Measurement of the ttH, H → bb decay with RunII data and prospects for RunIII

ICPPA-2020

5th International Conference on Particle Physics and Astrophysics

Charis-Kleio Koraka [1]

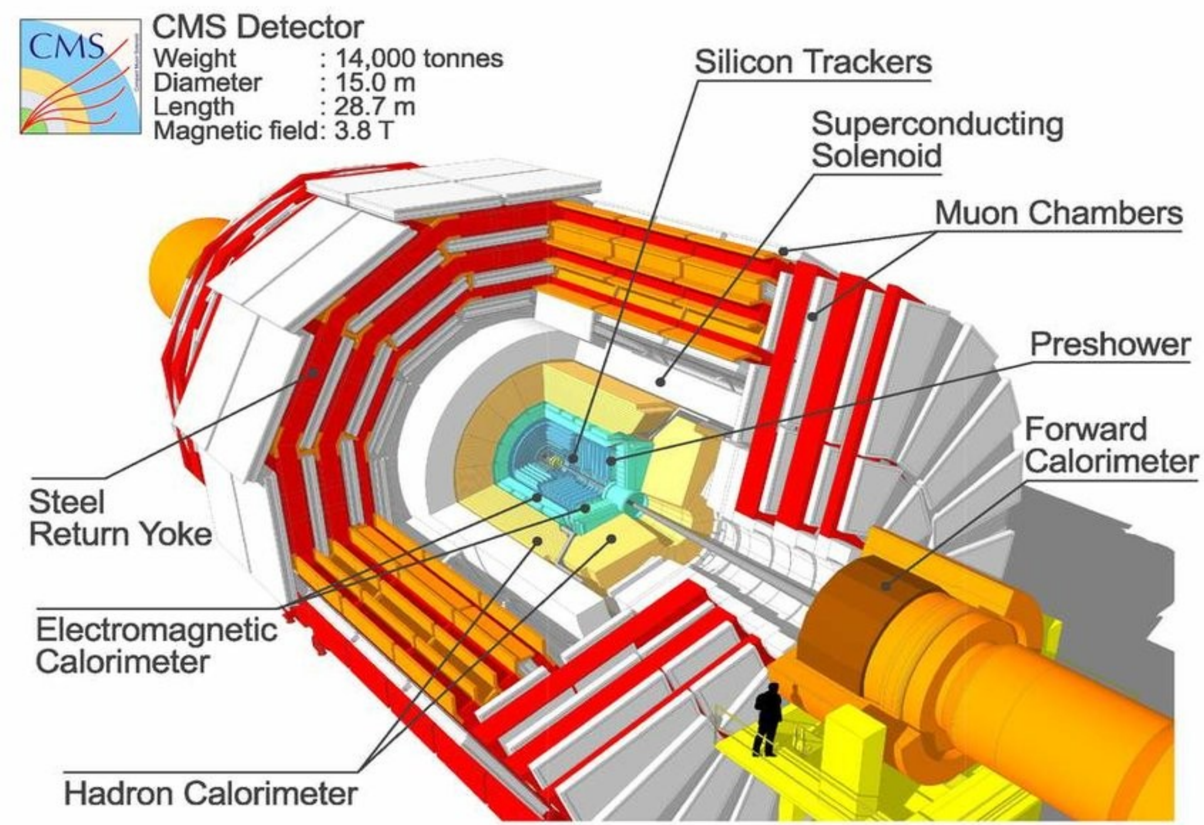
on behalf of the CMS Collaboration

[1] National and Kapodistrian University of Athens, Greece

Outline

- The CMS experiment @ the LHC
- The ttH , H→ bb associated production
- Analysis strategy
- Results for the 2016 & 2017 Data.
- New ideas and future prospects
- Summary

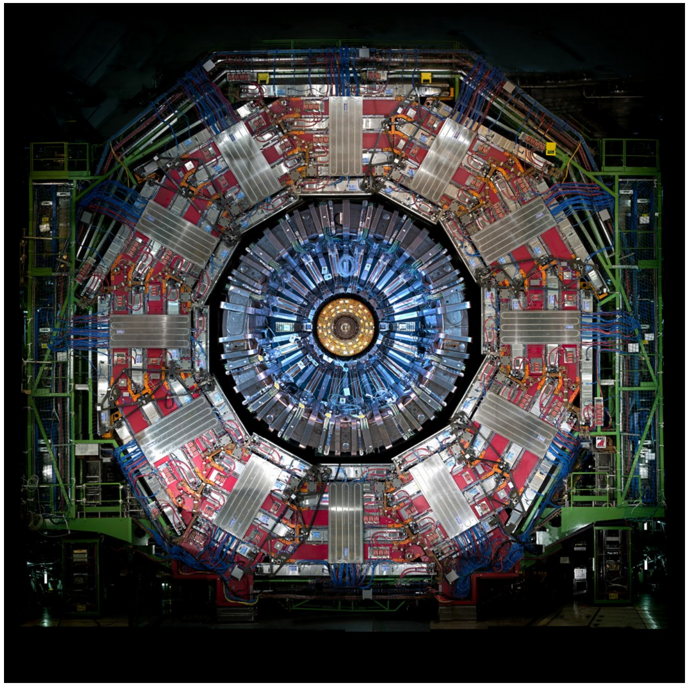
The CMS experiment @ the LHC



General-purpose detector with broad physics program

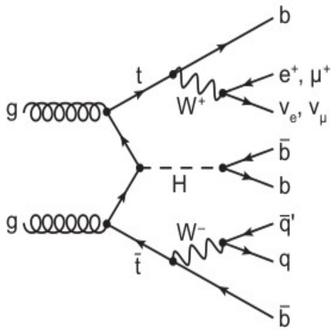
CMS sub-detector systems consist of :

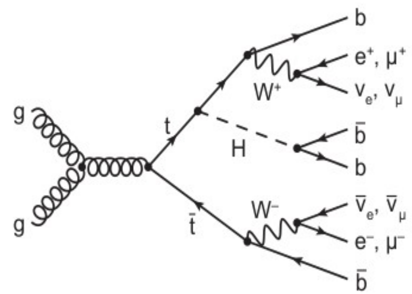
- Inner tracking system
- ECAL
- HCAL
- Magnet
- Muon System

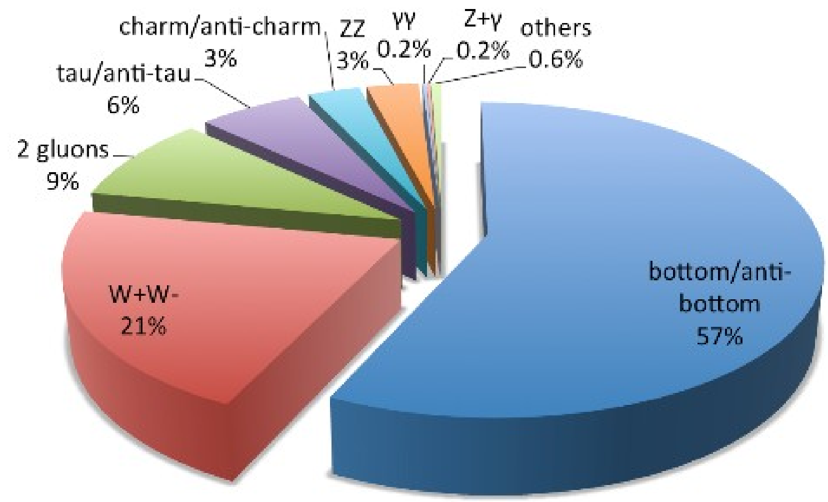


https://cds.cern.ch/record/1474902

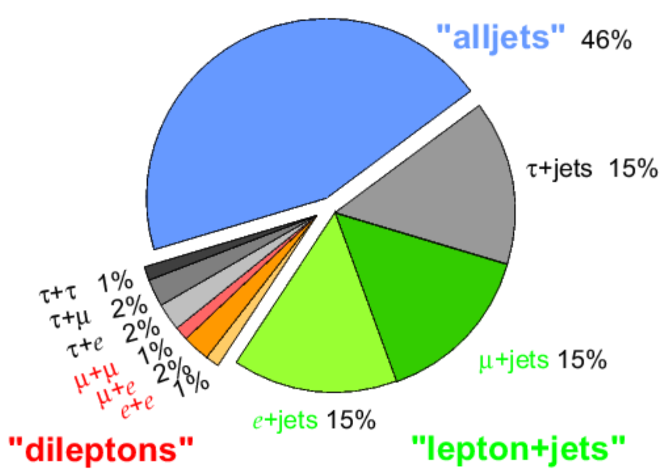
The ttH, H→bb associated production







- Top quark decays ~99 % to a W boson and a b-quark
- We can distinguish the following decay modes :
  - ➔ All hadronic (FH)
  - ➔ Semi-leptonic (SL)
  - ➔ Dileptonic (DL)



- Higgs boson decay to a bb pair is attractive as a final state because it features the **largest branching fraction of 0.58 ± 0.02 for a 125 GeV Higgs boson.**
- The top quark, the heaviest known fermion, is instrumental in testing the SM and constraining models of physics beyond it.
- The ttH associated production provides **a direct probe for measuring the top-Higgs coupling.**

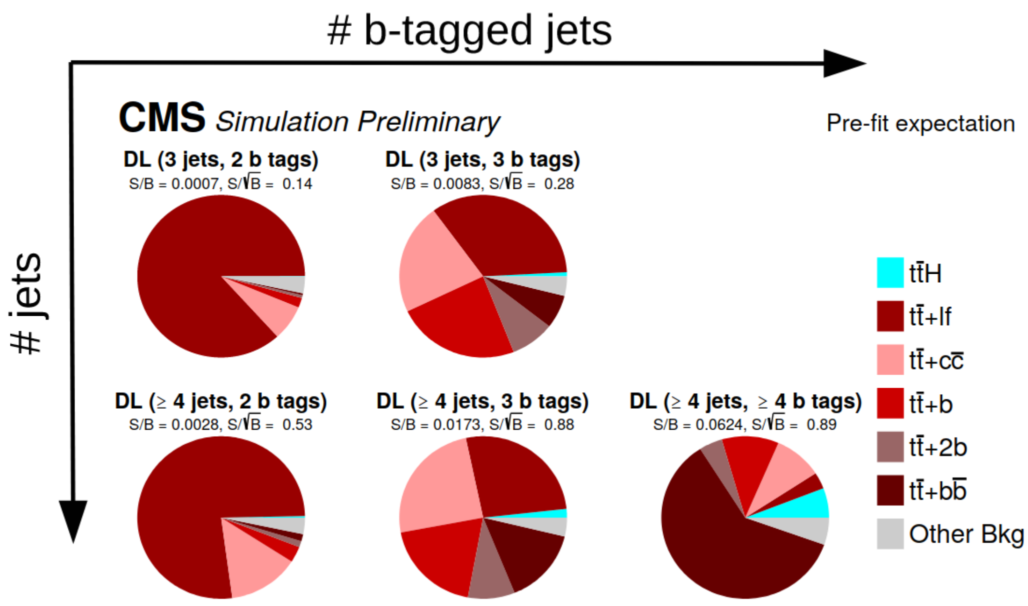
Baseline selection for ttH , H → bb signature

- The decays of the W bosons from the top quarks determine the specific tt signatures recorded in the detector.
- Dedicated event selection criteria and analysis methods are used for each channel.

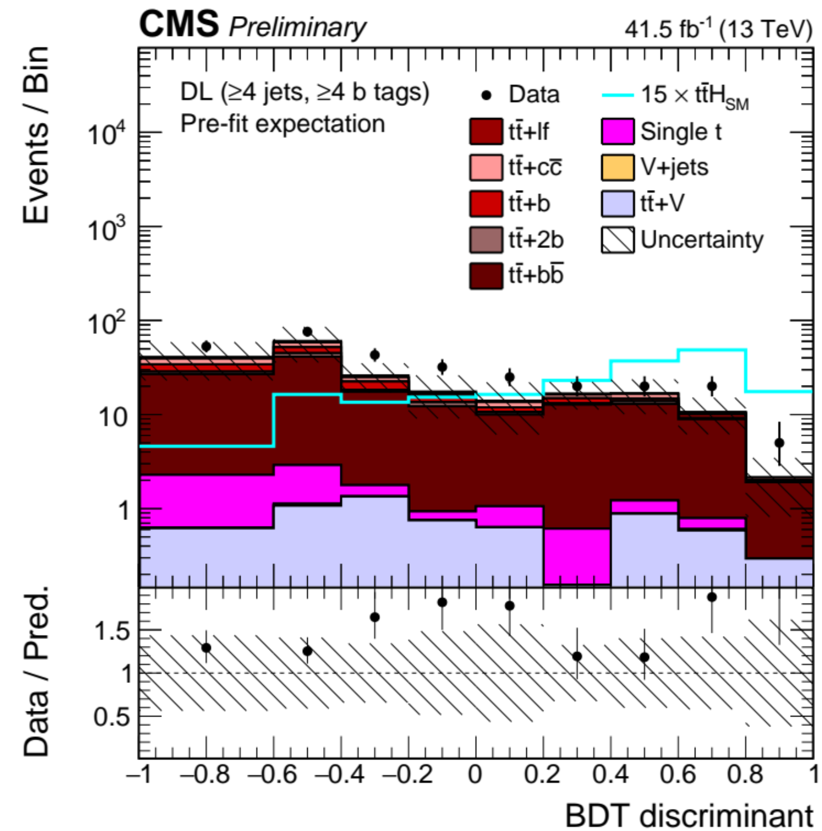
	FH channel	SL channel	DL channel
Number of leptons	0	1	2
p_{T} of leptons (e/μ) [GeV]	—	$> 30/29$	$> 25/25$ GeV
p_{T} of additional leptons [GeV]	< 15	< 15	< 15
$\lvert\eta\rvert$ of leptons	< 2.4	< 2.4	< 2.4
Number of jets	≥ 6	≥ 4	≥ 2
p_{T} of jets [GeV]	> 40	> 30	$> 30, 30, 20$
$\lvert\eta\rvert$ of jets	< 2.4	< 2.4	< 2.4
Number of b-tagged jets	≥ 2	≥ 2	≥ 1
$p_{\mathrm{T}}^{\mathrm{miss}}$	—	> 20 GeV	> 40 GeV

CMS PAS HIG-18-030

- For the FH channel, QCD dominant background contribution.
- For SL and DL channels, dominant background contributions from tt+jets production :
  - tt + light-flavor : one or more of jets misidentified as b-quark jets
  - tt+B and tt+cc : additional b or c quarks from QCD radiation or loop-induced QCD
- Remaining minor contributions from single-top quark, W+jets and Z+jets, tt+W and tt+Z ans di-boson production.

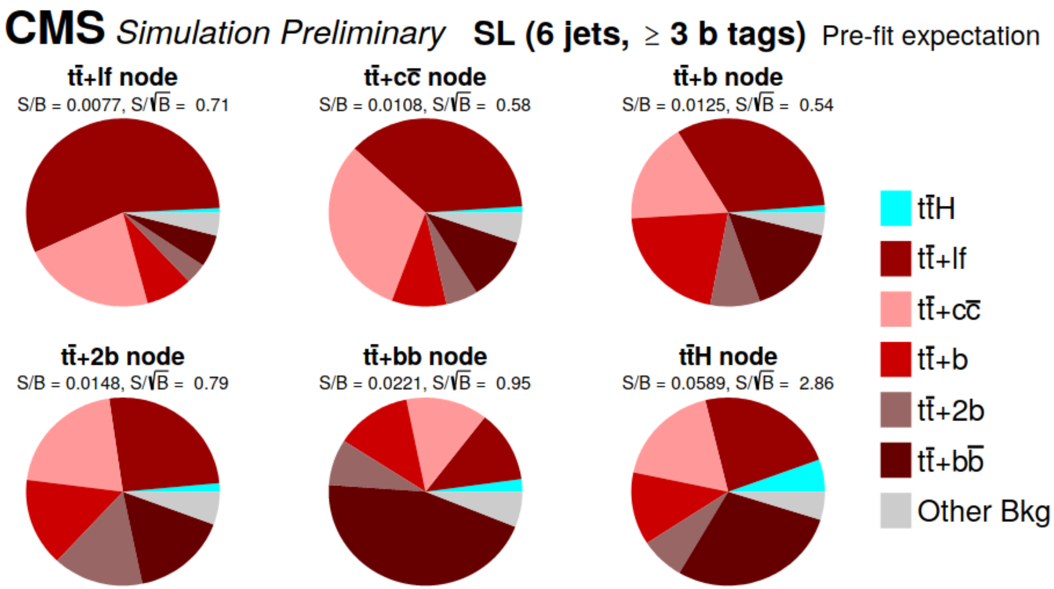


- Selected events are binned in jet and b-jet categories.
- 5 exclusive categories considered.
- **Boosted Decision Trees (BDTs)** used for each category to separate signal from tt+jets background events



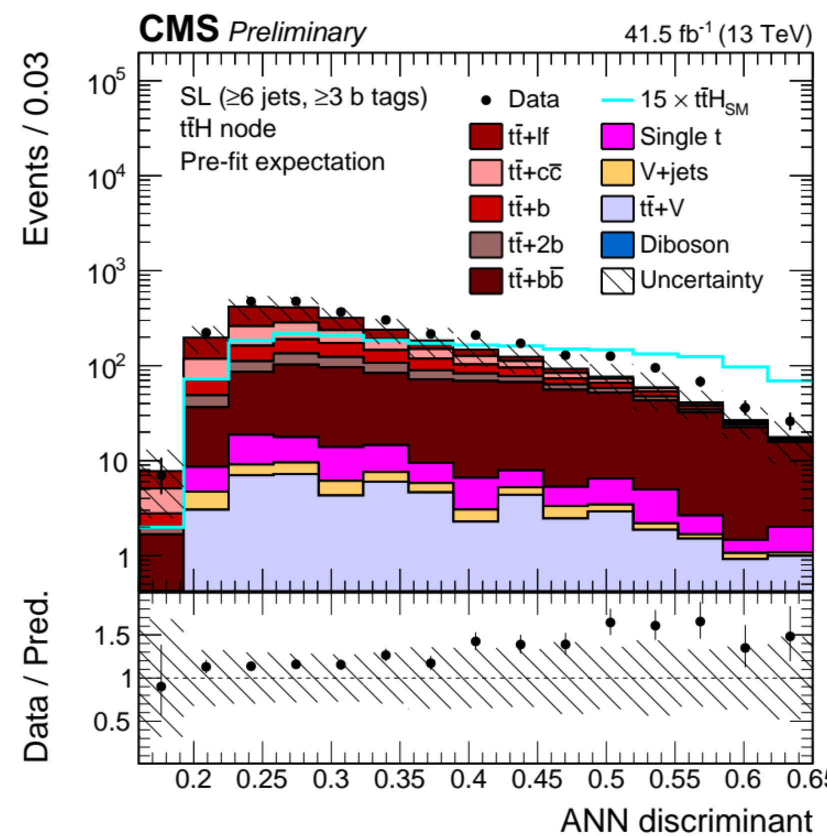
- BDT input variables :
  - Kinematic variables of jets and leptons
  - Invariant masses
  - Angular differences
  - b-tagging information
  - Matrix Element Method discriminant

CMS PAS HIG-18-030

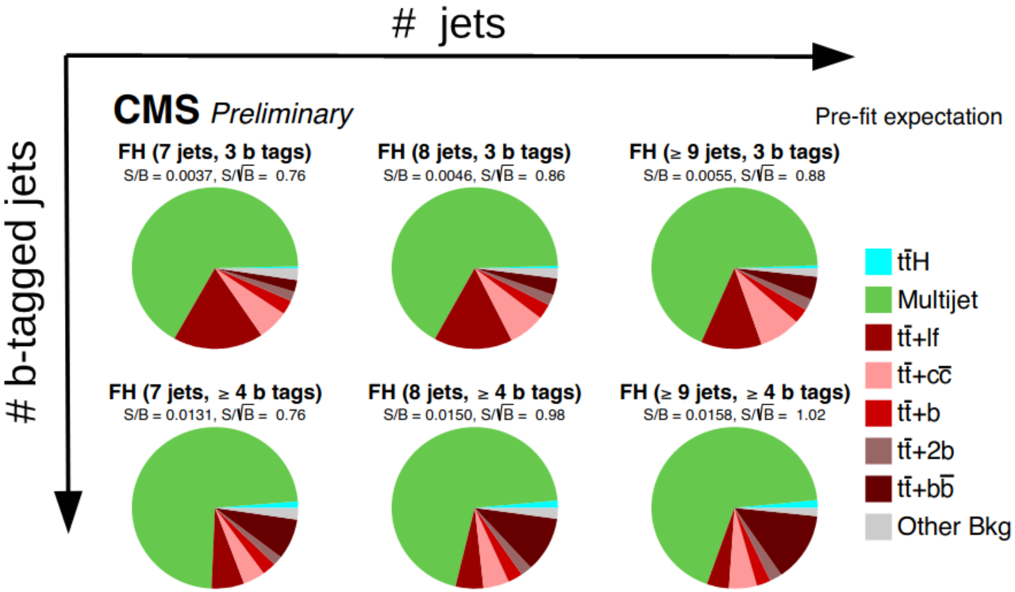


- 18 exclusive categories considered.

- Multi-classification performed using Artificial Neural Networks (ANNs). Event classified either as either signal or any of the five tt +jets background processes :
  - tt +bb
  - tt +2b
  - tt +b
  - tt +cc
  - tt +lf



CMS PAS HIG-18-030



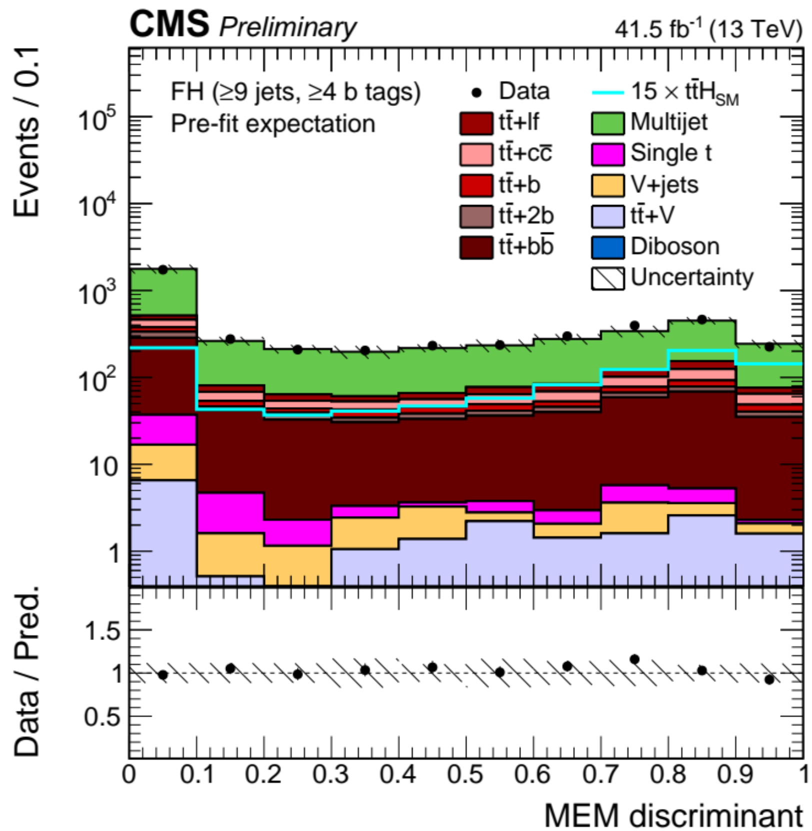
- Further event selection to reduce QCD background:
  - Select events where m_{qq} in W mass window
  - $\Delta\eta$ cut : QCD events larger angular separation
  - Cut on quark-gluon likelihood ratio : QCD has more gluon-initiated jets

- Matrix Element Method (MEM) approach used to construct final discriminant :

$$\frac{f(t\bar{t}H)}{f(t\bar{t}H) + \kappa \cdot f(t\bar{t} + b\bar{b})}$$

- 6 exclusive categories considered.

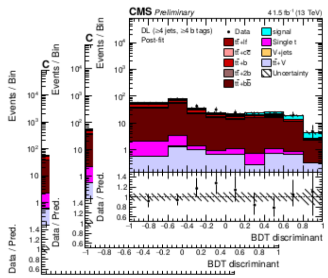
CMS PAS HIG-18-030



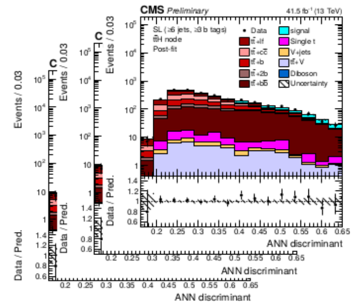
Pre-fit distributions

CMS PAS HIG-18-030

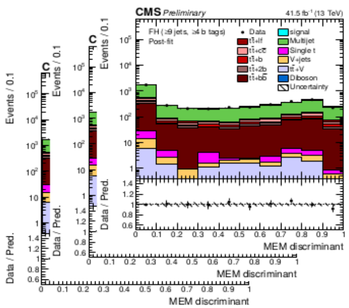
Post-fit distributions



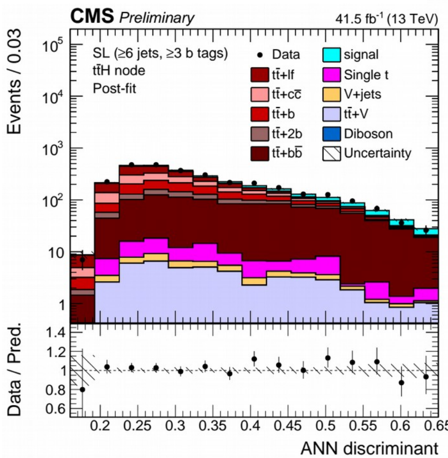
DL : 5 categories

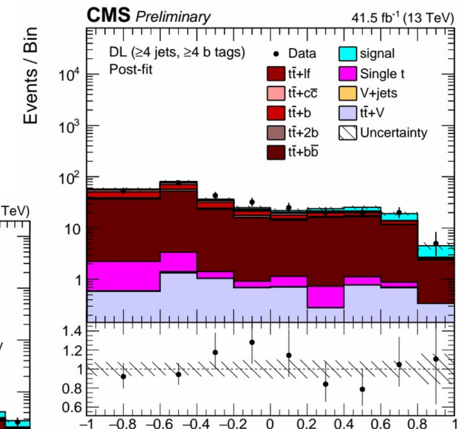


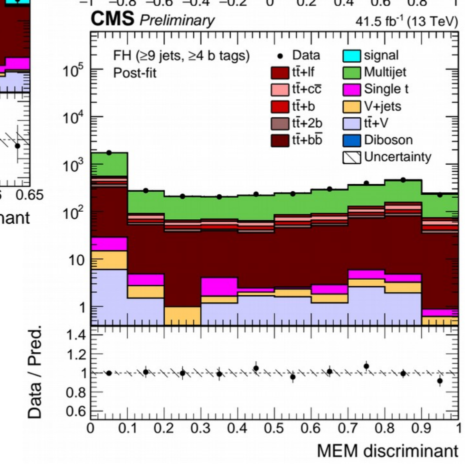
SL : 18 categories



FH : 6 categories

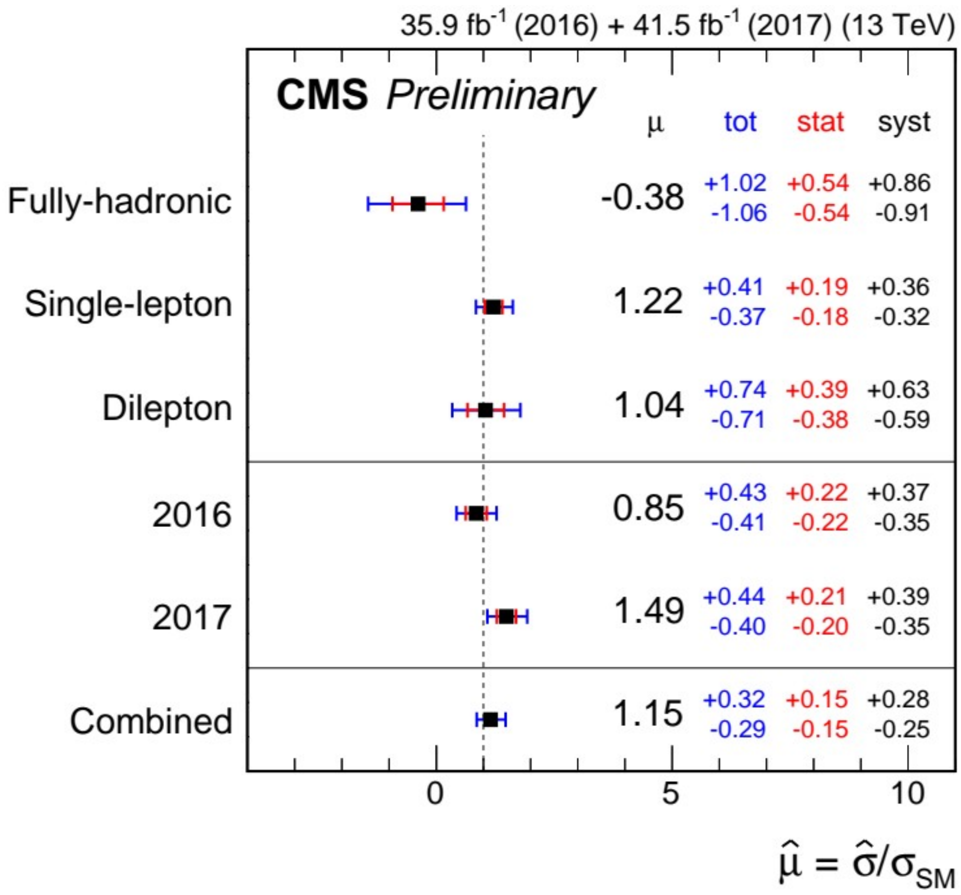






- The signal is extracted in a simultaneous signal plus background fit of all channels and corresponding categories.
- Sources of systematic uncertainties are incorporated as nuisance parameters (constrained or freely floating) in the final fit

2016+2017 Combined Results



CMS PAS HIG-18-030

The resulting **2016+2017** combined signal strength, for all decay channels, is :

$$\boldsymbol{\mu} = 1.15 \quad +0.32 \quad -0.29$$

which corresponds to an observed significance of :

3.9 σ significance (3.5 σ expected)

- **Result is systematics limited:**
  - ➔ Statistical uncertainty (Δμ = + 0.15/ − 0.15)
  - ➔ Systematic uncertainty (Δμ = + 0.28/ − 0.25)
- **Largest systematic uncertainties:**
  - ➔ tt+hf modeling (Δμ = + 0.14/ − 0.15)
  - ➔ simulated sample statistics (Δμ= +0.10/ −0.10)

Future improvements

- **It is of great importance to adopt new approaches, which would help to reduce the the systematic uncertainties described, especially in light of the increased luminosity expected during Run3.**

- **New ideas :**
  - Currently being utilized for the Full RunII results expected soon:
    - Prediction of the shape and rate of different simulated distributions using events with lower b-tag multiplicities in order to reduce MC statistical uncertainties.

  - In light of RunIII:
    - Provide a way to identify the b-jets originating from the Higgs decay and hence use the reconstructed m_{bb} mass as an observable.
    - Data – driven background prediction of the dominant tt+jets background of the DL ttH(bb) associated production, in order to entirely omit the use of simulation.

Summary

- The measurement of the associated ttH production is of great importance, as it is a direct probe for measuring the top-Higgs coupling.

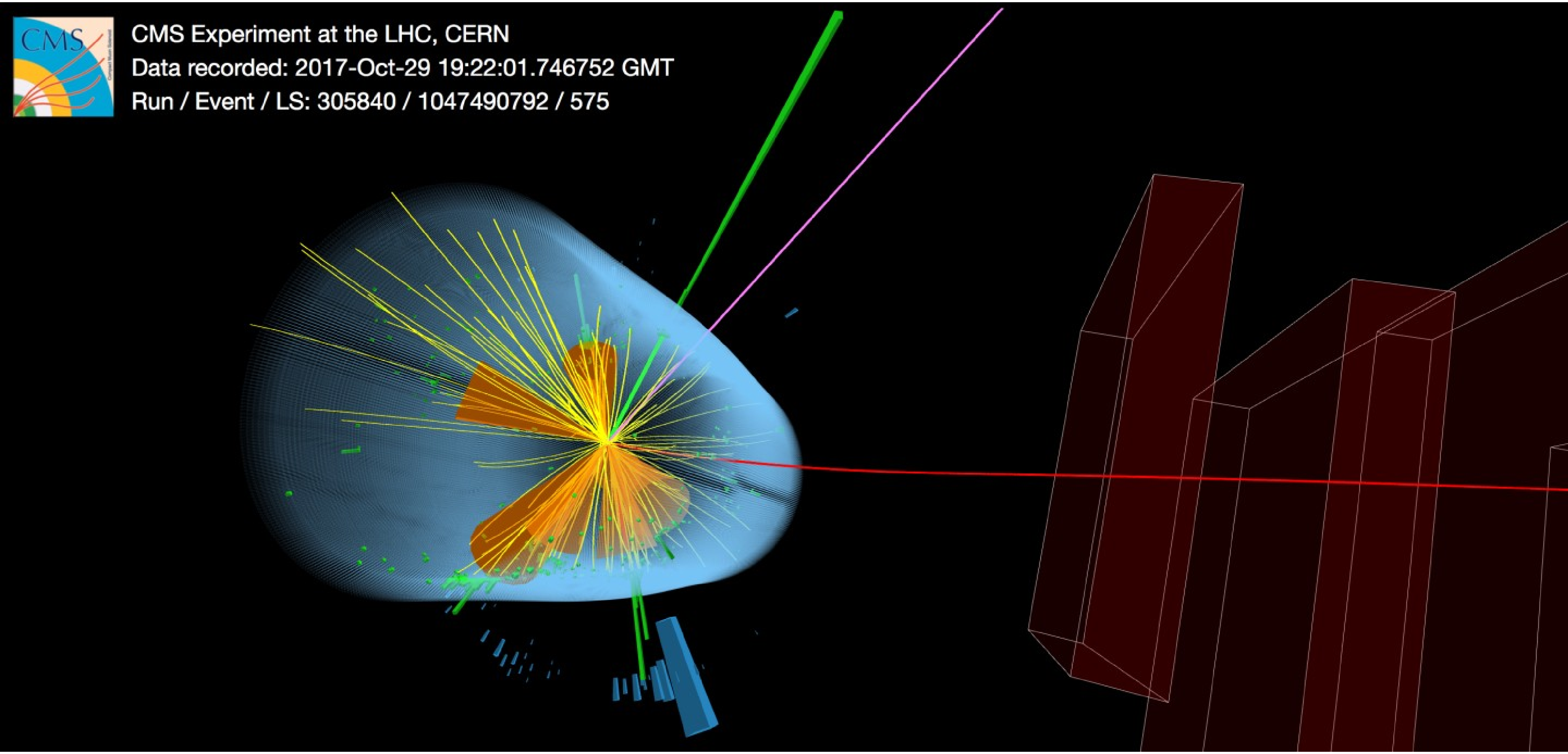
- The resulting signal strength obtained by combining data recorded in 2016 and 2017 correspond to an observed (expected) significance of 3.9 (3.5).

- The current results are systematics limited, with the dominant uncertainty source being this of estimating the tt+bb background

- For the full RunII anticipated results, improved methods and new approaches are being utilize expected to increase the current sensitivity

- In light of Run3 and beyond, we are developing methods to further reduce the uncertainties on the tt+bb simulation and to eventually predict the tt background in a data-driven way.

Thank you for your attention!



https://cds.cern.ch/record/2680325?ln=en

Back-up

References

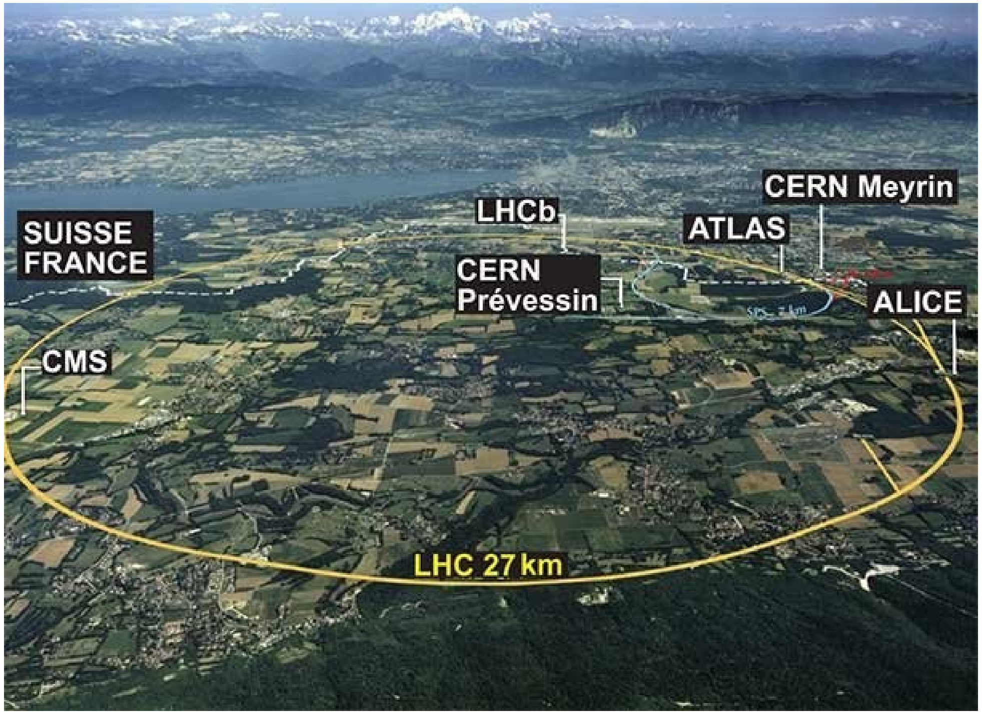
[1] CMS Collaboration, “Search for ttH production in the H → bb decay channel with leptonic tt decays in proton-proton collisions at √s = 13 TeV”, arXiv:1804.03682.
[2] CMS Collaboration Collaboration, “Measurement of ttH production in the H → bb decay channel in 41.5 fb $^{-1}$ of proton-proton collision data at √s = 13 TeV”, Technical Report CMS-PAS-HIG-18-030, CERN, Geneva, 2019.
[3] https://indico.cern.ch/event/807915/attachments/1864710/3073176/2019-06-25_tth-ttbb_CERN-LPC.pdf
[4] CMS Collaboration, “Search for ttH production in the all-jet final state in proton-proton collisions at √s = 13 TeV”, JHEP 06 (2018) 101, doi:10.1007/JHEP06(2018)101,arXiv:1803.06986.

Images:

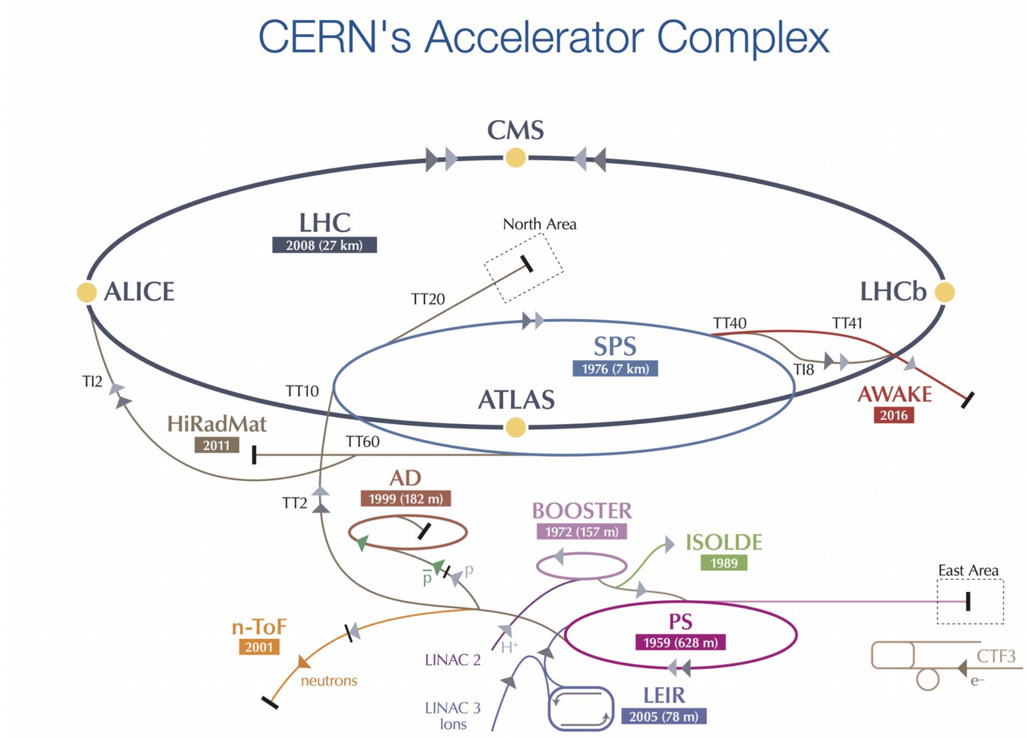
[1] https://cds.cern.ch/record/1474902

[2] https://cds.cern.ch/record/2680325?ln=en

The Large Hadron Collider (LHC)



- The Large Hadron Collider (LHC) in Geneva, Switzerland, is the most powerful collider built to date.
- It consists of two proton rings that run to collide in four different interaction points, including one where the CMS experiment is located.



- During Run-II (2016-2018), the LHC has successfully delivered $\sim$140 fb^{-1} of proton-proton collision data, under harsh experimental conditions, that include of a peak luminosity of 2.1×10^{34} cm^{-2} s^{-1}.

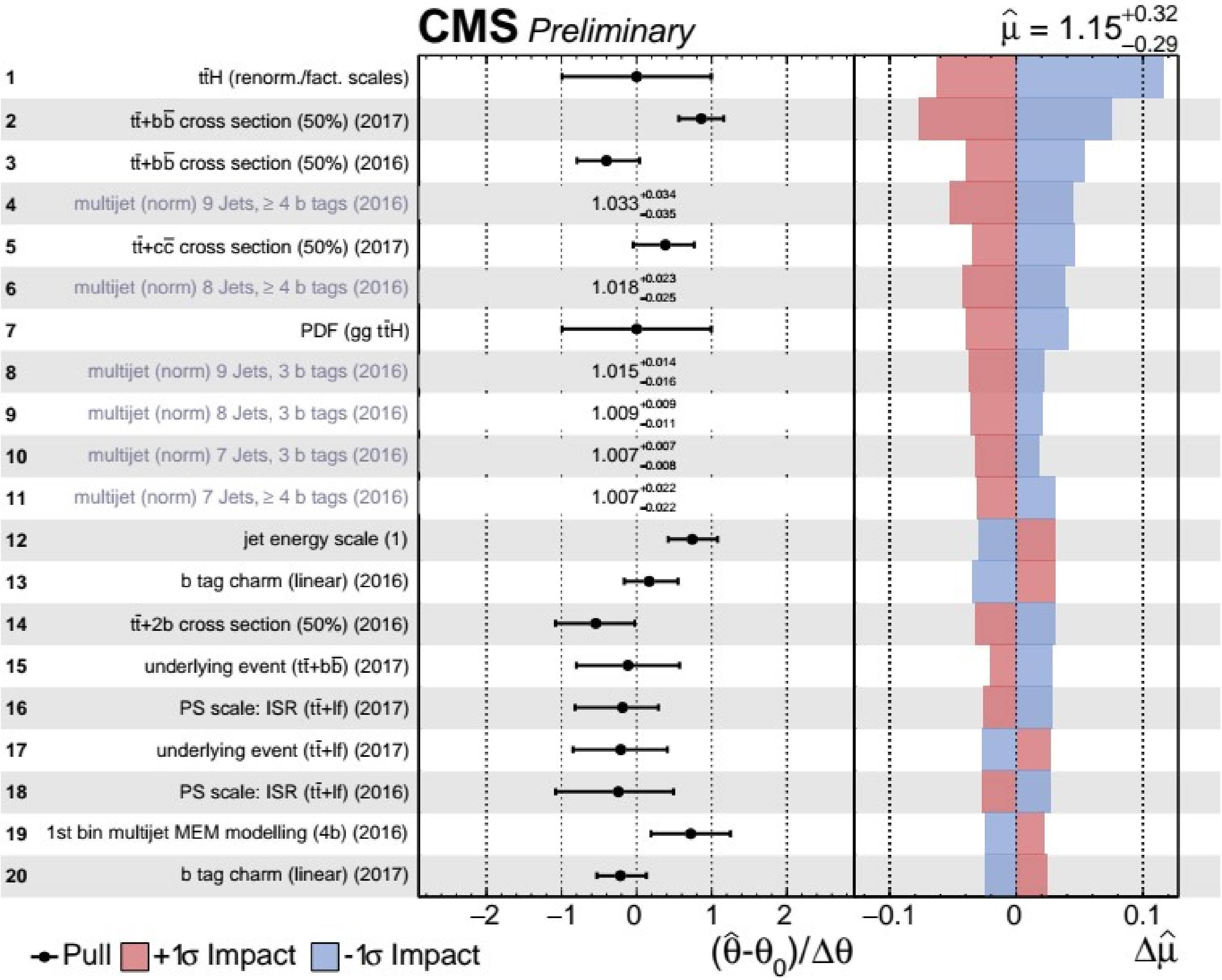
Matrix Element Method discriminant

- The Matrix Element Method discriminant (MEM) relies on the direct computation of per-event weights arising from the underlying scattering amplitudes integrated over the phase space, based only on the event kinematics such as jet and lepton momenta:
  - The observable detector-level quantities are connected to parton-level quantities using detector transfer functions taken from simulation.
  - The final MEM discriminator is constructed as likelihood ratio between the event weights for the signal and background hypothesis.

$$w(\mathbf{y}|\mathcal{H}) = \sum_{\text{perm}} \int \mathrm{d}\mathbf{x} \int \frac{\mathrm{d}x_a \mathrm{d}x_b}{2x_a x_b s} g(x_a, Q) g(x_b, Q) \times \delta^4\left(p_a + p_b - \sum_{k=1}^{8} p_k\right) |\mathcal{M}_{\mathcal{H}}(p_a, p_b, \mathbf{p})|^2 W(\mathbf{y}|\mathbf{p}),$$

where y represents the set of measured observables, i.e. jet momenta, and s is the square of the pp centre-of-mass energy

Sources of systematic uncertainties



Post-fit pull of the constrained (text in black) and unconstrained (text in gray) nuisance parameters included in the fit to the **2016** plus **2017** data as well as their impact on the signal strength μ, ordered by their impact.