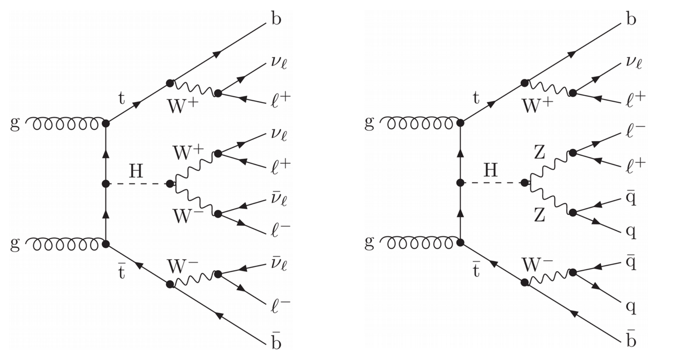
Search for Higgs boson production in association with top quarks in multilepton final states at √s = 13 TeV

Jing Li on behalf of CMS collaboration
Peking University, IPHC

Introduction

Signal process ($t\bar{t}H$):

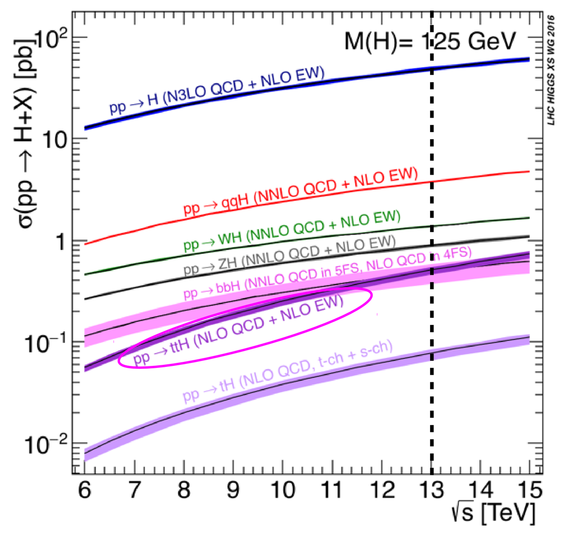
- Standard model Higgs boson in association with a top quark pair
  - → Higgs bosons decays to WW^*, $\tau\tau$, ZZ^*
  - → Leptonic decay of at least one of the top quarks
  - → Hadronic τ vetoed (measured in CMS-PAS-HIG-17-003)



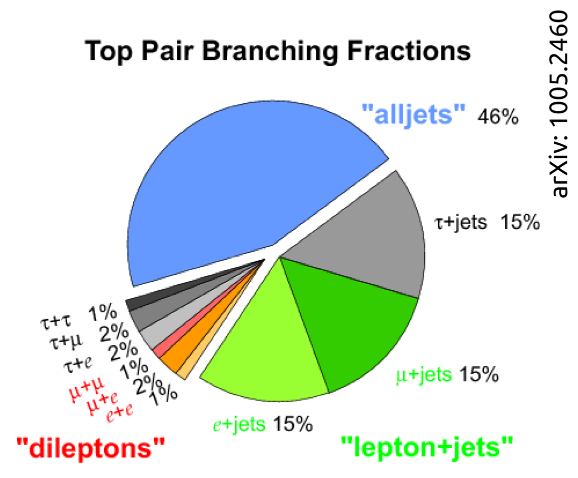
Multilepton final state: H → WW*, ττ, ZZ*
→ multileptons, large BR

Dataset used:

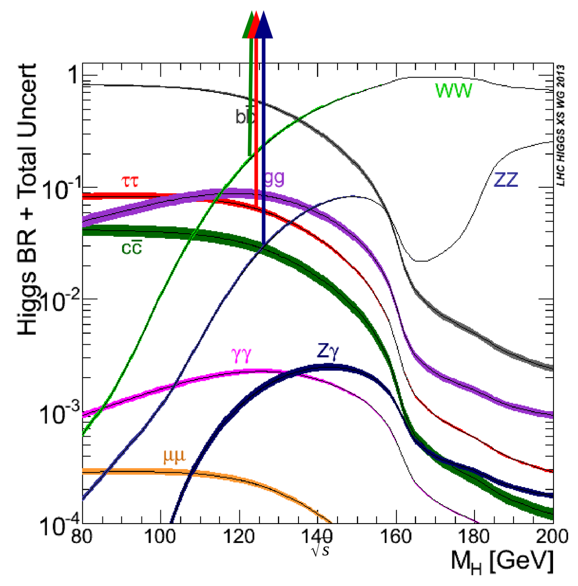
- pp collision data collected by CMS experiment at √s = 13TeV
- Integrated luminosity of 35.9 fb^{-1}



⊗

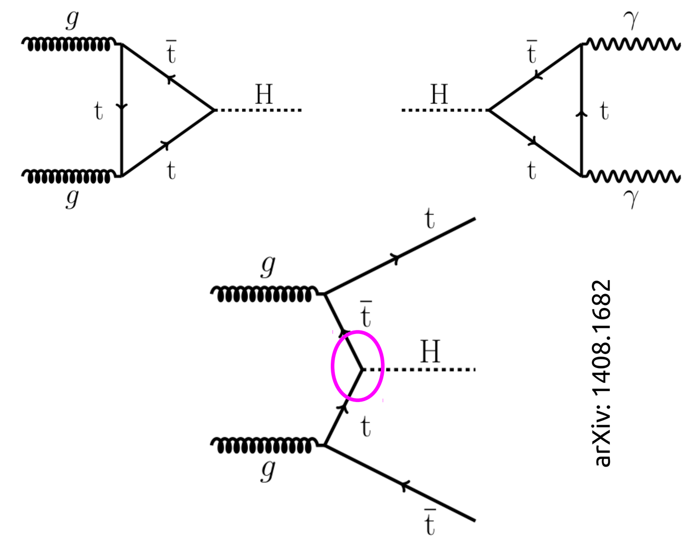


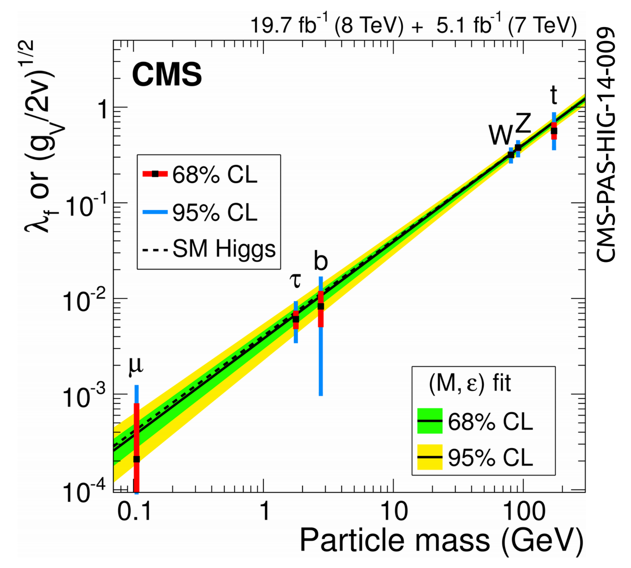
⊗



Motivation

- After the discovery of the Higgs boson, its mass has been accurately measured and its couplings to most standard model (SM) particles constrained
- Top quark is the most strongly coupled SM particle to Higgs (top Yukawa coupling $Y_t \approx 1$). Studies of various final states have been used to constrain the Y_t to about 30% accuracy
- Indirect constraints on the Y_t from loop diagrams where top quark contribute to gg → H production and H → γγ decay (assuming no new particles)
- Directly probe top-Higgs coupling through $t\bar{t}H$ production mechanism





Main backgrounds

Main reducible background:

- Non-prompt leptons from heavy flavor decays, misidentified light jets. Mainly comes from $t\bar{t}$.
  - → dedicated lepton MVA selection
  - → fake rate measured mainly from QCD data
  - → fake rate applied on relaxed selection
- Charge mis-assignment (in 2lss category)
  - → yields from fit of m_{ee} around the Z-peak in data
  - → measure the ratio of same sign to opposite sign events as a function of p_T, η

Main irreducible background:

- $t\bar{t}W/t\bar{t}Z/t\bar{t}\gamma^*/t\bar{t}\gamma$
  - → estimated from simulation
- Dibosons
  - → scale factor measured in 3l + >=2 jets and 0 b-jet control region

Event categorization

Lepton flavor

- **2lss** (2 same-sign leptons) / **3l (4l)** (>=3 leptons) (leptons=electrons, muons)

b-tight (2 CSVM) / **b-loose** (1 CSVM or 2 CSVL)

- b signal tagging efficiency: 85% for CSVL, 70% for CSVM

Sum of lepton charges (+/-)

- take advantage of $t\bar{t}W$ charge asymmetry

$\mu^\pm\mu^\pm$	$e^\pm\mu^\pm$	$e^\pm e^\pm$	3L
b tight +-	b tight +-	+-	b tight +-
b loose +-	b loose +-		b loose +-

Event selection

mET LD = mET*0.004 + mHT*0.003

- mHT is computed in the same way as mET, but using only the selected jets and leptons, recover from the performance degradation of mET due to pile-up
- mET LD has the effect of rejecting roughly a factor of two more Drell-Yan events than a simple mET requirement

2lss category

- $|m_{ll} - m_Z|$ >10 GeV (if has ee pair), m_{ll} >12 GeV
- mET LD > 0.2
- ≥ 4 hadronic jets

3l / 4l category

- $|m_{ll} - m_Z|$ >10 GeV (if has SF/OS lepton pairs), m_{ll} >12 GeV
- mET LD > 0.2 (if <4 jets), >0.3 (if SFOS)
- ≥ 2 hadronic jets

Matrix element method (MEM)

Conditional probability based on LO Feymann diagrams of theoretical process

MEM weight

Element of phase space corresponding to unmeasured quantities

Squared matrix element

$$w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha}\int d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k>2} p_k^\mu\right) \cdot \frac{f(x_1,\mu_F)\,f(x_2,\mu_F)}{x_1 x_2 s} \cdot \left|\mathcal{M}_\alpha(p_k^\mu)\right|^2 \cdot W(\Phi'|\Phi_\alpha),$$

Cross section

Enforcing 4-momentum conservation

Parton density function

Transfer functions

- Evaluate MEM weights under $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z/\gamma^*$ hypotheses
- MEM weights is the average weight of all possible lepton, jets, b-jets permutations
- Especially useful when a SFOS pair present
- Likelihood ratio of $t\bar{t}H$ vs $t\bar{t}V$ from MEM as input of the BDT in 3l category

$$\mathcal{L}_{t\bar{t}H\,vs\,t\bar{t}V} = -\log\left(\frac{\sigma_{t\bar{t}V}\, w_{t\bar{t}V}}{\sigma_{t\bar{t}H}\, w_{t\bar{t}H} + \sigma_{t\bar{t}V}\, w_{t\bar{t}V}}\right)$$

Advantages

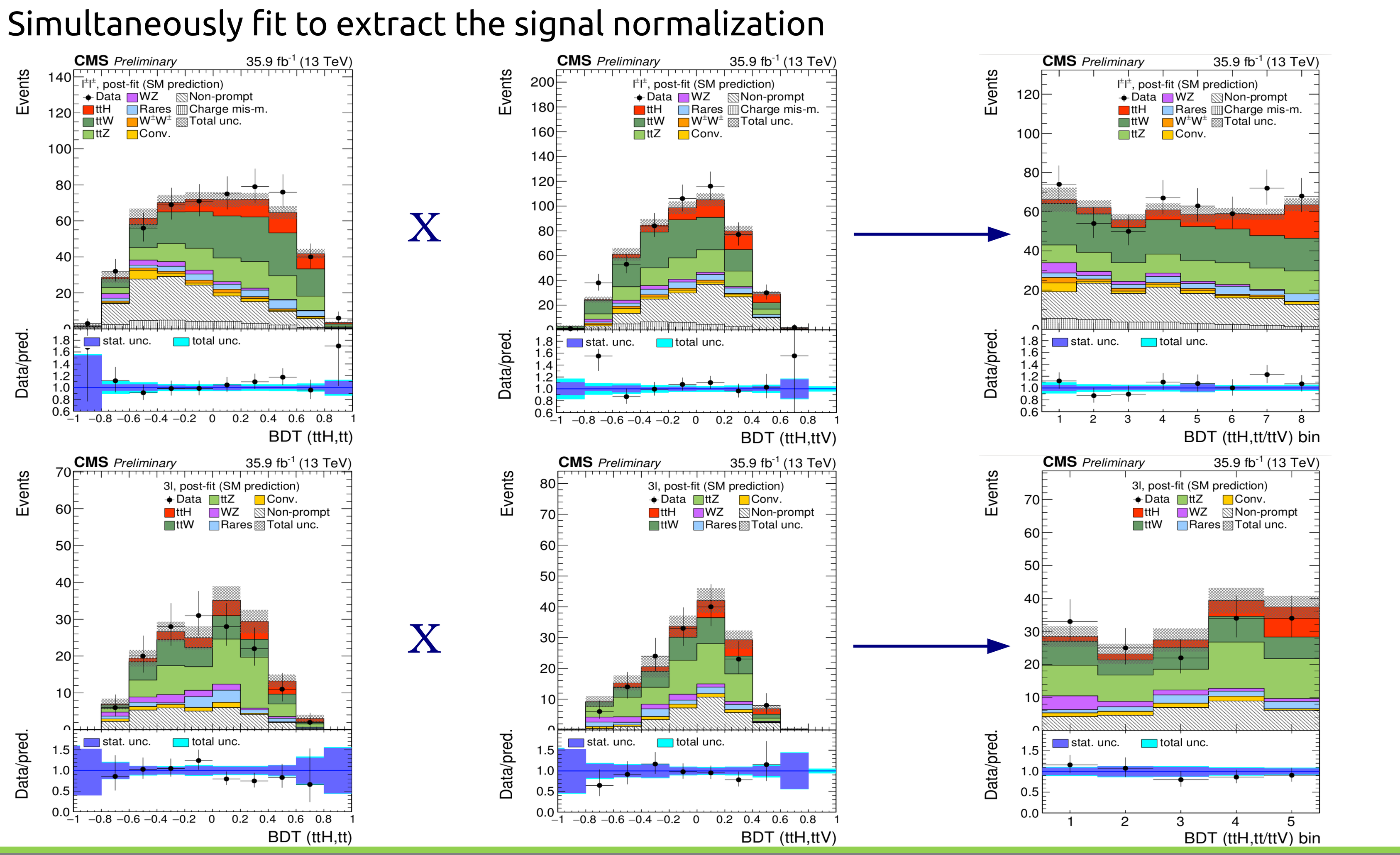
- Makes maximal use of both experimental information and the theoretical model on an event-by-event basis
- Good discrimination vs irreducible background, especially for complex final state

Limits

- CPU intensive
- Model dependent
- Matrix element LO only

Signal extraction strategy

- Build 2D kinematic BDT: MVA($t\bar{t}H$ vs $t\bar{t}$), MVA($t\bar{t}H$ vs $t\bar{t}V$)
- Unroll 2D BDT into a 1D distribution
  - Bin into a 1D distribution by ordered likelihood ratio value in the 2D BDT plane
  - Split in quantiles of background cumulative distribution
  - → obtain unrolled distributions with flat background from S-B likelihood ratio
- Simultaneously fit to extract the signal normalization



Other multivariate methods used

- Lepton prompt vs non-prompt MVA
  - → used for event selection to reduce non-prompt lepton background
- Hadronic top decay tagger
  - → target the hadronic top in signal signature
  - → used as input of the BDT in 2lss category to reduce the $t\bar{t}$ background
- Hj taggger:
  - → identify jets origination from the Higgs semi-leptonic decay
  - → use score of highest score jet input of the BDT in 2lss category to reduce the $t\bar{t}V$ background

Results

	Observed σ/σ_{SM} ± 1σ	Significance
2lss	1.7 (+0.6) (-0.5)	3.3 σ
3l	1.0 (+0.8) (-0.7)	1.4 σ
4l	0.9 (+2.3) (-1.6)	0.5 σ
Combined	1.5 (+0.5) (-0.5)	3.3 σ
Expected	1.0 (+0.5) (-0.4)	2.5 σ

Combine 2015+2016 results

- 2015 data (CMS-PAS-HIG-15-008): σ/σ_{SM} = 0.6 (+1.4) (-1.1)
- Combined (2015+2016 data): σ/σ_{SM} = 1.5 (+0.5) (-0.5); significance=3.3 σ

References

CMS Collaboration, "Search for Higgs boson production in association with top quarks in multilepton final states at √s = 13 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-17-004

jing.li@cern.ch