JHEP

Published for SISSA by Springer

Received: November 5, 2013
Accepted: February 7, 2014
Published: February 28, 2014

Buckets of Higgs and tops

Matthew R. Buckley,[a,b] **Tilman Plehn,**[c] **Torben Schell**[c] **and Michihisa Takeuchi**[d]

[a] *Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL, U.S.A.*

[b] *Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, U.S.A.*

[c] *Institut für Theoretische Physik, Universität Heidelberg, Heidelberg, Germany*

[d] *Theoretical Particle Physics and Cosmology Group, Department of Physics, King's College London, London WC2R 2LS, U.K.*

E-mail: mbuckley@physics.rutgers.edu, plehn@uni-heidelberg.de, schell@thphys.uni-heidelberg.de, michihisa.takeuchi@kcl.ac.uk

Abstract: We show that associated production of a Higgs with a top pair can be observed in purely hadronic decays. Reconstructing the top quarks in the form of jet buckets allows us to control QCD backgrounds as well as signal combinatorics. The background can be measured from side bands in the reconstructed Higgs mass. We back up our claims with a detailed study of the QCD event simulation, both for the signal and for the backgrounds.

Keywords: Monte Carlo Simulations, Hadronic Colliders

ArXiv ePrint: 1310.6034

doi:10.1007/JHEP02(2014)130

JHEP02(2014)130

Contents

JHEP02(2014)130

1 Introduction

Measuring the couplings of the recently discovered Higgs boson [1–5] to the Standard Model fermions is a critical part of the investigation of the electroweak symmetry breaking mechanism at the LHC (for an example analyses see [6, 7]). The Standard Model coupling to the top quark is expected to be of order unity, making it a prime target for studying the effects of many different new physics models in and beyond the Higgs sector [8]. Together with the Higgs self coupling it dominates the extrapolation of weak-scale Higgs physics to more fundamental energy scales (for a pedagogical introduction see [9]). Measuring this parameter will uniquely probe extensions of the Higgs sector at the weak scale [6, 7] as well as at high scales.

However, with a production cross section of only $\mathcal{O}(500\text{ fb})$ at the 13 TeV LHC, measurements based on the $t\bar{t}H$ channel are extremely difficult. Search strategies in the fully leptonic and semi-leptonic decay channels for the top have been suggested in combination with Higgs decays to $b\bar{b}$ (for an update based on jet substructure analyses see [10]), $\tau\tau$ [11–14], and W^+W^- [15–17]. These are challenging through a combination of combinatoric backgrounds, missing control regions, large QCD uncertainties on the backgrounds, and low rate. Typical luminosities required for a 5σ signal might well be in the $100\,\text{fb}^{-1}$ range for 13 TeV collider energy, with a signal-to-background ratio well below 1:1.

In this paper we provide a feasibility study for the fully hadronic channel of $t\bar{t}H$ production, i.e. a final state consisting of four b-jets plus up to four un-tagged jets. This channel has not been studied yet. In fact, there exist only a few analyses targeting Higgs or new physics searches in purely hadronic channels without missing energy or leptons, most notably some recent top pair resonance [18–22] and sgluon searches [23]. However, the fully hadronic decay channel of $t\bar{t}H$ has two possible advantages over the leptonic decay modes. First, hadronic decays of both the tops and the Higgs have the highest branching

ratios of any decay mode. Second, without neutrinos and their missing momenta, a full reconstruction of the $t\bar{t}H$ final state is possible, which allows for discrimination of signal and background and provides the best testing ground in the presence of possible experimental anomalies. In addition, this eight-parton final state has the highest jet multiplicity of any widely-considered Standard Model process at the LHC. Demonstrating our ability to understand and use such events is an important benchmark in our study of Standard Model physics at the LHC.

We separate the signal events from the large QCD background via a two-step process. For four b-jet events, we first apply global acceptance cuts, giving us an enriched sample of signal events. We then reconstruct the tops using the "bucket algorithm" [22], which closes the gap between kinematic top reconstruction at threshold and proper top taggers [24–27] by targeting slightly boosted top quarks, with

$$p_{T,t} \sim 100 - 300 \text{ GeV} . \tag{1.1}$$

The algorithm identifies the two top quarks in the event by permuting over jet assignments to three buckets, minimizing a distance metric on two of those buckets between the invariant mass of the jets in the buckets and the mass of the top. The remaining event contains two b-jets, which allow us to reconstruct the Higgs decay with a probability above 60%.

It should be noted that the results in this paper deliberately only rely on a simple cut-and-count method. It allows us to identify many opportunities for data-driven side band calibration of the backgrounds, which is crucial to such high-multiplicity searches. As the top-reconstruction technique provides a good approximation of the momenta of all the particles in the original event, more sophisticated techniques can be used to improve rejection of background and signal selection.

In the next section, we study the major backgrounds to the $t\bar{t}H$ search, including some global background rejection cuts. The bucket algorithm is introduced in section 3, where we also give a detailed estimate of the analysis performance. A detailed discussion of the QCD backgrounds and their simulation are included in appendix A. Finally, in appendix B we show the metrics for the top reconstruction.

2 Multi-jet backgrounds and global cuts

Our analysis aims to extract the fully hadronic final state of $t\bar{t}H$ production with the decay $H \to b\bar{b}$. For a Higgs mass of 125 GeV we assume the Standard Model branching ratio to the $b\bar{b}$ final state of 57.7% [28, 29]. We will require four b-tagged jets in the final state, no leptons, and at least two un-tagged hard jets. We assume jet-based triggers for hard multi-jet events, similar to all-hadronic $t\bar{t}$ searches [18–21]. The main four-b backgrounds, ordered by relevance, are

$$pp \to bb\bar{b}\bar{b} \qquad pp \to t\bar{t}b\bar{b} \qquad pp \to tt\bar{t}\bar{t} . \tag{2.1}$$

The corresponding fake-b channels are strongly suppressed if the experiments reach a 70% b-tagging efficiency with 1% mis-tagging probability for light-flavor and gluon jets. These

JHEP02(2014)130

tagging and mis-tagging rates are reasonable approximations to the experimental results in events with one or two b-jets; additional information from experiments would be necessary to estimate the rates for events with four b-jets [30, 31]. Our analysis is based on a simple calorimeter simulation with granularity of 0.1×0.1 in (η, ϕ). We sum the four momentum of all particles in the Monte Carlo data in each cell and rescale the resulting three-momentum such as to make the cells massless. The jets are reconstructed from the cells using the Cambridge/Aachen algorithm (CA) [48, 49], with $R = 0.5$. If a parton-level b-quark lies inside the radius of a reconstructed jet then a b-tag is randomly assigned 70% of the time. Light-flavor jets are assigned as a b-jet with a 1% mis-tag probability

The backgrounds shown in eq. (2.1) stand out from the large number of QCD background processes. For example, with an estimated cross section of 175 pb combined with a total 0.1% mis-tagging rate (due to combinatorial effects), the rate of the mis-tagged $b\bar{b}$+multi-jet background should have similar distributions to the leading $bb\bar{b}\bar{b}$ background and enhance the combined rate at the $\mathcal{O}(10\%)$ level. This is below the quoted uncertainties in the simulations of the primary background. For the same reason we can ignore the pure QCD events with four mis-tags. We did not treat charm quarks specially; the $bbccjj$ cross section is 20 pb; combined with a 10% mis-tag rate for charm quarks [30, 31] and the 1% mistag rate for other light jets, this channel also contributes at a rate below the error of the leading background. The $t\bar{t}jj$ background has been shown to be sub-leading as compared to the $t\bar{t}b\bar{b}$ background in the first boosted $t\bar{t}H$ analysis [10].

For our event simulation we rely on Alpgen [32] and Madgraph [33], both with a Pythia parton shower [34, 35], as well as on Sherpa [36]. For the $t\bar{t}H$ signal our main event sample includes zero and one hard extra jet merged in the Ckkw scheme [37] in Sherpa. In appendix A we compare the Sherpa results with the Madgraph simulation of $t\bar{t}H$ plus up to one hard jet merged in the Mlm scheme [38]. We confirm that the sensitivity to simulation and QCD issues is minimal. Similarly, for the $t\bar{t}b\bar{b}$ background, our main sample of events is produced by Sherpa and includes up to one hard QCD jet. A comparison with Alpgen samples in appendix A again shows negligible dependence on the simulation techniques. We normalize the merged event samples to the NLO results of 504 fb for the signal [29, 39, 40] and 1037 fb after generator cuts for the $t\bar{t}b\bar{b}$ background [41–43]. The $tt\bar{t}\bar{t}$ background from Alpgen is small compared to the primary $t\bar{t}b\bar{b}$ and $bb\bar{b}\bar{b}$ backgrounds, with a cross section of at maximum 5% of the signal. As a result, it does not require an extensive study of the theoretical and simulation uncertainties, and will not be considered in detail.

The critical background for the hadronic $t\bar{t}H$ signal with $H \to b\bar{b}$ decays is the QCD process $bb\bar{b}\bar{b}$+jets. Before any selection cuts, its total rate completely overwhelms the signal, with a cross section of 400 pb estimated by Alpgen after pre-selection cuts. However, as any QCD process it is dominated by soft b and un-tagged jets with an additional enhancement from the gluon splitting $g \to b\bar{b}$. To extract our signal we will require four hard, well separated b-tagged jets. We simulate these background events both in Alpgen and Sherpa with a hard process of (at least) four b-jets.

As we will see in section 3, our bucket reconstruction of two tops and the Higgs will require at least two additional hard un-tagged jets. Therefore, our central background

JHEP02(2014)130

simulation is defined by the hard process $bb\bar{b}\bar{b}jj$ plus parton shower in Alpgen, which results in a cross section of 2128 fb after pre-selection cuts. To ensure that our analysis is stable with respect to QCD uncertainties, we also simulate the background with Sherpa as $bb\bar{b}\bar{b}$ plus zero and one matrix element jet merged ($bb\bar{b}\bar{b}$+0/1j). The computational expense of two jet merging is prohibitive here, and so is not included. However, to have a measure for the underlying theory uncertainties, we vary the renormalization and factorization scales in the Sherpa simulation by a factor 1/2 to 2 around the central scale, to ensure that our conclusions hold independent of these choices. We carefully compare our two background estimates in appendix A, providing detailed information on kinematic distributions and the different simulation tools and hard processes.[1] There, we test a couple of important assumptions underlying our analysis. First, we demonstrate that the bucket analysis allows only background events with at least two hard un-tagged jets in our signal region. For this region the Alpgen estimate of the $bb\bar{b}\bar{b}jj$ rate is the appropriate and conservative choice. In addition, we demonstrate that our analysis is not too sensitive to describing the second un-tagged jet by either the hard matrix element (as in the Alpgen) or by the parton shower (as in Sherpa). Finally, the full merged $bb\bar{b}\bar{b}$+jets simulation would allow access to excellent control regions in the side band of the number of un-tagged jets, once these kinds of n_{jet} distributions are systematically evaluated by ATLAS and CMS [44–46].

Understanding the kinematics of the $bb\bar{b}\bar{b}$ background and reducing it using global kinematic cuts will be the central topic of this section. In the next section, we will find that several of these kinematic cuts can be replaced with the requirement of top reconstruction, which allows for an increased purity of signal over background. In the actual buckets analysis in section 3, we only quote the Alpgen results for the $bb\bar{b}\bar{b}$ background.

Compared to the signal rate, the raw QCD background (primarily $bb\bar{b}\bar{b}$+jets) is overwhelmingly large, and so we must apply selection cuts before the top-finding bucket algorithm can be employed. First, we require all events to have four b-tagged jets. These b-jets must be central and widely separated, to avoid the phase space regions with enhanced $g \to b\bar{b}$ splitting, with

$$p_{T,b} > 40 \text{ GeV}, \qquad |\eta_b| < 2.5, \qquad \Delta R_{bb} > 1.0 \qquad (4\times) \,. \tag{2.2}$$

In addition, we require at least two hard non-b jets with

$$p_{T,j} > 40 \text{ GeV}, \qquad |\eta_j| < 4.5, \qquad \Delta R_{jj} > 0.5 \qquad (2\times) \,. \tag{2.3}$$

These naive acceptance cuts are very inefficient, for example when compared to sub-jet methods. However, the aim of this paper is to show that the purely hadronic $t\bar{t}H$ process can be studied at the LHC, so we need to ensure that the pure QCD backgrounds can be reliably removed. Moreover, four individual b-tags cannot be treated as statistically independent unless we at least assume very widely separated b-jets. This necessitates the harsh cuts in this proof-of-concept analysis.

In the first line of table 1 we show the cross sections for the signal and two primary backgrounds at the 13 TeV LHC, after acceptance cuts. The $tt\bar{t}\bar{t}$ contribution is sub-percent

[1] We would like to thank the referees and the editor of ref. [22] for strongly supporting this kind of analysis and then allowing us to postpone it to this paper.

JHEP02(2014)130

	$t\bar{t}H$	$t\bar{t}b\bar{b}$	$bb\bar{b}\bar{b}jj$	S/B
After acceptance eqs.(2.2) and (2.3)	1.197	8.363	54.420	0.019
After global cuts eq. (2.5)	0.134	0.558	2.734	0.041
Mass window $m_{bb} = 90 - 130\,\text{GeV}$				
closest	0.096	0.299	1.577	0.051
hard	0.017	0.031	0.226	0.065
soft	0.060	0.173	0.893	0.056
min	0.071	0.246	1.143	0.051

Table 1. Cross section (in fb) of signal and background events after successive selection cuts. The $bb\bar{b}\bar{b}jj$ rate is based on the ALPGEN simulation. After the full set of cuts from eqs.(2.2), (2.3), and (2.5), we show several naive ways of selecting two b-jets to reconstruct the Higgs mass: the pair closest in invariant mass to the Higgs, the two hardest b-jets, the two softest b-jets, and the two b-jets with the minimum invariant mass. We assume a 70% b-tagging efficiency and neglect the small mis-tag backgrounds.

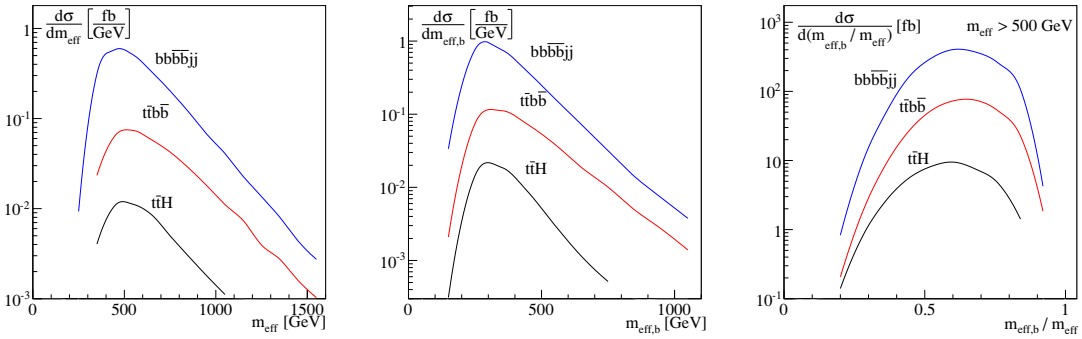

Figure 1. Signal and background distributions for the effective mass of the entire jet system, the four b-tagged jets, and their ratio. All jets fulfill eq. (2.2) and eq. (2.3). We require $m_{\text{eff}} > 500\,\text{GeV}$ in the selection cuts of eq. (2.5). For the $bb\bar{b}\bar{b}$ background we show the ALPGEN result with two additional hard jets plus parton shower.

level, and so not shown. At this stage the $bb\bar{b}\bar{b}$+jets cross section is still significantly larger than the signal, so an additional set of cuts is required. Once we introduce the top reconstruction technique, such cuts are not necessary, but it will be instructive to compare our later results to simple global cuts. We consider two global variables: the effective mass calculated by summing the scalar jet p_T over all jets, including those with b-tags, and its counterpart where the sum runs only over the four b-tagged jets,

$$m_{\text{eff}} = \sum_{\text{all jets}} p_T \,, \qquad m_{\text{eff},b} = \sum_{\text{four } b\text{-jets}} p_T \,. \tag{2.4}$$

Both of these observables will be sensitive to the kinematics of the multi-jet system. In figure 1 we show the distributions for both signal and backgrounds, normalized to the event rates after the cuts of eqs.(2.2) and (2.3). At this point, the signal-to-background

JHEP02(2014)130

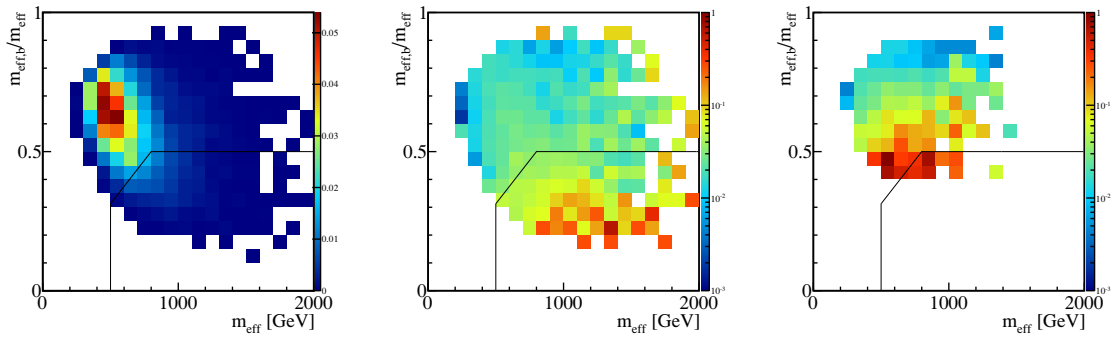

Figure 2. Two-dimensional distribution of m_{eff} vs. $m_{\text{eff},b}/m_{\text{eff}}$ for the $t\bar{t}H$ signal (left), and the ratio of the signal to the $bb\bar{b}\bar{b}jj$ ALPGEN background (center), and the ratio of signal to $bb\bar{b}\bar{b}+0/1j$ SHERPA simulation (right). The lines represent the cuts of eq. (2.5).

ratio is around 1:50. As we will discuss in appendix A, the fact that the m_{eff} distributions of the signal and the $bb\bar{b}\bar{b}$ backgrounds show a similar behavior is because our ALPGEN simulation requires two hard un-tagged jets. In other words, the background simulation shown in figure 1 anticipates the fact that we will only be interested in a reliable prediction of those background events which are kinematically similar to the signal. The right panel of figure 1 shows that after requiring $m_{\text{eff}} > 500\,\text{GeV}$ both m_{eff} and $m_{\text{eff},b}$ have similar shapes for signal and background.

It is more efficient to consider the 2-dimensional plane of the two effective mass variables defined in eq. (2.4). In figure 2, we plot these distributions for the signal and the ratios of signal-to-background against the primary ($bb\bar{b}\bar{b}$+jets) for both ALPGEN and SHERPA simulations. As can be seen, the SHERPA simulation has many fewer events in the high m_{eff} tail compared to the ALPGEN simulation, as expected due to the acceptance cut. The SHERPA simulation only generates up to one light-flavor or gluon jet from the hard matrix element, but the acceptance cuts require at least two hard un-tagged jets per event. As argued in appendix A we use the $bb\bar{b}\bar{b}$+jets sample from ALPGEN for a more conservative background estimate. In this proof-of-concept paper, we make the crude requirements that

$$m_{\text{eff}} > 500\ \text{GeV}\,, \qquad \frac{m_{\text{eff},b}}{m_{\text{eff}}} < 0.5\,, \qquad \frac{m_{\text{eff},b}}{m_{\text{eff}}} < \frac{m_{\text{eff}}}{1600\ \text{GeV}}\,. \tag{2.5}$$

This set of cuts brings the background rate to a manageable level, without a detailed analysis of the top and Higgs kinematics. For the specific $bb\bar{b}\bar{b}jj$ background modeling with ALPGEN we arrive at $S/B \sim 1/25$, as quoted in table 1.

From this point on, we are interested in identifying an excess of events that contain two b-jets which are clearly identified with the Higgs boson decay. This is complicated by the combinatorial background of picking the correct two b-jets out of the four in the event. First, we consider naive set of selection criteria for the two b-jets which have to lie in the Higgs mass window in table 1. We show that selecting the two b-jets closest in mass to $m_H = 125\,\text{GeV}$, the two b-jets with the softest p_T, the two hardest, and the two b-jets with the minimum invariant mass are all methods that fail to sufficiently increase signal over background.

JHEP02(2014)130

Clearly, a better solution to the reconstruction of the top and Higgs decay products and the combinatorics associated with this assignment is needed. We therefore turn to the bucket reconstruction [22] to rebuild the two top quarks in the event, using those events that have passed our initial selection criteria described by eqs.(2.2) and (2.3). This tags the two b-jets that come from the tops with a good degree of accuracy, identifying the Higgs decay products by exclusion. With this method of identifying the correct b-jets, the global cuts on m_{eff} variables do not improve the S/B ratio, and so we do not continue to apply the cuts of eq. (2.5). This simple algorithm is not meant to replace a full experimental likelihood analysis, but it shows that after simple kinematic cuts a purely hadronic $t\bar{t}H$ analysis can be a realistic possibility.

3 Top buckets

Following the arguments in the last section and the corresponding appendix A it should be possible to devise an analysis of the hadronic top and Higgs kinematics to reduce the backgrounds to a manageable level. Aside from the irreducible $t\bar{t}b\bar{b}$ background we need to extract the signal from the huge multi-jet $bb\bar{b}\bar{b}$ background. A more specific analysis of the multi-jet final state should be able to do better than the already promising global effective mass cuts in eq. (2.5). The key concern will be the signal efficiency, because of the limited $t\bar{t}H$ rate. For this reason, we choose the bucket reconstruction [22], which allows us to keep a larger fraction of signal events while removing significant parts of the background phase space identified by the global cuts analysis. The technical challenge is tracking the definition of the signal region and the corresponding background simulation.

After applying the jet-level selection cuts in the previous section, eqs.(2.2) and (2.3), we have a sample of events with four b-jets and additional extra jets. Of these four b-jets, two are presumed to come from the Higgs decay, and two from top decays. Without knowledge of the top decays, various Higgs reconstruction schemes could be tried. As discussed in the previous section, one could take the two b-jets with the highest or lowest p_T, the combination of b-jets with invariant mass that is closest to 125 GeV, the combination with the minimum invariant mass, or some other set based on simple jet kinematics. Taking the combination with the invariant mass closest to that of the Higgs in particular runs into a combinatorial problem: in both signal and background, one can often find two b-jets with invariant mass near that of the Higgs without the jets involved having originated with the Higgs. This shapes the background to look like signal. The multi-b combinatorics are the reason that the ATLAS $t\bar{t}H$ search in the early phase of LHC running was largely abandoned [47].

We can improve this situation if we find a better way to identify the b-jets that come from the Higgs. We approach this problem by first identifying the decay products of the tops, using the top bucket algorithm. The idea behind this algorithm is very simple and straightforward: we divide all jets in every event into three buckets. Two of the buckets correspond to the hadronic tops, while the third bucket consists of all jets in the event that cannot be associated with a top. In the original formulation of the algorithm [22] this last bucket was identified with initial state radiation (ISR). In the current analysis, this

JHEP02(2014)130

ISR bucket will contain two b-jets, which can — by exclusion — be identified as the decay products of the Higgs.

We start by seeding each of the two top buckets with a b-jet. We permute over all possible assignments of b-jets as top bucket seeds. We then cycle through every possible assignment of non-b-tagged jets to the three buckets, requiring at least one non-tagged jet in each of the top buckets. We use the distance metric

$$\Delta_{B_i} = |m_{B_i} - m_t| \qquad \text{with} \quad m_{B_i}^2 = \left(\sum_{k \in B_i} p_k \right)^2 , \tag{3.1}$$

where m_t is the top mass and the sum runs over all jets (including the b-jet) in the bucket B_i. We select the jet assignment that minimizes the distance measure $\Delta^2 = \omega \Delta_{B_1}^2 + \Delta_{B_2}^2$, where $\omega > 1$ is a factor chosen to stabilize the jet grouping. For this analysis, we choose $\omega = 100$, which essentially decouples the second bucket from the metric. Thus, bucket B_1 is the bucket with invariant mass closest to the top. After this construction, we have two buckets B_1 and B_2, with two or three jets, including the seed b-jet. Rarely, we find a bucket containing four or more jets, in which case we merge to three jets using the Cambridge/Aachen algorithm [48, 49].

To remove background events that do not contain real tops, we require the invariant masses of the two top buckets to lie in the window

$$155 \text{ GeV} < m_{B_{1,2}} < 200 \text{ GeV}. \tag{3.2}$$

Next, we require both B_1 and B_2 buckets to contain a hadronically decaying W boson candidate. We define a mass ratio cut, as in the HEPTopTagger [50],

$$\left| \frac{m_{k\ell}}{m_{B_i}} - \frac{m_W}{m_t} \right| < 0.15 \tag{3.3}$$

for at least one combination of the non-b-jets (denoted k and ℓ) in the bucket i. Buckets with only one b-tagged and one non-tagged jet by construction cannot satisfy eq. (3.3). We can therefore classify events in one of three categories:

- (t_w,t_w): both top buckets have W candidates as defined by eq. (3.3),
- (t_w,t_-) or (t_-,t_w): only the first or second top bucket has a W candidate,
- (t_-,t_-): neither top bucket has a W candidate.

The t_w or t_- status is ordered as (B_1, B_2), where again, B_1 is defined as the bucket closest in mass to the top. Buckets classified as t_w have to include at least three jets, while t_- buckets can include either three or two jets.

For buckets that fail the criteria of eq. (3.3), we can still attempt to reconstruct a top by replacing eq. (3.1) with an alternative distance metric,

$$\Delta_B^{bj} = \begin{cases} |m_B - 145 \text{ GeV}| & \text{if } m_B \leq 155 \text{ GeV} \\ \infty & \text{else} \end{cases} . \tag{3.4}$$

JHEP02(2014)130

	$t\bar{t}H$		$t\bar{t}b\bar{b}$	$b\bar{b}b\bar{b}jj$	S/B
After acceptance cuts eqs.(2.2) and (2.3)	1.197		8.363	54.420	0.019
2 tops tagged	0.894	(0.184)	5.882	29.356	0.025
$p_{T,t,1} > 100\,\text{GeV}$	0.709	(0.158)	4.868	20.838	0.028
$p_{T,t,1} > 200\,\text{GeV}$	0.289	(0.080)	2.189	5.194	0.039
$p_{T,t,1} > 300\,\text{GeV}$	0.089	(0.028)	0.724	0.917	0.054
Mass window $m_{bb} = 90 - 130\,\text{GeV}$					
2 tops tagged	0.259	(0.121)	0.859	5.424	0.041
$p_{T,t,1} > 100\,\text{GeV}$	0.208	(0.105)	0.688	3.600	0.048
$p_{T,t,1} > 200\,\text{GeV}$	0.091	(0.054)	0.265	0.679	0.096
$p_{T,t,1} > 300\,\text{GeV}$	0.028	(0.019)	0.072	0.082	0.182

Table 2. Cross sections (in fb) of events after the acceptance cuts of eqs.(2.2) and (2.3) and requiring two tops passing the bucket reconstruction. We also require one of the reconstructed tops to pass various p_T thresholds, with and without requiring the two remaining b-jets to have invariant mass inside the window 90-130 GeV. Number in parenthesis correspond to the events where the reconstructed Higgs lies within $\Delta R < 0.5$ of the true Higgs.

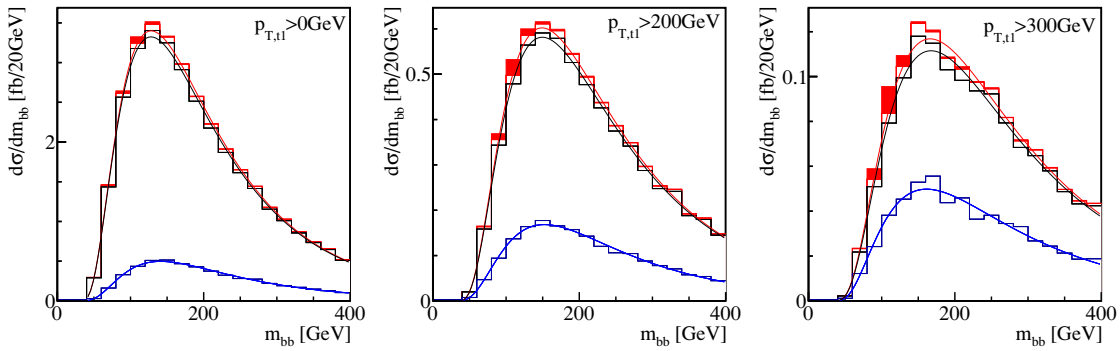

Figure 3. Stacked m_{bb} distribution, built from b-jets not used for top buckets, after top reconstruction and requiring the leading top $p_{T,t_1} > 0, 200$, and 300 GeV (from left). The two primary backgrounds are $t\bar{t}b\bar{b}$ (blue) and $b\bar{b}b\bar{b}$ (black). The signal distribution of reconstructed tops is shown in red. The subset of signal events where the reconstructed Higgs lies within $\Delta R < 0.5$ of the true Higgs are displayed as filled-in red regions.

We re-assign all b-tagged and un-tagged jets in the t_ bucket(s), combined with the jets in the ISR bucket to new buckets, irrespective of their original categorization. For this re-assignment we minimize $\sum_i \Delta^{bj}_{B_i}$. For a top candidate, we require at least one b/jet pair satisfying

$$75\ \text{GeV} < m_{bj} < 155\ \text{GeV}. \tag{3.5}$$

This revamped metric is intended to capture top events where the less energetic jet from W decay was lost in the detector.

The rate of $t\bar{t}H$ signal events passing the full top reconstruction, along with backgrounds, is listed in table 2. The accuracy of the reconstruction is addressed in appendix B.

JHEP02(2014)130

	$t\bar{t}H$		$t\bar{t}b\bar{b}$	$b\bar{b}\bar{b}\bar{b}jj$	S/B
After acceptance cuts eqs.(2.2) and (2.3)	1.197		8.363	54.420	0.019
2 tops tagged, $\Delta\eta$ cuts	0.587	(0.125)	2.762	10.654	0.044
$p_{T,t,1} > 100\,\text{GeV}$	0.485	(0.111)	2.392	8.364	0.045
$p_{T,t,1} > 200\,\text{GeV}$	0.207	(0.059)	1.153	2.541	0.056
$p_{T,t,1} > 300\,\text{GeV}$	0.064	(0.021)	0.405	0.507	0.071
Mass window $m_{bb} = 90 - 130\,\text{GeV}$					
2 tops tagged, $\Delta\eta$ cuts	0.170	(0.080)	0.376	1.864	0.076
$p_{T,t,1} > 100\,\text{GeV}$	0.144	(0.072)	0.317	1.396	0.084
$p_{T,t,1} > 200\,\text{GeV}$	0.066	(0.039)	0.129	0.338	0.142
$p_{T,t,1} > 300\,\text{GeV}$	0.021	(0.015)	0.034	0.043	0.276

Table 3. Cross sections (in fb) of events after successive selection cuts, as in table 2 but including the $\Delta\eta$ cuts of eq. (3.6).

With the strictest set of cuts, giving the highest purity of top reconstruction, 50% of the tops we reconstruct in the signal are identifiable with a parton-level top in the simulation. However, at this stage of our analysis, the reliable reconstruction of the top four momenta is not yet the main goal. What we need are the un-associated b-jets, which have to combine to the Higgs 4-momentum. Given the limited detector resolution, we require the invariant mass of these two b-jets to lie between 90 and 130 GeV. We find that, in the purest sample of reconstructed tops, 68% of the surviving signal events have the two remaining b-jets in the ISR bucket that are correctly assigned (that is, they correspond to parton-level b-quarks that originated in the decay of the Higgs). The signal-to-background ratio for hard top quarks can increase to around 1/10. Given the different background uncertainties this number is promising. However, it requires easily accessible side bands, in particular when we want to extract the top Yukawa coupling from such a rate measurement.

The obvious choice of side bands is, of course, the invariant mass of the two b-jets reconstructing the Higgs. While for the signal we expect a peak around 125 GeV, possibly shifted towards lower values by final state radiation escaping the momentum reconstruction, background (including combinatorics) can be well described by a log-normal distribution. In figure 3 we plot the invariant mass of the two un-associated b-jets of events that have survived the initial selection cuts of eqs.(2.2) and (2.3) and the top reconstruction algorithm. The relatively narrow mass peak of the signal is well separated from a broad feature of the backgrounds. Not all signal events can be associated with a parton-level Higgs momentum. In some cases the reason is missing final state radiation, in others it can be due to the b-jet combinatorics.

The situation significantly improves once we introduce a p_T^{min} cut on the reconstructed tops. Now the events under the Higgs peak are more and more dominated by the actual Higgs signal, and the signal peak separates cleanly from the broad maximum in the background shape. In table 2 we show the rate of signal and background events with m_{bb} in

JHEP02(2014)130

the mass window of 90-130 GeV using the top reconstruction method to identify the Higgs decay products. This can be directly compared to the naive reconstruction methods from the previous section, summarized in table 1. Requiring that one of the two reconstructed top quarks satisfy $p_{T,t}/m_t \gtrsim 1$ improves the signal-to-background ratio by a factor of two.

While at this stage the cut-and-count analysis is running out of steam, it might be useful to show that the bucket reconstruction of the two top decays gives us additional handles on the backgrounds. For example, we can require the reconstructed momenta of the reconstructed Higgs and tops to be central and not too widely separated in rapidity, as is typically the case for heavy particle production,

$$\Delta\eta(t_1, t_2) < 3 \,, \quad \Delta\eta(t_i, H) < 2 \,. \tag{3.6}$$

In table 3 we show the corresponding signal and background rates. As can be seen, the ratio of signal to background is greatly improved. Taking the most aggressive criteria, requiring the leading reconstructed top to have $p_T > 300\,\text{GeV}$ and the $\Delta\eta$ cut, we reach a $S/B = 1/3.6$, of which 70% of the signal b-jets in the mass window are correctly identified from the Higgs decay. In general, this shows that the top reconstruction provides two handles that can improve the signal strength. First, it gives us a more accurate method to assign b-jets to the Higgs decay, reducing the combinatorial background. Second, it gives us kinematic information for the event that can be exploited to discriminate signal events.

Clearly, the proposed bucket analysis is unlikely to be the last experimental word in extracting purely hadronic $t\bar{t}H$ events from QCD background. However, our analysis shows that QCD and combinatorial backgrounds do not render this channel hopeless. Reconstructing the top decay products preferably in the slightly boosted regime can solve both problems and even leave the analysis with simple side bands, like the m_{bb} distribution.

4 Conclusion

We have demonstrated a method to extract the associated production of the Higgs along with top pairs in the *fully hadronic channel*, using the top buckets method or ref. [22] to reconstruct the hadronic tops. Using this reconstruction technique nets us several useful advantages over more naive methods to reduce the very large backgrounds.

First, by having an accurate method of determining which two of the four b-jets in the event should be assigned to the top pair, we can cut through the combinatorial problem of identifying the two b-jets from the Higgs decay. Side bands with $m_{bb} \lesssim 100\,\text{GeV}$ or $m_{bb} \gtrsim 200\,\text{GeV}$ can be used to determine the background shape and extract the background cross section after cuts. Second, the bucket algorithm not only identifies the b-jets from the top decay, it also gives a good approximation of the top and Higgs momenta. This allows us to place additional cuts, for example on the transverse momenta of the top quarks or on the $\Delta\eta$ of the various parton-level objects. Both of them help to reject background. In particular, requiring a small boost of the top quarks eases the combinatorial problem [10]. Further, more detailed analyses may improve on the fairly crude cuts we have chosen in this proof-of-concept paper.

JHEP02(2014)130

In this paper, we concentrated on demonstrating the stability of our reconstruction technique despite the potentially large simulation and theoretical uncertainties inherent to a QCD background consisting of four b-jets with many extra un-tagged jets. Using both Alpgen and Sherpa, we have validated that the simulation issues are under control. However, our study also clearly shows that the theory uncertainties on this kind of backgrounds are hardly covered by a factor two on the rate prediction. These issues can be mitigated in the experiments by use of the ample side-bands that this analysis affords. In addition to the background dominated regions outside of the Higgs mass window, there are also many sidebands available, for example in the distribution of jet multiplicity.

Acknowledgments

MB and TP would like to thank the Aspen Center of Physics because the idea for this paper was born on a Snowmass ski lift. TP would furthermore like to thank the CCPP at New York University for their hospitality, which added Washington Square as a second location crucial to the progress of this paper. Fermilab is operated by Fermi Research Alliance, LLC, under contract DE-AC02-07CH11359 with the United States Department of Energy.

A Signal and background simulations

In this appendix we will confirm that the analysis described in this paper does not critically depend on uncertainties in the way we compute our signal and backgrounds. For the signal and the $t\bar{t}b\bar{b}$ background we primarily rely on Sherpa [36] predictions with up to one additional hard jet merged using the Ckkw approach [37]. For the $t\bar{t}H$ signal we test our results using Madgraph [33], with up to one hard jet included in the Mlm scheme [38]. Both event samples are normalized to the next-to-leading order rate (extrapolated to 13 TeV) of 504 fb [39, 40], times a Higgs branching ratio of 57.7%. This corresponds to 129 fb for the purely hadronic decay channel. In table 4 we observe a small difference in the normalization of the two event samples. The reason is that as a cross check in the Madgraph simulation, we do not require hadronic top decays in the simulation. As a result, the decays to hadronic taus contribute to the signal. Our default Sherpa simulation conservatively does not include these events.

For the $t\bar{t}b\bar{b}$ background we test the Sherpa simulation with up to one hard additional jet with an Alpgen [32] simulation without additional hard jets. Again, both samples are normalized to the next-to-leading order rate of 1037 fb [41–43] after the generator cuts $p_{T,b} > 35\,\text{GeV}$, $|\eta_b| < 2.5$, and $\Delta R_{bb} > 0.9$. This rate is approximate because, in the absence of a next-to-leading order prediction for $\sqrt{s} = 13\,\text{TeV}$, we are forced to first extract the K factor for 14 TeV and the cuts of ref. [41–43], including a regularizing cut on the invariant mass of the two bottom quarks. We then multiply our cross section at 13 TeV by this K factor. This approach is not ideal, but better then just using the leading order prediction.

In table 4 we see that the transverse momentum distributions for the "reconstructed" top quarks (which, for $b\bar{b}b\bar{b}$, do not correspond to any parton-level tops) from Alpgen

JHEP02(2014)130

are softer than for SHERPA. This effect comes from the generically harder jets of CKKW merging, compared to those from the parton shower. In order to be conservative, we use the merged SHERPA results for our analysis. On the other hand, the difference of less than 20% is well within the theory uncertainties for this background.

As argued in section 2, the most dangerous background events should be correctly described by our ALPGEN simulation of the $bb\bar{b}\bar{b}jj$ background plus PYTHIA parton shower, as the required extra jets must be hard, and therefore well-modeled by the matrix-level process. With our merged SHERPA simulation of $bb\bar{b}\bar{b} + 0/1$ jet, we test several aspects of our main background simulation:

1. We check if the events with two additional hard jets are indeed the leading background after the kind of global cuts proposed in section 2. This aspect is very important for the appropriate simulation of the QCD background in an actual analysis.

2. We test if our analysis depends on the simulation of the second un-tagged jet either with the hard matrix element or through the parton shower. In this way we can estimate an important source of theory uncertainties.

3. As a measure of the level of agreement between the two simulations we compute the merged SHERPA event rate with a consistent variation of the renormalization and factorization scales. Ideally, the two simulations should agree within this scale variation in the signal region of the buckets analysis. Because the merged SHERPA prediction includes some leading next-to-leading order contribution such a numerical agreement also indicates that our $bb\bar{b}\bar{b}jj$ simulations should not be plagued by huge QCD corrections.

This extensive list of tests should give a clear answer to the question if purely hadronic $t\bar{t}H$ searches can be done in the presence of the large QCD backgrounds. Finally, we point out that merged SHERPA simulations can define excellent side bands in the n_j distribution [44–46], which together with side bands in m_{bb} should be sufficient to control the background rate in the signal region in an experimental analysis. Our results suggest that such an approach would require a merged simulation of $bb\bar{b}\bar{b}$ with up to at least two hard light-flavor or gluon jets, which is beyond our CPU capabilities.

In figure 4 we first show the normalized transverse momenta of the four b-jets. The curves are set to unit normalization, as the significantly different cross section of the $bb\bar{b}\bar{b}jj$ and the merged $bb\bar{b}\bar{b} + 0/1$ jets is almost entirely due to the different number of un-tagged jets in the events. In the left panel we see that the leading b-jet agrees in the two approaches, while the second to fourth b-jets become increasingly harder in the ALPGEN $bb\bar{b}\bar{b}jj$ sample. This is because, with two additional hard jets, the available recoil momentum is slightly larger. The sensitivity to the proper simulation of the recoil is also the reason why the ALPGEN curves are not covered by the scale variation of the SHERPA simulation.

The results for the leading un-tagged jets in the right panel of figure 4 look much less promising. The very different integrated rates under the curves reflect the additional events with only $bb\bar{b}\bar{b}$ in the hard process plus any number of parton shower jets. This is particularly obvious for the first un-tagged jet, where the SHERPA simulation includes a

JHEP02(2014)130

	$t\bar{t}H$		$t\bar{t}b\bar{b}$	
	MADGRAPH (merged)	SHERPA (merged)	ALPGEN (shower)	SHERPA (merged)
After acceptance eqs.(2.2) and (2.3)	1.390	1.197	7.903	8.363
2 tops tagged	1.100	0.894	5.893	5.882
$p_{T,t,1} > 100\,\text{GeV}$	0.866	0.709	4.684	4.868
$p_{T,t,1} > 200\,\text{GeV}$	0.342	0.289	1.806	2.189
$p_{T,t,1} > 250\,\text{GeV}$	0.180	0.165	0.978	1.295
$p_{T,t,1} > 300\,\text{GeV}$	0.092	0.089	0.502	0.724
	Mass window $m_{bb} = 90 - 130\,\text{GeV}$			
2 tops tagged	0.337	0.259	1.016	0.859
$p_{T,t,1} > 100\,\text{GeV}$	0.274	0.208	0.780	0.688
$p_{T,t,1} > 200\,\text{GeV}$	0.112	0.091	0.260	0.265
$p_{T,t,1} > 250\,\text{GeV}$	0.058	0.050	0.128	0.144
$p_{T,t,1} > 300\,\text{GeV}$	0.032	0.028	0.059	0.072

Table 4. Signal and background cross sections (in fb) after successive selection cuts, showing the different ways of simulating the signal and the irreducible $t\bar{t}b\bar{b}$ background. All conventions correspond to the final result shown in table 2. We use the SHERPA results for our main analysis.

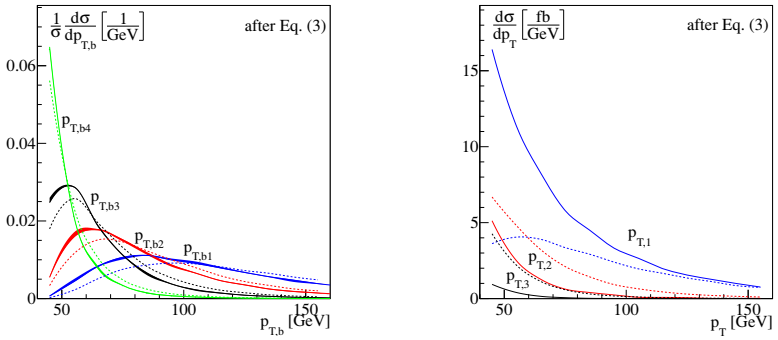

Figure 4. Normalized transverse momentum distributions of four b-tagged jets (left) and the leading three additional un-tagged jets (right). All events include four hard b-jets according to eq. (2.2), but no requirement on the number of additional un-tagged jets. The solid curves correspond to the merged $bb\bar{b}\bar{b} + 0/1$ jets simulation with SHERPA while the dashed curves show the $bb\bar{b}\bar{b}jj$ events from ALPGEN. The scale variation for the SHERPA result is indicated by the widths of the solid lines in the left panel.

majority of events with only one additional jet while the ALPGEN sample will always include a second hard jet together with the first. For the second un-tagged jet the integrated rates in the $p_{T,j}$ distributions are similar for the two samples. The ALPGEN simulation gives a significantly harder second jet from the matrix element while the second jet in our SHERPA sample (corresponding to the first jet from the parton shower) tends to be soft. The third un-tagged jet is the second parton shower jet in our SHERPA sample, while in the ALPGEN simulation it is the first parton shower jet radiated from a harder core process. Both effects combined result in a significantly harder $p_{T,j}$ spectrum for the ALPGEN sample. These distribution suggest that if our signal region should indeed require two or even three hard un-tagged jets to mimic top decay jets, the $bb\bar{b}\bar{b}jj$ sample from ALPGEN should be the appropriate, conservative estimate.

JHEP02(2014)130

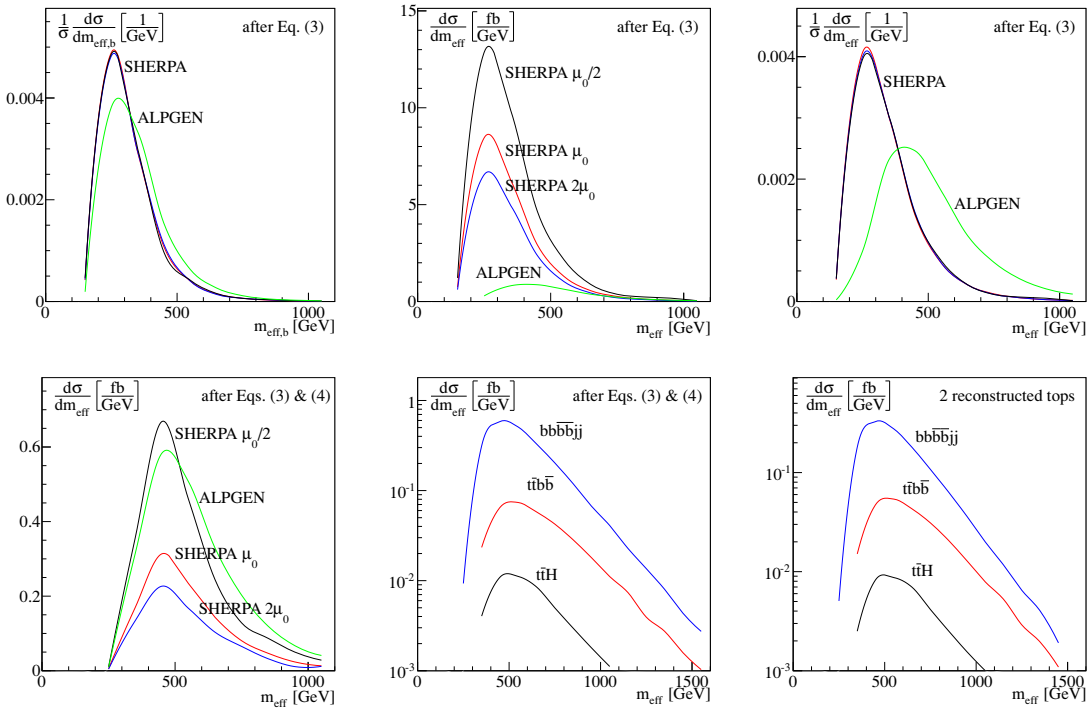

Figure 5. Effective mass distributions. In the upper panels we require four hard b-jets according to eq. (2.2), but no cut on the number of additional un-tagged jets. We show the normalized $m_{\text{eff},b}$ distribution for four b-jets (left), as well as the all-jet m_{eff} from SHERPA and ALPGEN normalized to the rates (center) and normalized to unity (right). In the lower panels we require two additional un-tagged jets fulfilling eq. (2.3). Here we show the ALPGEN vs SHERPA results (left), our default signal and background samples (center), and the default signal and backgrounds for all events with two valid top buckets (right).

In the upper panels of figure 5 we show the two relevant effective mass variables defined in eq. (2.4). We require four b-jets according to eq. (2.2) and any number of un-tagged jets after eq. (2.3). Unlike figure 1 we now only show the different results for the $bb\bar{b}\bar{b}$ background. In the upper left panel we again show the normalized observable from the multi-b sector. The conclusion follows from the discussion of the transverse momenta of the four b-jets: the dependence of the $m_{\text{eff},b}$ distribution on the simulation is small, clearly when we look at the SHERPA scale variation, but also in terms of the difference between ALPGEN and SHERPA. The only difference is that the $bb\bar{b}\bar{b}jj$ simulation with ALPGEN predicts slightly harder multi-b sectors.

In the upper central panel of figure 5 we see that in the background region the difference in rate between the two simulations is again dramatic, and certainly not covered by the scale variation of the $bb\bar{b}\bar{b}+0/1$ jets simulation with SHERPA. On the other hand, this result is entirely expected, and the difference becomes increasingly smaller once our analysis of the signal region requires something like $m_{\text{eff}} > 500\,\text{GeV}$. Increasing the cut to $m_{\text{eff}} > 700\,\text{GeV}$ brings them into agreement within scale uncertainties. In that regime the bulk of the $b\bar{b}b\bar{b}+0/1$ jet events do not contribute, so the two simulations should roughly agree within

JHEP02(2014)130

	$t\bar{t}H$ Sherpa	$bb\bar{b}\bar{b}jj$ Alpgen	$bb\bar{b}\bar{b}$+jets Sherpa		
			$2\mu_0$	μ_0	$\mu_0/2$
After acceptance eqs.(2.2) and (2.3)	1.197	54.420	18.825	25.812	50.974
2 tops tagged	0.894	29.356	7.507	10.091	20.473
$p_{T,t,1} > 100\,\text{GeV}$	0.709	20.838	5.049	7.283	13.843
$p_{T,t,1} > 200\,\text{GeV}$	0.289	5.194	1.155	1.419	3.018
$p_{T,t,1} > 250\,\text{GeV}$	0.165	2.213	0.488	0.717	1.361
$p_{T,t,1} > 300\,\text{GeV}$	0.089	0.917	0.218	0.351	0.645
	Mass window $m_{bb} = 90 - 130\,\text{GeV}$				
2 tops tagged	0.259	5.424	1.143	1.726	3.354
$p_{T,t,1} > 100\,\text{GeV}$	0.208	3.600	0.749	1.111	2.118
$p_{T,t,1} > 200\,\text{GeV}$	0.091	0.679	0.133	0.161	0.233
$p_{T,t,1} > 250\,\text{GeV}$	0.050	0.233	0.031	0.044	0.082
$p_{T,t,1} > 300\,\text{GeV}$	0.028	0.082	0.020	0.015	0.041

Table 5. Cross section (in fb) for signal and background events after successive selection cuts, showing the different ways of simulating the $bb\bar{b}\bar{b}$ background. All conventions correspond to the final result shown in table 2. Reconstructed tops for $bb\bar{b}\bar{b}$ do not correspond to any real parton-level object.

the scale variation of the Sherpa simulation. The upper right panel confirms that in the Alpgen simulation with the $bb\bar{b}\bar{b}jj$ hard process also gives a harder spectrum in m_{eff}.

In the lower panels of figure 5 we not only require four b-jets following eq. (2.2), but also at least two un-tagged jets fulfilling eq. (2.3). Two such additional jets are implicitly required for any event passing the bucket analysis described in section 3. First, we see in the left panel that simply asking for two un-tagged jets suppresses the central prediction from the merged $bb\bar{b}\bar{b}+0/1$ jets simulation to roughly half the $bb\bar{b}\bar{b}jj$ prediction. Both m_{eff} distributions peak around 500 GeV, and the Alpgen rate is covered by the scale variation of the Sherpa simulation. In the central lower panel we compare the distributions for our default signal and background simulations after requiring four b-tagged and two un-tagged jets. Finally, in the lower right panel we show the same distribution for all events passing the bucket analysis described in section 3. As compared to the acceptance cuts, eqs.(2.2) and (2.3), there is hardly any change, which means that the improvements by the bucket analysis are more promising than the global cuts proposed in section 2, with the added advantage of avoiding shaping the background distributions, such as m_{eff}.

From the above comparison we expect that for an actual top and Higgs analysis the two background simulations should be fairly consistent once we probe sufficiently hard multi-jet configurations. The scale uncertainty of our Sherpa simulation determines the numerical level of this consistency. Moreover, in the signal phase space region the $bb\bar{b}\bar{b}jj$ background simulation with parton shower should predict larger backgrounds and give us a conservative estimate. In table 5 we show the different $bb\bar{b}\bar{b}$ background rates after the buckets analysis. Indeed, the simulations agree roughly within the sizable scale uncertainties. The $b\bar{b}b\bar{b}jj$ simulation with Alpgen gives the largest rate, in particular once we require large, signal-

JHEP02(2014)130

like m_{bb} values and sizable p_T of the fake reconstructed top buckets. For the experimental analysis this implies firstly that the signal-to-background ratio for purely hadronic $t\bar{t}H$ events at 13 TeV can be of the order 1/3. Second, the remaining number of signal events will be an issue for a cut-and-count analysis. Lastly, the uncertainties on the background simulation will require a careful background determination from side bands and control regions. Any kind of $t\bar{t}H$ analysis which does not provide at least a slight mass peak in the m_{bb} distribution around 125 GeV would have a hard time convincing the authors of this study. Our buckets analysis will carefully ensure that this simple side band is clearly visible, in spite of the fact that this requirement might lead to a slightly reduced performance of our analysis.

B Top reconstruction

In this appendix, we provide metrics concerning the reconstruction of the top pair, using the bucket method originally proposed in ref. [22], and described in detail in section 3. As every event contains exactly four b-jets, reconstructing the tops leaves two b-jets that can be identified as coming from the decay of the Higgs, thus we can reconstruct its momentum as well. While cuts on the top and Higgs kinematics are not critical to the analysis presented in this paper, for example a multivariate version of the same analysis would immediately be able to benefit from a valid bucket reconstruction.

We can compare the magnitude and direction of the reconstructed top momenta to the true values of the parton-level tops, using Monte Carlo truth. As in ref. [22], the two kinematic variables we concentrate on are ΔR, defined between the parton-level top or Higgs and the closest of the two reconstructed tops or the reconstructed Higgs in the event, and $\Delta p_T/p_T^{\text{bucket}}$, again taking the difference in p_T between the parton-level object and the nearest bucket-reconstructed top or the reconstructed Higgs, normalized by the reconstructed p_T. Figure 6 shows these distributions for both the signal and the irreducible $t\bar{t}b\bar{b}$ background with real top quarks. Different lines correspond to the different reconstructed top $p_{T,t}^{\text{min}}$ cut for each bucket. For the Higgs plots (right column), these lines correspond to the different p_T cuts on the leading reconstructed top in an event.

Defining a "good" reconstruction as $\Delta R < 0.5$ and $|\Delta p_T/p_T| < 0.2$ for the tops, we give in table 6 the percentage of reconstructed tops in the $t\bar{t}H$ signal and the $t\bar{t}b\bar{b}$ background that are well-reconstructed. We also show the percentage of well-reconstructed Higgses in the signal sample. As can be seen, placing p_T cuts on the reconstructed tops increases the purity of well-reconstructed tops and Higgses, though clearly this sacrifices total cross section.

Finally, in figure 7, we show the efficiencies for this algorithm as a function of the parton-level top p_T, for several ranges of reconstructed p_T. The left panel displays the acceptances for the selection eqs.(2.2) and (2.3) as a function of $p_{T,t}$. Note that the acceptance for $t\bar{t}b\bar{b}$ sample is computed against the sample with the generation cut of $p_{T,b} > 35$ GeV and $R_{bb} > 0.9$. The central panel shows the efficiency for a single reconstructed top as a function of the parton-level top p_T relative to the number in the left panel. The efficiencies for $t\bar{t}H$ (black) and $t\bar{t}b\bar{b}$ (red) are shown. The contributions for the buckets with a parton-

JHEP02(2014)130

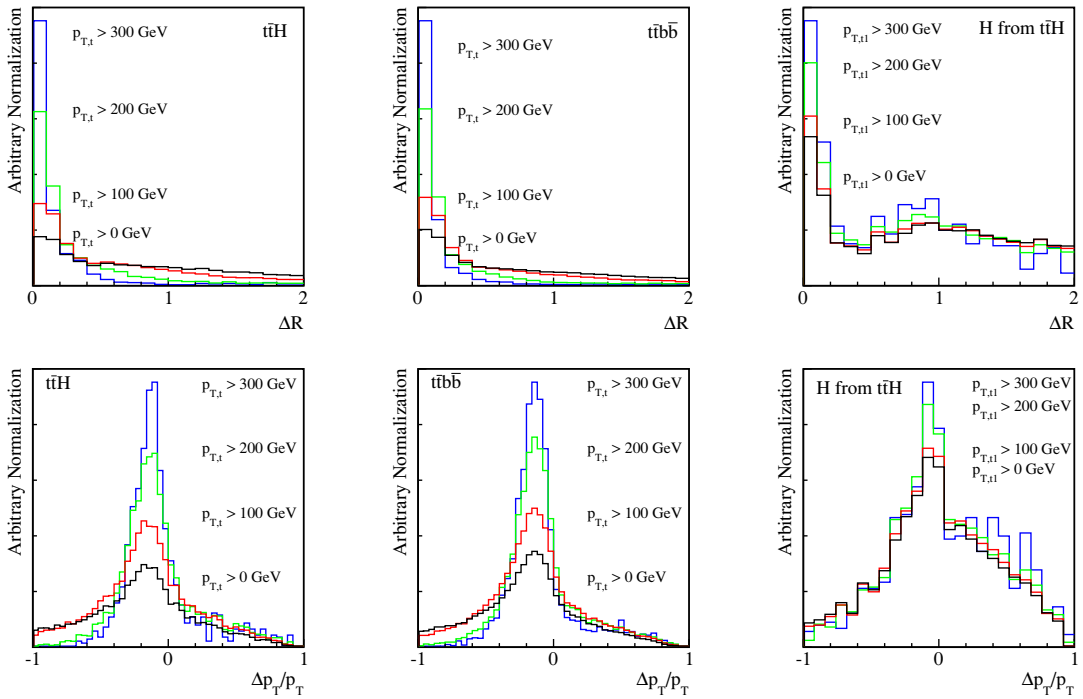

Figure 6. ΔR (top row) and $\Delta p_T/p_T$ (bottom row) distributions for tops in $t\bar{t}H$ (left) and $t\bar{t}b\bar{b}$ (center) samples, and for the Higgs in $t\bar{t}H$ (right).

		0 GeV	100 GeV	150 GeV	200 GeV	250 GeV	300 GeV
t from $t\bar{t}H$	$\Delta R < 0.5$	0.357	0.515	0.643	0.759	0.820	0.856
	$\lvert\Delta p_T/p_T\rvert < 0.2$	0.256	0.378	0.452	0.518	0.558	0.586
	$\Delta R < 0.5$ and $\lvert\Delta p_T/p_T\rvert < 0.2$	0.153	0.246	0.337	0.436	0.507	0.551
t from $t\bar{t}b\bar{b}$	$\Delta R < 0.5$	0.415	0.563	0.681	0.777	0.837	0.860
	$\lvert\Delta p_T/p_T\rvert < 0.2$	0.290	0.404	0.480	0.537	0.566	0.582
	$\Delta R < 0.5$ and $\lvert\Delta p_T/p_T\rvert < 0.2$	0.191	0.285	0.376	0.463	0.519	0.548
H from $t\bar{t}H$	$\Delta R < 0.5$	0.206	0.223	0.246	0.278	0.290	0.312
	$\lvert\Delta p_T/p_T\rvert < 0.2$	0.290	0.301	0.319	0.330	0.331	0.325
	$\Delta R < 0.5$ and $\lvert\Delta p_T/p_T\rvert < 0.2$	0.116	0.128	0.143	0.162	0.172	0.189

Table 6. Fraction of the buckets providing good momentum reconstruction for tops from $t\bar{t}H$ and $t\bar{t}b\bar{b}$, as well as Higgs from $t\bar{t}H$. Percentages of well-reconstructed tops are shown after placing a p_T cut on the top. For well-reconstructed Higgs, the percentages are shown for cuts on the highest-p_T reconstructed top.

level top found in $\Delta R < 0.5$ are indicated by dotted lines. We see both channels have similar efficiencies after the acceptance cut. The right panel gives the double tag efficiency as a function of the average of the parton-level transverse momenta of the two tops. Note that our algorithm always tags two tops and the resulting efficiencies are similar to the central panel in number.

JHEP02(2014)130

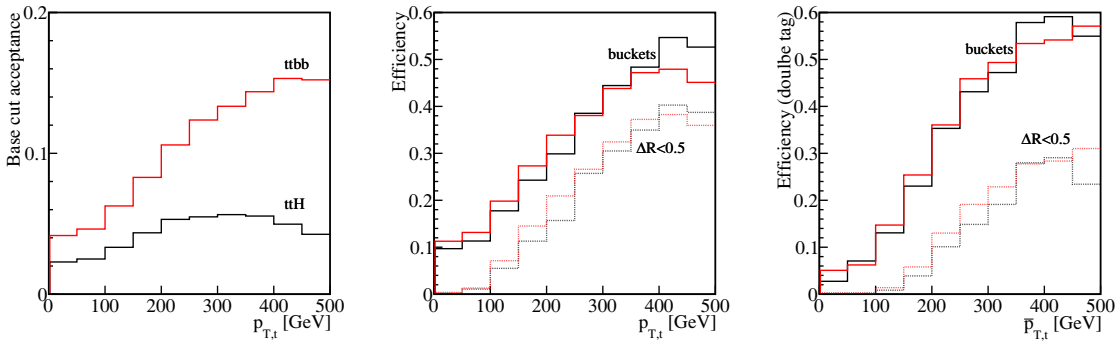

Figure 7. Acceptance efficiency for the basic selection cut as a function of $p_{T,t}$ (left). Efficiency for single/double bucket tag as a function of true transverse momentum of the top (center/right).

References

[1] P.W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132 [INSPIRE].

[2] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (1964) 508 [INSPIRE].

[3] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (1964) 321 [INSPIRE].

[4] 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 [arXiv:1207.7214] [INSPIRE].

[5] 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 [arXiv:1207.7235] [INSPIRE].

[6] M. Klute, R. Lafaye, T. Plehn, M. Rauch and D. Zerwas, *Measuring Higgs couplings from LHC data*, *Phys. Rev. Lett.* **109** (2012) 101801 [arXiv:1205.2699] [INSPIRE].

[7] D. López-Val, T. Plehn and M. Rauch, *Measuring extended Higgs sectors as a consistent free couplings model*, *JHEP* **10** (2013) 134 [arXiv:1308.1979] [INSPIRE].

[8] D.E. Morrissey, T. Plehn and T.M.P. Tait, *Physics searches at the LHC*, *Phys. Rept.* **515** (2012) 1 [arXiv:0912.3259] [INSPIRE].

[9] T. Plehn, *Lectures on LHC physics*, *Lect. Notes Phys.* **844** (2012) 1 [arXiv:0910.4182] [INSPIRE].

[10] T. Plehn, G.P. Salam and M. Spannowsky, *Fat jets for a light Higgs*, *Phys. Rev. Lett.* **104** (2010) 111801 [arXiv:0910.5472] [INSPIRE].

[11] A. Belyaev and L. Reina, $pp \to t\bar{t}H$, $H \to tau^+tau^-$: *toward a model independent determination of the Higgs boson couplings at the LHC*, *JHEP* **08** (2002) 041 [hep-ph/0205270] [INSPIRE].

[12] E. Gross and L. Zivkovic, $t\bar{t}H \to t\bar{t}\tau\tau$: *toward the measurement of the top-Yukawa coupling*, *Eur. Phys. J. Eur. Phys. J.* **59** (2009) 731 [INSPIRE].

JHEP02(2014)130

[13] C. Boddy, S. Farrington and C. Hays, *Higgs boson coupling sensitivity at the LHC using $H \to \tau\tau$ decays*, *Phys. Rev.* **D 86** (2012) 073009 [arXiv:1208.0769] [INSPIRE].

[14] P. Agrawal, S. Bandyopadhyay and S.P. Das, *Dilepton signatures of the Higgs boson with τ-jet tagging*, arXiv:1308.6511 [INSPIRE].

[15] F. Maltoni, D.L. Rainwater and S. Willenbrock, *Measuring the top quark Yukawa coupling at hadron colliders via $t\bar{t}H, H \to W^+W^-$*, *Phys. Rev.* **D 66** (2002) 034022 [hep-ph/0202205] [INSPIRE].

[16] D. Curtin, J. Galloway and J.G. Wacker, *Measuring the tth coupling from SSDL+2b measurements*, *Phys. Rev.* **D 88** (2013) 093006 [arXiv:1306.5695] [INSPIRE].

[17] P. Agrawal, S. Bandyopadhyay and S.P. Das, *Multilepton signatures of the Higgs boson through its production in association with a top-quark pair*, *Phys. Rev.* **D 88** (2013) 093008 [arXiv:1308.3043] [INSPIRE].

[18] ATLAS collaboration, *Search for resonances decaying into top quark pairs using fully hadronic decays in pp collisions with ATLAS at $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2012-102 (2012).

[19] ATLAS collaboration, *Search for resonances decaying into top-quark pairs using fully hadronic decays in pp collisions with ATLAS at $\sqrt{s} = 7$ TeV*, *JHEP* **01** (2013) 116 [arXiv:1211.2202] [INSPIRE].

[20] CMS collaboration, *Measurement of the top-quark mass in all-jets $t\bar{t}$ events in pp collisions at $\sqrt{s} = 7$ TeV*, arXiv:1307.4617 [INSPIRE].

[21] G. Kasieczka, *Early SUSY searches at the CMS experiment at CERN*, Ph.D. thesis, Heidelberg University, Heidelberg, Germany (2009).

[22] M.R. Buckley, T. Plehn and M. Takeuchi, *Buckets of tops*, *JHEP* **08** (2013) 086 [arXiv:1302.6238] [INSPIRE].

[23] S. Schumann, A. Renaud and D. Zerwas, *Hadronically decaying color-adjoint scalars at the LHC*, *JHEP* **09** (2011) 074 [arXiv:1108.2957] [INSPIRE].

[24] M.H. Seymour, *Searches for new particles using cone and cluster jet algorithms: a comparative study*, *Z. Phys.* **C 62** (1994) 127 [INSPIRE].

[25] T. Plehn and M. Spannowsky, *Top tagging*, *J. Phys.* **G 39** (2012) 083001 [arXiv:1112.4441] [INSPIRE].

[26] A. Abdesselam et al., *Boosted objects: a probe of beyond the standard model physics*, *Eur. Phys. J.* **C 71** (2011) 1661 [arXiv:1012.5412] [INSPIRE].

[27] A. Altheimer et al., *Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks*, *J. Phys.* **G 39** (2012) 063001 [arXiv:1201.0008] [INSPIRE].

[28] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: a program for Higgs boson decays in the standard model and its supersymmetric extension*, *Comput. Phys. Commun.* **108** (1998) 56 [hep-ph/9704448] [INSPIRE].

[29] LHC Higgs Cross Section Working Group collaboration, S. Heinemeyer et al., *Handbook of LHC Higgs cross sections: 3. Higgs properties*, arXiv:1307.1347 [INSPIRE].

[30] ATLAS collaboration, *Measurement of the mistag rate with 5 fb^{-1} of data collected by the ATLAS detector*, ATLAS-CONF-2012-040 (2012).

[31] CMS collaboration, *Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events*, CMS-PAS-BTV-13-001 (2013).

JHEP02(2014)130

[32] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001 [hep-ph/0206293] [INSPIRE].

[33] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: going beyond*, *JHEP* **06** (2011) 128 [arXiv:1106.0522] [INSPIRE].

[34] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[35] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[36] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007 [arXiv:0811.4622] [INSPIRE].

[37] S. Catani, F. Krauss, R. Kuhn and B. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063 [hep-ph/0109231] [INSPIRE].

[38] M.L. Mangano, M. Moretti, F. Piccinini and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, *JHEP* **01** (2007) 013 [hep-ph/0611129] [INSPIRE].

[39] W. Beenakker et al., *NLO QCD corrections to $t\bar{t}H$ production in hadron collisions*, *Nucl. Phys.* **B 653** (2003) 151 [hep-ph/0211352] [INSPIRE].

[40] S. Dawson, C. Jackson, L. Orr, L. Reina and D. Wackeroth, *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, *Phys. Rev.* **D 68** (2003) 034022 [hep-ph/0305087] [INSPIRE].

[41] A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, *NLO QCD corrections to $pp \to t\bar{t}b\bar{b} + X$ at the LHC*, *Phys. Rev. Lett.* **103** (2009) 012002 [arXiv:0905.0110] [INSPIRE].

[42] A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, *NLO QCD corrections to top anti-top bottom anti-bottom production at the LHC: 2. Full hadronic results*, *JHEP* **03** (2010) 021 [arXiv:1001.4006] [INSPIRE].

[43] G. Bevilacqua, M. Czakon, C. Papadopoulos, R. Pittau and M. Worek, *Assault on the NLO Wishlist: $pp \to t\bar{t}b\bar{b}$*, *JHEP* **09** (2009) 109 [arXiv:0907.4723] [INSPIRE].

[44] C. Englert, T. Plehn, P. Schichtel and S. Schumann, *Jets plus missing energy with an autofocus*, *Phys. Rev.* **D 83** (2011) 095009 [arXiv:1102.4615] [INSPIRE].

[45] E. Gerwick, T. Plehn, S. Schumann and P. Schichtel, *Scaling patterns for QCD jets*, *JHEP* **10** (2012) 162 [arXiv:1208.3676] [INSPIRE].

[46] S. El Hedri, A. Hook, M. Jankowiak and J.G. Wacker, *Learning how to count: a high multiplicity search for the LHC*, *JHEP* **08** (2013) 136 [arXiv:1302.1870] [INSPIRE].

[47] J. Cammin, *Study of a light standard model Higgs boson in the ttH channel with ATLAS and LHC decay mode independent searches for neutral Higgs bosons with OPAL at LEP*, BONN-IR-2004-06 (2004).

[48] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323] [INSPIRE].

[49] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, hep-ph/9907280 [INSPIRE].

[50] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop reconstruction with tagged tops*, *JHEP* **10** (2010) 078 [arXiv:1006.2833] [INSPIRE].

JHEP02(2014)130