL constraint of $[0.55, 1.49]$ ($[0.3, 1.7]$).

In this dissertation, we present the first search for nonresonant HH production in the *all-hadronic $b\bar{b}VV$ channel*, where one Higgs decays to two bottom quarks, while the other to two vector bosons (VV) both decaying hadronically to the four quark (4q) final state. Both the W and Z bosons are considered for the latter decay and collectively referred to as V bosons. The branching fractions for the $b\bar{b}$ and VV decays are 0.58 and 0.25 respectively, for a total branching fraction $\mathcal{B}(\mathrm{HH} \to b\bar{b}\mathrm{VV}) = 2 \cdot 0.58 \cdot 0.25 = 0.29$, which is the second-highest, behind only $b\bar{b}b\bar{b}$. The all-hadronic final state in particular has a branching fraction of 0.13. The analysis targets the boosted regime, which, as discussed above, has the two-fold advantage of 1) increasing sensitivity to κ_{2V} deviations and 2) exponentially reducing the dominant QCD multijet background.

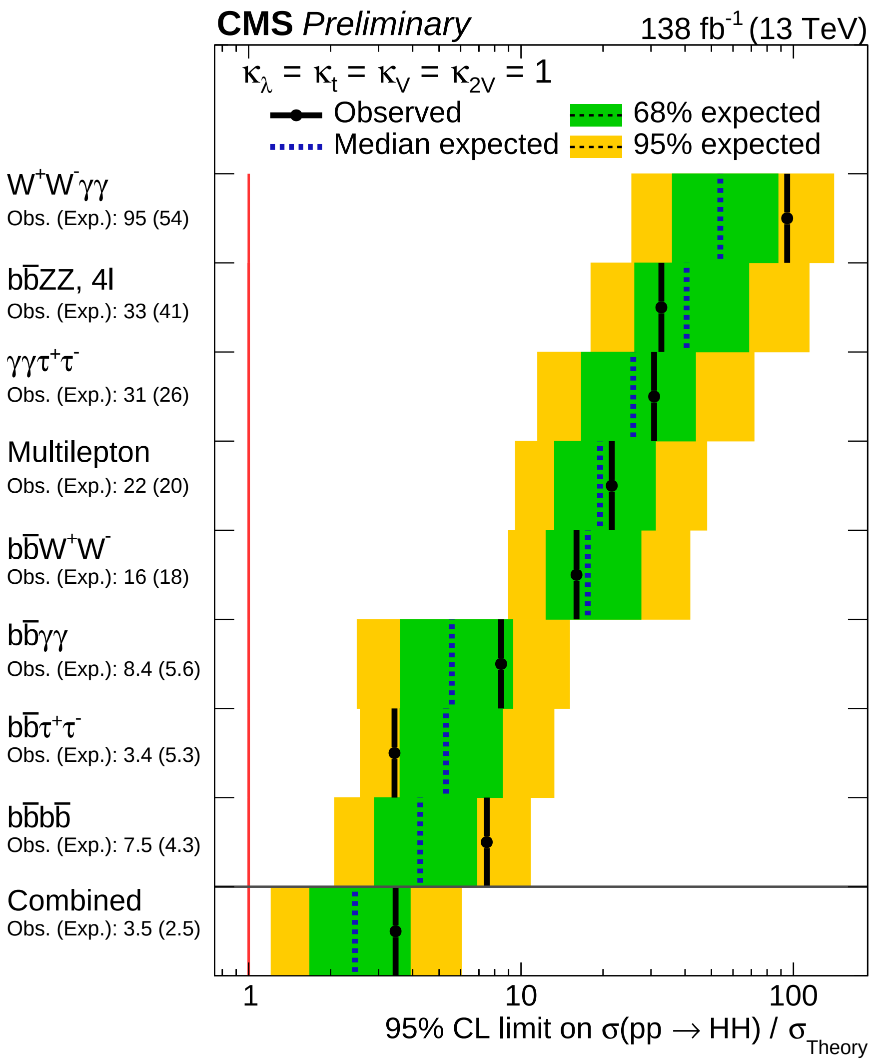


Figure 4.29. The expected and observed limits on the ratio of experimentally estimated production cross section and the expectation from the SM in searches using different final states and their combination, reproduced from Ref. [32]. The search modes are ordered, from upper to lower, by their expected sensitivities from the least to the most sensitive. The overall combination of all searches is shown by the lowest entry.

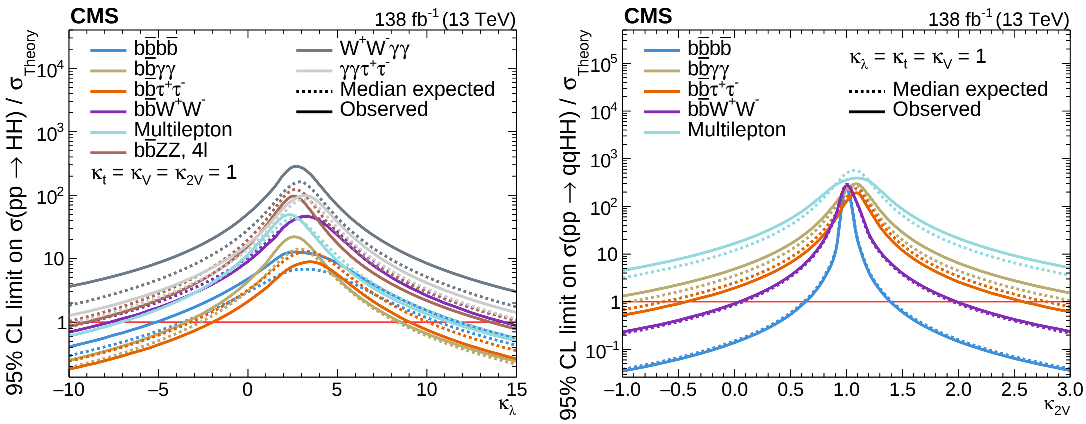


Figure 4.30. The expected and observed limits on the ratio of experimentally estimated production cross section and the theory expectation for different values of κ_λ (left) and κ_{2V} (right), reproduced from Ref. [32].

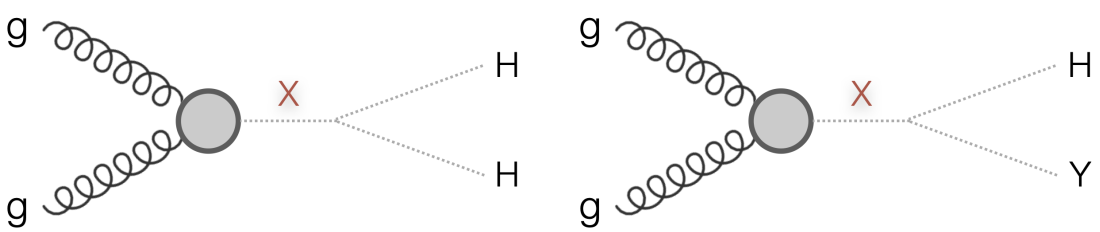


Figure 4.31. X → HY production in the symmetric (left) and asymmetric (right) cases.

4.3.4 BSM X → HY production

Many theoretical models predict a richer scalar sector than that in the SM to address aesthetic and observational inconsistencies with the SM, such as the Higgs mass hierarchy problem and the baryon asymmetry discussed above. These include two-Higgs doublet models (2HDM) [159] that add an additional scalar doublet to the SM, such as the minimal supersymmetric extension of the SM (MSSM) [147], which predicts two neutral CP-even scalars (H, h), one neutral CP-odd scalar (A), and two charged scalars ($H^{\pm}$), where one of the neutral CP-even scalars may be the discovered SM Higgs H_{125}. The next-to-minimal supersymmetric extension of the SM (NMSSM) [160] adds to this a complex scalar singlet, predicting two more CP-even (h_s) and CP-odd (a_s) neutral scalars. Finally, the two-real-singlet-model (TRSM) predicts two additional CP-even scalar fields. Depending on the kinematics, all these models allow for cascade decays of a heavier scalar to symmetric and asymmetric lighter scalars, such as $H \to H_{125}H_{125}$ and $H \to hH_{125}$, respectively, as shown in Figure 4.31.

We search for this broad class of signals, looking for generic decays of the form $X \to HY$, where X is the heavier and Y the lighter scalar resonance, with H decaying to $b\bar{b}$ and Y to $VV \to 4q$. Many models, such as the TRSM, predict branching ratios for the lighter scalar similar to or the same as the SM Higgs. In this case, the VV decay modes are dominant for $m_Y > 140$ GeV (Figure 4.18) and, hence, the $H \to b\bar{b}$ and $Y \to VV$ will be the dominant final states for the $X \to HY$ signal. Thus, the $b\bar{b}VV$ channel represents the highest BF in these models.

There are several published and ongoing CMS searches for $X \to HY$ production in a variety of regimes and final states with the Run 2 dataset, such as the boosted [161] $b\bar{b}b\bar{b}$ final state, the symmetric-only $b\bar{b}WW$ semi-leptonic final state [31], the resolved $b\bar{b}\gamma\gamma$ [162], and the resolved $b\bar{b}\tau\tau$ [163] final state. This dissertation presents the first search in the $b\bar{b}VV$ all-hadronic state, and the first in the $b\bar{b}VV$ state for the asymmetric case, representing a significant increase in the covered phase space for X → HY searches.

The search comprises two distinct topologies depending on the ratio of the X and Y masses: a highly-boosted fully-merged Y → VV topology for $m_X \gg m_Y$, with both VV bosons' decay products highly collimated into a single wide-radius jet; and a relatively less-boosted semi-merged topology, where the VV bosons are well separated and each V → qq decay is reconstructed as its own wide-radius jet. These two phases are illustrated in Figure 4.32, showing the fraction of Y → VV jets containing three or four generator-level quarks as a function of the X and Y boson masses, with the transition occurring around $m_X \approx 10 m_Y$. This dissertation focuses on a search for the fully-merged topology only, i.e. for $m_X \gtrsim 10 m_Y$, and is complementary to an ongoing CMS search in the semi-merged topology. Thus, in terms of the analysis strategy and techniques, this search is similar to the boosted nonresonant HH search in that they both target highly-boosted Higgs boson decays with single wide-radius jets for both H or Y bosons.

Acknowledgements

Chapters 4.3.2 and 4.3.4 are in part, currently being prepared for the publication of the material by the CMS collaboration. The dissertation author was the primary investigator and author of these papers.

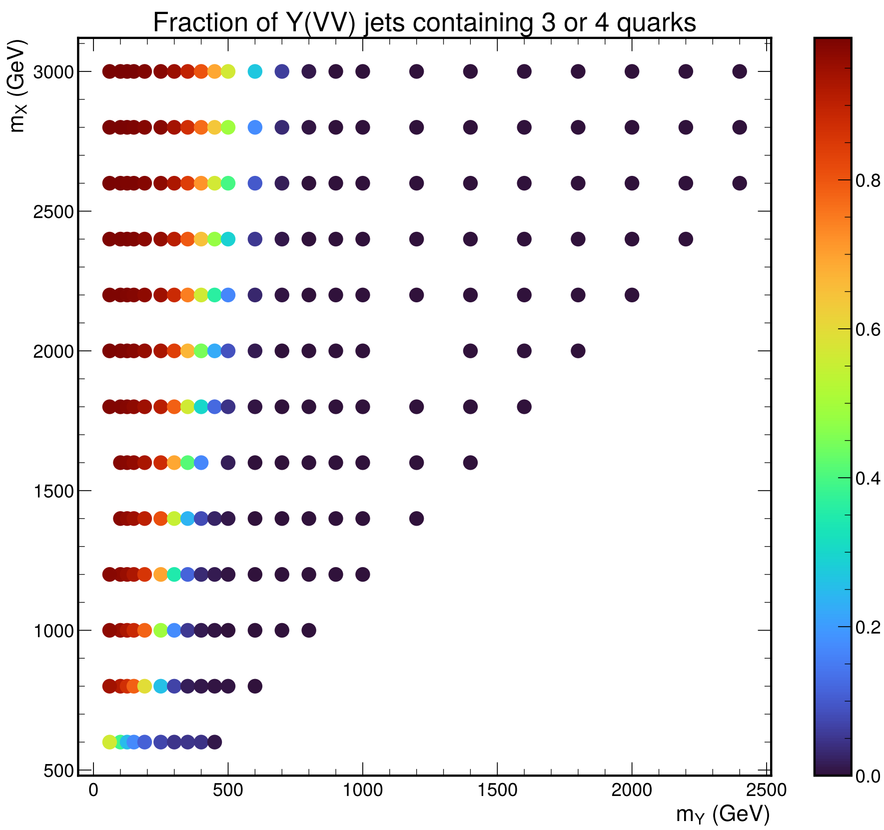


Figure 4.32. The fraction of $Y \to VV$ jets containing three or four generator-level quarks for the resonant $X \to (H \to b\bar{b})(Y \to VV)$ signal as a function of the X and Y boson masses.

Part II

Experimental Background

Chapter 5

The CERN Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton collider located at CERN on the border of Switzerland and France (Figure 5.1). It is the largest and highest energy particle accelerator in the world, with a circumference of 27.6 km and a center-of-mass (COM) energy of 13.6 TeV, reproducing energies in the universe 10^{-11} seconds after the Big Bang.

The tunnel was initially built for the large electron-positron collider (LEP), which operated from 1989 to 2000. Being point particles and not interacting with the strong force, electrons and positrons produce "clean" collisions (i.e., with low background) and can be simulated with relative ease; thus, LEP allowed high precision measurements of the electroweak sector of the standard model (SM), as discussed in Chapter 4. The drawback, however, is that due to the power loss from synchotron radiation, which scales as $\propto$ (mass of the accelerated particle)$^{-4}$, their low mass limits the COM energy that can be attained with electron-positron

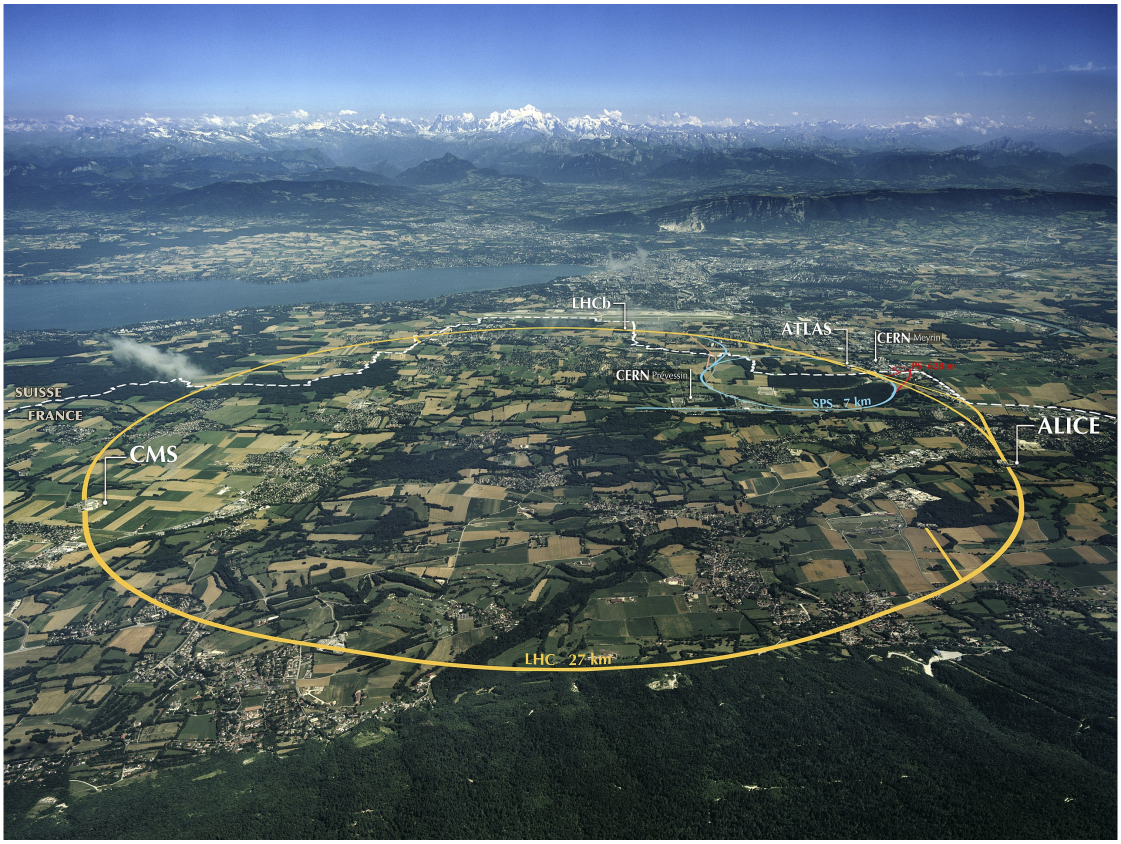


Figure 5.1. Outline of the LHC overlaid on a satellite image of Switzerland and France.

colliders.

Protons, on the other hand, are composite particles and produce “noisy”, high-multiplicity collisions, but are 2000× more massive and, hence, can be accelerated to much higher energies. This is why, from early on, the LEP tunnel had also been proposed as a site for a future *hadron-hadron* collider, which could achieve an order-of-magnitude greater energy than the previous energy-frontier machine, the Fermilab Tevatron. The LHC was eventually approved in 1994 and was built in collaboration with over 100 countries at CERN between 1998 and 2008. It is primarily a proton-proton collider, designed with the goal of accelerating each proton to 7 TeV, for a COM energy of 14 TeV, to explore the TeV energy scale for the first time. It also, less frequently, collides heavy ions to study QCD and the quark-gluon plasma inside nuclei.

The collisions occur at four interaction points around the ring and are observed by a total of nine detectors: two large general-purpose detectors, CMS and ATLAS, two more specialized detectors, ALICE and LHCb, for heavy-ion- and b-physics, respectively, and five smaller scale experiments, TOTEM, LHCf, MoEDAL, FASER, and SHiP. In this section, we describe the LHC accelerator in Section 5.1 and the overall number of collisions, quantified as "integrated luminosity", it has delivered and expects to deliver in Section 5.2.

5.1 The accelerator

The overall LHC accelerator complex is shown in Figure 5.2. Protons are first extracted from a hydrogen gas bottle through a duoplasmatron ion source [164] as a low energy beam of around 100 keV. They are then accelerated through a series of "injectors" (Figure 5.3): first through a linear accelerator (LINAC) up to 50 MeV; then a proton synchotron booster (PSB) up to 1.4 GeV; the proton synchotron (PS) up to 26 GeV; and finally through the super proton synchotron (SPS) up to 450 GeV, after which the protons are transferred to the LHC ring.

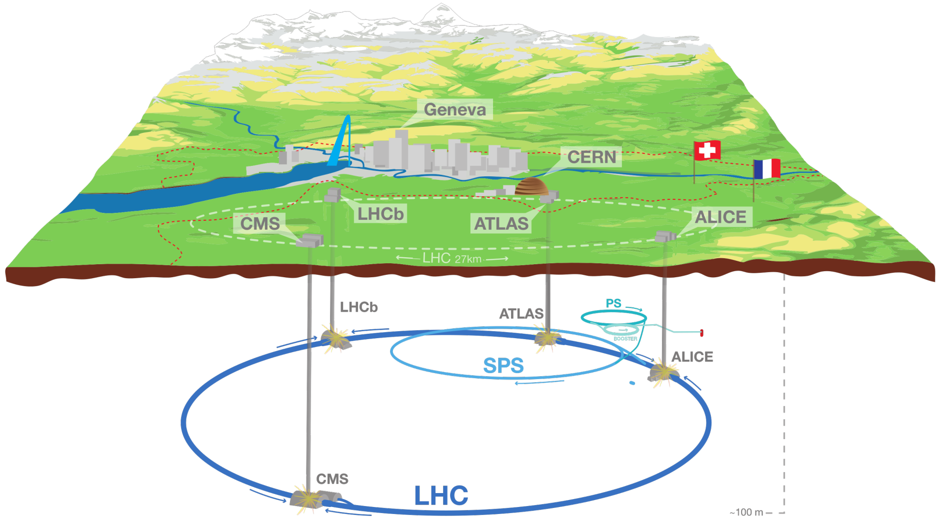


Figure 5.2. Diagram of the LHC accelerator complex adapted from Ref. [33], depicting the initial proton source (in red), LINAC, proton synchotron booster, PS, SPS, LHC, and the four main experiments: CMS, ATLAS, ALICE, and LHCb.

Unlike particle-antiparticle colliders, like the Tevatron, which can accelerate both beams in the same ring with the same magnet system, the proton-proton collisions at the LHC require opposite magnetic fields for the beams before their collision. The benefit, of course, is the ease of producing protons compared to antiprotons, allowing for far higher luminosities. Due to the small 3.7 m internal diameter of the existing LEP tunnel, it was not possible to install two separate rings for the two counter-rotating beams; instead, a twin-bore magnet design [165] was chosen to accommodate both in the same ring (Figure 5.4) with two separate vacuum chambers and superconducting coils.

A total of 1,232 such superconducting NbTi dipole magnets are installed around the ring to maintain the circular trajectory of the protons, as well as 392 quadrupole and higher multipole-order magnets to focus the beams. The

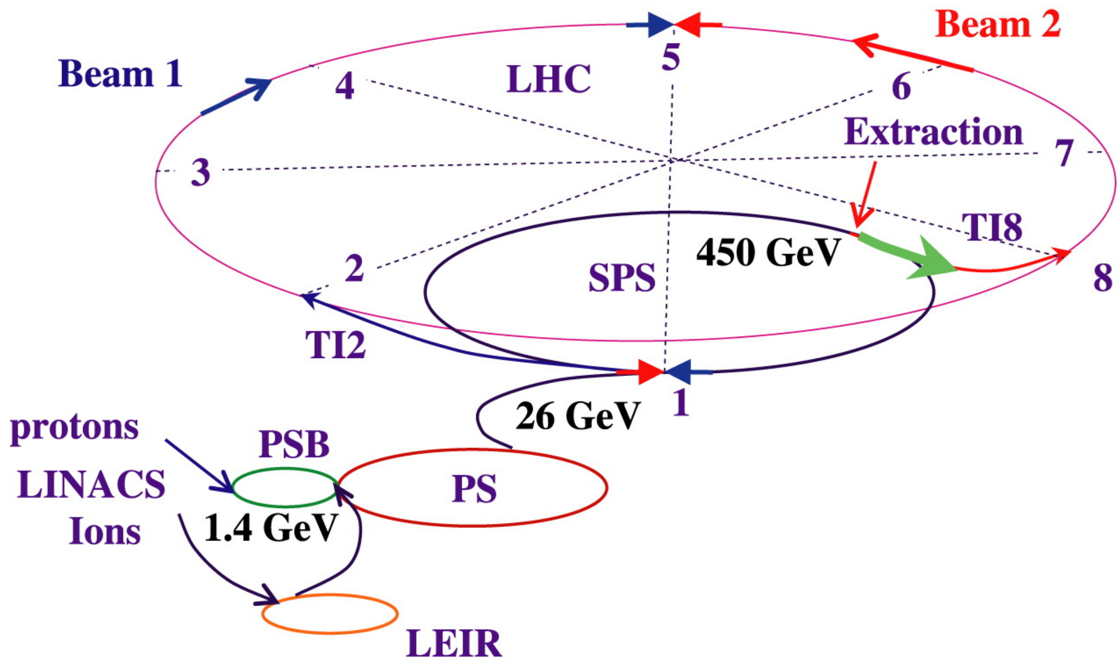


Figure 5.3. Schematic of the LHC injectors, reproduced from Ref. [34].

maximum beam momentum p is limited by the bending radius (ρ) and the bending field strength (B) of the dipole magnets, as [34]:

$$p[\,\mathrm{GeV}/c] = B[\,\mathrm{T}]\rho[\,\mathrm{m}]/3.336. \tag{5.1.1}$$

For the LHC tunnel, ρ is 2.8 km; hence, to achieve 7 TeV protons, the dipole magnets were designed to achieve a field strength of 8.33 T (requiring liquid helium cooling to a temperature of 1.9 K to maintain superconductivity). However, due to imperfections in some magnets, the LHC initially operated at 3.5–4 TeV per beam in Run 1 (2010–2012), then 6.5 TeV in Run 2 (2015–2018), and currently 6.8 TeV in Run 3 (2022–2026).

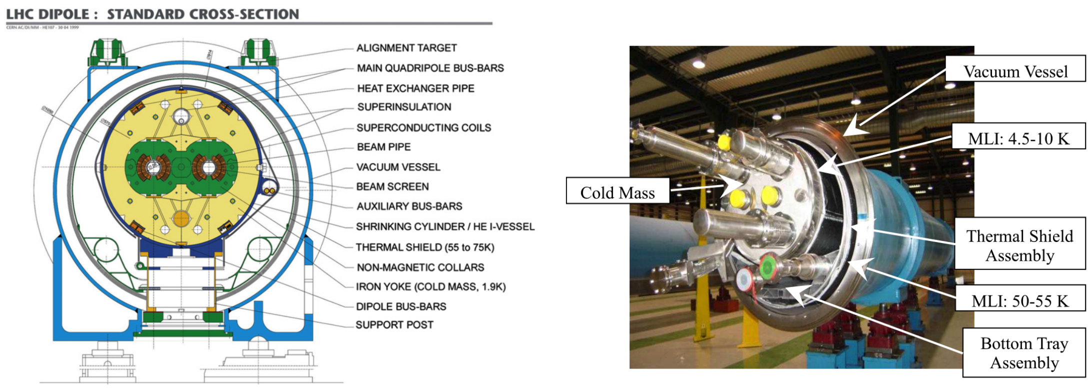


Figure 5.4. Diagram of the cross-section of the twin-bore LHC dipole magnets (left) and an image of an actual LHC dipole magnet (right), reproduced from Ref. [35].

The LHC layout comprises eight arcs and eight $\sim$ 500 m long straight sections. The two beams are diverted and collided in four of the straight sections, called "interaction points" (IPs), where the detectors are located (Figure 5.5). The other four straight sections are used for utilities, such as the beam dump and collimation systems.

The protons are accelerated and collided in "bunches" of 10^{11} protons each, with a separation of 25 ns between bunches. The greater the number of protons per bunch and frequency of bunches, the greater the total luminosity of the collider. Each bunch is accelerated and phase-focused longitudinally by a series of 16 superconducting radiofrequency (RF) cavities into separate "RF buckets". The RF-frequency of the cavities is 400 MHz, corresponding to a theoretical minimum spacing in time of 2.5 ns between RF buckets / bunches. The LHC opts for the 10-bucket spacing of 25 ns to avoid "parasitic" collisions between bunches [34]. With this spacing, the maximum number of bunches in the ring is 2808.

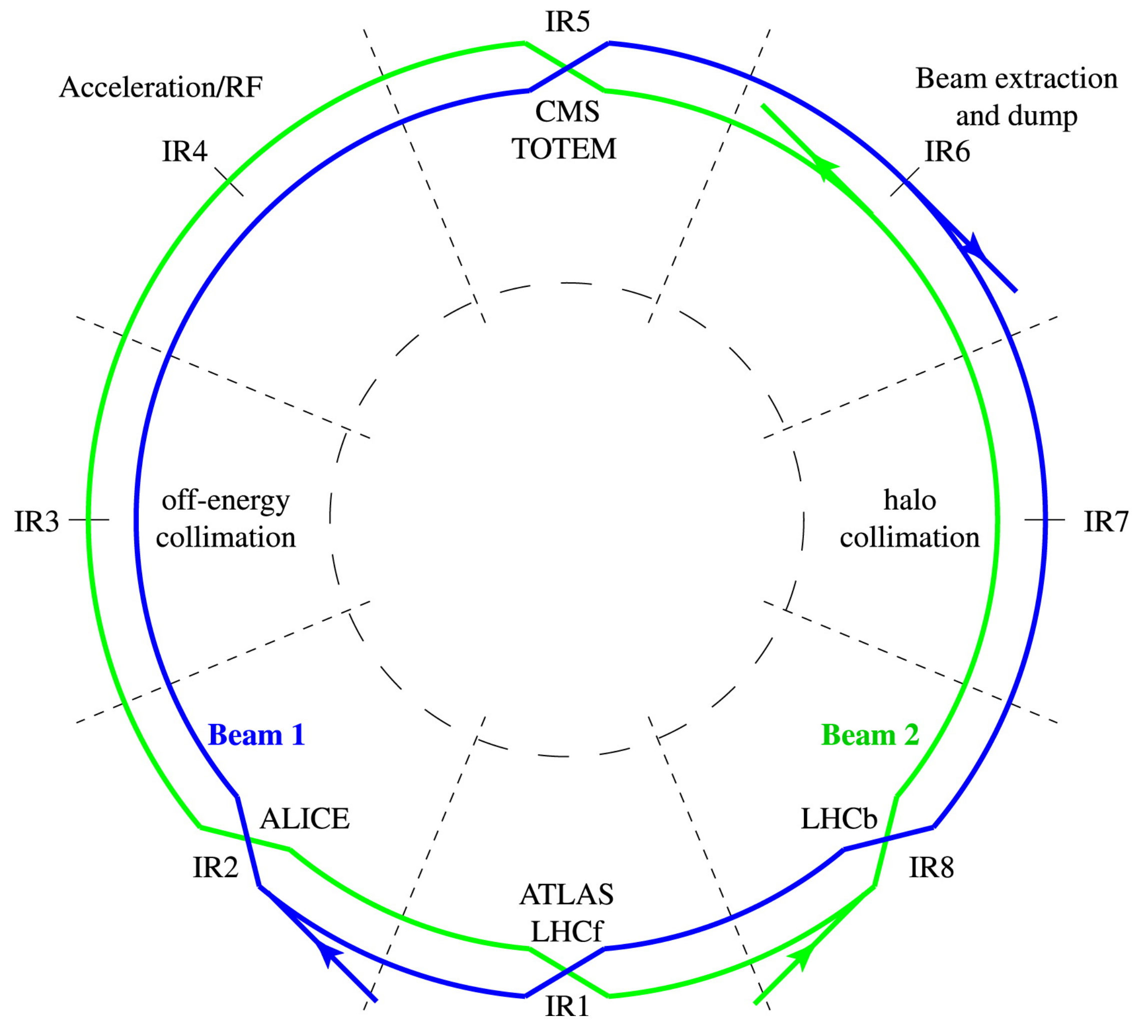


Figure 5.5. Schematic of the LHC layout showing the two proton beams in green and blue and its division into eight octants, reproduced from Ref. [34].

5.2 Luminosity and timeline

As discussed in Chapter 3, the number of scattering events we expect is the product of the scattering cross section and the *luminosity* (L) of the particle beams

(Eq. 3.2.16). Cross sections are typically given in units of barn (b), where 1 b = 10^{-28} m^2, and thus the luminosity often in inverse barns (b^{-1}). For a circular collider, the instantaneous luminosity is given by [34]:

$$L = \frac{N_1 N_2 n_b f_{\text{rev}}}{A}, \tag{5.2.1}$$

where N_1 and N_2 are the number of protons in each bunch, n_b is the number of bunches, f_{rev} is the revolution frequency of the beams, and A is the effective beam overlap area at the interaction point. This is why the LHC design aims to maximize the number of protons per bunch, the number of bunches, and the frequency of bunches, while focusing and aligning the beams as much as possible at the interaction point. The instantaneous luminosity of pp collisions at the LHC has increased steadily from a peak of 2.1×10^{32} cm^{-2}s^{-1} in 2010 to around 2.5×10^{34} cm^{-2}s^{-1} in 2022–24 [36]. Higher luminosity also leads to a higher rate of simultaneous pp collisions during a single bunch crossing, called *pileup*, which results in background noise to the detectors. As shown in Figure 5.6, the average rate of pileup in CMS has ranged from 10 in 2011 to around 57 in 2024.

The total luminosity delivered by the LHC is the integral of the above over time, called the *integrated luminosity*, and is shown in Figure 5.7 along with the projection up to 2041.[1] So far, the LHC has delivered around 60 fb^{-1} of integrated luminosity at 7 or 8 TeV COM to the CMS and ATLAS experiments in Run 1, 138 fb^{-1} at 13 TeV in Run 2, and is currently aiming for around 300 fb^{-1}

[1] Note that this projection has not been updated to reflect the decision made in September 2024 to extend Run 3 up to 2026 and delay the start of Run 4 to 2030.

at 13.6 TeV in Run 3. After this, (tentatively) between 2026 and 2030, the LHC will undergo a significant upgrade aiming to deliver an order of magnitude more luminosity in Runs 4–6, between $5\text{–}7 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ instantaneously, integrated to around $3000\,\mathrm{fb}^{-1}$! This is called the High-Luminosity LHC (HL-LHC) upgrade [**?**] (see Figure 5.8), and is expected to allow access to rare processes such as Higgs boson pair production; however, it also entails major accelerator, detector, and computational challenges to effectively deliver and exploit the increased luminosity.

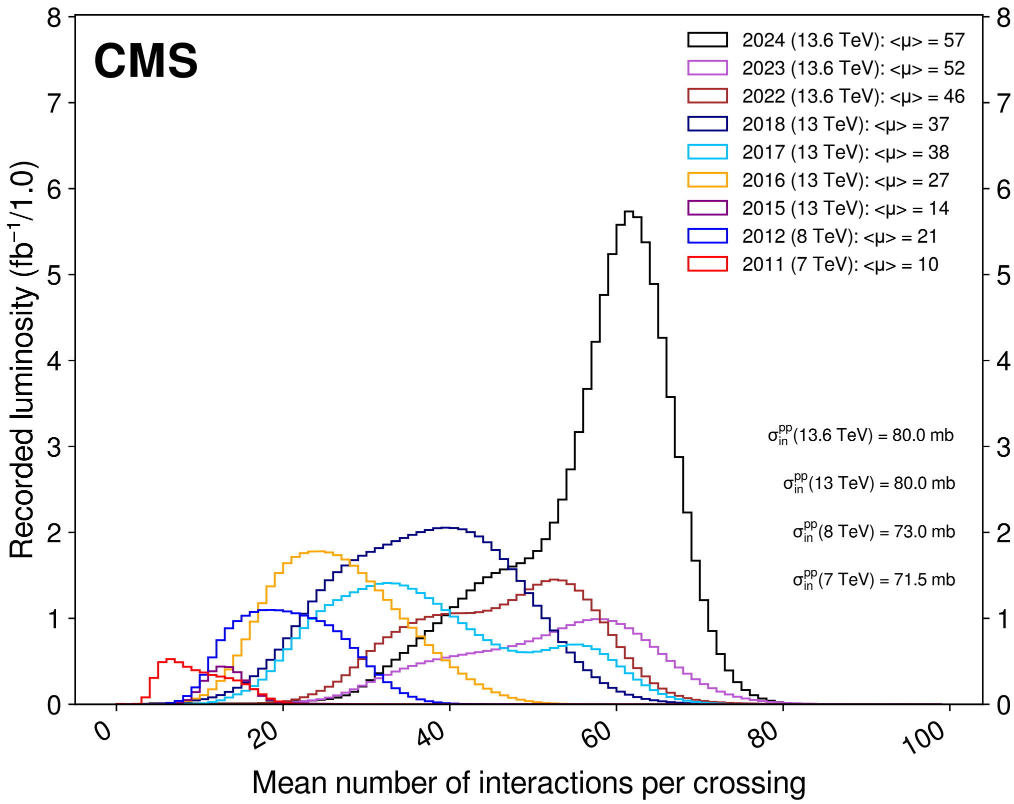


Figure 5.6. Mean number of interactions per crossing (pileup) in CMS between 2011–2024, reproduced from Ref. [36].

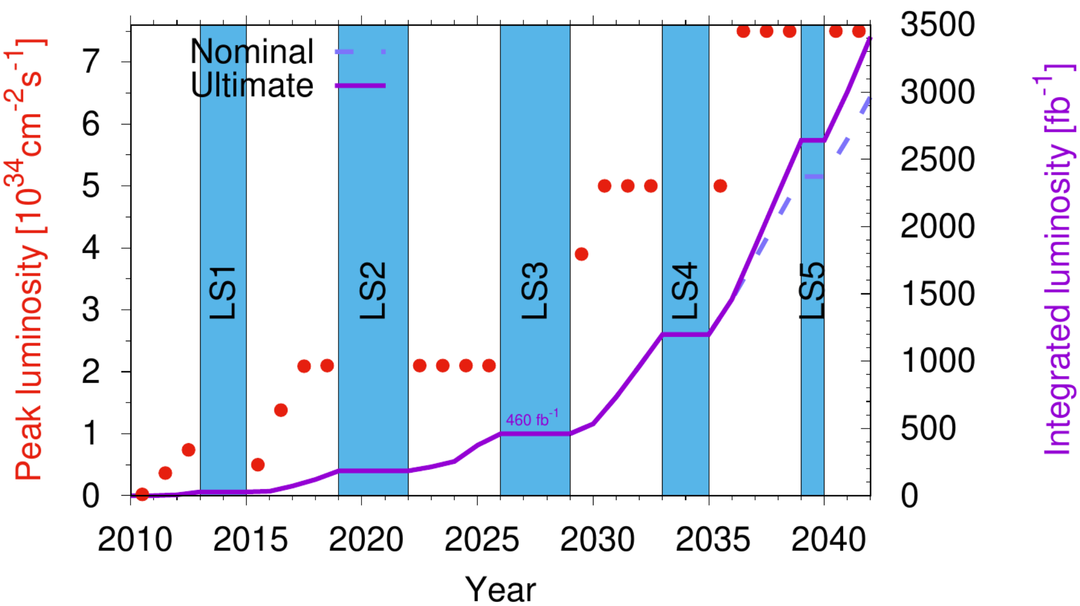


Figure 5.7. Integrated luminosity delivered by the LHC so far and the projection up to 2041, reproduced from Ref. [37].

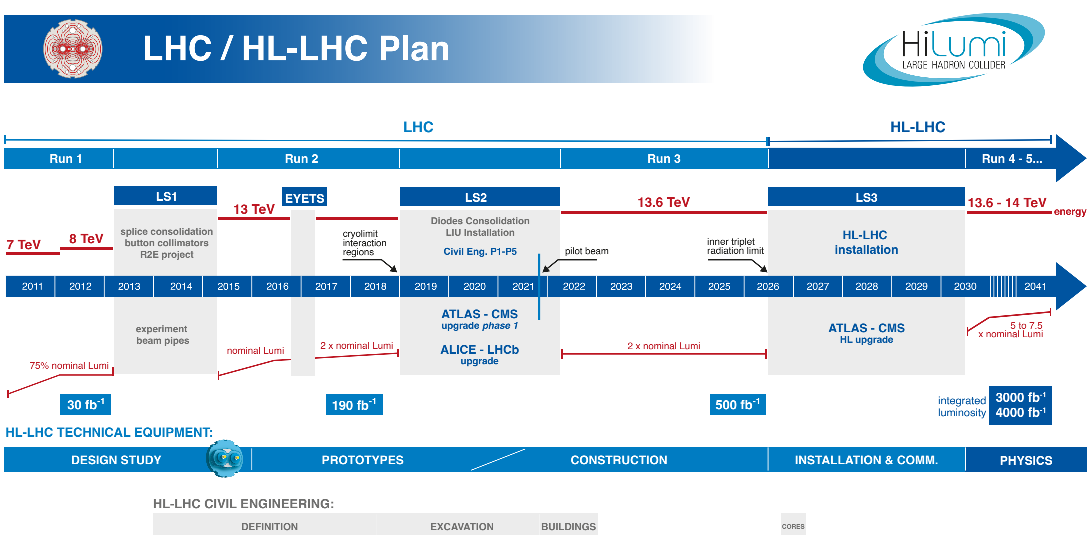


Figure 5.8. The LHC / HL-LHC operation and upgrade plan, reproduced from Ref. [38].

Chapter 6

The CMS detector

6.1 Overview

The experimental apparatus used in this dissertation is the Compact Muon Solenoid (CMS) detector (Figure 6.1), one of the two general-purpose detectors at the LHC. It is uniquely characterized by its strong, superconducting 3.8 T solenoid magnet, around which are arranged several subdetectors to measure the properties of particles produced in collisions at the LHC. Inside the solenoid, it contains an all-silicon tracker, to measure the momenta of charged particles and identify the collision vertex, and lead-tungstate crystal electromagnetic and brass and scintillator hadronic calorimeters to measure the energy of particles interacting through the electromagnetic and strong forces, respectively. Finally, outside the solenoid are gas-ionization detectors, interleaved with steel flux-return yoke plates, to track muons.

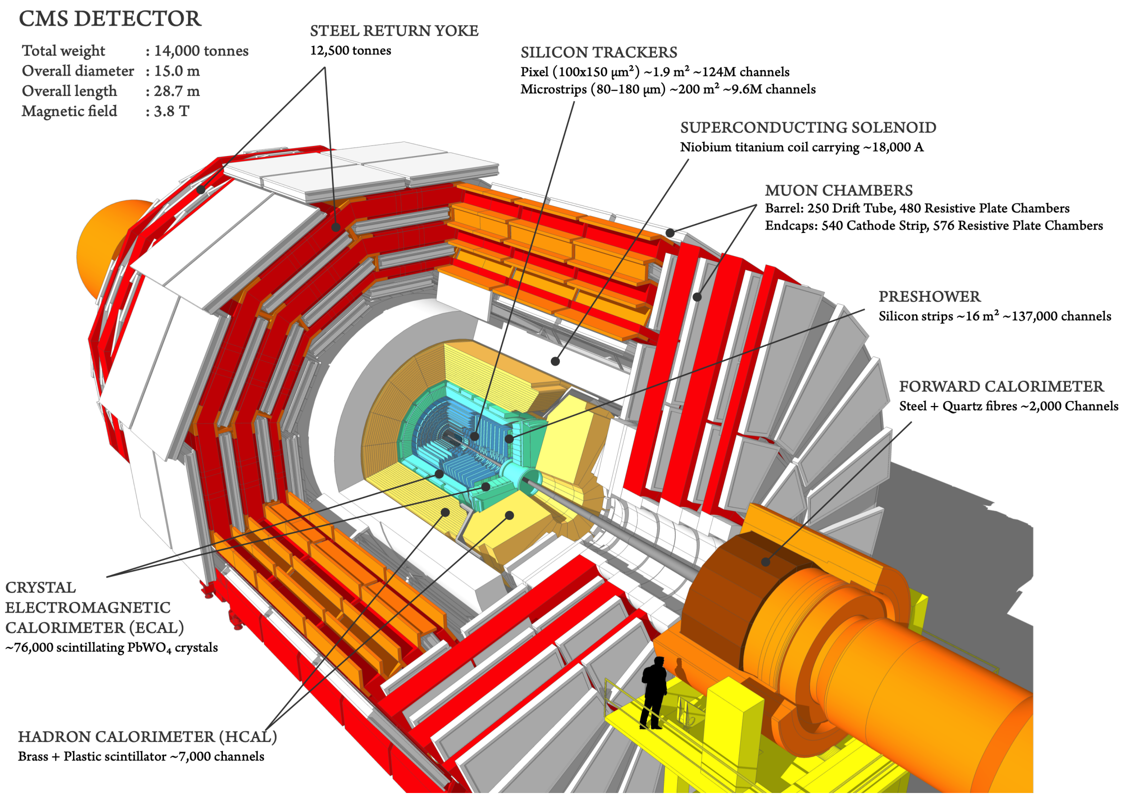


Figure 6.1. A cutaway view of the CMS detector showing the various subdetectors and the solenoid magnet, reproduced from Ref. [**?**].

The CMS detector design was strongly motivated by the potential for discovery of the Higgs boson as well as new physics at the TeV energy scale. Specifically, the original design requirements were [45]:

- Strong muon identification and momentum resolution, as well as good charge determination below $p < 1$ TeV;
- High charged-particle momentum resolution and reconstruction efficiency;
- Efficient triggering and good offline reconstruction of τ-leptons and b-jets;

- Strong and hermetic electromagnetic energy resolution for photons and electrons;
- Good missing transverse energy (MET) and jet mass resolutions.

In addition to the above, the detector also had to be robust against the high radiation environment and pileup at the LHC, as well as have a powerful online event selection system, called the *trigger*, to reduce the high raw 40 MHz data rate to something manageable for offline storage and analysis.

To satisfy the latter, events of interest in CMS are selected using a two-tiered trigger system. The first level (L1) uses custom hardware processors and information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of 4 μs [166]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing and reduces the event rate to around 1 kHz before data storage [167]. Both online and offline, the raw detector signals are processed and reconstructed first locally as hits in the individual subdetectors, then as tracks and calorimeter clusters, and finally as physics objects such as electrons, muons, jets, and missing energy using the particle-flow (PF) algorithm [47].

As we describe below, the CMS detector was able and continues to meet these ambitious requirements. The CMS collaboration not only discovered the Higgs boson in 2012 [143], but also has since performed a wide range of measurements of the Higgs sector and the SM as well as searches for diverse new physics, such as those described in this dissertation.

Looking ahead, however, the upcoming high-luminosity era of the LHC (Chapter 5.2) will bring forth considerable new challenges to the detector, with significantly higher radiation levels, occupancies, and pileup. To overcome them, nearly all CMS subdetectors will undergo a major upgrade after Run 3, known as the Phase-2 upgrade [42]. The L1 trigger latency will be increased from 4 to 12.5μs and the rate from 100 to 750 kHz, with an HLT rate of up to 10 kHz, to cope with the increased data rates (as well as incorporate tracking information at L1 for the first time) [168].

Along with this, the Phase-2 upgrade includes the addition of new timing layers and the high granularity endcap calorimeter (HGCAL), which is notable not only for its ambitious design, but also the computational challenges it poses in detector simulation and reconstruction. These challenges are a major motivation for the work described in Part IV, exploring machine learning innovations to accelerate these simulations in CMS.

In this chapter, we first introduce general concepts behind particle detectors in Section 6.2, before describing the individual CMS detector components in Section 6.3. The detector reconstruction and performance, as well as the PF algorithm is then discussed in Section 6.4. We conclude with the Phase-2 upgrade of CMS in Section 6.5, including the HGCAL in Section 6.5.4.

Coordinate system

The CMS detector uses a coordinate system illustrated in Figure 6.2, with the origin set at the interaction point within the detector. The x-axis is oriented toward the center of the LHC ring, the y-axis is perpendicular to the plane of the LHC ring, and the z-axis is parallel to the beamline. The azimuthal angle ϕ is measured in the x-y plane, relative to the x-axis and the polar angle θ in the x-z plane, relative to the z-axis. Typically, θ is converted to the pseudorapidity $\eta = -\ln\left[\tan\left(\theta/2\right)\right]$, which has a more useful scale for describing high energy collisions, and the angular separation between two particles is quantified using the variable $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. Finally, the transverse component of vectors, such as the transverse momentum p_{T}, are defined as projections onto the x-y plane.

6.2 Detecting particles

6.2.1 Particle interactions with matter

In Part I, we discussed the interpretation of fundamental particles as irreps of the Poincaré group and quantum excitations of fields. In experimental physics, we have yet another interpretation: "a particle is an object that interacts with your detector such that you can follow its track" (W. Riegler [169]).

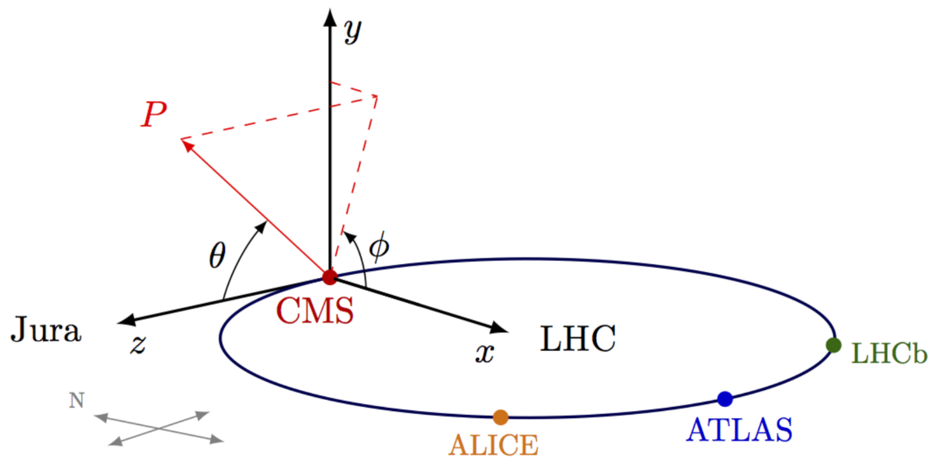


Figure 6.2. The conventional CMS coordinate system.

Which is to say, in order to detect particles, they must interact with the detector material and transfer energy in a way that we can measure. Out of the myriad particles produced in LHC proton-proton collisions, the are only eight particles stable enough to reach the CMS detector and be detected are listed in Table 6.1. Neutrinos are also stable, but are too weakly interacting to measure with the CMS detector, which means it is vital to measure the energy of all the other particles hermetically; the presence of neutrinos can then be inferred by energy conservation, or "missing energy" carried by neutrinos.

Charged particles

Out of these eight, the charged particles can interact electromagnetically with matter through:

Table 6.1. Particles which can reach and be detected by the CMS detector. Lifetimes are given in the rest frame for unstable particles.

Particle	**Mass (MeV/c^2)**	**Charge (e)**	**Lifetime (s)**
Photons	0	0	Stable
Electrons / positrons	0.511	± 1	Stable
Protons	938	$+1$	Stable
Neutrons	940	0	880
(Anti-)Muons	106	± 1	2.2×10^{-6}
Charged pions	140	± 1	2.6×10^{-8}
Neutral kaons	498	0	9×10^{-11}—5×10^{-8}
Charged kaons	494	± 1	1.2×10^{-8}

- Ionization and excitation of atoms: inelastic scattering with atomic electrons and elastic scattering from nuclei, respectively. The average energy loss per distance $\langle dE/dx \rangle$ of a particle due to ionization is given by the famous Bethe-Bloch formula [170].

- Bremsstrahlung: photon radiation because of (de-)acceleration in the electric field of nuclei;

- Cherenkov effect: photon radiation due to the particle moving faster than the speed of light in the medium;

- and transition radiation: photon radiation due to the crossing a boundary between two different dielectrics.

Generally, for "heavy" charged particles (of mass $\gg$ electron mass), electromagnetic interactions are dominated by ionization and excitation, while for electrons and positrons, Bremsstrahlung is dominant at higher energies.

The presence and angle of Cherenkov radiation depends on the particle momenta, a fact which is often exploited for "particle identification" (PID) by distinguishing particles of different masses at given momenta. For example, LHCb uses ring-imaging Cherenkov (RICH) detectors to distinguish hadrons, which is critical for b-physics [171]. CMS, on the other hand, did not prioritize PID and cannot accurately distinguish between the different hadrons beyond their charge.

Photons

Photons primarily interact through:

- The photoelectric effect: absorption by an atom causing the ejection of an electron, dominant at low energies, $\ll 1\,\text{MeV}$;
- Compton scattering: incoherent scattering off an atomic electron, dominant at intermediate energies, $\sim 1\,\text{MeV}$;
- and pair production: converting into electron-positron pairs in the Coulomb field of nuclei, dominant at high energies, $\gg 1\,\text{MeV}$.

The combination of these effects means that as high energy electrons and photons propagate through the detector material, they produce a cascade of secondary particles, called an *electromagnetic shower*, which are analyzed to infer the presence and overall energy of the originating particle.

It is often convenient to characterize detector materials by their *radiation length* (X_0), which is the mean distance into the material over which high-energy electrons lose $1/e$ of their energy due to Bremsstrahlung, and $7/9$ of the mean free path of photons before pair production.

Hadrons

Finally, high energy hadrons can also interact with atomic nuclei through the strong force, losing energy through further particle emissions which create their own *hadronic shower*. As a large fraction of these emitted particles are neutral pions, which decay immediately into photons, hadronic showers are also often accompanied by electromagnetic sub-showers. We can characterize materials for hadron detection similarly by their *nuclear interaction length* (λ), the mean free path between nuclear interactions.

6.2.2 Types of detectors

Tracking detectors

The earliest particle detectors were gaseous *ionization chambers*. Charged particles passing through these detectors ionize the gas along their trajectory, creating visible tracks, which can be captured, for example, by using an electric field to push the ions towards photographic film. The most notable examples are cloud chambers, which were prominent in the first half of the 20th century, and led to the discovery of the positron, muon, and kaon via cosmic rays. The Nobel Prize was awarded to Charles Wilson in 1927 and Carl Anderson in 1936 for the invention and development of the cloud chamber, respectively.

Tracking detectors have since continuously evolved, such as through the use of liquid media in *bubble chambers* and charged wires to produce electric fields and read out ionization signals electronically in *wire chambers*, both of which again led to Nobel Prizes for their inventors, Donald Glaser and Georges Charpak, respectively. The Gargamelle bubble chamber at CERN notably led to the discovery of weak neutral currents [138] (see Chapter 4.2.1).

More recently, a significant advancement in tracking detectors has been achieved through the use of semiconductors such as silicon. A *p*-*n* semiconductor diode [172, 173] effectively forms an ionization chamber as well, where charged particles passing through will create electron-hole pairs whose charge can be collected and recorded. Semiconductor detectors can have lower ionization energies, higher granularity, better position and time resolution, and strong radiation tolerance, while also being able to leverage innovations and state-of-the-art fabrication techniques from the semiconductor industry.

Thus, there has been a gradual shift towards their use, particularly in collider physics. Here, tracking detectors are crucial for (1) *vertexing* — measuring particle tracks precisely to determine the point of collision, or the "vertex" —

and (2) measuring the curvature of charged-particle trajectories in a magnetic field to determine their momenta. Silicon trackers, for example, were employed for vertexing in all LEP experiments and the CDF and DØexperiments at the Tevatron. The CMS detector is notably the first to use silicon for its entire tracking volume.

Semiconductor trackers are, however, more expensive per unit area. Hence, gaseous and liquid detectors remain prevalent in particle physics, particularly where large volumes are required, such as in neutrino experiments and the CMS muon system.

Calorimeters

Calorimeters are detectors designed primarily to measure the energy of particles. They can be either *homogeneous*, where the entire volume of the detector can both absorb *and* measure the energy of the shower; or, *sampling*, where separate "passive" layers which absorb energy and initiate the shower are interleaved with "active" layers to measure the energy. Sampling calorimeters are less precise than homogeneous calorimeters, but are more cost-effective, especially when large volumes are required. CMS employs both types of calorimeters, and primarily uses *scintillation* — photon emission due to atomic (de-)excitation from charged particles — to capture and measure energy.

Generally in collider physics trackers are designed to have short radiation lengths, to minimize particle energy loss, and calorimeters as long a radiation length as possible, to capture the entire energy of electromagnetic or hadronic showers. In addition to their radiation and nuclear interaction lengths, calorimeter materials are characterized as well by their *Molière radius* (R_M), which is the radius of a cylinder containing, on average, 90% of an incident electron or photon's electromagnetic shower energy. It is approximately related to X_0 as:

$$R_M = 0.0265 X_0 (Z + 1.2), \tag{6.2.1}$$

where Z is the atomic number of the material.

General-purpose detectors

Designing a general-purpose detector, such as CMS, requires a careful optimization of several factors, including high efficiency and resolution for as many of the particles in Table 6.1, radiation hardness, cost effectiveness, and more. A typical compromise in particle physics has been a "layered" detector design, as shown in Figure 6.3, with a thin (in radiation and nuclear interaction lengths) innermost tracker for precise vertexing and momentum measurements, followed by thick calorimeters to measure particle energies, and finally dedicated detectors to identify and measure high energy muons that are able to penetrate the previous layers. The CMS detector follows this general philosophy, as shown in Figure 6.4.

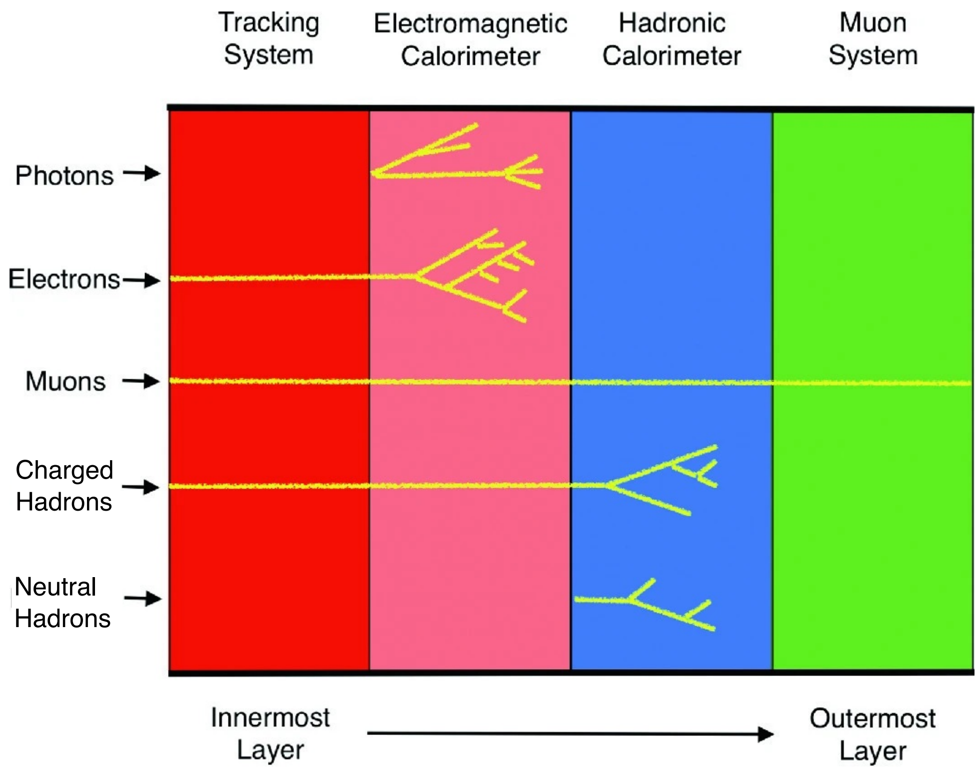


Figure 6.3. Layers in a typical general-purpose detector in particle physics and an illustration of the interactions of different particles.

6.3 CMS detector components

6.3.1 The magnet

The defining characteristic of the CMS detector is its strong 3.8 T solenoid magnet, with a 11.4 T m bending power (Figure 6.5). Its large 6 m diameter and 13 m length accommodate not only the tracker for momentum measurements, but also the

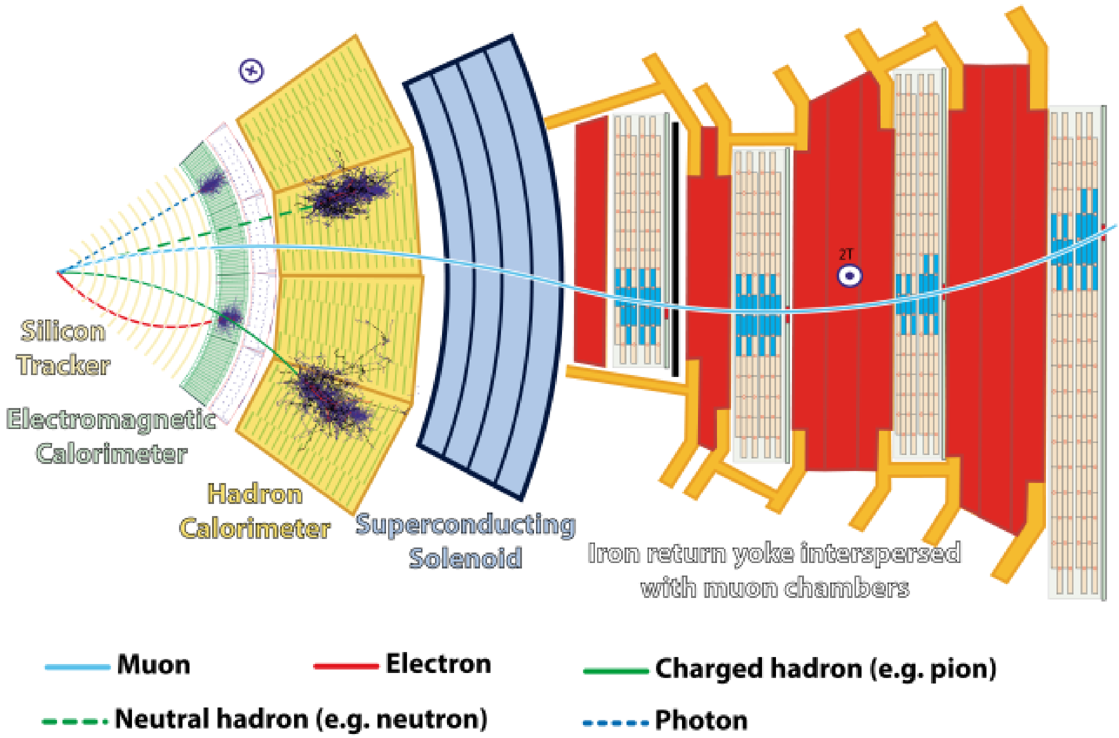


Figure 6.4. Illustration of the different detector layers in the CMS barrel region and the expected hits and energy deposits from various particles, reproduced from Ref. [39].

calorimeters in order to reduce the material in front of them. The strength of the magnet was chosen to allow precision measurements of charged-particle momenta and to achieve a target p_T resolution of 10% for 1 TeV muons.

As shown in Figures 6.1 and 6.4 in red, the solenoid is additionally surrounded by massive layers of steel "return yoke", weighing 12,000 tonnes in total, for several complementary reasons: (1) it confines the magnetic field, improving the efficiency and safety of the detector, by providing a low-reluctance path for the magnetic field lines to return to the solenoid; (2) it is interleaved

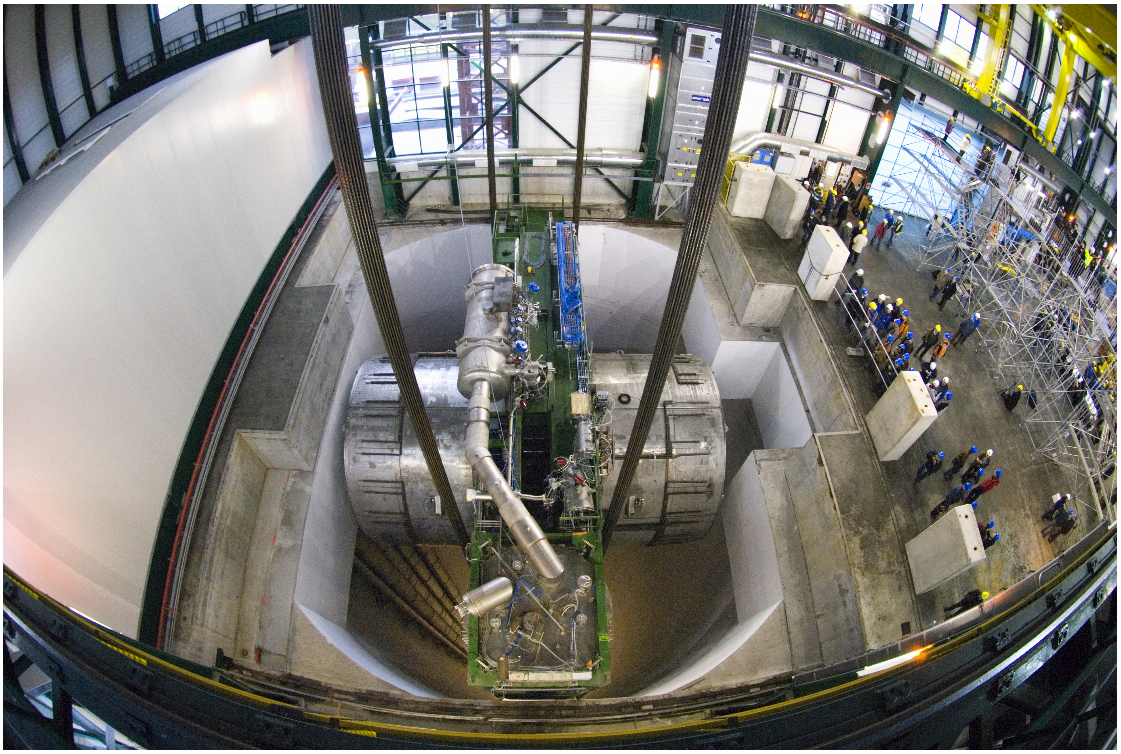


Figure 6.5. The CMS solenoid, as it was being lowered into the CMS cavern in 2007, reproduced from Ref. [40].

with the CMS muon system and provides a residual magnetic field for muon tracking; (3) it absorbs the remaining particles not been completely contained by the calorimeters; and, finally, (4) it provides structural support to the detector. The resulting magnetic field throughout CMS is shown in Figure 6.6, where we can see a uniform 3.8 T field within the solenoid and an ≈ 2 T field in the return yoke.

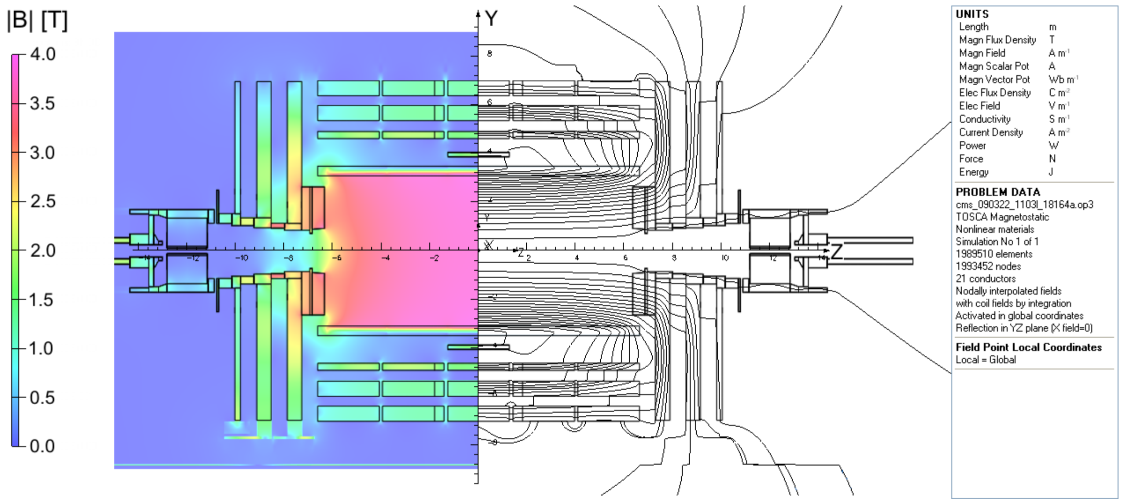


Figure 6.6. Measurement using cosmic rays (left) and illustration (right) of the CMS magnetic field, reproduced from Ref. [41].

6.3.2 Tracker

The CMS tracker is designed to achieve the key detector requirements of strong charged-particle momentum resolution and efficient online and offline τ and b-jet reconstruction. It must additionally operate at a high efficiency at the expected average pileup rate of 20-60 collisions per bunch crossing in Runs 1–3 of the LHC. To do so, as illustrated in Figure 6.7, it comprises relatively small and granular silicon *pixel* layers close to the interaction point for precise vertexing, followed by larger silicon *strip* layers. The tracker has an overall diameter of 2.5m and length of 5.8m, and is composed of a separate co-axial "barrel" region and two "endcap" regions perpendicular to the beamline, to provide pseudorapidity coverage of $|\eta| \lesssim 2.4$.

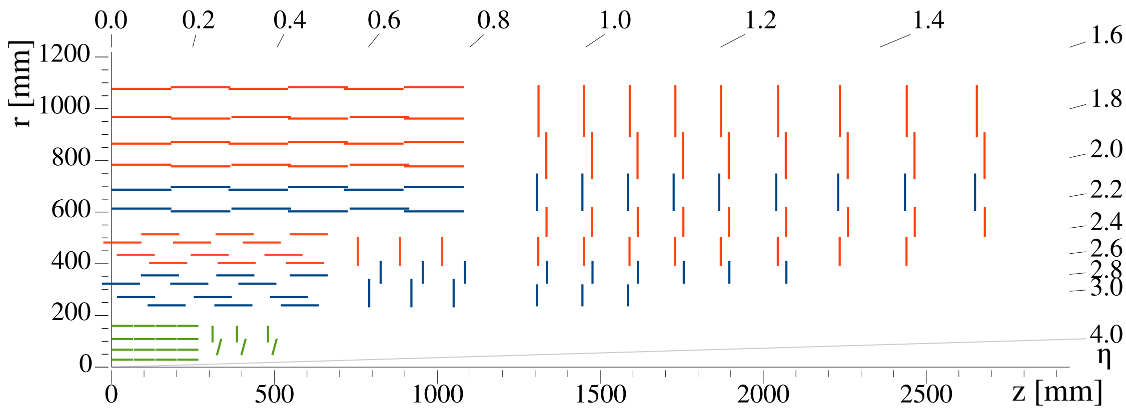


Figure 6.7. Schematic of one quarter of the Phase-1 CMS tracker in the r-z plane, reproduced from Ref. [42]. In green are the pixel detector layers, in red the single-sided strip modules, and in blue the double-sided strip modules.

Silicon has several advantages for tracking detectors: it can be made radiation hard to be placed close to the interaction point [174]; it has a low ionization energy of 3.6 eV; it can be made thin — the maximum radiation and nuclear interaction lengths over the entire CMS tracker are around 2 and 0.6 respectively (Figure 6.8); it is naturally abundant and widely used in the semiconductor industry; and it can be easily patterned to small dimensions for high granularity. This is why, as discussed in Section 6.3, silicon has gained popularity in particle physics, with CMS being the first to use it for the entire tracker.

The pixel tracker includes three barrel layers at radii of 4.4, 7.3, and 10.2 cm and two pairs of endcap disks at $z = \pm 34.5$ and ± 46.5 cm. It contains a total of 1440 modules and 66 million pixels, of size (or "pitch") $100 \times 150\,\mu\mathrm{m}^2$, and thickness 285 μm. The planar position of the sensor provides a third position coordinate as well. The pixel tracker yields an overall hit position resolution of 10–20 μm in the transverse direction — $r\phi$ in the barrel — and 20–40 μm in the

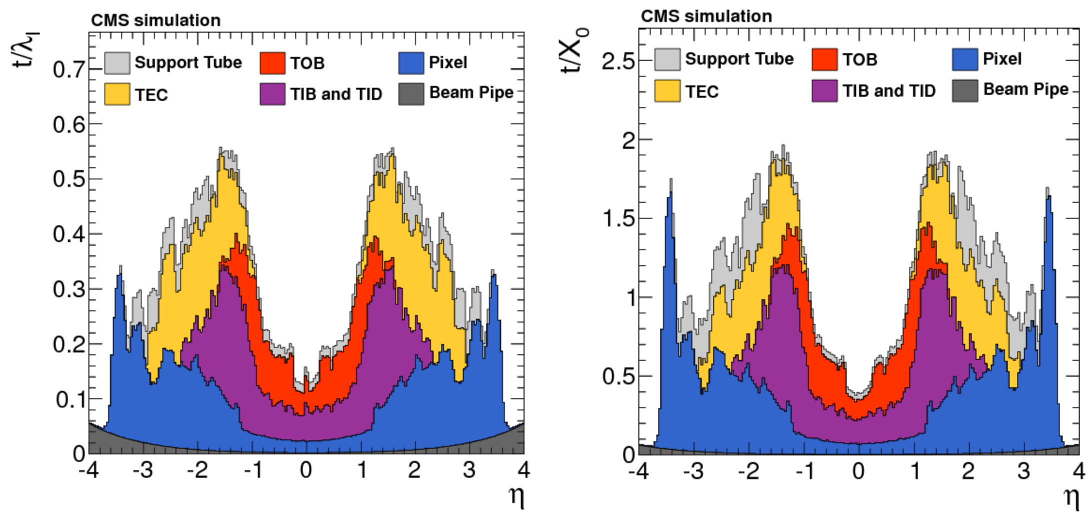


Figure 6.8. Total thickness t of the tracker material traversed by a particle produced at the nominal interaction point, as a function of pseudorapidity η, expressed in units of radiation length X_0 (left) and nuclear interaction length λ_I (right), reproduced from Ref. [175].

longitudinal direction — z in the barrel. The pixel orientation is optimized most for $r\phi$ resolution, as that is the plane in which charged particles bend from the CMS magnetic field.

Silicon strip modules are longer than pixels, providing high granularity only in one axis, but are more cost-effective; hence, they are used in the outer layers of the CMS tracker, which require a much larger area of coverage: 198 m^2 of active coverage versus 1.1 m^2 for the pixels. The strip tracker has 10 barrel layers and three small and nine large endcap disks, with a total of 15,148 modules and 9.3 million strips. It contains two types of strip modules: standard "single-sided" modules as well as "double-sided" modules mounted back-to-back at a stereo angle to effectively allow pixel-like 2D measurements as well, albeit at a lower granularity.

The strip modules in the barrel are aligned parallel to the beamline with a pitch ranging from 80–183 μm, while those in the endcaps are mounted in the radial direction with a pitch of 81–205 μm. Overall, the strips in the inner barrel and disk layers provide an $r\phi$ resolution of 13–38 μm, while the outer layers provide 18–47 μm.

6.3.3 ECAL

The CMS electromagnetic calorimeter (ECAL) (Figure 6.9) was designed to precisely measure energies of electrons and photons. It was particularly optimized for sensitivity to the Higgs-to-two-photon decay channel, which proved crucial to the discovery of the Higgs boson [143].

The ECAL is a homogeneous calorimeter made out of 75,848 lead-tungstate ($PbWO_4$) crystals (Figure 6.10), which are a type of highly transparent scintillators. $PbWO_4$ was chosen for its high density (8.28 g/cm^3), short radiation length (X_0 = 0.89 cm), and small Molière radius (R_M = 2.2 cm), which allows for a compact calorimeter with fine granularity. Additionally, its fast scintillation response ($\approx$ 10 ns) allows distinguishing between "out-of-time" (OOT) pileup — particles produced from adjacent bunch crossings — and particles from the primary interaction [176]. The scintillation light is detected by photodiodes glued to the back of each crystal.

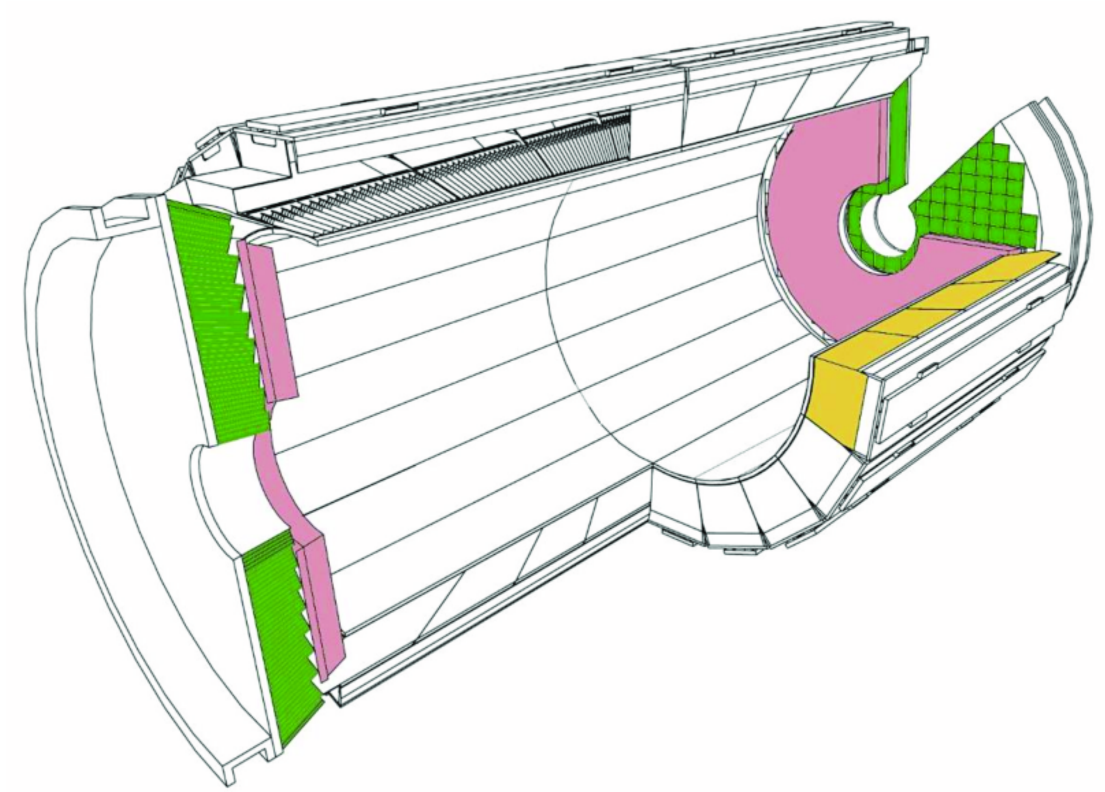


Figure 6.9. Layout of the CMS ECAL, reproduced from Ref. [43], with one of the 36 barrel regions highlighted in yellow, the preshower in pink, and the endcap regions in green.

Like the tracker, it has a barrel region, covering $|\eta| < 1.48$, and two endcap regions for $1.48 < |\eta| < 3.00$. They have thicknesses of $25.8X_0$ and $24.7X_0$, respectively, and a total nuclear interaction length of around 1. This is sufficient to contain >98% of the energy of $\leq$ 1 TeV electrons and photons, and causes around two thirds of charged hadrons to shower in the ECAL as well.

Additionally, the ECAL includes a “Preshower” sampling calorimeter in the endcap regions, which comprises two layers of $3X_0$ of lead to initiate

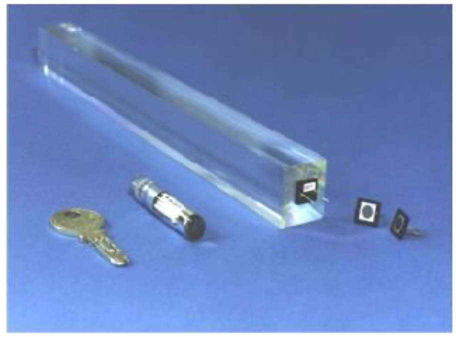


Figure 6.10. $PbWO_4$ crystals and photodiodes used in the CMS ECAL.

electromagnetic showers, interleaved with active silicon strip detectors for measurements. The purpose of the Preshower is primarily to distinguish between single photons and neutral pion decays into two, close-by, photons in the forward regions, whose energy deposits in the ECAL would otherwise overlap significantly.

6.3.4 HCAL

The CMS hadronic calorimeter (HCAL) sits roughly 30 cm outside the ECAL and is designed to measure the energies of neutral and charged hadrons. It is composed of four major sections: the HCAL barrel (HB), the HCAL endcap (HE), the HCAL outer (HO), and the HCAL forward (HF) (Figure 6.11). Due to their much greater volume compared to the ECAL, the HB, HE, and HO are all chosen to be sampling calorimeters, with alternating layers of absorber material and plastic scintillator. The HF extends the pseudorapidity coverage of CMS up to $|\eta| = 5.2$ and is a steel and quartz-fiber *Cherenkov calorimeter*.

The HB has 14 total layers of brass absorbers and scintillators, with additional steel front and back plates, covering $|\eta| < 1.4$ [45]. Due to its radial constraints, with the ECAL and magnet on either side, it has a thickness at $\eta = 0$ of only 5.8λ, with the ECAL adding another $\approx 1\lambda$.

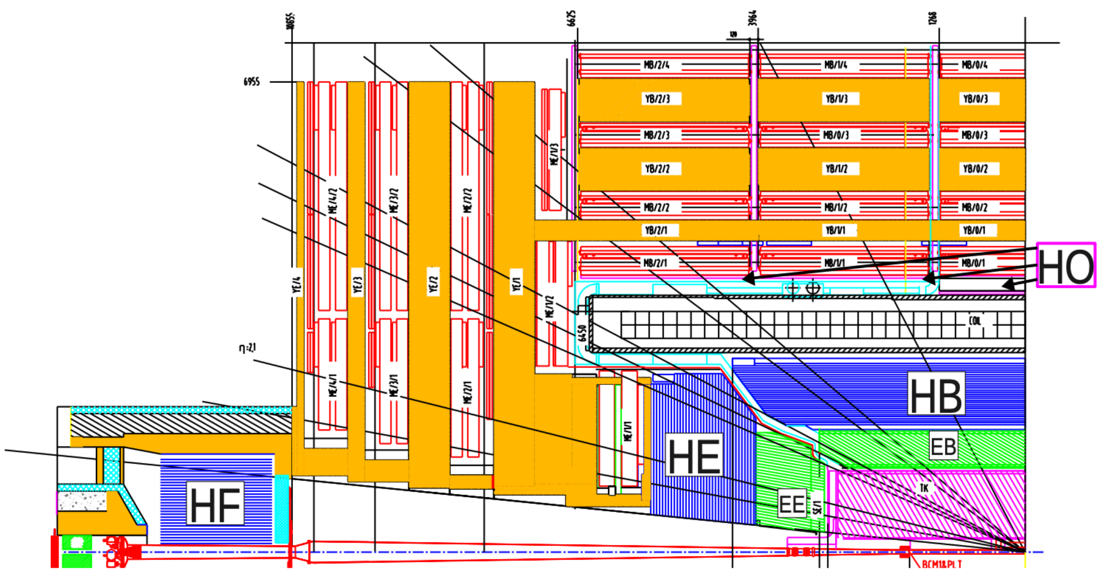


Figure 6.11. Layout of the CMS detector in the r-z plane with the four HCAL sections labeled, reproduced from Ref. [44].

To capture the remaining "tails" of hadronic showers in the barrel region, the HO is placed outside the solenoid. It uses the same scintillators and electronics as the HB, but uses the magnet and return yoke materials themselves as absorbers, adding up to 3λ more of material. Figure 6.12 demonstrates that the HO is crucial for capturing the entire energy of hadronic showers: without it, we see an excess of events with the measured energy of hadrons lower than their incident energy, implying a loss of energy with the HB alone.

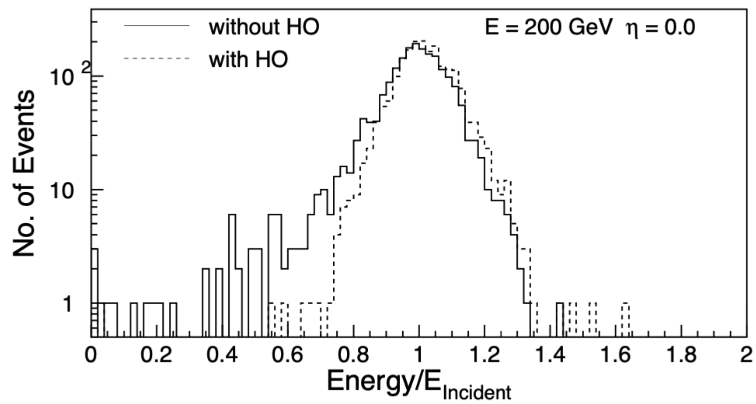


Figure 6.12. Simulation of the distribution of measured / incident energy for pions with incident energies of 200 GeV at $\eta = 0$, reproduced from Ref. [45]. Without the HO, there is an excess of events with the measured energy lower than the incident, while with the HO the distribution is a Gaussian centered at 1, implying recovery of the total pion shower energy.

The HE covers the pseudorapidity range $1.3 < |\eta| < 3.0$ and has 18 layers of brass and scintillators, for a total length of 10λ including the ECAL. The granularity of both the HB and HE for $|\eta| < 1.6$ is 0.087×0.087 in η-ϕ, while for $|\eta| > 1.6$ in the HE it increases to 0.17×0.17.

The HF is placed 11.2 m from the interaction point to cover the very forward range $3.0 < |\eta| < 5.2$. Its primary design constraints are the extremely hostile radiation levels in the high-rapidity region, and hence uses steel absorbers and quartz fibers that are radiation hard [177]. It contains 1.65 m of absorber material in each endcap, and measures energy through Cherenkov light emitted by charged particles moving through the longitudinal quartz fibers. It is hence more sensitive to the electromagnetic component of showers.

The HF is important in measuring the energy of an event hermetically and, thereby, the missing energy as well. Moreover, it is crucial in identifying highly forward jets like those produced through VBF-production of Higgs bosons, an important production-mode measured in this dissertation (see Chapters 4.3 and **??**).

6.3.5 Muon system

As implied by its name, detecting muons with high efficiency and precision was a key consideration in the design of CMS. This is because of their unique signature compared to the other particles in Table 6.1, the possibility of Higgs discovery in the $H \rightarrow ZZ^{*} \rightarrow 4\mu$ channel, and the relatively low isolated-muon background at the LHC. Indeed, the 4μ channel was crucial to the discovery of the Higgs boson [143].

Muons are detected through a combination of the silicon tracker inside the solenoid, and dedicated gas ionization chambers for tracking muons outside the solenoid, covering a total pseudorapidity range of $|\eta| < 2.4$. The muon system, being the outermost subdetector, needs to cover the largest amount of area — around 25,000m^2 of detector layers — and hence uses gaseous detectors to minimize cost while maintaining reliability and robustness. Three types of gas detectors are used: drift tubes (DTs) in the barrel region, cathode strip chambers (CSCs) in the endcap regions, and resistive plate chambers (RPCs) in both, all housed between the steel flux return yoke layers, as shown in Figure 6.13.

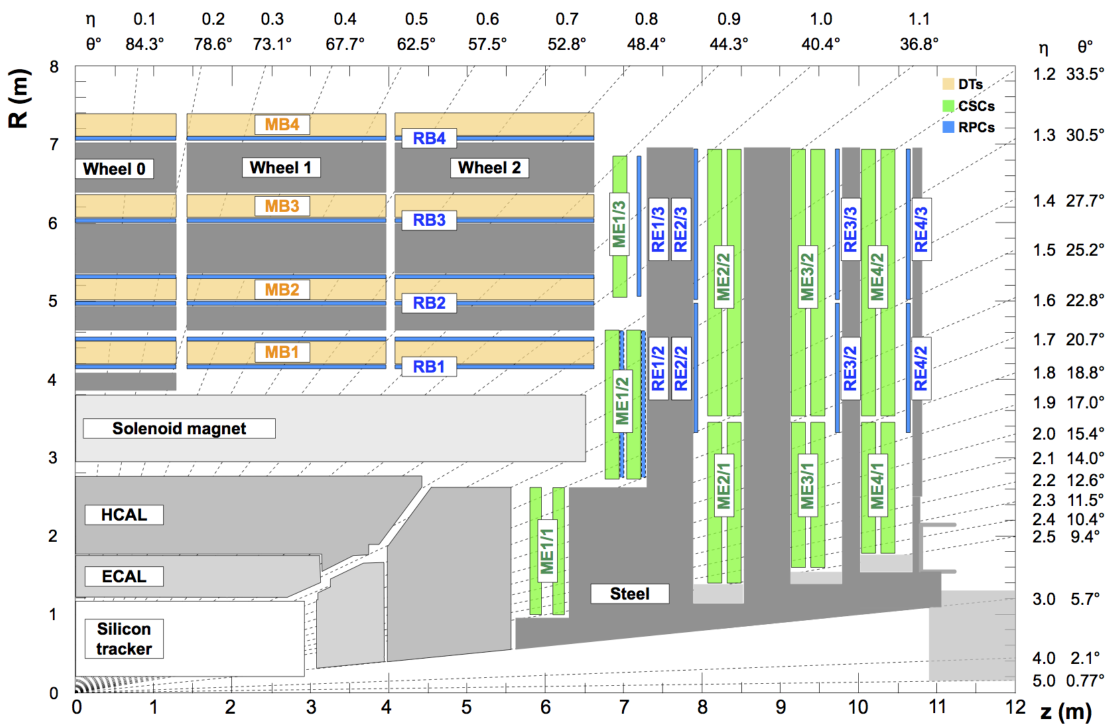


Figure 6.13. Layout of the CMS detector in the r-z plane, with the muon system highlighted and the steel return yoke in dark grey, reproduced from Ref. [46]. The drift tube stations (DTs) are labeled MB ("Muon Barrel") and the cathode strip chambers (CSCs) are labeled ME ("Muon Endcap"). Resistive plate chambers (RPCs) are mounted in both the barrel and endcaps of CMS, where they are labeled RB and RE, respectively.

Because of the lower radiation and background levels expected in the barrel region, standard DT chambers are used, which are known to have excellent spatial and timing resolution while remaining relatively inexpensive. There are four muon barrel (MB) radial layers, or "stations". The innermost three each contain eight chambers measuring the position in the r-ϕ plane and four measuring the z coordinate. The outermost contains only the r-ϕ chambers.

Each chamber comprises several layers of long aluminum *drift cells*, which have a transverse area of $42 \times 13\,\mathrm{mm}^2$ and are filled with a mixture of argon (Ar) and carbon dioxide (CO_2) gas. They contain a single anode wire in the center, and the drift time of the electrons to the wire is used to calculate the muon position. The average single-cell spatial resolution has been measured to be 170μm, with a combined per-chamber resolution of 100μm in r-ϕ [45]. DT chambers also provide a time resolution on the order of nanoseconds, allowing local, independent triggering on muon p_{T}.

The endcaps are subject to much higher radiation and hence use the more radiation-hard CSCs, which offer fast response times and fine segmentation as well but are more expensive. They can also tolerate the non-uniformity of the magnetic field in the endcap regions (see Figure 6.6). There are four muon endcap (ME) stations on each side, which are divided into "rings" in the r-direction, and labeled as ME1/2 for the second ring in the first station and so on.

The endcap muon system contains a total of 468 CSCs: 216 in ME1, 108 in ME2 and ME3 each, and 36 in ME4. A single CSC is composed of six layers of multi-wire proportional chambers, each containing several anode wires spaced between 2.5–3.2 mm apart and 80 cathode strips to read out position in the r-ϕ

plane [178]. The CSCs overall provide a spatial resolution in r-ϕ of 75μm in ME1/1 and ME1/2 and 150μm elsewhere, with a time resolution of <5 ns [45]. This means CSCs, like the DTs, independently allow local triggering on muon p_T with good efficiency and background rejection.

Finally, RPCs are included as well in both the barrel and endcap regions to provide complementary triggering capabilities. RPCs are double-gap chambers operated in *avalanche mode*, which means they primarily offer fast timing information, with around 1.5 ns resolution, but relatively poor spatial resolution [179]. There are four RPC barrel (RB) and three RPC endcap (RE) stations, complementing the DTs and CSCs with faster timing information and allowing the muon p_T trigger threshold to be lowered.

6.4 Detector reconstruction and performance

6.4.1 Tracker

Tracks are reconstructed from the hits in the tracker using an iterative algorithm called the Combinatorial Track Finder (CTF) [175], based on the Kalman filter [180]. CTF starts by finding the "easiest" tracks (e.g. of high p_T and produced near the interaction region), removes their hits from the search, and then repeats the process until all tracks have been found. Due to the high computational cost of track reconstruction, this cannot be performed at the L1 trigger and, hence, track information is not used for the L1 trigger decision currently in CMS.

Offline, CMS is able to reconstruct isolated muons of $p_{\mathrm{T}} > 0.9\,\mathrm{GeV}$ with 100% efficiency within the tracker acceptance of $|\eta| < 2.4$, with a p_{T} resolution of up to 2.8% for $p_{\mathrm{T}} \sim 100\,\mathrm{GeV}$ [175]. The pixel tracker can also achieve a vertex position resolution of 10–12 μm in all three spatial dimensions. B-hadrons produced in the pp collisions generally travel on the order of a few millimeters before decaying and, hence, the precise vertexing of the CMS tracker allows for efficient b-tagging [181] and even boosted bb-tagging [158], as is crucial for the analysis described in this dissertation.

6.4.2 ECAL

Signals in the ECAL crystals are reconstructed by fitting the signal pulse with template pulse shapes to distinguish OOT pileup, both offline and online [176]. The individual hits are then clustered to identify electromagnetic showers initiated by the same incident particle, and are further clustered into “superclusters” to account for photon conversions and bremsstrahlung losses [182]. Clusters are tested for compatibility with reconstructed tracks from both single electrons and pair-produced electrons by photons, and the combined information is used to identify electrons and photons. Cluster energies are calibrated based on differences in neutral pion to two-photon decays in data and simulation [47].

With the PF algorithm, electrons (isolated photons) are identified by the ECAL clusters, the presence (absence) of a corresponding track in the tracker, and a low relative energy deposit in the HCAL along the particle trajectory. The online triggers use a similar but simplified algorithm with tighter requirements on electron and photon identification. Offline, multivariate regression algorithms are used to correct the raw measured energy for inefficiencies due to energy loss before or in the ECAL.

Overall, the ECAL has been measured in data collected by CMS to have a reconstruction efficiency of >95% for $10 < E_{\mathrm{T}} < 500\,\mathrm{GeV}$, with an uncertainty on the electron and photon energy scale of 0.1% in the barrel and 0.3%, in the endcaps [182]. Electron energy resolution was measured to be between 2–5% in $Z \to e^+e^-$ decays.

6.4.3 HCAL

The energy of hits in the HCAL is estimated, and OOT PU rejected, by fitting pulse templates to the photodetector signals, both online and offline [183]. Corrections are applied as well based on measured reduction of the light output of the scintillators due to radiation damage and decrease in the photodetector efficiencies [184]. As the HCAL is a sampling calorimeter, the measured energy must be scaled to estimate the total energy of the hadronic shower. This scale factor is nonlinear with the energy of the incident particle, and is estimated through a variety of techniques using simulations and data for the different HCAL components and regions [184]. The overall energy scale is measured to a precision of <2% in the HB and HE, and <3% for the HO and HF.

A similar clustering algorithm to the ECAL's is used in all HCAL subdetectors, with the exception of the HF where a hit in a cell is directly considered a "cluster'. As hadrons deposit energy in the ECAL as well, cluster energies in both calorimeters are calibrated together for hadrons, using a sample of neutral kaons [47]. As for electrons and photons in the ECAL, the PF algorithm is used to identify hadrons based on a higher relative energy deposit in the HCAL versus ECAL, and, for the case of charged hadrons, a matching track in the tracker.

Figure 6.14 shows the *response* — the relative mean difference between the measured and true energy of a particle — and resolution for single neutral hadron energies in the barrel as a function of the true energy, before and after calibration. We see that the energy resolution is significantly worse than for charged particles and photons — >10% for all energies — due to the modest resolution of the HCAL compared to the tracker and ECAL. However, neutral hadrons on average comprise only 10% of event and jet energies (the rest coming from 65% charged hadrons and 25% photons), which means the overall contribution is at the percent level.

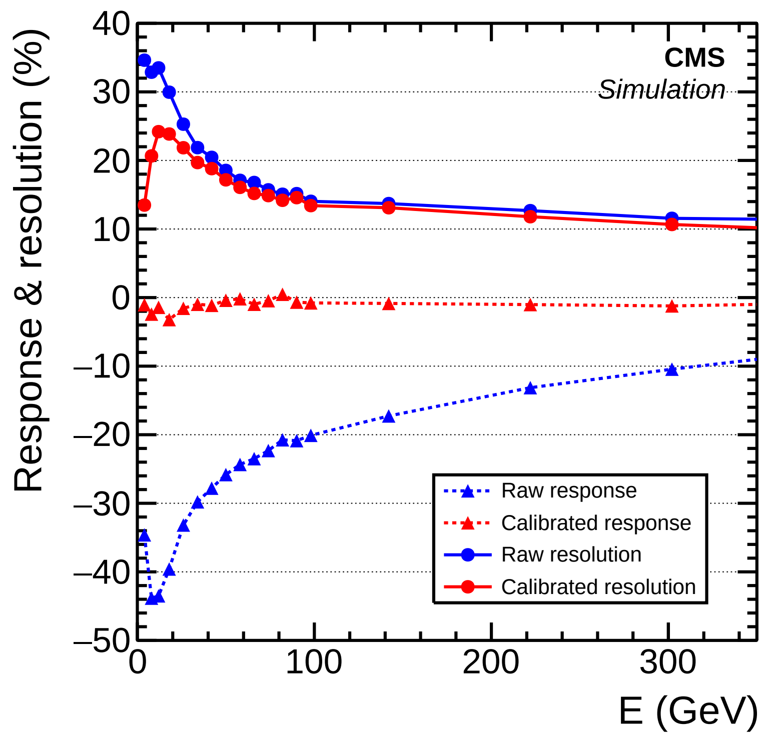


Figure 6.14. Response and resolution of single neutral hadron energies in the barrel as a function of the true energy, before and after calibration, reproduced from Ref. [47].

6.4.4 Muon system

The muon system is triggered using the independent and complementary timing information from the DTs and CSCs in the barrel and endcap, respectively, and the RPCs in both. Hits are first reconstructed locally based on the timing information from the RPCs and the position and timing information from the DTs and CSCs. Hits along the muon chambers are then combined to form *standalone-muon tracks* using a Kalman filter technique [180]. Additional *tracker muon tracks* and *global muon tracks* are formed by propagating tracker tracks to loosely matched DT or CSC hits, and matching the standalone-muon tracks to tracker tracks, respectively [46]. The combined track information is used by the global PF algorithm to optimize muon identification and determine their momenta [47].

Overall, the muon reconstruction and identification efficiency has been measured to be >96% [46]. For lower p_{T} muons ($p_{\mathrm{T}} < 200\,\mathrm{GeV}$), the momentum measurement is dominated by the inner tracker performance, with a resolution of approximately 1% in the barrel and 3% in the endcaps. For higher p_{T} muons, the combined tracker and muon system information is important, with a measured resolution of <6% at $p_{\mathrm{T}} \sim 1\,\mathrm{TeV}$.

6.4.5 Object reconstruction and particle flow

The PF algorithm [47] is used to reconstruct and identify each individual particle in an event, with an optimized combination of information from the different subdetectors. The energy of photons is obtained from the ECAL measurement. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies. The primary vertex (PV) is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in Ref. [185].

For each event, hadronic jets are clustered from these reconstructed particles using the infrared and collinear safe anti-k_{T} algorithm [186, 187] with a distance parameter of 0.4 (AK4 jets) or 0.8 (AK8 jets). Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_{T} spectrum and detector acceptance. For the analysis described in this dissertation, the charged-hadron subtraction [188] and pileup per particle identification [189, 190] algorithms are used to mitigate the effect of pileup on AK4 and AK8 jets, respectively, and further corrections are applied to their energy and mass scales and resolutions to correct for detector mismodeling.

Electrons falling within the tracker acceptance are reconstructed using momentum derived from the tracker, the energy from the corresponding ECAL cluster, and the collective energy of all bremsstrahlung photons spatially aligned with the electron track [191]. Muons falling within the muon chamber acceptance $|\eta| < 2.4$ are reconstructed as tracks in the central tracker which align with tracks or hits in the muon chambers [46]. For the analysis described in this dissertation, electron candidates are required to fall within the tracker acceptance of $|\eta| < 2.5$ and have $p_T > 20\,\text{GeV}$, while muon candidates are required to be within the muon chamber acceptance of $|\eta| < 2.4$ and have $p_T > 10\,\text{GeV}$. Both leptons are then required to pass additional identification criteria [46, 191] to improve purity and be isolated [47] to suppress those originating from bottom or charm hadron decays.

6.5 The Phase-2 Upgrade

6.5.1 Tracker

In the HL-LHC, the tracker will have to endure much higher radiation levels and help mitigate the larger 140–200 expected pileup interactions. The entire CMS tracker will thus be replaced during the long shutdown 3 (LS3) between 2026 and 2029 with the "Phase-2" tracker [42], comprised of an Inner Tracker of silicon pixels and an Outer Tracker of silicon strip and "macro-pixel" modules. Overall, the upgrade will improve its radiation hardness, granularity, as well as increase its forward acceptance.

A key additional novelty is the inclusion of dedicated "p_T modules" [192] in the Outer Tracker to efficiently and quickly detect high p_T ($\gtrsim 2\,\text{GeV}$) tracks. This will allow, for the first time, L1 trigger decisions to be based on tracker information. This is crucial to mitigate the increased pileup, and may perhaps even improve the trigger efficiency for objects such as b-jets and τ-leptons [193].

6.5.2 Timing layers

The Phase-2 upgrade of CMS will include a novel, thin layer of timing detectors between the tracker and calorimeters to provide a target resolution of 30–60 ps for charged particles. This precise timing information will be crucial for reducing OOT pileup particles not compatible with the time of the primary vertex, with an estimated effective pileup reduction from 200 to between 33–70 [194]. It may additionally aid particle identification (PID), and hence jet tagging, using time-of-flight measurements to calculate particle velocities and masses for given momenta [195].

The timing layers are based on minimum ionizing particle (MIP) timing detectors (MTDs). MIPs are high-energy particles which deposit a small fraction of their energy as they traverse and ionize the sensors. MTD sensors are designed for rapid signal collection and response to these interactions to achieve the target timing resolution. Two separate barrel and endcap timing layers (BTL and ETL, respectively) will be installed, using different sensor technologies based on the different geometries and radiation levels.

The BTL will be made out of 300,000 scintillating Cerium-doped Lutetium (LYSO) crystals, known for their fast response time and high light yield, read out by silicon photomultipliers (SiPMs), also known for speed and good photon detection efficiency (PDE). This combination has been measured to provide the desired time resolution of 30 ps in charged pion test beams [196]. Both the crystals and SiPMs are sufficiently radiation-hard for the barrel; however, SiPM PDE is expected to degrade over time due to increased *dark current* noise, reducing the timing resolution to about 50-60 ps by the end of the HL-LHC [196].

The more extreme levels of radiation in the endcap preclude the use of SiPMs. Instead, the ETL will use a more radiation-hard silicon sensor known as a low gain avalanche detector (LGAD). LGADs incorporate an extra gain layer into the typical *p*-*n* junction diode in order to rapidly amplify the signal by a factor of 10–30. They have been measured to allow single-hit resolution for MIPs at the level of 30–50 ps, even after the full expected radiation dose of the HL-LHC [194].

6.5.3 Barrel calorimeters

In the barrel, the $PbWO_4$ crystals and photodiodes of the ECAL are expected to perform well and will be retained, although the operating temperature will be lowered from 18 to 9C to counter increased noise in the photodiodes [43]. The electronics will be upgraded to be faster and more radiation tolerant, with a target time resolution of 30 ps for energy deposits greater than 50 GeV to mitigate pileup. The radiation damage to the HCAL barrel active material is expected to have a negligible impact on the physics performance and, hence, the HB scintillators and fibers will also be retained [197]. However, its back-end electronics will be similarly upgraded to sustain the higher 750 kHz L1-trigger rate.

6.5.4 HGCAL

Both the ECAL and HCAL calorimeter endcaps (CEs) will be replaced entirely by the new High-Granularity Calorimeter (HGCAL) [198] due to the extreme forward radiation levels expected. The HGCAL has been designed to not only withstand the increased radiation, but also to provide: (1) high lateral granularity, for better shower separation and narrow jet identification; (2) fine longitudinal granularity, for better shower shape and energy resolution; as well as (3) precision timing for pileup rejection. The latter means HGCAL will be a 5D calorimeter, able to measure the position, energy, and timing of hits.

The HGCAL will be a large sampling calorimeter with a total of 47 absorber and sensor layers, illustrated in Figure 6.15. The electromagnetic section (CE-E) will comprise 26 sensitive layers of $\approx$0.5–1 cm ^{2} silicon sensors interleaved with copper, copper-tungsten, and lead absorber plates, with a total thickness of $27.7X_0$ and 1.7λ. The hadronic section (CE-H) will contain 21 layers of sensors, with silicon in the high-radiation regions and $\approx$4–30 cm ^{2} plastic scintillating tiles read-out by SiPMs in the low-radiation regions (Figure 6.16), and stainless steel and lead absorbers, for a total thickness of 7λ. The entire HGCAL will be operated at -30C to keep electronic noise sufficiently low.

Silicon is again chosen for the benefits described in Section 6.3.2, as well as its fast response time, which is expected to provide time resolution of $20 - 60\,\mathrm{ps}$ depending on the energy of the hit. This has been shown to significantly aid in pileup rejection [199]. Cheaper plastic scintillators are used where the radiation levels are lower and, additionally, a hexagonal geometry is chosen to cover the more than $600\,\mathrm{m}^2$ of silicon area required in the most cost-effective manner, as shown in Figure 6.16.

Overall, the high-granularity, high-density, and fast timing calorimetry of the HGCAL is expected to significantly improve electron, photon, and jet efficiency and resolution, mitigate pileup, and allow more powerful trigger algorithms in HL-LHC [198]. However, the increased complexity and occupancy, and unorthodox geometry, will pose significant challenges not only in its design and construction, but also computationally with respect to its simulation and reconstruction. This is a strong motivation for the exploration of new computing techniques for fast and efficient CMS simulations, as we will discuss in Part IV.

6.5.5 Muon system

As with the barrel calorimeters, the gas detectors themselves in the muon system are expected to continue performing well at the HL-LHC with no significant degradation in the overall muon reconstruction performance [46]. The electronics, however, will be upgraded to be more radiation hard and sustain the higher trigger rate.

The higher radiation and pileup does pose considerable challenges to reliable muon triggering in the very forward regions, however. In fact, trigger inefficiencies were already anticipated in Run 3 of the LHC, which led to the installation of a new gas electron multiplier (GEM) station in the endcap regions of the muon system in the 2019–2022 long shutdown before Run 3, and recently another GEM station at the beginning of 2024 [200]. GEMs are popular gas detectors with high rate capability and radiation tolerance, and provide crucial additional hit information to improve the trigger efficiency and muon reconstruction in the forward $1.5 < |\eta| < 2.4$ regions.

During LS3, two more RPC stations for $1.8 < |\eta| < 2.4$ and one final GEM will be added in the endcap regions for HL-LHC. The RPCs will provide further forward timing information as well to complement the CSCs and recover single-muon trigger efficiencies [201].

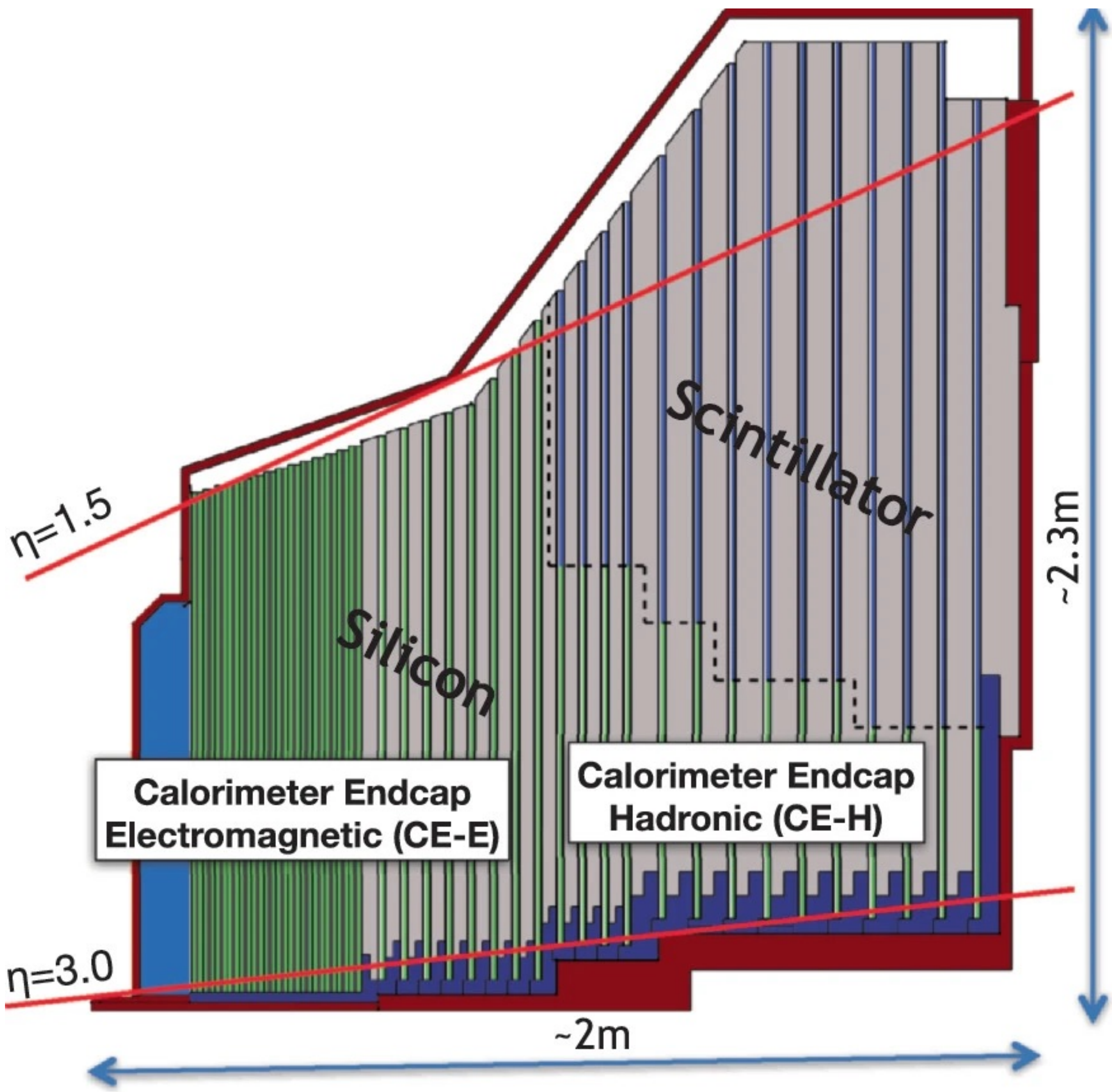


Figure 6.15. Layout of the CMS HGCAL, reproduced from Ref. [48].

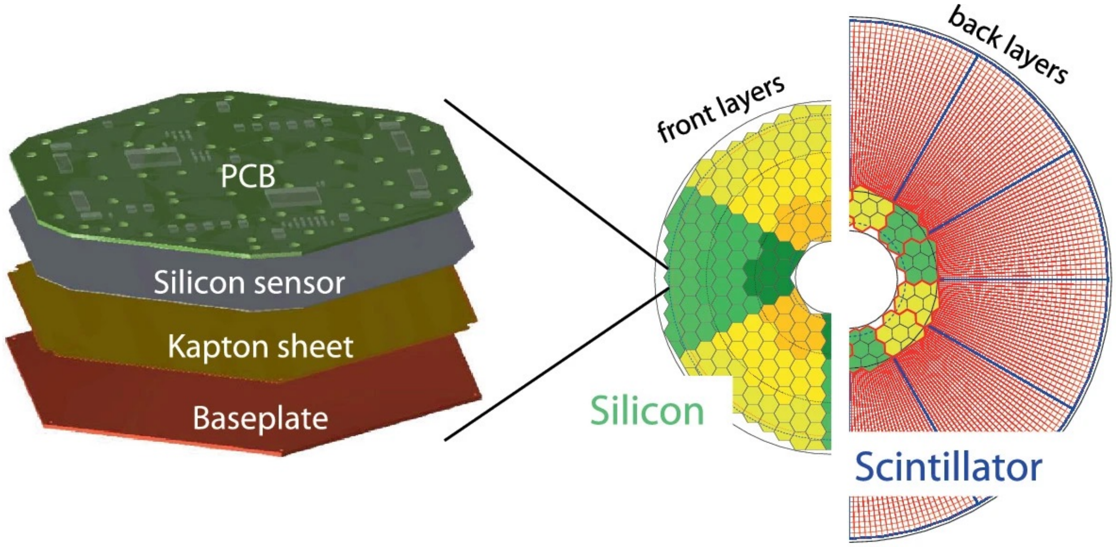


Figure 6.16. (Left) layers of the individual HGCAL modules and (right) their layout in the all-silicon and mixed layers, reproduced from Ref. [48].

Part III

AI/ML and Statistics Background

Chapter 7

Machine Learning for HEP

7.1 Introduction

Machine learning (ML) and deep learning (DL) are revolutionizing data analysis, computing, and even real-time triggers in high-energy physics (HEP). Significant contributions of this dissertation include ML advancements for Higgs boson searches and beyond and fast detector simulations for the HL-LHC. In this chapter, to motivate them, we introduce some core concepts of ML, especially as they relate to HEP applications.

ML refers to a general class of algorithms that "learn" from data to solve problems. This is in contrast to traditional, hand-engineered bespoke algorithms designed by domain experts to address specific tasks. A relevant example in HEP is selecting a high-purity, in terms of signal versus background, region of data: a traditional approach would be to manually define a set of selections on individual kinematic features based on physical reasoning; for example, when measuring Higgs boson production, we select for events exhibiting resonances around the Higgs mass.

However, as we enter the regime of extremely large quantities of high-dimensional data and measurements of ever more complex processes (such as the HH searches described in this dissertation), it soon becomes intractable, or at least suboptimal, to manually define selections over the $\mathcal{O}(10\text{–}100)$ event features that can help distinguish signal from background. This is where we turn to ML algorithms, such as boosted decision trees (BDTs) [202], which can automatically compute optimal, non-linear selections in this high-dimensional feature space.

More recently, the advent of artificial neural networks (ANNs) and deep learning (DL), along with increased data availability and computing power, has led to orders of magnitude increases in the dimensionality of data that can be exploited and the complexity and expressivity of the models built. A relevant example of their significant impact in HEP is in jet identification: jets are extremely high-dimensional objects, composed of hundreds of particles, tracks, and vertices each with several distinguishing features. Traditionally, this information had to be aggregated into hand-engineered, high-level features, such as the jet mass, number of prongs, and vertex displacements.

DL, on the other hand, allows us to leverage the full set of low particle- and vertex-level features. This leads to powerful classifiers that significantly outperform traditional methods and improve the sensitivity of our jet-based measurements. This is exemplified by the new HVV jet identification algorithm we introduce in Chapter **??** and apply to searches for HH production in Chapter **??**. As we argue in Section 7.3, DL also has the potential to alleviate the computational challenges we foresee in the HL-LHC era, particularly with respect to detector simulations, which are the focus of Part IV.

ANNs have proven to be extremely flexible building blocks out of which to construct diverse and sophisticated models for a variety of tasks in HEP, from classification and regression to simulation and anomaly detection, and more. Indeed, the development of DL algorithms in HEP is a rapidly growing subfield in its own right, and its various applications are visualized as a "nomological net" in Figure 7.1; a comprehensive "living" review is available in Ref. [203]. As we discuss below, however, with more complex data and models also comes the need for more sophisticated methods to validate, calibrate, and trust them; this is the subject of Chapters **??** and **??**, on evaluating generative models and calibrating HVV jet taggers, respectively.

In this chapter, we first provide a brief introduction to ML and DL, emphasizing key aspects relevant to HEP. These include the importance of: 1) generalization and calibration of models trained on simulations (Section 7.1.2); and 2) the importance of building thoughtful, physics-informed models and representations for our data (Section 7.1.4). In the same spirit, we then discuss *equivariant neural networks* in Section 7.2, which are designed to respect the symmetries of physical data, such as Lorentz transformations in HEP. Finally, we detail two types of unsupervised learning algorithms, autoencoders and generative models, that are relevant to the work in this dissertation in Section 7.3.

Figure 7.1. A "nomological net" of ML applications in HEP, reproduced from Ref. [49].

7.1.1 Basics of ML

Supervised and unsupervised learning

ML algorithms can be broadly categorized as *supervised* and *unsupervised* learning. The former involves learning a mapping between some input data $\mathbf{x}$ and a specific output $\mathbf{y}$; for example, classifying jets as originating from a Higgs boson or QCD background. Other examples include regression tasks where the target output is a continuous variable, such as predicting the mass of a jet or the energy of a particle. Algorithms used for supervised learning include support vector machines (SVMs) [204], (boosted) decision trees (BDTs) [202, 205], and neural networks. Such algorithms necessitate a *labeled* training dataset of input-output pairs $(\mathbf{x}_i, \mathbf{y}_i)$.

Tasks for which we do not have straightforward labeled data are considered unsupervised learning problems, in which the model must learn the properties and structure of the data $\mathbf{x}$ without explicit target outputs. Examples include clustering algorithms, which aim to group similar data points together, and generative and anomaly detection models, both of which aim to learn the underlying distribution of the data in some manner for the purposes of generating new data or identifying outliers, respectively. The latter two will be discussed in more detail in Section 7.3.

Note that these two categories are not mutually exclusive but rather two ends of a spectrum, with the middle ground including paradigms such as weakly-supervised [206] and self-supervised learning [207].

Linear models

Perhaps the simplest example of an ML task is linear regression, which entails fitting a linear model:

$$f(x|w) = \mathbf{w} \cdot \mathbf{x} \tag{7.1.1}$$

to a set of data points $(\mathbf{x}_i, y_i)$, where $\mathbf{w}$ are the model *weights* which need to be learned. To do so, we define a *loss function* L that quantifies the difference between the model's prediction and our desired output, such as the mean squared error:

$$L = \frac{1}{N} \sum_{i=1}^{N} (f(x_i|w) - y_i)^2. \tag{7.1.2}$$

The learning objective of our model is hence to minimize L with respect to the weights $\mathbf{w}$.

For linear regression, the minimum can in fact be found analytically to be:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{7.1.3}$$

where $\mathbf{X}$ is the matrix of input data and $\mathbf{y}$ the vector of target outputs. However, for more complex models (or even in linear regression when the matrix inversion is too expensive), numerical optimization techniques are required. The most common is *gradient descent*.

Gradient descent

Gradient descent is an optimization algorithm that iteratively adjusts the weights of a model in the direction of steepest descent, i.e., the gradient:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla_w L, \tag{7.1.4}$$

where $\mathbf{w}_t$ are the weights at iteration t, and η is the step size or *learning rate* (LR). This process is visualized for two learnable parameters in Figure 7.2.

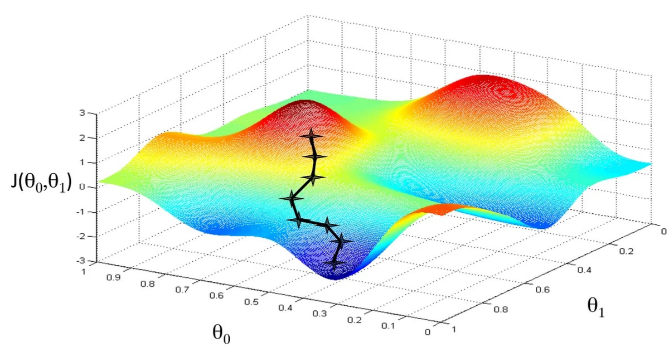


Figure 7.2. Illustration of gradient descent in a 2D parameter space of (θ_0, θ_1).

Gradient descent is the backbone of all deep learning optimization algorithms; though this basic idea is typically modified to improve convergence and efficiency. The most common variants are *stochastic* and *mini-batch* gradient descent, which compute the gradient on a subset of the data at each iteration. This has the dual benefit of computational efficiency and the introduction of stochasticity into the optimization process, which can help the model escape local minima of the loss function.

Other powerful ideas include *adaptive learning rates*, which adjust the LR during training based on the history of the gradients and/or number of iterations; and *momentum*, which retains some fraction of the previous gradients to smooth out oscillations in the optimization process. Popular optimizers which incorporate these techniques include RMSprop [208] and Adam [209], both of which are prominently used for the work in this dissertation.

7.1.2 The importance of generalization and calibration

It is crucial in ML that the model not only learns the training data but can also *generalize* to new, unseen data. This is what signifies that the model has effectively learned the underlying patterns and relationships, rather than merely memorizing, or *overfitting* to, the training samples.

A standard procedure to evaluate generalization is to split the available dataset into three subsets: training, validation, and testing. The former is the only dataset used to update the learnable parameters of the model themselves, and is typically the largest subset. The validation set is used to tune *hyperparameters* of the model — those parameters such as model size and learning rates that cannot be "learned" through gradient descent — as well as assess the model's performance on unseen data during training: if the performance on the validation set is significantly worse than on the training set, the model is likely overfitting. Finally, in case a bias is introduced by tuning the hyperparameters on the validation set, it is good practice to evaluate the model on the testing set at the end, which is never used to make decisions on the model.

The bias-variance tradeoff

Selecting the right model and hyperparameters involves making a *bias-variance tradeoff*. This is a fundamental concept in ML that describes the balance between two sources of error in a predictive model. *Bias* is the error due to overly simplistic assumptions in the learning algorithm — for example, using a linear model to capture non-linear relationships; while *variance* is the error due to a model which is too complex capturing noise in the training data.

A model with high bias may have systematic inaccuracies, or *underfit* the data, while a model with high variance may overfit and fail to generalize. Model selection involves using the performances on the training and validation datasets to find an optimal balance between these two errors. Common techniques to improve bias include improving the model design and increasing its complexity, while to address variance, there are several established *regularization* methods to reduce overfitting, such as early stopping [210], dropout [211], and batch normalization [212].

Model calibration

A related and unique aspect of ML in HEP is the reliance on theory and detector simulations to generate large quantities of labeled data for model training. The aim though, of course, is to deploy on and model correctly the real data collected by the experiments. It is hence crucial to verify how well the models generalize accurately to the latter, rather than overfitting to mismodeling in the former.

This process is sometimes referred to as *calibration*, where the performance of the ML model is compared between simulation and data to derive possible corrections to the model's predictions and quantify the systematic uncertainties associated with them. As models become more complex and high-dimensional, calibration becomes increasingly challenging (and often overlooked)! To this end, significant contributions of this dissertation are the development of novel methods to efficiently and sensitively validate the performance of ML-based simulations (Chapter **??**), and improving the calibration of HVV jet identification algorithms (Chapter **??**).

7.1.3 Artificial neural networks and deep learning

ANNs are ML models loosely inspired by the structure of the human brain. They were originally proposed in the 1940s, and improved over the 20th century through the perceptron [213] and backpropagation [214] algorithms, but had limited success in practical applications compared to algorithms like SVMs and decision trees.

Only in the 2010s was it recognized that their flexibility in both architecture and training makes them ideal for exploiting the recent exponential increase in data and computing power, propelling ANNs to the forefront of ML and sparking the so-called DL revolution. Through the development of large and innovative, so-called *deep* neural networks (DNNs), they have led to significant breakthroughs in the fields of computer vision, natural language processing, and indeed HEP. Specific types of models, or "architectures", include convolutional neural networks (CNNs) for image data, graph neural networks (GNNs) for graph data, and transformers for sets and sequences, all of which we discuss below.

Artificial neurons and multilayer perceptrons

The building blocks of ANNs are single "artificial neurons", or *perceptrons* [213]. They are similar to the linear models discussed above, but with an additional non-linear function σ — known as the *activation function*, applied to the output

(Figure 7.3, left):

$$f(x|w,b) = \sigma(\mathbf{w} \cdot \mathbf{x} + b), \tag{7.1.5}$$

where b is a constant, learned *bias* term. Common choices for the activation include the sigmoid, hyperbolic tangent, and piecewise linear functions.

By combining multiple perceptrons in a *multilayer perceptron* (MLP) architecture, i.e., an ANN, we can build a powerful and flexible model capable of learning complex, non-linear relationships in the data (Figure 7.3, right). In fact, the famous *universal approximation theorem* [215] states that, in theory, neural networks can approximate any continuous function to arbitrary accuracy given a sufficiently large number of neurons and layers (although in practice it is not so straightforward).

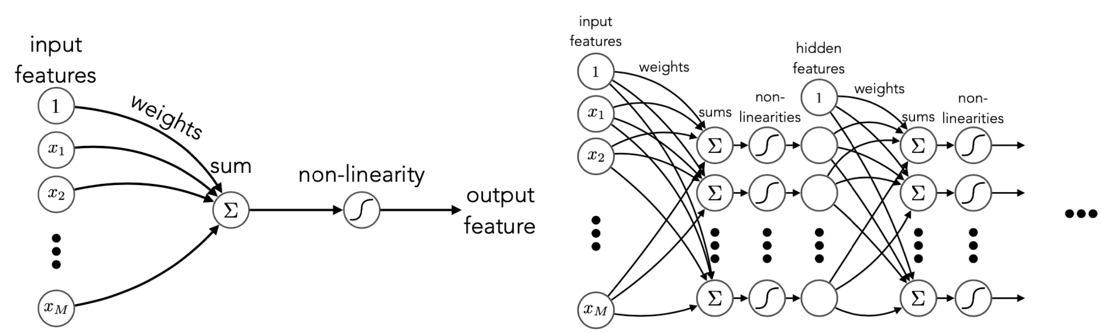


Figure 7.3. (Left) a single perceptron and (right) a neural network built using multiple layers of perceptrons (MLPs).

Another key characteristic of ANNs is their ability to learn hierarchical representations of the data, with each layer, in principle, learning progressively more abstract features from the previous layer's output. Intelligently designed "deep" networks with many layers can hence learn powerful, nonlinear, high-level representations of the high-dimensional input data, which can then be used to perform the desired task (assuming enough data and computing power to train them effectively). This is why this subfield of ML is also sometimes referred to as *representation learning*. As we discuss in Section 7.1.4, it is thus crucial to use representations and design architectures well-suited to the data and task at hand; naively adopting a specific architecture or input representation from another domain may not lead to the most optimal feature learning.

Backpropagation

Part of the effectiveness and popularity of DNNs is due to the backpropagation algorithm [214], which allows for efficient training of arbitrarily deep networks. Backpropagation is, essentially, the repeated application of the chain rule of calculus to iteratively propagate gradients of the loss function backwards through the network. For a simple two-layer network, for example:

$$f(x|w,b) = \sigma^{(2)}(\mathbf{w}^{(2)} \cdot (\sigma^{(1)}(\mathbf{w}^{(1)} \cdot \mathbf{x} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)}), \tag{7.1.6}$$

where the superscript denotes the layer of the network, the gradient of the loss function L with respect to $\mathbf{w}^{(1)}$ is:

$$\frac{\partial L}{\partial w^{(1)}} = \underbrace{\frac{\partial L}{\partial f}\frac{\partial f}{\partial \sigma^{(2)}}\frac{\partial \sigma^{(2)}}{\partial \mathbf{w}^{(2)}}}_{\partial L/\partial w^{(2)}}\underbrace{\frac{\partial \mathbf{w}^{(2)}}{\partial \sigma^{(1)}}\frac{\partial \sigma^{(1)}}{\partial \mathbf{w}^{(1)}}}_{\partial w^{(1)}/\partial w^{(2)}}. \tag{7.1.7}$$

This tells us that $\partial L/\partial w^{(1)}$ can be computed using the gradient with respect to $w^{(2)}$ — which needs to be calculated anyway — and, more generally, by walking backwards through the network operations and taking the product of the derivatives at each step. This simple but powerful idea scales well to large and diverse network architectures, and is why huge DNNs can be trained effectively with relative ease.

Convolutional neural networks

We now walk through some popular ANN architectures, starting with CNNs. CNNs are a type of NN designed to process grid-like data and, particularly, images. They contributed the first major breakthrough in DL by achieving impressive performances in computer vision tasks, with models such as AlexNet [216] in 2012 and ResNet [217] in 2016.

A single CNN convolutional layer convolves a set of discrete "kernels" (essentially, learnable matrices) through the input image or data (Figure 7.4), each of which detect useful features such as edges or textures. A CNN comprises multiple convolutional layers, interspersed with operations such as pooling or compression to reduce the spatial dimensions of the data, and then typically MLPs at the end as in Figure 7.4 to produce the final output.

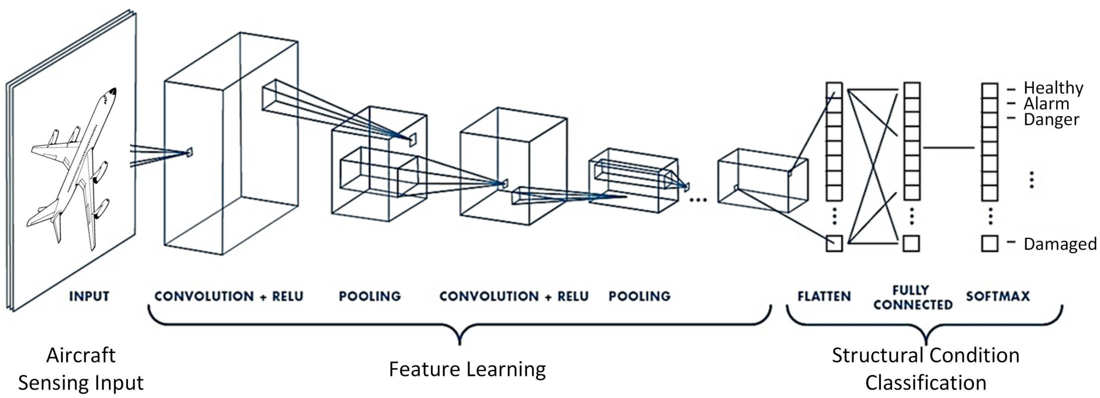


Figure 7.4. Schematic of a convolutional neural network, reproduced from Ref. [50].

Graph neural networks

GNNs are designed for graph-structured data, such as social networks or molecular structures. They are also useful for operating on *point clouds*: sets of unordered data points in some space, which we argue in Section 7.1.4 are the perfect data structures for representing particles in an event or hits in a detector. This is why GNNs have been extremely successful in HEP, generally outperforming standard MLP or CNN approaches.

The idea behind GNNs is to learn representations per-node or per-edge, based on information aggregated from their neighbors. Some generic methods to do so include local graph convolutions — similar to CNNs, but with graph-based kernels; and message-passing neural networks (MPNNs), which deliver and aggregate learned messages between nodes. An example of an MPNN is shown in Figure 7.5, and is the basis for a novel GNN generative model introduced in Chapter **??**.

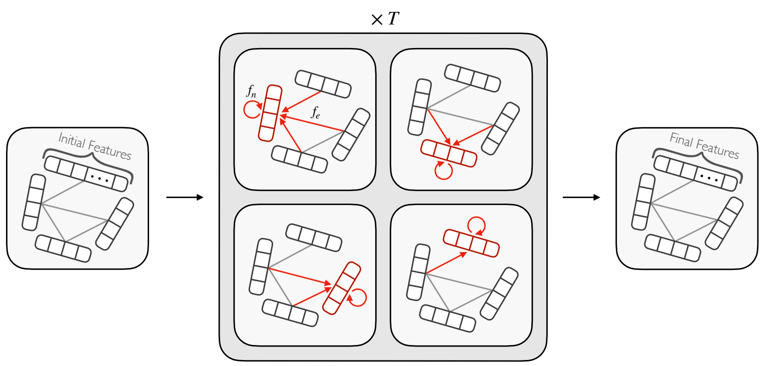


Figure 7.5. Schematic of a message passing graph neural network.

Attention and transformers

The final architecture we discuss is the transformer, introduced in 2017 [218], which is the powerhouse behind the recent revolution in natural language processing (NLP) and AI chatbots such as GPT-3 [219] and its successors. Transformers are built around the idea of *attention*, which encourages the model to learn to attend to different parts of an input set or sequence in each layer.

Explicitly, each element, or node's, features in the input set are first embedded via MLPs into *key* (K) and *value* (V) pairs, while each node in the output set is embedded into a *query* (Q). The attention mechanism is then defined as:

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V, \tag{7.1.8}$$

where d_k is the dimension of the keys and queries, and softmax $\left(QK^T/\sqrt{d_k}\right)$ are the "attention scores" between each pair of input and output nodes. This output A is finally used to update the features of the output nodes. Figure 7.6 shows a schematic of the special case of *self-attention*, in which the input set is also the output set; i.e., each node's features are updated based on the features of all other nodes.

Transformers can be thought of as a type of fully-connected GNN, with attention a (particularly efficient) form of message-passing. They have proven extremely successful and durable in NLP and other sequence-based tasks, and are also gaining prominence in computer vision and HEP. We introduce two novel transformer-based models for jet simulations and tagging in Chapters **??** and **??**, respectively.

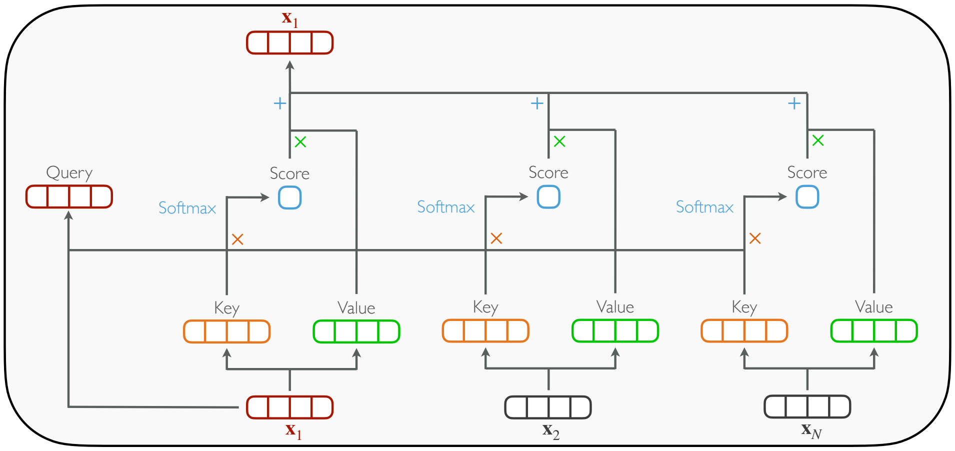


Figure 7.6. Schematic of set self-attention.

7.1.4 The importance of being physics-informed

The success of specific DNN models largely depends on the in-built *inductive biases* — assumptions or design choices — towards certain types of data. This is why it is important in HEP to build physics-informed models and representations that respect the symmetries and biases of our data. In this section, we outline the relevant properties of HEP data, such as jets and calorimeter showers, and the inductive biases of CNNs, GNNs, and transformers, arguing that the latter two are stronger fits.

The power of CNNs, in addition to their ease of computation, comes from their biases towards natural images, namely: *translation invariance* — the same features are learned regardless of input translations — and *locality* — the convolution operation is inherently local in space, suited to the structure of natural images. This led to CNNs leading the DL revolution in the 2010s and achieving results on par with or surpassing human performance in computer vision.

Consequently, this also led to early work in HEP applying CNNs to jets and calorimeter showers. Jets can, in principle, be represented as images by projecting the particle constituents onto a discretized angular η-ϕ plane, and taking the intensity of each "pixel" in this grid to be a monotonically increasing function of the corresponding particle p_{T} [220] (Figure 7.7, left). Showers can similarly be represented as 3D images of the energy deposited in the calorimeter cells (Figure 7.7, right).

At the time of the work of this dissertation, such image-based models were leading the field in tasks such as jet classification [221] and shower generation [222]. However, we argue that, despite these early successes, CNNs and images are not ideal for the physics and structure of our data, due to HEP data's following characteristics:

1. *Sparsity*: particles in a jet and hits in the detector tend to be extremely sparse relative to the total angular momentum phase space and total number of cells, respectively. Indeed, we see in Figure 7.7 that the resulting "images" tend to be extremely sparse, with typically fewer than 10% of pixels nonempty [223].
2. *High granularity*: LHC detectors are highly granular, which means the discretization process often lowers the spatial resolution (as with the ATLAS

FastCaloGAN [222]), unless the pixels are chosen to exactly match the detector cells; however, this is often computationally intractable due to the large number of cells, and the property we describe next.

3. *Irregular geometry*: jets and showers are not naturally square grid-like objects, and must be made to conform to this structure for use with CNNs. This is again often intractable or, at best, suboptimal.
4. *Global structure*: jets and particle showers each originate from a single or small set of sources, which leads to global correlations between the final-state particles and hits, independent of the spatial distance between them, that are vital to understanding the underlying physics.

Properties 1–3 strongly suggest that HEP data is not conducive to image-based representations. This is exemplified by the upcoming CMS HGCAL (Chapter 6.5.4): its high granularity, sparsity, hexagonal geometry, and non-uniform cell sizes all make HGCAL showers extremely challenging to represent as an image. Finally, Property 4 implies that local operations such as convolutions are ill-suited to the global structure of our data.

In contrast, GNNs and transformers are naturally: *sparse* — only the particles or hits need be represented, rather than a dense grid of mostly empty pixels; and *flexible* to the underlying geometry and granularity. Moreover, they are *permutation invariant* — learned features are independent of the order of the inputs, which means there is no need to impose an artificial ordering on particles or hits (as opposed to with an MLP, for example).

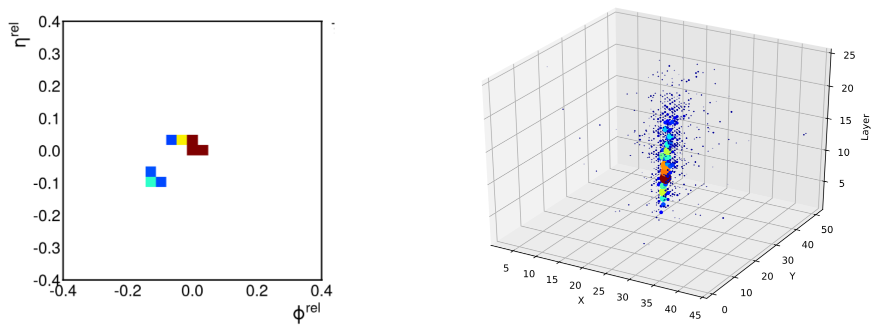


Figure 7.7. Examples of a jet (left) and calorimeter shower (right) represented as 2D and 3D images, respectively.

Finally, in the case of GNNs, the graph topology (i.e. the connections between nodes) can be tuned or even learned to reflect the physical nature of the data. For example, for local data, such as 3D point clouds of natural objects, connections can be defined based on the Euclidean distance between points, while in the case of jets or particle showers in a calorimeter, we can choose a fully-connected topology to reflect their global correlations (as we emphasize in Chapter **??**). The attention mechanism in transformers is by definition fully connected, and hence well-suited as well.

This is why we advocate for point-cloud representations and GNN and transformer models as natural choices for HEP data. Indeed, major contributions of this dissertation are the development of the first point-cloud based generative models for jet simulations (Chapter **??**), which achieve breakthrough performance for an ML simulator in terms of accuracy and efficiency, and the first transformer-based jet tagging algorithm (Chapter **??**) for HVV jet-tagging, powering a significant boost in the sensitivity of the HH search. Finally, in Chapter **??**, we push the inductive biases of ML models further by incorporating *equivariance* to Lorentz-symmetries, as we introduce next.

7.2 Equivariant neural networks

ANNs and DL have shown remarkable success in a wide range of computer vision and NLP tasks, motivating applications to the physical sciences. However, as highlighted in the previous section, the power of DL models is often derived from architectures tuned to the inductive biases of their domains.

A unique feature of physical data is its inherent physical *symmetries* (see Chapter 2), such as with respect to E(3) and Lorentz-transformations for molecules and high-energy collisions, respectively. It is hence desirable to develop NN architectures that themselves are intrinsically *equivariant* to the associated transformations, which can thereby be more data efficient, more easily interpretable, and perhaps ultimately more successful [224].

We have already encountered some forms of equivariance: to translations in CNNs and to permutations in GNNs. More recently, there has been work on building equivariance to a broader set of transformations, such as the symmetries mentioned above, which will be the focus of this section.

7.2.1 Equivariance

Let us first introduce precisely what we mean by "equivariance", adapting a definition from Refs. [225–228].

Definition 7.2.1. A feature map $f : \mathcal{X} \to \mathcal{Y}$ is considered **equivariant** to a group of transformations G if $\forall g \in G$ and some representation π there exists a representation π' satisfying

$$\pi'(g)f(x) = f(\pi(g)x), \tag{7.2.1}$$

i.e. the group operation commutes with the map f (and f therefore is an intertwiner). In this context, f generally represents a NN layer. Another way to think about this is that each transformation by a group element g on the input must correspond to a transformation by the same group element in the feature space (but with potentially different representations π and π').

Definition 7.2.2. **Invariance** is the particular case where π' is the trivial representation ($\pi'(g) = \mathbb{1}$), wherein transformations on x do not affect features at all.

While for many tasks, such as classification, *invariance* of the outputs is sufficient, Refs. [226, 227] argue that equivariance is more desirable at least in the intermediate layers, as it allows the network to learn useful information about the transformation g itself.

So far, we have discussed CNNs and GNNs / transformers, which are equivariant to the $\mathrm{T}(N)$ group (translations in N dimensions) and invariant to the S_N group (permutations of N objects), respectively. Next, we discuss the extension to broader symmetry groups.

7.2.2 Steerable CNNs for $\mathrm{E}(2)$-equivariance

We first describe the generalization of the translational invariance of CNNs to equivariance to not only translations, but *rotations and reflections* in 2D as well; i.e, the E(2) group. We make use of a general procedure, based on Refs. [225, 226], for extending 2D translational invariance (T(2)) to equivariance to a group $G = \mathrm{T}(2) \rtimes H$, where $\rtimes$ is the semi-direct product and H is a subgroup of G, meaning we can induce representations of G, Ind_H^G, from H.[1] For $G = \mathrm{E}(2)$, in particular, $H = \mathrm{O}(2)$, the group of distance-preserving transformations in 2D; i.e., rotations and reflections.

The key idea in developing a G-equivariant layer is to first find the set of maps $F \ni f$ which satisfy Eq. 7.2.1 for an element $h \in H$:

$$\rho_{\mathrm{out}}(h) f = f \rho_{\mathrm{in}}(h) \tag{7.2.2}$$

where ρ_{out} and ρ_{in} are reps of H. After this, Eq. 7.2.1 can be automatically satisfied

[1] See e.g. Chapter IV, p297 of Ref. [74] for induced representations of E(2).

using

$$\pi'(g)f = \mathrm{Ind}_H^G(g)f = \rho_{\mathrm{out}}(h)f(\rho_{\mathrm{in}}(h^{-1})(x-t)) \tag{7.2.3}$$

where $g = th$ for some 2D translation $t \in \mathrm{T}(2)$.

Since Eq. 7.2.2 is linear in f, we want a complete linear basis of functions that satisfy it. We can obtain this by restricting the convolutional filters of a standard CNN to circular harmonics [227]:[2]

$$W_m(r,\phi;R,\beta) = R(r)e^{i(m\phi+\beta)}, \tag{7.2.4}$$

where the radial component R and the filter phase β are learnable parameters. We can see that $m \in \mathbb{Z}$, these filters form a complete basis and satisfy Eq. 7.2.2 under convolutions ($*$) with an image $F(r,\phi)$ rotated by θ:

$$W_m * F(r,\phi+\theta) = e^{im\theta}W_m * F(r,\phi). \tag{7.2.5}$$

Here we took ρ_{in} to be the fundamental SO(2) rep acting on the image and ρ_{out} to be any one of the infinite complex reps. After discretizing these filters Ref. [227] demonstrates significant improvement in classification of rotated images compared to SOTA CNNs. Such networks are generally referred to as "Steerable CNNs", and, in practice, are implemented using a finite set of N such circular harmonic filters, with $m \in \{0, \frac{2\pi}{N}, \ldots, \frac{2\pi(N-1)}{N}\}$ (and possibly their reflections), which

[2] See Refs. [228, 229] for a more rigorous derivation.

are then pooled in a rotationally-invariant manner, as illustrated in Figure 7.8.

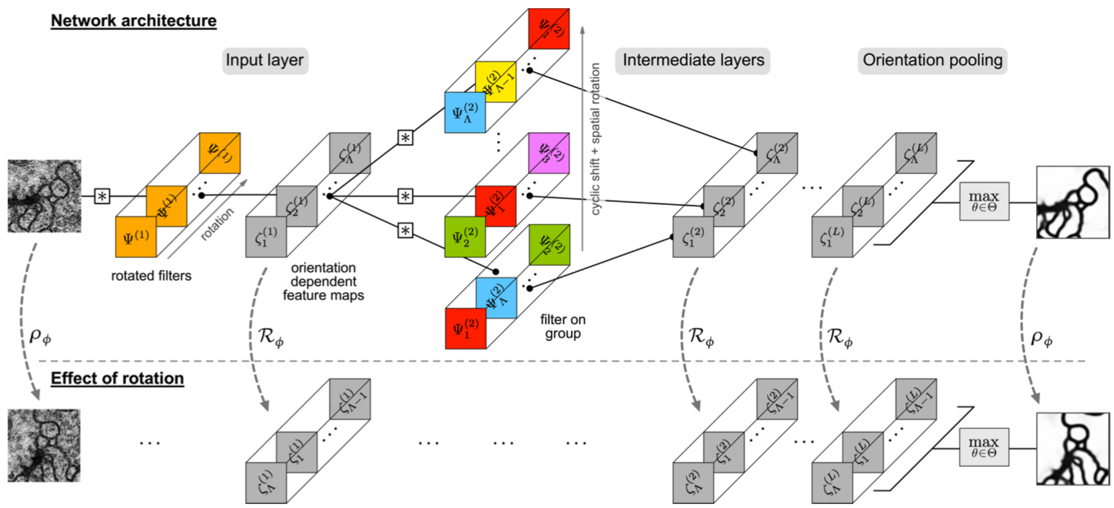


Figure 7.8. Schematic of a steerable CNN, reproduced from Ref. [51].

7.2.3 Tensor-field networks for E(3)-equivariance

Steerable CNNs have been extended to E(3)-equivariance — translations, rotations, and reflections in 3D — as well [229]. However, we will discuss a slightly different approach, applied to point-cloud data. This approach uses "Fourier decompositions" of the input, feature, and output spaces into irreducible representations (irreps) of the symmetry group, and is referred to as a "tensor-field network" [224]. In addition to their aforementioned applications to HEP, point clouds are also extremely useful representations of physical objects such as molecules and crystals, both of which are inherently E(3) invariant.

In the approach of Ref. [224], the input and intermediate network layers f take the set of coordinates $\mathbf{r}_a$ and features $\mathbf{x}_a$ for each point a in the point cloud and map them to the same set of coordinates with new learned features $\mathbf{y}_a$ ($f(\mathbf{r}_a, \mathbf{x}_a) = (\mathbf{r}_a, \mathbf{y}_a)$), with an equivariant f again having to satisfy Eq. 7.2.1. If necessary, the features are aggregated at the end across all points to produce the output. Translation equivariance is achieved directly by requiring f to only consider distances $\mathbf{r}_i - \mathbf{r_j}$ between points i and j (a global translation will not affect these).

For rotation equivariance, first the feature vectors $\mathbf{x}_a$ are decomposed according to how they transform under irreps of SO(3) — scalars, vectors or higher order tensors (the coordinates $\mathbf{r}_a$ already transform as vectors in $\mathbb{R}^3$ under the fundamental rep):

$$\mathbb{R}^3 \oplus \mathcal{X} = \bigoplus_l R_l^{m_l} \tag{7.2.6}$$

where the sum is performed over irreps R_l (with dimension $2l+1$) and m_l are the multiplicities. Thus, each point's features and coordinates have the corresponding decomposition:

$$\mathbf{r}_a \oplus \mathbf{x}_a = \bigoplus_l \bigoplus_{c=1}^{m_l} V_{ac}^l \tag{7.2.7}$$

where the V_{ac}^l are tensors which transform under the l irrep. Similar to steerable

CNNs, each of these tensors are individually acted upon by generalized convolutional filters with the form $R(r)Y^{l_f}(\hat{r})$, where R is a learned radial function, Y^l are the spherical harmonic tensors, and the set l_f corresponds to the set of desired irreps in feature space. The spherical harmonics are directly analogous to using circular harmonics for E(2) (except they have dimension $2l + 1$) and by the same argument they satisfy Eq. 7.2.1. This convolution effectively produces a tensor product representation of SO(3) $R_l \otimes R_{l_f}$, which is then decomposed via Clebsch-Gordan (CG) decomposition into irreps again.

A useful pedagogical example is of a network taking as input a collection of point masses and outputting the inertia tensor. The input features are the masses of each point, which are scalars under SO(3), and the inertia tensor transforms as the $0 \oplus 2$ representation, so we define this network to be of the type $0 \to 0 \oplus 2$.

Some interesting and successful applications include classifying molecules [230], predicting protein complex structures [231], and predicting the phonon density of states (DoS) in crystals [52]. A schematic of the architecture used for the latter is shown in Figure 7.9. Different crystals are represented geometrically as point clouds in $\mathbb{R}^3$, with individual atoms labeled via feature vectors $\mathbf{x}_a$ using mass weighted one-hot encoding. After a series of convolution layers the features are summed over all points to predict 51 scalars comprising the phonon DoS.

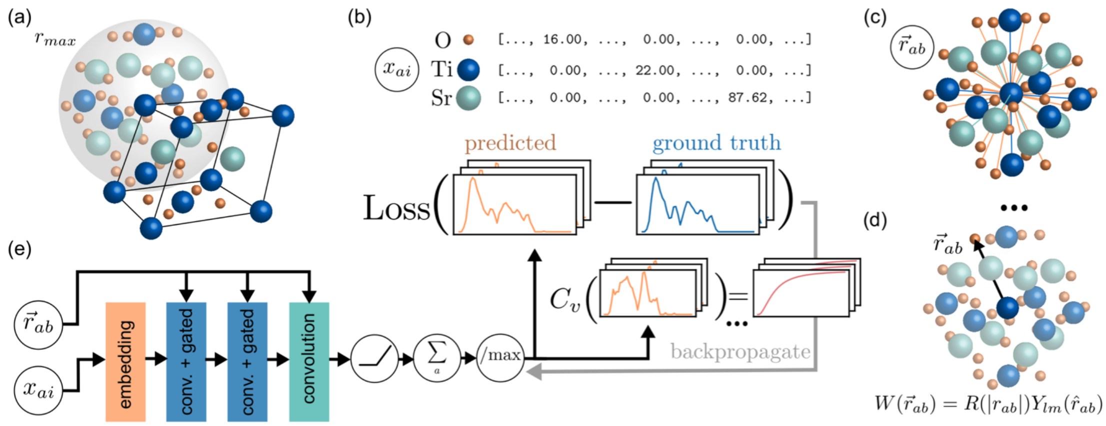


Figure 7.9. Schematic of the E(3)-equivariant neural network architecture used for predicting phonon density of states, reproduced from Ref. [52].

7.2.4 Lorentz-group-equivariant networks

Recently there has been some success in creating Lorentz-group-equivariant networks, which are desirable for DL applications to high energy data. The Lorentz group $O(3,1)$ comprises the set of linear transformations between inertial frames with coincident origins. Henceforth, we restrict ourselves to the special orthochronous Lorentz group $SO^+(3,1)$, which consists of all Lorentz transformations that preserve the orientation and direction of time. Equivariance to such transformations is a fundamental symmetry of the data collected out of high-energy particle collisions.

To our knowledge, there has been no generalization of steerable CNNs to the Lorentz group; however, Refs. [53, 232–234] propose an alternative, completely Fourier-based approach, again acting on point clouds, that shares some similarities with the E(3)-equivariant network discussed above.

The general method is to:

1. Decompose the input space into irreps of the group.
2. Apply an equivariant mapping (satisfying Eq. 7.2.1) to the feature space.
3. Take tensor products of the irreps and CG-decompose them again into irreps.
4. Repeat steps 2–3 until the output layer.

The crucial difference between this and the previous networks is that the mapping is no longer via convolutional filters; instead, it is chosen to be a simple linear aggregation across the nodes of the point clouds. Recall from Definition 7.2.1 that equivariant maps f must be intertwiners between input and output representations, which, according to Schur's Lemma, imposes strong restrictions on both the form of a linear f and its output $f(x)$. Namely: the outputs and inputs must have the same irrep decomposition (though the multiplicities are allowed to vary, akin to increasing/decreasing the "channels" in an image) and f must be a direct sum of learned matrices acting individually on each irrep. The transformation between f^{in} and f^{pre} in Figure 7.10 illustrates such a mapping.

To inject non-linearities into the network, Ref. [53] proposes to take tensor products between each pair of irreps after the mapping, and then perform a CG decomposition.[3] Another freedom available to us is acting with arbitrary learned

[3] See Ref. [235] for a detailed analysis of CG decomposition for the Lorentz group.

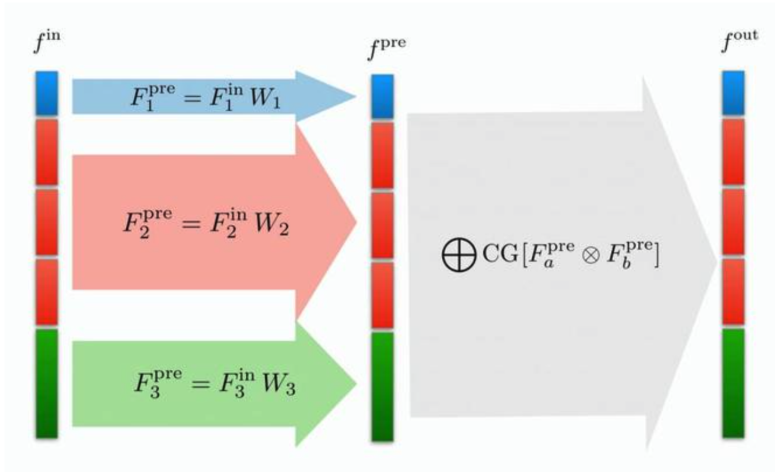


Figure 7.10. Schematic of a Lorentz group-equivariant network layer, reproduced from Ref. [53].

functions on any scalar irreps that result from the decomposition, since they are, by definition, Lorentz-invariants.

One successful application of this network has been to jet tagging: Ref. [53] successfully applied this "Lorentz-group network" (LGN) to top-quark identification, demonstrating a high (92.9%) accuracy, though they were unable to match the then-SOTA (93.8% using the ParticleNet GNN [223]).

Finally, we note that overall this is, in fact, a very general approach: applicable to any symmetry group. This includes the aforementioned E(2) and E(3) groups as well as potentially more exotic groups such as E_8 or G_2 which also arise in physics. The only group-dependent operations in such a network are the decompositions into irreps which can readily be calculated for any group (as opposed to steerable CNNs where one needs to derive group equivariant kernels/convolutional filters).

Summary

We reviewed three approaches to creating neural networks that are equivariant to physical symmetry groups: by extending the translation-equivariant convolutions in CNNs to more general symmetries with appropriately defined learnable filters as in Refs. [225, 236, 237], by operating in the Fourier space of the group [53], and a combination thereof [224]. Such networks are highly relevant to the physical sciences, where datasets often possess intrinsic symmetries, and, as demonstrated in some example tasks, they are promising alternatives and improvements to standard non-equivariant DL approaches. In particular, Lorentz-equivariant networks have shown promise in jet classification, a key task in HEP. In Chapter **??**, we will discuss the extension of these ideas to the first Lorentz-equivariant *autoencoder* for jets, with applications to data compression, anomaly detection, and potentially fast simulations as well.

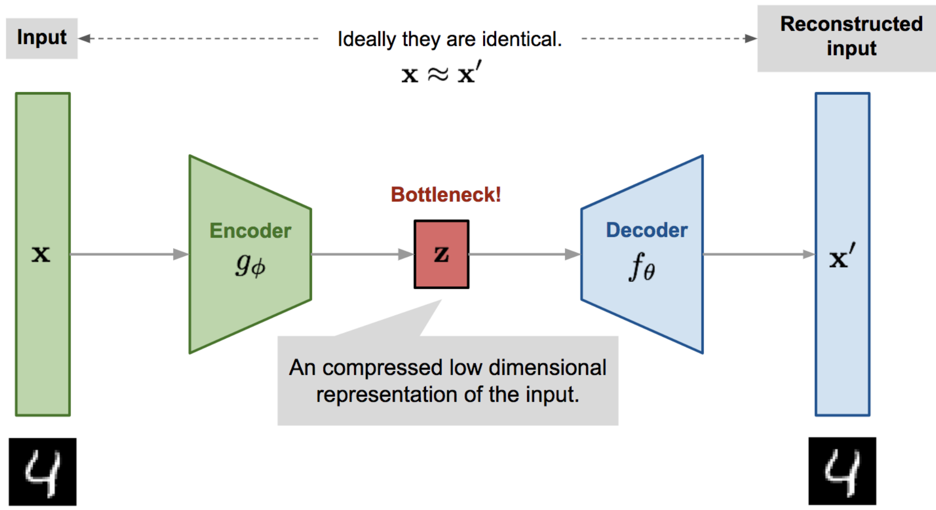


Figure 7.11. Diagram of an image autoencoder, reproduced from Ref. [54].

7.3 Autoencoders and generative models

7.3.1 Autoencoders and anomaly detection

In this final section, we discuss two paradigms of unsupervised learning relevant to this dissertation: autoencoders (AEs) and generative models. AEs are NN architectures composed of an *encoder* network, which maps the input into a typically lower dimensional latent space — called a "bottleneck" — and a *decoder*, which attempts to reconstruct the original input from the latent features (Figure 7.11). The bottleneck encourages AEs to learn a compressed representation of data that captures salient properties [238], which can be valuable in HEP for compressing the significant volumes of data collected at the LHC [239].

The learned representation can also be exploited for later downstream tasks, such as anomaly detection, where an autoencoder is trained to reconstruct data considered "background" to our signal, with the expectation that it will reconstruct the signal worse than the background. Thus, examining the reconstruction loss of a trained autoencoder may allow the identification of anomalous data. This can be an advantage in searches for new physics, since instead of having to specify a particular signal hypothesis, a broader search can be performed for data incompatible with the background. This approach has been successfully demonstrated in Refs. [240–248]. Two recent exciting examples from CMS include a model-agnostic search for di-jet resonances with Run 2 data [249], which prominently uses AEs for multiple search strategies, and a new AE-based online Level-1 trigger paths implemented in Run 3 [250, 251].

Furthermore, there are many possible variations to the general autoencoder framework for alternative tasks [252, 253], such as variational autoencoders (VAEs) [254], which we discuss in the next section. While there have been some recent efforts at GNN-based autoencoder models [62, 255], in this dissertation, we present the first Lorentz-equivariant autoencoder for jets in Chapter **??**. We focus on data compression and anomaly detection but note that our model can be extended to further applications, such as fast simulations in HEP.

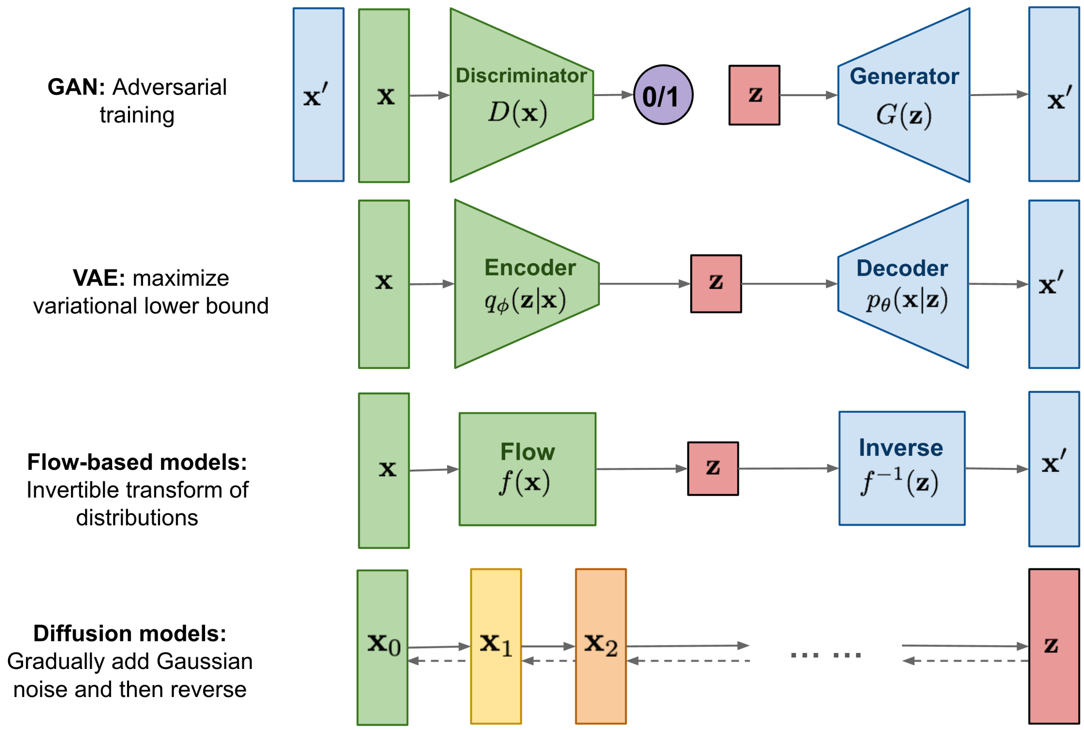


Figure 7.12. Summary of popular generative models, reproduced from Ref. [55].

7.3.2 Generative models

Generative models are a class of statistical models that aim to capture the probability distribution of the data $p(x)$ in order to generate new samples. This is a challenging problem, but one that has seen significant progress in recent years with DL, particularly in computer vision and NLP. We will briefly walk through four popular approaches, illustrated in Figure 7.12, which can be broadly categorized as *likelihood-based* or *implicit* models.

Likelihood-based models

Likelihood-based models attempt to directly learn the probability distribution of the data through some form of (approximate) likelihood maximization.[4] Flow-based models, for example, learn a series of invertible transformations to map a simple base distribution that is easy to sample from, such as a Gaussian, to the complex target data distribution. The most popular of these at the time of writing are "normalizing flows", which require each transformation to have a tractable Jacobian determinant with which to correctly normalize the result.

Normalizing flows have a number of advantages, such as their simple and intuitive training objective — maximizing the likelihood of each data point — and a tractable likelihood evaluation. These have led to successful applications to density estimation and generation tasks in both computer vision [256, 257] and HEP [258]. However, the constraint of invertible transformations with tractable Jacobians turns out to be extremely restrictive on the model design and expressivity in practice [259, 260], generally resulting in worse performance on high-dimensional data compared to the models we discuss below. Recently, over the last year, a related (in spirit) class of models without the normalization constraint, called "flow-matching" models, have emerged with extremely promising and, in some cases, state-of-the-art (SOTA) results on images [261, 262].

Another example of a likelihood-based model is the variational autoen-

[4]See the next chapter for an introduction to likelihoods.

coder (VAE) [254], which is structurally similar to an AE in that it has an encoder mapping an input data point into a latent representation, and a decoder mapped that back to the original. They key novelty, however, is that the latent space is encouraged through the loss function to follow a well-defined simple distribution to sample from — again, typically, a Gaussian. Explicitly, the VAE loss function is a combination of the reconstruction loss of a standard AE and the Kullback-Leibler divergence between the learned latent distribution and the assumed prior. Together, this can be shown to approximate the evidence lower bound (ELBO) of the true likelihood [254], which is why VAEs are thought of as likelihood-based.

VAEs were one of the early success stories in generative modeling, with a relatively simple implementation, training, and learning objective. However, they again are restrictive, this time due to the strong assumption imposed on the latent space, which actually competes with the reconstruction objective, and which, if incorrect, limits the performance. Indeed, our early studies in HEP showed that the learned latent space of VAEs is manifestly non-Gaussian for jets, leading to suboptimal performance with a Gaussian latent prior [263]. This is why VAEs also generally yield poorer performance than *generative adversarial networks* (GANs) [264], which we discuss next.

GANs

GANs are a type of implicit generative model. This means they learn to generate samples without directly learning the likelihood of the data. Instead, their loss function is effectively provided by a second neural network, called the discriminator or critic, which tries to distinguish between real and generated samples. The two generator and discriminator networks, with the former aiming to fool the latter, are trained iteratively and *adversarially*, forming a feedback loop and progressively improving each other. This continues until, ideally, the duo converge to a point where the generator produces samples indistinguishable from the real by the discriminator.

GANs have an interesting game-theoretic interpretation as a minimax game, where the Nash equilibrium, or global optimum, is achieved through minimizing the Jensen-Shannon divergence between the real and generated data distributions [264]. Several variations of GANs have also been proposed, including the Wasserstein-GAN [265], which instead aims to minimize the Wasserstein distance between the two distributions.

Due to the adversarial nature of the training, GANs are notoriously difficult to train [265–268]. However, their formulation poses no restrictions on the form of the generator while providing a powerful loss function and feedback mechanism. When trained successfully, this leads to expressive, flexible, and extremely successful generative models in a wide variety of domains. Indeed, at the time of the work of this dissertation, GANs were the SOTA in computer vision [269–271] and had shown promising signs in HEP as well [222, 272**? ? ?** , 273]. However, as we highlight below, there had been no successful application of GANs, or indeed any generative model, to point cloud data and GNNs or transformers in HEP.

Score-based diffusion models

Finally, we briefly note the recent development in the past two years of a new class of generative models, called *diffusion* or *score-based* models [274, 275]. These models iteratively "denoise" initial Gaussian noise into something resembling samples from the true data distribution; conceptually, this is related to diffusion in physical systems. The breakthrough with these models came from recognizing that, with the right learning objective, this denoising process is in fact equivalent to following the gradient of the log-likelihood function, AKA the *score*.

Diffusion models allow a likelihood-driven training objective, like flow-based models, but without the restrictive constraints (as the score does not need to be normalized!), thereby offering the flexibility of a GAN along with a far more stable training procedure. This, combined with several innovations in training and inference techniques, has led to diffusion models surpassing GANs in computer vision [276], and showing promising signs over the last year in HEP as well (in part enabled by the work in this dissertation, as we discuss in

Chapter **??**). However, so far, diffusion models remain computationally expensive, with inference naively requiring up to hundreds of denoising steps, which limits their application to fast simulations. Nevertheless, they are an exciting area for exploration in future work.

7.3.3 Previous work

Note: the following discussion represents the state of the field at the time of our first publications in 2021, to provide context. Since then, the field has evolved significantly, partly due to the work presented in this dissertation, as we discuss in Chapter **??**.

Generative modeling in HEP

Past work in generative modeling in HEP exclusively used image-based representations for HEP data. The benefit of images is the ability to employ CNN-based generative models, which have been highly successful on computer vision tasks. References [222, 272**? ? ?** , 273], for example, build upon CNN-based GANs, and Ref. [277] uses an autoregressive model, to output jet- and detector-data-images.

However, as highlighted in Section 7.1.4, the high sparsity of the images can lead to training difficulties in GANs, while the irregular geometry of the data poses a challenge for CNNs which require uniform matrices. While these challenges can be mitigated to an extent with techniques such as batch normalization [212] and using larger/more regular pixels [273], the approach we develop avoids both issues by generating particle-cloud-representations of the data, which are inherently sparse data structures and completely flexible to the underlying geometry.

Point cloud generative modeling

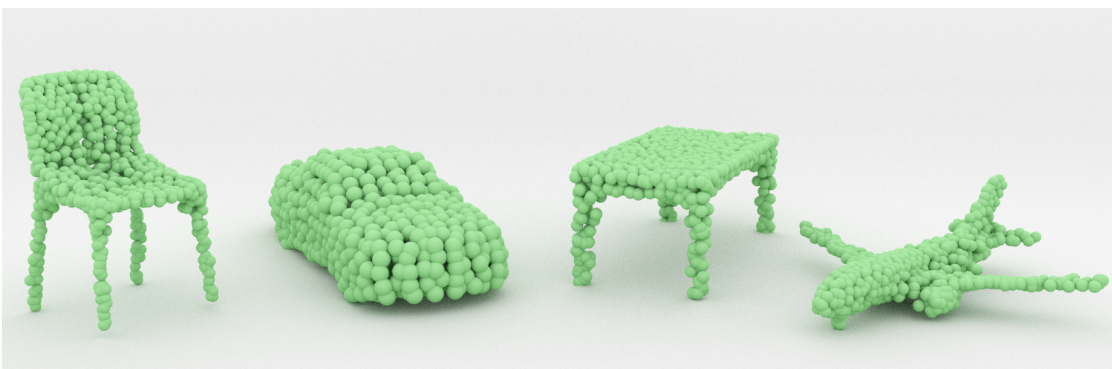


Figure 7.13. Sample point clouds from the ShapeNet dataset, reproduced from Ref. [56].

Prior to this work, point cloud generative approaches had not yet been developed in HEP; however, there had been some work in computer vision, primarily for 3D objects like those from the ShapeNet dataset [278]. As shown in Figure 7.13, ShapeNet comprises point clouds derived by sampling everyday objects in position space, and are thus naively analogous to the particle cloud representations in momentum space we employ for jets. However, as we note next, there are important differences in the inductive biases of the two datasets.

Firstly, jets have physically meaningful low- and high-level features such as particle momentum, total mass of the jet, the number of sub-jets, and n-particle energy correlations. These physical observables are how we characterize jets, and hence are important to reproduce correctly for physics analysis applications. Secondly, unlike the conditional distributions of points given a particular ShapeNet object, which are identical and independent, particle distributions within jets are highly correlated, as the particles each originate from a single source. The independence of their constituents also means that ShapeNet-sampled point clouds can be chosen to be of a fixed cardinality, whereas this is not possible for jets, which inherently contain varying numbers of particles due to the stochastic nature of particle production. Indeed, the cardinality is correlated with other jets features such as the jet mass and type.

Baseline models from computer vision

Still, particle clouds and point clouds have similarities insomuch as they represent sets of elements in some physical space, hence we first test existing point cloud GANs as baseline comparisons on JetNet. There are several published generative models in this area; however, the majority exploit inductive biases specific to their respective datasets, such as ShapeNet-based [279–282] and molecular [283–285] point clouds, which are not appropriate for jets. A more detailed discussion, including some experimental results, can be found in App. **??**.

There do exist some more general-purpose GAN models, namely r-GAN [286], GraphCNN-GAN [287], and TreeGAN [288], and we test these on JetNet. r-GAN uses a fully-connected (FC) network, GraphCNN-GAN uses graph convolutions based on dynamic k-nn graphs in intermediate feature spaces, and TreeGAN iteratively up-samples the graphs with information passing from ancestor to descendant nodes. In terms of discriminators, past work has used either an FC or a PointNet [289]-style network. Ref. [290] is the first work to study point cloud discriminator design in detail and finds amongst a number of PointNet and graph convolutional models that PointNet-Mix, which uses both max- and average-pooled features, is the most performant.

In Chapter **??**, we apply the three aforementioned generators and FC and PointNet-Mix discriminators to our dataset, but find jet structure is not adequately reproduced. GraphCNN's local convolutions make learning global structure difficult, and while the TreeGAN and FC generator + PointNet discriminator combinations are improvements, they are not able to learn multi-particle correlations, particularly for the complex top quark jets, nor deal with the variable-sized light quark jets to the extent necessary for physics applications. We thus aim to overcome limitations of existing GANs by designing novel generator and discriminator networks that can learn such correlations and handle variable-sized particle clouds.

Acknowledgements

This chapter is, in part, a reprint of the materials as they appear in R. Kansal. "Symmetry Group Equivariant Neural Networks," (2020); and NeurIPS, 2021, R. Kansal; J. Duarte; H. Su; B. Orzari; T. Tomei; M. Pierini; M. Touranakou; J.-R. Vlimant; and D. Gunopulos. Particle Cloud Generation with Message Passing Generative Adversarial Networks. The dissertation author was the primary investigator and author of these papers.

Chapter 8

Data Analysis and Statistical Interpretation

There are two possible outcomes: if the result confirms the hypothesis, then you've made a discovery. If the result is contrary to the hypothesis, then you've made a discovery. — Enrico Fermi

8.1 Introduction or: What is an analysis?

Once our data is collected by the CMS detector and reconstructed offline, it is analyzed to search and measure processes of interest. Typically, the raw data is entirely dominated by irrelevant background processes which we want to filter out in favor of the signal. The first step towards this is through appropriate online triggers, followed by offline selections to isolate the signal. The advent of machine learning, and later deep learning (DL), allows for more sophisticated selections, using increasingly lower-level information such as individual particles in jets, tracks and clusters, and even detector hits, as we introduced in Chapter 7.

Optimizing the event selection for all but a handful of data-driven searches requires simulations of the signal and background processes. Additionally, once the selections and phase space in which to perform the measurement have been finalized, the expected signal and background yields have to be carefully estimated, which often again necessitates simulations, as well as data-driven methods via unbiased control regions. Given the importance of simulations, it is critical to ensure sufficient quality and quantity of simulations in the HL-LHC era; Part IV will discuss efforts towards using DL.

Once we have our observations, and signal and background estimates, the final critical step is to interpret the results in a robust statistical framework. At the LHC, this is typically done using a frequentist, likelihood-based approach. In this chapter, this approach is introduced by way of simple experimental examples.

The chapter is organized as follows. Section 8.2.1 introduces the concepts of the likelihood functions and test statistics, with Section 8.2.2 discussing the framework for hypothesis testing, including p-values, significances, and the statistical definition of a "discovery". Sections 8.2.3 and 8.2.4 then describe frequentist confidence intervals and upper limits, and the important concepts of expected significances and limits, respectively. Finally, asymptotic approximations to simplify these computations are discussed Section 8.3.

The chapter is based primarily on the highly useful Refs. [57, 291]. The code for all the plots and results in this chapter is available at rkansal47.github.io/stats-for-hep; it makes extensive use of the `NumPy` [292], `SciPy` [293], and `matplotlib` [294] Python libraries.

8.2 Frequentist statistics at the LHC

8.2.1 The likelihood function and test statistics

The data model

Let us take the simplest possible case of a (one bin) counting experiment, where in our "signal region" we expect s signal events and b background events. The probability to observe n events in our signal region is distributed as a Poisson with mean $s + b$:

$$P(n; s, b) = \mathrm{Pois}(n; s + b) = \frac{(s + b)^n e^{-(s+b)}}{n!} \tag{8.2.1}$$

Since we have only one observation but two free parameters, this experiment is underconstrained. So, let's also add a "control region" where we expect no signal and b background events. The probability of observing m events in our control region is therefore:

$$\mathrm{Pois}(m; b) = \frac{b^m e^{-b}}{m!} \tag{8.2.2}$$

Combining the two, the joint probability distribution for n and m is:

$$P(n, m; s, b) = \mathrm{Pois}(n; s+b) \cdot \mathrm{Pois}(m; b) = \frac{(s+b)^n e^{-(s+b)}}{n!} \cdot \frac{b^m e^{-b}}{m!} \tag{8.2.3}$$

This is also called the model for the data and is plotted for sample s, b values in Figure 8.1.

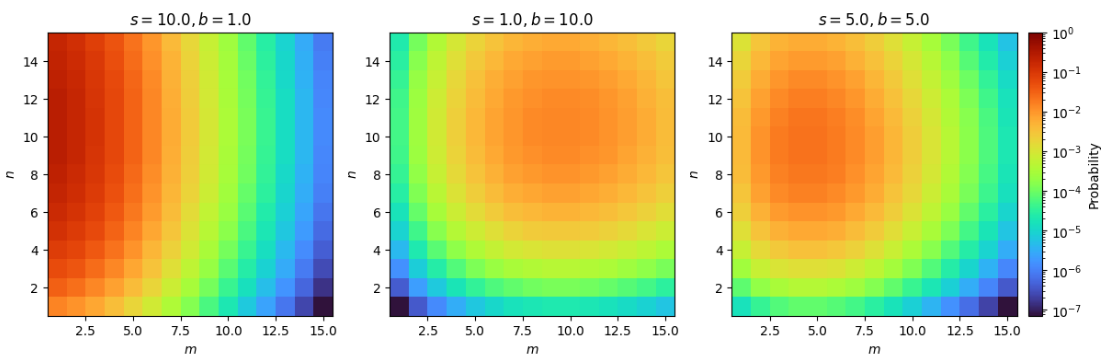


Figure 8.1. Sample 2D Poisson distributions.

The likelihood function

In the *frequentist* philosophy, however, all our parameters n, m etc. are simply fixed values of nature and, hence, don't have a probability distribution. Instead, we work with the *likelihood function*, which is a function only of our parameters of interest (POIs), s in our example, and "nuisance parameters" (b), given fixed values for n and m:

$$L(s, b) = P(n, m; s, b) = \frac{(s+b)^n e^{-(s+b)}}{n!} \cdot \frac{b^m e^{-b}}{m!}. \tag{8.2.4}$$

Importantly, this is not a probability distribution on s and b! To derive that, we would have to use Bayes' rule to go from $P(n, m; s, b) \to P(s, b; n, m)$; however, such probability distributions don't make sense in our frequentist world view, so we're stuck with this likelihood formulation. Often, it's more convenient to consider the negative log-likelihood:

$$-\ln L = \ln n! + \ln m! + s + 2b - n\ln(s+b) - m\ln b \tag{8.2.5}$$

The profile likelihood ratio

Fundamentally, the goal of any experiment is to test the compatibility of the observed data (n, m here) with a certain hypothesis H. We do this by mapping the data to a "test statistic" t, which is just a number, and comparing it against its distribution under H, $P(t|H)$. Our problem, thus, boils down to 1) choosing the most effective t for testing H, and 2) obtaining $P(t|H)$.

In the case of testing a particular signal strength, we use the "profile likelihood ratio":

$$\lambda(s) = \frac{L(s, \hat{\hat{b}}(s))}{L(\hat{s}, \hat{b})}, \tag{8.2.6}$$

where $\hat{s}, \hat{b}$ are the maximum-likelihood estimates (MLEs) for s and b, given the observations n, m, and $\hat{\hat{b}}(s)$ is the MLE for b given $n, m,$ *and* s. The MLE for a parameter is simply the value of it for which the likelihood is maximized, and will be discussed in the next section. The numerator of $\lambda(s)$ can be thought of as a way to "marginalize" over the nuisance parameters by simply values that maximize the likelihood for any given s, while the denominator is effectively a normalization factor, such that $\lambda(s) \leq 1$.

Again, it's often more convenient to use the (negative) logarithm:

$$t_s = -2 \ln \lambda(s) \tag{8.2.7}$$

Note that $\text{Max}[\lambda(s)] = 1 \Rightarrow \text{Min}[t_s] = 0$. $\lambda(s)$ and t_s are plotted for sample n, m values with $n - m = 10$ in Figure 8.2. The maximum (minimum) of the profile likelihood ratio (t_s) is at $s = n - m = 10$, as we expect; however, as the ratio between n and m decreases — i.e., the experiment becomes more noisy — the distributions broaden, representing the reduction in sensitivity, or the higher uncertainty on the true value of s.

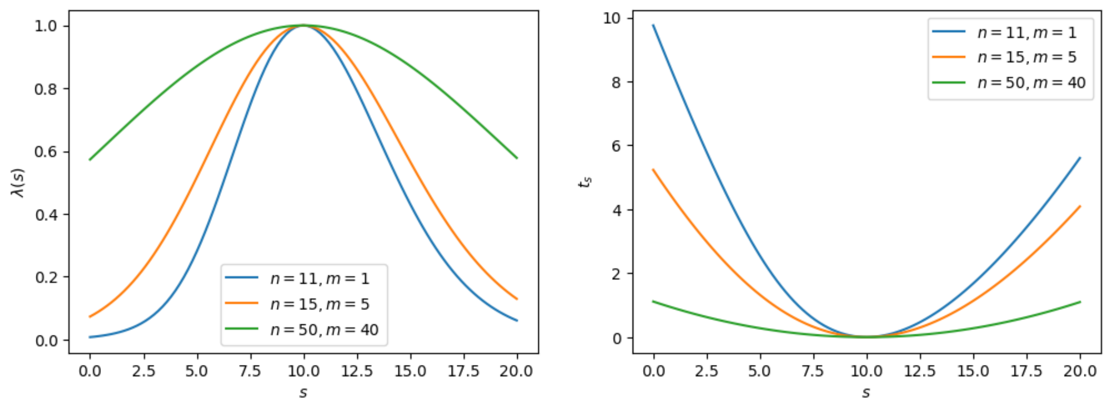


Figure 8.2. The profile likelihood ratio $\lambda(s)$ (left) and the t_s test statistic (right) for our one-bin Poisson model.

Note that the likelihood ratio and t_s are also broadened due to the nuisance parameter; i.e., because we are missing information about b. This can be demonstrated by plotting them with $b = m$, emulating perfect information of b (Figure 8.3), and indeed, we see the functions are narrower than in Figure 8.2. More generally, increasing (decreasing) the uncertainties on the nuisance parameters will broaden (narrow) the test statistic distribution. This is which is why experimentally we want to constrain them through auxiliary measurements as much as possible.

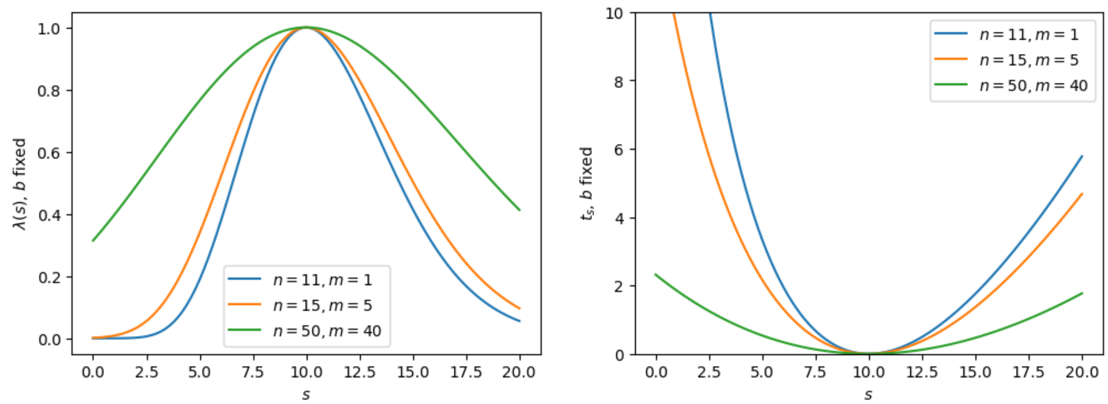


Figure 8.3. The profile likelihood ratio $\lambda(s)$ (left) and the t_s test statistic (right) with $b = m$, demonstrating the effect of decreasing uncertainties on our nuisance parameters.

Maximum-likelihood estimates

MLEs for s and b can be found for this example by setting the derivative of the negative log-likelihood to 0 (more generally, this would require numerical minimization):

$$\frac{\partial(-\ln L)}{\partial s} = 1 - \frac{n}{s+b} = 0 \tag{8.2.8}$$

$$\frac{\partial(-\ln L)}{\partial b} = 2 - \frac{n}{s+b} - \frac{m}{b} = 0 \tag{8.2.9}$$

Solving simultaneously yields, as you might expect:

$$\hat{b} = m, \hat{s} = n - m, \tag{8.2.10}$$

Or just for $\hat{\hat{b}}(s)$ from Eq. 8.2.8:

$$2b^2 + (2s - n - m)b - ms = 0 \tag{8.2.11}$$

Plugging this back in, we can get $\lambda(s)$ and t_s for any given s.

Alternative test statistic

So far, our construction allows for $s < 0$; however, physically the number of signal events can't be negative. Rather than incorporating this constraint in the model, it's more convenient to impose this in the test statistic, by defining:

$$\tilde{\lambda}(s) = \begin{cases} \frac{L(s,\hat{\hat{b}}(s))}{L(\hat{s},\hat{b})}, & \hat{s} \geq 0. \\ \frac{L(s,\hat{\hat{b}}(s))}{L(\hat{0},\hat{\hat{b}}(0))}, & \hat{s} < 0. \end{cases}, \tag{8.2.12}$$

and

$$\tilde{t}_s = -2 \ln \tilde{\lambda}(s) \tag{8.2.13}$$

The difference between the nominal and alternative test statistics is highlighted in Figure 8.4. For $n < m$, the $\tilde{\lambda}(s) = 1$ and $\tilde{t}_s = 0$ values are at $s = 0$, since physically that is what fits best with our data (even though the math says otherwise).

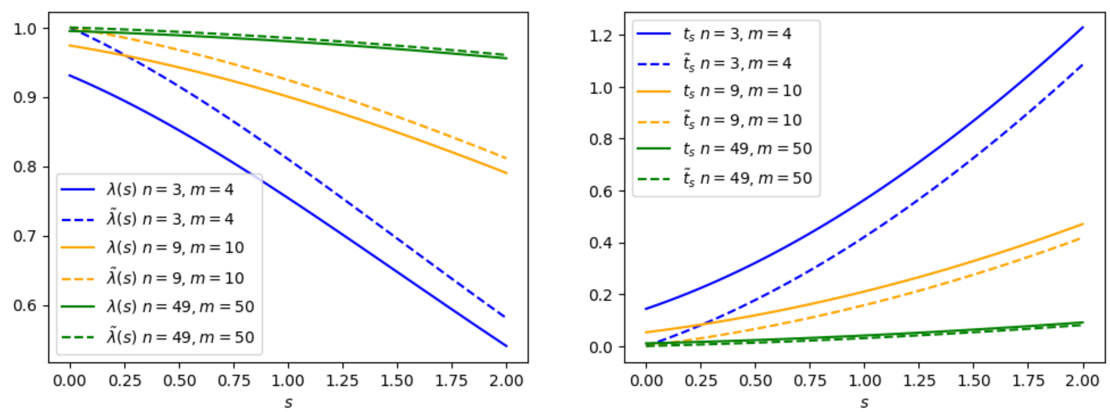


Figure 8.4. Comparing the nominal vs alternative test statistic.

Next, we want to translate this to a probability distribution of $\tilde{t}_s$ under a particular signal hypothesis (H_s) (i.e., an assumed value of s): $p(\tilde{t}_s|H_s)$, or just $p(\tilde{t}_s|s)$ for simplicity.

8.2.2 Hypothesis testing

The goal of any experiment is to test whether our data support or exclude a particular hypothesis H, and quantify the (dis)agreement. For example, to what degree did our search for the Higgs boson agree or disagree with the standard model hypothesis?

We have already discussed the process of mapping data to a scalar test statistic t that we can use to test H. However, we need to know the probability distribution of t under H to quantify the (in)consistency of the observed data with H and decide whether or not to exclude H.

We must also recognize that there's always a chance that we will exclude H even if it's true (called a Type I error, or a false positive), or not exclude H when it's false (Type II error, or false negative). The probability of each is referred to as α and β, respectively. This is summarized handily in Table 8.1.

Table 8.1. Table of error types, reproduced from Ref. [60].

<table>
<tr><th colspan="2" rowspan="2">Table of error types</th><th colspan="2">Null hypothesis (H_0) is</th></tr>
<tr><th>True</th><th>False</th></tr>
<tr><td rowspan="2">Decision about null hypothesis (H_0)</td><td>Fail to reject</td><td>Correct inference
(true negative)
(probability = $1-\alpha$)</td><td>Type II error
(false negative)
(probability = β)</td></tr>
<tr><td>Reject</td><td>Type I error
(false positive)
(probability = α)</td><td>Correct inference
(true positive)
(probability = $1-\beta$)</td></tr>
</table>

Before the test, we should decide on a probability of making a Type I error, α, that we are comfortable with, called a "significance level". Typical values are 5% and 1%, although if we're claiming something crazy like a new particle, we better be very sure this isn't a false positive; hence, we set a much lower significance level for these tests of 3×10^{-7}. (The *significance* of this value will be explained in Section 8.2.2 below.)

Deriving $p(\tilde{t}_s|s)$

We can approximate $p(\tilde{t}_s|s)$ by generating several pseudo- or "toy" datasets assuming s expected signal events. In this case, this means sampling possible values for n, m from our probability model. We will continue with our simple counting experiment (Section 8.2.1), for which such toy datasets are generated and then used to create histograms for $p(\tilde{t}_s|s)$ in Figure 8.5. Note that one complication in generating these toys is that the n, m distributions from which we want to sample (Eq. 8.2.3) also depend on the nuisance parameter b. However, we see from the figure that this does not matter as much as we might expect.

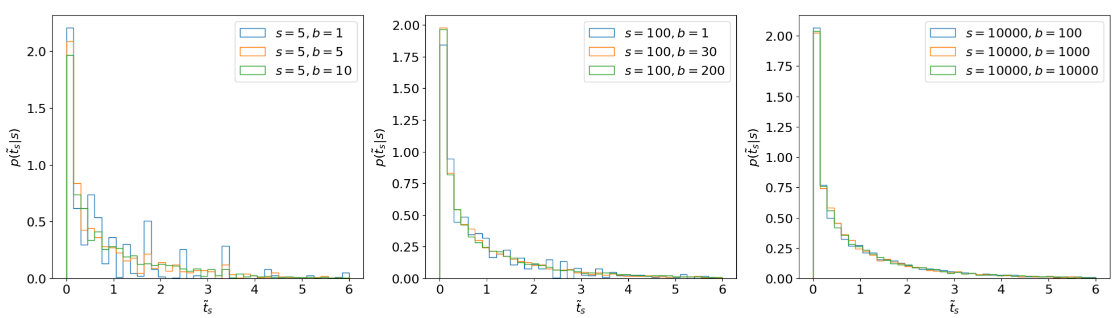


Figure 8.5. Estimating $p(\tilde{t}_s|s)$ through toys.

We make two important observations:

1. $p(\tilde{t}_s|s)$ does not depend on nuisance parameters as long as we have sufficiently large statistics (in this case, when b is sufficiently large). This is a key reason for basing our test statistic on the profile likelihood.

2. In fact, $p(\tilde{t}_s|s)$ doesn't even depend on the POI s! (Again, as long as s is large.)

Reference [291] shows that, asymptotically, this distribution follows a χ^2 distribution with degrees of freedom equal to the number of POIs, as illustrated in Figure 8.6.[1] We can see that the asymptotic form looks accurate even for s, b as low as ~ 5. Note that for cases where we can't use the asymptotic form, Ref. [57] recommends using $b = \hat{\hat{b}}(s)$ when generating toys, so that we (approximately) maximize the agreement with the hypothesis.

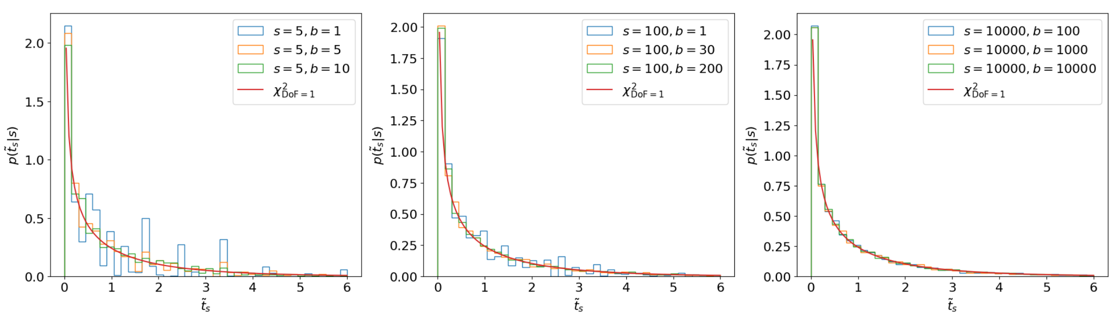


Figure 8.6. Asymptotic form of $p(\tilde{t}_s|s)$.

p-values and significance

Now that we know the distribution of the test statistic $p(\tilde{t}_s|H_s) \equiv p(\tilde{t}_s|s)$, we can finally test H_s with our experiment. We just need to calculate the "observed" test

[1]One can find the derivation in the reference therein; essentially, like with most things in physics, this follows from Taylor expanding around the minimum...

statistic $\tilde{t}_s^{\text{obs}}$ from our observations, and compare it to the $p(\tilde{t}_s|s)$.

Example 8.2.1. Let's say we're testing the hypothesis of $s = 10$ signal events in our model and we observe $n = 20, m = 5$ events. We can map this observation to our test statistic $\tilde{t}_s^{\text{obs}}(s = 10, n_{\text{obs}} = 20, m_{\text{obs}} = 5) = 1.07$, and see where this falls in our $p(\tilde{t}_s|s)$ distribution (Figure 8.7).

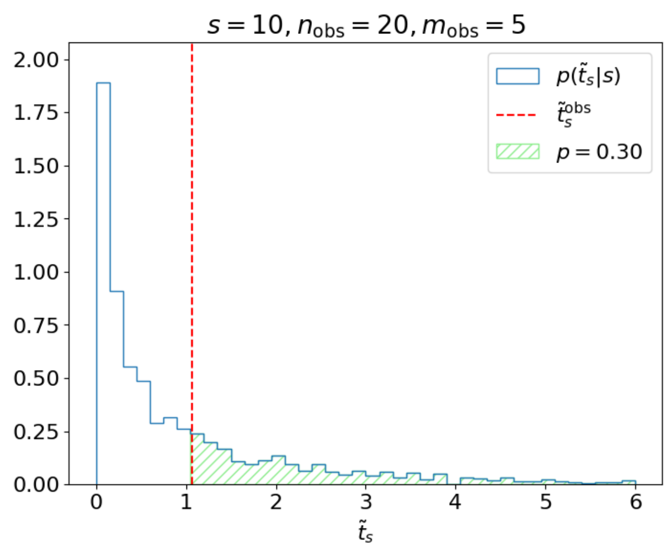


Figure 8.7. Testing H_s in Example 8.2.1.

Ultimately, we care about, given $p(\tilde{t}_s|s)$, the probability of obtaining $\tilde{t}_s^{\text{obs}}$ *or a value more inconsistent* with H_s; i.e., the green shaded region above. This is referred to as the *p-value* of the observation:

$$p_s = \int_{\tilde{t}_{\text{obs}}}^{\infty} p(\tilde{t}_s|s)\mathrm{d}\tilde{t}_s = 1 - F(\tilde{t}_{\text{obs}}|s), \tag{8.2.14}$$

which is 0.30 for this example, where

$$F(\tilde{t}_s|s) = \int_{-\infty}^{\tilde{t}_s} p(\tilde{t}'_s|s)\mathrm{d}\tilde{t}'_s \tag{8.2.15}$$

is the cumulative distribution function (CDF) of $\tilde{t}_s$. We reject the hypothesis if this p-value is less than our chosen significance level α; the idea being that if H_s were true and we repeated this measurement many times, then the probability of a false-positive (p-value $\leq \alpha$) is exactly α, as we intended.

The p-value is typically converted into a *significance* (Z), which is the corresponding number of standard deviations away from the mean in a Gaussian distribution:

$$Z = \Phi^{-1}(1-p), \tag{8.2.16}$$

where Φ is the CDF of the standard Gaussian. This is more easily illustrated in Figure 8.8, where φ is the standard Gaussian distribution:

The significance in Example 8.2.1 is, therefore, $\Phi^{-1}(1-0.30) = 0.53$. We sometimes say that our measurement is (in)consistent or (in)compatible with H at the 0.53σ level, or within 1σ, etc.

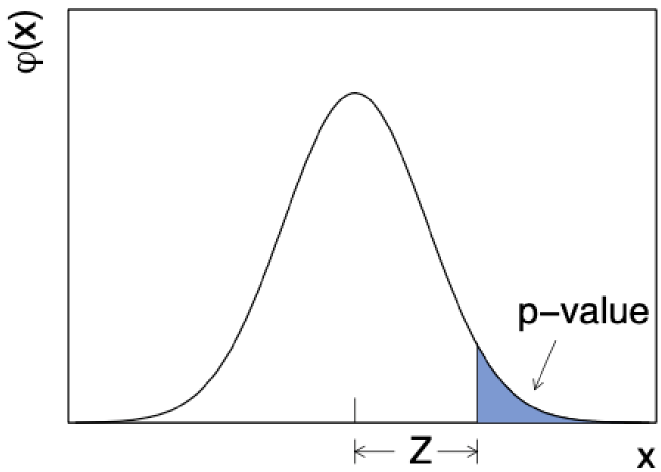


Figure 8.8. Relationship between significance Z and the p-value, reproduced from Ref. [57].

Signal discovery

So far, we have been testing the signal hypothesis, but usually when searching for a particle, we instead test the "background-only" hypothesis H_0 and decide whether or not to reject it. This means we want $\tilde{t}_0^{\text{obs}}$ and $p(\tilde{t}_0|0)$ (Figure 8.9).[2]

We could say for this experiment, therefore, that we exclude the background-only hypothesis at the "3 sigma" level. However, for an actual search for a new particle at the LHC, this is insufficient to claim a discovery, as the probability of a false positive at 3σ, $1/1000$, is too high. The standard is instead set at 5σ for discovering new signals, corresponding to the 3×10^{-7} significance level quoted earlier, as we really don't want to be making a mistake if we're claiming to have discovered a new particle! 3σ, 4σ, and 5σ are commonly

[2]Ref. [291] refers to the special case of the test statistic $\tilde{t}_s$ for $s = 0$ as q_0.

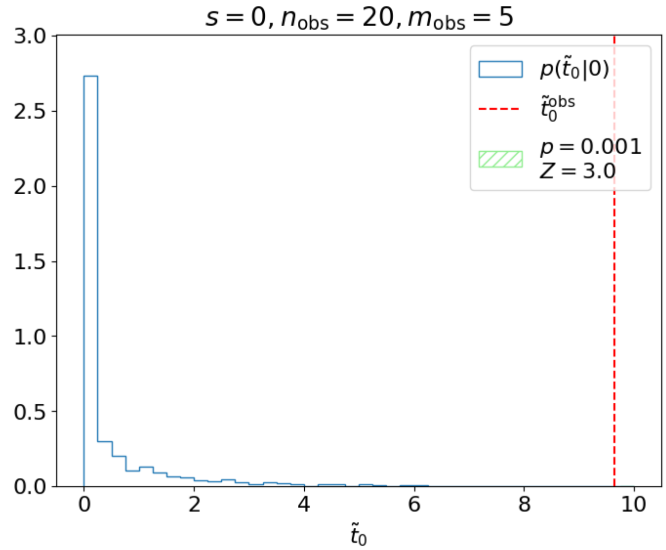


Figure 8.9. Testing the background-only hypothesis in Example 8.2.1.

referred to as "evidence", "observation", and "discovery", respectively, of the signals we're searching for.

In summary, the framework for hypothesis testing comprises:

1. Defining a test statistic t to map data $\mathbf{x}$ (in our example, $\mathbf{x} = (n, m)$) to a single number.
2. Deriving the distribution of t under the hypothesis being tested $p(t|H)$ by sampling from "toy" datasets assuming H.
3. Quantifying the compatibility of the observed data $\mathbf{x}_{\text{obs}}$ with H with the p-value or significance Z of t_{obs} relative to $p(t|H)$.

This p-value / significance is what we then use to decide whether or not to exclude H. A particularly important special case of this, as discussed above, is testing the background-only hypothesis when trying to discover a signal.

8.2.3 Confidence intervals and limits

Confidence intervals using the Neyman construction

Next, we discuss going beyond hypothesis testing to setting intervals and limits for parameters of interest. The machinery from Section 8.2.2 can be extended straightforwardly to extracting "confidence intervals" for our parameters of interest (POIs): a range of values of the POIs that are allowed, based on the experiment, at a certain "confidence level" (CL), e.g. 68% or 95%. Very similar to the idea of the significance level, the CL is defined such that if we were to repeat the experiment many times, a 95%-confidence-interval must contain, or *cover*, the true value of the parameter 95% of the time.

This can be ensured for any given CL by solving Eq. 8.2.14 for a p-value of $1 - \mathrm{CL}$:

$$p = 1 - \mathrm{CL} = \int_{\tilde{t}_s^{\mathrm{obs}}}^{\infty} p(\tilde{t}_s | s_\pm) \mathrm{d}\tilde{t}_s, \qquad (8.2.17)$$

where s_- and s_+ are the lower and upper limits on s, respectively.

This can be solved by scanning s and finding the values of s for which the RHS $= 1 - \mathrm{CL}$, as demonstrated in Figure 8.10 for the experiment in Example 8.2.1

($n_{\text{obs}} = 20, m_{\text{obs}} = 5$). This procedure of inverting the hypothesis test by scanning along the values of the POIs is called the "Neyman construction".

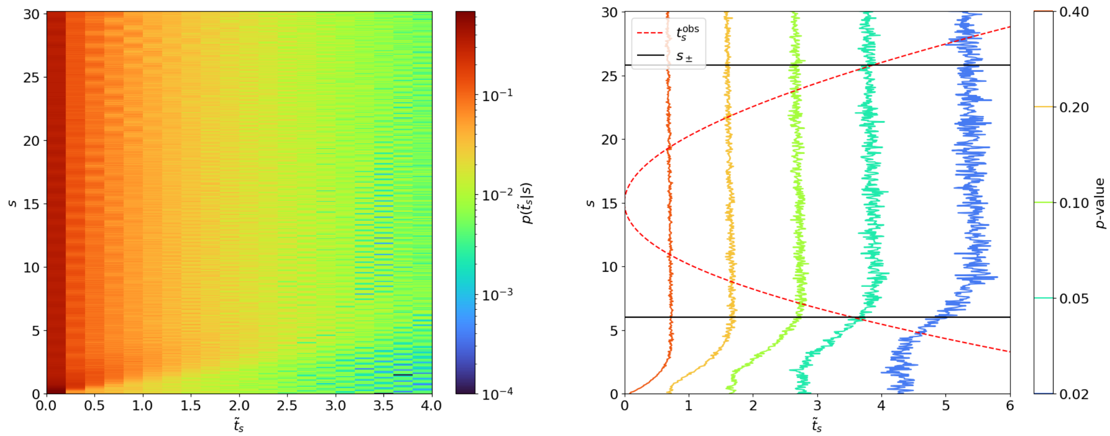


Figure 8.10. Demonstration of the Neyman construction for a 95% confidence interval for the experiment in Example 8.2.1 ($n_{\text{obs}} = 20, m_{\text{obs}} = 5$). Left: Scanning $p(\tilde{t}_s|s)$ using 10,000 toys each for different values of s. Right: Converting this to a contour plot of the p-values for different $\tilde{t}_s$'s as a function of s, with the observed t_s^{obs} in red. The points at which t_s^{obs} intersects with the p-value = 0.05 contour are marked in black and signify the limits of the 95% confidence interval for s - in this case, [6.0, 25.8].

One subtlety to remember is that, in principle, we should also be scanning over the nuisance parameters (b) when estimating the p-values. However, this would be very computationally expensive so in practice, we continue to use $b = \hat{\hat{b}}(s)$, to always (approximately) maximize the agreement with the hypothesis. Ref. [57] calls this trick "profile construction".

Upper limits

Typically if a search does not have enough sensitivity to directly observe a new signal, we instead quote an upper limit on the signal strength. This is similar in practice to the Neyman construction for confidence intervals, solving Eq. 8.2.17 only for the upper boundary. However, an important difference is that when setting upper limits, we have to modify the test statistic so that a best-fit signal strength *greater* than the expected signal ($\hat{s} > s$) does not lower the compatibility with H_s:

$$\tilde{q}(s) = \begin{cases} \tilde{t}(s), & \hat{s} < s. \\ 0, & \hat{s} \geq s. \end{cases} = \begin{cases} -2\ln\tilde{\lambda}(s), & \hat{s} < s. \\ 0, & \hat{s} \geq s. \end{cases} = \begin{cases} -2\ln\frac{L(s,\hat{\hat{b}}(s))}{L(0,\hat{\hat{b}}(0))}, & \hat{s} < 0. \\ -2\ln\frac{L(s,\hat{\hat{b}}(s))}{L(\hat{s},\hat{b})}, & 0 \leq \hat{s} < s. \\ 0, & \hat{s} \geq s. \end{cases} \tag{8.2.18}$$

The upper limit test statistic $\tilde{q}(s)$ is set to 0 for $\hat{s} > s$ so that this situation does not contribute to the p-value integral in Eq. 8.2.14. Figure 8.11 demonstrates this, and the difference between $\tilde{t}_s$ and $\tilde{q}_s$, for different sample observations.

Note that (as one may expect from Figure 8.11) the distribution $p(\tilde{q}_s|s)$ no longer behaves like a standard χ^2 but, instead, as a "half-χ^2". This is essentially a χ^2 plus a delta function at 0 (since, under the signal hypothesis, on average there will be an over-fluctuation half the time, for which $\tilde{q}_s = 0$), as shown in Figure 8.12.

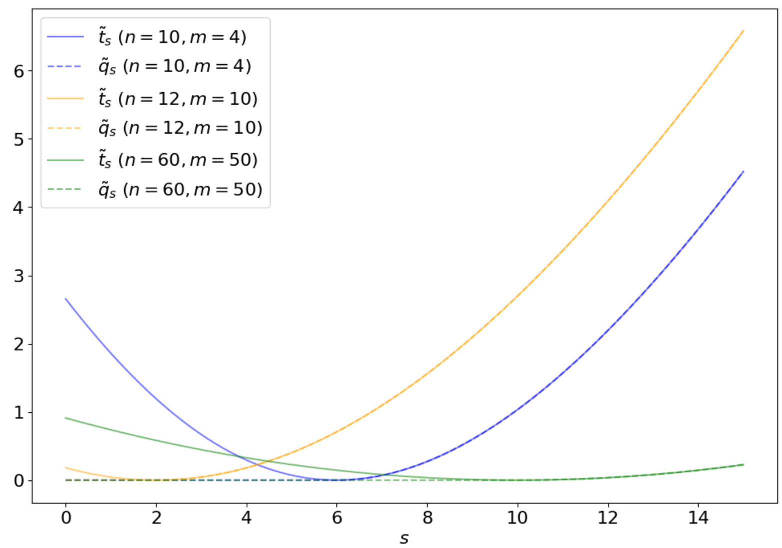


Figure 8.11. Comparing $\tilde{t}_s$ and $\tilde{q}_s$.

We can now revisit Example 8.2.1 to set an upper limit on s rather than a confidence interval (Figure 8.13). $p(\tilde{q}_s|s)$ is shifted to the left with respect to $p(\tilde{t}_s|s)$; hence, the upper limit of 24 is slightly lower than the upper bound of the 95% confidence interval we derived using $\tilde{t}_s$.

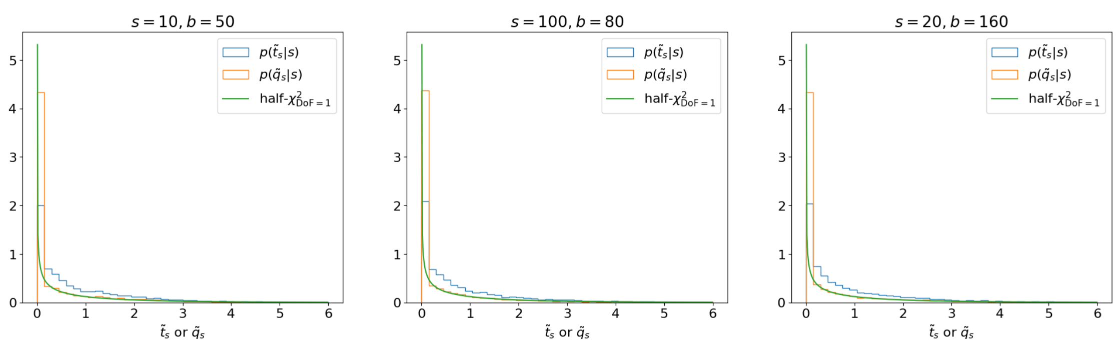


Figure 8.12. Comparing $p(\tilde{t}_s|s)$ and $p(\tilde{q}_s|s)$. $p(\tilde{q}_s|s)$ asymptotically follows a half-χ^2 distribution (green).

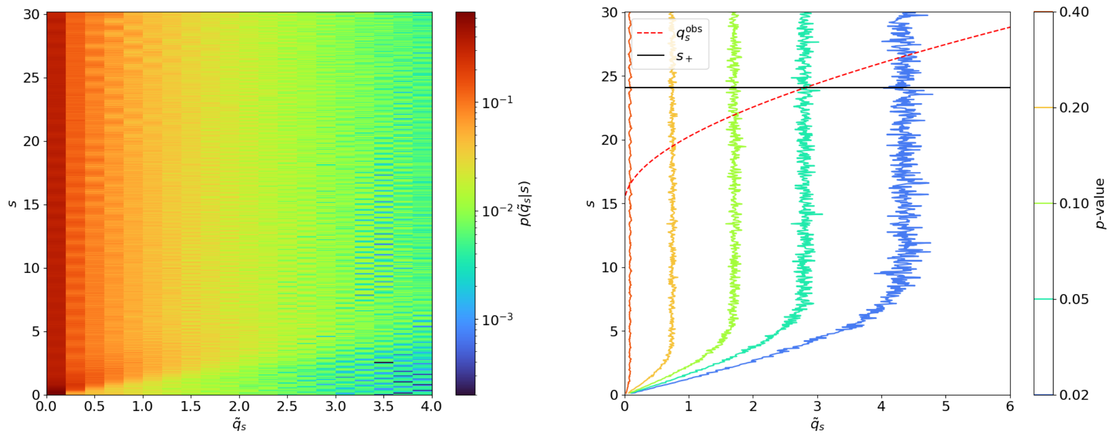


Figure 8.13. Extending the Neyman construction to an upper limit on s. Left: Scanning the upper limit test statistic distribution $p(\tilde{q}_s|s)$ using 10,000 toys each for different values of s. Right: Converting this to a contour plot of the p-values for different $\tilde{q}_s$'s as a function of s, with the observed q_s^{obs} in red. The point at which q_s^{obs} intersects with the p-value = 0.05 contour is marked in black and signifies the upper limit at 95% CL.

The CL$_s$ criterion

We now introduce two conventions related to hypothesis testing and searches in particle physics. Firstly (the simple one), the POI s is usually re-parametrized as $s \to \mu \cdot s$, where μ is now considered the POI, referred to as the "signal strength", and s is a fixed value representing the number of signal events we expect to see for the nominal signal strength μ of 1. For the example in Figure 8.13, if we expect $s = 10$ signal events, then we would quote the upper limit as $24/s = 2.4$ on μ at 95% CL.

The second, important, convention is that we use a slightly different criterion for confidence intervals, called "CL$_s$". This is motivated by situations where we have little sensitivity to the signal we're searching for, as in the below example.

Example 8.2.2. Let's say we expect $s = 10$ and observe $n = 70$, $m = 100$. Really, what this should indicate is that our search is not at all sensitive, since our search region is completely dominated by background and, hence, we should not draw strong conclusions about the signal strength. However, if we follow the above procedure for calculating the upper limit, we get $\mu \leq 0.001$ at 95% CL.

This is an extremely aggressive limit on μ, where we're excluding the nominal $\mu = 1$ signal at a high confidence level. Given the complete lack of sensitivity to the signal, this is not a sensible result. The CL_s method solves this problem by considering both the p-value of the signal + background hypothesis H_s (referred to as p_{s+b} or just p_μ for short), *and* the p-value of the background-only hypothesis H_0 (p_b), to define a new criterion:

$$p'_\mu = \frac{p_\mu}{1 - p_b} \tag{8.2.19}$$

In cases where the signal region is completely background-dominated, the compatibility with the background-only hypothesis should be high, so $p_b \sim 1$ and, hence, p'_μ will be increased. On the other hand, for more sensitive regions, compatibility should be lower $\Rightarrow p_b \sim 0$ and $p'_\mu \sim p_\mu$.

To be explicit, here

$$p_b = \int_{-\infty}^{\tilde{t}_{\text{obs}}} p(\tilde{t}_s|0)\mathrm{d}\tilde{t}_s, \tag{8.2.20}$$

where we should note that:

1. We're looking at the distribution of $\tilde{t}_s$ — *not* $\tilde{t}_0$ — under the background-only hypothesis, since the underlying test is of H_s, not H_0; and
2. We're integrating *up to* $\tilde{t}_{\text{obs}}$, unlike for p_s, because lower $\tilde{t}$ means greater compatibility with the background-only hypothesis.

The effect of the CL_s criterion is demonstrated in Figure 8.14 for Examples 8.2.1 and 8.2.2. In the former, the background-only distribution is shifted to the right of the $s + b$ distribution. This indicates that the experiment is sensitive to μ and, indeed, we find $p'_\mu \sim p_\mu$. In Example 8.2.2, however, the search is not sensitive and, hence, the background-only and $s + b$ distributions almost completely overlap, meaning $p_b \sim 1$ and $p'_\mu >> p_\mu$.[3]

Finally, if we repeat the Neyman construction using the CL_s criterion p'_μ instead of p_μ for Example 8.2.2, we can find an upper limit of $\mu \leq 1.2$ at 95% CL, which is indeed a looser, more conservative, upper limit. The upper limit for Example 8.2.1 remains unchanged at $\mu \leq 2.4$, as we would expect.

[3]Note that, unlike $p(\tilde{t}_s|s)$, $p(\tilde{t}_s|0)$ *doesn't* follow a simple χ^2; asymptotically, it is closer to a *noncentral* χ^2, as will be discussed in Section 8.3.2.

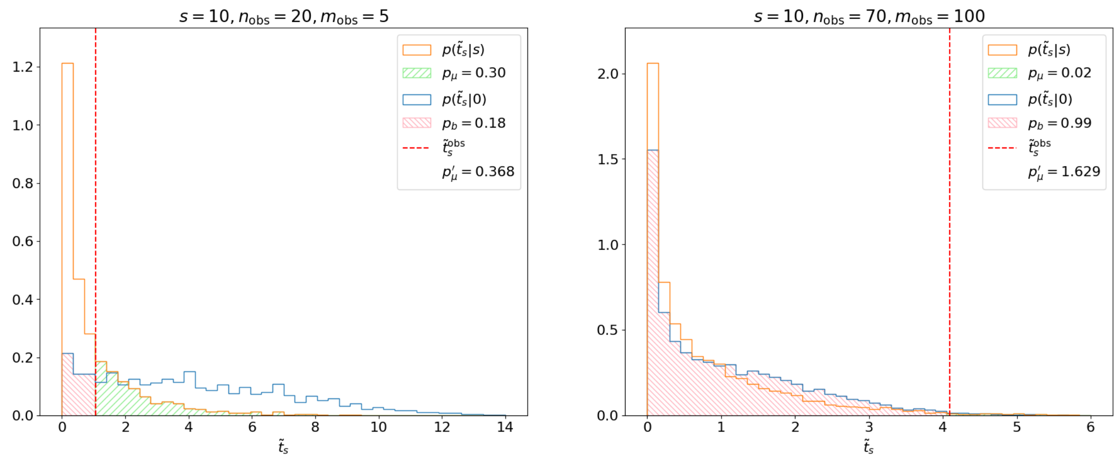


Figure 8.14. Demonstration of CL_s criterion for Examples 8.2.1 (left) and 8.2.2 (right).

8.2.4 Expected significances and limits

Expected significance

The focus so far has been only on evaluating the *results* of experiments. However, it is equally important to characterize the expected sensitivity of the experiment *before* running it (or before looking at the data).

Example 8.2.3. Concretely, we continue with the simple one-bin counting experiment (Section 8.2.1). Let's say we expect $b = 10$ background events and — at the nominal signal strength $\mu = 1$ — $s = 10$ signal events. How do we tell if this experiment is at all useful for discovering this signal, i.e., does it have any sensitivity to the signal?

One way is to calculate the significance with which we expect to exclude the background-only hypothesis if the signal were, in fact, to exist. Practically, this means we are testing H_0 and, hence, need $p(\tilde{t}_0|\mu = 0)$ as before. However, now we also need the distribution of the test statistic $\tilde{t}_0$ under the *background + signal* hypothesis $p(\tilde{t}_0|\mu = 1)$. Then, by calculating the significance for each sampled $\tilde{t}_0$ under $H_{\mu=1}$, we can estimate the distribution of expected significances. This is illustrated for Example 8.2.3 in Figure 8.15.

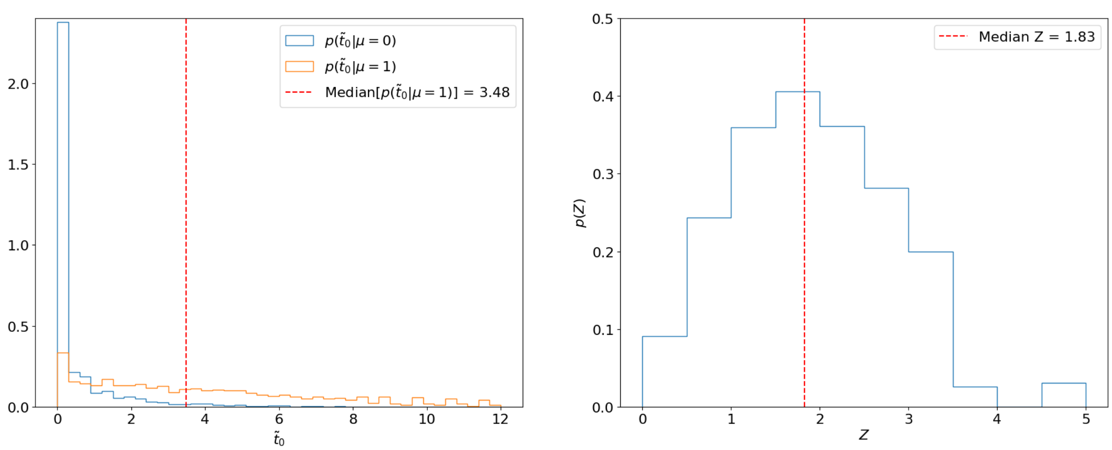


Figure 8.15. Left: Distributions of $\tilde{t}_0$ under the background-only and background + signal hypotheses using 30,000 toys each. The median of the latter is marked in red. Right: Distribution of the significances (with respect to the background-only hypothesis) of each sampled $\tilde{t}_0$ under the signal hypothesis.

Importantly, we usually quote the **median** of this distribution as the expected significance, since the median is "invariant" to monotonic transformations (i.e., the median p-value will always correspond to the median Z as well, whereas the mean p-value will not correspond to the mean Z). Similarly, we quote the

16%/84% and 2%/98% quantiles as the $\pm 1\sigma$ and $\pm 2\sigma$, respectively, expected significances. These quantiles correspond to the cumulative probabilities for a standard Gaussian (Figure 8.16). For Example 8.2.3, we thus find the median expected significance to be 1.83.

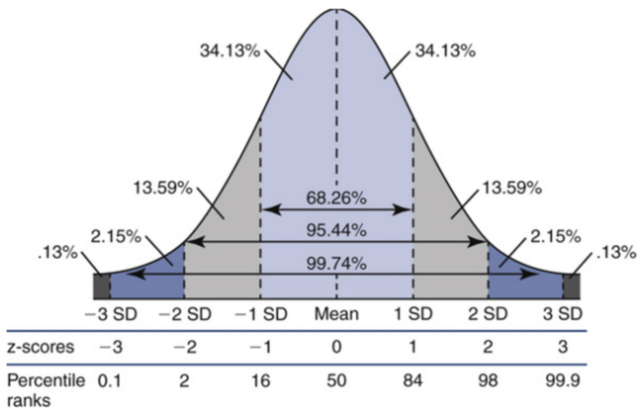


Figure 8.16. Gaussian quantiles, reproduced from Ref. [58].

Note that instead of converting each sampled $\tilde{t}$ under $H_{\mu=1}$ into a significance and finding the median of that distribution, as in Figure 8.15 (right), we can take advantage of the invariance of the median and directly use the significance of the median $\tilde{t}$ under $H_{\mu=1}$ (Figure 8.15, left). We will do this below for the expected limit.

Expected limits

The other figure of merit we care about in searches is the upper exclusion limit set on the signal strength. To derive the *expected* limit, we do the opposite of the above and ask, if the signal were not to exist, what value of μ would we expect to exclude at the 95% CL.[4]

This means we need:

1. The distribution $p(\tilde{q}_\mu|\mu)$ to solve for μ^+ in Eq. 8.2.17 and be able to do the upper limit calculation (as in Section 8.2.3);
2. $p(\tilde{q}_\mu|0)$ to get the median (and other quantiles') expected $\tilde{q}_\mu^{\text{obs}}$ for different signal strengths under the background-only hypothesis; and, furthermore,
3. To scan over the different signal strengths to find the μ that results in a median p-value of 0.05 — or, rather, p'_μ-value (Eq. 8.2.19), since we're using the CL_s method for upper limits (Section 8.2.3).

First, let's look at the first two steps for just the $\mu = 1$ signal strength in Example 8.2.3. These steps are similar to, and essentially an inversion of, the procedure for the expected significance: we're now finding the $p'_{\mu=1}$-value with respect to the signal + background hypothesis, for the median $\tilde{q}_\mu$ sampled under the background-only hypothesis. This is demonstrated in Figure 8.17.

The key difference with respect to calculating the expected significance is

[4] 95% is the standard CL for upper limits in HEP.

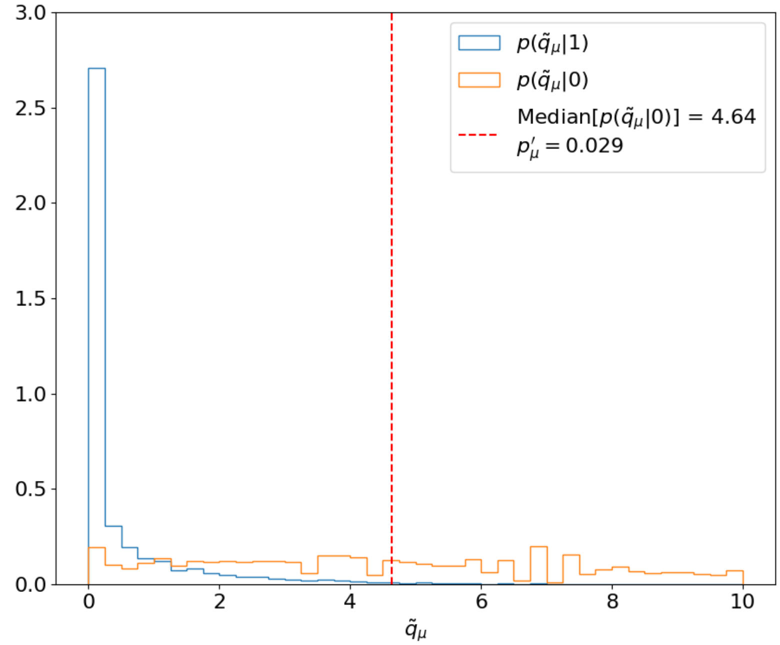


Figure 8.17. Calculating the median expected $p'_{\mu=1}$-value with respect to the signal + background hypothesis, for test statistics $\tilde{q}_\mu$ sampled under the background-only hypothesis. $p(\tilde{q}_\mu|1)$ and $p(\tilde{q}_\mu|0)$ are estimated using 30,000 toys each. Then, the median $p(\tilde{q}_\mu|0)$ (red) is used to calculate the p'_μ-value following the CL_s criterion.

step 3, in which this procedure has to be repeated for a range of signal strengths to find the value that gives a median (and $\pm 1\sigma, \pm 2\sigma$ quantile-) p'_μ of 0.05. This is thus the minimum value of μ that we expect to be able to exclude at 95% CL, as shown in Figure 8.18.

Thus, we have our expected limits. The right plot of Figure 8.18 is colloquially known as a "Brazil-band plot", and is the standard way of representing limits. For example, Figure 8.19 is the corresponding plot by ATLAS for the Higgs discovery (scanning over the Higgs mass).

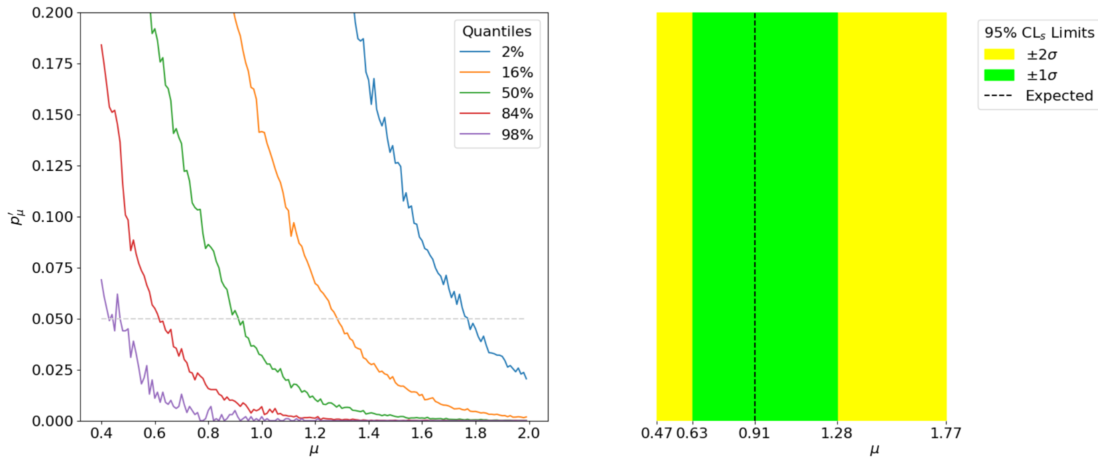


Figure 8.18. Left: The expected median and $\pm 1\sigma$, $\pm 2\sigma$ quantiles of p'_μ for different μ's. The intersection of these with $p'_\mu = 0.05$ (gray) corresponds to the expected exclusion limits. Right: The median and $\pm 1\sigma$, $\pm 2\sigma$ expected limits at 95% CL$_s$ on μ.

8.3 Asymptotic formulae

8.3.1 Asymptotic form of the MLE

So far, we have discussed how to extract meaningful statistical results from HEP experiments by making extensive use of pseudodata / toy experiments to estimate the sampling distributions of profile-likelihood-ratio-based test statistics. While this worked nicely for our simple counting experiment, generating a sufficiently large number of toys can quickly become computationally intractable for the more complex searches (and statistical combinations of searches) that are increasingly prevalent at the LHC, containing at times up to thousands of bins and nuisance

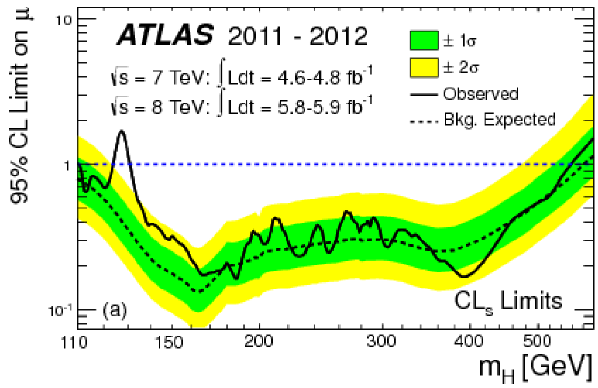


Figure 8.19. Expected and observed 95% CL$_s$ upper limits for the SM Higgs by ATLAS in 2012, for different hypothetical Higgs masses [59].

parameters. This and the following section discuss a way to approximate these sampling distributions without the need for pseudodata. This was introduced in the famous “CCGV” paper [291] in 2011 and has since become the de-facto procedure at the LHC.

As hinted at previously, such as in Figures 8.6 and 8.12, the distributions $p(\tilde{t}_\mu|\mu')$ and $p(\tilde{q}_\mu|\mu')$ (where, in general, $\mu' \neq \mu$) have similar forms regardless of the nuisance parameters (or sometimes even the POIs). This is not a coincidence: we will now derive their “asymptotic” — i.e., in the large sample limit — forms, starting first with the asymptotic form of the maximum likelihood estimator (MLE).

It is important to remember that the MLE $\hat{\mu}$ of μ is a random variable with its own probability distribution. We can estimate it as always by sampling toys, shown in Figure 8.20 for our counting experiment (Eq. 8.2.3). One can observe that $p(\hat{\mu})$ follows a Gaussian distribution as the number of events N increases, and indeed this becomes clear if we try to fit one to the histograms (Figure 8.21). We will now show this to be true generally, deriving the analytic distribution in Sections 8.3.1—8.3.1, and discussing the results and the important concept of the *Asimov* dataset for numerical estimation in Sections 8.3.1 and 8.3.1, respectively.

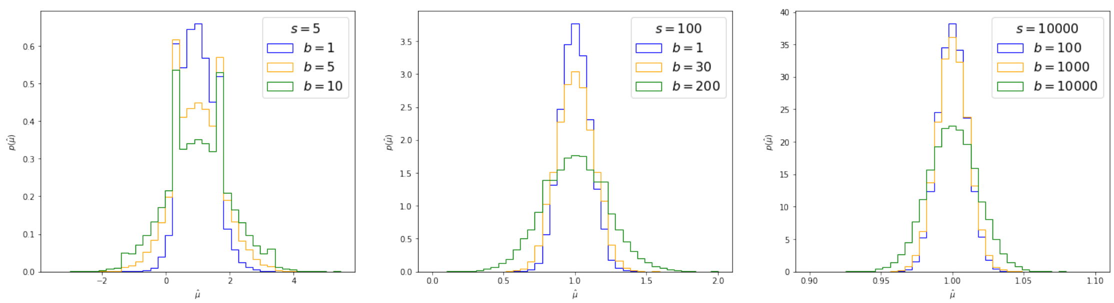


Figure 8.20. Distribution of the MLE of μ for different s and b produced using 30,000 toy experiments each. (Note the x-axis range is becoming narrower from the left-most to the right-most plot.)

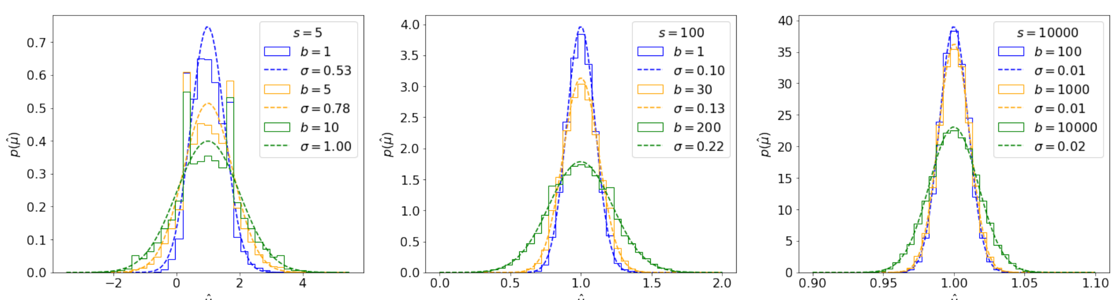


Figure 8.21. Gaussian fits to distributions of $\hat{\mu}$ for different s and b from Figure 8.20.

Statistics background

We first provide a lightning review of some necessary statistics concepts and results.

Definition 8.3.1. Let the negative log-likelihood (NLL) $-\ln L(\mu) \equiv -l(\mu)$. The derivative of the NLL $-l'(\mu)$ is called the **score** $s(\mu)$. It has a number of useful properties: [5]

1. Its expectation value at μ' $\mathbb{E}_{\mu=\mu'}[s(\mu')] = 0$.
2. Its variance $\mathrm{Var}[s(\mu)] = -\mathbb{E}[l''(\mu)]$.

Note that the expectation value here means an average over observations which are distributed according to a particular μ, which here we're calling the "true" μ: μ'.

Definition 8.3.2. $-\mathbb{E}[l''(\mu)] \equiv \mathcal{I}(\mu)$ is called the **Fisher information**. It quantifies the information our data contains about μ and importantly, as we'll see, it (approximately) represents the inverse of the variance of $\hat{\mu}$. More generally, for multiple parameters,

$$\mathcal{I}_{ij}(\mu) = -\mathbb{E}[\frac{\partial^2 l}{\partial \mu_i \partial \mu_j}] \tag{8.3.1}$$

is the Fisher information matrix. It is also commonly called the **covariance matrix**.

Theorem 8.3.1. Putting this together, by the central limit theorem [296], this

[5] See derivations in e.g. Ref. [295].

means $p(s(\mu'))$ follows a normal distribution with mean 0 and variance $\mathcal{I}(\mu')$, up to terms of order $\mathcal{O}(\frac{1}{\sqrt{N}})$:

$$s(\mu') \xrightarrow{\sqrt{N}>>1} \mathcal{N}(0, \sqrt{\mathcal{I}(\mu')}), \tag{8.3.2}$$

where N represents the data sample size.

The Fisher information

For our simple counting experiment, the Fisher information matrix $\mathcal{I}(\mu, b)$ can be found by taking second derivatives of the NLL (Eq. 8.2.5). The $\mathcal{I}_{\mu\mu}$ term, for example, is:

$$\mathcal{I}_{\mu\mu}(\mu, b) = -\mathbb{E}[\partial^{\mu}\partial^{\mu} l(\mu, b)] = \mathbb{E}\left[n \cdot \frac{s^2}{(\mu s + b)^2}\right] = \mathbb{E}[n] \cdot \frac{s^2}{(\mu s + b)^2} = \frac{(\mu' s + b')s^2}{(\mu s + b)^2}. \tag{8.3.3}$$

In the last step we use the fact that $\mathbb{E}[n]$ under true $\mu = \mu'$, $b = b'$, is $\mu' s + b'$. For the remainder of this section, $\mathcal{I}(\mu, b)$ will always be evaluated at the true values of the parameters,[6] so this can be simplified to $\mathcal{I}_{\mu\mu}(\mu', b') = \frac{s^2}{\mu' s + b'}$. This is plotted in Figure 8.22, where we can the Fisher information captures the fact that as b increases, we lose sensitivity to — or *information* about — μ.

For completeness (and since we'll need it below), the full Fisher informa-

[6]The reason for this is discussed shortly in Section 8.3.1.

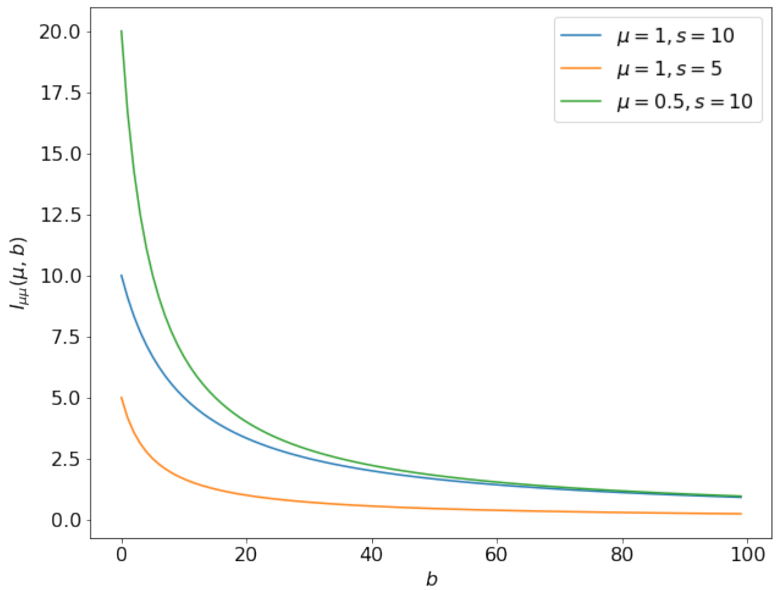


Figure 8.22. The Fisher information $\mathcal{I}_{\mu\mu}(\mu, b)$ for different μ and s, as a function of the expected background b.

tion matrix for our problem, repeating the steps in Eq. 8.3.3, is:

$$\mathcal{I}(\mu', b') = \begin{pmatrix} \mathcal{I}_{\mu\mu} & \mathcal{I}_{\mu b} \\ \mathcal{I}_{b\mu} & \mathcal{I}_{bb} \end{pmatrix} (\mu', b') = \begin{pmatrix} \frac{s^2}{\mu' s + b'} & \frac{s}{\mu' s + b'} \\ \frac{s}{\mu' s + b'} & \frac{1}{\mu' s + b'} + \frac{1}{b'} \end{pmatrix} \tag{8.3.4}$$

Derivation

We now have enough background to derive the asymptotic form of the MLE. We do this for the 1D case by Taylor-expanding the score of $\hat{\mu}$, $l'(\hat{\mu})$ - which we know to be = 0 - around μ':

$$l'(\hat{\mu}) = l'(\mu') + l''(\mu')(\hat{\mu} - \mu') + O((\hat{\mu} - \mu')^2) = 0 \tag{8.3.5}$$

$$\Rightarrow \hat{\mu} - \mu' \simeq -\frac{l'(\mu')}{l''(\mu')} \xrightarrow{\sqrt{N}>>1} \frac{1}{\mathcal{I}(\mu')} N(0, \sqrt{\mathcal{I}(\mu')}) = N\left(0, \frac{1}{\sqrt{\mathcal{I}(\mu')}}\right), \tag{8.3.6}$$

where we plugged in the distribution of $l'(\mu')$ from Eq. 8.3.2, claimed $l''(\mu')$ asymptotically equals its expectation value $\mathbb{E}[l''(\mu')] = \mathcal{I}(\mu')$ by the law of large numbers [297], and are ignoring the $O((\hat{\mu} - \mu')^2)$ term.[7]

For multiple parameters, $\mathcal{I}$ is a matrix so the variance generalized to the matrix inverse:

$$\hat{\mu} - \mu' \simeq N(0, \sqrt{\mathcal{I}^{-1}_{\mu\mu}(\mu', b')}), \tag{8.3.7}$$

Result

Thus, we see that $\hat{\mu}$ asymptotically follows a normal distribution around the true μ value, μ', with a variance $\sigma^2_{\hat{\mu}} = \mathcal{I}^{-1}_{\mu\mu}(\mu', b')$, up to $O(1/\sqrt{N})$ terms. Intuitively, from the definition of the Fisher information $\mathcal{I}$, we can interpret this as saying that the more information we have about μ from the data, the lower the variance should be on $\hat{\mu}$.

[7] For a more rigorous derivation, see e.g. Ref. [298].

Continuing with our counting experiment from Section 8.2.1, inverting $\mathcal{I}$ from Eq. 8.3.4 gives us

$$\sigma_{\hat{\mu}} = \sqrt{\mathcal{I}^{-1}_{\mu\mu}(\mu', b')} = \frac{\sqrt{\mu' s + 2b'}}{s}. \tag{8.3.8}$$

Note that, as we might expect, this scales as $\sim \sqrt{b}$, which is the uncertainty of our Poisson nuisance parameter b — showing mathematically why we want to keep uncertainties on nuisance parameters as low as possible. This is compared to the toy-based distributions from Section 8.3.1 in Figure 8.23 this time varying the true signal strength μ' as well, where we can observe that this matches very well for large s, b, while for small values there are some discrete differences.

We can also check the total per-bin errors between the asymptotic form and the toy-based distributions directly, as shown in Figure 8.24 (for $\mu' = 1$ only). Indeed, this confirms that the error scales as $\sim \frac{1}{\sqrt{s}}$ and $\sim \frac{1}{\sqrt{b}}$, as claimed above.

Numerical estimation and the Asimov dataset

In this section, because of the simplicity of our data model, we were able to derive the Fisher information $\mathcal{I}$ and, hence, the asymptotic form of $\hat{\mu}$ analytically. In general, this is not possible and we typically have to minimize l, find its second derivatives, and solve Eq. 8.3.3 etc. *numerically* instead.

However, when calculating the Fisher information, how do we deal with

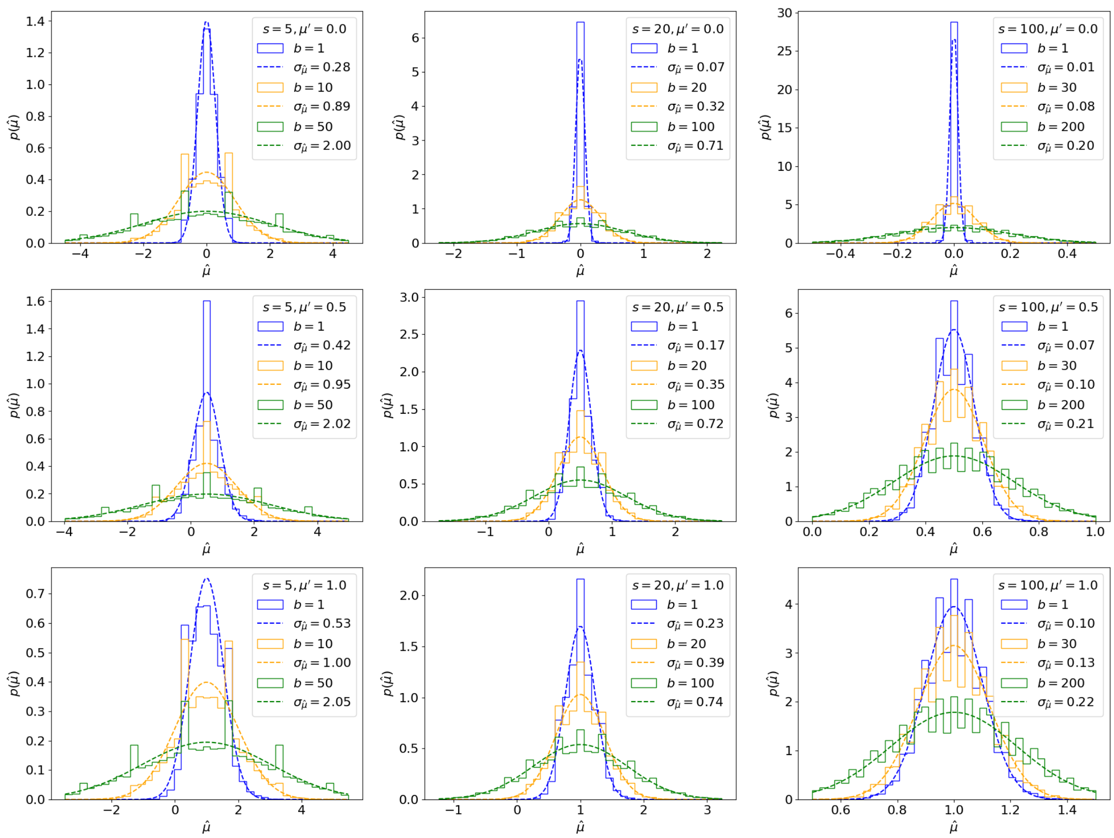


Figure 8.23. Asymptotic (dotted lines) and toy-based (solid lines) distributions, using 30,000 toys each, of the MLE of μ for different s, b, and true signal strengths μ'.

the expectation value over the observed data (n, m in our case)? Naively, this would require averaging over a bunch of generated toy n, m values again, which defeats the purpose of using the asymptotic form of $\hat{\mu}$!

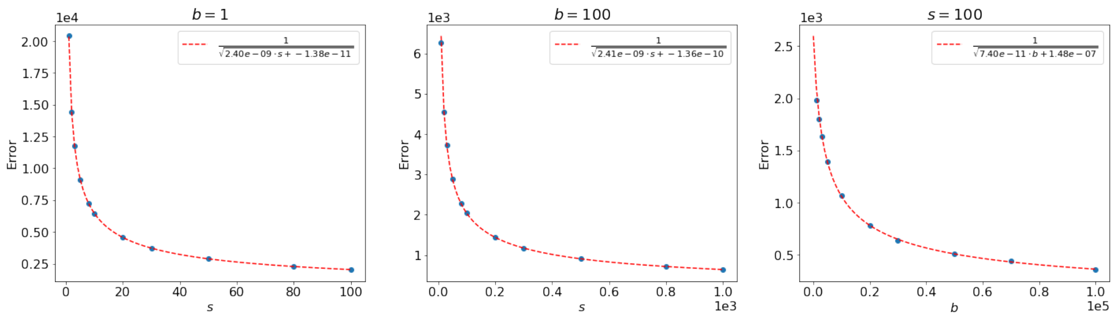


Figure 8.24. Error between the sampled toy distributions, using 50,000 toys each, and the asymptotic distributions of the MLE of μ for different s and b (blue), with $1/\sqrt{N}$ fits in red.

Instead, we can switch the order of operations in Eq. 8.3.3,[8] rewriting it as:

$$\mathcal{I}_{ij}(\mu, b) = -\mathbb{E}[\partial^i \partial^j l(\mu, b; n, m)] = -\partial^i \partial^j \mathbb{E}[l(\mu, b; n, m)] = -\partial^i \partial^j l(\mu, b; \mathbb{E}[n], \mathbb{E}[m]). \tag{8.3.9}$$

Importantly, this says we can find $\mathcal{I}$ by simply evaluating the likelihood for a dataset of observations equal to their expectation values under μ' instead of averaging over the distribution of observations and *then* getting its second derivatives.

Definition 8.3.3. Such a dataset is called the **Asimov** dataset, and $L(\mu; \mathbb{E}[n], \mathbb{E}[m]) \equiv L_A$ is referred to as the "Asimov likelihood".[9]

[8] We are able to do this because, as we saw above, the score is linear in n for Poisson likelihoods.

[9] The *Asimov* dataset is named after Isaac Asimov, the popular science fiction author, whose book *Franchise* is about a supercomputer choosing a single person as the sole voter in the U.S. elections, because they can represent the entire population.

8.3.2 Asymptotic form of the profile likelihood ratio

We can now proceed to derive the asymptotic form of the sampling distribution $p(t_\mu|\mu')$ of the profile likelihood ratio test statistic t_μ, under a "true" signal strength of μ'. This asymptotic form is extremely useful for simplifying the computation of (expected) significances, limits, and intervals; indeed, standard procedure at the LHC is to use it in lieu of toy-based, empirical distributions for $p(t_\mu|\mu')$.

Asymptotic form of the profile likelihood ratio

We start with deriving the asymptotic form of the profile likelihood ratio test statistic t_μ (Eq. 8.2.7) by following a similar procedure to Section 8.3.1 — and using the results therein — of Taylor expanding around its minimum at $\hat{\mu}$:[10]

$$
\begin{aligned}
t_\mu &= -2\ln\lambda(\mu) && (8.3.10)\\
&= -2l(\mu, \hat{\hat{b}}(\mu)) + 2l(\hat{\mu}, \hat{b}) && (8.3.11)\\
&\simeq \underbrace{-2l(\hat{\mu}, \hat{\hat{b}}(\hat{\mu})) + 2l(\hat{\mu}, \hat{b})}_{\hat{\hat{b}}(\hat{\mu})=\hat{b} \text{ so this is } 0} - \underbrace{2l'(\hat{\mu}, \hat{\hat{b}}(\hat{\mu}))(\mu-\hat{\mu})}_{l'(\hat{\mu},\hat{b})=0} - 2l''(\hat{\mu}, \hat{\hat{b}}(\hat{\mu}))\cdot\frac{(\mu-\hat{\mu})^2}{2} && (8.3.12)\\
&= -l''(\hat{\mu}, \hat{b})\cdot(\mu-\hat{\mu})^2 && (8.3.13)\\
&= \underbrace{-\mathbb{E}[l''(\hat{\mu}, \hat{b})]}_{\text{By law of large numbers}} \cdot(\mu-\hat{\mu})^2 && (8.3.14)\\
&= \underbrace{-\mathbb{E}[l''(\mu', b')]}_{\text{Since bias of MLEs} \sim 0} \cdot(\mu-\hat{\mu})^2 && (8.3.15)\\
&= \underbrace{\mathcal{I}_{\mu\mu}(\mu', b')}_{\text{From definition of Fisher information}} \cdot(\mu-\hat{\mu})^2 && (8.3.16)\\
\Rightarrow\ & \underbrace{\boxed{t_\mu \simeq \frac{(\mu-\hat{\mu})^2}{\sigma_{\hat{\mu}}^2}}}_{\text{Using } \sigma_{\hat{\mu}} \simeq \sqrt{\mathcal{I}^{-1}_{\mu\mu}(\mu',b')}} + \mathcal{O}((\mu-\hat{\mu})^3) + \mathcal{O}(\frac{1}{\sqrt{N}}). && (8.3.17)
\end{aligned}
$$

Here, just like in Eq. 8.3.6, we use the law of large numbers in Line 8.3.14 and take $l''(\hat{\mu}, \hat{b})$ to asymptotically equal its expectation value under the true parameter values μ', b': $l''(\hat{\mu}, \hat{b}) \xrightarrow{\sqrt{N}>>1} \mathbb{E}[l''(\hat{\mu}, \hat{b})]$. We then in Line 8.3.15 also use the fact that MLEs are generally unbiased estimators of the true parameter

[10] Note: this is not a rigorous derivation; it's just a way to motivate the final result, which is taken from Ref. [291]. (If you know of a better way, let me know!)

values in the large sample limit to say $\mathbb{E}[l''(\hat{\mu}, \hat{b})] \xrightarrow{\sqrt{N}>>1} \mathbb{E}[l''(\mu', b')]$. Finally, in the last step, we use the asymptotic form of the MLE (Eq. 8.3.7).

Asymptotic form of $p(t_\mu|\mu')$

Now that we have an expression for t_μ, we can consider its sampling distribution. With a simple change of variables, the form of $p(t_\mu|\mu')$ should hopefully be evident: recognizing that μ and $\sigma^2_{\hat{\mu}}$ are simply constants, while $\hat{\mu}$ we know is distributed as a Gaussian centered around μ' with variance $\sigma^2_{\hat{\mu}}$, let's define $\gamma \equiv \frac{\mu-\hat{\mu}}{\sigma_{\hat{\mu}}}$, so that

$$t_\mu \simeq \frac{(\mu - \hat{\mu})^2}{\sigma^2_{\hat{\mu}}} = \gamma^2, \tag{8.3.18}$$

$$\gamma \sim \mathcal{N}\left(\frac{\mu - \mu'}{\sigma_{\hat{\mu}}}, 1\right). \tag{8.3.19}$$

For the special case of $\mu = \mu'$, we can see that t_μ is simply the square of a standard normal random variable, which is the definition of the well-known χ^2_k distribution with $k = 1$ degrees of freedom (DoF):

$$p(t_\mu|\mu) \sim \chi^2_1. \tag{8.3.20}$$

In the general case where μ may not $= \mu'$, t_μ is the square of random variable with unit variance but *non-zero mean*. This is distributed as the similar, but perhaps less well-known, **non-central chi-squared** $\chi_k'^2(\Lambda)$, again with 1 DoF, and with a "non-centrality parameter"

$$\Lambda = \bar{\gamma}^2 = \left(\frac{\mu - \mu'}{\sigma_{\hat{\mu}}}\right)^2, \tag{8.3.21}$$

$$p(t_\mu|\mu') \sim \chi_1'^2(\Lambda). \tag{8.3.22}$$

The "central" vs. non-central chi-squared distributions are visualized in Figure 8.25 for $k = 1$. We can see that $\chi_k'^2(\Lambda)$ simply shifts towards the right as Λ increases (at $\Lambda = 0$ it is a regular central χ^2). As $\Lambda \to \infty$, $\chi_k'^2(\Lambda)$ becomes more and more like a normal distribution with mean Λ.[11]

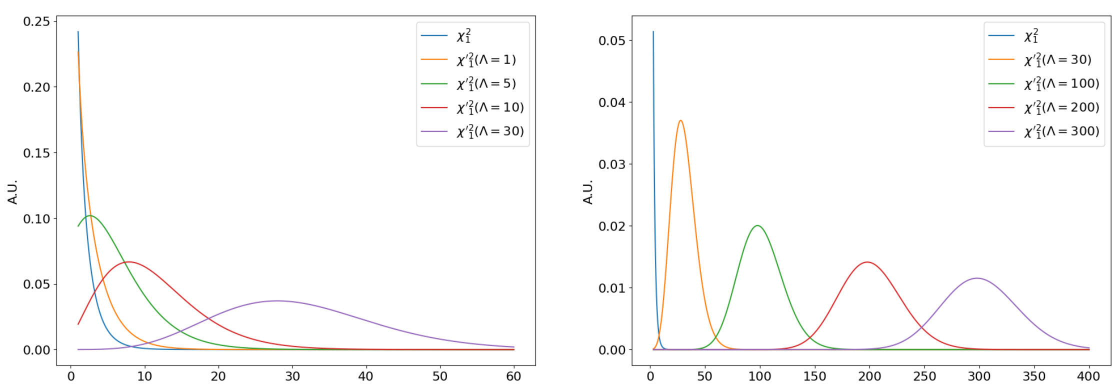


Figure 8.25. Central χ_k^2 and non-central $\chi_k'^2(\Lambda)$ distributions for Λ between $1 - 30$ (left) and $30 - 300$ (right).

By extending the derivation in Eq. 8.3.17 to multiple POIs, one can find

[11] More information can be found in e.g. Ref. [299].

the simple generalization to multiple POIs μ:

$$p(t_\mu|\mu') \sim \chi'^2_k(\Lambda), \tag{8.3.23}$$

where the DoF k are equal the number of POIs $\dim \mu$, and

$$\Lambda = (\mu - \mu')^T \cdot \tilde{\mathcal{I}}^{-1}(\mu') \cdot (\mu - \mu'), \tag{8.3.24}$$

where $\tilde{\mathcal{I}}^{-1}$ is $\mathcal{I}^{-1}$ restricted only to the components corresponding to the POIs.

Estimating $\sigma^2_{\hat{\mu}}$

The critical remaining step to understanding the asymptotic distribution of t_μ is estimating $\sigma^2_{\hat{\mu}}$ to find the non-centrality parameter Λ in Eq. 8.3.21. We now discuss two methods to do this.

Method 1: Inverting the Fisher information / covariance matrix

The first method is simply using $\sigma_{\hat{\mu}} \simeq \sqrt{\mathcal{I}^{-1}_{\mu\mu}(\mu', b')}$ as in Section 8.3.1.[12] This is shown in Figure 8.26 for our counting experiment, using the analytic form for $\sigma_{\hat{\mu}}$ from Eq. 8.3.8. We can see that this asymptotic approximation agrees well with the true distribution for some range of parameters, but can deviate significantly for others, as highlighted especially in the right plot.

[12] More generally, we'd need $\tilde{\mathcal{I}}^{-1}$ for Eq. 8.3.24.

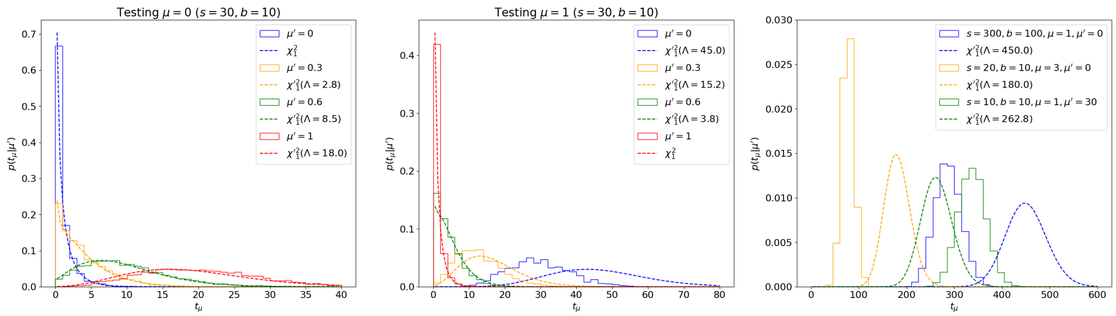


Figure 8.26. Comparing the distribution $p(t_\mu|\mu')$ (solid) with non-central $\chi_1'^2(\Lambda)$ distributions (dotted) for a range of s, b, μ, μ' values, with $\sigma_{\hat{\mu}}^2$ estimated using the inverse of the Fisher information matrix.

Interlude on Asimov dataset

While we are able to find the analytic form for $\sqrt{\mathcal{I}_{\mu\mu}^{-1}(\mu', b')}$ easily for our simple counting experiment, in general it has to be calculated numerically. As introduced in Section 8.3.1, to handle the expectation value under μ', b' in Eq. 8.3.1, we can make use of the **Asimov dataset**, where the observations n_A, m_A are taken to be their expectation values under μ', b', simplifying the calculation of $\mathcal{I}$ to Eq. 8.3.9.

Explicitly, for our counting experiment (Eq. 8.2.3), the Asimov observations are simply

$$n_A = \mathbb{E}[n] = \mu' s + b', \tag{8.3.25}$$

$$m_A = \mathbb{E}[m] = b'. \tag{8.3.26}$$

We'll now consider a second powerful use of the Asimov dataset to estimate $\sigma_{\hat{\mu}}^2$.

Method 2: The "Asimov sigma" estimate

Putting together Eqs. 8.2.10 and 8.3.26, we can derive a nice property of the Asimov dataset: the MLEs $\hat{\mu}$, $\hat{b}$ equal the true values μ', b':

$$\hat{b} = m_A = b' \tag{8.3.27}$$

$$\hat{\mu} = \frac{n_A - m_A}{s} = \frac{\mu' s + b' - b'}{s} = \mu'. \tag{8.3.28}$$

Thus, t_μ evaluated for the Asimov dataset is exactly the non-centrality parameter Λ that we are after!

$$t_{\mu,A} \simeq \left(\frac{\mu - \hat{\mu}}{\sigma_{\hat{\mu}}}\right)^2 = \left(\frac{\mu - \mu'}{\sigma_{\hat{\mu}}}\right)^2 = \Lambda. \tag{8.3.29}$$

While, not strictly necessary to obtain the asymptotic form for $p(t_\mu|\mu')$, we can also invert this to estimate $\sigma_{\hat{\mu}}$, as

$$\sigma_A \simeq \frac{(\mu - \mu')^2}{t_{\mu,A}}, \tag{8.3.30}$$

where σ_A is known as the "Asimov sigma".

The asymptotic distributions using $\Lambda = t_{\mu,A}$ are plotted in Figure 8.27. We see that this estimate matches the sampling distributions very well, even for cases where the covariance-matrix-estimate failed! Indeed, this is why estimating $\sigma_{\hat{\mu}} \simeq \sigma_A$ is the standard in LHC analyses, and that is the method we'll employ going forward.

Reference [291] conjectures that this is because the Fisher-information-approach is restricted only to estimating the second-order term of Eq. 8.3.17, while with $t_{\mu,A}$ we're matching the shape of the likelihood at the minimum which may be able capture some of the higher order terms as well.

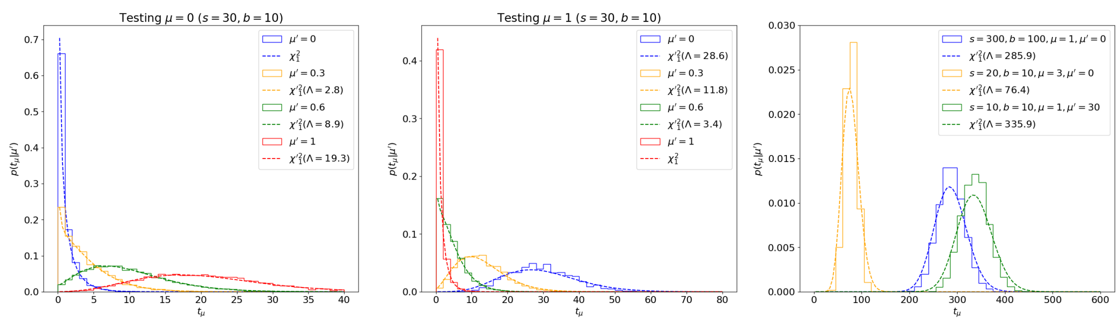


Figure 8.27. Comparing the sampling distribution $p(t_\mu|\mu')$ with non-central $\chi_1'^2(\Lambda)$ distributions for a range of s, b, μ, μ' values, with the Asimov sigma estimation for $\sigma_{\hat{\mu}}^2$.

Despite the pervasive use of the asymptotic formula at the LHC, it's important to remember that it's an *approximation*, only valid for large statistics. Figure 8.28 shows it breaking down for $s, b \lesssim 10$ below.

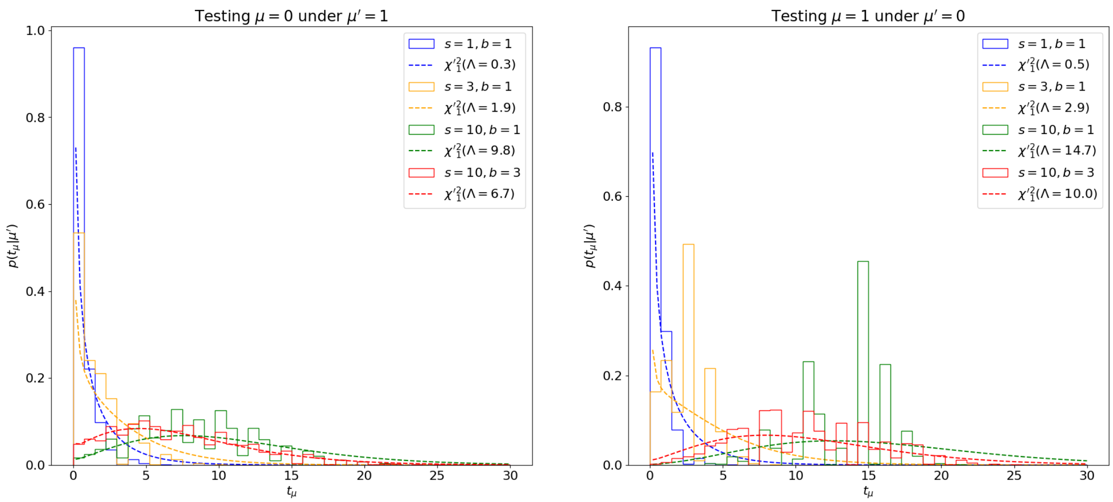


Figure 8.28. Comparing the sampling distribution $p(t_\mu|\mu')$ with non-central $\chi'^2_1(\Lambda)$ distributions for different $s, b \leq 10$, showing the break-down of the σ_A approximation for $\sigma^2_{\hat{\mu}}$ at low statistics.

The PDF and CDF

The probability distribution function (PDF) for a $\chi'^2_k(\Lambda)$ distribution can be found in e.g. Ref. [299] for $k = 1$:

$$p(t_\mu|\mu') \simeq \chi'^2_1(\Lambda) = \frac{1}{2\sqrt{t_\mu}}\left(\varphi(\sqrt{t_\mu} - \sqrt{\Lambda}) + \varphi(\sqrt{t_\mu} + \sqrt{\Lambda})\right), \tag{8.3.31}$$

where φ is the PDF of a standard normal distribution. For $\mu = \mu' \Rightarrow \Lambda = 0$, this simplifies to:

$$p(t_\mu|\mu) \simeq \chi^2 = \frac{1}{\sqrt{t_\mu}}\varphi(\sqrt{t_\mu}). \tag{8.3.32}$$

The cumulative distribution function (CDF) for $k = 1$ is:

$$F(t_\mu|\mu') \simeq \Phi(\sqrt{t_\mu} - \sqrt{\Lambda}) + \Phi(\sqrt{t_\mu} + \sqrt{\Lambda}) - 1, \tag{8.3.33}$$

where Φ is the CDF of the standard normal distribution. For $\mu = \mu' \Rightarrow \Lambda = 0$, again this simplifies to:

$$F(t_\mu|\mu) \simeq 2\Phi(\sqrt{t_\mu}) - 1. \tag{8.3.34}$$

From Eq. 8.2.14, we know the p-value p_μ of the observed t_μ^{obs} under a signal hypothesis of H_μ is

$$p_\mu = 1 - F(t_\mu^{\text{obs}}|\mu) = 2(1 - \Phi(\sqrt{t_\mu^{\text{obs}}})), \tag{8.3.35}$$

with an associated significance

$$Z = \Phi^{-1}(1 - p_\mu) = \Phi^{-1}(2\Phi(\sqrt{t_\mu^{\text{obs}}} - 1) \tag{8.3.36}$$

Application to hypothesis testing

Let's see how well this approximation agrees with the toy-based p-value we found in Example 8.2.1. For the same counting experiment example, where we expect $s = 10$ and observe $n_{\text{obs}} = 20$, $m_{\text{obs}} = 5$, we found the p-value for testing the $\mu = 1$ hypothesis $p_{\mu=1} = 0.3$ (and the associated significance $Z = 0.52$). Calculating

t_μ^{obs} for this example and plugging it into the asymptotic approximation from Eq. 8.3.35 gives:[13]

$$t_\mu^{obs} = 1.08 \tag{8.3.37}$$

$$\Rightarrow p_{\mu=1} = 2(1 - \Phi(\sqrt{1.08})) = 0.3 \tag{8.3.38}$$

$$\Rightarrow Z = 0.52. \tag{8.3.39}$$

We see that it agrees exactly!

The agreement more generally, with varying s, μ, n_{obs}, m_{obs}, is plotted in Figure 8.29. We observe generally strong agreement, except for low n, m where, as expected, the asymptotic approximation breaks down.

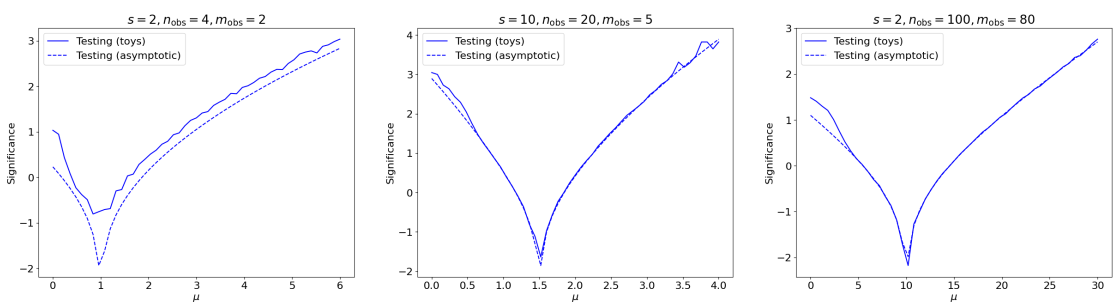


Figure 8.29. Comparing the significances, as a function of the signal strength μ of the hypothesis being tested, for simple counting experiments (Eq. 8.2.3) with different s, n_{obs}, m_{obs}'s, derived using 30,000 toys each (solid) to estimate the $p(t_\mu|\mu)$ distribution vs. the asymptotic approximation (dashed).

[13] Note that we're using t_μ here, not the alternative test statistic $\tilde{t_\mu}$; however, in this case since $\hat{\mu} > 0$, they are equivalent.

Summary

We have been able to find the asymptotic form for the profile-likelihood-ratio test statistic $t_\mu \simeq \frac{(\mu-\hat{\mu})^2}{\sigma_{\hat{\mu}}^2}$, which is distributed as a *non-central chi-squared* ($\chi_k'^2(\Lambda)$) distribution. We discussed two methods for finding the non-centrality parameter Λ, out of which the Asimov sigma σ_A estimation generally performed better. Finally, the asymptotic formulae were applied to simple examples of hypothesis testing to check the agreement with toy-based significances. These asymptotic formulae can be extended to the alternative test statistics for positive signals $\tilde{t}_\mu$ and upper-limit-setting $\tilde{q}_\mu$, as in Ref. [291], to simplify the calculation of both observed and expected significances, limits, and intervals.

With that, we conclude the overview of the statistical interpretation of LHC results. We will see practical applications of these concepts to searches in the high energy Higgs sector in Part V.

Part IV

Accelerating Simulations with AI

Part V

Searches for High Energy Higgs Boson Pairs

Part VI

AI for Jets

Part VII

Appendix

Bibliography

[1] Wikipedia contributors, Standard model — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=Standard_Model&oldid=1238968997 (2024), [Online; accessed 22-August-2024].

[2] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaboration, Precision electroweak measurements on the Z resonance, Phys. Rept. **427**, 257–454 (2006), arXiv:hep-ex/0509008 .

[3] Wikipedia contributors, Yang–mills theory — Wikipedia, the free encyclopedia (2024), [Online; accessed 16-October-2024].

[4] Alex Dias and V. Pleitez, Grand Unification and Proton Stability Near the Peccei-Quinn Scale, Physical Review D **70** (2004).

[5] M. J. Duff and K. S. Stelle, Sir Thomas Walter Bannerman Kibble CBE, Biogr. Mems. Fell. R. Soc. **70**, 225–244 (2021), arXiv:2011.13257 .

[6] Latham Boyle, Standard Model Of Particle Physics–Most Complete Diagram (2024), [Online; accessed 25-September-2024].

[7] Diogo Boito, Consequences of the renormalization group for perturbative quantum chromodynamics, Nature Phys. **19**, 1533–1535 (2023).

[8] F. Bissey, F-G. Cao, A. R. Kitson, A. I. Signal, D. B. Leinweber, B. G. Lasscock, and A. G. Williams, Gluon flux-tube distribution and linear confinement in baryons, Phys. Rev. D **76**, 114512 (2007), arXiv:hep-lat/0606016 .

[9] Derek B. Leinweber, Visualizations of Quantum Chromodynamics (2004), [Online; accessed 25-October-2024].

[10] Jun John Sakurai, *Invariance Principles and Elementary Particles* (Princeton University Press, 2015).

[11] Wikipedia contributors, Isospin — Wikipedia, the free encyclopedia (2024), [Online; accessed 26-October-2024].

[12] Wikipedia contributors, Deep inelastic scattering — Wikipedia, the free encyclopedia (2024), [Online; accessed 28-October-2024].

[13] ATLAS Collaboration, When protons collide (2024), [Online; accessed 28-October-2024].

[14] M. Krasny, F. Dydak, F. Fayette, W. Płaczek, and Andrzej Siodmok, $\Delta M_W \leq 10\,\mathrm{MeV}/c^2$ at the LHC: a forlorn hope?, European Physical Journal C **69**, 379–397 (2010).

[15] Charalampos Anastasiou, Lance J. Dixon, Kirill Melnikov, and Frank Petriello, High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO, Phys. Rev. D **69**, 094008 (2004), arXiv:hep-ph/0312266 .

[16] Monika Kollar, *Top-Quark and Top-Squark Production at Hadron Colliders at Electroweak NLO* (Technische Universitat Munchen, 2007).

[17] H1 Collaboration, Measurement of DIS cross section at HERA, Braz. J. Phys. **37**, 793–797 (2007).

[18] Henning Kirschenmann, *Jet Energy Scale Corrections and their Impact on Measurements of the Top-Quark Mass at CMS*, Ph.D. thesis, U. Hamburg, Dept. Phys. (2014).

[19] L. A. T. Bauerdick *et al.*, Event display for the visualization of CMS events, J. Phys. Conf. Ser. **331**, 072039 (2011).

[20] CMS Collaboration, *Search for highly energetic double Higgs boson production in the two bottom quark and two vector boson all-hadronic final state*, CMS Physics

Analysis Summary CMS-PAS-HIG-23-012 (2024).

[21] Gavin P. Salam, Elements of QCD for hadron colliders, in *2009 European School of High-Energy Physics* (2010) arXiv:1011.5131 .

[22] Stefan Prestel, *The Lund Hadronization Model* (NuSTEC workshop, Gran Sasso Science Institute, 2018) [Online; accessed 29-October-2024].

[23] CMS Collaboration, A portrait of the Higgs boson by the CMS experiment ten years after the discovery., Nature **607**, 60–68 (2022), [Erratum: Nature 623, (2023)], arXiv:2207.00043 .

[24] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN Yellow Reports: Monographs (CERN, Geneva, 2012).

[25] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs (CERN, Geneva, 2017).

[26] CMS Collaboration, A measurement of the Higgs boson mass in the diphoton decay channel, Phys. Lett. B **805**, 135425 (2020), arXiv:2002.06398 .

[27] Jason Veatch, Searches for Resonant Scalar Boson Pair Production Using Run 2 LHC Proton-Proton Collision Data, Symmetry **14**, 260 (2022).

[28] CMS Collaboration, Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at s=13 TeV, Phys. Rev. Lett. **129**, 081802 (2022), arXiv:2202.09617 .

[29] CMS Collaboration, Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at s=13 TeV, Phys. Lett. B **842**, 137531 (2023), arXiv:2206.09401 .

[30] CMS Collaboration, Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s}$ = 13 TeV, JHEP **03**, 257, arXiv:2011.12373 .

[31] CMS Collaboration, Search for Higgs boson pair production in the bbWW

decay mode in proton-proton collisions at $\sqrt{s} = 13$ TeV, Submitted to JHEP (2024), arXiv:2403.09430 .

[32] CMS Collaboration, *Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at $\sqrt{(s)} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-HIG-20-011 (CERN, Geneva, 2024).

[33] Philippe Mouche, Overall view of the LHC. Vue d'ensemble du LHC (2014), general Photo.

[34] O. Bruning, H. Burkhardt, and S. Myers, The Large Hadron Collider, Prog. Part. Nucl. Phys. **67**, 705–734 (2012).

[35] LHC Machine, JINST **3**, S08001.

[36] CMS Collaboration, CMS Luminosity Public Results, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (2024), [Online; accessed 13-November-2024].

[37] Rogelio Tomas Garcia, Luminosity forecasts, https://lhc-commissioning.web.cern.ch/schedule/hl-lhc-plots.htm (2024), [Online; accessed 10-November-2024].

[38] CERN, LHC / HL-LHC Plan, https://hilumilhc.web.cern.ch/content/hl-lhc-project (2024), [Online; accessed 3-December-2024].

[39] CMS Collaboration, Interactive Slice of the CMS detector (2016).

[40] Maximilien Brice and Claudia Marcelloni, Lowering the centre of CMS. Descente dans la caverne de la partie centrale du détecteur de particules CMS (Compact Muon Solenoid). (2007).

[41] CMS Collaboration, Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays, JINST **5**, T03021, arXiv:0910.5530 .

[42] CMS Collaboration, *The Phase-2 Upgrade of the CMS Tracker*, CMS Technical Design Report CERN-LHCC-2017-009, CMS-TDR-014 (2017).

[43] CMS Collaboration, Upgrade of the CMS Barrel Electromagnetic Calorimeter for the High Luminosity LHC, Instruments **6**, 29 (2022).

[44] CMS Collaboration, *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*, CMS Technical Design Report CERN-LHCC-2012-015, CMS-TDR-010 (2012).

[45] CMS Collaboration, The CMS experiment at the CERN LHC, JINST **3**, S08004.

[46] CMS Collaboration, Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$, JINST **13**, P06015, arXiv:1804.04528 .

[47] CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector, JINST **12**, P10003, arXiv:1706.04965 .

[48] Thorben Quast, CMS Calorimeter Endcap Upgrade (HGCAL), in *Beam Test Calorimeter Prototypes for the CMS Calorimeter Endcap Upgrade: Qualification, Performance Validation and Fast Generative Modelling* (Springer International Publishing, 2021) pp. 45–62.

[49] Don Lincoln, ed., *Instrumentation and Techniques in High Energy Physics* (World Scientific, Singapore, 2024).

[50] Iuliana Tabian, Hailing Fu, and Zahra Sharif Khodaei, A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures, Sensors **19**, 10.3390/s19224933 (2019).

[51] GEANT4 Collaboration, GEANT4: A Simulation toolkit, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vol. A506 (2003) p. 250.

[52] Zhantao Chen, Nina Andrejevic, Tess Smidt, Zhiwei Ding, Qian Xu, Yen-Ting Chi, Quynh T. Nguyen, Ahmet Alatas, Jing Kong, and Mingda Li, Direct Prediction of Phonon Density of States With Euclidean Neural Networks, Advanced Science **8**, 2004214 (2021).

[53] Alexander Bogatskiy, Brandon Anderson, Jan Offermann, Marwah Roussi,

David Miller, and Risi Kondor, Lorentz group equivariant neural network for particle physics, in *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 119 (PMLR, 2020) pp. 992–1002.

[54] Lilian Weng, From Autoencoder to Beta-VAE, lilianweng.github.io (2018).

[55] Lilian Weng, What are diffusion models?, lilianweng.github.io (2021).

[56] Matheus Gadelha, Subhransu Maji, and Rui Wang, Shape Generation using Spatially Partitioned Point Clouds, CoRR **abs/1707.06267** (2017), arXiv:1707.06267 .

[57] Kyle Cranmer, Practical Statistics for the LHC, in *2011 European School of High-Energy Physics* (2014) arXiv:1503.07622 .

[58] Caroline Stiller, *Baseline assessment and effect of a supplementary community-based nutrition intervention study on the prevention/treatment of anemia among young Adivasi children in West Bengal, India*, Ph.D. thesis, University of Hohenheim (2021).

[59] ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B **716**, 1–29 (2012), arXiv:1207.7214 .

[60] Wikipedia contributors, Type i and type ii errors — Wikipedia, the free encyclopedia (2024), [Online; accessed 7-December-2024].

[61] Raghav Kansal, Carlos Pareja, Zichun Hao, and Javier Duarte, JetNet: A Python Package for Accessing Open Datasets and Benchmarking Machine Learning Methods in High Energy Physics, J. Open Source Softw. **8**, 5789 (2023).

[62] Steven Tsan, Raghav Kansal, Anthony Aportela, Daniel Diaz, Javier Duarte, Sukanya Krishna, Farouk Mokhtar, Jean-Roch Vlimant, and Maurizio Pierini, Particle Graph Autoencoders and Differentiable, Learned Energy Mover's Distance, in *35th Conference on Neural Information Processing Systems* (2021) arXiv:2111.12849 .

[63] Zichun Hao, Raghav Kansal, Javier Duarte, and Nadezda Chernyavskaya, Lorentz group equivariant autoencoders, Eur. Phys. J. C **83**, 485 (2023), arXiv:2212.07347 .

[64] David E Morrissey and Michael J Ramsey-Musolf, Electroweak baryogenesis, New Journal of Physics **14**, 125003 (2012).

[65] F. Bazzocchi and M. Fabbrichesi, Flavor and electroweak symmetry breaking at the TeV scale, Nuclear Physics B **715**, 372–412 (2005).

[66] Spyros Argyropoulos, Oleg Brandt, and Ulrich Haisch, Collider Searches for Dark Matter through the Higgs Lens, Symmetry **13**, 10.3390/sym13122406 (2021).

[67] Lawrence M. Krauss and James B. Dent, A Higgs–Saw Mechanism as a Source for Dark Energy, Phys. Rev. Lett. **111**, 061802 (2013), arXiv:1306.3239 .

[68] H.R. Pagels, *The Cosmic Code: Quantum Physics as the Language of Nature*, Dover books on physics (Dover Publications, Incorporated, 2012) p. 295.

[69] David Tong, *Lectures on Quantum Field Theory* (University of Cambridge, 2019).

[70] David Tong, *Lectures on Gauge Theory* (University of Cambridge, 2018).

[71] David Tong, *Lectures on the Standard Model* (University of Cambridge, 2019).

[72] John McGreevy, *Lectures notes on Symmetry in physics* (University of California, San Diego, 2020).

[73] Frederic Schuller, *Lectures on Geometrical Anatomy of Theoretical Physics* (Friedrich-Alexander-Universität Erlangen-Nürnberg, 2016).

[74] Anthony Zee, *Group Theory in a Nutshell for Physicists* (Princeton University Press, USA, 2016).

[75] Anthony Zee, *Quantum field theory in a nutshell* (Princeton University Press, USA, 2003).

[76] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, Reading, USA, 1995).

[77] Hong Liu, *Relativistic Quantum Field Theory* (MIT OpenCourseWare, 2023).

[78] Ricardo D. Matheus, *Quantum Field Theory II* (Instituto de Física Teórica, 2023).

[79] Wikipedia contributors, Baker–campbell–hausdorff formula — Wikipedia, the free encyclopedia (2023), [Online; accessed 20-August-2024].

[80] C.G.J. Jacobi, Nova methodus, aequationes differentiales partiales primi ordinis inter numerum variabilium quemcunque propositas integrandi., Journal für die reine und angewandte Mathematik **60**, 1–181 (1862).

[81] Matthew D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, 2014).

[82] Eugene P. Wigner, On Unitary Representations of the Inhomogeneous Lorentz Group, Annals Math. **40**, 149–204 (1939).

[83] user1379857 (https://physics.stackexchange.com/users/157704/user1379857), Proof that $(1/2, 1/2)$ Lorentz group representation is a 4-vector, Physics Stack Exchange.

[84] Gerhart Lüders and Bruno Zumino, Connection between Spin and Statistics, Phys. Rev. **110**, 1450–1453 (1958).

[85] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman Lectures on Physics, Vol. 3* (Addison-Wesley, 1965) p. 4.1.

[86] Robert Foot, H. Lew, X. G. He, and Girish C. Joshi, Seesaw Neutrino Masses Induced by a Triplet of Leptons, Z. Phys. C **44**, 441 (1989).

[87] J. Schechter and J. W. F. Valle, Neutrino Masses in SU(2) x U(1) Theories, Phys. Rev. D **22**, 2227 (1980).

[88] Werner Rodejohann, Neutrino-less Double Beta Decay and Particle Physics, Int. J. Mod. Phys. E **20**, 1833–1930 (2011), arXiv:1106.1334 .

[89] CMS Collaboration, Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP **01**, 122, arXiv:1806.10905 .

[90] Steven Weinberg, The Search for Unity: Notes for a History of Quantum Field Theory, Daedalus **106**, 17–35 (1977).

[91] Aneesh V. Manohar, *Introduction to Effective Field Theories* (Les Houches Summer School on Effective Field Theories, 2018) arXiv:1804.05863 .

[92] Gino Isidori, Felix Wilsch, and Daniel Wyler, The standard model effective field theory at work, Rev. Mod. Phys. **96**, 015006 (2024), arXiv:2303.16922 .

[93] H. Lehmann, K. Symanzik, and W. Zimmermann, On the formulation of quantized field theories, Nuovo Cim. **1**, 205–225 (1955).

[94] Hideki Yukawa, On the Interaction of Elementary Particles I, Proc. Phys. Math. Soc. Jap. **17**, 48–57 (1935).

[95] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, Processes involving charged mesons, Nature **159**, 694–697 (1947).

[96] Zhong-Zhi Xianyu, A Complete Solution to Problems in "An Introduction to Quantum Field Theory" by Peskin and Schroeder (2016), [Online; accessed 29-August-2024].

[97] Dan-di Wu, A Brief Introduction to the Strong CP Problem, Z. Naturforsch. A **52**, 179–181 (1997).

[98] Thomas Mannel, Theory and phenomenology of CP violation, Nucl. Phys. B Proc. Suppl. **167**, 115–119 (2007).

[99] Chen-Ning Yang and Robert L. Mills, Conservation of Isotopic Spin and Isotopic Gauge Invariance, Phys. Rev. **96**, 191–195 (1954).

[100] H. David Politzer, Reliable Perturbative Results for Strong Interactions?, Phys. Rev. Lett. **30**, 1346–1349 (1973).

[101] David J. Gross and Frank Wilczek, Ultraviolet Behavior of Nonabelian

Gauge Theories, Phys. Rev. Lett. **30**, 1343–1346 (1973).

[102] Philip W. Anderson, Plasmons, Gauge Invariance, and Mass, Phys. Rev. **130**, 439–442 (1963).

[103] F. Englert and R. Brout, Broken Symmetry and the Mass of Gauge Vector Mesons, Phys. Rev. Lett. **13**, 321–323 (1964).

[104] Peter W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, Phys. Rev. Lett. **13**, 508–509 (1964).

[105] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Global Conservation Laws and Massless Particles, Phys. Rev. Lett. **13**, 585–587 (1964).

[106] Yoichiro Nambu and G. Jona-Lasinio, Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1., Phys. Rev. **122**, 345–358 (1961).

[107] Yoichiro Nambu and G. Jona-Lasinio, Dynamical model of elementary particles based on an analogy with superconductivity. II., Phys. Rev. **124**, 246–254 (1961).

[108] Sheldon L. Glashow, The renormalizability of vector meson interactions, Nucl. Phys. **10**, 107–117 (1959).

[109] Abdus Salam, Weak and Electromagnetic Interactions, Conf. Proc. C **680519**, 367–377 (1968).

[110] Steven Weinberg, A Model of Leptons, Phys. Rev. Lett. **19**, 1264–1266 (1967).

[111] S. Elitzur, Impossibility of Spontaneously Breaking Local Symmetries, Phys. Rev. D **12**, 3978–3982 (1975).

[112] Werner Heisenberg, Über den bau der atomkerne. i, Zeitschrift für Physik **77**, 1–11 (1932).

[113] Eugene P. Wigner, On the Consequences of the Symmetry of the Nuclear Hamiltonian on the Spectroscopy of Nuclei, Physical Review **51**, 106–119 (1937).

[114] Murray Gell-Mann, The Eightfold Way: A Theory of strong interaction symmetry (1961).

[115] Yuval Ne'eman, Derivation of strong interactions from a gauge invariance, Nucl. Phys. **26**, 222–229 (1961).

[116] Murray Gell-Mann, A Schematic Model of Baryons and Mesons, Phys. Lett. **8**, 214–215 (1964).

[117] George Zweig, An SU(3) model for strong interaction symmetry and its breaking. 1., Phys. Rev. Lett. **12**, 132–135 (1964).

[118] S. L. Glashow, J. Iliopoulos, and L. Maiani, Weak Interactions with Lepton-Hadron Symmetry, Phys. Rev. D **2**, 1285–1292 (1970).

[119] SLAC-SP-017 Collaboration, Discovery of a Narrow Resonance in e^+e^- Annihilation, Phys. Rev. Lett. **33**, 1406–1408 (1974).

[120] E598 Collaboration, Experimental Observation of a Heavy Particle J, Phys. Rev. Lett. **33**, 1404–1406 (1974).

[121] Makoto Kobayashi and Toshihide Maskawa, CP Violation in the Renormalizable Theory of Weak Interaction, Prog. Theor. Phys. **49**, 652–657 (1973).

[122] Martin L. Perl *et al.*, Evidence for Anomalous Lepton Production in e+ - e- Annihilation, Phys. Rev. Lett. **35**, 1489–1492 (1975).

[123] E288 Collaboration, Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions, Phys. Rev. Lett. **39**, 252–255 (1977).

[124] CDF Collaboration, Observation of top quark production in $\bar{p}p$ collisions, Phys. Rev. Lett. **74**, 2626–2631 (1995), arXiv:hep-ex/9503002 .

[125] D0 Collaboration, Observation of the top quark, Phys. Rev. Lett. **74**, 2632–2637 (1995), arXiv:hep-ex/9503003 .

[126] John C. Collins, Davison E. Soper, and George F. Sterman, Factorization of Hard Processes in QCD, Adv. Ser. Direct. High Energy Phys. **5**, 1–91 (1989), arXiv:hep-ph/0409313 .

[127] Torsten Pfoh, LHC Theory Lectures (2012), [Online; accessed 29-October-2024].

[128] Joey Huston, Introduction to QCD from an LHC perspective (2018), [Online; accessed 29-October-2024].

[129] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP **07**, 079, arXiv:1405.0301 .

[130] Katharina Danziger, Stefan Höche, and Frank Siegert, Reducing negative weights in Monte Carlo event generation with Sherpa, arXiv:2110.15211 (2021).

[131] Torbjörn Sjöstrand, Stefan Ask, Christiansen Jesper R., Richard Corke, Nishita Desai, Ilten Philip, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands, An introduction to PYTHIA 8.2, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 .

[132] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes), JHEP **01**, 010, arXiv:hep-ph/0011363 .

[133] Sherpa Collaboration, Event Generation with Sherpa 2.2, SciPost Phys. **7**, 034 (2019), arXiv:1905.09127 .

[134] Johan Alwall *et al.*, Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions, Eur. Phys. J. C **53**, 473–500 (2008), arXiv:0706.2569 .

[135] Stefan Höche, Introduction to parton-shower event generators, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders* (2015) pp. 235–295, arXiv:1411.4085 .

[136] Stefan Höche, *Introduction to Parton Showers* (School and Workshop on pQCD, West Lake, 2018) [Online; accessed 29-October-2024].

[137] E. Fermi, An attempt of a theory of beta radiation. 1., Z. Phys. **88**, 161–177 (1934).

[138] Gargamelle Neutrino Collaboration, Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment, Phys. Lett. B **46**, 138–140 (1973).

[139] UA1 Collaboration, Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s}$ = 540 GeV, Phys. Lett. B **122**, 103–116 (1983).

[140] UA1 Collaboration, Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider, Phys. Lett. B **126**, 398–410 (1983).

[141] UA2 Collaboration, Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider, Phys. Lett. B **122**, 476–485 (1983).

[142] UA2 Collaboration, Evidence for $Z^0 \longrightarrow e^+e^-$ at the CERN $\bar{p}p$ Collider, Phys. Lett. B **129**, 130–140 (1983).

[143] CMS Collaboration, Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC, Phys. Lett. B **716**, 30–61 (2012), arXiv:1207.7235 .

[144] Paul A. M. Dirac, New basis for cosmology, Proc. Roy. Soc. Lond. A **165**, 199–208 (1938).

[145] M. Gell-Mann, The interpretation of the new particles as displaced charge multiplets, Nuovo Cim. **4**, 848–866 (1956).

[146] J. Wess and J. Bagger, *Supersymmetry and supergravity* (Princeton University Press, Princeton, NJ, USA, 1992).

[147] Nathaniel Craig, Jamison Galloway, and Scott Thomas, Searching for Signs of the Second Higgs Doublet, arXiv:1305.2424 (2013).

[148] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt, Two-real-scalar-singlet

extension of the SM: LHC phenomenology and benchmark scenarios, Eur. Phys. J. C **80**, 151 (2020), arXiv:1908.08554 .

[149] Nathaniel Craig, *Naturalness and New Approaches to the Hierarchy Problem* (Prospects in Theoretical Physics, Institute for Advanced Study, 2017) [Online; accessed 30-October-2024].

[150] Ziro Maki, Masami Nakagawa, and Shoichi Sakata, Remarks on the unified model of elementary particles, Prog. Theor. Phys. **28**, 870–880 (1962).

[151] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties* (2013) arXiv:1307.1347 .

[152] Manuel Reichert, Astrid Eichhorn, Holger Gies, Jan M. Pawlowski, Tilman Plehn, and Michael M. Scherer, Probing baryogenesis through the Higgs boson self-coupling, Phys. Rev. D **97**, 075008 (2018), arXiv:1711.00019 .

[153] Massimiliano Grazzini, Gudrun Heinrich, Stephen Jones, Stefan Kallweit, Matthias Kerner, Jonas M. Lindert, and Javier Mazzitelli, Higgs boson pair production at NNLO with top quark mass effects, JHEP **05**, 059, arXiv:1803.02463 .

[154] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, and M. Spira, $gg \to HH$: Combined uncertainties, Phys. Rev. D **103**, 056002 (2021), arXiv:2008.11626 .

[155] Fady Bishara, Roberto Contino, and Juan Rojo, Higgs pair production in vector-boson fusion at the LHC and beyond, Eur. Phys. J. C **77**, 481 (2017), arXiv:1611.03860 .

[156] CMS Collaboration, Search for Nonresonant Pair Production of Highly Energetic Higgs Bosons Decaying to Bottom Quarks, Phys. Rev. Lett. **131**, 041803 (2023), arXiv:2205.06667 .

[157] ATLAS Collaboration, *Search for pair production of boosted Higgs bosons via vector-boson fusion production in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector*, ATLAS Conference Note ATLAS-CONF-2024-003 (2024).

[158] CMS Collaboration, *Performance of heavy-flavour jet identification in boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-BTV-22-001 (2023).

[159] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva, Theory and phenomenology of two-Higgs-doublet models, Phys. Rept. **516**, 1–102 (2012), arXiv:1106.0034 .

[160] Florian Domingo and Sebastian Paßehr, About the bosonic decays of heavy Higgs states in the (N)MSSM, Eur. Phys. J. C **82**, 962 (2022), arXiv:2207.05776 .

[161] CMS Collaboration, Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the four b quarks final state with boosted topology, Phys. Lett. B **842**, 137392 (2023), arXiv:2204.12413 .

[162] CMS Collaboration, Search for a new resonance decaying into two spin-0 bosons in a final state with two photons and two bottom quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV, Submitted to JHEP (2023), arXiv:2310.01643 .

[163] CMS Collaboration, Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau$bb final state at 13 TeV, JHEP **11**, 057, arXiv:2106.10361 .

[164] Bernhard Wolf, *Handbook of Ion Sources* (CRC Press, Boca Raton, FL, 1995).

[165] J. P. Blewett, 200-GeV Intersecting Storage Accelerators, eConf **C710920**, 501 (1971).

[166] CMS Collaboration, Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV, JINST **15**, P10017, arXiv:2006.10165 .

[167] CMS Collaboration, The CMS trigger system, JINST **12**, P01020, arXiv:1609.02366 .

[168] CMS Collaboration, *A CMS Level-1 Track Finder for the HL-LHC*, Tech. Rep. (2021) arXiv:2110.02826 .

[169] Werner Riegler, Particle Physics Instrumentation, https://indico.cern.ch/event/208901/contributions/1501073/attachments/323336/450953/Riegler_L1.pdf (2013), [Online; accessed 12-November-2024].

[170] Hans A. Bethe and Julius Ashkin, Passage of Radiations Through Matter, in *Experimental Nuclear Physics*, Vol. 1 (John Wiley & Sons, New York, 1953) pp. 166–357.

[171] LHCb Collaboration, The LHCb Detector at the LHC, JINST **3**, S08005.

[172] John Sparkes, *Semiconductor Devices*, 2nd ed. (CRC Press, Boca Raton, FL, 1994) p. 78.

[173] Andrei Nomerotski, Silicon detectors for tracking and vertexing, Nucl. Instrum. Meth. A **598**, 33–40 (2009).

[174] RD50 Collaboration, Radiation hard semiconductor devices for very high luminosity colliders, PoS **VERTEX2015**, 032 (2015).

[175] CMS Collaboration, Description and performance of track and primary-vertex reconstruction with the CMS tracker, JINST **9** (10), P10009, arXiv:1405.6569 .

[176] CMS Collaboration, Reconstruction of signal amplitudes in the CMS electromagnetic calorimeter in the presence of overlapping proton-proton interactions, JINST **15** (10), P10002, arXiv:2006.14359 .

[177] CMS Collaboration, The CMS-HF quartz fiber calorimeters, J. Phys. Conf. Ser. **160**, 012014 (2009).

[178] CMS Collaboration, The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC, JINST **8**, P11002, arXiv:1306.6905 .

[179] CMS Collaboration, *The Phase-2 Upgrade of the CMS Muon Detectors*, CMS Technical Design Report CERN-LHCC-2017-012, CMS-TDR-016 (2017).

[180] R. Fruhwirth, Application of Kalman filtering to track and vertex fitting, Nucl. Instrum. Meth. A **262**, 444–450 (1987).

[181] CMS Collaboration, *Identification of b quark jets at the CMS Experiment in the LHC Run 2*, CMS Physics Analysis Summary CMS-PAS-BTV-15-001 (2016).

[182] CMS Collaboration, Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC, JINST **16** (05), P05014, arXiv:2012.06888 .

[183] CMS Collaboration, Performance of the local reconstruction algorithms for the CMS hadron calorimeter with Run 2 data, JINST **18** (11), P11017, arXiv:2306.10355 .

[184] CMS Collaboration, Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s}$ = 13 TeV, JINST **15** (05), P05002, arXiv:1910.00079 .

[185] CMS Collaboration, *Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid*, CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02 (2015).

[186] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, The anti-k_{T} jet clustering algorithm, JHEP **04**, 063, arXiv:0802.1189 .

[187] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, FastJet user manual, Eur. Phys. J. C **72**, 1896 (2012), arXiv:1111.6097 .

[188] CMS Collaboration, *Pileup Removal Algorithms*, CMS Physics Analysis Summary CMS-PAS-JME-14-001 (2014).

[189] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran, Pileup Per Particle Identification, JHEP **10**, 059, arXiv:1407.6013 .

[190] CMS Collaboration, Pileup mitigation at CMS in 13 TeV data, JINST **15**, P09018, arXiv:2003.00503 .

[191] CMS Collaboration, Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV, JINST **10** (06), P06005, arXiv:1502.02701 .

[192] C. Foudas, A. Rose, J. Jones, and G. Hall, A Study for a Tracking Trigger at

First Level for CMS at SLHC (2005), arXiv:physics/0510227 .

[193] Aidan D. Chambers, *Neural-Network Based b-Tagging in the CMS Level 1 Trigger System*, Bachelor's thesis, Massachusetts Institute of Technology (2023).

[194] CMS Collaboration, CMS timing detectors in Phase-2, PoS **LHCP2019**, 250 (2019).

[195] CMS Collaboration, *Physics opportunities with a MIP Timing Detector in CMS at HL-LHC*, CMS Conference Report CMS-CR-2022-199 (CERN, Geneva, 2023).

[196] CMS Collaboration, *A MIP Timing Detector for the CMS Phase-2 Upgrade*, CMS Technical Design Report CERN-LHCC-2019-003, CMS-TDR-020 (2019).

[197] CMS Collaboration, *The Phase-2 Upgrade of the CMS Barrel Calorimeters*, CMS Technical Design Report CERN-LHCC-2017-011, CMS-TDR-015 (CERN, Geneva, 2017) this is the final version, approved by the LHCC.

[198] CMS Collaboration, *The Phase-2 Upgrade of the CMS Endcap Calorimeter*, CMS Technical Design Report CERN-LHCC-2017-023, CMS-TDR-019 (2017).

[199] CMS HGCAL Collaboration, Timing Performance of the CMS High Granularity Calorimeter Prototype, JINST **19** (04), P04015, arXiv:2312.14622 .

[200] CMS Collaboration, *Operation of CMS GE1/1 GEM detectors in Run-3*, CMS Conference Report CMS-CR-2024-210 (CERN, Geneva, 2024).

[201] CMS Muon Collaboration, The upgrade of the CMS muon system for the high luminosity LHC, JINST **19** (02), C02077, arXiv:2312.05939 .

[202] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman, Boosting and Additive Trees, in *The Elements of Statistical Learning* (Springer, New York, 2009) pp. 337–384, 2nd ed.

[203] HEP ML Community, A Living Review of Machine Learning for Particle Physics.

[204] Corinna Cortes and Vladimir Vapnik, Support-Vector Networks, Machine Learning **20**, 273–297 (1995).

[205] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone, *Classification and Regression Trees* (Wadsworth International Group, 1984).

[206] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, eds., *Semi-Supervised Learning* (MIT Press, Cambridge, MA, 2006).

[207] Randall Balestriero, Mark Ibrahim, Vlad Sobal, Ari Morcos, Shashank Shekhar, Tom Goldstein, Florian Bordes, Adrien Bardes, Gregoire Mialon, Yuandong Tian, Avi Schwarzschild, Andrew Gordon Wilson, Jonas Geiping, Quentin Garrido, Pierre Fernandez, Amir Bar, Hamed Pirsiavash, Yann LeCun, and Micah Goldblum, A Cookbook of Self-Supervised Learning (2023), arXiv:2304.12210 .

[208] Geoffrey Hinton, Neural Networks for Machine Learning – Lecture 6e – RMSProp, Coursera Lecture Slides (2012), https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

[209] Diederik P. Kingma and Jimmy Ba, Adam: A Method for Stochastic Optimization, in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2015).

[210] Lutz Prechelt, Early Stopping—But When?, in *Neural Networks: Tricks of the Trade*, Lecture Notes in Computer Science, Vol. 7700 (Springer, Berlin, Heidelberg, 2012) pp. 53–67.

[211] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research **15**, 1929–1958 (2014).

[212] Sergey Ioffe and Christian Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in *Proceedings of the 32nd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 37 (PMLR, Lille, France, 2015) pp. 448–456.

[213] Frank Rosenblatt, The Perceptron: A Probabilistic Model for Information

Storage and Organization in the Brain, Psychological Review **65**, 386–408 (1958).

[214] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, Learning Representations by Back-Propagating Errors, Nature **323**, 533–536 (1986).

[215] Kurt Hornik, Maxwell Stinchcombe, and Halbert White, Multilayer Feedforward Networks are Universal Approximators, Neural Networks **2**, 359–366 (1989).

[216] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems*, Vol. 25 (2012) pp. 1097–1105.

[217] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, Deep Residual Learning for Image Recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) pp. 770–778.

[218] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, Attention Is All You Need, in *Advances in Neural Information Processing Systems*, Vol. 30 (2017) pp. 5998–6008.

[219] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei, Language Models are Few-Shot Learners, Advances in Neural Information Processing Systems **33**, 1877–1901 (2020).

[220] Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman, Jet-images — deep learning edition, JHEP **07**, 069, arXiv:1511.05190 .

[221] CMS Collaboration, Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques, JINST **15** (06), P06005,

arXiv:2004.08262 .

[222] ATLAS Collaboration, *Deep generative models for fast shower simulation in ATLAS*, Tech. Rep. ATL-SOFT-PUB-2018-001 (CERN, Geneva, 2018).

[223] Huilin Qu and Loukas Gouskos, ParticleNet: Jet Tagging via Particle Clouds, Phys. Rev. D **101**, 056019 (2020), arXiv:1902.08570 .

[224] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley, Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, arXiv:1802.08219 (2018).

[225] Taco Cohen and Max Welling, Group Equivariant Convolutional Networks, in *Proceedings of The 33rd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 48 (PMLR, New York, New York, USA, 2016) pp. 2990–2999.

[226] Taco S. Cohen and Max Welling, Steerable CNNs, in *International Conference on Learning Representations* (2017).

[227] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, and G. J. Brostow, Harmonic Networks: Deep Translation and Rotation Equivariance, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, Los Alamitos, CA, USA, 2017) pp. 7168–7177.

[228] Maurice Weiler and Gabriele Cesa, General E(2)-Equivariant Steerable CNNs, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).

[229] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S Cohen, 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data, in *Advances in Neural Information Processing Systems*, Vol. 31 (Curran Associates, Inc., 2018).

[230] Benjamin Kurt Miller, Mario Geiger, Tess E. Smidt, and Frank Noé, Relevance of Rotationally Equivariant Convolutions for Predicting Molecular Properties (2020) arXiv:2008.08461 .

[231] Stephan Eismann, Raphael J.L. Townshend, Nathaniel Thomas, Milind

Jagota, Bowen Jing, and Ron O. Dror, Hierarchical, rotation-equivariant neural networks to select structural models of protein complexes, Proteins: Structure, Function, and Bioinformatics **89**, 493–501 (2021), arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26033 .

[232] Risi Kondor, Zhen Lin, and Shubhendu Trivedi, Clebsch-Gordan Nets: a fully Fourier space spherical convolutional neural network (2018).

[233] Brandon Anderson, Truong-Son Hy, and Risi Kondor, Cormorant: Covariant Molecular Neural Networks (2019).

[234] Maurice Weiler and Gabriele Cesa, General E(2)-equivariant steerable CNNs, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).

[235] I.M. Gelfand, R.A. Minlos, and Z.Y. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Dover Publications, 2018).

[236] Taco S Cohen, Mario Geiger, and Maurice Weiler, A General Theory of Equivariant CNNs on Homogeneous Spaces, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).

[237] Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson, Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data, in *37th International Conference on Machine Learning* (2020) pp. 3146–3157, arXiv:2002.12880 .

[238] G. E. Hinton and R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, Science **313**, 504 (2006).

[239] Giuseppe Di Guglielmo *et al.*, A Reconfigurable Neural Network ASIC for Detector Front-End Data Compression at the HL-LHC, IEEE Trans. Nucl. Sci. **68**, 2179 (2021), arXiv:2105.01683 .

[240] Theo Heimel, Gregor Kasieczka, Tilman Plehn, and Jennifer M. Thompson, QCD or What?, SciPost Phys. **6**, 030 (2019), arXiv:1808.08979 .

[241] Marco Farina, Yuichiro Nakai, and David Shih, Searching for New Physics with Deep Autoencoders, Phys. Rev. D **101**, 075021 (2020), arXiv:1808.08992

.

[242] Olmo Cerri, Thong Q. Nguyen, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant, Variational Autoencoders for New Physics Mining at the Large Hadron Collider, JHEP **05**, 036, arXiv:1811.10276 .

[243] Thorben Finke, Michael Krämer, Alessandro Morandini, Alexander Mück, and Ivan Oleksiyuk, Autoencoders for unsupervised anomaly detection in high energy physics, JHEP **06**, 161, arXiv:2104.09051 .

[244] Gregor Kasieczka *et al.*, The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, Rept. Prog. Phys. **84**, 124201 (2021), arXiv:2101.08320 .

[245] Ekaterina Govorkova *et al.*, Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider, Nat. Mach. Intell. **4**, 154 (2022), arXiv:2108.03986 .

[246] Adrian Alan Pol, Victor Berger, Gianluca Cerminara, Cecile Germain, and Maurizio Pierini, Anomaly Detection With Conditional Variational Autoencoders, in *18th International Conference on Machine Learning and Applications* (2020) arXiv:2010.05531 .

[247] Vishal S. Ngairangbam, Michael Spannowsky, and Michihisa Takeuchi, Anomaly detection in high-energy physics using a quantum autoencoder, Phys. Rev. D **105**, 095004 (2022), arXiv:2112.04958 .

[248] Barry M. Dillon, Gregor Kasieczka, Hans Olischlager, Tilman Plehn, Peter Sorrenson, and Lorenz Vogel, Symmetries, safety, and self-supervision, SciPost Phys. **12**, 188 (2022), arXiv:2108.04253 .

[249] CMS Collaboration, *Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-EXO-22-026 (CERN, Geneva, 2024).

[250] CMS Collaboration, *Anomaly Detection in the CMS Global Trigger Test Crate for Run 3*, CMS Detector Performance Note CMS-DP-2023-079 (2023).

[251] CMS Collaboration, *2024 Data Collected with AXOL1TL Anomaly Detection at*

the CMS Level-1 Trigger, CMS Detector Performance Note CMS-DP-2024-059 (2024).

[252] Dor Bank, Noam Koenigstein, and Raja Giryes, Autoencoders, arXiv:2003.05991 (2020).

[253] Michael Tschannen, Olivier Bachem, and Mario Lucic, Recent Advances in Autoencoder-Based Representation Learning, arXiv:1812.05069 (2018).

[254] Diederik P Kingma and Max Welling, Auto-Encoding Variational Bayes, in *2nd International Conference on Learning Representations, ICLR, Conference Track Proceedings* (2014) arXiv:1312.6114 .

[255] Oliver Atkinson, Akanksha Bhardwaj, Christoph Englert, Vishal S. Ngairangbam, and Michael Spannowsky, Anomaly detection with convolutional Graph Neural Networks, JHEP **08**, 080, arXiv:2105.07988 .

[256] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio, Density estimation using Real NVP, in *International Conference on Learning Representations* (2017).

[257] Diederik P. Kingma and Prafulla Dhariwal, Glow: Generative Flow with Invertible 1×1 Convolutions, in *Advances in Neural Information Processing Systems*, Vol. 31 (2018) pp. 10215–10224.

[258] Claudius Krause and David Shih, Fast and accurate simulations of calorimeter showers with normalizing flows, Phys. Rev. D **107**, 113003 (2023), arXiv:2106.05285 .

[259] Tor Fjelde, Emile Mathieu, and Vincent Dutordoir, An Introduction to Flow Matching (2024), online; accessed 2024-11-26.

[260] Scott H. Hawley, Flow With What You Know (2024), online; accessed 2024-11-26.

[261] Xingchao Liu, Chengyue Gong, and qiang liu, Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow, in *The Eleventh International Conference on Learning Representations* (2023).

[262] Sangyun Lee, Zinan Lin, and Giulia Fanti, Improving the Training of Rec-

tified Flows, in *The Thirty-eighth Annual Conference on Neural Information Processing Systems* (2024).

[263] Breno Orzari, Nadezda Chernyavskaya, Raphael Cobe, Javier Duarte, Jefferson Fialho, Dimitrios Gunopulos, Raghav Kansal, Maurizio Pierini, Thiago Tomei, and Mary Touranakou, LHC hadronic jet generation using convolutional variational autoencoders with normalizing flows, Machine Learning: Science and Technology **4**, 045023 (2023).

[264] Ian J. Goodfellow, Jean Pouget-Abadie, Mirza Mehdi, Bing Xu, David Warde-Farley, Ozair Sherjil, Aaron Courville, and Yoshua Bengio, Generative Adversarial Nets, in *Advances in Neural Information Processing Systems*, Vol. 27 (Curran Associates, Inc., 2014) p. 2672, arXiv:1406.2661 .

[265] Martin Arjovsky, Soumith Chintala, and Léon Bottou, Wasserstein generative adversarial networks, in *Proceedings of the 34th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 70 (PMLR, 2017) pp. 214–223.

[266] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen, Improved techniques for training GANs, in *Advances in Neural Information Processing Systems*, Vol. 29 (2016).

[267] Martin Arjovsky and Léon Bottou, Towards Principled Methods for Training Generative Adversarial Networks, in *International Conference on Learning Representations* (2017).

[268] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin, Which training methods for GANs do actually converge?, in *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80 (2018) pp. 3481–3490.

[269] Tero Karras, Samuli Laine, and Timo Aila, A Style-Based Generator Architecture for Generative Adversarial Networks, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) pp. 4401–4410.

[270] Andrew Brock, Jeff Donahue, and Karen Simonyan, Large scale GAN training for high fidelity natural image synthesis, in *International Conference on Learning Representations (ICLR)* (2019).

[271] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila, Analyzing and improving the image quality of StyleGAN, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020) pp. 8110–8119.

[272] Michela Paganini, Luke de Oliveira, and Nachman Benjamin, CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks, Phys. Rev. D **97**, 014021 (2018), arXiv:1712.10321 .

[273] ATLAS Collaboration, *Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks*, ATLAS Software Public Note ATL-SOFT-PUB-2020-006 (2020).

[274] Jonathan Ho, Ajay Jain, and Pieter Abbeel, Denoising Diffusion Probabilistic Models, in *Advances in Neural Information Processing Systems*, Vol. 33 (2020) pp. 6840–6851.

[275] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, Score-Based Generative Modeling through Stochastic Differential Equations, in *International Conference on Learning Representations* (2021).

[276] Prafulla Dhariwal and Alex Nichol, Diffusion models beat GANs on image synthesis, in *Advances in Neural Information Processing Systems*, Vol. 34 (2021) pp. 8780–8794.

[277] Yadong Lu, Julian Collado, Daniel Whiteson, and Pierre Baldi, Sparse autoregressive models for scalable generation of sparse images in particle physics, Phys. Rev. D **103**, 036012 (2021), arXiv:2009.14017 .

[278] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu, *ShapeNet: An Information-Rich 3D Model Repository*, Technical Report (2015) arXiv:1512.03012 .

[279] Chun-Liang Li, Manzil Zaheer, Yang Zhang, Barnabas Poczos, and Ruslan Salakhutdinov, Point cloud GAN, in *2019 ICLR Workshop Deep Generative Models for Highly Structured Data* (2018) arXiv:1810.05795 .

[280] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan, PointFlow: 3D point cloud generation with continuous normalizing flows, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019) p. 4540, arXiv:1906.12320 .

[281] Roman Klokov, Edmond Boyer, and Jakob Verbeek, Discrete Point Flow Networks for Efficient Point Cloud Generation, in *2020 European Conference on Computer Vision (ECCV)* (Springer International Publishing, Cham, 2020) p. 694, arXiv:2007.10170 .

[282] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan, Learning Gradient Fields for Shape Generation, in *2020 European Conference on Computer Vision (ECCV)* (Cham, 2020) p. 364, arXiv:2008.06520 .

[283] Jonas Köhler, Leon Klein, and Frank Noe, Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities, in *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 119 (PMLR, 2020) p. 5361, arXiv:2006.02425 .

[284] Gregor N. C. Simm, Robert Pinsler, Gábor Csányi, and José Miguel Hernández-Lobato, Symmetry-Aware Actor-Critic for 3D Molecular Design, in *International Conference on Learning Representations* (2021).

[285] Niklas Gebauer, Michael Gastegger, and Kristof Schütt, Symmetry-adapted generation of 3D point sets for the targeted discovery of molecules, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc., 2019) p. 7566, arXiv:1906.00957 .

[286] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas, Learning representations and generative models for 3D point clouds, in *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80 (PMLR, 2018) p. 40.

[287] Diego Valsesia, Giulia Fracastoro, and Enrico Magli, Learning localized generative models for 3D point clouds via graph convolution, in *International Conference on Learning Representations (ICLR) 2019* (2019).

[288] Dong Wook Shu, Sung Woo Park, and Junseok Kwon, 3D point cloud

generative adversarial network based on tree structured graph convolutions, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019) p. 3858, arXiv:1905.06292 .

[289] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas, PointNet: Deep learning on point sets for 3D classification and segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

[290] He Wang, Zetian Jiang, Li Yi, Kaichun Mo, Hao Su, and Leonidas J Guibas, Rethinking sampling in 3D point cloud generative adversarial networks, arXiv:2006.07029 (2020).

[291] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C **71**, 1554 (2011), [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv:1007.1727 .

[292] Charles R. Harris *et al.*, Array programming with NumPy, Nature **585**, 357–362 (2020).

[293] Pauli Virtanen *et al.*, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods **17**, 261–272 (2020).

[294] J. D. Hunter, Matplotlib: A 2D graphics environment, Computing in Science & Engineering **9**, 90–95 (2007).

[295] Wikipedia contributors, Informant (statistics) — Wikipedia, the free encyclopedia (2024), [Online; accessed 7-December-2024].

[296] Wikipedia contributors, Central limit theorem — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=Central_limit_theorem&oldid=1257009135 (2024), [Online; accessed 7-December-2024].

[297] Wikipedia contributors, Law of large numbers — Wikipedia, the free encyclopedia, https://en.wikipedia.org/w/index.php?title=Law_of_large_numbers&oldid=1256071749 (2024), [Online; accessed 7-December-2024].

[298] Gregory Gundersen, Asymptotic Normality of MLE, https://gregorygundersen.com/blog/2019/11/28/asymptotic-normality-mle/

(2019), [Online; accessed 7-December-2024].

[299] Wikipedia contributors, Noncentral chi-squared distribution — Wikipedia, the free encyclopedia (2024), [Online; accessed 7-December-2024].