Boosted $t\bar{t}H$ analyses in CMS and ATLAS

E. Le Quilleuc

CEA Saclay

May 4, 2017

Outline

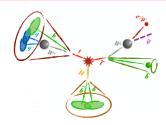# Outline

Introduction

Boosted analysis

Targets $t\bar{t}H(b\bar{b})$ events with high momenta Higgs/tops

Large-radius jets are used to catch the Higgs/top decay products

Advantages

- ▶ reduced combinatorics when assigning reconstructed objects to the Higgs decay products
- ▶ recover topologies where small radius jets start to merge

Drawback small fraction of the phase space, low signal statistics

Both CMS and ATLAS are considering a boosted analysis to improve their sensitivity to $t\bar{t}H$ in the lepton+jets channel

Current results

    CMS last results with 2015 dataset ▸ CMS PAS HIG-16-004

   ATLAS no public results available yet

Outline

CMS Boosted Analysis Selection

selection $1e/\mu$, ≥ 4 jets and ≥ 2 b-tags[1], **1 Higgs-tagged** large-R jet and **1 top-tagged** large-R jet

large-R jets C/A algorithm of $R = 1.5$, $p_T > 200$ GeV

Higgs- & top-tagging

- find subjets compatible with $H \to b\bar{b}$ for Higgs-tag (▸ arXiv:0802.2470v2) and with $t \to bq\bar{q}$ for top-tag (▸ arXiv:1006.2833) by redoing/reverting the jet clustering

- tag according to kinematics and b-tagging of subjets with multivariate technics

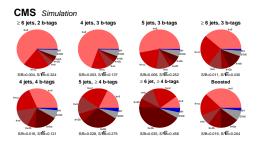**Boosted events** moderatly boosted, low number of b-tags

---

[1] 70%(1%) b-($light$-)jet efficiency

lepton+jets phase space is divided into 8 regions

- ▶ boosted region
- ▶ remaining events are classified wrt the number of jets/b-tags

CMS *Simulation*

Boosted region large contribution from $t\bar{t}+$ light flavour, comparable with $t\bar{t}+$ heavy flavour

CMS Yields at 2.7/fb (2015)

Process	4 jets, $\geq$ 4 b-tags	5 jets, $\geq$ 4 b-tags	$\geq$ 6 jets, $\geq$ 4 b-tags	boosted
$t\bar{t}$+lf	17.8 ± 10.8	17.7 ± 10.9	17.6 ± 11.3	45.1 ± 9.4
$t\bar{t} + c\bar{c}$	11.6 ± 8.2	22.1 ± 15.4	35.9 ± 24.9	21.8 ± 12.0
$t\bar{t}$+b	8.4 ± 4.4	14.8 ± 7.7	20.0 ± 10.9	10.3 ± 5.5
$t\bar{t}$+2b	3.5 ± 1.9	6.9 ± 3.7	12.3 ± 6.9	12.3 ± 6.6
$t\bar{t} + b\bar{b}$	10.1 ± 4.9	28.8 ± 13.9	73.4 ± 36.6	17.0 ± 8.4
Single Top	2.5 ± 1.1	4.3 ± 1.4	5.5 ± 2.0	7.0 ± 1.7
V+jets	1.0 ± 0.8	0.9 ± 0.8	1.4 ± 0.7	2.5 ± 0.8
$t\bar{t}$+V	0.3 ± 0.1	0.7 ± 0.3	1.6 ± 0.6	0.9 ± 0.3
Diboson	0.0 ± 0.0	0.1 ± 0.1	0.0 ± 0.0	0.1 ± 0.1
Total bkg	55.2 ± 23.0	96.5 ± 37.6	167.6 ± 65.7	117.0 ± 24.9
$t\bar{t}$H	0.9 ± 0.2	2.7 ± 0.6	5.9 ± 1.4	2.2 ± 0.3
Data	75	104	150	104
S/B	0.017	0.028	0.035	0.019
Data/B	1.4 ± 0.5	1.1 ± 0.4	0.9 ± 0.4	0.9 ± 0.2

The boosted region has a

- relatively good acceptance, compared with signal regions with $\geq$ 4 b-tags
- S/B almost twice lower than the most sensitive region $\geq$ 6 jets $\geq$ 4 b-tags
- smaller bkg (sg) relative uncertainty ($\sim 21\%(\sim 14\%)$) than $\geq$ 4 b-tags regions ($\sim 40\%(\sim 23\%)$), in part because less sensitive to b-tagging uncertainties
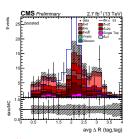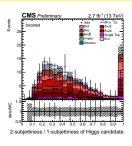
BDT Input Variables List for Boosted

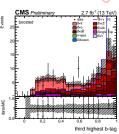BDT is trained to improve the sensitivity of the boosted channel

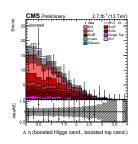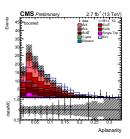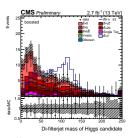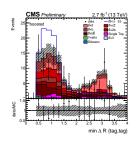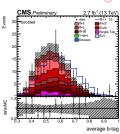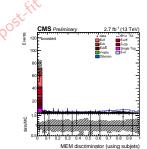| BDT event variable	Description
avg ΔR(tag,tag)	Average ΔR between b-tagged jets
τ_2/τ_1 of Higgs cand.	substructure variable, help to distinguish
	2 against 1 prong Higgs cand.
third highest CSV	Third highest CSVv2IVF value among all jets
fourth highest CSV	Fourth highest CSVv2IVF value among all jets
$\Delta\eta$(top,Higgs)	η difference between the Higgs cand. and the
	top cand.
aplanarity	$3/2\lambda_1$ (λ_i: eigeinvalues of momentum tensor)
m(Higgs, di-filteredjet)	Invariant mass of boosted Higgs cand.
	reconstructed from filtered subjets B1 and B2
min ΔR(tag,tag)	ΔR between the two closest b-tagged jets
avg CSV (all)	Average b-tag discriminator value of all jets
MEM discriminator	MEM discriminator using subjets from the
	top cand.
b-tagging likelihood ratio	no info.
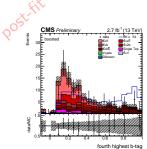
BDT Input Variables for Boosted, 2)
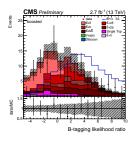
best discrimination from

- b-tagging information
- Higgs cand. mass

fairly good agreement data/MC
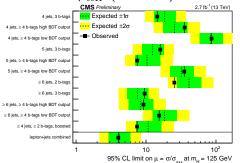
Each of 8 regions has its own fitted distribution – constructed to separate sg. from bkg. – which is fitted to data to extract the signal strengh $\mu_{t\bar{t}H}$ (< 4 b-tags : BDT, ≥ 4 b-tags : MEM)

No systematic uncertainties specific to boosted topologies are considered in the combination (see backup)

Limit on $\mu_{t\bar{t}H}$ (95% CL)

≥ 4 b-tags regions splitted into high/low BDT output

Category	Observed	Expected
boosted	7.5	$10.7^{+5.9}_{-3.5}$
combined	4.0	$4.1^{+1.8}_{-1.2}$

Boosted channel one of the most sensitive region

Outline

large-R jet anti-k_T R=1.0, inputs : standard small-R jets

simpler than large-R jets reconstructed from calorimeter clusters $\rightarrow$ no additional systematic uncertainties

b-tag jet passing b-tagging weight with 85%(3%) b-($light$-)jet efficiency

1 lepton electron or muon

1 Higgs-tag large-R jet $p_T > 200$ GeV, ≥ 2 b-tags[2]

1 top-tag large-R jet $p_T > 250$ GeV, ≥ 1 b-tag & ≥ 1 non b-tag

1 additional b-tag outside Higgs- and top-tag

Similar strategy than CMS 2 large-R jets moderately boosted, quite loose in term of b-tagging in the event

---

[2]with ΔR(large-R jet, b-tag) < 1.0

Study of Higgs and Top Purity

Study performed with slightly a different selection, with large-R jets reconstructed from calorimeter clusters

Higgs purity fraction of H-tags that contain a truth H $\to b\bar{b}$
    59%

top purity

quark matching	$t\bar{t}$ non	$t\bar{t}H$
	all-hadronic	semi-leptonic
at least b-quark from had. top	39%	69%
all partons from had. top	20%	40%

Good purity for both top-tag and H-tag with the selection 1 H-tag 1 top-tag

BDT Input Variables

$\sum_{jets} b$-tag weight   Sum of b-tag weights of all b-tagged jets in event (binned)

$\Delta R_{Higgs,top}$ ΔR between Higgs and top

$\sum_{jetsoutsideH/top} \text{mv2c10} / \sum_{jets} \text{mv2c10}$ Sum of b-tag weights of all b-tagged subjets outside t & H, divided by sum of b-tag weights of all jets

Higgs-tag d_{12} Higgs cand. $\sqrt{d_{12}}$

top-tag d_{12} Top cand. $\sqrt{d_{12}}$

large-R jet η Eta of leading large-R jet

$\Delta R_{b\bar{b}^{max_{p_T}}}$ ΔR between two leading b-jets

$\Delta R_{b\bar{b}^{Higgs}}$ ΔR between two leading b-jets in H

$\Delta R_{Higgs,lep}$ ΔR between Higgs and the lepton

$\Delta R_{Higgs,add.b}$ ΔR between H and leading b-jet outside t & H

$\Delta R_{top,add.b}$ ΔR between top and leading b-jet outside t & H

Conclusion

In Run 2, boosted analyses are part of $t\bar{t}H$ lepton+jets analyses

The moderately boosted signal makes the boosted analysis challenging

In CMS/ATLAS, the event selection is loose in b-tagging in order to

- ▶ improve the acceptance of boosted $t\bar{t}H$ events
- ▶ tag Higgs and top candidates with good purity

In CMS, the boosted analysis of the lepton+jets analysis is one of the most sensitive channels
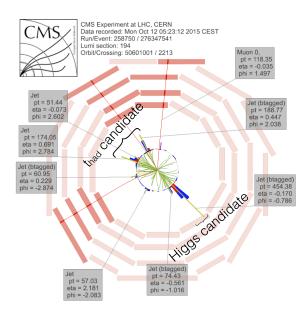
We expect that boosted analysis will play an important role in Run 2 and beyond with increasing luminosity

Backup

CMS Systematics (2015)

Source	Type	Remarks
Luminosity	rate	Signal and all backgrounds
Lepton ID/trigger efficiency	shape	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
b-tag HF fraction	shape	Signal and all backgrounds
b-tag HF stats (linear)	shape	Signal and all backgrounds
b-tag HF stats (quadratic)	shape	Signal and all backgrounds
b-tag LF fraction	shape	Signal and all backgrounds
b-tag LF stats (linear)	shape	Signal and all backgrounds
b-tag LF stats (quadratic)	shape	Signal and all backgrounds
b-tag charm (linear)	shape	Signal and all backgrounds
b-tag charm (quadratic)	shape	Signal and all backgrounds
QCD scale ($t\bar{t}$H)	rate	Scale uncertainty of NLO $t\bar{t}$H prediction
QCD scale ($t\bar{t}$)	rate	Scale uncertainty of NLO $t\bar{t}$ prediction
QCD scale ($t\bar{t}$+hf)	rate	Additional scale uncertainty of NLO $t\bar{t}$+hf predictions
QCD scale (t)	rate	Scale uncertainty of NLO single t prediction
QCD scale (V)	rate	Scale uncertainty of NNLO W and Z prediction
QCD scale (VV)	rate	Scale uncertainty of NLO diboson prediction
pdf (gg)	rate	Pdf uncertainty for gg initiated processes except $t\bar{t}$H
pdf (gg $t\bar{t}$H)	rate	Pdf uncertainty for $t\bar{t}$H
pdf (q$\bar{q}$)	rate	PDF uncertainty of q$\bar{q}$ initiated processes ($t\bar{t}$ W, W, Z)
pdf (qg)	rate	PDF uncertainty of qg initiated processes (single t)
Q^2 scale ($t\bar{t}$)	shape	Renormalization and factorization scale uncertainties of the $t\bar{t}$ ME generator, independent for additional jet flavors
PS Scale ($t\bar{t}$)	shape	Renormalization and factorization scale uncertainties of the parton shower (for $t\bar{t}$ events), independent for additional jet flavors

CMS Limits lepton+jets (2015)

Category	Observed	Expected
4 jets, 3 b-tags	14.5	$18.6^{+8.2}_{-5.5}$
4 jets, $\geq$ 4 b-tags high BDT output	35.7	$25.6^{+13.4}_{-8.1}$
4 jets, $\geq$ 4 b-tags low BDT output	86.6	$84.2^{+41.3}_{-25.8}$
5 jets, 3 b-tags	16.0	$12.3^{+5.5}_{-3.6}$
5 jets, $\geq$ 4 b-tags high BDT output	7.5	$10.3^{+5.6}_{-3.4}$
5 jets, $\geq$ 4 b-tags low BDT output	35.2	$31.9^{+16.1}_{-9.9}$
$\geq$ 6 jets, 2 b-tags	25.4	$41.1^{+21.1}_{-13.1}$
$\geq$ 6 jets, 3 b-tags	9.6	$7.6^{+3.3}_{-2.2}$
$\geq$ 6 jets, $\geq$ 4 b-tags high BDT output	9.2	$8.3^{+4.4}_{-2.7}$
$\geq$ 6 jets, $\geq$ 4 b-tags low BDT output	15.4	$18.3^{+9.6}_{-5.8}$
$\geq$ 4 jets, $\geq$ 2 b-tags, boosted	7.5	$10.7^{+5.9}_{-3.5}$
lepton+jets combined	4.0	$4.1^{+1.8}_{-1.2}$