RESEARCH PROPOSAL

JIN CHOI[*]

CONTENTS

LIST OF FIGURES

LIST OF TABLES

* *Department of Physics and Astronomy, Seoul National University, Seoul, Republic of Korea*

1 INTRODUCTION

In 2012, a spin 0 particle with its mass of 125 GeV has been discovered, which can be interpreted as Higgs Boson(H) in the Standard Model(SM). The discovery of the particle completed the theory, but there are still many anomalies that cannot be explained in SM. For example, the mass hierachy problem, matter-antimatter asymmetry, or the existence of the dark matter are still not configured yet. Therefore, many theories so called Beyond the Standard Model(BSM) have been constructed and many of them extend the singlet Higgs sector to be doublet. In general two Higgs doublet model(2HDM), the theory predicts 5 physical Higgs eigenstates: two charged Higgs bosons($H^{\pm}$), two neutral Higgs bosons(H, h), and one pseudoscalar Higgs boson(A). The already founded particle can be one of the neutral Higgs boson, so if it is the light neutral Higgs boson, then it is called normal scenario(NS), otherwise inverted scenario(IS).

Because of 5 physical Higgs boson eigenstates, various decay modes of $H^{\pm}$ can be established. The most actively searched decay modes in theLHC are fermionic decay modes such as $H^{+} \to t\bar{b}$, $H^{+} \to c\bar{s}$, and $H^{+} \to \tau^{+}\nu$. These channels are used to set upper limits on the mass of H^{+} and $\tan\beta$. There are also bosonic decay modes such as $H^{+} \to hW^{+}$, $H^{+} \to HW^{+}$, and $H^{+} \to W^{+}A$. Especially, $H^{\pm}W^{\mp}A$ coupling does not depend on $\tan\beta$, so if it is kinematically allowed, $H^{+} \to W^{+}A$ can be the most dominant decay mode and highly suppress the fermionic decay modes. The importance of the $H^{+} \to W^{+}A$ decay mode was well known in the LEP era. The DELPHI and OPAL experiment set the lower bound in the $m_{H^{+}}$ at 72 GeV in type 1 2HDM with $m_A \in$[12, 70] GeV using $e^{+}e^{-} \to H^{+}H^{-}$ production. In Tevatron, the CDF experiment searched the $H^{+} \to W^{+}a \to W^{+}\tau^{+}\tau^{-}$ decay mode with $m_{H^{+}} \in$ [90, 160] GeV and $m_A \in$ [4, 9] GeV, setting the upper limit on branching fraction $B(t \to bH^{+} \to bW^{+}a \to bW^{+}\tau^{+}\tau^{-})$ as 8 to 50%. At the LHC Run1, there was no search for $H^{+} \to W^{+}A$ decay mode, but the Atlas experiment searched another bosonic decay mode $H^{+} \to W^{+}Z$. In Run 2, only the CMS experiment searched $H^{+} \to W^{+}A$ decay mode from top quark pair production using 35.9fb^{-1} data collected in 2016. The research have set the upper limit on $\sigma(pp \to t\bar{t}) \times B(t \to bH^{+} \to bW^{+}A \to bW^{+}\mu^{+}\mu^{-})$ below 15fb, assuming $m_{H^{+}} \in$ [70, 160] GeV and $15\text{GeV} < m_A < m_t - m_{H^{+}}$.

In this research, we will use the full Run2 LHC data collected by the CMS experiment to enhance the sensitivity searched in the previous analysis. In addition, we will also investigate the off-shell decay of $H^{+} \to W^{+}A$ decay mode ($m_{H^{+}} < m_A + m_{W^{+}}$). Previous searches have been avoided this pseudoscalar Higgs mass region because of the large background from Z associated productions(e.g. $t\bar{t} + Z$).

2 DATA AND SIMULATED SAMPLE

Table 1: Sample Information

Name		
First name	Last Name	Grade
John	Doe	7.5
Richard	Miles	2

3 SEARCH STRATEGY

In this analysis, we investigate H^+ and A with the mass range of 70 GeV < m_{H^+} < 160 GeV and 15 GeV < m_A < m_H^+ - 5 GeV using 139 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV collected during the LHC Run 2 by the CMS experiment. Here, the mass of H^+ has been set to be greater than the lower bound from LEP experiment and less than that of the top quark. The mass of A has been set to avoid the low mass resonances but allow the off-shell decay from H^+. The most dominant production channel in this mass range is via top quark pair production, and we search for the cascade decay chain $H^+ \to W^+A \to W^+\mu^+\mu^-$. Assuming at least one of the W bosons decay leptonically, the final state contains $e\mu\mu$ of $\mu\mu\mu$ + b-jet signature. The dominant background in this channel is leptonically decaying top quark pair production, with one of the jets being misidentified as a muon or an electron(non-prompt background). Prompt contribution from $WZ \to 3l + \nu$ and $ZZ \to 4l$(VV), $TT + X(X = W, Z, H)$ are also contributed to the dominant background sources. Drell-Yan processes with bremsstrahlung of final state leptons, referred to as conversion background, are also estimated. Some rare processes such as $H \to ZZ \to 4l$ with H produced either by gluon fusion or vector boson fusion, $VVV(V = W, Z)$, TTTT are also considered.

Baseline selection of events has been build based on the signal event topology:

- 3 muons 1 electron + 2 muons with at least one OS muon pair
- least two jets
- at least one b-tagged jet

To maximize the sensitivity, graph-neural-network(GNN) based discriminators are trained to separate the signal events and the non-prompt or VV backgrounds. Graph based data structure is the most natural way to represent the collider events, since the graph can easily represent the event topology and no intrinsic order between objects in a single event is needed. Other ML techniques such as boosted decision tree(BDT) or deep neural network(DNN) frequently use a high-level inputs contructed from the objects and should represent the event in a list, which have to make a manual ordering(e.g. PT ordering in jets). In GNN, each object in an event can be represented as a node and their relation such as ΔR can be represented as an edge of two objects. The network aggregate the information in the last layer to determine the score(or probability) of the event to be signal or background. The performance of the GNN discriminator classifying $(m_H^+, m_A) = (130, 90)$ GeV vs leptonically decaying top-pair production is given in figure 1.

The event rate for each background source is measured in different manner. For the nonprompt background, the contribution is estimated using tight-loose method. In this method, we use single lepton events to measure the probability of nonprompt leptons passing loose ID criteria but not tight ID criteria. The probability is used to estimate the event rate in application regions by extrapolating the event rates that not all the leptons passing the tight ID criteria. For the conversion background, since the MC modelling of the conversion leptons might be ill-modelled, the scale factors for $Z + \gamma$ samples are measured in separated control region which is orthogonal with the signal region. The other prompt backgrounds are estimated directly from the MC samples.

For each signal mass point, the sensitivity will be maximized by giving cuts on the dimuon mass spectrum and signal versus background GNN discriminator score. The result will be interpreted as the signal production rate in each mass point using profile likelihood method.

4 CURRENT STATUS AND FUTURE PLANS

In this analysis, electrons, muons, and jets are used to describe the events. The object definition are finalized. For muons, measuring the ID efficiency in data and simulation has been done with corresponding trigger efficiency, and the correction factor has been applied to each MC samples to describe the data better. Also, the nonprompt probability has been measure to estimate the nonprompt background using tight-loose method. For electrons, even if the definition for tight and loose ID criteria has been finalized, its efficiency and nonprompt probability have not been measured yet. Jets and missing transverse momenta definitions are provided from the CMS central POG team.

Since we have not measured the correction factors and nonprompt probability of electrons yet, training the GNN discriminator for $\mu\mu\mu$

channel has been done in first. We will use the discrimination score and the dimuon mass spectrum to check the preliminary sensitivity soon.

After measuring the electron efficiency and nonprompt probability, the result from $e\mu\mu$ channel will be supplemented. Also, systematic sources have not been considered properly, for example the b-tagging efficiency correction for jet energy resolution. It will also be accounted in the final interpretation.

The result will be published in the name of CMS collaboration within an year.

REFERENCES