Karlsruhe Institute of Technology

Background modeling in tHq

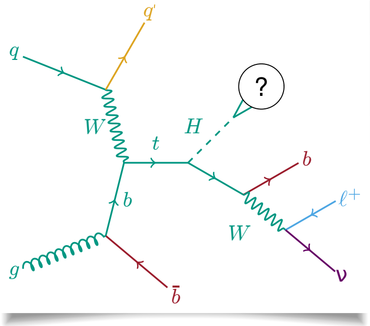
HXSWG meeting – ttH/tH subgroup

Simon Fink for the CMS tHq working group | 02.02.2015

INSTITUT FÜR EXPERIMENTELLE KERNPHYSIK

Event Topology

- Event topology:
  - Two b quarks
  - One isolated lepton
  - One light forward jet
  - Missing transverse energy
  - Two Higgs decay products



Different final states

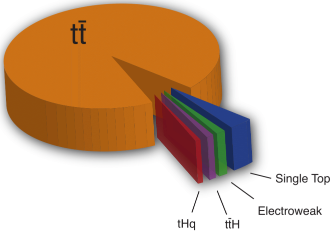
- $H \to b\bar{b}$
  - Highest possible branching ratio
  - Low signal to background ratio
  - Public result: *HIG-14-015*

- $H \to \gamma\gamma$
  - High purity
  - Very low cross section
  - Public result: *HIG-14-001*

tHq, $H \rightarrow b\bar{b}$

$H \to b\bar{b}$: Backgrounds

- $t\bar{t}$ MADGRAPH – Dominant background
- Single Top POWHEG
- $t\bar{t}H$ PYTHIA6
- W/Z+jets MADGRAPH
- Di-Boson PYTHIA6
- QCD PYTHIA6
  - Adapted $\not{E}_T$ cut
  - Negligible, shown by ABCD method
- tZq PRIVATE PRODUCTION, aMC@NLO
  - Very low cross section, negligible for $\kappa_f = -1$ analysis



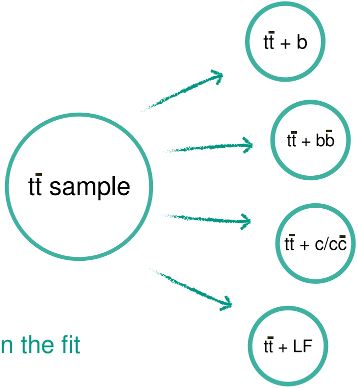
Main effort put into understanding $t\bar{t}$ background

MC approach

DD approach

$t\bar{t}$ heavy flavor separation

- Modeling of the $t\bar{t}b\bar{b}$ component difficult
- Separated $t\bar{t}$ sample into different subsets depending on additional flavors
- Splitting procedure adopted from other CMS analyses
  - **TOP-12-024** - Measurement of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ at 7TeV
  - **HIG-13-019** - Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons



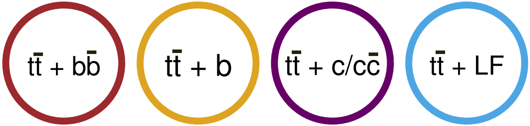
- Used differently in MC and DD approach:
  - **MC** : $t\bar{t}$ components are allowed to float in the fit independently with 50% unc.
  - **DD** : splitting and uncertainty input before building the dd template

HF Separation: Jet-parton matching

- Start with a list of status 2 partons of the event
  - Dismiss everything, but b and c quarks
  - Dismiss b and c quarks which were produced in EW interactions (t $\rightarrow$ bW and W $\rightarrow$ cs)
    - Exclude those status 2 partons, which are ΔR-matched to status 3 quarks originating from EW decays
  - Exclude quarks whose mother is a beam particle
- Every parton of the reduced list gets matched to a reconstructed jet
  - Use a ΔR matching with maximal ΔR is 0.5
  - Only the closest jet is assigned to the parton
  - More than one parton can be assigned to one jet

HF Separation: Event categories

- Event is categorized according to the first matching category from top to bottom:
- Defined categories:
  - `TwoBottom:` at least two jets are matched to b quarks
  - `DoubleBottom:` one jet is matched to two or more b quarks
  - `SingleBottom:` only one jet is matched to a b quark
  - `TwoCharm:` at least two jets are matched to c quarks
  - `DoubleCharm:` one jet is matched to two or more c quarks
  - `SingleCharm:` only one jet is matched to a c quark
  - `UnmatchedBottom:` there is an unmatched b quark in the event
  - `UnmatchedCharm:` there is an unmatched c quark in the event
  - `NoHF:` there are no b or c quarks in the event
- Group similar categories into larger ones



Data-driven model of $t\bar{t}$ background

- A data-driven approach is adopted as a cross check
  - The data template is taken from $t\bar{t}$ control region
    - Purity is $\sim 83\%$ (dileptonic $t\bar{t}$ contributes for another $\sim 10\%$)
  - Each event is assigned a weight to reflect the probability that an event with the same jet momenta and flavours ends up in a signal region
  - The probability that an event containing n jets (with momenta p_i and flavours f_i) obtains m b-tags is

$$\mathcal{P}_m = \sum_{\text{comb}} \prod_{i=1}^{m} \epsilon(p_i, f_i) \cdot \prod_{j=m+1}^{n} (1 - \epsilon(p_j, f_j)),$$

    where ϵ is b-tagging efficiency and the sum is taken over all $\binom{n}{m}$ ways to choose m tagged jets

- The weight is calculated as

$$w = \mathcal{P}_3/\mathcal{P}_2 \quad \text{or} \quad \mathcal{P}_4/\mathcal{P}_2$$

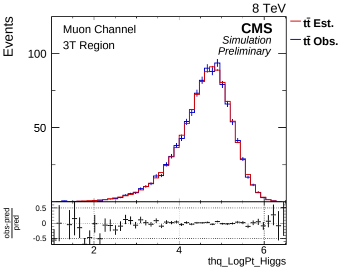
in 3T or 4T region resp.

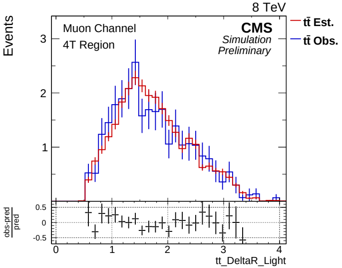
Data-driven model of $t\bar{t}$ background

- The method is designed to reproduce both shape and normalisation
- Checked with a closure test in simulation
  - Normalisation and shapes are reproduced fairly

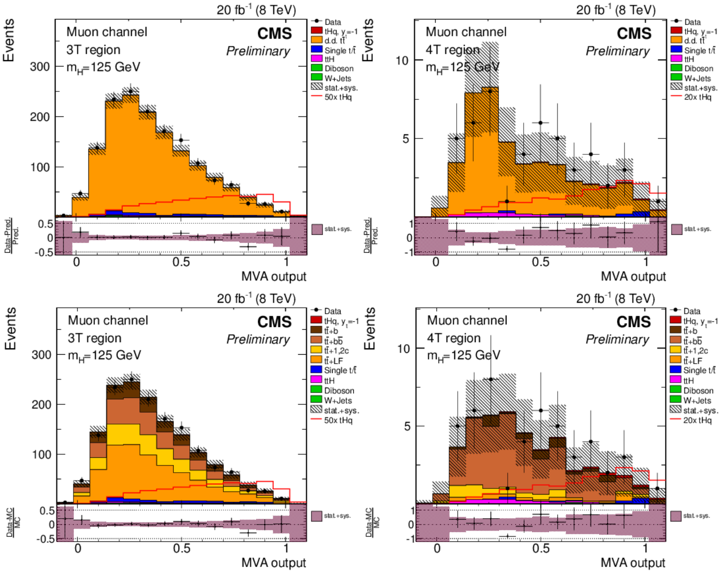
NT	Predicted	Observed	Ratio
3	1093.6 ± 1.9	1082.5 ± 8.9	0.990 ± 0.008
4	29.71 ± 0.45	28.16 ± 1.39	0.948 ± 0.049

- Additional residual reweighting is applied
  - Forces the closure test
  - Accounts for impurity of the $t\bar{t}$ control region
  - Corresponding conservative systematics assigned





Comparing DD- and MC-driven



$H \to b\bar{b}$: Uncertainties

- Q^2 scale
- Matching
- Theoretical pdf and QCD scale uncertainties
- Top p_T reweighting
  - Already well established in CMS
  - Also see older talks of this forum, e.g. CMS Experimental tt+jets Talk by Wuming Luo
- $t\bar{t}$ HF rates
  - 50% on each template
  - Is used as normalization uncertainty in the MC approach
  - splitting and uncertainty input before building the DD template

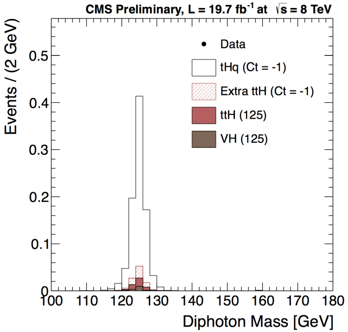
Process	pdf			QCD Scale			
	gg	$q\bar{q}$	qg	$t\bar{t}$	V	VV	$t\bar{t}H$
tHq			2%				
$t\bar{t}H$	9%						12.5%
$t\bar{t}$	2.6%			3%			
Single top			4.6%	2%			
W+jets		4.8%			1.3%		
Z+jets		4.2%			1.2%		
Dibosons						3.5%	

$t\bar{t}$ HF rates main reason why DD approach is more susceptible to systematic variations

tHq, H $\rightarrow \gamma\gamma$

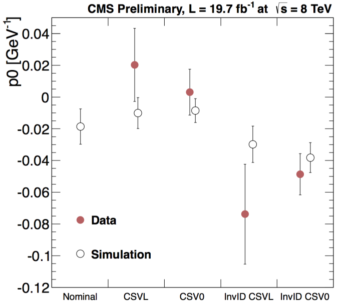
$H \to \gamma\gamma$: Backgrounds

- Dealing with two different types of backgrounds
  - Resonant backgrounds
    - BGs with $H \to \gamma\gamma$
    - Appear under the signal peak
    - dominated by $t\bar{t}H$ PYTHIA
    - Suppression of $t\bar{t}H$ with a 5-variable likelihood product discriminant
    - Small VH PYTHIA contribution
    - Taken from simulation
  - Non-resonant backgrounds
    - Main backgrounds: $\gamma\gamma$+jets SHERPA, γ+jets PYTHIA, $t\gamma\gamma$ WHIZARD, $t\bar{t}\gamma\gamma$ WHIZARD
    - Evaluated from data in the $m_{\gamma\gamma}$ sidebands



$H \rightarrow \gamma\gamma$: non-resonant backgrounds

- No data observed in sidebands of $m_{\gamma\gamma}$
- Assume exponential shape for non-resonant background
- Fit shape in four different control regions to data (checked in MC)
- Background shape f_{bg} used to predict influence in signal region with scale factor α
  - $\alpha = \frac{\int_{\text{signal region}} f_{bg}(m_{\gamma\gamma})dm}{\int_{\text{sideband region}} f_{bg}(m_{\gamma\gamma})dm}$
  - Shape from CSV0 control region taken as nominal background shape
  - Difference to taking the InvID CSV0 shape is used as uncertainty



Results in a 33% uncertainty on the continuous background

H $\rightarrow \gamma\gamma$: Systematic uncertainties

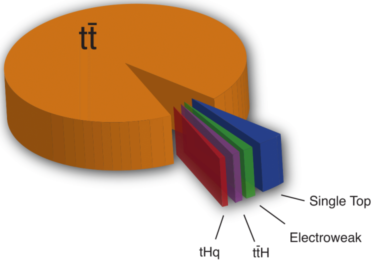
	tHq	$t\bar{t}$H	VH	Continuous BG
Luminosity	$\pm$2.6%	$\pm$2.6%	$\pm$2.6%	-
PDF	+3.1/-2.5 %	$\pm$8%	$\pm$11%	-
QCD Scale	+4.8/-4.3 %	+11/-14 %	$\pm$2.3%	-
Signal Model	$\pm$5.5%	-	-	-
Photon Energy Resolution	+4/-2 %	+4/-2 %	+4/-2 %	-
Photon Energy Scale	+1/-4 %	+1/-4 %	+1/-4 %	-
Photon ID Efficiency	$\pm$2%	$\pm$2%	$\pm$2%	-
Vertex Efficiency	$\pm$0.1%	$\pm$0.1%	$\pm$0.1%	-
HLT	$<$ 0.1%	$<$ 0.1%	$<$ 0.1%	-
JEC	$\pm$1.5%	+3/-5 %	$\pm$8%	-
JER	$\pm$0.5%	$\pm$3%	+8/-0 %	-
b-tagging	$\pm$2%	$\pm$1.5%	$\pm$0.1%	-
PU ID	$\pm$2%	$\pm$0.5%	$\pm$2%	-
Lepton Reconstruction	$\pm$1%	$\pm$1%	$\pm$1%	-
BG shape	-	-	-	33%

Conclusions

- Both analyses face completely different obstacles concerning background modeling
- Subjects were tackled carefully and well accepted solutions were found
- Run-II will change the picture again
- Already doing preparation work, like thorough study of the single top+jets background
- Will have to see what treatment serves our purposes the best

BACKUP

Yields



- Integrated Luminosity of $19.7 fb^{-1}$
- tt̄ dominating background
  - Cross check uses a data-driven tt̄ template

Process	Muon channel	Electron channel
tt̄	1058±5	718±4
Single top	39±3	27±3
Electroweak	17^{+7}_{-5}	11±7
tt̄H	12.87±0.17	9.35±0.15
Total background	1128±9	767±10
tHq, $y_t = -1$	7.54±0.03	5.15±0.02
S/B ratio	0.7%	0.7%

Process	Muon channel	Electron channel
tt̄	29.1±0.8	19.8±0.7
Single top	$1.1^{+0.8}_{-0.6}$	1.2±1.0
Electroweak	4^{+6}_{-4}	5^{+6}_{-4}
tt̄H	1.72±0.06	1.43±0.05
Total background	37^{+6}_{-4}	29^{+7}_{-4}
tHq, $y_t = -1$	0.835±0.010	0.580±0.009
S/B ratio	2.3%	2.0%

*Only statistical errors are shown in these tables

Impact of systematic source - H $\rightarrow b\bar{b}$
MC approach

Source	Type	impact as exclusive source on final limit [%]	improvement of final limit after removal [%]
JES	shape	17	3
JER	shape	< 1	< 1
BTag light flavor	shape	13	< 1
BTag heavy flavor	shape	17	< 1
Pile up	normalization	< 1	< 1
Unclustered energy	shape	3	1
Lepton efficiency	normalization	5	< 1
Luminosity	normalization	10	< 1
Cross section (PDF)	normalization	8	< 1
Cross section (Scale)	normalization	9	< 1
MC Bin-by-Bin unc.	shape	< 1	< 1
Q^2 scale (tHq + $t\bar{t}$)	shape	20	4
Matching	shape	2	2
Top p_T reweighting	shape	19	2
$t\bar{t}$ HF rates (b)	normalization	13	< 1
$t\bar{t}$ HF rates ($b\bar{b}$)	normalization	15	< 1
$t\bar{t}$ HF rates (c / $c\bar{c}$)	normalization	13	1

Impact of systematic source - H $\rightarrow$ b$\bar{b}$
DD approach

Source	Type	impact as exclusive source on final limit [%]	improvement of final limit after removal [%]
JES	shape	< 1	< 1
JER	shape	< 1	< 1
BTag light flavor (MC)	shape	< 1	< 1
BTag heavy flavor (MC)	shape	6	3
Pile up	normalization	< 1	< 1
Unclustered energy	shape	< 1	< 1
Lepton efficiency	normalization	< 1	< 1
Luminosity	normalization	< 1	< 1
Cross section (PDF)	normalization	< 1	< 1
Cross section (Scale)	normalization	< 1	< 1
Q^2 scale (tHq)	shape	< 1	< 1
MC Bin-by-Bin unc.	shape	< 1	< 1
BTag light flavor (DD $t\bar{t}$)	shape	4	< 1
BTag heavy flavor (DD $t\bar{t}$)	shape	< 1	< 1
$t\bar{t}$ contamination	shape	9	16
Method bias	shape	9	3
Scale	shape	< 1	< 1
$t\bar{t}$ HF rates (b)	shape	12	3
$t\bar{t}$ HF rates ($b\bar{b}$)	shape	15	5
$t\bar{t}$ HF rates (c / $c\bar{c}$)	shape	< 1	< 1