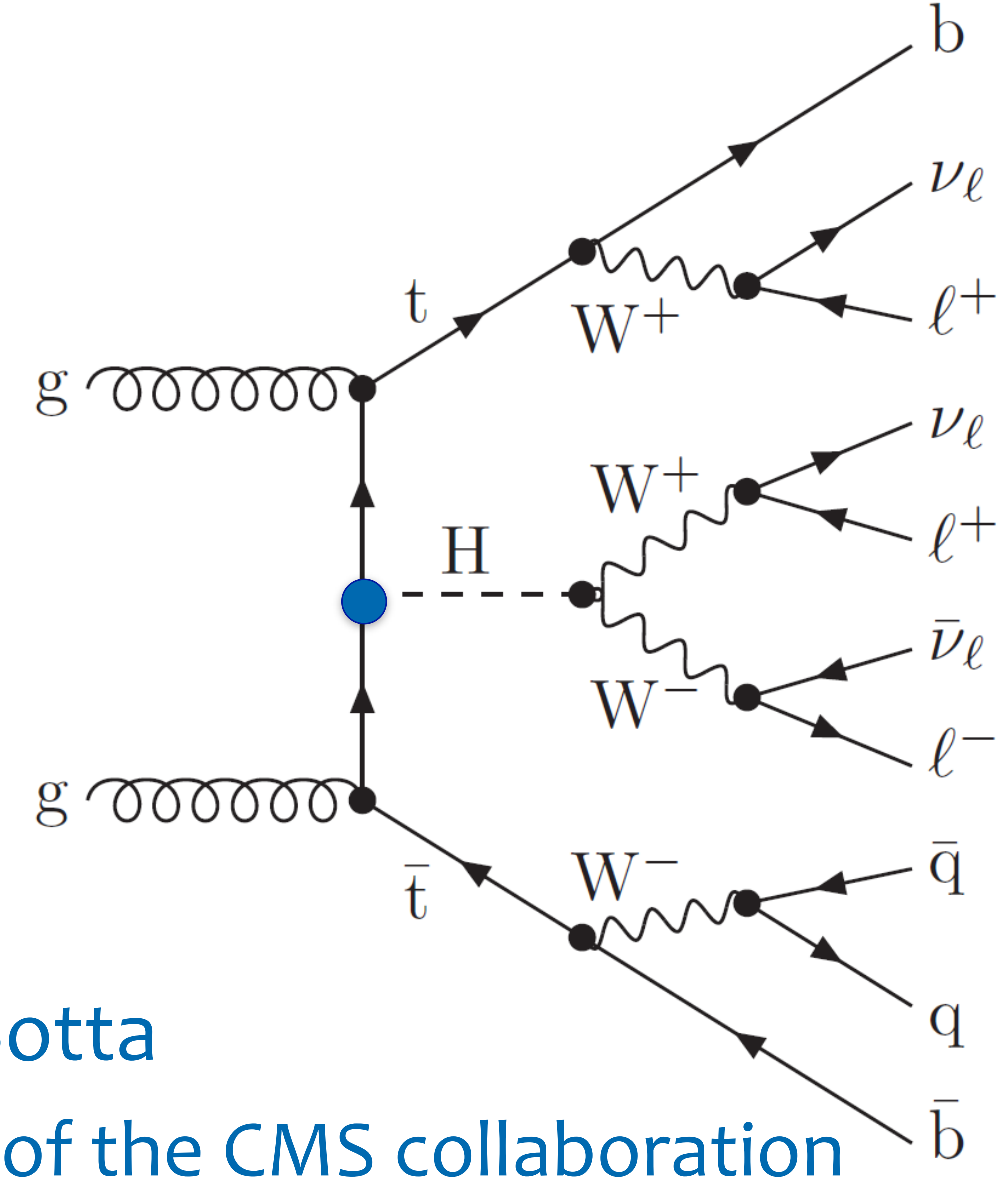
Search for $t\bar{t}H$ associated production at CMS



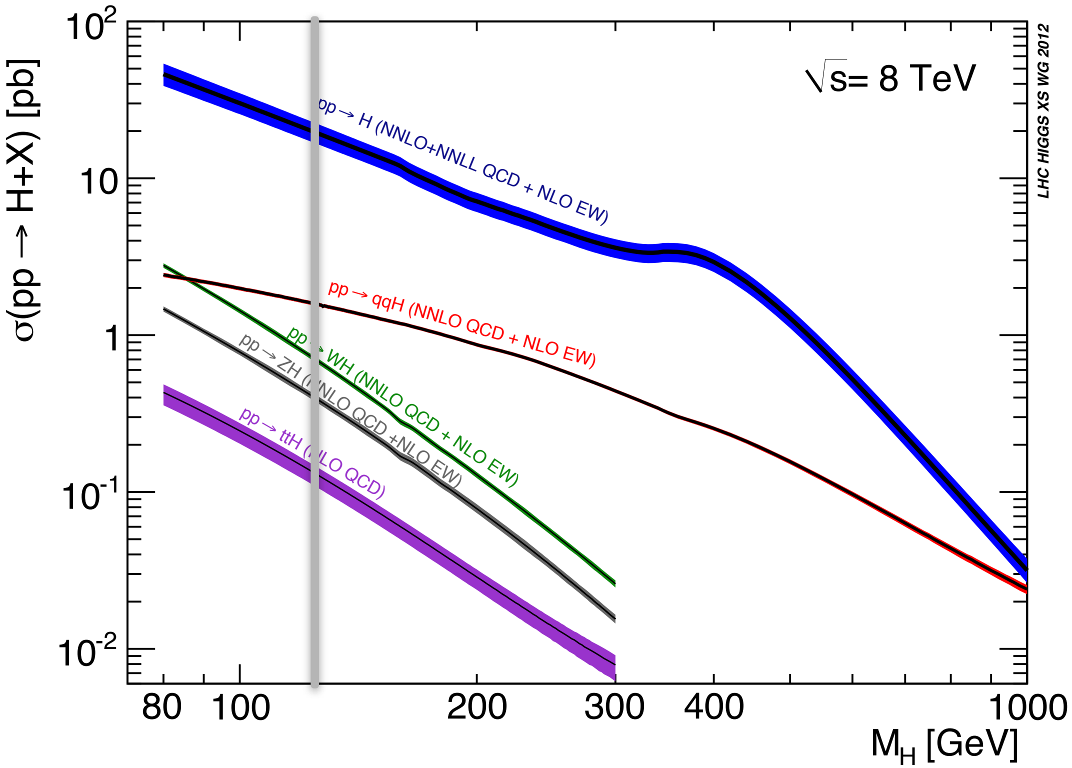
Cristina Botta
on behalf of the CMS collaboration

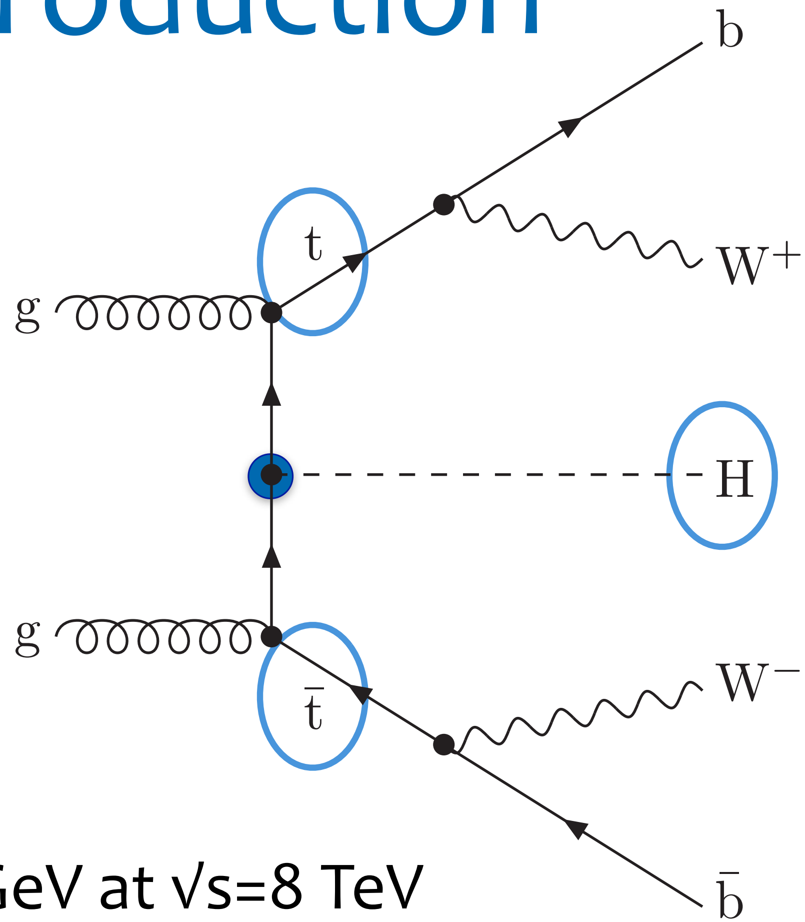
Rencontres de Moriond on “EW Interactions and Unified Theories”
15-22.03.2014, La Thuile (Italy)

The top Yukawa coupling

- Two main probes of $t\bar{t}H$ coupling at LHC:
  - gluon fusion production cross section ($\sigma \sim |y_t|^2$), assuming no BSM particles in the loop.
  - associated production cross section, a tree level process proportional to $|y_t|^2$
- The first is pretty well known: already now the experimental accuracy on y_t is ~20%
- Significant progress from the experimental side on the second point in the last year.

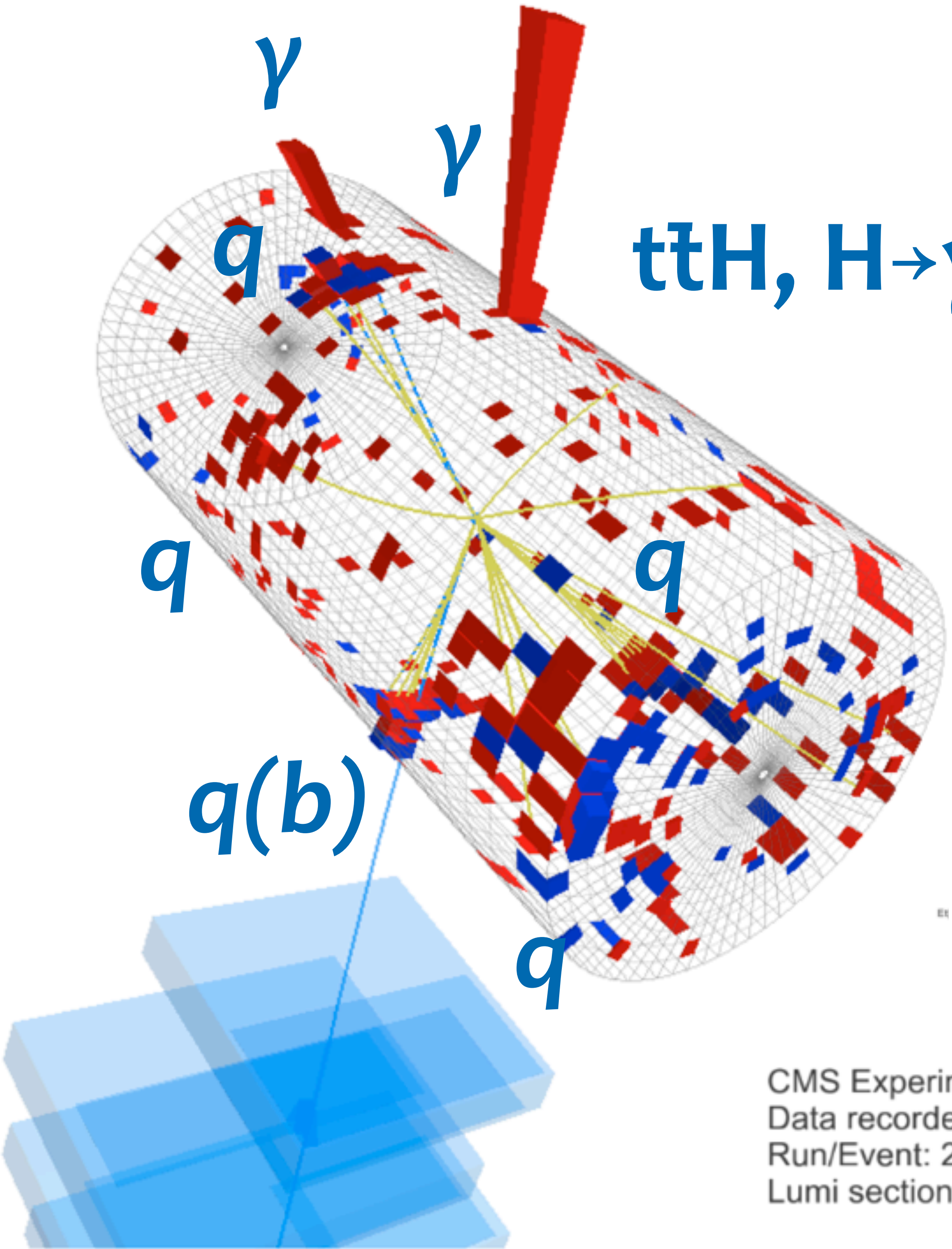
tt̄H associated production





- ttH cross section is **130 fb** for m_H = 126 GeV at √s=8 TeV
- Higgs decays at this mass:
  - BR H→bb̄ ~60%, H→WW$^{(*)}$ ~20% and H→γγ, H→ττ, H→ZZ$^{(*)}$ significantly smaller BR but produce experimentally accessible signatures
- ttH events are crowded due to the presence of additional b-jets, jets/leptons from the top quarks decays

tt̄H, H→γγ candidate



CMS Experiment at LHC, CERN
Data recorded: Sat Nov 24 19:16:36 2012 CEST
Run/Event: 207889 / 771018991
Lumi section: 783

ttH searches at CMS

- **tt + b-jets,** to search for **H→bb**
  - old staple. 7 and 8 TeV data analyzed.
  - high rate but big tt+bb bkg and complex multi-jet final state
  - Public document: **HIG-12-035 / HIG-13-019** (together with the low sensitivity channel **tt + ττ: with hadronically decaying taus**)

H→hadrons

- **tt + γγ,** to search for **H→γγ:**
  - 8 TeV data analyzed.
  - low rate. important for the high-lumi projection as systematics play a negligible role in it
  - Public document: **HIG-13-015**

H→photons

- **tt + leptons,** to search for **H→ WW, ZZ, ττ (τ → ℓ)**
  - 8 TeV data analyzed. Latest CMS search
  - low rate. clean and low bkg signatures with 2,3,4 leptons
  - Public document: **HIG-13-020**

H→leptons

H→hadrons

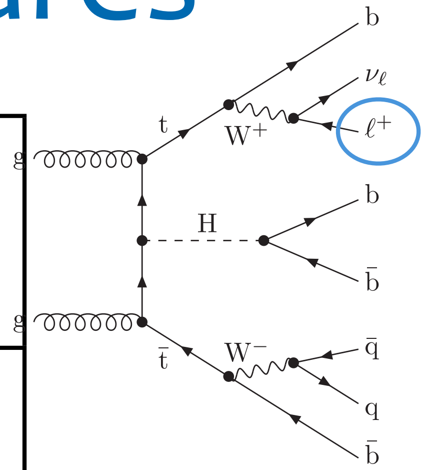
19.3 fb^{-1} @ 8 TeV

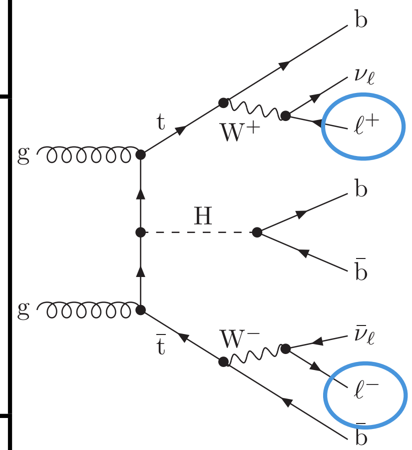
the old analysis for 5.1 fb^{-1} @ 7 TeV is not discussed here but included in the final results

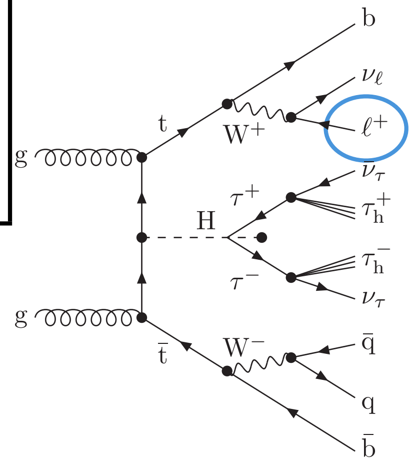
HIG-12-035 / HIG-13-019

Channels and Signatures

H decay	top pair decay	trigger	signature
bb	semileptonic	single lepton (p_T >27(e), 24(μ) GeV)	1 e/μ, p_T>30 GeV ≥4 jets (≥2b-jets), p_T>30 GeV **(#sig~90 sig/bkg~0.004)**
bb	dileptonic	double lepton (p_T >17,8 GeV)	1 e/μ, p_T>20 GeV 1 e/μ, p_T>10 GeV ≥3 jets (≥2b-jets), p_T>30 GeV **(#sig~30 sig/bkg~0.002)**
$\tau_h\tau_h$	semileptonic	single lepton (p_T >27(e), 24(μ) GeV)	1 e/μ, p_T>30 GeV 2 τ, p_T>20 GeV ≥4jets (1-2b-jets), p_T>30 GeV **(#sig~2 sig/bkg~0.003)**



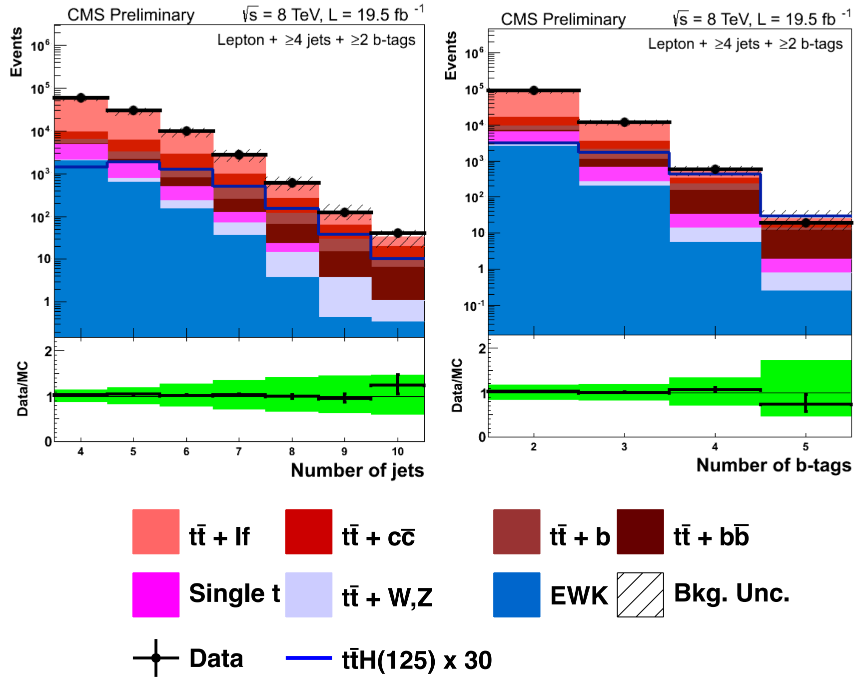




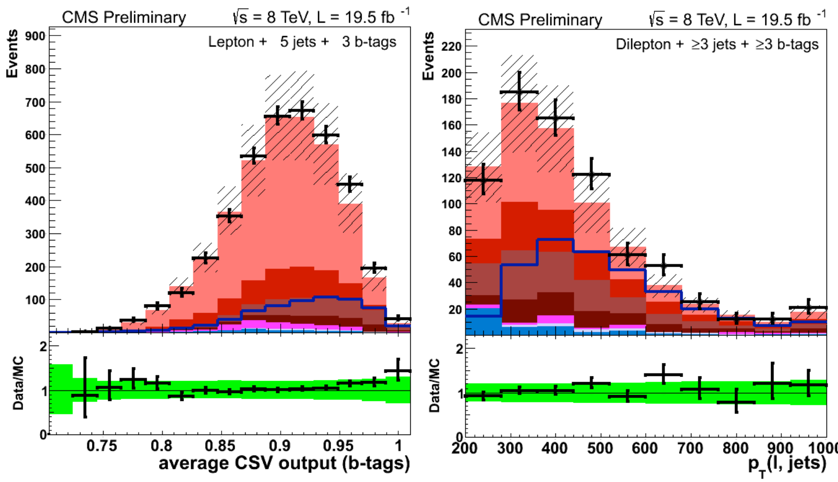
Analysis strategy

bb-semileptonic

- Further **categorization in NJets and Nb-Jets**
  - S/B increases with Njets and Nbjets:
    - 16 categories, best S/B = 0.03
- All bkg are modeled with MC **simulation**
  - **main bkg tt+jets** (MADGRAPH, tt+3 extra parton)
    - separated in sub-samples: (tt+lf, tt+bb,tt+b,tt+cc)
  - data in bins with low S/B are used to constraint these bkg sources
- **BDTs** are used to further improve sensitivity
  - input variables related to: **reconstructed object kinematic, event shape, b-tagging discriminator value**

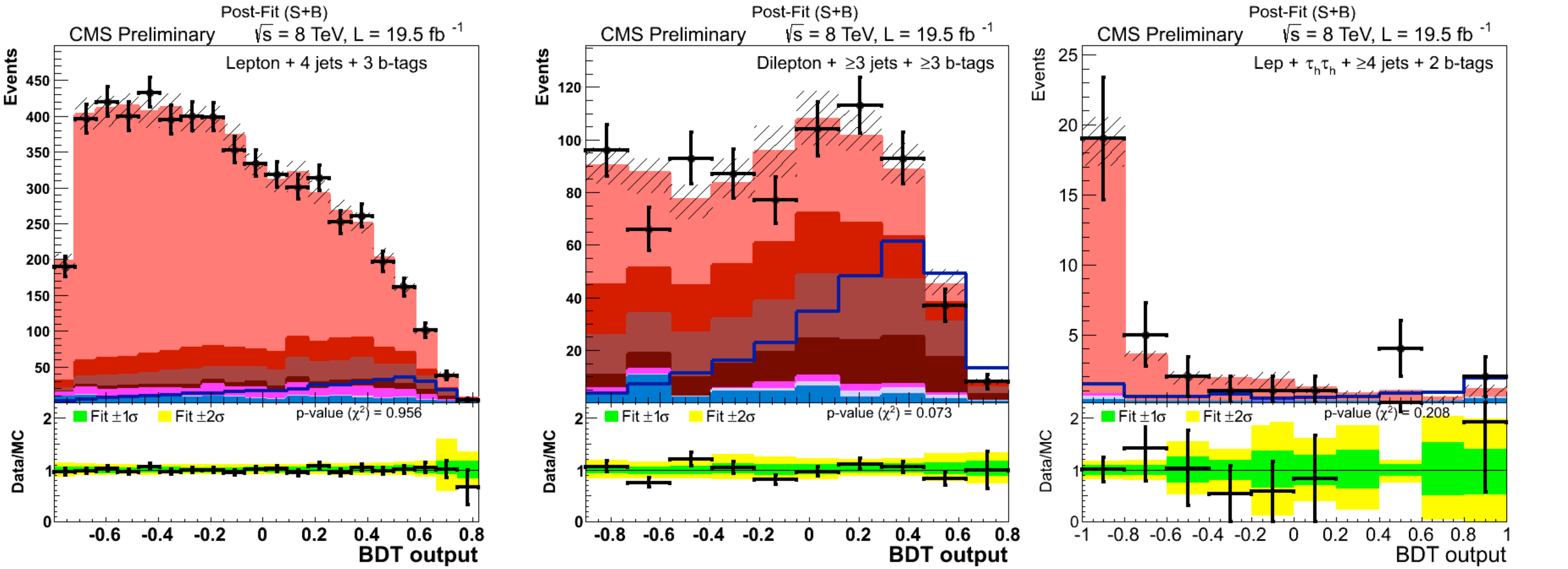


1 Lep + 5 Jets + 3 b-tags | **2 Lep ≥ 3 Jets ≥ 3 b-tags**



Final distributions

- Signal extraction is performed fitting the **BDT outputs** (3 example out of 16 distributions)



bb: 1 Lep + 4 Jets + 3 b-tags

bb: 2 Lep ≥ 3 Jets ≥ 3 b-tags

ττ: 1 Lep ≥ 4 Jets ≥ 2 b-tags

uncertainty bands: systematic on bkg+signal prediction, after a fit to the data assuming SM ttH (μ=1)

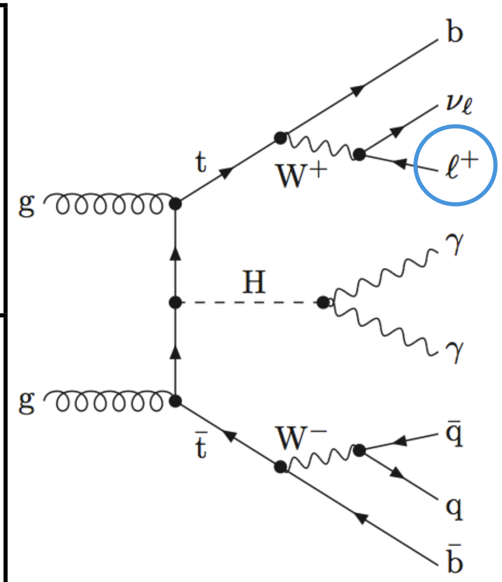
H→photons

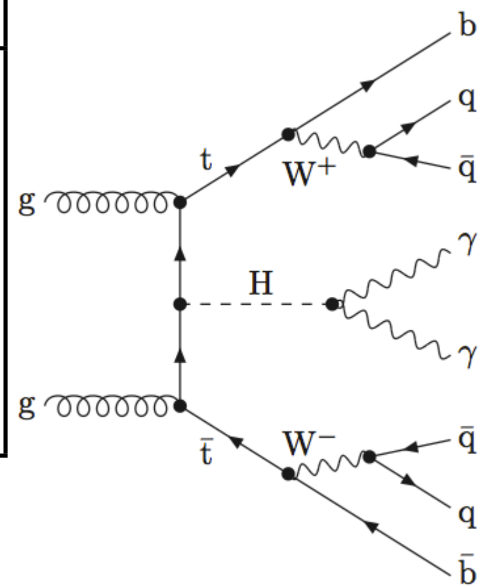
19.6 fb^{-1} @ 8 TeV

HIG-13-015

Channels and Signatures

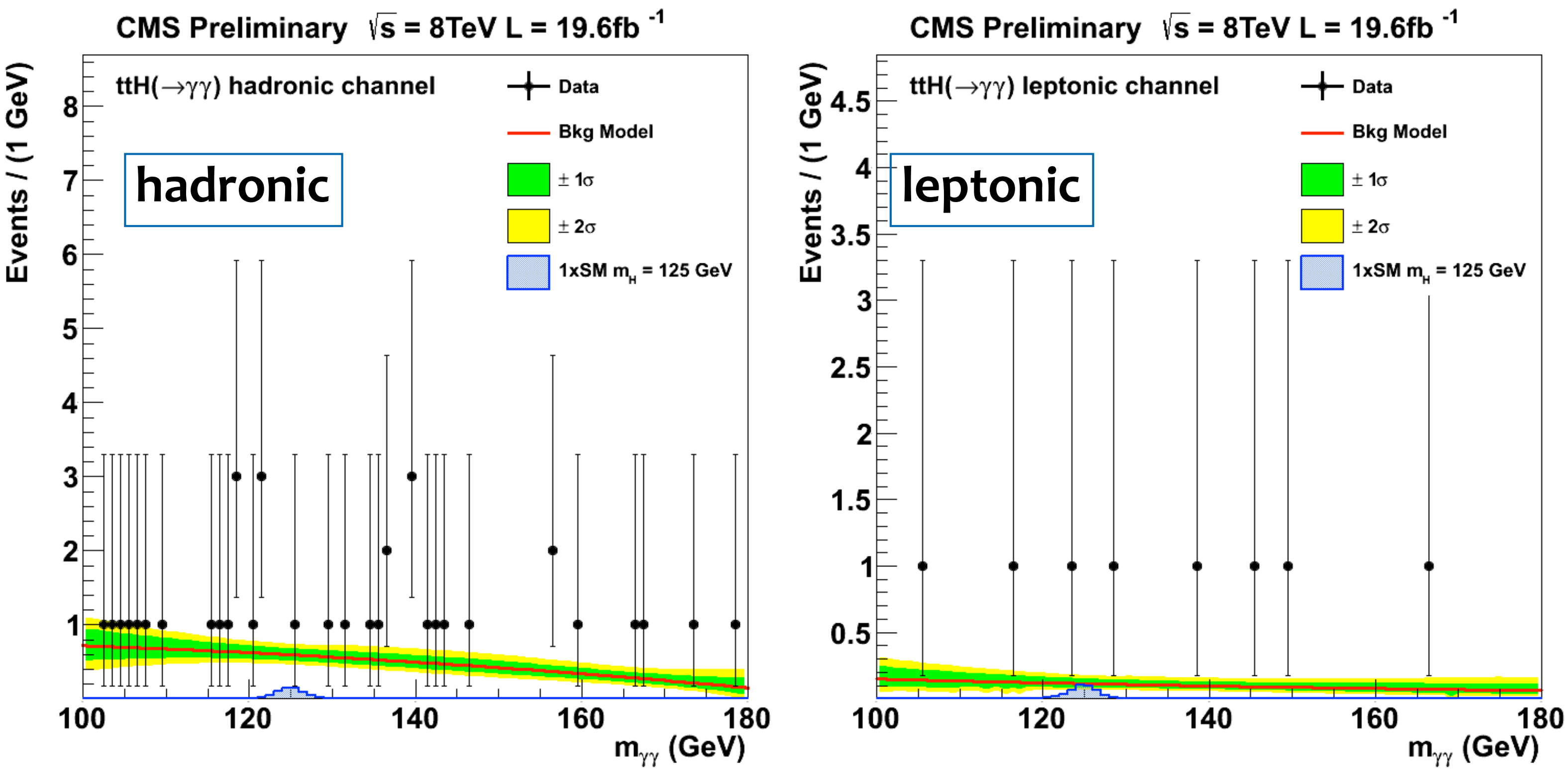
H decay	top pair decay	trigger	signature
γγ	semileptonic or dileptonic	di-photon (p_T >36,22 GeV)	$p_{T(\gamma 1)} > m_{\gamma\gamma}/2$ $p_{T(\gamma 2)} >$ 25GeV ≥ 1 e/μ, p_T>20 GeV ≥2 jets (≥1b-jets), p_T>25 GeV **(#sig~0.3 sig/bkg~1 under signal peak)**
γγ	hadronic	di-photon (p_T >36,22 GeV)	$p_{T(\gamma 1)} > m_{\gamma\gamma}/2$ $p_{T(\gamma 2)} >$ 25GeV 0 e/μ, p_T>20 GeV ≥5 jets (≥1b-jets), p_T>25 GeV **(#sig~0.5 sig/bkg~0.3 under signal peak)**





Analysis strategy

- Analysis **limited by statistic** (low BR H→γγ) but **distinctive signature**:
  - two energetic photons, narrow Higgs peak over falling bkg in $M_{\gamma\gamma}$ distribution
  - the only channel that can eventually confirm that an excess is due to h(126)
- Strategy: **fit the $M_{\gamma\gamma}$ distribution** using the diphoton spectrum **sidebands to fit the bkg**
  - Data fitted with simple exponential (second order polinomial) in the leptonic (hadronic) channel



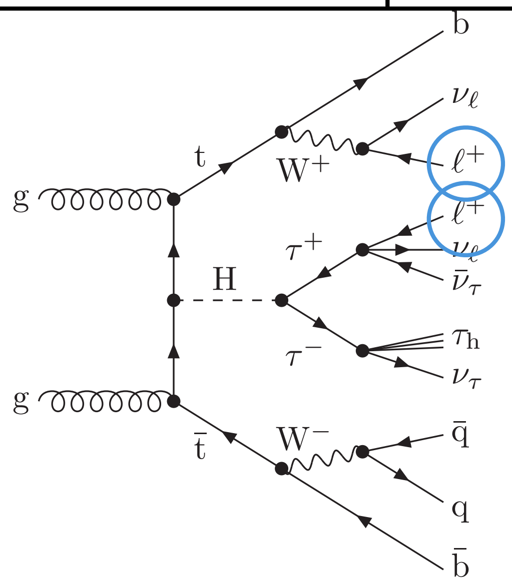
H→leptons

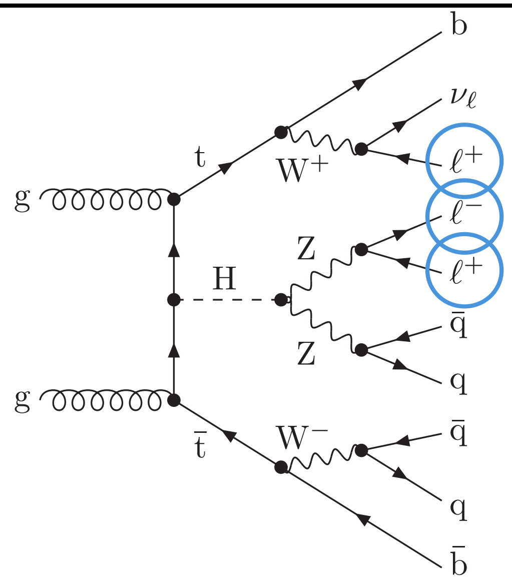
19.6 fb^{-1} @ 8 TeV

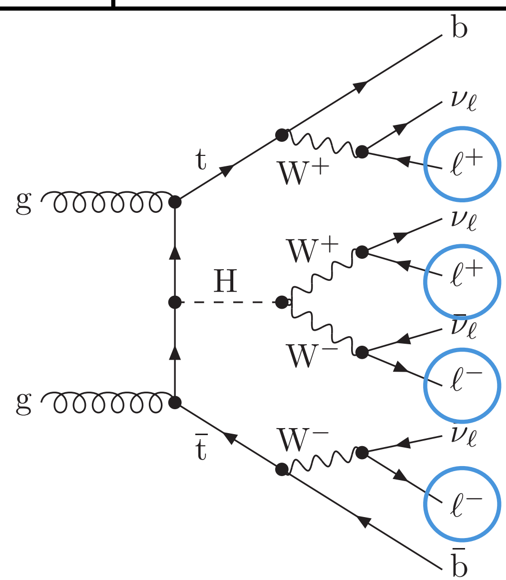
HIG-13-020

Channels and Signatures

H decay	top pair decay	trigger
WW, ZZ, ττ	semileptonic or dileptonic	double lepton (p_T >17,8 GeV)
signature		
2 same-sign leptons (ee,em,mm) 2 e/μ, p_T>20 GeV ≥4 jets (≥1b-jet), p_T>25 GeV **(#sig~8 sig/bkg~0.08)**	**3 leptons** 1 e/μ, p_T>20 GeV 1 e/μ, p_T>10 GeV 1 e(μ), p_T>7(5) GeV ≥2 jets (≥1b-jet), p_T>25 GeV no resonant Z->ll **(#sig~4 sig/bkg~0.07)**	**4 leptons** 1 e/μ, p_T>20 GeV 1 e/μ, p_T>10 GeV 2 e(μ), p_T>7(5) GeV ≥2 jets (≥1b-jet), p_T>25 GeV no resonant Z->ll **(#sig~0.5 sig/bkg~0.2)**

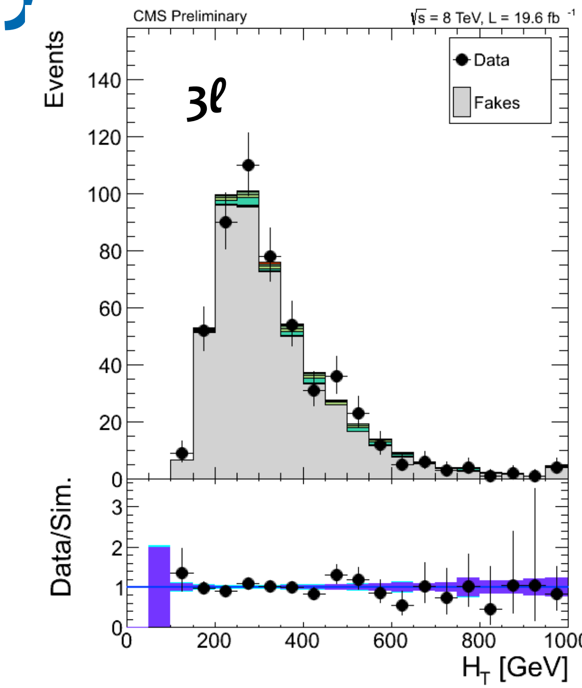




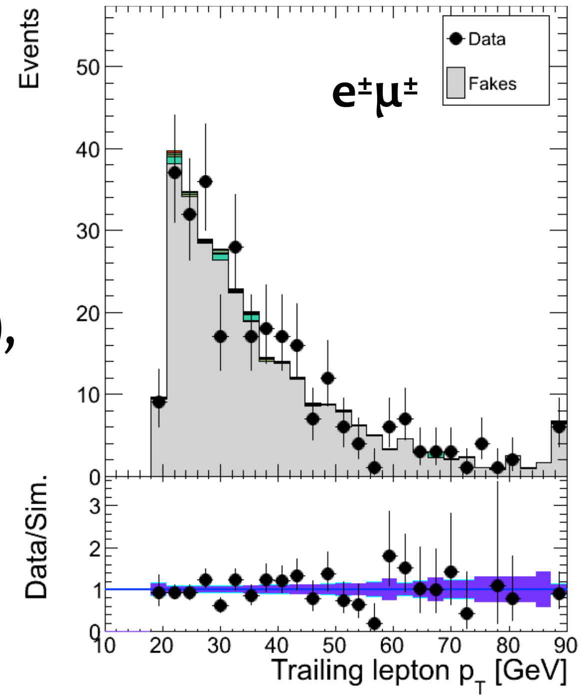


Analysis Strategy

- Main focus: suppress and control **reducible background** (~up to 2/3 of the total bkg after selection)
  - tt with fake ℓ from b-jets
    Dedicated lepton ID (MVA) developed to suppress it.
  - data-driven estimate: measurement of the probability for a lepton from b-jet to pass the MVA ID requirement

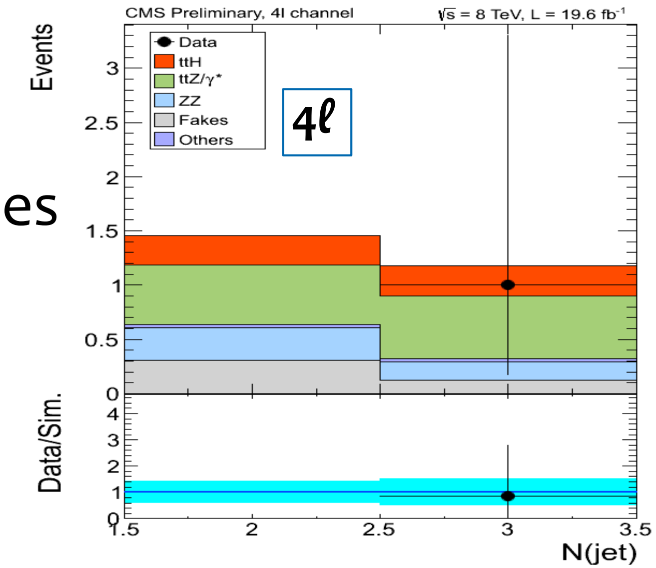


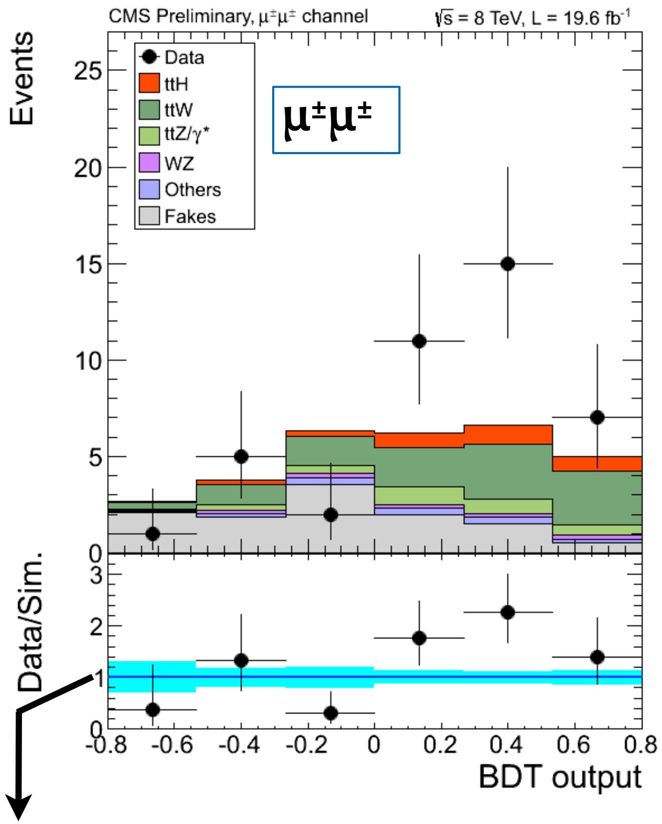
- Inclusive selection to preserve signal efficiency.
  Full event kinematic cannot be reconstructed
  - **to improve sensitivity:**
    - categorize events (for 2ℓ, 3ℓ) in **positive and negative total lepton charge** (ttW, WZ and Wjets are asymmetric), **5%** gain in sensitivity
    - combine partial kinematic variables in a **BDT** (for 2ℓ, 3ℓ), **10%** gain in sensitivity

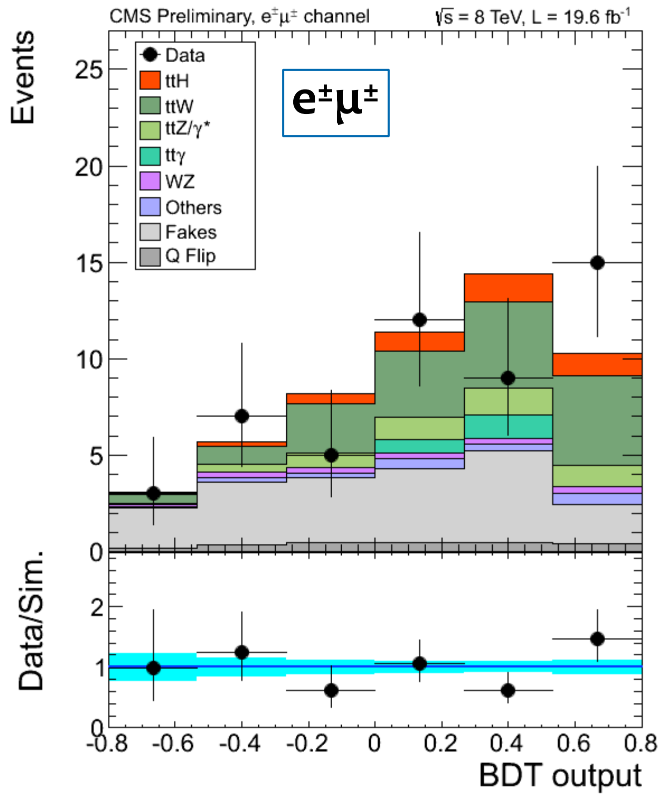


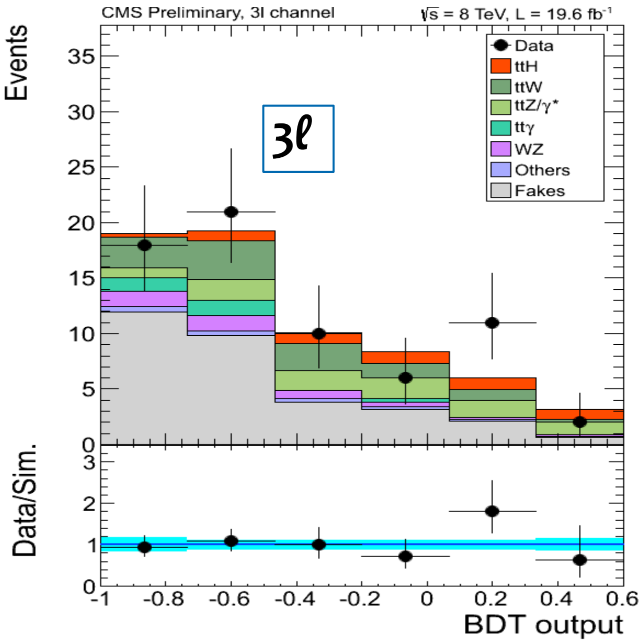
Final distributions

- signal extraction, in each category:
  - **2ℓ, 3ℓ: simple BDT** with few kinematic variables
  - **4ℓ: just use N(jet)**, since yields are small.
  - N(jet) used also as cross-check in 2ℓ, 3ℓ







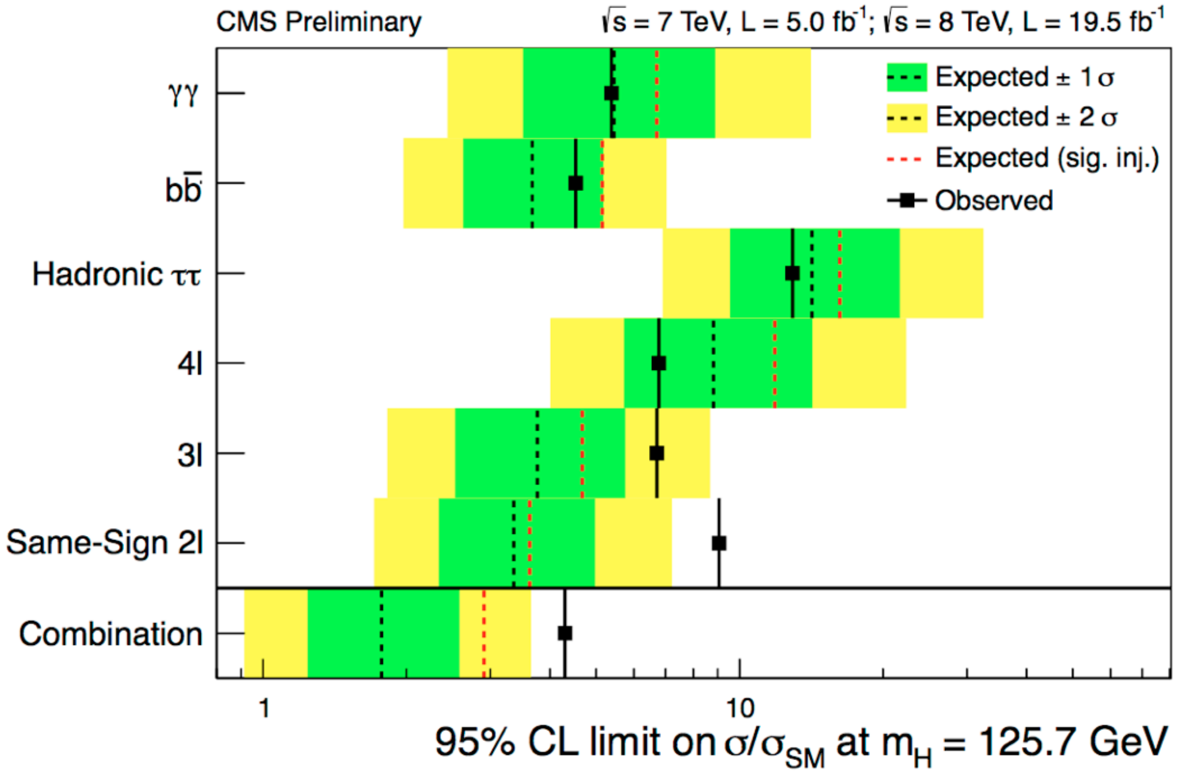


uncertainty bands: systematic on bkg+signal prediction, after a fit to the data assuming SM ttH (μ=1)

Statistical Interpretation

https://twiki.cern.ch/twiki/bin/view/CMSPublic/ttHCombinationTWiki

Limits on $\mu=\sigma/\sigma_{SM}$

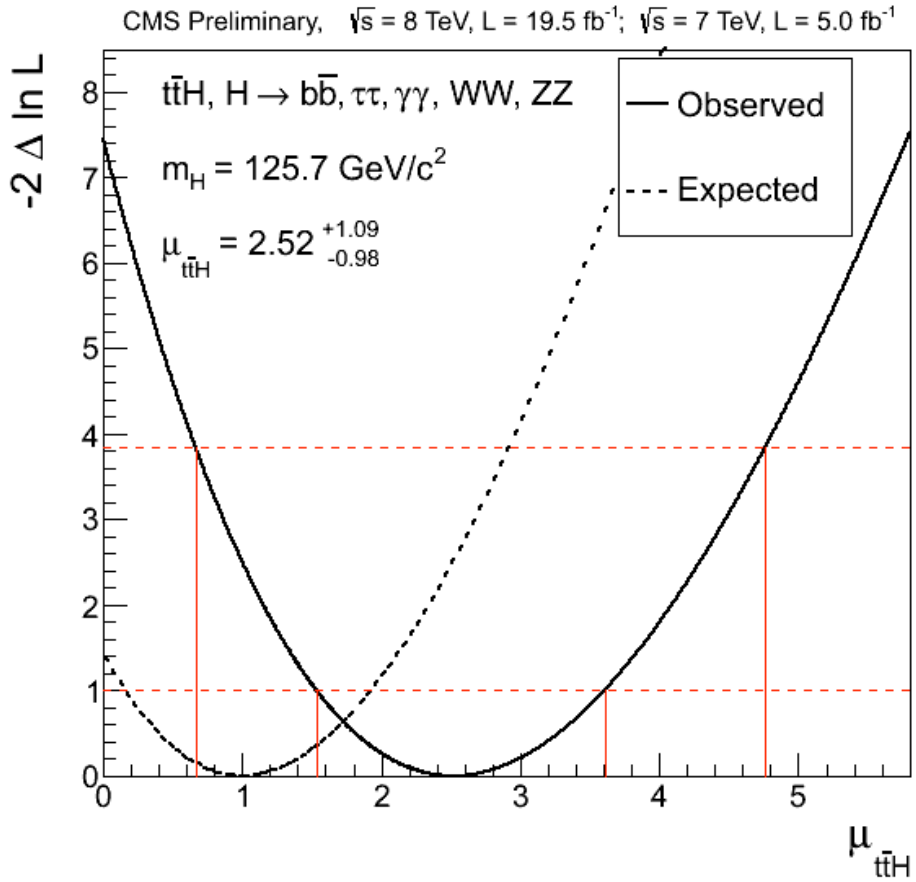


ttH Channel	95% CL upper limits on $\mu=\sigma/\sigma_{SM}$ ($m_H = 125.7$ GeV)				
		Expected			
	Observed	Median Signal Injected	Median	68% CL Range	95% CL Range
$\gamma\gamma$	5.4	6.7	5.5	[3.5,8.9]	[2.4,14.1]
$b\bar{b}$	4.5	5.2	3.7	[2.6,5.2]	[2.0,7.0]
$\tau\tau$	12.9	16.2	14.2	[9.5,21.7]	[6.9,32.5]
4l	6.8	11.9	8.8	[5.7,14.2]	[4.0,22.4]
3l	6.7	4.7	3.8	[2.5,5.8]	[1.8,8.7]
Same-sign 2l	9.1	3.6	3.4	[2.3,5.0]	[1.7,7.2]
Combined	4.3	2.9	1.8	[1.2,2.6]	[0.9,3.6]

- Observed yields and distributions are compared to the expectation for **a SM Higgs with m_H = 125.7**
- Median expected combined UL on μ:
  - in the absence of ttH signal: **1.8** at 95% CL
  - with the SM ttH production: **2.9** at 95% CL
- Observed UL is **4.3** at 95% CL
  - mainly driven by the observed excess in the **ss μμ channel**

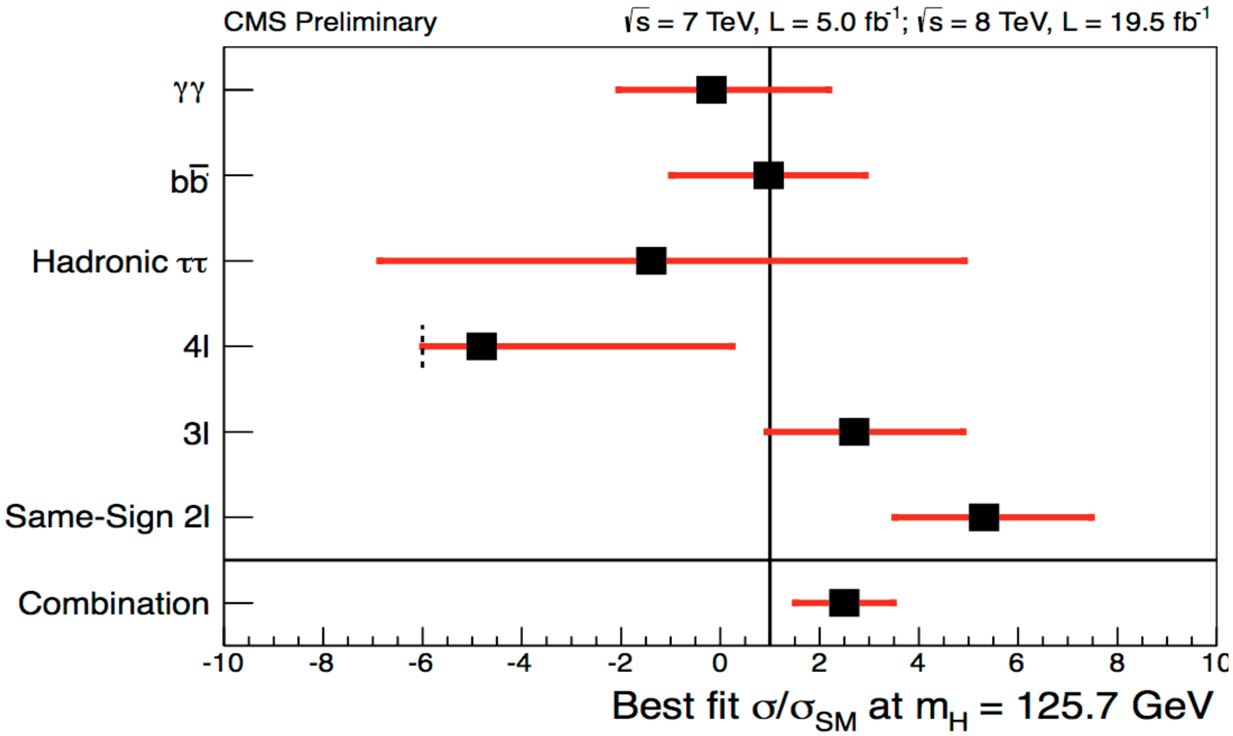
Best fit $\mu=\sigma/\sigma_{SM}$

ttH Channel	$\mu = \sigma/\sigma_{SM}$ ($m_H = 125.7$ GeV)
$\gamma\gamma$	$-0.2^{+2.4}_{-1.9}$
$b\bar{b}$	$+1.0^{+1.9}_{-2.0}$
$\tau\tau$	$-1.4^{+6.3}_{-5.5}$
4l	$-4.8^{+5.0}_{-1.2}$
3l	$+2.7^{+2.2}_{-1.8}$
Same-sign 2l	$+5.3^{+2.2}_{-1.8}$
Combined	$+2.5^{+1.1}_{-1.0}$

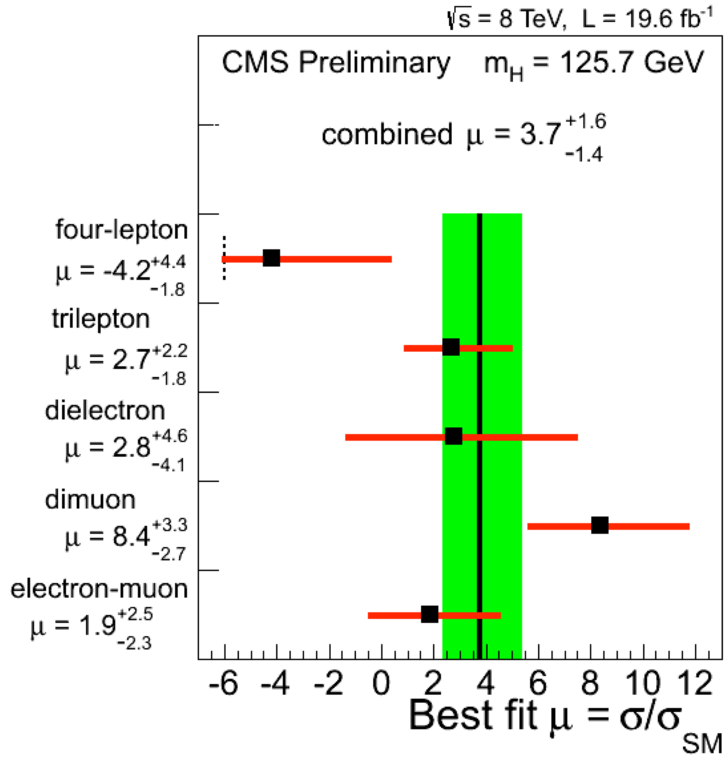


- results from bb, γγ, $\tau_h\tau_h$, and multi-lepton final states : **reached 1×SM sensitivity on μ(ttH)!** (was 2.6×SM at Moriond'13)
- direct access to the y_t coupling

Compatibility



Leptonic channels



- The internal consistency of the 6 best-fit signal-strengths with a common value: **22%**
- The observed p-value relative to μ=1 is **1.6 σ**
  - 2.7 σ relative to μ=0 (1.2 σ expected)

Conclusions

- Significant progress on the search for $t\bar{t}H$ in the past year
  - several signatures have been explored: $tt+bb$, $tt+\tau_h\tau_h$, $tt+\gamma\gamma$, tt+leptons
  - all the channels analysed on the full 8 TeV dataset
- We reached 1xSM sensitivity on $\mu(t\bar{t}H)$
  - we fit $\mu(t\bar{t}H)=2.5^{+1.1}_{-1.0}$ compatible with the SM Higgs prediction ($\mu=1$) at 1.6 σ
  - the excess is mainly driven by the same-sign $\mu\mu$ channel

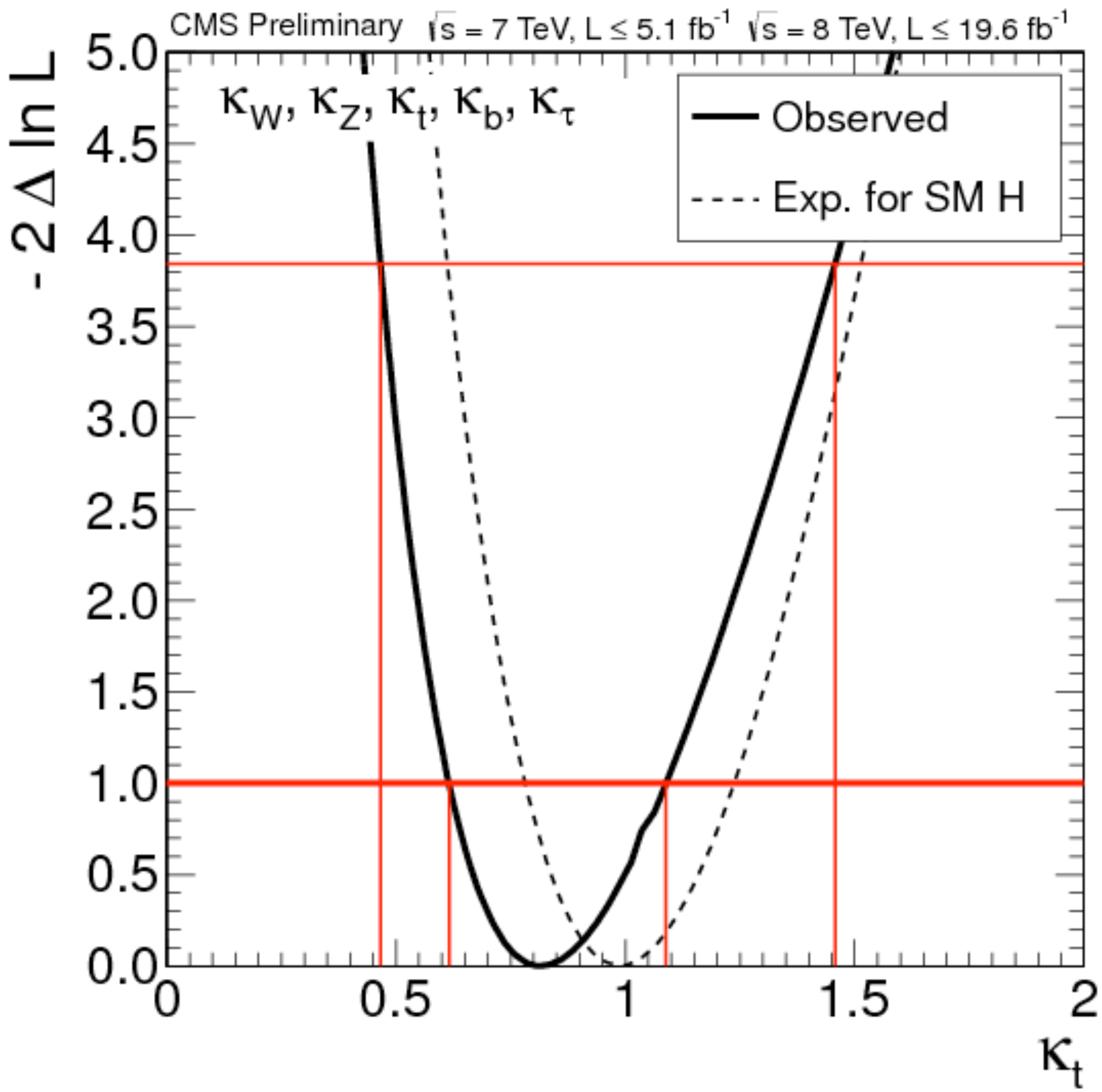
Backup

Systematics

- Statistical methodology based on a binned likelihood spanning all channels
- The amount of signal is characterized by the parameter **$\mu=\sigma/\sigma_{SM}$**
- **Uncertainties in signal and bkg predictions are incorporated by means of nuisance parameters**
  - if same source of uncertainty impact more channels, only one nuisance is used to get the correlation in the uncertainty
  - nuisances are profile, allowing high-statistic signal-poor regions on data to constrain them

Source	Rate Uncertainty Signal	Rate Uncertainty Backgrounds	Shape
Experimental			
Luminosity	2.2–2.6%	2.2–2.6%	No
Jet Energy Scale	0.0–8.4%	0.1–11.5%	Yes
CSV B-Tagging	0.9–21.7%	3.0–29.0%	Yes
Lepton Reco. and ID	0.3–14.0%	1.4–14.0%	No
Lepton Fake Rate ($H \to$ leptons)	N/A	35.1–45.7%	Yes
Tau Reco. and ID ($H \to$ hadrons)	11.3–14.3%	24.1–28.8%	Yes
Photon Reco. and ID ($H \to$ photons)	1.6–3.2%	N/A	Yes
MC Statistics	N/A	0.2–7.0%	Yes
Theoretical			
NLO scale and pdf	9.7–75.0%	3.4–14.7%	No
MC Modeling	2.3–5.1%	1.7–24.2%	Yes
H Contamination ($H \to$ photons)	36.7–41.2%		No

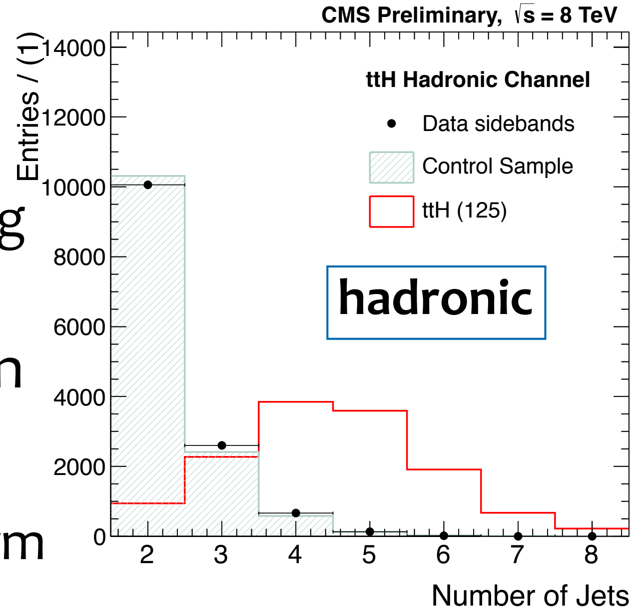
top coupling from combination

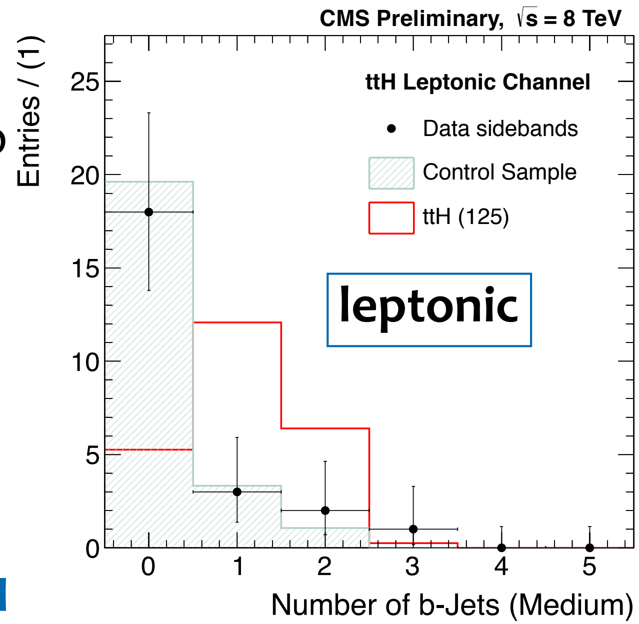


HIG-13-005

Analysis strategy γγ

- Analysis **limited by statistic** (low BR H→γγ) but **distinctive signature**:
  - two energetic photons, narrow Higgs peak over falling bkg in $M_{\gamma\gamma}$ distribution
- Strategy: **fit the $M_{\gamma\gamma}$ distribution** using the diphoton spectrum **sidebands to fit the bkg**
  - independent control sample defined in data to perform studies of the expected bkg
    - single photon trigger
    - inverted identification requirements on one of the two photons
    - two-dimensional reweighting procedure applied to match photon pT and η of the signal region
- Very pure category in ttH
  - signal contamination from other production modes: hadronic: **9% gg→H, 3% WH/ZH, 1% VBF H**, leptonic: **3% VBF H**





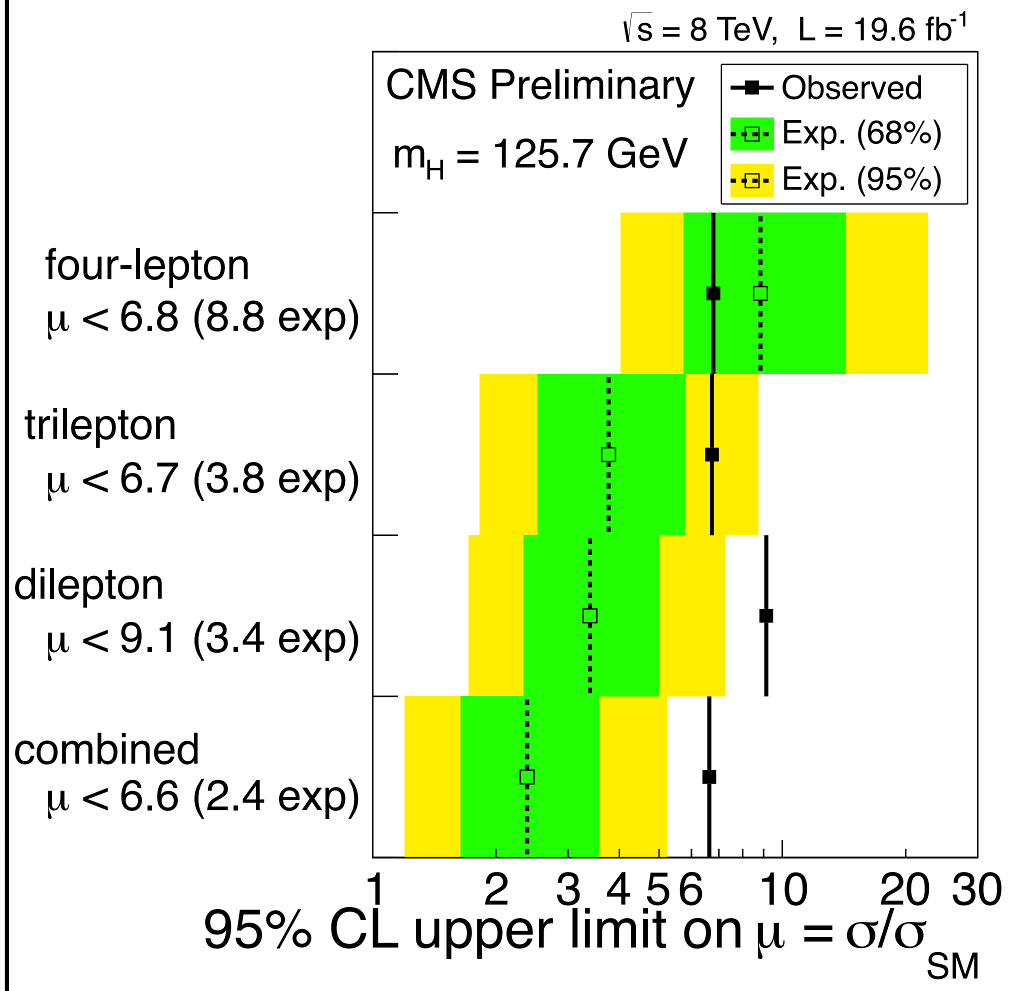
Analysis strategy bb

Source	Shape	Remarks
Luminosity	No	Signal and all backgrounds
Lepton ID/Trigger efficiency	No	Signal and all backgrounds
Pileup	No	Signal and all backgrounds
Top p_T reweighting	Yes	Only $t\bar{t}$ background
Jet Energy Resolution	No	Signal and all backgrounds
Jet Energy Scale	Yes	Signal and all backgrounds
b-Tag bottom-flavor contamination	Yes	Signal and all backgrounds
b-Tag bottom-flavor statistics (linear)	Yes	Signal and all backgrounds
b-Tag bottom-flavor statistics (quadratic)	Yes	Signal and all backgrounds
b-Tag light-flavor contamination	Yes	Signal and all backgrounds
b-Tag light-flavor statistics (linear)	Yes	Signal and all backgrounds
b-Tag light-flavor statistics (quadratic)	Yes	Signal and all backgrounds
b-Tag Charm uncertainty (linear)	Yes	Signal and all backgrounds
b-Tag Charm uncertainty (quadratic)	Yes	Signal and all backgrounds
QCD Scale ($t\bar{t}H$)	No	Scale uncertainty for NLO $t\bar{t}H$ prediction
QCD Scale ($t\bar{t}$)	No	Scale uncertainty for NLO $t\bar{t}$ and single top predictions
QCD Scale (V)	No	Scale uncertainty for NNLO W and Z prediction
QCD Scale (VV)	No	Scale uncertainty for NLO diboson prediction
PDF (gg)	No	Parton distribution function (PDF) uncertainty for gg initiated processes ($t\bar{t}$, $t\bar{t}Z$, $t\bar{t}H$)
PDF ($q\bar{q}$)	No	PDF uncertainty for $q\bar{q}$ initiated processes ($t\bar{t}W$, W, Z).
PDF (qg)	No	PDF uncertainty for qg initiated processes (single top)
Madgraph Q^2 Scale ($t\bar{t}$+0p,1p,2p)	Yes	Madgraph Q^2 scale uncertainty for $t\bar{t}$+jets split by parton number. There is one nuisance parameter per parton multiplicity and they are uncorrelated.
Madgraph Q^2 Scale ($t\bar{t}$+b/$b\bar{b}$/$c\bar{c}$)	Yes	Madgraph Q^2 scale uncertainty for $t\bar{t}$+b/$b\bar{b}$/$c\bar{c}$.
Madgraph Q^2 Scale (V)	No	Varies by jet bin.
Extra $t\bar{t}$+hf rate uncertainty	No	A 50% uncertainty in the rate of $t\bar{t}$+b, $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$.
τ Energy Scale	Yes	Tau signal and background
τ ID efficiency	Yes	Tau signal and background
τ Jet Fake Rate	Yes	Tau signal and background
τ Electron Fake Rate	Yes	Tau signal and background

Limits on $\mu=\sigma/\sigma_{SM}$ H->leptons

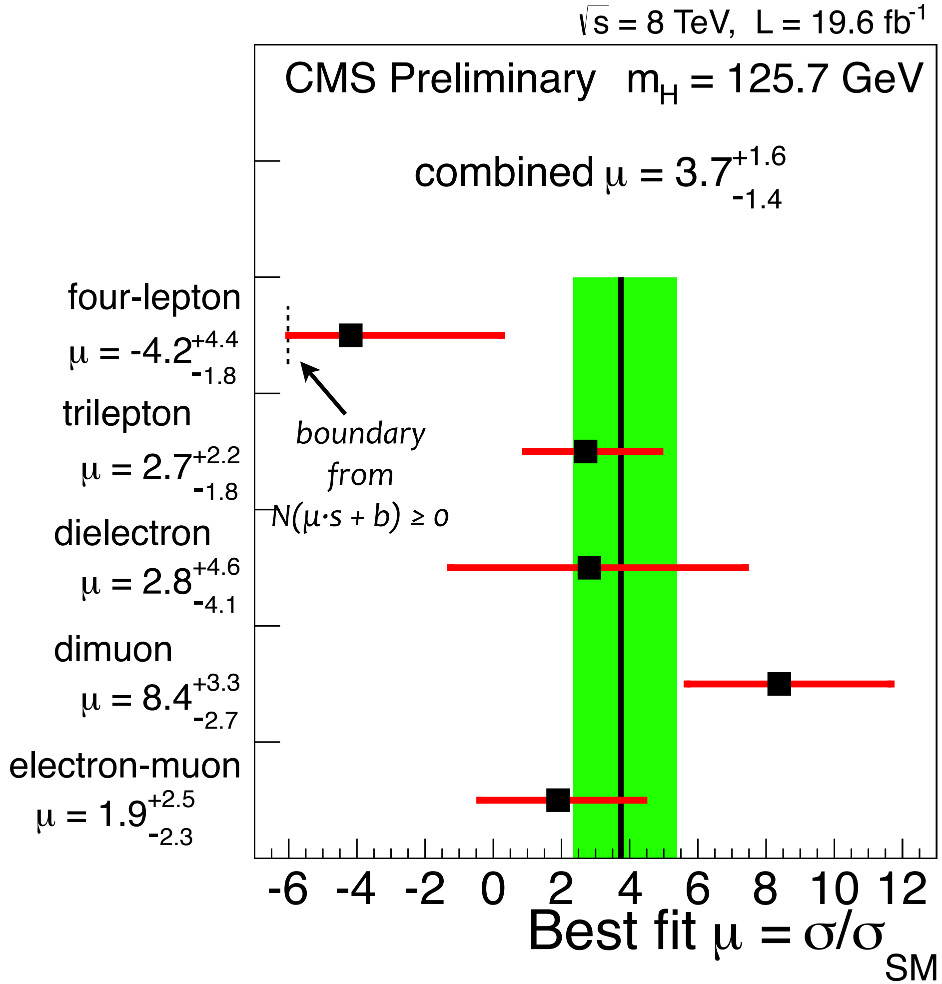
- Observed yields and distributions are compared to the expectation for **a SM Higgs with m_H = 125.7**
- Median expected combined UL:
  - in the absence of ttH signal: **2.4** at 95% CL
  - with the SM ttH production: **3.5** at 95% CL
- Observed UL is **6.6** at 95% CL

mainly driven by the observed excess in the **ss μμ channel**



Best fit $\mu=\sigma/\sigma_{SM}$ H->leptons

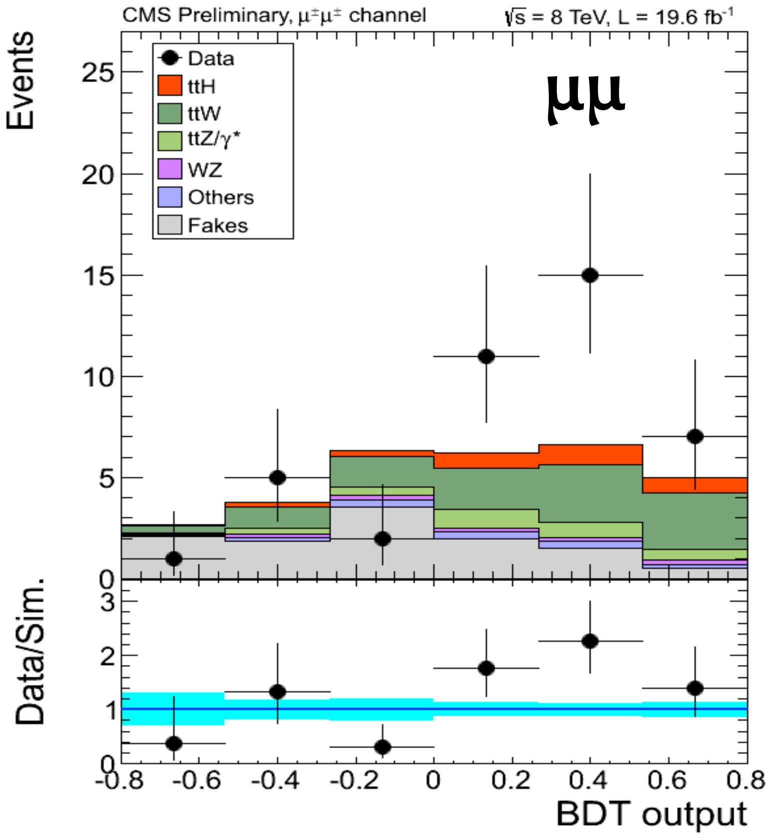
- Observed yields and distributions are compared to the expectation for **a SM Higgs with m_H =125.7**
- The internal consistency of the 5 best-fit signal-strengths with a common value: **16%**
- The fit to the combination yields: **$\mu = 3.7^{+1.6}_{-1.4}$**
- The combined μ is compatible with the SM Higgs boson prediction μ = 1 at the **3%** level
- (The signal extraction is repeated using just **N_{jets}** : the result is compatible with the nominal one but worse sensitivity)



Anatomy of the $\mu^{\pm}\mu^{\pm}$ excess

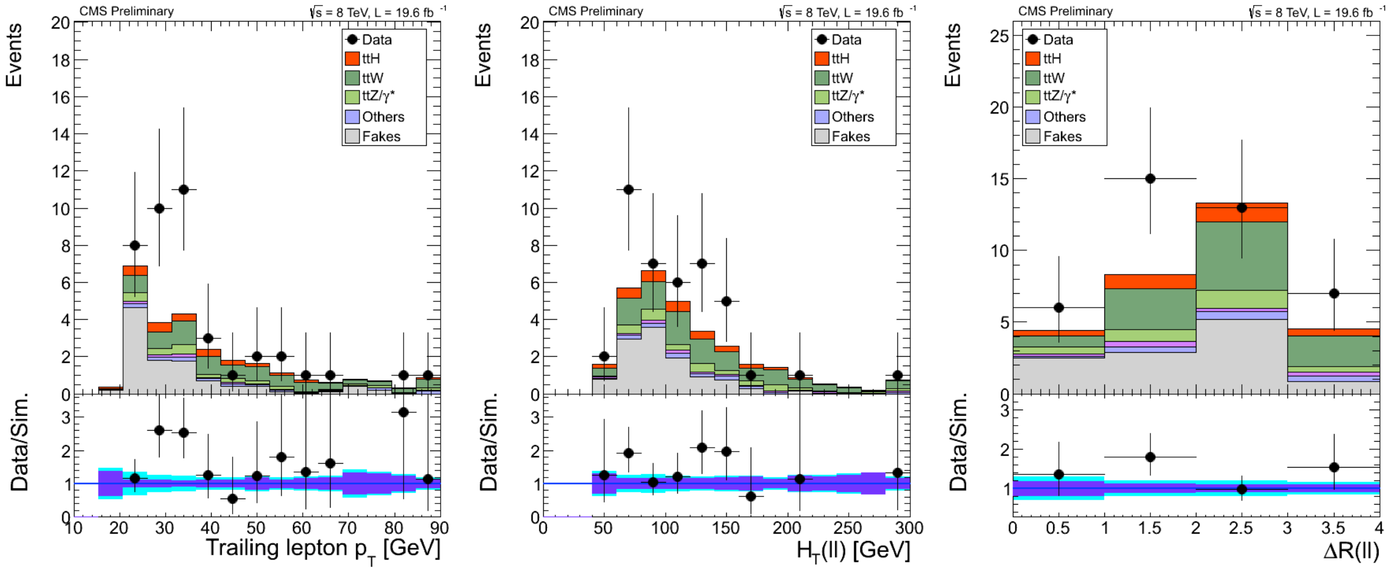
Nominal result

- The results in the different channels are fairly close to the SM Higgs predictions except in the $\mu^{\pm}\mu^{\pm}$ final state
  - **Excess of events** compared to the expectations, in the **signal-like region** of the final BDT discriminator

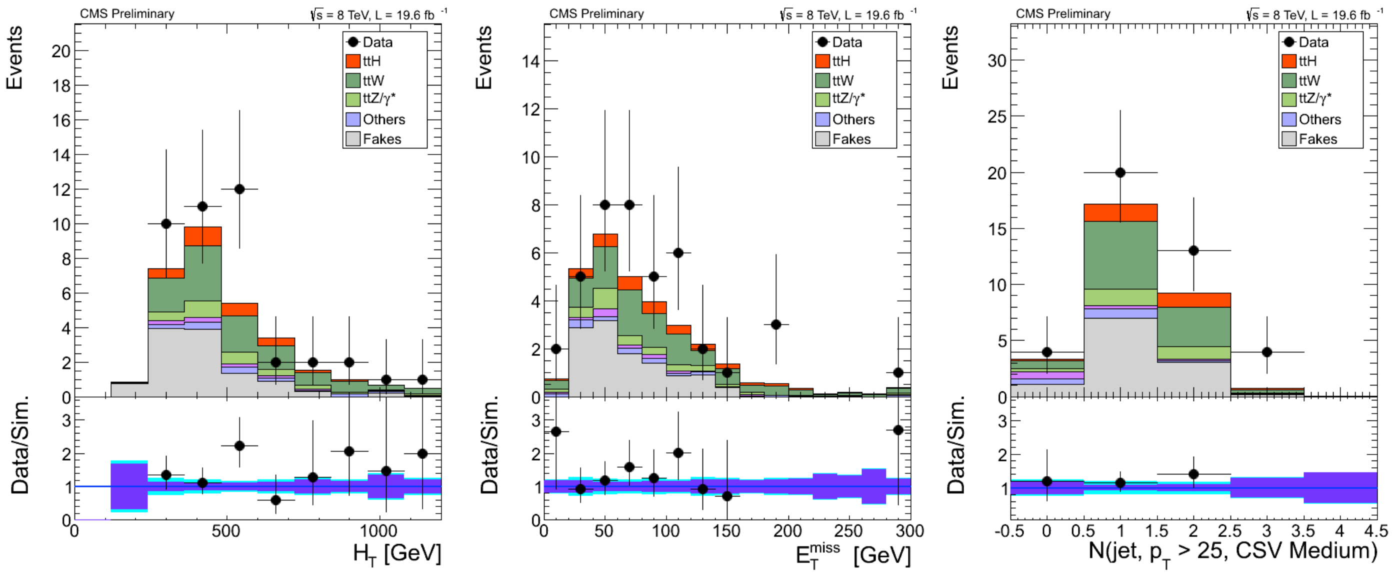


Process	Expected ± syst.
ttH	2.7 ± 0.4
ttW	8.2 ± 1.4
ttZ/γ*	2.5 ± 0.5
WZ	0.8 ± 0.9
Others	1.4 ± 0.1
Reducible	10.8 ± 4.8
Data	41

Event kinematics



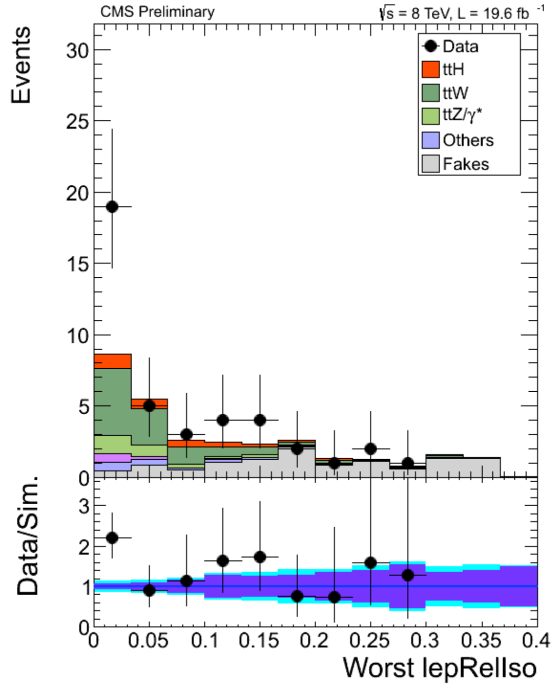
- The **kinematic of the leptons** in the events does not show anomalies and is compatible with that of signal or ttV events



- **Jets and E_T^{miss}** are more compatible with signal or ttV.
- The multeplicity of **b-tags** is also signal-like (while the reducible background has more often only 1 b-tag since the other b-jet is misidentified as a lepton)

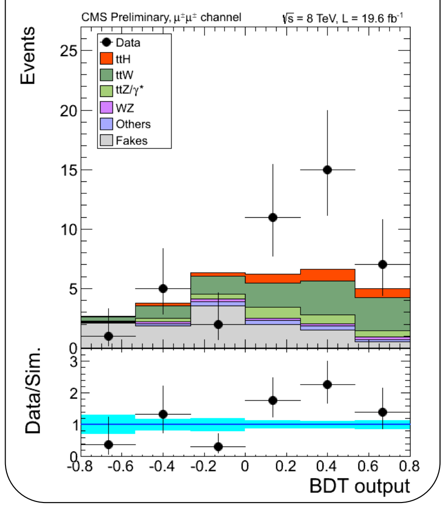
Leptons

- The events in excess are characterized by having both leptons **very well isolated.**
- Scrutiny of the events also confirms that both leptons are **well reconstructed** in the tracker and muon system, and that their charge is correctly assigned
- The analysis was also repeated using a **looser working point of the lepton MVA**
  - the excess is visible only when both leptons pass the tight MVA wp
  - the rest of the sample is well described by the background model
- The analysis was also repeated with a **cut-based muon selection.** The result is compatible with the nominal one but the sensitivity is worse

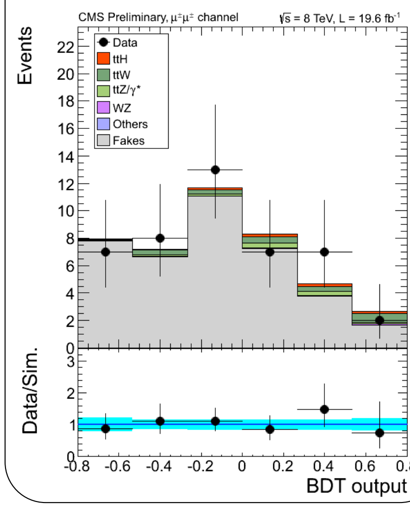


both muons pass loose MVA WP

both muons pass the tight MVA WP



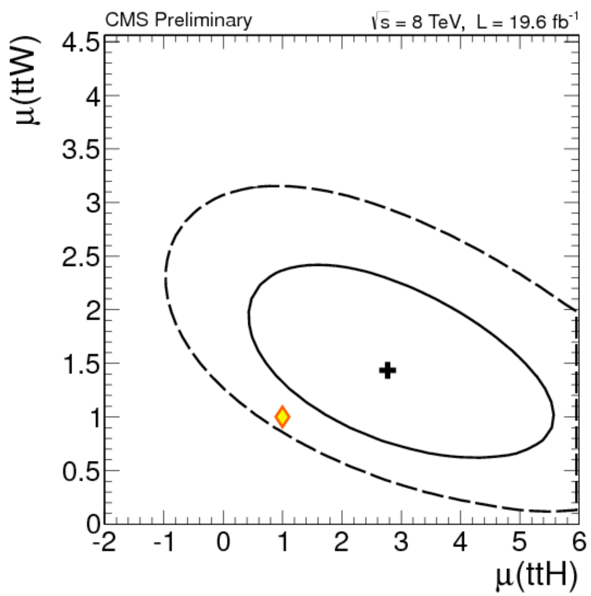
at least one muon fails tight MVA WP



Irreducible bkg check

- A **more general fit** is performed:
  - leaving unconstrained the yields of ttW, ttZ, and reducible background (for fake e, μ separately)
  - including additional control regions in the fit: trilepton events with one Z candidate (mostly ttZ), and dilepton events with 3 jets (ttW & red. bkg.)
- Results **compatible with the nominal ones** (but ~20% worse sensitivity)
- All backgrounds yields remain **within 1σ from their input value:** no indication of issues with ttW & ttZ
  - results for ttH and ttW are correlated, all the others are well resolved

parameter	expected	observed
$\mu(\mathrm{ttH})$	$1.0^{-1.3}_{+1.5}$	$2.8^{-1.6}_{+1.8}$
$\mu(\mathrm{ttW})$	$1.0^{-0.5}_{+0.5}$	$1.4^{-0.5}_{+0.6}$
$\mu(\mathrm{ttZ})$	$1.0^{-0.3}_{+0.4}$	$1.1^{-0.3}_{+0.4}$
$\mu(\mathrm{fake}\ \mu)$	$1.0^{-0.3}_{+0.3}$	$0.7^{-0.3}_{+0.4}$
$\mu(\mathrm{fake\ e})$	$1.0^{-0.3}_{+0.3}$	$0.9^{-0.3}_{+0.3}$



Projections

Projection for LHC and HL-LHC

CMS Projections for 14 TeV, 300 & 3000 fb^{-1}

- Including bb, $\tau_h\tau_h$, $\gamma\gamma$ & multilept.
  the multilepton is new wrt the ECFA studies.
- Two scenarios:
  1. Pessimistic: keep systematics as in 8 TeV analysis
  2. Optimistic: assume experimental systematics improve as √L, theoretical systematics halved

	300 fb^{-1}		**3000 fb^{-1}**	
	scenario 1	*scenario 2*	*scenario 1*	***scenario 2***
Δy_t from ggH	8%	6%	5%	3%
Δy_t from ttH	12%	9%	8%	4%
without multilepton →			10%	7%

ttH leptonic analysis at Run II

- Current analysis designed for 20 fb^{-1} at 8 TeV
  - inclusive selection to **preserve signal efficiency**
  - main focus to **contain the reducible bkg** (no serious attempt to separate ttV/ttH)
- **The picture will change at 14 TeV and high luminosity**
  - the ttH cross section will increase by a factor 5
    - 2.6 for gg and VBF
    - **ttH is the Higgs production channel already interesting with the first year data**
  - higher event yields will allow isolating categories of higher S/B events, or whose **kinematic** can be **reconstructed**