A boost to Higgs Physics: new regimes at high energy

Silvia Biondi

University & INFN of Bologna

silvia.biondi@cern.ch / silvia.biondi@bo.infn.it

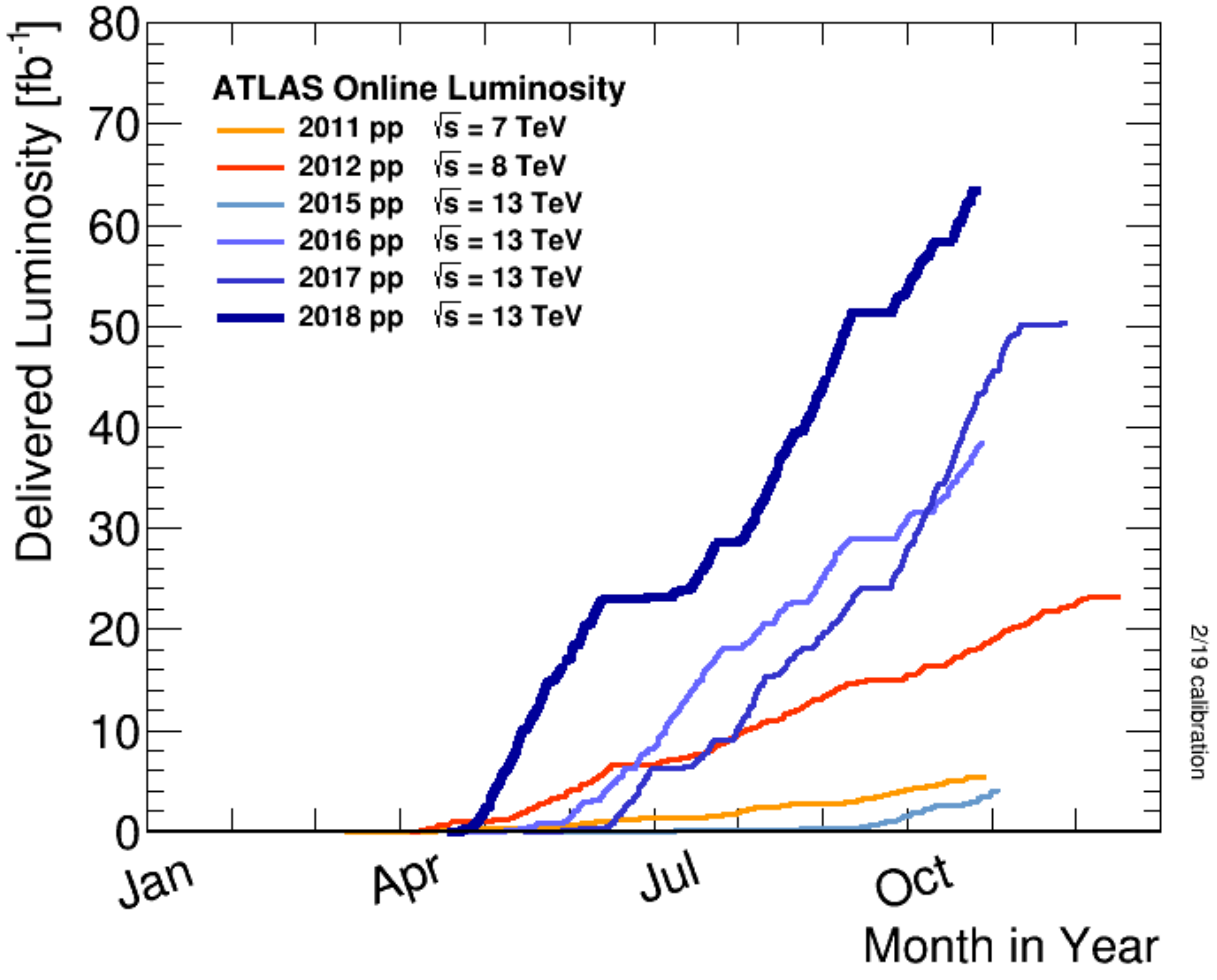
Corso di Dottorato - 2019/2020

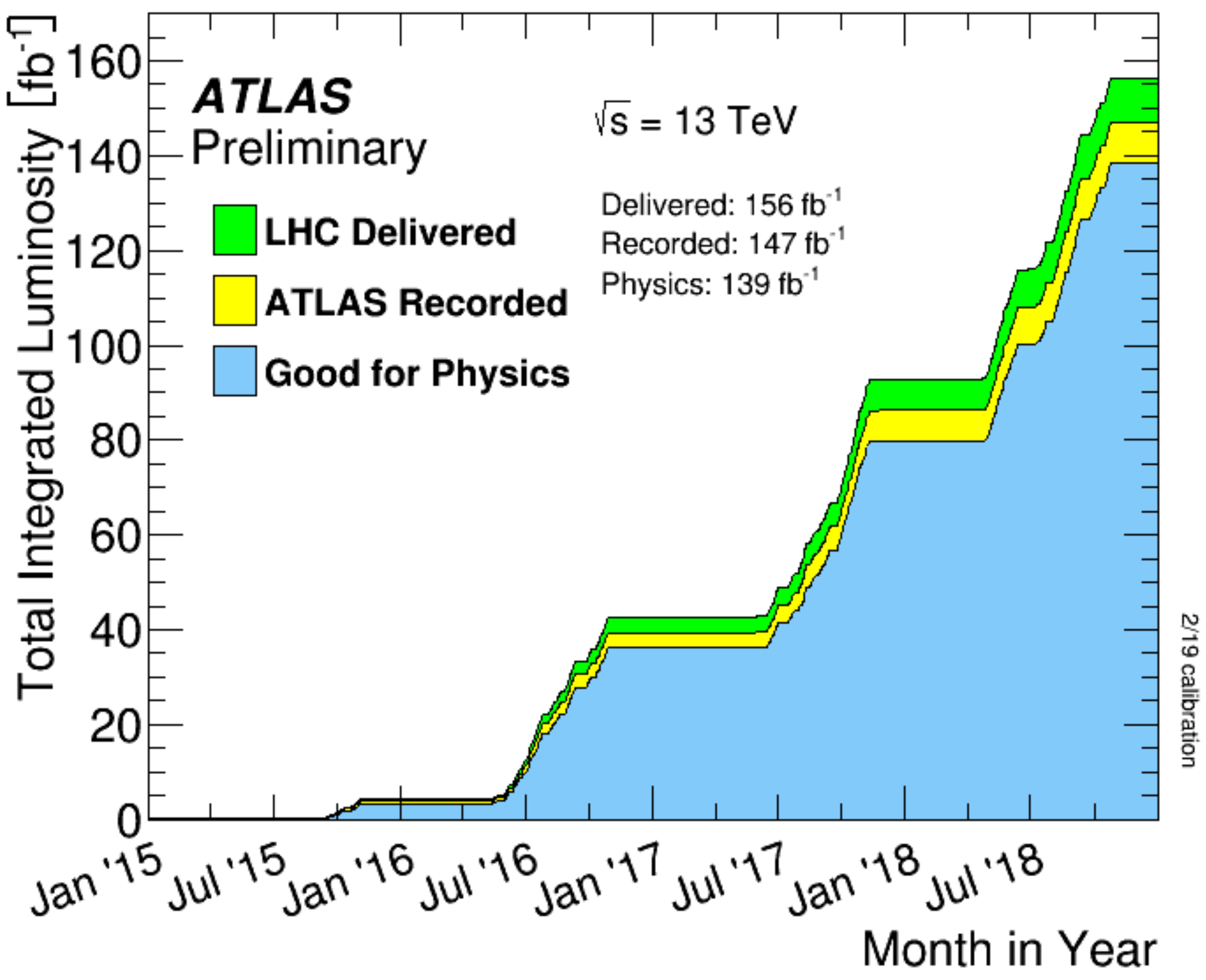
Course outline

- Theory reminder
- Higgs boson production and decay modes
- Higgs boson discovery by ATLAS and CMS
- Higgs boson mass measurement by ATLAS and CMS
- Overview of ATLAS and CMS analyses about Higgs
- Signal/background discrimination techniques
  - boosted regimes
    - tagging, large-radius jets substructure, re-clustering
  - multivariate analysis and deep neural network
- Signal extraction techniques
  - likelihood and test statistic
  - CLs method
- **ttH analysis: an example**

ttH analysis: an example

Integrated luminosity



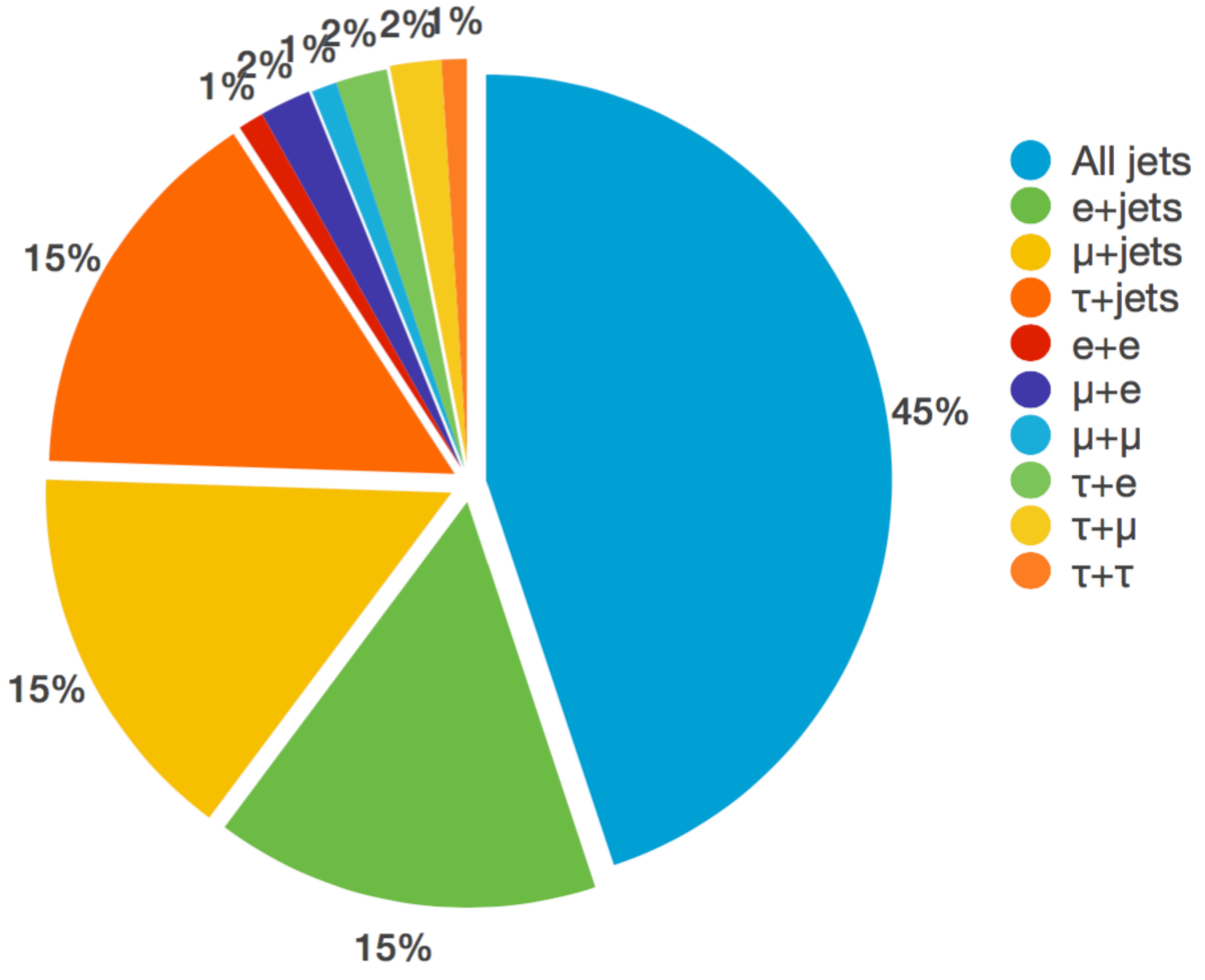


- highest luminosity ever reached in few months;
- 2015 + 2016 + 2017 + 2018 = **Run 2**
- 139 fb^{-1} used in the analyses.

- **delivered luminosity** from the start of stable beams until ATLAS goes to standby mode for the beam dump;
- **recorded luminosity** reflects the data acquisition inefficiency;
- **good for physics** reflects the criteria applied to ensure the quality of data for analyses.

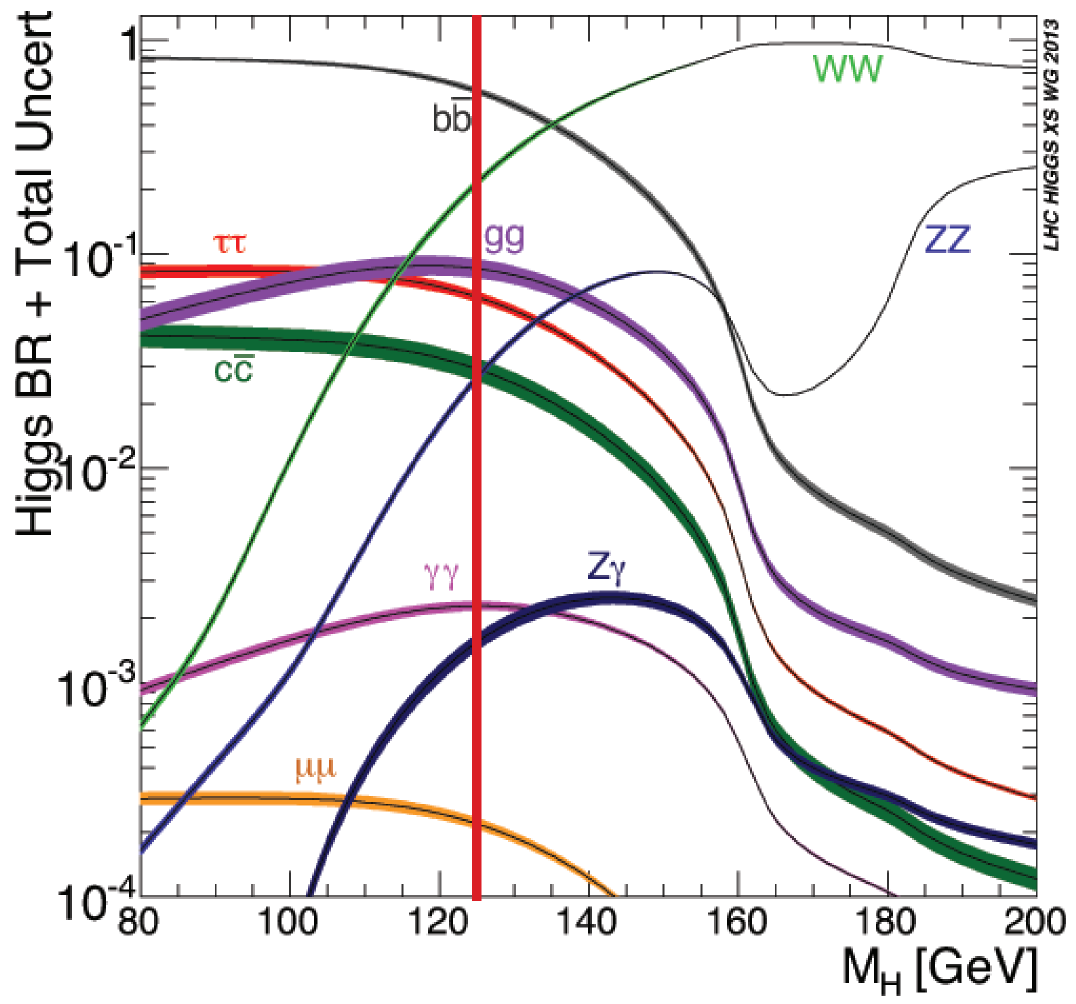
The ttH (H→bb) channel

BR (tt->lνqq) ~35%



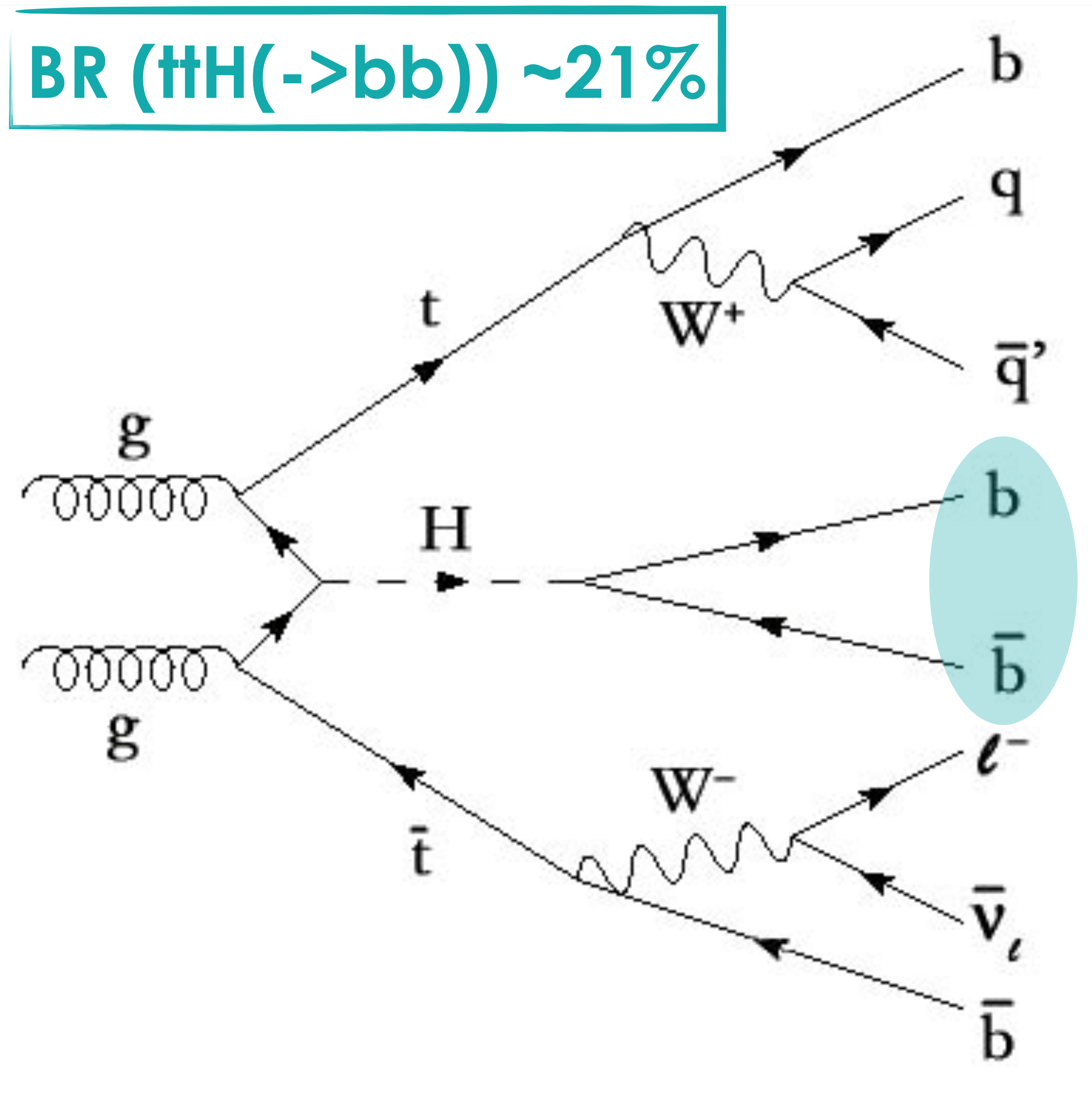
+

BR (H->bb) ~60%



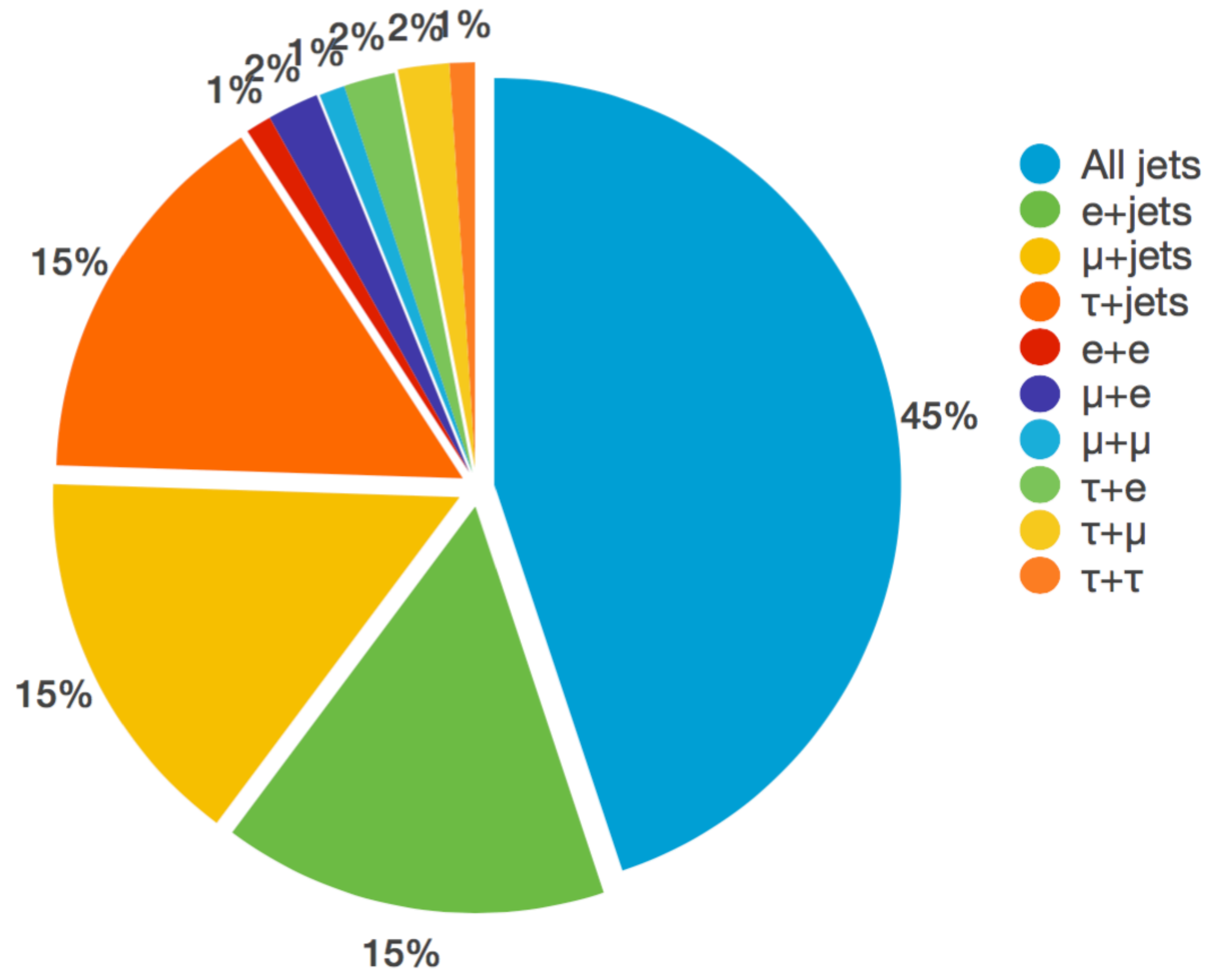
=

BR (ttH(->bb)) ~21%



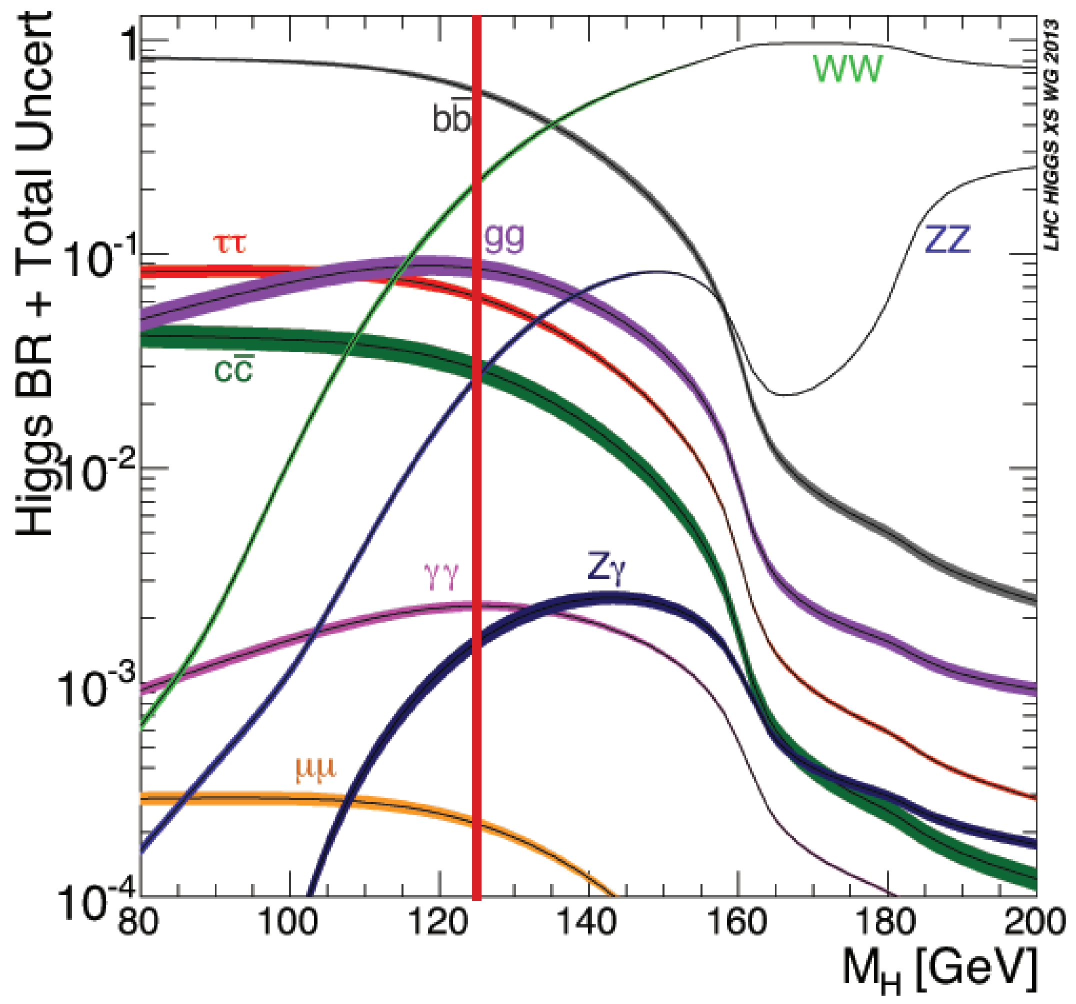
The ttH (H→bb) channel

BR (tt->lvqq) ~35%



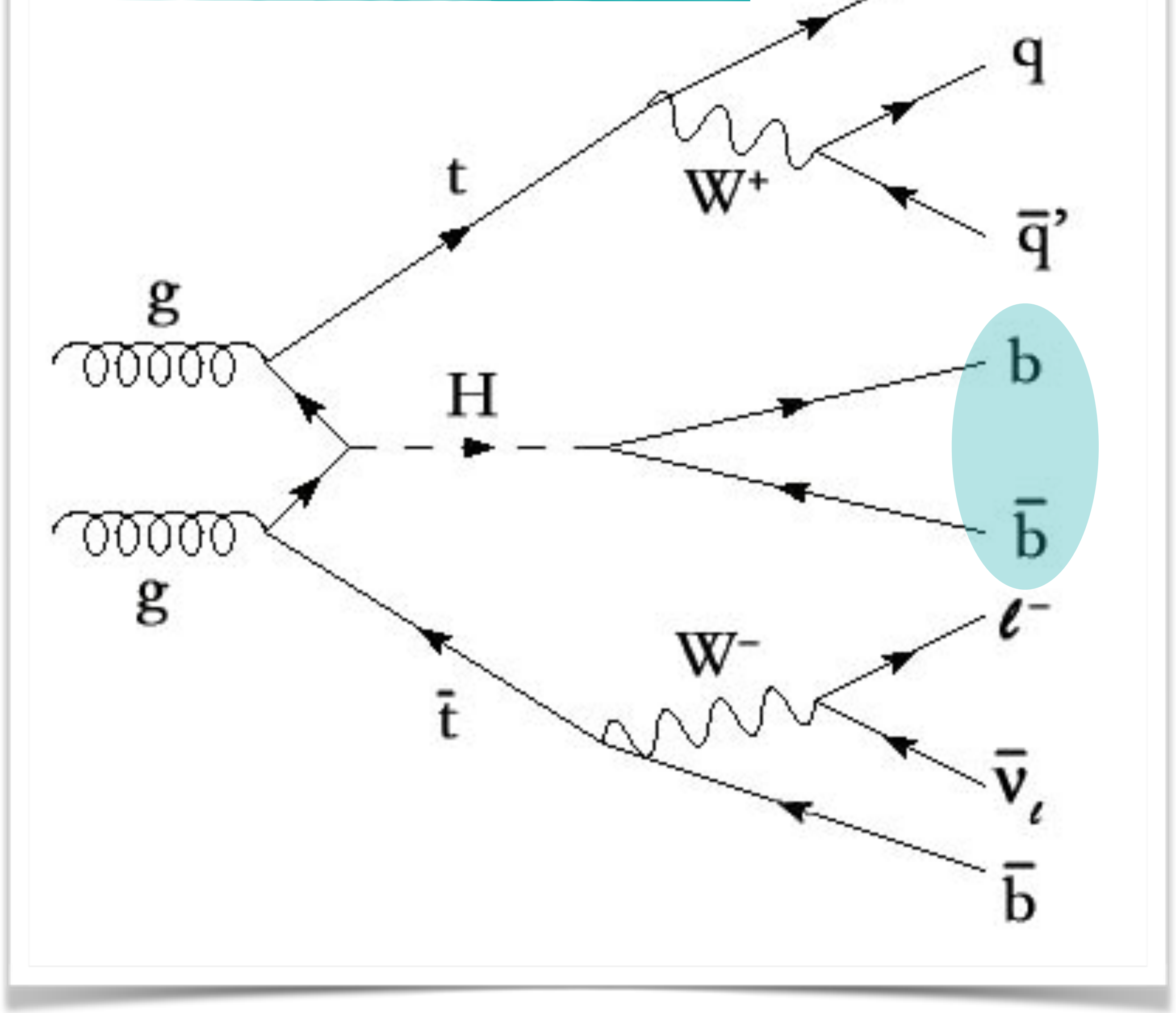
+

BR (H->bb) ~60%



=

BR (ttH(->bb)) ~21%



σ(pp->**tt**) ~ 8x10³ pb

80 events each billion

NB! the main background contribution to our channel

σ(pp->**ttH**) ~ 0.5 pb

~1% of total Higgs production only!

σ(pp->**HX**) ~ 45 pb

1 event each 2 billions

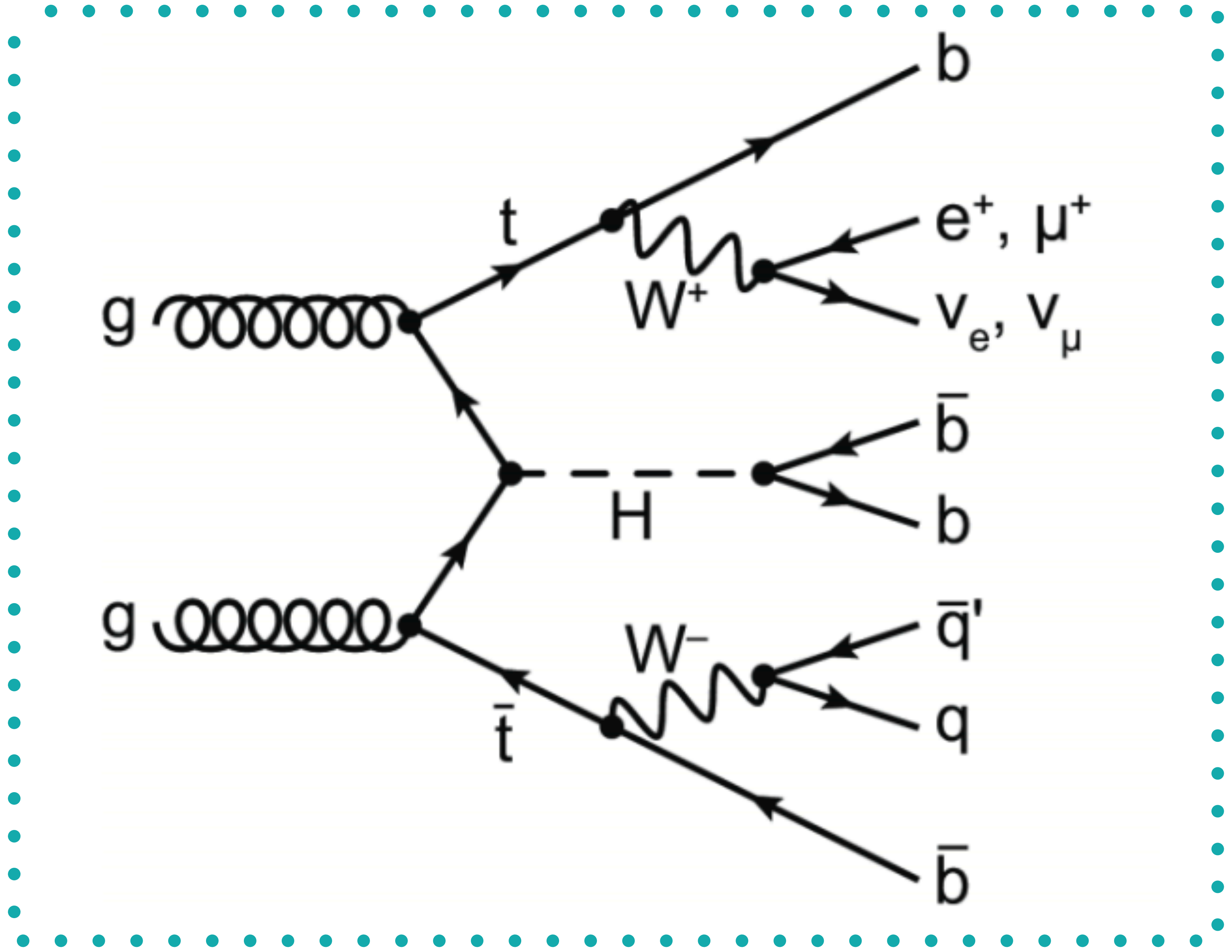
Rare event!

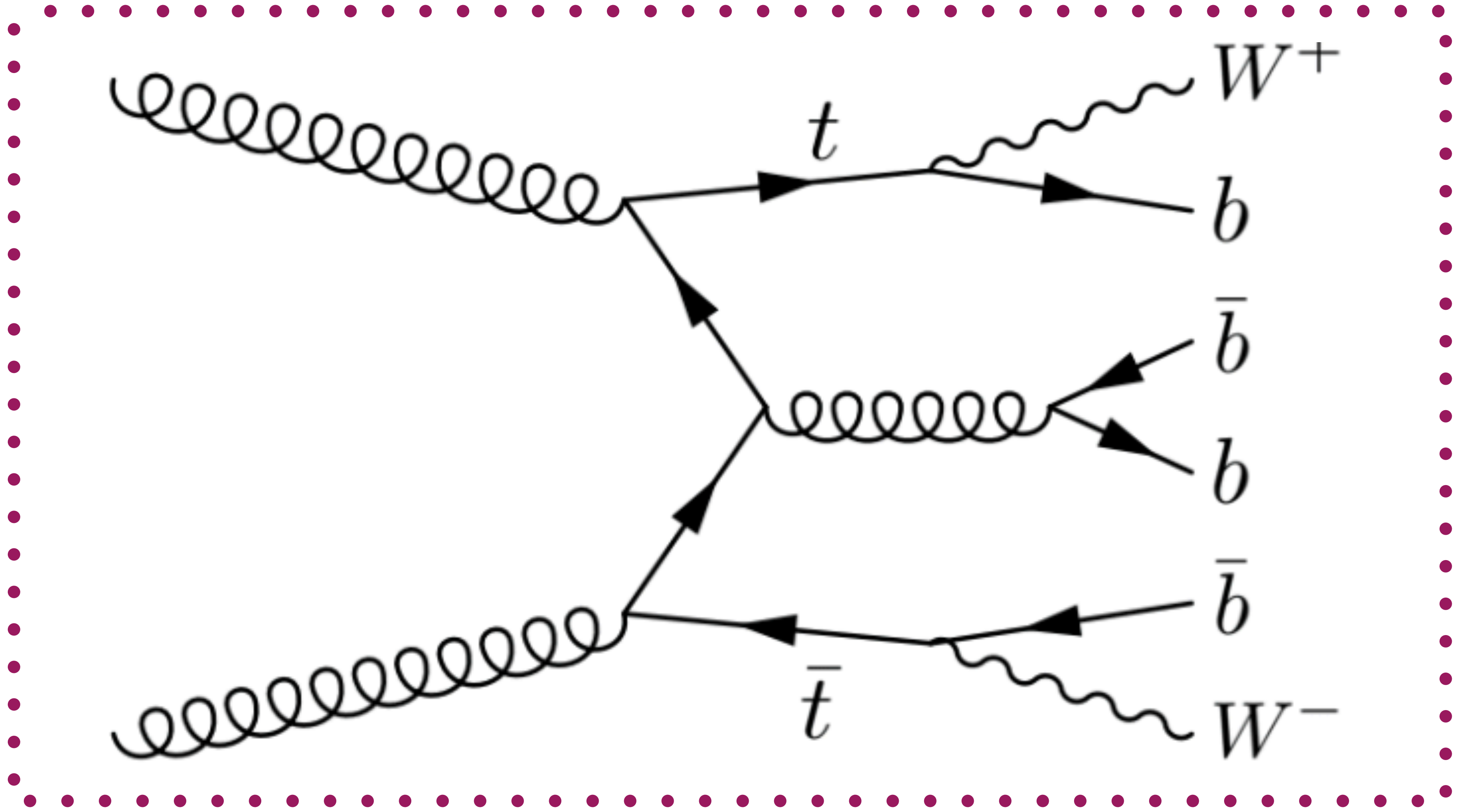
- **PROS**: highest BR, help against the extremely low production xsec, lepton helps to detect the event;
- **CONS**: combinatorics background, due to the large number of jets in the final state.

NB!

ttH process observed only in 2017!

The ttH (H→bb) channel





- **Characteristics:**
  - **0/1/2 leptons** (e, mu) from W-decays;
  - at least **8/6/4 jets**;
  - at least **4 b-jets** from top and Higgs decays;
  - Higgs boson reconstruction possible but challenging due to the combinatorics.

- **Major challenge:**
  - **irreducible tt+bb background** has large theory uncertainty in leptonic channels;
  - **QCD multijet background** in all-hadronic channel.

ttH analysis - Objects definition

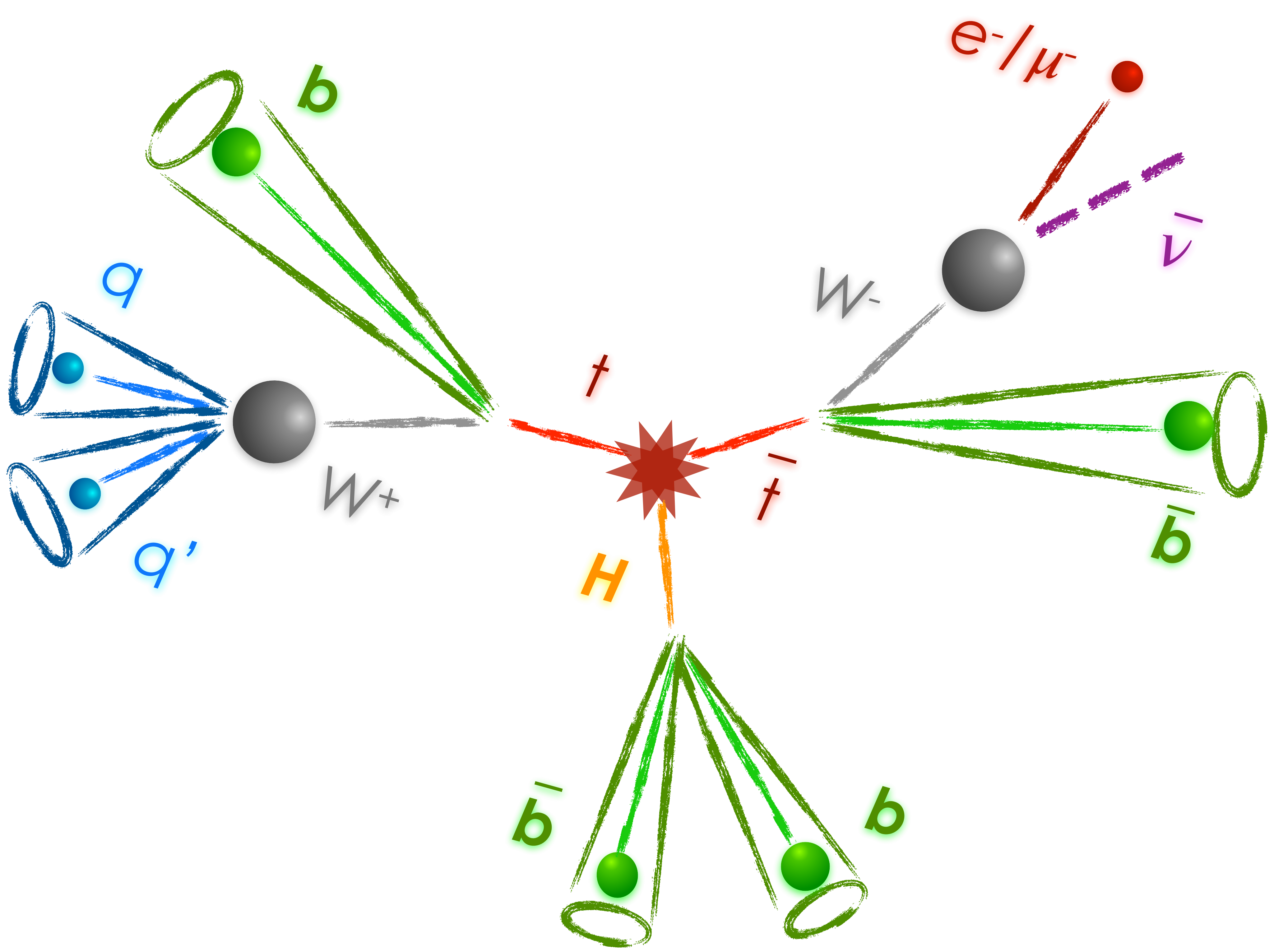
- **Identification: to determine if reco candidates are signal-like or background-like objects**
  - LH discriminant: observables related to the shower shape in the calo and to the track matching the electromagnetic cluster (for **electrons**);
  - algorithms matching tracks from ID and MS (for **muons**) or from ID and topological clusters in calo (for **jets**);
  - track multiplicity and a multivariate discriminant based on the track collimation, further jet substructure, and kinematic information (for **taus**).

- **Isolation: to disentangle prompt objects from others**
  - algorithms using angular and clusters informations between object which tells how much an object is isolated from all the others close to it and how much it matches with the **primary vertex**;

- **Quality: to further discriminate signal from background**
  - requirements on tracks energy, distance from primary vertex, number of tracks and signals in the detectors.

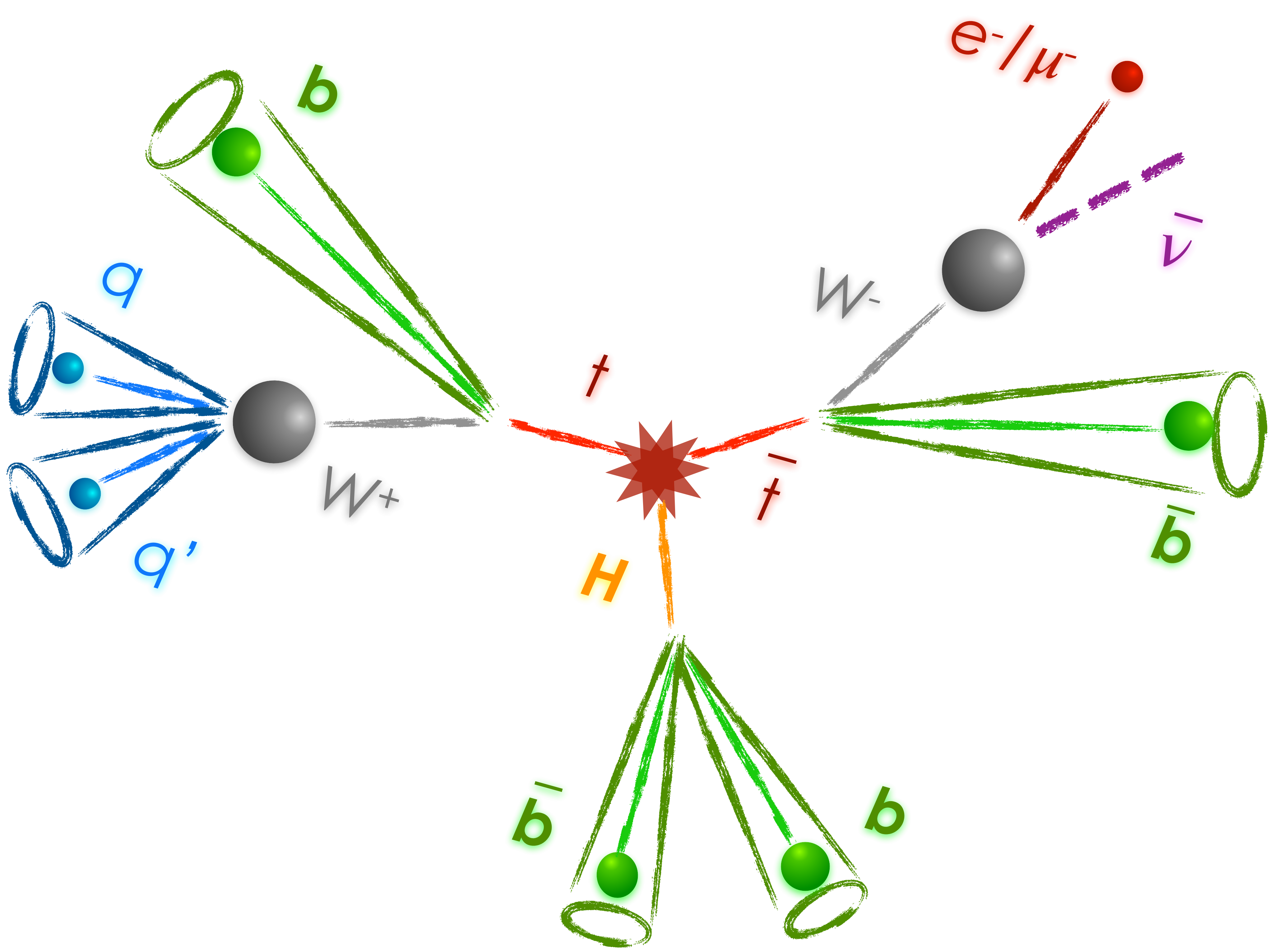
	Leading p_T	Subleading p_T	η	Isolation	Quality
Electron	27 GeV	10 GeV	2.47	Gradient	TightLH
Muon			2.5	FCTightTrackOnly	Medium
Jet	(EMTopo) 25 GeV		2.5	JVT	
Tau	25 GeV		2.5	JetID : Medium, EleID : Loose	

ttH analysis - Resolved vs boosted

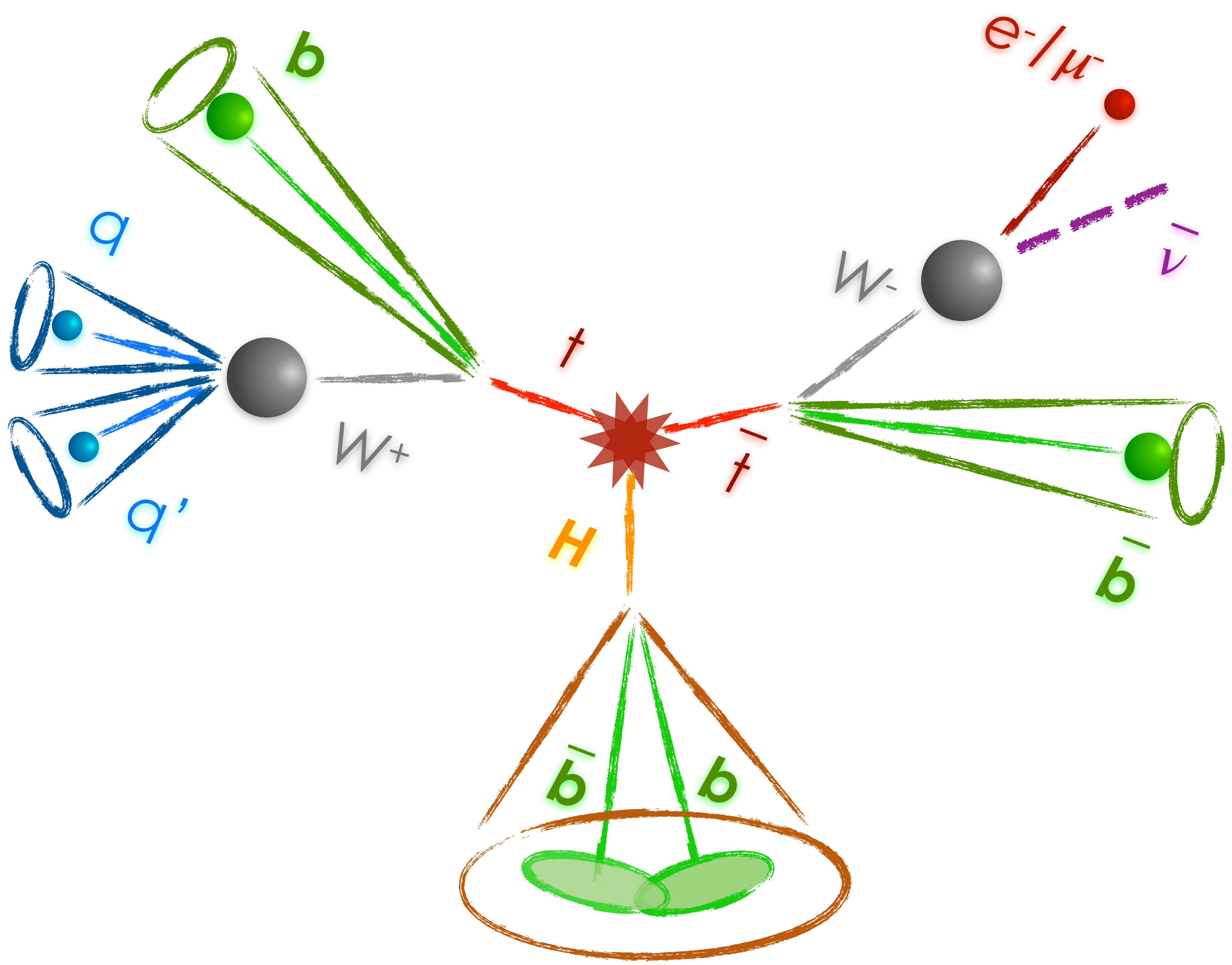


- 6 well separated small-R jets ("**resolved**"):
  - standard jet reconstruction algorithms (anti-k_t Δ**R < 0.4**);
  - significant combinatorial background.

ttH analysis - Resolved vs boosted



- 6 well separated small-R jets ("**resolved**"):
  - standard jet reconstruction algorithms (anti-k_t $\Delta R < 0.4$);
  - significant combinatorial background.



- at least 1 reclustered large-R jet ("**boosted**"):
  - re-clustering techniques algorithm (anti-k_t $\Delta R < 1.0$);
  - high background contamination into large-R jet;
  - additional algorithms for jet grooming and identification.

ttH analysis - Event pre-selection

Resolved

Dilepton channels
- 2 leptons;
- at least 3 jets;
- at least 3 b-tagged jets @ 60% WP in 3-jet category;
  & at least 3 b-tagged jets @ 70% WP in 4-jet category;
- $m_{ll} > 15$ GeV;
- $m_{ll} < 83$ GeV or $m_{ll} > 99$ GeV.

Single-lepton channels
- 1 lepton;
- at least 5 jets;
- at least 4 b-tagged jets @ 70% WP.

Boosted

Single-lepton channel
- 1 lepton;
- at least 4 jets;
- at least 3 b-tagged jets @ 85% WP;
- at least 1 Higgs candidate*:
  - $p_T > 300$ GeV;
  - mass in [100,140) GeV;
  - exactly 2 b-tagged jets @85% WP associated;
  - P(H) > 0.6 .

* **re-clustered jet:**
anti-k_T small-R jets (R=0.4) used to re-cluster the large-R jets (R = 1.0, $p_T > 200$ GeV, $|\eta| < 2$, m > 50 GeV) in this analysis.

ttH analysis - ML techniques in all channels

technique	ML discriminant	goal
event selection in different channels	→	definition of training samples
intermediate multivariate discriminants	reconstruction BDT (resolved dilepton channel) reconstruction BDT and LHD (resolved single-lepton channel) DNN (boosted channels)	reconstruction of objects and discrimination of ttH wrt tt
final multivariate discriminants	classification BDTs	separation of signal from background and final discriminants for the profile likelihood fit

analysis result

☆ **Input variables**:
important to check the Data/MC agreement looking for any mis-modelling that can affect the performance of the ML technique.

ttH analysis - ML techniques in all channels

technique	ML discriminant	goal
event selection in different channels	→	definition of training samples
intermediate multivariate discriminants	reconstruction BDT (resolved dilepton channel) reconstruction BDT and LHD (resolved single-lepton channel) DNN (boosted channels)	reconstruction of objects and discrimination of ttH wrt tt
final multivariate discriminants	classification BDTs	separation of signal from background and final discriminants for the profile likelihood fit

analysis result

Input variables:
important to check the Data/MC agreement looking for any mis-modelling that can affect the performance of the ML technique.

ttH analysis - Event categorisation strategy

Event categorisation strategy

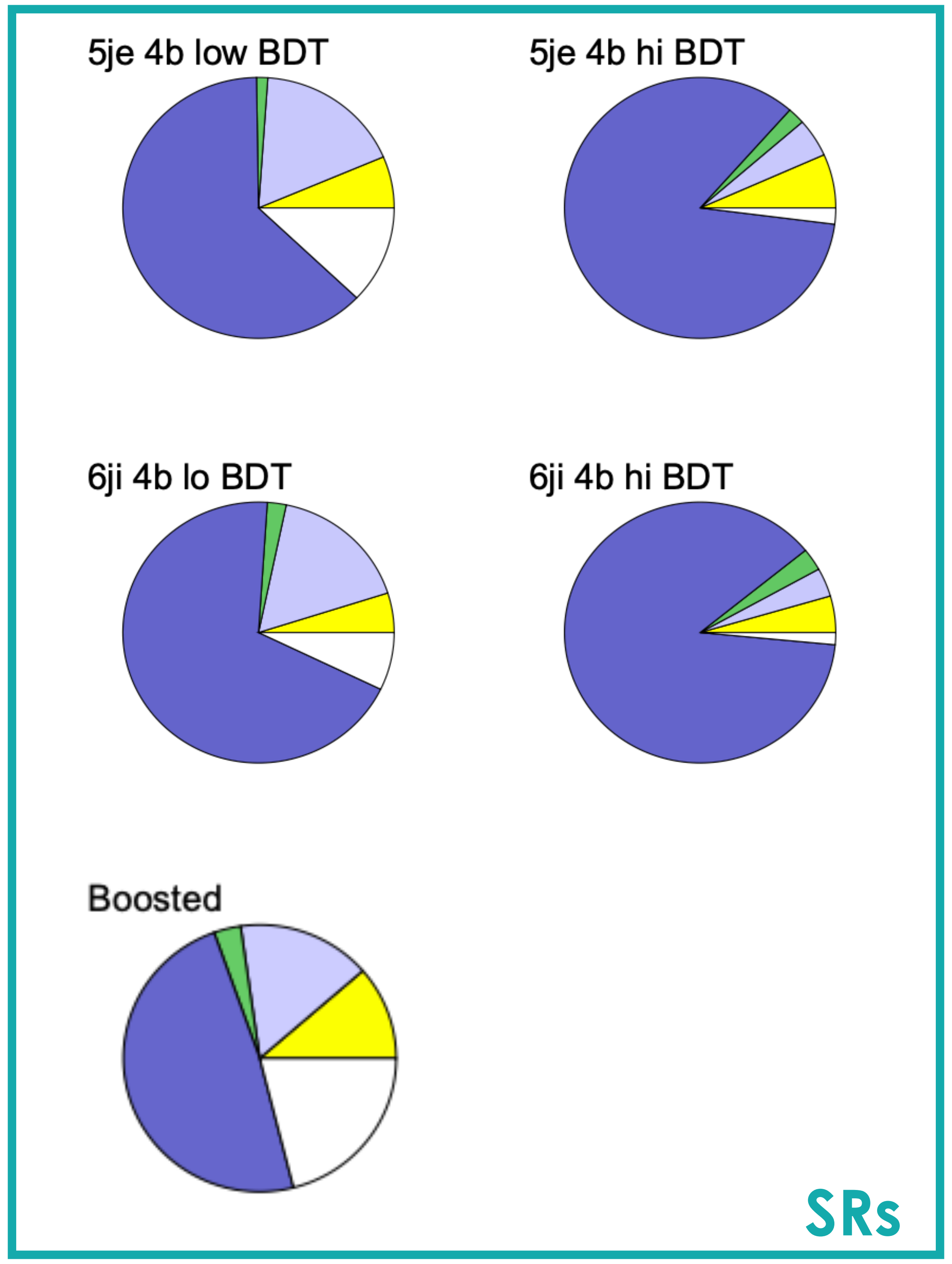
- regions where ttH and tt+bb are enhanced relative to the other backgrounds: **'signal regions' (SR)**
  - multivariate techniques used to further separate ttH signal from background events;
- remaining analysis categories: **'control regions' (CR)**
  - no attempt is made to separate signal from background in these regions,
  - but they provide stringent constraints on backgrounds normalisations and systematic uncertainties in a combined fit with the signal regions.

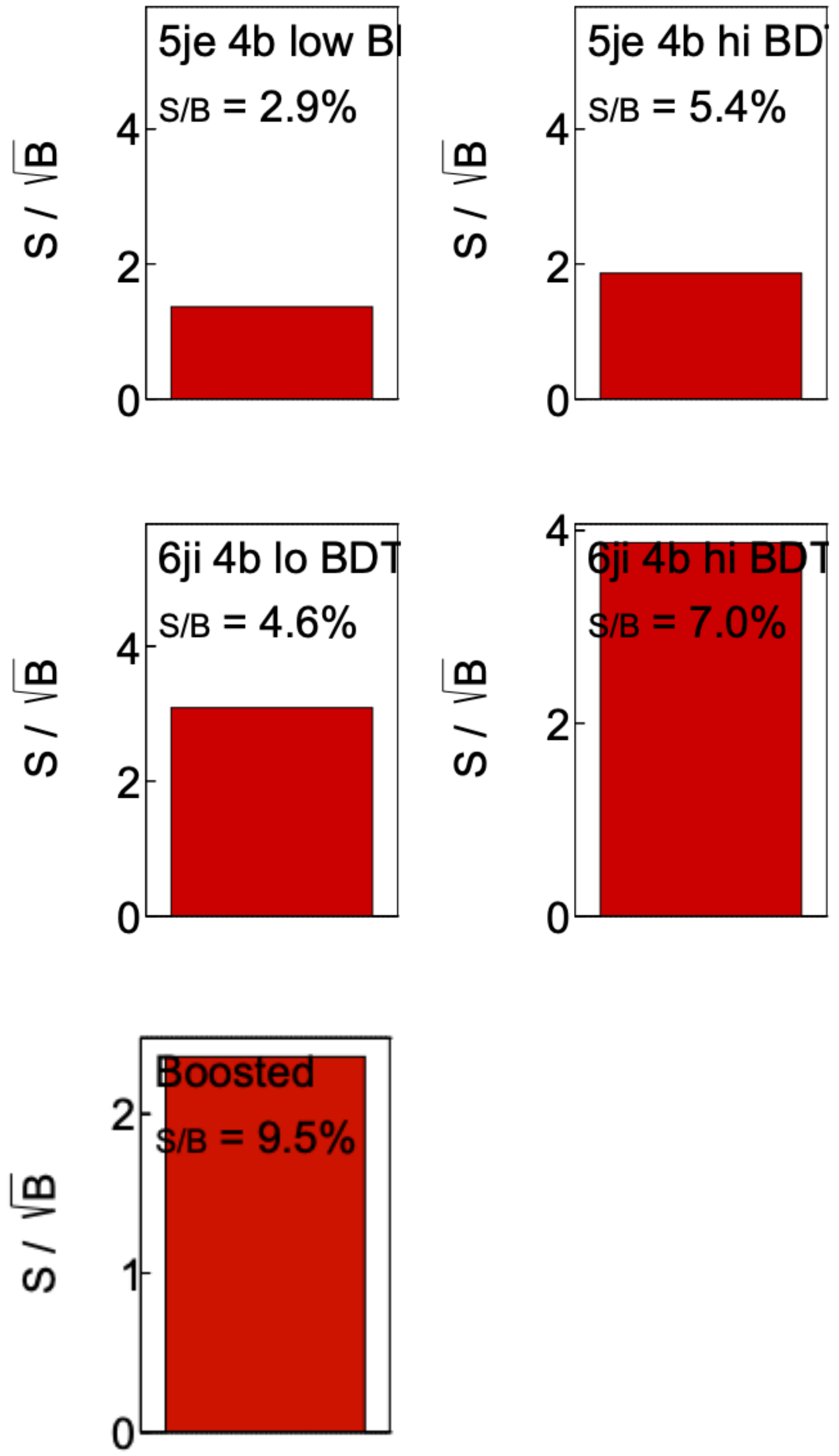
<table>
<tr><th rowspan="2">Region</th><th rowspan="2">#leptons</th><th rowspan="2">#jets</th><th colspan="3">#b-tag</th><th rowspan="2"># boosted Higgs candidates</th></tr>
<tr><th>@60%</th><th>@70%</th><th>@85%</th></tr>
<tr><td>$SR^{2\ell}_{\geq 4j \geq 4b}$</td><td rowspan="4">= 2</td><td rowspan="3">≥ 4</td><td>≥ 4</td><td rowspan="2">≥ 4</td><td rowspan="4">–</td><td rowspan="4">–</td></tr>
<tr><td>$SR^{2\ell}_{\geq 4j \geq 4b\text{low}}$</td><td>< 4</td></tr>
<tr><td>$CR^{2\ell}_{\geq 4j 3b}$</td><td rowspan="2">= 3</td><td rowspan="2">= 3</td></tr>
<tr><td>$CR^{2\ell}_{3j 3b}$</td><td>= 3</td></tr>
<tr><td>$SR^{1\ell}_{\geq 6j \geq 4b}$</td><td rowspan="5">= 1</td><td rowspan="2">≥ 6</td><td>≥ 4</td><td rowspan="4">≥ 4</td><td rowspan="5">≥ 3</td><td rowspan="4">0</td></tr>
<tr><td>$SR^{1\ell}_{\geq 6j \geq 4b\text{low}}$</td><td>< 4</td></tr>
<tr><td>$SR^{1\ell}_{5j \geq 4b}$</td><td rowspan="2">= 5</td><td>≥ 4</td></tr>
<tr><td>$SR^{1\ell}_{5j \geq 4b\text{low}}$</td><td>< 4</td></tr>
<tr><td>$SR^{1\ell}_{\text{boosted}}$</td><td>≥ 4</td><td colspan="2">–</td><td>≥ 1</td></tr>
</table>

- **Orthogonality between the regions ensured by**
  - **number of leptons** for the dilepton and single-lepton regions;
  - **number of boosted Higgs candidates** for the single-lepton boosted and resolved regions;
  - **number of jets and of b-tagged jets** using the 60% or 70% WPs for the different regions of each resolved channels;
  - **Boosted veto**: events which fall in the boosted category are removed from the single-lepton resolved regions.

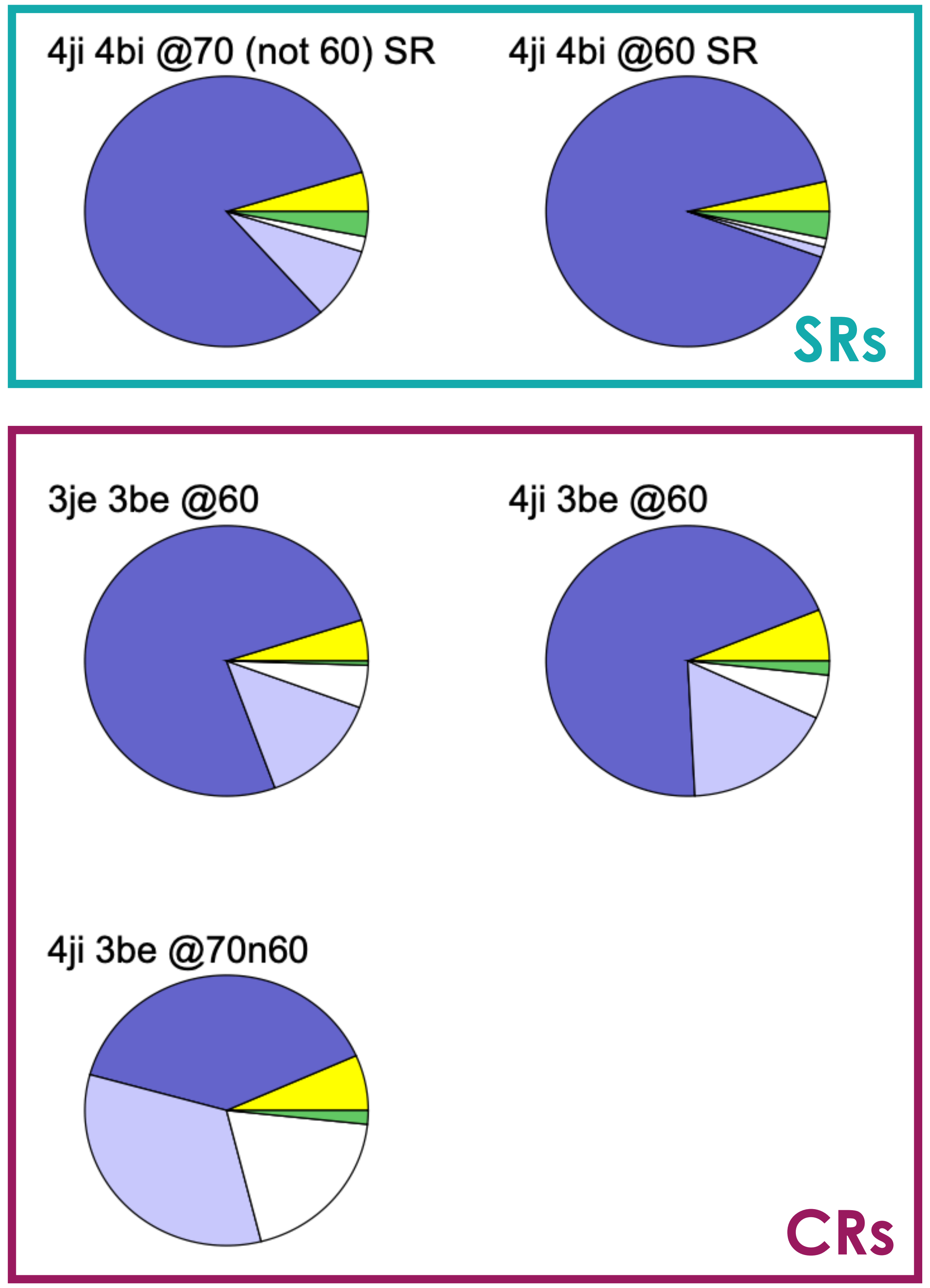
ttH analysis - background composition

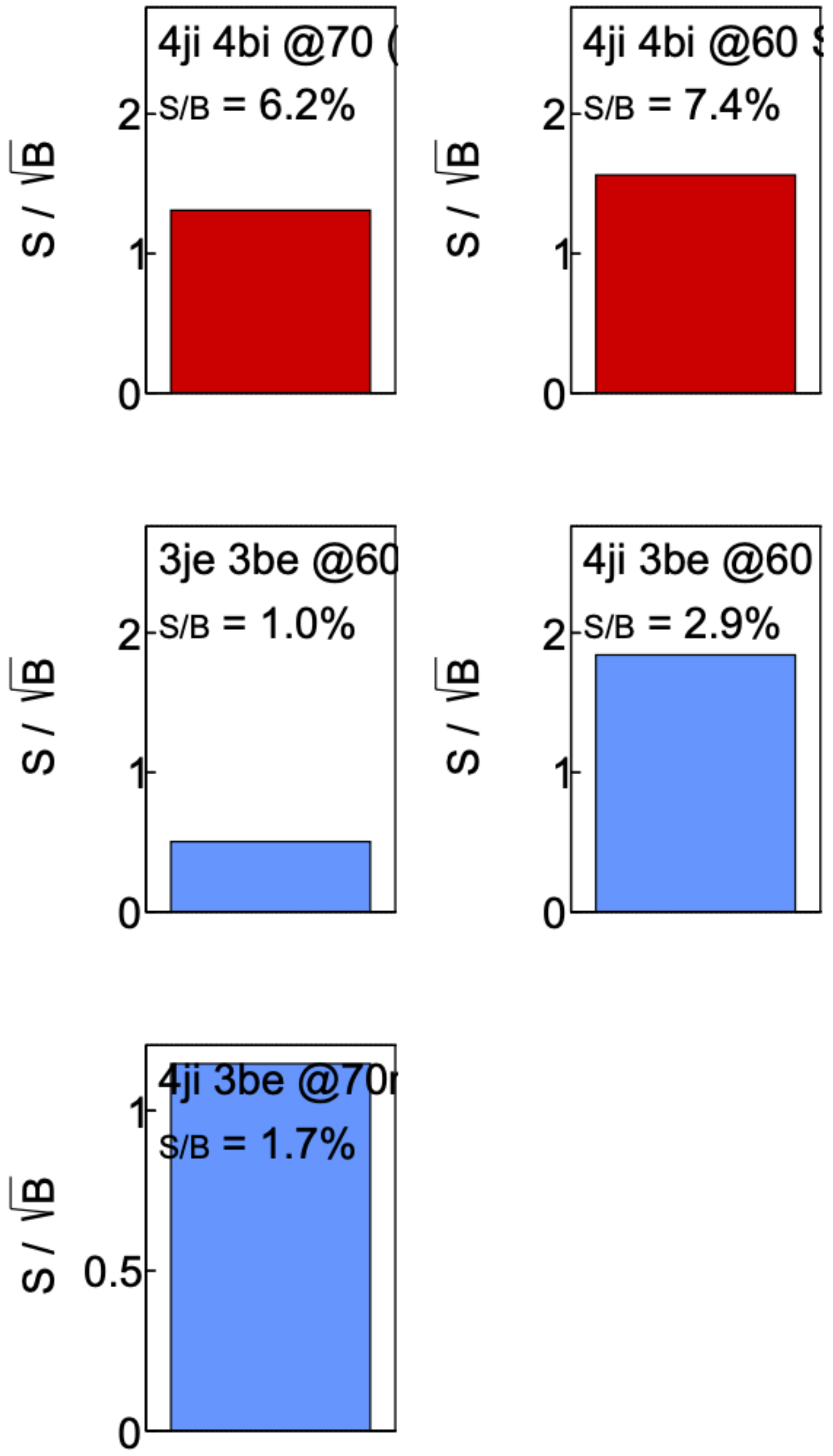
l+jets categories

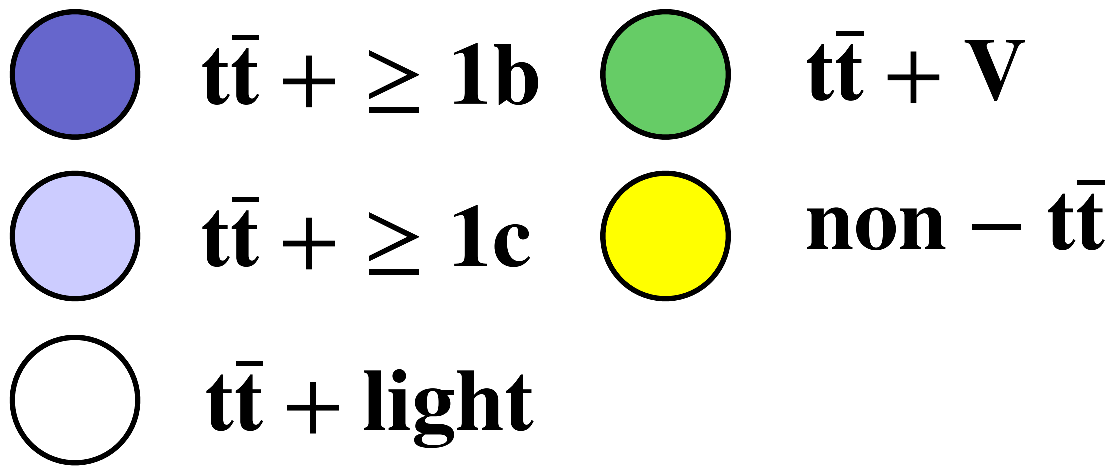




dilepton categories







- $t\bar{t} + \geq 1b$ is the main background;
- non tt = W/Z+jets, single top, diboson production processes.

ttH analysis - ML techniques in all channels

technique	ML discriminant	goal
event selection in different channels	→	definition of training samples
intermediate multivariate discriminants	reconstruction BDT (resolved dilepton channel) reconstruction BDT and LHD (resolved single-lepton channel) DNN (boosted channels)	reconstruction of objects and discrimination of ttH wrt tt
final multivariate discriminants	classification BDTs	separation of signal from background and final discriminants for the profile likelihood fit

analysis result

☆ **Input variables**:
important to check the Data/MC agreement looking for any mis-modelling that can affect the performance of the ML technique.

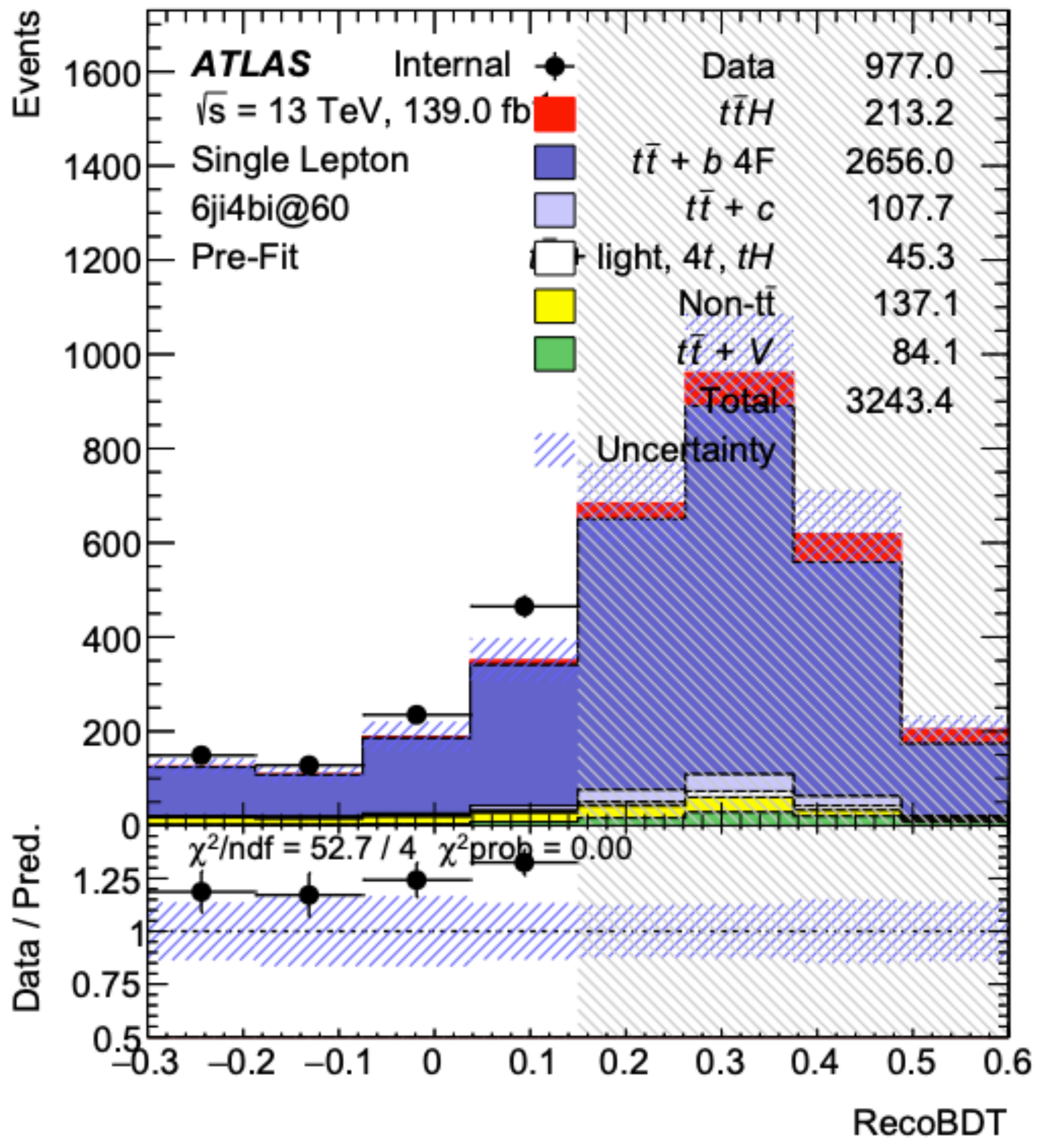
ttH analysis - reconstruction with a BDT

* only resolved channels

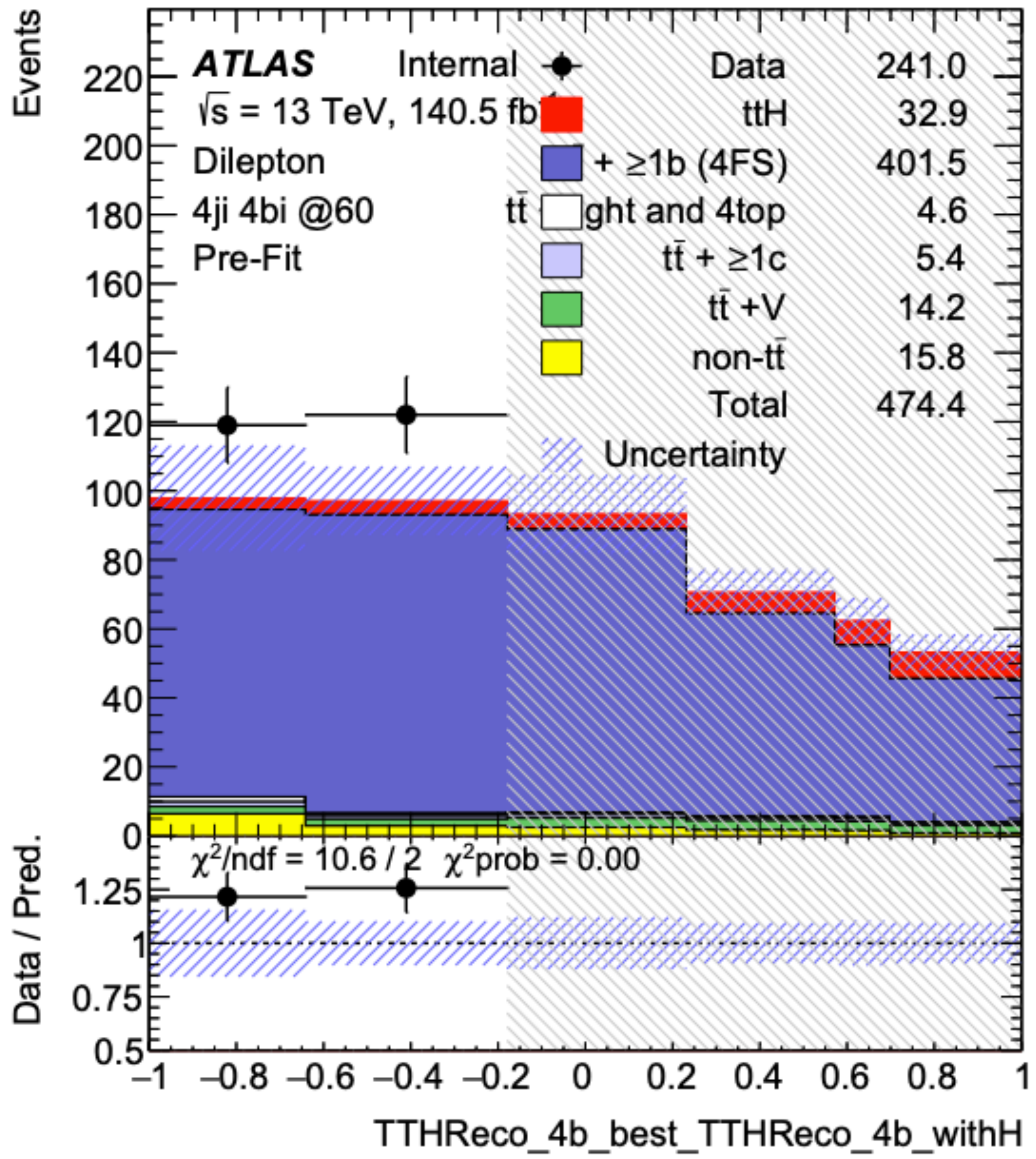
Select the best combination of jet–parton assignments in each event and build the Higgs-boson and top-quark candidates

- BDTs trained to **match reconstructed jets to the partons emitted from top-quark and Higgs-boson decays**;
  - **distinguish between correct and incorrect jet assignments**, using invariant masses and angular separations in addition to other kinematic variables as inputs;
- W-boson, top-quark and Higgs-boson candidates are built from combinations of jets and leptons;
- simulated ttH events are used to **iterate over all allowed combinations**.
- In each event, a specific combination of jet–parton assignments, corresponding to the best BDT output, is chosen in order to **compute kinematic and topological information of the top-quark and Higgs-boson candidates to be input to the classification BDT**.

dilepton channel



single-lepton channel



ttH analysis - reconstruction with a BDT

* only resolved channels

- **best possible reconstruction performance** can be obtained by including **information related to the Higgs boson** (i.e. invariant mass):
  - **possible bias in the bkg distributions** of Higgs-boson-related observables towards the signal expectation, reducing their ability to separate signal from background;
- two versions of the reconstruction BDT are used:
  - **one with and one without the Higgs-boson information** and the resulting jet–parton assignments from one, the other or both are considered when computing input variables for the classification BDT.

dilepton channel

Variables	BDT with Higgs info.	BDT w/o Higgs info.
Topological information from $t\bar{t}$		
Mass of top	✓	✓
Mass of anti-top	✓	✓
Mass difference between top and anti-top	✓	✓
$\Delta R(\ell, b)$ from top	✓	✓
$\Delta R(\ell, b)$ from anti-top	✓	✓
$\|\Delta R(\ell, b)$ from top - $\Delta R(\ell, b)$ from anti-top$\|$	–	✓
$\Delta R(b$ from top, b from anti-top$)$	✓	–
$\Delta\phi(b$ from top, b from anti-top$)$	–	✓
p_T b from top	–	✓
p_T b from anti-top	–	✓
Min. $\Delta\eta(\ell, b$ from top or anti-top$)$	–	✓
Topological information from the Higgs-boson candidate		
Max. $\Delta R($Higgs, b from top or anti-top$)$	✓	–
Mass of Higgs	✓	–
$\Delta R($Higgs$, t\bar{t})$	✓	–
$\Delta R(b_1$ from Higgs, b_2 from Higgs$)$	✓	–

single-lepton channel

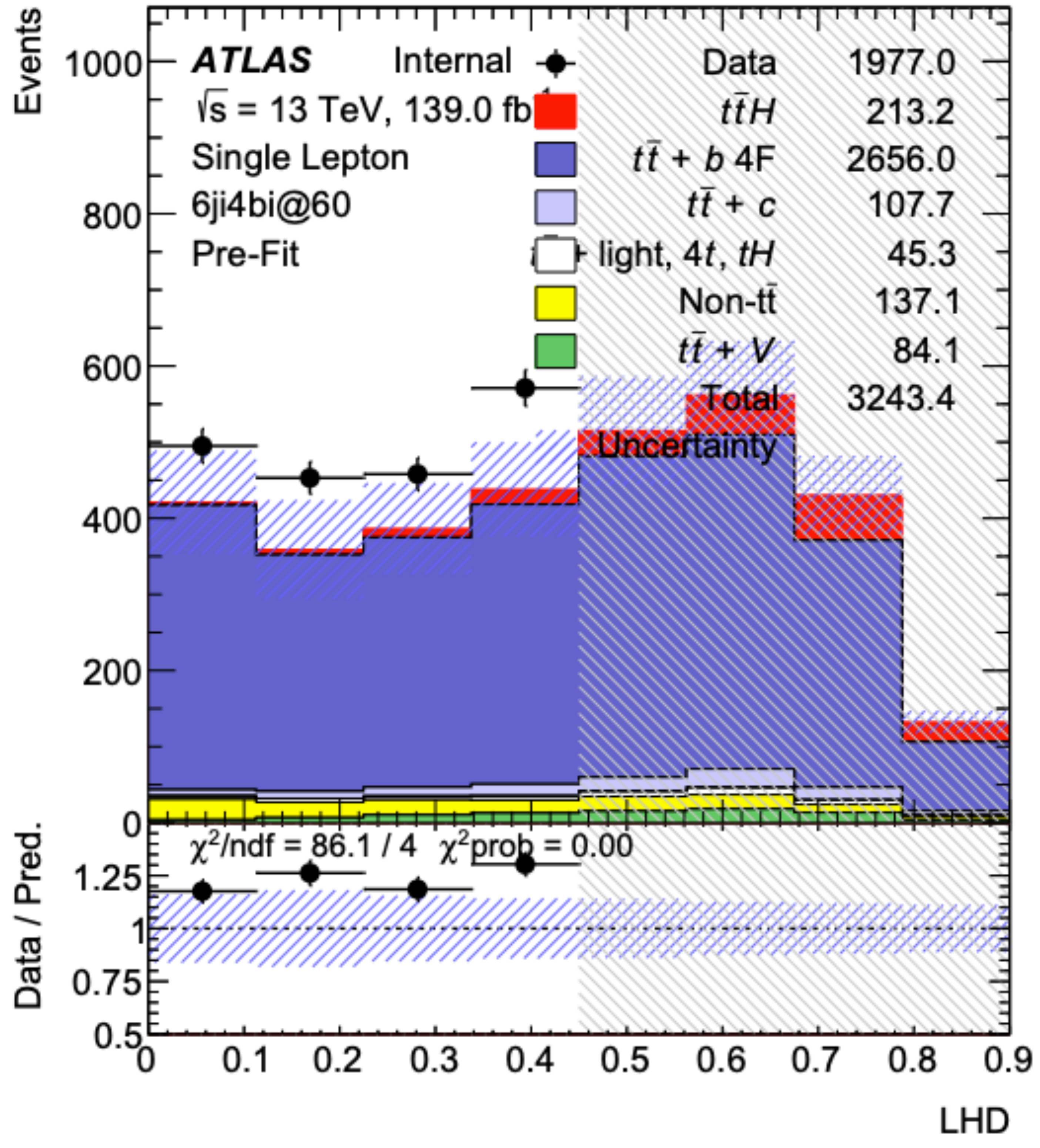
Variable	6 jets	5 jets
Topological information from $t\bar{t}$		
Mass of top$_{\text{lep}}$	✓	✓
Mass of top$_{\text{had}}$	✓	–
Mass of q_1 from W_{had} and b from top$_{\text{had}}$	–	✓
Mass of W_{had}	✓	–
Mass of W_{had} and b from top$_{\text{lep}}$	✓	–
Mass of q_1 from W_{had} and b from top$_{\text{lep}}$	–	✓
Mass of W_{lep} and b from top$_{\text{had}}$	✓	✓
$\Delta R(W_{\text{had}}, b$ from top$_{\text{had}})$	✓	–
$\Delta R(q_1$ from W_{had}, b from top$_{\text{had}})$	–	✓
$\Delta R(W_{\text{had}}, b$ from top$_{\text{lep}})$	✓	–
$\Delta R(q_1$ from W_{had}, b from top$_{\text{lep}})$	–	✓
$\Delta R(\ell, b$ from top$_{\text{lep}})$	✓	✓
$\Delta R(\ell, b$ from top$_{\text{had}})$	✓	✓
$\Delta R(b$ from top$_{\text{lep}}, b$ from top$_{\text{had}})$	✓	✓
$\Delta R(q_1$ from W_{had}, q_2 from $W_{\text{had}})$	✓	–
$\Delta R(b$ from t_{had}, q_1 from $W_{\text{had}})$	✓	–
$\Delta R(b$ from t_{had}, q_2 from $W_{\text{had}})$	✓	–
Min. $\Delta R(b$ from top$_{\text{had}}, q_i$ from $W_{\text{had}})$	✓	–
$\Delta R($lep, b from top$_{\text{lep}})$ - min. $\Delta R(b$ from top$_{\text{had}}, q_i$ from $W_{\text{had}})$	✓	✓
Topological information from the Higgs-boson candidate		
Mass of Higgs	✓	✓
Mass of Higgs and q_1 from W_{had}	✓	✓
$\Delta R(b_1$ from Higgs, b_2 from Higgs$)$	✓	✓
$\Delta R(b_1$ from Higgs, lepton$)$	✓	✓
$\Delta R(b_1$ from Higgs, b from top$_{\text{lep}})$	–	✓
$\Delta R(b_1$ from Higgs, b from top$_{\text{had}})$	–	✓

ttH analysis - reconstruction with a LHD

* only resolved single-lepton channel

Additional input variable for the classification BDT

- **product of one-dimensional pdfs**, for the signal and the background hypotheses:
  - built for various invariant masses and angular distributions from reconstructed jets and leptons and from the missing transverse momentum;
- **Two background hypotheses** are considered:
  - production of tt+≥ 2 b-jets and tt+1 b-jet;
  - likelihoods for both hypotheses are weighted by their relative fractions in simulated tt+jets events;
- The **probabilities p_{sig} and p_{bkg}**, for signal and background hypotheses, respectively, are obtained as the **product of the pdfs for the different kinematic distributions**.



5 jets	6 jets
$M_H(b_1, b_2)$	$M_H(b_1, b_2)$
$M_{t_l}(l, \nu, b_l)$	$M_{t_l}(l, \nu, b_l)$
$M_{W_h}(q_1, q_2)$	–
$[M_{t_h} - M_{W_h}](b_h, q_1, q_2)$	$M_{t_h}(b_h, q_1)$
$[M_{t_h t_l} - M_{t_h} - M_{t_l}](l, \nu, b_l, b_h, q_1, q_2)$	$[M_{t_h t_l} - M_{t_h} - M_{t_l}](l, \nu, b_l, b_h, q_1)$
$[M_{t_h t_l b_1 b_2} - M_{t_l t_h} - M_H](l, \nu, b_l, b_h, q_1, q_2, b_1, b_2)$	$[M_{t_h t_l b_1 b_2} - M_{t_l t_h} - M_H](l, \nu, b_l, b_h, q_1, b_1, b_2)$
$\cos\theta^*_{b_{1/2},H}(b_1, b_2)$	$\cos\theta^*_{b,H}(b_1, b_2)$
$\cos\theta^*_{b_1 b_2, t_h t_l b_1 b_2}(l, \nu, b_l, b_h, q_1, q_2, b_1, b_2)$	$\cos\theta^*_{b_1 b_2, t_h t_l b_1 b_2}(l, \nu, b_l, b_h, q_1, b_1, b_2)$

- For each event, the discriminant is defined as the **ratio of the probability p_{sig} to the sum of p_{sig} and p_{bkg}** and added as an input variable to the classification BDT;
- takes advantage of **all possible combinations in the event**, but it does not fully account for correlations between variables in one combination.

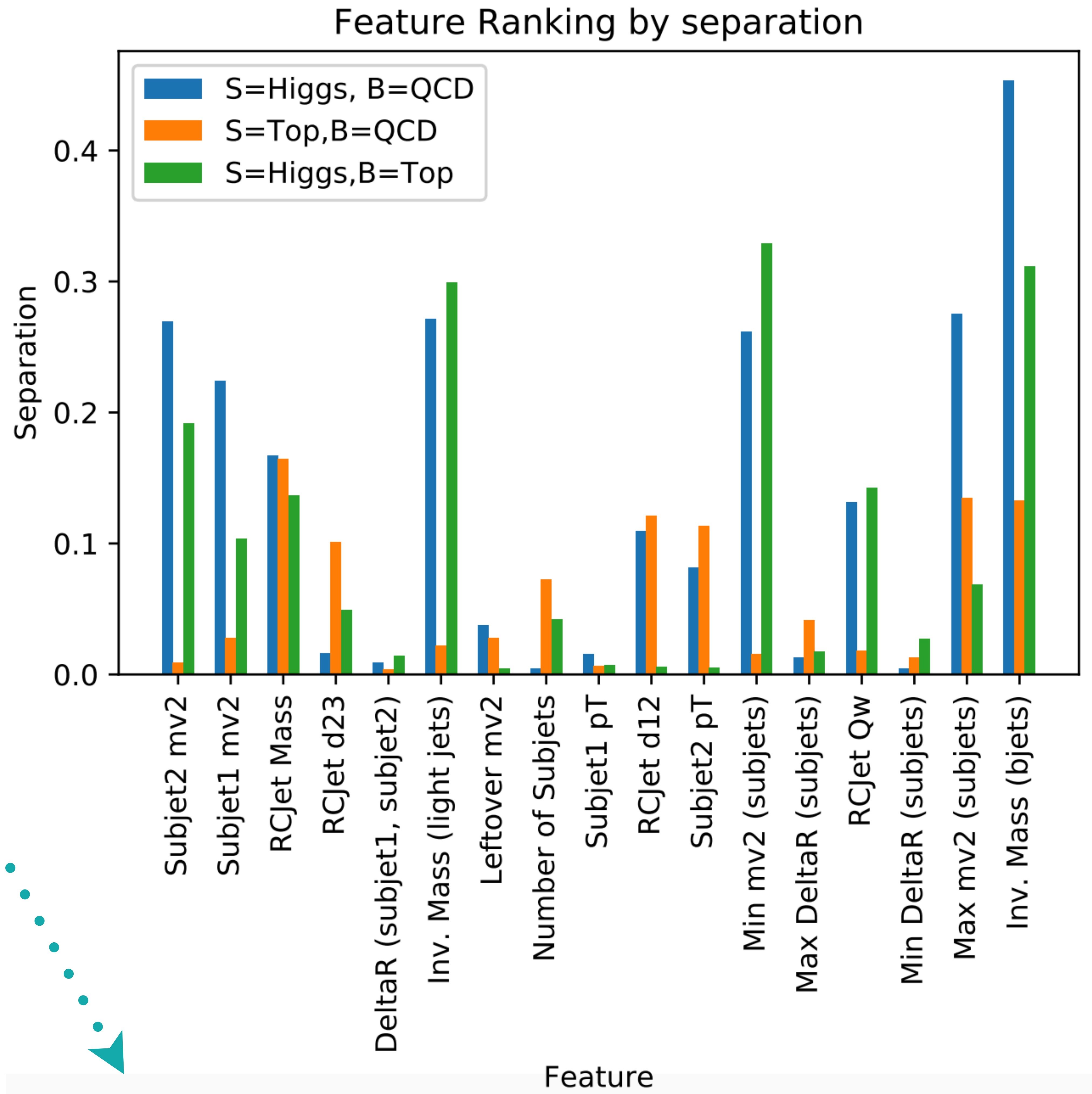
ttH analysis - reconstruction with a DNN

* only boosted channel

New region definition with new techniques

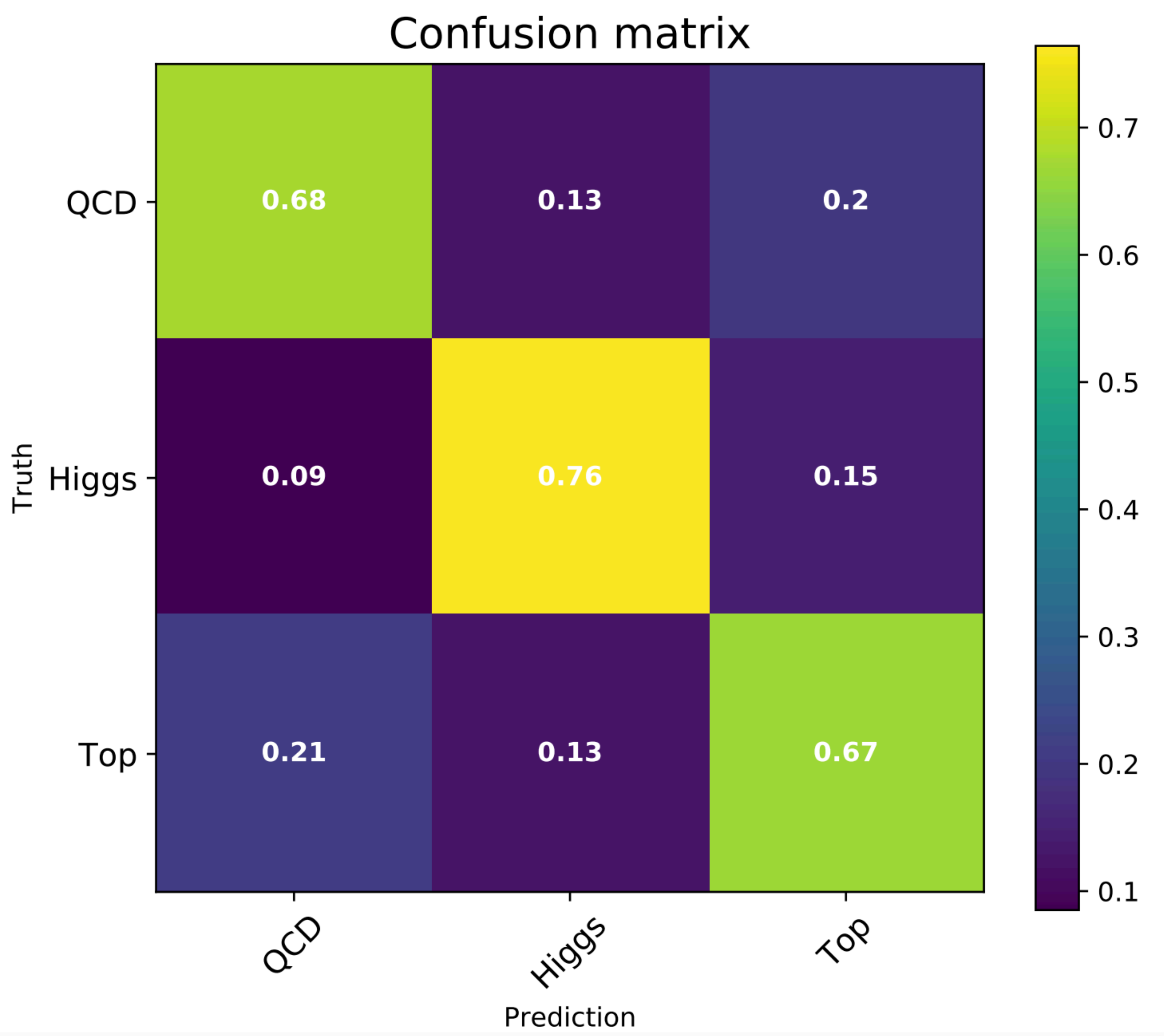
- **multi-class deep neural network (DNN)** is trained to identify the most likely parent particle of the RC jets, distinguishing between 3 categories: **Higgs-boson**, **top-quark** and **QCD** jets;
- improvement of the reconstruction matching of Higgs and top final objects wrt the previous analysis strategy;
- the output is used as **input for the classification BDT** and as an **additional requirement of the boosted signal selection**.

Variable	Description
m^{RCjet}	mass of reclustered jet
$\sqrt{d_{12}}$	first splitting scale
$\sqrt{d_{23}}$	second splitting scale
Q_W	minimum invariant mass of constituent pairs
$n_{\mathrm{constituents}}$	number of constituents in the RC jet
$p_{\mathrm{T}}^{\mathrm{const}_1}$	p_{T} of constituent leading in pseudo-continuous b-tagging score (mv2)
$p_{\mathrm{T}}^{\mathrm{const}_2}$	p_{T} of constituent sub-leading in mv2
$\mathrm{mv2}^{\mathrm{const}_1}$	mv2 score of constituent leading in mv2
$\mathrm{mv2}^{\mathrm{const}_2}$	mv2 score of constituent sub-leading in mv2
$\Delta R(\mathrm{const1},\mathrm{const2})$	angular separation between leading and sub-leading constituents in mv2 score
$m^{b-\mathrm{jets}}$	invariant mass of all b-tagged constituents
$m^{\mathrm{light-jets}}$	invariant mass of all untagged constituents
$\mathrm{mv2}_{\mathrm{min}}$	minimum constituent mv2 score
$\mathrm{mv2}_{\mathrm{max}}$	maximum constituent mv2 score
$\Delta R(\mathrm{consts})_{\mathrm{max}}$	maximum angular separation between two constituents
$\Delta R(\mathrm{consts})_{\mathrm{min}}$	minimum angular separation between two constituents
$\mathrm{mv2}^{\mathrm{rest}}$	mv2 score of all constituents except the leading and sub-leading in mv2 score



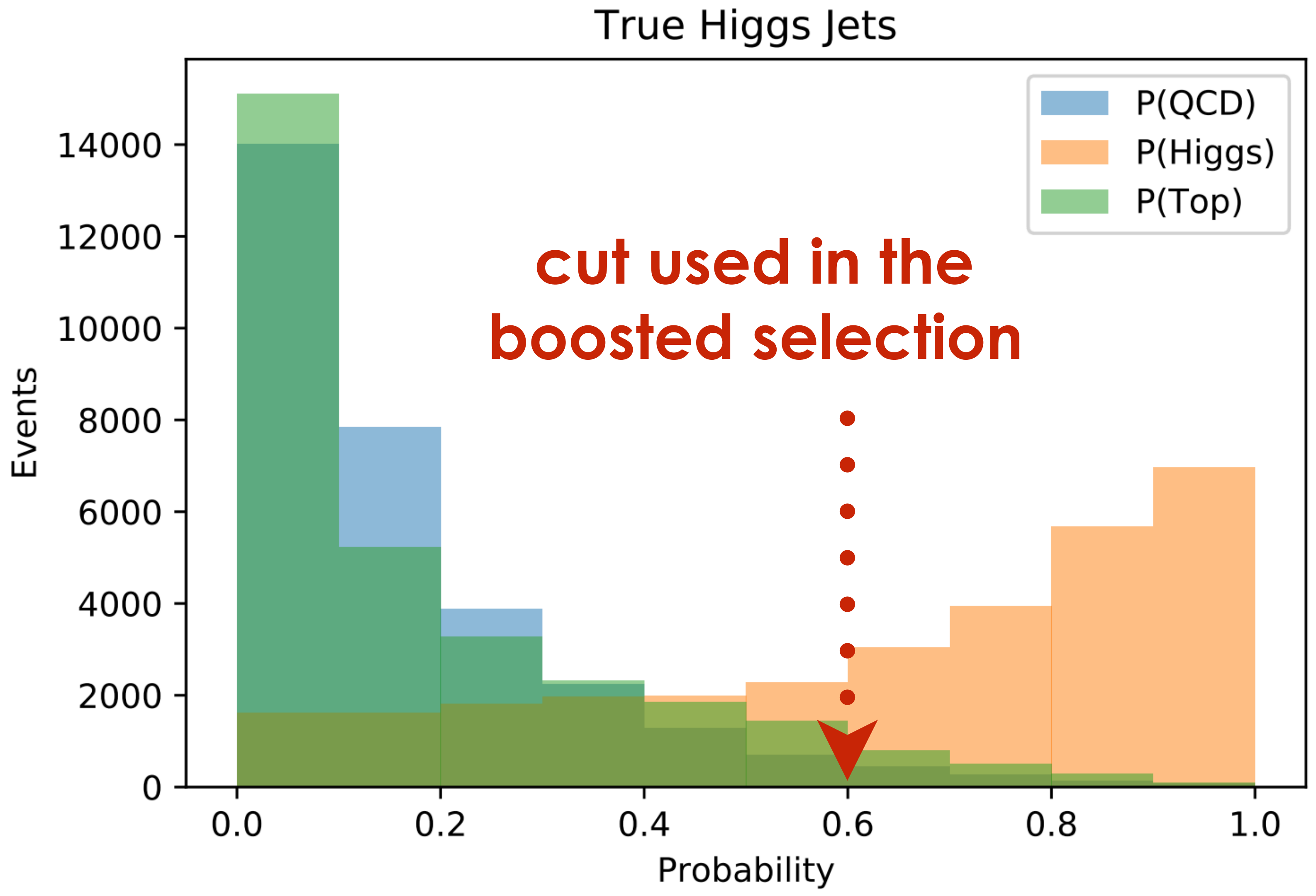
Higgs jets are correctly identified 76% of the time
top jets are correctly identified 67% of the time

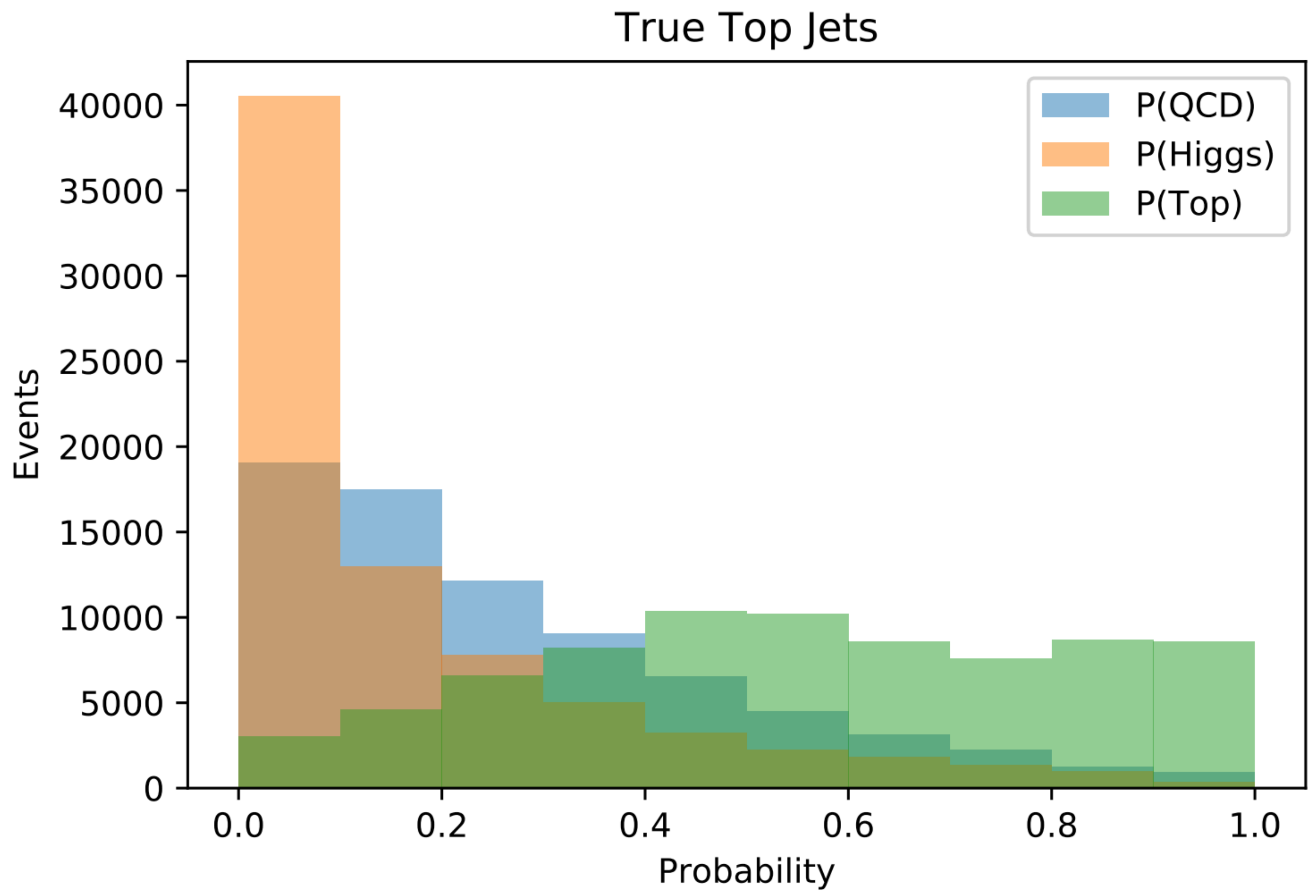
ttH analysis - reconstruction with a DNN

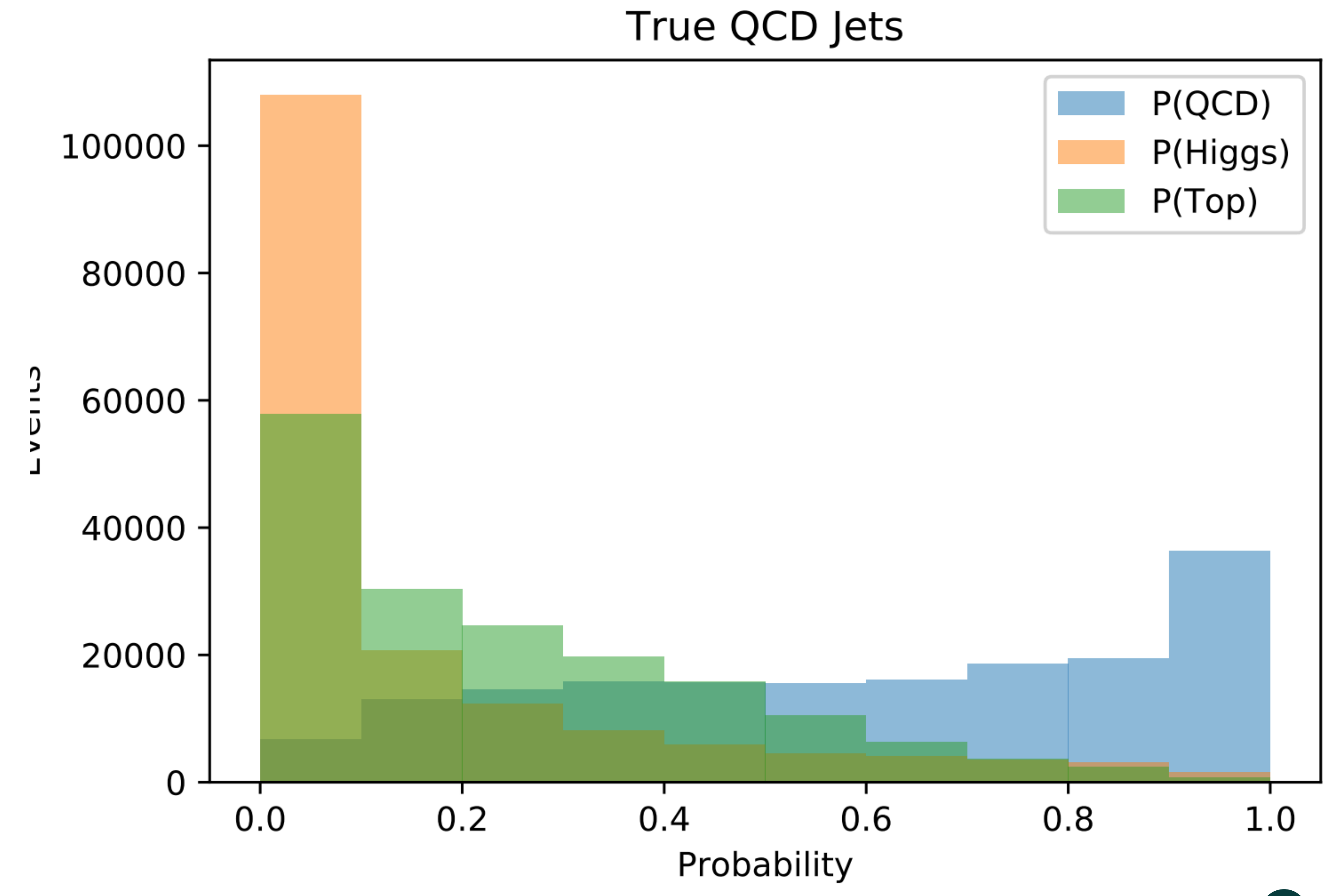


DNN training

- **Truth matching** between the RC jets to a hadronically decaying boson or top quark:
    - for the **Higgs boson**, two b-quarks have been required to match the constituents of the RC jet within a cone of $\Delta R = 0.4$;
    - for the **top quark**, one b-quark and at least one W-boson decay product have been required to match the constituents of the RC jet within a cone of $\Delta R = 0.4$;
    - all the other RC jets are labeled as **QCD jets**.
- **Trained jet by jet**, using a ttH sample;
- events are selected requiring at least **one isolated lepton with $p_T > 27$ GeV** and **two RC jets, each with $p_T > 200$ GeV, m > 50 GeV and at least two constituents**.







ttH analysis - ML techniques in all channels

technique	ML discriminant	goal
event selection in different channels	→	definition of training samples
intermediate multivariate discriminants	reconstruction BDT (resolved dilepton channel) reconstruction BDT and LHD (resolved single-lepton channel) DNN (boosted channels)	reconstruction of objects and discrimination of ttH wrt tt
final multivariate discriminants	classification BDTs	separation of signal from background and final discriminants for the profile likelihood fit

analysis result

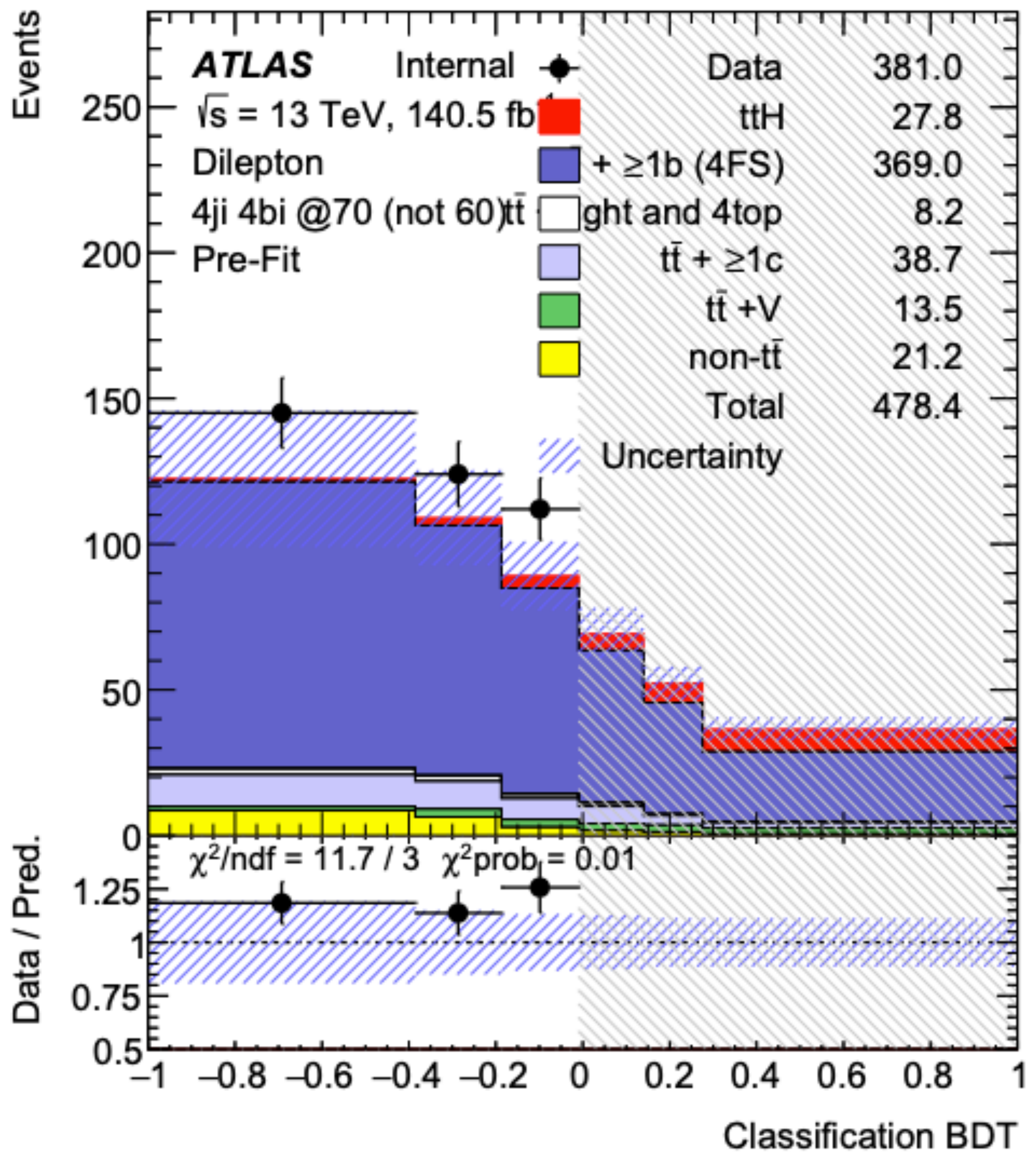
☆ **Input variables**: important to check the Data/MC agreement looking for any mis-modelling that can affect the performance of the ML technique.

ttH analysis - classification BDTs

Events used in the training

- inclusive selection of events with **at least 4 jets**;
- **at least 4 of which are b-tagged using the 70%** working point.

- For variables depending on b-tagged jets, only jets b-tagged using the 70% WP are considered.



dilepton channel

Variable	Definition	
General kinematic variables		
m_{bb}^{min}	Minimum invariant mass of a b-tagged jet pair	✓
$m_{bb}^{\text{min}\,\Delta R}$	Invariant mass of the b-tagged jet pair with minimum ΔR	✓
$m_{jj}^{\text{max}\,p_T}$	Invariant mass of the jet pair with maximum p_T	✓
$m_{bb}^{\text{max}\,p_T}$	Invariant mass of the b-tagged jet pair with maximum p_T	✓
$\Delta\eta_{bb}^{\text{avg}}$	Average $\Delta\eta$ for all b-tagged jet pairs	✓
$N_{bb}^{\text{Higgs 30}}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass	✓
Variables from reconstruction BDT		
BDT output	Output of the reconstruction BDT	✓**
m_{bb}^{Higgs}	Higgs candidate mass	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*
$\Delta R_{H,\ell}^{\text{min}}$	Minimum ΔR between Higgs candidate and lepton	✓
$\Delta R_{H,b}^{\text{min}}$	Minimum ΔR between Higgs candidate and b-jet from top	✓

*** variables from reconstruction BDT using Higgs-boson information**

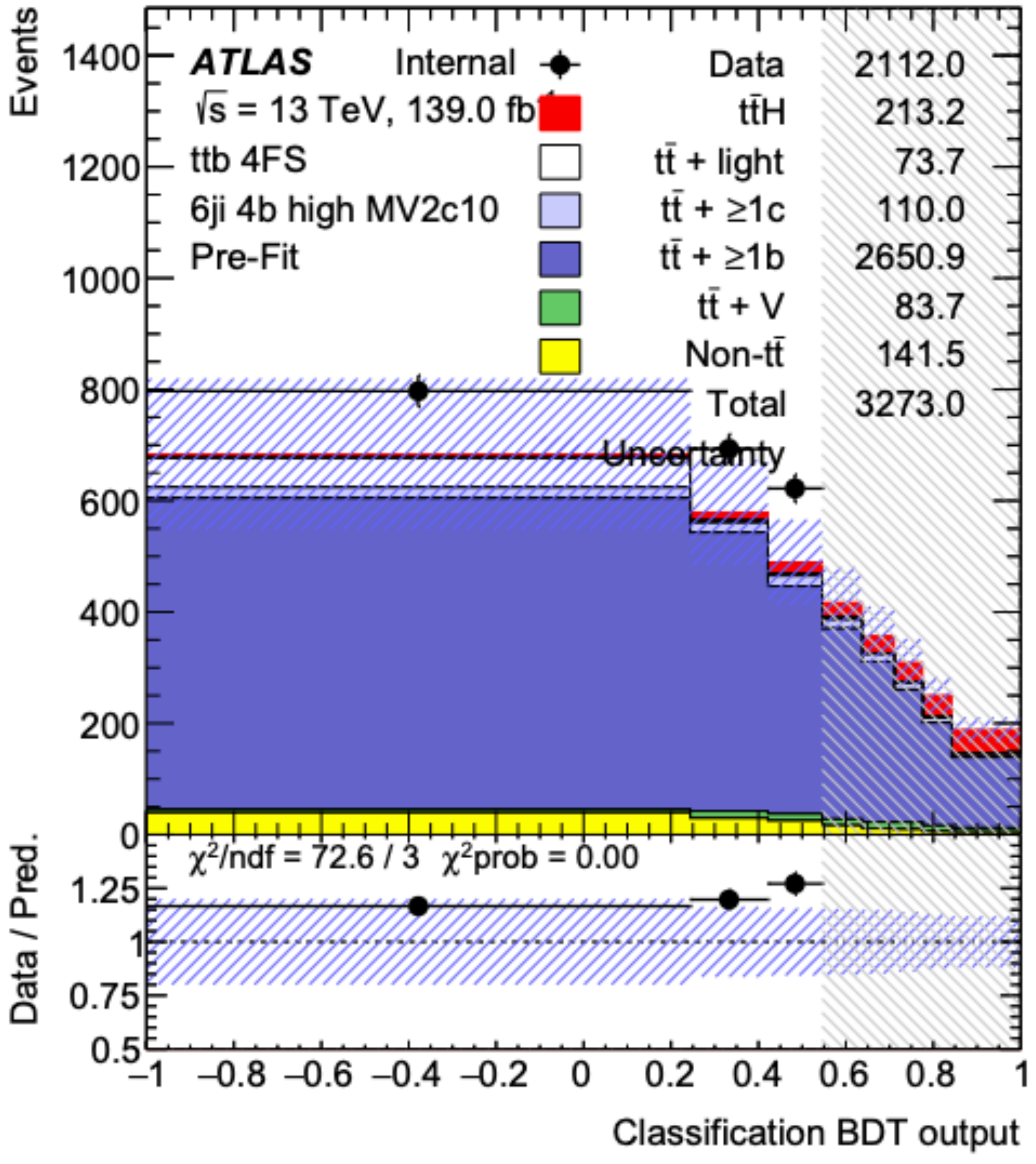
**** variables from both recoBDT (with and without Higgs information)**

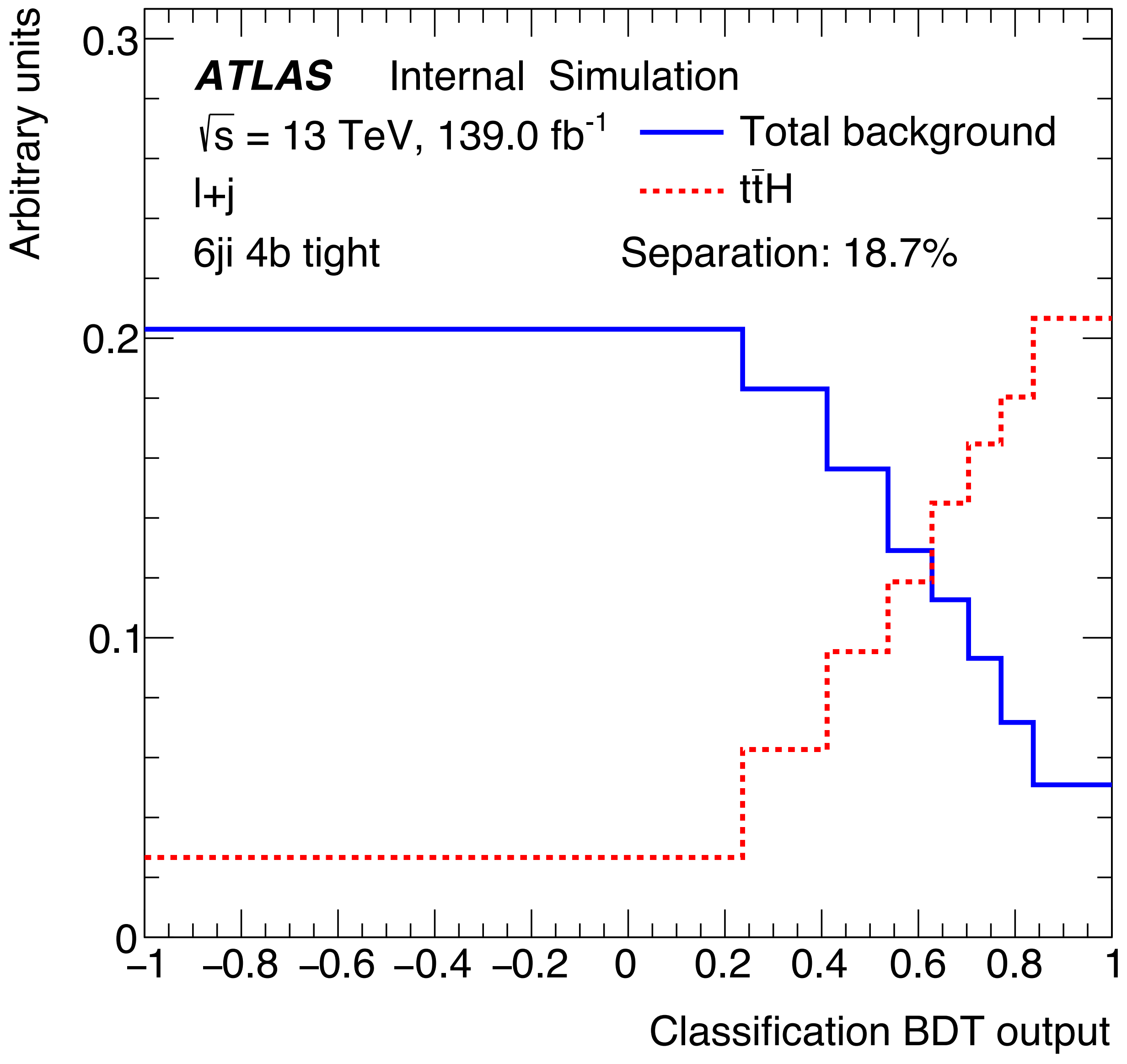
ttH analysis - classification BDTs

Events used in the training of the 5 jets (6 jets) classification BDT

- inclusive selection of events with **at least 5 (6) jets**;
- **4 of which are b-tagged using the 85%** working point.

- For variables depending on b-tagged jets, jets are sorted by their pseudo-continuous b-tag (PCB) score, and by their p_T when they have the same b-tag score;
- **LHD** and **recoBDT** outputs is included for both the BDTs.





*** variables from reconstruction BDT using Higgs-boson information**

single-lepton channel

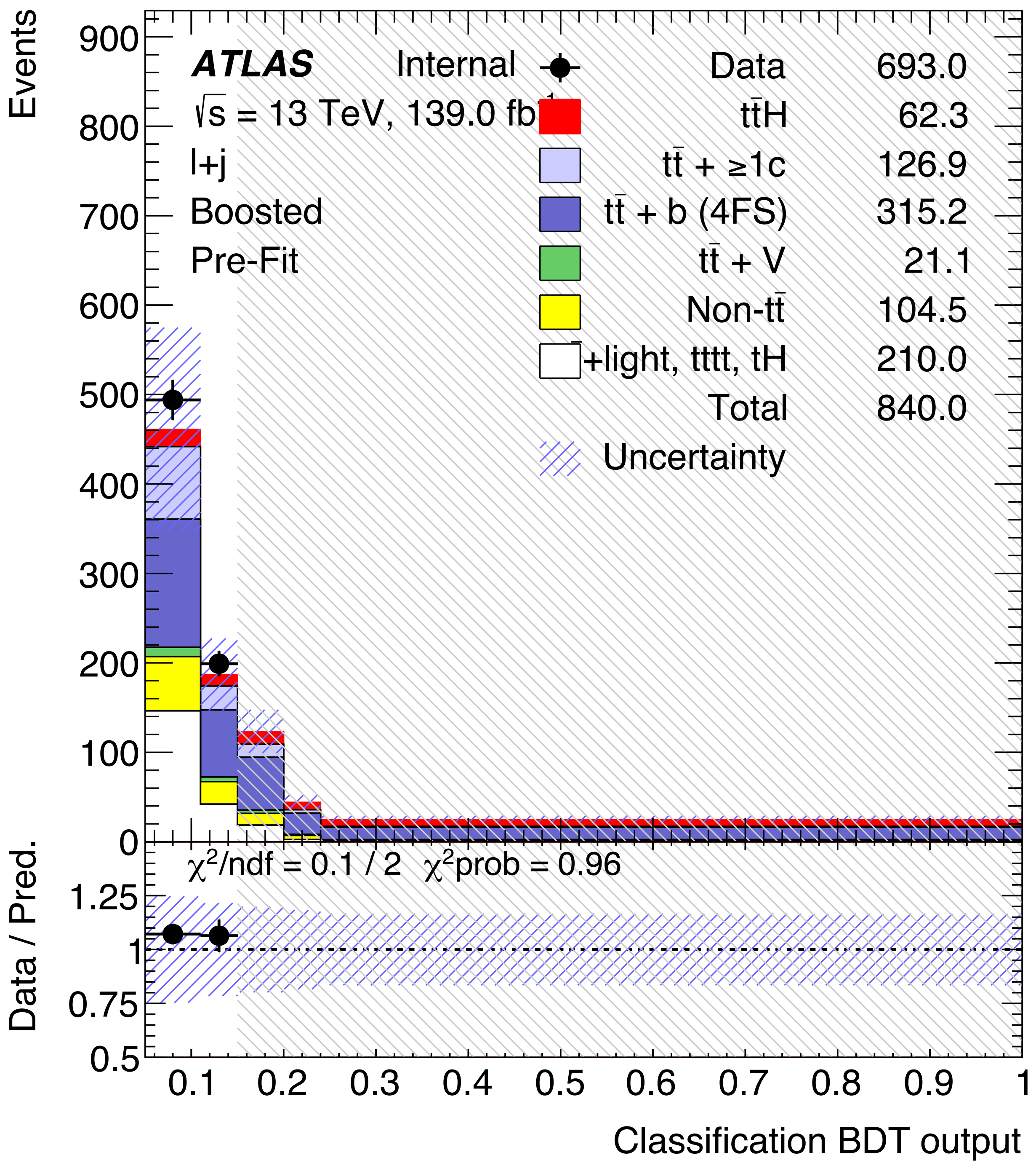
Variable	Definition	6 jets	5 jets
General kinematic variables			
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b-tagged jet pairs	✓	✓
$\Delta R_{bb}^{\max p_T}$	ΔR between the two b-tagged jets with the largest vector sum p_T	✓	–
$\Delta\eta_{jj}^{\max}$	Maximum $\Delta\eta$ between any two jets	✓	✓
$m_{bb}^{\min \Delta R}$	Mass of the combination of two b-tagged jets with the smallest ΔR	✓	–
$m_{jj}^{\min \Delta R}$	Mass of the combination of any two jets with the smallest ΔR	–	✓
$N_{bb}^{\text{Higgs } 30}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass	✓	✓
H_T^{had}	Scalar sum of jet p_T	–	✓
$\Delta R_{\ell,bb}^{\min}$	ΔR between the lepton and the combination of the two b-tagged jets with the smallest ΔR	–	✓
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [117] built with all jets	✓	✓
H_1	Second Fox–Wolfram moment computed using all jets and the lepton	✓	✓
Variables from reconstruction BDT			
BDT output	Output of the reconstruction BDT	✓*	✓*
m_{bb}^{Higgs}	Higgs candidate mass	✓	✓
$m_{H,b_{\text{lep top}}}$	Mass of Higgs candidate and b-jet from leptonic top candidate	✓	–
$\Delta R_{bb}^{\text{Higgs}}$	ΔR between b-jets from the Higgs candidate	✓	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*	✓*
$\Delta R_{H,\text{lep top}}$	ΔR between Higgs candidate and leptonic top candidate	✓	–
$\Delta R_{H,b_{\text{had top}}}$	ΔR between Higgs candidate and b-jet from hadronic top candidate	–	✓*
Variables from likelihood calculations			
LHD	Likelihood discriminant	✓	✓
Variables from b-tagging (not in $\text{SR}^{1\ell}_{\geq 6j \geq 4b60}$)			
$w_{b\text{-tag}}^{\text{Higgs}}$	Sum of b-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT	✓	✓
B_{jet}^{3}	3rd largest jet b-tagging discriminant	✓	✓
B_{jet}^{4}	4th largest jet b-tagging discriminant	✓	✓
B_{jet}^{5}	5th largest jet b-tagging discriminant	✓	✓

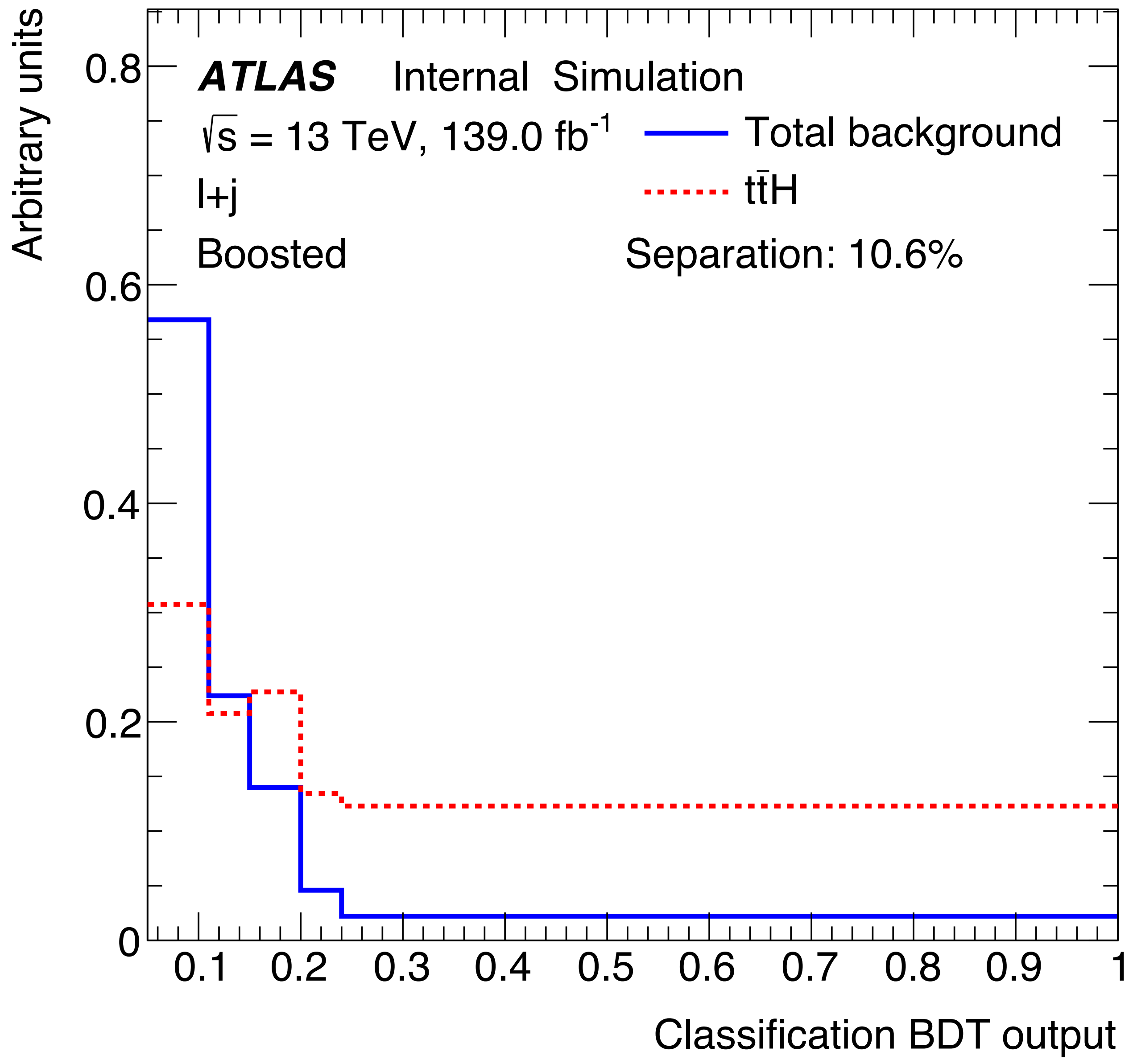
ttH analysis - classification BDTs

Events used in the training

- events passing the **event selection of the channel**;

- For variables depending on b-tagged jets, jets are sorted by their PCB score, and by their p_T when they have the same b-tag score;
- **DNN** output also included in the input variables list.





boosted channel

Variable	Description
m_{Higgs}	Higgs candidate mass
p_T^{Higgs}	Higgs candidate transverse momentum
$\eta_{\text{Higgs}}^{\text{lep}}$	η of the Higgs candidate relative to the lepton
$\text{P(H)}_{\text{Higgs}}$	DNN Higgs probability for the Higgs candidate
m_{hadTop}	hadronic top candidate mass
p_T^{hadTop}	hadronic top candidate transverse momentum
$\eta_{\text{hadTop}}^{\text{lep}}$	η of the hadronic top candidate relative to the lepton
$\text{PCB}_{\text{hadTop}}^{\text{jet}_i}$	PCB score of the i^{th} jet associated to the hadronic top
m_{lepTop}	leptonic top candidate mass
p_T^{lepTop}	leptonic top candidate transverse momentum
$\text{PCB}_{\text{lepTop}}^{\text{jet}}$	PCB score of the jet associated to the leptonic top
n_{jets}	small-R jets multiplicity
$\Delta R(\text{Higgs, hadTop})$	ΔR between the Higgs and the hadronic top candidates
$\Delta R(\text{Higgs, lepTop})$	ΔR between the Higgs and the leptonic top candidates
$\Delta R(\text{hadTop, lepTop})$	ΔR between the hadronic top and the leptonic top candidates
$p_T^{t\bar{t}H}$	transverse momentum of the $t\bar{t}H$ system
$p_T^{t\bar{t}}$	transverse momentum of the $t\bar{t}$ system
PCB^{sum}	PCB score sum of the jets associated to the Higgs, hadronic and leptonic top
$\text{PCB}^{\text{add jet}}$	PCB score of the additional jet in the event

ttH analysis - Systematic model

Systematic uncertainty	Type	Comp.
Experimental uncertainties		
Luminosity	N	1
Pileup modeling	SN	1
Physics Objects		
Electrons	SN	7
Muons	SN	15
Jet energy scale	SN	31
Jet energy resolution	SN	9
Jet vertex tagger	SN	1
E_T^{miss}	SN	3
***b*-tagging**		
Efficiency	SN	45
Mis-tag rate (c)	SN	20
Mis-tag rate (light)	SN	20

Signal and background modeling		
Signal		
$t\bar{t}H$ cross-section	N	2
H branching fractions	N	3
$t\bar{t}H$ modeling	SN	4
$t\bar{t}$ Background		
$t\bar{t}$ cross-section	N	1
$t\bar{t} + \geq 1c$ normalization	N	1
$t\bar{t} + \geq 1b$ normalization	N (free floating)	1
$t\bar{t}$ + light modeling	SN	4
$t\bar{t} + \geq 1c$ modeling	SN	4
$t\bar{t} + \geq 1b$ modeling	SN	4

Systematic sources treatment

- a single independent nuisance parameter is assigned to each source of systematic uncertainty in the statistical analysis;
- some of the systematic uncertainties are **decomposed into several independent sources** (≥ 1 NP);
- each individual source has a **correlated effect across all the channels, analysis categories, signal and background samples**
- **tt background modelling** requires a careful and complex treatment.

Other Backgrounds		
$t\bar{t}W$ cross-section	N	2
$t\bar{t}Z$ cross-section	N	2
$t\bar{t}W$ modeling	SN	1
$t\bar{t}Z$ modeling	SN	1
Single top cross-section	N	3
Single top modeling	SN	?
W+jets normalization	N	3
Z+jets normalization	N	3
Diboson normalization	N	1
$t\bar{t}t\bar{t}$ cross-section	N	1
Small backgrounds cross-sections	N	?
Multijets normalisation	SN	8

ttH analysis - tt background model

- **tt+≥1b, tt+≥1c and tt+light processes affected by different types of uncertainties:**
  - **tt+light**: profits from relatively precise measurements in data;
  - **tt+≥1b** and **tt+≥1c** can have similar or different diagrams depending on the flavour scheme used for the PDF, different mass of the c- and the b-quark contribute to additional differences between these two processes.

Uncertainty source	Description		Components
$t\bar{t}$ cross-section	Up or down by 6%		$t\bar{t}$ + light
$t\bar{t} + \geq 1b$ normalisation	Free-floating		$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1c$ normalisation	Up or down by 50%		$t\bar{t} + \geq 1c$
NLO matching	MadGraph5_aMC@NLO +Pythia8 $t\bar{t}$ (5FS)	vs. PowhegBox+Pythia8 $t\bar{t}$ (5FS)	All
PS & hadronisation	PowhegBox+Herwig7 $t\bar{t}$ (5FS)	vs. PowhegBox+Pythia8 $t\bar{t}$ (5FS)	All
ISR	Varying α_S^{ISR} (PS), μ_{R}&μ_{F} (ME)	in PowhegBox+Pythia8 $t\bar{t}b\bar{b}$ (4FS)	$t\bar{t} + \geq 1b$
		in PowhegBox+Pythia8 $t\bar{t}$ (5FS)	$t\bar{t} + \geq 1c$, $t\bar{t}$ + light
FSR	Varying α_S^{FSR} (PS)	in PowhegBox+Pythia8 $t\bar{t}b\bar{b}$ (4FS)	$t\bar{t} + \geq 1b$
		in PowhegBox+Pythia8 $t\bar{t}$ (5FS)	$t\bar{t} + \geq 1c$, $t\bar{t}$ + light

ttH analysis - tt background model

- **tt+≥1b, tt+≥1c and tt+light processes affected by different types of uncertainties:**
  - **tt+light**: profits from relatively precise measurements in data;
  - **tt+≥1b** and **tt+≥1c** can have similar or different diagrams depending on the flavour scheme used for the PDF, different mass of the c- and the b-quark contribute to additional differences between these two processes.

- **Systematic uncertainties on the acceptance and shapes (nominal vs different MC samples and settings)**
  - Such comparisons would change significantly the fractions of tt+≥1b in the phase-space selected in this analysis when comparing to these alternative setups.
  - **the normalisation of this sub-process is measured on data** by the profile likelihood fit (free-floating);
  - **reweighing of the alternative predictions** is applied in such a way to have the same fraction of tt+≥1b as the nominal sample.

Uncertainty source	Description		Components
$t\bar{t}$ cross-section	Up or down by 6%		$t\bar{t}$ + light
$t\bar{t} + \geq 1b$ normalisation	Free-floating		$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1c$ normalisation	Up or down by 50%		$t\bar{t} + \geq 1c$
NLO matching	MadGraph5_aMC@NLO +Pythia8 $t\bar{t}$ (5FS)	vs. PowhegBox+Pythia8 $t\bar{t}$ (5FS)	All
PS & hadronisation	PowhegBox+Herwig7 $t\bar{t}$ (5FS)	vs. PowhegBox+Pythia8 $t\bar{t}$ (5FS)	All
ISR	Varying α_S^{ISR} (PS), μ_{R}&μ_{F} (ME)	in PowhegBox+Pythia8 $t\bar{t}b\bar{b}$ (4FS)	$t\bar{t} + \geq 1b$
		in PowhegBox+Pythia8 $t\bar{t}$ (5FS)	$t\bar{t} + \geq 1c$, $t\bar{t}$ + light
FSR	Varying α_S^{FSR} (PS)	in PowhegBox+Pythia8 $t\bar{t}b\bar{b}$ (4FS)	$t\bar{t} + \geq 1b$
		in PowhegBox+Pythia8 $t\bar{t}$ (5FS)	$t\bar{t} + \geq 1c$, $t\bar{t}$ + light

ttH analysis - tt background model

- tt+≥1b, tt+≥1c and tt+light processes affected by different types of uncertainties:
  - **tt+light**: profits from relatively precise measurements in data;
  - **tt+≥1b** and **tt+≥1c** can have similar or different diagrams depending on the flavour scheme used for the PDF, different mass of the c- and the b-quark contribute to additional differences between these two processes.

- Systematic uncertainties on the **acceptance and shapes** (nominal vs different MC samples and settings)
  - Such comparisons would change significantly the fractions of tt+≥1b in the phase-space selected in this analysis when comparing to these alternative setups.
  - **the normalisation of this sub-process is measured on data** by the profile likelihood fit (free-floating);
  - **reweighing of the alternative predictions** is applied in such a way to have the same fraction of tt+≥1b as the nominal sample.

- **Modelling uncertainties** of tt+≥1b (tt+≥1c and tt+light) by the nominal prediction MC sample
  - need to distinguish **different effects in the modelling**, while comparing, for each component, different MC setups with the same process generated in the ME and with sufficient stats.

Uncertainty source	Description		Components
$t\bar{t}$ cross-section	Up or down by 6%		$t\bar{t}$ + light
$t\bar{t} + \geq 1b$ normalisation	Free-floating		$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1c$ normalisation	Up or down by 50%		$t\bar{t} + \geq 1c$
NLO matching	MadGraph5_aMC@NLO +Pythia8 $t\bar{t}$ (5FS)	vs. PowhegBox+Pythia8 $t\bar{t}$ (5FS)	All
PS & hadronisation	PowhegBox+Herwig7 $t\bar{t}$ (5FS)	vs. PowhegBox+Pythia8 $t\bar{t}$ (5FS)	All
ISR	Varying α_S^{ISR} (PS), μ_R&μ_F (ME)	in PowhegBox+Pythia8 $t\bar{t}b\bar{b}$ (4FS)	$t\bar{t} + \geq 1b$
		in PowhegBox+Pythia8 $t\bar{t}$ (5FS)	$t\bar{t} + \geq 1c$, $t\bar{t}$ + light
FSR	Varying α_S^{FSR} (PS)	in PowhegBox+Pythia8 $t\bar{t}b\bar{b}$ (4FS)	$t\bar{t} + \geq 1b$
		in PowhegBox+Pythia8 $t\bar{t}$ (5FS)	$t\bar{t} + \geq 1c$, $t\bar{t}$ + light

ttH analysis - signal extraction

Ingredients

- **BDT** distributions in SRs;
- **Yields** in CRs;
- **P** depends on estimated number of events in each bin (function of μ);
- **θ** = set of parameters to model the systematics uncertainties (nuisance parameters);
- **hypotesis**: S+B or onlyB?

Recipe

In order to **test for signal presence** in the channel:

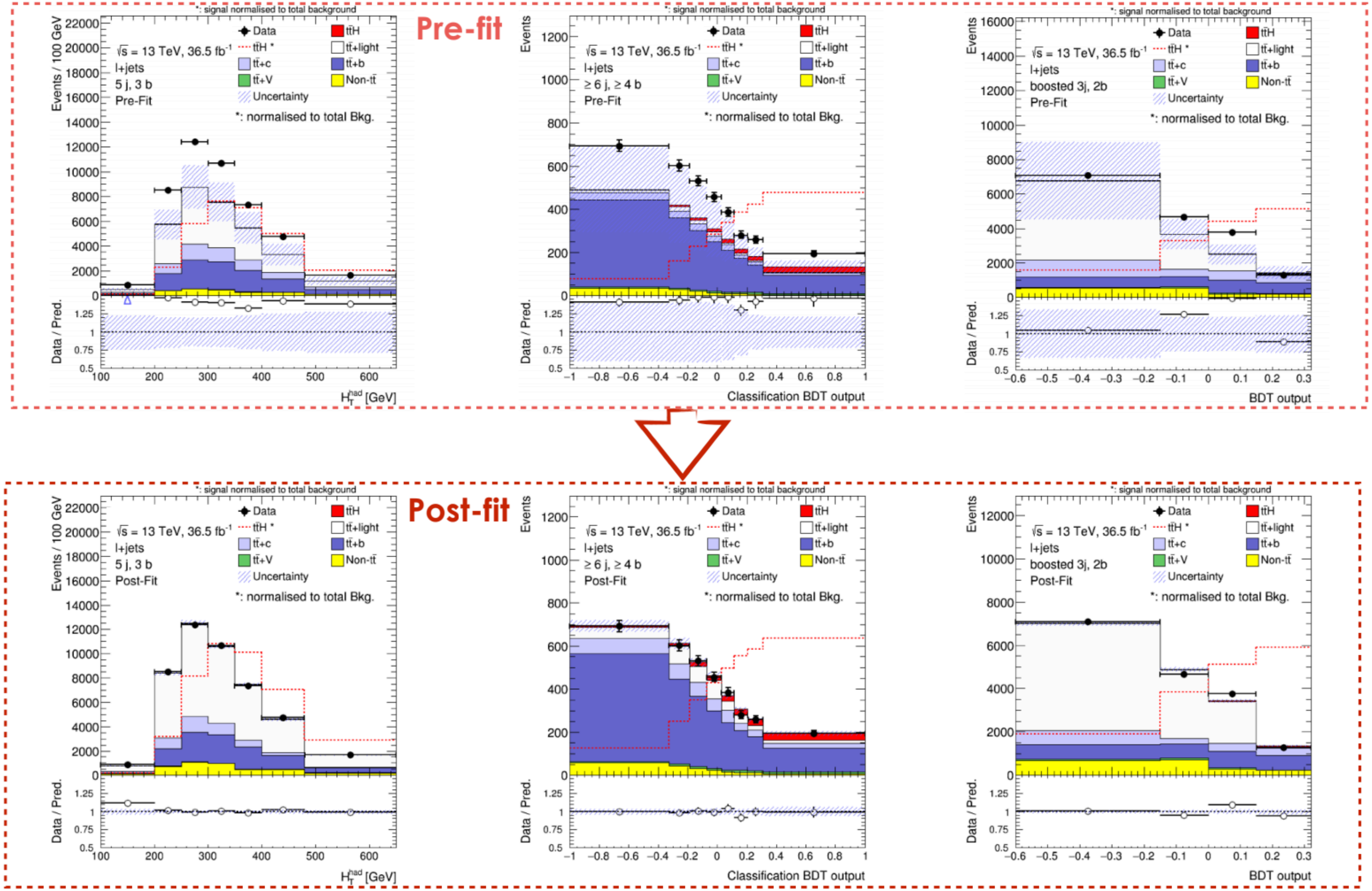
1. build a **likelihood** as a product of P terms over all the bins of the distributions;
2. perform a **simultaneous fit** in the signal and control regions;
3. put a **limit on the signal strength** **$\mu = \sigma/\sigma_{SM}$** . (if the process has not been observed yet)

Region	Fitted observable
$SR^{2\ell}_{\geq 4j \geq 4b}$	Classification BDT
$SR^{2\ell}_{\geq 4j \geq 4b\,low}$	Classification BDT
$CR^{2\ell}_{\geq 4j 3b}$	Event yield
$CR^{2\ell}_{\geq 4j 3b\,low}$	Event yield
$CR^{2\ell}_{3j 3b}$	Event yield
$SR^{1\ell}_{\geq 6j \geq 4b}$	Classification BDT
$SR^{1\ell}_{\geq 6j \geq 4b\,low}$	Classification BDT
$SR^{1\ell}_{5j \geq 4b}$	Classification BDT
$SR^{1\ell}_{5j \geq 4b\,low}$	Classification BDT
$SR^{1\ell}_{boosted}$	Classification BDT

- **NPs θ** adjust the expectations for S and B according to the corresponding syst uncertainties, and their fitted values correspond to the amount that best fits the data:
  - impact of syst uncertainties on the search sensitivity can be reduced by taking advantage of the **highly populated background-dominated CRs included in the likelihood fit**.
- **Normalisation of each background** is determined from the fit simultaneously with μ:
  - Contributions from backgrounds are constrained by the **theoretical uncertainties**, the uncertainty on the **luminosity**, and the **data themselves**.
- **Statistical uncertainties** in each bin of the discriminants are taken into account by **dedicated parameters in the fit**.

ttH analysis - signal extraction

- **Fit effect on distributions and systematic uncertainties:**
  - significant improvement in data/MC agreement from the pre-fit plot to the post-fit one;
  - estimated number of events is in agreement with the number of data in all the regions;
  - the systematics bands are reduced significantly with the fit.



* just for educational purposes!

* these plots are from my PhD thesis and not related to this analysis anymore

ttH analysis - Recent results: evidence

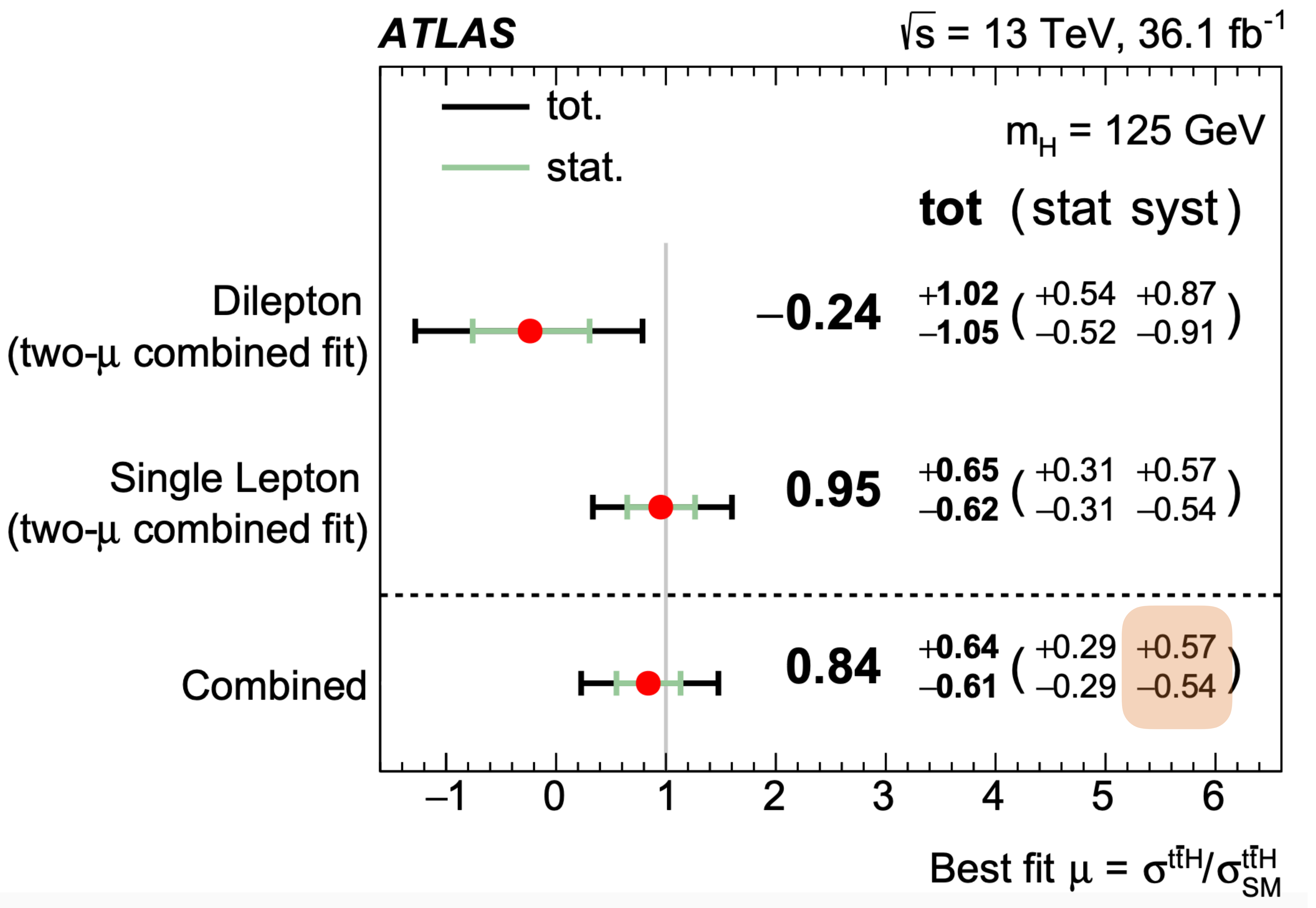
ttH(bb) analysis - 2018

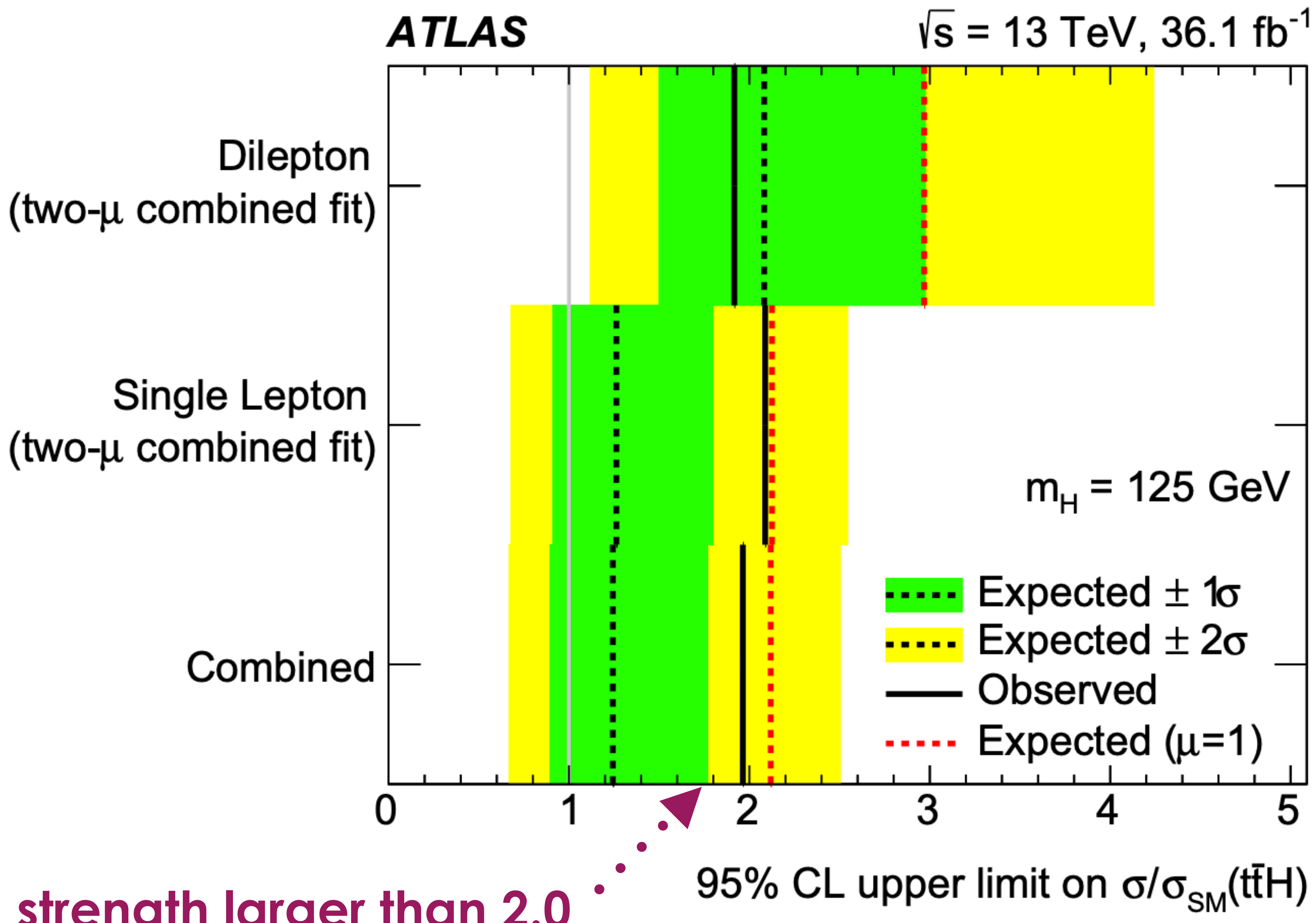
Evidence of ttH process - 2018

$H \rightarrow \gamma\gamma, ZZ^*, b\bar{b}, \mathrm{ML}$

36 fb^{-1}

- Results from the **first evidence of the ttH process at LHC**;
- Free-floating normalisation factors for tt+Heavy Flavour jets:
  - **tt + ≥1b : 1.24 ± 0.10**
  - **tt + ≥1c : 1.63 ± 0.23**
- Best-fit: $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}/\sigma_{SM} = 0.84^{+0.64}_{-0.61}$
- Precision limited by **systematic uncertainty on tt + ≥1b simulation**.





signal strength larger than 2.0 excluded at the 95% CL

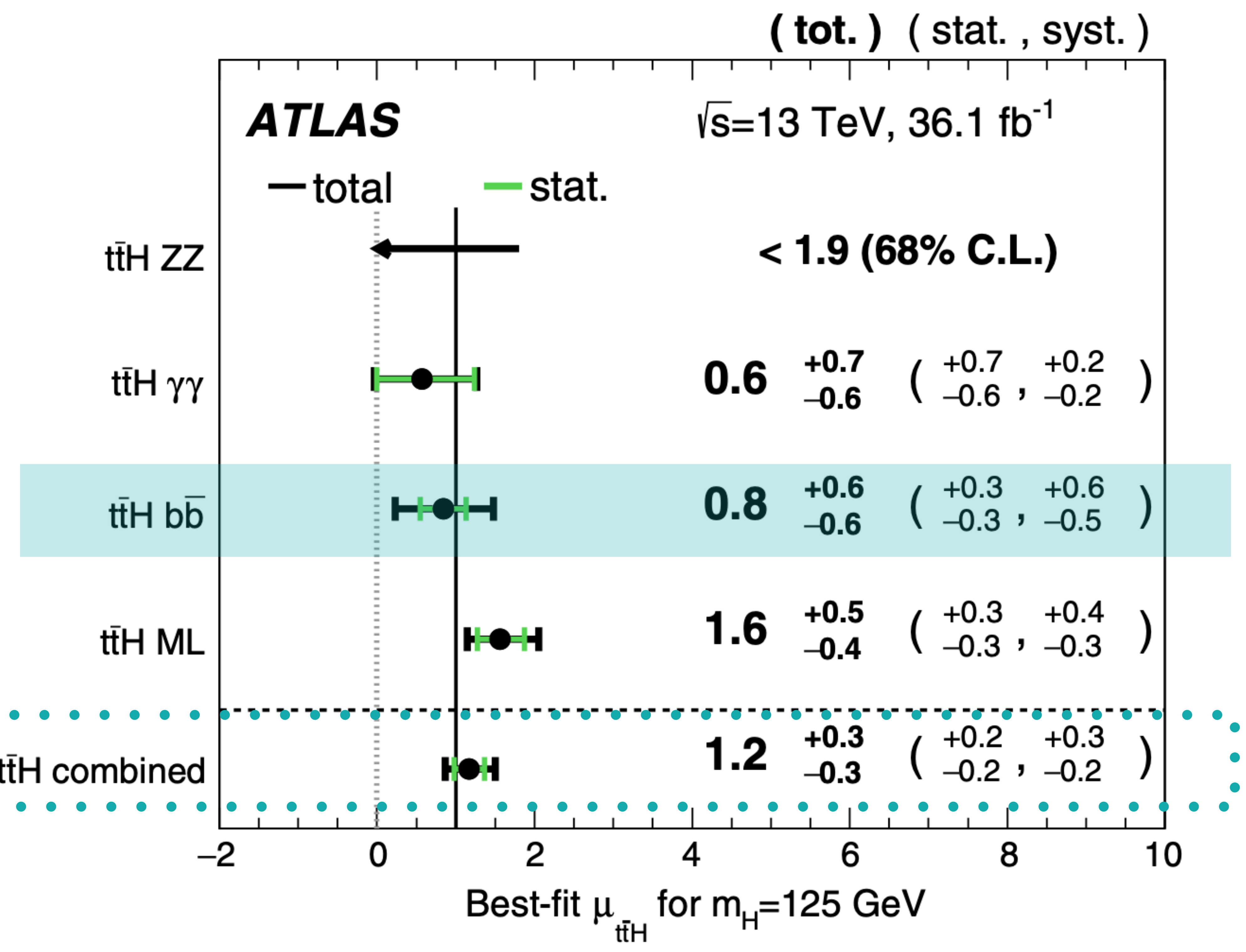
ttH analysis - Recent results: evidence

ttH(bb) analysis - 2018

Evidence of ttH process - 2018

Results

- combination of different channels allowed to reach the first evidence of thee ttH production modes;
- Both systematic and statistical uncertainties limit the measurements;
- measured cross-section compatible with SM prediction.



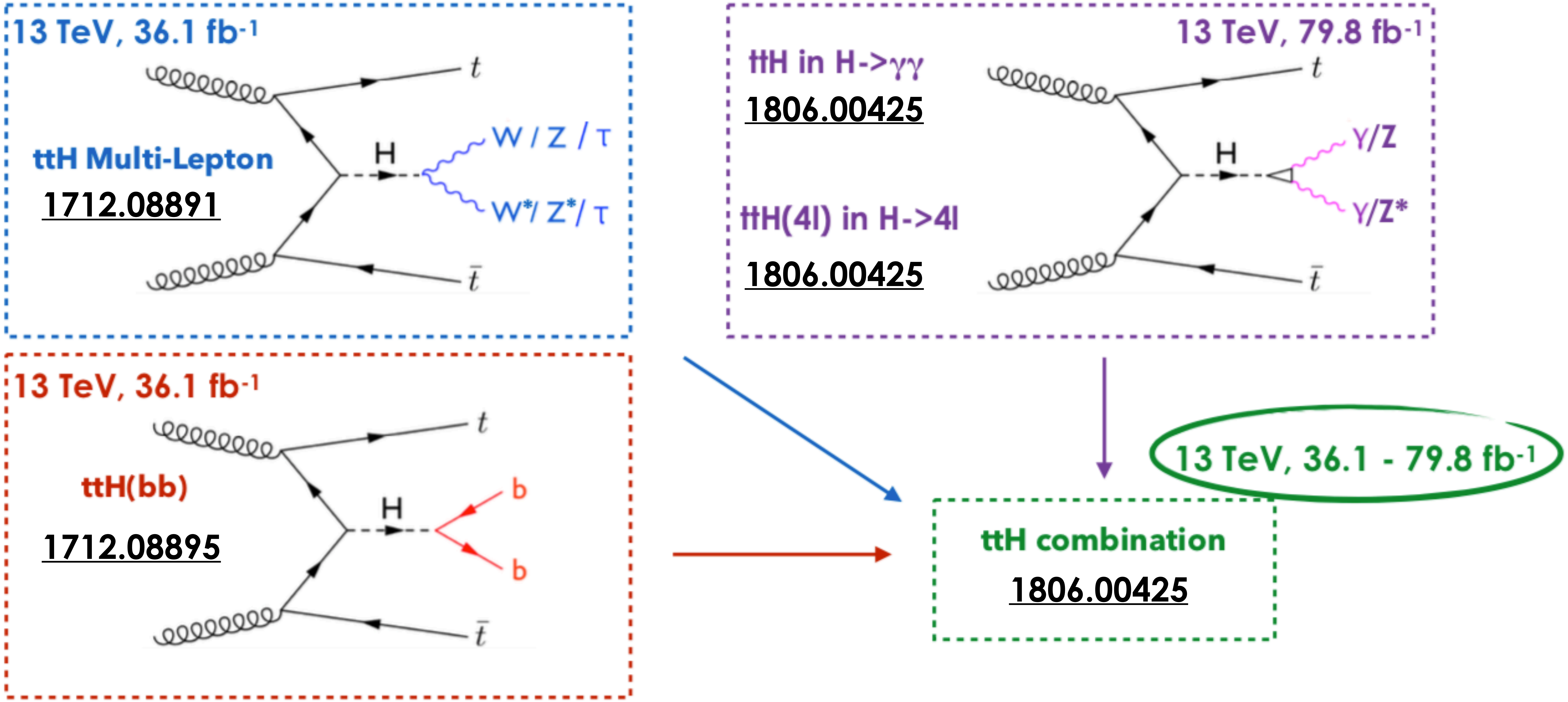
➡ **measured σ_{ttH} = 590 + 160 - 150 fb (SM prediction 507 + 35 -50 fb).**

Channel	Best fit μ Observed	Best fit μ Expected	Significance Observed	Significance Expected
Multilepton	$1.6^{+0.5}_{-0.4}$	$1.0^{+0.4}_{-0.4}$	4.1σ	2.8σ
$H \to b\bar{b}$	$0.8^{+0.6}_{-0.6}$	$1.0^{+0.6}_{-0.6}$	1.4σ	1.6σ
$H \to \gamma\gamma$	$0.6^{+0.7}_{-0.6}$	$1.0^{+0.8}_{-0.6}$	0.9σ	1.7σ
$H \to 4\ell$	< 1.9	$1.0^{+3.2}_{-1.0}$	—	0.6σ
Combined	$1.2^{+0.3}_{-0.3}$	$1.0^{+0.3}_{-0.3}$	4.2σ	3.8σ

Uncertainty source	$\Delta\mu$	
$t\bar{t}$ modeling in $H \to b\bar{b}$ analysis	+0.15	−0.14
$t\bar{t}H$ modeling (cross section)	+0.13	−0.06
Nonprompt light-lepton and fake τ_{had} estimates	+0.09	−0.09
Simulation statistics	+0.08	−0.08
Jet energy scale and resolution	+0.08	−0.07
$t\bar{t}V$ modeling	+0.07	−0.07
$t\bar{t}H$ modeling (acceptance)	+0.07	−0.04
Other non-Higgs boson backgrounds	+0.06	−0.05
Other experimental uncertainties	+0.05	−0.05
Luminosity	+0.05	−0.04
Jet flavor tagging	+0.03	−0.02
Modeling of other Higgs boson production modes	+0.01	−0.01
Total systematic uncertainty	+0.27	−0.23
Statistical uncertainty	+0.19	−0.19
Total uncertainty	+0.34	−0.30

ttH analysis - Recent results: observation

- Results from the **first observation of the ttH process at LHC**

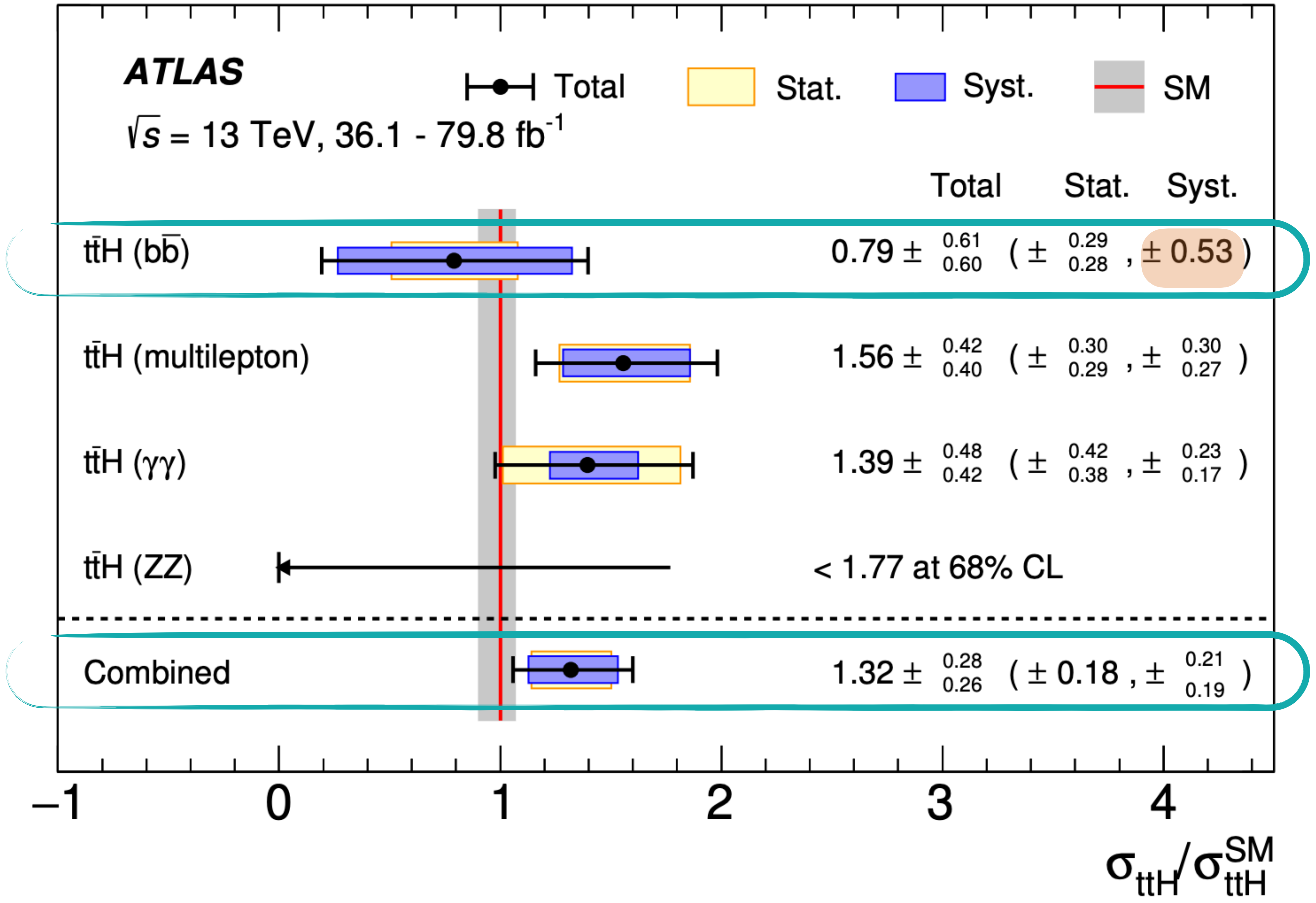


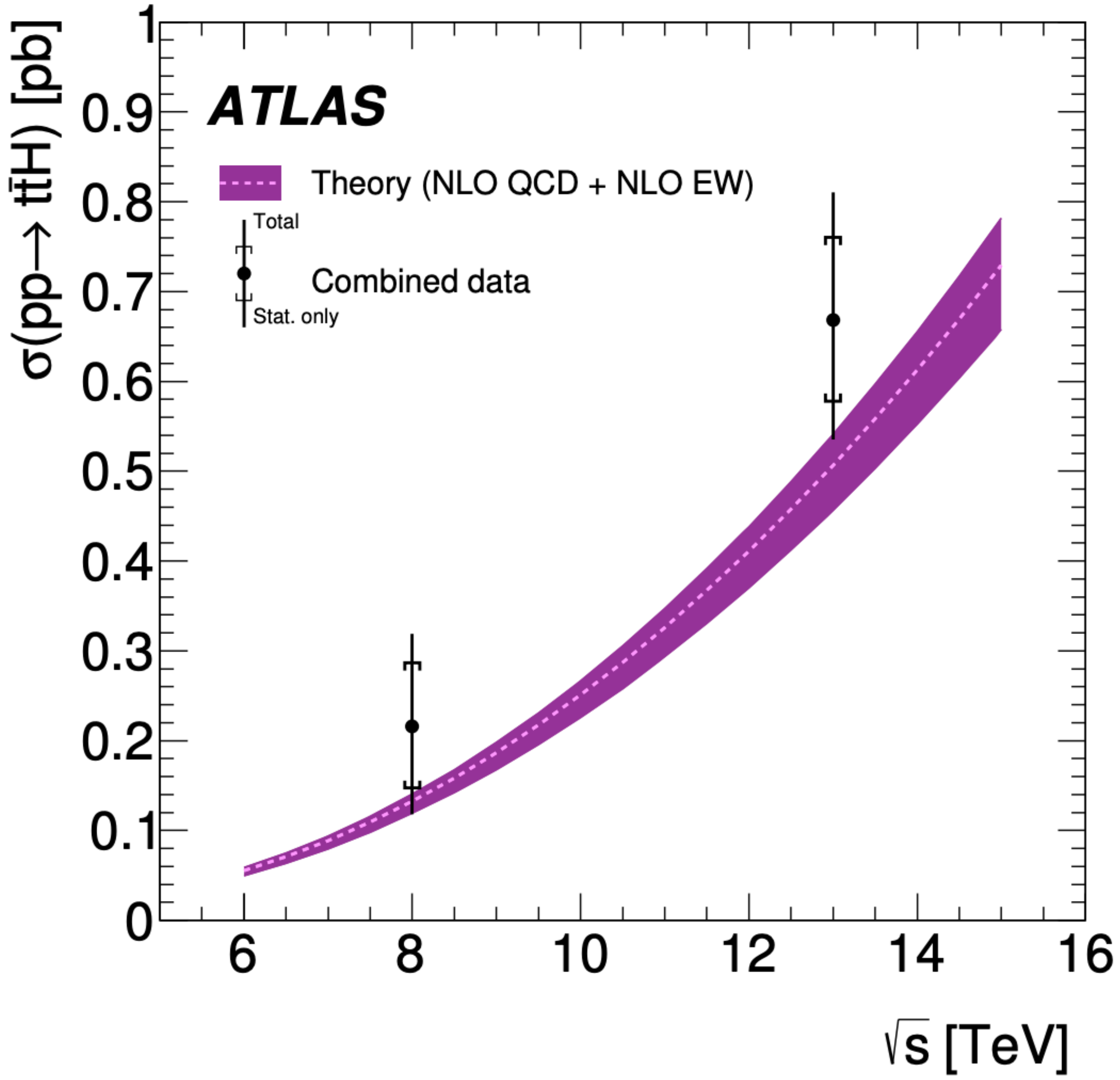
ttH analysis - Recent results: observation

Observation of ttH process - 2018

$H \to \gamma\gamma, ZZ^*, b\bar{b}, \mathrm{ML}$

- Results from the **first observation of the ttH process at LHC**;
  - 36 fb^{-1} for ttH(bb) and ttH multilepton, 79.8 fb^{-1} for ttH(gg) and ttH(4l);
  - combination with Run 1 data (4.5 fb^{-1} @ 7 TeV, 20.3 fb^{-1} @ 8 TeV).
- Best-fit: $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}/\sigma_{SM} = 1.32^{+0.28}_{-0.26}$ → **ttH cross section 1.32 times higher wrt the SM one**
  ➡ still compatible in 20% of the measurement precision
- ttH(bb) precision limited by **systematic uncertainty on tt + ≥1b simulation**.





ttH analysis - Recent results: observation

Observation of ttH process - 2018

- **Summary of the systematic uncertainties affecting the combined ttH cross-section measurement at 13 TeV**
  - only systematic uncertainty sources with at least 1% impact are listed;
  - MC statistical uncertainty is due to limited numbers of simulated events.

Uncertainty source	$\Delta\sigma_{t\bar{t}H}/\sigma_{t\bar{t}H}$ [%]
Theory uncertainties (modelling)	11.9
$t\bar{t}$ + heavy flavour	9.9
$t\bar{t}H$	6.0
Non-$t\bar{t}H$ Higgs boson production	1.5
Other background processes	2.2
Experimental uncertainties	9.3
Fake leptons	5.2
Jets, $E_{\mathrm{T}}^{\mathrm{miss}}$	4.9
Electrons, photons	3.2
Luminosity	3.0
τ-leptons	2.5
Flavour tagging	1.8
MC statistical uncertainties	4.4

- **Measured total ttH production cross sections at 13 TeV**
  - Since no event is observed in the H→ZZ*→4l decay channel, an observed upper limit is set at 68% CL on the ttH cross section in that channel using pseudo-experiments

Analysis	Integrated luminosity [fb^{-1}]	$t\bar{t}H$ cross section [fb]	Obs. sign.	Exp. sign.
$H \to \gamma\gamma$	79.8	$710\ ^{+210}_{-190}$ (stat.) $^{+120}_{-90}$ (syst.)	4.1σ	3.7σ
$H \to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	4.1σ	2.8σ
$H \to b\bar{b}$	36.1	$400\ ^{+150}_{-140}$ (stat.) $\pm$ 270 (syst.)	1.4σ	1.6σ
$H \to ZZ^* \to 4\ell$	79.8	<900 (68% CL)	0σ	1.2σ
Combined (13 TeV)	36.1–79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	5.8σ	4.9σ
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1–79.8	–	6.3σ	5.1σ

observed significance 5.8σ (4.9σ exp)

ttH analysis - Recent results: observation

Observation of ttH process - 2018

- **Summary of the systematic uncertainties affecting the combined ttH cross-section measurement at 13 TeV**
  - only systematic uncertainty sources with at least 1% impact are listed;
  - MC statistical uncertainty is due to limited numbers of simulated events.

Uncertainty source	$\Delta\sigma_{t\bar{t}H}/\sigma_{t\bar{t}H}$ [%]
Theory uncertainties (modelling)	11.9
$t\bar{t}$ + heavy flavour	9.9
$t\bar{t}H$	6.0
Non-$t\bar{t}H$ Higgs boson production	1.5
Other background processes	2.2
Experimental uncertainties	9.3
Fake leptons	5.2
Jets, $E_{\mathrm{T}}^{\mathrm{miss}}$	4.9
Electrons, photons	3.2
Luminosity	3.0
τ-leptons	2.5
Flavour tagging	1.8
MC statistical uncertainties	4.4

- **Measured total ttH production cross sections at 13 TeV**
  - Since no event is observed in the H→ZZ*→4l decay channel, an observed upper limit is set at 68% CL on the ttH cross section in that channel using pseudo-experiments.

Analysis	Integrated luminosity [fb^{-1}]	$t\bar{t}H$ cross section [fb]	Obs. sign.	Exp. sign.
$H\to\gamma\gamma$	79.8	$710\,^{+210}_{-190}$ (stat.) $^{+120}_{-90}$ (syst.)	4.1σ	3.7σ
$H\to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	4.1σ	2.8σ
$H\to b\bar{b}$	36.1	$400\,^{+150}_{-140}$ (stat.) $\pm$ 270 (syst.)	1.4σ	1.6σ
$H\to ZZ^*\to 4\ell$	79.8	<900 (68% CL)	0σ	1.2σ
Combined (13 TeV)	36.1–79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	5.8σ	4.9σ
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1–79.8	–	6.3σ	5.1σ

Combination with Run 1

↓

observed significance 6.3σ (5.1σ exp)

Supporting material

ttH analysis - fitted observables

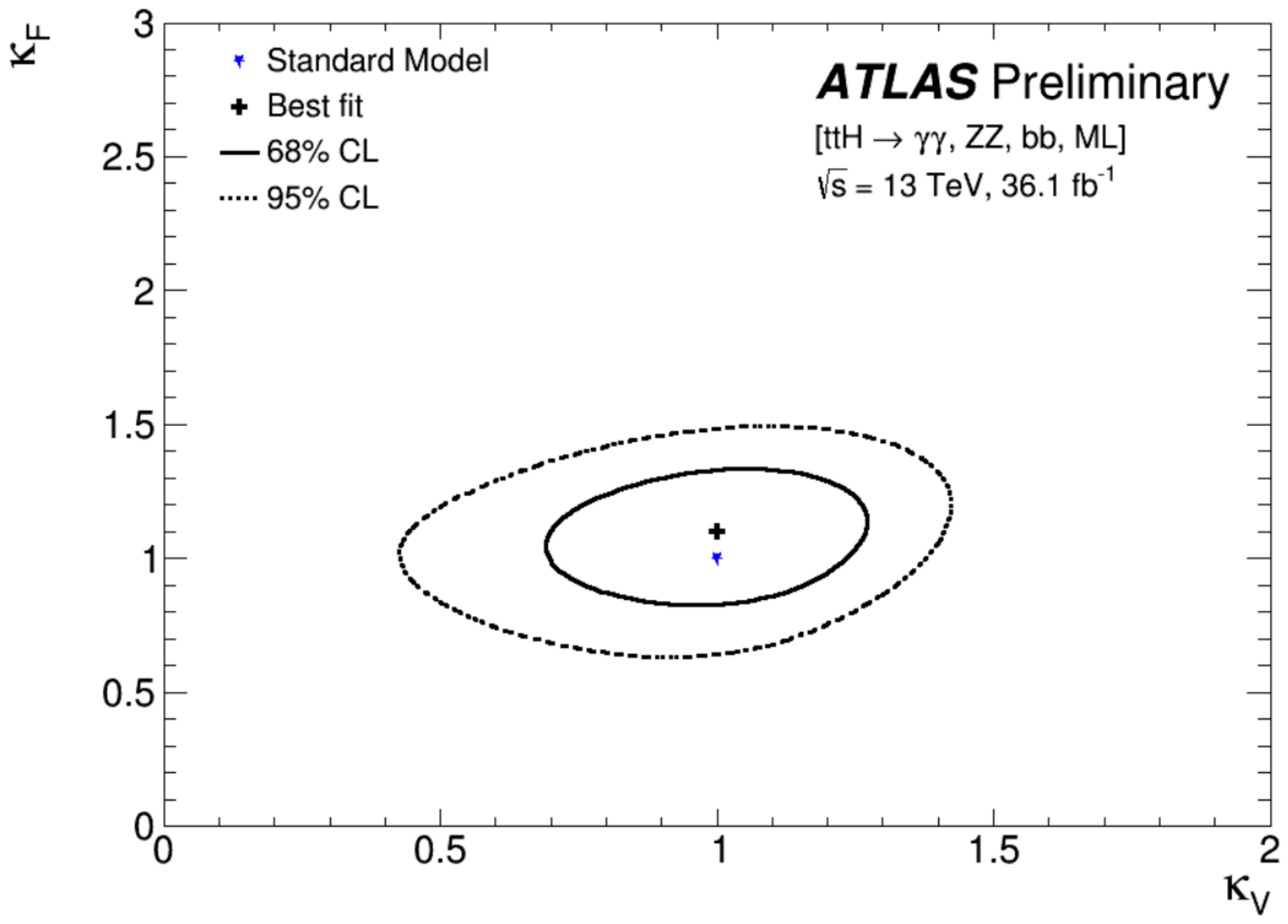
Years	$\int L dt$ [fb^{-1}]	Uncertainty [%]
	Leptonic channels	
2015-2016	36.2	2.1
2017	44.3	2.4
2015-2017	80.5	2.0
2018	58.5	2.0
2015-2018	139.0	1.7

ttH analysis - Recent results: evidence

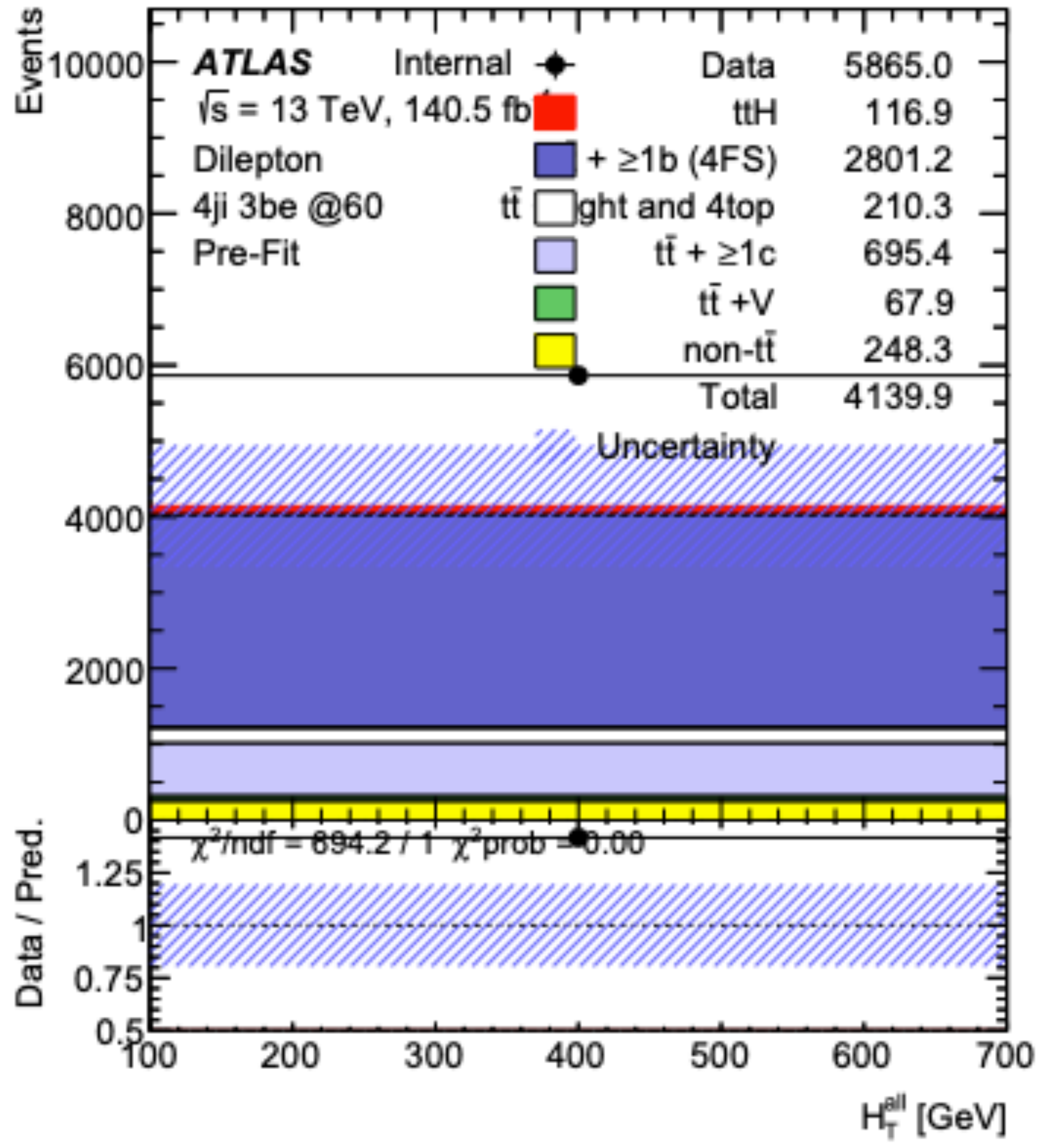
$$H \to \gamma\gamma, ZZ^*, b\bar{b}, \mathrm{ML}$$

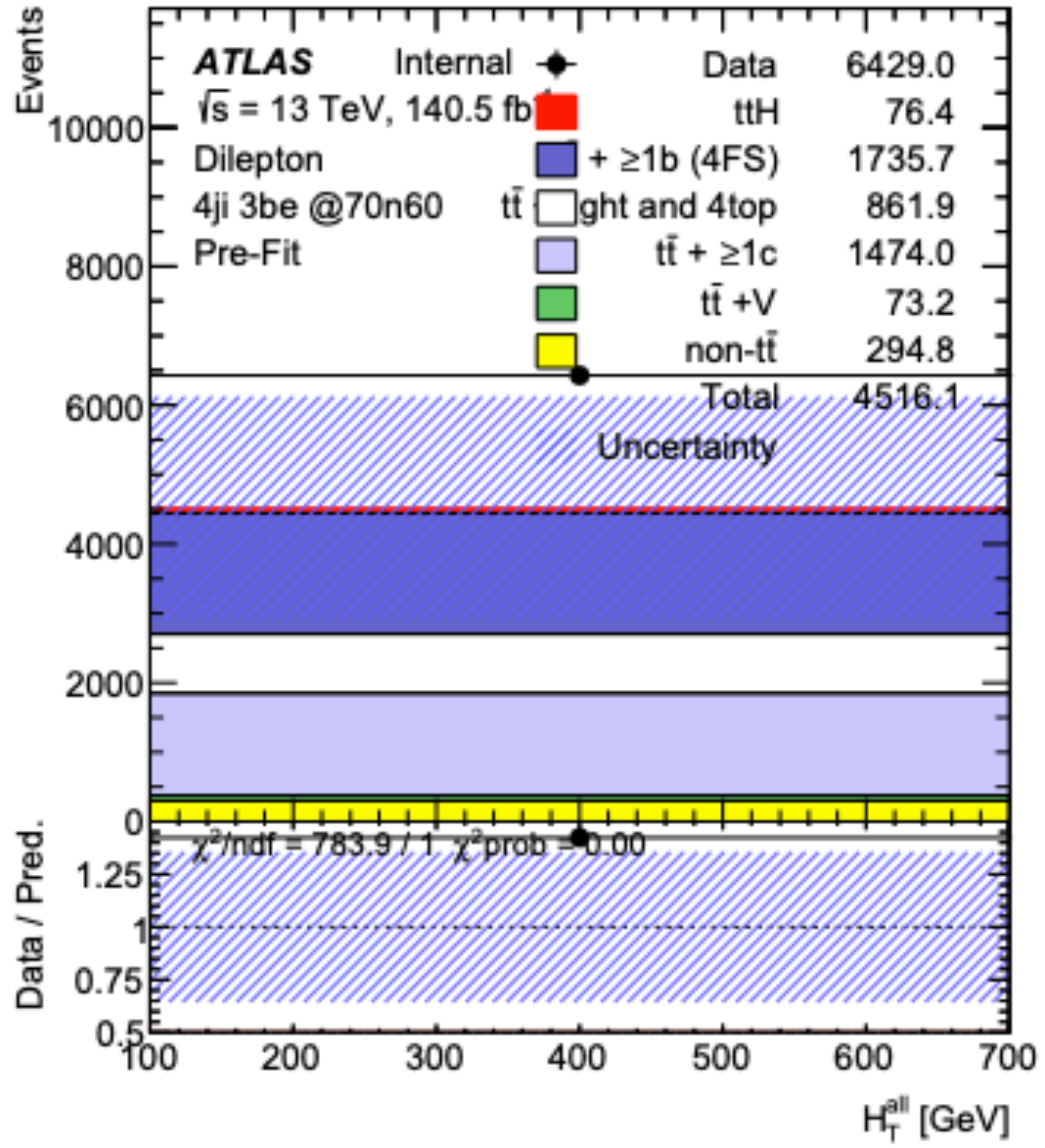
Coupling studies performed as well

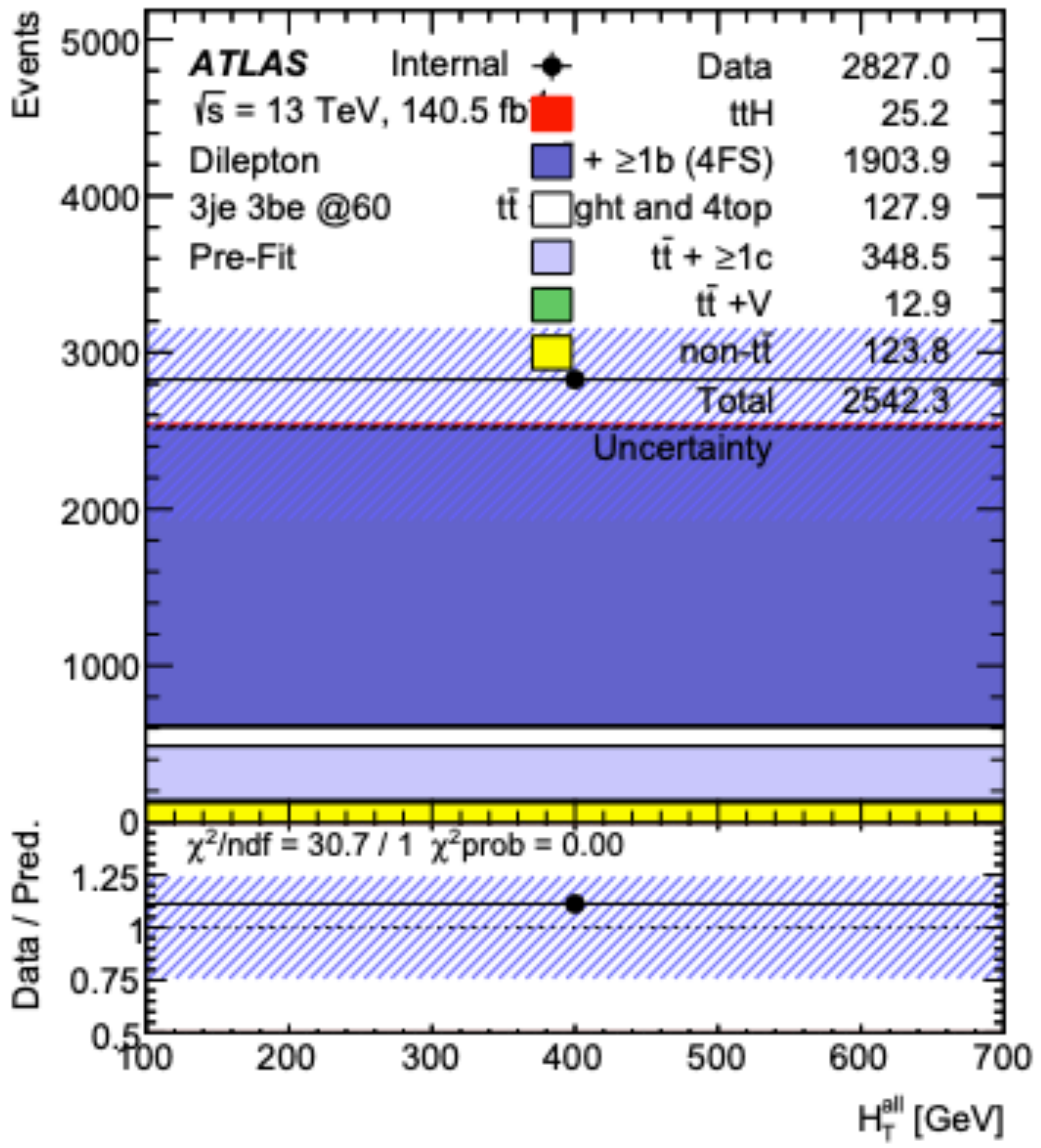
- scan in the k_F-k_V plane from the combination of all ttH channels;
- assuming Higgs boson not to couple to any BSM particles;
- good agreement with the SM prediction.



ttH analysis - fitted observables







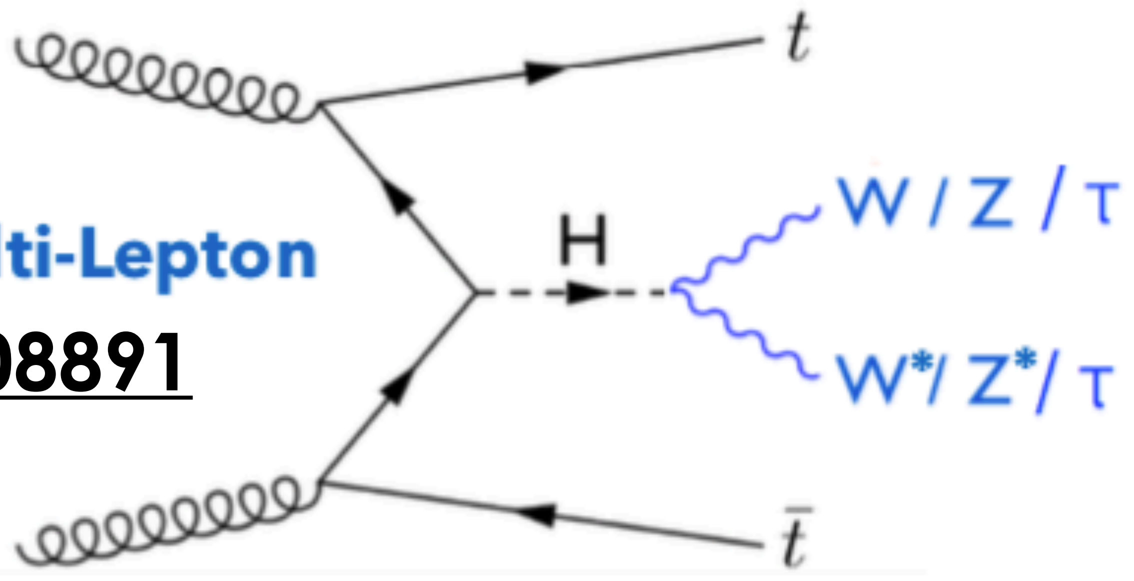
- Event yields in CRs of dilepton channel.

ttH analysis - observation

13 TeV, 36.1 fb^{-1}

ttH Multi-Lepton

1712.08891



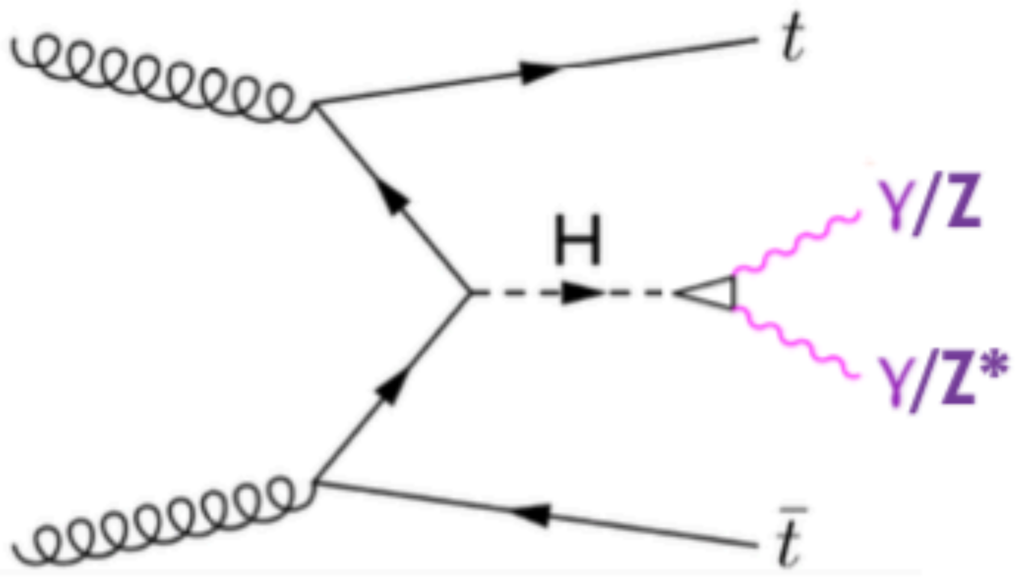
13 TeV, 79.8 fb^{-1}

ttH in H->γγ

1806.00425

ttH(4l) in H->4l

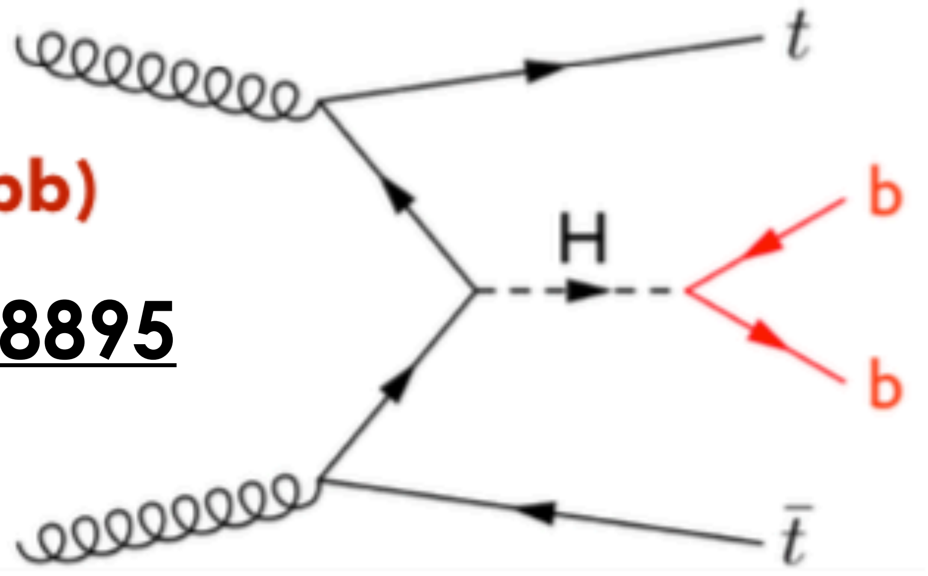
1806.00425



13 TeV, 36.1 fb^{-1}

ttH(bb)

1712.08895



13 TeV, 36.1 - 79.8 fb^{-1}

ttH combination

1806.00425

ttH analysis - observation

Bin	Expected $t\bar{t}H$ (signal)	Non-$t\bar{t}H$ Higgs	Non-Higgs	Total	Observed Total
			$H \to \gamma\gamma$		
Had 1	4.2 ± 1.1	0.49 ± 0.33	1.8 ± 0.5	6.4 ± 1.3	10
Had 2	3.4 ± 0.7	0.7 ± 0.6	7.5 ± 1.1	11.6 ± 1.5	14
Had 3	4.7 ± 0.9	2.0 ± 1.7	32.9 ± 2.2	39.6 ± 3.2	47
Had 4	3.0 ± 0.5	3.2 ± 3.1	55.0 ± 2.8	61 ± 5	67
Lep 1	4.5 ± 1.0	0.24 ± 0.09	2.2 ± 0.6	6.9 ± 1.2	7
Lep 2	2.2 ± 0.4	0.27 ± 0.10	4.6 ± 0.9	7.1 ± 1.0	7
Lep 3	0.82 ± 0.18	0.30 ± 0.13	4.6 ± 0.9	5.7 ± 0.9	5
			$H \to ZZ^* \to 4\ell$		
Had 1	0.169 ± 0.031	0.021 ± 0.007	0.008 ± 0.008	0.198 ± 0.033	0
Had 2	0.216 ± 0.032	0.20 ± 0.09	0.22 ± 0.12	0.63 ± 0.16	0
Lep	0.212 ± 0.031	0.0256 ± 0.0023	0.015 ± 0.013	0.253 ± 0.034	0