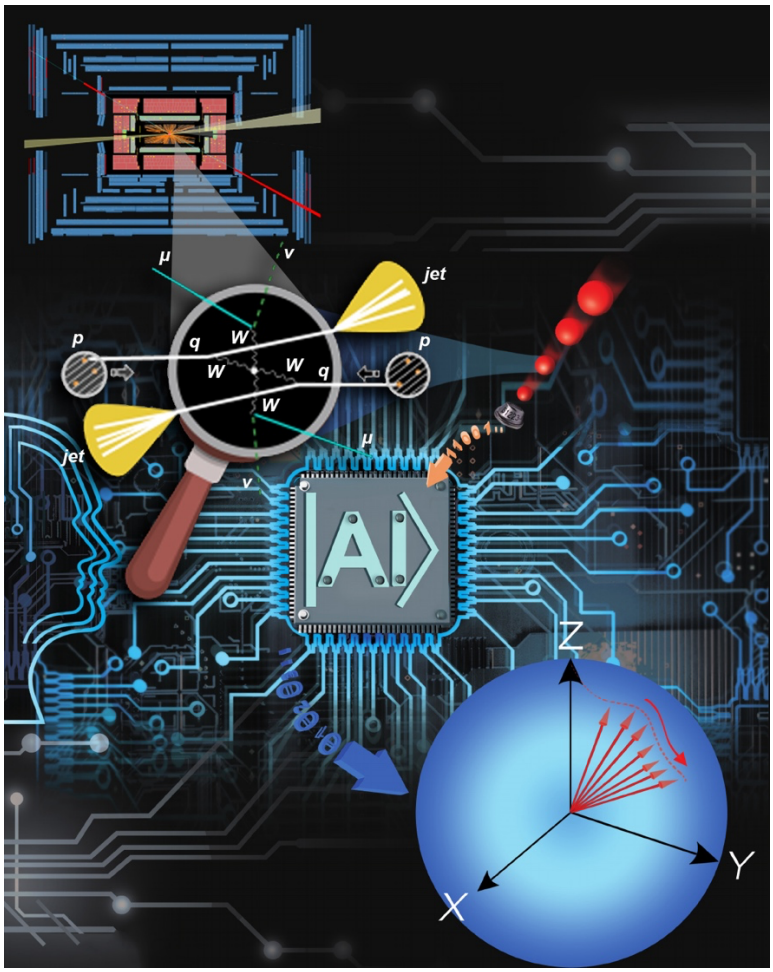
UNIVERSITÀ DEGLI STUDI DI PAVIA
DOTTORATO DI RICERCA IN FISICA – XXXIII CICLO

W Boson Polarization Studies for Vector Boson Scattering at LHC: from Classical Approaches to Quantum Computing

Michele Grossi



Tesi per il conseguimento del titolo

Università degli Studi di Pavia
Dipartimento di Fisica

DOTTORATO DI RICERCA IN FISICA – XXXIII CICLO

W Boson Polarization Studies for Vector Boson Scattering at LHC: from Classical Approaches to Quantum Computing

Michele Grossi

Submitted to the Graduate School of Physics in partial fulfilment of the requirements for the degree of

DOTTORE DI RICERCA IN FISICA

DOCTOR OF PHILOSOPHY IN PHYSICS

at the

University of Pavia

Supervisor: Prof. Daniela Marcella Rebuzzi

Cover: artistic representation of Vector Boson Scattering observation with the ATLAS detector and data analysis integration flow with artificial intelligence and quantum computing methodology

***W* Boson Polarization Studies for Vector Boson Scattering at LHC: from Classical Approaches to Quantum Computing**

Michele Grossi

PhD thesis - University of Pavia
Pavia, Italy, April 2021

Final Examination Committee

External Referees

Aleandro NISATI (INFN, Roma 1)
Pietro GOVONI (University of Milano Bicocca)

External Collaborators

Borut KERSEVAN (University of Ljubljana)
Jakob NOVAK (University of Ljubljana)

This PhD work has been carried out in the framework of an industrial doctoral path collaboration between the University of Pavia and IBM Italia (Executive PhD)

IBM Supervisor

Federico Mattei (IBM Italia)
Ivano TAVERNELLI (IBM Research Zurich)

Final Examination

8$^{\text{th}}$ April, 2020
CANDIDATE: Michele GROSSI
Università degli Studi di Pavia, Dipartimento di Fisica,
Pavia, Italy

Our virtues and our failures are inseparable, like force and matter. When they separate, man is no more.

Nikola Tesla

Contents

Introduction

> Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.
>
> Marie Curie

The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) has, in the recent years, delivered unprecedented high-energy proton-proton collisions that have been collected and studied by two multi-purpose experiments, ATLAS and CMS. The two collaborations are exploiting these data to explore the fundamental nature of matter and the basic forces that shape the Universe, testing the Standard Model of particle physics, our current understanding of elementary particles and their fundamental interactions, in regimes never investigated before. In this thesis, we focus on one physics process in particular, the Vector Boson Scattering (VBS), which is one of the keys to probe the ElectroWeak sector of the Standard Model in the TeV regime and to shed light on the mechanism of ElectroWeak symmetry breaking.

LHC has operated at 7 and 8 TeV center of mass energies during the so called "Run 1" data taking period (2010-2012). The landmark Higgs boson discovery, in July 2012, presented us with a wonderful rich and diverse physics program. In "Run 2", namely the data taking period at 13 TeV which followed, we learnt a lot more about the Higgs boson, notably how it couples to bosons and to the heaviest generation of fermions. Thanks to "Run 2" data, we know the masses of the Higgs boson, of the top quark and of the W boson with considerably greater precision. Such measurements are important to constrain the Standard Model as a (meta) stable theory. In this perspective, LHC will be upgraded to the High Luminosity LHC project starting from 2027, according to the current schedule.

The quantity of data collected are keeping the analysis teams busy in the current long-shutdown period which precedes "Run 3" data taking period. On the phenomenological side, this moment can offer the great opportunity to

develop tools and knowledge for the exploration of rare processes, like VBS.

VBS measurement is extremely challenging, because of its low signal yields, complex final states and large backgrounds. Its understanding requires a coordinated effort of theorists and experimentalists, to explore all possible information about inclusive observables, kinematics and background isolation. The present work wants to contribute to Vector Boson Scattering studies by exploring the possibility to disentangle among W boson polarizations when analyzing a pure VBS sample.

This work is organized as follows. In Chapter 1, we overview the main concepts related to the Standard Model of particle physics. We introduce the VBS process from a theoretical perspective in Chapter 2, underlying its role with respect to the known mechanism of ElectroWeak Symmetry Breaking. We emphasize the importance of regularizing the VBS amplitude by canceling divergences arising from longitudinally polarized vector bosons at high energy. In the same Chapter, we discuss strategies to explore how to identify the contribution of longitudinally polarized W bosons in the VBS process. We investigate the possibility to reconstruct the event kinematics and to thereby develop a technique that would efficiently discriminate between the longitudinal contribution and the rest of the participating processes in the VBS. In Chapter 3, we perform a Montecarlo generator comparison at different orders in perturbation theory, to explore the state-of-art of VBS Montecarlo programs and to provide suggestions and limits to the experimental community. In the last part of the same Chapter we provide an estimation of PDF uncertainty contribution to VBS observables. Chapter 4 introduces the phenomenological study of this work. We perform an extensive study on polarization fraction extraction and on reconstruction of the W boson reference frame. We first make use of traditional kinematic approaches, moving then to a Deep Learning strategy. Finally, in Chapter 5, we test a new technological paradigm, the Quantum Computer, to evaluate its potential in our case study and overall in the HEP sector.

This work has been carried on in the framework of a PhD Executive project, in partnership between the University of Pavia and IBM Italia, and has therefore received supports from both the institutions.
This work has been funded by the European Community via the COST Action *VBSCan*, created with the purpose of connecting all the main players involved in Vector Boson Scattering studies at hadron colliders, gathering a solid and multidisciplinary community and aiming at providing the worldwide phenomenological reference on this fundamental process.

Chapter 1

The Standard Model

The Standard Model (SM) of elementary particles is a Lorentz invariant gauge theory based on the following symmetry group:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.1}$$

It is a quantum field theory that couples quantum mechanics with special relativity, providing a coherent description of strong, weak and electromagnetic forces. It represents our current understanding of elementary particles and their interactions.
At present, as it will be emphasized through this Chapter, all experimental results at colliders show a substantial satisfactory agreement with the SM predictions. Besides the fact that it is proved to be incomplete, the SM is, at our current knowledge, the best self-consistent picture of the fundamental building blocks of matter and their interactions. Across this Chapter, we intend to review its main principles and aspects.

1.1 The Model

1.1.1 Particle Content

The elementary particles which compose the matter sector of the SM are organized in three families of fermions, represented in violet and green colours in Fig. (1.1). Fermions are divided into two groups of six: those that are sensitive to the electromagnetic and weak interactions are called *leptons*, while those affected also by the strong interactions are called *quarks*. The latest are known to be physically bound into triplets and doublets. The triplets are physical particles called *baryons*, while the doublets are referred as *mesons*.
Charged particles can interact through the exchange of photons, the carriers of the electromagnetic force. Photons are massless, uncharged, and have an unlimited range. The gauge theory that describes the interactions of charged

particles through the exchange of photons is known as Quantum ElectroDynamics (QED).
Quarks, furthermore, can interact among other because they possess a characteristic property known as *colour* (or colour charge). Quarks come in three colours: red, green, and blue. Coloured particles are bound together by gluons, the mediators of the strong interaction, and they can also interact with each other, keeping quarks together through their exchange. The peculiar property of gluon self-interaction, which originates from the structure of $SU(3)_C$ gauge group, generates the peculiarities of the strong interactions, such as the so-called confinement that prohibits gluons and quarks to manifest themselves alone outside the hadrons. The gauge theory developed to describe the interactions of coloured particles through the exchange of gluons is known as Quantum ChromoDynamics (QCD).
In analogy, the theory of the weak interaction is sometimes called Quantum FlavourDynamics (QFD), but in practice this definition is rarely used because the weak force is better understood in terms of unified ElectroWeak theory (EW). Weak interactions occur via the exchange of W or Z bosons (in red in Fig. (1.1) like all the other force carriers). This force controls radioactive decays and thermonuclear reactions.

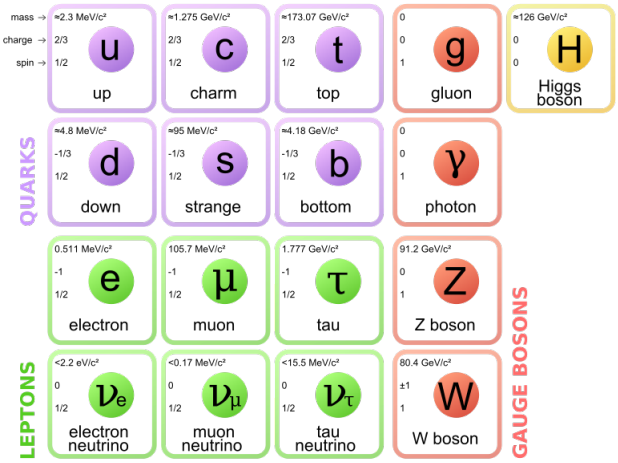


Figure 1.1: The Standard Model of particle physics [Courtesy to Wikipedia: "Standard Model of Elementary Particles" by MissMJ].

Nature is symmetric under the group of Lorentz transformations, rotations[1] and space-time translations which all together form the Poincaré group. Particles are classified according to their properties, in particular by their spin: scalars (spin 0), fermionic spinors (semi-integer spin), and vector bosons (integer spin). They correspond to irreducible representations of the Poincaré group.
While the gauge symmetry of the SM is considered as fundamental (and will be discussed in the next section), there are also accidental symmetries that

[1]Rotations are actually a sub-group of Lorentz transformations.

arise only as a consequence of imposing gauge invariance and renormalizability with a given particle content. This is a very powerful argument when constraining physics beyond the SM, since the new physics can in principle violate accidental symmetries, but the violation is typically subject to strong experimental constraints. Practically, this means that new physics either has to contain these symmetries as accidental symmetries as well, or they have to be promoted to fundamental symmetries.
The SM is a chiral theory i.e. left and right-handed fermion components transform differently under the SM gauge group. The transformation property is determined by the quantum numbers that depend on the representation of the group to which particles belong. The representations of the multiplets with respect to $SU(3)_C$, $SU(2)_L$, $U(1)_Y$ respectively, are for leptons:

$$L^f = (\mathbf{1},\mathbf{2})_{-1/2} = \begin{bmatrix} \nu_L^f \\ l_L^f \end{bmatrix} \qquad l_R^f = (\mathbf{1},\mathbf{1})_{-1} \tag{1.2}$$

where f stays for leptons: $f = e, \mu, \nu$, and we get accidental lepton symmetries L_e, L_μ, L_τ (three lepton numbers) because of SM massless neutrinos. Quarks, instead, are organized as follows:

$$Q_L^f = (\mathbf{3},\mathbf{2})_{1/6} = \begin{bmatrix} u_L^f \\ d_L^f \end{bmatrix} \qquad u_R^f = (\mathbf{3},\mathbf{1})_{2/3} \qquad d_R^f = (\mathbf{3},\mathbf{1})_{-1/3} \tag{1.3}$$

where f refers to the different quarks:

$$\begin{bmatrix} u & c & t \\ d & s & b \end{bmatrix}$$

and we get accidental symmetry B (baryon number), the conservation of which is strongly bound by proton decay[2].

1.1.2 Gauge Structure and Global Symmetry

The gauge structure of the Standard Model is composite and non–abelian. The gauge fields of the Standard Model transform in the adjoint representation of the corresponding simple subgroup. For the moment, we consider both gauge bosons and fermions as massless: any mass term explicitly added to the Lagrangian would indeed break the gauge invariance of the theory. In the following, we will give an explicit formulation for the gauge field evaluating global symmetry properties.

Consider fermion fields described by Dirac spinors (4-component complex objects) ψ, the kinetic term for such fields is given by:

$$\mathcal{L}^{\psi}{}_{kin} = i\bar{\psi}(x)\gamma^\mu\partial_\mu\psi(x); \qquad \bar{\psi}(x) = \psi^\dagger(x)\gamma^0. \tag{1.4}$$

[2]The quark quantum numbers, S, C, B, T corresponding to strange, charm, bottom and top are not conserved in weak interactions.

Global transformations of these fields can be defined as:

$$\psi(x) \to \exp\left(iq^k T^k\right)\psi(x); \qquad \bar{\psi}(x) \to \exp\left(-iq^k T^k \bar{\psi}(x)\right); \tag{1.5}$$

where generators T^k belong to the representation where the fermion transforms and are Hermitian, and q^k are a set of real parameters (therefore $U = \exp(iq^k T^K)$ is a unitary matrix, so $UU^\dagger = 1$). Under these transformations that are space-time independent, the Lagrangian is invariant thanks to the unitarity of the matrix U:

$$\mathcal{L}^\psi{}_{kin} \to i\bar{\psi}(x)U^\dagger\gamma^\mu\partial_\mu U\psi(x) = \mathcal{L}^\psi{}_{kin}\,. \tag{1.6}$$

However when the transformations are space-dependent, the Eq. (1.5) looks different and *local gauge transformations* are introduced:

$$\psi(x) \to \exp\left(iq^k(x)T^k\right)\psi(x); \qquad \bar{\psi}(x) \to \exp\left(-iq^k(x)T^k\right)\bar{\psi}(x). \tag{1.7}$$

Now $U(x) = \exp\left(iq^k(x)T^k\right)$ is a unitary, **space-time dependent** matrix. Under these transformations, the Lagrangian is not invariant anymore. The problem arises with the transformation law of the derivative of the fields:

$$\begin{aligned}\mathcal{L}^\psi{}_{kin} \to i\bar{\psi}(x)U^\dagger(x)\gamma^\mu\partial_\mu(U(x)\psi(x)) = \mathcal{L}^\psi{}_{kin} + \\ i\bar{\psi}(x)\gamma^\mu U^\dagger(x)\partial_\mu U(x)\psi(x)\end{aligned} \tag{1.8}$$

$$\partial_\mu\psi(x) \to \partial_\mu(U(x)\psi(x)) = U(x)\partial_\mu\psi(x) + \partial_\mu U(x)\psi(x) \tag{1.9}$$

To overcome this problem, the *covariant derivative* is introduced:

$$D_\mu\psi(x) = \partial_\mu\psi(x) + igA^k_\mu(x)T^k\psi(x) \to U(x)D_\mu\psi(x). \tag{1.10}$$

Provided that new vector gauge fields are introduced and transform according to:

$$A^k_\mu(x)T^k \to U(x)A^k_\mu(x)T^kU^\dagger(x) + \frac{i}{g}U(x)\partial_\mu U^\dagger(x), \tag{1.11}$$

the invariance under local gauge transformations is restored. In this way we can define an invariant kinetic term for fermions, and this procedure makes the Lagrangian invariant under the local (gauge) transformation. The gauge fields however, do not transform covariantly under the gauge group.
We can construct a covariant quantity out of the gauge fields by noticing that

$$[D_\mu, D_\nu] \to U(x)[D_\mu, D_\nu]U^\dagger(x)\,. \tag{1.12}$$

This quantity is now a differential operator acting on fermions (all derivatives on the fermion fields cancel) and can be written as:

$$\boxed{\begin{aligned}[D_\mu, D_\nu] &= -igF^k_{\mu\nu}T^k \\ F^k_{\mu\nu}T^k &= \left(\partial_\mu A^k_\nu - \partial_\nu A^k_\mu\right)T^k - ig[A^k_\mu T^k, A^l_\nu T^l]\end{aligned}} \tag{1.13}$$

In the abelian case, everything simplifies since the operator has no index and the generators are just equal to 1, so that the commutator in the last term also vanishes.
The gauge-invariant kinetic term for all SM fermions is then given by:

$$\mathcal{L}^{\psi}{}_{kin} = \sum_{f} \sum_{\psi=L,e_R,Q_L,u_R,d_R} i\psi_f\bar{(}x)\gamma^{\mu} D_{\mu}\psi_f(x); \tag{1.14}$$

where the covariant derivative takes the form:

$$D_{\mu}\psi = \left(\partial_{\mu} - igW^{i}_{\mu}T^{i} - ig'YB_{\mu} - ig_S G^{a}_{\mu}T^{a}\right)\psi \tag{1.15}$$

where G^a, B_μ, W_μ are the $SU(3)$, $U(1)$, $SU(2)$ gauge vector fields respectively, a is the index spanning over the color charges of the QCD gauge field tensor, i spans over the weak isospin components, and g, g' and g_S are the coupling constants of the electromagnetic, weak and strong interactions respectively.
The propagation and the interactions of **massless** fermions via gauge vector boson interactions are now described by a gauge invariant kinetic term.
To summarize all the steps above, we can write down the kinetic terms for the gauge fields, according to the group commutation properties:

Field-Strength:

Abelian: $F_{\mu\nu} = \partial_\mu A^k_\nu - \partial_\nu A^k_\mu$
Non-Abelian: $F_{\mu\nu} = \partial_\mu A^k_\nu - \partial_\nu A^k_\mu + g f^{klm} A^l_\mu A^m_\nu$

where f^{klm} is the totally antisymmetric tensor and A^μ are the gauge fields.
After a proper normalization of the kinetic terms, we obtain:

$$\mathcal{L}^{kin}_{V} = \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^{i}_{\mu\nu}W^{\mu\nu i} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu a} \tag{1.16}$$

where $B^{\mu\nu}$, $W^{\mu\nu}$ and $G^{\mu\nu}$ refers to the gauge tensor of the tree generator groups defined in Tab. (1.1).
Notice that the non-covariant transformation law of the gauge fields does not allow for any gauge invariant bilinear in the fields without derivatives. Therefore no gauge boson mass is allowed by gauge invariance. On the fermionic side, instead, since left and right fermion components transform differently under the gauge group, a mass term of the form:

$$m\bar{\psi}\psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$$

is not invariant under gauge transformations:

$$\begin{matrix}\psi_L \to U_L\psi_L \\ \psi_R \to U_R\psi_R\end{matrix} \Longrightarrow \; m\bar{\psi}\psi \; \to m(\bar{\psi}_L U^{\dagger}_L U_R\psi_R + h.c.). \tag{1.17}$$

The left and right matrices are different unitary transformations so that the mass term is not compatible with gauge invariance.

Lie group	Quantum number	Gauge Field
$SU(3)$	color	G^a_μ $a = 1, ..., 8$
$SU(2)$	weak isospin	W^a_μ $a = 1, ..., 3$
$U(1)$	hypercharge	B_μ

Table 1.1: Standard Model gauge structure for the strong, weak and electromagnetic interactions.

A field theory based on an $SU(N)$ gauge invariance predicts the presence of $N^2 - 1$ different vector fields, one for each gauge *degree of freedom*. Thus, the SM involves eight fields for the $SU(3)$ sector, three for the $SU(2)$ sector and one last for the $U(1)$ sector. The gauge fields of the SM are summarized in Table (1.1).

Notice that all we did was to write down gauge-invariant kinetic terms for fermions and gauge bosons. We never attempted at writing down interactions. As a result, we got that from the requirement of gauge invariance of kinetic terms, we *automatically* derived all interactions. Gauge invariance gave us interactions "for free". As a matter of fact, anyway, our theory does not describe the observations at all. Most of the particles that we observe have masses and weak interactions have short-range, i.e. the mediators are massive. The gauge invariance that proved to be so powerful seems to be now the main problem.

1.1.3 Spontaneous Symmetry Breaking

What we have defined so far was necessary but leaves us with the problem of generating masses. In this section we will elaborate on the solution that is provided by the mechanism of Spontaneous Symmetry Breaking (SSB) of gauge invariance [1], [2]. The name sounds a little bit misleading since, as we will see, gauge invariance is not broken at all, it is just "realized" differently. SSB could explain the generation of gauge boson masses both in the case of weak dynamics, where we expect new light degrees of freedom, and in the case of strong dynamics, where we expect to see resonances close to the cut-off scale. In the case of weak dynamics, SSB is said to be *linearly* realized, while in the case of strong dynamics it is said to be *non-linearly* realized.

To better understand the details about the SSB we would like to introduce a fundamental theorem which goes under the name of Goldstone Theorem [3].

Goldstone Theorem

A spontaneously broken global symmetry implies the existence of massless scalars called Goldstone bosons. There are as many Goldstone bosons as broken symmetry generators.

Examples

QCD with 2 flavors (u, d) quarks:

$$SU(2)_L \otimes SU(2)_R \to SU(2)_{L+R} = SU(2)_V.$$

Each $SU(2)$ group has three generators, therefore in total three generators are broken so we have 3 massless Goldstone bosons: $\pi^0, \pi^\pm$, triplet of $SU(2)_V$. The symmetry is approximate and it is broken by light quark masses and therefore the pions are not exactly massless. Since the quark masses are much lighter than the QCD confinement scale ($m_u \approx 2$ MeV, $m_d \approx 5$ MeV) the symmetry is a good approximation and in fact pions are lighter ($m_\pi \approx 135$ MeV) than the other QCD resonances.
The same approach can be used in case of:

$$SU(3)_L \otimes SU(3)_R \to SU(3)_{L+R} = SU(3)_V;$$

$SU(3)$ has eight generators, therefore in total eight massless Goldstone bosons. Now, while this symmetry is a very good approximation for the two light quarks, it is less exact for the strange quark ($m_s \approx 100$ MeV, $m_k \approx 500$ MeV) so that strange mesons are heavier.

Continuous Global Symmetry

The simplest example of SSB is a self-interacting real scalar field filling the vacuum with discrete Z_2 symmetry. Its formulation can be found in App. (A.1). A slightly more illustrative example is the one related to a global continuous symmetry.
Consider the following Lagrangian for a complex scalar field (2 real components):

$$\mathcal{L}_\phi = \partial_\mu \phi^\dagger \partial^\mu \phi - V(\phi); \qquad V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda}{4}(\phi^\dagger \phi)^2$$

We need to impose the condition $\lambda \geq 0$, for the potential to be bounded from below. This new theory is invariant under continuous $U(1) \sim SO(2)$ transformations:

$$\phi \to U\phi; \qquad \phi^\dagger \to U^\dagger \phi^\dagger; \qquad U = e^{i\alpha}.$$

Looking at the minimum of the potential we get:

$$\frac{\partial V}{\partial \varphi} = 0 \quad \Rightarrow \quad \begin{cases} a)\ \mu^2 > 0 & \langle \phi \rangle = 0 \\ b)\ \mu^2 < 0 & \langle \phi \rangle = \sqrt{\frac{-2\mu^2}{\lambda}} e^{i\theta} \end{cases}$$

Case a) represents a single (symmetric) vacuum state with two degenerate states with masses $m^2 = V''(\langle \phi \rangle) = \mu^2$.
Case b) gives us an infinite class of equivalent vacuum states the spectrum of which consists of a massive and a massless real scalar.

Looking closer at the broken phase and expanding the field as quantum fluctuations around a classical vacuum expectation value (where the potential is minimum), one obtains, using a parametrization of the complex field as an argument, σ, and a phase π:

$$\phi = \left(\sqrt{\frac{-2\mu^2}{\lambda}} + \frac{1}{\sqrt{2}}\sigma\right) e^{\frac{i\pi}{f_\pi}}.$$

The potential becomes:

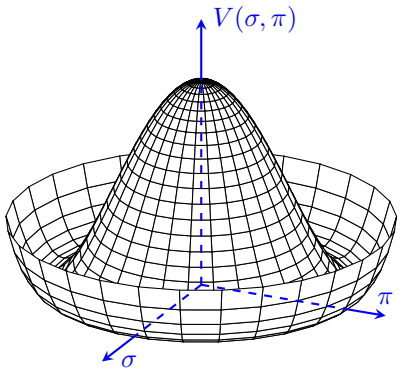
$$\begin{aligned} V(\phi) = \mu^2 \left(\sqrt{\frac{-2\mu^2}{\lambda}} + \frac{1}{\sqrt{2}}\sigma\right)^2 + \frac{\lambda}{4}\left(\sqrt{\frac{-2\mu^2}{\lambda}} + \frac{1}{\sqrt{2}}\sigma\right)^4 \\ \simeq -\frac{\mu}{4} - \mu^2\sigma^2 + \frac{\sqrt{-\lambda\mu^2}}{2}\sigma^3 + \frac{\lambda}{16}\sigma^4; \end{aligned}$$

where only σ has a potential and has mass $m_\sigma^2 = -2\mu^2$ but π cannot have disappeared. We should also look at the kinetic terms, so the Lagrangian reads:

$$\mathcal{L}_\phi = \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma + \left(\sqrt{\frac{-2\mu^2}{\lambda}} + \frac{1}{\sqrt{2}}\sigma\right)\frac{1}{f_\pi^2}\partial_\mu\pi\partial^\mu\pi - V(\phi). \tag{1.18}$$

The arbitrary dimensionful constant f_π can be fixed requiring π to have a normalized kinetic term

$$f_\pi = 2\sqrt{\frac{-\mu^2}{\lambda}} = \sqrt{2}|\langle\phi\rangle| = v.$$



Therefore the field π propagates and interacts with σ only with derivative couplings, but does not have any potential (it is massless). The field is said to be a *Goldstone Boson* (GB). Now, the original symmetry is realized as a shift symmetry $\pi \to \pi + c$, that is the Nambu-Goldstone realization.

The non-linear Realization

To identify the Goldstone boson and the radial excitation it is useful to consider the limit:

$$\mu \to \infty; \qquad \lambda \to \infty; \qquad v = 0.$$

In this limit the radial mode decouples and the Lagrangian reduces to the free theory of a scalar with a shift symmetry:

$$\mathcal{L} = \frac{1}{2}\partial_\mu \pi \partial^\mu \pi.$$

This is a "trivial" example of what is known as a *non-linear* sigma model where SSB is said to be realized non-linearly and the radial degree of freedom is heavy and it is not part of the light spectrum of the theory. *Non-linear* means that now the field ϕ is just a non-linear function of the Goldstone bosons:

$$\phi = \sqrt{\frac{-2\mu^2}{\lambda}} e^{\frac{i\pi}{v}}.$$

Linear and *non-linear* realizations can be distinguished by looking at interactions.

We have seen so far the implications of SSB on global discrete and continuous symmetries. The Goldstone Theorem is a crucial result in QFT and predicts the existence of massless, or approximately massless, scalars. In the context of SSB, gauge invariance is radically different and we can understand this by just counting degrees of freedom:

Unbroken phase: Massless vector has 2 physical degrees of freedom. They correspond for instance to the two transverse polarizations of photons.

Broken phase: Massive vector has 3 physical degrees of freedom. There is a new longitudinal polarization for massive vectors.

In the transition from the unbroken to the broken phase, it seems that SSB can change the number of degrees of freedom. There should be a reconciliation that preserve the number of *dof*. It follows that the Goldstone Theorem is violated by SSB of "local" (gauge) invariance. The Goldstone Bosons are not anymore new massless degrees of freedom, but provide the degrees of freedom corresponding to the longitudinal polarizations of gauge bosons. This is at the basis of the Higgs mechanism [2].
In the Standard Model, the SSB is linearly realized through weak dynamics and the light degree of freedom is the **Higgs boson**[3].

[3]Notice that this does not exclude that the Higgs boson itself cannot be generated by some other strong dynamics responsible for EWSB.

1.1.4 SSB of gauge invariance: scalar QED

SSB of gauge invariance is very simply understood in a model of scalar QED, that is the electrodynamics of a charged scalar (Abelian Higgs model):

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger D_\mu\phi - V(\phi); \qquad V(\phi) = -\mu^2|\phi|^2 + \frac{\lambda}{4}|\phi|^4$$
$$D_\mu\phi = (\partial_\mu + ieA_\mu)\,\phi; \qquad (D_\mu\phi)^\dagger = (\partial_\mu - ieA_\mu)\,\phi^\dagger.$$

This theory is invariant under $U(1)$ gauge transformations:

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu q(x); \qquad \phi \to e^{iq(x)}\phi; \qquad \phi^\dagger \to e^{-iq(x)}\phi^\dagger;$$

where we have chosen the sign of the mass term in such a way that for $\mu^2 > 0$ SSB occurs as in the case of the linear sigma model:

$$|\langle\phi\rangle| = \sqrt{\frac{2\mu^2}{\lambda}} = \frac{c}{\sqrt{2}}.$$

Expanding the field around the vacuum configuration as:

$$\phi = \left(\frac{v+\sigma}{\sqrt{2}}\right)e^{\frac{i\pi}{v}};$$

we get the Lagrangian:

$$\begin{aligned}\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \left(\frac{v+\sigma}{\sqrt{2}}\right)^2\left(\frac{\partial_\mu\sigma}{v+\sigma} - i\frac{\partial_\mu\pi}{v} - ieA_\mu\right)\\ \left(\frac{\partial^\mu\sigma}{v+\sigma} + i\frac{\partial^\mu\pi}{v} + ieA_\mu\right) - \left(-\frac{\mu}{4} + \mu^2\sigma^2 + \frac{\sqrt{\lambda}\mu}{2}\sigma^3 + \frac{\lambda}{16}\right).\end{aligned} \tag{1.19}$$

As expected, interactions of the Goldstone boson are again only derivative interactions. In order to understand the spectrum, let us focus on the quadratic part of the Lagrangian:

$$\begin{aligned}\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma + \frac{1}{2}\partial_\mu\pi\partial^\mu\pi\\ + \frac{e^2v^2}{2}A_\mu A^\mu + ev\partial_\mu\pi A^\mu - \mu^2\sigma^2.\end{aligned} \tag{1.20}$$

It follows that we have successfully generated a mass term for the gauge boson, the kinetic terms are canonical, but we have an unpleasant kinetic mixing between the Goldstone boson and the gauge field. This, of course, can be removed by properly choosing a gauge.
Let's add to the original Lagrangian a gauge-fixing term (a generalisation of the Lorentz gauge):

$$\mathcal{L} = -\frac{1}{2\xi}\left(\partial_\mu A_\mu - \xi ev\pi\right)^2. \tag{1.21}$$

With the introduction of this gauge-fixing, the quadratic part of the Lagrangian becomes:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma + \frac{1}{2}\partial_\mu\pi\partial^\mu\pi + \frac{e^2v^2}{2}A_\mu A^\mu + \cancel{ev\partial_\mu\pi A^\mu} - m^2\sigma^2 + \frac{1}{2\xi}\left(\partial_\mu A_\mu\right)^2 + \frac{\xi^2}{2}e^2v^2\pi^2 + \cancel{ev\partial_\mu A^\mu\pi} \tag{1.22}$$

Now, the two kinetic mixing terms are equal, up to an integration by parts, where the total derivative is irrelevant, so they cancel out.
In order to understand the spectrum, it is useful to write down the propagators of the three fields appearing in our Lagrangian:

$$\frac{i}{k^2 - e^2v^2}\left(-g_{\mu\nu}(1-\xi)\frac{k_\mu k_\nu}{k^2 - \xi e^2v^2}\right)$$

$$\frac{i}{k^2 - 2\mu^2}$$

$$\frac{i}{k^2 - \xi e^2v^2}$$

As a result we can summarize the main points:

- The gauge boson has acquired a mass $m_A = ev$;
- The sigma field, as before, has a mass $m_\sigma = \sqrt{2}\mu$;
- The mass of the Goldstone boson π depends on the gauge parameter ξ, signaling the fact that this field cannot be a physical independent degree of freedom anymore;
- One can see that π can be completely removed from the spectrum by going in the so-called unitary gauge, which corresponds to $\xi \to \infty$. In this gauge, it becomes infinitely heavy and its propagator vanishes, while the boson propagator goes back to the propagator of the massive vector that we introduced previously;
- This shows how the longitudinal polarization of the gauge field and the Goldstone boson are tied together (the GB is "eaten-up" by the corresponding Gauge boson).

Of course, there exist other known gauge choices, e.g. *Landau Gauge* $\xi = 0$, and *Feynman Gauge* $\xi = 1$. All ξ's gauges are called **renormalizable gauges or 't Hooft gauges**.

The important conclusion here is that SSB of gauge invariance is the solution to the problem of generating gauge boson masses. As discussed above, gauge invariance is not actually broken, but it is realized via Nambu-Goldstone mechanism. The Goldstone Boson is eaten up by the corresponding gauge field providing its longitudinal component and it is not anymore an independent degree of freedom.

1.1.5 SSB of gauge invariance in the SM

As a natural step, we would like to complete our study and determine the SSB for the SM. The ElectroWeak Symmetry Breaking (EWSB), represented by this transformation $SU(2)_L \otimes U(1)_Y \to U(1)_q$ is the non-abelian generalization of the abelian case we have just seen.
Considering a Lagrangian density of a free, massive, electromagnetic-charged, spin-1/2 particle [4]:

$$\mathcal{L} = i\overline{\Psi}\gamma^\mu\partial_\mu\Psi - m\overline{\Psi}\Psi; \tag{1.23}$$

where $\Psi \equiv \Psi(x)$ and x is the space-time four-vector; the first and the second terms are the kinetic and the mass term respectively.
As the SM is a chiral theory, for each spinor it is possible to define a left-handed (LH) and a right-handed (RH) part, which behave differently under the SM gauge group. For this reason it is usually more convenient to adopt a two-component Weyl spinor representation instead of a four-component Dirac field, since a Dirac spinor Ψ is composed of a left-handed Weyl spinor Ψ_L and a right-handed Weyl spinor Ψ_R,

$$[\Psi] = \begin{pmatrix} \Psi_L \\ \Psi_R \end{pmatrix}.$$

As anticipated above, leptons and quarks are grouped into generations or families of left-handed $SU(2)_L$-doublets of *weak isospin* **T**, three in the lepton sector, $(\nu_l, l)_L$, with $l = e; \mu; \tau$; and three in the quark sector, $(u, d')_L; (c, s')_L; (t, b')_L$ (see Eq. (1.2,1.3)) in the same way there are nine right-handed $U(1)_R-$singlets of *hypercharge* Y composed by the charged leptons l_R, up-like quarks u_R and down-like quarks d'_R. Now we require, as we did several times before, that the Lagrangian remains unchanged under the *global* and *local* gauge or phase transformation.
To reach the *unification*, the same concepts used previously for the electromagnetic interactions have to be extended to the weak interactions and a symmetry breaking sector must be added.
The local gauge transformations of the weak hypercharge Y unitary group in one dimension $U(1)_Y$ and the local gauge transformations of the weak isospin **T** special unitary group in two dimensions $SU(2)_L$ are:

$$U(1)_Y = \exp\left(i\vartheta(x)\frac{Y}{2}\right); \qquad SU(2)_L = \exp\left(i\vec{\alpha}(x)\cdot\vec{T}\right);$$

where $U(1)_Y$ group is equal to the $U(1)_q$ previously considered in QED, while the $SU(2)_L$ is not Abelian since the generators of the group[4] do not commute and the transformations are *special*[5].
The composition gives the group $SU(2)_L \otimes U(1)_Y$ of the local gauge transformations of the weak isospin and hypercharge, for hypothetically 0-mass particles. Now we can add to the SM Lagrangian of Eq. (1.19) a doublet H of $SU(2)_L$ fields with hypercharge Y= 1/2 described by the following Lagrangian:

$$\begin{aligned} \mathcal{L}_{\mathcal{H}} = (D_\mu H)^\dagger D_\mu H - V(H); \qquad V(H) = -\mu^2 H^\dagger H + \lambda (H^\dagger H)^2 \\ D_\mu H = \left(\partial_\mu - i\frac{g}{2} W_\mu^i \sigma^i - i\frac{g'}{2} B_\mu \right) H. \end{aligned} \tag{1.24}$$

We choose to parametrize the field in a way analogous to the previous linear sigma model example through a radial excitation times a phase:

$$H = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} e^{\frac{i\pi^a \sigma^a}{v}}. \tag{1.25}$$

From the kinetic term of $\mathcal{L}_H$ we immediately find the gauge boson spectrum:

$$\mathcal{L}_{mass} = \frac{g^2 v^2}{2} \left((W_\mu^1)^2 + (W_\mu^2)^2 + \left(\frac{g'}{g} B_\mu - W_\mu^3 \right)^2 \right). \tag{1.26}$$

Looking at the previous formula we see that the states W^i and B are neither mass nor charge eigenstates. The fields W_μ^i, W_μ^2 do not have definite charge, but can be combined to give the charge eigenstates in this way:

$$\boxed{W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right)} \tag{1.27}$$

The fields W^3 and B instead are both neutral, but have a mass mixing, and they need to be diagonalized to determine the spectrum. The relations between the weak and the electromagnetic coupling constants suggest the construction of a unified ElectroWeak theory with a free parameter ϑ_W (Weinberg angle) which is related to the weak isospin coupling g and to the weak hypercharge g' by the following relations:

$$\cos\vartheta_W = \frac{g}{\sqrt{g^2 + g'^2}} \qquad \text{and} \qquad \sin\vartheta_W = \frac{g'}{\sqrt{g^2 + g'^2}}.$$

A simple 2×2 rotation of ϑ_W, is sufficient to identify the photon and the Z boson

$$\boxed{\begin{aligned} A_\mu &= \sin\vartheta_W W_\mu^3 + \cos\vartheta_W B_\mu \qquad (1.28) \\ Z_\mu &= \cos\vartheta_W W_\mu^3 - \sin\vartheta_W B_\mu \qquad (1.29) \end{aligned}}$$

[4] We are referring to the 2×2 Pauli matrices $T_k = \tau_k/2$.
[5] I.e. with $det(U) = 1$, or similarly $Tr(T_k) = 0$.

In terms of g and g' couplings or, alternatively in terms of gauge couplings, the gauge boson masses can be written:

$$m_W = \frac{gv}{2} \approx 80\,\text{GeV}; \qquad m_Z = \frac{\sqrt{g^2 + g'^2}v}{2} = \frac{m_W}{\cos\vartheta_W} \approx 91\,\text{GeV}.$$

For contributions to the above unification of the weak and electromagnetic interactions between elementary particles, called GSW theory, Abdus Salam, Sheldon Glashow and Steven Weinberg were awarded the Nobel Prize in Physics in 1979.
The existence of the ElectroWeak interactions was experimentally established in two stages, the first being the discovery of neutral currents in neutrino scattering by the Gargamelle collaboration in 1973, and the second in 1983 the discovery of the W and Z gauge bosons at the CERN proton-antiproton collider by the UA1 and the UA2 collaborations. For the discovery of the ElectroWeak mediators, C. Rubbia and S. Van der Meer were awarded the Nobel Prize in Physics in 1984. In addition, in 1999, Gerardus'T Hooft and Martinus Veltman were also awarded the Nobel prize for showing that the ElectroWeak theory is renormalizable.

Higgs interactions

Looking at the Lagrangian involving the Higgs boson, we find:

$$\begin{aligned}\mathcal{L}_H = &\frac{1}{2}\partial_\mu h \partial^\mu h - \frac{m_h^2}{2}h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 + \\ &\left(m_W^2 W_\mu^+ W^{\mu-} + \frac{m_Z^2}{2} Z_\mu Z^\mu\right)\left(\frac{h^2}{v^2} + \frac{2h}{v}\right).\end{aligned} \tag{1.30}$$

From Eq. (1.30) we can observe that the Higgs field couples to gauge bosons proportionally to their masses, and this is a key prediction of EWSB. The Lagrangian has several independent parameters to be determined from experiments, namely:

m_H The Higgs mass (measured now at 0.1% precision [5]);

$m_{W,Z}$ The W and Z boson masses, connected to the ϑ_W by $\sin^2\vartheta_W = 1 - (m_w^2/m_z^2) \simeq 0.229(30)$. They are known with very good accuracy from LEP and LHC precision measurements and are defined to be EW precision observables;

v This parameter represents the vacuum expectation value of the Higgs field and it is experimentally connected to the Fermi constant $G_F = 1/\left(\sqrt{2}v^2\right)$, determined from muon decay;

λ This parameter governs Higgs self-interactions and it is related in the SM to the mass and the v of the Higgs boson. It is still directly unmeasured and it represents a missing (challenging) test of the SM, from the experimental point of view.

At this point we have successfully found a mechanism to generate gauge boson masses, yet we know that also SM fermions have a mass. Notice that after the introduction of the Higgs field, we can write new renormalizable interactions between it and (some of) the SM fermions, via the so-called Yukawa vertices, for instance:

$$\mathcal{L}_{Yuk} = -y\bar{L}He_R + h.c. \Longrightarrow$$
$$-y\frac{h+v}{\sqrt{2}}\bar{e}e = -m_e\left(1+\frac{h}{v}\right)\bar{e}e.$$

As a result, after the SSB, also fermions acquire masses, proportional to their Yukawa coupling $m = \frac{yv}{\sqrt{2}}$. As we anticipated, Higgs interacts proportionally to the mass (squared mass for bosons) of the particle[6].

SM non-linear Realization

We have cosidered before that SSB can be realized non-linearly, with the radial mode decoupled from the light spectrum. Experimental evidences [6] show that the Higgs boson exists and is light ($m_H = 125,18 \pm 0,16$ GeV [7]), however we could ask ourselves if it is really the radial mode or whether EWSB is realized *non-linearly* and the Higgs is just an accidentally light scalar (singlet under the SM). The rest of the Lagrangian would remain the same, but Higgs boson interactions would not be predicted anymore by EWSB (they would be free parameters).

In Fig. (1.2), we can appreciate how measurements of the Higgs normalized couplings[7] to SM particles, measured at the LHC, reproduce very well the SM expectation. The relation between the masses of the SM particles and their interactions with the Higgs boson is linear, suggesting that the linear-realization is, at least to a very good level of accuracy, correct.

With the introduction of the Higgs field and the assumption that SSB occurs, we have solved several problems at once. We have successfully generated gauge boson masses in a way compatible with gauge invariance, and we have generated fermion masses proportional to new Yukawa couplings. The Higgs boson is predicted to couple to all other particles proportionally to their masses. Most of these predictions have, by now, been verified experimentally.

We conclude this section reporting the compact version of the SM Lagrangian, including the Higgs mechanism:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + D_\mu\Phi^\dagger D^\mu\Phi - V(\Phi) + \bar{\Psi_L}\hat{Y}\Phi\Psi_R + h.c. \tag{1.31}$$

[6] Notice that some Yukawa couplings would need the Higgs to have opposite hypercharge to be invariant.

[7] This statement is different with respect to the particle families considered so that the proportionality could be linear or quadratical.

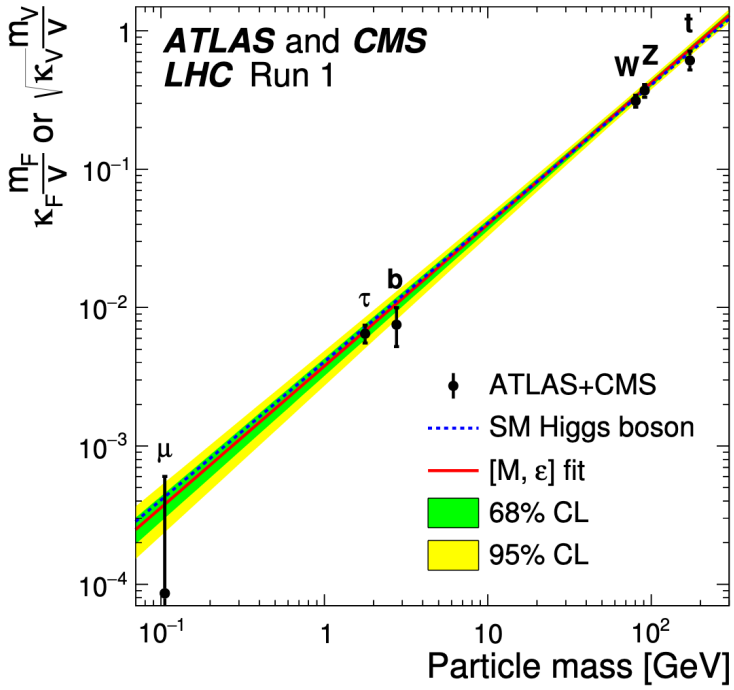


Figure 1.2: Best fit values as a function of particle mass for the combination of ATLAS and CMS Run 1 data. The dashed (blue) line indicates the predicted dependence on the normalized particle mass in the case of the SM Higgs boson [8].

In Chapter 3 we will discuss how rare processes, as the VBS, could help physicists to solve some important missing parts of the EWSB mechanism: the measurement of the Higgs self-coupling and of the Yukawa couplings with fermions.

1.2 The QCD and EW Sector of the SM

In this section, we will discuss two main approaches to SM that allow physicists to perform calculation and estimation about physics observables. Since we presented already QED and EW interactions, here we would like to elaborate more some *Quantum ChromoDynamics* concepts.
When probed at very short wavelengths, QCD is essentially a theory of free "partons", quarks and gluons. They scatter off one another through relatively small quantum corrections, that can be systematically calculated in perturbative expansion of the strong coupling α_s.
At longer wavelengths, of the order of the proton size $\sim 1\ fm = 10^{-15}\ m$, however, we see evidences of hadron resonances, with string-like potentials, if we try to separate their partonic constituents. In this regime, the running coupling constant is too big to act as a parameter for the expansion. Due to our inability to perform analytic calculations in strongly coupled field theories, QCD is therefore still only partially solved. Nonetheless, all its features, across

all distance scales, are believed to be encoded in a single one-line formula of alluring simplicity: the Lagrangian of QCD, reported here below, in Eq. (1.32). The consequence for collider physics is that some parts of QCD can be calculated in perturbation theory, whereas others (the non perturbative parts) must be expressed through models or functions whose effective parameters are not a priori calculable but which can be constrained by fits to data.
For its features, QCD can be inferred from a few general requirements and basic observations. A simplified list reads as follows:

- ◇ Our understanding of chiral symmetry as a *spontaneously* realized one which allows treating pions as Goldstone bosons implying various soft pion theorems, requires vector couplings for the gluons.
- ◇ The ratio $R = \sigma(e^+e^- \longrightarrow had.)/\sigma(e^+e^- \longrightarrow \mu^+\mu^-)$ and the amplitude decays Γ of $\pi^0 \to \gamma\gamma, \tau^- \to e^-\overline{\nu}_e\nu_\tau$ and $B \to l\upsilon X_c$ point to the need for three colour charges. Moreover, the presence of an even function describing antisymmetric particles like in the $J^P = 3/2^+$ lowest mass baryons decuplet, for example the Δ^{++} and the Ω^-, makes this requirement a necessity.
- ◇ Colour has to be implemented as an *un*broken symmetry.
- ◇ Combining confinement with asymptotic freedom requires a *non-abelian* gauge theory.

All of the above, require QCD to be described by the gauge group $SU(3)$, the Special Unitary group in 3 (complex) dimensions.
In the context of QCD, we represent this group as a set of unitary 3×3 unitary matrices (the Gell-Mann matrices). This is called the adjoint representation and can be used to represent gluons in colour space. Since there are 9 linearly independent unitary complex matrices, the determinant -1 acts as a constraint and consequently there are a total of 8 independent directions in the adjoint colour space, i.e., the gluons are arranged in a colour octet. In QCD, these matrices can operate both on each other (gluon self-interactions) and on a set of complex 3-vectors, the fundamental representation of quarks in the colour space. The fundamental representation has one linearly independent basis vector per degree of $SU(3)$, and hence the quarks are triplets.
The QCD density of Lagrangian can be written as:

$$\mathcal{L}_{QCD} = -\frac{1}{4}G_a^{\mu\nu}G^a_{\mu\nu} + \sum_f \overline{q}_f(i\gamma^\mu D_\mu - m_f)q_f. \tag{1.32}$$

where $G^{\mu\nu}$ is the gauge field tensor, a is the colour index, and D_μ is the covariant derivative introduced before.
Despite the conciseness of this formula, the QCD theory conceals many pitfalls that deserve to be analyzed in more detail [9]. One important property of QCD

is that the coupling constant decreases when the momentum transferred in the interaction increases or, in other words, at short distances, oppositely to what happens in QED interactions. This particular behaviour allows us to fully understand the QCD in a perturbatively approach, namely for hard processes. That is to say that the larger is the distance, the stronger is the interaction. This propriety is well-known as *asymptotic freedom*.
While in case of QED interactions, the running of α_{em} with the energy scale produces a *screening effect* of the charge (this means that at large distances, the electric charge is screened while its real *bare* value can be reached getting closer to it), in QCD the strong attractive force gets weaker the closer the particles is. The net effect of polarization of virtual gluons in the vacuum does not screen the field but it augments it. This is called *anti screening effect*. There is therefore a clear difference between QCD and QED, which is ultimately related to their non-Abelian/Abelian nature: photons do not interact with other photons, but gluons can interact with and even create/absorb other gluons.

In addition, in high-energy scattering involving hadrons in the initial state, we immediately face the complication that hadrons are composite, with a time-evolving structure; there are partons within clouds of further partons, constantly being emitted and absorbed. Thus, before we can use perturbatively calculated partonic scattering matrix elements, we must first address the partonic structure of the colliding hadron(s).
What happens at larger distance can be explained by saying that when the distance increases, the colour field can be seen as a tube connecting the quarks, thus it becomes energetically favorable to convert the (increasing) energy stored in the colour tube to a new $q\overline{q}$ *pair*. This kind of processes (and in general the phenomenology of confinement) can be described by lattice calculations or bag models, inspired by QCD.
However, since the interactions between quarks and gluons become weaker at small distances, it might be possible, by creating a high density/temperature extended system composed by a large number of quarks and gluons, to create a "deconfined" phase of matter and this is what the ALICE experiment at LHC is pursuing.

In the following two paragraphs, we will detail the perturbative and non perturbative parts of the calculation of any observable analyzed at an hadron collider.

1.2.1 Perturbative Ingredients

According to what we presented in the previous Sections, the Standard Model is based on a symmetry groups that describes 3 interactions. When the coupling constants of EW interactions and of QCD at small distances are small, they can be used as the power expansion to compute most of the high-energy observables. Indeed, the principal technique for computations of particle scatterings at high energy is *perturbation expansion*.

In the formalism of perturbation theory, any observable, like the cross section, for instance, can be calculated by:

$$\hat{\sigma}^{k+n} = \alpha^k \sum_{n=0}^{k} C^{(n)} \alpha^n; \tag{1.33}$$

where α represents the coupling constant. When $n = 0$ we find the so called Leading Order (LO) process, the lowest order possible at which the process can occur.
In a LO diagram we have vertices that bring an α contribution each in the calculation, as shown, for instance, in Fig. (1.3).
The local gauge invariance imposed to our Lagrangian provides the main ingredient of our theory: the existence of a conserved quantity given by the conserved symmetry. Considering the QED Lagrangian in $U(1)$ group theory, for instance, we obtain the conservation of electric charge. Each vertex in a tree-level diagram then contributes with a square charge e that is then multiplied by the charged or neutral current (vectorial) part of the process.
Without repeating all the steps, we could arrive at the same conclusion also for QCD, with the difference that $SU(3)_C$ is not abelian, its generators do not commute so that the gluons are charged and can interact together. A QCD

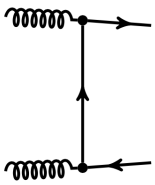


Figure 1.3: A generic leading order process.

tree-level diagram is proportional to α_s^n where n represents the number of vertices in the lowest order process.
Considering a LO process diagram, the computation of a LO cross section, the probability distribution of observing a given event can be computed as:

$$\sigma = \frac{1}{F} \int dPSP |\mathcal{M}|^2; \tag{1.34}$$

where:

F: flux factor $F = |\mathbf{v_a} - \mathbf{v_b}| 2E_a 2E_b$, where $\mathbf{v_i} = \mathbf{p_i}/E_i$;

dPSP: phase-space weight with parton-distribution function (PDF) factors;

$\mathcal{M}$: matrix element.

NLO corrections

A LO diagram in which an additional vertex is added is called Next-to-Leading Order (NLO) diagram. In this section we want to look into LO and NLO corrections, both in QCD and EW power expansion. ElectroWeak interactions are carried by $W^{\pm}, Z$, γ and $Higgs$ gauge bosons. The coupling constant is related to the $U(1)$ conserved quantity, the electric charge e, by the following formula:

$$\alpha = e^2/4\pi \qquad \text{where} \qquad e = \frac{gg'}{\sqrt{g^2 g'^2}}$$

where g and g' are the weak isospin coupling and the weak hypercharge coupling respectively, as discussed above. Higher order corrections include both real and virtual contributions, as shown in Fig. (1.4) on the left and right respectively. In the left diagram, an emission of a real photon off one of the two outgoing fermions is represented. This is a NLO correction to the LO Born diagram. In the right diagram, a Next-to-Next-to-Leading Order (NNLO) correction is represented, namely the exchange of a virtual photon from the two fermionic legs. This diagram has an α^2 contribution (one α for each additional vertex) with respect to the LO process. EW corrections account for both QED

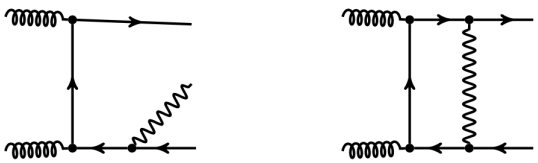


Figure 1.4: Higher order process: real emission of a photon at NLO (left), exchange of a virtual photon from final fermions at NNLO (right).

and weak contributions, i.e. in the diagrams above, the emitted and exchanged particle can also be a gluon. These contributions to the observable can be analytically computed in perturbation QCD and EW theory.
The matrix element evaluation to be used in Eq. (1.34 for a general QCD process, with an EW correction added is:

$$\begin{aligned} |\mathcal{M}|^2 &= \mathcal{M} \times \mathcal{M}^* \\ &= (\mathcal{M}_b + \mathcal{M}_v)\,(\mathcal{M}_b^* + \mathcal{M}_v^*) \\ &= |\mathcal{M}_b|^2 + \mathcal{M}_v\mathcal{M}_b^* + \mathcal{M}_b\mathcal{M}_v^* + |\mathcal{M}_v|^2 \\ &= \underbrace{|\mathcal{M}_b|^2}_{\sim\alpha_s^2(LO)} + \underbrace{2Re\,\{\mathcal{M}_b\mathcal{M}_v^*\}}_{\sim\alpha\alpha_s^2(NLO)} + \underbrace{|\mathcal{M}_b|^2}_{\sim\alpha^2\alpha_s^2(NNLO)} \end{aligned} \tag{1.35}$$

where the index b refers to Born process and v indicates the virtual contribution.
QCD corrections can be obtained in a similar way, replacing the coupling constant $\alpha \to \alpha_s$, the photon (γ) with the gluon (g) and electric charge with colour, and keeping in mind that three or four-gluons vertices are allowed.

Correction Properties

There exist situations in which the matrix element squared of formula (1.35) becomes infinite. This is the case of the so called *infrared* (IR) singularities, which are obviously not physical. An infrared observable is said to be *safe* if it is insensitive to soft radiation, so that adding any number of infinitely soft particles should not change the value of the observable. This means that it can be calculated reliably using perturbative methods. In case of EW corrections, it happens when the photon is soft ($p \to 0$) or collinear ($p_T \to 0$).
The way these singularities are cancelled out is by the inclusion of real corrections through the Bloch-Nordsiek/KLN theorem. The first theorem (1937) states that soft IR divergences cancel in QED when summing over final state photons with finite energy resolution, whereas the KLN theorem (1962-64) states that S-matrix elements squared are IR finite when summing over final states (f) and initial states (i) within some energy window:

$$\sum_{f,i\in[E-E_0,E+E_0]} |\langle f|S|i\rangle^2| < \infty.$$

This condition has been put in a stronger version in 2018:

Strong KLN Theorem
S-matrix elements squared are IR finite when summing overfinal states or initial states:
$\sum_f |\langle f|S|i\rangle^2 = \langle i|S^\dagger \sum_f |f\rangle\langle f|S|i\rangle = \langle i|i\rangle \propto 1 < \infty.$

For unitarity reasons we can state that all probabilities sum to one that is IR finite and all diagrams (to any non-zero order α) sum to zero.

If we aim at constructing IR-safe objects, we define *dressed fermions* when we consider only photon radiation from quarks line. Diagrams squared are then of order: $|\text{real}|^2 \sim \alpha_s^2\alpha = O(\text{LO}) \times \alpha$.
The real correction has $n+1$ particles in the final state:

$$\delta\sigma^{NLO} = \int d\phi_{n+1}\mathcal{M}^2_{0,n+1} + 2\int d\phi_n Re\left[\mathcal{M}^*_{1,n}\mathcal{M}_{0,n}\right]. \tag{1.36}$$

In addition to what has been introduced above, EW corrections are not necessarily made of EW radiations (real) and the inclusion of EW particle *interferences* must be included also in order to obtain an IR finite result.
Up to now we understood (without proofing) that adding real and virtual corrections ensures IR finiteness, however (as in QCD), other type of divergences can arise when computing virtual corrections: *ultraviolet* (UV) divergences. They can arise in the loops (emission and absorption of a particle) when the momentum which enters goes into infinity.
These (unphysical) divergences are removed by the procedure of *renormalization*. A well detailed explanation of this procedure can be found in [10].

As a general rule, in order to get a finite contribution for (NLO) EW corrections computation one follows this scheme: compute real corrections, compute virtual corrections, possibly including interference corrections and finally compute UV counterterms.
Previous conclusions are partial because, if we compute radiations of massive EW gauge boson, we obtain terms such as: LO $\times \log(M_v^2/s)$ and LO $\times \log^2(M_v^2/s)$ which are divergent for low s, center of mass energy of the process. These logarithms are defined *Sudakov logarithms*.

1.2.2 Non Perturbative Ingredients

At this point it is clear that the way particles interact depends on their energy. For most of the calculations, we operate in relativistic regime and high transverse momentum limit, where partons behave as quasi free in the collision. But partons are part of the colliding protons, therefore the perturbative matrix-element calculation representing the hard scattering, must be convoluted with the probability to extract the partons from the nucleons.
We can write a generic cross section for an hard process in hadron-hadron collisions as:

$$\boxed{\sigma = \int dx_1 f_{q/p}(x_1, \mu^2) \int dx_2 f_{q/p}(x_2, \mu^2) \hat{\sigma}\left(x_1 p, x_2 p, \mu^2\right)} \tag{1.37}$$

where $\hat{s} = x_1 x_2 s$ is the cross-section of the partonic process and μ^2 represents the energy scale, p is the incoming hadronic momentum and x_1, x_2 are the fraction of these momenta carried by the colliding partons. The total cross-section is factorized into an hard part $\hat{\sigma}\left(x_1 p, x_2 p, \mu^2\right)$ computed in perturbative regime and a normalization part, the parton distribution functions (PDF) f_i which are non-perturbative and represent the probability to extract a parton with a fraction $x_{1,2}$ of the total momentum from the hadron. The range of energy of the two cross sections, the hadronic one and the partonic one, is completely different.
PDFs functions are not perturbative, they are universal and evolve as a function of μ^2 according to specific system of coupled integro/differential equation of the SM which go under the name of DGLAP evolution equation by Dokshitzer, Gribov-Lipatov, Altarelli, Parisi [11]. They are extracted from data at low energy and represent the probability to find a parton in a hadron with a fraction x of the longtudinal momentum[8]. PDF extraction from data is out of the scope of this work.
As a part of this discussion, we include the definition of *factorization* and *renormalization* scales (μ_F, μ_R). These two factors are often set as equal to simplify the calculations, but in principle they acts as a cut off for two complemently different situations.

[8] Transverse momentum is neglected.

The renormalization scale μ_R controls which corrections enter explicitly in the vertex definition and it therefore explains why the coupling depends on the energy. The factorization scale μ_F is supposed to cure soft and collinear divergences, as these divergences can be absorbed into the parton distribution functions. Essentialy it is the scale where you declare these soft divergences to fail. Above that scale, the regime is perturbative and can rely on Feynman diagram calculations. The definition of the factorization scale is then essential as it represents the scale that divides the processes entering the partonic cross section in Eq. (1.37) ($\hat{s}^2 > \mu_F$) from those that are part of the hadronic wave function and therefore enter the PDF ($\hat{s}^2 < \mu_F$). The scale of the process represents our understanding to derive predictions for the structure functions, by separating the long-distance non-perturbative behaviour from short-distance perturbative behaviour computable in QCD. Both μ_F and μ_R are unphysical scales that enter in the observables because the perturbation expansion is truncated. Were we able to calculate *all the orders* in perturbation theory they would lose any meaning.
In Sec. (3.5) we will focus on studies of PDF for VBS process.

1.3 Precision tests of the SM

Higher Order Accuracy

At the time of writing this thesis, we lack of some ingredients in perturbative calculation, in particular, for any of the processes under study at LHC there are no (pure) NNLO EW corrections available, as well as mixed NNLO QCD-EW corrections. However some pure NNLO EW estimation corrections can be approximatively estimated as: nNLO EW $\sim$ (nLO EW)^2.
Like in the EW case, we can express the correction for QCD as a global external correction term:

$$\sigma^{NLO}_{QCD} = \sigma^{Born} + \delta\sigma^{NLO}_{QCD}.$$

where the correction factor can be applied to rescale both inclusive observables and differential distribution, according to the process. There are mainly two different approaches to approximate higher order correction: *additive* and *multiplicative*.

$$\textbf{Additive: } \sigma^{NLO}_{QCD+EW} = \sigma^{Born} + \delta\sigma^{NLO}_{QCD} + \delta\sigma^{NLO}_{EW}$$

$$\textbf{Multiplicative: } \sigma^{NLO}_{QCD\times EW} = \sigma^{NLO}_{QCD}\left(1 + \frac{\delta\sigma^{NLO}_{EW}}{\sigma^{Born}}\right) = \sigma^{NLO}_{EW}\left(1 + \frac{\delta\sigma^{NLO}_{QCD}}{\sigma^{Born}}\right)$$

These two approaches give different results according to the magnitude of each corrections. In particular, looking at Fig. (1.5), for observables with small EW corrections we can appreciate small differences between prescriptions, whereas

for observables with large EW corrections visible differences between prescriptions are present. A general rule suggest that when typical scales of QCD and EW corrections are well separated, then multiplicative recipe is the most reliable prediction. This happens in case of soft QCD interactions and Sudakov logarithms at high energy, as the two effects factorise.

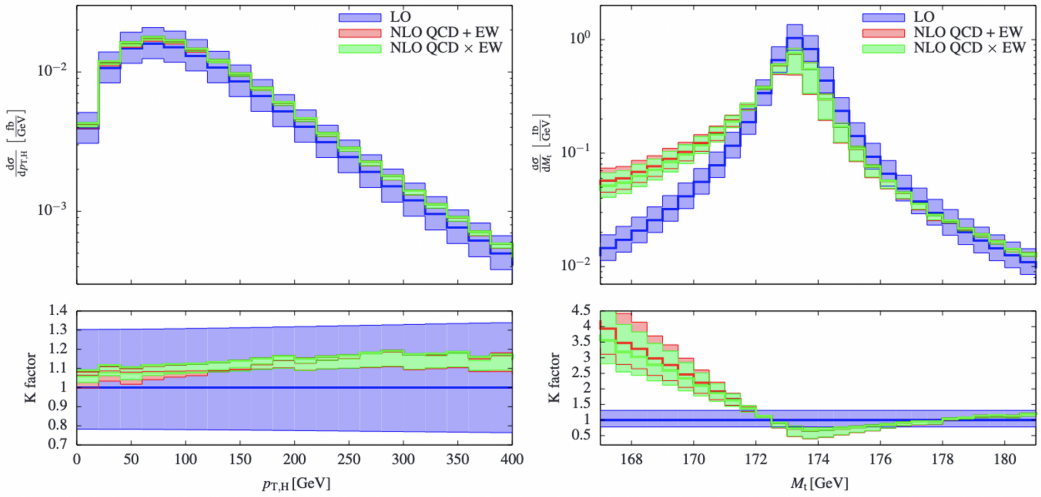


Figure 1.5: Differential distributions for $pp \to e^+\nu_e\mu^-\bar{\nu}_\mu b\bar{b}H$ [12]: transverse momentum of the Higgs boson (left) and the invariant mass of the top quark (right). The additive and multiplicative combination of the NLO QCD and NLO EW corrections are shown.

(N)NLO Implementations and Availabilities in Modern Programs

The evaluation of the NLO cross-section defined in Eq. (1.36) is not obvious; the two terms, the real and virtual components are separately divergent. From an analytical point of view this is not a problem, as with a regulator they can be computed separately, then added together so that the regulator dependence disappears. Differently, it is numerical impossible to integrate it (without regulator), mainly because the two terms have different integration variables. To overcome this problem, without entering into details, we just mention that there exist two approaches: *Slicing method* and *Subtraction method*.

If we consider the QCD case, NLO QCD corrections stabilises cross section as the LO suffers from large uncertainties. The inclusion of NLO QCD corrections leads to more stable predictions, but on the other hand the details of the parton-shower matching[9] programs cause differences which are considerably larger than those observed at fixed.
NLO QCD corrections remained a big challenge for more than twenty years but now is solved thanks to theoretical conceptual breakthrough ideas. In particular there are some different areas of contributions: from connection between

[9]This will be investigated in the next Chapter.

NLO amplitudes and LO ones, input from supersymmetry/string theory, sophisticated algebraic methods and connections with formal theory and pure mathematics.

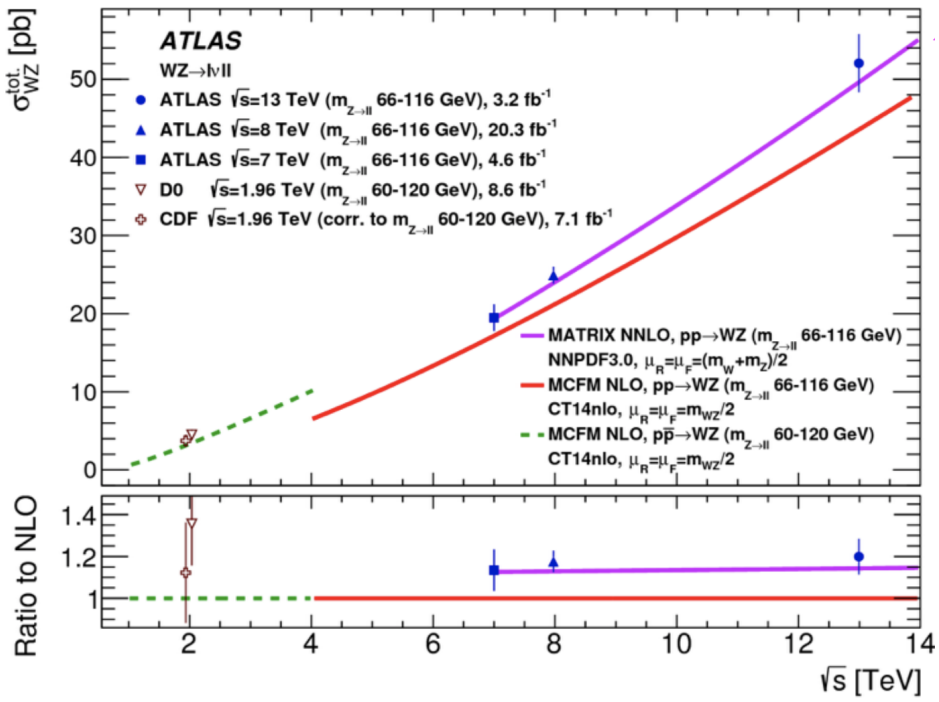


Figure 1.6: Cross-section data for the process $WZ \to l\nu ll$ [13]. Red line represents NLO prediction while purple line represents NNLO predictions. The blue points represent ATLAS Run1 and Run2 inclusive results.

Not only NLO calculation has been performed for most of the interesting processes, but also an explosion of NNLO results appeared in the last two years, at least form $2 \to 2$ SM process. Observing Fig. (1.6) we see that LHC data are clearly better described by NNLO calculations. With more and more data coming from LHC run, NLO is likely to be insufficient. Obviously, NNLO calculations are even more complicated because of two-loop master integrals (when many scales are involved) and the cancellation (overlapping) of divergences before integration.
We would also like to mention that few LHC processes are now known at N3LO, namely: fully inclusive Higgs production via gluon fusion (large m_T limit) [14] and [15], Higgs rapidity evaluation [16] and Higgs production via vector boson fusion in the DIS approximation [17] (factorised approximation).

1.4 Future Challenges Beyond the SM

Even though the SM describes almost all existing data very well, and even if it successfully passes tests at a precision level of $10^{-3} \div 10^{-4}$, we know that this

Model is incomplete, as it supplies no answer to some fundamental questions.

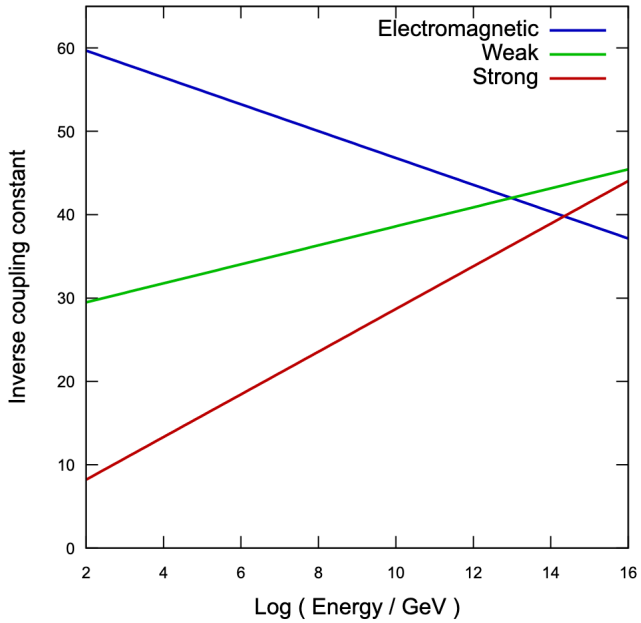


Figure 1.7: Evolution of the coupling constants of the three interactions described in the SM as a function the energy scale. Image taken from `https://commons.wikimedia.org/wiki/File:Running_coupling_constants.svg`.

During the last century, scientists have tried to find a unifying principle to explain Mother Nature behaviour, they have been wondering whether it could exist a general force that comprises all those already discovered. It was found that, for the ElectroWeak interaction to allow particles and mediators to acquire mass (as was described previously), the symmetry that characterises the interaction needs to be broken by a mechanism. The scientific community, in an attempt to overcome this scenario and to provide an alternative explanation, has been involved in the research of a general symmetry which includes all others. We know that all the fundamental interactions are proportional to their coupling constants that, despite its name, as already discussed, are running parameters which varies with energy. Considering Fig. (1.7) and taking into account the three different behaviours of the strong, electromagnetic and weak couplings, as a function of the energy scale, one may find an energy value where they meet at a point. Unfortunately, there exists a desert of 13 orders of magnitude from todays reachable energies and the putative unification one. In addition, it seems that, according to the current models, they do not meet at the same point. It is needful to stress here, that the coupling constant evolutions are obtained by considering only the currently observed particles and assuming absence of "new physics" through all the energy-scale considered.
There are some additional problems related to the SM. As already discussed in this chapter, the *unitary problem* related to the divergences in the probability amplitudes with increasing energies in the scattering matrices and the

hierarchy problem i.e., the large discrepancy between aspects of the weak force and gravity. A hierarchy problem occurs when the fundamental parameters, such as coupling constants or masses are vastly different than the parameters measured by experiments. This can happen because of the renormalization, where parameters are closely related to the fundamental ones, and this is a consequence of a gauge theory using massless gauge bosons.
Hierarchy problems are also related to fine-tuning and to problems of naturalness. We are talking about higher order corrections in the amplitude scattering matrices, which are required to achieve a high degree of precision in the theoretical predictions to be compared with experimental results. Performing those calculation means that we must have perfect knowledge of several parameters such as the mass of Higgs boson, of the *top* quark and especially their coupling constants.

Naturalness Problem

The Naturalness problem of a theory arises if its underlying parameters are all of the same size in appropriate units. A more precise definition involves the notion of an effective field theory – the idea that a given quantum field theory might only describe nature at energies below a certain scale, or cutoff. Certainly, it can be phrased just as a question about the origin of the Higgs mechanism (and the Higgs potential). The Standard Model (SM) is an effective field theory because it cannot be valid up to arbitrarily high energies even in the absence of gravity. The Higgs mass operator is the only operator with dimension less than 4 in the SM (relevant operator). In fact it is the only one that requires the introduction of a dimensionful parameter:

$$\mathcal{L}_H = (D_\mu H)^\dagger D^\mu H - V(H); \qquad V(H) = -\mu^2 H^\dagger H + \lambda \left(H^\dagger H\right)^2 .$$

If we switch off the Higgs mass term, then there are no scales in the SM. For QCD scale, which is dynamically generated, the situation is different, i.e. we did not require it, QCD dynamics just generates it. Then, if any dimensionful scale is present in the ultra violet (UV), Naive Dimensional Analysis (NDA) would set the Higgs mass parameter equal to this scale:

$$\boxed{\mathcal{L}_{NDA} = c\Lambda^2 H^\dagger H} \tag{1.38}$$

We do have some indication of the existence of fundamental high scales:

- ⋆ Grand Unification Scale: is the scale where gauge couplings (almost) converge toward a single value $M_{GUT} = 10^{16}$ GeV;
- ⋆ Planck Scale: the scale of quantum gravity $M_{Planck} = 10^{19}$ GeV.

If any of these new physics scales exists (and is associated to the propagation of new states), then NDA tells us that the Higgs mass should be of the order of

these scales. Of course it is not, and indeed all our statements were classical, i.e. did not include any radiative (loop) corrections and radiative corrections should then be responsible for the Higgs mass we observe.

$$\boxed{m_h^2 = c\Lambda^2 + \delta m_h^2} \tag{1.39}$$

However, this would require an extremely finely tuned c constant in order to cancel very big numbers (notice the square) and give rise to the observed Higgs mass. This does not seem natural and it is therefore referred to as a problem of *naturalness*.
There are several directions to address this problem:

- New physics at TeV scale: the most studied solution. It aims at lowering Λ to the TeV scale. Then the required cancellation becomes much smaller;
- Anthropic's: assumes the existence of many different universes (multiverse) with different values of m_H and aims at proving that our universe is possible only for the observed value;
- No UV scale: one can try to construct theories that do not have any high energy scale associated to new scales, so that $\Lambda = m_h$ is just the NDA prediction.

Trying to address this problem, new models and theories came out in the last 40 years, among these we cite Supersymmetry (SUSY), Extra-dimensions and Strong dynamics (the composite Higgs model). At the time of this writing, there is no evidence yet for any sign of new physics and scientists start wondering about our understanding of Naturalness. For sure this remains one of the biggest puzzles in fundamental physics.

1.5 Summary

In this Chapter, we reviewed the Standar Model of particle physics in its main aspects: particle content, gauge sector, perturbative and non perturbative ingredients, SSB and capability of prediciting observables. We realized that symmetries are an incredibly powerful tool in physics and SSB is not an actual breaking, rather a different realization of the symmetry. In the SM, the SSB of gauge invariance is responsible for all masses of elementary particles (apart from neutrinos) and it is realized linearly, with the Higgs boson identified as the radial mode.

In LHC Run 1 phase, with center-of-mass energies of 7 and 8 TeV, all particles predicted by the Standard Model have been observed and their main parameters determined. In 2015, Run 2 has started with an increased center-of-mass energy of 13 TeV. This data taking period and the subsequent runs,

approved for two further decades (including a high-luminosity upgrade), will allow to test the predictions of the SM at high precision and to explore the possibility to find physics beyond the Standard Model. With this increase in the energy and luminosity we have already entered a *precision-era*, where we can access a class of processes which strongly benefit from these unique and unprecedented conditions.
One among them is the **Vector Boson Scattering** which will be the main actor of next Chapter and of this thesis work.

Chapter 2

Vector Boson Scattering

Vector Boson Scattering (VBS) and Vector Boson Fusion (VBF) are an important class of processes under investigation at the Large Hadron Collider at CERN. The ElectroWeak sector of the Standard Model predicts self-interactions between W and Z gauge bosons through triple and quartic gauge couplings. The quartic coupling in particular can be probed at colliders only via rare processes such as tri-boson production or Vector Boson Scattering.
In Fig. (2.1), a general diagram for the VBS process is given, where many different interactions can occur in the grey blob. The common signature for VBS processes is therefore two jets which originate from the outgoing quarks, plus additional particles from the decay of the two bosons.
VBF has a similar structure, but here an Higgs boson emerges from the fusion of 2 vector bosons. VBF is the second production process of the Higgs boson at LHC in terms of cross section value.
In the following paragraphs, we will focus on VBS process, leaving VBF to dedicated Higgs boson papers, as [18, 19, 20, 21].

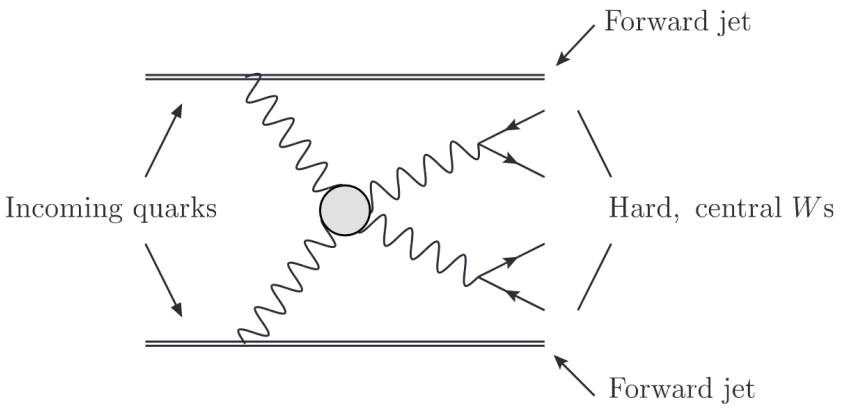
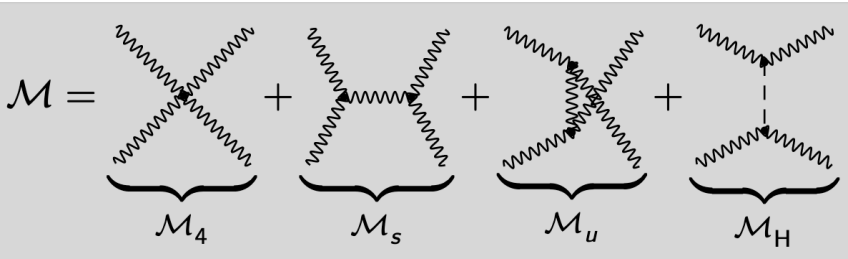


Figure 2.1: Schematic representation of a Vector Boson process.

In the following we will summarize the main characteristics of the VBS process, highlighting the role of the Higgs boson in its regularization and discussing how this process can act as a portal for possible new physics.

2.1 Process Definition

As already mentioned, Vector Boson Scattering at a hadron collider usually refers to the interaction of massive vector bosons ($W^{\pm}, Z$) radiated by partons (quarks) originating from the incoming protons. Counting out the 2 tagging jets, the hard VBS is a $2 \to 4$ process with two ElectroWeak bosons in the final state, therefore the typical signature of VBS events is characterised by two energetic jets coming from quarks, and four fermions, originating from the decay of the two vector bosons. It is a pure ElectroWeak process and among all possible diagrams the scattering process can be mediated by a Higgs boson. Let's consider the following contributions to the matrix element $\mathcal{M}$ of a generic VBS process:



$$\mathcal{M}_4 \propto -s^2 - u^2 - 4su + 2\left(M_W^2 + M_Z^2\right)\frac{s^2 + 6su + u^2}{s+u} + \dots$$
$$\mathcal{M}_s \propto s^2 + 2su - 2M_W^2\frac{3su + u^2}{s+u} - 2M_Z^2\frac{2u^2 + 3su - s^2}{s+u}$$
$$\mathcal{M}_v \propto u^2 + 2su - 2M_W^2\frac{3su + s^2}{s+u} - 2M_Z^2\frac{2s^2 + 3su - u^2}{s+u} - \frac{m_Z^4 U}{M_W^2} + \dots$$
$$\mathcal{M}_H \propto -\frac{M_Z^4}{M_W^2} + \dots + \dots$$

$$\boxed{\mathcal{M} \Longrightarrow \propto -\frac{m_Z^4}{M_W^2}(s+u) \qquad \text{if} \qquad m_H \to +\infty.} \tag{2.1}$$

Where we used the Mandelstam variables:

$$s = (p_1 + p_2)^2 \quad t = (p_1 - p_3)^2 \quad u = (p_1 - p_4)^2.$$

Assuming $|t| >> M_H^2, M_Z^2, M_W^2$, we are able to estimate the maximal effect of different Higgs couplings. The Higgs exchange diagram form an essential

part of the amplitude, because they cancel the remaining contribution. In addition, if the Higgs boson does not have the exact couplings to vector bosons as predicted by the SM, then the necessary divergent contributions cancellations will not occur and unitarity cannot be guaranteed in the high energy scenario. The Higgs boson plays a crucial role in restoring the unitarity of the scattering amplitude by interfering with all the other particle mediated diagrams. As will be emphasized later on, such cancellations are not needed when the scattering involves transverse polarized gauge bosons, whereas for **longitudinally polarized** bosons ($W_LW_L \to W_LW_L$) it is of crucial interest, because the corresponding matrix elements feature unitarity cancellations that strongly depend on the actual structure of the Higgs sector of the Standard Model. Therefore, with precision studies of VBS process, we access to Higgs couplings at a very different energy scale with respect to the Higgs boson mass value.
According to the intermediate propagators (boson) that we consider, we can define several different VBS processes. The main work carried out in this thesis is related to VBS mediated by same sign W bosons, even though an interesting part about polarization extraction has been tested on ZZ mediated VBS process (see Chapter 4). The VBS process involving two same-sign W bosons has the largest signal-to-background ratio at the LHC.
VBS processes can also be classified according to the vector boson decays: in the *semi-leptonic* case, one of the W bosons decays into light leptons, while in *fully-leptonic* case, both W bosons decay leptonically. The latter signature is experimentally preferred due to a lower background contamination and more efficient data acquisition triggers. The scattered quarks then hadronize to jets, resulting in two hadronic jets, not colour connected as will be discussed later. VBS processes are currently under investigation by several experimental groups at the two main general purpose experiments at CERN (ATLAS [22] and CMS [23]).

The typical cross-section of VBS processes is of $\mathcal{O}$ (1 fb), and this explains why large $\sqrt{s}$ and luminosity $\mathcal{L}$ are needed to access experimental evidences. This can be seen in Fig. (2.2), where VBS signature is depicted at the very right of the summary plots, together with other rare processes. This plot summarises the main SM cross sections, as measured by ATLAS on Run 1 + Run 2 data, together with their experimental and theoretical uncertainties.

Leading Order Contribution

VBS receives contributions at LO from three different sets of diagrams, namely:

$$\mathcal{O}(\alpha^6), \qquad \mathcal{O}(\alpha_s^2\alpha^4) \qquad \text{and} \qquad \mathcal{O}(\alpha_s\alpha^5).$$

The first one represents the EW contribution, see Fig. (2.3) on the left. This is what is normally referred as pure VBS contribution. In the same figure also s-channel contributions with non-resonant vector bosons (centre diagram)

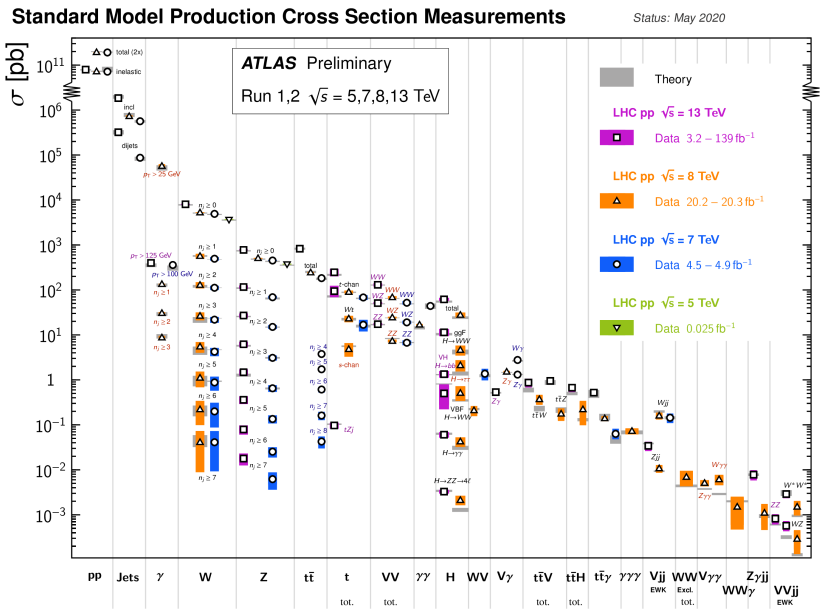


Figure 2.2: Summary of several Standard Model (total and fiducial) production cross section measurements, performed by ATLAS experiment. Vector Boson Scattering appears on the very right of the table, where rare processes are confined [24].

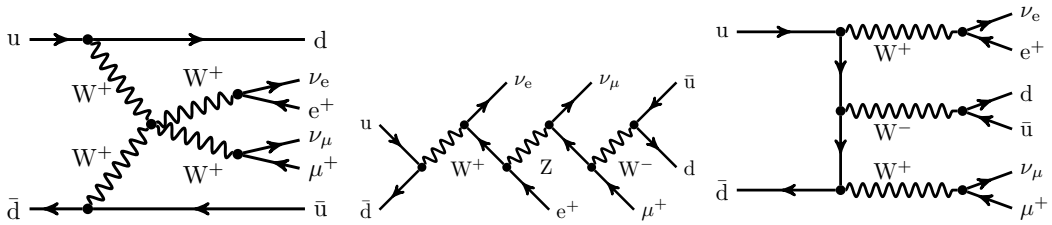


Figure 2.3: Sample tree-level diagrams that contribute to the process $pp \to \mu^+\nu_\mu e^+\nu_e jj$ at order $\mathcal{O}(\alpha^6)$. In addition to typical pure VBS contributions (left), this order also possesses s-channel contributions such as decay chain (middle) and tri-boson contributions (right).

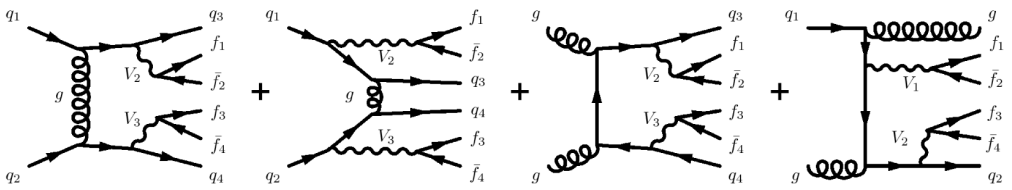


Figure 2.4: QCD-induced contribution.

and triple-boson production (right diagram) are represented. Here, s,t,u contributions are defined according to the quark lines, and their names refer to the corresponding Mandelstam variable. The s-channel groups all Feynman diagrams where the two initial-state partons are connected by a continuous fermion line, while in the $t-$ and u-channel the fermion lines connect initial state quarks to final state quarks. The u-channel refers to contributions with

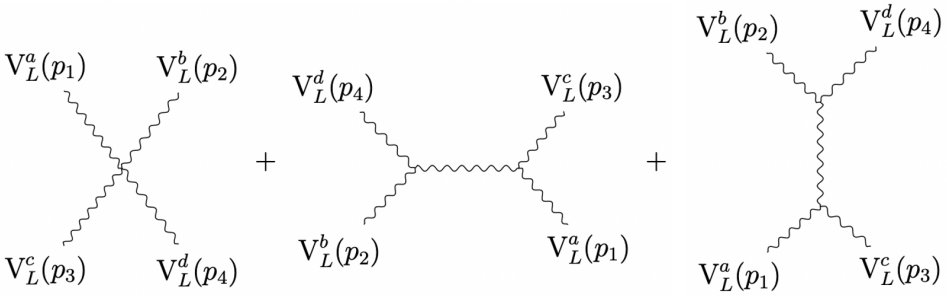
crossed fermion lines with respect to t-channel, which appears for identical (anti) quarks in the final state.

The second contribution to LO VBS process, $\mathcal{O}(\alpha_s^2\alpha^4)$, comes from gluon connecting the two quark lines, with the W bosons radiated off the quark lines. An important note here is that this contribution has a different colour structure that features a different kinematic behaviour than pure VBS, nonetheless it shares the same final state, and therefore constitutes an **irreducible background** to the EW process. Some QCD-induced contributions of the order $\mathcal{O}(\alpha_s^2\alpha^4)$ are also represented in Fig. (2.4). The production of a single top quark or a tt pair represents a clear example. The *top* quarks decay into a b quark and W boson and produce the same signature of VBS. The cross-section of this background is high at LHC but it can be reduced thanks to b-tagging strategies of the detector. In general, the largest background contribution comes for the production of a W vector boson plus jets at order $\mathcal{O}(\alpha_s^2\alpha^4)$. This is a **reducible background** since the final state is different from VBS ones (one vector boson is missing) but can be misidentified since vector bosons are not directly detected. Different multiplicities of jets enter in the definition of the matrix-elements of this background since the QCD radiation can add more jets in the final state.

The last component, $\mathcal{O}(\alpha_s\alpha^5)$, refers to the mixed contribution of the previous processes, namely the *interference*, which represents a few percent with respect to the total of LO calculation [25].
In the attempt of isolating the contribution we are interested in, namely the pure VBS $\mathcal{O}(\alpha^6)$ diagram, we can define kinematic cuts (VBS cuts) that are designed both to enhance the EW contribution over the QCD one and to suppress the interference, as will be discussed in Chapter 3.

Scattering Amplitude

As an example, we calculate the scattering amplitude in the specific case of four longitudinal vector bosons[1], specifying the different s-, t-, and u- contributions.



$$\mathcal{A}\left(V_L^a(p_1)V_L^b(p_2) \to V_L^c(p_3)V_L^d(p_4)\right) = \frac{g^2 s}{4m^2}\delta^{ab}\delta^{cd} + \frac{g^2 t}{4m^2}\delta^{ac}\delta^{bd} + \frac{g^2 u}{4m^2}\delta^{ad}\delta^{bc}$$

[1]The reason beneath this choice will be clear in the next Chapter.

From Eq. (2.1) we know that tree-level amplitude grows with energy so that we can interpret this with a coupling strength varying with the energy. At the same time, this coupling constant cannot become arbitrarily large, as it would cause the amplitude to diverge.

$$\mathcal{A} \approx g(s)^s$$

In particular, considering the expansion in the coupling constant:

$$\mathcal{A} = \mathcal{A}^{tree}(s) \left(1 + c_1 \frac{g(s)^2}{(4\pi)^2} + c_2 \frac{g(s)^4}{(4\pi)^4} + ...\right);$$

when the coupling $g(s)$ becomes of order 4π, all terms in the perturbative expansion become of the same order and perturbation theory breaks down. This problem is usually referred to as *unitarity violation*, and it is related to the fact that unitarity, in this probabilistic interpretation of the scattering matrix, can be violated at high energy. In this regime, perturbation theory does not hold anymore. Indeed the scattering matrix, if it exists, it is unitary by construction even though it cannot always be computed perturbatively. According to what has been pointed out we can now introduce the *no-lose theorem*.
The constraint on perturbativity can tell us something about the ElectroWeak theory, if we interpret the massive vector as the W boson. Knowing its mass and coupling strength, we can get information on what to expect for new physics. The constraints are:

$$g(s) = \frac{g\sqrt{s}}{2m_W} \lesssim \quad \Longrightarrow \quad \sqrt{s} \lesssim \frac{8\pi m_W}{g} \frac{4\pi}{\sqrt{\sqrt{2}G_F}} \approx 3 \text{ TeV}.$$

This is a clear indication that the existence of massive vector bosons in the Standard Model requires new phenomena at a scale almost fully accessible at collider experiments.
Putting together all of these observations, we come to the conclusion that in the attempt to introduce masses for the weak gauge bosons we discovered that no matter what the mechanism generating these masses is, new phenomena are expected at the TeV scale. If these new phenomena enter at low scales, where their size is small (perturbative) we define the new dynamics *weak*, while if they enter close to our bound we refer to them as a new *strong* dynamics (non-perturbative).

2.1.1 Process Selection

Pure VBS process is characterized by the presence of two initial and final quarks without intermediate strong (colour) interactions. The pure EW contribution defined above is characterized by two jets with large rapidities in the forward region of the detector, by a reduced jet activity in the central region as

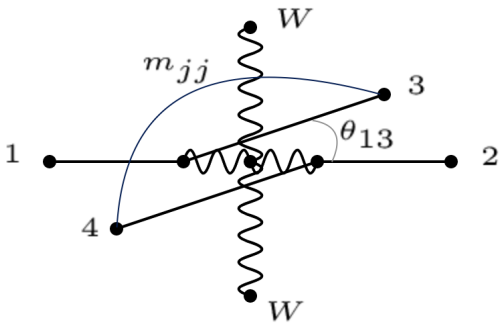


Figure 2.5: LO kinematics representation for VBS process.

well as a large di-jet invariant mass. The two W bosons are mostly produced centrally and the event selection adopted usually involves rapidity separation ($\Delta\eta$) and invariant-mass (m_{jj}) cuts for the jets. We define *tagging jets* the two quarks in the final state, which appear as hadronic jets in the detector and are expected to have large energies that allow cut selection. The kinematics representation is shown in Fig. (2.5) where quarks are located into detector regions with fairly small scattering angle or large pseudorapidity. A useful correlation between the jet variables is:

$$m_{jj}^2 \simeq 2p_{T,j_1}p_{T,j_2}\left(\cosh(\Delta(y_{jj}) - \cos(\Delta(\phi_{jj}))\right);$$

where $\Delta(\phi_{jj})$ denotes the azimuthal angle difference between the two jets and $\Delta(y_{jj})$ indicates their rapidity separation. In LO approximation, the above relation becomes exact as the two jets are massless.
QCD contributions, like the one shown in Fig. (2.4), instead, favours jets in the central region due to colour connection between the two tagging quarks. Therefore, despite sharing the same signature with the pure EW VBS, the kinematics of this process is different with respect to the pure EW VBS case. Here, there is no preferred region for jet emission and the W is radiated from quark lines.

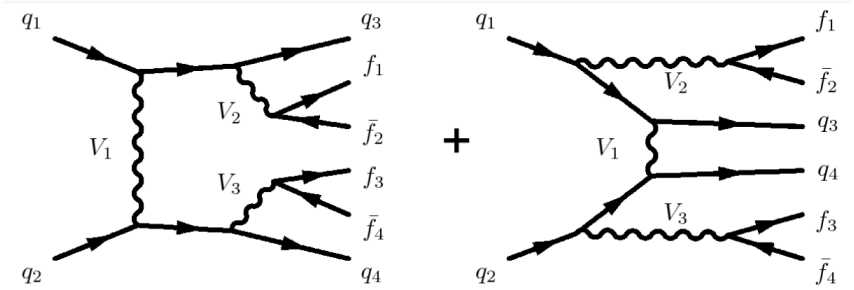


Figure 2.6: VBS backgound diagrams of order $\mathcal{O}(\alpha^6)$.

Additional diagrams contribute as a background process, in particular for W, Z events, those are of order $\mathcal{O}(\alpha^6)$, like the one shown in Fig. (2.6). This

class of processes, that mimic VBS-like events in the final states, are mostly suppressed by means of kinematic cuts according to the low rapidity of the two jets.
Despite the kinematic selections that one can apply to events, signal extraction represents a challenge for experimental collaborations due to low statistics of the phase-space regions where EW contribution is dominating [26]. This is explicitly shown in Tab. (2.1) where cross section values for different VBS channels are compared. QCD and EW contributions are process dependent, so that pure EW selection in W^+Z process could become challenging for experimental collaborations.

2.2 NLO corrections for VBS

In this section, we focus mainly on higher order corrections for VBS process. In order to obtain precise predictions in perturbation theory, as explained in the previous Chapter, the inclusion of higher-order corrections is necessary. The complete calculations for different channel are presented in: WW [25], WZ [27] and ZZ [28]. As already state, there are several ways to define NLO

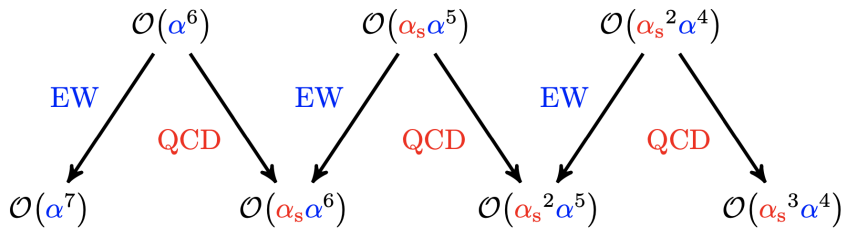


Figure 2.7: Schematic representation of NLO correction to VBS process, both from QCD and EW contribution.

corrections: they can either be virtual, which contribute via an interference with the Born matrix element, or real-emission diagrams, where an additional gluon or photon is radiated off the quark line. An illustrative representation of VBS NLO corrections is shown in Fig. (2.7). We have four possible NLO contributions: $\mathcal{O}(\alpha^7)$, $\mathcal{O}(\alpha_s\alpha^6)$, $\mathcal{O}(\alpha_s^2\alpha^5)$ and $\mathcal{O}(\alpha_s^3\alpha^4)$. We refer to the first as EW corrections to the EW-induced LO processes and to the last as QCD corrections to the QCD-induced LO processes (whose diagrammatic representation is depicted in Fig. (2.8)). Regarding the two remaining orders, a simple separation of the EW-induced process and the QCD-induced process is not

Process	Order	$\mathcal{O}(\alpha^6)$	$\mathcal{O}(\alpha_s^2\alpha^4)$
W^+W^+ [25]	σ_{LO} [fb]	1.4178(2)	0.17229(5)
W^+Z [26]	σ_{LO} [fb]	0.25416(6)	0.9912(2)

Table 2.1: Theoretical LO cross sections at order $\mathcal{O}(\alpha^6)$ and $\mathcal{O}(\alpha_s^2\alpha^4)$ for two different VBS channels: W^+W^+, W^+Z.

Order	$\mathcal{O}(\alpha^7)$	$\mathcal{O}(\alpha_s\alpha^6)$	$\mathcal{O}(\alpha_s^2\alpha^5)$	$\mathcal{O}(\alpha_s^3\alpha^4)$	Sum
$\delta\sigma_{NLO}$ [fb]	$-0.2169(3)$	$0.0568(3)$	$-0.00032(13)$	$-0.0063(4)$	$-0.2804(7)$
$\delta\sigma_{NLO}/\sigma_{LO}$ [%]	-13.2	-3.5	0.0	-0.4	-17.1

Table 2.2: NLO fiducial cross sections for $pp \to \mu^+\nu_\mu e^+\nu_e jj$ [25].

possible any more. From a numerical perspective, α_s is ~ 0.1 while α is $\sim 0.76 \times 10^{-2}$, so we can deduce that NLO EW $\sim$ NNLO QCD corrections[2], in size.
At LO we need to consider all possible diagrams to preserve the gauge invariance, similarly, related NLO diagrams are gauge invariant group-wise, yet not IR finite. Because of that, when considering a full computation calculation, we should consider an IR finite set collecting all the contribution from virtual and real ones.

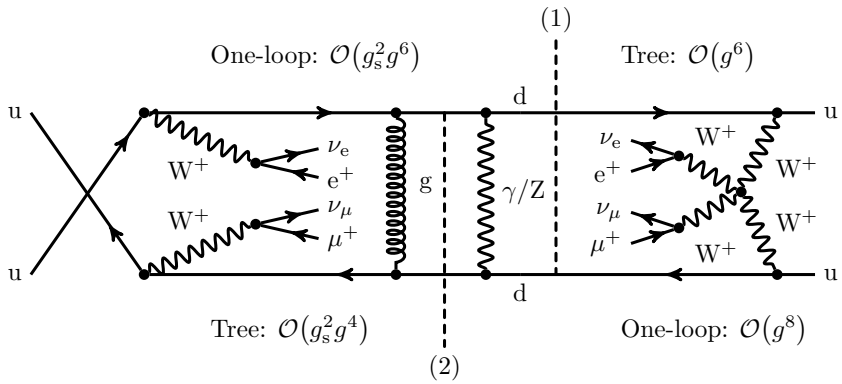
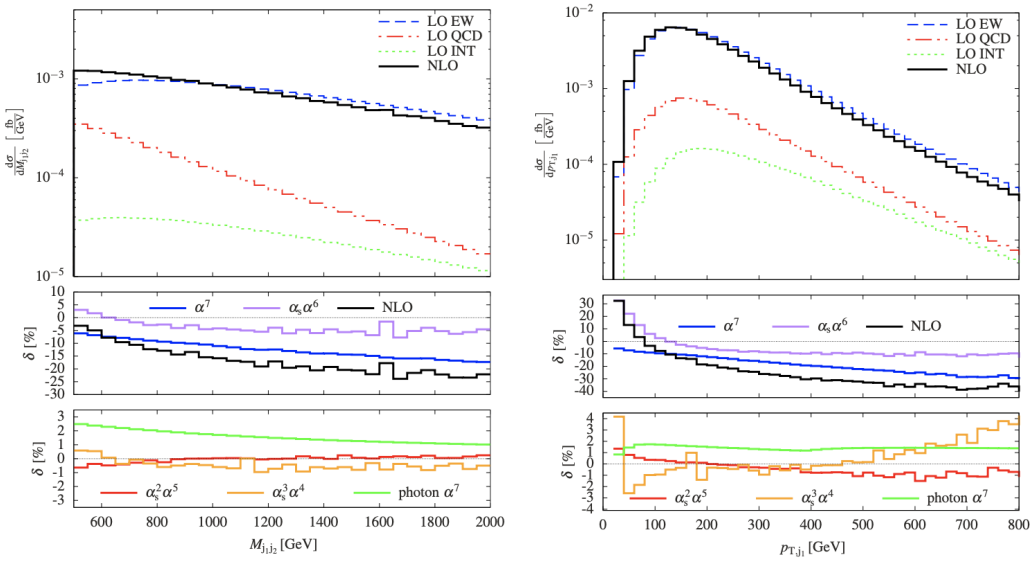


Figure 2.8: Diagrammatic version of the contribution $\mathcal{O}(\alpha_s\alpha^6)$, QCD corrections to QCD induced process. It can be viewed as an amplitude of order $\mathcal{O}(g_s^2 g^6)$ interfered with the LO EW amplitude [cut (1)]. On the other hand, it can be seen as an EW correction to the EW amplitude interfered with the LO QCD amplitude [cut (2)]. Owing to the colour structure, the illustrated contractions necessarily connect t- and u-channel contributions.

The goal of the whole theory community here is to find a match between theoretical predictions and what then is measured experimentally, to avoid biased interpretation due to model dependency. In particular, separated contribution from EW, QCD and interference, give different results with respect to combined prediction computed against full computation. Looking at Tab. (2.2), we observe large $\mathcal{O}(\alpha^7)$ and negative $\mathcal{O}(\alpha_s\alpha^6)$ EW corrections that distinguish the process considered from others[3].

In Fig. (2.9a), transverse momentum of the anti-muon distribution coming from a $pp \to \mu^+\nu_\mu e^+\nu_e jj$ process is displayed. EW-induced contribution is dominant over the whole phase space (top panel). Considering the relative NLO corrections in the lower panel, the largest contribution is the one

[2]Data are extrapolated at 100 GeV.
[3]In the same paper [25], the author also calculated the photon-induced NLO contributions.



(a) (b)

Figure 2.9: Differential distributions at a centre-of-mass energy $\sqrt{s} = 13$ TeV at the LHC for $pp \to \mu^+\nu_\mu e^+\nu_e jj$. Transverse-momentum for the anti-muon (left) and invariant mass for the two leading jets (right). The upper panels show the three LO contributions as well as the sum of all NLO predictions. The two lower panel show the relative NLO corrections with respect to the full LO [25]

of $\mathcal{O}(\alpha^7)$, we see the typical large corrections for high transverse momenta of these kind of process, namely the Sudakov logarithms. The second largest contribution is the QCD correction to EW process ($\mathcal{O}(\alpha_s^2\alpha^5)$).
On another hand, Fig. (2.9b) shows the distribution in the transverse momentum of the leading jet. The clear hierarchy between the LO contributions observed before does not hold anymore, here the LO interference becomes comparable to the QCD-induced process around 800 GeV. We have similar behaviour instead for the NLO corrections with just a different contribution in sign of the order $\alpha_s\alpha^6$ with respect to kinematic energy. Talking about combined prediction, again in [25] (see, for instance, Fig. 7, in this reference), we have a clear evidence of large negative corrections for the full process, where total corrections dominated by EW contribution to EW process show that bands do not overlap, confirming the prediction that the NLO EW corrections to VBS represent a large fraction of the NLO corrections.
The EW corrections are typically added to observables via an overall delta factor, which can be expressed in the following way:

$$\sigma_{NLO}^{EW} = \sigma + \alpha \times \delta_{NLO}^{EW}$$

$$\sigma_{NNLO}^{EW} = \sigma + \alpha \times \delta_{NLO}^{EW} + \alpha^2 \delta_{NNLO}^{EW}$$

In a different paper [29], EW corrections to $W^\pm W^\pm$ are evaluated: they sum up to -16%, see Fig. (2.10). Same conclusion are provided in [27] for the WZ

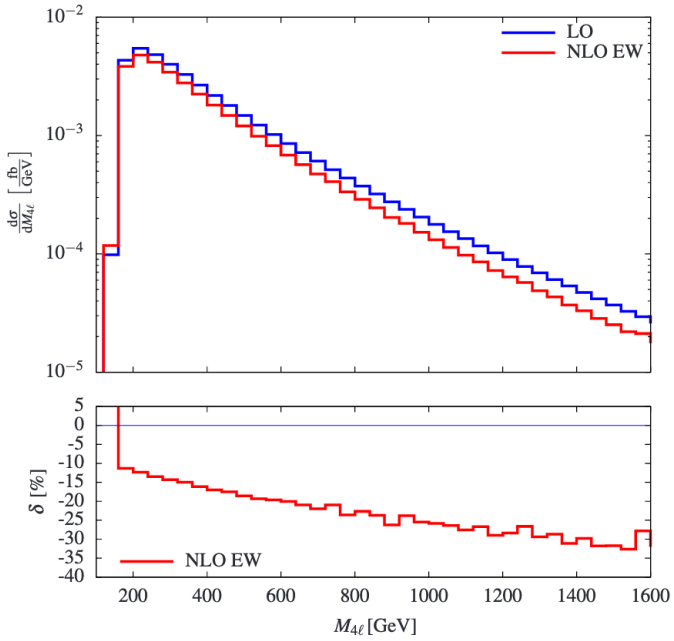


Figure 2.10: Invariant-mass distribution of the four leptons in $pp \to \mu^+\nu_\mu e^+\nu_e jj$ including NLO EW corrections (upper panel) and relative NLO EW corrections (lower panel) [29].

channel. The authors confirmed how the ElectroWeak corrections feature the typical Sudakov behaviour towards high energy ($\log^2\left(\frac{Q^2}{M_W^2}\right)$).
Both papers confirm the hyphotesis that large EW corrections are indeed an intrinsic feature of VBS at the LHC and are not related to VBS typical kinematic cuts: $m_{jj} < 500$ GeV and $|\Delta y_{jj}| > 2.5$. This will be discussed extensively in the next Chapter.

Double Pole Approximation

We introduce an important technique [30], an approximation with respect to the full computation calculation that allows for the description of a process like VBS using leading contribution of an expansion about the resonance poles. This method, which goes under the name of *Double Pole Approximation* (DPA), is used when one is only interested in the resonance regions, which is typically the case in many-particle processes with low cross sections. It eases the calculation, avoiding a proliferation of terms induced by the numerous Feynman diagrams contributing only in off-shell regions.
The pole scheme suggests to isolate the gauge-invariant residues of the resonance poles and to introduce propagators with complex masses M only there, while keeping the remaining parts untouched. Restricting this general procedure to resonant contributions, defines the pole approximation (PA), which is adequate if only the off-shell behaviour of cross sections near resonances is relevant, but contributions deep in the off-shell region are negligible. Corrections are classified into separately gauge-invariant factorizable and non-factorizable,

where the former can be attributed to the production and decay of unstable particles on their mass shell.
A representation of both a general and VBS-like process is shown in Fig. (2.11). On the left we can see a generic process with multiple resonances, where final states result from the decay of a resonant particle or they are directly produced without intermediate resonant state that we can describe with an expansion about the resonance poles. On the right we have a VBS DPA representation. The process involves the emission of two W bosons which then decay into fermions: the approximation allows the numerator of the amplitude to be written with the momenta projected in such a way that the boson is on–shell, while the Ws propagators are kept off–shell. In both diagrams, the bubbles represent the procedure to isolate the gauge-invariant resonant contributions, residues of the resonance poles, as a solution to the ElectroWeak radiative corrections of high-multiplicity final states with several intermediate resonances. The remaining non-factorizable corrections are induced by the exchange of soft photons between different production and decay subprocesses. Double Pole Approximation gives an agreement within 1% with full calculation where dominant terms are factorizable because of large corrections driven by the scattering process.

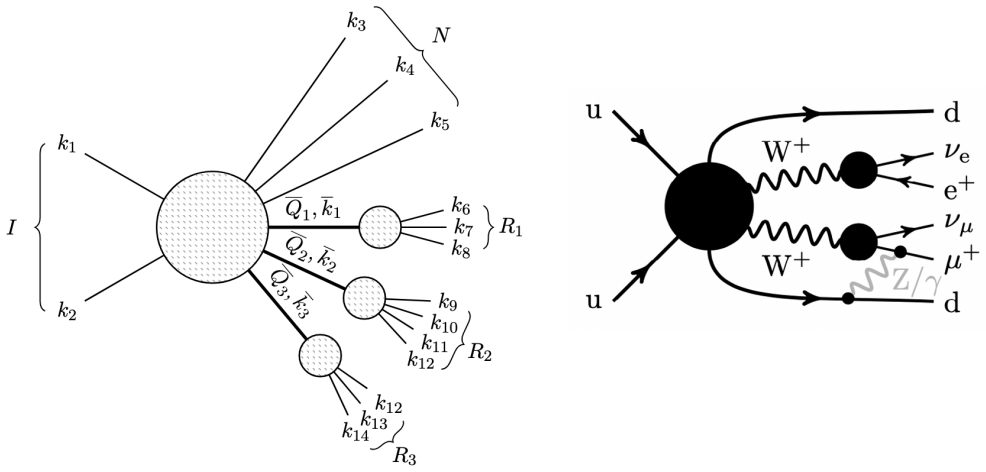


Figure 2.11: Diagram for a typical process with multiple resonances,outgoing particles either result from the decay of a resonant particle, $i \in R$ with electric charges $\bar{Q}_j$and momenta $\bar{k_j}$ or are directly produced without intermediate resonant state,[30](left). Double-pole approximation for VBS non-factorizable correction process (right).

Effective Vector Boson approximation

An additional technique widely used in VBS studies is the Effective Vector Boson approximation [31, 32]. This approximation focuses on the inner bubble sketched in Fig. (2.12), whose contribution can be analytically computed by

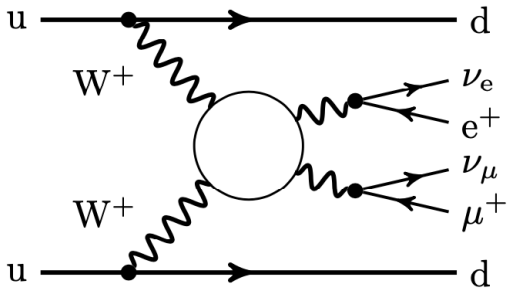


Figure 2.12: Diagram representation of Vector Boson Approximation (VBA) for the VBS simplified process: $W^+W^+ \to W^+W^+$.

introducing the *Leading logarithm approximation* [33]:

$$\sigma_{LL} = \sigma_{LO}\left[1 - \frac{\alpha}{4\pi}4C_W^{EW}\log^2\left(\frac{Q^2}{M_W^2}\right) + \frac{\alpha}{4\pi}2b_W^{ew}\log\left(\frac{Q^2}{M_W^2}\right)\right]; \qquad (2.2)$$

where Q is the typical scale of the process, so that $\sigma_{LL}^{EW} = -16\%$[4].

$$Q > 2M_W; \qquad Q = s = (p_1+p_2)^2 = \left(\sum_l p_l\right)^2 = \langle m_{4l}\rangle \sim 390 \text{ GeV}$$

$$\implies \sigma_{LL} \simeq \sigma_{NLO}^{EW}$$

This result shows that σ_{NLO}^{EW} for VBS process is almost 4 times the σ_{NLO}^{EW} of di-boson process ($q\bar{q} \to W^+W^+$). This observation is related to the fact that C^{EW} is larger for bosons (W, Z) than for fermions and quarks[5], and that $\langle m_{4l}\rangle$ is larger for VBS due to massive t-channel. In [34], the reader can find a full detailed analysis of this effect where the t-channel contributions originates from the s-channel contributions via crossing the outgoing W^- boson with the incoming W^+ boson.

2.3 Summary

In this Section, we focused on the VBS process and on how it is known in perturbation theory. All diagrams which contribute to the LO description are well know and all together are gauge invariant.
At NLO, this process receives two main contributions on the EW side, namely QED radiations near kinematic threshold and weak corrections in high-energy limit (Sudakov logarithms). These predictions are relevant for SM measurements and BSM searches. The second largest contribution is represented by the

[4]Eq. (2.2) takes into account several contributions: the double EW logs, the collinear single EW logs and single logs including the parameter for the renormalisation. Angular-dependant logarithms are omitted.

[5]With typical $\langle m_{4l}\rangle \sim 250$ GeV.

QCD corrections to EW process, while mixed contributions depend strongly on the size of QCD-induced process at LO. Of course, NLO calculation does not necessarily give NLO-accurate observables, however an accurate combination of NLO EW with QCD NLO matched with parton shower is important, as it will be emphasized in the next Chapter.

The state-of-art for VBS calculation for the fully leptonic case can be summarized in Table 2.3, 2.4 and 2.5. No results beyond LO are known for the fully or semi hadronic cases. Processes matched to PS are known in VBS approximation, i.e. $t-u$ interferences and tri-boson contributions neglected and for both QCD/EW-induced process.

In the next Section, theoretical features related to vector boson polarization will be presented.

2.4 Longitudinal Boson Scattering

In this Section, we would like to re-compute the scattering amplitude of longitudinal W bosons making use of the *equivalence theorem*. The theorem is described in App. (B.3) and it allows to greatly simplify the calculations of scattering amplitude. In particular, we will perform the calculation before and after the Higgs boson, to show that the presence of the Higgs field (the SM Higgs) makes the theory stable at all energies.

In order to work under the approximation defined by the Goldstone theo-

Order	$\mathcal{O}(\alpha^7)$	$\mathcal{O}(\alpha_s\alpha^6)$	$\mathcal{O}(\alpha_s^2\alpha^5)$	$\mathcal{O}(\alpha_s^3\alpha^4)$
NLO	✓[29] [25]	✓[25]	✓[25]	✓[25]
NLO + PS	✓[35]	✓[36]	×	✓[37]

Table 2.3: State of the art for the $W^\pm W^\pm$ process computation at NLO and NLO matched to PS for different EW and QCD orders.

Order	$\mathcal{O}(\alpha^7)$	$\mathcal{O}(\alpha_s\alpha^6)$	$\mathcal{O}(\alpha_s^2\alpha^5)$	$\mathcal{O}(\alpha_s^3\alpha^4)$
NLO	✓[27]	✓[27] [38]	×	✓[39]
NLO + PS	×	✓[40]	×	✓[39]

Table 2.4: State of the art for the $W^\pm Z$ process computation at NLO and NLO matched to PS for different EW and QCD orders.

Order	$\mathcal{O}(\alpha^7)$	$\mathcal{O}(\alpha_s\alpha^6)$	$\mathcal{O}(\alpha_s^2\alpha^5)$	$\mathcal{O}(\alpha_s^3\alpha^4)$
NLO	×	✓[41]	×	✓[42]
NLO + PS	×	✓	×	✓[42]

Table 2.5: State of the art for the ZZ process computation at NLO and NLO matched to PS for different EW and QCD orders.

rem (high energy behaviour, see App. (B.3)), we can set the gauge-less limit[6] and consider that Feynman rules applied to Goldstone Bosons are entirely contained in the Higgs kinetic term. Looking at Fig. (2.13) we can write down the

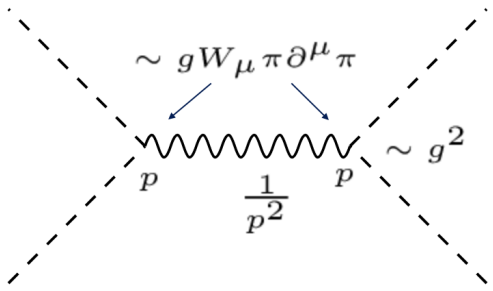


Figure 2.13: Feynman diagram representation of a Higgs-like contribution.

corresponding Lagrangian density expanding the exponential of the Goldstone bosons up to quartic Goldstone boson interactions[7]:

$$\boxed{\mathcal{L} = \frac{1}{2}\partial_\mu h \partial^\mu + \frac{m_h^2}{2}h^2 + \frac{v^2}{4}Tr\left(\partial_\mu U \partial^\mu U\right)\left(1 + 2a + \frac{h}{v} + b\frac{h^2}{v^2}\right)} \tag{2.3}$$

where $U = e^{\frac{i\pi^a\sigma^a}{v}}$, a, b are free parameters and h is the Higgs field. Now, if we consider all possible channels related to the longitudinal WW scattering shown in Fig. (2.14), we can then calculate the scattering amplitude as:

$$\begin{aligned}&\mathcal{A}\left(\pi^a(p_1) + \pi^b(p_2) \to \pi^c(p_3) + \pi^d(p_4)\right)\\ &= \left[\frac{s}{v^2} - \frac{a^2}{v^2}\left(\frac{s^2}{s - m_h^2}\right)\right]\delta^{ab}\delta^{cd} + \text{perm.}\end{aligned} \tag{2.4}$$

If $s >> m_h^2$, we can expand Eq. (2.4) and obtain a simplified expression for the amplitude:

$$\boxed{\mathcal{A} = \left(1 - a^2\right)\frac{s}{v^2} - \frac{a^2 m_h^2}{v^2} = \left(1 - a^2\right)\frac{s}{v^2} - 8a^2\lambda} \tag{2.5}$$

[6]The gauge couplings is set to zero.

[7]Since the diagram involves gauge couplings it does not grow with energy.

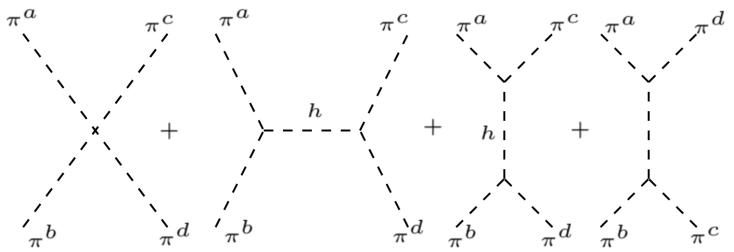


Figure 2.14: Tree–level diagrams for the longitudinal WW scattering.

where λ is the dimensionless self-Higgs-coupling constant.

According to the SM, this amplitude is just a constant at high energy, so a is predicted to be equal to 1. This parameter is assigned a different value in other BSM models and this would reflect into a modification of Higgs couplings to gauge bosons, couplings which are now constrained by experiments at the 10% level.
These considerations were of fundamental importance before, and even more now, after the discovery of the Higgs boson. In a precision era like the one we are living in, they set a strong constraint on the Higgs coupling to gauge bosons. Deviation from SM prediction would indicate undoubtedly presence of new physics.

To summarize the two main aspects of the above discussion:

1. the total cross section of a simple tree level process like the one in Fig. (2.14) is a function of the center of mass energy;

2. its behaviour depends on both the Higgs mass and the Higgs couplings to vector bosons.

These facts can be appreciated and summarised in Fig. (2.15) [43], where the dependence of the cross-section for W^+W^+ scattering process, according to different polarization flavour, is shown as a function of the center of mass energy for different values of the Higgs mass (top) and Higgs couplings (bottom).

In the next Section we will see how, in particular, longitudinal component of the Vector Boson Scattering could be a probe candidate, starting from theoretical assumptions and then focusing on experimental aspects.

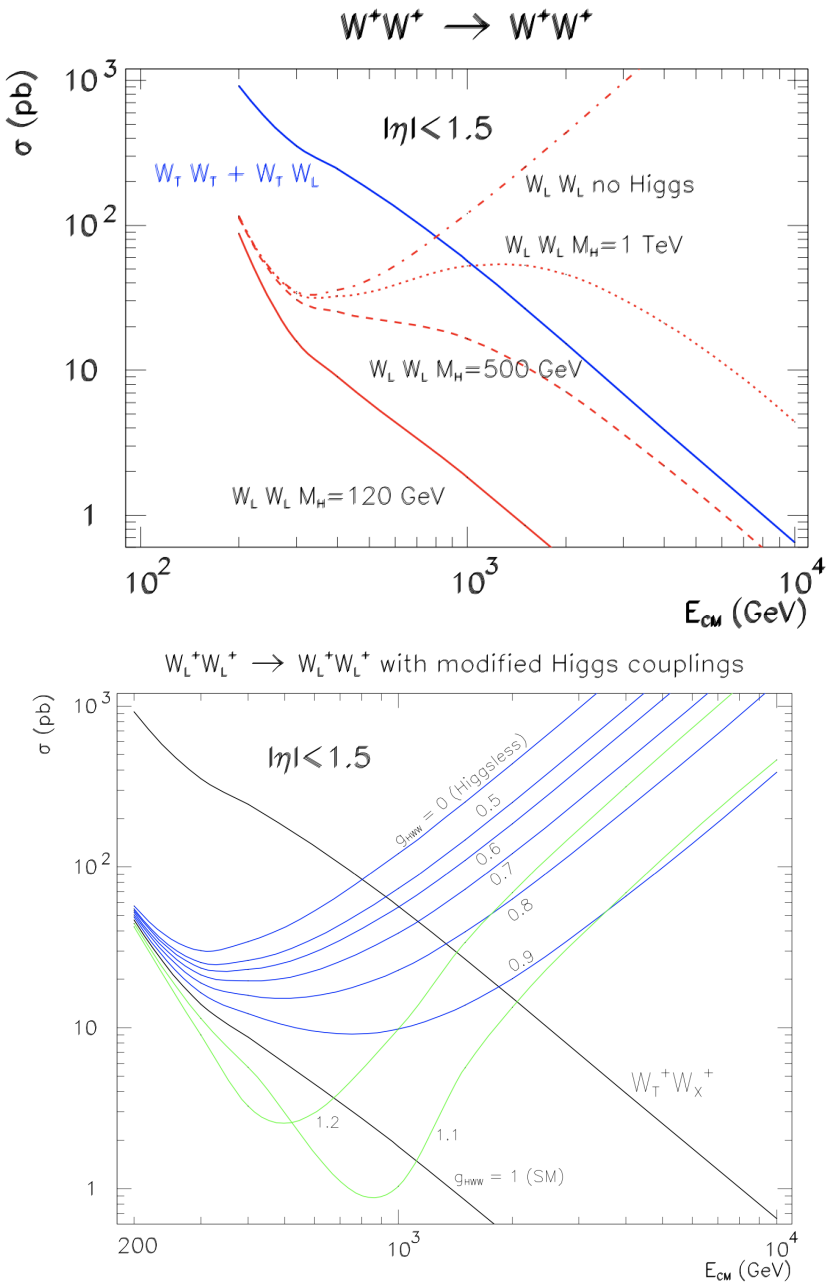


Figure 2.15: Top figure: The total W^+W^+ scattering cross sections as a function of the center of mass energy for different final (and initial) state polarizations and for different Higgs masses. The Higgsless case is also included.
Bottom figure: total $W_L^+W_L^+$ scattering cross sections as a function of the center of mass energy for different values of the HWW coupling g_{HWW}. Coupling $g_{HWW} = 1$ (lower black curve) corresponds to the Standard Model. Blue curves represent $g_{HWW} < 1$, the curve for $g_{HWW} = 0$ is equivalent to the Higgsless case. Green curves represent $g_{HWW} > 1$.
Both figure obtained from MadGraph5_aMC@NLO calculation, cutting on the scattering angle i.e. pseudorapidity of ± 1.5 [43].

2.5 Polarization in VBS

As discussed previously, the scattering of longitudinally polarized vector bosons is a process highly sensitive to the presence of new physics. Being able to experimentally isolate the longitudinal component of the W polarization is a key feature to indirectly access the Higgs sector and the mechanism of EWSB. This is the main topic addressed by this thesis work.
In the following, we will focus on polarization from the theoretical point of view, in order to achieve phenomenological prescriptions and to prepare tools that can be used in experimental analyses, as it will be discussed in the next Chapters.

2.5.1 Polarization Degree of Freedom

Let's consider a massive vector field Lagrangian:

$$\mathcal{L} = -\frac{1}{4}V_{\mu\nu}V^{\mu\nu} + \frac{m^2}{2}V_\mu V^\mu. \tag{2.6}$$

A massive vector has three propagating degrees of freedom (polarizations). Here, the vector field V_μ has four components, and the $4-$dimensional vector representation of the Lorentz group decomposes into a scalar (1 d.o.f.) and a spin-1 particle (3 d.o.f.) under $3-$dimensional rotations. The scalar degree of freedom (due to the non-compact nature of the Lorentz group) has a negative kinetic term (negative norm). If it were physical, it would spoil the probabilistic interpretation of quantum mechanics. To avoid this problem, it is necessary to project away the scalar (time-like) component of the vector field. This is done by requiring all vector polarizations to be orthogonal to the four momentum as a constraint from the equation of motion:

$$\partial_\mu V^\mu = 0 \qquad \Longrightarrow \qquad p_\mu \epsilon_i^\mu = 0.$$

The three physical polarizations of our vector field (in the rest frame) can be chosen, without losing generalisation, aligned to the three axes (x, y, z):

$$\boxed{k_\mu = (m, 0, 0, 0) \qquad \epsilon_\mu^1 = (0, 1, 0, 0) \qquad \epsilon_\mu^2 = (0, 0, 1, 0) \qquad \epsilon_\mu^3 = (0, 0, 1, 0)} \tag{2.7}$$

To identify the transverse and longitudinal components with respect to the three momentum, we need to preform a boost, according to general convention in the z-direction. The first two polarizations remain unchanged and, composed, give rise to the transverse ones, while the third becomes the longitudinal polarization mode.

$$\boxed{k_\mu = \left(\sqrt{k_z^2 + m^2}, 0, 0, 0\right) \qquad \epsilon_\mu^3 = \left(k_z, 0, 0, \sqrt{k_z^2 + m^2}\right)} \tag{2.8}$$

We see that the longitudinal polarization vector grows with the energy and it is at the origin of the growing with energy of the amplitude, already confirmed by previous calculation.
The energy growing amplitude is signalling the need, at some energy scale, of new dynamics and/or new degrees of freedom, to constrain its unitarity. We have already found the solution to this problem, which is the same of generating masses through SSB, namely the Higgs mechanism discussed in Chapter 1. We recall here the only aspects of SSB where polarization components play a crucial role.
As just recalled, while a massive vector is described by three degrees of freedom, a massless (gauge) vector has only two degrees of freedom. If we try to take the zero mass limit, the massive case presents the problem of the *longitudinal* polarization blowing up to infinity. Starting from the simpler case of electromagnetism we have:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

where $F^{\mu\nu}$ is the usual differential 2-form that is an antisymmetric rank-2 tensor field-on Minkowski space that describes the electromagnetic field. In Chapter 1, we proved that this Lagrangian is invariant under $U(1)$ local gauge transformations:

$$A_\mu \to A_\mu + \frac{1}{e} \partial_\mu q(x).$$

We can use this gauge invariance to impose the constraint of removing the time-like polarization (Lorentz gauge):

$$\partial_\mu A^\mu = 0$$

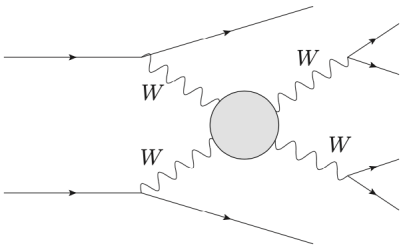
However, now the **rest frame** four-momentum is different and the three polarization vectors satisfying equation above become:

$$k_\mu = (E, 0, 0, E) \qquad \epsilon^1_\mu = (0, 1, 0, 0) \qquad \epsilon^2_\mu = (0, 0, 1, 0) \qquad \epsilon^3_\mu = (0, 0, 0, 1)\,.$$

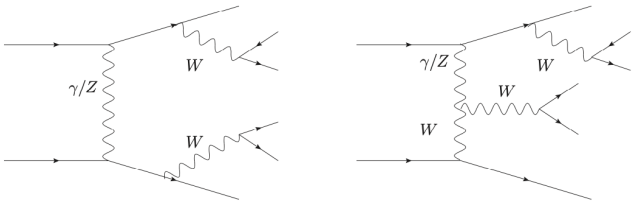
The third polarization vector is unphysical. In fact, it is proportional to the four momentum and again describes a scalar degree of freedom:

$$A_\mu = \partial_\mu \phi \tag{2.9}$$

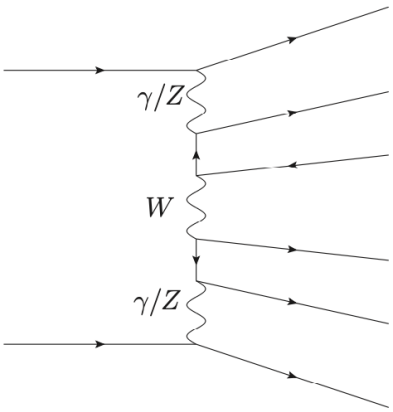
However, A_μ is, by gauge invariance, arbitrarily defined up to the divergence of a scalar function, so gauge equivalent to $A_\mu = 0$.
This proofs that the third polarization vector is not physical (is said to be a pure gauge configuration) and that massive vector fields have three propagating degrees of freedom, with the constraint $k_\mu V^\mu = 0$.



(a) Resonant diagrams: signal



(b) Resonant diagrams: irreducible background.



(c) Non resonant diagrams.

Figure 2.16: WW scattering diagram classes.

2.6 Polarization Discrimination

The most direct way to get access to polarizations of vector bosons is the analysis of angular distributions of their decay products: it is a model independent approach, which has already been proposed for EWSB studies, e.g. in [44] for the opposite sign WW scattering in the semi–leptonic decay.

2.6.1 Amplitude Estimation

We start our polarization studies for VBS from the definition of cross section for polarized Ws, dropping all non resonant diagrams, and projecting on shell the resonant ones, thus preserving gauge invariance [45]. A vector boson production at tree level, in general, receives contributions from different classes of diagrams, both resonant and non resonant, as represented in Fig. (2.16a, 2.16c).

Considering a single intermediate W^+ which decays leptonically, each diagram includes one EW propagator of time-like momentum. In this case the amplitude can be written, in the unitary gauge, as:

$$\mathcal{M} = \mathcal{M}_\mu \frac{i}{k^2 - M^2 + i\Gamma_w M} \left(-g^{\mu\nu} + \frac{k^\mu k^\nu}{M^2}\right) \left(\frac{-ig}{2\sqrt{2}} \bar{\psi}_l \gamma_\nu (1-\gamma^5) \psi_{\nu_l}\right); \quad (2.10)$$

where M and Γ are the W and the W width, respectively.[8]
Let's consider a frame in which an off shell W boson propagates along the z-axis, with momentum k, energy E and invariant mass $\sqrt{Q^2} = \sqrt{E^2 - \kappa^2}$.
It can be shown, writing the tensor explicitly, that the tensor part of a vector boson propagator can be written as a sum over all available polarization states of the boson, i.e. in terms of four polarization vectors [47]:

$$-g^{\mu\nu} + \frac{k^\mu k^\nu}{M^2} = \sum_{\lambda=1}^{4} \epsilon_\lambda^\mu(k)\epsilon_\lambda^{\nu*}(k); \tag{2.11}$$

where the polarizations read:

$$\begin{aligned}
\epsilon_L^\mu =& \frac{1}{\sqrt{2}}(0, +1, -i, 0) \text{ (left)};\\
\epsilon_R^\mu =& \frac{1}{\sqrt{2}}(0, -1, -i, 0) \text{ (right)};\\
\epsilon_0^\mu =& (\kappa, 0, 0, E)/\sqrt{Q^2} \text{ (longitudinal)};\\
\epsilon_A^\mu =& \sqrt{\frac{Q^2 - M^2}{Q^2 M^2}}(E, 0, 0, \kappa) \text{ (auxiliary)}.
\end{aligned} \tag{2.12}$$

In this framework, standard orthogonality constraints hold: for longitudinal and transverse polarizations $\epsilon_i \cdot \kappa = 0$, $\epsilon_i \cdot \epsilon_j^* = -\delta_{ij}, i, j = 0, L, R$.
Regarding the auxiliary component that, as we discussed in Sec. (2.5.1), is not physical, we have:

$$\epsilon_A \cdot \epsilon_i^* = 0, L, R,\ \epsilon_A \cdot \epsilon_A^* = (Q^2 - M^2)/M^2 \text{ and } \epsilon_A \cdot \kappa = \sqrt{(Q^2 - M^2)Q^2/M^2}.$$

In case of an on-shell W ($Q^2 = M^2$) the fourth polarization vanishes and other components transform accordingly. The most general case, in which the W propagates along a generic direction, can easily be obtained by a rotation. The decay amplitudes of the W, as explicitly calculated using the helicity amplitude method by Hagiwara-Zeppenfeld in App. (D.3), depends on its polarization and is given by:

$$\mathcal{M}_\lambda^{\mathcal{D}} = \frac{-ig}{2\sqrt{s}}\bar{\psi}_l \epsilon_\lambda^{\mu*} \gamma_\mu (1 - \gamma^5)\psi_{\nu_l}. \tag{2.13}$$

The main result of that calculation is that, in the rest frame of the W, we have:

$$\begin{aligned}
\mathcal{M}_0^{\mathcal{D}} =& ig\sqrt{2}E \sin\vartheta\\
\mathcal{M}_{R/L}^{\mathcal{D}} =& igE(1 \pm \cos\vartheta)e^{\pm i\phi}
\end{aligned} \tag{2.14}$$

Each physical polarization is uniquely associated with a specific angular distribution of the charged lepton, regardless of the W energy [46], and it is due to the W spin-1 nature, typical of a massive gauge bosons.

[8] This formulation can be generalized to account for Z boson [46].

2.6.2 Polarization Fractions

In the following, we will focus on the charged lepton angular distributions and their relation to the polarization parameters. This part will connect the previous theoretical discussion to important aspects of experimental observables. From a general perspective, the W boson is an arbitrary superposition of the 3 polarization states, so we can define a polarized production amplitude:

$$\mathcal{M}_{\lambda}^{\mathcal{P}} = \mathcal{M}_{\mu}\epsilon_{\lambda}^{\mu}; \tag{2.15}$$

and the full amplitude can be written as:

$$\mathcal{M} = \sum_{\lambda=1}^{3} \mathcal{M}_{\lambda}^{\mathcal{P}} \frac{i}{k^2 - M^2 + i\Gamma_{\mu} M} \mathcal{M}_{\lambda}^{\mathcal{D}} = \sum_{\lambda=1}^{3} \mathcal{M}_{\lambda}^{\mathcal{F}}; \tag{2.16}$$

where $\mathcal{M}_{\lambda}^{\mathcal{F}}$ is the full amplitude with a single polarization for the intermediate W.
To evaluate the differential cross-section of the process we need the squared amplitude of the matrix element that can be write as:

$$\underbrace{|\mathcal{M}|}_{\text{coherent sum}} = \underbrace{\sum_{\lambda} |\mathcal{M}_{\lambda}^{\mathcal{F}}|^2}_{\text{incoherent sum}} + \underbrace{\sum_{\lambda \neq \lambda'} \mathcal{M}_{\lambda}^{\mathcal{F}^*} \mathcal{M}_{\lambda'}^{\mathcal{F}}}_{\text{interference term}} \tag{2.17}$$

The interference terms in Eq. (2.17) are not, in general, zero. They cancel only when the squared amplitude is integrated over the full range of the angle ϕ, however, the acceptance and kinematical cuts on leptons compromise this aspect [45]. Therefore, there is no exact cancellation and the incoherent sum of polarized squared amplitudes does not correspond exactly to the squared amplitude of the unpolarized process.
A step forward can be done by extracting the polarization fractions in the W leptonic decay. Working in the W boson rest frame, where (ϑ, ϕ) are the charged lepton angles with respect to the W boson flight direction (in the laboratory frame), we can define the integral of fully differential cross section as:

$$\frac{d\sigma(X)}{dX} = \int d\cos\vartheta \frac{d\sigma(\vartheta, X)}{d\cos\vartheta dX}; \tag{2.18}$$

where the integral in Eq. (2.18) extends to the entire $\phi-$angle range $(0, 2\pi)$ and X stands for all additional phase space variables.
Combining Eq. (2.14) with Eq. (2.18) and considering that the boson polarization fraction are not known in principle, we can write the differential cross section with respect to $\cos\vartheta$ introducing polarization fractions, to keep track

from which vector polarization each term is originated:

$$\frac{1}{\frac{d\sigma(X)}{dX}}\frac{d\sigma(\vartheta,X)}{d\cos\vartheta dX} = \frac{3}{8}(1\mp\cos\vartheta)f_L(X) + \frac{3}{8}(1\pm\cos\vartheta)^2 f_R(X) + \frac{3}{4}\sin^2 f_0(X) \tag{2.19}$$

polarization fractions f_i depend on the variables X which are not integrated over, so we can impose a normalization such that $f_0 + f_L + f_R = 1$.
An interesting output of the work reported in [45] is that, under the hypothesis that led us to write Eq. (2.19), the polarized components can be extracted from the differential angular distribution by a projection on the first three Legendre polynomials:

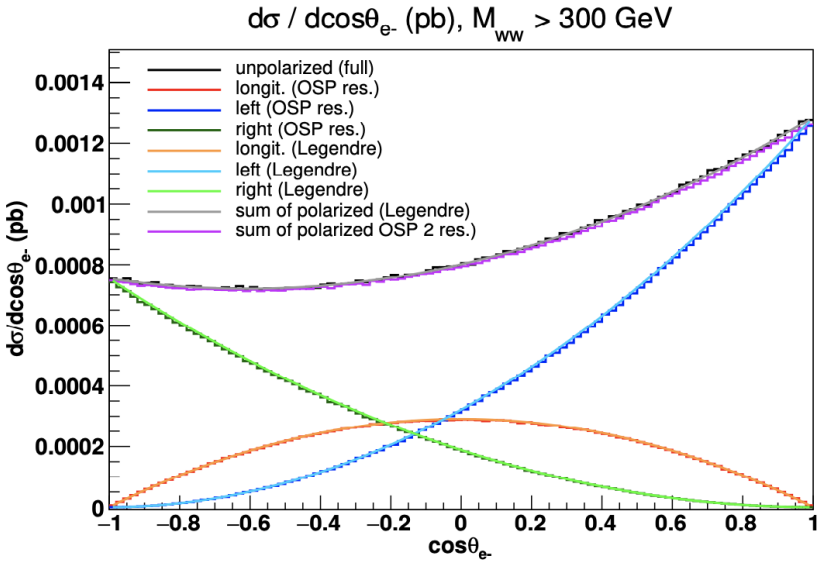
$$\frac{1}{\sigma}\frac{d\sigma}{d\cos\vartheta} = \sum_{l=0}^{2}\alpha_l P_l(\cos\vartheta), \qquad \alpha_l = \frac{2l+1}{2}\int_{-1}^{1} d\cos\vartheta \frac{1}{\sigma}\frac{d\sigma}{d\cos\vartheta}P_l(\cos\vartheta). \tag{2.20}$$

Then, the polarization fractions can be written, according to the W boson sign, as:

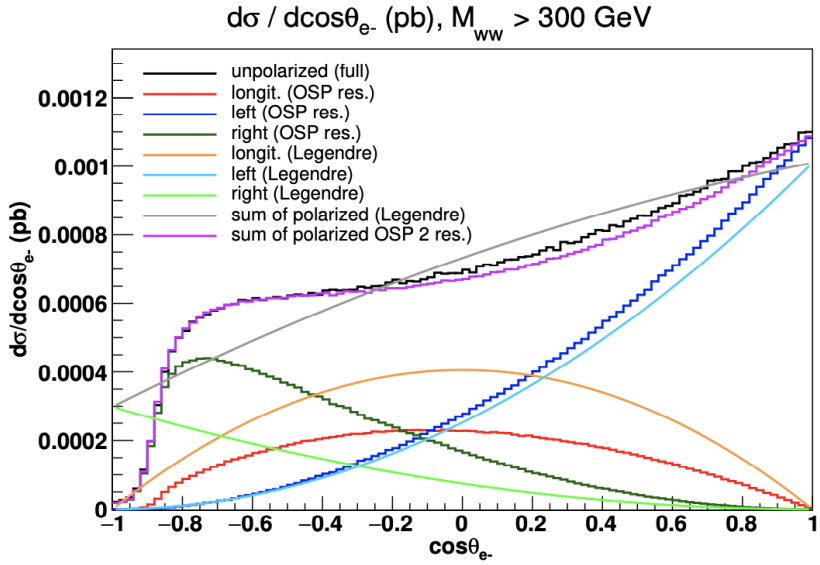
$$\boxed{\begin{aligned} f_0 &= \frac{2}{3}(\alpha_0 - 2\alpha_2); \\ f_L &= \frac{2}{3}(\alpha_0 \mp \alpha_1 + \alpha_2); \\ f_R &= \frac{2}{3}(\alpha_0 \pm \alpha_1 + \alpha_2). \end{aligned}} \tag{2.21}$$

Unfortunately, this elegant mathematical representation, which would be also an interesting experimental handle, does not hold anymore when acceptance cuts are imposed on charged leptons. While the cancellation of the interference terms in Eq. (2.17) is a necessary condition for the validity of Eq. (33) in App. (A.1), this is by no means sufficient.
It is a matter of fact that events generation, by means of any of the existing MC programs, should take into account cuts that include a realistic representation of the detector acceptance, if not also dead, i.e. not sensitive, phase space regions. Integrating over different phase space variables, in the presence of cuts, results in a different theta dependence in the formulas above. As a consequence, the measured lepton decay distribution is not described any more by Eq. (2.19) and the polarization fractions cannot be computed as in Eq. (2.21). In order to separate the polarized components in the data, it is necessary to compute the individual amplitudes $\mathcal{M}_{\mathcal{F}}^{\lambda}$ in Eq. (2.16).

In Fig. (2.17) (taken from [45]) we can appreciate how the introduction of simple cuts on lepton momentum and pseudorapidity completely spoils the Legendre polynomial behaviour described by Eq. (2.20) (upper plot: polarization distribution when no cuts a part from $M_{WW} < 300$ GeV are applied; lower plot: same variables plotted when kinematical cuts on the lepton are



(a) $\cos\theta_e$



(a) $\cos\theta_e$

Figure 2.17: Distribution of electron $\cos\vartheta$ in the W^- reference frame. (top) Polarization components obtained by expanding the full angular distribution on Legendre polynomials are shown in lighter colors. The darker histograms are obtained integrating the polarized amplitudes squared. (bottom) Same distribution in case of the following cuts: $p_t^e > 20$ GeV, $|\eta_e| < 2.5$ [45]. Different polarization component are shown: unpolarized full computation (black), unpolarized On Shell Projection (grey) (see App. (D.3) for its definition), longitudinal-longitudinal (red), longitudinal-transverse (light green), transverse-longitudinal (dark green), transverse-transverse (blue). The sum of the doubly-polarized distributions is shown in orange. In both figures the black and grey curves perfectly overlap.

imposed). The most prominent feature in comparison to the curves on the left is the depletion at $\cos\vartheta_e = -1$ in the angular distribution. The right handed component is the one mostly affected, with the resulting shape substantially different from the $(1-\cos\vartheta)^2$ behaviour displayed in the absence of cuts. The effect is largely related to the preferred direction of emission of the charged leptons from right handed W^-s, which tend to produce leptons with smaller transverse momentum.

Final Remarks

We conclude this Chapter by remarking that, despite the difficulties connected to the impositions of cuts to VBS event, is it still possible to extract polarization fractions using singly polarized SM Monte Carlo templates. The same templates can be used to measure the polarization of the W with reasonable accuracy, even if new physics is present.
In the next Chapter, we will present a detailed study of the Vector Boson Scattering process starting from a systematic comparison of the various approximations against the complete calculation, at LO and NLO QCD accuracy.

Chapter 3

Montecarlo for VBS

The VBS process involving two same-sign W bosons both leptonically decaying, is, among all VBS contributions, the one with the largest signal-to-background ratio at the LHC and it is thus of prime importance to provide predictions, with systematic uncertainties at least comparable to the current and envisaged experimental precision. The W^+W^+ scattering, is also the simplest VBS process to calculate, because the double-charge structure of the leptonic final state limits the number of partonic processes and the total number of Feynman diagrams for each process. Nonetheless, it possesses all features of VBS, therefore, it is the ideal candidate for a comparative study of the different simulation tools. Recently, the complete NLO corrections to W^+W^+ have been evaluated in [25], making it possible for the first time to study in detail the quality of the VBS approximations at NLO QCD.

This Chapter starts with the description of the physical process and the various computer codes used, in Sec. (3.1). In Sec. (3.2) we show results coming from a leading-order study of the different contributions which lead to the production of two same-sign W bosons and two jets. There, predictions for VBS from different Montecarlo tools are compared, at the level of the cross section and differential distributions. In Sec. (3.3) we extend the comparison to the NLO predictions. Finally, the effect of the inclusion of matching LO and NLO computations to parton shower is discussed in Sec. (3.4).

Thanks to a Short Term Scientific Mission (STSM) supported by the VBS Cost Action VBSCan, I had the possibility to spend several days at Nikhef (Dutch National Institute for Subatomic Physics) to set the basis, discuss and consolidate, with the colleagues involved in the Action, the analysis of the data presented in this Chapter. The main outcome of the work done during that period flew into a paper which has been published in *EPJ* [37]. In this paper, we presented a detailed study of the Vector Boson Scattering process starting from a systematic comparison of the various approximations against the complete calculation, at LO and NLO QCD accuracy. This Chapter is inspired by part of these studies.

3.1 Problem Statement

As discussed in the previous Chapter, the process under evaluation is VBS, which in W^+W^+ declination is characterized by the following signature:

$$\boxed{pp \to \mu^+\nu_\mu e^+\nu_e\, jj + \mathrm{X}.} \tag{3.1}$$

From the previous Chapter we know that, at LO, this process can proceed via three different coupling-order combinations: $\mathcal{O}(\alpha^6)$, $\mathcal{O}(\alpha_s^2\alpha^4)$, and $\mathcal{O}(\alpha_s\alpha^5)$. Among several channels (u, t, s), we will discuss now how the s-channel contributions play a particular role in the study of the various diagrams.
In this work, we focus on LO or NLO(-QCD) predictions at order $\mathcal{O}(\alpha^6)$ and $\mathcal{O}(\alpha_s\alpha^6)$, respectively. From now on, unless otherwise specified, we will refer to these NLO corrections.
An important aspect to take into account when using approximations, or working with a specific order ($\mathcal{O}(\alpha^6)$) is that a gauge-invariant subsets of diagrams must be considered to obtain physically meaningful results. In this sense, considering just those diagrams which contain the $2 \to 2$ process for the scattering of two W bosons, which includes the quartic gauge-boson vertex, represents a non gauge invariant approximation of the full process. In order to ensure gauge invariance, an on-shell projection of the incoming and outgoing W bosons should be performed.
We make use of an improved version of a technique already introduced in Chapter 2 called *effective vector boson approximation* (EVBA), where we consider all t- and u-channel diagrams and we square them separately. We neglect interference contributions between the two classes (both phase-space and colour suppressed, see our original work [37]) and we discard s-channel squared diagrams and any interferences between them and the t/u-channels. This procedure allow us to define a *VBS approximation*, which can be considered a gauge-invariant approximation.
The VBS approximation can be extended to NLO in a straightforward manner for what concerns the virtual contributions. For the real-emission contributions special care must be adopted for the gluon-initiated processes. The initial-state gluon and initial-state quark must not couple together, otherwise infrared (IR) divergences proportional to s-channels appear, which do not cancel with the ones found in the virtual contributions. Consequently, a further refinement is to add the squared matrix element of the s-channel contributions to this approximation.
As already discussed in Chapter 2 and in particular in [25], QCD and EW corrections cannot be separated any more on the basis of Feynman diagrams, since some loop diagrams contribute to both types of corrections. This is the case when considering the full NLO corrections of order $\mathcal{O}(\alpha_s\alpha^6)$: besides real and virtual QCD corrections to the EW tree-level contribution of order $\mathcal{O}(\alpha^6)$, real and virtual EW corrections to the LO interference of order $\mathcal{O}(\alpha_s\alpha^5)$ have to be taken into account.

3.1.1 Montecarlo Tools

In the following, we provide an overview of all the Montecarlo generators used in this comparison study and how they have been tuned. They are both VBS specific and general purpose generators. Some of them are capable of just LO calculation, others also allow NLO (QCD and/or EW) implementation.
Montecarlo event generators are essential tools for the comparison of theory and experiment in High Energy Physics. This is due to their capability of producing hard interaction between partons, usually calculated at fixed order, via matrix elements. Hard interactions generate resonances, such as W bosons, that decay into different products according to the branching ratio of the selected process, and are managed by the generator itself with a mother-daughter elaboration convention.

PHANTOM

The Montecarlo event generator mainly used and investigated during this thesis studies is: PHANTOM [48]. PHANTOM is a Montecarlo event generator capable of performing exact calculations of matrix-element at tree level, without using any production times decay approach for the W bosons interactions. Tree-level matrix elements are generated in the complex-mass scheme [49, 50] computed via the modular helicity formalism [51, 52]. The integration uses a multi-channel approach and an adaptive strategy [53]. It is computationally efficient[1] since it is a dedicated generator of tree level process for six parton final states at $O(\alpha^6)$ and $O(\alpha^4\alpha_s^2)$ in perturbation theory, including possible interferences between the two sets of diagrams. It can generate events for any set of processes and it is interfaced to parton shower and hadronization packages via the latest Les Houches Accord protocol [54]. Not only it is efficient, but also enables several physical contribution like exact spin correlations between the decays of different heavy particles, effect of the non resonant background, relevance of the offshellness of boson decays, interferences between different sub-amplitudes which cannot be tackled without a complete calculation. PHANTOM generates unweighted events at parton level for both the SM and a subset of beyond the Standard Model (BSM) theories.

WHIZARD

WHIZARD [55, 56] is a multi-purpose event generator with LO matrix-element generator O'Mega. For QCD amplitudes it uses the colour flow formalism [57]. For NLO QCD calculations [58], where WHIZARD is in the final validation phase, it provides FKS subtraction terms [59], while virtual matrix elements are provided externally mainly by Recola [60]. Furthermore, WHIZARD can automatically provide POWHEG matching to parton shower

[1] The authors declare differences in CPU time of one order of magnitude or more with respect to other MC generator like `Madgraph` [48].

[61]. WHIZARD allows to simulate a huge number of BSM models as well, in particular in terms of higher-dimensional operators for VBS processes including means to provide unitarity limits [62].

BONSAY

The program Bonsay [63] consists of a general-purpose Montecarlo integrator and matrix elements taken from different sources: Born matrix elements are adapted from the program Lusifer [64], which have been generalised to calculate also real matrix elements. Virtual matrix elements have been calculated using an in-house matrix-element generator. One-loop integrals are evaluated using the Collier library [65, 66]. For the results presented here, it uses the VBS approximation at LO and NLO. The virtual corrections are additionally approximated using a double-pole approximation where the final state leptons are assumed to originate from the decay of two resonant W-bosons; see Ref. [67] for the on-shell projection used. At LO the exact matrix elements can also be used.

VBFNLO

VBFNLO [68] is a flexible parton-level Montecarlo for processes which involve EW bosons. It allows for the calculation of VBS processes at NLO QCD in the VBS approximation. The s-channel contributions are available separately as tri-boson processes with semi-leptonic decays. Besides the SM, also a variety of new-physics models including anomalous couplings of the Higgs and gauge bosons can be simulated.

MG5_aMC

MadGraph5_aMC@NLO [69] (MG5_aMC) is an automatic meta-code (a code that generates codes) which makes it possible to simulate any scattering process including NLO QCD corrections both at fixed order and including matching to parton showers, using the MC@NLO method [70]. Finally, scale and PDF uncertainties can be obtained in an exact manner via reweighting at negligible additional CPU cost [71].
No approximation is performed for the Born and real-emission matrix elements.

MoCaNLO+Recola

The program MoCaNLO+Recola is made of a flexible Montecarlo program dubbed MoCaNLO and of the matrix-element generator Recola [60, 72]. It can compute arbitrary processes at the LHC with both NLO QCD and EW accuracy in the SM. This is made possible by the fact that Recola can compute arbitrary processes at tree and one-loop level in the SM. In MoCaNLO+Recola no approximation is performed neither at LO nor at NLO.

It implies that also contributions stemming from EW corrections to the interference are computed. The full NLO corrections to VBS and its irreducible background that we mention in Sec. (2.2) [25] have been obtained thanks to this tool.

Parton Showering and Hadronization

Dealing with real data coming from experiments means that we must be able to generate the whole production chain, from the hard interaction to detector simulation. Hard interaction creates only a fraction of the particles included in the final state observed by a detector at hadronic colliders. The non-perturbative phenomena, such as initial and final state QCD radiation, multi-parton interaction, and hadronization are simulated thanks to other Montecarlo distributions like PYTHIA [73]. Similar to PYTHIA, in this work we made use also of HERWIG7 [74]. Finally, the stable (or long-living particles such as muon, pions and K) particles are clustered in jets by the `Fastjet` package, included in DELPHES [75] fast simulation.
DELPHES is a framework for fast simulation of collider experiments. It mimics the detector effects on the measured quantities, such as final state physics objects. This fast simulation includes a generic tracking system embedded into a magnetic field, electromagnetic and hadronic calorimeters and a muon system. Each of the subsystems can be customized to reproduce the effect of any specific experiment sub-detector. This last step of the simulation chain, has been used for a study that will be presented in Chapter 4.

3.1.2 VBS Selection Parameters

In Tab. (3.1) we summarise the main characteristics of the MC codes used in this work, shortly described in the previous paragraph. In particular, we highlight additional specifications for each tool, i.e. phase space integration, channel approximation and virtual or real corrections assumptions that are required when the order of calculation move from LO to NLO. In particular, it is specified whether:

- all s-, t-, u-channel diagrams are included;
- interferences between diagrams of different types are included at LO;
- diagrams which do not feature two resonant W bosons are included;
- the so-called non-factorisable (NF) QCD corrections, *i.e.* the corrections where (real or virtual) gluons are exchanged between different quark lines, are included;
- EW corrections to the interference of order $\mathcal{O}(\alpha^5\alpha_s)$ are included. These corrections are of the same order as the NLO QCD corrections to the contribution of order $\mathcal{O}(\alpha^6)$ term.

Table 3.1: Summary of the different precision implemented in the Montecarlo codes involved in the comparison.

Code	$\mathcal{O}(\alpha^6)$ s,t,u	$\mathcal{O}(\alpha^6)$ interf.	Non-res.	NLO	NF QCD	EW corr. to order $\mathcal{O}(\alpha_s\alpha^5)$
PHANTOM	s,t,u	Yes	Yes	No	-	-
WHIZARD	s,t,u	Yes	Yes	No	-	-
BONSAY	t,u	No	Yes, virt. No	Yes	No	No
POWHEG	t,u	No	Yes	Yes	No	No
VBFNLO	s,t,u	No	Yes	Yes	No	No
MG5_AMC	s,t,u	Yes	Yes	Yes	virt. No	No
MOCANLO+RECOLA	s,t,u	Yes	Yes	Yes	Yes	Yes

The hadronic scattering processes are simulated at the LHC with a centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$. The NNPDF 3.0 parton distribution functions with five massless flavours[2], NLO-QCD evolution, and a strong coupling constant $\alpha_s(M_Z) = 0.118$ are employed in the present calculation. The EW coupling is renormalised in the G_μ scheme according to:

$$\alpha = \frac{\sqrt{2}}{\pi} G_\mu M_W^2 \left(1 - \frac{M_W^2}{M_Z^2}\right), \tag{3.2}$$

with

$$G_\mu = 1.16637 \times 10^{-5}\text{GeV}^{-2}, \tag{3.3}$$

The Cabibbo–Kobayashi–Maskawa matrix is assumed to be diagonal, meaning that the mixing between different quark generations is considered negligible. The complex-mass scheme [76] is used throughout to treat unstable intermediate particles in a gauge-invariant manner.
The central value of the renormalisation and factorisation scales is set to

$$\mu_R = \mu_F = \sqrt{p_{T,j_1}\, p_{T,j_2}}, \tag{3.4}$$

defined via the transverse momenta of the two hardest jets (identified with the procedure outlined in the following), event by event.

Event selection used in the present study includes the following kinematic criteria:

- The two same-sign charged leptons are required to fulfil the following cuts on transverse momentum, rapidity, and separation in the rapidity–azimuthal-angle separation:

$$p_{T,l} > 20\,\text{GeV}, \qquad |y_l| < 2.5, \qquad \Delta R_{ll} > 0.3\,. \tag{3.5}$$

[2]For the process considered, no bottom (anti-)quarks appear in the initial or final state at LO and NLO, as they would lead to top quarks, and not light jets, in the final state.

- The total missing transverse momentum, computed from the vectorial sum of the transverse momenta of the two neutrinos, is required to be

$$p_{\mathrm{T,miss}} > 40\,\mathrm{GeV}. \tag{3.6}$$

- QCD partons (light quarks and gluons) are clustered together using the anti-k_T algorithm [77], possibly using the FastJet implementation [78], with distance parameter $R = 0.4$. We impose cuts on the jet transverse momenta, rapidities, and their separation from leptons, according to the following:

$$p_{T_j} > 30\,\mathrm{GeV}, \qquad |y_j| < 4.5, \qquad \Delta R_{jl} > 0.3\,. \tag{3.7}$$

  VBS cuts are applied on the two jets with largest transverse momentum, unless otherwise stated. In particular, we impose a cut on the invariant mass of the di-jet system, as well as on the rapidity separation of the two jets:

$$m_{jj} > 500\,\mathrm{GeV}, \qquad |\Delta y_{jj}| > 2.5, \tag{3.8}$$

  if not explicitly stated otherwise.

- When EW corrections are computed, real photons and charged fermions are clustered together using the anti-k_T algorithm [77] with radius parameter $R = 0.1$. In this case, leptons and quarks are defined to be *dressed fermions*.

3.2 LO VBS Approximation

As an investigation of the state-of-art in terms of Montecarlo predictions, we performed a comparison among the different MC generators at LO precision in QCD and EW. Results are summarised here below.
Cross sections calculation done at matrix element with PHANTOM are reported in Tab. (3.2). In the table, the three LO contributions are separately shown, together with their integration errors. They are obtained without applying the VBS cuts of Eq. (3.8) on the variables m_{jj} and $|\Delta y_{jj}|$. The EW, QCD and interference contributions amount to 57%, 37%, and 6% of the total inclusive cross section, respectively. It follows that the interference between EW and QCD contributions is small, due to colour suppression, but not negligible. The three above-mentioned LO contributions are shown in Fig. (3.1) in terms of differential distributions of the di-jet invariant mass m_{jj} and the rapidity difference $|\Delta y_{jj}|$, separately and summed. For the di-jet invariant-mass distribution (left), one can observe that the EW contribution peaks around an invariant mass of about 80 GeV. This is due to diagrams where the two jets originate from the decay of a W boson. They come from s-channel contributions that are not present in calculations relying on the VBS approximation.

Table 3.2: Cross sections at LO accuracy for the three contributions to the process pp $\to$ $\mu^+\nu_\mu e^+\nu_e jj$, obtained with exact matrix elements. These results are for the set-up described in Sec. (3.1.2) but no cuts on m_{jj} and $|\Delta y_{jj}|$ are applied. The uncertainties shown refer to the estimated statistical errors of the Montecarlo integrations.

Order	$\mathcal{O}(\alpha^6)$	$\mathcal{O}(\alpha_s^2\alpha^4)$	$\mathcal{O}(\alpha_s\alpha^5)$
σ[fb]	2.292 ± 0.002	1.477 ± 0.001	0.223 ± 0.003

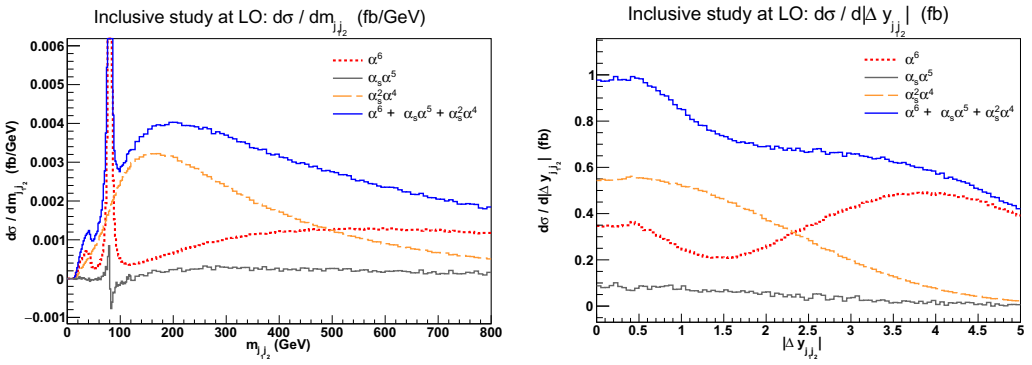


Figure 3.1: Differential distribution in the di-jet invariant mass m_{jj} (left) and the difference of the jet rapidities $|\Delta y_{jj}|$ (right) for the three LO contributions to the process pp $\to$ $\mu^+\nu_\mu e^+\nu_e jj$. The EW contribution is in red, the QCD one in orange, and the interference one in grey. The sum of all the contributions is in blue. The cuts applied are the ones of Sec. (3.1.2) but no cuts on m_{jj} and $|\Delta y_{jj}|$ are applied.

The EW contribution becomes dominant for di-jet invariant mass larger than 500 GeV, likewise for jet-rapidity difference larger than 2.5 (right). Kinematic cuts which we want to employ to select the EW contributions with respect to the QCD one should be based on this distributions. A further confirmation can be obtained from the plots in Fig. (3.2), where the three contributions are displayed in double-differential distributions in the di-jet invariant mass and jet rapidity difference. Again, the region with low di-jet invariant mass is dominated by tri-boson contributions and should be avoided in VBS studies. For this reason, to focus on LO inclusive study, we adopted thie following cut: $m_{jj} > 200\,\text{GeV}$.

In Fig. (3.3), ratios of double-differential cross sections are shown in the $(m_{jj}, |\Delta y_{jj}|)$ plane. For low m_{jj} and low $|\Delta y_{jj}|$, we have significant s-channel contributions and good approximation in fiducial region of the process. Two different situations are displayed: the ratios of $|t|^2+|u|^2$ (left) and $|s|^2+|t|^2+|u|^2$ (right) approximations, both expressed as fractions of the full calculation. In the first case, the approximation is good within $\pm 10\%$ over the whole range, apart from the low invariant-mass region at both low and large rapidity difference. The low rapidity-difference region possesses remnants of the tri-bosons contribution that have a di-jet invariant mass around the W-boson mass. It is therefore expected that the $|t|^2 + |u|^2$ approximation fails in this region. The right plot, where the $|s|^2 + |t|^2 + |u|^2$ approximation is considered, displays a

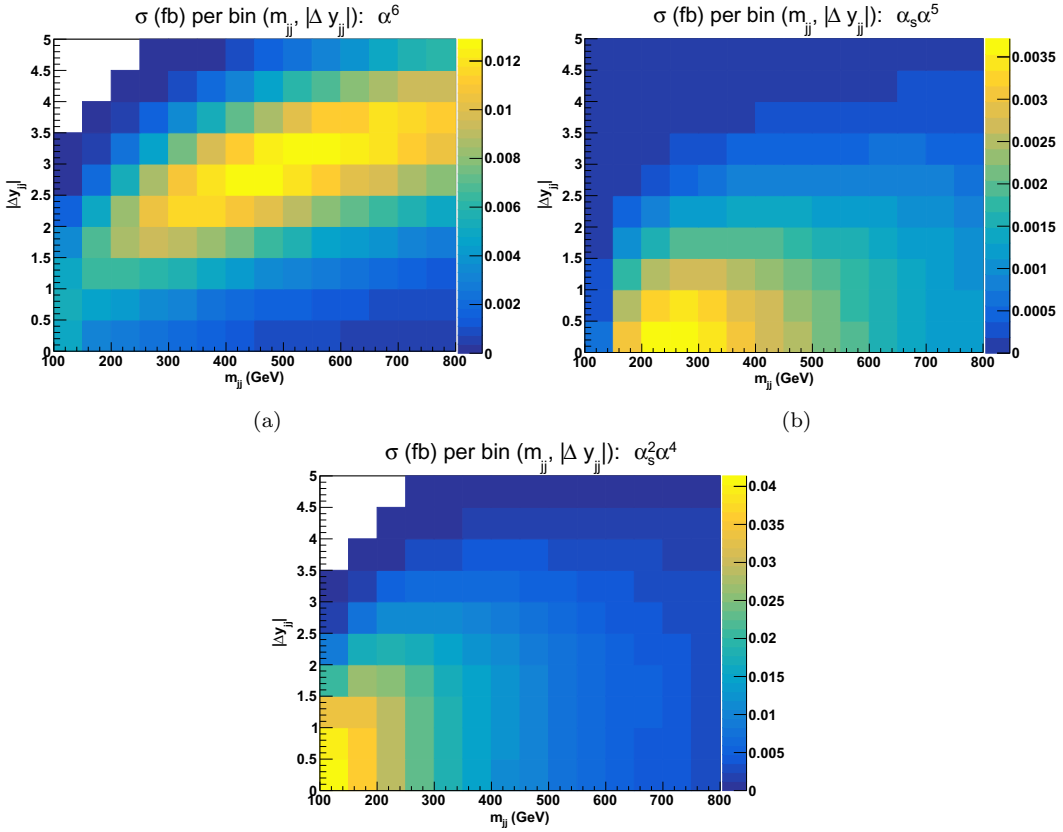


Figure 3.2: Double-differential distributions in the variables m_{jj} and $|\Delta y_{ii}|$ for the three LO contributions of orders $\mathcal{O}(\alpha^6)$ (a), $\mathcal{O}(\alpha_s\alpha^5)$ (b), and $\mathcal{O}(\alpha_s^2\alpha^4)$ (c).

better behaviour in the previously mentioned region.

3.2.1 Fiducial Region

In Tab. (3.3), we report the total rates at LO accuracy at order $\mathcal{O}(\alpha^6)$ obtained in the fiducial region described in Sec. (3.1.2). The different underlying VBS approximations have an impact below 1% in this inclusive calculation[3], performed over the typical phase-space volume used by experimental collaborations. However, the four complete predictions (WHIZARD, PHANTOM, MG5_AMC, and MoCaNLO+Recola) exhibit different statistical errors. After checking the point-wise agreement of the matrix-element, we realised how different implementation of features, like the estimation of the Monte-carlo uncertainty or a non-perfect mapping of the six-body phase-space, could account for a level of ambiguity up to 0.5%. Yet, this is a satisfactory result, if compared to the larger differences observed at NLO or when including

[3]The two most-distant predictions (Powheg-box and MG5_AMC) are only 0.7% apart.

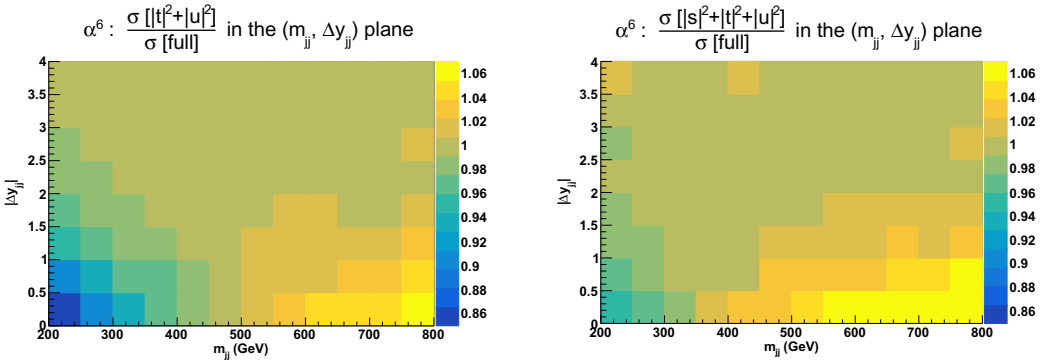


Figure 3.3: Ratios for double-differential distributions in the variables m_{jj} and $|\Delta y_{jj}|$ at LO *i.e.* at order $\mathcal{O}(\alpha^6)$ of approximated squared amplitudes over the full matrix element. The approximated squared amplitudes are computed as $|\mathcal{A}|^2 \sim |t|^2 + |u|^2$ (left) and $|\mathcal{A}|^2 \sim |s|^2 + |t|^2 + |u|^2$ (right). No cuts on m_{jj} and $|\Delta y_{jj}|$ are applied.

Table 3.3: Cross sections at LO accuracy of order $\mathcal{O}(\alpha^6)$. The complete $2 \to 6$ matrix-element, without any approximation, is implemented in by PHANTOM, WHIZARD, MG5_aMC, and MoCaNLO+Recola. The predictions are obtained in the fiducial region described in Sec. (3.1.2). The uncertainties shown refer to the estimated statistical errors of the Montecarlo integrations.

Code	σ[fb]
Bonsay	1.43636 ± 0.00002
Powheg-Box	1.44092 ± 0.00009
VBFNLO	1.43796 ± 0.00005
PHANTOM	1.4374 ± 0.0006
WHIZARD	1.4381 ± 0.0002
MG5_aMC	1.4304 ± 0.0007
MoCaNLO+Recola	1.43476 ± 0.00009

matching to parton shower.

A relatively good agreement among the various tools is confirmed also in Fig. (3.4), where the distributions in the invariant mass (left) and the rapidity difference (right) of the two tagging jets are shown, as key observables for VBS measurements. In both cases, the absolute distributions is plotted in the upper part, while the lower part displays the ratio over the predictions of MoCaNLO+Recola, take, as a benchmark, together with its scale-uncertainty band. This seven-points variation is defined in Eq. (3.11) of [25] where renormalisation and factorisation scales are set dynamically.
The same level of agreement (1% or below in each bin) holds for other standard differential distributions such as rapidity, invariant mass, or transverse momentum, confirming that contributions from s-channel diagrams as well as interferences are suppressed in the fiducial region. At LO, the VBS approximation works well and it completely excludes the phase-space region where tri-boson contributions could have a noticeable impact.

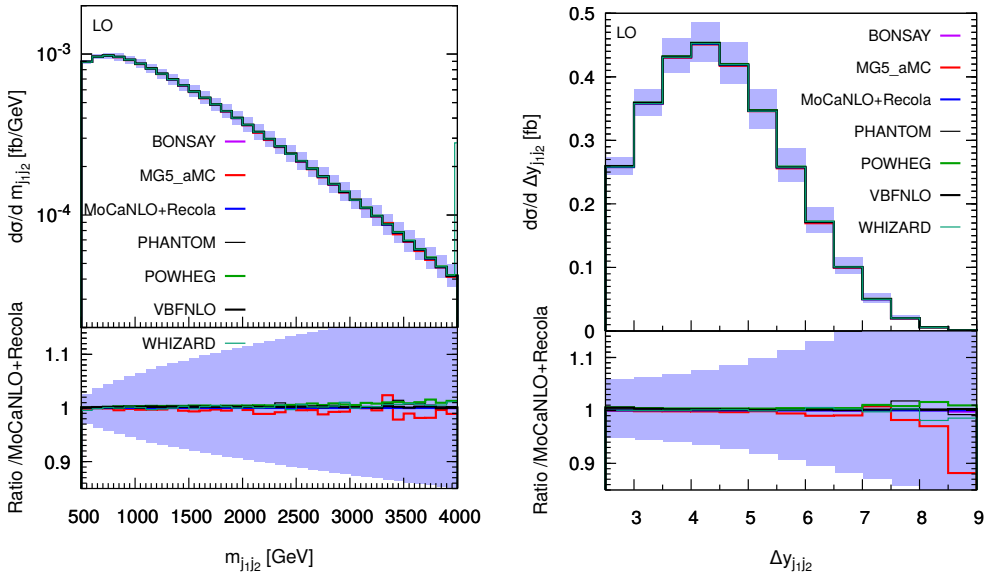


Figure 3.4: Differential distributions in the invariant mass (left) and rapidity difference of the two tagging jets (right) at LO accuracy at order $\mathcal{O}(\alpha^6)$. The description of the different programs used can be found in our work [37]. The upper plots provides the absolute value for each prediction while the lower plots presents all predictions normalised to MoCaNLO+Recola which is one of the programs that provides the full prediction. The band corresponds to scale-uncertainty band defined in Eq. (3.11) of [25]. The predictions are obtained in the fiducial region described in Sec. (3.1.2).

3.3 NLO VBS Approximation

VBS approximation at LO fails dramatically in the region $m_{jj} < 200$ GeV, $|\Delta y_{jj}| < 2$ as can be appreciated in Fig. (3.2) and Fig. (3.3). Therefore, we performed an inclusive[4] study at NLO QCD for the EW component, namely the order $\mathcal{O}(\alpha_s\alpha^6)$.

According to the different Montecarlo generator characteristics, we compare three different predictions at NLO QCD: the VBS approximation implemented in Bonsay ($|t|^2 + |u|^2$), the VBS approximation with the s-channel contributions from VBFNLO ($|s|^2 + |t|^2 + |u|^2$), and the full computation. The latter employs exact matrix elements, meaning that $t/u/s$ interferences, factorisable and non-factorisable QCD corrections, as well as EW corrections to the order $\mathcal{O}(\alpha_s\alpha^5)$ are included. The total cross sections, under the above-mentioned kinematic cuts are shown in Tab. (3.4). The $|t|^2 + |u|^2$ approximation for NLO QCD predictions is lower by about 6% than the full calculation. The inclusion of s-channel diagrams improves the approximate prediction, leading to an excess at the 3% level.

To better understand how the VBS approximation is performing and where the inclusion of s-channel contributions cures partly the NLO QCD predictions,

[4]Here, inclusive means that we apply only the following two selections: $m_{jj} > 200\,\text{GeV}$ and $|\Delta y_{jj}| > 2$.

Table 3.4: Cross sections at NLO QCD *i.e.* at order $\mathcal{O}(\alpha_s\alpha^6)$ for the full computation and two approximations. In addition to the cuts of Sec. (3.1.2), the VBS cuts include $m_{jj} > 200$ GeV and $|\Delta y_{jj}| > 2$. The uncertainties shown refer to the estimated statistical errors of the Montecarlo programs.

Prediction	σ_{tot} [fb]	δ[%]
full	1.733 ± 0.002	-
$\lvert t\rvert^2 + \lvert u\rvert^2$	1.6292 ± 0.0001	-6.0
$\lvert s\rvert^2 + \lvert t\rvert^2 + \lvert u\rvert^2$	1.7780 ± 0.0001	$+2.6$

we study the double-differential distribution in the variables m_{jj} and $|\Delta y_{jj}|$. The jet-pair kinematics is further investigated in Fig. (3.5), where we compute in each bin the ratios of the approximated cross sections over the full one $[\sigma(|t|^2+|u|^2)/\sigma(\text{full})$ and $\sigma(|s|^2+|t|^2+|u|^2)/\sigma(\text{full})]$. In the low invariant-mass and low rapidity-separation regions of the jet pair ($200\,\text{GeV} < m_{jj} < 500\,\text{GeV}$, $2 < |\Delta y_{jj}| < 2.5$), the VBS approximation fails significantly (by more than 40%). Including the s-channel contributions leads to a difference of less than 10% in this very region. However, in the region of large di-jet invariant mass and low rapidity separation of the jets, the $(|s|^2 + |t|^2 + |u|^2)$ approximation overestimates the full computation by more than 40%.[5] Again, this seems to support the fact that interferences and non-factorisable corrections can be non-negligible in this region.
In conclusion, both the loose minimum di-jet invariant-mass cut and the inclusion of QCD radiative corrections cause the s-channel contributions to be less suppressed than at LO, making their inclusion mandatory, in order to provide trustworthy predictions at NLO accuracy. In the inclusive region studied here, neglecting s-channel contributions, non-factorisable corrections and EW corrections can lead to discrepancies of up to 30% with respect to the full computation. In the typical VBS region where $m_{jj} \gtrsim 500\,\text{Gev}$ and $|\Delta y_{jj}| \gtrsim 2.5$, the VBS approximation shows a good agreement with the full computation as documented in the following.

3.3.1 NLO Fiducial Region

In Tab. (3.5), we present the cross sections as computed by the Montecarlo programs which implement a NLO-QCD accuracy, evaluated in the fiducial volume defined in Sec. (3.1.2). In contrast with Tab. (3.3), the NLO predictions differ visibly according to the different approximations implemented in the codes. The first observation is that the predictions featuring two versions of the VBS approximation (Bonsay and the Powheg-Box) are relatively close[6]. Bonsay uses a double-pole approximation for the virtual matrix ele-

[5] The bin in the top-left corner of the right-hand-side plot of Fig. (3.5) suffers from large uncertainty (30%) while the other errors are at the percent level.

[6] The VBFNLO prediction omitting s-channel contributions amounts to 1.3703 ± 0.0001 fb. This differs from the Powheg-Box prediction mainly due to the different choice of

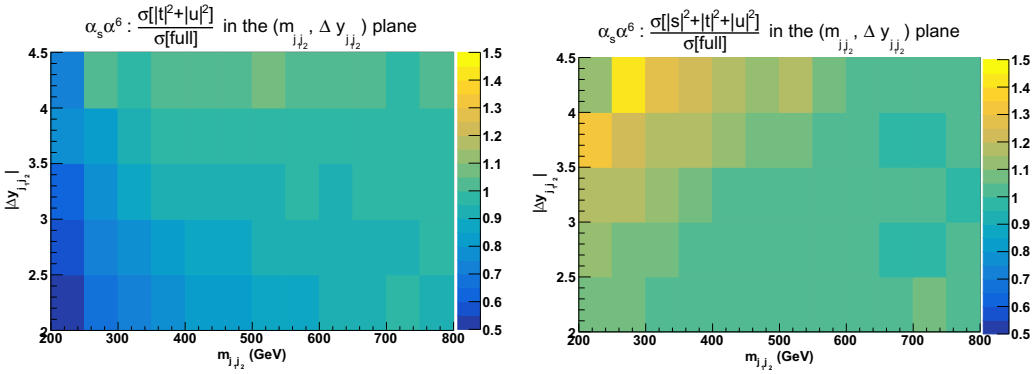


Figure 3.5: Ratios for double-differential distributions in the variables m_{jj} and $|\Delta y_{jj}|$ at NLO QCD *i.e.* at order $\mathcal{O}(\alpha_s\alpha^6)$ of the approximated squared amplitudes over the full matrix element. The approximated squared amplitudes are computed as $|\mathcal{A}|^2 \sim |t|^2 + |u|^2$ (left) and $|\mathcal{A}|^2 \sim |s|^2 + |t|^2 + |u|^2$ (right). In addition to the standard LO cuts, the VBS cuts take the values $m_{jj} > 200\ GeV$ and $|\Delta y_{jj}| > 2$.

Table 3.5: Cross sections at NLO accuracy and order $\mathcal{O}(\alpha_s\alpha^6)$. The predictions are obtained in the fiducial region described in Sec. (3.1.2). The uncertainties shown refer to estimated statistical errors of the Montecarlo integrations.

Code	σ[fb]
Bonsay	1.35039 $\pm$ 0.00006
Powheg-Box	1.3605 $\pm$ 0.0007
VBFNLO	1.3916 $\pm$ 0.0001
MG5_aMC	1.363 $\pm$ 0.004
MoCaNLO+Recola	1.378 $\pm$ 0.001

ment, and it is worth noticing that this approximation seems to be accurate at 1% level as compared to the Powheg-Box. This means that the double-pole approximation on the two W bosons used in Bonsay constitutes a good approximation of the VBS-approximated virtual corrections implemented in the Powheg-Box. Both predictions differ by about 2% with respect to the full computation (MoCaNLO+Recola).

Without losing generality, in the following we report just one differential distributions performed at NLO accuracy at the order $\mathcal{O}(\alpha_s\alpha^6)$. In the upper panel, the absolute predictions are shown while in the lower panel, the ratio with respect to the full predictions are displayed. The band corresponds to a seven-points variation of the factorisation and renormalisation scales (as defined in Eq. (3.11) of [25]). Indeed, in Fig. (3.6) we show the transverse momentum (left) and rapidity (right) of the hardest jet.
While MG5_aMC is very close to the full prediction for low transverse momentum, it departs from it at larger transverse momentum by about 10%.

scales used.

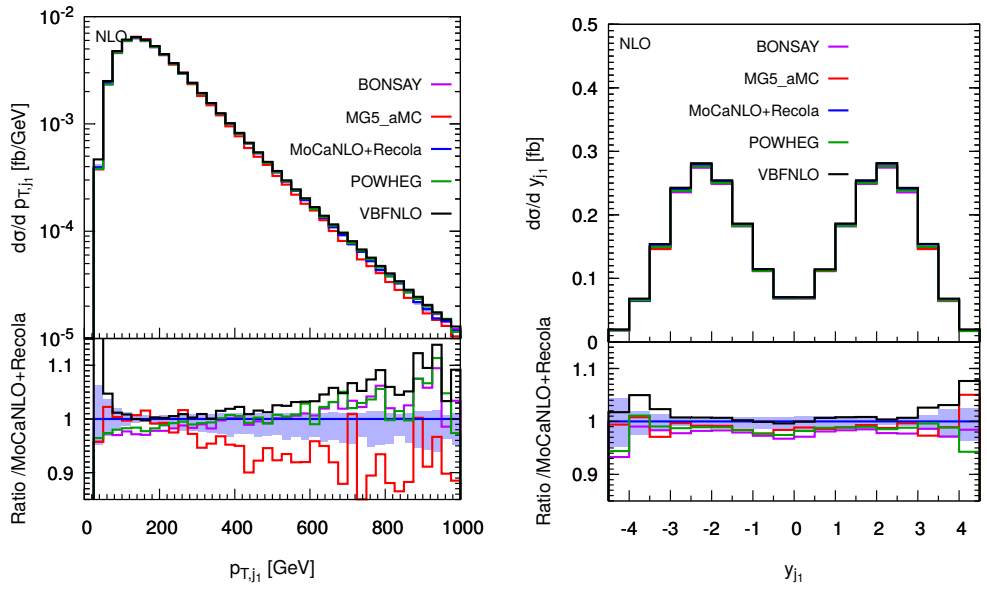


Figure 3.6: Differential distributions in the transverse momentum (left) and rapidity (right) of the hardest jet at NLO accuracy *i.e.* at order $\mathcal{O}(\alpha_s\alpha^6)$. The upper plots provide the absolute value for each prediction while the lower plots present all predictions normalised to MoCaNLO+Recola which is the full prediction. The band corresponds to a seven-point variation of the renormalisation and factorisation scales. The predictions are obtained in the fiducial region described in Sec. (3.1.2).

This is in contrast with the VBS-approximated predictions such as Bonsay, Powheg, and VBFNLO which are lower than the full computation at low transverse momentum and higher for larger transverse momentum. The difference at high transverse momentum between the latter predictions and the full computation can be attributed to EW Sudakov logarithms that become large in this phase-space region. However, for the transverse momentum of the second hardest jet (not shown here), the predictions from MG5_aMC are in good agreement with the other VBS-approximated predictions. In the end, the quality of the VBS approximations is good up to 10% in the fiducial region. These differences are larger than those at LO. The contributions from the s-channel amplitude can be sizeable especially at low invariant mass for the two tagging jets, explained by the fact that s-channel contributions are less suppressed at NLO.
Finally, the effect of EW corrections and non-factorisable contributions in the virtual corrections are usually small, but they can be relatively large (about 10%) for large transverse momentum of the hardest jet. These high-energy regions of the phase space are where EW Sudakov logarithms become large. Nonetheless these regions are rather suppressed and therefore these effects are hardly visible at the level of the cross section.

3.4 Parton Shower

As an additional step, for a meaningful comparison between theory and experiment, we need to dress the fixed-order parton-level cross-sections and observables which are all relative to matrix element calculations, with the non perturbative ingredients described in the Sec. (1.2.2), namely parton shower and PDFs.
At LO, the final-state partons are directly identified with streams of clustered hadrons (jet), observed in the detector. At NLO we have the so-called jet algorithm, so that when the radiation becomes soft or collinear, asymptotically we approach Born configuration ensuring the cancellation of the poles with their corresponding counterparts from the virtual expression - defining inclusive observables with hard and well-separated jets. An extra jet can only appear for hard, wide-angle emissions known as the infrared-safeness of observables, where the typical scale of the hard process is given by the factorization scale.
From the computational point of view, the approach to match the hard process with the generation of additional radiation is a unitary procedure which goes under the name of *parton shower* (PS). If hard events, computed at fixed order, as described above, are given as input to a PS code, we expect larger discrepancies among the codes than what we found at fixed order, as a consequence of the different matching schemes, parton showers employed, and of other details of the matching (such as the choice of the parton shower initial scale).
Among the codes capable of providing fixed-order results, presented before, MG5_aMC, the Powheg-Box, and VBFNLO can also provide results at (N)LO+PS accuracy. For VBFNLO matched to Herwig and the Powheg-Box, we restrict ourselves to show results only in the VBS approximation, *i*.e. the *s*-channel contributions are neglected here. Besides, also PHANTOM and WHIZARD are used to provide LO+PS results.
Without entering into details of the matching technique for each generator, we present the results of predictions matched to parton showers, as total rates within VBS cuts, in Tab. (3.6) and Tab. (3.7), at LO and NLO accuracy respectively.

Results are presented within the fiducial region defined by the cuts described in Sec. (3.1.2), applied after shower and hadronisation (this implies that jets are obtained by clustering stable hadrons, and not QCD partons). It follows that at the event-generation level, looser cuts (or even no cuts at all) must be employed in order not to bias the results. At NLO accuracy, for MG5_aMC + Pythia8 and VBFNLO+Herwig7-Dipole, we also quote theoretical uncertainties. For the former, we show both PDF and scale uncertainties[7], obtained via exact reweighting [71] by varying independently the renormalisation and factorisation scales by a factor of two around the central value of Eq. (3.4). Theory uncertainties on the total rates are very small, while

[7] A specific study of PDF uncertainties in VBS is presented in the last section of this chapter.

Table 3.6: Cross sections at LO+PS accuracy. The uncertainties shown refer to estimated statistical errors of the Montecarlo integrations.

Code	σ[fb]
MG5_aMC+Pythia8	1.352 ± 0.003
MG5_aMC+Herwig7	1.342 ± 0.003
MG5_aMC+Pythia8, Γ_{resc}	1.275 ± 0.003
MG5_aMC+Herwig7, Γ_{resc}	1.266 ± 0.003
PHANTOM+Pythia8	1.235 ± 0.001
PHANTOM+Herwig7	1.258 ± 0.001
VBFNLO+Herwig7-Dipole	1.3001 ± 0.0002
WHIZARD+Pythia8	1.229 ± 0.001

larger discrepancies appear for differential observables, which we discuss in the following.

Here we show a selection of these observables. The first one is the exclusive jet multiplicity, shown in Fig. (3.7). Looking at the LO+PS predictions, one can appreciate that the main effects are driven by the parton shower that is employed (Herwig7 [74] or Pythia8 [73]), with the clear tendency of producing more radiation for the latter, leading to higher jet multiplicities. Differences among tools that employ the same parton shower are typically smaller, and can be traced back to different values of the initial scale of the parton shower (the `scalup` entry of the Les Houches Event file [54]).The main effect of NLO corrections for this (rather inclusive) observable is to stabilise the predictions for the two-jet bin, where discrepancies among tools are reduced to about 10%. For the three-jet bin, which is described only at LO accuracy, differences among tools remain large, and are possibly related to the underlying approximation performed[8]. Despite the fact that the same parton shower is employed, the

[8]MG5_aMC is the only tool that makes use of the full matrix element for the real

Table 3.7: Cross sections at NLO+PS accuracy. The MG5_aMC results with Γ_{resc} are rescaled to account for the effect related to the W-boson width computed by MadSpin (see the text for details). For VBFNLO+Herwig7-Dipole, the three-point scale uncertainties are shown, while for MG5_aMC+Pythia8 the two displayed uncertainties are respectively the nine-point scale uncertainty and the PDF one. The uncertainties shown refer to estimated statistical errors of the Montecarlo integrations.

Code	σ[fb]
MG5_aMC+Pythia8	$1.491^{+1\%+2\%}_{-2\%-2\%}$ ± 0.004
MG5_aMC+Herwig7	1.427 ± 0.003
MG5_aMC+Pythia8, Γ_{resc}	1.407 ± 0.003
MG5_aMC+Herwig7, Γ_{resc}	1.346 ± 0.002
Powheg-Box+Pythia8	1.3642 ± 0.0004
VBFNLO+Herwig7-Dipole	$1.3389^{+0\%}_{-1\%}$ ± 0.0006
VBFNLO+Herwig7	1.3067 ± 0.0006

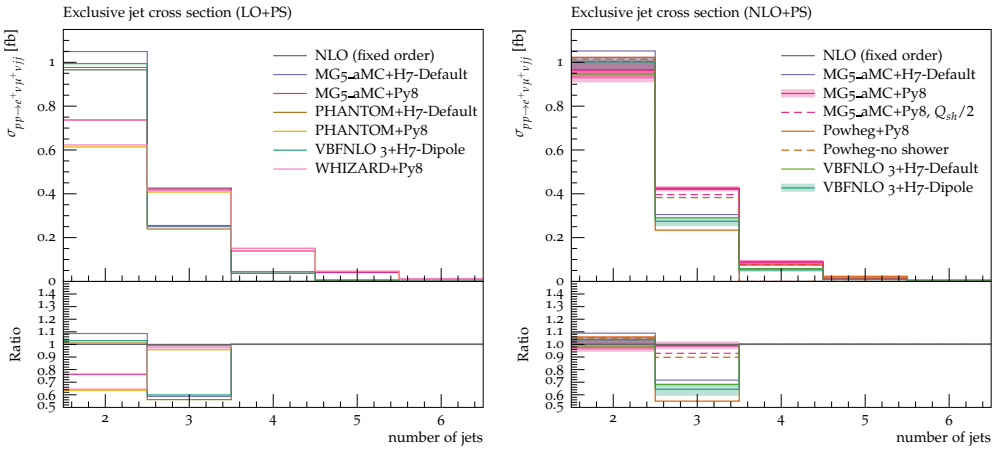


Figure 3.7: Differential distribution in the exclusive jet multiplicity from predictions matched to parton showers, at LO (left) or NLO (right) accuracy (upper plot), compared with the fixed-NLO result computed with VBFNLO (lower plot). At NLO+PS accuracy, for VBFNLO+Herwig7-Dipole, the three-point scale uncertainties are shown, while for MG5_aMC+Pythia8 the darker and lighter bands correspond respectively to the nine-point scale uncertainty and the scale and PDF uncertainties combined linearly[9]. The predictions are obtained in the fiducial region described in Sec. (3.1.2).

way emissions are treated is different among the tools.

The next observable is the rapidity difference between the two tagging jets, shown in Fig. (3.8). At LO+PS all predictions show the tendency to deplete the large-separation region with respect to the fixed-order prediction. At NLO+PS, when the extra radiation is described by the matrix element, such an effect is greatly reduced. A notable exception is the Powheg-Box prediction, which still shows a suppression at large separations. Since such a suppression is already there for the Powheg-LHE sample, it is very likely that it is driven by the way the first emission is generated. Finally, we would like to focus on the third jet, concluding the list of differential observables by showing in Fig. (3.9) the Zeppenfeld variable [79] defined as:

$$z_{e^+} = \frac{y_{e^+} - \frac{y_{j_1}+y_{j_2}}{2}}{|\Delta y_{jj}|} \,. \tag{3.9}$$

This variable is closely related to the third jet rapidity, and small (large) values of z correspond to central (peripheral) rapidities. In general, for observables which involve the third jet, one can clearly see a degradation of the agreement among the various tools, because of the poorer perturbative description of these observables.

radiation.

[9]The three-point scale uncertainties are obtained by considering correlated variations of the renormalisation, factorisation, and shower starting scale. The nine-point variations are obtained via exact reweighting [71] by varying independently the renormalisation and factorisation scales by a factor of two around the central value, Eq. (3.4).

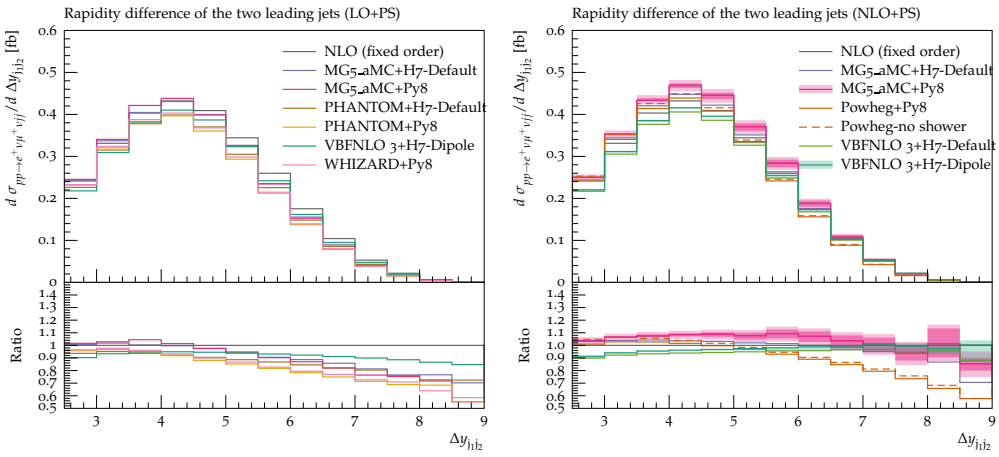


Figure 3.8: Differential distribution in the rapidity separation of the two tagging jets from predictions matched to parton showers, at LO (left) or NLO (right) accuracy (upper plot), compared with the fixed-NLO result computed with VBFNLO (lower plot). At NLO+PS accuracy, for VBFNLO+Herwig7-Dipole, the three-point scale uncertainties are shown, while for MG5_aMC+Pythia8 the darker and lighter bands correspond respectively to the nine-point scale uncertainty and the scale and PDF uncertainties combined linearly. The predictions are obtained in the fiducial region described in Sec. (3.1.2).

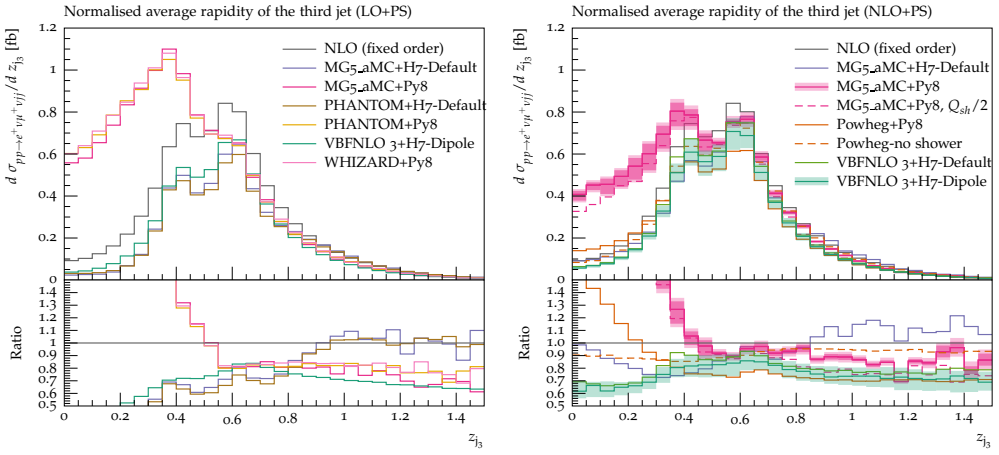


Figure 3.9: Differential distribution in the Zeppenfeld variable of the third-hardest jet from predictions matched to parton showers, at LO (left) or NLO (right) accuracy (upper plot), compared with the fixed-NLO result computed with VBFNLO (lower plot). At NLO+PS accuracy, for VBFNLO+Herwig7-Dipole, the three-point scale uncertainties are shown, while for MG5_aMC+Pythia8 the darker and lighter bands correspond respectively to the nine-point scale uncertainty and the scale and PDF uncertainties combined linearly. The predictions are obtained in the fiducial region described in Sec. (3.1.2).

The Zeppenfeld variable is a striking example: both at LO and NLO, the tendency of PYTHIA8 to generate more hard and central radiation, corresponding to low values of z, is clearly visible. Such an effect, which is related to the way PYTHIA8 deals with the recoil of the radiation in VBF(VBS)-type processes, can be mitigated by setting `SpaceShower:dipoleRecoil = on` in the PYTHIA8 input file[10]. A similar behaviour of PYTHIA8 has also been observed in the study of EW production of a Z boson in association with two jets (see the recent CMS measurement, [80] Figure 12), where the experimental data seem to prefer the description by HERWIG++.

In conclusion, the comparison of tools (eventually matched to parton shower) clearly shows the benefits of the inclusion of NLO corrections: for most observables described effectively at NLO accuracy, differences between tools are at (or below) the 10% level. Some exceptions exist, *e.g.* the rapidity separation of the two tagging jets, which, on the one hand, clearly suggest not to rely on a single tool/parton shower. On the other this makes worth investigating more in detail the way QCD radiation is generated, when fully-differential computations at NNLO will become available.

Key Results

In this Chapter, we have reviewed the state-of-art in terms of the Montecarlo programs available to simulate VBS events, and we performed comparisons among their predictions. We have precisely quantified the differences that arise on several physical observables, both inclusive and exclusive. Besides the study at fixed-order, we have also investigated the impact of parton showering. To this extend, several LO and NLO event generators which are able to perform matching to parton showers have been investigated, and various observables have been thoroughly compared. While, in general, on observables which are described at NLO accuracy, a reasonable agreement among the tools is obtained, larger differences appear for observables described at a lower accuracy, such as those that involve the third jet. We understood a peculiar behaviour in the central-rapidity region that triggered a deep and useful discussion with authors of PS generators regarding the different treatment of recoil.

[10] This requires a version of the code $\geq$ 8.230.

As discussed in the previous Chapter and in [25], the NLO EW corrections of order $\mathcal{O}(\alpha^7)$ are the dominant NLO contribution to the process $pp \to \mu^+\nu_\mu e^+\nu_e\, jj + \mathrm{X}$. It is thus highly desirable to combine them with NLO-QCD predictions matched with parton shower. Large EW corrections originate from the Sudakov logarithms which factorise, suggesting to combine them with QCD corrections in a multiplicative way.
For the typical fiducial region used by experimental collaborations for their measurements, the agreement between the approximations and the full calculation is satisfactory. At LO, the interference term can be included in the background component, at NLO the separation of EW and QCD components becomes more blurred. Therefore, a combined measurement including the EW, QCD, and interference contributions is desirable. The inclusion of NLO QCD corrections gives a better control of extra QCD radiation and reduces the ambiguities related to the matching details and/or the parton shower employed.
Very recently, in [81] the authors present a combined analysis of $pp \to \mu^+\nu_\mu e^-\bar{\nu}_e$ and $pp \to \mu^+\nu_\mu e^-\bar{\nu}_e\, j$ at next-to-leading order, including both QCD and electroweak corrections. In addition, they provide all-order prediction for $pp \to \mu^+\nu_\mu e^-\bar{\nu}_e$ + jets using merged parton-shower simulations that also include approximate EW effect.

In the final section of this Chapter we will introduce a dedicated study about PDF uncertainties for VBS.

3.5 PDF uncertainties

The uncertainty due to parton distribution functions (PDFs) is an important contribution to the theoretical uncertainty, as pointed out in Sec. (1.2.2). Theory uncertainties due to the terms used to write the cross sections of a hard process can be written as a calculable parton interaction convoluted with the parton densities, Eq. (1.37). The model uncertainties are related to the assumptions made in PDF extraction. Therefore, it is very important to evaluate such an uncertainty carefully as an essential ingredient for the prediction of observables at LHC. In the following, we discuss about some methods adopted to estimate their contribution to the process.

As a matter of fact, one of the uncertain parameters included in the PDF is the coupling constant α_S. The experimental value of this parameter is based on several experiments. The current PDG average is [82]:

$$\alpha_S(m_Z^2) = 0.1184 \pm 0.0007. \tag{3.10}$$

The α_S variation is typically not a part of the standard PDF variation. Instead, it is appended to certain PDF sets in the form of two additional members. The

upper and the lower limits of the confidence interval, taken into account in PDF sets are slightly more conservative than in Eq. (3.10): $\alpha_S(m_Z^2) = 0.118 \pm 0.0015$. The combined uncertainty can be calculated as:

$$\delta^{\text{pdf}+\alpha_S}\sigma = \sqrt{(\delta^{\text{pdf}}\sigma)^2 + (\delta^{\alpha_S}\sigma)^2}, \qquad \text{where} \qquad \delta^{\alpha_S} = \frac{1}{2}\left[\sigma(\alpha_S^+) - \sigma(\alpha_S^-)\right], \tag{3.11}$$

where the $\sigma(\alpha_S^+)$ and $\sigma(\alpha_S^-)$ are the cross-sections calculated with the values $\alpha_S^+(m_Z^2) = 0.1195$ and $\alpha_S^-(m_Z^2) = 0.1165$ and the central values of the rest of parameters.

PDF uncertainties for VBS

In the following we show some results regarding the study of the PDF uncertainty for the process $pp \to \mu^+\nu_\mu e^+\nu_e jj$ generated with PHANTOM [48]. All the details about the calculation can be found in App (B.4), in particular about the reweighting technique (Eq. (39)).
The kinematical cuts used in generation are:

- $p_T^\ell > 20$ GeV, $|\eta^\ell| < 2.5$, $p_T^{\text{miss}} > 40$ GeV;
- $p_T^j > 30$ GeV, $|\eta^j| < 4.5$, $|\Delta\eta_{jj}| > 2.5$, $m_{jj} > 500$ GeV;
- $\Delta R_{\ell\ell} > 0.3$, $\Delta R_{j\ell} > 0.3$;

There exist two leading representations of the uncertainties of PDF: Montecarlo [83] (MC) and Hessian [84]. The MC representation contains an ensemble of replicas, which are the instances of uncertain PDF parameters, sampled according to the Gaussian distribution, around its central values. To roughly estimate how much does the specific distribution resemble to the Gaussian, we performed the Shapiro-Wilk Gaussianity test [85]. The Shapiro-Wilk normality test on the cross-section distribution of our MC PDF set gives $W = 0.9891$ and p-value $p = 0.5928$. These values indicate that the distribution is Gaussian triggering a specific calculation of the PDF uncertainty for the MC PDF set (see Eq. (36) in App (B.4)).
As opposed to a MC PDF set, an Hessian PDF set is not composed by the random replicas, but each of its members coincides with one eigenvalue and eigenvector of the covariance matrix in the parameter space.
The `LHAPDF` library [86] offers a wide range of PDF sets from different groups. Because different PDF sets are based on different experimental data and use different assumptions, it is better to take into account more than just one PDF set as a part of the PDF uncertainty as well. For this purpose statistical combinations can be used.
The `LHAPDF` library currently contains one statistical combination: PDF4LHC15 [87]. This is a combination of CT14 [88], MMHT2014 [89], and NNPDF3.0 [90] sets. The only input PDF set into the PDF4LHC15, represented with the MC

replicas, is NNPDF3.0, while the CT14 and MMHT2014 are Hessian representations. PDF4LHC15 is distributed in three options, which use different reduction algorithms as indicated in the brackets:

- Montecarlo (CMC-PDFs),
- Hessian with 30 eigenvectors (META-PDFs),
- Hessian with 100 eigenvectors (MCH-PDFs).

The resulting uncertainties of the total cross-sections, calculated with the NLO representations of the PDF4LHC15 set, are listed in Table (3.8).

PDF type	Total LO xsection [fb]	PDF unc MC (a) [%]	PDF unc MC (b) [%]	PDF unc Hess [%]	α_S unc [%]	Combined unc [%]
PDF4LHC15_nlo_100	2.15271	-	-	1.76815	-	-
PDF4LHC15_nlo_30_pdfas	2.15298	-	-	1.6248	1.3131×10^{-2}	1.6249
PDF4LHC15_nlo_mc_pdfas	2.15333	1.8329	1.92104	-	1.1046×10^{-2}	1.8329

Table 3.8: PDF uncertainties of the total cross-section for the process $pp \to \mu^+\nu_\mu e^+\nu_e jj$. The PDF uncertainty is observed to be around 2%, while the α_S uncertainty is two orders of magnitude lower. The difference between the uncertainties calculated from the Eq. (36) and Eq. (37) is small, which is consistent with the Shapiro-Wilk test.

We note that the α_S uncertainty is significantly smaller than the PDF uncertainty. Uncertainty bands, calculated with the same PDF sets, are presented in the Fig. (3.10).

We can then conclude this part highlighting that different representations give consistent predictions. The PDF uncertainty is observed to be more or less uniform along the whole phase space and it amount to $\sim 2\%$. Uncertainties due to α_s variation are two orders of magnitude smaller.

The content of this section represents a reworking of a conference paper we published [91] as a summary of a mid-year work of the whole VBS Cost Action initiative. This part about PDF, in particular, represents our contribution. In the next Chapter, we will introduce a method originally developed within these thesis studies, to discriminate polarization. As we will discuss, it relies more on experimental than theoretical approaches.

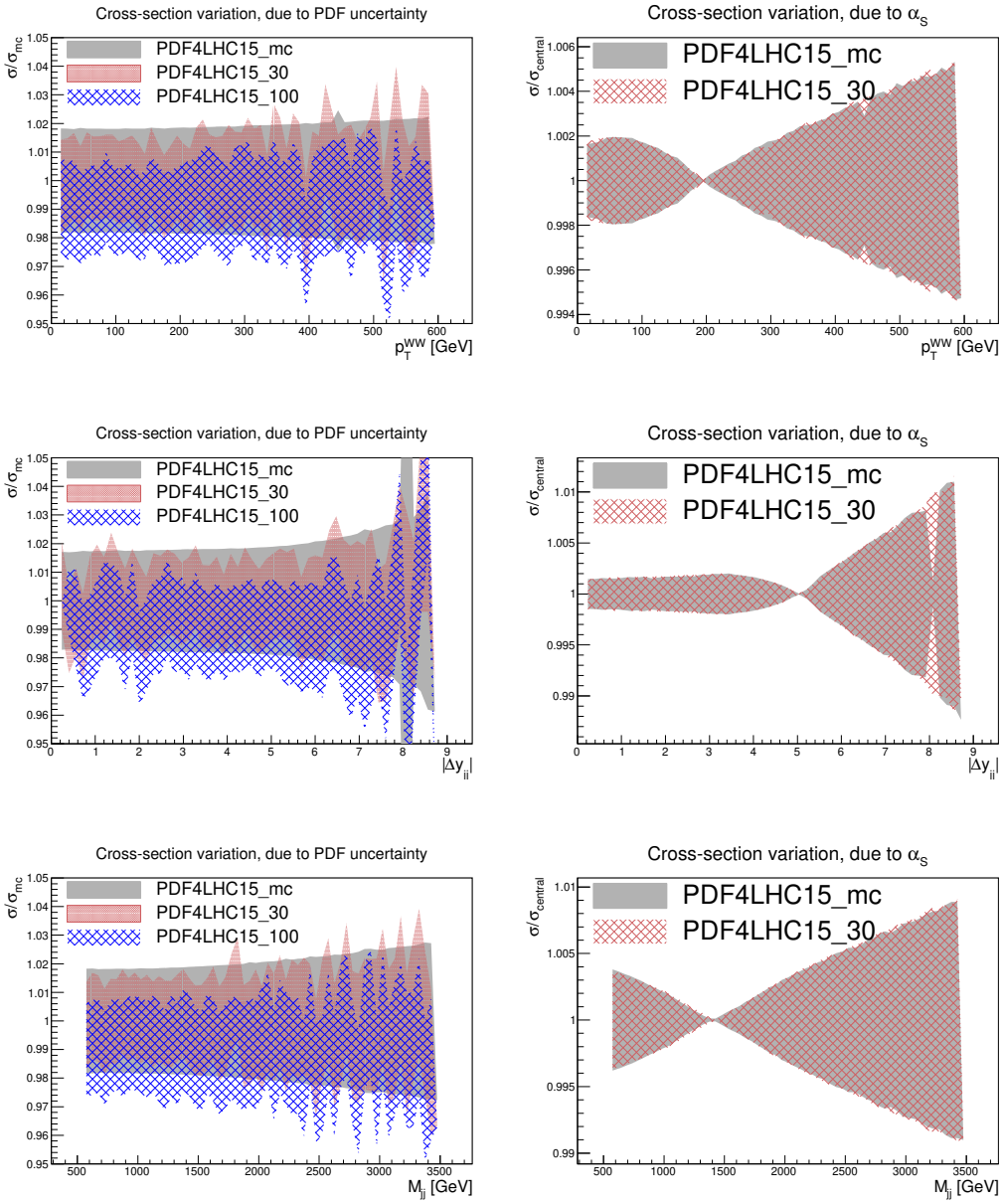


Figure 3.10: PDF uncertainty bands (left) and α_S uncertainty bands (right), along the variable p_{T}^{WW} (top), Δy_{jj} (center) and along the variable m_{jj}(bottom), calculated with the PDF sets `PDF4LHC15_nlo_mc`, `PDF4LHC15_nlo_30` and `PDF4LHC15_nlo_100`. The cross-sections without α_S variation are normalized to the central PDF of MC PDF set, while the PDFs from α_S variation are normalized to the central PDF of their own set. The spikes, in the shape of the uncertainty bands are the consequence of the statistical error. The differences among different PDF sets are consistent with the calculated uncertainties.

Chapter 4

W Boson Polarization Reconstruction

In this Chapter, we approach vector boson polarizations in the direction of disentangling the longitudinal component most meaningful from the physical point of view, from the others, here considered as background. In Chapter 2, we have discussed about the importance of measuring longitudinally polarized vector boson scattering and about how a theoretical method based only on $\cos\vartheta$ distribution could fail in presence of kinematic cuts. Here, we propose alternative solutions, and in this direction we investigate several approaches, from traditional ones up to advanced deep neural network structures. We compare their ability to reconstruct the W boson reference frame and to consequently measure the longitudinal fraction W_L in both semi-leptonic and di-leptonic (fully-leptonic) WW decay channels. We did not consider the fully-hadronic channel that is characterized by higher statistics but suffers from a huge background due to QCD multi–jet events and it presents more difficulties in trigger. The main results have been published in [92] and they represent the phenomenological part of this thesis work.
The last part of the Chapter is dedicated to the investigation of the possibility to directly discriminate boson polarized components both for the WW di-leptonic decay channels and for the ZZ channel.

This is not meant to be an analysis of VBS, this is a proof test that longitudinally polarized vector boson could be separated in VBS processes. The analysis of Run 2 data performed by both CMS and ATLAS collaborations led to a first observation of VBS in proton-proton collisions at the LHC [93], [94]. However, because of the limited statistics, no polarization study has been performed.
An attempt to study polarized VBS has been made, in this thesis, through Monte Carlo simulations. To this extent, we selected the pure VBS sample, neglecting the complexity of the signal extraction from background. Typical approaches employs several Monte Carlo event generators that are used to simulate the signal and background contributions, from generator level to full

detector simulation taking into accounts all different source of errors. The simulated samples are employed to optimize the event selection, to develop the multivariate discriminator like BDT [94] and [93], and to estimate the irreducible background yields while the reducible one is estimated via lepton identification and isolation in different control regions. The BDT distribution of the events in the selection is used to extract the significance of the EW signal via a maximum-likelihood fit. The expected distributions for the signal and the irreducible backgrounds are taken from the simulation while the reducible background is estimated from the data. From the referenced results ,the semileptonic channel sets better limits and possibly the usage of Deep networks that in general performs better than boosted decision trees can be adopted.
From a complete analysis point of view, this study should be considered after a detailed effort in extracting the VBS signal from the background discussed at the beginning of Chapter 2. In a complex environment such as the future High Luminosity LHC (HL-LHC), the high multiplicity of collisions occurring at the same bunch crossing would affect the sensitivity to polarized VBS events, for example when identifying the forward jets (VBS tagging jets), since isolation requirements on leptons and jets need to be applied [95].

Dataset Description

The dataset used across this Chapter has been generated according to the following prescriptions:

- Montecarlo event generator: PHANTOM v.1.6
- Parton distribution function: `NNPDF30_nnlo_as_0118`;
- Calculation type: α_{ew}^6 at 13 TeV CM energy;
- Scale choice: (invariant mass of the 2 central jets and of 2 leptons)/$\sqrt{2}$;
- Parton shower generator: PYTHIA;
- Fast detector simulation: DELPHES.

We will provide additional details about kinematic cuts applied in the generation of the events according to the physical process. A more complete description of all the code functionalities can be found in Sec. (3.1.1). In the adopted version of PHANTOM generator the only option for the reference frame in which to compute polarization is the laboratory. For VBS events the polarization fractions calculated in the laboratory frame and those calculated in the partonic centre–of–mass frame are very similar [45]. Nevertheless, in VBS processes at the LHC, the most convenient frames are usually the partonic center-of-mass frame or the center-of-mass frame[1].

[1]The possibility of choosing the reference system for the polarization calculation was only introduced in the latest PHANTOM v.1.7 version.

4.1 Traditional Kinematic Approach

Due to the short lifetime of vector bosons, a direct measurement of their polarization is unfeasible. $W^{\pm}W^{\pm}$ is non-resonant as long as there are no bosonic isospin triplets, and hence doubly charged bosons, in nature. They decay before reaching the detector, therefore only a study of their decay products can be feasible.
The $W^{\pm}W^{\pm}$ final state, with its ± 2 total electric charge carries unique features that make it of particular interest at the LHC. We have already seen that same-sign WW scattering is the only process for which the cross-talk amplitudes, $W_T W_X \to W_L W_L$ and $W_L W_L \to W_T W_X$, are completely negligible, mostly due to lack of any s-channel graphs that contribute to the process. This has also other consequences: contrary to other diboson states with two accompanying jets, production of the $jjW^{\pm}W^{\pm}$ state in the lowest order is dominated by only one physical mechanism at the quark level, namely a quark-quark interaction associated with a $W^{\pm}$ emission from each colliding quark.
In Sec. (2.6.2) we described a possible strategy in extracting polarization fractions from the angular distribution of leptons in the W rest frame. In order to reconstruct the W boson rest frame, one needs to reconstruct the full event kinematics, including the momenta of the neutrinos resulting from W boson decays.
In this Section, we will discuss how to possibly reconstruct the W boson rest frame using kinematic cuts of different observables. This procedure is what we define *traditional* approach. The result of this approach will be compared in the remaining part of the Chapter with the one obtained by applying several machine learning techniques.

4.1.1 Semi-leptonic VBS channel

The VBS Semi-leptonic channel as been defined in Sec. (2.1). With only one neutrino present, transverse neutrino components are measurable, i.e. they can be evaluated from the measured missing transverse momentum (MET) in an event, but longitudinal one is not. Indeed, with a single W boson decaying leptonically we can impose four-momentum conservation of the W boson decay products in the ultra-relativistic limit. Solving the constraint for the longitudinal components of the neutrino ($p_{\nu L}$), a second-order equation can be easily obtained:

$$\begin{aligned} &\underbrace{\left(p_{\ell L}^2 - E_\ell^2\right)}_{a} p_{\nu L}^2 \\ &+ \underbrace{\left(m_{\mathrm{W}}^2 p_{\ell L} + 2p_{\ell L}(\vec{p}_{\ell T} \cdot \vec{p}_{\nu T})\right)}_{b} p_{\nu L} \\ &+ \underbrace{\frac{m_{\mathrm{W}}^4}{4} + (\vec{p}_{\ell T} \cdot \vec{p}_{\nu T})^2 + m_{\mathrm{W}}^2(\vec{p}_{\ell T} \cdot \vec{p}_{\nu T}) - E_\ell^2 \vec{p}_{\nu T}^{\,2}}_{c} = 0 \end{aligned} \tag{4.1}$$

where m_{W} is the W boson invariant mass, $p_{\ell L}$, $\vec{p}_{\ell T}$ are respectively the longitudinal and transverse components of the lepton momentum and E_ℓ represents its energy, while $\vec{p}_{\nu T}$ represents the transverse component of the momentum of the neutrino.
Eq. (4.1) defines the unknown $p_{\nu L}$ as the zeros of a second-order polynomial. The problem of *negative* discriminant should be considered. Some *ad-hoc* solutions have been adopted by the experimental analysis groups, such as setting the discriminant to zero or recalculating the discriminant by applying a W transverse mass constraint. Speculations on the meaning of the imaginary solutions are also considered (see Fig. (4.4)).
In the core of present study, without any loss of completeness of the problem, we decided to ignore the negative discriminant cases, and we focus instead only on the ambiguity of sign $(+/-)$ in the two solutions for the positive discriminant ($\Delta = b^2 - 4ac > 0$) case. Relativistic kinematics gives a priori no physical reason to prefer one solution instead of the other.
In the attempt of selecting the correct solution for Eq. (4.1), sign ambiguity in longitudinal neutrino momentum can be partially resolved by adopting selection criteria, that tend to isolate the appropriate sign choice in the equation. The decision criteria impose requirements for the correct solution to pass a given kinematic cut. If both solutions pass or fail this kinematic cut, a random solution is chosen. Selection criteria considered are the following (a and b represent the parameters of the quadratic equation given by Eq. (4.1)):

- Random Selection: sign is chosen randomly (this is the worst-case scenario);
- Selection 1: the correct solution is required to have the absolute value of the scalar product of the reconstructed neutrino three-momentum with the reconstructed W three-momentum smaller than 5000 GeV^2- random solution is chosen if both solutions pass/fail this criterion;
- Selection 2: the correct solution is required to have the absolute value $|p_{\nu L}|$ smaller than 50 GeV- random solution is chosen if both solutions pass/fail this criterion;
- Selection 3: the correct solution is required to have the absolute value of the scalar product of the reconstructed neutrino three-momentum with the reconstructed W three-momentum, multiplied by a/b, smaller than 25 GeV- random solution is chosen if both solutions pass/fail this criterion;
- Selection 4: the correct solution is required to lie above the parabola minimum, i.e. $p_{\nu L} > -b/(2a)$ or, alternatively, in a normalized way: $p_{\nu L} \cdot a/b > -0.5$- random solution is chosen if both solutions pass/fail this criterion.

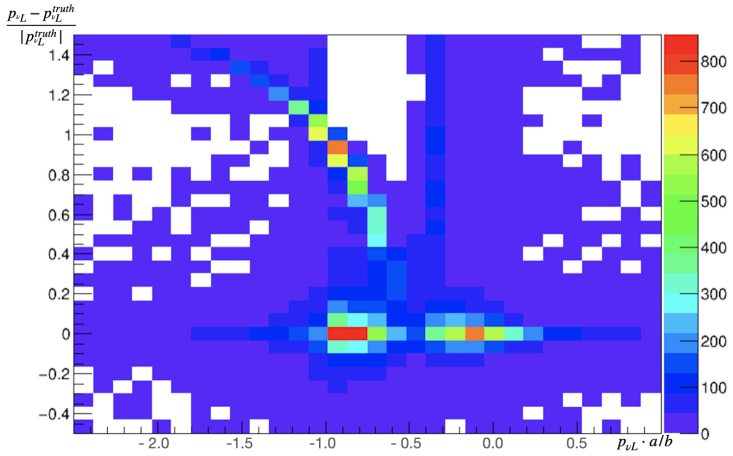


Figure 4.1: Relative deviation of the longitudinal neutrino momentum $p_{\nu L}$ from the true (generator-level) value $p_{\nu L}^{\text{truth}}$ is shown as a function of the discriminating variable defined in *Selection 4* above.

The application of each selection criterion results in a reduction of events statistics. This mainly leads to two possibilities: either discard the events permanently and reduce the statistics further or to adopt a common choice. Since this dataset is balanced with respect to the choice of the two solutions, it was decided to adopt a random choice with respect to the discarded solutions. This allowed us to keep the sample unbiased by maintaining the statistics comparable between the various criteria. A different choice, dictated by some kinematic parameter would have been combined in a non-obvious way with respect to the selected criterion. We determined thresholds for kinematic selection by visualizing, in a 2D plot, the number of events as a function of relative error and specific discriminating variable defined previously. For instance, we can selects positive or negative solution by means of a visual analysis, as depicted in Fig. (4.1), which represents Selection 4 criterion.

The effect of the different selection criteria on the reconstruction of the longitudinal neutrino boson momentum $p_{\nu L}$ from truth level quantities can be appreciated in Fig. (4.2), which shows the relative deviation of this reconstructed quantity from the true (generator-level) value $p_{\nu L}^{\text{truth}}$. A qualitative analysis of this plot demonstrates that all selections enhance the peak at one, whereby Selection 1 coincides with (lies under) Selection 3. As a distinctive case, the Selection 4 achieves a higher peak at one by substantially reducing the high-value tails, while the other Selections reduce the contributions which are less than one. In general, the widths are limited from below by the resolution on the reconstructed W momentum.
Comparisons of truth and reconstructed lepton angular distributions from generator-level measurable quantities are shown in Fig. (4.3). Here, both the longitudinal W polarization (left) and the transverse polarization (right) are shown.
One may ask where the wrong solutions are located and what is their nature.

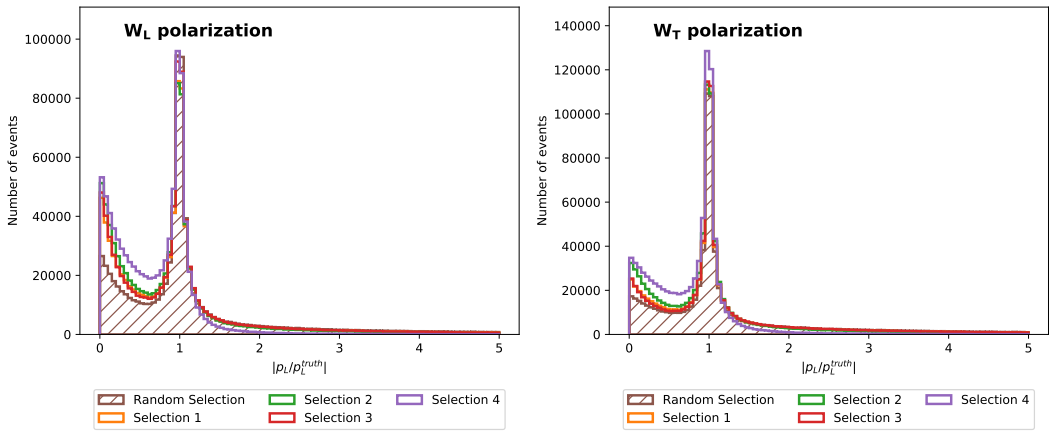


Figure 4.2: Reconstruction of longitudinal neutrino momentum using different Selection criteria described in the text. The random Selection is shown as the worst-case scenario. The ratio between the longitudinal neutrino momentum $p_{\nu L}$ and the true (generator-level) value $p_{\nu L}^{\text{truth}}$ is shown. The left plot shows results obtained assuming the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario.

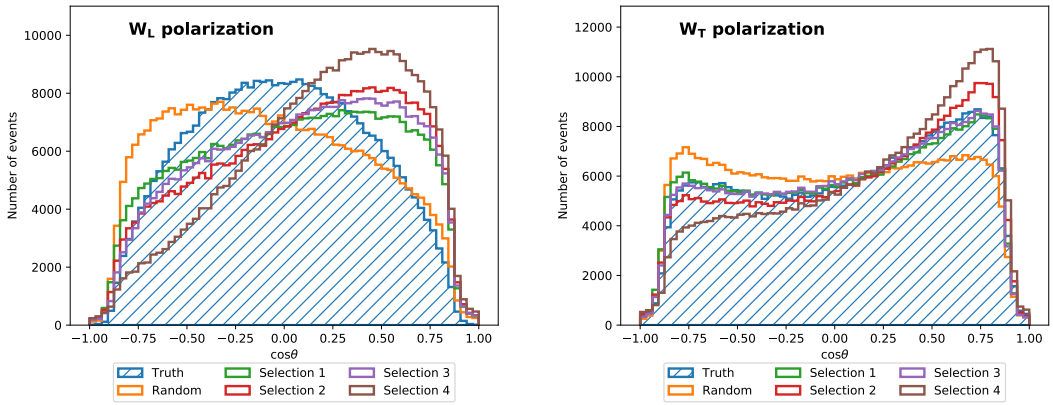


Figure 4.3: Shape of the reconstructed, normalized lepton angular distribution in the reconstructed W boson rest frame. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$. The left plot shows results obtained assuming the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario. Selection criteria 1–4 and random solutions, have been studied as strategies for choosing the correct solution, as described in the text.

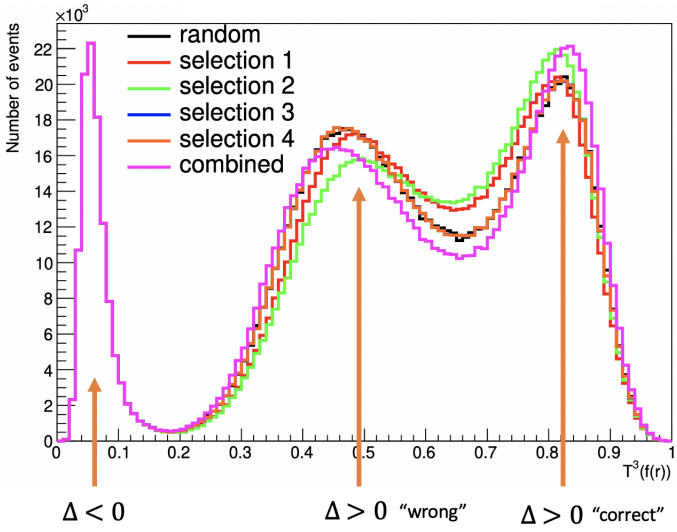


Figure 4.4: Phase space distribution of neutrino relative error after the transformation defined in Eq. (4.2).

For this reason, we can try to visualize the whole range of relative error by introducing a transformation $f(r)$ with the following properties:

- for wrong solutions of Eq. (4.1) $f(r) \Longrightarrow 0$;
- for correct solutions of Eq. (4.1): $f(r) \Longrightarrow 1$.

Under this assumption, the function assumes the form:

$$f(r) = \frac{2}{\frac{1}{r} + r}; \qquad \text{where } r = \frac{p_L^{reco}}{p_L^{truth}};$$

However, in order to enhance separation between selection criteria, we better implement a phase space transformation Φ, defined as $T^3(f(r))$ where:

$$T(x) = \arcsin(2x - 1)/\pi + \frac{1}{2}; \qquad \text{with} \qquad x = f(r). \tag{4.2}$$

In order to solve the equation mathematically and analyse all possible cases, we need to define some boundary conditions. The mass of the W is off-shell and therefore defined, event by event, by the truth MC generation. If the discriminant is negative, we change the given mass value to the derived *transverse* one, recalculating the solutions of the equation. The action of the previous transformation can be appreciated in Fig. (4.4) where the first peak contains events with negative discriminant, the second peak contains wrong reconstructed events with positive discriminant and finally the third peak groups the correctly reconstructed events.

An alternative metric to quantitatively asses the performance of the selection criteria is the fraction of *correct* solution choices, where *correct* denotes a solution giving $p_{\nu L}$ with a smaller absolute error with respect to the true

Table 4.1: Summary of the fractions of correct choices in the semi-leptonic channel for different selection criteria, applied on samples containing a purely longitudinally polarized (W_{L}) leptonically decaying W boson, a purely transversely polarized (W_{T}) boson, an unpolarized W boson generated in the OSP PHANTOM framework (Un.OSP) and the full computation with an unpolarized W boson and with non-resonant production modes included (Full comp.) is shown.

Type	W_{L}	W_{T}	Un. OSP	Full comp.
Selection 1	0.503	0.544	0.527	0.551
Selection 2	0.518	0.580	0.554	0.576
Selection 3	0.528	0.553	0.539	0.564
Selection 4	0.565	0.645	0.607	0.638

value. The results for different polarizations of the leptonically decaying W boson (W_{L}, W_{T}), unpolarized W boson generated in the OSP framework[2], and full computation (unpolarized W boson, with non-resonant production modes included) are reported in Table (4.1).

Key Results

In the semi-leptonic channel, the reconstruction of the W reference frame can be done up to a neutrino sign ambiguity. The use of selection criteria presented here to select the correct solution is not fully effective and could be improved by using machine learning techniques, as investigated in Sec. (4.2).

4.1.2 Fully-leptonic VBS channel

The VBS Fully-leptonic channel, that will be studied in this Section, has been defined in Sec. (2.1). The strategies described for the semi-leptonic channel can be generalized to the fully-leptonic polarized VBS.

Information on WW polarizations stays encoded in the kinematics of the two outgoing quarks, unless they are disturbed by a subsequent quark or gluon interaction. If only we knew the energies of the colliding quarks, appropriate cuts on the angles and transverse momenta of the two tagging jets would increase the probability of choosing a $W_L W_L$ state - regardless of its final kinematics and the rest of the process. Of course, these kind of detailed information is not accessible experimentally, in fully-leptonic decays, where both neutrinos contribute to the measured MET. The kinematic reconstruction is consequently even more challenging that for the semi-leptonic case. There

[2]OSP refers to a method called On Shell Projection (OSP), an approximation implemented by PHANTOM to compute the amplitude of resonant contributions projecting on shell (in the numerator) the four momenta of the decay particles (see D.3 for details).

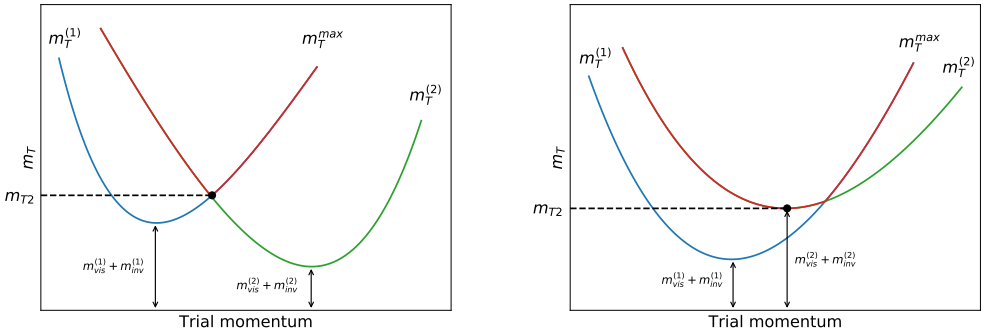


Figure 4.5: The balanced and unbalanced solution classes of the MAOS minimization problem. The two transverse masses $m_{\mathrm{T}}^{(1,2)}$ as a function of the trial momenta are shown in blue and green and the resulting functional $m_{\mathrm{T}}^{\mathrm{max}}$ is shown in red. In the balanced case (left) the solution $m_{\mathrm{T}2}$ is in the intersection of the two m_{T} curves and in the unbalanced case (right) the $m_{\mathrm{T}2}$ is in one of the minima of the m_{T} curves. The $m_{\mathrm{vis}}^{(1,2)}$ and $m_{\mathrm{inv}}^{(1,2)}$, represent the masses of the visible and invisible final state particles. In the case studied in this paper, where the masses of the final state particles are negligible in comparison to the mass of the decaying heavy particle, the $m_{\mathrm{T}}^{(1)}$ and $m_{\mathrm{T}}^{(2)}$ have their minima values approximately at zero and therefore only the balanced solution topology is possible.

are eight unknown parameters[3] and only six equations defining kinematic constraints, 2 for each neutrino, as specifically shown in Eq. (4.3):

$$\boxed{\begin{aligned} \vec{p}_T^{\,\nu_\mu} + \vec{p}_T^{\,\nu_e} &= \vec{p}_T^{\,miss} \\ (p_l + p_\nu)^2 &= m_W^2 \\ p_\nu^2 &= 0 \end{aligned}} \tag{4.3}$$

Several algorithms have been developed to fully reconstruct the kinematics of a process with two invisible particles in the final state. The one considered here is the `MT2`-Assisted On-Shell (MAOS) [96], an approach that exploits the fact that most of the processes involving an exchange of the W boson (or another heavy decaying particle) prefer the phase space region with low heavy particle transverse mass. This algorithm has been successfully used in SUSY searches involving vector bosons, and also in $H \to WW$ reconstruction [97, 96, 98].

In order to simplify the notation, let us restrict ourselves to a VBS process where one W decays to a muon and the other to an electron. The MAOS algorithm introduces a functional expression of trial neutrino transverse momenta $\vec{p}_1$ and $\vec{p}_2$ as:

$$m_{\mathrm{T}}^{\mathrm{max}}(\vec{p}_1, \vec{p}_2) = \max[m_{\mathrm{T}}^{(1)}(\vec{p}_1), m_{\mathrm{T}}^{(2)}(\vec{p}_2)], \tag{4.4}$$

with:

$$\begin{aligned} m_{\mathrm{T}}^{(1)}(\vec{p}_1) &= 2(|\vec{p}_{\mathrm{T}}^{\,\mu}||\vec{p}_1| - \vec{p}_{\mathrm{T}}^{\,\mu} \cdot \vec{p}_1), \\ m_{\mathrm{T}}^{(2)}(\vec{p}_1) &= 2(|\vec{p}_{\mathrm{T}}^{\,e}||\vec{p}_2| - \vec{p}_{\mathrm{T}}^{\,e} \cdot \vec{p}_2), \end{aligned} \tag{4.5}$$

[3]The eight unknown parameters are: p_X, p_Y, p_Z, E, for each neutrino.

where $\vec{p}_{\mathrm{T}}^{\,\mu}$ and $\vec{p}_{\mathrm{T}}^{\,e}$ represent muon and electron transverse momentum respectively. The minimum of this functional defines the m_{T2} variable and represents a compromise for minimization of both W transverse masses at the same time:

$$m_{\mathrm{T2}} \equiv \min_{\vec{p}_1+\vec{p}_2=\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}} m_{\mathrm{T}}^{\mathrm{max}}(\vec{p}_1,\vec{p}_2) = m_{\mathrm{T}}^{\mathrm{max}}|_{\vec{p}_{\mathrm{T}}^{\,\nu_\mu},\vec{p}_{\mathrm{T}}^{\,\nu_e}}\,, \tag{4.6}$$

with the solutions $\vec{p}_{\mathrm{T}}^{\,\nu_e}$ and $\vec{p}_{\mathrm{T}}^{\,\nu_\mu}$ giving the estimates for the neutrino transverse momenta.

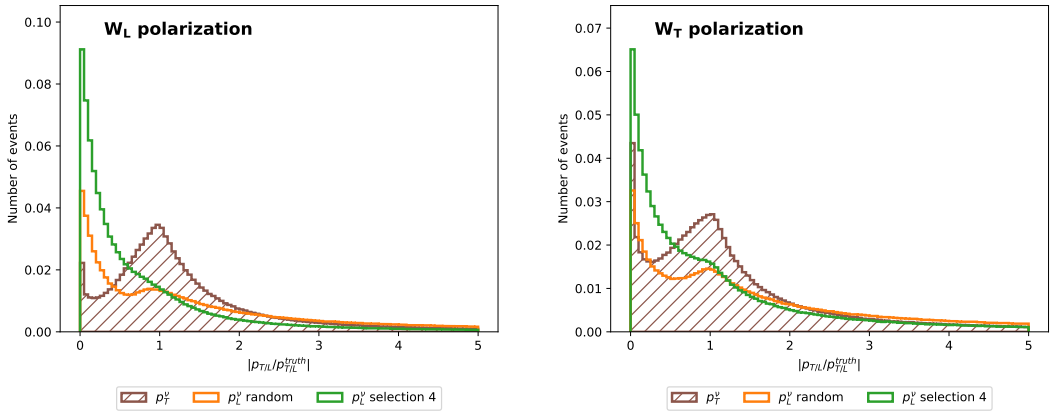


Figure 4.6: Reconstruction of the neutrino momenta in the fully-leptonic channel at the truth (generator) level using the MAOS approach. The ratio between the reconstructed and true neutrino momentum is shown for the longitudinal and transverse components. The left plot shows results obtained assuming the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario. While the transverse momenta are reconstructed using the MAOS procedure, the W mass constraint from the semi-leptonic case is adopted for reconstruction of both longitudinal neutrino momenta. The random choice and Selection 4, as described in Sec. (4.1.1), are applied as strategies to resolve solution ambiguities.

In general, there are two classes of solutions to the minimization problem of Eq. (4.6) - *balanced* and *unbalanced*. The difference between the two cases is diagrammatically shown in Fig. (4.5). When the masses of the final state particles are negligible with respect to the W mass, the unbalanced case is not feasible and the minimum has to lie on the intersection of the $m_{\mathrm{T}}^{(1)}(\vec{p}_1)$ and $m_{\mathrm{T}}^{(2)}(\vec{p}_{\mathrm{T}}-\vec{p}_1)$. In case of the leptonic W decay we find ourselves in this regime. The intersection curve can be expressed analytically, while the m_{T2} solution has to be estimated numerically along this curve. The derivation of an analytic solution for the intersection curve can be found in App. (D.1).
The next kinematic reconstruction steps of the longitudinal neutrino momenta follow the same approach adopted for the semi-leptonic case and described in Sec. (4.1.1), for each W boson separately. The performance is evaluated on purely longitudinally (W_{L}) and purely transversely (W_{T}) polarized samples with both W bosons decaying leptonically. Neutrino distributions are shown inclusively for both Ws in the event. In Fig. (4.6) one can observe the success in the reconstruction of the neutrino momenta using the MAOS approach at

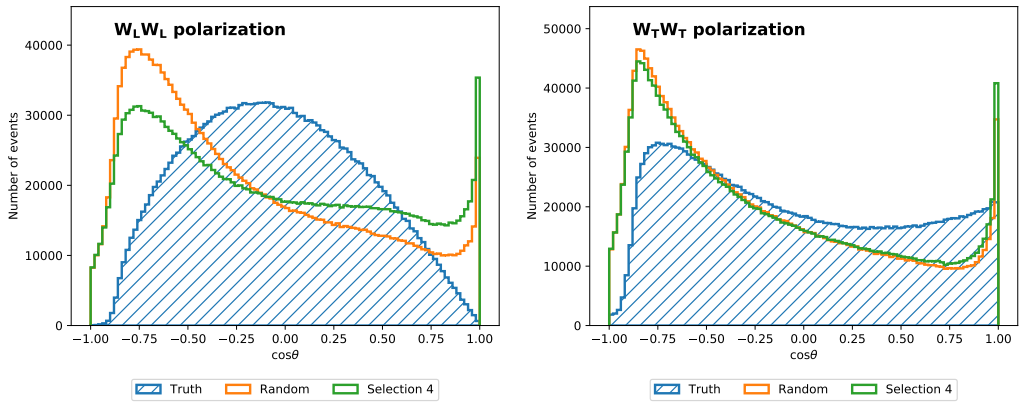


Figure 4.7: MAOS reconstructed inclusive lepton (electron and muon) angular distribution in the W boson rest frame, compared to true parton level distribution, in the fully-leptonic channel. The left plot shows the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario. Longitudinal neutrino momenta are calculated from the W mass constraint, using random solution and selection 4 strategies, described in Sect. (4.1.1). Truth distributions use the generator-level (true) $\cos\vartheta$.

the truth (generator) level. Ratios between the reconstructed longitudinal and transverse neutrino momentum components $p_{\nu L,T}$ and true (generator-level) values $p^{\mathrm{truth}}_{\nu L,T}$ are shown inclusively for both neutrinos, since a complete symmetry is expected. A comparison of the true and reconstructed lepton angular distributions for the longitudinally and transversely polarized samples is shown in Fig. (4.7). The complexity of this channel compared to the semileptonic one lies in the need to reconstruct two components: transverse and longitudinal. We adopted a strategy for the transverse part (MAOS) and tested different selection criteria for the longitudinal part. The combination of these methods shows that in this case the reconstruction of the two components of the polarization is far from the truth.

For sake of completeness, we point out that the whole analysis described above has been performed also including detector effects. As explained in Sec. (3.1.1), to include these effects, we processed the parton level events generated with PHANTOM with a parton shower tool (PYTHIA) and then we closes the production toolchain by simulating detector contributions by means of a fast simulation program, called DELPHES. This can be helpful to understand how much the smearing due to the detector can affect particles and therefore polarization reconstruction.

We evaluated also the impact of different detector smearings, by applying it to muon and to MET component in different combinations: gradually smearing only the muon component then the MET component and eventually both.

According to our result, it is the MAOS approach that intrinsically brings the largest uncertainty in the reconstruction of observable variables. The effect of detector smearing, especially in the fully-leptonic channel, is sub-leading and overall the main source of uncertainty comes from the reconstruction algorithm.

Key Results

While the MAOS framework considered here performs relatively well compared to other (non-machine-learning) approaches, it turns out that W polarization separation in the lepton angular distributions (evaluated in the reconstructed W reference frame) is still quite poor, obscuring the differences between transverse and longitudinal polarizations. Consequently, as in the semi-leptonic case, exploiting a machine learning approach seems worthy of investigation.

4.2 Deep-Learning Techniques

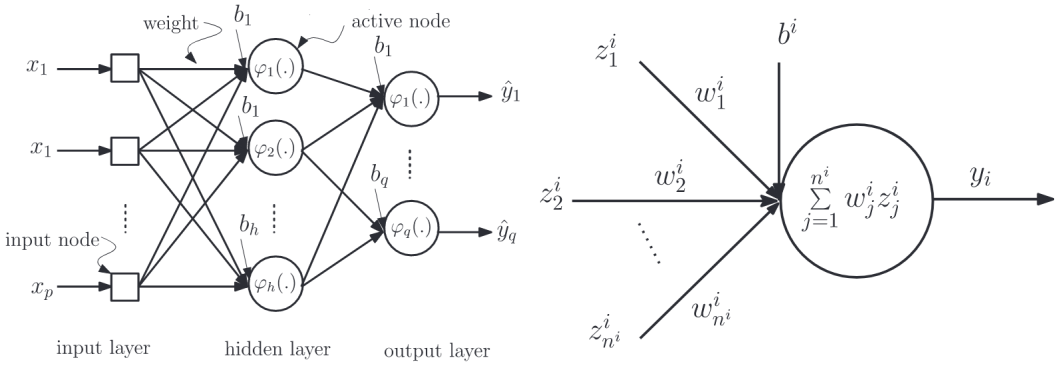
In this Section we will investigate a deep-learning multivariate technique to reconstruct the W boson rest frame(s) and consequently to estimate the vector boson polarization fractions.
Particular attention has been dedicated to the determination of the optimal network architecture and configuration of the neural network, in terms of performance and reliability. We will start giving a brief and schematic overview of this topic from a general point of view. The remaining part of the Chapter is dedicated to the application to our specific use case: Vector Boson Scattering.

4.2.1 Human and Machine Learning

From mid 1980s something radically changed in the applied statistics community with the introduction of Neural Networks. They marked a shift of predictive modelling towards computer science and machine learning. They are still used every day to make critical decisions in medical diagnosis, stock trading, energy load forecasting, and more.
Neural networks were reincarnated around 2010 with "deep learning" as a flashier name, largely a result of much faster and larger computing systems, plus a few new ideas. They have been shown to be particularly successful in the difficult task of classifying natural images, using what is known as a convolutional architecture.

Machine learning is a subfield of artificial intelligence. Deep learning is a subfield of machine learning, and neural networks make up the backbone of deep learning algorithms. In fact, it is the number of node layers, or depth, of neural networks that distinguishes a single neural network from a deep learning algorithm, which must have more than three. Abundant literature (for instance [99], [100] and [101]) exists about machine learning (ML) and deep learning (DL). The high energy physics community contributed to this effort as well with studies, including the specific case of VBS addressed here, see for instance ([102] and [103]).

Machine learning uses two types of techniques: *supervised learning*, which trains a model on known input and output data so that it can predict future



(a) Three-layer feed forward Neural Network. (b) Node of the Network.

Figure 4.8: Three-layer feed-forward neural network (a), where input layer has p input nodes, hidden layer hash activation functions, and output layer has q nodes [104].

outputs, and *unsupervised learning*, which finds hidden patterns or intrinsic structures in input data.
The first one builds a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data. The second one finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses. Unsupervised learning finds hidden patterns or intrinsic structures in data and it draws inferences from datasets consisting of input data without labeled responses.
Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data. Common algorithms for performing clustering include k-means and k-medoids, hierarchical clustering, Gaussian mixture models, and hidden Markov models. *Supervised learning* uses *classification* and *regression* techniques to develop predictive models. **Classification techniques** predict discrete responses and classify input data into categories. Common algorithms for performing classification include *support vector machine (SVM)*, boosted and bagged decision trees, k-nearest neighbour, Naive Bayes, discriminant analysis, logistic regression, and *neural networks*. **Regression techniques** predict continuous responses and they are used when dealing with a data range or if the nature of your response is a real number. Common regression algorithms include linear model, nonlinear model, regularization, stepwise regression, boosted and bagged decision trees, *neural networks*.

Our problem falls under the specifics of a supervised learning, hence neural network is our choice: a tool which is suited to analyse large sets of high-dimensional data in order to reproduce one or more outputs. A neural network (NN) is a highly parametrized model, inspired by the architecture of the human brain, that was widely promoted as a universal approximator - a machine

that with enough data could learn any smooth predictive relationship. Its basic units are the artificial neurons (or *nodes*) which receive many inputs and produce a single output. Neurons are organised in a sequence of *layers* and the output of each neuron is passed to every neuron in the successive layer.
Fig. (4.8a) shows a specific topology: a *feed-forward* neural network (FNN), the computational models adopted in this work. The FNNs have a specific structural configuration (architecture) in which the nodes at a layer have forward connections from the nodes at its previous layer in a fully connected way with many independent weights. In particular we have predictors or inputs[4] $\vec{x} = \langle x_1, x_2, \ldots, x_p \rangle$ and n-dimensional real-valued weight vector $\vec{w} = \langle w_1, w_2, \ldots, w_n \rangle$. We have then intermediate layers called *hidden layers* and an output layer.

The computing unit of a neural network is the *perceptron*, a node capable of processing information coming through the connection weights, combining them linearly with a set of weights w and produces a single output following a chosen *activation function*, as shown in Fig. (4.8b).
Mathematically, the output is computed as:

$$y_j(\vec{x}) = \phi_i \left(\sum_{i=1}^{n^i} w_j^i z_j^i + b^i \right) \tag{4.7}$$

where i is the total incoming connections, z^i is the input, w^i is the weight, b^i is the bias, and $\phi_i(\cdot)$ is the *activation function* at the i-th node that limits the amplitude of the output node into a certain range.
The activation transforms the signal non-linearly and it is this non-linearity that allows us to learn arbitrarily complex transformations between the input and the output. Note that without the nonlinearity in the hidden layer, the neural network would reduce to a generalized linear model.
The structure of a FNN, or in general of a NN, can be thought as a function $f(\vec{x}, \vec{w})$ as a solution to the problem. Therefore, the items to be addressed when solving a problem using a FNN are: discover an appropriate function f (i.e. the architecture optimization) and to discover an appropriate weight vector w (i.e. the weights optimization) using some learning algorithm performed using analytic derivatives and stochastic gradient descent as will be clarify in the following. This part is the so-called *training-phase* which essentially is the task to find a desired output y through a subsequent iteration steps from a y' initial value. The difference between the predicted output and the desired output $(y' - y)$ is converted to a metric known as the *loss function*. The goal of the training process is to find the weights and bias that minimise the loss function over the training set.

At any point in the training process, the partial derivatives of the loss function with respect to the weights are functions to iteratively find their local

[4]The number of predictors are equal to the number of training features used for the problem.

minima. This is called gradient descent, where the weights are updated using a method called **backpropagation**. Basically, each value of the input layer is passed to each node of the next layer with the corresponding weight. The output is then passed to the next layer and so on. Finally, the last output layer returns the results of the computation of the neural network: the number of output node correspond to the discriminant values used as a test statistic (which is one for a binary classification). Then, the error is back-propagated in the network using derivatives. This is done in an iterative manner till a moment when the derivative of the loss vanishes indicating a minimum finding, which can be a local or a global minimum with different implications. However, there are some ways to avoid or eliminate local minima in FNN optimization. For instance, if the weights and training patterns are assigned randomly to a three-layer FNN that contains h neurons at the hidden layer, then gradient-descent algorithm can avoid trapping into local minima [105], [106].
Several different loss functions can be used[5], for example the *mean squared error*, defined as:

$$L(w) = \sum_{i=1}^{N} \left(y_i^{true} - y(\vec{x}_i)\right)^2 .$$

The backpropagation procedure can be then formulated as:

$$w_{ij}^n \to w_{ij}^n - \alpha \frac{\partial(w)}{\partial w_{ij}^n};$$

where α is called *learning rate*: a the parameter that controls how large is the change to the weight w at each iteration. The weight is then updated proportionally to its influence on the final results. Formally, a neural network defines a function of composed operations:

$$\boxed{f_L(\vec{w}_L, f_{L-1}(\vec{w}_{L-1}, \ldots, f_1(\vec{w}_{1,\vec{x}})))}. \tag{4.8}$$

Gradient descent requires starting values for all the weights.
Typically one would use random starting weights, close to zero, random uniform or Gaussian weights are common. There are a multitude of "tricks of the trade" in fitting a neural network, and many of them are connected with gradient descent. Rather than process all the observations before making a gradient step, it can be more efficient to process smaller batches at a time. This technique, called *stochastic gradient descent*, is characterised by the weights that are iteratively modified after a bunch of training observations and the previous formula reads:

$$w_{ij}^n \to w_{ij}^n - \alpha \tilde{\Delta}_w L(w). \tag{4.9}$$

This speeds up the training process and introduces some noise in the minimization that can be useful to exit local minima of the loss function. When all

[5]For a complete list of predefined loss function the reader can visit: `https://keras.io/api/losses/`.

the samples have been used, the dataset is shuffled and another *epoch* of the training is started with new bunches. The number of samples in each bunch is also a tunable hyper-parameter of the network and its called *batch size*.

Apart from the many details associated with gradient descent, there are several other important structural and operational aspects of neural networks that have to be specified and has to be optimized in order to obtain the best performance. The most important ingredient for the success of a machine learning analysis is the origin and the structure of the sample data. The general receipt suggest to create at least three separate samples for the model definition:

- **Training Dataset**: the sample of data used to fit the model. This a portion of the whole generated dataset that we use to train the model (weights and biases in the case of a Neural Network). The model sees and learns from this data. The training dataset is even shuffled before each epoch to avoid any possible data-related updates in the network.

- **Validation Dataset**: the sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. We use this data to fine-tune the model hyperparameters. Hence, the model occasionally sees this data, but never does it "learn" from this.

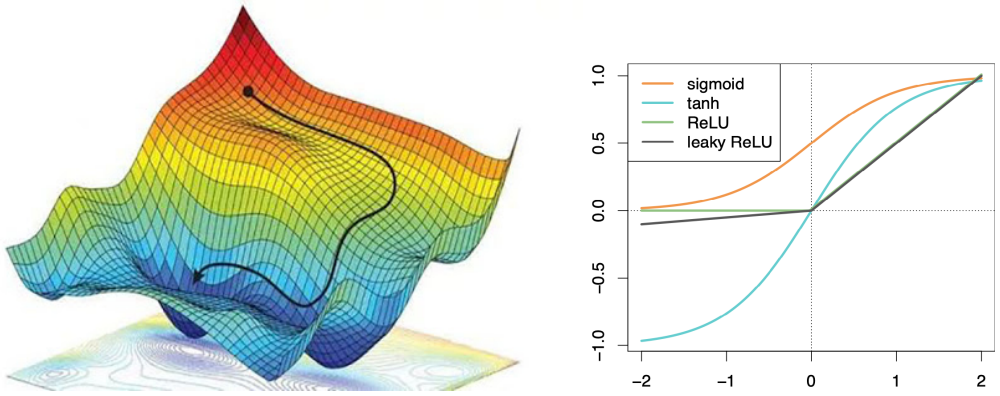
- **Test Dataset**: it provides the gold standard used to evaluate the model. It is only used once a model is completely trained (using the train and validation sets) and it is the source of our plot used to evaluate the goodness of our method.

Activation Function

Activation functions, of course, are also a fundamental element in neural network definition. There are several in current use, see a selection of them in Fig. (4.9b).

Considering hidden layers, *sigmoid* activation function (like the logistic function and hyperbolic tangent) have proven to work well indeed, yet, these suffer from vanishing gradients when the networks become too deep. Namely, the gradient will be near zero and it causes very slow or no learning during backpropagation. In that case, ReLUs [108] have become popular and have been widely used in this work . The choice of activation function is even more important at the output layer and often depends on the loss function due to the complex and iterative chain of backpropagation. For quantitative regression is typically the *identity* function, and for a binary response it is the *sigmoid* or *softmax* function.

To be more precise and to better understand the choices adopted in this work we summarize the main properties:



(a) (b)

Figure 4.9: (a) Path-evolution of gradient descent [107]: finite-length steps iterative optimization for finding a local minimum that exhibit a typical zigzagging nature of the method according to the gradient vector direction at each iteration. (b) Collection of the most common activation functions on the y axis as a function of their input (see Eq. (4.7)) [99].

- *Softmax* outputs produce a vector that is non-negative and sums to 1. It is used in case of mutually exclusive categories like the possible multiple polarization states LL, TT, LT.
- *Sigmoid* outputs produce a vector where each element is a probability. Usually adopted in situation where we have as many sigmoid neurons as we have categories, and labels should not be mutually exclusive. Of course, in case of a single output node, sigmoid can be used as a mutually-exclusive binary problem since the Kolmogorov axioms imply that when y is binary, we have $1 - p(y = 1) = p(y = 0)$.
- *Identity* function as an output can be helpful when outputs are unbounded.
- *ReLu* instead, can be helpful when the output is bounded above (below). this is typically used in hidden layer.
- tanh function can be set if the outputs are somehow constrained to lie in $[-1, 1]$.

Overtraining and Hidden Layers

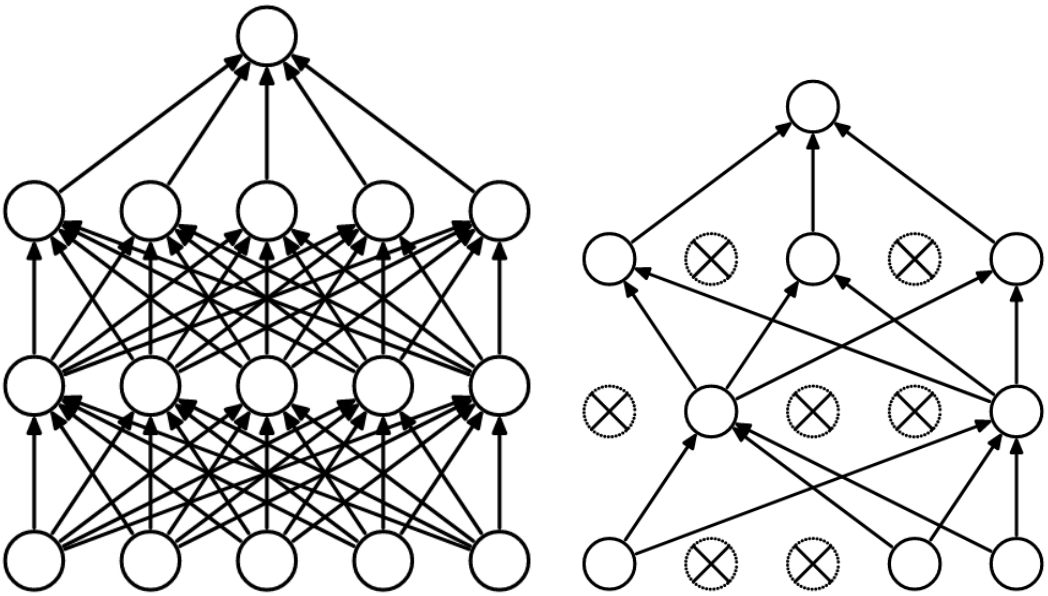
A potential general problem with the training of machine learning algorithms is the possibility that the algorithm tries to exploit artefacts of the necessarily finite size of training samples that are not representative of the actually expected distributions, also because NN are typically over-parametrized. The presence of overtraining can be checked looking at the difference in performance between the training dataset and an independent validation dataset as discussed in App. (D.2). This behaviour will be recalled during the discussion about VBS NN procedure.

Usually, the training loss keeps decreasing for each epoch, whereas the loss calculated on the validation dataset reaches a minimum and then start increasing as an effect of the overtraining. To achieve the best performance, a safe training procedure is to stop the training when the validation loss reaches a plateau. This technique is called *Early Stopping*. Originally, early stopping was set up as the primary tuning parameter, and the stopping time was determined using a held-out set of validation data. Nevertheless, in modern networks the regularization is tuned adaptively to avoid overfitting, and hence it is less of a problem.

We talked about the properties of some of the hyper-parameters that can be tuned to get the best result in our analysis such as the activation function of the hidden layer nodes, the activation function of the output node and the loss function. It is of great importance trying to get the best balance in terms of number of layers and nodes per layer that characterize the network. With a single hidden layer, the number of hidden units determines the number of parameters. In principle, one could treat this number as a tuning parameter, which could be adjusted to avoid overfitting. The current collective wisdom suggests it is better to have an abundant number of hidden units, and control the model complexity instead by weight regularization. Having deeper networks (more hidden layers) provides a hierarchy of layers that build up increasing levels of abstraction from the space of the input variables to the output variables, however the correct number tends to be task specific.

Dropout

Another method that can be adopted to reduce the risk of overtraining is *dropout*. The term refers to dropping out units (hidden and visible) in a neural network. This is a form of regularization that is performed when learning a network, typically at different rates at the different layers. It applies to all networks, in fact, it appears to work better when applied at the deeper, denser layers [99]. This technique significantly reduces overfitting and gives major improvements over other regularization methods as clearly discussed in the original paper [109].
Considering a collection of 2^n possible possible weights of a n-node networks as in Fig. (4.10a), during test phase, if a node is retained with a probability p during training, the outgoing weights of that unit are multiplied by p. This ensures that for any hidden unit the expected output during training is the same as the actual output at test time. With dropout, the idea is to randomly set, each of the nodes to zero, with probability given by the dropout-rate. Using the dropout technique applied to the hidden layer we simulates a sparse activation from a given layer which, interestingly, encourages the network to actually learn a sparse representation. Hence, for this observation, those nodes that survive have to stand in for those omitted, Fig. (4.10b). This can be shown to be a form of *ridge* regularization, and when done correctly improves performance.



(a) Standard Neural Network. (b) Neural Network after applying dropout.

Figure 4.10: Left: A standard neural network with 2 hidden layers. Right: An example of a thinned network produced by applying dropout. Crossed units have been dropped [109].

Practically, in terms of implementation, we specify the probability at which outputs of the layer are dropped out, or inversely, the probability at which outputs of the layer are retained. In addition, perhaps dropout breaks-up situations where network layers co-adapt to correct mistakes from prior layers, in turn making the model more robust.

We conclude this section by summarizing its main points: the effectiveness of FNN training primarily depends on data quality. This is governed by the following data quality assurance parameters: accuracy, reliability, timeliness, relevance, completeness, currency, consistency, flexibility, and precision. One problem related to data-driven modelling (FNN learning) is the data itself, which can be insufficient, imbalanced, incomplete, high-dimensional, or abundant.

In this work we adopted a grid search approach, selecting different high-granular NN configurations and trying to spot the macro parameters that were performing better, moving then to a micro tuned grid search. From a general point of view, the design of the correct architecture for a deep-learning network, along with the various choices at each layer, requires experience and trial and error.

4.2.2 Semi-leptonic VBS channel

For the semi-leptonic VBS channel, two different NN approaches have been adopted: Binary classification (Section 4.2.2) and Regression (Section 4.2.2). In both cases, the same dataset for training, testing and validation phases is used, the only difference lies in the neural network problem formulation. The dataset characteristics are reported at the beginning of this Chapter, other details about dataset dimension, variables and assumptions can be found in App. (D.2).

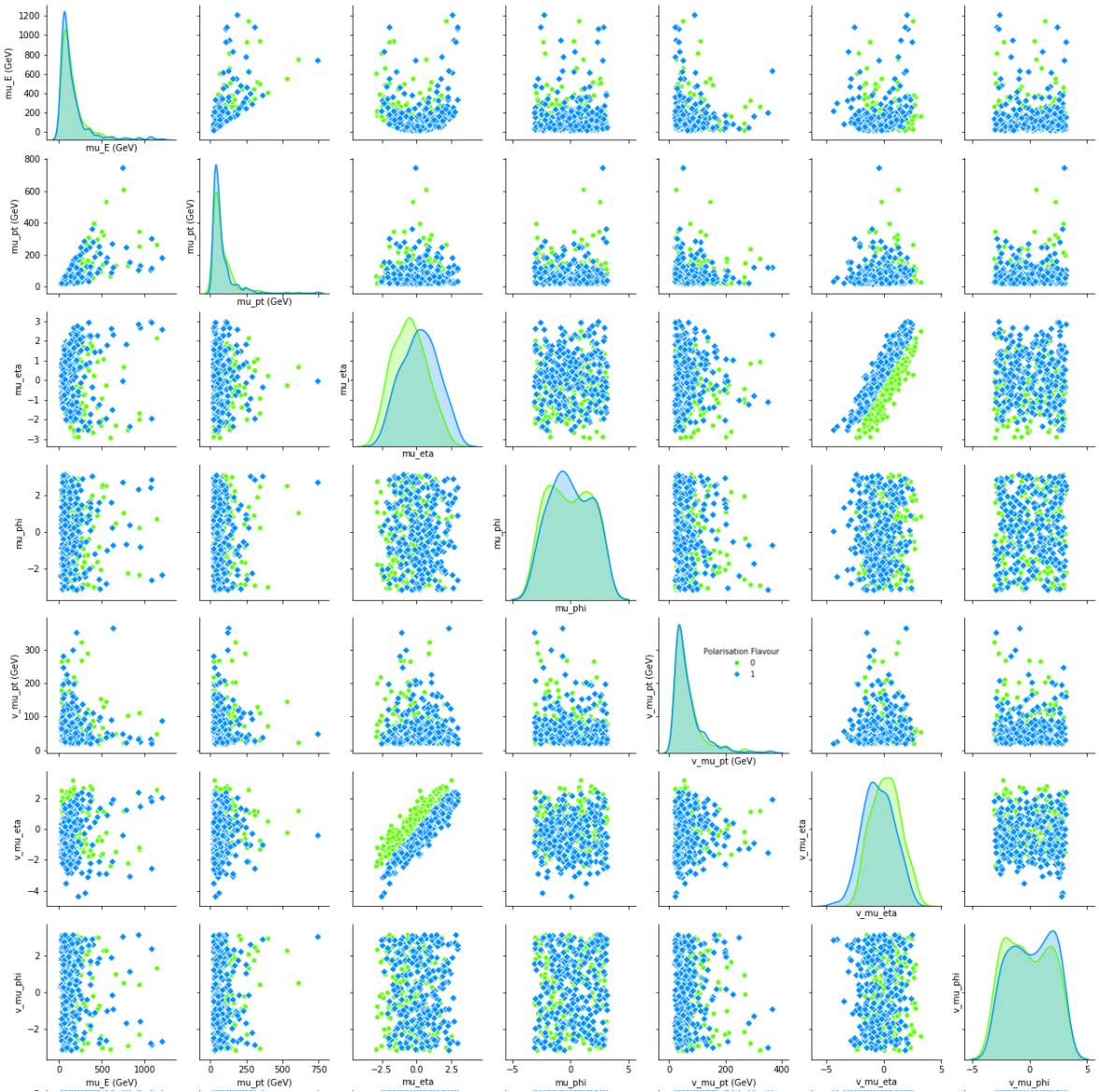


Figure 4.11: Pairplot of a selection of variables from semileptonic channel. Events are colored according to the correct 0 or 1 solution. The separation power of the two classes is really low, and this explains the difficulty of the problem.

A Binary Discriminator

The first DL implementation, namely Deep Neural Networks (DNN), that we explored to study VBS makes use of a binary classifier to determine first the correct solution of Eq. (4.1), and then once the neutrino momentum is identified, to reconstruct the W boson reference frame. Our binary classifier is a FNN with one node as the output layer. The code deduces which of the two $(+/-)$ solutions is chosen by nature (addressed in the following as the *correct solution*), based on training variables. In this way the kinematics of the boson decay is fully reconstructed, as discussed in Section (4.1.1), but with a set of weighted parameters configured by the DL approach.

The DNN input consists of 27 variables, among which we include both simple variables related to the momentum and geometrical disposition of the physical objects and more complex objects that combine several features. The detailed list is given in App. (D.2).
In our binary classifier, the label is assigned on event level basis following the criterion: events with negative (positive) sign giving the result closer to truth $p_{\nu L}$ are assigned label 0 (1), events with negative discriminant are discarded[6]. The truth is determined event by event by using the full information available in PHANTOM Les Houches file [54]. The main results are shown in Fig. (4.11).

Our DNN is trained on truth solution labels, with a single output node, as

[6]In this approximation the mass of W boson is fixed at the value of 80.38 GeV, according to the Particle Data Book [7].

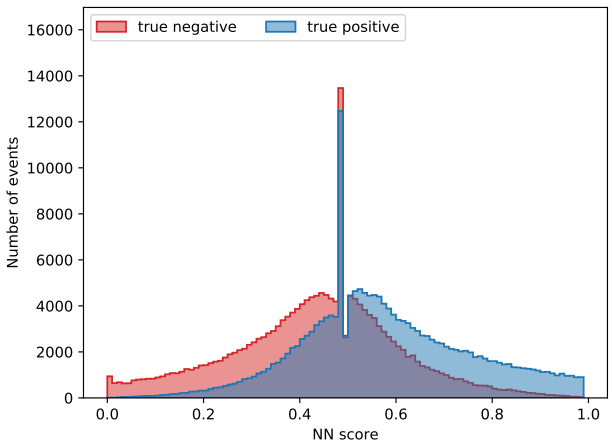


Figure 4.12: The DNN binary classification score in the semi-leptonic channel, determining the appropriate sign to use in neutrino momentum reconstruction. Here results are shown separately for events with true positive and true negative solutions, evaluated on events with transverse W polarization. The DNN model is composed of 4 hidden layers and 60 neurons in each of them, the training was performed on an unpolarized PHANTOM sample. The sharp peak at 0.5 represents events for which the classifier cannot make a decision.

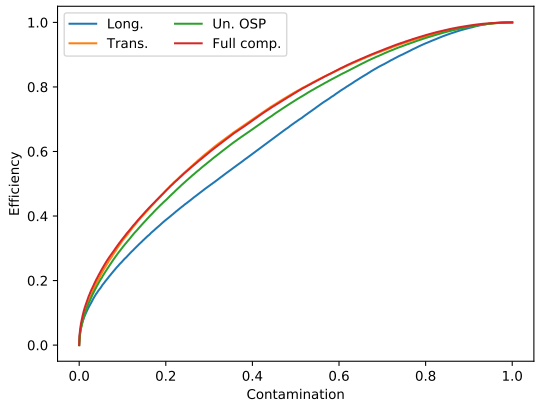


Figure 4.13: ROC curves in the semi-leptonic channel, representing the fraction of correct positive solution choices (efficiency) versus the fraction of negative solutions falsely labelled as positive (contamination), obtained on PHANTOM samples containing a purely longitudinally polarized (Long.) leptonically decaying W boson, a purely transversely polarized (Trans.) boson, an unpolarized W boson generated in the OSP PHANTOM framework (Un. OSP) and the full computation with an unpolarized W boson and with non-resonant production modes included (Full comp.) are shown. The optimal working point is defined as the threshold giving the shortest distance to point (0,1).

a binary classifier, with the following parameters:

- Hidden Layer Activation Function: *ReLu*;
- Output Layer Activation Function : *Sigmoid* $g(x) = \frac{1}{1+e^{-x}}$ to provide a score bounded between 0 and 1;
- Loss function: *binary_crossentropy*: $L(p, y) = -(y \ln(p) + (1-p) \ln(1-p))$ where y is the target label (0 background or 1 signal) and p is the predicted score $p \in [0, 1]$;
- Optimizer: *Adam gradient descent* classical stochastic gradient minimizers use only one learning rate for all the weights in the networks, Adam instead maintains a learning rate for each weight and separately adapts it as learning unfolds.

The score distribution, obtained when training only on the transversely polarized PHANTOM sample is shown in Fig. (4.12). Optimal working points for classification have been selected for each model based on the Receiver Operating Characteristic (ROC) curves [110]. An example of ROC curves for the DNN model using 60 neurons and 4 hidden layers is shown in Fig. (4.13). Values of area under curve (AUC) and fractions of correct solution choices are summarized in Table (4.2).

An important aspect already discussed in the previous section is the performance of the NN with respect to the data dimensionality. On this crucial point a dedicated study has been preformed on our network. Different sizes of

Table 4.2: Classifier performance in the semi-leptonic channel, using 60 neurons and 4 hidden layers. Results obtained on samples containing a purely longitudinally polarized (W_{L}) leptonically decaying W boson, a purely transversely polarized (W_{T}) boson, an unpolarized W boson generated in the OSP PHANTOM framework (Un. OSP) and the full computation with an unpolarized W boson and with non-resonant production modes included (Full comp.) are shown.

	W_{L}	W_{T}	Un. OSP	Full comp.
Working point	0.497	0.498	0.500	0.491
AUC	0.655	0.716	0.698	0.717
Corr. choices	0.596	0.651	0.636	0.650

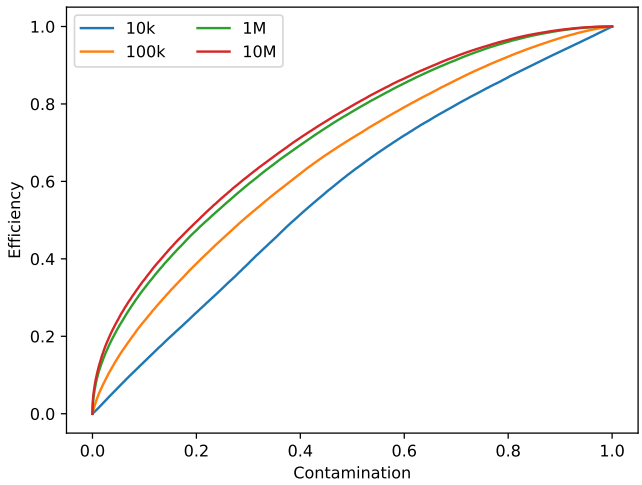


Figure 4.14: ROC curve dependency in the semi-leptonic channel, showing on the training unpolarized OSP dataset size (10^4, 10^5, 10^6, 10^7 events). Number of neurons is fixed to 60 and number of hidden layers to 4.

the training sets have been considered, spanning from ten thousands events to ten millions events. The resulting ROC curves are shown in Fig. (4.14). One can see that the performance improves with the increase of the training set size, but it seems to converge quite well at size one million. It can be noticed that difference between one million and ten millions is practically negligible. Given this feature, we adopted a training set size of one million, to perform further optimizations of the DNN parameters.

Following Eq. (4.1), $p_{\nu L}$ and E_ν in the laboratory frame are calculated by the NN. Finally, the lepton momentum is boosted into reconstructed W rest frame, defined by the vector sum of neutrino and lepton momentum. The resulting angular distribution in the W boson rest frame, obtained using the binary classification technique, is shown in Fig. (4.15) as a function of $\cos\vartheta$, for a DNN with a fixed number of neurons (60) and different hidden-layer configurations.

In addition to the standard open-source platforms, we used a specific one

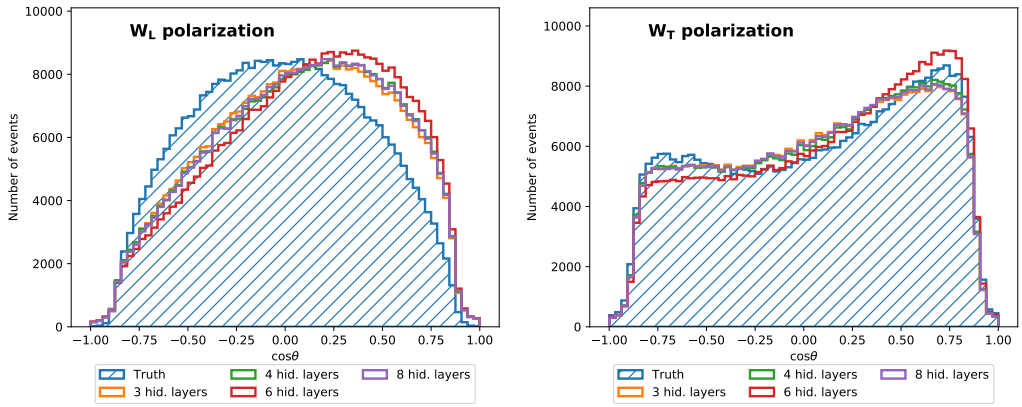


Figure 4.15: Angular distribution in the W boson rest frame as a function of $\cos\vartheta$, obtained using the binary classification technique in the semi-leptonic channel. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$. The left plot shows results obtained assuming the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario. The results for a DNN with a fixed number of neurons (60) and different hidden-layer configurations are shown.

available on IBM Cloud, named AutoAI, that exploits artificial intelligence routines to generate candidate model pipelines customized for predictive modelling problem. These model pipelines are created iteratively, AutoAI analyses datasets and discovers data transformations, algorithms, and parameter settings that work best for the problem setting. It covers all the steps of a typical machine learning project from data pre-processing, to model selection, feature optimization and model deployment. In particular, it automatically detects and categorizes features based on data type, such as categorical or numerical. Then it transforms the raw data into the combination of features that best represents the problem to achieve the most accurate prediction. AutoAI uses a unique approach that explores various feature construction choices in a structured, non-exhaustive manner, while progressively maximizing model accuracy using *reinforcement learning*.

Finally, a hyper-parameter optimization step refines the best performing model pipelines. AutoAI uses a novel hyper-parameter optimization algorithm that enables fast convergence to a good solution despite long evaluation times of each iteration. The model found by Watson AutoAI, as the best one, is LGBM Classifier as can be seen in bottom part of Fig. (4.16). This is a Gradient boosting framework that uses leaf-wise (horizontal) tree-based learning algorithm. It can handle the large size of data and takes lower memory to run. Light GBM is sensitive to overfitting and can easily overfit small data (less than 10k events). A complicated thing is parameter tuning: Light GBM covers more than 100 parameters and the advantage of AutoAI implementation is the automation in this hyperparameters optimization. Light GBM grows tree vertically while other algorithm grows trees horizontally meaning that Light GBM grows tree leaf-wise while other algorithm grows level-wise. It will choose the leaf with max delta loss to grow. When growing the same leaf, Leaf-wise

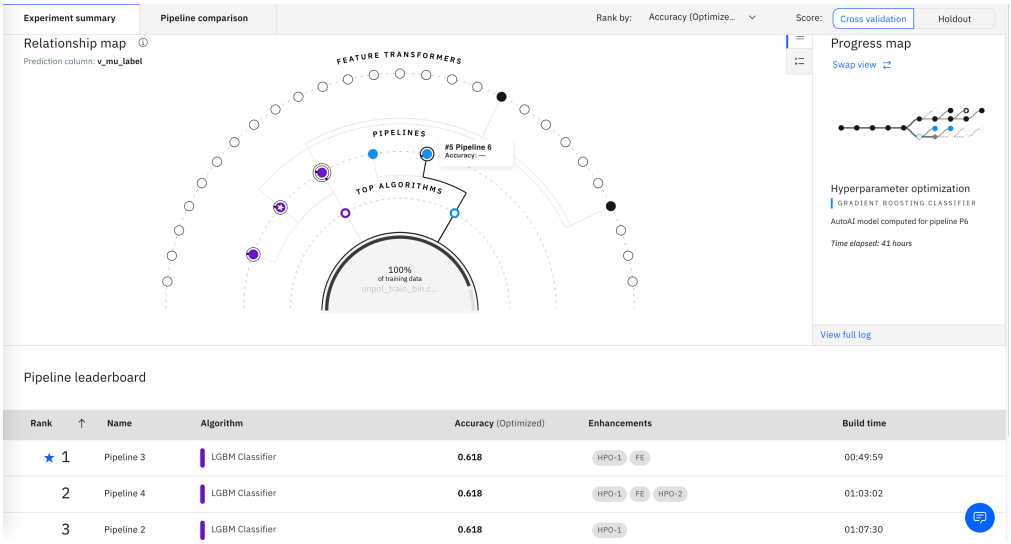


Figure 4.16: Screenshot taken from the AutoAI tool where we can appreciate the collection of all possible model parametrization tested by the machine (top) and the ranking of all model according to the selected metrics (accuracy), in the bottom part.

	Holdout Score	Cross Validation Score
Accuracy	0.618	0.618
Area Under ROC Curve	0.671	0.670
Precision	0.620	0.620
Recall	0.616	0.612
F_1 Measure	0.618	0.616
Average Precision	0.667	0.665
Log Loss	0.644	0.644

Figure 4.17: Screenshot taken from the AutoAI tool of the LGBM classifier with the highest accuracy score. We can see the plot of the ROC curve on the left and the table collecting all possible metrics obtained by the trained model on the right.

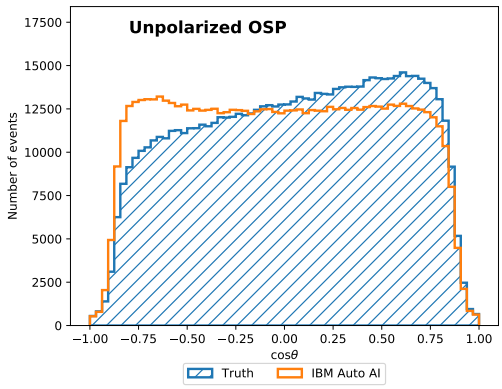


Figure 4.18: Angular distribution in the W boson rest frame as a function of $\cos\vartheta$, obtained using the IBM auto AI tool as a binary classifier. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$.

algorithm can reduce more loss than a level-wise algorithm.
The performance of the model are shown in Fig. (4.17), where not only the ROC curve is presented but also an exhaustive table with several metrics is given. These evaluation metrics are provided with two approaches: hold-out method, where the whole dataset is splitted into training and validation ones according to the fraction indicated and cross-validation or "k-fold cross-validation" where the dataset is randomly split up into k groups. One of the groups is used as the test set and the rest are used as the training set, so the model is trained on the training set and scored on the test set. Then the process is repeated until each unique group as been used as the test set. Cross-validation is usually the preferred method because it gives your model the opportunity to train on multiple train-test splits. The hold-out method instead, is good to use when one has a very large dataset or is on a time crunch, or in case one is starting to build an initial model for a given data science project.

Comparing Fig. (4.15) and Fig. (4.18) we see that, in employing the IBM AutoAI tool, there is no further improvement in our binary classification as we already extracted all possible information belonging to training data. We made this comparison for binary classification where the AUC value represents a clear statistical indicator, although we got a similar conclusion also for the technique that will be discussed in the next paragraph.

Regression Technique

Instead of using the neural network to address the solution of Eq. (4.1), we could train the algorithm to determine directly the reconstructed angular distributions of the lepton in W rest frame, with respect to the W direction in

the laboratory frame. The DNN input is composed by the same set of 27 variables of the previous case. Differently from the binary classification approach, here the DNN is trained with the true value of the $\cos\vartheta$ variable, which it aims at reproducing. With this approach, we tested as many different neural network topologies as in the case of binary classification, albeit changing the relevant parameters that convert a classification problem into a regression one, as described in the following:

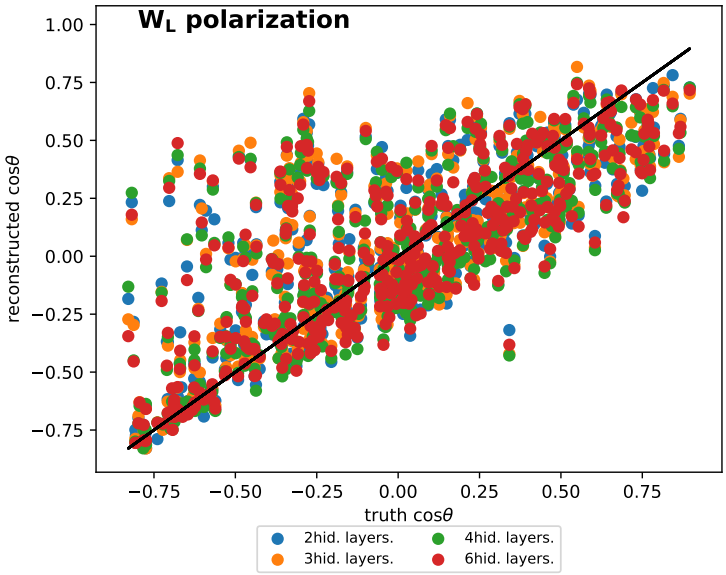


Figure 4.19: Scatter plot: generator-level (true) $\cos\vartheta$ value w.r.t. the reconstructed one in the semi-leptonic channel, using different network topologies. The closeness to the black line can qualitatively estimate the goodness of the topology adopted.

- Activation Function for hidden layer: *ReLu*;
- Activation Function for output layer: *Linear*.
  It creates a continuous output signal proportional to the input $g(x) = x$;
- Loss function: *mean squared error*: $L(\cos\vartheta) = \frac{1}{N}\sum_{i=1}^{N}(\cos\vartheta_{1,i}^{truth} - \cos\vartheta_{1,i}^{pred})^2$ where N is the number of events per batch, $\cos^{pred}$ and $\cos^{truth}$ are the predicted score and truth label respectively.
- Optimizer: *Stochastic gradient descent* (SGD). It is a variant of gradient descent. Instead of performing computations on the whole dataset, only computes on a small subset or random selection of given data.

One additional way to evaluate the performance of the model for different neural network topologies is presented in Fig. (4.19), where the distance of points from the main diagonal represents the ability of the network to reconstruct $\cos\vartheta$. The resulting angular distribution in the W boson rest frame

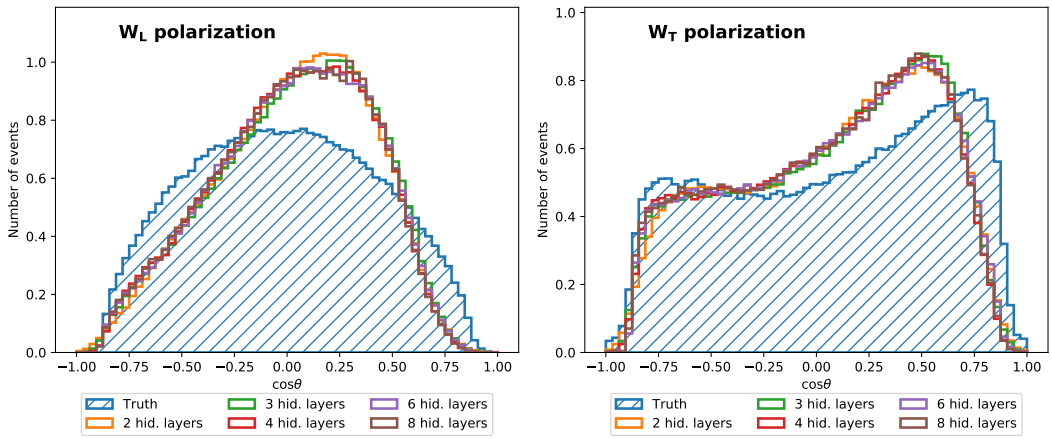


Figure 4.20: Angular distribution in the W boson rest frame as a function of $\cos\vartheta$, obtained using $\cos\vartheta$ regression in the semi-leptonic channel. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$. The left plot shows results obtained assuming the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario. Results for a DNN with a fixed number of neurons (60) and different hidden-layer configurations are shown.

as a function of $\cos\vartheta$ is shown in Fig. (4.20) for a DNN with a fixed number of neurons (60) and different hidden-layer configurations. To evaluate the performance of this approach, the distribution of the reconstructed $\cos\vartheta$ for longitudinally and transversely polarized samples is compared to the truth values.
Similarly to binary classification, a dedicated study has been made to evaluate the dependence of the DNN on the dimension of the training set. Different sizes of the training sets have been considered, spanning from 10k events to 10M events.

The performances of the regression technique as a function of the training sets are shown in Fig. (4.21). As expected, the larger the training set size, the better the performance, although one should also consider the possibility of over-fitting of the DNN. For a qualitatively comparison in the reconstruction of $\cos\vartheta$ for longitudinally and transversely polarized samples, between the DNNs used as regressor and the DNNs used as a binary classifier, one should compare Fig. (4.20) and Fig. (4.15).

4.2.3 Fully-leptonic VBS channel

In the previous Sections, we tested a NN approach to semi-leptonic channel. Here we focus on the fully-leptonic channel. From a technical perspective, this channel can be considered as a direct extension of the semi-leptonic case, even though only the regression strategy hold, since the kinematic reconstruction is not related to a sigle sign ambiguity anymore. The fully-leptonic channel

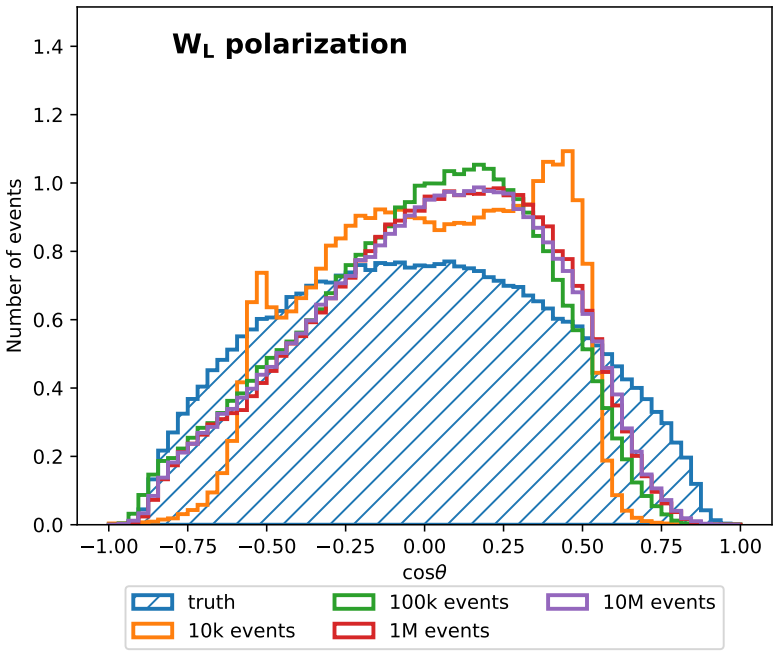


Figure 4.21: Reconstructed lepton angular distribution in the semi-leptonic channel, for the longitudinal W boson polarization, as a function of $\cos\vartheta$ with a specific network topology (60 neurons, 4 hidden layers) for model trained with different training dataset sizes (10^4, 10^5, 10^6, 10^7 events). Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$.

kinematics reconstruction can be supplemented with the MAOS algorithm, as described in Sec. (4.1.2). Since we want to check whether the information obtained from the MAOS could improve the DNN reconstruction, we included MAOS algorithm outputs among the training features, namely MAOS reconstructed neutrino transverse momenta ($p_{\mathrm{T}}^{\nu_\ell}$ from Eq. (4.6)) and m_{T2}.

Direct Approach

In the fully-leptonic channel, our methodology extends the possibility for the network to reconstruct both $\cos\vartheta$ for the muon and electron part directly, then combines these two contributions to get the total angular distribution of a certain polarization component. We refer to this approach as a *direct* approach.

Indirect Approach

As another step in the process of improving our ability to extract a model for angular distribution, an *indirect* approach can be defined by introducing an intermediate step: in this case the DNN training labels are now the six components of neutrino momenta and the neural network now has to solve a regression problem with a six-dimensional output. Once the DNN model outputs the reconstructed six neutrino components we calculate and reconstruct

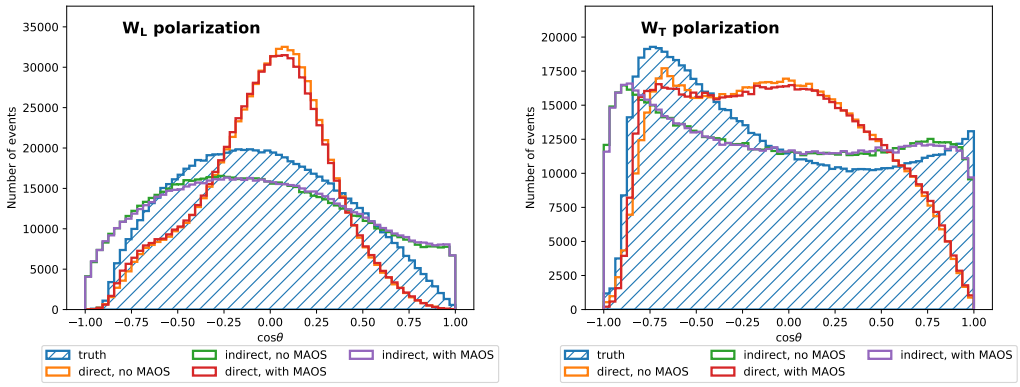


Figure 4.22: Shape of the reconstructed lepton (electron and muon) angular distribution in the W boson rest frame in the fully-leptonic channel. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$. The left plot shows the the (purely) longitudinally polarized W_{L} and the right plot the (purely) transversely polarized W_{T} sample. Performance of the direct and indirect approach can be appreciated qualitatively. 60 neurons and 3 hidden layers have been used for all the approaches. Other DNN topologies give similar results.

the $\cos\vartheta$ distribution. From NN point of view, direct and indirect approach are built as a regression problem. For this reason we adopted as in the semi leptonic case a linear activation function for the output node, likewise for loss function extending the formula with respect to the number of output node, two and six for the direct and indirect[7] approach respectively.

A summary of the performance of both reconstruction approaches, with and without the inclusion of the MAOS quantities among the training features, is shown in Fig. (4.22). To help the comparison, these results are superimposed to those of Fig. (4.7), and they appear in the summary plot in Fig. (4.23). It is evident how, in particular for the longitudinal components, the DNN approach outperforms the classical kinematics approaches. In Fig. (4.24), we appreciate the performance of the NN in the indirect approach to reconstruct its continuos distribution of neutrinos missing momenta.

4.3 Direct Polarization Discriminator

The approach followed in this Section involves Deep Neural Networks for the development of a classifier to discriminate between the different polarized events both for semi-leptonic and fully-leptonic case.
In the semi-leptonic case a binary classifier to discriminate between longitudinal (L) and transverse (T) is built, while for the fully-leptonic one a multi-classifier able to discriminate between longitudinal (LL), mixed (LT) and trans-

[7] In this case, the variables need to be scaled in the range $[0, 1]$.

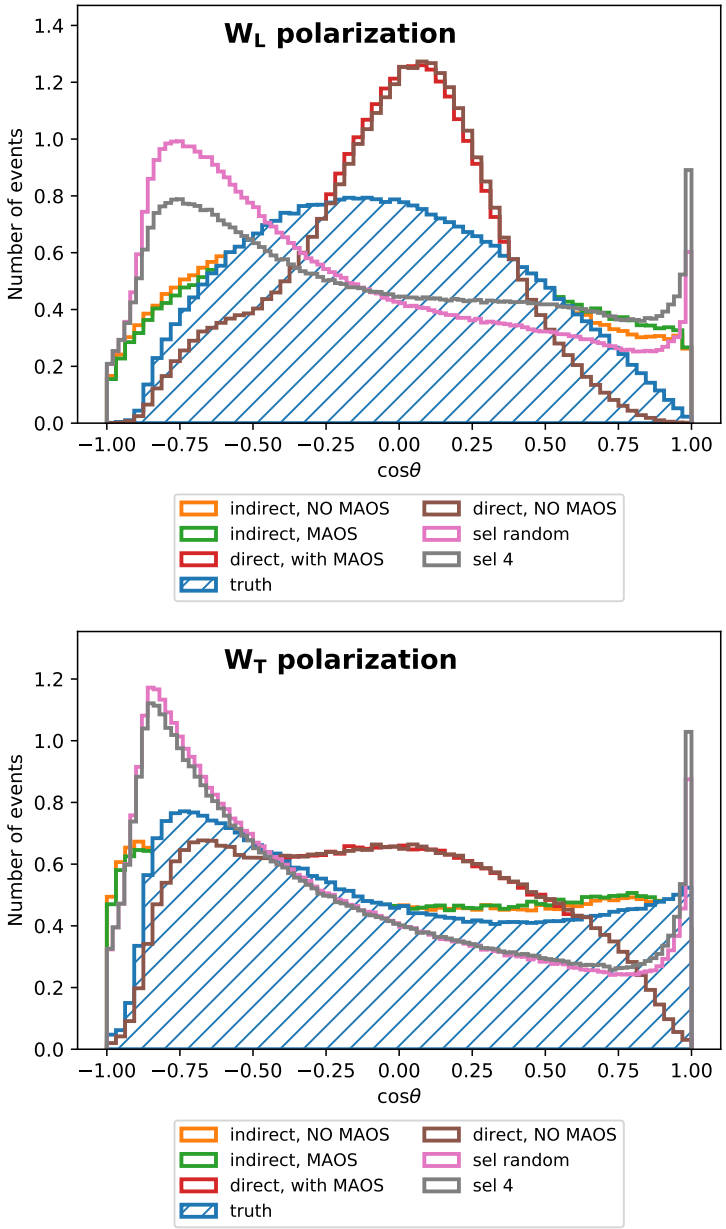


Figure 4.23: Summary plot of the reconstructed lepton (electron and muon) angular distribution in the W boson rest frame in the fully-leptonic channel. Standard approach (selection 4 strategies described in Sect. (4.1.1)), random solution, and the DNN approach (direct, indirect with/out MAOS variables inclusion) are compared to the Parton-level (truth) distributions, using the generator-level (true) $\cos\vartheta$. Top plot shows the the (purely) longitudinally polarized W_{L} and the bottom plot shows the (purely) transversely polarized W_{T} sample.

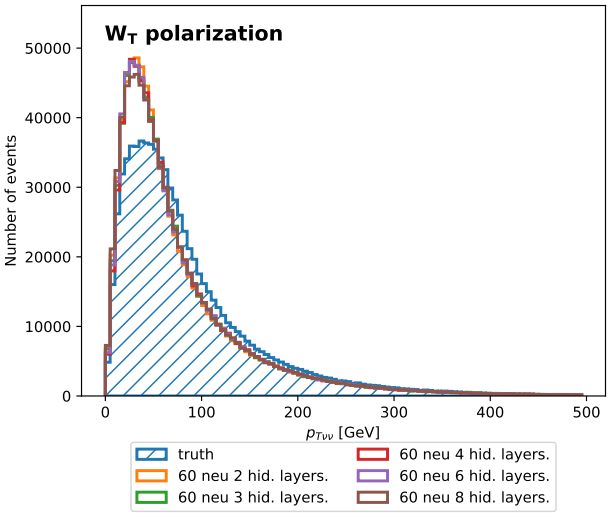


Figure 4.24: Reconstruction of $p_{\mathrm{T}\nu\nu}$ for different neural network topology with the same batch-size, applied on the fully-leptonic channel using a transversely polarized W_{T} scenario.

verse (TT) scattering events is defined. Here again, we ask machine learning techniques to bring out correlations between variables and to extract as much information as possible from the finite dataset.
In addition, a further consideration is that the composition of the training sample is a mixture of data generated according to different polarization flavours. The mixed sample can be created essentially in two ways: *balanced* and *weighted*. In the first case, we create a dataset with same number of events for each category whereas, in the weighted case, the number of events for each polarization are weighted according to relative SM cross section that are reported in App. (D.3).
The two compositions are quite different, in particular for the fully-leptonic case, the ultimate goal of the polarization classification is to extract the longitudinal fraction, which is the important component from a physical point of view and should be considered as the signal in our discrimination. However the LL component represents a small fraction of the total cross section, accounting for around 9% of the total cross-section, and has observables that are largely similar to the unpolarized process. This means that working with balanced dataset requires enhancement of longitudinal component. Details about dataset composition as well as model loss consideration can be found in App. (D.2).

From classical machine learning point of view, the composition of a dataset is crucial. Unbalancing is a well known problem, typically related to rare events. This is the case for classic machine learning techniques like Logistic Regression, Discriminant analysis and Ensemble methods that need to keep track of the original sample composition and are not recommended in case of unbalanced methods[8]. Apparently, from neural network point of view, we

[8]Linear and Quadratic Discriminant Analysis are even more suppressed due to rigid

realised that working with the abstraction given by the deep inner structure of the NN, does not affect the final performance using enhancement of longitudinal (L, LL) components.

Semi-leptonic VBS channel

To better understand the complex scenario we are dealing with, we refer to Fig. (4.25) where a pairplot representation of the generated variables is shown according to the two classes (colour). From a qualitatively point of view the power discrimination seems to be a bit better with respect to the classification problem defined to solve the ambiguity of Eq. (4.1), whose distribution for each class is reported in Fig. (4.11). Our ability to reconstruct the angular distribution as a function of $\cos\vartheta$ is shown in Fig. (4.26).

Fully-leptonic VBS WW channel

Here we present the result concerning the polarization discrimination for the fully-leptonic channel. The goal here is to develop a multilabel classifier which is able to recognize mutually exclusive components LL, TL, TT. Each polarization category is tagged with one hot encoding where, to keep the structure of a classification problem, we define three binary variables which are associated to the LL, LT and TT classes, respectively. The value is 1 or 0, depending on whether the event belongs to the class or not so that the DNN target variables are these three labels, and the output consists in three DNN scores distributed between 0 and 1.
From NN point of view, the main difference is represented by the *softmax* output activation function that allow the mutual exclusivity of the choosen class. This is compulsory for this particular problem where each event cannot be assigned to multiple classes whereas other multi-classification problems allow overlap between categories so that *sigmoid* is the chosen output activation function.
The distribution of the variables according to the different classes is shown in the pair plot of Fig. (4.27). To quantitatively analyse the performance of our NN we need to extend the metrics used for the binary case, in particular the ROC Curve.

In a multi-classification problem, one can study the ROC curve for each class or define an "average" ROC curve, in order to quantify the overall performance of the classifier. In the present analysis, our *macro average* ROC curve is defined as follows:

$$\epsilon_i = \frac{TP_i}{TP_i + FP_i} \Longrightarrow \epsilon_{macro} = \frac{1}{N}\sum_{i=1}^{N} \epsilon_i;$$

dataset assumption like dataset normality and variance equality.

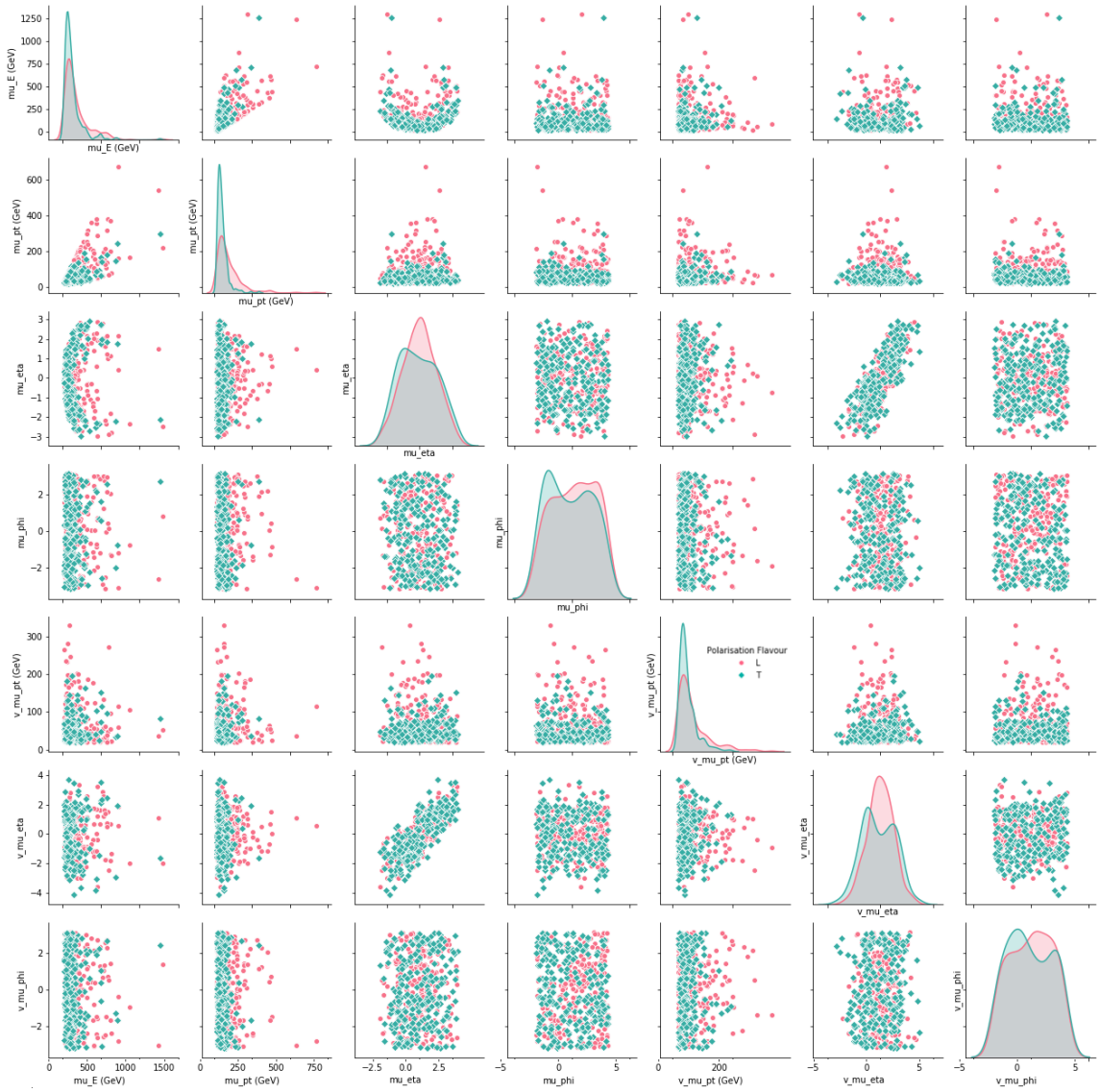


Figure 4.25: Pairplot of a selection of variable from semileptonic balanced polarized channel. Events are colored according to the polarization (L,T). The separation power of the two classes is really low which explain the difficulty of the problem.

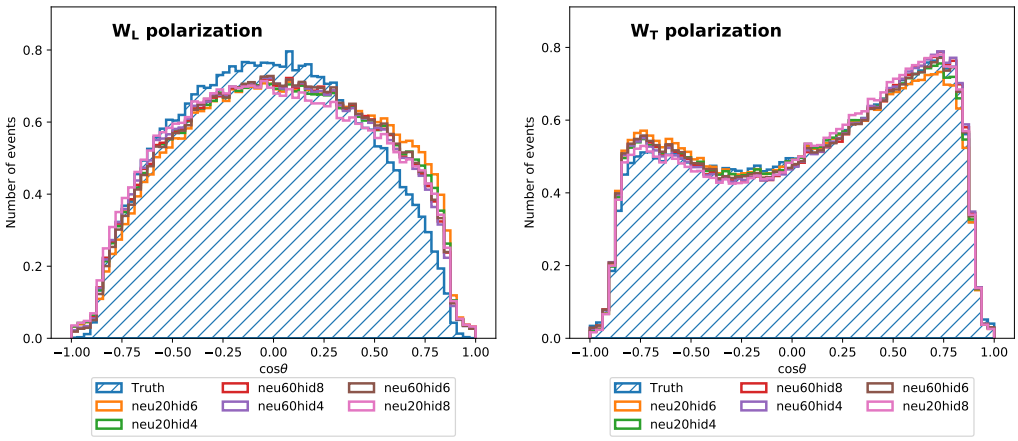


Figure 4.26: Angular distribution in the W boson rest frame as a function of $\cos\vartheta$, obtained using the binary classification technique trained on the polarization species, in the semi-leptonic channel. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$. The left plot shows results obtained assuming the purely longitudinal (W_{L}) and the right plot the purely transverse (W_{T}) scenario. The results for different number of neurons and hidden-layer configurations are shown.

where TP, FP denote the number of true and false positive of the class i, respectively, and N is the number of classes. The resulting ϵ_{macro} is nothing else than the average of the efficiency ϵ_i over all classes.

The pattern shown in Fig. (4.28) highlights that the categorization performs better for the combination with 100, 60 neurons and 4 hidden layers. The difference is not too sharp among the different models, apart from the one with 40 neurons and 2 hidden layers which is excluded by a wide margin. Again, in these results, we can appreciate the contribution of deep network topology with respect to large ones. It seems that a deeper solution increases our performance to disentangle a complex classification like the one we are willing to solve.
The ROC curves for the single classes with the optimized parameters are also shown in Fig. (4.29). We observe that the LL class has the best performance, followed by the TT class and finally the LT class which results to be more difficult to classify. This can be explained by the fact that the LT class has a mixed topology which shares properties with both the other two classes.

The output of the network consists in a score distributed between 0 and 1, one for each class. Events belonging to a given class are more likely peaked around 1 in the corresponding score distribution as can be seen in Fig. (4.30) where the distributions of the three classes of the multilabel classifier, namely, LL, LT and TT scores are reported, with the events separated according to their true polarization. The LL and TT score distributions show a nice separation of the LL and TT components, while the LT component is more distributed between 0 and 1. The LT score has less separation power than the other two, as the distributions of the different polarized components overlap, in analogy to what we pointed out looking at the ROC curve.

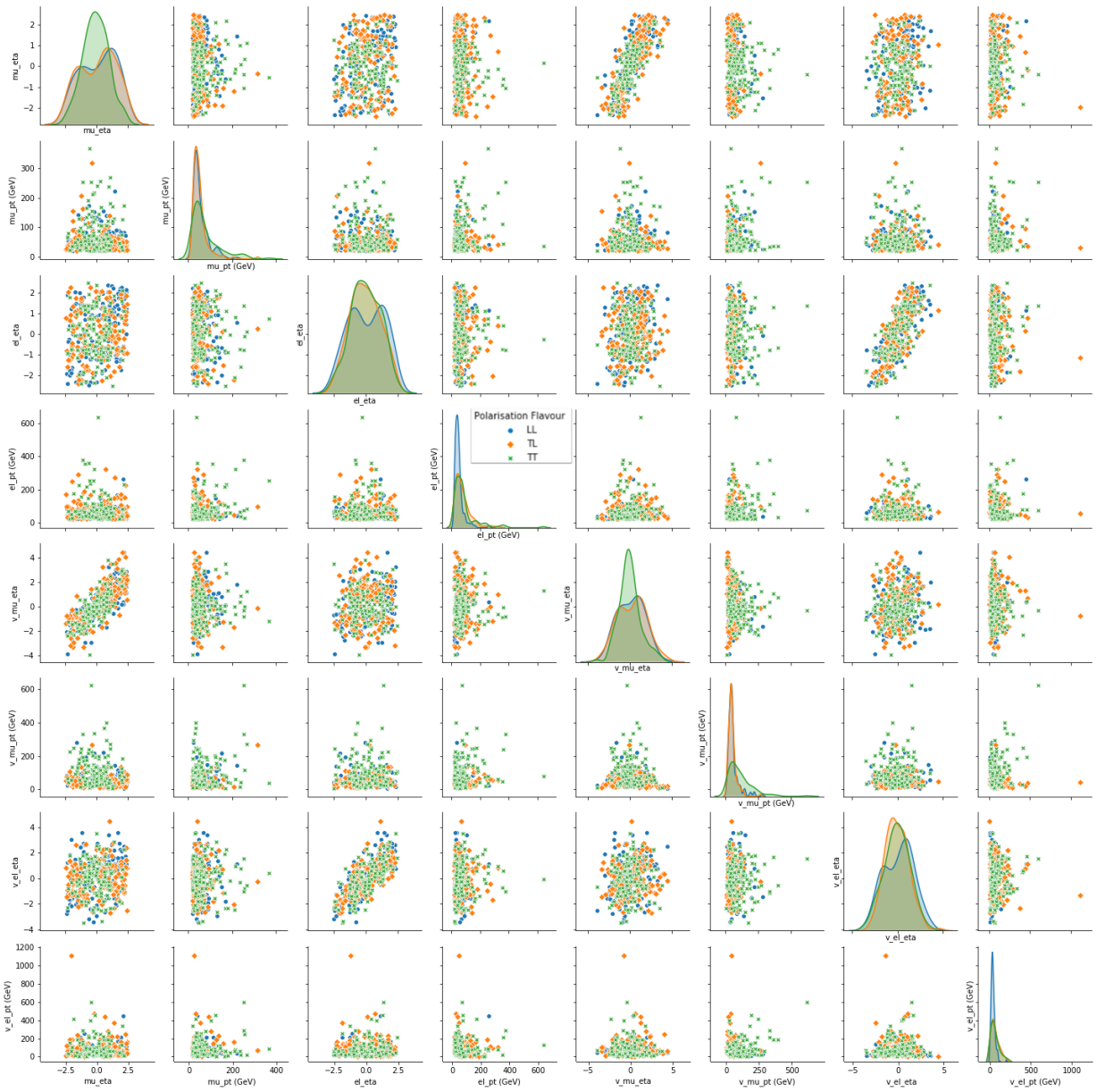


Figure 4.27: Pairplot of a selection of variable from fully-leptonic balanced polarized sample. Events are colored according to their polarization (LL, LT, TT). Plot of single variable is shown along the diagonal.

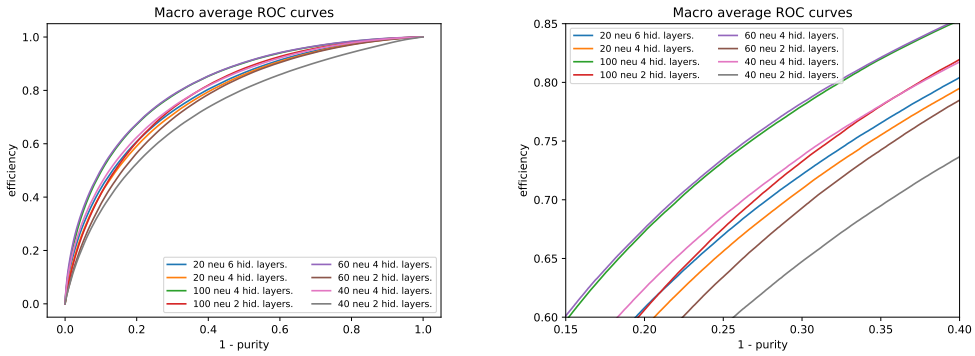


Figure 4.28: Macro average ROC curves representing the fraction of correct positive solution choices versus the fraction of negative solutions falsely labelled as positive according to the macro average definition for the fully-leptonic WW channel obtained with PHANTOM mixed samples containing LL, TL and TT contributions. Scan for different combination of neurons and hidden layer are shown (left). Figure represents a zoomed version of the top left part of the left figure to better appreciate the classification power of the NN (right).

In addition, to provide a complete description, we tested the NN on a sample generated with full computation, all the polarization states plus the interference and non-resonant contributions (Fig. (4.30d)). The DNN scores allow us to reproduce again the angular distribution as a function of $\cos\vartheta$ for both leptons (e, μ).

By comparing Fig. (4.31) with Fig. (4.22) we could infer that direct polarization discrimination performs better than the reconstruction of the $\cos\vartheta$ distribution, for both leptonic components.

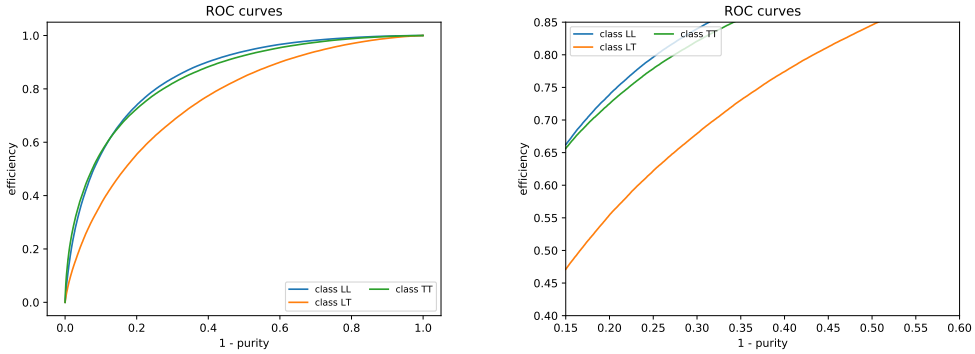
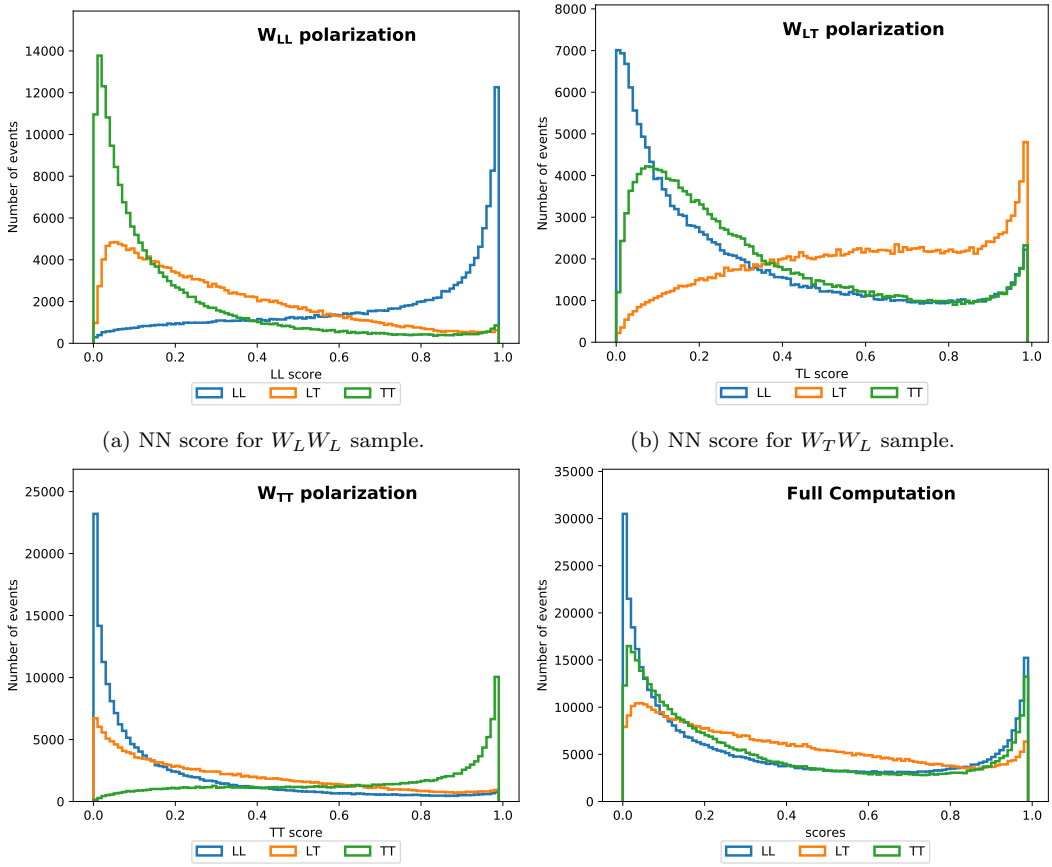


Figure 4.29: ROC curves representing the fraction of correct positive solution choices versus the fraction of negative solutions falsely labelled as positive for the single LL, TL and TT contributions of the fully-leptonic WW channel obtained with PHANTOM. Figure represents a zoomed version of the top left part of the left figure to better appreciate the classification power of the NN (right).



(a) NN score for $W_L W_L$ sample.

(b) NN score for $W_T W_L$ sample.

(c) NN score for $W_T W_T$ sample.

(d) NN score for full computation sample.

Figure 4.30: NN scores obtained for each classes evaluating different fully-leptonic polarized channel obtained with PHANTOM. Output distributions of the LL, LT, TT and full computation scores, for a DNN with 4 hidden layers composed of 100 neurons each and batch size = 128.

4.3.1 Fully-leptonic VBS ZZ channel

The performance of the framework developed and described in the previous Section has been tested also on a different VBS channel. We built a multilabel classifier for the ZZ channel, that, among VBS processes, is of particular interest as well, in particular the $ZZ \to 4l$.
This channel has not been investigated depth wise across this work as the WW one, nevertheless here we report its main properties.

ZZ VBS Process

The ZZ cross-section is lower than the WW, however, it provides a fully reconstructable final state with no missing energy, due to the absence of neutrinos in the Z decay products. Here, QCD-induced production is an irreducible back-

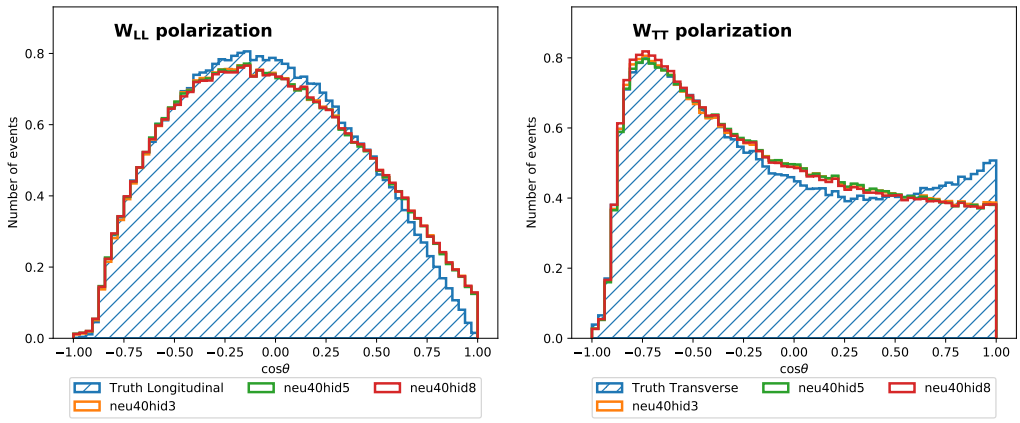


Figure 4.31: Angular distribution in the W boson rest frame as a function of $\cos\vartheta$, obtained using the multilabel classifier technique trained on different polarization species, in the fully-leptonic channel. Parton-level (truth) distributions are shown for comparison, using the generator-level (true) $\cos\vartheta$. The left plot shows results obtained assuming the purely longitudinal (W_{LL}) and the right plot the purely transverse (W_{TT}) scenario. The results for 40 neurons and different number of hidden-layer configurations are shown.

ground for the EW production of $lll'l'jj$. Fully leptonic final state, with both Z bosons decaying into charged lepton pairs, is a clean signature to tag the event in the detector, with four prompt leptons and two forward jets in the event. The invariant mass M_{ZZ} of the diboson system can be directly computed from the leptons momenta, allowing the determination of the center-of-mass energy of the scattering. Furthermore, the measurement of the spin correlations of the reconstructed fermions allows the extraction of the longitudinal contribution that, as well as in the WW channel, is a crucial measurements.

In this study, we want to test the NN performance going beyond parton level. Therefore, we generated detector level events following the chain described at the beginning of the Chapter. For sake of completeness, we report here the main steps and kinematical cuts.
The ZZ channel is generated with PHANTOM by means of the following process: $pp \to jje^+e^-\mu^+\mu^-$ at $\sqrt{s} = 13$ TeV. For this channel we generated only tree level electroweak contributions as well ($\mathcal{O}(\alpha^6)$), neglecting processes involving b and t quarks.
We applied the following kinematic cuts:

- lepton transverse momentum $p_T^l > 10$ GeV;
- leading lepton transverse momentum $p_{T,1}^l > 20$ GeV;
- lepton pseudorapidity $|\eta_l| < 2.5$;
- jet transverse momentum $p_T^j > 20$ GeV;
- invariant mass of the system of the two tagging jets $M_{jj} > 500$ GeV;

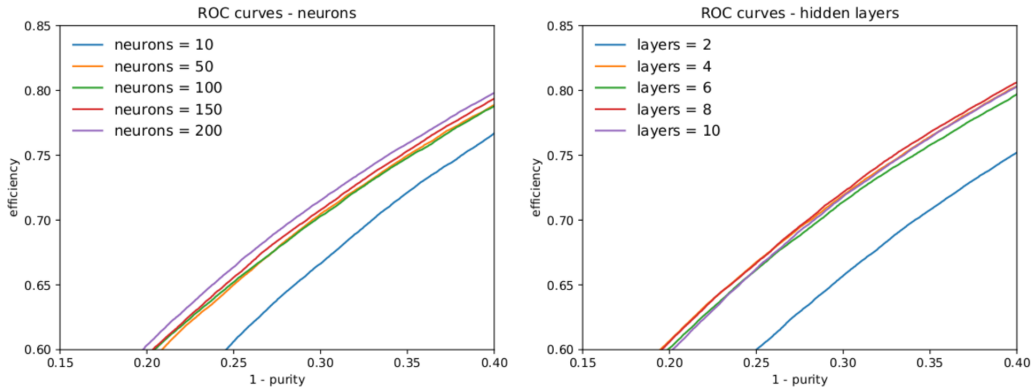


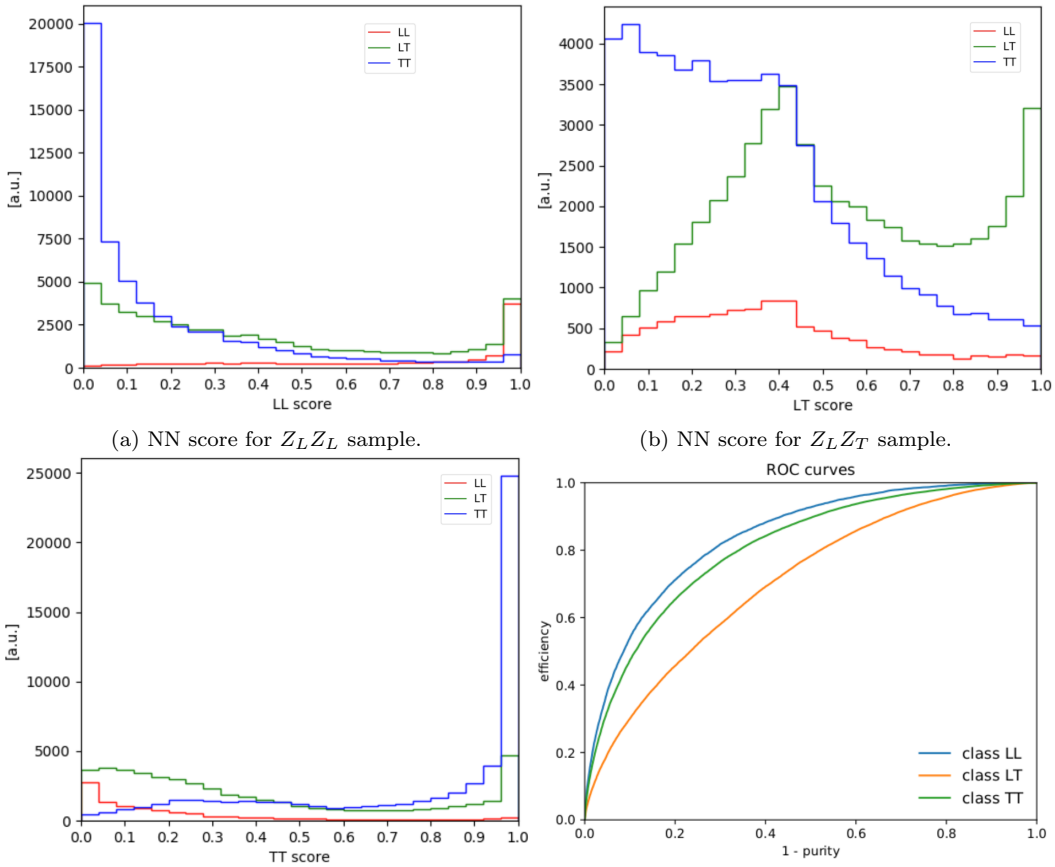
Figure 4.32: ROC curves representing the fraction of correct positive solution choices versus the fraction of negative solutions falsely labelled as positive for the single LL, TL and TT contributions of the fully-leptonic ZZ channel obtained with DELPHES. Zoomed scan for different combination of neurons (left) and hidden layer (right) are shown to better appreciate the classification power of the NN.

- pseudorapidity separation between the two tagging jets $|\Delta\eta_{jj}| > 2.5$;
- invariant mass of the four lepton system $M_{4l} > 200$ GeV.

Similarly to what we discussed for the WW channel, the cut on dijet invariant mass suppresses the triboson contribution and the cut on dijet pseudorapidity separation enhances the VBS topology. In addition, here the cut on four-lepton invariant mass M_{4l} excludes the Higgs boson peak at 125 GeV and the off-shell Z tail.

The interesting test for our NN framework in this case comes from the origin of the data. Since we want to evaluate polarization fraction at detector level, we need to define few conditions. Detector-level polarized distributions could be affected by two main contributions: pileup and detector efficiency.
Event pile-up spoils the event reconstruction because of the multiple collisions decay products overlapping in the detector.
Detector efficiency, instead, limits the acceptance and the particles reconstruction performance. From PHANTOM events we move to showered phase where jets are fragmented and hadronized with PYTHIA8 before passing through the jet reconstruction and detector simulation using DELPHES. The EW production of ZZ bosons is searched in a subset of events that pass the ZZ selection, namely those that feature at least two jets separated from the leptons by $\Delta R > 0.4$. As we found in the WW channel, the resulting detector-level angular polarized distribution is strongly affected by the detector effects. The analysis-level cuts is described in this conference paper [91], to which we contributed for the kinematic reconstruction overview.

In the following, we shortly discuss on the main results in terms of NN ability to separate polarization components for the ZZ channel. The composition of the dataset here is different, because of some technicalities, the amount of



(a) NN score for $Z_L Z_L$ sample.

(b) NN score for $Z_L Z_T$ sample.

(c) NN score for $Z_T Z_T$ sample.

(d) ROC curves for the LL, LT and TT classes.

Figure 4.33: NN score obtained for each classes evaluating different fully-leptonic polarized channel obtained with DELPHES. Output distributions of the LL, LT, TT scores, for a DNN with 8 hidden layers composed of 200 neurons each and batch size of 64 and corresponding ROC Curve.

data used for training and test are lower with respect to the WW case and the dataset composition is not balanced. According to SM fractions, the LL component accounts for around 9% of the total cross-section and LT, TT for the remaining 91%.

Here we use a training dataset with enhanced longitudinal component, with 30% of the LL contribution. The training dataset is composed of roughly 100k detector-level events, and a corresponding dataset of the same size is used for the validation of the models. In order to test the performance of the DNN on a SM-like dataset, the evaluation instead has been performed on same size detector-level events with SM expected LL fraction. The input variables include all the available observables related to the final state objects, namely the four-momentum of leptons and tagging jets, together with constructed observables such as the Z bosons four-momenta and their invariant mass, m_{Z1} and m_{Z2}, the four-momentum of the ZZ system and its invariant mass m_{ZZ}, and

the dijet invariant mass m, for a total of 34 input variables.
Each polarization category is tagged in the same way as we did for the WW case with one-hot encoding technique. Regarding the ROC curve for the single classes with the optimized parameters shown in Fig. (4.32), we can draw same conclusion as we found for the charged boson case, we observe that the LL class has the best performance, followed by the TT class and finally the LT class which results to be more difficult to classify. This can be appreciated also in Fig. (4.33) where the LL and TT score distributions show a nice separation of the LL and TT components, while the LT component has less separation power than the other two.
If we look at the performance of the NN built for the ZZ detector level problem we see two main differences with respect to the WW generator level case. Here, we trained the network with a smaller, unbalanced dataset consisting of more variables. From a qualitative perspective we got same classification performance among the 3 classes but we made use of a deeper and wider network topology. This fact can be understood from the fact that the detector level dataset was smaller and it is intrinsically more complex.
With all the experience developed in managing DNN in so many different physics scenarios, we show in App. (D.2) how the number of epochs affects our performance comparing the loss function for training and validation datasets. We realized that the loss functions in some cases were flat, in some cases followed a negative exponential decay, in other cases they were made by a composition of step losses or we had even to increase the number of training epochs to reach a convergence. In particular, we did not use too wide NN but deep ones, suggesting the needs for a complex convergence structure related to the problem.

In this Section, a multiclassifier has been developed and optimized to obtain the best separation among polarized categories. We tested the DNN framework developed to discriminate among polarization states for the WW VBS channel from the previous Section, on a different one: the ZZ VBS channel. We went beyond parton level, by classifying detector level events.
Despite the different dimension and composition of the dataset with respect to the WW case, this categorization approach shows promising results, showing that the best performance is in the LL component classification. Of course, considering a practical usage of this framework, we should work on real data obtained after a clear background discrimination. Background to signal discrimination can be tackled using a DNN approach as well, so that the final procedure could be a sequence of DNN discriminator. We have different backgrounds that have to be discriminated from the VBS signal, so that a neural network can be trained separately for each background with respect of the signal and then combined together to provide a global discriminator. This study was out of scope of this work, even if a preliminary test on different machine learning and DNN techniques has been done before moving to the land of polarization discrimination.

4.4 Polarization Fraction Extraction

This last part of the present Chapter is dedicated to how the performance of our different NN experiments reflects into the extraction of polarization fractions. This is done by means of a maximum likelihood fit of the polarized templates to the full computation data.
As a matter of fact, event-by-event polarization is not well defined, the boson lives in a quantum superposition of different polarization states (longitudinal, transverse). The choice of the variable for the polarization shape analysis is more or less arbitrary. Basically, any variable for which transversely and longitudinally polarized samples populate different phase space regions is a candidate. The larger the statistical difference between the transverse and longitudinal components, the smaller will be final errors on the polarization fractions. In some cases, polarization sensitive variables are probably process dependent, like lepton angle in the W rest frame with respect to the W direction in the laboratory frame or $p_{T_{WW}}, R_{pT}$, see Fig. (4.34).
The full computation not only contains admixtures of the longitudinal and transverse polarization, but also interference among them and non-resonant contributions. Since the interference can be constructive as well as destructive, we adopted an approach in which all the non-purely longitudinal components (transverse polarization, interference and non-resonant contribution) are summed up and treated as a single distribution. They are obtained by sub-

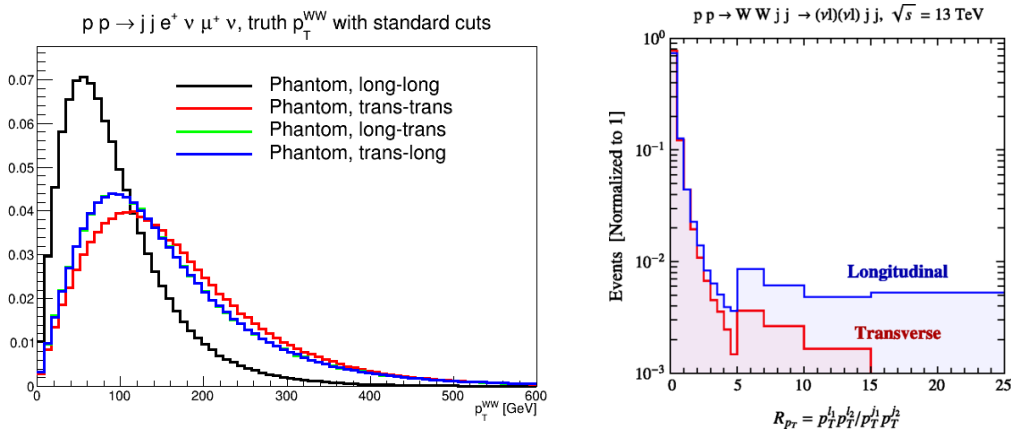


Figure 4.34: VBS Full-leptonic distribution of the kinematic variable p_T of WW (left) according to different WW polarizations. In black $p_{T_{WW}}$ of W_LW_L boson, in red W_TW_T, in green W_LW_T which is superimposed to the blue one with W_TW_L. VBS Full-leptonic distribution of the kinematic variable $R_{p_T} = p_T^{l1}p_T^{l2}/p_T^{j1}p_T^{j2}$ (right). In blue (red) the events with leptons coming from the decay of longitudinal (transverse) polarized W bosons (plot taken from [111]).

Table 4.3: Summary of the longitudinal polarization fraction extraction from the fit of the longitudinal (W_L) and background distributions to the full computation. The background is defined as the difference between the full computation with an unpolarized W boson and with non-resonant production modes included (Full comp.) and the longitudinal (W_L) contribution. The best fit value, for all the reconstruction methods, matches the truth polarization fraction, $\sigma_L/\sigma_\mathrm{full} = 0.257$, within the sub-percent level. CI widths at the 95 % CL for background and longitudinal polarization fraction are reported. The numbers are obtained at 139.0 fb^{-1} luminosity, which is the value of the integrated ATLAS luminosity for the 2015–2018 data taking period.

Type	Background	W_L
Random	0.122	0.119
Selection 1	0.137	0.135
Selection 2	0.125	0.122
Selection 3	0.130	0.128
Selection 4	0.098	0.095
Binary	0.105	0.102
Binary-LT	0.044	0.036
Regression	0.076	0.071
Truth	0.104	0.101

traction of the longitudinal distribution from the full computation. With this approach we avoid fitting histograms with the negative entries, which cannot be handled in the maximum likelihood fit. From here on, the mixture of these three components will be referred to as the *background* distribution.

In the fit, the full computation is hypothesized to be well described by the sum of the longitudinal and background distribution and its normalizations are

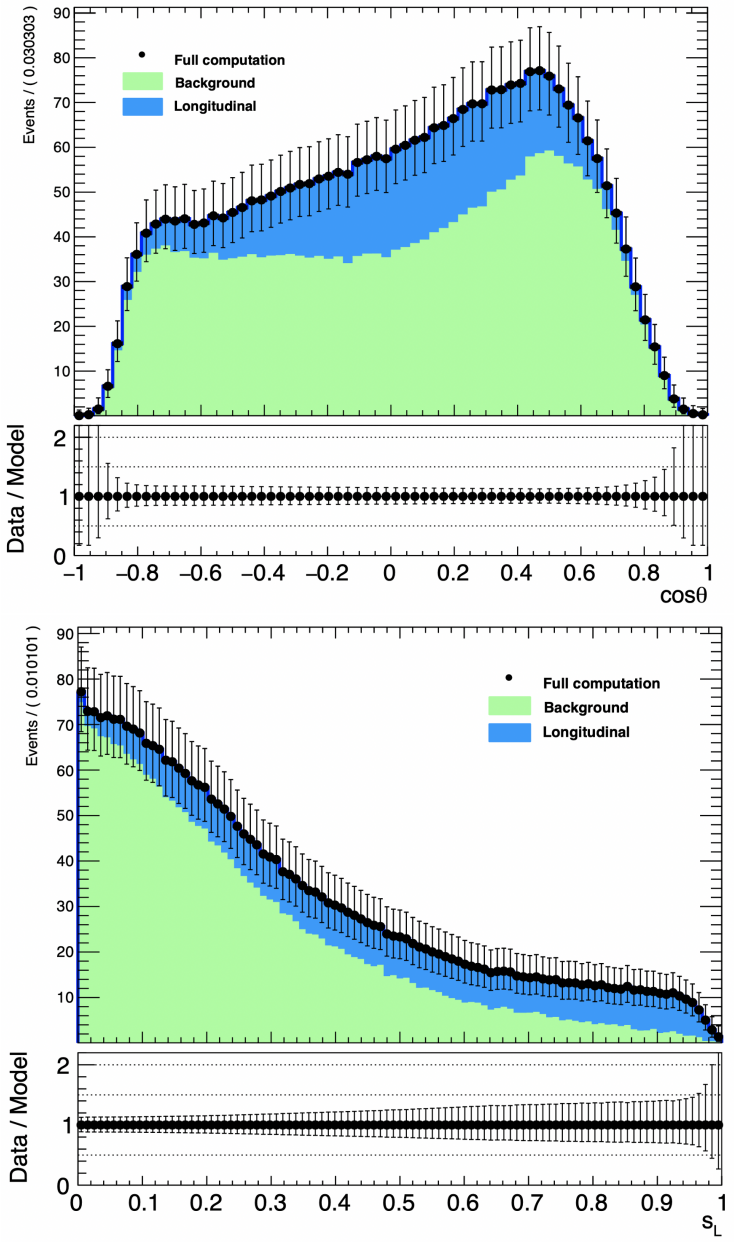


Figure 4.35: Top plot: background (green) and longitudinal (blue) cos ϑ distribution semileptonic polarization as obtained by NN. Bottom plot: background (green) and longitudinal (blue) score distribution of NN semileptonic polarization score classifier according to the two templates used in the fitting procedure to full computation sample (black dots). In both cases the lower panel shows the ratio of the data to model.

chosen to be the fit parameters. There is of course a very high anti-correlation between the fitted parameters because the benchmark is a direct sum of both. The latter has as a consequence also the exact convergence of the best fit values to the truth polarization fractions, as predicted by the generator.
This analysis takes into account only the statistics given by the MC sample produced but the histograms are normalized to 139.0 fb^{-1}, which is the value of the integrated luminosity collected by ATLAS during the 2015–2018 data taking period. Interesting values are the confidence intervals (CI) for longitudinal and background normalization. The resulting normalization CI widths at 95% CL using different sources of angular distributions and methodologies for the semileptonic channel (random solution choice, binary classification, binary polarization classifier (LT), regression and truth) are summarized in Table (4.3). The CI provided in this analysis are calculated as the product of the cross section ratio between the process considered (e.g. semileptonic or full leptonic) and the full computation contribution times the error given by the fit procedure. In the upper Fig. (4.35) we show fit results obtained in the semileptonic regression case.
Both approaches, analytical with binary classification and regression, show an improvement with respect to the random solution choice, as expected. An interesting observation emerges from the comparison of the fit performance employing different reconstruction techniques with the truth information. While the binary classifier very closely approaches the truth performance, the regression model even outperforms the fit using the truth distributions. What we might have considered as a *failure* of the DNN in reconstructing the truth angular distribution seems to encode some further important information for the distinction of the longitudinal polarization from the background, which might not be simply extracted from the truth angular distribution with respect to only one variable.

We summarize the performance of the different methods in Tables (4.4, 4.5) where RMS errors between the true and reconstructed $\cos\vartheta$ is used as a general metrics. DNN technique gives smaller values with respect to classic selection criteria, confirming the potential of the method according to different vector boson scattering processes.
In Fig. (4.35), in the bottom, we show fit results obtained in the semileptonic case where a binary classifier is built to directly discriminate longitudinal versus transverse polarization components (score variable). Different training has been performed on the admixture of the longitudinal and transverse polarization and the output score (s_L) can be interpreted as the event longitudinal polarization fraction. The error bars shown in the plot are related to the uncertainty provided by the fit procedure where no measurement effects are included and they represents the statistics uncertainty of the chosen integrated luminosity. The capability here is to discriminate the blu distribution from the green one. The fit extraction is almost two times narrower (CI at 95% see Table 4.3)

Table 4.4: RMSE between the the true and reconstructed $\cos\vartheta$, calculated as a quantitative parameter to evaluate the performance of applied selection-criteria and DNN techniques in the semi-leptonic channel. We evaluate RMSE on samples containing a purely longitudinally polarized (W_{L}) leptonically decaying W boson, a purely transversely polarized (W_{T}) boson and an unpolarized W boson generated in the OSP PHANTOM framework (Un. OSP). RMSE is equal to 0 in the case of truth.

	W_{L}	W_{T}	Un. OSP
Truth	0.0	0.0	0.0
Random	0.440	0.347	0.359
Selection 1	0.420	0.303	0.328
Selection 2	0.401	0.269	0.305
Selection 3	0.390	0.274	0.305
Selection 4	0.370	0.228	0.275
Binary	0.357	0.241	0.270
Regression	0.277	0.200	0.217

Table 4.5: RMSE between the the true and reconstructed $\cos\vartheta$ calculated as a quantitative parameter to evaluate the performance of applied DNN techniques. We are comparing neural networks trained in the fully-leptonic channel. Models are trained and evaluated on samples of purely longitudinally polarized (W_{L}) and a purely transversely polarized (W_{T}) leptonically decaying W bosons and then direct and indirect approaches are used, with an optional inclusion of the MAOS and MT2 variables in the training, as described in the text. RMSE is equal to 0 in the case of truth.

DNN type	W_{L}	W_{T}
direct	0.291	0.262
direct+MAOS	0.291	0.261
indirect	0.528	0.605
indirect+MAOS	0.436	0.493

with respect to the confidence level obtained in case of regression technique ($\cos\vartheta$).
Similarly, we show in Fig. (4.36) a comparison of fit result obtained in the fully-leptonic channel. It is evident that the error bars are much larger than in the previous semileptonic case. Here we trained our NN with an admixture of pure longitudinal (LL), pure transverse (TT) and mix polarization (TL) sample. On top we show the fit result coming from a regression performed on this dataset (direct approach defined before) whereas at the bottom we can appreciate the fit distribution coming from a multilabel classifier. In this case the output score (s_L) can be interpreted as the event pure longitudinal polarization fraction.

Table 4.6: Summary of the longitudinal polarization fraction extraction from the fit of pure longitudinal ($W_{\mathrm{L}}W_{\mathrm{L}}$) and background distributions to the full computation in two different scenario: regression and multilabel classifier. The background is defined as the difference between the full computation with two unpolarized W boson and with non-resonant production modes included (Full comp.) and pure longitudinal ($W_{\mathrm{L}}W_{\mathrm{L}}$) contribution. The best fit value, for all the reconstruction methods, matches the truth polarization fraction, $\sigma_L/\sigma_{\mathrm{full}} = 0.072$, within the sub-percent level. CI widths at the 95 % CL for background and longitudinal polarization fraction are reported. The numbers are obtained at 139.0 fb^{-1} luminosity, which is the value of the integrated ATLAS luminosity for the 2015–2018 data taking period.

Type	Background	$W_{\mathrm{L}}W_{\mathrm{L}}$
Run 2 Multilabel	0.243	0.206
Run 2 Regression	0.324	0.294

As we can appreciate in Table (4.6), the CI widths are rather large, despite the fact that for example LL fraction[9] is 0.072, which implies that the LL polarization fraction in the fully-leptonic channel, unless we find a better variable than the classifier score, would not be measurable at the current ATLAS luminosity. The widths are higher with respect to semi-leptonic channel mainly due to smaller cross-section.

For these reasons, we wanted to push this analysis using the predicted data for LHC Run 4[10] and determine the discovery potential related to these two polarization determination approaches. The result is shown in Fig. (4.37) where, comparing it with Fig. (4.36), one can see a clear decrease of the error bars related to narrower CIs, that for the longitudinal components are quantified in 0.089 (regression) and 0.063 (multilabel). For sake of comparison, current VBS WW same sign published results [93] give us an estimation of systematics and statistical uncertainties. The measured fiducial signal cross section is $\sigma_{fid} = 2.89^{+0.51}_{-0.48}(stat)^{+0.24}_{-0.22}(exp\,syst)^{+0.24}_{-0.22}(mod\,syst)^{+0.08}_{-0.06}(lumi)\,fb$ where the uncertainties correspond to the statistical, experimental systematic, theory modelling systematic, and luminosity uncertainties, respectively. Having a pure VBS signal is already very challenging: the uncertainties in the QCD background normalization and the jet energy scale are the dominant systematic uncertainties in the measurement, therefore misidentification of leptons and in the integrated luminosity determination should be considered. The current analyses are, at the time being, mainly limited by the large statistical uncertainty.

[9] Single polarize fraction is given by $\sigma_{LL}/\sigma_{fullcomp}$.

[10] Definition of different LHC Run and their evolution for the HL-LHC scenario can be found in `https://www.desy.de/~ameyer/trento_hllhc_190909.pdf`.

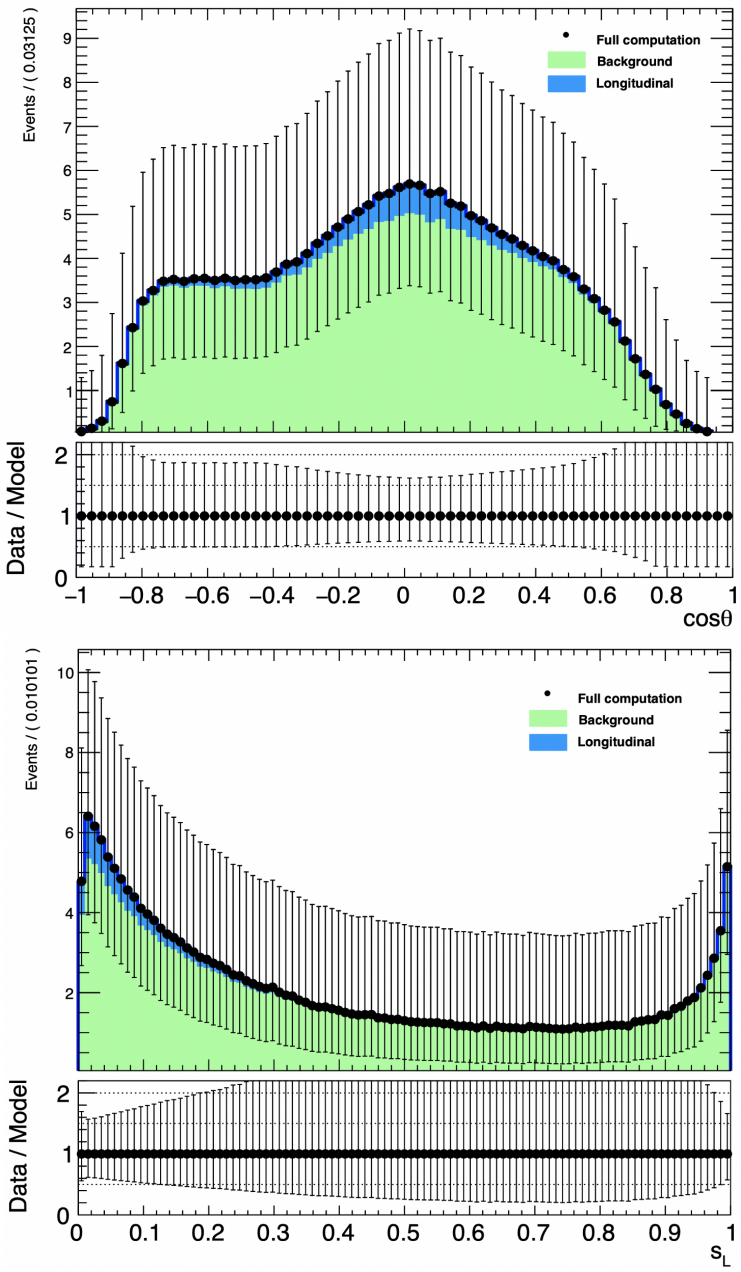


Figure 4.36: Top figure: background (green) and longitudinal (blue) $\cos\vartheta$ distribution of NN fully-leptonic polarization. Bottom figure: Background (green) and longitudinal (blue) score distribution of NN full leptonic polarization multilabel classifier according to the two templates used in the fitting procedure to full computation sample (black dots). In both cases the lower panel shows the ratio of the data to model.

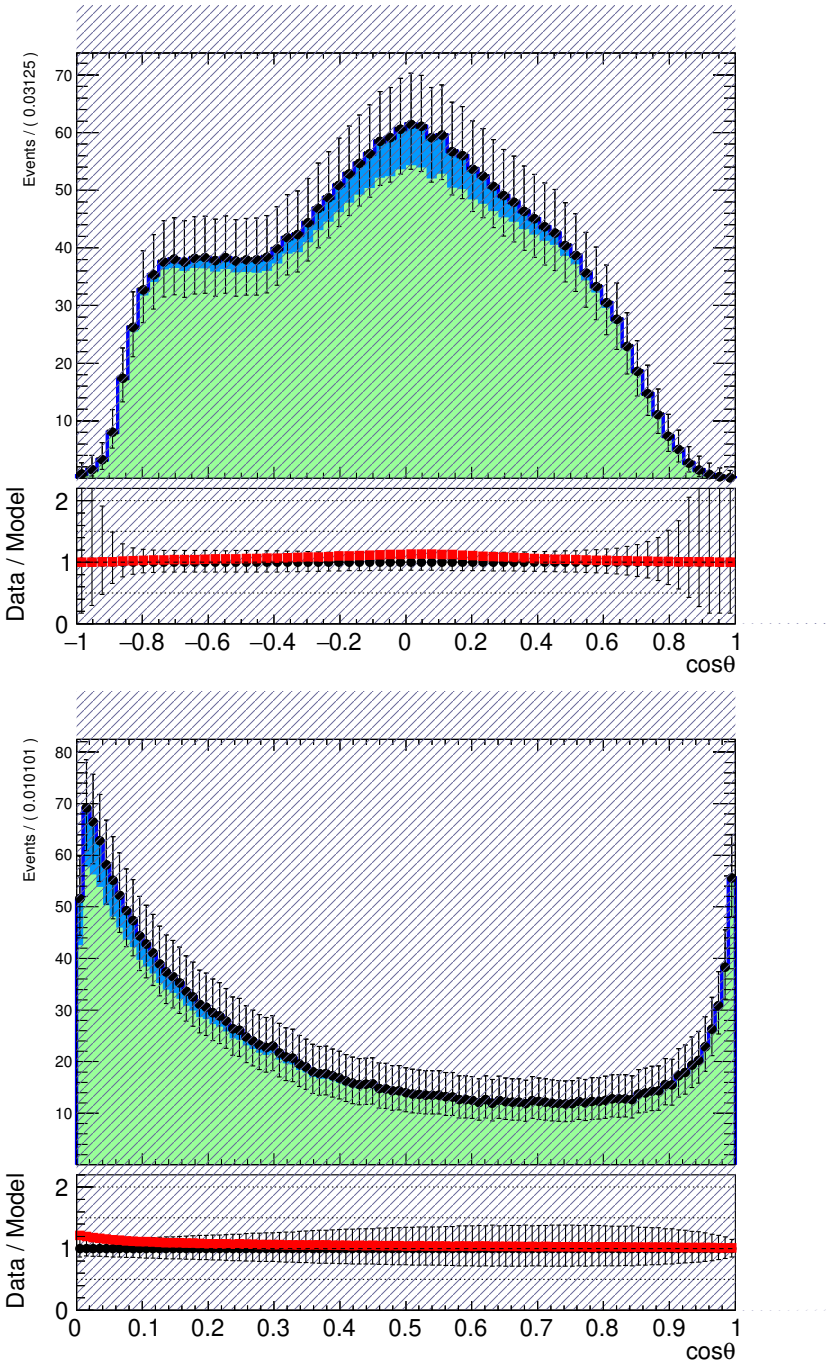


Figure 4.37: Top plot: background (green) and longitudinal (blue) $\cos\vartheta$ (score) distribution of NN fully-leptonic polarization. Bottom plot: Background (green) and longitudinal (blue) score distribution of NN full leptonic polarization multilabel classifier according to the two templates used in the fitting procedure to full computation sample (black dots). The numbers are obtained for a projection at 1500.0 fb^{-1} luminosity, which is the foreseen value of the integrated ATLAS luminosity for the 2026–2030 data taking period. In both cases the lower panel shows the ratio of the data to model. Red line represents the ratio between the blue and green contribution.

Key Results

We transformed the performance of our different NN experiments into the extraction of polarization fractions by fitting polarized templates to the data.
NN categorization approach shows promising results, showing that the best performance is in the LL component classification.
According to fit extraction, the score s_L seems to be a candidate for a more optimal variable for the fit of the polarization fractions both in the semileptonic channel and in fully-leptonic channel.
This FIT analysis represents a first step into the discrimination of the polarization components, mainly the longitudinal one. A further step would be to include systematics effects which affect any signal to background discrimination, both in simulated and real data. At the current ATLAS luminosity, LL polarization fraction would not be measurable yet, because of the limited statistics. However, an improvement of the fit result can be appreciated using the expected Run 4 integrated luminosity estimated at 1500.0 fb^{-1}.

In the next Chapter we introduce a new paradigm in the computation: the *Quantum Computing*. We will try to use quantum algorithms as an additional technique to solve VBS classification problem addressed here.

Chapter 5

Quantum Computing and VBS

This final Chapter is dedicated to a pioneering attempt of approaching the W polarization discrimination, using a new technology paradigm that is becoming of great importance in several scientific fields: *Quantum Computing* (QC). The study of this emerging technology and of its application in the High Energy Physics sector covered a significative fraction of my PhD.
We begin by introducing Quantum Computing and its main characteristics in Sec. (5.1), while in Sec. (5.2) we focus on its intersection with Machine Learning field, by investigating some Quantum Machine Learning strategies. In Sec. (5.3), we discuss the application of QC to VBS, comparing classical and quantum algorithm performances as a natural alternative to DNN methods. We close the Chapter with a review of the main present and future applications of QC in HEP, focusing not only on quantum machine learning, but also on the way this new paradigm could revolutionize the field, from group theory to detector simulation.

5.1 Quantum Computing in a Nutshell

Quantum mechanics is the gateway towards novel and potentially disruptive approaches to scientific and technical computing. The origin of quantum computing is always referred to a famous statement from Richard Feynman: *Nature isn't classical, . . . , and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy* [112].
Several years of theoretical and experimental efforts by the scientific community has been devoted to the formulation of this new paradigm.

Quantum mechanical models and their simulation are often conceptually troublesome and technically demanding. We will highlight the main aspect of this discipline, but we point to dedicate books or reviews (as [113]) for a complete and exhausting introduction.
Classical computation is becoming more and more extreme and the approxi-

Figure 5.1: Image credits: `https://qiskit.org/documentation/qc_intro.html`.

mations given by classical computing machines, once acceptable, now become sources of insurmountable errors. The cost and space effects of minimisation of microprocessor components to reach such extreme performance, described by the well-known *Moore's law*, is approaching its physical saturation [114]. Digital information is stored on devices magnetically, as known, using a series of ones and zeros. The trade-off between density and speed reached a level in which classical determinism of these devices is no longer guarantee. Indeed, the failure of classical simulation methods is especially manifest when strong quantum correlations between subsystems become dominant in determining the properties of the object under investigation, which is indeed the case in many interesting and generally open problems. Hence, the need of quantum computation is manifested, despite the many alternative methods and tools which have been developed over the last century such as quantum Montecarlo [115] or tensor networks [116].
Quantum Computers offer an essential speed advantage over classical computers. In particular, the idea of efficient and inefficient algorithms was made mathematically precise by the field of computational complexity. Roughly speaking, an efficient algorithm is one which runs in time polynomial in the size of the problem solved. In contrast, an inefficient algorithm requires super-polynomial (typically exponential) time. QC are expected to solve unsolvable classical problem. When and if this will occur, we could claim the so called *quantum advantage*. But quantum advantage can manifest itself also in solution of problems which are classically unsolvable.

Quantum computing were formally defined for the first time by David Deutsch in his 1985 seminal paper [117], where he introduced the notion of quantum Turing machine, a universal quantum computer based on qubits and quantum circuits. Here, we focus on digital quantum simulators that are programmable and general purpose quantum devices, which promise a larger flex-

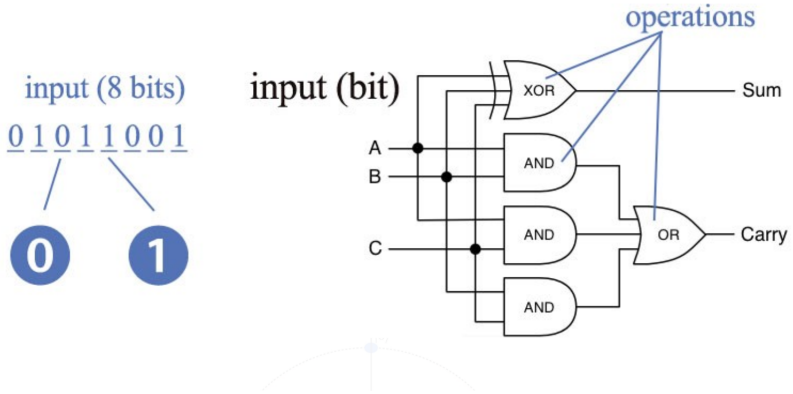


Figure 5.2: Representation of a classic circuit, with standard bit notation.

ibility on the models to be solved. Strictly speaking, Quantum Computer is a quantum machines satisfying the known DiVincenzo criteria [118] for universal quantum computation. Those are five criteria that any candidate quantum computer implementation must satisfy. Namely, the device must have well-defined qubits (a two-level quantum system), they should be initialized to pure states, and the complete set of available operations (gates) must be complete. The machine must allow single qubits measurements and allows long coherence time[1].

5.1.1 Quantum Bit

Bits are designed to be the world simplest alphabet. With only two characters, 0 and 1, we can classically code any piece of information. This is how all information is represented in (classic) computers. Whether numbers, letters, images, or sound, all exists in the form of binary strings.
Computation can be thought as a diagram, or the simplest programs with very few bits, it is useful to represent this process in a diagram known as a *circuit diagram*. This has inputs on the left, outputs on the right, and operations represented by arcane symbols in between, as shown, for instance, in Fig. (5.2)[2]. In QC, these operations are called *gates* and they represent the effective algebra of quantum registers made of qubits. Qubits (quantum-bit) are our basic variable in quantum computers: a quantum variant of the bit. These have exactly the same restrictions as normal bits: they can store only a single binary piece of information, and can only ever give us an output between 0 or 1. However, they can be manipulated in ways that can only be described by quantum mechanics.

Just as a classical bit has a state – either 0 or 1 – a qubit also has a state. Two possible states for a qubit are the states $|0>$ and $|1>$ where we are using

[1]The coherence time summarizes many aspects of qubit state degradation.

[2]Image credits [https://jonathan-hui.medium.com/qc-control-quantum-computing-with-unitary-operators-interference-entanglement]

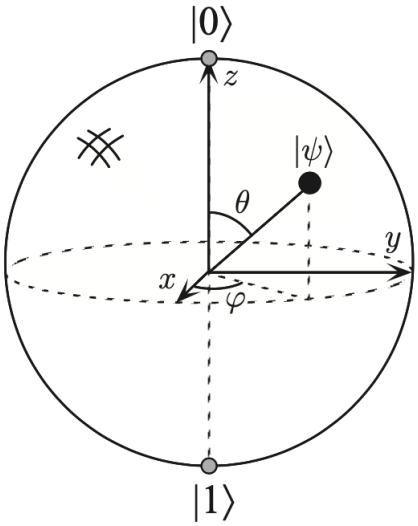


Figure 5.3: Bloch sphere representation of a qubit.

the typical (quantum) notation called *Dirac notation*. A qubit is a vector in a two-dimensional complex vector space and, differently from an ordinary classical bit, can be $|0>$ and $|1>$, and in any linear combinations of these two states, often called *superpositions*:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle;$$

where α, β are complex numbers. The representation above of a general state $|\psi\rangle$ is referred as *computational basis* and $|0\rangle$ and $|1\rangle$ form an orthonormal basis for this vector space. Superposition is the first quantum mechanics feature that is exploited by QC. The other two are *entanglement* and *interference*, as it will be described in the following.
According to quantum mechanics, when we measure a qubit, we cannot determine its quantum state but we have access to either the result 0 or 1 with the probabilities normalization constraint that $|\alpha^2| + |\beta^2| = 1$.
An useful way of describing a qubit state is the following geometric representation.

$$|\psi\rangle = e^{i\gamma}\left(\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle\right) \tag{5.1}$$

where θ, ϕ and γ are real numbers. The factor $e^{i\gamma}$ is an overall phase and it has no observable effects, therefore it is normally omitted in usual formulation.

A very common representation of a qubit is the one represented in Fig. (5.3) which goes under the name of *Bloch Sphere*. The state $|0\rangle$ is placed at the North Pole, while the state $|1\rangle$ occupies the South pole. Any other state $|\psi\rangle$ lives as a linear combination of $|0\rangle$ and $|1\rangle$ on the Sphere surface and can be represented by the polar spherical coordinates θ and ϕ of Eq. (5.1), with $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$.

Hilbert space is a big place, and we can explore it by considering multi-qubit states. A two qubit system, has four computational basis states denoted $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. This system can exist in superpositions of them, with complex coefficients α – sometimes called *amplitudes* – so that the state vector describing the two qubits is typically denoted by:

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_1|11\rangle. \tag{5.2}$$

More generally, we may consider a system of n qubits. The computational basis states of this system are of the form $|x_1x_2...x_n\rangle$ and so a quantum state of such a system is specified by 2^n amplitudes. Trying to store all these complex numbers (e.g. $n > 60$) would not be possible on any conceivable classical computer [113]. This enormous potential computational power is something we would very much like to fully exploit it.

5.1.2 Quantum Circuits

Single Qubit Operations

Analogously to classical computer, which is built by electrical circuits containing wires and logic gates, a Quantum Computer is built from quantum circuits containing wires and elementary quantum gates (which are the analogous of classical operations), to carry around and manipulate the quantum information. Quantum gates on a single qubit can be described by 2×2 matrices U. As an example, we list two common quantum gates acting on each single qubit:

$$\text{Z-Gate:} \qquad Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{5.3}$$

which leaves $|0\rangle$ unchanged, and flips the sign of $|1\rangle \rightarrow -|1\rangle$;

$$\textit{Hadamard} \text{ gate:} \qquad H \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{5.4}$$

This last gate creates a superposition states:

$$H|0\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle) \text{ and } H|1\rangle = 1/\sqrt{2}(|0\rangle - |1\rangle),$$

which are referred as $|+\rangle$ and $|-\rangle$ respectively, and form themselves an orthogonal bases.
The most general single qubit $SU(2)$ operation has the form:

$$U(\theta, \phi, \lambda) = \begin{bmatrix} \cos(\theta/2) & -e^{i\lambda}\sin(\theta/2) \\ e^{i\phi}\sin(\theta/2) & e^{i\lambda+\phi}\cos(\theta/2) \end{bmatrix} \tag{5.5}$$

which represents a single-qubit arbitrary $U(\theta, \phi, \lambda)$ rotations via the identity:

$$U(\theta, \phi, \lambda) = R_z(\phi)R_x(\theta)R_z(\lambda)$$

where, using the Pauli matrices notation, we define rotations as:

$$R_\alpha(\theta) = \exp(-i\frac{\theta}{2}\sigma_\alpha) \qquad \alpha = x, y, z, \tag{5.6}$$

performed araound each of the three axes of the Bloch Sphere. One can easily verify that $R_Z(\pi) = Z$, the Z-gate of Eq. (5.3).

Two or more qubits

Single qubits are interesting, but individually they offer no computational advantage. In Eq. (5.2) we represented a multiple qubits state. We can generalize the way we operate on multi-qubit states using the tensor product to calculate matrices that act on these statevectors. We can represent, for instance, a simultaneous operation like:

$$X \otimes H = \begin{bmatrix} 0 & H \\ H & 0 \end{bmatrix} \tag{5.7}$$

where H represents the Hadamard gate introduced in Eq. (5.4). From a matrix perspective, a 2-qubits gate is given by a 4×4 matrix acting con the $\mathbb{C}^4$ computational basis.
Computational advantage is reached when qubits interact with each other. To make this happen we introduce the `CNOT` gate which creates *entangled states*. Given a two input qubits, this gate is a conditional gate (in analogy to XOR gate in the classic case) that performs an X-gate on the second qubit (*target*), if the state of the first qubit (*control*) is $|1\rangle$. Any multiple qubit logic gate, for example, in superconducting qubits connected via cross-resonance interactions, may be composed from `CNOT` and single qubit gates. The `CNOT` matrix representation is:

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5.8}$$

depending on which qubit is the control and which is the target. Different books, simulators and papers order their qubits differently. This matrix swaps the amplitudes of $|01\rangle$ to $|11\rangle$, as can be easily checked.
Here, we see at work the second fundamental property of quantum computing always derived from coming from quantum mechanics: the *entanglement*. Via application of `CNOT`, we create entangled states where two subspaces are coupled in a non-classical way. Entanglement induces a change in the state of the controlling qubit; correlations of measurements performed independently on the two qubits show outcomes that cannot be explained classically. In particular, an entangled state provides a detailed description of the system and no further information can be obtained about the system.

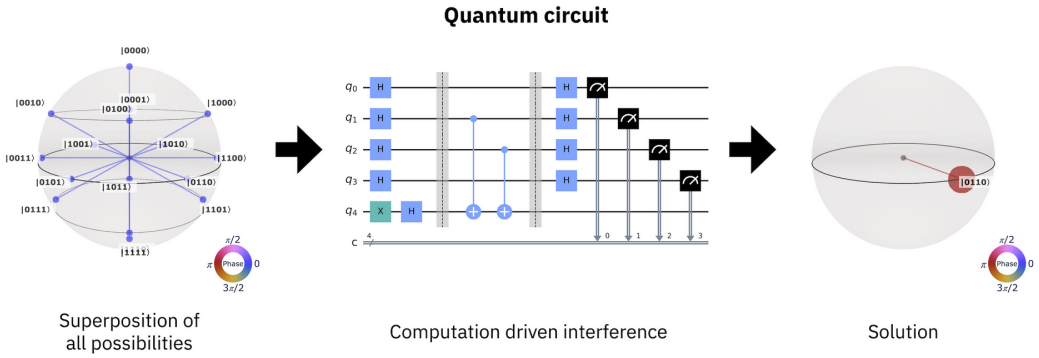


Figure 5.4: Representation (QSphere) of 5-qubits superposition states (left), quantum circuit representation with single (X, H), 2-qubits gate (`CNOT`) and measurement operations (center), QSphere visualization of the final state obtained from the circuit (right) [`https://qiskit.org/documentation/qc_intro.html`].

A peculiar property of quantum computing is that unitary quantum gates are *always* time reversible, since the inverse of a unitary matrix is also a unitary matrix, and thus a quantum gate can always be inverted by another quantum gate. Understanding how to do classical logic in this *reversible* or invertible sense will be a crucial step in understanding how to harness the power of quantum mechanics for computation. More generally, an arbitrary quantum computation on any number of qubits can be generated by a finite set of gates that is said to be *universal* for quantum computation.

In the final step of any simulation run, measurements are applied to retrieve information. To find the probability of measuring a state $|\psi\rangle$ in the state $|z\rangle$, according to quantum mechanics, we calculate the following probability:

$$p(|z\rangle) = |\langle z|\psi\rangle|^2. \tag{5.9}$$

Here we have the so called collapse of the state of the qubit. The amplitudes contain information about the probability of us finding the qubit in a specific state, but once we have measured the qubit, we know with certainty what the state of the qubit is. This readout procedure is done, for the IBM system, for instance, in the computational basis, i.e. we are given the occupation probability of the eigenstates of σ_z for each qubit. All the components described so far are shown in Fig. (5.4): starting with 5 qubits, we apply successive single gates to create superposition of all possibilities, then we create entangled states between $q1 - q4$ and $q2 - q4$. In the third part then, we apply again Hadamard gates and finally the measurement operations are put on the first 4 qubits and we visualize the result on the QSphere on the right as a solution to the problem.

We define *quantum algorithm* the application of a finite number of quantum gates (unitary operations) to an initial quantum state and the following measurement of the expectation value for the final state in a given basis. There

are a growing number of quantum algorithms[3], among these we recall the most famous ones: Shor's algorithm [119] to factor integers exponentially faster than our best classical algorithm, Grover's algorithm [120] invented in 1996 to search an element in an unstructured database with a quadratic speed-up, the Harrow-Hassidim-Lloyd (HHL) algorithm [121], invented in 2008 to solve linear systems of equations to quantum phase estimation (QPE) algorithm [113] to estimate the phase (or eigenvalue) of an eigenvector of a unitary operator.

5.1.3 IBM Quantum Computing

An important aspect is that all of these algorithms require large-scale fault-tolerant quantum computers to be useful, while current and near-term quantum devices are characterized by at least three major drawbacks [122]:

- *Noise*: the coherence time (lifetime) of a qubit and the fidelity of each gate (accuracy of the computation) are currently very low in all devices due to the interaction of each qubit with its surrounding environment, limiting the depth of practical quantum circuits that can be run on current machines.

- Small number of qubits: most near-term quantum computers have between 50 and 100 qubits, which is not enough for traditional algorithms such as Shor's or Grover's to achieve a quantum advantage over classical algorithms.

- Low connectivity: qubits disposal on the chip cannot be all-to-all, and usually only nearest-neighbors can interact. While it is theoretically possible to run any algorithm on a device with limited connectivity —by "swapping" quantum states from qubit to qubit with a good transpiling software routine—the quantum advantage of some algorithms can be lost due to the dimension of the resulting circuit.

For these reasons, a new class of algorithms, the so-called Near-term Intermediate-Scale Quantum (NISQ) algorithms [123], have started to emerge. They are based on the concept of *variational circuits* which are fixed-size circuits with variable parameters that can be optimized to solve a given task. Performance of a QC is measured with a metrics called Quantum Volume (QV) [124]. This metric takes into account a combination of different elements, from hardware related errors (cross-talks, decoherence) and software ability to compile the circuit, transpiling logical gates into physical native gates. Just as Moore's law was used to track the performance of classical systems, Quantum Volume is how the performance of quantum computers is tracked and IBM is on the path of doubling its Quantum Volume every year.

[3]See for example `https://quantumalgorithmzoo.org/`.

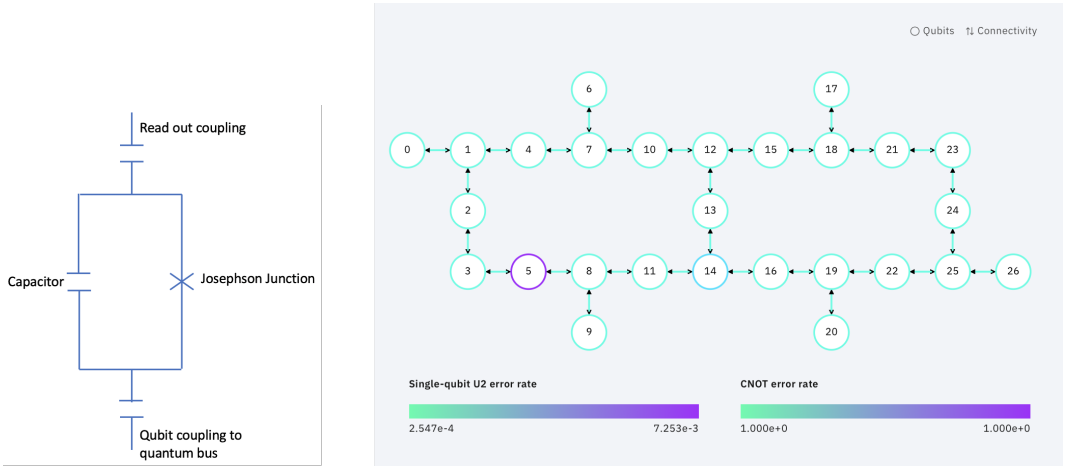


Figure 5.5: (left) Schematic of a quibit, LC-like circuit [125], (right) hexagonal qubit arrangement for IBM Quantum Falcon processor [*ibmq_paris* v1.3.6, IBM Quantum team. Retrieved from `https://quantum-computing.ibm.com` (2020)].

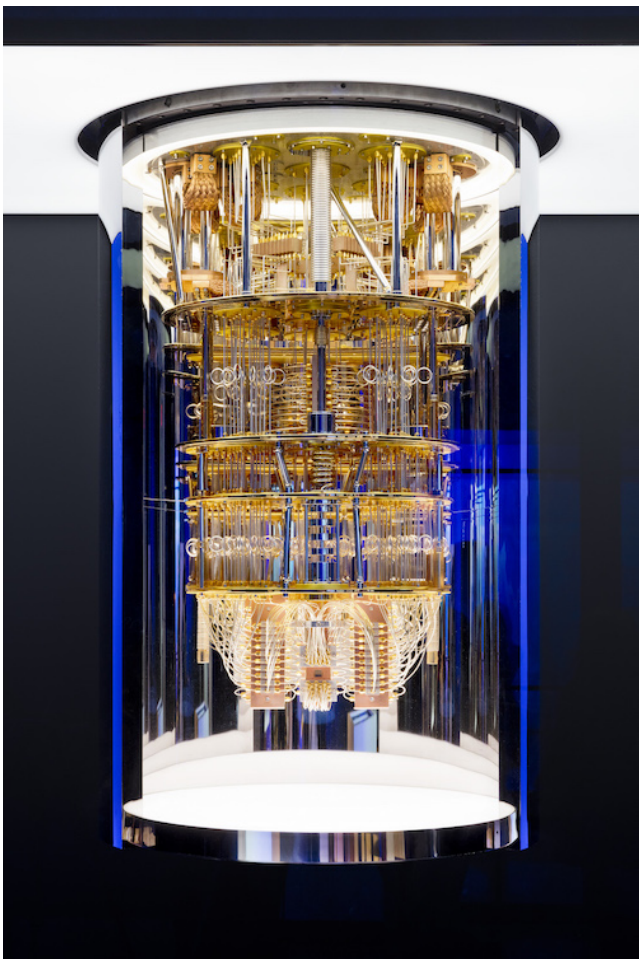
Hardware

How qubits are physically made and connected is out of scope of this work, however here below we review the main hardware realizations of a quantum computer. Some of the more common qubit platforms are neutral atoms, Quantum dots, Nitrogen-vacancy (NV) centers in diamond, trapped ions, and superconducting qubits. Out of these platforms, it is the superconducting qubits that are used by IBM and other big player in this Sector [125]. A fair comparison of performances between competing platforms can be found in [126, 127].

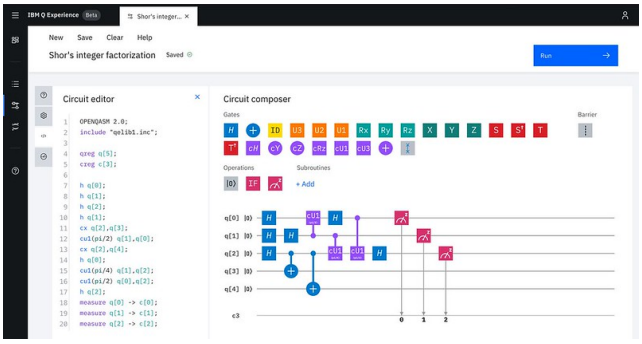
In the following we focus on the IBM Quantum Computing approach, based on superconducting technology that belongs to the class of Noisy Intermediate Scale Quantum prototypes. Superconductors that are made up of a combination of niobium and aluminium are at the base of the qubit. There, they are used as the basic charge carriers that comprise a pair of electrons, more commonly referred to as Cooper pairs. This is different from other conductors that generally use single electrons [128]. At the core of the qubit, we have a capacitor and Josephson Junction (very similar to an Inductance Capacitance (LC) circuit) which is an anharmonic oscillator. This allows us to easily differentiate between the different energy levels (DiVincenzo's criterion), which we map as the states of the qubit. On the ends of the qubit, there are two external coupling points, the top of which is the read out (or read-in) resonator. This resonator is used to perform operations on the qubit from your quantum circuit and read out the measured result of the qubit when you wish to obtain the collapsed result (0 or 1). At the other end of the qubit is the coupling to a neighbouring qubit, which is used to connect qubits together to create connectivity between them. The layout of the qubits can be set up in a configuration that can vary from one device to another, see for instance

Fig. (5.5). Here we can appreciate the topology and the quality of a 27-qubits IBM Q Device, each qubit and connection among them is characterized by a colour which corresponds to specific calibration parameters as can be seen in Fig. (5.5). Error rates ranges are illustrated by single-qubit $U3$ error rate (colored bar on the left), and `CNOT` error rate (colored bar on the right) for each device that is calibrated at least once a day or so, and each time they are calibrated, they calculate the average error rates. Each qubit has its own read out coupling so that we can perform operations on individual qubits (DiVincenzo's criterion). They also have their own read out couplings because they include bus couplings to complete the full connection space of the qubits.
IBM was the Tech company that firstly made available a quantum computing platform for free cloud access in 2016. At the time of writing this thesis, 29 different quantum chips are available, with different size and performance ranging from 5 to 65 qubits. The platform has the potential for implementing universal quantum computation because of universal sets of single- and two-qubit gates. The platform is called IBM Quantum Experience and it can be accessed at `https://quantum-computing.ibm.com`, while `Qiskit` documentation and updates are hosted at `https://qiskit.org/`. Users can submit quantum circuits to be run by using graphical interfaces and the `Qiskit` python development kit [129].

Over the last 5 years, the number of publications exploiting quantum computing in several application fields has grown exponentially. The following are some examples that go from quantum chemistry and material sciences [130, 131], to quantum field theories [132, 133], high energy physics [134] and dark matter [135], from the dynamics of many-body quantum systems [136] to imaginary-time evolution [137, 138]. Other contributions can be found for lattice simulation [139, 140], genetic algorithms [141] and variational time evolution approaches directly inspired by the action minimization principle [142]. Lots of different publications are defined within real use cases belonging to different industries, including finance, material science and optimization [143, 144, 145]. I gave my personal contribution to this publication [146], reproducing from quantum computed data, the cross section for four-dimensional inelastic neutron scattering, a common tool employed in the analysis of molecular magnetic clusters. Through an innovative combination of quantum algorithms with classical post-processing and error mitigation protocols, we demonstrate in practice full digital quantum simulation of spin-spin dynamical correlation functions, reporting experimental results obtained on IBM Q devices.
At the time of writing, there is no evidence of *quantum advantage* in real application, except ad-hoc situation [147, 148]. With this naming convention we refer to problems where classical computation is really overwhelmed.



(a) A view inside the IBM Quantum System One.



(b) Screenshot of the IBM Quantum Experience platform.

Figure 5.6: Different elements of IBM Quantum Computer, hardware and software components taken from `https://www.flickr.com/photos/ibm_research_zurich/albums/72157663611181258`.

5.2 Quantum Machine Learning

Quantum Machine Learning (QML) is a discipline which aims at establishing a productive interplay between the parallel revolutions brought by quantum

computing and artificial intelligence.
The most natural point of contact between quantum information processing and machine learning is probably linear algebra. Indeed, many of the most demanding protocols in artificial intelligence and data analysis feature matrix manipulations in high-dimensional vector spaces, and they are often identified with the acronym q-BLAS [149], for quantum Basic Linear Algebra Subroutines. In many cases, there exist theoretical proofs showing an exponential advantage of q-BLAS procedures over classical counterparts.
As described in the previous Chapter, classification of data is one of the typical tasks that supervised and unsupervised machine learning systems are programmed for. Several quantum classifiers have been recently proposed and tested [150, 151, 152, 153, 154, 155], relying on different forms of encoding and data processing and sometimes mimicking different classical machine learning algorithms. Oracle-based quantum classification has also been put forward as a possible route to the demonstration of quantum advantage on noisy quantum processors [156]. Estimating distances between feature vectors, thus translating in mathematical terms the common wisdom that similar properties are a proxy for group identity, is at the basis of some simple yet effective classification algorithms. For example, the k-nearest neighbours method [157] where efficient evaluations of vector distances is evaluated as an algebraic problem, calculating the fidelity between two quantum states. In [158], a quadratic speedup with respect to classical algorithms in computing the inner product of two vectors is reported. Quantum-enhanced support vector machines, that will be widely analyzed in the following paragraph, has been proposed in [153]: an extension of its classic version (SVM) method that searches for the best separating hyperplane discriminating between two class regions and acting as a boundary for classification of unknown inputs.

Research on quantum computing in the NN field is also very active, lot of solutions have been proposed with the hope of recognizing and generating atypical data patterns. In Chapter 4, we presented and applied FNN to the physical use case under investigation across this work. Interestingly, there are efforts to achieve fully trainable quantum feed-forward neural networks [159] which starts from their classical implementation. The effective practical potential of quantum neural networks, however, is still subject of ongoing research. In the following, we shortly discuss some really interesting proposal presented in [160] and further developed in [161, 162]. The authors start from the simplest model of an artificial neuron, the so called "perceptron" and design a quantum procedure to closely mimic the functionality of a binary valued artificial neuron. They implement an original procedure to generate multipartite entangled states based on quantum information principles that allows to crucially scale down the quantum computational resources.
In the second work [161], they move from single layer artificial neuron design to complete feed-forward networks, with several copies of the quantum register, implementing the quantum artificial neuron. By measuring the output

layer only, the activation state of the network can be assessed, and several computational tasks, that are in principle inaccessible to a single neuron, can be performed.
A further development, in the third work [162], generalizes that model to the case of a quantum circuit accepting also continuously valued input vectors. Continuous inputs are not possible in conventional, digital computers, where they are usually rendered by using bit strings. This is a model for a continuous quantum neuron which can be used for pattern recognition in grey scale images without the need to increase the number of qubits to be employed.

QML still suffers from a series of bottlenecks and caveats that need to be overcome. One of the dominating factor is the exponential costs for loading classical data onto quantum registers and the equally hard task of extracting the processed information stored in quantum states. As a matter of fact, the true power of QML may ultimately reside in the analysis of quantum data, which is by definition a difficult or inaccessible task for classical systems.

Understanding expressibility and trainability of quantum models-and quantum neural networks has been clearly addressed in [163], where the authors show that quantum neural networks are able to achieve a significantly better effective dimension than comparable classical neural networks. Defining a notion of *expressibility* for both classical and quantum NN, the latter offers an advantage over classical neural networks through a higher effective dimension and faster training ability.

5.2.1 From Classic to Quantum Classification

An interesting question is whether there are ways to apply quantum machine learning to High Energy Physics problems. Quantum algorithms have recently been proposed to tackle the computational challenges faced in particle physics data processing and analysis. In this paragraph we provide a detailed description about the possibility to use quantum algorithm to solve classification tasks.

As mentioned before, in the context of classification, a very common technique called *support vector machine* is used. Given a dataset with two or more classes, this supervised classification method is used to find an optimal *separating hyperplane* among those classes. In a p-dimensional space, a hyperplane is a flat affine subspace of dimension $p-1$. In two dimension a hyperplane is a one-dimensional subspace (a line). The mathematical definition of a hyperplane is quite simple. In two dimensions, a hyperplane is defined by the equation:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0; \tag{5.10}$$

where $\mathbf{X} = (X_1, X_2)^T$ is a point on the hyperplane and β_i are free parameters. Given a new test point to be classified, placing $\mathbf{X_T}$ into Eq. (5.10), we can verify to which side of the hyperplane it belongs (e.g. if Eq. (5.10)> 0 or < 0). This approach can be easily generalized for a $n \times p$ data matrix $\mathbf{X}$ that consists

of n training observations in $p-$dimensional space [99, 164]. Mathematically, if our data can be perfectly separated using a hyperplane, then there will in fact exist an infinite number of such hyperplanes as shown for a simplified version on the top for Fig. (5.7). A natural choice is the *maximal margin hyperplane*, which is the separating hyperplane that is farthest from the training observations. We can compute the (perpendicular) distance from each training observation to a given separating hyperplane; the smallest of these distances from the observations to the hyperplane is known as *margin*. The separating hyperplane for which the margin is largest becomes our separating hyperplane. Nevertheless, in many cases no separating hyperplane exists, and the concept of a separating hyperplane is extended in order to develop a hyperplane that almost separates the classes, by means of a so-called *soft margin*. Looking at the bottom of Fig. (5.7), we see three training observations that are equidistant from the maximal margin hyperplane and lie along the dashed lines indicating the width of the margin. We define those points as *support vectors*, whose term indicates precisely the role of supporting the hyperplane position.
In the following analysis we give some mathematical details of the Support Vector Classifier (SVC). The role of SVC is to separate most of the training observations into the two classes, but it may misclassify a few observations. It represents the solution to the following optimization problem:

$$\begin{aligned} &\max M(\beta_1, \ldots, \beta_p, \epsilon_1, \ldots, \epsilon_n), \\ &\text{subject to} \qquad \sum_{j=1}^{p} \beta_j^2 = 1, \\ &y_i \left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\right) \geq M(1 - \epsilon_i), \\ &\epsilon_i \geq 0; \qquad \sum_{1=1}^{n} \epsilon_i \leq C. \end{aligned} \tag{5.11}$$

where C is a non negative tuning parameter and M is the width of the margin, quantity which should be made as large as possible. Penalization terms, $\epsilon_1, \ldots, \epsilon_n$ are *slack* variables that take into account the possibility of having individual observations on the wrong side of the margin (hyperplane).
The classification of a test observation x^* is done by determining the sign of $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$. The tuning parameter C can be thought as a "budget" for the amount which the margin can be violated - observation on the wrong side- by the n observations. The optimization problem defined in Eq. (5.11) has a very interesting property: only observations that either lie on the margin or that violate the margin will affect the hyperplane [164]. Observations that lie directly on the margin, or lie on the wrong side of the margin for their class, are known as *support vectors* and do affect the support vector classifier.
This method is quite robust because decision rule is based only on a potentially small subset of the training observations (the support vectors) and not on the amount of data that are far away from the hyperplane. Everything described

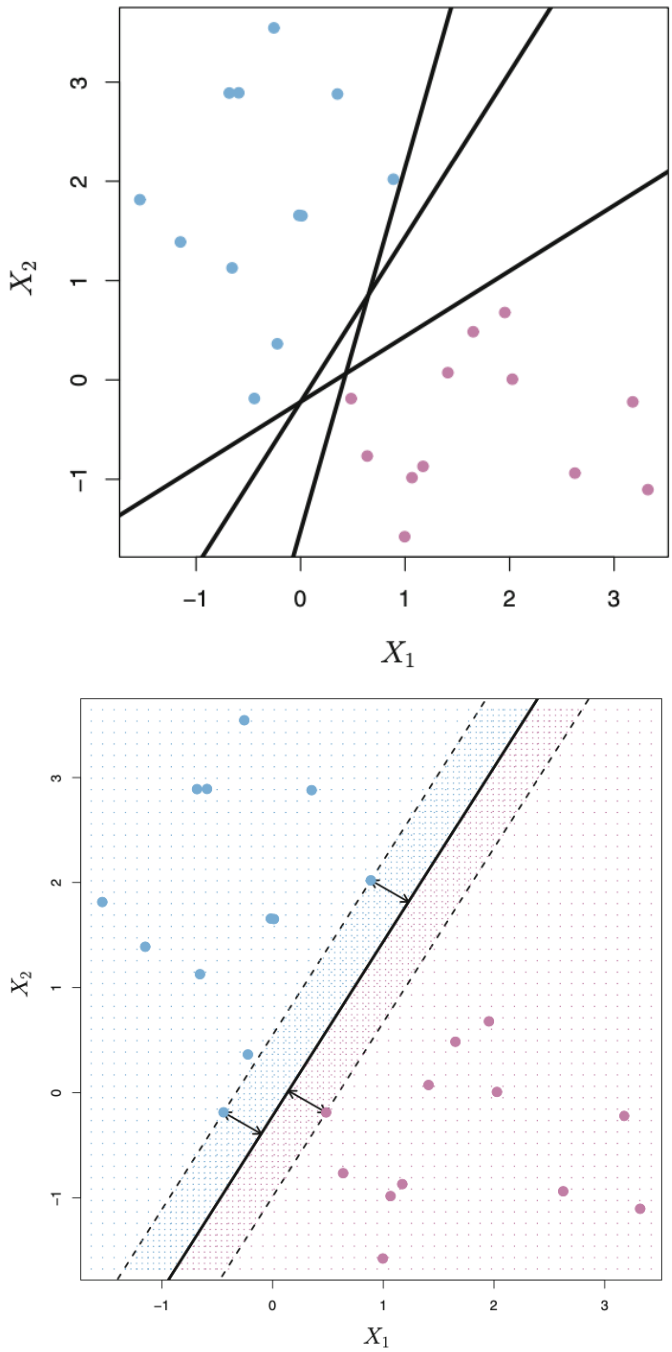


Figure 5.7: (top) Three separating hyperplanes, out of many possible, are shown in black for two classes of observations (blue and purple). (bottom) Maximal margin hyperplane shown as a solid line for two classes of observations (blue and purple). The two blue points and the purple point that lie on the dashed lines are the support vectors.

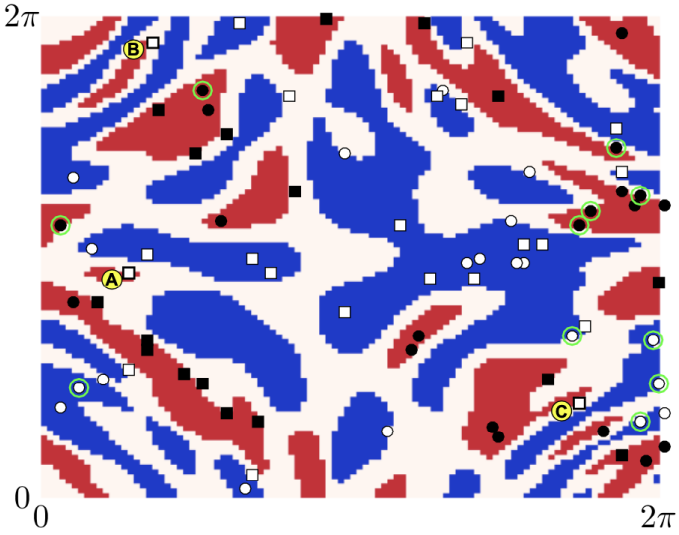


Figure 5.8: Dataset used for the binary classification problem described in [153]. The data labels are red (1) and blue (-1). Label training set (20 points) is shown as white and black circles. Support vectors (green circles) are obtained with the quantum kernel estimation method. Classified test set are represented in white and black squares.

so far works well, if the boundary between the two classes is linear. It means that, considering the easiest case with a dataset described by two variables, we can easily separate the classes with a line. However, there exist non-linear boundaries between classes, like the situation shown in Fig. (5.8). To perform a classification here we need to enlarge the feature space, exploiting higher-order polynomial terms of the training variables such as $X_1, X_1^2, X_j \cdot X_j'$. This procedure could end up with a huge number of features, causing the computations to become unmanageable.
The algorithm that allows enlarging the feature space of the support vector classifier using a *kernel* function is called **support vector machine**. This would generally lead to more efficient computations.
We skip here all calculation details, referring to [99] for a rigorous discussion. It can be show, anyway, that the solution to the support vector classifier problem defined in Eq. (5.11) involves only the inner products of the observations, and that the linear support vector classifier can be represented as:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x_i, x_i' \rangle, \tag{5.12}$$

where $\langle x, x_i' \rangle$, also defined as *gram matrix*, is the inner product between the new point x and each of the training points x_i, where i is the number of training observations. It can be shown that α_i, which is the Lagrange dual, is nonzero only for the support vectors in the solution, feature that is expected, being Eq. (5.12) a different formulation of Eq. (5.11). For this reason, the sum in Eq. (5.12) can be replaced with $\sum_{i \in \mathcal{S}}$ where $\mathcal{S}$ is the collection of indices of these support points.

The *kernel* function, is generalization of the inner product defined above:

$$K(x_i, x_i') = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d; \tag{5.13}$$

$$K(x_i, x_i') = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2). \tag{5.14}$$

There exist several versions of this function, i.e. the formulation Eq. (5.13) is called polynomial kernel while the formulation in Eq. (5.2.1) is *radial* kernel. When $d = 1$, the first one become a *linear* kernel because the support vector classifier is linear in the features. This kernel is a function that quantifies the similarity of two observations using Pearson (standard) correlation. When the support vector classifier is combined with a non-linear kernel as in Eq. (5.13) and Eq. (5.2.1), the resulting classifier is a *support vector machine*.

In the following we introduce the role of quantum computing when the feature space becomes large, and the kernel functions become computationally expensive to estimate.
A core element to computational speed-ups provided by quantum algorithms is the exploitation of an exponentially large quantum state space through controllable entanglement and interference. In this work we made use of two novel methods developed on superconducting processor [153]. The first one is variational classifier that operates using a variational quantum circuit to classify a training set in direct analogy to conventional SVMs. On the other hand, the second method features a quantum kernel estimator, where the kernel function is calculated and optimized directly by the classifier.

5.2.2 Variation Quantum Classifier

In the context of digital quantum computing, QML can be seen as a set of trainable circuits that represent our models. This is often written as a function $f(x, \theta)$ that depends on an input data point x – for example describing the pixels of an image or a vectorized text document – as well as trainable parameters θ. The result of this quantum circuit is a measurement with a probabilistic output. For this reason, the final expectation value is given by multiple run (*shoots*) of the experiment. At this point, the machine learning model depends on the parameter θ which is translated into a rotation of the qubit in the circuit.
As discussed at the beginning of this Chapter, the evolution of the system is governed by an overall unitary transformation $U(x, \theta)$ so that the expectation value of the circuit is given by:

$$f_q(x, \theta) = \langle 0|U(x, \theta)^\dagger M U(x, \theta)|0\rangle. \tag{5.15}$$

The circuit is trained by adjusting the parameters θ with a classical optimization routine. It learns from data, minimizing a standard cost function with the

same logic of traditional NN, such as mean square loss. Trainable circuits are also known as *variational* or parametrized circuits. This interplay of quantum and classical computation goes under the name of *hybrid architecture*.

The Variation Quantum Classifier (VQC) exploits a variational circuit to generate a hyperplane in a *quantum feature space*, separating data that have been mapped non-linearly to a quantum state through a *feature map* $\mathcal{U}_{\Phi(\mathbf{x})}$. This protocol follows four steps: mapping of the data to quantum space, definition of the variational circuit, measurement phase and decision updates.
The feature map represents a unitary transformation of the type: $|\Phi(\mathbf{x})\rangle = \mathcal{U}_{\Phi(\mathbf{x})}|0\rangle$ that, in this formulation, maps each data feature to a qubit state. The basic components of the feature map are: the Hadamard gates defined in Eq. (5.4) applied to all qubits ($H^{\otimes n}$), single qubit rotations and different combination of entangling operations. This procedure is executed by the `Qiskit Aqua` library, with different options for data encoding to circuits[4]. These circuits are called `Pauli Expansion`, and they perform transformation of the input data $\mathbf{x} \in \Re^n$ as:

$$\boxed{\mathcal{U}_{\Phi(\mathbf{x})} = \exp\left(i \sum_{S\subseteq[n]} \phi_S(\mathbf{x}) \prod_{i\in S} P_i\right);} \tag{5.16}$$

where the variable $P_i \in I, X, Y, Z$ denotes the identity and the Pauli matrices σ_x, σ_y and σ_z. The index S describes connectivities between different qubits or datapoints. Per default, the data-mapping ϕ_S is:

$$\phi_S(\mathbf{x}) = \begin{cases} x_0 & \text{if } k = 0 \\ \prod_{j\in S}(\pi - x_i) & \text{otherwise.} \end{cases} \tag{5.17}$$

The algorithm flow proceeds for the second step with the quantum states, representing the encoded input data, to be processed via the application of further quantum gates. This sequence of operation is defined by an *ansatz* state, an euristhic parametric function chosen for the variational part. This final state is then measured, given a specific parameter value, to produce the output:

$$p_y(\mathbf{x}) = \langle\Phi(\mathbf{x})|W^\dagger(\theta)M_y W(\theta)|\Phi(\mathbf{x})\rangle. \tag{5.18}$$

The variational circuit is a parameterized quantum circuit $W(\theta)$, which depends on a set of circuit parameters θ, as well as an objective function defining a score for the set of parameters. After each query, the measurement output bit string in the z-base is mapped to a label $f(z) = y$. It can be understood as a separating hyperplane in the quantum state space. In the training phase, the objective function estimates the distance between the expected classification

[4]PauliFeatureMap: `https://qiskit.org/documentation/apidoc/circuit_library.html`.

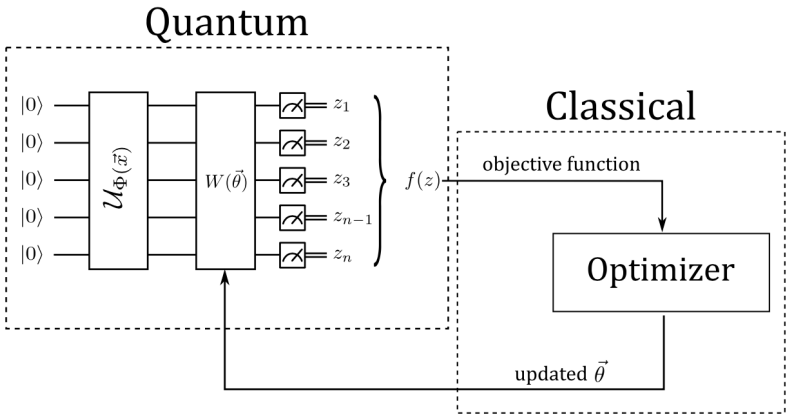


Figure 5.9: The VQC circuit comprised of a feature map $\mathcal{U}_{\Phi(\mathbf{x})}$ followed by a variational circuit $W(\theta)$ and the mapping function $f(z) = y$ on the measured bit string z.

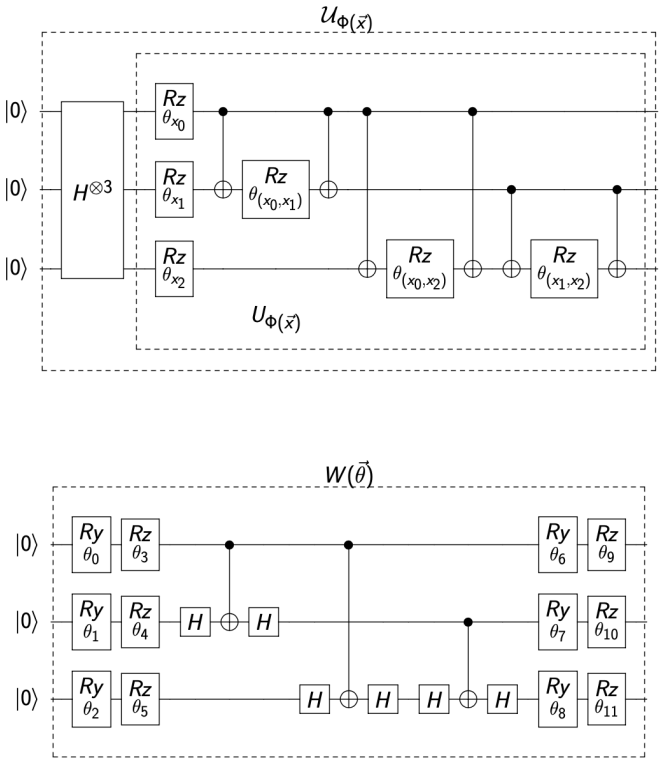


Figure 5.10: (top) `ZZFeatureMap` for a data sample $\mathbf{x}$ with three features. The circuit has only one $\mathcal{U}_{\Phi(\mathbf{x})}$ bloch, but there could be many consecutive $\mathcal{U}_{\Phi(\mathbf{x})}$. (bottom) $RYRZ$ variational circuit for a data sample $\mathbf{x}$ with three features. The circuit above has only one $W(\theta)$ block, but there could be many consecutive $W(\theta)$.

results for a training set and the labels obtained from the variational circuit. This estimated distance is used by the optimizer leveraging classical optimization tools such as `COBYLA` and `SPSA` to update circuit parameters[5]. The whole structure of the VQC is shown in Fig. (5.9) while an example of the feature map and of the variational circuit (ansatz) are shown in Fig. (5.10), top figure and bottom figure respectively. The feature map starts by creating a superposition among all the three qubits, then a series of R_Z rotations are applied followed by `CNOT` gates interleaved by R_Z rotation creating specific entanglement blocks between all possible qubit couples. This construction allows the exploration of a large non-classical computational space. On the other hand, the variation circuit consists of successive R_Y and R_Z rotations parametrized by θ, then a typical entangling block made with a `CNOT` gate put between two Hadamard rotations. This block is applied to all possible qubit couples and the ansatz is closed with a symmetric sequence of R_Y and R_Z rotations parametrized by different angles.
Looking at Eq. (5.12), we can say that this variational approach exploits quantum algorithm to improve the prediction on α_i. In the following, we introduce a complementary approach, that performs the optimal parameter finding by employing the afore mentioned kernel methods.

5.2.3 Quantum Kernel Method

The second classifier introduced in [153] is called Quantum Support Vector Machine (QSVM). As anticipated before, a SVM models the learning problem as an optimization problem with an objective function, the core information of which relies in the inner products of the data points to be classified. The kernel trick is then introduced to solve the classification for non-linearly separable classes. Classically, having to handle high dimensional dataset, we should compute the feature map transformation and perform the inner product induced by the kernel function. Here, instead, we will use the natural induced inner product of the Hilbert space. Technically, after having mapped the dataset to quantum states, like in the first step of the VQC[6], we perform the inner product as:

$$K(\mathbf{x}, \mathbf{x'}) = \langle \phi(\mathbf{x}) | \phi(\mathbf{x'}) \rangle. \tag{5.19}$$

The problem of Eq. (5.12) can be written, performing the minimization, as a loss function[7]:

$$L_D(\alpha) = \sum_{i=1}^{t} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{t} y_i y_j \alpha_i \alpha_j K(\mathbf{x}, \mathbf{x'}) \tag{5.20}$$

[5] All details can be found here: `https://qiskit.org/documentation/apidoc/qiskit.aqua.components.optimizers.html`.
[6] The notation x is a shot form for $|\phi(x)\rangle$
[7] All the steps of differentiating the minimization problem can be found in [99].

where $\sum_{i=1}^{t} \alpha_i y_i = 0$ and $\alpha_i \geq 0$ for each i.
This problem is concave[8], whenever $K(\mathbf{x}, \mathbf{x'})$ is a positive definite matrix, and the final solution is given by the vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_t)$.
Quantum information processing techniques can then be used to build kernels which may be hard to compute classically. Conversely, it can be used to speed up the evaluation of inner products between feature vectors, getting the same result directly from the original data points.
Considering the feature map as the unitary: $\mathcal{U}_{\Phi(\mathbf{x})}$ and the conventional initial state for all qubits $|0^{\otimes n}\rangle$, Eq. (5.19) becomes:

$$\boxed{K(\mathbf{x}, \mathbf{x'}) = |\langle 0^{\otimes n}|\mathcal{U}^{\dagger}_{\Phi(\mathbf{x})}\mathcal{U}_{\Phi(\mathbf{x'})}|0^{\otimes n}\rangle|^2} \tag{5.21}$$

which is the probability of observing the state $|0^{\otimes n}\rangle$ in the Z basis. The resulting probability distribution is obtained with N shoots up to a sampling error $\tilde{\epsilon} = \mathcal{O}(1/\sqrt{N})$. Recalling the geometric interpretation of a classification as the assignment of a new data point to a specific side of the hyperplane, we can summarize the procedure in the following way. The optimal solution α^* for a new datum $\mathbf{s}^* \in \mathcal{S}$ with all the support vectors, is given by the kernel, calculated on the quantum computer as:

$$\boxed{\tilde{m}(\mathbf{s}) = sign\left(\sum_{i=1}^{t} y_i \alpha_i^* K(\mathbf{x_i}, \mathbf{s}) + b\right)} \tag{5.22}$$

where $\tilde{m}$ is the label, or the class assigned to the new datum. The possibility given by this quantum enhanced classification allows for the classification of complex scenario as the one depicted in Fig. (5.8).

5.3 QML for VBS

In this section we provide details about the possibility to use quantum algorithm to solve classification problems related to VBS, largely discussed in the previous Sections. We briefly define the initial settings and eventually we show results and comparisons with classical methods.

The classification problems we aim at solving are: *a)* the binary ones related to the W rest frame reconstruction to determine the correct solution of Eq. (4.1), *b)* the determination of the polarization class (L,T) for the semi-leptonic VBS channel. Training variables used for the second case are shown, as a reference, in Fig. (5.11).
Compared to the possibilities provided by NNs, here we have to optimise the number of variables because of the limited number of qubits that can be used

[8]With this term we refer to the usual definition of a concave (non-convex) functional form.

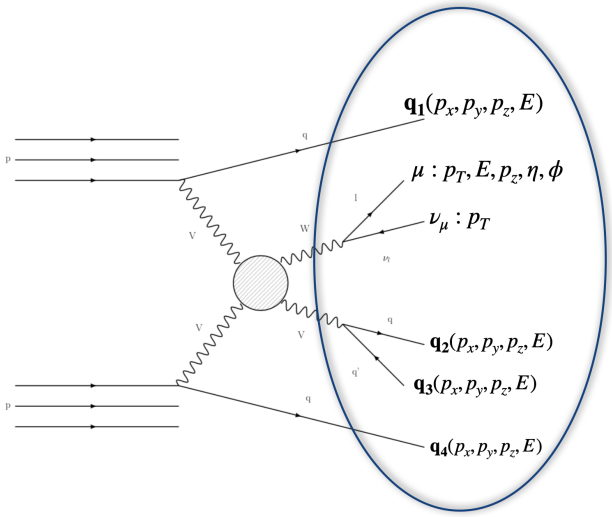


Figure 5.11: Representation of VBS semi-leptonic channel and relative variables used for the training.

in the feature mapping. Having more than twenty variables results in a quantum circuit depth[9] that is not affordable for NISQ devices at the moment. The bigger the number of qubits and circuit gates, the higher the quantum volume (performance) required to get meaningful results. In this scenario, QSVM and VQC become part of an hybrid computation architecture, where the quantum algorithm can be applied after a dimensionality reduction of the input variables.
There are many techniques that can be used for dimensionality reduction, among these we have: matrix factorization method, where techniques from linear algebra can be used for dimensionality reduction, and autoencoder methods, where deep learning neural networks can be constructed to perform the task.

In this work, we focus on the first kind of techniques, using the Principle Component Analysis (PCA). For this reason, before proceeding, we recall its main properties.

PCA

Qualitatively, with a large set of correlated variables, principal components allow us to summarize this set with a smaller number of representative variables that collectively explain most of the variability in the original set. To obtain this goal, variables that describe the system are transformed into a new set of variables (the principal components), obtained as linear combination of the original ones. These new variables are constructed so that they have the interesting property of being uncorrelated among each other.

[9]The depth of a circuit is a metric that calculates the longest path between the data input and the output in terms of single, two-qubits gate. Each gate counts as a unit.

PCA is carried on through the following steps. The first principal component is built in such a way that it maximizes the variance among the data. Anyway, before applying this method, one has to manipulate the variables, centering and scaling them to obtain a standard Gaussian distribution with 0 mean and unit variance. A standardization process is needed to avoid huge (in absolute value) fluctuations due to different variable scales representation.
The first principal component of our training variables $X_1, \ldots, X_p$ is the normalized linear combination of the features:

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p; \tag{5.23}$$

that has the largest variance, where Z is defined as the *scores* of the first principal component. Normalized means that $\sum_{j=1}^{p} \phi_{j1}^2 = 1$ where ϕ_{j1} are defined as the *loadings* of the first principal component.
Geometrically, this linear combination represents a selection of a new coordinate system obtained by rotating the original one. This geometric interpretation, starting from the first principal component, suggests that the loading vector $\boldsymbol{\phi_1}$ defines a direction in feature space along which the data vary the most. Then, the second principal component can also be computed, as the vector Z_2 that has maximal variance out of all linear combinations that are *uncorrelated* with Z_1. Again, geometrically, this condition is equivalent to constraining the direction of $\boldsymbol{\phi_2}$ to be orthogonal to $\boldsymbol{\phi_1}$. The third component is orthogonal to both the first and the second and maximize the variance among the allowed directions. Similarly, for all the other components. Principal components depends solely on the covariance matrix Σ (or the correlation matrix ρ) of the original training dataset. Indeed, the principal component directions $\boldsymbol{\phi}_1, \ldots \boldsymbol{\phi}_p$ are the ordered sequence of eigenvectors λ_i of the matrix $\mathbf{X}^T\mathbf{X}$ where the variances of the components are the eigenvalues. See [165] for additional details.

We applied the above described PCA analysis to our simulated VBS semileptonic dataset, already introduced in Chapter 4. We show in Fig. (5.12) the heatmap corresponding to the correlation matrix of the two problems investigated. In the top plot the classification task related to determine the correct solution of Eq. (4.1), while in the bottom plot the correlation for the dataset used in the polarization class (L,T) determination. What has been done so far regarding PCA is trying to find a low-dimensional *linear* subspace where the data are confined to. Yet, it might be that the data are confined to a low-dimensional *non linear* subspace. Likewise the SVM discussion, dealing to a non linearity suggests to move the problem to higher dimensionality. For this reason, a *kernel PCA* has been introduced in [166] and we tested it for our problem.
We can see in Fig. (5.13) a theoretical comparison between the two situations. The data points on the right are located mostly along a curve in 2D. PCA cannot reduce the dimensionality, because the points are not located along a straight line. But still, the data are located around a one-dimensional nonlinear curve. So while PCA fails, kernel PCA can find this non-linear manifold

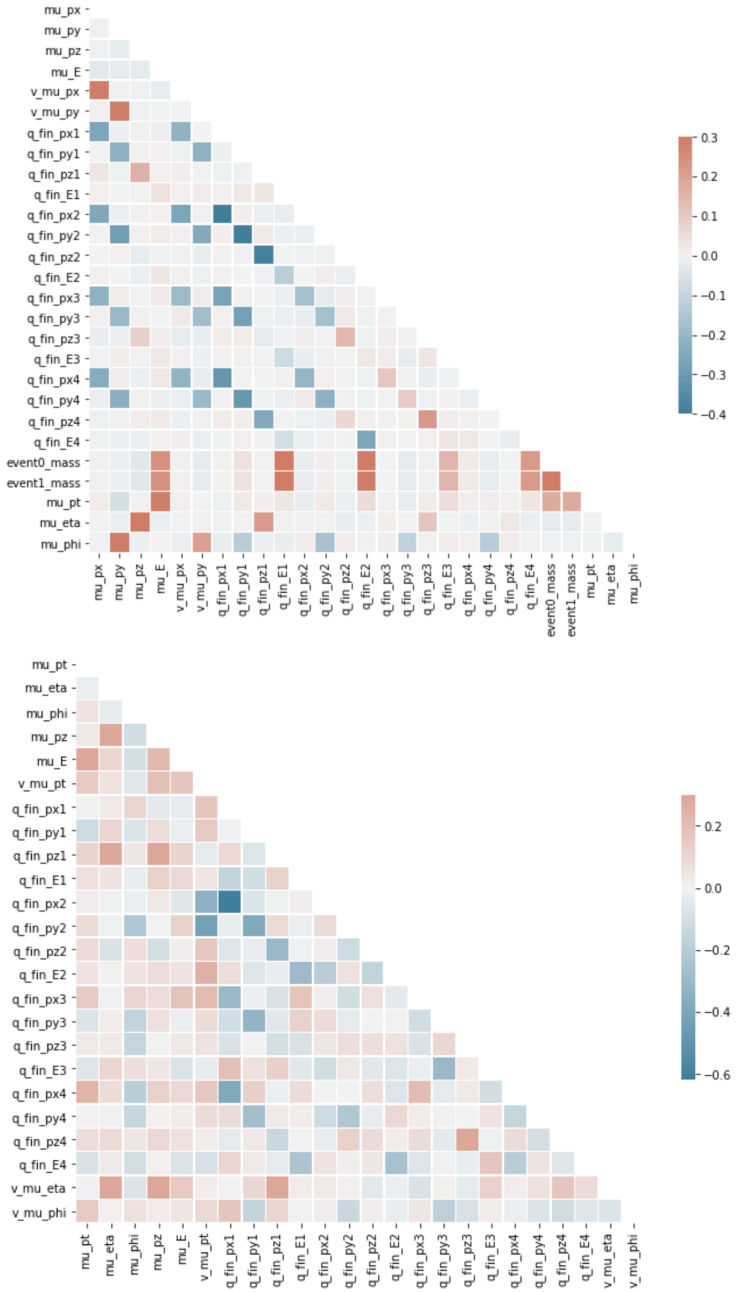


Figure 5.12: Correlation matrix representation for the two classification problems analyzed. Training variables related to the W rest frame reconstruction to determine the correct solution of Eq. (4.1) (top). Training variables related to a binary classifier to determine the polarization class (L,T) for the semi-leptonic VBS channel (bottom).

and discovering that the data are in fact nearly one-dimensional. As a matter of fact we need to increase the dimensionality in order to be able to decrease it.

Reducing the number of variables leads to a reduction in the total explainability of our original dataset. We need to find a balance between the number

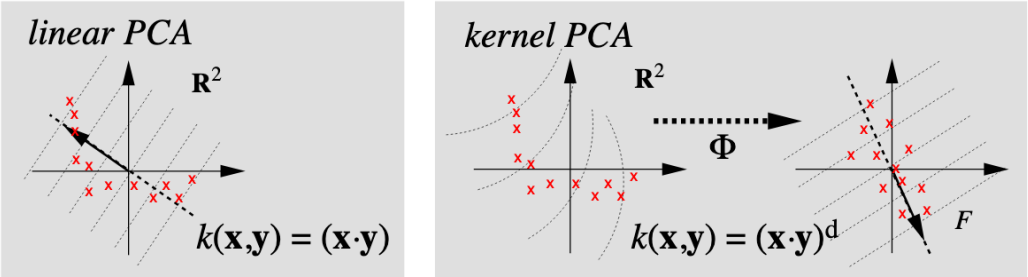


Figure 5.13: Basic idea of kernel PCA: by using a non linear kernel function k instead of the standard dot product, we implicitly perform PCA in a possibly high dimensional space F which is non linearly related to input space. The dotted lines are contour lines of constant feature value [166].

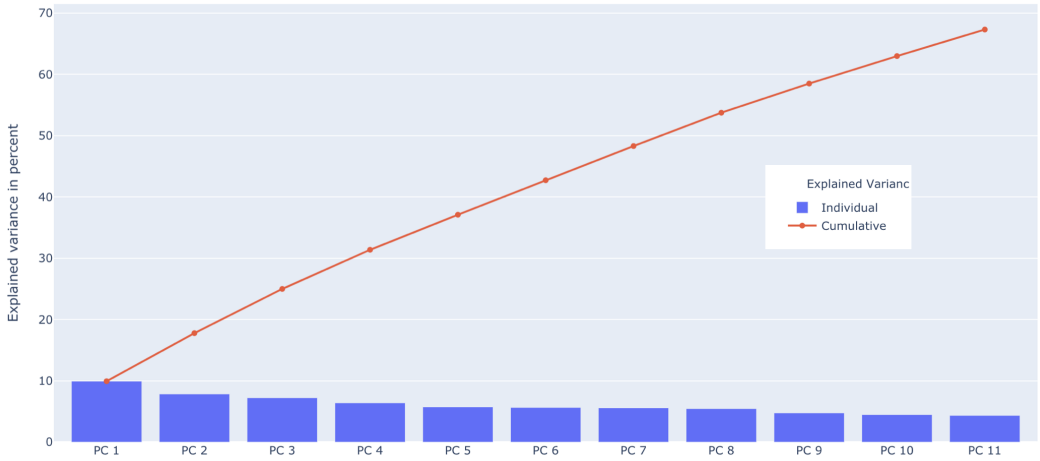


Figure 5.14: Representation of the cumulative variance distribution for the principal components. The dataset refers to VBS semi-leptonic channel.

of principal components that we keep, to fully capture the major pattern in the data, the number of qubit we may use, and how much of the information is lost. In this sense we can calculate the proportion of variance explained (PVE) by each principal component, and in particular the cumulative proportion of variance explained. In a very lucky scenario, two or three out of tens of features provide a cumulative variance explained of roughly 90%. This happens when the entire dataset variance is contained in the first and in the second components. In our case, the situation is completely different, as we can see in Fig. (5.14): unluckily, we do not have predominant principal components and to get a fair description of the original problem - looking at a cumulative variance greater than 75% - we must keep more than ten components.

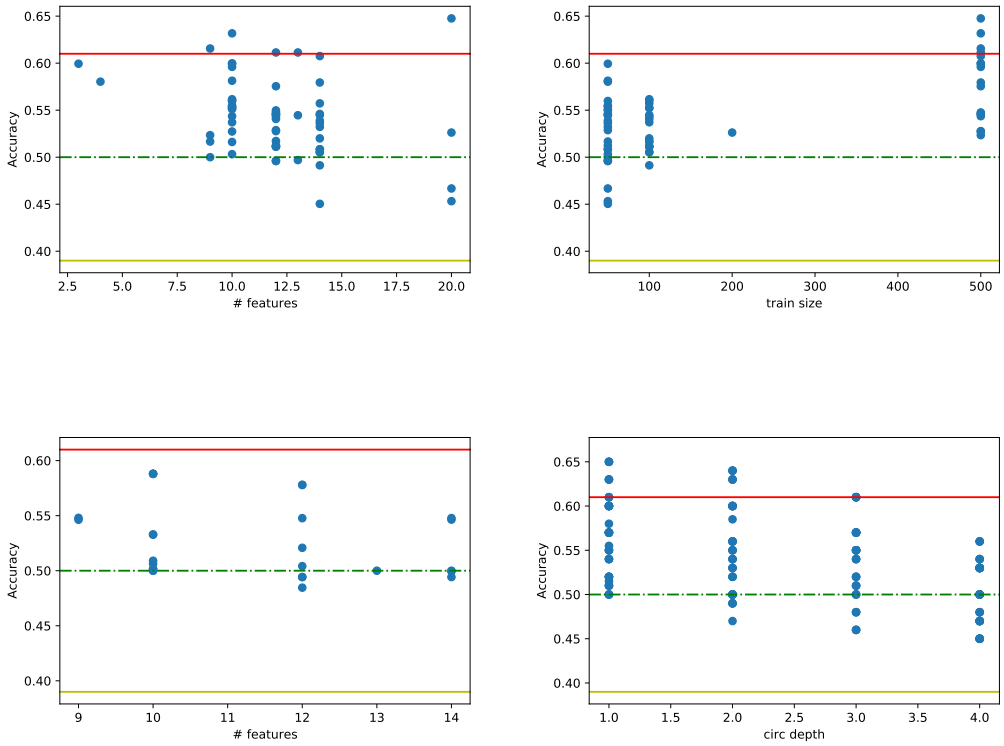


Figure 5.15: Evolution of the quantum classifier performance according to different parameters. Accuracy evolution as a function of features (qubits) as a solution of the Eq. (4.1) (top left). Accuracy evolution as a function of training size as a solution of the Eq. (4.1) (top right). Accuracy evolution as a function of features (qubits) for the polarization classification (T,L) (bottom left). Accuracy evolution as a function of circuit depth (common behaviour among the two classifier) (bottom right).

5.3.1 Algorithm Performance

In this paragraph, we discuss the main results obtained using quantum algorithms and we provide a comparison, tuning several parameters, to the classical elaboration.

In Fig. (5.15) we can appreciate the performance of QSVM classifier as a function of different parameters. Results are obtained using a linear entanglement between the qubits (features), running the algorithm on a simulated backend (*statevector_simulator*). This is the standard approach to quantum computing, because one has to find the combination of parameters that mathematically guarantee the best performance. Only at that point it make sense to run the algorithm on a real processor, keeping in mind that this involves further study of errors and noise due to the technology adopted. The performances are measured in terms of accuracy, namely the number of correct predictions on the total number of attempt. We evaluated the performance of this technique, as well as for the VQC, when increasing the number of features (qubits), the dimensionality of the training sample, and when changing some quantum specific parameters like the feature map. Concerning the VQC, we

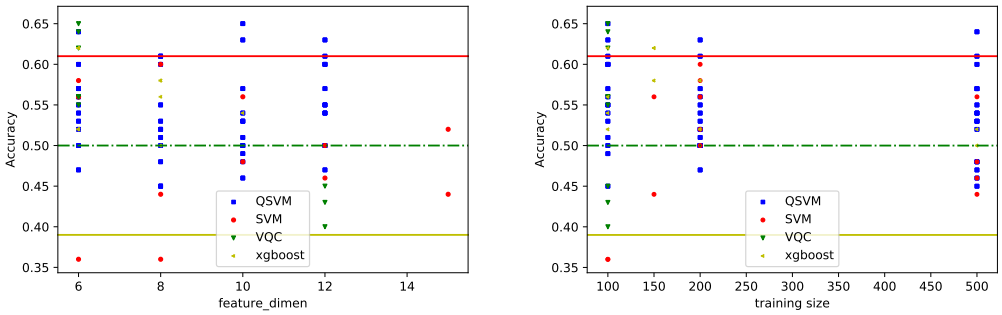


Figure 5.16: Comparison of binary classification performance using different models as a solution of the W rest frame reconstruction. (Left) Evolution of the accuracy of the model as a function of the features size. (right) Evolution of the accuracy of the model as a function of the training size. Red and yellow lines define a poor classification performance area, namely between 0.4 and 0.6 in accuracy.

performed several tests also studying different classical optimizers and their specifications. The structure of the quantum circuits adopted in the present test is described in [153], the mathematical steps of which have been translated into the specific QSVM library of `Qiskit`. At the time of writing this thesis, there is no proven link between the choice of algorithm parameters and the properties of the dataset. This led us to perform a multidimensional scan against several variables, and through a process of reverse engineering, we were able to reconstruct the best performing circuit.
A common behaviour that can be inferred from the plots is the deterioration of the performance with respect to the depth of the circuit. The depth is defined as the number of repetitive applications of the chosen feature map $\mathbf{U}_{\mathbf{\Phi(x)}}$, as shown in Fig. (5.10) on the top. This feature map can be composed of only single qubit rotations (`Z Pauli Expansion`) or even `CNOT` gates to create not only a superposition state for each qubit but also non-classical entangled states (`ZZ Pauli Expansion`[10]).
In Fig. (5.16), instead, we compare the performance of different classification techniques, both classic and quantum. This plot represents a collection of different and separated optimizations obtained for each technique. We have two classic methods, a standard SVM and XGBoost, as well as the two quantum algorithms, the QSVM and VQC. In Fig. (5.17) we show a comparison of the performance for the two different classification problems. Results are aligned to previous findings where, according to the parameter tested, there is not a clear evidence of classification improvement over 10 features while a better classification is achieved from 50 to 100 and to 500 training samples.
It is worth stressing here that all the results presented here are run on quantum simulator and not on real hardware. Quantum enhanced kernel calculation requires a number of quantum circuits to execute that are a function of the

[10]The Z rotation represent the default one but the feature map can be created with different combiantion of the Pauli representation (X, Y, XX, XY... etc).

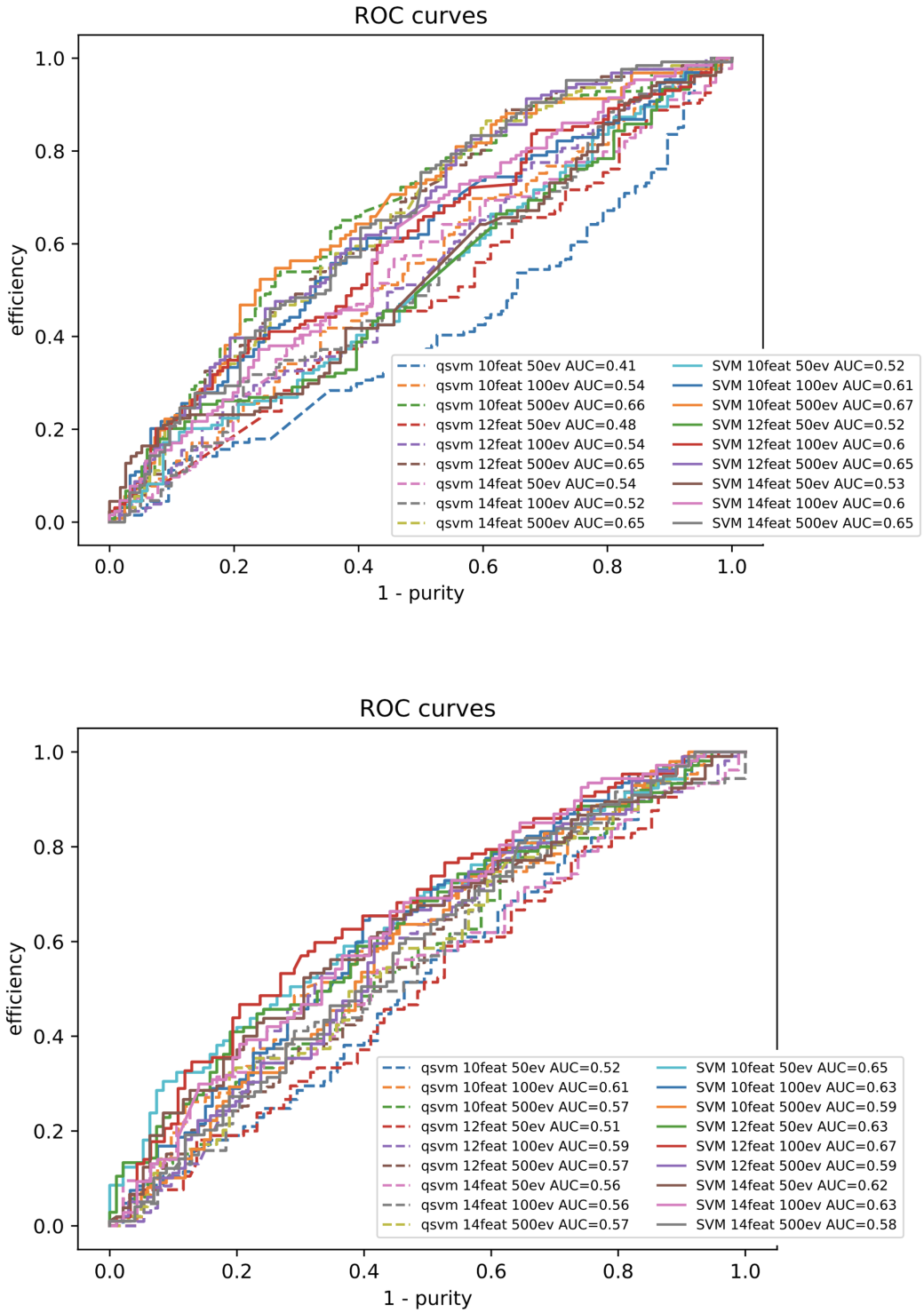


Figure 5.17: Comparison of binary classification performance using the ROC Curve for SVM and QSVM techniques. Different combination of training size and number of feature. (Top) ROC curve evolution for different models as a solution of the W rest frame reconstruction, (Bottom) ROC curve evolution for different models as a solution to the classification of W polarized states (T,L). Dashed curves represent QSVM result whereas continuos lines are classic SVM. Naming convention for each model is: number of features, training sample dimension and area under the ROC curve.

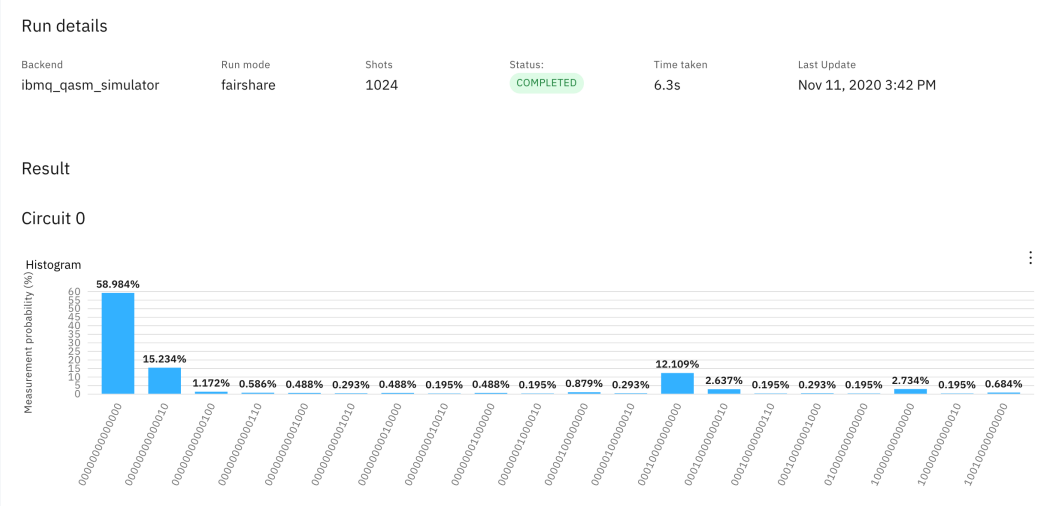


Figure 5.18: Represention of the probability distribution obtained for a specific circuit run on the IBM Quantum Experience platform.

square of the number of input vectors, which may not be a good selection for classifying huge number of events. This is why we are looking for good accuracy for a small dataset training set. Regarding the available simulators we remind here that we used a *statevector_simulator* only to get an idea about the accuracy of the current configuration of the algorithm, since it is much faster than the full quantum simulator named *qasm_simulator* [125]. In fact, roughly comparing the computation times, the first one required, on average, a factor of 30 less than a full quantum simulation. This evidences the exponential time required by a classical computer to simulate a quantum one.
In Fig. (5.18) an example of QSVM circuit results is shown. We can see how the algorithm produce a distribution with higher probability for the states that represent the solution of the kernel evaluation.

A natural evolution of this study would be to run the same circuit on a real quantum hardware. In this case, the whole computation made of several circuit execution, would be affected by error. A discussion about possible error mitigation techniques is out of the scope of this work, however in `Qiskit` there is a specific package that allows the user to mitigate errors with respect to the specific hardware used [167].
A quantitative comparison between classical SVM and QSVM, is summarized in Table (5.1). All the model features and results are presented in terms of accuracy and AUC. These results should be compared to AUC values obtained with DNN, see Table (4.2). A final investigation that we propose, is to check how a DNN architecture, as proposed in the previous Chapter, performs under the same conditions exposed in this Chapter, namely, using reduced variables and a much smaller number of events than those traditionally used in Deep Learning. For this reason, in Fig. (5.19), we show the usual ROC curve where

Table 5.1: This table shows the comparison, in terms of *accuracy* and AUC, among classical SVM, and SVM with data preprocessing executed with kernel PCA (k SVM), and QSVM as a solution of the W rest frame reconstruction. We have different dataset compositions in terms of number of features and training dataset size. The two metrics used as a comparison are the *accuracy* of the model and AUC.

Feat.	Train	QSVM ACC	QSVM AUC	SVM ACC	SVM AUC	k SVM ACC	k SVM AUC
10	50	0.5	0.52	0.57	0.52	0.55	0.39
10	100	0.56	0.56	0.67	0.61	0.72	0.56
10	500	0.63	0.66	0.64	0.67	0.65	0.31
12	50	0.55	0.55	0.57	0.52	0.57	0.51
12	100	0.54	0.56	0.68	0.6	0.69	0.58
12	500	0.61	0.65	0.64	0.65	0.64	0.7
14	50	0.53	0.53	0.61	0.53	0.55	0.5
14	100	0.56	0.53	0.67	0.6	0.66	0.57
14	500	0.61	0.65	0.64	0.65	0.64	0.7

it is evident how the various architectural combinations do not deviate from the diagonal of the plot, in agreement to the values of AUC reported in the label of the captions. The presented approach provides results which are not comparable with the proposed quantum model. Considering the results shown so far, there are two possible directions for further improvements of the applied methods. The first is linked to the classic aspect of this architecture, that is, the feature reduction. The second is related to the possible customizations of the quantum feature map. Actually, the greatest customization can be done when choosing a second order feature map, with the possibility to exploit entangled states. Several tests performed, however, show that under current conditions, the performance of the first-order feature map is higher.
During this study, several second order tests were performed, in which different Pauli matrix combinations were used. These attempts did not show any particular improvements. Concerning instead the first aspect of variable reduction, some research has been done moving from linear PCA to kernel PCA. As can be seen in Fig. (5.20), under the same condition of standard PCA application, we have a clear improvement of the performance for classic SVM, in particular in case of higher number of features (14). Every time we look for a kernel method, as in classic SVM, the best way to define the kernel is to adopt a parametrized grid search[11] that performs an exhaustive evaluation over a specified list of parameters. We implemented this technique both for SVM and kernel PCA.

A different approach, that has not been implemented, which consists in tackling the feature dimension reduction, is known as *autoencoder*. An autoencoder is a special neural network for computing a type of non-linear principal-component decomposition. Classical autoencoders can be trained

[11] `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html`.

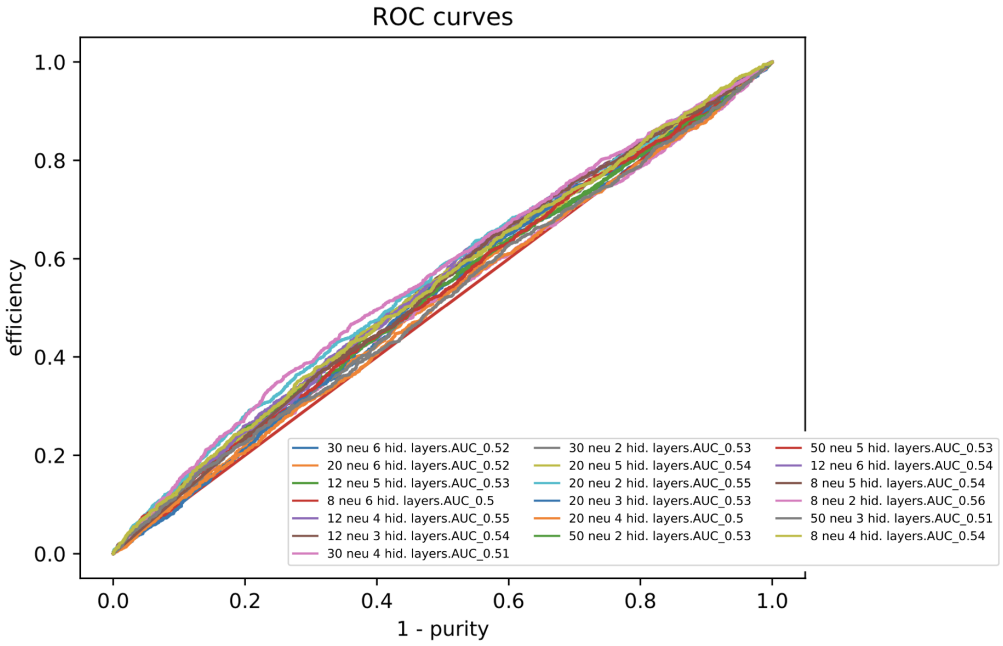


Figure 5.19: ROC curve evolution for different DNN models as a solution of the W rest frame reconstruction using 12 PCs variables trained using 100 events. Each model is described with the following naming convention: number of neurons - number of hidden layers - AUC value.

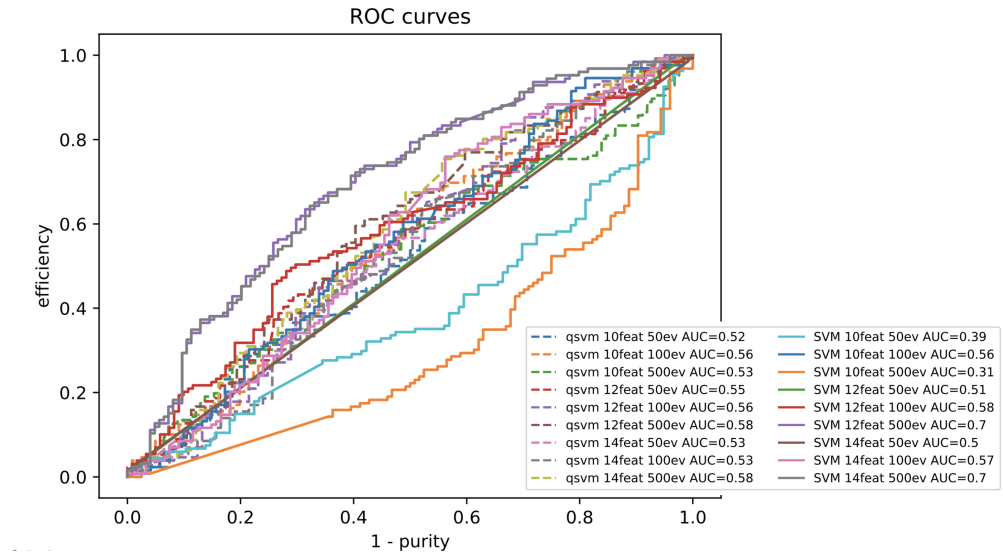


Figure 5.20: ROC curve evolution for different models as a solution of the W rest frame reconstruction using the Kernel PCA variable reduction method. Each model is described with the following naming convention: algorithm - number of features (PCs) - training set dimension - AUC value.

to extract a low-dimensional representation (in the so called *latent space*) of higher-dimensional data and, in reverse, constitute the basis for generating new plausible samples not originally included in the training set.
There exist also a quantum version, namely *variational quantum autoencoders* that fits nicely into the current stream of heuristic and semi-heuristic approaches, and which seems particularly well suited for noisy intermediate-scale applications. Quantum autoencoders have been described primarily as tools to compress quantum states into a smaller Hilbert space [168], thus enabling a more effective use of quantum resources. For HEP classification this requires a lot of encoding and measuring. Further technological improvements, such as quantum memory, might be necessary to cope with big dataset.

5.4 QC for HEP: future applications

As a conclusion, in this section we discuss additional examples about QC application in HEP, looking at the main tasks that are already touched by this technology.
The ambitious high luminosity LHC (HL-LHC) program, in the next two decades and beyond, will require enormous computing resources. New technologies such as quantum computing could possibly help to overcome this computational challenge, not only for QML, but on more generally exploiting the possibility to map quantum effects on quantum circuits.

An interesting line of research concerns generative models, such as Boltzmann Machines, Variational Auto-Encoders and Generative Adversarial Networks, and their quantum counterparts.
Regarding classical generative models that are being investigated by the HEP community as solutions to speedup Montecarlo simulation, their training is a difficult task, computing intensive. Quantum Graph Neural Networks for particle track reconstruction promise to greatly speed-up search in large parameter spaces due to quantum optimization procedure (QUBO).
Tracking is the root of event reconstruction in particle physics. Because of their ability to evaluate a very large number of states simultaneously, these techniques may play an important role in the future of track reconstruction in particle physics experiments. State-of-the-art algorithms rely, today, on a Kalman filter based approach [169]. They are considered a robust approach and they provide good physics performance, however they are expected to scale worse than quadratically with the increasing number of simultaneous collisions.
During the last period of this work we started to evaluate another interesting approach of QC to HEP. Recently, some literature has been proposed to speedup Montecarlo integration process in the finance sector [145]. This is a very active research field, Quantum Amplitude Estimation (QAE) can compute an integral with a smaller number of iterative calls of the quantum circuit which calculates the integrand, than classical methods call the integrand sub-

routine. This could be crucial in particular for the computation of high order in perturbation theory which are not feasible today. As already mentioned, in [134], a first real application of quantum algorithm applied to parton shower is studied and evaluated. The state-of-art parton shower modelling is based on classical Markov Chain MC (MCMC) algorithms to efficiently generate high multiplicity radiation patterns. Classical MCMC is not able to capture the important quantum interference effects, and a full classical calculation scales exponentially with the number of steps. In future work the analysis of final state radiation would need to include those quantum effects and QC seems to be the right candidate. The authors, in fact, developed a polynomial time quantum final state shower that accurately models the effects of intermediate spin states similar to those present in high energy electroweak showers. Quantum algorithm accounts for the amplitude estimation of final states when there is more than one emission during the evolution. In this case the classic matrix formalism is insufficient and one must compute the full amplitude for which there are $\mathcal{O}(2^N)$ possible histories.
Similar approach for classification in HEP has been used in [170, 171]. Usually the performance of the quantum algorithm is compared with classic methods like BDT and DNN. This is basically the file-rouge of this work: a comprehensive study of different classification methods from classical to quantum algorithms.

Key Results

In this Chapter we have introduced the new paradigm of Quantum Computing. We started with the definition of simplest quantum system featuring the peculiarities of quantum mechanics, the qubit. We have defined the basis of the computation, the unitary gates and we have assembled them together creating quantum circuits focusing on the main principle of quantum computation: superposition, entanglement, interference and measurement. We have realised how this circuits are representation of innovative quantum algorithms that are currently used in different fields, from quantum chemistry, to finance, optimization and to HEP. We have defined what the primary objective is, that is to prove a *quantum advantage* in terms of algorithm precision, time to execution and resolution of problems that are classically impossible to deal with.
The HEP represents an excellent test bed for QC: a sector capable of providing complex problems, linked to quantum mechanics and which put classical computation in crisis. In this direction, we have taken up our VBS problem, benchmarking our DNN framework, described in the Chapter 4, to a quantum classifier.
Considering just the semi-leptonic channel, the classification problems tackled with quantum algorithm are the one used to solve the ambiguity of Eq. (4.1) for the reconstruction of the W reference frame, and the one that discriminates between longitudinal (L) and transverse (T) W polarized state. Since the QSVM and VQC are the quantum version of the classic SVM algorithm, we provided a direct comparison among the two approaches, building a common framework for the initial feature reduction data preparation (PCA, KPCA). We adopted two metrics for the comparison, the *accuracy* of the model and AUC and we changed mainly parameters: the number of events in the training sample, the number of feature. For each classifier, we also implemented an hyperparametrization of internal parameters like the feature map, the variational form and kernel type.

From a general perspective, comparable performances between classic SVM, QSVM and DNN are obtained for the W reference frame classifier (AUC $\sim$ 0.7), whereas DNN works better in classifying the polarization state. In this case, the AUC is $\sim$ 0.8 and $\sim$ 0.7 for the DNN and QSVM/SVM respectively.
This main difference between SVM/QSVM and DNN approach is the dimension of the training sample, at least three orders of magnitude, that allows DNN to learn and classify better the problem with the related time-cost of computation. Indeed, we checked how a DNN architecture performs under the same conditions we adopted in the usage of quantum algorithm, using reduced variables and a much smaller number of events. The results obtained are poor and the classification does not deviate from a random result, confirming the potential of the results obtained with a quantum approach. On the other hand, having a comparable performance for the reconstruction of the W reference frame highlights the complexity of the problem, an ambiguity that cannot be solved up to that level of accuracy.
Important differences arise between SVM and QSVM: apart from internal QSVM tuning (depth, expansion), the quantum method does not show deterioration of the performance while increasing the number of features, indeed showing a slight improvement. On the contrary, SVM suffers from an increasing computational demand which would imply a flattening of performance when typical HEP problems with tens of variables are involved. The same effect is expected considering the training size but this cannot be compared clearly with QSVM due to the current technical difficulties in ingesting a big dataset into the quantum algorithm structure. QSVM and VQC, in addition to a clever variable manipulation, would represent a good and quick response in building robust classifiers.

Conclusions and Future Perspectives

If we do not take advantage of our opportunities, it is our own fault.

Thomas J. Watson

The Standard Model of particle physics describes the known fundamental particles and their interactions. Despite the SM success, culminated in the last Nobel Prize in 2013 for the discovery of the Higgs boson, the last missing piece of the Standard Model formulation, important questions about the underlying nature of our Universe still remain unresolved. This pushes the current experiments at hadron colliders to search for evidence of new physics directly and to investigate known processes at higher precision, to search for possible deviations from the SM predictions. One of the fundamental aspects to be clarified is the inner nature of the breaking of the ElectroWeak symmetry which, in the SM, occur via the Higgs mechanism. ElectroWeak Symmetry Breaking also rules the scattering of vector bosons in the SM, avoiding its divergence at high energy. The rate of occurrence of VBS processes predicted by the SM is very low and new physics can disturb this delicate balance, leading to potentially large enhancements of the VBS rate.

In this thesis work, we have approached the study of VBS by exploiting both classical approaches and innovative solutions. At first, we studied the state-of-art of this process description in Montecarlos, by means of an extensive study of the same-sign WW process. We provided a systematic comparison between different generators on inclusive observables and differential distributions, and we investigated the impact of parton shower matching. The main outcome of this work flew into a paper which has been published in EPJ [37].
Then, we focused on the discrimination of W longitudinally polarised W bosons. boson polarization. Indeed, the scattering of longitudinally-polarized vector bosons is maximally sensitive to any new phenomena in the Higgs sector. The effects searched are expected to be very small, therefore sophisticated

analysis strategies and tools are required. Starting from a purely theoretical approach developed in [172], we highlighted its limits, confirming the expectations reported by the authors themselves in [45]. With this in mind, we subsequently developed several strategies, in order to extract information about Ws polarization, and we developed a technique to experimentally determine its reference system. We analysed the VBS process in semi-leptonic and fully leptonic channels, starting with a kinematic study and then exploiting Deep Neural Network techniques. We finally tried to approach VBS studies by means of Quantum Computing and Quantum Machine Learning. This last part represents the most innovative part of this work and it covered the largest fraction of my PhD work.

Outcomes from these studies could be particularly relevant for experimental measurements which exploit the full Run 2 dataset and for the upcoming high-luminosity phase of the LHC. Already now, part of our phenomenological studies and tools has been adapted in searches for Higgs boson pairs from the ATLAS collaboration [173]. The Montecarlo routines developed in our work [37] have been used in [174] to apply different kinematic selections.
Concerning the polarisation extraction, we got promising results for the longitudinal components in terms of classification power and fit. Deep Neural Network approach shows a better performance with respect to a classic kinematic analysis. This is evident, in particular, for the classifier used to directly discriminate between the different polarized events, both for the semi-leptonic and the fully-leptonic cases, whereas the DNN model that perform the reconstruction of the W angular distribution is less precise. The main results have been published in [92] and they represent the phenomenological part of this thesis work, that is taken as a benchmark, for example in [175]. At the current ATLAS luminosity, unfortunately, LL polarisation fraction would not be measurable because of the limited statistics, but in the future we expect VBS process to reach enough sensitivity for this study.
Considering the Quantum Computing approach to VBS semi-leptonic channel, in the present work we found comparable performances between classic SVM, QSVM and DNN methodologies for the binary problem of the W rest frame reconstruction. We evidenced, instead, lower performance for the direct polarization classifier despite the similar parametrization tested. At the present time and on this particular use case, we could observe pros and cons about adopting quantum computation: the quantum approach does not show deterioration of the performance with increasing the number of features, and it converges quickly with few samples, whereas a bottleneck is present with a big dataset because of the quantum algorithm architecture. Unfortunately, could not evidence here any aspect of a possible quantum advantage.

Looking ahead at future prospects, the effort of the scientific community could be split in different directions. On the theoretical side, a full glory Montecarlo generator which implements full NLO QCD and EW corrections would be of extreme utility. On the phenomenological side, studies to further

enhance our capability to extract information about gauge boson polarizations would be desirable. To this extent, there is room for improving and increasing the interconnection between standard ML techniques and QC. The extensive usage of these innovative methods would eventually lead to a possible quantum advantage and they could therefore open frontiers to unexplored discoveries.

This PhD work is the result of an industrial doctoral path collaboration between the University of Pavia and IBM Italia (Executive PhD). This project aims at exploiting the connection between fundamental research actors, like the Universities, and an environment where scientific discoveries are strengthened and applied to the industrial scenario. This PhD work has been developed within the framework of a COST Action supported by the EU Framework Programme Horizon 2020 and named *Vector Boson Scattering Coordination and Action Network* (VBSCan). The European Community has appointed the study of the VBS, the process under study in this thesis works, as one of the top priorities for the LHC phase 2 era.

Appendix 1

A.1 Discrete Symmetry

The simplest example of SSB is a self-interacting real scalar field filling the vacuum with discrete Z_2 symmetry, whose Lagrangian is:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \varphi\right)^2 - V(\varphi) \qquad \text{with} \quad V(\varphi) \equiv \mu^2\varphi^2 + \lambda\varphi^4 \quad \text{and} \quad \lambda > 0$$

$\mathcal{L}$ is invariant under the transformation $\varphi \leftrightarrow -\varphi$. Since the vacuum φ_0 is defined as the value where $V(\varphi_0)$ has a minimum, for φ_0, corresponding $\mu^2 > 0$ and $\mu^2 < 0$, respectively:

$$\frac{\partial V}{\partial \varphi} = (2\mu^2 + 4\lambda\varphi^2)\varphi = 0 \quad \Rightarrow \quad \begin{cases} a)\ \mu^2 > 0 & \varphi_0 = 0 \\ b)\ \mu^2 < 0 & \varphi_0^2 = \frac{-\mu^2}{2\lambda} \end{cases}$$

Case a) is the unbroken phase that corresponds to the usual definition $\varphi_0 = 0$; the Lagrangian describes a single (symmetric) vacuum state with a scalar field with mass $\mu\sqrt{2}$ (for a scalar field the mass term is $-\frac{1}{2}m^2\varphi^2$) and self-interacting in a four−particle vertex.
Case b) is the broken phase that presents two equivalent vacuum states for $\varphi_0 = \pm\sqrt{\frac{-\mu^2}{2\lambda}}$.
Since it is $\varphi_0 \neq 0$, one has to redefine the scalar field φ as a small oscillation η around the vacuum, as in Eq. (24), which can be arbitrarily chosen as $\varphi_0 = +\sqrt{\frac{-\mu^2}{2\lambda}}$, giving rise to a *symmetry breaking* since the potential $V(\varphi)$ and consequently the Lagrangian $\mathcal{L}$ are no more symmetric in φ or (η). However it is only realized in a different way (it is not a symmetry of the vacuum):

$$\begin{aligned} \text{Wigner-Weyl realization:} \qquad & \varphi = \eta + \varphi_0 \\ \text{Nambu-Goldstone realization:} \qquad & \langle\eta\rangle = 0 \end{aligned} \tag{24}$$

$$\begin{aligned} V(\eta) &= \left(\mu^2 + 6\lambda\varphi_0^2\right)\eta^2 + \left(2\mu^2 + 4\lambda\varphi_0^2\right)\varphi_0\eta + O(\eta^3, \eta^4, \text{cost}) \\ &= -2\mu^2\eta^2 + O(\eta^3, \eta^4, \text{cost}) \end{aligned} \tag{25}$$

where the terms in η^3 e η^4 are negligible and the constant term is not relevant.

We left the mass term $-2\mu^2\eta^2$ corresponding to the mass value $m_\eta = \sqrt{-4\mu^2}$. By choosing $V(\varphi)$ to have a non-zero vacuum value, the scalar field η appears to acquire a mass. This is the simplest way in which the SSB could be described, the redefinition of the vacuum and the choice of the Lagrangian are the main steps of the mechanism together with the field transformation connecting to the group of symmetry considered.
One important theorem connected with the SSB we will introduce later on is the Goldstone theorem. It assert that in the attempting to generate massive gauge bosons, the SSB gives rise to a mass-less scalar particles.

Despite this choice and the redefinition of the Lagrangian under the previous local gauge transformation, mass-less Goldstone boson is still present. One way out, therefore, is to change the local gauge transformation, including the generator of the transformation, such as the charge q in the case of $U(1)_q$ symmetry group, and the Lagrangian accordingly by choosing the phase $\beta(x)$ in such a way that a new field $h(x)$ (which replaces $\eta(x)$) is real and described a massive particle ($\equiv$ *Higgs boson*). This is the *Higgs mechanism*: the Goldstone boson gets canceled from the theory and the corresponding degree of freedom contributes to the mass of the gauge bosons, in other words, the Goldstone boson is replaced by one more polarization state. For what concerned the $SU(2)_L \otimes U(1)_Y$ group of local gauge transformation one has to apply the same procedure with some additional complication of the involved mathematics.
An attractive feature of the SSB is that the same Higgs doublet, which generates the W and Z masses, is sufficient to give mass to the leptons and quarks.

Appendix 2

B.1 W Amplitude Estimation

In this section we want to calculate the amplitude of a single tree level W decay. Let's consider W+ boson decay represented in Fig. (21).

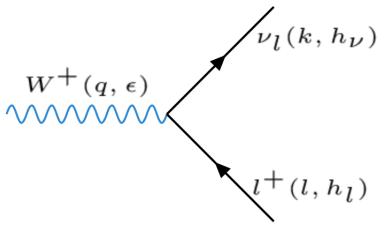


Figure 21

The calculation is performed in the *W rest frame*, using the helicity amplitude method by Hagiwara–Zeppenfeld [176] which is based on the explicit representation of Dirac matrices and Weyl spinors. In this method the polarization vectors of a spin–1 boson with momentum q and mass M can be written as:

$$\epsilon^\mu(\hat{q}, \lambda = \pm 1) = \frac{1}{\sqrt{2}}(0, \mp\cos\alpha\cos\beta + i\sin\beta, \mp\cos\alpha\sin\beta - i\cos\beta, \pm\sin\alpha)$$
$$\epsilon^\mu(\hat{q}, \lambda = 0) = \frac{1}{M}(|\vec{q}|, E_q\sin\alpha\cos\beta, E_q\sin\alpha\sin\beta, E_q\cos\alpha)$$

where $\hat{q} = \sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha$ and λ is the boson helicity.
Under Standard Model hypothesis the amplitude is given by:

$$\mathcal{M} = \left(\frac{-ig_{ew}}{\sqrt{2}}\right)\bar{u}_{h_\nu}(k)\gamma_\nu\left(\frac{1-\gamma^5}{2}\right)\epsilon^\mu_\lambda(q)v_{h_l}(l) \tag{26}$$

where h_ν,h_l are the fermions helicities. We can then simplify the previous equation to the following:

$$\mathcal{M} = \left(\frac{-ig_{ew}}{\sqrt{2}}\right)u^\dagger\not{\epsilon}v_- \tag{27}$$

with this notation:

$$
\begin{aligned}
u_{-}(k,h=-) &= \frac{(E_k|\mathbf{k}|)^{1/2}}{(2|\mathbf{k}|(|\mathbf{k}|+k_z))^{1/2}}\begin{pmatrix}-k_x+ik_y\\ |\mathbf{k}|+k_z\end{pmatrix}\\
v_{-}(l,h=+) &= \frac{(E_l|\mathbf{l}|)^{1/2}}{(2|\mathbf{l}|(|\mathbf{l}|+l_z))^{1/2}}\begin{pmatrix}-l_x+il_y\\ |\mathbf{l}|+l_z\end{pmatrix}\\
\not{\epsilon}_{-} &= \begin{pmatrix}\epsilon_0+\epsilon_z & \mp\epsilon_x-i\epsilon_y\\ \pm\epsilon_x+i\epsilon_y & \epsilon_0-\epsilon_z\end{pmatrix}
\end{aligned}
\tag{28}
$$

In the W's rest frame and considering massless final state fermions the neutrino momentum is $k = E(1, \sin\theta\cos\phi, \sin\theta\sin\phi, \cos\phi)$ and $k = -l$. In this condition, considering a W boson with momentum along the positive z–axis, the polarization vectors are:

$$
\begin{aligned}
\epsilon^{\mu}_{\pm} &= \frac{1}{\sqrt{2}}(0, \mp 1, -i, 0) \qquad \text{transverse;}\\
\epsilon^{\mu}_{0} &= (0,0,0,1) \qquad\qquad \text{longitudinal.}
\end{aligned}
$$

Now we are ready to evaluate the final amplitude for different choices of the initial boson polarization vector with the following convention: (θ, ϕ) are the lepton angles relative to the boson direction defined above. In case of longitudinal polarization[12]:

$$
\mathcal{M}(\lambda = 0) = ig_{ew}\sqrt{2}E\sin\theta \tag{29}
$$

while, in case of transverse polarization,

$$
\mathcal{M}(\lambda = \pm 1) = ig_{ew}E(1 \pm \cos\theta)e^{\pm i\phi} \tag{30}
$$

B.2 Off–diagonal terms in differential cross–section

The matrix element can be generalized as:

$$
|\mathcal{M}|^2 = \sum_{\lambda\lambda'} c\lambda\lambda'\mathcal{M}^{*}_{\lambda}\mathcal{M}_{\lambda'} \tag{31}
$$

where $c_{\lambda\lambda'}$ are real weights for the polarization combinations. Combining Eq. (31) and the specific representation results obtained in Eq. (2.14) we have:

$$
\begin{aligned}
|\mathcal{M}|^2 = \big(g_{ew}E)^2(c_{00}2\sin\theta^2 &+ c_{LL}(1-\cos\theta)^2 + c_{RR}(1-\cos\theta)^2\\
&+\sqrt{2}\sin\theta(1-\cos\theta)(c_{0L}e^{-i\phi} + c_{L0}e^{i\phi})\\
&+\sqrt{2}\sin\theta(1+\cos\theta)(c_{0R}e^{-i\phi} + c_{R0}e^{i\phi})\\
&+(1-\cos\theta^2)(c_{LR}e^{-2i\phi} + c_{RL}e^{2i\phi})\big)
\end{aligned}
\tag{32}
$$

At this step we are ready to give a parameterization of the normalized differential cross–section including the phase–space and flux factors as in Eq. (1.34). The convention adopted here is:

[12] In case of W^- the only difference is in the sign of the transverse component.

- θ is the angle between the charged lepton and the $\hat{z}$ axis;
- ϕ is the azimuth angle between the charged lepton and the $\hat{x}$ axis;
- $\hat{x}$ axis lies in the plane defined by the incoming protons in the W rest frame pointing in the flight direction of the proton which forms the smaller angle with the $\hat{z}$ axis.

$$\begin{aligned}\frac{1}{\sigma}\frac{d\sigma}{d\phi d\cos\theta} =& \frac{1}{16\pi}\big((1+\cos\theta)^2 + a_0\frac{1}{2}(1-3\cos\theta^2) + a_1\sin 2\theta\cos\phi \\ &+ a_2\frac{1}{2}\sin\theta^2\cos 2\phi + a_3\sin\theta\cos\phi + a_4\cos\theta \\ &+ a_5\sin\theta\sin\phi + a_6\sin 2\theta\sin\phi + a_7\sin\theta^2\sin 2\phi\big)\end{aligned} \tag{33}$$

Integrating Eq. (33) over the entire $\cos\theta$ range, we finally get the differential distribution:

$$\boxed{\frac{1}{\sigma}\frac{d\sigma}{d\phi} = \frac{3}{16\pi}\left(\frac{8}{3} + \frac{2}{3}a_2\cos 2\phi + \frac{\pi}{2}a_3\cos\phi + \frac{\pi}{2}a_5\sin\phi + \frac{4}{3}a_7\sin 2\phi\right).} \tag{34}$$

B.3 Equivalence Theorem

Here we want to give more details about the *equivalence theorem* used in Sec. (2.4). Given the fact that Goldstone bosons are related to one of the degrees of freedom of massive vector fields, at high energies, the EW symmetry is "restored", so that the gauge bosons become effectively a massless degree of freedom (2 transverse polarizations) while the Goldstone boson carries all the information on the longitudinal polarization. In Fig. (22) we have a diagrammatic representation of the theorem. In the high energy limit, the scattering problem formulation moves from a complete description with the massive vector field to a more simplified treatment with the massless or low-energy effective field theory representation. We see the change in the boson mediators, from massive $V(W,Z)$ ones to πs. This allows for greatly simplify the calculations of scattering amplitude, and it is formalized in the following.

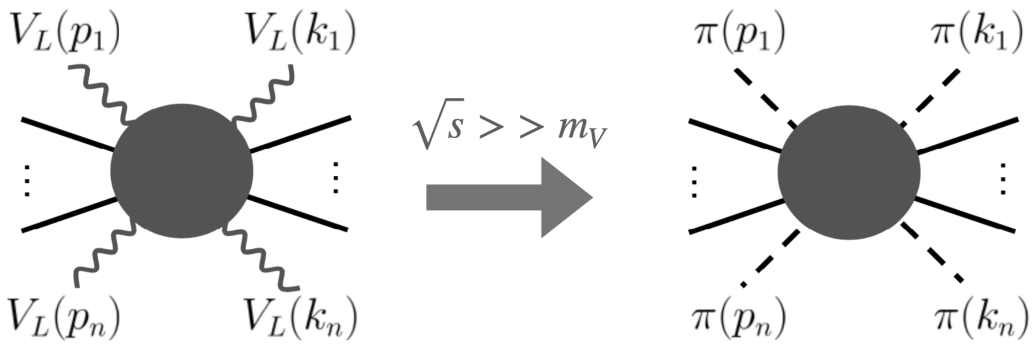


Figure 22: Schematic representation of Goldstone approximation where under the specified condition, the gauge bosons become effectively massless.

Equivalence Theorem

In the limit of large center of mass energy $\sqrt{s} >> m_V$, the scattering amplitude for a process involving n initial and m final longitudinally polarized vector bosons is related to the corresponding amplitude for the scalar Goldstone bosons by:

$$\begin{aligned}\mathcal{A}\left((V_L(p_1)) + \cdots + (V_L(p_n)) + X \to (V_L(k_1)) + \cdots + (V_L(k_m)) + Y\right) \\ = (-i)^{n+m}\mathcal{A}\left(\pi(p_1) + \ldots \pi(p_n) + X \to \pi(k_1) + \ldots \pi(k_m) + Y\right) \\ \left(1 + O\left(\frac{m_V}{\sqrt{s}}\right)\right)\end{aligned} \tag{35}$$

B.4 Polarization and Lorentz Transformation

In Chapter 3 we discussed about the origin of polarization components from QFT principles, in App(B.1) we evaluated the cross section for a simple W boson decay focusing on polarization fraction. Putting together all of this insights we want to highlight an important concept. Though the polarization vectors are built as Lorentz four–vectors, longitudinal and transverse polarization vectors depend on the reference frames in which they are defined, as well as the polarization fractions. This means that if we calculate them in a certain frame A (W momentum in the A frame) and then we boost them in another frame B, we obtain a different result than we would obtain by boosting the W momentum from A to the B frame and then building the polarization vectors in it.

Considering a W boson moving along the positive z–axis in the laboratory frame we can switch its momentum into another reference frame by means of

a Lorentz boost:

$$p^{\mu} = (E, 0, 0, p) \rightarrow p^{'\mu} = \Lambda^{\mu}_{\nu} p^{\mu} = (\gamma E, \gamma\beta E, 0, p)$$

where:

$$\Lambda^{\mu}_{\nu} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The longitudinal polarization vector is given by $\epsilon^{\mu} = (p/M, 0, 0, E/M)$, while the one built in the boosted frame is:

$$\rightarrow \epsilon^{'\mu} = \frac{1}{M}\left(\sqrt{p^2 + (\gamma\beta E)^2}, \gamma E \frac{\gamma\beta E}{\sqrt{p^2 + (\gamma\beta E)^2}}, 0, \gamma E \frac{p}{\sqrt{p^2 + (\gamma\beta E)^2}}\right)$$

In case we boost ϵ^{μ} we obtain a completely different expression with respect to the longitudinal polarization built starting from boosted momentum:

$$\epsilon^{''\mu} = \frac{1}{M}(\gamma p, \gamma\beta p, 0, E).$$

Appendix 3

C.1 PDF uncertainties calculation for VBS

In this part we will integrate what has been discussed in Sec. (3.5).
From a general point of view, recent progress achieved by the different PDF groups includes:

- better assessment of uncertainties (e.g. different groups now agree at the 1σ level where data is available);
- exploit new information coming from LHC Run I and Run II measurements;
- progress in tools and methods to include these data in the fits;
- inclusion of PDFs for photons and top quarks.

However, still some issues are present, on which the collaboration are working hard:

- homogeneity of the data to include in the fits (and how to deal with incompatible data);
- enhance relevance of some data (reduce effect of inconsistent data sets);
- heavy-quark treatment and masses;
- parametrization for PDFs (theoretical bias, reduced in Neural Network PDFs);
- inclusion of theoretical improvement (e.g. resummation) for some observables;
- unphysical behaviour close to $x = 0$ and $x = 1$;
- meaning of uncertainties;
- α_s as external input or fitted with PDFs.

We said that there exists two leading representations of the uncertainties of PDF: Monte Carlo [83] (MC) and Hessian [84].
The MC representation contains an ensemble of replicas, which are the instances of uncertain PDF parameters, sampled according to the Gaussian distribution, around its central values. Central PDF is the average of the PDF set, while the PDF uncertainty is its standard deviation. Under the assumption of Gaussian distribution of the cross-sections, obtained from the PDF set, the same treatment can be applied to the cross-section:

$$\delta^{\text{pdf}}\sigma = \sqrt{\frac{1}{N_{\text{mem}}-1}\sum_{k=1}^{N_{\text{mem}}}(\sigma^{(k)}-\langle\sigma\rangle)^2}, \qquad \langle\sigma\rangle = \frac{1}{N_{\text{mem}}}\sum_{k=1}^{N_{\text{mem}}}\sigma^{(k)}, \tag{36}$$

where $\sigma^{(k)}$ represents the cross-section calculated from the k-th member of the set and N_{mem} represents the number of members. On the other hand, when the distribution of cross-sections differs significantly from Gaussian, it is better to employ the robust measures. This means that 16% of the largest and 16% of the smallest cross-sections fall out of the 68% C.L. interval, the rest of them lies within. Symmetric PDF uncertainty is calculated as

$$\delta^{\text{pdf}}\sigma = \frac{\sigma^{(N_{\text{mem}}84/100)} - \sigma^{(N_{\text{mem}}16/100)}}{2}; \qquad \sigma^{(1)} \leq \sigma^{(2)} \leq \ldots \leq \sigma^{(N_{\text{mem}})}. \tag{37}$$

This kind of treatment better account for the outliers. To roughly estimate how much does the specific distribution resemble to the Gaussian, we can compare the values calculated from Eq. (36) and Eq. (37). If they give similar results, the distribution can be approximated by the Gaussian and uncertainty from Eq. (36) can be taken as a result. In the opposite case, it is better to use Eq. (37). In this study we performed the Shapiro-Wilk Gaussianity test.
As opposed to a MC PDF set we have an Hessian PDF set. Central PDF cannot be calculated from its members and have to be appended separately to a set. In `LHAPDF` sets, this is always the first member. The equation for the calculation of the cross-section uncertainty is:

$$\delta^{\text{pdf}}\sigma = \sqrt{\sum_{k=1}^{N_{\text{mem}}}(\sigma^{(k)}-\sigma^{(0)})^2}, \tag{38}$$

where $\sigma^{(0)}$ represents the cross-section of the central PDF and N_{mem} represents the number of members of the set (without central).
Because it is necessary to generate $N_{\text{mem}}+1$ MC samples to evaluate the cross-section uncertainty due to PDF uncertainty ($N_{\text{mem}}+3$ for the combined uncertainty), the reweighting of the events is very useful, which can significantly reduce the computational time. For LO reweighting of the sample, produced with the PDF set A, to the PDF set B, the corresponding weights are

$$w_{A\to B} = \frac{f_1^B(x_1,Q)f_2^B(x_2,Q)}{f_1^A(x_1,Q)f_2^A(x_2,Q)}, \tag{39}$$

where the $f_i^A(x_i, Q)$ are the PDFs of the two incoming partons from the sample A and $f_i^B(x_i, Q)$ are the PDFs of the two incoming partons from the sample B. For the NLO reweighting the use of built-in generator routines is necessary to calculate the weights. NLO reweighting is, for example, implemented in the generators `MadGraph5_aMC@NLO` [177], `POWHEG` [178], `Sherpa` [179].

The `LHAPDF` library [86] offers a wide range of PDF sets from different groups. Because different PDF sets are based on different experimental data and use different assumptions, it is better to take into account more than just one PDF set. For this purpose statistical combinations can be used. This approach is competitive with the older PDF envelope method, by which the PDF uncertainty has to be calculated for each input PDF set in order to obtain uncertainty bands for the PDF. The combined PDF uncertainty is defined as the envelope of these bands. On the other hand, the statistical combination of the PDF sets already contains characteristics of several input sets. Uncertainty has to be calculated only on the statistical combination, while the obtained uncertainty takes into account the uncertainties of all the input sets. The PDF uncertainty from the envelope method is often overestimated, but this is not the case for the statistical combination method. The `LHAPDF` library currently contains one statistical combination: PDF4LHC15 [87]. This is a combination of CT14 [88], MMHT2014 [89], and NNPDF3.0 [90] sets. If the input PDF sets are represented by MC replicas, construction of the statistical combination is pretty straightforward. Namely, the input PDF sets, with the equal number of replicas, can simply be merged into one larger PDF set [180], which is called the prior set. The only input PDF set into the PDF4LHC15, represented with the MC replicas, is NNPDF3.0, while the CT14 and MMHT2014 are Hessian representations. Therefore, the first step of construction of the statistical combination is transformation of the Hessian representations into MC. This is done by sampling along directions of each eigenvector, according to the corresponding eigenvalue, assuming Gaussian distribution.

The authors of the statistical combination PDF4LHC15 tested different sizes of the prior set: $N_{\text{rep}} = 300$, $N_{\text{rep}} = 900$, and $N_{\text{rep}} = 1800$. After the comparison of the central values and the uncertainty for different partons, they concluded that there is (small) difference, between the sets with 300 and 900 replicas, while the differences between 900 and 1800 are completely negligible [181]. Therefore prior set with 900 replicas has been adopted.

Since the prior set is too large to be handled in the analysis, reduction methods are applied to it. PDF4LHC15 is distributed in three options, which use different reduction algorithms as indicated in the brackets:

- Monte-Carlo (CMC-PDFs),
- Hessian with 30 eigenvectors (META-PDFs),
- Hessian with 100 eigenvectors (MCH-PDFs).

In the case of Monte-Carlo compression method (CMC-PDFs) the number of MC replicas is reduced in a way, that agreement between the certain statistical

PDF set	Reduct. algo.	Pert. order	Uncertainty type	N_{mem}	α_S var.
PDF4LHC15_nlo_100	MCH-PDFs	NLO	symhessian	100	No
PDF4LHC15_nlo_30	META-PDFs	NLO	symhessian	30	No
PDF4LHC15_nlo_mc	CMC-PDFs	NLO	replicas	100	No
PDF4LHC15_nlo_30_pdfas	META-PDFs	NLO	symhessian+as	32	Yes
PDF4LHC15_nlo_mc_pdfas	CMC-PDFs	NLO	replicas+as	102	Yes
PDF4LHC15_nnlo_100	MCH-PDFs	NNLO	symhessian	100	No
PDF4LHC15_nnlo_30	META-PDFs	NNLO	symhessian	30	No
PDF4LHC15_nnlo_mc	CMC-PDFs	NNLO	replicas	100	No
PDF4LHC15_nnlo_30_pdfas	META-PDFs	NNLO	symhessian+as	32	Yes
PDF4LHC15_nnlo_mc_pdfas	CMC-PDFs	NNLO	replicas+as	102	Yes

Table 2: The list of the main PDF4LHC15 sets, in 5 quark flavour scheme, available in the LHAPDF library. Currently α_S variation is available only in the META-PDFs and the CMC-PDFs sets.

characteristics of the prior and reduced set is the best [182]. For transformation of the MC representation into the Hessian, the authors of PDF4LHC15 used two different methods. META-PDFs is based on fitting a functional form to the set of MC replicas, while MCH-PDFs uses singular value decomposition, followed by principal component analysis [183, 184]. META-PDFs offers more accurate description of the prior set at smaller numbers of eigenvectors in the reduced set, while the opposite holds for the larger reduced sets. For this reason the representation with 30 eigenvectors is reduced with the META-PDFs and the representation with 100 eigenvectors is reduced with the MCH-PDFs. The latter is more appropriate, when the highest accuracy is desired, on the other hand the META-PDFs is useful, when simple statistical analysis is the priority. The main PDF4LHC15 sets, available in the LHAPDF library are gathered in Table 2.

Appendix 4

D.1 MAOS solution

Equality $m_{\mathrm{T}}^{(1)}(\vec{p}_1) = m_{\mathrm{T}}^{(2)}(\vec{p}_{\mathrm{T}} - \vec{p}_1)$ defines a curve in the phase space of the first neutrino trial momentum $\vec{p}_1$. This curve equation can be expressed analytically in the polar coordinates (p_1, φ). Let us introduce the parameters:

$$\begin{aligned} \varphi_0 &= \arccos\left(\frac{\vec{p}_{\mathrm{T}}^{\,\ell\ell} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}{|\vec{p}_{\mathrm{T}}^{\,\ell\ell}||\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}|}\right), \\ c &= (\vec{p}_{\mathrm{T}}^{\,e} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})^2 - |\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}|^2 |\vec{p}_{\mathrm{T}}^{\,e}|^2, \end{aligned}$$

where $\vec{p}_{\mathrm{T}}^{\,\ell\ell} = \vec{p}_{\mathrm{T}}^{\,\mu} + \vec{p}_{\mathrm{T}}^{\,e}$. The intersection curve can be written as:

$$\begin{aligned} f(\varphi) &= \left(|\vec{p}_{\mathrm{T}}^{\,\mu}|^2 + |\vec{p}_{\mathrm{T}}^{\,\ell\ell}|^2 \cos^2\varphi - 2|\vec{p}_{\mathrm{T}}^{\,\mu}||\vec{p}_{\mathrm{T}}^{\,\ell\ell}|\cos\varphi - |\vec{p}_{\mathrm{T}}^{\,e}|^2\right), \\ g(\varphi) &= \big(2(|\vec{p}_{\mathrm{T}}^{\,\mu}| - |\vec{p}_{\mathrm{T}}^{\,\ell\ell}|\cos\varphi)\vec{p}_{\mathrm{T}}^{\,e} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \\ &\qquad + 2|\vec{p}_{\mathrm{T}}^{\,e}|^2 |\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}|\cos(\varphi + \varphi_0)\big), \\ p_1 &= \frac{-g(\varphi) \pm \sqrt{g(\varphi)^2 - 4cf(\varphi)}}{2f(\varphi)}. \end{aligned}$$

m_{T2} can be searched numerically by defining an arbitrary number of equidistant points in φ and searching for a minimal value of $m_{\mathrm{T}}^{(1)}(\vec{p}_1(\varphi))$. This point defines $\vec{p}_{\mathrm{T}}^{\,\nu_\mu}$, $\vec{p}_{\mathrm{T}}^{\,\nu_e}$ can be obtained from the minimization bond as $\vec{p}_{\mathrm{T}}^{\,\nu_e} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} - \vec{p}_{\mathrm{T}}^{\,\nu_\mu}$.

D.2 Neural Network Details

In this section we provide additional details about NN used in this work. The first ingredient is the number of events used in the different phase, namely training, validation and test or evaluation phase. This information is given in Table (5) and (4). The training variables used for the semi-leptonic case in training are:

- μ: $E, p_x, p_y, p_z, p_t, \eta, \phi$,
- ν_μ:p_x, p_y

- q_1: E, p_x, p_y, p_z
- q_2:,E, p_x, p_y, p_z,
- q_3:,E, p_x, p_y, p_z,
- q_4:,E, p_x, p_y, p_z,
- $event0\_mass, event1\_mass$;

where q_i refers to outgoing quarks and $event0\_mass$, $event1\_mass$ refer to the fact that we have two possibilities for the neutrino z-component of momenta and energy, eventually leading to a double measurable invariant masses of the event.
The training variables used for the full-leptonic case, instead are:

- μ: $E, p_x, p_y, p_z, p_t, \eta, \phi$,
- ν_μ:p_x, p_y, η, ϕ
- e: $E, p_x, p_y, p_z, p_t, \eta, \phi$,
- ν_e:p_x, p_y, η, ϕ
- q_3:,E, p_x, p_y, p_z,
- q_4:,E, p_x, p_y, p_z,
- pt_vv,pv_xx,pv_yy,$mt2$;

where pt_vv,pv_xx,pv_yy refer to the sum of both neutrinos missing momenta and components x and y. While, when MAOS variables are included, $mt2$ come from reconstructed neutrino transverse momenta ($p_{\mathrm{T}}^{\nu_\ell}$ from Eq. (4.6)).

In this study deep neural networks have been used. In general, we pointed out that increasing the number of parameters can help the neural network to accomplish more difficult classification tasks because its approximation power increases [185]. The number of parameter and a simple NN structure is shown in Fig. (23), where we can see the detailed structure of the NN in terms of number of hidden layers (8), neurons for each layer (200) and the output layer with dimesion 3, which is the case of the multiclassifier (LL,LT,TT). The number of parameters managed by the network is 286.403.

Before starting the training procedure, a Gaussian standardization technique is applied to the whole training dataset: each variable is centered in 0 and scaled to unit variance independently. This procedures helps the neural network training so that each variables can contribute equally to the classification regardless its measurement unit.

```
model.summary()None
--
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 200)               4400
_________________________________________________________________
dropout_1 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_2 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_2 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_3 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_3 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_4 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_4 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_5 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_5 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_6 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_6 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_7 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_7 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_8 (Dense)              (None, 200)               40200
_________________________________________________________________
dropout_8 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_9 (Dense)              (None, 3)                 603
=================================================================
Total params: 286,403
Trainable params: 286,403
Non-trainable params: 0
_________________________________________________________________
```

Figure 23: Print of a model summary provided by Keras Python package.

Table 3: Number of events used for each semi-leptonic sample in the training, validation and test datasets.

sample	training	validation	test
unpolarized	530219	530219	706959
longitudinal	264052	264052	352070
transverse	269769	269769	359692
full-computation	535236	535236	713648

Table 4: Number of events used for each full-leptonic sample in the training, validation and test datasets.

sample	training	validation	test
longitudinal	299999	300000	300000
transverse	299999	300000	300000
full-computation	NA	NA	400000
WW mix polarization	1199997	1200000	1200000
ZZ mix polarization	113933	116152	116152

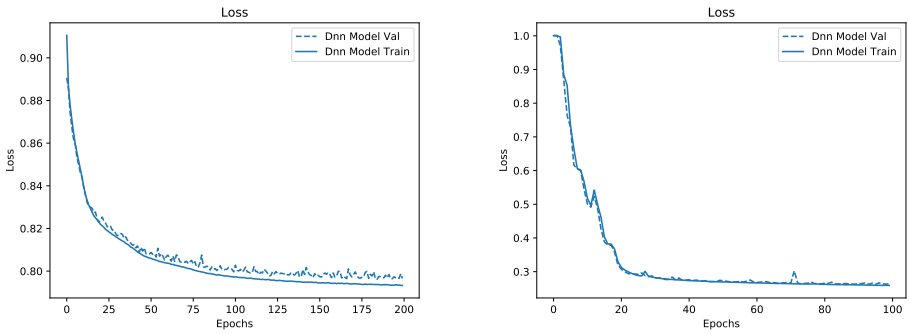


Figure 24: Shape of the loss function related to a specific neural network topology used for regression in the semi-leptonic and fully-leptonic channel. The decreasing of loss values a minimum value can be appreciated so that represents a good convergence of the neural network parameters according to the number of hidden layers, neurons and batch-size.

For sake of completeness we include representative plots of loss functions in Fig. (24) and Fig. (25) to highlight the pros and cons according to different DNN configurations. In Fig. (26) we can appreciate an example of performance worsening as neurons increase in the topology. The NN starts learning in the very beginning (as a function of epochs) but all of a sudden the loss function explodes due to the fact that the weights are not under control any more and we lost completely any chance of convergence. In the last plot instead, we have a standard example of overtraining behaviour in contrast with the right plot of Fig. (24).

All activities have been performed in two main steps. First all the events were generated on a PC computing cluster using the HTCondorCE that allows the user to parallelize all the computation splitting it into several nodes.
Secondly, once all the events were created, we converted them into Root files to perform common data preparation and finally we converted them into pandas arrays to get train, test and validation dataset. These data was migrated to an IBM Power9 machine to perform all the training and evaluation steps. IBM Power System AC922 is widely used in general context for its performance for analytics, artificial intelligence (AI), and modern HPC. The Power AC922 is engineered to be the most powerful training platform available, providing the data and compute-intensive infrastructure needed to deliver faster time to insights. Data scientists get to use their favourite tools without sacrificing speed and performance, while IT leaders get the proven infrastructure to accelerate time to value.[13].

[13] `https://www.ibm.com/us-en/marketplace/power-systems-ac922`

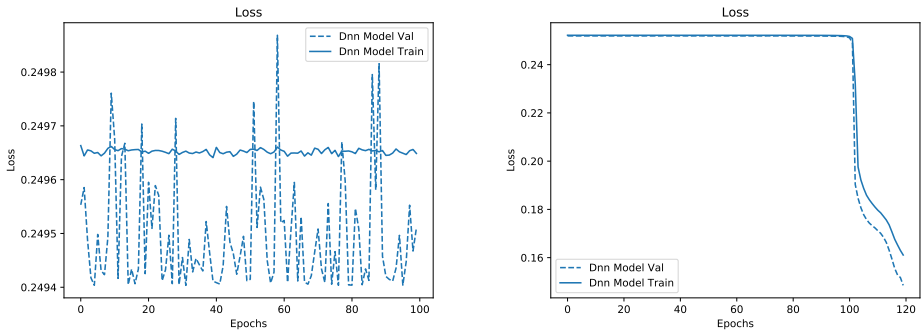


Figure 25: Shape of the loss function related to a specific neural network topology used for regression in the semi-leptonic and full-leptonic channel. In this case the loss function is not converging, in particular in the first case the network is not improving its parameters at all giving a flat loss function and relatively unstable results for the validation dataset. In the second case the neural networking is not improving and only after a relatively high number of epochs the loss is starting to evolve. For the optimal treatment of each particular convergence scenario an early stop automation can be employed so that the number of training epochs in each case can be tuned according to the learning rate and loss function behavior.

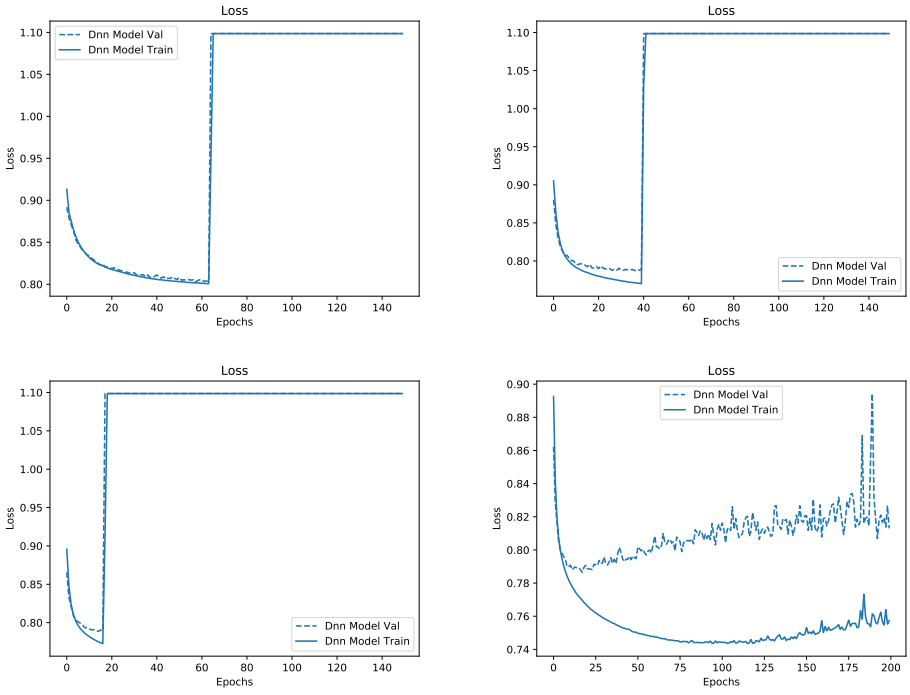


Figure 26: Example of performance worsening in case of neurons increase in the topology, top figures and figure in the bottom left. Last plot is a classic example of NN overtraining behaviour.

D.3 PHANTOM generator (in a nutshell)

PHANTOM is an event generator which is capable of simulating any set of reactions with six partons in the final state at pp colliders at order $O(\alpha_{\rm EW}^6)$ and $O(\alpha_{\rm EW}^4\alpha_s^2)$ including possible interferences between the two sets of diagrams. This includes all purely electroweak contributions as well as all contributions with one virtual or two external gluons.
In the semi-leptonic channel we define the channel `mu vm_` whereas for full leptonic case the channel is defined in the configuration card with this parsing syntax `mu_ e_ vm ve`[14].
PHANTOM performs an exact calculation of the matrix-element at tree level, without using any production times decay approach for the W bosons interaction. PHANTOM implements a method called On Shell Projection (OSP) to compute the amplitude of resonant contributions projecting (in the numerator) the four momenta of the decay particles on shell, as shown in Eq. (41).

$$A_f = \sum_\lambda \frac{A^\mu_{p,RES}(p,k)\epsilon^\lambda_\mu \epsilon^{\lambda *}_\nu A^\nu_d(k,q)}{k^2 - M_W^2 + i\Gamma_W M_W} + A_{NONRES} \tag{40}$$

$$\to \sum_\lambda \frac{A^\mu_{p,RES}(p,k_{OSP})\epsilon^\lambda_{\mu,OSP} \epsilon^{\lambda *}_{\nu,OSP} A^\nu_d(k_{OSP},q_{OSP})}{k^2 - M_W^2 + i\Gamma_W M_W} \tag{41}$$

With this method, PHANTOM conserves the total four-momentum of the WW system, the direction of the two W bosons in the WW center-of-mass frame and the direction of each charged lepton in his W center-of-mass frame. In the case of semi leptonic single polarization/resonant study we select the polarization for only one boson whereas the other boson is kept unpolarized.

The formula can be seen as an on shell production times decay modulated by the Breit-Wigner width shape with all exact spin correlations. between decay products of any unstable resonance. This gives an excellent framework to make comprehensive analyses of VBS signal together with its irreducible backgrounds. The amplitude are computed in a semi–automatic way, through the program PHACT, which is based on the helicity formalism. Due to final states with a large number of particles there may be a huge number of Feynman diagrams which contribute to a process; despite being a challenging computational issue, the calculation can be simplified: the flavour mixing matrix in the quark sector (CKM matrix) is taken diagonal and light quarks and leptons can be considered massless. These approximations induce weak–interactions universality (symmetry under family conjugation), CP conservation and crossing symmetry, which strongly reduce the number of amplitudes to evaluate. Employing exact matrix elements at the tree level, we cannot separate different physical processes just selecting the final state, but we can impose kinematic cuts to extract the physics of interest from a given final state.
Events generation takes place in two steps: the program produces one or more

[14] Where the _ stands for antiparticle.

Table 5: Cross sections at LO accuracy obtained with exact matrix elements at $\mathcal{O}(\alpha^6$ contribution for the process: $pp \to jjjj\mu^+\nu_\mu$ (semi-leptonic) and $pp \to jj\mu^+\nu_\mu e^+\nu_e$ (full-leptonic) at $\sqrt{s} = 13$ TeV. The uncertainties shown refer to the estimated statistical errors of the Monte Carlo integrations.

process	computation	σ[pb]	systematics
semi-leptonic	unpolarized	$1.912\,E-002$	$1.010\,E-005$
semi-leptonic	longitudinal	$5.695\,E-003$	$7.265\,E-006$
semi-leptonic	transverse	$1.352\,E-002$	$6.527\,E-006$
semi-leptonic	full-computation	$2.215\,E-002$	$9.742\,E-006$
full-leptonic	W_LW_L polarization	$1.092\,E-004$	$1.717\,E-007$
full-leptonic	W_TW_T polarization	$8.373\,E-004$	$6.575\,E-007$
full-leptonic	W_TW_L polarization	$2.945\,E-004$	$3.815\,E-007$
full-leptonic	full-computation	$1.516\,E-003$	$1.156\,E-006$

integration grids depending on the number of channels required for a good mapping of phase space for the selected process. This is called *thermalization*. It determines the relative weight of each channel in the multichannel integration and it produces a first instance of PHANTOM space grids, one per channel. Then the program generates unweighted events for a set of processes, which is specified by the user, using the previously produced grids. All the integration grids for each selected process must be included for a meaningful generation. In this thesis we used mainly 2 version of the package and we contributed to discover and discuss physics related bugs, in particular the problem, which gave some rare events in which the color flow was inconsistent with the mothers assignment.
In fact, version 1.6 differs from the previous one 1.3.2, in particular for the introduction of the possibility of computing only resonant W and/or Z diagrams, computing polarized cross sections for W and/or Z diagrams, projecting on shell the above contributions.

Generation Summary

In the main text we underline the fact that only tree level electroweak contributions $\mathcal{O}(\alpha^6$ are considered, and partonic processes involving b and t quarks are neglected. If we include in the analysis events which involve at least one $b(b)$–quark as a tagging jet, the kinematics of the final state is altered. The inclusion of b–quark as a tagging jet add many more channels in the generations, when only electroweak contributions are considered. This probability is even enhanced if also QCD contributions are included. Allowing b–quarks in the final state induces the propagation of top quarks.

References

[1] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.321

[2] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.508

[3] J. Goldstone, "Field theories with superconductor solutions," *Nuovo Cimento*, vol. 19, no. CERN-TH-118, pp. 154–164, Aug 1960. [Online]. Available: http://cds.cern.ch/record/343400

[4] F. Mandl and G. Shaw, *Quantum Field Theory*, 1 1985.

[5] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, M. Dragicevic, J. Erö, A. Escalante Del Valle, M. Flechl, R. Frühwirth, and et al., "A measurement of the higgs boson mass in the diphoton decay channel," *Physics Letters B*, vol. 805, p. 135425, Jun 2020. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2020.135425

[6] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and 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*, vol. 716, no. 1, p. 1–29, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.020

[7] M. Tanabashi and e. a. Hagiwara, "Review of particle physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.98.030001

[8] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, O. AbouZeid, N. Abraham, H. Abramowicz, H. Abreu, R. Abreu, Y. Abulaiti, B. Acharya, L. Adamczyk, D. Adams, J. Adelman, S. Adomeit, T. Adye, A. Affolder, and N. Woods, "Measurements of the higgs boson

production and decay rates and constraints on its couplings from a combined atlas and cms analysis of the lhc pp collision data at $\sqrt{s} = 7$ and 8 tev," *Journal of High Energy Physics*, vol. 2016, 08 2016.

[9] G. Dissertori, I. G. Knowles, and M. Schmelling, *Quantum chromodynamics: high energy experiments and theory*, ser. International series of monographs on physics. Oxford: Oxford Univ. Press, 2002. [Online]. Available: https://cds.cern.ch/record/579254

[10] M. Peskin and D. Schroeder, *An Introduction To Quantum Field Theory*, ser. Frontiers in Physics. Avalon Publishing, 1995. [Online]. Available: https://books.google.it/books?id=EVeNNcslvX0C

[11] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language," *Nuclear Physics B*, vol. 126, no. 2, pp. 298 – 318, 1977. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0550321377903844

[12] A. Denner, J.-N. Lang, M. Pellen, and S. Uccirati, "Nlo qcd and ew corrections to processes involving off-shell top quarks," *PoS*, vol. RADCOR2017, p. 060, 2017.

[13] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, O. AbouZeid, N. Abraham, H. Abramowicz, and et al., "Measurement of the $w \pm z$ boson pair-production cross section in pp collisions at $s = 13$ tev with the atlas detector," *Physics Letters B*, vol. 762, p. 1–22, Nov 2016. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2016.08.052

[14] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger, "High precision determination of the gluon fusion higgs boson cross-section at the lhc," 2016.

[15] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, "Higgs boson pair production via gluon fusion at n3lo in qcd," *Physics Letters B*, vol. 803, p. 135292, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0370269320300964

[16] F. Dulat, B. Mistlberger, and A. Pelloni, "Precision predictions at n^3LO for the higgs boson rapidity distribution at the lhc," *Phys. Rev. D*, vol. 99, p. 034004, Feb 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.99.034004

[17] F. A. Dreyer and A. Karlberg, "Vector-boson fusion higgs production at three loops in qcd," *Phys. Rev. Lett.*, vol. 117, p. 072001, Aug 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.072001

[18] S. Dittmaier *et al.*, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," 1 2011.

[19] L. Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, S. Alekhin, J. Alwall, E. Bagnaschi, A. Banfi, J. Blumlein, S. Bolognesi, N. Chanon, T. Cheng, L. Cieri, A. Cooper-Sarkar, M. Cutajar, S. Dawson, G. Davies, N. Filippis, and G. Zanderighi, "Handbook of lhc higgs cross sections: 2. differential distributions," 01 2012.

[20] S. Heinemeyer, C. Mariotti, G. Passarino, R. Tanaka, J. R. Andersen, P. Artoisenet, E. A. Bagnaschi, A. Banfi, T. Becher, F. U. Bernlochner, S. Bolognesi, P. Bolzoni, R. Boughezal, D. Buarque, J. Campbell, F. Caola, M. Carena, F. Cascioli, N. Chanon, T. Cheng, S. Y. Choi, A. David, P. De Aquino, G. Degrassi, D. Del Re, A. Denner, H. Van Deurzen, S. Diglio, B. Di Micco, R. Di Nardo, S. Dittmaier, M. Duhrssen, R. K. Ellis, G. Ferrera, N. Fidanza, M. Flechl, D. De Florian, S. Forte, R. Frederix, S. Frixione, S. Gangal, Y. Gao, M. V. Garzelli, D. Gillberg, P. Govoni, M. Grazzini, N. Greiner, J. Griffiths, A. . V. Gritsan, C. Grojean, D. C. Hall, C. Hays, R. Harlander, R. Hernandez-Pinto, S. Hoche, J. Huston, T. Jubb, M. Kadastik, S. Kallweit, A. Kardos, L. Kashif, N. Kauer, H. Kim, R. Klees, M. Kramer, F. Krauss, A. Laureys, S. Laurila, S. Lehti, Q. Li, S. Liebler, X. Liu, E. Logan, G. Luisoni, M. Malberti, F. Maltoni, K. Mawatari, F. Maierhoefer, H. Mantler, S. Martin, P. Mastrolia, O. Mattelaer, J. Mazzitelli, B. Mellado, K. Melnikov, P. Meridiani, D. J. Miller, E. Mirabella, S. O. Moch, P. Monni, N. Moretti, A. Muck, M. Muhlleitner, P. Musella, P. Nason, C. Neu, M. Neubert, C. Oleari, J. Olsen, G. Ossola, T. Peraro, K. Peters, F. Petriello, G. Piacquadio, C. T. Potter, S. Pozzorini, K. Prokofiev, I. Puljak, M. Rauch, D. Rebuzzi, L. Reina, R. Rietkerk, A. Rizzi, Y. Rotstein-Habarnau, G. P. Salam, G. Sborlini, F. Schissler, M. Schonherr, M. Schulze, M. Schumacher, F. Siegert, P. Slavich, J. M. Smillie, O. Stal, J. F. Von Soden-Fraunhofen, M. Spira, I. W. Stewart, F. J. Tackmann, P. T. E. Taylor, D. Tommasini, J. Thompson, R. S. Thorne, P. Torrielli, F. Tramontano, N. V. Tran, Z. Trocsanyi, M. Ubiali, M. V. Acosta, T. Vickey, A. Vicini, W. J. Waalewijn, D. Wackeroth, C. Wagner, J. R. Walsh, J. Wang, G. Weiglein, A. Whitbeck, C. Williams, J. Yu, G. Zanderighi, M. Zanetti, M. Zaro, P. M. Zerwas, C. Zhang, T. J. . E. Zirke, and S. Zuberi, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties," CERN, Research Report, 2013. [Online]. Available: http://hal.in2p3.fr/in2p3-00850353

[21] D. Florian *et al.*, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," vol. 2/2017, 10 2016.

[22] V. Özcan, "Vector boson scattering at ATLAS," *Balk. Phys. Lett.*, vol. 17, pp. 155–159, 2009.

[23] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, M. Dragicevic, J. Erö, A. Escalante Del Valle, R. Frühwirth, M. Jeitler, and et al., "Measurements of production cross sections of wz and same-sign ww boson pairs in association with two jets in proton-proton collisions at s=13tev," *Physics Letters B*, vol. 809, p. 135710, Oct 2020. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2020.135710

[24] "Standard Model Summary Plots Spring 2020," CERN, Geneva, Tech. Rep. ATL-PHYS-PUB-2020-010, May 2020. [Online]. Available: http://cds.cern.ch/record/2718937

[25] B. Biedermann, A. Denner, and M. Pellen, "Complete nlo corrections to w+w+ scattering and its irreducible background at the lhc," *Journal of High Energy Physics*, vol. 2017, no. 10, Oct 2017. [Online]. Available: http://dx.doi.org/10.1007/JHEP10(2017)124

[26] J. Bendavid, F. Caola, V. Ciulli, R. Harlander, G. Heinrich, J. Huston, S. Kallweit, S. Prestel, E. Re, K. Tackmann, J. Thaler, K. Theofilatos, J. R. Andersen, J. Bellm, N. Berger, D. Bhatia, B. Biedermann, S. Bräuer, D. Britzger, A. G. Buckley, R. Camacho, G. Chachamis, S. Chatterjee, X. Chen, M. Chiesa, J. R. Currie, A. Denner, F. Dreyer, F. Driencourt-Mangin, S. Forte, M. V. Garzelli, T. Gehrmann, S. Gieseke, E. W. N. Glover, P. Gras, N. Greiner, C. Gütschow, C. Gwenlan, M. Heil, M. Herndon, V. Hirschi, A. H. Hoang, S. Höche, A. Huss, S. P. Jones, D. Kar, A. Karlberg, Z. Kassabov, M. Kerner, J. Klappert, S. Kuttimalai, J. N. Lang, A. Larkoski, J. M. Lindert, P. Loch, K. Long, L. Lönnblad, G. Luisoni, A. Maier, P. Maierhöfer, D. Maître, S. Marzani, J. A. McFayden, I. Moult, M. Mozer, S. Mrenna, B. Nachman, D. Napoletano, C. Pandini, A. Papaefstathiou, M. Pellen, L. Perrozzi, J. Pires, S. Plätzer, S. Pozzorini, S. Quackenbush, K. Rabbertz, M. Rauch, C. Reuschle, P. Richardson, A. G.-D. Ridder, G. Rodrigo, J. Rojo, R. Röntsch, L. Rottoli, D. Samitz, T. Samui, G. Sborlini, M. Schönherr, S. Schumann, L. Scyboz, S. Seth, H. S. Shao, A. Siódmok, P. Z. Skands, J. M. Smillie, G. Soyez, P. Sun, M. R. Sutton, F. J. Tackmann, S. Uccirati, S. Weinzierl, E. Yazgan, C. P. Yuan, and F. Yuan, "Les houches 2017: Physics at tev colliders standard model working group report," 2018.

[27] A. Denner, S. Dittmaier, P. Maierhöfer, M. Pellen, and C. Schwan, "Qcd and electroweak corrections to wz scattering at the lhc," *Journal of High Energy Physics*, vol. 2019, no. 6, Jun 2019. [Online]. Available: http://dx.doi.org/10.1007/JHEP06(2019)067

[28] A. Denner, R. Franken, M. Pellen, and T. Schmidt, "Nlo qcd and ew corrections to vector-boson scattering into zz at the lhc," 2020.

[29] B. Biedermann, A. Denner, and M. Pellen, "Large electroweak corrections to vector-boson scattering at the large hadron collider," *Physical Review Letters*, vol. 118, no. 26, Jun 2017. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.118.261801

[30] S. Dittmaier and C. Schwan, "Non-factorizable photonic corrections to resonant production and decay of many unstable particles," *The European Physical Journal C*, vol. 76, no. 3, Mar 2016. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-016-3968-1

[31] S. Dawson, "The Effective W Approximation," *Nucl. Phys.*, vol. B249, pp. 42–60, 1985.

[32] M. J. Duncan, G. L. Kane, and W. W. Repko, "W W Physics at Future Colliders," *Nucl. Phys.*, vol. B272, pp. 517–559, 1986.

[33] A. Denner and S. Pozzorini, "One-loop leading logarithms in electroweak radiative corrections," *The European Physical Journal C*, vol. 18, no. 3, p. 461–480, Jan 2001. [Online]. Available: http://dx.doi.org/10.1007/s100520100551

[34] A. Denner and T. Hahn, "Radiative corrections to $w^+w^- \to w^+w^-$ in the electroweak standard model," *Nuclear Physics B*, vol. 525, no. 1-2, p. 27–50, Aug 1998. [Online]. Available: http://dx.doi.org/10.1016/S0550-3213(98)00287-9

[35] M. Chiesa, A. Denner, J.-N. Lang, and M. Pellen, "An event generator for same-sign W-boson scattering at the LHC including electroweak corrections," *Eur. Phys. J. C*, vol. 79, no. 9, p. 788, 2019.

[36] B. Jäger and G. Zanderighi, "Nlo corrections to electroweak and qcd production of w+w+ plus two jets in the powheg box," *Journal of High Energy Physics*, vol. 2011, no. 11, Nov 2011. [Online]. Available: http://dx.doi.org/10.1007/JHEP11(2011)055

[37] A. Ballestrero, B. Biedermann, S. Brass, A. Denner, S. Dittmaier, R. Frederix, P. Govoni, M. Grossi, B. Jäger, A. Karlberg, and et al., "Precise predictions for same-sign w-boson scattering at the lhc," *The European Physical Journal C*, vol. 78, no. 8, Aug 2018. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-018-6136-y

[38] G. Bozzi, B. Jäger, C. Oleari, and D. Zeppenfeld, "Next-to-leading-order qcd corrections to W^+z and W^-z production via vector-boson fusion," *Phys. Rev. D*, vol. 75, p. 073004, Apr 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.75.073004

[39] F. Campanario, M. Kerner, L. D. Ninh, and D. Zeppenfeld, "wz production in association with two jets at next-to-leading order in qcd," *Phys. Rev. Lett.*, vol. 111, p. 052003, Aug 2013. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.111.052003

[40] B. Jäger, A. Karlberg, and J. Scheller, "Parton-shower effects in electroweak wzjj production at the next-to-leading order of qcd," *The European Physical Journal C*, vol. 79, no. 3, Mar 2019. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-019-6736-1

[41] B. Jäger, A. Karlberg, and G. Zanderighi, "Electroweak zzjj production in the standard model and beyond in the powheg-box v2," *Journal of High Energy Physics*, vol. 2014, no. 3, Mar 2014. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2014)141

[42] F. Campanario, M. Kerner, L. D. Ninh, and D. Zeppenfeld, "Next-to-leading order qcd corrections to zz production in association with two jets," *Journal of High Energy Physics*, vol. 2014, no. 7, Jul 2014. [Online]. Available: http://dx.doi.org/10.1007/JHEP07(2014)148

[43] M. Szleper, "The higgs boson and the physics of ww scattering before and after higgs discovery," 2014.

[44] 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*, vol. 2010, no. 3, Mar 2010. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2010)082

[45] A. Ballestrero, E. Maina, and G. Pelliccioli, "W boson polarization in vector boson scattering at the lhc," *Journal of High Energy Physics*, vol. 2018, no. 3, Mar 2018. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2018)170

[46] E. Maina, "Vector boson polarizations in the decay of the standard model higgs," 2020.

[47] A. Kadeer, J. G. Körner, and U. Moosbrugger, "Helicity analysis of semileptonic hyperon decays including lepton-mass effects," *The European Physical Journal C*, vol. 59, no. 1, p. 27–47, Nov 2008. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-008-0801-5

[48] A. Ballestrero, A. Belhouari, G. Bevilacqua, V. Kashkan, and E. Maina, "PHANTOM: A Monte Carlo event generator for six parton final states at high energy colliders," *Comput. Phys. Commun.*, vol. 180, pp. 401–417, 2009.

[49] A. Denner *et al.*, "Predictions for all processes $e^+e^- \to 4\,\text{fermions} + \gamma$," *Nucl. Phys.*, vol. B560, pp. 33–65, 1999.

[50] A. Denner and S. Dittmaier, "The complex-mass scheme for perturbative calculations with unstable particles," *Nucl. Phys. Proc. Suppl.*, vol. 160, pp. 22–26, 2006, [,22(2006)].

[51] A. Ballestrero, "PHACT: Helicity amplitudes for present and future colliders," in *High energy physics and quantum field theory. Proceedings, 14th International Workshop, QFTHEP'99, Moscow, Russia, May 27-June 2, 1999*, 1999, pp. 303–309.

[52] A. Ballestrero and E. Maina, "A New method for helicity calculations," *Phys. Lett.*, vol. B350, pp. 225–233, 1995.

[53] G. P. Lepage, "A new algorithm for adaptive multidimensional integration," *J. Comput. Phys.*, vol. 27, p. 192, 1978.

[54] J. Alwall, A. Ballestrero, and P. B. et all., "A standard format for les houches event files," *Computer Physics Communications*, vol. 176, no. 4, pp. 300 – 304, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010465506004164

[55] W. Kilian, T. Ohl, and J. Reuter, "WHIZARD: Simulating Multi-Particle Processes at LHC and ILC," *Eur. Phys. J.*, vol. C71, p. 1742, 2011.

[56] M. Moretti, T. Ohl, and J. Reuter, "O'Mega: An Optimizing matrix element generator," pp. 1981–2009, 2001.

[57] W. Kilian, T. Ohl, J. Reuter, and C. Speckner, "QCD in the Color-Flow Representation," *JHEP*, vol. 10, p. 022, 2012.

[58] B. Chokoufé Nejad, W. Kilian, J. M. Lindert, S. Pozzorini, J. Reuter, and C. Weiss, "NLO QCD Predictions for off-shell $t\bar{t}$ and $t\bar{t}H$ Production and Decay at a Linear Collider," *JHEP*, vol. 12, p. 75, 2016.

[59] S. Frixione, "A General approach to jet cross-sections in QCD," *Nucl. Phys.*, vol. B507, pp. 295–314, 1997.

[60] S. Actis *et al.*, "Recursive generation of one-loop amplitudes in the Standard Model," *JHEP*, vol. 04, p. 037, 2013.

[61] J. Reuter, B. Chokoufé Nejad, A. Hoang, W. Kilian, M. Stahlhofen, T. Teubner, and C. Weiss, "Automation of NLO processes and decays and POWHEG matching in WHIZARD," *J. Phys. Conf. Ser.*, vol. 762, no. 1, p. 012059, 2016.

[62] W. Kilian, T. Ohl, J. Reuter, and M. Sekulla, "High-Energy Vector Boson Scattering after the Higgs Discovery," *Phys. Rev.*, vol. D91, p. 096007, 2015.

[63] S. Dittmaier, P. Maierhöfer, C. Schwan, and F. Siegert, "Vector-boson Scattering at the LHC," in: *PoS* RADCOR2017.

[64] S. Dittmaier and M. Roth, "LUSIFER: A LUcid approach to six FERmion production," *Nucl. Phys.*, vol. B642, pp. 307–343, 2002.

[65] A. Denner, S. Dittmaier, and L. Hofer, "COLLIER - A fortran-library for one-loop integrals," *PoS*, vol. LL2014, p. 071, 2014.

[66] ——, "Collier: a fortran-based complex one-loop library in extended regularizations," *Computer Physics Communications*, vol. 212, 04 2016.

[67] S. Dittmaier and C. Schwan, "Non-factorizable photonic corrections to resonant production and decay of many unstable particles," *Eur. Phys. J.*, vol. C76, no. 3, p. 144, 2016.

[68] K. Arnold *et al.*, "VBFNLO: A Parton level Monte Carlo for processes with electroweak bosons," *Comput. Phys. Commun.*, vol. 180, pp. 1661–1670, 2009.

[69] J. Alwall *et al.*, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP*, vol. 07, p. 079, 2014.

[70] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP*, vol. 06, p. 029, 2002.

[71] R. Frederix *et al.*, "Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties," *JHEP*, vol. 02, p. 099, 2012.

[72] S. Actis *et al.*, "RECOLA: REcursive Computation of One-Loop Amplitudes," *Comput. Phys. Commun.*, vol. 214, pp. 140–173, 2017.

[73] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[74] J. Bellm, S. Gieseke, D. Grellscheid, S. Plätzer, M. Rauch, C. Reuschle, P. Richardson, P. Schichtel, M. H. Seymour, A. Siódmok, and et al., "Herwig 7.0/herwig++ 3.0 release note," *The European Physical Journal C*, vol. 76, no. 4, Apr 2016. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-016-4018-8

[75] S. Ovyn, X. Rouby, and V. Lemaitre, "DELPHES, a framework for fast simulation of a generic collider experiment," 3 2009.

[76] A. Denner and S. Dittmaier, "The complex-mass scheme for perturbative calculations with unstable particles," *Nuclear Physics B - Proceedings Supplements*, vol. 160, p. 22–26, Oct 2006. [Online]. Available: http://dx.doi.org/10.1016/j.nuclphysbps.2006.09.025

[77] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[78] ——, "Fastjet user manual," *The European Physical Journal C*, vol. 72, no. 3, Mar 2012. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-012-1896-2

[79] D. Rainwater, R. Szalapski, and D. Zeppenfeld, "Probing color-singlet exchange inz+2-jet events at the cern lhc," *Physical Review D*, vol. 54, no. 11, p. 6680–6689, Dec 1996. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.54.6680

[80] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al., "Electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV," *The European Physical Journal C*, vol. 78, no. 7, Jul 2018. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-018-6049-9

[81] S. Bräuer, A. Denner, M. Pellen, M. Schönherr, and S. Schumann, "Fixed-order and merged parton-shower predictions for ww and wwj production at the lhc including nlo qcd and ew corrections," *Journal of High Energy Physics*, vol. 2020, p. 159, 10 2020.

[82] S. Bethke, "World Summary of α_s (2012)," 2012, [Nucl. Phys. Proc. Suppl.234,229(2013)].

[83] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo, "Unbiased determination of the proton structure function F(2)**p with faithful uncertainty estimation," *JHEP*, vol. 03, p. 080, 2005.

[84] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, "Uncertainties of predictions from parton distribution functions. 2. The Hessian method," *Phys. Rev.*, vol. D65, p. 014013, 2001.

[85] S. S. SHAPIRO and M. B. WILK, "An analysis of variance test for normality (complete samples)†," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 12 1965. [Online]. Available: https://doi.org/10.1093/biomet/52.3-4.591

[86] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rufenacht, M. Schönherr, and G. Watt, "LHAPDF6: parton density access in the LHC precision era," *Eur. Phys. J.*, vol. C75, p. 132, 2015.

[87] M. Botje *et al.*, "The PDF4LHC Working Group Interim Recommendations," 2011.

[88] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, "New parton distribution functions from a global analysis of quantum chromodynamics," *Phys. Rev.*, vol. D93, no. 3, p. 033006, 2016.

[89] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs," *Eur. Phys. J.*, vol. C75, no. 5, p. 204, 2015.

[90] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali, "A Determination of parton distributions with faithful uncertainty estimation," *Nucl. Phys.*, vol. B809, pp. 1–63, 2009, [Erratum: Nucl. Phys.B816,293(2009)].

[91] R. Bellan, J. Beyer, C. Bittrich, G. Boldrini, I. Brivio, L. S. Bruni, D. B. Franzosi, C. Charlot, V. Ciulli, R. Covarelli, D. Giljanovic, G. Gonella, P. Govoni, P. Gras, M. Grossi, T. Herrmann, J. Kalinowski, A. Karlberg, K. Kallonen, E. Kasimi, A. K. Topaksu, B. Kersevan, H. Kirschenmann, M. Kobel, K. Kordas, A. Leisos, D. Lelas, P. Lenzi, A. Li, K. Lohwasser, N. Lorenzo-Martinez, J. Manjarres, D. Mapelli, A. Marantis, M. Marchegiani, A. Mascellani, I. Maznas, H. Mildner, M. U. Mozer, M. Neukum, J. Novak, G. Ortona, K. Ozdemir, M. Pellen, G. Pelliccioli, C. Petridou, S. Plätzer, C. Quaggio, M. Rauch, D. Rebuzzi, J. Reuter, D. Sampsonidou, S. Schramm, A. Siodmok, M. Slawinska, M. Szleper, A. Tarabini, C. I. Thorburn, S. Todt, S. E. Tzamarias, A. Tsirigotis, D. Valsecchi, L. Wuest, and M. Zaro, "Vbscan thessaloniki 2018 workshop summary," 2019.

[92] M. Grossi, J. Novak, B. Kerševan, and D. Rebuzzi, "Comparing traditional and deep-learning techniques of kinematic reconstruction for polarization discrimination in vector boson scattering," *The European Physical Journal C*, vol. 80, no. 12, Dec 2020. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-020-08713-1

[93] "Observation of electroweak production of a same-sign w boson pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *Phys. Rev. Lett.*, vol. 123, p. 161801, Oct 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.123.161801

[94] "Observation of electroweak production of same-sign w boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev. Lett.*, vol. 120, p. 081801, Feb 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.081801

[95] C. Goy, "VBF and VBS measurements at HL-LHC/HE-LHC," Dec 2019. [Online]. Available: https://cds.cern.ch/record/2704641

[96] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park, "Measuring superparticle masses at hadron collider using the transverse mass kink," *JHEP*, vol. 02, p. 035, 2008.

[97] L. Sonnenschein, "Analytical solution of ttbar dilepton equations," *Phys. Rev.*, vol. D73, p. 054015, 2006, [Erratum: Phys. Rev.D78,079902(2008)].

[98] K. Choi, S. Choi, J. S. Lee, and C. B. Park, "Reconstructing the Higgs boson in dileptonic W decays at hadron collider," *Phys. Rev.*, vol. D80, p. 073010, 2009.

[99] B. Efron and T. Hastie, *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, ser. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2016.

[100] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," *Nature Communications*, vol. 5, no. 1, p. 4308, 2014. [Online]. Available: https://doi.org/10.1038/ncomms5308

[101] P. Sadowski, P. Baldi, and D. Whiteson, "Searching for higgs boson decay modes with deep learning," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2393–2401, 01 2014.

[102] J. Lee, N. Chanon, A. Levin, J. Li, M. Lu, Q. Li, and Y. Mao, "Polarization fraction measurement in zz scattering using deep learning," *Phys. Rev. D*, vol. 100, p. 116010, Dec 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.100.116010

[103] J. Searcy, L. Huang, M.-A. Pleier, and J. Zhu, "Determination of the ww polarization fractions in $pp \to W^{\pm}W^{\pm}jj$ using a deep machine learning technique," *Phys. Rev. D*, vol. 93, p. 094033, May 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.93.094033

[104] V. K. Ojha, A. Abraham, and V. Snášel, "Metaheuristic design of feedforward neural networks: A review of two decades of research," *Engineering Applications of Artificial Intelligence*, vol. 60, pp. 97 – 116, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197617300234

[105] L. F. A. Wessels and E. Barnard, "Avoiding false local minima by proper initialization of connections," *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 899–905, 1992.

[106] K. . Toh, "Deterministic global optimization for fnn training," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 33, no. 6, pp. 977–983, 2003.

[107] Machine learning website. [Online]. Available: https://towardsdatascience.com/coding-deep-learning-for-beginners-linear-regression-gradient-descent-fcd5e0fc077d

[108] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, p. 807–814.

[109] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[110] T. Fawcett, "Introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 06 2006.

[111] M. Fabbrichesi, M. Pinamonti, A. Tonero, and A. Urbano, "Vector boson scattering at the lhc: A study of the $ww \to ww$ channels with the warsaw cut," *Phys. Rev. D*, vol. 93, p. 015004, Jan 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.93.015004

[112] R. P. Feynman, "Simulating physics with computers," *Int. J. Theor. Phys*, vol. 21, no. 6/7, 1982.

[113] M. A. Nielsen, I. L. Chuang, and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000, no. 2.

[114] V. Zhirnov, R. III, J. Hutchby, and G. Bourianoff, "Limits to binary logic switch scaling - a gedanken model," *Proceedings of the IEEE*, vol. 91, pp. 1934 – 1939, 12 2003.

[115] J. Gubernatis, N. Kawashima, and P. Werner, *Quantum Monte Carlo Methods*. Cambridge University Press, 2016.

[116] S. Montangero, *Introduction to tensor network methods: numerical simulations of low-dimensional many-body quantum systems*. Cham: Springer, 2018. [Online]. Available: https://cds.cern.ch/record/2650847

[117] D. Deutsch and R. Penrose, "Quantum theory, the church–turing principle and the universal quantum computer," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 400, no. 1818, pp. 97–117, 1985. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1985.0070

[118] D. P. DiVincenzo, "The physical implementation of quantum computation," *Fortschritte der Physik: Progress of Physics*, vol. 48, no. 9-11, pp. 771–783, 2000.

[119] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134.

[120] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 1996, pp. 212–219.

[121] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Physical Review Letters*, vol. 103, no. 15, Oct 2009. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.103.150502

[122] W. Guan, G. Perdue, A. Pesah, M. Schuld, K. Terashi, S. VALLECORSA, and j.-r. vlimant, "Quantum machine learning in high energy physics," *Machine Learning: Science and Technology*, Oct 2020. [Online]. Available: http://dx.doi.org/10.1088/2632-2153/abc17d

[123] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: https://doi.org/10.22331/q-2018-08-06-79

[124] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, and et al., "Quantum optimization using variational algorithms on near-term quantum devices," *Quantum Science and Technology*, vol. 3, no. 3, p. 030503, Jun 2018. [Online]. Available: http://dx.doi.org/10.1088/2058-9565/aab822

[125] R. Loredo, *Learn Quantum Computing with Python and IBM Q Experience: A Hands-On Introduction to Quantum Computing and Writing Your Own Quantum Programs with Python*. Packt Publishing, 2020. [Online]. Available: https://books.google.it/books?id=jKy3zQEACAAJ

[126] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, "Experimental comparison of two quantum computing architectures," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, p. 3305–3310, Mar 2017. [Online]. Available: http://dx.doi.org/10.1073/pnas.1618020114

[127] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, "Validating quantum computers using randomized model circuits," *Physical Review A*, vol. 100, no. 3, Sep 2019. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.100.032328

[128] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Optimized driving of superconducting artificial atoms for improved single-qubit gates," *Phys. Rev. A*, vol. 82, p. 040305, Oct 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.82.040305

[129] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, L. Bello, Y. Ben-Haim, J. Bucher, A. Chen, C. Chen *et al.*, "Qiskit: An open-source framework for quantum computing (2019)."

[130] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, "Quantum algorithms for electronic structure calculations: Particle-hole hamiltonian and optimized wave-function expansions," *Phys. Rev. A*, vol. 98, p. 022322, Aug 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.022322

[131] J. D. Guimarães, C. Tavares, L. S. Barbosa, and M. I. Vasilevskiy, "Simulation of nonradiative energy transfer in photosynthetic systems using a quantum computer," *Complexity*, vol. 2020, p. 1–12, Sep 2020. [Online]. Available: http://dx.doi.org/10.1155/2020/3510676

[132] F. Fillion-Gourdeau, S. MacLean, and R. Laflamme, "Algorithm for the solution of the dirac equation on digital quantum computers," *Physical Review A*, vol. 95, no. 4, Apr 2017. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.95.042343

[133] N. Klco, M. J. Savage, and J. R. Stryker, "Su(2) non-abelian gauge field theory in one dimension on digital quantum computers," *Physical Review D*, vol. 101, no. 7, Apr 2020. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.101.074512

[134] C. W. Bauer, W. A. de Jong, B. Nachman, and D. Provasoli, "A quantum algorithm for high energy physics simulations," 2019.

[135] M. J. Cervia, A. B. Balantekin, S. N. Coppersmith, C. W. Johnson, P. J. Love, C. Poole, K. Robbins, and M. Saffman, "Exactly solvable model as a testbed for quantum-enhanced dark matter detection," 2020.

[136] W. A. de Jong, M. Metcalf, J. Mulligan, M. Płoskoń, F. Ringer, and X. Yao, "Quantum simulation of open quantum systems in heavy-ion collisions," 2020.

[137] K. Yeter-Aydeniz, R. C. Pooser, and G. Siopsis, "Practical quantum computation of chemical and nuclear energy levels using quantum imaginary time evolution and lanczos algorithms," 2019.

[138] S.-N. Sun, M. Motta, R. N. Tazhigulov, A. T. K. Tan, G. K.-L. Chan, and A. J. Minnich, "Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution," 2020.

[139] T. V. Zache, F. Hebenstreit, F. Jendrzejewski, M. K. Oberthaler, J. Berges, and P. Hauke, "Quantum simulation of lattice gauge theories using wilson fermions," *Quantum Science and Technology*, vol. 3, no. 3, p. 034010, jun 2018. [Online]. Available: https://doi.org/10.1088%2F2058-9565%2Faac33b

[140] S. V. Mathis, G. Mazzola, and I. Tavernelli, "Toward scalable simulations of lattice gauge theories on quantum computers," *Physical Review D*, vol. 102, no. 9, Nov 2020. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.102.094501

[141] U. Las Heras, U. Alvarez-Rodriguez, E. Solano, and M. Sanz, "Genetic algorithms for digital quantum simulations," *Physical Review Letters*, vol. 116, no. 23, Jun 2016. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.116.230504

[142] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Phys. Rev. X*, vol. 7, p. 021050, Jun 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.7.021050

[143] N. Stamatopoulos, D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner, "Option pricing using quantum computers," *Quantum*, vol. 4, p. 291, Jul 2020. [Online]. Available: http://dx.doi.org/10.22331/q-2020-07-06-291

[144] S. Mugel, C. Kuchkovsky, E. Sanchez, S. Fernandez-Lorenzo, J. Luis-Hita, E. Lizaso, and R. Orus, "Dynamic portfolio optimization with real datasets using quantum processors and quantum-inspired tensor networks," 2020.

[145] K. Kaneko, K. Miyamoto, N. Takeda, and K. Yoshino, "Quantum speedup of monte carlo integration in the direction of dimension and its application to finance," 2020.

[146] A. Chiesa, F. Tacchino, M. Grossi, P. Santini, I. Tavernelli, D. Gerace, and S. Carretta, "Quantum hardware simulating four-dimensional inelastic neutron scattering," *Nature Physics*, vol. 15, no. 5, p. 455–459, Mar 2019. [Online]. Available: http://dx.doi.org/10.1038/s41567-019-0437-4

[147] F. Arute, K. Arya, R. Babbush, D. Bacon, J. Bardin, R. Barends, R. Biswas, S. Boixo, F. Brandao, D. Buell, B. Burkett, Y. Chen, Z. Chen,

B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, and J. Martinis, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, pp. 505–510, 10 2019.

[148] S. Bravyi, D. Gosset, and R. König, "Quantum advantage with shallow circuits," *Science*, vol. 362, no. 6412, p. 308–311, Oct 2018. [Online]. Available: http://dx.doi.org/10.1126/science.aar3106

[149] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, p. 195–202, Sep 2017. [Online]. Available: http://dx.doi.org/10.1038/nature23474

[150] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning," 2013.

[151] M. Schuld, M. Fingerhuth, and F. Petruccione, "Implementing a distance-based classifier with a quantum interference circuit," *EPL (Europhysics Letters)*, vol. 119, no. 6, p. 60002, sep 2017. [Online]. Available: https://doi.org/10.1209%2F0295-5075%2F119%2F60002

[152] E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," 2018.

[153] V. Havlícek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, p. 209–212, Mar 2019. [Online]. Available: http://dx.doi.org/10.1038/s41586-019-0980-2

[154] M. Schuld and N. Killoran, "Quantum machine learning in feature hilbert spaces," *Physical Review Letters*, vol. 122, no. 4, Feb 2019. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.122.040504

[155] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, Feb 2020. [Online]. Available: http://dx.doi.org/10.22331/q-2020-02-06-226

[156] D. Ristè, M. P. da Silva, C. A. Ryan, A. W. Cross, A. D. Córcoles, J. A. Smolin, J. M. Gambetta, J. M. Chow, and B. R. Johnson, "Demonstration of quantum advantage in machine learning," *npj Quantum Information*, vol. 3, no. 1, Apr 2017. [Online]. Available: http://dx.doi.org/10.1038/s41534-017-0017-3

[157] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879

[158] N. Wiebe, A. Kapoor, and K. Svore, "Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning," 2014.

[159] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner, and M. S. Kim, "Quantum generalisation of feedforward neural networks," *npj Quantum Information*, vol. 3, no. 1, Sep 2017. [Online]. Available: http://dx.doi.org/10.1038/s41534-017-0032-4

[160] F. Tacchino, C. Macchiavello, D. Gerace, and D. Bajoni, "An artificial neuron implemented on an actual quantum processor," *npj Quantum Information*, vol. 5, no. 1, Mar 2019. [Online]. Available: http://dx.doi.org/10.1038/s41534-019-0140-4

[161] F. Tacchino, P. Barkoutsos, C. Macchiavello, I. Tavernelli, D. Gerace, and D. Bajoni, "Quantum implementation of an artificial feed-forward neural network," *Quantum Science and Technology*, vol. 5, no. 4, p. 044010, Oct 2020. [Online]. Available: http://dx.doi.org/10.1088/2058-9565/abb8e4

[162] S. Mangini, F. Tacchino, D. Gerace, C. Macchiavello, and D. Bajoni, "Quantum computing model of an artificial neuron with continuously valued input data," *Machine Learning: Science and Technology*, vol. 1, no. 4, p. 045008, Oct 2020. [Online]. Available: http://dx.doi.org/10.1088/2632-2153/abaf98

[163] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," 2020.

[164] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R.* Springer Publishing Company, Incorporated, 2014.

[165] R. Johnson and D. Wichern, *Applied multivariate statistical analysis*, 6th ed. Upper Saddle River, NJ: Prentice Hall, 2013. [Online]. Available: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+330798693&sourceid=fbw_bibsonomy

[166] B. Scholkopf, A. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," 1996.

[167] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," *Nature*, vol. 567, no. 7749, p. 491–495, Mar 2019. [Online]. Available: http://dx.doi.org/10.1038/s41586-019-1040-7

[168] J. Romero, J. P. Olson, and A. Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," *Quantum Science and Technology*, vol. 2, no. 4, p. 045001, Aug 2017. [Online]. Available: http://dx.doi.org/10.1088/2058-9565/aa8072

[169] Summers, Sioni and Rose, Andrew, "Kalman filter track reconstruction on fpgas for acceleration of the high level trigger of the cms experiment at the hl-lhc," *EPJ Web Conf.*, vol. 214, p. 01003, 2019. [Online]. Available: https://doi.org/10.1051/epjconf/201921401003

[170] J. Chan, W. Guan, S. Sun, A. Z. Wang, S. L. Wu, C. Zhou, M. Livny, F. Carminati, and A. D. Meglio, "Application of Quantum Machine Learning to High Energy Physics Analysis at LHC using IBM Quantum Computer Simulators and IBM Quantum Computer Hardware," *PoS*, vol. LeptonPhoton2019, p. 049, 2019.

[171] K. Terashi, M. Kaneda, T. Kishimoto, M. Saito, R. Sawada, and J. Tanaka, "Event classification with quantum machine learning in high-energy physics," 2020.

[172] A. Ballestrero, E. Maina, and G. Pelliccioli, "W boson polarization in vector boson scattering at the LHC," *JHEP*, vol. 03, p. 170, 2018.

[173] "Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector," 2019.

[174] "Comparison of ATLAS and CMS VBS Monte Carlo simulation," 10 2020.

[175] J. Li, S. Yang, and R. Zhang, "Detecting anomaly in vector boson scattering," 10 2020.

[176] K. Hagiwara and D. Zeppenfeld, "Helicity Amplitudes for Heavy Lepton Production in e+ e- Annihilation," *Nucl. Phys. B*, vol. 274, pp. 1–32, 1986.

[177] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer, "MadGraph/MadEvent v4: The New Web Generation," *JHEP*, vol. 09, p. 028, 2007.

[178] C. Oleari, "The POWHEG-BOX," *Nucl. Phys. Proc. Suppl.*, vol. 205-206, pp. 36–41, 2010.

[179] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP*, vol. 02, p. 007, 2009.

[180] S. Forte, "Parton distributions at the dawn of the LHC," *Acta Phys. Polon.*, vol. B41, pp. 2859–2920, 2010.

[181] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," *J. Phys.*, vol. G43, p. 023001, 2016.

[182] S. Carrazza, J. I. Latorre, J. Rojo, and G. Watt, "A compression algorithm for the combination of PDF sets," *Eur. Phys. J.*, vol. C75, p. 474, 2015.

[183] J. Gao and P. Nadolsky, "A meta-analysis of parton distribution functions," *JHEP*, vol. 07, p. 035, 2014.

[184] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, and J. Rojo, "An Unbiased Hessian Representation for Monte Carlo PDFs," *Eur. Phys. J.*, vol. C75, no. 8, p. 369, 2015.

[185] L. Lista, "Statistical Methods for Data Analysis in Particle Physics," *Lect. Notes Phys.*, vol. 941, pp. 1–257, 2017.

List of Publications

Journal Articles

- A. Chiesa, F. Tacchino, M. Grossi, P. Santini, I. Tavernelli, D. Gerace and S. Carretta: **Quantum hardware simulating four-dimensional inelastic neutron scattering**. – *Nature Physics*, 15, 455 (2019), DOI: 10.1038/s41567-019-0437-4, Pre-print at arXiv:1809.07974.

- A. Ballestrero, P. Govoni, M. Grossi, D. Rebuzzi, et al.: **Precise predictions for same-sign W-boson scattering at the LHC**. – *Eur. Phys. J. C*, 78(8), Aug 2018.

- A. Ballestrero, P. Govoni, M. Grossi, D. Rebuzzi, et al.: **Vector boson scattering: Recent experimental and theory developments**. – *Reviews in Physics*, November 2018, Pages 44-63.

- M. Grossi, J. Novak, D. Rebuzzi, B. Kersevan: **Comparing Traditional and Deep-Learning Techniques of Kinematic Reconstruction for polarisation Discrimination in Vector Boson Scattering**. – *Eur. Phys. J. C*, 80(12), Dec 2020.

Conference Procedings

- F. Tacchino, M. Grossi, D. Gerace, A. Chiesa, P. Santini, S. Carretta, and I. Tavernelli: **Efficient Quantum Simulation of Dynamic Correlations on Superconducting Quantum Computers**. – *Conference on Lasers and Electro-Optics, OSA Technical Digest* (Optical Society of America, 2019), paper JTh5C.3 (postdeadline).
DOI: 10.1364/$CLEO\_AT$.2019.$JTh5C$.3.

Pre-print manuscripts

- A. Ballestrero, P. Govoni, M. Grossi, D. Rebuzzi, et al.: **VBSCan Mid-Term Scientific Meeting**. – Pre-print at arXiv:2004.00726.

- R. Bellan, J. Beyeri, M. Grossi, et al.: **VBSCan Thessaloniki 2018 Workshop Summary**. – Pre-print at arXiv:1906.11332.

IP Publications

- M. Grossi et all.: **Method and system for AI Visual Recognition training**. – *Prior Art Database*, `https://priorart.ip.com/IPCOM/000258681`.

- M. Grossi, L. Crippa, A. Aita, G. Agliardi, F. Accetta, F. Tramonto: **Method and system to create and deploy Cloud containerized quantum-based web applications using API-exposed quantum computers as back-end**. – *Prior Art Database*, `https://priorart.ip.com/IPCOM/000258553`.