國立清華大學

物理系

碩士學位論文

基於 SPA-Net 的雙頂夸克全強子衰變事件重建

Event reconstruction of all hadronic Top-quark-pair decays using SPA-Net

系所組別: 物理所物理組

學號姓名:

108022517 何大維 (Ta-Wei Ho)

指導教授 :

張敬民　教授 (Prof. Kingman Cheung)

徐士傑　教授 (Prof. Shih-Chieh Hsu)

中　華　民　國　一　一　〇　年　五　月

Event reconstruction of all hadronic Top-quark-pair decays using SPA-Net

A Thesis Presented to
the Department of Physics at
National Tsing Hua University
in Partial Fulfillment for the Requirement of
the Master of Science Degree Program

By
Ta-Wei Ho
Advisor
Dr. Kingman Cheung
Dr. Shih-Chieh Hsu
May 2021

Abstract

The top quarks produced by proton-proton collisions in the Large Hadron Collider(LHC) undergo a very complicated process of formation. To date, the decay products of top quarks are still not well-classified. In this project, we present a novel approach to the "all hadronic decay" process of top quarks based on the neural networks with attention mechanism, we refer to as the "Symmetry Preserving Attention Networks"(SPA-Net). These networks identify the decay products of each quark unambiguously and without combinatorial explosion. This approach performs outstandingly compared to the generally accepted state-of-the-are method. Our networks can correctly assign all hadronic decay in 93.0% of 6 jets, 87.8% of 7 jets, and 82.6% of ≥ 8 jets event respectively.

摘要

在大型強子對撞機 (LHC) 實驗中，經由質子對撞所產生的頂夸克對具有非常複雜的過程以及產物，至今仍無法被非常正確的判別以及重建。在本研究中，我們提出了一個利用新穎的機器學習方法來對雙頂夸克全強子衰變過程進行重建。此方法基於 Attention mechanism，我們稱之為 Symmetry Preserving Attention Networks(SPA-Net)。這個模型架構可以在避免組合性爆炸的前提下對所有的衰變產物進行辨識以及重建。此方法對比於傳統的 χ^2 重建方式，表現出了非常巨大的差異。本方法可以在一、存在 6 jets 條件下正確的重建 93% 的事件；二、存在 7 jets 條件下正確的重建 87% 的事件；三、存在大於 8 jets 條件下正確的重建 82.6% 的事件。

Acknowledgements

I am deeply grateful for the help from my colleagues, advisor, collaborators and friends who unstintingly helped me whenever I grappled with research and technical issues. Their kindness helps me to overcome the advantages and receive the degree.

Professor Kingman Cheung who is my dear advisor and Professor Shih-Chieh Hsu, my respected collaborator, provided me with invaluable support without which I would have struggled to accomplish this project. To my colleague, Yi-Lun Chung, who helped me solve the problem of simulation packages and theoretical problems, I also express my deepest gratitude.

I am profoundly indebted to all the distinguished overseas collaborators who participated in this project, namely Mike Fenton, Alexander Shmakov, Daniel Whiteson, and Pierre Baldi. I could not have finished this project without their support. The way that Alexander constructed a machine learning model and the way that Mike described a physics concept really taught me a lot. The guidance of Daniel, Shih-Chieh, and Pierre inspired me to explore more deeply.

我想感謝我的同事、教授、朋友以及合作者。若沒有他們的鼎力相助，我肯定無法順利地完成這個研究。在我遇到各種問題，在困難中掙扎時，他們友善的協助讓我跨越難關，最後終於取得了學位。

張敬民教授以及徐士傑教授作為我的指導教授以及合作者，不論在物理上或是機器學習上，都非常大方地給予我非常多的建議。這些建議讓我能夠以足夠的知識來執行這個研究。他們的建議，每每都在我遇到困難時發揮關鍵的作用，也在我陷入迷霧時給予了我一盞明燈。

鍾沂倫，研究室的博士班學長，也在研究中給予了我相當大的幫助。無論是在各種環境的設定以及工作站的維護，乃至模擬軟體的設定以及物理模型的參數設定，都給了我非常多的幫助。他的協助讓我在維護工作站以及管理上輕鬆了非常多，真的是非常感謝他的幫忙。

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Inside the Large Hadron Collider(LHC), two protons collide with very high energy and produce many kinds of products. A process whereby *pp* collision produces a pair of top quarks and results in the 6 jets final state is called **All Hadronic Top-quark-pair Decay**. This process has a very complicated signature due to a large number of combinations producde by the possible permutation of final state jets. These jets produced by the top quark pair are hard to tag as a specific daughter of top quarks correctly. The traditional method is to reconstruct the event using χ^2 reconstruction, but it takes a long time to compute and the result have a low accuracy (about 30% or less). The investigation of top quark and its full hadronic decay channel is 1. Top quark is the most heaviest fundamental particle in the standard model and will decay before hadronization, 2. The branching ratio of full hadronic decay is the biggest component of Top quark decay(46%).

For a problem that contains a large amount of data and highly require complex and intensive computing resources, machine learning can widely provide powerful support on solving the problem and helps to reduce the CPU time. The machine learning method can facilitate the study and discovery of physics phenomena,an example of which is the remarkable discovery of the Higgs Boson. Both CMS and ATLAS groups apply machine learning methods to promote the search for the Higgs Boson. [1][2]

In this thesis, we developed a novel architecture for the parton-jet assignment problem. This method is base on the state-of-the-art machine learning technology, Attention mechanism.[3] We call this novel ML model **Symmetry Preserving Attention**

NETworks (SPA-NET). By applying attention networks, the SPA-NET is capable of outstanding performance compared to traditional methods while avoiding combinatorial explosion. And thanks to the natural properties of the attention network, the network reflects the permutation symmetry naturally and provides a chance to explore the model with permutation symmetry.

This project was accomplished in collaboration with distinguished physicists from the University of Washington(Shih-Chieh Hsu), University of California Irvine(Mike Fenton, Alexander Shmakov, Daiel Whiteson, and Pierre Baldi). Mike provided an idea of the suitable process to investigate. My jobs was focused on the physics concept, designed the data format, and generate datasets. Also, the traditional event reconstruction method is implemented by my effort. Alexander provide a technical support and machine learning network setup. This project has been submit to the arxiv and under the review of PR. D.[1,2]

Top physics and the concept of machine learning in chapter 2; event generation and simulation configuration are explained in chapter 3. Dataset and event reconstruct using the traditional method and ML approach is explained in chapter 4. Result are shown and discussed in chapter 5; summary and conclusion are presented in chapter 6.

[1] There are two version of submission, https://arxiv.org/pdf/2010.09206.pdf and https://arxiv.org/pdf/2106.03898.pdf

[2] A full code repository containing a general library, the specific configuratio used, and a complete dataset release, are avaiable at https://github.com/Alexanders101/SPANet

Chapter 2

The Top Physics and Machine Learning

2.1 The Top Physics

Top quark, the most massive fundamental particle in Standard Model(SM), is the only quark that decays semi-weakly (i.e. decay into a W bonson and bottom quark). Its large mass leads to a short lifetime and decay before hadronization occurs. Top quark contains many interesting properties, such as its mass, couplings, and cross-section, etc. The accurate measurement these properties will facilitate understanding of fundamental interactions and provide the key to Beyond Standard Model.[4]

In the Standard Model, Top quark pair produced by *pp* collision has three decay modes, **all hadronic channel**, **semi-leptonic channel**, and **dileptonic channel**. The branching ratios of each channel are shown in Table 2.1. The decay width of Top quark predicted in SM is[5]:

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} \left(1 - \frac{M_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_t^2}\right) \times \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] \tag{2.1}$$

Table 2.1: Top quark pair decay process[4]

Decay Channel	Process	Branch Ratio(%)
All-hadronic	$t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'bq''\bar{q}'''\bar{b}$	45.7
Semi-leptonic	$t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'b\ell^-\bar{\nu}_\ell\bar{b} + \ell^+\nu_\ell bq''\bar{q}'''\bar{b}$	43.8
Dileptonic	$t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\nu_\ell b\ell'\bar{\nu}_{\ell'}\bar{b}$	10.5

In recent study, the most precise result of Top quark mass is measured in the lepton+jets channel due to its good signal-to-background ratio and the presence of one neutrino final state.[8] Although the all-hadronic channel has the most probability to appears in the top quark pair decay process, its poor signal-to-background ratio renders inaccurate mass measurements owing to the difficult QCD background. The CMS and ATLAS group approach a precision of Top mass measurement using the all-hadronic channel with uncertainties of 0.65% and 1.1%.[6][7]

In this project, we focus on the **jet-parton assignment problem in all hadronic decay channel**, becuase of the resolved 6 jets signature and the potential of the machine learning method to apply to the ambiguous event reconstruction problem. There exist 6 jets in the final state, 2 b-jets and 4 quark jets, they can be separated into two groups $(b, q, \bar{q})$ and $(\bar{b}, q, \bar{q})$. The schematic of the decay products is shown in 2.1.

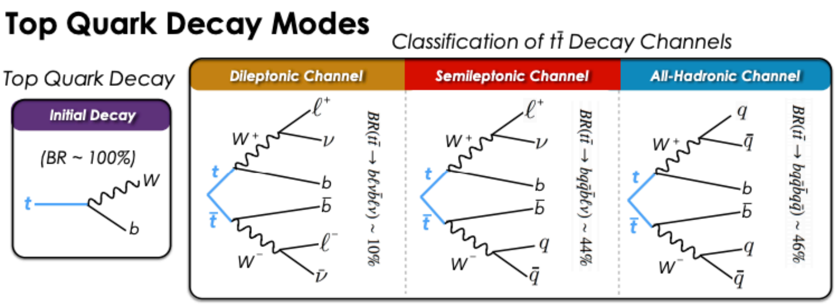

Figure 2.1: The schematic of Top quark decay channels.[8]

2.2 Machine Learning and its application on Particle Physics

Machine Learning techniques have practical applications in many fields(e.g. Computer Vision, Solving PDE, Medical analysis, etc.) in recent age, including particle physics. From the search of the higgs boson(neural network and BDT) to the b-tagging technology(BDT[9]), physicists have already applied several kinds of machine learning methods to recent research.

In a nutshell, machine learning can break into several cases; it can help to do classification, regression, and clustering problems. Machine learning (ML) can not only accelerate the computation of well-defined problems, and can help find new path to unsolved area. This project utilize state-of-the-art machine learning technology, the attention mechanism[3], a technology based on the evolution of Recurrent Neural Networks(RNN).[3] The attention mechanism not only consider the local relationship and the sequence neighbor but also calculates the global relation base on the self-attention calculation shown in Figure 2.2. Using this novel architecture, we will train on the relationship between each jet and try to figure out the correct pair information.

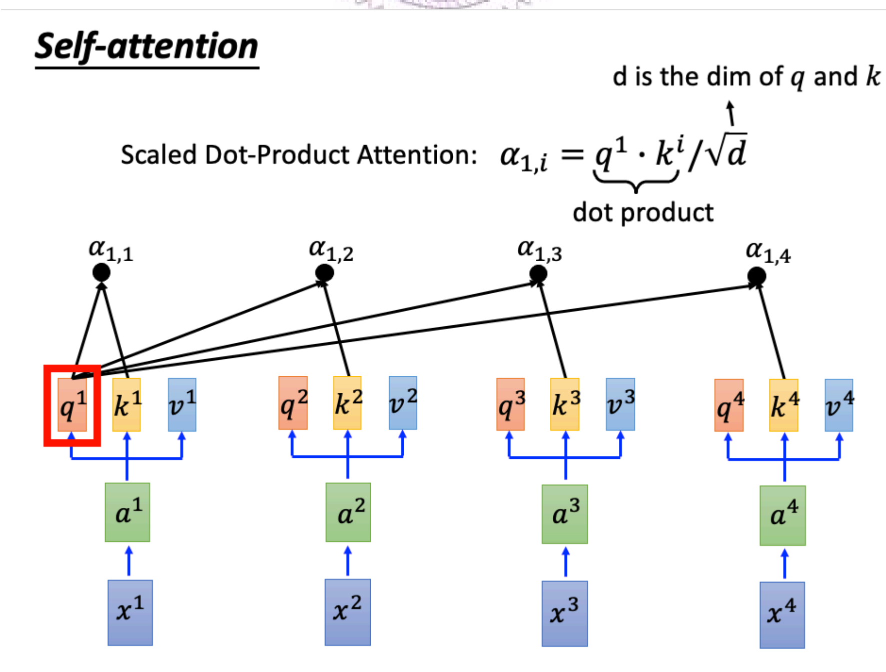

Figure 2.2: How self-attention works.[10]

Chapter 3

Event Generation

3.1 MC samples

For data preparation, we generate our dataset using a custom docker image with MadGraph_aMC@NLO(v2.7.2), Pythia8(v.8.2), and Delphes(v3.4.2) for showering, hadronization, and detector simulation. We apply the ATLAS parametrization during detector simulation. The data are generated at Leading Order(LO) quantum chromodynamics(QCD) and using the PDF set NNPDF23_lo_as_0130_qed. The top mass is configured as $m_{\text{top}} = 173$ GeV. The W quark decay is forced hadronically into a (q, q') pair. The following is our configuration:

```
generate p p > t t~ QED=0, (t > W+ b, W+ > j j), (t~ > w- b~, w- > j j )
output <file_path>
launch <file_path>
shower=Pythia8
detector=Delphes
analysis=OFF
done
set nevents = 10000
set iseed = 1
Delphes/cards/delphes_card_ATLAS.tcl
done
exit
```

Listing 3.1: Configuration for generating samples. The "iseed" is just a placeholder, it will be changed when generating samples.

To get a more general performance, we scan the iseed value from 1 to 30000, each value has 10 thousand events before event selection. The reason for scanning iseed value is that the iseed value is the key to the random data generation. Originally, the program will choose the iseed value randomly and generate different samples. By scanning the issed value, we can make sure the seed is not reused.

Chapter 4

Data analysis and Event reconstruction

In this chapter, we will first explain the event selection, then how we apply the truth matching, traditional reconstruct method, and the machine learning approach.

4.1 Data analysis

4.1.1 Event selection

The top all hadronic decay channel has 2 b-jets and 4 quark jets. In our configuration, all of them are not in the boosted region, which means the daughters of top quarks will not appear with high transverse momentum. Following the event selection used in [8], we apply an event selection that an event should at least exists **2 b-jets** and **4 quark jets** satisfied p_{T} larger than **25 GeV** and $|\eta|$ less than **2.5**. A cutflow table and figure help us to understand how many events are killed by the selections. We may apply 5 cuts and see the evolution of surviving event numbers. The rule of cuts is shown in Table 4.1, and the cutflow is shown in Figure 4.1.

Table 4.1: Rule of cuts. All the cuts require a kinematic limitation that $p_T > 25$ GeV and $|\eta| < 2.5$.

#Cut	Number of b-jets	Number of quark jets
C1	0	4
C2	0	5
C3	0	6
C4	1	6
C5	2	6

The b-tagging and jet information we used here is provided by the Delphes, a detector simulation package.[11] The b-tagging is a method of jet flavor tagging used in CMS and ATLAS.[12][13] This method base on the b-hadron properties, such as the displaced vertex from the primary vertex, large b-hadron mass, large impact parameters(d0), and semi-leptonic e/μ decay of B-hadron. This is also related to the track reconstruction and secondary vertex reconstruction. The Delphes package decide a jet is b-tagged or not based on an efficiency table and returns an array to indicate a jet is b-tagged or not. The number of surviving events is an important factor in our event generation. If the surviving events are very rare, it would mean that our cuts are too tight and the events desired may also possibly be ignored. Also, the lack of surviving events may slow down our data generation efficiency. In Figure 4.1, it shows that there are around $18 \sim 20\%$ events that surviving after C5 cut. This is an acceptable number because we obtain that the events in our dataset will at least contain 2 b-tag jets and 6 jets that passed the kinematics selection without ignoring too many events.

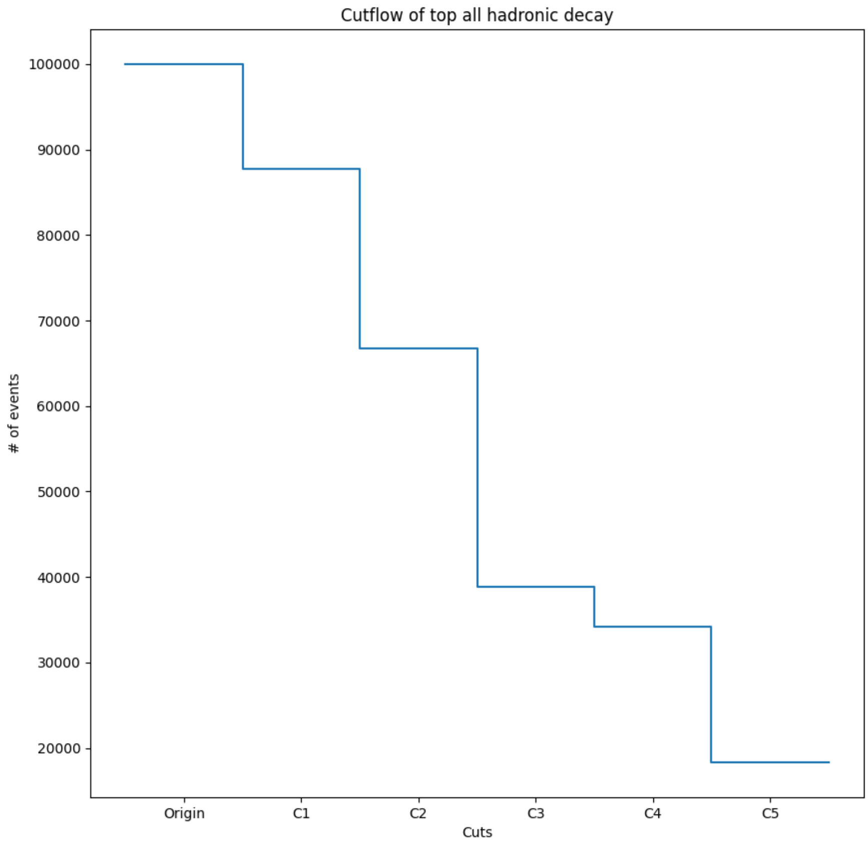

Figure 4.1: Cutflow of all hadronic top decay.

4.1.2 Truth matching

The **truth matching**, which is also called **ΔR matching**, matches the detector simulation(i.e. jet information generated by Delphes) data to truth record(i.e. Parton level information). To calculate the ΔR value, we will find the daughters of top quarks, W boson, and b quark. Once the daughters of top quarks are found, we can find the daughters of W bosons. Finally, we obtain six partons that come from the decay of top quark pairs. These six partons can match the jets identically by considering their distances. The formula of calculating ΔR is:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{4.1}$$

By using the kinematic properties provided in parton level and detector simulation information, we can calculate the ΔR value between each parton and jets. Using the result of the calculation, we may assign each parton to a specific jet.

4.1.3 Custom barcode system

To specify the relation between each parton, and the relation between mothers and daughters, we design a barcode system that helps us to declare the relationship.

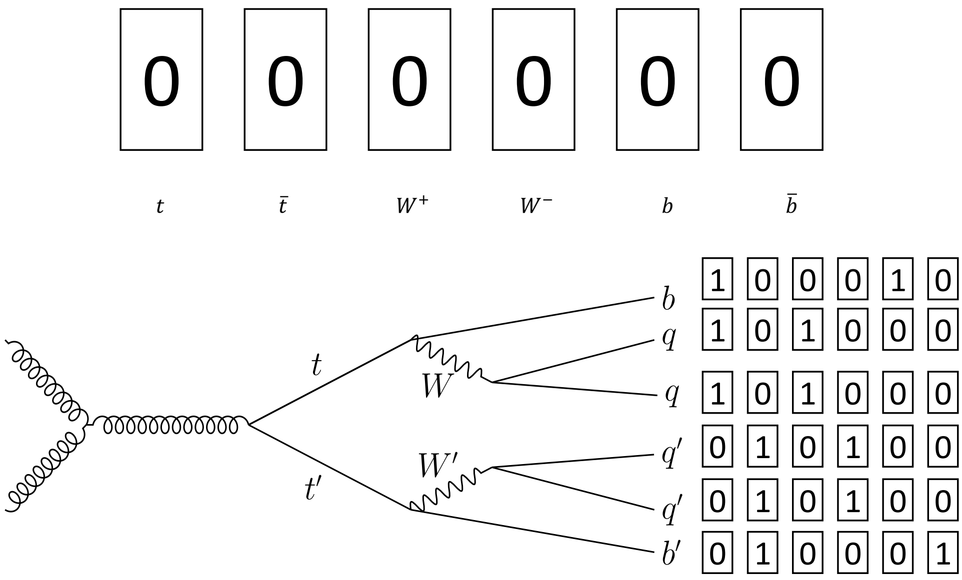

Figure 4.2: Barcode design. These barcodes can provide a pari-wide information to our network.

In Figure 4.2, we define a six-digit barcode, the first two digits are to show which top quark is the mother of this parton and the last four digits of the sequence declare which daughter of the top quark is the mother of parton. Our barcode system breaks six parton (jet) candidates into two subsets which contain 3 elements. Our barcode system is able to specify the relation without losing the information, and also provide a permutation relationship to our network. This will discussed this in the following section.

4.2 Event reconstruction

4.2.1 χ^2 minimization method

The χ^2 minimization method is a traditional method to reconstruct an event. An event that exists 6 jets, it has about $6!/(2 \times 2 \times 2) = 90$ possible combinations, the first two in the denominator is contributed by two b-tag jets, and the middle one and last one is the contributions from two pairs of quark produce by W bosons. The number of possible combinations is proportional to the number of existing jets in an event. The χ^2 minimization method can calculate all the candidates and will try to find the candidate which has the smallest χ^2 value. This method based on the mass of the W boson and top quark. The origin equation of χ^2 minimization in this model is:

$$\chi^2 = \frac{(m_{bqq'} - m_t)^2}{\sigma_t^2} + \frac{(m_{\bar{b}q''q'''} - m_t)^2}{\sigma_t^2} + \frac{(m_{qq'} - m_W)^2}{\sigma_W^2} + \frac{(m_{q''q'''} - m_W)^2}{\sigma_W^2} \tag{4.2}$$

The equation 4.2 has four part. Each parts is a "pull" that is contributed by the component of observables. The parameter σ_W and σ_t is obtained by applying a fitting to the distribution of reconstructed invariant mass of W boson and top quarks. The mass of the W boson and top quark is provided by the recent experiment result. To avoid the bias of top quark candidates, we may combine the first two term into one term by substituting $m_t = \frac{m_{bqq'} + m_{\bar{b}q''q'''}}{2}$, then the equation reduces to:

$$\chi^2 = \frac{(m_{bqq'} - m_{\bar{b}q''q'''})^2}{\sigma^2_{\Delta_{m_{bqq'}}}} + \frac{(m_{qq'} - m_W)^2}{\sigma_W^2} + \frac{(m_{q''q'''} - m_W)^2}{\sigma_W^2} \tag{4.3}$$

Note that there are some events in which the three-jets invariant mass can be far away from the top quark mass but still generate a small χ^2 value. This is because we only consider the difference between two three-jets invariant mass. This can be improved by applying a constraint to the invariant mass[8], but we didn't apply such a constraint in this study.

In this project, we force the b quark candidates in equation 4.3 must be b-tagged jets. This may help to reduce the number of permutations. This restriction may also lead to a incorrect assignment since a jet can be tagged incorrectly.

4.2.2 Machine Learning Approach

For a machine learning model, equivariance and invariance are important properties that may affect the performance of the model. Such as a computer vision problem, the object should be invariant under translation to prevent affect the prediction. The translation, rotation, and shift of the position should not change the prediction of the model because the object remains the same object. The Convolutional Neural Network(CNN) can produce object recognition outcomes that are invariant under translations. The properties of invariance can be generalized to another geometry structure, e.g. manifolds and groups. In all hadronic top decay, we have two subsets with the same elements (b, q, q') and $(\bar{b}, q'', q''')$. These subsets should remain invariant under permutations of the input jets order. The reason that the order of input jets should not affect the result is because the permutation symmetry is not base on the order of jets but the pair-wise permutation.

By the invariant feature of attention architecture, rearranging the elements in a sequence leaves the attention weight unchanged. This permutation symmetry present in the attention-based model may be used to render the efficient reconstruction of the all hadronic top decay process. In this case, the network output of the all hadronic top decay process should identify two distinct interchangeable subsets, and each contains an interchangeable qq' pair produced by the W bosons. This invariant property on the output is the unique feature of our dataset and the model should take into account.

We propose an attention-based network, called **Symmetry Preserving Attention NETwork(SPA-NET)** in this project. Its structure is shown in Figure 4.3. The input of SPA-NET is a list of unsorted jet information, with their 4-momentum $(p_{\mathrm{T}}, \eta, \phi, M)$ as well as the b-tag information provided by Delphes. We take the logarithm to M and p_T and normalize all the components to have zero mean and unit variance. The input jets will be sent to the network and be embedded into a D-dimensional latent space representation. This D-dimensional latent space is obtained by progressively increasing the latent dimensionality of the input jets up to the final dimensionality D(This operation is done by the embedding blocks in the whole stack). We target this latent space dimensionality D to a value 128 with the following sequence: 8 → 16 → 32 → 64 →128. After embedding, the output will be sent to a stack of transformer encoder layers, this layer will learn the relationship between each element in the input sequence. When the encoding is finished, the encoded output will be forwarded into an important architecture in this network - a two-branched structure with each branch able to compute the output individually. Each branch has a transformer encoder layer whichl extract the information from the top quark and a tensor attention layer which produce the top quark distribution. The output distribution predicts the top quark triplets. Figure 4.4 is an example of network outputs; note that the attention mechanism is an architecture that allows the network to propagate the information selectively by using a "mask". This enable the neural network to learn from partial information and update the parameters with selected information. Using the "mask" also allow the network to infer the relationships between different elements in a sequence.

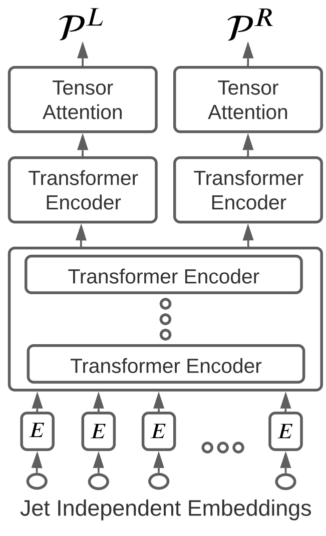

Figure 4.3: High-level structure of SPA-NET.

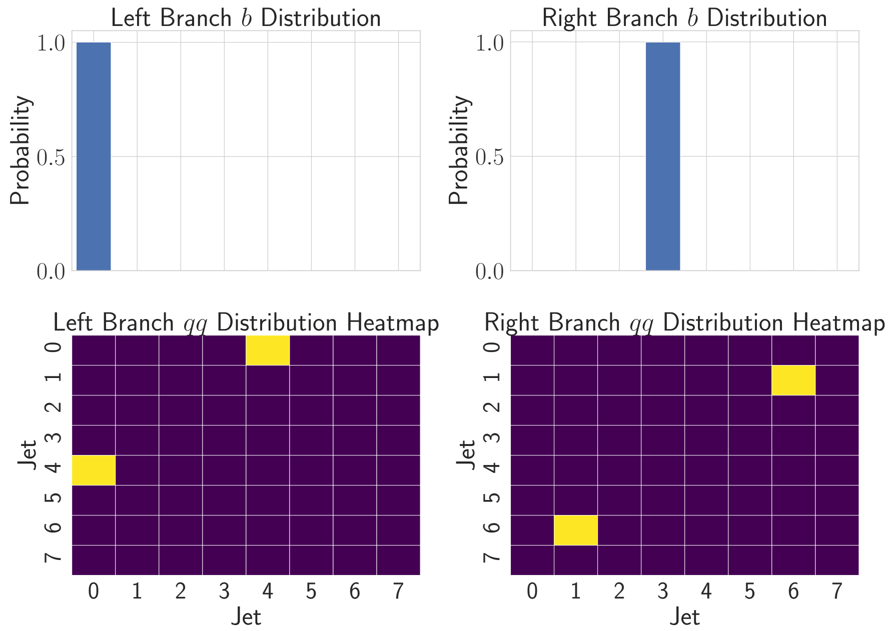

Figure 4.4: A visualization of the example single event output produced by SPA-NET. The top two plots are the projected b quark distribution, and the bottom two plots are the qq distribution respectively.

The most important part in this network is the **Symmetry Preserving Tensor Attention**. Consider a set of weights $\theta \in \mathbb{R}^{D\times D\times D}$, this θ is not inherently symmetric at all. To make the θ become an invariant attention weighting, we apply the following transformation(eq. 4.4). This transformation will transform the θ into an auxiliary weights tensor $S^{ijk} \in R^{D\times D\times D}$. Using S^{ijk} and the embedded jets tensor $X \in \mathbb{R}^{N\times D}$(N is the number of jets), we can calculate the dot-product attention. The dot-product works in flat Euclidean space and produces the output tensor O^{ijk}. The summation product tensor S^{ijk} guarantees that the interchange of the first two dimensions of S will be symmetric and ensures that $O^{ijk} = O^{jik}$. These properties enforce the qq' invariance.

$$
\begin{aligned}
S^{ijk} &= \frac{1}{2}\left(\theta^{ijk} + \theta^{jik}\right) \\
O^{ijk} &= X_n^i X_m^j X_l^k S^{nml}
\end{aligned}
\tag{4.4}
$$

To obtain the probability distributions P^L and P^R, a 3-dimensional softmax is applied on O^{ijk} to generate the joint triplet probability distribution.

$$P(i,j,k) = \frac{exp\ O^{ijk}}{\sum_{ijk} exp\ O^{ijk}} \tag{4.5}$$

Equation 4.5 is used to produce the individual probability distribution of two top quarks and to produce the single triplet from each by selecting the peak of these distributions.

During the training, a suitable loss function is needed to deal with the double output probability distributions. We design the loss function based on the cross-entropy between the output probability and truth distribution on the all hadronic top decay. The loss function must ensure the symmetry of the top quark pairs which are invariant concerning the permutation $tt' \leftrightarrow t't$. A symmetry loss function $\mathcal{L}$ by the following function:

$$\mathcal{L} = min(\mathcal{L}_1(P^L, T_1, P^R, T_2), \mathcal{L}_1(P^L, T_2, P^R, T_1)) \tag{4.6}$$

$$\mathcal{L}_1(P_1, T_1, P_2, L_2) = \mathcal{H}(T_1, P_1) + \mathcal{H}(T_2, P_2) \tag{4.7}$$

Where $\mathcal{H}$ is the general cross-entropy. It is possible that both branches produce the same output pairs. To make sure the network produces unique predictions, the one with the higher probability is selected and the other one is re-evaluated;the loss function is then computed. Figure 4.4 is an example of the output produced by SPA-NET.

Chapter 5

Result and Discussion

In this chapter, we will discuss our results and evaluate the performance of the machine learning approach compared to the traditional method.

5.1 Invariant mass and reconstruct efficiency

Before discussing the result, we should define the category for the reconstruction efficiency. Treating the truth matching result as the target, the prediction of χ^2 or SPA-NET may produce three kinds of results:

- Correct matched: An event in which both top quarks are correctly predicted by the reconstruction method.
- Incorrect matched: An event that one or both top quarks is incorrectly predicted by the reconstruction method.
- Unmatched: An event that one of the truth match results contains an element that does not match to any jets.

Based on the category above, the reconstruction efficiency can be computed. The reconstruction efficiency is separated into two components: event-based efficiency and quark-based efficiency. The event-based efficiency is calculated based on how many **events** are matched correctly, incorrectly, and unmatched. The quark-based efficiency is based on how much **top quarks** are assigned correctly. The efficiency is shown in the table below:

Table 5.1: Using ϵ as the symbol of efficiency. This table performs the efficiencies of the χ^2 and SPA-NET assignments assessed by per-event efficiency ϵ^{event} and per-top efficiencies ϵ^{top} inclusively and by jet multiplicity N_{jets}. The subscript of ϵ_1^{top} and ϵ_2^{top} is stands for the one/two reconstructable events.

	χ^2 Method			SPA-NET		
N_{jets}	ϵ^{event}	ϵ_2^{top}	ϵ_1^{top}	ϵ^{event}	ϵ_2^{top}	ϵ_1^{top}
6	61.8%	65.0%	24.2%	80.7%	84.1%	56.7%
7	40.8%	50.4%	24.6%	66.8%	75.7%	56.2%
≥8	23.2%	35.5%	20.2%	52.3%	66.2%	52.9%
Inclusive	**37.7%**	**47.0%**	**23.0%**	**63.7%**	**73.5%**	**55.2%**

On overall events, the χ^2 has performed a 37.7% efficiency, while SPA-NET archieves an efficiency 63.7%. The χ^2 method had the best performance on the 6 jets category but perform worse where an event contains more than 8 jets. The SPA-NET performs much better than the χ^2 in all categories. For the event with two identifiable top quarks, the χ^2 method achieved an efficiency ϵ_2^{top} 65.0%, but SPA-NET archives an ϵ_2^{top} 84.1%. Since we train the SPA-NET with the events that contain two identifiable top quarks, it is reasonable that the ϵ_1^{top} of SPA-NET is lower than ϵ_2^{top}. Moreover, since our definition of equation 4.2 is based on the difference between reconstructing invariant masses of 2 two quark candidates, it may be difficult for the χ^2 method to assign the jet properly in an event that only contains one identifiable top quark. Note that in our evaluation dataset, 8.1% of events in which both tops are identifiable have at least one b-quark matched to non-b-tagged jets, which are impossible for the χ^2 method to reconstruct correctly. These quarks were reconstructed by SPA-NET with an efficiency of 29.4%.

5.1.1 Reconstructed invariant mass

Using χ^2 minimization, we may obtain the best assignment under the constraint of parameters. In this project, the parameters are configured as: $m_W = 81.3$ GeV, $\sigma_W = 12.3$ GeV, and $\sigma_{m_{bjj}} = 26.3$ GeV. The parameters σ_W and $\sigma_{m_{bjj}}$ are found by fitting the mass distribution of the W boson and top quark. While computing χ^2 value, we use the b-tagging information to assign the b quark candidates.
The following plots is the reconstructed mass distribution of W boson and top quark using χ^2 minimization method and SPA-NET.

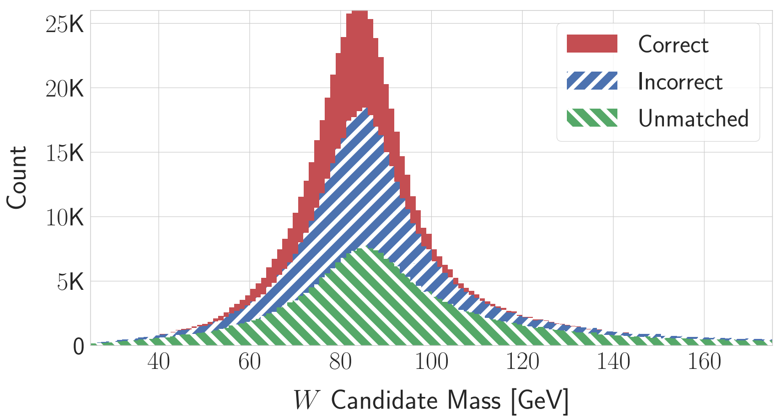

Figure 5.1: W boson mass reconstructed by χ^2 minimization method

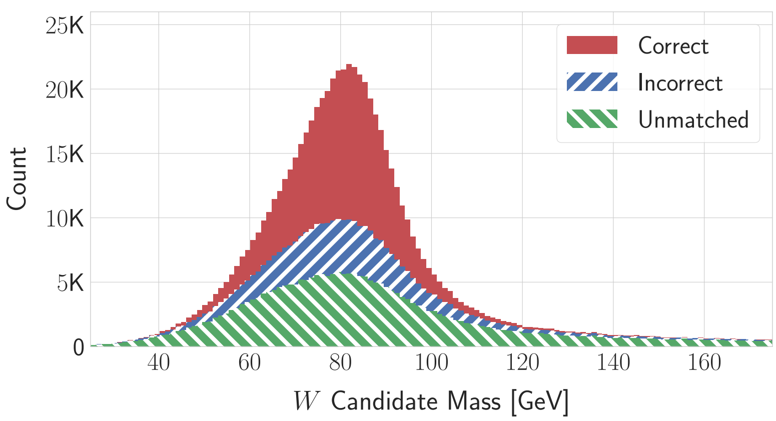

Figure 5.2: W boson mass reconstructed by SPA-NET

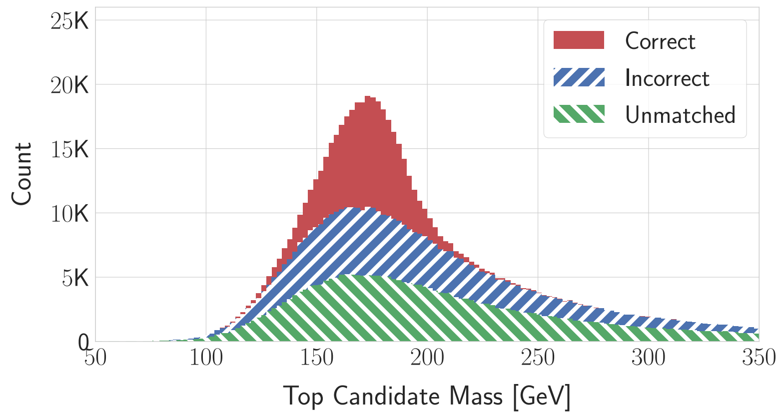

Figure 5.3: Top quark mass reconstructed by χ^2 minimization method

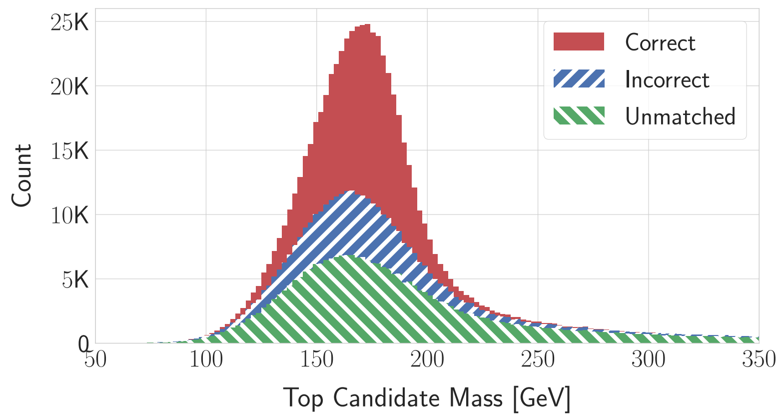

Figure 5.4: Top quark mass reconstructed by SPA-NET

Comparing the result shown in Figure 5.1 and Figure 5.2, we found the χ^2 method has a narrower peak around W boson mass than the SPA-NET. This narrower peak is due to the incorrect and unmatched events and can be explained by the presence of m_W in equation 4.2. Figure 5.4 and Figure 5.3 show that the SPA-NET has the more peaked distribution compared to the χ^2 method with comparable incorrect/unmatched events.

5.1.2 ROC curve

The Receiver operating characteristic (ROC) curve is a useful tool to evaluate the performance of a machine learning model. The ROC curve of SPA-NET applied on events with one and two reconstructable top quarks is shown in Figure 5.5 and Figure 5.6. Note that the targets are defined as 1 if the prediction is correct; 0 if the prediction is incorrect.

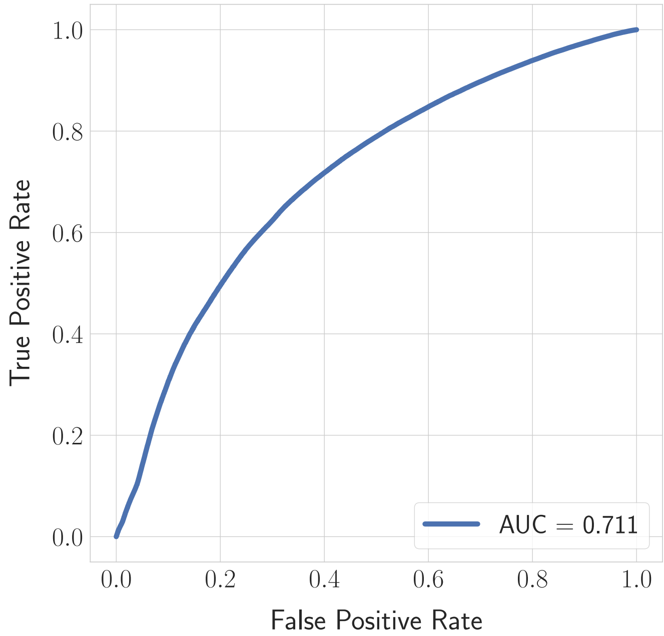

Figure 5.5: ROC curve of SPA-NET applies on events with one reconstructable top.

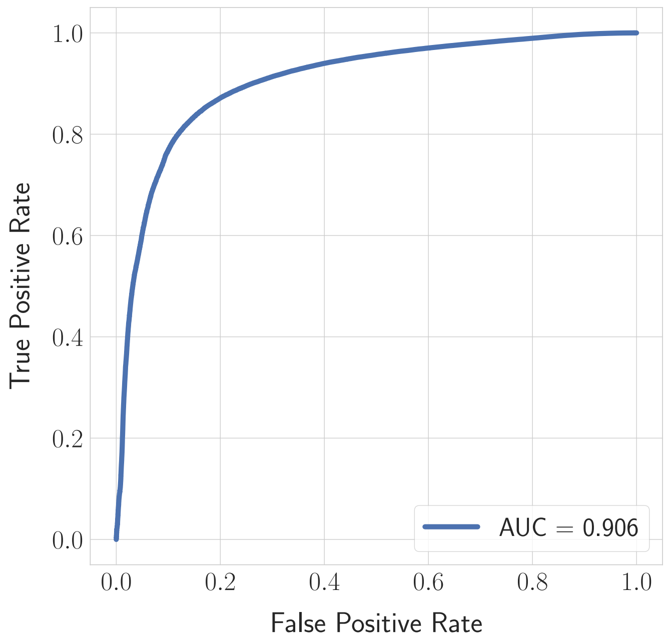

Figure 5.6: ROC curve of SPA-NET applies on events with two reconstructable top quarks.

With regard to the more efficient performance of SPA-NET discussed in the last paragraph of 5.1, our network achieved a lower AUC value in 5.5 and performs a remarkable ROC value in 5.6.A possible solution we did not implement in this study is to train with the “partial” events by using the “mask”.[14][15]

5.2 Reduction of computing time

The time required to compute SPA-NET is much lower than the χ^2 needed. We may compare the time they needed by considering the time complexity. The χ^2 has a time complexity that scales approximately as $P(N, 6) = \mathcal{O}(N^6)$ where N is the number of jets in an event. The time complexity is χ^2 proportional to the number of jets and this makes a limitation of maximum jets indeed. Considering a 2019 DELL XPS13 computer with Core i7-1065G7 1.30GHz CPU, the SPA-NET took an average of 4.4 ms for one event. The χ^2 took an average 20 ms in 6 jets events and 369 ms in ≥ 8 jets events.

5.3 Outlook

Based on the design of SPA-NET, the input is a sequence of jet information and their relationships. The network considers the permutation relation between each subset and possible permutations. In case, a physics model which has a permutation relationship might be a good item to explore with SPA-NET. For example, ttH all hadronic decay process or all hadronic four top decay process may be well-suited to study with SPA-NET. Furthermore, SPA-NET may have potential applications not only for the parton-jet assignment problem but for the clustering problem, graph matching problem, and so does another problem which contains a permutation symmetry. We plan to explore the SPA-NET with some interesting BSM problem, such as the tri-Higgs production in 2HDM model.[16]

Chapter 6

Conclusion

We accomplishd a novel approach to parton-jet assignment using a symmetry preserving attention mechanism. This network is able to learn the permutation symmetry that appears in a physics process and achieves a remarkable performance. Utilizing this architecture can significantly improve performance and reduce the cost of computing time. We were also able to show that a well-trained deep learning network can replace permutation-based algorithms and avoid the combinatorial explosion by estimating symmetry-aware pairwise similarities.
Also, this novel architecture can not only enables more efficient study of hadronic top decay parton-jet assignment problems; it also can be generalized to other models which also contain the permutation symmetry.

Bibliography

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," Phys. Lett. B **716**, 1-29 (2012) [arXiv:1207.7214 [hep-ex]]. https://arxiv.org/abs/1207.7214.

[2] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," Phys. Lett. B **716**, 30-61 (2012) [arXiv:1207.7235 [hep-ex]]. https://arxiv.org/abs/1207.7235.

[3] A. Vaswani *et al.*, "Attention is all you need," Advances in Neural Information Processing Systems, NIPS (2017) [arXiv:1706.03762[sc.CL]]. https://arxiv.org/abs/1706.03762.

[4] P. A. Zyla *et al.* [Particle Data Group], "Review of Particle Physics," PTEP **2020**, no.8, 083C01 (2020) doi:10.1093/ptep/ptaa104

[5] A. Quade, "Top quark physics at hadron colliders," doi:10.1007/978-3-540-71060-8. https://www.springer.com/gp/book/9783540710592.

[6] CMS Collaboration, "Measurement of the top quark mass in the all-jets final state at $\sqrt{s} = 13$ TeV and combination with the lepton+jets channel," Eur. Phys. J. C **79**, no.4, 313 (2019) [arXiv:1812.10534 [hep-ex]]. https://arxiv.org/abs/1812.10534.

[7] ATLAS Collaboration, "Top-quark mass measurement in the all-hadronic $t\bar{t}$ decay channel at $\sqrt{s} = 8$ TeV with the ATLAS detector," JHEP **09**, 118 (2017) [arXiv:1702.07546 [hep-ex]]. https://arxiv.org/abs/1702.07546.

[8] T. Mccarthy, "Measurement of the Top Quark Mass in the All-Hadronic Top-Antitop Decay Channel Using Proton-Proton Collision Data from the ATLAS Experiment at a Centre-of-Mass Energy of 8 TeV," CERN-THESIS-2015-275. https://cds.cern.ch/record/2125782.

[9] ATLAS Collaboration, "Machine Learning Algorithms for *b*-Jet Tagging at the ATLAS Experiment," J. Phys. Conf. Ser. **1085**, no.4, 042031 (2018) [arXiv:1711.08811 [hep-ex]]. https://arxiv.org/abs/1711.08811.

[10] Hung-Yi, Lee, "NTU course lecture note-Transformer" https://speech.ee.ntu.edu.tw/ hylee/ml/ml2021-course-data/self_v7.pdf

[11] J. de Favereau *et al.* [DELPHES 3], "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," JHEP **02**, 057 (2014) [arXiv:1307.6346 [hep-ex]]. https://arxiv.org/abs/1307.6346.

[12] ATLAS Collaboration, "Optimisation of the ATLAS *b*-tagging performance for the 2016 LHC Run," ATL-PHYS-PUB-2016-012. https://cds.cern.ch/record/2160731

[13] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," JINST **13**, no.05, P05011 (2018) [arXiv:1712.07158 [physics.ins-det]]. https://arxiv.org/abs/1712.07158

[14] M. J. Fenton, A. Shmakov, T. W. Ho, S. C. Hsu, D. Whiteson and P. Baldi, "Permutationless Many-Jet Event Reconstruction with Symmetry Preserving Attention Networks," [arXiv:2010.09206 [hep-ex]]. https://arxiv.org/abs/2010.09206.

[15] A. Shmakov, M. J. Fenton, T. W. Ho, S. C. Hsu, D. Whiteson and P. Baldi, "SPANet: Generalized Permutationless Set Assignment for Particle Physics using Symmetry Preserving Attention," [arXiv:2106.03898 [hep-ex]]. https://arxiv.org/abs/2106.03898.

[16] I. Low, N. R. Shah and X. P. Wang, "Higgs Alignment and Novel CP-Violating Observables in 2HDM," [arXiv:2012.00773 [hep-ph]]. https://arxiv.org/pdf/2012.00773.pdf.