이학박사 학위논문

Application of Machine Learning on Searches at Large Hadron Collider

대형 강입자 가속기 기반 탐색에서 기계 학습 응용

2021 년 2 월

서울대학교 대학원

물리천문학부 물리학전공

이 동 섭

Application of Machine Learning
on Searches at Large Hadron Collider

대형 강입자 가속기 기반 탐색에서의 기계 학습 응용

지도교수 김 형 도

이 논문을 이학박사 학위논문으로 제출함

2021 년 1 월

서울대학교 대학원

물리천문학부 물리학 전공

이 동 섭

이동섭의 이학박사 학위논문을 인준함

2020 년 12 월

위 원 장 정성훈 (인)

부 위 원 장 김형도 (인)

위 원 이현민 (인)

위 원 양윤기 (인)

위 원 최선호 (인)

Abstract

Application of Machine Learning on Searches
at Large Hadron Collider

LEE, Dongsub
Department of Physics and Astronomy
The Graduate School
Seoul National University

In this thesis, we study the application of machine learning for searches at the Large Hadron Collider without a sharp resonance peak. First, we use machine learning to find the best observables for the broad resonance search. A vector resonance from the composite Higgs models in $t\bar{t}$ final state is considered as a benchmark. Various approaches are adopted to interpret the abstracted information by the machine, and we conclude that the resonance energy is still important for the broad resonance search, while the angular distributions and the transverse momenta of the decayed products have also great importance. Second, we use machine learning to extract information about the resonance from other than the final state. We show the correlation between the kinematics of jets from initial state radiation and the resonance particle. To demonstrate the experimental feasibility we perform the searching for invisible decay of Higgs by using machine learning. As a result, we show that the bound from gluon-fusion production mechanism can be improved even stronger than the other production mechanisms due to the correlation.

Keywords: Large Hadron Collider, Machine Learning, Broad Resonance, Higgs, jet, jet substructure, Initial State Radiation
Student Number: 2015-20343

Contents

List of Figures

List of Tables

Chapter 1

Introduction

By the discover of Higgs particle [1–3] at the Large Hadron Collider(LHC) of European Organisation for Nuclear Research(CERN) in 2012, all of the particle contents of the standard model are found. The standard model of particle physics becomes the most successful description of various interactions and phenomena of fundamental particles in nature at scale larger than 10^{-19} m($\sim$ 10 TeV). However, the standard model cannot be the theory of everything because of problems which cannot be answered by itself. One of the problems is the absence of dark matter candidate in the standard model. To explain cosmological phenomena such as rotation of galaxy [4] or collision of bullet-clusters [5–7], non-luminous matters should be clustered around those astrophysical structures. From recent observations [8], most of those additional components are 'cold', i.e has no pressure and mass larger than the temperature scale of the current universe. But there are no such particle contents in the standard model; non-relativistic, electromagnetically neutral, and stable or at least having a very long lifetime. Another problem is the so-called naturalness problem of Higgs

mass. Higgs field, Ψ, is two complex scalar fields in a doublet of $SU(2)_L$ with the following potential

$$V(\Psi) = -\mu^2|\Psi|^2 + \lambda|\Psi|^4, \tag{1.1}$$

at the classical level. If the ultra-violet(UV) cut-off of the standard model is at $\Lambda_{\rm UV}$, then μ receives quantum correction from the cut-off scale to electroweak scale as

$$\Delta\mu^2 = \frac{\lambda}{8\pi^2}\left(\Lambda_{\rm UV}^2 + 2\mu^2 \ln\left(\frac{\Lambda_{\rm UV}}{\mu}\right)\right) + \mathcal{O}\left(\lambda^2\right) \tag{1.2}$$

at 1-loop order. The other fermionic field contents never gives quadratic correction, while gauge bosons give similar quadratic corrections. Since there is no evidence of new physics up to $\mathcal{O}(1)$ TeV and renormalised μ is $\mathcal{O}(100)$ GeV, this correction should yield at least $\lesssim 1\%$ cancellation between bare μ^2 at the cut-off scale and the cut-off scale itself. Unless we accept such finely tuned cancellation, it would be natural to expect the dynamical origin of the hierarchy between the electroweak scale and cut-off scale and their fine-tuning. There are many other problems, such as the absence of neutrino mass terms in the standard model, naturalness problem of cosmological constant and feebleness of θ-term in quantum chromodynamics(QCD), but we would not describe those problems here.

The solutions for two problems aforementioned, in many cases, include the extension of particle contents and internal symmetry group. Such new particles or extra symmetry would provide additional scattering process in addition to those with standard model interactions of particle. Therefore the observation of new physics beyond the standard model in collider experiment could be done by searching for an excess on the data or finding deviation of parameters from the standard model prediction with precision measurements. This may not be an easy task in hadron collider since the collisions are initiated by particles with non-trivial colour charge. QCD is an asymptotically free theory at high energy, but at the same time, it has a non-perturbative regime in the infra-red(IR) regime. The propagation of any particle with colour charge

across relatively long-range scale $\gtrsim 1/\Lambda_{\mathrm{QCD}}(\sim 10^{-12}\mathrm{cm}$ with $\Lambda_{\mathrm{QCD}}\sim 300$ MeV) cannot be explained by the perturbative expansion of the standard model. Colour-charged particles accompany a large amount of emission and confinement in a long-range scale and become a bunch of baryons and mesons. Those fragmentation and hadronisation processes make identification of a particle or an event hard, and we could not isolate the scattering process of interest asides from the other scattering process exactly.

Yet there is a possibility to identify objects in the IR regime and reconstruct the hard scattering processes occurred in the perturbative regime. The objects are called as 'jets' which are bunches of hadrons clustered with specific algorithms. The understanding of jets is, therefore, one of the most important issues in collider physics to understand QCD itself and to reconstruct the hard scattering process. Recent studies using EFT [9–11] or Monte-Carlo simulation have lead progress in the understanding of jets and the IR physics through many observables, such as jet substructure variables [12–17] or event shape variables [18]. But a large number of the variables increases the complexity of statistical analysis of the data, and utilisation of all variables at once in a conventional way is nearly impossible in the practical sense.

On the other hand, machine learning(ML) has been acquired attention with its great performance and is frequently used in many fields based on the accumulated progress in computer science and hardware performance, and so as in the collider physics. In collider physics, many machine learning architectures such as boosted decision tree(BDT) or neural network(NN) have also been adopted for various analysis. For the collider physics, machine learning is mainly used for building discrimination models between signal and background events [19, 20], and also for regression [21] and event generation [22], too. Among various architecture and purpose of using, neural-network-based discrimination model is one of the most frequently adopted tool for the analysis. Also the limitation of the statistical analysis with a large number of

variables is alleviated with the machine learning since it has capability to express various functions and to abstract the given data.

It is natural to apply machine learning technique for research in collider physics if the final state contains a jet or jets in that sense. Jets are objects reconstructed from more than one particle, and they have large complexity. Using the limited number of variables describing jets may not contain the full information for the event reconstruction, and it is same for the shape of events with jets, too. For searches at hadron collider, however, it remain answered that how much it would be improved with machine learning, and which information has the largest impact for the machine learning algorithm for each search. To get quantitative answers to the questions, we studied two applications of machine learning in this thesis. The first one is an application to searching for composite vector resonance expected from minimal composite Higgs scenario. The hypothesised vector resonance in this example can have broad width compared to its mass($\Gamma/m \gtrsim \mathcal{O}(10)\%$), and conventional search strategy of a narrow-width signal may not be the optimal one for the search of broad width signal. We tried, thus, to expand the input features to cover more variables and checked the enhancement of the sensitivity. Then two methods are used to understand what had been learned by the machine. The first approach is using a certain measure to assess the impact of each variable to the final output from the neural network and the second one is planing away one information from the input data. The measure we used in the first case is called 'variable importance' and defined as the square-sum of weights between each input node to the nodes in the first hidden layer of the neural network. There is an issue that whether this raking does properly appraise the impact of respective input on the output from the neural network. The other hidden layers may have a strong impact on the output of the neural network compared to the first hidden layer, considering the complexity of the neural network. Thus it should be addressed first, whether the impact of the first hidden layer is larger compared to

the impact of the other hidden layers on the output. We checked the gradient of the loss function, which is the minimisation target of the ML algorithm, with respect to internal parameters of respective hidden layers during the training stage. Resulted gradient value at the first hidden layer was larger than the gradient values from the other hidden layers in all training in this application example. This is indirect evidence indicating that the impact of the first hidden layer on the output is stronger compared to the impact of the other hidden layers. The second method is 'planing'. Respective events are weighted to have a flat distribution of certain input variable. By this method, information corresponding to the variable planed away is intentionally washed out. If the degradation from this planing is significant, then we could conclude that the input variable planed away contains information crucial to the output from the neural network.

The second application is on the search for the invisible decay of Higgs. We started from the motivation that different event structure may produce different jets, quark jet or gluon jet as the initial state radiation(ISR). The difference between signal and background processes, such as the spin or interaction of the resonance particles, could be imprinted other than the final state. Let's consider the general searches of Higgs decay as an example. The decay of Z boson becomes background process in many decay channel of Higgs. However, there is no characteristic difference in the final state between decay from Higgs and from Z boson. The masses of Higgs and Z boson are similar to each other, and only the spins are different. But the difference of spin-correlation in the final state does not yield enough separation in the distribution of variables between signal and background processes. Even worse, there is no available information at all in the invisible final state. At the LHC, however, the leading(in transverse momentum ordering) ISR jet associated to Higgs production via gluon-fusion is soft gluon jet in the central region, while Z boson production carries relatively harder quark jet in the relatively forward region. So if we could

access to the further information of jets indicating that if the jet was originated from quark or gluon, we could exploit the information from ISR jet to various searches. We demonstrate this search strategy on the search for the invisible decay of Higgs. As a result, we checked the current upper limit on the ratio between combination of production cross section and production from standard model prediction with 95% confidence level can be improved a lot, in gluon-fusion production channel.

This thesis is organised as follow. In Chapter 2 we review the standard model and neural network first. In Chapter 3, we discuss the first example of an application of machine learning on the searching for the composite vector resonance from composite Higgs model. The second example with invisible decay of Higgs is demonstrated at Chapter 4. At last in Chapter 5, we conclude.

Chapter 2

Reviews on the Standard Model and Neural Network

In this section we shortly review the standard model at Section 2.1 and basic concept of neural network at Section 2.2. The mostly positive metric, i.e, $\eta_{mn} = \text{diag}(-1, 1, 1, 1)$ is used, and alphabets $m, n, \cdots$ indicate space-time indexes. If alphabetical indexes start from $a, b, \cdots$ then those components correspond to spatial indexes, only.

2.1 The Standard Model

In this section, we shortly review the standard model. The standard model is a gauge theory with the gauge group, $SU(3)_C \times SU(2)_L \times U(1)_Y$ [23]. The subscripts C, L and Y mean that respective group is internal symmetry group of colour(C), left-handed fermions(L) and hyper-charge(Y). Precisely, $SU(2)_L$ is the symmetry group of chiral spinor, rather than left-handed fermions. However, for the case of massless

representation of Poincaré group, of which $P^2 = 0$, the chirality coincides with the helicity(or handedness) of spinors. It manifests in Weyl basis, in which γ-matrices are written as

$$\gamma^m = \begin{pmatrix} 0 & \sigma^m \\ \bar{\sigma}^m & 0 \end{pmatrix} \tag{2.1}$$

with $\sigma^m = (\mathbb{1}_{2\times 2}, \vec{\sigma})$ and $\bar{\sigma}^m = (\mathbb{1}_{2\times 2}, -\vec{\sigma})$. In this basis, the representation of Lorentz group generators are $\gamma^{mn} \equiv \frac{i}{4}[\gamma^m, \gamma^n]$, and already diagonalised as

$$\gamma^{mn} = \begin{pmatrix} \sigma^{mn} & 0 \\ 0 & \bar{\sigma}^{mn} \end{pmatrix}, \tag{2.2}$$

where,

$$\sigma^{mn} = \frac{i}{4}(\sigma^m\bar{\sigma}^n - \sigma^n\bar{\sigma}^m) \quad \text{and} \quad \bar{\sigma}^{mn} = \frac{i}{4}(\bar{\sigma}^m\sigma^n - \bar{\sigma}^n\sigma^m). \tag{2.3}$$

Remind that chirality is defined whether the Weyl spinor, two-component spinor, is transformed by σ^{mn} or $\bar{\sigma}^{mn}$; one transformed by σ^{mn} is called chiral spinor, and the other one is anti-chiral spinor. In this basis, therefore, a general Dirac spinor Ψ is written with chiral spinor ψ and anti-chiral spinor $\bar{\chi}$ as,

$$\Psi = \begin{pmatrix} \psi_\alpha \\ \bar{\chi}^{\dot{\alpha}} \end{pmatrix}. \tag{2.4}$$

And we have chiral projection operators, $P_{L,R}$, which are,

$$P_L = \frac{1}{2}(1 - \gamma^5) \quad \text{and} \quad P_R = \frac{1}{2}(1 + \gamma^5), \tag{2.5}$$

with $\gamma^5 = \frac{\sqrt{\det\eta}}{4!}\epsilon_{mnlo}\gamma^m\gamma^n\gamma^l\gamma^o$. In Weyl-basis they are,

$$P_L = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad P_R = \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1} \end{pmatrix}. \tag{2.6}$$

Now let's consider the helicity of massless spinors. In massless case, each component satisfies following respective free field equation of motion,

$$\begin{aligned} 0 &= p_m\,(\bar{\sigma}^m)^{\dot{\alpha}\alpha}\,\psi_\alpha \\ 0 &= p_m\,(\sigma^m)_{\alpha\dot{\alpha}}\,\bar{\chi}^{\dot{\alpha}}. \end{aligned} \tag{2.7}$$

For on-shell field ψ_α,

$$0 = (p_0 - \vec{p}\cdot\vec{\sigma})\,\psi = -\,|\vec{p}|\,(1 + \hat{p}\cdot\vec{\sigma})\psi \tag{2.8}$$

with $\hat{p} \equiv \vec{p}/|\vec{p}|$ and it is clearly a left-handed spinor. At the same time, $\bar{\chi}$ is a right-handed spinor, because

$$0 = |\vec{p}|\,(-1 + \hat{p}\cdot\vec{\sigma})\bar{\chi}. \tag{2.9}$$

This shows that the chirality does coincide with the helicity as their eigenspaces are the same. And, before the spontaneous electro-weak symmetry all spinors are massless, so any terms can be used, helicity(handedness) or chirality.

Returning to the standard model, the particle contents should be clarified first to construct the Lagrangian. There are 5 kinds of Weyl spinors, depending on the charge under the internal symmetry group. Three from those five Weyl fermions have colour charge as fundamental representation, and the other two are singlets under $SU(3)_C$. Among colour charged fermions, one is a doublet in $SU(2)_L$, and the other two are singlets. The other two Weyl spinors without charge under $SU(3)_C$ are again classified by the charge under $SU(2)_L$; one is a doublet and the other one is a singlet. And, each of them has 3 copies corresponding to three generations.

Note that the dimensionality of $SU(N)$ is $N^2 - 1$ and $U(N)$ is N^2 dimensional group. Therefore, from three gauge group(local symmetry group), we have $(3^2 - 1) + (2^2 - 1) + 1^2$ massless dynamical gauge bosons. Among 12 gauge fields, 11 of them come from non-Abelian groups, $SU(3)_C$ and $SU(2)_L$ and the left one comes from Abelian group $U(1)_Y$. In addition to that, we have four more bosonic degrees of freedom. Those are in a $SU(2)_L$ doublet with two complex scalar fields. The full list of fermionic field contents are summarised in the Table 2.1 and bosonic fields are in Table 2.2.

Name	Notation	Poincaré	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
left-handed quark	Q_L^i	$(\mathbf{2}\ ,\ \mathbf{1})$	$\mathbf{3}$	$\mathbf{2}$	$\frac{1}{6}$
up-type quark	u_R^i	$(\mathbf{2}\ ,\ \mathbf{1})$	$\bar{\mathbf{3}}$	$\mathbf{1}$	$-\frac{2}{3}$
down-type quark	d_R^i	$(\mathbf{2}\ ,\ \mathbf{1})$	$\bar{\mathbf{3}}$	$\mathbf{1}$	$\frac{1}{3}$
left-handed lepton	L_L^i	$(\mathbf{2}\ ,\ \mathbf{1})$	$\mathbf{1}$	$\mathbf{2}$	$-\frac{1}{2}$
right-handed lepton	e_R^i	$(\mathbf{2}\ ,\ \mathbf{1})$	$\mathbf{1}$	$\mathbf{1}$	1

Table 2.1 The particle contents of fermionic degrees of freedom of the standard model. Indexes i label the generation, and three generation of Weyl fermions had been discovered for each fields. Here Q_L and L_L consist with two Weyl spinors, so that $Q_L^i = \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix}$ and $L_L^i = \begin{pmatrix} \nu^i \\ e_L^i \end{pmatrix}$. Since no right-handed components of neutrino fields have been discovered yet, the neutrino fields do not need labels to be specified. Here all fermions are written as they are left-handed Weyl spinors.

Name	Notation	Poincaré	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
Higgs	Φ	$(\mathbf{1}\ ,\ \mathbf{1})$	$\mathbf{1}$	$\mathbf{2}$	$-\frac{1}{2}$
Gluon	$\{G_m^a\}_{a=1,\cdots,8}$	$(\mathbf{2}\ ,\ \mathbf{2})$	$\mathbf{8}$	$\mathbf{1}$	0
Weak-gauge boson	$\{W_m^i\}_{i=1,2,3}$	$(\mathbf{2}\ ,\ \mathbf{2})$	$\mathbf{1}$	$\mathbf{3}$	0
Hyper-charge gauge boson	B_m	$(\mathbf{2}\ ,\ \mathbf{2})$	$\mathbf{1}$	$\mathbf{1}$	0

Table 2.2 The particle contents of bosonic degrees of freedom of the standard model. Here a and i labels the generators of respective symmetry group $SU(3)_C$ and $SU(2)_L$.

The full Lagrangian of the standard model, say $\mathcal{L}_{\rm SM}$, can be decomposed as follow.

$$\mathcal{L}_{\rm SM} = \mathcal{L}_{\rm gauge} + \mathcal{L}_{\rm fermion} + \mathcal{L}_{\rm Higgs} + \mathcal{L}_{\rm Yukawa} + \mathcal{L}_{\rm g.f.} + \mathcal{L}_{\rm ghost} \tag{2.10}$$

Each term represents the gauge kinetic terms, fermion kinetic terms, Higgs kinetic term and its potential, Yukawa terms, gauge fixing terms(g.f.) and the terms from associated ghosts to the gauge fixing. It contains all operators with dimension not higher than 4 except three boundary terms from gauge sector.

Let's discuss from the first term, the gauge kinetic term. We have two non-Abelian gauge fields and one Abelian gauge field as in Table 2.2. For the general non-Abelian gauge field A^a_m, the field strength tensor is defined as the commutation between covariant derivatives,

$$F_{mn} = i[D_m, D_n] = i[\partial_m - iA_m, \partial_n - iA_n], \tag{2.11}$$

where $D_m = \partial_m - iA_m$ is covariant derivative, $A_m \equiv gA^a_m T_a$ with corresponding coupling g and generators $\{T_a\}_{a=1,\cdots,\dim\rho[G]}$ of representation ρ of gauge group G for coupled source. In component-wise expression for $F_{mn} \equiv gF^a_{mn}T_a$,

$$F^a_{mn} = \partial_m A^a_n - \partial_n A^a_m + gA^b_m A^c_n f_{bc}{}^a. \tag{2.12}$$

Here $f_{bc}{}^a$ is fully anti-symmetric structure constant of the gauge group. The canonically normalised kinetic terms would be

$$-\frac{1}{4g^2 D_\rho}\mathrm{tr}F_{mn}F^{mn}, \tag{2.13}$$

with $F_{mn} = gF^a_{mn}T_a$ and Dynkin index D_ρ of source field representation ρ. In standard model, all matter fields are in fundamental representations, thus $D_\rho = \frac{1}{2}$. For each components, therefore,

$$\begin{aligned} -\frac{1}{4}F^a_{mn}F^{mn}_a &= -\frac{1}{2}\partial_m A^a_n\left(\partial^m A^n_a - \partial^n A^m_a\right) \\ &\quad -gf_{abc}\partial_m A^a_n A^{b\,m}A^{c\,n} \\ &\quad -\frac{g^2}{4}f^a{}_{bc}f_{aef}A^b_m A^c_n A^{e\,m}A^{f\,n} \end{aligned} \tag{2.14}$$

The only difference of Abelian gauge field B_m is triviality of the representation, and

$$-\frac{1}{4}B_{mn}B^{mn} = -\frac{g_1^2}{2}\partial_m B_n(\partial^m B^n - \partial^n B^m), \tag{2.15}$$

with $B_{mn} = g_1(\partial_m B_n - \partial_n B_m)$.

Combining them all, the gauge kinetic terms in the standard model Lagrangian are,

$$\mathcal{L}_{\text{guage}} = -\frac{1}{2g_3^2}\text{tr}G_{mn}G^{mn} - \frac{1}{2g_2^2}\text{tr}W_{mn}W^{mn} - \frac{1}{4g_1^2}B_{mn}B^{mn}, \tag{2.16}$$

with couplings g_3, g_2 and g_1 of respective $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$.

The second term, $\mathcal{L}_{\text{fermion}}$, the kinetic terms of fermionic degrees of freedom as in Table 2.1 consists of $\bar{\psi}\not{D}\psi$ like terms. The covariant derivative depends on the charge of given fermion field as,

$$\begin{aligned}
\bar{Q}_L\not{D}Q_L &= \bar{Q}_L\bar{\sigma}^m(\partial_m - i\frac{g_3}{2}G_m^a\lambda_a - i\frac{g_2}{2}W_m^i\tau_i - i\frac{g_1}{6}B_m)Q_L \\
u_R\not{D}\bar{u}_R &= u_R\sigma^m(\partial_m - i\frac{g_3}{2}G_m^a\lambda_a - i\frac{2g_1}{3}B_m)\bar{u}_R \\
d_R\not{D}\bar{d}_R &= d_R\sigma^m(\partial_m - i\frac{g_3}{2}G_m^a\lambda_a + i\frac{g_1}{3}B_m)\bar{d}_R
\end{aligned} \tag{2.17}$$

for quark sector, and

$$\begin{aligned}
\bar{L}_L\not{D}L_L &= \bar{L}_L\bar{\sigma}^m(\partial_m - i\frac{g_2}{2}W_m^i\tau_i + i\frac{g_1}{2}B_m)L_L \\
e_R\not{D}\bar{e}_R &= e_R\sigma^m(\partial_m + ig_1B_m)\bar{e}_R,
\end{aligned} \tag{2.18}$$

for the lepton sector. Here $\{\lambda_a\}_{a=1,\cdots,8}$ are Gell-Mann matrices and $\{\tau_i\}_{i=1,2,3}$ are Pauli matrices. Note that the all right-handed components can be written in left-handed convention as, for example of e_R,

$$e_R\not{D}\bar{e}_R = \bar{e}_R\bar{\sigma}^m(\partial_m - ig_1B_m)e_R - \partial_m(\bar{e}_R\bar{\sigma}^m e_R). \tag{2.19}$$

This can be easily checked that as follow.

$$e_R\sigma^m\bar{e}_R = e_R^\alpha\sigma_{\alpha\dot{\alpha}}^m\bar{e}_R^{\dot{\alpha}} = -\epsilon^{\dot{\alpha}\dot{\beta}}\bar{e}_{R\dot{\beta}}\sigma_{\alpha\dot{\alpha}}^m\epsilon^{\alpha\beta}e_{R\beta} = -\bar{e}_R\bar{\sigma}^m e_R. \tag{2.20}$$

Remind that $\bar{e}_R$ is right-handed and e_R is left-handed, so that $\bar{e}_R$ carries dotted index, and here we used the identity,

$$\epsilon^{\dot{\alpha}\dot{\beta}}\epsilon^{\alpha\beta}\sigma^m_{\beta\dot{\beta}} = \bar{\sigma}^{m\dot{\alpha}\alpha}. \tag{2.21}$$

Combining them all, the fermion kinetic terms $\mathcal{L}_{\text{fermion}}$ is

$$\mathcal{L}_{\text{fermion}} = \bar{Q}_L \not{D} Q_L + u_R \not{D} \bar{u}_R + d_R \not{D} \bar{d}_R + \bar{L}_L \not{D} L_L + e_R \not{D} \bar{e}_R, \tag{2.22}$$

and covariant derivatives are as in eq. (2.17) and eq. (2.18).

The next term, Higgs sector Lagrangian $\mathcal{L}_{\text{Higgs}}$ is written as,

$$\mathcal{L}_{\text{Higgs}} = -D^m\Phi^\dagger D_m\Phi + \mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2, \tag{2.23}$$

with positive μ^2 and the covariant derivative,

$$D_m\Phi = (\partial_m - i\frac{g_2}{2}W^i_m\tau_i + i\frac{g_1}{2}B_m)\Phi. \tag{2.24}$$

The Higgs sector Lagrangian contains $+\mu^2\Phi^\dagger\Phi$ with positive μ^2, hence its potential around $\Phi \simeq 0$ has negative curvature. Furthermore, the global minima of the potential is not at $\Phi = 0$, but $|\Phi|^2 = \mu^2/(2\lambda)$. This two characteristics, negative curvature at the origin and global minima away from the origin, make theory expanded around $\Phi = 0$ unstable at the classical level. The tachyonic state is not regulated, but exponentially decay or growth in on-shell solution. The global minima away from the origin yield non-zero vacuum expectation value of the field Φ. Non-zero vacuum expectation value of given field makes creation operator to produce linear combination of single particles and the ground, so that the Lehmann-Symanzik-Zimmermann(LSZ) reduction formula is not-well defined. Therefore, the field Φ has to be redefined around the true minima. After the redefinition of Φ to have null expectation value from vacua, the field excitation orthogonal to the ground state configuration seems not to respect the full symmetry $SU(2)_L \times U(1)_Y$, but only has $U(1)$ symmetry. This is called spontaneous symmetry breaking.

There are three real modulus for constant-$|\Phi|$ space. If we write, with $v = \mu/\sqrt{\lambda}$, we can parametrise Φ to manifest the modulus as,

$$\Phi \equiv \frac{1}{\sqrt{2}}(v+H)\chi(\xi,\theta,\varphi) \equiv \frac{1}{\sqrt{2}}(v+H)e^{i\xi}\begin{pmatrix}\cos\frac{\theta}{2}\\ -\sin\frac{\theta}{2}e^{i\varphi}\end{pmatrix}. \tag{2.25}$$

Here only the H field can change the absolute value of Φ, and θ, φ, ξ never change neither $|\Phi|$ nor the potential value. In this parametrisation, the covariant derivatives is,

$$D_m\Phi = \frac{1}{\sqrt{2}}\left(\partial_m H\cdot\chi + (v+H)\partial_m\chi - i(v+H)A_m^{(g)}\chi\right), \tag{2.26}$$

with abbreviates $A_m^{(g)}$, the summation of all gauge fields, and eq. (2.23) becomes,

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} &= -\frac{1}{2}\partial_m H\partial^m H - \frac{\lambda}{4}(2vH+H^2)^2 + \frac{\lambda}{4}v^4 \\ &\quad -\frac{1}{2}(v+H)^2\partial_m\chi^\dagger\partial^m\chi + (v+H)^2\chi^\dagger A_m^{(g)}A^{(g)m}\chi \\ &\quad +i(v+H)^2\left(\chi^\dagger A_m^{(g)}\partial^m\chi - \partial^m\chi^\dagger A_m^{(g)}\chi\right).\end{aligned} \tag{2.27}$$

This is merely redefinition of fields, so that there still is $SU(2)_L\times U(1)_Y$ symmetry. The only difference is that the gauge transform now becomes non-linear in fields due to the redefinition, and this is the reason why it is called as 'spontaneous' symmetry breaking.

Those angular fields ξ, θ and φ are said to be Nambu-Goldstone bosons [24, 25]. In certain gauge choice of $SU(2)_L\times U(1)_Y$, which is called, 'unitary gauge', those Nambu-Goldstone bosons are absorbed by $W_m^{1,2}$ and one linear combination of W_m^3 and B_m. For this, we rather re-parametrise Φ in Cartesian-like form, rather than as in eq. (2.25), as

$$\Phi = \begin{pmatrix}\frac{1}{\sqrt{2}}(v+H(x)+i\varphi^Z(x))\\ \varphi^-(x)\end{pmatrix}, \tag{2.28}$$

with $\varphi^-(x)\in\mathbb{C}$ and $H(x),\varphi^Z(x)\in\mathbb{R}$. Put aside the details until the discussion about the gauge fixing terms, let us assume that it is possible to fix the gauge in

which φ^- and φ^Z do not appear, after the spontaneous symmetry breaking. Then the covariant derivative eq. (2.24) becomes,

$$D_m\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} \partial_m H - \frac{i}{2}(g_2 W_m^3 - g_1 B_m)(v+H) \\ -i\frac{g_2}{\sqrt{2}} W_m^-(v+H) \end{pmatrix}, \tag{2.29}$$

and corresponding kinetic term would be as follow.

$$\begin{aligned} -D_m\Phi^\dagger D^m\Phi &= -\frac{1}{2}(\partial H)^2 - \frac{g_2^2}{4} W_m^- W^{+m}(v+H)^2 \\ &\quad -\frac{1}{8}(g_2 W_m^3 - g_1 B_m)(g_2 W^{3m} - g_1 B^m)(v+H)^2 \end{aligned} \tag{2.30}$$

Here we defined $W_m^\pm = \frac{1}{\sqrt{2}}(W_m^1 \mp W_m^2)$. From eq. (2.30), it seems natural to redefine gauge fields (W_m^3, B_m) in to another basis, (Z_m, A_m), so that Z_m solely receives effect of v^2. The corresponding transform which keep unitarity is as follow,

$$\begin{pmatrix} Z_m \\ A_m \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_m^3 \\ B_m \end{pmatrix}, \tag{2.31}$$

where $\tan\theta_W = g_1/g_2$. In this basis, eq. (2.30) becomes simple as,

$$-\frac{1}{2}(\partial H)^2 - \frac{g_2^2}{4} W_m^- W^{+m}(v+H)^2 - \frac{g_2^2+g_1^2}{8} Z_m Z^m (v+H)^2, \tag{2.32}$$

and induce the tree-level gauge boson masses,

$$m_W = \frac{1}{2} g_2 v \quad \text{and} \quad m_Z = \frac{\sqrt{g_2^2+g_1^2}}{2} v = \frac{1}{2\cos\theta_W} g_2 v, \tag{2.33}$$

for respective $W_m^\pm$ and Z_m.

Since we changed basis from (W_m^3, B_m) to (Z_m, A_m), it is natural to redefine the corresponding couplings, too. Let's consider a single component field ψ with $\frac{\tau_3}{2}\psi = T_3^\psi \psi$ and Y^ψ charge under $U(1)_Y$. The reason why we don't need to know eigenvalue of $\tau_{1,2}/2$ is because that, the pattern of field redefinition only touches $\tau_3/2$ component and identity components. For this field ψ,

$$\begin{aligned} T_3^\psi g_2 W_m^3 + Y^\psi g_1 B_m &= (T_3^\psi g_2 \cos\theta_W - Y^\psi g_1 \sin\theta_W) Z_m \\ &\quad + (T_3^\psi g_2 \sin\theta_W + Y^\psi g_1 \cos\theta_W) A_m. \end{aligned} \tag{2.34}$$

Note that, from $\tan\theta_W = g_1/g_2$, we have $g_2 \sin\theta_W = g_1 \cos\theta_W$. If we set $e \equiv g_2 \sin\theta_W = g_1 \cos\theta_W$, then the term corresponding to A_m and Z_m would be,

$$T_3^\psi g_2 W_m^3 + Y^\psi g_1 B_m = \frac{g_2}{\cos\theta_W}(T_3^\psi - Q^\psi \sin^2\theta_W) Z_m + Q^\psi e A_m, \tag{2.35}$$

where $Q^\psi \equiv T_3^\psi + Y^\psi$. After the spontaneous symmetry breaking with eq. (2.28), there are Dirac spinors which consists of a pair of left- and right-handed Weyl spinors, rather than respective two Weyl spinors. Those Dirac spinors are interaction eigenstates for unbroken $U(1)_{\rm em}$ gauge group with A_m, i.e, left- and right-handed Weyl spinors have the same Q. Therefore, the Dirac spinor Ψ with left-handed component ψ and right-handed component $\bar{\chi}$ has following interaction withg Z_m,

$$\begin{aligned} Z_m \frac{g_2}{\cos\theta_W}(T_3^\psi - Q^\psi \sin^2\theta_W)(\bar{\sigma}^m \psi + \sigma^m \bar{\chi}) \\ \rightarrow Z_m \frac{g_2}{\cos\theta_W}(T_3^\psi - Q^\psi \sin^2\theta_W)\gamma^m (P_L \Psi + P_R \Psi) \\ = Z_m (g_V^\psi \gamma^m - g_A^\psi \gamma^m \gamma^5)\Psi \end{aligned} \tag{2.36}$$

with

$$g_V = \frac{g_2}{\cos\theta_W}\left(\frac{1}{2}T_3 - Q\sin^2\theta_W\right) \quad \text{and} \quad g_A = \frac{g_2}{2\cos\theta_W} T_3. \tag{2.37}$$

The other interaction terms with $W_m^\pm$ are relevant only for left-handed fermions. To discuss the picture, let's denote upper component of the arbitrary left-handed fermions as ψ_+ and the lower component as ψ_-. Then the following 'charged-current interaction' terms, such that

$$-i\frac{g_2}{\sqrt{2}}\begin{pmatrix}\bar{\psi}_+ & \bar{\psi}_-\end{pmatrix}\begin{pmatrix}0 & W_m^+ \\ W_m^- & 0\end{pmatrix}\bar{\sigma}^m \begin{pmatrix}\psi_+ \\ \psi_-\end{pmatrix}, \tag{2.38}$$

could be written in terms of Dirac spinors $\Psi_\pm = {}^{\rm t}(\psi_\pm \ \bar{\psi}_\pm)$ corresponding to respective $\psi_\pm$ as,

$$-i\frac{g_2}{\sqrt{2}}(W_m^- J_+^m + W_m^+ J_-^m) \tag{2.39}$$

with the current operators $J^m_\pm$,

$$J^m_\pm = \overline{\Psi}_\mp \gamma^m P_L \Psi_\pm. \tag{2.40}$$

Now we discuss about $\mathcal{L}_{\text{Yukawa}}$. Yukawa coupling terms include all possible charge-neutral dimension-4 operators made by Φ and fermions as follow.

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}} &= -Y^{(u)I}{}_J(\Phi^\dagger)^\alpha(Q_{L\,\alpha})_I u^J_R - Y^{(d)I}{}_J\epsilon^{\alpha\beta}\Phi_\alpha(Q_{L\,\beta})_I d^J_R \\ &\quad -Y^{(\ell)I}{}_J\epsilon^{\alpha\beta}\Phi_\alpha(L_{L\,\beta})_I e^J_R + (\text{h.c.})\end{aligned} \tag{2.41}$$

Here (h.c.) means Hermitian conjugates of all explicitly shown terms, and I, J are generation indexes, while α and β are $SU(2)_L$ indexes. Note that three Yukawa matrices $Y^{(u)}$, $Y^{(d)}$, and $Y^{(\ell)}$ do not have to be neither unitary or Hermitian matrices. The reality condition of Lagrangian is automatically respected by adding Hermitian conjugates of explicitly shown terms as in eq. (2.41). Hence the Yukawa matrices are general $N_{\text{gen}} \times N_{\text{gen}}$ complex matrices having $2N^2_{\text{gen}}$ real degrees of freedom with the number of generation, N_{gen}. Still the Yukawa matrices can be diagonalised via bi-unitary transforms as, for example of up-type Yukawa matrix,

$$Y^{(u)} \mapsto U^\dagger_L Y^{(u)} U_R, \tag{2.42}$$

by the following global transform,

$$u_{L\,I} \mapsto u_{L\,J}(U_L)^J{}_I \quad \text{and} \quad \bar{u}_{R\,I} \mapsto \bar{u}_{R\,J}(U_R)^J{}_I. \tag{2.43}$$

The similar things happen for $Y^{(d)}$, the down-type Yukawa matrix with unitary matrices D_L and D_R for respective d_L and d_R quarks. Among four unitary matrices, the global transforms from U_R and D_R do not change anything. The two unitary matrices rotate the kinetic terms of right-handed quarks u_R and d_R as because,

$$\delta^J{}_I \psi^I_R \sigma^m D_m \bar{\psi}_{R\,J} \mapsto (V_R)^N{}_J \delta^J{}_I (V^\dagger_R)^I{}_M \psi^M_R \sigma^m D_m \bar{\psi}_{R\,N}, \tag{2.44}$$

and $V_R V_R^\dagger$=1, for $\psi_R = u_R$ or d_R and $V = U$ or D. Since there is no other term at which right-handed quarks appear, the U_R and D_R only diagonalise respective Yukawa matrices and cancelled in the kinetic terms of respective right-handed quarks. This is not true for left-handed quarks. The left-handed quarks with different generation labels(or flavours) are mixed through the charged current interaction which is eq. (2.38). For the generalised current operator $(J^m_\pm)^I{}_J$ such that,

$$(J^m_\pm)^I{}_J = \bar{\psi}^I_\pm \bar{\sigma}^m \psi_{\mp\, J} = \overline{\Psi}^I_\mp \gamma^m P_L \Psi_{\pm J}, \tag{2.45}$$

the interaction terms in eq. (2.38) change as,

$$\begin{aligned} -i\frac{g_2}{\sqrt{2}}(W^-_m (J^m_+)^I{}_J + W^+_m (J^{m-})^I{}_J)\delta^J{}_I \qquad & (2.46) \\ \mapsto \quad -i\frac{g_2}{\sqrt{2}}(W^-_m (J^m_+)^I{}_J V^J{}_I + W^+_m (J^m_-)^I{}_J (V^\dagger)^J{}_I), & \end{aligned}$$

with $V^I{}_J = (U_L D_L^\dagger)^I{}_J$. The matrix V is called Cabibbo-Kobayashi-Maskawa (CKM) matrix [26,27]. Since the CKM matrix is also an unitary matrix, it includes $\frac{1}{2}N_{\rm gen}(N_{\rm gen}-1)$ angles from $SO(N_{\rm gen})$, and the others are complex phases with the number of $\frac{1}{2}N_{\rm gen}(N_{\rm gen}+1)$ parameters. Among the complex phases, $(2N_{\rm gen}-1)$ can be absorbed by relative phase difference between u^I_L and d^J_L while the Yukawa matrices are kept to be diagonal. Therefore, CKM matrix has total $\frac{1}{2}N_{\rm gen}(N_{\rm gen}-1)$ angles and,

$$\frac{1}{2}N_{\rm gen}(N_{\rm gen}+1) - (2N_{\rm gen}-1) = \frac{1}{2}(N_{\rm gen}-1)(N_{\rm gen}-2) \tag{2.47}$$

complex phases relevant for observables. The standard model, there are 3 generations, so that 3 angles and 1 complex phase parametrise the CKM matrix. There is no such complication for the leptons, since there is only one Yukawa matrix, $Y^{(\ell)}$. Since there is no Yukawa with right-handed neutrinos, $Y^{(\ell)}$ can be diagonalised without mixes left-handed neutrinos and charged leptons.

Now, let's discuss the gauge fixing terms, $\mathcal{L}_{\rm g.f.}$ and ghosts terms $\mathcal{L}_{\rm ghost}$ together. Without the gauge fixing terms, the path integral is not be well-defined since it sums

up the infinitely many configuration which are equivalent via gauge transform. It makes the propagator of the gauge field non-invertible. To make the path integral well-defined, we need to normalise the total contribution from the equivalent field configuration in path integral up-to gauge transform. It does not mean that the gauge should be fixed strictly with a specific choice, but just the total the contribution from physically equivalent configuration should be normalised. Hence, the following term is inserted on the total action $S_{\rm SM}$ of the standard model,

$$S_{\rm SM} \mapsto S_{\rm SM} + i\sum_G \int d^dx \frac{1}{\xi} {\rm tr}\omega_G\omega_G, \tag{2.48}$$

where $\omega_G \equiv \omega_G^a T_a$ is gauge fixing function of group G with coupling g. The ω_G would be eventually replaced by arbitrary gauge fixing functional $\mathcal{F}(A)$. For example, if ω is identified with $\mathcal{F}(a) = \partial_m A^m$ in $\xi \to 0$ limit, the resulted path integral only allows field configuration in Lorentz gauge, in which $\partial_m A^m = 0$. For the general path integral $\mathcal{Z}$, this can be done by δ-functional in path integral as,

$$\mathcal{Z} = \int \mathcal{D}\psi\mathcal{D}\omega\mathcal{D}\theta\delta[\mathcal{F}(A) - \omega]e^{iS}. \tag{2.49}$$

Here $\mathcal{D}\psi$ is abbreviation of the functional integration measure of all possible field contents of the theory, and ω is for gauge fixing terms as in eq. (2.48). θ is a set of gauge transform parameters. The functional integral over ω will copy the gauge fixing condition $\mathcal{F}(A)$ from the δ-function, and the integral over θ makes the gauge field to satisfy the gauge fixing condition. This can be again generalised using δ-function identity as

$$\mathcal{Z} = \int \mathcal{D}\psi\mathcal{D}\omega\mathcal{D}\theta {\rm det}\left(\frac{\delta G}{\delta\theta}\right)\delta[G(A)]e^{iS}, \tag{2.50}$$

with $G(A) = \mathcal{F}(A) - \omega$. The Jacobian determinant, which is called as the Faddeev-Popov determinant, can be lifted up into the path integral by introducing two Grassmann fields c^a and $\bar{c}^a$ with in the same representation space where gauge fields belong.

$${\rm det}\left(\frac{\delta G}{\delta\theta}\right) \propto \int \mathcal{D}\bar{c}\,\mathcal{D}c\, e^{-i\int d^dx \bar{c}\frac{\partial\omega}{\partial\theta}c}\Big|_{\omega=\mathcal{F}(A)} \tag{2.51}$$

The factor i in front of the exponential in the integrand only yields the normalisation factor so it does not change the result. Note that here $\bar{c}$ is not Hermitian conjugate of c. They just share the similar labels for the field only for the notational simplicity.

The conventional choice would be Lorentz-like gauge, which is called R_ξ gauge. The gauge fixing terms in R_ξ-gauge for the symmetry generators are introduced in the same manner with the Lorentz gauge. For the spontaneously broken symmetry generators, however, it would carries additional terms to absorb the Nambu-Goldstone boson. To manifest this picture, let's consider the toy example with spontaneous symmetry breaking of $U(1)$ symmetry with following Lagrangian $\mathcal{L}$.

$$\mathcal{L} = -\frac{1}{4e^2}F_{mn}F^{mn} - D_m\phi^* D^m\phi - \frac{\lambda}{4}\left(|\phi|^2 - \frac{v^2}{2}\right)^2 \tag{2.52}$$

with complex scalar field ϕ and Abelian gauge field A_m and associated field strength tensor, covariant derivative and charge e of ϕ under $U(1)$. Then reparametrisation $\phi \equiv \frac{1}{\sqrt{2}}(v + h + i\varphi)$ with gauge fixing term $\omega = \mathcal{F}(A)$ such that,

$$\mathcal{F}(A) = \partial_m A^m - \xi e v \varphi \tag{2.53}$$

yields $\mathcal{L}$ as follow.

$$\begin{aligned}
\mathcal{L} \;=\; & -\frac{1}{2}A_m\left((\partial^2 - e^2v^2)\eta^{mn} - \left(1 - \frac{1}{\xi}\right)\partial^m\partial^n\right)A_n && (2.54)\\
& -\frac{1}{2}(\partial h)^2 - \frac{1}{2}(\partial\varphi)^2 - \frac{\lambda}{16}(2vh + h^2 + \varphi^2)^2 - \frac{\xi}{2}e^2v^2\varphi^2 \\
& -e\varphi(2\partial^m h A_m + h\partial^m A_m) - \frac{1}{2}e^2(2vh + h^2 + \varphi^2)A_m A^m \\
& -\bar{c}(\partial^2 - \xi e^2 v(v + h))c && (2.55)
\end{aligned}$$

If the $U(1)$ was not broken, then the ghost would be decoupled from the other fields, but due to additional gauge fixing condition it is still coupled to h field. Furthermore, even though the fields similar with φ, the Nambu-Goldstone boson, never has the mass term as in eq. (2.27), here it becomes massive with $m_\varphi^2 = \xi e^2 v^2$. It is easy to

find the origin, the gauge fixing term, as it has ξ. The gauge fixing term with non-zero ξ accounts all contribution from field configuration which does not respect the strict gauge fixing condition, $\omega = 0$. With $\xi \to 0$ limit, however, only the field configuration which respect the gauge fixing condition contributes. This manifests as $m_\varphi^2 = 0$ in this limit. We, therefore, call this specific case, R_ξ-gauge in $\xi \to 0$ limit, as unitary gauge.

The similar things happen in the standard model. As eq. (2.50), the standard model includes gauge fixing terms $\mathcal{L}_{\text{g.f.}}$ as

$$\mathcal{L}_{\text{g.f.}} = \frac{1}{\xi'}\text{tr}\omega_C\omega_C + \frac{1}{\xi}\text{tr}\omega_L\omega_L + \frac{1}{2\xi}\omega_Y\omega_Y, \tag{2.56}$$

where the auxiliary filed are exactly solved as $\omega_C^a = \mathcal{F}_C^a$, $\omega_L^i = \mathcal{F}_L^i$ and $\omega_Y = \mathcal{F}_Y$ with

$$\begin{aligned} \mathcal{F}_C^a(G) &= \partial_m G^{a\,m} \\ \mathcal{F}_L^i(W) &= \partial_m W^{i\,m} - \frac{\xi}{2} g_2 v \varphi^i \\ \mathcal{F}_Y(B) &= \partial_m B^m - \frac{\xi}{2} g_1 v \varphi^3. \end{aligned} \tag{2.57}$$

Here φ^i's are Nambu-Goldstone bosons, such that,

$$\varphi^- \equiv -\frac{i}{\sqrt{2}}(\varphi^1 + i\varphi^2) \quad \text{and} \quad \varphi^3 \equiv \varphi^Z, \tag{2.58}$$

and φ^- and φ^Z parametrise Φ as in eq. (2.28). The gauge fixing terms for W_m^i and B_m can be re-written in terms of $W_m^\pm$, Z_m and A_m as,

$$\begin{aligned} \mathcal{F}^\pm(W) &= \partial_m W^{\pm m} \mp i\xi m_W \varphi^\pm \\ \mathcal{F}^Z(Z) &= \partial_m Z^m - \xi m_Z \varphi^Z \\ \mathcal{F}^A(A) &= \partial_m A^m, \end{aligned} \tag{2.59}$$

where $\mathcal{F}^\pm = \frac{1}{\sqrt{2}}(\mathcal{F}_L^1 \mp i\mathcal{F}_L^2)$ and

$$\begin{pmatrix} \mathcal{F}^Z \\ \mathcal{F}^A \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} \mathcal{F}_L^3 \\ \mathcal{F}_Y \end{pmatrix}. \tag{2.60}$$

For the masses of gauge bosons, please refer to eq. (2.33). The eq. (2.59) more manifest the symmetry breaking pattern of the standard model compared to eq. (2.59). In this basis, the gauge fixing terms $\mathcal{L}_{\text{g.f.}}$ from eq. (2.56) will be given by,

$$\mathcal{L}_{\text{g.f.}} = \frac{1}{2\xi'}\mathcal{F}_C^a\mathcal{F}_C^b\delta_{ab} + \frac{1}{\xi}\mathcal{F}^-\mathcal{F}^+ + \frac{1}{2\xi}\mathcal{F}^Z\mathcal{F}^Z + \frac{1}{2\xi}\mathcal{F}^A\mathcal{F}^A. \tag{2.61}$$

The Faddeev-Popov determinants from gauge fixing terms induce total 12 ghost fields and 12 anti-ghost fields. 8 of each, say $\bar{c}_a$ and c^b, account to the $SU(3)_C$ symmetry, and the left correspond to $SU(2)_L \times U(1)_Y$. Following the basis used in eq. (2.61), those ghost fields from $SU(2)_L \times U(1)_Y$ are labelled as c_+, c_-, c_Z and c_A, and similar for anti-ghost fields. Then the ghost terms are,

$$\mathcal{L}_{\text{ghost}} = \sum_{a,b}\bar{c}_a\frac{\partial\mathcal{F}^a(G)}{\partial\theta^b}c^b + \sum_{i,j}\bar{c}_i\frac{\partial\mathcal{F}^i}{\partial\alpha^j}c^j, \tag{2.62}$$

where a, b take values from $1, \cdots, 8$ and i and j run over $\{+, -, Z, A\}$. To evaluate those terms, let's first consider the gauge transform of arbitrary gauge field. The gauge transform of general non-Abelian gauge field, $A_m \mapsto U(G_m + i\partial_m)U^\dagger$, gives following infinitesimal transform with parameters θ,

$$\delta A_m^a = \partial_m\theta^a - g f_{bc}{}^a\theta^b A_m^c = (D_m^{\text{adj}})^a{}_b\theta^b. \tag{2.63}$$

Here D_m^{adj} is the covariant derivative of the adjoint representation. Therefore, the functional derivative of Lorentz-like gauge fixing terms with respect to group parameters is

$$\frac{\delta}{\delta\theta^b}\partial^m A_m^a = \delta(\theta)\partial^m D_m^{\text{adj}\ a}{}_b. \tag{2.64}$$

Therefore, the gauge fixing of $SU(3)_C$ have following ghost terms,

$$\bar{c}_a\frac{\partial\mathcal{F}^a(G)}{\partial\theta^b}c^b = \bar{c}_a\partial^m D_m^{\text{adj}\ a}{}_b c^b. \tag{2.65}$$

In electro-weak sector, we have additional gauge fixing terms from the subtraction of Nambu-Goldstone bosons. At the same time, the basis used for eq. (2.61) is non-trivial. We have to know the infinitesimal gauge transform of $SU(2)_L \times U(1)_Y$ gauge

bosons and related Nambu-Goldstone bosons from Φ. The suitable parameters would be $(\alpha^+, \alpha^-, \alpha_Z, \alpha_A)$, rather than $\vec{\alpha}_L \oplus \alpha_Y$. The definitions are

$$\alpha^\pm \equiv \frac{1}{\sqrt{2}}(\alpha_L^1 \mp i\alpha_L^2), \tag{2.66}$$

and

$$\begin{pmatrix} \alpha_Z \\ \alpha_A \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} \alpha_L^3 \\ \alpha_Y \end{pmatrix}. \tag{2.67}$$

The infinitesimal gauge transform of gauge fields $W_m^\pm$, Z_m and A_m with above parameters are,

$$\begin{aligned} \delta W_m^\pm &= \partial_m \alpha^\pm \mp i g_2 \alpha^\pm (\cos\theta_W Z_m + \sin\theta_W A_m) \\ &\quad \pm i g_2 (\cos\theta_W \alpha_Z + \sin\theta_W \alpha_A) W_m^\pm, \end{aligned} \tag{2.68}$$

and

$$\begin{aligned} \delta Z_m &= \partial_m \alpha_Z - i g_2 \cos\theta_W (\alpha^- W_m^+ - \alpha^+ W_M^-) \\ \delta A_m &= \partial_m \alpha_A - i g_2 \sin\theta_W (\alpha^- W_m^+ - \alpha^+ W_M^-). \end{aligned} \tag{2.69}$$

At the same time, corresponding infinitesimal transform of Φ would be as follow.

$$\begin{aligned} \delta H &= -\frac{g_2}{2\cos\theta_W} \alpha_Z \varphi^Z - g_2 \mathrm{Im}\left[\alpha^+ \varphi^-\right] \\ \delta \varphi^Z &= \frac{g_2}{2\cos\theta_W} \alpha_Z (v + H) + g_2 \mathrm{Re}\left[\alpha^+ \varphi^-\right] \\ \delta \varphi^\pm &= \mp \frac{i}{2} g_2 \alpha^+ (v + H \mp i\varphi^Z) \pm i \left(\frac{g_2 \cos 2\theta_W}{2\cos\theta_W} \alpha_Z + e\alpha_A \right) \varphi^\pm \end{aligned} \tag{2.70}$$

Now the gauge fixing terms from eq. (2.59) and infinitesimal transforms from eq. (2.68), eq. (2.69) and eq. (2.70), determine the ghost terms.

Combining them all, gauge kinetic terms from eq. (2.16), fermion kinetic terms from eq. (2.22), Higgs sector Lagrangian from eq. (2.23), Yukawa terms from eq. (2.41), gauge fixing terms from eq. (2.61) and ghost terms from eq. (2.62) complete the standard model Lagrangian.

There could be additional dimension-4 terms which are

$$\frac{\theta_3}{64\pi^2}\epsilon^{klmn}G^a_{kl}G^b_{mn}\delta_{ab} + \frac{\theta_2}{64\pi^2}\epsilon^{klmn}W^i_{kl}W^j_{mn}\delta_{ij} + \frac{\theta_1}{64\pi^2}\epsilon^{klmn}B_{kl}B_{mn}. \tag{2.71}$$

To manipulate these terms, let's consider a general non-Abelian gauge field $A_m \equiv gA^a_m T_a$ with Hermitian generators T_a under certain irreducible representation. Let $A^a{=}A^a_m dx^m$ be 1-form, and $A{=}gA^aT_a$ be matrix-valued one-form. This notation simplifies the field strength tensor as follow,

$$G \equiv \frac{1}{2}G_{mn}dx^m \wedge dx^n = dA - iA \wedge A. \tag{2.72}$$

Now, with the oriented volume-form $\omega = dx^0 \wedge \cdots \wedge dx^3$ in 4-dimensional Minkowski space-time,

$$\epsilon^{klmn}\mathrm{tr}(G_{kl}G_{mn})\omega = 4\mathrm{tr}(G \wedge G). \tag{2.73}$$

The trace $G \wedge G$, then contains four terms as below.

$$\begin{aligned} dA \wedge dA &= d(A \wedge dA) - (-1)^1 A \wedge d^2 A \\ \mathrm{tr}(dA \wedge (A \wedge A)) &= \frac{1}{3}\mathrm{tr}(d(A \wedge A \wedge A)) \\ &= (-1)^2\mathrm{tr}((A \wedge A) \wedge dA) \\ \mathrm{tr}((A \wedge A) \wedge (A \wedge A)) &= 0 \end{aligned} \tag{2.74}$$

The invariance of trace over cyclic permutation is used other than the first line above. From the above manipulation, the terms in eq. (2.71) is written as,

$$\begin{aligned} \epsilon^{klmn}\mathrm{tr}(G_{kl}G_{mn})\omega &= 4d\left[\mathrm{tr}\left(A \wedge dA - \frac{2}{3}iA \wedge A \wedge A\right)\right] \\ &= 2g^2\epsilon^{klmn}\partial_k\left(A^a_l\partial_m A_{an} + \frac{1}{3}gf_{abc}A^a_l A^b_m A^c_n\right)\omega. \end{aligned} \tag{2.75}$$

Using the Stokes' theorem, this term is represented as surface integral of a current. In 4-dimensional space-time, we can identify the current

$$J_{\mathrm{CS}} = \mathrm{tr}\left(A \wedge dA - \frac{2}{3}iA \wedge A \wedge A\right), \tag{2.76}$$

with so-called Chern-Simons current J_{CS}. In the Minkowski space-time manifold $\mathcal{M}$, therefore,

$$\int_{\mathcal{M}} \mathrm{tr}(G \wedge G) = \oint_{\partial\mathcal{M}} J_{\mathrm{CS}}, \tag{2.77}$$

and this is merely surface term. Yet, if we consider the finite gauge transform of this term, this would gives

$$\int_{\mathcal{M}} \mathrm{tr}(G \wedge G) \to \oint_{\partial\mathcal{M}} \left[2J_{\mathrm{CS}} + \frac{2}{3}\mathrm{tr}\Big(UdU^\dagger \wedge UdU^\dagger \wedge UdU^\dagger \Big) \right], \tag{2.78}$$

and the last term gives the numbers proportional to the quantised winding number elements in $\pi_3(G)$, the third homotopy group of the gauge group G. In case of $G = SU(N)$ or, in genera, simple Lie group, this is $\mathbb{Z}$. To maintain gauge invariance under finite gauge transform, the original term should also have its value from $\mathbb{Z}$, and, hence, it is quantised. At the same time, as long as θ_i the coefficients are not small, the term can give $\mathcal{O}(1)$ effect compared to the other terms. This is sever problem especially for QCD, of which coupling constant diverges in low energy physics. Unless θ_3 is small, this gives large charge-conjugation(C) and parity(P) violation at the same time, which is not the case of real world. To resolve this 'QCD θ-problem', we need additional mechanism which includes dynamical θ term, or axion field [28].

The applications of the standard model on IR physics, such as the parton evolution [29–31], or simulation of jet [32], we left references without further discussion.

2.2 Neural Network

The neural network is one of the machine learning architecture frequently used for building a discrimination program. Due to its large capability and growing hardware performance, it becomes one of the most widely-used algorithms for discrimination problem. To understand its output and the mechanism, here we shortly describe the internal structure of the neural network and the training procedure.

For the structure, as its name manifests, the neural network imitates the neural system of the brain to express arbitrary function of input features. The neural system consists of neurons and the connections between neurons. Each of those neurons delivers the simple output to next neuron depending on the stimulation. The whole system, however, can return the complicate output from the stimulation by combining simple response of each neuron. For example, let's consider the two neurons with the Heaviside step function as their response functions and one-dimensional real-valued input. Now, if we bypass the input to each neuron with different level, say x and $x-1$, and combine the output from two neurons in opposite direction, we have response function $f(x)$,

$$f(x) = \Theta(x) - \Theta(x-1). \tag{2.79}$$

This neural system activates only if the input stimulation x is in $[0, 1]$. Like this, even for a more complicated system can be constructed with many neurons by introducing non-trivial weights and biases to the connections between neurons.

In the context of machine learning with the neural network, such neurons are called as 'perceptrons'. Each perceptron has output as a function of inputs and its

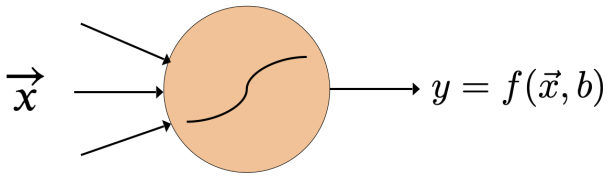


Figure 2.1 Schematic description of perceptron. Here $\vec{x}$ is an array of inputs from the previous perceptrons or bare input. b is bias inherited on its perceptron and and y is output with functional form of $f(\vec{x}, b)$.

own bias as depicted in Figure 2.1. The array of inputs $\vec{x}$ in Figure 2.1 acts as weighted sum of input and bias shifts the this input as Figure 2.2. There are many candidates for those functional form f. Those f are called as 'activation function'.

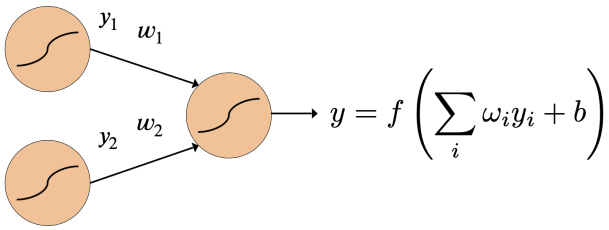


Figure 2.2 Schematic description of perceptron. b is bias inherited on its perceptron and and y is output. y_i's are outputs from previous perceptrons, and ω_i's are weights between previous i-th perceptron and itself.

Rectified-linear unit(or ReLU) which is $f(x) = \max(x, 0)$ is one of the most frequently used one, and there are also binary step($f(x)$=$\Theta(x)$), logistic(or sigmoid, $f(x)$=$(1 + e^{-x})^{-1}$) etc. Combining them all, the neural network is constructed as in Figure 2.3. Figure 2.3 is one of an example architecture which is called a deep neural network

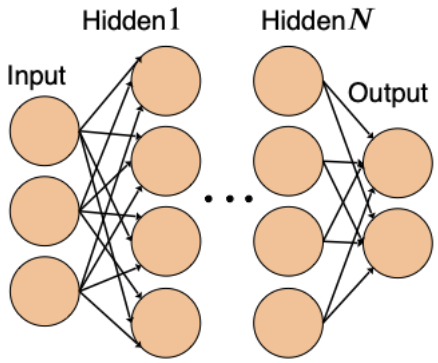


Figure 2.3 Schematic figure of deep neural network. Usually neural network means neural network with single hidden layer, and deep neural network is the one with more than one hidden layers.

with fully connected dense layers. Sometimes, to prevent an over-fitting problem, some of those connections are randomly cut, or additional layers are inserted between one hidden layer and its next hidden layers and normalised all weights between two layers.

Furthermore, sometimes additional layers are inserted to make more complicated processing, such as convolution with kernel function defined a priori. For the output from the discrimination model with n categories, one can place n output nodes with softmax output, which is $e^{x_i}/\sum_j e^{x_j}$ or with the other activation functions.

The details of the neural network, such as weights and biases are determined during the training procedure. Training is the procedure by which model parameters, weights and biases, are adjusted by minimisation of 'loss function'. The loss functions used for building discrimination model are minimised when the neural network could label input data with correct categories and maximised if labelling is maximally random. To be more specific, let us consider the binary classification case with label 0 and 1. In this case, so-called cross entropy function L is defined as,

$$\begin{aligned} L(\omega_{ij}, b_i|\mathrm{D}) &= -\sum_i [t_i \ln f(\vec{x}_i|\omega_{mn}, b_m) \\ &\qquad +(1-t_i)\ln\left(1-f(\vec{x}_i|\omega_{mn}, b_m)\right)]\,. \end{aligned} \tag{2.80}$$

Here i is label of each entries in data D, and $\vec{x}_i$ is array of input feature describing i-th entry. t_i are true label of each entry indicating the category to which entry belongs. Respective ω_{mn} and b_m are weights between m and n nodes(perceptrons) and bias associated to m-th node. Nor let's assume, for the simplicity, machine learning output $f(\vec{x}_i) \equiv f(\vec{x}_i|\omega_{mn}, b_m)$ with respect to any entry is restricted between $[0, 1]$. Then, the loss function L is minimised when all t_i are identical with $f(\vec{x}_i)$. In this case, L is 0. However, if the neural network just return $f(\vec{x}_i) = 1/2$ no matter what the input was, then $L = N\ln 2$ with the size of data N. Since $L \geq 0$, $L = 0$ is one of the global minimum, and w_{mn} and b_m giving $L = 0$ for any set of data models the perfect discriminator in this case. Now the problem is well-defined as minimisation of the given loss function, and the degree of 'correctness' of the neural network can be quantified. This minimisation is done by a stochastic gradient-based algorithm with a given data set. This is called stochastic gradient-descent method. In this method, the

neural network is initialised with random weights and biases, and evaluate stochastic gradient $\vec{\nabla}_{\vec{\omega}} L(\vec{\omega}$ is an array of model parameters including weights and biases) with a proper subset d of the full data set D. Then the model parameters are replaced by $\vec{\omega}\,'$ such that,

$$\vec{\omega}\,' = \vec{\omega} - \eta\, \vec{\omega} \cdot \vec{\nabla}_{\vec{\omega}} L. \tag{2.81}$$

Here η is called 'learning rate'. This is repeated until the loss function is saturated to a certain value. Note that, this neural network architecture can always over-fit the specific data set, D. Therefore this procedure is monitored with independent data set, called 'validation set'. The validation set is never used for evaluation of stochastic gradient, but it is only used for observing the value of loss function. If over-fitting occurred, the loss function evaluated with the validation set can increase, even though the loss function from the training set is kept decreasing. Hence the performance is watched during this back-propagation of the gradient, and training is usually cut at the certain 'epoch' with a minimum value of loss function while the loss function from the training set and validation set coincides with each other.

Chapter 3

Broad resonance in $t\bar{t}$ Final State

In this chapter, we discuss the application of machine learning on the search for broad resonance from beyond standard model. Different from a narrow-width case, there is a sharp resonance peak in case of broad resonance. Therefore, searching for a bump from the data in the invariant mass distribution of the final state may not be relevant. To study the alternative observables which are relevant for broad resonance search, other than such invariant mass distribution, we used machine learning, specifically deep neural network(DNN) architecture. DNN is used to build discrimination model between the signal from broad resonance and its irreducible background, and we tried to extract information from trained DNN by embedding in higher dimensional input space, planing a column or evaluation of a measure to represent the impact of each column of data on the output.

3.1 Introduction

Discovering new physics through a new resonance is one of the most exciting opportunities. A "narrow" resonance peak, being sharply localized in the energy spectrum, allows for the most efficient discovery above continuum backgrounds as well as for precision measurements of the particle mass, width and other properties. However, the widths of new (and presumably heavier) resonances in new physics can be easily much larger than those of the Standard Model (SM) particles. The width generally grows with the mass of resonance, and a new strong coupling may induce rapid decay as in composite Higgs models [33–36] or warped extra-dimensional models [37, 38]. Also, more decay channels to lighter beyond-SM particles may open up, which further increases the width.

The large width causes several difficulties in collider experiments. Above all, without a sharp peak, the discovery becomes challenging, as the signal becomes spread over a large range of energy above continuum backgrounds. For example, the ATLAS result based mostly on the invariant mass distribution [39] shows that for a $M = 1$ TeV Kaluza-Klein gluon, the measured (expected) cross-section upper limit $\sigma(pp \to g_{KK} \to t\bar{t})$ increases from 1.4 (1.2) pb to 4.7 (2.7) pb when the width-to-mass ratio Γ/M varies from 10% to 40%. In addition, the phenomenological study in Ref. [36] shows that for the minimal composite Higgs model with the third generation left-handed quark $Q_L = {}^{\mathrm{t}}(t_L, b_L)$ being fully composite, a vector $t\bar{t}$ resonance as light as $M = 1$ TeV is still allowed by the direct search in the $\Gamma/M \gtrsim 20\%$ region.

Secondly, broad resonance shape is more susceptible to the energy dependences of parton luminosity and the width, interferences with backgrounds or other resonances, and mixing and overlap with nearby resonances. These effects make discoveries further challenging and complicated. In particular, the complex interference (the one with imaginary parts in amplitudes) in supersymmetric or two-Higgs doublet models

can make broad heavy Higgs bosons decaying to $t\bar{t}$ generally appear not as a pure resonance peak [40–43] but even as pure dips or nothing [43]. And nearly degenerate heavy Higgs bosons can overlap significantly, producing complicated resonance shapes [44–46].

Many of these new broad resonances are just beyond the current reach of the LHC. Thus, it is imperative to study the physics of broad resonances and develop efficient discovery methods. However, broad-resonance searches have been studied only in limited cases, e.g., phenomenologically in Refs. [36] (third-generation quark pair, $\ell^+\ell^-$), [37] ($\mu^+\mu^-$), and experimentally in Refs. [39, 47] ($t\bar{t}$), [48, 49] (jj) [50]($\ell^+\ell^-$). In all the cases, the invariant mass had still been used as a main observable, but the question of "how do we (best) discover a broad resonance without a peak?" had not been answered thoroughly [1].

This question might be a problem appropriate to use DNN technique to answer. It is because the answer is not so obvious, *a priori*, and even small improvements will be significant. Machine learning has indeed been applied to various problems in particle physics. For example, bump-hunting resonance searches were improved with DNN [52, 53]. The DNN is one of machine learning algorithms. Coming with various network structures such as fully-connected network [54–59], convolutional neural network [15, 60–62] and others [63–67], DNN had shown remarkable performances in the exploration of physics beyond the SM, often better than other machine learning algorithms such as boosted decision tree (BDT). We refer to Refs. [68, 69] and references therein for reviews of the DNN applications in LHC physics.

Here we consider a spin-1 broad $t\bar{t}$ resonance at the LHC (Section 3.2). Being the heaviest particle in the SM, the top quark has been regarded as an important portal to new physics. As a first step toward a more general study of broad resonances, we ignore

[1] If a broad resonance can decay to multi-top/W channels, it can be searched using the same-sign di-lepton final state [35, 36, 51], which doesn't rely on the reconstruction of the invariant mass.

any interference effects and nearby resonances (Section 3.3). We use fully-connected DNN to explore answers beyond common knowledge (Section 3.3). Finally, we assess whether and what DNN can learn, even beyond what we know well (Section 3.4).

3.2 Benchmark Model

For simplicity, here we consider a gauge singlet vector resonance ρ interacting strongly with the SM right-handed top quark t_R, and the relevant Lagrangian is

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}\rho_{mn}\rho^{mn} - \frac{m_\rho^2}{2g_\rho^2}(g_\rho\rho_m - g_1 B_m)^2 \\ &\quad -\bar{t}_R\gamma^m t_R(g_\rho\rho_m - g_1 B_m),\end{aligned} \tag{3.1}$$

where $\rho_{mn} = \partial_m\rho_n - \partial_n\rho_m$, and g_1 is the SM hypercharge gauge coupling. This model is also considered in Refs. [34, 51, 70]. Note that the ρ_m mixes with the SM gauge field B_m in eq. (3.1). Given $g_\rho \gg g_1$, the mixing angle is $\sin\theta \approx g_1/g_\rho$ before the electroweak symmetry breaking (EWSB). Therefore, after transforming to the mass eigenstates, the interactions between ρ resonance and SM fermions will be $\sim g_\rho$ for t_R and $\sim Y g_1^2/g_\rho$ for other fermions (including t_L and other light quarks), with Y being the hypercharge of the corresponding fermion. The physical mass of ρ is $M_\rho = m_\rho$. EWSB gives $\mathcal{O}(v^2/m_\rho^2)$ corrections to above picture, and the details can be found in Appendix C of Ref. [51].

Due to the large coupling g_ρ, the ρ resonance decays to $t\bar{t}$ with a branching ratio $\sim 100\%$, and the width-to-mass ratio is

$$\frac{\Gamma_\rho}{M_\rho} \approx \frac{\Gamma_{\rho\to t\bar{t}}}{M_\rho} \approx \frac{g_\rho^2}{8\pi}. \tag{3.2}$$

For $g_\rho = 3$ and 4, this ratio reaches 36% and 64%, respectively. Thus a broad ρ is easily realized in the model described by Eq. (3.1). Note that $\Gamma_{\rho\to t\bar{t}} \leqslant \Gamma_\rho$, if ρ has other strong dynamical decay channels such as the decay to low-mass top partners (which are not listed in our simplified model), typically Γ_ρ is several times larger than

$\Gamma_{\rho\to t\bar{t}}$, thus a large Γ_ρ/M_ρ can be obtained even for smaller g_ρ. We consider $M_\rho = 1$ and 5 TeV as two benchmarks, and for each mass point the width-to-mass ratios Γ_ρ/M_ρ = 10%, 20%, 30% and 40% are considered. The corresponding benchmark cases are then identified as M$i\Gamma j$, with $i = 1$ or 5 denoting the mass (in unit of TeV) and $j = 1, 2, 3, 4$ being $10 \times \Gamma_\rho/M_\rho$. For example, M1$\Gamma$4 is the benchmark for $M_\rho = 1$ TeV and $\Gamma_\rho/M_\rho = 40\%$.

At the LHC, the ρ resonance can be produced via the Drell-Yan process ($q\bar{q} \to \rho$) through the ρ-light quark interaction. Among the various decay channels of the $t\bar{t}$, we choose to focus on the semi-leptonic final state

$$pp \to \rho \to t\bar{t} \to \ell^\pm \nu b\bar{b}jj. \tag{3.3}$$

The dominant background is then the SM $t\bar{t}$ process, which contributes 81% $\sim$ 88% of the total backgrounds [39]. For simplicity, we only consider this background. It should be emphasized that although we provide a benchmark model as a physical motivation here, our results are general for all heavy singlet spin-1 resonances with top quark portal.

3.3 Searching for a Broad $t\bar{t}$ Resonance

In this section, we describe the technical details of our work and show the final cross-section limits. First, we describe how we parameterize a broad resonance, and how we build learning datasets and train DNN for each benchmark signal case. Then we derive improved cross-section upper limits.

3.3.1 Breit-Wigner Parametrisation

We assume a single, isolated broad resonance far away from any other resonances and thresholds, and ignore any interference effects. Then we use the following Breit-

Wigner description of the propagator of a broad resonance

$$\frac{1}{s - \hat{M}_\rho^2(s) + i\sqrt{s}\hat{\Gamma}_\rho(s)} \approx \frac{1}{s - M_\rho^2 + iM_\rho\Gamma_\rho}, \tag{3.4}$$

where the nominal resonance mass M_ρ and the width Γ_ρ are fixed constants. The energy dependence of the mass $\hat{M}_\rho(s)$ from the real part of the self-energy correction is higher-order, hence small irrespective of the large width. On the other hand, the energy dependence of the width $\hat{\Gamma}_\rho(s) \propto \sqrt{s}$ from the imaginary part can induce corrections as large as ∼100 (10)% for broad resonances considered here $\Gamma_\rho/M_\rho \sim$ 40 (20)%. But, within this range of the width, the resonance shape remains relatively undistorted albeit some shifts of the peak and height [36,37,71]. Also, the fixed mass and width have been used in LHC searches of broad resonances [39,47]. Thus, we use Eq. (3.4) with fixed M_ρ and Γ_ρ, both for simplicity and for comparison purpose.

3.3.2 Preparation of Training Data

The model described by eq. (3.1) is written in the universal `FeynRules` output file [72]. We generate parton-level events of the signals and background using 5-flavour scheme within the `MadGraph5_aMC@NLO` [73] package. All spin correlations of the final state $\ell^\pm\nu b\bar{b}jj$ objects are kept. The phase space integrate region is set to $|\sqrt{s} - M_\rho| \leqslant 15 \times \Gamma_\rho$, which is large enough for us to simulate the full on- and off-shell effect of the ρ resonance. The interference between $pp \to \rho \to t\bar{t}$ and the SM $t\bar{t}$ background is negligible [39], thus not considered here. We normalize the SM $t\bar{t}$ cross section with the the next-to-next-to-leading order with next-to-next-to-leading logarithmic soft-gluon resummation calculation from the `Top++2.0` package [74–79], and the K-factor is 1.63. The parton-level events are matched to +1 jet final state and then interfaced to `Pythia 8` [80] and `Delphes` [81] for parton shower and fast detector simulation. As for the detector setup, we mainly use the CMS configuration, but with following modifications: the isolation ΔR parameters for electron, muon and

Process	Event number	Cut 1	Cut 2	Cut 3	Efficiency
M1Γ1	5.00×10^6	3.32×10^6	3.02×10^6	1.81×10^6	36.3%
M1Γ2	5.00×10^6	3.29×10^6	2.98×10^6	1.79×10^6	35.8%
M1Γ3	3.85×10^6	2.52×10^6	2.23×10^6	1.36×10^6	35.3%
M1Γ4	5.00×10^6	3.25×10^6	2.93×10^6	1.75×10^6	34.9%
SM $t\bar{t}$	4.98×10^6	2.60×10^6	2.21×10^6	1.39×10^6	28.0%

Table 3.1 The cut flows of the signals and background in resolved region. The events are generated at $1\ell^{\pm}+\not{E}_T+$ jets final state, where ℓ denotes e and μ. M$i\Gamma j$ denotes the benchmark case with $M_\rho = i$ TeV and $\Gamma_\rho/M_\rho = 0.1 \times j$.

jet are set to 0.2, 0.3 and 0.5 respectively. The b-tagging efficiency (and mis-tag rate for c-jet, light-flavour jets) is corrected to 0.77 (and 1/6, 1/134) according to Ref. [82]. We generate $\gtrsim 5 \times 10^6$ events for the background and each signal benchmark.

We defined two kinematic regions. The first one is called the resolved region, in which the decay products of the top quark (i.e $\ell^{\pm}\nu b\bar{b}jj$) are identified as individual objects. This region is defined as follows

1. Exactly one charged lepton $\ell^{\pm} = e^{\pm}$ or $\mu^{\pm}$ with $p_T^{\ell} > 30$ GeV and $|\eta^{\ell}| < 2.5$. Events containing a second lepton with $p_T^{\ell} > 25$ GeV are vetoed.

2. $\not{E}_T > 20$ GeV and $\not{E}_T + M_T^W > 60$ GeV, where the W-transverse mass is defined as
$$M_T^W = \sqrt{2p_T^{\ell}\not{E}_T\left[1 - \cos\Delta\phi(p_T^{\ell}, \not{E}_T)\right]}.$$

3. At least four jets with $p_T^j > 25$ GeV and $|\eta^j| < 2.5$, and at least one of the leading four jets is b-tagged.

The cuts are mainly based on Ref. [39], but with some simplifications. The cut flows of the signals and backgrounds are listed in Table 3.1. We only consider the

$M_\rho = 1$ TeV benchmark cases in this kinematic region. The SM $t\bar{t}$ cross section is 68.9 pb taken into account the K-factor.

Process	Event number	Cut 1	Cut 2	Cut 3	Cut 4	Efficiency
M1Γ1	5.00×10^6	3.32×10^6	3.02×10^6	1.17×10^6	9.61×10^5	19.2%
M1Γ2	5.00×10^6	3.29×10^6	2.98×10^6	1.08×10^6	8.85×10^5	17.7%
M1Γ3	5.00×10^6	3.27×10^6	2.96×10^6	1.02×10^6	8.34×10^5	16.7%
M1Γ4	5.05×10^6	3.28×10^6	2.96×10^6	9.92×10^5	8.06×10^5	15.9%
M5Γ1	5.00×10^6	2.53×10^6	2.36×10^6	1.15×10^6	8.41×10^5	16.8%
M5Γ2	5.00×10^6	2.72×10^6	2.52×10^6	1.19×10^6	8.76×10^5	17.5%
M5Γ3	5.00×10^6	2.81×10^6	2.59×10^6	1.19×10^6	8.85×10^5	17.7%
M5Γ4	5.00×10^6	2.86×10^6	2.64×10^6	1.20×10^6	8.90×10^5	17.8%
SM $t\bar{t}$ (`xptj = 150`)	1.99×10^7	1.22×10^7	1.08×10^7	1.41×10^6	1.21×10^6	6.10%

Table 3.2 The cut flows of the signals and background in boosted region. The events are generated at $1\ell^{\pm} + \not{E}_T$ + jets final state, where ℓ denotes e and μ. M$i\Gamma j$ denotes the benchmark case with $M_\rho = i$ TeV and $\Gamma_\rho/M_\rho = 0.1 \times j$. The setup `xptj = 150` is to improve the event generating efficiency of the background, see the text for details.

The second kinematic region is the boosted region, in which the hadronic decay products of the top quark are combined into a fat jet. The corresponding event selection criteria is

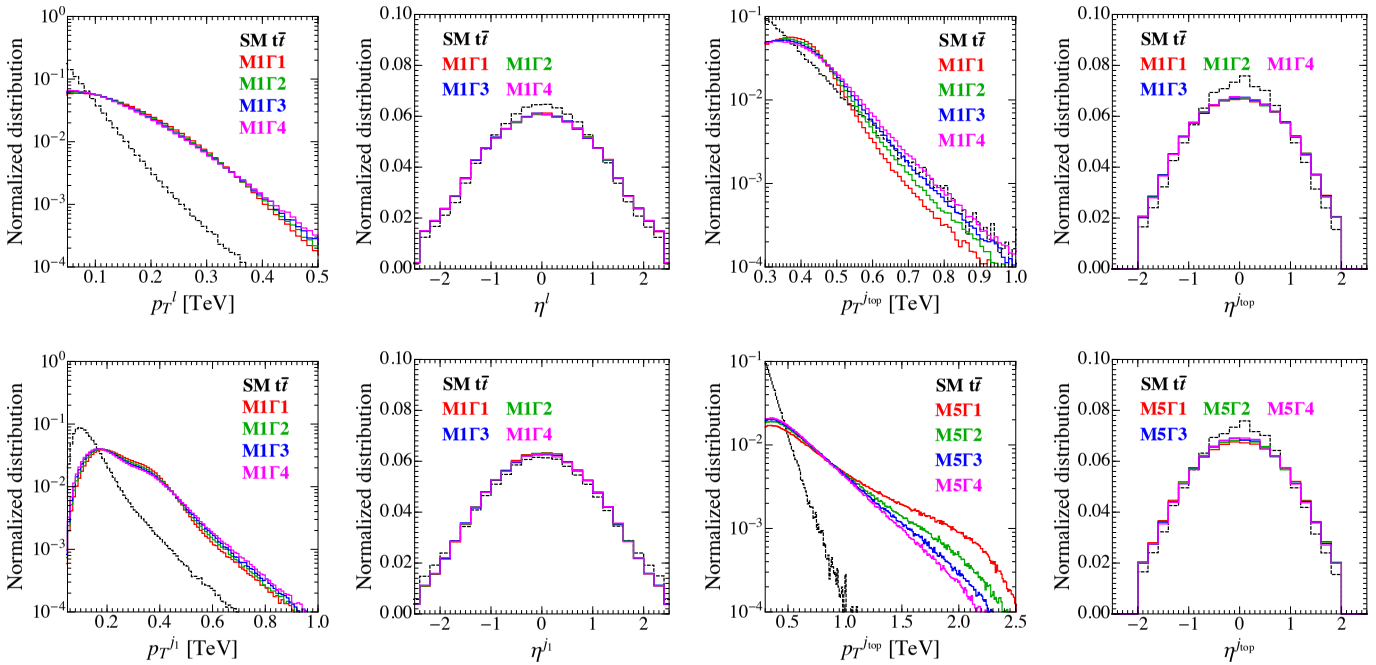
1. Exactly one charged lepton $\ell^\pm = e^\pm$ or $\mu^\pm$ with $p_T^\ell > 30$ GeV and $|\eta^\ell| < 2.5$. Events containing a second lepton with $p_T^\ell > 25$ GeV are vetoed.

2. $\not{E}_T > 20$ GeV and $\not{E}_T + M_T^W > 60$ GeV.

3. Exactly one top-jet with $p_T^{j_{\rm top}} > 300$ GeV and $|\eta^{j_{\rm top}}| < 2.0$, and satisfies $\Delta\phi(j_{\rm top}, \ell^\pm) > 2.3$. The top-jet is reconstructed with a $R = 1.0$ cone in anti-k_t algorithm, and is trimmed with $R_{\rm cut} = 0.2$ and $f_{\rm cut} = 0.05$ [83]. We use a simplified top-tagging procedure in event selection. The top-tagging efficiency and the mistag-rate are set to 80% and 20% respectively, based on Ref. [84], which makes use of jet invariant mass and N-subjettiness [85–90].

4. Exactly one selected jet with $p_T^{j_{\rm sel}} > 25$ GeV and $|\eta^{j_{\rm sel}}| < 2.5$. In addition, the selected jet should have $\Delta R(j_{\rm sel}, j_{\rm top}) > 1.5$ and $\Delta R(j_{\rm sel}, \ell) < 1.5$.

The cuts here are again mainly based on Ref. [39]. and the cut flows for signals and background are listed in Table 3.2. In this region, we consider both $M_\rho = 1$ and 5 TeV signals. To increase the event generating efficiency of the background events, in this region we require the SM $pp \to t\bar{t} \to \ell^\pm \nu b \bar{b} j j$ process has at least one final state parton (including the b-parton) with $p_T > 150$ GeV. This is done by setting `xptj = 150` in `MadGraph5_aMC@NLO`. We have checked that this setup doesn't lose the generality, but improves the event generating efficiency by a factor of ~ 6. The background cross section after cuts is 2.88 pb taken into account the K-factor.



(a) Examples of distributions in the resolved region.

(b) Examples of distributions in the boosted region.

Figure 3.1 Distributions of some low-level observables used to train a DNN. Selection cuts in Table 3.1 and Table 3.2 are applied.

The events after cuts are collected to make training and validation/test datasets. For the resolved region, we have $1\ell^{\pm} + \not{E}_T + 4$ jets in total 6 reconstructed objects in the final state, and 26 low-level kinematic observables can be used as input features: E^{ℓ}, p_T^{ℓ}, η^{ℓ} and ϕ^{ℓ} from the charged lepton; $\not{E}_T$, $\phi^{\not{E}_T}$ from the missing transverse momentum; E^{j_i}, $p_T^{j_i}$, η^{j_i}, ϕ^{j_i} and b^{j_i} from the 4 leading jets, with $i = 1, 2, 3, 4$. Here b^j is the b-tagging observable, which is 1 for a b-tagged jet and 0 otherwise. Some examples of the low-level observables distributions are shown in Figure 3.1(a). For each benchmark case (i.e. M1Γ1$\sim$M1Γ4), we build a training dataset and a validation/test dataset. Both of those two datasets have 1,000,000 events, which contain nearly equal signal and background events.

For the boosted region, $1\ell^{\pm} + \not{E}_T + 1$ top-jet $+ 1$ selected jet in total 4 objects are reconstructed, and we can extract 15 low-level observables as input features: the first 6 are from ℓ and $\not{E}_T$, same as the resolved region; the other 9 insist of $E^{j_{\rm sel}}$, $p_T^{j_{\rm sel}}$, $\eta^{j_{\rm sel}}$, $\phi^{j_{\rm sel}}$, $b^{j_{\rm sel}}$ from the selected jet, and $E^{j_{\rm top}}$, $p_T^{j_{\rm top}}$, $\eta^{j_{\rm top}}$, $\phi^{j_{\rm top}}$ from the top-jet. Some examples of the low-level observables distributions are illustrated in Figure 3.1(b). For each benchmark case (i.e. M1Γ1$\sim$M1Γ4, and M5Γ1$\sim$M5Γ4), we randomly mix equal number of signal and background events to get 800,000 events for training and another 800,000 events for validation/test.

3.3.3 Training the DNN

The DNN classifier is implemented using the `Keras` [91] package (with `Tensorflow` [92] as the backend). The architecture of the DNN is as follows,

$$[N_{\rm in}, \underbrace{N_{\rm node}, N_{\rm node}, \cdots, N_{\rm node}}_{N_{\rm hidden}}, 2], \tag{3.5}$$

where $N_{\rm hidden}$ and $N_{\rm node}$ are the numbers of hidden layers and the number of neurons per hidden layer, respectively. The number of input features $N_{\rm in} = 26$ (15)

for the resolved (boosted) region. All the input features are rescaled to have average 0 and standard deviation 1 before training. We label the events with column matrices to match the two neurons in output layer:

$$\text{signal} \to \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ \text{background} \to \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{3.6}$$

The Rectified Linear Unit (`ReLU`) activation function is used for all the hidden layers, while the `softmax` activation function is adopted for the output layer. The loss function is `categorical_crossentropy`, and the optimizer is `Adam`. To get the best configuration of the DNN, we try various choices of the hyper-parameter combination as follows,

$$\begin{aligned} &N_{\text{hidden}} = 4,\ 5; && N_{\text{node}} = 200,\ 300; \\ &L_r = 0.001,\ 0.003; && D_r = 0.1,\ 0.2,\ 0.3; \\ &B_s = 10^3,\ 10^4; \end{aligned} \tag{3.7}$$

where L_r is the initial learning rate, D_r is the dropout rate, and B_s is the batch size. For each benchmark case, there are in total 48 different DNN configurations, in which we select the best one based on the learning curves with the following criteria:

1. If the validation/test accuracy curve achieves its maximum when crossing with the training accuracy curve, and meanwhile the validation/test loss curve reaches its minimum and crosses with the training curve, we select that configuration and cut the training at that epoch. This early stop is to prevent over-fitting.

2. If more than one configurations have the behaviors mentioned above, then we select the one with the higher validation/test accuracy and lower validation/test

loss; if still there remain more than one networks, we choose the one with learning curves having less fluctuation.

The details of training and the chosen configurations are listed in Table 3.3 and Table 3.4 for $M_\rho = 1$ and 5 TeV, respectively. The epochs when we cut the training are listed in the forth columns of the tables. We left short remark that, for a individual signal benchmark in a given kinematic region, the DNN with low-level observables usually requires a longer training epoch than the DNN with all observables, if they have the same configurations. That is because the DNN needs more time to learn about the physics in the signal process, if no hint is given to it. The classification accuracies (on the validation/test data) of the networks are given in the fifth columns.

For the $M_\rho = 1$ TeV models, the DNN can reach a classification accuracy of $\geqslant 80\%$ in the resolved region and of $\geqslant 65\%$ in the boosted region. While for the $M_\rho = 5$ TeV case, the accuracy is $\geqslant 76\%$ in the boosted region.

Benchmark case	Kinematic region	Observables	N_{hidden}, N_{node}, L_r, D_r, B_s, N_{epoch}	Classification accuracy
M1Γ1	resolved	low-level	5, 200, 0.001, 0.2, 10^3, 150	85.2%
		all	5, 200, 0.001, 0.2, 10^3, 100	85.1%
	boosted	low-level	5, 200, 0.001, 0.2, 10^4, 55	67.9%
		all	4, 300, 0.001, 0.3, 10^3, 40	70.1%
M1Γ2	resolved	low-level	4, 300, 0.003, 0.2, 10^3, 35	83.2%
		all	5, 200, 0.001, 0.2, 10^3, 60	83.2%
	boosted	low-level	5, 200, 0.001, 0.2, 10^4, 45	65.8%
		all	4, 300, 0.003, 0.3, 10^4, 40	68.2%
M1Γ3	resolved	low-level	4, 300, 0.001, 0.2, 10^3, 30	81.6%
		all	5, 200, 0.001, 0.2, 10^3, 40	81.6%
	boosted	low-level	4, 300, 0.003, 0.2, 10^4, 30	65.1%
		all	4, 300, 0.003, 0.3, 10^4, 40	67.0%
M1Γ4	resolved	low-level	5, 200, 0.001, 0.2, 10^3, 80	80.8%
		all	5, 200, 0.001, 0.2, 10^3, 40	80.6%
	boosted	low-level	4, 300, 0.001, 0.2, 10^4, 20	64.3%
		all	4, 300, 0.001, 0.3, 10^3, 40	66.7%

Table 3.3 The selected networks for $M_\rho = 1$ TeV. N_{epoch} is the epoch number when we cut the training.

Benchmark case	Kinematic region	Observables	N_{hidden}, N_{node}, L_r, D_r, B_s, N_{epoch}	Classification accuracy
M5Γ1	boosted	low-level	4, 300, 0.001, 0.2, 10^3, 20	79.5%
		all	5, 200, 0.001, 0.1, 10^4, 30	80.5%
M5Γ2	boosted	low-level	4, 300, 0.003, 0.2, 10^3, 40	78.2%
		all	5, 200, 0.001, 0.1, 10^4, 45	79.1%
M5Γ3	boosted	low-level	4, 300, 0.003, 0.2, 10^3, 30	77.4%
		all	4, 300, 0.003, 0.3, 10^4, 40	78.4%
M5Γ4	boosted	low-level	4, 300, 0.003, 0.2, 10^3, 45	76.8%
		all	5, 200, 0.001, 0.3, 10^4, 45	77.8%

Table 3.4 The selected networks for $M_\rho = 5$ TeV. N_{epoch} is the epoch number when we cut the training.

The `softmax` activation function for the output layer guarantees the output responses of the 0th neuron (r_0) and the 1st neuron (r_1) satisfy

$$0 < r_0,\ r_1 < 1, \quad r_0 + r_1 \equiv 1. \tag{3.8}$$

Therefore, we can consider r_1 only, and denote it as r. Due to the label definition in eq. (3.6), If the DNN is well trained, the distribution of r should have a peak around 1.0 (0.0) for the signal (background), for both the training data and the validation/test data. Figure 3.2 shows the distributions of the validation/test data for benchmark cases with $\Gamma_\rho/M_\rho = 40\%$ as an illustration. The DNN for M1Γ4 shows worse performance in boosted region compare to the one in resolved region. This is because that two peaks in neuron output from signal and SM background are not separated well. In fact, this is a generic feature for all $M_\rho = 1$ TeV benchmark cases. It is mainly due to the the boosted region cuts, which require a top-jet with $p_T^{j_{\rm top}} > 300$ GeV. As a result, most of the SM $t\bar{t}$ background events are round this value. However, for a $M_\rho = 1$ TeV resonance, its decay product $t/\bar{t}$ acquires a transverse momentum ~ 500 GeV, quite similar to the cut threshold. Therefore, the signal and background look similar (see the $p_T^{j_{\rm top}}$ distribution in Figure 3.1(b)), and thus the separation is not efficient. On the other hand, for a $M_\rho = 5$ TeV resonance, $p_T^{j_{\rm top}} \sim 2.5$ TeV, the DNN works very well, as plotted in the bottom of Figure 3.2.

3.3.4 Setting Bounds for the Signal

We treat the neuron output r as an observable, and fit its distribution shape to get the cross section upper limit of $pp \to \rho \to t\bar{t}$ for a given integrated luminosity. For the $M_\rho = 1$ TeV benchmark cases, we use a binned χ^2 fitting method by dividing the $0 < r < 1$ range into 50 bins. While for the $M_\rho = 5$ TeV benchmarks, as the

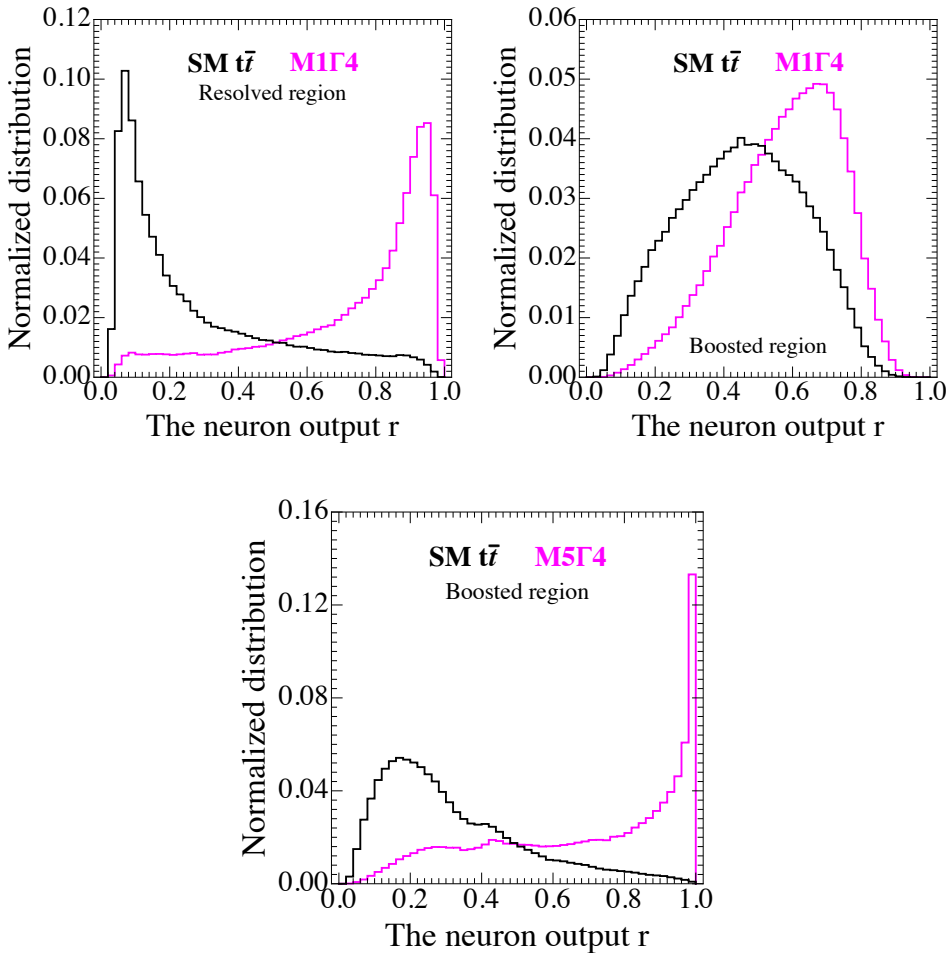


Figure 3.2 The final DNN output r distributions that we use to obtain cross section upper limits. Benchmark cases with $\Gamma_\rho/M_\rho = 40\%$ and $t\bar{t}$ backgrounds.

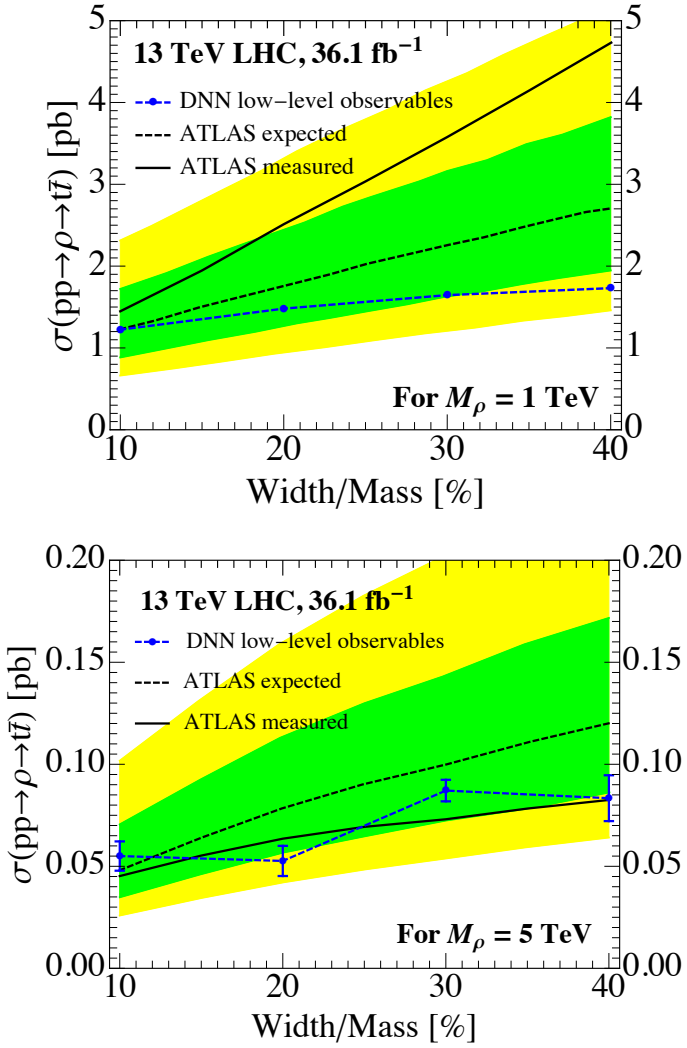


Figure 3.3 The DNN-improved cross section upper limits at 95% C.L., obtained by fitting DNN output r distributions in Figure 3.2. The latest ATLAS results [39] are also shown for comparison. The vertical error bars of the DNN results are training uncertainties, which are derived by running the same network for 15 times.

signal cross sections are expected to be tiny, to improve the efficiency we use the un-binned fitting method described in Refs. [93, 94]. In each case, we consider the statistic uncertainty and assume a 12% systematic uncertainty for the background. To include the effect of other subdominant backgrounds besides $t\bar{t}$ (i.e. W + jets, multi-jet, etc), we further rescale the cross section by a factor of $1.23 = 1/0.81$ and $1.14 = 1/0.88$ for the resolved and boosted regions, respectively. Those factors come from the fact that $t\bar{t}$ contributes 81% (88%) of the total background for resolved (boosted) region [39]. This simple rescaling could overestimate final contributions from subdominant backgrounds, and result in somewhat conservative estimations of cross section bounds.

The signal strength upper limits are derived for the unfolded parton-level cross section $\sigma(pp \to \rho \to t\bar{t})$, which can be compared with the final results in experimental papers, e.g. Refs. [39, 47]. Our results are shown in Figure 3.3, in which the expected and measured upper limits of Ref. [39] are also plotted as references, as they use the same final state and similar selection cuts. One can read that the DNN results are rather insensitive to the width of the ρ resonance compare to the traditional approach, achieving better constraints in the large width region [2]. For the $M_\rho = 1$ TeV benchmark, the result is obtained by the combined fitting of both resolved and boosted regions. Individually, the resolved and boosted regions respectively yield cross sections ~ 3 pb and ~ 1 pb. Although networks in the resolved region have a higher accuracy ($\geqslant 80\%$) than those in the boosted region ($\geqslant 65\%$) in Table 3.3, they actually give a worse measurement of the cross section. This is because the boosted cuts can remove lots of background events and hence improve the fitting performance. That is also the reason why we only consider the boosted region for $M_\rho = 5$ TeV: the production rate for such a high mass ρ is so small that we have to use the boosted region to suppress the background. The DNN bounds for 5 TeV signal benchmark are comparable to

[2] We also checked that the DNN results are better than those from more traditionally used BDT.

the experimentally measured ones, but still better than the experimentally expected ones. As the training uses random number for the initialization of weights and biases, even for a given DNN configuration, the final results are slightly different for different running. To take into account this training uncertainty, we repeat 15 times of running the chosen DNN configuration for each benchmark case. For the $M_\rho = 1$ TeV case, the relative fluctuation is small thus not shown; while for the $M_\rho = 5$ TeV case, the standard deviations of the runs are shown as vertical error bars in Figure 3.3.

3.4 Figuring out What the Machine Had Learned

In this section, we attempt to assess information learned by DNN using three methods, each of which will be discussed in each subsection. As a result, we can figure out not only which information has been learned, but also which information is most important.

3.4.1 Testing High-level Observables

It is important to know whether a DNN had learned well-known useful but complicated features. In fact, it has been argued that some machine learning methods such as jet image [15] do not efficiently capture invariant mass features [61].

Our approach is to train another set of DNNs using additional high-level observables, of which features we want to test. By comparing the performances of these new DNNs with the original DNNs trained with only low-level observables, we can test whether those particular high-level features (i.e. physically-motivated) have been effectively learned [3] or not. This "saturation approach" has been widely used in particle physics research [11, 55].

To construct high-level observables, we first reconstruct the t and $\bar{t}$. The longi-

[3] The definition of "learned" could be ambiguous, but we use subjective criteria discussed in text.

tudinal momentum of the neutrino is solved by requiring the leptonically decaying W to be on-shell, i.e. $M_{\ell\nu} = M_W$. For the resolved region, the assignment of the 4 reconstructed jets are done by minimizing

$$\chi^2 = \frac{(M_{jj} - M_W)^2}{\sigma_W^2} + \frac{(M_{jjj} - M_t)^2}{\sigma_t^2} + \frac{(M_{j\ell\nu} - M_t)^2}{\sigma_t^2},$$

for various jet permutations, where $\sigma_W = 0.1 \times M_W$ and $\sigma_t = 0.1 \times M_t$. For the boosted region, a top quark is identified as the top-jet and the other is reconstructed from the combination of $\ell^{\pm}\nu j_{\rm sel}$. Once the t and $\bar{t}$ are reconstructed, we are able to define the following 7 high-level observables for the signal $pp \to \rho \to t\bar{t}$:

1. The invariant mass $M_{t\bar{t}}$ of the $t\bar{t}$ system.

2. The polar angle and azimuthal angle in the Collins-Soper frame [95]. We label the leptonic and hadronic decaying tops with subscripts "tl" and "th", respectively. Hence we have $\cos\theta_{\rm tl}^{\rm CS}$, $\cos\theta_{\rm th}^{\rm CS}$, $\phi_{\rm tl}^{\rm CS}$ and $\phi_{\rm th}^{\rm CS}$ in total 4 observables.

3. The polar angles in the Mustraal frame [96], $\cos\theta_1^{\rm Mus.}$ and $\cos\theta_2^{\rm Mus.}$.

The first observable reveals the resonance feature, while the latter 6 observables reflect the spin-1 nature of the ρ resonance. For the boosted region, to take into account the features of the top jet, we introduce 3 additional high-level observables, i.e.

1. The invariant mass $M_{j_{\rm top}}$ of the top jet.

2. The N-subjettiness observables τ_{21} and τ_{32} of the top jet [86–90].

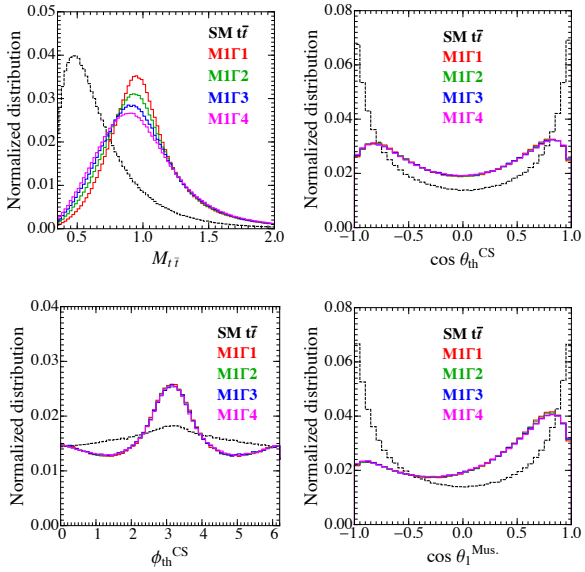
Those observables are shown to be important in identifying the colour structure of the hard process [18,90,97]. In our scenario, the signal results from a colour-singlet

resonance, while the background comes from QCD process, and the jet mass and N-subjettiness can help to reveal this difference [18]. Moreover, such jet substructures can be more independent on resonance characteristics and kinematics.

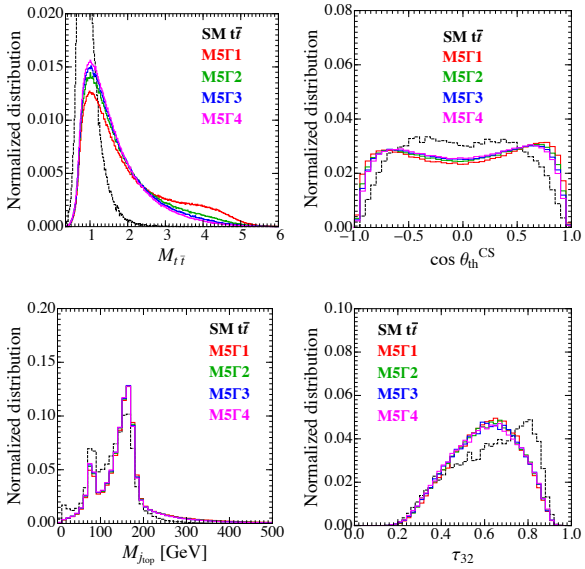
Some distributions of these high-level observables are shown in Figure 3.4. Note that the spin correlations as well as the jet substructure observables are rather insensitive to the width of ρ, as expected. For the 5 TeV resonance, the mass peak of $M_{t\bar{t}} \sim 5$ TeV almost disappears for $\Gamma_\rho/M_\rho \geqslant 10\%$; instead, there is a peak ~ 1 TeV,due to the parton-distribution support of off-shell effects and hard p_T cuts. Most identified top-jets in both signal and background originate correctly from the top quark, thus the differences shown in the distributions of $M_{j_{\text{top}}}$ and τ_{32} come from the colour structure of the hard process. For example, the background's $M_{j_{\text{top}}}$ distribution is slightly broader and the τ_{32} is slightly bigger than the signals. This is because the top-jets from QCD $t\bar{t}$ are colour connected with the initial state, consequently having more radiations. Using these "all observables" (i.e. sum of low- and high-level observables) as inputs, we train a new set of DNNs; best network configurations are again surveyed and detailed in Table 3.3 and Table 3.4.

We compare the performances of original and new DNNs using receiver operating characteristic (ROC) curves. The area under curve (AUC) is used as a metric of the performance. Some of the comparisons are shown in Figure 3.5. First, in the resolved region as shown in the top panel, we found that there is only little change on ROC curves by adding high-level observables. Not only AUC, but also background efficiencies show small change. This means that the inclusion of high-level observables does not yield the improvement of accuracy; the original DNN had learned those high-level features successfully from low-level inputs.

In the boosted region, while the $M_{t\bar{t}}$, $M_{j_{\text{top}}}$ and spin correlations can be derived from the four momenta of reconstructed objects, the N-subjettiness cannot be inferred



(a) Examples of distributions in the resolved region.



(b) Examples of distributions in the boosted region.

Figure 3.4 Distributions of high-level observables in the (a) resolved and (b) boosted regions. We use these to train a new set of DNNs to test whether such high-level features were learned.

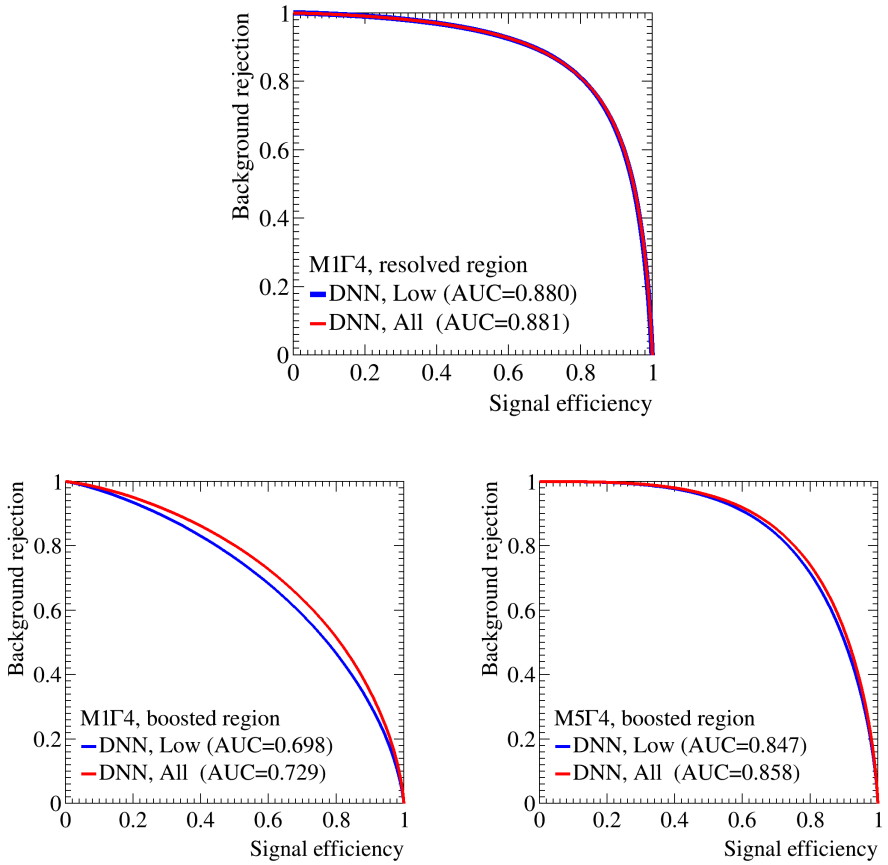


Figure 3.5 ROC curves, comparing the performance with ("all") and without ("low") high-level observables used to train DNNs. The AUC of each curve is also shown inside parenthesis.

from the low-level inputs. Therefore, adding high-level features can bring improvements. As shown with ROC in the bottom two panels of Figure 3.5, the improvement is sizable for M1Γ4, while, however, relatively small for M5Γ4. This may be because the event topology of M5 boosted cases becomes so simple that many features are more correlated.

3.4.2 Ranking Input Observables by Importance

Which information has been used most usefully by DNN in distinguishing a broad resonance against continuum background? To answer this, we attempt to identify which connections between which neurons and layers are weighted most importantly. Following Ref. [98], we define the learning speed of the j-th hidden layer as

$$v^{(j)} = \left| \vec{b} \cdot \left[\vec{\nabla} \mathcal{L}_{\text{loss}}(w, b) \right]_{w=0} \right|, \tag{3.9}$$

where $\vec{b}^{(j)}$ is the bias vector of the j-th hidden layer, while $\mathcal{L}_{\text{loss}}$ is the loss function. As the target of machine learning is to find the global minimum of $\mathcal{L}_{\text{loss}}$, the $v^{(j)}$ approximately reflects the training sensitivity of a specific layer. When training the DNN, the larger $v^{(j)}$ a layer acquires, the more important it is. We found that for all individual benchmark cases MiΓj, the first hidden layer has the highest learning speed several times larger than that of other layers. For example, for M1Γ4 case in the resolved region, the learning speed is $v^{(1)} = 0.457$, $v^{(2)} = 0.086$, $v^{(3)} = 0.033$, $v^{(4)} = 0.016$ and $v^{(5)} = 0.008$. This means that good features are typically learned most efficiently in the first hidden layer.

For our DNN architecture described in eq. (3.5), the weights of the first hidden layer form a $N_{\text{in}} \times N_{\text{node}}$ matrix, whose element is denoted as $w^{(1)}_{mn}$ with $m = 1, \cdots, N_{\text{in}}$ and $n = 1, \cdots, N_{\text{node}}$. As all the input features are rescaled to have average 0 and standard deviation 1, the magnitude of the weight $w^{(1)}_{mn}$ reflects the correlation strength between the m-th input and the n-th neuron in the first hidden layer. Motivated by

this, we further define

$$W_m = \mathcal{N} \left[\sum_{n=1}^{N_{\text{node}}} \left(w_{mn}^{(1)} \right)^2 \right]^{\frac{1}{2}}, \tag{3.10}$$

as a measure of the importance of the m-th input feature. The normalization $\mathcal{N}$ is such that

$$\sum_{m=1}^{N_{\text{in}}} W_m = 1. \tag{3.11}$$

Figure 3.6 and Figure 3.7 show the W_m's of each input observable from the DNN trained using both low- and high-level observables. Above all, the $M_{t\bar{t}}$ – that we expected to be less useful for a broad resonance – is still one of the most important observables even when the resonance is broad. This is particularly true for a low-mass broad resonance in the resolved region (upper panel). In the case of a heavy-resonance in the boosted region (lower panel), its importance is relatively reduced, partly because some invariant-mass information has been used in the selection of the boosted region. In such cases, the top-jet mass and transverse momentum which are somewhat correlated with $M_{t\bar{t}}$ and width can significantly complement the search, as shown in the bottom panel. In addition, the invariant mass of the top-jet is another important input feature because it reflects the colour flow difference between signals and background. On the other hand, N-subjettinesses again turn out to be relatively less useful.

Remarkably, there are much other useful information, particularly from angular distributions $\eta^{\ell,j}$ and $\cos\theta_{1,2}^{\text{Mus.}}$. From Figure 3.1 and Figure 3.4, we can see that these observables are relatively uncorrelated with the resonance width. We have indeed checked that the cross entropies [99] between these observables and $M_{t\bar{t}}$, which can quantify their correlations, are not so high. As we will see in the next subsection,

these information are useful even in the off-shell region away from the resonance, hence less correlated with the width. Thus, these features are useful in search of broad resonances. This may also imply that narrow-resonance searches can be improved by adding off-resonance information; this is partly because a large fraction of signals is still from low-energy off-resonance region where parton-luminosity support is much larger (although buried under larger backgrounds). We leave this for a future study.

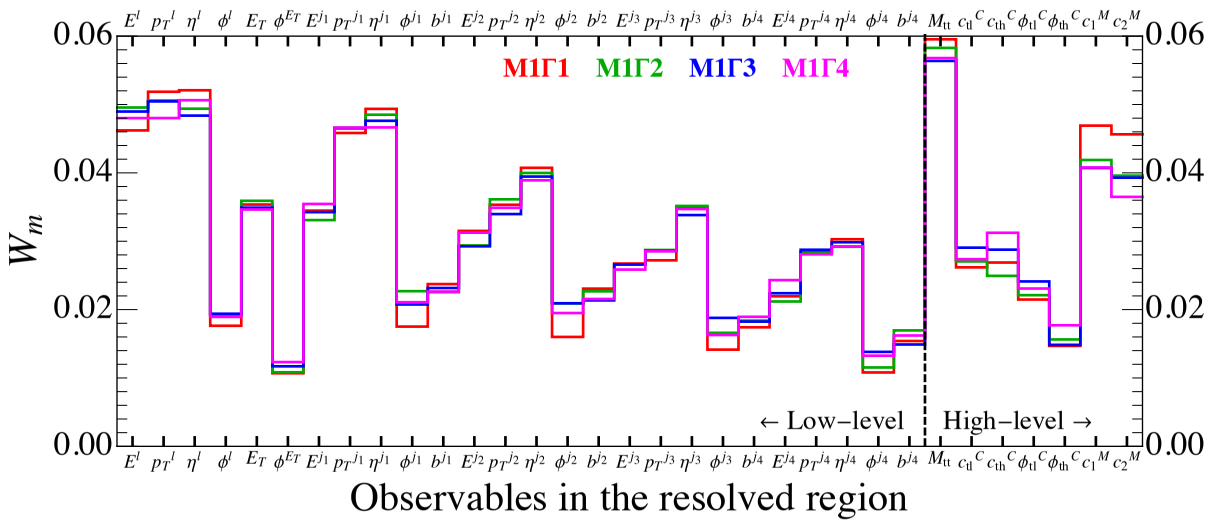


Figure 3.6 The weight of each observable in the first layer of DNN, W_m defined in eq. (3.10). $M_\rho = 1$ TeV in the resolved region. High-level observables are also used in training in order to get those weights. For notations, E_T and ϕ^{E_T} in the figures stand for $\not{E}_T$ and $\phi^{\not{E}_T}$, respectively. $c^{\rm C}_{\rm tl,th}$ and $c^{\rm CS}_{\rm tl,th}$ are short for $\cos\theta^{\rm CS}_{\rm tl,th}$, and $c^{\rm M}_{1,2}$ and $c^{\rm Mus}_{1,2}$ for $\cos\theta^{\rm Mus.}_{1,2}$.

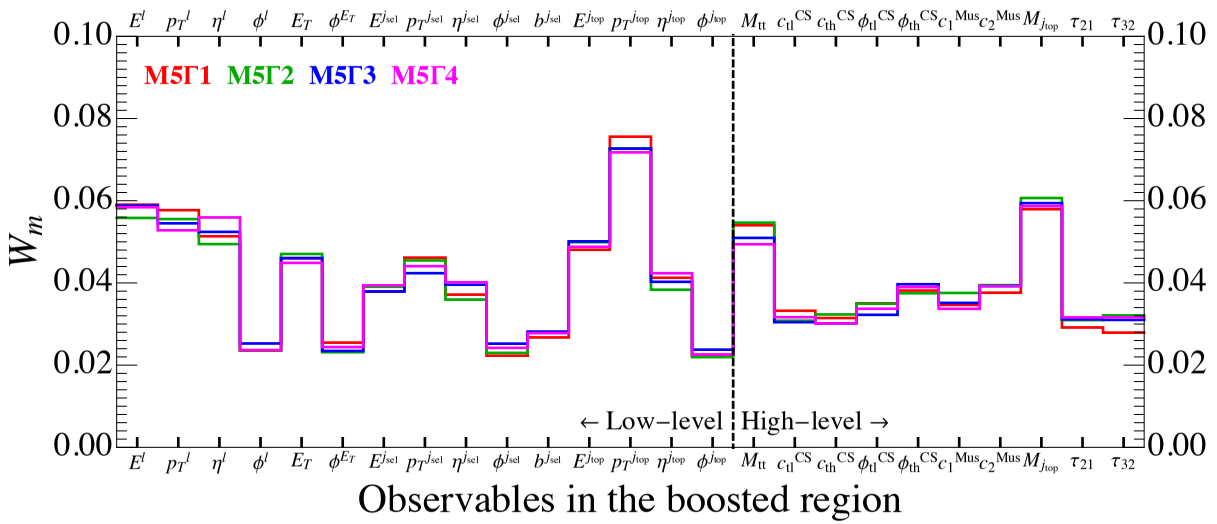


Figure 3.7 The weight of each observable in the first layer of DNN, W_m defined in eq. (3.10). $M_\rho = 5$ TeV in the boosted region. High-level observables are also used in training in order to get those weights. For notations, E_T and ϕ^{E_T} in the figures stand for $\not{E}_T$ and $\phi^{\not{E}_T}$, respectively. $c^{\rm C}_{\rm tl,th}$ and $c^{\rm CS}_{\rm tl,th}$ are short for $\cos\theta^{\rm CS}_{\rm tl,th}$, and $c^{\rm M}_{1,2}$ and $c^{\rm Mus}_{1,2}$ for $\cos\theta^{\rm Mus.}_{1,2}$.

3.4.3 Planing Away $M_{t\bar{t}}$

We have observed that $M_{t\bar{t}}$ is still important, but there are indeed uncorrelated useful information. How much is discovery capability attributed to those uncorrelated (whether known or unknown) information? Using the data planing method [61, 100], we wash out the feature in the invariant mass spectrum. We attach a weight to each event so that the weighted distribution of $M_{t\bar{t}}$ becomes flat for both signals and backgrounds. The details of chosen network configurations are described in Table 3.5 for both before and after planing away the key observable $M_{t\bar{t}}$. The data of the second row, i.e. the accuracies before planing, are taken from the fifth columns of Table 3.3 and Table 3.4. While the accuracies after planing listed in the third row are obtained by the weighted training.

A new set of DNNs trained with such planed data must learn information uncorrelated with $M_{t\bar{t}}$, and the difference between the performance with/without $M_{t\bar{t}}$ offers a quantitative answer to the question "how much information it is beyond the invariant mass".

In practice, to avoid large fluctuations, we use only $M_{t\bar{t}} \in [0.5, 3]$ TeV region with 20 GeV bin size for all signal cases. This means that for 5 TeV signals, we consider only off-resonance events; note that the majority of signal is from the low-energy region supported by larger parton luminosities.

After $M_{t\bar{t}}$ planed away, the classification accuracies reduce from $\geqslant 80\%$ to $\geqslant 73\%$ for $M_\rho = 1$ TeV in the resolved region and from $\geqslant 65\%$ to $\geqslant 62\%$ in the boosted region. For $M_\rho = 5$ TeV cases, accuracies reduce from $\geqslant 76\%$ to $\geqslant 63\%$ in the boosted region. As accuracies are still significantly higher than random guess (i.e. 50%), we conclude that DNNs still have some capabilities to distinguish signals from background, even though they are blind to $M_{t\bar{t}}$ and most events are from off-resonance region (for 5 TeV cases). Clearly, on top of $M_{t\bar{t}}$ and width, the original DNNs had

learned extra information (such as aforementioned angular correlations).

Indeed, we have checked that the weights W_m for various anglular and angular-correlation observables, after planing the $M_{t\bar{t}}$, are relatively high. From Figure 3.1 and Figure 3.4, one can also see that they are largely independent on the width. The helicity conservation (hence, angular correlations) can hold somewhat independently of the invariant mass, as the range of the invariant mass considered is always much larger than the top mass. Thus, we conclude that much of the angular information can be from off-resonance region, and such off-resonance information (although buried under larger backgrounds) can enhance discovery power. As a result, as shown in Figure 3.3, final performance is not only improved but became rather insensitive to the resonance width.

A final remark is that there could still be *unknown* (to us) useful information that are not identified in our analysis.

Models	M1Γ1		M1Γ2		M1Γ3		M1Γ4	
Kinematic region	resolved	boosted	resolved	boosted	resolved	boosted	resolved	boosted
Low-level input	85.2%	67.9%	83.2%	65.8%	81.6%	65.1%	80.8%	64.3%
Planing away $M_{t\bar{t}}$	76.8%	63.7%	75.3%	62.7%	74.1%	62.1%	73.0%	62.3%

Models	M5Γ1	M5Γ2	M5Γ3	M5Γ4
Kinematic region	boosted	boosted	boosted	boosted
Low-level input	79.5%	78.2%	77.4%	76.8%
Planing away $M_{t\bar{t}}$	65.3%	65.8%	63.7%	64.1%

Table 3.5 The accuracy reach of the chosen neural networks before and after planing away $M_{t\bar{t}}$. The configurations of the DNN's are listed in Table 3.3 and Table 3.4.

3.5 Conclusion

We have found that, in an attempt to develop methods to discover broad $t\bar{t}$ resonances, $M_{t\bar{t}}$ is still one of the most important observables, but additional information from both on- and off-resonance regions can significantly enhance discovery capability. As a result, the cross section upper limits can be improved by $\sim 60\%$ for $\Gamma_\rho/M_\rho \sim 40\%$, and the improved LHC sensitivities do not strongly depend on the width of a resonance. As resonances in new physics beyond the SM are easily broad, our learnings and techniques can be used to efficiently search for them.

The most useful observables turn out to be $M_{t\bar{t}}$ (even for broad resonances), $p_T^{j_{\rm top}}$, $M_{j_{\rm top}}$, angular distributions and colour correlations. The usefulness of $M_{t\bar{t}}$ even for broad-resonance searches is not necessarily obvious, *a priori*. But correlated observables such as $p_T^{j_{\rm top}}$ are found to further complement. Angular information (some of whose contributions come from off-resonance region) and $M_{j_{\rm top}}$ (which can measure colour flow structures irrespective of resonance characteristics) are relatively uncorrelated with the width and $M_{t\bar{t}}$, making improved LHC sensitivities less dependent on the width. Lastly, as we trained using only low-level inputs, our results also show that high-level observables such as $M_{t\bar{t}}$ are effectively well learned by DNN.

We have assessed these machine-learned information in three ways: by explicitly testing those high-level observables, by ranking input (low and/or high) observables using weights of the network, and by planing away features correlated with $M_{t\bar{t}}$. Notably, after all, there can still be unknown useful information that are not easily identified in our analysis. Thus, being able to communicate more efficiently with networks will enable better explorations of the nature, beyond what we know.

Chapter 4

Invisible Higgs Decay

In this chapter, we show the new strategy for searches at hadron collider for the non-hadronic final state with associated jets from initial state radiation(ISR). If a signal process, which is the process of interest we want to find, has a different preference in the parton contents of ISR jets, whether the ISR jets are initiated from quark or gluon[1], compared to its irreducible background, we can apply the jet-tagging technique to isolate the signal from its background in expended phase space. To demonstrate this search strategy, we considered the search for invisible decay of Higgs especially produced from the gluon-fusion process.

4.1 Introduction

After the discovery of Higgs at the LHC in 2012 [1–3, 101–103], the precision measurement of the SM Higgs couplings is one of the most important tasks for probing

[1]This is not a well-defined concept. But for now, let's consider a quark jet or gluon jet as the jet produced from fragmentation and hadronisation of respective quark or gluon.

new physics and the dynamics of electroweak symmetry breaking of the universe at future collider experiments [104–106]. However, Higgs precision measurements are highly non-trivial tasks in the existence of huge irreducible backgrounds which mainly involve the electroweak gauge bosons. In many searches, production mechanisms of electroweak gauge bosons consist of a large portion of those irreducible backgrounds for Higgs signal in the various final states. This is because, first of all, massive bosons, such as $Z/W^{\pm}$, have masses which are similar with the mass of Higgs boson, and second, one or more than one gauge bosons can decay into a pair of fermion-anti-fermion as Higgs boson. Also, Higgs boson decay into a pair of gauge bosons with branching ratio around 20%.

Among various production mechanism of Higgs at the LHC, Higgs production via gluon-fusion(ggH) [107, 108] comprises the dominant contribution(90%) to the total production rate of Higgs at the LHC. ggH is a very unique process in which the the state of QCD force carriers transform into the electroweak bosons via quark loops, and not leaving any other QCD remnants at the leading order (LO), so its event topology can basically be the same with the electroweak vector boson(EWVB) productions from the leading orders. In result, tagging the Higgs from gluon fusion has been suffering from the irreducible backgrounds much more than the other sub-dominant productions including vector boson fusion (VBF), Higgsstrahlung (VH), and $t\bar{t}H$, as it does not have associated objects with fixed particle identity good for tagging the whole process.

For this reason the most stringent constraints for probing the $H\psi\bar{\psi}$ couplings have usually been obtained via the non-ggH processses, e.g. in $H \to b\bar{b}$ [109,110], $c\bar{c}$ [111], $\tau^+\tau^-$ [112–114], $\mu^+\mu^-$ [115,116], e^+e^- [117,118].

The same argument also applies to the searches of Higgs pair production via gluon fusion against the EWVB backgrounds, but in this case, the situation can

get worse as the dominant ggH contribution increases (93%). Here we revisit and generalise an overlooked property, and investigate a new possibility for boosting Higgs searches via ggH. We discuss the differences in the quark-gluon composition of the central ISR jets between the general ggH productions and their irreducible EWVB backgrounds. And by tagging those central gluon jets from ISR, we provide useful discrimination power to overall Higgs searches based. The difference was stated earlier in [119] without no attention, and the new possibility on the difference was claimed in [120] for $H \to \mu\mu$, and studied [121] for a monojet analysis. As the new method can have big impacts, here we generalize the property for (multiple) Higgs and EWVB productions, emphasizing that in the central region of the detector, leading ISR jet from ggH is mostly a gluon jet. To prove its experimental feasibility, we apply the new method in the search for invisible Higgs decays, and show that the limit on the Higgs invisible decay branching ratio can be improved significantly ($60\% \to 5\%$) for the most dominant gluon fusion, to be the most constraining channel, which has been not so useful compared to the other channels at the LHC. In Section 4.2 we discuss the general motivation and quantitative estimation on the leading jet population through the leading order calculation or corresponding amplitudes. We then discuss the relevant phenomenology about invisible Higgs decay briefly in Section 4.3, and introduce the setup for multi-variate analysis in Section 4.4. With the statistical analysis procedure written in Section 4.5, the result is presented in Section 4.6.

4.2 Estimation on Leading Jet Kinematics

Referring to the Figure 4.1, we expect the initial particle of jets as initial state radiation(ISR jets) has a correlation with both factorized parton(PFP) and parton initiated the hard process(PIP). What we call the 'hard process' is the scattering process of interest without any other associated emission. Since the distributions of the

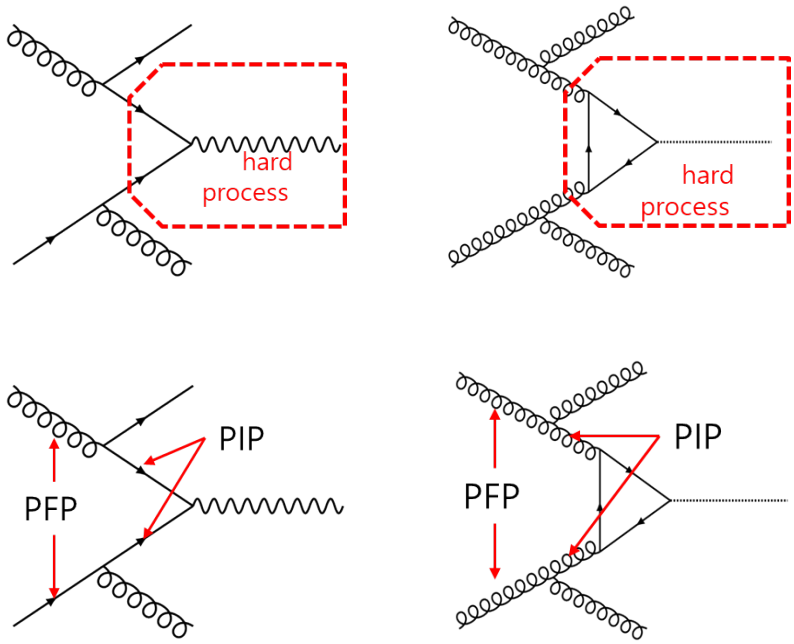


Figure 4.1 Two diagrams on the top are(left) Drell-Yan(DY)-like process and (right)gluon-gluon-funsion-Higgs production(ggH)-like diagrams with associated initial-state radiated(ISR) jets. Here we call the 'hard process' as the scattering process after emission of ISR jets. Two diagrams in the bottom line show our definition of 'parton from proton(PFP)' and 'process-initiating parton(PIP)'. As it can be seen from figures, the particle of PIP and ISR jets are related through the PFP parton.

PIPs at the LHC are already known in terms of parton distribution function(PDF), PIP can be figured out with the PDF as long as we know the particle initiated ISR jet. Therefore, the particles produced the resonance can be found, and also the resonance particle is constrained. This is our goal; to tag 'process', beyond the tagging of each jets using the information of ISR jet. In this study, we did not carry jet tagging for each jet separately but carried multi-variate analysis(MVA) by using 'jet substructure' variables in addition to the conventionally used variable related to the dynamics. Jet substructure variables are selected from a few studies on quark/gluon tagging of jets. Before showing the demonstration of this idea on invisible Higgs search, we establish the motivation with the analytic calculation of the leading jet characteristic from ggH and DY process based on the lowest order matrix element.

4.2.1 Higgs Produced via Gluon-Fusion

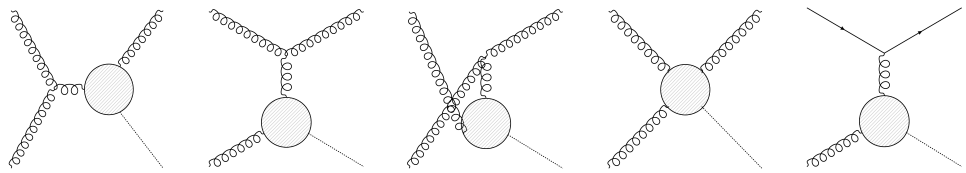


Figure 4.2 Diagrams of gluon-fusion Higgs production(ggH) process with an additional emission. From the left to right, respective diagrams are s, t, u channels of gluon emission from ggH vertex, gluon emission from $gggH$ vertex and quark emission with ggH vertex.

Here we show the kinematic property of the leading quark/gluon jet associated to ggH process. We first consider effective vertexes for ggH and $gggH$ in $m_t \to \infty$ limit for this discussion. The relevant Feynman rules are,

$$ggH \quad : \quad \kappa \frac{g_3^2}{v}(p_2^n p_1^m - p_1 \cdot p_2 \eta^{mn})\delta_{ab} \tag{4.1}$$

$$gggH \quad : \quad -i\kappa \frac{g_3^3}{v} f_{abc} \, (\eta^{mn}(p_1 - p_2)^o + \eta^{on}(p_2 - p_3)^m + \eta^{om}(p_3 - p_1)^n) \tag{4.2}$$

with external gluon fields $\epsilon_m^a(p_1)$, $\epsilon_n^b(p_2)$ and $\epsilon_o^c(p_3)$ with colour indexes a, b and c. And the common κ is purely numerical one which is $-1/(48\pi^2)$.

Taking massless limit for the associated quarks, spin-summed matrix element squares, say $\frac{d}{d\hat{t}}\hat{\sigma}^{(\mathrm{HEFT})}_{gg\to Hq}$ and $\frac{d}{d\hat{t}}\hat{\sigma}^{(\mathrm{HEFT})}_{gq\to Hq}$ for respective $gg \to Hg$ process and $gq \to Hq$ process as in Figure 4.2, are as follow.

$$\frac{d}{d\hat{t}}\hat{\sigma}^{(\mathrm{HEFT})}_{gg\to Hg} = \frac{\alpha_3^3}{144\pi^2\hat{s}^2v^2}\frac{C_A}{N_C^2-1}\frac{\hat{s}^4+\hat{t}^4+\hat{u}^4+m_H^8}{\hat{s}\hat{t}\hat{u}} \tag{4.3}$$

$$\frac{d}{d\hat{t}}\hat{\sigma}^{(\mathrm{HEFT})}_{gq\to Hq} = \frac{\alpha_3^3}{144\pi^2\hat{s}^2v^2}\frac{C_F}{N_C^2-1}\left(-\frac{\hat{s}^2+\hat{u}^2}{\hat{t}}\right) \tag{4.4}$$

These amplitudes could be found in [122]. Here $\alpha_3 = g_3^2/4\pi$, and C_A and C_F are quadratic Casimir of respective adjoint and fundamental representation of $SU(N_C)$. Hence, $C_A = N_C = 3$, and $C_F = T_F(N_C^2-1)/N_C = 4/3$, with Dynkin index $T_F = 1/2$, in case of QCD. Note that we intentionally dropped the contribution from gluon emission from $q\bar{q} \to Hg$ process, because corresponding $q\bar{q}$ parton luminosity at the LHC is $\mathcal{O}(10^{-1})$ smaller than the other parton luminosity.

By naive estimation, $\frac{d}{d\hat{t}}\hat{\sigma}^{(\mathrm{HEFT})}_{gg\to Hg}$ is $C_A/C_F = 2.25$ times larger than $\frac{d}{d\hat{t}}\hat{\sigma}^{(\mathrm{HEFT})}_{gq\to Hq}$. But, as can be seen in the left of Figure 4.3, the parton luminosity of gg is smaller than the parton luminosity of gq at the LHC, and this gives larger weight on $\sigma_{gq\to Hq}$rather than $\sigma_{gq\to Hq}$. In the central region, however, the imbalanced parton luminosity of gq and spin-correlation between initial and final state makes associated quark in $gq \to Hq$ process to be in rather a forward region than where gluon lies in the $gg \to Hg$ process(See the right of Figure 4.3). Combining then all, the gluon jet population in the central region is larger then quark jet population in Higgs production with a jet.

The difference between rapidity distribution of quark jets and gluon jets associated to the ggH process gets significant when one consider dynamical effect of the top-quark loop with finite mass of top quark. Corresponding amplitudes with finite top mass and dynamical loop could be found in [107]. For the case of $gg \to Hg$, the amplitude

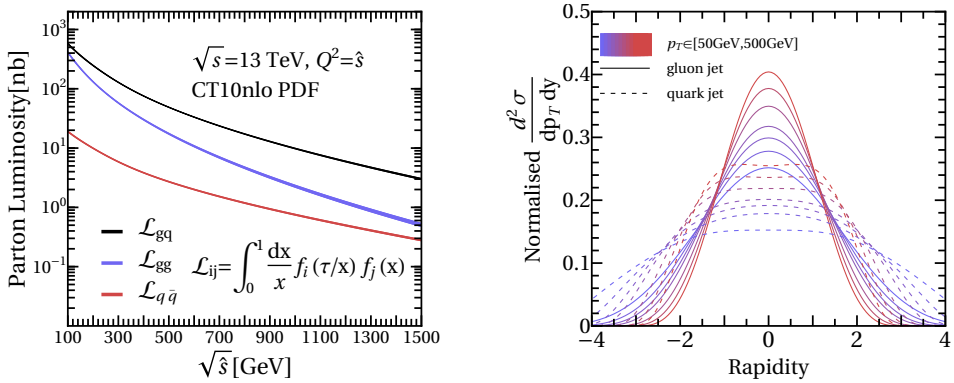


Figure 4.3 The left plot shows the parton luminosity of the LHC at $\sqrt{s} = 13$ TeV. The plot on the right show rapidity distribution of emitted quark(dashed)/gluon(solid) from H+jet process. Colour represents the transverse momenta of emitted quark/gluon, and varied from 50GeV(blue) to 500GeV(red). The formulae used for calculation of differential cross section could be found in Appendix B.

would be,

$$\frac{d}{d\hat{t}}\hat{\sigma}_{gg\to Hg} = \frac{\alpha_3^3}{16\pi^2\hat{s}^2v^2}\frac{4C_A}{N_C^2-1}\frac{m_H^8}{\hat{s}\hat{t}\hat{u}}\Big(\big|A_2(\hat{s},\hat{t},\hat{u})\big|^2+\big|A_2(\hat{t},\hat{u},\hat{s})\big|^2 + \big|A_2(\hat{u},\hat{s},\hat{t})\big|^2+\big|A_4(\hat{s},\hat{t},\hat{u})\big|^2\Big) \tag{4.5}$$

where the definition of the loop functions, A_2 and A_4 are discussed in Appendix C.

Note that, using the asymptotic expansion of $W_{1,2,3}$ around $m_t \to \infty$ restores the HEFT limit in which

$$A_2(s,t,u) \to -\frac{s^2}{3m_H^4} \qquad \text{and} \qquad A_4(s,t,u) \to -\frac{1}{3}, \tag{4.6}$$

as $m_t \to \infty$.

Similarly, the quark jet associated amplitude $d\hat{\sigma}_{gq\to Hq}/d\hat{t}$ of $gq \to Hq$ process would be,

$$\frac{d}{d\hat{t}}\hat{\sigma}_{gq\to Hq} = \frac{\alpha_3^3}{64\pi^2\hat{s}^2v^2}\frac{C_F}{N_C^2-1}\left(-\frac{\hat{s}^2+\hat{u}^2}{\hat{t}}\right)\frac{m_H^4}{(\hat{s}+\hat{u})^2}\big|A_5(\hat{t},\hat{s},\hat{u})\big|^2, \tag{4.7}$$

The argument of the function A_5 in eq. (4.7) is $(\hat{t},\hat{s},\hat{u})$, not $(\hat{s},\hat{t},\hat{u})$, because this A_5 is calculated for $q\bar{q} \to Hg$ amplitude, and we used crossing symmetry to evaluate the

amplitude of $gq \to Hq$ process. Specifically, due to the crossing symmetry, the amplitudes $d\hat{\sigma}_{q\bar{q}\to Hg}/d\hat{t}$ of $q\bar{q} \to Hg$ process and $d\hat{\sigma}_{gq\to Hq}/d\hat{t}$ would satisfy the following relation.

$$\frac{d}{d\hat{t}}\hat{\sigma}_{q\bar{q}\to Hg}(\hat{s},\hat{t},\hat{u}) = -\frac{N_C^2-1}{N_C}\frac{d}{d\hat{t}}\hat{\sigma}_{gq\to Hq}(\hat{t},\hat{s},\hat{u}) \tag{4.8}$$

Figure 4.4 shows the corresponding result. The left side of Figure 4.4 shows the rapidity distributions evaluated with the amplitudes $d\hat{\sigma}_{gg\to Hg}/d\hat{t}$ and $d\hat{\sigma}_{gq\to Hq}/d\hat{t}$, and the one at the right shows the parton level simulation result from MadGrpah5 aMC@NLO. Both plots in Figure 4.4 assume 13 TeV, and all quarks other than top are massless. Also they used the same PDF, CT10NLO with factorisation scale $\mu_F = \sqrt{m_H^2 + p_T^2}$ with Higgs mass m_H and transverse momentum of Higgs, p_T.

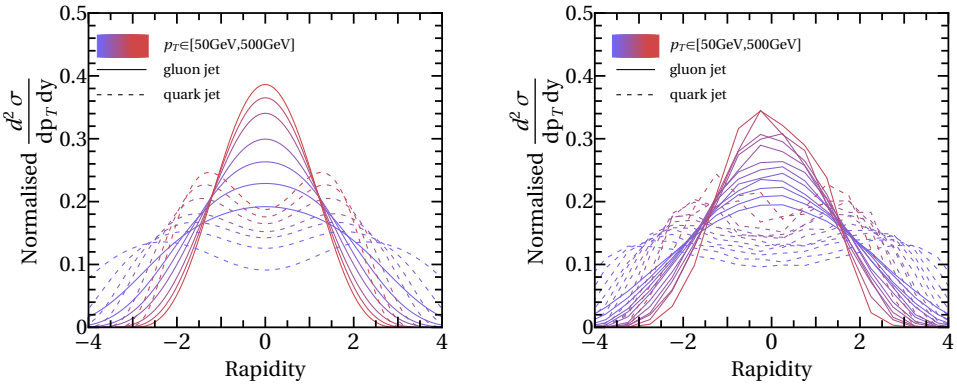


Figure 4.4 The normalised rapidity distribution of associated gluon(solid) and quark(dashed) to Higgs produced via gluon-fusion. The figure on the left panel shows the distribution evaluated from the analytic form of corresponding amplitudes, while the figure on the right panel shows the similar one from the simulated samples at the parton level with aMC@NLO. The details of the simulation set up could be found in the text.

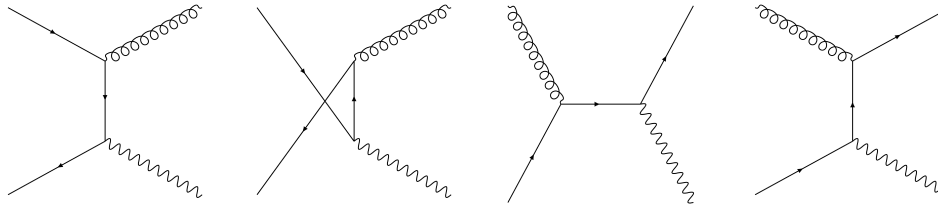


Figure 4.5 Diagrams of DY process with an addtional emission. From the left to right, respective diagrams are t, u channels of gluon emission from $\bar{\psi}\not{G}\psi$ vertex, and quark emission from s, t channels from respective $\bar{\psi}\not{Z}\psi$ and $\bar{\psi}\not{G}\psi$ vertexes.

4.2.2 Massive Gauge Boson Production

In case of DY process, $\frac{d}{d\hat{t}}\hat{\sigma}_{q\bar{q}\to Zg}$, the spin-summed matrix element square of $q\bar{q} \to Zg$ process is [123],

$$\frac{d}{d\hat{t}}\hat{\sigma}_{q\bar{q}\to Zg} = \frac{\alpha_3}{16\hat{s}^2}\frac{C_F}{N_c}\sum_q \left(g_V^{q\,2} + g_A^{q\,2}\right)\frac{\hat{t}^2+\hat{u}^2+2\hat{s}m_Z^2}{\hat{t}\hat{u}}, \tag{4.9}$$

with weak coupling constant g_2, Weinberg angle θ_W and $g_{V/A}$, the vector/axial coupling of quark with Z, i.e,

$$g_V = \frac{1}{\cos\theta_W}\left(\frac{1}{2}T_3 - Q\sin^2\theta_W\right) \quad \text{and} \quad g_A = \frac{1}{2\cos\theta_W}T_3, \tag{4.10}$$

with third-component of weak-isospin generator T_3 and electromagnetic charge Q for each fermion. The summation($\sum_q$) is done for all quarks except top quark.

$d\hat{\sigma}_{gq\to Zq}/d\hat{t}$, the similar one for $gq \to Zq$ process is [123],

$$\frac{d}{d\hat{t}}\hat{\sigma}_{gq\to Zq} = \frac{\alpha_3}{16\hat{s}^2}\frac{T_F}{N_c}\sum_q \left(g_V^{q\,2} + g_A^{q\,2}\right)\left(-\frac{\hat{s}^2+\hat{u}^2+2\hat{t}m_Z^2}{\hat{s}\hat{u}}\right). \tag{4.11}$$

The difference between associated colour factor comes from the initial state. Even though those two process have the crossing symmetry, $q\bar{q} \to Zg$ will summed over all possible external gluon at the final state, and have $1/N_C^2$ factor from taking the average of initial state colours. Therefore, it has following factor.

$$\frac{1}{N_C^2}\delta^{ab}\mathrm{tr}(t_a t_b) = \frac{1}{N_C}T_F\frac{N_C^2-1}{N_C} = \frac{C_F}{N_C}, \tag{4.12}$$

where $t_{a,b}$ are the generator of fundamental representation.

On the other hand, for $gq \to Zq$, averaging factor comes from one gluon and one quark, so corresponding colour factor would be

$$\frac{1}{N_C(N_C^2-1)}\delta^{ab}\mathrm{tr}(t_a t_b) = \frac{T_F}{N_C}. \tag{4.13}$$

This makes $C_F/T_F \simeq 2.7$ factor enhancement on $q\bar{q} \to Zg$ compared to $gq \to Zq$. But still, as can be seen in the left of Figure 4.3, the parton luminosity of $q\bar{q}$ is $\mathcal{O}(10^{-1})$ times smaller than the parton luminosity of gq. Also, due to crossing symmetry, there is no big difference in dynamics of associated jets(Figure 4.6), and, hence, quark jet population is larger in Z-boson production.

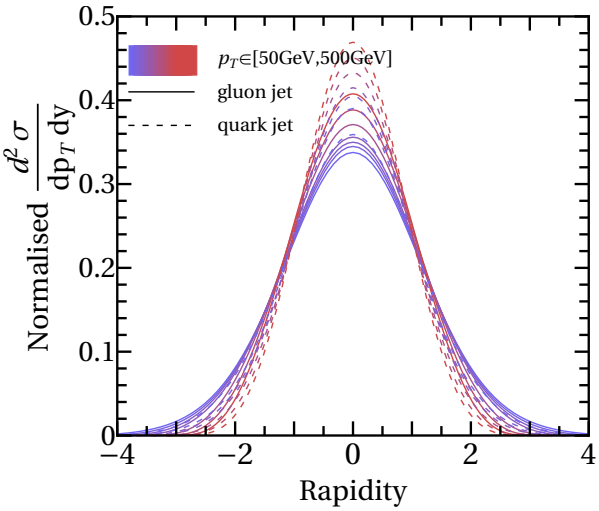


Figure 4.6 The rapidity distribution of emitted quark(dashed)/gluon(solid) from Z+jet(right) process. Colour represents the transverse momenta of emitted quark/gluon, and varied from 50GeV(blue) to 500GeV(red).

4.2.3 Multiple Production

Here we generalise the discussion into the case of multi-Higgs production and multi-vector-boson production. Figure 4.7 shows the leading diagrams of general

Higgs signal productions from gluon fusion (ggH^n+jets) in bold lines and EWVB productions (V^n+jets) as irreducible backgrounds with an emission of ISR(s), for three different initial parton configurations (gq, gg, $q\bar{q}$). Here the n can be larger than one for multi Higgs/EWVB productions, and additional Higgs or EWVB productions with $n \geq 2$ is also represented. The gluon lines in grey indicate extra gluon emissions and the V^n+jets diagram (bottom-centre) from gg initial states is drawn also in grey as it is sub-leading to the other 5 diagrams in α_s.

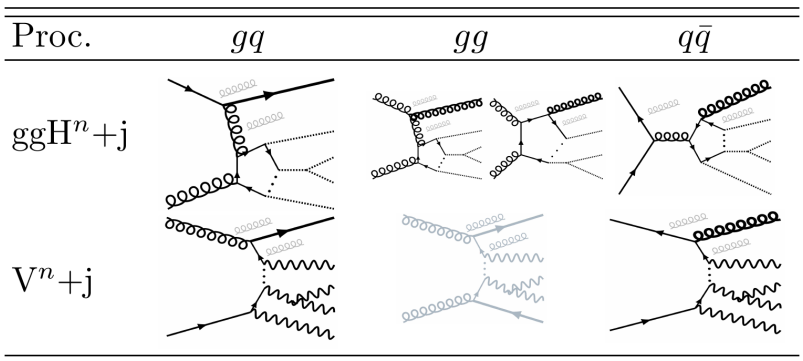


Figure 4.7 Leading diagrams (bold) of the (multi) Higgs productions from gluon fusion (ggH^n+jets), against the corresponding irreducible (multi) EWVB backgrounds (V^n+jets) with additional ISR(s) for 3 parton initial states (gq,gg,$q\bar{q}$).

It should be noted that the flavour of ISRs emitted in the leading diagrams are uniquely fixed since the other final state (H^n/V^n) accompanied with is a colourless non-QCD particle in the 2 to (n+1) processes, for the given initial parton configurations. It is also noticeable that the parton luminosity functions, $\mathcal{L}_{gq,gg,q\bar{q}}$ referring to the left of Figure 4.3, for the three initial partonic states are hierarchical - $\mathcal{L}_{gq} > \mathcal{L}_{gg} \gg \mathcal{L}_{q\bar{q}}$, e.g. $\mathcal{L}_{gq}$:$\mathcal{L}_{gg}$:$\mathcal{L}_{q\bar{q}} \sim$ 2:1:0.07 at $\sqrt{\hat{s}} \sim 100$ GeV, and such a hierarchy persists to higher energy scale as shown in the same plot. Based on these above two observations, the dominant flavour of leading ISR jet from the whole ggH^n+jets and V^n+jets processes can be predicted and their quark-gluon compositions can be compared.

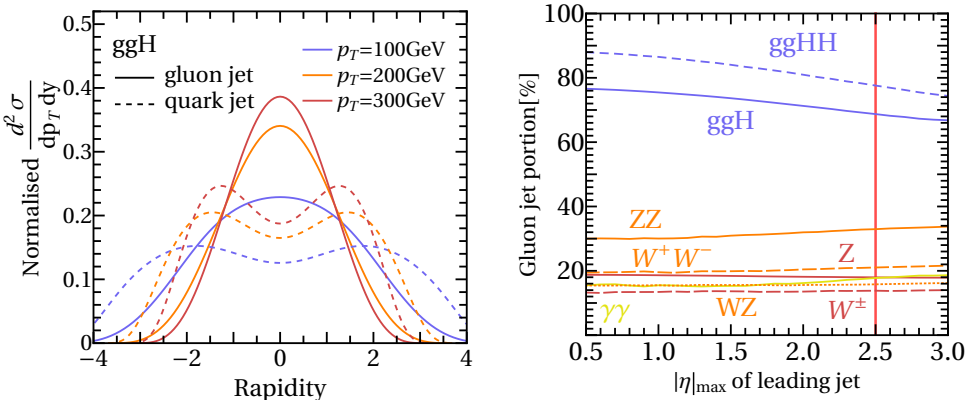


Figure 4.8 The left panel shows the rapidity distribution of leading gluon and quark ISR jets from ggH. The right panel shows the gluon portion over $|\eta^{j_1}|_{max}$ ($p_T^{j_1} > 100$ GeV) of the leading ISR jet associated with ggH^n and V^n productions. Here, for the truth tagging, we used internally developed algorithm as depicted in Appendix D.

For the V^n+jets EWVB processes, the leading ISR jet is mostly quark jet. Even though there is non-negligible gluon jet contributions from, the first, the emissions in subleading diagrams to the leading V^n+jets(gq,gg), and, the second, the emission from the V^n+jets($q\bar{q}$) at leading order with suppression by small parton luminosity, the leading ISR jet is expected most likely to be a quark jet mainly because of the hierarchical parton luminosities between $gq + gg$ and $q\bar{q}$ initial states. Gluon jet portion, R^g, of the leading ISR ($|\eta^{j_1}| \leq 1$ and $p_T^{j_1} > 100$ GeV), are $R^g_{(WW,WZ,ZZ)} \approx$ (0.20,0.16,0.30) for V$^{n=2}$+jets processes, $R^g_{(W,Z)} \approx$ (0.13,0.19) for V$^{n=1}$+jets, and (3) $R^g_{\gamma\gamma} \approx 0.15$ for prompt di-photon+jet processes, as in the right of Figure 4.8.

The leading ISR jet is more gluon like in the ggH^n+jets processes. There exist two enhancement factors of the gluon jet leading ggH^n(gg)+jets process ($gg \to Hg$ at LO) over the quark jet leading, ggH^n(gq)+jets ($qg \to Hq$ at LO). In the diagrams of the ggH^n(gg)+jets in Figure 4.7, a gluon is emitted as the leading ISR from the gluon parton or from the top loop, and all of the results end up with the enhancement of QCD colour factor ratio, $C_A/C_F = 9/4$ in the cross-section compared to the

ggH$^n(gq)$+jets, where C_A (C_F) is the quadratic Casimir of the adjoint (fundamental) representation. Furthermore, in comparison to the quark ISR from the ggH$^n(gq)$+jets, the gluon ISR from the ggH$^n(gg)$+jets is more likely to be emitted in central region with small absolute rapidity as in the left of Figure 4.8 (from the cross section of gluon-fusion Higgs production at NLO [107, 108] with finite top mass effect), due to the balanced momentum profile of the gg initial states, while in addition the quark ISR prefers to be aligned in beam direction (backward) more than the gluon, with spin angular momentum conservation, showing the rapidity-gaps in the quark ISR distributions. Although the parton luminosity function in Figure 4.3 is larger for the gq state compared to the gg (by a factor of 2 at $\sqrt{\hat{s}} \sim 100$ GeV), all in all the leading ISR from the ggH^n+jets becomes more gluon jet like in central region with the R^g of the leading ISR ($|\eta^{j_1}| \leq 1$ and $p_T^{j_1} > 100$ GeV), are $R^g_{ggH} \approx 0.75$ for ggH$^{n=1}$+jets, and $R^g_{ggHH} \approx 0.87$ for a pair of Higgs production, as in the right of Figure 4.8.

If p_T cut is lowered to $p_T^{j_1} > 50$ GeV, the R^g for the ggH^n(V^n) process increases(decreases), respectively, by $\sim$ 2-5% in the $|\eta^{j_1}|_{max}$ range of the right panel in Figure 4.8.

Like as the two quark jets in the forward region from the VBF, and as the extra Z/W from the VH, now the ggH^n + jets also has such a unique property accompanied the gluon-like ISR jets in the central region. In this regard, if some relevant techniques using quark-gluon tagging of ISR jets are employed, one can improve the constraints from the most dominant gluon fusion channel for a broad range of Higgs signatures which are buried in the irreducible EWVB backgrounds matched with.

4.3 Phenomenology of Invisible Decay of Higgs

As mentioned at the beginning of the chapter, we focus on the search for invisible decay of Higgs to test the experimental feasibility of using the gluonic ISR jets.

Historically, there had been lots of studies on the possibility of invisible Higgs decays along with the developments of the Standard Model and the beyond. Based on the standard model, the only possible invisible decay of Higgs is the final state with two pairs of $\nu\bar{\nu}$, through two Z bosons, and corresponding branching ratio is almost $\mathcal{O}(10^{-3})$. However, beyond the standard model, in principle, could have extra invisible channels so that the branching ratio into the invisible final state can deviate from the standard model prediction. The early proposals include the models in diversity e.g. with Majorons [124,125], supersymmetries [126–135], heavy neutrinos with radiatively generated masses [136], large extra dimensions [137–140], the 4th generations [141–143], and so on, while the recent interpretations are mainly based on the effective singlet extensions of the SM, in the context of so-called Higgs-portal models with dark matters [144–146]. For these proposals, there exist numerous phenomenological and experimental researches in search for the invisible Higgs decays, via the production channels, including VH [119, 147, 148], $t\bar{t}H$ [149, 150], ggH (monojet) [119, 151, 152], and VBF [153–156] which has been presented the most sensitive limits on the invisible Higgs decay BR at the LHC.

There also have been interesting surveys via diffractive Higgs productions [157], total decay width [158], Higgs rare B decays [159], di-Higgs [160, 161], Higgs off-shell decays [162–164], lepton colliders [165–168], including global analysis [169–174]. Throughout the searches, the dominant ggH channel has never been competitive to the other sub-dominant channels. However, employing the new method we show that the most stringent constraint can be obtained from the ggH channel for the invisible Higgs decays, as is demonstrated in the latter part. More detailed reference through this section can be found in [20].

4.4 Data Preparation and Multi-variate Analysis

Assuming the Higgs production cross-section of the SM, we perform the analysis in the search for the invisible Higgs decays in E_T^{miss}+jets signature via the ggH+jets channel. Samples are generated by Monte Carlo simulated proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV at the LHC, for 36 fb^{-1} using MadGraph5 aMC@NLO v2.6.2 [73] interfaced with Pythia v8.235 [80] for hadronisation and fragmentation. Delphes v.3.4.1 is used for detector simulation [81]. The signal process (ggH+jets) is generated with up to extra 1 jet at LO taking into account finite top mass effects [175] with $M_H = 125$ GeV, and backgrounds are generated at NLO in QCD. We use FxFx scheme with k_T-algorithm and $\Delta R = 1$ for jet merging [176]. For jet clustering, FastJet v3.2.1 [177] is used with anti-k_T algorithm with $\Delta R = 0.4$, and CT10NLO [178] is used for parton distribution function.

Among the relevant background processes - $V(Z(\nu\nu), W(\ell\bar{\nu}))$+jets, Diboson, top quarks, $Z/\gamma \to \ell\bar{\ell}$, QCD multijets, where the leptons (ℓ) in $W/Z/\gamma$ decays are misidentified, we only included the most dominant irreducible EWVB backgrounds - V+jets, while the others take just $O(1)\%$ level for the event selection criteria as follows [179]:

- $p_T^{j_1} > 100$GeV, $|\eta^{j_1}| < 2.5$, $E_T^{\text{miss}} > 200$GeV, $\min_{j\in\{\text{jets}\}} \Delta\phi(\vec{p}_T^{\,\text{miss}}, \vec{p}_T^{\,j}) \geq 0.5$, $N_{\text{jet}} \geq 1$.

The 1st (2nd) cut on the transverse momentum (pseudorapidity) of the leading jet is imposed to suppress all of the backgrounds, the 3rd cut on the missing transverse energy is mainly to reduce the QCD and top quarks, and the 4th cut with the missing transverse momentum, $\vec{p}_T^{\,\text{miss}}$ suppresses the QCD multijets very efficiently [148].

There also exist contributions from other Higgs productions, VBF and VH with yield rates (ggH:VBF:VH $\sim$ 70:20:10%). However as the leading jets from VBH are

most likely quark jets opposed to the gluonic leading jets in the ggH+jets, we checked that the VBF can be easily separated from the ggH+jets by tagging gluonic central leading jet in addition to the forward jet tagging for VBF. As for the VH which also has quark-jet like leading ISRs according to the same argument with the V+jets, it can have additional selection criteria [87, 180] for identifying jets from hadronically decaying vector bosons. In this regard, to demonstrate the main idea without making event selection scheme too complicated, we simply consider the ggH+jets as the only signal versus the V+jets as the main background in this analysis, without loss of consistency in applying the flavour information for discrimination of gluon-jet rich ggH signal from general quark-jet rich backgrounds.

We use a set of jet substructure variables [181], say $S^{\,\mathrm{jet}}$, in our analysis as the following,

- $S^{\,\mathrm{jet}} \equiv \{$ n_{tk} (track multiplicity) [13], Girth [12,13], Broadening [9], EEC (energy-energy correlation) [17] with $\beta = 0.2$ [14], RMS-p_T [13] $\}$,

which contain the information on jet flavours. Here the definition of those 5 jet substructure variables, track multiplicity(n_{tk}), girth(G), broadening(B), EEC(C_1^{β},) with $\beta = 0.2$ and RMS-p_T of jet are in eq. (4.14).

$$
\begin{aligned}
\text{girth} : G &= \frac{1}{p_T^{\,\mathrm{jet}}} \sum_{i \in \{\mathrm{const.}\}} p_T^i \, |\Delta \vec{r}_i| \\
\text{broadening} : B &= \frac{1}{\sum_i |\vec{p}^{\,i}|} \sum_i \left| \vec{p}^{\,i} \times \hat{p}^{\,\mathrm{jet}} \right| = \frac{1}{\sum_i |\vec{p}^{\,i}|} \sum_i \left| \vec{k}_T^{\,i} \right| \\
\text{EEC} : C_1^{\beta} &= \frac{1}{\left(\sum_i p_T^i\right)^2} \sum_{i<j} p_T^i p_T^j \, (\Delta R_{ij})^{\beta} \\
\text{RMS-}p_T : \sqrt{\langle p_T^2 \rangle} &= \frac{1}{p_T^{\,\mathrm{jet}}} \sqrt{\frac{1}{n_{\mathrm{tk}}} \sum_i \left(p_T^i\right)^2}
\end{aligned}
\tag{4.14}
$$

The set {const.} is the set of 'constituents' of jet, so that in includes the tracks of which jet consists. The track multiplicity($n_{\rm tk}$) could be defined as $\sum_{i\in\{\rm const.\}} 1$, in this sense.

This part, in general could be replaced with the other inputs from raw data, such as jet images [15,16] for deep learning. Among the five jet substructure variables used, the Girth as the linear radial moment of a jet reflects a fatness/radius of a jet. As gluon jets tend to have more showers and be fatter by the colour factor enhancement, $C_A(g \to gg)/C_F(q \to gq)$, such a property can be checked in the Girth distribution of the leading jet from ggH and V+jet processes in the top right panel of Figure 4.9. For the other jet substructure variables, broadening, EEC, RMS-p_T and track multiplicity, see Figure 4.10.

Jet substructure observables have been used to build a jet tagger, $\mathrm{P}_{q/g}(S^{\rm jet})$, while the kinematic observables, such as reconstructed four-momenta of jets have been used to build an event classifier, $\mathrm{P}_{\rm S/B}(\{p^{\rm jet}, ...\})$. However, as can be seen from $d^2\sigma/dp_T^{\rm jet}dy^{\rm jet}$ in the left of Figure 4.8, the flavour of a jet can have a correlation with kinematic information depending on the scattering process. This observation motivates us to build $\mathrm{P}_{\rm S/B}(\{p^{\rm jet}, ...\} \cup S^{\rm jet})$, rather than a factorized classifier, $\mathrm{P}_{\rm S/B}(\{p^{\rm jet}, ...\}) \otimes \mathrm{P}_{q/g}(S^{\rm jet})$.

As results, for respective ggH signal and V+jets background processes, Figure 4.9 shows the normalized distribution of $E_T^{\rm miss}$(top left), Girth(top right), $\mathrm{P}_{\rm S/B}(S^{\rm jet})$(bottom left), and $\mathrm{P}_{\rm S/B}$ ($S^{\rm jet} \cup \{E_T^{\rm miss}, p_T^{\rm jet}, \eta^{\rm jet}\}$)(bottom right).

The two event classifiers $\mathrm{P}_{\rm S/B}$ in the bottom left of Figure 4.9 and the bottom right of Figure 4.9 are obtained by training neural networks with the specified input features as above. The number of data, method of preprocessing on the data, number of layers, nodes, drop out rate and the other details of DNN structure can be found in Table 4.1.

	$\mathbf{P}_{\mathrm{S/B}}(\mathbf{S}^{\mathrm{jet}})$	$\mathbf{P}_{\mathrm{S/B}}(\mathbf{S}^{\mathrm{jet}} \cup \{\mathrm{kin.}\})$
Training data	1M(0.5M each for sig/bg)	
Validation data	1M(0.5M each for sig/bg)	
Preprocessing	Standard-Scaler	
Environment	Keras [91] with TensorFlow backend	
NN structure	Sequentially connected Dense Layers	
Batch normalisation	Used	
Drop out	30%	10%
NN Structure	2 layers with 300 nodes each	4 layers with 200 nodes each
Optimiser	Adam	
Loss Function	Categorical cross-entropy	
Learning rate	0.001	0.001
Batch size	50,000	
Activation Function	ReLU for intermediate layers, soft-max for output layer	
Initialisation	HE normal	

Table 4.1 Specification of the DNN structure. Here the number of layer in NN structure does not include output layer.

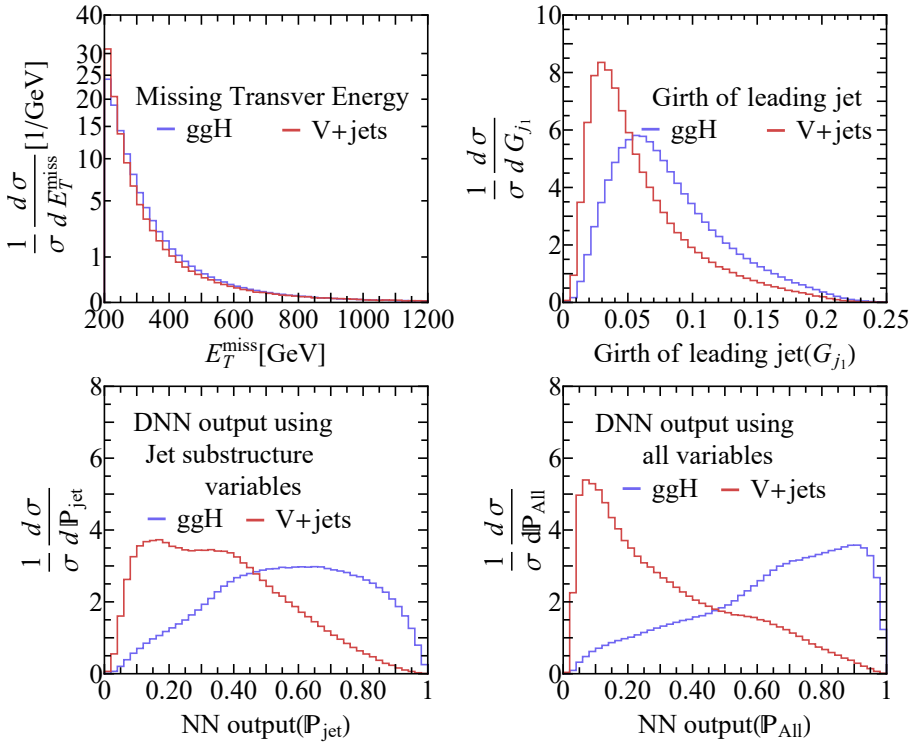


Figure 4.9 Signal and background profiles in various templates, (a) $E_T^{\rm miss}$, (b) Girth of leading jet, (c) event classifier $\mathrm{P_{S/B}}(S^{\rm jet})$ (1:ggH-like, 0:V+jet) trained using the jet substructure observables $S^{\rm jet}$, and (d) $\mathrm{P_{S/B}}$ ($S^{\rm jet} \cup$ {kin.}) using all features.

We also shows the receiver operating characteristic(ROC) curves and $\epsilon_{\rm sig}/\sqrt{\epsilon_{\rm bg}}$ derived from the ROC curve at Figure 4.11.

It is noticeable that the event classifier using the set of jet substructures alone can provide a much better separation of signal and background compared to the one $E_T^{\rm miss}$ as in the bottom left of Figure 4.9. Combining them all we get the best separation as is clearly seen in the bottom right of Figure 4.9.

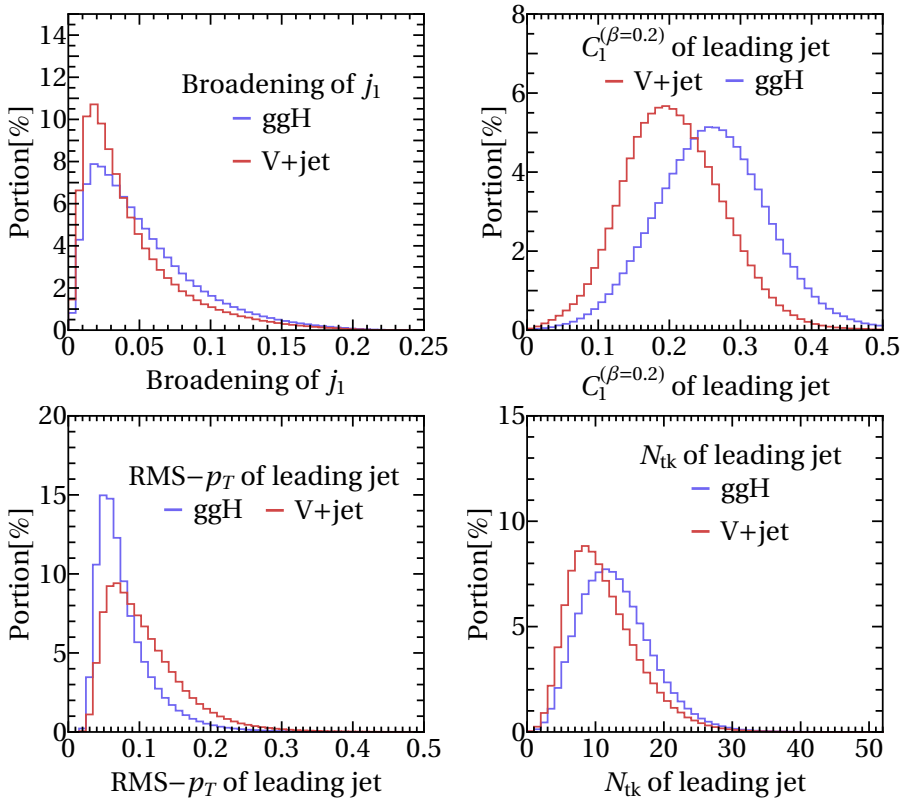


Figure 4.10 Signal and background profiles in various jet substructure variables; the broadening(top left), EEC(top right), RMS-p_T(bottom left) and the track multiplicity(bottom right).

4.5 Analysis Method

The result obtained the previous section can be used to discover invisible Higgs decay or put constraints on the invisible branching ratio of Higgs. After selecting the events with the criteria, we performed the profile likelihood ratio test following the procedure in [182] with the four template distributions in Figure 4.9. The likelihood function is given,

$$\mathcal{L} = \prod_{i=1}^{N_{\mathrm{bin}}} \frac{\hat{n}_i^{n_i}}{n_i!} e^{-\hat{n}_i} \times \frac{1}{2\pi} e^{-\frac{1}{2}\left(\theta_s^2+\theta_b^2\right)}, \tag{4.15}$$

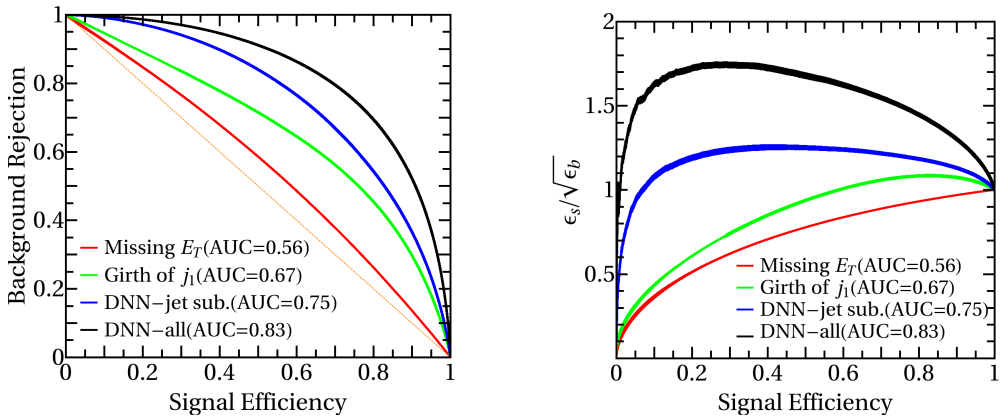


Figure 4.11 Receiver operating characteristic(ROC) curves(left) and $\epsilon_{\rm sig}/\sqrt{\epsilon_{\rm bg}}$ curves(right) derived from each ROC curve. In each curve, ROC derived from $E_T^{\rm miss}$ distribution in drawn with red, and one from girth is green. $\mathrm{P_{S/B}(S^{\,jet})}$ and $\mathrm{P_{S/B}(S^{\,jet} \cup \{kin.\})}$ are blue and black colour, respectively.

where n_i is the number of events (or pseudo events) in i-th bin, and $\hat{n}_i$ is the number of expected events with branching ratio parameter $\mu = \frac{\sigma}{\sigma_{\rm SM}} \times BR(h \to \text{inv})$ (production x-section of invisibly decaying Higgs over the total x-section of the SM Higgs - $\sigma_{\rm SM}$), i.e.,

$$\hat{n}_i = \mu N_s P_s(i)(1+f_s)^{\theta_s} + N_b P_b(i)(1+f_b)^{\theta_b}. \tag{4.16}$$

Here the $P_{s(b)}(i)$ is the expected event rate in i-th bin, given the total number of events $N_{s(b)}$ survived the cut, and the $\theta_{s(b)}$ in the Poisson and prior probabilities denotes a nuisance parameter associated to the systematic uncertainty $f_{s(b)}$, of signal (background). As a global variation of event rates in signal and background distributions, we tested the $f_{s(b)}$ in 5-20% [179], which changes median of expected upper limit in their 3% at most, and set $f_{s(b)}$ to 10%. More details of statistical method can be found in Appendix A. Signal cross-section (ggH+X) is taken from [183] computed at NNLO+NNLL QCD and NLO EW, and we applied the efficiency on the selection criteria evaluated using simulated event samples, for the fiducial signal yield. For the background process (V+jets) we take both of the cross-section and efficiency from

our MC simulation of the Z+jets, and the fiducial background yield was obtained by a K-factor (1.53) for the Z+jet, to take the W+jets into account simply, reproducing the expected limit using missing transverse energy [148].

4.6 Result and Conclusion

We obtained the upper limits on the branching ratio of the invisible Higgs decay in 95% confidence level(C.L.) with the integrated luminosity 36fb^{-1} at the LHC. The upper limits are evaluated via profile likelihood ratio test with the four templates from Figure 4.9; 'Missing E_T', 'Girth'(of leading jets), 'jet substructure variables' and ''All variables'. There are the results from the four templates in the right panel of Figure 4.12, while the existing experimental results [184] are in the left panel. The left-most column, 'combined', from the left panel of Figure 4.12 shows the combined results from the CMS experiment. Besides 'combined', there are two columns, 'VBF-tag' and 'ggH-tag', which are the results evaluated with different selection criteria optimised for the two production mechanism, VBF and ggH, respectively. The jet flavour information is not used for any results in the left panel, while the three templates other than 'Missing E_T' from the right panel use the new feature of the jet flavour. 'Missing E_T' and 'ggH-tag', two red-boxed columns are directly matched with each other since they are evaluated with the same selection criteria and the same template. Hence the results from the other three templates can be compared to the existing limit from the ggH-tagged events, and summarised in the second row of Table 4.2. The errors in the table are rescaled with factors matching two error band from 'ggH-tag' and 'Missing E_T', as they can be directly compared to each other. From the result, the limit on Higgs invisible decays from the ggH is significantly improved from 60% down to 5% if sub-jet level information of the leading ISR is employed for the analysis. It is interesting that the jet substructures alone provide stronger constraints

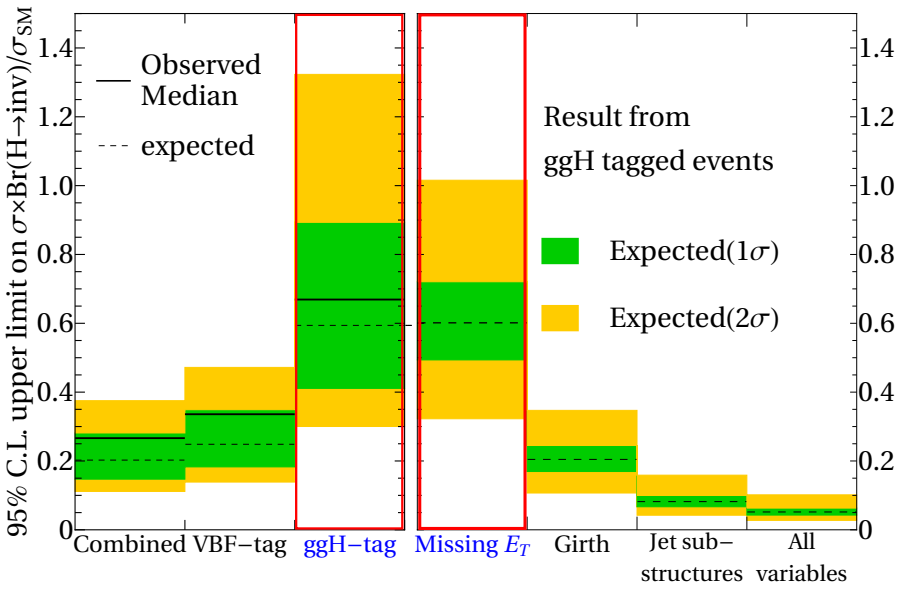


Figure 4.12 Upper limit in 95% of confidence level(C.L.) on $\frac{\sigma}{\sigma_{\rm SM}} \times {\rm Br}(H \rightarrow {\rm inv})$ with the integrated luminosity 36fb^{-1}. Here we used 5,000 ensembles of pseudo data set which consists of background events only. Left panel shows the results from the experiment [184].

(8%) than the missing transverse energy in Higgs invisible search. Moreover, combining the features in two kinds, we end up with the best sensitive result (5%) only from the ggH, much lower than the one obtained from VBF ($\sim$ 20%). To address the impact of the one-binned systematic uncertainties, we also considered the bin-by-bin freely varying systematic uncertainties. The similar reference value of the upper limit from the Missing E_T can be maintained with 5% error for all individual bins for the respective ggH signal and V+jet background. The resulted upper limits are shown in the third row of Table 4.2. Compared to the one-binned 10% systematic uncertainties, the result using 'Missing E_T' template only changes by 1%. On the other hand, the result using Girth template increases more than 50%, and also the other two templates give less sensitive results with 5% bin-by-bin systematic uncertainties. Still

the best upper limit comes from the template using 'All variables' with the neural network, which is around 7%.

Systematic Uncertainty	E_T^{miss}	Girth	DNN(jet sub)	DNN(all)
Correlated 10%	$60.2^{+30.0}_{-18.3}$ %	$20.4^{+10.1}_{-5.99}$ %	$8.3^{+4.46}_{-2.55}$ %	$5.2^{+2.83}_{-1.54}$ %
Un-correlated 5%	59.0%	35.1%	12.2%	7.4%

Table 4.2 Summary on the upper limits in 95% of confidence level on $\sigma/\sigma_{\text{SM}} \times \text{BR}(\text{H} \rightarrow \text{inv.})$ for the integrated luminosity 36 fb^{-1} at the LHC, from the four template distributions in Figure 4.9. The errors were rescaled by the factors projecting the 'Missing ET' band to the reference 'ggH-tag' band in Figure 4.12 for the 'Correlated' systematic uncertainty. The first line, (Correlated 10%) shows the result with 10% systematic uncertainties on the shape of the respective ggH signal and V+jet background. The second line (Un-correlated 5%) shows the result with 5% systematic uncertainties for individual bins.

Though more sophisticated understanding and treatment of systematic errors are necessary to obtain a firm number for the expected limit on the Higgs invisible decays, the exercise we did in this study strongly suggests that 1% (2%) precision for Higgs invisible branching ratio at the end of the LHC running with 3 ab^{-1} (300 fb^{-1}) is a plausible expectation from the gluon fusion solely. To understand the limitation of the analysis, let's consider a simple example with one-bin analysis to set the bound on cross section $\sigma(i \rightarrow f)$. In that case, the signal fraction can never be bounded below the fluctuation of the background background events. More precisely, if we consider the signal and background with the expected total cross sections σ_{sig} and

$\sigma_{\rm bkg}$, respectively, the branching fraction, μ=$\sigma(i \to f)/\sigma_{\rm sig}$ has lower bound[2]

$$\sigma(i \to f) = \mu\sigma_{\rm sig} \geq f_{\rm sys}\sigma_{\rm bkg}. \tag{4.17}$$

In the analysis with more than one bin, the lower bound at i-th bin with $P_{\rm sig}(i)$ portion of the signal and $P_{\rm bkg}(i)$ portion of the background changes as

$$\mu \geq \frac{\sigma_{\rm bkg}P_{\rm bkg}(i)}{\sigma_{\rm sig}P_{\rm sig}(i)}f_{\rm sys}. \tag{4.18}$$

The lower limit, therefore, can be mitigated in bins with $P_{\rm bkg}(i)/P_{\rm sig}(i)$ larger than the unity. Corresponding lower bounds in individual bin from the four templates in Figure 4.9 are shown in Figure 4.13. As in the figure, the lower bounds of the upper limit from individual bins in the 'Missing E_T' template are already higher than 100%. It indicates that there is no single bin dominates the signal/background discrimination, and the final upper bound($\sim$60%) is achieved with all 50 bins together. On the contrary, the results with neural network outputs are dominated by few bins at which the signal portion is much larger than the background portion. In those bins, the lower bound with is $\mathcal{O}(1) \times f_{\rm sys}$ %. The 'Girth j_1' template is similar with neural network outputs, so that it has bins with manifestly smaller lower limits compared to the other bins, yet the difference is not significant. Note that Figure 4.13 does not give any prediction on the final upper limit. It just shows the possible lower limit of the upper bound from every single bins in each template, if a bin is used for the analysis. The final number of the upper bound is affected by all 50 bins, and the correlation between the bins. So that it is impossible to estimate the final results from neither Figure 4.9 nor Figure 4.13. Still, Figure 4.13 can be used for the validation of the results in Table 4.2. From the figure, the minimum of lower limit of the upper bounds

[2]This assumes that the integrated luminosity or the total cross section of the background process is large enough so that the statistical fluctuation can be ignored. Otherwise, the fluctuation due to respective systematic uncertainty and statistics contributes at the same time, and the lower limit of μ is changed as $\sigma_{\rm sig}^{-1}\sqrt{f_{\rm sys}^2\sigma_{\rm bkg}^2 + \mathcal{L}_{\rm int}^{-1}\sigma_{\rm bkg}}$ with the integrated luminosity $\mathcal{L}_{\rm int}$.

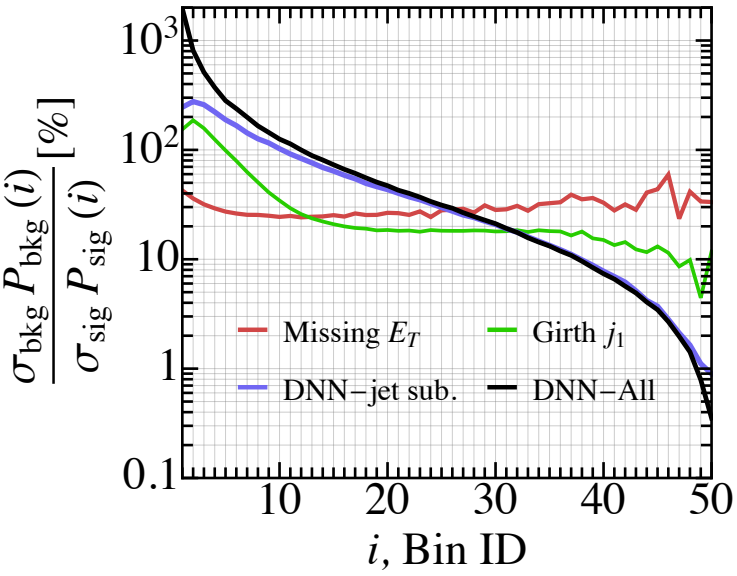


Figure 4.13 The bin-by-bin lower bound from the four templates from Figure 4.9. Here the bin-by-bin lower bound is given by $(\sigma_{\rm bkg}P_{\rm bkg}(i))$ / $(\sigma_{\rm sig}P_{\rm sig}(i))$. This corresponds to the case with $f_{\rm sys}$=1%. The efficiency of the selection criteria is taken into account for the two total cross section values for the ggH signal and the V+jets background. Note that the x-axis is the ordering of the bin, the labels of the bins, and does not have any other indication.

with $f_{\rm sys}$=10% are around 200%, 40%, 10% and 3% from the respective 'Missing E_T', 'Girth j_1', $\mathbb{P}_{\rm jet}$ and $\mathbb{P}_{\rm All}$ templates. The final results in Table 4.2 show the similar hierarchy with the one between the upper bounds from respective templates, and, the results are not absurd.

We revisited and generalized the property - *the gluon rich leading ISR jets in central rapidity region from gluon fusion Higgs productions versus the quark rich EWVB backgrounds*, and proposed the idea *to improve general Higgs searches produced from ggH by tagging the central gluonic ISR jets*. Applying the new method to the searches of invisible decays of Higgs, we showed that the ggH can be the best channel with the improved limit on invisible Higgs decay branching ratio (60% $\rightarrow$ 5%), significantly exceeding the best limit given by the other channels - VBF ($\sim$ 20%) and VH

($\sim 40\%$). It is also expected that the limit can significantly be improved again if it is combined with the results from VBF and other processes. The physics and methods in this analysis can also be applied to a broad range of new resonance and Higgs productions induced by gluon fusion, e.g. in search for exotic/rare Higgs decays and di-Higgs productions, concurrently with their irreducible EWVB backgrounds mostly containing quark jet dominant ISRs.

Chapter 5

Conclusion

In this thesis, we considered two possible application of machine learning on searches at LHC. Specifically, we focused on the searches without narrow resonance peak from the final state. There are two possibilities of the absence of narrow resonance. The first case is when the width of the resonance particle is not small compared to its pole mass. In such case, Breit-Wigner suppression due to width is not severe and the signal is distributed over a wide range of resonance energy. Furthermore, as the pole mass gets heavier, the differential cross-section with respect to resonance energy at IR regime increases due to large parton luminosity. The IR enhancement makes the mass pole insignificant and the signal similar to the continuum background at the hadron collider. Another possible case is the resonance search in the invisible final state. In this case, there is no information from the final state at all. The existence of the signal can be proven only through the measurement of the deviation of the data from the expected background.

To study the first case, searching for a broad resonance, we choose a benchmark

scenario with heavy vector resonance with mass $\sim \mathcal{O}(1)$ TeV which is widely expected from minimal composite Higgs models. Focusing on $t\bar{t}$ final state, we considered 1 TeV and 5 TeV resonance and 10%, 20%, 30% and 40% of Γ/m for respective mass. We considered two kinematic regions for 1 TeV depending on whether the b quark and W boson formed jets independently or made a single jet since large boost of the top quark, while 5 TeV resonance scenario includes boosted region only. We tried three different methods to examine what is the best observables to discriminate the broad resonance signal from its continuum background in this benchmark analysis, using machine learning with deep neural network structure.

The first method is to compare the result from two deep neural network structures, while the one is trained only with simple detector-level(low-level) observables and the other one having high-level observables in addition to the low-level observables. The high-level observables are physically motivated observables which require the additional procedures, such as combinatoric optimisation or tagging, to reconstruct. If the enhancement exists along with the adding observables, it implies that the high-level observables provide additional information relevant for the signal discrimination. In all benchmark points of mass, width and kinematic regions, we checked that there is enhancement from adding observables but it is $\mathcal{O}(1\%)$. Also, the enhancement tends to be large in the boosted region compared to the resolved region. This means that, in the resolved region, the information of high-level observables are already reflected by low-level observables, while in the high-level observables, some of it is lost while passing selection criteria.

The second method is raking all observables by a measure, the variable importance, which is defined as a normalised square-sum of all weights between the corresponding input node and all the other nodes at the first hidden layer. From the definition, an observable with larger weights is considered to have more importance

than the others with relatively small weights. We checked that $M_{t\bar{t}}$ is top-ranked in resolved kinematic regions, while the angular distributions take great importance compared to the $M_{t\bar{t}}$, and p_T of top-jet has even higher importance than the other two.

The last approach is the so-called planing method, which is an intentional drop-out of information from an observable. If the planing of an observable deteriorates the discrimination power to zero, or 50% accuracy, it indicates that the all information solely comes from the observable planed away. To check the residual information from other than $M_{t\bar{t}}$, we planned the $M_{t\bar{t}}$ from the data and recast the training neural networks. As results, we check that there still is discrimination power in data with planned away $M_{t\bar{t}}$. The degree of deterioration was larger in the resolved region for 1 TeV resonance, and the degree was similar for 5 TeV resonance.

Combining them all, we conclude that the invariant mass or resonance energy is still useful in the case of broad resonance, while p_T or angular distributions also have great importance on the search.

For the second study with the invisible final state, we first examined the correlation between the partonic contents of leading ISR jets and characteristic of resonance particle, such as spin, colour charge etc. This means that the parton content of ISR jet can be used for signal/background discrimination in certain condition. The condition requires that the signal and its background have different populations of quark/gluon for the ISR, and all jets can be tagged as ISRs. As a benchmark analysis, we considered the search for the invisible decay of Higgs. In searching for invisible decay of Higgs, the gluon-fusion production channel does not set strong bound on the $\frac{\sigma}{\sigma_{\rm SM}} \times \mathrm{Br}(H \to \mathrm{inv})$, since it has no characteristic feature compared to the background, $Z \to \nu\bar{\nu}$. The other production mechanism, such as vector-boson-fusion and Higgs-strahlung, have stronger bound compared to the gluon-fusion channel since they have respective

handles; two forward jets for vector-boson-fusion and fat-jet from massive boson in Higgs-strahlung. However, using parton contents of leading ISR jets as additional information by including jet substructure variables with the deep neural network, we improved the bound from gluon-fusion channel even strong than vector-boson-fusion channel which gives the strongest bound in current experiment status.

The reason why the machine learning technique is adopted in two analysis is to exploit many physical observables at once in the analysis. As a multi-variate analysis tool, machine learning with the deep neural network structure abstracts the feature from the space of input observables and projects the multi-dimensional distributions into a single direction, the neural network output. The neural network output from a properly trained neural network is the most powerful discriminant between the signal and the background processes.

There are many directions for the future works. Here we would briefly discuss two among them. The first one is about the machine learning architecture. Yet what the neural networks had learned is the non-linear distributions of the signal and the background process in multi-dimensional space of input observables. The machines did not learn about the physics nor nature, and it is not certain whether the trained neural networks is valid outside the coverage of training data set. This makes the interpretation of the trained neural networks limited. Therefore we need an architecture which is possible to learn the governing dynamics and which we can interpret.

The second direction is hadron physics. As in [121], some of effective description on jets does not match with the others. The discrepancy between effective approaches would remain as theoretical uncertainties, and it would limit the sensitivity of the study. For this we need better understanding on QCD and relevant effective descriptions.

Bibliography

[1] ATLAS collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1 [1207.7214].

[2] CMS collaboration, S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30 [1207.7235].

[3] CMS collaboration, S. Chatrchyan et al., *Observation of a New Boson with Mass Near 125 GeV in pp Collisions at* $\sqrt{s}$ *= 7 and 8 TeV*, *JHEP* **06** (2013) 081 [1303.4571].

[4] H. W. Babcock, *The rotation of the Andromeda Nebula*, *Lick Observatory Bulletin* **498** (1939) 41.

[5] M. Markevitch, A. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman et al., *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, *Astrophys. J.* **606** (2004) 819 [astro-ph/0309303].

[6] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones et al., *A direct empirical proof of the existence of dark matter*, *Astrophys. J. Lett.* **648** (2006) L109 [astro-ph/0608407].

[7] E. Hayashi and S. D. White, *How Rare is the Bullet Cluster?*, *Mon. Not. Roy. Astron. Soc.* **370** (2006) L38 [astro-ph/0604443].

[8] Planck collaboration, Y. Akrami et al., *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *Astron. Astrophys.* **641** (2020) A1 [1807.06205].

[9] S. Catani, G. Turnock and B. R. Webber, *Jet broadening measures in e^+e^- annihilation*, *Phys. Lett.* **B295** (1992) 269.

[10] T. Becher and G. Bell, *NNLL Resummation for Jet Broadening*, *JHEP* **11** (2012) 126 [1210.0580].

[11] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073 [1704.08249].

[12] J. Gallicchio, J. Huth, M. Kagan, M. D. Schwartz, K. Black and B. Tweedie, *Multivariate discrimination and the Higgs + W/Z search*, *JHEP* **04** (2011) 069 [1010.3698].

[13] J. Gallicchio and M. D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [1106.3076].

[14] A. J. Larkoski, G. P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [1305.0007].

[15] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [1407.5675].

[16] P. T. Komiske, E. M. Metodiev and M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110 [1612.01551].

[17] B. Bhattacherjee, S. Mukhopadhyay, M. M. Nojiri, Y. Sakaki and B. R. Webber, *Quark-gluon discrimination in the search for gluino pair production at the LHC*, *JHEP* **01** (2017) 044 [1609.08781].

[18] K. Joshi, A. D. Pilkington and M. Spannowsky, *The dependency of boosted tagging algorithms on the event colour structure*, *Phys. Rev.* **D86** (2012) 114016 [1207.6066].

[19] S. Jung, D. Lee and K.-P. Xie, *Beyond $M_{t\bar{t}}$: learning to search for a broad $t\bar{t}$ resonance at the LHC*, *Eur. Phys. J. C* **80** (2020) 105 [1906.02810].

[20] W. S. Cho, H. D. Kim and D. Lee, *Boosting Invisible Higgs Searches by Tagging a Gluon Jet for Gluon Fusion Process*, *Phys. Rev. D* **102** (2020) 115007 [2003.06822].

[21] R. T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning Multivariate New Physics*, 1912.12155.

[22] ATLAS collaboration, A. Ghosh, *Deep generative models for fast shower simulation in ATLAS*, *J. Phys. Conf. Ser.* **1525** (2020) 012077.

[23] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264.

[24] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, *Physical Review* **117** (1960) 648.

[25] J. Goldstone, *Field theories with Superconductor solutions*, *Il Nuovo Cimento* **19** (1961) 154.

[26] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (1963) 531.

[27] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, *Progress of Theoretical Physics* **49** (1973) 652.

[28] R. D. Peccei and H. R. Quinn, *CP conservation in the presence of pseudoparticles*, *Phys.Rev.Lett.* **38** (1977) 1440.

[29] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, *Sov. Phys. JETP* **46** (1977) 641.

[30] V. Gribov and L. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438.

[31] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys. B* **126** (1977) 298.

[32] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31.

[33] D. Barducci, A. Belyaev, S. De Curtis, S. Moretti and G. M. Pruna, *Exploring Drell-Yan signals from the 4D Composite Higgs Model at the LHC*, *JHEP* **04** (2013) 152 [1210.2927].

[34] D. Greco and D. Liu, *Hunting composite vector resonances at the LHC: naturalness facing data*, *JHEP* **12** (2014) 126 [1410.2883].

[35] D. Barducci and C. Delaunay, *Bounding wide composite vector resonances at the LHC*, *JHEP* **02** (2016) 055 [1511.01101].

[36] D. Liu, L.-T. Wang and K.-P. Xie, *Broad composite resonances and their signals at the LHC*, *Phys. Rev. D* **100** (2019) 075021 [1901.01674].

[37] R. Kelley, L. Randall and B. Shuve, *Early (and Later) LHC Search Strategies for Broad Dimuon Resonances*, *JHEP* **02** (2011) 014 [1011.0728].

[38] S. Ask, J. H. Collins, J. R. Forshaw, K. Joshi and A. D. Pilkington, *Identifying the colour of TeV-scale resonances*, *JHEP* **01** (2012) 018 [1108.2396].

[39] ATLAS collaboration, M. Aaboud et al., *Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Eur. Phys. J.* **C78** (2018) 565 [1804.10823].

[40] K. J. F. Gaemers and F. Hoogeveen, *Higgs Production and Decay Into Heavy Flavors With the Gluon Fusion Mechanism*, *Phys. Lett.* **146B** (1984) 347.

[41] D. Dicus, A. Stange and S. Willenbrock, *Higgs decay to top quarks at hadron colliders*, *Phys. Lett.* **B333** (1994) 126 [hep-ph/9404359].

[42] N. Craig, F. D'Eramo, P. Draper, S. Thomas and H. Zhang, *The Hunt for the Rest of the Higgs Bosons*, *JHEP* **06** (2015) 137 [1504.04630].

[43] S. Jung, J. Song and Y. W. Yoon, *Dip or nothingness of a Higgs resonance from the interference with a complex phase*, *Phys. Rev.* **D92** (2015) 055009 [1505.00291].

[44] S. Y. Choi, J. Kalinowski, Y. Liao and P. M. Zerwas, *H/A Higgs mixing in CP-noninvariant supersymmetric theories*, *Eur. Phys. J.* **C40** (2005) 555 [hep-ph/0407347].

[45] J. R. Ellis, J. S. Lee and A. Pilaftsis, *CERN LHC signatures of resonant CP violation in a minimal supersymmetric Higgs sector*, *Phys. Rev.* **D70** (2004) 075010 [hep-ph/0404167].

[46] M. Carena and Z. Liu, *Challenges and opportunities for heavy scalar searches in the $t\bar{t}$ channel at the LHC*, *JHEP* **11** (2016) 159 [1608.07282].

[47] CMS collaboration, A. M. Sirunyan et al., *Search for resonant $\mathrm{t\bar{t}}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Submitted to: JHEP* (2018) [1810.05905].

[48] CMS collaboration, A. M. Sirunyan et al., *Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles*, *JHEP* **08** (2018) 130 [1806.00843].

[49] ATLAS collaboration, M. Aaboud et al., *Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett.* **B759** (2016) 229 [1603.08791].

[50] ATLAS collaboration, M. Aaboud et al., *Search for new high-mass phenomena in the dilepton final state using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **10** (2017) 182 [1707.02424].

[51] D. Liu, L.-T. Wang and K.-P. Xie, *Prospects of searching for composite resonances at the LHC and beyond*, *JHEP* **01** (2019) 157 [1810.08954].

[52] J. H. Collins, K. Howe and B. Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning*, *Phys. Rev. Lett.* **121** (2018) 241803 [1805.02664].

[53] J. H. Collins, K. Howe and B. Nachman, *Extending the search for new resonances with machine learning*, *Phys. Rev.* **D99** (2019) 014038 [1902.02634].

[54] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty Detection Meets Collider Physics*, 1807.10261.

[55] P. Baldi, P. Sadowski and D. Whiteson, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, *Nature Commun.* **5** (2014) 4308 [1402.4735].

[56] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, *Eur. Phys. J.* **C76** (2016) 235 [1601.07913].

[57] H. Luo, M.-x. Luo, K. Wang, T. Xu and G. Zhu, *Quark jet versus gluon jet: deep neural networks with high-level features*, 1712.03634.

[58] J. Lee, N. Chanon, A. Levin, J. Li, M. Lu, Q. Li et al., *Polarization fraction measurement in same-sign WW scattering using deep learning*, *Phys. Rev. D* **99** (2019) 033004 [1812.07591].

[59] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet Constituents for Deep Neural Network Based Top Quark Tagging*, 1704.02124.

[60] J. Guo, J. Li, T. Li, F. Xu and W. Zhang, *Deep learning for R-parity violating supersymmetry searches at the LHC*, *Phys. Rev.* **D98** (2018) 076017 [1805.10730].

[61] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [1511.05190].

[62] G. Li, Z. Li, Y. Wang and Y. Wang, *Improving the measurement of Higgs boson-gluon coupling using Convolutional Neural Networks at e^+e^- colliders*, 1901.09391.

[63] K. Fraser and M. D. Schwartz, *Jet Charge and Machine Learning*, *JHEP* **10** (2018) 093 [1803.08066].

[64] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, *JHEP* **01** (2019) 057 [1702.00748].

[65] M. Abdughani, J. Ren, L. Wu and J. M. Yang, *Probing stop pair production at the LHC with graph neural networks*, *JHEP* **08** (2019) 055 [1807.09088].

[66] J. Ren, L. Wu and J. M. Yang, *Unveiling CP property of top-Higgs coupling with graph neural networks at the LHC*, *Phys. Lett. B* **802** (2020) 135198 [1901.05627].

[67] I. Henrion, J. Brehmer, J. Bruna, K. Cho, K. Cranmer, G. Louppe et al., *Neural message passing for jet physics*, .

[68] D. Guest, K. Cranmer and D. Whiteson, *Deep Learning and its Application to LHC Physics*, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161 [1806.11484].

[69] M. Abdughani, J. Ren, L. Wu, J. M. Yang and J. Zhao, *Supervised deep learning in high energy phenomenology: a mini review*, 1905.06047.

[70] D. Liu and R. Mahbubani, *Probing top-antitop resonances with $t\bar{t}$ scattering at LHC14*, *JHEP* **04** (2016) 116 [1511.09452].

[71] H. An, X. Ji and L.-T. Wang, *Light Dark Matter and Z' Dark Force at Colliders*, *JHEP* **07** (2012) 182 [1202.2894].

[72] 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 [1310.1921].

[73] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *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 [1405.0301].

[74] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014) 2930 [1112.5675].

[75] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through* $O(\alpha_S^4)$, *Phys. Rev. Lett.* **110** (2013) 252004 [1303.6254].

[76] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, *JHEP* **01** (2013) 080 [1210.6832].

[77] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, *JHEP* **12** (2012) 054 [1207.0236].

[78] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to* $q\bar{q} \to t\bar{t} + X$, *Phys. Rev. Lett.* **109** (2012) 132001 [1204.5201].

[79] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, *Phys. Lett.* **B710** (2012) 612 [1111.5869].

[80] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

[81] DELPHES 3 collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [1307.6346].

[82] ATLAS collaboration, M. Aaboud et al., *Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *JHEP* **12** (2018) 039 [1807.11883].

[83] *Performance of jet substructure techniques in early* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2015-035, CERN, Geneva, Aug, 2015.

[84] *Boosted hadronic top identification at ATLAS for early 13 TeV data*, Tech. Rep. ATL-PHYS-PUB-2015-053, CERN, Geneva, Dec, 2015.

[85] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, *JHEP* **07** (2008) 092 [0806.0023].

[86] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, *Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks*, *Phys. Rev. Lett.* **101** (2008) 142001 [0806.0848].

[87] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [1011.2268].

[88] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [1108.2701].

[89] T. Plehn and M. Spannowsky, *Top Tagging*, *J. Phys.* **G39** (2012) 083001 [1112.4441].

[90] G. Kasieczka, T. Plehn, T. Schell, T. Strebler and G. P. Salam, *Resonance Searches with an Updated Top Tagger*, *JHEP* **06** (2015) 203 [1503.05921].

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

[92] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean et al., *Tensorflow: a system for large-scale machine learning.*, in *OSDI*, vol. 16, pp. 265–283, 2016.

[93] D. Yang and Q. Li, *Probing the Dark Sector through Mono-Z Boson Leptonic Decays*, *JHEP* **02** (2018) 090 [1711.09845].

[94] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011.

[95] J. C. Collins and D. E. Soper, *Angular Distribution of Dileptons in High-Energy Hadron Collisions*, *Phys. Rev.* **D16** (1977) 2219.

[96] E. Richter-Was and Z. Was, *Separating electroweak and strong interactions in Drell–Yan processes at LHC: leptons angular distributions and reference frames*, *Eur. Phys. J.* **C76** (2016) 473 [1605.05450].

[97] D. E. Soper and M. Spannowsky, *Finding physics signals with event deconstruction*, *Phys. Rev.* **D89** (2014) 094005 [1402.1189].

[98] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press USA, 2015.

[99] T. Roxlo and M. Reece, *Opening the black box of neural nets: case studies in stop/top discrimination*, 1804.09278.

[100] S. Chang, T. Cohen and B. Ostdiek, *What is the Machine Learning?*, *Phys. Rev.* **D97** (2018) 056009 [1709.10106].

[101] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321.

[102] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132.

[103] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508.

[104] K. Cheung, J. S. Lee and P.-Y. Tseng, *Higgs Precision (Higgcision) Era begins*, *JHEP* **05** (2013) 134 [1302.3794].

[105] M. Cepeda et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*, vol. 7, pp. 221–584. 12, 2019. 1902.00134. 10.23731/CYRM-2019-007.221.

[106] J. de Blas et al., *Higgs Boson Studies at Future Particle Colliders*, *JHEP* **01** (2020) 139 [1905.03764].

[107] R. Ellis, I. Hinchliffe, M. Soldate and J. van der Bij, *Higgs Decay to tau+ tau-: A Possible Signature of Intermediate Mass Higgs Bosons at the SSC*, *Nucl. Phys. B* **297** (1988) 221.

[108] A. Djouadi, M. Spira and P. Zerwas, *Production of Higgs bosons in proton colliders: QCD corrections*, *Phys. Lett. B* **264** (1991) 440.

[109] ATLAS collaboration, M. Aaboud et al., *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, *Phys. Lett. B* **786** (2018) 59 [1808.08238].

[110] CMS collaboration, A. M. Sirunyan et al., *Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) 421 [1809.10733].

[111] ATLAS collaboration, M. Aaboud et al., *Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment*, *Phys. Rev. Lett.* **120** (2018) 211802 [1802.04329].

[112] CMS collaboration, A. M. Sirunyan et al., *Search for the associated production of the Higgs boson and a vector boson in proton-proton collisions at $\sqrt{s} = 13$ TeV via Higgs boson decays to τ leptons*, *JHEP* **06** (2019) 093 [1809.03590].

[113] ATLAS collaboration, M. Aaboud et al., *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **99** (2019) 072001 [1811.08856].

[114] CMS collaboration, A. M. Sirunyan et al., *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector*, *Phys. Lett. B* **779** (2018) 283 [1708.00373].

[115] CMS collaboration, A. M. Sirunyan et al., *Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **122** (2019) 021801 [1807.06325].

[116] ATLAS collaboration, M. Aaboud et al., *Search for the dimuon decay of the Higgs boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *Phys. Rev. Lett.* **119** (2017) 051802 [1705.04582].

[117] CMS collaboration, V. Khachatryan et al., *Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC*, *Phys. Lett. B* **744** (2015) 184 [1410.6679].

[118] ATLAS collaboration, G. Aad et al., *Search for the Higgs boson decays $H \to ee$ and $H \to e\mu$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **801** (2020) 135148 [1909.10235].

[119] D. Choudhury and D. P. Roy, *Signatures of an invisibly decaying Higgs particle at LHC*, *Phys. Lett.* **B322** (1994) 368 [hep-ph/9312347].

[120] D. Lee, W. S. Cho and H. D. Kim, *Boosting the Searches of Gluon Pair Initiated Processes using Deep Learning of ISR Jets*, *PoS* **ICHEP2018** (2019) 796.

[121] G. Kasieczka, N. Kiefer, T. Plehn and J. M. Thompson, *Quark-Gluon Tagging: Machine Learning vs Detector*, *SciPost Phys.* **6** (2019) 069 [1812.09223].

[122] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, *Phys. Rept.* **457** (2008) 1 [hep-ph/0503172].

[123] C. G. Tully, *Elementary particle physics in a nutshell*. Princeton University Press, 2011.

[124] R. E. Shrock and M. Suzuki, *Invisible Decays of Higgs Bosons*, *Phys. Lett.* **110B** (1982) 250.

[125] A. S. Joshipura and S. D. Rindani, *Majoron models and the Higgs search*, *Phys. Rev. Lett.* **69** (1992) 3269.

[126] K. Griest and H. E. Haber, *Invisible Decays of Higgs Bosons in Supersymmetric Models*, *Phys. Rev. D* **37** (1988) 719.

[127] J. F. Gunion and H. E. Haber, *Higgs Bosons in Supersymmetric Models. 3. Decays Into Neutralinos and Charginos*, *Nucl. Phys. B* **307** (1988) 445.

[128] A. Djouadi, P. Janot, J. Kalinowski and P. Zerwas, *SUSY decays of Higgs particles*, *Phys. Lett. B* **376** (1996) 220 [hep-ph/9603368].

[129] A. Djouadi and M. Drees, *Higgs boson decays into light gravitinos*, *Phys. Lett. B* **407** (1997) 243 [hep-ph/9703452].

[130] G. Belanger, F. Boudjema, A. Cottrant, R. M. Godbole and A. Semenov, *The MSSM invisible Higgs in the light of dark matter and g-2*, *Phys. Lett.* **B519** (2001) 93 [hep-ph/0106275].

[131] M. Hirsch, J. Romao, J. Valle and A. Villanova del Moral, *Invisible Higgs boson decays in spontaneously broken R-parity*, *Phys. Rev. D* **70** (2004) 073012 [hep-ph/0407269].

[132] J.-J. Cao, Z. Heng, J. M. Yang and J. Zhu, *Higgs decay to dark matter in low energy SUSY: is it detectable at the LHC ?*, *JHEP* **06** (2012) 145 [1203.0694].

[133] A. Butter, T. Plehn, M. Rauch, D. Zerwas, S. Henrot-Versillé and R. Lafaye, *Invisible Higgs Decays to Hooperons in the NMSSM*, *Phys. Rev. D* **93** (2016) 015011 [1507.02288].

[134] R. K. Barman, G. Belanger, B. Bhattacherjee, R. Godbole, G. Mendiratta and D. Sengupta, *Invisible decay of the Higgs boson in the context of a thermal and nonthermal relic in MSSM*, *Phys. Rev. D* **95** (2017) 095018 [1703.03838].

[135] K. Wang and J. Zhu, *Funnel annihilations of light dark matter and the invisible decay of the Higgs boson*, *Phys. Rev. D* **101** (2020) 095028 [2003.01662].

[136] A. Pilaftsis, *Radiatively induced neutrino masses and large Higgs neutrino couplings in the standard model with Majorana fields*, *Z. Phys. C* **55** (1992) 275 [hep-ph/9901206].

[137] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263 [hep-ph/9803315].

[138] K. R. Dienes, E. Dudas and T. Gherghetta, *Neutrino oscillations without neutrino masses or heavy mass scales: A Higher dimensional seesaw mechanism*, *Nucl. Phys. B* **557** (1999) 25 [hep-ph/9811428].

[139] N. Arkani-Hamed, S. Dimopoulos, G. Dvali and J. March-Russell, *Neutrino masses from large extra dimensions*, *Phys. Rev. D* **65** (2001) 024032 [hep-ph/9811448].

[140] G. F. Giudice, R. Rattazzi and J. D. Wells, *Graviscalars from higher dimensional metrics and curvature Higgs mixing*, *Nucl. Phys. B* **595** (2001) 250 [hep-ph/0002178].

[141] D. Fargion, Y. Golubkov, M. Khlopov, R. Konoplich and R. Mignani, *Possible effects of the existence of the fourth generation neutrino*, *JETP Lett.* **69** (1999) 434 [astro-ph/9903086].

[142] V. A. Khoze, *Comment on an invisible Higgs boson and 50 GeV neutrino*, hep-ph/0105069.

[143] K. Belotsky, D. Fargion, M. Khlopov, R. Konoplich and K. Shibaev, *Invisible Higgs boson decay into massive neutrinos of fourth generation*, *Phys. Rev.* **D68** (2003) 054027 [hep-ph/0210153].

[144] V. Silveira and A. Zee, *SCALAR PHANTOMS*, *Phys. Lett. B* **161** (1985) 136.

[145] H. Davoudiasl, R. Kitano, T. Li and H. Murayama, *The New minimal standard model*, *Phys. Lett. B* **609** (2005) 117 [hep-ph/0405097].

[146] A. Djouadi, A. Falkowski, Y. Mambrini and J. Quevillon, *Direct Detection of Higgs-Portal Dark Matter at the LHC*, *Eur. Phys. J. C* **73** (2013) 2455 [1205.3169].

[147] ATLAS collaboration, M. Aaboud et al., *Search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a Z boson in pp collisions at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector*, *Phys. Lett. B* **776** (2018) 318 [1708.09624].

[148] CMS collaboration, A. M. Sirunyan et al., *Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at* $\sqrt{s} = 13$ TeV, *Phys. Rev.* **D97** (2018) 092005 [1712.02345].

[149] J. F. Gunion, *Detecting an invisibly decaying Higgs boson at a hadron supercollider*, *Phys. Rev. Lett.* **72** (1994) 199 [hep-ph/9309216].

[150] *First constraints on invisible Higgs boson decays using* $\mathrm{t\bar{t}H}$ *production at* $\sqrt{s} = 13$ TeV, 3, 2019.

[151] CMS collaboration, S. Chatrchyan et al., *Search for New Physics with a Mono-Jet and Missing Transverse Energy in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *Phys. Rev. Lett.* **107** (2011) 201804 [1106.4775].

[152] ATLAS collaboration, G. Aad et al., *Search for new phenomena with the monojet and missing transverse momentum signature using the ATLAS detector in $\sqrt{s} = 7$ TeV proton-proton collisions*, *Phys. Lett. B* **705** (2011) 294 [1106.5327].

[153] O. J. P. Eboli and D. Zeppenfeld, *Observing an invisible Higgs boson*, *Phys. Lett.* **B495** (2000) 147 [hep-ph/0009158].

[154] H. Davoudiasl, T. Han and H. E. Logan, *Discovering an invisibly decaying Higgs at hadron colliders*, *Phys. Rev. D* **71** (2005) 115007 [hep-ph/0412269].

[155] ATLAS collaboration, G. Aad et al., *Search for invisible decays of a Higgs boson using vector-boson fusion in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **01** (2016) 172 [1508.07869].

[156] CMS collaboration, *Search for invisible decays of a Higgs boson produced via vector boson fusion at $\sqrt{s} = 13$ TeV.*, .

[157] K. Belotsky, V. Khoze, A. Martin and M. Ryskin, *Can an invisible Higgs boson be seen via diffraction at the LHC?*, *Eur. Phys. J. C* **36** (2004) 503 [hep-ph/0406037].

[158] I. Low, P. Schwaller, G. Shaughnessy and C. E. Wagner, *The dark side of the Higgs boson*, *Phys. Rev. D* **85** (2012) 015009 [1110.4405].

[159] C. Kim, S. C. Park, K. Wang and G. Zhu, *Invisible Higgs decay with $B \to K$ nu anti-nu constraint*, *Phys. Rev. D* **81** (2010) 054004 [0910.4291].

[160] S. Banerjee, B. Batell and M. Spannowsky, *Invisible decays in Higgs boson pair production*, *Phys. Rev. D* **95** (2017) 035009 [1608.08601].

[161] C. Englert and J. Jaeckel, *Probing the Symmetric Higgs Portal with Di-Higgs Boson Production*, *Phys. Rev. D* **100** (2019) 095017 [1908.10615].

[162] M. Endo and Y. Takaesu, *Heavy WIMP through Higgs portal at the LHC*, *Phys. Lett. B* **743** (2015) 228 [1407.6882].

[163] N. Craig, H. K. Lou, M. McCullough and A. Thalapillil, *The Higgs Portal Above Threshold*, *JHEP* **02** (2016) 127 [1412.0258].

[164] M. Ruhdorfer, E. Salvioni and A. Weiler, *A Global View of the Off-Shell Higgs Portal*, *SciPost Phys.* **8** (2020) 027 [1910.04170].

[165] O. J. Eboli, M. Gonzalez-Garcia, A. Lopez-Fernandez, S. Novaes and J. Valle, *Searching for an invisibly decaying Higgs boson in e^+e^-, $e\gamma$ and $\gamma\gamma$ collisions*, *Nucl. Phys. B* **421** (1994) 65 [hep-ph/9312278].

[166] ILC collaboration, G. Aarons et al., *International Linear Collider Reference Design Report Volume 2: Physics at the ILC*, 0709.1893.

[167] Y. Kato, *Probing the dark sector via searches for invisible decays of the Higgs boson at the ILC*, 2, 2020, 2002.12048.

[168] Y.-L. Tang, C. Zhang and S.-h. Zhu, *Invisible Higgs Decay at the LHeC*, *Phys. Rev. D* **94** (2016) 011702 [1508.01095].

[169] C. Englert, T. Plehn, M. Rauch, D. Zerwas and P. M. Zerwas, *LHC: Standard Higgs and Hidden Higgs*, *Phys. Lett. B* **707** (2012) 512 [1112.3007].

[170] J. R. Espinosa, M. Muhlleitner, C. Grojean and M. Trott, *Probing for Invisible Higgs Decays with Global Fits*, *JHEP* **09** (2012) 126 [1205.6790].

[171] G. Belanger, B. Dumont, U. Ellwanger, J. F. Gunion and S. Kraml, *Status of invisible Higgs decays*, *Phys. Lett.* **B723** (2013) 340 [1302.5694].

[172] T. Corbett, O. J. P. Eboli, D. Goncalves, J. Gonzalez-Fraile, T. Plehn and M. Rauch, *The Higgs Legacy of the LHC Run I*, *JHEP* **08** (2015) 156 [1505.05516].

[173] GAMBIT collaboration, P. Athron et al., *Global analyses of Higgs portal singlet dark matter models using GAMBIT*, *Eur. Phys. J. C* **79** (2019) 38 [1808.10465].

[174] C. Arina, A. Beniwal, C. Degrande, J. Heisig and A. Scaffidi, *Global fit of pseudo-Nambu-Goldstone Dark Matter*, *JHEP* **04** (2020) 015 [1912.04008].

[175] V. Hirschi and O. Mattelaer, *Automated event generation for loop-induced processes*, *JHEP* **10** (2015) 146 [1507.00020].

[176] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061 [1209.6215].

[177] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896 [1111.6097].

[178] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin et al., *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010) 074024 [1007.2241].

[179] CMS collaboration, V. Khachatryan et al., *Searches for invisible decays of the Higgs boson in pp collisions at $\sqrt{s}$ = 7, 8, and 13 TeV*, *JHEP* **02** (2017) 135 [1610.09218].

[180] S. D. Ellis, C. K. Vermilion and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev.* **D81** (2010) 094023 [0912.0033].

[181] A. J. Larkoski, I. Moult and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, *Phys. Rept.* **841** (2020) 1 [1709.04464].

[182] *Procedure for the LHC Higgs boson search combination in summer 2011*, 2011.

[183] LHC Higgs Cross Section Working Group collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922.

[184] CMS collaboration, A. M. Sirunyan et al., *Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *Phys. Lett.* **B793** (2019) 520 [1809.05937].

Appendix A

Profile Likelihood Ratio Test

The profile likelihood ratio test is one way of hypothesis test frequently used in many collider search. It assumes that, event by event, variables describing an event in collider are independent and follow identical distribution(i.i.d.). More precisely, if we consider, for example, event-by-event p_T with certain selection criteria to collect relevant collisions, then all of those p_T^i from i-th event would follow the distribution which is same with the $\frac{1}{\sigma_{\text{cut}}}\frac{d\sigma_{\text{cut}}}{dp_T}$, and independent from each others. Here the 'cut' in subscript means that this value was evaluated with the selection criteria. In general those distribution could be multi-dimensional, and also the combination of more than one observable can be used as random variable. So, let's say $\vec{x}_i$ is a vector or observable describing the i-th event among the set of the full data D. Therefore, $\mathrm{D} = \{\vec{x}_i \in \mathbb{R}^n | i = 1, \cdots, N_{\text{event}}\}$, where n is the dimensionality of event variables $\vec{x}_i$, and N_{event} is the number of total events in the data set.

Now, we could consider the probability density function(PDF) $P_i(\vec{x}_i)$ of those $\vec{x}_i$. If those $\vec{x}_i$ could be assumed to be i.i.d., then $P_i = P_j$ for all i and j, so we

can drop the index for PDF, and denote it as $P(\vec{x}_i)$. We could consider more than one hypothetical PDF as candidates for PDF of population. Also, in many cases, those PDFs can be deformed each others by introducing moduli $\vec{\mu}$. In this way, we could introduce more than one hypothesis by varying the moduli μ. For example, we could consider hypothesis H_0 and H_1 as the two PDFs with moduli value $\vec{\mu}_0$ and $\vec{\mu}_1$, respectively, i.e,

$$\begin{aligned} \mathrm{H}_0 \quad &: \quad P(\vec{x}_i) = P(\vec{x}_i|\vec{\mu} = \vec{\mu}_0), \\ \mathrm{H}_1 \quad &: \quad P(\vec{x}_i) = P(\vec{x}_i|\vec{\mu} = \vec{\mu}_1). \end{aligned}$$

Therefore, the hypothesis test could be considered as the test on moduli $\vec{\mu}$. To help understanding, we left one example of moduli $\vec{\mu}$. Let's we consider a new physics effect in the final state with two photons. As an observable, if we could consider differential cross section with respect to di-photon invariant mass, $m_{\gamma\gamma}$, then $\vec{x}_i = (m_{\gamma\gamma,i})$ for every i-th event. We could parametrise the strength of new physics with μ, by saying,

$$\frac{d\sigma}{dm_{\gamma\gamma}} = \frac{d\sigma_{\mathrm{SM}}}{dm_{\gamma\gamma}} + \mu\frac{d\Delta\sigma}{dm_{\gamma\gamma}},$$

where $\frac{d}{dm_{\gamma\gamma}}\sigma_{\mathrm{SM}}$ is differential cross section expected while assuming the standard model only, and $\frac{d}{dm_{\gamma\gamma}}\Delta\sigma$ is the deviation from the standard model due to given new physics. Then, $\mu = 0$ gives null-hypothesis that support no models but the SM, and $\mu \neq 0$ indicates the existence of new physics. There could be model-dependence issue on $\Delta\sigma$ or its differentials, but we would not discuss about it.

Now, we could define a 'likelihood', $\mathcal{L}$, of the data D with given hypothesis $\vec{\mu}$ as the probability to have data D based on hypothetical PDF $P(\vec{x}|\mu)$. Since we assumed every event is i.i.d.,

$$\mathcal{L}(\mathrm{D}|\vec{\mu}) \propto \prod_{i=1}^{N_{\mathrm{event}}} P(\vec{x}_i|\vec{\mu}). \tag{A.1}$$

Also, we need to consider the probability to have N_{event}-size of data. N_{event} may not coincides with $\hat{N}_{\mathrm{event}} = \hat{N}_{\mathrm{event}}(\vec{\mu})$, which is the expected number of data set assuming

hypothesis $\vec{\mu}$. Therefore, the full likelihood would be,

$$\mathcal{L}(\mathrm{D}|\vec{\mu}) = \mathrm{Pois}(N_{\mathrm{event}}|\hat{N}_{\mathrm{event}}(\vec{\mu})) \prod_{i=1}^{N_{\mathrm{event}}} P(\vec{x}_i|\vec{\mu}), \tag{A.2}$$

where $\mathrm{Pois}(n|\lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$ is Poisson probability at which yield is n when the expected yield is λ.

Now, let's consider a hypothesis test with null-hypothesis H_0 and alternative hypothesis H_1, where,

$$\begin{aligned} \mathrm{H}_0 \quad &: \quad \vec{\mu} = \vec{0} \qquad \text{(say background hypothesis)} \\ \mathrm{H}_1 \quad &: \quad \vec{\mu} = (\mu, \vec{m}_*) \qquad \text{(signal hypothesis).} \end{aligned}$$

Here we can consider μ as a signal strength(such as coupling to new resonance) deviated from the background, and $\vec{m}_*$ as modulus which determine the shape of signal(such as mass of new resonance). Now, for simplicity, let's consider the background PDF $p_{\mathrm{bkg}}(\vec{x})$ and signal PDF $p_{\mathrm{sig}}(\vec{x})$ which are

$$\begin{aligned} p_{\mathrm{bkg}}(\vec{x}) &= P(\vec{x}|\vec{\mu} = 0), \\ p_{\mathrm{sig}}(\vec{x}|\mu) &\propto P(\vec{x}|\vec{\mu} = (\mu, \vec{m}_*)) - P(\vec{x}|\vec{\mu} = 0). \end{aligned}$$

Similarly, let's say the total yield of respective 'signal($\vec{\mu} \neq 0$)' and 'background($\vec{\mu} = 0$)' as $\hat{N}_{\mathrm{sig}}$ and $\hat{N}_{\mathrm{bkg}}$,

$$\hat{N}_{\mathrm{bkg}} = \hat{N}_{\mathrm{event}}(\vec{\mu} = 0) \quad \text{and} \quad \hat{N}_{\mathrm{sig}}(\vec{\mu}) = \hat{N}_{\mathrm{event}}(\vec{\mu}) - \hat{N}_{\mathrm{bkg}}.$$

Then, we could have intuitive expression for $P(\vec{x}_i|\vec{\mu})$, which is,

$$P(\vec{x}_i|\vec{\mu}) = \frac{\hat{N}_{\mathrm{bkg}}}{\hat{N}_{\mathrm{sig}} + \hat{N}_{\mathrm{bkg}}} p_{\mathrm{bkg}}(\vec{x}_i) + \frac{\hat{N}_{\mathrm{sig}}}{\hat{N}_{\mathrm{sig}} + \hat{N}_{\mathrm{bkg}}} p_{\mathrm{sig}}(\vec{x}_i|\mu).$$

Furthermore, one can include the effect of uncertainties from various sources, such as measurement or limited power of theoretical prediction, in terms of 'nuisance parameters'. Nuisance parameters, at the end of stage, would be marginalised. This

procedure enhances the capability of likelihood function to interpret the given data. More precisely, one can introduce nuisance parameters θ associated to error ϵ on predicted profiles, $p_{\rm sig}$ or $p_{\rm bkg}$, i.e,

$$p_{\rm sig}(\vec{x}|\mu) \to p_{\rm sig}(\vec{x}|\mu,\epsilon,\theta) \quad \text{and} \quad p_{\rm bkg}(\vec{x}) \to p_{\rm bkg}(\vec{x}|\epsilon,\theta).$$

One example is $p(\vec{x}) = (1+\epsilon)^{\theta} p(\vec{x})$ for each signal and background profile. Here ϵ should not have to be constant, and, in general, can depend on $\vec{x}$. Also we can introduce a nuisance parameter for every single event. Now, with those profile with auxiliary nuisance parameters, we can write $\mathcal{L}(\mathrm{D}|\mu,\epsilon,\theta)$ as,

$$\mathcal{L}(\mathrm{D}|\mu,\epsilon,\theta) = \mathrm{Pois}(N_{\rm event}|\hat{N}_{\rm event}(\mu,\epsilon,\theta))P(\theta) \prod_{i=1}^{N_{\rm event}} P(\vec{x}_i|\mu,\epsilon\theta). \tag{A.3}$$

Here $P(\theta)$ is a penalty-term which limits the variation through nuisance parameters. If we include trivial penalty term, $P(\theta) = 1$, then there is no cost to deform the signal or background, and one can always perfectly fit the data to predicted model. This may not be appropriate, and we need to penalise too large variation through $P(\theta)$. With reasonable choice of penalty term, we marginalise iθ from this likelihood function as,

$$\hat{\mathcal{L}}(\mathrm{D}|\mu,\epsilon) = \max_{\theta} \mathcal{L}(\mathrm{D}|\mu,\epsilon,\theta), \tag{A.4}$$

and it would give more conservative likelihood compared to the case without no nuisance parameters.

Now with this $\hat{\mathcal{L}}$, we can define test statistics $t(\mu)$ as

$$t(\mu) = -2\ln \frac{\hat{\mathcal{L}}(\mathrm{D}|\mu,\epsilon)}{\hat{\hat{\mathcal{L}}}(\mathrm{D}|\epsilon)}. \tag{A.5}$$

Here $\hat{\hat{\mathcal{L}}}$ is best fit likelihood, so that,

$$\hat{\hat{\mathcal{L}}} = \max_{\mu} \hat{\mathcal{L}}(\mathrm{D}|\mu,\epsilon) \equiv \hat{\mathcal{L}}(\mathrm{D}|\hat{\mu},\epsilon). \tag{A.6}$$

Note that, in ideal case the likelihood ratio could be approximated as Gaussian probability, so that we can approximate

$$t(\mu) \simeq -2 \ln e^{-\frac{(\mu-\hat{\mu})^2}{2\sigma_\mu^2}} = \frac{(\mu-\hat{\mu})^2}{\sigma_\mu^2}, \tag{A.7}$$

and it makes clear the reason why it is called as test statistics.

Appendix B

Collider Phenomenology

Here we summarised the formulae for collider physics, especially for hadron collider.

B.1 Parton Density Function

Hadron colliders use composite particles, such as proton or anti-proton. The initial states of them could consist with more than one elementary field in the Standard Model(SM). Those non-trivial structure of hadron is studied in the context of the parton distribution function, or simply PDF. This PDF is a probabilistic distribution at which corresponding fundamental field takes specific portion of hadrons momenta. Therefore, with the partonic cross section in one hand, we can convolute the PDF on two incoming particles and get the full hadronic cross section. Here the partonic cross section is the cross section evaluated with fundamental fields in SM both at initial and final state.

Let's consider the PDF $f_i(x, \mu_F)dx$ which is the probability that a particle with

label i takes momentum fraction x from the hadron. Here μ_F is factorisation scale at which we factored out the partonic cross section from hadronic cross section. Then the full hadronic cross section σ could be written in terms of partonic cross section $\hat{\sigma}$ as,

$$\begin{aligned}&\sigma((\text{hadron}) \to (\text{final})) \\ &\quad = \sum_{i,j} \int_0^1 dx_1 f_i(x_1, \mu_F) \int_0^1 dx_2 f_j(x_2, \mu_F) \frac{\hat{\sigma}(ij \to (\text{final}))}{1 + \delta_{ij}}\end{aligned} \tag{B.1}$$

Here the factor $1/(1 + \delta_{ij})$ prevents the double counting when i and j are same. For simplicity, let's abbreviate the summation of initial particle($\Sigma_{i,j}$) and that factor $(1/(1 + \delta_{ij}))$ from now on.

In high energy environment($\gtrsim$ 1TeV), taking massless limit of quarks is good approximation except for the top-quark. And fortunately, current energy scale of Large Hadron Collider(LHC) suppress the top-quark PDF in protons, so we can simply think all initial particles are massless. In such limit, we can approximate the four-momenta of initial particles as

$$x_1(p, 0, 0, p) \qquad \text{and} \qquad x_2(p, 0, 0 - p) \tag{B.2}$$

with two four-momenta of protons $(p, 0, 0, \pm p)$, which are also in massless limit. This simplifies the center-of-mass(CM) energy, $\hat{s}$, of initial particles and the longitudinal rapidity, y, of the CM frame with respect to the laboratory(LAB) frame as follow.

$$\hat{s} = x_1 x_2 s \qquad \text{and} \qquad y = \frac{1}{2} \ln \frac{x_1}{x_2} \tag{B.3}$$

Here $s \equiv 4p^2$ is CM energy of two protons[1]. This allows us the change of variables from (x_1, x_2) to (τ, y), each of which are $\tau = x_1 x_2$ and y as shown in eq. (B.3). Corresponding Jacobian determinant is trivial, and $x_1 = \sqrt{\tau} e^y$, $x_2 = \sqrt{\tau} e^{-y}$, so we

[1] $s = -(p_1 + p_2)^2$ and $y = \frac{1}{2} \ln \frac{E+p^3}{E-p^3}$ are used.

can rewrite eq. (B.1) as follow.

$$\sigma = \int d\tau dy f_i(\sqrt{\tau}e^{y}) f_j(\sqrt{\tau}e^{-y})\hat{\sigma} \tag{B.4}$$

Integration would be taken over $[0,1]$ for τ, and $\left[\frac{1}{2}\ln\tau, -\frac{1}{2}\ln\tau\right]$ for y.

B.2 Partonic Cross Section

In Appendix B.1, I summarised how we carry PDF convolution on partonic cross section. The hadronic cross section is already suitable for the evaluation of differential cross section with respect to CM energy(τ). This would be resonant energy in s-channel collision and, hence, relevant for resonance search. However in many case, other observables, such as rapidity or transverse momenta of particles in final state are extremely useful, too. Those kinematic information would be inherited in the partonic cross section, and here I summarise it for 2-to-2 scattering case.

Consider general 2-to-2 scattering, and let $m_{1,2}$ be the respective masses of the particles in the final state. It is well known that, in such simple scattering process, the partonic cross section would be,

$$d^2\hat{\sigma} = \frac{1}{\varphi}\frac{d^3\vec{p_1}}{(2\pi)^3}\frac{1}{2E_1}\frac{d^3\vec{p_2}}{(2\pi)^3}\frac{1}{2E_2}(2\pi)^4\delta^{(4)}(p_{\text{in}} - p_1 - p_2)\overline{|\mathcal{M}|^2}, \tag{B.5}$$

with spin-summed(and also taken average for initial spins) squared-matrix-element $\overline{|\mathcal{M}|^2}$ and flux of incident beams φ. Here φ in arbitrary boost of head-on collision would be

$$4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}. \tag{B.6}$$

Note that, even though I denoted as $d^2\hat{\sigma}$ on the left-hand-side(lhs) of eq. (B.5), it is not a double differential of function $\hat{\sigma}$. Rather than that, it is kind of notation emphasising that it would be defined as double integral of right-hand-side(rhs) of eq.

(B.5)[2].

From eq. (B.5), we can integrate out $d^3\vec{p_2}$ exactly using three of four Dirac-δ function. It is equivalent to solving 2-to-2 scattering kinematics. In this scattering, we can determine the four-momenta of one particle as long as we know about another one. Then,

$$d^2\sigma = \frac{1}{\varphi}\frac{d^3\vec{p_1}}{(2\pi)^2}\frac{1}{4E_1E_2}\delta(E_{\rm in}-E_1-E_2)\overline{|\mathcal{M}|^2} \tag{B.7}$$

Here $d^3\vec{p_1}$ could be written in (p_T, y_1, ϕ) basis as

$$d^3\vec{p_1} = dp_T d\phi dp_{1z}\; p_T = dp_T d\phi dy_1\; m_{1T}p_T\cosh y_1, \tag{B.8}$$

since $p_{1z} = m_{1T}\sinh y_1$ with transverse mass $m_{1T} = \sqrt{m_1^2+p_{1z}^2}$. Now, integrating over ϕ gives

$$d\sigma = dp_T dy_1 \frac{1}{\varphi}\frac{m_{1T}p_T\cosh y_1}{8\pi E_1E_2}\delta(E_{\rm in}-E_1-E_2)\overline{|\mathcal{M}|^2}. \tag{B.9}$$

Here $E_{\rm in}$ is energy injected from the initial state. So if we fix a direction of particle labelled by 1, i.e, fix y_1, then, for fixed $E_{\rm in}$, p_T would be determined automatically. Or, in opposite way, to have particle 1 in direction with rapidity y_1 and transverse momentum p_T, then $E_{\rm in}$ is required to be certain value. Using this fact, we will integrate over τ in hadronic cross section formula to get differential cross section with respect to p_T or y_1.

Before going further, let's write Mandelstam variables with values we have, first.

$\hat{s}$ was already quite simple, as $\hat{s} = \tau s$. With $p_\pm = (p, 0, 0, \pm p)$, we can write $\hat{t} \equiv -(p_1 - x_1 p_+)^2$ as,

$$\hat{t} = m_1^2 + 2x_1 p\,(-E_1 + p_{1z}) = m_1^2 - x_1\sqrt{s}m_{1T}e^{-y_1}. \tag{B.10}$$

[2]To help the understanding, it would be good to image ϕ, azimuthal angle, integration or differential cross section. Neither differential cross section nor cross section would depend on ϕ, so one might think $d\hat{\sigma}/d\phi = 0$, but it is not true, obviously.

Considering another equivalent definition, $\hat{t} = -(p_2 - x_2 p_-)^2$ and similar things for $\hat{u}$, we have Automatically, $\hat{u} \equiv -(p_2 - x_1 p_+)^2$ would be,

$$\begin{aligned} \hat{t} &= m_1^2 - x_1\sqrt{s}m_{1T}e^{-y_1} = m_2^2 - x_2\sqrt{s}m_{2T}e^{y_2} \\ \hat{u} &= m_1^2 - x_2\sqrt{s}m_{1T}e^{y_1} = m_2^2 - x_1\sqrt{s}m_{2T}e^{-y_2}. \end{aligned} \tag{B.11}$$

Note that, $\hat{s}$ can be written fully in terms of final state variables,

$$\hat{s} = m_{1T}^2 + m_{2T}^2 + 2m_{1T}m_{2T}\cosh(y_1 - y_2). \tag{B.12}$$

Or, using $\hat{s} + \hat{t} + \hat{u} = m_1^2 + m_2^2$,

$$\hat{s} = \left[m_{1T}\cosh(y - y_1) + \sqrt{m_{1T}^2\cosh^2(y - y_1) - m_1^2 + m_2^2}\right]^2. \tag{B.13}$$

In the expression of eq. (B.13), there is no other variables but p_T, y_1 and y are used. Using this again simplifies $\hat{t}$ and $\hat{u}$ in eq. (B.11), and we can write all $\hat{s}$, $\hat{t}$ and $\hat{u}$ in terms of $(p_T,\ y_1,\ y)$. Since $\overline{|\mathcal{M}|^2}$ could be expressed in terms of Mandelstam variables, so, now we have $\overline{|\mathcal{M}|^2}$ in terms of $(p_T,\ y_1,\ y)$, too. Last, injected flux would be nothing but $4|p_1 \cdot p_2| = 2\hat{s}$, so, again it would be written in the same basis.

Now, from eq. (B.9), we can write all of them in terms of $(p_T,\ y_1,\ y)$ except E_2. But E_2 is simple, because, as long as we fix τ to give appropriate $\hat{s}$, then $E_{\text{in}} = \sqrt{\hat{s}}\cosh y$, hence,

$$E_2 = \sqrt{\hat{s}}\cosh y - m_{1T}\cosh y_1. \tag{B.14}$$

But we should careful that we didn't fix τ yet. But we do know which τ we need to have appropriate (p_T, y_1) at certain value of y.

B.3 Hadronic Cross Section

Using eq. (B.4) and eq. (B.9), now we have,

$$\sigma = \int d\tau dy dp_T dy_1 \left[f_i f_j \frac{1}{\varphi}\frac{m_{1T}p_T\cosh y_1}{8\pi E_1 E_2}\delta(E_{\text{in}} - E_1 - E_2)\overline{|\mathcal{M}|^2}\right]. \tag{B.15}$$

Remind that $E_{\rm in} = \sqrt{\tau s}\cosh y$ and $E_1 + E_2 = \sqrt{\hat{s}_*}\cosh y$, where $\hat{s}_*$ is required value to have (p_T, y_1, y). Therefore,

$$\int d\tau \delta(E_{\rm in} - E_1 - E_2) = \frac{2}{s}\frac{\sqrt{\hat{s}_*}}{\cosh y}. \tag{B.16}$$

Using $E_1 = m_{1T}\cosh y_1$,

$$\sigma = \int dy dp_T dy_1 \left[f_i f_j \frac{1}{\varphi}\frac{p_T}{8\pi E_2}\frac{2}{s}\frac{\sqrt{\hat{s}_*}}{\cosh y}\overline{|\mathcal{M}|^2}\right]. \tag{B.17}$$

And combining them all,

$$\sigma = \int dy dp_T dy_1 \left[f_i f_j \frac{1}{8\pi s}\frac{p_T}{\sqrt{\hat{s}}\cosh y}\frac{1}{\sqrt{\hat{s}}\cosh y - m_{1T}\cosh y_1}\overline{|\mathcal{M}|^2}\right]. \tag{B.18}$$

Here the integral range of y would be $[y_{\rm min}, y_{\rm max}]$ are derived from boundary of (x_1, x_2) plane. $y_{\rm min}$ gives $x_2 = 1$ and $y_{\rm max}$ gives $x_1 = 1$. Since $y = \frac{1}{2}\ln(x_1/x_2)$, we have $(x_1, x_2) = (e^{2y_{\rm min}}, 1)$ at $y = y_{\rm min}$, and $(x_1, x_2) = (1, e^{2y_{\rm max}})$ at $y = y_{\rm max}$.

From eq. (B.11), at the same time, we have

$$\hat{t} + \hat{u} = 2m_1^2 - x_1\sqrt{s}m_{1T}e^{-y_1} - x_2\sqrt{s}m_{1T}e^{y_1}, \tag{B.19}$$

while $\hat{s} + \hat{t} + \hat{u} = m_1^2 + m_2^2$ and $\hat{s} = x_1 x_2 s$.

This gives condition on $y_{\rm min,max}$ as follow.

$$\begin{aligned} m_1^2 + m_2^2 - se^{2y_{\rm min}} &= 2m_1^2 - \sqrt{s}m_{1T}e^{-y_1}e^{2y_{\rm min}} - \sqrt{s}m_{1T}e^{y_1} \\ m_1^2 + m_2^2 - se^{-2y_{\rm max}} &= 2m_1^2 - \sqrt{s}m_{1T}e^{-y_1} - \sqrt{s}m_{1T}e^{y_1}e^{-2y_{\rm max}} \end{aligned} \tag{B.20}$$

So, if we denote $y_+ \equiv y_{\rm max}$ and $y_- \equiv y_{\rm min}$, then

$$\left(s - \sqrt{s}m_{1T}e^{\pm y_1}\right)e^{\mp y_\pm} = m_2^2 - m_1^2 + \sqrt{s}m_{1T}e^{\mp y_1}, \tag{B.21}$$

or more explicitly,

$$y_\pm = \mp\frac{1}{2}\ln\left(\frac{m_2^2 - m_1^2 + \sqrt{s}m_{1T}e^{\mp y_1}}{s - \sqrt{s}m_{1T}e^{\pm y_1}}\right). \tag{B.22}$$

Therefore, double-differential cross section with respect to p_T and y_1 is as follow.

$$\frac{d^2\sigma}{dp_T dy_1} = \frac{1}{8\pi s}\int_{y_-}^{y_+} dy f_i f_j \frac{p_T}{\sqrt{\hat{s}}\cosh y(\sqrt{\hat{s}}\cosh y - m_{1T}\cosh y_1)}\overline{|\mathcal{M}|^2} \tag{B.23}$$

From eq. (B.15), we can also derive $d\sigma/d\tau$ or equivalently $d\sigma/d\hat{s}$. To write this, we need to integrate the last δ-function with other than τ. The y gives simple form, since fixed τ, p_T and y_1 determine y exactly. In case $m_{1,2} = 0$, p_T also gives simple form, since $m_{1,2} = 0$ implies

$$E_i = |\vec{p_i}| \qquad \text{for} \quad i = 1,\ 2. \tag{B.24}$$

Using this condition, we can write E_2 as

$$E_2 = \left|-\vec{p}_T + \left(\sqrt{\hat{s}}\sinh y - p_T \sinh y_1\right)\mathbf{e}_3\right|. \tag{B.25}$$

Therefore, for fixed τ, y and y_1, we can perform integration over p_T as

$$\int dp_T \delta(E_{\text{in}} - E_1 - E_2) = \int dp_T \frac{E_2(p_T^*)}{\sqrt{\hat{s}}\cosh(y_1 - y)}\delta(p_T - p_T^*), \tag{B.26}$$

where p_T^* is the solution of δ-function, so that,

$$p_T^* = \frac{\sqrt{\hat{s}}}{2\cosh(y_1 - y)}. \tag{B.27}$$

Combining them all, as long as $m_1 = m_2 = 0$,

$$\frac{d\sigma}{d\tau} = \int dy dy_1 \left[\frac{f_i f_j}{32\pi\hat{s}}\frac{1}{\cosh^2(y - y_1)}\overline{|\mathcal{M}|^2}\right]. \tag{B.28}$$

Note that eq. (B.28) restores simple formula if there were no cut on the phase space. Without cut, we can take integration over y_1 exactly,

$$\int_{-\infty}^{\infty} dy_1 \frac{1}{\cosh^2(y_1 - y)} = 2, \tag{B.29}$$

therefore,

$$\frac{d\sigma}{d\tau} = \left[\int dy f_i f_j\right]\frac{1}{16\pi\hat{s}}\overline{|\mathcal{M}|^2}. \tag{B.30}$$

Now let's consider a cut on phase space, say $\mathcal{P}(\Pi)$, which is 1 for allowed phase space $\Pi = (p_{1T}, p_{2T}, y_1, \cdots)$ and 0 for forbidden region. Then, for differential cross section with cut $\mathcal{P}$ is,

$$\frac{s}{2\sqrt{\hat{s}}}\frac{d\sigma}{d\sqrt{\hat{s}}} = \int dy dy_1 \left[\mathcal{P}(\Pi)\frac{f_i f_j}{32\pi\hat{s}}\frac{1}{\cosh^2(y-y_1)}\overline{|\mathcal{M}|^2}\right]. \tag{B.31}$$

Appendix C

Loop Functions

We summarise the relevant functions for Higgs production via gluon-fusion with an associated jet at the lowest order in quantum chromodynamics(QCD). First, $A_2(s,t,u)$ and $A_4(s,t,u)$ are the functions relevant for $gg \to Hg$ processes and defined as follow.

$$\begin{aligned} A_2(s,t,u) &= b_2(s,t,u) + b_2(s,u,t) \\ A_4(s,t,u) &= b_4(s,t,u) + b_4(t,u,s) + b_4(u,s,t) \end{aligned}$$

The function b_2 and b_4 are defined as,

$$\begin{aligned} b_2(s,t,u) &= \frac{m_t^2}{m_H^4}\left[\frac{s(u-s)}{u+s} + \frac{2ut(u+2s)}{(u+s)^2}\left(W_1(t) - W_1(m_H^2)\right)\right. \\ &+ \left(m_t^2 - \frac{1}{4}s\right)\left(\frac{1}{2}W_2(s) + \frac{1}{2}W_2(m_H^2) - W_2(t) + W_3(s,t,u,m_H^2)\right) \\ &+ s^2\left(\frac{2m_t^2}{(s+u)^2} - \frac{1}{2(s+u)}\right)\left(W_2(t) - W_2(m_H^2)\right) \\ &\left. + \frac{ut}{2s}\left(W_2(m_H^2) - 2W_2(t)\right) + \frac{1}{8}\left(s - 12m_t^2 - \frac{4ut}{s}\right)W_3(t,s,u,m_H^2)\right] \end{aligned} \tag{C.1}$$

$$b_4(s,t,u) \tag{C.2}$$
$$= \frac{m_t^2}{m_H^2}\left[-\frac{2}{3} + \left(\frac{m_t^2}{m_H^2} - \frac{1}{4}\right)\left(W_2(t) - W_2(m_H^2) + W_3(s,t,u,m_H^2)\right)\right]$$

with m_t, mass of top-quark, while assuming all other quarks other than top are massless. The W_1, W_2 and W_3 are loop functions, which are [107].

$$\begin{aligned} W_1(s) &= 2 + \int_0^1 dx \ln\left(1 - x(1-x)\frac{s}{m_t^2} - i\varepsilon\right) \\ W_2(s) &= 2\int_0^1 \frac{dx}{x}\ln\left(1 - x(1-x)\frac{s}{m_t^2} - i\varepsilon\right) \\ W_3(s,t,u,v) &= I_3(s,t,u,v) - I_3(s,t,u,s) - I_3(s,t,u,u) \\ I_3(s,t,u,v) &= \int_0^1 dx \left(\frac{m_t^2 t}{us} + x(1-x)\right)^{-1} \ln\left(1 - x(1-x)\frac{v}{m_t^2} - i\varepsilon\right), \end{aligned} \tag{C.3}$$

with $\varepsilon \to +0$ limits.

There is another form-factor A_5 for $gq \to Hq$ or $q\bar{q} \to Hg$ processes, which is defined as,

$$\begin{aligned} A_5(s,t,u) &= \frac{m_t^2}{m_H^2}\left[4 + \frac{4s}{t+u}\left(W_1(s) - W_1(m_H^2)\right)\right. \\ &\qquad \left. + \left(1 - \frac{4m_t^2}{t+u}\right)\left(W_2(s) - W_2(m_H^2)\right)\right] \end{aligned} \tag{C.4}$$

$W_{1,2,3}$ and hence I_3 can also be written in the closed form after integration.

$$W_1(s) = \begin{cases} 2\left(1 - \frac{4m_t^2}{s}\right)^{\frac{1}{2}} \operatorname{arcsinh}\left(\frac{\sqrt{-s}}{2m_t}\right), & s \in (-\infty, 0) \\ 2\left(\frac{4m_t^2}{s} - 1\right)^{\frac{1}{2}} \arcsin\left(\frac{\sqrt{s}}{2m_t}\right), & s \in \left[0, 4m_t^2\right) \\ 2\left(1 - \frac{4m_t^2}{s}\right)^{\frac{1}{2}} \left[\operatorname{arccosh}\left(\frac{\sqrt{s}}{2m_t}\right) - i\pi\right]. & s \in \left[4m_t^2, \infty\right) \end{cases} \tag{C.5}$$

$$W_2(s) = \begin{cases} 4\operatorname{arcsinh}^2\left(\frac{\sqrt{-s}}{2m_t}\right), & s \in (-\infty, 0) \\ -4\arcsin^2\left(\frac{\sqrt{s}}{2m_t}\right), & s \in \left[0, 4m_t^2\right) \\ \left[2\operatorname{arccosh}\left(\frac{\sqrt{s}}{2m_t}\right) - i\pi\right]^2. & s \in \left[4m_t^2, \infty\right) \end{cases} \tag{C.6}$$

For I_3, if $v < 0$,

$$\begin{aligned} I_3(s,t,u,v) &= \frac{2}{2\beta-1}\left[-\mathrm{Li}_2\left(\frac{\gamma}{\gamma+\beta-1}\right)+\mathrm{Li}_2\left(\frac{\gamma-1}{\gamma+\beta-1}\right)\right. \\ &\quad +\mathrm{Li}_2\left(\frac{\beta-\gamma}{\beta}\right)-\mathrm{Li}_2\left(\frac{\beta-\gamma}{\beta-1}\right) \\ &\quad \frac{1}{2}\left(\log^2\beta-\log^2(\beta-1)\right)+\log\gamma\log\left(\frac{\gamma+\beta-1}{\beta}\right) \\ &\quad \left.\log(\gamma-1)\log\left(\frac{\beta-1}{\gamma+\beta-1}\right)\right] \end{aligned} \tag{C.7}$$

for $v \in \left[0, 4m_t^2\right)$,

$$I_3(s,t,u,v) = \frac{2}{2\beta-1}\left[2\mathrm{Li}_2^{\mathrm{im}}(r,\theta)-2\mathrm{Li}_2^{\mathrm{im}}(r,\phi)+(\phi-\theta)(\phi+\theta-\pi)\right] \tag{C.8}$$

and for $v \in \left[4m_t^2, \infty\right)$,

$$\begin{aligned} I_3(s,t,u,v) &= \frac{2}{2\beta-1}\left[-\mathrm{Li}_2\left(\frac{\gamma}{\gamma+\beta-1}\right)+\mathrm{Li}_2\left(\frac{\gamma-1}{\gamma+\beta-1}\right)\right. \\ &\quad +\mathrm{Li}_2\left(\frac{\gamma}{\gamma-\beta}\right)-\mathrm{Li}_2\left(\frac{\gamma-1}{\gamma-\beta}\right) \\ &\quad \left.+\log\left(\frac{\gamma}{1-\gamma}\right)\log\left(\frac{\gamma+\beta-1}{\beta-\gamma}\right)-i\pi\log\left(\frac{\gamma+\beta-1}{\beta-\gamma}\right)\right] \end{aligned} \tag{C.9}$$

Here β, γ, r, ϕ and θ are defined as in eqeq. (C.10),

$$\begin{aligned} \beta(s,t,u) &= \frac{1}{2}\left(1+\sqrt{1+\frac{4\,t\,m_t^2}{s\,u}}\right), \\ \gamma(v) &= \frac{1}{2}\left(1+\sqrt{1-\frac{4m_t^2}{v}}\right), \\ \alpha(v) &= \sqrt{\frac{4m_t^2}{v}-1}, \\ r^2 &= \frac{\alpha^2+1}{\alpha^2+(2\beta-1)^2}, \\ \cos\phi &= \frac{r\left(\alpha^2+2\beta-1\right)}{1+\alpha^2}, \\ \cos\theta &= \frac{r\left(\alpha^2-2\beta+1\right)}{1+\alpha^2}, \end{aligned} \tag{C.10}$$

while $\phi, \theta \in (0, \pi)$.

Also, Li_2 is usual di-logarithmic function, so that,

$$\mathrm{Li}_2(x) = -\int_0^x dz \frac{\log(1-z)}{z}, \tag{C.11}$$

and $\mathrm{Li}_2^{\mathrm{m}}$ is modified one of di-logarithmic function,

$$\mathrm{Li}_2^{\mathrm{m}}(r, \theta) = -\frac{1}{2}\int_0^r dz \frac{\log(1 - 2\cos\theta z + z^2)}{z}. \tag{C.12}$$

Appendix D

Jet Tagging Algorithm for Simulated Events

We summarize the algorithm to find the initial particle of jets after fragmentation/hadronisation and jet clustering. After passing through the fragmentation/hadronization program like Pythia8 [80], the full history of each event is recorded. In the event history, each particle has information about the particles which produced itself as their 'mother'. There are `mother1` and `mother2` for each particle, and the ordering between `mother1` and `mother2` and the status of the particle indicate different interactions by which the particle is produced. For example, it is $2 \to n$ process when (`mother1` >`mother2`), and $2 \to 1$ process if (`mother1` < `mother2`). If the status of the particle is in between 81 and 86 while (`mother1` < `mother2`), it is the fragmentation of the hadrons, so that all particles from `mother1` to `mother2` are considered to be its mothers. There are five more cases, and the other cases can be found from the manual of Pythia8 [80].

To find an origin of given jet, the algorithm tracks the event history starting from

the particles, say $\{C_1^{(j)}, \cdots, C_n^{(j)}\}$, which are clustered as the jet. The label (j) at the superscript means that it the particle $C_i^{(j)}$ is a constituent of the jet, (j), and subscript i labels the constituents. The first step of algorithm is to collect all mothers of constituents following the event history. The collection is said to be 'candidates', and

$$\begin{aligned}&\text{(STEP 1)}\\ &\text{candidates} = \left\{M_1\left(C_1^{(j)}\right), \cdots, M_2\left(C_n^{(j)}\right)\right\}.\end{aligned}$$

Note that, there can be duplication of mother particles from, for example $2 \to n$ process or fragmentation of hadrons. We, therefore, remove all duplication after finding the candidates array. A similar procedure is repeated for the candidates; the mothers of candidates are collected, and duplications are removed. This procedure is terminated if all particles in candidates have `status` from {4, 14, 15, 23, 24, 63}. Those `status` are recorded along the fragmentation/hadronisation and indicate the stage at which the particle is produced. The particle comes from the hard process or subsequent process if `status` $\in$ {4, 14, 15, 23 or 24}, and it is a remnants of the beam if `status`=63. This step may not be terminated even though all particles in event history are iterated. Then the algorithm consider the corresponding jet is not able to be tagged with particle(s) from hard scattering process, and assign '`0`'. If the step ended with more than zero particle, say,

$$\text{candidates} = \left\{P_1^{(j)}, \cdots, P_m^{(j)}\right\}, \tag{D.1}$$

we remove all colour singlet particles from the candidates (STEP 2). This may remove all particles, if the jet is not a QCD jet. Then, again, the algorithm returns `0`, for the jet. For the left particles, some of them are excluded from the candidates depending on the angular separation ΔR between the particle and the jet (STEP 3). If ΔR of a particle is greater than ΔR_{cut}, which is input parameter of the algorithm, the

corresponding particle is excluded. The algorithm then evaluate '`score`' for the left particles in the candidates(STEP 4), where the `score` is defined as

$$\texttt{score} = \frac{100}{1 + \epsilon_{p_T} \Delta R}, \tag{D.2}$$

with $\epsilon_{p_T} \equiv \frac{1}{p_T^{(j)}} \left(p_T - p_T^{(j)} \right)$. From its definition, the `score` distributes between 0 and 100, and it is close to 100 as the three momentum of the particle is close to the three momentum of the jet. Hence the algorithm consider the particle with the highest score as the origin of the jet, and assign the particle ID to the jet.

The algorithm is validated with the monojet samples from respective $pp \to gZ$ and $pp \to qZ$ processes where the Z bosons are forced to decay only into neutrino pairs. With these clean samples, the efficiency and the accuracy of algorithm was checked while varying the input parameter ΔR_{cut}. The efficiency is defined to be ratio between the number of jets of which the algorithm successfully found the origin and the number of total input jets, i.e,

$$\text{(tagging efficiency)} = \frac{\#\text{(tagged jets)}}{\#\text{(total jets)}}.$$

The accuracy is the rate at which the jets are tagged with its correct origin when it is tagged with the algorithm successfully, so that

$$\text{(tagging accuracy)} = \frac{\#\text{(correctly tagged jets)}}{\#\text{(tagged jets)}}.$$

The samples are generated at leading order of QCD. Monte-Carlo simulation was done via MadGraph5 aMC@NLO at 13 TeV. The showering and hadronisation are done by Pythia8 [80], and Delphes3 [81] is used for the detector simulation. The MLM matching was done to avoid any ambiguity. Jets are reconstructed with anti-k_T algorithm with the angular parameter $\Delta R = 0.4$, and CMS setup for the fast detector simulation was used. The result can be found from Fig. D.1. The two plots on the top panels of Fig. D.1 shows the efficiency of gluon jet tagging(left) andthe efficiency

of quark jet tagging(right). Two plots at the bottom line show the accuracy, of each cases.

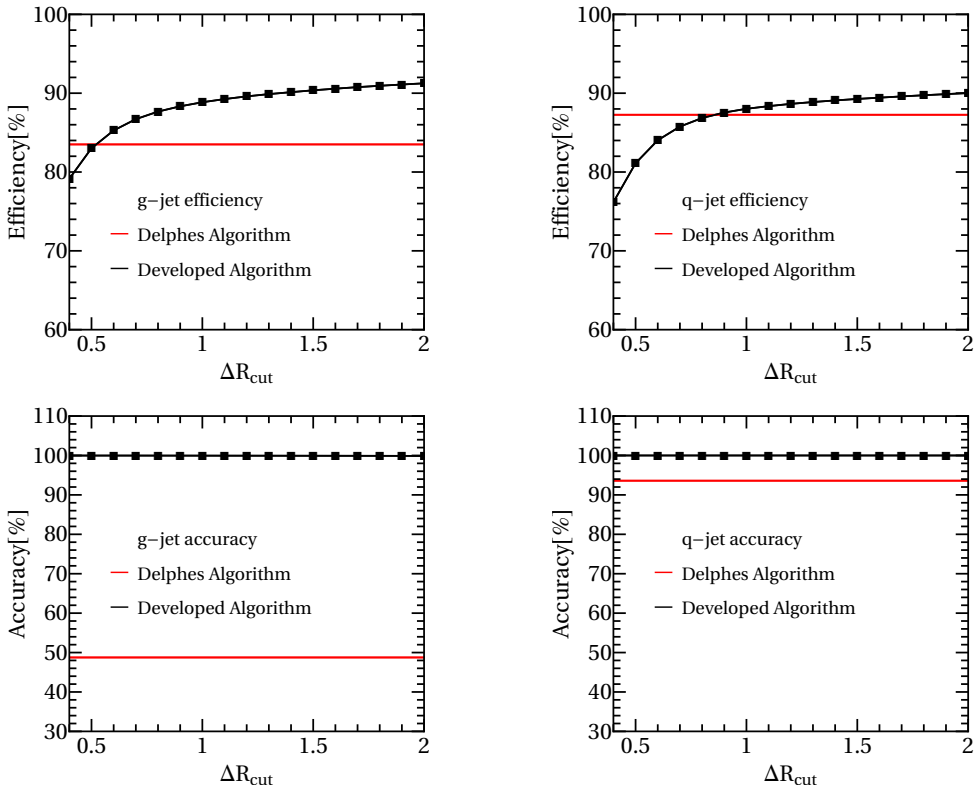


Figure D.1 Efficiency(top line) and accuracy(bottom line) measure with respective $p\,p \to Z\,g$ sample(left) and $p\,p \to Z\,q$ sample(right). The definition of the efficiency and the accuracy could be found in the text. Here ΔR_{cut} is the parameter of our algorithm while the Delphes algorithm does not depend on it.

초록

이 논문은 대형 강입자 가속기에서의 탐색 중, 중간 입자의 날카로운 공명 정점이 없는 경우에 대한 기계 학습의 응용을 다룬다. 응용의 한가지 예로써, 기계 학습을 이용하여 탐색하고자 하는 중간 입자의 공명의 폭이 넓을 때 가장 적합한 관측량이 무엇인가에 대하여 조사한다. 보다 자세한 연구를 위하여, 합성 힉스 모형의 전반에서 예측 되는 무거운 합성 벡터 중간자가 위 쿼크와 위 쿼크의 반입자로 공명 붕괴하는 과정을 생각한다. 다양한 방법론을 통하여 기계 학습을 통해 추상화된 정보를 해석하여, 결과적으로 공명 폭이 넓은 경우에도 재구축 된 중간 입자의 질량이 유용하고 붕괴한 입자들의 각도 분포 및 직교 운동량 또한 해당 탐색에서 중요성을 갖고 있음을 확인한다. 두 번째 예로써, 기계 학습을 이용하여 공명을 유발하는 중간 입자의 붕괴 후 최종 상태가 아닌 곳에서 방출 된 입자를 통해 중간 입자에 대한 정보를 추출하는 방법을 연구한다. 이를 이해하기 위하여 초기 상태 복사로 나온 제트와 중간 입자 사이의 상호 관련성이 있음을 보인다. 제시된 분석 방법의 실제 실험에 대한 응용 가능성을 실증하기 위하여, 힉스 입자의 관측 불가능 입자들로의 붕괴에 대한 연구를 기계학습을 이용하여 재현한다. 결과적으로 글루온 융합으로 힉스입자가 생성되는 과정에서 얻을 수 있는 힉스입자의 관측 불가능 입자로 붕괴하는 확률에 대한 구속 조건이 크게 개선되며, 다른 생성 과정보다 강한 구속 조건을 줄 수 있음을 보인다.

주요어: 대형 강입자 가속기, 기계학습, 힉스 보존, 탑-파트너, 제트, 제트 내부구조, 초기 상태복사

학번: 2015-20343