Study of the associated $t\bar{t}H(H \rightarrow b\bar{b})$ production with the CMS detector

Tobias Verlage, V. Zhukov

I. Physikalisches Institut B, RWTH Aachen

December 2nd, 2014

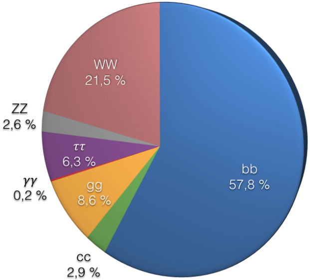
Terascale 2014 - 8th Anual Workshop

Introduction

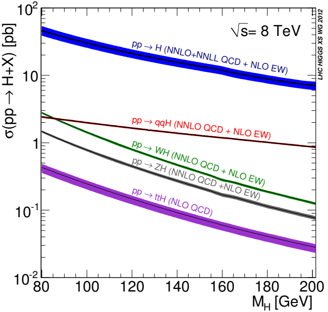
- $t\bar{t}H(\rightarrow b\bar{b})$ is an important channel to measure the properties of the Higgs, because
  - it's the only channel to measure the Higgs coupling only to quarks.
  - it's cross section measurement gives access to the Yukawa couplings of the top and bottom quarks

- experimental challenge posed by:
  1. low signal cross section
     - ★ good separation between signal and bkg needed.
     - ★ use MultiVariateAnalysis-technics to maximize the separation power.
  2. large (irreducible) SM background
     - ★ need a good knowledge of the background.
     - ★ control regions needed to constrain the background components.
  3. complicated final state
     - ★ reconstruct the initial event topology ($t\bar{t}$)
     - ★ increases signal/background separation.

SM BR for $m_{Higgs} = 125\,\mathrm{GeV}$



Higgs-Boson production



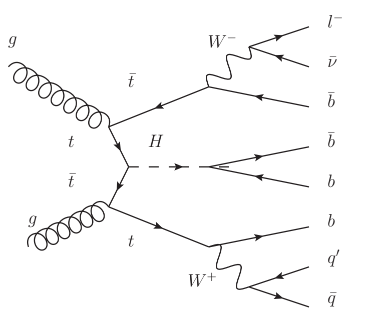
Introduction - Analysis Strategy

Analysis flow:

Consider semi-leptonic(μ/e) $t\bar{t}$ decay

1. preselection :
   - 1 lepton + $\geq$ 4 jets ($\geq$ 2 b-tagged)
2. jets association to $t\bar{t}$- and Higgs-decay
   - **Kinematic constrained fit** identifies products of the top decays.
   - **M**ulti**V**ariate**A**nalysis **& b-tagging** separate further correct and wrong associations.
3. Control- and signal-regions definitions
4. Signal - background separation based on MVA's
5. Systematics & Limit calculation

semi-leptonic $t\bar{t}H$ decay



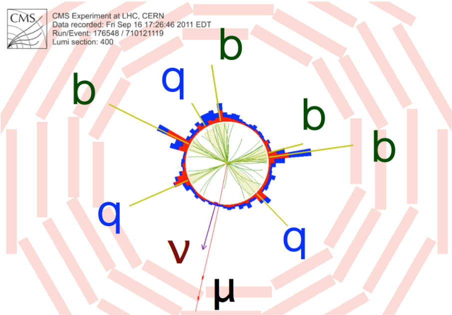
The CMS 2012 dataset is used with $L = 19.6\,\mathrm{fb}^{-1}$ @ $\sqrt{s} = 8\,\mathrm{TeV}$

Preselection - object definitions

- trigger stream: single isolated μ/e -trigger with $p_t > 24/27\,\mathrm{GeV}$

- exactly 1 tight lepton(e/μ):
  - $p_t > 30\,\mathrm{GeV}$
  - $|\eta| < 2.5/2.1$
  - tight ID & isolation

- no further loose lepton(e/μ):
  - $p_t > 10\,\mathrm{GeV}$
  - $|\eta| < 2.5/2.1$
  - loose ID & isolation



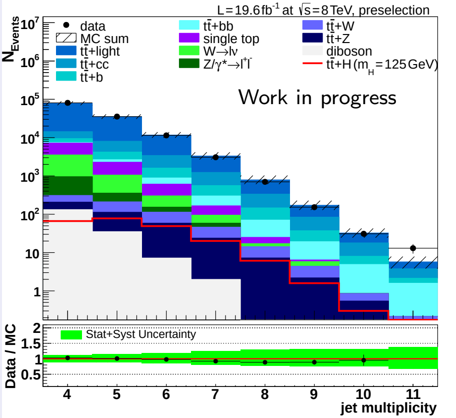
- at least 4 jets:
  - $p_t^{jet} > 30\,\mathrm{GeV}$
  - $|\eta| < 2.4$
  - looseID

- at least 2 b-tagged jets:
  - combined SV algorithm
  - efficiency: $\sim 60\%$
  - mis-tag rate: $1 - 10\%$

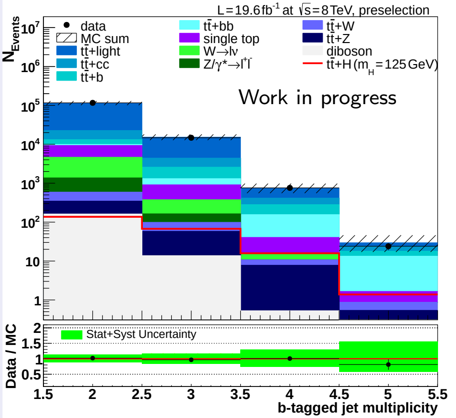
Corrections applied due to MC/data differences in efficiencies (trigger, leptons, b-tagging,..), PileUp, Top p_T, jet resolution, ...

Preselection - control plots

jet multiplicity after preselection



b-tagged jet multiplicity after preselection

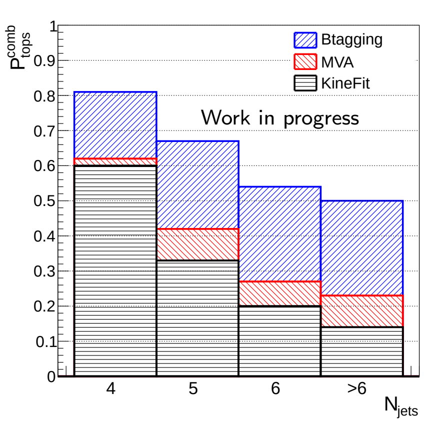


data-driven method to estimate QCD ($< 2\%$) is ongoing.

Jets association to the top-pair decay

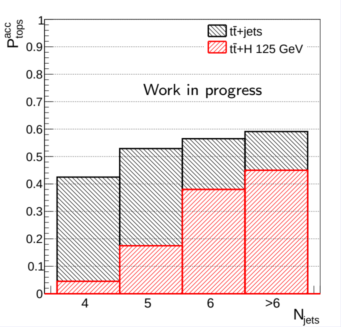
For each event, build all 4-jet combinations ⇒ Kinematic fit of W and top masses ⇒ MVA trained with correct vs. wrong combinations ⇒ Prefer combinations where jets assigned to the b-quarks of the $t\bar{t}$ decay are b-tagged.

Probability of the correct jets association to the top-pair



×

$P_{acceptance}$: reconstruct all jets from $t\bar{t}$



=

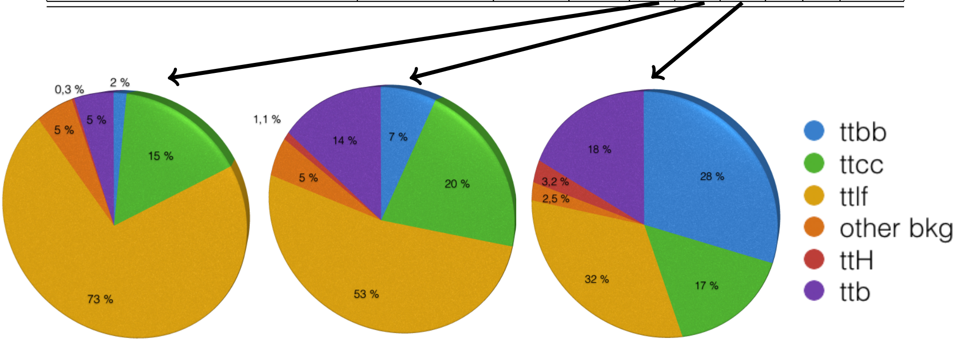
$$P_{total} = P_{acceptance} \times P_{correct\ association}$$

N_{jets}	4	5	6	≥ 7
P_{total}	0.35	0.36	0.31	0.30

Categorization of control and signal regions

the categorization benefits from the $t\bar{t}$ jets-association

number of jets	4	5		6			≥ 7		
$\mathbf{t\bar{t}}$(4 jets) $+n$ add. jets	**0**	**1**		**2**			≥ 3		
additional b-tagged jets	/	**0**	**1**	**0**	**1**	**2**	**0**	**1**	≥ 2



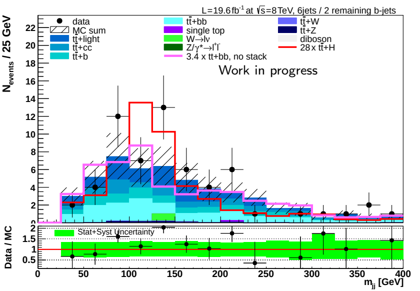
- $t\bar{t}$+light flavor and $t\bar{t}+c\bar{c}$ distribute in all categories.
  - can be constrained in categories with low b-tags.
- $t\bar{t}+b\bar{b}$ and $t\bar{t}+b$ contribute mainly only in signal enriched categories.
  - get no good handle on controlling these backgrounds.

$t\bar{t}bb$ enriched control regions

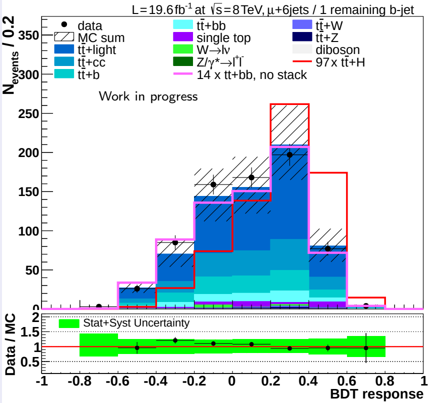
MVA's are used to separate $t\bar{t}$H and $t\bar{t}b\bar{b}$ processes in all categories with at least 2 jets in addition to the $t\bar{t}$-pair

- two types of input variables are used:
  - **„global event topology"** variables (e.g. FoxWolfram moments)
  - **„jets association related"** variables (e.g. mass of the Higgs candidate)

$t\bar{t} + 2$ jets (2 b-tags): mass(Higgs candidate)



$t\bar{t} + 2$ jets (1 b-tags): MVA ($t\bar{t}$H vs $t\bar{t}b\bar{b}$)



cut on MVA ($t\bar{t}$H vs $t\bar{t}b\bar{b}$) defines:

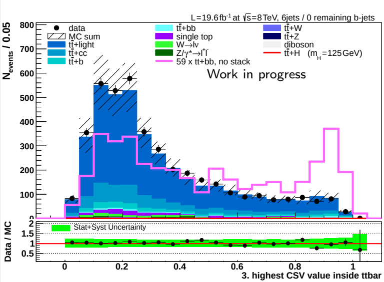
- signal region with enriched $t\bar{t}$H (signal fraction up to 5%)
- control region with a large $t\bar{t}b\bar{b}$ component (up to 34%).

control regions: ttbb / tt+jets separation based on MVA's

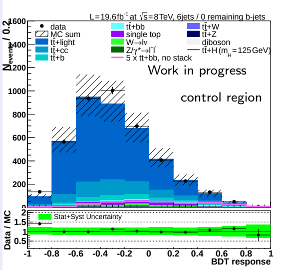
MVA's are used to separate $t\bar{t}b\bar{b}$ and $t\bar{t}$+jets(light, cc) processes.

- input variables (in average 8):
  - **„global event topology"** variables
  - **„jets association related"** variables
- MC simulations of $t\bar{t}b\bar{b}$ and $t\bar{t}$+jets (light, cc) are used for the training.
- separately done for each control-region.

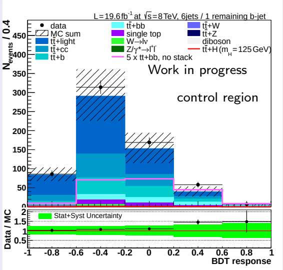
$t\bar{t} + 2\,(0$ b-tags): 3^{rd} highest b-tag in $t\bar{t}$



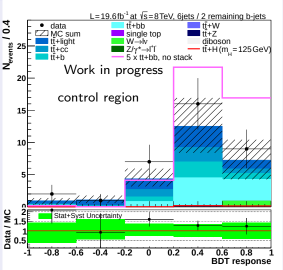
$t\bar{t} + 2$ jets (0 b-tags)



$t\bar{t} + 2$ jet (1 b-tags)



$t\bar{t} + 2$ jet (2 b-tag)

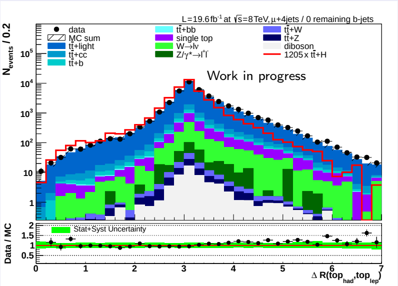


signal regions: ttH / background separation based on MVA's

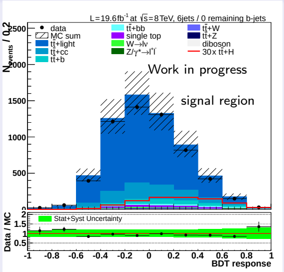
MVA's are used to separate signal and background processes.

- input variables (in average 10):
  - **„global event topology"** variables
  - **„jets association related"** variables
  - **„MVA ($t\bar{t}$H vs $t\bar{t}b\bar{b}$) response"**
- MC simulations of $t\bar{t}H$ and $t\bar{t}$+jets (lf, cc, b, bb) are used for the training.
- separately done for each signal-region.

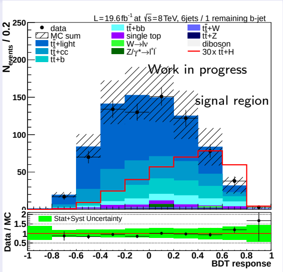
$t\bar{t}$ + 0 jets: ΔR(Tops)



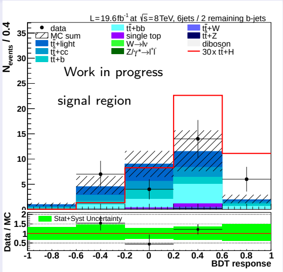
$t\bar{t}$ + 2 jets (0 b-tags)



$t\bar{t}$ + 2 jet (1 b-tags)



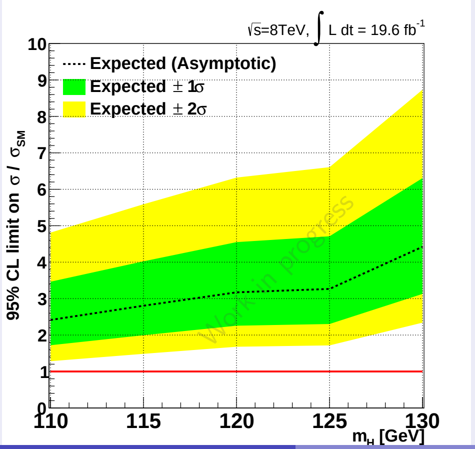
$t\bar{t}$ + 2 jet (2 b-tag)



Limit calculations

MVA responses are used in a combined model for all signal and control regions taking into account correlations of the systematic uncertainties.

$t\bar{t}H$: 95% CL. limit on $\sigma^{Higgs}/\sigma^{Higgs}_{SM}$



Systematics	ttH	Bkg. total	treatment
luminosity	2.6%	2.6%	norm.
pileup	$< 2\%$	$< 1\%$	shape
efficiency trigger	1%	1%	norm.
efficiency μ	1%	1%	norm.
efficiency ele	1%	1%	norm.
top p_T reweighting	0%	$1-5\%$	shape
b-tagging	$5-17\%$	$4-21\%$	shape
JER	$1-2\%$	$1-3\%$	shape
JES	$2-8\%$	$3-12\%$	shape
PDF gg	8%	$2.6-9\%$	norm.
PDF $q\bar{q}$		$4.2-7\%$	norm.
PDF qg		4.6%	norm.
$Q^2(t\bar{t}+0,..,3p)$		$2-22\%$	shape
$Q^2(t\bar{t}+bb)$		$4-18\%$	shape
$Q^2(t\bar{t}+cc)$		$3-21\%$	shape
$\sigma(t\bar{t}+Higgs)$	12.5%		norm.
$\sigma(t\bar{t})$		3%	norm.
$\sigma(t\bar{t}+V)$		15%	norm.
$\sigma(t\bar{t}+bb)$		50%	norm.
$\sigma(t\bar{t}+b)$		50%	norm.
$\sigma(t\bar{t}+cc)$		50%	norm.
$\sigma(single\ t)$		2.0%	norm.
$\sigma(V+jets)$		$1.2-1.3\%$	norm.
$\sigma(VV+jets)$		3.5%	norm.

- analysis is still blinded
- 95% CL. expected limit on $\sigma^{Higgs}/\sigma^{Higgs}_{SM}$ $(m_H = 125\,\mathrm{GeV})$:
  - for 2012 CMS-data: 3.2
- ATLAS 2012-data results:
  - expected limit: 3.1
  - observed limit: 4.2

summary

- presented an optimized $t\bar{t}H(\rightarrow b\bar{b})$ analysis
  - which handles the experimental challenges
    - ★ developed a method to associate jets with the $t\bar{t}$-decay, which improves the signal/background separation.
    - ★ in total 16 control regions and 14 signal regions are used.
    - ★ control regions are used to constrain all background components.
    - ★ shape analysis with dedicated trained MVA's for each region.
  - sets an expected upper limit on σ_{Higgs} for a Higgs mass of $m_H = 125\,\mathrm{GeV}$ to 3.2 times SM predictions with L=19.6 fb^{-1} at $\sqrt{s} = 8\,\mathrm{TeV}$
- ongoing work:
  - independent $t\bar{t}b\bar{b}/t\bar{t}jj$ estimation, which reduces the systematic uncertainty from 50% to 32%.