Constraining some SMEFT operator using LHC data

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Prasham Jain

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2020

Supervisor: Dr. Diptimoy Ghosh

Certificate

This is to certify that this dissertation entitled Constraining some SMEFT operator using LHC data towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Prasham Jain at Indian Institute of Science Education and Research under the supervision of Dr. Diptimoy Ghosh, Assistant Professor, Department of Physics , during the academic year 2019-2020.

Prasham Jain

Dr. Diptimoy Ghosh

Committee:

Dr. Diptimoy Ghosh

Dr. Arun Thalapillil

This thesis is dedicated to my late grandmother, Mrs. Vijaya Jain.

Declaration

I hereby declare that the matter embodied in the report entitled Constraining some SMEFT operator using LHC data are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Diptimoy Ghosh and the same has not been submitted elsewhere for any other degree.

Prasham Jain

Dr. Diptimoy Ghosh

Acknowledgments

Firstly, I would like to express my deepest gratitude to my supervisor Dr. Diptimoy Ghosh for his invaluable guidance. Every step of the way, we had countless enlightening and wonderful physics discussions. I have learned to look at physics problems with a wider perspective and deeper understanding. Not only has he helped me become a better physicist but also a responsible researcher.

Numerous factors have assisted my academic and personal growth. I have been fortunate enough to be surrounded by excellent professors at IISER Pune. I am grateful to Dr. Arun Thalapillil for his valuable contributions to this thesis. Discussion sessions with him were certainly very stimulating. A special thanks to Dr. Sourabh Dube, who has been more than a great academic influence on me. He has been very encouraging, supportive, and friendly. I would also like to thank Dr. Seema Sharma for giving me pep talks at the times I needed them the most.

My family has always stood with me with their constant support throughout any ups and downs. My mother and father have shown overwhelming faith in me and my decisions. I am the person I am because of them.

Finally, I would like to thank my close friends and colleagues: Meher, Rahul, Prasanna, Shomik, Niramay, Raj, Palash and Abhishek for always supporting and tolerating me. You guys are awesome. I will always cherish our discussions, debates, coffee breaks, meals, and silly moments. Thank you.

Abstract

The discovery of the Higgs boson at the Large Hadron Collider (LHC) is the crowning achievement of the Standard Model (SM) of particle physics. The SM is now proven to be a consistent field theory describing a vast range of phenomena at the electroweak scale and below.

In this success of theoretical and experimental efforts, however, lies a pressing question: what is the underlying physics at length scales shorter than the electroweak (EW) scale $\sim$100 GeV^{-1}? Or turning it around, till what length scale are the rules of SM valid? One way to study this question in a relatively model independent approach is to look for non-zero coefficients of local higher dimensional operators of mass dimension greater than four. These coefficients, the so-called Wilson Coefficients (WCs), can be directly constrained from experimental data collected at the LHC. Another approach to put bounds on these WCs is developed using Machine Learning.

In the master's thesis, my aim is to put bounds on some WC of the Standard Model Effective Field Theory (SMEFT), using processes involving the top quark. This thesis sheds light on some preliminary theory, followed by Cut and Count analysis to obtain bounds on the operator, and then methods in Machine Learning to improve the bound. For $\Lambda = 1$ TeV, the strongest bound at $\mathscr{L}$=35.8 fb^{-1} obtained using Cut and Count analysis with $M_{t\bar{t}} - p_{\mathrm{T}}^{t}$ distribution is $C_G \in [-0.259, 0.225]$.

List of Tables

List of Figures

Contents

Chapter 1

Introduction

The Standard Model (SM) of particle physics has been very successful in explaining strong and electroweak interactions at currently accessible energies. However, a few phenomena/anomalies indicate that SM is not a complete description of nature. It does not account for dark matter, dark energy, does not explain gravitational force, strong CP problem, etc. Consequently, we look for possible new physics (NP) particles and interactions at higher energy scales.

Beyond SM or BSM physics can be analyzed and parametrized without loss of generality using SMEFT (SM Effective Field Theory) [3, 4]. While preserving the SM gauge symmetries, higher-dimensional operators can be augmented. This approach is justified when the scale upto which the EFT is valid is given by Λ. This scale Λ is taken to be constrained as $v/\Lambda < 1$ and $s/\Lambda^2 < 1$ where v is the Electro-Weak Symmetry Breaking scale in the SM and $\sqrt{s}$ is the experimental measurement scale [5].

Existence of NP at higher energy scales which can modify interactions seen at hadron colliders includes an interesting operator in the dimension-six gluon sector. Using gluon field strengths given by $G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g_s f_{abc} G_b^\mu G_c^\nu$, one can build a gauge invariant CPT protecting operator [6]:

$$\mathcal{O}_G = f_{abc} G_\mu^{a\nu} G_\nu^{b\rho} G_\rho^{c\mu}. \tag{1.1}$$

Top quark pair production serves as a clean and sensitive probe for $\mathcal{O}_G$ [7]. Significant impact at LHC energies can be observed even for small values of its Wilson coefficient. It dominates over

other quark-involving operators with comparable coefficients. Also, contamination from higher dimension gluon terms (dimension-8 and above) is small.

The parton distribution function of $t\bar{t}$ production has highest contribution from di-gluon initial states. So, in this thesis, we have approximated $pp \to t\bar{t}$ with $gg \to t\bar{t}$. The unfolded values of differential cross-section at the parton level for kinematic variables in the $t\bar{t}$ process provided by CMS [1] collaboration enable us to perform all our analysis using parton level information.

Several important topics for this thesis are briefly described in the following sections. In chapter 2, precision measurements in high energy $t\bar{t}$ observables are performed to put constraints on C_G, the Wilson coefficient of $\mathcal{O}_G$. We call it Cut and Count analysis. We calculate Observed, Expected and Linear (D6) bounds on C_G at the luminosity of CMS data, $\mathscr{L}_{exp} = 35.8\,\text{fb}^{-1}$. Assuming scaling of only the experimental uncertainties at high luminosities, we estimate high luminosity bounds.

In chapter 3, we try to improve bounds on C_G using Machine Learning. Convolutional Neural Networks (CNNs) are used as a classifier to distinguish signal against background. We perform event-by-event classification in section 3.1, once with NP generated at a fixed energy scale and then for three different energy scales with each sample as a separate classification category. Finally, we describe a novel, physically intuitive method of bin-based classification in section 3.2.

In this thesis, I will be using Natural Units, where we take Planck's constant and speed of light to be unity, $h = c = 1$. This simplifies the dimensions of relevant physical quantities in terms of mass or energy dimensions (refer to Table 1.1).

Quantity	**Dimensions in SI units**	**Dimensions in Natural Units**	**Formula**
Mass	kg	E	$E = mc^2$
Length	m	E^{-1}	$E = hc/\lambda$
Time	s	E^{-1}	$c = 1$, $[L] = [E]^{-1}$
Energy	$kg.m^2/s^2$	E	–
Momentum	$kg.m^2/s$	E	$E = pc$

Table 1.1: Few physical quantities in SI and Natural Units.

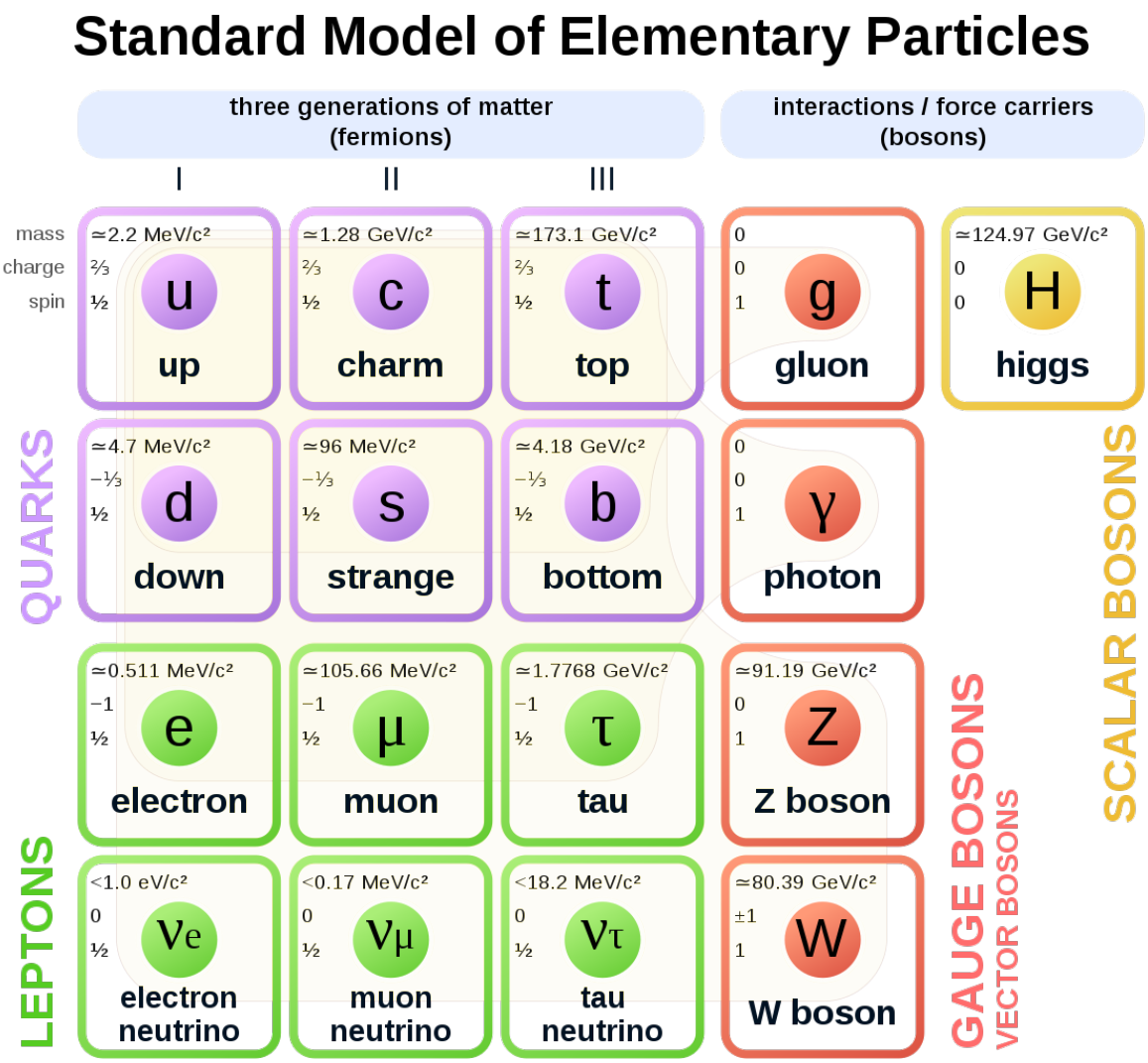


Figure 1.1: Standard Model of elementary particles [2].

1.1 Standard Model Particles

In order to develop an intuition for SM particles' behaviour, we look at their mass, full decay width and mean lifetime values. Stability of the particle can be inferred from lifetime and decay width. The relation between them is given by

$$\tau = \frac{\hbar}{\Gamma}. \tag{1.2}$$

Planck's constant is given by $\hbar \approx 6.58 \times 10^{-16}$ eV.s. The experimental values for the following tables are taken from PDG book, 2018 edition [8].

Quark	**Mass**	**Quark**	**Mass**
u	$2.2^{+0.5}_{-0.4}$ MeV	c	$1.275^{+0.025}_{-0.035}$ GeV
d	$4.7^{+0.5}_{-0.3}$ MeV	b	$4.18^{+0.04}_{-0.03}$ GeV
s	95^{+9}_{-3} MeV	t	173.1 ± 0.9 GeV(Pole mass)

Table 1.2: Quark masses with experimental uncertainties.

Boson	**Mass**	**Full Decay Width (Γ)**
W	80.379 ± 0.012 GeV	2.085 ± 0.042 GeV
Z	91.1876 ± 0.0021 GeV	2.4952 ± 0.0023 GeV
γ	$m_\gamma < 1 \times 10^{-18}$ eV	–
H^0	125.18 ± 0.16 GeV	$\Gamma < 0.013$ GeV(CL= 95%)

Table 1.3: Boson masses and their full decay width (Γ) with experimental uncertainties.

Lepton	Mass	Mean Lifetime (τ)
e	$0.5109989461 \pm 0.0000000031$ MeV	$\tau > 6.6 \times 10^{28}$ yr (CL= 90%)
μ	$105.6583745 \pm 0.0000024$ MeV	$(2.1969811 \pm 0.0000022) \times 10^{-6}$ s
τ	1776.86 ± 0.12 MeV	$(290.3 \pm 0.5) \times 10^{-15}$ s
ν	$m_\nu < 2$ eV(tritium decay)	$\tau/m > 15.4$ s/eV(CL= 90%, accelerator)

Table 1.4: Lepton masses and their lifetimes (τ) with experimental uncertainties.

1.2 Large Hadron Collider

The Large Hadron Collider or LHC is the world's highest energy collider. This superconducting collider contains two proton beams in separate pipes throughout most of the circumference and brings them together in a single pipe at particular interaction points (IP). The beam pipes are kept at ultrahigh vacuum. Beams contain protons in bunches with short bunch spacing. At each IP, approximately 30 head-on collisions happen through 120 m of common pipe [8]. There are 1.2×10^{11} protons per bunch and 2808 bunches per proton beam.

The number of events produced at the LHC are given by

$$N_{exp} = \sigma_{exp} \times \int \mathscr{L}(t) dt \,, \tag{1.3}$$

where N_{exp} : number of events, σ_{exp} : cross-section of interest and $\mathscr{L}(t)$: instantaneous luminosity.

Around the accelerator ring, there are four particle detectors − ATLAS, CMS, ALICE and LHCb where the beams are guided to collide.

1.2.1 CMS

CMS (formerly Compact Muon Solenoid) is a general-purpose detector in the LHC. The detector has a large superconducting solenoid, 6m in diameter, which generates very high magnetic field $\sim$3.8 T. Within the volume of solenoid, silicon tracker, electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL) are present. The tracker is made up of silicon pixels and strips. It detects charged particle trajectories. ECAL is constituted of lead tungstate crystal while HCAL

is made of brass and scintillator. Forward endcaps extend pseudorapidity coverage. Muons are detected outside the solenoid in the gas-ionization chambers inside the steel magnetic flux-return yoke. A two-tier trigger system based on hardware and software controls does event selection.

1.2.2 ATLAS

ATLAS is another general-purpose detector in LHC. It consists of an inner detector (ID), electromagnetic (EM) and hadronic calorimeters and a muon spectrometer (MS) [9, 10]. ID is located within 2T axial magnetic field. It can measure charged particle momenta with $-\pi \leq \phi \leq \pi$ and $|\eta| < 2.5$. The calorimeter system covers pseudorapidity range $|\eta| < 4.9$. The EM calorimeters are high-granularity and consist of a barrel section ($|\eta| < 1.475$) and two endcap sections ($1.375 < |\eta| < 3.2$). Hadronic calorimeters are scintillator-tile barrel calorimeter ($|\eta| < 1.7$) and a copper liquid Argon endcap ($1.7 < |\eta| < 3.2$). The MS is capable of triggering and tests for muons. It is located inside a toroidal magnetic field. Two zero-degree calorimeters (ZDCs) are present to detect neutral particles with $|\eta| > 8.5$. Three-tier trigger system selects required events. L1, L2 and EF (event filter) are based on hardware, online and offline reconstruction algorithms respectively.

1.3 Kinematics

At LHC, a $2E$=13 TeV center of mass energy proton-proton collision has incoming protons in z-direction with respective momenta $P_1^\mu = (E,0,0,E)$ and $P_2^\mu = (E,0,0,-E)$. A proton is constituted mostly by soft gluons and a few hard gluons, quarks and anti-quarks. On $pp-$collision, we get a lot of soft gluon scattering which give minimum bias events. We are usually interested in hard $qg-$scattering that might give detectable final particles transverse to the beam direction.

In a hard scatter, for the partons that collide, we can write their momenta in terms of proton momenta as $p_1^\mu = x_1 P_1^\mu$ and $p_2^\mu = x_2 P_2^\mu$. Here, x_i gives the fraction of i^{th} proton's momentum carried by its parton. According to the parton model, each x_i is probabilistic and independent distribution.

The angular information in an event can be captured using azimuthal angle ϕ ranging from 0 to 2π transverse to the beam line and polar angle θ ranging from 0 to π with the beam line. We define a

kinematic quantity called rapidity (y)

$$y = \frac{1}{2}\log\frac{E+p_z}{E-p_z}.$$

Difference in rapidities, $\Delta y = y_1 - y_2$ is boost-invariant while y, by itself, is not. Hence, angular separation is defined as $R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$ (boost-invariant). For massless particles, $E = |\vec{p}|$ which gives

$$y = \frac{1}{2}\log\frac{E+p_z}{E-p_z} = \frac{1}{2}\log\frac{1+\cos\theta}{1-\cos\theta} = \frac{1}{2}\log\frac{2\cos^2\theta/2}{2\sin^2\theta/2} = \frac{1}{2}\log\cot\frac{\theta}{2} \qquad \text{(massless particles)}.$$

This motivates the definition of pseudorapidity η,

$$\eta = \frac{1}{2}\log\cot\frac{\theta}{2}. \tag{1.4}$$

η a geometric quantity and is only boost-invariant for massless particles.

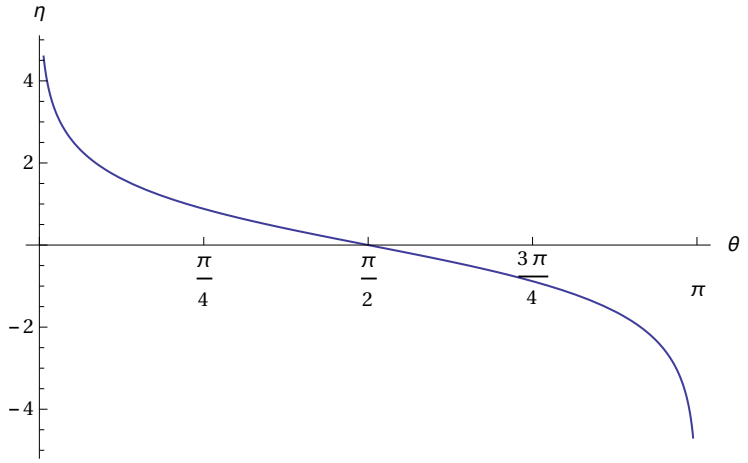


Figure 1.2: This plot shows pseudorapidity η as a function of polar angle θ.

For a boost along z-direction, we have transverse momenta p_x and p_y. We use these to define boost-invariant quantities $\vec{p}_{\mathrm{T}} \equiv (p_x, p_y)$ and the scalar $p_{\mathrm{T}} \equiv |\vec{p}_{\mathrm{T}}|$. Azimuthal angle ϕ can also be defined as $\phi = \tan^{-1}(p_x/p_y)$ which is also boost-invariant.

1.4 Standard Model

The Lagrangian in the Standard Model (SM), obeying $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetries, is given by

$$\begin{aligned}\mathscr{L}_{\rm SM} = &-\frac{1}{4}G^A_{\mu\nu}G^{A\mu\nu} - \frac{1}{4}W^I_{\mu\nu}W^{I\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + (D_\mu H)^\dagger (D^\mu H) + \mu^2 H^\dagger H - \lambda (H^\dagger H)^2 \\ &+ i\left(\overline{L} D_\mu \gamma^\mu L + \overline{e}_R D_\mu \gamma^\mu e_R + \overline{Q} D_\mu \gamma^\mu Q + \overline{u}_R D_\mu \gamma^\mu u_R + \overline{d}_R D_\mu \gamma^\mu d_R\right) \\ &- \left(\overline{L} Y_e e_R H + \overline{Q} Y_u u_R \widetilde{H} + \overline{Q} Y_d d_R H + \text{h.c.}\right), \end{aligned} \tag{1.5}$$

where $\widetilde{H} = i\sigma_2 H^*$, and the gauge-field strengths are defined as

$$G^A_{\mu\nu} = \partial_\mu G^A_\nu - \partial_\nu G^A_\mu + g_s f^{ABC} G^B_\mu G^C_\nu , \tag{1.6}$$

$$W^I_{\mu\nu} = \partial_\mu W^I_\nu - \partial_\nu W^I_\mu + g\varepsilon^{IJK} W^J_\mu W^K_\nu , \tag{1.7}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu , \tag{1.8}$$

where f^{ABC} and ε^{IJK} are the structure constants of $SU(3)$ and $SU(2)$: $[T^A, T^B] = if^{ABC}T^C$ with $f^{ABC} = i(T^A)^{BC}$ for $SU(3)$, and $[\sigma^I/2, \sigma^J/2] = i\varepsilon^{JIK}\sigma^K/2$ with ε^{JIK} being the totally antisymmetric tensor. We define the covariant derivative as

$$D_\mu q_L = (\partial_\mu - ig_s T^A G^A_\mu - igS^I W^I_\mu - ig' Y_q B_\mu) q_L \tag{1.9}$$

for the left-handed quark, where $T^A = \lambda^A/2$, $S^I = \sigma^I/2$, and the hyper charge is normalized as $Y_q = 1/6$. The gauge quantum numbers of each field are summarized in Table 1.5.

The electroweak (EW) symmetry is spontaneously broken as the Higgs field acquires the non-zero vacuum expectation value (VEV), $v = \mu/\sqrt{\lambda}$. In unitary gauge, the Higgs field H is written in terms of v and the physical Higgs field h:

$$H = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 + \dfrac{h}{v} \end{pmatrix} . \tag{1.10}$$

For a general 2x2 matrix between $H^\dagger$ and H,

$$H^\dagger \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & (v+h)^\dagger \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} ,$$

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
G_μ^A	**8**	**1**	0
W_μ^I	**1**	**3**	0
B_μ	**1**	**1**	0
$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	**1**	**2**	$-1/2$
e_R	**1**	**1**	-1
$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	**3**	**2**	$1/6$
u_R	**3**	**1**	$2/3$
d_R	**3**	**1**	$-1/3$
$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}$	**1**	**2**	$1/2$

Table 1.5: Gauge quantum numbers of the SM fields.

$$= \frac{1}{2}|v+h|^2 a_{22}. \tag{1.11}$$

The Higgs kinetic term then yield the mass terms of the gauge bosons as well as hVV and $hhVV$ interactions:

$$\begin{aligned}
(D_\mu H)^\dagger (D^\mu H) &= \left(\partial_\mu H - ig\frac{\sigma^I}{2}W_\mu^I H - ig'\frac{1}{2}B_\mu H\right)^\dagger \left(\partial^\mu H - ig\frac{\sigma^J}{2}W^{J\mu} H - ig'\frac{1}{2}B^\mu H\right), \\
&= (\partial_\mu H)^\dagger(\partial^\mu H) + \frac{g^2}{4}H^\dagger W_\mu^{I\dagger}\sigma^I\sigma^J W^{J\mu}H + \frac{g'^2}{4}H^\dagger B_\mu^\dagger B^\mu H \\
&\quad + \left[i\frac{g}{2}H^\dagger W_\mu^{I\dagger}\sigma^I(\partial^\mu H) + i\frac{g'}{2}H^\dagger B_\mu^\dagger(\partial^\mu H) + \frac{gg'}{4}H^\dagger W_\mu^{I\dagger}\sigma^I B^\mu H + \text{h.c.}\right], \\
&= \frac{1}{2}(\partial_\mu h)^\dagger(\partial^\mu h) + \frac{g^2}{8}|v+h|^2 W_\mu^{I\dagger}(\sigma^I\sigma^J)_{22}W^{J\mu} + \frac{g'^2}{8}|v+h|^2 B_\mu^\dagger B^\mu \\
&\quad + \left[i\frac{g}{4}(v+h)^\dagger W_\mu^{I\dagger}(\sigma^I)_{22}(\partial^\mu h) + i\frac{g'}{4}(v+h)^\dagger B_\mu^\dagger(\partial^\mu h)\right. \\
&\quad \left. + \frac{gg'}{8}|v+h|^2 W_\mu^{I\dagger}(\sigma^I)_{22}B^\mu + \text{h.c.}\right],
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}(\partial_\mu h)^\dagger(\partial^\mu h) + \frac{g^2}{8}|v+h|^2\big(W_\mu^{1\dagger}W^{1\mu} + W_\mu^{2\dagger}W^{2\mu} + W_\mu^{3\dagger}W^{3\mu} - iW_\mu^{1\dagger}W^{2\mu} \\
&\quad + iW_\mu^{2\dagger}W^{1\mu}\big) + \frac{1}{8}|v+h|^2\big(g'^2B_\mu^\dagger B^\mu - gg'B_\mu^\dagger W^{3\mu} - gg'W_\mu^{3\dagger}B^\mu\big) \\
&\quad - \left[\frac{i}{4}(v+h)^\dagger\big(gW_\mu^{3\dagger} - g'B_\mu^\dagger\big)(\partial^\mu h) + \text{h.c.}\right], \\
&= \frac{1}{2}(\partial_\mu h)^\dagger(\partial^\mu h) + \frac{g^2}{8}|v+h|^2\big(W_\mu^1 - iW_\mu^2\big)^\dagger\big(W_\mu^1 - iW^{2\mu}\big) \\
&\quad + \frac{1}{8}|v+h|^2\big(gW_\mu^3 - g'B_\mu\big)^\dagger\big(gW^{3\mu} - g'B^\mu\big) \\
&\quad - \left[\frac{i}{4}(v+h)^\dagger\big(gW_\mu^3 - g'B_\mu\big)^\dagger(\partial^\mu h) + \text{h.c.}\right], \\
&= \frac{1}{2}(\partial_\mu h)^\dagger(\partial^\mu h) + \frac{g^2v^2}{4}\left(1 + \frac{2h}{v} + \frac{h^2}{v^2}\right)W_\mu^{+\dagger}W^{+\mu} \\
&\quad + \frac{1}{2}\frac{(g^2+g'^2)v^2}{4}\left(1 + \frac{2h}{v} + \frac{h^2}{v^2}\right)Z_\mu^\dagger Z^\mu \\
&\quad \underbrace{- \left[\frac{i(g^2+g'^2)^{1/2}v}{4}\left(1 + \frac{h^\dagger}{v}\right)Z_\mu^\dagger(\partial^\mu h) + \text{h.c.}\right]}_{\text{vanishes because } Z_\mu^\dagger = Z_\mu}, \\
&= \frac{1}{2}(\partial_\mu h)^\dagger(\partial^\mu h) + M_W^2 W_\mu^- W^{+\mu} + \frac{1}{2}M_Z^2 Z_\mu Z^\mu \\
&\quad + \frac{g^2v^2}{4}\left(\frac{2h}{v} + \frac{h^2}{v^2}\right)W_\mu^- W^{+\mu} + \frac{1}{2}\frac{(g^2+g'^2)v^2}{4}\left(\frac{2h}{v} + \frac{h^2}{v^2}\right)Z_\mu Z^\mu\,, \qquad (1.12)
\end{aligned}
$$

where the physical gauge-boson fields are defined as

$$
W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)\,, \qquad (1.13)
$$

$$
\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \qquad (1.14)
$$

with the weak mixing angle θ_W defining through the relations,

$$
c_W \equiv \cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}\,, \qquad (1.15)
$$

$$
s_W \equiv \sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}}\,, \qquad (1.16)
$$

$$
e^2 = g^2\sin^2\theta_W = g^2 s_W^2\,. \qquad (1.17)
$$

For the electric charge, we adopt the sign convention $e = g s_W$. In the lagrangian density, mass term for real vector field Z is $\frac{1}{2}m^2 Z_\mu Z^\mu$ and that for complex vector field W is $m^2 W^*_\mu W^\mu$. The masses of the W and Z bosons are then given by

$$M_W^2 = \frac{g^2 v^2}{4}, \qquad M_Z^2 = \frac{(g^2 + g'^2)v^2}{4}, \tag{1.18}$$

which imply $c_W^2 = M_W^2/M_Z^2$. The Higgs VEV v can be estimated as

$$v = \frac{1}{(\sqrt{2}G_F)^{1/2}} \approx 246 \text{ GeV}, \tag{1.19}$$

where $G_F = \sqrt{2}g^2/(8M_W^2)$.

The kinetic terms of the gauge bosons involve the triple and quartic gauge-boson couplings:[1]

$$\begin{aligned}
-\frac{1}{4}G^A_{\mu\nu}G^{A\mu\nu} &= -\frac{1}{4}(\partial_\mu G^A_\nu - \partial_\nu G^A_\mu + g_s f^{ABC} G^B_\mu G^C_\nu)(\partial^\mu G^{A\nu} - \partial^\nu G^{A\mu} + g_s f^{ADE} G^{D\mu} G^{E\nu}), \\
&\supset -\frac{g_s}{2} f^{ABC}(\partial_\mu G^A_\nu - \partial_\nu G^A_\mu) G^{B\mu} G^{C\nu} - \frac{g_s^2}{4} f^{ABC} f^{ADE} G^B_\mu G^C_\nu G^{D\mu} G^{E\nu}, \\
&= -g_s f^{ABC}(\partial_\mu G^A_\nu) G^{B\mu} G^{C\nu} - \frac{g_s^2}{4} f^{ABC} f^{ADE} G^B_\mu G^C_\nu G^{D\mu} G^{E\nu}, \qquad (1.20) \\
-\frac{1}{4}W^I_{\mu\nu}W^{I\mu\nu} &= -\frac{1}{4}(\partial_\mu W^I_\nu - \partial_\nu W^I_\mu + g\varepsilon^{IJK} W^J_\mu W^K_\nu)(\partial^\mu W^{I\nu} - \partial^\nu W^{I\mu} + g\varepsilon^{ILM} W^{L\mu} W^{M\nu}), \\
&= -\frac{1}{4}(\partial_\mu W^I_\nu - \partial_\nu W^I_\mu)(\partial^\mu W^{I\nu} - \partial^\nu W^{I\mu}) \\
&\quad - \frac{g}{2}\varepsilon^{IJK}(\partial_\mu W^I_\nu - \partial_\nu W^I_\mu) W^{J\mu} W^{K\nu} - \frac{g^2}{4}\varepsilon^{IJK}\varepsilon^{ILM} W^J_\mu W^K_\nu W^{L\mu} W^{M\nu}, \\
&= -\frac{1}{4}(\partial_\mu W^I_\nu - \partial_\nu W^I_\mu)(\partial^\mu W^{I\nu} - \partial^\nu W^{I\mu}) \\
&\quad - g(\partial_\mu W^1_\nu - \partial_\nu W^1_\mu) W^{2\mu} W^{3\nu} - g(\partial_\mu W^2_\nu - \partial_\nu W^2_\mu) W^{3\mu} W^{1\nu} \\
&\quad - g(\partial_\mu W^3_\nu - \partial_\nu W^3_\mu) W^{1\mu} W^{2\nu} - \frac{g^2}{4}(\delta^{JL}\delta^{KM} - \delta^{JM}\delta^{KL}) W^J_\mu W^K_\nu W^{L\mu} W^{M\nu}, \\
&= -\frac{1}{2}(\partial_\mu W^+_\nu - \partial_\nu W^+_\mu)(\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}) - \frac{c_W^2}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu Z^\nu - \partial^\nu Z^\mu) \\
&\quad - \frac{s_W c_W}{2}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) - \frac{s_W^2}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&\quad + ig(\partial_\mu W^+_\nu - \partial_\nu W^+_\mu) W^{-\mu} W^{3\nu} - ig(\partial_\mu W^-_\nu - \partial_\nu W^-_\mu) W^{+\mu} W^{3\nu}
\end{aligned}$$

[1] Here we employ $W^1_\mu = (W^+_\mu + W^-_\mu)/\sqrt{2}$, $W^2_\mu = i(W^+_\mu - W^-_\mu)/\sqrt{2}$, $W^3_\mu = c_W Z_\mu + s_W A_\mu$, and $W^J_\mu W^J_\nu = W^+_\mu W^-_\nu + W^-_\mu W^+_\nu + c_W^2 Z_\mu Z_\nu + s_W c_W (Z_\mu A_\nu + A_\mu Z_\nu) + s_W^2 A_\mu A_\nu$.

$$
\begin{aligned}
&+ ig(\partial_\mu W^3_\nu - \partial_\nu W^3_\mu)W^{+\mu}W^{-\nu} - \frac{g^2}{4}W^J_\mu W^{J\mu}W^K_\nu W^{K\nu} + \frac{g^2}{4}W^J_\mu W^J_\nu W^{K\mu}W^{K\nu},\\
&= -\frac{1}{2}(\partial_\mu W^+_\nu - \partial_\nu W^+_\mu)(\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}) - \frac{c^2_W}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu Z^\nu - \partial^\nu Z^\mu)\\
&- \frac{s_W c_W}{2}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) - \frac{s^2_W}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu)\\
&+ i\sum_{V=Z,\gamma} g_{WWV}\Big[(\partial_\mu W^+_\nu - \partial_\nu W^+_\mu)W^{-\mu}V^\nu - (\partial_\mu W^-_\nu - \partial_\nu W^-_\mu)W^{+\mu}V^\nu\\
&\qquad + (\partial_\mu V_\nu - \partial_\nu V_\mu)W^{+\mu}W^{-\nu}\Big]\\
&+ \frac{g^2}{2}\left(W^+_\mu W^{+\mu}W^-_\nu W^{-\nu} - W^+_\mu W^{-\mu}W^+_\nu W^{-\nu}\right)\\
&+ g^2 c^2_W\left(W^+_\mu Z^\mu W^-_\nu Z^\nu - W^+_\mu W^{-\mu}Z_\nu Z^\nu\right) + g^2 s^2_W\left(W^+_\mu A^\mu W^-_\nu A^\nu - W^+_\mu W^{-\mu}A_\nu A^\nu\right)\\
&+ g^2 s_W c_W\left(W^+_\mu Z^\mu W^-_\nu A^\nu + W^+_\mu A^\mu W^-_\nu Z^\nu - 2W^+_\mu W^{-\mu}Z_\nu A^\nu\right), \qquad (1.21)\\
-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} &= -\frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)(\partial^\mu B^\nu - \partial^\nu B^\mu),\\
&= -\frac{s^2_W}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu Z^\nu - \partial^\nu Z^\mu) + \frac{s_W c_W}{2}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu)\\
&- \frac{c^2_W}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu), \qquad (1.22)
\end{aligned}
$$

where the triple gauge boson couplings are written as $g_{WWZ} = gc_W = ec_W/s_W$ and $g_{WW\gamma} = gs_W = e$.

Moreover, the Yukawa terms read

$$
\begin{aligned}
\mathscr{L}_{\text{Yukawa}} &= -\left(\overline{L}Y_e e_R H + \overline{Q}Y_u u_R \widetilde{H} + \overline{Q}Y_d d_R H\right) + \text{h.c.},\\
&= -\frac{v+h}{\sqrt{2}}\left(\bar{e}_L Y_e e_R + \bar{u}_L Y_u u_R + \bar{d}_L Y_d d_R\right) + \text{h.c.} \qquad (1.23)
\end{aligned}
$$

The interactions between the gauge bosons and the left-handed fermions are then given by

$$
\begin{aligned}
\bar{f}_L i\not{D} f_L \supset \bar{f}_L\gamma^\mu &\left(g_s T^A G^A_\mu + g I^I_f W^I_\mu + g' Y_{f_L} B_\mu\right) f_L,\\
&= \bar{f}_L\gamma^\mu\left\{g_s T^A G^A_\mu + \frac{e}{\sqrt{2}s_W}\left[(I^1_f + iI^2_f)W^+_\mu + (I^1_f - iI^2_f)W^-_\mu\right]\right.\\
&\qquad \left. + \frac{e}{s_W c_W}(I^3_f - Q_f s^2_W)Z_\mu + eQ_f A_\mu\right\} f_L, \qquad (1.24)
\end{aligned}
$$

where $I^I_f = \sigma^I/2$ and $Q_f = I^3_f + Y_{f_L}$. Similarly, the right-handed fermions have the following

interactions:

$$\begin{aligned}\bar{f}_R\, i\not{D} f_R &\supset \bar{f}_R \gamma^\mu \left(g_s T^A G^A_\mu + g' Y_{f_R} B_\mu\right) f_R\,, \\ &= \bar{f}_R \gamma^\mu \left[g_s T^A G^A_\mu + \frac{e}{s_W c_W}(-Q_f s_W^2) Z_\mu + e Q_f A_\mu\right] f_R\,. \end{aligned} \tag{1.25}$$

We define the effective couplings g_V^f and g_A^f for the interaction of the Z boson with the neutral current as

$$\mathscr{L}_{\text{eff}}^{Zf\bar{f}} = \frac{e}{2 s_W c_W} Z_\mu \sum_f \bar{f} \gamma^\mu \left(g_V^f - g_A^f \gamma_5\right) f = \frac{e}{s_W c_W} Z_\mu \sum_f \bar{f} \gamma_\mu \left(g_L^f P_L + g_R^f P_R\right) f\,, \tag{1.26}$$

where $P_L = (1-\gamma_5)/2$ and $P_R = (1+\gamma_5)/2$, and the tree-level couplings are given by $g_V^f = I_f^3 - 2Q_f s_W^2$, $g_A^f = I_f^3$, $g_L^f = I_f^3 - Q_f s_W^2$ and $g_R^f = -Q_f s_W^2$ with $g_L^f = (g_V^f + g_A^f)/2$ and $g_R^f = (g_V^f - g_A^f)/2$.

Transformation of $\widetilde{H}$ under SU(2)

We want to see how $\widetilde{H}$ transforms when $H \to UH$ where U is an order 2 unitary matrix. We had defined $\widetilde{H} = i\sigma_2 H^*$ which transforms as

$$\widetilde{H} = i\sigma_2 H^* \to i\sigma_2 (UH)^* = i\sigma_2 U^* H^*\,. \tag{1.27}$$

For an infinitesimal transformation,

$$\begin{aligned} U = e^{i\vec{\sigma}.\vec{\theta}/2} &= \mathbb{1}_{2\text{x}2} + \frac{i\vec{\sigma}.\vec{\theta}}{2} + \left(\frac{i\vec{\sigma}.\vec{\theta}}{2}\right)^2 + \ldots \approx \mathbb{1} + \frac{i\vec{\sigma}.\vec{\theta}}{2}\,, \\ U^* = \mathbb{1} - \frac{i}{2}\sigma_i^* \theta_i &= \mathbb{1} - \frac{i}{2}(\sigma_1\theta_1 - \sigma_2\theta_2 + \sigma_3\theta_3)\,, \end{aligned}$$

where $\sigma_1^* = \sigma_1, \sigma_3^* = \sigma_3$ but $\sigma_2^* = -\sigma_2$. Therefore, equation (1.27) becomes

$$\begin{aligned} \widetilde{H} &\to i\sigma_2 \left(\mathbb{1} - \frac{i}{2}(\sigma_1\theta_1 - \sigma_2\theta_2 + \sigma_3\theta_3)\right) H^*\,, \\ &= i\sigma_2 H^* - i\sigma_2 \frac{i}{2}(\sigma_1\theta_1 - \sigma_2\theta_2 + \sigma_3\theta_3) H^*\,, \\ &= \widetilde{H} - \frac{i}{2}(i\sigma_2\sigma_1\theta_1 - i\sigma_2^2\theta_2 - i\sigma_2\sigma_3\theta_3) H^*\,, \\ &= \widetilde{H} - \frac{i}{2}(\sigma_3\theta_1 - i\theta_2 - \sigma_1\theta_3) H^*\,, \qquad\qquad (\text{using } \ (\sigma_i)^2 = \mathbb{1} \ \ \& \ \sigma_i\sigma_j = i\varepsilon_{ijk}\sigma_k) \end{aligned}$$

$$
\begin{aligned}
&= \widetilde{H} - \frac{1}{2}(\sigma_3\theta_1 - i\theta_2 - \sigma_1\theta_3) i(\sigma_2\sigma_2) H^*, && \text{(introducing } \sigma_2^2 = \mathbb{1}) \\
&= \widetilde{H} - \frac{1}{2}(\sigma_3.\sigma_2\theta_1 - i\sigma_2\theta_2 - \sigma_1.\sigma_2\theta_3)\widetilde{H}, \\
&= \widetilde{H} - \frac{1}{2}(-i\sigma_1\theta_1 - i\sigma_2\theta_2 - i\sigma_3\theta_3)\widetilde{H}, \\
&= \widetilde{H} + \frac{i}{2}\sigma_i\theta_i\widetilde{H}, \\
&= U\widetilde{H}. && (1.28)
\end{aligned}
$$

Thus equation (1.28) tells us that $\widetilde{H}$ transforms similar to H under SU(2) symmetry group. The need for $\widetilde{H}$ arises because of hypercharge and not SU(2) symmetry.

1.4.1 CKM matrix

In the Standard Model lagrangian, $\mathscr{L}_{\text{SM}}$, (eq. 1.5) we can observe that the yukawa matrices are not necessarily diagonal. No invariance condition on $\mathscr{L}_{\text{SM}}$ imposes diagonal form on them. So, we can assume yukawa matrices to be of general complex form, not necessarily diagonal. We can diagonalize them so that the mass terms may be read out directly from the lagrangian.

We know that any hermitian operator can be diagonalized using a unitary operator. But a general complex matrix can be diagonalized using bi-unitary transformation.

We diagonalize the yukawa matrix for the leptonic part of the lagrangian by doing a basis transformation.

$$
\mathscr{L}_{\text{lep}} \supset \overline{L}^i Y_e^{ij} e_R^j H + i\overline{L}^i D_\mu \gamma^\mu L^i + \overline{e}_R^i D_\mu \gamma^\mu e_R^i . \tag{1.29}
$$

Let prime superscript denote initial basis. Operator transformations are given by:

$$
L' \to U_L L, \qquad e'_R \to U_R e_R .
$$

where U_L and U_R are unitary matrices. Now,

$$
\overline{L}' Y_e e'_R H \to \overline{L} U_L^\dagger Y_e U_R e_R H = \overline{L} Y_e^D e_R H .
$$

We can choose U_L and U_R such that $U_L^\dagger Y_e U_R = Y_e^D$ is diagonal. If we do the same transformation on other terms of the above lagrangian(eq. 1.29), then we see that they are still diagonal since

$U_L^\dagger U_L = U_R^\dagger U_R = 1$.

$$\begin{aligned} i\overline{L}' D_\mu \gamma^\mu L' &\to i\overline{L} U_L^\dagger D_\mu \gamma^\mu U_L L, \\ &= i\overline{L} U_L^\dagger U_L D_\mu \gamma^\mu L, \\ &= i\overline{L} D_\mu \gamma^\mu L. \end{aligned} \tag{1.30}$$

$$\begin{aligned} \overline{e}'_R D_\mu \gamma^\mu e'_R &\to \overline{e}_R U_R^\dagger D_\mu \gamma^\mu U_R e_R, \\ &= \overline{e}_R U_R^\dagger U_R D_\mu \gamma^\mu e_R, \\ &= \overline{e}_R D_\mu \gamma^\mu e_R. \end{aligned} \tag{1.31}$$

Thus, we see that all the terms of the leptonic lagrangian become diagonal under basis transformation.

$$\mathscr{L}_{\text{lep}} \supset \overline{L} Y_e^D e_R H + i\overline{L} D_\mu \gamma^\mu L + \overline{e}_R D_\mu \gamma^\mu e_R . \tag{1.32}$$

Since Y_e is now diagonal, we can directly get mass terms from the lagrangian,

$$\begin{pmatrix} \overline{L}_1 & \overline{L}_2 & \overline{L}_3 \end{pmatrix} \begin{pmatrix} Y_e & 0 & 0 \\ 0 & Y_\mu & 0 \\ 0 & 0 & Y_\tau \end{pmatrix} \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} H \to \text{gives mass terms}.$$

There are no off-diagonal terms indicating flavor mixing. Hence, SM does not allow any vertex that changes the flavor of lepton.

Now, we do a similar transformation for the quarks.

$$\mathscr{L}_{\text{q}} \supset i\big(\overline{Q} D_\mu \gamma^\mu Q + \overline{u}_R D_\mu \gamma^\mu u_R + \overline{d}_R D_\mu \gamma^\mu d_R\big) - \big(\overline{Q} Y_u u_R \widetilde{H} + \overline{Q} Y_d d_R H + \text{h.c.}\big) \tag{1.33}$$

$$\overline{Q}' Y_u u'_R \widetilde{H} = \overline{u}'_{L_i} Y_{u_{ij}} u'_{R_j} (v+h)/\sqrt{2} \tag{1.34}$$

$$\overline{Q}' Y_d d'_R H = \overline{d}'_{L_i} Y_{d_{ij}} d'_{R_j} (v+h)/\sqrt{2} \tag{1.35}$$

Again, let prime superscript denote initial basis. Operator transformations are given by:

$$\begin{aligned} u'_L &\to V_L^u u_L, \qquad & u'_R &\to V_R^u u_R \\ d'_L &\to V_L^d d_L, \qquad & d'_R &\to V_R^d d_R \end{aligned} \tag{1.36}$$

where V_L and V_R are unitary matrices. Now,

$$\begin{aligned}
\mathscr{L}_{\rm q} &\supset \overline{Q}' Y_u u'_R \widetilde{H} + \overline{Q}' Y_d d'_R H \,, \\
&= \frac{v+h}{\sqrt{2}} \big(\overline{u}'_{L_i} Y_{u_{ij}} u'_{R_j} + \overline{d}'_{L_i} Y_{d_{ij}} d'_{R_j} \big) \,, \\
&\to \frac{v+h}{\sqrt{2}} \big(\overline{u}_L (V_L^{u\dagger} Y_u V_R^u) u_R + \overline{d}_L (V_L^{d\dagger} Y_d V_R^d) d_R \big) \,, \\
&= \frac{v+h}{\sqrt{2}} \big(\overline{u}_L Y_u^D u_R + \overline{d}_L Y_d^D d_R \big) \,.
\end{aligned} \tag{1.37}$$

Thus, above terms get diagonalized under above basis transformation. Also, the kinetic terms for the right-handed quarks get diagonalized for the same transformation:

$$\begin{aligned}
\mathscr{L}_{\rm q} &\supset i \big(\overline{u}'_R D_\mu \gamma^\mu u'_R + \overline{d}'_R D_\mu \gamma^\mu d'_R \big) \,, \\
&\to i \big(\overline{u}_R V_R^{u\dagger} D_\mu \gamma^\mu V_R^u u_R + \overline{d}_R V_R^{d\dagger} D_\mu \gamma^\mu V_R^d d_R \big) \,, \\
&= i \big(\overline{u}_R (V_R^{u\dagger} V_R^u) D_\mu \gamma^\mu u_R + \overline{d}_R (V_R^{d\dagger} V_R^d) D_\mu \gamma^\mu d_R \big) \,, \\
&= i \big(\overline{u}_R D_\mu \gamma^\mu u_R + \overline{d}_R D_\mu \gamma^\mu d_R \big) \,.
\end{aligned} \tag{1.38}$$

However, if we focus on the kinetic term for the left-handed quarks,

$$\begin{aligned}
\mathscr{L}_{\rm q} &\supset i \overline{Q}' D_\mu \gamma^\mu Q' \\
&= \begin{pmatrix} \overline{u}'_L & \overline{d}'_L \end{pmatrix} \left(\partial_\mu \mathbb{1}_{2\text{x}2} + g \frac{\sigma^I}{2} W_\mu^I + g' Y B_\mu \mathbb{1}_{2\text{x}2} \right) \gamma^\mu \begin{pmatrix} u'_L \\ d'_L \end{pmatrix} \,.
\end{aligned} \tag{1.39}$$

In above expression, the terms involving $\mathbb{1}_{2\text{x}2}$ and σ^3 do not give any off-diagonal elements. So we write the terms with σ^1 and σ^2 explicitly while neglecting the constant factors:

$$\begin{aligned}
\begin{pmatrix} \overline{u}'_L & \overline{d}'_L \end{pmatrix} \left[\begin{pmatrix} 0 & W_\mu^1 \\ W_\mu^1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -iW_\mu^2 \\ iW_\mu^2 & 0 \end{pmatrix} \right] \gamma^\mu \begin{pmatrix} u'_L \\ d'_L \end{pmatrix} &= \overline{u}'_L (W_\mu^1 - iW_\mu^2) \gamma^\mu d'_L \\
&\quad + \overline{d}'_L (W_\mu^1 + iW_\mu^2) \gamma^\mu u'_L \,, \\
&= \overline{u}'_L W_\mu^+ \gamma^\mu d'_L + \overline{d}'_L W_\mu^- \gamma^\mu u'_L \,, \\
&= \overline{u}'_L W_\mu^+ \gamma^\mu d'_L + \text{h.c.} \,, \\
&\to \overline{u}_L V_L^{u\dagger} W_\mu^+ \gamma^\mu V_L^d d_L + \text{h.c.} \,, \\
&= \overline{u}_L W_\mu^+ \gamma^\mu (V_L^{u\dagger} V_L^d) d_L + \text{h.c.} \,, \\
&= \overline{u}_L W_\mu^+ \gamma^\mu V_{\rm CKM} d_L + \text{h.c.}
\end{aligned} \tag{1.40}$$

where V_{CKM} is a 3x3 unitary matrix called Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix. It is an off-diagonal matrix which gives flavor mixing of left-handed quarks.

$$V_{\text{CKM}} \equiv V_L^{u\dagger} V_L^d \,, \qquad\qquad (V_{\text{CKM}}^\dagger V_{\text{CKM}} = \mathbb{1}) \,. \tag{1.41}$$

The matrix V_{CKM} can be written as

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} . \tag{1.42}$$

In conclusion, $W^\pm$ can change flavors of left-handed quarks $-$ Flavor Changing Charged Current (FCCC) $-$ while right-handed quarks can't change flavors. It is the only source of flavor and generation mixing within the SM. W interactions violate parity. Also, regardless of chirality, a neutral current cannot change flavor of a quark. Flavor Changing Neutral Currents (FCNCs) are not present in SM at tree level.

1.5 Effective Field Theory

SM of particle physics has been quite successful as it agrees fairly well with experimental observations but it has its flaws. One of them is that it does not predict number of particles. Hence, at higher energy scales, one may discover new particles or interactions. We can construct an effective lagrangian keeping the symmetries/conservation laws in mind ($SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetries, baryon number conservation, etc.) without loss of generality [4]. New physics (NP) phenomena and particles can be added to the SM lagrangian in full generality using higher dimensional operators with dimension greater than or equal to four [11–13]. These effects may be too heavy for direct access at the LHC. The effective lagrangian can be written as

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_{\text{SM}} + \frac{1}{\Lambda} \sum_i C_i^{(5)} O_i^{(5)} + \frac{1}{\Lambda^2} \sum_i C_i^{(6)} O_i^{(6)} + \ldots \tag{1.43}$$

where $O_i^{(n)}$ denote dimension-n operators. The corresponding dimensionless coupling constants or Wilson coefficients are denoted by $C_i^{(n)}$. The operators are suppressed by powers of an unknown NP scale Λ.

The EFT constructed out of SM fields is called SMEFT. We can use it for BSM (Beyond SM) physics searches and to analyze deviations from SM. Considering a particular high-energy theory, all the Wilson coefficients can be determined by integrating out the heavy fields.

Reasons for using EFTs

Consider a scalar field ϕ which lies beyond NP scale Λ,

$$\begin{aligned}\mathscr{L} = \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}\lambda_2\phi^2 - \lambda_4\phi^4 - \lambda_6\phi^6 - ... \\ - \lambda_{4,2}\phi^2(\partial\phi)^2 ... - \lambda_{6,2}\phi^4(\partial\phi)^2 ... - \lambda_{4,4}(\partial\phi)^4 ...,\end{aligned} \tag{1.44}$$

where $\lambda_{n,m}$ stands for n fields and m derivatives present in the term. Above equation (eq. 1.44) can be rewritten using a super index I which counts all possible higher dimensional terms:

$$\mathscr{L} = \frac{1}{2}(\partial\phi)^2 + \sum_I \lambda_I O_I\,. \tag{1.45}$$

Let $dim\ (O_I) = \Delta_I$ and $\lambda_I = g_I/\Lambda^{\Delta_I - 4}$ where g_I is dimensionless and very small in order to keep perturbation theory valid at scale Λ ($g_I < O(1)$, neglecting orders of 4π).

From eq. 1.44, we can observe that $\lambda_2 = m^2 = g_2\Lambda^2$. If order of mass is the same as that of fundamental energy scale, *i.e.* $O(m) \approx O(\Lambda)$, then in relativistic theory, no different physics exists at that scale because there exists no quanta with energy much less than m to scatter.

Assuming the lagrangian is perturbative, we compute $2 \to 2$ scattering amplitude. At tree level, $\mathscr{A}_{2\to2}$ is dimensionless. Taking the term $\lambda_{4,2}\phi^2(\partial\phi)^2$ as an example, it clearly has dimension of $1/m^2$ and thus contributes a factor of E^2 in the amplitude. Other terms can be written similarly.

$$\mathscr{A}_{2\to2} = \mathscr{A}_{4-\text{legs}} = \lambda_4 + \lambda_{4,2}E^2 + \lambda_{4,4}E^4 + ... \tag{1.46}$$

For very small energies, the terms of $\mathscr{O}(E^2)$ and higher are negligible compared to the leading term. Now, if we look at $\mathscr{A}_{2\to4}$, the term with coefficient λ_4^2 has a propagator and it goes like

λ_4^2/E^2. Other terms behave similar to the way described earlier.

$$\begin{aligned}\mathscr{A}_{2\to 4} = \mathscr{A}_{6-\text{legs}} &= \lambda_4^2 \frac{1}{E^2} + \lambda_6 + \lambda_{6,2} E^2 + \dots, \\ &= \frac{1}{E^2}\left(\lambda_4^2 + \lambda_6 E^2 + \lambda_{6,2} E^4 + \dots\right). \end{aligned} \tag{1.47}$$

Again, if we go to low enough energy, only one parameter is dominating the amplitude.

A simple analogy is that a complicated structure looks like a simple point object at large distances and one cannot infer details about it without a very precise measurement. An experiment with finite precision will only give leading term(s).

All the operators with $\Delta_I \geq 4$, affect $\mathscr{A}_{n-\text{legs}}$ as

$$\mathscr{A}_{n-\text{legs}} = E^N \mathscr{P}(\lambda_I E^{\Delta_I - 4}), \tag{1.48}$$

where N depends on number of legs and $\mathscr{P}(\lambda_I E^{\Delta_I - 4})$ is some polynomial such that couplings have the right powers of E to make the input dimensionless. Hence, the coefficients of operators of higher dimensionality will have increasingly lower contribution to amplitude.

Equation 1.48 implies that as we go to lower energy scales, all operators with $\Delta_I > 4$ have lesser impact on the amplitude. Operators of dimension 4 remain unaffected by change of scale. Operators of dimensionality lower than mass are negligible at very large energy but become relevant as energy scale approaches that of mass.

Using only dimensional analysis, we can conclude that for a complicated lagrangian with all the terms, we only need terms with operator dimension less than or equal to 4 ($\Delta_I \leq 4$) to reliably compute the scattering amplitude at sufficiently low energies.

Advantages of using EFTs:

- EFTs deal with one energy scale at a time which makes it computationally easier to work with. We can evaluate different terms at their own energy scale instead of using messy complicated functions.

- It makes symmetries manifest. One can derive the consequences of manifest symmetry more easily.

- EFTs provide a unified framework to efficiently describe all new physics theories using

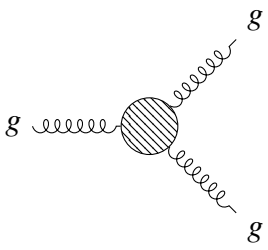
higher dimensional operators and their coefficients. Since all necessary symmetries are incorporated in the terms, finding bounds on Wilson coefficients meaningfully characterizes the new physics.

1.6 Higher Dimensional Operators

In order to incorporate a NP interaction at certain energy scale, a higher dimension EFT operator is added to the SM lagrangian density. As an example, we take a NP model for the process $gg \to t\bar{t}$ given by an EFT operator of dimension-six containing a new tri-gluon interaction vertex [11, 12, 14]. Let this triple gluon field strength operator be $\mathscr{O}_G$where

$$\mathscr{O}_G = f_{abc} G_\mu^{a\nu} G_\nu^{b\rho} G_\rho^{c\mu} . \tag{1.49}$$

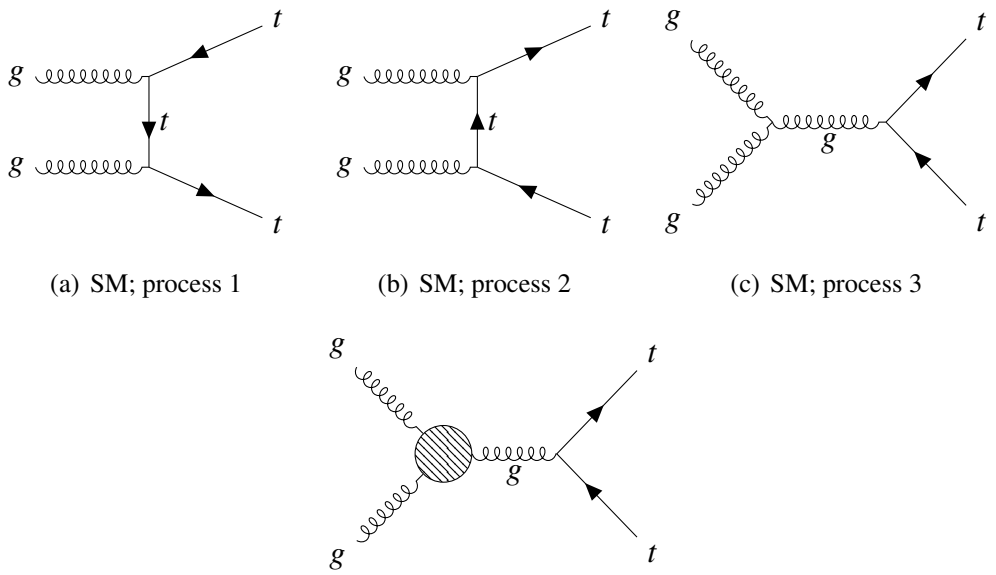
NP corresponds to the vertex:



The lagrangian density becomes

$$\mathscr{L}_{\text{eff}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_{\text{G}} = \mathscr{L}_{\text{SM}} + \frac{1}{\Lambda^2} C_G \mathscr{O}_G . \tag{1.50}$$

The corresponding processes are given by following four Feynman diagrams as shown in fig. 1.6. Three of these are SM processes and one is a NP process.



(a) SM; process 1 (b) SM; process 2 (c) SM; process 3

(d) NP

Figure 1.3: Feynman diagrams of all the possible vertices for the process $gg \to t\bar{t}$. Figures 1.3(a), 1.3(b) and 1.3(c) are tree level SM diagrams while 1.3(d) comes from NP interaction.

1.6.1 cross-section

For the process $gg \to t\bar{t}$, with proton center of mass (CM) energy $\sqrt{s} = 2E = 13$ TeV, let x_1 be fraction of proton momentum carried by gluon 1 and x_2 be that carried by gluon 2.

$$p_1^{\text{lab}} = x_1(E,0,0,E)\,, \tag{1.51}$$

$$p_2^{\text{lab}} = x_2(E,0,0,-E)\,, \tag{1.52}$$

$$\begin{aligned} \hat{s} &= \left((x_1+x_2)^2-(x_1-x_2)^2\right)E^2\,, \\ &= 4x_1x_2E^2\,, \end{aligned} \tag{1.53}$$

$$m_{t\bar{t}}^2 = (p_3+p_4)^2\,, \tag{1.54}$$

Here we know mass of top quark $m_t = 173.8$ GeV and $\hat{s},\hat{t},\hat{u}$ are Mandelstam variables. In the gluon CM frame, let $\hat{E}$ be the CM energy for gluons and θ be the angle between outgoing top quark (p_3) and collision axis (z). Let $E_1 = E_2 = \hat{E}$ and $E_3 = E_4 = E_{34}$.

$$p_1^{\text{CM}} = (\hat{E},0,0,\hat{E})\,, \tag{1.55}$$

$$p_2^{\rm CM} = (\hat{E}, 0, 0, -\hat{E})\,, \tag{1.56}$$

$$p_3^{\rm CM} = (E_{34}, \sqrt{E_{34}^2 - m_t^2}\,\sin\theta, 0, \sqrt{E_{34}^2 - m_t^2}\,\cos\theta)\,, \tag{1.57}$$

$$p_4^{\rm CM} = (E_{34}, -\sqrt{E_{34}^2 - m_t^2}\,\sin\theta, 0, -\sqrt{E_{34}^2 - m_t^2}\,\cos\theta)\,. \tag{1.58}$$

Above momenta give

$$\hat{s} = 4\hat{E}^2\,, \tag{1.59}$$

$$m_{t\bar{t}}^2 = 4E_{34}^2\,, \tag{1.60}$$

$$\hat{t} = m_t^2 - 2\hat{E}E_{34} + 2\hat{E}\sqrt{E_{34}^2 - m_t^2}\cos\theta\,, \tag{1.61}$$

$$\hat{u} = m_t^2 - 2\hat{E}E_{34} - 2\hat{E}\sqrt{E_{34}^2 - m_t^2}\cos\theta\,. \tag{1.62}$$

In the gluon CM frame, the 2-body phase space is given by

$$\begin{aligned}
\int d\phi_2(p_3, p_4) &= \int \frac{d^3p_3}{(2\pi)^3 2E_3}\frac{d^3p_4}{(2\pi)^3 2E_4}(2\pi)^4\delta(p_1^\mu + p_2^\mu - p_3^\mu - p_4^\mu)\,, \\
&= \int \frac{d^3p_3}{(2\pi)^3 2E_{34}}\frac{d^3p_4}{(2\pi)^3 2E_{34}}(2\pi)\delta(2\hat{E} - 2E_{34})(2\pi)^3\delta(\vec{p}_3 + \vec{p}_4)\,, \\
&= \int \frac{d^3p}{(2\pi)^3 4E_{34}^2}(2\pi)\delta(2\hat{E} - 2E_{34})\,, \\
&= \int \frac{p^2\; dp\; d\cos\theta\; d\phi}{(2\pi)^3 4E_{34}^2}(2\pi)\delta(2\hat{E} - 2E_{34})\,, \\
&= \int_0^{2\pi} d\phi \int \frac{p\; E_{34}\; dE_{34}\; d\cos\theta}{(2\pi)^3 4E_{34}^2}(2\pi)\delta(2\hat{E} - 2E_{34})\,, \\
&= (2\pi)\int \frac{\sqrt{E_{34}^2 - m_t^2}\; E_{34}\; dE_{34}\; d\cos\theta}{(2\pi)^2 4E_{34}^2}\delta(2\hat{E} - 2E_{34})\,,
\end{aligned} \tag{1.63}$$

where we have used $|\vec{p}_3| = |\vec{p}_4| = p$ and $d^3p = p^2\; dp\; d\cos\theta\; d\phi = pE_{34}\, dE_{34}\; d\cos\theta\; d\phi$.

Now,

$$\begin{aligned}
d\hat{\sigma} &= \frac{1}{2E_1 2E_2|\vec{v}_1 - \vec{v}_2|}|\mathscr{M}|^2 d\phi_2\,, \\
&= \frac{1}{8\hat{E}^2}|\mathscr{M}|^2 d\phi_2\,,
\end{aligned}$$

$$= \frac{1}{8\hat{E}^2}|\mathscr{M}|^2 \frac{\sqrt{E_{34}^2 - m_t^2}\, dE_{34}\, d\cos\theta}{(2\pi)4E_{34}} \delta(2\hat{E} - 2E_{34})\,. \tag{1.64}$$

Using equation (1.54 and 1.60),

$$d\hat{\sigma} = \frac{1}{2\hat{s}}|\mathscr{M}|^2 \frac{\sqrt{1-(2m_t/m_{t\bar{t}})^2}}{8\pi} \frac{d\cos\theta}{2} dm_{t\bar{t}}\, \delta(\sqrt{\hat{s}} - m_{t\bar{t}})\,, \tag{1.65}$$

$$\frac{d\hat{\sigma}}{dm_{t\bar{t}}} = \frac{1}{2\hat{s}} \frac{\sqrt{1-(2m_t/m_{t\bar{t}})^2}}{8\pi}\, \delta(\sqrt{\hat{s}} - m_{t\bar{t}}) \int |\mathscr{M}|^2 \frac{d\cos\theta}{2}\,. \tag{1.66}$$

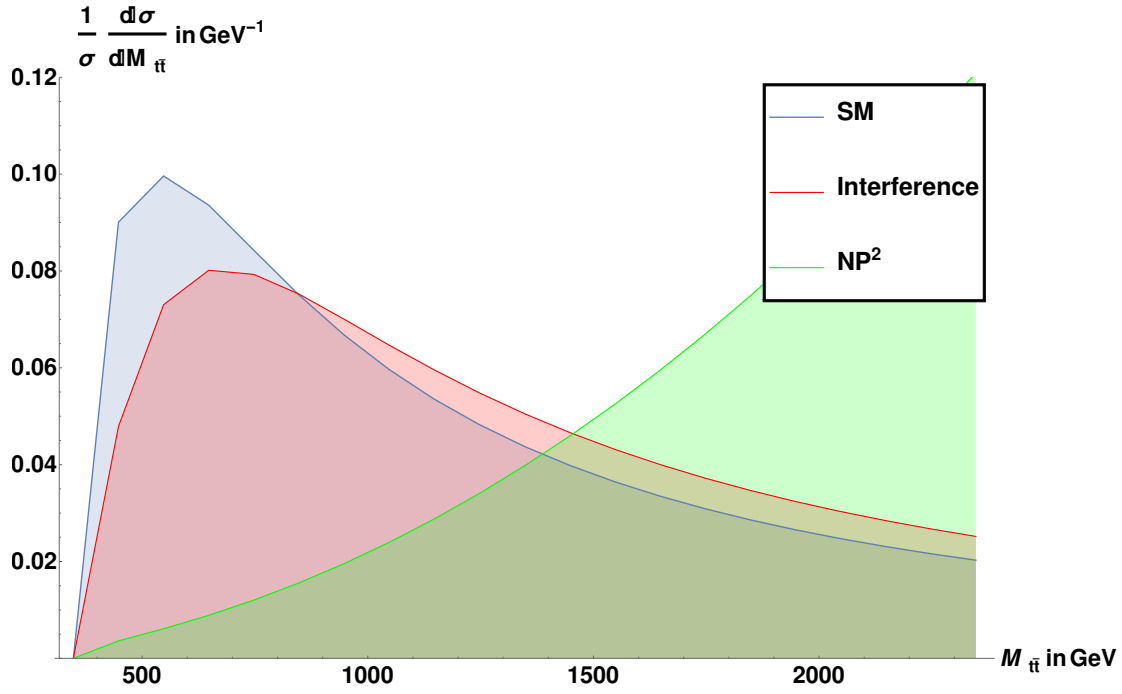


Figure 1.4: Normalized overlay plot for calculated parton-level cross-section of process $gg \to t\bar{t}$ for SM, New Physics and Interference terms in Lagrangian. On the x-axis, $m_{t\bar{t}}$ is plotted with intercept at $2m_t$ and steps of 50 GeV.

Consider $f(x)$ to be parton distribution function, then total cross-section at proton level can be written as

$$\hat{\sigma} = \int f(x_1) f(x_2) d\hat{\sigma}\,,$$

$$= \int f(x_1) f(x_2) \frac{1}{2\hat{s}} \frac{\sqrt{1-(2m_t/m_{t\bar{t}})^2}}{8\pi} \, \delta(\sqrt{\hat{s}} - m_{t\bar{t}}) dm_{t\bar{t}} |\mathscr{M}|^2 \, \frac{d\cos\theta}{2} \,. \tag{1.67}$$

Total $gg \to t\bar{t}$ amplitude squared, averaged over initial, summed over final spins and colors is given by Ref. [7]. Here, subscript 'int' stands for interference term between SM and NP.

$$\begin{aligned} |\overline{\mathscr{M}}|^2_{\rm SM}(gg \to t\bar{t}) = g_s^4 \Bigg\{ & \frac{3}{4} \frac{(m_t^2 - \hat{t})(m_t^2 - \hat{u})}{\hat{s}^2} - \frac{1}{24} \frac{m_t^2(\hat{s} - 4m_t^2)}{(m_t^2 - \hat{t})(m_t^2 - \hat{u})} \\ & + \frac{1}{6} \Big[\frac{\hat{t}\hat{u} - m_t^2(3\hat{t} + \hat{u}) - m_t^4}{(m_t^2 - \hat{t})^2} + \frac{\hat{t}\hat{u} - m_t^2(3\hat{u} + \hat{t}) - m_t^4}{(m_t^2 - \hat{u})^2} \Big] \\ & - \frac{3}{8} \Big[\frac{\hat{t}\hat{u} - 2\hat{t}m_t^2 + m_t^4}{(m_t^2 - \hat{t})\hat{s}} + \frac{\hat{t}\hat{u} - 2\hat{u}m_t^2 + m_t^4}{(m_t^2 - \hat{u})\hat{s}} \Big] \Bigg\} , \end{aligned} \tag{1.68}$$

$$|\overline{\mathscr{M}}|^2_{\rm int}(gg \to t\bar{t}) = \frac{1}{\Lambda^2} \left\{ \frac{9}{8} C_G \frac{m_t^2(\hat{t} - \hat{u})^2}{(m_t^2 - \hat{t})(m_t^2 - \hat{u})} \right\} , \tag{1.69}$$

$$|\overline{\mathscr{M}}|^2_{\rm NP^2}(gg \to t\bar{t}) = \frac{1}{\Lambda^4} \left\{ \frac{27}{4} C_G^2 (m_t^2 - \hat{t})(m_t^2 - \hat{u}) \right\} . \tag{1.70}$$

1.7 Convolutional Neural Network

Deep Convolutional Neural Networks or CNNs are powerful Machine Learning (ML) tools for visual object recognition. For training, supervised learning is most common, where we show a bunch of pictures to the machine for it to identify and learn patterns or motifs. Architectures with multiple layers learn through simple stochastic gradient descent [15]. Chain rule of derivatives (or gradients) gives a keen insight on this sort of training incorporating weights using backpropagation.

The following layers make up the architecture of a CNN. **Convolutional (Conv) layer** does most of the computing involved. It detects local features, correlations, or conjunctions of an image through learnable filters or kernels [16]. Kernel is a weights matrix created by the layer. Neurons/Units here are arranged in feature maps. Each neuron is connected to local spots on feature maps of the previous layer through a set of weights called filter banks. A non-linearity such as a rectified linear unit or ReLU is then passed through the result of this local weighted sum. **Pooling layer** merges similar features into one. It reduces the global computational complexity. **Dense or Fully Connected layer** is like a regular NN, where all units are connected to all of the previous layer's units. As opposed to Conv layers, neurons do not have a spatial organization here. **Dropout**

layer prevents overfitting and trains more robust parameters by dropping a fraction of input units. **Flatten layer** reduces the dimensionality of input layer to prepare a vector for Fully Connected layer.

ReLU activation returns element-wise $max(x,0)$. ReLU has sparsity property and suffers less from diminishing gradient flow [17]. Softmax activation function, $S(y_i) = e^{y_i}/\sum_j e^{y_j}$, gives a vector output representing the probability distributions of a list of potential outcomes.

In classification, the model fits parameters to predict the probability of an image being one of given categories. The loss or objective function helps it evaluate the prediction probabilities. Higher the loss function, worse are the predictions. For example, binary cross-entropy or log loss function performs quite well for binary classifications. It is given by,

$$-\frac{1}{N}\sum_{i=1}^{N}\left[y_i.log(P(y_i)) + (1-y_i).log(1-P(y_i))\right]. \tag{1.71}$$

Adam optimizer algorithm [18] can be used as an alternative/extension to stochastic gradient descent for computational efficiency and other advantages. An epoch is an iteration over the entire data provided. Batch size gives the number of samples per gradient update during training or testing.

Chapter 2

Cut and Count Analysis

The New Physics (NP) model is incorporated into the SM lagrangian using FeynRules [19] (refer to Appendix A) to produce a model file. Here, I have taken the relevant energy scale, or scale of NP, as $\Lambda = 1$ TeV and the Wilson coefficient $C_G = 1$. This model file can be imported in MadGraph [20] to generate samples of the process $gg \to t\bar{t}$ incorporating the NP interaction. The aim of the exercise is to put bounds on the Wilson coefficient of operator $\mathcal{O}_G$ (eq. 1.49), denoted by C_G. This analysis has $N = 10^6$ events each for three different kinds of samples and the values are scaled to correspond to integrated luminosity of 35.8 fb^{-1}. MadGraph samples generated include

1. only SM processes (gives $\sigma_{\text{SM\_LO}}$)[1],
2. only NP processes (gives σ_{NP^2}),
3. combined sample with both SM and NP processes[2] (gives σ_{Full}).

Let $\sigma_{Exp}^{(I)}$ be experimentally observed binwise differential cross-section for the I^{th} bin in the distribution of kinematic variables like $M_{t\bar{t}}$, ${p_{\text{T}}}^t(L)$ and $|y(t)|$ given by CMS [1]. $M_{t\bar{t}}$ denotes the invariant top mass, ${p_{\text{T}}}^t(L)$ denotes Leading or highest p_{T} top quark in the event, and $y(t)$ denotes rapidity of the leading top quark. Total experimental error on the bin, including statistical and systematic, is denoted by $\Delta\sigma_{Exp}^{(I)}$, where

$$\Delta\sigma_{Exp}^{(I)} = \sqrt{(\Delta\sigma_{stat}^{(I)})^2 + (\Delta\sigma_{sys}^{(I)})^2}\,. \tag{2.1}$$

[1] LO stands for Leading Order.
[2] includes three different amplitudes: only SM, only NP and SM+NP interference.

We construct the following cross-sections using the generated samples:

$$\sigma_{int}^{(I)} = \sigma_{\text{Full}}^{(I)} - \sigma_{\text{SM\_LO}}^{(I)} - \sigma_{\text{NP}^2}^{(I)}\,, \tag{2.2}$$

$$\sigma_{\text{NP}}^{(I)} = \frac{C_G^2}{(\Lambda/\text{ TeV})^4}\sigma_{\text{NP}^2}^{(I)} + \frac{C_G}{(\Lambda/\text{ TeV})^2}\sigma_{int}^{(I)}\,. \tag{2.3}$$

Now we take three cases for the analysis: Observed, Expected and D6/Linear. In the Observed case, the constant term we add to $\sigma_{\text{NP}}^{(I)}$ is $\kappa\,\sigma_{\text{SM\_LO}}$ (where $\kappa = \sigma_{\text{SM\_NNLO}}/\sigma_{\text{SM\_LO}} = 1.80269$) [21] while in the Expected case, the constant term added is σ_{Exp}. κ modifies SM contrto be NNLO (Next-to-Next-to-Leading Order). The third case, D6 represents the (linear) cross-section when we exclude the *only NP* contribution (C_G^2 term) from the Full Expected cross-section σ_{Extd} (eq. 2.6). This ensures that the NP contribution comes purely from dimension-six term, *i.e.*, the SM and NP interference term in this (D6) case.

$$\begin{aligned}\sigma_{Full}^{(I)} &= \kappa\sigma_{\text{SM\_LO}}^{(I)} + \sigma_{\text{NP}}^{(I)}\,,\\ &= \kappa\sigma_{\text{SM\_LO}}^{(I)} + \frac{C_G}{(\Lambda/\text{ TeV})^2}\sigma_{int}^{(I)} + \frac{C_G^2}{(\Lambda/\text{ TeV})^4}\sigma_{\text{NP}^2}^{(I)}\,,\end{aligned} \tag{2.4}$$

$$\sigma_{Obs}^{(I)} = \kappa\;\sigma_{\text{SM\_LO}}^{(I)} + \sigma_{\text{NP}}^{(I)}\,, \qquad (\text{where } \kappa = 1.80269) \tag{2.5}$$

$$\sigma_{Extd}^{(I)} = \sigma_{Exp}^{(I)} + \sigma_{\text{NP}}^{(I)}\,, \tag{2.6}$$

$$\sigma_{D6}^{(I)} = \sigma_{Exp}^{(I)} + \frac{C_G}{(\Lambda/\text{ TeV})^2}\sigma_{int}^{(I)}\,. \tag{2.7}$$

In order to obtain constraints on C_G, the following chi-square statistic (as a function of C_G and Λ) is constructed to evaluate goodness of fit:

$$\chi^2\left(C_G/\Lambda^2\right) = \sum_I \frac{(\sigma_{Full}^{(I)} - \sigma_{Exp}^{(I)})^2}{(\Delta\sigma_{Exp}^{(I)})^2}\,, \tag{2.8}$$

$$\Delta\chi^2\left(C_G/\Lambda^2\right) = \chi^2\left(C_G/\Lambda^2\right) - \chi^2_{min}\left(C_G/\Lambda^2\right)\,. \tag{2.9}$$

The following tables (Table 2.1 - Table 2.5) show obtained MadGraph values for the cross-section variables. Note that CMS measures cross-section multiplied by branching ratio, $\sigma \times \text{BR}_\ell$, of semi-leptonic $t\bar{t}$ events. At parton level, $\text{BR}_\ell \approx 0.29$ [11].

Bin No.	$M_{t\bar{t}}$ [GeV]	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]
1	300 – 360	18.76	13.21	1.78	3.77	51.10 ± 11.91
2	360 – 430	154.49	128.55	14.89	11.05	260.93 ± 22.31
3	430 – 500	132.55	107.54	15.22	9.79	190.93 ± 17.01
4	500 – 580	99.64	76.80	14.92	7.92	133.79 ± 8.98
5	580 – 680	72.88	52.11	14.77	6.01	90.00 ± 7.03
6	680 – 800	47.63	30.52	13.34	3.76	51.72 ± 4.22
7	800 – 1000	37.95	20.15	15.03	2.77	32.90 ± 2.49
8	1000 – 1200	17.34	6.79	9.42	1.13	11.24 ± 0.99
9	1200 – 1500	12.20	3.30	8.23	0.66	5.02 ± 0.64
10	1500 – 2500	10.34	1.43	8.78	0.13	2.14 ± 0.45

Table 2.1: This table gives the differential cross-section binned in the variable $M_{t\bar{t}}$ for $gg \to t\bar{t}$ process. These cross-section variables are described in section 2. The numbers are scaled to integrated luminosity 35.8 fb^{-1}. Note that C_G=1 and Λ=1 TeV for the Full sample. To scale SM LO value to NNLO, we use the κ factor ($\kappa \approx$1.8).

Bin No.	$p_T{}^t$ (L) [GeV]	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]
1	$0-40$	53.63	44.37	3.41	5.85	45.93 ± 4.42
2	$40-80$	128.21	105.86	9.05	13.30	171.59 ± 13.33
3	$80-120$	133.11	109.12	12.15	11.84	201.38 ± 15.23
4	$120-160$	98.92	78.70	12.80	7.42	167.31 ± 13.02
5	$160-200$	63.09	46.89	12.08	4.13	107.17 ± 7.37
6	$200-240$	38.38	25.44	10.70	2.22	64.55 ± 4.33
7	$240-280$	23.78	13.47	9.24	1.07	35.59 ± 2.25
8	$280-330$	18.40	8.44	9.49	0.48	23.28 ± 1.62
9	$330-380$	11.92	3.99	7.62	0.31	11.60 ± 0.92
10	$380-430$	8.17	1.95	6.10	0.12	5.93 ± 0.69
11	$430-500$	7.86	1.26	6.56	0.04	3.79 ± 0.39
12	$500-800$	14.40	0.91	13.22	0.31	3.27 ± 0.37

Table 2.2: This table gives the differential cross-section binned in the variable $p_T{}^t$ (leading top or L) for $gg \to t\bar{t}$ process. These cross-section variables are described in section 2. The numbers are scaled to integrated luminosity 35.8 fb^{-1}. Note that C_G=1 and Λ=1 TeV for the Full sample. To scale SM LO value to NNLO, we use the κ factor ($\kappa \approx 1.8$).

Bin No.	$\lvert y(t)\rvert$	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]
1	0.0 – 0.2	74.46	51.99	16.82	5.65	100.35 ± 6.51
2	0.2 – 0.4	72.47	51.17	16.31	4.95	99.66 ± 6.58
3	0.4 – 0.6	69.20	49.29	15.44	4.48	99.48 ± 6.03
4	0.6 – 0.8	65.39	46.55	14.12	4.72	89.45 ± 6.09
5	0.8 – 1.0	59.96	43.16	12.52	4.27	80.69 ± 5.61
6	1.0 – 1.2	54.03	39.14	10.58	4.31	73.45 ± 5.41
7	1.2 – 1.4	47.35	34.87	8.86	3.62	64.34 ± 4.59
8	1.4 – 1.6	40.62	30.30	7.12	3.20	53.86 ± 4.59
9	1.6 – 1.8	33.75	25.51	5.43	2.81	46.14 ± 3.76
10	1.8 – 2.0	26.94	20.64	4.01	2.30	36.69 ± 3.36
11	2.0 – 2.5	42.69	33.07	5.27	4.35	56.72 ± 5.11

Table 2.3: This table gives the differential cross-section binned in the variable $|y(t)|$ (of leading top) for $gg \to t\bar{t}$ process. These cross-section variables are described in section 2. The numbers are scaled to integrated luminosity 35.8 fb^{-1}. Note that C_G=1 and Λ=1 TeV for the Full sample. To scale SM LO value to NNLO, we use the κ factor ($\kappa \approx 1.8$).

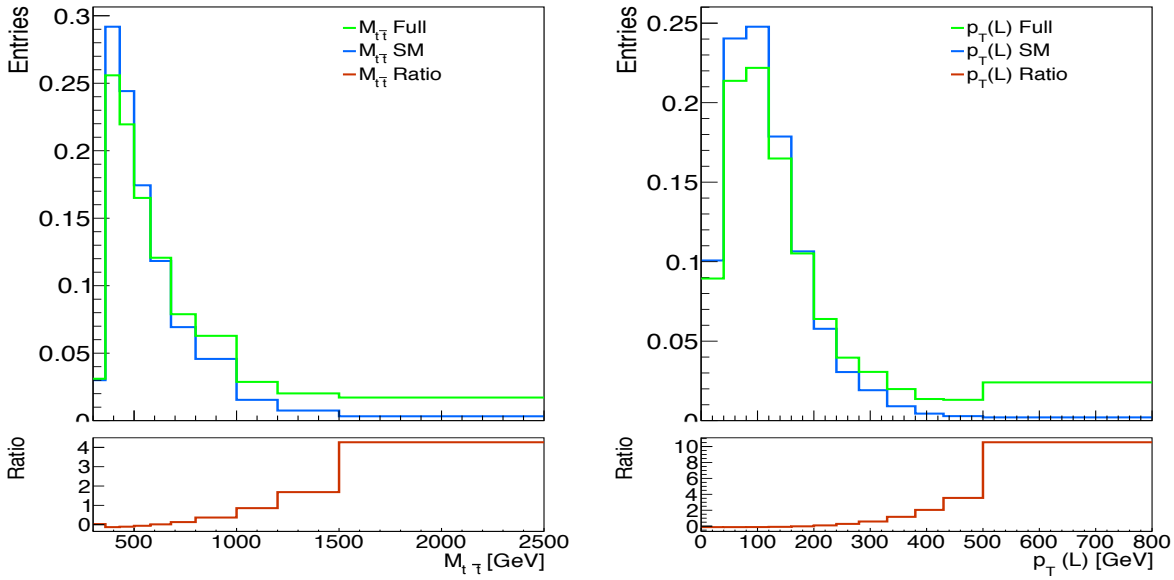
$M_{t\bar{t}}$ [GeV]	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]
	$0 < {p_{\text{T}}}^t < 90$					$90 < {p_{\text{T}}}^t < 180$				
$300-360$	18.76	13.21	1.78	3.77	42.83 ± 7.86	0	0	0	0	3.43 ± 0.59
$360-430$	111.11	94.03	9.47	7.60	175.31 ± 13.30	43.38	34.52	5.41	3.45	84.51 ± 6.36
$430-500$	41.28	34.04	2.27	4.96	64.74 ± 8.08	91.09	73.47	12.91	4.71	113.62 ± 9.60
$500-580$	21.77	17.99	0.96	2.82	34.01 ± 3.27	49.33	39.26	5.95	4.12	64.30 ± 5.27
$580-680$	12.26	10.15	0.45	1.66	20.23 ± 1.89	25.25	20.03	2.29	2.93	35.50 ± 3.06
$680-800$	6.39	5.19	0.21	0.99	10.39 ± 1.38	12.55	9.90	0.99	1.67	18.21 ± 2.13
$800-1000$	3.68	3.01	0.12	0.55	5.96 ± 1.22	7.20	5.76	0.51	0.93	10.68 ± 1.24
$1000-2000$	1.78	1.46	0.05	0.26	3.51 ± 1.11	3.45	2.74	0.21	0.50	5.25 ± 1.27
	$180 < {p_{\text{T}}}^t < 270$					$270 < {p_{\text{T}}}^t < 800$				
$300-430$	0	0	0	0	4.24 ± 1.19	0	0	0	0	0.74 ± 0.29
$430-500$	0.19	0.02	0.04	0.12	12.45 ± 0.92	0	0	0	0	1.80 ± 0.30
$500-580$	28.54	19.55	8.01	0.97	33.02 ± 2.42	0	0	0	0	2.87 ± 0.49
$580-680$	31.99	20.31	10.25	1.43	29.08 ± 2.36	3.38	1.61	1.78	-0.01	6.56 ± 0.65
$680-800$	13.25	8.96	3.33	0.97	14.41 ± 1.26	15.43	6.48	8.81	0.14	10.18 ± 0.76
$800-1000$	7.26	4.97	1.47	0.82	8.32 ± 0.84	19.82	6.41	12.93	0.48	9.47 ± 0.75
$1000-1200$	2.18	1.51	0.37	0.30	2.71 ± 0.44	11.90	2.63	8.88	0.40	3.75 ± 0.38
$1200-2000$	1.30	0.95	0.20	0.14	1.29 ± 0.42	14.59	1.99	12.23	0.38	2.95 ± 0.33

Table 2.4: This table gives the differential cross-section binned in two variables $M_{t\bar{t}}$ and ${p_{\text{T}}}^t(L)$ for $gg \to t\bar{t}$ process. These cross-section variables are described in section 2. The numbers are scaled to integrated luminosity 35.8 fb^{-1}. Note that C_G=1 and Λ=1 TeV for the Full sample. To scale SM LO value to NNLO, we use the κ factor ($\kappa \approx 1.8$).

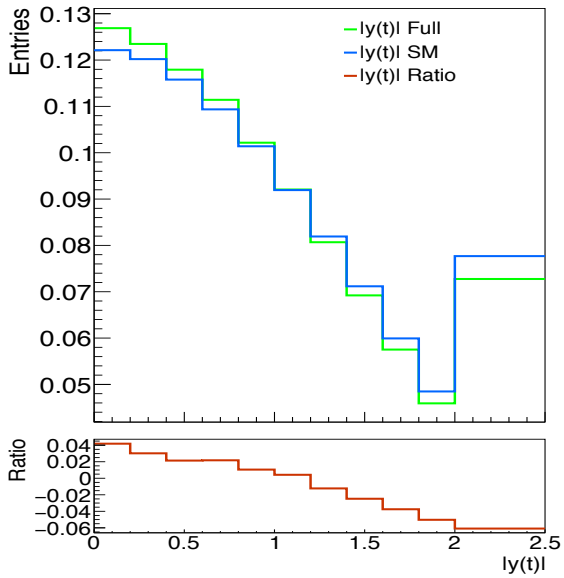
$p_T{}^t$ [GeV]	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]	$\sigma^{(I)}_{\text{Full}}$ [pb]	$\sigma^{(I)}_{\text{SM\_LO}}$ [pb]	$\sigma^{(I)}_{\text{NP}^2}$ [pb]	$\sigma^{(I)}_{int}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]
	$0 < \lvert y(t)\rvert < 0.5$					$0.5 < \lvert y(t)\rvert < 1$				
$0-40$	14.61	12.19	0.96	1.46	26.34 ± 1.81	13.16	10.84	0.87	1.45	23.24 ± 1.88
$40-80$	35.61	29.33	2.57	3.70	58.62 ± 4.02	31.96	26.59	2.32	3.05	52.34 ± 3.75
$80-120$	37.71	30.98	3.53	3.20	59.31 ± 4.16	33.97	27.87	3.19	2.91	52.83 ± 3.88
$120-160$	28.85	23.14	3.86	1.86	42.90 ± 2.99	25.65	20.51	3.42	1.72	37.79 ± 3.05
$160-200$	19.27	14.15	3.75	1.38	27.17 ± 1.87	16.87	12.69	3.28	0.90	23.03 ± 1.67
$200-240$	12.20	8.06	3.43	0.71	15.52 ± 1.04	10.51	7.02	3.02	0.48	13.18 ± 0.98
$240-280$	7.72	4.38	3.11	0.23	8.80 ± 0.53	6.73	3.76	2.64	0.32	7.31 ± 0.48
$280-330$	6.30	2.78	3.30	0.23	5.77 ± 0.37	5.32	2.47	2.80	0.05	4.84 ± 0.31
$330-380$	4.27	1.37	2.74	0.16	2.96 ± 0.22	3.58	1.20	2.34	0.04	2.47 ± 0.22
$380-450$	4.01	0.89	3.09	0.03	1.81 ± 0.18	3.29	0.76	2.58	-0.05	1.58 ± 0.19
$450-800$	8.37	0.66	7.58	0.12	1.56 ± 0.18	6.51	0.51	5.83	0.17	1.07 ± 0.13
	$1 < \lvert y(t)\rvert < 1.5$					$1.5 < \lvert y(t)\rvert < 2.5$				
$0-40$	10.70	9.03	0.69	0.99	18.55 ± 1.54	12.46	10.16	0.75	1.55	20.69 ± 2.11
$40-80$	25.91	21.21	1.84	2.86	41.59 ± 3.20	29.02	24.02	2.00	3.00	43.86 ± 3.63
$80-120$	27.06	22.24	2.49	2.33	40.21 ± 3.19	29.42	23.98	2.58	2.87	42.62 ± 3.90
$120-160$	20.45	16.15	2.65	1.66	28.55 ± 2.43	21.08	16.61	2.59	1.88	29.52 ± 3.06
$160-200$	12.99	9.65	2.53	0.80	17.38 ± 1.26	12.61	9.43	2.31	0.87	16.41 ± 1.54
$200-240$	7.95	5.29	2.23	0.43	9.86 ± 0.77	7.13	4.69	1.90	0.55	8.22 ± 0.76
$240-280$	4.99	2.83	1.92	0.25	5.31 ± 0.43	4.11	2.34	1.50	0.27	4.12 ± 0.45
$280-330$	3.77	1.75	1.95	0.07	3.29 ± 0.26	2.90	1.39	1.40	0.11	2.48 ± 0.27
$330-380$	2.43	0.83	1.56	0.04	1.52 ± 0.13	1.61	0.58	0.96	0.06	1.03 ± 0.18
$380-450$	2.20	0.49	1.61	0.10	0.92 ± 0.10	1.20	0.31	0.89	0.01	0.57 ± 0.11
$450-800$	3.56	0.33	3.14	0.09	0.71 ± 0.14	1.32	0.17	1.14	0.01	0.32 ± 0.07

Table 2.5: This table gives the differential cross-section binned in two variables $p_T{}^t$ and $|y(t)|$ (of leading top) for $gg \to t\bar{t}$ process. These cross-section variables are described in section 2. The numbers are scaled to integrated luminosity 35.8 fb^{-1}. Note that C_G=1 and Λ=1 TeV for the Full sample. To scale SM LO value to NNLO, we use the κ factor ($\kappa \approx$1.8).

The following normalized plots of kinematic variables $M_{t\bar{t}}$, $p_{\mathrm{T}}{}^{t}$ and $|y(t)|$, provide an estimate of NP contribution in their distribution by the quantity defined as Ratio=(Full-SM)/SM.



(a) $M_{t\bar{t}}$ (b) $p_{\mathrm{T}}{}^{t}$



(c) $|y(t)|$

Figure 2.1: In this figure, for the process $gg \to t\bar{t}$, upper plots give overlay of SM and Full (C_G=1, Λ=1 TeV) samples for variables $M_{t\bar{t}}$, $p_{\mathrm{T}}{}^{t}$ and $|y(t)|$. Lower plots show Ratio = (Full-SM)/SM. Upper plots are normalized to 1.

2.1 Cut and Count Results

Constraints obtained using Cut and Count analysis are tabulated here (Table 2.6 and Table 2.7). Note that in the case of $M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$ (high p_{T}), the bounds are calculated using only the high p_{T} bins of the 2-D distribution, *i.e.*, in the range $270 < p_{\mathrm{T}} < 800$.

We assume that only the experimental cross-section uncertainties scale corresponding to higher luminosities as $(\mathscr{L}_{high}/\mathscr{L}_{exp})^{-1/2}$ for calculating high luminosity bounds. For the Observed cases, when the χ^2 statistic gives a split region for bound on C_G, one of them is reported here.

$\mathscr{L}$ [fb^{-1}]	**Variable**	**C_G range** $(\times 10^1)$ Observed case	Expected case	D6/linear case
35.8	$M_{t\bar{t}}$	$[-2.83, 2.11]$	$[-3.75, 3.13]$	$[-12.2, 12.2]$
	${p_{\mathrm{T}}}^{t}(L)$	$[-5.39, -3.70]$	$[-2.85, 2.60]$	$[-13.6, 13.6]$
	$\lvert y(t)\rvert$	$[-1.07, 6.61]$	$[-8.94, 5.26]$	$[-12.3, 12.3]$
	$M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$	$[-2.54, 2.23]$	$[-3.56, 3.18]$	$[-16.3, 16.3]$
	${p_{\mathrm{T}}}^{t}(L)$ & $\lvert y(t)\rvert$	$[-3.64, 3.58]$	$[-3.05, 2.77]$	$[-15.5, 15.5]$
	$M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$ (high p_{T})	$[-1.57, 1.20]$	$[-2.59, 2.25]$	$[-16.8, 16.8]$
300	$M_{t\bar{t}}$	$[-1.35, 0.58]$	$[-2.28, 1.70]$	$[-4.23, 4.23]$
	${p_{\mathrm{T}}}^{t}(L)$	$[-4.90, -4.33]$	$[-1.70, 1.46]$	$[-4.69, 4.69]$
	$\lvert y(t)\rvert$	$[-8.74, -5.73]$	$[-6.18, 2.58]$	$[-4.27, 4.27]$
	$M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$	$[-1.07, 0.79]$	$[-2.14, 1.78]$	$[-5.62, 5.62]$
	${p_{\mathrm{T}}}^{t}(L)$ & $\lvert y(t)\rvert$	$[1.45, 2.74]$	$[-1.83, 1.56]$	$[-5.35, 5.35]$
	$M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$ (high p_{T})	$[-0.69, 0.31]$	$[-1.59, 1.26]$	$[-5.81, 5.81]$
3000	$M_{t\bar{t}}$	$[-0.71, -0.08]$	$[-1.30, 0.81]$	$[-1.34, 1.34]$
	${p_{\mathrm{T}}}^{t}(L)$	$[-4.71, -4.53]$	$[-0.96, 0.74]$	$[-1.48, 1.48]$
	$\lvert y(t)\rvert$	$[-7.88, -6.96]$	$[-4.37, 1.06]$	$[-1.35, 1.35]$
	$M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$	$[-0.43, 0.16]$	$[-1.22, 0.89]$	$[-1.78, 1.78]$
	${p_{\mathrm{T}}}^{t}(L)$ & $\lvert y(t)\rvert$	$[1.98, 2.37]$	$[-1.05, 0.79]$	$[-1.69, 1.69]$
	$M_{t\bar{t}}$ & ${p_{\mathrm{T}}}^{t}(L)$ (high p_{T})	$[-0.35, -0.03]$	$[-0.98, 0.65]$	$[-1.84, 1.84]$

Table 2.6: For process $gg \to t\bar{t}$, this table shows the bounds obtained on C_G and that estimated at high luminosities. These are obtained using $\Delta\chi^2 \leq 10$ for 1-D analyses and using $\Delta\chi^2 \leq n_b$ (where n_b = no. of bins) for 2-D analyses.

$\mathscr{L}$	**Variable**	**Expression equated to C_G/Λ^2**	
		$1/\tilde{\Lambda}^2$	$(1/16\pi^2)(g_3^3/\tilde{\Lambda}^2)$
[fb^{-1}]		[TeV]	[TeV]
35.8	$M_{t\bar{t}}$	$\tilde{\Lambda} > 1.79$	$\tilde{\Lambda} > 0.14$
	$p_{\mathrm{T}}{}^t$	$\tilde{\Lambda} > 1.96$	$\tilde{\Lambda} > 0.16$
	$\lvert y(t)\rvert$	$\tilde{\Lambda} > 1.38$	$\tilde{\Lambda} > 0.11$
	$M_{t\bar{t}}$ & $p_{\mathrm{T}}{}^t(L)$	$\tilde{\Lambda} > 1.77$	$\tilde{\Lambda} > 0.14$
	$p_{\mathrm{T}}{}^t(L)$ & $\lvert y(t)\rvert$	$\tilde{\Lambda} > 1.90$	$\tilde{\Lambda} > 0.15$
	$M_{t\bar{t}}$ & $p_{\mathrm{T}}{}^t(L)$ (high p_{T})	$\tilde{\Lambda} > 2.11$	$\tilde{\Lambda} > 0.17$
300	$M_{t\bar{t}}$	$\tilde{\Lambda} > 2.43$	$\tilde{\Lambda} > 0.19$
	$p_{\mathrm{T}}{}^t$	$\tilde{\Lambda} > 2.62$	$\tilde{\Lambda} > 0.21$
	$\lvert y(t)\rvert$	$\tilde{\Lambda} > 1.97$	$\tilde{\Lambda} > 0.16$
	$M_{t\bar{t}}$ & $p_{\mathrm{T}}{}^t(L)$	$\tilde{\Lambda} > 2.37$	$\tilde{\Lambda} > 0.19$
	$p_{\mathrm{T}}{}^t(L)$ & $\lvert y(t)\rvert$	$\tilde{\Lambda} > 2.52$	$\tilde{\Lambda} > 0.20$
	$M_{t\bar{t}}$ & $p_{\mathrm{T}}{}^t(L)$ (high p_{T})	$\tilde{\Lambda} > 2.82$	$\tilde{\Lambda} > 0.22$
3000	$M_{t\bar{t}}$	$\tilde{\Lambda} > 3.51$	$\tilde{\Lambda} > 0.28$
	$p_{\mathrm{T}}{}^t$	$\tilde{\Lambda} > 3.68$	$\tilde{\Lambda} > 0.29$
	$\lvert y(t)\rvert$	$\tilde{\Lambda} > 3.07$	$\tilde{\Lambda} > 0.24$
	$M_{t\bar{t}}$ & $p_{\mathrm{T}}{}^t(L)$	$\tilde{\Lambda} > 3.35$	$\tilde{\Lambda} > 0.27$
	$p_{\mathrm{T}}{}^t(L)$ & $\lvert y(t)\rvert$	$\tilde{\Lambda} > 3.56$	$\tilde{\Lambda} > 0.28$
	$M_{t\bar{t}}$ & $p_{\mathrm{T}}{}^t(L)$ (high p_{T})	$\tilde{\Lambda} > 3.92$	$\tilde{\Lambda} > 0.31$

Table 2.7: For process $gg \to t\bar{t}$, the energy scales for NP process are constrained using kinematic variable bounds on C_G (Expected case) by equating C_G/Λ^2 (where $\Lambda = 1$ TeV) to expressions mentioned in this table. Note that $g_3 = 1$.

Chapter 3

Machine Learning Analysis

In this chapter, I describe our approach to improve Cut and Count analysis results using Machine Learning (ML). We construct Convolutional Neural Network (CNN) using Keras [22]. It is trained as a classifier to distinguish signal against background. In general, for the following analyses, signal represents Full while background represents pure SM sample. Details are mentioned in the respective sections.

I describe two different approaches for training our CNN algorithm for such classification. These approaches are henceforth named as event-based analysis and bin-based analysis. Monte Carlo samples are generated using MadGraph for the parton-level gluons to top quarks ($gg \to t\bar{t}$) process along with the same process including one extra jet ($gg \to t\bar{t}g$). This allows us to skip the κ factor since $gg \to t\bar{t}$ including $gg \to t\bar{t}g$ gives a closer approximation to an NNLO $t\bar{t}$ production sample.

3.1 Event-based Analysis

In event-based analysis, the CNN is trained on images in which each pixel corresponds to the value of a kinematic variable at parton-level. Since we are considering four such variables, each event can be visualised as a square greyscale image with four pixels. In order to make the event kinematic variables represent an image, the values of these variables are rescaled to values in the range 0 to 256 such that they correspond to the greyscale intensity at their assigned pixel location.

A variable's original value, say a, is rescaled as

$$a \mapsto 256 \times \left(\frac{a_{max} - a}{a_{max} - a_{min}} \right). \tag{3.1}$$

At the time of analysis, these rescaled values are further normalized to 1.

$$\text{Image as a } 2\times2 \text{ matrix} : \begin{bmatrix} p_{\mathrm{T}}^{t}(L) & p_{\mathrm{T}}^{t\bar{t}} \\ |y(t)| & M_{t\bar{t}} \end{bmatrix}$$

Here, $p_{\mathrm{T}}^{t}(L)$ is p_{T} of leading top, $p_{\mathrm{T}}^{t\bar{t}}$ is invariant top p_{T} (p_{T} of four-vectorially added tops), $|y(t)|$ is rapidity of leading top, and $M_{t\bar{t}}$ is invariant top mass.

The default MadGraph selections are used in addition to the following event selections at generation level:

- $p_{\mathrm{T}}(t) > 400$ GeV
- $M_{t\bar{t}} > 1000$ GeV

Note that the MadGraph cross-section for these processes are $\sigma_{SM} \approx 6.44$ pb and $\sigma_{Full}^{2\ \mathrm{TeV}} \approx 11.8$ pb.

3.1.1 Binary Training

The 2×2 image inputs are categorized into following two categories:

- **SM** represents events for $gg \to t\bar{t}$ including $gg \to t\bar{t}g$ processes considering only SM interactions, and
- **Full** represents events for above processes considering SM plus NP ggg vertex upto first order at $\Lambda = 2$ TeV.

After training, the CNN algorithm can make predictions for the probability of a given event to be SM or Full (2 TeV). We test our trained algorithm on statistically different MC samples of both categories with 350,000 events each.

Label	SM	Full (2TeV)
Train	500,000	500,000
Validation	150,000	150,000
Test	350,000	350,000

Table 3.1: No. of samples/images used for event-based binary training of the CNN algorithm.

CNN Architecture:

- Model $\rightarrow$ Sequential.
- 2-D Conv layer with 32 filters, 2$\times$2 kernel, ReLU activation, and same padding.
- 2-D Conv layer with 32 filters, 2$\times$2 kernel, ReLU activation, and same padding.
- 2-D Conv layer with 64 filters, 1$\times$1 kernel, ReLU activation, and same padding.
- Dropout layer dropping 25% units.
- Flatten layer.
- Dense layer with 128 units and ReLU activation.
- Dense layer with 50 units and ReLU activation.
- Dropout layer dropping 10% units.
- Dense layer with 2 units and softmax activation.
- Compiling $\rightarrow$ Adam optimizer, binary cross-entropy loss function, and accuracy metric.
- Fitting/Training $\rightarrow$ 50 epochs with batch size of 150.

3.1.2 Results for Binary Training

Label	Accuracy (%)
Train	71.18
Validation	72.56
Test	72.51

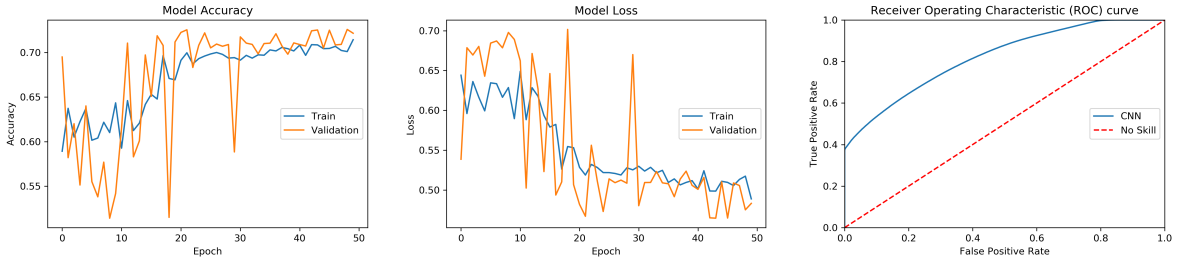
Table 3.2: Accuracy obtained on binary training of the CNN algorithm.

The probability of a test event to be predicted as SM event is given by P(SM).P(SM)=0 implies Full event predicted and P(SM)=1 implies SM event predicted. The default threshold for binary classification is 0.50, *i.e.*, events with P(SM)$>$ 0.50 are predicted to be SM events by the trained CNN algorithm. Fig. 3.2 shows P(SM) plotted for various test samples. Table 3.3 shows P(SM) while varying thresholds.

Label	**Threshold of P(SM)**			
	0.50	**0.80**	**0.90**	**0.95**
Signal events	222,717	135,767	132,231	130,590
Background events	127,283	214,233	217,769	219,410
$S/\sqrt{B}$	624.26	293.33	283.36	278.79

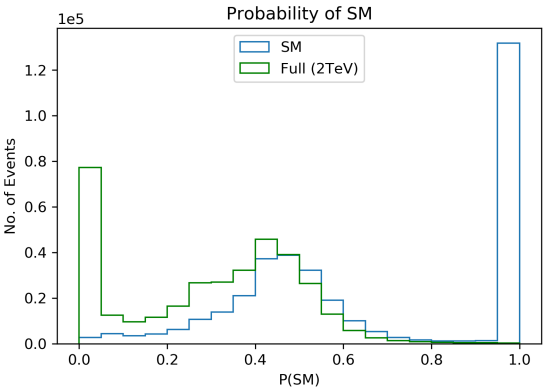
Table 3.3: This table shows the change in $S/\sqrt{B}$ while varying threshold for classification. The sample being tested is Full (2 TeV) with 350,000 events. Signal is predicted Full event while background is predicted SM event.

The Receiver Operating Characteristic (ROC) plot gives Area Under Curve (AUC) = 0.8263.

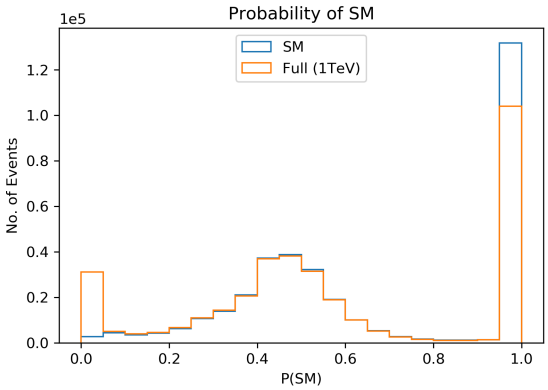


(a) Model Accuracy (b) Model Loss (c) ROC curve

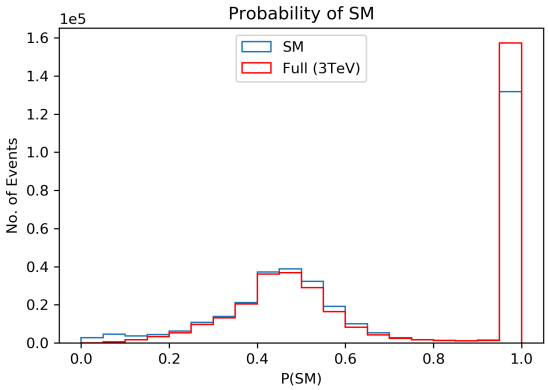
Figure 3.1: These plots depict the performance of CNN for the event-based analysis. CNN training accuracy and loss are shown in fig. 3.1(a) and 3.1(b) respectively with the no. of epochs in the $x-$axis. Fig. 3.1(c) shows the ROC curve.



(a) 2 TeV



(b) 1 TeV



(c) 3 TeV

Figure 3.2: P(SM) for binary event-based analysis − Trained model gives the prediction probability P(SM) on the test samples. Test is performed on MC SM and Full samples at different energy scales of Λ=1, 2 and 3 TeVwith 350,000 events each.

3.1.3 Multi-category Training

The 2×2 image inputs are categorized into following four categories:

- **SM** represents events for $gg \to t\bar{t}$ including $gg \to t\bar{t}g$ processes considering only SM interactions,
- **Full (1, 2, 3 TeV)** represent events for above processes considering SM plus NP ggg vertex upto first order at three different energy scales, $\Lambda = 1, 2, 3$ TeV.

Label	SM	Full (1TeV)	Full (2TeV)	Full (3TeV)
Train	500,000	500,000	500,000	500,000
Validation	150,000	150,000	150,000	150,000
Test	350,000	350,000	350,000	350,000

Table 3.4: No. of samples/images used for event-based multi-category training of the CNN algorithm.

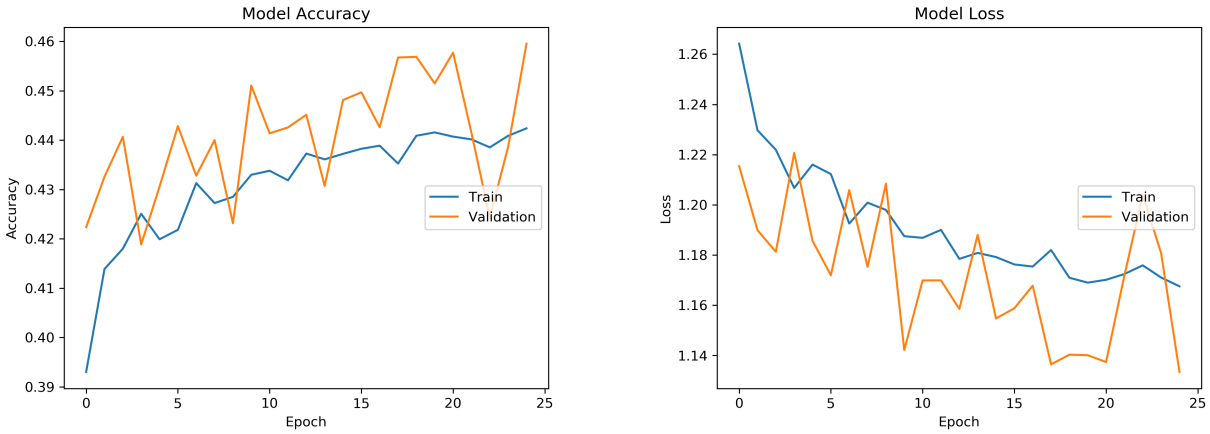
CNN Architecture:

- Model $\rightarrow$ Sequential.
- 2-D Conv layer with 32 filters, 2$\times$2 kernel, ReLU activation, and same padding.
- 2-D Conv layer with 32 filters, 2$\times$2 kernel, ReLU activation, and same padding.
- 2-D Conv layer with 64 filters, 1$\times$1 kernel, ReLU activation, and same padding.
- Dropout layer dropping 25% units.
- Flatten layer.
- Dense layer with 128 units and ReLU activation.
- Dense layer with 50 units and ReLU activation.
- Dropout layer dropping 20% units.
- Dense layer with 2 units and softmax activation.
- Compiling $\rightarrow$ Adam optimizer, binary cross-entropy loss function, and accuracy metric.
- Fitting/Training $\rightarrow$ 25 epochs with batch size of 50.

3.1.4 Results for Multi-Category Training

Label	Accuracy (%)
Train	44.24
Validation	45.95
Test	45.95

Table 3.5: Accuracy obtained on multi-category training of the CNN algorithm.



(a) Model Accuracy

(b) Model Loss

Figure 3.3: These plots depict the performance of CNN for the multi-category training for event-based analysis. CNN training accuracy and loss are shown in fig. 3.3(a) and 3.3(b) respectively with the no. of epochs in the $x-$axis.

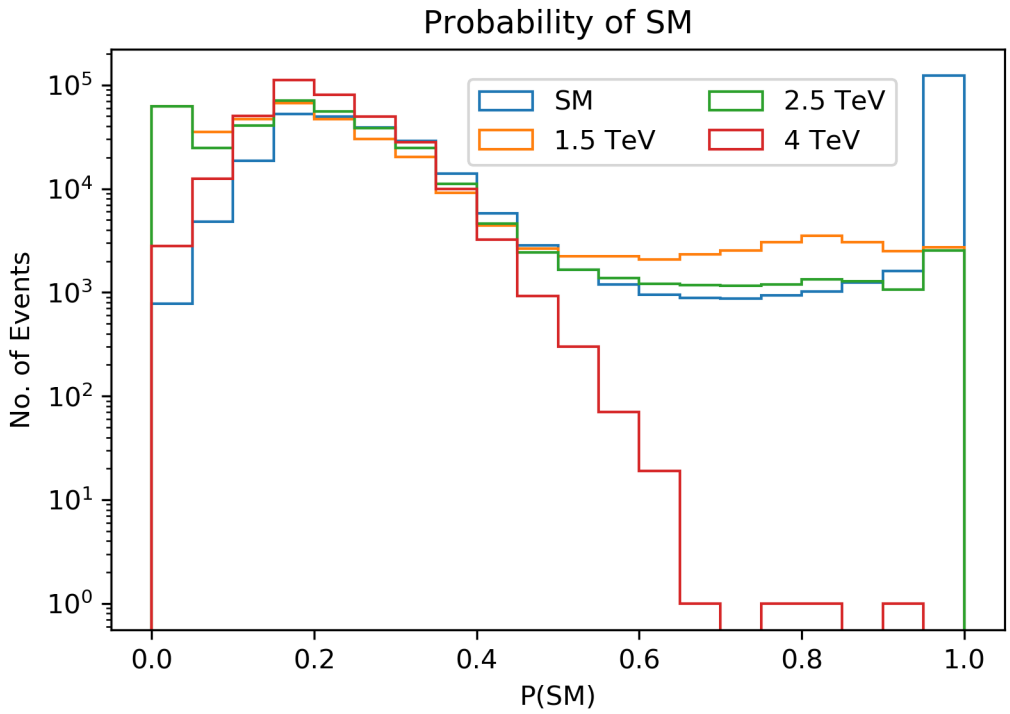


Figure 3.4: P(SM) for multi-category event-based analysis − Trained model gives the prediction probability P(SM) on the test samples. Test is performed on MC SM and Full samples at different energy scales of Λ=1.5, 2.5 and 4 TeVwith 350,000 events each.

3.2 Bin-based Analysis

The NP vertex is enhanced at high energies and momenta, giving a difference in distribution of kinematic variables at high-energy tails. We used this property to do the Cut and Count analysis. In the bin-based approach, we exploit this property to get a better classification out of CNN algorithm. In event-based analysis, one image represented an event but in bin-based analysis, each image represents a 2-D $M_{t\bar{t}} - p_T(L)$ distribution made using 500,000 events. One pixel has information about the number of events in specific bin ranges. We take last eight bins at the tail of the distribution and normalize them to 1. These serve as the value of each pixel. Hence, each image can be seen as a proxy for a normalized 2-D histogram made using 500,000 events. Given below

is the input visualised in 2×4 matrix form (variable ranges in GeV).

$$\text{Image as a } 2\times4 \text{ matrix}: \begin{array}{c} \begin{matrix} M_{t\bar{t}}\in[680,800] & M_{t\bar{t}}\in[800,1000] & M_{t\bar{t}}\in[1000,1200] & M_{t\bar{t}}\in[1200,2000] \end{matrix} \\ \begin{bmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \end{bmatrix} \end{array} \begin{matrix} \\ p_T(L)\in[180,270] \\ p_T(L)\in[270,800] \end{matrix}$$

The default MadGraph selections are used in addition to the following event selections at generation level:

- 180 GeV $< p_{\mathrm{T}}(t) <$ 800 GeV
- $M_{t\bar{t}} >$ 680 GeV

Note that the cross-section for these processes with above event selections are $\sigma_{SM} \approx 74.8$ pb, $\sigma_{Full}^{1\text{ TeV}} \approx 224$ pb, $\sigma_{Full}^{2\text{ TeV}} \approx 84$ pb, $\sigma_{Full}^{2.5\text{ TeV}} \approx 78$ pb, and $\sigma_{Full}^{3\text{ TeV}} \approx 76$ pb.

Sample images (Normalized to 1) :

$$\text{SM}: \begin{bmatrix} 0.1958 & 0.1102 & 0.0331 & 0.0198 \\ 0.2383 & 0.2280 & 0.0982 & 0.0767 \end{bmatrix}, \text{Full}^{1\text{ TeV}}: \begin{bmatrix} 0.0985 & 0.0538 & 0.0162 & 0.0095 \\ 0.1980 & 0.2530 & 0.1596 & 0.2114 \end{bmatrix},$$

$$\text{Full}^{2\text{ TeV}}: \begin{bmatrix} 0.1958 & 0.1102 & 0.0331 & 0.0198 \\ 0.2383 & 0.2280 & 0.0982 & 0.0767 \end{bmatrix}, \text{Full}^{2.5\text{ TeV}}: \begin{bmatrix} 0.1904 & 0.1071 & 0.0326 & 0.0190 \\ 0.2355 & 0.2294 & 0.1015 & 0.0845 \end{bmatrix},$$

$$\text{Full}^{3\text{ TeV}}: \begin{bmatrix} 0.1935 & 0.1087 & 0.0334 & 0.0192 \\ 0.2371 & 0.2288 & 0.0999 & 0.0796 \end{bmatrix}.$$

Range of values for uniformly distributed values within 1-σ uncertainty (Data):

$$\text{Data within 1-}\sigma: \begin{bmatrix} [0.1698, 0.2195] & [0.0935, 0.1274] & [0.0242, 0.0426] & [0.0114, 0.0282] \\ [0.2191, 0.2593] & [0.2090, 0.2507] & [0.0878, 0.1100] & [0.0671, 0.0855] \end{bmatrix}.$$

Cross-section (MadGraph):

$$\sigma_{SM}: \begin{bmatrix} 14.6 & 8.2 & 2.5 & 1.5 \\ 17.8 & 17.1 & 7.3 & 5.8 \end{bmatrix}, \sigma^{2\ \text{TeV}}_{Full}: \begin{bmatrix} 15.2 & 8.6 & 2.6 & 1.5 \\ 19.2 & 19.5 & 9.1 & 8.2 \end{bmatrix}.$$

Label	**SM**	**Full(1TeV)**	**Full(2TeV)**	**Full(2.5TeV)**	**Full(3TeV)**
Train	50	25	50	25	25
Validation	15	10	15	10	10
Test	35	15	35	15	15

Table 3.6: No. of Samples/Images used for bin-based training of CNN algorithm.

After training the CNN, we use it to make predictions on publicly available CMS data. Similar to the expected bounds in Cut and Count analysis, we use only the uncertainty on each bin by fixing the central value of the bin to average of the value corresponding to MC SM sample. A set of images is generated with central value fixed as described above and uncertainty on bin as reported by CMS assuming a uniform distribution in that uncertainty band. Assuming uniform distribution within $1-\sigma$ uncertainty, a set of 1000 such images is generated representing data. Fig. 3.6 shows the results for the same. The average values for P(SM) are reported in Table 3.9.

CNN Architecture:

- Model $\rightarrow$ Sequential.
- 2-D Conv layer with 16 filters, 2$\times$2 kernel, ReLU activation, and same padding.
- 2-D Conv layer with 32 filters, 2$\times$2 kernel, ReLU activation, and same padding.
- 2-D Conv layer with 64 filters, 1$\times$1 kernel, ReLU activation, and same padding.
- Dropout layer dropping 20% units.
- Flatten layer.
- Dense layer with 128 units and ReLU activation.
- Dense layer with 50 units and ReLU activation.
- Dropout layer dropping 20% units (only for 3 TeV case).
- Dense layer with 2 units and softmax activation.

- Compiling → Adam optimizer, binary cross-entropy loss function, and accuracy metric.
- Fitting/Training → 50 epochs[1] with batch size of 5.

$M_{t\bar{t}}$ [GeV]	$\sigma^{(I)}_{\text{Full, 2TeV}}$ [pb]	$\sigma^{(I)}_{\text{SM}}$ [pb]	$\sigma^{(I)}_{Exp}$ [pb]
	$180 < {p_{\text{T}}}^t < 270$		
$680-800$	15.2	14.6	14.41 ± 1.26
$800-1000$	8.6	8.2	8.32 ± 0.84
$1000-1200$	2.6	2.5	2.71 ± 0.44
$1200-2000$	1.5	1.5	1.29 ± 0.42
	$270 < {p_{\text{T}}}^t < 800$		
$680-800$	19.2	17.8	10.18 ± 0.76
$800-1000$	19.5	17.1	9.47 ± 0.75
$1000-1200$	9.1	7.3	3.75 ± 0.38
$1200-2000$	8.2	5.8	2.95 ± 0.33

Table 3.7: This table gives the double differential cross-sections binned in $M_{t\bar{t}}$ and ${p_{\text{T}}}^t$ for the processes $gg \to t\bar{t}$ & $gg \to t\bar{t}g$. It also contains experimentally obtained values for the double differential cross-section at the parton level as a function of $M_{t\bar{t}}$ and ${p_{\text{T}}}^t$ [1]. The numbers are scaled to integrated luminosity 35.8 fb^{-1}. Note that $C_G = 1$ & $\Lambda = 2$ TeV.

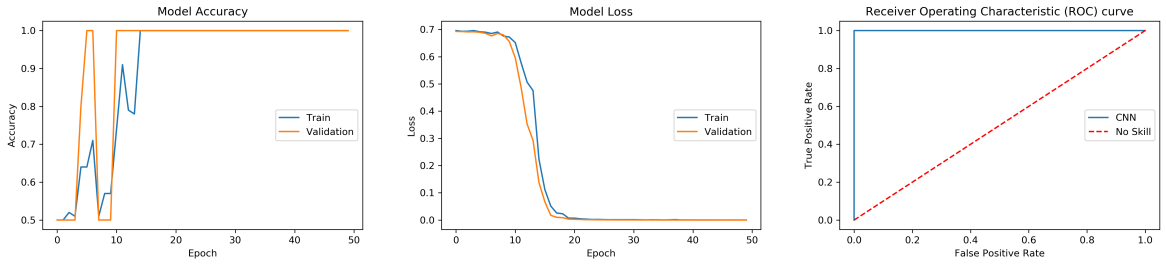
3.2.1 Results for bin-based analysis

Label	**Accuracy (%)**	**Accuracy (%)**	**Accuracy (%)**	**Accuracy (%)**
Train	100	100	100	54
Validation	100	100	100	50
Test	100	100	100	50
Trained on:	SM v 1TeV Full	SM v 2TeV Full	SM v 2.5TeV Full	SM v 3TeV Full

Table 3.8: Accuracy results for bin-based analysis.

[1] epochs=150 & 500 for the 2.5 & 3 TeV case respectively

The Receiver Operating Characteristic (ROC) plot gives Area Under Curve (AUC) = 1.

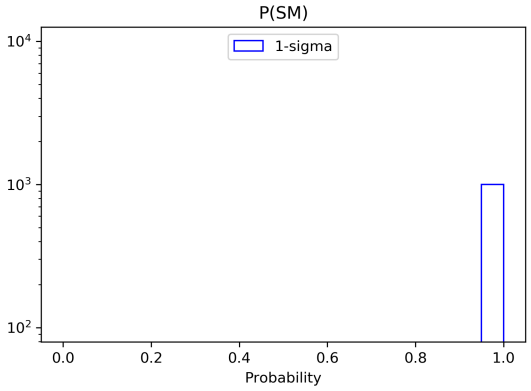


(a) Model Accuracy (b) Model Loss (c) ROC curve

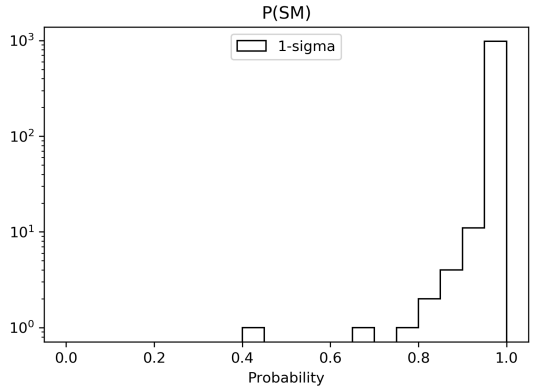
Figure 3.5: These plots depict the performance of CNN for the bin-based analysis. CNN training accuracy and loss are shown in fig. 3.5(a) and 3.5(b) respectively with the no. of epochs in $x-$axis. Fig. 3.5(c) shows the ROC curve.

Training Categories	**P(SM) (%)**
SM and 1 TeV Full	99.9989
SM and 2 TeV Full	99.6224
SM and 2.5 TeV Full	84.8864
SM and 3 TeV Full	50.4207

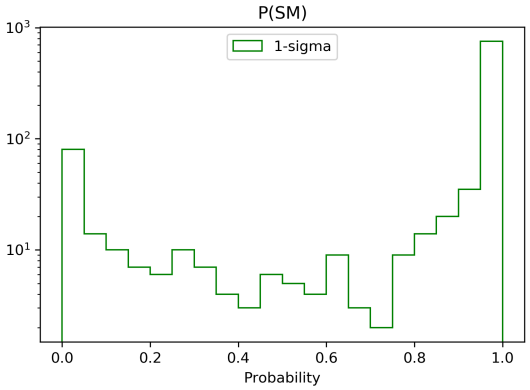
Table 3.9: Predictions on data images within 1-σ uncertainty, averaged for 1000 such images. All the Full samples are trained against SM individually.



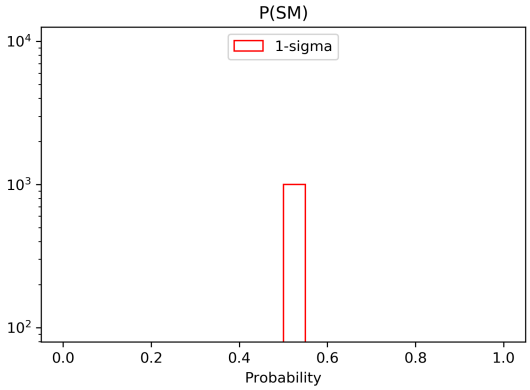
(a) SM and 1 TeV Full



(b) SM and 2 TeV Full



(c) SM and 2.5 TeV Full



(d) SM and 3 TeV Full

Figure 3.6: P(SM) for the bin-based analysis− Trained model gives the prediction probability P(SM) on the test sample. These plots represents the model's prediction on 1000 images representing CMS data for the 2-D $M_{t\bar{t}} - p_{\mathrm{T}}^{t}(L)$ distribution within $1-\sigma$ uncertainty.

Chapter 4

Discussions and Outlook

In the Cut and Count analysis, we have reported three different bounds: Observed, Expected, and Linear/D6 bounds. The Observed bounds are calculated using tree-level MadGraph simulations representing the SM contribution. For top pair production, leading order (LO) simulations give SM cross-section $\sigma_{\text{SM\_LO}} \sim 400$ pb while NNLO QCD cross-section is $\sigma_{\text{SM}} \sim 800$ pb. Even though we have tried to correct for this discrepancy using the κ factor, it is not good enough. The κ factor corrects for the overall cross-section while an ideal correction would be using bin-by-bin cross-section correction factor. This could not be done since these numbers are not explicitly reported [21]. With these numbers or QCD NNLO simulations, one can find a better and more reliable bound in the Observed case.

The Expected bounds are computed by substituting the MadGraph SM cross-section values by corresponding bin central values measured in the CMS experiment [1]. The Expected bound results are fairly consistent using different kinematic variable distributions. The strongest bound on C_G and Λ is observed using the differential cross-section binned in 2-D distribution of $M_{t\bar{t}}$ & $p_{\text{T}}^{t}(L)$ in the high p_{T} regime ($270 < p_{\text{T}} < 800$ GeV). The NP energy scale we can roughly eliminate using this analysis is $\Lambda < 2$ TeV. We report reliable conclusions from the Cut and Count analysis using the Expected bound results.

The Linear bounds are calculated by omitting the higher order $\mathcal{O}(\Lambda^{-4})$ terms in the Expected case. This leaves the SM $\times$ NP interference term as major contributor to the bound. They are fairly weaker than the Expected bounds. This indicates that the significant contribution to the bound does not come from the interference term of $\mathcal{O}(\Lambda^{-2})$.

Next, we tried Machine Learning (ML) techniques for improving our results. For the event-based analysis, firstly, we trained the CNN for binary classification under SM and Full categories. Each event was fed as a 2×2 image with the pixels as normalized values of some important kinematic variables. The hope was that the algorithm will learn some correlations or other hidden features/motifs in these images which are not usually observed. This was successful to an extent, giving an accuracy of $\sim 70-75\%$ for classification. Since we did not have event-by-event data information publicly available, the trained algorithm could not be tested on it. However, the qualitative results look quite promising and one can put bounds once eventwise data is available. We also checked the performance of algorithm trained on 2 TeV Full sample against SM on other generated Full samples of 1 and 3 TeV NP scale. The performance on these samples was observed to be poor which indicated that the algorithm is not learning Full vs SM but it is exclusively learning Full sample at 2 TeV scale. This led us to try the multi-category analysis.

The multi-category classification was performed on SM and Full 1, 2 and 3 TeV samples. The accuracy obtained was $\sim 40-45\%$ which is relatively good when compared to purely random prediction with 25% accuracy. The algorithm was tested on Full samples of different energy scales: 1.5, 2.5 and 4 TeV. The performance got poorer with increasing energy. It was best on 1.5 TeV, then 2.5 TeV, and lastly on 4 TeV sample. One may expect this intuitively since $\Lambda \to \infty$ would give purely SM sample, meaning that with increasing Λ, we get closer to SM.

Finally, in the bin-based analysis, we thought of incorporating the fact that our NP vertex contributes to high energy tails of the observables. This information was not obviously present in the event-based analysis. A histogram with a sufficiently populated high energy tail can do the trick. Our image input should reflect that. Hence, our image contains 8 pixels representing normalized values for no. of events in the last 8 bins of 2-D $M_{t\bar{t}} - p_{\mathrm{T}}^{t}(L)$ distribution. This method is computationally quite heavy but it produced nice qualitative results. Three different binary classification trainings were done as an attempt to observe how the algorithm performs at different NP energy scales (1, 2, 2.5 and 3 TeV). Since we had data available for these bins, we asked our trained model to predict the probability of data within 1-σ uncertainty being purely SM. Based on these results, we can roughly estimate that $\Lambda > 2.5$ TeV. But proper quantification techniques have to be worked out to estimate final bounds.

Translating the ML results into C_G bounds is tricky and needs more sophisticated method(s) of quantification and error estimation. We expect stronger bounds than Cut and Count using ML.

Appendix A

FeynRules

Following lines were added from top EFT model (∗.fr file) by author C. Degrande [23] to the SM in order to incorporate new tri-gluon interaction ($\mathscr{L}_G$):

- Lambda== {
  ParameterType -> External,
  ParameterName -> Lambda,
  BlockName -> DIM6,
  InteractionOrder -> {NP,-1},
  Value -> 1000,
  TeX -> Λ,
  Description -> ”Scale of the new physics”},

- CG== {
  ParameterType -> External,
  ParameterName -> CG,
  BlockName -> DIM6,
  Value -> 1,
  TeX -> Subscript[C,G],
  Description -> ”coefficient of OG”},

- LG := CG/Lambda^2 Module[{aa, bb, cc, mu, nu, rho}, f[aa, bb, cc] FS[G, mu, nu, aa] FS[G, nu, rho, bb] FS[G, rho, mu, cc]];

Bibliography

[1] A. M. Sirunyan *et al.*, [CMS], "Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev.* **D97** no. 11, (2018) 112003, arXiv:1803.08856 [hep-ex].

[2] F. Dave, "The standard model of particle physics: A lunchbox's guide," 2008.

[3] S. Weinberg, "Phenomenological Lagrangians," *Physica* **A96** no. 1-2, (1979) 327–340.

[4] W. Buchmuller and D. Wyler, "Effective Lagrangian Analysis of New Interactions and Flavor Conservation," *Nucl. Phys.* **B268** (1986) 621–653.

[5] V. Hirschi, F. Maltoni, I. Tsinikos, and E. Vryonidou, "Constraining anomalous gluon self-interactions at the LHC: a reappraisal," *JHEP* **07** (2018) 093, arXiv:1806.04696 [hep-ph].

[6] E. H. Simmons, "Dimension-six Gluon Operators as Probes of New Physics," *Phys. Lett.* **B226** (1989) 132–136.

[7] P. L. Cho and E. H. Simmons, "Searching for G3 in $t\bar{t}$ production," *Phys. Rev.* **D51** (1995) 2360–2370, arXiv:hep-ph/9408206 [hep-ph].

[8] M. Tanabashi *et al.*, [Particle Data Group], "Review of Particle Physics," *Phys. Rev.* **D98** no. 3, (2018) 030001.

[9] G. Aad *et al.*, [ATLAS], "Measurement of $W^{\pm}$ boson production in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV with the ATLAS detector," arXiv:1907.10414 [nucl-ex].

[10] G. Aad *et al.*, [ATLAS], "Measurement of K^0_S and Λ^0 production in $t\bar{t}$ dileptonic events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," arXiv:1907.10862 [hep-ex].

[11] M. Farina, C. Mondino, D. Pappadopulo, and J. T. Ruderman, "New Physics from High Energy Tops," *JHEP* **01** (2019) 231, arXiv:1811.04084 [hep-ph].

[12] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian," *JHEP* **10** (2010) 085, arXiv:1008.4884 [hep-ph].

[13] A. V. Manohar, "Introduction to Effective Field Theories," in *Les Houches summer school: EFT in Particle Physics and Cosmology Les Houches, Chamonix Valley, France, July 3-28, 2017*. 2018. arXiv:1804.05863 [hep-ph].

[14] C. Zhang and S. Willenbrock, "Effective-Field-Theory Approach to Top-Quark Production and Decay," *Phys. Rev.* **D83** (2011) 034006, arXiv:1008.3869 [hep-ph].

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521** (2015) 436–444.

[16] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," *2017 International Conference on Communication and Signal Processing (ICCSP)* (2017) 0588–0592.

[17] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudk, eds., vol. 15 of *Proceedings of Machine Learning Research*, pp. 315–323. PMLR, Fort Lauderdale, FL, USA, 11–13 apr, 2011. http://proceedings.mlr.press/v15/glorot11a.html.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[19] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, "FeynRules 2.0 - A complete toolbox for tree-level phenomenology," *Comput. Phys. Commun.* **185** (2014) 2250–2300, arXiv:1310.1921 [hep-ph].

[20] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. 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," *JHEP* **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[21] S. Catani, S. Devoto, M. Grazzini, S. Kallweit, and J. Mazzitelli, "Top-quark pair production at the LHC: Fully differential QCD predictions at NNLO," *JHEP* **07** (2019) 100, arXiv:1906.06535 [hep-ph].

[22] F. Chollet *et al.*, "Keras." https://keras.io, 2015.

[23] C. Degrande, J.-M. Gerard, C. Grojean, F. Maltoni, and G. Servant, "Non-resonant New Physics in Top Pair Production at Hadron Colliders," *JHEP* **03** (2011) 125, arXiv:1010.6304 [hep-ph].