

# Constraining the Higgs Boson Interactions with Effective Field Theory using Machine Learning

or

## Mining ‘Gold’ in the ‘Golden Channel’

Irene Cortinovis

Supervised by Dr. Roberto Salerno

High Energy Physics

Joint Master Program

Academic Years 2018-2020

École Polytechnique - ETH Zürich

### Abstract

In this thesis we investigate innovative methods to set exclusion limits to the Standard Model in the framework of Effective Field Theory. To this end, we introduce Madminer, a powerful neural network tool which provides various modern multivariate inference techniques. We choose the Vector Boson Fusion in the  $H \rightarrow ZZ \rightarrow 4\ell$  channel ( $\ell = \mu, e$ ), also known as the ‘golden channel’, as a playground for our study, or in other words for mining some ‘gold’. Our findings reveal that the approach of MadMiner in Effective Field Theory is not only feasible, but also more efficient in performance than the traditional methods.

*To Saurabh and Mihir*



*“A Mad Miner Mining Gold in the Golden Channel”*

Illustrated by Saagnik Paul

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	A short recap on Effective Field Theory . . . . .	5
1.2	Classical methods to constrain EFT parameters . . . . .	6
1.3	Constraining EFT with Machine Learning . . . . .	8
1.4	The Fisher information . . . . .	10
<b>2</b>	<b>The experimental apparatus and the <math>H \rightarrow ZZ \rightarrow 4\ell</math> channel</b>	<b>12</b>
2.1	LHC and the CMS detector . . . . .	12
2.2	The Vector Boson Fusion in the ‘Golden Channel’: playground for this study . . . . .	12
2.3	Event generation . . . . .	13
2.4	Reconstruction and selection . . . . .	15
2.5	The High-Luminosity LHC (HL-LHC) scenario . . . . .	17
<b>3</b>	<b>Study of the event kinematics</b>	<b>18</b>
3.1	Effect of the EFT operators on the kinematic variables . . . . .	18
3.2	Exploring the Fisher information . . . . .	20
<b>4</b>	<b>The MadMiner approach</b>	<b>25</b>
4.1	Sampling . . . . .	25
4.2	Likelihood ratio estimation with ALICES . . . . .	25
4.3	Score estimation with SALLY . . . . .	26
<b>5</b>	<b>Results</b>	<b>28</b>
<b>6</b>	<b>Conclusions</b>	<b>32</b>
	<b>References</b>	<b>33</b>

# 1 Introduction

Before we elaborate on our plans to mine ‘gold’ in the ‘golden channel’, it would be wise to introduce the reader to the idea of Effective Field Theory (EFT), a knowledge indispensable in our pursuit to discover the ‘gold’. We describe the classical methods with which the parameters of EFT can be constrained, and then we will discuss the possibility of an innovative method, a means to mine the ‘gold’: the MadMiner [1, 2, 3]. Finally, in order to gain some physical insight into the mining process, we discuss the Fisher information.

## 1.1 A short recap on Effective Field Theory

The Standard Model (SM), as it stands today, has been confirmed by most of the experiments in particle physics. While from theory we know that the SM is not the answer to all our questions, in experiments we have not been able to find any clear hints of Beyond the Standard Model (BSM) physics (yet). One possible reason could be that the new physics might be hiding at a very high energy scale, a domain we have not been able to access directly with our experiments so far. However, despite the impossibility of reaching high energy scales, it is reasonable to argue that we might be able to peek through and observe some effects in the current measurements themselves, given a sufficiently high precision. Not to miss out on what is already barely visible, rather than trying to narrow down on very specific BSM theories, it is best to rely on as few assumptions as possible and parametrize the theory as generally as one can.

One possible approach with the essence of these ideas is an Effective Field Theory (EFT) framework. If the typical energy scale of the new physics interactions  $\Lambda$  is much higher than the scale of Standard Model, one can supplement the Lagrangian  $\mathcal{L}$  beyond the SM with operators  $O_i^{(d)}$  of canonical dimension  $d > 4$ :

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + \mathcal{L}^{(7)} + \dots , \quad \mathcal{L}^{(d)} = \sum_{i=1}^{n_d} \frac{c_i^{(d)}}{\Lambda^{d-4}} O_i^{(d)}, \quad (1)$$

where  $c_i^{(d)}$  are the expansion coefficients, called Wilson coefficients.

The above expression parametrizes the effects of the new physics at the scale of our experiments ( $\ll \Lambda$ , which in this work is assumed to be 1 TeV).

A critical assumption on which this framework is built upon is that the operators which can capture the new physics are a function of the Standard Model fields alone. While the operators  $O_i^{(d)}$  can be expressed using various existing bases which can be mapped on to each other, the choice of basis generally depends on the phenomenological interpretation of the operators, and should represent the studied scenario as best as it can. Furthermore, the local SM symmetry  $SU(3) \times SU(2) \times U(1)$  is conserved.

The aim of this work is to investigate innovative methods to set limits for the Wilson coefficients. In particular, I will focus on operators of dimension 6, as dimension-5 operators violate the lepton number and higher order operators are more suppressed by powers of  $1/\Lambda$ .

## 1.2 Classical methods to constrain EFT parameters

Originally, the experiments at LHC were designed to explore a large range of energy scales to make discoveries. Since the historical discovery of the Higgs boson, physicists have concentrated on making precise Standard Model measurements, as well as setting stronger limits on BSM physics. However, as of now, no hints of new physics beyond the SM have been observed. In this regard, the Effective Field Theory has been able to successfully provide a framework that can be used to set limits around the SM.

The classical procedure for setting limits is by following a frequentist approach, based on hypothesis testing. In this case, the alternative hypothesis  $H_1$  comprises the SM scenario in which all  $c_i^{(d)}$  are zero, and the null hypothesis  $H_0$  is a BSM scenario, where the Wilson coefficients are not all zero [4].

Figure 2 sketches the general workflow for an example where we want to constrain the space spanned by the Wilson coefficients of two operators, around the SM ( $\theta_{ref} = (0, 0)$ ). For each point  $\theta$  in the parameter space, a set of events is generated using a simulation. The distribution of the likelihood ratio with respect to SM is obtained from the asymptotic properties [4], and subsequently utilized to calculate the p-value corresponding to the observed data. Thereafter, with the calculated p-values, the exclusion contours are drawn at the required confidence level.

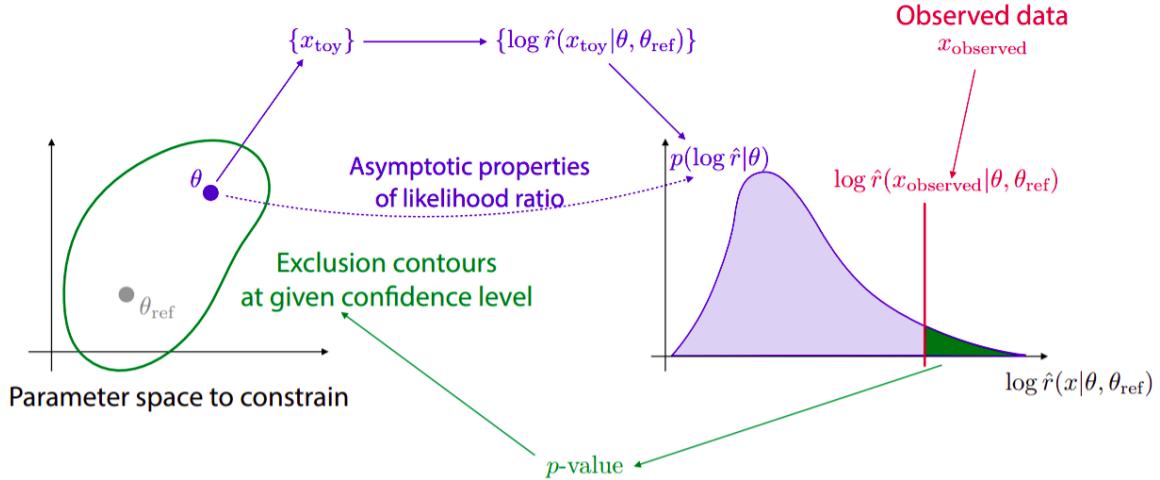


Figure 2: Workflow for the standard LHC procedure of limits setting, in the case of a 2D parameter space [5]

This inference procedure relies on the likelihood ratio which, according to the Neyman-Pearson Lemma, is the most powerful test statistic to discriminate between two simple hypotheses. However, the likelihood function for a process as complicated as a high energy collision in a detector, is not easily computable. In fact, the stochastic effects of the shower splittings and detector interactions depicted in Figure 3, make the likelihood completely intractable.

More specifically, the likelihood function factorizes to:

$$\textcolor{red}{p}(x|\theta) = \int dz_d \int dz_s \int dz_p \textcolor{red}{p}(x|z_d) \textcolor{red}{p}(z_d|z_s) \textcolor{red}{p}(z_s|z_p) p(z_p|\theta), \quad (2)$$

where  $z_d$  describes the effect of the interactions in the detector,  $z_s$  indicates the parton shower trajectories,  $z_p$  are the parton-level momenta, and  $\theta$  are the theory parameters, namely the Wilson coefficients. The red color is here used to denote intractability. We note that the only tractable part is the probability density of the parton-level momenta, conditional on the theory parameters:  $p(z_p|\theta) = 1/\sigma(\theta)d\sigma/d\theta$ .

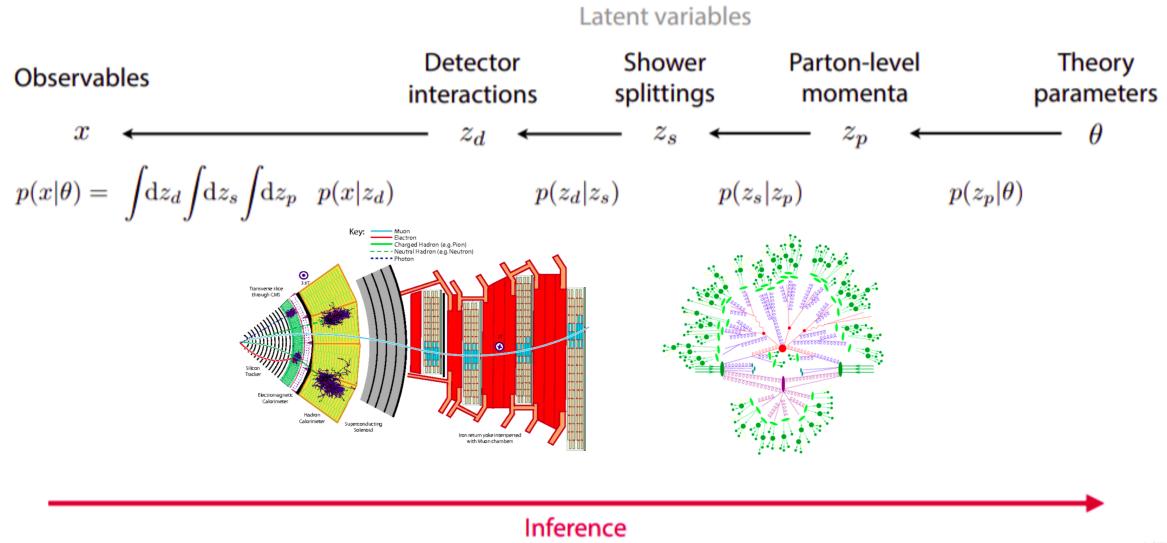


Figure 3: Schematic overview of the evolution of the likelihood function, starting from the theory parameters on the right (Wilson coefficients) until the observables at detector level, on the left. The more challenging reverse process of inferring the likelihood is indicated by the red arrow [5]

Hence the call for the so-called likelihood-free inference techniques, which do not make use of the analytical expression of the likelihood function, but rather try to approximate it in different ways.

Amongst the most utilized methods in the literature, we find three significant approaches:

- Carefully selecting a few summary statistics which contain the most information possible and estimating the likelihood function with histograms of the same. The compromise here is between having a more convenient lower dimensional space to work with, and inevitably losing information in the process.
- Using directly the matrix element to estimate the likelihood ratio (Matrix Element Method [6, 7], Optimal Observables [8]). The downside of this class of approaches is that the stochastic contribution to the full likelihood is neglected or approximated with transfer functions which are not fully realistic. Additionally, although the remaining integral is computable, it is still numerically expensive.

- Training neural networks according to different parameter points to be able to infer without a tractable likelihood function. These methods do not have the restrictions in the number of observables studied, but traditionally do not exploit the knowledge of the matrix element associated to the physics process.

It follows naturally that a fusion of these complementary approaches would improve the quality of likelihood-free inference. In this regard, we use the MadMiner tool, to combine the power of neural networks and the physics insight of the matrix element.

### 1.3 Constraining EFT with Machine Learning

The general workflow of an analysis using MadMiner<sup>1</sup> is illustrated in Figure 4. First, as per routine, a *simulation* realizes the original physics parameters (in this case the Wilson coefficients) into final state variables. The novelty here is that, in addition to the usual observables such as  $p_T$  of the Higgs boson or the missing transverse energy, some key information is also extracted from the simulator using the latent variables  $z$ , the parton level four-momenta. Subsequently, the output of the simulation (including the *augmented data*) is fed into a *Machine Learning* model (neural network) which aims to yield an estimator for the likelihood ratio. The estimator is then used for *inference* and limit setting, as outlined in Section 1.2.

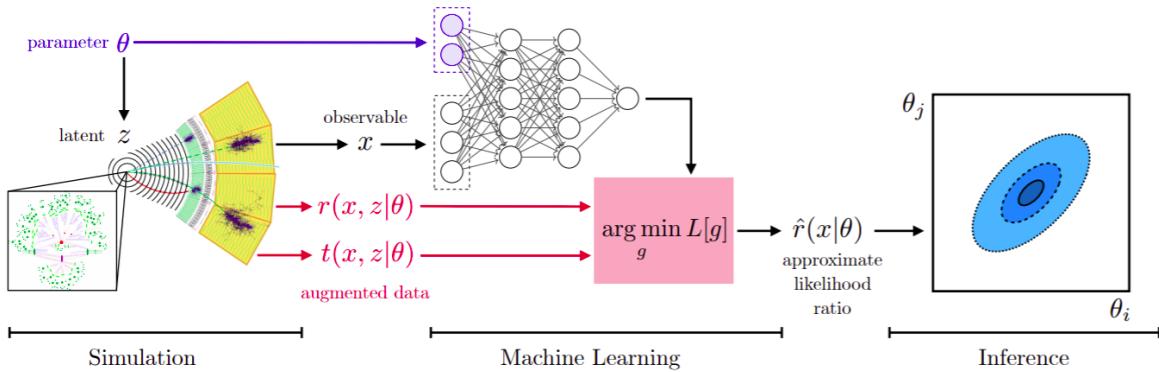


Figure 4: Schematic overview of the MadMiner workflow [3]

The augmented data is the ‘gold’ we ‘mine’ with MadMiner. The gold comprises a joint likelihood ratio  $r(x, z|\theta)$  and a joint score  $t(x, z|\theta)$ , for each single event. We will now discuss the nature and relevance of these physical quantities.

As seen in Eq. 2, the likelihood  $p(x|\theta)$  is an intractable quantity, and therefore so is the likelihood ratio  $r(x|\theta_0, \theta_1) = p(x|\theta_0)/p(x|\theta_1)$ . However, if we consider the joint likelihood  $p(x, z_d, z_s, z_p|\theta)$ , where the latent variables are frozen throughout the development of the event, we realise that there is no intractable integration measure anymore. The expression can be factorized according to the different consequent steps in the event:  $p(x|z_d)p(z_d|z_s)p(z_s|z_p)p(z_p|\theta)$ . It is now evident that the initial parameters  $\theta$  have an effect only on the parton level likelihood  $p(z_p|\theta)$ , which can be evaluated as a matrix element, while the rest does not depend on the

<sup>1</sup>The curious case of the name of *MadMiner* can be understood by appreciating the analogy with miners extracting gold, not too different from the mining of the augmented data itself [9].

EFT. Defining the joint likelihood ratio for an event as in the following equation, we can see that the only part that does not cancel out is the ratio of the matrix elements squared, thus making this quantity tractable.

$$r(x, z|\theta_0, \theta_1) \equiv \frac{p(x, z_d, z_s, z_p|\theta_0)}{p(x, z_d, z_s, z_p|\theta_1)} = \frac{p(x|z_d)}{p(x|z_d)} \frac{p(z_d|z_s)}{p(z_d|z_s)} \frac{p(z_s|z_p)}{p(z_s|z_p)} \cdot \frac{p(z_p|\theta_0)}{p(z_p|\theta_1)} \sim \frac{|\mathcal{M}(z_p|\theta_0)|^2}{|\mathcal{M}(z_p|\theta_1)|^2}. \quad (3)$$

The second quantity which is extracted and used to enhance the performance of the neural network is the joint score. The score  $t(x|\theta_0)$  is defined as the relative change of the likelihood under infinitesimal changes in the parameter space:

$$t(x|\theta_0) = \nabla_\theta \log p(x|\theta)|_{\theta_0}. \quad (4)$$

A straight forward expansion of the log likelihood in the neighborhood of the Standard Model reveals that the first order term is indeed the score:

$$\log p(x|\theta) = \log p(x|\theta_{SM}) + \nabla_\theta \log p(x|\theta)|_{\theta_{SM}} \cdot (\theta - \theta_{SM}) + \mathcal{O}((\theta - \theta_{SM})^2). \quad (5)$$

This makes the score a test statistic which is as powerful as the full likelihood, provided that it is used in the vicinity of the Standard Model. Unfortunately, this quantity is again intractable, because it depends on the full likelihood. However the joint score, defined as  $t(x, z|\theta_0) \equiv \nabla_\theta \log p(x, z_d, z_s, z_p|\theta)|_{\theta_0}$  can be computed, as the intractable parts which do not depend on the EFT parameters cancel out.

Thus we see that the joint likelihood ratio and the joint score are remarkable in the sense that they can be computed for every single event, unlike the other intractable physical quantities. Our aim would therefore be to utilize these special joint quantities for estimating the full likelihood ratio and score.

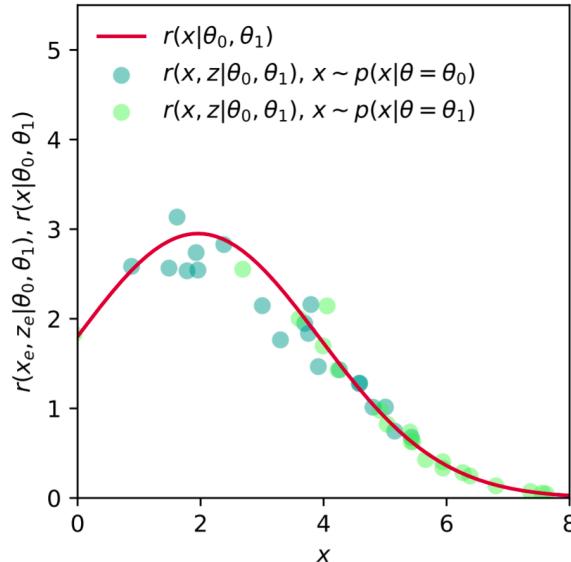


Figure 5: Regression on the joint likelihood ratios  $r(x, z|\theta_0, \theta_1)$  of simulated BSM events (green dots), converging to the likelihood ratio  $r(x|\theta_0, \theta_1)$  (red line) [1]

Extensive studies on this topic, such as [1], (Figure 5)], have shown that the joint quantities are scattered around the true functions. This critical condition provides a great handle for exploiting augmented data in our study: the problem of estimating the likelihood ratio and score can now be reduced to the minimization of loss functionals. For the purpose of explaining the basic concept, we can define them as:

$$L[\hat{r}(x|\theta_0, \theta_1)] = \frac{1}{N} \sum_{(x,z) \sim p(x,z|\theta_1)} |r(x, z|\theta_0, \theta_1) - \hat{r}(x|\theta_0, \theta_1)|^2, \quad (6)$$

and

$$L[\hat{t}(x|\theta)] = \frac{1}{N} \sum_{(x,z) \sim p(x,z|\theta)} |t(x, z|\theta) - \hat{t}(x|\theta)|^2. \quad (7)$$

Loss functionals based on these are then implemented in fully connected feed-forward neural networks, which after an upfront training phase will allow to quickly use the regressed estimators to draw conclusions about any initial parameters.

## 1.4 The Fisher information

Another interesting quantity that makes use of the likelihood function and which helps to give a different perspective on the analysis is the Fisher information matrix  $I_{ij}$ . It is a  $n \times n$  matrix, where  $n$  is the number of theory parameters, and intuitively represents the quadratic coefficient of the local approximation of the expected log likelihood ratio at  $\theta_{ref}$ :

$$\mathbb{E}[-2 \log r(x|\theta)|\theta_{ref}] = I_{ij}(\theta_{ref}) \times (\theta - \theta_{ref})_i (\theta - \theta_{ref})_j + \dots \quad (8)$$

This means that a higher Fisher information matrix entry corresponds to a steeper log likelihood ratio curve, which translates into a better knowledge of the phenomenon.

It can be shown that the expression reduces to:

$$I_{ij}(\theta_{ref}) \equiv \mathbb{E} \left[ \frac{\partial \log p(x|\theta)}{\partial \theta_i} \frac{\partial \log p(x|\theta)}{\partial \theta_j} \middle| \theta_{ref} \right]. \quad (9)$$

We note here that  $I_{ij}$  is a priori intractable, given that it depends on  $p(x|\theta)$ . However, it can be approximated with an estimator of the likelihood or score, as obtained with MadMiner, or by any other traditional method.

We now discuss some special properties of the Fisher information matrix:

- The inverse of the Fisher information gives the lower bound for the expected error of the best-fit parameter:  $cov[\hat{\theta}|\theta_0] \geq I_{ij}^{-1}(\theta_0)$  (Cramér-Rao bound). This is a mathematical reformulation of the intuitive idea that large entries are associated with directions in the parameter space that can be measured well.
- After removing blind directions (that is, eigenvectors with zero eigenvalues), it is a symmetric and positive definite rank-two tensor which can define a Riemannian metric on the Wilson parameters space. This allows us to define a local distance of any parameter point from the Standard Model:  $d_{loc}(\theta_{SM}, \theta) = \sqrt{(\theta_{SM} - \theta)_i I_{ij}(\theta_{SM} - \theta)_j}$ , which reveals how likely it would be to measure the parameter  $\theta$ , given the true value  $\theta_{SM}$ . This can be utilized to quickly draw limit contours, in the approximation of a neighborhood of the Standard Model.

- The information is additive over phase space regions, which implies that the sum of all the information defines the minimum covariance matrix. It is also useful to study how the differential information is distributed over the phase space. This is also a quick way to get physical insight into where the power of the analysis is coming from.
- It is also additive between experiments. In order to see the power of combined results, it is sufficient to add the Fisher information matrices.
- The Fisher information approach can be readily extended to take into account the uncertainties just by adding the nuisance parameters to the parameter space and by adding the constraint terms to the likelihood.
- It is independent of reparameterizations of  $x$  and covariant under reparameterizations in  $\theta$ .

Having outlined all the ideas and building blocks, I will now proceed to my implementation which aims to set EFT limits in the  $H \rightarrow ZZ \rightarrow 4\ell$  channel at CMS.

## 2 The experimental apparatus and the $H \rightarrow ZZ \rightarrow 4\ell$ channel

In this chapter, we first introduce the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector. Thereafter, we discuss the  $H \rightarrow ZZ \rightarrow 4\ell$  channel, also known as the ‘golden channel’, which we use as a playground for this study. Following this, we describe the process of generation of events and subsequent selection. Finally, we project our study to the specific case of the High Luminosity LHC phase.

### 2.1 LHC and the CMS detector

The Large Hadron Collider is the last element of the CERN accelerator complex (Figure 6, left): it is a superconducting hadron accelerator and collider, sitting in a 27-kilometer ring underground tunnel. It collides protons at a center of mass energy of  $\sqrt{s} = 13$  TeV. The Compact Muon Solenoid (CMS) experiment [10] (Figure 6, right) is a general-purpose particle detector located at one of the collision points of the LHC. As its name suggests, its distinctive part is the superconductive solenoid: it generates a magnetic field of around 4 Tesla, which is kept confined in the detector thanks to a heavy steel yoke. The strong magnetic field bends the particles in the detector, enabling the measurement of their momentum. The various subdetectors of CMS are shown in the right panel of Figure 6; radially from the center they are: the Pixel, the Silicon Tracker, an Electromagnetic Calorimeter (ECAL) made of lead tungstate crystals, a sampling hadronic calorimeter (HCAL), and outside the solenoid, the Muon Detectors.

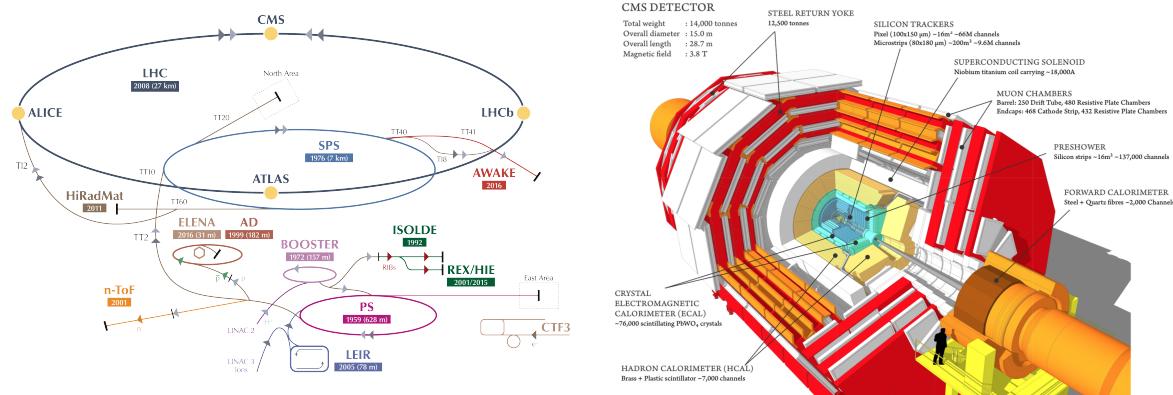


Figure 6: Schematics of CERN accelerator complex (left) and CMS detector overview (right).

### 2.2 The Vector Boson Fusion in the ‘Golden Channel’: playground for this study

The  $H \rightarrow ZZ \rightarrow 4\ell$  channel ( $\ell = e, \mu$ ) has been called the ‘golden channel’ because of its extremely neat signature, the cleanest of all Higgs boson decay modes. Its final state products can be fully reconstructed, therefore it is possible to attain a great resolution for the signal, that is a narrow peak in the distribution of the invariant mass of the final states. The achieved excellent signal to background ratio has made this channel stand out since the very beginning, being one of the very first channels for the Higgs boson discovery [11]. To date, the increased integrated luminosity available and the developments in the analysis techniques allow high

precision measurements. It is only natural to exploit the level of precision in the ‘golden channel’, in order to reach for something beyond the Standard Model as well.

In fact, using the framework of Effective Field Theory, the ATLAS experiment at LHC has recently moved in this direction, successfully setting limits for a few dimension-6 operators with the observed data [12].

In the context of Fisher information on EFT parameters, this channel has been studied in [13] and [14], although only at parton level. The aim of this thesis is, therefore, to utilize the newer techniques described in Section 1 (MadMiner and Fisher information) to study the  $H \rightarrow ZZ \rightarrow 4\ell$  process at the CMS experiment. In particular, I concentrate on the Vector Boson Fusion (VBF) production mode because of its clean signature. Moreover, the possibility to study the same operators (relevant to the coupling of the Higgs boson to vector bosons) in the production and decay of the Higgs boson increases the sensitivity of the channel to such operators, thus making it ideal for this study.

### 2.3 Event generation

In order to study the effect of EFT operators, I generated data with MadGraph5 [15], utilizing the model EWdim6, which implements the weak boson EFT described in [16], using the HISZ basis [17]. Figure 7 shows a representative Feynman diagram for the input process to MadGraph5.

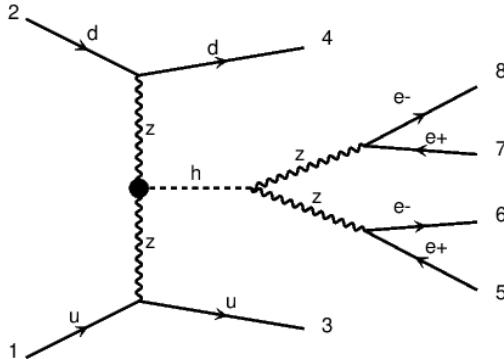


Figure 7: One of the possible Feynman diagrams from MadGraph5, for the  $H \rightarrow ZZ \rightarrow 4\ell$  channel with VBF production.

The process has been generated with the following mg5@NLO command:

```
generate u d > u d h QCD=0 QED=99 NP=2,
(h > z z QCD=0 QED=99 NP=0, (z > 1+ 1- QCD=0 QED=99 NP=0))
```

where `NP=2` corresponds to the request of a new physics vertex in production only (black dot in Figure 7). This choice follows from a set of physical considerations: firstly, the selection of more than one vertex at a time can lead to the possibility of negative event weights due to interference. Second, it can be argued that the new physics would have a greater contribution at the Higgs boson production vertex than the decay vertex, because the momenta in the former are higher and more unconstrained.

In my study I will consider two EFT operators:

$$\mathcal{O}_W = (D_\mu \Phi)^\dagger W^{\mu\nu} (D_\nu \Phi) \quad (10)$$

$$\mathcal{O}_{\widetilde{W}} = (D_\mu \Phi)^\dagger \widetilde{W}^{\mu\nu} (D_\nu \Phi), \quad (11)$$

where  $\Phi$  is the Higgs boson doublet field,  $W^{\mu\nu}$  denotes the full field strength of the  $W$  gauge field.  $\mathcal{O}_W$  is CP-even, while  $\mathcal{O}_{\widetilde{W}}$  is CP-odd. These two operators have an effect on the tree-level Higgs boson coupling with Vector Bosons.

The brute force approach to generate data would be to scan the entire parameter space of the two Wilson coefficients, with a very fine grid. However, this would be computationally extremely expensive. One way to circumvent this problem, or at least make it computationally less expensive, is to use a technique called morphing. Morphing consists of choosing a sufficient number of basis points in the parameter space that can serve to analytically map the entire phase space. An application of the technique of morphing can be for computing the likelihood at parton level (the principle holds also for the full likelihood), as shown in the following equation:

$$p(z|\theta) = \sum_c w_c(\theta) p(z|\theta_c). \quad (12)$$

Here, the weights  $w_c(\theta)$  can be calculated from the amplitude of the process. Figure 8 shows the parameter space of my study, which consists of the values of the Wilson coefficients  $c_W$  and  $c_{\widetilde{W}}$ , multiplied by the square of the vacuum expectation value of the Higgs field,  $v \equiv (\sqrt{2}G_F)^{-1/2} \approx 246$  GeV. The chosen basis points are in black and the color map represents the resulting squared sum of the weights. On one hand it is preferable to have low values for the weights everywhere in the parameter space. On the other hand, by moving away from the basis points, the weights would inevitably increase.

One way to reach a balance between these two self-competing aspects is to try to minimise the sum of the weights in the parameter space, by an appropriate selection of the basis points. In this particular case, I chose the basis points in a way such that a sufficiently large region around the Standard Model would have a low value for the weights (yellow).

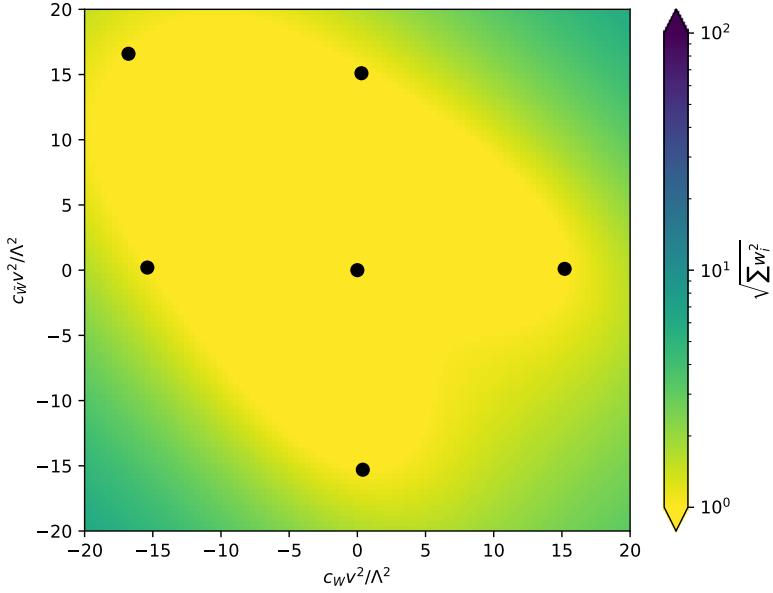


Figure 8: Basis points  $\theta_c$  (black dots) and square root of the squared sum of the morphing weights (color map), in the entire 2D Wilson coefficients parameter space.

Following the generation of events, the hadronization of the process is modelled with Pythia8 [18], and the events pass through a simulation of the CMS detector implemented with Delphes 3.4.2 [19].

Delphes is a framework designed to perform fast and realistic phenomenological detector simulations. The choice of Delphes over a more realistic simulator such as Geant4 stems from the fact that MadMiner is already interfaced with Delphes.

## 2.4 Reconstruction and selection

After generating the events at the detector level, the output of Delphes provides all the reconstructed objects. Thereafter, I implemented an algorithm to analyze the reconstructed objects for each event and select the Higgs boson candidates, based on the logic of the published CMS analysis [20]:

1. Find all  $Z$  candidates: opposite sign - same flavour pairs of leptons such that their invariant mass satisfies  $12 < m_{\ell\ell} < 120$  GeV.
2. Form all possible  $ZZ$  candidates, defining  $Z_1$  as the candidate which has mass closest to the on-shell  $Z$  mass, and  $Z_2$  the other one.
3. Apply the following kinematical cuts:
  - on the invariant mass of the on-shell  $Z$  boson:  $m_{Z_1} > 40$  GeV
  - defining  $\ell_1$  and  $\ell_2$  the two candidate leptons with highest  $p_T$ :  $p_T(\ell_1) > 20$  GeV and  $p_T(\ell_2) > 10$  GeV

- on the invariant mass of the 4 candidate leptons:  $m_{4\ell} > 70$
  - on the invariant mass for opposite flavour candidate lepton pairs  $m_{\ell\ell} > 4$  GeV
  - reject  $4\mu$  and  $4e$  candidates where the alternate pairing  $Z_aZ_b$  (where  $Z_a$  is closer to the nominal Z mass) satisfies  $|m(Z_a) - m(Z)| < |m(Z_1) - m(Z)|$  and  $m(Z_b) < 12$  GeV
4. If more than one ZZ candidate pass the cuts, the pair where  $Z_1$  has mass closest to  $m(Z)$  is chosen.

After applying these cuts, we are left with about 15% of the originally generated events. The loss is due to different factors: the acceptance of the detector, the reconstruction efficiency, and the selection efficiency.

When the analysis is performed inclusively on all the production channels, it is useful to have pure reconstructed categories in each production mechanism. This is already implemented in the current CMS Run 2  $H \rightarrow ZZ \rightarrow 4\ell$  analysis, where the Higgs boson candidates are further categorized in exclusive production categories. In particular, for the case discussed in this thesis, we could enhance the VBF production mechanism from the irreducible backgrounds and from the other production mechanisms, simply by requiring 2 reconstructed jets and a lower bound of 700 GeV for the invariant mass of the two leading jets. As shown in Figure 9, based on the published analysis, this selection allows to attain a purity of around 80% in the VBF channel. These cuts have for these reasons been also considered in my analysis.

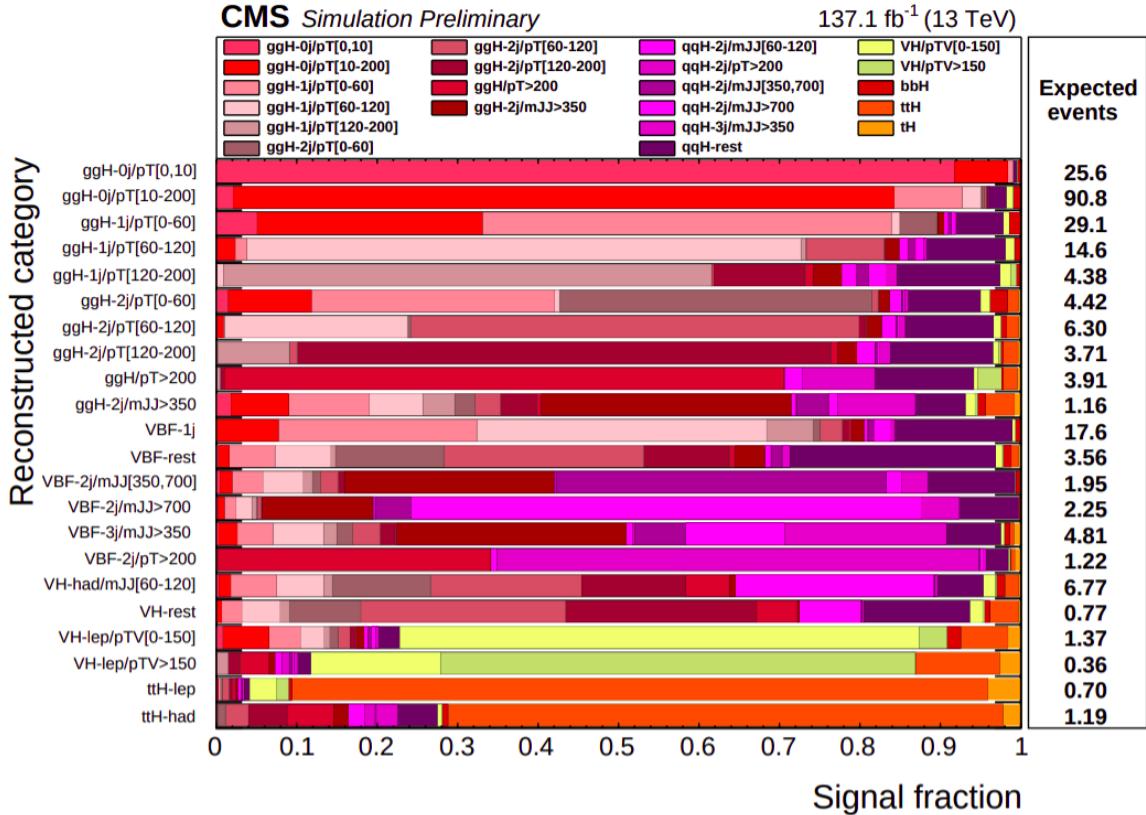


Figure 9: Signal relative purity of the 22 reconstructed event categories in the  $118 < m_{4\ell} < 130$  GeV mass window, for the published CMS  $H \rightarrow ZZ \rightarrow 4\ell$  analysis [20]. The category selected in my study is VBF-2jet-tagged- $m_{jj} > 700$  GeV.

## 2.5 The High-Luminosity LHC (HL-LHC) scenario

Figure 9 tells us that based on the currently available amount of data ( $137 \text{ fb}^{-1}$ ), one would expect the 80% of 2.25 final events ( $\sim 1.8$ ), to fall in our chosen signal category. However, in the context of EFT, it is much more meaningful to work with as many events as possible. This is why we project this study in the final scenario of the High Luminosity Large Hadron Collider (HL-LHC) phase, where  $3000 \text{ fb}^{-1}$  are expected to be collected. Scaling linearly, this would result in around 40 expected events for this case.

Such an estimate is in fact a conservative guess, because a linear extrapolation for the number of expected events based on the increase in available data does not take into account the developments of techniques which will improve the selection efficiency of the analysis. On one hand, it is true that the conditions in the detector will be harsher and the pile-up will increase, but there will be also improvements in the detector itself. For example, the CMS experiment intends to replace the current endcap calorimeters with a new High-Granularity Calorimeter (HGCal) [21], in the context of the HL-LHC upgrade program. The HGCal will allow for a better quark/gluon jet discrimination. This will help to remove the main background for the VBF channel, which consists of gluon-fusion events with two additional

jets originating from gluons.

### 3 Study of the event kinematics

This chapter elaborates on the selection of appropriate kinematical observables, which can subsequently be fed into the neural network. The extent to which the EFT operators can influence these variables would then be studied in the context of the Fisher information.

#### 3.1 Effect of the EFT operators on the kinematic variables

We now need to choose meaningful variables which can capture the deviations from the SM that we have ‘activated’ with the EFT operators in the  $H \rightarrow ZZ \rightarrow 4\ell$  channel, with VBF production mode. This set of variables will represent each single event as it is fed into the neural network, and therefore they should carry as much information as possible. It is natural to choose observables related to the kinematics of the reconstructed Higgs boson candidates, the  $Z$  bosons in which they decay, and the two additional jets  $j_1$  and  $j_2$ , which unequivocally define the VBF signature. In particular, I have selected the following 17 variables:

- invariant mass of the Higgs boson and the off-shell  $Z$  boson,  $Z_2$  (2 variables);
- transverse momentum  $p_T$  of the Higgs boson,  $Z_1$ ,  $Z_2$  and  $j_1$  (4 variables);
- pseudorapidity  $\eta$  of the Higgs boson,  $Z_1$ ,  $Z_2$ ,  $j_1$  and  $j_2$  (5 variables);
- $\Delta\phi$ ,  $\Delta\eta$  between the two  $Z$  bosons, and between the two leading jets (4 variables);
- number of jets (1 variable);
- missing transverse energy (1 variable).

By plotting the distributions of these variables, we can compare the SM prediction with the two pure CP-even and pure CP-odd EFT scenarios, respectively where  $(cwv^2, c_{\widetilde{W}}v^2)$  are  $(10,0)$  and  $(0,10)$ , and look for visible changes in their shape and normalization. As representative histograms of observables, in Figure 10 we plot the signed azimuthal angle difference between the two leading jets  $\Delta\phi_{jj}$  (top left), the transverse momentum of the reconstructed Higgs boson ( $p_T$ ) (top right), the  $Z_1$  boson (bottom left), and the  $Z_2$  boson (bottom right).

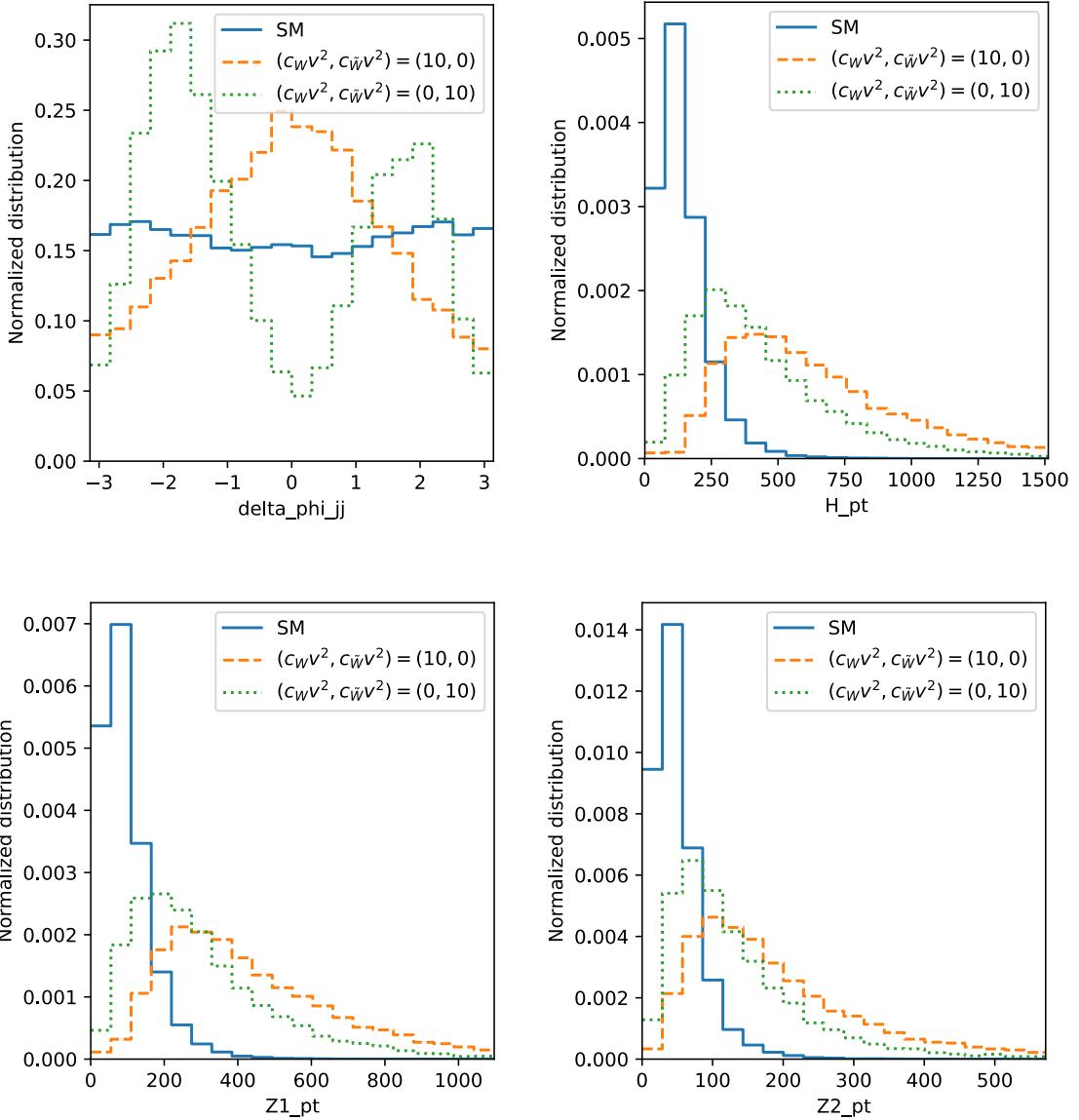


Figure 10: One-dimensional histogram of observables for SM (blue), the CP-even pure scenario  $O_W$  (orange) and the CP-odd pure scenario  $O_{\tilde{W}}$  only (green), normalized to the total cross section. Top left: signed azimuthal angle difference between the two leading jets. Top right:  $p_T$  of the reconstructed Higgs boson. Bottom left:  $p_T$  of the on-shell decaying Z boson. Bottom right:  $p_T$  of the off-shell decaying Z boson. ( $\Lambda$  is always assumed to be 1 TeV)

From Figure 10 we can observe that the shapes of the normalized distributions are indeed different for each of the three different scenarios. This establishes the crucial point that the EFT operators, as chosen in this case, do have an impact on the phenomenology of this channel.

We start by first discussing Figure 10, top left panel: in the case of the  $\Delta\phi_{jj}$  distribution,

the SM (blue line) predicts that the variable is fairly independent of the signed azimuthal angle. Conversely, the CP-even scenario (orange) shows that the majority of the events occur at zero azimuthal angle. Furthermore, the distribution exhibits a preservation of the CP symmetry. On the other hand, the CP-odd scenario (green) shows minima at the azimuthal angle values of  $0$ ,  $+\pi$  and  $-\pi$ . Curiously, the distribution is not symmetric around zero, which is in fact a signature of CP violation. In general, one must be careful about such an interpretation, because absorptive phases originating from an imaginary Wilson coefficient could also lead to CP asymmetries [14].

Moving on to Figure 10, top right panel, we see the distribution of the Higgs boson transverse momentum. Here we note that the majority of the events predicted by the SM (blue line) occurs at a lower value of  $p_T$ , when compared to the BSM scenarios (orange and green lines). This can possibly be attributed to the additional boost given from the off-shell particles that exist at higher energy scales, with which the Higgs boson couples in higher order loops.

Finally, in Figure 10 bottom left and right panels, we plot the transverse momentum of the reconstructed Z1 and Z2 bosons, decaying from the Higgs boson. We expect the Higgs boson products to keep the memory of the additional boost, and indeed, the distributions of the  $p_T$  of the two Z bosons are not contrary to our intuition.

### 3.2 Exploring the Fisher information

In this section, we try to gain some physical insight into the amount of information regarding each EFT operator that the chosen observables can carry. As outlined in the introduction, the Fisher information matrix  $I_{ij}$  is the tool that serves this precise function. To compute  $I_{ij}$ , the likelihood is estimated classically, that is, from the histograms of the variables, both taken singularly and/or combined in 2D histograms.

Figure 11 provides an overview of the quantity of Fisher Information carried by some of the single variables which are chosen to be fed to the neural network, and some of their 2D combinations, compared to all the information present in the full event at parton level.  $I_{ij}$  is evaluated at the Standard Model, for a HL-LHC projected integrated luminosity of  $3000 \text{ fb}^{-1}$ . As a reference to Figure 11, we provide in Table 1 a summary of the Fisher information contained in each one of the 17 chosen variables.

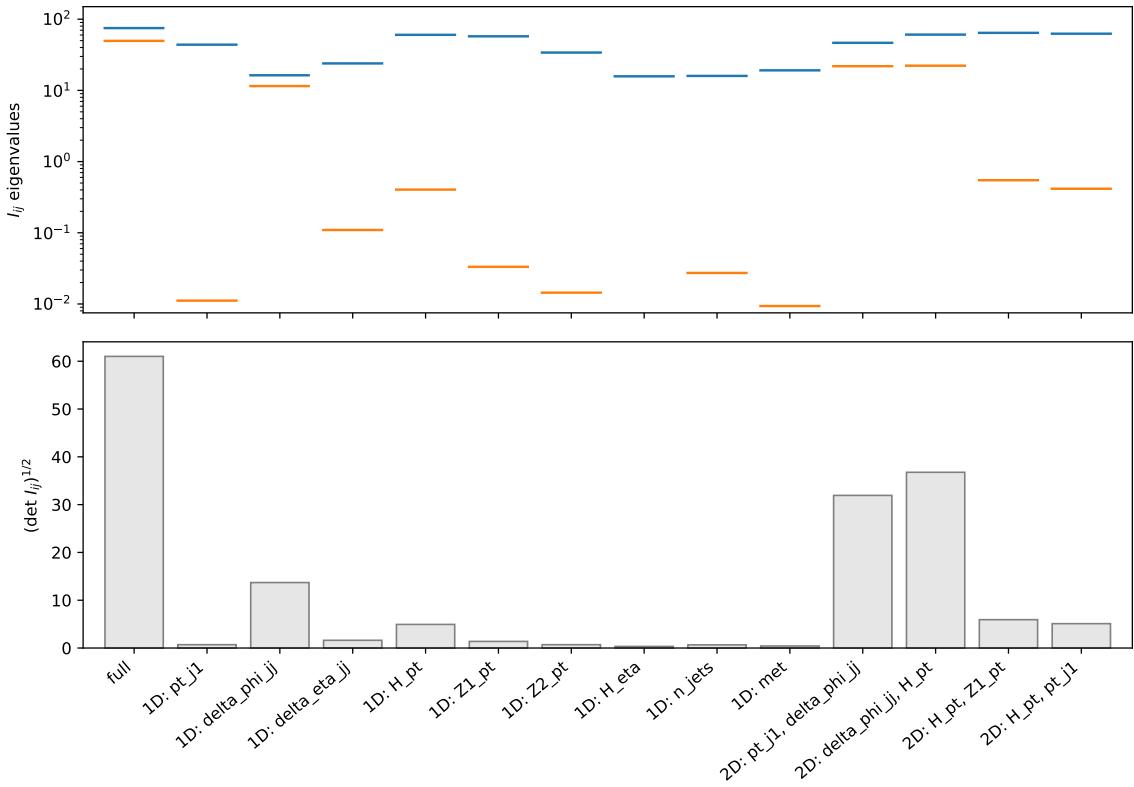


Figure 11: Comparison of the Fisher information content in different 1D or 2D variables, with the full information. Given that this is a feasibility study, the binning of the histograms has not been fully optimized, therefore small improvements in the reach of these variables can be obtained after optimization. Bottom half: the bars show the determinants of Fisher matrix for each variable. Upper half: eigenvalues and the composition of the corresponding eigenvectors for each variable: in blue  $O_W$ , in orange  $O_{\widetilde{W}}$ .

1D observable	$(\det I_{ij})^{-1/2}$	1D observable	$(\det I_{ij})^{-1/2}$
delta_phi_jj	13.7	eta_j2	0.5
H_pt	4.9	met	0.4
delta_eta_jj	1.6	H_m	0.4
Z1_pt	1.4	H_eta	0.3
Z2_pt	0.7	Z1_eta	0.2
pt_j1	0.7	Z2_eta	0.2
n_jets	0.7	Z2_m	0.2
delta_phi_zz	0.5	m_jj	0.1
eta_j1	0.5		

Table 1: Fisher information contained in each one of the 17 observables chosen as an input for the neural networks.

In the top panel of Figure 11, we see the eigenvalues of the Fisher information matrices, with the colors orange and blue corresponding to the eigenvector decomposition of CP-even  $O_W$  and CP-odd  $O_{\widetilde{W}}$  respectively. The fact that every decomposition is single-colored is a testament to the fact that the Fisher matrices are mostly diagonal. Moreover, the eigenvalue corresponding to the CP-even operator being always greater than that of the CP-odd operator, demonstrates that the sensitivity for  $O_W$  is in general more than that of  $O_{\widetilde{W}}$ .

In the bottom panel, we see vertical bars representing the amount of information carried by each variable: the more the information, the higher the potential reach in the new physics, when the analysis is restricted to that sole variable (1D or 2D). Furthermore, when comparing the information contained in the parton-level full event to that carried by single variables and 2D combinations, we observe that by restricting the analysis to only one variable, we definitively miss out on a considerable amount of information. This helps us make the critical observation that the EFT does not capture the new physics in one single observable, rather in a more diffuse nature, where different variables capture various aspects of the phenomena. We further note that the signed azimuthal angle between the two hardest jets,  $\Delta\phi_{jj}$ , stands out amongst the other observables, thanks to its ability to better represent the effects of the CP-odd operator,  $O_{\widetilde{W}}$  - in line with the observations from Figure 10.

A careful look at the 2D variables reveals that, if constructed by a judicious choice of combination, they can be more effective in carrying the information, in comparison to its constituent single variables. As an illustrative example we look at the bottom panel of Figure 11 and select the 2D variable 2D: `pt_j1, delta_phi_jj`, and compare it to its constituents 1D: `pt_j1` and 1D: `delta_phi_jj`. From the magnitude of the vertical bars themselves we know that the 2D variable contains more information than the single ones. We can further provide an example of a naive combination, where the 2D variable is evidently not as effective. This is the case of 2D: `H_pt, Z1_pt`, constructed by the combination of the  $p_T$  of the Higgs boson and of the on-shell decaying Z boson, where the two quantities are highly correlated (as already discussed in the previous section, Figure 10).

Once the appropriate variables have been selected, we can now use the local Fisher distance to draw a first order approximation of the optimal EFT limits. In fact, we can draw equidistant contours based on the 2D variable and its constituents in the phase space of the Wilson coefficients, and directly compare how effective they are in constraining the parameters.

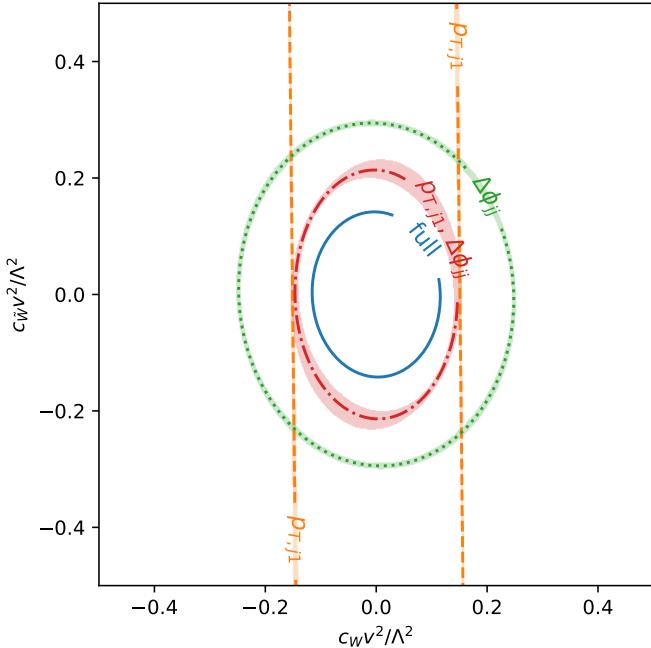


Figure 12: First order approximation of optimal 68% CL contours, computed with the Fisher information evaluated at the Standard Model, for an integrated luminosity of  $3000 \text{ fb}^{-1}$ . The shaded error bands are calculated with plain Gaussian error propagation.

In Figure 12 we see that the 68% CL contours for the  $p_T$  of the leading jet (yellow) taken singularly does not constrain  $O_{\widetilde{W}}$ , at least for the first order approximation at the SM. While the contour for  $\Delta\phi_{jj}$  (green) is able to constrain both the parameters, we see that the 2D combination (red) is far more effective for the same. Of course, the most effective in constraining the parameters is the contour corresponding to the full information at parton-level (blue). However, practically this knowledge is not available.

This is where the Fisher information becomes extremely useful, as it provides a quick and quantitative measure of the relative importance of the different physical observables, with regards to the extent of the information they can carry.

As a side note, it can also be interesting to study how the information is distributed over the phase space of the observables. More specifically, this knowledge can drive the choice of selections in the analysis, by quantitatively showing the loss of information resulting from any cut. We illustrate this in the left panel of Figure 13, which shows the distribution of the  $\Delta\phi_{jj}$  variable in the SM hypothesis (black), and the corresponding Fisher information (red). The figure clearly shows that the Fisher information peaks at  $\Delta\phi_{jj} = \pm\pi/2$ , which intuitively implies that in the normalized distribution of  $\Delta\phi_{jj}$ , it would be easier to distinguish between the SM and BSM scenarios at these values.

We demonstrate this in the right panel of Figure 13, where we plot the normalized distri-

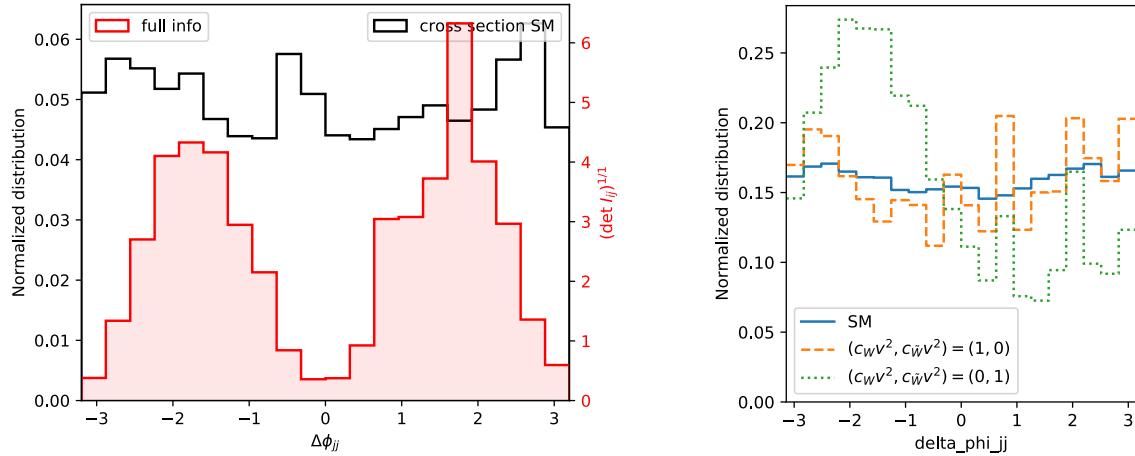


Figure 13: On the left: distribution of the signed azimuthal angle between the two leading jets, and distribution of Fisher information for each corresponding bin. On the right: One-dimensional histogram of  $(\Delta\phi_{jj})$  for SM (blue), the CP-even pure scenario  $c_W v^2 = 1$  (orange) and the CP-odd pure scenario ( $\tilde{c}_W v^2 = 1$ ) only (green), normalized to the total cross section. ( $\Lambda$  is always assumed to be 1 TeV)

butions of the CP-odd, CP-even and SM events. This is in fact the same plot as Figure 10, but with a value of the Wilson coefficients one order of magnitude smaller.

A high value of Wilson coefficients accentuates the deviations from the SM, and thus makes it difficult to distinguish where the distributions are more sensitive to BSM effects. This can be seen in Figure 10, where we plot the normalized distributions of the CP-odd, CP-even and SM events for a Wilson coefficient value of 10. The resulting BSM distributions give the impression of being equally sensitive for all values of the signed azimuthal angle. However, when plotted with a Wilson coefficient of 1, as in the right panel of Figure 13, we see that the deviations seem more prominent at  $\Delta\phi_{jj} = \pm\pi/2$ , as suggested by the distribution of the Fisher information itself.

## 4 The MadMiner approach

In the previous chapter we have utilized the Fisher information to estimate EFT limits based on a local approximation. This however neglects the non linear effects of the theory parameters on the likelihood function. In this section, we discuss how we can compute the full likelihood ratio function in order to set exclusion limits accounting for the non-linear effects as well. We do this based on the MadMiner approach, where the likelihood ratio and/or the score are estimated with neural networks, following implementation of different models. In particular, here I use the model ALICES (Approximate Likelihood with Improved Cross-entropy Estimator and Score, [22]) for the likelihood ratio estimation, and SALLY (Score Approximates Likelihood LocallY [1]) for the score estimation.

### 4.1 Sampling

The first step involves preparing the input for the neural network. This requires splitting the events in what is typically called *training dataset* (80% of the total events) and *test datasets* (20% of the total events). The training dataset will be used to train the neural networks, while the results will be obtained using the test dataset. The events in the training sample are unweighted, that is, the events will not carry the individual weights associated to them during the Monte Carlo generation, but instead will follow the  $p(x|\theta)$  distribution. Following this, the data is augmented by supplementing each event with the joint likelihood ratio and/or the joint score (the ‘gold’).

We now discuss the details of the two different approaches.

### 4.2 Likelihood ratio estimation with ALICES

The training sample for the ALICES model consists of  $10^5$  unweighted events. The values of the two components of the 2D vector of EFT parameters  $\theta_0$  (appearing in the numerator of the joint likelihood ratio) are drawn randomly from a two-dimensional gaussian prior with mean 0 and RMS 0.5, while  $\theta_1$  remains fixed at the SM (0,0). The joint score is also computed for each event, and it then serves as a guide for the estimator.

To explain how this works in practice, we can now consider the loss function for this neural network:

$$L_{\text{ALICES}}[\hat{s}(x|\theta_0, \theta_1)] = -\frac{1}{N} \sum_{(x_i, z_i) \sim p(x_i, z_i)} \left[ s(x_i, z_i|\theta_0, \theta_1) \log(\hat{s}(x_i)) + (1 - s(x_i, z_i|\theta_0, \theta_1)) \log(1 - \hat{s}(x_i)) \right. \\ \left. + \alpha(1 - y_i) \left| t(x_i, z_i|\theta_0, \theta_1) - \nabla_\theta \log \left( \frac{1 - \hat{s}(x_i|\theta, \theta_1)}{\hat{s}(x_i|\theta, \theta_1)} \right) \right|_{\theta_0}^2 \right], \quad (13)$$

where  $s(x, z|\theta_0, \theta_1) = p(x, z|\theta_1)/(p(x, z|\theta_0) + p(x, z|\theta_1))$ ,  $y_i = 0$  for samples  $(x_i, z_i) \sim p(x, z|\theta_0)$  and  $y_i = 1$  for samples  $(x_i, z_i) \sim p(x, z|\theta_1)$ .

The loss function is built with the sum of two terms: the first one is related to the likelihood itself, and the second one to the score. The underlying idea is that the joint score is scattered around the score calculated from the likelihood estimator, and therefore both terms are simultaneously minimized by the best estimator. The hyperparameter  $\alpha$  can also be optimized, but for the sake of simplicity, in this study I have chosen its value to be 10.

The ALICES model is now implemented by MadMiner in a fully connected feed-forward neural network. In this case, the network has two hidden layers of 60 units each, activated by the tanh function. It is trained for 10 epochs using the amsgrad optimizer, with a learning rate of 0.001 which decays to 0.0001.

The trained model is then validated by plotting the expected log likelihood ratio estimated from a test sample of 1000 SM events, in a grid of  $25 \times 25$  of the parameter space (see Figure 14). As expected, the best fit falls very close to  $(0,0)$  and the log likelihood ratio increases as the distance from the SM increases.

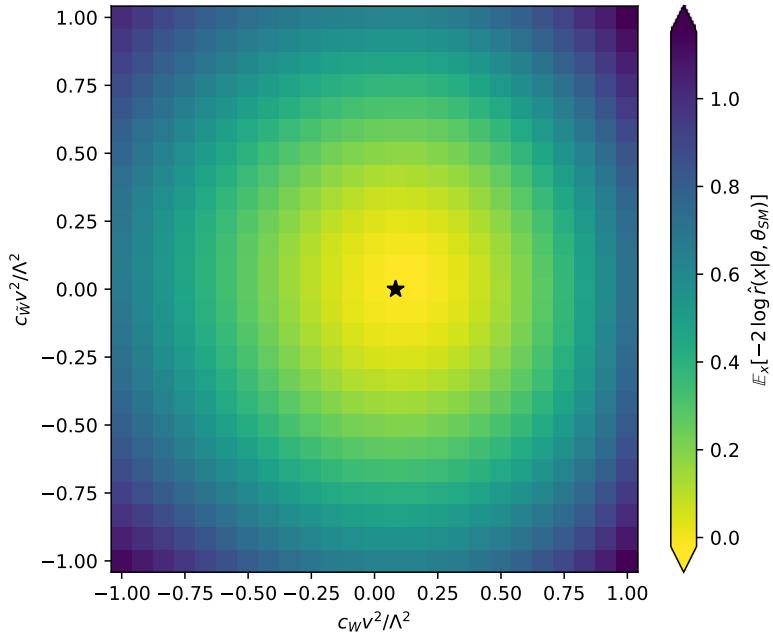


Figure 14: Expected log likelihood ratio, estimated from SM samples. The star indicates the best fit and the color map the value of the expected log likelihood ratio with respect to the SM. The purpose is to validate the trained ALICES model.

Now that the model has been validated against the SM, it can now be used for drawing exclusion limits for EFT.

### 4.3 Score estimation with SALLY

SALLY, as introduced before, is a local score estimator.

Following Eq. 5, we see that the score must be estimated at a particular reference parameter point, which in this case is the Standard Model. Therefore, the training data here

must also be sampled from the Standard Model alone. This provides a clear advantage to this model, by considerably easing the event generation process. The model is also implemented in a fully connected neural network, with two hidden layers of 30 units each, trained with the amsgrad optimizer for 50 epochs. The loss function is the one described in Section 1.3, for the score regression.  $5 \times 10^5$  unweighted events are used for the training.

Once the model is trained, it is instructive to study how the score is related to the observables. The score is estimated on the same test dataset that was previously used to validate the ALICES model. This dataset comprises 1000 events drawn from the SM.

In order to illustrate this we go back to our classical example from Chapter 3, where we had chosen the variables  $p_{T,j1}$  and  $\Delta\phi_{jj}$ . In Figure 15, we plot each component of the estimated score in the phase space spanned by the two variables - left:  $\hat{t}_0$ , right:  $\hat{t}_1$ . It follows from Eq. 5 that the two components are related to the CP-even and CP-odd operators respectively ( $O_W$  and  $\widehat{O_W}$ ).

The Figure 15 shows that, for a given azimuthal angle, both the CP-even and CP-odd components are sensitive to the transverse momentum. On the other hand, for a given transverse momentum, the CP-odd component exhibits a stronger dependence on the azimuthal angle, than the CP-even one. This is fairly consistent with our observations from Figure 11 , where we see that for the CP-odd operator, the eigenvalue corresponding to  $\Delta\phi_{jj}$  is by far the largest within the 1D variables. Furthermore, the magnitudes of the eigenvalues of the variables in the CP-even operator are not sufficiently different, therefore it is not entirely unexpected that the CP-even component would be roughly equally sensitive to both the variables, as seen in the left panel of Figure 15.

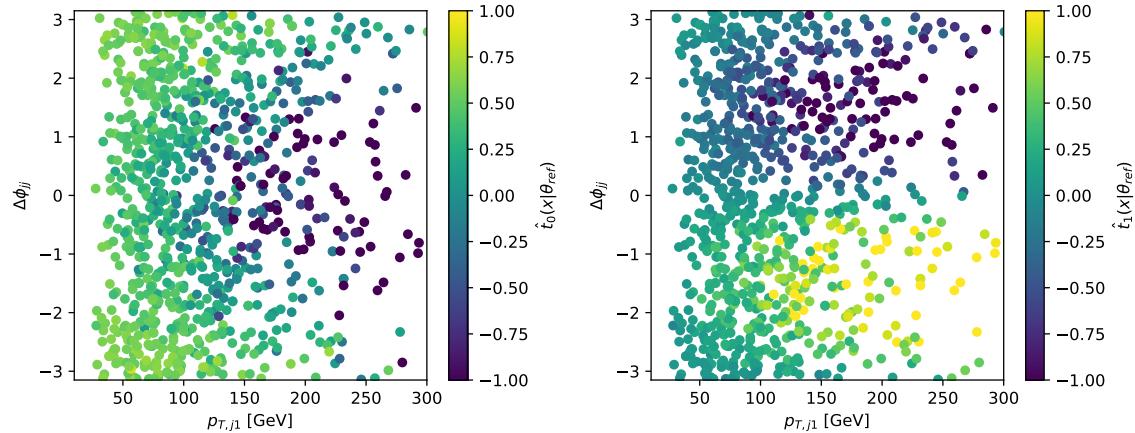


Figure 15: Score estimator components in the phase space spanned by  $p_{T,j1}$  and  $\Delta\phi_{jj}$ .

Now, having estimated the score components, we can use them as locally optimal observables for inference, in this case by filling 2D histograms to calculate the likelihood ratio.

## 5 Results

After having validated the two estimator models ALICES and SALLY, we can use them to reach our final objective and compute exclusion limits in the considered EFT parameter space.

It is instructive, for understanding the improvement brought by the two estimator models based on machine learning methods, to compare them with the traditional method based on a simple histogram. As an example, we choose the transverse momentum of the leading jet,  $p_{T,j1}$ , as the one-dimensional variable to be used in the histogram method. In Figure 16, we provide the histograms of the  $p_{T,j1}$  distributions corresponding to four different values of the CP-even Wilson coefficient  $c_W$  (colored lines) and the SM value  $c_W = 0$  (green), as well as the expected data based on an Asimov dataset following the SM hypothesis (black dots).

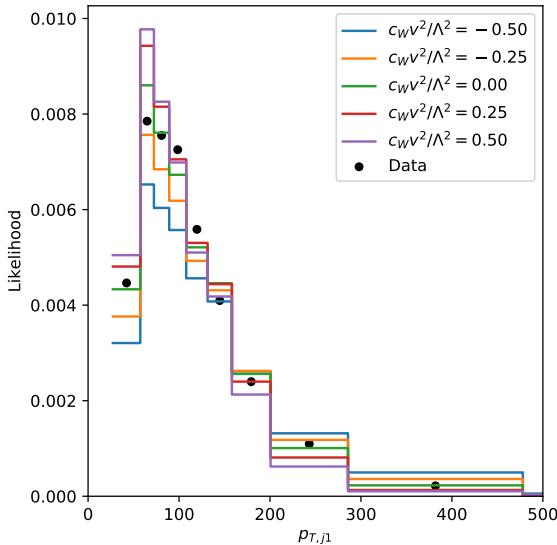


Figure 16: 1D-histogram showing the distribution of the likelihood for the transverse momentum of the leading jet. Four different values of the EFT Wilson coefficients  $c_W$  are shown, and the SM. The expected data for the SM hypothesis (black dots) are based on an Asimov dataset, comprising of 20% of the selected events at detector level.

We note that the histograms based on different  $c_W$  values are all different and distinct from each other. This reveals that the variable  $p_{T,j1}$  has a ‘discriminating power’, which can be exploited to infer EFT constraints.

The different likelihood ratio estimators are computed on the same grid of  $25 \times 25$  values in the range of  $[-1, 1]^2$ , in the EFT parameter space of the two considered operators ( $c_W v^2, c_{\widetilde{W}} v^2$ ), for the three methods studied (traditional based on histograms, ALICES and SALLY neural networks). These are then evaluated, for 2 different cases: first on an Asimov dataset based on the Standard Model, corresponding to an integrated luminosity of  $3000 \text{ fb}^{-1}$ , and second following a BSM hypothesis of  $(c_W v^2, c_{\widetilde{W}} v^2) = [0.5, 0.5]$ , considering 40 events (which is approximately the number of events expected at the SM). Subsequently, the p-values

are computed using the asymptotic approach [4].

In Figure 17, we plot the 68% CL expected exclusion limits in the phase space of the Wilson parameters, comparing the three approaches: based on a histogram of  $p_{T,j1}$  with 10 bins (blue), the SALLY method (green) and ALICES technique (orange). We also provide the best-fit points, indicated with a star, as estimated by minimizing of the log likelihood ratio. Furthermore the black and white color map indicates the expected p-values assuming the SM hypothesis, calculated by the ALICES method.

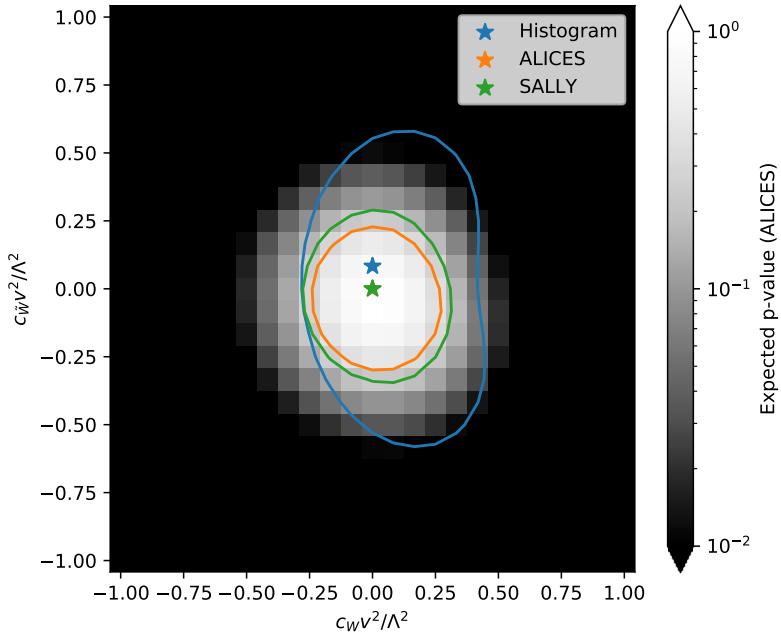


Figure 17: 68% CL expected contours in the EFT parameter space, with corresponding best-fit points using the histogram, ALICES and SALLY methods.

There are some critical observations that follow from Figure 17.

First, we note that the blue contour is much larger than the orange and the green ones. More quantitatively, when compared to the orange contour, it is approximately twice as large in the direction of the CP-odd operator, and five quarters in the direction of the CP-even operator. Also, we can deduce from here that the histogram method is evidently not as efficient as the ALICES and SALLY methods. Nevertheless, the histogram method does succeed in creating a closed contour, unlike the two unbounded vertical lines in Figure 12. This reveals that the histogram method is at least capable of constraining both the Wilson parameters, which the approximated Fisher information approach fails to accomplish. This is because the Fisher information matrix is in fact a linear approximation, and thus fails to capture the interference effects not in the vicinity of the SM.

It is worth mentioning here that the best-fit point of the traditional method does not exactly coincide with the origin (SM). This is related to the fact that the Asimov dataset does

not exactly match the SM template (Figure 16, green line), being the test dataset restricted to a fifth of the total dataset. However, the deviation from the origin is just one ‘pixel’ length distance, and hence can be considered to be sufficiently close to the origin.

Second, the orange contour is slightly smaller than the green contour, which reveals that the ALICES model performs better than the SALLY technique. This is not without reason: the SALLY technique uses the SM as its reference point. It follows that the optimal variables found with SALLY would be optimal only in the vicinity of the SM. On the contrary, in the ALICES method, the neural network learns the full likelihood ratio function throughout the entire parameter space. Thus, if trained sufficiently, the ALICES method would be sensitive even further away from the SM, thereby providing better EFT limits. This is also consistent with the observations in [2], where they compare the ALICES and SALLY techniques with the histogram technique, for a different physics case.

In our particular example, ALICES provides an expected confidence interval at 68% CL of the order of  $0.25/v^2 \sim 4 \text{ TeV}^{-2}$  in both directions for both coefficients. This improves the constraints from the traditional histogram method to around 1.25 times better in one direction (CP-even), and up to 2 times better in the other direction (CP-odd) with respect to the histogram method. This implies that the improvement, as manifested by the neural network approach, can be matched by the traditional method when supplemented with an equivalent integrated luminosity. Nevertheless, current analyses as in the case of ATLAS [12], for instance, are still based on traditional methods and can, therefore, be improved upon significantly.

Now that we have established ALICES as a more effective technique, we can inquire how well it can resolve data Beyond the Standard Model. Figure 18 compares the SM expectation estimated by ALICES model (orange contours) with the BSM expectation contours based on a BSM signal (red contours), also estimated with ALICES.

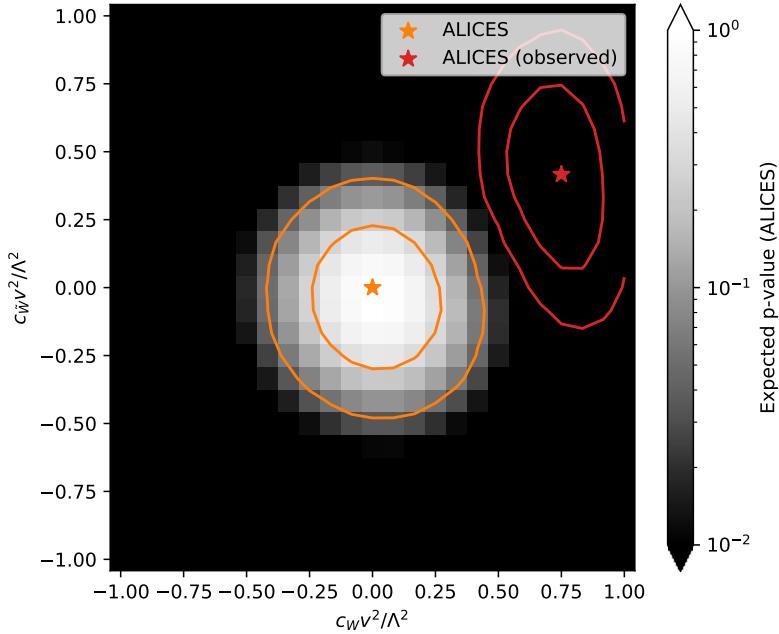


Figure 18: 68% CL and 95% CL expected contours for SM, and BSM observed contours in the EFT parameter space, corresponding to data generated at [0.5,0.5]. The contours and the best-fit points are calculated using the ALICES method.

We note that the true initial parameter [0.5,0.5] does not coincide exactly with the best-fit point. However, what is important to realise here is that it is contained within the 68% CL contour. Moreover, the two ALICES contours are well separated, and the best-fit point is more than  $2\sigma$  away from the SM. What we can infer from here is that the ALICES method is capable of distinguishing between different hypothesis, and thus becomes a powerful tool, specifically in the context of Beyond the Standard Model evidence.

## 6 Conclusions

One of the broad aims of this thesis was to find out how machine learning techniques can be utilized in particle physics, especially in the context of Beyond the Standard Model predictions. More specifically, I wanted to investigate how MadMiner, which of late has emerged as a powerful neural network tool, can be used in Effective Field Theory, to set exclusion limits for the Wilson coefficients.

I began this journey by first selecting the  $H \rightarrow ZZ \rightarrow 4\ell$  channel (where  $\ell = e, \mu$ ), also known as the ‘golden channel’, with the intent of ‘mining’ some ‘gold’ with MadMiner. The gold in this case refers to the joint likelihood ratio and the joint score, which are generated by the simulator and mined by MadMiner. For this process, I started by generating the raw Monte Carlo data at detector level, and developing a code for the event reconstruction and selection. Subsequently, the neural network models (ALICES and SALLY) were trained on the selected events, augmented with the ‘gold’. Finally, the estimators are utilized to draw EFT exclusion limits.

Furthermore, to underline the efficiency of MadMiner, we also provide a comparison with the traditional histogram method for drawing exclusion limits. Also, the Fisher information matrix method is explored, to provide additional physical insight into the chosen observables and EFT operators.

From the results it can be inferred that the approach of MadMiner is not only feasible, but also more efficient in performance than the traditional methods.

There of course remains scope for much more work to be done. For instance, we know that in the VBF production mode the chosen EFT operators affect both the production and the decay couplings of the Higgs boson. We can therefore study the effect of the new physics in the decay vertex, and subsequently combine the results with this study. Furthermore, we can optimize the neural networks to substantially improve the ability to draw exclusion limits. Additionally, while we chose the ‘golden channel’, in the context of ‘mining gold’, it would be worth exploring other channels too, golden or not. This also holds true for the different production modes which we have not investigated in this work. Moreover, for the sake of simplicity we have not considered the backgrounds and the systematics, which need to be considered for a more complete study. Also, this approach scales well for high dimensional parameter spaces. The present study, as well as the previous works, has not considered more than 2 operators at a time, for estimating exclusion limits. This clearly leaves the possibility of studying a considerable number of EFT operators simultaneously, something that this approach is very ideal for.

## References

- [1] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. A guide to constraining effective field theories with machine learning. *Physical Review D*, 98(5):052004, 2018.
- [2] Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer. Madminer: Machine learning-based inference for particle physics. *Computing and Software for Big Science*, 4(1):1–25, 2020.
- [3] Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. Constraining effective field theories with machine learning. *Physical Review Letters*, 121(11):111801, 2018.
- [4] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2):1–19, 2011.
- [5] Cranmer, Kyle. Constraining EFTs with machine learning. <https://indico.cern.ch/event/817757/contributions/3712517/attachments/1998425/3334673/PREFIT20-lecture2.pdf>, 3 2020.
- [6] Kunitaka Kondo. Dynamical likelihood method for reconstruction of events with missing momentum. i. method and toy models. *Journal of the Physical Society of Japan*, 57(12):4126–4140, 1988.
- [7] Till Martini and Peter Uwer. The matrix element method at next-to-leading order qcd for hadronic collisions: Single top-quark production at the lhc as an example application. *Journal of High Energy Physics*, 2018(5):141, 2018.
- [8] M. Diehl and O. Nachtmann. Optimal observables for the measurement of three gauge boson couplings in  $e+ e- \rightarrow W+ W-$ . *Z. Phys. C*, 62:397–412, 1994.
- [9] Johann Brehmer, Gilles Louppe, Juan Pavez, and Kyle Cranmer. Mining gold from implicit models to improve likelihood-free inference. *Proceedings of the National Academy of Sciences*, 117(10):5242–5249, 2020.
- [10] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [11] The CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, Sep 2012.
- [12] The ATLAS collaboration. Higgs boson production cross-section measurements and their eft interpretation in the  $4\ell$  decay channel at  $\sqrt{s} = 13$  tev with the atlas detector. 2020.
- [13] Johann Brehmer, Kyle Cranmer, Felix Kling, and Tilman Plehn. Better higgs boson measurements through information geometry. *Physical Review D*, 95(7):073002, 2017.
- [14] Johann Brehmer, Felix Kling, Tilman Plehn, and Tim MP Tait. Better higgs-c p tests through information geometry. *Physical Review D*, 97(9):095017, 2018.

- [15] Johan Alwall, R Frederix, S Frixione, V Hirschi, Fabio Maltoni, Olivier Mattelaer, H-S Shao, T Stelzer, P Torrielli, and M Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, 2014.
- [16] Celine Degrande, Nicolas Greiner, Wolfgang Kilian, Olivier Mattelaer, Harrison Mebane, Tim Stelzer, Scott Willenbrock, and Cen Zhang. Effective field theory: a modern approach to anomalous couplings. *Annals of Physics*, 335:21–32, 2013.
- [17] Kaoru Hagiwara, Sumio Ishihara, R Szalapski, and D Zeppenfeld. Low energy effects of new interactions in the electroweak boson sector. *Physical Review D*, 48(5):2182, 1993.
- [18] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, 2008.
- [19] J De Favereau, Christophe Delaere, Pavel Demin, Andrea Giammanco, Vincent Lemaitre, Alexandre Mertens, Michele Selvaggi, Delphes 3 Collaboration, et al. Delphes 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2):57, 2014.
- [20] Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV. Technical Report CMS-PAS-HIG-19-001, CERN, Geneva, 2019.
- [21] CMS Collaboration. The phase-2 upgrade of the cms endcap calorimeter. Technical report, 2017.
- [22] Markus Stoye, Johann Brehmer, Gilles Louppe, Juan Pavez, and Kyle Cranmer. Likelihood-free inference with an improved cross-entropy estimator. *arXiv preprint arXiv:1808.00973*, 2018.

## Acknowledgements

First and foremost, I would like to thank my advisors Dr. Roberto Salerno and Matteo Bonanomi, who have guided me into this journey with great dedication, even from distance. The confinement has surprised us all, and I consider myself extremely lucky to have gone through it with their constant support.

I am also truly grateful to the MadMiner team: Kyle Cranmer, Johann Brehmer and Felix Kling. Their help has been of key importance in the progress and understanding of my work, so I would like to thank them for their kindness, up until the very last minute and beyond.

Words cannot be enough to thank my friends in Maison de l'Inde, with whom I shared the entirety of my days: without them, nothing could have been possible. In particular, special credits to my friend Saagnik, for his amazing illustration of the *Mad Miner*.

A final special thank you goes to my family, for their continuous loving support and encouragement in my studies.