Study of the Top Quark and Antiquark Pair plus Four b-Quarks Production as a Background in the Top-Higgs Sector Searches

Leônidas Fernandes do Prado,[1,2] Andrej Saibel,[3] Aurore Savoy-Navarro,[1]
Matthias Schröder,[4] Sebastian Wieland,[4]

[1]IRFU - CEA, University Paris-Saclay, DPhP and CNRS/IN2P3, France
[2]Instituto de Física Teórica – UNESP, São Paulo, Brazil
[3]DESY, Germany
[4]Karlsruhe Institute of Technology, Germany

E-mail: leonidas.prado@cern.ch, andrej.saibel@cern.ch,
aurore.savoy.navarro@cern.ch, matthias.schroeder@kit.edu,
sebastian.wieland@cern.ch

Abstract. This note describes the ongoing study for the production of a top quark-antiquark pair associated to four b-quarks as a background in particular for the search of top quark-antiquark pair associated to a pair of Higgs bosons, both decaying into b-quarks. The detailed theoretical computation, at the NLO level, of the top quark-antiquark pair associated to two b-quarks production has proven to be important in the background estimate of the top quark-antiquark pair associated to a Higgs boson decaying into a pair of b-quarks process. Likewise the top quark-antiquark pair associated to four b-quarks production is of similar importance for the search of the top quark-antiquark pair associated to a pair of Higgs bosons, both decaying into b-quarks. However, because no such theoretical computation is available yet and while waiting for it, this study is attempting to estimate as best as possible with the currently available tools this process. This will be useful as well for a number of other processes in the Top-Higgs sector, at the femtobarn cross-section level, being searched for in the years to come.

Contents

1 Introduction

The top pair production ($t\bar{t}$) process is a dominant background for both the $t\bar{t}H$ and $t\bar{t}HH$ processes. Therefore this $t\bar{t}HH$ analysis uses the same overall $t\bar{t}$ background sample used for the $t\bar{t}H$ analysis. But in addition the $t\bar{t}H$ ($H\to b\bar{b}$) analysis has proven the specific importance of the $t\bar{t}+b\bar{b}$ background component in the $t\bar{t}H$ ($H\to b\bar{b}$) channel.

The $t\bar{t}+b\bar{b}$ component has been recently benefiting from a complete theoretical description at NLO, motivated by the $t\bar{t}H$ analysis as reported in [1]. This $t\bar{t}+b\bar{b}$ description is further implemented within a POWHEG based simulation and used in the $t\bar{t}H$ analysis for a better modeling of the $t\bar{t}+b\bar{b}$ background [2, 3].

The $t\bar{t}HH$ (4b) process has 6 b's in the final state. Thus the relevance of the $t\bar{t}+4b$ background for $t\bar{t}HH$ (4b) is similar to $t\bar{t}+b\bar{b}$ for $t\bar{t}H$ ($b\bar{b}$). Furthermore, the importance of $t\bar{t}+4b$ goes even beyond the $t\bar{t}HH$ analysis as other processes might well benefit from this improved estimate as: VBF(HH)=HH+jets; ZHH; WHH; $t\bar{t}H$; $t\bar{t}HH$; 4 tops. All as well in the range of 1 to a few tens of fb's and without mentioning a number of BSM processes.

2 A preliminary $t\bar{t}+b\bar{b}b\bar{b}$ study

Since there is no theoretical description of the $t\bar{t}+4b$ process at the moment a preliminary study is therefore done, using the sample at disposal (see last line of Table 2) to b compared with the overall $t\bar{t}$ sample of the third line of the same table. The $t\bar{t}+4b$ content provided by each of these two samples are compared. To achieve this, new $t\bar{t}$ event categories are implemented for selecting the $t\bar{t}+4b$ events and separating them from the $t\bar{t}+b\bar{b}$ event sample. Ne new categories are summarized below:

	SL channel
Number of leptons	1
p_T of leptons (e/μ) [GeV]	$> 30/29$
p_T of additional leptons [GeV]	< 15
$\|\eta\|$ of leptons	< 2.4
Number of jets	≥ 4
p_T of jets [GeV]	> 30
$\|\eta\|$ of jets	< 2.4
Number of b tagged jets	≥ 2

Table 1. Baseline event selection criteria in the single-lepton (SL) channel.

- $t\bar{t}+b\bar{b}$: exactly 2 extra b's
- $t\bar{t}+bbb$: more than 2 extra b's (excluding the case of exactly 4 extra b's)
- $t\bar{t}+4b$: exactly 4 extra b's.

Sample	MiniAOD events	Selected events
/TTHHTo4b_5f_LO_TuneCP5_13TeV_madgraph_pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/	9,800,000	781,129
/ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/	8,000,000	239,246
/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/	110,014,744	584,123
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/	43,732,445	232,264
/TTbb_Powheg_Openloops/asaibel-RunIIFall17MiniAODv2-PU2017_12Apr2018_ new_pmx_94X_mc2017_realistic_v14-v1-18783c0a07109245951450a1a4f55409/	5,311,500	212,226

Table 2. Signal and Background Simulation samples used in the $t\bar{t}HH$ analysis [4].

Process	Description	Number of Events	
		TTToSemiLeptonic	TTbb_Powheg_Openloops
$t\bar{t}+b\bar{b}$	Exactly 2 extra b's	40,935	78,117
$t\bar{t}+bbb$	More than 2 extra b's (excluding the case of 4 extra b's)	1,529	3,046
$t\bar{t}+4b$	Exactly 4 extra b's	202	496

Table 3. Number of events falling on the $t\bar{t}+b\bar{b}$, $t\bar{t}+bbb$, and $t\bar{t}+4b$ categories after applying the baseline selection of Table 1 an requiring ≥ 3 b tags

Table 3 shows the number of events, after applying the event selection used in the $t\bar{t}HH$ analysis [4]. For completeness, the number of events of the remaining categories of the ttbb_Powheg_Openloops_new_pmx sample is listed below:

- $t\bar{t}+b$: 74,918 events
- $t\bar{t}+2b$: 47,621 events
- $t\bar{t}+LF$: 6,601 events, ttcc: 1,427 events

The small statistics in tt+4b does not allow a good description of this background. Our DNN classifies these events mostly on the ttHH node, but not enough available statistics to make it possible to plot a defined shape as is observed in Figure 1. The increase in statistics of $t\bar{t}+4b$ in the dedicated TTbb_Powheg_Openloops sample is not enough to improve these plots. Figure 2 shows the signal and $t\bar{t}$ background split the new categories. The $t\bar{t}+4b$ category i small and almost not visible. It is seen that $t\bar{t}+b\bar{b}$ still dominates.

3 $t\bar{t}+4b$ Background: Conclusion and Perspectives

A theoretical description of $t\bar{t}+4b$ is required, similar to the one performed for $t\bar{t}+b\bar{b}$ in the $t\bar{t}H$ case. This is a long term study (2-3 years) and preliminary contacts have been taken with theoreticians.

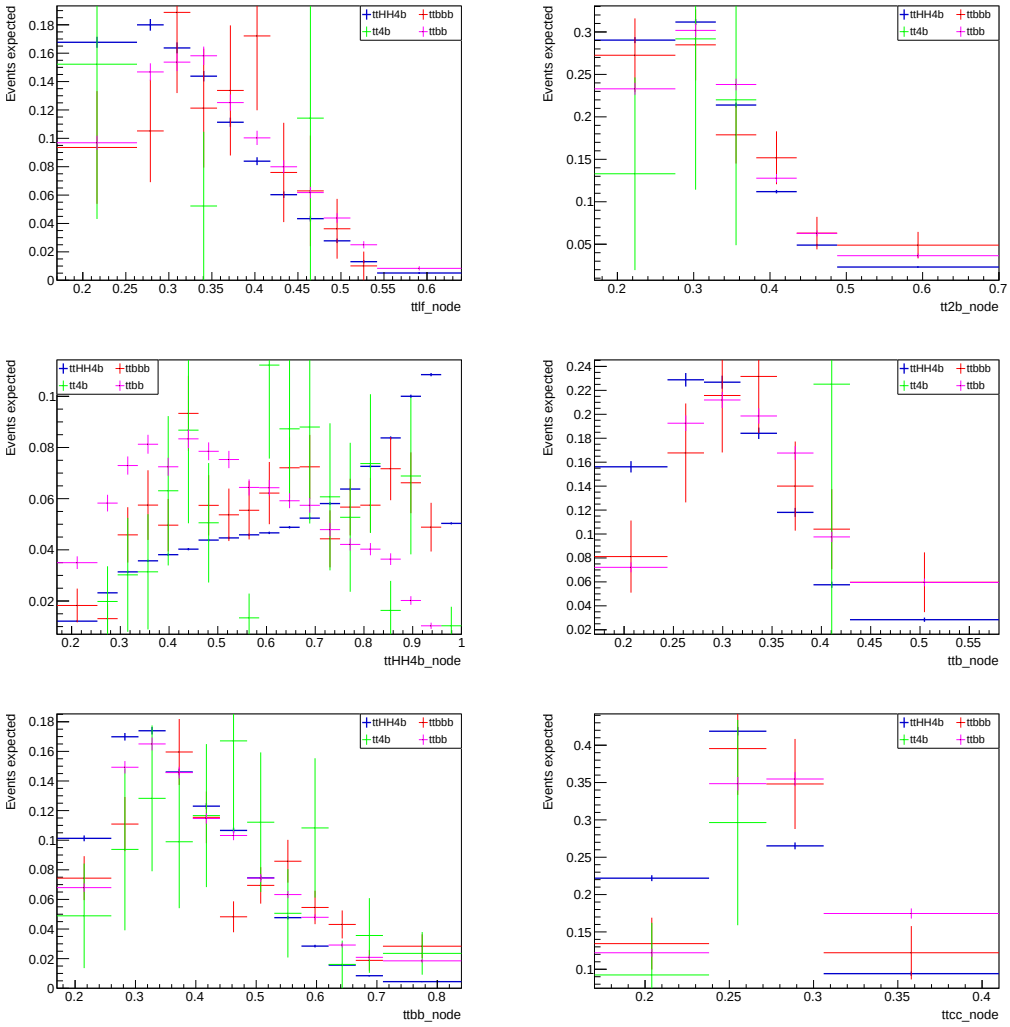

Figure 1. Histograms of $t\bar{t}HH$ (4b) events (blue), $t\bar{t}+b\bar{b}$ events (pink), $t\bar{t}+bbb$ events (red) and $t\bar{t}+4b$ events (green). All histograms are normalized to one in order to compare the shapes.

The $t\bar{t}+b\bar{b}$ (Powheg + Openloops) sample has already an increased number of $t\bar{t}+4b$, but not enough statistics. We think that a dedicated $t\bar{t}+4b$ sample at tree level will already provide a better estimate of this background, while waiting for the theoretical description. For the present $t\bar{t}$ (SL) HH(4b) analysis based on 2017 data, the overall $t\bar{t}$ sample is used without including a dedicated $t\bar{t}+4b$ category; the current $t\bar{t}+b\bar{b}$ category contains top pairs plus 2 or more b-quarks.

While waiting for a theoretical description of $t\bar{t}+4b$ as performed for $t\bar{t}+b\bar{b}$, a dedicated $t\bar{t}+4b$ MC sample at tree level will provide a better estimate of this background. Such a sample of 3M simulated events has been requested and is currently under processing [5] and will be used for the overall Run 2 analysis. More $t\bar{t}$ events and a dedicated $t\bar{t}+4b$ sample with NLO calculations will both be important already for the LHC Run 3.

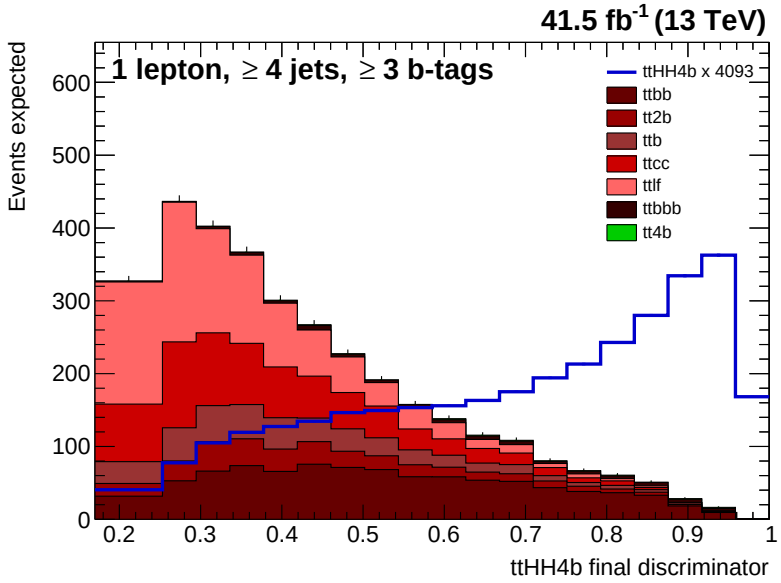

Figure 2. Final discriminant distribution for the signal node, constructed with signal and background processes using the $t\bar{t}$ overall sample with $t\bar{t} + b\bar{b}$ split in the categories of Table 3 and normalized to 41.5 fb^{-1} luminosity.

Acknowledgments

This work was supported by the São Paulo Research Foundation (FAPESP) under Grants No. 2016/01343-7, No. 2013/01907-0, No. 2015/26624-6 and No. 2018/11505-0 and by Science Without Borders/CAPES for UNESP-SPRACE under the Grant No. 88887.116917/2016-00.

References

[1] T. Ježo, J. M. Lindert, N. Moretti and S. Pozzorini, *New NLOPS predictions for $t\bar{t}+b$ -jet production at the LHC*, *Eur. Phys. J.* **C78** (2018) 502 [1802.00426].

[2] CMS Collaboration, *Measurement of tth and th in the h→bb channel with the full run 2 data sample*, CMS Analysis Note 2019/094, CERN, 2019. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/094.

[3] M. M. Horzela, *Merging of $t\bar{t}$ and $t\bar{t}+b\bar{b}$ simulations for an improved background modeling for $t\bar{t}H(b\bar{b})$ measurements*, MSc thesis, Karlsruher Institut für Technologie (KIT), 2019. ETP-KA/2019-13.

[4] CMS Collaboration, *Search for* $\mathrm{t\bar{t}HH}$ *in the semileptonic decay of the top pair and the higgs pair decay into b-quarks, using the 2017 data sample*, CMS Analysis Note 2019/173, CERN, 2019. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/173.

[5] S. Wieland and A. Saibel, "tt+4b simulation request: https://cms-pdmv.cern.ch/mcm/requests?prepid=HIG-RunIIFall17wmLHEGS-04395&page=0&shown=127."