Probability and Statistics

additional comments

random order

The Likelihood Function

Consider the Poisson distribution - describes a discrete event count n for a real-valued mean μ.

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

The **likelihood** of μ given n is the same equation evaluated as a function of μ

- Now it's a continuous function
- But it is not a pdf!

$$L(\mu) = Pois(n|\mu)$$

Common to plot the -ln *L* (or -2 ln *L*)

- helps avoid thinking of it as a PDF
- connection to χ^2 distribution

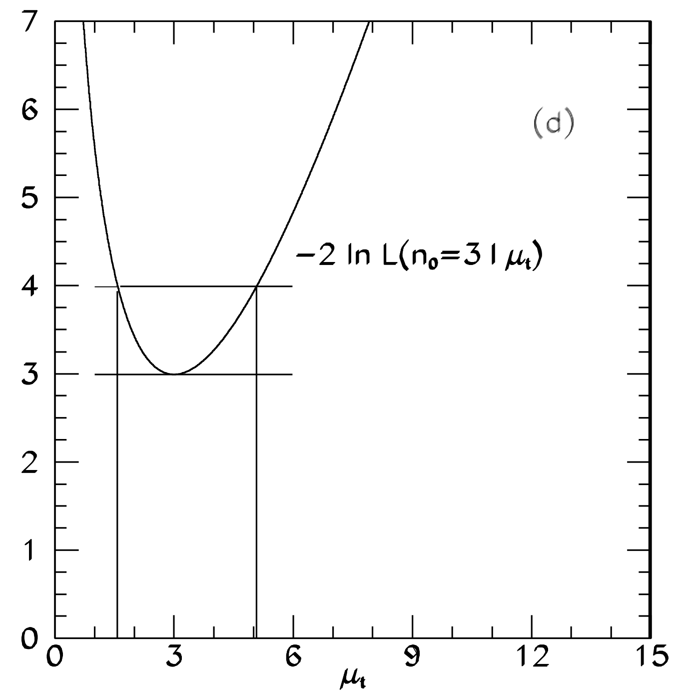


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

Repeated observations

In particle physics we are usually able to perform repeated observations of x that are independent and identically distributed

- These repeated observations are written $\{x_i\}$
- and the likelihood in that case is

$$L(\alpha) = \prod_i f(x_i|\alpha)$$

- and the log-likelihood is

$$\log L(\alpha) = \sum_i \log f(x_i|\alpha)$$

Estimators

Given some model $f(x|\alpha)$ and a set of observations $\{x_i\}$ often one wants to estimate the true value of α (assuming the model is true).

An estimator is function of the data written $\hat{\alpha}(x_1, \ldots x_n)$

- Since the data are random, so is the resulting estimate
- often it is just written $\hat{\alpha}$ where the x-dependence is implicit
- one can compute expectation of the estimator

$$E[\hat{\alpha}(x)|\alpha] = \int \hat{\alpha}(x) f(x|\alpha) dx$$

Properties of estimators.

- **bias** $E[\hat{\alpha}(x)|\alpha] - \alpha$ (unbiased means bias=0)
- **variance** $E[(\hat{\alpha}(x) - \alpha)^2|\alpha] = \int (\hat{\alpha}(x) - \alpha)^2 f(x|\alpha) dx$
- **asymptotic bias** limit of bias with infinite observations

Maximum likelihood estimators

There are many different possible estimators, but the most well-known and well-studied is the maximum likelihood estimator (**MLE**)

$$\hat{\alpha}(x) = \text{argmax}_\alpha L(\alpha) = \text{argmax}_\alpha f(x|\alpha)$$

This is just the value of α that maximizes the likelihood

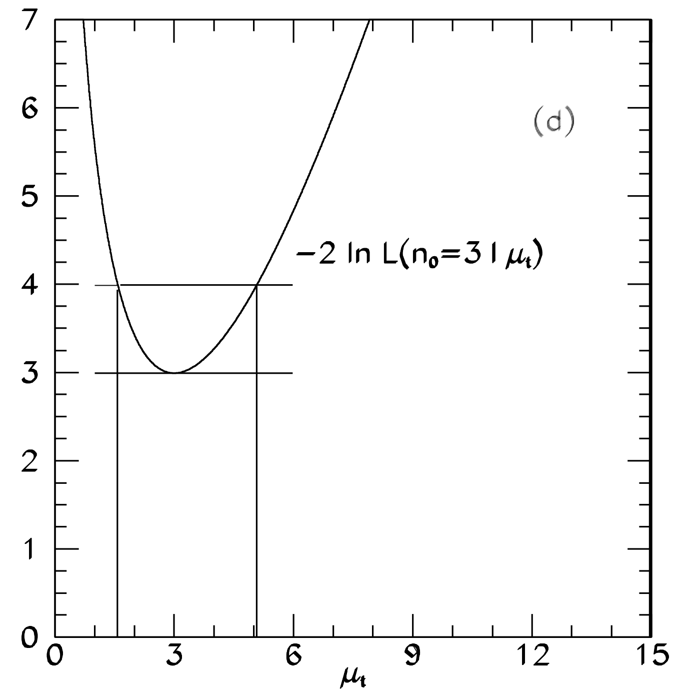


Figure from R. Cousins,
Am. J. Phys. 63 398 (1995)

Example: the Poisson distribution

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

Maximizing $L(\mu)$ is the same as minimizing -ln $L(\mu)$

$$-\frac{d}{d\mu} \ln L(\mu)\Big|_{\hat{\mu}} = 0 = \frac{d}{d\mu}\left(\mu - n\ln\mu + \underbrace{\ln n!}_{\text{const}}\right) = 1 - \frac{n}{\mu}$$

$$\Rightarrow \hat{\mu} = n$$

In this case, the MLE is unbiased b/c E[n]=μ

A second example

Consider a set of observations $\{x_i\}$ and we want to estimate the mean of a Gaussian with known σ

$$G(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

which gives

$$-\frac{d}{d\mu}\ln L(\mu)\Big|_{\hat{\mu}} = 0 = \frac{d}{d\mu}\left(\sum_i \frac{(x_i-\mu)^2}{2\sigma^2} + \underbrace{\ln\sqrt{2\pi}\sigma}_{\text{const}}\right) = \sum_i \frac{(x_i-\mu)}{\sigma^2}$$

$$\Rightarrow \hat{\mu} = \frac{1}{N}\sum_i x_i$$

(an unbiased estimator) .

However, the MLE $\hat{\sigma}^2 = \frac{1}{N}\sum_i (x_i-\mu)^2$ is biased

It can be shown that $\hat{\sigma}^2 = \frac{1}{N-1}\sum_i (x_i-\mu)^2$ is unbiased

Thus, the MLE is **asymptotially unbiased .**

Covariance & Correlation

Define covariance cov[x,y] (also use matrix notation V_{xy}) as

$$\text{cov}[x,y] = E[xy] - \mu_x \mu_y = E[(x-\mu_x)(y-\mu_y)]$$

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{\text{cov}[x,y]}{\sigma_x \sigma_y}$$

If x, y, independent, i.e., $f(x,y) = f_x(x) f_y(y)$, then

$$E[xy] = \int\int xy\, f(x,y)\, dx dy = \mu_x \mu_y$$

$\rightarrow$ $\text{cov}[x,y] = 0$ x and y, 'uncorrelated'

N.B. converse not always true.

RS- Problem with limits of integration

[G. Cowan]

Change of variable x, change of parameter θ

- For pdf $p(x|\theta)$ and change of variable from x to $y(x)$: $p(y(x)|\theta) = p(x|\theta) / |dy/dx|$.

  Jacobian modifies probability *density*, guaranties that

  $P(\, y(x_1) < y < y(x_2)\,) = P(x_1 < x < x_2\,)$, i.e., that

  Probabilities are invariant under change of variable x.

  – Mode of probability *density* is *not* invariant (so, e.g., criterion of maximum probability density is ill-defined).

  – Likelihood *ratio* is invariant under change of variable x. (Jacobian in denominator cancels that in numerator).

- For likelihood $\mathcal{L}(\theta)$ and reparametrization from θ to $u(\theta)$:

  $\mathcal{L}(\theta) = \mathcal{L}(u(\theta))$ (!).

  – Likelihood $\mathcal{L}(\theta)$ is invariant under reparametrization of parameter θ (reinforcing fact that $\mathcal{L}$ is *not* a pdf in θ).

skip

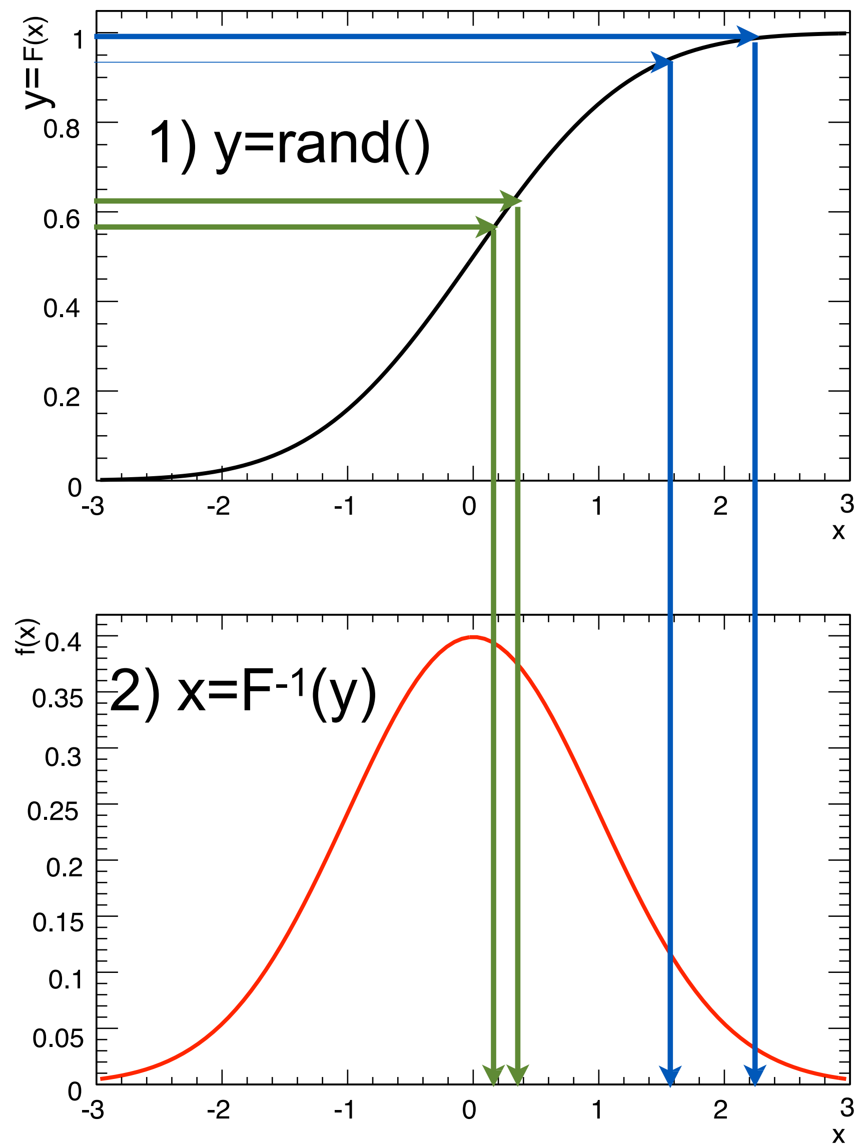
A more efficient Monte Carlo technique

No inefficiency

Requires inverse of cumulative $F^{-1}(y)$

Recall

$$f(x) = \frac{\partial F(x)}{\partial x}$$



skip

Different definitions of Probability

Frequentist

- defined as limit of long term frequency
- probability of rolling a 3 := limit of (# rolls with 3 / # trials)
  - you don't need an infinite sample for definition to be useful
  - sometimes ensemble doesn't exist
    - eg. P(Higgs mass = 120 GeV), P(it will snow tomorrow)
- Intuitive if you are familiar with Monte Carlo methods
- compatible with orthodox interpretation of probability in Quantum Mechanics. Probability to measure spin projected on x-axis if spin of beam is polarized along +z

$$|\langle \leftarrow | \uparrow \rangle|^2 = \frac{1}{2}$$

Subjective Bayesian

- Probability is a degree of belief (personal, subjective)
  - can be made quantitative based on betting odds
  - most people's subjective probabilities are not ***coherent*** and do not obey laws of probability

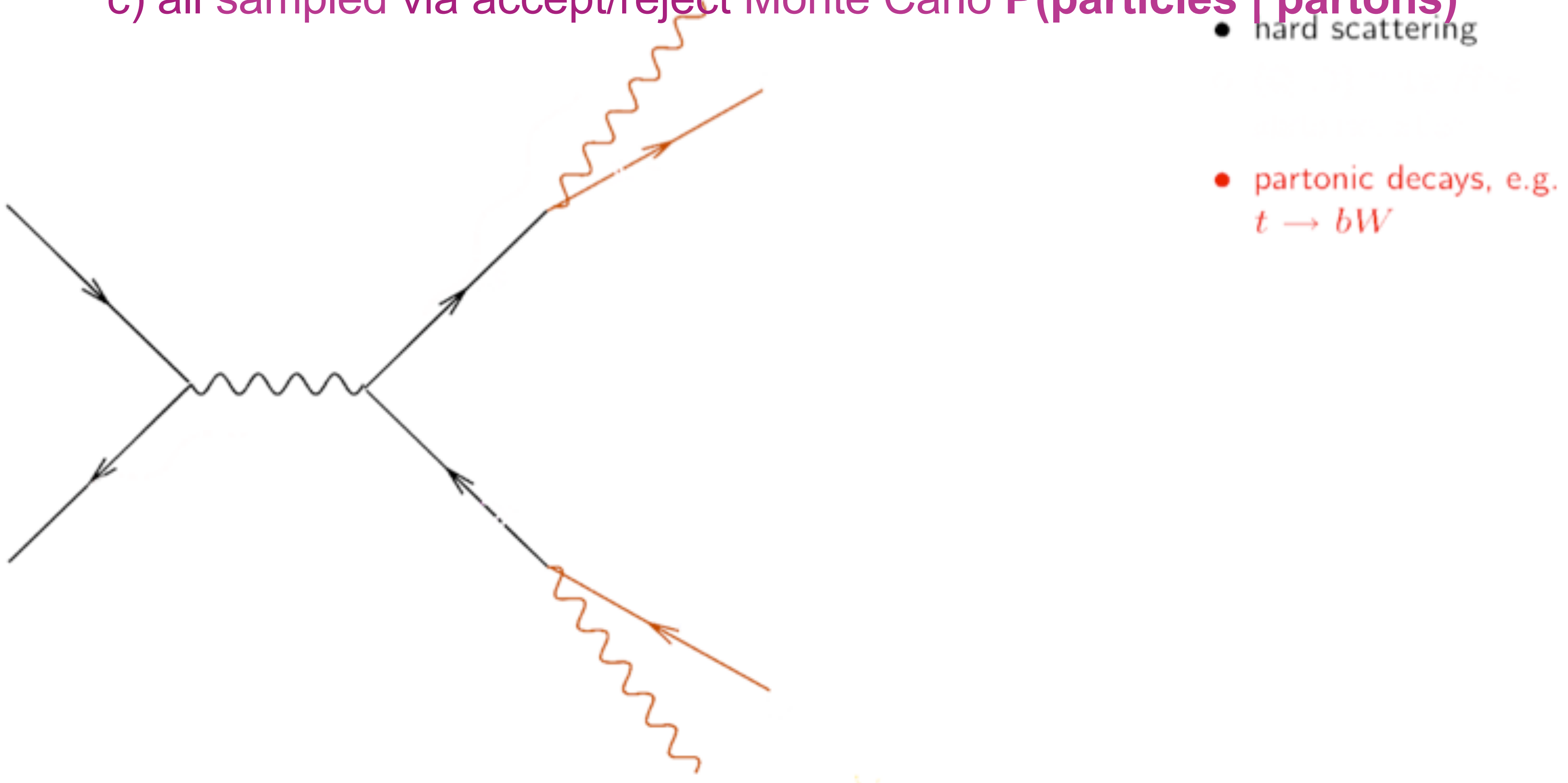
http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/#3.1

simulation

a) Perturbation theory used to systematically approximate the theory.

b) splitting functions, Sudokov form factors, and hadronization models

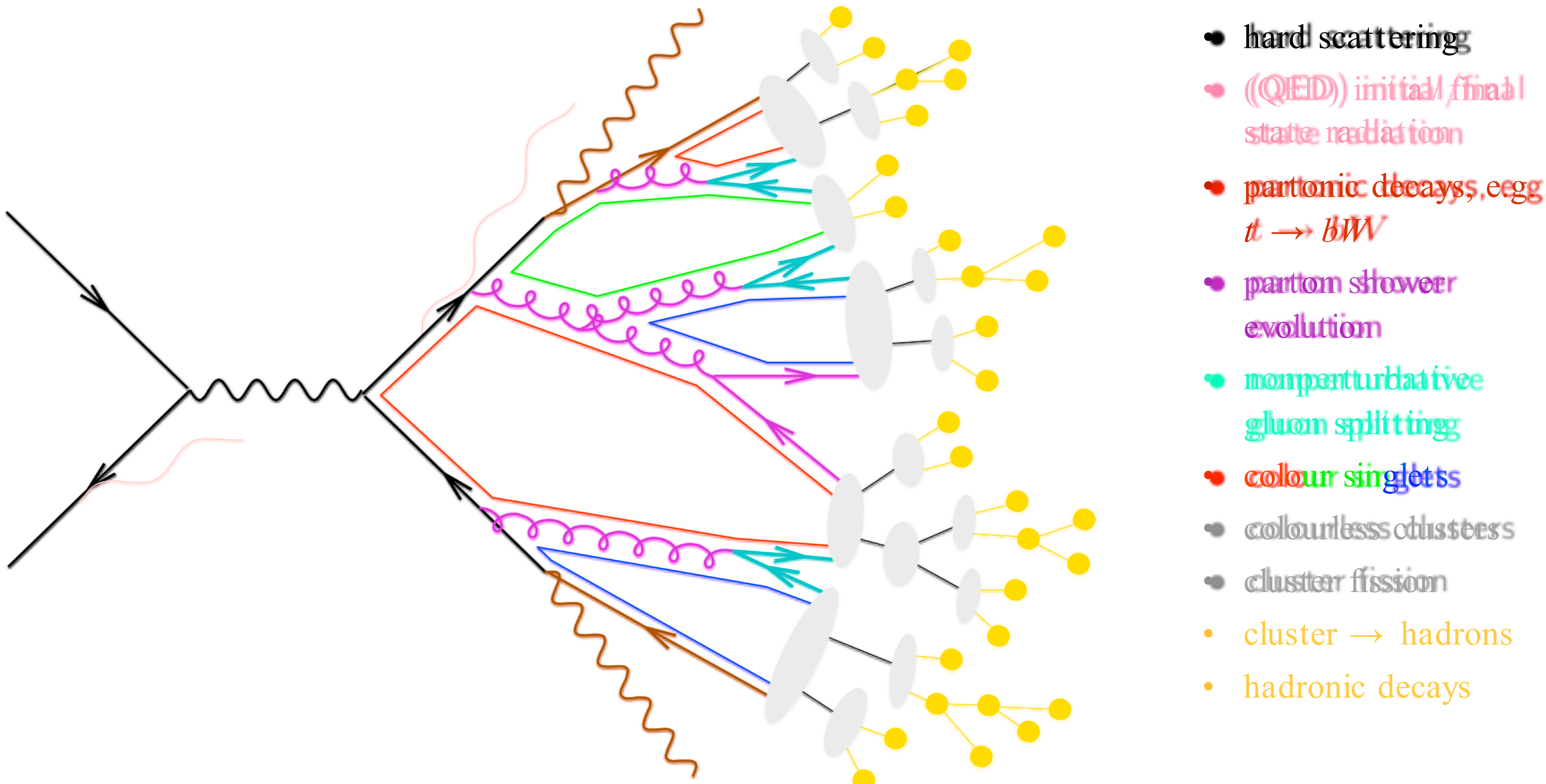
c) all sampled via accept/reject Monte Carlo **P(particles | partons)**



a) Perturbation theory used to systematically approximate the theory.

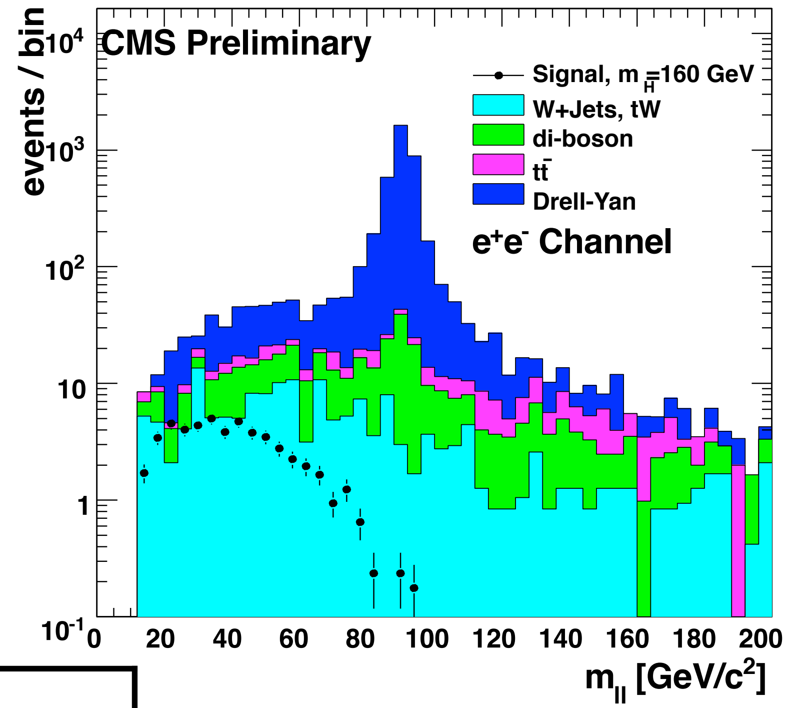
b) splitting functions, Sudokov form factors, and hadronization models

c) all sampled via accept/reject Monte Carlo **P(particles | partons)**



Regions in the data with negligible signal expected are used as control samples

- simulated events are used to estimate extrapolation coefficients
- extrapolation coefficients may have theoretical and experimental uncertainties



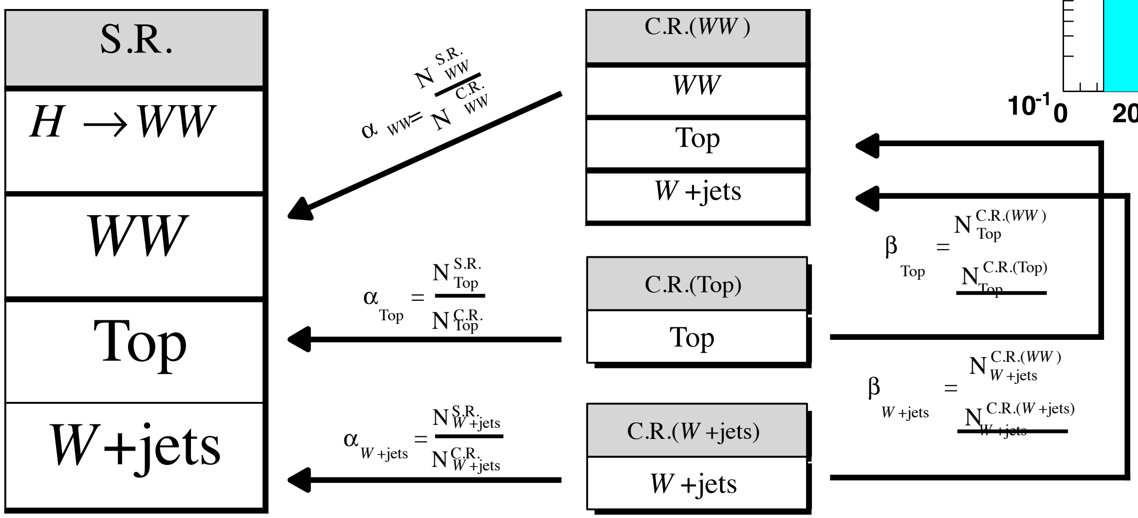
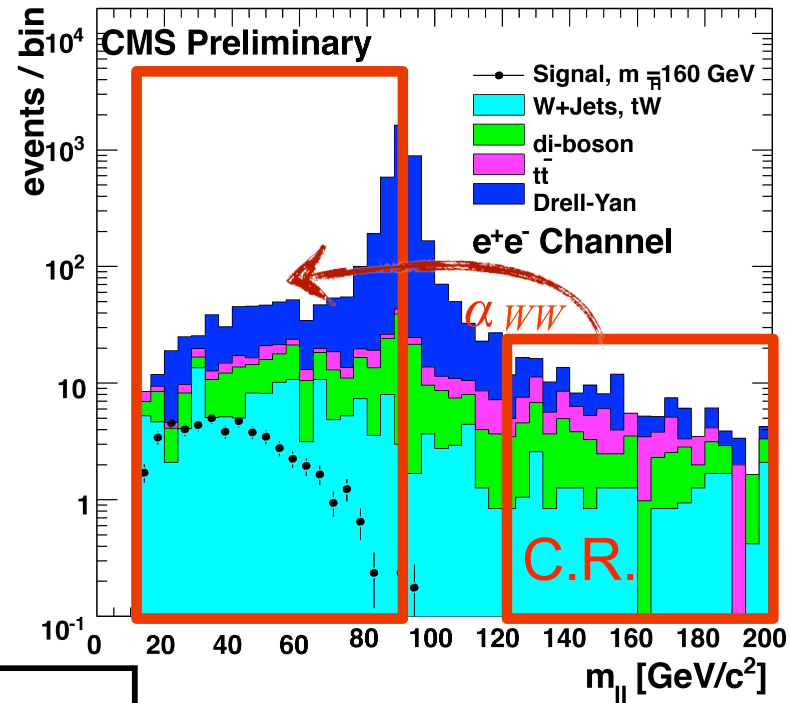


Figure 10: Flow chart describing the four data samples used in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. S.R and C.R. stand for signal and control regions, respectively.

Regions in the data with negligible signal expected are used as control samples

- simulated events are used to estimate extrapolation coefficients
- extrapolation coefficients may have theoretical and experimental uncertainties



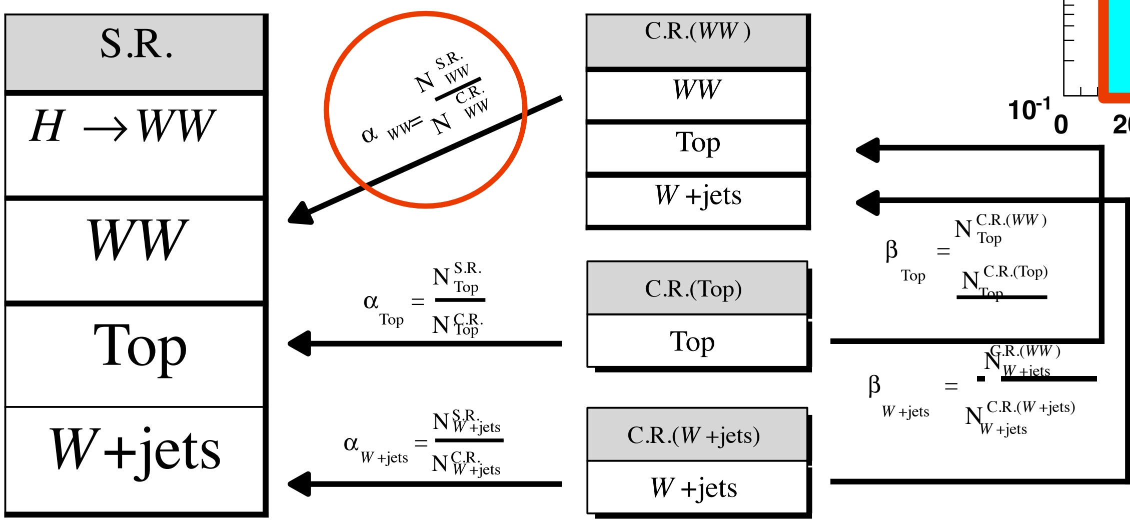


Figure 10: Flow chart describing the four data samples used in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. S.R and C.R. stand for signal and control regions, respectively.

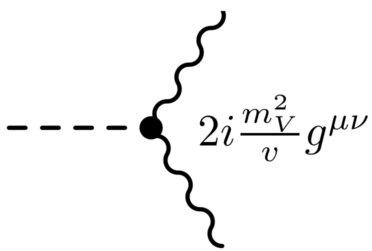
Application to Higgs discovery

Standard Model Lagrangian - Higgs sector

$$L_{SM} = D_\mu H^+ D_\mu + \mu^2 H^+ H - \frac{\lambda}{2}(H^+ H)^2 - (y_{ij} H \overline{\psi_i} \psi_j + h.c.)$$

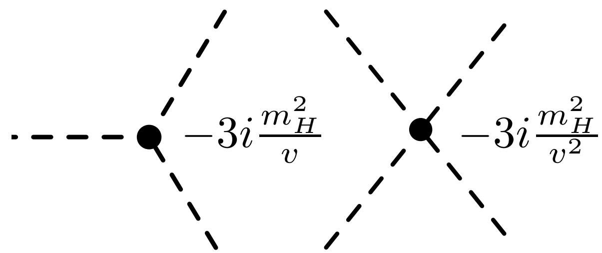
Couplings to EW gauge bosons

$$[m_W^2 W^{\mu+} W_\mu^- + \tfrac{1}{2} m_Z^2 Z^{\mu 0} Z_\mu^0] \cdot (1 + \tfrac{h}{v})^2$$



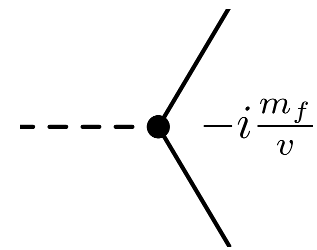
Higgs self-couplings

$$-\mu^2 h^2 - \tfrac{\lambda}{2} v h^3 - \tfrac{1}{8} \lambda h^4$$



Couplings to fermions

$$-\sum_f m_f \bar{f} f \left(1 + \frac{h}{v}\right)$$



$$m_H = \sqrt{2}\mu = \sqrt{\lambda} v$$

$v =$ vacuum expectation value

m_H – only parameter not fixed in SM

LHC Goals

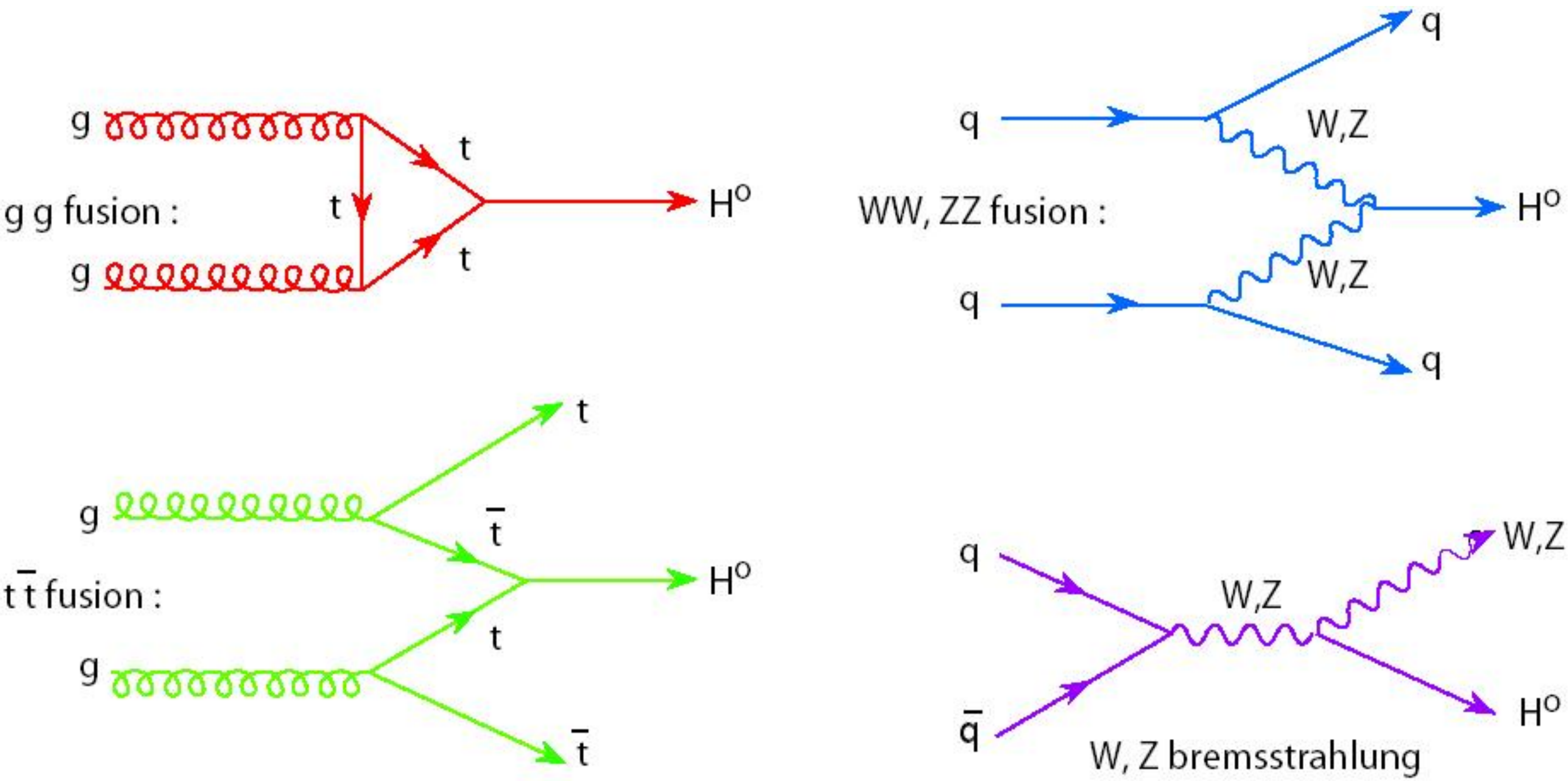
- verify Standard Model Lagrangian
- measure Higgs boson parameters
- search for physics beyond the Standard Model

Production

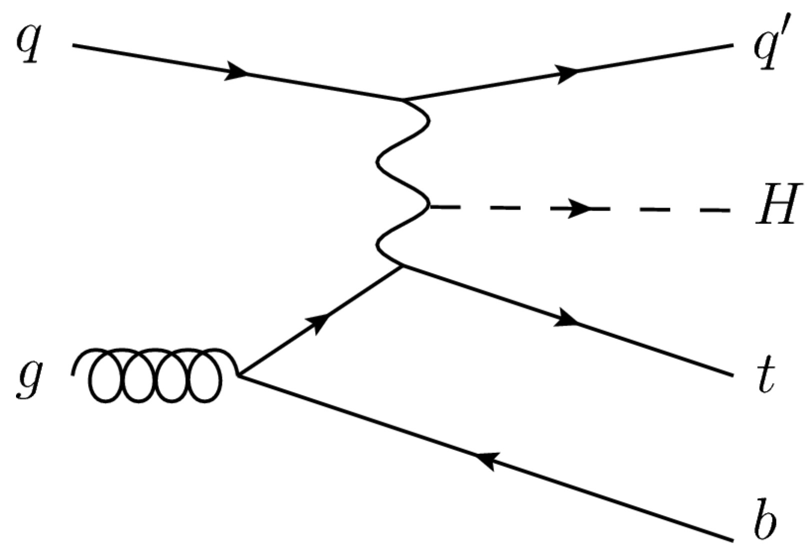
and

Decays

Production - dominant processes



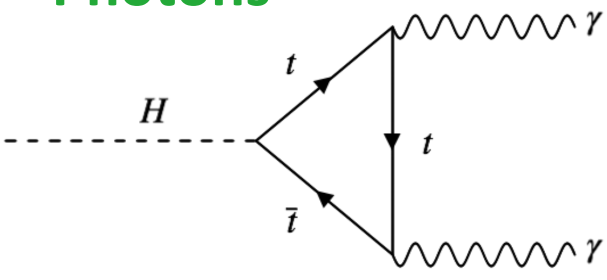
More complex diagram are possible with a penalty of multiple couplings

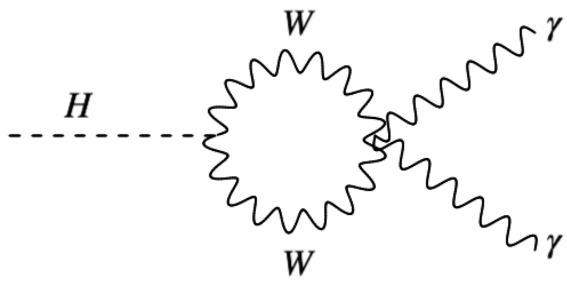


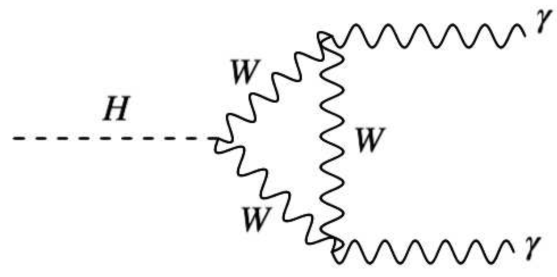
Important higher order corrections

Decays

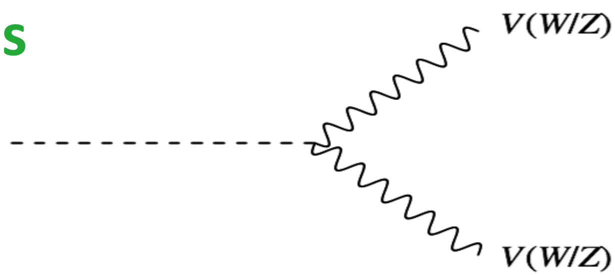
Photons



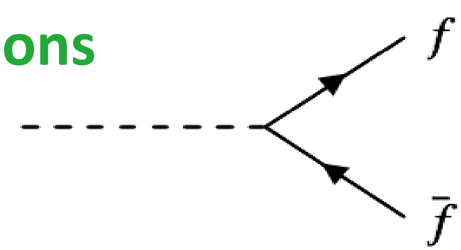




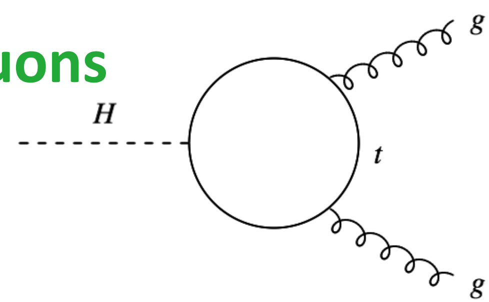
Vector bosons



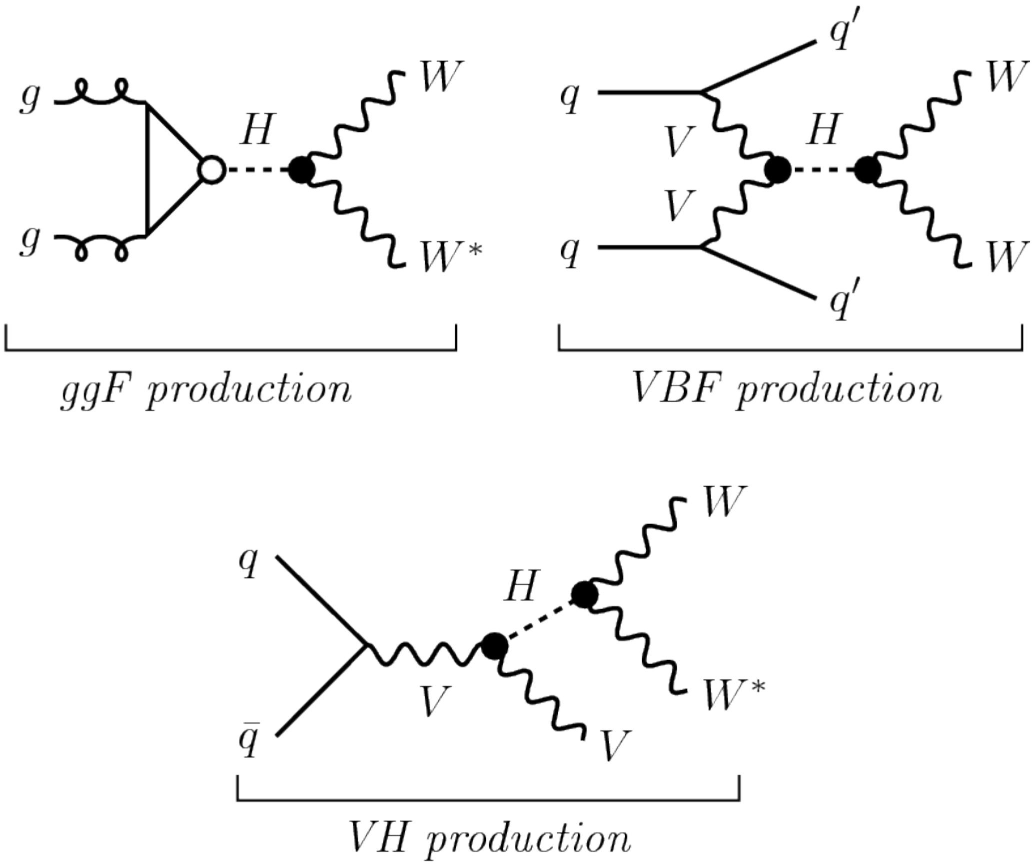
Fermions



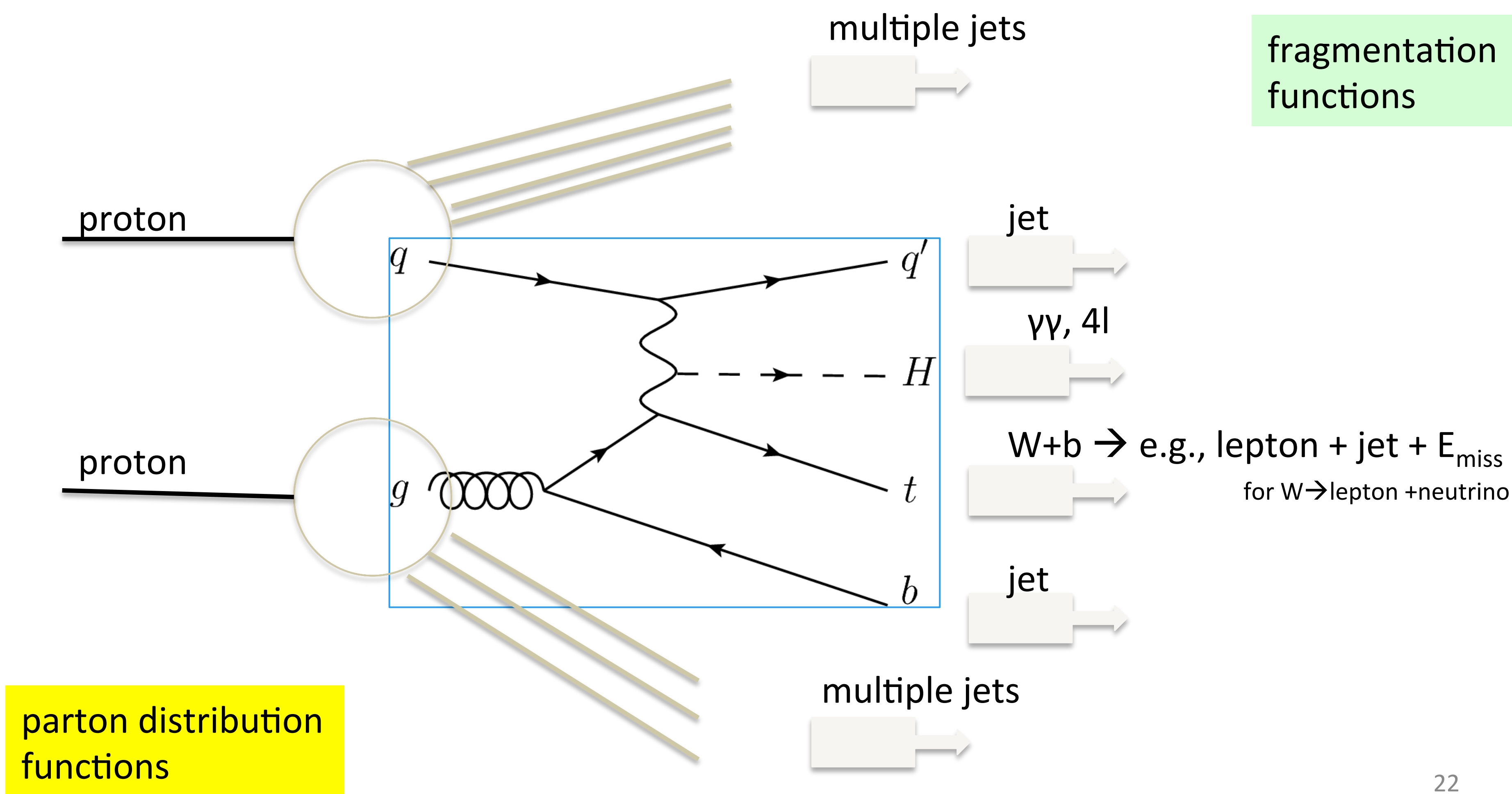
Gluons



Production + decay - theorist's view



Production + decays – experimenter's view



Event Classification

Higgs decay is independent of the production mechanism. However, different production mechanisms imply different kinematical distributions and therefore different acceptances and detection efficiencies.

→ For precision measurement it is important to separate Higgs production channels.

Difficult and possible only for a fraction of cases.

Topology of events (extra jets, additional leptons or missing energy) allows for partial separation of production mechanisms.

Statistical method

Extended likelihood function for (signal + background): *L*(α,ν)

$$-\ln L(\alpha,\nu)=(n_s+n_b)-\sum_e\Big[\underbrace{n_s\cdot f_s(x_e\,|\,\alpha,\nu_s)}_{\text{signal pdf}}+\underbrace{n_b\cdot f_b(x_e\,|\,\nu_b)}_{\text{background pdf}}\Big]-\underbrace{\sum_k \ln \pi_k(\nu_k)}_{\text{ancillary pdfs}}$$

n_s, n_b - signal / background yields
x_e - observables
f_s, f_b - signal / background pdfs
α - parameter of interest (mass, couplings, cross-section,…)
ν - "nuisance parameters" (shape parameters, systematics,….)
π_k - pdfs obtained from auxiliary measurements

Many variables + many signal + background processes → many terms

Likelihood fits

Confidence intervals (value ± error), limits and significances are based on the **Profile Likelihood Ratio**

$$\Lambda(\alpha) = \frac{L(\alpha, \hat{\hat{\nu}}(\alpha))}{L(\hat{\alpha}, \hat{\nu})}$$

likelihood for fixed α and profiled ν

maximum likelihood for free α, ν

$\hat{\hat{V}}$ is the conditional best fit for a particular value of α

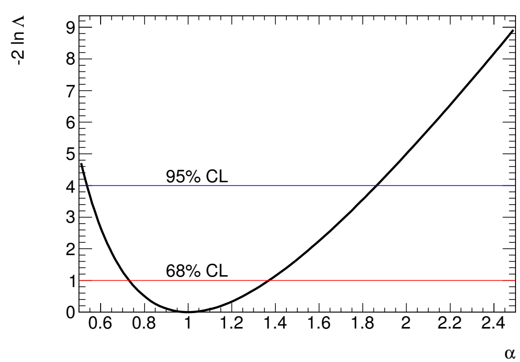
$\hat{\nu}$ is the best-fit α

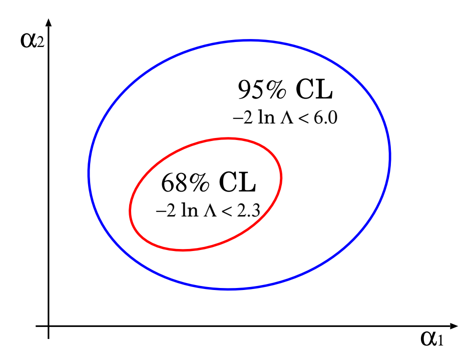
To **combine** – multiply the likelihood terms

Test statistics: $q_\alpha = -2 \ln \Lambda(\alpha)$

Wilks theorem: if $\alpha = \alpha_{true}$, then q_α follows a χ^2 distribution

compute **confidence interval**





Pseudo-experiments

Selected candidate events represent a small subsample of all produced signal events. They may all be in a tail of a distribution of a particular discriminant.

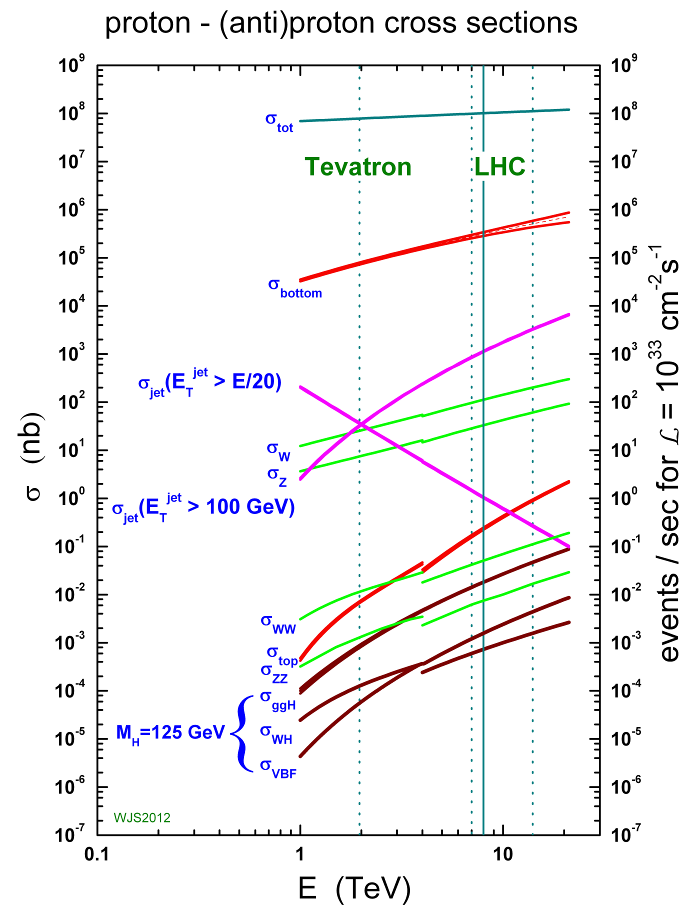
-> Need to estimate the probability of this selection, e.g., for VBF process how often there are two separated jets fulfilling selection criteria.

Estimated by generating large number (~10^4) of Monte Carlo data sets with the same number of events with full reconstruction and applying selection criteria.

Cross sections at LHC

Higgs cross section overwhelmed by QCD

process	cross section (pb) at √s = 8 TeV	events/s at L-10^{33} cm^{-2} s^{-1}
low-Q^2 QCD (minimum bias)	≈10^{11}	≈ 10^8
high-Q^2 QCD	≈10^9	≈ 10^6
W production	≈10^5	≈ 100
Z production	≈$5 \cdot 10^4$	≈ 50
ttbar production	≈240	≈ 0.24
SM Higgs	≈22	≈ 0.022



Need to apply several filters starting from the on-line trigger and then in off-line analysis

selection based on isolated leptons, photons, jets with high p_T and large missing energy

Higgs Boson Production Rates

Run 1 integrated luminosity: ~ 5 fb^{-1} at √7 TeV and ~20 fb^{-1} at √s = 8 TeV

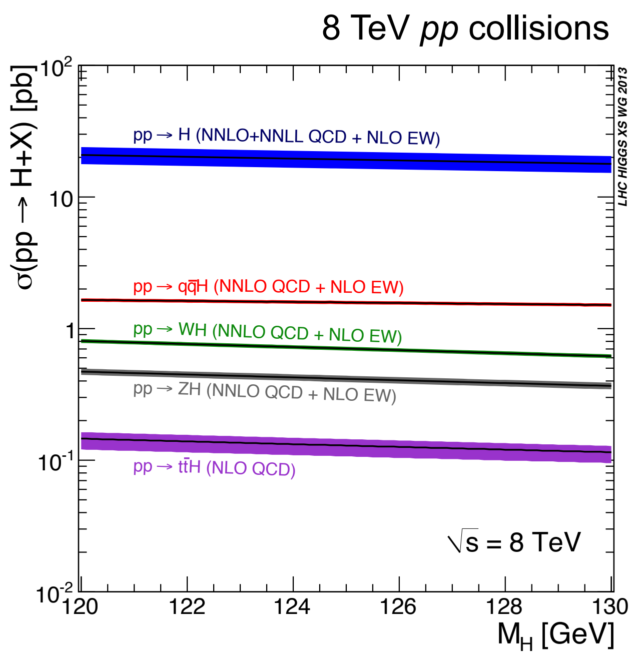
At √8 TeV: total pp cross section ~ 70 mb

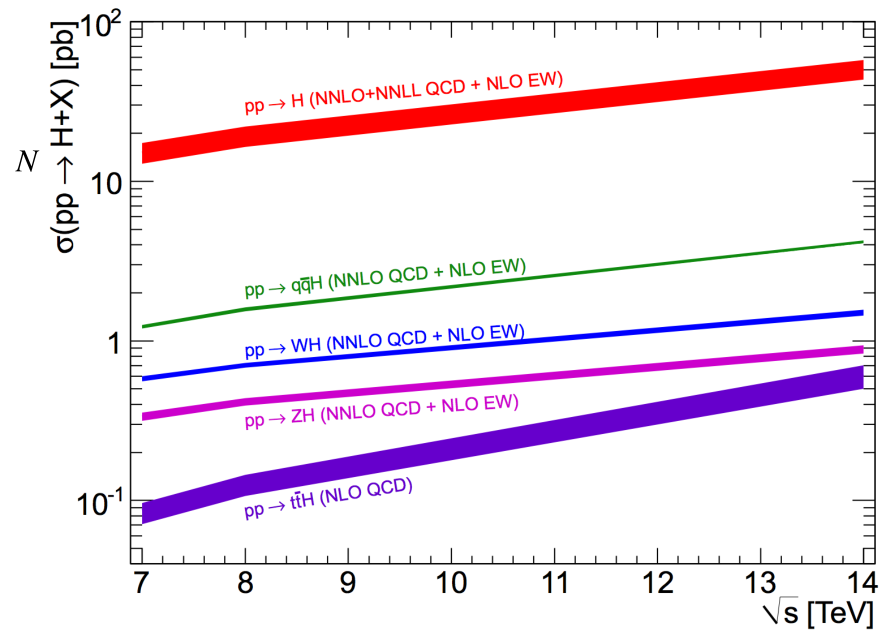
total Higgs production cross section ~ 22 pb

★ ~500,000 Higgs produced in Run 1

$$N_{ev} = \sigma \cdot \int L \cdot A \cdot E_{ff}$$

★ only 1 in 10^{10} events contains Higgs





Small dependence on Higgs mass

Factor 2-4 increase with energy for Run 2.
Large phase space increase for ttH.

Production cross sections and decay rates

Production process	Cross section (pb) $\sqrt{s} = 7$ TeV	Cross section (pb) $\sqrt{s} = 8$ TeV
ggF	15.0 ± 1.6	19.2 ± 2.0
VBF	1.22 ± 0.03	1.57 ± 0.04
WH	0.573 ± 0.016	0.698 ± 0.018
ZH	0.332 ± 0.013	0.412 ± 0.013
bbH	0.155 ± 0.021	0.202 ± 0.028
ttH	0.086 ± 0.009	0.128 ± 0.014
tH	0.012 ± 0.001	0.018 ± 0.001
Total	17.4 ± 1.6	22.3 ± 2.0

Decay channel	Branching ratio (%)
$H \to b\bar{b}$	57.1 ± 1.9
$H \to WW^*$	22.0 ± 0.9
$H \to gg$	8.53 ± 0.85
$H \to \tau\tau$	6.26 ± 0.35
$H \to c\bar{c}$	2.88 ± 0.35
$H \to ZZ^*$	2.73 ± 0.11
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to Z\gamma$	0.157 ± 0.014
$H \to \mu\mu$	0.022 ± 0.001

Handbook of LHC Higgs cross sections http://arxiv.org/pdf/1307.1347

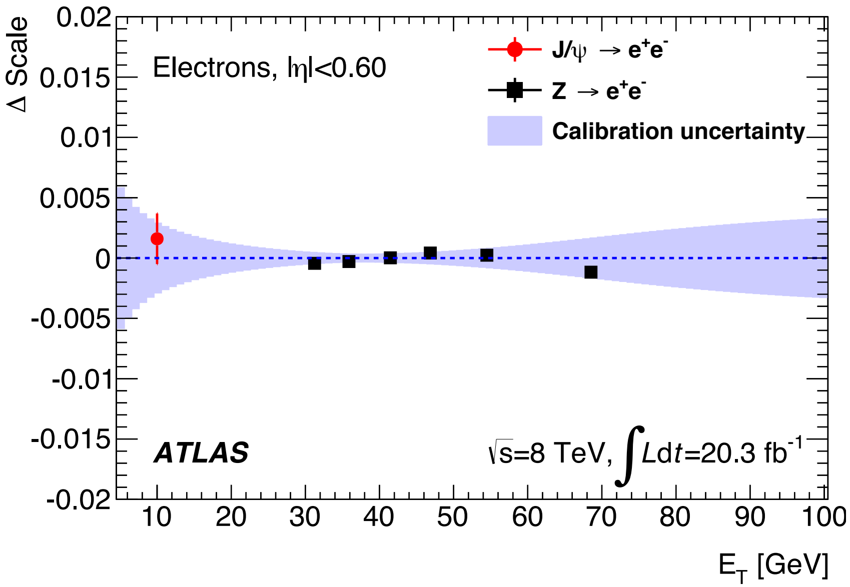
Each Higgs decay channel suffers (after filters) from QCD backgrounds with rates that are typically 10^5 – 10^6 higher than rates expected for the signal.

Higgs Mass
and
Cross Section

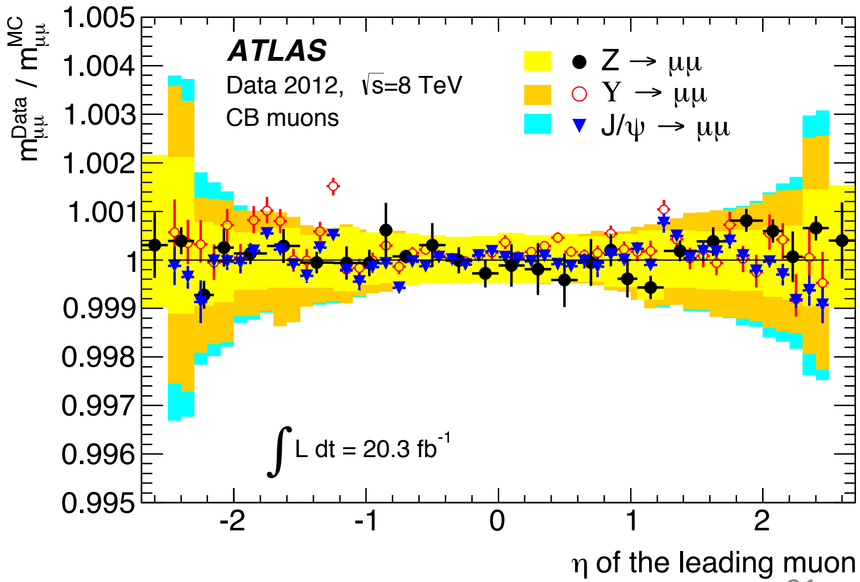
Higgs Mass and Production Rates Experimental Details

- Mass - Most precisely determined with H -> γγ and H –> 4 leptons channels
- Precise measurements of low p_T leptons down to 5-7 GeV are important
- Detector calibrations: ECAL (e/γ) and muon systems extremely important.
  ->ATLAS calibration reached precision below few per mille
- Energy scale from J/ψ, Υ, Z decays to e+e- and μ+μ-

Electron calibration



Muon calibration



Phys. Rev. D. 90, 052004 (2014)

Event selection

H -> γγ **-** large signal, clean but with large irreducible background

- several categories of photons:
  - unconverted
  - converted to e+e- with two tracks reconstructed
  - converted with one track reconstructed
- several classes of production mechanisms

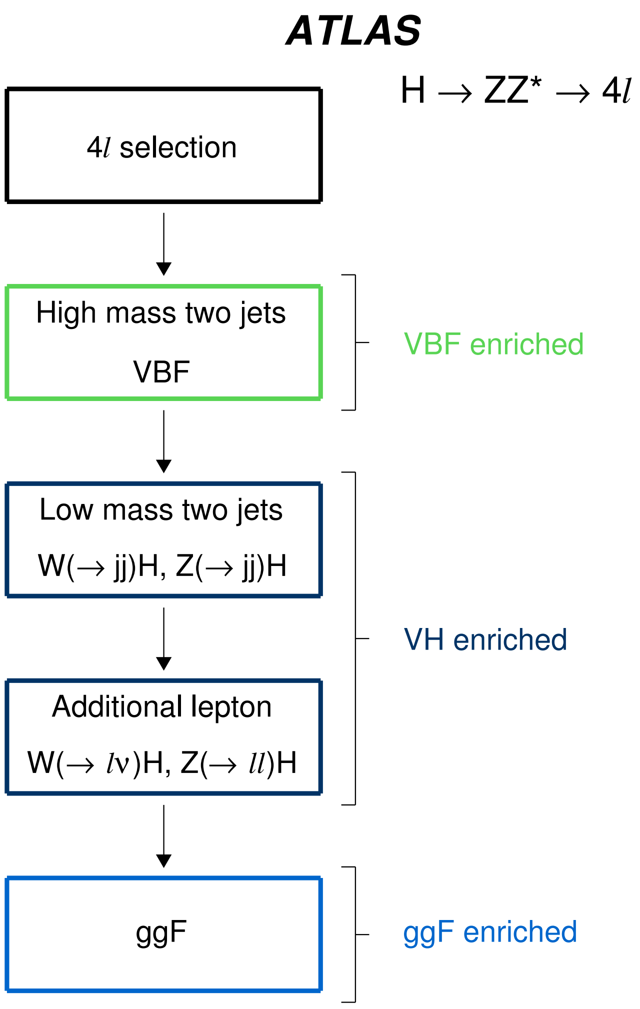
H -> ZZ*-> 4 leptons **-** small statistics but large signal/background ratio

- four separate final state channels:
  - ZZ* -> 4 electrons
  - ZZ* -> 4 muons
  - Z -> 2 electrons, Z* -> 2 muons
  - Z -> 2 muons, Z* -> 2 electrons
- Several classes of production mechanisms

For each category and decay channel there are different efficiencies, backgrounds and different systematic errors

Higgs -> 4 lepton event selection

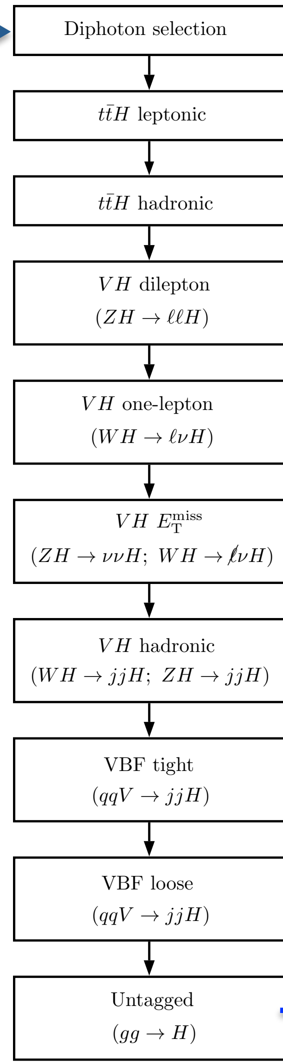
- Separate out most likely candidates for low-rate processes.
- Put everything else into dominant category.
- Introduce selection uncertainty into error estimate



Diphoton event selection and classification

photon selection

Unconverted photons
Two-track photon conversions
“One-track” photon conversions



event selection

detector region selection

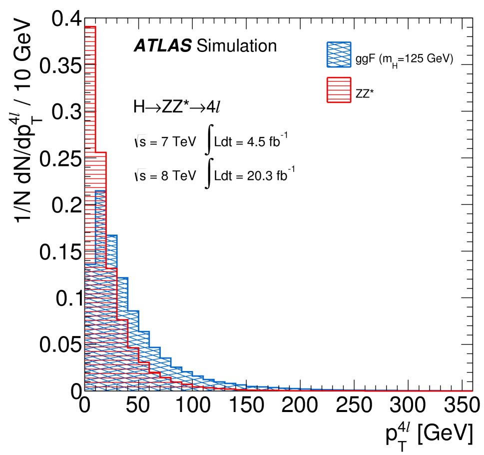
Central/forward
Low pt/high pt

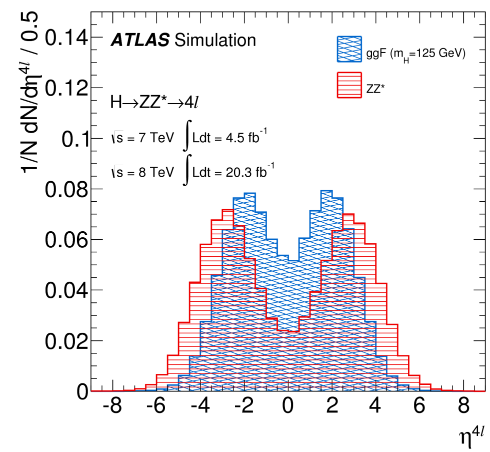
Sum individual contributions from 14 different categories with weights corresponding to selection efficiencies (see later ~300 fit parameters)

BDT -Multivariate discriminant construction

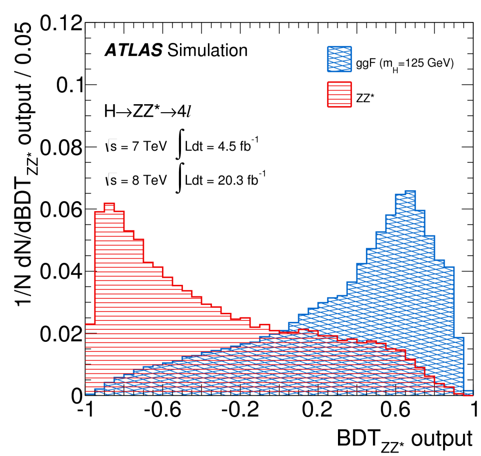
Need to separate H->ZZ* signal from ZZ* background and separate ggF production from VBF production mode. Use MC simulations using matrix element calculations (MadGraph5).

signal vs background distributions





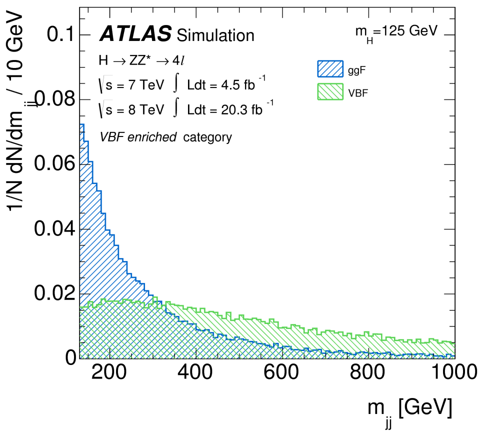
signal vs background separation

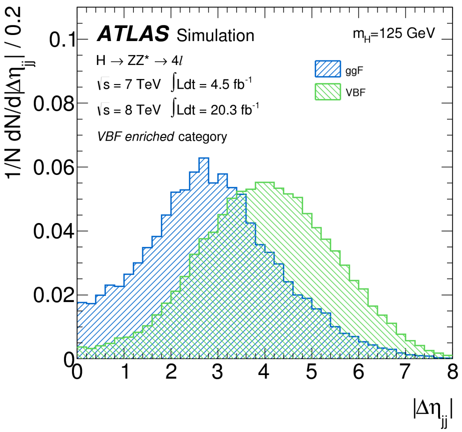


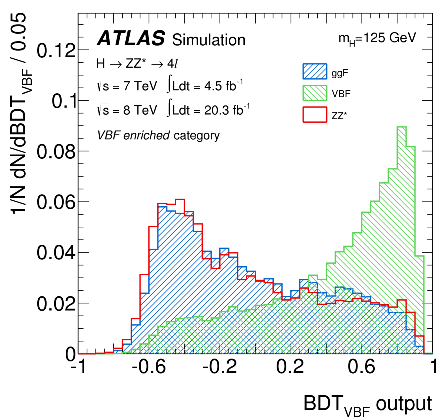
Discriminant of ggF vs VBF

VBF - 5 additional variables for extra jets:
invariant mass of two jets, Δη separation of jets, p_T of each jet, η of leading jet

simulated distributions for ggF vs VBF



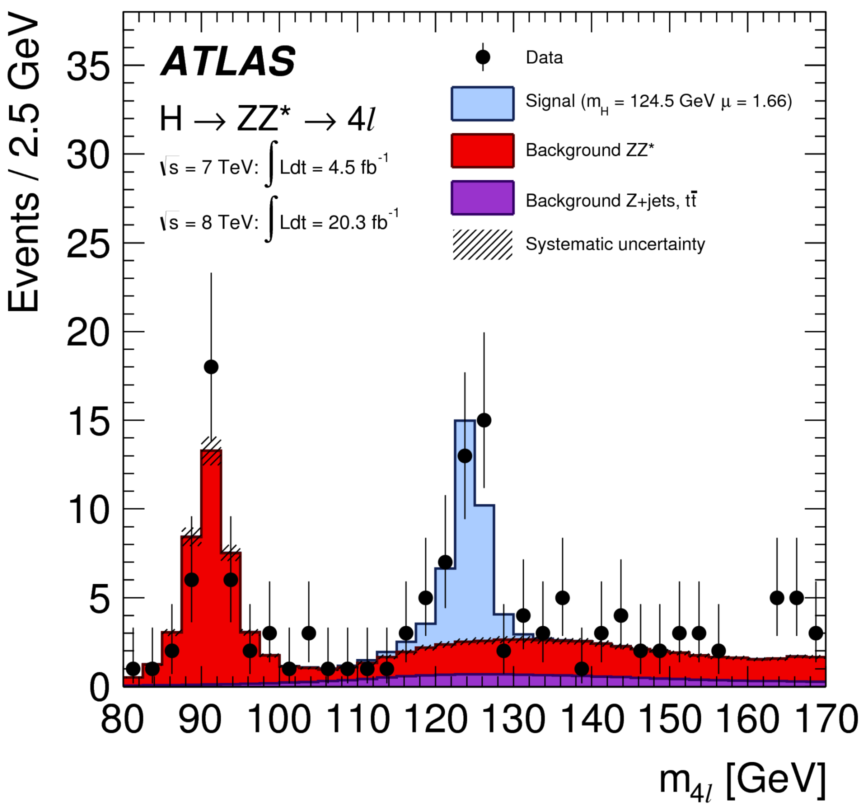


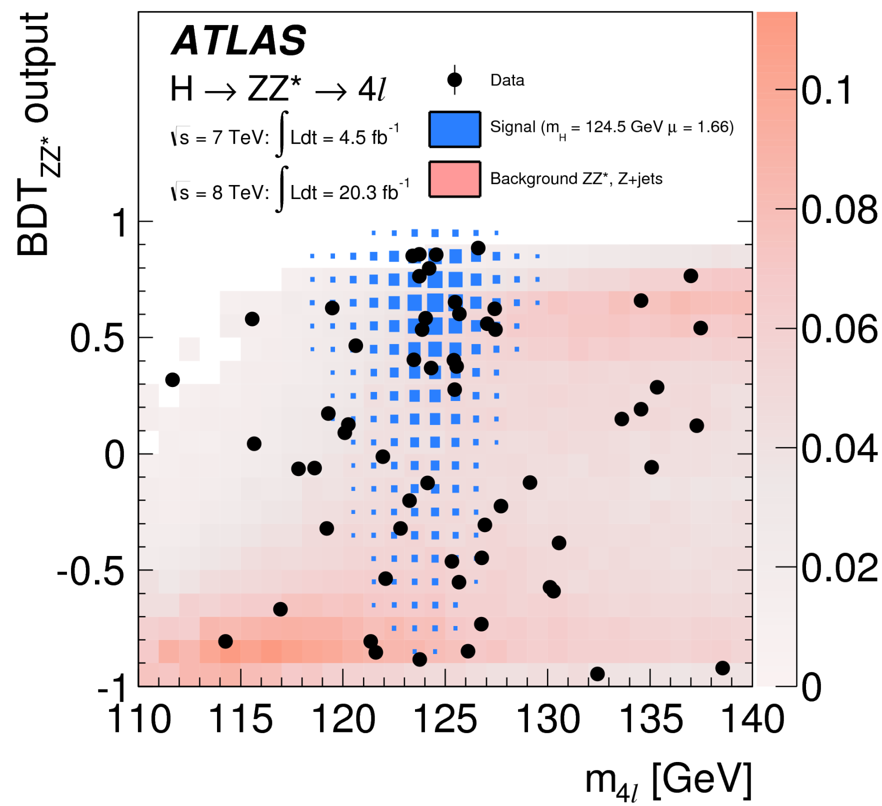


final discriminant

H→ ZZ* → 4l

Boosted Decision Tree (BDT) 2D analysis trained on simulated signal and ZZ* background events





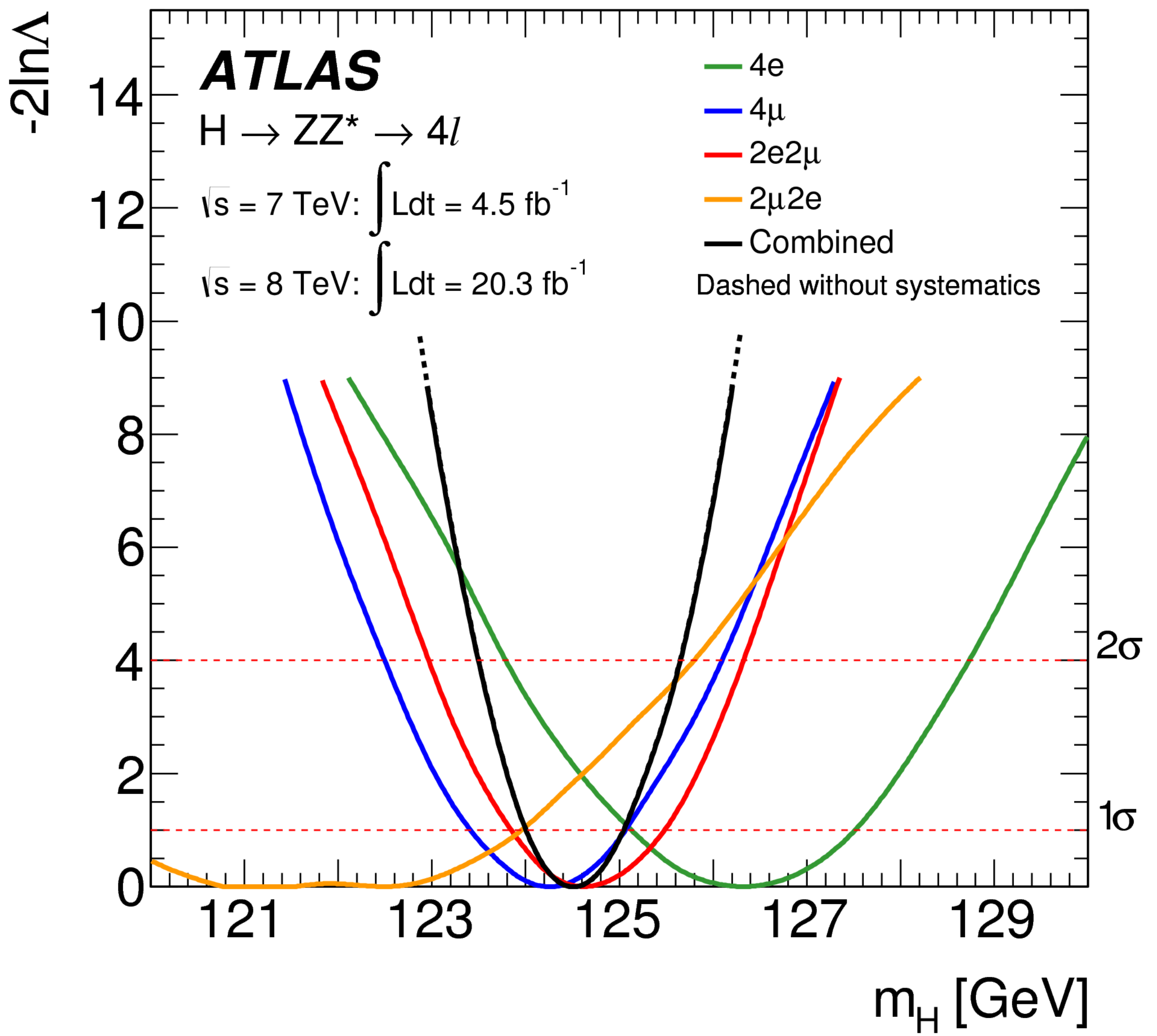
ATLAS $m_H = 124.51 \pm 0.52\ (\pm 0.52\ (stat) \pm 0.04\ (syst))$ **GeV**

CMS $m_H = 125.59 \pm 0.45\ (\pm 0.42\ (stat) \pm 0.17\ (syst))$ GeV

ATLAS: Phys. Rev. D. 90, 052004 (2014)
CMS: Phys. Rev. D. 89, 092007 (2014)

Detail check - Does mass depend on the 4l decay mode?

→No significant mass difference between different 4 lepton channels

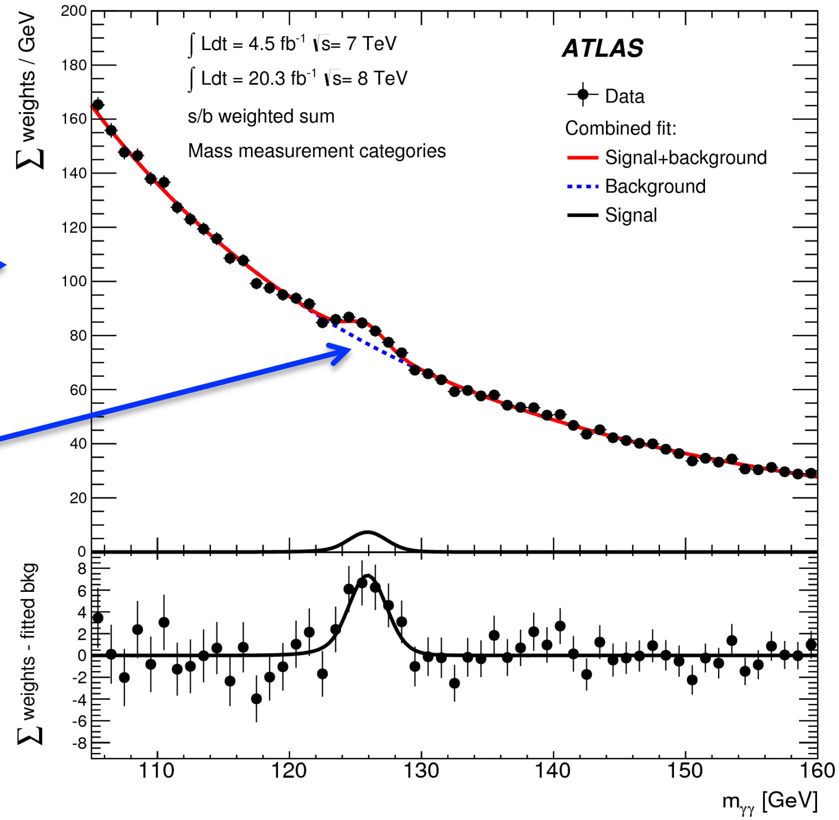


H→γγ

weights derived independently for each category

~300 events

observed weighted signal - background

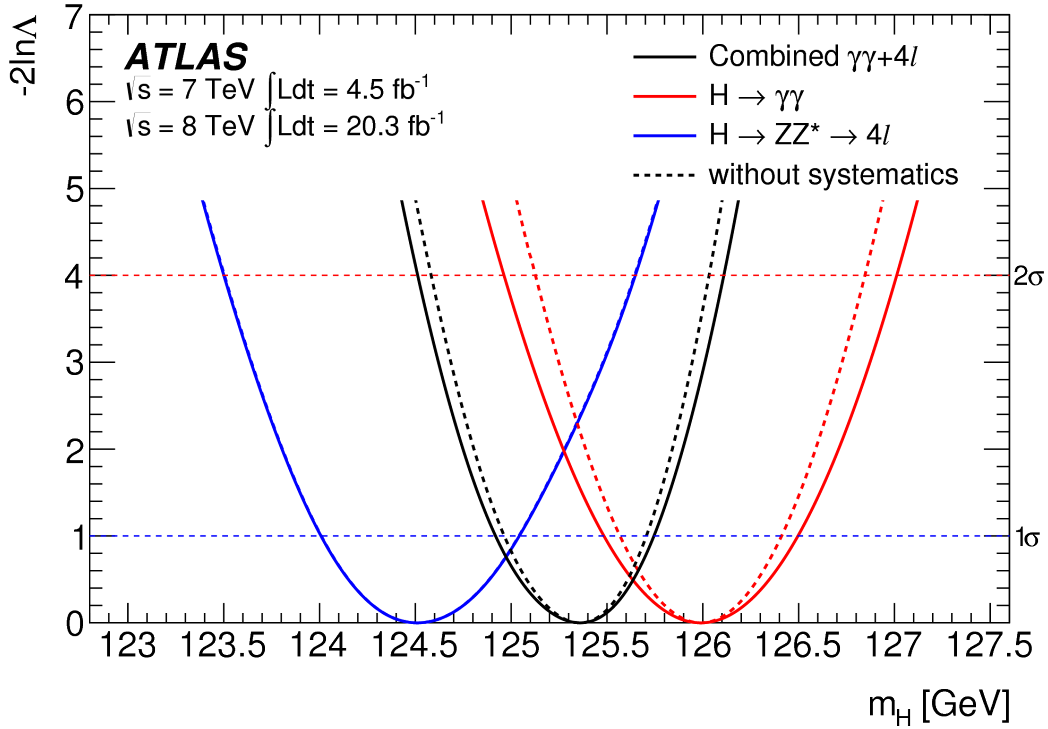


ATLAS **m_H = 126.02 ± 0.51 (± 0.43 (stat) ± (0.27 (syst)) GeV**

CMS m_H = 124.70 ± 0.45 (± 0.31 (stat) ± (0.15 (syst)) GeV

ATLAS: Phys. Rev. D. 90, 052004 (2014)
CMS: Eur. Phys. J. C. 74, 3076 (2014)

$H \rightarrow ZZ^* + H \rightarrow \gamma\gamma$ combination



No significant mass difference between $H \rightarrow \gamma\gamma$ and 4 lepton channels

ATLAS: $\Delta m_H(\gamma\gamma\text{-}4l) = +1.47 \pm 0.67$ (stat.) ± 0.28 (syst.) GeV (1.98σ)

CMS: $\Delta m_H(\gamma\gamma\text{-}4l) = -0.89 \pm 0.57$ GeV (1.6σ)

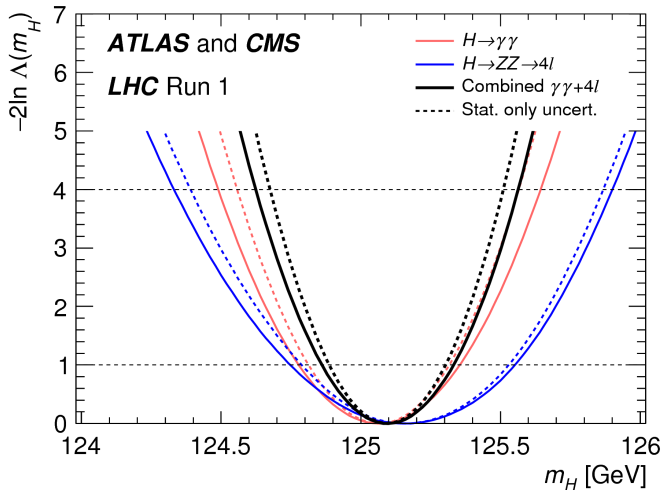
ATLAS $m_H = 125.36 \pm 0.37$ (stat) ± 0.18 (syst) GeV

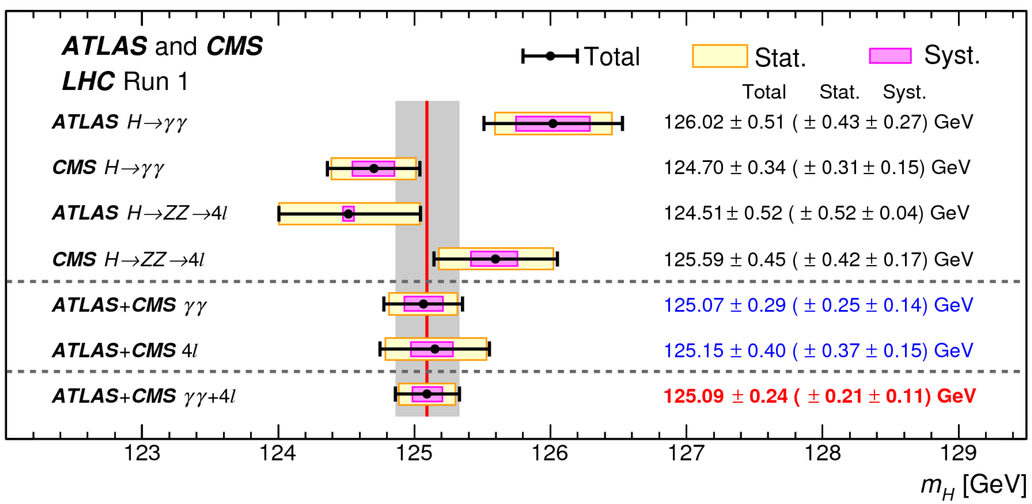
CMS $m_H = 125.03^{+0.26}_{-027}$ (stat) $^{+0.13}_{-0.15}$ (syst) GeV

New: ATLAS/CMS combination

- Maximum of the profile-likelihood fits using signal probability density functions derived from modeling and background probability distributions derived from the data
- Includes interference between signal and backgrounds (EW only)

arXiv:1503.07589



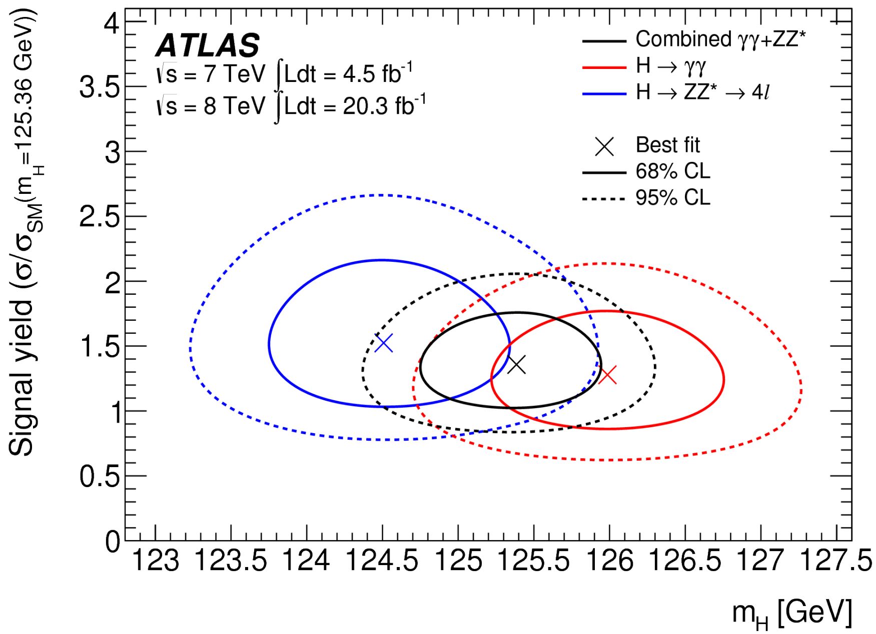


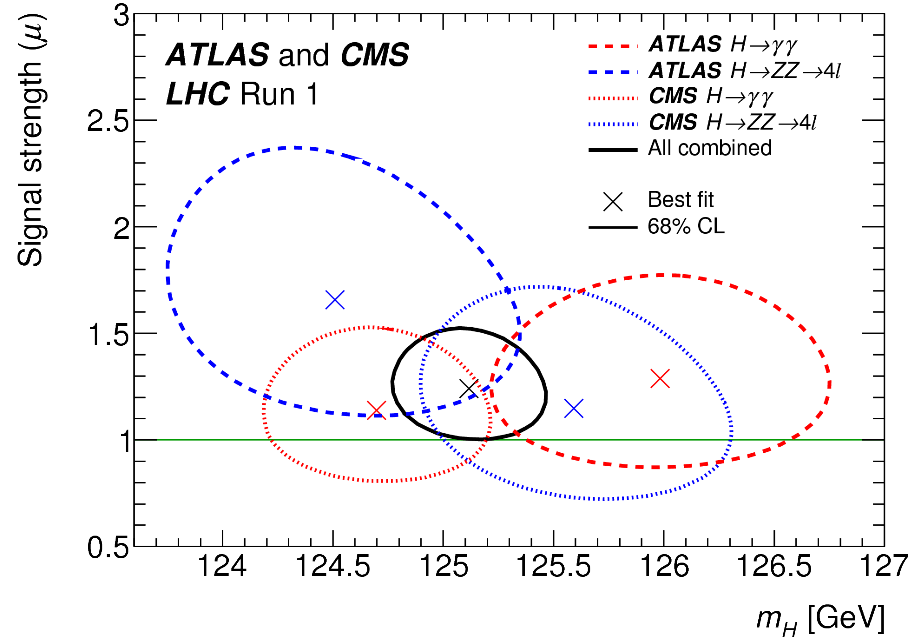
$$m_H = 125.09 \pm 0.24\ (\pm 0.21\ (\text{stat.}) \pm 0.11\ (\text{syst.}))\ \text{GeV}$$

★ **I offer a beer for best/craziest explanation why this value is so close to 5^3 in GeV**

Production rate

- Derived from the same 2D fit as the Higgs mass using 4l and γγ decays.
- **Caution:** For photon channel there are about 300 nuisance parameters with about 100 fitted parameters describing shapes and normalization of background models and about 200 parameters describing experimental and theoretical systematic uncertainties.





Results from different channels are consistent within 2σ and are consistent with signal strength expected from Standard Model.

Higgs Production Rates - other channels

Similar procedures as that for ZZ and γγ:

- Signal selection using leptons, b-jets, missing energy and tau hadronic decays
- Background minimization using kinematic properties
- Comparison with signal expected from various Monte Carlos
- Identification of systematic and theoretical uncertainties

Channels studied (tags)

H→W W→lν lν (l = e or μ, missing energy carried by neutrinos)
H→τ τ (τ hadronic and leptonic decays: lepton-lepton, lepton-jet, jet-jet topology)
H→b b (b jets tagged by 70% likelihood of identifying separated vertices)
H→Z γ (reconstructed Z ->ee and Z->μμ)
VBF (Higgs reconstruction applied in all decay channels + 2 separated hadronic jets)
VH (Higgs reconstructed in all channels, W tagged by lepton + missing energy,
Z reconstructed from leptonic decays)
ttH (Higgs reconstructed in bb, γγ, WW->lνlν, additional leptons from top decays)

Evidence for VBF process

- Second largest expected rate, low theory uncertainty.
- Distinctive topology with two jets widely separated in η and suppressed QCD activity between them
- Hints consistent with SM expectations in several channels.
- Combined analysis based o profile likelihood ratio test statistics Probability densities used for in are derived from MC for the signal and MC and data for the backgrounds.

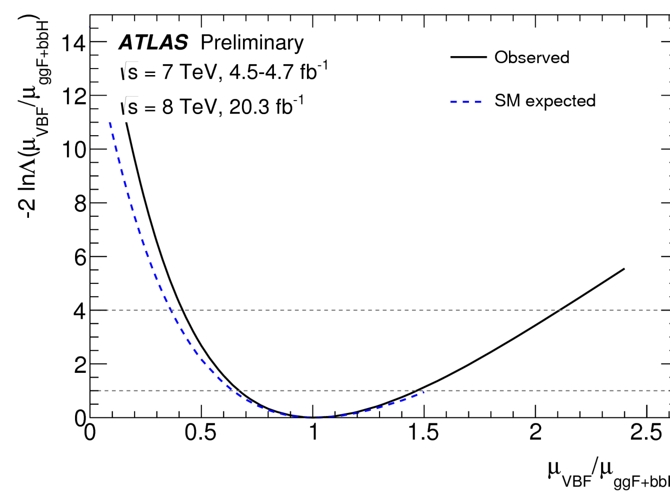
H->γγ 2 photons with $E_T/m_{\gamma\gamma}$ > 0.35 and 0.25 plus 2 jets

H->4l 2 pairs of same flavor, opposite charge leptons plus 2 jets with m_{jj} 130 GeV

H->WW* leptonic W decays - lvlv (same and opposite lepton charges) plus N_{jet}> 2

H->ττ leptonic and hadronic tau decays plus 2 jets separated by pseudorapidity

H->μμ opposite charged muon pair plus N_{jet}> 2



Likelihood contours

($\mu^f_{ggF+ttH}$, μ^f_{VBF+VH}) plane for Higgs mass m_H=125.36 GeV.

μ – ration of observed yield wrt SM expectation

Solid lines - 68% CL contours, dashed lines – 95% CL contours.

Standard Model expectation - star at (1,1).