

DRAFT-001
January 20, 2015

Feasibility Study for the Final State $hh \rightarrow (b\bar{b})(b\bar{b})$

This is a feasibility study for the final state $hh \rightarrow (b\bar{b})(b\bar{b})$ using reconstruction techniques for both resolved and boosted topologies as well as multivariate methods.

1 Introduction

2 Samples

2.1 Background

All background samples are generated with the **SHERPA** event generator, version 2.1.1. For the explicit runcards used in the generation, see Appendix A.

The NNPDF 3.0 $n_f = 4$ LO set with strong coupling $\alpha_S = 0.118$ is used for all samples. At the generator level the following basic cuts are applied. Each final state particle must have $p_T \geq 20$ GeV, and be located within $|\eta| \leq 3.0$. Factorisation and renormalisation scales are set as $\mu_F = \mu_R = H_T/2$. Total cross-sections and details of the samples generated are shown in Table 1.

| Process | Filename | Generator | N_{evt} | σ_{tot} |
|-----------------------------|-----------------------|--------------|------------------|---------------------------|
| $pp \rightarrow HH$ | HH.sm.eft.100K.lhe | MG5_aMC@NLO | 100K | 1.729×10^{-2} pb |
| $pp \rightarrow bbbb$ | SHERPA_QCD_4b.hepmc | SHERPA 2.1.1 | 3M | 1.016×10^3 pb |
| $pp \rightarrow b\bar{b}jj$ | SHERPA_QCD_2b2j.hepmc | SHERPA 2.1.1 | 3M | 4.076×10^5 pb |
| $pp \rightarrow jjjj$ | Running | SHERPA 2.1.1 | 3M | ??? |
| $pp \rightarrow t\bar{t}$ | | | | |

Table 1: Summary of generated samples to date

Although suffering from a large theory uncertainty, we can compare the result of our background samples against those presented in the MG5_aMC@NLO paper [1]. Here for comparison we require in all samples four anti- k_T $R = 0.5$ jets with $p_T \geq 80$ GeV, and the leading jet must have $p_T \geq 100$ GeV. All jets must be within an acceptance of $|\eta| \leq 2.5$. In the case of the samples with b quarks in the final state, these requirements are extended to the appropriate number of b -jets. For example, in the $2b2j$ sample there must be at least two b -jets that pass the cuts outlined above.

In Table 2 this comparison is summarised for the $2b2j$ and $4b$ samples. Considering the large theory errors, agreement is reasonable in both instances.

| Process | σ aMC@NLO | σ Oxford (SHERPA) |
|--------------------|---------------------------|---------------------------|
| $b\bar{b}b\bar{b}$ | 5.050×10^{-1} pb | 3.473×10^{-1} pb |
| $b\bar{b}jj$ | 1.852×10^2 pb | 5.193×10^2 pb |

Table 2: Comparison of LO Oxford SHERPA cross-sections with those of the aMC@NLO paper. The aMC@NLO cross-sections come with a quoted 50% theory uncertainty.

3 Object and Event Selection

3.1 Resolved Topology

The resolved selection requires the presence of at least four b-tagged anti- k_T $R = 0.4$ jets with $p_T > 25$ GeV and $|\eta| > 2.5$. *Pre-cut* histograms are written out for various kinematic distributions for all events with at least four b-tagged jets but before applying further kinematic cuts. *Post-cut* histograms contain only events where the four jets also pass the above kinematic cuts.

The di-Higgs system is reconstructed by considering all possibilities of forming two pairs of jets with invariant masses m_{j1j2} and m_{j3j4} , respectively, and choosing the configuration that minimises their difference $|m_{j1j2} - m_{j3j4}|$. Only the four leading- p_T jets are considered here.

3.1.1 Resolved b -jet tagging efficiencies

In order to perform an initial validation of the b -tagging procedure used in our resolved analysis we consider first a simple flat b -jet tagging efficiency of $\epsilon_b = 0.8$ and a light jet mis-tag rate of $\epsilon_l = 0.01$. In the UCL diHiggs analysis [2] the $2b2j$ background was discarded by applying a naive selection factor of $\epsilon_l^2 \epsilon_b^2$ to a similar $2b2j$ cross-section as in Table 2 and finding the resulting cross-section to be considerably smaller than the equivalent selection factor applied to the $4b$ sample.

In practice we have found with these simple efficiencies, that the actual selection efficiencies differ considerably from the naive factors of $\epsilon_b^{n_b} \epsilon_l^{n_l}$. In Table 3 the results of this simple b -tagging test are demonstrated. Shown in the table are the total cross-sections before and after the application of the described b -tagging procedure, the resulting percentage of the cross-section passing the b-tag, and the naive selection percentage that would result from applying powers of $\epsilon_{b/l}$ to the total cross-section. The pass rate for samples with four b quarks in the final state is therefore considerably lower in prac-

tice than one would naively expect, and the pass rate for the $2b2j$ sample considerably higher.

| Process | σ_{tot} | After b-tag | % selected | % Naive selection |
|--------------|---------------------------|---------------------------|---------------|-------------------|
| HH | 1.729×10^{-2} pb | 1.565×10^{-3} pb | 9.050% | 40.96% |
| $bbbb$ | 1.016×10^3 pb | 93.46 pb | 9.199% | 40.96% |
| $b\bar{b}jj$ | 4.076×10^5 pb | 103.904 pb | 0.026% | 0.0064% |

Table 3: b -tagging analysis performed upon the samples described in Section 2. Cross-sections are given for the samples before and after the application of the b -tagging procedure whereby the four hardest jets in the event are used, and the flat selection efficiency and mis-tag rates applied. The percentage of cross-section surviving the selection is shown, along with the naive expectation ($\epsilon_b^{n_b} \epsilon_l^{n_l}$) for comparison.

These differences are due to the choices made in the selection procedure. In practice the four hardest jets in the event often do not correspond directly to the four parton level final state particles. In Table 4 the number of true b -jets (that is, prior to the application of selection efficiencies and fake rates) present in the hardest four jets per event is shown. Indeed, for the samples with four b -quarks in the parton level final state, the most common scenario has only three b -jets in the hardest four jets, significantly affecting the selection efficiency.

A more accurate estimation for the passing cross-section may be obtained by multiplying the various fractions of b -jets by the corresponding efficiencies, for example the 3 b -jet rate by $\epsilon_b^3 \epsilon_l^1$. Summing the various scenarios we then have a better estimate for the percentage of cross-section passing the selection, shown in the final column of Table 4. This agrees well with the observed rates shown in Table 3.

| Process | 0 b -jets | 1 b -jet | 2 b -jets | 3 b -jets | 4 b -jets | % selected |
|--------------|-------------|------------|-------------|-------------|-------------|---------------|
| $H\bar{H}$ | 0.049% | 2.448% | 23.49% | 52.58% | 21.43% | 9.050% |
| $bbbb$ | 0.429% | 5.477% | 25.53% | 46.67% | 21.89% | 9.208% |
| $b\bar{b}jj$ | 6.213% | 40.56% | 51.93% | 1.227% | 0.072% | 0.039% |

Table 4: Breakdown of the number of true b -jets in the hardest four clustered jets for various samples. The final column shows the predicted pass rate for the b -tagging given the distribution of events in the preceding columns.

3.1.2 b -jet tagging efficiencies with UCL-style selection

In the UCL analysis their jet selection does not proceed as described in the previous section, presumably as requiring the four hardest jets in the event to be b -tagged cut too aggressively into the signal process. Rather, the selection proceeds by taking all anti- k_T $R = 0.5$ jets with $p_T > 40$ GeV and running the b -tagging algorithm on them. The four hardest successfully b -tagged jets are then selected for further analysis. With such a method implemented for our samples and with our previous selection efficiencies and fake rates, Table 5 summarises the results.

| Process | σ_{tot} (4 Jets $p_T > 40$ GeV) | After b -tag | % selected |
|--------------|---|----------------------------|------------|
| $H\bar{H}$ | 1.081×10^{-2} pb | 8.1147×10^{-4} pb | 7.467% |
| $bbbb$ | 7.239×10^1 pb | 3.961 pb | 5.472% |
| $b\bar{b}jj$ | 2.137×10^4 pb | 4.779 pb | 0.0224% |

Table 5: Summary of the b -tagging procedure of the UCL analysis. Presented as in Table 3. In this case the total cross-sections are presented after a cut requiring four anti- k_T $R=0.5$ jets with $p_T > 40$ GeV in the final state. The percentage selected column is given for the b -tagging stage *only* for comparison.

Once again, it is clear that more of the $b\bar{b}jj$ sample is making it past the b -tagging selection than originally expected. It seems unlikely that this background can be neglected in the resolved analysis based upon a simple factor of $\epsilon_b^2 \epsilon_l^2$ as used by the UCL group.

3.2 Boosted Topology

In the boosted topology, the decay products of each Higgs boson are merged into a single large- R jet with a two-prong substructure. These Higgs jets are reconstructed as anti- k_T $R = 1.0$ jets for which two substructure variables, the first k_T splitting scale d_{12} and the 2-subjettiness ratio τ_{21} , are calculated. *Pre-cut* histograms are written out after requiring the presence of at least two large- R in the event but without applying any further kinematic cuts. *Post-cut* histograms are filled after the following additional cuts: Both jets are required to have $p_T > 100$ GeV and $|\eta| > 2.5$. Moreover, each jet is required to have at least two b -tagged anti- k_T $R = 0.3$ jets matched to it via *ghost association*: To this end, the constituents of a given large- R jet is reclustered using the jet algorithm and radius parameter of the original jet and so-called *ghost jets*, each corresponding to a small- R jet in the event, are added to the input for the

cluster sequence. The 4-vector of a ghost jet is obtained from a given small- R jet by setting its transverse momentum and mass to negligibly small values but retaining its direction in η and ϕ . A small- R jet is considered matched to the large- R jet if its ghost is found among the constituents of the reclustered jet. Ghost association allows to unambiguously match small- R to large- R jets, even in dense environments where a simple matching based on the distance ΔR between jets may lead to multiple matchings.

3.3 Boosted Topology with Variable- R Jets

The same selection as in Subsection 3.2 is applied but this time using Variable- R jets with the following parameters: $\rho = 500$ GeV, $R_{max} = 1.0$, $R_{min} = 0.2$.

4 Multivariate Tools

5 Results

6 Conclusion

References

- [1] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao and T. Stelzer *et al.*, JHEP **1407** (2014) 079 [arXiv:1405.0301 [hep-ph]].
- [2] D. Wardrop, E. Jansen, N. Konstantinidis, B. Cooper, R. Falla and N. Norjoharuddeen, “Non-resonant Higgs pair production in the $b\bar{b}b\bar{b}$ final state at the LHC,” arXiv:1410.2794 [hep-ph].

A Event sample runcards

A.1 Background: QCD 4b

```
(run){
  EVENTS 3M;
  EVENT_GENERATION_MODE U;
  ME_SIGNAL_GENERATOR Comix;

  EVENT_OUTPUT HepMC_Short[SHERPA_QCD_4b];

  BEAM_1 2212; BEAM_ENERGY_1 7000;
  BEAM_2 2212; BEAM_ENERGY_2 7000;

  FRAGMENTATION=Off # disable hadronisation
  MI_HANDLER=None # disable multiple parton interactions

  SCF:=1; ### default scale factor
  SCALES VAR{SCF*H_T2/2};

  PDF_LIBRARY LHAPDFSherpa;
  PDF_SET NNPDF30_lo_as_0118_nf_4.LHgrid;
  PDF_SET_VERSION 0;

  MASSIVE[5] 1;
  MASS[5] 4.75;

}(run);
(processes){
  Process 93 93 -> 5 -5 5 -5;
  Order_EW 0;
  End process;

}(processes);
(selector){
  PT 5 20 7000
  PT -5 20 7000
  PseudoRapidity 5 -3.0 3.0
  PseudoRapidity -5 -3.0 3.0
}(selector);
```

A.2 Background: QCD 2b2j

```
(run){
  EVENTS 3M;
  EVENT_GENERATION_MODE U;
  ME_SIGNAL_GENERATOR Comix;

  EVENT_OUTPUT HepMC_Short[SHERPA_QCD_2b2j];

  BEAM_1 2212; BEAM_ENERGY_1 7000;
  BEAM_2 2212; BEAM_ENERGY_2 7000;

  FRAGMENTATION=Off # disable hadronisation
  MI_HANDLER=None # disable multiple parton interactions

  SCF:=1; ### default scale factor
  SCALES VAR{SCF*H_T2/2};

  PDF_LIBRARY LHAPDFSherpa;
  PDF_SET NNPDF30_lo_as_0118_nf_4.LHgrid;
  PDF_SET_VERSION 0;

  MASSIVE[5] 1;
  MASS[5] 4.75;

}(run);
(processes){
  Process 93 93 -> 93 93 5 -5;
  Order_EW 0;
  End process;
}(processes);
(selector){
  PT 5 20 7000
  PT -5 20 7000
  PseudoRapidity 5 -3 3
  PseudoRapidity 5 -3 3

  PT 93 20 7000
  PseudoRapidity 93 -3 3
```



```
}(selector);
```

A.3 Background: QCD 4j

```
(run){
  EVENTS 3M;
  EVENT_GENERATION_MODE U;
  ME_SIGNAL_GENERATOR Comix;

  EVENT_OUTPUT HepMC_Short[SHERPA_QCD_4j];

  BEAM_1 2212; BEAM_ENERGY_1 7000;
  BEAM_2 2212; BEAM_ENERGY_2 7000;

  FRAGMENTATION=Off # disable hadronisation
  MI_HANDLER=None # disable multiple parton interactions

  SCF:=1; ### default scale factor
  SCALES VAR{SCF*H_T2/2};

  PDF_LIBRARY LHAPDFSherpa;
  PDF_SET NNP30_lo_as_0118_nf_4.LHgrid;
  PDF_SET_VERSION 0;

  MASSIVE[5] 1;
  MASS[5] 4.75;

}(run);
(processes){
  Process 93 93 -> 93 93 93 93;
  Order_EW 0;
  End process;
}(processes);
(selector){
  PT 93 20 7000
  PseudoRapidity 93 -3 3
}(selector);
```