Master's thesis

Master's Programme in Particle Physics and Astrophysical Sciences

Elementary Particle Physics and Cosmology

Modeling Vector Boson Scattering in the Hadronic Final State with `MadGraph5_aMC@NLO`

Alexi Stadnitski

30 July 2024

Supervisor: Henning Kirschenmann

Examiners: Henning Kirschenmann, Nurfikri Norjoharuddeen

University of Helsinki

Faculty of Science

Faculty: Faculty of Science

Degree programme: Master's Programme in Particle Physics and Astrophysical Sciences

Study track: Elementary Particle Physics and Cosmology

Author: Alexi Stadnitski

Thesis title: Modeling Vector Boson Scattering in the Hadronic Final State with `MadGraph5_aMC@NLO`

Level: Master of Science

Month and year: July 2024

Number of pages: 79

Keywords: Vector boson scattering (VBS), Large Hadron Collider (LHC), Compact Muon Solenoid (CMS), high-energy physics (HEP), beyond the Standard Model (BSM), electroweak symmetry breaking (EWSB)

Supervisor: Henning Kirschenmann

Where deposited: HELDA – Helsinki University Digital Library

Abstract

Vector boson scattering (VBS) is a process that occurs when vector bosons, the spin-1 fundamental particles of the Standard Model (SM), are radiated by quarks accelerated to high energies in a collider environment. These bosons may interact with each other before decaying due to their instability. This process is of interest because of its strong sensitivity to physics Beyond the Standard Model (BSM): occurring in the high-energy regime at the cutting edge of modern technology, VBS is one of few processes able to probe the coupling of the gauge fields to each other, as well as to the Higgs field, both of which are sensitive to new physics which may spoil the delicate cancellations required to maintain unitarity. In particular, the polarization states of the massive vector bosons—the $W^{\pm}$ and Z^0, carriers of the weak interaction—are acquired as a direct consequence of electroweak symmetry breaking (EWSB). EWSB has been at the frontier of known physics since the Higgs boson was confirmed to exist in 2012. VBS has three possible decay channels: fully leptonic, semi-leptonic, and all-hadronic. We focus on the all-hadronic channel, in which each boson begets two quarks, to take advantage of its higher branching ratio: nearly half of all VBS events decay through this channel. With the Large Hadron Collider (LHC) in a period of long shutdown as it is upgraded to reach higher luminosities, one of the major steps in VBS analyses is to separate the BSM-sensitive electroweak VBS from chromodynamically-induced (QCD) diboson production, in which vector bosons do not interact; furthermore, a deeper understanding of EWSB requires the various polarization states to be distinguished from each other. To these ends, simulation is a critical component of analyses, as VBS is among the rarest processes at the LHC and overwhelmed by backgrounds with larger cross sections. We investigate the viability of employing the `MadGraph5_aMC@NLO` suite to generate VBS events, successfully producing electroweak, QCD, and mixed samples of all possible combinations of weak bosons: the same-sign $W^{\pm}W^{\pm}$, opposite-sign W^+W^-, $W^{\pm}Z^0$, and Z^0Z^0 channels. Furthermore, we create working samples of these VBS channels with enforced polarization: both bosons longitudinally polarized, both transversely polarized, and one of each. We provide histograms that report differences in the angular distributions of the produced events, showing how to distinguish the modes of production and polarizations based on kinematic topology.

Contents

Acknowledgements

I would like to acknowledge the guidance of my thesis supervisor, Henning Kirschenmann, without whom this research work would not have been possible. I have known Henning since the second year of my bachelor's studies from the course *Applications of Quantum Physics II*, where I was first introduced to many fascinating topics and concepts that motivated me to pursue a career in experimental high-energy physics. I cannot understate the input of the other members of the ForVVard project team, Nurfikri Norjoharuddeen and Nico Toikka, who were always ready to answer any questions I may have had. I worked closely with Nurfikri on every step of this study, and much of the content in this thesis was direcly informed by him. I extend additional thanks to members of the HIP team for lecturing courses which have helped me to understand the complex topics covered throughout my master-level education: Mikko Voutilainen, for supervising my bachelor's thesis and lecturing the course *Introduction to Particle Physics*; Kenneth Österberg, for lecturing the courses *Statistical Methods* and *Particle Physics Phenomenology*, as well as organizing the Seminar in Particle Physics and Astrophysical Sciences, all of which were invaluable resources for this project; Sami Lehti, whose course *Computing Methods in High Energy Physics* was critical in developing the skills with Python and ROOT that I needed for the data analysis components of this work; Erik Brücken and Timo Hildén, whose practical courses with radiation detectors informed my understanding of the instrumentation used in the state of the art of particle physics, such as the CMS inner tracker and its continued development, and added greatly enjoyable hands-on laboratory work to my final year of master's studies. I would also like to thank Patin Inkaew and Mikael Myllymäki for their guidance during my time working at CERN over the summer of 2023. Finally, I extend my gratitude to my parents for their unconditional support throughout my academic career, as well as to my extended family, who were always interested to hear about what I had achieved in my studies.

1 Introduction

1.1 Foreword

The research comprising this thesis was conducted as part of the ForVVard project at the Helsinki Institute of Physics (HIP), a Finnish research institute working closely with the European Organization for Nuclear Research (CERN). HIP is made up of the University of Helsinki, Aalto University, the University of Jyväskylä, the Lappeenranta-Lahti University of Technology, and Tampere University; the Finnish Radiation and Nuclear Safety Authority has been an interim member since 2018. At CERN, HIP primarily collaborates with the Compact Muon Solenoid (CMS) collaboration, although some results from A Toroidal LHC Apparatus (ATLAS) are also referenced in this work.

The structure of this document is the following: a discussion of the physics paradigms currently in favor is intended to introduce the reader to terms and concepts explored in detail in the thesis, and challenges to these paradigms are used to explain the motivation behind the research. The theoretical background is established as the foundation for the experimental and computational components of the work. The relevant physical instrumentation in use at CERN is briefly covered as the source of previous results and foundation for event simulation. The simulation pipeline is examined; with all of the preceding knowledge, the results are thoroughly discussed. The thesis concludes with a summary of the presented research, encountered challenges, and an outlook for the future of high-energy physics related to vector boson scattering. Natural units ($\hbar = c = 1$) are used throughout the work to simplify quantities.

1.2 Motivation

Investigations into the subatomic structure of the universe have their frontier at the Large Hadron Collider (LHC), an underground accelerator complex operated by CERN at the border of Switzerland and France. Proton bunches are accelerated through circular tunnels 27 km in circumference [1], reaching progressively higher centre-of-mass energies of $\sqrt{s}$ = 7, 8, 13, and 13.6 TeV in three successive runs. These high energies allow researchers to probe otherwise prohibitively unlikely interactions, and various specially-designed detectors placed around the tunnel record these processes as experimental data, which is used to confirm and expand theories about the nature of matter and energy.

As the LHC has not directly revealed any behavior expected of new physics at currently achievable energies, attention has turned to rarer interactions that probe Beyond the Standard Model (BSM) phenomena at the higher tail of the achievable energy range [2]. Vector boson scattering (VBS), for example, is attractive due to its high sensitivity to electroweak symmetry breaking, including the ability to probe both the Higgs coupling to the vector bosons as well as anomalous quartic gauge couplings, which may exhibit BSM effects [3] [4]. VBS is expected to offer insight into effective field theories, a set of simple extensions to the Standard Model [2]. The second and third runs of the LHC are the first to reach sufficiently high energies and record enough data for proper Standard Model sensitivity; moreover, incidence of VBS events is thought to increase with increasing centre-of-mass energy [5]. It is the success of this machine, alongside the CMS detector, that makes the content of this thesis possible.

The research presented in this thesis focuses on the massive vector bosons: the charged W^+ and W^- and neutral Z^0 bosons, mediators of the weak interaction. The key component of the research is event simulation with Monte Carlo methods and analysis with Python of the resulting data. This is required because of the incredibly small cross section of VBS; even with the 140 fb^{-1} of data analyzed by CMS over the three years of the LHC's Run 2, only some hundred VBS events were registered.

Motivation behind a kinematic analysis of VBS is strong. With the discovery of the Higgs boson in 2012 and the LHC gearing up for its high-luminosity upgrade, interest is focused around the nature of electroweak symmetry breaking (EWSB) and the Higgs mechanism, both of which VBS and the closely related vector boson fusion (VBF) processes are able to probe. A better understanding of VBS kinematics is instrumental in separating the few events available in the data from overwhelming backgrounds, first from the various gluonic processes that generate similar final states, and then from chromodynamically-induced vector scattering events without sensitivity to new physics.

In this thesis, we have chosen to work with the all-hadronic final state, as it is historically the least thoroughly researched. We leverage the all-hadronic channel's main strength: its branching fraction of nearly 50% in all combinations of $W^\pm$ and Z^0 bosons. In particular, we focus on $W^\pm W^\pm$ scattering with same-sign W bosons, as it has strong leading-order contributions from couplings sensitive to the electroweak and Higgs sectors, which form the expected domain of new physics. However, most of the techniques

we use can be easily extended to the other hadronic VBS combinations, or with a little modification to the leptonic channels.

Over the course of the work, we have aimed to investigate the applicability of the `MadGraph5_aMC@NLO` Monte Carlo suite to Standard Model measurements as well as new physics searches in the context of VBS. We explore the kinematics of the events it is able to generate, enabling us to compare the topologies of different polarization channels at leading order.

1.3 Previous results

Although studies of VBS predate the operation of the LHC, it took until its second run at $\sqrt{s} = 13$ TeV for the process to be experimentally discovered. The first public VBS results were released by the ATLAS collaboration, where boosted decision trees were trained for the $4l$ and $2l2\nu$ channels of ZZjj production. The signal strength was found to be $\mu = 1.35 \pm 0.34$; by comparing to the Standard Model background, the first observation of VBS at the LHC was made with a significance of 5.3σ. Similar studies by CMS found a signal strength of $\mu = 1.39 \pm 0.34^{+0.86}_{-0.65}$ with an observed significance of 2.7σ [2].

Since then, the majority of published studies have focused on the leptonic decay channels of VBS; these channels are appealing due to their superior signal-to-background ratios. However, the small cross section of leptonic decays severely limits the sample size of available data. Previous studies thus heavily rely on simulated events to bolster statistics. Despite the large branching ratio of the all-hadronic decay channel, VBS in general is one of the rarest processes one can expect to see in the LHC, so Monte Carlo event simulation remains a crucial part of the research work. Additionally, events in this channel tend more towards the higher end of the energy spectrum, where BSM events are expected to manifest, than leptonic decays.

The most recent component of our experimental analysis is a comparison of the polarization states of the vector bosons. The polarization states of a weak gauge boson are immediate consequences of its acquisition of mass via the Higgs mechanism, necessary to maintain unitarity in line with the Standard Model, making this sector a prime target for new physics searches. We compare the angular distributions of VBS processes with both particles longitudinally polarized, both transversely polarized, or with one of each. Most of the previous analyses have been reconstruction-level; the research in this work is

performed instead with generator-level event simulation, which offers a better view of the underlying physics. Establishing an understanding of the kinematic topology of hadronic VBS events is necessary to filter out the signal from a substantial background in real data.

2 Theory

2.1 Standard Model

2.1.1 Fundamental particles and their properties

From the middle of the twentieth century to the modern day, the Standard Model (SM) has been the main paradigm in use by physicists to explain the natural world at the subatomic level. The SM framework has made numerous theoretical predictions that have been experimentally confirmed once instrumentation developed sufficiently; the most recent such confirmation was the existence of the Higgs boson in 2012, nearly half a century after its initial prediction in theory [6].

The Standard Model is itself built on quantum field theory (QFT), a framework that interprets the elementary particles as being excitations of fields, which have certain values at every point in space and time. As a whole, the Standard Model interactions obey the $\mathcal{G} = \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \times \mathrm{SU}(3)_C$ gauge group, with each symmetry corresponding to the quantum numbers of weak isospin, hypercharge, and color charge [7]. Specific quantum field theories include quantum chromodynamics for fields with color charge, quantum electrodynamics for fields with electric charge, and quantum flavordynamics for the study of particles changing flavor. In practice, the latter two are combined into electroweak theory, as they are now understood to be two aspects of the same fundamental field, unifying at an energy scale of 246 GeV, corresponding to the vacuum expectation value of the Higgs field; this value is set by observation of the $\mathrm{W}^{\pm}$ mass [8] [9] [10]. The strong force, and possibly gravity, are also expected to join this union, but at still higher energies. Understanding these energy scales are among the goals of a potential Grand Unified Theory (GUT) and Theory of Everything (TOE), respectively.

The fundamental SM particles are distinguished from each other by intrinsic properties such as mass and electrical charge. Another such property is spin, a form of angular

momentum; it is the metric by which particles are categorized as bosons or fermions. Bosons have integer spin values, while the spins of fermions are always halfway between two integers. As a consequence of quantum mechanics (QM), a major difference between the two types is their observance of the Pauli exclusion principle: multiple identical fermions are unable to occupy the same quantum state, while multiple bosons may do so. All fundamental fermions—that is, the quarks and leptons, shown in table 1—have a spin of exactly 1/2. Composite fermions may have higher spins. These particles are so named because they follow Fermi-Dirac statistics, characterized by the distribution function

$$f(E) = \left[A \exp\left(\frac{E - E_F}{k_B T}\right) + 1\right]^{-1}$$

where E_F is the Fermi energy, k_B is the Boltzmann constant, and T is the temperature of the system. A is a normalization constant, ensuring that the sum of all probabilities is unity. The function takes as an argument the energy E of the particle.

Fermions		Generation I	Generation II	Generation III
Quarks	Up-type	[u]	[c]	[t]
		Up quark	Charm quark	Top quark
	Down-type	[d]	[s]	[b]
		Down quark	Strange quark	Bottom quark
Leptons	Charged	[e]	[μ]	[τ]
		Electron	Muon	Tau
	Neutrinos	[ν_e]	[ν_μ]	[ν_τ]
		Electron neutrino	Muon neutrino	Tau neutrino

Table 1: Experimentally confirmed fundamental SM fermions. As a general rule, mass increases left to right with increasing generation, although in the case of the neutrinos an alternate hierarchy may be possible.

Quarks are the indivisible components of matter. Atomic nuclei are composed of protons and neutrons, which are further formed exclusively of the first generation of quarks. Heavier quarks may form short-lived mesons with their antiparticles or exotic baryons with other quarks. All quarks (except for the top quark, which has a lifetime shorter than the timescale of strong interactions owing to its high mass) hadronize immediately due to color confinement, and are observed as jets—showers of particles registered across

multiple pixels of a detector—rather than individual, pointlike particles. VBS events in which the vector bosons decay to a quark and antiquark are termed hadronic. This thesis focuses on all-hadronic decays, in which each vector boson decays into a quark and an antiquark. Additionally, at the generator level of event simulation with which this work is concerned, quarks may be studied as single particles, prior to hadronization.

Like the quarks, leptons are split across three generations by increasing mass, and further categorized as either charged leptons or their corresponding neutrinos. These are produced in weak decays along with their partners to conserve leptonic quantum number. The muon and tau are massive enough to decay into lighter particles, while the electron and all neutrinos are expected to be stable. The masses of the neutrinos are unknown, and they cannot be generated from the Higgs mechanism [11]. All known neutrinos are left-handed in terms of chirality. In the context of VBS, a semi-leptonic or partially leptonic decay channel is one in which one vector boson decays hadronically and one leptonically; fully leptonic decays see the two vector bosons become four leptons.

Bosons are so named because they obey Bose-Einstein statistics:

$$f(E) = \left[A \exp\left(\frac{E}{k_B T} \right) - 1 \right]^{-1}$$

Bosons with a spin of one are known as vector bosons, as according to QFT, the fields they represent have both magnitude and direction at every point in spacetime. In the SM framework, each vector boson is also a gauge boson, a carrier of a fundamental force. Gauge bosons are exchanged whenever an interaction of the corresponding type occurs. These correspondences are delineated in table 2.

Bosons	Strong	Electromagnetic	Weak		Higgs
Vector bosons	[g] Gluon	[γ] Photon	[$W^{\pm}$] W bosons	[Z^0] Z boson	
Scalar bosons					[H^0] Higgs boson

Table 2: Experimentally confirmed fundamental SM bosons. A $W^{\pm}$ boson may be either positively or negatively charged, with the two W^+ and W^- bosons being the antiparticles of each other. Furthermore, many SM extensions postulate the existence of further Higgs bosons, but so far only the neutral charge conjugation parity (CP)-even H^0 has been detected.

The gluon and photon are both massless and electrically neutral. For the photon, which couples to particles with nonzero electric charge, this means that it does not interact with other photons. In contrast, all gluons have one of the three color charges (or "anticolors" in the case of antigluons), so self-interaction between gluons is common. Gluons inexorably join systems of zero net color; the energy required to overcome the potential barrier holding them together is greater than that needed to produce replacement particles, which again lead to a colorless state, meaning that free gluons are never observed. This phenomenon is termed color confinement.

The remaining vector bosons, responsible for mediating the weak interaction, are all massive as a consequence of electroweak symmetry breaking (EWSB). This failure to maintain local gauge symmetry produces massless Goldstone vector bosons from three of the four longitudinal degrees of freedom belonging to the Higgs field; the coupling of the electroweak field to the Higgs allows the Goldstone particles' longitudinal polarizations to be absorbed by the $W^{\pm}$ and Z^0 bosons, granting them mass in the process. This is known as the Higgs mechanism [8]. As the polarization states of the weak bosons are direct consequences of the Higgs field, longitudinally-polarized VBS is a strong probe of the Higgs sector [3] [12].

<table>
<tr><th colspan="3">Vector bosons</th><th>Charge [e]</th><th>Mass [GeV]</th><th>Decay width [GeV]</th><th>ID</th></tr>
<tr><td colspan="2">Strong</td><td>g</td><td rowspan="3">0</td><td rowspan="2">0</td><td rowspan="2">Stable</td><td>21</td></tr>
<tr><td rowspan="4">Electroweak</td><td>EM</td><td>γ</td><td>22</td></tr>
<tr><td rowspan="3">Weak</td><td>Z^0</td><td>91.1880 ± 0.0020</td><td>2.4955 ± 0.0023</td><td>23</td></tr>
<tr><td>W^+</td><td>+1</td><td rowspan="2">80.3692 ± 0.0133</td><td rowspan="2">2.085 ± 0.042</td><td>24</td></tr>
<tr><td>W^-</td><td>−1</td><td>−24</td></tr>
</table>

Table 3: Selected properties of the SM vector bosons. Numerical values are obtained from the 2024 Particle Data Group review [13] and Monte Carlo IDs refer to the 2019 revision of the Particle Data Group numbering scheme [14].

The Higgs differs from the gauge bosons in that it transforms as a scalar under Lorentz transformations due to having zero net spin. Fermions that couple to the Higgs field gain mass based on the strength of their Yukawa coupling, a phenomenon separate from that responsible for the weak boson masses. In the two-Higgs doublet model (2HDM), a currently popular extension to the Standard Model, five Higgs particles are predicted to exist in total: the A boson, which violates charge conjugation parity (CP) symmetry,

switching sign under parity transformation; two charged bosons, H^+ and H^-; and two which are both CP-even and neutral, the light h^0 and heavy H^0 boson, of which the latter is thought to correspond with the Higgs boson experimentally confirmed in 2012. Due to self-coupling, all Higgs bosons have nonzero mass [8].

The Standard Model also allows for the existence of tensor bosons, which would have a spin of two. This category would include the graviton, a hypothetical interaction mediator for gravitation; following from the fact that gravity propagates at the speed of light, the graviton must, like the photon, be massless. However, gravitons interact so weakly that they are for all intents and purposes impossible to experimentally confirm. The spin-2 nature of the graviton also means there is no renormalizable quantum theory for gravity [11]. Studies on the nature of gravity instead revolve around the detection of gravitational waves from cataclysmic astrophysical events, such as black hole mergers, which are the only sources that produce enough gravitational radiation to be observable with ground-based instruments such as the Laser Interferometer Gravitational-Wave Observatory (LIGO).

2.1.2 The electroweak gauge

As its name suggests, the electroweak interaction is the union of the electromagnetic and weak interactions, which function as a single entity at scales higher than the Higgs field's vacuum expectation value. Although energies beyond this electroweak scale have not been achieved since the Big Bang, quantum electrodynamics and flavordynamics are understood well enough to explain both with a single theory.

The electromagnetic interaction affects all particles that are not electrically neutral, manifesting as a mutual repulsion between similarly charged objects and an attraction between oppositely charged ones. The strength of the force depends on the magnitude of the charges, and varies with the inverse of the square of their separation. Because the excitation of the quantum electromagnetic field, the photon, is massless, these interactions propagate at the speed of light, and their range is unlimited; every charged particle in existence has a mutual action with every other charged particle, regardless of how far apart they are in the universe.

Conversely, the range of the weak interaction is extremely short, only 10^{-17} m, limited by the distances the $W^{\pm}$ and Z^0 bosons are able to travel before decaying due to their

heavy masses. The quantum number conserved by the weak force is known as weak isospin. All fundamental fermions, as well as the electroweak bosons, have nonzero weak isospin and hence couple to the weak interaction; exchange of the weak vector bosons facilitates the change from one flavor to the next [11]. The lepton flavors all couple with the same strength to the weak interaction. For quarks, the probability to go from any given flavor to another is described by the unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{bmatrix}$$

The square of each quantity V_{ij} is proportional to the probability of conversion between the quark flavors i and j; up-type quarks may only become down-type and vice versa. From the near-unity values of the matrix's diagonal elements, it can be seen that conversions are far more likely to occur between quarks of the same generation, although conversions across generations are not so rare as to be negligible. The weak charged-current interaction, mediated by the exchange of the W^+ and W^- bosons, does not respect charge-conjugation symmetry; it is also the only SM interaction to violate parity symmetry, as it couples differently to left-handed and right-handed chiral states. Charge conjugation refers to the inversion of particles with their antiparticles, while parity transformations invert spatial coordinates, performing the operation $\hat{P}\psi(\vec{x},t) = \psi(-\vec{x},t)$ [8].

The behavior of the $W^\pm$ bosons is described by the non-Abelian SU(2) gauge group. The electromagnetic and weak interactions are able to be combined because the photon and $W^\pm$ bosons are all part of one SU(2) multiplet. However, fields in the SU(2) gauge are massless; to explain the required massive nature of the $W^\pm$ bosons, the SU(2) symmetry must be broken. This is done by an extension to SU(2)×U(1), corresponding to the Higgs mechanism. Parity is broken, as chirally right-handed fermions form singlets while left-handed ones form doublets. This revelation also led to the theoretical prediction of the massive Z^0 boson's inclusion into the Standard Model; its existence is a consequence of the fourth degree of freedom of the Goldstone boson (the first three are acquired by the W^+, W^-, and Higgs bosons) [3]. The $W^\pm$ polarization states we study in this thesis are direct manifestations of those degrees of freedom, taken on by the electroweak bosons

in acquiring mass via the Higgs mechanism.

2.1.3 Quartic couplings

A quartic gauge coupling (QGC) is a four-way interaction vertex in which all participating particles are excitations of gauge fields. They are important potential indicators of new physics and are hence a popular target for electroweak studies. Feynman diagrams of the four SM-permitted vector boson QGCs are illustrated in figure 1. VBS is one of a small number of processes that may contain this type of coupling [15].

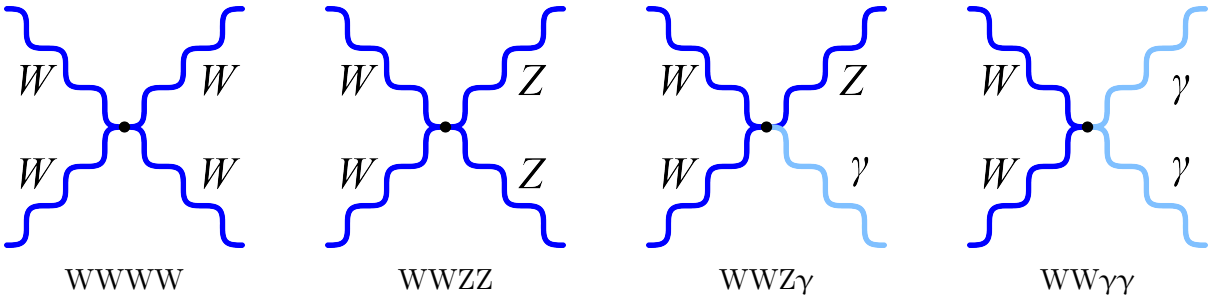


Figure 1: Feynman diagrams of the four permitted electroweak quartic couplings [3]. The two diagrams on the left were studied over the course of our research, with particular focus on the quadruple-W coupling due to its greater sensitivity to the Higgs and electroweak sectors.

Unitarity is the requirement that, for a system in a given initial state, the probabilities of all possible final states sum to one. Theories are expected to be unitary in order to have any physical meaning. Quartic couplings in VBS would be expected to violate unitarity if not for cancellations due to gauge invariance and the presence of a Higgs boson with a mass below 1 TeV [3]. Indeed, this is the source of the so-called LHC "no-lose theorem": either the Higgs boson would be found within the aforementioned regime, or the discovery of its absence would imply some alternative new physics needed to maintain unitarity [16]. The 2012 experimental confirmation of the Higgs boson at 125.09 GeV offers a satisfying reasoning for unitarization, and the influence of the Higgs field on vector processes means that VBS research continues to be a sensitive probe of Higgs-sector physics [6] [10] [12].

An anomalous quartic gauge coupling (aQGC) is a QGC that deviates from known physics when something interferes with the stringent cancellations predicted by the Standard Model [16]. These deviations scale with the fourth power of energy. Representative diagrams at different energy scales are demonstrated in figure 2. High statistics from ATLAS and CMS without detected aQGCs allow for limits to be placed in terms of energy

[3]. Published results have found upper limits on the WW$\gamma\gamma$ and WWZγ aQGCs at the 95% confidence level, but no such results yet exist for the WWWW vertex, meaning its limits are still quite wide [17]. These limits are used to constrain the directions of potential divergences, helping to focus future searches; moreover, theoretical calculations of positivity conditions have narrowed down the viable space to only 2.1% of the initial parameter space [4]. These searches are important because aQGCs could provide evidence of new physics, and are well probed by all types of VBS [3].

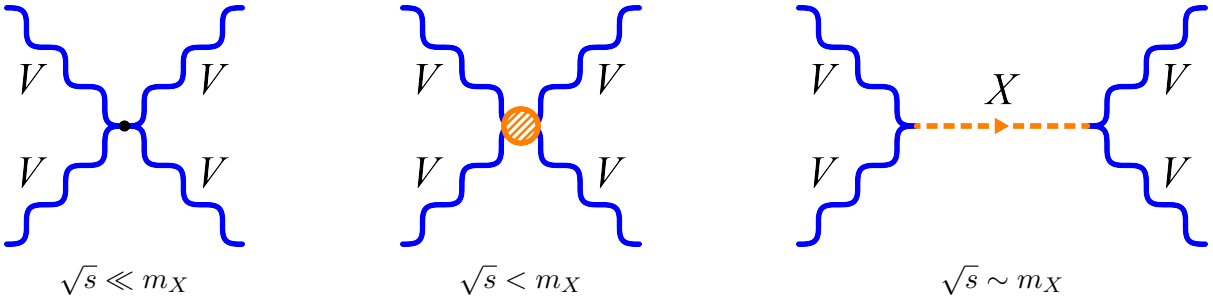


Figure 2: Three Feynman diagrams of the same process as it would appear at different energy scales. X represents some hypothetical particle that couples to the vector bosons and has a mass m_X greater than the highest centre-of-mass energies achieved at the LHC. At very low energy scales, $\sqrt{s} \ll m_X$, the process corresponds to the SM predictions for the QGC. At higher energies, but still below the m_X scale, evidence of its interaction may manifest as the aQGC. At energies comparable to m_X, the coupling of X to gauge bosons would be recovered, in this example producing an s-channel diagram [18].

2.2 Beyond the Standard Model

While every particle in the Standard Model has been experimentally confirmed, it is not a final theory of physics as there are many phenomena—such as unlikely coincidences, delicate cancellations, or exotic signatures—that it fails to explain. A notable example from astronomical evidence is dark matter (DM), which vastly outnumbers regular matter in the universe, but does not clearly correspond to any SM particle; in turn, there is a great asymmetry of regular matter over antimatter, which also has no known explanation. Another is the hierarchy problem, regarding the enormous difference in strength of the fundamental interactions, for which the Standard Model offers no mechanism [7]. Additionally, the Standard Model does not incorporate gravity in its current state, while also appearing to be incompatible with general relativity, which for its part does not take into account the quantum nature of the universe. These are the attractive unsolved problems at the forefront of modern physics.

Theories that seek to address the questions left open by the Standard Model are termed BSM physics. Some popular extensions advanced by theorists include supersymmetry, which offers compelling DM candidates and a potential explanation for the hierarchy problem, and string theory, which endeavors to join the Standard Model and relativity by reconsidering particles as one-dimensional "strings"; however, neither of these theories have been supported by any evidence from the LHC, and interest in them has tended to wane over time. From an experimentalist's perspective, research at the frontier of science involves searching for evidence of previously unseen physics, and comparing to see if it lines up with theoretical BSM predictions. The main motivation behind the drive for ever larger accelerators is the hope of expanding the accessible parameter space to regions in which these effects may manifest. Experimental results from CMS have set a lower limit to BSM behavior near the TeV scale [4].

In practice, there are two main ways to observe new physics. One is to study a new part of the phase space, in this case diboson hadronic decay, while treating the known physics as an expected background. Experimentally, the background should be minimized by applying clever cuts at the analysis level [5]. Subtracting the irreducible background from the observed distributions may reveal a leftover signal. The other is to indirectly search for new physics by its influences on known phenomena; in our analysis, such a phenomenon might be the QGC. In this case, a known background is changed to something new. Modern extensions to the Standard Model fall into one of three categories depending on their selected energy range: model-independent (or effective) field theories, simplified models leveraging additional resonances for diboson production, and ultraviolet (UV)-complete models, designed to extend to even higher scales [2].

2.3 Effective field theories

An effective field theory (EFT) is a basic QFT expansion to the Standard Model that allows a model-independent prediction of new physics [19]. It is especially useful when direct observation of new particles is unreasonable or unlikely. EFT operators (or more technically, constraints on the coefficients of these operators from electroweak precision tests) can be used to qualify BSM effects on VBS processes [20]. Such theories take advantage of the hierarchies of physical quantities to isolate scales, working with one at a time. The key components of an EFT are the degrees of freedom relevant to the selected

physical system, the symmetries defining its dynamics, and the expansion parameters that restrict the impact of physics beyond the scale being considered [21].

With these tools, an EFT can simplify calculations by enacting a modular approach: it may treat the Standard Model as just a theory of the low-energy behavior of some more general physical model that is UV-complete, meaning that it will remain valid at arbitrarily high energies [22]. As their scales of choice, EFTs usually target the lower energies just beyond the currently explored phase space [2]. Limiting the scale of the system may produce symmetries that were invisible when the entire physics were taken into account [21].

Of interest, Higgs effective field theory (HEFT) is an effective theory that treats the longitudinal polarizations of the weak boson fields as a sigma model of the Goldstone bosons and, unlike in EWSB as it is generally understood, introduces the scalar H^0 boson as a separate field. In this model, the Goldstone interactions are generally non-linear. Linearizing the model means taking the four fields together to form the SU(2)×U(1) gauge multiplet; this is a special case within HEFT called Standard Model effective field theory (SMEFT) [2]. This model introduces the Lagrangian

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \frac{\mathcal{L}_{\text{dim-6}}}{\Lambda^2} + \frac{\mathcal{L}_{\text{dim-8}}}{\Lambda^4} + ...$$

via a Taylor expansion of the SM Lagrangian with higher-dimensional terms. The SMEFT scale Λ is taken to be far above the electroweak scale [22]. The location of Λ in the denominator of the coefficients demonstrates that larger scales required for BSM effects to manifest lead to smaller BSM contributions to the overall Lagrangian, consistent with the suppression of these events at low ($\sqrt{s} \sim 8$ TeV) energies.

In order to preserve unitarity when working with higher-order operators, the EFT may be chosen to hold up to a chosen energy bound, beyond which the SM prediction takes over. This spliced-together model can then be compared to experimental evidence [18]. Dimension-6 operators modify the behavior of both the triple gauge coupling (TGC) and the QGC, while dimension-8 operators only act on QGCs [19]. There are 3045 dimension-6 operators and 44807 dimension-8 operators; at yet higher dimensionalities, the count of operators continues to increase exponentially, but their suppression by the higher power of the Λ scale makes their impact negligible. Indeed, without even any experimental confirmation at $D = 6$, it is unreasonable to expect such high-order effects

to appear in observations [23]. Limits on these operators may be recovered experimentally through searches in VBS and vector boson fusion, as well as more general diboson and triboson production [18]. VBS diboson production in particular is the most promising for constraints on dimension-8 operator couplings due to its high branching ratio [2]. Moreover, evidence of EFTs can be probed by studying polarized VBS particles, as higher-dimension EFT operators may induce polarization [24].

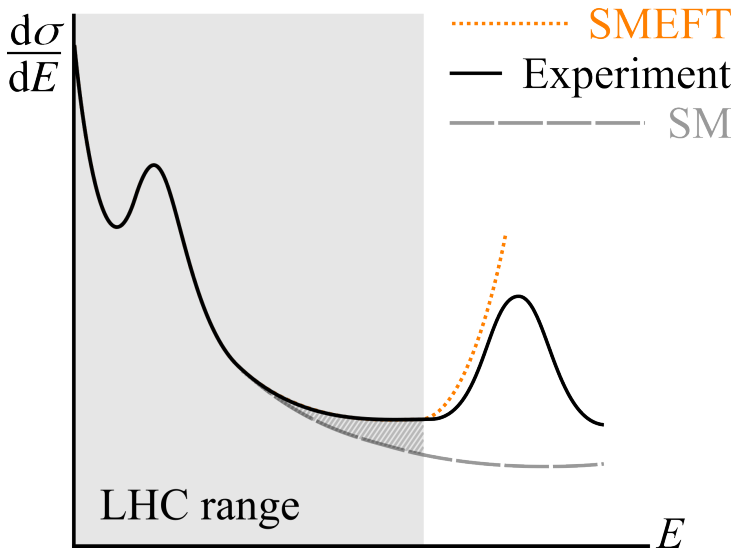


Figure 3: A simple example of divergent behavior at high energies near the upper limit of the LHC's accessible range. Observed experimental behavior hewing more closely to an EFT than the Standard Model would be evidence in favor of BSM physics. Original figure by C. Severi, M. Thomas, and E. Vryonidou [25] [26].

The primary targets of our study, with the help of bootstrapped data, are the distributions of kinematic variables at high energy in hopes of detecting enhancements in the tail behavior of SM physics. Figure 3 is a popular example illustrating how the appearance of SMEFT effects depends on energy. An absence of these resonances would imply the SMEFT scale to be significantly above the electroweak scale [23].

2.4 Vector boson scattering

2.4.1 Occurrence and BSM prospects

At the parton level, VBS occurs when a quark from each accelerated proton bunch emits a vector boson; if the emitted particle is a $W^{\pm}$ boson, the quark changes flavor according to the CKM matrix. The two vector bosons then interact with each other in one of many possible processes, eventually decaying into a total of four fermions. In our Monte Carlo

(MC) simulations, permitted processes are specified with cuts and other parameters in input files prior to event generation. Regarding the final state, in a detector, between two and six jets are registered; the quarks from the protons always hadronize, while the bosons' decay products may hadronize if they are quarks, trigger calorimeters if they are charged leptons, or exit without interacting if they are neutrinos. In practice, these would be observed as two (leptonic), four (semi-leptonic), or six jets (hadronic). The events we simulate in the experimental component of our research should in principle register a signature of six jets, although highly boosted regimes may cause the two jets of two decay products to be resolved as a single, wide-cone jet.

The combinations of weak bosons in the types of VBS studied in our research are $W^{\pm}W^{\pm}$ (two charged-current weak bosons of the same sign), $W^{+}W^{-}$ (two of opposite sign), $W^{\pm}Z^{0}$, and $Z^{0}Z^{0}$. Events with Z^{0} bosons are rarer, with cross sections significantly lower than those of $W^{\pm}W^{\pm}$ or $W^{+}W^{-}$ scattering [27]. The $W^{\pm}W^{\pm}$ process has been found to return the highest signal-to-background ratio [3] [28]. At leading order (LO), the main contributions to $W^{\pm}W^{\pm}$ VBS come from the WWWW quartic vertex, Z^{0}/γ exchange, and Higgs exchange. In the opposite-sign case, where *s*-channel diagrams are allowed, there are five processes contributing at LO [3]. Feynman diagrams of the possible SM processes are shown in figure 4.

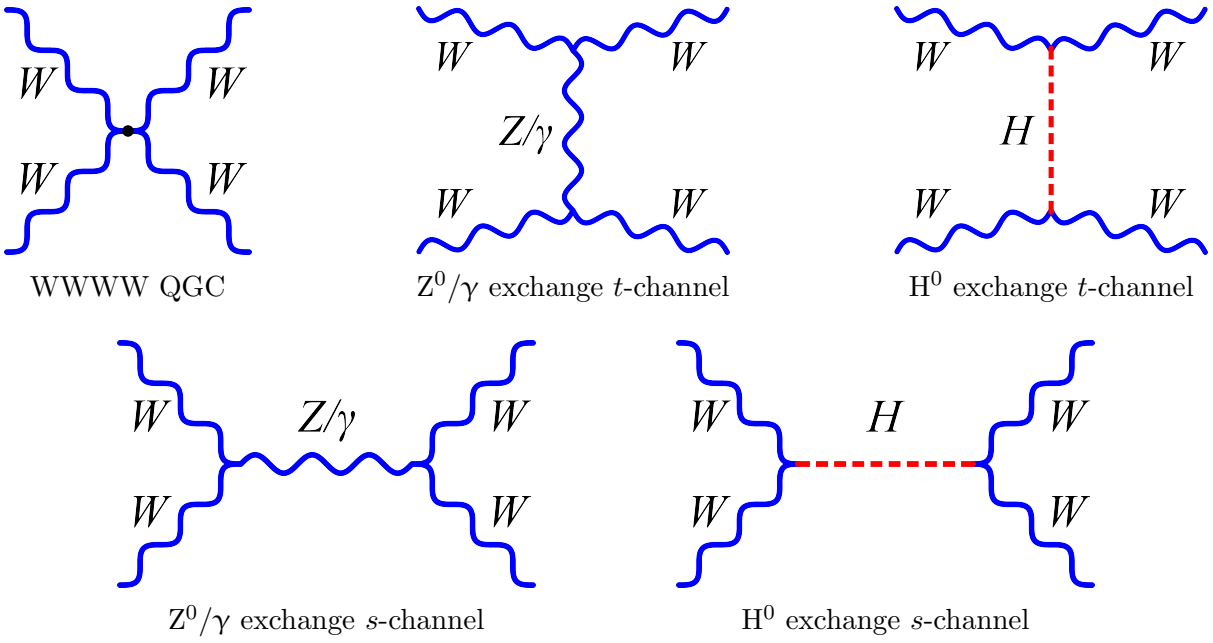


Figure 4: LO WW→WW processes permitted in the SM framework. The top row holds the three $W^{\pm}W^{\pm}\rightarrow W^{\pm}W^{\pm}$ processes, while all five can occur in $W^{+}W^{-}\rightarrow W^{+}W^{-}$.

We have chosen to focus our analysis on $W^{\pm}W^{\pm} \to W^{\pm}W^{\pm}$ in particular, as with three allowed LO diagrams instead of five, the QGC and t-channel Higgs diagrams probing the electroweak and Higgs sector for new physics each have greater contributions.

Every successive LHC upgrade has vastly improved the odds of encountering VBS events. In addition to monitoring aQGCs for BSM effects, a further increase in the cross section of VBS events at higher centre-of-mass energies may be a sign of new physics. This may be due to extended Higgs sectors; even though only one Higgs particle, a 125 GeV neutral scalar, has been experimentally confirmed by LHC experiments, the Standard Model places few limits on how many additional excited modes of the Higgs field may manifest as fundamental particles. Instead, improved data-taking during Run 2 and Run 3 after the initial 2012 discovery without additional detections have served to constrict the parameter space of possible extensions to the Higgs sector. Some extended-Higgs models currently being explored include

- The complex singlet extension to the Standard Model (CxSM): two additional CP-even scalar Higgs bosons arising from a soft breaking of a global U(1) symmetry, for a total of three Higgs states.

- The CP-conserving two-Higgs doublet model (2HDM): introduces the neutral pseudoscalar A boson, another neutral scalar Higgs h_0 (unless it is more massive than 125 GeV, in which case it inherits the H^0 symbol), and two oppositely-charged scalar bosons, for a total of five Higgs states. This is the extension favored by the Minimal Supersymmetric Standard Model (MSSM).

- The CP-violating two-Higgs doublet model (C2HDM): a more general version of the 2HDM, with complex terms in the potential so that none of the five Higgs bosons have a defined CP behavior.

- The singlet extension to the two-Higgs doublet model (N2HDM): adds an extra real singlet to the 2HDM in the form of another CP-even scalar, for a total of six Higgs states.

- The Next-to-Minimal Supersymmetric Standard Model extension (NMSSM): extends the doublets of the 2HDM with a complex singlet superfield, generating an

additional CP-even and CP-odd boson, both electrically neutral. This model has a total of seven Higgs states.

All of the above models are compatible with the confirmed 125 GeV Higgs, but none of them have so far benefited from additional experimental evidence [10]. The sensitivity of VBS, in particular its longitudinal polarizations, to the underlying EWSB mechanism makes it a strong probe of potential BSM extensions to the Higgs sector.

As a relatively simple model, the 2HDM is a popular starting point used to identify BSM parameters. A major search for evidence in support of the 2HDM at the LHC focuses on heavy Higgs bosons that decay to lighter Higgs bosons and bottom quarks. Notably, the unitarity of the model imposes constraints on possible manifestation of BSM effects in VBS: tighter measurements of the HWW and HZZ couplings indicate smaller possible deviations from SM behavior [2]. The Type-II formulation of the 2HDM, identified by which symmetries the fermionic field Yukawa couplings follow, hinges on a total of six independent parameters:

- Masses of the scalar bosons: m_{h^0}, m_{h^1}, m_A, and $m_{H^\pm}$
- Mixing angle between the CP-even scalars, α
- Arctangent of ratio between the two vacuum expectation values, β

A 2020 study focusing on the W^+W^-jj leptonic and semi-leptonic VBS channels demonstrates a method that is sensitive to a wide range of anomalies in polarizations and kinematic distributions within the Type-II 2HDM framework. The method applies a classification network and autoencoder to reduce the parameter space to three dimensions without a significant drop in discrimination power; the abilities to separate W^+W^-jj from other processes, and to distinguish polarization states within W^+W^-jj, are maintained [20].

Along with extensions to the Higgs sector, a larger VBS cross section may be indicative of new resonances [20]. In the context of particle physics, a resonant state is the signature of an unstable heavy particle, rapidly decaying into lighter components. Resonances form peaks in experimental data, with the peak's full width at half-maximum (FWHM) corresponding to the decay width $\Gamma \equiv \tau^{-1}$, where τ is the mean lifetime of the heavy particle [8] [29]. Different models propose resonances close to the TeV-scale kinematic

limits of the LHC, including heavy vector W' and Z' bosons or the scalar radions and tensor bulk gravitons of the Randall-Sundrum model. Constraints on the parameters of these models are computed by choosing reasonable permutations of the parameters and simulating events to produce the desired resonances [2] [30].

Much like the aforementioned extended Higgs sectors, diboson production at the LHC may be used to probe for electroweak extensions. Simply, the W' and Z' particles are heavy spin-1 resonances that decay to the SM vector bosons:

$$pp \rightarrow \text{Z'} \rightarrow \text{WW} \qquad pp \rightarrow \text{W'} \rightarrow \text{WZ}$$

$$pp \rightarrow \text{Z'} \rightarrow \text{ZH} \qquad pp \rightarrow \text{W'} \rightarrow \text{WH}$$

Research on these diboson final states can be used to constrain the masses and mixing parameters of the heavy primed resonances, even with data below their mass scale [31]. The W' and Z' bosons are included in the Heavy Vector Triplet (HVT) model, which is determined by four parameters: mass of the W' and Z' resonance m_V, scalar c_F controlling the coupling of the extra bosons to fermions, scalar c_H controlling the coupling of the extra bosons to the Higgs and longitudinally polarized SM weak bosons, and the strength of the extra interaction g_V. Current studies constrain the masses of HVT vector bosons by a lower limit of 4.8 TeV [30].

The Randall-Sundrum (RS) model seeks to explain the hierarchy problem of why the Planck scale ($M_{\text{Pl}} \sim 10^{15}$ TeV) is so many orders of magnitude greater than the electroweak scale ($\rho \sim 1$ TeV). It describes the spatial extent of the universe as Anti-de Sitter (AdS) space and proposes a warped additional spatial dimension, accounting for the hierarchy gap by changing the spacetime metric exponentially along its curvature [32]. The extra dimension is accompanied by a tower of unevenly-spaced Kaluza-Klein (KK) resonances, with the lightest being on the order of the electroweak scale ρ and the energy range 1 TeV $\ll m_{\text{KK}} \lesssim 100$ TeV remaining a region of interest [33]. Extra-dimensional theories are attractive because of their potential to solve the hierarchy problem without the need for supersymmetry [34].

Radions, also referred to as graviscalars, are among the lightest proposed BSM particles [33]. In the RS model, fluctuations in distance between two branes, one playing

host to the SM fields and the other to gravity, manifest as the massive radion field; the fluctuations correspond to warping in the additional space dimension [32]. The independent parameters for the radion model are the compactification radius r_c and high-energy threshold Λ_R of the theory. Radions, if they exist, have been experimentally excluded from having masses below 1.4 TeV, identifying a slightly more stringent lower mass limit in the search for BSM states [30].

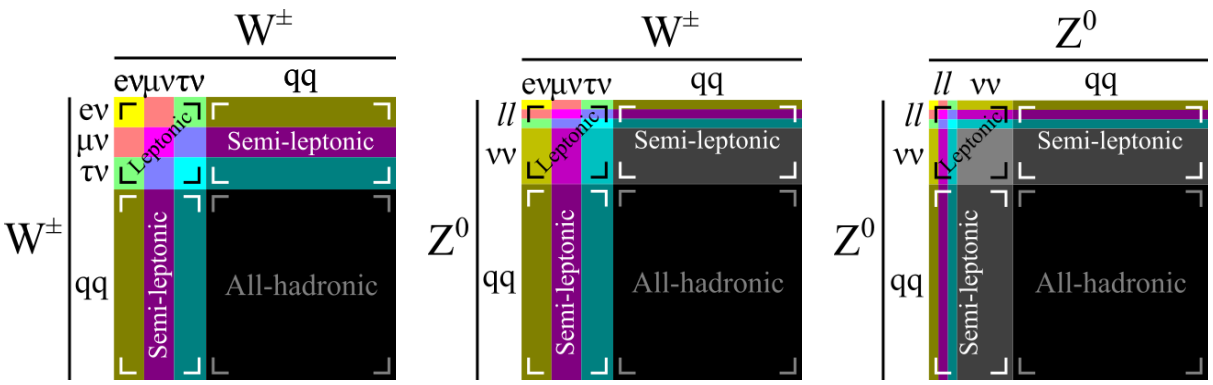
Gravitons in the RS model are unique in that they are much more likely to be found in a separate brane from the SM fields; gravitons can propagate through the bulk to the other brane, resulting in a "gravitational redshift", but the SM fields are restricted to it [13]. They are also heavy, unlike the hypothetical SM graviton. The weakness of gravitational attraction compared to the other fundamental interactions is explained by its strength being diluted over an additional dimension [32]. The parameters characterizing RS gravitons are their masses $m_{G_{\text{bulk}}}$ and the ratio $\tilde{\kappa} \equiv \frac{\kappa\sqrt{8\pi}}{M_{\text{Pl}}}$, where κ is the curvature scale of the additional space dimension and M_{Pl} is the Planck mass. The curvature ratio $\tilde{\kappa}$ also influences the decay width of the graviton. Currently, the most stringent lower mass limit for bulk gravitons is 2.7 TeV [30].

A 2023 search for heavy resonances based on 138 fb^{-1} of CMS Run-2 data taken between 2016 and 2018 targeted new decays to the following massive boson pairs: WW, WZ, ZZ, WH, and ZH. The phase space was limited to the all-hadronic final state, much like this thesis. Investigating the mass range between 1.3 and 6 TeV, the study found no significant difference from the SM background; at most, the mass peaks at 2.1 and 2.9 TeV corresponded to a local significance of 3.6σ and a global significance of 2.3σ. As a result of the findings, it has been concluded with 95% confidence that heavy resonance production by vector boson fusion (VBF) can have a cross section no greater than 0.1 fb [30].

2.4.2 Decay channels

As any vector boson in a VBS process can decay into either a pair of leptons or a pair of hadrons, there are three possible combinations, each of which defines a decay channel. All of the weak bosons are more likely to decay into hadrons; the proportion of events that proceed hadronically is $(67.41 \pm 0.27)\%$ in the $\text{W}^{\pm}$ case and $(69.911 \pm 0.056)\%$ in the Z^0 case [13]. This means that the ranking of VBS final states in increasing order

of commonness is two leptons, one lepton and one hadron, and two hadrons. Figure 5 presents a graphical comparison of each channel's branching ratios.



$W^{\pm}W^{\pm}$, 45.3% all-hadronic $W^{\pm}Z^{0}$, 47.1% all-hadronic $Z^{0}Z^{0}$, 48.9% all-hadronic

Figure 5: Phase space diagrams showing the possible vector boson decays in the $W^{\pm}W^{\pm}$, $W^{\pm}Z^{0}$, and $Z^{0}Z^{0}$ channels; *ll* refers to the charged leptons e, μ, and τ, as listing each individually would require a font too small to be legible [13] [16].

Fully leptonic processes are by far the most extensively researched type of VBS even despite their rarity. The four-lepton $pp \rightarrow W^{\pm}W^{\pm}jj$ decay channel was the first VBS process to be observed in Run 1 of the LHC; it has been referred to as a "golden channel" due to its clean signature, that is, its high signal-to-background ratio. Said background is dominated by top-antitop ($t\bar{t}$) processes, which have a cross section around a hundred times larger than that of VBS [20]. The $pp \rightarrow Z^{0}Z^{0}jj$ decay channel is similarly clean, but suffers from a smaller cross section [5]. A common downside to the fully leptonic decays is that they operate in a lower-energy regime, below the diboson invariant mass, such that they cannot provide limits on the energy scale of EFT operators, which makes the other channels attractive targets for exploration [2].

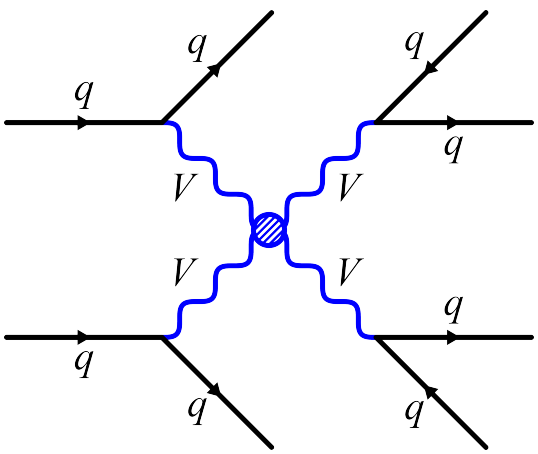
Semi-leptonic decays constitute the middle ground, lying between the fully leptonic and fully hadronic channels not only because of the particles in the final state; they have lower statistics and higher backgrounds in LHC data than the leptonic channel, but higher statistics and lower backgrounds than the hadronic channel. As in fully leptonic decays, the background is formed mainly by $t\bar{t}$ processes [20]. The semi-leptonic channel in particular is the strongest probe of dimension-8 EFT operators. Study of the semi-leptonic channel benefits from "jet pruning" methods to distinguish the electroweak-based signal from quantum chromodynamics (QCD)-based background, although these methods have not been made compatible with CMS's anti-k_T reconstruction algorithm [3].

A major advantage of the all-hadronic channel is its high branching ratio: almost half of all VBS events at the LHC are fully hadronic, especially useful when studying the rare case in which both vector bosons are longitudinally polarized. Another is the fact that all four decay products (and their kinematics) are visible, unlike in the leptonic channels, where neutrinos escape undetected, making it difficult to recover the polarizations of the parent vector bosons. At higher energies, where BSM physics are more likely to have an influence, a stronger forward boost causes both quarks resulting from a vector boson decay to coalesce to a single, larger jet; a detector would then register the signature of this process as two wide central jets along with two forward jets from the quarks that initially emitted the vector bosons travelling in their altered trajectories, strongly separated in pseudorapidity [5] [15]. However, the notable disadvantage that has often prevented earlier analysis is the poor signal-to-noise ratio due to high QCD multi-jet backgrounds.

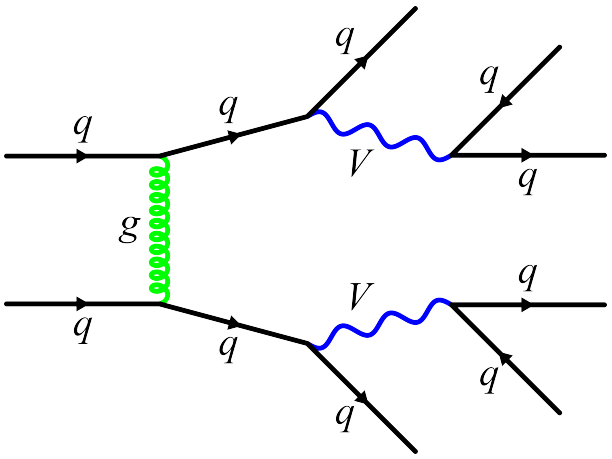
2.4.3 Sources of background

An overbearing background exists for all VBS channels [16]. It is caused by QCD processes, which in Run 2 had cross sections of orders over 10^5 pb, at least eight orders of magnitude larger than VBS at around 10^{-3} pb. This background does not involve vector bosons, but produces multi-jet final states that are difficult to distinguish from those of VBS processes. The background is strongly suppressed in the $pp \rightarrow W^{\pm}W^{\pm}jj$ channel, in which the vector boson decay products do not result in jets, and the only hadronic components are the quarks used for tag jets. However, other backgrounds involving diboson production and top decays to the weak bosons may mimic VBS signatures, contributing to all channels and accounting for up to 10^3 pb of irreducible background in Run 2 [26].

Furthermore, it is also important to distinguish two different types of processes described as VBS, depending on whether they are caused by the electroweak (EWK) interaction or the QCD interaction [2]. Unlike in the electroweak case, the two vector bosons in QCD VBS do not interact with each other. Such processes are of less interest and may present an unwanted, irreducible background, as they are not sensitive to potential BSM physics; for example, the absence of vector boson co-interaction means that quartic couplings are unable to be probed for anomalies. This difference is illustrated with LO Feynman diagrams in figure 6.



Electroweak-induced diboson event.



QCD-induced diboson event.

Figure 6: Comparison of diboson events generated through electroweak and QCD radiation. Notably, in the QCD case, the vector bosons do not actually interact with each other.

Interference between the electroweak and QCD modes of production must be taken into account. In total, the sources to be considered in a VBS analysis contribute with orders

- Electroweak VBS signal: $\mathcal{O}(\alpha^6)$
- Interference: $\mathcal{O}(\alpha_s\alpha^5)$
- QCD background: $\mathcal{O}(\alpha_s^2\alpha^4)$

where the fine structure constant α and strong coupling constant α_s are related by $\alpha = \alpha_s/137$ [7]. Differences in the kinematics of the electroweak and QCD production modes make it possible to identify regions in the parameter space where the signal is amplified. For instance, electroweak production favors large rapidity separations Δy_{jj} between the jet quarks and larger masses M_{jj}; for the background, the opposite is true [35].

Kinematic cuts are a leading strategy to distinguish signal from background; differences in their QCD structures cause the two types of events to cluster in different parts of the phase space [7]. For the fully leptonic channel, the cross section of the QCD background is 2.12 fb to the electroweak signal's 1.10 fb, or 192% of the signal. Cuts placed at $M_{\mathrm{jj}} > 600$ GeV and $|\delta y_{\mathrm{jj}}|$ were found to reduce the QCD cross section to 7.4 ab, while the signal was reduced to 0.201 fb; with cuts applied, the QCD cross section equals only 3.7% that of the signal [35]. A goal of our research is to identify similar cuts for the all-hadronic channel.

2.4.4 Polarization

A particle is described as longitudinally polarized when its spin lies along the direction in which it moves, and transversely polarized when its spin is perpendicular to that direction. Transverse polarization is either right-handed or left-handed, with the two states related by a reflection across the direction of motion. Performing this same reflection for a longitudinally-polarized particle is equivalent to a mere rotation in the plane perpendicular to the momentum vector, so there is only one distinguishable longitudinal state. Choosing the direction of movement to lie along the z-axis, the normalized polarization four-vectors for the three states are

- Longitudinal: $\varepsilon_L^\mu = \frac{1}{m}(p_z, 0, 0, E)$
- Right-handed: $\varepsilon_+^\mu = \frac{1}{\sqrt{2}}(0, 1, i, 0)$
- Left-handed: $\varepsilon_-^\mu = \frac{1}{\sqrt{2}}(0, 1, -i, 0)$

It should be noted that the longitudinal polarization four-vector requires division by the particle's mass m, which is unphysical for particles with mass 0. As a result, the massless photons and gluons can only be polarized in the right- or left-handed states, and not longitudinally, to maintain gauge invariance [3]. A somewhat metaphorical visualization of these phenomena is presented in figure 7.

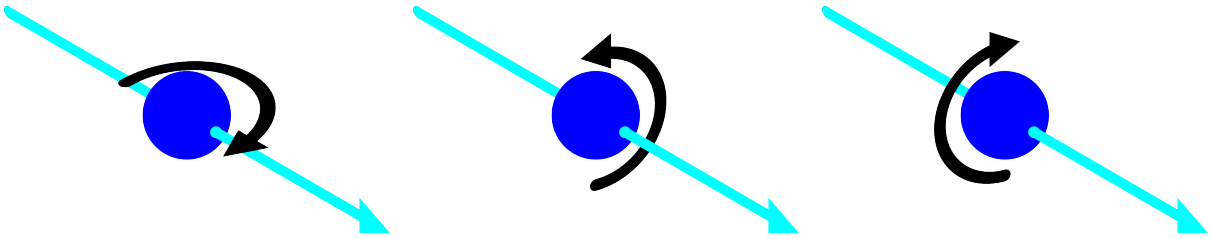


Figure 7: Conceptual representation of possible polarization states in massive vector bosons. In this model, where the boson is analogous to a thrown, spinning ball, longitudinal polarization occurs when the axis around which it rotates is perpendicular with the direction of motion; transverse polarization occurs when that axis and direction of motion coincide.

The polarization states may also be described in terms of helicity, a pseudoscalar quantity manifesting as the projection of the particle's spin onto its direction of motion. Transverse polarizations correspond to helicities of $h = +1$ (right-handed) and $h =$

-1 (left-handed), while longitudinal polarization corresponds to a helicity of $h = 0$ [3]. The weak bosons are always found with one of these three helicities. In the context of weak VBS, anomalous effects are more apparent for longitudinally polarized vector bosons, due to their origin being the EWSB mechanism, than for transversely polarized ones. Hence, discrimination between the two types of polarized states is a key effort within VBS research. This thesis aims to investigate whether requiring the massive vector bosons to be longitudinally or transversely polarized changes the kinematic topology of scattering events. The different polarization states interfere with each other, and cancel out completely when all azimuthal angles are integrated over. This needs to be performed with MC data due to the lack of LHC statistics, as well as due to the cancellation being disrupted when kinematic cuts are applied [7] [20].

The polarization of a vector boson is encoded in the kinematics of its decay products [24]. In LHC data, this information can be extracted from the detected decay products to chart the polarization fraction of the process [27]. In the case of fully leptonic decays, most of the bosons are known to be transversely polarized [36]. For all-hadronic decays, there are two steps in obtaining the polarizations: the first is to separate the electroweak-induced signal from the QCD background, and the second is to separate longitudinal polarizations from transverse ones. This would test the Standard Model by checking the $\mathrm{W}_L\mathrm{W}_L$ cross section against theoretical predictions as well as surveying the helicity fractions of VBS events [5]. However, difficulty arises as hadronic decays tend towards the more highly boosted regime, where decay products have smaller angular separations, making them harder to resolve from each other [26]. This regime is crucial because it is thought to be more strongly influenced by BSM physics. Potential approaches to remedy this challenge include shrinking the radius of the subjets with increasing transverse momentum, or using a completely different method to deconstruct jets specifically in high-boost regions [36].

2.5 Vector boson fusion

VBF is a scattering process in which two energetic quarks emit vector bosons, which then make contact and fuse into a H^0 Higgs boson. If the H^0 decays back into two vector bosons, this kind of VBF is also a VBS process. Of the four main types of processes that produce observable Higgs bosons at the LHC, VBF is the second-most common, but still

an order of magnitude rarer than gluon-gluon fusion (ggF), the dominant form of Higgs production [37]. Feynman diagrams of all four types of event are presented in figure 8 for comparison.

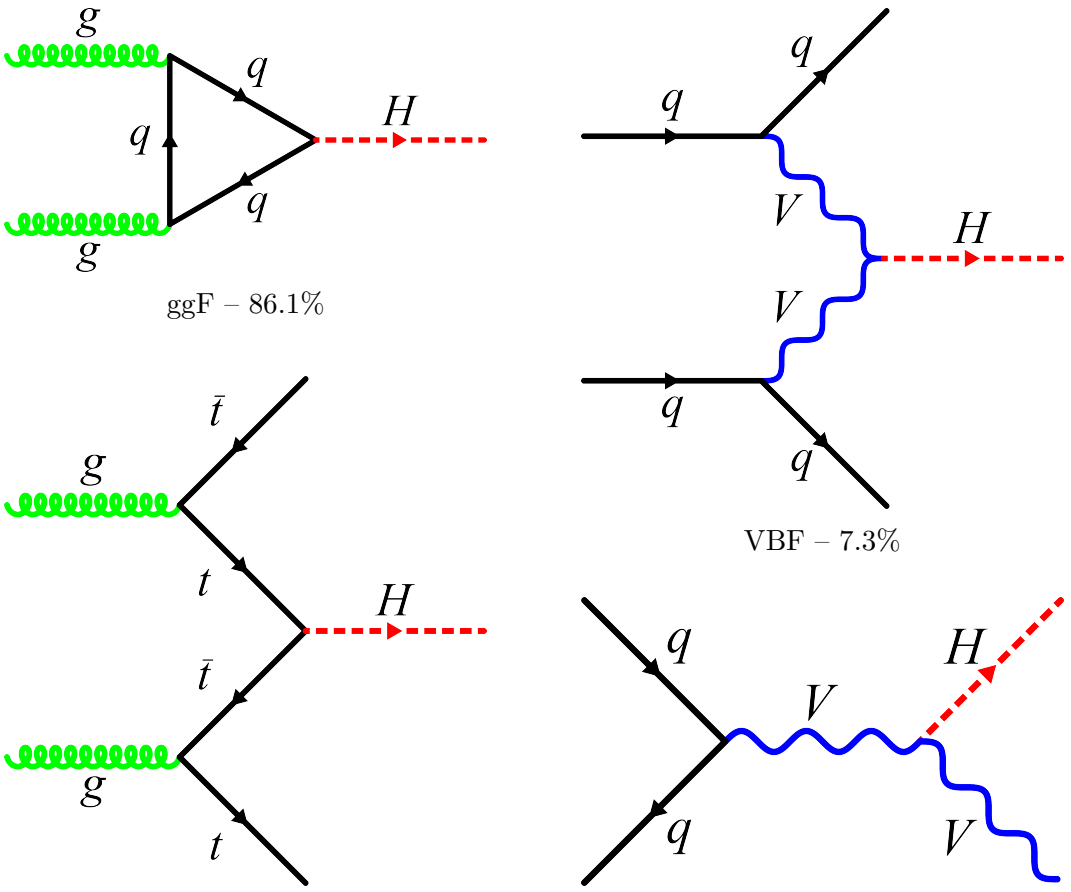
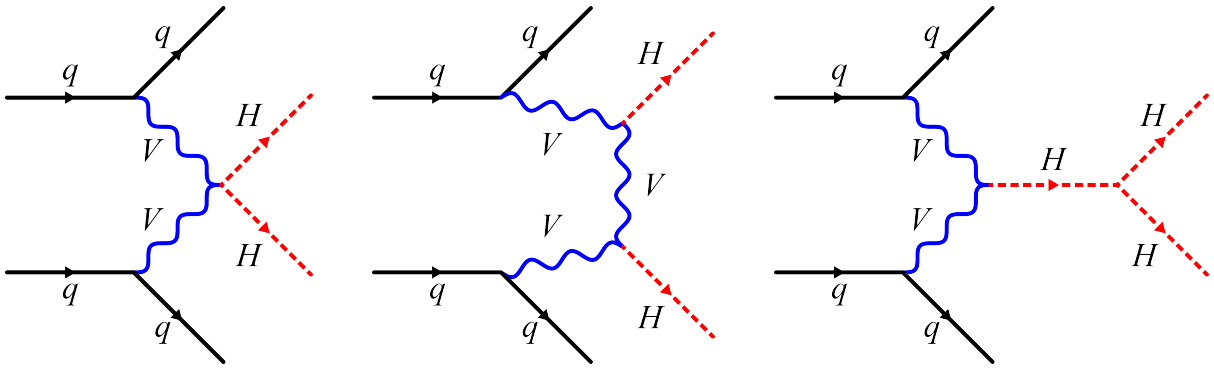


Figure 8: Feynman diagrams of the main processes responsible for Higgs production at the LHC, clockwise from top left in order of increasing rarity. Percentages in subcaptions indicate each process's fractional contribution to the total Higgs production [37].

These processes are key to studying potential BSM extensions to the Higgs sector; VBF and Higgs-strahlung would additionally be primary modes of Higgs production in e^+e^- collisions. Kinematic analysis of these processes relies heavily on the reconstruction of leading, subleading, and trailing jets.

VBF in particular is expected to be a strong probe of di-Higgs production [2]. One of its advantages over ggF is its sensitivity to BSM behavior connected to the electroweak symmetry, as new physics affecting the coupling between the Higgs and vector bosons will greatly enlarge the cross section of di-Higgs production as the invariant mass of the system m_{HH} increases [38]. The main contributions to di-Higgs production due to VBF are shown in figure 9.



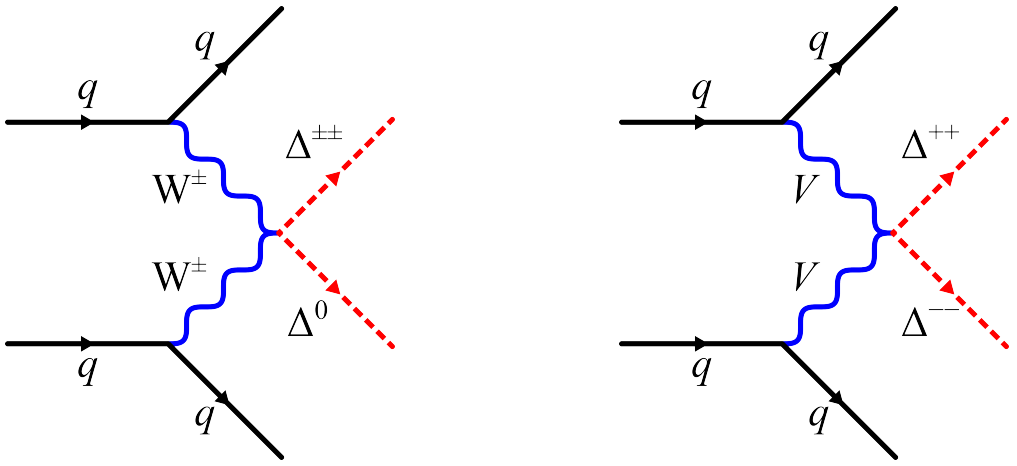
VV→HH coupling Vector-boson t-channel Trilinear Higgs coupling

Figure 9: Primary modes of Higgs pair production due to VBF at the LHC [38].

VBF can also be used to search for a specific type of DM candidate called the pseudo-Nambu-Goldstone boson (pNGB) [2]. A pNGB would be a weakly interacting massive particle (WIMP), compatible with the Lambda cold dark matter (ΛCDM) model, a leading cosmological model uniting regular matter, dark matter, and dark energy (responsible for driving the continued expansion of the universe and represented by the cosmological constant Λ). DM in the model is described as "cold" because it does not propagate at relativistic speeds. WIMPs are so called because they are permitted to interact only up to the weak scale, consistent with DM having well-documented gravitational effects (gravity being weaker than the weak interaction) but giving off no light (electromagnetism being stronger). There are two scenarios introducing pNGBs from BSM physics. The first is accomplished by adding a new complex scalar field which is invariant under the U(1) gauge group, and when the U(1) symmetry is broken by the addition of a mass term, the angular mode of the scalar field becomes the pNGB [39]. In the second, a real scalar gauge field η is instead added that obeys parity symmetry; the composite particle of a light Higgs particle and the gauge singlet η is then the pNGB. This model addresses the naturalness of light scalar bosons comprising DM by presenting pNGBs as strongly coupled composite particles [40]. The significant cross section of VBF when compared to other modes of Higgs production makes it a crucial target for experimental studies of pNGBs.

VBF is also in use in searches for the doubly charged Higgs [2]. This particle is included in the model of the Type II Seesaw mechanism, a gauge-invariant approach intended to account for the vanishingly small masses of neutrinos and their oscillation. Unlike the Type I Seesaw mechanism, it does so without necessitating the presence of

right-handed neutrinos, for the existence of which there is otherwise no evidence. The Type II Seesaw mechanism predicts EWSB and SM mixing to produce the following seven states: two uncharged CP-even particles h and Δ^0, an uncharged CP-odd χ, the singly charged $\Delta^{\pm}$, and the doubly charged $\Delta^{\pm\pm}$. These extensions to the Higgs sector are probed effectively by $W^{\pm}W^{\pm}$ VBF due to their coupling weakly to the SM fermion fields [2] [41]. VBF contributions to hypothetical doubly charged Higgs pair production are shown diagramatically in figure 10.



(a) $W^{\pm}W^{\pm}$ fusion (b) Electroweak boson fusion

Figure 10: Potential doubly-charged Higgs pair production in VBF processes [41].

3 Experimental science

3.1 Accelerator physics

3.1.1 Proton-proton collisions

Each proton is a chromodynamic system of quarks and gluons, collectively known as partons. A parton distribution function (PDF) is used to describe the interactions between these particles; it corresponds to the probability distribution of finding a parton within a proton given the fraction x it carries of the proton's four-momentum at a scale of Q^2 [42]. It cannot be predicted theoretically due to the strong interaction's large coupling constant α_s, and instead must be drawn from experimental results [8]. PDFs are different depending on the type of collisions being considered, but those for protons are the most well-researched [42]. The structure of a hadron is further complicated by a sea of virtual partons: in addition to the three valence quarks and their gluonic bindings,

there are quark-antiquark pairs and gluons that are constantly appearing and disappearing within the hadron's interior. While high values of x suppress the production of sea quarks, at low values of x, they are instead the dominant contribution to the proton's hadron's momentum. Taking all of this into account, PDFs allow for the calculation of hard scattering cross sections in proton-proton collisions. The use of PDFs is also crucial in simulations, where they inform the underlying event and initial-state parton showers. We use the Les Houches Accord parton distribution function (LHAPDF) in our event generation parameters.

The LHC is a proton-proton collider occupying the tunnel originally constructed for the Large Electron-Positron (LEP) collider, 27 km in circumference. During a run of the LHC, bunches of protons are accelerated in opposite directions around the collider, which defines the x-z plane in the CMS and ATLAS coordinate systems. Each bunch contains on the order of 10^{11} protons [43] [44]. The accelerating beams are not continuous, with spacings of 25 ns between bunches [43]. Shorter bunch spacings help to reduce the amount of pileup, or unwanted secondary interactions [45]. Once they achieve $\sqrt{s} = 13.6$ TeV of centre-of-mass energy, corresponding to speeds within 1 ms^{-1} of the speed of light, the bunches are allowed to collide. The cross section of proton-proton collisions is still rather small, so few bunches actually collide per crossing; the others continue around their orbits until they reach the interaction point again every 89.1 μs. The event rate is the product of the collisions' luminosity and cross section [43]. This high-energy regime is the perfect environment in which to study hadronic VBS.

Another type of hadron collider is the proton-antiproton collider, such as the now-decommissioned Tevatron at the Fermi National Accelarator Laboratory, so named because it was the first particle collider to cross the centre-of-mass energy threshold of 1 TeV. Notably, Tevatron experimentally confirmed the top quark in 1995 [46], elucidating its Yukawa coupling to the Higgs field [47]; this, in addition to its precision measurements of the W mass [8], has helped to further cement the Standard Model as the leading fundamental physics paradigm. However, no current proton-antiproton colliders are able to achieve energies high enough to be of interest to VBS research [43]. Additionally, many leading future accelerator proposals are oriented towards accommodating electron collisions with attention having turned towards EWSB and the Higgs sector.

In comparison to hadron colliders, electron-positron colliders are able to measure the

couplings of the Higgs field to the fermion and gauge fields due to a more favorable signal-to-background ratio; however, they are unable to reach the same high centre-of-mass energies at comparable circumferences due to the electron being some three orders of magnitude lighter than the proton [8] [45]. Future electron-positron colliders—proposals include the Compact Linear Collider (CLIC) and International Linear Collider (ILC), as well as an e^+e^- component to the Future Circular Collider (FCC)—have strong prospects as Higgs factories, the next major collider goal after the discovery of the Higgs boson [48]. Unlike in proton-proton collisions, where gluon-gluon fusion is the main mode of Higgs production, in electron-positron colliders, Higgs-strahlung is the main mode at low energies, while VBF takes over at higher energies, where the incoming leptons are more likely to radiate out weak bosons [49]. This makes electron-electron colliders promising for the study of VBF, although the low centre-of-mass energies proposed for future colliders of this type mean the LHC—especially with its High Luminosity upgrade—will continue to be at the forefront of VBS research. In the context of event generation, hard scattering data from electron-positron colliders can be used to find fragmentation functions when simulating jet behavior [50].

In summary, the record centre-of-mass energy and integrated luminosity provided by the LHC remain unmatched. Unlike proton-antiproton colliders, the LHC is also available for the study of heavy ion physics, data from which can also be applied for fragmentation functions [50]. It outclasses in its kinematic reach specialized heavy-ion colliders such as the Relativistic Heavy Ion Collider (RHIC) [51]. For VBS processes in particular, the data collected by the LHC's CMS and ATLAS detectors is the only data with sensitivity to BSM signals. The unparalleled amount of information produced by the LHC comes from three data-taking periods, or runs, with ever-increasing centre-of-mass energy $\sqrt{s}$ and integrated luminosity $\mathcal{L}$.

- Run 1 (2011-2012): $\sqrt{s} = 7\text{-}8$ TeV, $\mathcal{L} \sim 30$ fb^{-1}
- Run 2 (2015-2018): $\sqrt{s} = 13$ TeV, $\mathcal{L} > 150$ fb^{-1}
- Run 3 (2022-2022): $\sqrt{s} = 13.6$ TeV, $\mathcal{L} > 300$ fb^{-1}

Before and after each Run is an intermediate period of long shutdown, during which systems are maintained and upgraded; the accelerator is in LS3 at the time of publication of

this document [52]. The next periods of data-taking will begin after the High-Luminosity Large Hadron Collider (HL-LHC) upgrade is completed, promising to achieve energies of 14 TeV and integrated luminosities of 3-5 ab^{-1} [24] [53]. Bunch spacing will remain at 25 ns, nevertheless leading to a far greater data-taking rate due to the higher luminosity. This massive jump in radiation exposure will demand improved hardness in the various LHC detectors, including CMS [54].

3.1.2 CMS detector

The primary instrument of the CMS collaboration is the detector of the same name, a layered, 12500-ton cylindrical device 21.6 m in length and 14.6 m in diameter. The detector is designed to allow reconstruction of events with particle flow (PF) algorithms that track particles as they move away from the beam interaction region [55]. Each type of layer is intended to trace the path followed by a different type of particle; from this information, the physics of the collision can be inferred. The main cylindrical part of the detector is called the barrel, and it is closed at each face by the endcaps, which have different types and arrangements of detector modules. A transverse view of the detector barrel is presented in figure 11.

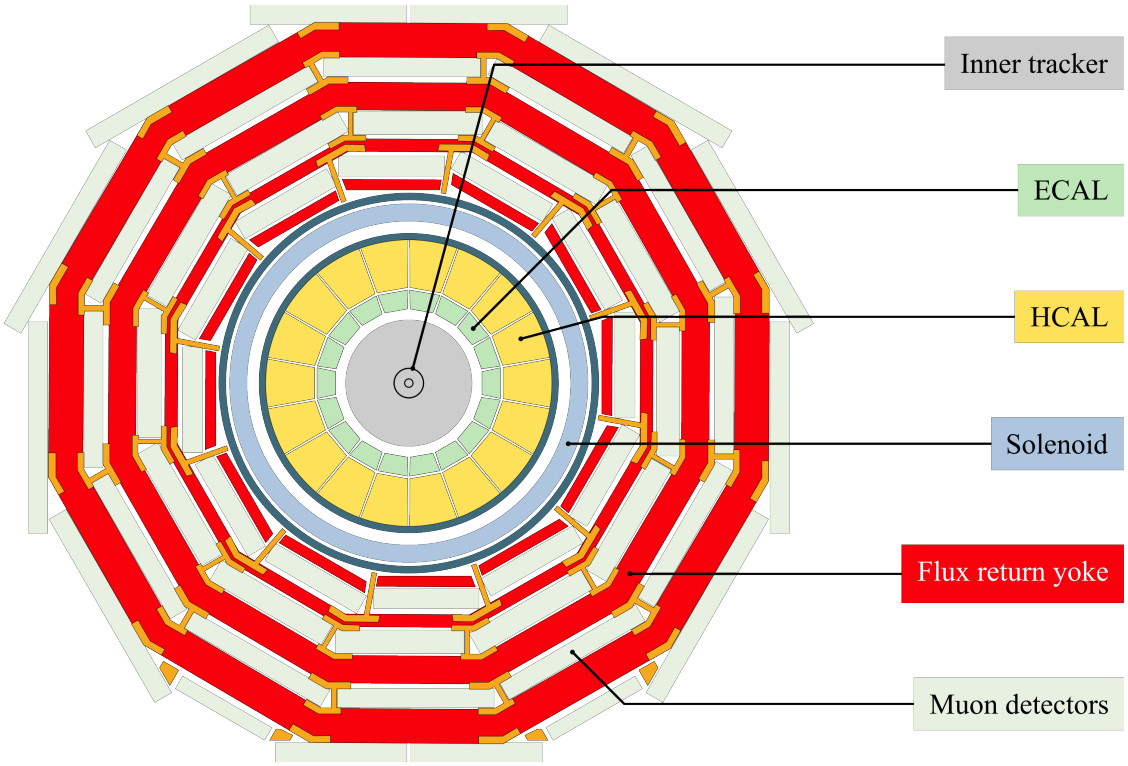


Figure 11: Cross section of the CMS detector perpendicular to the barrel axis [55].

The purpose of each layer follows:

- The inner tracker is a narrow tube surrounding the 23-mm beam pipe [56]. It is composed of inner layers of silicon pixels and outer layers of silicon strips, offering extremely fine particle trajectory information as close to the beam region as possible. The tracker is explained more thoroughly in its own section.

- The electromagnetic calorimeter (ECAL) has a pseudorapidity coverage of $|\eta| < 1.479$ in the barrel and $1.479 < |\eta| < 3.0$ in the endcaps. It is built with lead tungstate, and is capable of detecting photons, electrons, and both electrically neutral and charged hadrons; it is long enough to capture over 98% of the energy of electrons and photons below 1 TeV and triggering showers for about two thirds of hadrons. ECAL's high granularity leads to a positional resolution of 5 cm, allowing individual particles comprising a jet to be identified. The endcaps of ECAL also contain specialized, finer subdetectors called preshowers, made out of lead and silicon; the lead is intended to trigger showers from photons and electrons, which are then immediately detected by the silicon layer. However, due to high systematic uncertainties, data points from the preshowers are not analyzed by themselves; they are either added to nearby ECAL clusters, or ignored completely if none exist [55].

- The hadron calorimeter (HCAL) covers pseudorapidities of $|\eta| < 1.3$ in the barrel and $1.3 < |\eta| < 3.0$ in the endcaps. It is a structure of brass and plastic designed to detect exiting hadrons; the 4-mm plastic tiles act as scintillators, generating light which is absorbed as a signal by the brass [57]. Additional absorbers for HCAL are installed outside of the solenoid, as well as a 20-mm layer of steel to record low-pseudorapidity ($|\eta| < 0.25$) events [55]. However, due to a larger cell size, it does not offer the same high resolution as the layers interior to it, but is nevertheless fine enough to distinguish neutral hadrons from charged ones [30]. Forward calorimeters, made out of steel plating and quartz fibres, extend coverage into the forward region, detecting particles with pseudorapidities up to $|\eta| < 5$ [55] [58].

- The superconducting solenoid is a large cylindrical magnet, with an inner diameter of 6 m [59], generating a 3.8 T magnetic field [30]. The presence of this magnetic field immediately allows electrically charged particles to be distinguished from neutral ones depending on whether or not their trajectories are deflected by the field.

Furthermore, the amount of deflection can be used to identify the momenta of the particles; a particle carrying a higher momentum will suffer less deflection than one with a lower momentum. The solenoid's influence extends across the entire transverse cross section of the detector, producing oppositely-oriented fields of 3.8 T inside the magnet and 2 T outside it; as a result, the direction in which a charged particle's trajectory is bent inverts as it passes through the solenoid [55].

- The muon detectors are ionized gas chambers embedded in a steel flux-return yoke. There are three layers of steel and four of the detectors: drift tubes are used in the barrel region and cathode strip chambers in the endcaps. These detectors are specifically designed to reconstruct the paths of any muons which have managed to pass unopposed through all of the interior layers; muons with momenta up to around 200 GeV are registered by the inner tracker. In addition to acting as the flux return, the steel yoke is the absorber for the muon detectors and the frame of the entire device, accounting for a hundred thousand tons of its weight [30].

Due to the unstable nature of vector bosons as heavy resonances, it is impossible to analyze them directly; rather, their decay products must be studied [20]. All of the information about a scattering process must be gleaned from its observed final state, which may contain jets, photons, and charged leptons; additionally, the absence of expected momentum is evidence of neutrino production [7].

3.1.3 CMS inner tracker

The CMS tracker is the innermost layer of the instrument, fit around the beam axis and hence nearest to the interaction region. Steeped in the magnetic field generated by the solenoid, the inner tracker allows measurement of the momentum and electric charge of particles right as they leave the beam pipe, data which is used to reconstruct the vertices from which they originated [55]. All of the sensitive modules it contains are made out of silicon; when a particle passes through a module, it ionizes the doped bulk material, creating electron-hole pairs that are separated and drawn to either end to produce a signal [60]. The tracker is able to register jets with transverse momenta up to $p_T \sim 1$ TeV [55]. Measuring 5.6 m along its axis and 1.2 m in diameter, it has three active regions: the barrel and an endcap on either side, with different arrangements of detecting modules.

Four layers of silicon pixels and ten layers of silicon strip detectors constitute the barrel. In each endcap, there are instead three layers of pixels and 12 layers of strips [56]. Pixels are finer than strip detectors; the closer a layer of CMS is to the beam, the greater its fidelity, achieved with a finer grid made up of a higher number of modules. The machine has a total of 66 million pixels and 9.6 million strips [55].

A pixel module is designed around an n-type substrate, with more heavily n-doped implants on one side and a heavily p-doped base on the other, creating a p-n junction. This is known as an n-in-n sensor. The modules used in the barrel are known as barrel pixels (BPIX) and those in the endcap disks as forward pixels (FPIX). The BPIX are positioned in four concentric rings at 29, 68, 109, and 160 mm of radial distance from the interaction region. The FPIX are placed at 291, 396, and 516 mm from the centre of CMS. Together, they provide a net surface area of 1.9 m^2. In this configuration, the coverage of the pixel detectors extends over the pseudorapidity range $|\eta| < 2.5$. Both types of pixels are held in graphite and carbon-fibre frames, and receive cooling through CO_2 carried by stainless-steel tubes, necessary for operation at -20 °C [54] [56]. they differ primarily in shape, optimized for arrangement either along the length of the cylinder or into disks at either end [56].

Unlike pixels, the operating principle for a silicon strip involves the p-in-n approach: p-n junctions are formed by heavily p-doped implants in an n-type bulk, mounted to a heavily n-doped aluminum backplane [60]. In the barrel, the ten layers of strips are split into four layers of tracker inner barrel (TIB) strips and six layers of tracker outer barrel (TOB) strips. Perpendicularly oriented relative to the barrel are the tracker inner discs (TID) and tracker endcaps (TEC). The different regions each have specially shaped components in order to maximize detector coverage. Rectangular shapes are used for the longitudinally-oriented barrel strips, while the endcaps and discs use wedge-shaped strips. The areas covered by the layers are designed to overlap for more accurate track reconstruction [60]. The diagonally nested arrangement is presented in figure 12. As with pixels, there is no strip coverage for pseudorapidities in the region $|\eta| > 2.5$, with extended coverage being a priority for the upcoming upgrade [61].

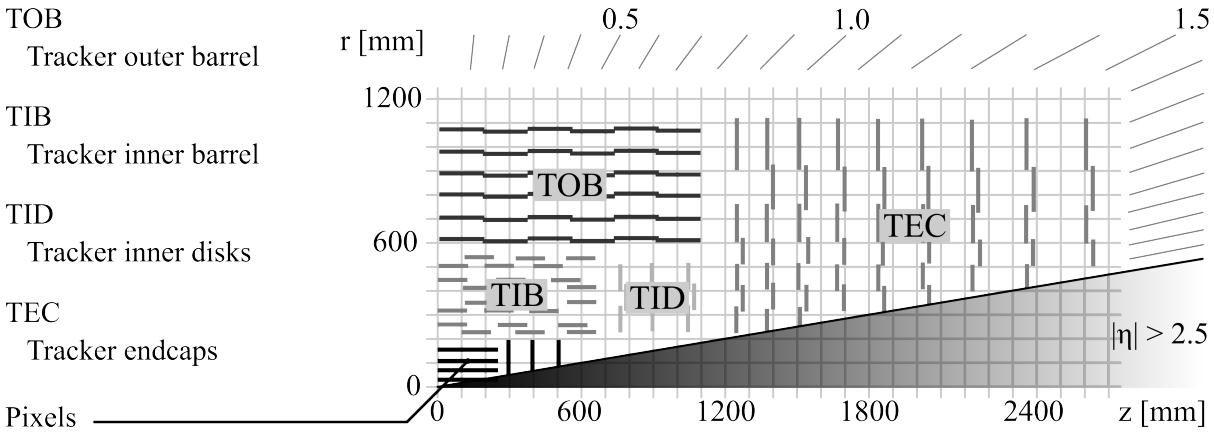


Figure 12: Pseudorapidity coverage of the various CMS inner tracker elements [61].

The greater HL-LHC upgrade taking place after Run 3 is intended to improve the precision of SM measurements and allow for better investigation of the Higgs sector [62]. For CMS, this will include installation of the phase-2 inner tracker; it will be a completely new device taking the place of the current tracker, as the components currently present will not be able to survive the high radiation outputs of the next runs [54]. Doped silicon remains the favored material for the new pixels and strip detectors, but the phase-2 tracker is intended to be radiation-hard enough to require replacement only once in ten years, coping with the vast increase in luminosity [63].

Material considerations for the upgrade include using Magnetic Czochralski (MCz) silicon due to its higher amount of impurities compared to the float-zone (FZ) and deep diffused float-zone (ddFZ) methods; primarily oxygen, these impurities have been found to lead to stronger resistance to radiation damage. Additionally, whereas the current pixels and strips are n-in-n and p-in-n planar sensors, respectively, a switch to an n-in-p structure would lead to components that do not degrade as quickly, avoiding an additional contribution of noise that results in fake hits [64]. All of the previously mentioned combinations are examples of planar structures; in comparison to the planar design, 3D silicon offers better radiation resistance due to shorter charge collection distances, although this comes at increased cost per sensor. As a result, the 3D sensors are to be employed only in the regions with high fluence [62].

The tracker upgrade will bring with it notable improvements in almost all metrics. Most relevant for VBS is the extension of the accessible pseudorapidity region to $|\eta| < 4$. Pileup in this region will be mitigated by design in the layout of the modules, as well as more precise timing measurements and better granularity [52]. This is a critical

consideration as pileup will be increased to around 140-200 secondary events per bunch crossing [54], with a hit rate of 3.2 GHz/cm^2 up from 0.58 GHz/cm^2 [65]. The decrease in the size of the tracker modules by a factor of six will allow for finer granularity and hence improved resolution. A smaller amount of material constituting the structure of the tracker will reduce multiple scattering, also contributing to a better resolution. Finally, replacement will be made easy with the implementation of a simple installation and removal scheme [63].

3.1.4 CMS kinematics

The CMS coordinate system is defined by the following orthogonal basis, pictured in figure 13:

- The x-axis points towards the centre of the accelerator
- The y-axis points up out of the ground
- The z-axis is collinear with the beam pipe, counterclockwise

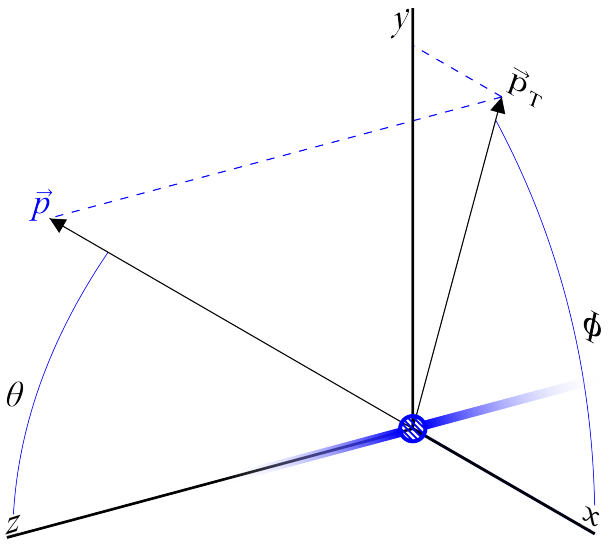


Figure 13: The coordinate system convention in use by the CMS experiment [58].

The following kinematic variables are easily extracted from ROOT data in order to plot angular distributions. First, the transverse momentum of a particle, $\vec{p}_T$, is defined to be

$$\vec{p}_T \equiv \sqrt{p_x^2 + p_y^2}$$

where p_i is the component of its momentum vector $\vec{p}$ parallel to axis i [8]. This is understood to be the projection of the particle's momentum vector onto the x-y plane. For two momentum vectors with the same magnitude, the one with the higher p_T is said to be more central, and the one with the lower p_T is more forward, corresponding to a higher-pseudorapidity region. Pseudorapidity, η, is the quantity

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right)$$

where θ is the polar angle the momentum vector makes with the z-axis, as shown in figure 13. Pseudorapidity is a more useful metric than the scattering angle θ due to its being invariant under Lorentz transformations, as it is often required to convert between the centre-of-mass and laboratory reference frames. Particles with low pseudorapidity are more transverse, that is, they are closer to the x-y plane; angular distributions favoring low pseudorapidities are described as being central. In contrast, particles travelling closer to the z-axis are described as being in the forward regions, with one travelling along the z-axis itself ($\theta = 0$) having a pseudorapidity of $\eta = \infty$. These particles are more difficult to study as the detector's coverage there is limited, which is always a prominent goal for improvement with every upgrade, as jets are approximately constant with pseudorapidity [8].

The azimuthal angle ϕ is the angle the transverse momentum vector $\vec{p}_T$ makes with the x-axis; angular distributions of ϕ are expected to be uniform, as particles are emitted from the interaction point isotropically with respect to the x-y plane. The final element of the ROOT vector is the mass m of the particle, useful for invariant mass calculations or the study of jets.

3.1.5 Jets

Jets are physics objects corresponding to conical bursts of hadrons and photons originating from a common origin, such as a high-energy quark. The mass of a jet corresponds to the invariant mass of the system of particles comprising it, and in high-energy regimes, is small compared to its energy [8]. Jets are the only way that quarks can be observed in detector experiments, as they immediately hadronize due to color confinement [66]. In CMS, about 65% of jets have sufficiently low momenta, $p_T < 1$ TeV, to be reconstructed by the inner tracker; another 25% are accounted for by ECAL and the final 10% by

HCAL. The capability to extract the constituents of a jet depends on the granularity of the detector; CMS's layers are as fine as possible so that multiple particles are less likely to be counted by a single cell as one particle [55].

Jet clustering is the process of classifying nearby particles registered by the detector into groups based on which jet they belong to. The common types of algorithms for this purpose are hierarchical (or sequential recombination) algorithms and cone algorithms [67]. Cone algorithms identify jets by placing particles into regions with area πR^2 in the $\eta \times \phi$ space, where R is the hyperparameter defining the radius of the cone, and the cone axis lies along the centroid of all the particles within it, weighted by transverse energy $E_T = E \sin\theta$ [66] [68]. High-energy particles are selected as seeds to form viable cones. However, this procedure does not restrict cones from overlapping, which has the downside that some particles will be counted as belonging to more than one jet. Additionally, cone algorithms bring with them an unwanted sensitivity to soft radiation at next-to-next-to-leading order (NNLO), meaning that particles belonging to two different jets may be counted as belonging to the same one [68].

Hierarchical algorithms, on the other hand, work by successively grouping pairs of particles or pseudojets by their relative transverse momenta. They hold advantages over cone algorithms by ensuring that particles are grouped as belonging to no more than one jet [68] and dropping the assumption that every jet can be modeled by a precise cone, an approach that includes unphysical constants in theory [69]. A generalization of some popular hierarchical algorithms is set using the following distance parameters:

$$d_{ij} \equiv \min\left(p_{Ti}^{2p}, p_{Tj}^{2p}\right) \cdot \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \qquad \text{and} \qquad d_{iB} \equiv p_{Ti}^{2p}$$

where y_i is the rapidity of particle i. Here, p is not the momentum but a parameter identifying the type of algorithm in effect. First, all the distances are calculated between each of the particles (d_{ij}, distance between particles i and j) and between each particle and the beam axis (d_{iB}). If the nearest distance is a d_{ij}, the elements i and j are grouped as one element for subsequent steps. If it is a d_{iB}, element i is understood to be a jet and excluded for subsequent steps. This repeats until all elements have been identified as parts of a jet (inclusive clustering) [67] or until a desired number of jets is reached (exclusive clustering) [69].

The case $p = 0$ is named the Cambridge-Aachen algorithm. It simplifies the distance

parameters to

$$d_{ij} = \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \qquad \text{and} \qquad d_{iB} = 1$$

meaning it depends exclusively on spatial distribution, independent of the transverse momenta of the elements. However, this means it is not infrared-safe, that is, it is vulnerable to the influence of soft radiation [69]. In practice, the Cambridge-Aachen algorithm is applied in reverse to "de-cluster" a jet, allowing its substructure to be effectively determined [70]. Substructure studies of boosted jets from hadronic decays in the WW and WZ channels are especially important as the source of the most stringent EFT parameter constraints [22].

The case $p = 1$ is known as the k_T algorithm, with the corresponding distance parameters

$$d_{ij} \equiv \min\left(p_{Ti}^2, p_{Tj}^2\right) \cdot \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \qquad \text{and} \qquad d_{iB} \equiv p_{Ti}^2$$

which are dominated by low p_T. This makes it more vulnerable to soft radiation than the Cambridge-Aachen algorithm, but also means it has better resolution regarding subjets.

The case $p = -1$ corresponds to the anti-k_T algorithm,

$$d_{ij} \equiv \min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}}^2\right) \cdot \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \qquad \text{and} \qquad d_{iB} \equiv \frac{1}{p_{Ti}^2}$$

As its name suggests, this mode of operation is converse to that of the k_T algorithm; whereas k_T favors low p_T, anti-k_T clusters high-p_T particles first. As a result, it is the most resistant to the effects of soft radiation, making it the preferred overall algorithm for jet reconstruction [69]. Although it is a hierarchical clustering algorithm, anti-k_T behaves in some aspects like an idealized cone algorithm [67].

Anti-k_T jet clustering has different established configurations based on R. The AK4 algorithm is the case of anti-k_T for which $R = 0.4$, used for small-R jets, such as those produced by single-quark decays. Conversely, the AK8 ($R = 0.8$) algorithm is used for large-R jets [66]. One kind of large-R jet is the fat jet, occurring in the highly boosted regime, where the angular separation ΔR between the immediate products of heavy resonance decays decreases. This collimation leads to the particles being registered as a single jet with a relatively larger cone size [70]. This is one way to identify vector boson

decays at the detector level and hence is a key part of the experimental study of VBS. Our simulation uses the AK8 algorithm for this purpose, as we are interested in comparing the kinematics of the leading and subleading jets manifesting from VBS decays.

The next step, after jet reconstruction with a clustering algorithm, is tagging. Tagging is a classification method that uses machine learning (ML) to identify which type of particle serves as the origin of a jet. Its purpose is to differentiate relevant jets from those caused by less interesting processes. Usually, the separation is between background jets in soft processes and the jets of decaying unstable resonances, such as top quarks or the heavy bosons; in our case, such a signal is of course provided by the electroweak bosons [71]. Tagging is also essential in identifying the main experimental signature of VBS: those of the scattered quarks, creating two jets with high invariant mass and separated by a large gap in rapidity [15]. The currently established state-of-the-art tagger for Run-2 and Run-3 data is ParticleNet, an architecture that treats jets as point clouds, or unordered sets of particles [72]. The next-generation Particle Transformer (ParT) has already shown great improvements over ParticleNet and is favored for future studies [26] [71].

3.2 Simulation of events

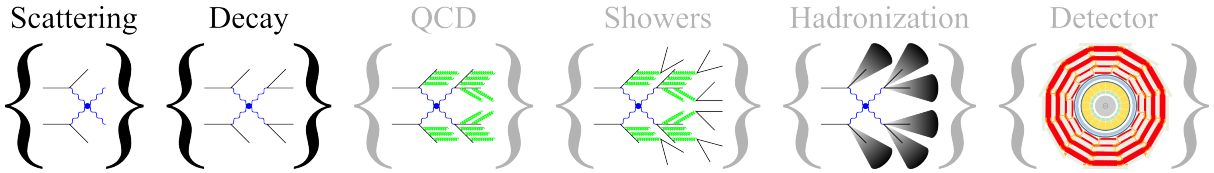


Figure 14: The stages through which event generation progresses. Highlighted stages are included in our generator-level simulation [73].

The MC event chain comprises the following levels, as portrayed in figure 14:

- Scattering: the underlying, "hard" process being simulated. In the electroweak VBS case, this is the LO $W^{\pm}W^{\pm}\rightarrow W^{\pm}W^{\pm}$ process, with contributions from the $W^{\pm}$ QGC, the t-channel vector boson exchange, and the t-channel Higgs exchange.

- Resonance decays: unstable scattered particles decay into lighter states; in our case, each $W^{\pm}$ boson decays into a quark and antiquark. The decay into quarks is the same for both electroweak and QCD VBS, but we will see whether the kinematics of the decay products are different.

- QCD: In particle collisions, gluons are being constantly emitted both before the hard scattering as initial-state radiation (ISR) and after it as final-state radiation (FSR). These radiations partially deplete the energy of the particles in the initial or final state, respectively, and may additionally interfere with each other [74].

- Parton showers: these are the "soft" processes by which the energy scale decreases from around 100 GeV for hard scattering scattering to the hadronization scale, Λ_{QCD}. The initial collision triggers numerous secondary interaction vertices, as each energetic parton loses energy by radiating out gluons, forming showers of partons [73].

- Hadronization: also called fragmentation. In this stage, the final-state quarks are bound by gluons, hadronizing to form jets of colorless baryons and other composite particles as a consequence of color confinement. This stage may be modeled phenomenologically, or described by a fragmentation function (FF), much as PDFs are used in the simulation of showers.

- Detector: the jets resulting from the previous steps interact with the detector on their way out of the interaction region. They penetrate the detector to different lengths, and with different behavior depending on the simulated detector material; in a CMS simulation, jets with low energy and pseudorapidity will be intercepted by the inner tracker, with particles that make it further interacting instead with ECAL or HCAL. Outside effects and imperfections required to compare simulation to experiment are added via MC. Commonly applied for this purpose is the Geant toolkit; the current version is Geant4.

An analysis containing all of the above steps is described as reconstruction- or reco-level. It is the most accurate to real-world experimental physics, in which the properties of the invisible processes and short-lived particles unfolding in the interaction region of a detector must be reconstructed based only on the information carried by particles passing through that detector.

Our generator-level analysis contains the the the first two of the above steps: scattering, decay, and QCD. The generator level is also called the truth level because it enables study of the underlying physics of processes of interest, such as by treating quarks as

free particles outside of color-confined hadronic jets, and tracking exactly the behavior of neutrinos, which pass unnoticed through detector material. The disadvantage of generator-level analyses is that they cannot be directly compared to observational findings; instead, they may be used to point towards which part of the phase space should be investigated in future searches.

All parts of our analysis are done on CERN's LXPLUS environment. To simulate events, we use the `MadGraph5_aMC@NLO` (`MG5aMC`) suite, which produces outputs in the form of compressed files called gridpacks. Other popular event simulators in use at CMS include Pythia, Sherpa, and Herwig, and we have chosen MadGraph due to its suitability for BSM searches [73]. The samples presented in this report are all based on just the LO contributions, as this is the only way to generate polarized samples with `MG5aMC` [12]. To instruct MadGraph on what types of processes to include, we modify two different input cards per desired sample.

The process card is used to specify the kind of process to be simulated. The process card for same-sign VBS, `WPMJJWPMJJjj_EWK_4f_LO_proc_card.dat`, is

```
generate p p > w+ w+ j j QED=4 QCD=0, w+ > j j @ 1
add process p p > w- w- j j QED=4 QCD=0, w- > j j @ 2
output WPMJJWPMJJjj_EWK_4f_LO -nojpeg
```

where we have first enabled a process that goes from an initial state, `p p`, with two protons, to a state with two W^+ bosons and two jets, `w+ w+ j j`, where each W^+ decays hadronically by `w+ > jj`. On the next line, the same is done for the W^-W^- channel with `w- w-`. We use `QED=4 QCD=0` to suppress the QCD contributions to the scattering process. The last line names the output. The process card can be modified to create polarized samples; for example, in `WPMJJWPMJJjj_EWK_4f_LO_TL_proc_card.dat`:

```
generate p p > w+{T} w+{0} j j QED=4 QCD=0, w+ > j j @ 1
add process p p > w-{T} w-{0} j j QED=4 QCD=0, w- > j j @ 2
output WPMJJWPMJJjj_EWK_4f_LO_TL -nojpeg
```

We have added the symbols `T` and `0` to induce transverse and longitudinal polarization, respectively. Each sample produced with `MG5aMC` requires such a file, as well as another input card called a run card.

The run card allows the user to set the parameters—including beam energy, number of events, kinematic cuts—for the simulation. In this case, we use the same run card for every sample so that we can directly compare the results. We have selected a centre-of-mass energy of 13.6 TeV, corresponding to LHC Run 3. We generate 500000 events per sample, enough of a sample size for differences between the samples to be readily visible in the histogram comparisons. Parton distributions are evaluated with the LHAPDF. An example of a section within the card is shown below:

```
#*********************************************************************
# Minimum and maximum invariant mass for pairs                       *
#*********************************************************************
150.0  = mmjj    ! min invariant mass of a jet pair
150.0  = mmbb    ! min invariant mass of a b pair
 -1.0  = mmaa    ! min invariant mass of gamma gamma pair
 -1.0  = mmll    ! min invariant mass of l+l- (same flavour) lepton pair
 -1.0  = mmjjmax ! max invariant mass of a jet pair
 -1.0  = mmbbmax ! max invariant mass of a b pair
 -1.0  = mmaamax ! max invariant mass of gamma gamma pair
 -1.0  = mmllmax ! max invariant mass of l+l- (same flavour) lepton pair
```

We have opted to place cuts on the dijet invariant mass M_{jj} and ditop invariant mass M_{tt}, at 150 GeV, far enough away from the mass of the $\mathrm{W}^{\pm}$ boson at 80 GeV to suppress the generation of triboson events. The goal is to focus exclusively on diboson VBS.

There are two different factorization schemes in use for this type of simulation: the four-flavor scheme (4FS) and the five-flavor scheme (5FS), based on which types of partons are accompanied by PDFs [75]. In the 4FS, the flavors are the gluon, up/down/strange quark, charm quark, and top quark [76]. The 5FS extends this to the bottom quark, which is treated as strictly massive in the 4FS, where it is not included in the initial-state proton PDF. 5FS can be enabled by adding the line `import model sm-ckm_no_b_mass` to the beginning of the process card. For the purposes of our research, we apply the 4FS as it is sufficient for our LO analysis [77].

MadSpin is a utility that allows for the conservation of spin correlation between the parent and decay product in complicated calculations. For example, `MG5aMC` is not able

to handle polarization at next-to-leading order (NLO), nor when the 5FS is selected due to the more complicated diagrams that need to be taken into account. With our setup involving the 4FS at leading order, MadGraph is sufficient for the simulation, and we may choose to proceed with or without MadSpin toward the same result. We have decided to include MadSpin for completeness.

The command to run is submitted to CRAB via a Python script in order to take advantage of distributed resources. The output of the CRAB job is collated into a compressed file called a gridpack or tarball. Extracting the data from the tarball yields a directory with one hundred ROOT files, each containing a portion of the generated dataset. In order to mitigate the large file size, we submit a new job with the HTCondor software to convert the data to the skimmed NanoAOD format.

NanoAOD is a ROOT-based data format designed for CMS that significantly reduces the file size of LHC events, resulting in a skimmed dataset. It is 20 times more compact than MiniAOD, the next-smallest format, with a conversion rate of around ten events per second from MiniAOD to NanoAOD. Each event occupies at most 2 kB of disk space. Each NanoAOD file centres around a single ROOT tree with one entry per event; branch names group variables into shared collections based on type. The collections we access in our analysis are generator-level particles (`NanoGenPart`) and generator-level fat jets (`NanoGenJetAK8`). The skimming for NanoAOD is performed by dropping systematic variations in observables, as they may be easily calculated at any point in the analysis using calibration data, if required. File size is further reduced as, instead of holding all of the data required for event identification purposes, the format only stores top-level discrimination scores. In addition to the main tree holding entries for each individual event, further information, such as the sums of generator weights, is stored in coarser supplementary trees [78].

3.3 Analysis

Our analysis was done in Python, making use of the `PyROOT` interface to read and write data stored in ROOT files, a ubiquitous format in high-energy physics. The scripts we employ are able to produce an event record by invoking the method `analyze()` of the `GenPartDumper` class to print out a detailed event record. An example is presented below, with many of the rows that are not relevant to our analysis excluded for brevity.

---------- Event 770 ----------

index	pdgId	moth	mothId	dR	pt	eta	status	prompt	last copy
0	1				0.00	999.00	21	True	True
1	3				0.00	-999.00	21	True	True
2	-24	0	1	999.00	93.47	0.04	22	True	False
3	-24	0	1	999.00	103.86	1.16	22	True	False
4	2	0	1	999.00	190.74	2.94	23	True	False
5	4	0	1	999.00	41.85	-0.98	23	True	False
6	-24	2	-24	0.01	94.26	0.05	44	True	False
7	-24	3	-24	0.04	102.31	1.19	44	True	False
8	-24	6	-24	0.08	90.55	0.01	44	True	False
9	-24	7	-24	0.02	104.46	1.20	44	True	False
10	-24	8	-24	0.00	90.51	0.01	62	True	True
11	-24	9	-24	0.00	104.57	1.20	62	True	True
12	1	10	-24	0.34	86.32	0.33	23	True	False
13	-2	10	-24	1.96	11.96	-1.58	23	True	False
14	3	11	-24	0.22	92.32	1.30	23	True	False
15	-4	11	-24	1.48	23.07	0.04	23	True	False
16	4	5	4	0.18	39.94	-0.82	71	True	True
17	411	5	4	0.17	39.36	-0.83	2	True	True
18	2				2.96	-4.00	71	True	True
19	-3				0.98	-6.49	71	True	True
20	-421	5	4	0.99	2.26	-0.81	2	True	True
21	513				4.59	2.68	2	True	True
22	-513				6.66	2.08	2	True	True
23	3	14	3	0.00	89.62	1.30	71	True	True
24	-4	15	-4	0.00	22.85	0.04	71	True	True
25	-413	14	3	1.68	18.46	0.05	2	True	True
26	1	12	1	0.01	82.53	0.33	71	True	True
27	-2	13	-2	0.34	7.14	-1.84	71	True	True
28	421				3.57	-0.28	2	True	True
29	2	4	2	0.11	111.49	2.96	71	True	True

...	...	...	...	...	...	...	...	...	...
55	421	50	413	0.01	4.34	1.73	2	False	True

The initial-state quarks responsible for the emission of the W bosons have indices `0` and `1`; in this case, they are a down quark (`pdgId 1`) and a strange quark (`pdgId 3`). At indices `2-5`, we see that both quarks have emitted a W^- boson (`pdgId -24`), causing the down quark to change flavor to an up quark (`pdgId 2`) and the strange quark to change flavor to a charm quark (`pdgId 4`).

Multiple entries may appear for the same particle as it scatters around, and the `lastcopy` flag is used to indicate the last appearance of the particle in the event record before either the particle decays or the event record ends. This flag is used to identify the particles for which we would like to plot kinematic distributions. Additionally, we need to ensure that the particles we are selecting originate from the main scattering event and not from some post-hoc secondary collisions caused by descendant particles that we are not interested in; this is the purpose of the `prompt` flag.

The final copies of the scattered up and charm quark are indexed at `29` and `16`, respectively, confirmed by checking that the corresponding values in the `moth` column map to `4` and `5` respectively. The final copies of the W^- bosons are indexed at `10` and `11`. Right below them, at indices `12-15`, we can see that the first boson decayed into a down quark (`pdgId 1`) and an anti-up quark (`pdgId -2`), while the second decayed into a strange quark (`pdgId 3`) and an anti-charm quark (`pdgId -4`). The final copies of these quarks are located at indices `23`, `24`, `26`, and `27`.

In this manner we have identified all of the particles from the collection `NanoGenPart` whose kinematics we aim to study. The remaining relevant entities are the fat jets representing the W bosons' hadronized decay products; we obtain these simply by accessing the first two elements of the `NanoGenJetAK8` collection, which is already sorted by p_T in decreasing order. If there are less than two elements in the collection, no fat-jet data is recorded for the given event; for this reason, the histograms and ROOT files produced by our analysis script will have less than 500000 entries saved in each fat jet branch.

Once this process is repeated for every event of every sample, `plot_vbs_kinematics.py` is ready to begin the plotting routine. Our program is equipped to handle the properties shown in table 4.

Variable	Property	Arguments
p_T	`pT`	1
η	`eta`	1
ϕ	`phi`	1
E	`E`	1
m	`m`	1–2
Boosted $\cos\theta$	`boosted_cos_theta`	2
$\cos\theta$	`cos_theta`	2
$\Delta R = R_1 - R_0$	`delta_R`	2
$\Delta\eta$	`delta_eta`	2
$\lvert\Delta\eta\rvert$	`abs_delta_eta`	2
$\eta_0 \cdot \eta_1$	`eta_product`	2
$\lvert\Delta\phi\rvert$	`delta_phi`	2

Table 4: List of kinematic variables that `plot_vbs_kinematics.py` can plot from the NanoAOD format. In this context, θ is the angle made by the vectors of the two particles, and not the angle from the z-axis as in the CMS coordinate system. The Arguments column shows how many arguments—particles or jets—are accepted by each function.

The plots may be configured to compare properties either across multiple samples for the same particle or for different particles within the same sample. These selections are made in `.json` files. A simple example of such a file, `WW_example.json`, plotting selected kinematics of two $W^{\pm}$ bosons, is provided in the appendix. An explanation of the fields in the input file and their purposes follows therein.

To save the plots, the program creates a directory with the same name as the input file, in this case `WW_example`. This directory is subdivided depending on which plots were requested; at most, there will be subdirectories for each individual sample and a subdirectory `COMPARE` for plots in the `same_particle_different_sample` section. Finally, each of these directories is further split into subdirectories for each required property. All of the plots in section 4 have been produced in this fashion.

In addition to producing plots, the program also saves the relevant information to ROOT files, organized into trees by property (kinematic variable) and with each leaf corresponding to a single kinematic variable of one or more particles. Our overall MC chain, from event generation to analysis, is shown in figure 15.

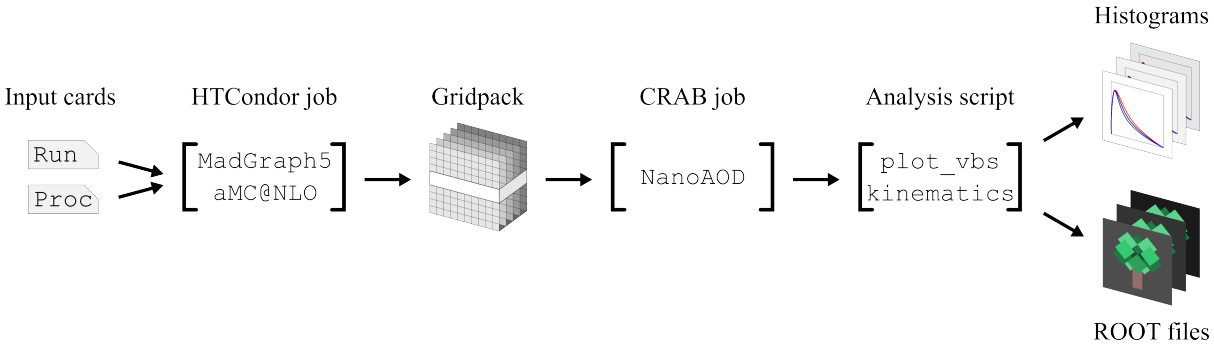


Figure 15: The pipeline showing the steps we take to get from the input cards to the plots of kinematic distributions.

4 Results

4.1 LXPLUS9 sample production

We were able to successfully produce gridpacks for the following leading-order samples, each with 500000 events, using `MG5aMC` on LXPLUS9:

- `WPJJWMJJjj_EWK_LO`: electroweak-based opposite-sign W^+W^- scattering
- `WPJJWMJJjj_QCD_LO`: QCD-based opposite-sign W^+W^- scattering
- `WPMJJWPMJJjj_EWK_LO`: electroweak-based same-sign $W^\pm W^\pm$ scattering
- `WPMJJWPMJJjj_QCD_LO`: QCD-based same-sign $W^\pm W^\pm$ scattering
- `ZJJWPMJJjj_EWK_LO`: electroweak-based $W^\pm Z^0$ scattering
- `ZJJWPMJJjj_QCD_LO`: QCD-based $W^\pm Z^0$ scattering
- `ZJJZJJjj_EWK_LO`: electroweak-based Z^0Z^0 scattering
- `ZJJZJJjj_QCD_LO`: QCD-based Z^0Z^0 scattering

Then, with the parameters of the run cards set to exclude triboson production (M_{jj}, $M_{tt} >$ 150 GeV), we additionally produced

- `WPMJJWPMJJjj_EWK_LO`: unpolarized electroweak $W^\pm W^\pm$ scattering
- `WPMJJWPMJJjj_EWK_LO_LL`: both bosons longitudinally polarized

- `WPMJJWPMJJjj_EWK_LO_TL`: one boson longitudinally and one transversely polarized
- `WPMJJWPMJJjj_EWK_LO_TT`: both bosons transversely polarized
- `WPMJJWPMJJjj_QCD_LO`: unpolarized QCD $W^{\pm}W^{\pm}$ scattering

in order to compare the electroweak and QCD-induced VBS types, as well as the different polarization state combinations. These samples have then been skimmed with NanoAOD; the skimmed data is then used to draw our histograms in the following subsections.

4.2 Kinematics by production mechanism

We first looked for differences between the kinematics of electroweak-based VBS and those of QCD-based VBS. The invariant mass distributions of both samples are shown in a log/log plot in figure 16. We find that with enough of a sample size, there is a noticeable difference between the two types of production, with QCD VBS favoring higher diboson mass sums.

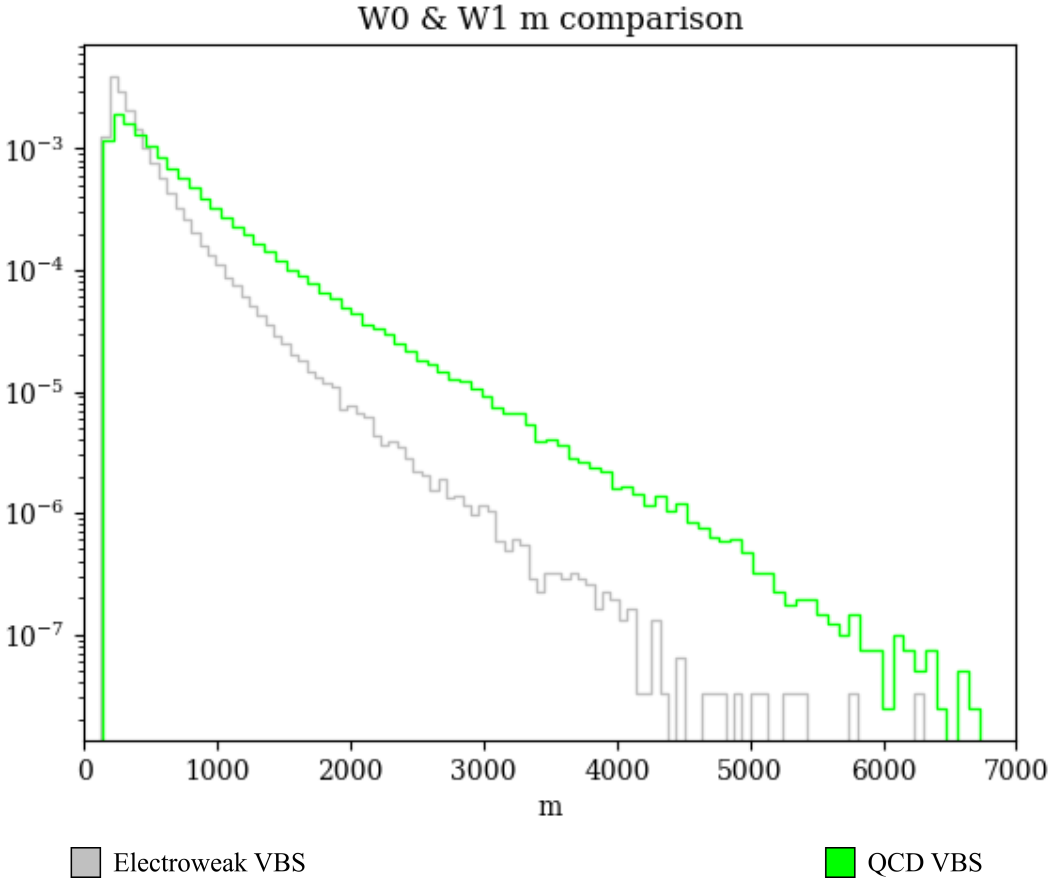


Figure 16: Invariant mass of the diboson system comparing electroweak and QCD VBS.

Generally, and unless stated otherwise, the index i after the letter is used to rank particles of the same type by transverse momentum in descending order, following the convention used to sort the `GenJetAK8` collection. In the preceding example, W_0 is the leading boson and W_1 subleading, that is, lower in transverse momentum than W_0.

We next compare the masses of the fat jets, corresponding to the decay products of each vector boson, in figure 17. Each of the distributions shows a resonance at around 85 GeV, which is most prominent in the leading jet of the electroweak VBS sample. Additionally, the main peak is more prominent in both jets for the electroweak sample.

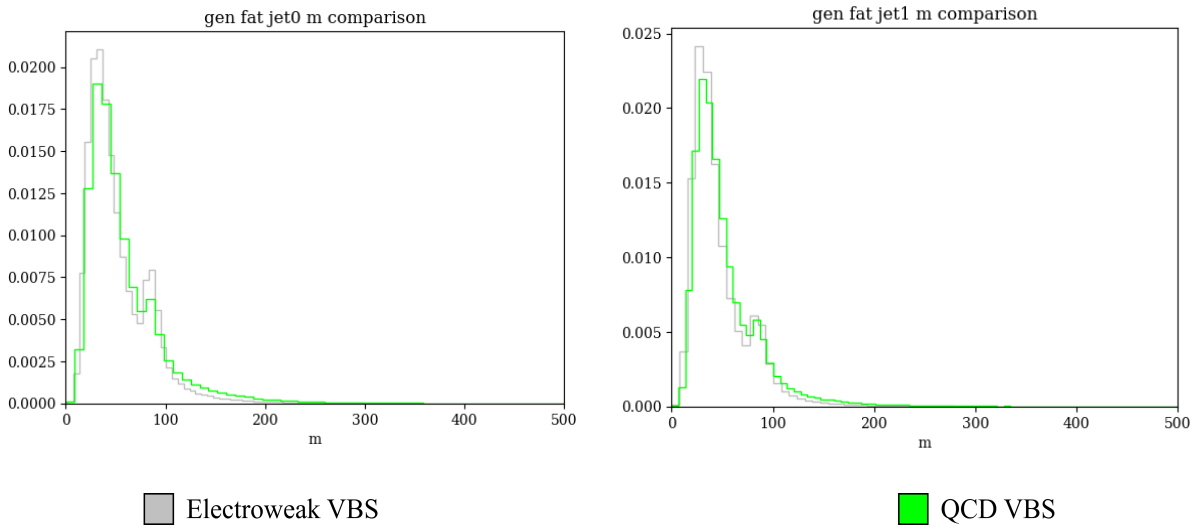


Figure 17: Mass distributions of the leading and subleading fat jets in electroweak- and QCD-based VBS. Event entries where NanoAOD contains less than two jets are not included, so the electroweak and QCD sides have different amounts of data points; the histograms have been normalized to have areas of unity so that the distributions may be compared.

We find that QCD VBS prefers higher absolute pseudorapidities, so such events are more likely to be found in the forward region comopared to electroweak VBS, which favors a more central regime. This difference is apparent in the pseudorapidity distributions of all the W bosons and their decay products, as can be seen in figure 18. Notably, the distribution of the leading boson is slightly concave for QCD compared to that of electroweak VBS, which exhibits no concavity. In the case of the subleading boson, both samples are concave, but the QCD sample is still noticeably more forward.

All pseudorapidity distributions in figure 18 are symmetric, consistent with the uniform distributions of azimuthal angle shown in figure 19. This confirms the isotropy of VBS in the x-y plane, expected to follow from the interaction region's cylindrical symmetry.

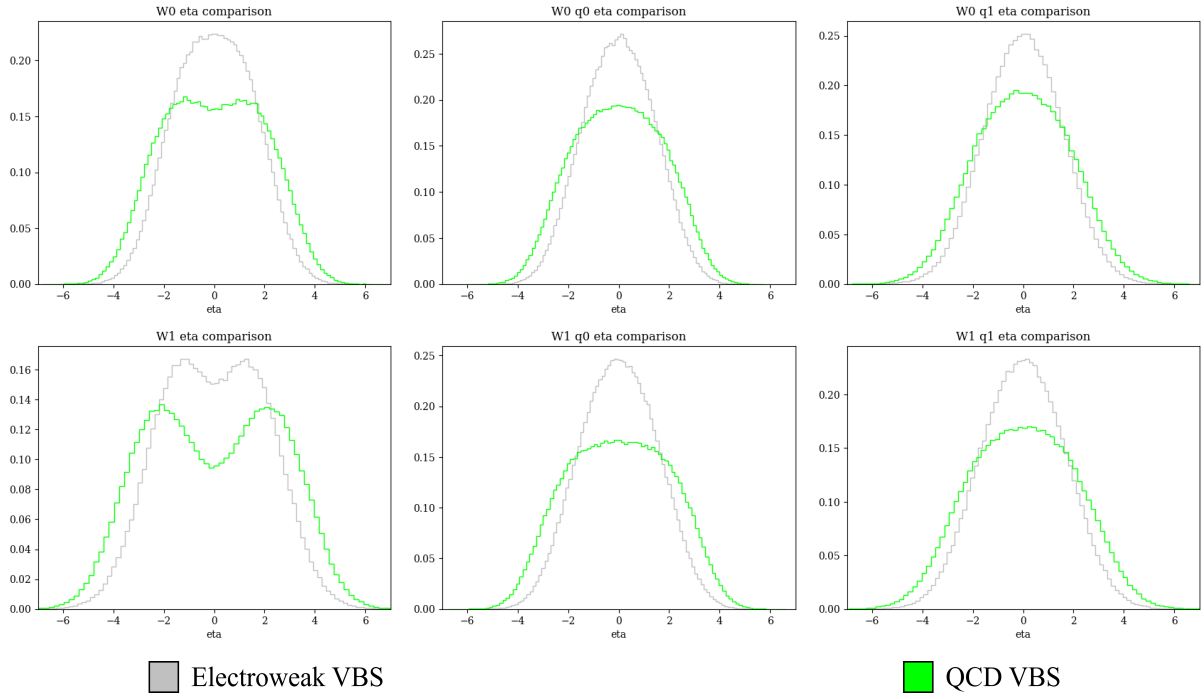


Figure 18: Pseudorapidity distributions of the vector bosons and the quarks that comprise their decay products. The top row shows the leading boson and its decay products ordered by transverse momentum, and the second row does the same for the subleading boson.

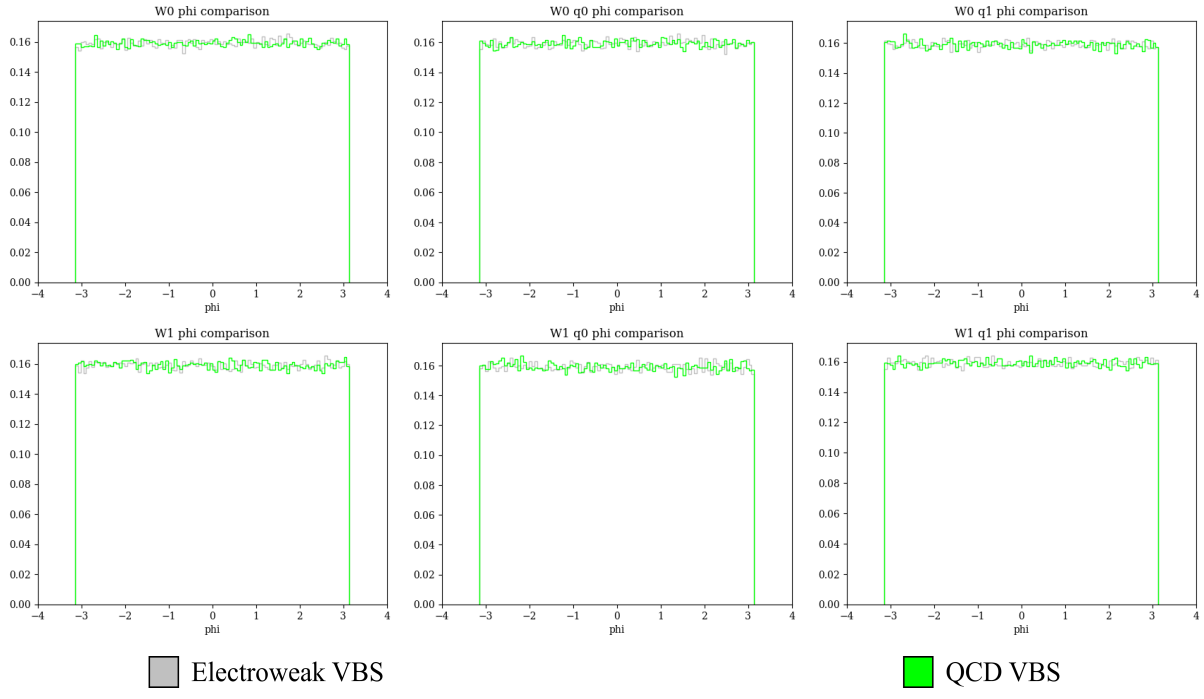


Figure 19: Distributions of the azimuthal angle of the vector bosons and their decay products. The top row shows the leading boson and its decay products ordered by transverse momentum, and the second row does the same for the subleading boson. All combinations are uniformly distributed, as decay products are equally likely to be emitted in any direction from the interaction point.

The decay product quarks may be organized either by their parent particle (W_0q_0, W_0q_1, W_1q_0, W_1q_1), as above, or disregarding the parent and simply sorted by momentum (q_0, q_1, q_2, q_3). We demonstrate the second approach in figure 20, in which there is no clear difference between the distributions of transverse momenta for the daughter quarks, meaning that this is not a useful metric by which to distinguish electroweak VBS from QCD VBS.

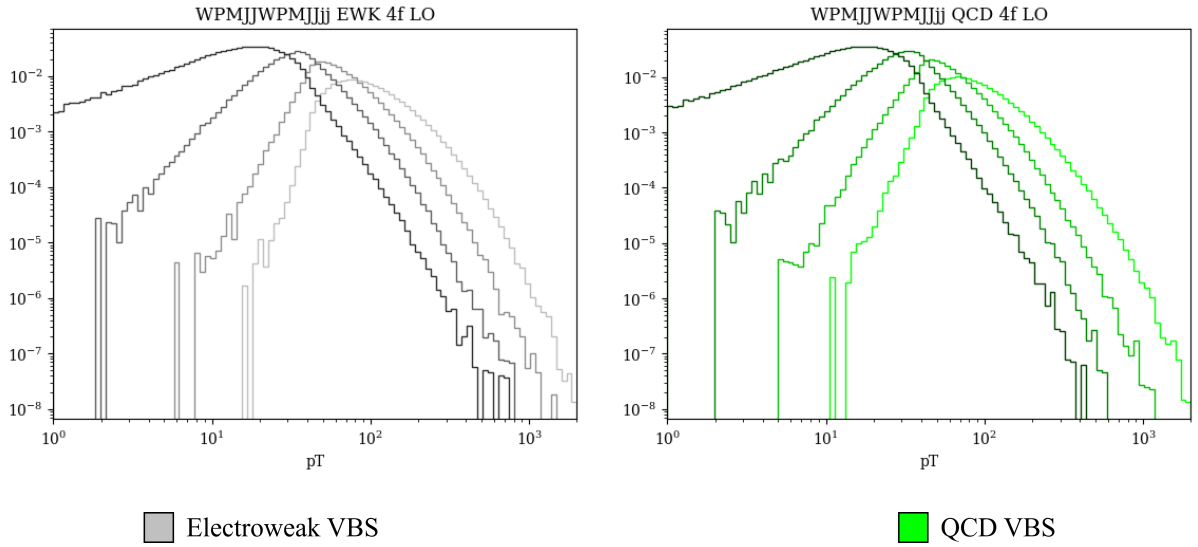


Figure 20: Transverse momentum distribution for the four quarks in the final states of electroweak (left) and QCD (right) VBS. Lighter colors correspond to higher values of transverse momentum. The differences in shape between the two samples are too minor to be able to conclude which is which.

4.3 Kinematics by polarization state

Here we compare the same electroweak sample from the previous section with three electroweak samples that require the following vector boson polarizations: both longitudinal, both transverse, and one in each state. In figure 21, a logarithmic plot of the diboson invariant mass of all four samples, we show that longitudinal polarization, with a sharper peak in its distribution, favors lower invariant mass, while transverse polarization favors higher mass. The transverse/longitudinal (TL) sample distribution is closer to that of the longitudinal/longitudinal (LL) one, while the unpolarized sample is closer to that of the transverse/transverse (TT) one. The latter implies a higher proportion of transversely polarized bosons compared to longitudinally polarized ones.

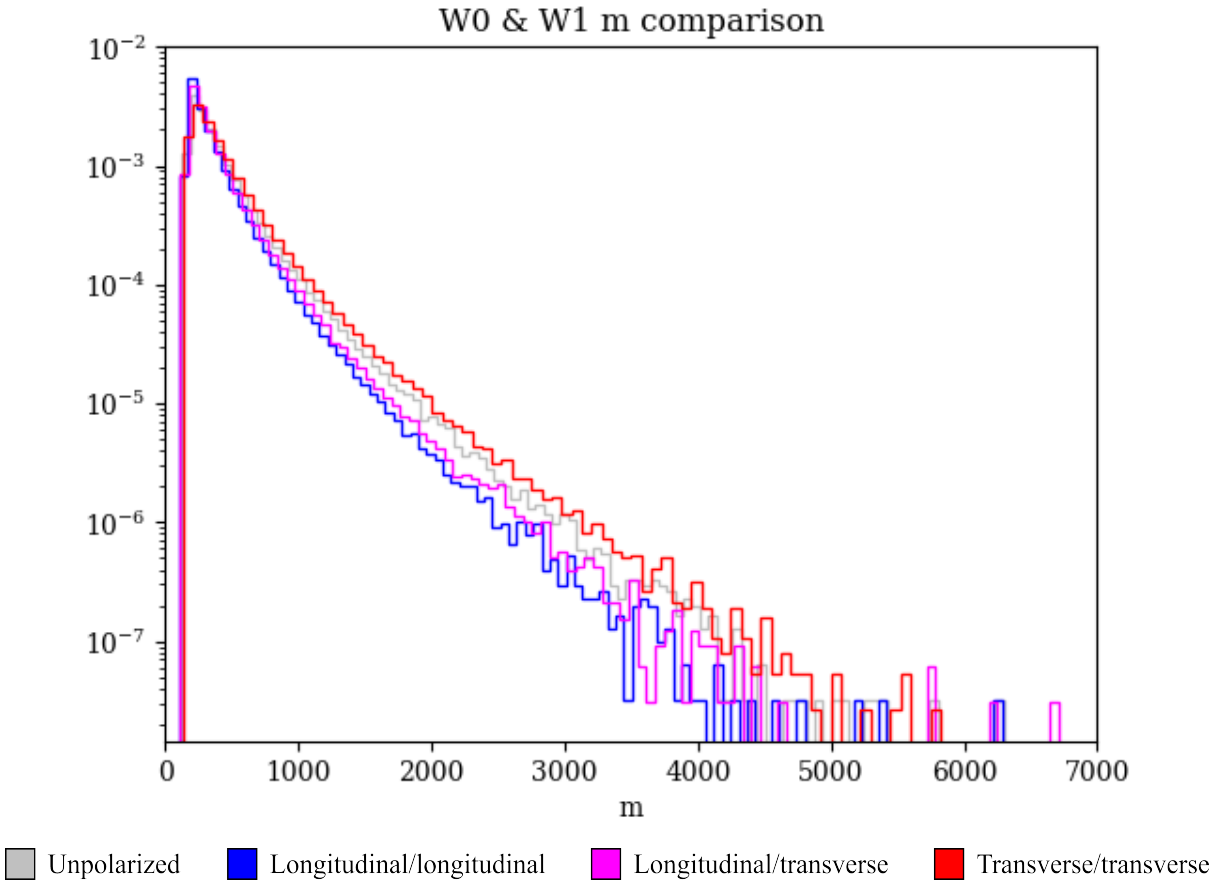


Figure 21: Invariant mass of the diboson system comparing the differently polarized samples.

In the fat jet mass distributions, shown in figure 22, the $\sim$ 85 GeV resonance is significantly more prominent relative to the main peak in TT compared to LL.

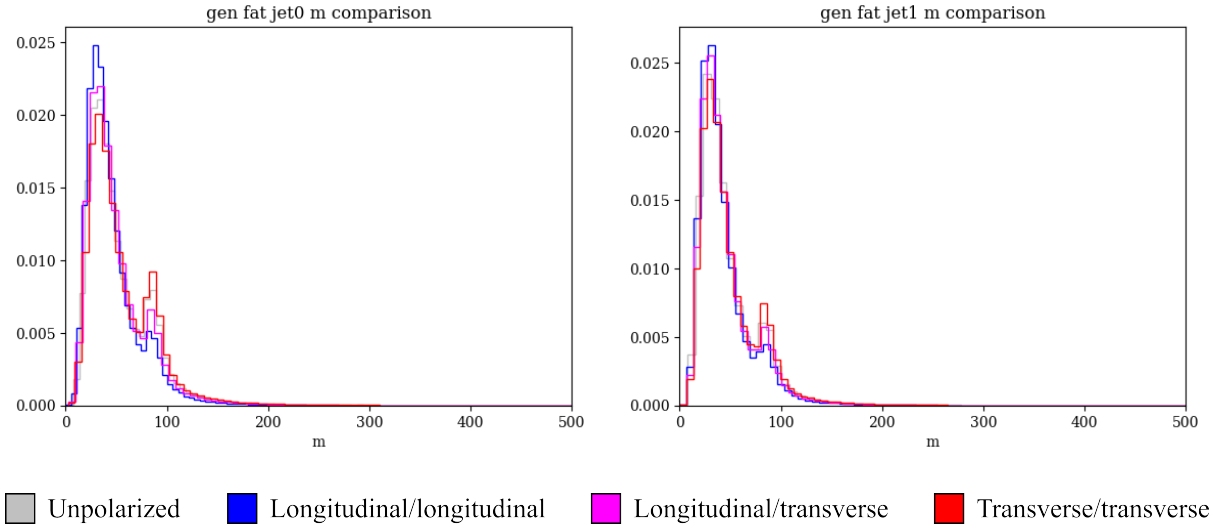


Figure 22: Mass distributions of the leading and subleading fat jets in the different polarization channels. As before, all histograms are normalized to have unity area to compare the shapes of the distributions.

In terms of pseudorapidity, LL electroweak VBS was found to favor higher absolute η, while TT was more central, with a FWHM about 60% that of LL. The other samples are intermediate between LL and TT. This comparison is shown figure 23. We do not show the azimuthal angular distributions as they are simply uniform.

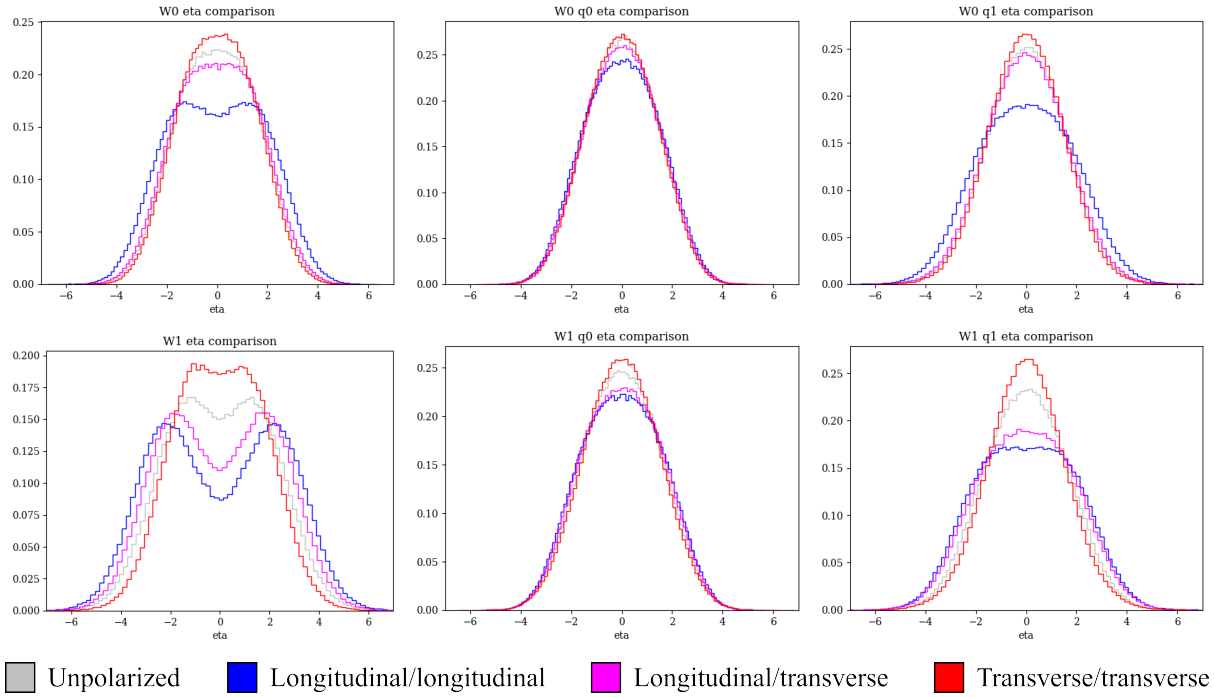


Figure 23: Pseudorapidity distributions of the vector bosons and the quarks that comprise their decay products. The top row shows the leading boson and its daughters ordered by transverse momentum, and the second row does the same for the subleading boson.

In order to compute the polar angle θ between the vector boson and its child quark, we must first boost the quark to the rest frame of the boson [24]. This is done by taking advantage of ROOT's four-vector object types: in our function `boost_to_rest()`, the x, y, and z component are each divided by the energy of the boson and their sign is inverted. The function returns the boosted four-momentum, from which we use the spatial three-momentum to plot $\cos\theta$. The definition of the inner product may be rearranged to recover the desired variable:

$$\vec{p}_V \cdot \vec{\tilde{p}}_q \equiv |\vec{p}_V||\vec{\tilde{p}}_q| \cos\theta \qquad \Rightarrow \qquad \cos\theta = \frac{\vec{p}_V \cdot \vec{\tilde{p}}_q}{|\vec{p}_V||\vec{\tilde{p}}_q|}$$

where the tilde indicates that the momentum of the particle (in this case that of the child quark) is in the rest frame of the boson [24]. The resulting $\cos\theta$ distributions are displayed in figure 24.

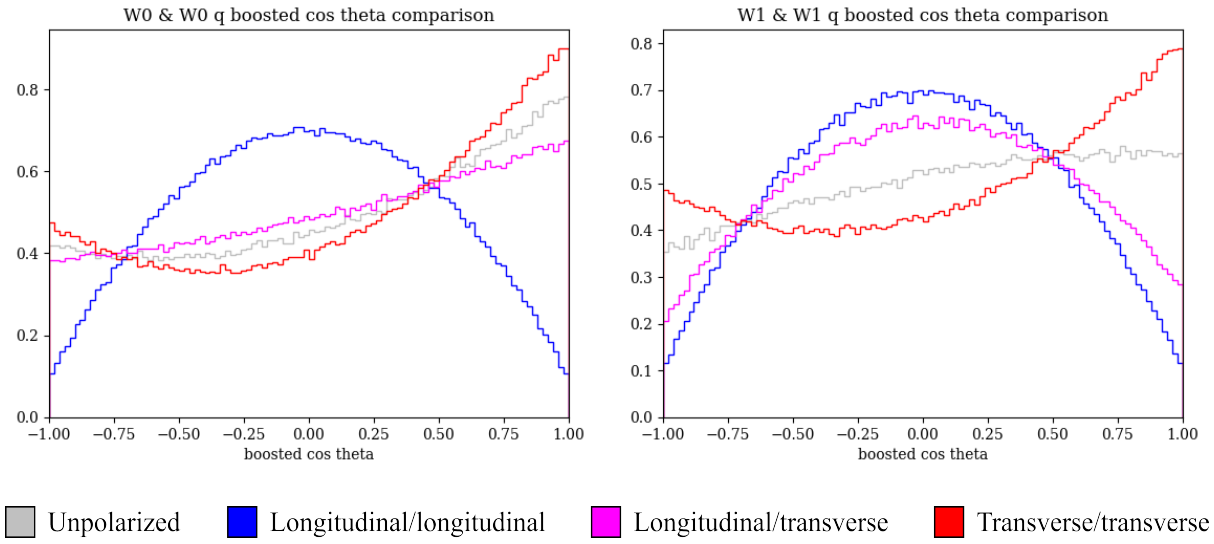


Figure 24: Distributions of $\cos\theta$ between vector bosons and their daughter quarks, the momenta of both boosted to the rest frame of the boson. The same property with the antiquarks is not shown.

We are also interested in the trajectories of scattered quarks. In this case, the "nearest" scattered quark (q_0) to each vector boson is identified in terms of pseudorapidity. First, a check is made by computing $\eta_W \cdot \eta_q$ to determine if each quark is in the same hemisphere as the vector boson, as the product will only be negative if both pseudorapidities have opposite signs. This is further explained in figure 25.

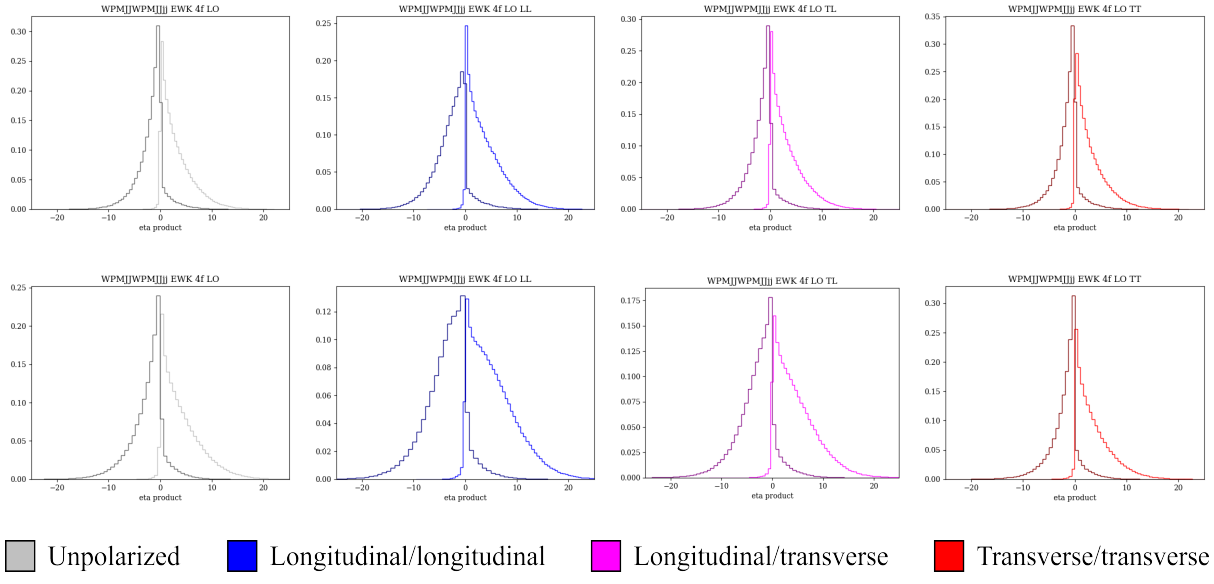


Figure 25: Distributions of pseudorapidity products between a vector boson and the scattered quarks. The lighter color indicates the closer quark in terms of pseudorapidity. Positive values in the x-axis mean that the quark is in the same hemisphere as the boson and negative values mean it is in the opposite hemisphere. The top row is with the subleading boson and the bottom row is with the subleading boson.

Hence, if one product is positive and one is negative, we take the quark corresponding to the positive product to be q_0 and the one responsible for the negative product to be q_1. Otherwise, both are in the same hemisphere; in this case the one corresponding to the lower value of $|\Delta\eta| = |\eta_W - \eta_q|$ is selected for q_0. It can be seen from figure 25 that by far the most common occurrence is for a boson's nearest quark to be in the same hemisphere as the boson and the further one to be in the opposite hemisphere, but this is not always the case. Nonzero bins of the light histogram represent instances in which both quarks are in the opposite hemisphere, and nonzero bins of the dark histogram represent instances in which they are both in the boson's hemisphere. To summarize, based on the figure, it is far less likely for both quarks to be in the opposite hemisphere from the boson than to share the boson's hemisphere, but both cases are in turn far less likely than one quark being in each hemisphere.

This categorization produces some interesting histograms of $|\Delta\phi|$, shown in figure 26. The common peaks are at 0 and $\pm\pi$, but the major variation in the prominence of these peaks distinguishes the samples from each other.

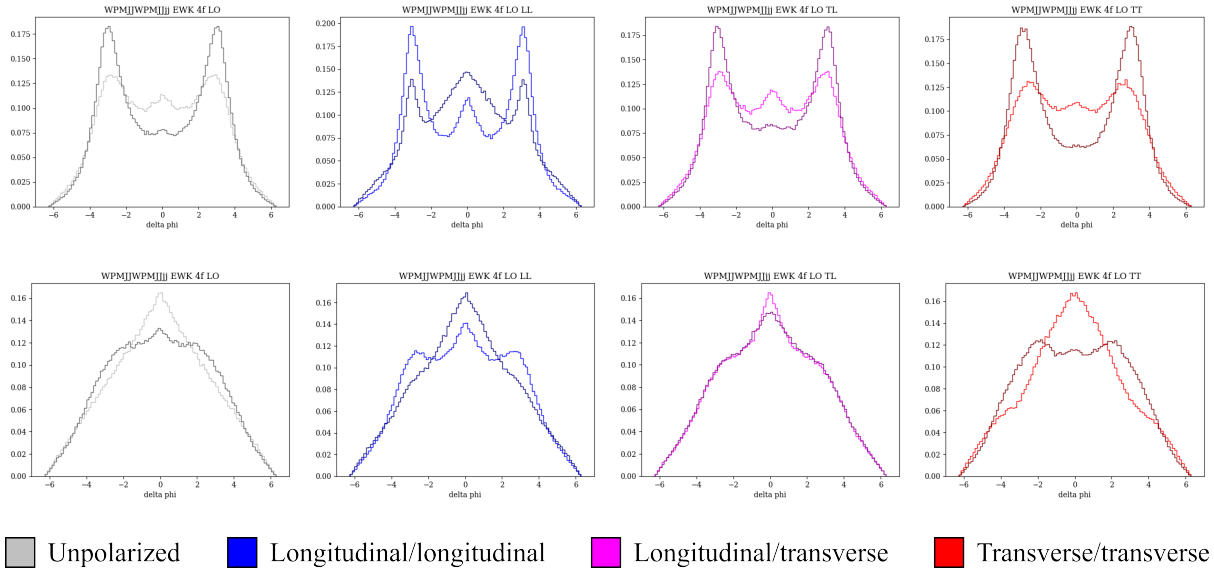


Figure 26: Distributions showing the difference in azimuthal angle between the vector bosons and scattered quarks. The lighter color indicates the closer quark in terms of pseudorapidity. The top row is with the subleading boson and the bottom row is with the subleading boson.

We may also compute $\cos\theta$ between the vector bosons and the scattered quarks, as we did with the decay products, shown in figure 27. However, in this case, we take the angle θ made by their momenta in the lab frame; no Lorentz boost is required.

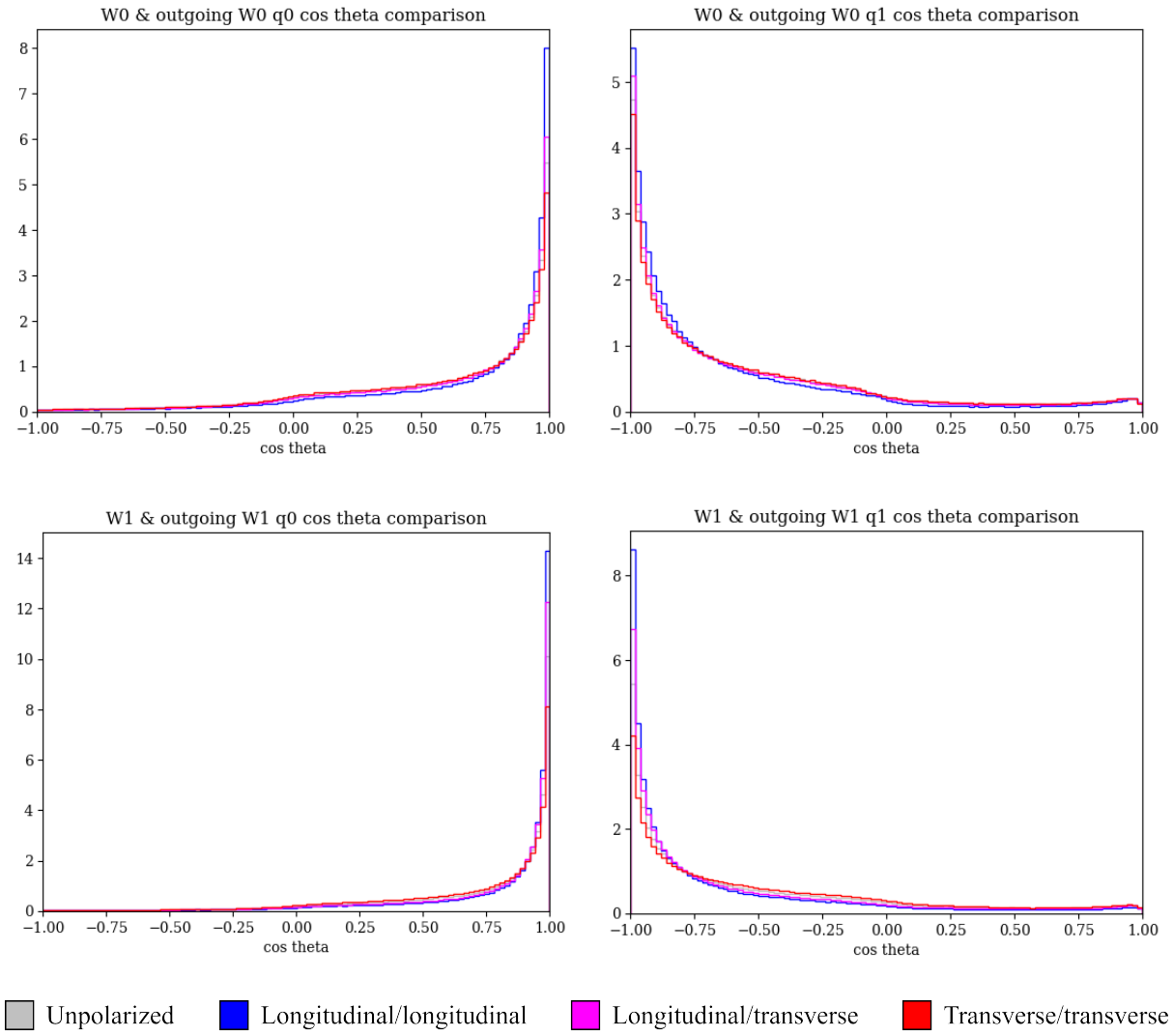


Figure 27: Distributions of $\cos\theta$ between the vector bosons and outgoing quarks. The top row shows the leading W_0 and the outgoing quark nearer to it (W_0q_0) and further from it (W_0q_1) in terms of pseudorapidity. The bottom row shows the subleading W_1 and the outgoing quark nearer to it (W_1q_0) and further from it (W_1q_1).

5 Conclusions

5.1 Summary

We have presented an investigation of the kinematic topology of different types of all-hadronic VBS, an understanding of which will allow us to better discriminate events of interest from difficult backgrounds in the search for new physics. This is especially important considering the rarity of VBS as a process within the collider environment, which is further complicated by the contributions of substantial irreducible backgrounds. As there are currently no published results on the all-hadronic channel of VBS due to

these difficulties, we hope that our analysis will expand the understanding of a process with strong potential to provide evidence of BSM physics.

We include an overview of the Standard Model, an influential physics paradigm that has correctly predicted the existence of fundamental particles well before there existed technology powerful enough to detect them. Accentuated are the components central to VBS: the eponymous vector bosons, the quarks they decay into, and the Higgs boson that serves as a window to physics beyond the Standard Model. VBS is sensitive to the Higgs and EWSB mechanism, for example via deviations in tail behavior, and is one of few processes sensitive to quartic couplings. We describe the need for new physics to address phenomena that the Standard Model does not encompass, and briefly discuss some leading extensions to the Standard Model including additional Higgs bosons, heavier W and Z particles, and EFTs. The study of VBS is expected to provide deeper insight into these extended theories.

The nature of proton-proton collisions involved in VBS production is governed by the dynamics of fundamental partons. We describe the environment of high-energy collisions, in which bunches of protons are accelerated at the LHC and recorded by the CMS detector, which includes a massive electromagnet to bend the trajectories of collision products and a layered structure to register different types of particles. The nature of jets, and the leading processes in place to cluster and identify them, are briefly explained. The information obtained from these experiments is used to design accurate simulations, which we take advantage of in our analysis to study rare processes. A description of the pipeline through which event generation proceeds, and an overall look at our own setup for event generation and subsequent analysis is included.

We conclude the thesis with a presentation of the results of the study. First, the BSM-sensitive electroweak-based VBS is compared to QCD-induced VBS, and differences in the kinematic distributions of the two samples that can be used to differentiate signal from background are identified. Next the focus moves specifically to electroweak VBS, for which we have produced additional samples with specified polarization channels: both particles longitudinally polarized, both transversely polarized, and one of each. Our interest in the polarization of the bosons is due to the phenomenon being a direct consequence of the Higgs mechanism, making it a strong choice for probing the spontaneous symmetry breaking at the threshold of new physics. We indicate the observables that

will allow us to distinguish differently-polarized vector bosons in accelerator experiments. The histograms we have produced offer a comprehensive overview of the kinematics of particles involved in all-hadronic VBS.

5.2 Challenges

Experimentally, the biggest challenge in VBS physics is the minute cross section of the process; this is part of the reasoning behind choosing to look specifically at the all-hadronic channel, which constitutes almost half of all VBS events by occurrence. Nevertheless, even the all-hadronic cross section is some six orders of magnitude smaller than that of non-VBS diboson production, and eight orders smaller than that of QCD multi-jet processes, which dominate at the LHC. The small cross section leads to very high statistical uncertainty. Other uncertainties include that of jet energy resolution, that of background estimation, and finally theoretical uncertainties from loose constraints on parameters of the various BSM models [22].

As only the final state of a process is accessible in a collider experiment, it is impossible to immediately distinguish the signatures of VBS in detectors. The all-hadronic channel in particular is difficult due to its complex final state of up to six jets [20]. Discrimination can be done in advance through cuts—as we have done over the course of this project, by enforcing minima on dijet masses in order to suppress triboson production—or by using MC simulations to find backgrounds and subtracting them from LHC data. However, such cuts may end up distorting the shapes of angular distributions relative to truth-level processes, creating an additional challenge [20].

The difficulties of studying jets are amplified for all-hadronic VBS in comparison to the other channels, as leptons do not shower as jets. Decisions need to be made about which clustering algorithm to apply for reconstruction and which tagger to use for substructure studies. A topic that has not been addressed in this thesis, but can be targeted with a similar analysis framework as the one applied here, is how to tell apart the signatures of W^+ from W^- from Z^0 bosons. This challenge becomes an optimization problem of finding the middle ground between high efficiency (cuts that exclude as few of the VBS events as possible) and a minimized rate of mis-tagging. The preference tag jets tend to have for boosted regimes further complicates matters, as they correspond to the forward regions, coverage of which is currently lacking in the CMS detector.

5.3 Outlook

The most promising future development for VBS will be the HL-LHC upgrade, which is in progress as of the publication of this work and planned for completion in 2027 [79]. Data taking is expected to begin in 2028 [54]. This phase is intended to produce an integrated luminosity of $\mathcal{L} = 4.5 - 5$ ab^{-1} of data [2] [24] with around 200 interactions per bunch crossing [45]. The HL-LHC stands to particularly improve the value of the hadronic and semi-leptonic final states of VBS, for which a lack of statistics is currently among the most major obstacles. The main limitation is instead expected to be systematic uncertainty, the reduction of which will be a key focus for LHC researchers during and after the transition [5].

In addition to improved luminosity of the collider, the corresponding upgrades to CMS will annul the limitations caused by years of radiation damage to the aging detector and, crucially, expand coverage of the interaction region to higher pseudorapidities. The phase-2 upgrade for the inner tracker, involving a complete replacement of the instrument, will increase the maximum accessible pseudorapidity from ± 2.5 to ± 4, improving background rejection in the identification of both jets and leptons [5]. This will make it easier to detect VBS events in the first place, as the signature of the two tag jets originating from the scattered quarks favors more highly boosted regimes.

With the HL-LHC and phase-2 tracker upgrades, theoretical uncertainties are expected to be halved, while experimental uncertainties will scale as the inverse of the square root of the integrated luminosity. Beyond the inner tracker, the installation of high-granularity calorimeters (with pseudorapidity range up to ± 3) in the endcaps of CMS will allow for finer jet resolution with which come cleaner triggers for narrow-cone VBS jets. Accessible level-1 triggering for VBS itself leads to better discrimination between jets perpetuated by quarks and gluons at high transverse momenta, and further from low-p_T soft jets as an effective method of pileup mitigation, important as the higher luminosities will inevitably bring with them a higher rate of production for such soft processes [5].

Although the specific results reported in this thesis involve same-sign $W^{\pm}W^{\pm}$ scattering in its various polarized states, the research framework we used has been designed such that it may be applied with little modification to study the kinematics of VBS events involving Z^0 bosons. Comparisons between electroweak, QCD-induced, and mixed

samples, or between the polarization combinations, are equally supported by our analysis program for $W^{\pm}W^{\pm}$, $W^{+}W^{-}$, $W^{\pm}Z^{0}$, and $Z^{0}Z^{0}$ channels.

In polarization studies with collider data, the problem of extracting the polarization fraction of a sample includes first separating electroweak-induced VBS from QCD-induced diboson production, after which longitudinally polarized vector bosons must be discriminated from transversely polarized ones. The plots in this report show that this can be accomplished by reviewing the particle kinematics; in practice, deep neural networks are a promising potential solution for this task [5]. The research presented here did not include any ML component. ML is ubiquitous in modern particle physics (boosted decision trees were a key component in many previous VBS results) and its influence will only continue to grow; it is especially in demand for jet tagging, with the state-of-the-art ParticleNet and next-generation ParT promising more efficient identification of VBS events via forward jets from the scattered quarks [71]. Additionally being explored is the use of autoencoders to detect anomalies in VBS corresponding to new physics [20].

In essence, despite the inherent difficulties associated with studying a process this elusive, the potential of BSM evidence being encoded in the behavior of VBS makes it well worth the effort. With much of the parked data from Run 3 yet to be combed, bolstered by high-precision MC methods making available unlimited sample sizes, this topic remains active even with the LHC in a period of maintenance. The upcoming luminosity upgrades, coupled with more advanced detector instrumentation, will help to overcome the dearth of statistics currently associated with VBS; the rapid rate of improvement in ML models will further ensure more efficient processing of signals and more dependable discrimination from backgrounds. VBS certainly still has much to offer, and we continue to investigate the tools and methods which will allow us to maximize its impact at the frontier of high-energy physics.

Appendix: Example input file

```
{   "type": "png",
    "organize": "",
    "scheme": "Nano",
    "samples": [["WPMJJWPMJJjj_EWK_4f_LO", "WW", [0.75, 0.75, 0.75]],
                ["WPMJJWPMJJjj_EWK_4f_LO_LL", "WW", [0, 0, 1]],
                ["WPMJJWPMJJjj_EWK_4f_LO_TL", "WW", [1, 0, 1]],
                ["WPMJJWPMJJjj_EWK_4f_LO_TT", "WW", [1, 0, 0]]],
    "pT": ["V0", "V1"], "eta": ["V0", "V1"], "phi": ["V0", "V1"],
    "m": [["V0", "V1"]],
    "same_sample_different_particle": {
        "pT": [["V0", "V1"]],
        "eta": [["V0", "V1"]],
        "phi": [["V0", "V1"]] },
    "same_particle_different_sample": {
        "pT": ["V0", "V1"],
        "eta": ["V0", "V1"],
        "phi": ["V0", "V1"],
        "m": [["V0", "V1"]] },
    "plot_parameters": {
        "pT": { "x-range": [1, 2000],
                "x-scale": "log",
                "y-scale": "log" },
        "eta": { "x-range": [-7, 7],
                 "x-scale": "linear",
                 "y-scale": "linear" },
        "phi": { "x-range": [-4, 4],
                 "x-scale": "linear",
                 "y-scale": "linear" },
        "V0 V1 m": { "x-range": [0, 7000],
                     "x-scale": "linear",
                     "y-scale": "log" } } }
```

The `type` field takes any of the file extensions supported by `matplotlib`, including `svg`, `png`, and `pdf`. The `organize` field defines whether the vector are identified and sorted by particle type or by transverse momentum p_T; `"organize": "type"` toggles type-sorting, while any other input toggles the default behavior of p_T-sorting. For instance, sorting a WZ sample by `type` will designate the bosons `W` and `Z`, while sorting by momentum designates them `V0` and `V1`. The `scheme` field tells the program to process the NanoAOD data with the argument `Nano` or to process the unskimmed data when an empty argument is passed.

The `samples` field holds a list of the samples to be analyzed. Each item in the list follows the format `[sample name, boson types, plot color]` where `sample name` is the name of the folder containing ROOT files for the relevant sample, `boson types` is a string containing the types of the bosons, and `plot color` is a string representing the color selected for the sample's histogram for cross-sample comparisons. Here we have chosen gray for the unpolarized sample, blue for the sample with both bosons longitudinally polarized, red for the sample with both bosons transversely polarized, and magenta for the sample with one of each, as magenta is an equal mixture of blue and red. The following string values for `sample name` are accepted:

- To name the bosons `W0` and `W1`: `WW`, `SS`, `WWSS`, `SSWW`
- To name the bosons `Wp` and `Wm`: `OS`, `WWOS`, `OSWW`, `W+W-`, `W-W+`
- To name the bosons `W` and `Z`: `WZ`, `ZW`
- To name the bosons `Z0` and `Z1`: `ZZ`

The above input file may be provided to our analysis program, `plot_vbs_kinematics.py`, as a positional argument, for instance: `python3 plot_vbs_kinematics.py WW_example.json`. The program does not run without an acceptable input file in the above format.

It should be noted that the names will revert to the default `V0` and `V1` if a cross-sample comparison of boson combinations of different types are requested. For instance, when `"organize": "type"` is chosen, a plot comparing the pseudorapidity η of W^+W^- and $W^\pm Z^0$ VBS will designate W^+ and W^- as `V0` and `V1` from the first sample and $W^\pm$ and Z^0 as `V0` and `V1` from the second sample when making plots containing both samples. In within-sample plots, however, the names `Wp`, `Wm`, `W`, and `Z` would be kept.

The next few fields tell the program which kinematic variables to record. In this case, the transverse momentum p_T, pseudorapidity η, and azimuthal angle ϕ are saved for each vector boson. The same could be done for the mass m, but it is more useful to record the invariant mass of the diboson system, which is done by supplying an array with both vector bosons. The mass field (`m`) accepts both individual particles and arrays, but this is not true for all properties. Properties that only make sense with two particles, such as $\cos\theta$, will only accept arrays of length two, for example.

Once the desired properties for each particles have been selected, the next sections tell the program which plots to make. In `same_sample_different_particle`, each subfield corresponding to a property accepts a nested list as an argument: for example, passing `"pt" : [["V0", "V1"]]`, as we have done, will produce one plot with two histograms on top of each other.

In `same_particle_different_sample`, there is an important distinction: it is not possible to plot multiple particles in the same plot, so nested lists behave differently. In our example, `"pT": ["V0", "V1"]` will produce two plots, one for each particle, while `"m": [["V0", "V1"]]` produces a single plot for a joint property of the two particles.

Finally, `plot_parameters` is used to set the scales and display ranges of the axes. The program splits the given x-range into 100 bins by default, and scales the bin widths appropriately if the x-axis scale is chosen to be logarithmic. Accepted axis scales are those supported by `matplotlib`: `linear`, `log`, `symlog`, and `logit`. The outputs of the example file are collected in figures 28-32.

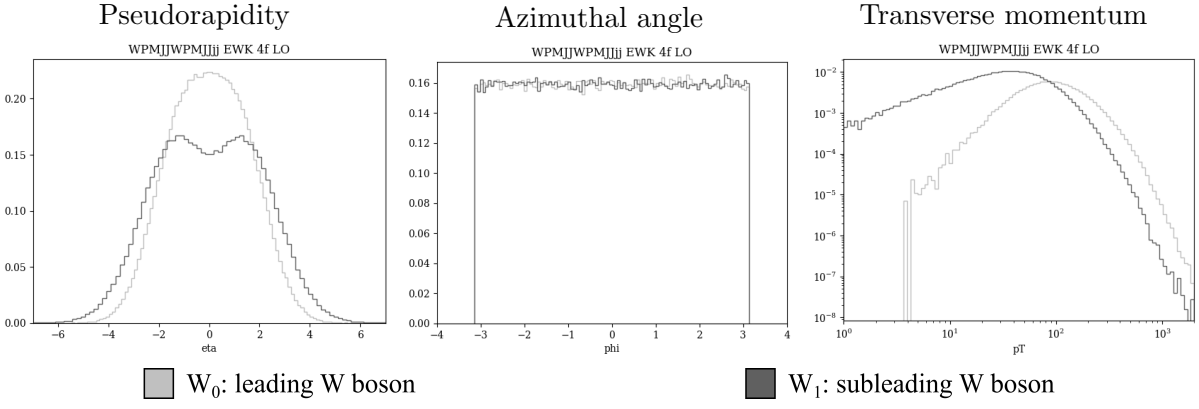


Figure 28: Kinematic comparisons of the vector bosons for the unpolarized sample.

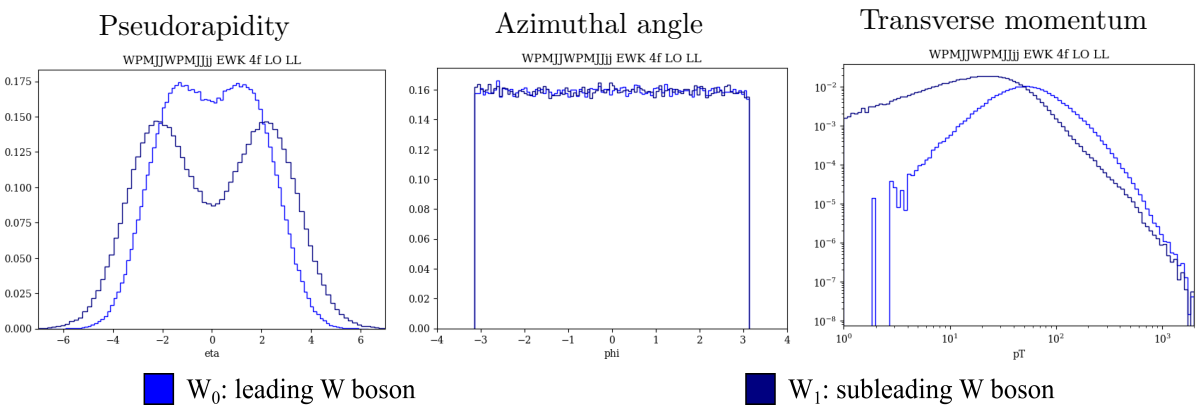


Figure 29: Kinematic comparisons of the vector bosons for the longitudinal sample.

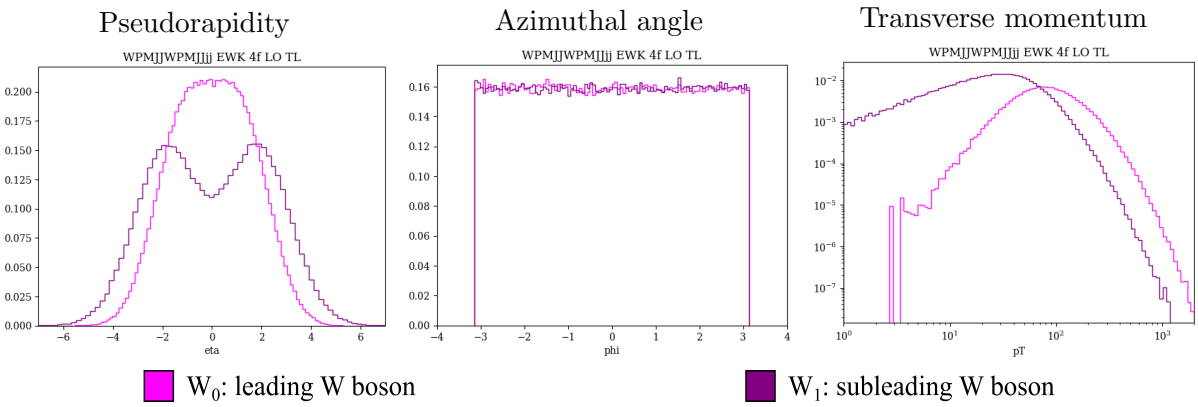


Figure 30: Kinematic comparisons of the vector bosons for the hybrid sample.

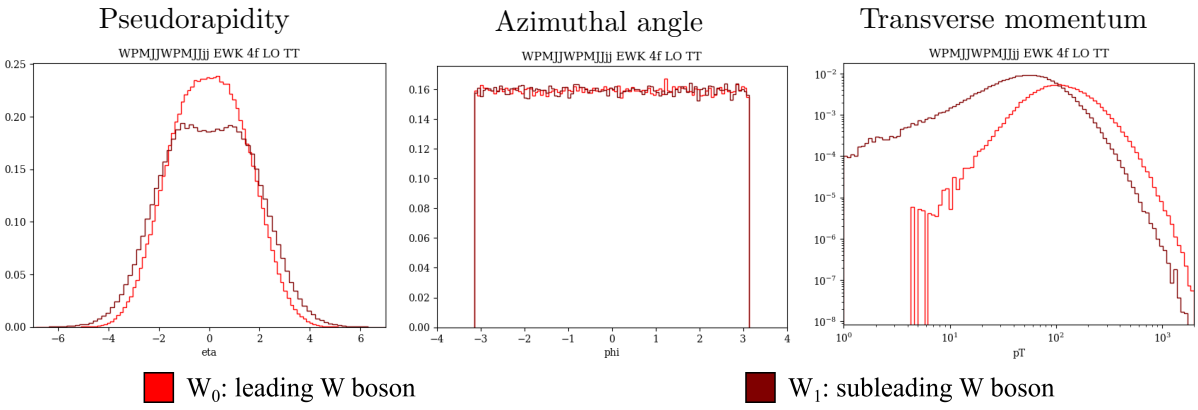


Figure 31: Kinematic comparisons of the vector bosons for the transverse sample.

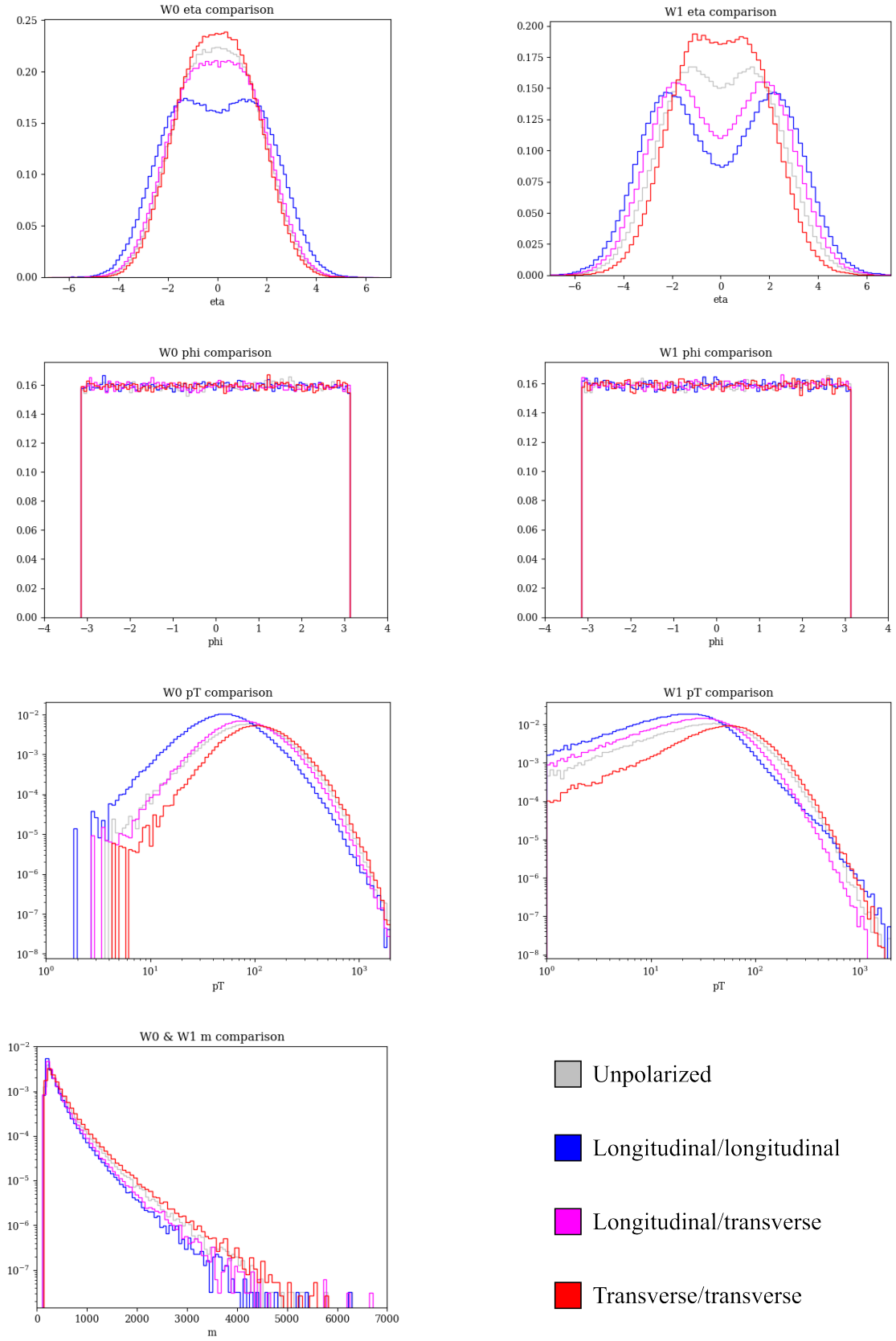


Figure 32: Plots produced by the arguments in `same_particle_different_sample`.

References

[1] P. Arpaia, G. Azzopardi, F. Blanc, X. Buffat, L. Coyle, E. Fol, F. Giordano, et al. Beam Measurements and Machine Learning at the CERN Large Hadron Collider. *IEEE Instrumentation & Measurement Magazine*, 24(9), December 2021, doi:10.1109/MIM.2021.9620023.

[2] M. Gallinaro, K. Long, J. Reuter, R. Ruiz, D. Bachas, L. Barak, F. Bishara, et al. Beyond the Standard Model in Vector Boson Scattering Signatures. In *International Workshop on BSM models in Vector Boson Scattering processes*, number DESY-PROC-2020-002, ISBN 978-3-945931-33-2, ISSN 1435-8077, CP3-20-17, VBSCAN-PUB-04-20, May 2020.

[3] M. Szleper. The Higgs boson and the physics of WW scattering before and after Higgs discovery. December 2014, doi:10.48550/arXiv.1412.8367.

[4] Q. Bi, C. Zhang, and S.-Y. Zhou. Positivity constraints on aQGC: carving out the physical parameter space. *Journal of High Energy Physics*, 2019(6), June 2019, doi:10.1007/jhep06(2019)137.

[5] D. Buarque Franzosi, M. Gallinaro, R. Ruiz, T. K. Aarrestad, F. Cetorelli, M. Chiesa, A. Costantini, et al. Vector boson scattering processes: Status and prospects. *Reviews in Physics*, 8:100071, June 2022, doi:10.1016/j.revip.2022.100071.

[6] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, September 2012, doi:10.1016/j.physletb.2012.08.020.

[7] R. Covarelli, M. Pellen, and M. Zaro. Vector-Boson scattering at the LHC: Unraveling the electroweak sector. *International Journal of Modern Physics A*, 36(16):2130009, May 2021, doi:10.1142/s0217751x2130009x.

[8] M. Thomson. *Modern Particle Physics*. Cambridge University Press, New York, 2013, doi:10.1017/CBO9781139525367.

[9] J. Davighi, G. Isidori, and M. Pesut. Electroweak-flavour and quark-lepton unification: a family non-universal path. *Journal of High Energy Physics*, 2023(4), April 2023, doi:10.1007/jhep04(2023)030.

[10] D. Azevedo, P. Ferreira, M. M. Mühlleitner, R. Santos, and J. Wittbrodt. Models with extended Higgs sectors at future e^+e^- colliders. *Physical Review D*, 99(5), March 2019, doi:10.1103/physrevd.99.055013.

[11] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli. The standard model of particle physics. *Rev. Mod. Phys.*, 71:S96–S111, March 1999, doi:10.1103/RevModPhys.71.S96.

[12] M. Lizzo. ATLAS and CMS di-boson polarization measurements. SMP Workshop, May 2024.

[13] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022, doi:10.1093/ptep/ptac097.

[14] P. A. Zyla, R. M. Barnett, J. Beringer, O. Dahl, D. A. Dwyer, D. E. Groom, C.-J. Lin, et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), August 2020, doi:10.1093/ptep/ptaa104.

[15] S. Solomon. Electroweak Vector Boson Scattering Measurements at the LHC. Technical Report ATL-PHYS-PROC-2023-005, CERN, Geneva, March 2023.

[16] H. Kirschenmann. Overview of the ForVVard project. HIP CMS Experiment Group Retreat, March 2024.

[17] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, et al. A search for WWγ and WZγ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 8$ TeV. *Physical Review D*, 90(3), August 2014, doi:10.1103/physrevd.90.032008.

[18] C. Roland. Triple and quartic gauge boson couplings. 41st International Symposium on Physics in Collision, September 2022.

[19] M. Rauch. Vector-Boson Fusion and Vector-Boson Scattering. October 2016, doi:10.48550/arXiv.1610.08420.

[20] J. Li, S. Yang, and R. Zhang. Detecting anomalies in vector boson scattering. *Chinese Physics C*, 45(7):073104, July 2021, doi:10.1088/1674-1137/abf829.

[21] R. Penco. An Introduction to Effective Field Theories. June 2020, doi:10.48550/arXiv.2006.16285.

[22] A. Ballestrero, R. Bellan, B. Biedermann, C. Bittrich, I. Brivio, A. Cardini, G. Gomez-Ceballos, et al. Vector boson scattering: Recent experimental and theory developments. *Reviews in Physics*, 3:44–63, November 2018, doi:10.1016/j.revip.2018.11.001.

[23] A. Falkowski. Lectures on SMEFT. *Eur. Phys. J. C*, 83(7):656, July 2023, doi:10.1140/epjc/s10052-023-11821-3.

[24] D. B. Franzosi, O. Mattelaer, R. Ruiz, and S. Shil. Automated predictions from polarized matrix elements. *Journal of High Energy Physics*, 2020(4), April 2020, doi:10.1007/jhep04(2020)082.

[25] E. Celada. Triboson production in the SMEFT. HEFT 2024, June 2024.

[26] N. Norjoharuddeen. Vector Boson Scattering Measurement in the All-Hadronic Final State. HIP CMS Experiment Group Retreat, March 2024.

[27] S. De, V. Rentala, and W. Shepherd. Measuring the polarization of boosted, hadronic W bosons with jet substructure observables. August 2020, doi:10.48550/arXiv.2008.04318.

[28] K. Ozdemir. Study of Kinematics of the Jets in VBS Process at the Large Hadron Collider. May 2023, doi:10.48550/arXiv.2305.16552.

[29] F. Close. *Particle Physics: A Very Short Introduction*. Oxford University Press, May 2004, doi:10.1093/actrade/9780192804341.001.0001.

[30] A. Tumasyan, W. Adam, J.W. Andrejkovic, T. Bergauer, S. Chatterjee, K. Damanakis, M. Dragicevic, et al. Search for new heavy resonances decaying to WW, WZ, ZZ, WH, or ZH boson pairs in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 844:137813, September 2023, doi:10.1016/j.physletb.2023.137813.

[31] P. Osland, A. A. Pankov, and I. A. Serenkova. Bounds on the mass and mixing of Z' and W' bosons decaying into different pairings of W, Z, or Higgs bosons using CMS data at the LHC, June 2022.

[32] C. Csáki, M. L. Graesser, and G. D. Kribs. Radion dynamics and electroweak physics. *Physical Review D*, 63(6), February 2001, doi:10.1103/physrevd.63.065002.

[33] E. Megías, G. Nardini, and M. Quirós. Radion dynamics, heavy Kaluza–Klein resonances and gravitational waves. *International Journal of Modern Physics A*, 37(33), September 2022, doi:10.1142/s0217751x22400231.

[34] H. Cai. Radion dynamics in the multibrane Randall-Sundrum model. *Physical Review D*, 105(7), April 2022, doi:10.1103/physrevd.105.075009.

[35] A. Denner. VBS precision predictions. 15th Vienna Central European Seminar, November 2019.

[36] T. Han, D. Krohn, L.-T. Wang, and W. Zhu. New physics signals in longitudinal gauge boson scattering at the LHC. *Journal of High Energy Physics*, 2010(3), March 2010, doi:10.1007/jhep03(2010)082.

[37] Claudius G. Krause. Higgs Effective Field Theories - Systematics and Applications. September 2016, doi:10.5282/edoc.19873.

[38] F. Bishara, R. Contino, and J. Rojo. Higgs pair production in vector-boson fusion at the LHC and beyond. *The European Physical Journal C*, 77(7), July 2017, doi:10.1140/epjc/s10052-017-5037-9.

[39] M. Ruhdorfer, E. Salvioni, and A. Weiler. A global view of the off-shell Higgs portal. *SciPost Physics*, 8(2), February 2020, doi:10.21468/scipostphys.8.2.027.

[40] M. Frigerio, A. Pomarol, F. Riva, and A. Urbano. Composite scalar dark matter. *Journal of High Energy Physics*, 2012(7), July 2012, doi:10.1007/jhep07(2012)015.

[41] B. Fuks, M. Nemevšek, and R. Ruiz. Doubly charged Higgs boson production at hadron colliders. *Physical Review D*, 101(7), April 2020, doi:10.1103/physrevd.101.075022.

[42] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. *SciPost Phys. Codebases*, November 2022, doi:10.21468/SciPostPhysCodeb.8.

[43] D. Valsecchi. First evidence of VBS in semileptonic decays with $WVjj \to l\nu qqjj$ final state and optimization of the CMS electromagnetic calorimeter for Run III. Other thesis, 2022, Milan Bicocca U., January 2022.

[44] J. P. Farmer, A. Caldwell, and A. Pukhov. Preliminary Investigation of a Higgs Factory based on Proton-Driven Plasma Wakefield Acceleration. January 2024, doi:10.48550/arXiv.2401.14765.

[45] C. M. Bhat, P. C. Bhat, W. Chou, E. Gianfelice-Wendt, J. Lykken, G. L. Sabbi, T. Sen, and R. Talman. Proton-proton and electron-positron collider in a 100 km ring at Fermilab. In *Snowmass 2013: Snowmass on the Mississippi*, June 2013.

[46] S. Abachi, B. Abbott, M. Abolins, B. S. Acharya, I. Adam, D. L. Adams, M. Adams, et al. Observation of the Top Quark. *Physical Review Letters*, 74(14):2632–2637, April 1995, doi:10.1103/physrevlett.74.2632.

[47] A. Giammanco and R. Schwienhorst. Single top-quark production at the Tevatron and the LHC. *Reviews of Modern Physics*, 90(3), July 2018, doi:10.1103/revmodphys.90.035001.

[48] R. K. Ellis et al. Input for the European Strategy for Particle Physics Update 2020. *Physics Briefing Book*, October 2019, doi:10.48550/arXiv.1910.11775.

[49] J. de Blas, M. Cepeda, J. D'Hondt, R.K. Ellis, C. Grojean, B. Heinemann, F. Maltoni, et al. Higgs Boson studies at future particle colliders. *Journal of High Energy Physics*, 2020(1), January 2020, doi:10.1007/jhep01(2020)139.

[50] A. Butter, T. Plehn, S. Schumann, S. Badger, S. Caron, K. Cranmer, F. Bello, et al. Machine learning and LHC event generation. *SciPost Physics*, 14, April 2023, doi:10.21468/SciPostPhys.14.4.079.

[51] L. Cunqueiro and A. M. Sickles. Studying the QGP with Jets at the LHC and RHIC. *Progress in Particle and Nuclear Physics*, 124:103940, May 2022, doi:10.1016/j.ppnp.2022.103940.

[52] V. O'Dell. CMS at the HL-LHC: The technological challenges for Higgs physics. CPAD Instrumentation Frontier Meeting 2016, October 2016.

[53] D. Contardo, M. Klute, J. Mans, L. Silvestris, and J. Butler. Technical Proposal for the Phase-II Upgrade of the CMS Detector. Technical Report CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02, CERN, Geneva, 2015.

[54] E. Migliore. The CMS Tracker Upgrade for the High-Luminosity LHC. November 2019, doi:10.48550/arXiv.1911.02438.

[55] A.M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, et al. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, October 2017, doi:10.1088/1748-0221/12/10/p10003.

[56] W. Adam, T. Bergauer, D. Blöch, M. Dragicevic, R. Frühwirth, V. Hinger, H. Steininger, et al. The CMS Phase-1 Pixel Detector Upgrade. *JINST*, 16(02):P02027, 2021, doi:10.1088/1748-0221/16/02/P02027.

[57] LHC Experiments Committee. The CMS hadron calorimeter project: Technical Design Report. Technical Report CERN-LHCC-97-031, CMS-TDR-2, CERN, Geneva, June 1997.

[58] S. Chatrchyan, G. Hmayakyan, V. Khachatryan, A. M. Sirunyan, R. Adolphi, G. Anagnostou, R. Brauer, et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008, doi:10.1088/1748-0221/3/08/S08004. Also published by CERN Geneva in 2010.

[59] N. Amapane and V. Klyukhin. Development of the CMS Magnetic Field Map. *Symmetry*, 15(5):1030, May 2023, doi:10.3390/sym15051030.

[60] J.-L. Agram. CMS Silicon Strip Tracker Performance. Technology and Instrumentation in Particle Physics 2011, June 2011.

[61] J.-L. Agram. CMS Silicon Strip Tracker Performance. *Physics Procedia*, 37(CMS-CR-2011-147), 2012, doi:10.1016/j.phpro.2012.02.423. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).

[62] M. Musich. Upgrades of CMS inner tracker for HL-LHC (pixels). 23rd International Workshop on Vertex Detectors, Charles University, September 2014.

[63] S. Orfanelli on behalf of CMS collaboration. The CMS Inner Tracker electronics system development. *Journal of Instrumentation*, 17(08):C08003, August 2022, doi:10.1088/1748-0221/17/08/C08003.

[64] A. Tumasyan et al. The Phase-2 Upgrade of the CMS Tracker. (CERN-LHCC-2017-009, CMS-TDR-014, CERN-LHCC-2017-009, CMS-TDR-014), June 2017, doi:10.17181/CERN.QZ28.FLHW.

[65] S. M. Beolé. Upgrade of Trackers at LHC. 10th Annual Large Hadron Collider Physics Conference, May 2022.

[66] M. Verzetti. Jet classification techniques in CMS. Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada., December 2019.

[67] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, April 2008, doi:10.1088/1126-6708/2008/04/063.

[68] G. C. Blazey, J. R. Dittmann, S. D. Ellis, V. D. Elvira, K. Frame, S. Grinstein, R. Hirosky, et al. Run II Jet Physics. In *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*, number FERMILAB-CONF-00-092-E, May 2000.

[69] R. Atkin. Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645(1):012008, September 2015, doi:10.1088/1742-6596/645/1/012008.

[70] C. Beluffi. b jet Identification in CMS. *Nuclear and Particle Physics Proceedings*, 273-275:2491–2493, 2016, doi:10.1016/j.nuclphysbps.2015.09.435. 37th International Conference on High Energy Physics (ICHEP).

[71] H. Qu, C. Li, and S. Qian. Particle Transformer for Jet Tagging. February 2022, doi:10.48550/arXiv.2202.03772.

[72] H. Qu and L. Gouskos. Jet tagging via particle clouds. *Physical Review D*, 101(5), March 2020, doi:10.1103/physrevd.101.056019.

[73] T. Plehn, A. Butter, B. Dillon, T. Heimel, C. Krause, and R. Winterhalder. Modern Machine Learning for LHC Physicists. November 2024, doi:10.48550/arXiv.2211.01421.

[74] C. Charlot. LHC Physics. Sarajevo School of High Energy Physics 2018, October 2018.

[75] S. Forte, D. Napoletano, and M. Ubiali. Higgs production in bottom-quark fusion: Matching beyond leading order. *Physics Letters B*, 763:190–196, December 2016, doi:10.1016/j.physletb.2016.10.040.

[76] N. Toikka. Jet flavor: background and current status. HIP CMS Experiment Group Retreat, March 2024.

[77] F. Maltoni. Discussion: 4F vs 5F schemes. Berkeley Workshop on flavor production, January 2013.

[78] M. Peruzzi, G. Petrucciani, and A. Rizzi. The NanoAOD event data format in CMS. *Journal of Physics: Conference Series*, 1525(1):012038, April 2020, doi:10.1088/1742-6596/1525/1/012038.

[79] P. Chang, P. Elmer, Y. Gu, V. Krutelyov, G. Niendorf, M. Reid, B. V. Sathia Narayanan, et al. Segment Linking: A Highly Parallelizable Track Reconstruction Algorithm for HL-LHC. *Journal of Physics: Conference Series*, 2375(1):012005, November 2022, doi:10.1088/1742-6596/2375/1/012005.