

# Search for heavy neutrinos in a 3-lepton final-state

## Applications using supervised machine learning

by

William Hirst

THESIS

for the degree of

MASTER OF SCIENCE



Faculty of Mathematics and Natural Sciences  
University of Oslo

Autumn 2022



Search for heavy neutrinos in a 3-lepton final-state

Applications using supervised machine learning

William Hirst



© 2022 William Hirst

Search for heavy neutrinos in a 3-lepton final-state

<http://www.duo.uio.no/>

Printed: Reprosentralen, University of Oslo



# Abstract

This will be the abstract.





# Acknowledgments

First of all I would like to thank my supervisor, Are Raklev, for giving me the trust and encouragement to explore what I found most fascinating on the subject of this thesis. Your enthusiasm for physics and dedication to detail has been of enormous value. I am grateful that you always found time to answer my questions and put me back on track. I also wish to thank Anders Kvellestad, Jeriek Vda and Eli Rye for all the help and encouragement.

Thanks to the whole theory group for making these last years so enjoyable, and a special thanks to Cecilie Glittum, a good friend and vital physics-partner the last five years. I also would like to thank Kristian Wold, Lars Dean and Jonathan Waters for all the fruitful knowledge exchanges and especially for the warm friendships. Many thanks to my family, for always supporting and believing in me.

Finally, huge thanks to my girlfriend Elin. Words cannot describe how grateful I am for all her love, support and patience. This thesis would not have been possible without her.



# Contents

<b>Introduction</b>	<b>1</b>
<b>1 The Standard model of elementary particles</b>	<b>3</b>
1.1 Phenomenology - What is it?	3
1.2 The background channels	3
<b>2 Beyond the standard model - Heavy neutrinos</b>	<b>5</b>
2.1 Why look beyond?	5
2.2 Neutrinos	5
2.3 Dirac and Majorana	5
2.3.1 How to distinguish the two?	5
<b>3 Statistical and multi-variabel analysis</b>	<b>7</b>
3.1 Discovery and significance	7
3.2 Exclusion	7
3.2.1 Likelihood	7
3.3 High- and Low-level features	7
<b>4 Introduction to supervised and unsuperised machine learning</b>	<b>9</b>
4.1 Neural Networks in physics	9
4.1.1 Deep vs Shallow networks	9
4.2 Gradient Boosting and decision trees	9
<b>5 Implementation</b>	<b>11</b>
5.1 Features	11
5.1.1 Cuts and triggers	11
5.1.2 Lepton variables selection	11
5.1.3 Jet variables selection	11
<b>Appendices</b>	<b>13</b>
<b>Appendix A</b>	<b>15</b>
A.1 Light-Cone Coordinates	15



# Introduction

The standard model (SM) is perhaps one of the most successful scientific theories ever created. It accurately explains the interactions of leptons and quarks as well as the force carrying particles which mediate said interactions. In 2012 the SM achieved one of its crowning achievements when we discovered the Higgs boson. Much of the accolade was rightfully given to the theoretical work on the SM, but another aspect of the discovery was equally important. Data analysis was and is a crucial part of any new discovery in physics. One of the most important and exiting tools is machine learning.

## **Outline of the Thesis**



# Chapter 1

## The Standard model of elementary particles

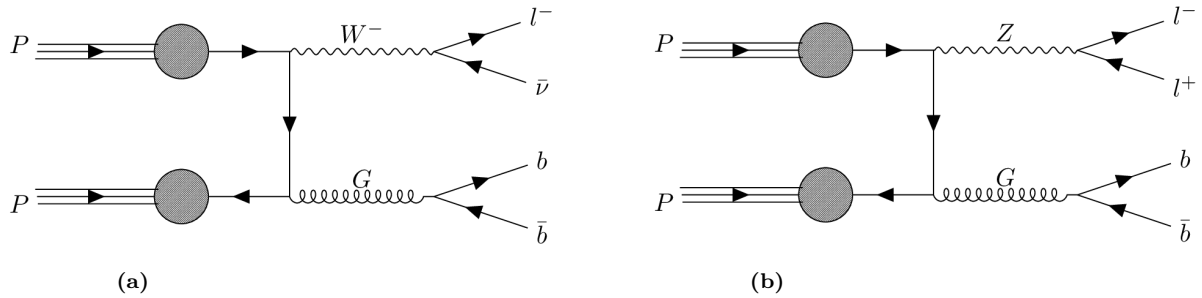
The standard model (SM) is perhaps one of the most successful scientific theories ever created. It accurately explains the interactions of leptons and quarks as well as the force carrying particles which mediate said interactions. In 2012 the SM achieved one of its crowning achievements when we discovered the Higgs boson.

### 1.1 Phenomenology - What is it?

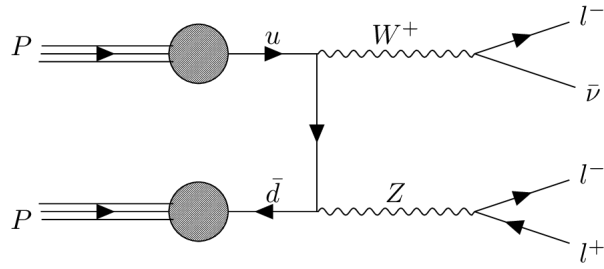
### 1.2 The background channels

The dominant SM backgrounds can be divided into two categories: (i) from leptonic  $\tau$  decays and (ii) from fake leptons. In the first category, the dominant process is the pair production of  $WZ$  with  $W$  decaying leptonically and  $Z \rightarrow \tau\tau$ . The trilepton final states with no-OSSF pairs can arise from the subsequent leptonic decay of  $\tau$ 's. We estimate this background process via Monte Carlo simulations.

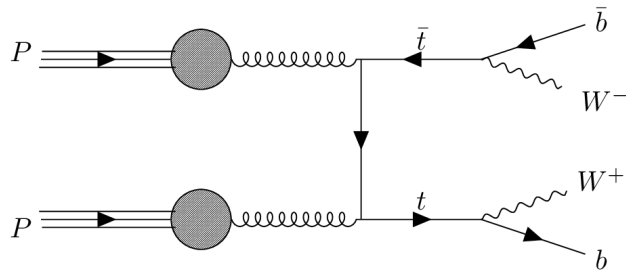
The dominant processes of the second category are  $\gamma^*/Z + \text{jets}$  and  $t\bar{t}$ , where two leptons come from  $\gamma^*/Z \rightarrow \tau\tau$  or the prompt decay of  $t$  and  $\bar{t}$ , and a third lepton is faked from jets containing heavy-flavor mesons.



**Figure 1.1:** The Feynman diagram of both the  $W$ +jets [1.1a](#) and the  $W$ +jets [1.1b](#).



**Figure 1.2:** The Feynman diagram of the diboson WZ-channel.



**Figure 1.3:** The Feynman diagram of the  $t\bar{t}$ -channel.



## Chapter 2

# Beyond the standard model - Heavy neutrinos

### 2.1 Why look beyond?

### 2.2 Neutrinos

### 2.3 Dirac and Majorana

#### 2.3.1 How to distinguish the two?



## Chapter 3

# Statistical and multi-variabel analysis

The standard model (SM) is perhaps one of the most successful scientific theories ever created. It accurately explains the interactions of leptons and quarks as well as the force carrying particles which mediate said interactions. In 2012 the SM achieved one of its crowning achievements when we discovered the Higgs boson. Much of the accolade was rightfully given to the theoretical work on the SM, but another aspect of the discovery was equally important. Data analysis was and is a crucial part of any new discovery in physics. One of the most important and exiting tools is machine learning.

### 3.1 Discovery and significance

### 3.2 Exclusion

#### 3.2.1 Likelihood

### 3.3 High- and Low-level features



## Chapter 4

# Introduction to supervised and unsuperised machine learning

This will give a brief intoruction to the concept of machine learning as well as the difference between supervised and unsuperised.

### 4.1 Neural Networks in physics

#### 4.1.1 Deep vs Shallow networks

### 4.2 Gradient Boosting and decision trees

In this rapport I will use the XGBoost-classifier which uses gradient-boosted trees. Gradient-boosting is a machine learning algorithm which uses a collective of "weak" classifiers in order to create one strong classifier. In the case of gradient-boosted trees the weak classifiers are a collective of shallow trees, which combine to form a classifiers that allows for deeper learning. As is the case for most gradient-boosting techniques, the collecting of weak classifiers is an iterative process.

We define an imperfect model  $\mathcal{F}_m$ , which is a collective of m number of weak classifiers, estimators. A prediction for the model on a given data-points,  $x_i$  is defined as  $\mathcal{F}_m(x_i)$ , and the observed value for the aforementioned data is defined as  $y_i$ . The goal of the iterative process is to minimize some cost-function  $\mathcal{C}$  by introducing a new estimator  $h_m$  to compensate for any error,  $\mathcal{C}(\mathcal{F}_m(x_i), y_i)$ . In other words we define the new estimator as:

$$\tilde{\mathcal{C}}(\mathcal{F}_m(x_i), y_i) = h_m(x_i), \quad (4.1)$$

where we define  $\tilde{\mathcal{C}}$  as some relation defined between the observed and predicted values such that when added to the initial prediction we minimize  $\mathcal{C}$ .

Using our new estimator  $h_m$ , we can now define a new model as

$$\mathcal{F}_{m+1}(x_i) = \mathcal{F}_m + h_m(x_i). \quad (4.2)$$

The XGBoost [?] framework used in this analysis enables a gradient-boosted algorithm, and was initially created for the Higgs ML challenge. Since the challenge, XGBoost has become a favorite for many in the ML community and has later won many other ML challenges. XGBoost often outperforms ordinary decision trees, but what is gains in results it looses in interpretability. A single tree can easily be analysed and dissected, but when the number of trees increases this becomes harder.



# Chapter 5

## Implementation

### 5.1 Features

The choice of which features to study and which to neglect are crucial in a search for new physics. This is particularly true in the case of applying machine learning. The general motivation for including a given feature can be based on several factors. The first being its ability to provide a trend which we as researches can exploit when creating our regions. By this I mean that it is a variable where there is diversity in distribution between the different channels. The second motivation is grounded in physics. Often we as physicists tend to lean towards variables we know have some effect on the physics we are studying. For example the variable  $E_T^{miss}$ , can be directly used to either include or discard events where we do or do not expect final states with sufficient missing energy. The final motivation is grounded in the MC-simulations ability to represent the variable. If there seems to be a clear deviation between the real and MC-data which does not stem from any new physics, we tend to discard them from the analysis.

#### 5.1.1 Cuts and triggers

#### 5.1.2 Lepton variables selection

Now we will have a look at what variables from the leptons that were included in the analysis. All low level information on the momentum of the leptons were added into the dataset: i.e the transverse momentum  $P_t$ , the pseudo rapidity  $\eta$  and the azimuthal angle  $\phi$ . All momentum features were represented individually for each lepton. For example  $P_t$  was added as three columns,  $P_t(l_1)$ ,  $P_t(l_2)$  and  $P_t(l_3)$ , where the ordering of the leptons were based on the momentum from highest ( $l_1$ ) to lowest ( $l_3$ ). Similarly I added information regarding the charge ( $\pm$ ) and flavor (electron, muon) of each lepton. Based on the momentum variables the transverse mass  $m_t$  of each lepton was calculated and included along with the missing transverse energy  $E_t^{miss}$  of the final state.

The variables described in the section above are often considered as low-level features. These are very useful in many (if not all) searches and contribute little to no bias to your analysis. But, in the case of final-state specific searches such as mine, one can allow one self of adding physics motivated higher-level physics. The higher level features calculated in this thesis were inspired by [1] (ATLAS 2022).

Firstly I added different mass variables, namely  $m_{ll}$  and  $m_{ll}(OSSF)$ . The first being the dilepton invariant mass and the latter being the dilepton invariant mass of the pair with OSSF. In the case of more than one possible OSSF-pair, the pair with the highest invariant mass was chosen. Secondly I added variables composed of the sum of different set of momentum. These variables are the sum of all three leptons  $H_t(lll)$ , of the pair with opposite sign  $H_t(SS)$  and the sum of the momentum for all three leptons added with the missing transverse energy  $H_t(lll) + E_t^{miss}$ .

#### 5.1.3 Jet variables selection





# Appendices



# Appendix A

## A.1 Light-Cone Coordinates

Light-cone coordinates is specifically useful in high energy scattering processes where one want to decompose the momentum of the involving particles. For a general four vector  $p^\mu$ , one defines

$$p^\mu = (p^+, p^-, p_\perp), \quad (\text{A.1})$$

where

$$p^+ = \frac{1}{\sqrt{2}}(p^0 + p^3) \quad (\text{A.2})$$

$$p^- = \frac{1}{\sqrt{2}}(p^0 - p^3) \quad (\text{A.3})$$

$$p_\perp = (p^1, p^2). \quad (\text{A.4})$$

Scalar products are given by

$$p \cdot k = p^+ k^- + p^- k^+ - p_\perp \cdot k_\perp \quad (\text{A.5})$$

$$p^2 = 2p^+ p^- - p_\perp^2, \quad (\text{A.6})$$

where the transverse contraction  $p_\perp \cdot k_\perp$  is understood from the definition of the transverse vector and must not be mistaken as the same as the four momentum contraction  $p \cdot k$ . We will usually parametrize our momenta in terms of plus-components and from Eq. (A.6) it follows that the minus component can be written as

$$p^- = \frac{p^2 + p_\perp^2}{2p^+}. \quad (\text{A.7})$$

The  $d$ -dimensional Jacobian takes the form

$$d^d p = dp^+ dp^- d^{d-2} p_\perp. \quad (\text{A.8})$$

From the above relations the light-cone metric takes the form

$$g_{\text{LC}}^{\mu\nu} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (\text{A.9})$$

where the index runs over  $\mu = +, -, 1, 2$ . We will elsewhere drop the subscript LC as it will always be clear from the context when we are using light-cone coordinates. One can also define light-like basis vectors

$$n_+^\mu = (1^+, 0^-, 0_\perp), \quad n_{+\mu} = (0^+, 1^-, 0_\perp), \quad (\text{A.10})$$

$$n_-^\mu = (0^+, 1^-, 0_\perp), \quad n_{-\mu} = (1^+, 0^-, 0_\perp), \quad (\text{A.11})$$

giving

$$n_+^2 = 0, \quad n_-^2 = 0, \quad n_+ \cdot n_- = 1. \quad (\text{A.12})$$

These basis vectors project out the following components of a vector

$$p \cdot n_+ = p^- , \quad p \cdot n_- = p^+ . \quad (\text{A.13})$$

We can also construct a transversal metric

$$g_{\perp}^{\mu\nu} = g^{\mu\nu} - (n_+^\mu n_-^\nu + n_+^\nu n_-^\mu) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} , \quad (\text{A.14})$$

from which it follows that

$$g_{\perp}^{\mu\nu} g_{\perp\,\mu\nu} = 2 . \quad (\text{A.15})$$

We can also define the gluon polarization sum in light-cone gauge, i.e  $A^+ = 0$ , as

$$\sum_{\text{pol}} \varepsilon_\alpha(k') \varepsilon_\beta^*(k') = -g_{\alpha\beta} + \frac{k'_\alpha n_{-\beta}}{k' \cdot n_-} + \frac{k'_\beta n_{-\alpha}}{k' \cdot n_-} . \quad (\text{A.16})$$

# Bibliography

- [1] M. Franchini, K. H. Mankinen, G. Carratta, F. Scutti, A. Gorisek, E. Lytken et al., *Search for type-III seesaw heavy leptons in dilepton final states in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, .