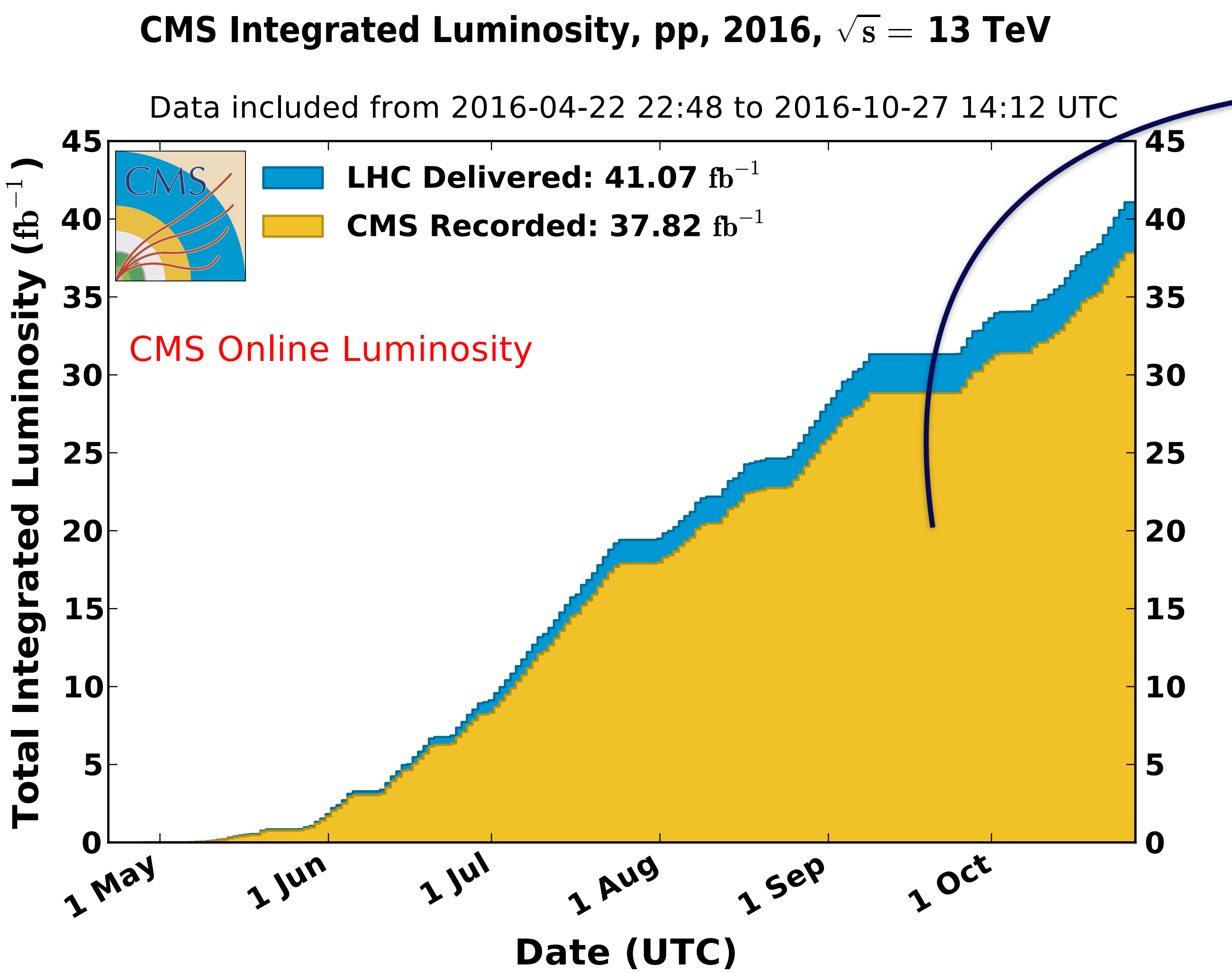
Imperial College London

Status of analyses with Higgs + dijets at CMS

Ed Scott, on behalf of the CMS collaboration

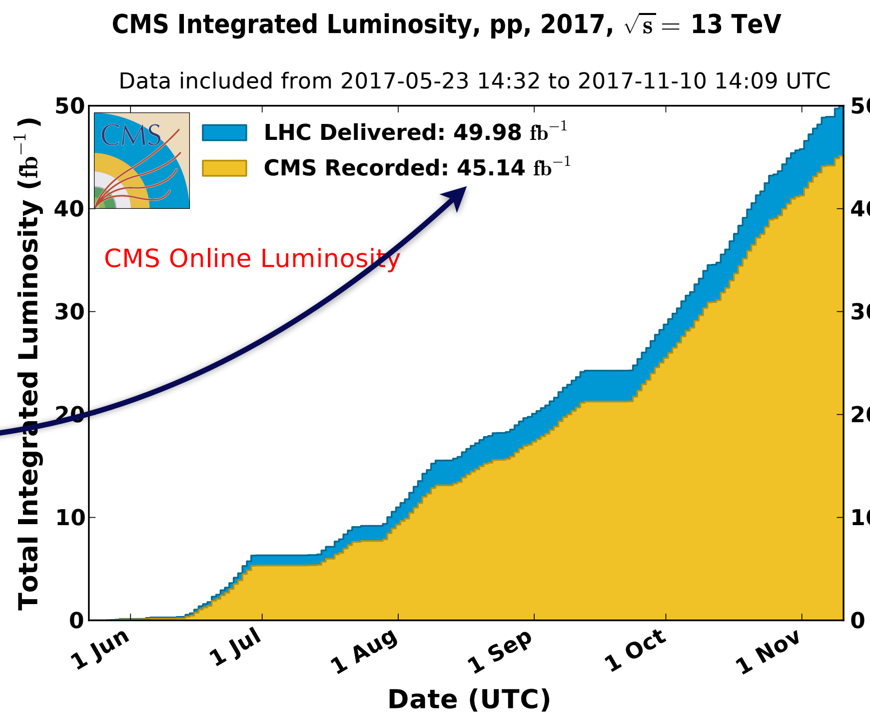
IPPP Higgs + dijets workshop, 10^{th} January 2018

Where are we?

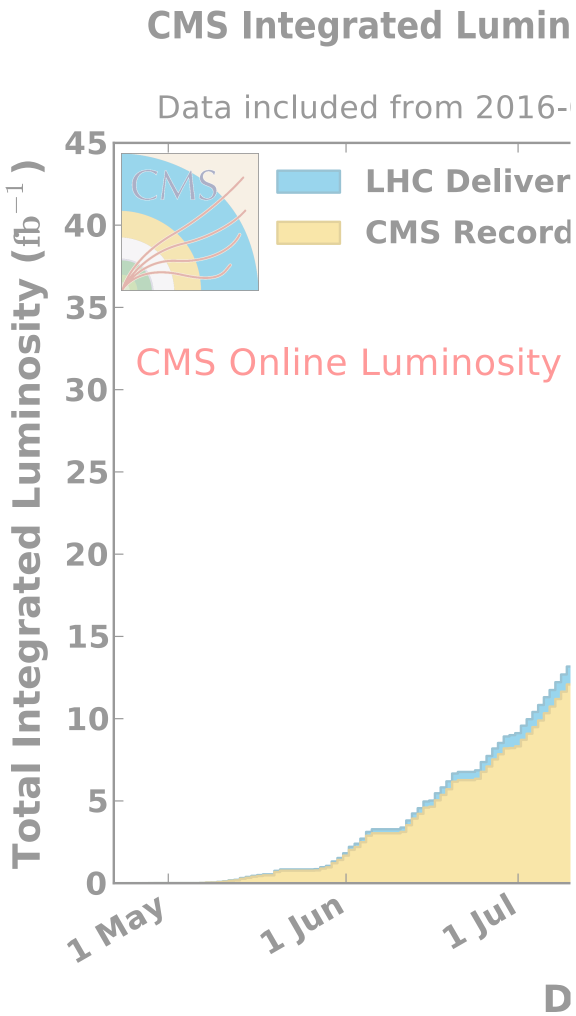


- First full year of Run 2 data-taking in 2016: collected $35.9fb^{-1}$
  - most Higgs analyses now public with this dataset

- An even larger dataset collected in 2017



Where are we?



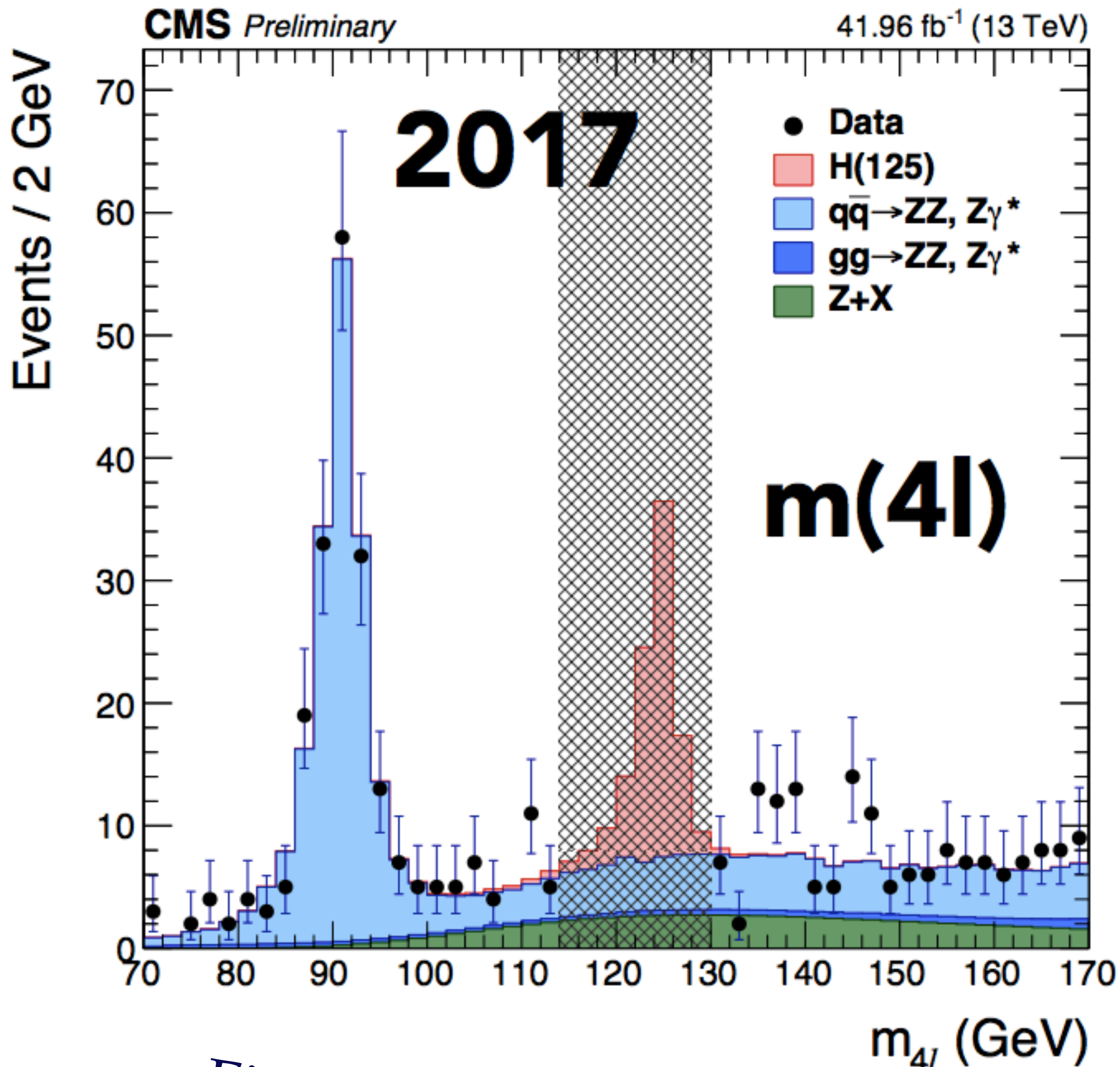
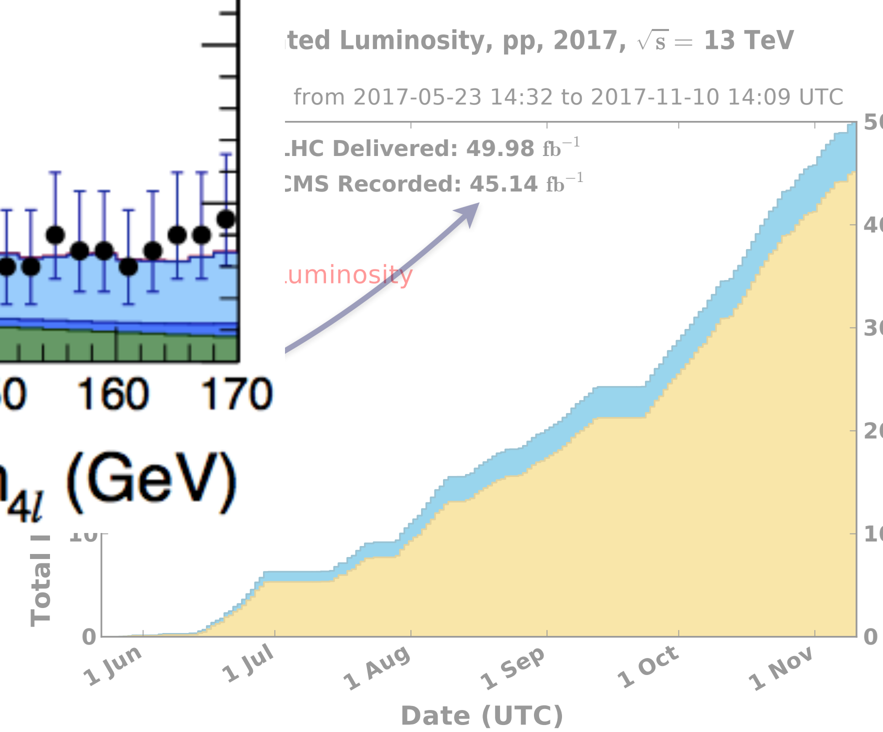


Figure from here



n 2 data-taking
.9fb
yses now public

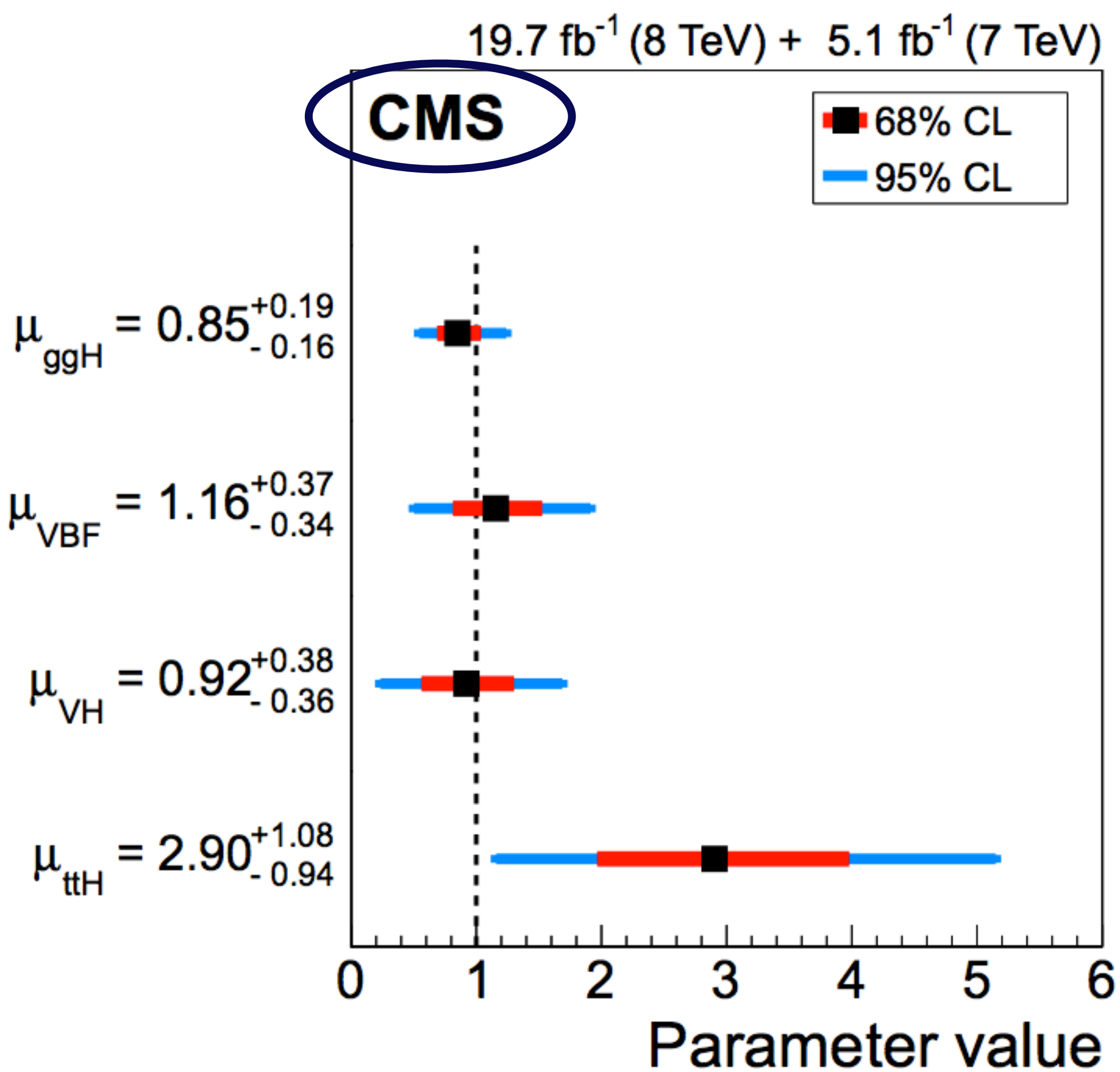
passing Run 1
st cases

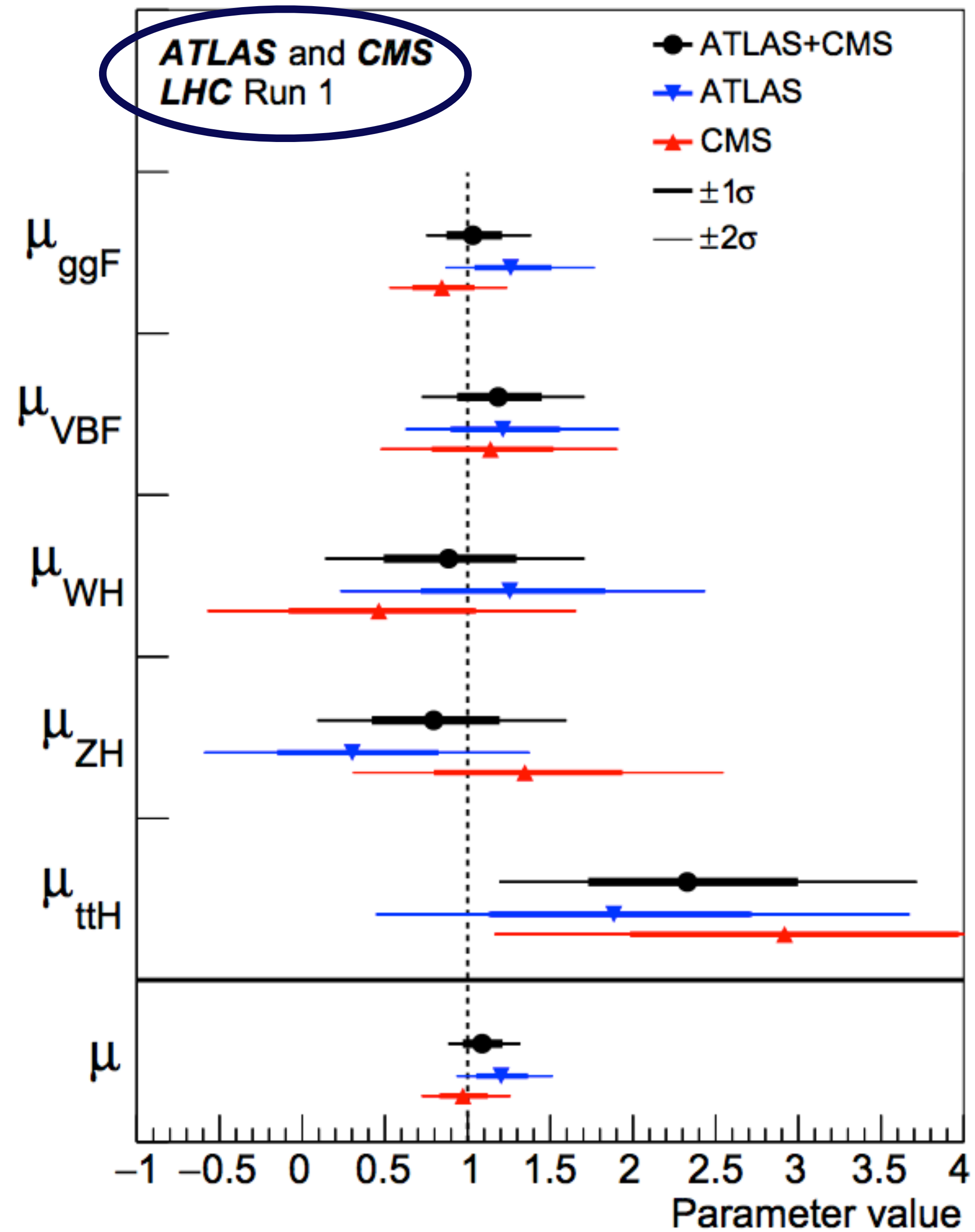
n the works

- An even larger

Where did we come from?

- CMS combination of Run 1 results
- Includes γγ, ZZ, WW, ττ and bb channels
- **Overall measurement of μ_{VBF} was**
  μ_{VBF} = 1.16 +0.37 -0.34





- Most precise results from ATLAS + CMS couplings combination
- **Value of μ_{VBF} was 1.18 +0.25 -0.23**

Imperial College London

Status of 2016 analyses

CMS

Papers accepted / submitted:

- ★ H→ZZ→4ℓ (arXiv:1706.09936)
- ★ H→ττ (arXiv:1708.00373)
- ◆ VH→bb (arXiv:1709.07497)
- ◆ Boosted H→bb (arXiv:1709.05543)

Physics analysis summary (PAS):

- ★ H→γγ (HIG-16-040, CDS link)
- ◆ H→μμ (HIG-17-019, CDS link)
- ◆ ttH→multi-lepton (HIG-17-004, CDS link)
- ◆ ttH→ττ (HIG-17-003, CDS link)

In preparation:

- ★ H→WW
- ◆ ttH→bb (hadronic and leptonic)

- Four key channels which drive sensitivity to VBF process (★)
- All public except H→WW
- H→γγ paper update expected shortly

Key channels for VBF

Higgs → ZZ: expected uncertainty (2016 dataset) on μ_{VBF} ~ 1.0

- Very low statistics; close to zero background
- One and two jet categories

Higgs → ττ: expected uncertainty (2016 dataset) on μ_{VBF} ~ 0.4

- VBF comprises large fraction of signal, expect ~40 events
- One VBF category for each τ decay mode

Higgs → γγ: expected uncertainty (2016 dataset) on μ_{VBF} ~ 0.5

- Reasonable stats, with ~25 VBF signal events expected
- Three MVA-driven VBF categories

Higgs → WW: expected uncertainty (Run 1) on μ_{VBF} ~ 0.6

Talk outline

Key observables and how we measure them

- The CMS detector
- Definition of observables
- Common uncertainties

Strategies and results of key analyses (ZZ, $\tau\tau$, $\gamma\gamma$)

- Overall analysis strategy
- Methods used to target the VBF signature

Discussion of future plans

- How current measurements will evolve with more data
- And how they will be affected by uncertainties

Measurements

$$\mu_i = \frac{\sigma_i}{(\sigma_i)_{\mathrm{SM}}}$$

$$\kappa_j^2 = \sigma_j / \sigma_j^{\mathrm{SM}}$$

$$\text{or} \quad \kappa_j^2 = \Gamma^j / \Gamma^j_{\mathrm{SM}}$$

- Traditional per-process coupling modifiers μ_i, for i = ggH, VBF, ttH, etc.
- LO-motivated κ framework that modifies Higgs' couplings to SM particles
  - applies for both production and decay
  - additional effective coupling modifiers κ_g and κ_γ describe the loop processes for ggH production and γγ decay respectively

Introduced here and in YR3

Measurements

$$\mu_i = \frac{\sigma_i}{(\sigma_i)_{\text{SM}}}$$

$$\kappa_j^2 = \sigma_j / \sigma_j^{\text{SM}}$$

or $$\kappa_j^2 = \Gamma^j / \Gamma^j_{\text{SM}}$$

- Traditional per-process coupling modifiers μ_i, for i = ggH, VBF, ttH, etc.
- LO-motivated κ framework that modifies Higgs' couplings to SM particles

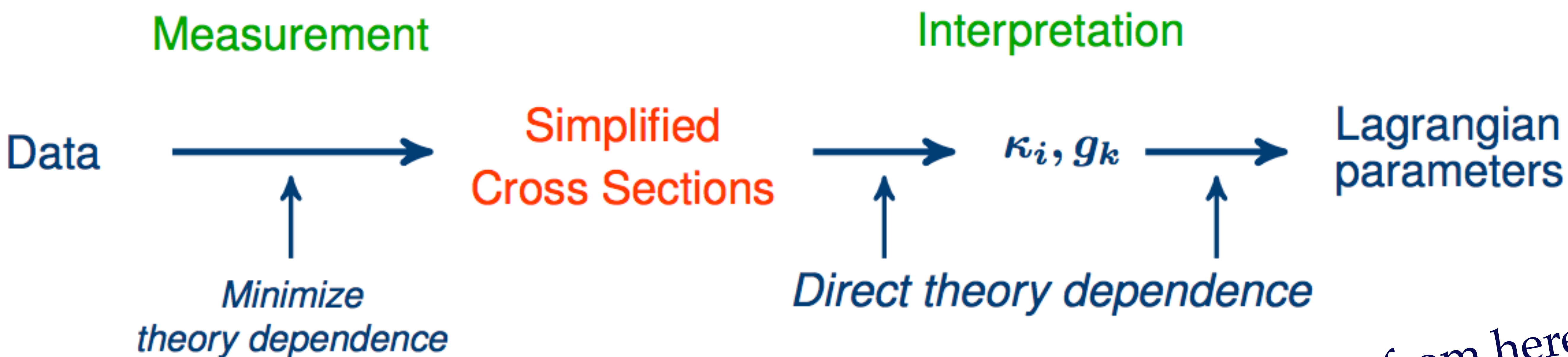
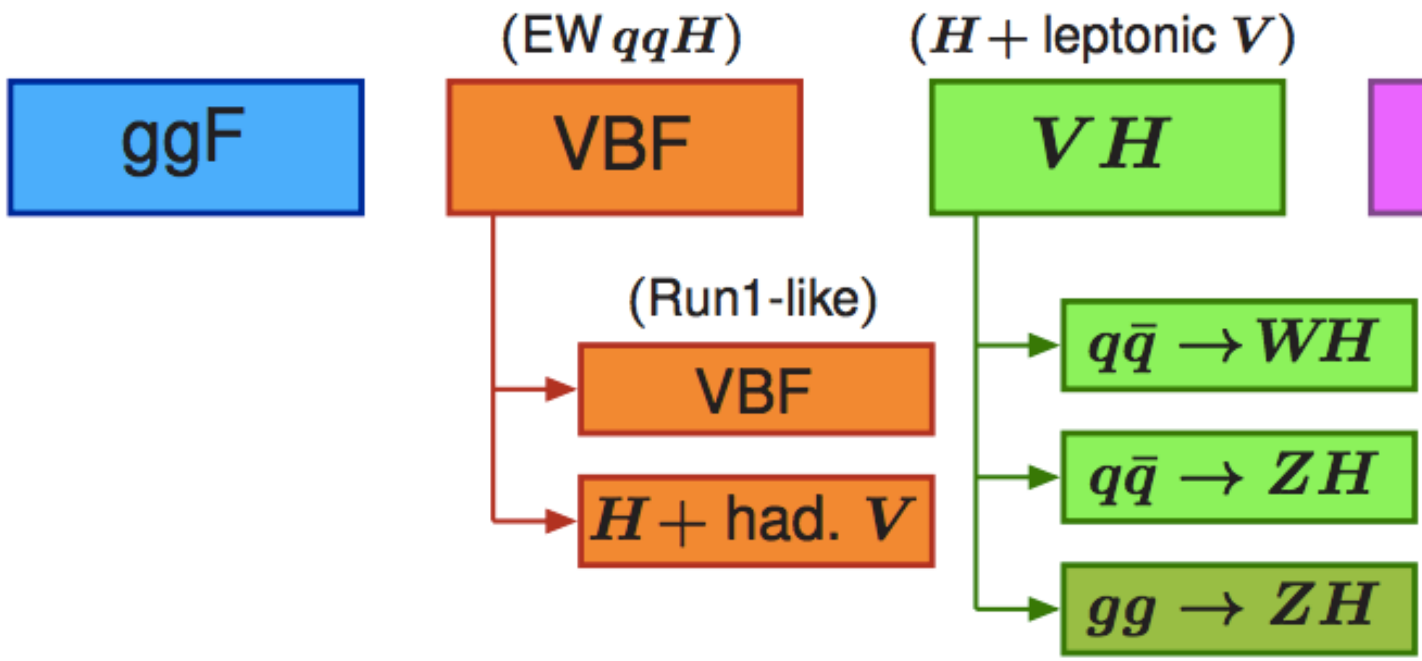


Diagram from here

STXS



$$|Y_H| < 2.5$$

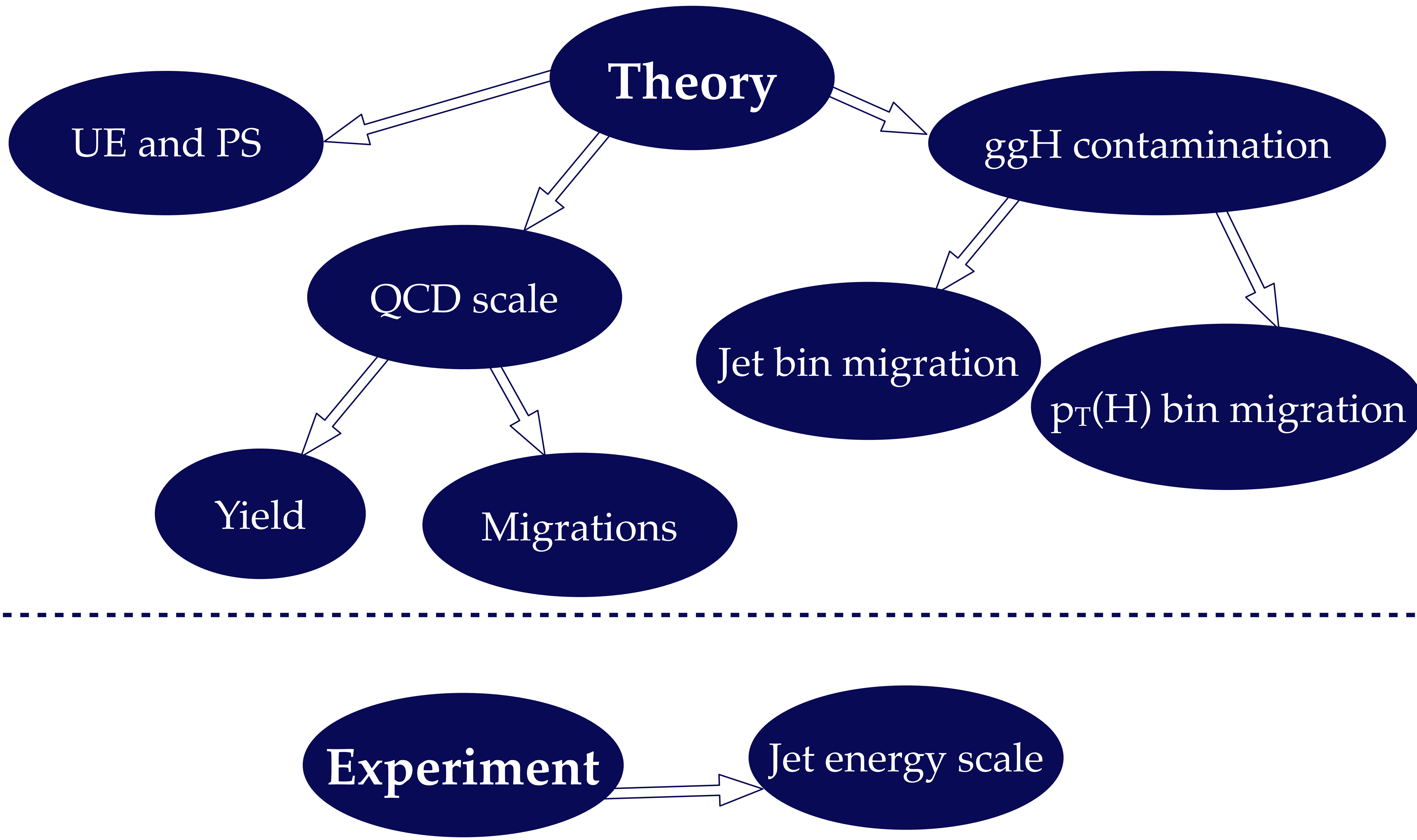
From YR4

- STXS bin definition aims to minimise measurements' dependence on theory
- Useful in long-term, especially for re-interpretation

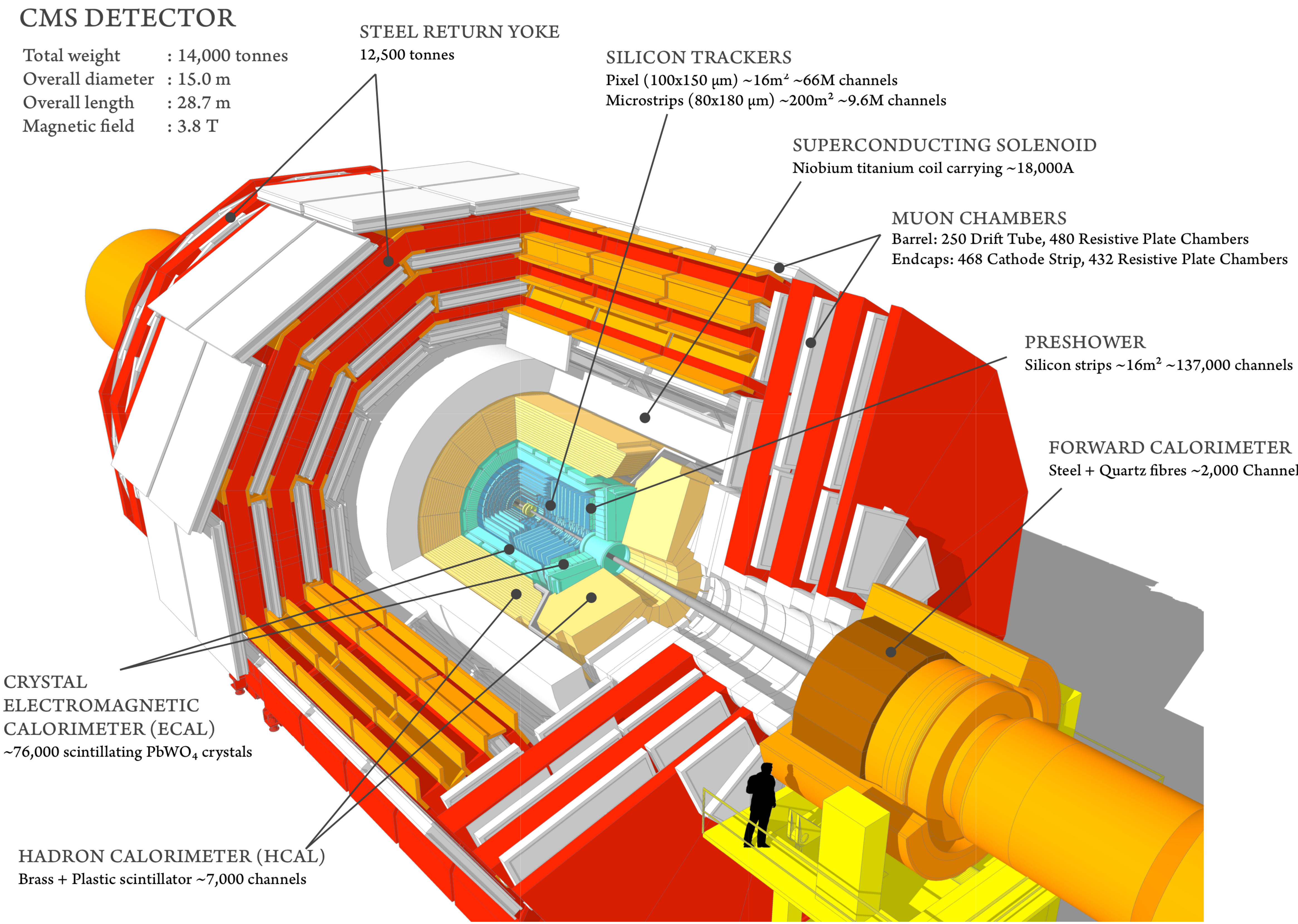
- Stage 0 bins closely mirror Run 1 process definitions
  - CMS results generally include these for the 2016 results
  - theory uncertainties on overall yield factored out of measurement

- With more statistics, move to Stage 1 → finer binning using p_T, nJets

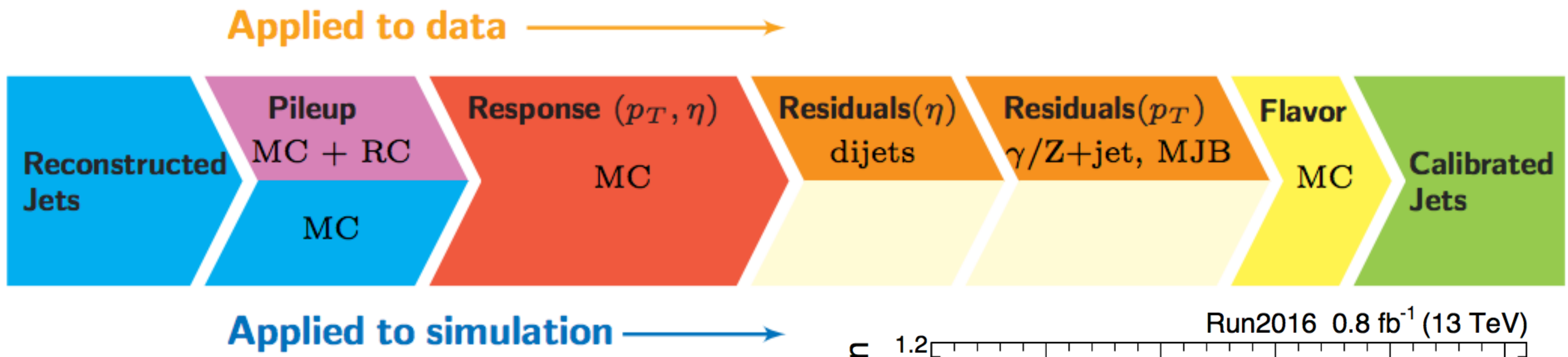
Common systematics



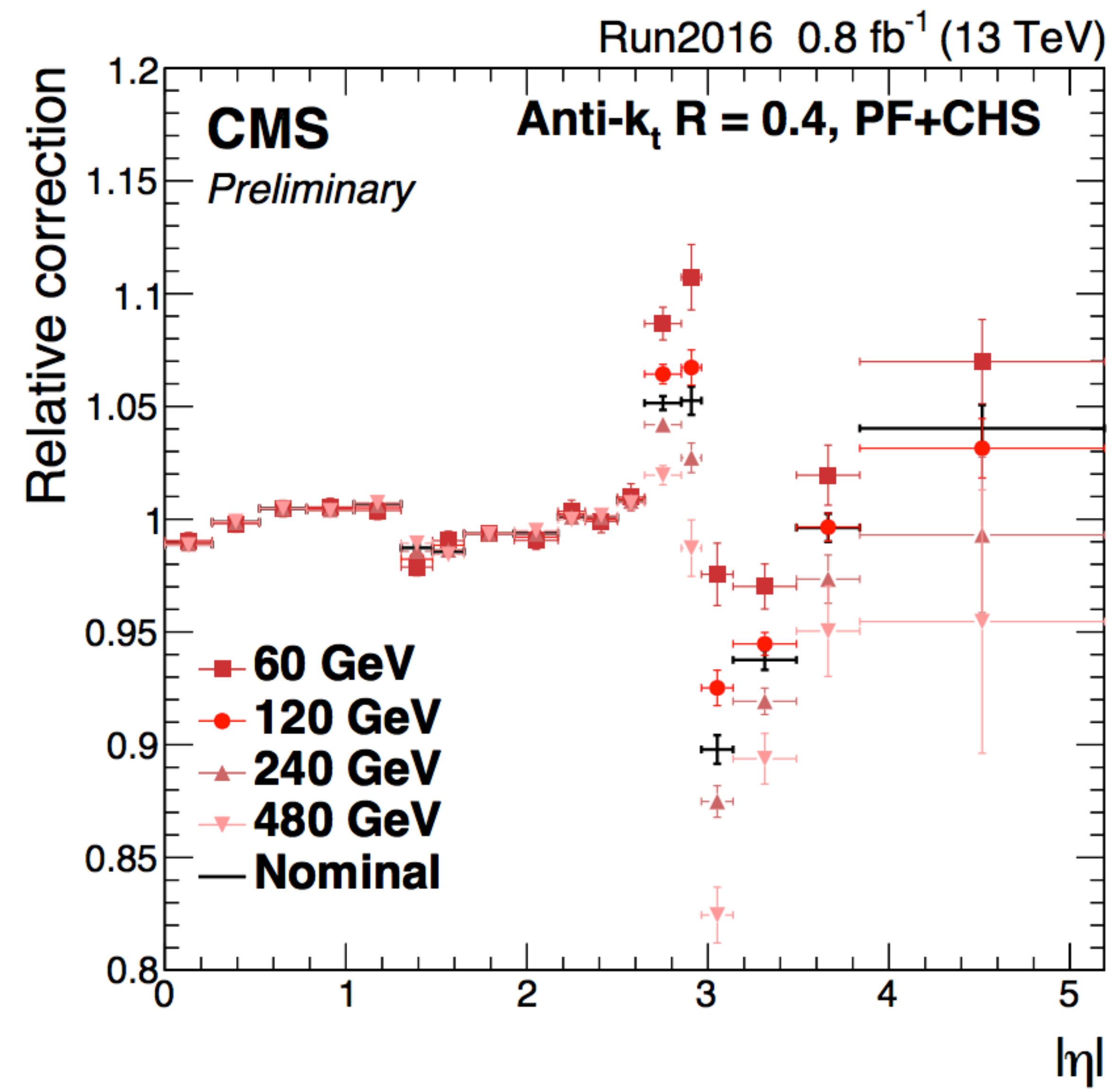
CMS detector



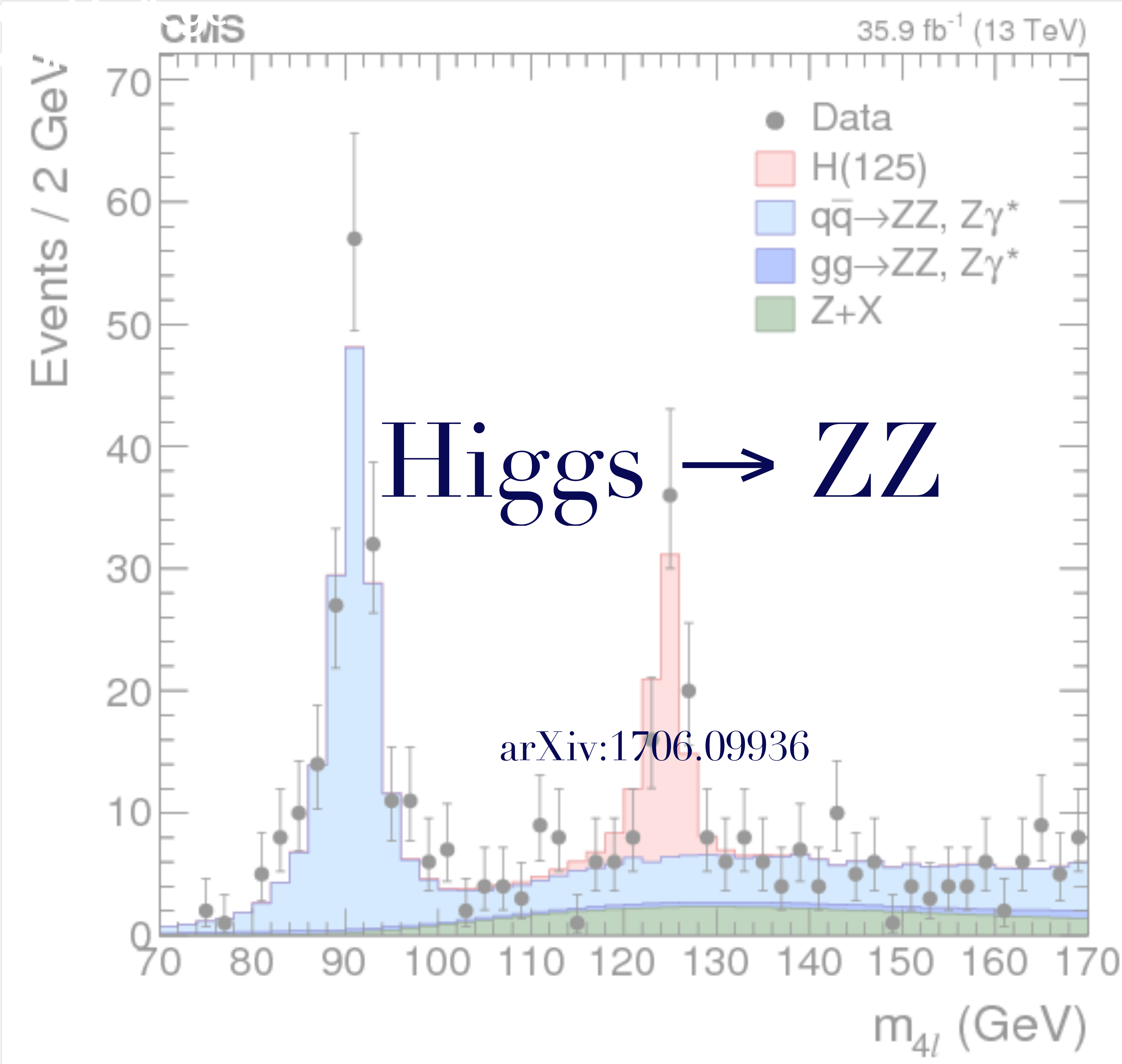
Jet energy scale



- Corrections to the JES applied in several steps for both data and MC
- Associated uncertainties dominate experimental component in current VBF measurements
  - particularly large in the forward region

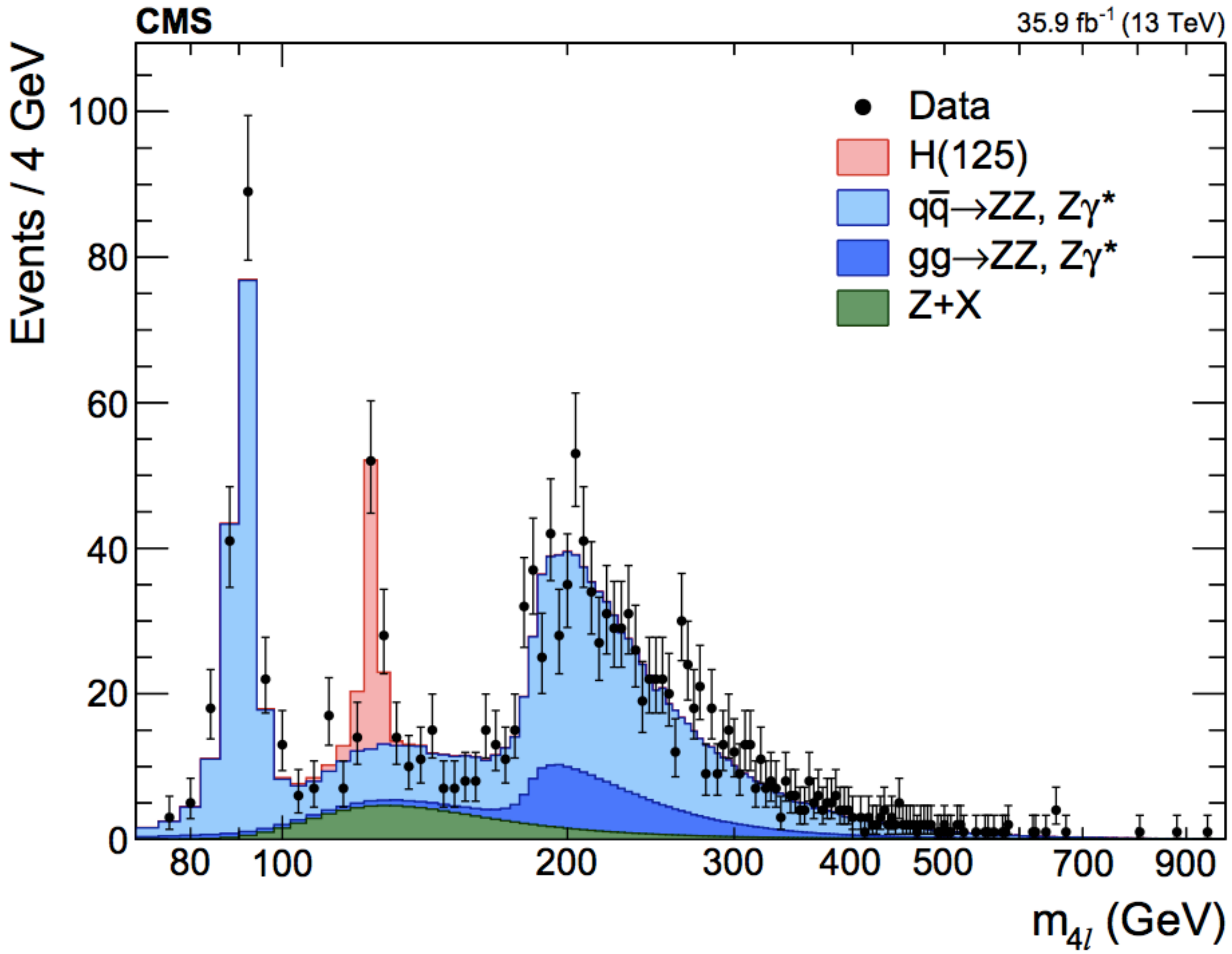


Higgs → ZZ



arXiv:1706.09936

H→ZZ: Overview



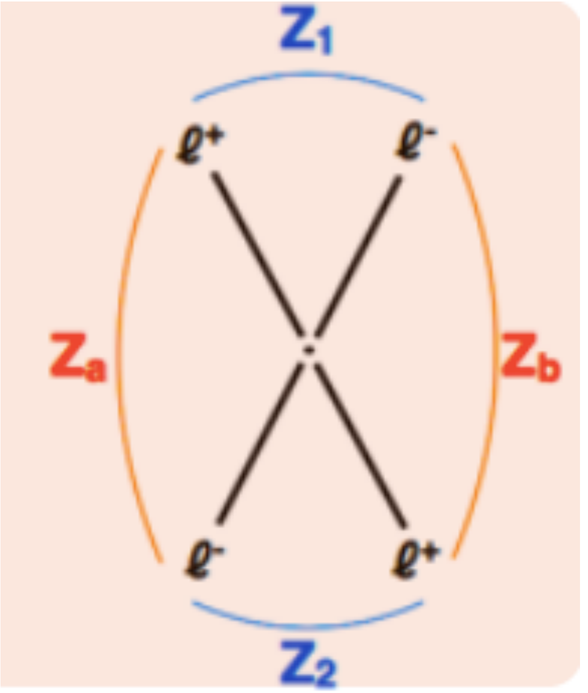
- Search for **H→ZZ→4l (l=e,μ)**
- Clean signature with large S/B
- Narrow resonance gives excellent resolution
- Small background from non-resonant ZZ* production
- Very low statistics (BR ~0.01%)

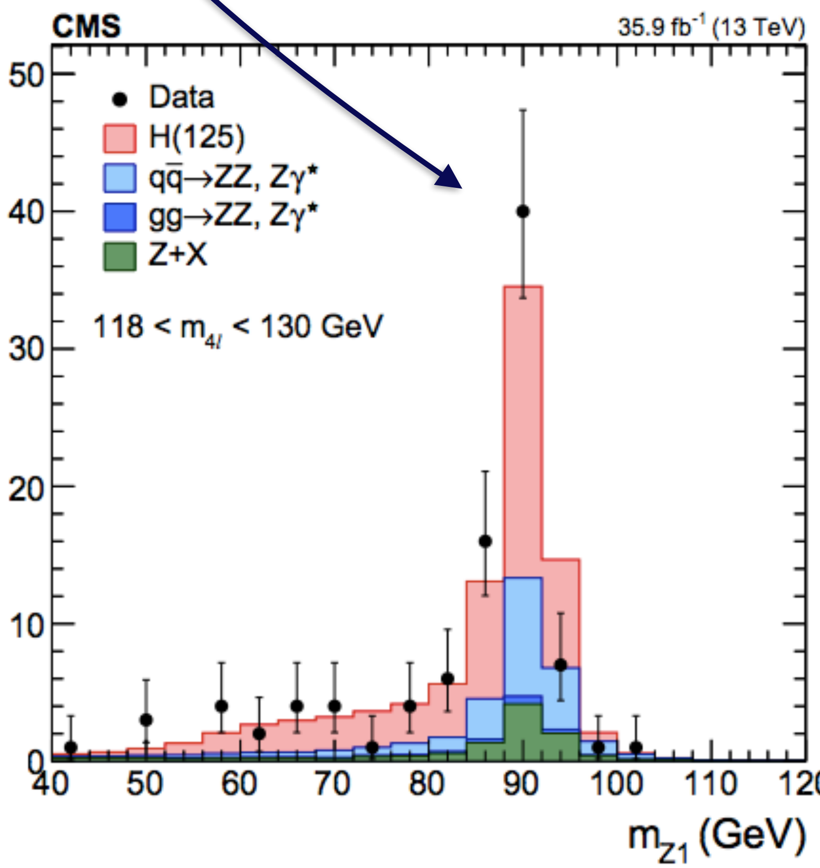
- Select same flavour, opposite sign lepton pairs
- Categorise according to main SM production modes
- Perform 2D likelihood fit to extract signal

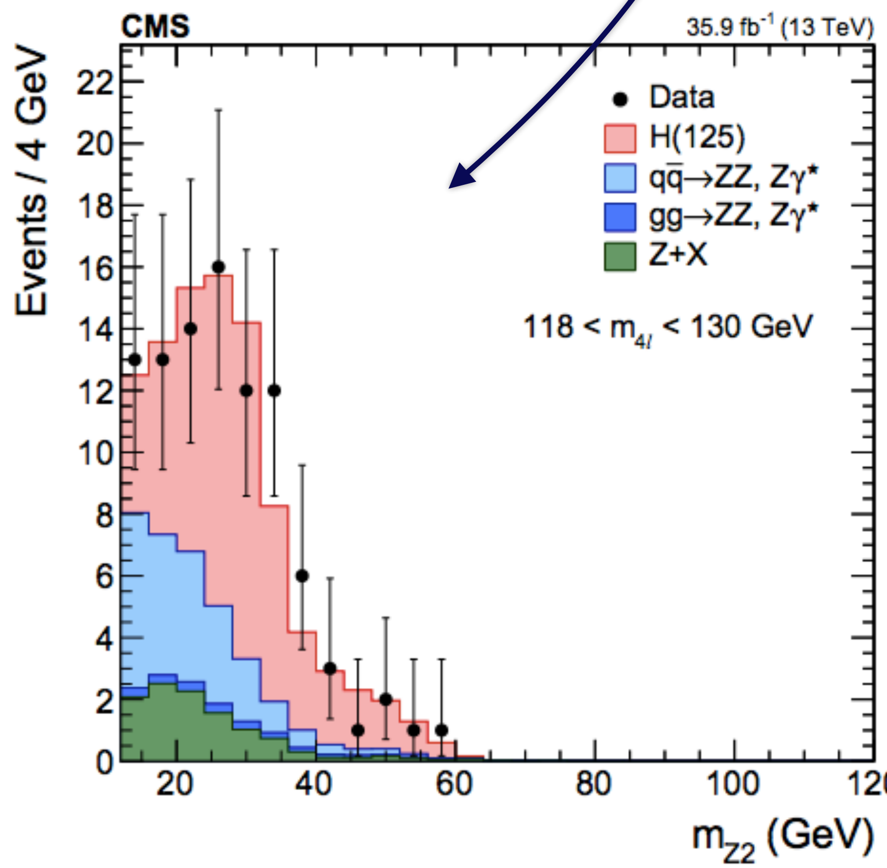
$$\mathcal{L}_{2D}(m_{4\ell}, \mathcal{D}^{\text{kin}}_{\text{bkg}}) = \mathcal{L}(m_{4\ell})\mathcal{L}(\mathcal{D}^{\text{kin}}_{\text{bkg}}|m_{4\ell})$$

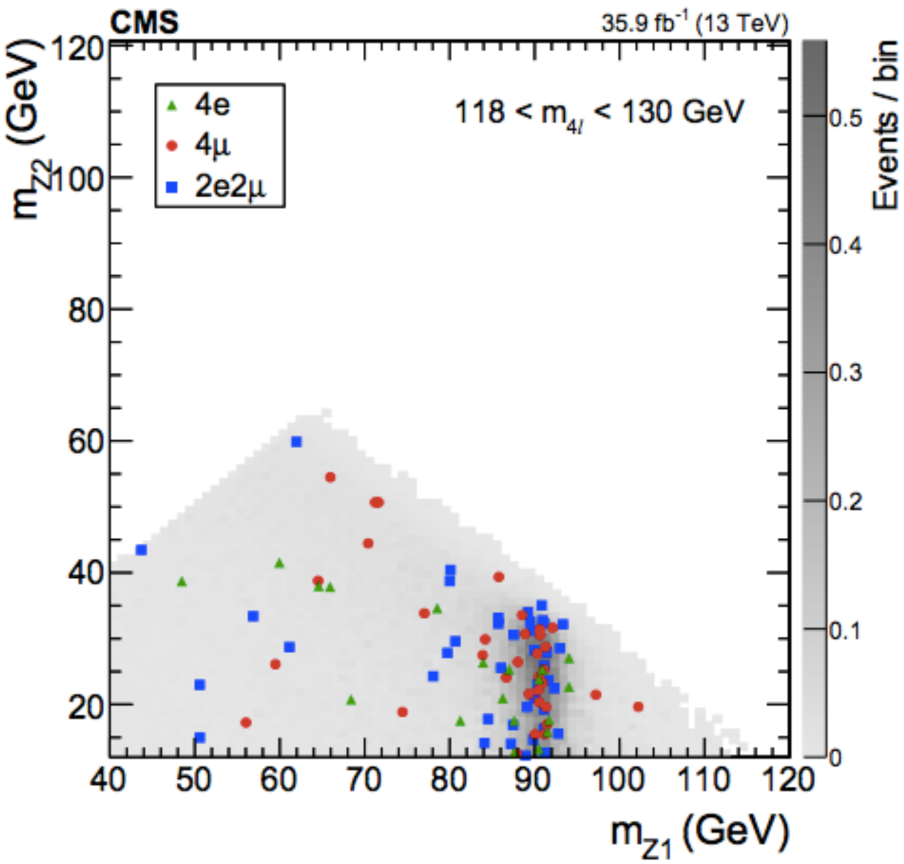
H→ZZ: Signal

- Three categories defined by lepton pairing: **ee, eμ, μμ**
- Identify one lepton pair as on-shell Z, other off-shell
  - both with 12 < m(Z) < 120 GeV
  - daughter electrons (muons) $p_T > 7$ (5) GeV, $|\eta| < 2.4$ (2.5)
  - **FSR photon recovery,** using nearest photon with low $\Delta R/p_T$
  - jets require $p_T > 30$ GeV, $|\eta| < 4.7$









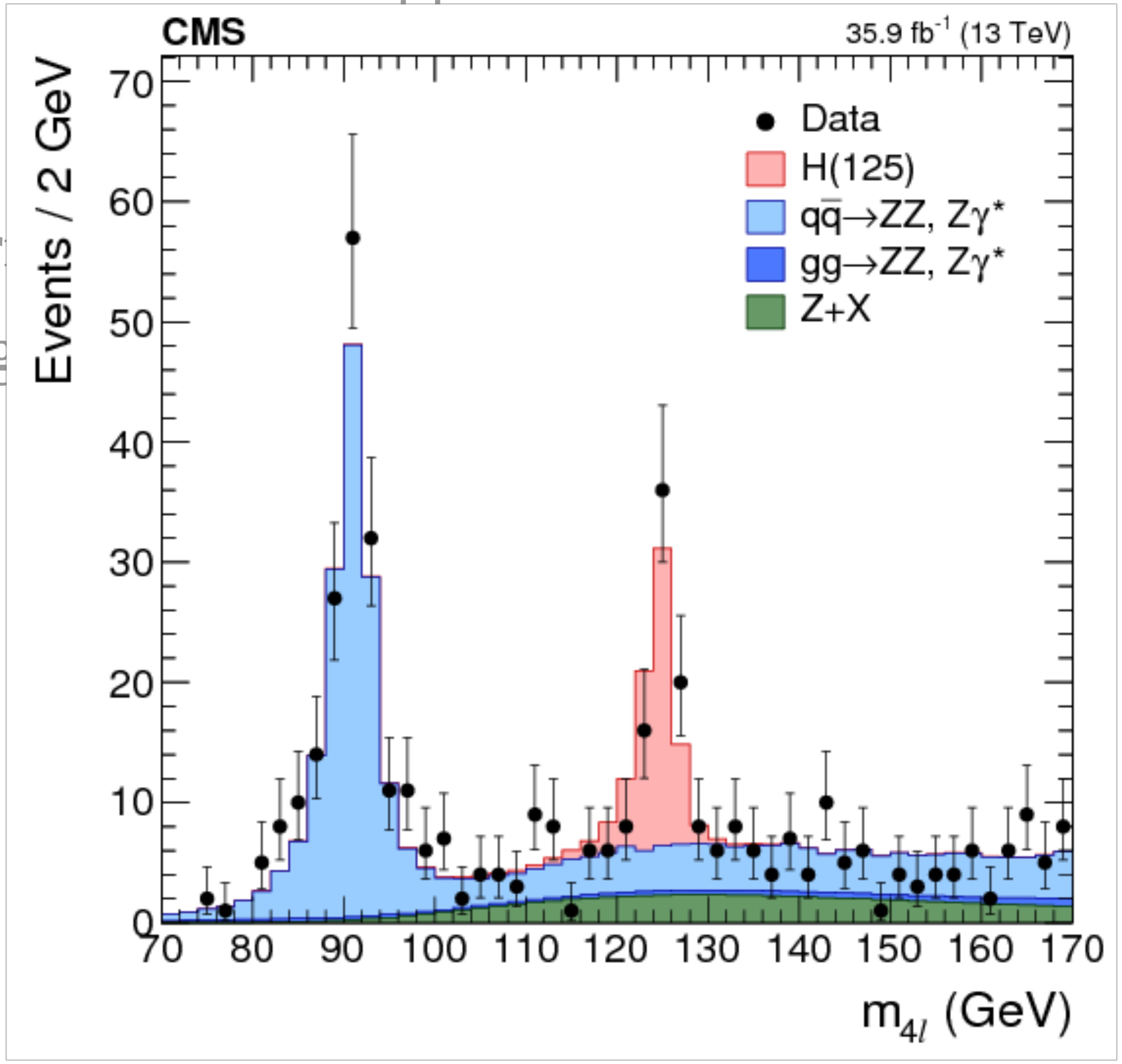
H→ZZ: Backgrounds

- **Irreducible: non-resonant qq→ZZ and gg→ZZ**
  - both obtained from simulation: NLO in pQCD for qq, LO for gg
  - corrected using k-factors (taken from signal for gg→ZZ; additional EW factors for qq→ZZ) - details in paper
  - NNLO/NLO ~1.0-1.2 for qq→ZZ, NNLO/LO ~2.0-2.6 for gg→ZZ
- **Reducible: Z→ll + jets, tt, WZ**
  - **two data-driven methods** using independent control regions
  - one using opposite sign regions, other same sign; take weighted average

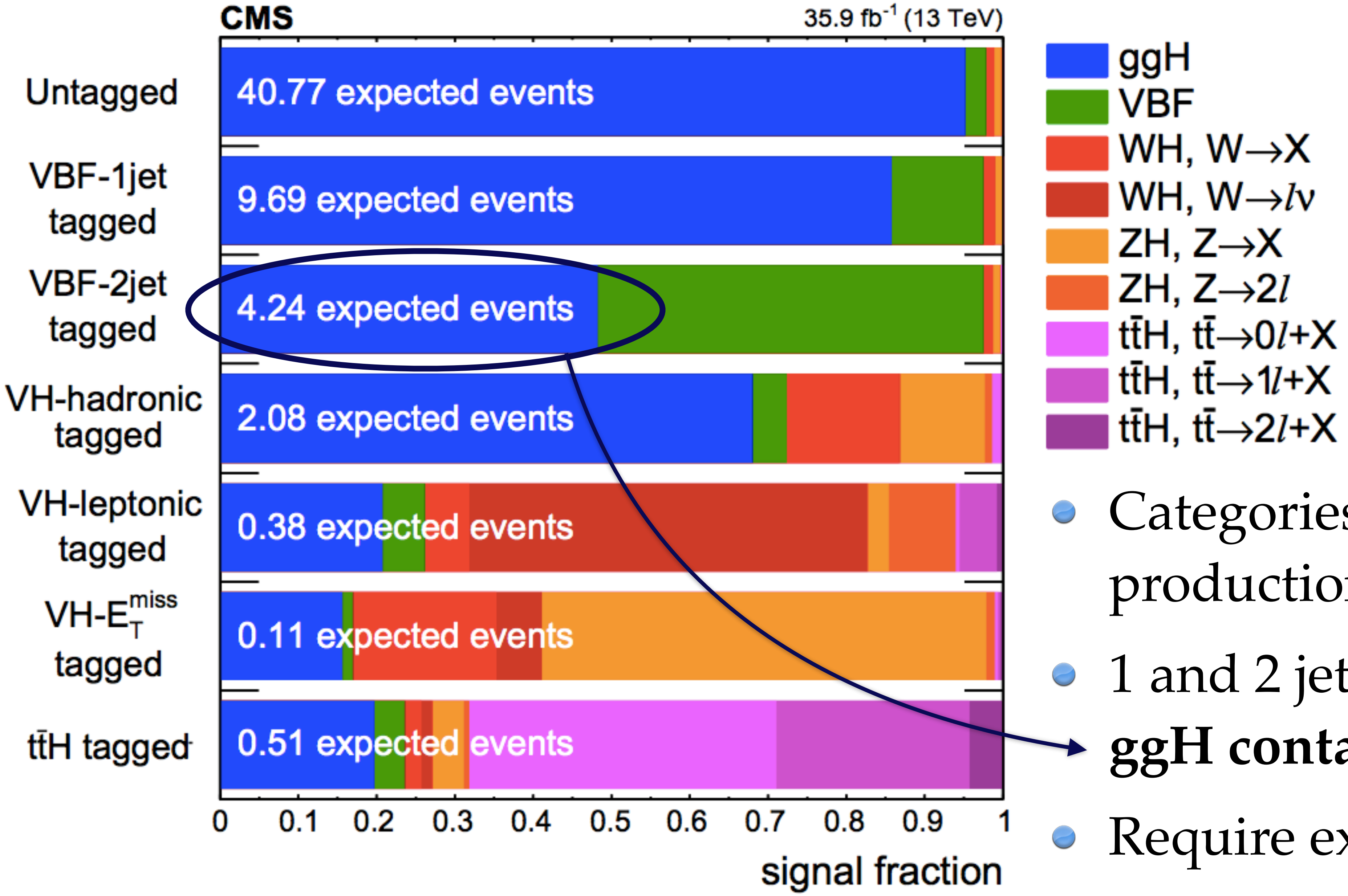
Channel	4e	4μ	$2e2\mu$	4ℓ
$q\bar{q} \to ZZ$	193^{+19}_{-20}	360^{+25}_{-27}	471^{+33}_{-36}	1024^{+69}_{-76}
$gg \to ZZ$	$41.2^{+6.3}_{-6.1}$	$69.0^{+9.5}_{-9.0}$	102^{+14}_{-13}	212^{+29}_{-27}
Z+X	$21.1^{+8.5}_{-10.4}$	34^{+14}_{-13}	60^{+27}_{-25}	115^{+32}_{-30}
Sum of backgrounds	255^{+24}_{-25}	463^{+32}_{-34}	633^{+44}_{-46}	1351^{+86}_{-91}
Signal	$12.0^{+1.3}_{-1.4}$	23.6 ± 2.1	30.0 ± 2.6	65.7 ± 5.6
Total expected	267^{+25}_{-26}	487^{+33}_{-35}	663^{+46}_{-47}	1417^{+89}_{-94}
Observed	293	505	681	1479

H→ZZ: Backgrounds

- **Irreducible: non-resonant qq**
  - both obtained from simulation: NLO for qq, LO for gg
  - corrected using k-factors from signal, as a function of m(4l)
  - NNLO/NLO ~1.0-1.2 for qq
- **Reducible: Z**
  - **two data-dr**
  - **opposite sig**
  - **same sign:**



H→ZZ: Categories

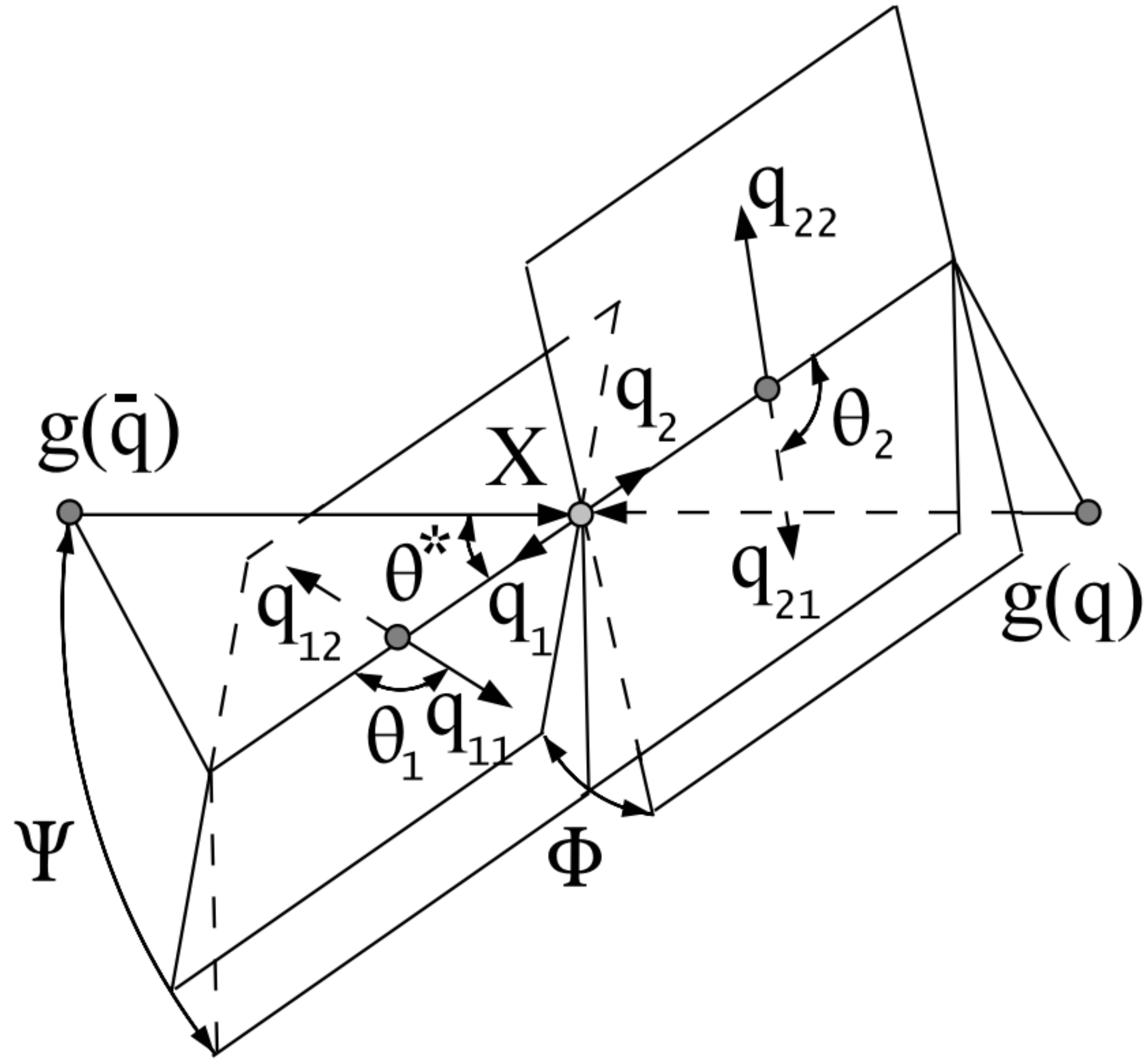


- Categories for each main SM production mode
- 1 and 2 jet categories for VBF; **ggH contamination ~50% in 2j**
- Require exactly 4 leptons in both
- In VBF-2jet category, 2-3 jets with at most one b jet, or >= 4 jets with 0 b jets
- Plus kinematic discriminants

$$\mathcal{D}^{\text{kin}}_{\text{bkg}} = \left[1 + \frac{\mathcal{P}^{q\bar{q}}_{\text{bkg}}(\vec{\Omega}^{\text{H}\to 4\ell}|m_{4\ell})}{\mathcal{P}^{gg}_{\text{sig}}(\vec{\Omega}^{\text{H}\to 4\ell}|m_{4\ell})}\right]^{-1}$$

H→ZZ: Discriminants

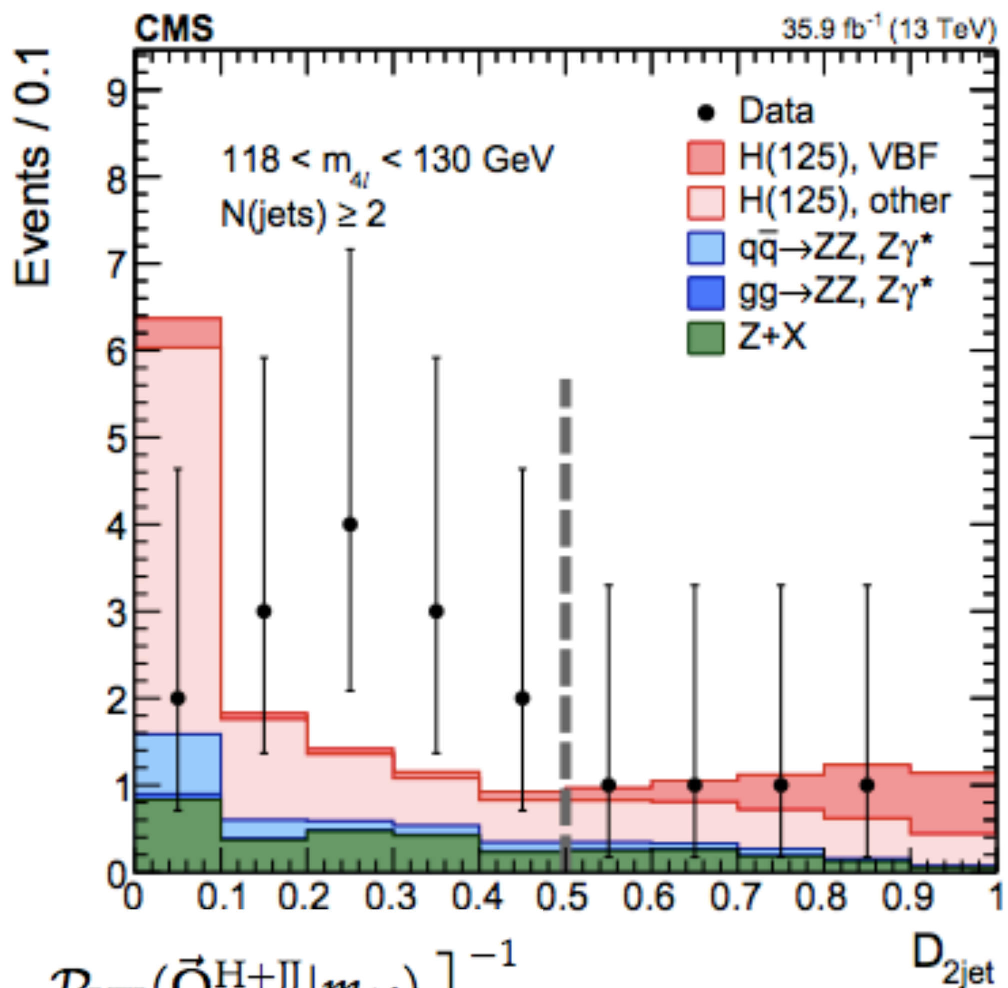
Imperial College London

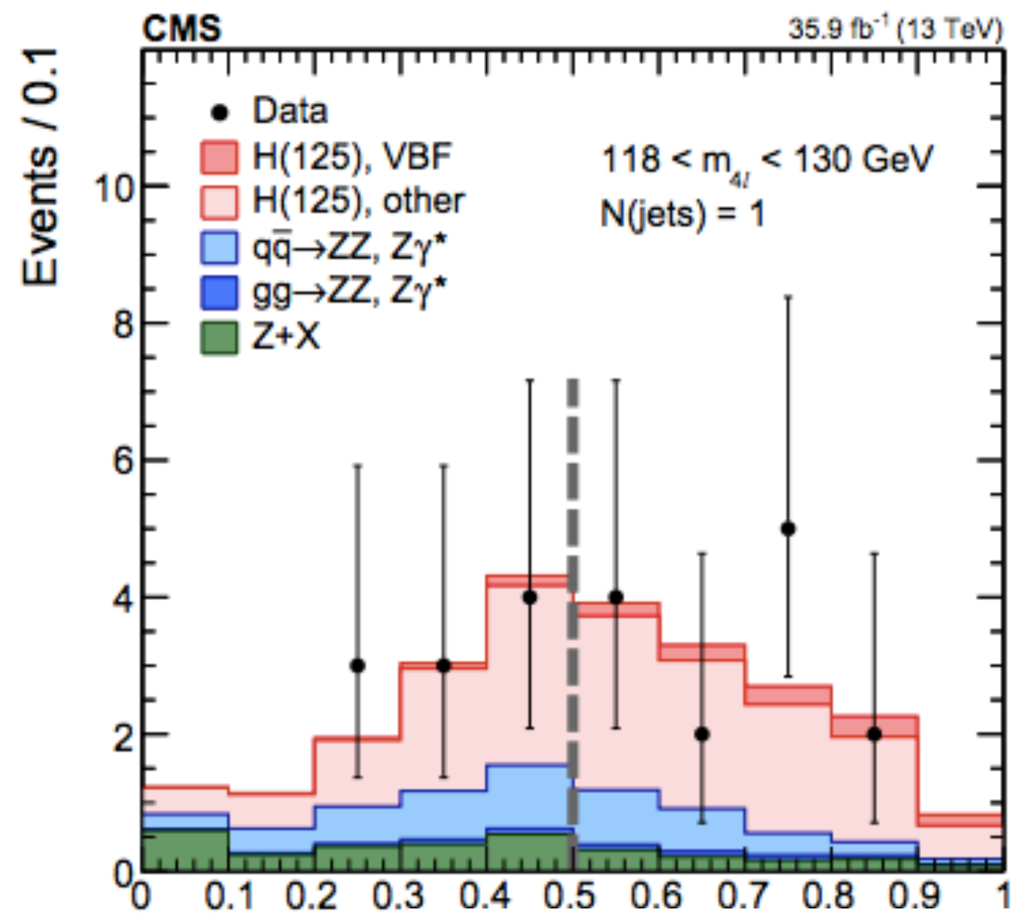


- Kinematic discriminants based on matrix elements
- Inputs rely on the four vectors of each decay product

Imperial College London

H→ZZ: Discriminants



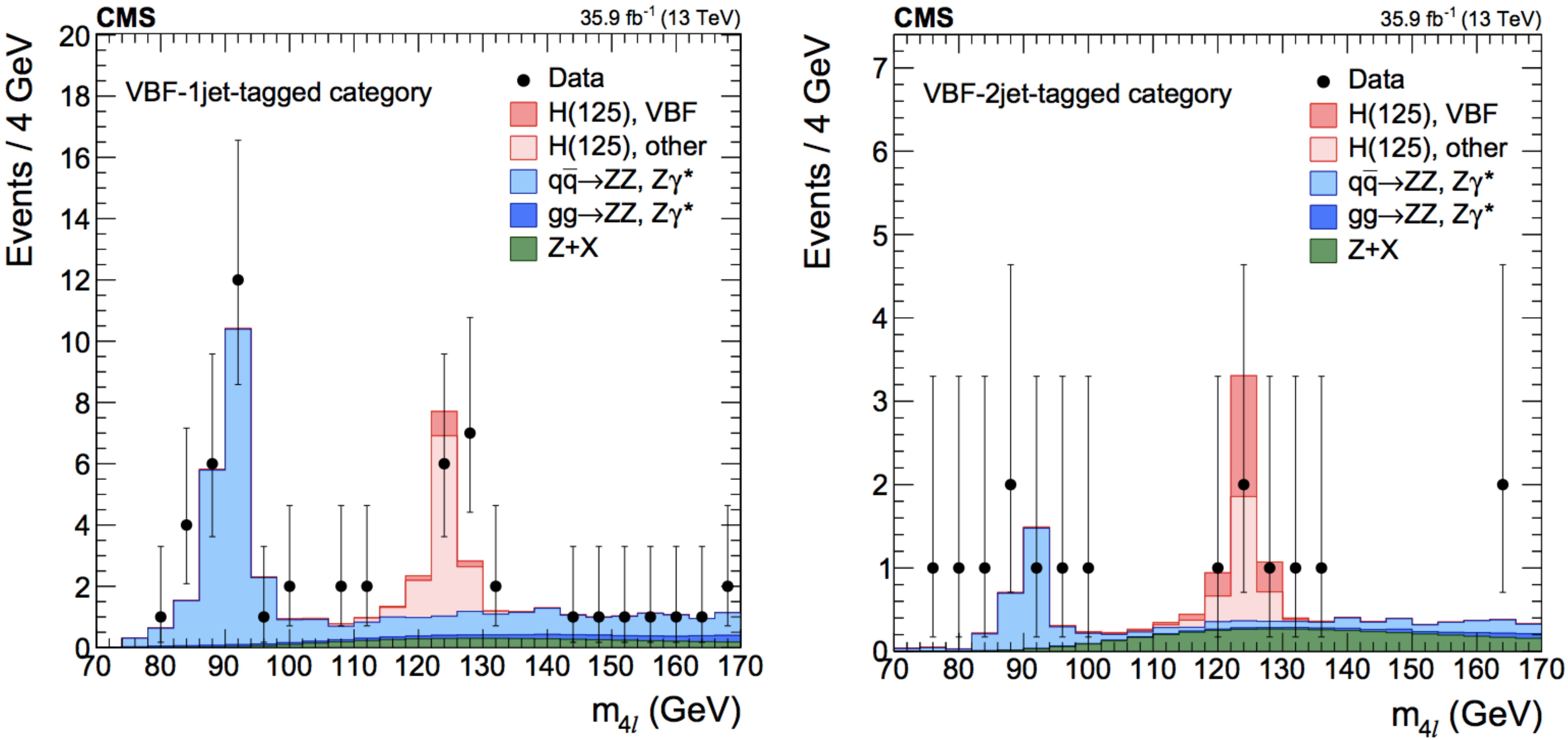


$$\mathcal{D}_{2\,\mathrm{jet}} = \left[1 + \frac{\mathcal{P}_{\mathrm{HJJ}}(\vec{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}{\mathcal{P}_{\mathrm{VBF}}(\vec{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}\right]^{-1}$$

$$\mathcal{D}_{1\,\mathrm{jet}} = \left[1 + \frac{\mathcal{P}_{\mathrm{HJ}}(\vec{\Omega}^{\mathrm{H+J}}|m_{4\ell})}{\int \mathrm{d}\eta_{\mathrm{J}} \mathcal{P}_{\mathrm{VBF}}(\vec{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}\right]^{-1}$$

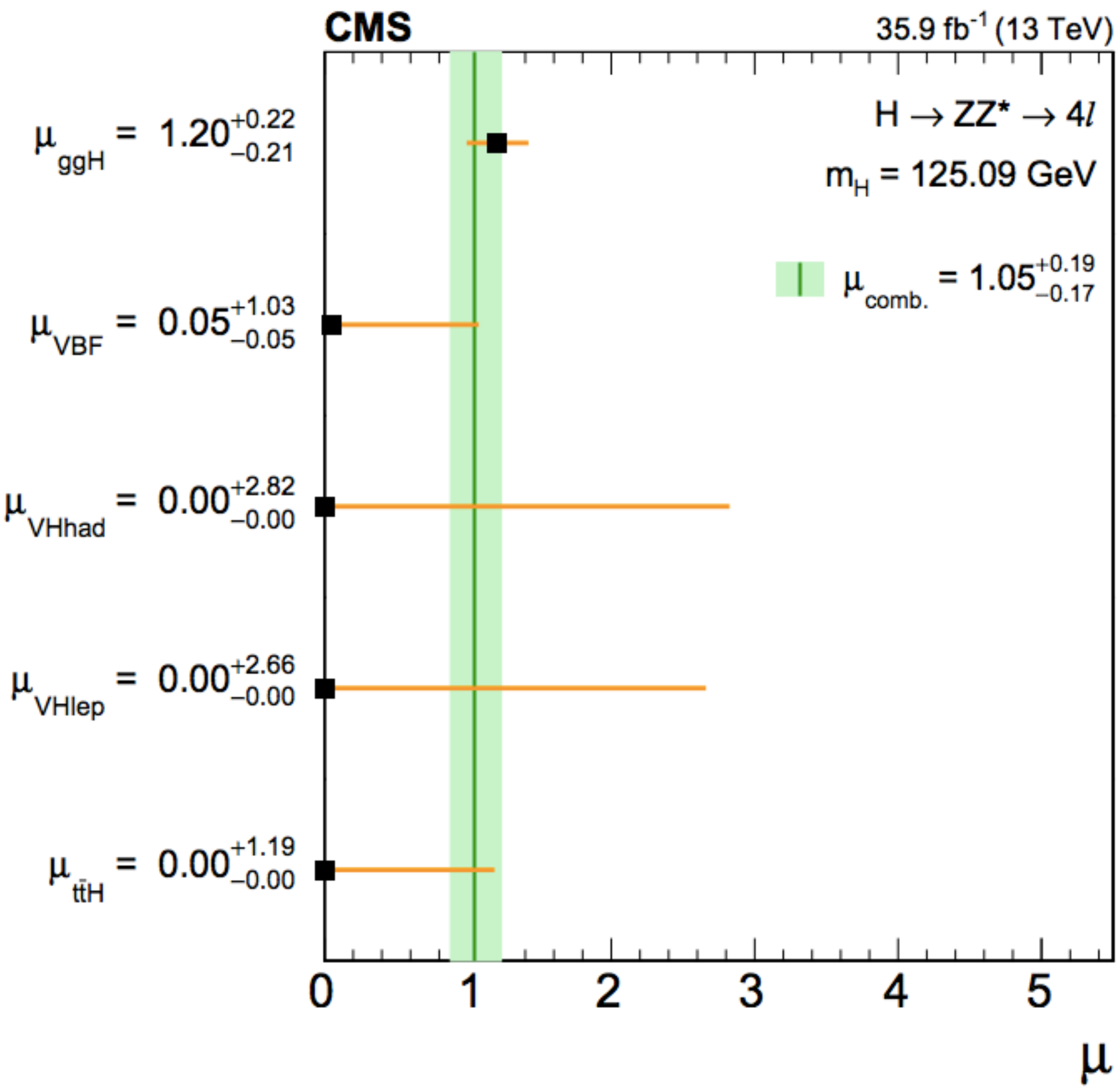
- One discriminant for background rejection, several for production modes
- Cuts placed at 0.5 for both

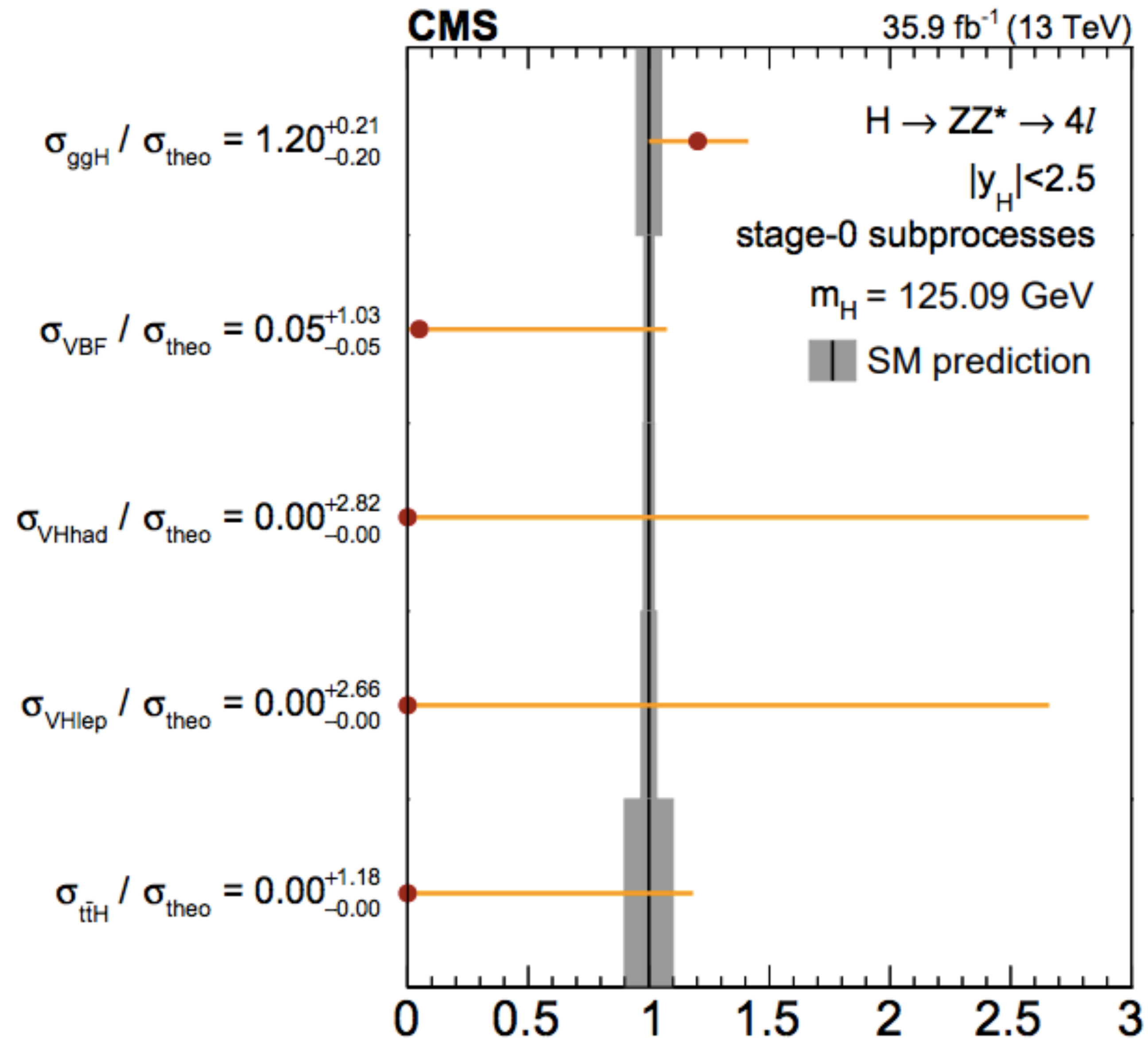
H→ZZ: VBF



- Both VBF categories are very low-stat
- Downward fluctuation in the VBF-2jet category

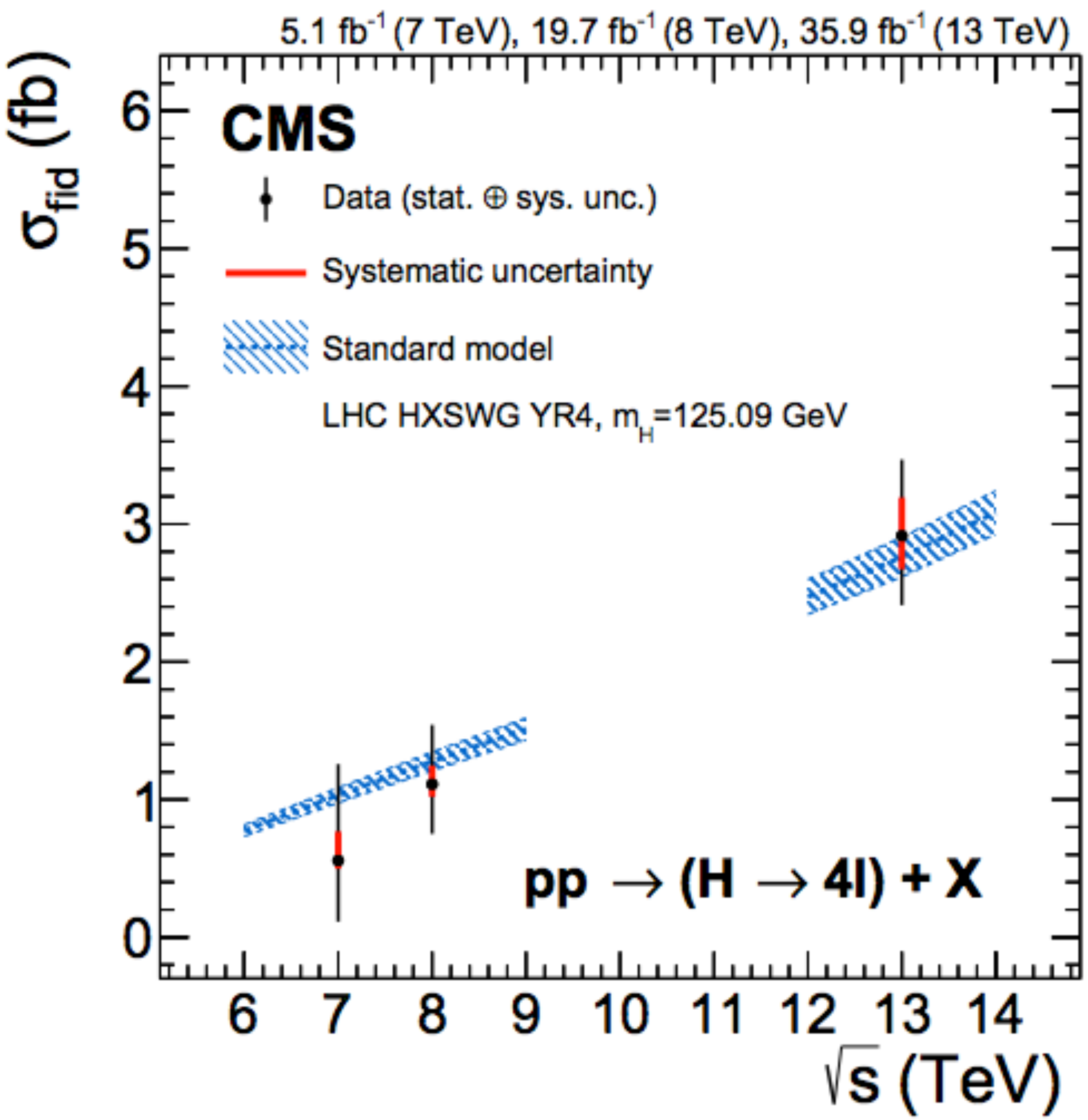
H→ZZ: Results

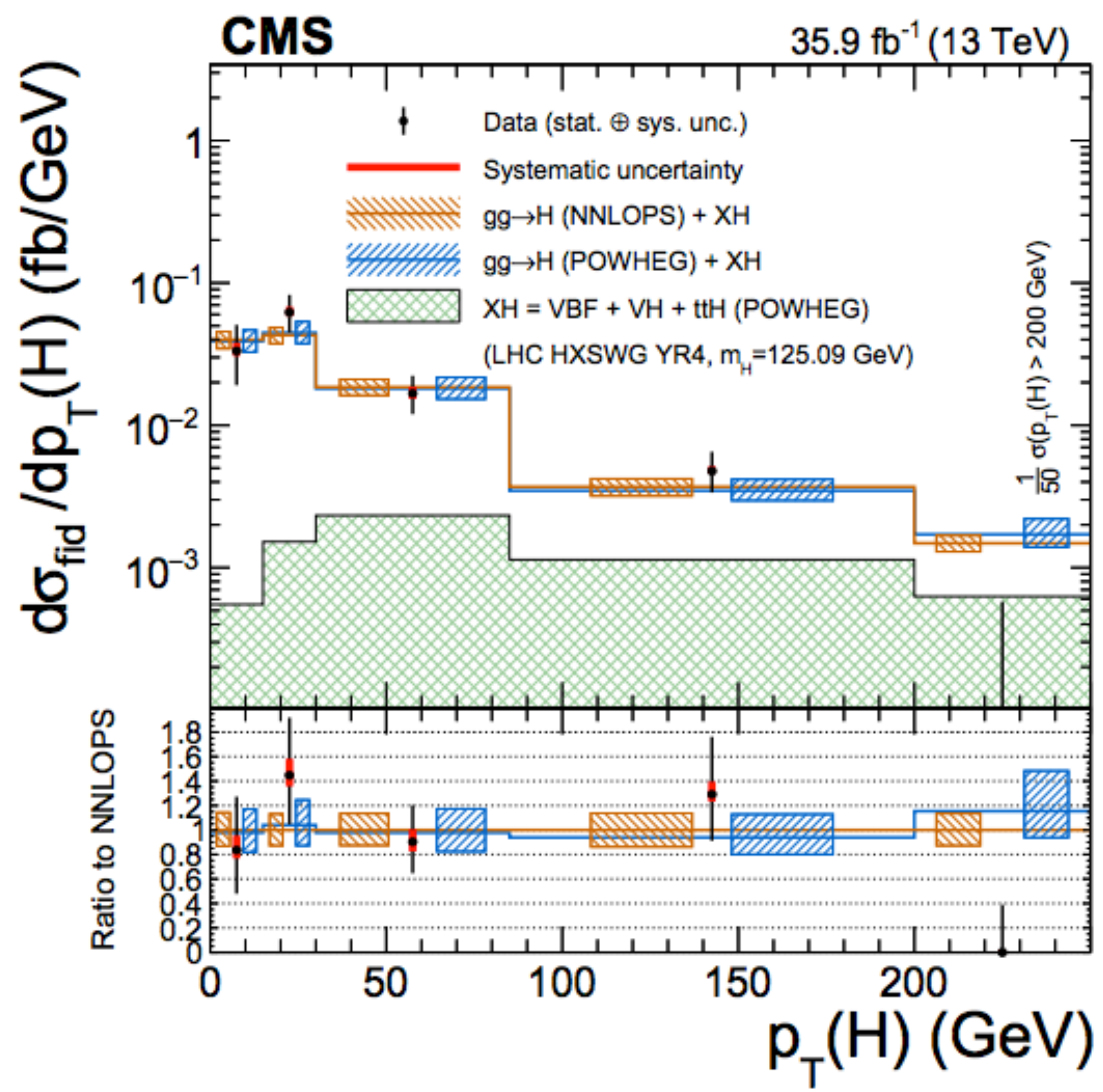




- Usual signal strength modifiers on left, STXS on right
- Low value partly due to upward fluctuation in ggH

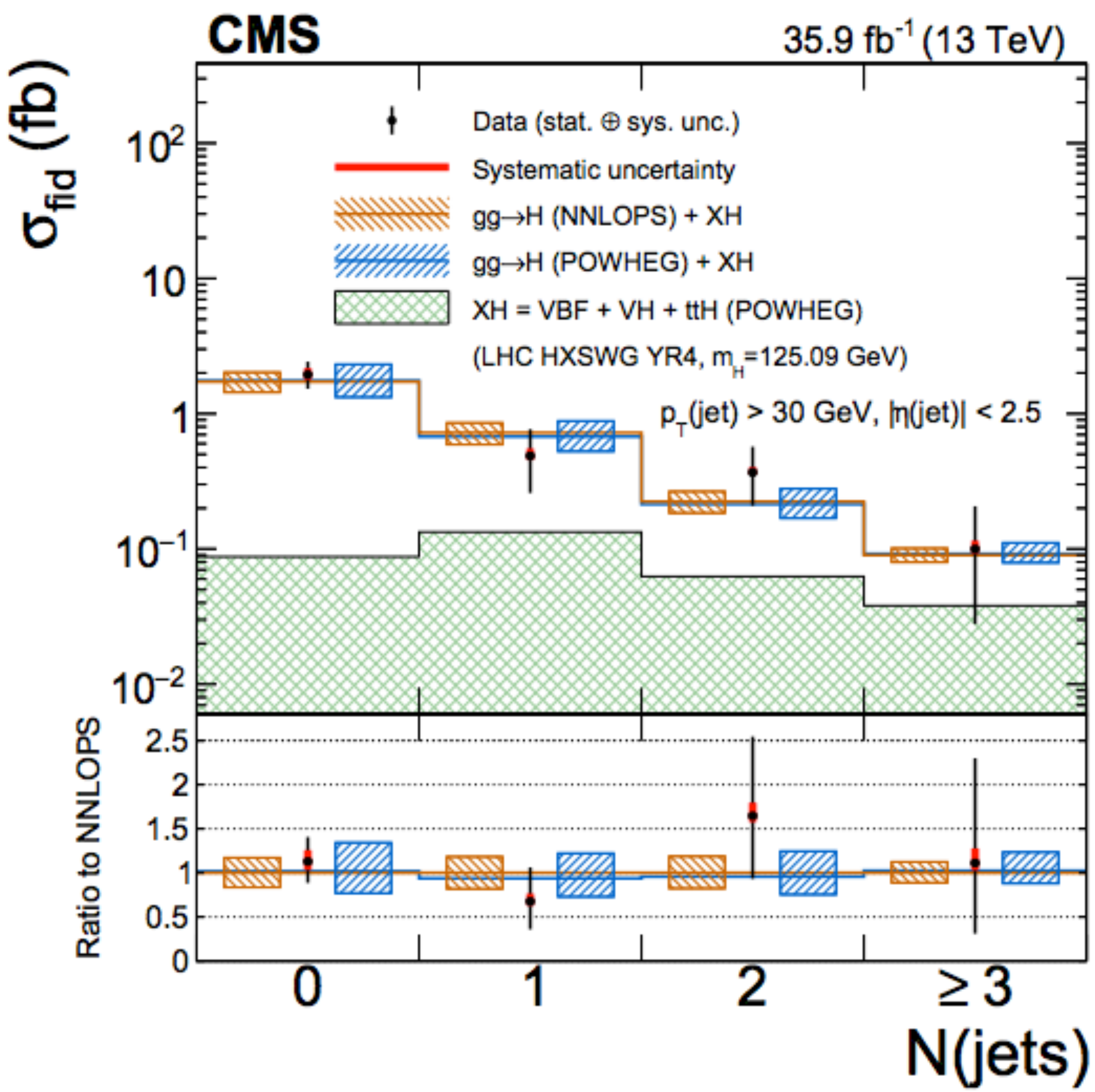
H→ZZ: Differential

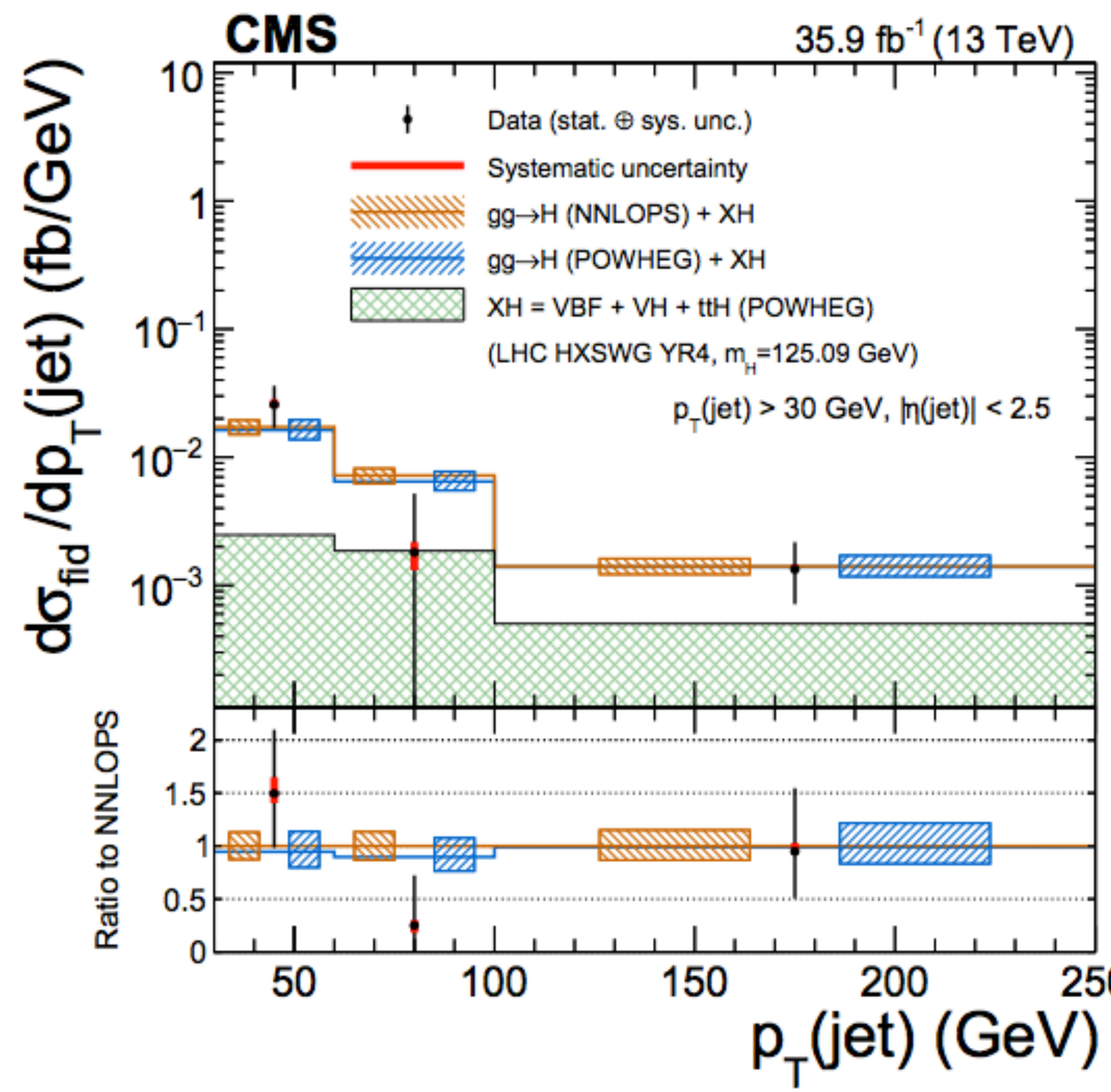




- Fiducial region defined with 105 < m(4l) < 140 GeV; **2.92 +0.56 -0.50 fb**
- **Minimise model dependence:** no cut on $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$, and BR to the three different final states (4e, 2e2μ, 4μ) is left unconstrained

H→ZZ: Differential

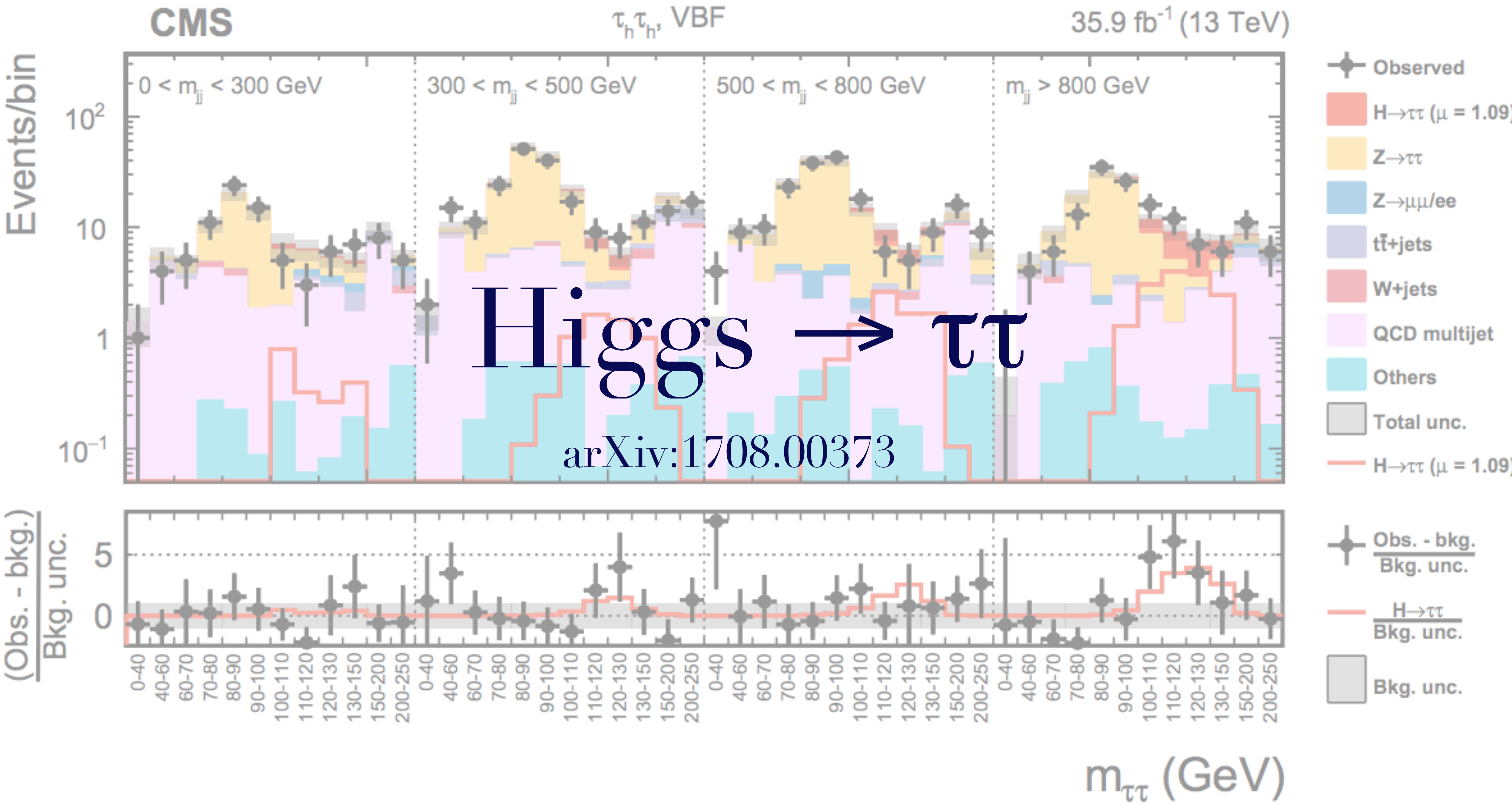




- In addition to $p_T(H)$, results presented for nJets and p_T(lead jet)
- Jet selection is $pT > 30$ GeV and $|\eta| < 2.5$

Higgs → ττ

arXiv:1708.00373



$H\to\tau\tau$: Overview

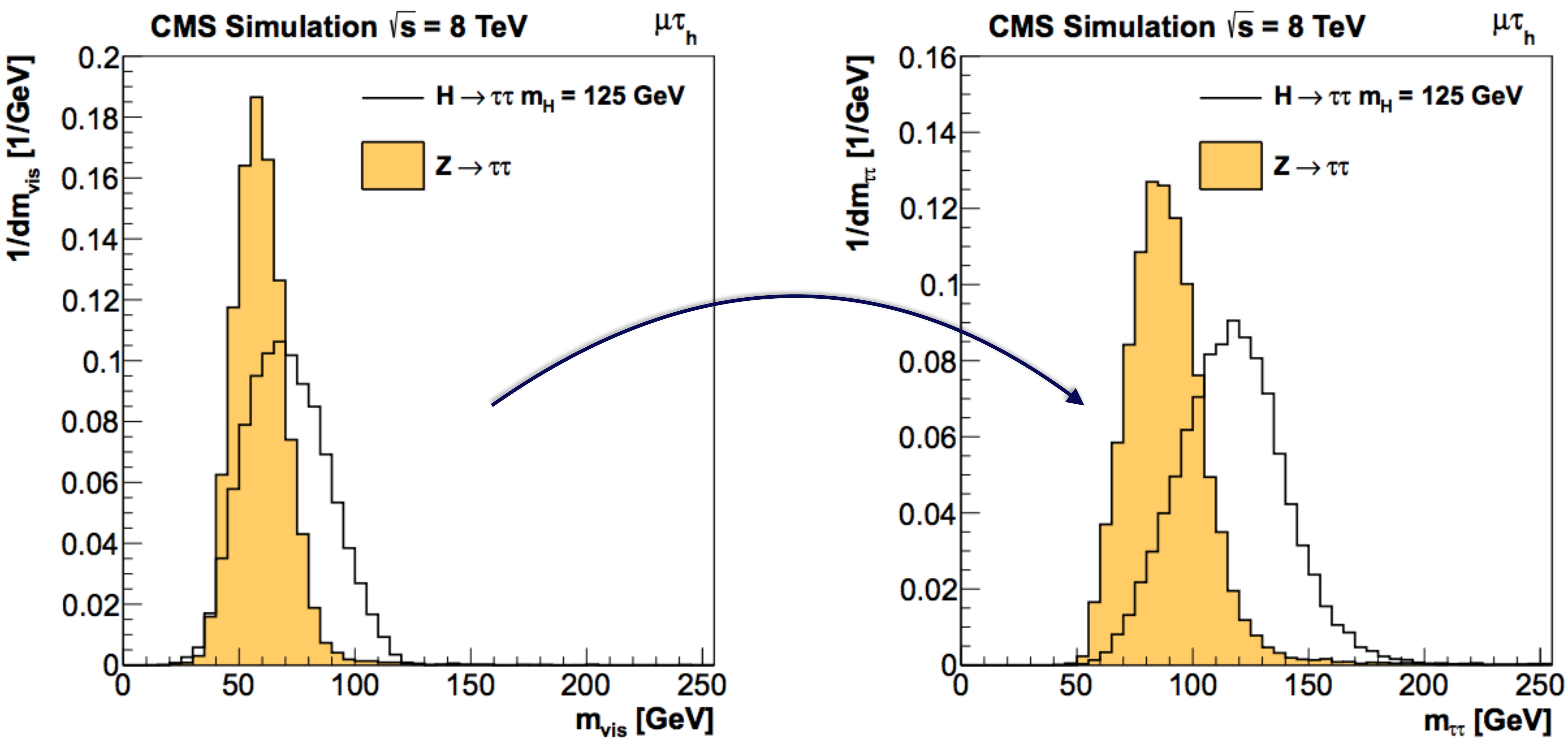
channel	pT
eτ	e>26 τ>30
μτ	μ>20 τ>30
ττ	τ>50 τ>40
eμ	e>13(24) μ>24(15)

The limit is extracted in 2 dimensions.
A mass variable is "unrolled" in a second dimension (variable) to extract the limit targeting a specific production mode.

- Search for **$H\to\tau\tau$** in four different channels: **$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $e\mu$**
- Three categories:
  - **0-jet**, targeting ggH production
  - **Boosted**, for ggH with H recoiling against a jet
  - **VBF** via cut-based dijet analysis
- VH also included as signal, with $H\to WW$ as background
- Different strategy to Run 1: 2D fits
- One of the two is always $m_{\tau\tau}$
- Other is used to target production modes: can be tau p_T, Higgs p_T or dijet mass
- Additional control regions included in the fit

H→ττ: Signal

- The di-tau mass ($m_{\tau\tau}$) is reconstructed using SVFit in most categories
  - Allows for separation of Z→ττ and H→ττ, shifts the H→ττ peak to 125 GeV
- Inputs are MET, its uncertainty, and the four vector of each tau candidate



$H\to\tau\tau$: Background

$Z\to\tau\tau$ and $Z\to\mu\mu$ (irreducible)

- Taken from DY MC, binned in nJets
- Corrections applied using $Z\to\mu\mu$ control region

W+jets

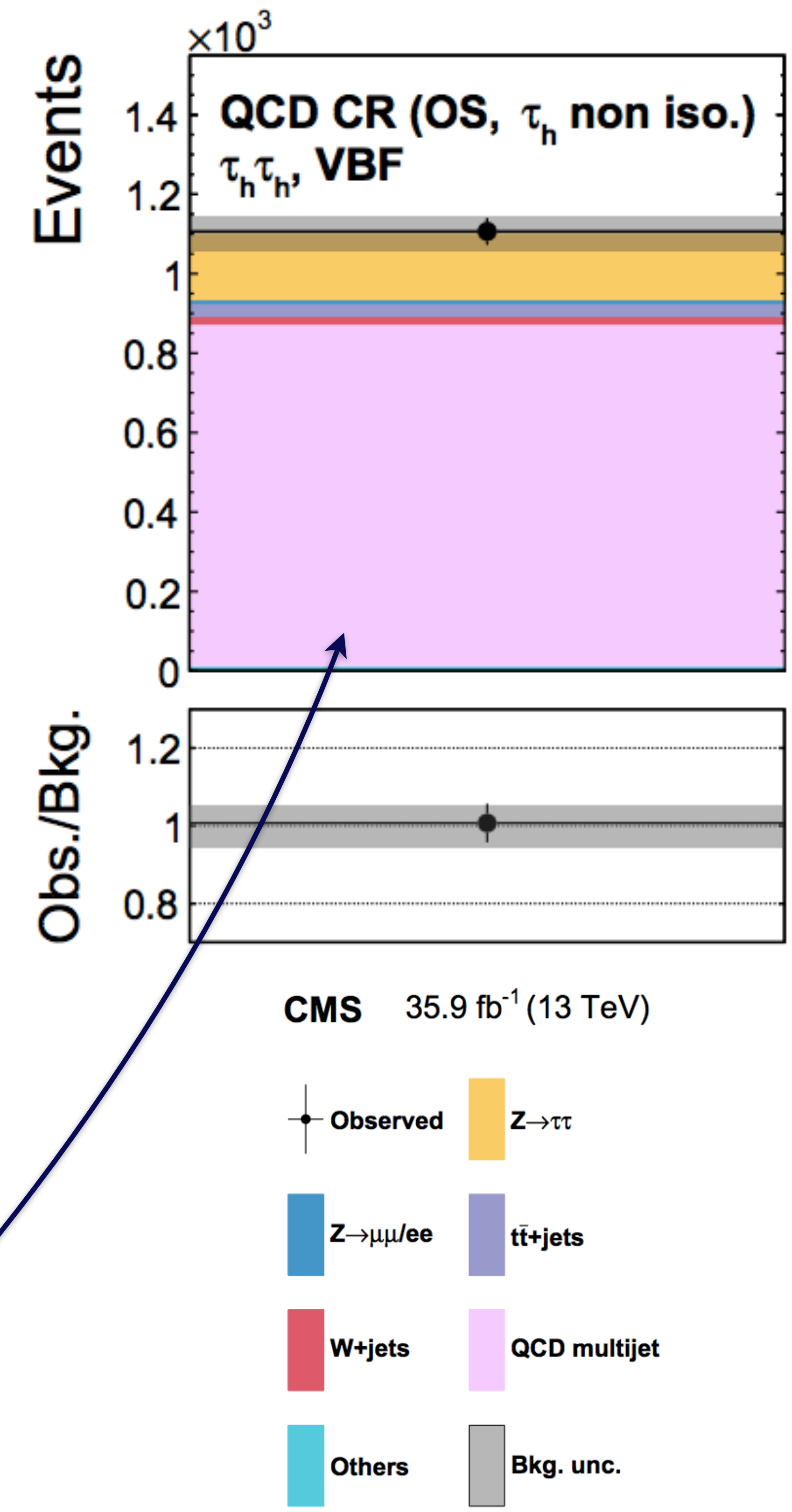
- $e\tau$ and $\mu\tau$ channels use CR defined by high m_T; normalisation included in fit

tt

- constrained using $e\mu$ channel CR, with low $D\zeta$ (geometric variable); norm included in fit

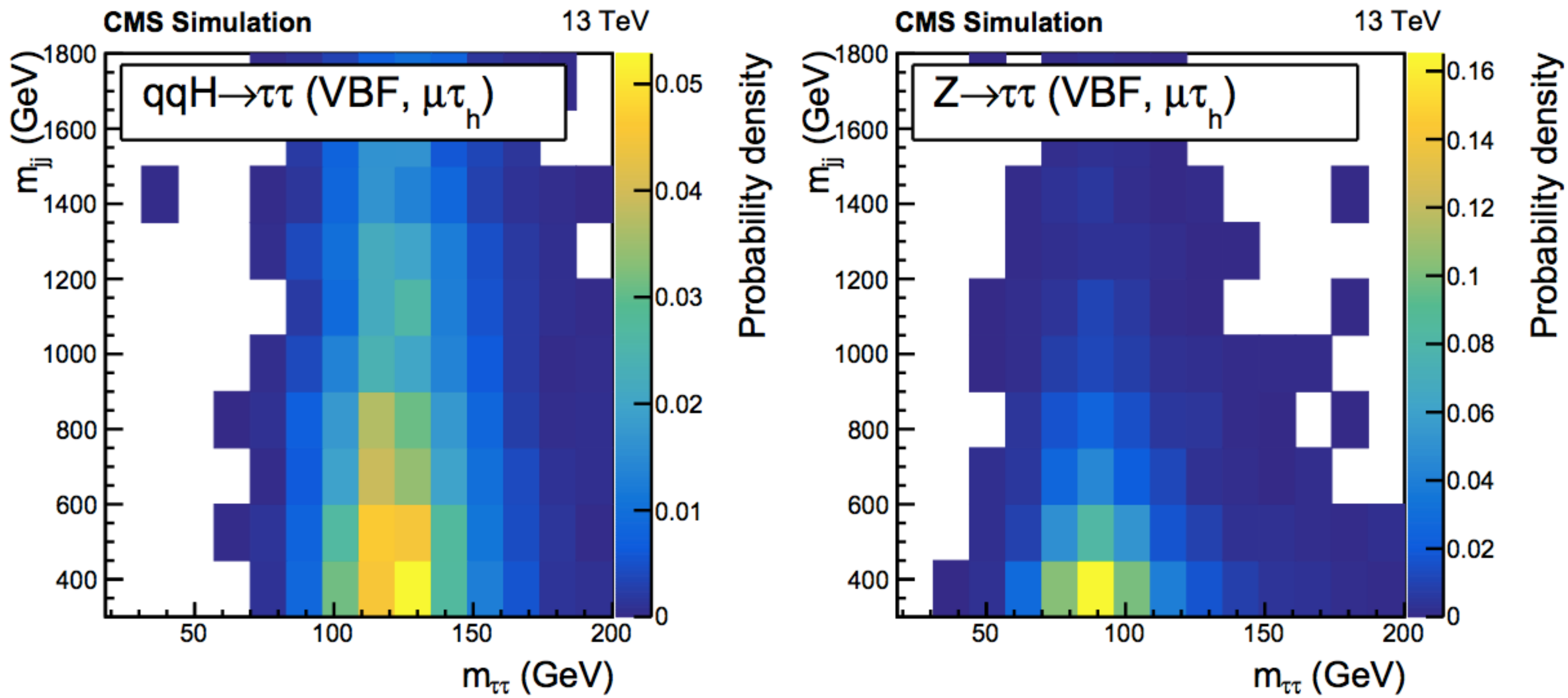
QCD

- same sign CR used for $e\tau$ and $\mu\tau$
- opposite sign relaxed isolation CR for $\tau_h\tau_h$

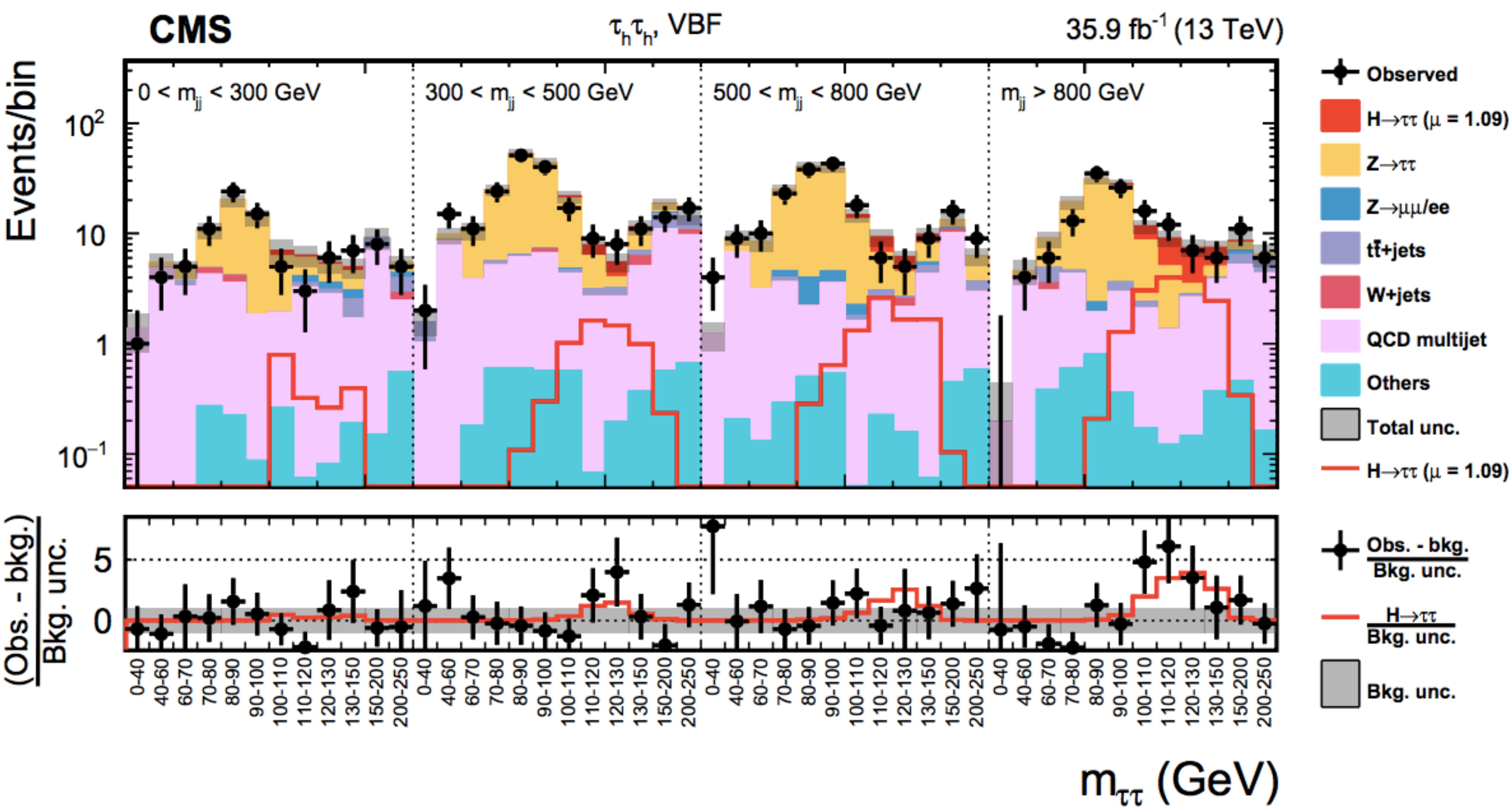


H→ττ: Categories

	0-jet	VBF	Boosted
		Selection	
$\tau_h\tau_h$	No jet	$\geq$2 jets, $p_T^{\tau\tau} > 100$ GeV, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300$ GeV, $p_T^{\tau\tau} > 50$ GeV, $p_T^{\tau_h} > 40$ GeV	Others
$e\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300$ GeV, $p_T^{\tau\tau} > 50$ GeV	Others
$e\mu$	No jet	2 jets, $m_{jj} > 300$ GeV	Others
		Observables	
$\tau_h\tau_h$	$m_{\tau\tau}$	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$\mu\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\mu$	p_T^{μ}, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$

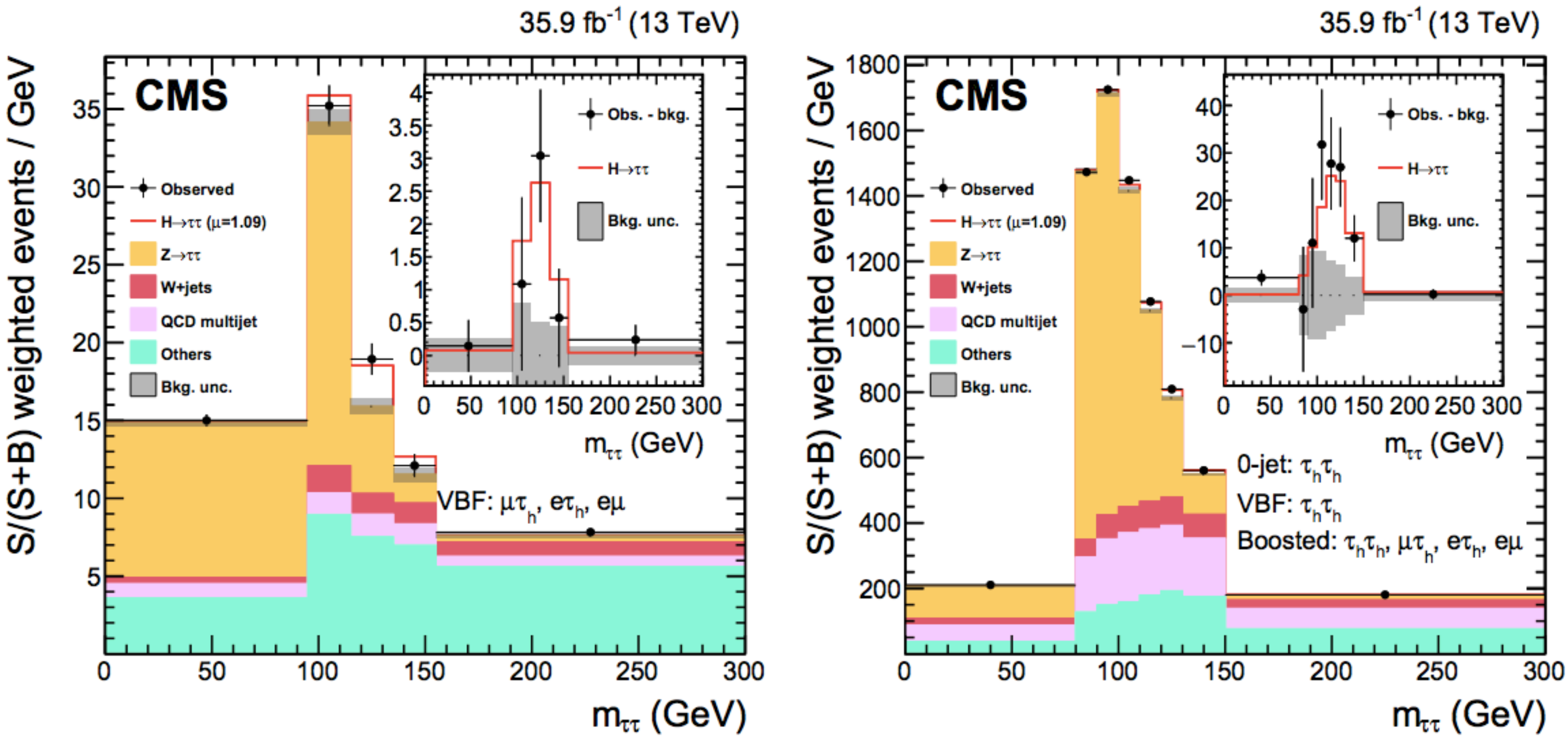


H→ττ: VBF



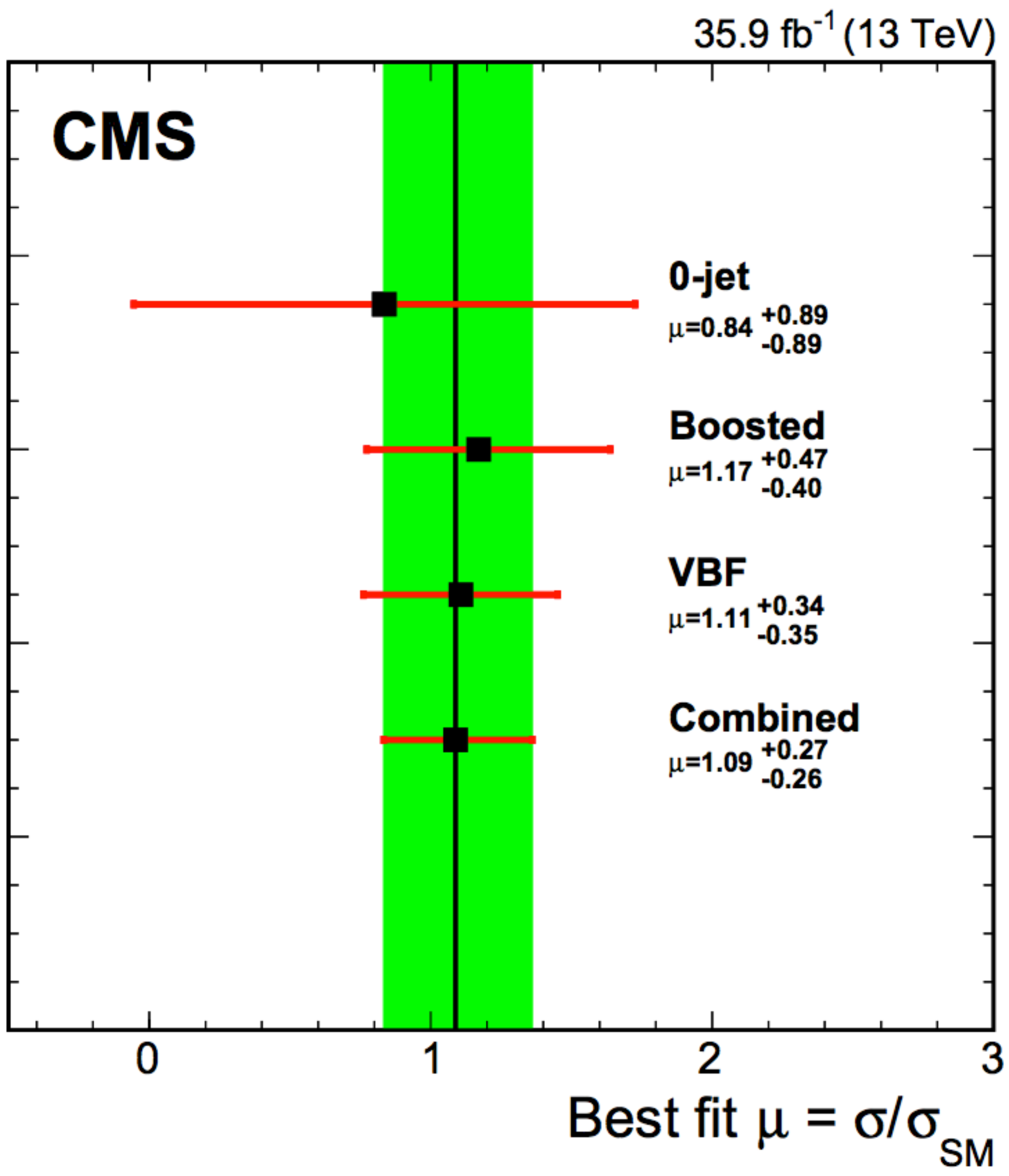
- Majority of the signal in larger m_{jj} bins

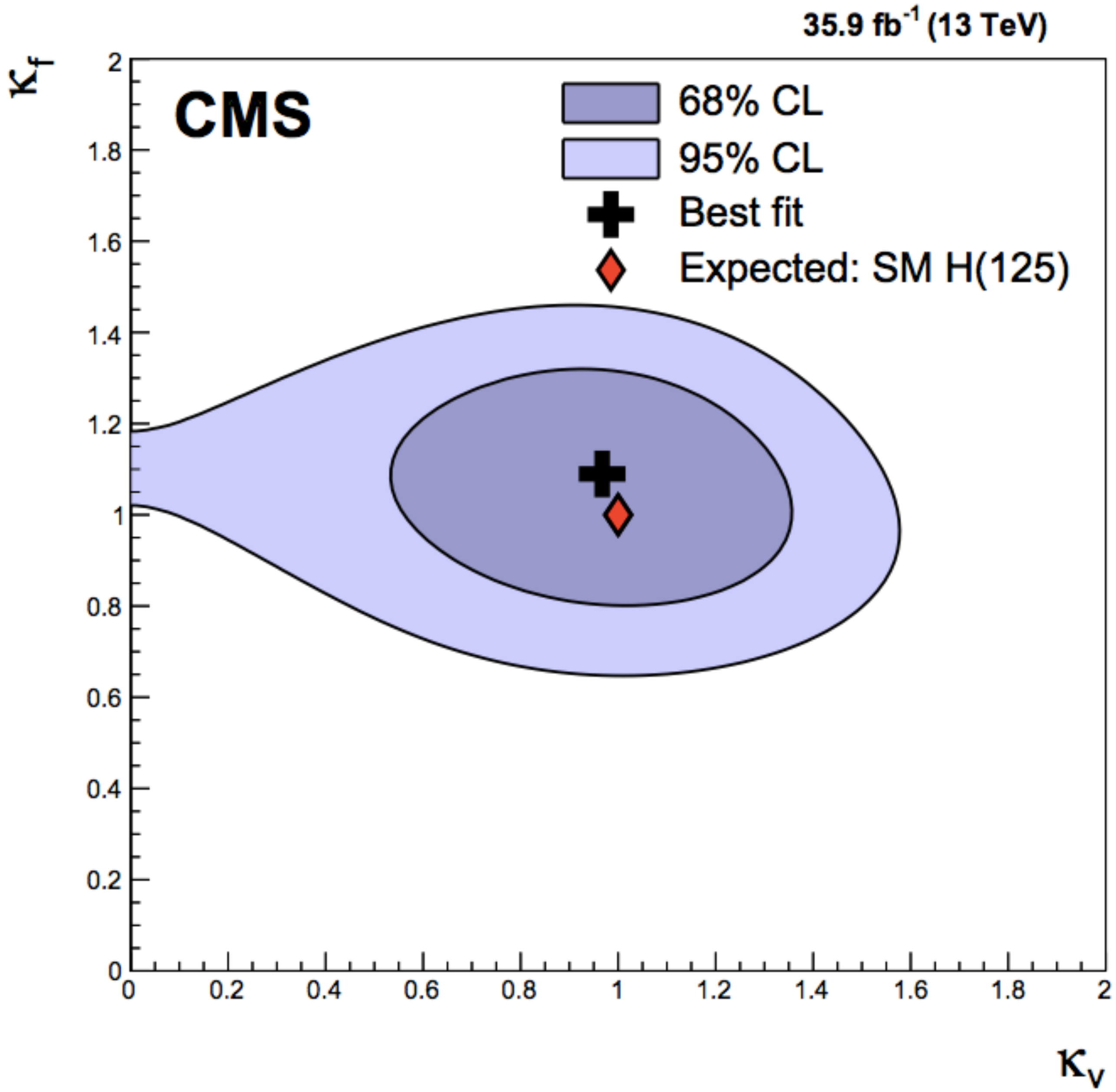
H→ττ: Results



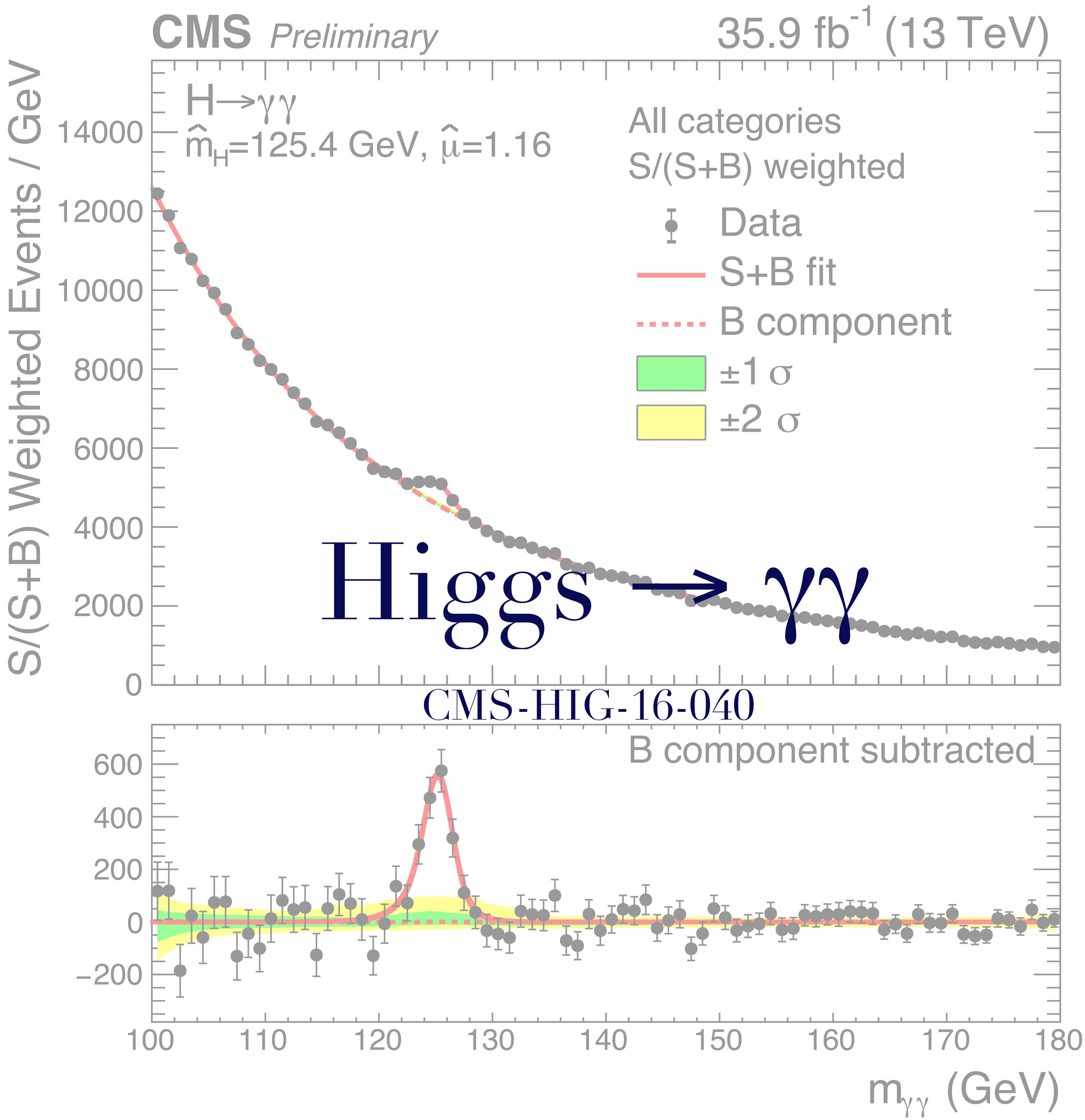
- Signal peak is visible by eye after background subtraction
- Events weighted by S/S+B

H→ττ: Results





- H→WW included as signal for 2D kappa scan



Higgs → γγ

CMS-HIG-16-040

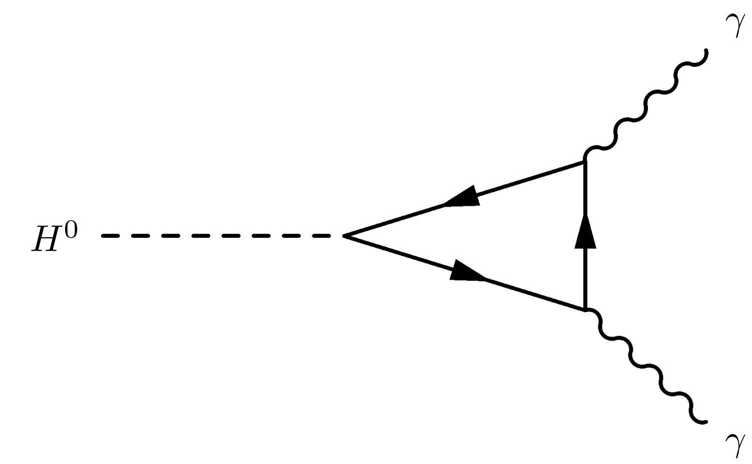
H→γγ: Overview

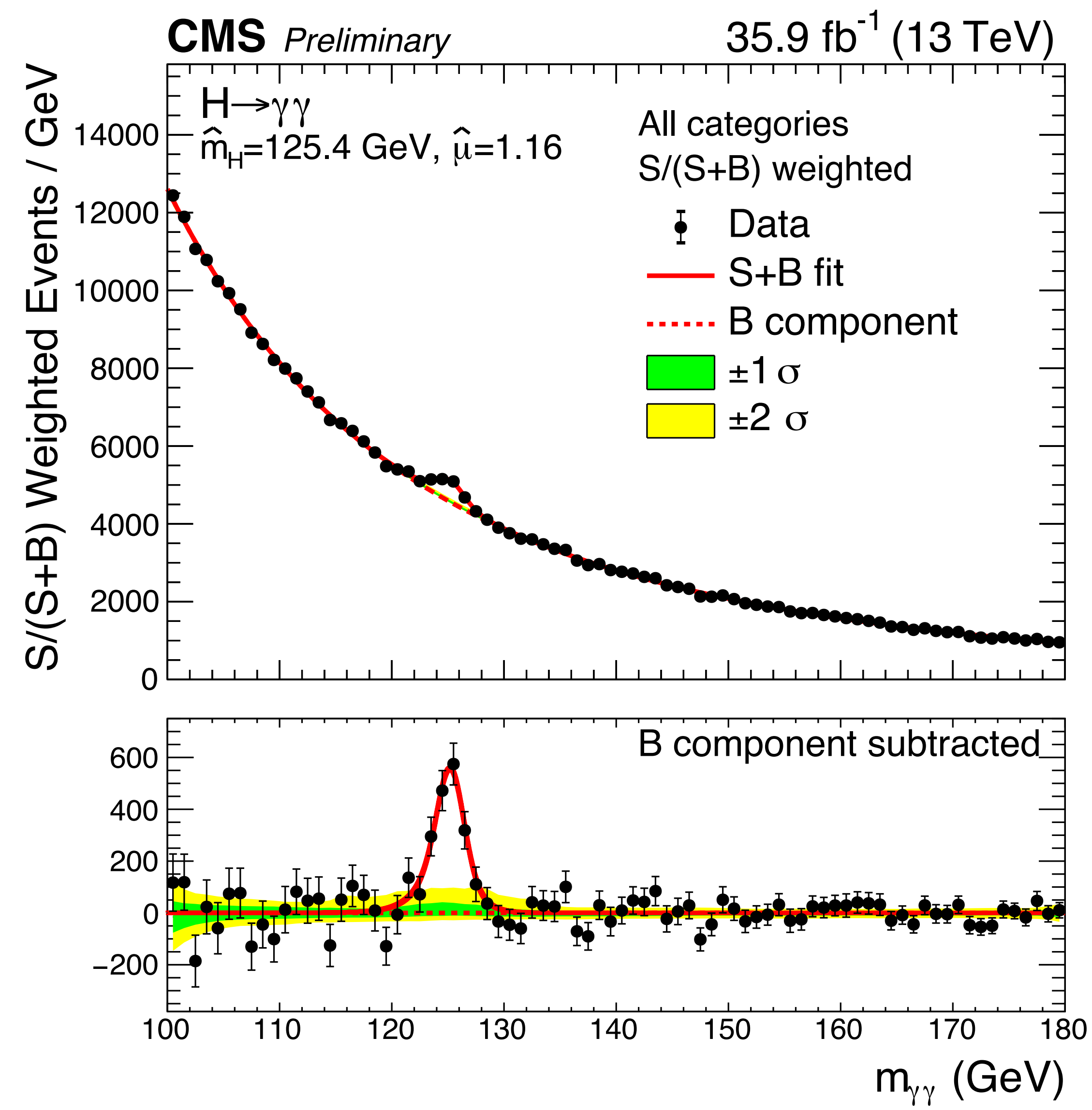
- High resolution channel
- Key to discovery despite low BR

- PAS with 2016 dataset released
- Proceeding to release of a paper

- Select events with two well-isolated photons with $p_T > 30$ (20) GeV for (sub)leading, $|\eta| < 2.5$
- Additional scaled cut on (sub)lead photon $p_T > m_{\gamma\gamma}/3$ ($m_{\gamma\gamma}/4$)

- Fit small signal peak on large falling background in categories



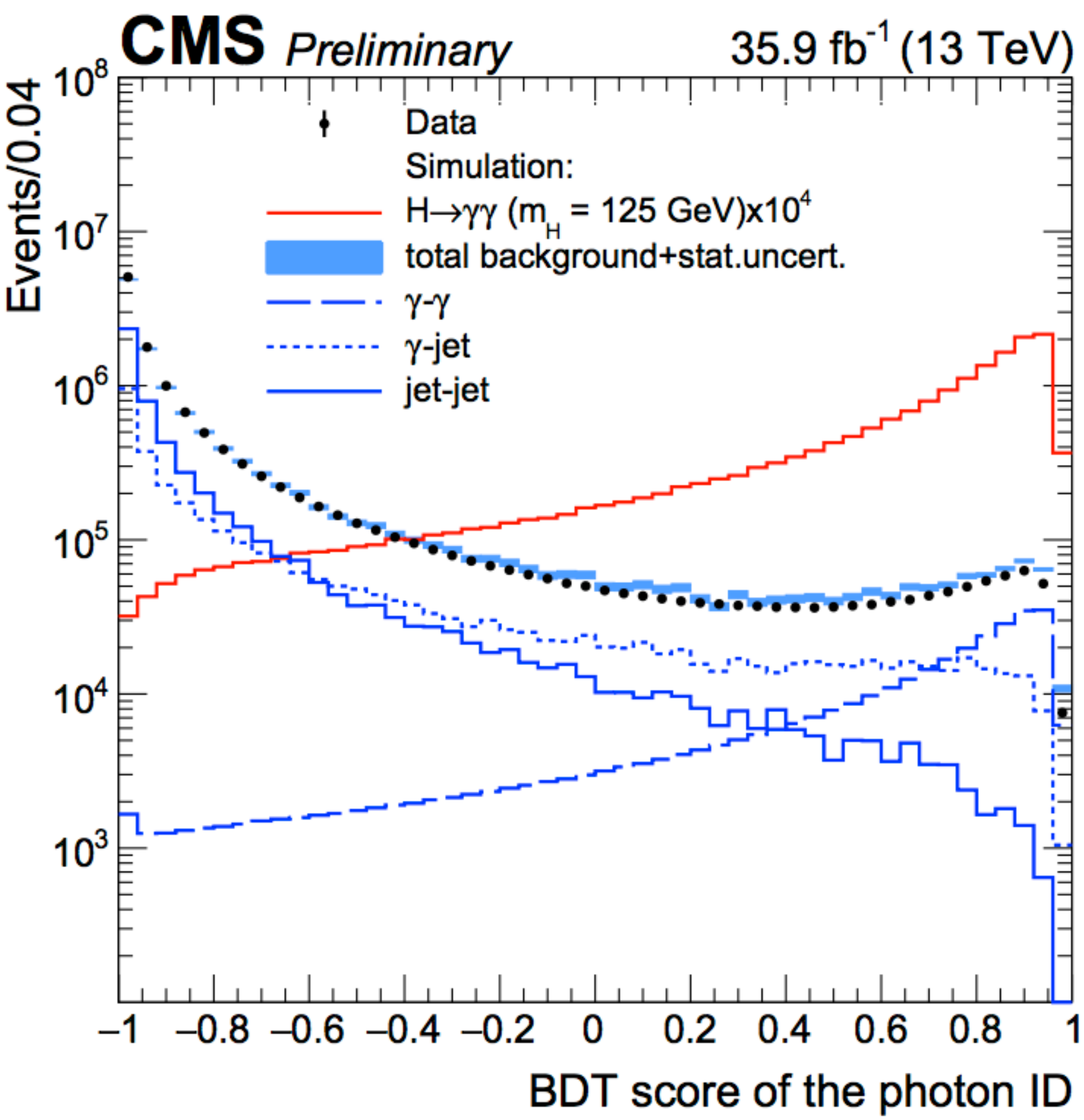


H→γγ: Strategy

- **Identify photons using a BDT**
- Discriminates between real and fakes using shower shape & isolation

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1-\cos\theta)}$$

- **Another BDT for vertex selection**
- Inputs are track recoil variables
- Negligible effect on resolution if within 1cm of true position

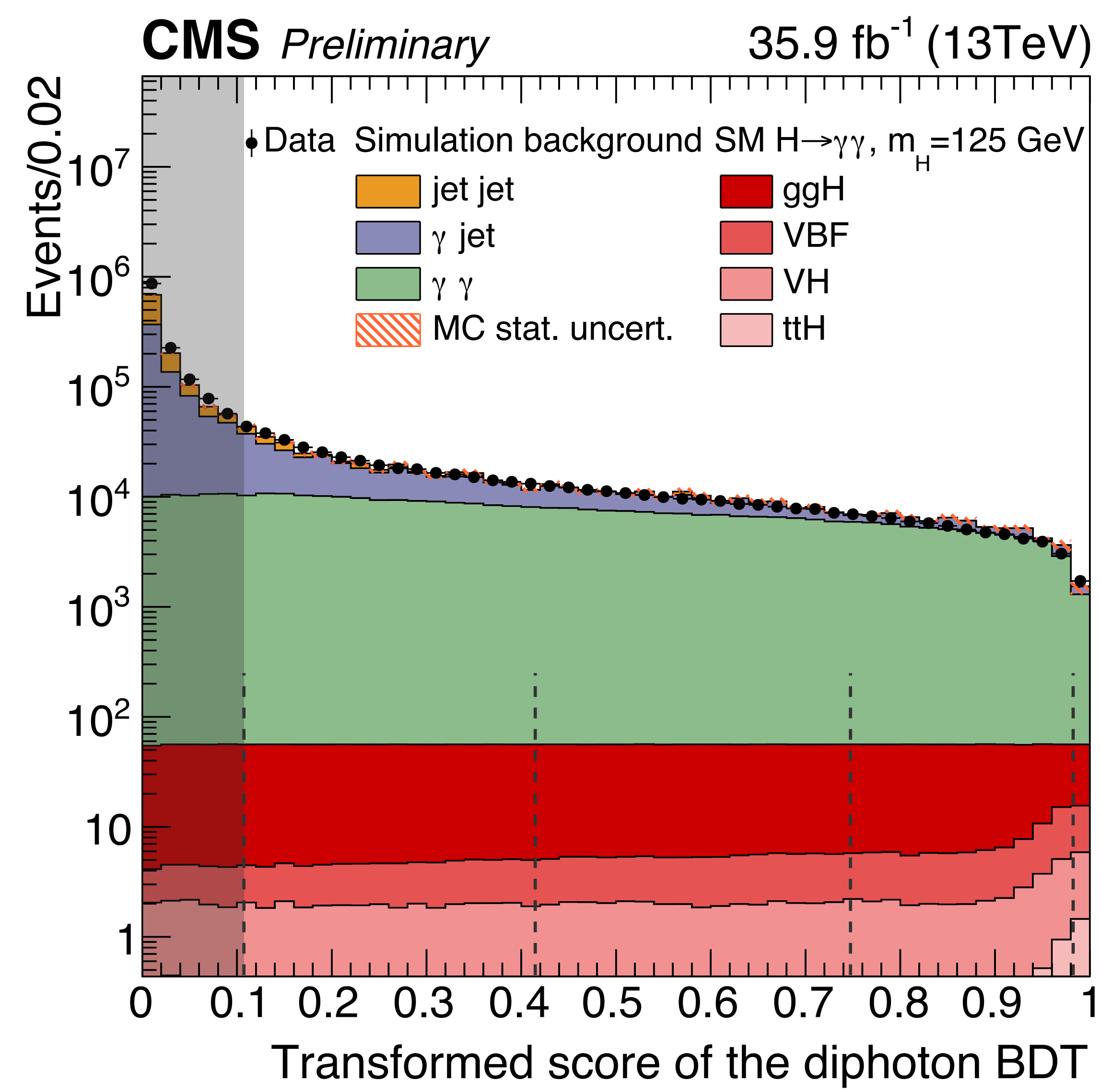


- **Quality of diphotons quantified with third BDT**
- Used to classify events by S/B

Imperial College London

H→γγ: Strategy

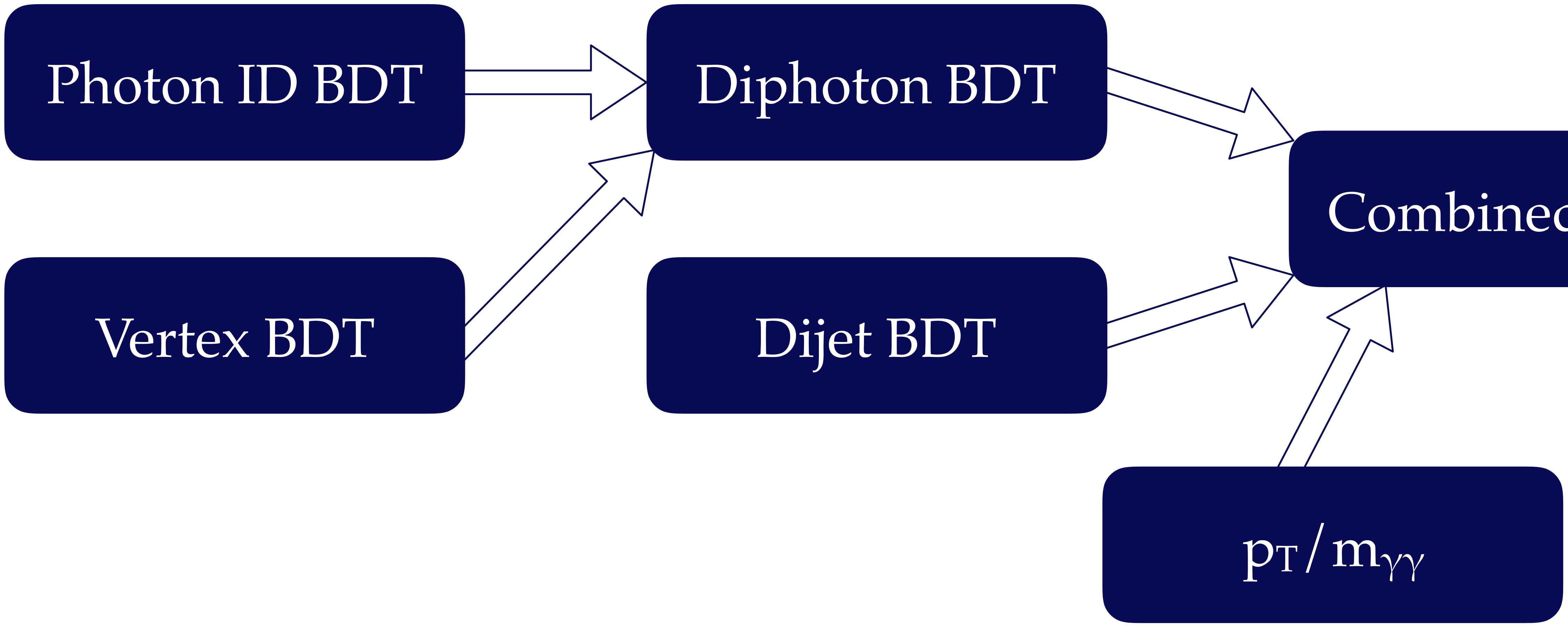
- **Event Classification:** tag events using additional objects present
  → target different production modes
- Enable measurement of per-process signal strengths
- Untagged events (mostly ggH) further separated by S/B
  → improves overall sensitivity



ttH leptonic → VH leptonic → ttH hadronic →

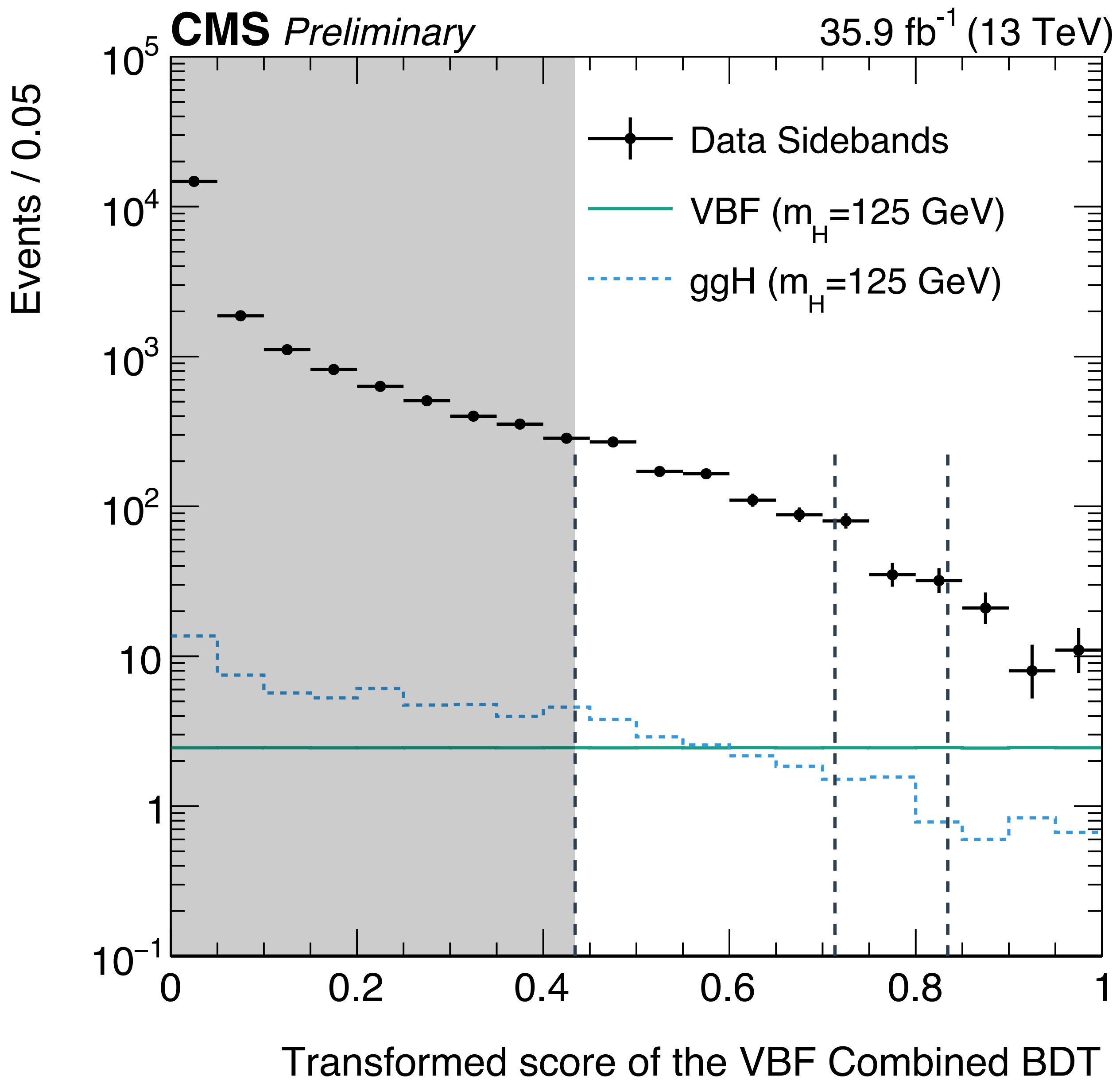
→ VBF 0-2 → VH hadronic → Untagged 0-3

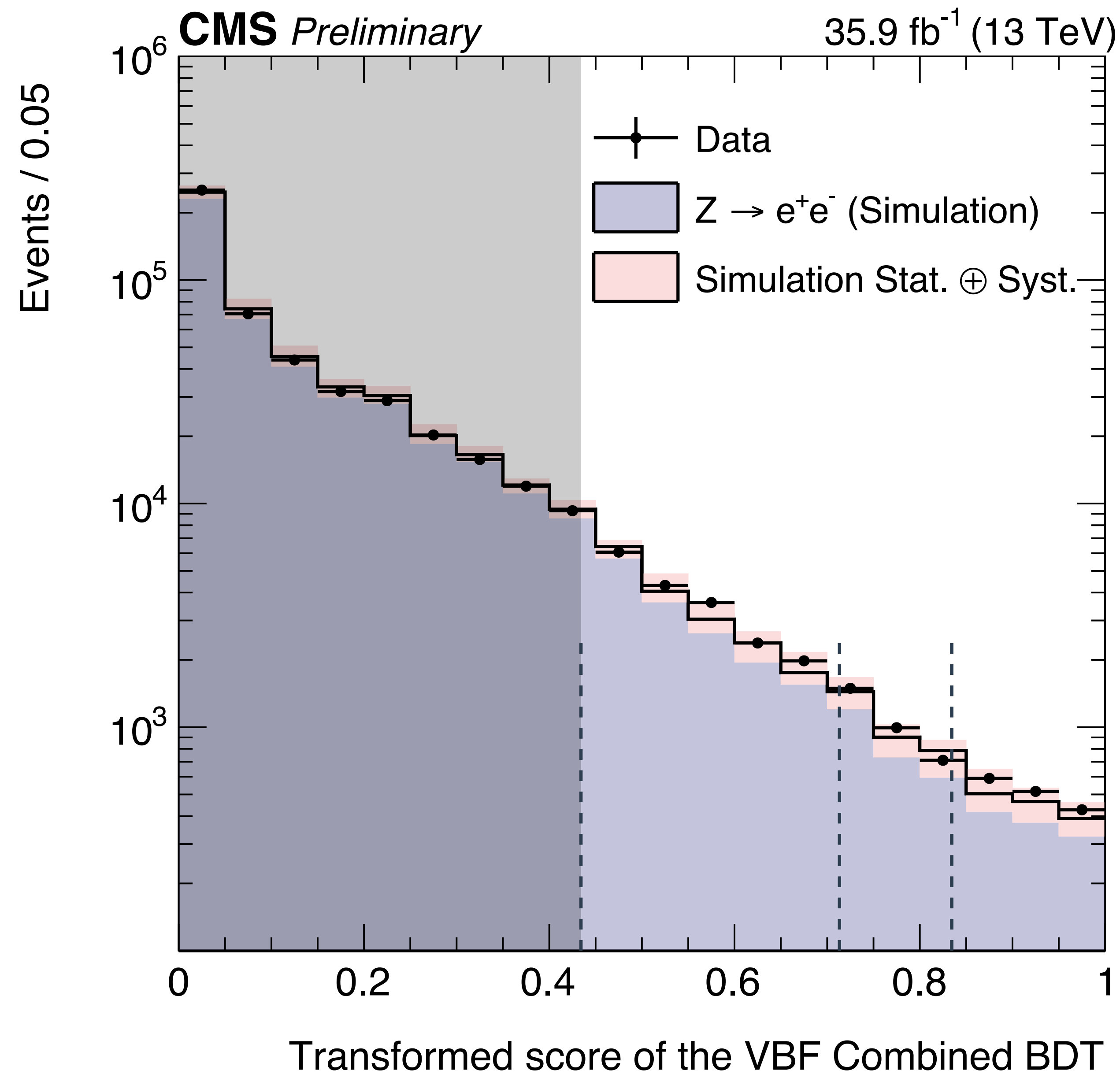
$H\rightarrow\gamma\gamma$: VBF tags



- Diphoton and dijet BDT combined to classify VBF events
- Inputs to dijet BDT include jet p_T and $\Delta\eta$, m_{jj}, additional angular variables

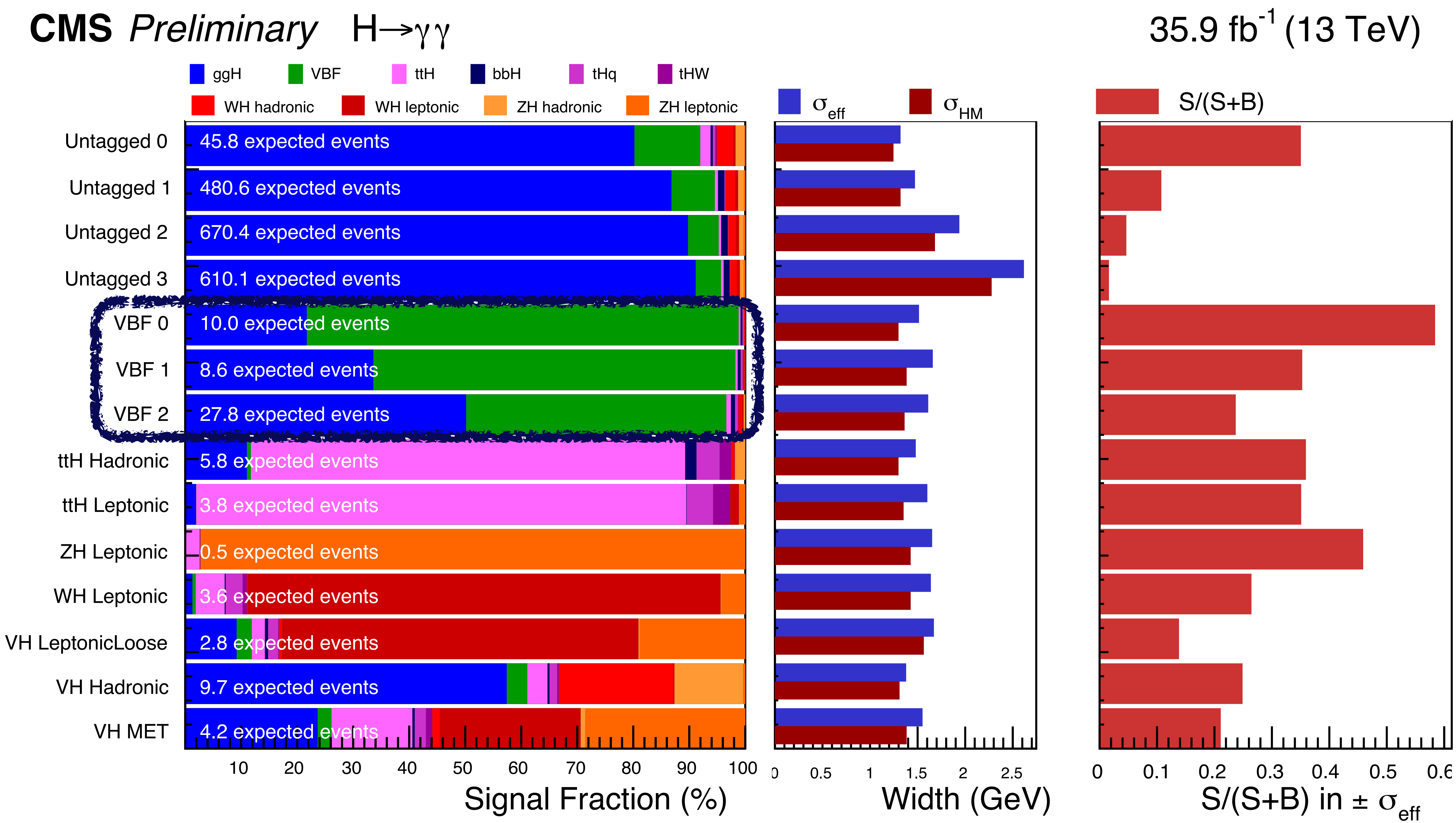
H→γγ: VBF tags





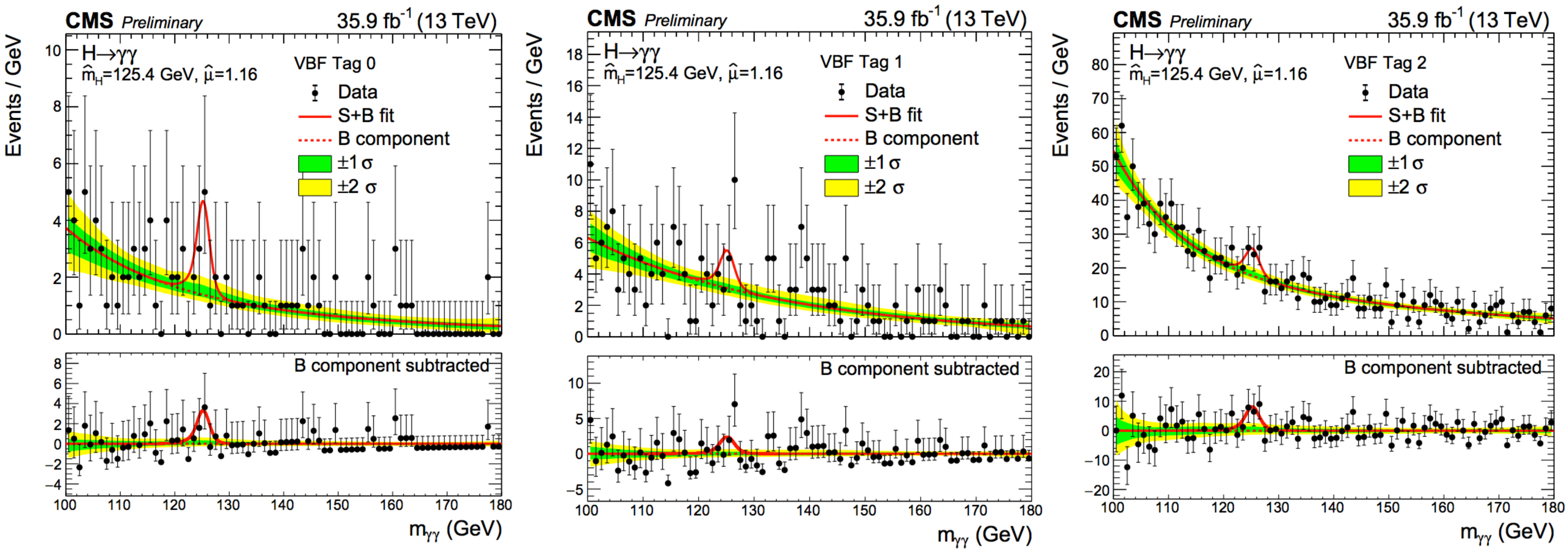
- VBF tags defined using two-step BDT process, where the dijet BDT is combined with the diphoton BDT - cut on resulting distribution
- Validation using both $m_{\gamma\gamma}$ sidebands and Z→ee events with dijets

H→γγ: Signal Model



Imperial College London

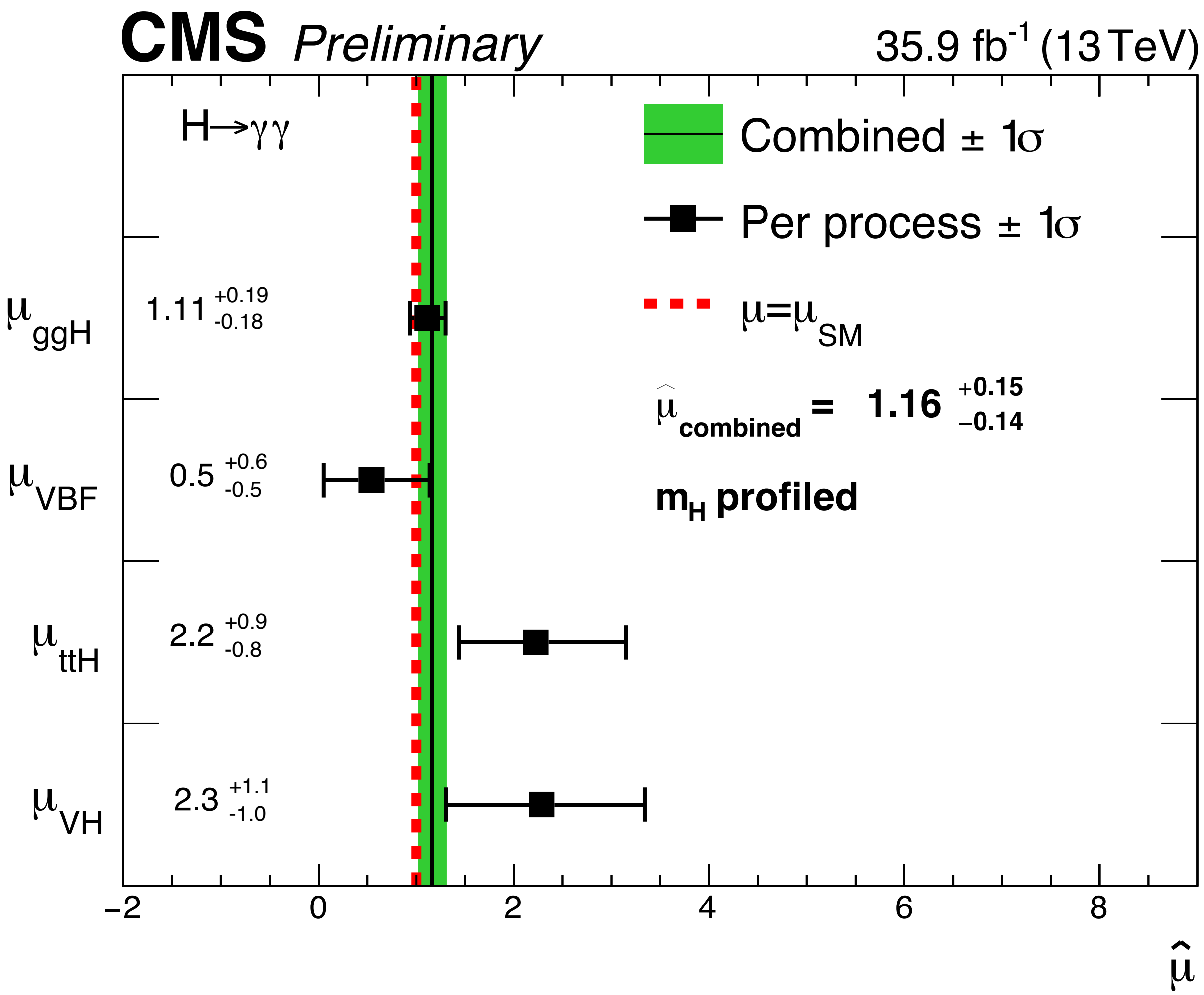
$H \rightarrow \gamma\gamma$: VBF tags

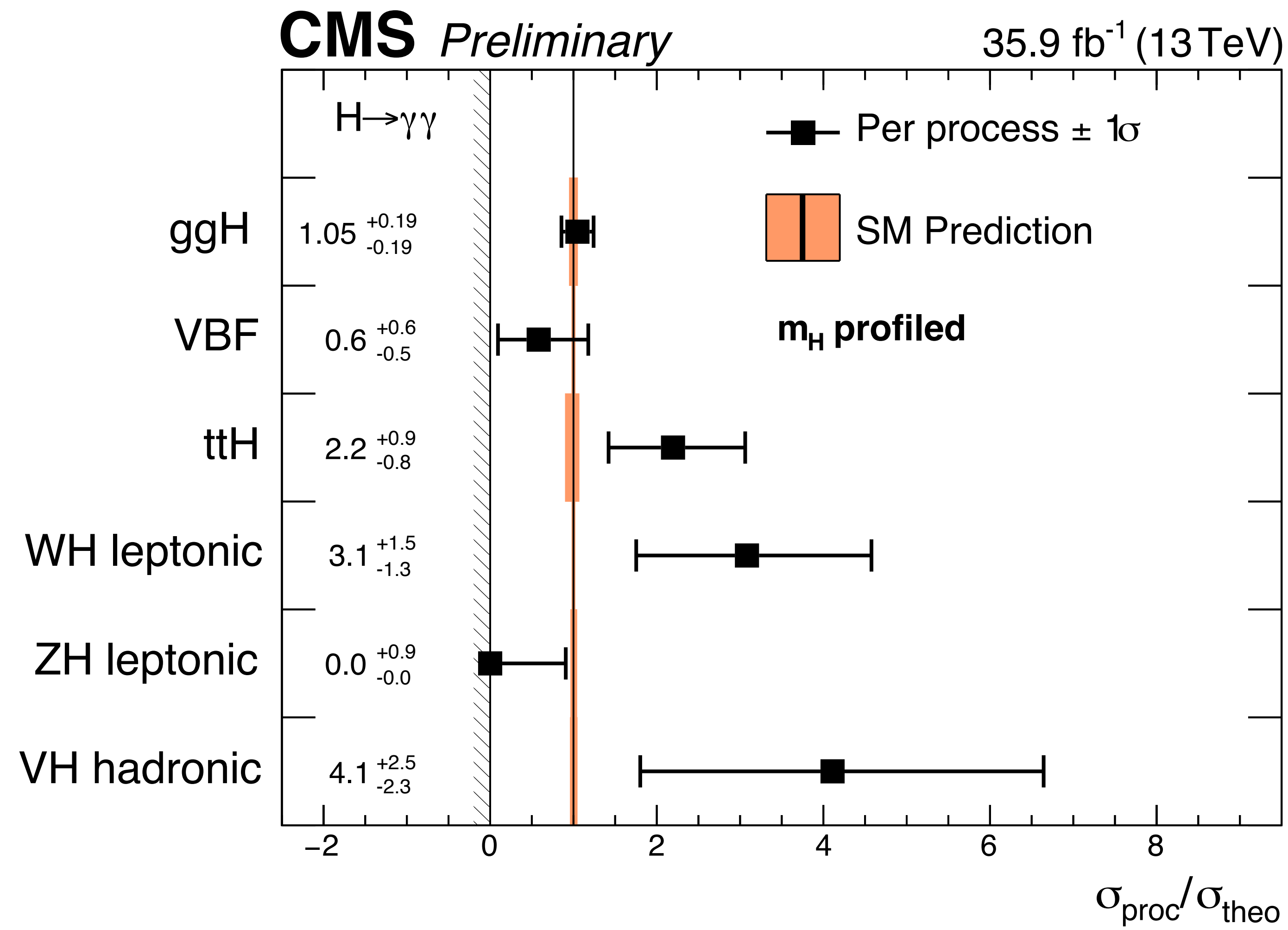


- VBF categories have relatively high S/B
- Data-driven background model uses $m_{\gamma\gamma}$ sidebands
- In many categories Higgs peak now visible by eye

Imperial College London

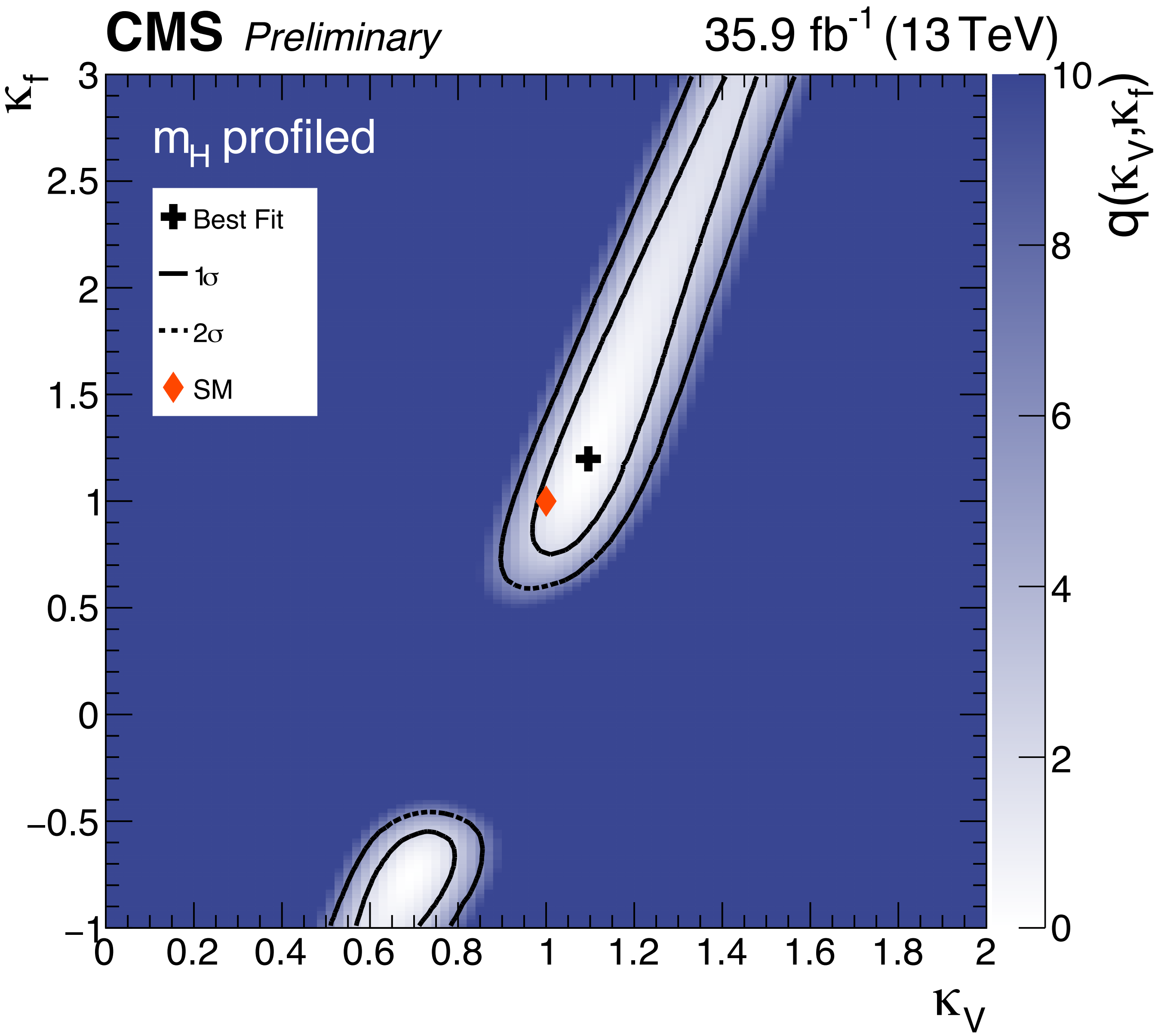
$H\rightarrow\gamma\gamma$: Results

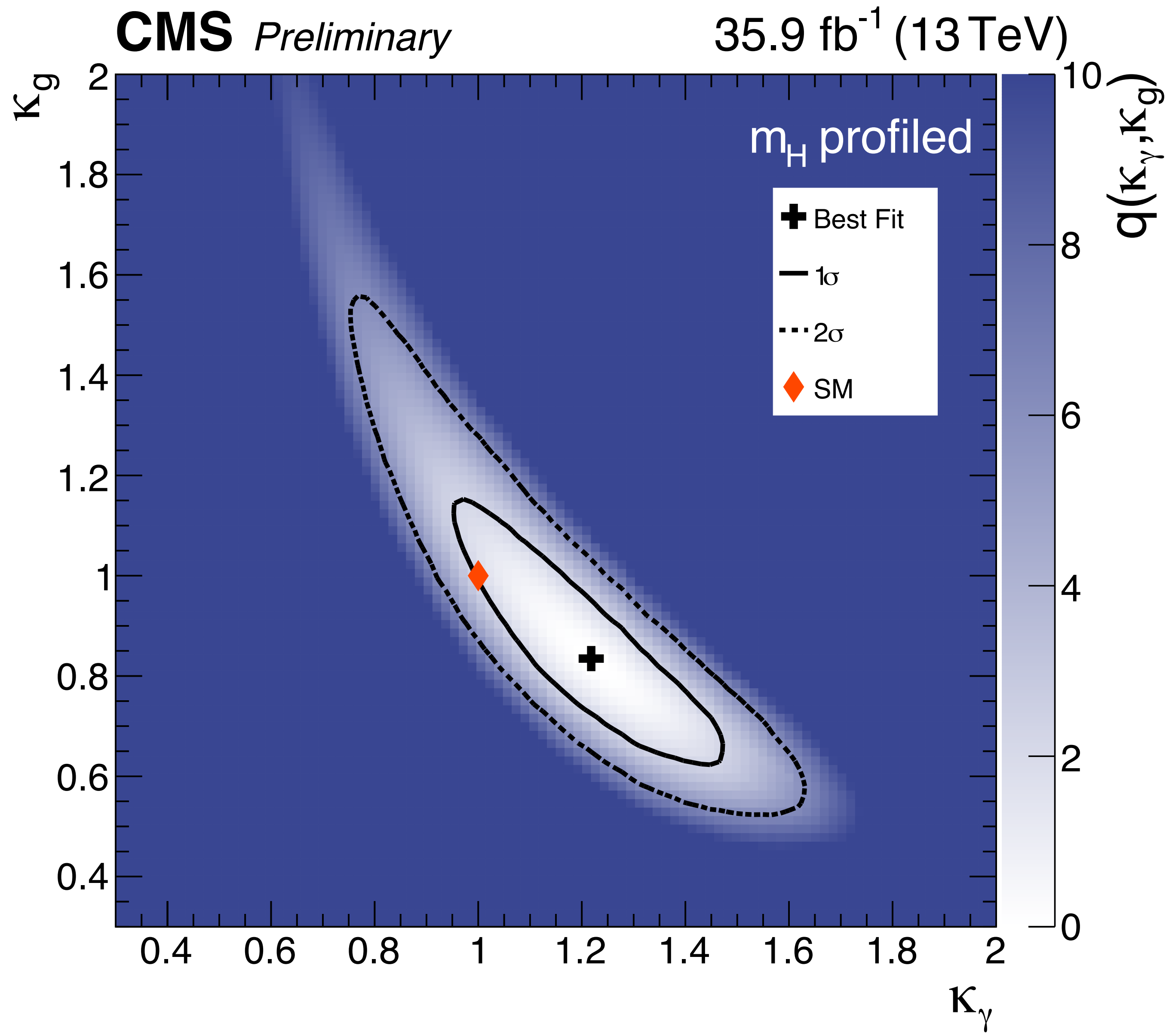




- Simultaneous fit to all categories yields **μ_{VBF} = 0.5 +0.6 -0.5**
- Per-process μ on LHS, including 3.3σ significance for ttH (wrt μ=0)
- Stage 0 Simplified Template Cross-Section (STXS) measurement on RHS

H→γγ: Results

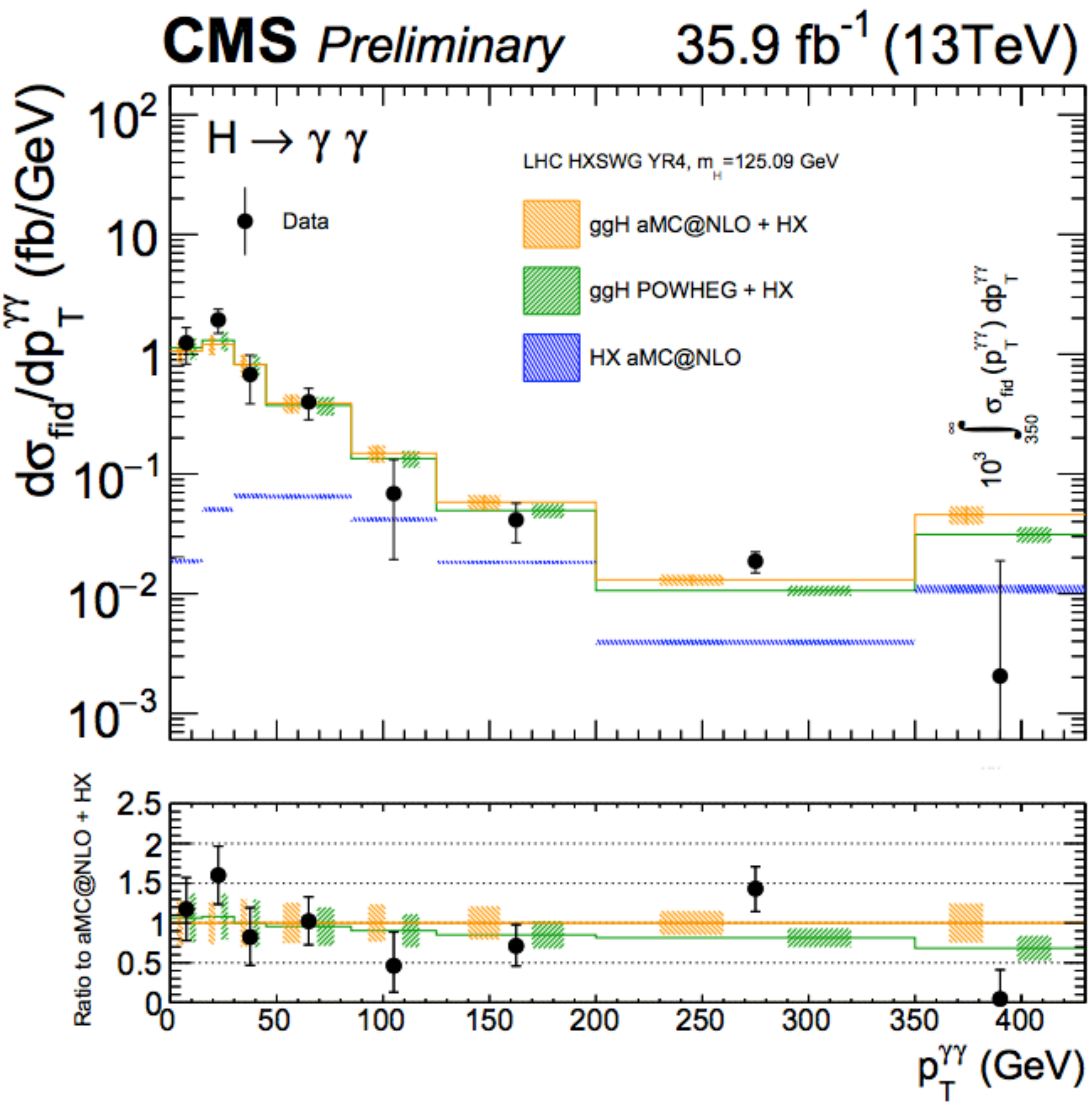


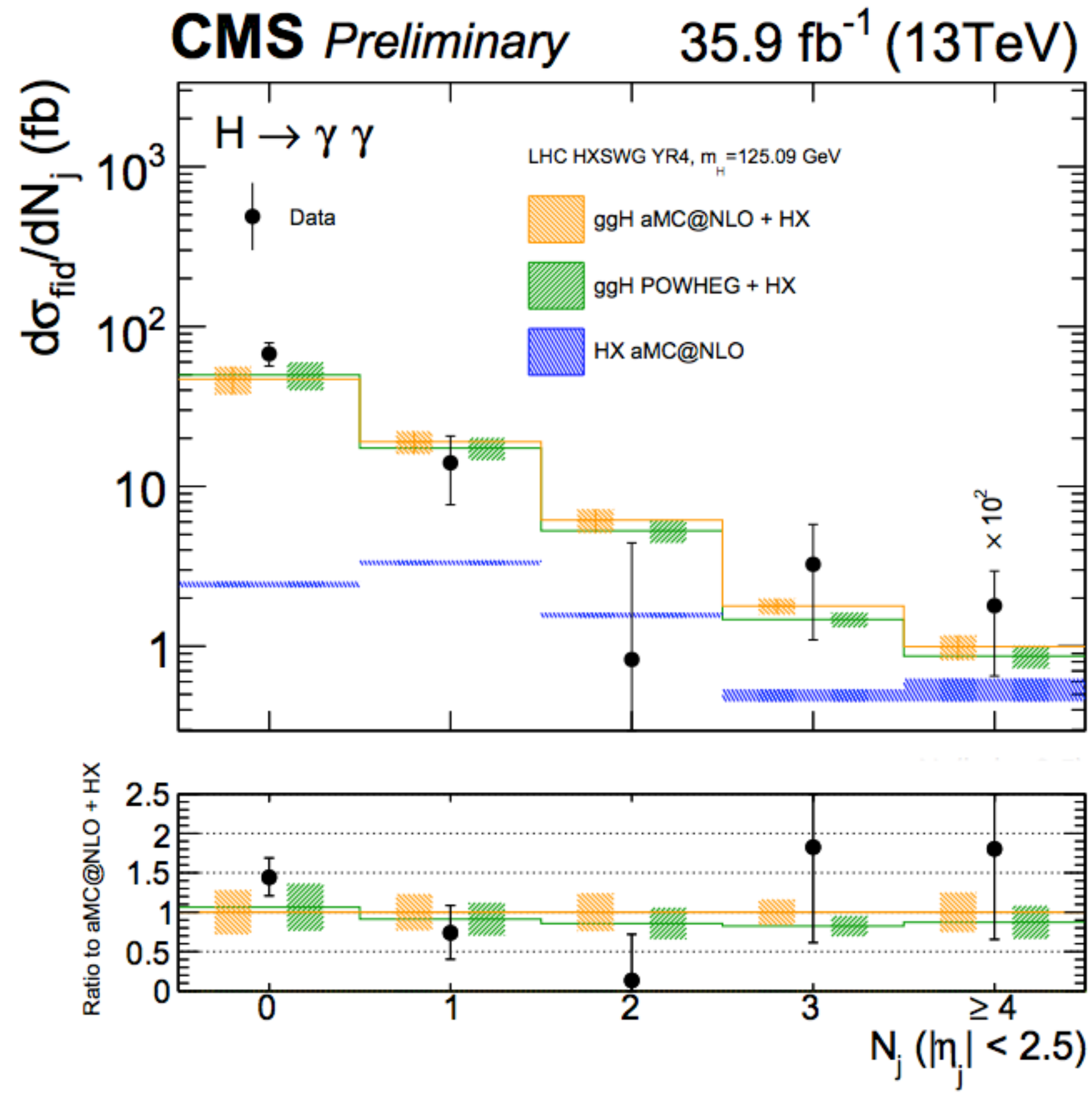


- Coupling to fermions vs vector bosons on LHS
- Effective coupling to gluons vs photons on RHS

Imperial College London

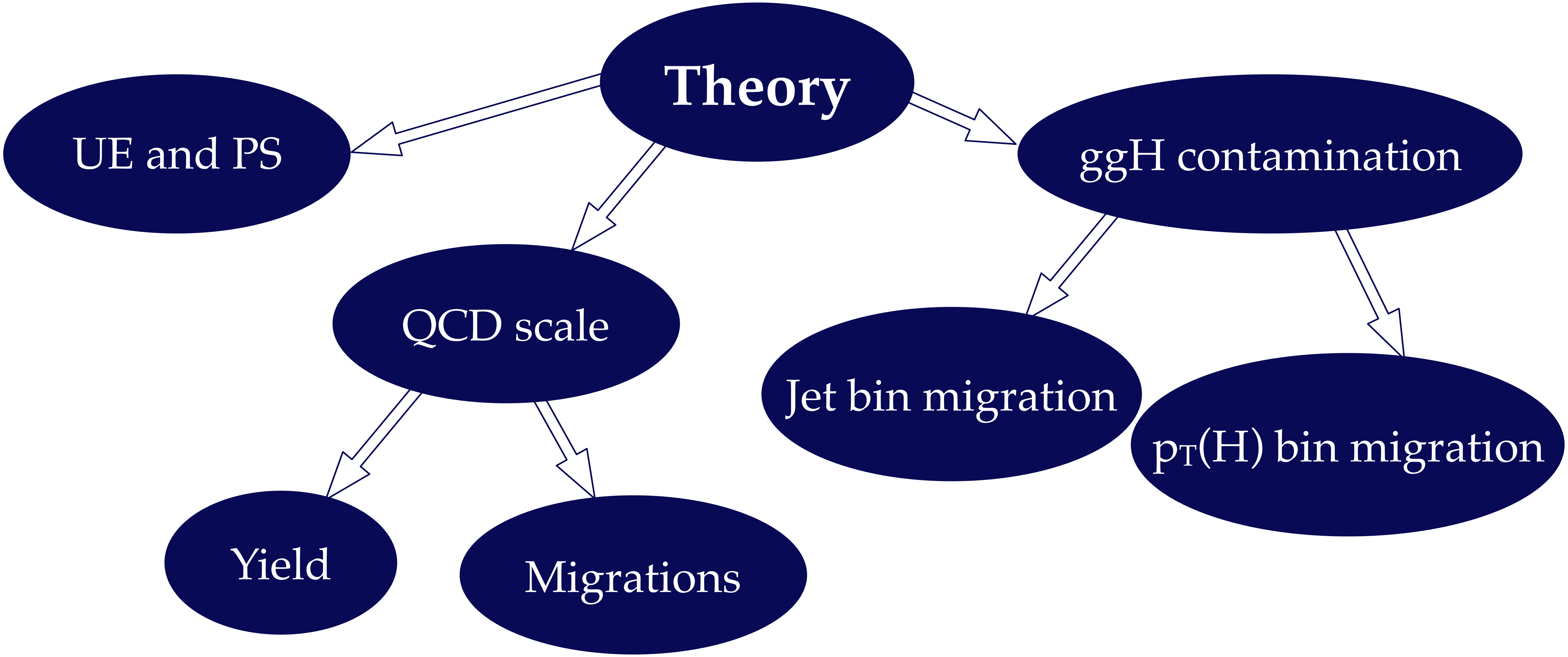
$H \rightarrow \gamma\gamma$: Differential

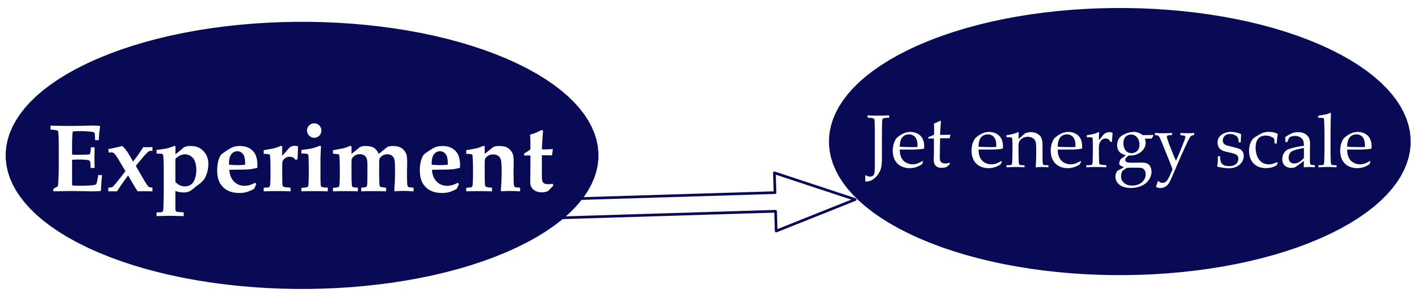




- Separate PAS (HIG-17-015) available in CDS here
- Fiducial region of isolated photons with $\mathbf{p_T/m_{\gamma\gamma} > 1/3\ (1/4),\ |\eta| < 2.5}$

Common systematics





Impact of systematics

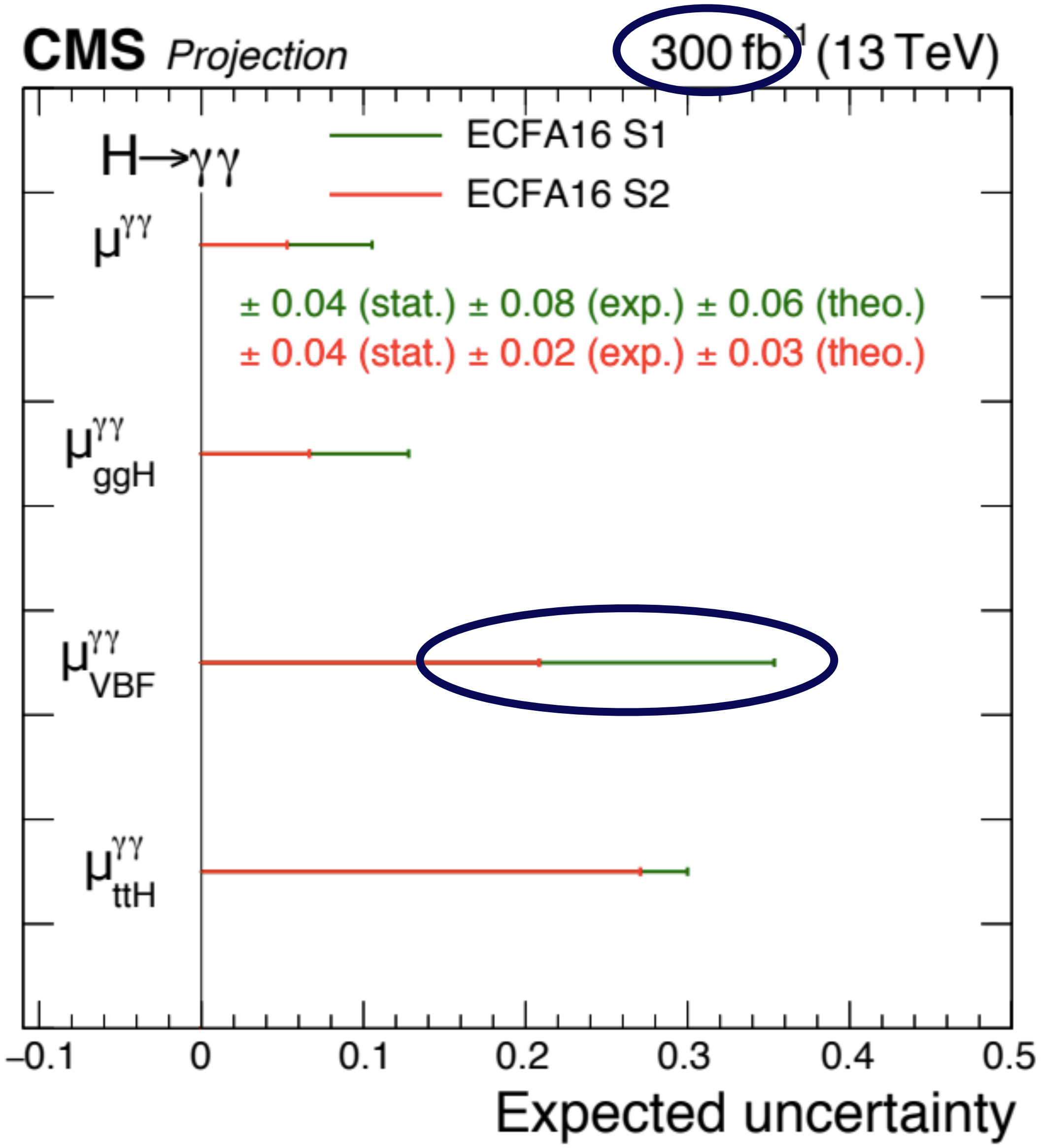
- **Statistical uncertainties are still dominant** for all channels with the 35.9fb^{-1} 2016 dataset

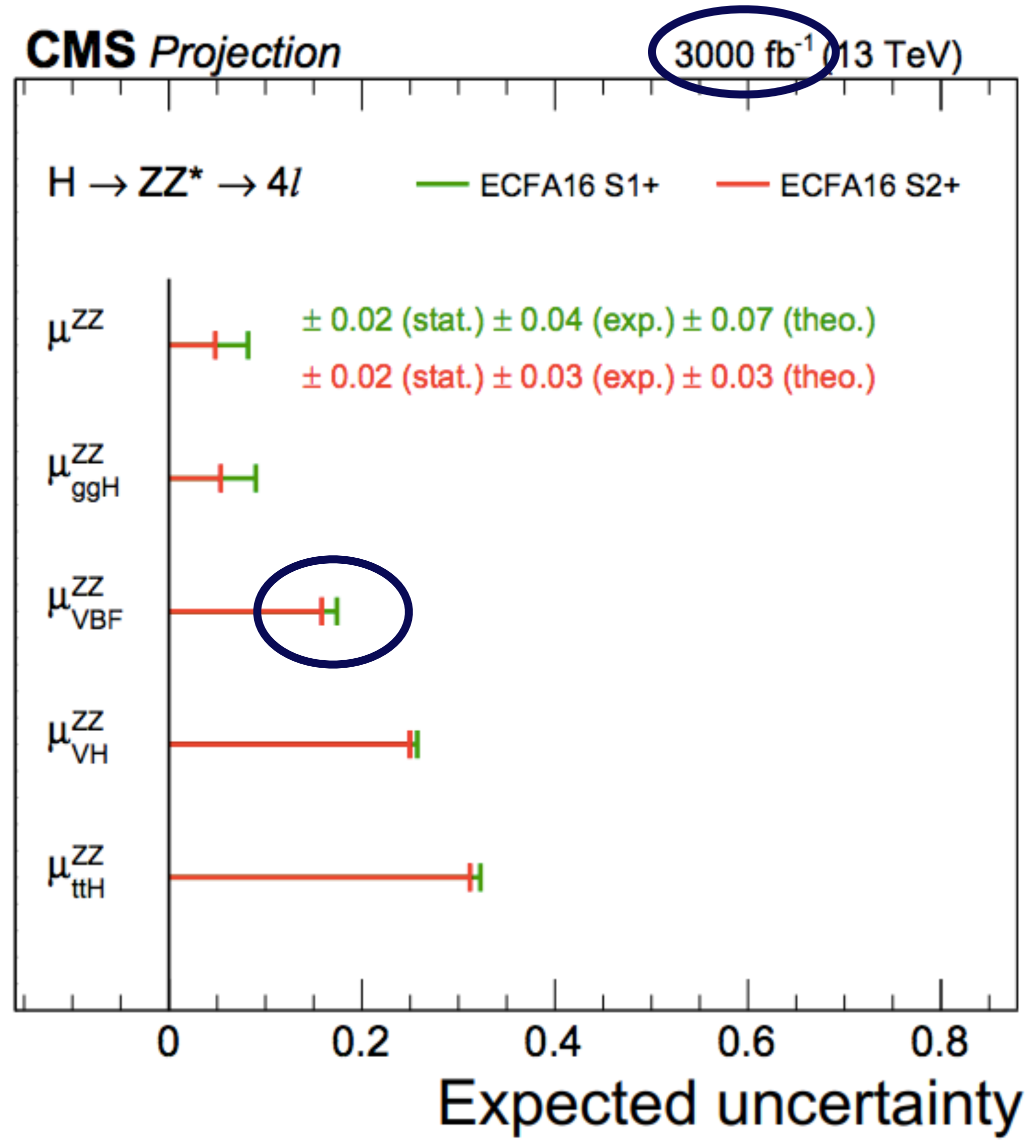
- **Systematics will play an important role soon** in some analyses
  - Currently majority from experimental uncertainties (esp jet energy scale)
  - These should decrease - then comparable to theory components

- For ZZ, the statistical component dominates even at 3000fb^{-1}
- However, in $\gamma\gamma$, systematics have large effect at just 300fb^{-1}

Impact of systematics

Imperial College London





- Projections based previous versions of the Run 2 analyses produced for ECFA 16, in PAS FTR-16-002 (here)
- Scenario 1: fixed systematic uncertainties
- Scenario 2: **expt uncertainties scale with √L, theory uncertainties halved**

Impact of systematics

- For each channel, total signal theory uncertainties are up to 20%, depending on process and category
  - highest for ggH signal in VBF categories → significant effect on μ_{VBF}

- In $\gamma\gamma$, uncertainty on μ_{VBF} : stat ±~0.5, syst ±~0.3, total ±~0.6
- Significant individual contributions:
  - Jet energy scale
  - ggH contamination (using Stewart-Tackmann method as described in YR3)
  - UE and PS (estimated by varying generator tunes)
  - Category migrations from QCD scale variation

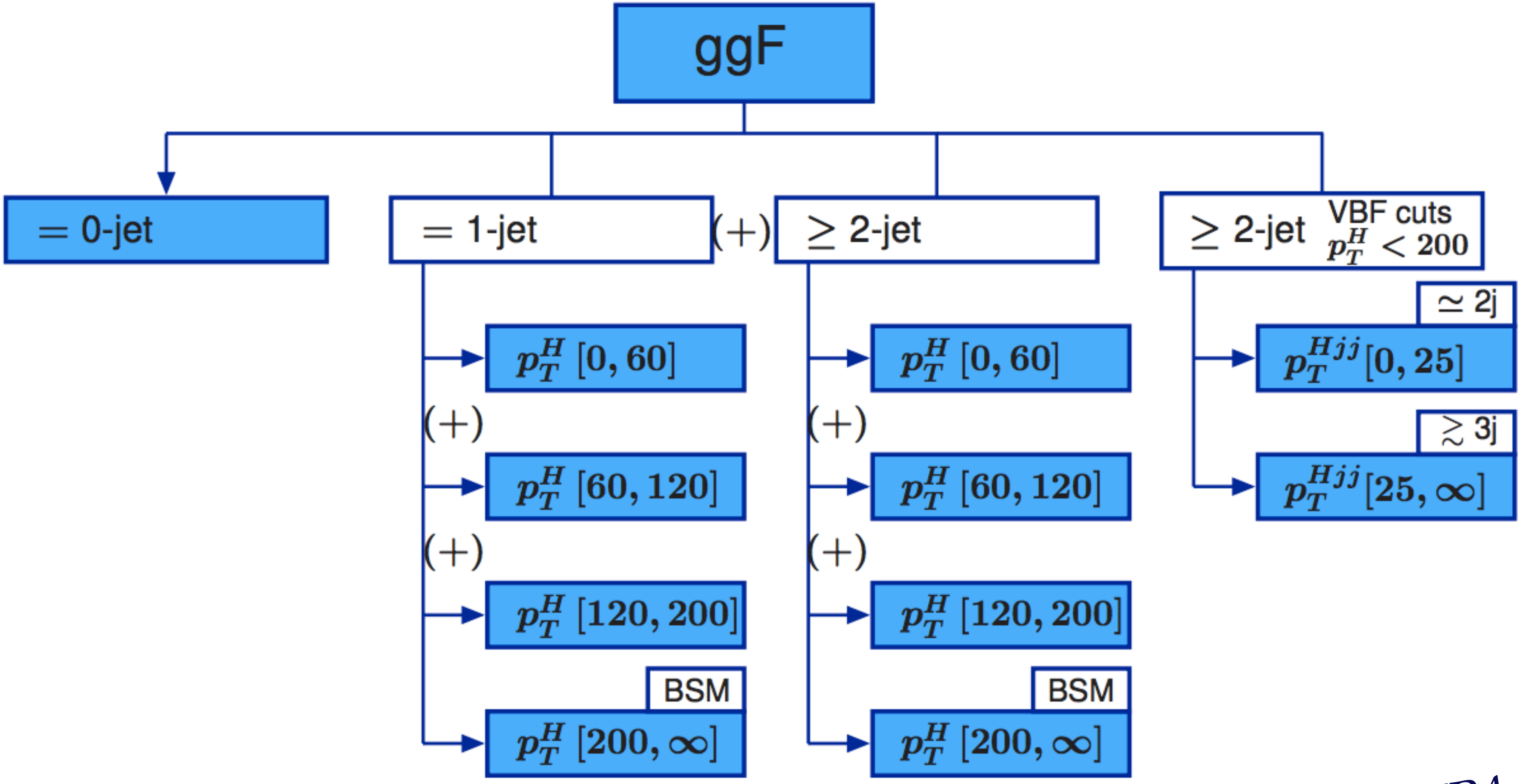
Impact of systematics

- Moving forward, will adopt the latest "2017" recommendation from WG1 for the treatment of ggH uncertainties, and NNLOPS for signal model MC
  - upcoming paper for $\gamma\gamma$ will include these developments

- Nine nuisances accounting for overall cross section, migration between jet and p_T phase spaces, and finite top mass corrections at high p_T
  - new VBF 2-jet and 2-jet veto sources that make VBF phase space uncorrelated from the exclusive 0 and 1 jet bins
  - replaces conservative ST jet veto approach in $\gamma\gamma$

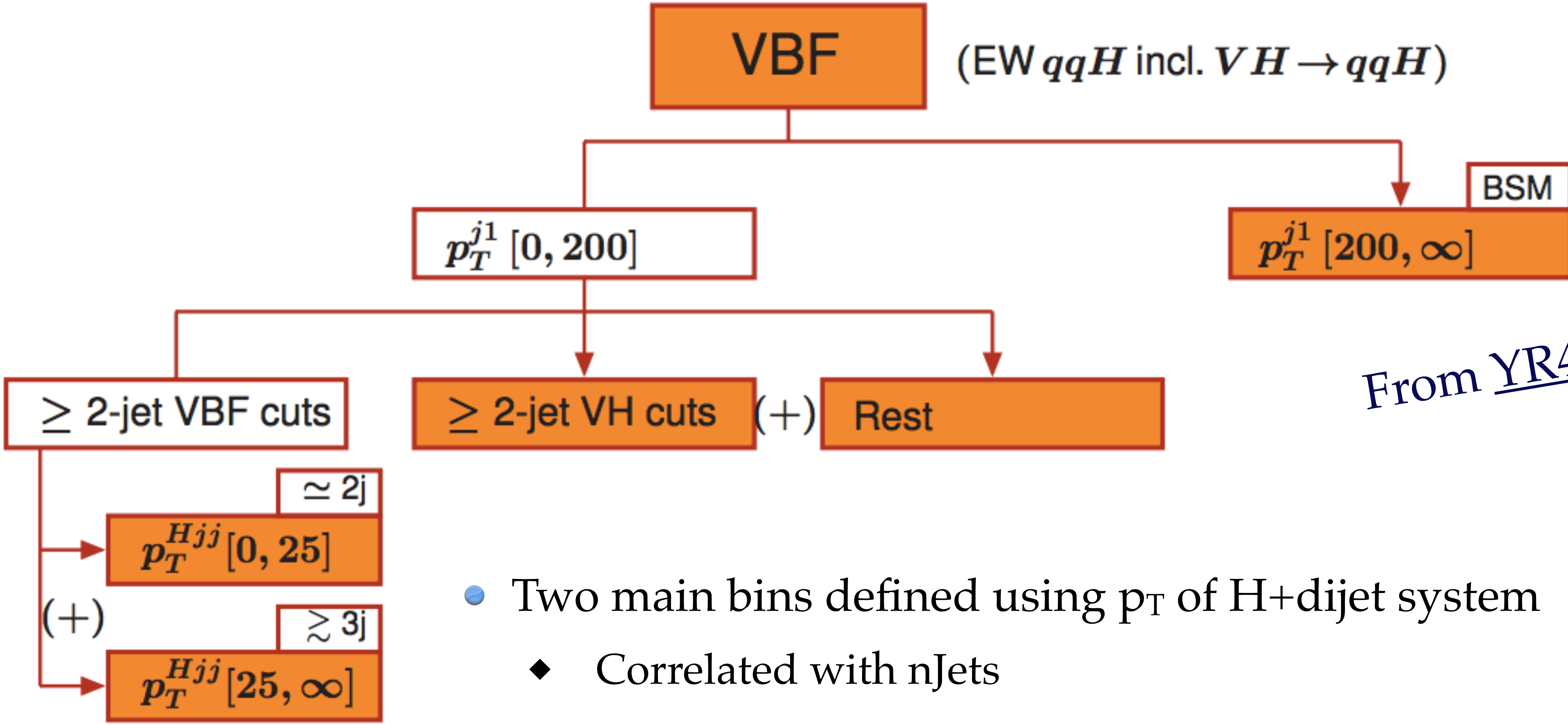
- Analyses are most sensitive to with high p_T Higgs, often with associated jets; change to signal model can affect this region substantially

STXS stage 1 - ggH



From YR4

STXS stage 1 - VBF



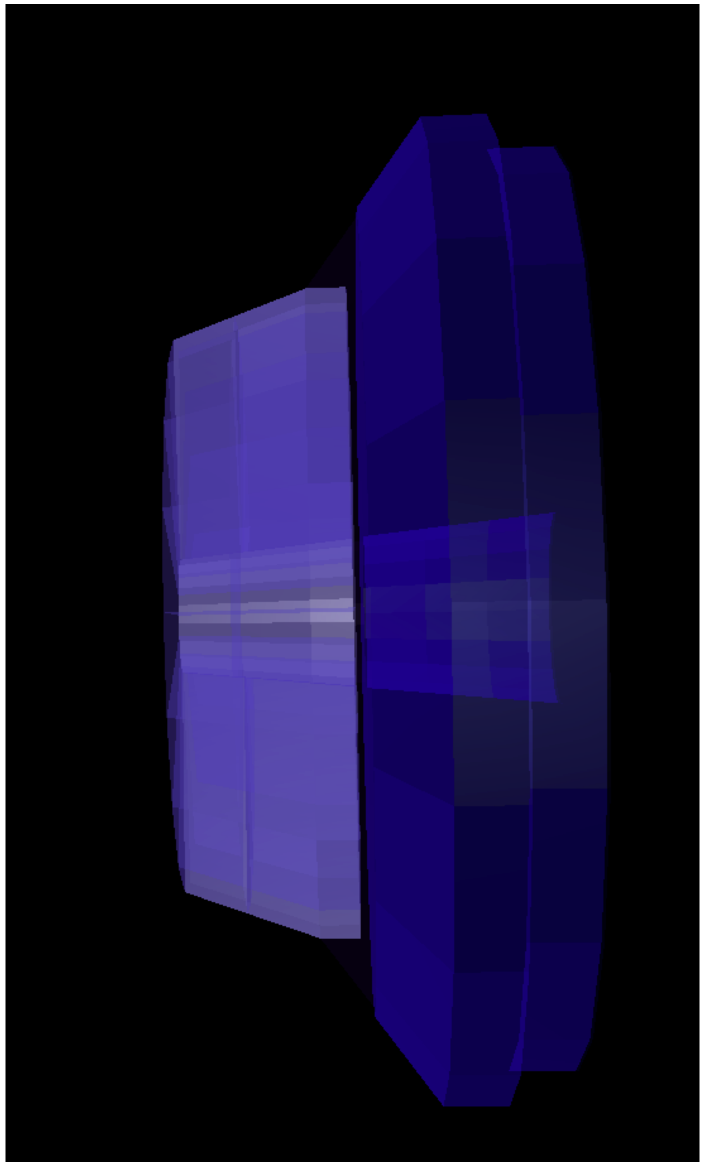
From YR4

- Two main bins defined using p_T of H+dijet system
  - Correlated with nJets
- Interesting for future analyses: discrimination between VBF and ggH+2-jets

Imperial College London

Phase 2 and HGCAL

- In the longer term, focus is on improving detector capability for future VBF measurements
- At CMS the High Granularity Calorimeter (HGCAL) will provide unprecedented amounts of information on forward jets
  - very fine segmentation
  - addition of depth info; 52 layers total
  - most likely some use of timing in addition



- Hopefully will open up new avenues for ggH vs VBF discrimination
  - e.g. improved ability to distinguish quark and gluon jets

Summary

- Several new CMS results containing VBF measurements using 2016 data
  - $H \rightarrow ZZ$: $\mu_{VBF} = 0.05^{+1.03}_{-0.05}$
  - $H \rightarrow \tau\tau$: $\mu_{VBF\text{-}tag} = 1.11^{+0.34}_{-0.35}$
  - $H \rightarrow \gamma\gamma$: $\mu_{VBF} = 0.5^{+0.6}_{-0.5}$
- Have now moved from discovery to precision measurements
- With more Run 2 data, provide STXS Stage 1 results
  - with further splitting of processes, into bins of $p_T(H)$, nJets and $p_T(Hjj)$
- And in long term, Phase 2 upgrade will bring new possibilities
  - prospect of reducing uncertainties to few per-cent level

Thank you