How charming is the Higgs boson?

L. Mastrolorenzo[1]

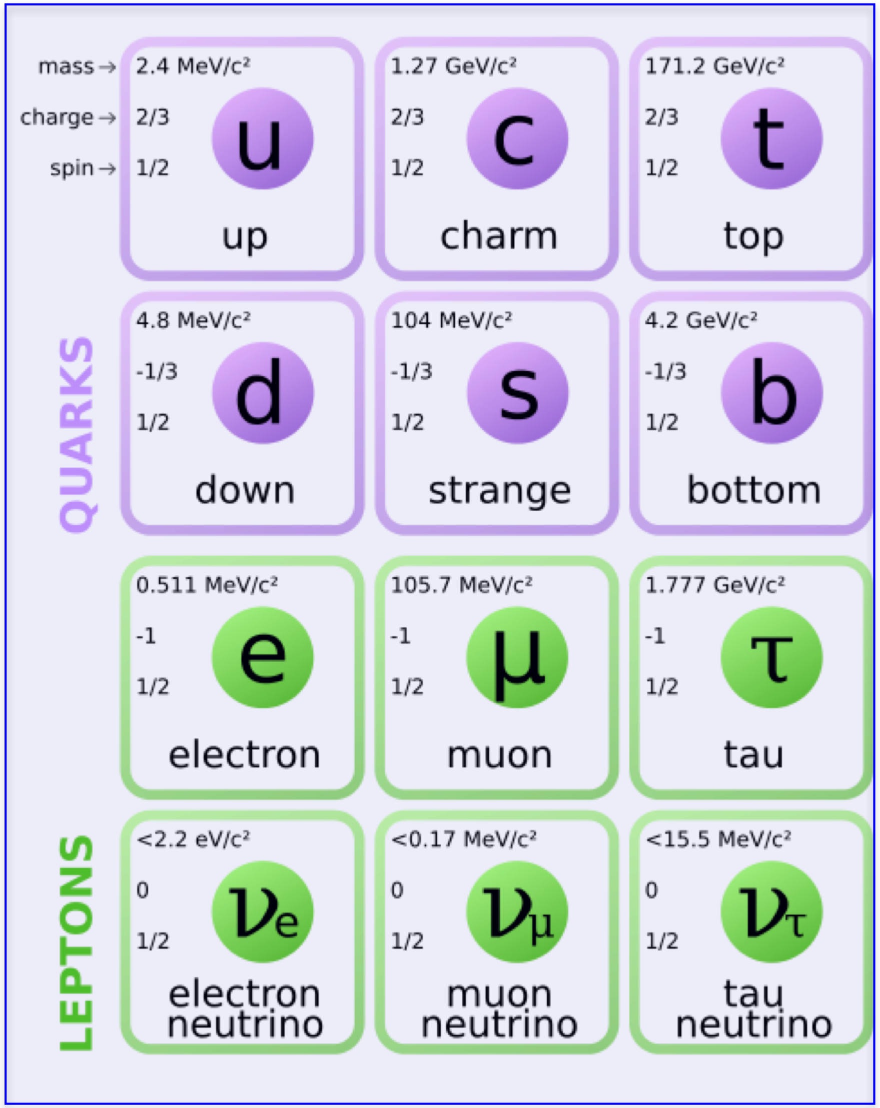
[1]RWTH Aachen University

Seminar

LLR, Ecolè Polytechnique, Palaiseau (France)

14 October 2019

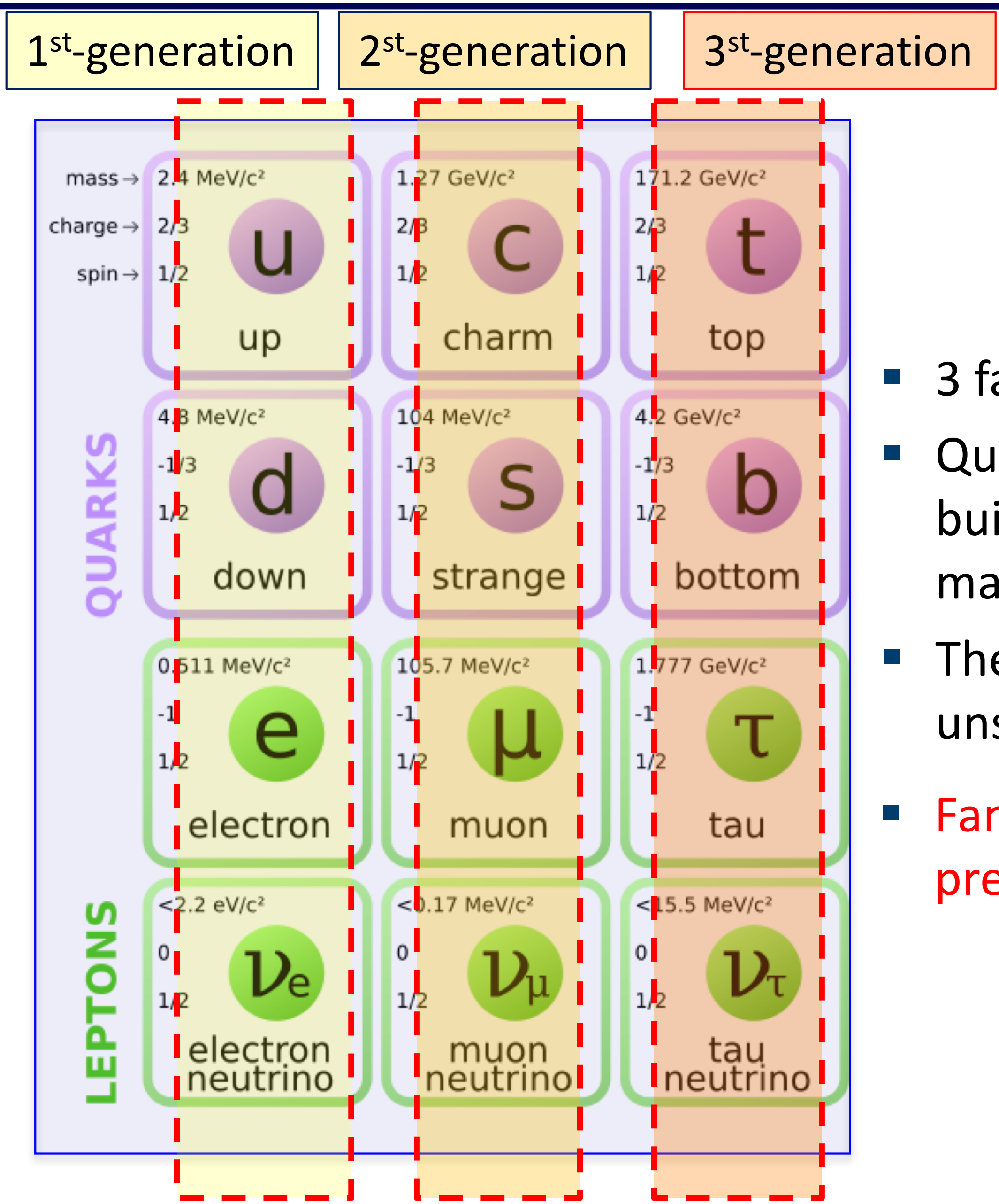
The Standard Model – brief summary



Matter Particles
Spin=1/2

- 3 families
- Quark u, d and electrons are the building brick of the ordinary matter
- The muon (μ) and tau (τ) are unstable leptons

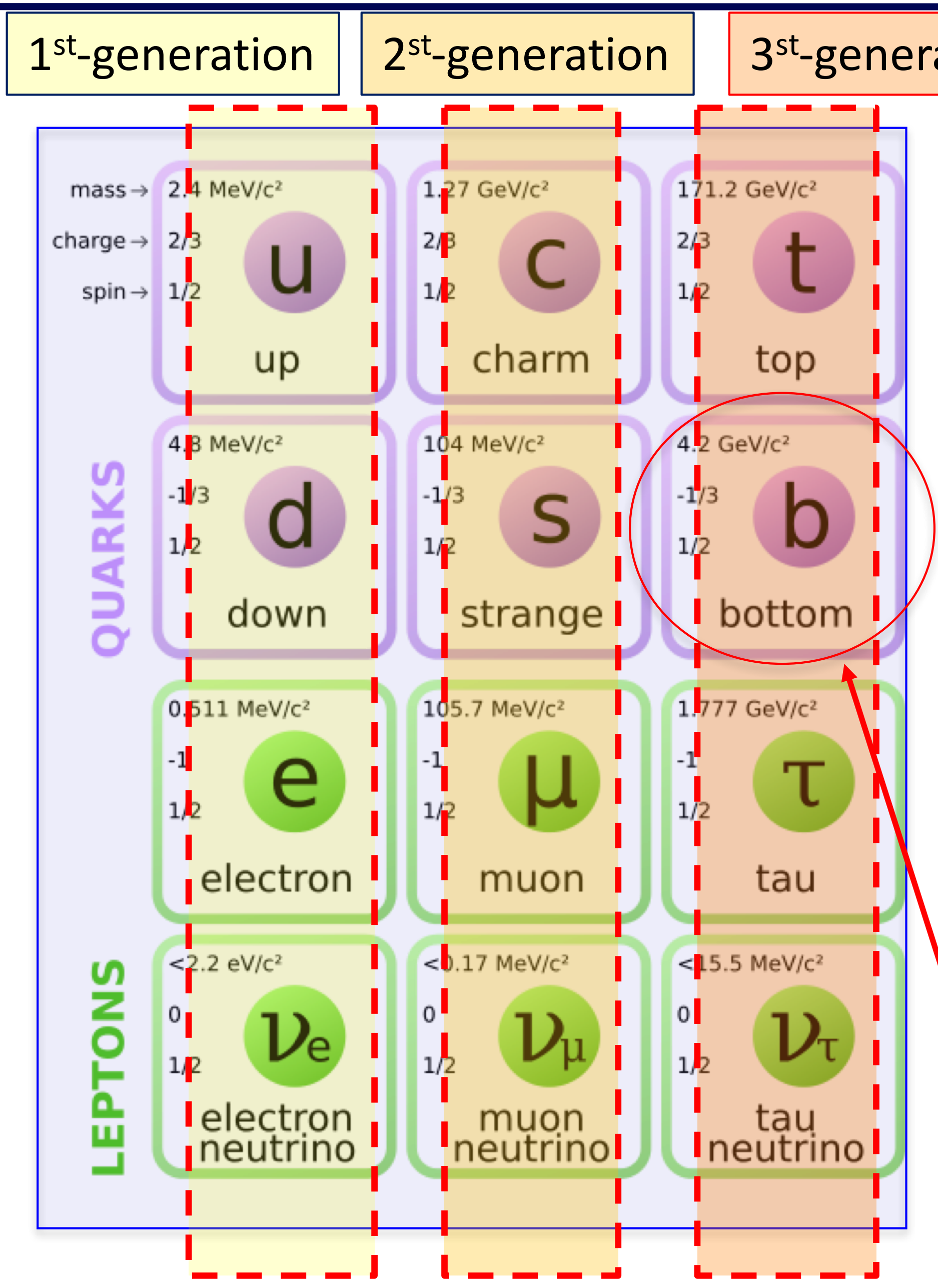
The Standard Model – brief summary



Matter Particles
Spin=1/2

- 3 families
- Quark u, d and electrons are the building brick of the ordinary matter
- The muon (μ) and tau (τ) are unstable leptons
- Families are grouped in a precise hierarchy or generation

The Standard Model – brief summary



Matter Particles
Spin=1/2

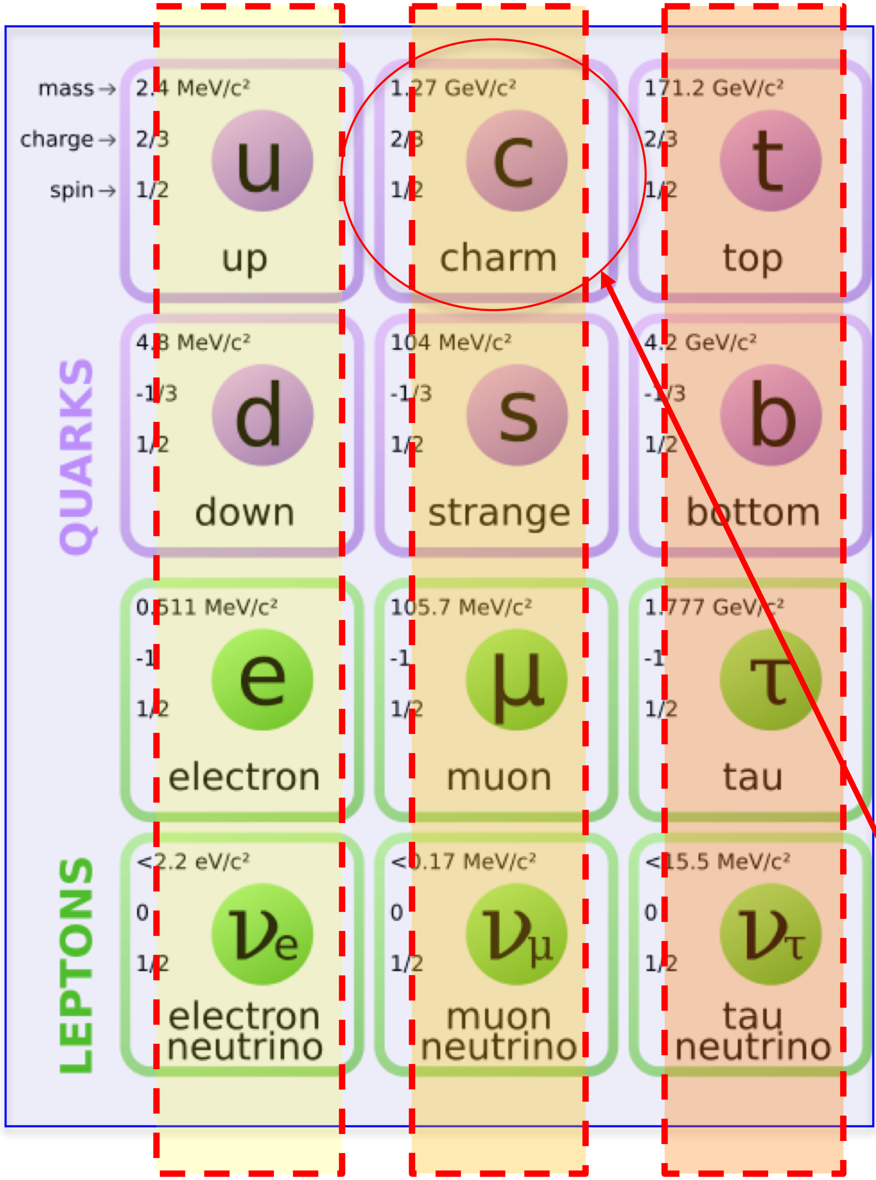
- 3 families
- Quark u, d and electrons are the building brick of the ordinary matter
- The muon (μ) and tau (τ) are unstable leptons

Focus on b-quark:

- 1977 by L. M. Lederman
- **3rd generation quark**
- $m_b = 4.2$ GeV/c^2
- "low-transition rate" ~10^{-12} s
- product in ~all top decays

The Standard Model – brief summary

1st-generation | 2st-generation | 3st-generation



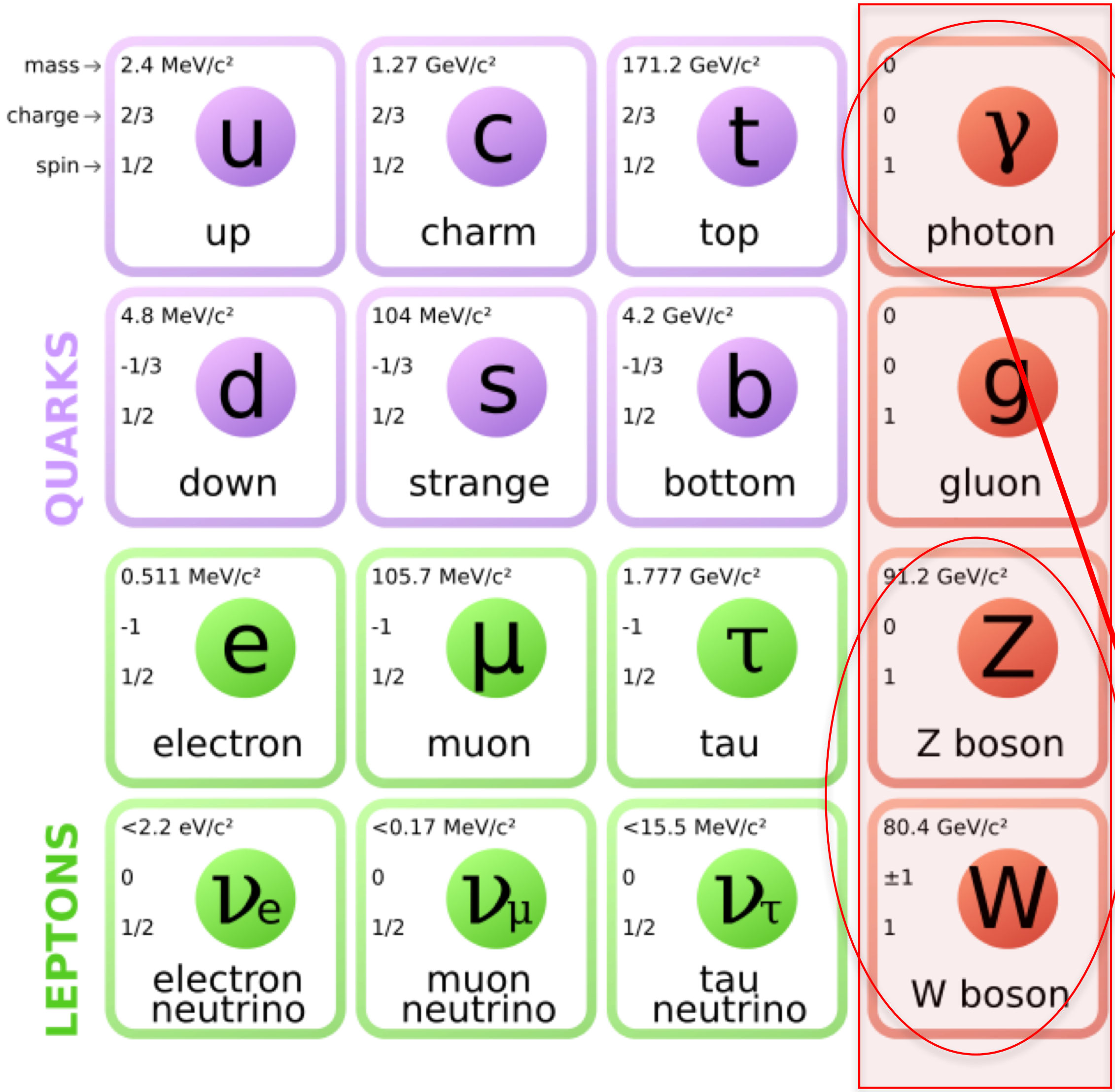
Matter Particles
Spin=1/2

- 3 families
- Quark u, d and electrons are the building brick of the ordinary matter
- The muon (μ) and tau (τ) are unstable leptons

Focus on c-quark:

- 1974 by B. Richter, S. Ting
- **2rd generation quark**
- $m_b = 1.3$ GeV/c^2
- transition rate ~10^{-13} s

The Standard Model – brief summary



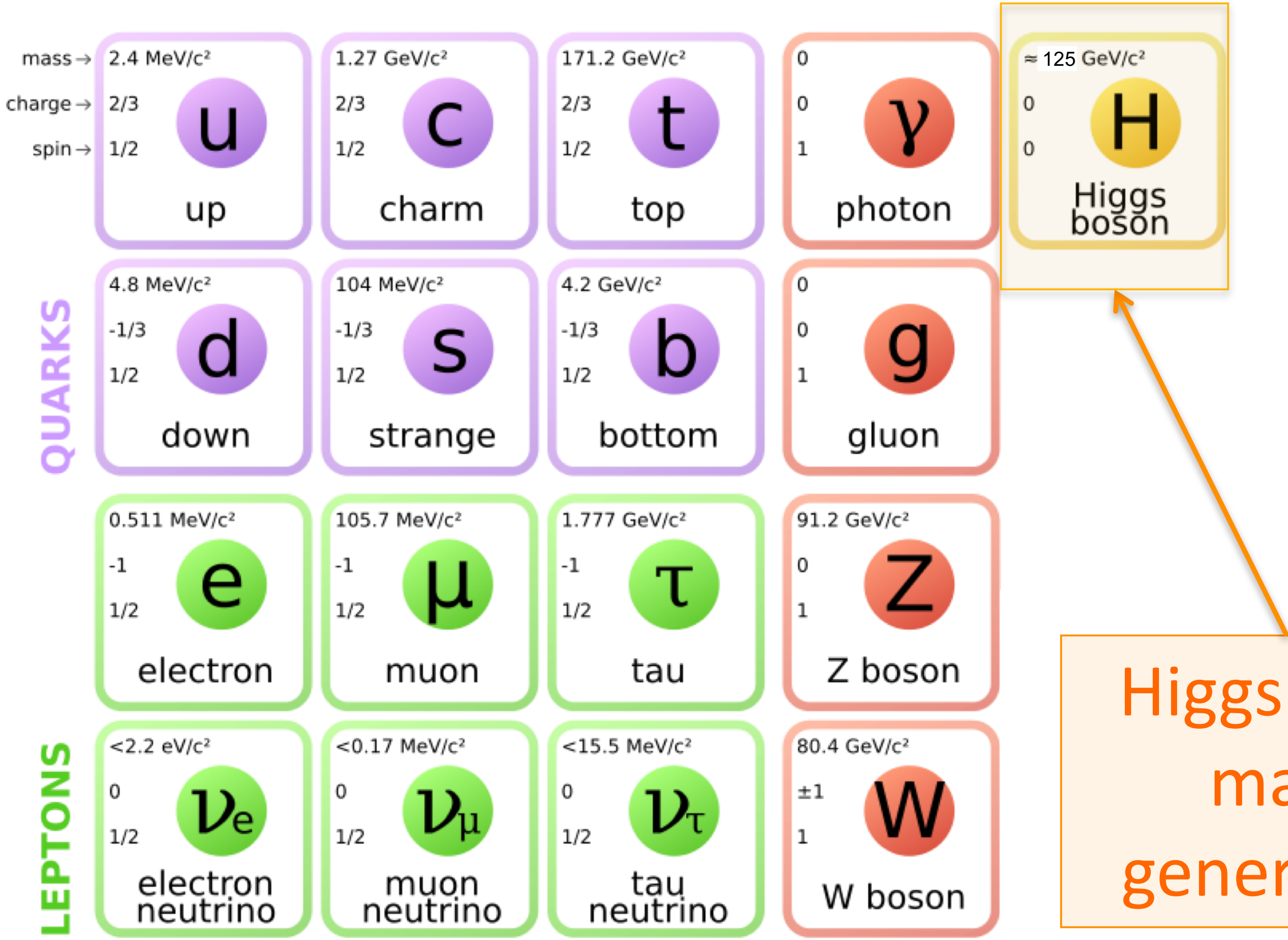
Interactions

Gauge bosons
Spin=1

EWK unification:

- EM and Weak interactions are considered as a single force

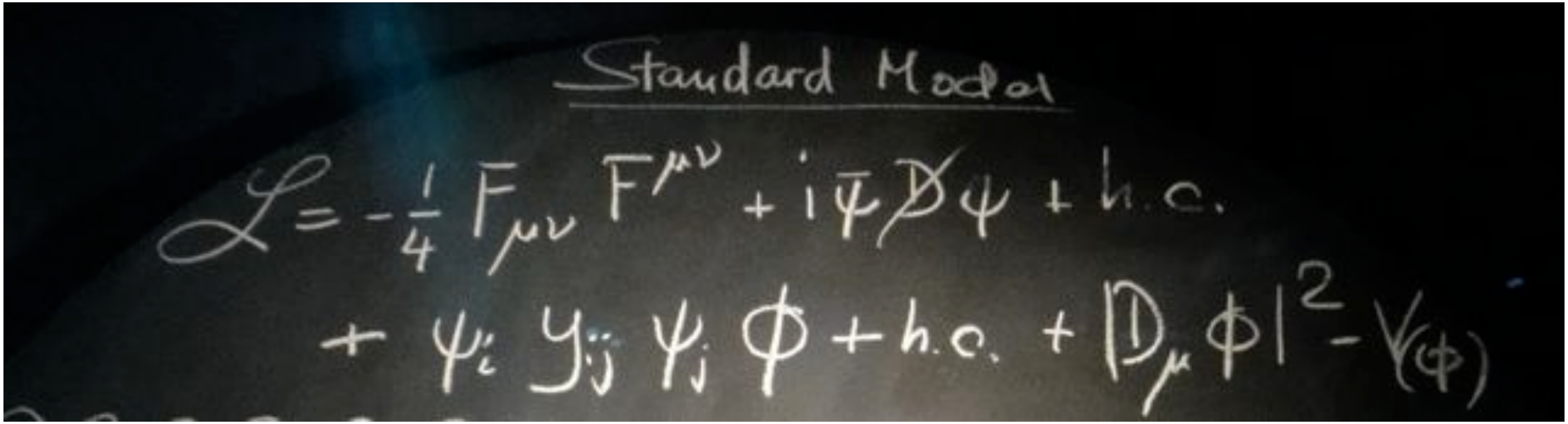
The Standard Model – brief summary



Higgs field: mass generation

- **The Standard Model of Paricle Physics**
  - The SM is a non-abelian, locally gauge invariant, quantum field theory (QFT) symmetric under local gauge transformation of the group:

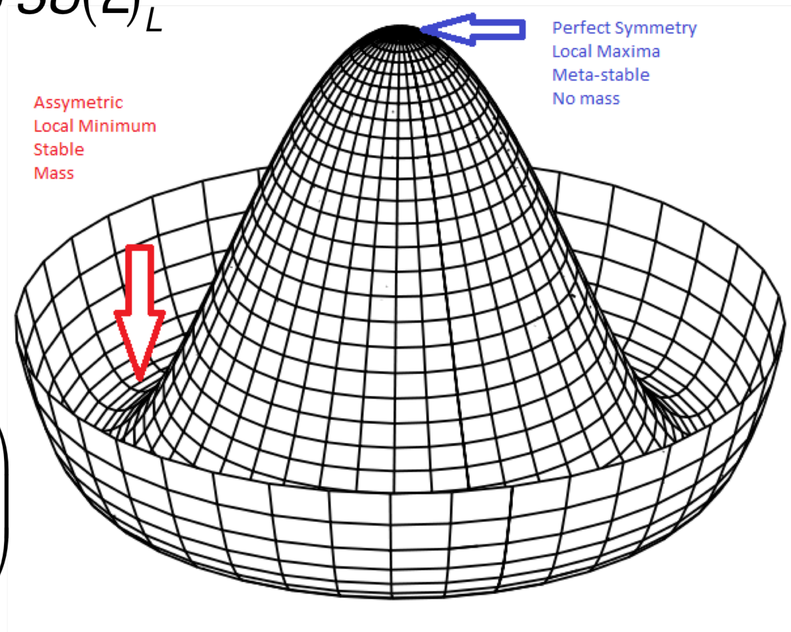
$$U(1)_Y \otimes SU(2)_L \otimes SU(3)$$



Brout-Englert-Higgs Mechanism

- **No explicit mass term in the SM lagrangian**
  - Adding by "hand" such terms ($m\Psi\bar{\Psi}$) would spoil the renormalizability of the SM
  - Particle can gain mass through the electroweak symmetry breaking mechanism

- **Introducing the "*Higgs potential*":** $V(\Phi) = -\mu^2\Phi\,\Phi + \lambda(\Phi\,\Phi)^2$
  - Invariant under local transformation $U(1)_Y \otimes SU(2)_L$
  - It must preserve Lorentz invariance
  - It breaks $U(1)_Y \otimes SU(2)_L \rightarrow U(1)_{em}$



$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}_L \xrightarrow[\text{Choice of a ground state}]{\text{small oscillations around the ground state}} \Phi = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

$$\Phi_0 = \begin{pmatrix} 0 \\ v \end{pmatrix}$$

vacuum expectation value ≠ 0

Brout-Englert-Higgs Mechanism

- **When the symmetry is spontaneously broken:**
  - The mass terms for the vector bosons **naturally appear** → $m_W = \frac{vg}{2}$, $m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}$
  - A **new massive particle** emerges: the Higgs boson → $m_H = \sqrt{2\lambda}v$
  - Fermion mass generation → **Yukawa couplings**

$$L_Y = f_l \overline{\chi}_L \phi l_R + f_u \overline{q}_L \tilde{\phi} u_R + f_d \overline{q}_L \phi d_R + \text{h.c.}$$

$$\phi = \begin{pmatrix} 0 \\ v + h \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}$$

$$L_Y = \frac{v f_l}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) + \frac{v f_u}{\sqrt{2}} (\bar{u}_L u_R + \bar{u}_R u_L) + \frac{v f_l}{\sqrt{2}} (\bar{d}_L d_R + \bar{d}_R d_L)$$

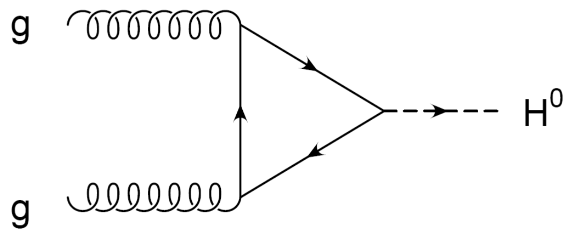
$$f_i = \frac{m_i}{v}\sqrt{2}$$

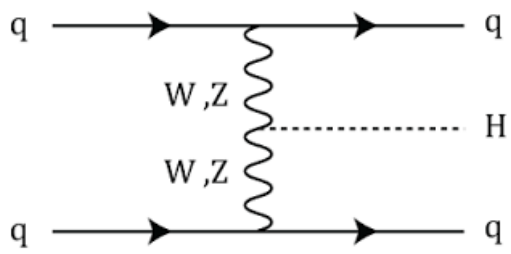
The Yukawa couplings bring new non-gauge interactions!

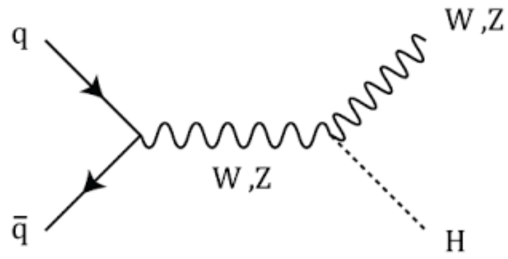
Represents something never probed before

The Higgs boson searches at the LHC

- **Main Higgs boson production mechanism at the LHC:**



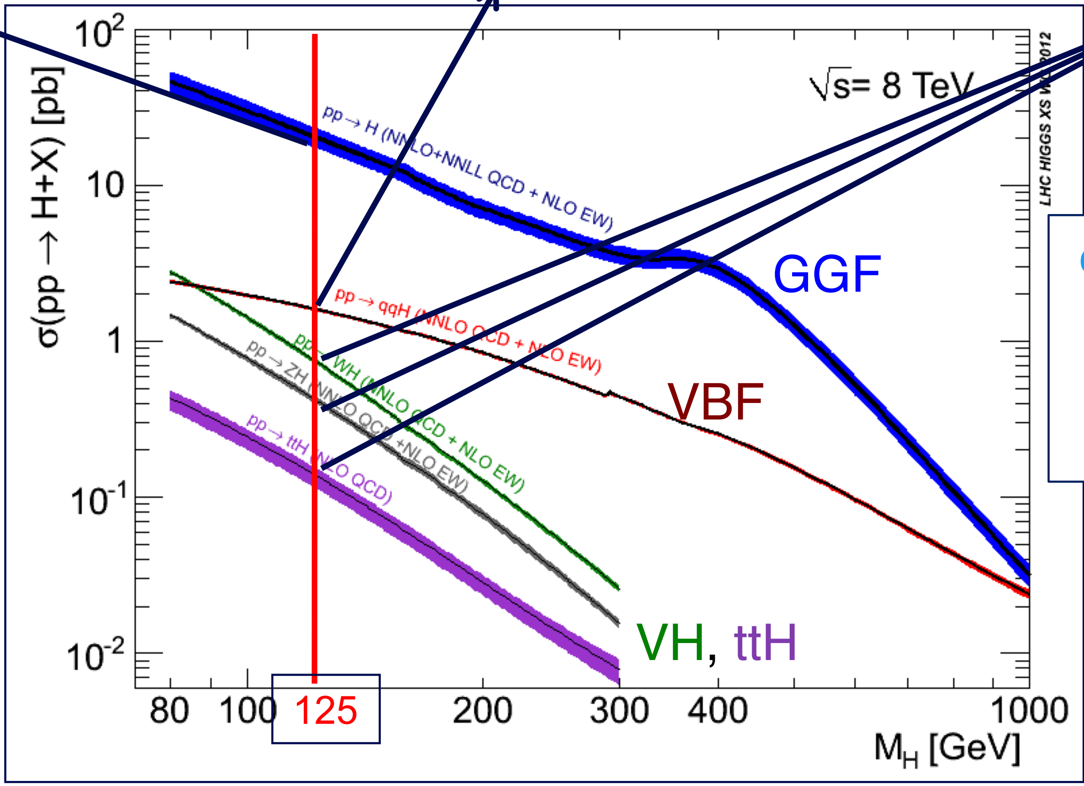




GGF: ~87%

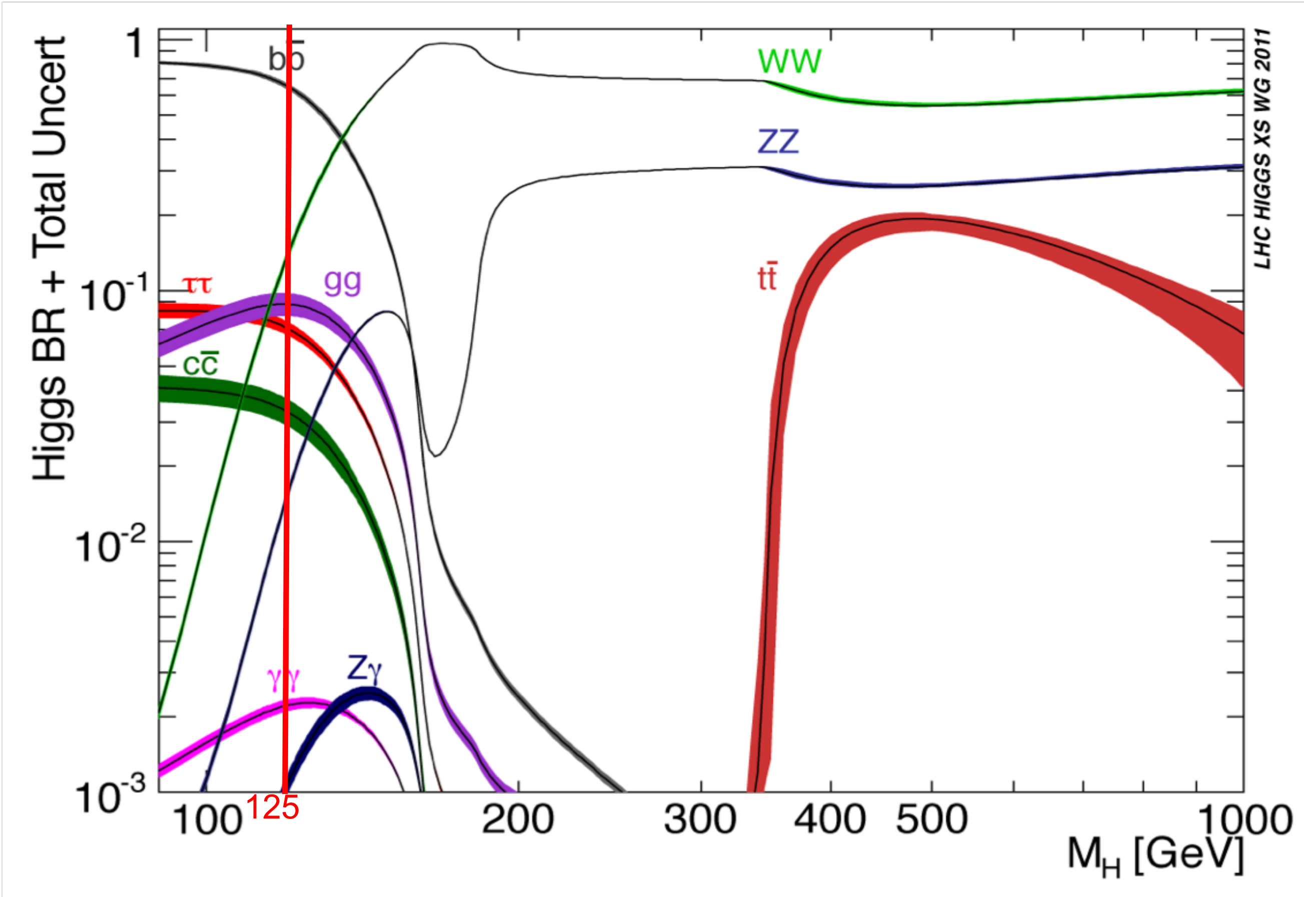
VBF: ~7%

WH: ~3%
ZH: ~2%
ttH: ~1%



crucial role in searches for VH(H→bb) and VH(H→cc) at LHC

The SM Higgs boson decay channels



- At 125 GeV, the highest branching ratio is into H→bb (~60%), followed by the WW channel (~20%). Then, the other sensitive channels also studied at the LHC are ττ (~6%), ZZ and γγ
- The most sensitive channels are ZZ→4l, γγ, WW→lνlν

Higgs boson discovery

At the time of my thesis!

Analysis in the main H decay channels

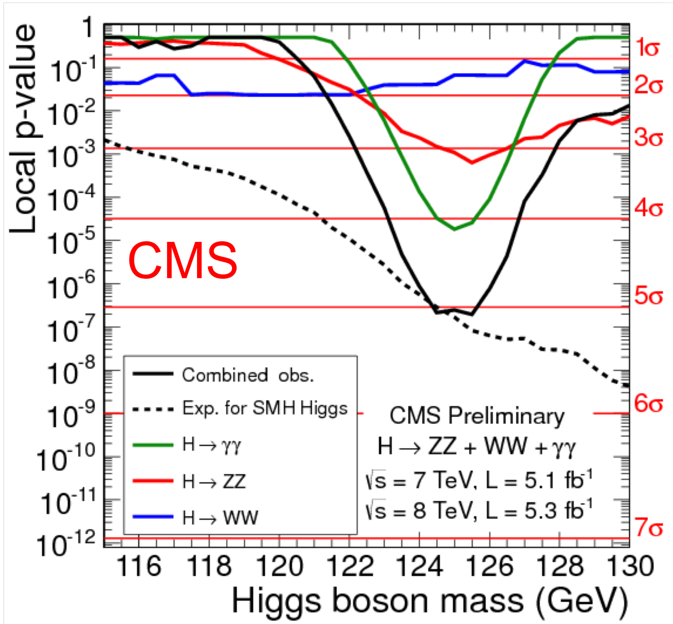
- $H \to ZZ \to 4l$
- $H \to \gamma\gamma$
- $H \to VV$
- $H \to b\bar{b}$ No evidence in 2012
- $H \to \tau\tau$ No evidence in 2012

Discovery in the bosonic decays 5.1σ

1st observation: ATLAS+CMS 2015

The CMS full combination in the five main decay modes 4.9σ

m_H=125.3±0.6 GeV

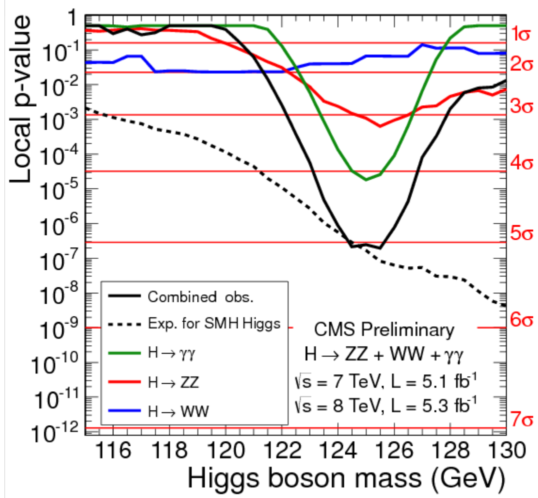


One year later…

P.Higgs and F. Englert were awarded the Nobel Prize in Physics

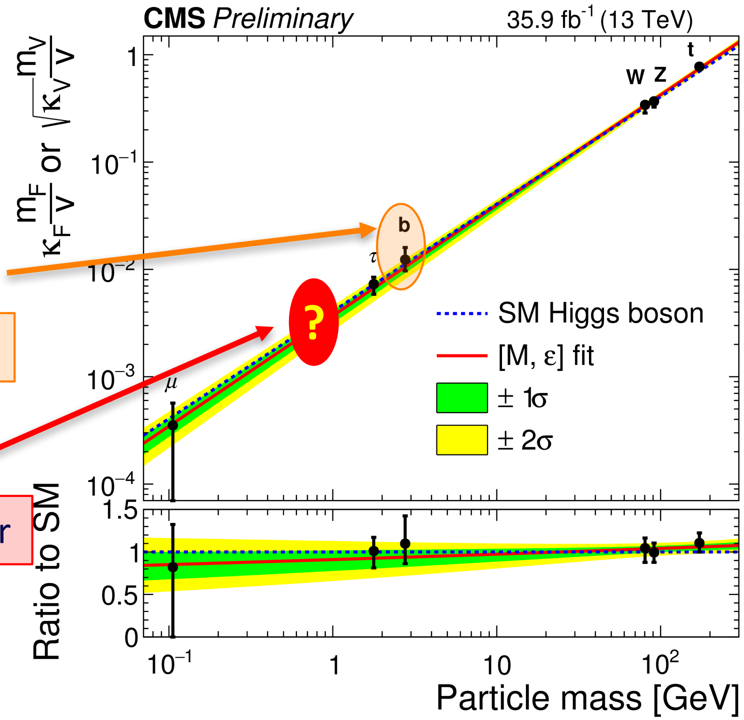
What we know today on the Higgs boson

- **Higgs discovery in 2012 ➔ characterization**
  - **Mass:** $125.09 \pm 0.21\,(\text{stat.}) \pm 0.11\,(\text{syst.})\ \text{GeV}$
    ATLAS+CMS: PRL 114 (2015) 191803
  - **Spin/Parity:** 0^+
    ATLAS: EPJC 75 (2015) 476
    CMS: PRD 92 (2015) 012004
  - **Width:** on-shell + off-shell searches comb .<3.2MeV
    CMS: JHEP 11 (2017) 047
    ATLAS: arXiv:1808.01191 submitted to PLB



- **Observed direct coupling to:**
  - **Vector bosons**
    ATLAS: PLB 716 (2012) 1-29
    CMS: PLB 716 (2012) 30
  - **τ leptons**
    ATLAS: ATLAS-CONF-2018-021
    CMS: PLB 779 (2018) 283
  - **top quarks**
    ATLAS: PLB 784 (2018) 173
    CMS: PRL 120 (2018) 231801
  - **bottom quarks**
    PhysRevLett. 121.121801
  - **charm quarks**
    No direct search in CMS so far

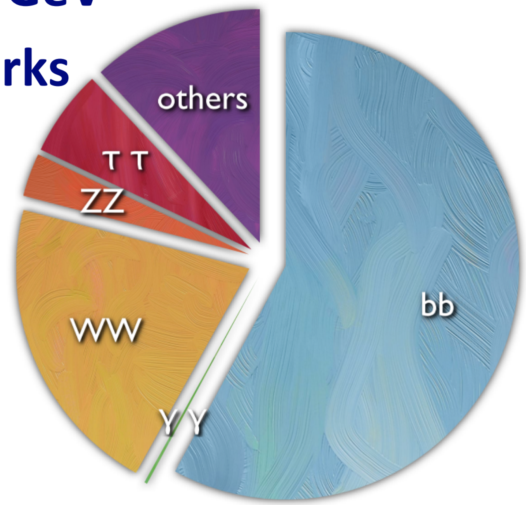
Today's talk!



So far, all measurements compatible with SM predictions!

H→bb physics case

- **H→bb has the largest branching fraction (58%) for m_H=125 GeV**
- Unique final state to measure **coupling with down-type quarks**
- **Drives the uncertainty on the total Higgs boson width**
- Limits the sensitivity to BSM contributions
- **Not observed until this Summer**

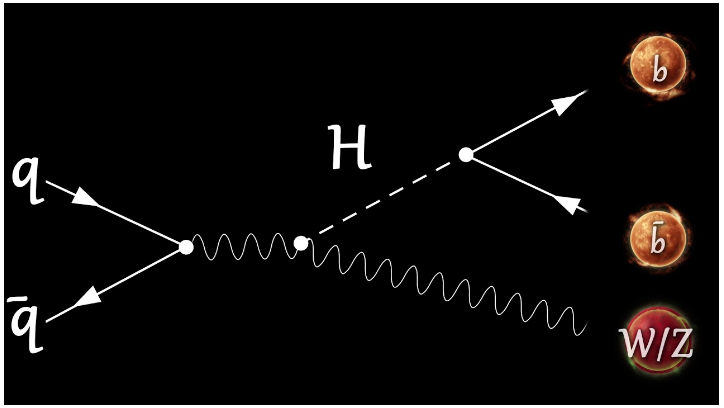


Many feature similar to searches for H→cc!

- **High BR**
- **Low mass resolution**
- **Low S/B**

→

- **Highly efficient b-jets identification**
- **Improved resolution on m(bb)**
- **Full event information to increase S/B**



Higgs-Strahlung (Associated production)

- 4% of Higgs production mechanism
- Benefit from leptons triggers
- Further reduce background requiring high V-p_T

➔ **Provides the most sensitive channel**

First H→bb searches started at LEP…

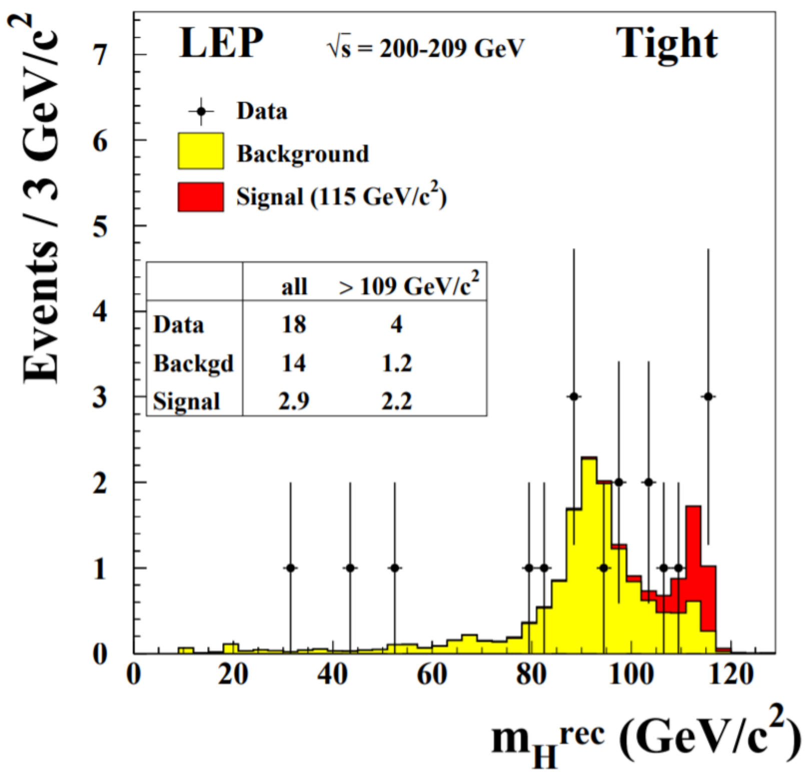
Physics Letters B 565 (2003) 61–75

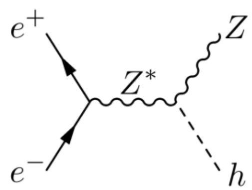
Search for the Standard Model Higgs boson at LEP

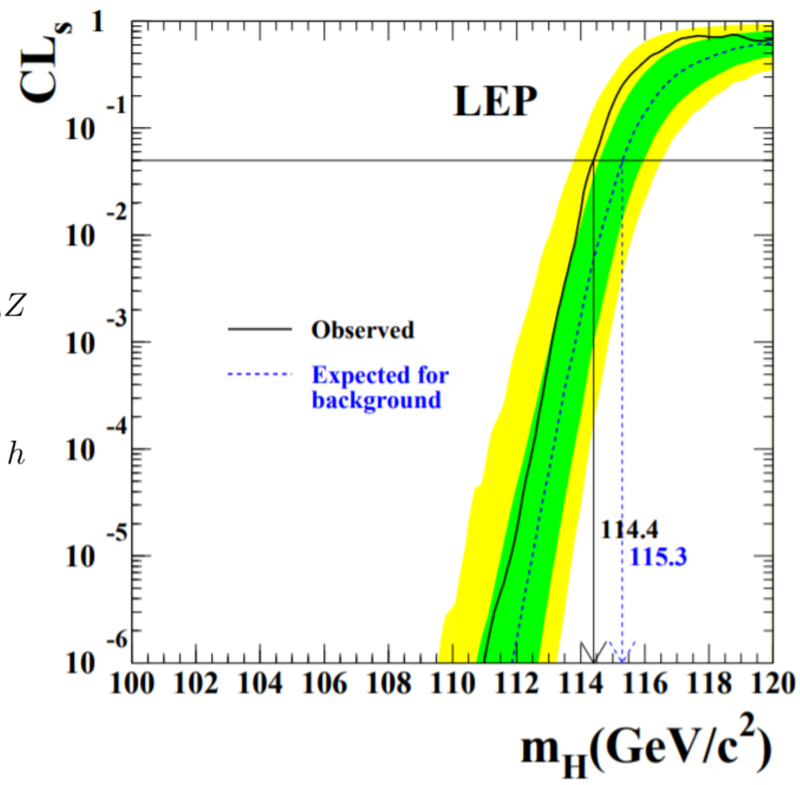
ALEPH Collaboration [1] DELPHI Collaboration [2] L3 Collaboration [3] OPAL Collaboration [4]
The LEP Working Group for Higgs Boson Searches [5]

ELSEVIER
PHYSICS LETTERS B

m_H > 114.4 GeV @ 95%CL





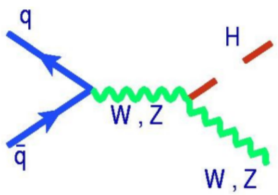


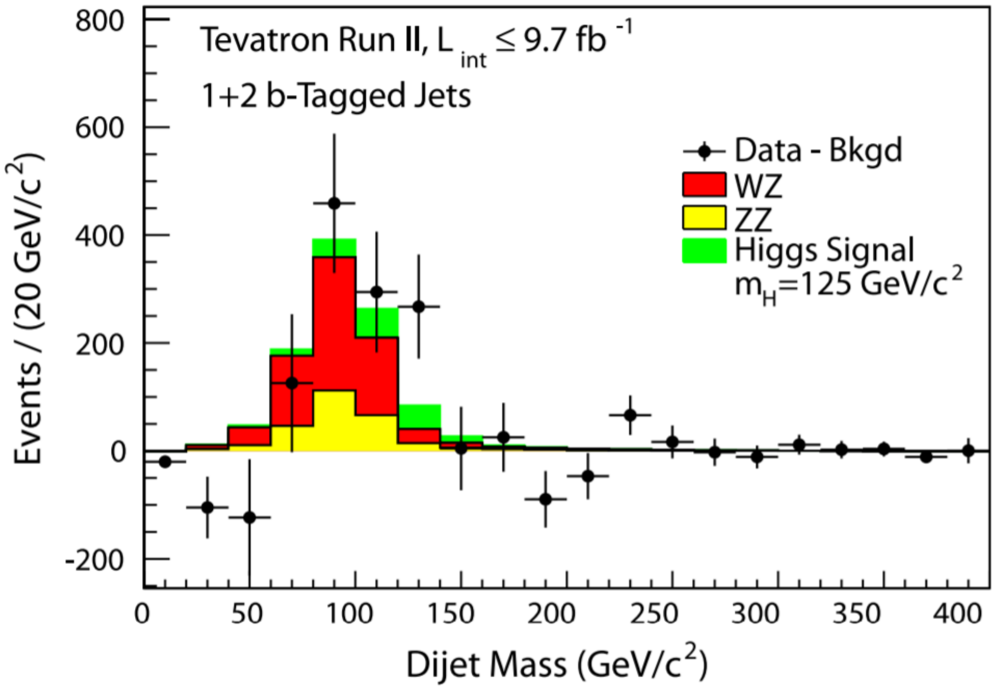
…and continued at Tevatron…

PRL **109,** 071804 (2012) PHYSICAL REVIEW LETTERS week ending 17 AUGUST 2012

Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron

(*CDF Collaboration)
(†D0 Collaboration)





Significance

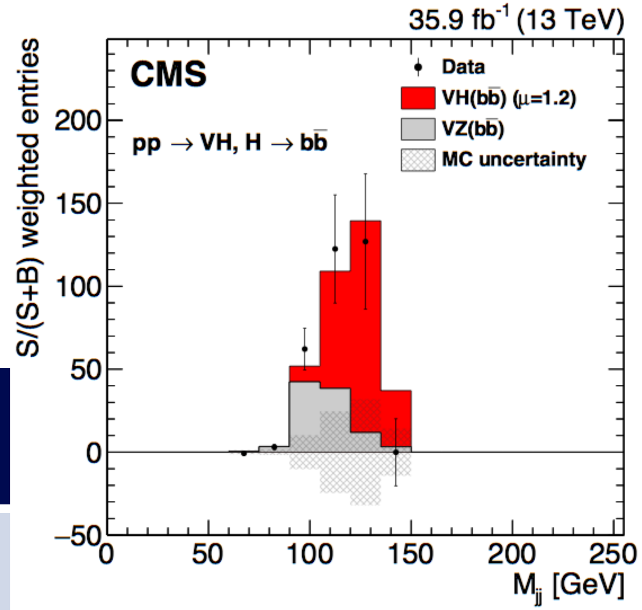
2.8σ observed @ 125 GeV

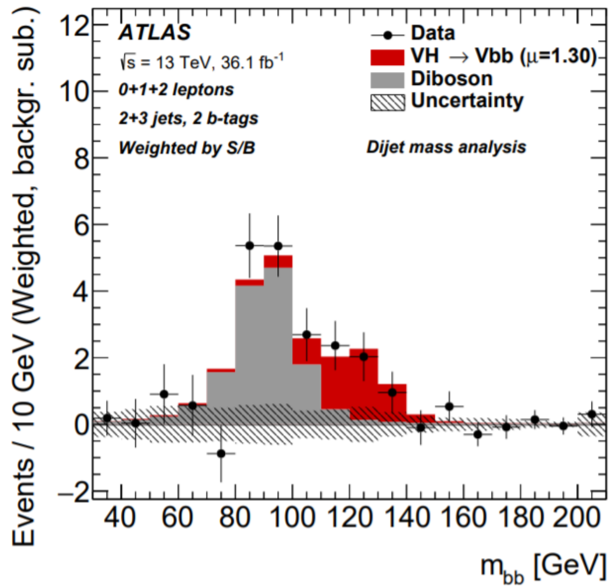
...and also at the LHC!

- **VH(bb) evidence at LHC established with 2016 data by both ATLAS and CMS**
  - Detectors demonstrated ability to deal with very high PU
  - For 2016 analyses used ~40fb^{-1}
- **Signal strength uncertainty ~40%**

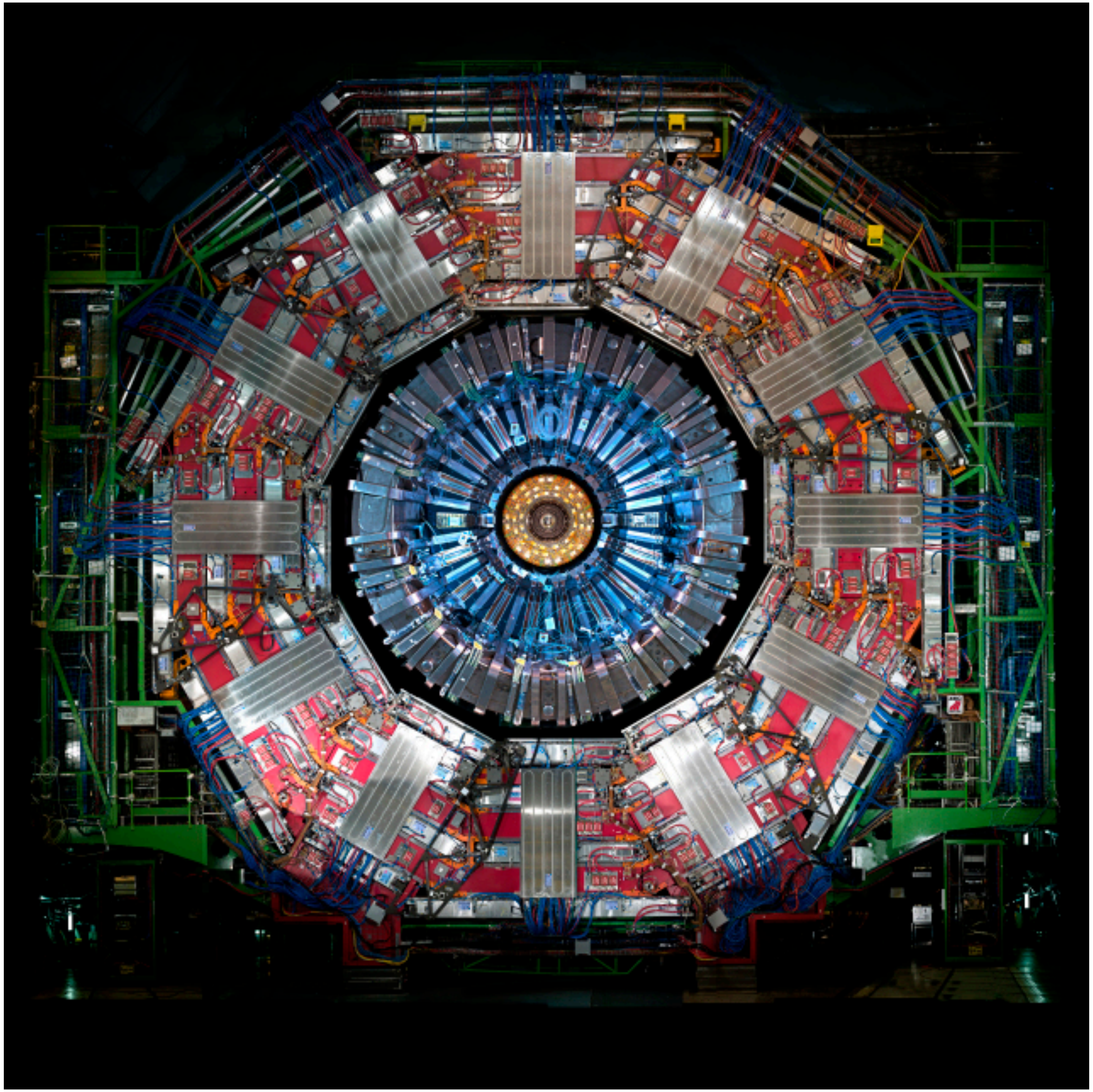
	μ	Significance (exp.)	Significance (obs.)
ATLAS Run 1 [1]	$0.52^{+0.40}_{-0.37}$	2.6σ	1.4σ
CMS Run 1 [2]	$0.89^{+0.47}_{-0.44}$	2.5σ	2.1σ
ATLAS+CMS Run 1 [3]	$0.79^{+0.29}_{-0.27}$	3.7σ	2.6σ
ATLAS 2015+2016 [4]	$1.20^{+0.42}_{-0.36}$	3.0σ	**3.5σ**
CMS 2016 [5]	$1.19^{+0.40}_{-0.38}$	2.8σ	**3.3σ**

[1] JHEP 01 (2015) 069
[2] JHEP 08 (2016) 045
[3] JHEP 08 (2016) 045
[4] JHEP 12 (2017) 024
[5] PLB 780 (2018) 501



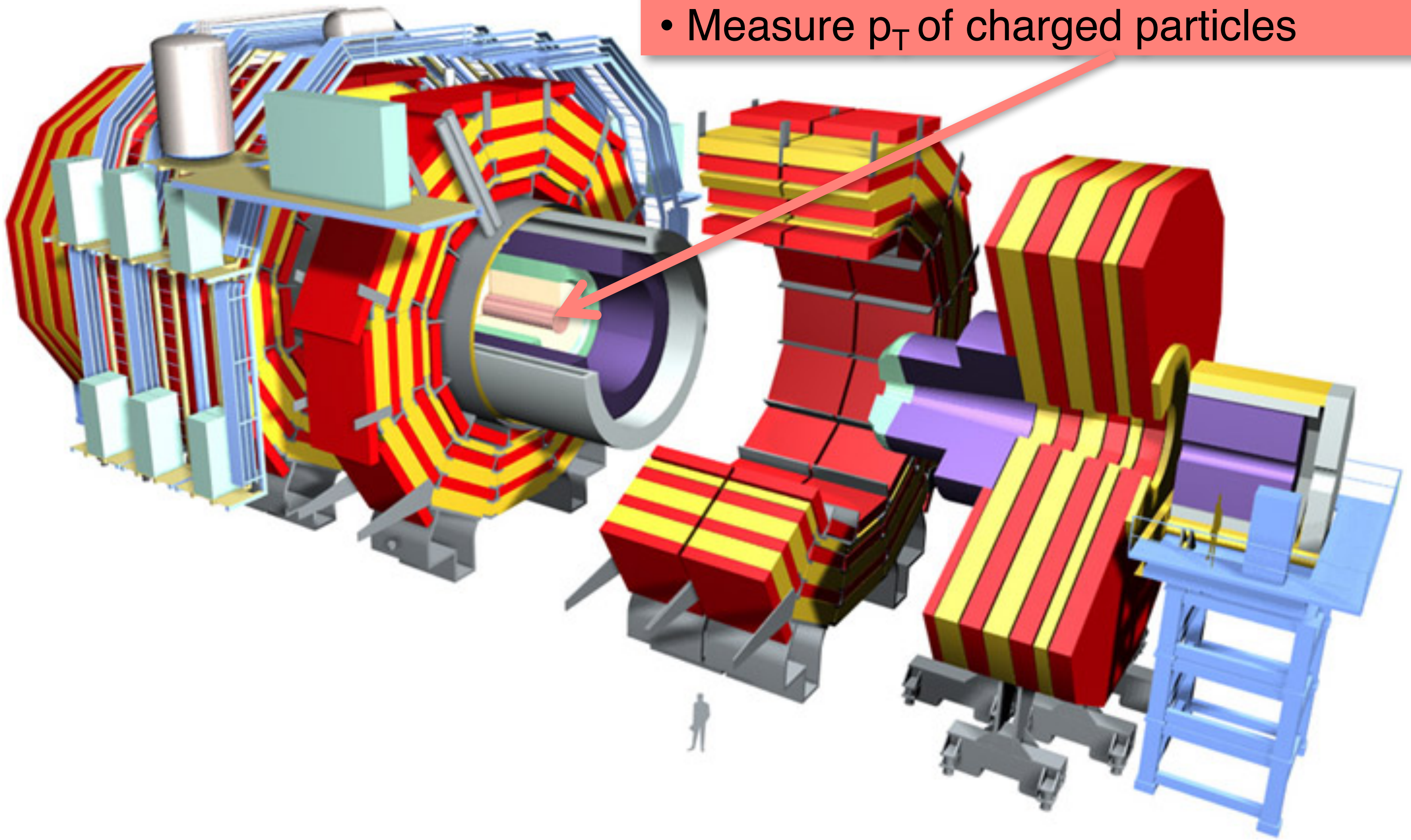


The Compact Muon Solenoid (CMS) detector



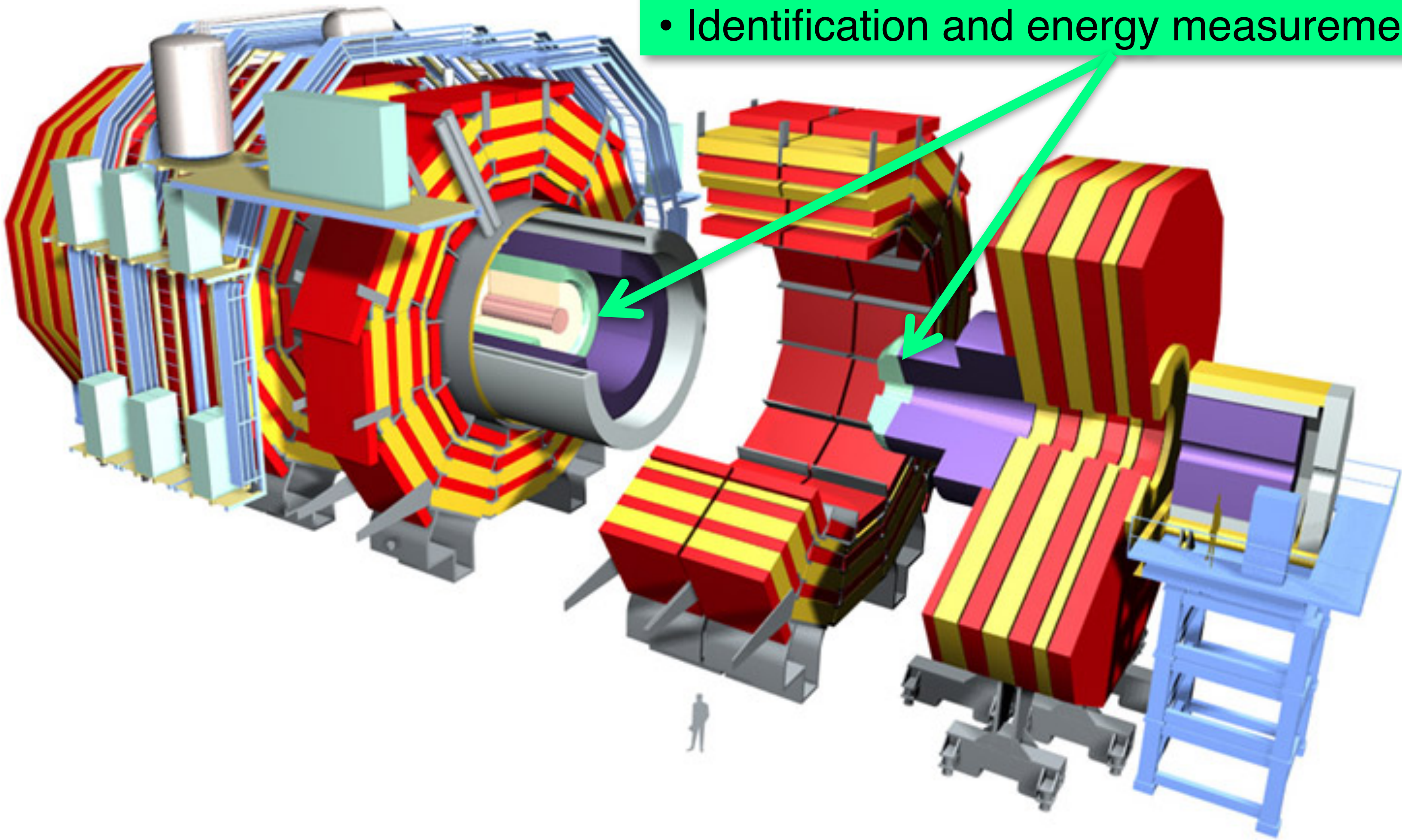
The Compact Muon Solenoid (CMS) detector

Tracker:
- Length = 6m, diameter = 2.4 m
- Silicon detectors (100µm x 150µm x 250µm)
- Measure p_T of charged particles

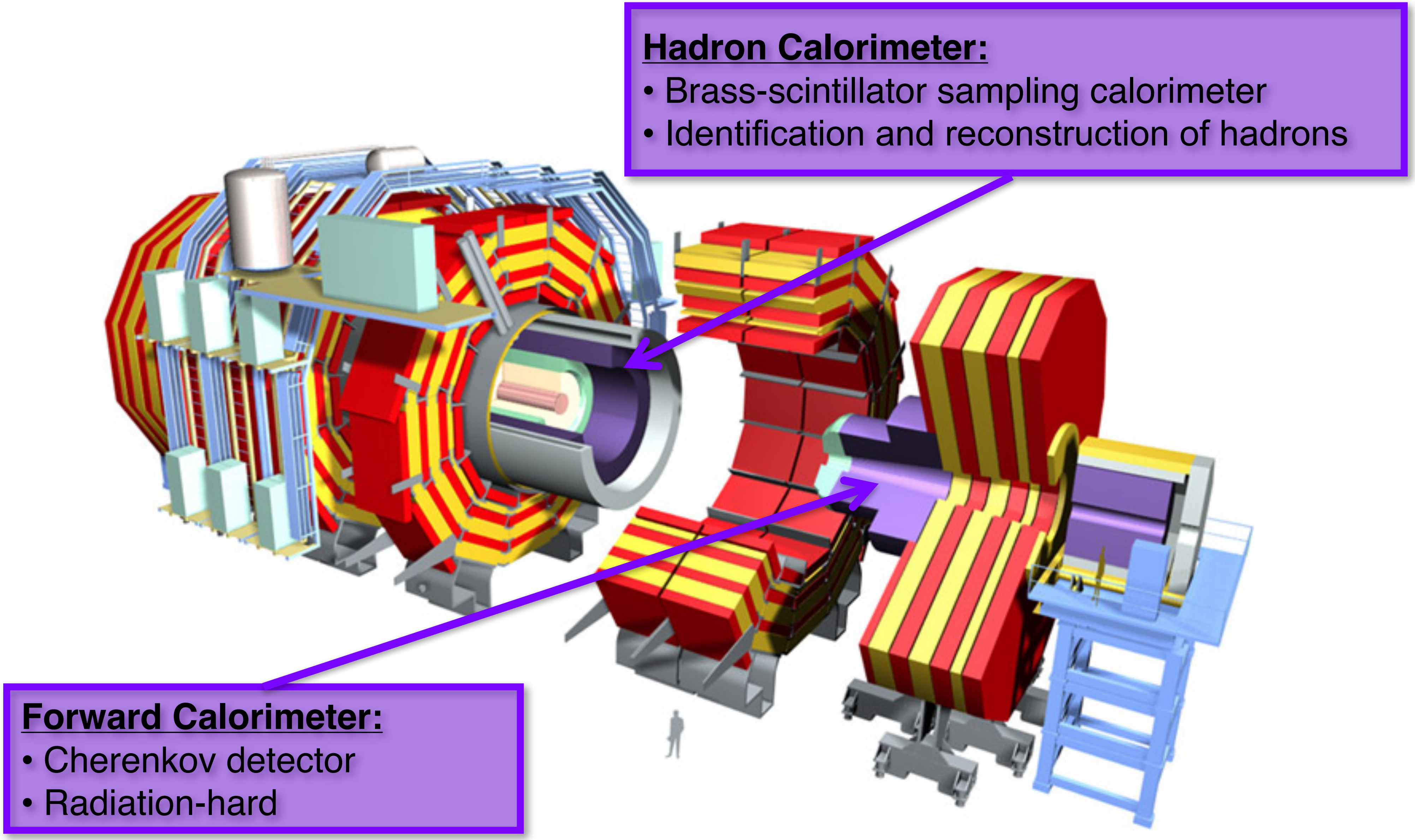


The Compact Muon Solenoid (CMS) detector

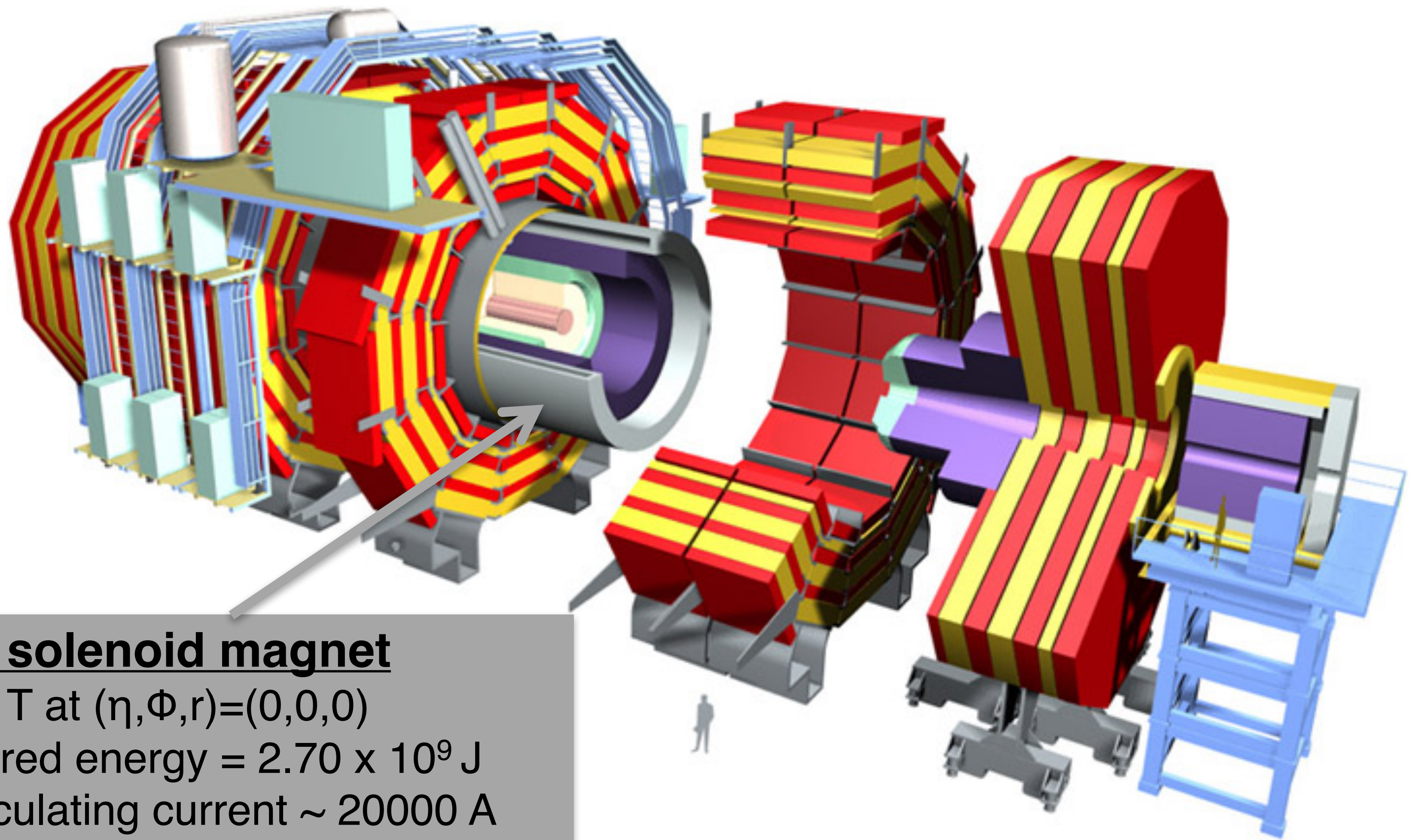
Electromagnetic Calorimeter:
- $PbWO_4$ scintillator
- X_0=0.89cm, R_M=21.9mm
- Identification and energy measurement of e/γ



The Compact Muon Solenoid (CMS) detector



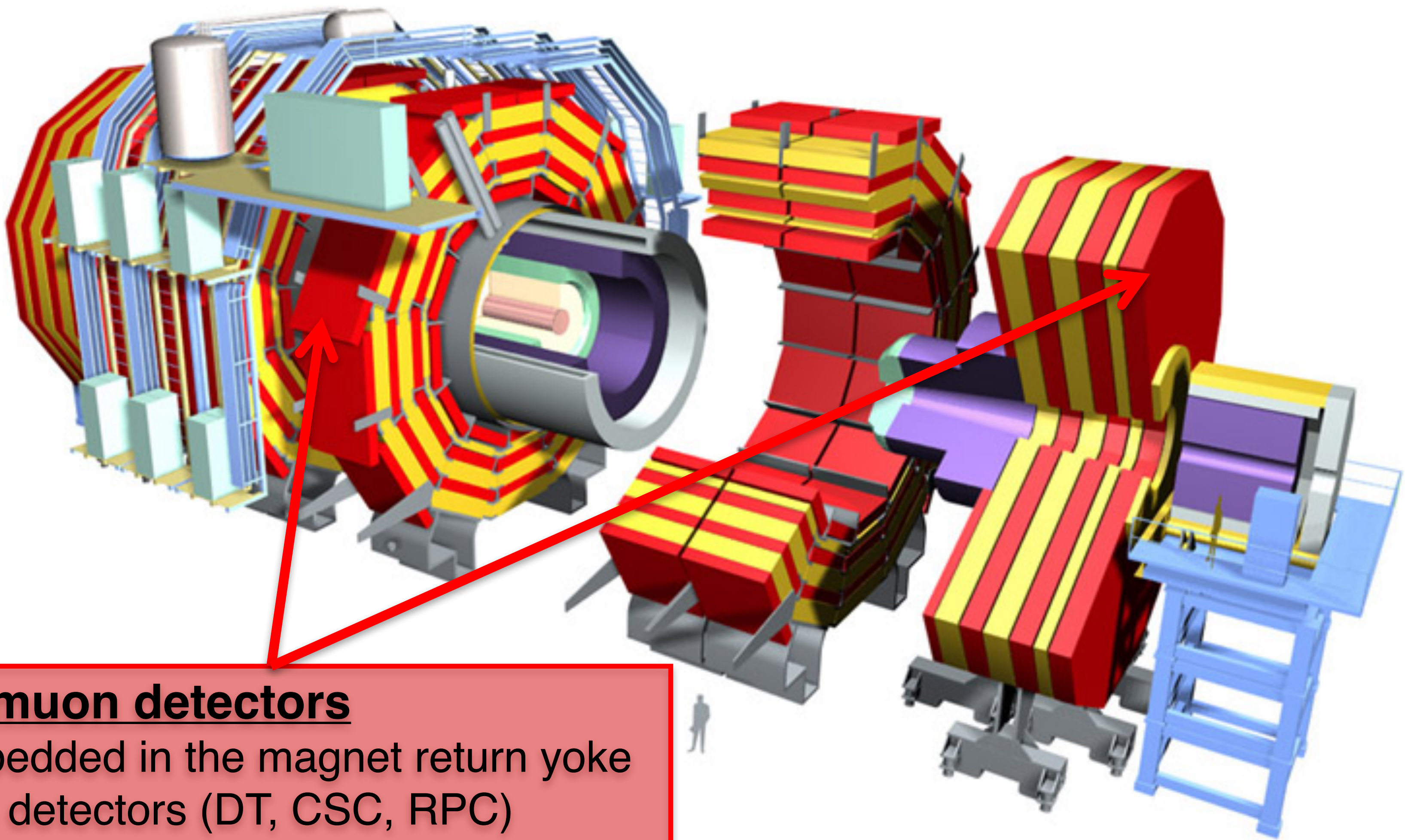
The Compact Muon Solenoid (CMS) detector



The solenoid magnet

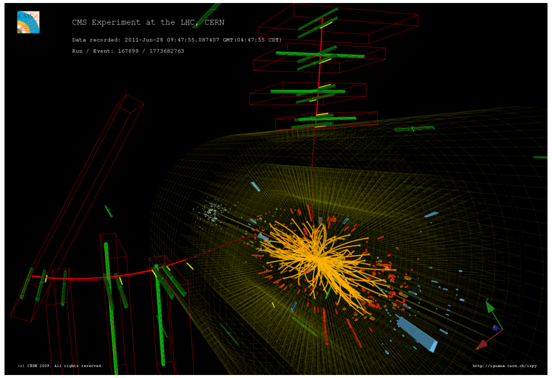
- 3.8 T at $(\eta,\Phi,r)=(0,0,0)$
- Stored energy = 2.70×10^9 J
- Circulating current ~ 20000 A
- bend charged particle trajectory

The Compact Muon Solenoid (CMS) detector



The muon detectors
- Embedded in the magnet return yoke
- Gas detectors (DT, CSC, RPC)
- Muon detection and p_T measurement

CMS trigger system



- CMS → ~70Mpixel
- high resolution high speed photocamera

- 1 MB / event
- LHC bunch frequency: 40 MHz

⇒ 40 TB/s=> ~420 EB/year

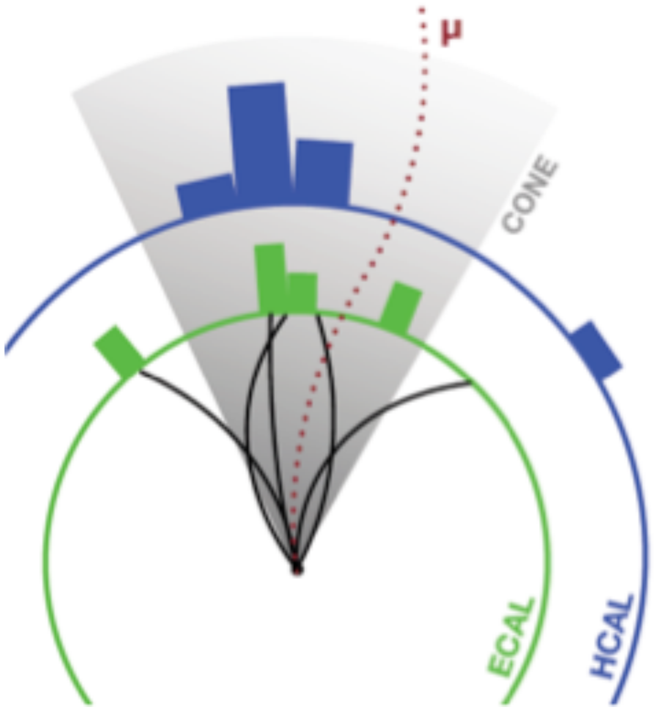
We can’t store all the events. We need to select the interesting picture on the fly!

Trigger system – 2 levels

- **Hardware trigger (L1):**
  - decrease the rate down to O(100)KHz
  - ~100GB/s → ~2000 computers
- **Software trigger (HLT)**
  - further decrease the rate down to O(100)Hz
  - 300MB/s (20Tb/day)

Physics object reconstruction in CMS

- Combines the information from the different CMS sub-detectors to identify all the stable particles in the event: $e^{\pm}$, $\mu^{\pm}$, γ, $h^{\pm}$, h^{0}



Inputs to build

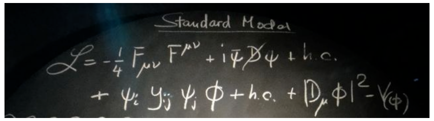
Jets, E_T-miss, τ_h,
Lepton/photon Isolation

Exploiting:

- The **excellent tracking** capability of CMS
- The **very good ECAL granularity** and **resolution**

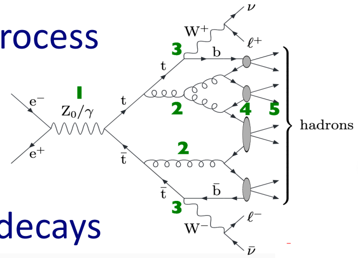
Tipical data-analysis workflow

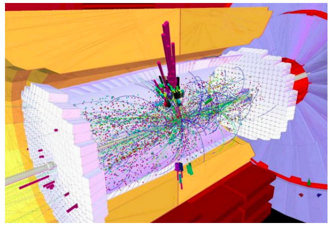
Physics theory



Monte Carlo simulation:

1. Hard-Scattering process
2. Parton-shower
3. Parton-decay
4. Hadronization
5. Unstable hadron decays

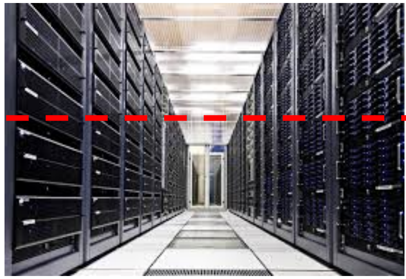




Detector Simulation (GEANT4):

1. Interaction with detector material
2. Digitization process

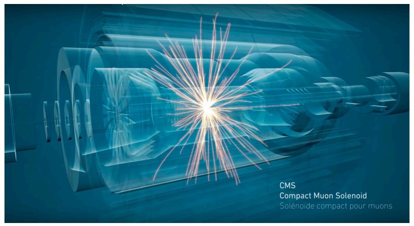
Trigger + Object Reconstruction



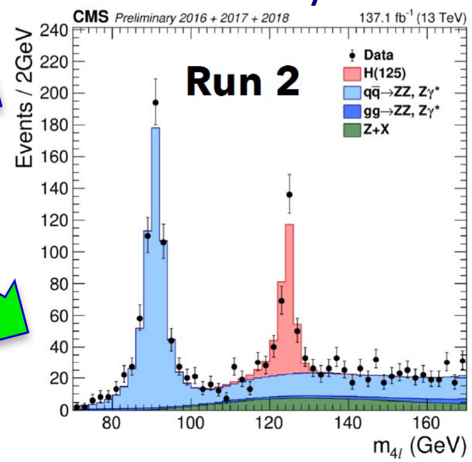
SIMULATION

REAL LIFE

proton-proton collision



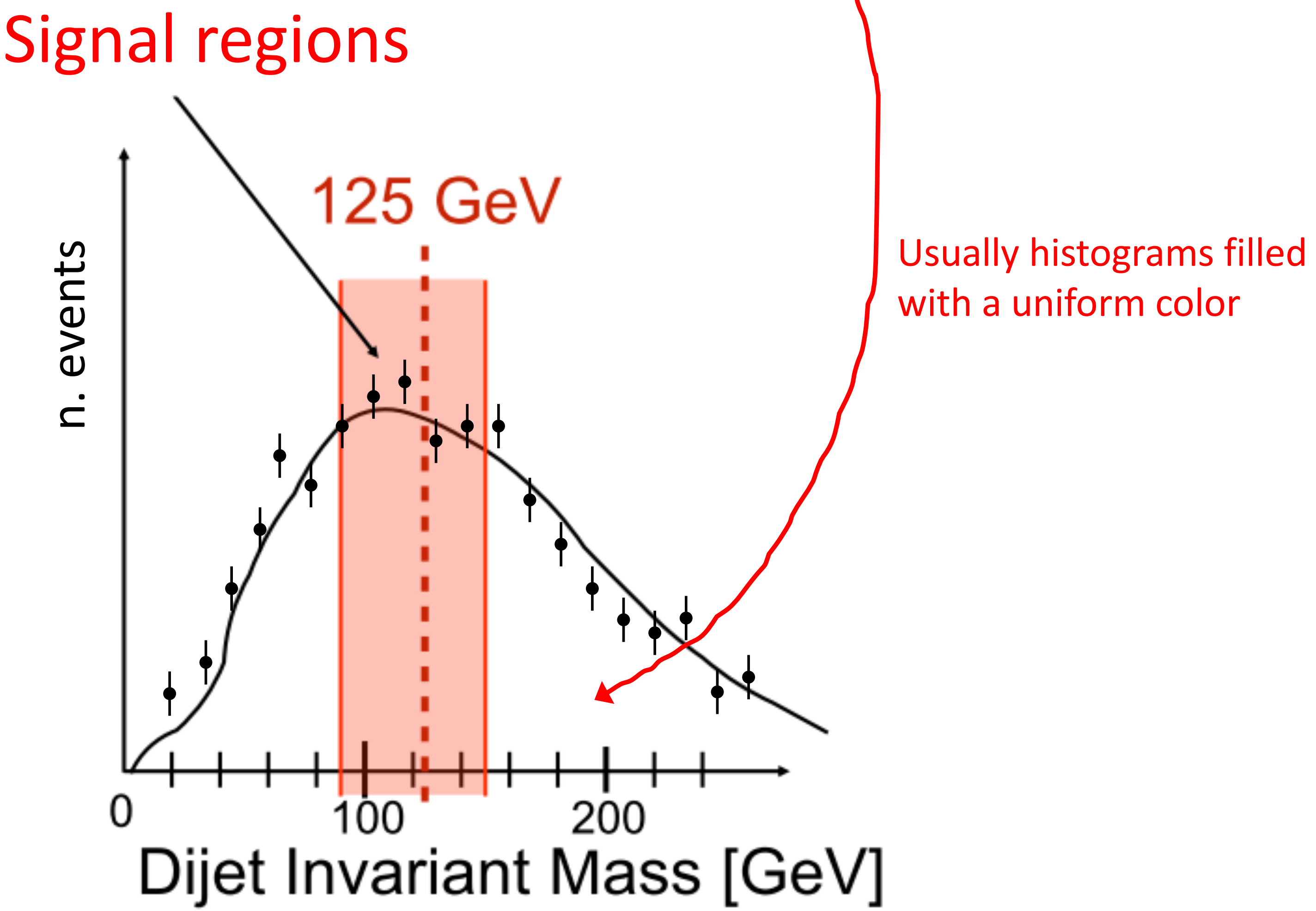
Data Analysis



Typical data-analysis workflow

Template Analysis – Approach used in VHbb and VHcc analyses

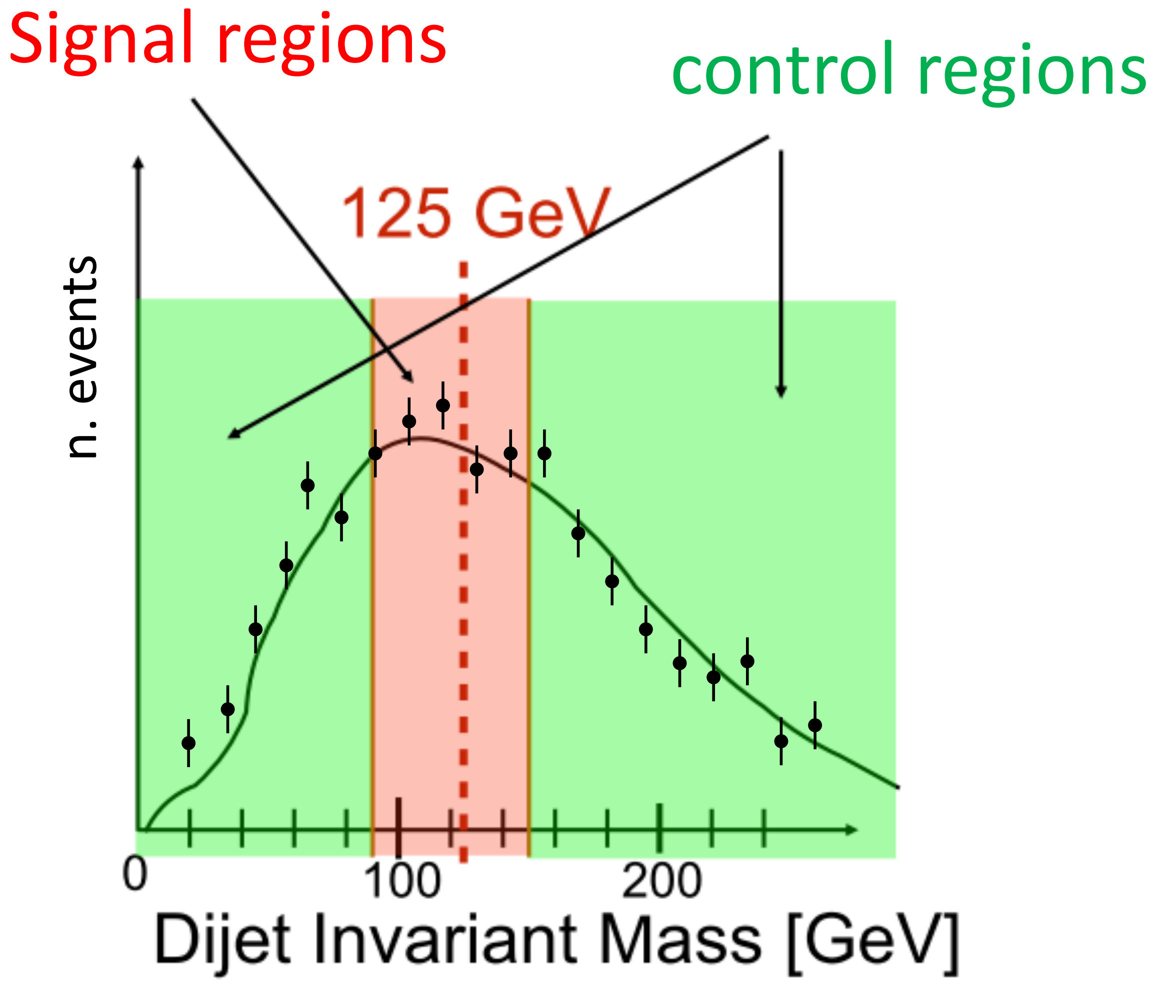
1. Signal and Background samples are simulated with MC



Typical data-analysis workflow

Template Analysis

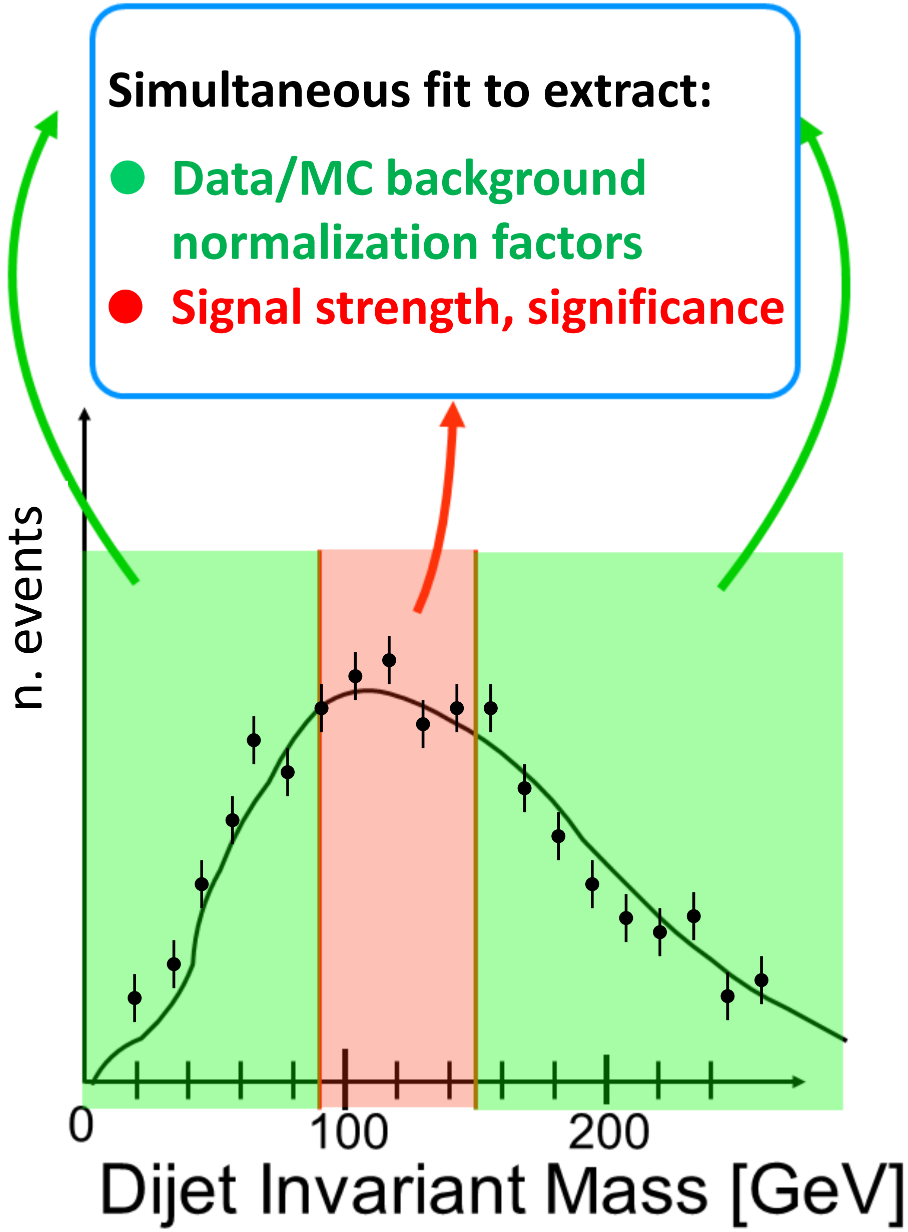
2. Definition of **control region** or ”side-bands” to evaluate the **backgrounds yields**
3. **Fit MC** samples **to data** and **extract** the best-fit values for the **parameters**,



Typical data-analysis workflow

Template Analysis

4. Fit MC samples to data and extract the best-fit values for the parameters



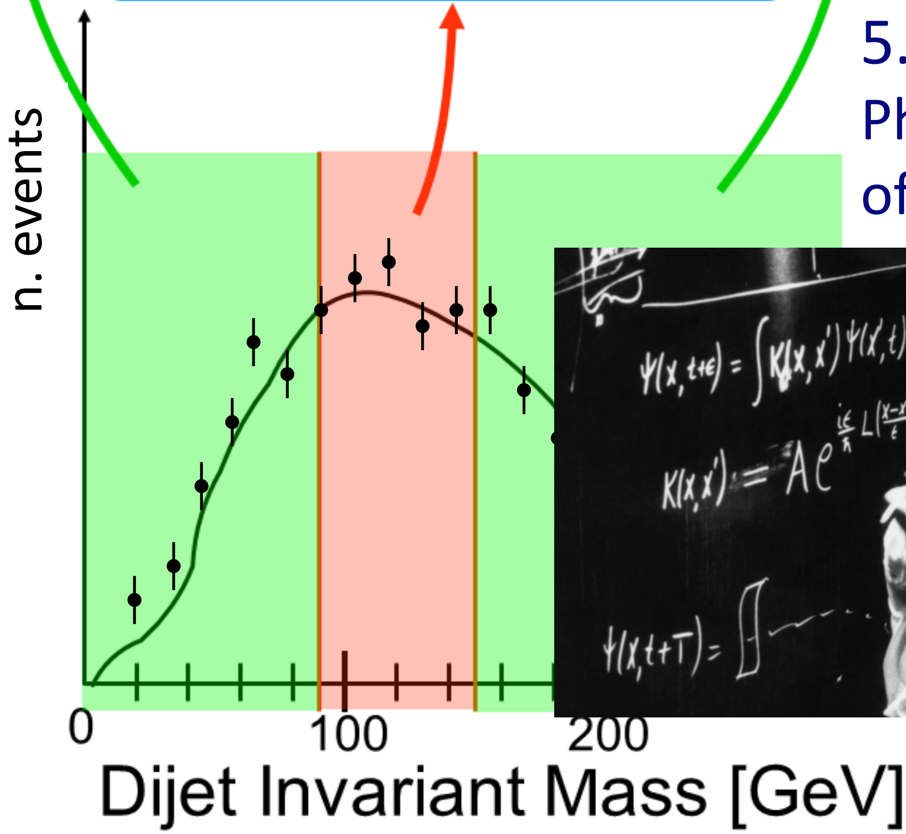
Typical data-analysis workflow

Template Analysis

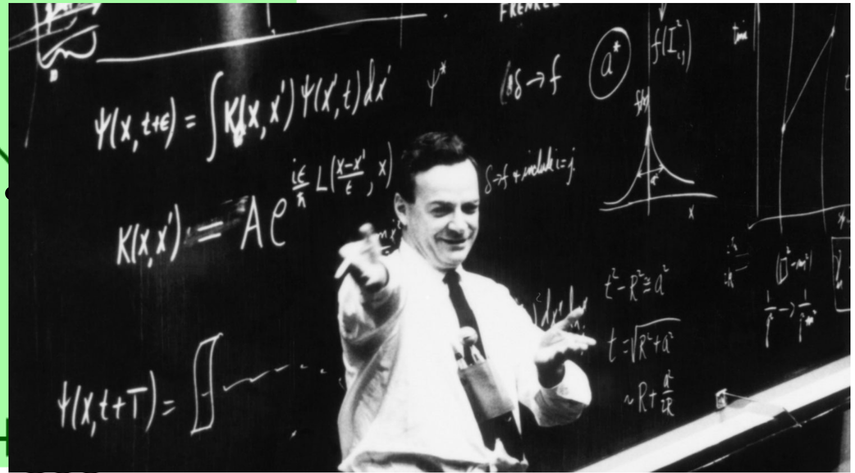
4. Fit MC samples to data and extract the best-fit values for the parameters

Simultaneous fit to extract:

- **Data/MC background normalization factors**
- **Signal strength, significance**



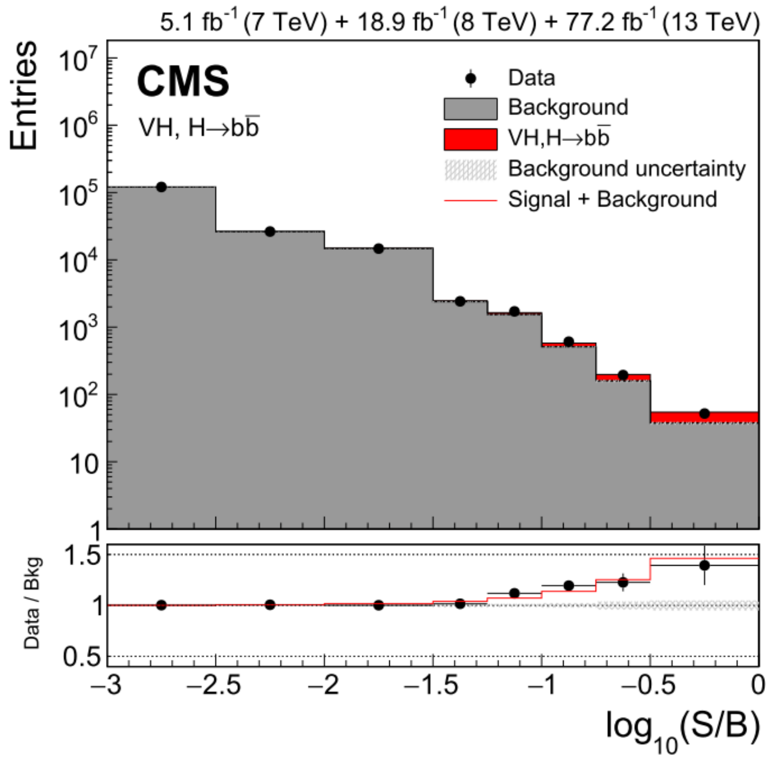
5. <u>The fun part:</u> Physical interpretation of the results

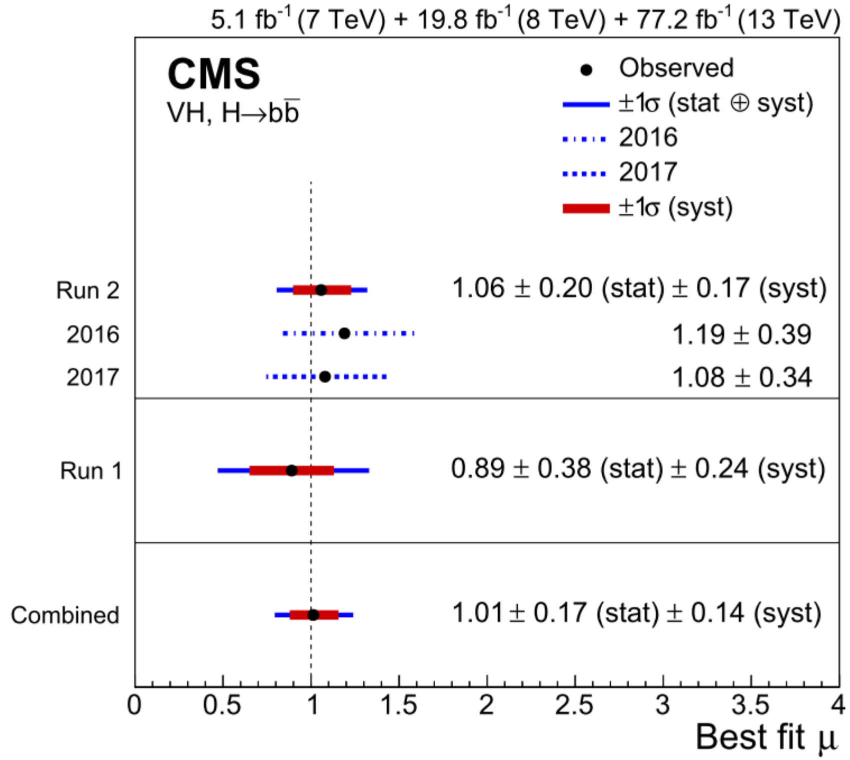


Combination of VH(H→bb) measurements

- **Combination of VH(H→bb) measurement**

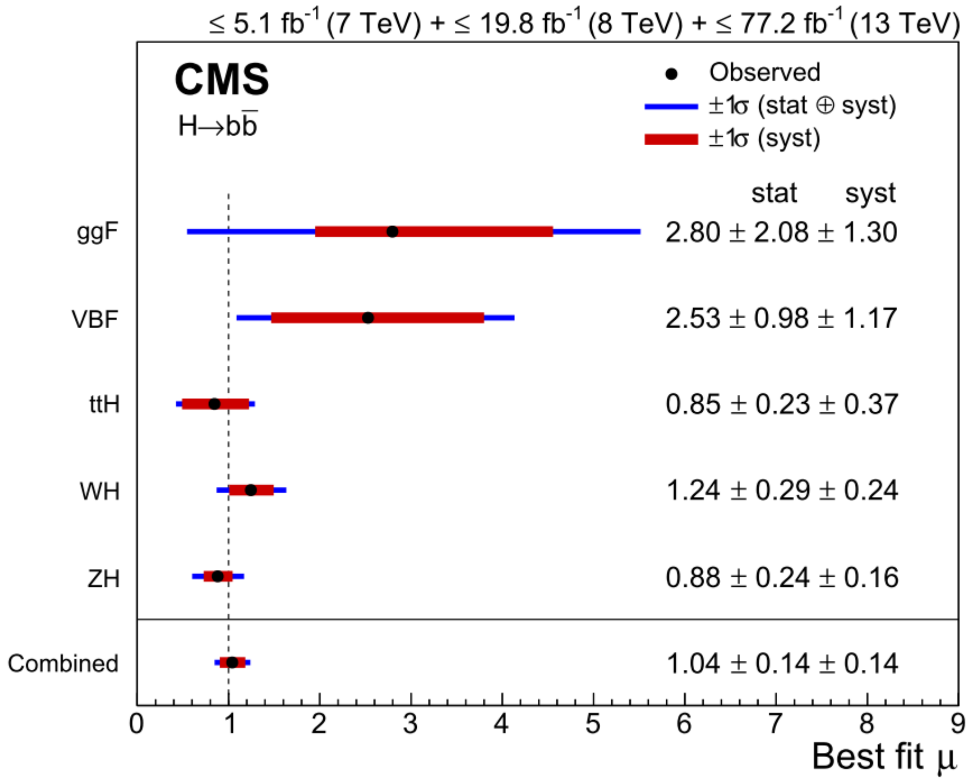
	Significance (σ)		
Data set	Expected	Observed	Signal strength
2017	3.1	3.3	1.08 ± 0.34
Run 2	4.2	4.4	1.06 ± 0.26
Run 1 + Run 2	4.9	4.8	1.01 ± 0.23





Combination of H→bb measurements

- **Combination of all CMS H→bb measurements**
  - VH, boosted ggH, VBF, ttH
  - Most sources of systematic uncertainty are treated as uncorrelated
  - Theory uncertainties are correlated between all processes and data sets



Significance:
5.5σ expected
5.6σ observed

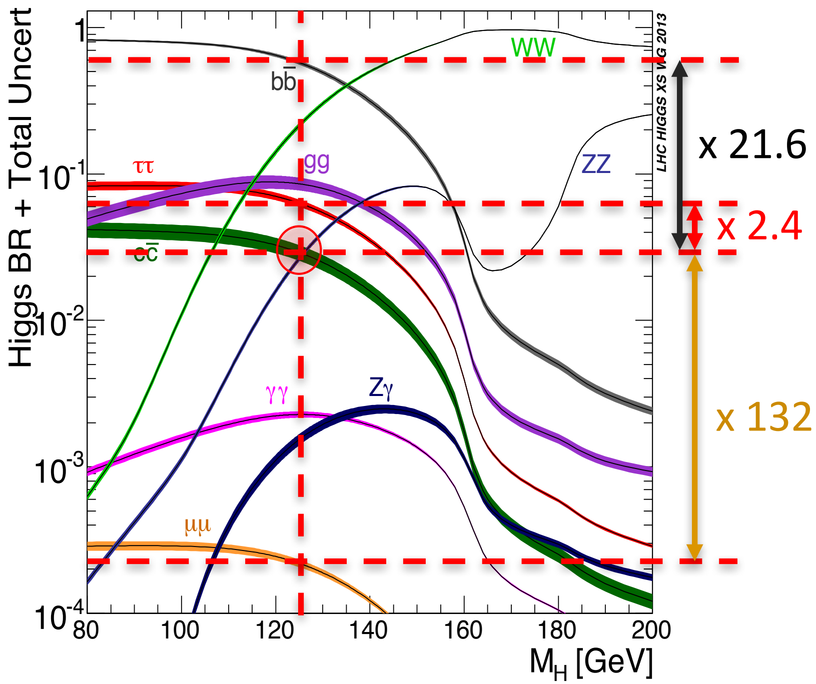
Observation of the H→bb decay by the CMS Collaboration

Measured signal strength:
$\mu = 1.04 \pm 0.20$

Search for the associated production of Higgs boson with W/Z decaying to Charms

Status of H→cc searches at the LHC

- **Objective: Probe Higgs couplings to up-type, 2nd-generation quarks**
  - Higgs-charm coupling can be significantly modified by the presence of BSM



Direct H→cc search:
- ATLAS in Z(LL)H channel [2016]
  UL(μ) < 110 (150) Obs (Exp)

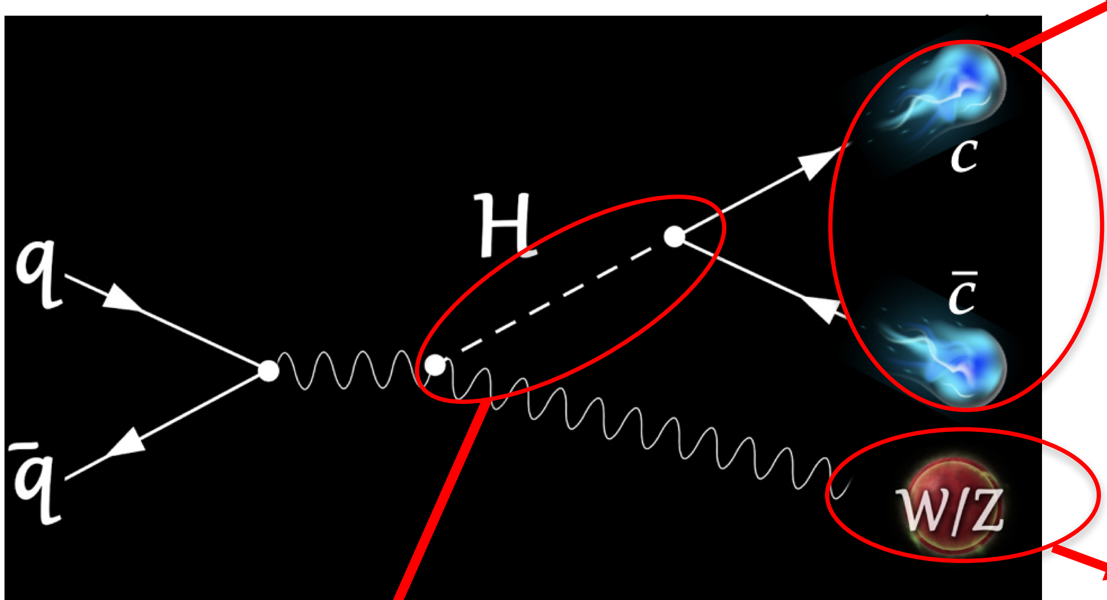
Exclusive decay modes with H→J/ψγ
- ATLAS: 120 (100) x BR obs(exp)
- CMS: 220 (160) x BR obs(exp)

Indirect bounds:
- $\kappa c = yc/yc_{SM}$ from global fit to existing data: $\kappa c < 6.2$ results also from CMS

- H→cc: very challenging to hunt at the LHC
  - Small BR: 2.9×10^{-2} + large backgrounds + H→bb is a background in this search
- c-tagging more challenging than b-tagging

First search for direct H→cc decay in CMS

- **Higgs boson produced in association with W/Z bosons**
  - Low production cross section (~4% of tot x-sec)
  - Cleaner experimental signature



- 2 charm-tagged jets
- 1 fat jet tagging boosted di-charm

➔ **c-tagging plays a crucial role**

- Exploiting leptonic decays of W/Z
- Handle to trigger efficiently events
- W/Z boost to suppress background

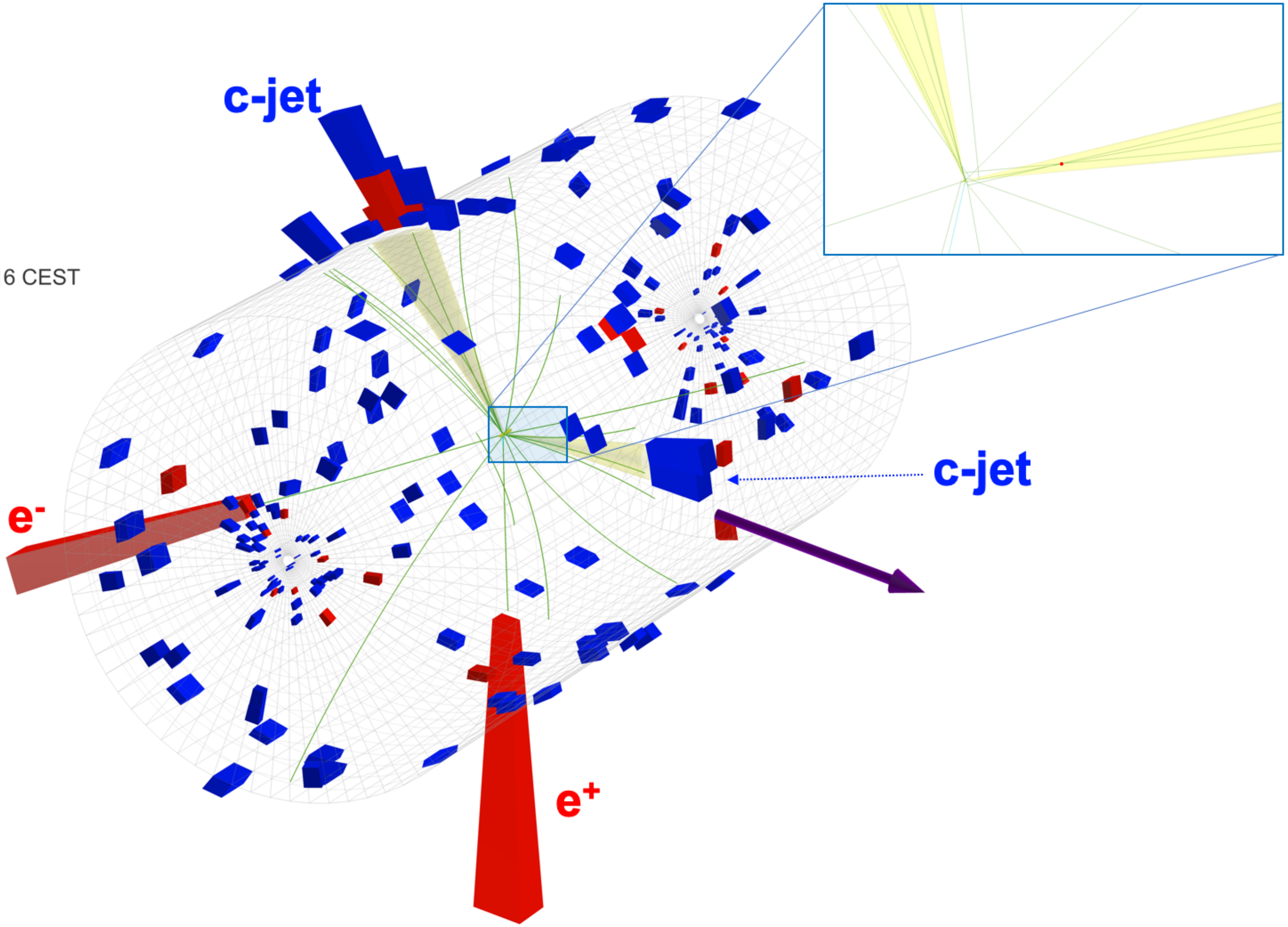
Depending on the pT of the vector, two analysis strategies are deployed

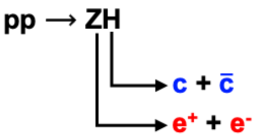
Resolved analysis ➔ regimes of **moderate p_T(H)**, H decays reconstructed in 2 AK4 jets

Boosted analysis ➔ regimes of **high p_T(H)**, H decays reconstructed in 1 AK15 jets

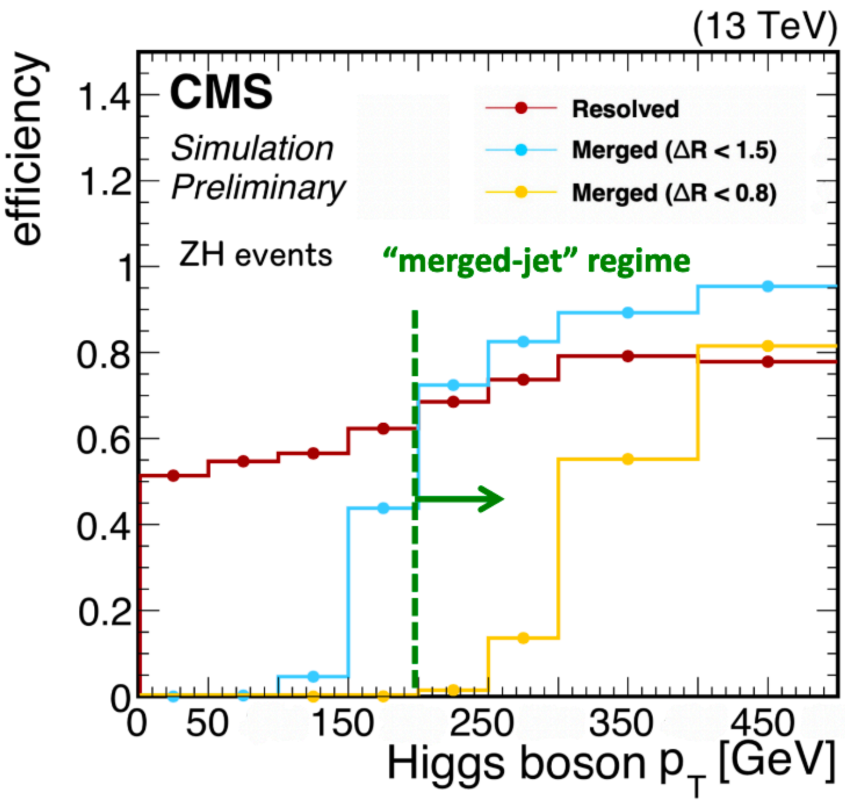
VH(H→cc) candidate - Event Display

CMS Experiment at LHC, CERN
Data recorded: Tue May 31 11:26:24 2016 CEST
Run/Event: 274250 / 1058807020
Lumi section: 543
Orbit/Crossing: 142305803 / 593





VH(H→cc) General Analysis Strategy



- **Resolved-jet topology**
  - Higgs decay products resolved in two AK4 (R=0.4) jets (di-jet)
  - Probe larger fraction of the available signal cross-section (95% of events have $p_T(V)$<200 GeV)
- **Merged-jet topology**
  - A single AK15 (R=1.5) jet to reconstruct the H→cc decay
  - Allows to better exploit the correlations between the two charms

Channel	Resolved-jet	Merged-jet
$Z(\nu\nu)H(cc)$: 0L	$p_T(Z) > 170$ GeV	$p_T(V) > 200$ GeV
$W(\ell\nu)H(cc)$: 1L	$p_T(W) > 100$ GeV	
$Z(\ell\ell)H(cc)$: 2L	$p_T(Z) > 50$ GeV	

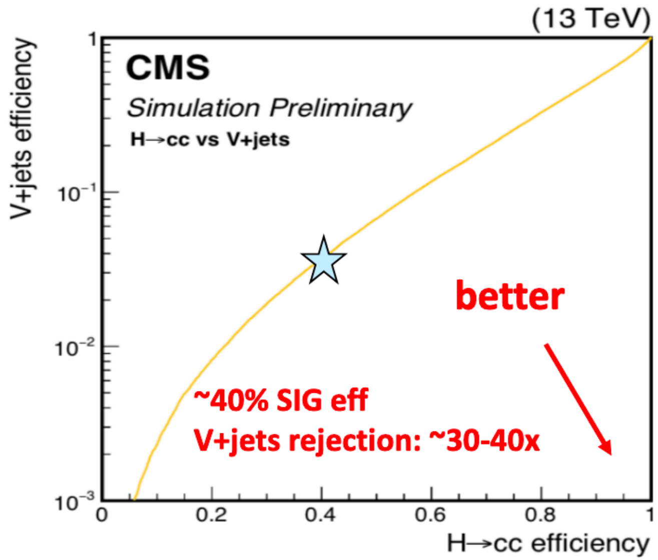
Final results: combination of the two topologies to maximise the sensitivity

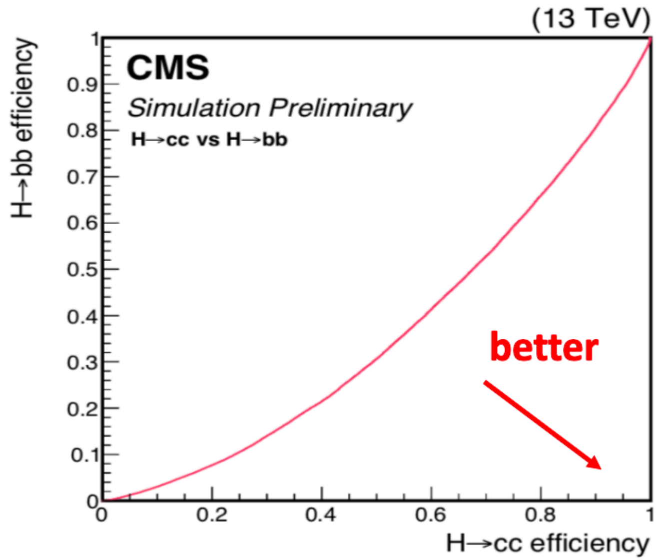
Heavy flavour tagger for AK15: DeepAK15

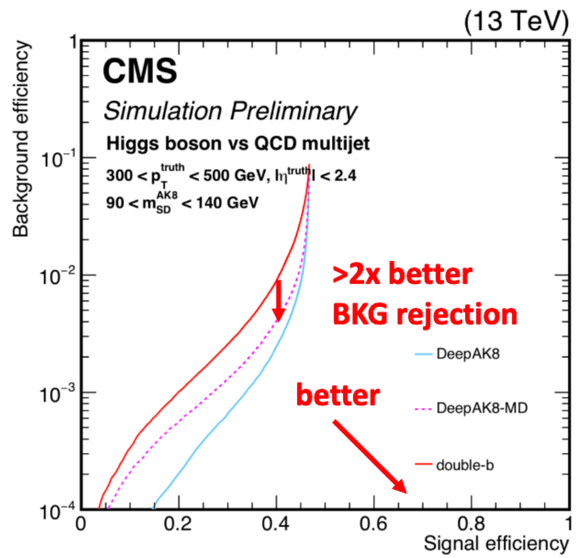
- **DeepAK15 tagger – cornerstone of the boosted VHcc analysis**
  - Reconstruction of moderately to largely boosted Higgs
  - DeepAK15: good compromise between signal purity and acceptance p_T>200 GeV
- **Boosted jet tagger "DeepAK8" adapted on AK15 jets**
  - DNN multiclassifier for top, W, Z, Higgs, and QCD jets
  - Mass decorrelation techniques to mitigate mass sculpting
  - Validation in data using proxy jets from g→cc

More information
→ Huilin talk

CMS-DP-2017-049
NIPS 2017 paper,
CMS-JME-18-002







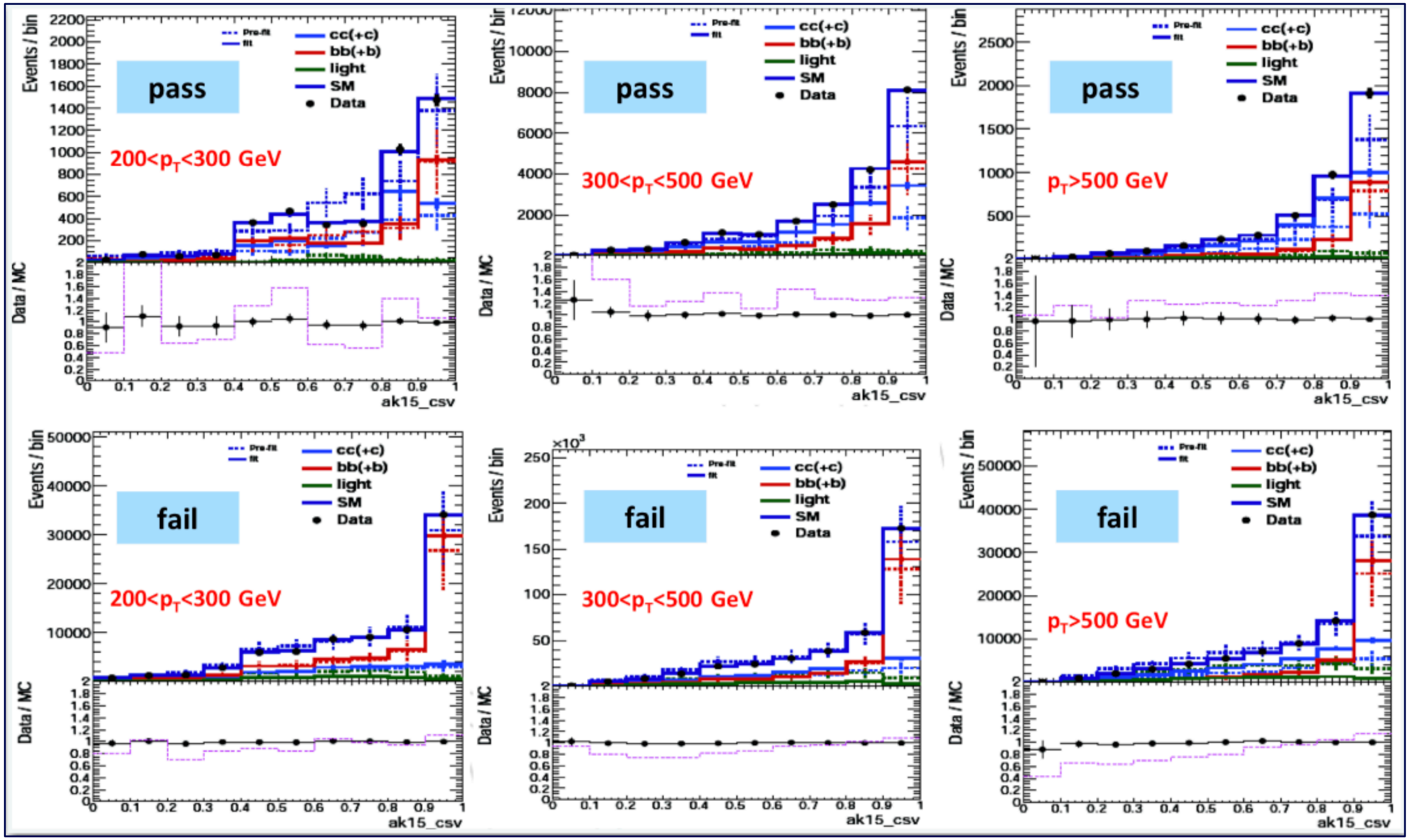
$$\frac{\text{score}(Z \to c\bar{c}) + \text{score}(H \to c\bar{c})}{\text{score}(Z \to c\bar{c}) + \text{score}(H \to c\bar{c}) + \text{score}(\text{QCD})}$$

Significant gain in performance
[Even larger @high p_T]

Heavy flavour tagger for AK15: DeepAK15

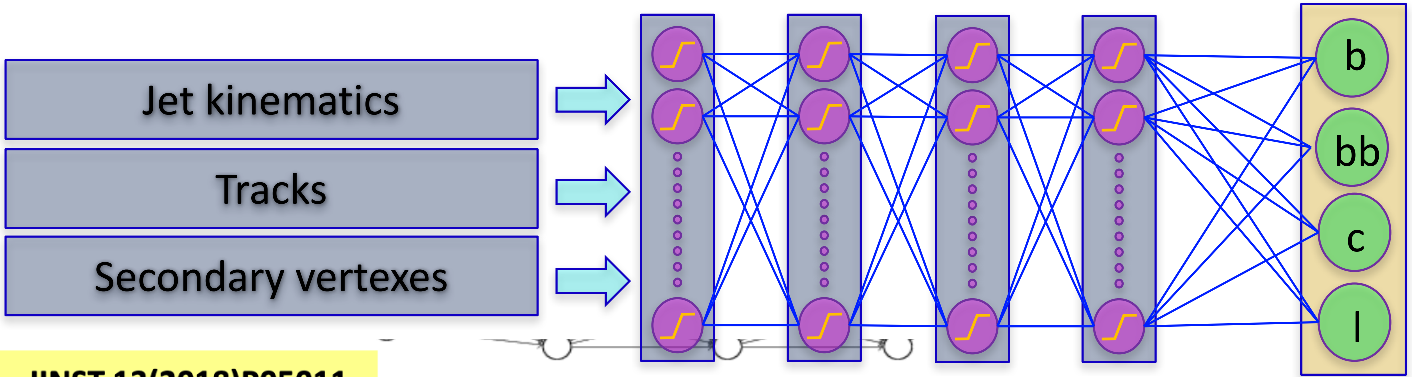
- **Scale factor derivation**
  - Template fit of the tagger for the 3 chosen WP
  - Loose selection + constraint on SV

	p_T range in GeV:		
	200–300	300–500	>500
LP WP	$1.12^{+0.17}_{-0.13}$	$1.18^{+0.07}_{-0.06}$	$1.11^{+0.07}_{-0.06}$
MP WP	$1.33^{+0.40}_{-0.20}$	$1.01^{+0.06}_{-0.06}$	$0.90^{+0.06}_{-0.06}$
HP WP	$1.40^{+0.27}_{-0.19}$	$0.85^{+0.08}_{-0.07}$	$0.95^{+0.08}_{-0.07}$

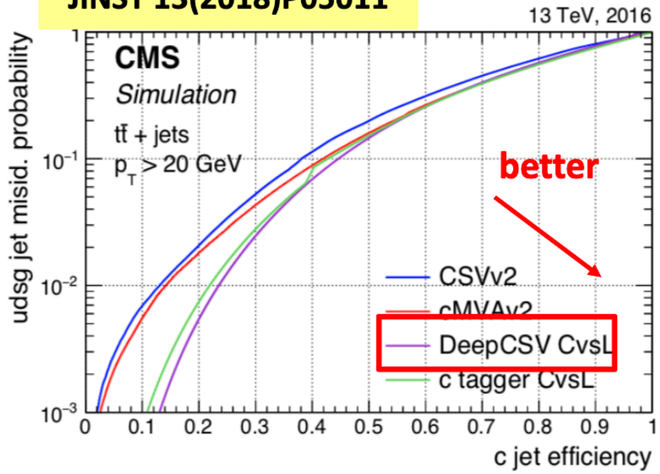


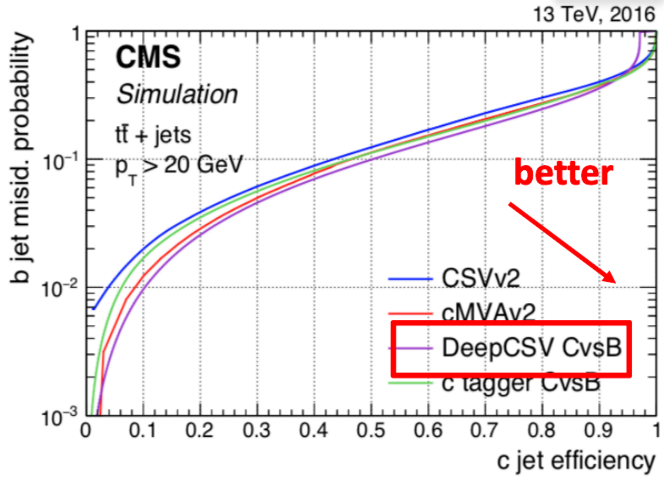
Heavy flavour tagger for AK4: DeepCSV

- **Tagging c-jets is challenging → intermediate properties vs light- and b-jets**
- **<u>DeepCSV: DNN architecture</u>**
  - Input variables go through 4 fully connected layers, each layer has 100 nodes
  - Output layer → softmax activation function → multiclassification
  - Returns 4 scores interpreted as a prob. for a given jet to be originated by a b, bb, c and l



JINST 13(2018)P05011





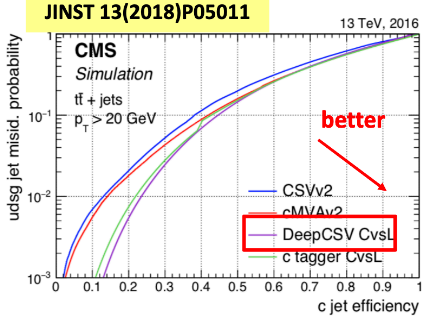
Heavy flavour tagger for AK4: DeepCSV

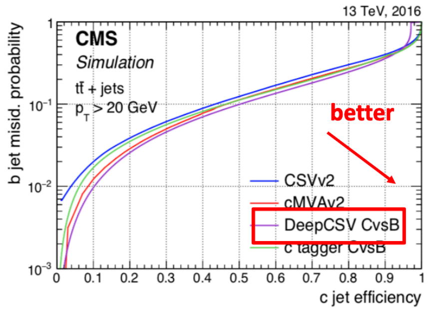
- **Define two discriminants to separate c-jets from light and b-jets**

$$CvsL = \frac{p(c)}{p(c) + p(light)}$$

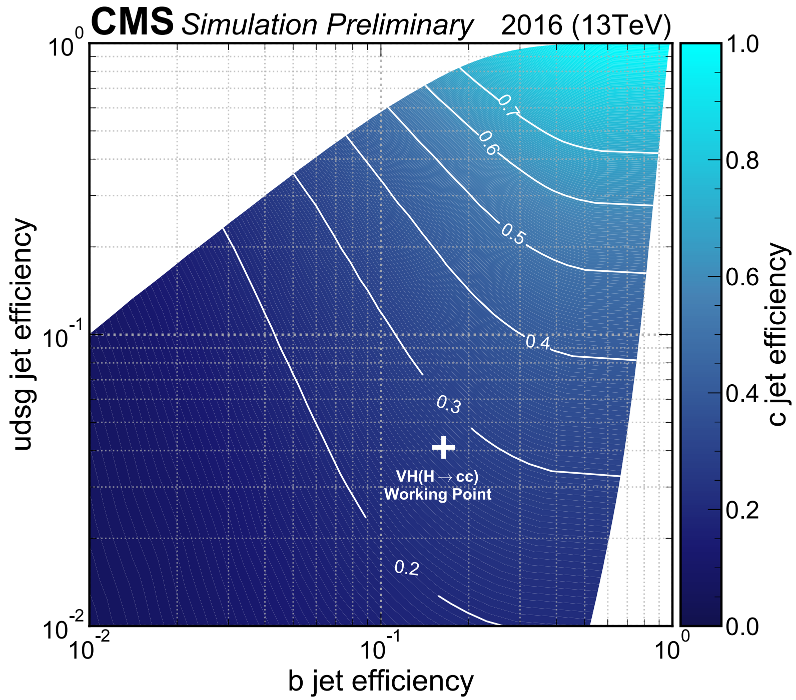
$$CvsB = \frac{p(c)}{p(c) + p(b)}$$

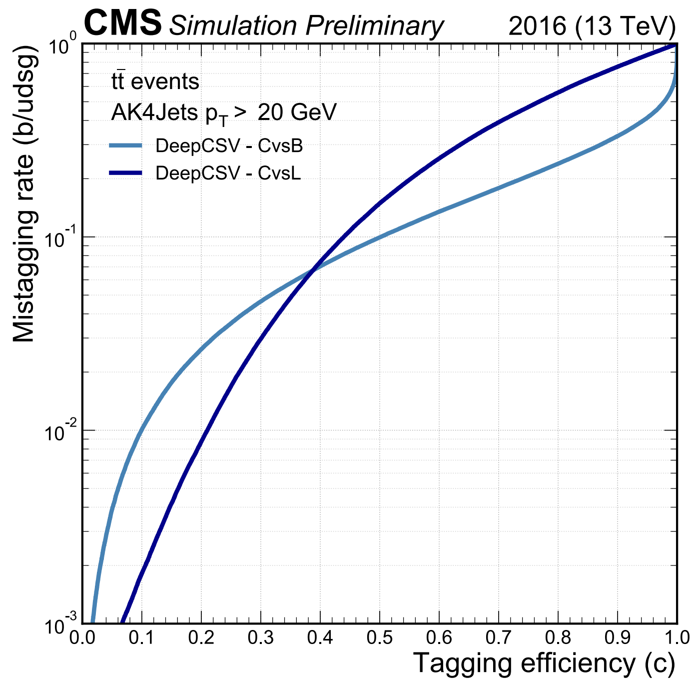
JINST 13(2018)P05011





More information
→ Spandan talk

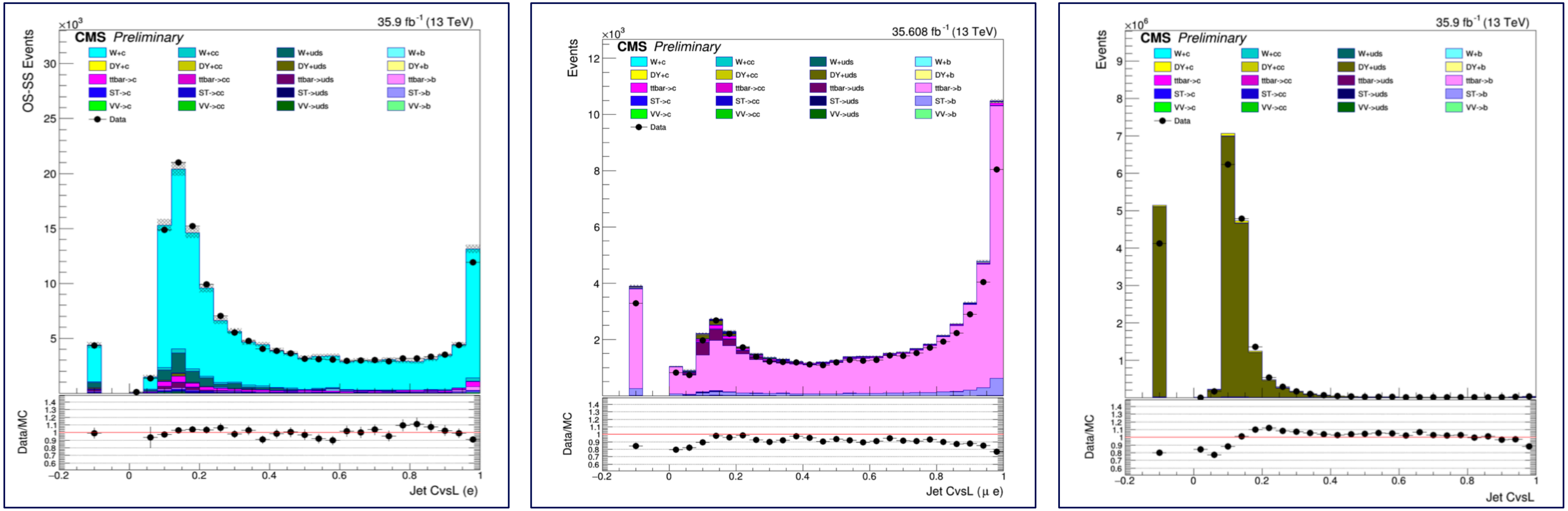




- Taggers working point used in the analysis allow for ~**28%** efficiency for **charm** jet while keeping the rate from **b-jet ~15%** and from **light ~4%**

New SFs measurement technique for AK4

- **Strategy in a nutshell**
  - Iterative fit to the CvsL-CvsB plane in 3 data samples enriched in different jet-flavours
- **Event Selections**
  - **c-jet:** OS-SS W+jets selection, looking to leptonic decay of the W boson + soft muon
  - **b-jet:** Both semileptonic tt+jets (less pure) and dileptonic tt+jets (~5x less statistics)
  - **light-jet:** leading jet in a DY+jets(Z→$\mu\mu$) selections

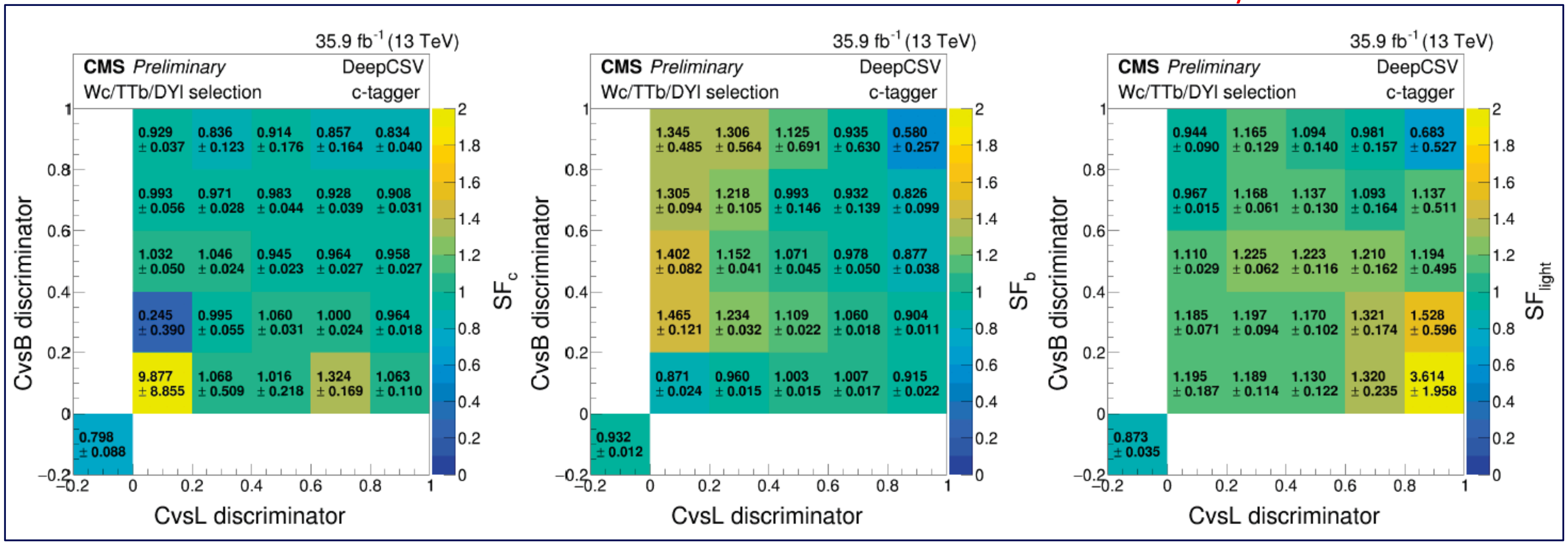


New SFs measurement technique for AK4

- **Strategy in a nutshell**
  - Iterative fit to the CvsL-CvsB plane in 3 data samples enriched in different jet-flavours
- **Event Selections**
  - **c-jet:** OS-SS W+jets selection, looking to leptonic decay of the W boson + soft muon
  - **b-jet:** Both semileptonic tt+jets (less pure) and dileptonic tt+jets (~5x less statistics)
  - **light-jet:** leading jet in a DY+jets(Z→$\mu\mu$) selections
- **Reshaping scale factor central values**

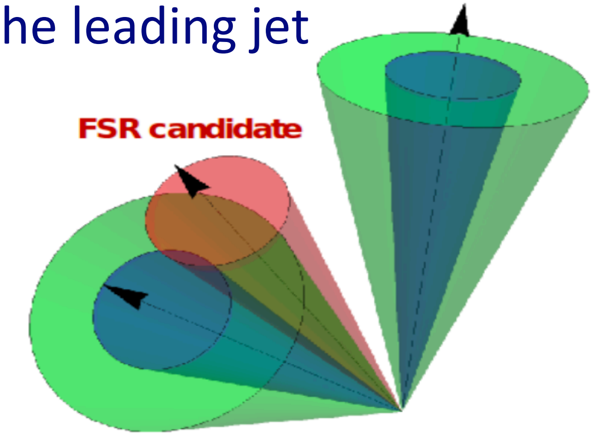
$$w_i = \prod_{i=1}^{jets} sf_i(CvsL, CvsB)$$

Errors account for both statistical and systematics uncertainties



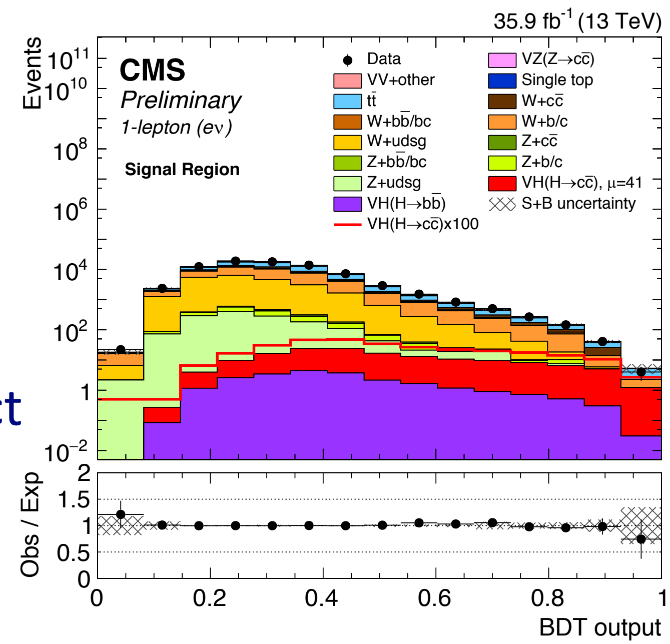
Resolved-jet: Search Strategy

- **Higgs boson reconstruction**
  - Pair of jets with the highest CvsL-score ➔ build Higgs candidate 4-vector
  - Further require: CvsL(max) >0.4 & CvsB(min)>0.2 for the leading jet



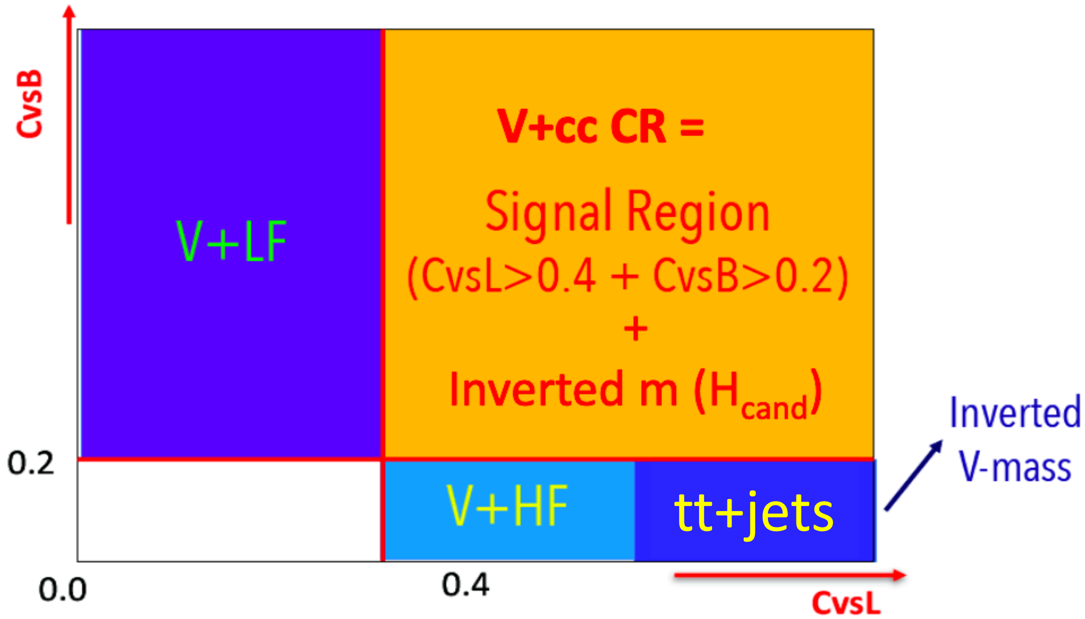
- **Final State Radiation (FSR) recovery**
  - Improve dijet invariant mass resolution by a few %

- **Multivariate analysis for final signal extraction**
  - BDT to further discriminate signal from backgrounds
  - Dedicated training in each channel
  - Input variables: H properties, V boson properties, c-tagging discriminants, event kinematics & object correlations



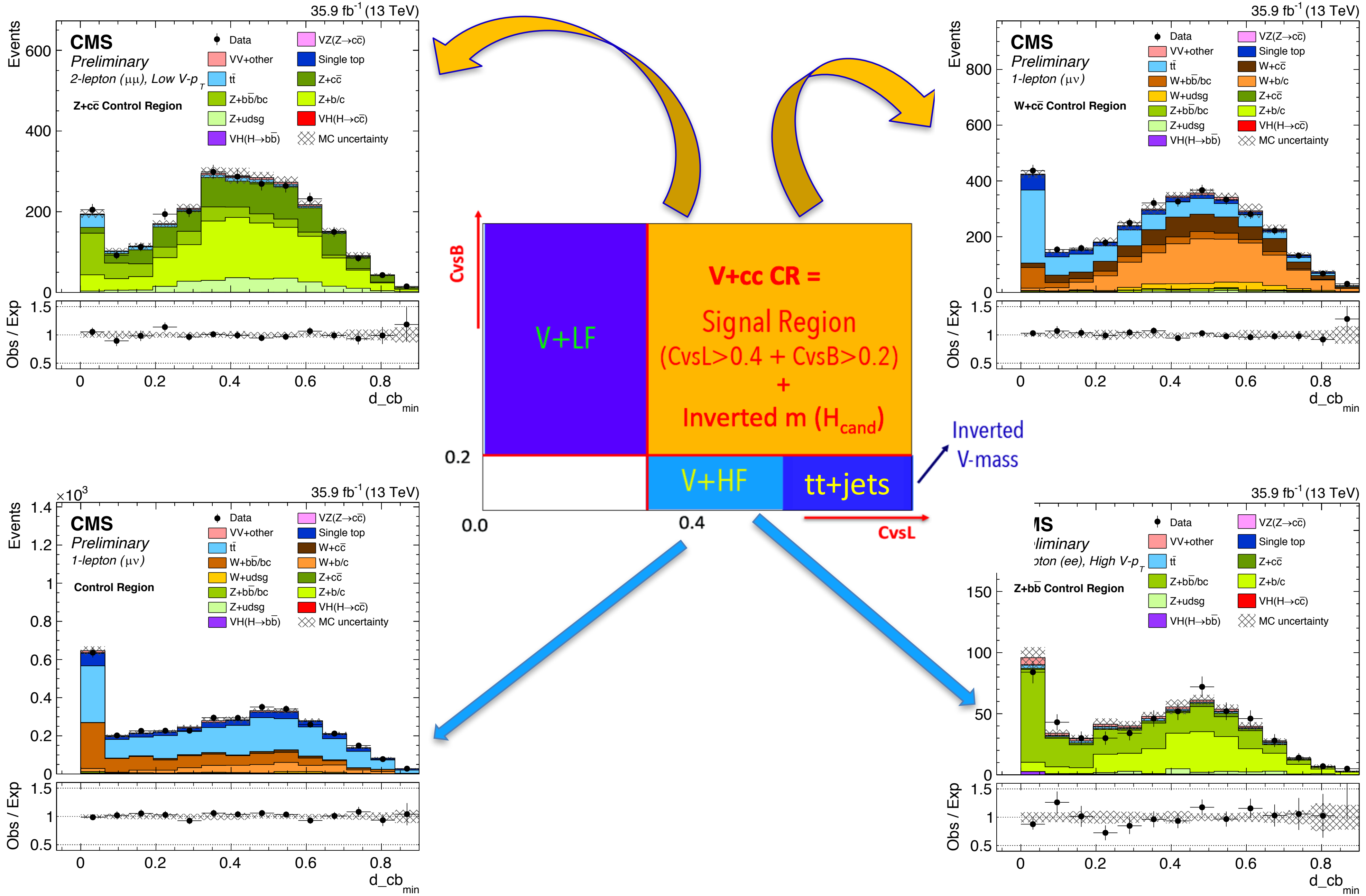
Resolved-jet: Background estimation (I)

- Main backgrounds normalization (V+jets and tt+jets) estimated from data
  - The shapes are taken from simulation (LO samples used for V+jets)
  - 4 control regions are defined per each analysis category and channel
  - V+jets: split based on flavour composition (V+cc, V+bb/bc, V+bl/cl, V+udsg)



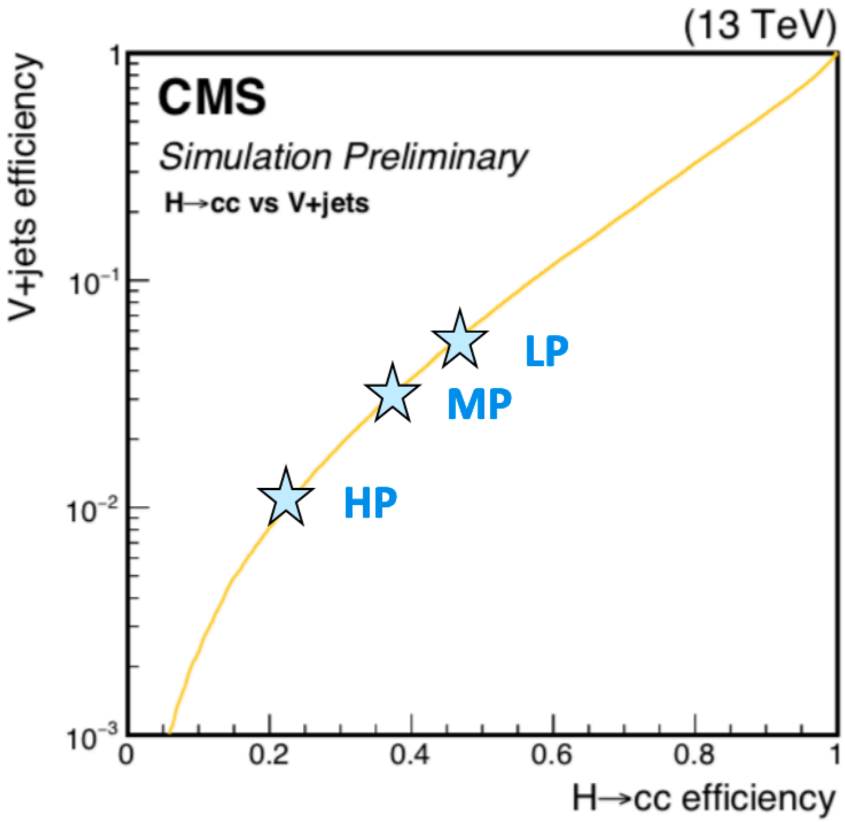
- The control region are fitted simultaneously with the SR
  - The shape of the CvsB/CvsL is fitted in the control region

Resolved-jet: Background estimation (II)



Merged-jet: Search Strategy (I)

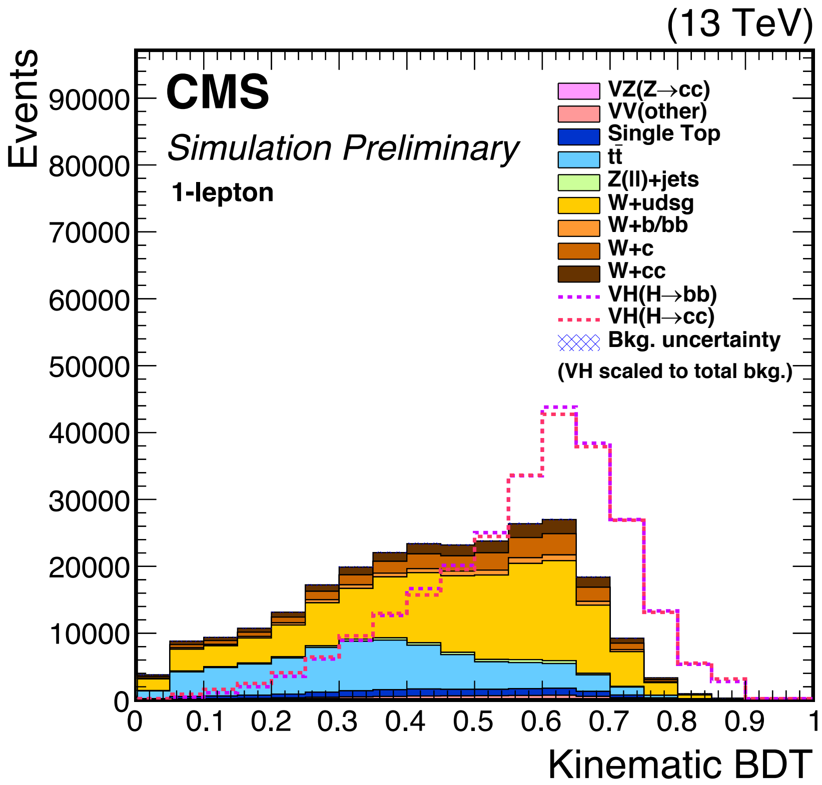
- H reconstruction: highest pT AK15 jet [pT>200 GeV, 50 < m(jet) < 250 GeV]
- Events classified into three mutually exclusive categories based on the three WPs of the cc-discriminant: [High / Medium / Low purity (HP, MP, LP)]

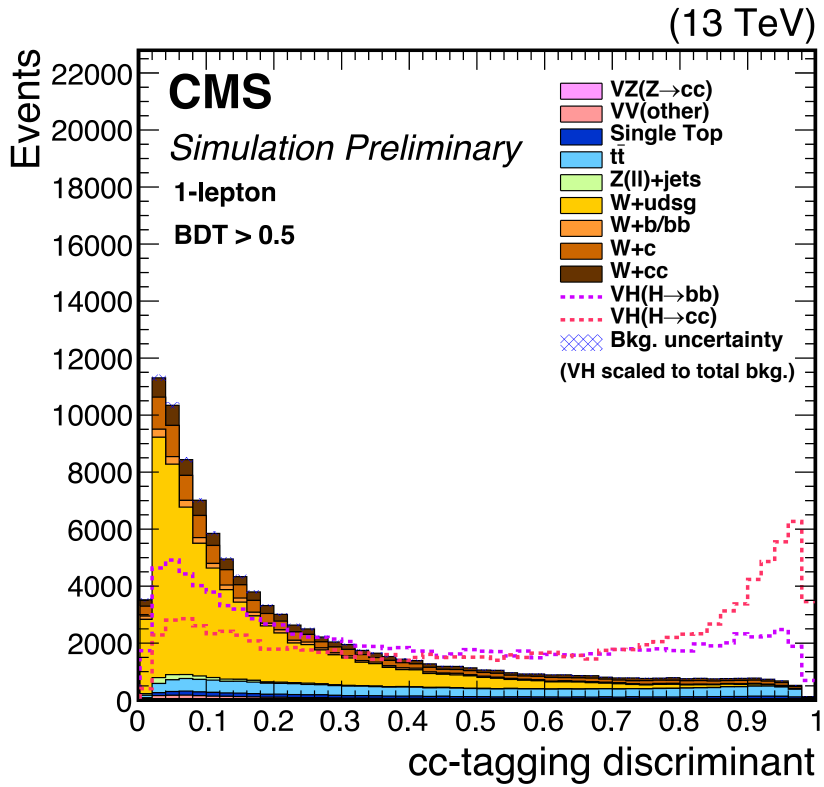


cc-discriminant	>0.72	>0.83	>0.91
$\varepsilon(H\rightarrow cc)$	46%	35%	23%
ε(V+jets)	5%	2.5%	1%
$\varepsilon(H\rightarrow bb)$	27%	17%	9%

Merged-jet: Search Strategy (II)

- Event-level separation: BDT to suppress major backgrounds
  - Use only event kinematics, NOT the intrinsic properties (flavour/mass) of H
  - Search region: BDT > 0.5 [same for all channels]



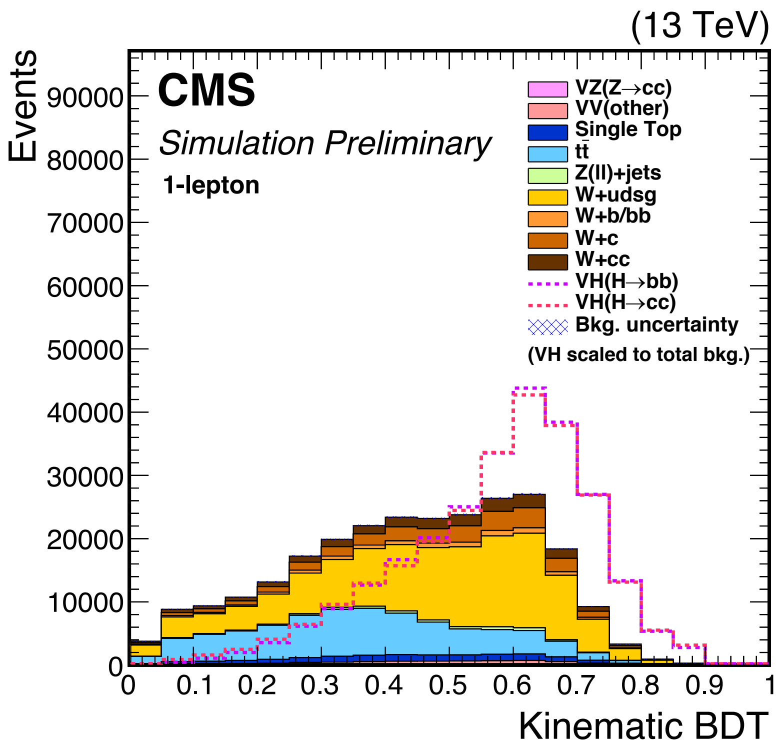


- BDT largely uncorrelated with Higgs candidate mass and cc-discriminant

➔ The variable used in the final fit is the m(H) = m(jet)

Merged-jet: Background estimation

- Major backgrounds (i.e. V+jets and ttbar) estimated from data CRs
  - V+jets CR: low BDT score [i.e. BDT<0.5] + one overall normalization for V+jets (in each of the HP/MP/LP categories)
  - ttbar CR: As the SR but invert N_{AK4}(NB: N_{AK4}<2 requirement applied in SR)
- CRs are designed to have similar flavour composition as SRs
  - same cc-tagging requirement as the corresponding SR



Full analysis validated in two data samples:

→ Low $p_T(V)$

→ Low values of the cc-tagger

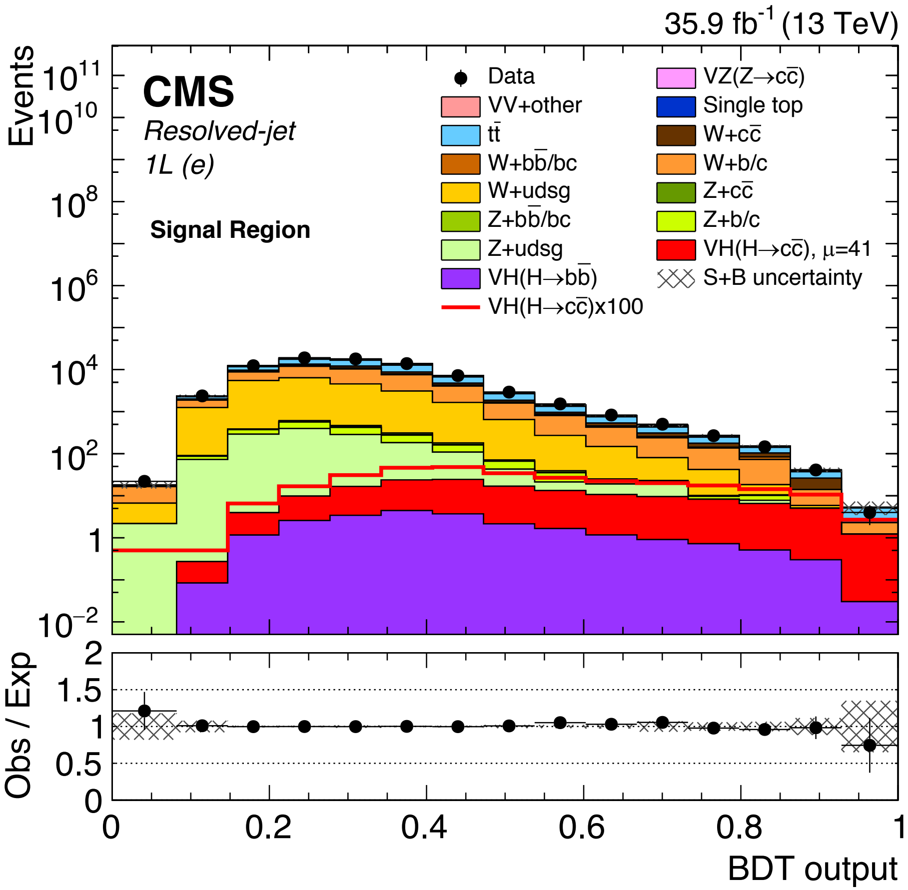
Systematic Uncertainties

Source	Type	0-lepton	1-lepton	2-lepton
Size of simulated samples	shape	✓	✓	✓
Jet energy scale	shape	✓	✓	✓
Jet energy resolution	shape	✓	✓	✓
MET unclustered energy	shape	✓	✓	
c tagging efficiency	shape	✓	✓	✓
Lepton identification efficiency	shape (rate)		✓	✓
Pileup reweighting	shape	✓	✓	✓
top p_T reweighting	shape	✓	✓	✓
$p_T(V)$ reweighting	shape	✓	✓	✓
PDF	shape	✓	✓	✓
Renormalization and factorization scales	shape	✓	✓	✓
VH: $p_T(V)$ NLO EWK correction	shape	✓	✓	✓
Luminosity	rate	2.5%	2.5%	2.5%
MET trigger efficiency	rate	2%		
Lepton trigger efficiency	shape (rate)		✓	✓
Single top cross section	rate	15%	15%	15%
Diboson cross section	rate	10%	10%	10%
VH: cross section (PDF)	rate	✓	✓	✓
VH: cross section (scale)	rate	✓	✓	✓

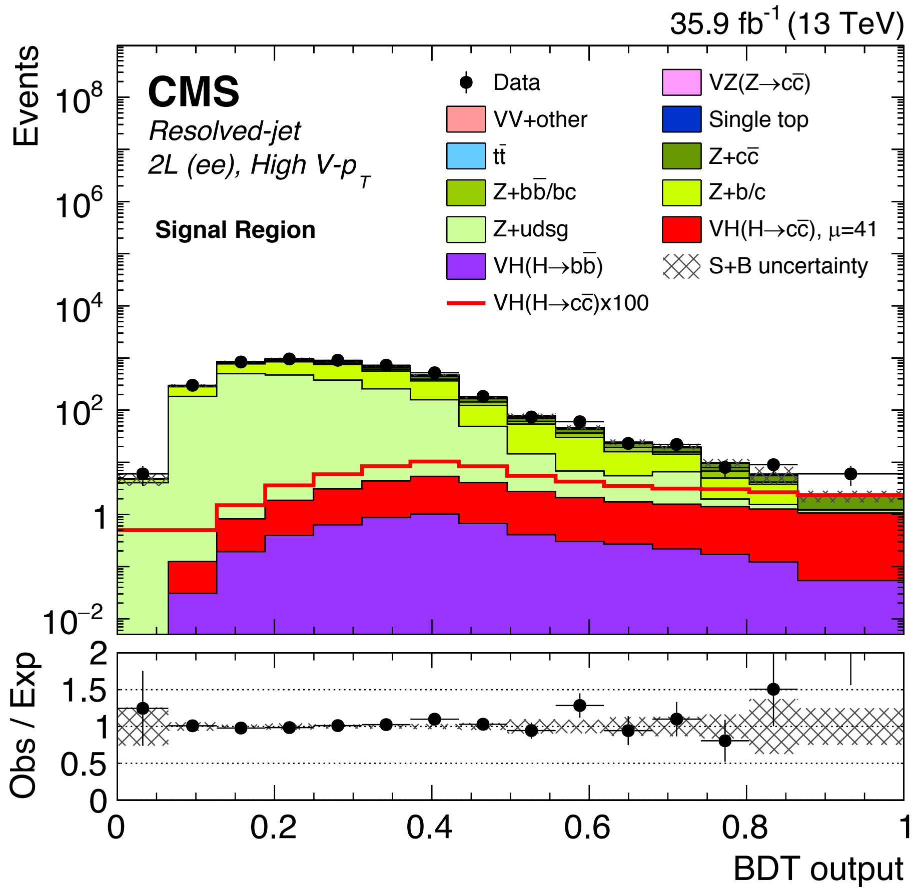
- Dominant sources:
  - statistical uncertainty, c/cc-tagging and MC modelling

Post-fit distribution: resolved-jet

1-lepton – W(eν)

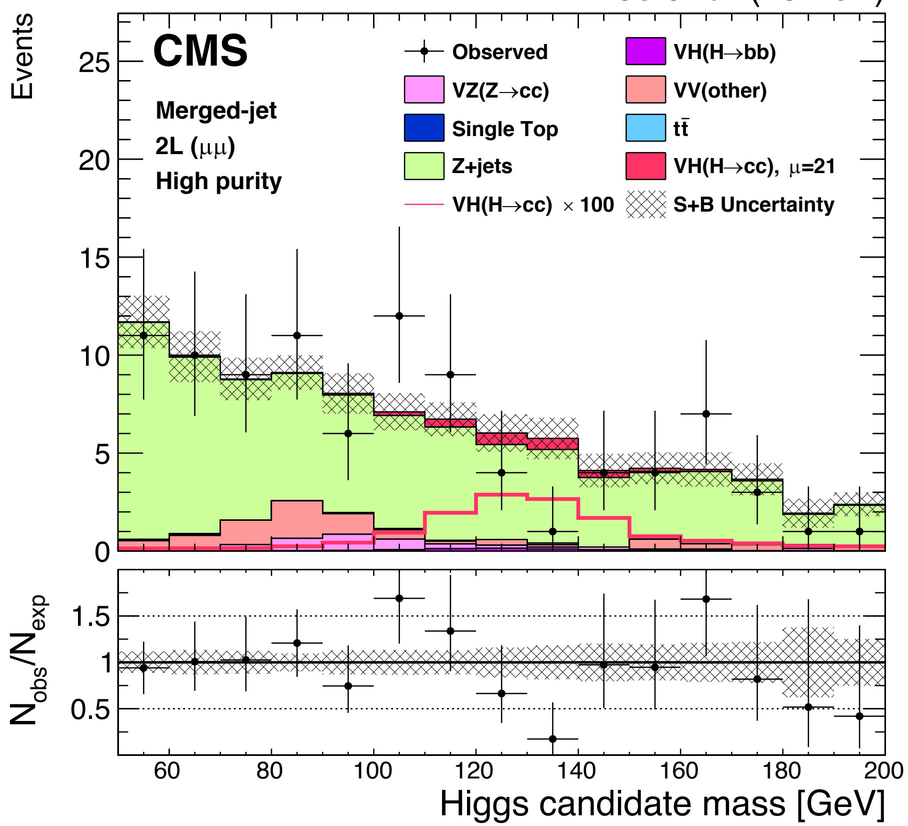


2-leptons High-pT(V) – Z(ee)

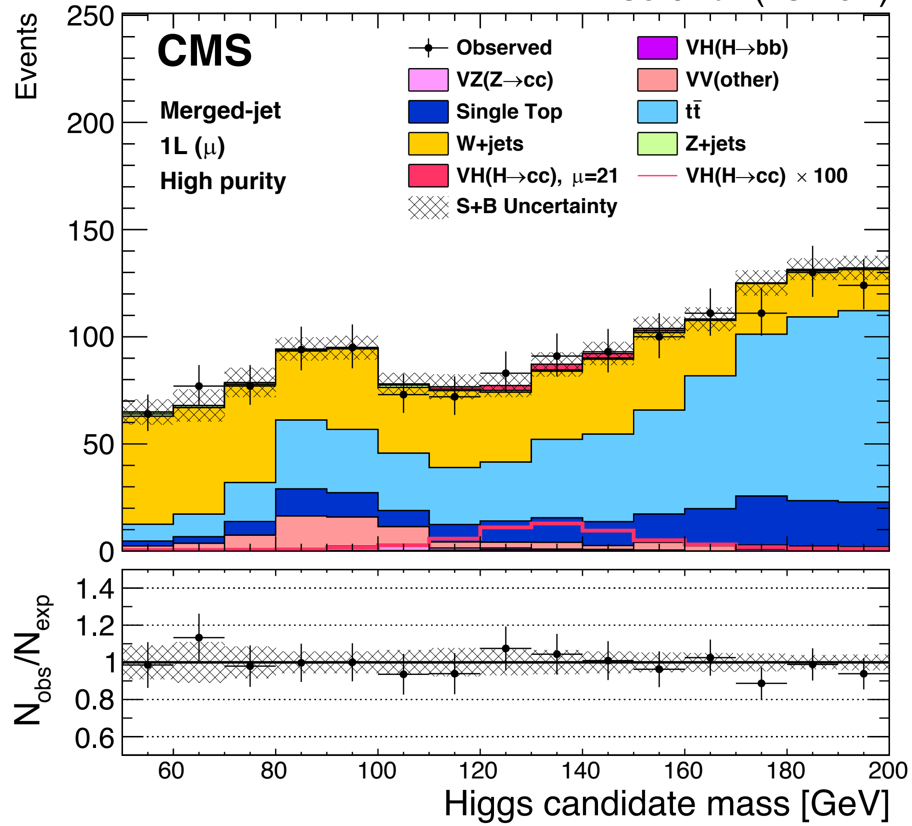


Post-fit distribution: merged-jet

2-leptons– Z($\mu\mu$)



1-leptons– W($\mu\nu$)



Resolved- and Merged-jet results

- **Both the analysis have been validated measuring VZ(Z→cc)**
  - Same analysis as VH(H→cc) but the VZ(Z→cc) has been considered signal

Topology	$\mu_{VZ(Z\to cc)}$	Significance Obs. (Exp.)
Resolved-jet	$1.35^{+0.94}_{-0.95}$	1.5 (1.2)
Merged-jet	$0.69^{+0.89}_{-0.75}$	0.9 (1.3)

- **Results for VH(H→cc):**

95% C.L Exclusion Limit on the signal strength

	Resolved-jet (inclusive)				Merged-jet (inclusive)			
	01	1L	2L	All Ch.	0L	1L	2L	All Ch.
Exp.	84	79	59	**38**	81	88	90	**49**
Obs.	66	120	116	**75**	74	120	76	**71**

μ<75 obs. (38^{+16}_{-11} exp.) μ<71 obs. (49^{+24}_{-15} exp.)

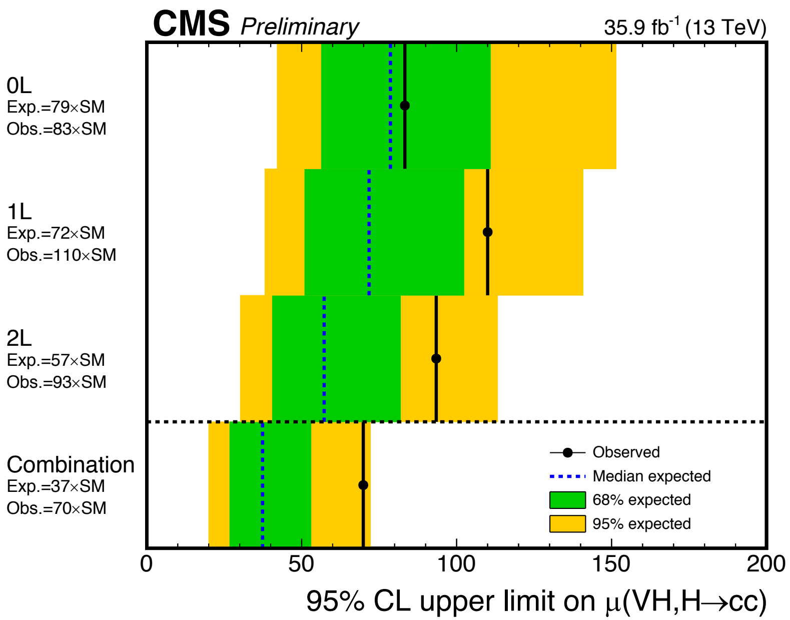
Best-fit signal strength

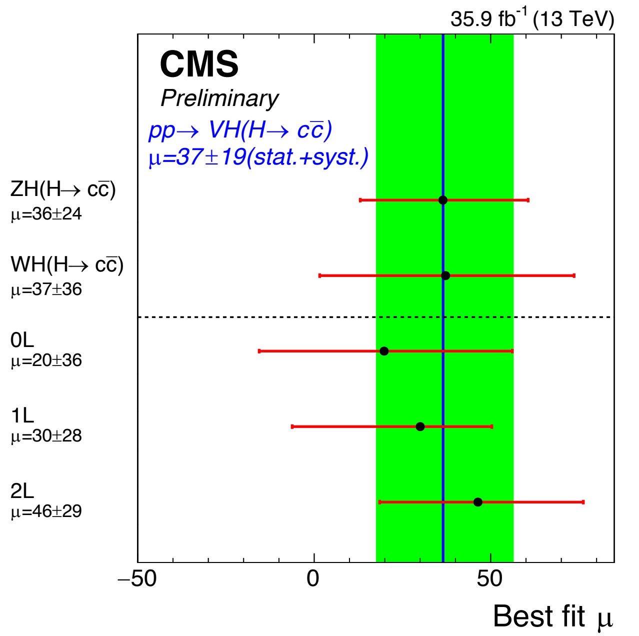
Topology	$\mu_{VH(H\to cc)}$
Resolved-jet	41^{+20}_{-20}
Merged-jet	21^{+26}_{-24}

VH(H→cc) Combination

- Combination: resolved-jet: $p_T(V)$ < 300 GeV / merged-jet: $p_T(V)$ > 300 GeV
  - Systematics: correlated, but: c/cc-tagging efficiency & PDF, μR, μF for V+jets
- Validation with VZ(Z→cc): $\mu_{VZ(Z\to cc)} = 0.55^{+0.86}_{-0.84}$ with 0.7σ obs. (1.3σ exp.)

95% C.L. Exclusion Limits						
	Resolved-jet	Boosted-jet	Combination			
	$p_T(V)$<300 GeV	$p_T(V)$>300 GeV	0L	1L	2L	**All. Ch.**
Exp.	45^{+18}_{-13}	73^{+34}_{-22}	79^{+32}_{-22}	72^{+31}_{-21}	57^{+25}_{-17}	**$37^{+16\,(+35)}_{-11\,(-17)}$**
Obs.	86	75	83	**110**	93	**70**



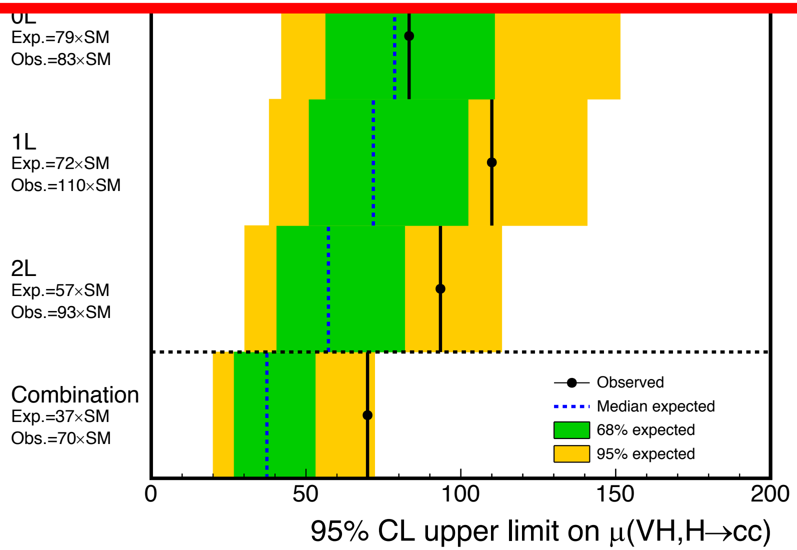


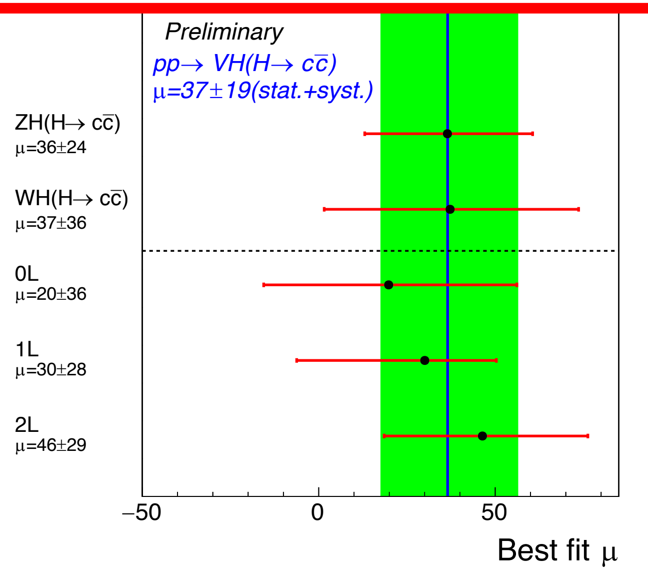
VH(H→cc) Combination

- Combination: resolved-jet: $p_T(V) < 300$ GeV / merged-jet: $p_T(V) > 300$ GeV
  - Systematics: correlated, but: c/cc-tagging efficiency & PDF, μR, μF for V+jets
- Validation with VZ(Z→cc) : $\mu_{VZ(Z\rightarrow cc)} = 0.55^{+0.86}_{-0.84}$ with 0.7σ obs. (1.3σ exp.)

95% C.L. Exclusion Limits						
	Resolved-jet	Boosted-jet	Combination			
	$p_T(V)<300$ GeV	$p_T(V)>300$ GeV	0L	1L	2L	**All. Ch.**
Exp						17)
Obs						

$\sigma(VH) \times \mathcal{B}(H\rightarrow cc) < 4.5$ obs. $(2.4^{+1.0}_{-0.7}$ exp.) pb





Conclusion

- **Direct search for H→cc decay (new in CMS!)**
  - Looking to VH production mode with 2016 only dataset
  - Two strategies: Resolved and Boosted, looking to different $p_T(H)$ regimes
  - Improved results for VH(H→cc): Exp. limits on μ~37
  - First H→cc analysis in CMS [HIG-18-031]
- **Main challenge: tagging charm quarks**
  - Two different approaches in boosted and resolved analysis
  - DeepAK15 for boosted and DeepCSV-based likelihoods for resolved
  - A new method to measure the c-tagger SFs from resolved
- **CMS search for H→$\mu\mu$ decay**
  - Most recent CMS results from 2016 data analysis are shown [HIG-17-019]
  - Results are combined with Run-1, leading to measure μ=1.0±1.0 with an observed (expected) significance of 0.9σ (1.0σ)
  - CMS plans for full Run-2: not only upgrade the dataset but also incorporate as many improvements as possible

Conclusions (I)

- **CMS has achieved a 5.6σ observation of the H→bb decay, with signal strength μ = 1.04 ± 0.20**
  - Combination of several production channels, dominated by VH(H→bb)
  - Result contained in **arXiv:1808.08242** and **published in Physical Review Letter**

- **SM** assumption on **Yukawa coupling to b's is confirmed** within uncertainty (~20%)

  ➔ **All 3rd generation fermion couplings are now observed!**

- Future is exiting and challenging: **reduce systematics in 2017** analysis, exploit full MC statistics @NLO, include 2018 data → **increase precision in H-b coupling**

- **DNN** plays **key role** in the 2017 analysis: **b-Reg, b-tagging, signal extraction**

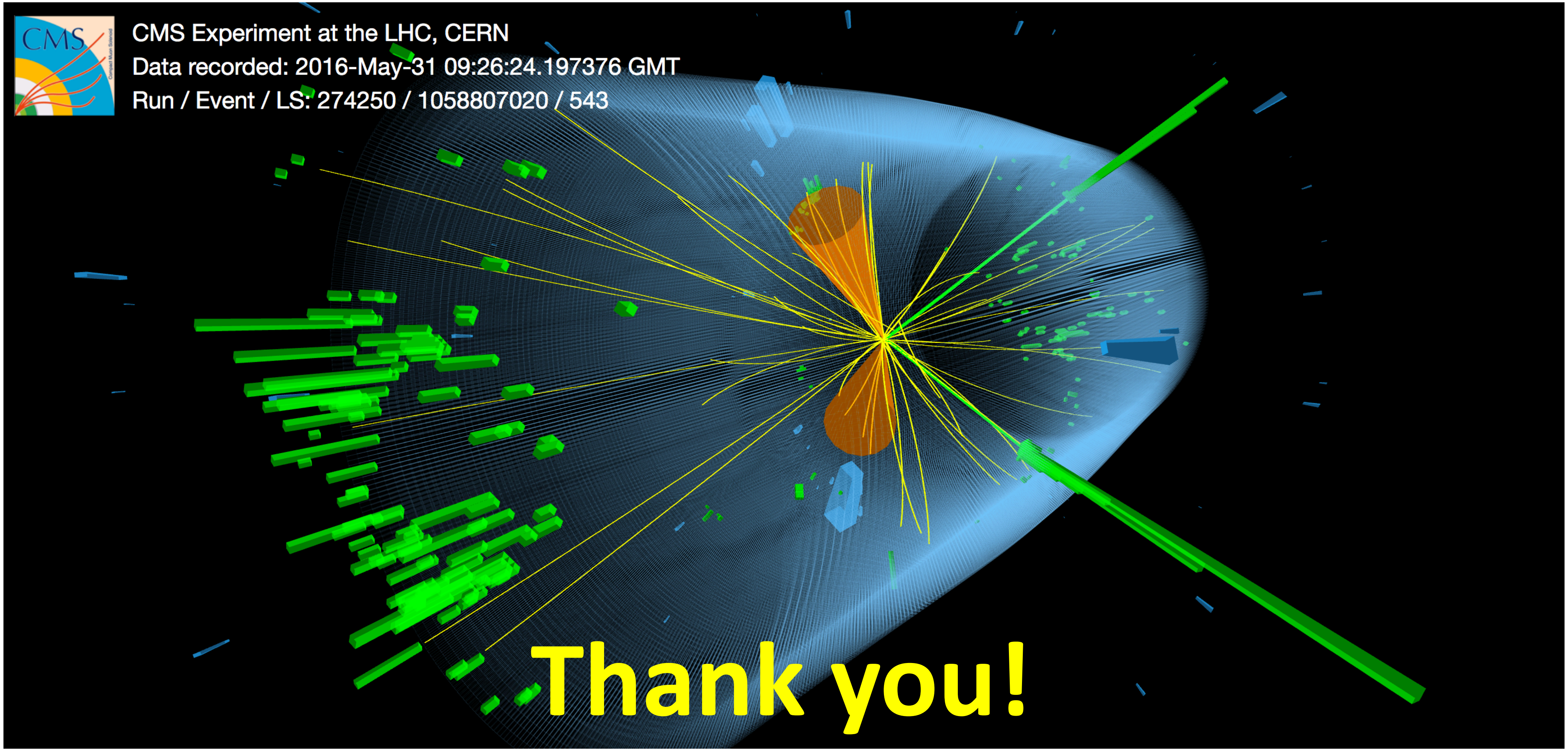
  ➔ b-Reg and b-tag in particular largely benefit from DNN

- **Looking forward: prepare for HL-LHC:** This analysis and the techniques developed to maximally increase the significance (b-reg, b-tag, kin.-fit,FSR-rec.,DNN) can represent a benchmark for other analysis looking to H→bb, e.g HH→bbXX (X=b, τ)

Conclusion (II)

- **Direct search for H→cc decay (new in CMS!)**
  - Looking to VH production mode with 2016 only dataset
  - Two strategies: Resolved and Boosted, looking to different $p_T(H)$ regimes
  - Improved results for VHcc: Exp. limits on μ~37 (ATLAS Exp. limits on μ~150)
  - First H→cc analysis in CMS [HIG-18-031]
- **Main challenging: tagging charm quarks**
  - Two different approaches in boosted and resolved analysis
  - DeepAK15 for boosted and DeepCSV-based likelihoods for resolved
  - A new method to measure the c-tagger SFs from resolved [AN-19-028]
- **What's next?**
  - Energy regression for charm initiated jets started to be investigated + kin-fit
  - Possible switch to DeepJet for Ak4 and further optimize DeepAK15
  - Analyze the full Run-2 + optimize signal extraction methods
  - Very simple projection with 140fb^{-1} ==> 95% CL. Exp. limits on μ <19

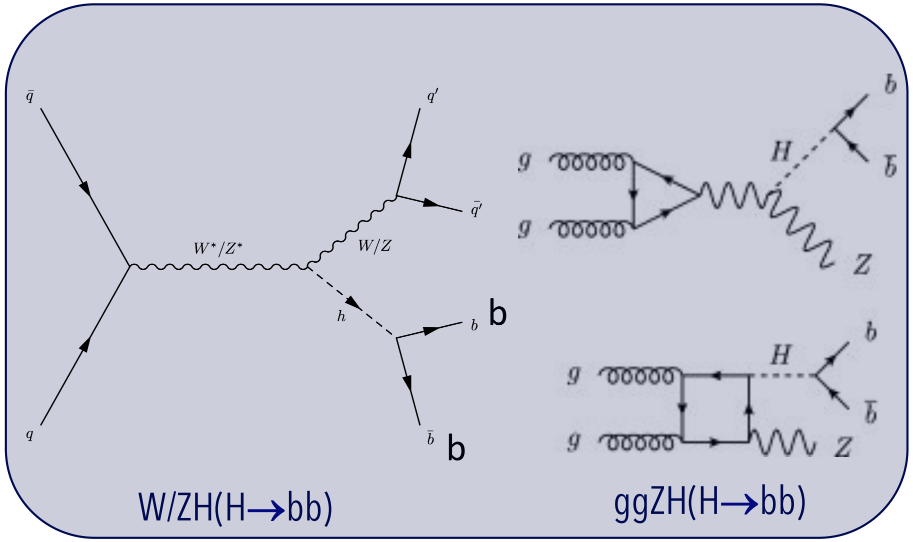
VH(H→cc) candidate - Event Display



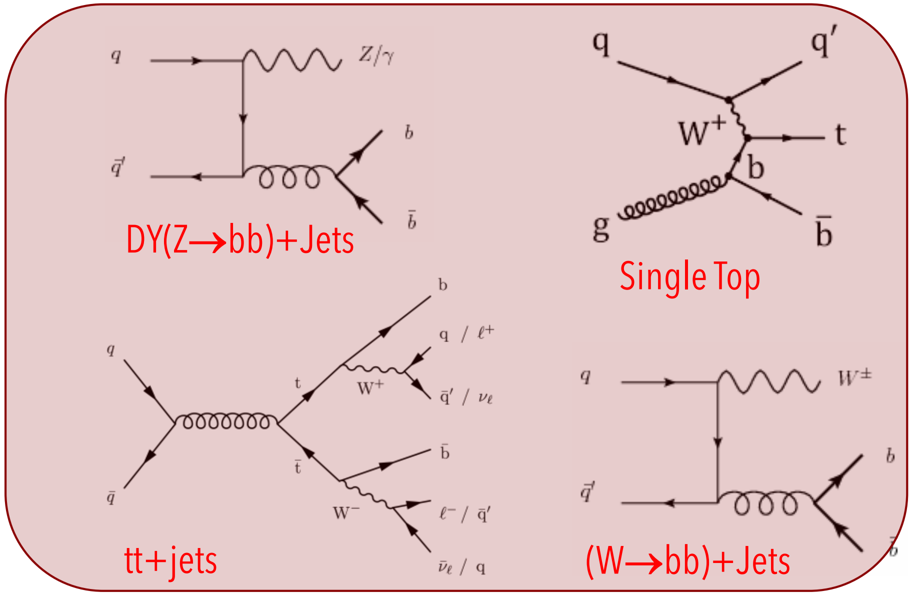
Back-Up

VH(H→bb) Analysis Strategy - 2017

SIGNAL PROCESSES



IRREDUCIBLE BACKGROUNDS



- **3 channels** with 0, 1, and 2 leptons and 2 b-tagged jets
  - To target **Z(νν)H(bb)**, **W(lν)H(bb)** and **Z(ll)H(bb)** processes
- **Signal region designed to increase S/B**
  - **Large boost** for vector boson
  - **Multivariate analysis** exploiting the most discriminating variables ($m_{b\bar{b}}$, b-tag,…)
- **Control regions to validate backgrounds and constrain normalizations**
- **Signal extraction:** binned maximum likelihood fit of final MVA distribution performed simultaneously in all the channels of all the categories in SR and CRs

Event Selection+Categorization

Triggers*

0-Lepton	1-Lepton	2-Lepton
MET>120+MHT>120	μ-pT>27 e-pT>32	μ-pT>17+μ-pT>8 e-pT>23+e-pT>12

*thresholds in GeV

Preselection

-V-p_T>100 GeV (V-p_T>50 GeV for 2-lep)
-2 Jets with p_T>20 GeV not overlapping with leptons

Selection

2-Lepton

Zee: two isolated opposite charge electron with | η |<2.5 and p_T>20 GeV
Zmm: two isolated opposite charge muon with |η|<2.4 and p_T>20 GeV

1-Lepton

Wen: one isolated electron with |η|<2.5 and p_T>30 GeV
Wmn: one isolated, tight-id muon with |η|<2.4 and p_T>25 GeV

0-Lepton

Znunu: Missing transverse energy > 170 GeV

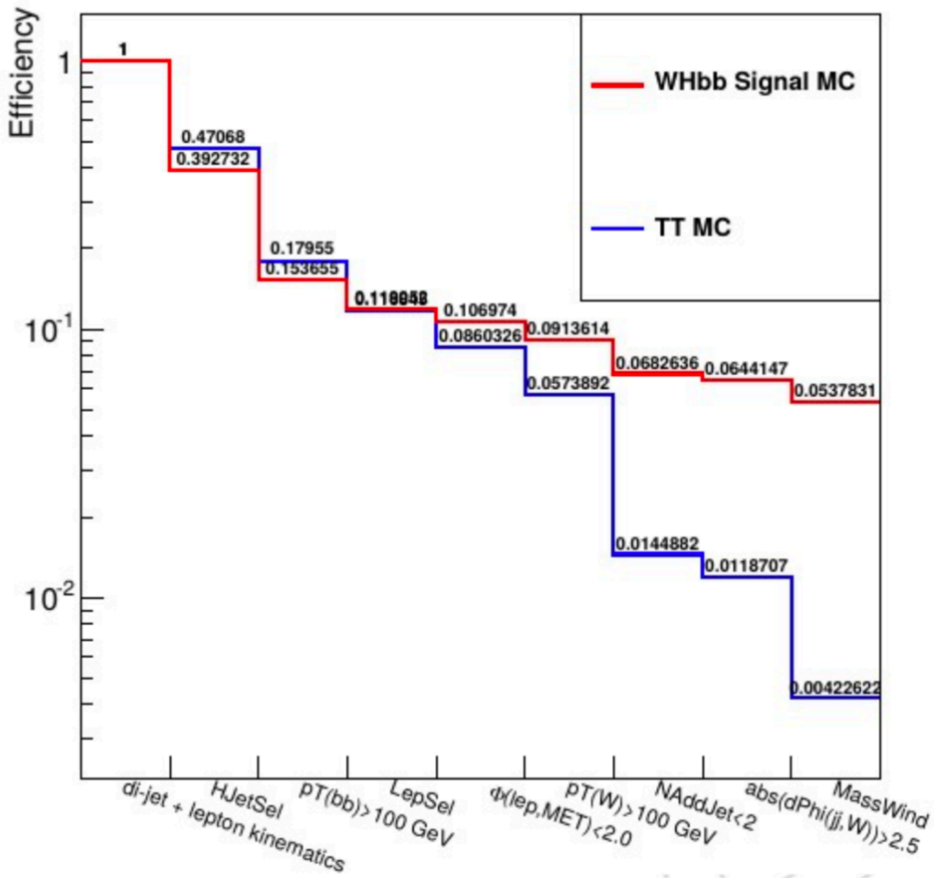
+ 2 pu-filtered jets with p_T>25GeV and |η|<2.5 fulfilling ID against PU

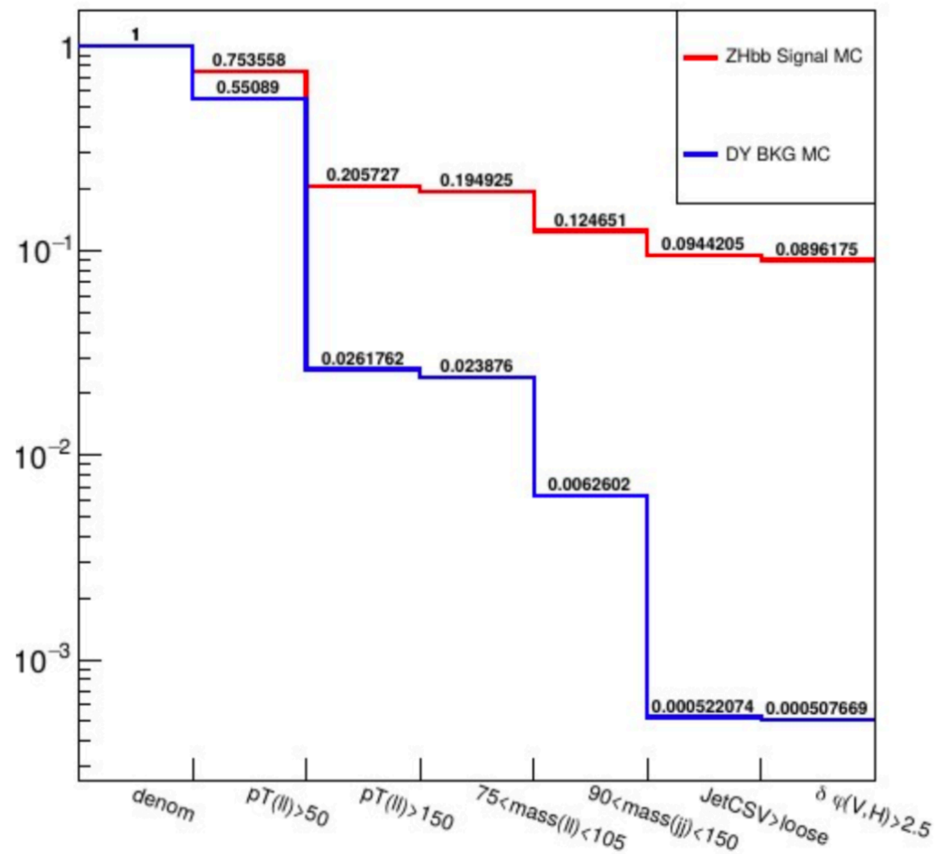
SR efficiency

Efficiencies:

1-lep 5% signal – 0.5% bkg (TT)

2-lep ~10% signal – 0.005% bkg (DY)





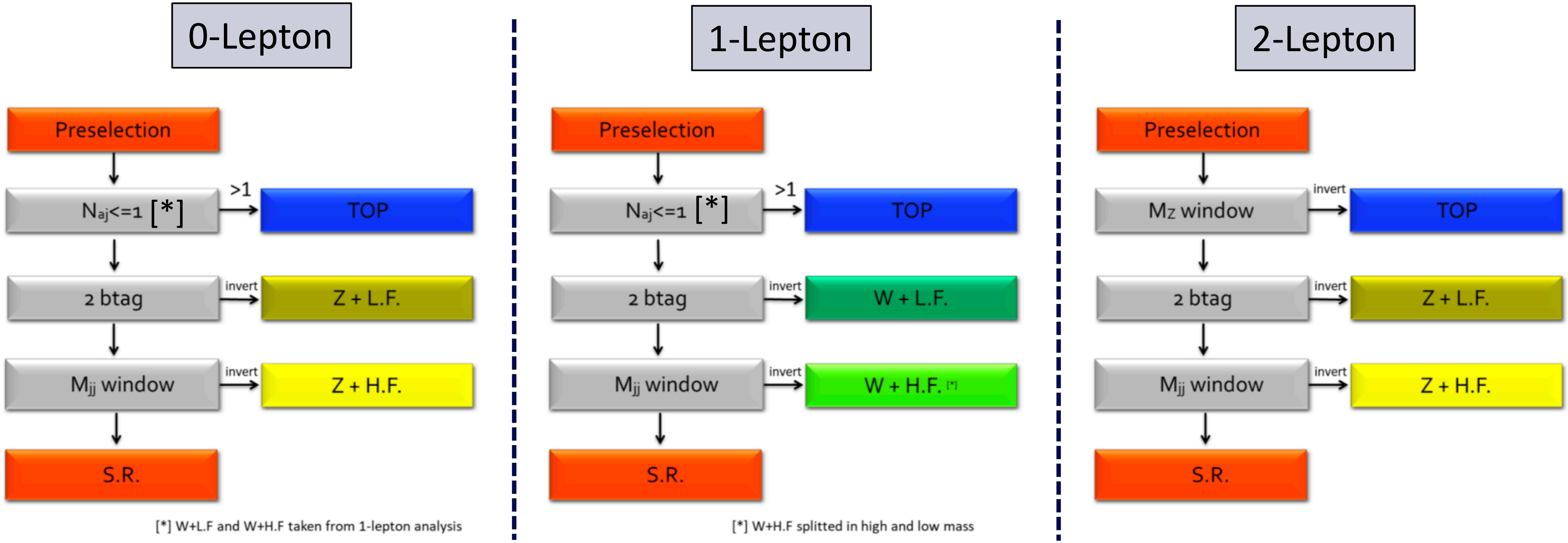
Event Selection+Categorization

- **Selections** (jets, leptons, b-tagging) **optimizd separately by channel**
  - 4 analysis categories:
    - 0-lepton: **p_T(Z) > 170 GeV**
    - 1-lepton: **p_T(W) > 150 GeV**
    - 2-lepton High-Vp_T: **p_T(Z) > 150 GeV**
    - 2-lepton Low-Vp_T: **50 GeV < p_T(Z) < 150 GeV**

- **Control regions** designed to map closely each signal region
  - Inverted selections to **enhance purity in targeted backgrounds:**

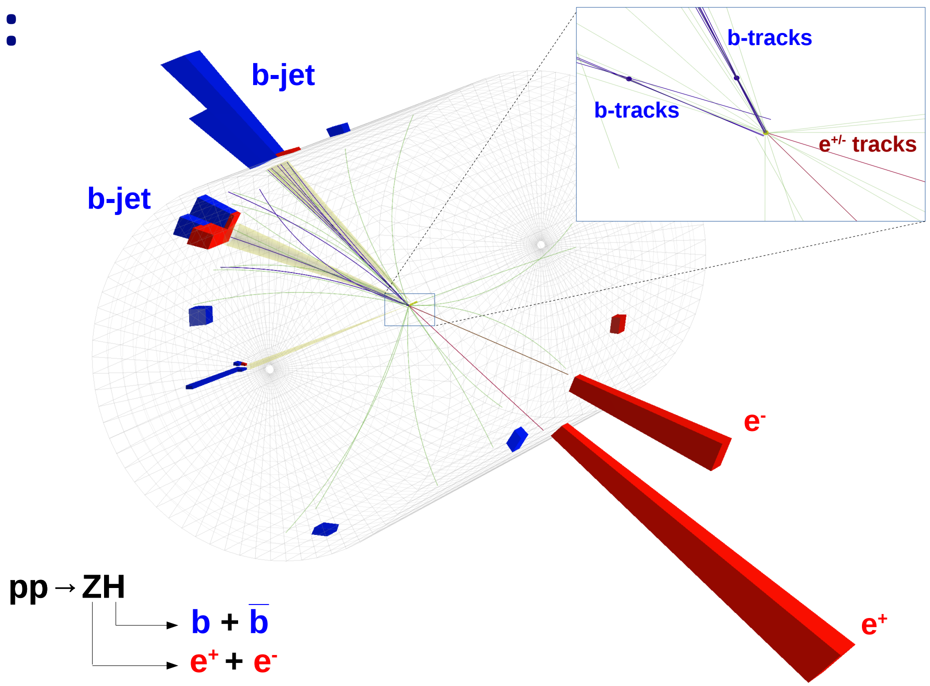
tt, V+light flavor, and V+heavy flavor

[*] Number of additional jets in the event



Improvements in VH(H→bb) 2017 analysis

- **Improved mass resolution from:**
  - Better **b-jet identification**
    - ➔Thanks to **improved b-tagger**
    - ➔**+ new pixel detector**
  - New **b-jet energy regression**
  - FSR jet recovery
  - **Kinematic fit in 2-lepton channel**

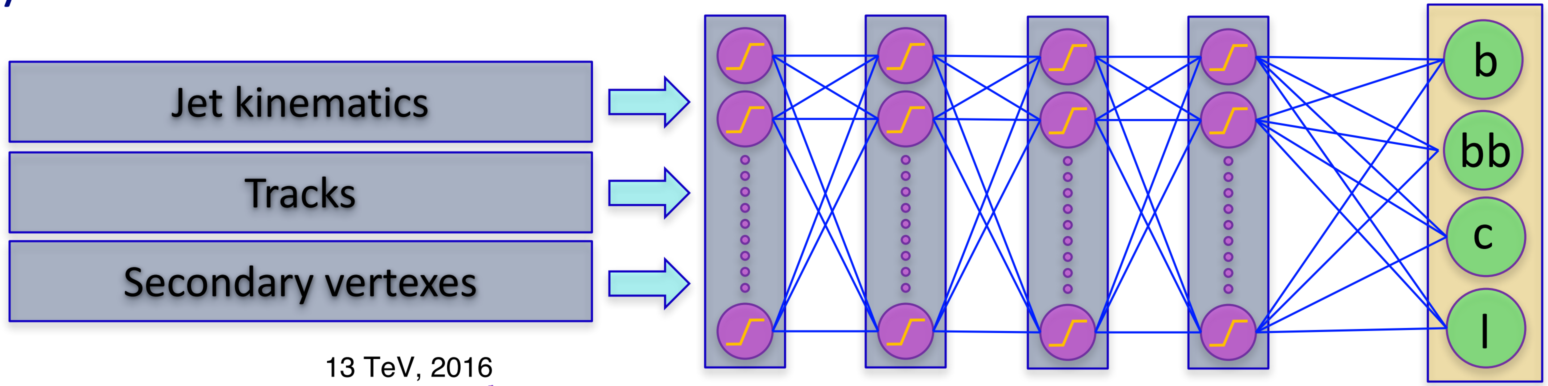


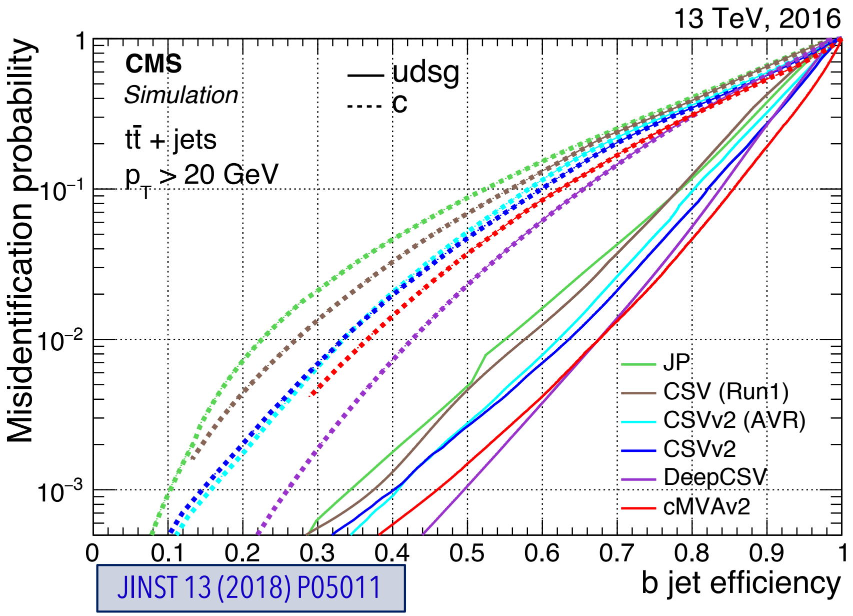
- Use of **deep neural network** (DNN) to discriminate:
  - Signal from background, in Signal Regions
  - Background components among each other, in Control Regions

- Combined effect: +O(5-10%) in the analysis sensitivity wrt 2016

State of the art b-jet identification

- **DeepCSV: Deep Neural Network architecture**
  - Input variables go through 4 fully connected layers, each layer has 100 nodes
  - ReLu activation function used in each of the hidden nodes
  - Output layer → softmax activation function → multiclassification





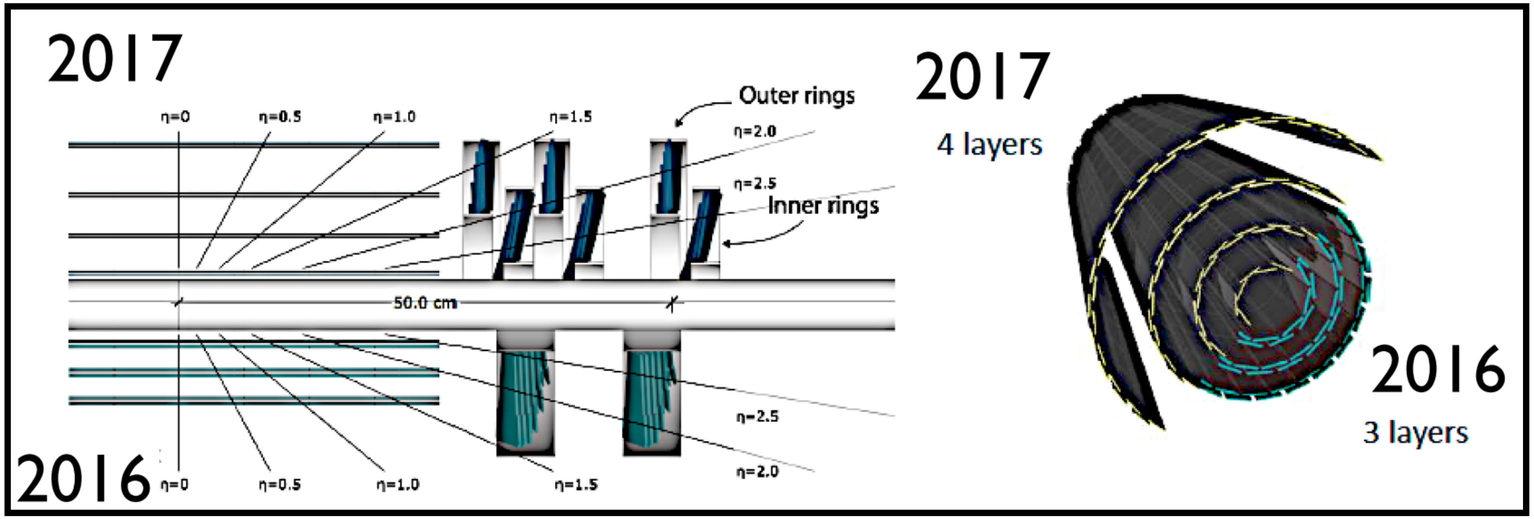
JINST 13 (2018) P05011

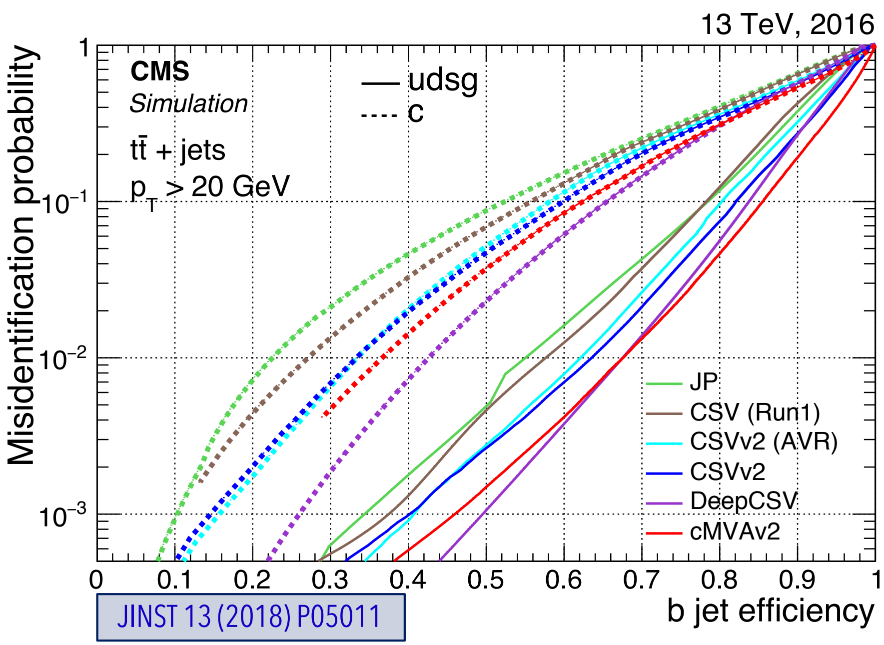
- Three working points commissioned with data
- Available set of data/MC SF for full 2017 run

Tagger	Working point	ε_b (%)	ε_c (%)	ε_{udsg} (%)
	DeepCSV L	84	41	11
Deep combined secondary vertex (DeepCSV) $P(b) + P(bb)$	DeepCSV M	68	12	1.1
	DeepCSV T	50	2.4	0.1

State of the art b-jet identification

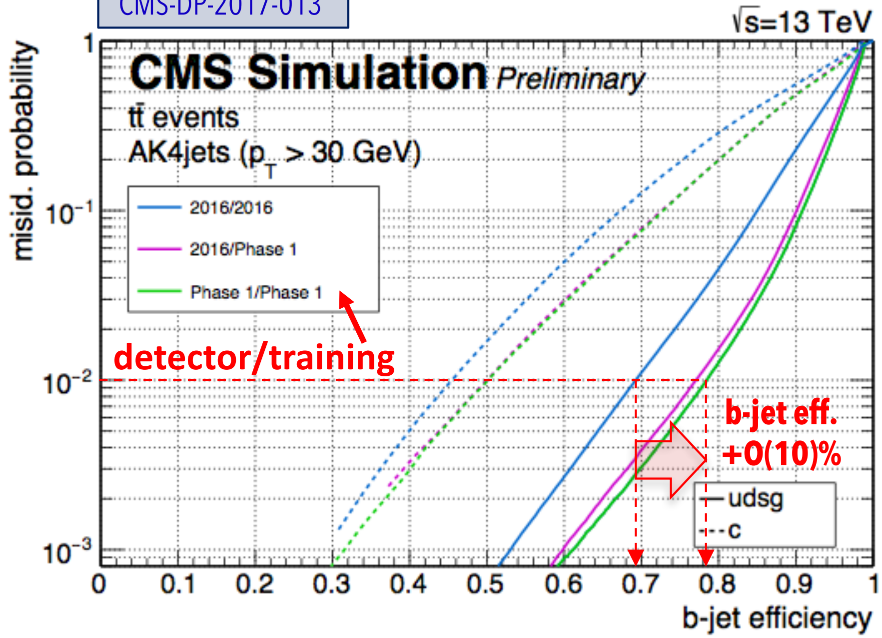
- **Upgraded pixel detector**





JINST 13 (2018) P05011

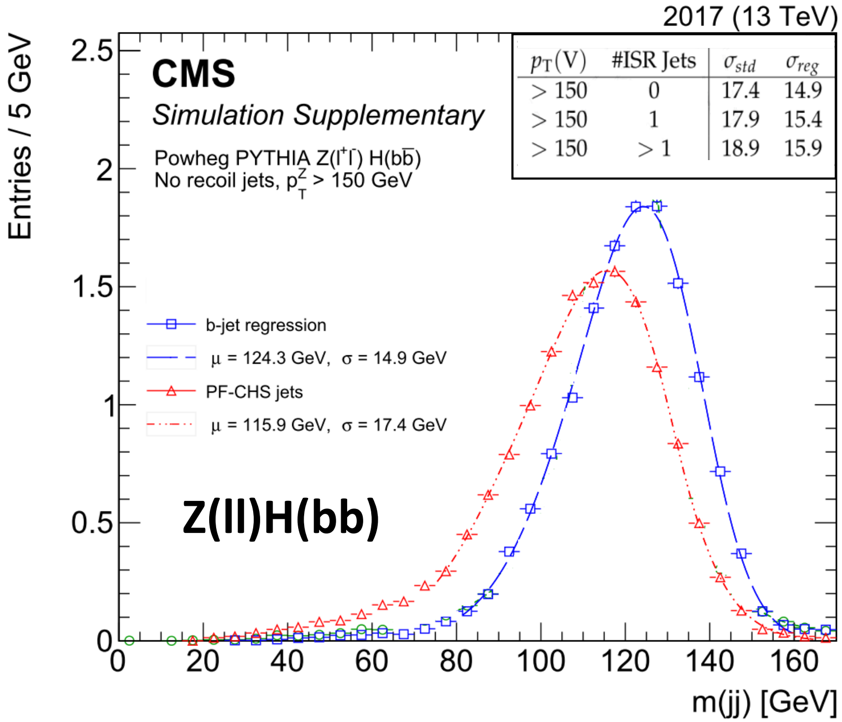
CMS-DP-2017-013



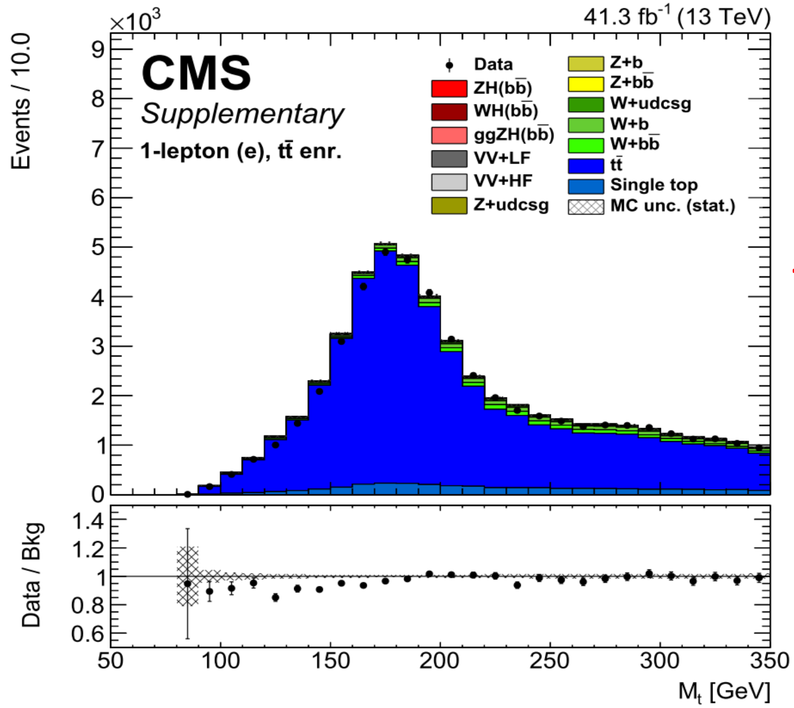
b-jet energy regression

- **Regression mainly recovers missing energy in the jet due to neutrino**
  - Switch from Boosted Decision Trees to DNN algorithm
- **Extended set of input variables** now including lepton flavor (μ/e), jet mass and energy fractions in DR rings
- **Significant m_{bb} resolution improvement without mass sculpting**
  - σ/peak down to 11.9% in 2017 wrt 13.2% in 2016 ➔ + O(10%)
  - dedicated calibration of b-jets with Z+b events + measure JER

Higgs mass in 2-lep SR



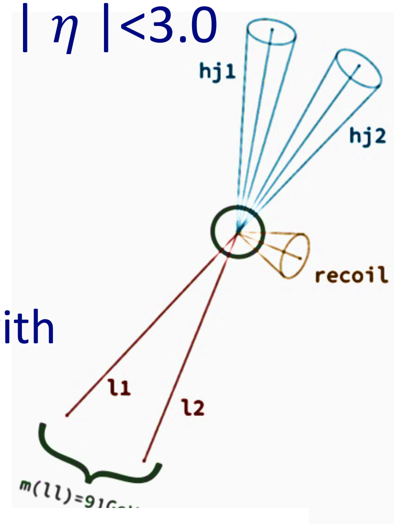
$p_T(V)$	#ISR Jets	σ_{std}	σ_{reg}
> 150	0	17.4	14.9
> 150	1	17.9	15.4
> 150	> 1	18.9	15.9

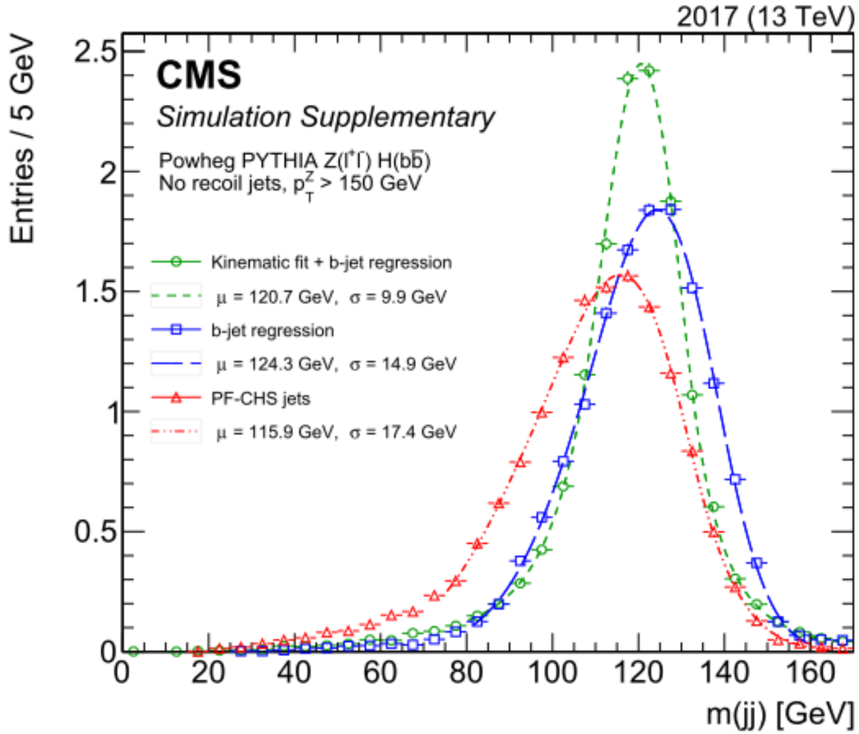


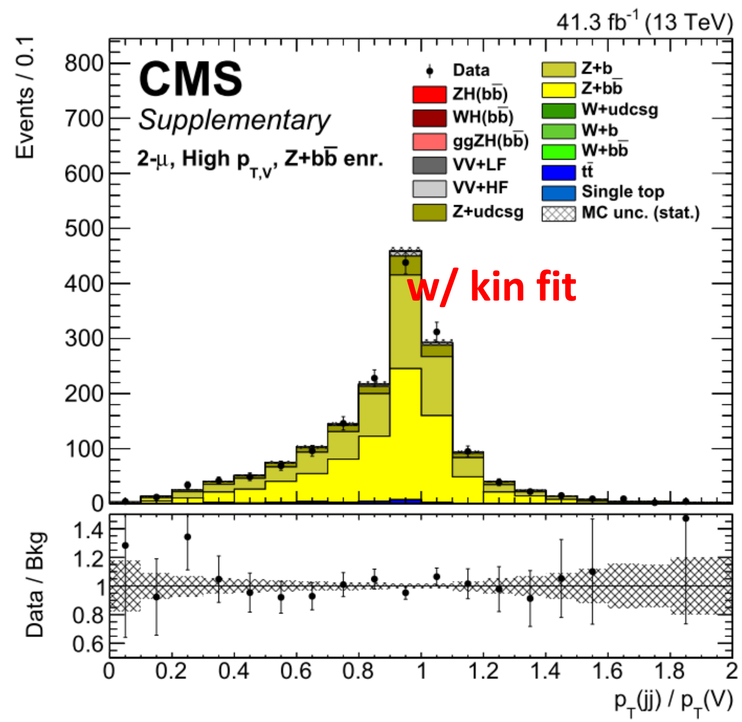
Top mass in 1-lep tt CR

FSR+Kinematic fit in 2-lepton channel

- **FSR-recovery:** additional jets in dR<0.8 cone with p_T>20GeV and | η |<3.0
- **No intrinsic missing energy in the Z(ll)H(bb) process**
  - Constrain di-lepton system to Z mass
  - **Balance the ll+bb+(jet) system in the (p_x,p_y) plane**
  - lepton and jet p_T's adjust within their experimental uncertainties with the constraint that the MET is 0 within resolution
  - Improve m(bb) resolution up to 36%



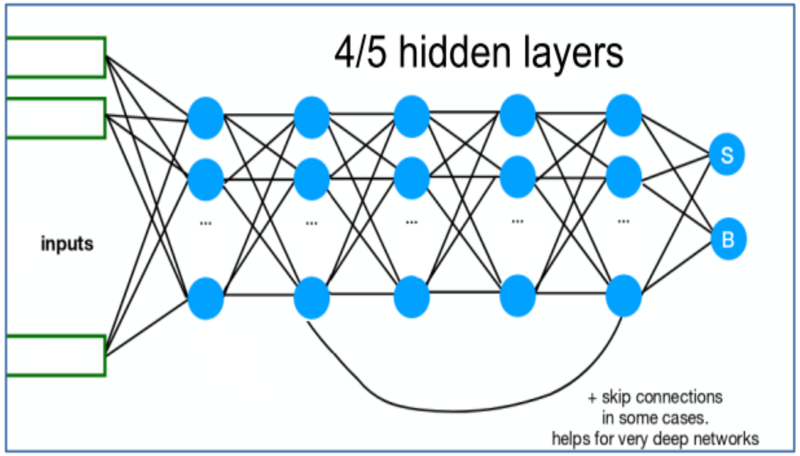




Signal vs Background discriminator

- To increase sensitivity, use **DNN discriminator to extract signal**
  - DNN outperforms BDT due to network depth
    - Same input variables as 2016 (b-jet properties, di-jet kinematics, event topology)
  - Trained separately in each channel to discriminate VH(bb) from the weighted sum of all backgrounds
  - Parameters optimized to maximize the sensitivity in each channel

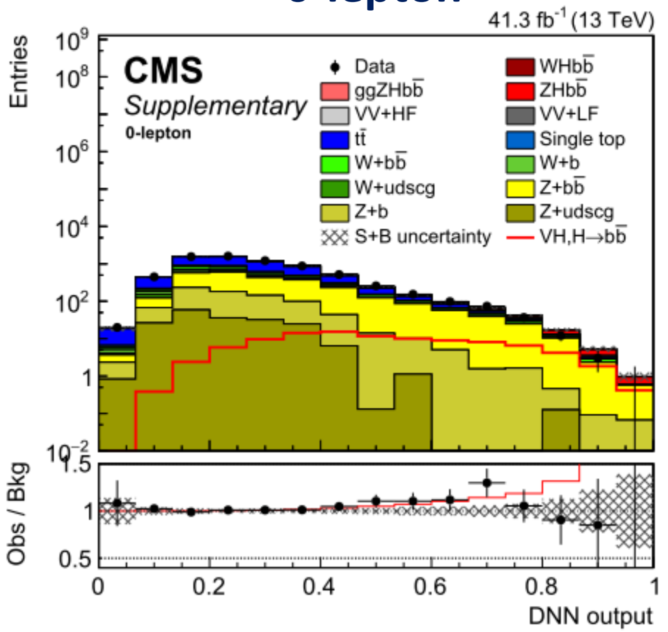
Variable	Description	0-lepton	1-lepton	2-lepton
$M(\mathrm{jj})$	dijet invariant mass	✓	✓	✓
$p_T(\mathrm{jj})$	dijet transverse momentum	✓	✓	✓
$p_T(\mathrm{j}_1)$, $p_T(\mathrm{j}_2)$	transverse momentum of each jet	✓		✓
$\Delta R(\mathrm{jj})$	distance in η–ϕ between jets			✓
$\Delta\eta(\mathrm{jj})$	difference in η between jets	✓		✓
$\Delta\varphi(\mathrm{jj})$	azimuthal angle between jets	✓		
$p_T(\mathrm{V})$	vector boson transverse momentum		✓	✓
$\Delta\phi(\mathrm{V,H})$	azimuthal angle between vector boson and dijet directions	✓	✓	✓
$p_T(\mathrm{jj})/p_T(\mathrm{V})$	p_T ratio between dijet and vector boson			✓
M_Z	reconstructed Z boson mass			✓
$btag_{max}$	value of the b-tagging discriminant (DeepCSV) for the jet with highest score	✓		✓
$btag_{min}$	value of the b-tagging discriminant (DeepCSV) for the jet with second highest score	✓	✓	✓
$btag_{add}$	value of b-tagging discriminant for the additional jet with highest value	✓		
E_T^{miss}	missing transverse momentum	✓	✓	✓
$\Delta\phi(E_T^{miss},\mathrm{j})$	azimuthal angle between E_T^{miss} and closest jet with $p_T > 30\,\mathrm{GeV}$	✓		
$\Delta\phi(E_T^{miss},\ell)$	azimuthal angle between E_T^{miss} and lepton		✓	
m_T	mass of lepton $\vec{p}_T$ + E_T^{miss}		✓	
M_t	reconstructed top quark mass		✓	
N_{aj}	number of additional jets		✓	✓
$p_T(\mathrm{add})$	transverse momentum of leading additional jet	✓		
SA5	number of soft-track jets with $p_T > 5\,\mathrm{GeV}$	✓	✓	✓



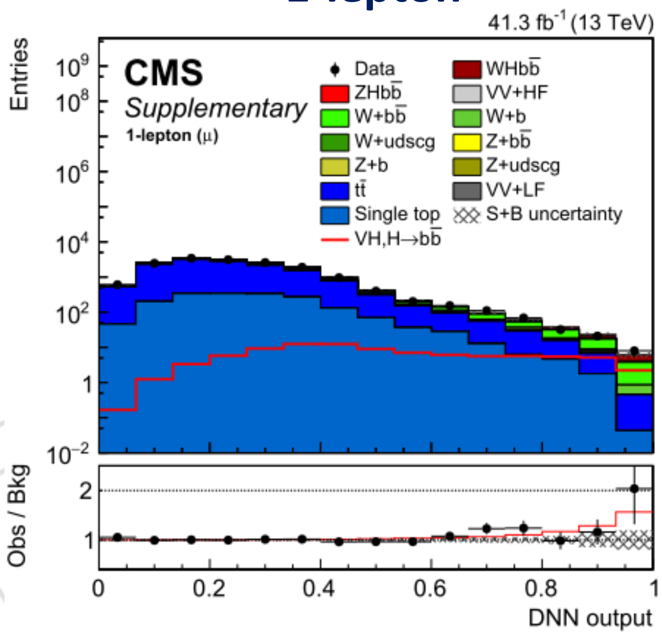
Signal vs Background discriminator

- **DNN discriminator used to extract signal**
  - **Input variables: b-jet properties, di-jet kinematics, event topology,** carefully validated through data/MC comparison
  - Trained separately in each channel
  - Performance optimization with blind analysis

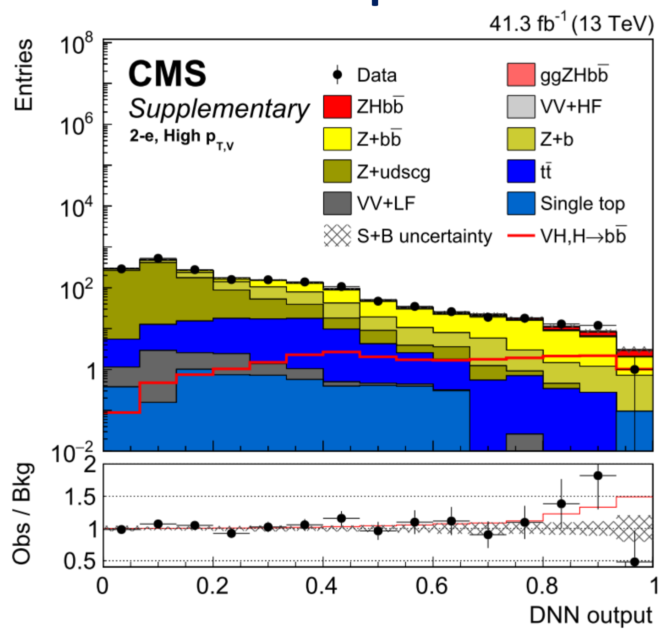
0-lepton



1-lepton



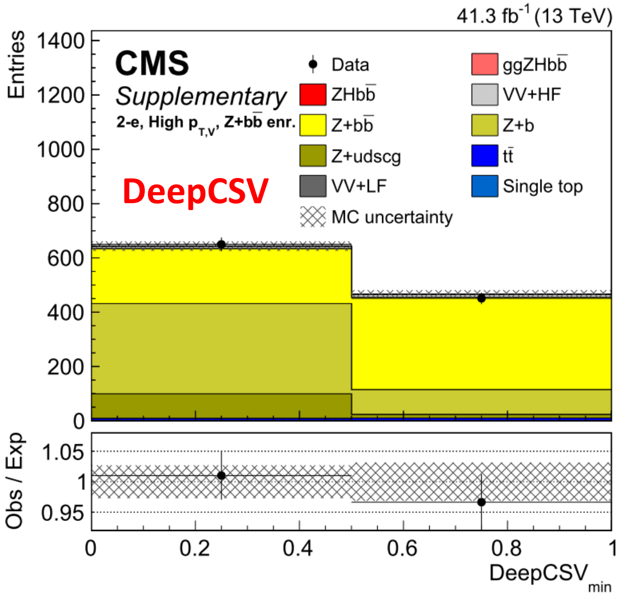
2-lepton



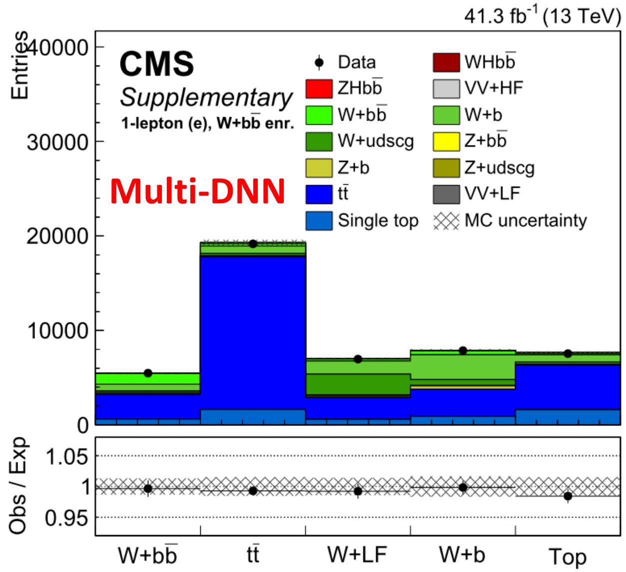
Heavy Flavor control region discriminators

- Reminder: leading systematic uncertainty from normalization of V+(b)b
- **2-lepton channel control region very pure**
  - Fit **b-tag** shape (DeepCSV) to discriminate processes
- **0- and 1-lepton channel control regions less pure**
  - Fit **DNN multi-categorizer** to distinguish among background components
    - Use same input variables as Signal vs Background discriminator

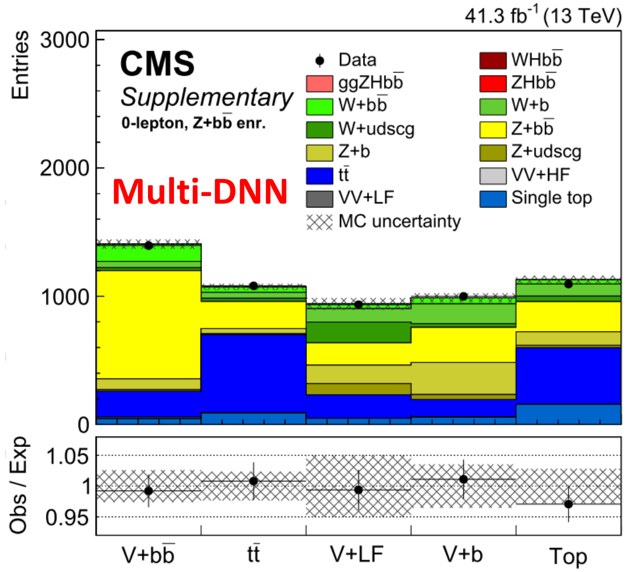
2-lepton



1-lepton



0-lepton



Post-fit normalization + systematics

- MC shapes floated within constraints from systematic uncertainties through nuisance parameters in the final fit
- MC normalization truly float ➔ fitted SFs in agreement with those measured in 2016 analysis

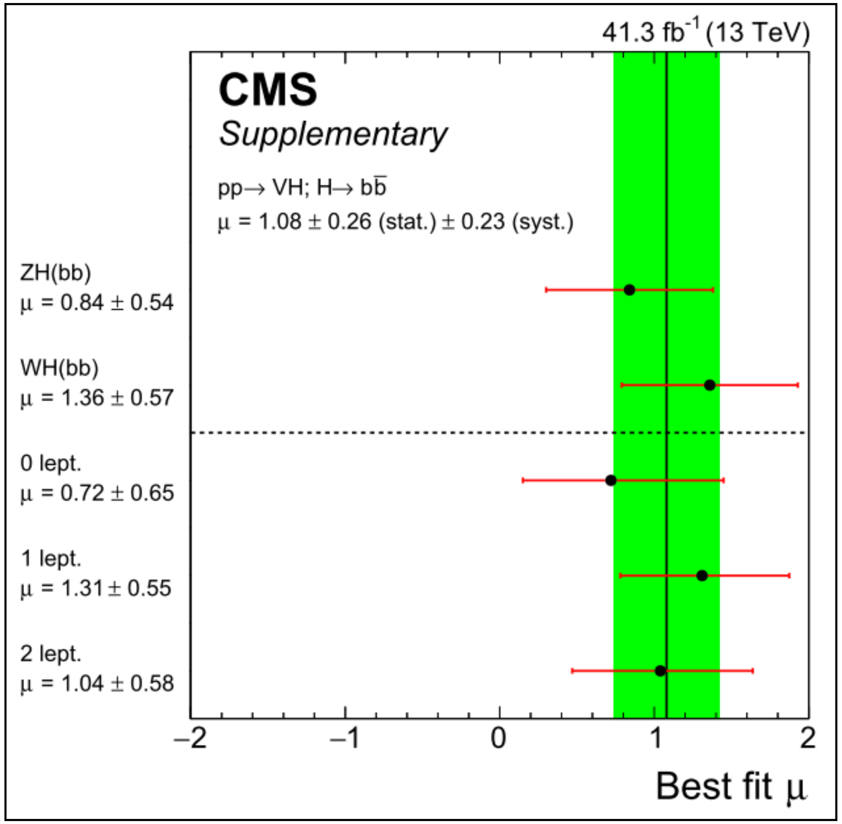
Process	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H low-p_T	Z($\ell\ell$)H high-p_T
W + udscg	1.04 ± 0.07	1.04 ± 0.07	–	–
W + b	2.09 ± 0.16	2.09 ± 0.16	–	–
W + b$\bar{b}$	1.74 ± 0.21	1.74 ± 0.21	–	–
Z + udscg	0.95 ± 0.09	–	0.89 ± 0.06	0.81 ± 0.05
Z + b	1.02 ± 0.17	–	0.94 ± 0.12	1.17 ± 0.10
Z + b$\bar{b}$	1.20 ± 0.11	–	0.81 ± 0.07	0.88 ± 0.08
t$\bar{t}$	0.99 ± 0.07	0.93 ± 0.07	0.89 ± 0.07	0.91 ± 0.07

- **Total uncertainty on μ~34%**
- Major sources of systematic uncertainties:
  - background normalization
  - background modeling
  - b-tagging
  - MC sample size

Uncertainty source	$\Delta\mu$	
Statistical	+0.26	−0.26
Normalization of backgrounds	+0.12	−0.12
Experimental	+0.16	−0.15
b-tagging efficiency and misid	+0.09	−0.08
V+jets modeling	+0.08	−0.07
Jet energy scale and resolution	+0.05	−0.05
Lepton identification	+0.02	−0.01
Luminosity	+0.03	−0.03
Other experimental uncertainties	+0.06	−0.05
MC sample size	+0.12	−0.12
Theory	+0.11	−0.09
Background modeling	+0.08	−0.08
Signal modeling	+0.07	−0.04
Total	+0.35	−0.33

VH(H→bb) Results with 2017 data

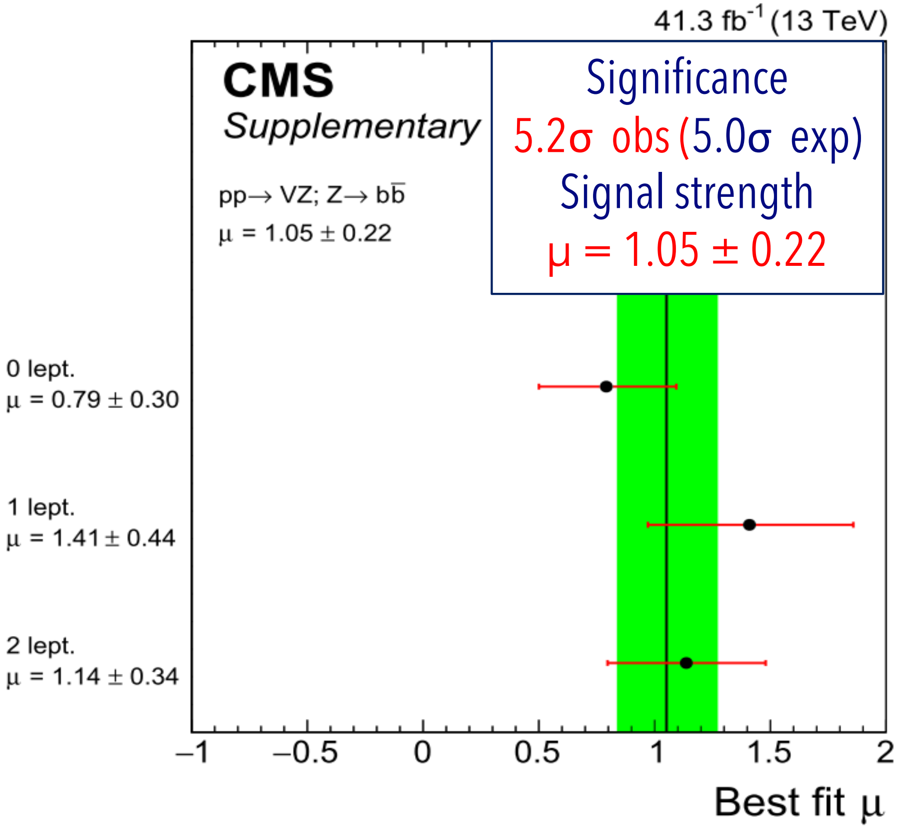
Data set	Significance (σ) Expected	Observed	Signal strength
2017			
0-lepton	1.9	1.3	0.73 ± 0.65
1-lepton	1.8	2.6	1.32 ± 0.55
2-lepton	1.9	1.9	1.05 ± 0.59
Combined	3.1	3.3	1.08 ± 0.34
2016	2.8	3.3	1.2 ± 0.4



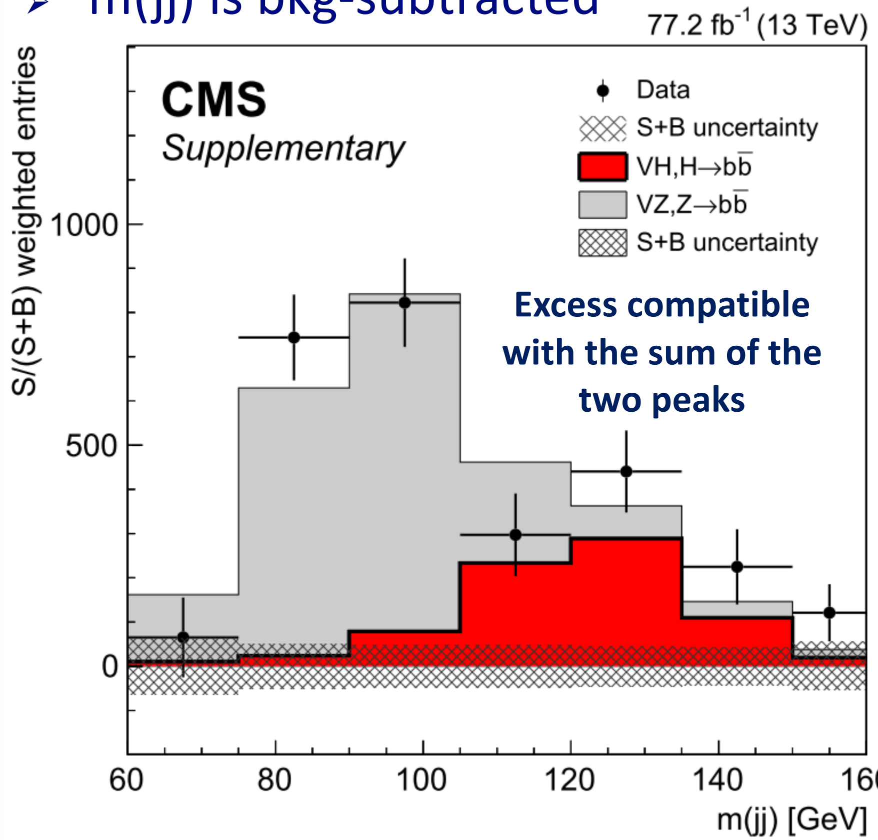
- **Standalone evidence for H→bb with 2017 data**
  - **Observed significance 3.3σ, signal strength 1.08 ± 0.34**
  - O(5-10%) increase in analysis sensitivity wrt 2016, depending on channel
  - Signal strengths extracted from each channels are compatible

Validation (VZ(Z→bb)) and Visualization (m_{jj})

- **VZ analysis** using Z(bb) standard candle
- **Same "technology" as used for VH(bb)**
  - Same DNN inputs and CRs
  - VH(bb) normalized to SM
  - Larger m(bb) window in SR



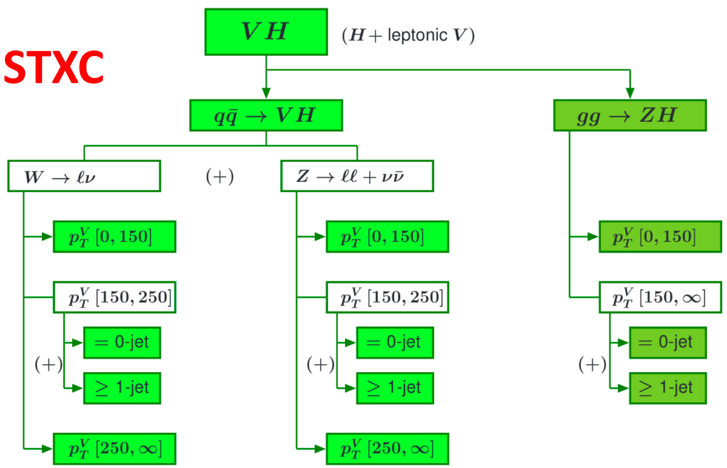
- **Fit to the m(jj)**:
  - Lower sensitivity
  - direct visualization of the signal
  - m(jj) distributions combined and weighted by S/(S + B)
  - m(jj) is bkg-subtracted



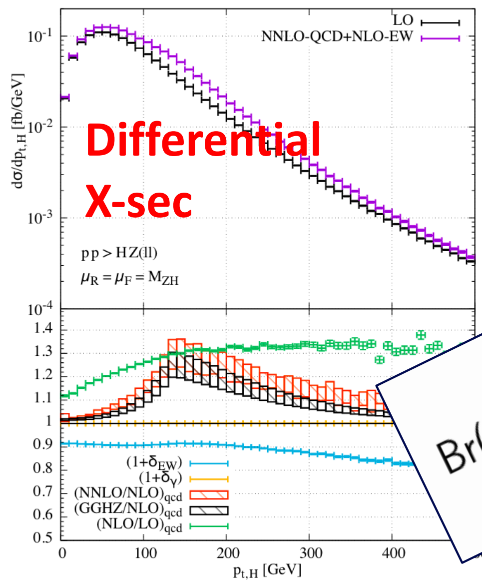
What's next?

Experimental	+0.16	−0.15
b-tagging efficiency and misid	+0.09	−0.08
V+jets modeling	+0.08	−0.07
Jet energy scale and resolution	+0.05	−0.05
Lepton identification	+0.02	−0.01
Luminosity	+0.03	−0.03
Other experimental uncertainties	+0.06	−0.05
MC sample size	+0.12	−0.12
Theory	+0.11	−0.09
Background modeling	+0.08	−0.08
Signal modeling	+0.07	−0.04

STXC



Differential X-sec



$$Br(h \to b\bar{b}) = \frac{(\kappa_b^2 + \tilde{\kappa}_b^2)Br(h \to b\bar{b})_{SM}}{1 + (\kappa_b^2 + \tilde{\kappa}_b^2 - 1)Br(h \to b\bar{b})_{SM}}$$

Anomalous couplings

SF comparison: 2016 vs 2017

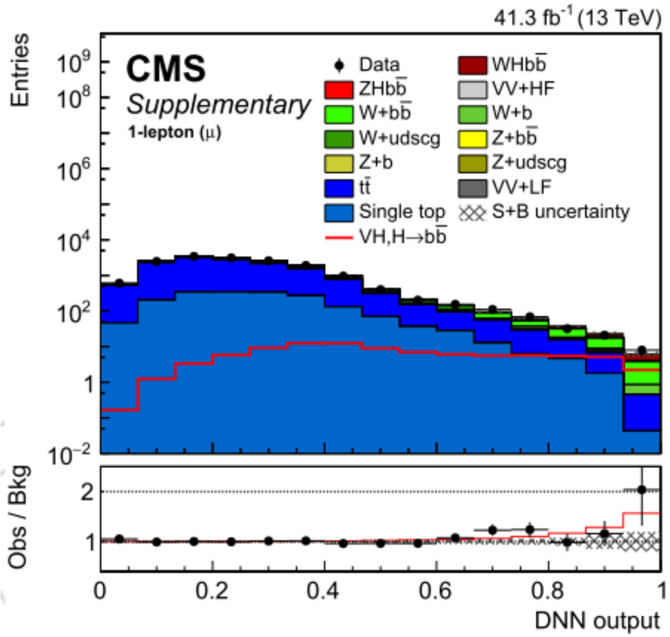
2016

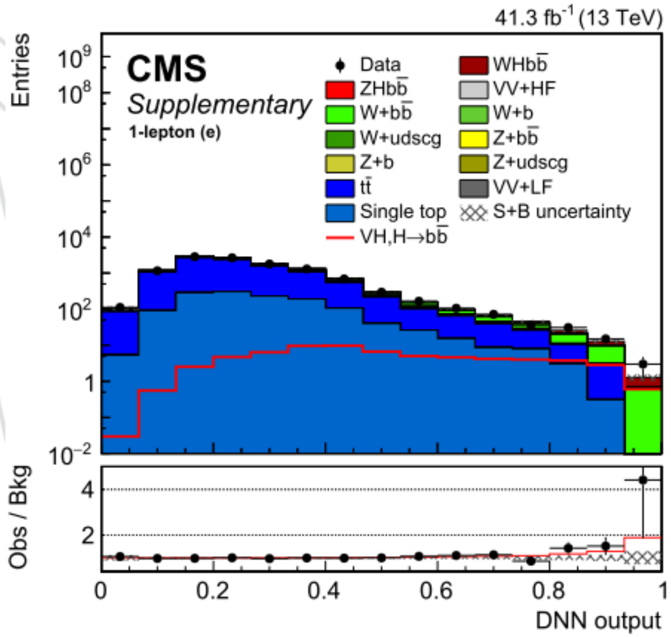
Process	0-lepton	1-lepton	2-lepton low-p_T(V)	2-lepton high-p_T(V)
W0b	1.14 ± 0.07	1.14 ± 0.07	—	—
W1b	1.66 ± 0.12	1.66 ± 0.12	—	—
W2b	1.49 ± 0.12	1.49 ± 0.12	—	—
Z0b	1.03 ± 0.07	—	1.01 ± 0.06	1.02 ± 0.06
Z1b	1.28 ± 0.17	—	0.98 ± 0.06	1.02 ± 0.11
Z2b	1.61 ± 0.10	—	1.09 ± 0.07	1.28 ± 0.09
$t\bar{t}$	0.78 ± 0.05	0.91 ± 0.03	1.00 ± 0.03	1.04 ± 0.05

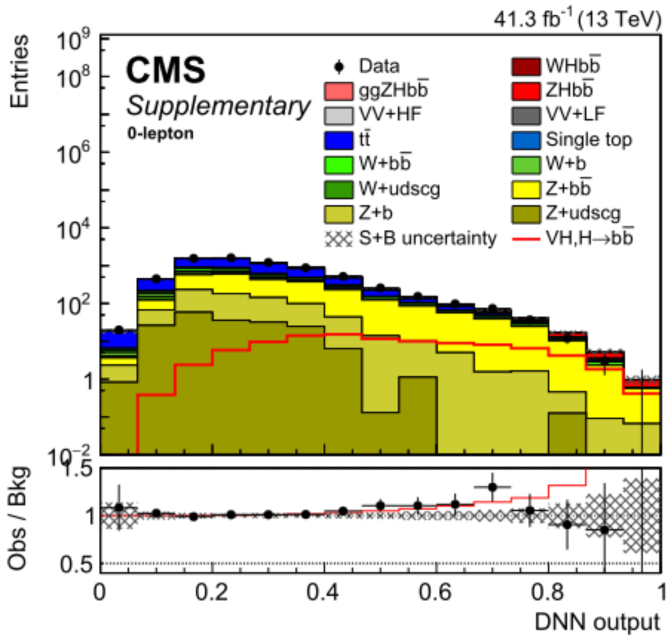
2017

Process	$Z(\nu\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$ low-p_T	$Z(\ell\ell)H$ high-p_T
W + udscg	1.04 ± 0.07	1.04 ± 0.07	–	–
W + b	2.09 ± 0.16	2.09 ± 0.16	–	–
$W + b\bar{b}$	1.74 ± 0.21	1.74 ± 0.21	–	–
Z + udscg	0.95 ± 0.09	–	0.89 ± 0.06	0.81 ± 0.05
Z + b	1.02 ± 0.17	–	0.94 ± 0.12	1.17 ± 0.10
$Z + b\bar{b}$	1.20 ± 0.11	–	0.81 ± 0.07	0.88 ± 0.08
$t\bar{t}$	0.99 ± 0.07	0.93 ± 0.07	0.89 ± 0.07	0.91 ± 0.07

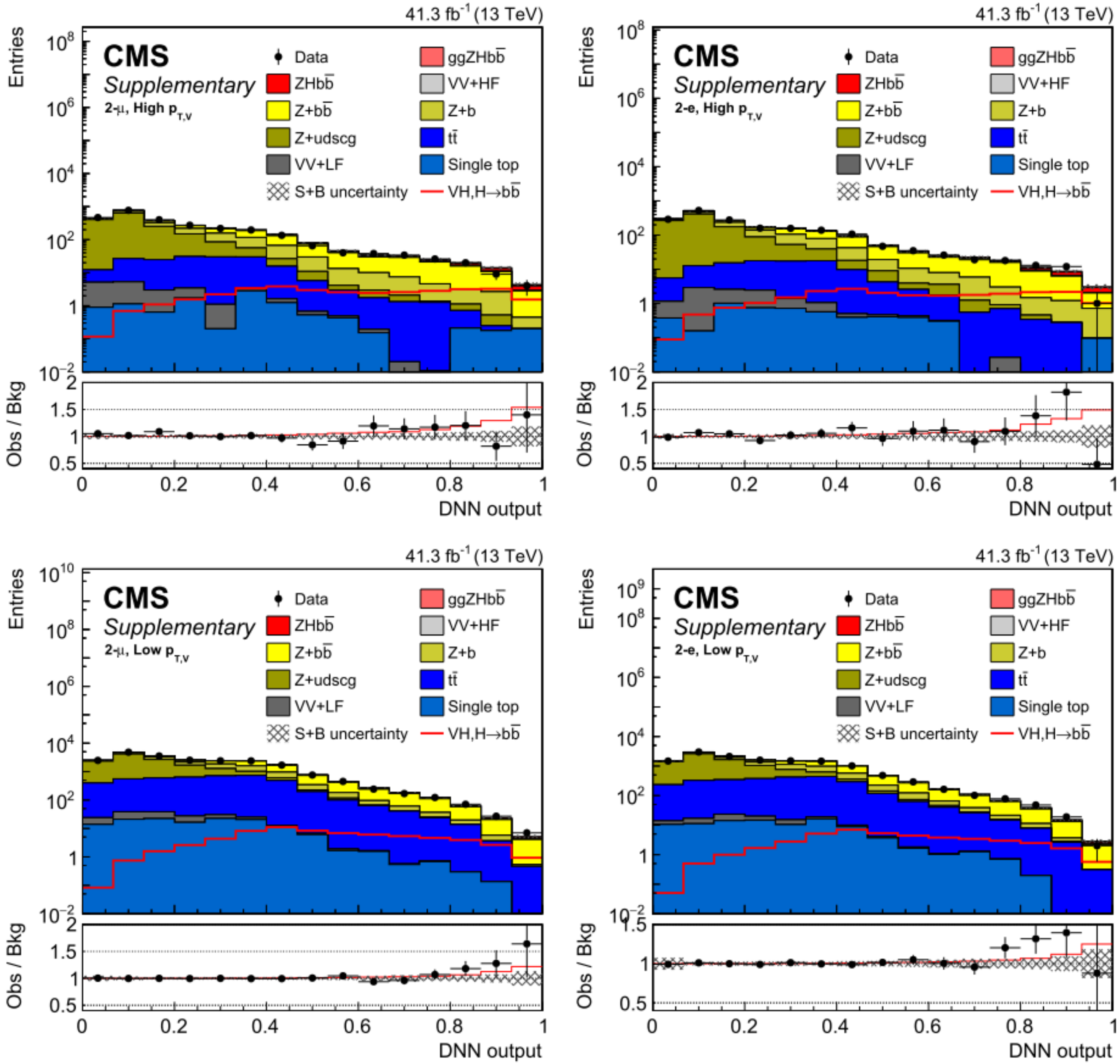
0- and 1-lepton signal regions' DNN







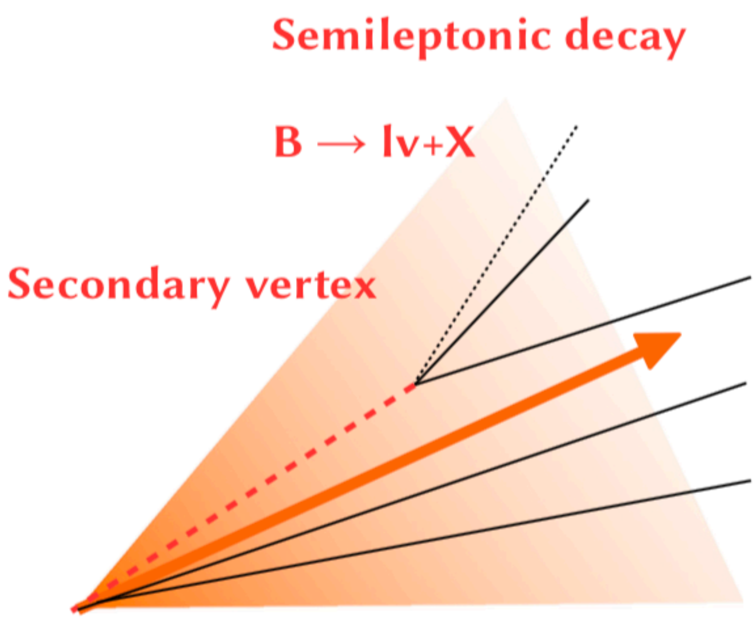
2-lepton signal regions' DNN



Regression inputs

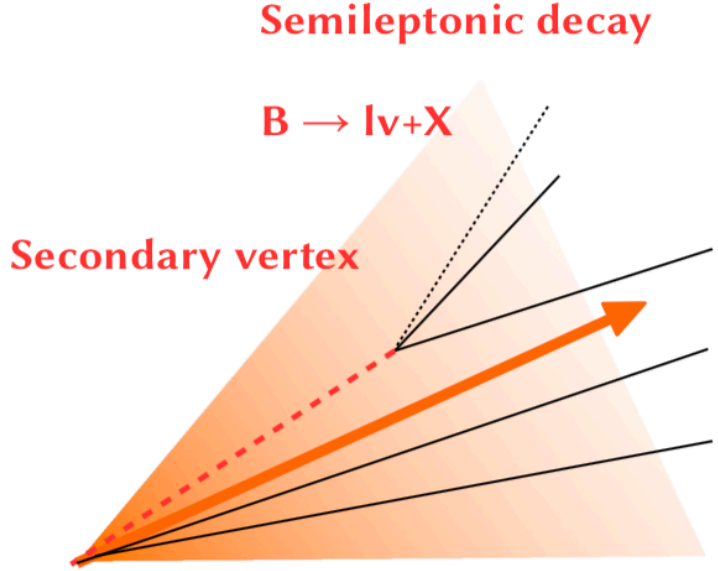
Optimized set of inputs

- Jet kinematics
  - jet p_T, η, and transverse mass
- PU information
  - nPVs or rho
- Jet energy fractions
- Jet leading track and soft lepton track
  - p_T component and distance relative to the jet axis of the soft-lepton candidate
- Secondary vertex
  - p_T, mass and # of charged tracks associated to the secondary vertex, decay length and uncertainty of the secondary vertex

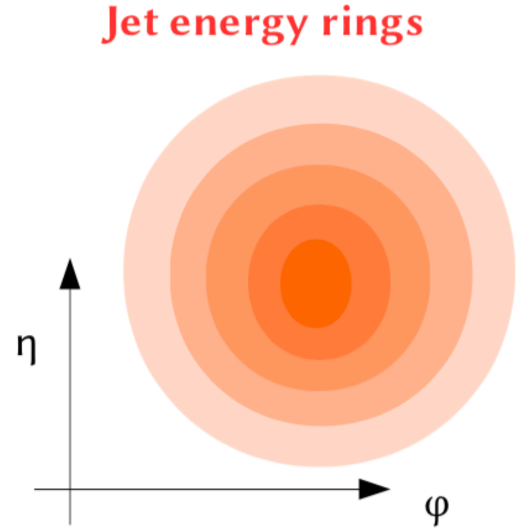


What's new?

- **Optimized set of inputs**
  - Jet kinematics → **uncorrected 4-vector**
  - Pile-Up information
  - Jet energy fractions
  - Leading track, soft lepton track, SV



- **New inputs :**
  - **Jet shape:** energy fractions in rings of dR, energy spread ($p_T D$)
  - Multiplicity of jet constituents
  - Lepton ID (e/μ)
  - Jet p_T rel wrt to lepton, jet mass



Validation on data

- p_T^{reco} i.e. the TARGET is a *"MC variable"*
  - L123 jet energy corrections are used, but no resolution scale factor is applied
  - Resolution to be compared in MC and data after the regression, as a function of p_T, η, ...

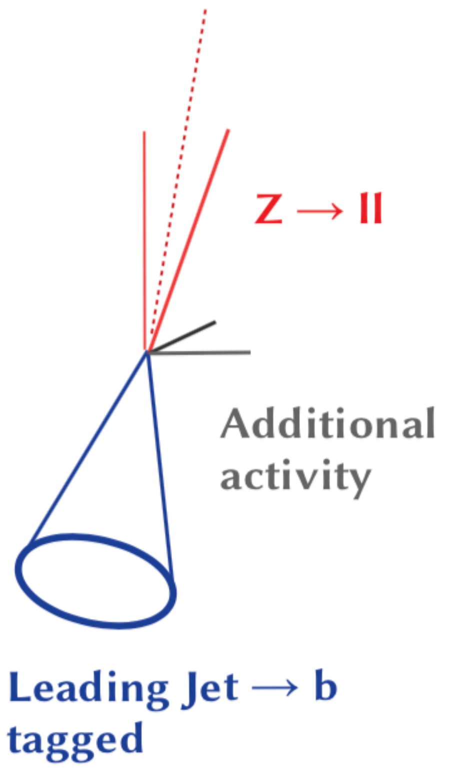
- Aim of this effort is reducing the JES uncertainty
  - B-jets are better measured thanks to the regression
  - We may be able to reduce the JER scale factor and the uncertainty (not in 2017 data)

Validation on data for $H \to b\bar{b}$

- Performance in data evaluated with **p_T balance** in $Z \to \mu\mu/ee$+b-jet topology
  - Leading jet collinear with Z ($|d\Phi|>2.8$)
  - NO Additional activity: ($\alpha = p_T$ 2nd jet / p_T Z >0.3) and α binning

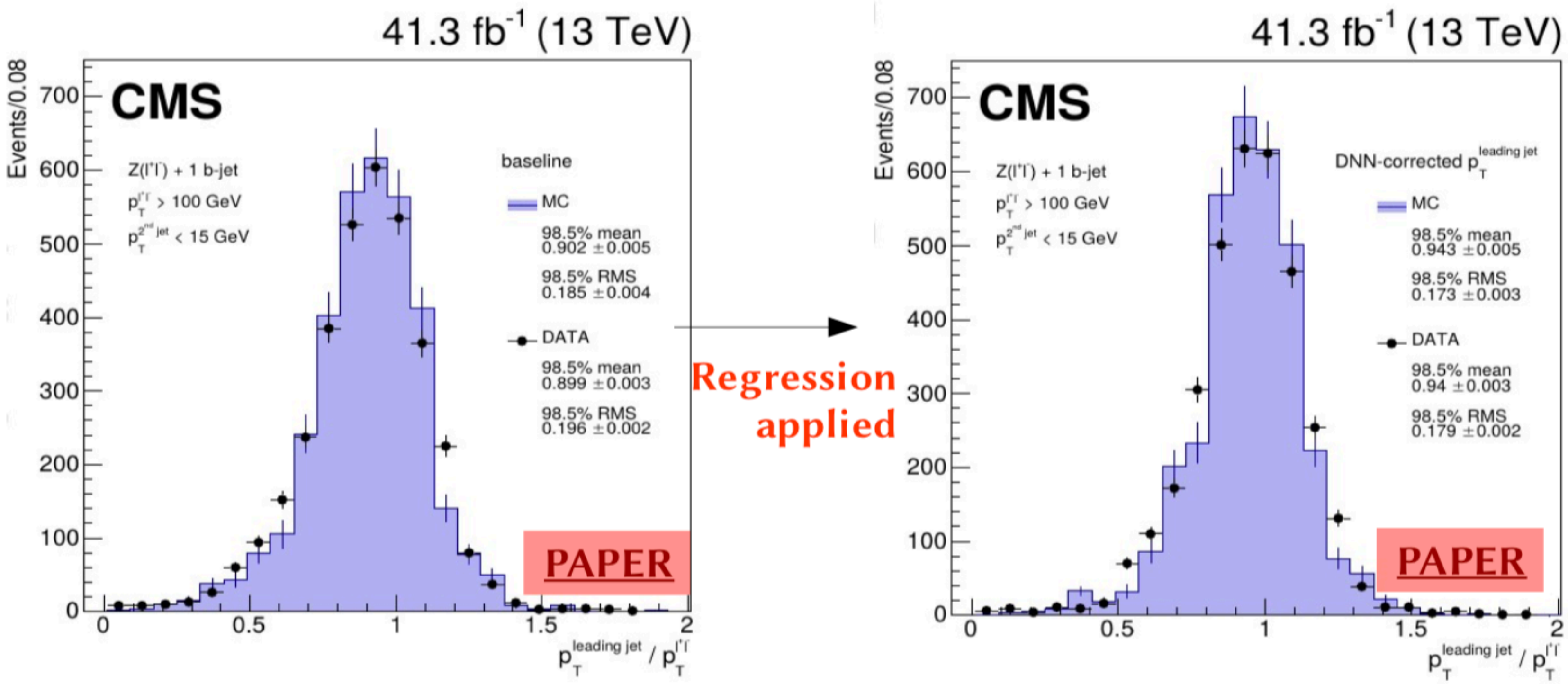
    $\longrightarrow$ Extrapolation in α to estimate JER scale factor, as prescribed by JME (CMS AN-2011/004, JME-10-014), truncated RMS used
  - leading jet p_T and $|\eta|$ fiducial region

    $p_T>100$ GeV, $|\eta|<2.0$
  - b-jet enriched region $\longrightarrow$ b-tagged leading jet, (deepCSV medium WP)

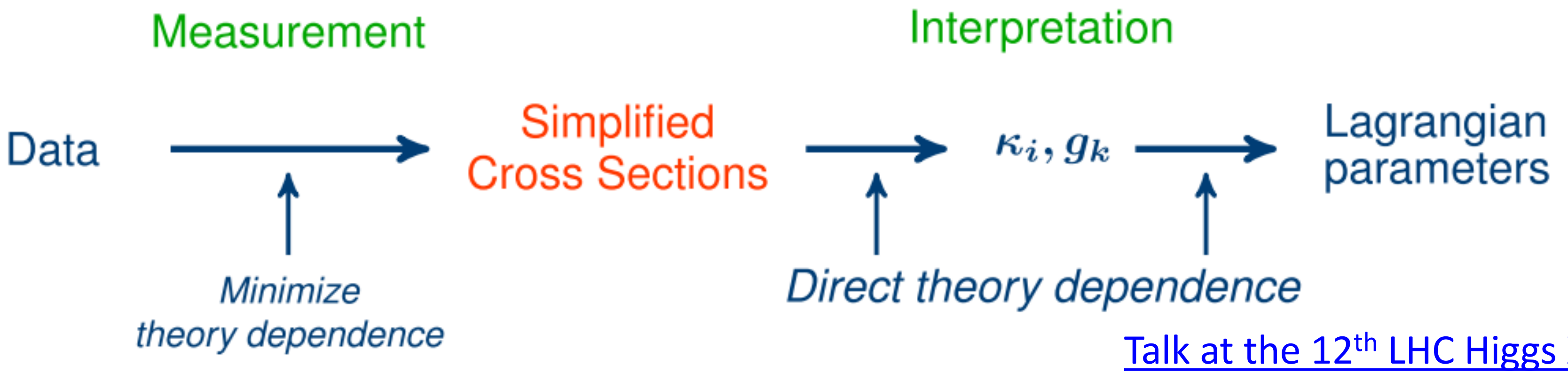


Validation on data for H→b$\bar{b}$

- Performance in well balanced events (extra jets p_T <15 GeV) – not used in extrapolation:
  - Truncated (98.5%) mean consistent in MC and data (0.9 → 0.94)
  - Truncated (98.5%) RMS improvement in MC and data
    - ~10% JER scale factor needed to account for the different resolution
      → same as standard JER scale factor provided by JME



Separating Measurement from Interpretation.



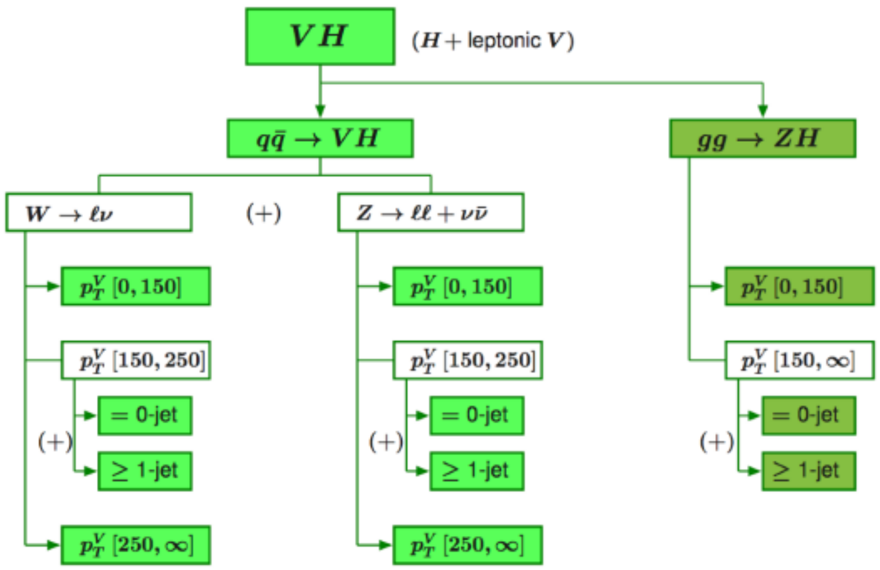
Talk at the 12th LHC Higgs XSWG
Oct. 2016

Goals

- Minimize theory systematics in measurements
  - Clearer and systematically improvable treatment at interpretation level
- Minimize model dependence in measurements
  - Decouples measurements from assumption of underlying physics model (SM, (non)linear EFT, BSM models)
- Measurements stay long-term useful
- Allows easy further (re)interpretation with different theory inputs/assumptions
  - Improved theory predictions/uncertainties
  - μ_i, κ_i, anomalous couplings, EFT coefficients, specific BSM scenarios

STXS for VH - short intro

- Stage-1 bin split mostly based on VH(bb) analysis categories / variables



- "VH" bins include leptonic VH (H undecayed)
- qq → V(qq)H as part of "VBF" bins
- gg → Z(qq)H as part of "ggF"

- Feedback on the bin split is still welcome, not set in stone!

STXS ≠ fiducial XS (and complementary)
[fid/diff XS minimize theory dependence and acceptance corrections, decayed Higgs, ...]

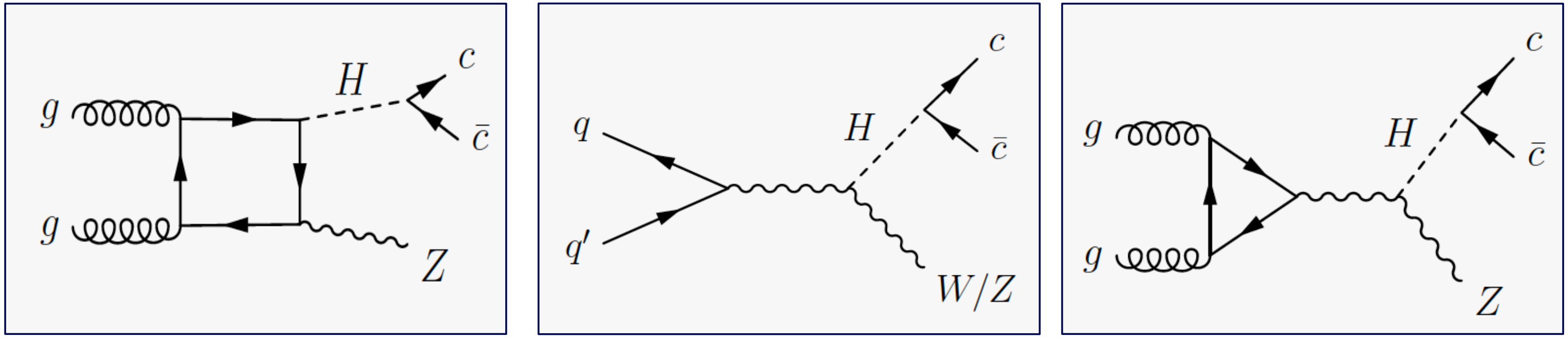
- optimized for analysis sensitivity (e.g. in this case driven by VH(bb) categorization)
- reducing dominant theory dependence in the measurement (by moving it to the interpretation stage)
- reduced residual theory uncertainties within the measurement of each bin (if residual th. uncertainties become large in the exp. acceptance for a bin, the bin the be further split in sub-categories)

(reference from LesHouches2017)

Talk at the VH LHC Higgs XSWG soubgroup Apr. 2018

Analysis Strategy - 2016 dataset

- **Signals targeted**



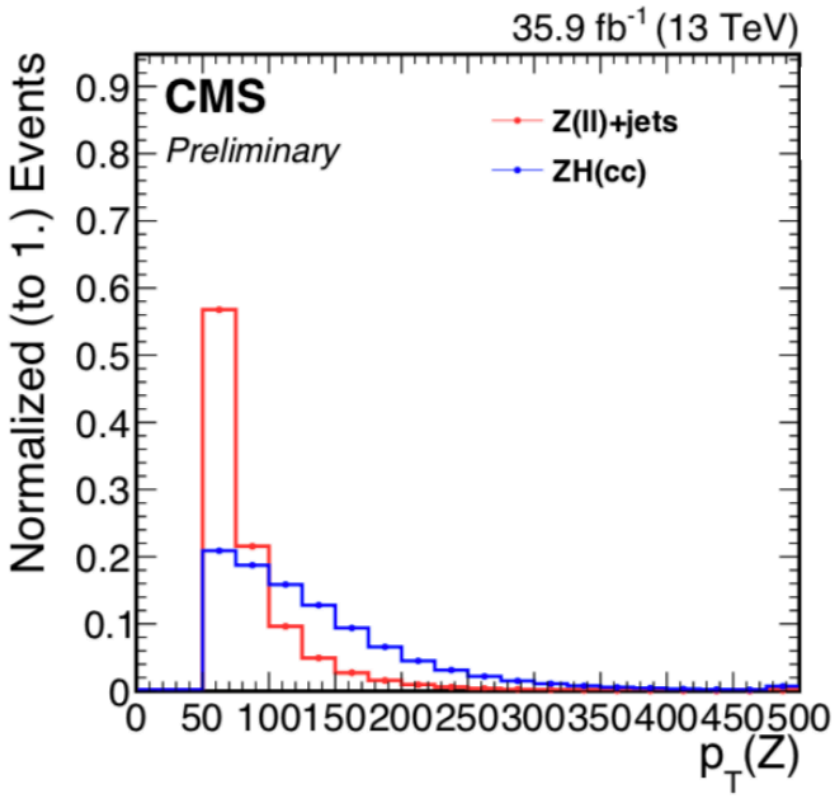
- **Main backgrounds:**
  - Z/W+jets, tt+jets, single-top

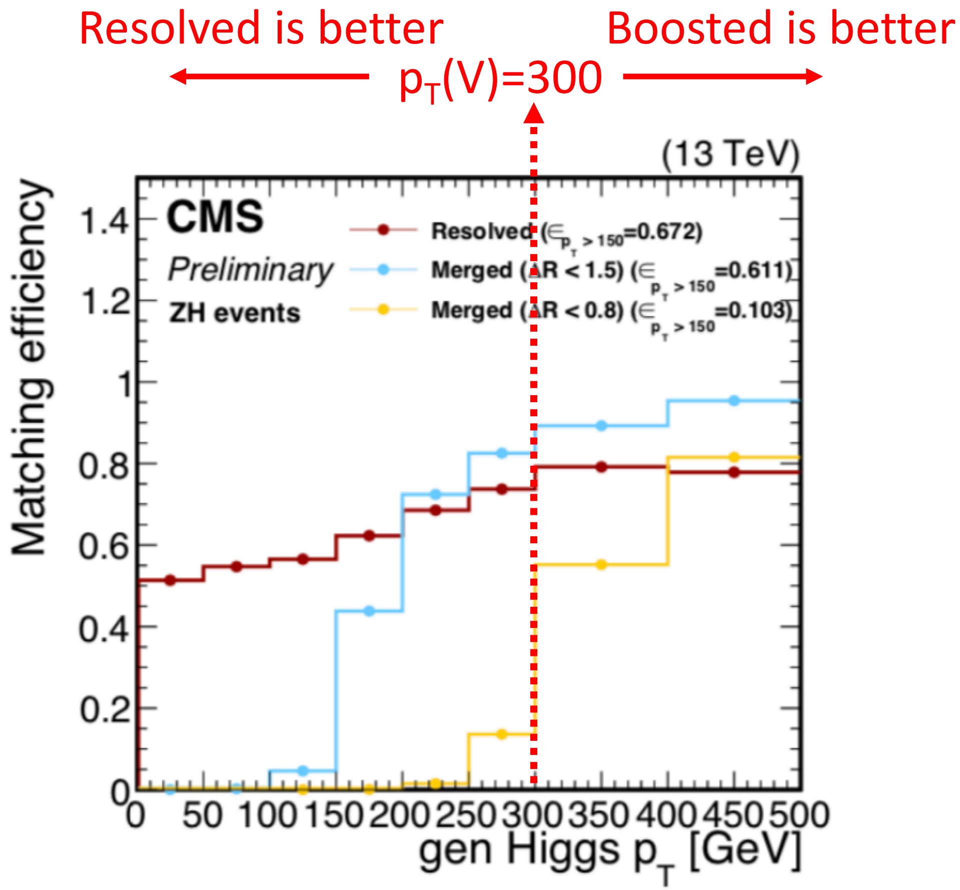
- **Vector bosons and Higgs boson reconstruction**
  - Same flavor lepton with pT > 20 GeV and 75 < m(Z) < 105 GeV
  - Single lepton with pT > 25 GeV, pT(W) > 100 GeV
  - PF MET > 170 GeV
  - RESOLVED: Reconstructed from the two leading CvsL jets + FSR AN-18-275
  - BOOSTED: Reconstructed from highest-score-fatJet AN-18-243

Analysis Strategy - 2016 dataset

- **Why two strategies?**
  - Quickly falling $p_T(V)$ spectrum of both signal and background
  - Around 200 GeV, similar efficiency of resolved and merged in AK15

➔ **Maximize analysis sensitivity**





Analysis Strategy – VH Boosted

- **Categorization of events**
  - According to the number of leptons in final state: 0-, 1- and 2-lepton category

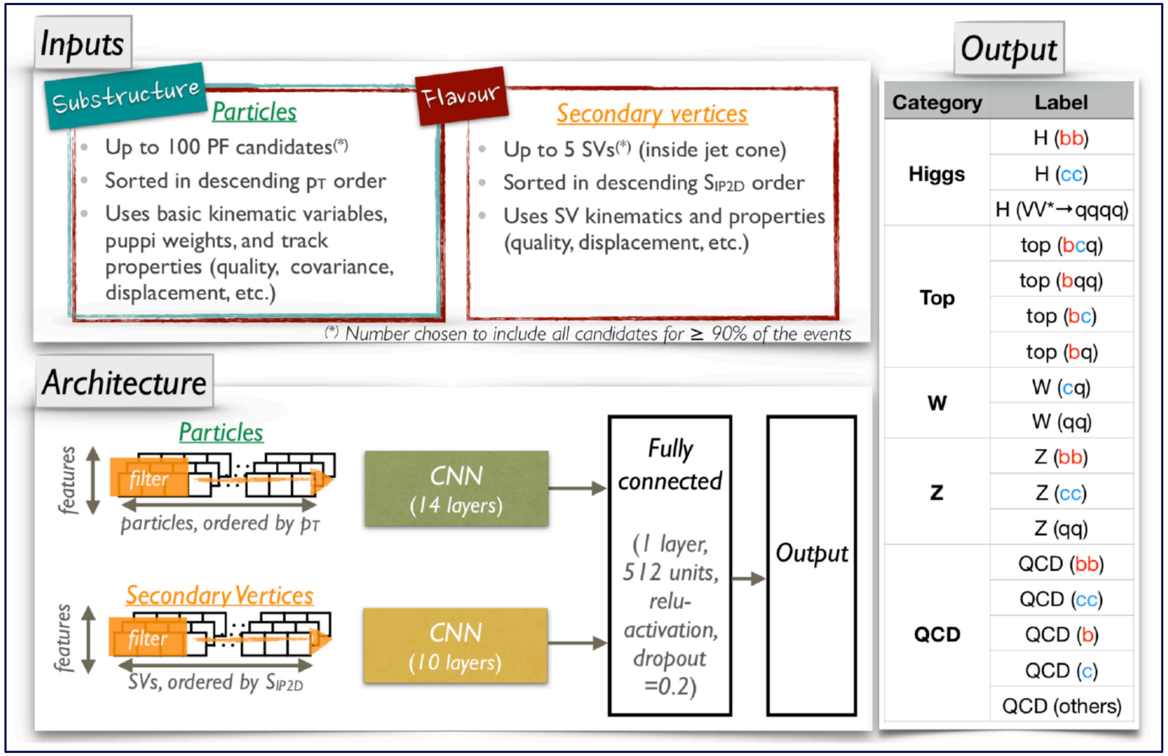
- **Further categorization according to charm-tagger score**
  - A further split into 3 more categories is performed based on the c-tagger score
  - Improve the sensitivity isolating regions with jets with higher c-tagger score

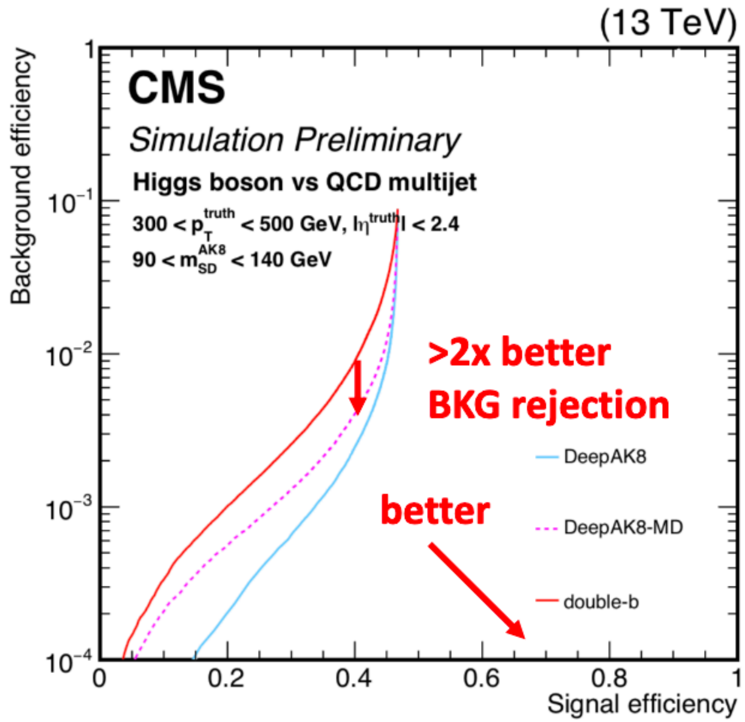
- **Signal region and control region definition**
  - A kinematic-BDT, orthogonal to charm tagger score, is trained
  - Signal and control regions are defined cutting on the Kinematic-BDT score

- **Final fit**
  - Binned max. lik. fit in all the categories/channels in CRs + SRs
  - The **fat-jet invariant mass** shape is fitted in the SRs and in the CRs

- **DeepAK15 tagger – cornerstone of the boosted VHcc analysis**
  - Reconstruction of moderately to largely boosted Higgs
  - DeepAK15: good compromise between signal purity and acceptance >200 GeV
- **Boosted jet tagger "DeepAK8" adapted on AK15 jets**
  - DNN multiclassifier for top, W, Z, Higgs, and QCD jets
  - Mass decorrelation techniques to mitigate mass sculptinG

CMS-DP-2017-049
NIPS 2017 paper,
CMS-JME-18-002





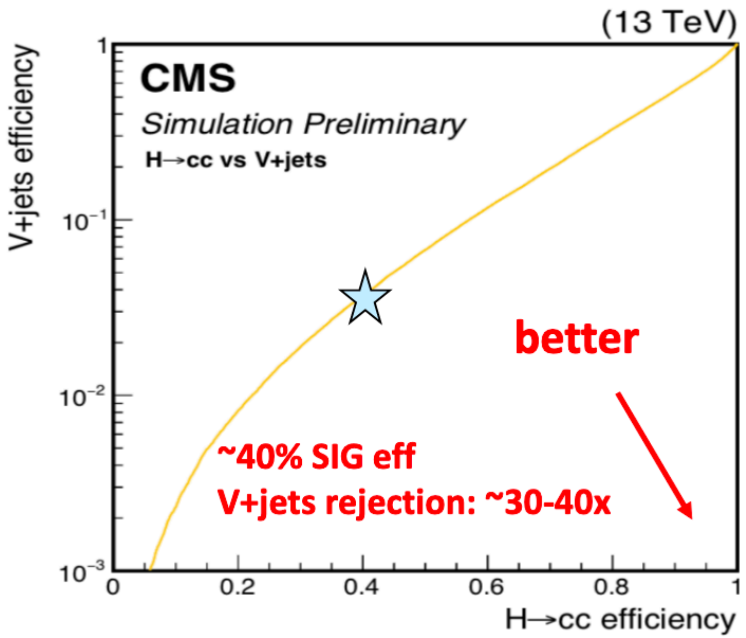
Heavy flavour tagger for AK15: DeepAK15

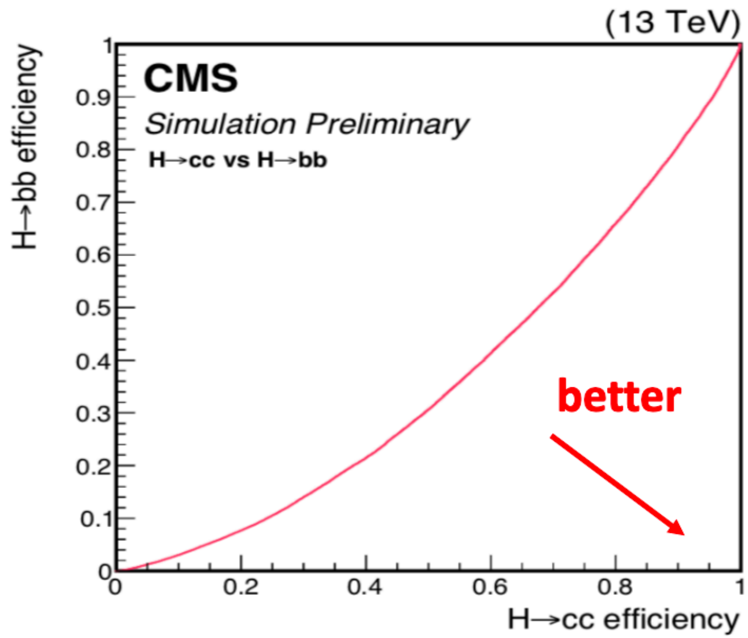
■ DeepAK15: DNN architecture

- cc –tagging discriminant defined as:

$$\frac{\mathrm{score}(Z \to c\bar{c}) + \mathrm{score}(H \to c\bar{c})}{\mathrm{score}(Z \to c\bar{c}) + \mathrm{score}(H \to c\bar{c}) + \mathrm{score}(QCD)}$$

- Performance evaluated with MC simulation

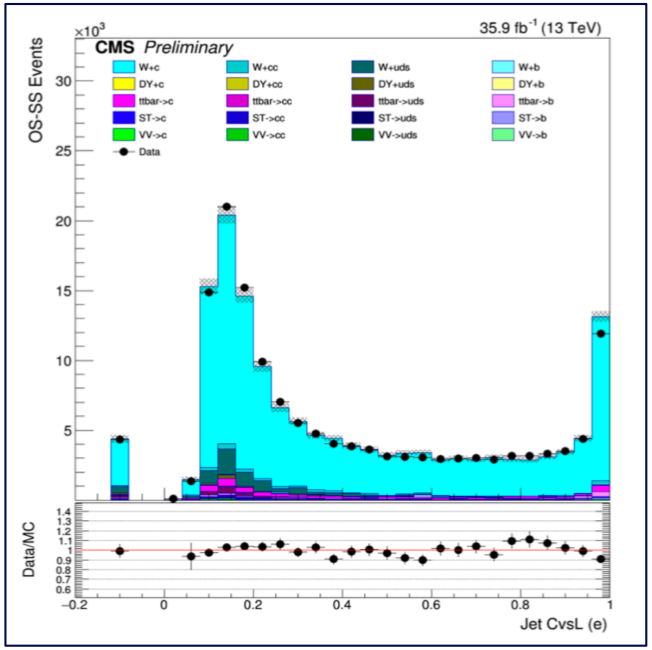


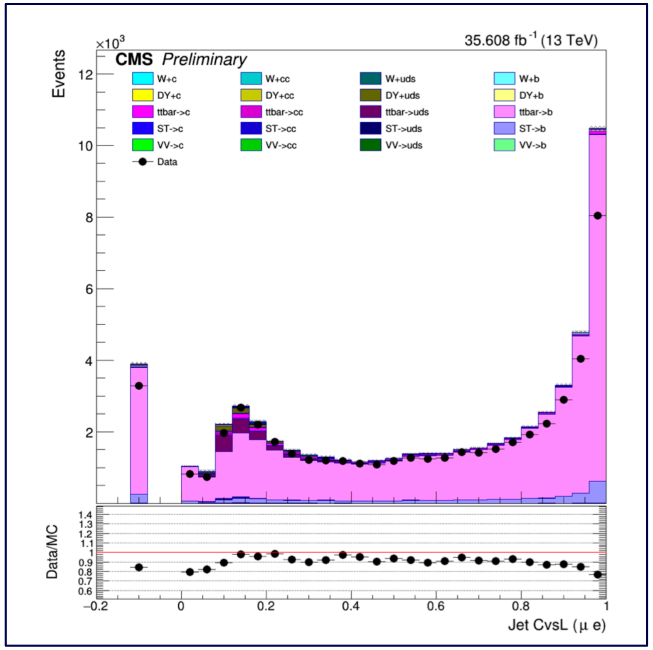


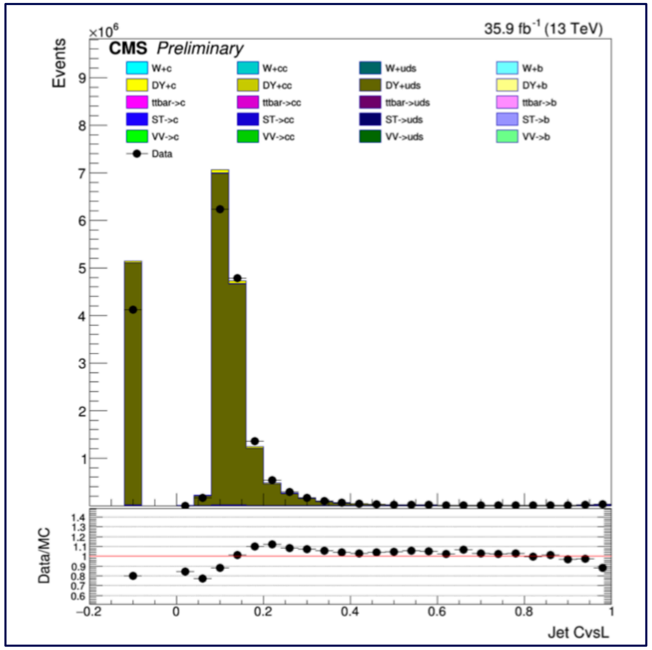
- Validation in data using proxy jets from g→cc

A new method to measure charm-tagger SFs

- <u>**Selections**</u>
  - **c-jet:** OS-SS events after W+jets is selected, looking to leptonic decay of the W boson and to the presence of a soft muon inside the jet
  - **b-jet:** Attempts have been made looking to semileptonic tt+jets (less pure) and to dileptonic tt+jets (~5x less statistics) ➔ at the end an inclusive region has been considered
  - **light-jet:** leading jet in a DY+jets(Z→$\mu\mu$) selections







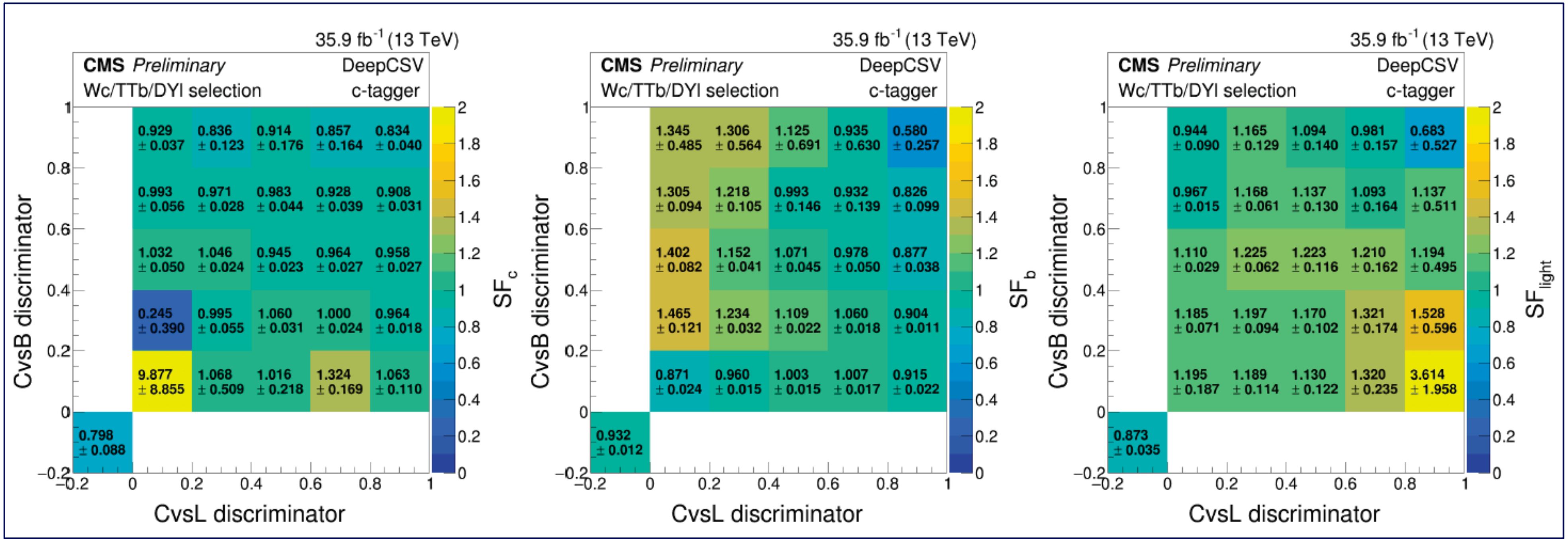
Scale Factor for CvsL/CvsB

■ Reshaping scale factor central values

- Events with CvsL and CvsB = -1 are considered in the normalization and in the fit
- The central values have been then used to define an event-by-event weight
- Such a weight is finally used to reshape the tagger distribution

$$w_i = \prod_{i=1}^{jets} sf_i(CvsL, CvsB)$$

Errors account for both statistical and systematics uncertainties

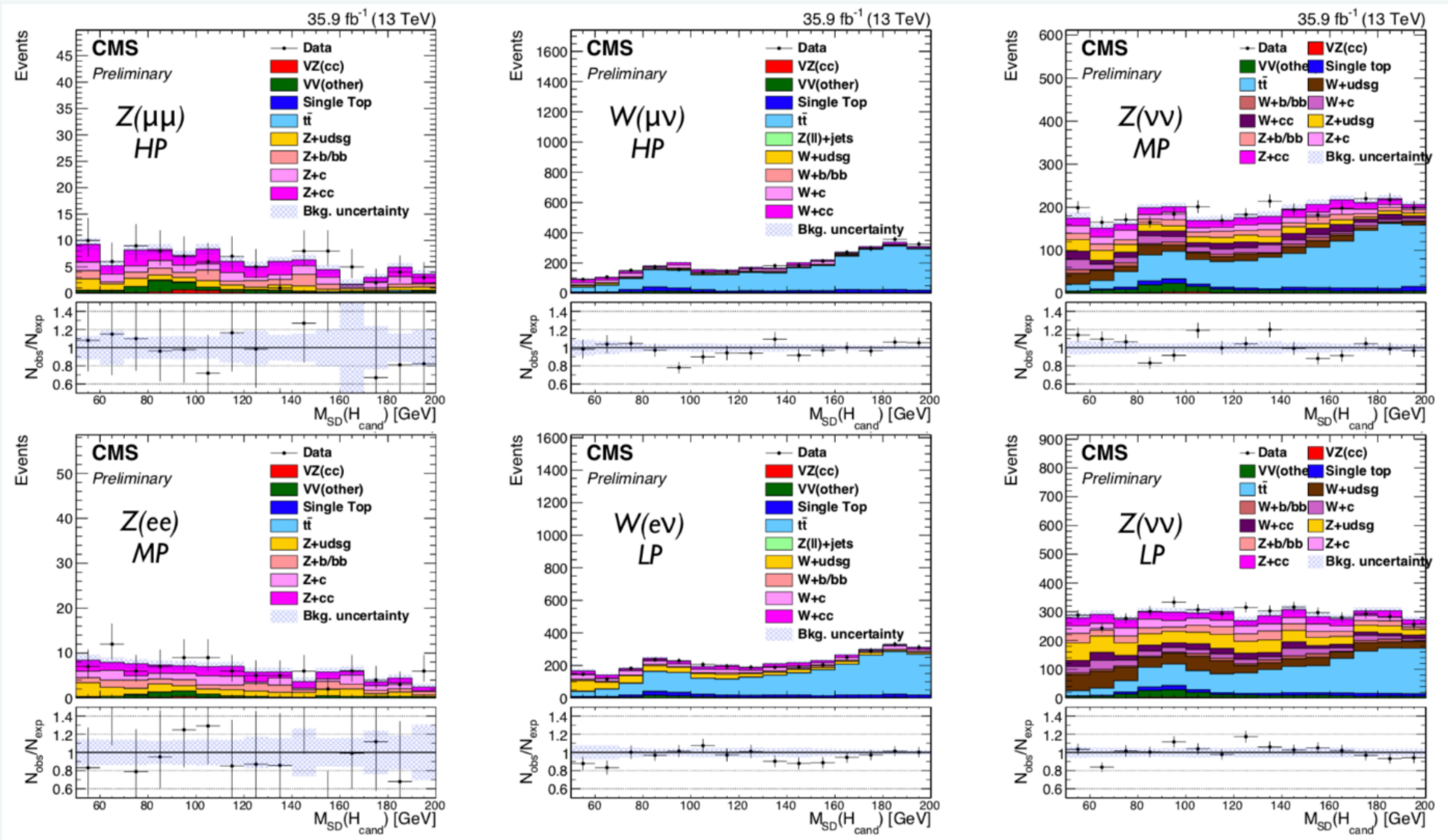


Scale Factor for CvsL/CvsB

- **Systematics considered in the scale factor derivation:**
  - Lepton ID/Iso
  - Pile-Up weight
  - Renormalization and factorization scale
  - Inclusive JES
  - JER
  - Cross-sections up/down variation (assumed fully uncorrelated among the processes)
  - MC statistics
  - Data statistics

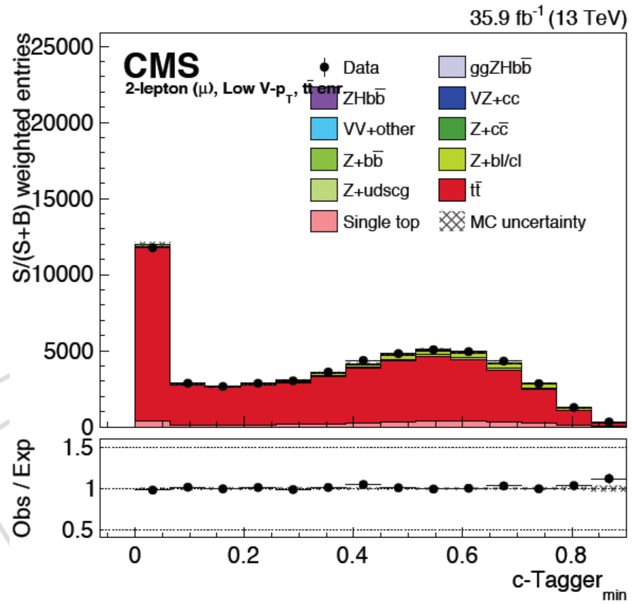
- **Documentation:**
  - SFs have been approved by BTV
  - The whole method is fully detailed in AN-19-028

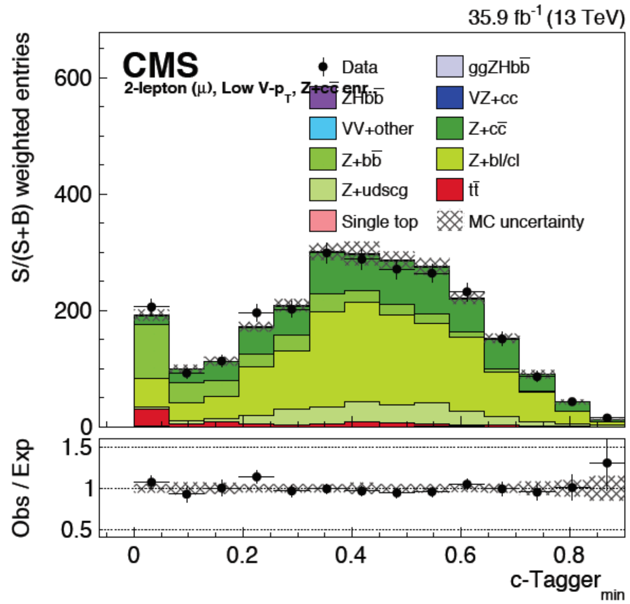
Control Region – VH boosted

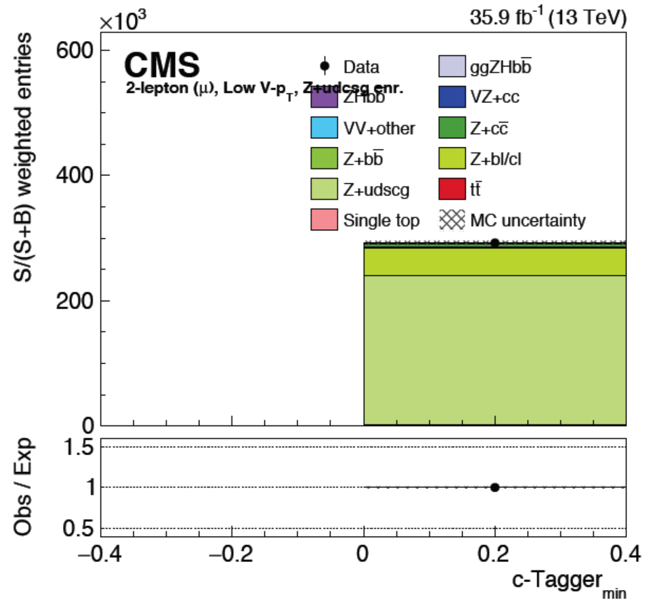


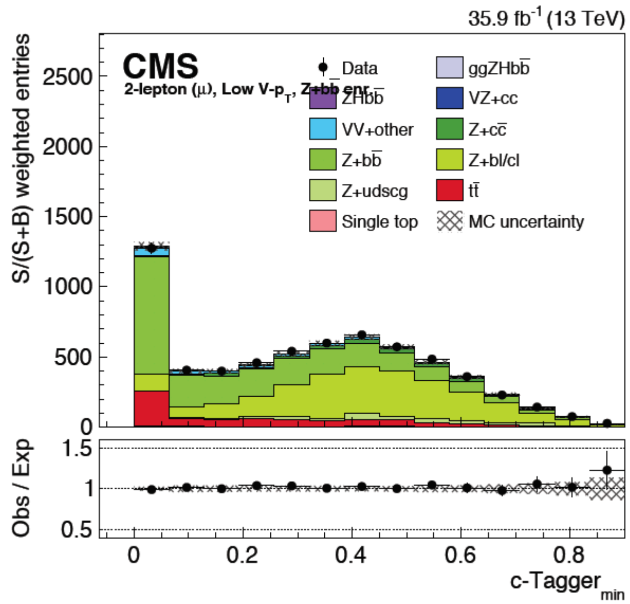
Control Region – VH resolved

2-lepton
Low-$p_T(V)$

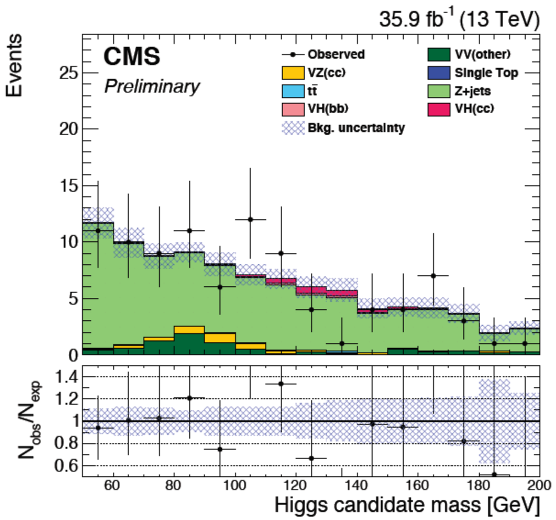


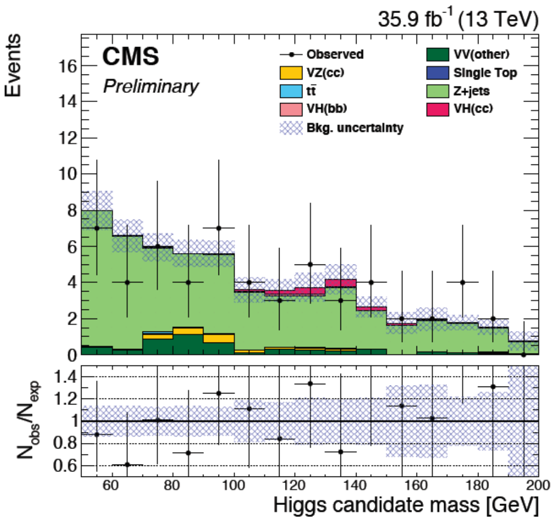


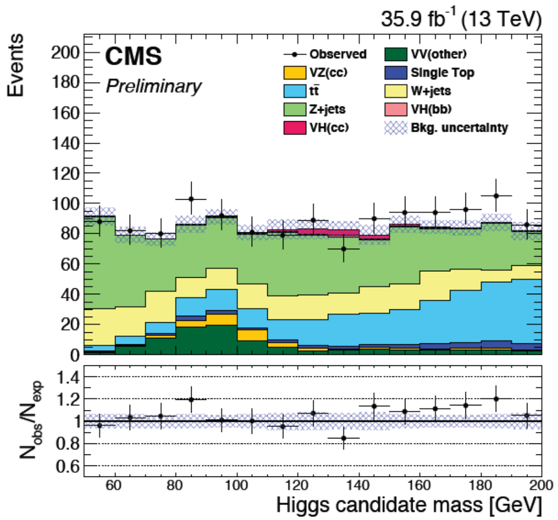


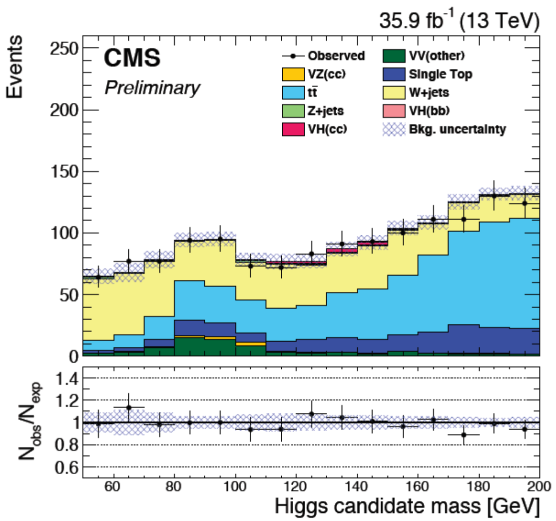


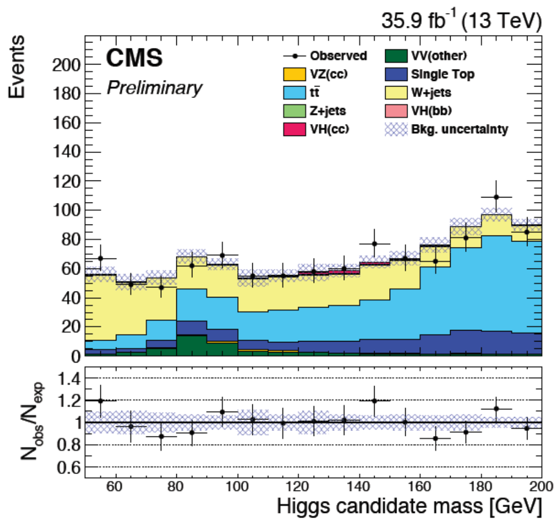
Signal Region – VH resolved



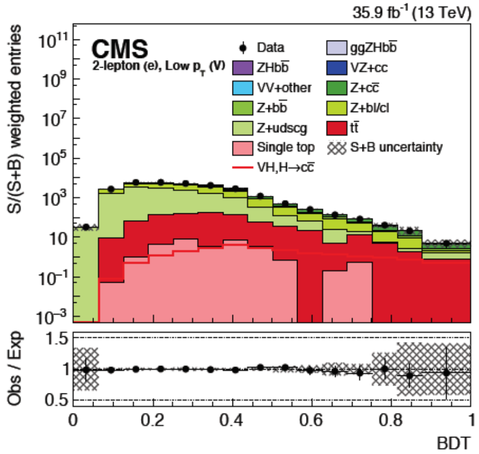


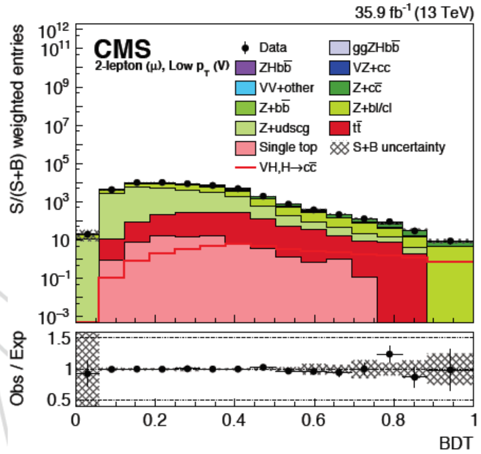


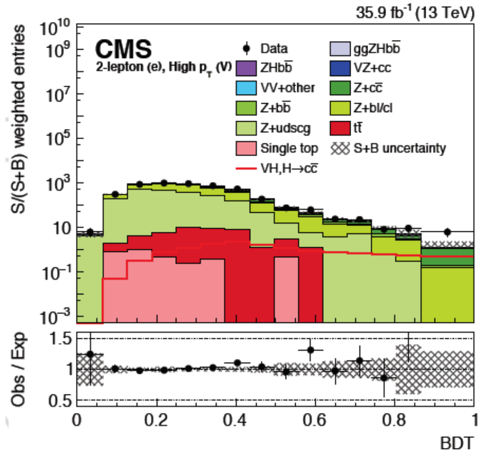


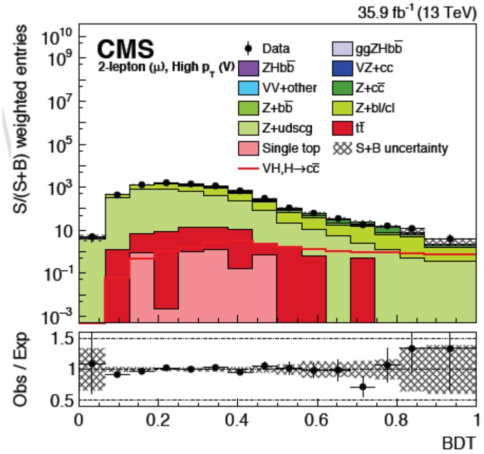


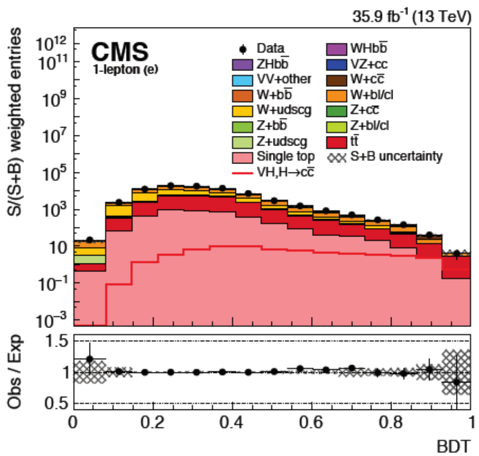
Signal Region – VH resolved

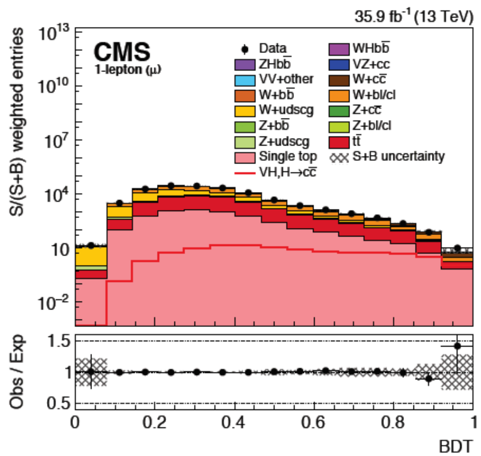


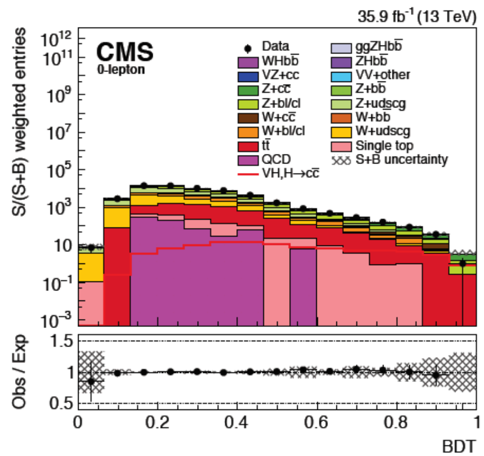






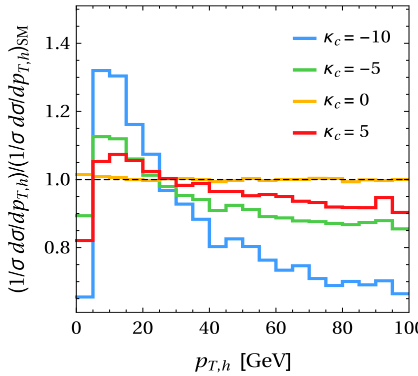






Looking forward…

- **Very challenging channel… lead improve analysis techniques**
  - possibility to improve many tools, e.g. c-taggers
  - Z→cc analysis possible with ~0 changes to the H→cc analysis:
    - Targeting VZ(Z→cc) evidence with full Run-2 (would be 1st time at had. coll.)
- **Full Run-2 "rule of thumb" prediction:**
  - Lumi. 2016:2017:2018=36:41:80 + assuming 1./Exp. L scale in quadrature
  - Assuming no improvement in the analysis neither on the c-taggers side
    - Projection on 95% CL. Exp. Limit on $\mu \sim 18$
  - Working also on ggH(H→cc): possibility to combine
  - With full Run-2, sensitivity to H→cc can be in the O(sensitivity on HH)
- **H→cc as a probe for new physics**
  - Potentially sensitive to BSM modification to H-charm coupling



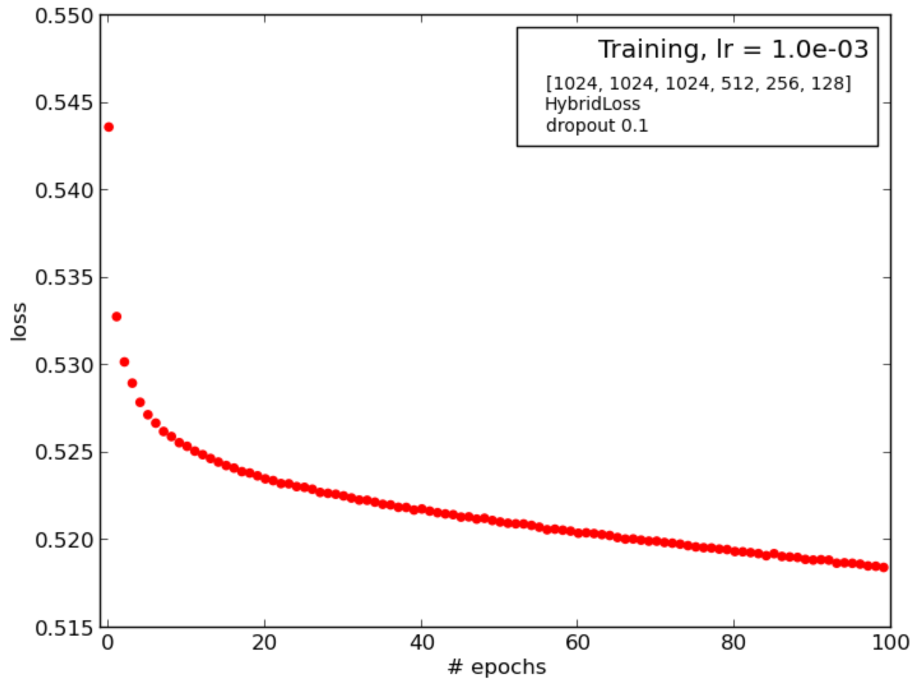
VHcc: What's next?

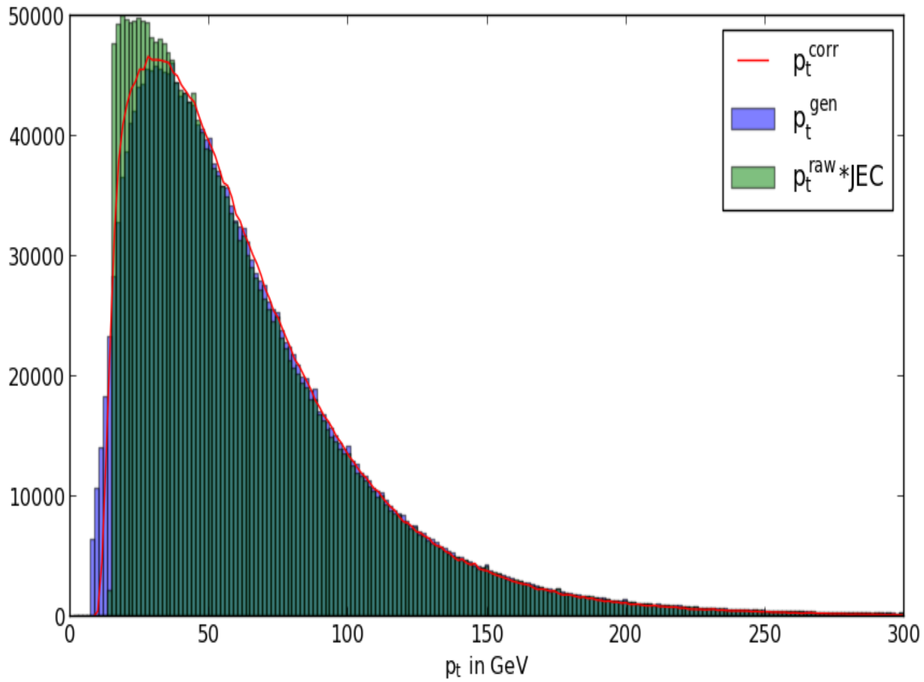
- **Possible improvement to the current analysis**
  - C-jet energy regression (work in progress)
  - Kinematic fit in the 2-lepton categories
  - Study what's the gain in deploying DeepJet
  - Add 2017 and 2018 dataset
- **Possible benefits from interplay with VHbb**
  - Fit simultaneously VHbb and VHcc
  - How correlate the systematics?
    - Different flavour splitting for V+jets → different rate parameters
  - Open discussion…

▪ DNN architecture and training

NEW

- Relying on the ETH training for b-jet energy regression (thanks Nadya!)
- The regression is trained on 2.3 millon of c-jets from hadronic tt+jets
- Preselections: p_T>15GeV && 1 GeV<gen-p_T<6 TeV && $|\eta|$<2.5
- DNN Input variables same as in b-jet energy regression
- Trainig with a batch size of 1024. This NN I have trained over 100 epochs

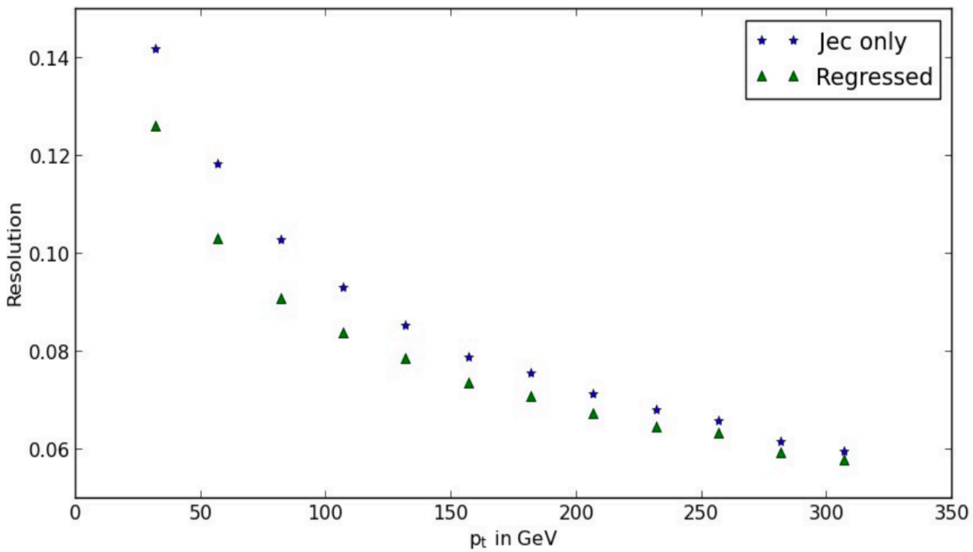


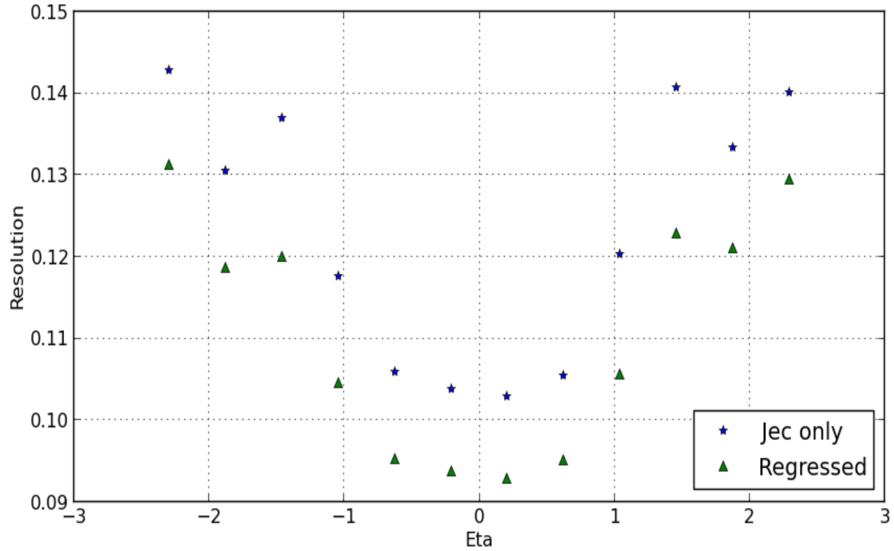


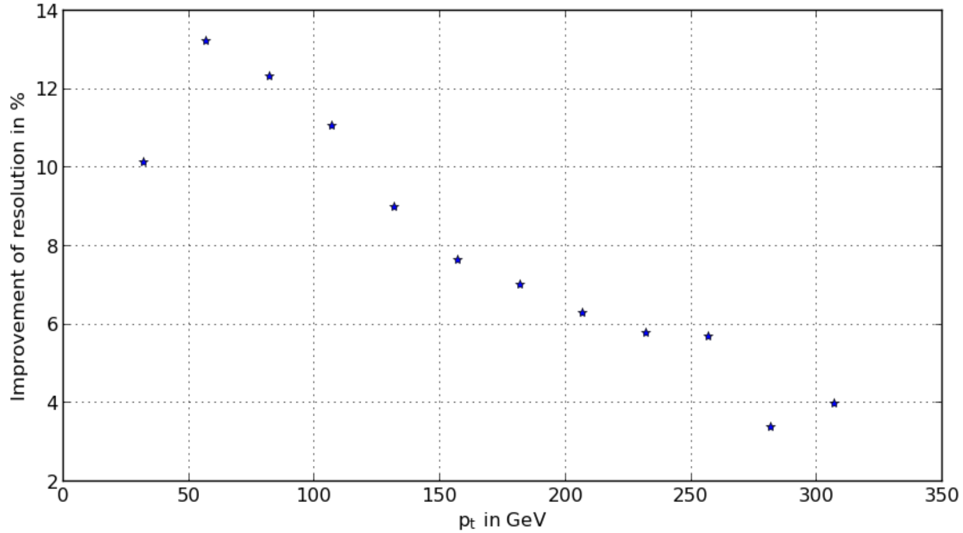
VHcc: c-jet energy regression

Preliminary Performance

NEW







<u>Improvement on single jet energy resolution:</u>

- >10% in pT range [30, 120]
- 5%-10% in pT range [120, 250]

Looking forward to assess improvements on mjj