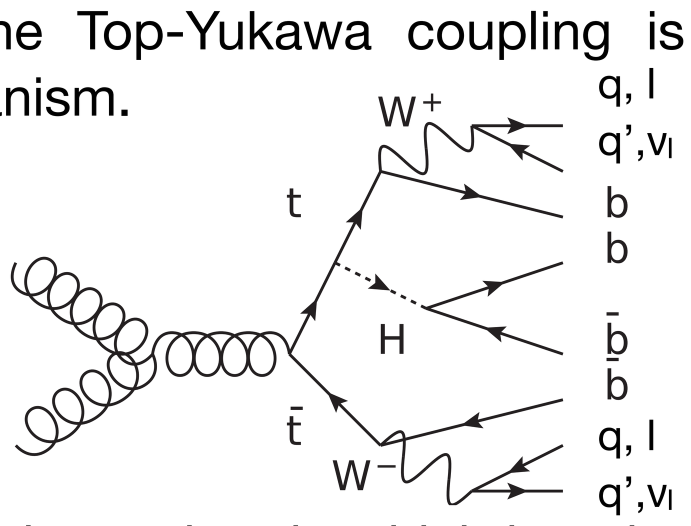
Search for $t\bar{t}H(b\bar{b})$ in the all hadronic channel at CMS

L. Caminada, F. Canelli, S. Donato, D. Salerno, K. Schweiger *

$t\bar{t}H$, H → $b\bar{b}$ in a nutshell

The production of the Higgs boson in association with a top-quark pair is one of the main production modes of the Higgs boson in the standard model of particle physics (SM), which has not been experimentally observed and it also is the only channel that offers the possibility to directly measure the Top-Yukawa coupling. Within the Higgs mechanism the Yukawa coupling describes how fermions get mass. Since the coupling strength is predicted to be proportional to the fermion mass and the top quark is the heaviest fermion, the Top-Yukawa coupling is fundamental to the SM and the Higgs mechanism.



Challenges of the search for $t\bar{t}H$:

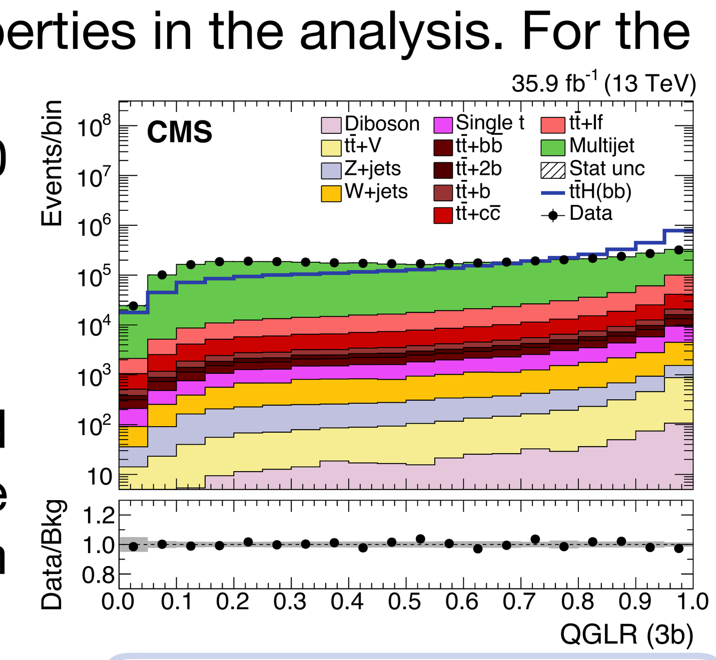
- rare, roughly once in 5 billion collisions
- many objects (jets, leptons, neutrinos) in the final signature
- many background events mimicking the signal

We study the decay of the Higgs boson to a b-quark pair, which has the largest branching faction (58%) of all Higgs-boson decays. The final signature is defined by both W bosons decaying into quarks — referred to as the **all hadronic (AH) channel**. It represents 46% of all $t\bar{t}H(b\bar{b})$ events and has a expected signature of 8 jets including 4 b-tagged jets. Unfortunately, the AH channel suffers from a large QCD multijet background and a large ambiguity of assigning the correct jets to the original quarks.

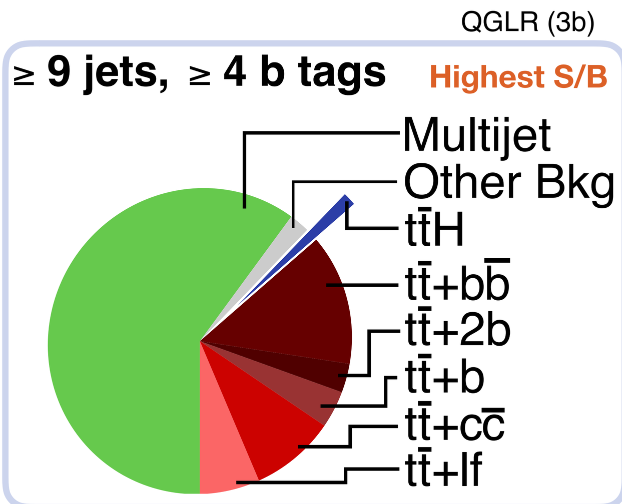
Event Selection

Other SM processes with similar signatures form backgrounds to the $t\bar{t}H$ signal. Therefore, events are selected to increase signal purity. First during data taking using hardware and software triggers and then by cutting on event and reconstructed jet properties in the analysis. For the signal region these are:

- ≥ 7 jets with ≥ 3 b-tagged jets (p_T > 30 GeV, 6 jets with p_T > 40 GeV, $|\eta| < 2.4$)
- HT > 500 GeV (p_T sum of all jets)
- No lepton in the final state
- Cut on a Quark-gluon-likelihood discriminant (QGLR) used to discriminate light-quark jets in signal events from gluon jets in multijet events



Events are **categorized by the number of jets and b tags** (7 jets, 3 b tags to ≥ 9 jets, ≥ 4 b tags) to increase the overall sensitivity. Categories with high signal to background ratio (S/B) drive the sensitivity while other categories help to better understand the backgrounds.



Dominant Backgrounds

Because the production rate of all backgrounds is much higher than the signal, finding background events closely emulating the signal is very likely. The main backgrounds after selection are the following.

$t\bar{t}+b\bar{b}$:

The best way to distinguish the production of a $t\bar{t}$ pair with an additional pair of b quarks is by comparing the event properties with sophisticated methods like the Matrix Element Method.

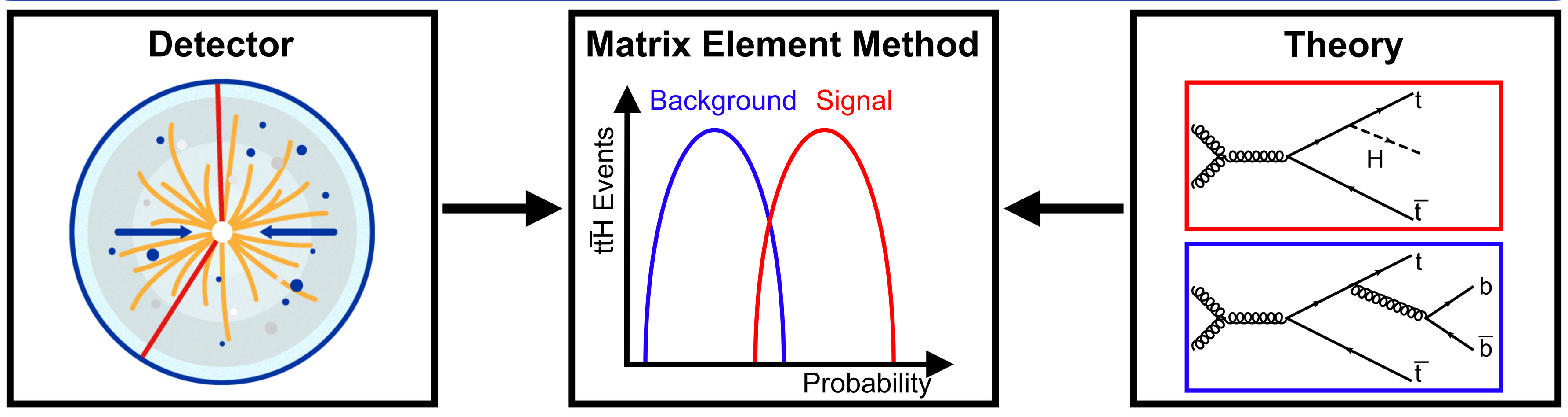
$t\bar{t}$+jets:

Final states with $t\bar{t}$ pairs and additional light jets can pass the event selection because of light jets misidentified as b jets.

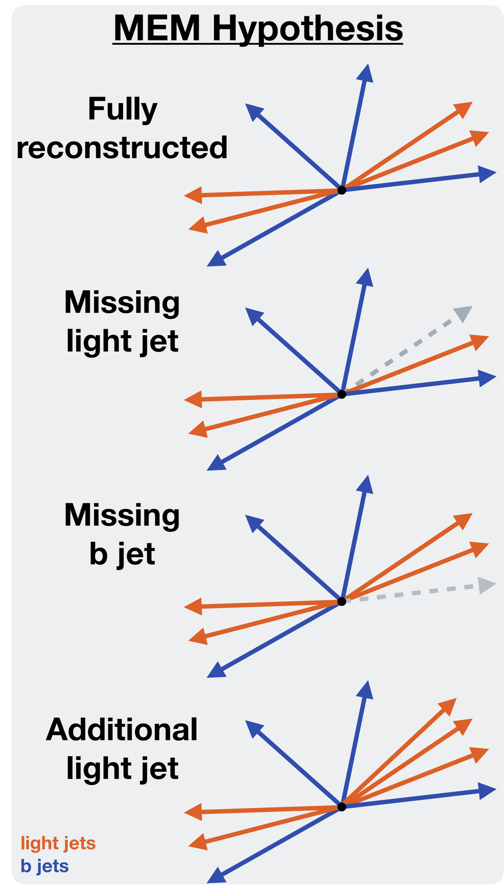
Multijet:

Since no leptons are present in the final state, the signal regions are dominated by QCD multijet background which is **estimated with a data driven approach**. For this, a control region with low signal contribution is defined by requiring fewer b jets. The distribution of a given variable in the signal region is estimated from this control region after applying jet-by-jet corrections using data and simulated $t\bar{t}$ events. The normalization is set by the final fit to data.

Matrix Element Method in $t\bar{t}H(b\bar{b})$



The Matrix Element Method (MEM) is a powerful tool to separate signal and background by using theoretical knowledge from the standard model of particle physics.



The MEM calculates the probability density ω that the final state of the **event originated from either $t\bar{t}H$ or $t\bar{t}+b\bar{b}$**. Furthermore, different hypotheses are used depending on the number of reconstructed light- and b jets.

$$\omega_{s,b} = \sum_{\text{perm}} \int d\Phi_8 \ \text{PDF} \otimes |\mathcal{M}_{s,b}|^2 \otimes W$$

The MEM requires 3 base components:

- The gluon PDF, describing the momentum of the initial state gluons
- The scattering and decay amplitudes $\mathcal{M}_{s,b}$ for the signal and background processes
- The transfer function *W* for taking into account detector effects

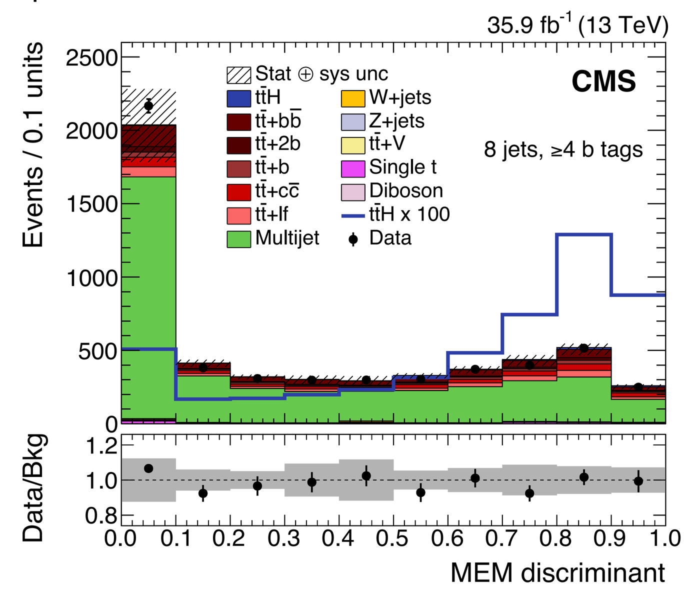
A sum over all possible jet assignments and an integration over the full phase space of the final-state particles is performed.

Analysis Strategy

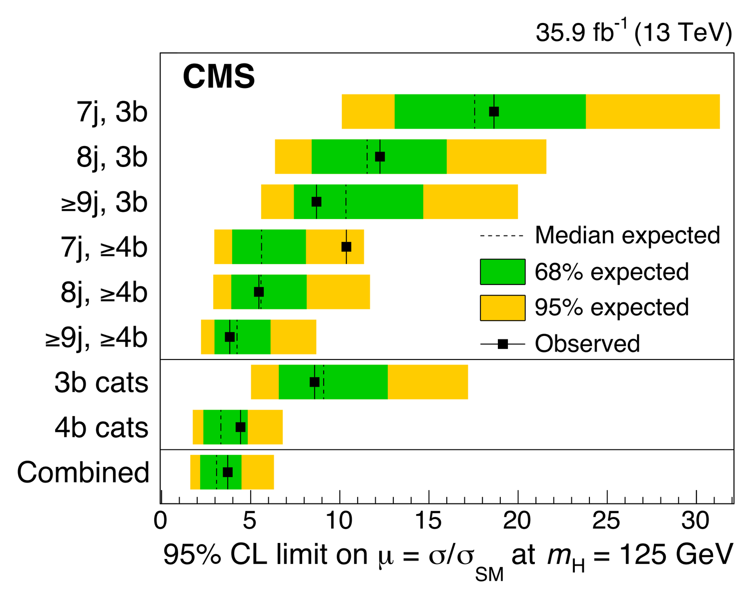
To extract the signal strength modifier μ, defined as σ/σ_{SM} with the production cross section σ and the SM prediction σ_{SM}, a discriminator is formed from the MEM probability densities for $t\bar{t}H$ ω_S and $t\bar{t}+b\bar{b}$ ω_B:

$$P_{sb} = \frac{\omega_S}{\omega_S + \kappa_{sb} \cdot \omega_B}$$

Distribution of the MEM discriminant for $t\bar{t}H$ signal and non-negligible background processes in simulation.



This discriminant provides **optimal separation between signal and background**. Events with a P_{sb} value close 1 are $t\bar{t}H$-like while events close to 0 are $t\bar{t}+b\bar{b}$-like. Since the $t\bar{t}+b\bar{b}$ hypothesis discriminates well against QCD multijet events no dedicated ME for this process is considered.



A simultaneous fit to the MEM discriminator distribution is performed in all six analysis categories under the assumption of a signal+background and background-only hypothesis. Because the sensitivity with the current data, is too low to measure the production cross section directly, an upper limit on μ is calculated. With 35.9 fb^{-1} of data recored with the CMS experiment in 2016 we obtain an observed (expected) upper exclusion limit μ < 3.7 (3.1) and a best-fit μ value of $0.84^{+1.50}_{-1.48}$, which is compatible with the SM expectation.

Future Plans

To improve the separation power of the MEM, adding a representative 2→8 ME process for QCD is under investigation.

The $t\bar{t}H(b\bar{b})$ analyses are systematically dominated, mostly because of the uncertainty on the $t\bar{t}+b\bar{b}$ cross section. A better measurement of this will make a large improvement to the sensitivity. In particular, a dedicated measurement in the AH channel is planned.

The dominant uncertainty of this analysis is related to the QCD multijet background. Methods to better understand this background and to reduce its contribution will be investigated.