

Table of Contents

CMS Single Top + Higgs analysis with bb decay.....	1
Introduction.....	1
Analysis goals.....	1
MC generation.....	2
Orthogonal Category Yields.....	2
Software framework(s).....	3
Synchronization.....	3
Objects definitions and event selection.....	3
Signal Categories.....	3
Control regions.....	4
Jet association.....	5
MVA optimization.....	5
Systematics.....	5
Limit setting.....	5
Blinding policy.....	6
Open tasks.....	6
Analysis Note 2013/113.....	6

CMS Single Top + Higgs analysis with bb decay

Introduction

This wiki is intended to help the coordination of the teams involved in the tHq search with $H \rightarrow bb$ decay.

People involved:

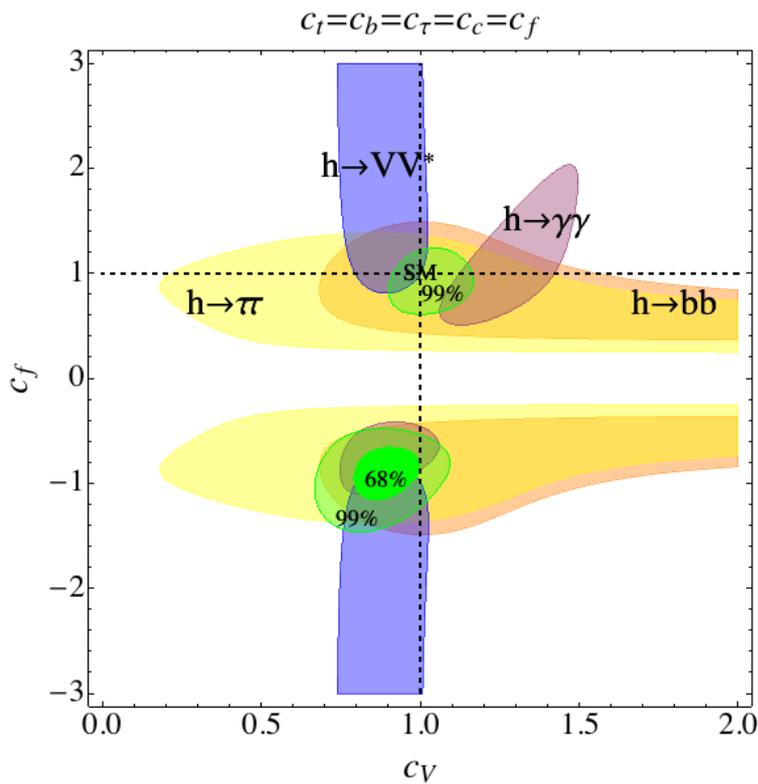
- Kansas: Alice Bean, Daniel Noonan
- Karlsruhe: Benedikt Maier, Christian Boeser, Simon Fink, Thorsten Chwalek, Jeannine Wagner-Kuhr
- Louvain: Andrey Popov, Matthias Komm, Andrea Giammanco
- Malaya: Afiq Anuar, Jyothsna Komaragiri
- Nebraska: Dan Knowlton, Rebeca Gonzalez Suarez, Ken Bloom

We meet every Friday at 14:00 CET, the agenda is under Top physics/Top other meetings [☞](#) category. Additional information on meetings is provided in a dedicated page.

Analysis goals

1) set limits on cross section times branching ratio in two idealized scenarios, "SM" (i.e. the rate of this process is kept floating but the Higgs has the usual SM properties) and "SM" (i.e. as SM but $y_t = -1$); this will probably require a MVA optimized for SM and one optimized for SM; an MVA without model-dependent inputs would be less sensitive but it would be interesting to have as well.

2) in the case of the MVA optimized for SM, we will strive to push the sensitivity until we discover or exclude the "island of wrong sign" in this plane:



3) set limits on other new-physics scenarios giving the same signature; for example, are we able to set an analysis that is independent from having or not having additional diagrams with new heavy particles exchanged internally?

MC generation

Signal MC:

We have a specific wiki for MC generation aspects.

It currently contains only details related to MadGraph generations.

ACTION for Benedikt: please add in that wiki also material for aMC@NLO: instructions, comments, useful links, ...

Background MC:

- tqZ: Simon is working on this using aMC@NLO
- t-channel+jets: Karlsruhe plans to work on this using aMC@NLO

Orthogonal Category Yields

This section is temporary, used only to estimate the number of events that are needed in a full simulation. Below, you will find three tables. First, a view at the preselection prior to creating orthogonal categories. Second, preselection yields of orthogonal categories separated by b-tagging working point and number of tagged jets in the event (more on this in a later section). This preselection includes jet cuts as well. Lastly, a similar table showing a full selection. The full selection includes the following variables and was optimized using TMVA's

SingleTopHiggsBB < Main < TWki

Cuts Method: absolute eta of forward light jet, minimum invariant mass of a b-jet and forward light jet, invariant mass of higgs candidate and forward light jet, invariant mass of all objects in final state, and the sphericity.

Preselection	tqh SM	tt-semi	tt-di	t-tchan	tbar-tchan	tth
AODSIM	499971	25274818	12119013	3915598	1711403	994697
GRID Sel.	299619	12780858	8515970	1459407	664480	435045
Lepton Cuts	223498	9224746	4039696	1161044	531013	208065

Preselection	tqh SM	tt-semi	tt-di	t-tchan	tbar-tchan	tth
CSVMB	10897	70055	13844	606	330	4488
CSV T3	9227	14379	2970	121	64	2625
CSV M4	5698	5093	977	18	11	5601
CSV T4	2894	619	205	4	2	1852

Full Selection	tqh SM	tt-semi	tt-di	t-tchan	tbar-tchan	tth
CSVMB	7035	9477	4101	307	159	658
CSV T3	4406	1021	438	40	22	138
CSV M4	3178	645	189	5	3	462
CSV T4	1072	24	12	2	0	65

Software framework(s)

We have two: the one developed by Andrey (Louvain/Moscow's framework) and used also in Nebraska, and the Karlsruhe one. The UZH group is adapting the skimming code from the main single top t- and s-channel analysis.

- Louvain's framework

Synchronization

- SingleTopHiggsBBSynchEx

Objects definitions and event selection

ACTION: we are using single-lepton triggers for simplicity, but we would gain statistics (without worsening S/B) by using cross-triggers like l+3j or l+b, and correspondingly loosening the offline lepton thresholds. But this would bring the complication of dealing with the turn-on curves of the hadronic part of the trigger, and additional systematics. Nevertheless, it is likely that in the end we would gain. Volunteers needed to study the optimal trigger option!

Analysis setup, definitions of physics objects, and reference event selection are described in SingleTopHiggsBBEventSel

Signal Categories

The signal categories define signal-enriched regions in the phase space which try to account for the topology of a single top + Higgs boson

SingleTopHiggsBB < Main < TWiki

event, with the Higgs boson decaying to a pair of b quarks and a leptonically decaying W. The topology is dominated by the presence of one isolated lepton (electron, muon), missing energy and three or four b jets, depending on whether the b quark from the gluon splitting enters the detector in a tag-able range. In addition, a jet formed by a light quark is expected. Accordingly, four different and **orthogonal** categories have been defined here; they serve as a baseline selection, suitable for synchronisation at an early stage of the analysis and being a starting point from which one can define additional cuts to further improve the S/B ratio and in which one can train MVA tools for a better separation of signal and background.

INFO: If a quantity is **not** cut on, this is denoted by a "-". In contrast, "= 0" would mean that the variable must be zero to pass the event selection.

Variable	CSV B	CSV M	CSV T 3	CSV T 4
# isolated leptons	= 1	= 1	= 1	= 1
# of jets	= 4	>= 5	= 4	>= 5
# of jets w/ CSV > CSV M	= 3	= 4	-	-
# of jets w/ CSV > CSV T	< 3	< 4	= 3	= 4

The CSV~~M~~ and CSV~~T~~ working points are defined as usual:

- CSV~~M~~ = 0.679
- CSV~~T~~ = 0.898

Given the low statistics in the two categories requiring 4 b tags (CSV~~M~~, CSV~~T~~4), one may think about not using a MVA in these high purity regions (just contributing to the global fit by the event yield).

But this strategy must be compared **as soon as possible** (i.e., as soon as the limit setting procedure with systematics is ready) with alternative strategies like, e.g., having only two categories (3 or at least 4 tags passing a single threshold) and/or using b-tagging as input to the MVA (which means a special treatment for b-tagging systematics).

Control regions

Regions with "one less b-tagging", e.g. 4j2t and 5j3t, are dominated by our main background (ttbar) and depleted in signal and are very close to our signal regions (4j3t, 5j4t), therefore it is quite natural to consider them in the final fit as handles to constrain ttbar normalization.

Other possible uses of control regions:

- to extract shapes? The problem is that in some cases the derived variables in the tHq assumption make no sense if there are less tags...
- to validate the background modeling; in this case, when a variable is not defined with a different number of tags, one can randomly pick one of the untagged jets and randomly associate it to H or top, and this would act as a proxy for the purpose of ttbar MC validation.

KB: Here's an idea: We are thinking at the moment that the bulk of the background is going to be $t\bar{t}$ with mistags. So, we should be able to take the events in the 4j2t and 5j3t (or maybe 5j2t) regions described above and use them to model the signal region. Take each untagged jet in those events and weight by the mistag probability. I think it gives you the right kinematic shapes (if done correctly, admittedly there are some tricky aspects that must be thought through) and it is also instantly correctly normalized. Systematic uncertainties on this come from the mistag probabilities and also the fact that we are ignoring that there can be $t\bar{t}$ plus real heavy-flavor production, but the latter can be estimated from MC. The technique can be tested entirely in MC, too, by doing it on the MC events and then comparing kinematic shapes to those in the MC events in signal regions.

Jet association

Jets are associated to top quarks and Higgs bosons according to a $t\bar{t}$ hypothesis and a tHq hypothesis. A dedicated neural network is used in both cases.

ACTION for Andrey: insert list of inputs for the two interpretations.

MVA optimization

For the final discrimination between signal and background we currently use NNs.

ACTION for Andrey: insert current list of inputs.

Systematics

See for example this wiki for a set of recipes of the most standard systematics in top-like analyses.

Special for this analysis:

- Signal modeling: compare aMC@NLO 4FS with 5FS (or reweight the key distributions at generator level, and compare with/without reweighting; this can be preferable to save computation time)
- top pt reweighting ([link](#))
- b-tagging: having to deal with tagging and anti-tagging in the same selection means that the "simple" approach known in CMS as "Rizzi's recipe" demands quite cumbersome formulas (see for example TOP-10-008 and associated notes). Formulas become more awkward if we split in mutually exclusive medium and tight categories. And if we use b-tagging as input for a MVA, we have to consider a procedure as in $tH(->bb)$.
- If we use cross-triggers: modeling of turn-on curves.

Limit setting

Consensus is on using "Combine" (the official Higgs PAG tool) for the final result, but Theta for quick evaluations (for example when comparing several options for fitting) because it is faster.

The current Theta setup is described in these slides [↗](#).

Binding policy

To be discussed asap.

Proposal by AG: until the freezing of the documentation, look at the data only in the control regions, and in the left-hand half of the final NN in the signal regions.

Open tasks

A list of open tasks in the analysis is available in page SingleTopHiggsBBTaskList.

Analysis Note 2013/113

```
> svn co -N svn+ssh://svn.cern.ch/repos/tdr2 myDir # where myDir is a placeholder for a name of your choice
> cd myDir
> svn update utils
> svn update -N notes
> svn update notes/AN-13-113
> eval `./notes/tdr runtime -csh` # for tcsh. use -sh for bash.
> cd notes/AN-13-113/trunk
# (edit the template, then to build the document)
> tdr --style=an b AN-13-113
```

You can commit your changes with

```
> svn commit -m "commit message"
```

New files will first need to be added with

```
> svn add NewFileNames
```

before they can be committed.

Note: I committed a script MAKENOTE for compiling without having to remember the exact command line.

This topic: Main > SingleTopHiggsBB

Topic revision: r17 - 2014-07-18 - AndreyPopov



Copyright © 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback