Note

Feng-Yang Hsieh

1 Signal

We consider a simple extension of the standard model (SM) [1], which includes a vector-like dark fermion $(\overline{\chi}, \chi)$ and a complex singlet scalar S. A signature of CP violation could come from the Higgs-to-Higgs decays, $h_3 \to h_2 h_1$, where $h_3/h_2/h_1$ are the heaviest scalar, second heaviest scalar, and the SM-like 125 GeV Higgs, respectively.

The signal process is the triple production of 125 GeV Higgs bosons via the gluon fusion:

$$gg \to h_3 \to h_2 h_1 \to h_1 h_1 h_1$$

The Higgs boson h_1 would further decay to the $b\bar{b}$ pair. We consider the banchmark point 1 (BP1), where $m_{h_3} = 450$ GeV, $m_{h_2} = 280$ GeV, $m_{h_1} = 125$ GeV. This process is generated at $\sqrt{s} = 13$ TeV. Following are the MadGraph scripts for generating signal samples:

```
import model cxSM_VLF_EFT
generate g g > h h h
output MG5/gghhh_bsm
launch MG5/gghhh_bsm

shower=Pythia8
detector=Delphes
analysis=OFF
madspin=ON
done

set param_card mh1 125
set param_card mh2 280
set param_card mh3 420
set param_card theta12 0.73
```

```
set param_card theta13 1.67079632679
set param_card theta23 -0.73
set param_card vs 200
set param_card delta2 0
set param_card Rdelta3 0
set param_card Idelta3 -3.5
set param_card b2 0
set param_card Rc1 0
set param_card Ic1 0
set param_card Rc2 0
set param_card Ic2 0
set param_card Rd3 0
set param_card Id3 0
set param_card msq -5033.406281907266
set param_card lam 0.13850082540690806
set param_card Rdelta1 -47.561525227572744
set param_card Idelta1 853.05384671134
set param_card Rb1 -70476.6380004269
set param_card Ib1 -30486.140015405872
set param_card Rd1 -2.562109886826132
set param_card Id1 2.257859679994403
set param_card d2 6.340799300844676
set param_card gh1ggr -0.00005478952893059635
set param_card gh1gagar -0.00003270447254456052
set param_card gh1Zgar -0.00005871986046374793
set param_card gh2ggr -1.4279972541632635e-7
set param_card gh2gagar -8.237715486808595e-8
set param_card gh2Zgar -1.3984990232267825e-7
set param_card gh3ggr -6.031835872118092e-6
set param_card gh3gagar -1.1377279177203616e-6
set param_card gh3Zgar -2.2999597941282603e-6

set param_card decay 102 auto
set param_card decay 103 auto

set run_card nevents 100000
```

```
set run_card ebeam1 6500.0
set run_card ebeam2 6500.0

set run_card ptb 24
set run_card etab 2.6

set spinmode none
decay h > b b~

done
```

2 SPANet pairing

We employ the novel neural network structure Spa-Net [2, 3, 4] to identify the correct pairings among the jets in the final states.

2.1 Training dataset preparation

Preselection: ≥ 6 jets with transverse momentum $p_{\mathrm{T}} \geq 25$ GeV in range $|\eta| < 2.5$.

The input features for the Spa-Net are a list of jets, each represented by its 4-component vector $(p_{\mathrm{T}}, \eta, \phi, m)$ as well as a boolean b-tag. We only keep each event's 15 highest p_{T} jets. For each event, we define the correct jet assignments by matching the jets to the simulated truth quarks within an angular distance of $\Delta R < 0.4$. If a simulated truth quark is matched to more than one jet, such an event will be dropped. Furthermore, some simulated truth quarks may not be matched to any jet, in which case the event will not be used in training either.

After the selection and matching, we could obtain the following results from 1M events:

- Total sample size: 522,899
- 1h sample size: 184,769
- 2h sample size: 161,476
- 3h sample size: 94,464

Here, the 1h sample is where we could define the correct jet assignments for 1 Higgs boson.

2.2 Training results

- Training sample:
  - Total sample size: 470,609
  - 1h sample size: 166,490
  - 2h sample size: 145,309
  - 3h sample size: 84,913
  - 5% used on validation
- Testing sample:
  - Total sample size: 52,290
  - 1h sample size: 18,279
  - 2h sample size: 16,167
  - 3h sample size: 9,551

Some useful definitions for evaluating jet assignment performance:

- Event Efficiency

$$\epsilon^{\text{event}} \equiv \frac{\text{number of events with and all Higgs are correctly identified}}{\text{number of events}} \tag{1}$$

- Higgs Efficiency

$$\epsilon^{\text{h}} \equiv \frac{\text{number of correctly identified Higgs}}{\text{number of identifiable Higgs}} \tag{2}$$

The training results are shown in Table 1.

Table 1: SPA-NET pairing efficiencies on 3h events.

N_{Jet}	Event Fraction	Event Efficiency	Higgs Efficiency
$= 6$	0.077	0.532	0.650
$= 7$	0.057	0.345	0.536
≥ 8	0.052	0.237	0.452
Total	0.186	0.375	0.548

3 χ^2 pairing

χ^2 method considers all possible combinations of final jets and selects the configuration that minimizes the mass difference between Higgs candidates and SM Higgs, i.e., minimizes this:

$$\chi^2 = [m(j_1 j_2) - m_h]^2 + [m(j_3 j_4) - m_h]^2 + [m(j_5 j_6) - m_h]^2 \tag{3}$$

where $m(j_i j_j)$ is the invariant mass of jet i, j and $m_h = 125$ GeV.

Table 2 is the performance of the χ^2 method.

Table 2: χ^2 pairing efficiencies on 3h events.

N_{Jet}	Event Fraction	Event Efficiency	Higgs Efficiency
$= 6$	0.077	0.403	0.450
$= 7$	0.057	0.158	0.281
≥ 8	0.052	0.000	0.077
Total	0.186	0.215	0.294

4 Estimate cross-section of background process

Besides the 6 b background, we need to consider the backgrounds that come from the mis-tagging of light jets or charm-jets to b-jets. We assume that the probability of a charm-jet being misidentified as b-jet is $\mathcal{P}_{c\to b} = 0.1$ and that of light jets is $\mathcal{P}_{j\to b} = 0.01$. The b-tagging efficiency is assumed to be $\mathcal{P}_{b\to b} = 0.7$.

Table 3 shows the cross-section computed from `MadGraph` and the cross-section times the mis-tagging probabilities $\mathcal{P}_{c\to b}$ and $\mathcal{P}_{j\to b}$. Table 4 shows the same results with kinetic cuts. We require the transverse momentum p_{T} of each jet greater than 24 GeV and in the range $|\eta| < 2.6$ at the `MadGraph` level. The $6b$ process contributes much more than the processes containing charm jets and light jets.

5 Compute pairing efficiency

To understand the pairing performance with different pairing methods, we compute how many events where 1h/2h/3h bosons are reconstructed correctly.

The pairing performance of SPA-NET are shown in Table 5. Table 6 is the performance of the χ^2 method. For both cases, we found the number of events where only two Higgs are

Table 3: The cross-sections of $6b$ and mis-tagging background processes. The cross-sections are computed from the MadGraph at $\sqrt{s} = 13$ TeV.

process	σ (pb)	$\sigma \times \mathcal{P}$(tagging efficiency) (pb)
$(b\bar{b})(b\bar{b})(b\bar{b})$	2.53×10^3	2.97×10^2
$(b\bar{b})(b\bar{b})(c\bar{c})$	2.72×10^2	6.54×10^{-1}
$(b\bar{b})(c\bar{c})(c\bar{c})$	3.73×10^1	1.83×10^{-3}
$(b\bar{b})(b\bar{b})(jj)$	7.44×10^4	1.79

Table 4: The cross-sections of $6b$ and mis-tagging background processes. The cross-sections are computed from the MadGraph at $\sqrt{s} = 13$ TeV. We require the transverse momentum p_{T} of each jets greater than 24 GeV in range $|\eta| < 2.6$.

process	σ (fb)	$\sigma \times \mathcal{P}$(tagging efficiency) (fb)
$(b\bar{b})(b\bar{b})(b\bar{b})$	9.63×10^2	113.35
$(b\bar{b})(b\bar{b})(c\bar{c})$	1.67×10^3	4.02
$(b\bar{b})(c\bar{c})(c\bar{c})$	1.06×10^3	5.19×10^{-2}
$(b\bar{b})(b\bar{b})(jj)$	4.16×10^5	9.98
$(b\bar{b})(jj)(jj)$	1.50×10^7	7.73×10^{-2}

paired correctly is very small, which means if we can pair two Higgs bosons correctly, then we have a high chance to pair the final Higgs correctly.

Note that Higgs Efficiencies of SPA-NET are inconsistent with Table 1. This issue needs more checking.

Table 5: SPA-NET pairing efficiencies on different categories.

	Correctly reconstructed Higgs				
N_{Jet}	3h	2h	1h	0h	Higgs Efficiency
$= 6$	0.532	0.000	0.119	0.348	0.572
$= 7$	0.345	0.021	0.166	0.469	0.414
≥ 8	0.237	0.022	0.186	0.554	0.314
Total	0.375	0.014	0.156	0.455	0.436

Table 6: χ^2 pairing efficiencies on different categories.

	Correctly reconstructed Higgs				
N_{Jet}	3h	2h	1h	0h	Higgs Efficiency
$= 6$	0.403	0.000	0.143	0.455	0.450
$= 7$	0.158	0.070	0.228	0.544	0.281
≥ 8	0.000	0.000	0.231	0.769	0.077
Total	0.215	0.022	0.194	0.570	0.294

6 SPANet pairing and classification

We train a SPA-NET to identify the correct pairings and perform the signal/background classification at the same time.

6.1 Training dataset

The selection and matching process for the jet pairing is the same as Section 2.1. We prepare the signal and background samples of the same size for classification.

For the jet assignment part,

- Training sample:
  - Total sample size: 1,800,000
  - 1h sample size: 318,053
  - 2h sample size: 277,876
  - 3h sample size: 162,444
  - 5% used on validation
- Testing sample:
  - Total sample size: 200,000
  - 1h sample size: 35,372
  - 2h sample size: 30,853
  - 3h sample size: 18,004

For event classification,

- Training sample:

  - Total sample size: 1,800,000
  - Signal sample size: 900,000
  - Background sample size: 900,000
  - 5% used on validation

- Testing sample:
  - Total sample size: 200,000
  - Signal sample size: 100,000
  - Background sample size: 100,000

This training takes around 10 hours on our server.

6.2 Training results

The training results are presented in Table 7. This result is better than Table 5 since we use larger training datasets.

Table 7: SPA-NET training results on the tri-Higgs samples. SPA-NET is trained on jet pairing and event classification tasks at the same time.

	Correctly reconstructed Higgs				
N_{Jet}	3h	2h	1h	0h	Higgs Efficiency
$= 6$	0.656	0.000	0.082	0.262	0.684
$= 7$	0.436	0.017	0.168	0.379	0.504
≥ 8	0.341	0.018	0.173	0.468	0.411
Total	0.478	0.012	0.142	0.368	0.533

Table 8 presents the classification training results. We use the accuracy (ACC) and the area under the Receiver Operating Characteristic (ROC) curve (AUC) as two metrics.

Table 8: The SPA-NET classification training results with tri-Higgs sample.

	ACC	AUC
SPA-NET	0.822	0.900

6.3 3h training dataset

We only consider 3h events in pairing tasks in this subsection. We prepare the signal and background samples of the same size for classification.

For the jet assignment part,

- Training sample:
  - Total sample size: 1,800,000
  - 1h sample size: 0
  - 2h sample size: 0
  - 3h sample size: 900,000
  - 5% used on validation
- Testing sample:
  - Total sample size: 200,000
  - 1h sample size: 0
  - 2h sample size: 0
  - 3h sample size: 100,000

For event classification,

- Training sample:
  - Total sample size: 1,800,000
  - Signal sample size: 900,000
  - Background sample size: 900,000
  - 5% used on validation
- Testing sample:
  - Total sample size: 200,000
  - Signal sample size: 100,000
  - Background sample size: 100,000

The training results are presented in Table 9. This result is similar to Table 7.

However, some issues should be resolved when we try to use all 3h events for combining training. The loss values are not reasonable.

Table 9: SPA-NET training results on the tri-Higgs samples, where we only consider 3h events. SPA-NET is trained on jet pairing and event classification tasks at the same time.

	Correctly reconstructed Higgs				
N_{Jet}	3h	2h	1h	0h	Higgs Efficiency
$= 6$	0.680	0.000	0.084	0.236	0.708
$= 7$	0.477	0.014	0.150	0.359	0.536
≥ 8	0.311	0.027	0.184	0.477	0.391
Total	0.491	0.014	0.139	0.356	0.547

7 Review the pairing methods

In Refs. [5, 6]: We select the 6 b-tagged jets with the highest transverse momentum. The requirements for the transverse momentum and pseudo-rapidity are applied. We subsequently make use of the observable:

$$\chi^{2,(6)} = \sum_{qr \in J} (m_{qr} - m_h)^2 \tag{4}$$

where $J = \{j_1 j_2, j_3 j_4, j_5 j_6\}$ is the set of all possible 15 pairings of 6 b-tagged jets. Out of all the possible combinations we pick the one with the smallest value $\chi^{2,(6)}_{\text{min}}$. The pairing of b-jets defining $\chi^{2,(6)}_{\text{min}}$ is our best candidate for the reconstruction of the three Higgs bosons, h. No pairing efficiency is provided.

In Ref. [7]: We select the 6 b-tagged jets with the highest transverse momentum and form pairs in different combinations, with the aim of first reconstructing individual SM-like Higgs bosons, h_1, and subsequently the two scalars h_2 and h_3. To this end, we introduce two observables:

$$\chi^{2,(4)} = \sum_{qr \in I} (m_{qr} - m_h)^2 \tag{5}$$

$$\chi^{2,(6)} = \sum_{qr \in J} (m_{qr} - m_h)^2 \tag{6}$$

where we have defined the sets $I = \{i_1 i_2, i_3 i_4\}$ and $J = \{j_1 j_2, j_3 j_4, j_5 j_6\}$, constructed from different pairings of 4 and 6 b-tagged jets, respectively, and where m_{qr} denotes the invariant mass of the respective pairing, qr. Note that the set I that defines $\chi^{2,(4)}_{\text{min}}$ should be a subset of the arrangement J.

We select the combinations of b-tagged jets entering in I and J based on the minimization of the sum

$$\chi^{2,(4)} + \chi^{2,(6)} \tag{7}$$

We then "identify" candidates for the scalars h_2 and h_3 with the pairing configurations I_{min} and J_{min} which minimise $\chi^{2,(4)}$ and $\chi^{2,(6)}$ respectively. Note that this procedure does not guarantee that I_{min} indeed reconstructs to h_2; in fact, we found this to be the case in about 40% on average for all benchmark samples, being slightly higher than a "blind guess" that would lead to a probability of 1/3.

References

[1] T.-K. Chen, C.-W. Chiang, and I. Low, "Simple model of dark matter and CP violation," *Phys. Rev. D*, vol. 105, no. 7, p. 075025, 2022.

[2] 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," *Phys. Rev. D*, vol. 105, p. 112008, Jun 2022.

[3] M. J. Fenton, A. Shmakov, H. Okawa, Y. Li, K.-Y. Hsiao, S.-C. Hsu, D. Whiteson, and P. Baldi, "Extended Symmetry Preserving Attention Networks for LHC Analysis," 9 2023.

[4] 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," *SciPost Phys.*, vol. 12, p. 178, 2022.

[5] A. Papaefstathiou, G. Tetlalmatzi-Xolocotzi, and M. Zaro, "Triple Higgs boson production to six b-jets at a 100 TeV proton collider," *Eur. Phys. J. C*, vol. 79, no. 11, p. 947, 2019.

[6] A. Papaefstathiou and G. Tetlalmatzi-Xolocotzi, "Multi-Higgs Boson Production with Anomalous Interactions at Current and Future Proton Colliders," 12 2023.

[7] A. Papaefstathiou, T. Robens, and G. Tetlalmatzi-Xolocotzi, "Triple Higgs Boson Production at the Large Hadron Collider with Two Real Singlet Scalars," *JHEP*, vol. 05, p. 193, 2021.