# IL PROGRAMMA DI FISICA DI LHC: QUALI ANALISI, COME FARLE E PERCHE'?

## INCONTRI DI FISICA DI FRASCATI 2019

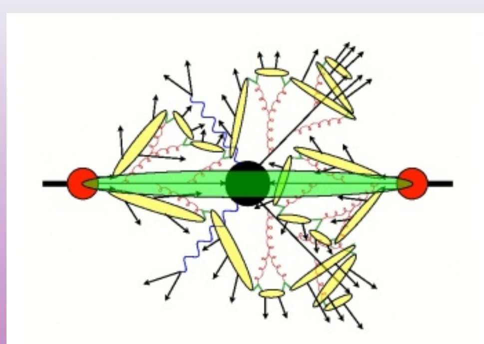
Patrizia Azzi - INFN Padova

# OUTLINE

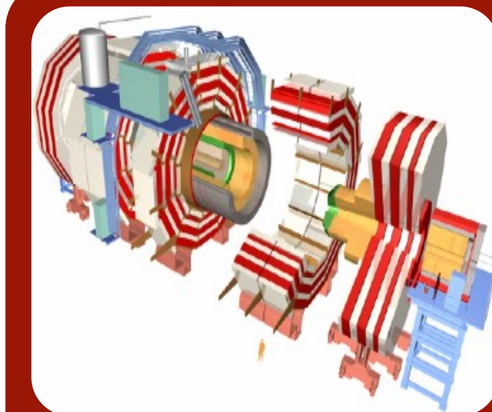
- The LHC physics program
- What is important in a physics analysis:
  - Know your signal and backgrounds. Example: H->bb
  - Control your uncertainties. Example: the W mass
  - Expand your searches (also in the future). Example: BSM search
- Conclusions

# THE ANALYSIS FLOW

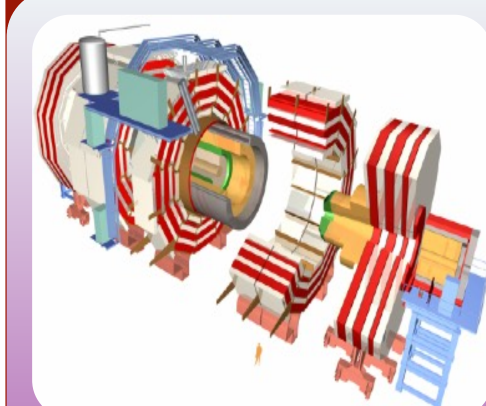

**Collisions!**



**Monte Carlo Event Generator**



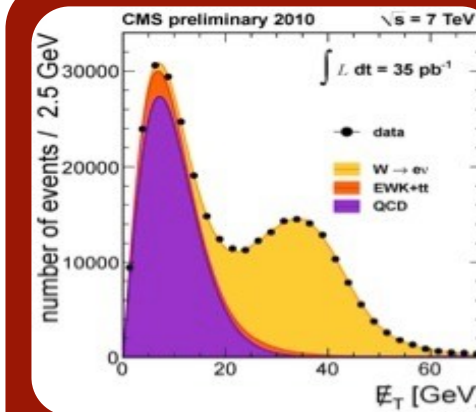
**Detector Response**



**Simulation of the Detector Response**

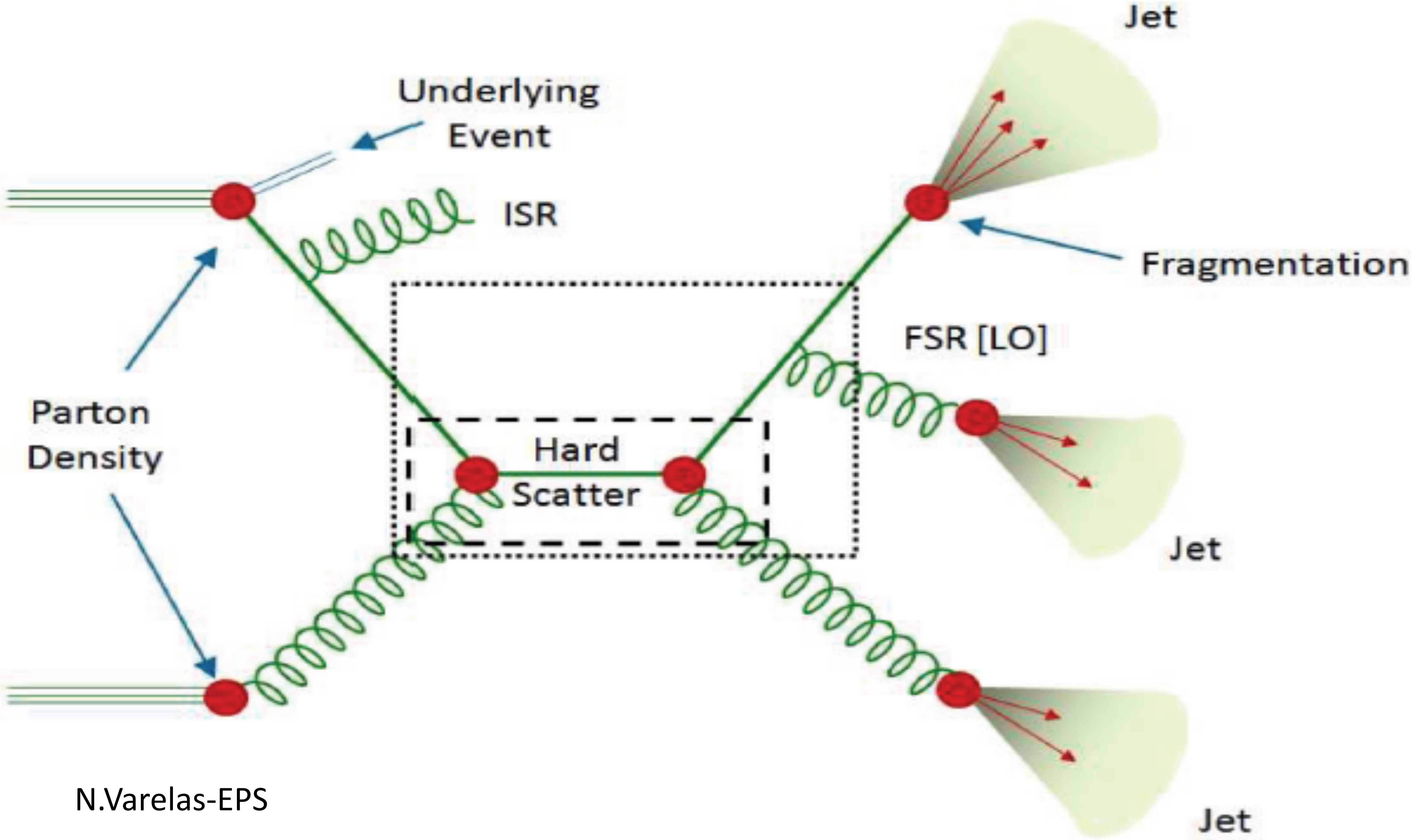

**Trigger**

**Event Reconstruction**



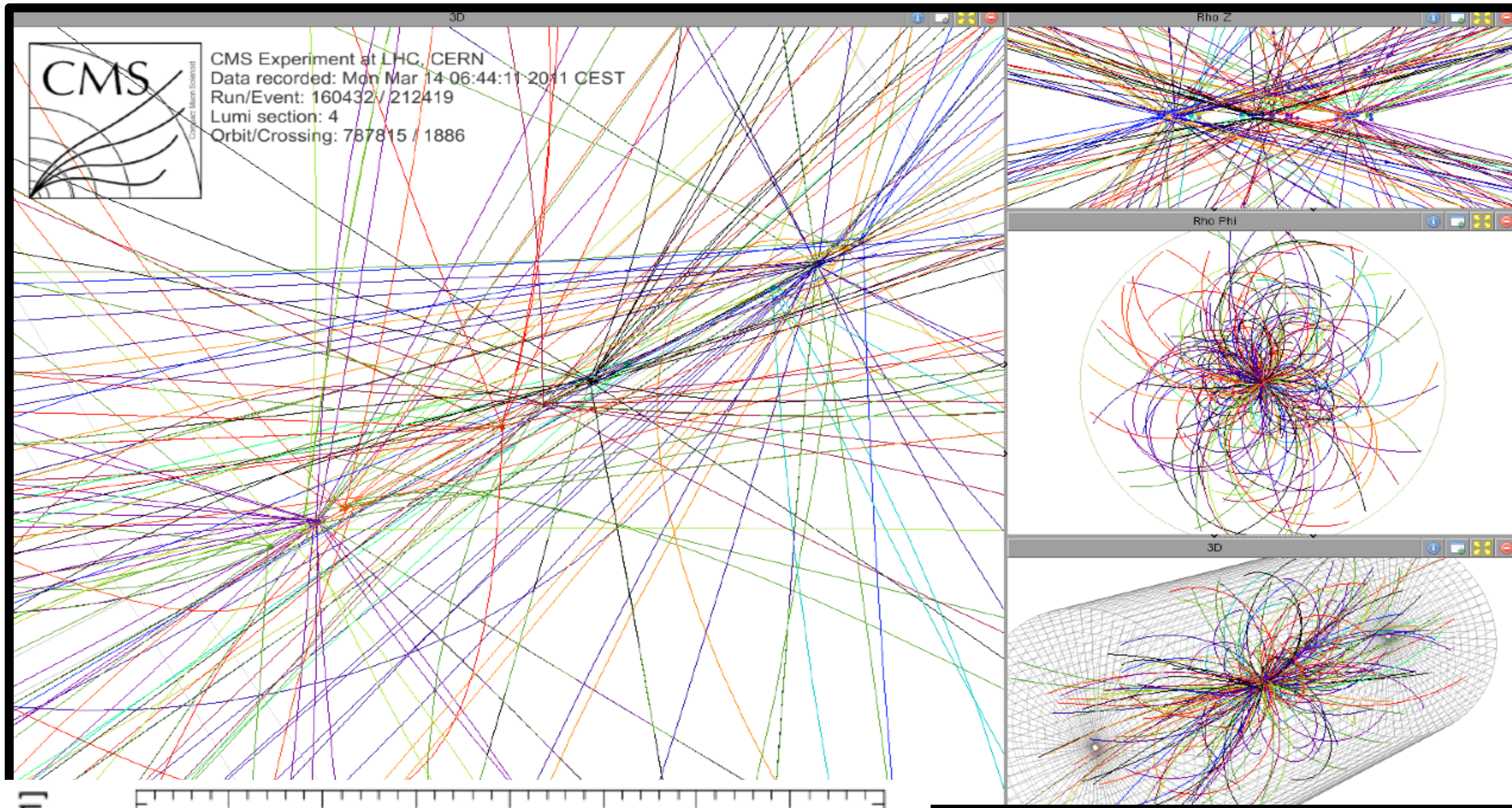

**Physics Analysis**

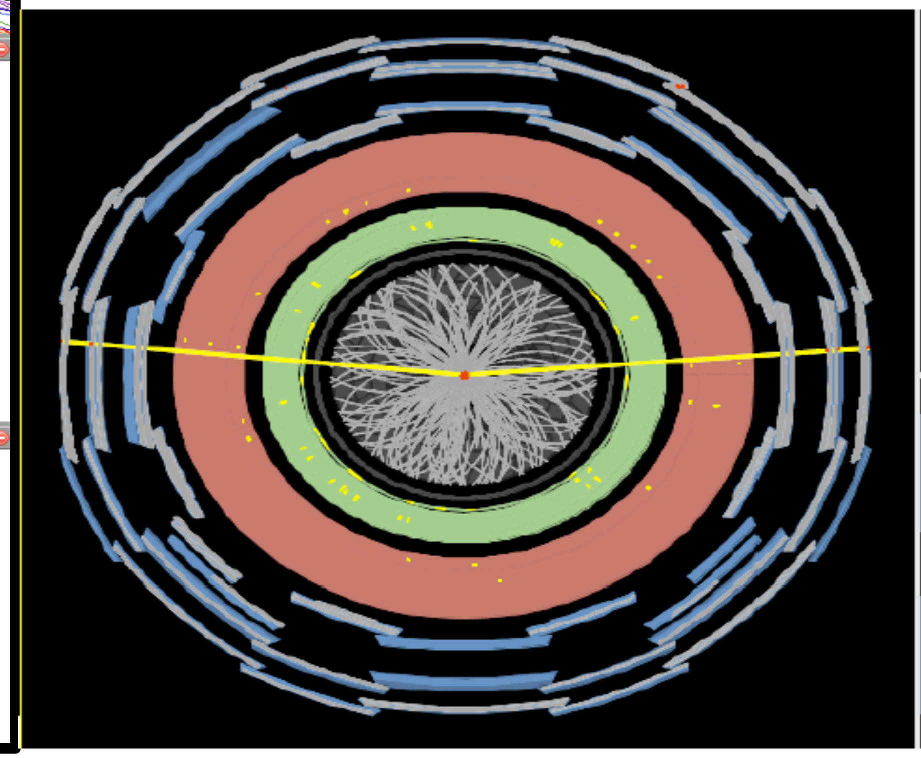
# QCD AT LHC

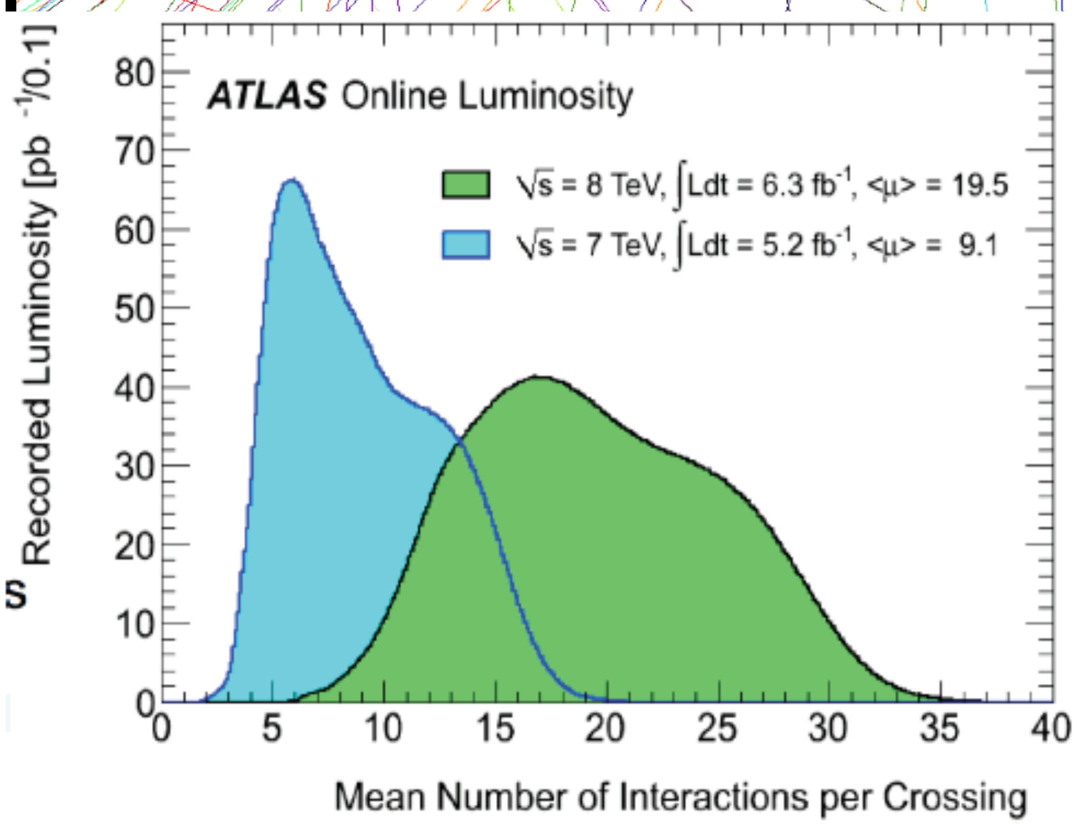


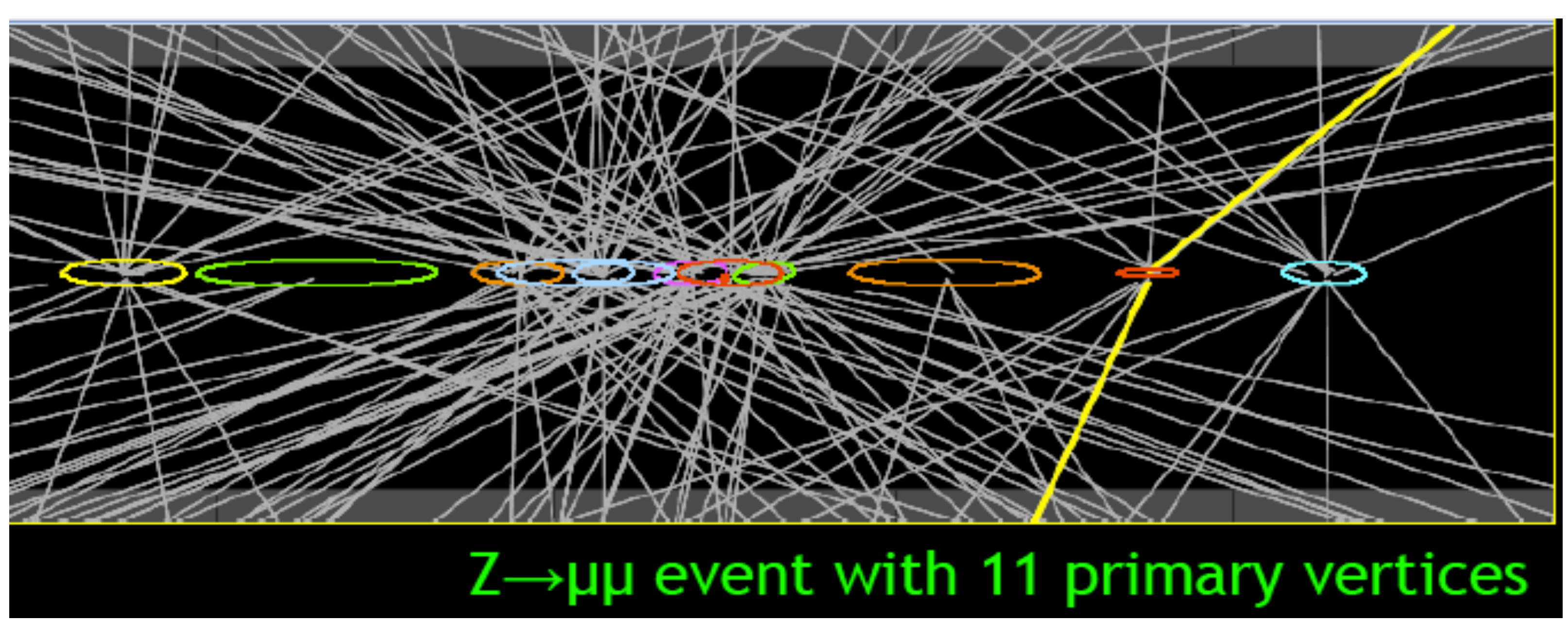

N.Varelas-EPS

# PILE-UP


CMS
CMS Experiment at LHC, CERN
Data recorded: Mon Mar 14 06:44:11 2011 CEST
Run/Event: 160432 / 212419
Lumi section: 4
Orbit/Crossing: 787815 / 1886
3D
Rho Z
Rho Phi
3D

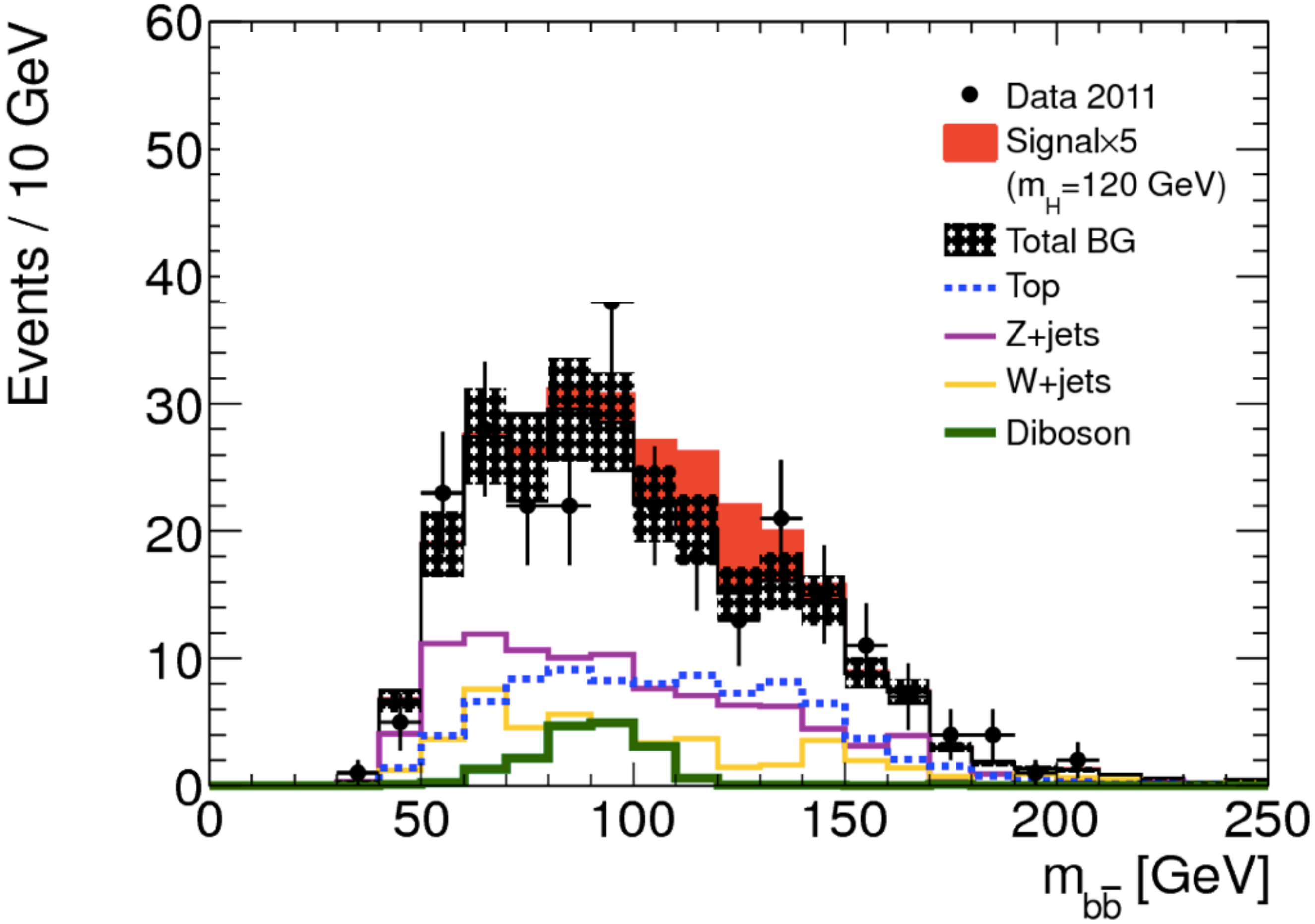



ATLAS Online Luminosity
√s = 8 TeV, ∫Ldt = 6.3 fb⁻¹, <μ> = 19.5
√s = 7 TeV, ∫Ldt = 5.2 fb⁻¹, <μ> = 9.1
Recorded Luminosity [pb⁻¹/0.1]
Mean Number of Interactions per Crossing
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30
35
40


Z→μμ event with 11 primary vertices

# BASICS OF A TYPICAL ANALYSIS

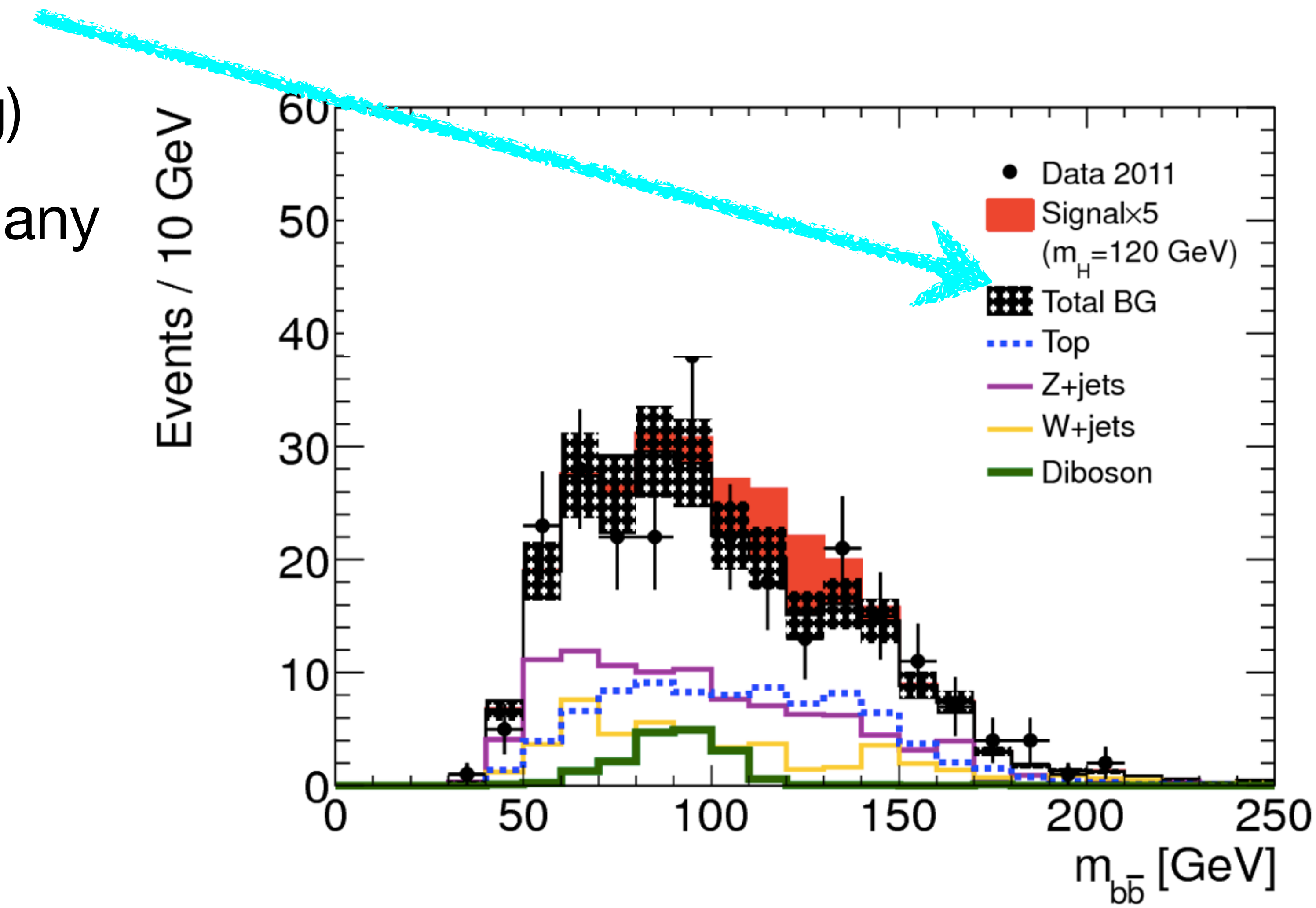
# BASICS OF A TYPICAL ANALYSIS



# BASICS OF A TYPICAL ANALYSIS

Design a selection at a given mass maximizing an estimator (eg s/√bkg)

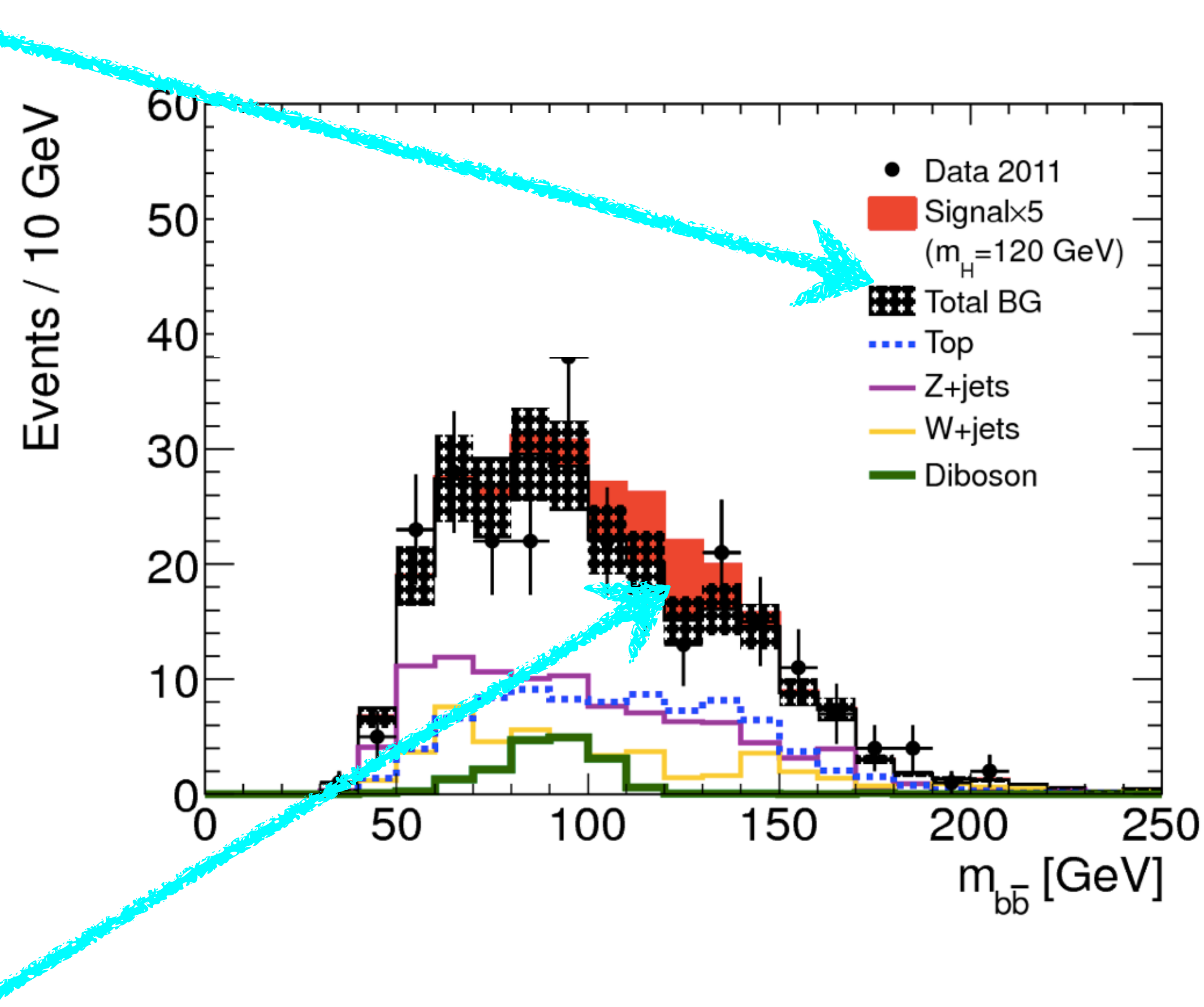
Often cutting the phase-space in many regions



# BASICS OF A TYPICAL ANALYSIS

Design a selection at a given mass maximizing an estimator (eg s/√bkg)

Often cutting the phase-space in many regions

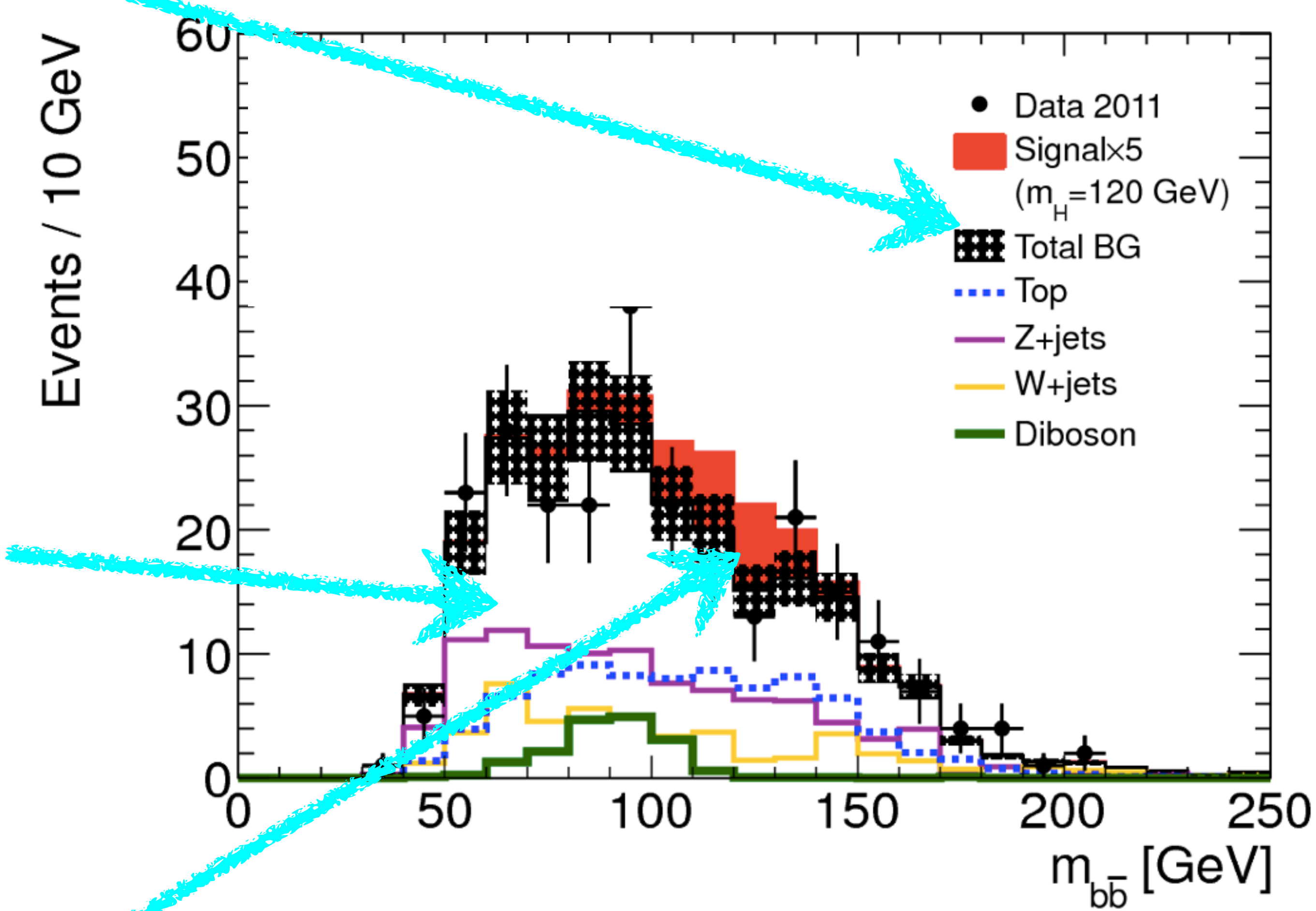



Evaluate the signal efficiency using SM Higgs MC simulation

# BASICS OF A TYPICAL ANALYSIS

Design a selection at a given mass maximizing an estimator (eg s/√bkg)

Often cutting the phase-space in many regions



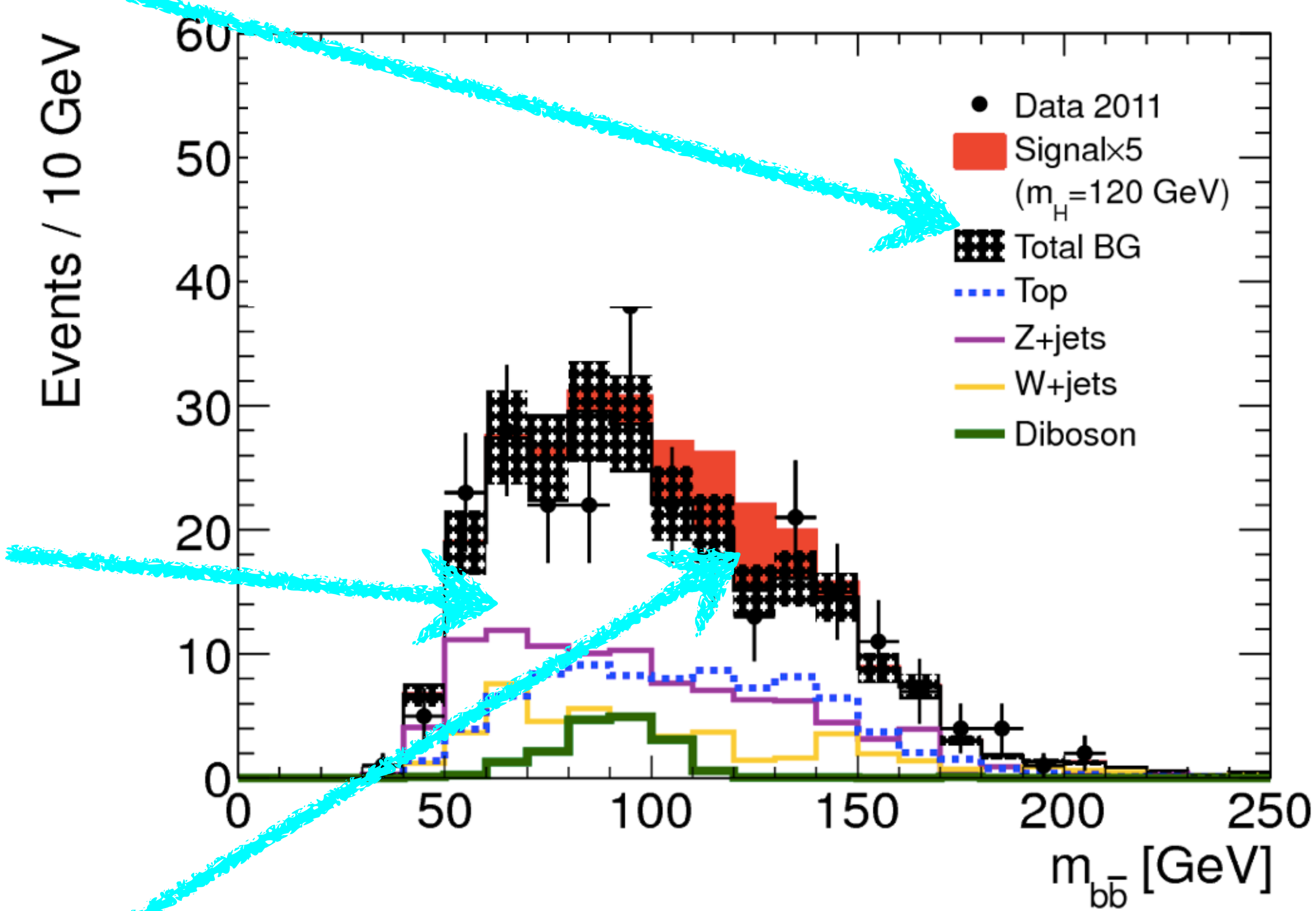

Compute the expected SM background from control samples side-bands, etc. often with the help from MC simulation (shapes). Assess the systematic error.

Evaluate the signal efficiency using SM Higgs MC simulation

# BASICS OF A TYPICAL ANALYSIS

Design a selection at a given mass maximizing an estimator (eg s/√bkg)

Often cutting the phase-space in many regions




Compute the expected SM background from control samples side-bands, etc. often with the help from MC simulation (shapes). Assess the systematic error.

Evaluate the signal efficiency using SM Higgs MC simulation

Compute with statistical methods the largest signal cross section one can accommodate in the data.

# A CHALLENGING HIGGS PHYSICS ANALYSIS H->BB

# MAIN CONCEPTS TO FOCUS ON

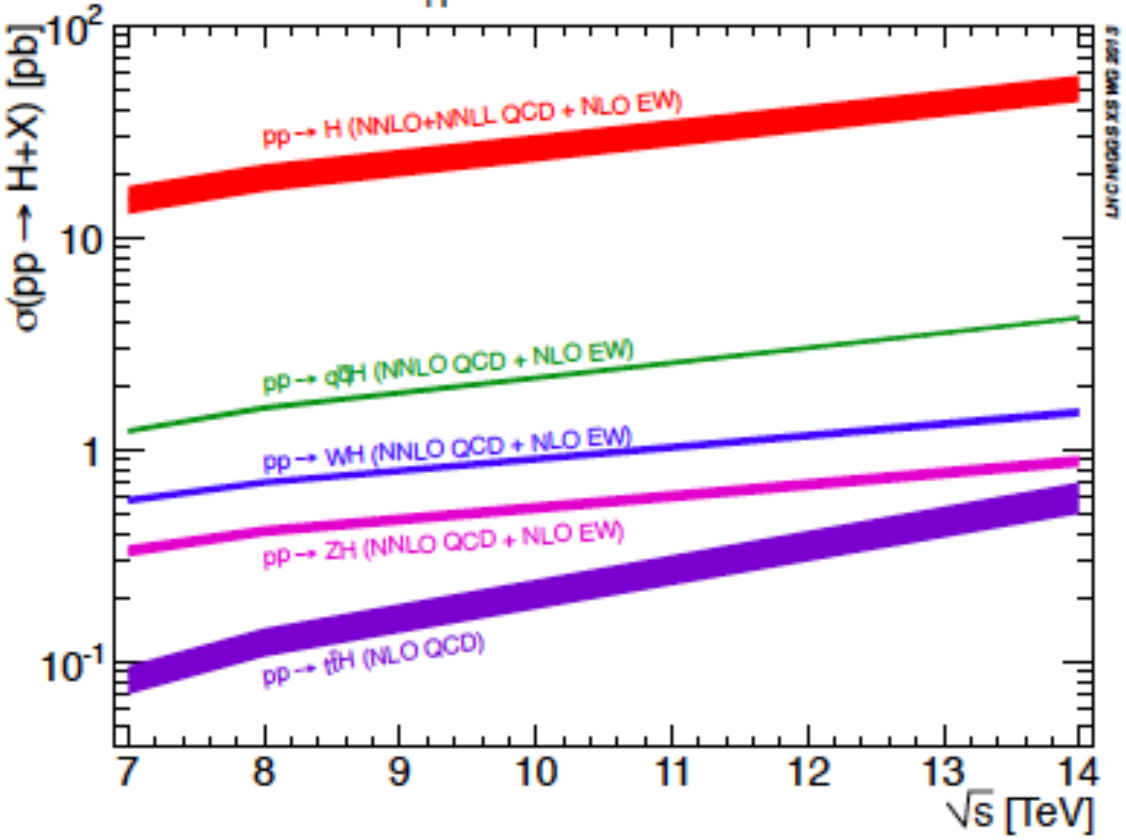
- We want to measure a signal for the first time. we know it's there, but it is difficult to measure.

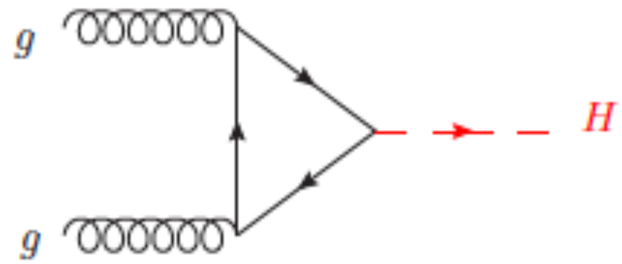
- Background >> signal
  - Need to find ways to extract the signal from a lot of « background »
  - Need to find ways to make sure we know precisely how much « background » is left to measure how much signal we have

**learn what is « background » and how to control it**

# Higgs Production Modes

k for $m_H$ = 125.5 GeV

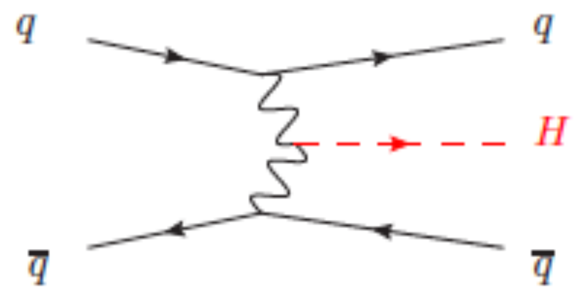



Gluon fusion process

NNnLO ~O(10%)

σ(prod)=48.6pb

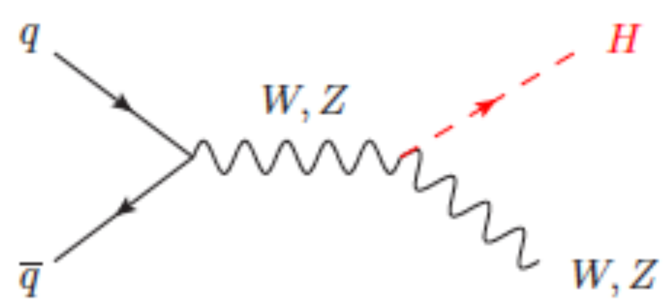



Vector Boson Fusion

NLO TH uncertainty ~O(5%)

Two forward jets and a large rapidity gap

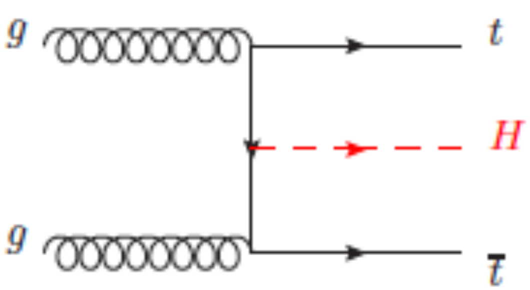
σ(prod)=3.78pb




W and Z Associated Production

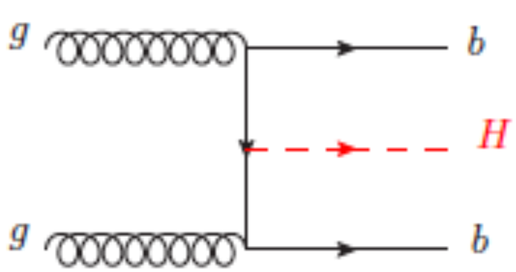
NNLO TH uncertainty ~O(5%)

σ(prod)WH=1.37pb
σ(prod)ZH=0.88

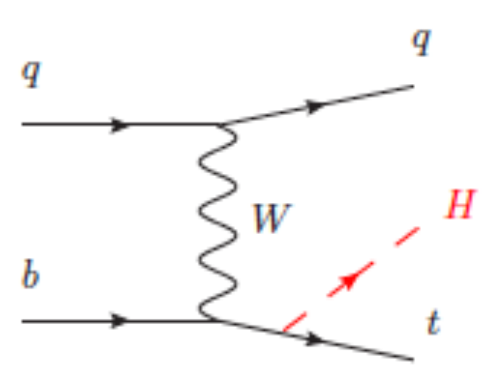



Top Assoc. Prod.

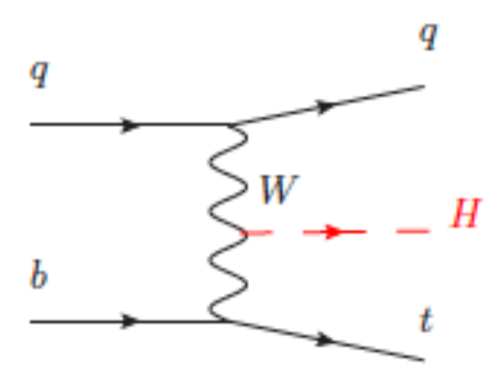
σ(prod)=0.51pb




B-quark Assoc. Prod.




tH




σ(prod)=0.074pb

# Higgs Decay Channels

SWEET SPOT@125 GeV!

- Dominant: bb (57%)

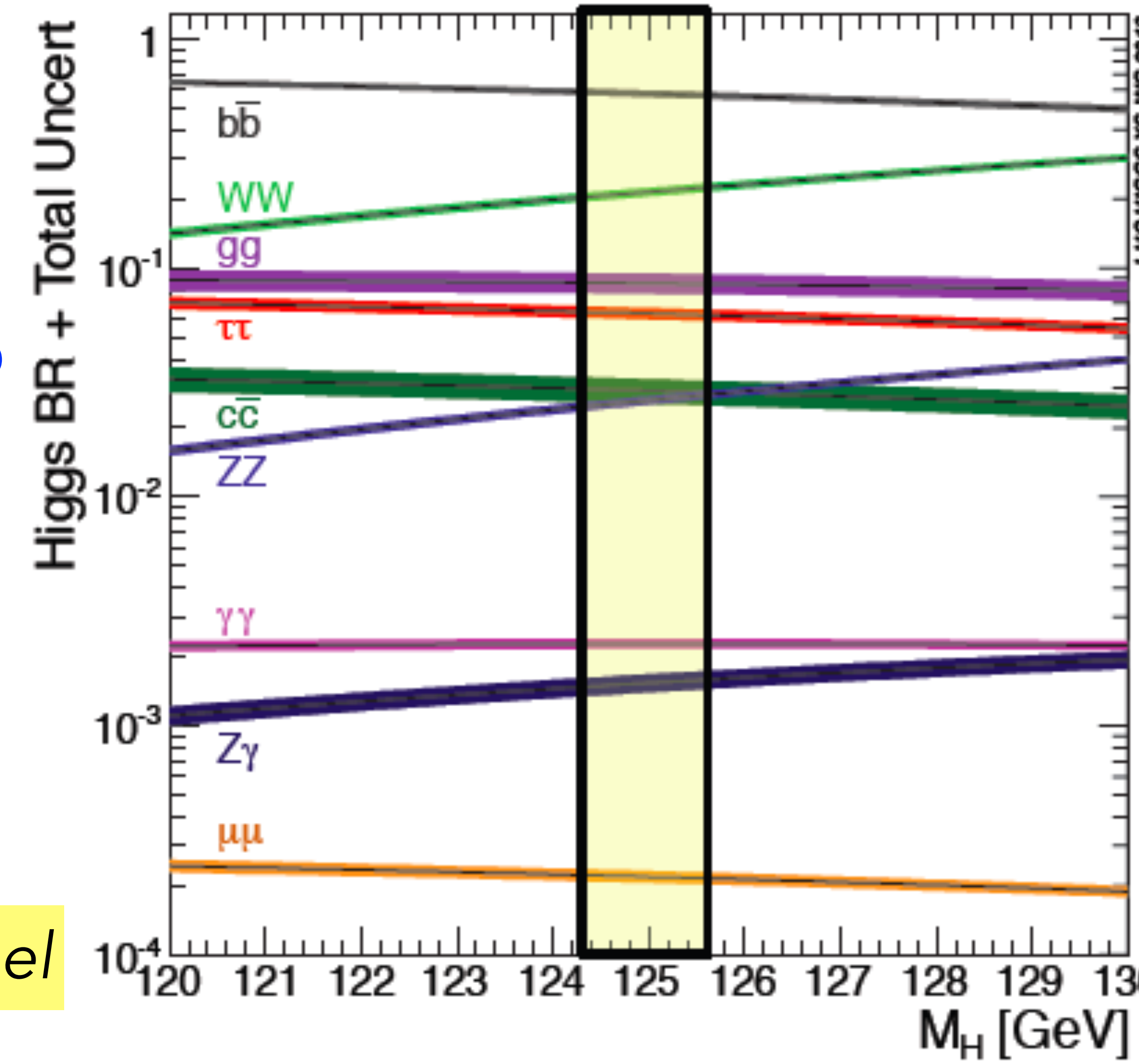
- WW channel (22%)
  but $10^{-2}$ in 2lep2neutrino

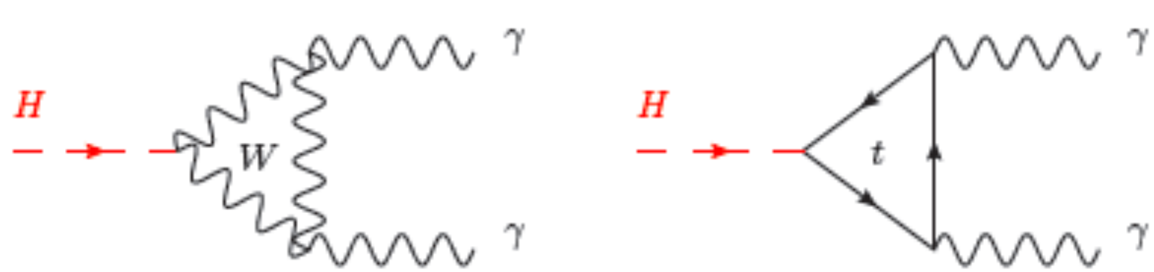
- ττ channel (6.3%)

- ZZ channel (3%) *Discovery channel*
  but $1.2 \times 10^{-4}$ in 4leptons (e,μ)

- cc channel (3%)
  Extremely difficult

- The γγ channel (0.2%) *Discovery channel*



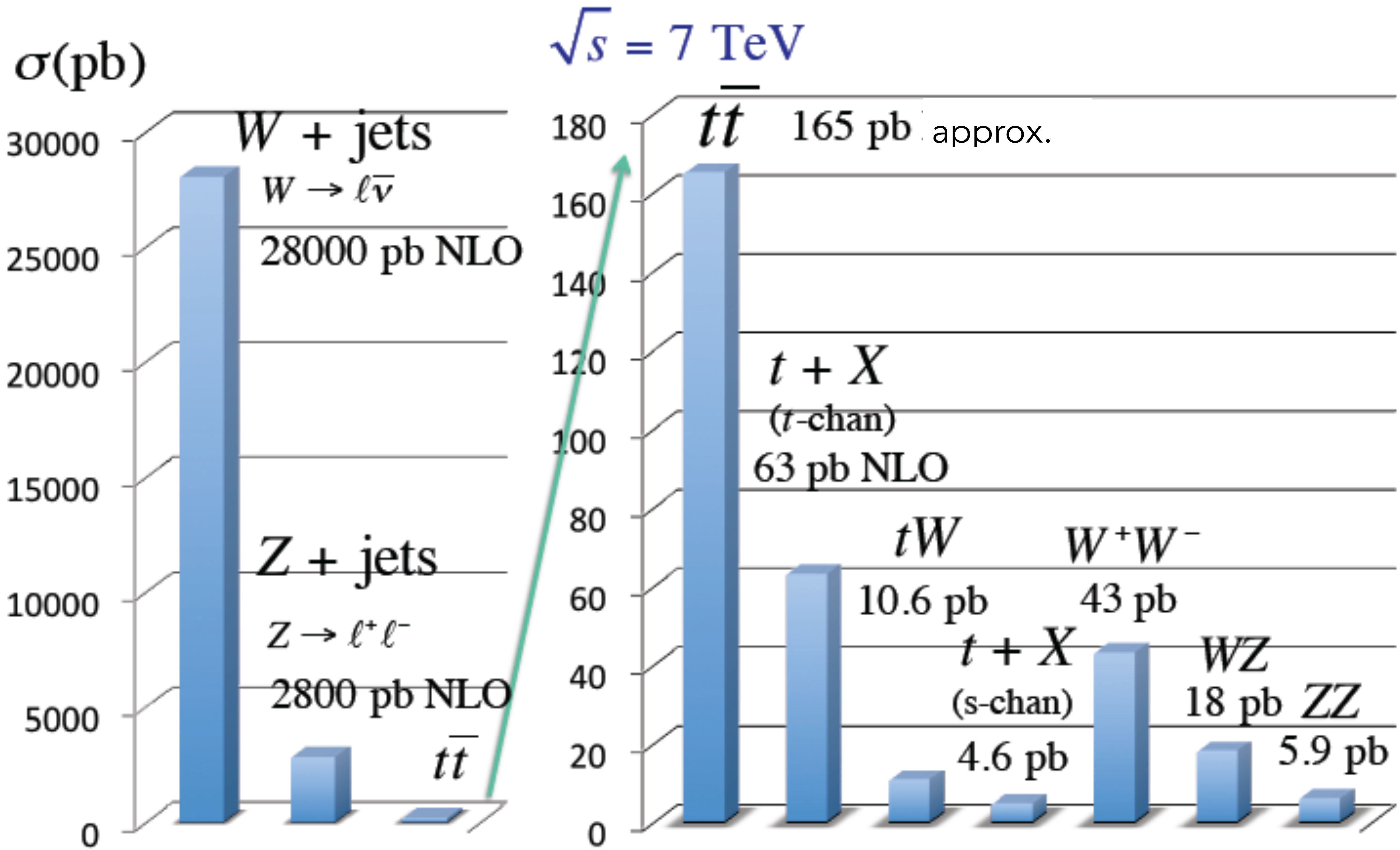




(when assuming no BSM charged in the loop)

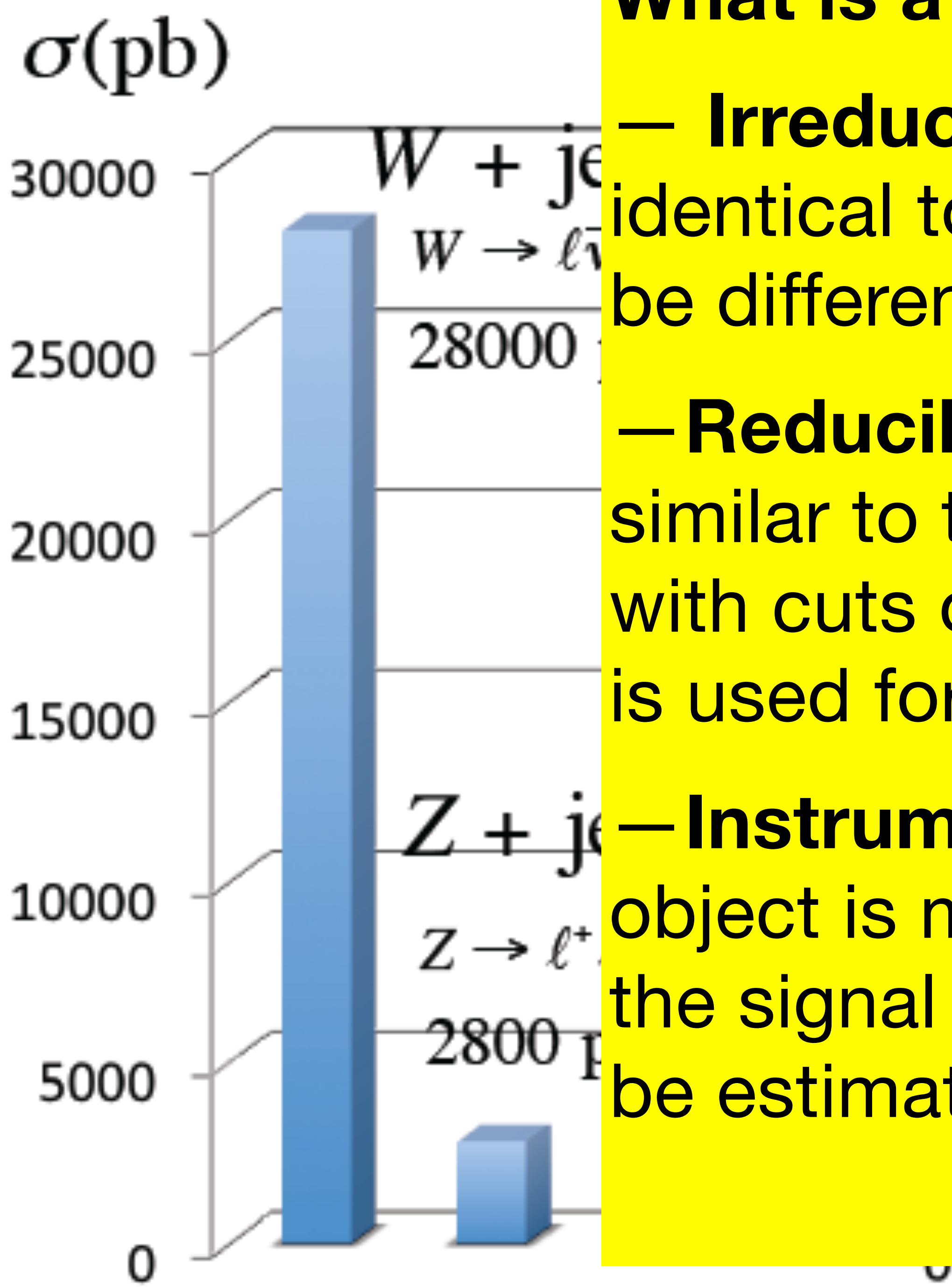
- The Zγ (0.2%)

- The μμ channel (0.02%)

# KEY SM (BACKGROUND) PROCESSES
# AT LHC



# KEY SM (BACKGROUND) PROCESSES
## AT LHC




**What is a background:**

— **Irreducible:** a SM process with a final state identical to the one of our signal. Kinematic could be different. MC is used for estimate.

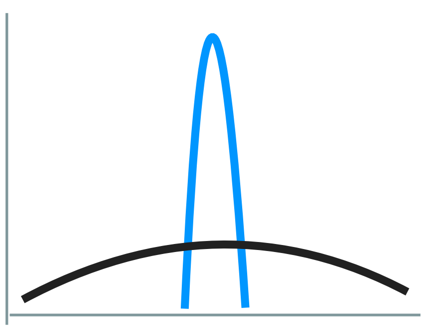
—**Reducible:** a SM process with a final state very similar to the one of our signal. Can be reduced with cuts on the object presence/properties. MC is used for estimate.

—**Instrumental/fakes:** a SM process where one object is misreconstructed to look like one a in the signal final state. These backgrounds need to be estimated from the data themselves.

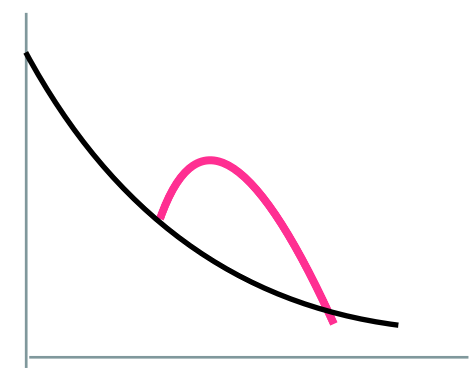
# CHALLENGES OF THE H($b\bar{b}$) MODE AT THE LHC

➤ The final state where the H->bb is the one with the highest probability but it suffers of an overwhelming bkg from QCD ($10^7$ times bigger) if the gluon fusion process is considered.

Comparison with one of the discovery channels



| | $H \to 4\ell$ | $H \to b\bar{b}$ |
|---|---|---|
| Branching | 0,03% | 58% |
| mass | 1% | 10% |
| S/B | 2 | 0.05 |



H($b\bar{b}$) searches need:

- good b-quark identification performance
- best possible resolution on m($b\bar{b}$)
- to exploit all possible information from the event to improve S/B

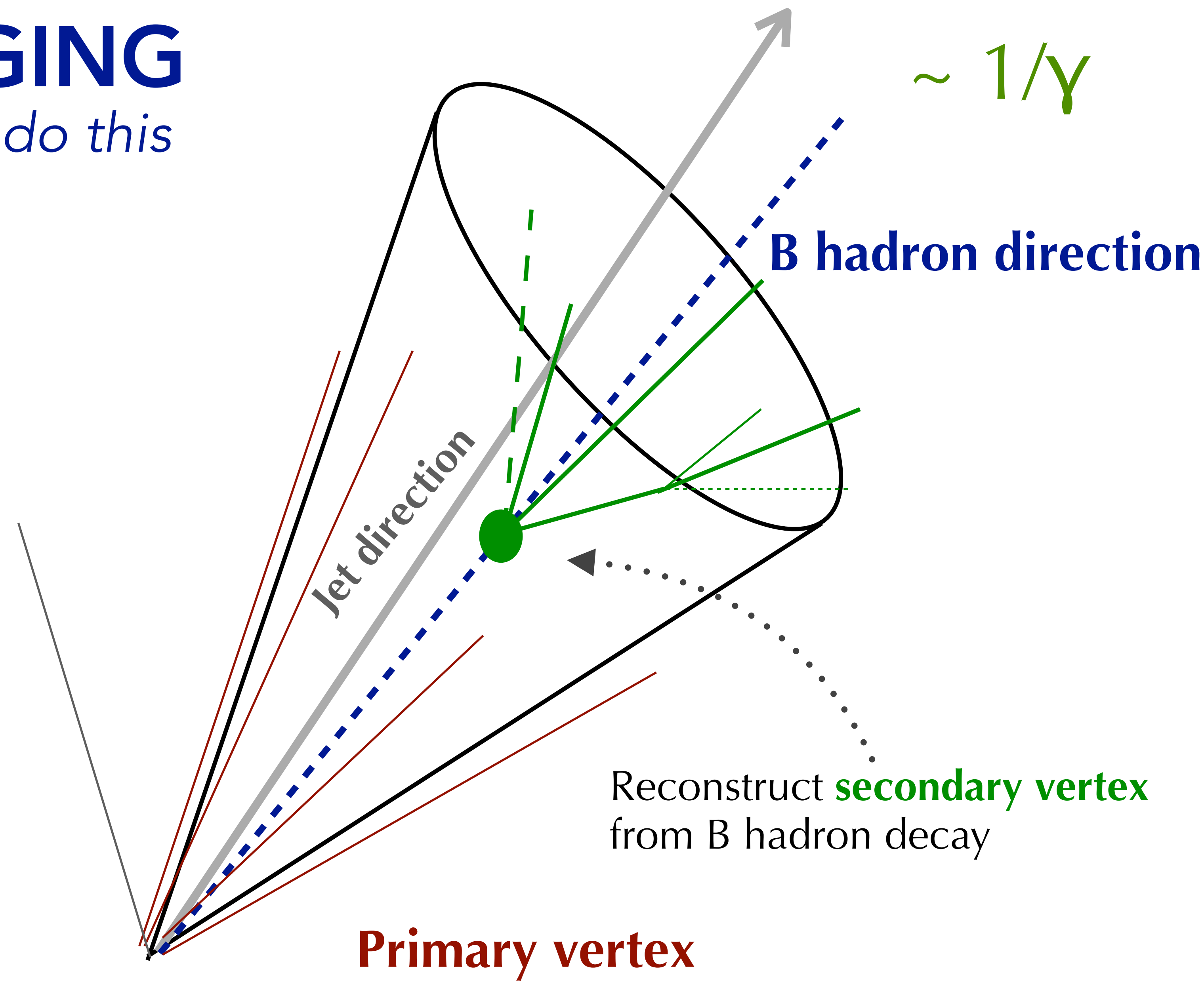
ATLAS-PHYS-PUB-2017-013

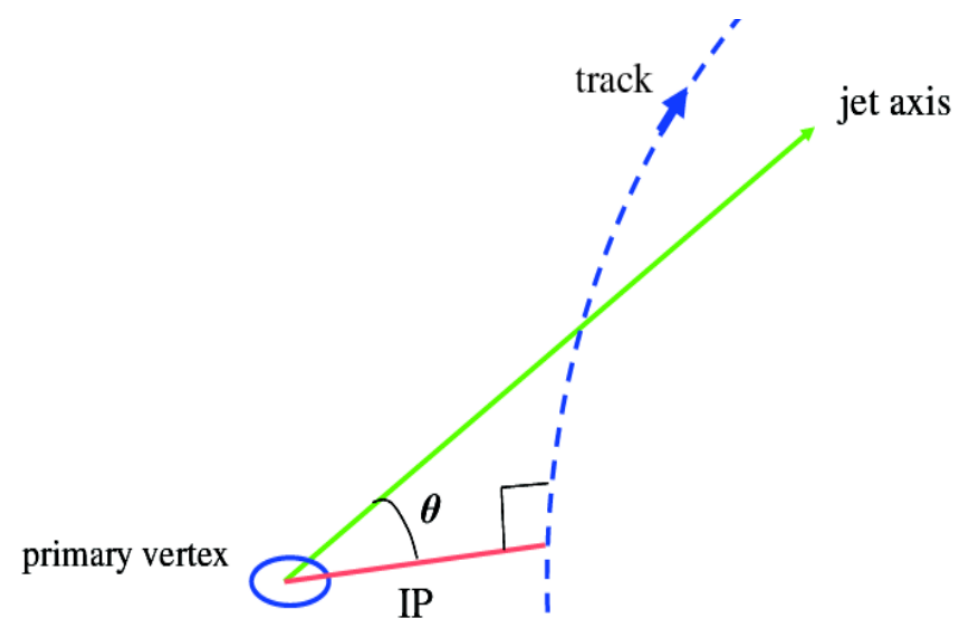
CMS-JINST 13 (2018) P05011

# DIGRESSION: B-TAGGING

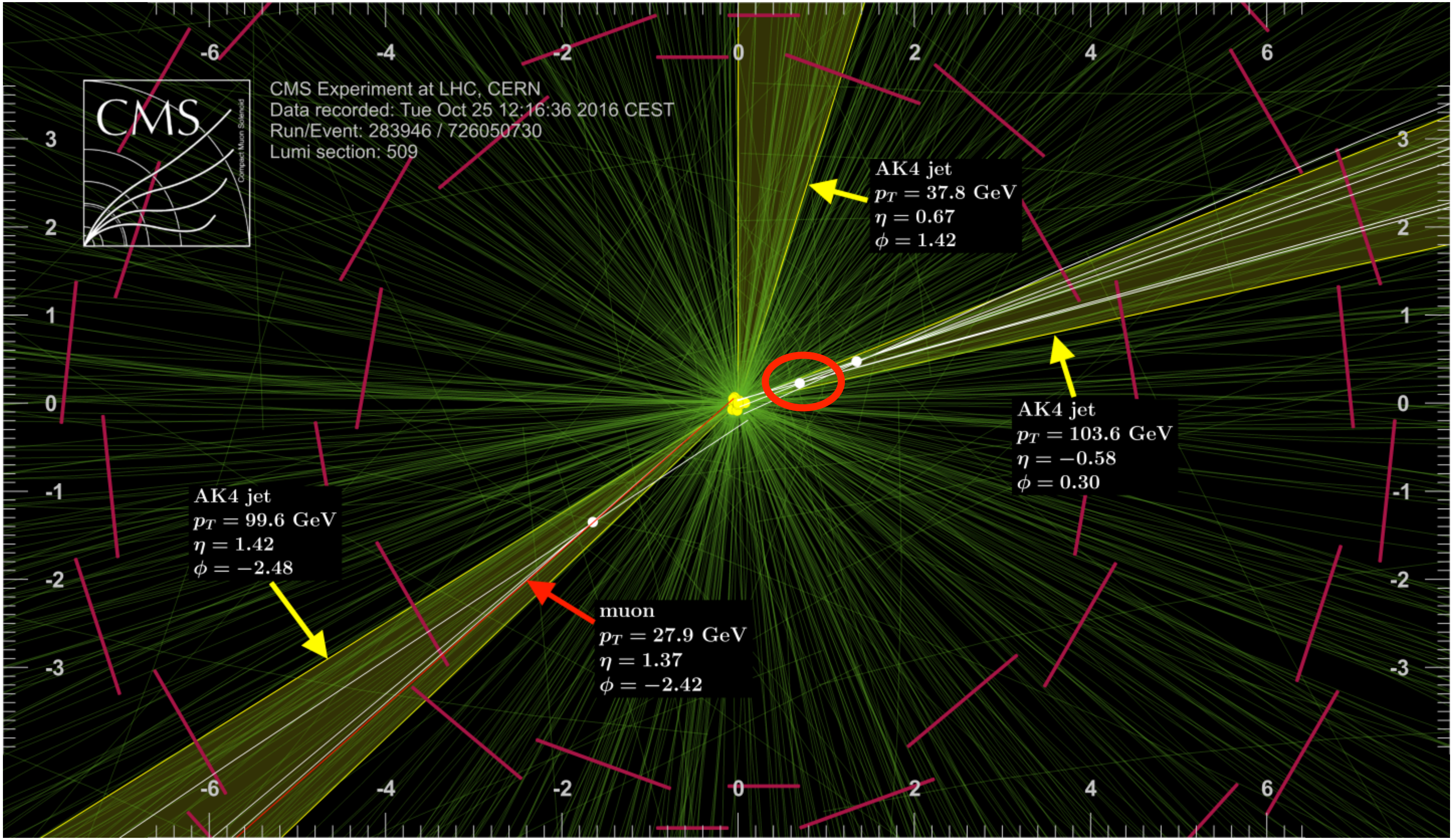
*because it's cool we know how to do this*

- b-quark fragments to B hadrons about 90% of the time. the B-hadron takes about 70% of the quark energy
  - b-hadron decays is typically into ~5 stable charged decay products
- 10% to semileptonic decays.
  - « soft » electrons or muons inside the jet. Specific reconstruction techniques.
- B-quark lifetime ~1.5ps for a cτ~ 0.5mm.
  - Observed distance of the decay vertex is boosted so that for a 50GeV b-hadron dist = βγcτ ~5mm

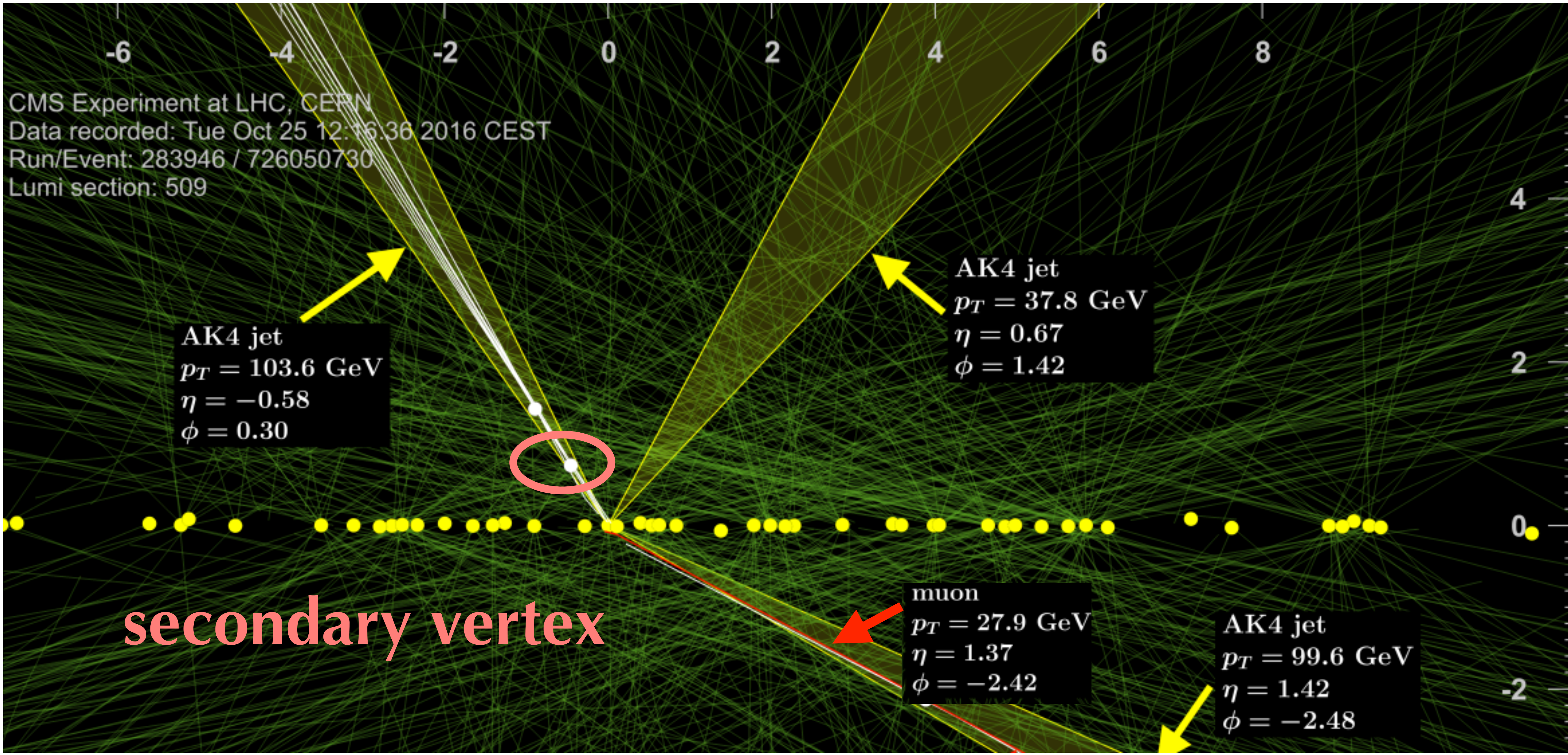


# B TAGGING IN CMS

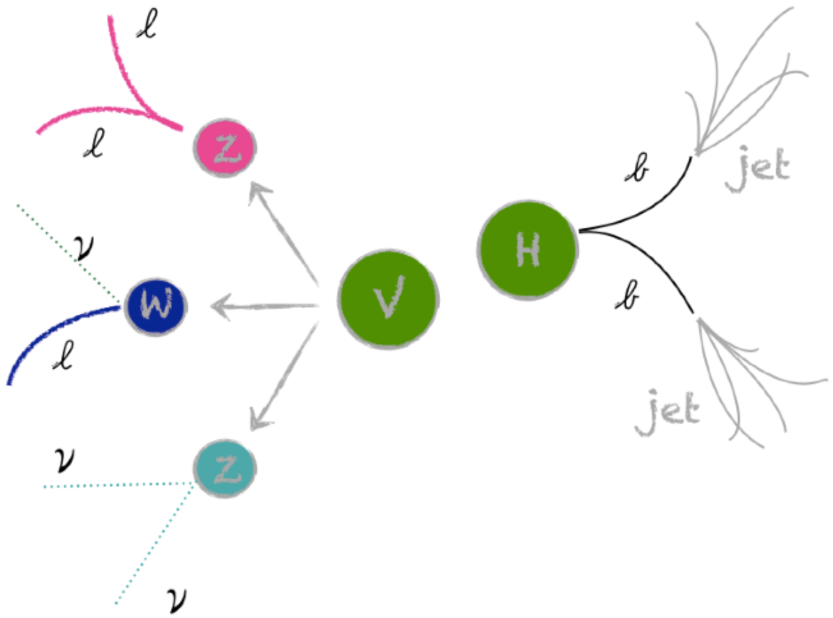


# B TAGGING IN CMS



# ASSOCIATED PRODUCTION VH($B\bar{B}$) TOPOLOGY

## signal

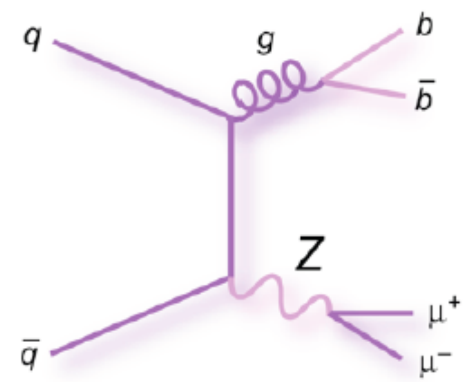


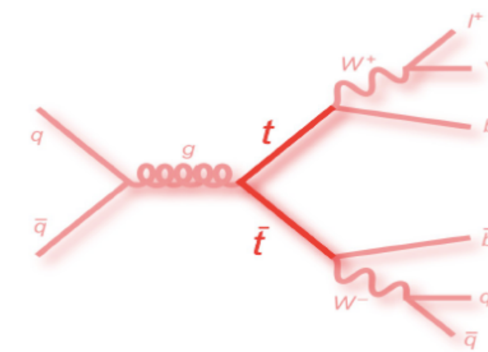

**0-lepton (MET)**

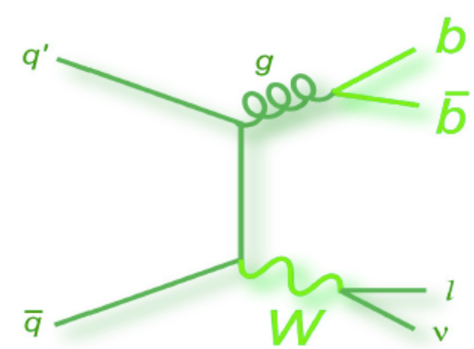
**1-lepton [e,μ]**

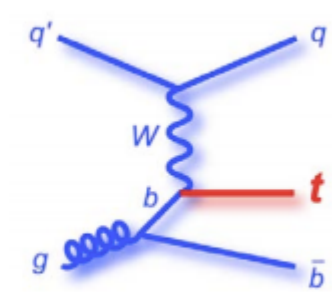
**2-OSSF leptons [ee,μμ]**

## irreducible backgrounds



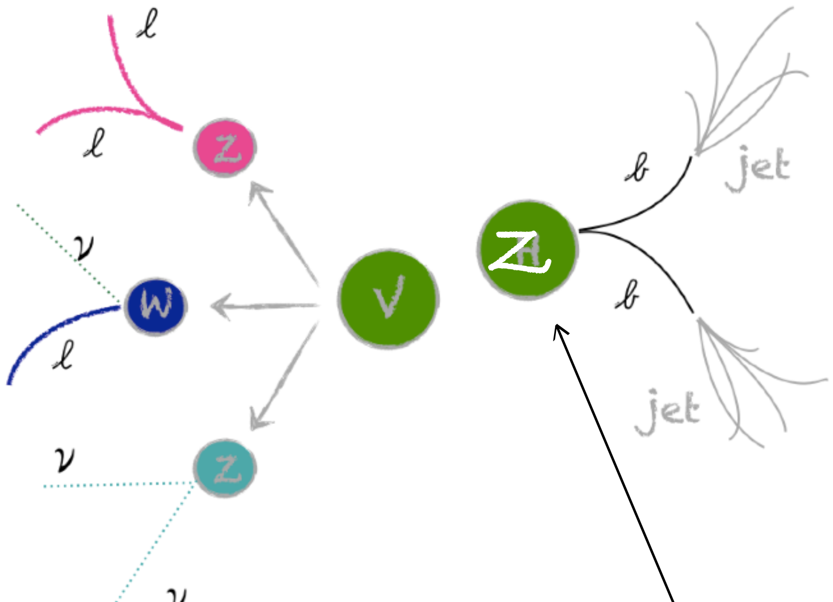




normalization from data, shapes from MC

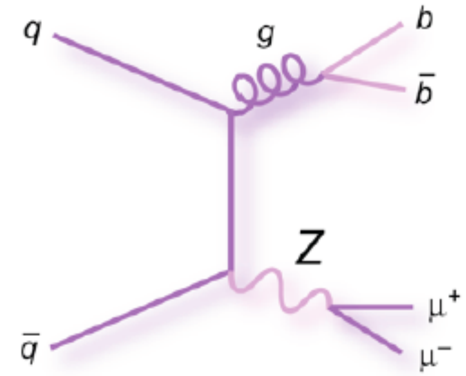
# ASSOCIATED PRODUCTION VH(BB̄) TOPOLOGY

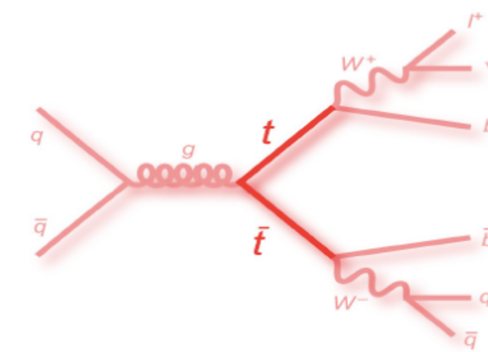
**signal**

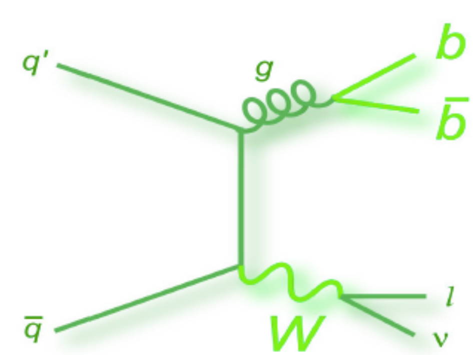


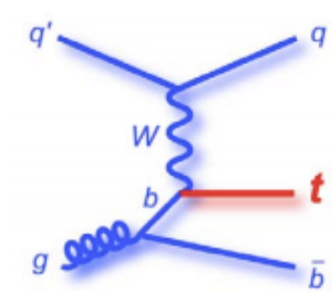

0-lepton (MET)
1-lepton [e,μ]
2-OSSF leptons [ee,μμ]

**irreducible backgrounds**



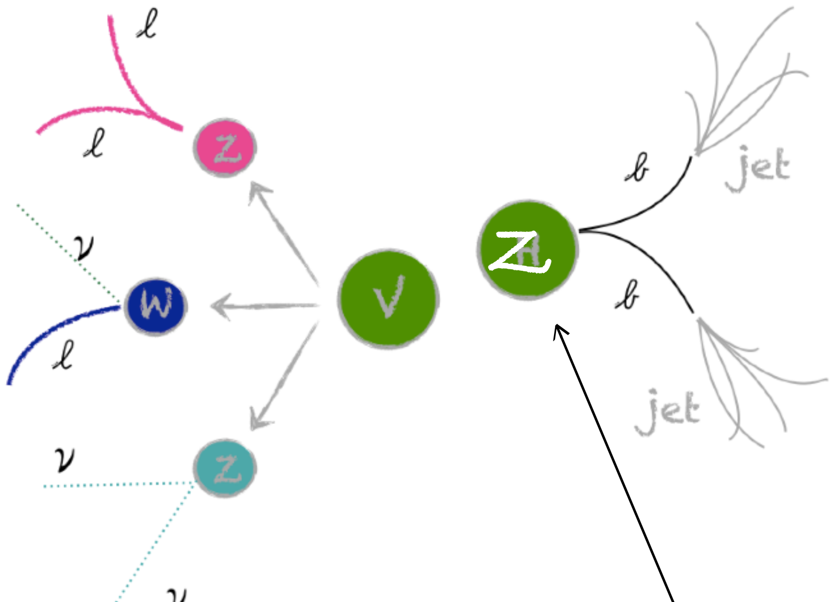




normalization from data, shapes from MC

**and diboson, of course**

# ASSOCIATED PRODUCTION VH(BB̄) TOPOLOGY

**signal**

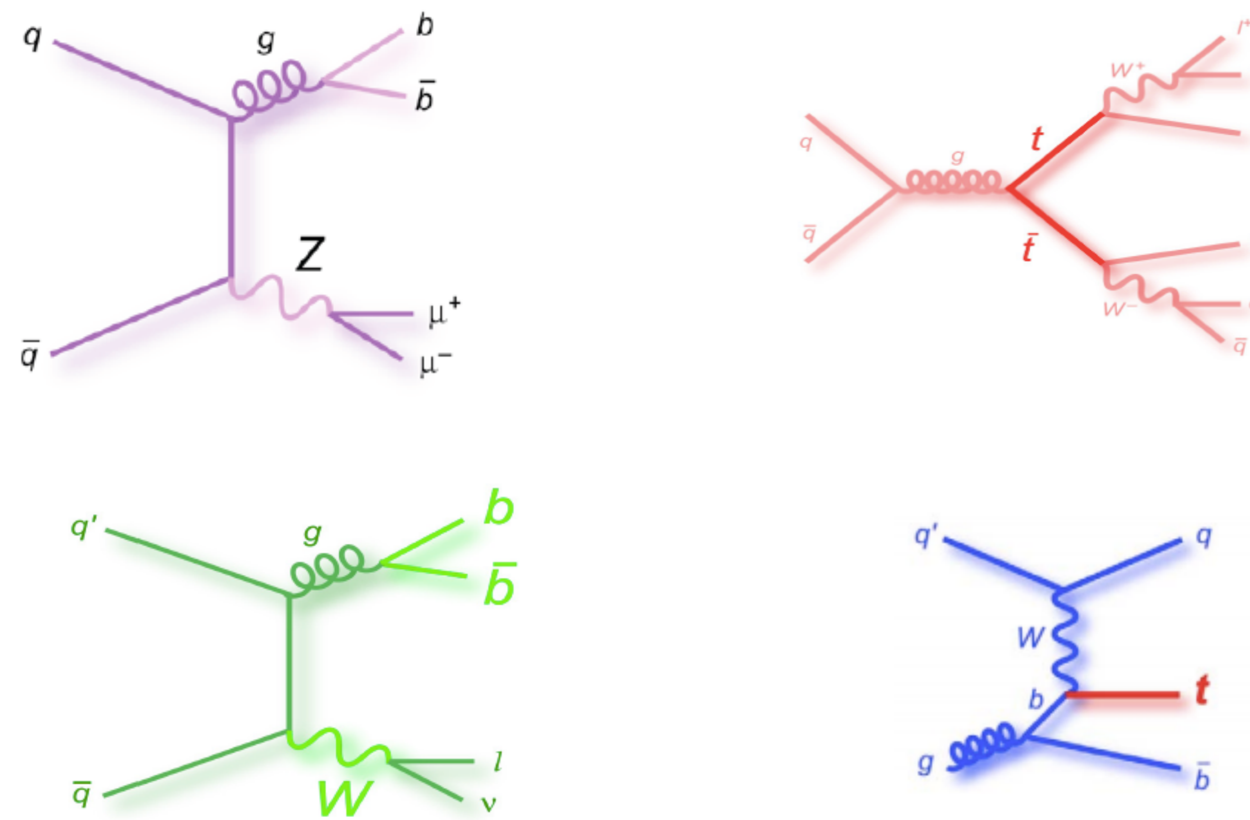



0-lepton (MET)

1-lepton [e,μ]

2-OSSF leptons [ee,μμ]

**irreducible backgrounds**




normalization from data, shapes from MC

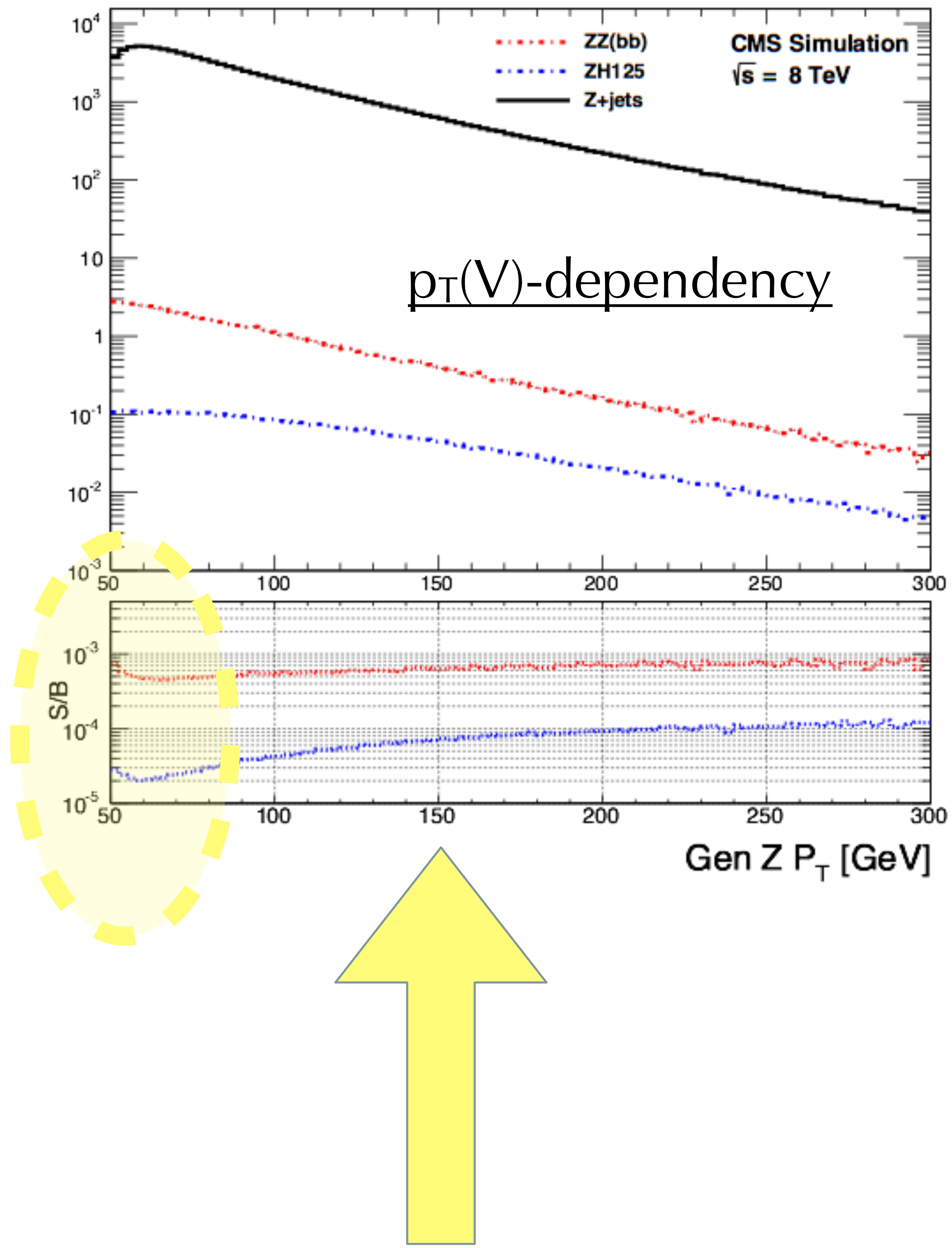
**and diboson, of course**

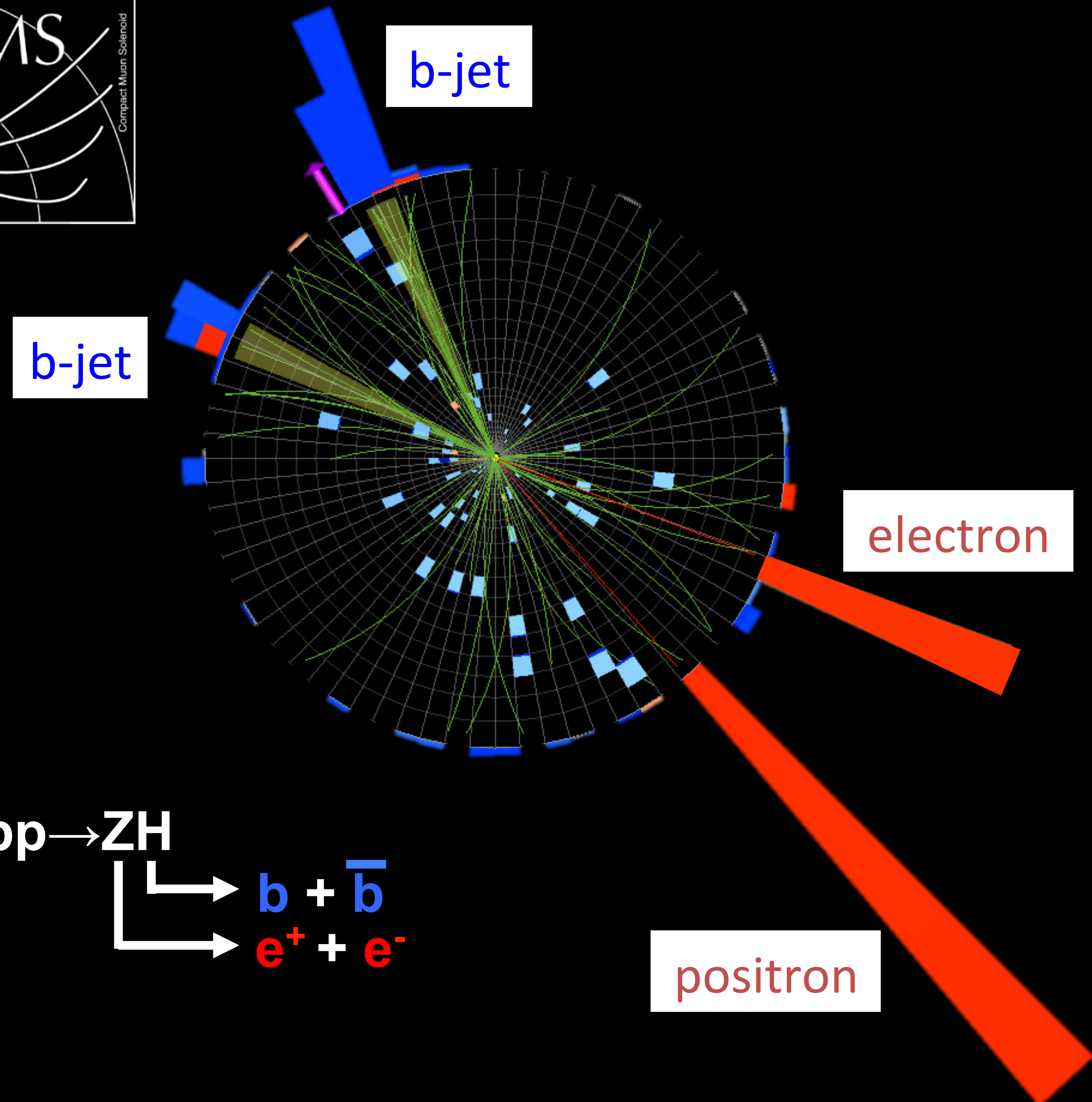
Used to validate the analysis strategy

# ANALYSIS STRATEGY

- Require W/Z to have large boost (~150 GeV)
  - multi-jet QCD background is highly suppressed
- Extract normalization for the dominant backgrounds from the data

  **V+0b/1b/2b and top pair production**
- b-jet energy specific corrections
- Multivariate analysis (DNN) to separate signal and background(s)






$pp \to ZH$

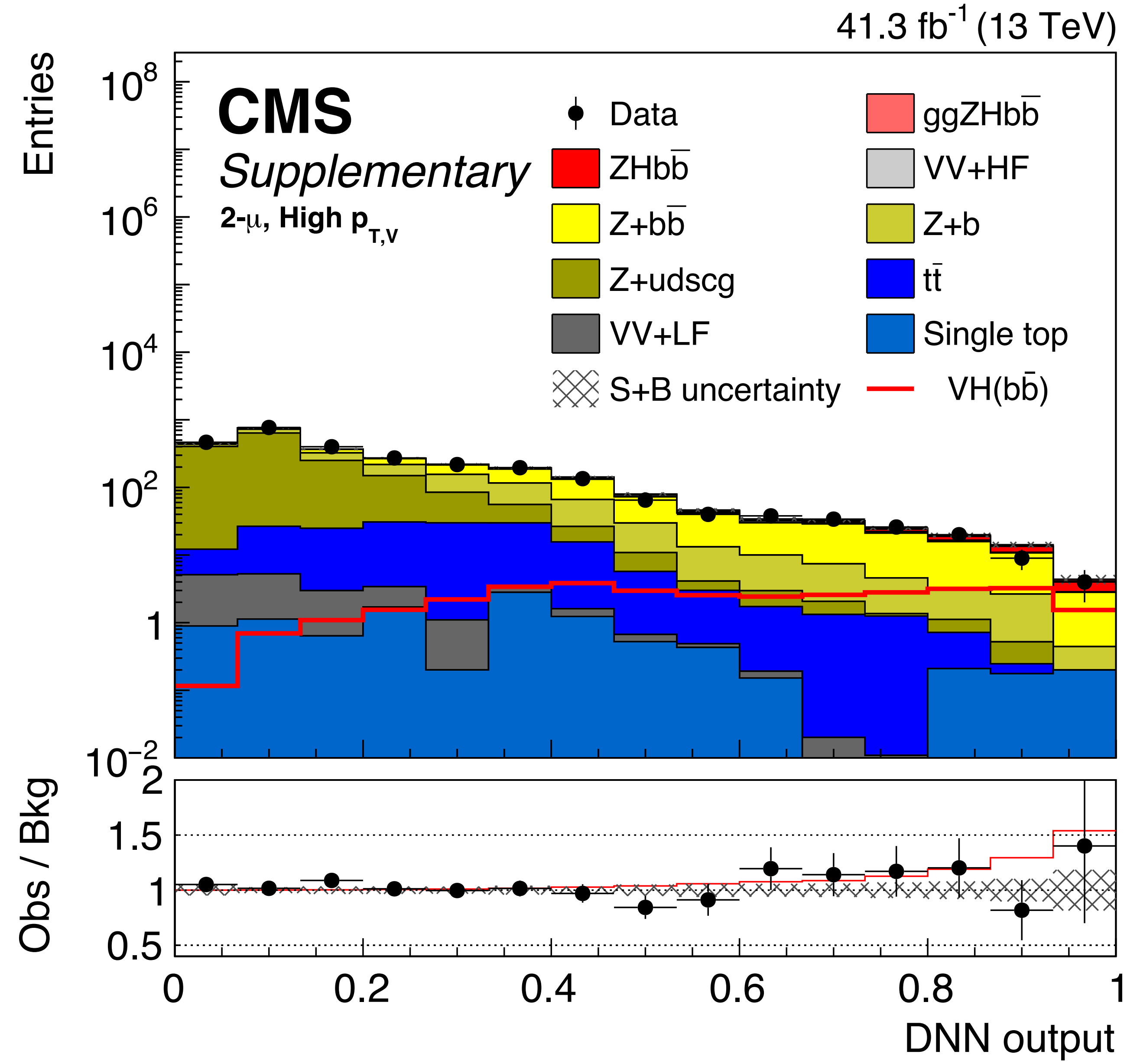
$\quad \to b + \bar{b}$

$\quad \to e^{+} + e^{-}$

# MACHINE LEARNING FOR SIGNAL VS. BACKGROUND

- In the past we used to compare different variables of the event to discriminate the signal from the background.
- Now we have Machine Learning tools that take all the variables and compared them in a more global way, taking into account correlations, and generally having a better discrimination power.
- About 15 input variables describing the kinematics of the events are used depending on the regions
    - Combined into a DNN
- The disadvantage is that for the final results we look at the output « discriminant » variable.

***$m(b\bar{b})$, $\Delta\eta(b\bar{b})$ and b-tagging are the most discriminant***




(this is only one signal region)

# VH BDT/DNN INPUTS

| Variable | Description | Channels |
|---|---|---|
| $M(\mathrm{jj})$ | dijet invariant mass | All |
| $p_{\mathrm{T}}(\mathrm{jj})$ | dijet transverse momentum | All |
| $p_{\mathrm{T}}(\mathrm{j}_1)$, $p_{\mathrm{T}}(\mathrm{j}_2)$ | transverse momentum of each jet | 0- and 2-lepton |
| $\Delta R(\mathrm{jj})$ | distance in $\eta$–$\phi$ between jets | 2-lepton |
| $\Delta\eta(\mathrm{jj})$ | difference in $\eta$ between jets | 0- and 2-lepton |
| $\Delta\phi(\mathrm{jj})$ | azimuthal angle between jets | 0-lepton |
| $p_{\mathrm{T}}(\mathrm{V})$ | vector boson transverse momentum | All |
| $\Delta\phi(\mathrm{V},\mathrm{jj})$ | azimuthal angle between vector boson and dijet directions | All |
| $p_{\mathrm{T}}(\mathrm{jj})/p_{\mathrm{T}}(\mathrm{V})$ | $p_{\mathrm{T}}$ ratio between dijet and vector boson | 2-lepton |
| $M(\ell\ell)$ | reconstructed Z boson mass | 2-lepton |
| $\mathrm{CMVA}_{\mathrm{max}}$ | value of CMVA discriminant for the jet with highest CMVA value | 0- and 2-lepton |
| $\mathrm{CMVA}_{\mathrm{min}}$ | value of CMVA discriminant for the jet with second highest CMVA value | All |
| $\mathrm{CMVA}_{\mathrm{add}}$ | value of CMVA for the additional jet with highest CMVA value | 0-lepton |
| $p_{\mathrm{T}}^{\mathrm{miss}}$ | missing transverse momentum | 1- and 2-lepton |
| $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\mathrm{j})$ | azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and closest jet ($p_{\mathrm{T}} > 30\,\mathrm{GeV}$) | 0-lepton |
| $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\ell)$ | azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and lepton | 1-lepton |
| $m_{\mathrm{T}}$ | mass of lepton $\vec{p}_{\mathrm{T}} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ | 1-lepton |
| $m_{\mathrm{top}}$ | reconstructed top quark mass | 1-lepton |
| $N_{\mathrm{aj}}$ | number of additional jets | 1- and 2-lepton |
| $p_{\mathrm{T}}(\mathrm{add})$ | transverse momentum of leading additional jet | 0-lepton |
| SA5 | number of soft-track jets with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ | All |

# CROSS CHECKS: $M_{B\bar{B}}$ ANALYSIS

CMS-PRL 120 (2018) 231801

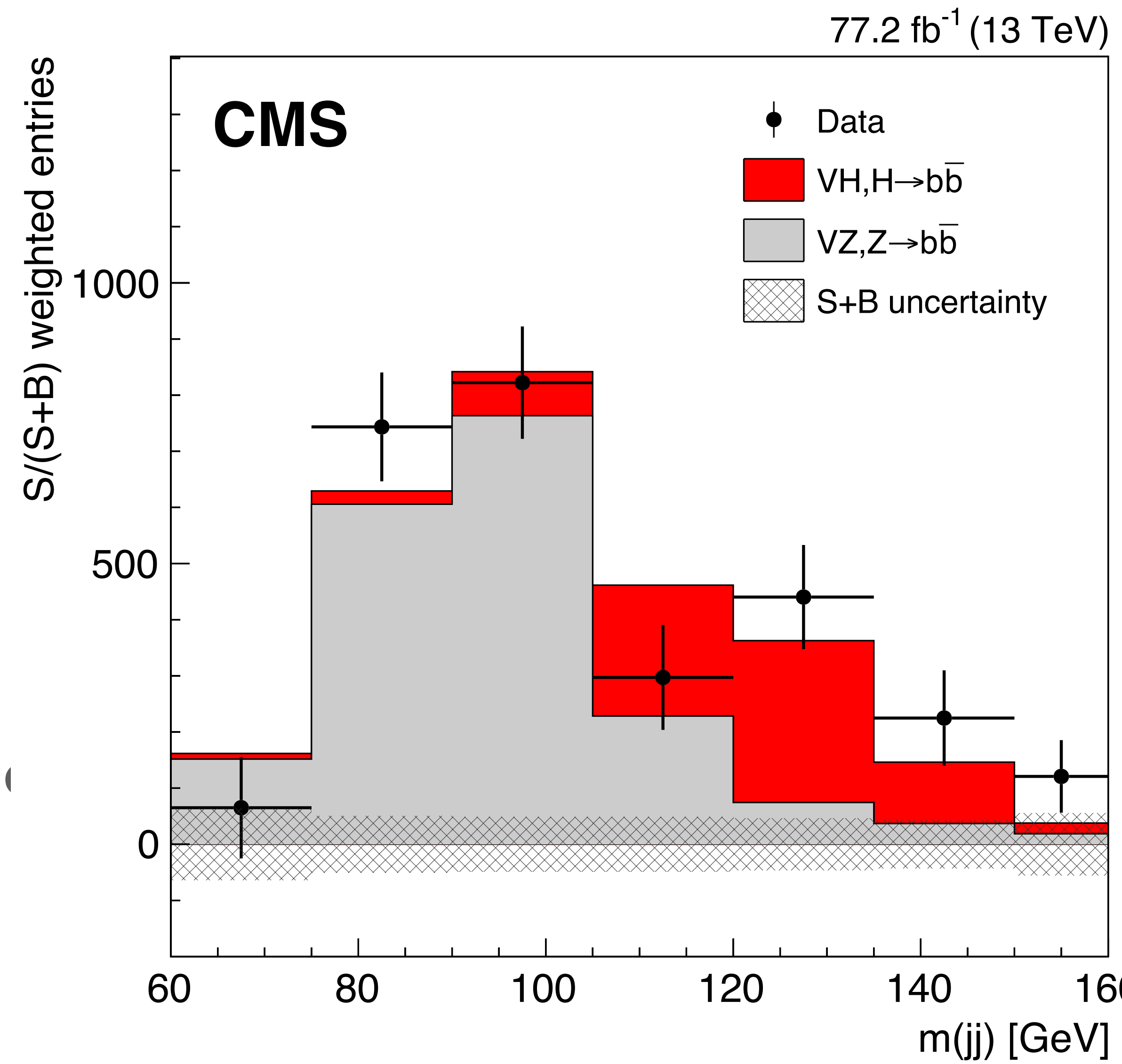
- As a cross check it is good to look at some direct variables instead, for instance the M(bb) mass which should show the enhancement from the Higgs

- Re-derive DNN in signal regions to discriminate **VZ(b$\bar{b}$) signal**

***Consistent with SM expectations***
*Run-2 2017* ***5.2 (5.0) σ***
***$\mu = 1.05^{+0.22}_{-0.21}$***

- Re-derive DNN removing $m_{b\bar{b}}$ dependen
  - Split each channel signal region into four categories based on **massless DNN** score

*Run-2 2016+2017*
***2.7 (3.0) σ   $\mu = 0.91^{+0.35}_{-0.34}$***




M(jj) distribution for events in DNN signal region

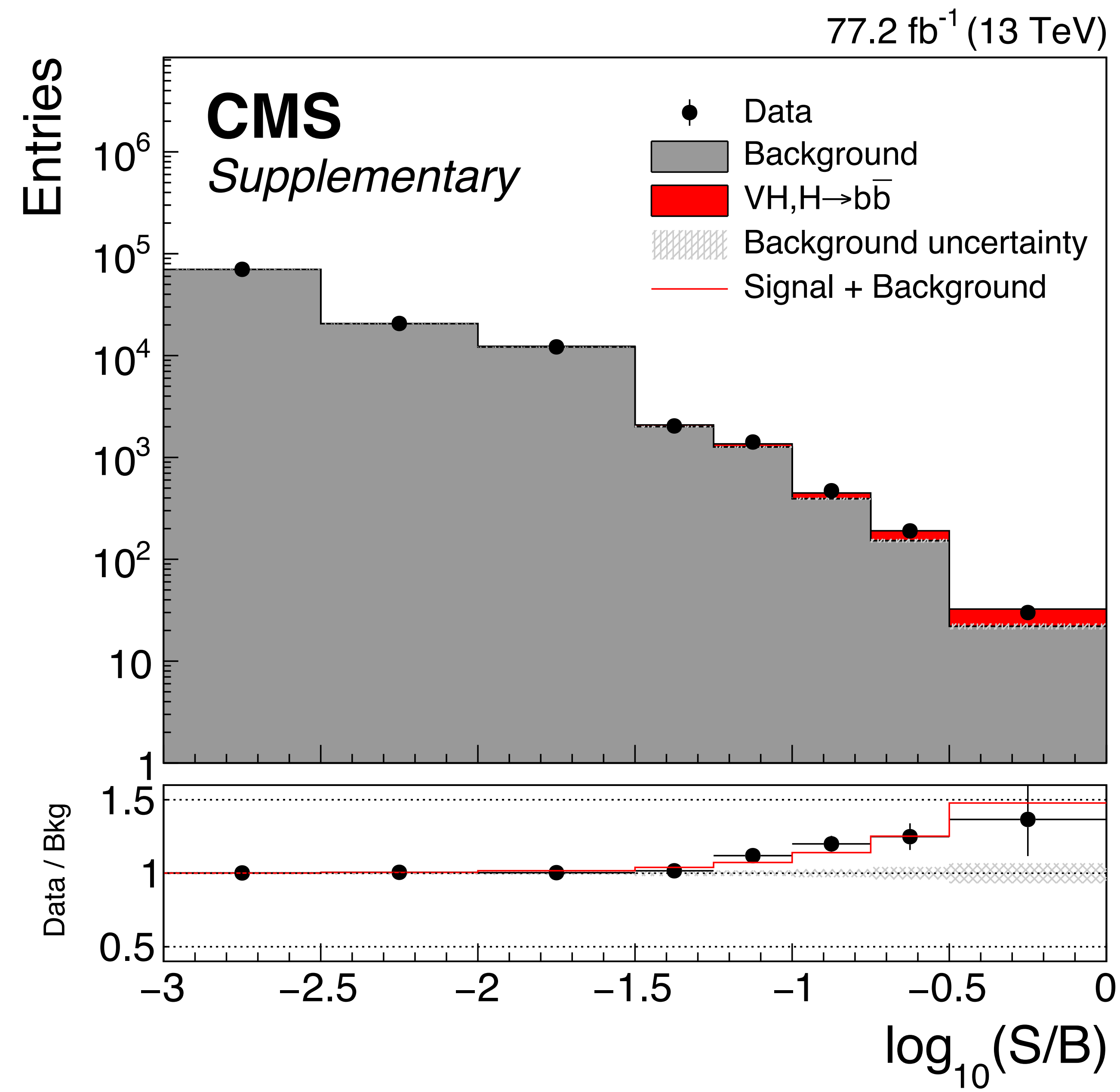
# VH(BB̄) FINAL RESULTS

- The data set has been separated in many regions with different S and B content
- For the final results a simultaneous fit to the DNN output of all regions is performed to extract signal strength

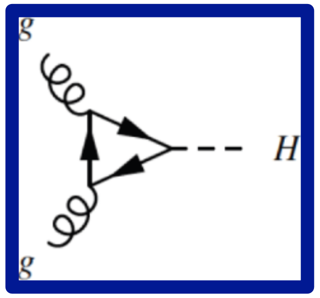
*Run 2 2016+2017*

***4.8 (4.5) σ***

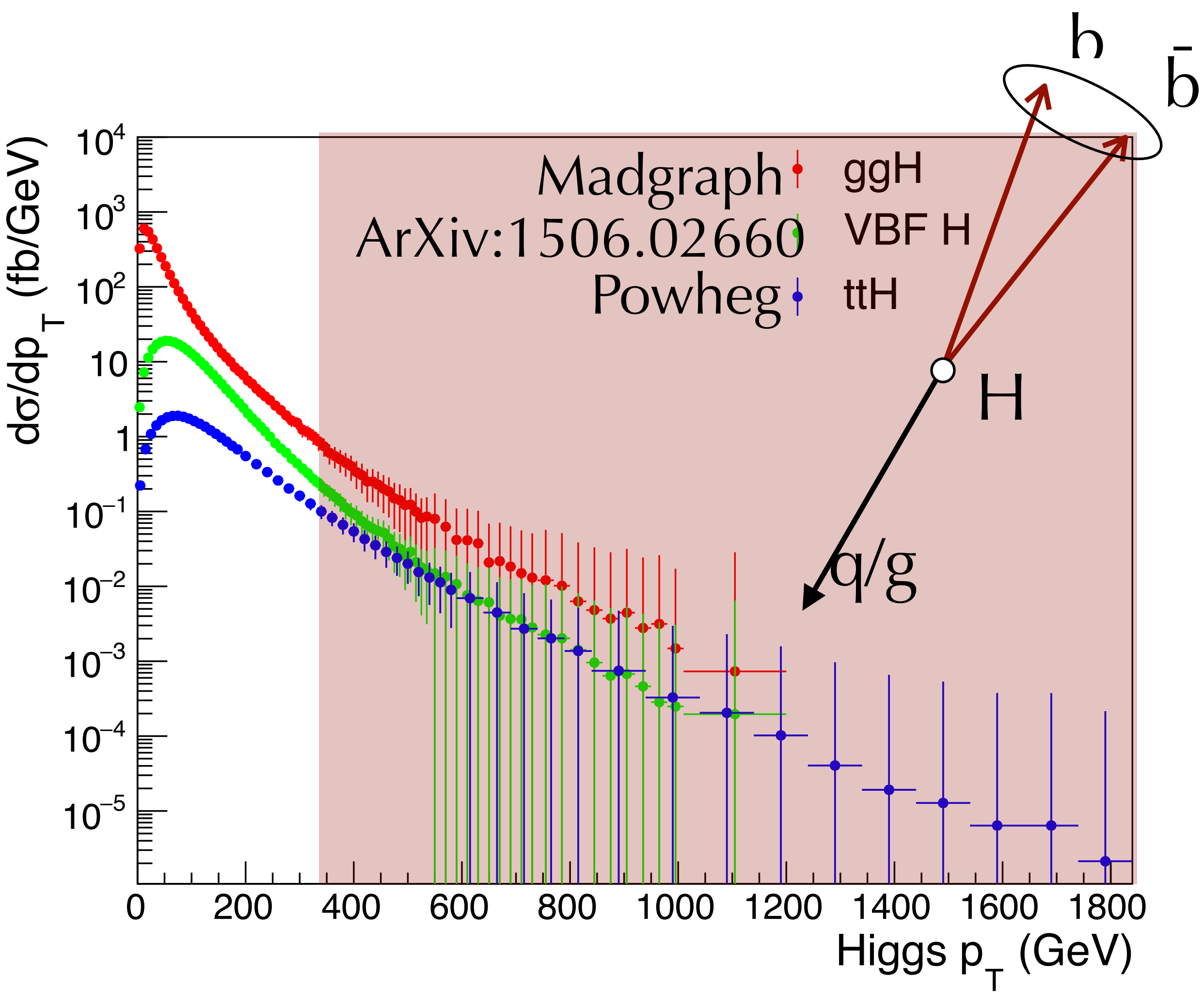
***$\mu = 1.06^{+0.26}_{-0.25}$***



# GLUON FUSION H(B$\bar{B}$)

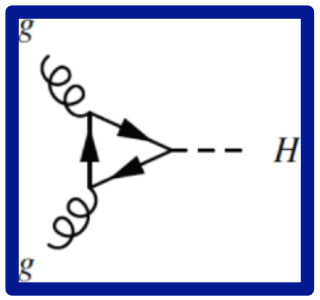



**Gluon Fusion (87%)**

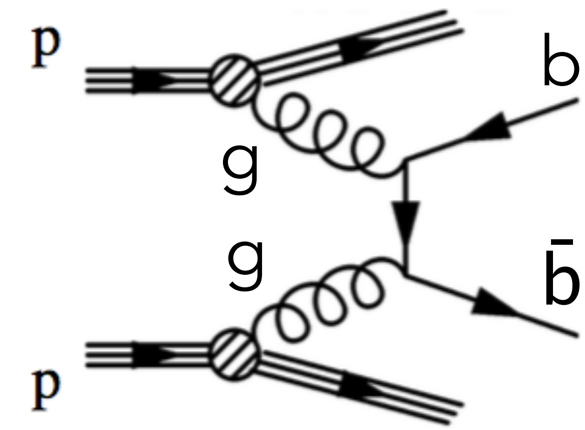


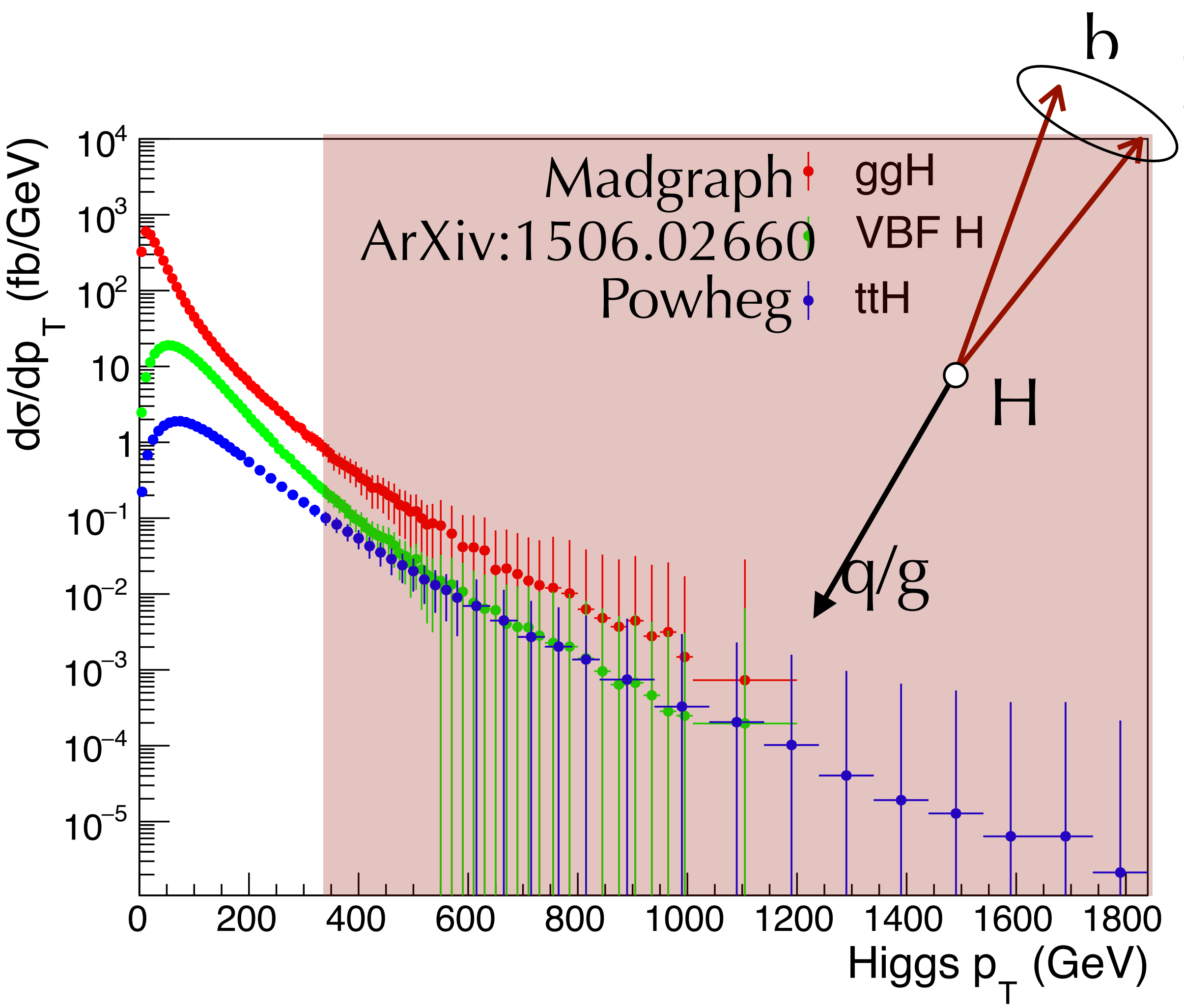

- We can access gluon fusion H(b$\bar{b}$) in the boosted dijet topology
- Look for boosted H boson in a single jet mass distribution
  - Use the Z boson as calibration of the analysis technique
    - b-tagging to resolve W/Z

# GLUON FUSION H(BB̄)



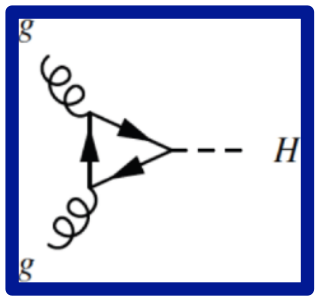

## Gluon Fusion (87%)



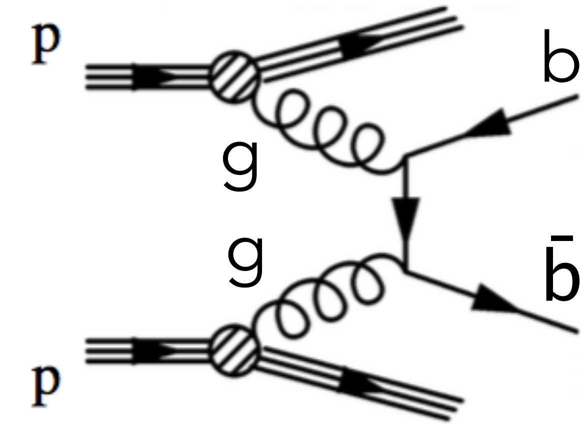

- ➤ We can access gluon fusion H($b\bar{b}$) in the boosted dijet topology
- ➤ Look for boosted H boson in a single jet mass distribution
  - ➤ Use the Z boson as calibration of the analysis technique
    - ➤ b-tagging to resolve W/Z

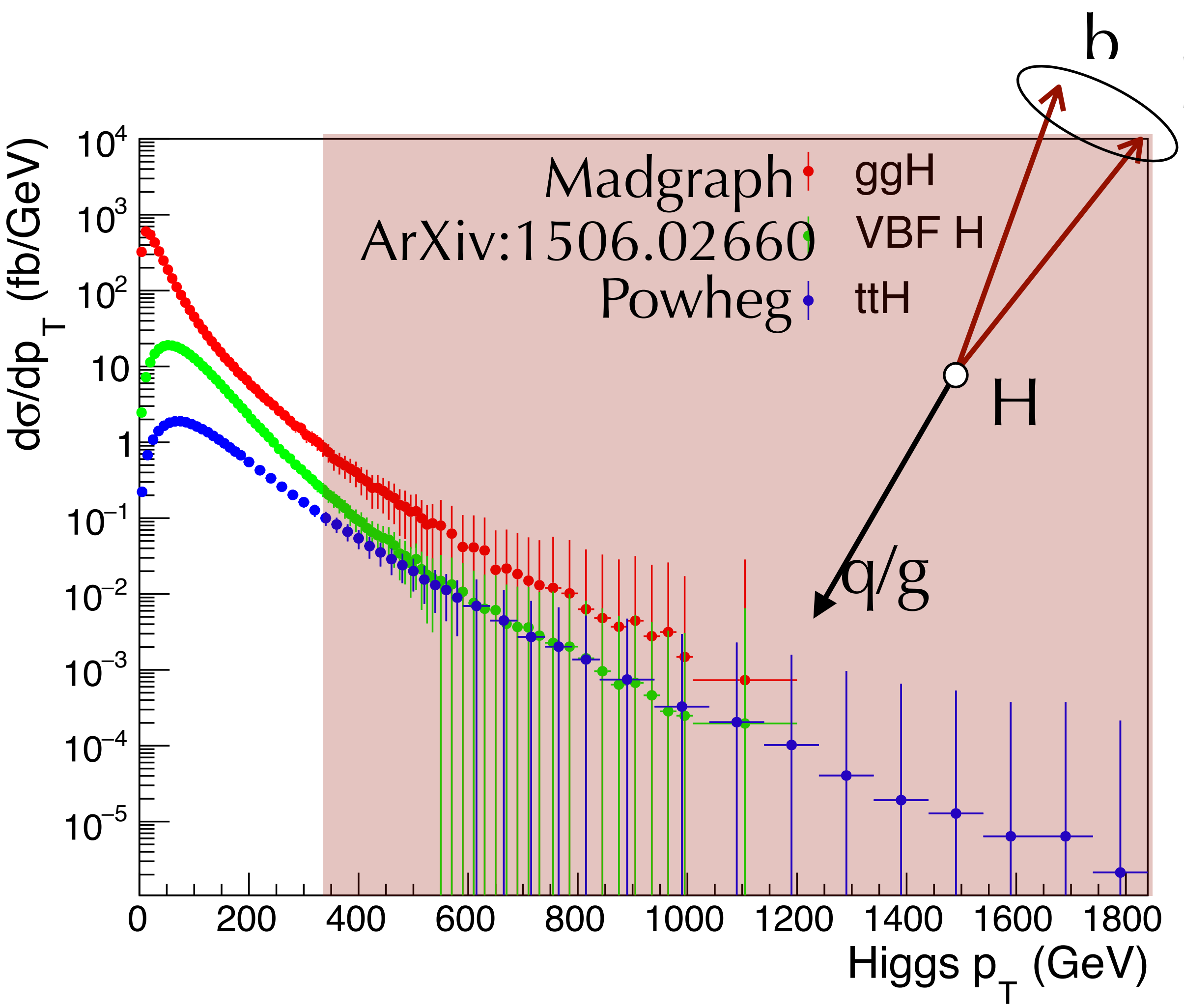
# GLUON FUSION H(BB̄)




**Gluon Fusion (87%)**

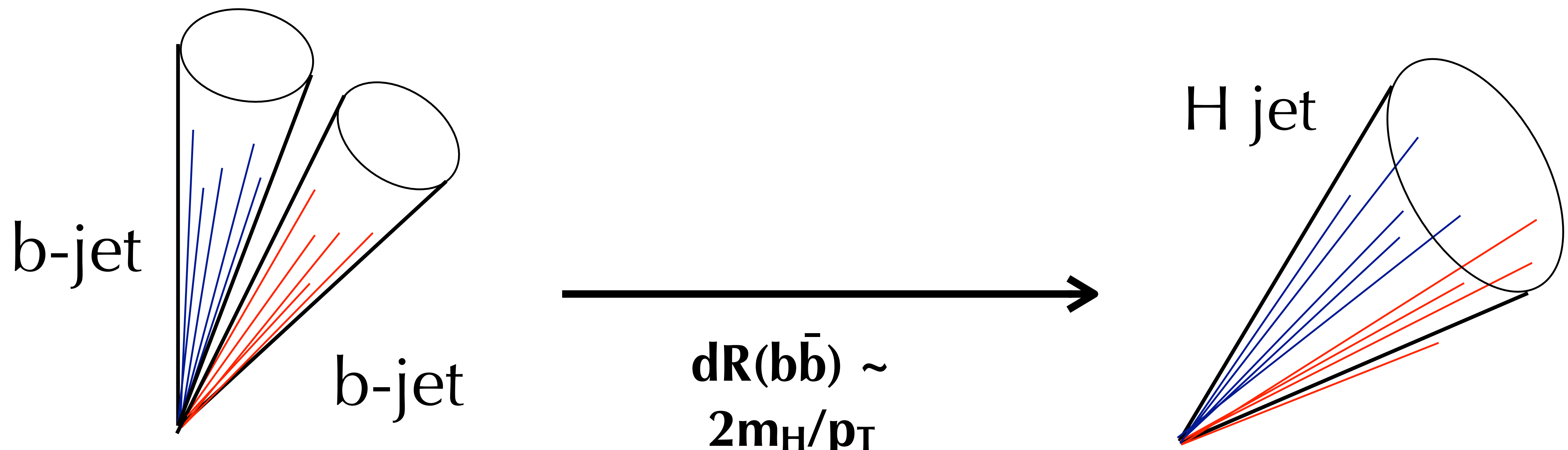
Overwhelming ($10^7$ larger) background of b-quark production due to strong interactions




- We can access gluon fusion H(bb̄) in the boosted dijet topology
- Look for boosted H boson in a single jet mass distribution
  - Use the Z boson as calibration of the analysis technique
    - b-tagging to resolve W/Z

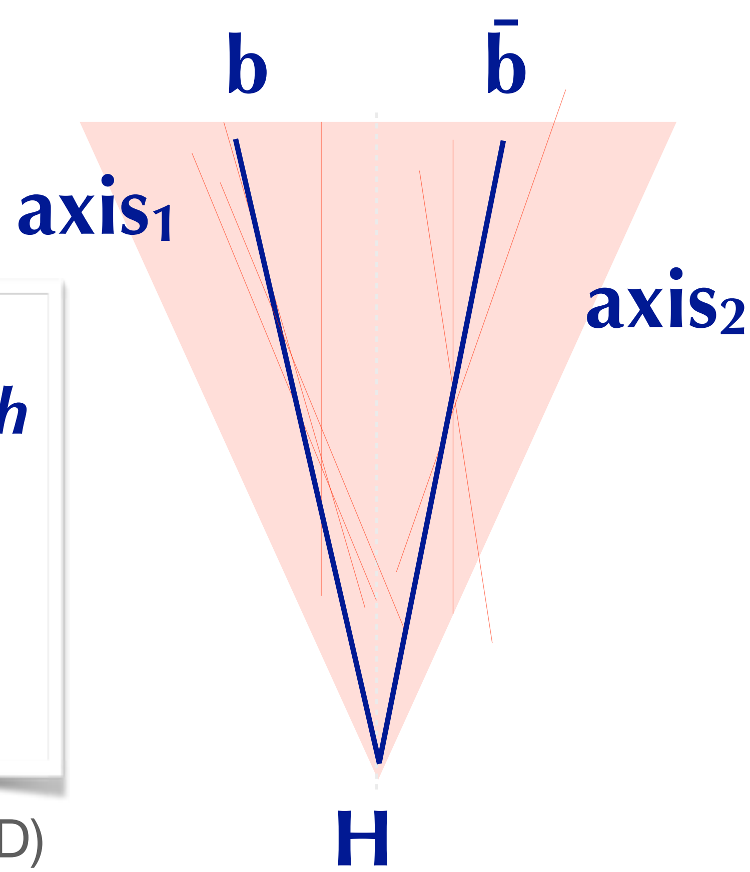
# BOOSTED H(B$\bar{B}$)




**H(b$\bar{b}$)**

two-separate b-jets
(R = 0.4)

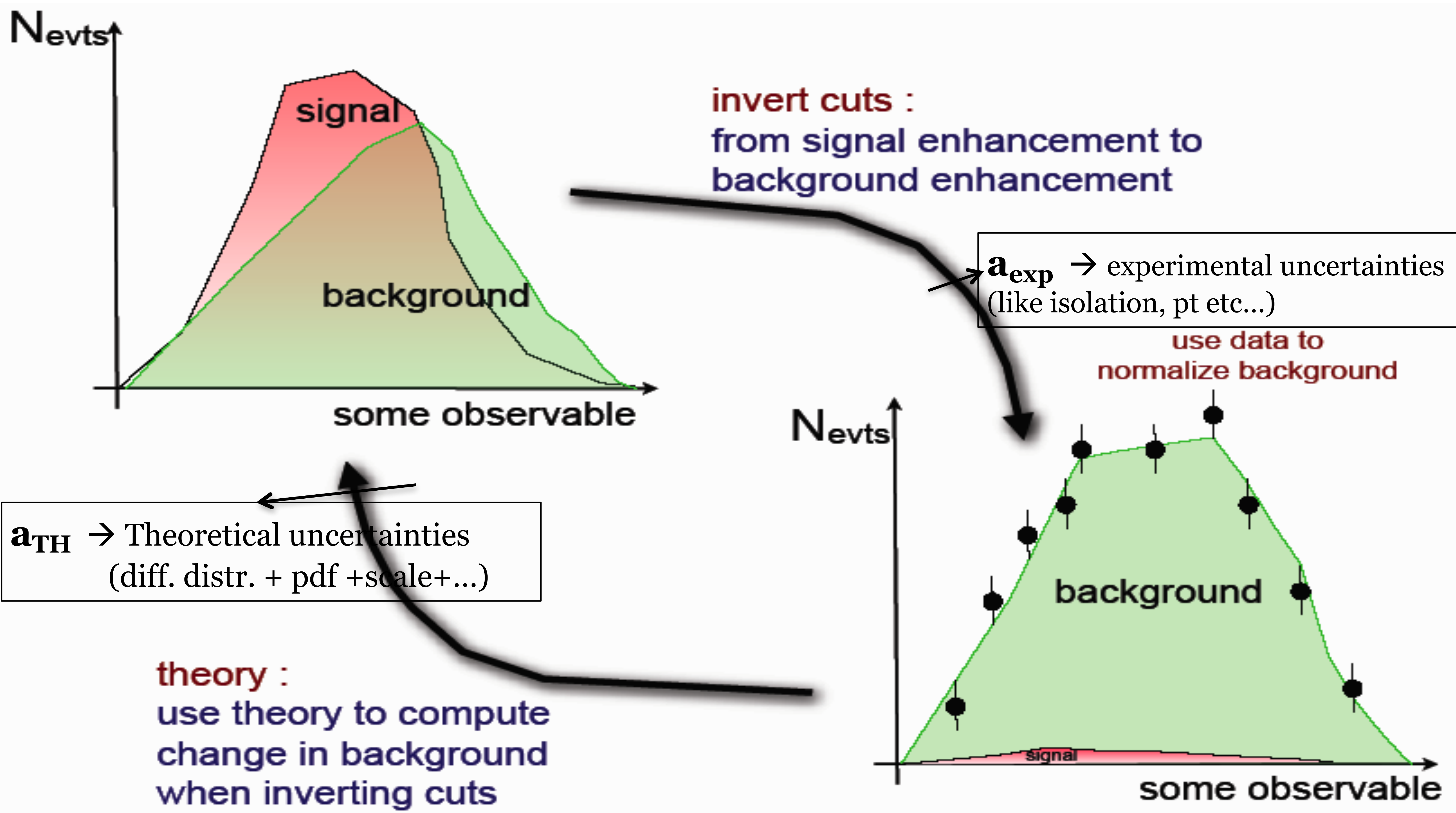
**H(b$\bar{b}$)** one single large-cone (fat) jet (R = 0.8)



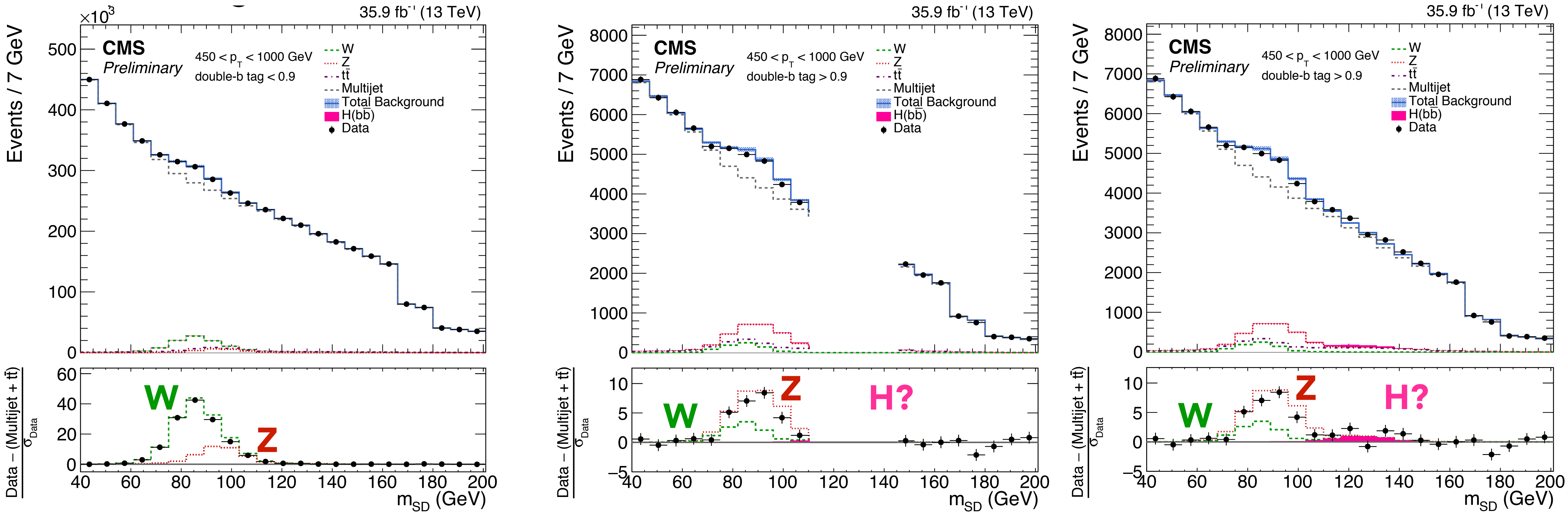

***two axes associated with the two constituents with highest momentum***

- Developed a novel approach, double-b tagger
  - Identifies the two B hadron decay chains from b and $\bar{b}$ within the same fat jet
  - It targets the b$\bar{b}$ signal aiming to be independent of the jet mass or pT

# ONE TECHNIQUE TO CONTROL THE BACKGROUND



# RESULTS H->bb ANALYSIS




SM candles: Z(bb) peak provides in-situ constraint of H(bb) signal systematics

**observed Z(bb) significance: 5.1σ, $\mu_Z = 0.78^{+0.23}_{-0.19}$**

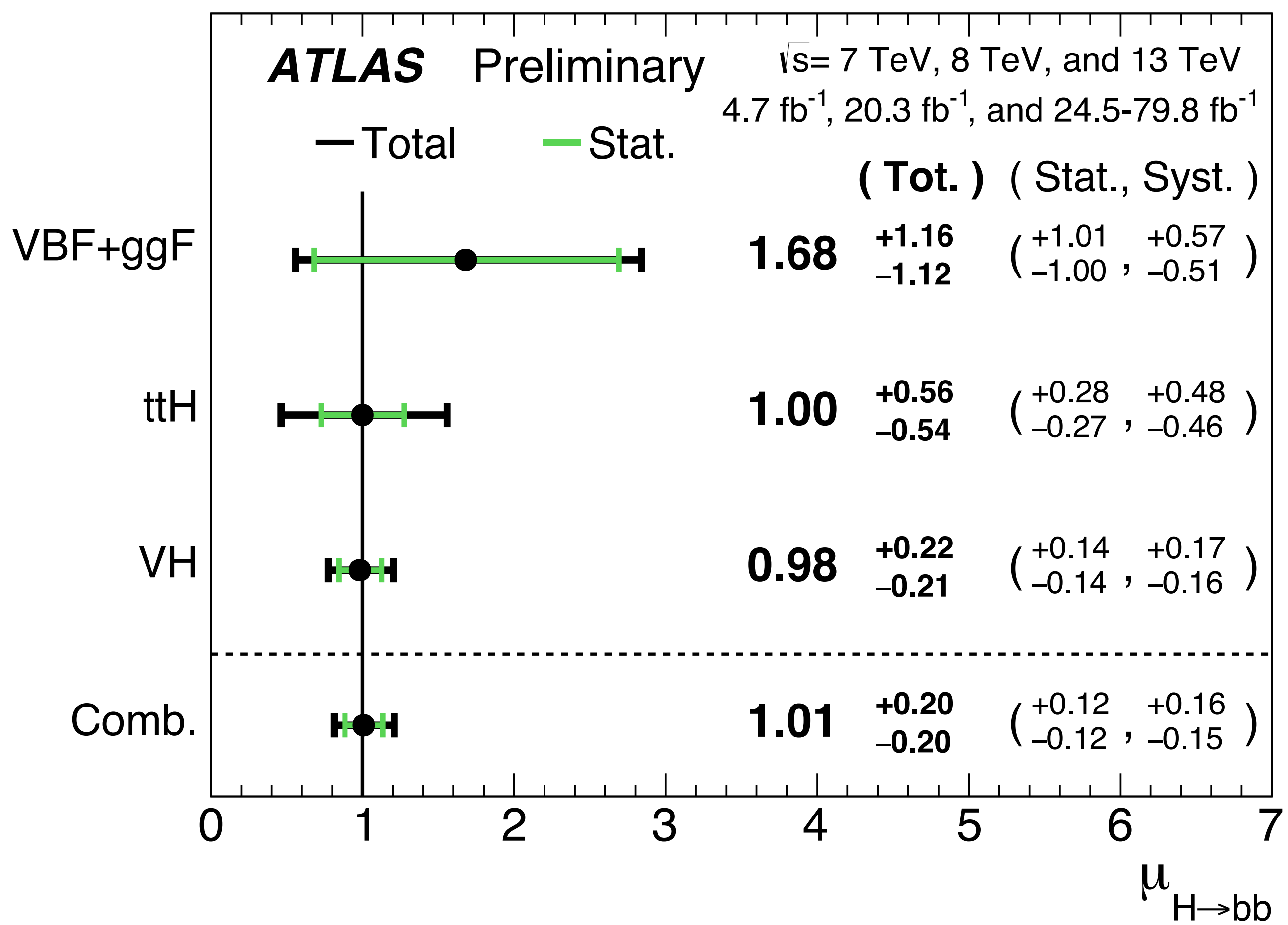
**observed H(bb) significance: 1.5σ, $\mu_H = 2.3^{+1.8}_{-1.6}$**

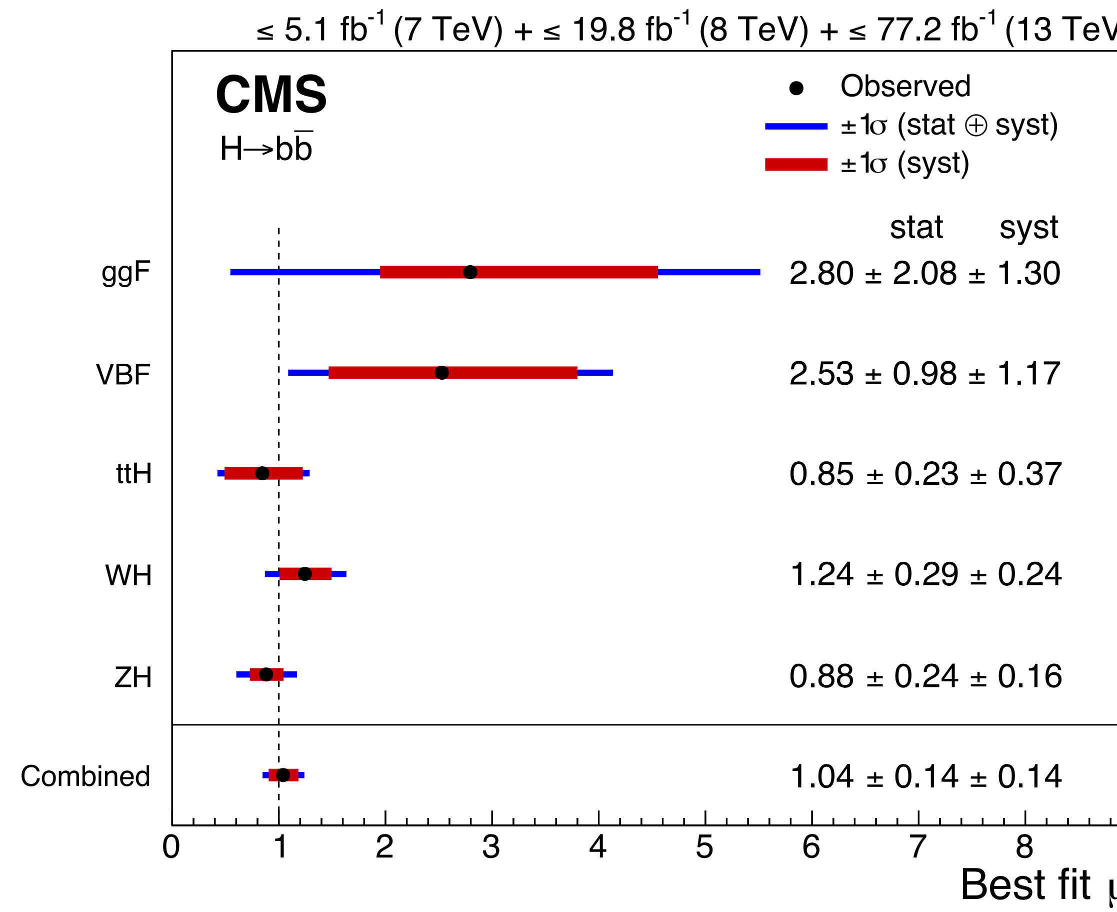
Require that the content of the jets is due to light quarks: enhance the W content

Require that the content of the jets is due to bottom quarks: enhance the Z content

Require that the content of the jets is due to bottom quarks: enhance the Z and the Higgs content

# H(B$\bar{B}$) OBSERVATION



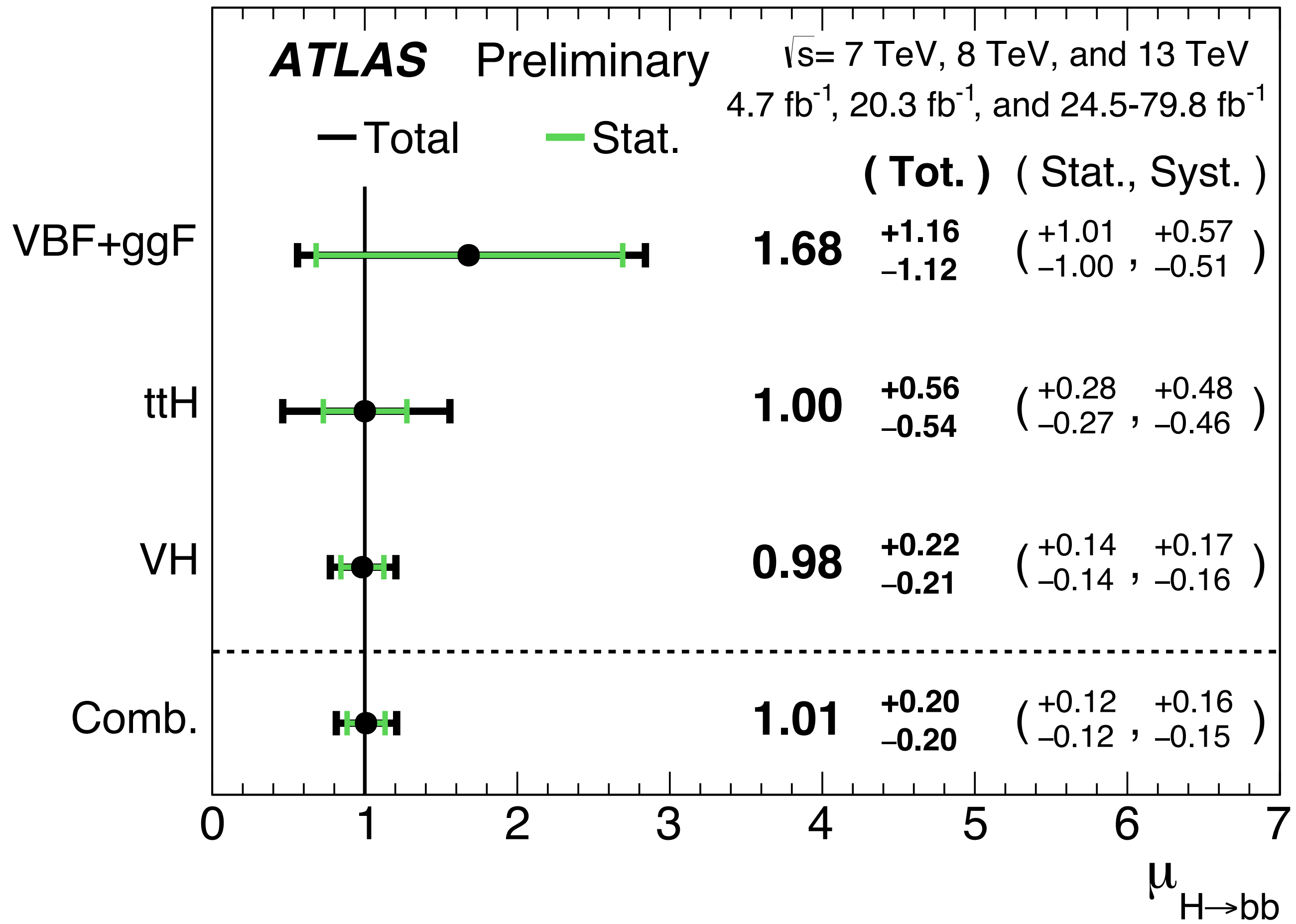

***First observation of H(b$\bar{b}$) decay***

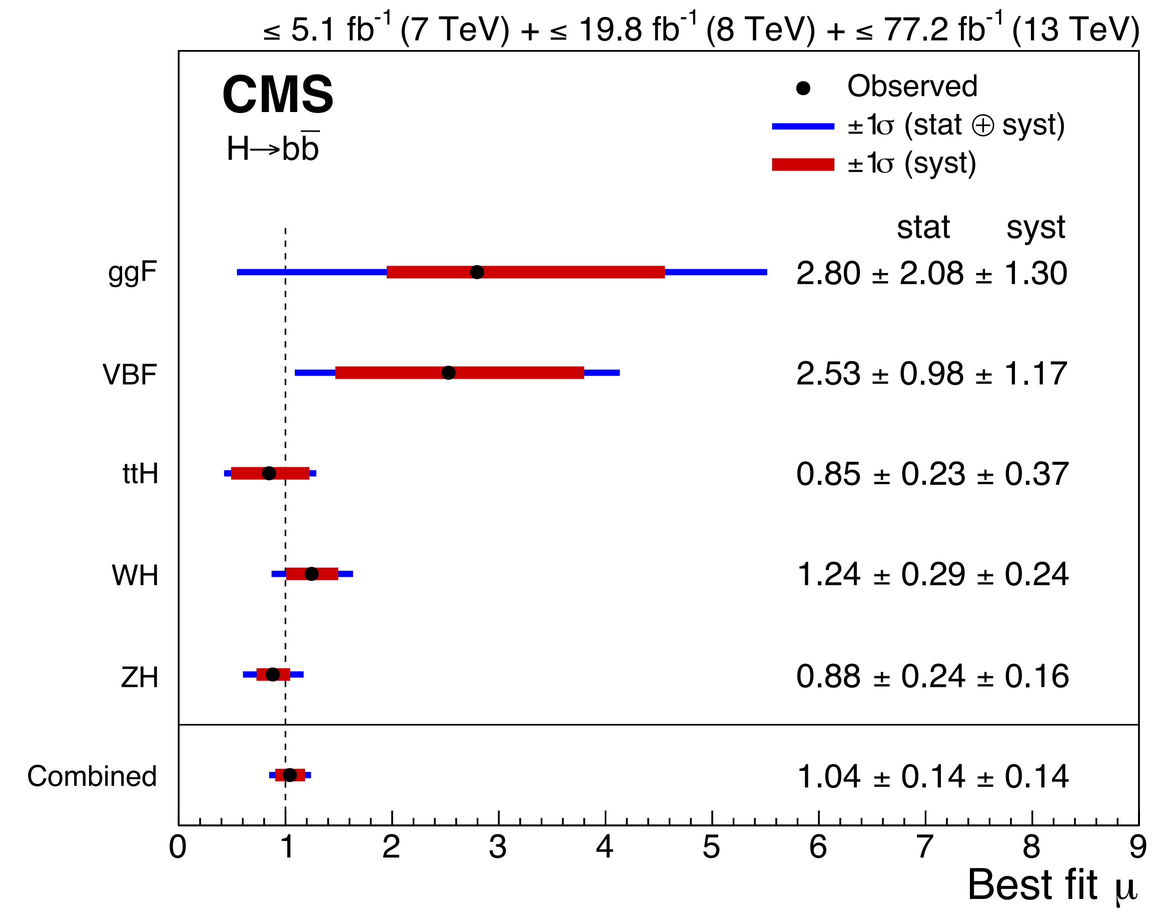
***CMS*** *Run 1+2:* ***5.6 (5.5) σ***

***ATLAS*** *Run 1+2:* ***5.4 (5.5) σ***

CMS-PRL 120 (2018) 231801
ATLAS-Phys. Lett. B 786 (2018) 59

# H($B\bar{B}$) OBSERVATION



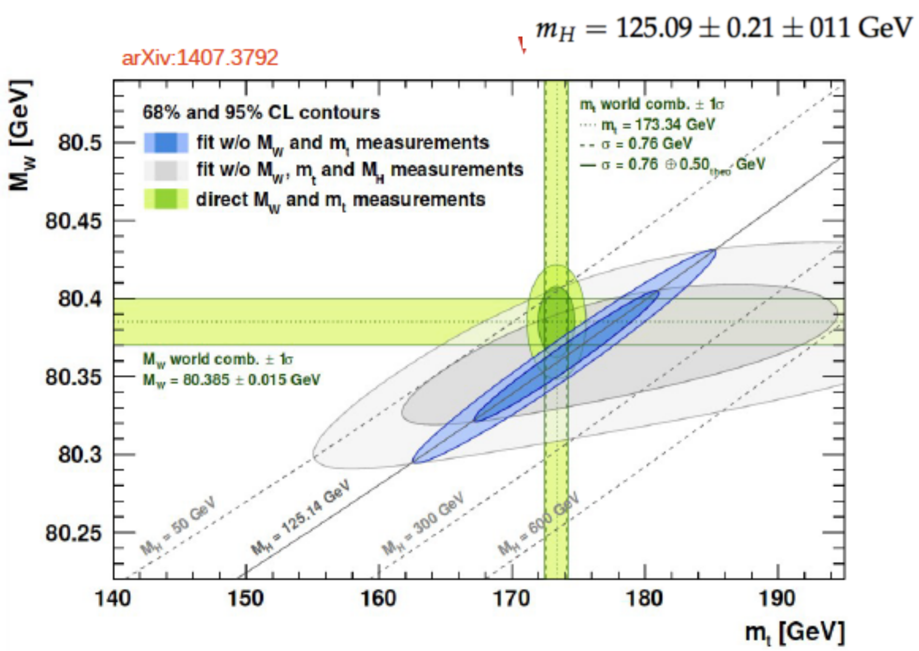

***First observation of H($b\bar{b}$) decay***

***CMS Run 1+2: 5.6 (5.5) σ***
***ATLAS Run 1+2: 5.4 (5.5) σ***

# A STANDARD MODEL PRECISION MEASUREMENT THE W MASS

# MAIN CONCEPTS TO FOCUS ON

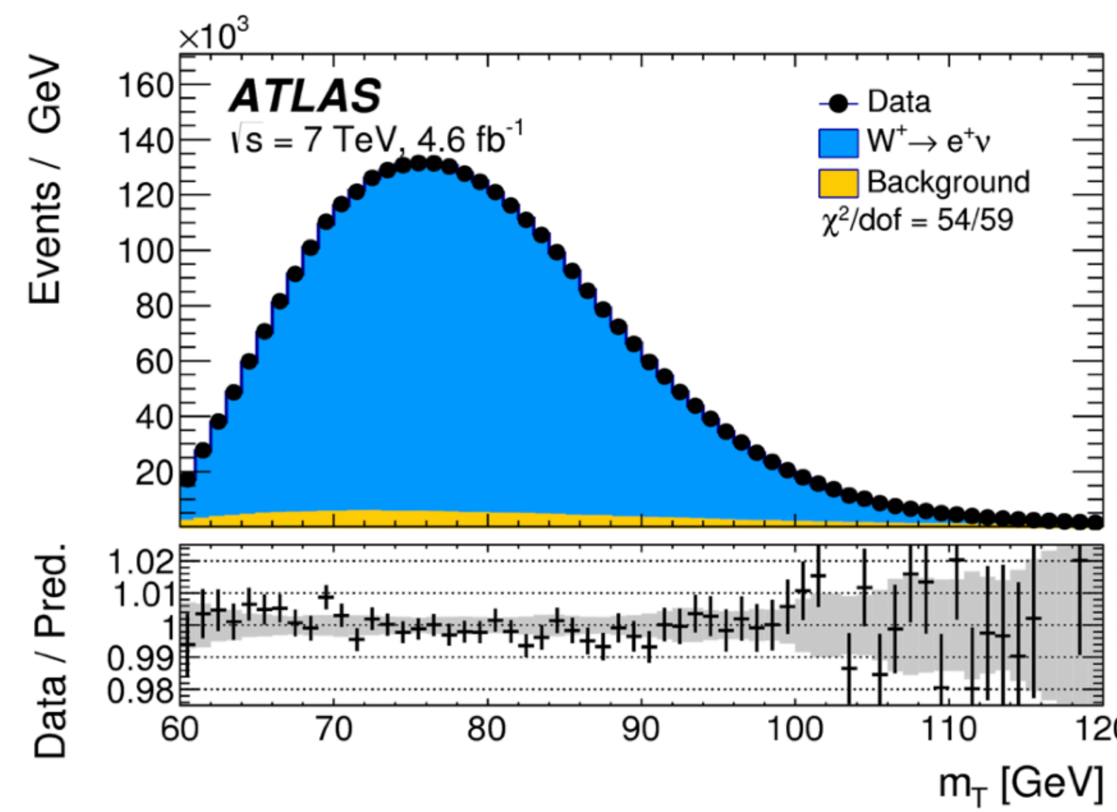
- a very visible Standard Model signal: lots of statistics!
- need to measure very precisely a fundamental parameter of the SM
- absolute control of all the sources of uncertanties: experimental and theoretical




arXiv:1608.01509

| | measurement | posterior | SM prediction |
|---|---|---|---|
| $m_H$ | $125.09 \pm 0.24$ | $125.09 \pm 0.24$ | $102.8 \pm 26.3$ |
| $m_{top}$ | $173.34 \pm 0.76$ | $173.61 \pm 0.63$ | $176.6 \pm 2.5$ |
| $m_W$ | $80.385 \pm 0.015$ | $80.3644 \pm 0.0061$ | $80.3604 \pm 0.0066$ |

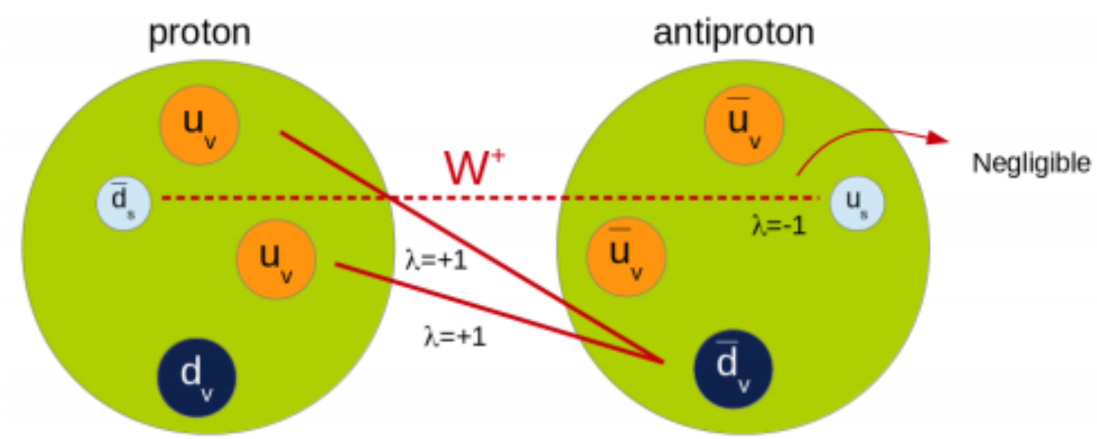
GeV



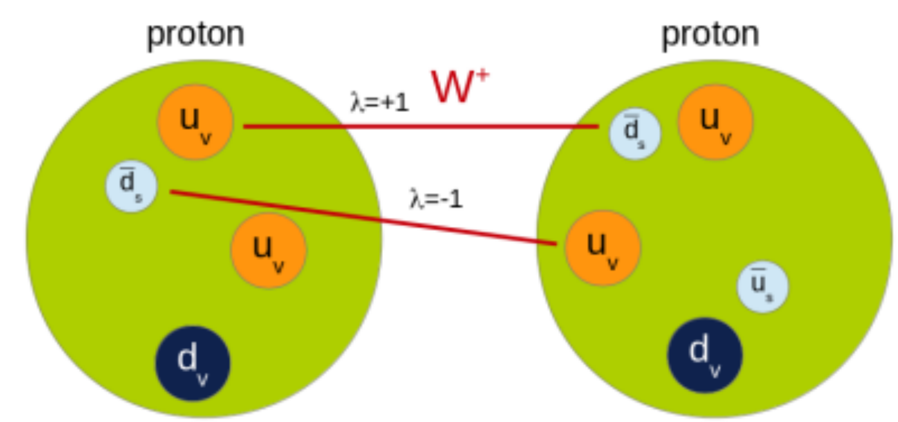

learn the challenges of a precision measurement

# WHY IT IS HARD(ER) AT THE LHC

A proton-proton collider is the most challenging enviroment to measure $m_W$, worse compared to e+e- and proton-antiproton




In $p\bar{p}$ collisions W bosons are mostly produced in the same helicity state




In pp collisions they are equally distributed between positive and negative helicity states

Large PDF-induced W-polarisation uncertainty affecting the $p_T$ lepton distribution

Further QCD complications

25% vs 5%

- Heavy-flavour-initiated processes
- W+, W- and Z are produced by different light flavour fractions

40% more $W^+$

- Larger gluon-induced W production

Larger Z samples, available for detector calibration given the precisely known Z mass → most of the measurement is then the transfer from Z to W

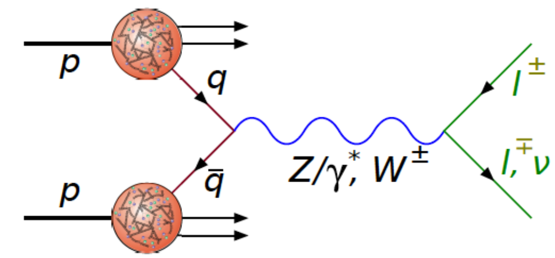
# STRATEGY OF THE MEASUREMENT

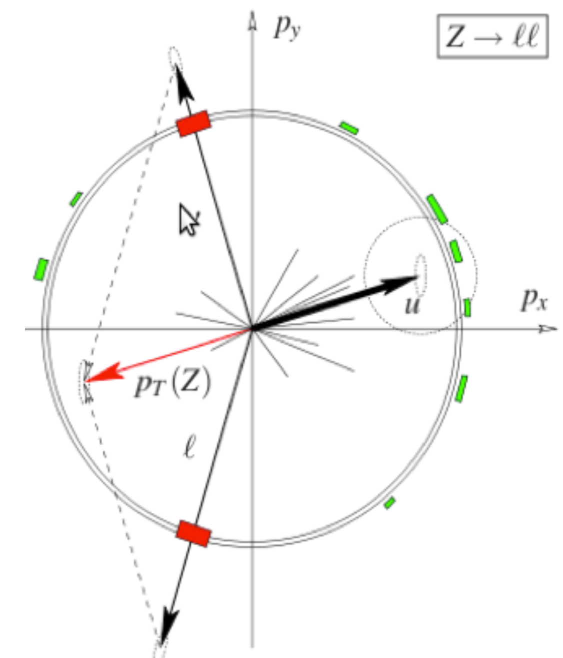
Not possible to fully reconstruct W mass

Sensitive final state distributions: $p_T^l$, $m_T$, $p_T^{miss*}$

$$\vec{p}_T^{\,\rm miss} = -\left(\vec{p}_T^{\,\ell} + \vec{u}_T\right), \quad m_T = \sqrt{2 p_T^{\ell} p_T^{\rm miss}(1 - \cos\Delta\phi)}$$

$u_T$ being the recoil




Benefit from the fully reconstructed mass in Z-boson sample to validate the analysis and to provide significant experimental *(lepton and recoil calibration using resp. $m_Z$ measured at LEP and expected momentum balance with $p_T^{ll}$)* and theoretical constraints *(ancilliary measurements).*

The whole analysis is checked by performing a measurement of the Z-boson mass and comparing to the LEP value, also a cross-check Z mass measurement in "W-like" i.e removing the 2nd lepton and treating it like a neutrino

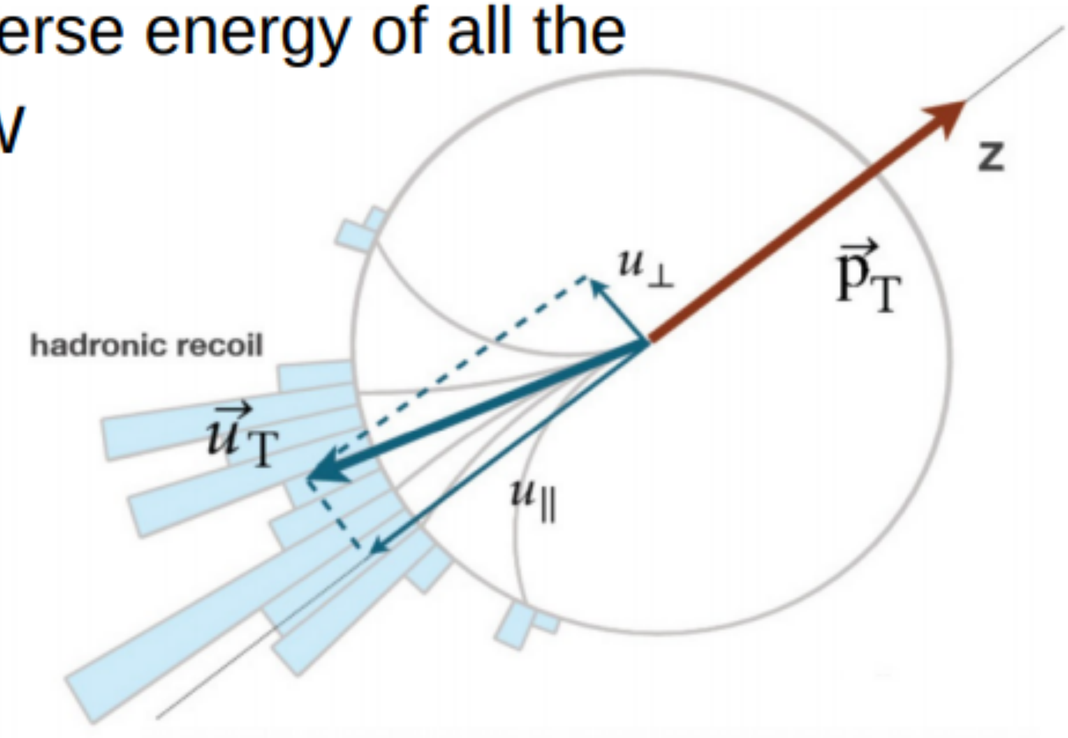
Need to consider additional systematics for W mass measurement (*theory uncertainties, Z—>W extrapolation and background)*

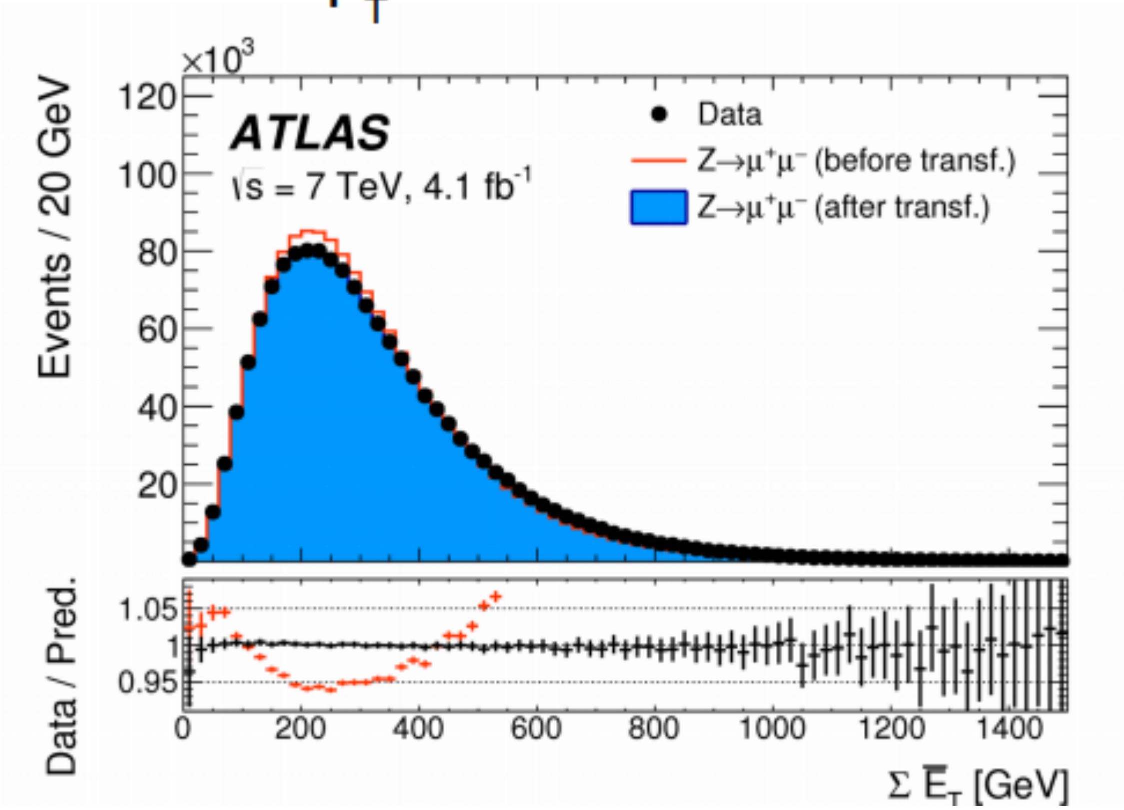
# THE RECOIL

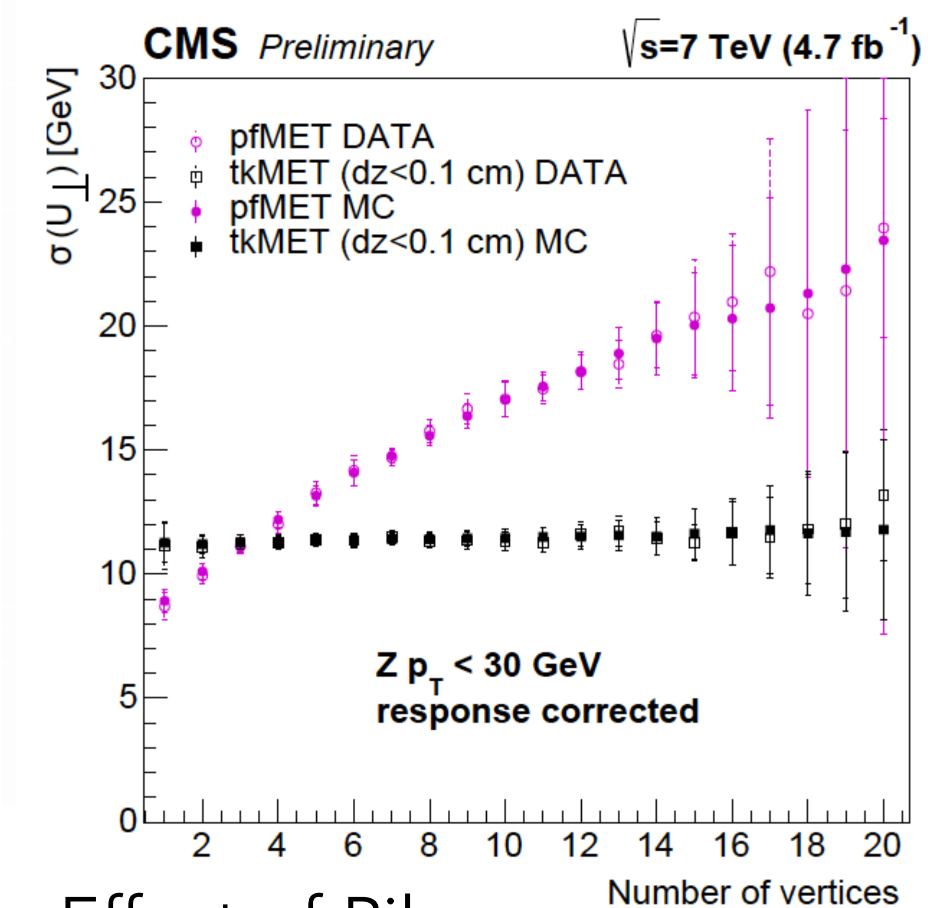
The recoil $u_T$ is the vector sum of the transverse energy of all the calorimeter clusters: $u_T$ is a measure of $p_T$ W

Calibration steps:

- Correct pile-up multiplicity in MC to match the data
- Correct for residual differences in the $\Sigma E_T$ distribution
- Derive scale and resolution corrections from the $p_T$ balance in Z events



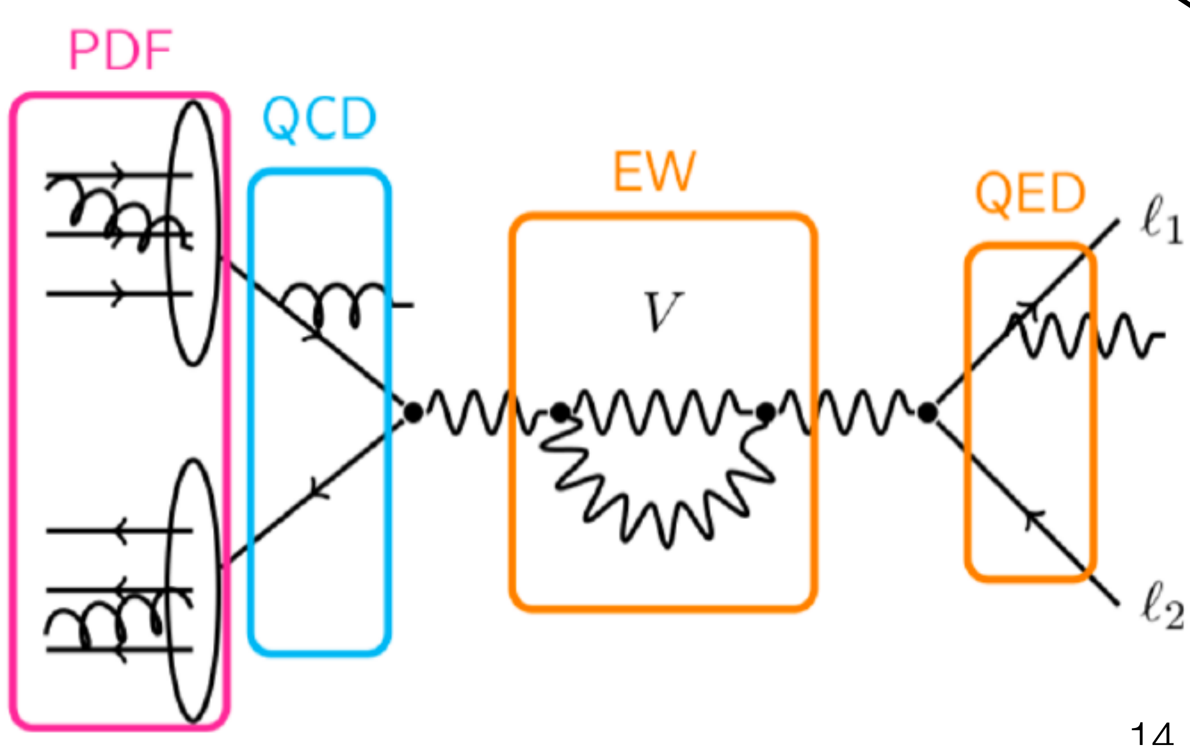

Effect of Pileup

# PHYSICS MODELLING

No single generator able to describe all observed distributions.

Start from the Powheg+Pythia8 and apply corrections. Use ancillary measurements of Drell-Yan processes to validate (and tune) the model and assess systematic uncertainties.

Physics modelling corrections




Electroweak corrections

- *QED FSR and ISR (included)*
- *missing higher order effects and FSR pair production (uncertainties)*

QCD corrections

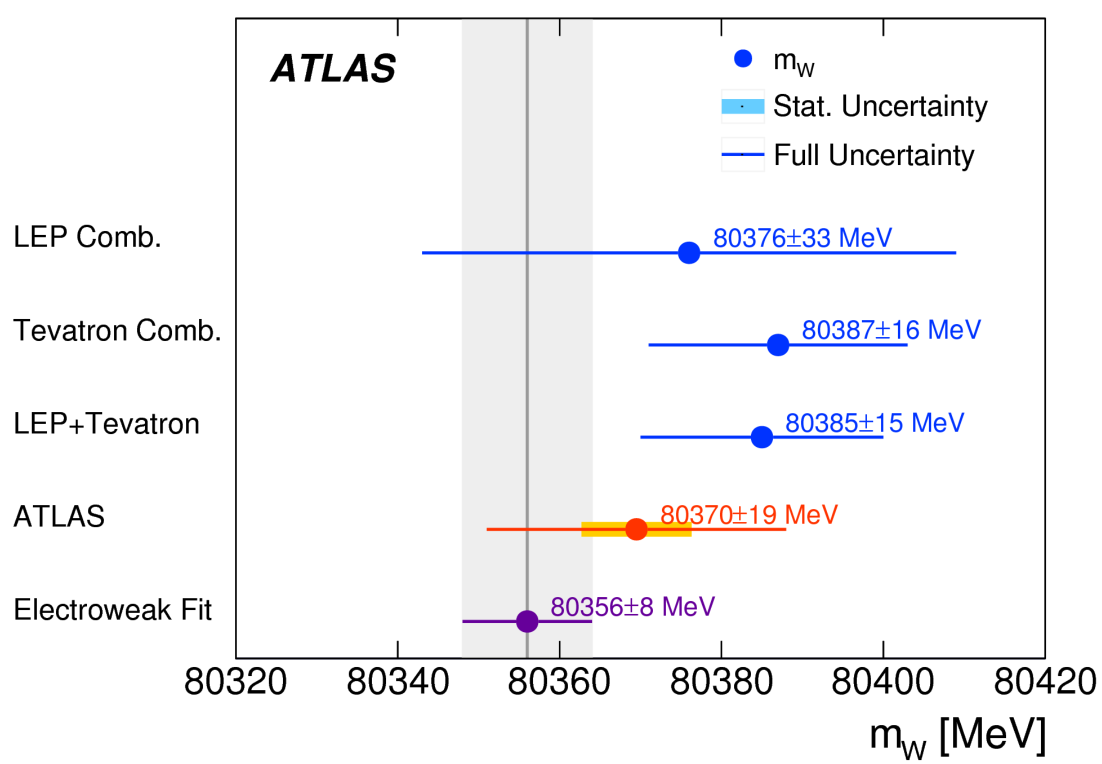
- *pT distribution*
- *polarisation*
- *rapidity*

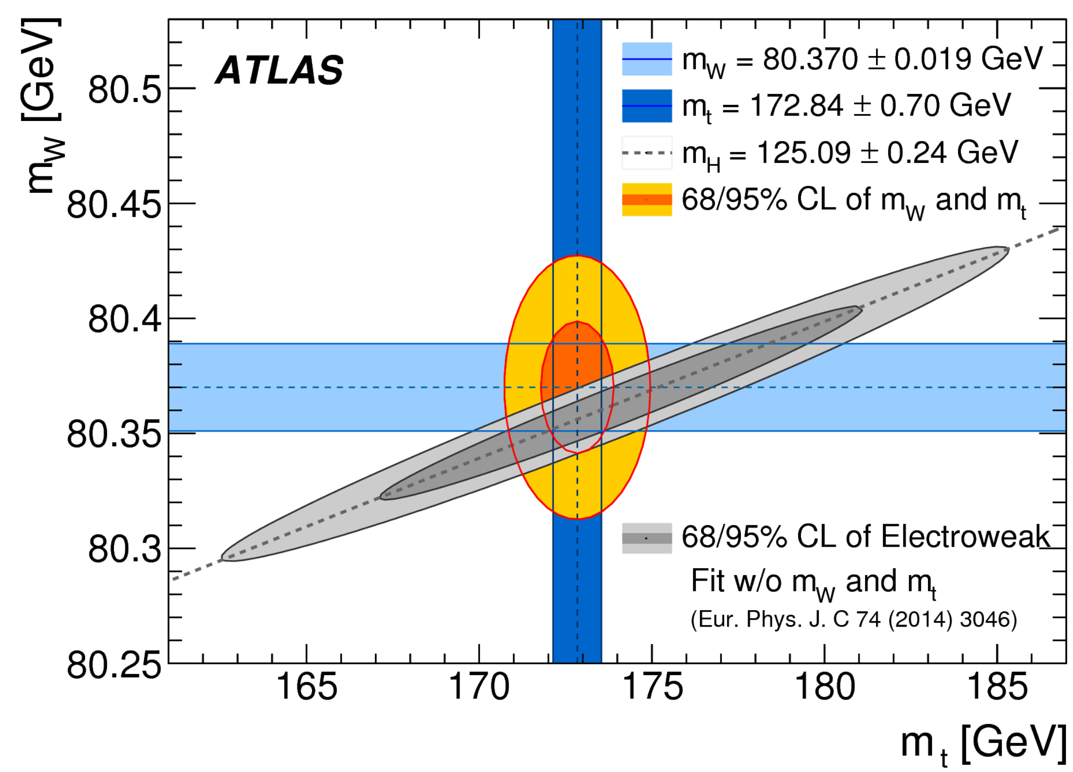
# RESULTS

7.8M W->μν
5.8M W->eν

$$m_W = 80369.5 \pm 6.8\text{ MeV(stat.)} \pm 10.6\text{ MeV(exp. syst.)} \pm 13.6\text{ MeV(mod. syst.)}$$
$$= 80369.5 \pm 18.5\text{ MeV},$$

| Combined categories | Value [MeV] | Stat. Unc. | Muon Unc. | Elec. Unc. | Recoil Unc. | Bckg. Unc. | QCD Unc. | EWK Unc. | PDF Unc. | Total Unc. | $\chi^2$/dof of Comb. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $m_T$-$p_T^\ell$, $W^\pm$, e-$\mu$ | 80369.5 | 6.8 | 6.6 | 6.4 | 2.9 | 4.5 | 8.3 | 5.5 | 9.2 | 18.5 | 29/27 |




The result is consistent with the SM expectation, compatible with the world average and competitive in precision to the currently leading measurements by CDF and D0

# A SEARCH FOR AN EXOTIC SIGNATURE

# MAIN CONCEPTS TO FOCUS ON

- searching for the production of a new exotic particle beyond the Standard Model
  - theory model to drive the analysis strategy
- heavy particle: will be on the tail of the kinematical distribution. Small background but also small signal
- how do we define our sensitivity?

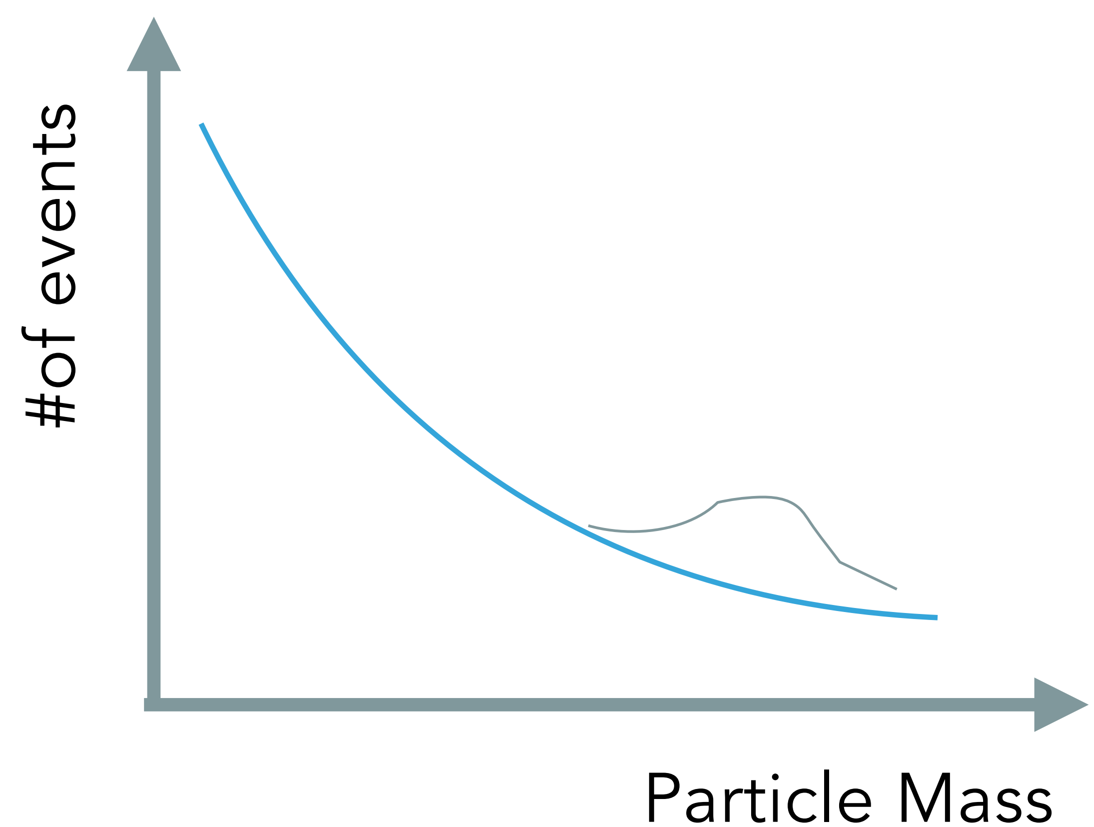
**how to optimize the search strategy and determine the discovery reach or an exclusion limit**

*Ignoring here « difficult » searches with unusual signatures requiring original reconstruction techniques*

# SEARCHES FOR HIGH-MASS/PT SIGNATURES

- When looking for physics BSM the high-mass/high-pt signal region is the one with the cleanest signatures.
  - « Bumps » for resonance production in relevant distributions like the new particle mass for instance
- Typically higher energy preferred for these searches (100 TeV pp machine, FCC-hh)
- Searches could take advantage from intensity frontier as well, helps processes with small production cross section

Plethora of NP models predicting resonances studied at LHC and HL-LHC (and future colliders)

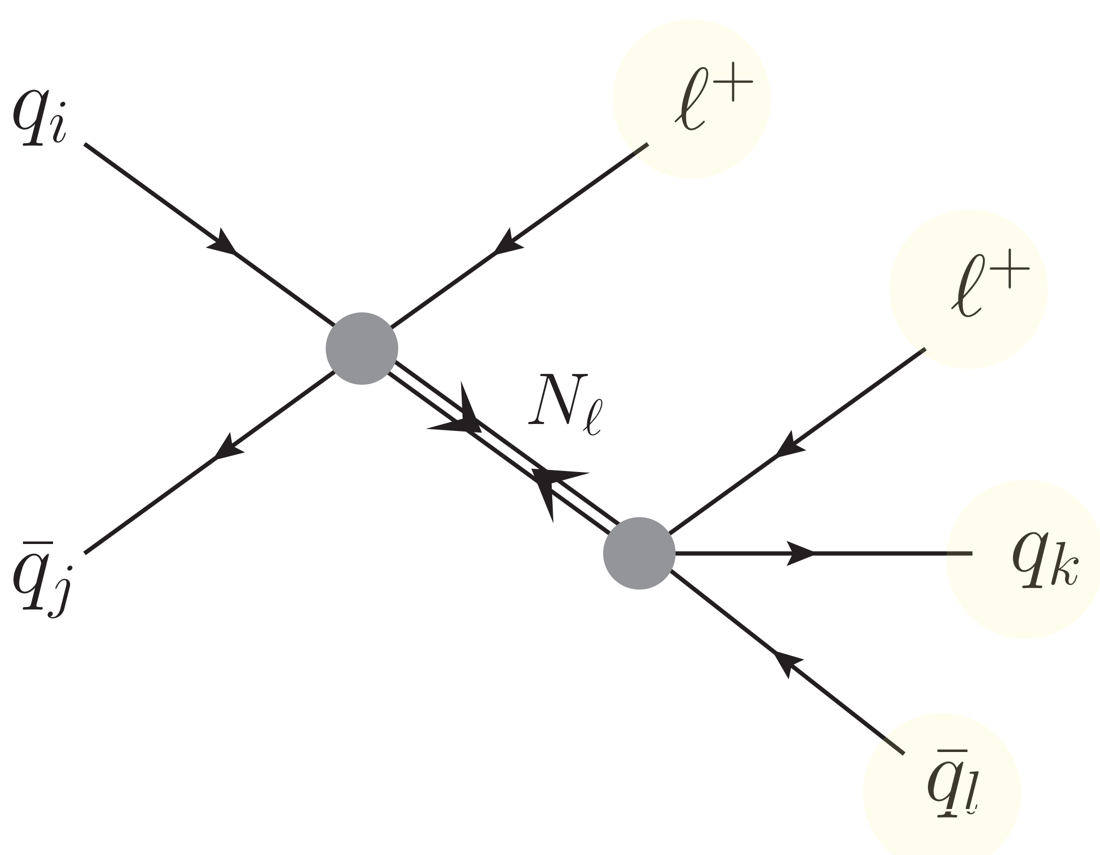


# CHARACTERISTICS & STRATEGY

- Usually these types of searches look for a small number of event on the « tail » of the known SM processes distribution.
- The strategy is to use the theory model developed by the theorists and implemented in MC generators to derive the kinematical properties of the BSM events and their signature.
  - Goal is to maximize the signal acceptance while keeping a way to control the background.
- In these cases the « background » is small, and subject to statistical fluctuations.
  - Need to devise strategies to be able to extrapolate the background from a large statistics region (and usually different kinematical regime) to the signal one

# A PRACTICAL EXAMPLE: HEAVY COMPOSITE NEUTRINOS

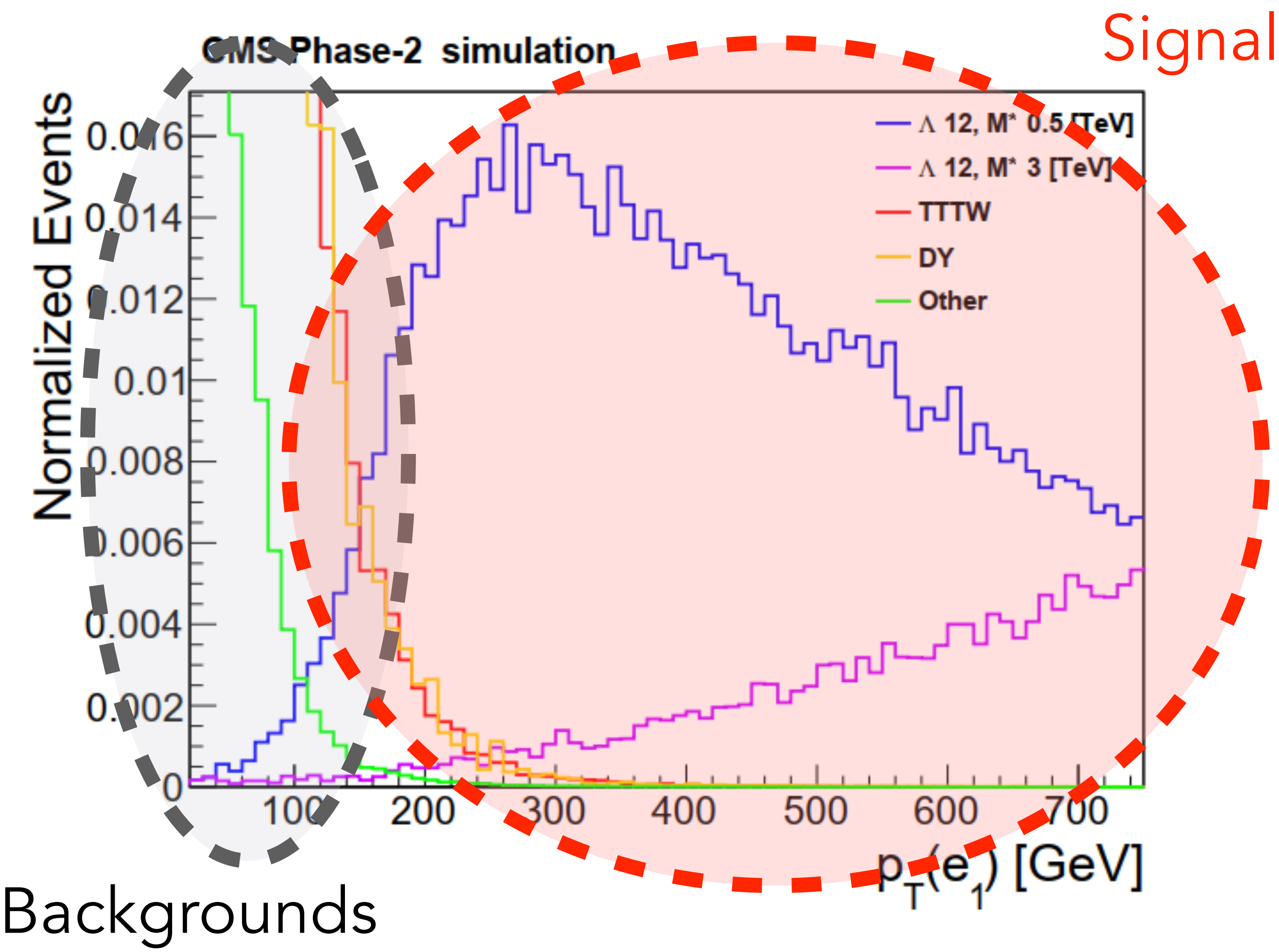
- Compositeness of leptons and quarks is one possible BSM scenario. Fermions are assumed to have an internal substructure
- If quarks and leptons are composite we can expect excited states of quarks and leptons like the Heavy Composite Majorana Neutrino (HCMN).
  - They will be massive otherwise we would have seen them already!
- In this example the new heavy particle HCMN:
  - is produced along with a lepton
  - it decays in two quarks and one lepton




the final state is a di-leptons+di-jets signature lljj, l=e, μ

Considering the sensitivity for a search at the HL-LHC, $\sqrt{s}$=14 TeV and 3ab$^{-1}$

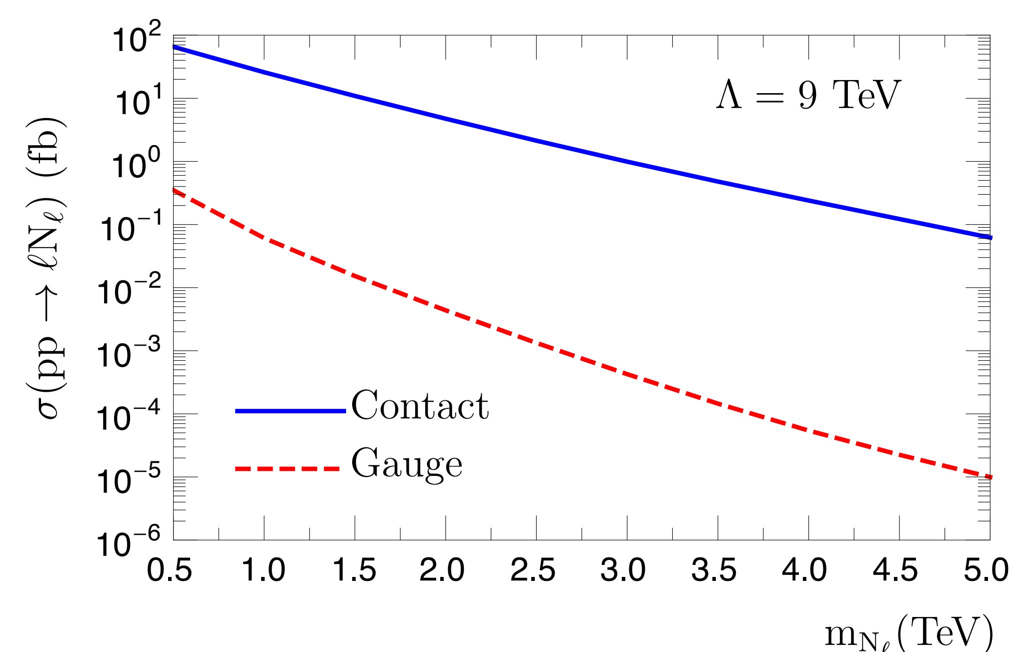
# KINEMATIC OF THE SIGNAL AND BACKGROUND




pT distribution for the leading lepton

(plot normalized to unity)

- The kinematic of the background is very different from the signal
- The difference is bigger as a function of the new particle mass

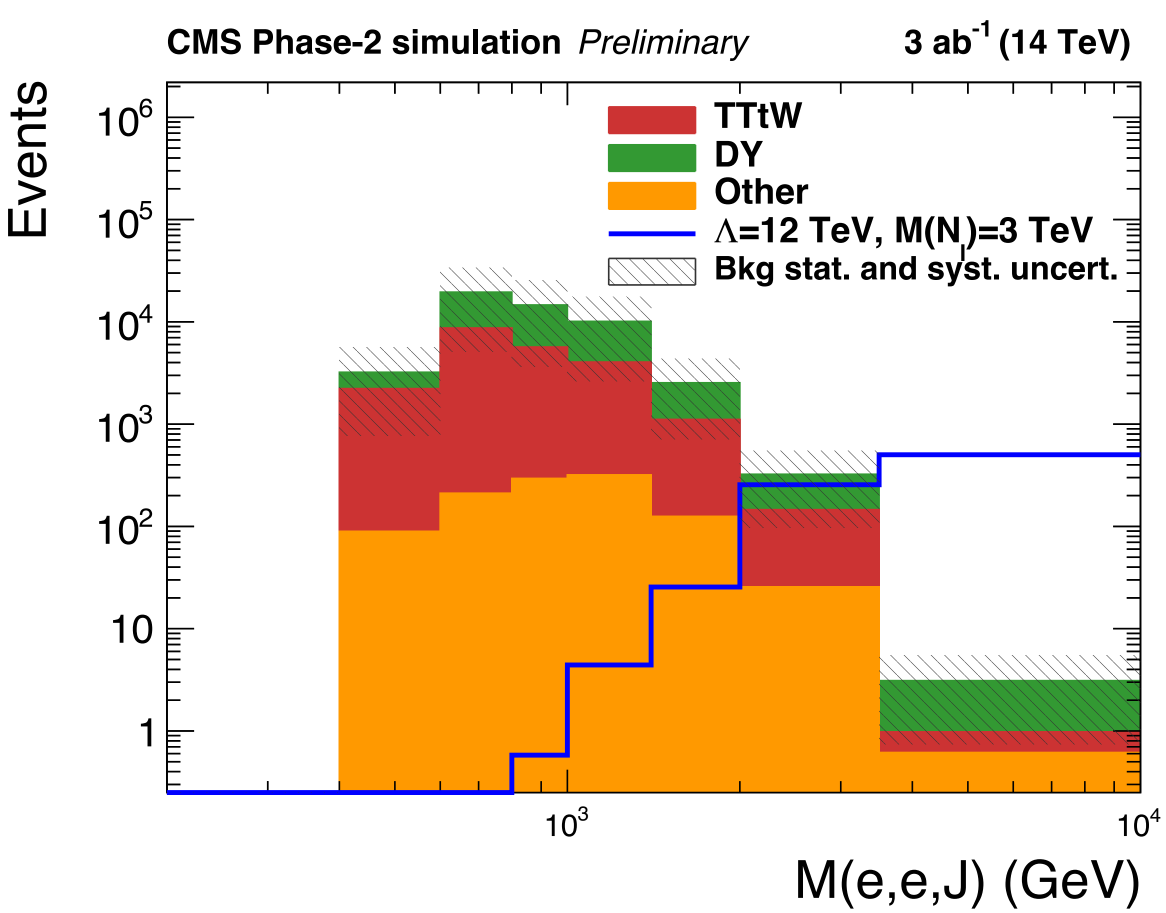



However, the signal production cross section goes down with the mass:

the number of expected signal events becomes small too

# ANALYSIS STEPS

- Studying the kinematical distributions of signal and background an optimal selection is defined. i.e. kinematical and topological cuts on the different event variables
  - Efficiencies are evaluated for the signal
  - Various methods are employed to estimate the expected background (some data driven, some MC only)
  - Uncertainties are assigned to all components
- The invariant mass M(llj) corresponding to the hypotetical HCN is chosen as a discriminating variable for the final fit.




given our observed data we fit the distribution to our knowledge of expected background.

In this example we have only MC as it is a study for a future machine, the HL-LHC

# DIGRESSION: MONTECARLO ONLY STUDIES

- MC only studies are very useful in the case of searches to estimate a priori the sensitivity to the specific signal in specific conditions
  - particle collider, center of mass energy, integrated luminosity, detector choices
  - These studies are essential for the design and choices of new machines and detectors
- A middle way is the concept of «RECAST »:
  - sometimes a final state can correspond to several different models
  - A published analysis optimized on the model X can be can be used to extract results on model Y. This is done studying the behavior of model Y in MC with the cuts of the analysis.
  - It is a procedure used frequently mostly by theorists to check new ideas against real data

# SYSTEMATICS FOR A FUTURE ANALYSIS

- How to estimate uncertainties for the future? Making reasonable assumptions.
- When we want to know how an analysis will perform in the future we need to put ourselves in the conditions of advanced knowledge and statistics of the future.
- See our example here:

| Obtained from | (Run 2) Value Sig (Bkg) [%] |
|---|---|
| UPGAnalysisSystematics | 2.7 (2.7) |
| UPGAnalysisSystematics | 2 (4) |
| UPGAnalysisSystematics | 2 (1) |
| UPGAnalysisSystematics | - (2) |
| UPGAnalysisSystematics | 2 (3) |
| UPGAnalysisSystematics | - (2) |
| UPGAnalysisSystematics | 0.4 (2) |
| UPGAnalysisSystematics | 1 (3) |
| EXO-16-026 | - (15) |
| EXO-16-026, scaled by 0.5 | - (8) |

| Source | Value |
|---|---|
| Luminosity | 1% |
| Pileup | 2% |
| Electron ID | 0.5% |
| Electron scale | 0.5% |
| Muon ID | 0.5% |
| Muon scale | 0.5% |
| Jet energy scale | 1% |
| Jet energy resolution | 1% |
| Background | 0.3% |
| Drell-Yan (theory) | 4% |

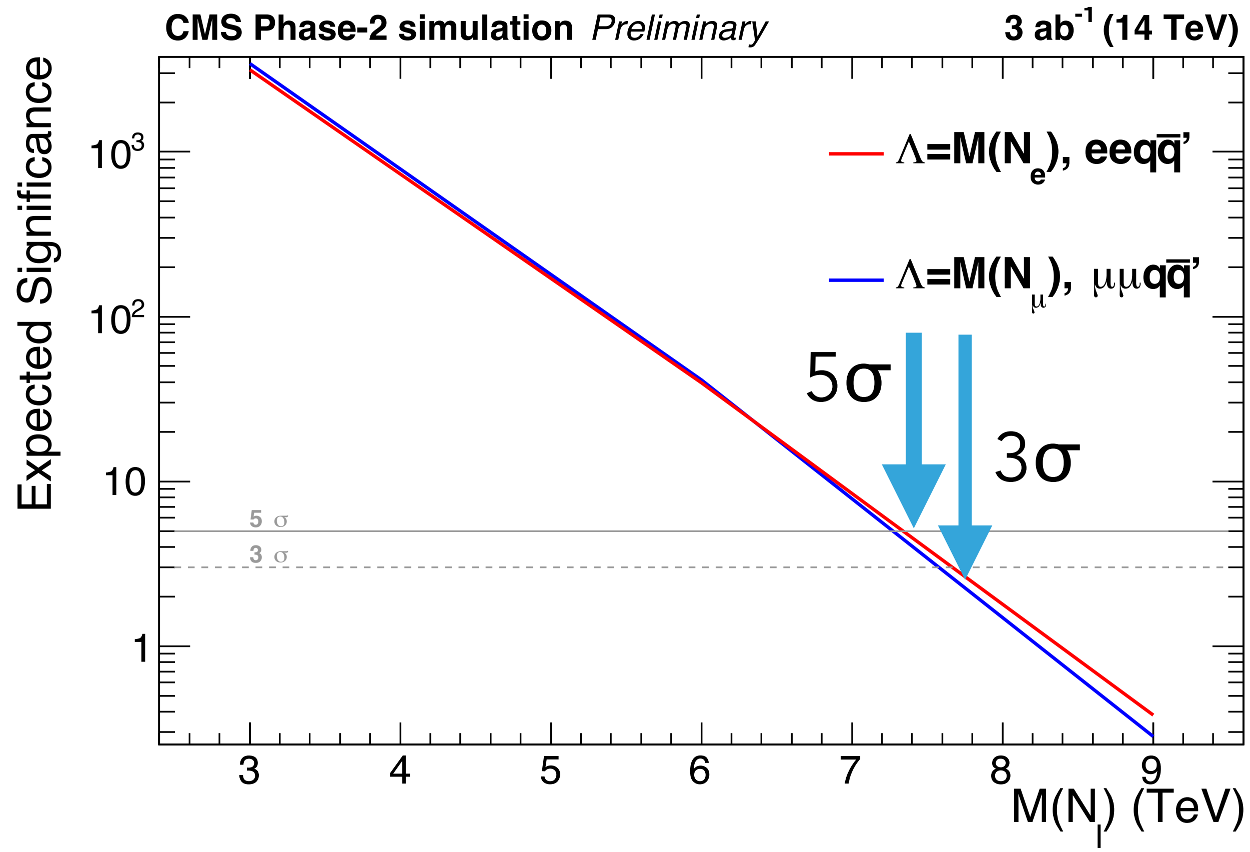
# DISCOVERY OR EXCLUSION?

- When searching for a new particle physicist are confronted with two cases:
  - For discovery: H0=Background only and H1=Background+signal
  - For exclusion: H0=Background+Signal and H1=Background only
- The probability (p-value) of the null hypotesis (H0) is calculated,
  - i.e. the probability of finding data of equal or greater incompatibility with the prediction of H0
- Hypotesis is excluded for different pre-established thresholds:
  - for exclusion: $p<0.05$ (i.e. 95% exclusion limit)
  - for discovery: $p<0.003$ (3 sigma) and $p<3\text{x}10^{-7}$ (5sigma)
    - (the concept of sigma comes from the standard deviation of a normal distribution, 68% of the data is withing 1sigma, 95% within 2sigma etc)
- Special case to deal with those cases where the data have a downward fluctuation than the background only and avoid exclusions that are « too good » $CL_s$:
  - $CL_s=p_{s+b}/(1-p_b)$ the exclusion limit becomes more conservative. We dilute the compatibility with S+B with the incompatibility of B-only hypotesis.

# PROSPECTS FOR THE HCMN SEARCH

- The results are expressed in terms of sensitivity for a discovery or extensions of the excluded region.
- The expected statistical significance for both the eeqq¯′ and μμqq¯′ channel for the case Λ = M(Nl)

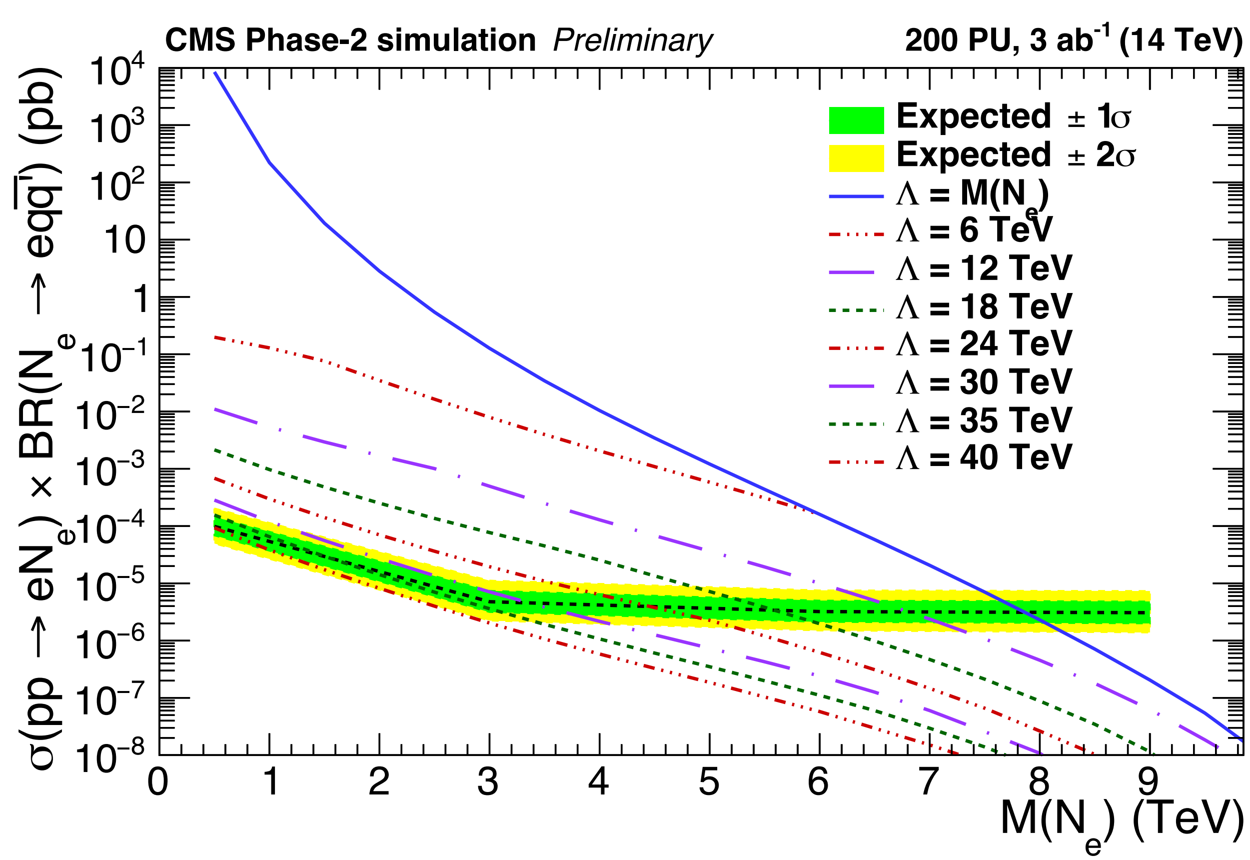
Discovery up to Λ = M($N_l$) =7 TeV




- Tthe expected 95% CL upper limits on σ(pp → lNl) × B(Nl → lqq¯′) for the eeqq¯′ as a function of the HCMN mass

Exclusion up to Λ = M($N_l$) =8 TeV

(run 2 was 4.6TeV)

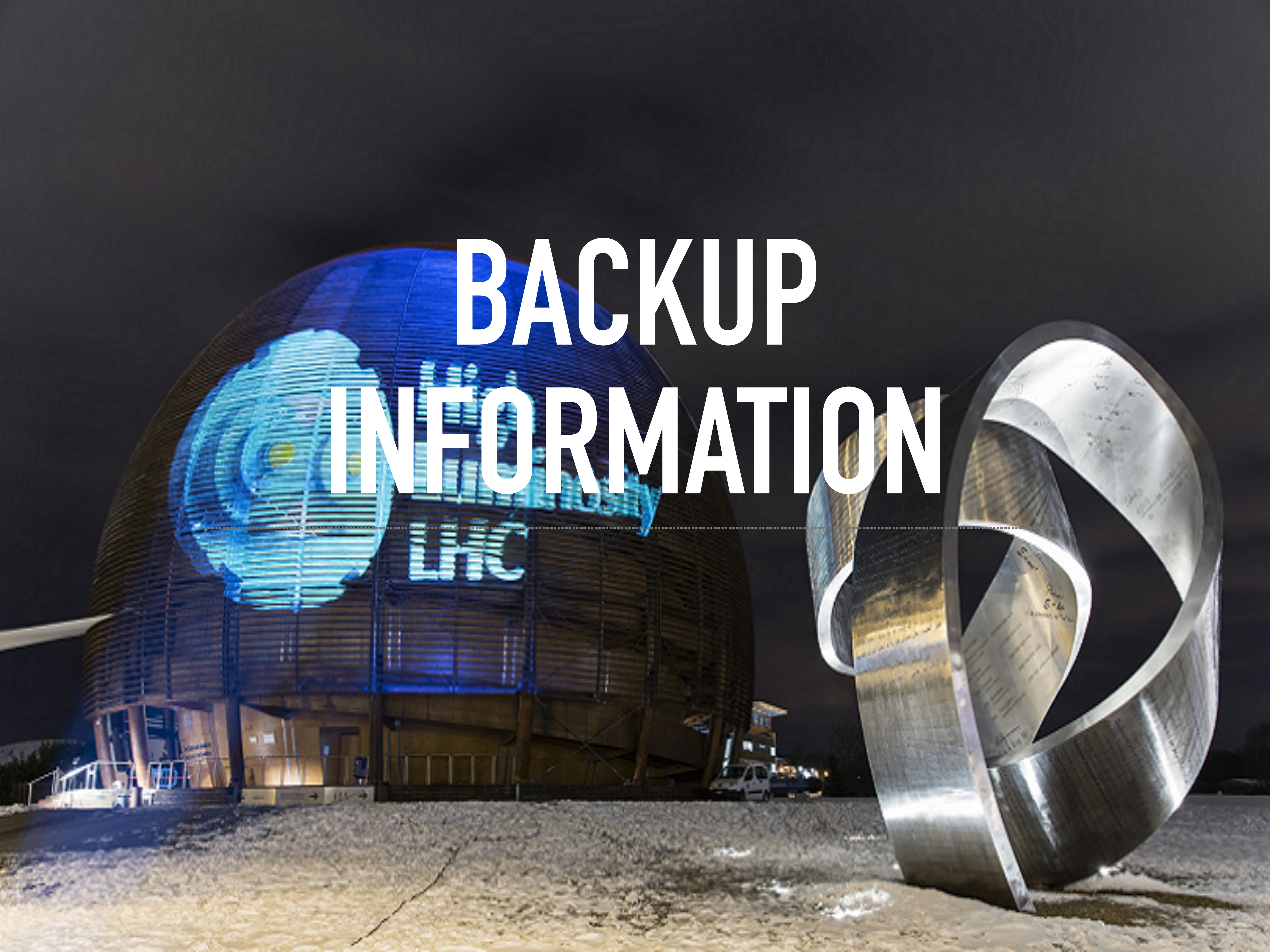


# SUMMARY

- Very quick tour of the analysis strategies employed at the LHC. Hopefully clarification of some of the concepts and jargon that we can find on published research papers
- The spectrum of approaches is extremely different depending on the type of measurement we are interested into
  - studying a new signal
  - measuring a quantity very precisely
  - searching for a signature beyond the Standard Model
- For each of these approaches there is a vaste number of actual implementation of algorithms and techniques depending on the specific case

*the challenge to devise a smarter and more efficient method to obtain a better result or a discovery is possibly the most fascinating aspect of the job of an experimental particle physicst*

# BACKUP INFORMATION



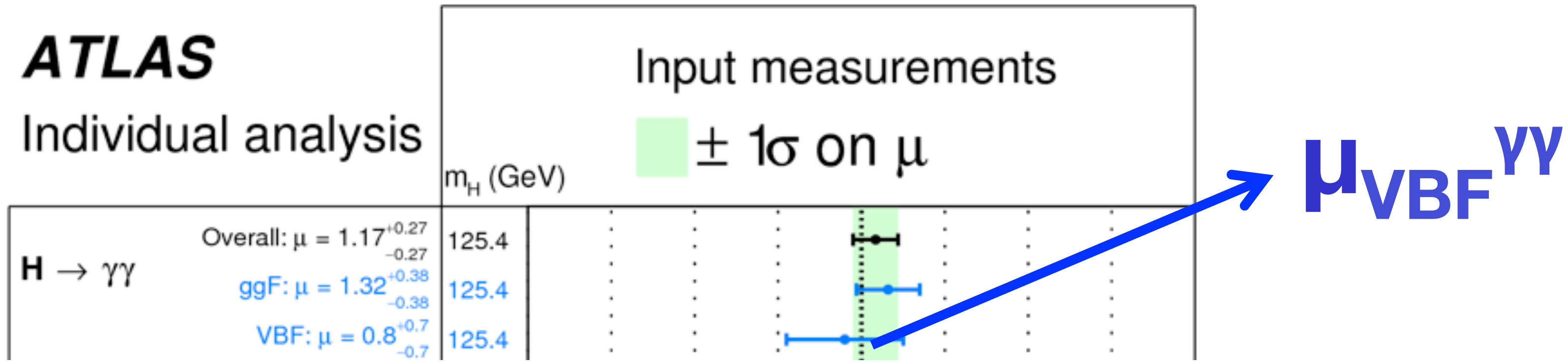
# VH EVENT SELECTION

| | Variable | 0-lepton | 1-lepton | 2-lepton |
|---|---|---|---|---|
| Boson momentum | $p_{\mathrm{T}}(\mathrm{V})$ | $>170$ | $>100$ | $[50,150], >150$ |
| Dilepton Mass | $M(\ell\ell)$ | — | — | $[75,105]$ |
| | $p_{\mathrm{T}}^{\ell}$ | — | $(>25, >30)$ | $>20$ |
| | $p_{\mathrm{T}}(\mathrm{j}_1)$ | $>60$ | $>25$ | $>20$ |
| | $p_{\mathrm{T}}(\mathrm{j}_2)$ | $>35$ | $>25$ | $>20$ |
| | $p_{\mathrm{T}}(\mathrm{jj})$ | $>120$ | $>100$ | — |
| | $M(\mathrm{jj})$ | $[60,160]$ | $[90,150]$ | $[90,150]$ |
| | $\Delta\phi(\mathrm{V},\mathrm{jj})$ | $>2.0$ | $>2.5$ | $>2.5$ |
| | $\mathrm{CMVA}_{\mathrm{max}}$ | $>\mathrm{CMVA}_{\mathrm{T}}$ | $>\mathrm{CMVA}_{\mathrm{T}}$ | $>\mathrm{CMVA}_{\mathrm{L}}$ |
| | $\mathrm{CMVA}_{\mathrm{min}}$ | $>\mathrm{CMVA}_{\mathrm{L}}$ | $>\mathrm{CMVA}_{\mathrm{L}}$ | $>\mathrm{CMVA}_{\mathrm{L}}$ |
| | $N_{\mathrm{aj}}$ | $<2$ | $<2$ | — |
| | $N_{\mathrm{a}\ell}$ | $=0$ | $=0$ | — |
| | $p_{\mathrm{T}}^{\mathrm{miss}}$ | $>170$ | — | — |
| | $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\mathrm{j})$ | $>0.5$ | — | — |
| | $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\vec{p}_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{trk}))$ | $<0.5$ | — | — |
| | $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\ell)$ | — | $<2.0$ | — |
| | Lepton isolation | — | $<0.06$ | $(<0.25, <0.15)$ |
| | Event BDT | $>-0.8$ | $>0.3$ | $>-0.8$ |

# Understanding signal strengths for process i → H → f

- Signal strength μ is observed rated normalized by SM prediction

$$\mu_i^f \equiv \frac{\sigma_i \cdot \mathrm{BR}^f}{(\sigma_i \cdot \mathrm{BR}^f)_{\mathrm{SM}}} = \mu_i \times \mu^f$$




- Disentangling production ($\mu_i$) & decay ($\mu_f$) always **requires assumption of narrow Higgs width.**
- Additional assumptions required when combining measurements, e.g

$$\mu_i = \frac{\sigma_i}{\sigma_i^{\mathrm{SM}}} \quad \text{and} \quad \mu^f = \frac{\mathrm{BR}^f}{\mathrm{BR}^f_{\mathrm{SM}}}.$$

*Assumes SM value of decay BRs*

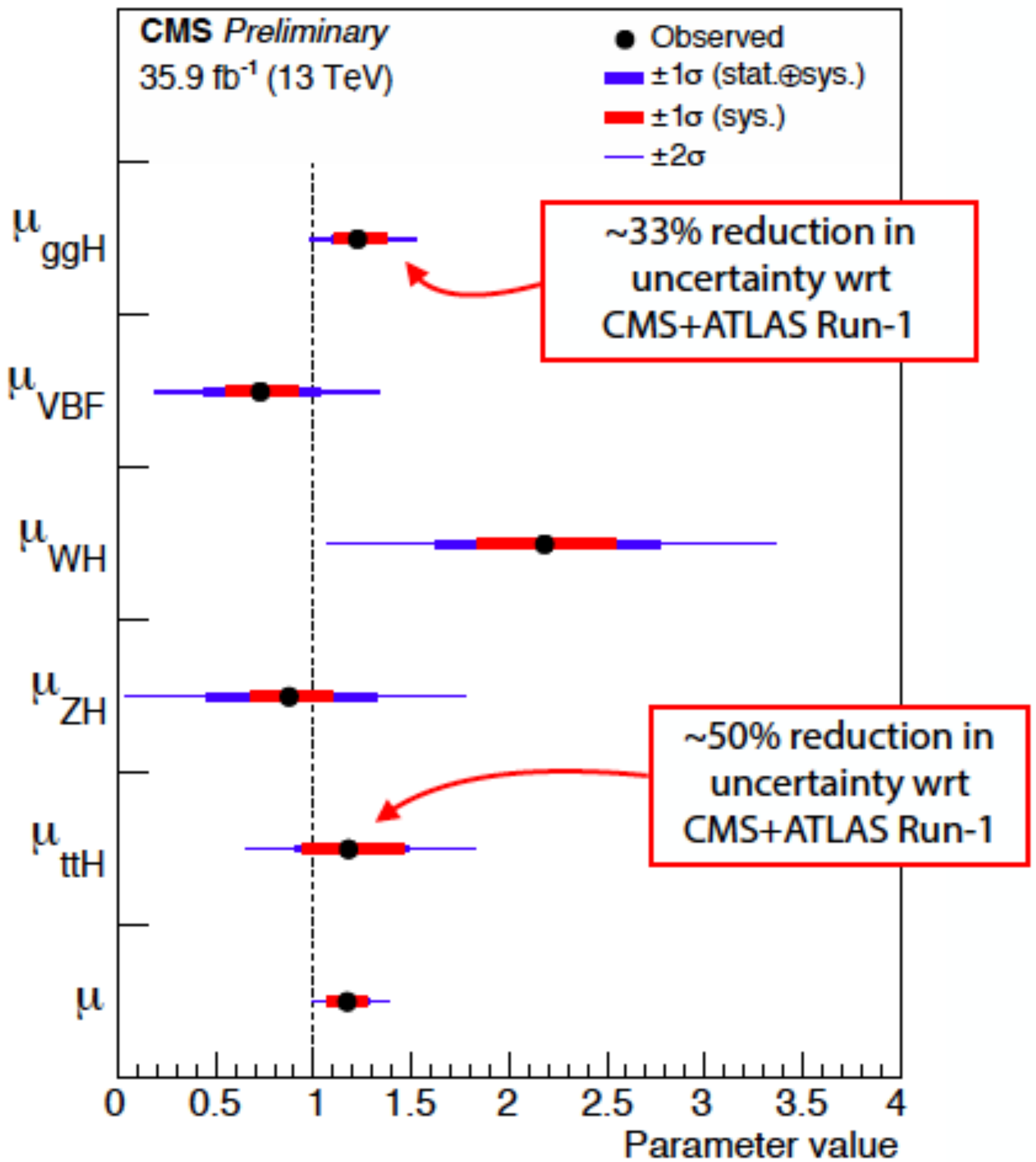
*Assumes SM value of production σ's*

# Signal strength measurements

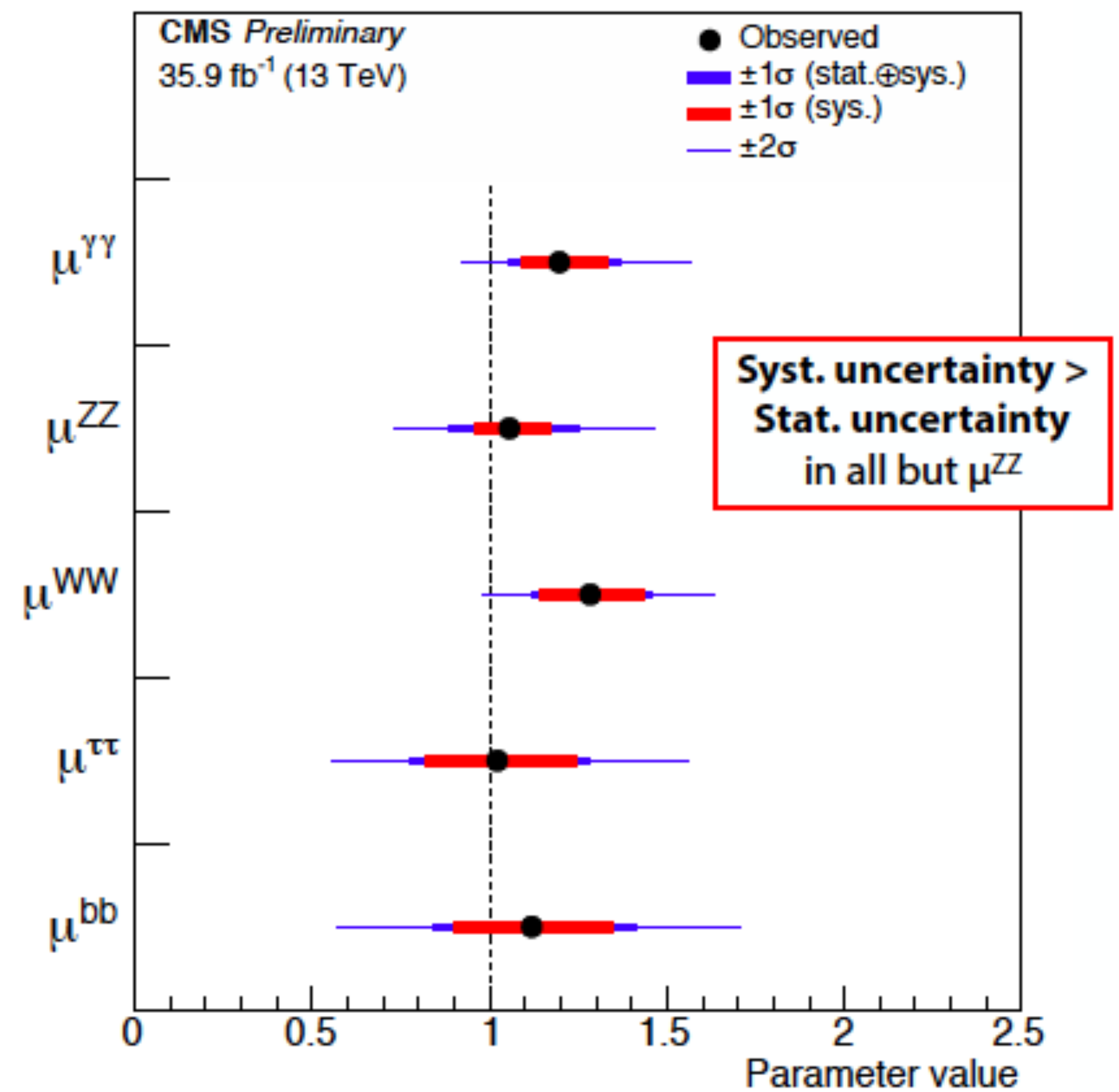
$$\mu = 1.17^{+0.10}_{-0.10} \quad = 1.17^{+0.06}_{-0.06}\,(\text{stat.})\,^{+0.06}_{-0.05}\,(\text{sig. th.})\,^{+0.06}_{-0.06}\,(\text{other sys.})$$

- Split by production or decay mode:

Run 2 summary

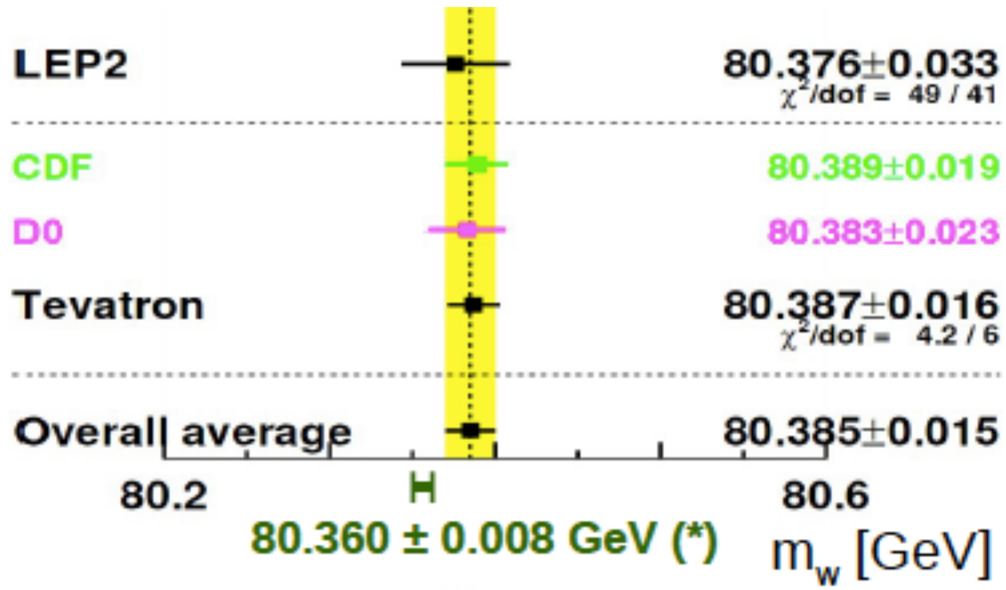



**Per production mode**




**Per decay mode**

# COMPARISON OF SYSTEMATICS




New ATLAS Measurement
december 2016

**mW = 80.370 ± 19 MeV**

CDF

| Source | Uncertainty |
|---|---|
| Lepton energy scale and resolution | 7 |
| Recoil energy scale and resolution | 6 |
| Lepton tower removal | 2 |
| Backgrounds | 3 |
| PDFs | 10 |
| $p_T(W)$ model | 5 |
| Photon radiation | 4 |
| Statistical | 12 |
| Total combination | 19 |

$$625{,}000\ W \to \mu\nu$$
$$470{,}000\ W \to e\nu$$

ATLAS

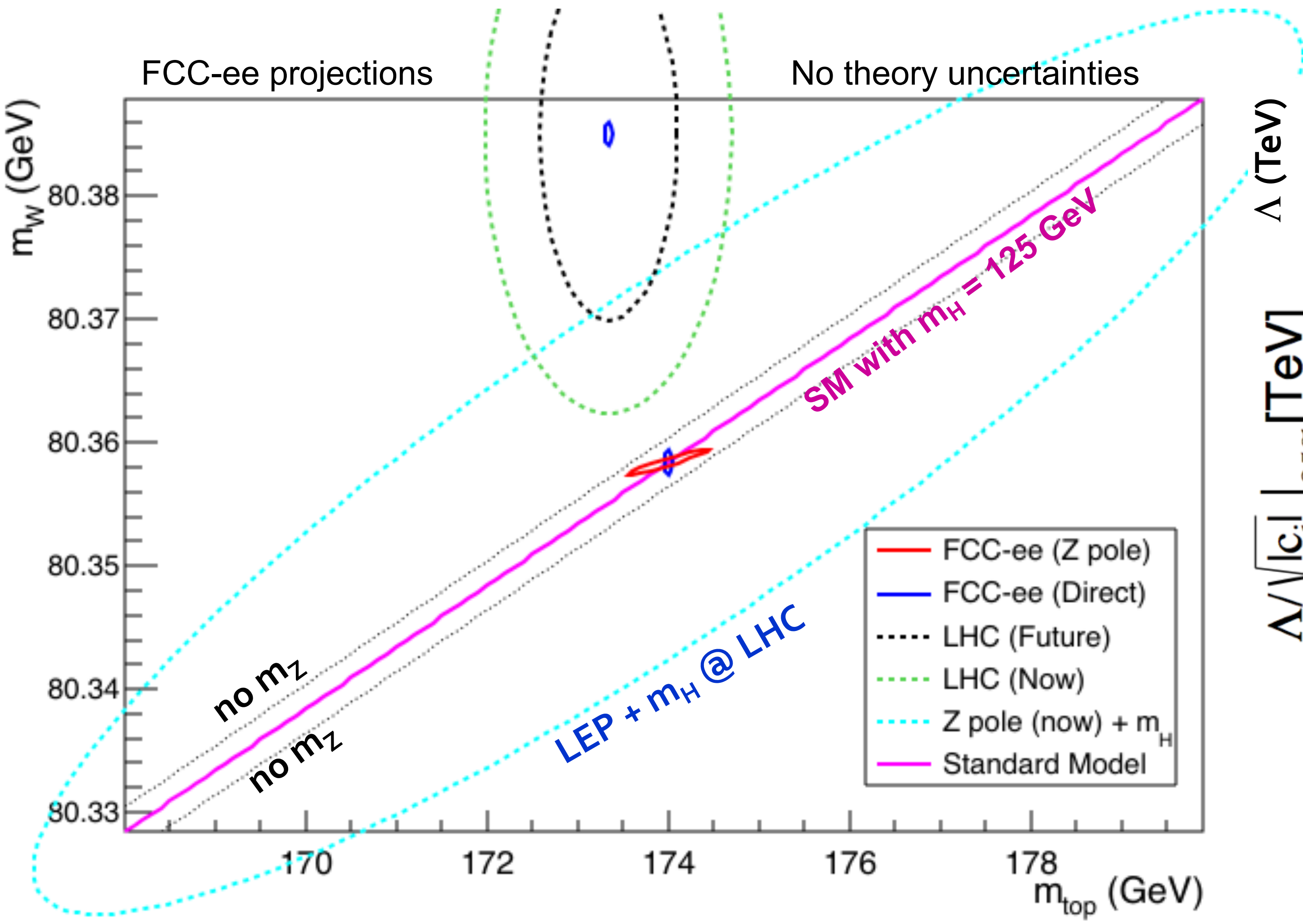
| Source | Uncertainty (Gigi's Breakdown) | ATLAS Numbers |
|---|---|---|
| Lepton energy scale and resolution | 9-14 | |
| Recoil energy scale and resolution | 2-13 | **11** |
| Lepton tower removal | | |
| Backgrounds | 6-12 | |
| PDFs | 8-9 | **14** |
| $p_T(W)$ model | 8-10 | |
| Photon radiation | 5-3 | |
| Statistical | 7 | **7** |
| Total combination | 19 | |

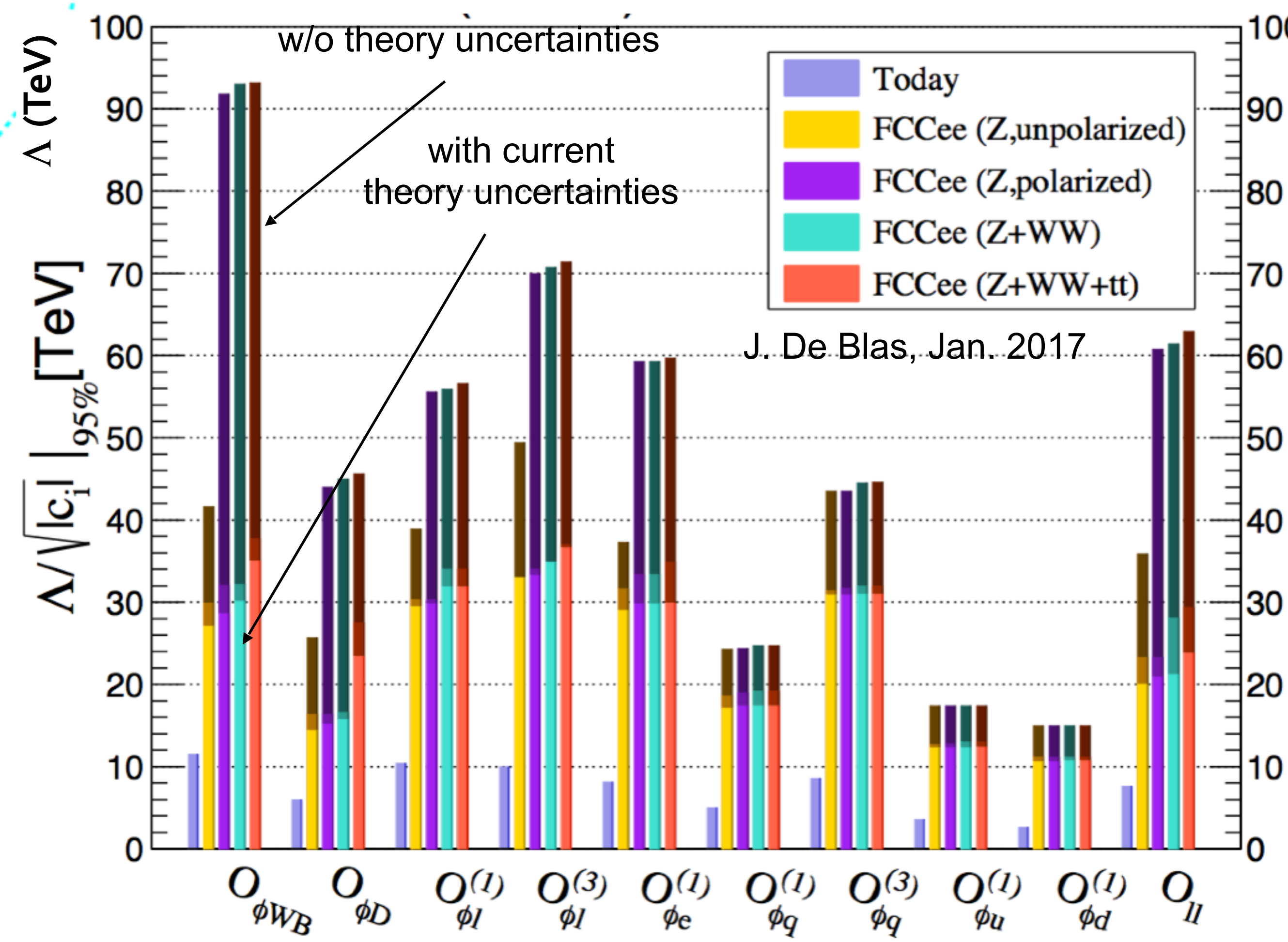
$$7.8\ M\ W \to \mu\nu$$
$$5.8\ M\ W \to e\nu$$

# PRECISION ⇔ DISCOVERY

- Combining all FCC-ee EW measurements
  - In the context of the SM … and beyond

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i$$




**Requires 10-fold improved theory calculations**

Points to the physics to be looked for at FCC-hh

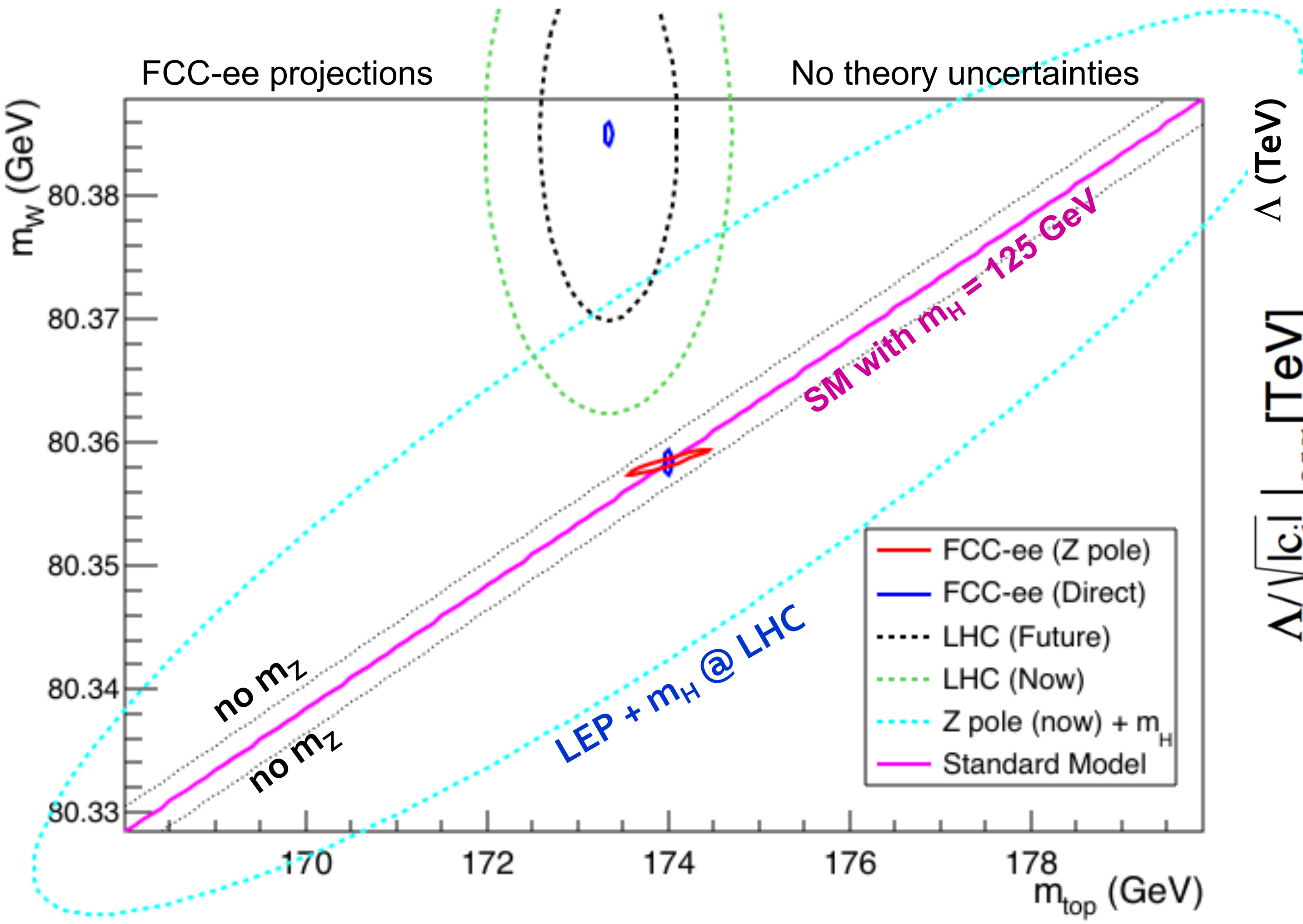
Today: $\Lambda/\sqrt{c}$ **> 5-10 TeV**

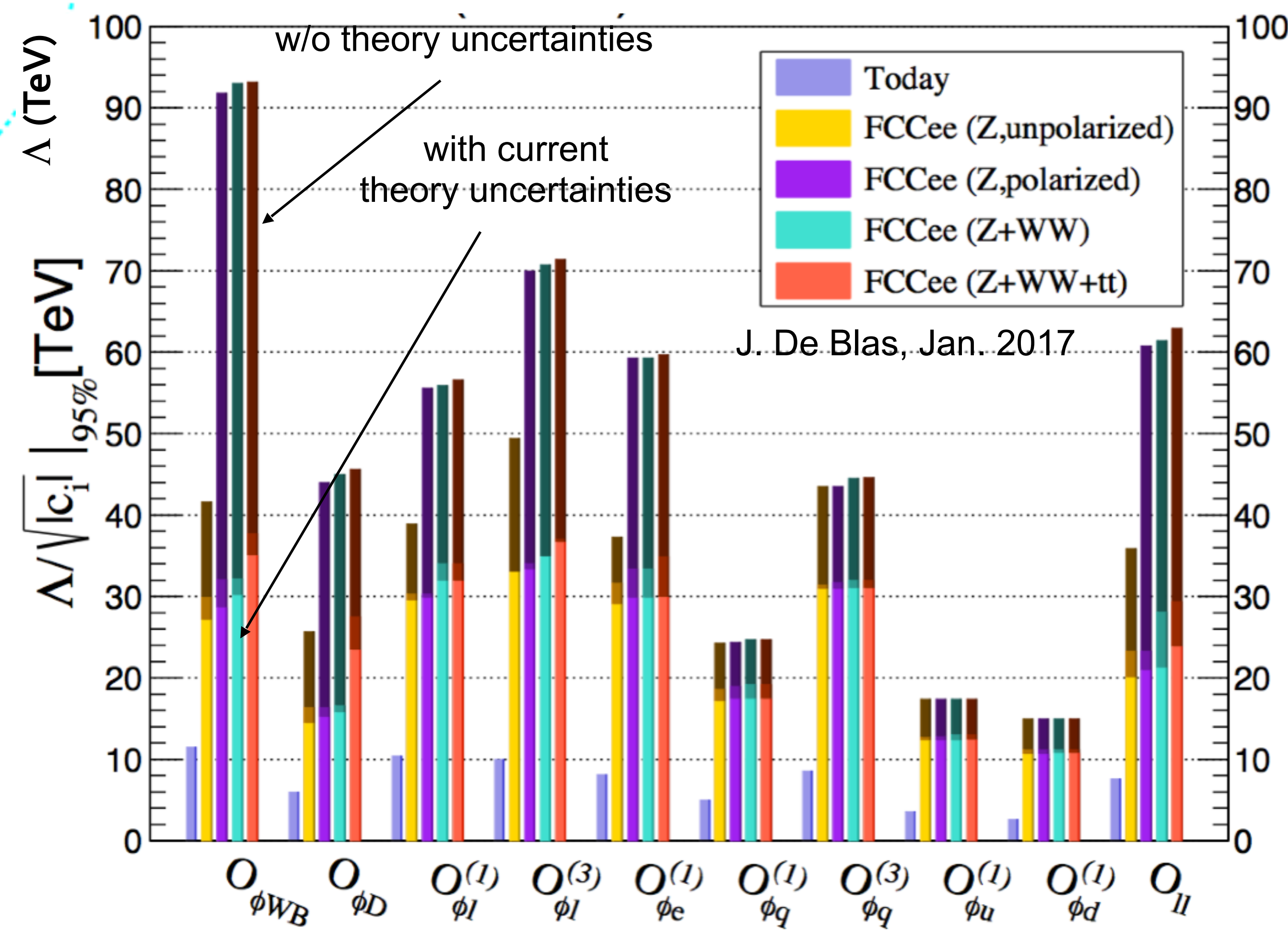
- New physics: blue and red ellipses may not overlap
  - Or even better, data may not fit to the SM

# PRECISION ⇔ DISCOVERY

- Combining all FCC-ee EW measurements
  - In the context of the SM … and beyond

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i$$




**Requires 10-fold improved theory calculations**

Points to the physics to be looked for at FCC-hh

Today: $\Lambda/\sqrt{c}$ **> 5-10 TeV**

- New physics: blue and red ellipses may not overlap
  - Or even better, data may not fit to the SM

**After FCC-ee: $\Lambda/\sqrt{c}$ > 50-100 TeV ?**

**What do we mean by "Sensitivity to NP up the scale of N TeV?" e.g.**

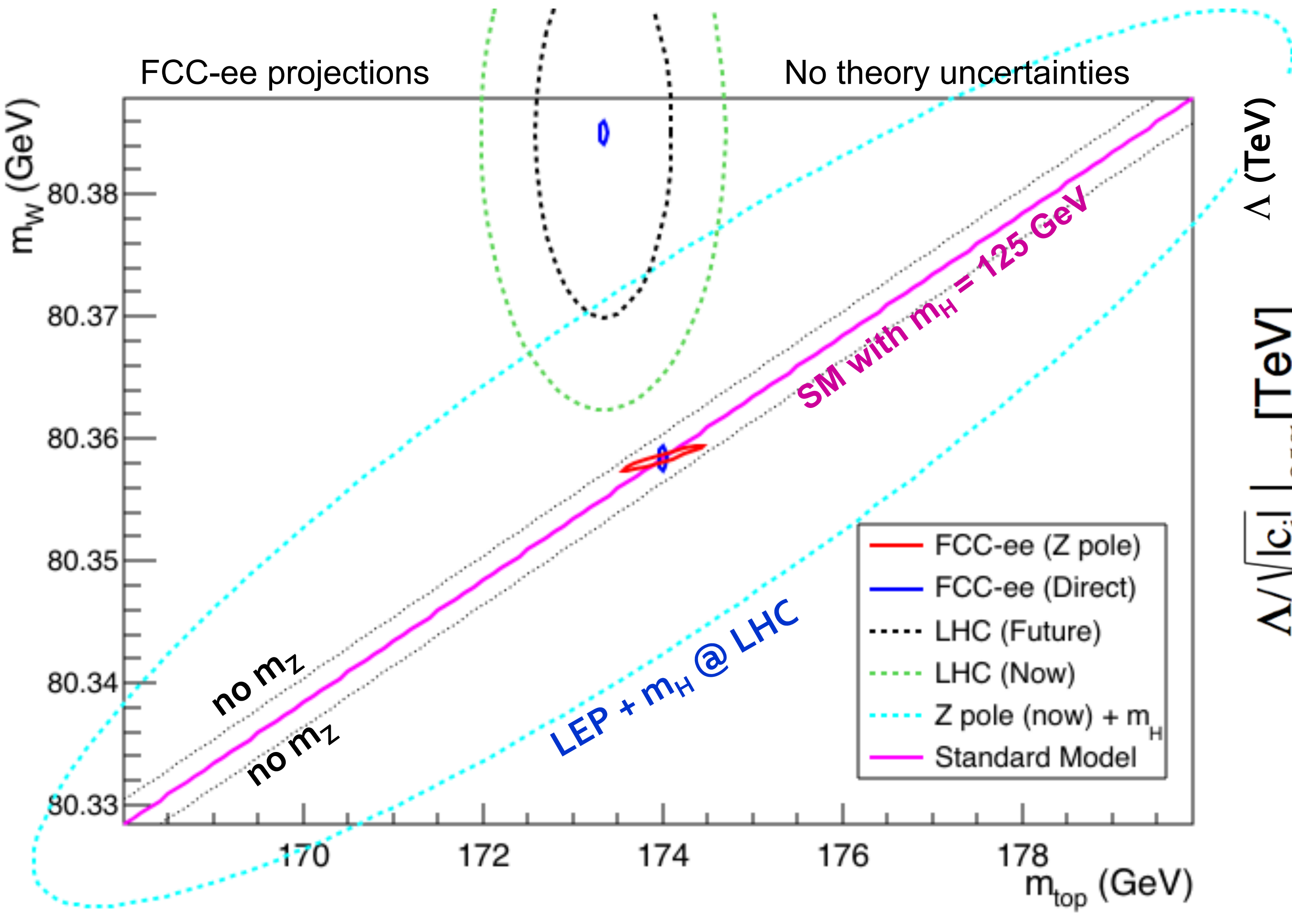
RY

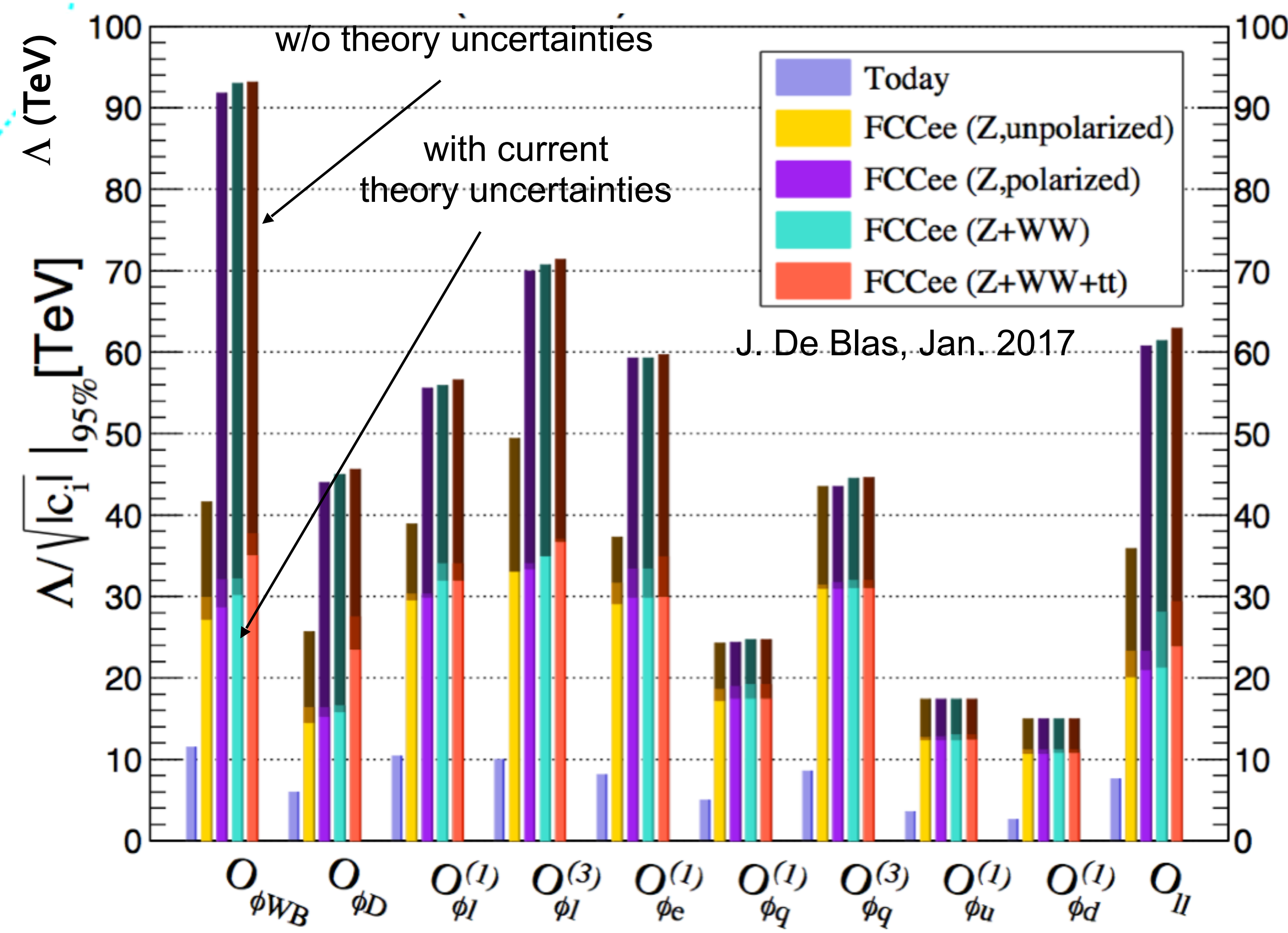
$$\frac{c}{\Lambda^2} \sim \frac{g_{NP}^2}{M_{NP}^2} < 0.01\ \mathrm{TeV}^{-2} \longrightarrow M_{NP} > 10\, g_{NP}\ \mathrm{TeV}$$

( **Weakly coupled NP** $M_{NP} > 10\ \mathrm{TeV}\quad (g_{NP} \sim 1)$ )

- Combining all FCC-ee EW measurements
  - In the context of the SM … and beyond

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i$$




**Requires 10-fold improved theory calculations**

Points to the physics to be looked for at FCC-hh

Today: $\Lambda/\sqrt{c}$ **> 5-10 TeV**

- New physics: blue and red ellipses may not overlap
  - Or even better, data may not fit to the SM

**After FCC-ee: $\Lambda/\sqrt{c}$ > 50-100 TeV ?**