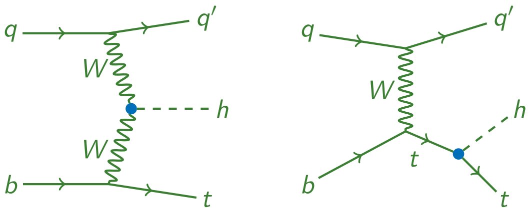
Search for Higgs boson production in association with single top quark in $h \to b\bar{b}$ decay channel

A. Giammanco, M. Komm, **A. Popov** (Louvain-la-Neuve)
C. Böser, T. Chwalek, B. Maier, S. Fink, J. Wagner–Kuhr (KIT)
K. Boom, D. Knowlton, R. Gonzalez Suarez (Nebraska)
A. Aize, J. Komaragiri–Shroff (Malaya), A. Bean, D. Noonan (Kansas)
A. De Cosa, D. Pinna (Zürich)

Scalar Search and Study in Belgium
24 Jan 2014

Associated *th* production

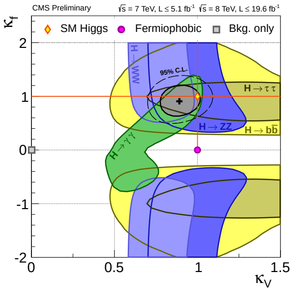
- There are two dominant diagrams for *thq* production in SM:

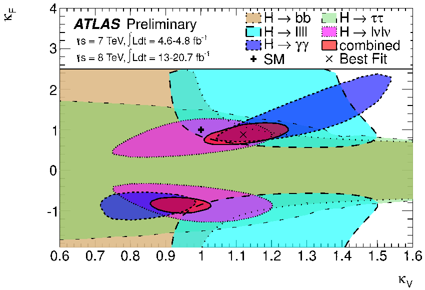


  - Destructive interference in SM ⇒ cross section 18.3 fb
  - With an inversed sign of Yukawa coupling $\kappa_t = -1$ the interference is constructive, $\sigma = 234$ fb ($\times 13$ enhancement)
- Phenomenological studies:
  - M. Farina, C. Grojean, F. Maltoni, E. Salvioni, A. Thamm, JHEP 1305 (2013) 022, arXiv:1211.3736
  - S. Biswas, E. Gabrielli, B. Mele, JHEP 1301 (2013) 088, arXiv:1211.0499

Constraints on Higgs coupings from LHC

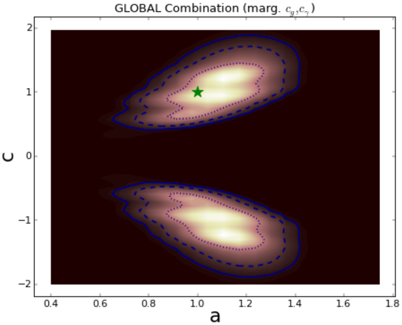
- Interest to *thq* process was originally motivated by presence of two allowed regions in the plane of higgs couplings to bosons and fermions
  - $h \to \gamma\gamma$ is the only channel sensitive to the sign of κ_f
  - However, in the Moriond'13 update (below) CMS nearly ruled out possibility of $\kappa_f = -1$; ATLAS still allowed it





Higgs couplings with BSM contribution allowed

- But if BSM contributions to $h\gamma\gamma$ or hgg (loop-induced) couplings are allowed, $\kappa_f = -1$ is still tolerated
  - J. Ellis, T. You, JHEP 06 (2013) 103, arXiv:1303.3879
  - Analysis is based on combination of CMS, ATLAS, and Tevatron measurements
  - The Fig. shows constraints while marginalising over possible BSM contributions to $h\gamma\gamma$ or hgg (meaning of factors a, c is similar to κ_V, κ_f resp.)

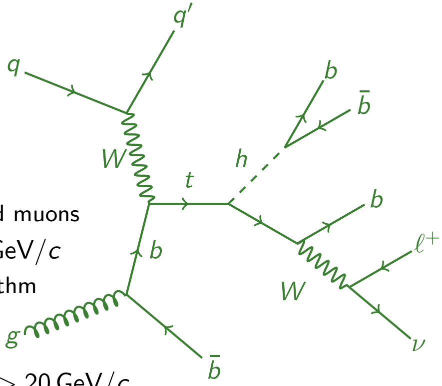


Analysis strategy

- The ultimate goal is to derive constraints on higgs couplings in the (κ_t, κ_W) plane
- For the first result we concentrate on the $\kappa_t = -1$ case
- Focus on decay $h \to b\bar{b}$ for its large branching
  - There is a complementary analysis of th, $h \to \gamma\gamma$, but it will not be addressed in the talk
- Exploit full 8 TeV data
  - In future will profit from Run II: larger cross section, a bit better S/B ratio, larger int. luminosity

Baseline event selection

- Decay channel:
  - $t \to b\ell\nu$, $\ell = e, \mu$ ($\mathcal{B} \approx 0.22$)
  - $h \to b\bar{b}$ ($\mathcal{B} \approx 0.58$)
- Single-lepton triggers
- A tight muon (electron) with $p_T > 26\ (30)\,\mathrm{GeV}/c$
  - Veto additional loose electrons and muons
- $3 \div 4$ b-tagged jets with $p_T > 20\,\mathrm{GeV}/c$
  - Tight working point of CSV algorithm
  - $3t$ and $4t$ regions considered independently
- At least one untagged jet with $p_T > 20\,\mathrm{GeV}/c$
- Additionally, $p_T(j_4) > 30\,\mathrm{GeV}/c$
  - To further suppress QCD and $W+\mathrm{jets}$



Expected event yield

Expected number of events at 19 fb^{-1} in $e+\mu$ channel:

Process	**3t region**		**4t region**	
	Expected	**MC stat.**	**Expected**	**MC stat.**
thq, $\kappa_f = -1$	17	15.8 k	1.5	1.5 k
$t\bar{t}h$	21	10.1 k	2.2	1.1 k
$t\bar{t}$	1940	32.7 k	25.1	460
Single top, t-ch.	23	280	0.3	5
Single top, tW-ch.	45	121	—	0
W + jets	15	46	—	0
Total background	2040	—	27.6	—
S/B	0.8%	—	5.4%	—

Modelling of $t\bar{t}$ background

- (Semileptonic) $t\bar{t}$ is by far the dominant background
  - Reliable and precise modelling of it is essential
- A data-driven approach is adopted as an alternative to MC
  - The data template is taken from $t\bar{t}$ control region with two b-tagged jets
  - 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

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

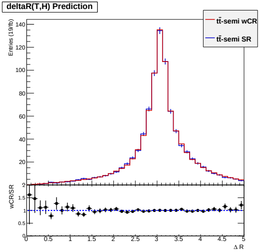
    where ϵ is b-tagging efficiency and the sum is taken over all $\binom{n}{m}$ ways to choose m tagged jets
- The weight is calculated as

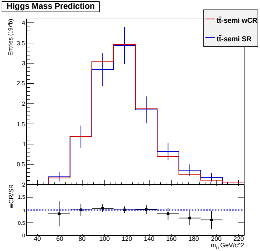
$$w = \mathcal{P}_3/\mathcal{P}_2 \quad \text{or} \quad \mathcal{P}_4/\mathcal{P}_2$$

  in $3t$ or $4t$ region resp.

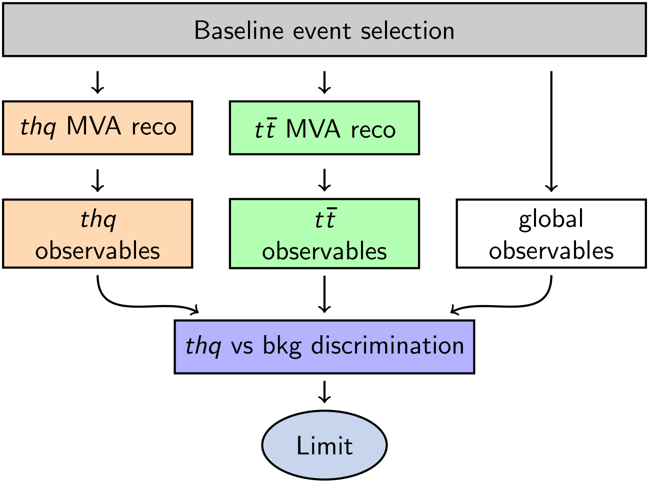
Modelling of $t\bar{t}$ background

- The method is designed to reproduce both shape and normalisation
- Verified with a closure test in simulation
  - Normalisation reproduced within statistical uncertainties in both 3t and 4t regions
  - Good modelling of shape (Fig.)
- A drawback is that the method does not allow to use values of b-tags in subsequent analysis
  - These variables discriminate thq from $t\bar{t}$ relatively well $\Rightarrow$ would like to exploit them
  - The method still allows to use boolean information: whether a jet is b-tagged or not
- We exploit MC simulation in parallel with the data-driven approach
  - Can make use of values of b-tags
  - Cross-check the data-driven method

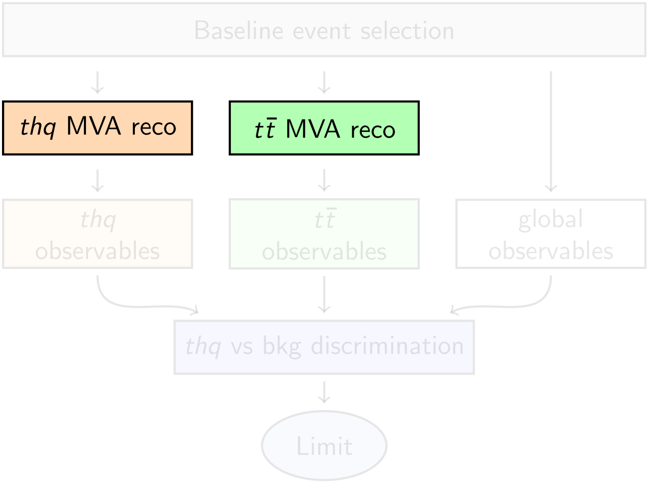




Analysis scheme

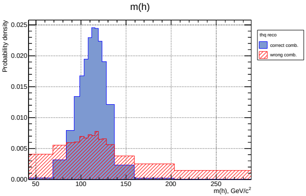


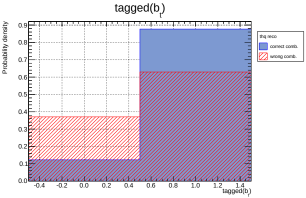
Jet assignment



Jet assignment under signal hypothesis

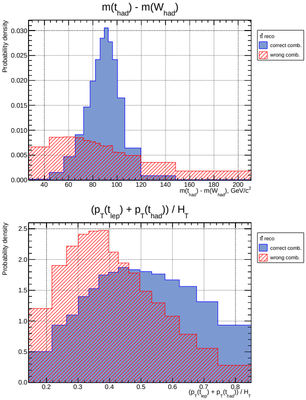
- Consider all the ways to match four reco jets to quarks in $thq \to \ell\nu 3bq$
- Train an MVA to find the correct match
  - MVA is used in classification mode
  - Only events in which the correct match is present, are considered
    - It might happen that a quark does not give birth to reco jet
  - If at least one jet is mismatched, the match is a background
  - In this talk BFGS NNs from TMVA are used
- Jet assignment with trained MVA
  - In each event consider all possible matches
  - Choose the one with largest MVA response



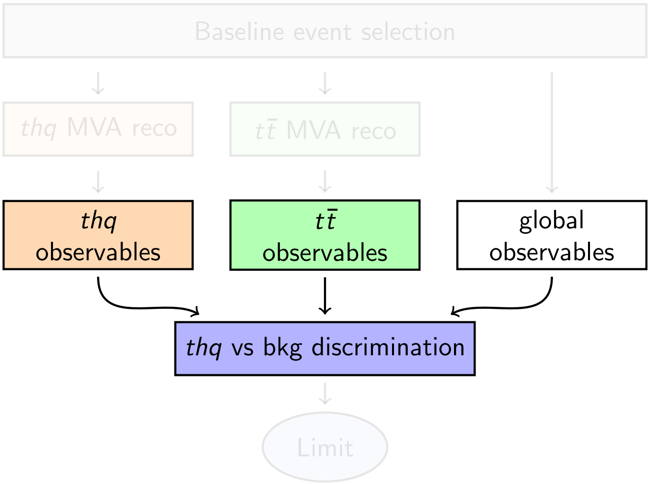


Jet assignment under $t\bar{t}$ hypothesis

- Because of huge amount of $t\bar{t}$, it is important to exploit observables intrinsic to $t\bar{t}$ events
  - Hence the dedicated reconstruction
- The jet assignment is performed in the same way as for signal:
  - Train an MVA to choose the correct match
  - Identify the jets according to the match with largest MVA response
  - Only semileptonic $t\bar{t}$ is considered

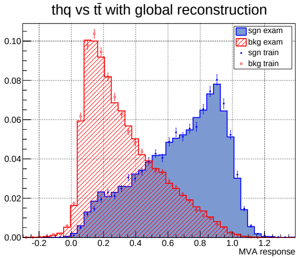


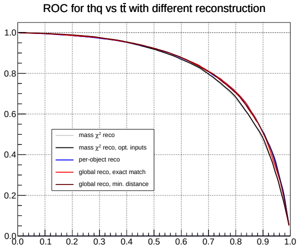
Signal extraction



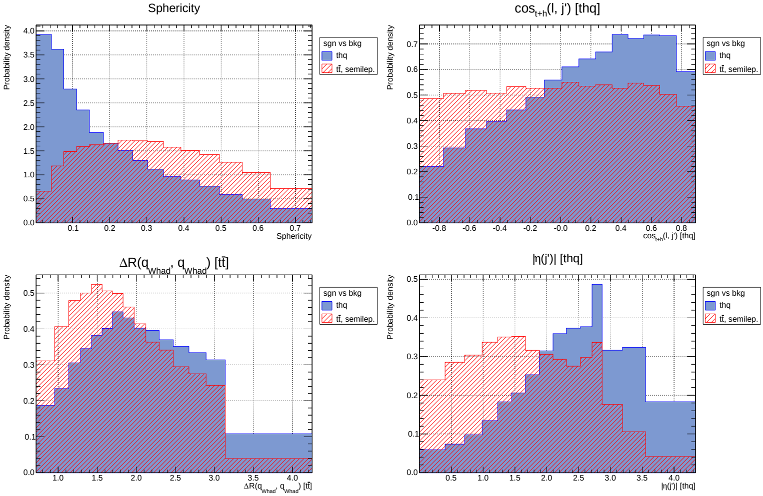
Signal extraction

- Signal is discriminated from backgrounds with a dedicated MVA
- Three sets of input variables are used (examples in next slide)
  - Defined under *thq* hypothesis
  - Defined under $t\bar{t}$ hypothesis
  - Variables that do not rely on jet assignment



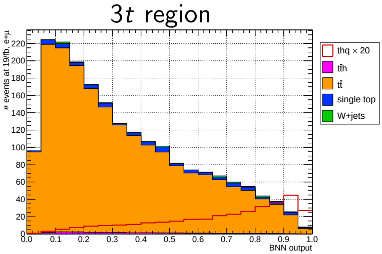


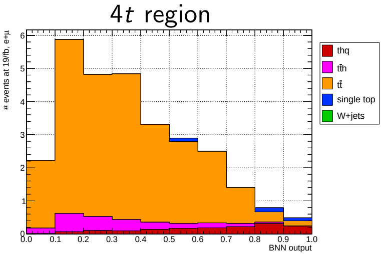
Signal extraction



Signal extraction

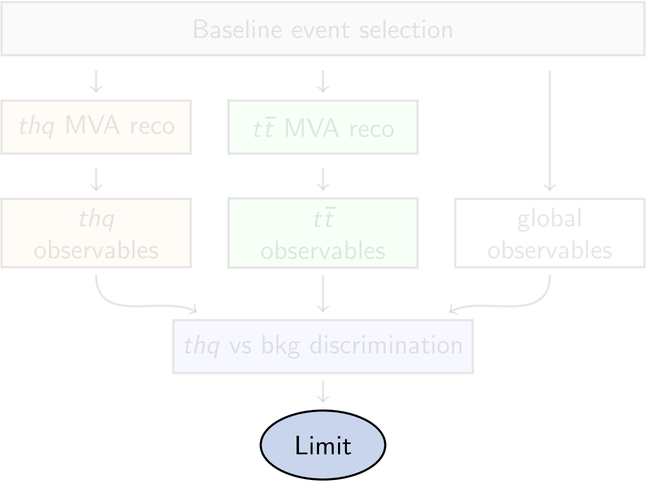
- Expected distributions over MVA output with the whole 8 TeV data in $e + \mu$ channels:





- Normalisation of signal process is deduced from distributions over MVA output

Limit calculation



Limit calculation

- At the moment the signal region is kept blinded
- We evaluate expected CL_s limits $(\kappa_t = -1)$
  - Use the `theta` package for prototyping, will obtain final results with `combine`
  - Nuisance parameters are fitted to data
- Included systematical variations:
  - uncertainties of cross sections,
  - b-tagging (tag and mistag rate),
  - JEC and JER,
  - unclustered $\not{E}_{\mathrm{T}}$,
  - pile-up

Region	$\sigma_{95\%}/\sigma_{\mathrm{exp}}$
$3t$	$3.9^{+1.9}_{-1.1}$
$4t$	$5.5^{+2.3}_{-1.5}$
$3t+4t$	$3.0^{+1.5}_{-0.7}$

- Variations currently missing:
  - factorisation/renormalisation scale (expected to be significant),
  - ME/PS matching scale for $t\bar{t}$,
  - dedicated uncertainties of the data-driven method,
  - scale factors for triggers and lepton IDs, luminosity, ...

Summary and status

- Associated *th* production provides a handle to access additional information on higgs couplings
- Analysis scheme is proven to work, all key components validated
  - Baseline event selection
  - Data-driven modelling of the dominant $t\bar{t}$ background
  - MVA approach to jet assignment
  - Signal extraction with an MVA
- Analysis is being reloaded after the Grid step has been redone
  - Re-reco of data, new datasets (full sim. of *thq*, systematics for $t\bar{t}$), additional information (jet charge, pull angle), bug fixes
- Pending issues:
  - Careful check of data modelling in control region(s)
  - Incorporation of missing systematics
  - Documentation
- We aim at Moriond'14

BACKUP

Cross sections

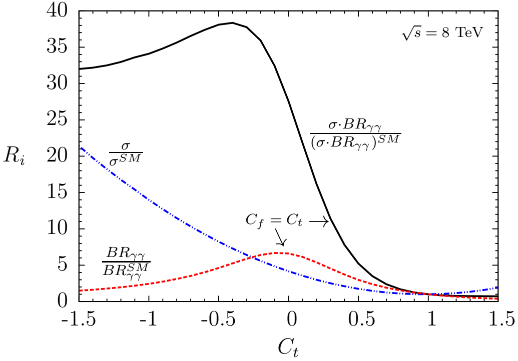
- Cross section is challengely small
  - The main background is $t\bar{t}$; its cross section is provided for comparison

Cross-section	8 TeV	14 TeV
thq, $\kappa_t = +1$ (SM)	18.3 ± 0.4 fb	$88.2^{+1.7}_{-0.0}$ fb
thq, $\kappa_t = -1$	$233.8^{+4.6}_{-0.0}$ fb	980^{+30}_{-0} fb
$t\bar{t}$	245^{+9}_{-10} pb	950^{+40}_{-30} pb

thq cross sections are cited according to M. Farina et al., JHEP 1305 (2013) 022 [arXiv:1211.3736]. Cross-sections for $t\bar{t}$ are calculated in M. Czakon, P. Fiedler, Phys. Rev. Lett. 110 (2013) 252004 [arXiv:1303.6254]. Uncertainties are combined following R. Barlow, arXiv:physics/0306138

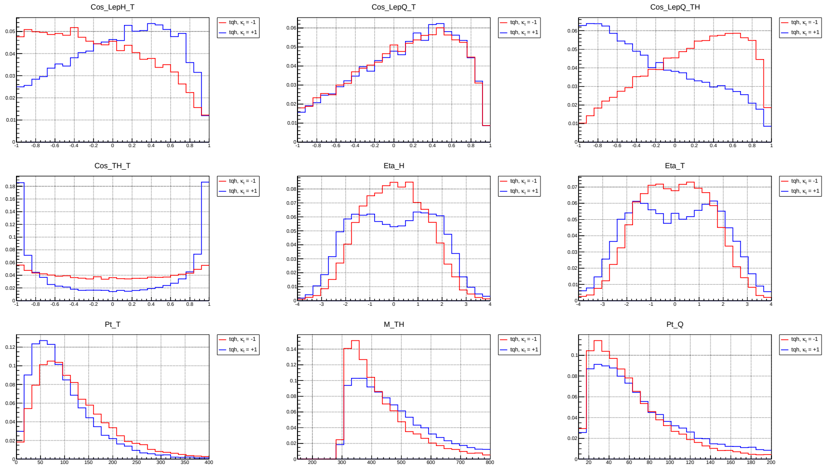
Sensitivity of $\mathcal{B}(h \to \gamma\gamma)$ to κ_t

- Dependence of $\mathcal{B}(h \to \gamma\gamma)$ on κ_t is moderate
  - Note that the fermiophobic case $\kappa_t \sim 0$, in which the branching ratio is maximal, is well excluded by LHC data

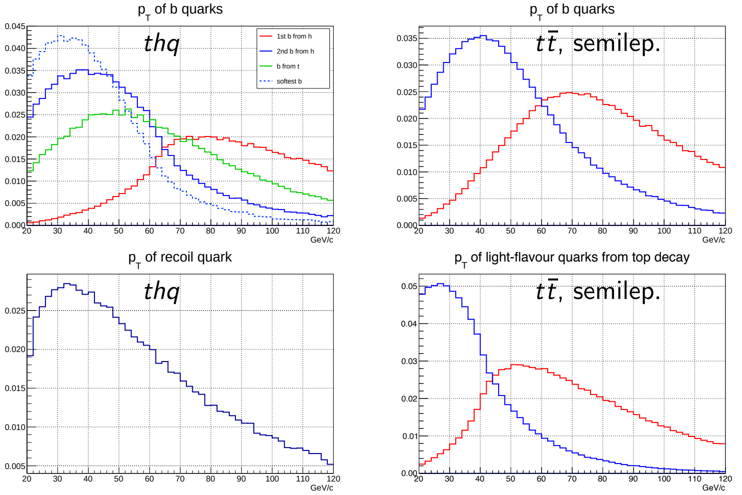


S. Biswas et al., JHEP 1301 (2013) 088, arXiv:1211.0499

Generator-level comparison of $\kappa_t = \pm 1$ cases

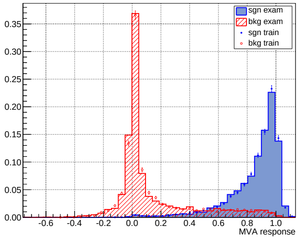


Transverse momenta of jets matched to partons

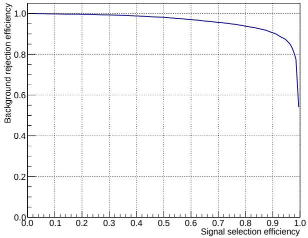


Individual MVAs in global reconstruction

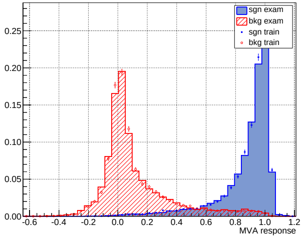
Global reconstruction of thq events



Global reconstruction of thq events



Global reconstruction of semilep. $t\bar{t}$ events



Global reconstruction of semilep. $t\bar{t}$ events