

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

Copyright
Raghav Kansal, 2024
All rights reserved.

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 गूहळमगेर प्रकेतं सलिलं सर्वाऽइदम् ।
तुच्छ्येनाभ्वपिहितं यदासीत्तपसस्तन्महिनाजायतैकम् ॥३॥

- नासदीय सूक्त

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Epigraph	v
Table of Contents	vi
List of Figures	vii
List of Tables	viii
Acknowledgements	ix
Vita	xiv
Abstract of the Dissertation	xvii
 Introduction	 1
 Part I Theoretical Background	
Bibliography	8

LIST OF FIGURES

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

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 ??—?? 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 ?? 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 ?? 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 ?? on the JETNET library and Lorentz-equivariant ML represents a collection of work [1–3] 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 ?? 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 ?? 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 ?? and ?? and Part ??, 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 ?? 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 ac-

cessing 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 <i>Summa cum laude</i>
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 $H \rightarrow VV$ 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*

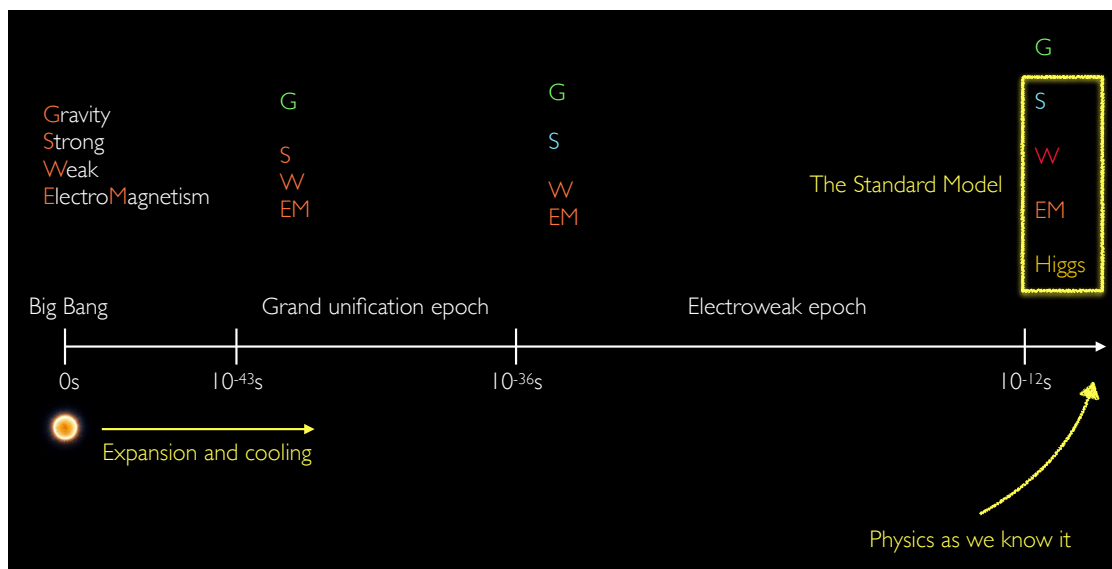


Figure 1. The timeline and evolution of forces in the early universe.

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 *asymmetry* 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 [4] and flavor symmetries [5] as well. The Higgs boson may also be the connection between the SM and the dark sector [6], while the “Higgs-Saw” [7] 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 $< 1\text{ps}$ 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 addi-

tionally 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 ?? introduces the theoretical basis for this dissertation, starting with the mathematical framework behind symmetries in physics (Chapter ??) and of quantum field theory (Chapter ??) before detailing the SM of particle physics (Chapter ??). Part ?? then describes the experimental apparatus used in this dissertation: the LHC (Chapter ??) and the CMS experiment (Chapter ??). Part ?? concludes the background material with an introduction to ML in HEP (Chapter ??), as well as the data analysis and statistical framework used in this dissertation (Chapter ??).

Parts ??—?? comprise the novel contributions of this dissertation. Part ?? 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 ?? 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 ?? 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

Bibliography

- [1] 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 \(2023\)](#).
- [2] 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](#) .
- [3] Zichun Hao, Raghav Kansal, Javier Duarte, and Nadezda Chernyavskaya, Lorentz group equivariant autoencoders, [Eur. Phys. J. C 83, 485 \(2023\)](#), [arXiv:2212.07347](#) .
- [4] David E Morrissey and Michael J Ramsey-Musolf, Electroweak baryogenesis, [New Journal of Physics 14, 125003 \(2012\)](#).
- [5] F. Bazzocchi and M. Fabbrichesi, Flavor and electroweak symmetry breaking at the TeV scale, [Nuclear Physics B 715, 372–412 \(2005\)](#).
- [6] Spyros Argyropoulos, Oleg Brandt, and Ulrich Haisch, Collider Searches for Dark Matter through the Higgs Lens, *Symmetry* **13**, [10.3390/sym13122406](#) (2021).
- [7] 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](#) .