RESEARCH PROPOSAL

JIN CHOI*

CONTENTS

1 Introduction 2
2 Data and Simulated samples 3
3 Object Definition 3
4 Search Strategy 4
5 Current Status and Future Plans 6

LIST OF FIGURES

Figure 1 Feynman diagram for the signal events. 5

LIST OF TABLES

Table 1 Background Samples 4

* *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[1][2], which can be interpreted as Higgs Boson(H) in the Standard Model(SM)[3]. 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 understood well 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)[4], 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).

Since there are no mass relation in 5 physical eigenstats of the Higgs bosons, various decay modes of $H^{\pm}$ can be established. The most actively searched decay modes in the LHC are fermionic decay modes such as $H^+ \to t\bar{b}$[5], $H^+ \to c\bar{s}$[6], and $H^+ \to \tau^+\nu$[7]. These channels are used to set upper limits on the mass of H^+ and $\tan\beta$, depending on the type of each 2HDM model. There are also bosonic decay modes[8] 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[9]. 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[10][11]. 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 the branching fraction $B(t \to bH^+ \to bW^+a \to bW^+\tau^+\tau^-)$ as 8 to 50%[12]. 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 $35.9\,fb^{-1}$ data collected in 2016. The research have set the upper limit on $B(t \to bH^+ \to bW^+A \to bW^+\mu^+\mu^-) \leqslant 2.9\%$ at 95% CL, assuming $m_{H^+} \in [70, 160]$ GeV and $15GeV < m_A < m_t - m_{H^+}$[13].

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 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. $t\bar{t} + Z$), but increased number of events and using novel MVA methods made the search possible.

2 DATA AND SIMULATED SAMPLES

LHC full Run2 proton-proton collision data, which is total integrated luminosity of $139 fb^{-}1$, collected by the CMS experiments is used in this analysis. The data has been triggered by either doublemuon triggers or electron-muon triggers depending on the final state leptons as described in section 4. More detail description of the CMS detector and the triggering system can be found in Ref. [14] [15].

The signal events are generated using Madgraph_aMC@NLO v5.2.6.5[16] in leading order. General 2HDM model cards restricted by type 1 scenario has been used. The mass of the heavy Higgs boson is set to be 5TeV to avoid its production or decay, while the mass of the light Higgs boson is set to be 125 GeV. The mass of charged Higgs boson is set to be between 70 GeV and 160 GeV, and the mass of the pseudoscalr Higgs boson is set to be between 15 GeV and $m_{H^+} - 5$ GeV. Each sample is normalized to 15fb. Background samples are generated using Powheg Box v2[17] or Madgraph_aMC@NLO v5.2.6.5[16] up to leading order(LO) or next-leading order(NLO) in matrix element calculation, and some samples are normalized up to higher order precision. More information of the background sample generation can be found in Table 1. For all samples, the actual detection process was implemented through Geant4 detector simulation[18].

3 OBJECT DEFINITION

Final objects are reconstructed by particle flow(PF) algorithm[19], which not only use the subdetector information but use all the subdetector information for better identify the final objects. The major objects we use in this analysis is electrons, muons, and jets. Electrons use tracker to measure it's position and momentum combined with the ECAL information[20]. Tracks that matched to the muon stations can be identified as muons[21]. Other objects, such as charged hadrons, neutral hadrons or tau lepton can also be identified, while neutrinos escape from the detector but it's transverse momentum can be measured as negative vector sum of the p_T of all the PF candidates. Jets use all the PF candidates using anti-kt algorithm[22] with cone size of 0.4. To identify b-jets, all the constituents are used to train the tagger[23], which gives scores to be light, c or b. Here, b-tagged jets are called b-jets.

Table 1: Background Samples

Sample	ME Generator	Xsec(pb)	Precision
TTLL	Powheg	88.29	NNLO
Z+jets	MC@NLO	6077.22	NNLO
$WZ \to 3l\nu$	MC@NLO	5.213	NLO
$ZZ \to 4l$	Powheg	1.256	NLO
$Z\gamma \to ll\gamma$	MC@NLO	51.1	NLO
TT+W ($W \to l\nu$)	MC@NLO	0.2043	NLO
TT+Z ($Z \to 2l2\nu$)	MC@NLO	0.2529	NLO
TT+H ($H \nrightarrow bb$)	Powheg	0.2151	NNLO QCD+NLO EW
tZq	MC@NLO	0.0758	NLO
tHq	Madgraph	0.7464	LO
VVV(V = W, Z)	MC@NLO	0.44333	NLO
TTTT	MC@NLO	0.009103	NLO
VBF $H \to ZZ \to 4l$	Powheg	0.001034	NLO
ggH$\to$ZZ$\to$4l	Powheg	0.01212	NLO

4 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 charged Higgs production channel in this mass range is through 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 $\mu\mu\mu$ + multijet + b-tagged jet signature. The Feynman diagram of the signal event has been shown in figure 1. 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. If m_{H^+} m_t, b-jets branching from the top quark associated

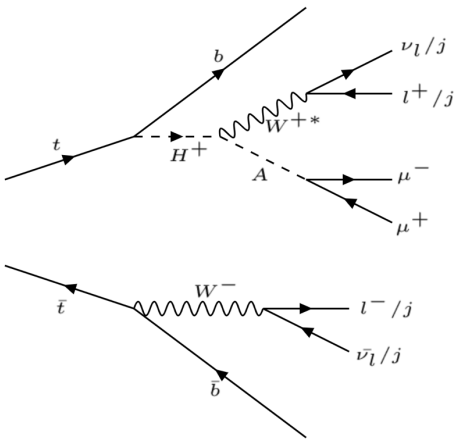

Figure 1: Feynman diagram for the signal events.

with the charged Higgs can be very soft. Not to loose such signal events, instead of requiring 2 b-jets, we only require single b-jet. The final baseline selection is as below:

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

By using the opposite signed muon pair, it is easy to reconstruct the pseudoscalar candidates. However, to reconstruct the charged Higgs candidate is not a simple job. Instead of reconstructing the charged Higgs, we decided to use the final objects' kinematics to discriminated irreducible background further. Graph-neural-network(GNN) based discriminators[24][25] are trained to separate the signal events and the nonprompt or $TT + X$ backgrounds. To make the network sensitive to the final states' kinematics, Graph based data structure is the most natural way to represent the data, since the graph is length-free and permutation invariant data structure. Here, we only used P_T, η, ϕ, mass, charge, PID(=muons, electrons, jets), b-tag score for the discrimination. Other ML techniques such as boosted decision tree(BDT) or deep neural network(DNN) frequently use high-level features constructed from several objects and represent an event as a list, which have to assume a manual ordering(e.g. PT ordering in jets) and have to drop some information or zero-padded the list to avoid the blank in the list. In GNN, each object in an event can be represented as a node and their relation such as ΔR or reconstructed mass can be represented in the edge attribute of two objects. The network convolute the nearby objects and generate a new graph with different number of node attributes and edge attributes. The information of the final graph is aggregated in the readout layers and make a fixed length of

list, so can be flow to the final probability, which is the score to be a signal event.

The event rate for each background source is measured in a different manner. For the nonprompt background, the contribution is estimated using tight-loose method from the data. In this method, we use single lepton events to measure the probability of nonprompt leptons passing the loose ID but not tight ID. The probability is used to estimate the expected event rate in the application regions by extrapolating the event rates that not all the leptons passing the tight ID. 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 the separated control region which is dominated by $Z + \gamma$ events and orthogonal to the baseline selection. The other prompt backgrounds such as VV or $TT + X$ are estimated directly from the MC samples.

In the final sensitivity calculation, the most dominant backgrounds, TT+nonprompt and VV, $TT + X$, will be suppressed using the GNN response of each event. The final limit will be extracted using the dimuon mass spectrum, which is the candidate of the pseudoscalar Higgs boson. The result will be interpreted as the signal production rate in each mass point using profile likelihood method[26].

5 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

[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 (2012) 1*.

[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 (2012) 30*.

[3] P.W Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett. 13*, pages 508–509, 1964.

[4] G.C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". *Phys. Rep. 516*, 2012.

[5] The CMS Collaboration. "Search for a charged Higgs boson decaying into top and bottom quarks in proton-proton collisions at $\sqrt{s}$ = 13 TeV in events with electrons and muons". *JHEP 01*, 2020.

[6] The CMS Collaboration. "Search for a light charged Higgs boson in the $H^{\pm} \to cs$ channel in proton-proton collisions at $\sqrt{s}$ = 13 TeV". *Phys. Rev. D 102, 072001*, 2020.

[7] The ATLAS Collaboration. "Search for charged Higgs bosons decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in the τ+lepton final states with 36 fb^{-1} of pp collision data recorded at $\sqrt{s}$ = 13 TeV with ATLAS experiment". *JHEP 09*, 2018.

[8] S. Moretti A. Arhrib, R. Benbrik. "Bosonic decays of charged Higgs bosons in a 2HDM type-1". *Eur. Phys. J. C. 77*, 2017.

[9] Prasenjit Sanyal. "Limits on the Charged Higgs Parameters in the Two Higgs Doublet Model using CMS $\sqrt{s}$ = 13 TeV Results". *Eur. Phys. J. C79*, 2019.

[10] DELPHI Collaboration. "Search for charged Higgs bosons at LEP in general two Higgs doublet models". *Eur. Phys. J. C. 34*, 2004.

[11] OPAL Collaboration. "Search for charged Higgs bosons in e^+e^- collisions at $\sqrt{s}$ = 189-209 GeV". *Eur. Phys. J. C 72*, 2012.

[12] CDF Collaboration. "Search for a very light CP-odd Higgs boson in top quark decays from $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV". *Phys. Rev. Lett. 107*, 2011.

[13] CMS Collaboration. "Search for a light charged Higgs boson decaying to a W boson and a CP-odd Higgs boson in final states with eμμ or μμμ in proton-proton collisions at $\sqrt{s}$ = 13 TeV". *Phys. Rev. Lett. 123, 131802*, 2019.

[14] CMS Collaboration. "The CMS experiment at the CERN LHC". *JINST 3 (2008) S08004*.

[15] CMS Collaboration. "The CMS trigger system". *JINST 12 (2017) P01020*.

[16] Olivier Mattelaer et al. "Madgraph 5: Going Beyond". *Journal of High Energy Physics 128 (2011)*.

[17] Carlo Oleari. "The POWHEG-BOX". *Nucl. Phys. Proc. Suppl. 205-206:46-41 (2010)*.

[18] GEANT4 Collaboration. "GEANT4-a simulation toolkit". *Nucl. Instrum. Methods Phys. Res. Sect. A 506 (2003) 250*.

[19] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". *JINST 12 (2017) P10003*.

[20] CMS Collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}$ = 8TeV".

[21] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV". *JINST 13 (2018) P06015*.

[22] M. Cacciari et al. "The anti-k_T jet clustering algorithm". *JHEP 04 (2008) 063*.

[23] Emil Bols et al. "Jet Flavour Classification Using DeepJet". *JINST 15 (2020) P12012*.

[24] Zonghan Wu et al. "A Comprehensive Survey on Graph Neural Networks". *IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4-24 (2021)*.

[25] Loukas Gouskos Huilin Qu. "ParticleNet: Jet Tagging via Particle Clouds". *Phys. Rev. D 101, 056019 (2020)*.

[26] A. L. Read. "Presentation of search results: CL_s technique". *J. Phys. G 28 (2002) 2693*.