Final Year Project Proposal

Top Quark Reconstruction with The CMS Detector

SIF3004 Project

Department of Physics, Universiti Malaya

Mohamad Zuhaily Mahmud

U2004495

Supervisor

Dr. Nurfikri Nor Joharuddeen

24th October 2022

Table of Contents

1 Introduction

1.1 Problem Statement

What is the most efficient methodology to reconstruct hadronically decaying top quark.

1.2 Objectives

Reconstruct hadronically decaying top quark using various method such as chi-square fit and machine learning.

1.3 CERN, LHC & CMS

Conseil Européen pour la Recherche Nucléaire or CERN is an European laboratory for particle physics located near Geneva, Switzerland. They consist of thousands of scientists and engineers aiming to study the basic constituent of matter (Sutton, 2022).

CERN make use the world largest and most complex scientific equipment to achieve this which is the Large Hadron Collider (LHC). Using collision of subatomic particle accelerated near the speed of light inside the LHC, they study how these particles interact with each other.

The LHC first ran is on 10th September 2008. From its first run until today, the operation process has been stopped three times. Once due to technical problem and the other two are for upgrading the LHC equipment. It was first stopped due to an electrical problem in the cooling system, causing a rise in the equipment temperature at about 100˚C on 18th September 2008 and later resume operation on 20th November 2009. Later in March 2010, CERN announce that there is a problem with the superconducting wire, requiring it to be operated only at half-energy (7TeV). It is shutdown in February 2013 to fix the problem. Year 2010 until 2013 is considered the first run of LHC. It then resumes operation in April 2015. In December 2018, there was another long shutdown in order to upgrade the LHC equipment which ended recently in July 2022. (Jones, 2022). We can summarise these into three runs which is Run 1 in 2010-2013, Run 2 in 2015-2018 and Run 3 is currently running.

LHC is a powerful particle accelerator that consists of 27 kilometres ring of superconducting magnets and is located around 50-175 metres underground. Two particle beam, which are usually proton, is accelerated in opposite direction near the speed of light in the ring to be collided and is guided using these superconducting magnets at ultra-high vacuum space. This require the magnets to be chilled to -271.3˚C using a distribution system of liquid helium. Plus, there is another set of special magnets near the collision area to “squeeze” the particle close together to increase the probability of collision (CERN, The Large Hadron Collider, n.d.).

LHC consists of multiple detectors. One of them is the Compact Muon Solenoid (CMS). It is located at one of the collisions point in the LHC. CMS act as a giant, high speed camera, which capture particle collision from all direction up to 40 million times each second (CERN, CMS, n.d.). The transverse slice of CMS is shown in figure 1.

2 Literature Review

2.1 Top Quarks

Top quarks are one of the fundamental particles in the SM with a mass of 173 ±0.27 GeV/c^2 with +2/3 charge and is located in the third generation particle. The invariant mass of top quarks is an interesting quality as it contributes a large portion of the electroweak radiative correction for the particles Z, Higgs and photon (Naderi, 2017).

By studying the top quarks particle, we can understand and explain the nature of the Higgs particles in more details. Higgs is relatively new particles that had just been discovered back in 2012 by the ATLAS and CMS experiments at the LHC, CERN near Geneva, Switzerland. The Higgs boson is a special particle that gives mass to fundamental particles and itself. The Higgs decays predominantly to $b\bar{b}$, yielding signals that is quite like $t\bar{t}$. This entails that we can integrate the $t\bar{t}$ identification into the Higgs boson analysis.

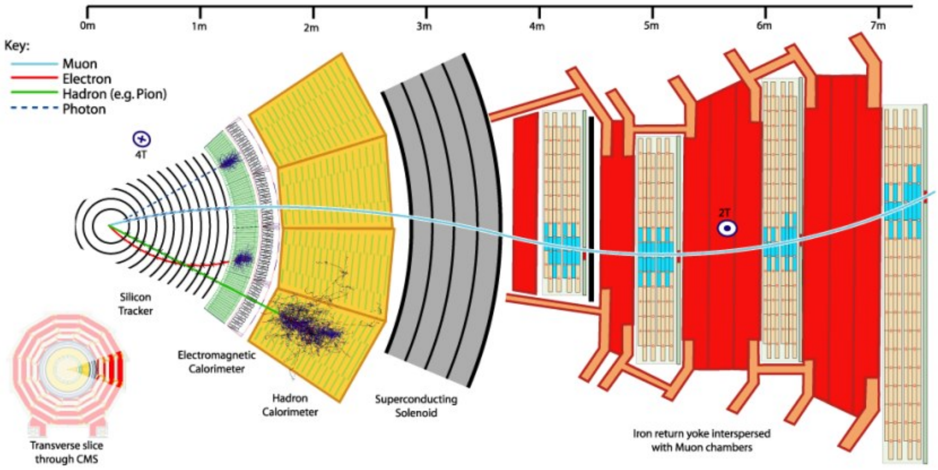

Figure 1: Transverse slice of CMS, showing the different sub-detectors and how different particles interact

2.2 Top Quark Production

So far, $t\bar{t}$ production that have been recorded are from Tevatron and LHC. From the CDF and D0 experiments at Tevatron, they are able to conclude a combined value of cross-section for $t\bar{t}$ production which is $\sigma_{t\bar{t}}$ = (7.60 ± 0.20 (stat) ± 0.36 (syst)) pb, assuming a top-quark mass of m_t = 172.5 GeV with precision of 5.4%. Meanwhile, in 2015 at the LHC, a $p\bar{p}$ collisions at a center-of-mass energy of 5 TeV results in an inclusive $t\bar{t}$ production. The inclusive $t\bar{t}$ cross section was measured using a data sample corresponding to an integrated luminosity of 26 pb−1.

2.3 Top Quark Decay

The decay of $t\bar{t} \rightarrow W^+bW^-\bar{b}$ has a branching ratio (BR) close to unity described by the Cabibo-Kobayashi-Maskawa (CKM) matrix, $V_{tb} \approx 0.9992$, meaning that this decay is exclusively used for top-quark measurement. There is also a very small chances that it would decay into strange quark and down quark, $V_{tb} \approx 0.0387$ and $V_{tb} \approx 0.0084$ respectively. The W can decay hadronically or leptonically. This leads to the final states of $t\bar{t}$ events to be either "lepton+jets" ($t\bar{t} \rightarrow \ell^+\text{vbq}\bar{q}'\bar{b}$ and $\text{qq}'\text{b}\ell^-\bar{\text{v}}\text{b}$), "all jets" ($t\bar{t} \rightarrow \text{q}\bar{\text{q}}'b\bar{q}q'\bar{b}$) and "dileptons" ($t\bar{t} \rightarrow \ell^+\text{vb}\ell^-\bar{v}\bar{b}$). The measurement performed at LHC of the top quark decay process results in a solid agreement with the SM prediction (Altonen & et. al, 2012).

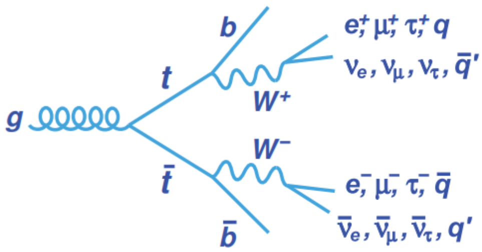

Figure 2: Leading-order Feynman diagram for $t\bar{t}$ decay

Based on the figure 3, top quark is most commonly produced in high energy collision. (R. L. Workman et al. (Particle Data Group), 2022)

2.4 Current Studies on Top Quarks

Today, most top quarks investigations are performance studies and precision measurement. What technique is the best to reconstruct the top quarks with the most precision efficiently. By efficiency, we are talking about how fast the analysis can be, how precise it is, how many memory storage consumption can we reduce and many more. Slight difference in measurement could indicate a hint to new particle or new physics that describe fundamental particle interaction beyond the Standard Model (SM). However, in the search of something new beyond the SM, there are a much more complicated works that need to be done. Some of the challenges in reconstructing top quark decay are:

i. $t\bar{t}$ that decay leptonically produce neutrino that is hard to detect. This led to reconstruction based only on conservation of momentum.
ii. Assigning jets to their origin is a challenge by itself as one collision produced multiple jets output.

Various studies have been done prior to this date in term of reconstructing top quark production efficiently. Most of them turned to deep machine learning or neural network in reducing the complexity of reconstructing top quark. They use kinematic fitting to identify which leptons and jets that is produced via $t\bar{t}$ decay. Naderi (2017) for example make use of the δR method which select two closest jet to reconstruct the decay. Meanwhile Greif and Lannon (2019) make use of Lorentz's four momentum concept as an inspiration for their Lorentz Neural Network (LNN). However, most of the works make use of the chi-squared method (χ^2). They pair up two to three jets and test their combination whether they belong to the top quark and took the best-fit combination as the solution.

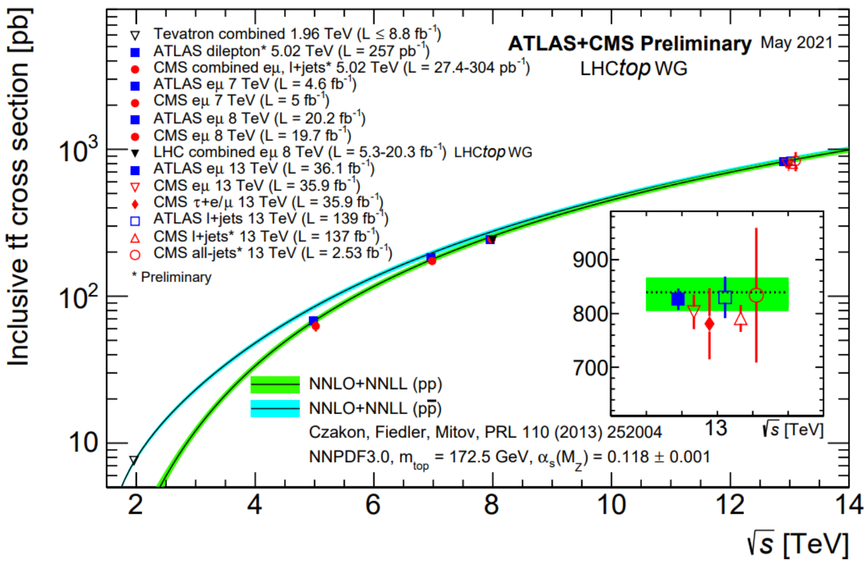

Figure 3: Measured and predicted $t\bar{t}$ production cross sections from Tevatron energies in $p\bar{p}$ collisions to LHC energies in pp

Measurement	$\sigma_{t\bar{t}}$ (pb)	stat (%)	exp (%)	th (%)	lumi (%)
ATLAS $e\mu$ 7 TeV	182.9	1.7	2.3	2.0	1.8
CMS $e\mu$ 7 TeV	173.6	1.2	$^{+2.6}_{-2.3}$		2.2
ATLAS $e\mu$ 8 TeV	242.4	0.7	2.3	3.1	1.7
CMS $e\mu$ 8 TeV	244.9	0.6	$^{+2.6}_{-2.2}$		2.6
ATLAS $e\mu$ 13 TeV	818	1.0	3.3		2.7
CMS e/μ+jets 13 TeV	834.6	0.3	2.7		2.7

Table 1: Summary of most precise inclusive $t\bar{t}$ cross section measurements from the ATLAS and CMS

3 Brief Project Overview

3.1 Analytical Techniques

Based on literature review, the most effective way to reconstruct top quarks are by using machine learning techniques. Due to this, the project aims to use machine learning to reconstruct hadronically decaying top quarks. Techniques from Naderi (2017), the δR and the χ^2 techniques are a good starting point to start this project. Especially the χ^2 technique since majority of the works are using it. Other than that, we can also use top quark analysis object for jets such as the jet algorithm, jet energy scale and resolution, b-tagging algorithm, missing transvers momentum and the flow particle approach (Husemann, 2017) are also few of the option approaches to reconstruct the hadronically decaying top quark.

3.2 Computational Studies

Since most of the previous studies concluded that machine learning is the best way to approach the problem, the need to understand and utilize machine learning is an absolute. Using python language, two suitable libraries are the Scikit and TensorFlow. Both are a free machine learning libraries for Python. The reconstruction of top quark also will make use of the ROOT library designed by CERN specifically for high energy physics research. This will be run on Windows subsystem Linux (WSL). Secure Shell Protocol (SSH) to CERN machine would also help in running the program since thousands to millions of data would be used in this project to train the machine learning.

3.3 Monte-Carlo samples

This project will use Monte-Carlo generated samples to train the machine learning program to recognize the jets that belongs to the top quark. Performance studies will then be used to rate how a certain analysing technique performed.

3.4 Resources

This project is under computational physics category as it studies the computational method to reconstruct the top quark decay. Thus, this project will be solely being done on the computer whether it is locally or using CERN server as the hardware to run the programs. Fortunately, Dr Nurfikri Norjoharuddeen has provide the CMS CERN membership to all his undergraduate final year project students. Being one of the official CMS members, we are now able to utilize some of the utility provided by CMS.

The programs that are specifically installed to run this project are in table 2.

No	Program
1	Windows Subsystem Linux (WSL)
2	Visual Studio Code (VS Code)
3	Miniconda environment
4	ROOT library
5	WinSCP

Table 2: Computer programs that are going to be use in this project

All the programs that are listed in table 1 are free to use. Another important equipment that is required is a good internet connection as this project will use a lot on CERN online utilities applications. The most important one is the CERN Box.

5 Expected Result

The final goal of this project is to build a machine learning based top quark reconstruction analysis. Being able to create a few of the analysis techniques and compare their performance efficiency is already a good enough result.

From the literature review, we can hypothesis that the χ^2 techniques will probably be the most efficient one since most of the papers conclude so. Meaning that this will likely end up as a test on χ^2 performance against other methods.

References

Husemann, U. (2017). Top-Quark Physics: Status and Prospects. *Progress in Particle and Nuclear Physics, 95*, 48-97. doi:https://arxiv.org/abs/1704.01356v2

Altonen, T., & et. al. (2012). Combination of the top-quark mass measurements from the Tevatron collider. *Phys. Rev. D, 86*. doi:http://arxiv.org/abs/1207.1069v4

Brooijmans, G. (2008). Top Reconstruction Algorithms in ATLAS. *3rd Top Workshop, Grenoble, Oct 23-25 2008.*

CERN. (2012). Measurement of the Top Quark Mass with the Template Method in the t[bar t] → lepton + jets Channel Using ATLAS Data. *The European Physical Journal C, 72.6*. doi:http://dx.doi.org/10.1140/epjc/s10052-012-2046-6

CERN. (n.d.). *CMS*. Retrieved from CERN: https://home.cern/science/experiments/cms

CERN. (n.d.). *The Large Hadron Collider*. Retrieved from CERN: https://www.home.cern/science/accelerators/large-hadron-collider

Erdman, J., Kallage , T., Kröninger, K., & Nackenhorst, O. (2019). From the bottom to the top—reconstruction of $t\bar{t}$ events with deep learning. *Journal of Instrumentation, 14*. doi:https://doi.org/10.1088/1748-0221/14/11/P11015

Gallinaro, M. (2013). Top quark physics: A tool for discoveries. *Journal of Physics: Conference Series, 447*. doi:doi:10.1088/1742-6596/447/1/012012

Geron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow.* CA, United States of America: O'Reilly.

Greif, K., & Lannon, K. (2019). Physics Inspired Deep Neural Networks for Top Quark Reconstruction. *EPJ Web of Conferences, 245*. doi:https://doi.org/10.1051/epjconf/202024506029

Jones, D. G. (31 August, 2022). *Large Hadron Collider*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/technology/Large-Hadron-Collider

Naderi, K. (2017). *Deep Learning Techniques for Top-Quark Reconstruction.* Sharif University of Technology.

R. L. Workman et al. (Particle Data Group). (2022). Particle Physics Booklet. *Prog. Theo. Exp. Phys., 2022*. Retrieved from https://pdg.lbl.gov/order

R. L. Workman et al. (Particle Data Group). (2022). Review of Particle Physics. *Prog. Theor. Exp. Phys., 2022*. doi:DOI: 10.1093/ptep/ptac097

Riley, K. F., Hobson, M. P., & Bence, S. J. (2006). *Mathematical Method for Physics and Engineering Third Edition.* Cambridge University Press.

Santos, R., Webster, J., Ryu, S., Adelman, J., Chekanov, S., & Zhou, J. (2016). Machine learning techniques in searches for tth¯ in the h → b¯b decay channel. doi:https://arxiv.org/abs/1610.03088v4

Schramm, S. (n.d.). Jet reconstruction (experimental) Part 1. University of Geneva.

Sutton, C. (2022). *CERN*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/topic/CERN