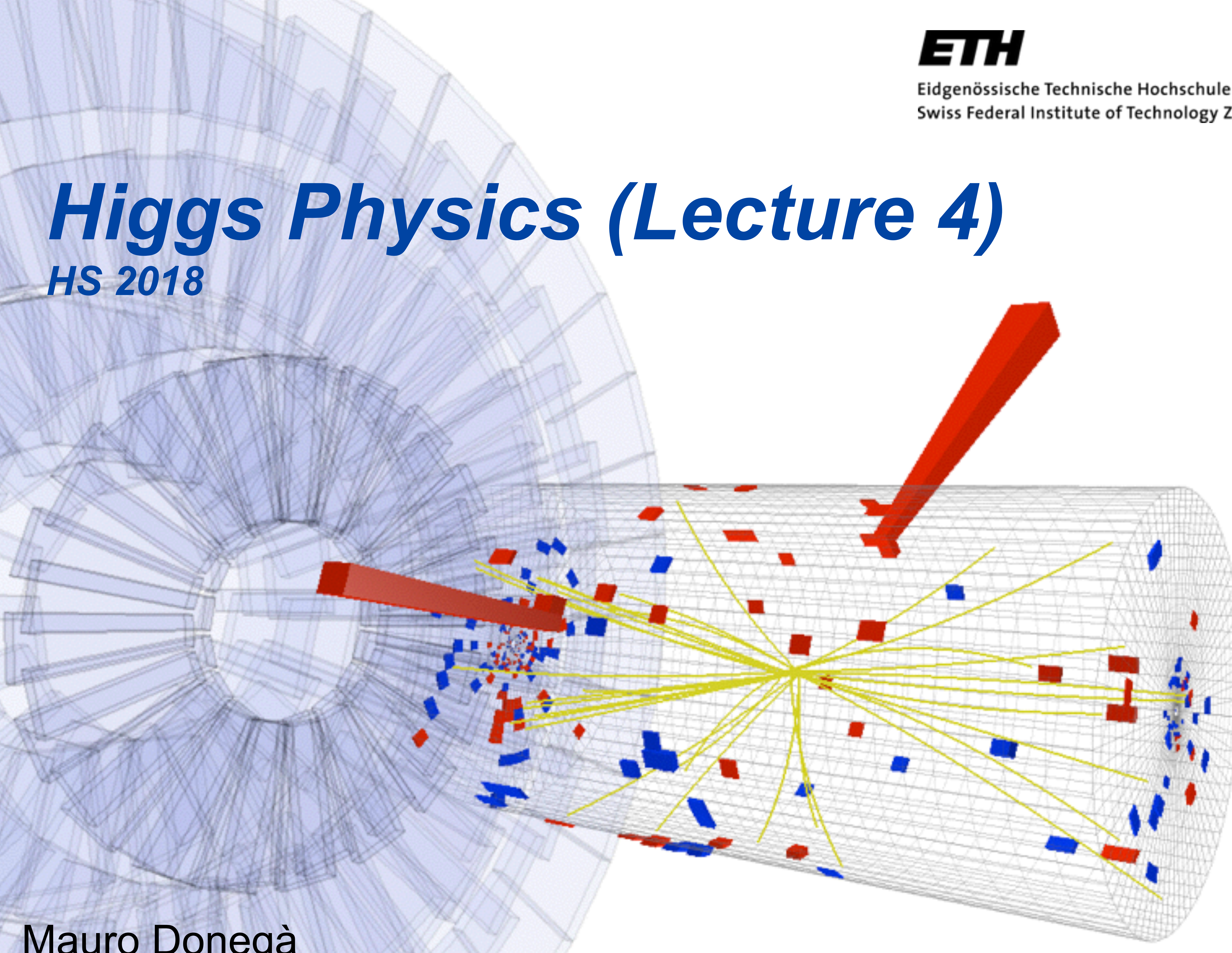
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Higgs Physics (Lecture 4)

HS 2018



Mauro Donegà

Outline

Lectures

- 1
  - Introduction
  - Accelerators
  - Detectors
  - EW constraints
- 2
  - Search at LEP1 / LEP 2
  - Statistics: likelihood and hypothesis testing
- 3
  - Searches at TeVatron
    - Channels overview
    - Neural Networks
    - Results
- 4
  - Boosted Decision Trees
  - Statistics at the LHC
- 5
  - LHC
    - Dissect one analysis H→γγ
    - Channels overview
- 6
  - Higgs properties

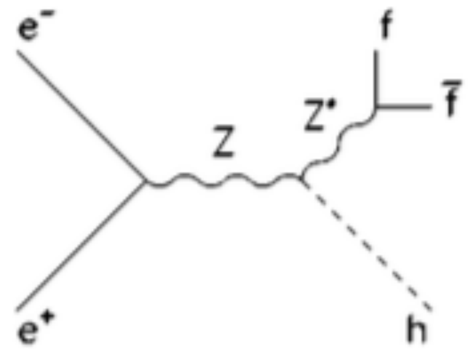
Exercises

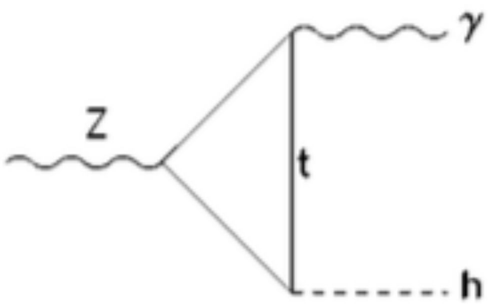
- W mass
- CLs
- Hbb TeVatron
- SciKit MVA
- Toy CVCF

Higgs searches at LEP

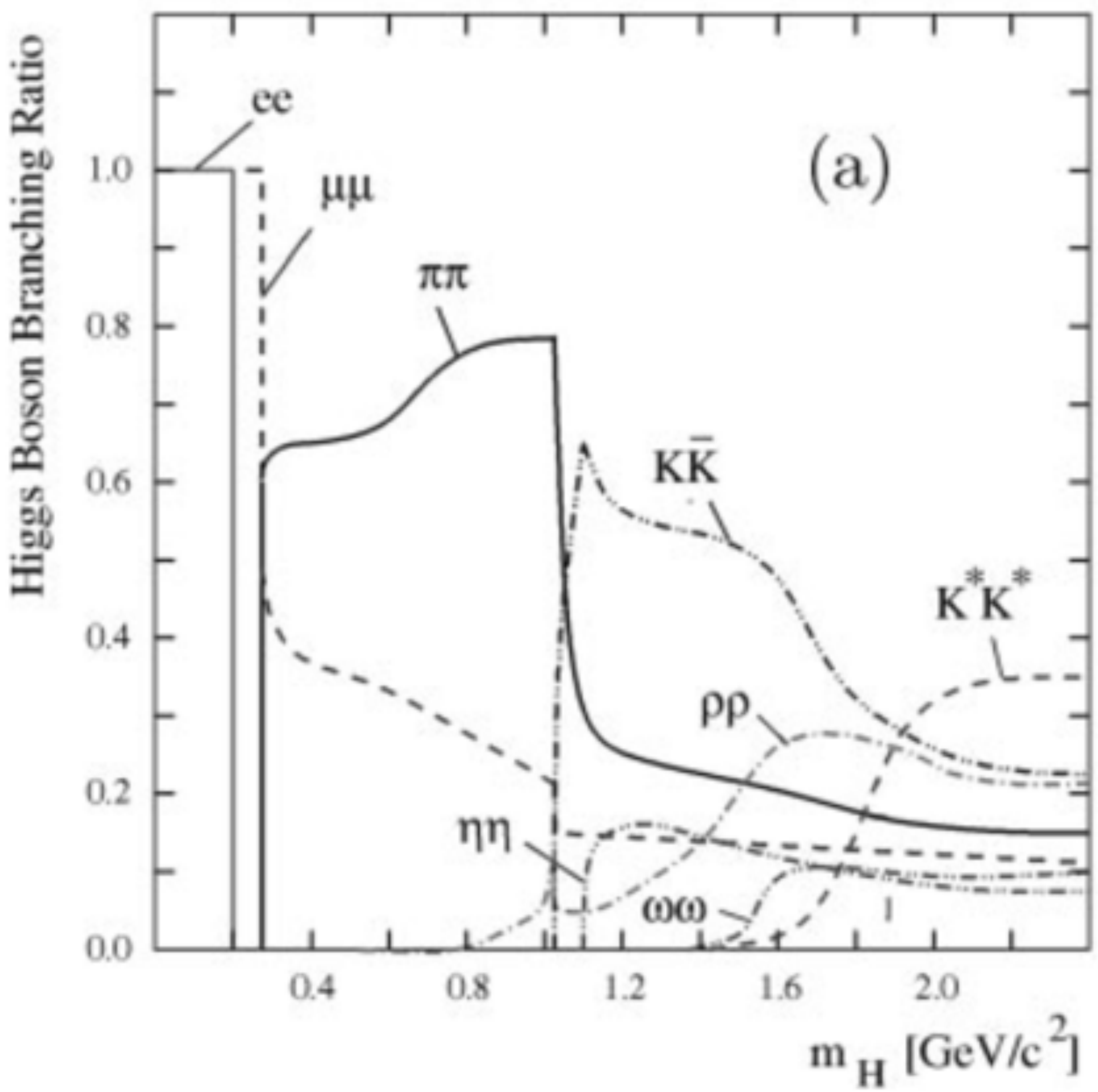
LEP 1 searches (1989-1995)

Production mechanisms :





(much smaller cross section
larger backgrounds ee→qqγ)

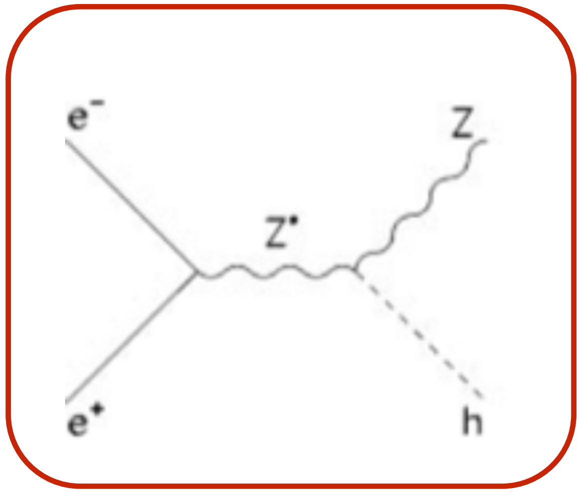


$2 < m_H < 20$ GeV
look for heaviest fermion kinematically allowed

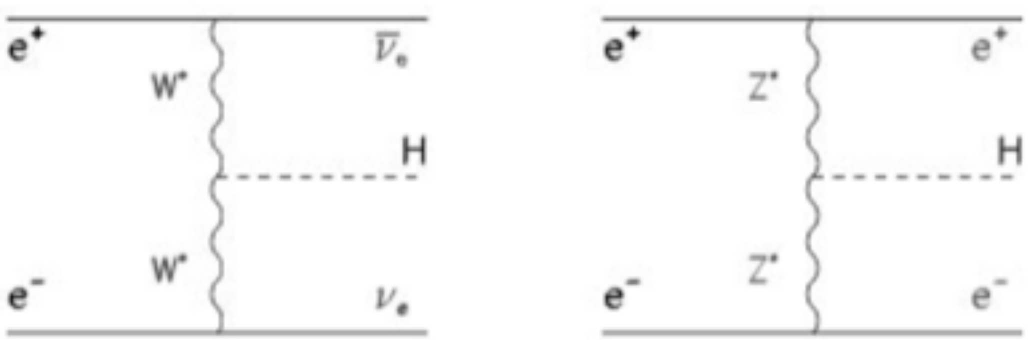
$m_H > 20$ GeV
Higgsstrahlung: $Z \to \nu\nu$ or $Z \to e^+e^-$ ($\mu^+\mu^-$)
$H \to bb$
10 signal events expected in 13 10^6 events
(all other channels swamped by $Z \to qq$ bkgs)

LEP 1 limit : mH > 65.6 GeV @ 95% CL

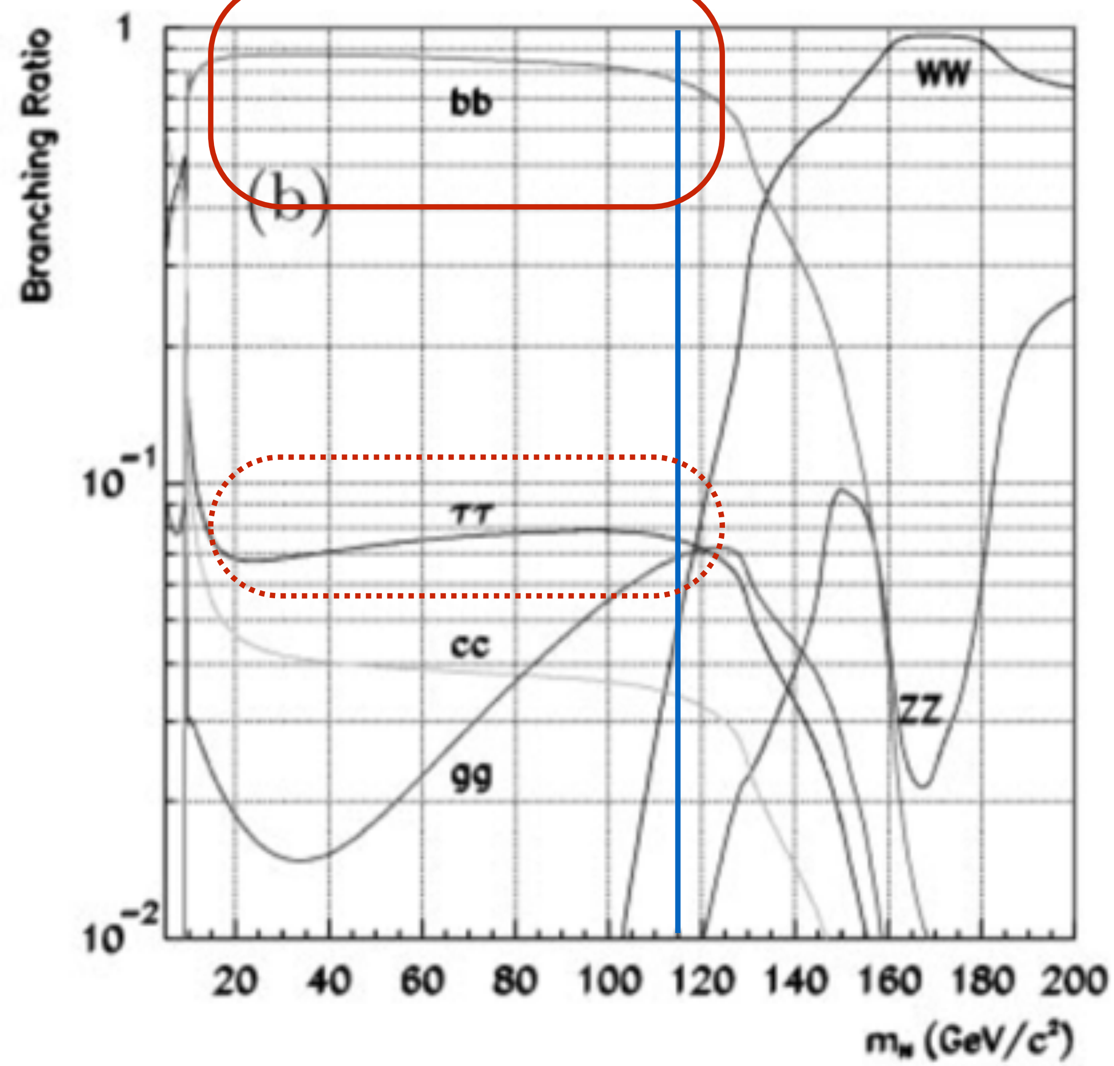
Production and decay



Higgsstrahlung



Vector Boson Fusion (VBF)



Compare this with an updated plot

(and interference between the two production modes $Z \to e^+e^-$, $Z \to \nu\nu$)

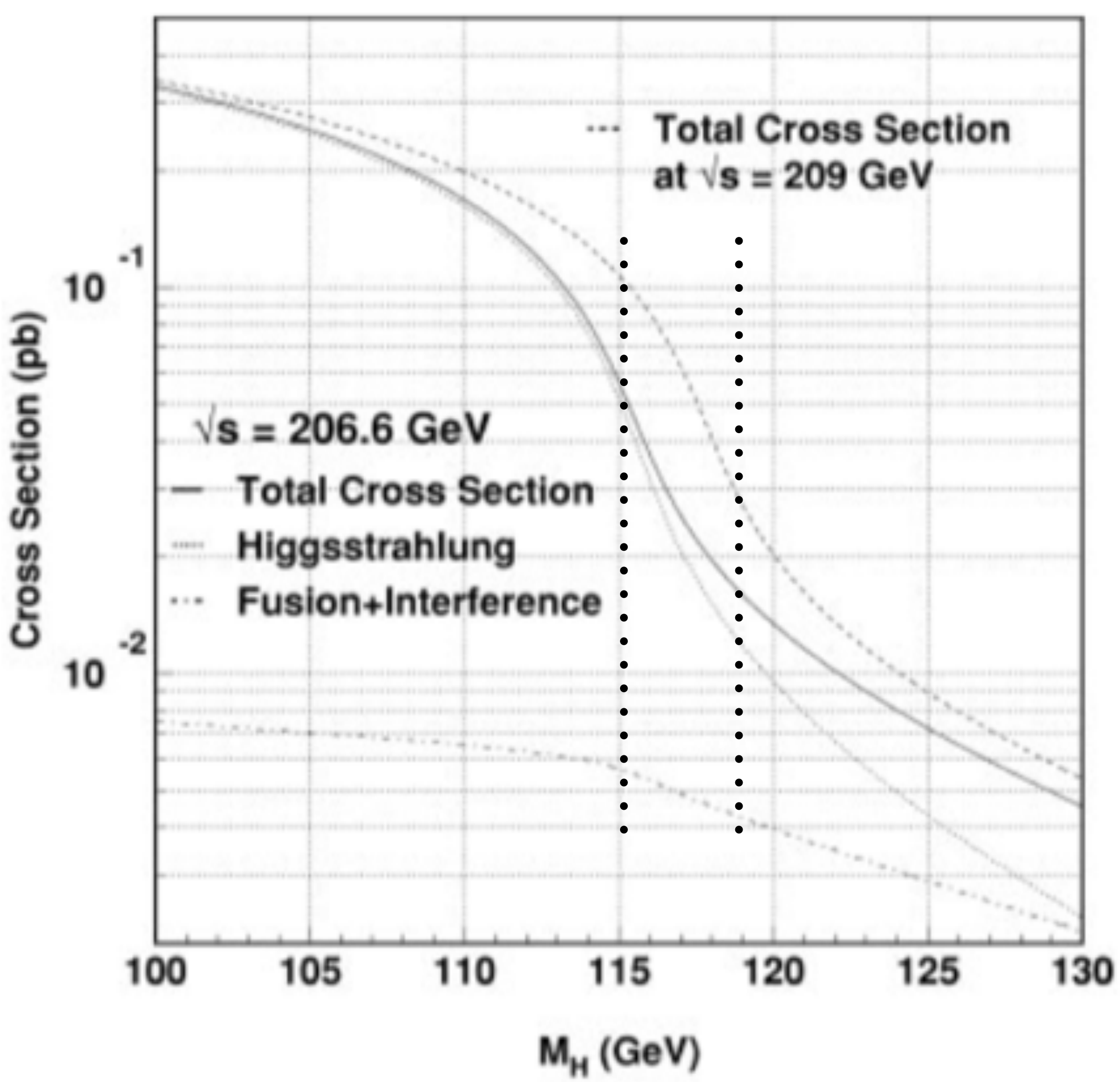
LEP 2 (1995-2000)

Threshold effect : $m_{threshold} = \sqrt{s} - m_Z$

Confront this with an hadron collider

⇒ the LEP Higgs sensitivity

depends dramatically on $\sqrt{s}$



Machine design highest energy:
RF: 6 MV/m, $\sqrt{s}$ = 192 GeV
mH sensitivity up to ~100 GeV

Upgrade and be smart to surpass the design capabilities:

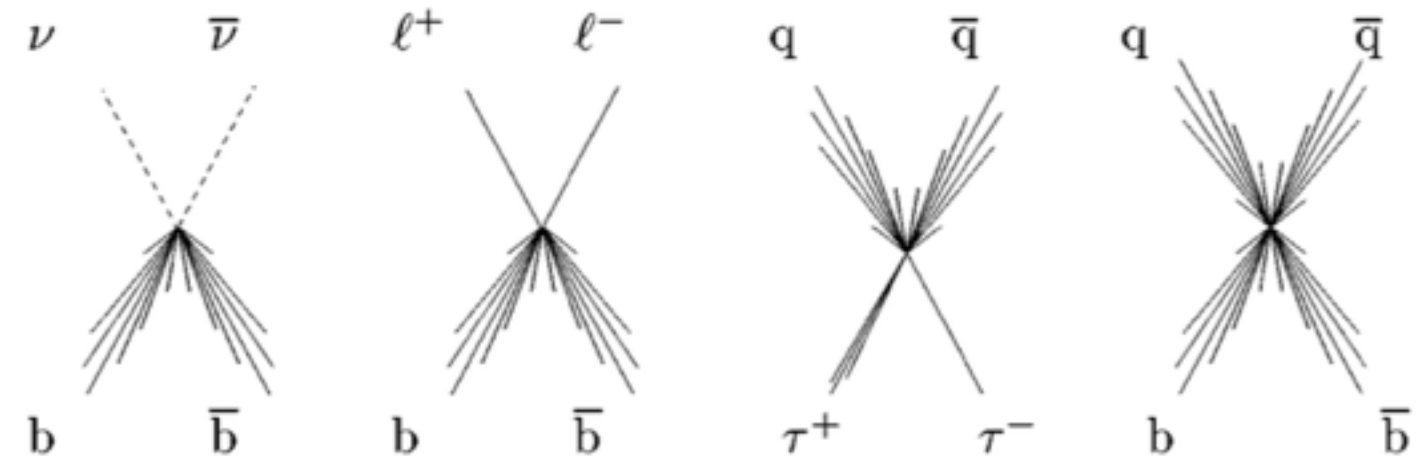
- upgrade cryogenics: RF up to 7.5 MV/m (increased stability)
  $\sqrt{s}$ = 204 GeV ; mH < 112 GeV
- run with one klystron margin (average trip frequency reduced below 1h thanks to improved stability)
  $\sqrt{s}$ = 205.5 GeV ; mH < 113 GeV
- reduce 350MHz RF by 100Hz (different orbit effectively more bending from quadrupoles)
  $\sqrt{s}$ = 206.6 GeV ; mH < 113.6 GeV
- Unused orbit correctors used as dipole
  $\sqrt{s}$ = 207.3 GeV ; mH < 113.85 GeV
- reinstall 8 old Cu cavities from LEP 1
  $\sqrt{s}$ = 207.7 GeV ; mH < 114 GeV
- miniramps (tradeoff between energy/stability/fill time)
  $\sqrt{s}$ = ~209 GeV ; mH ~ 115 GeV

Signal topologies / Backgrounds

Signal:

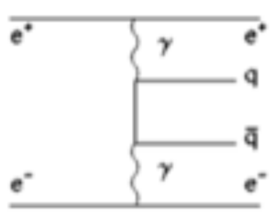
Z →

H →



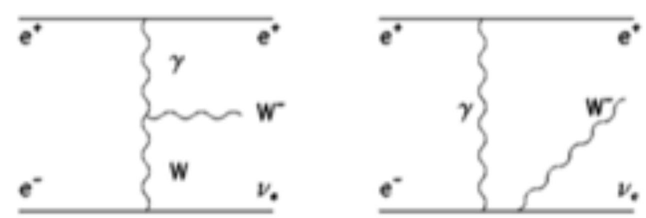
this means both hadronic and leptonic tau decays

Background:



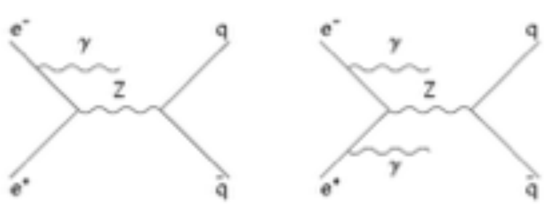
Two photon process

$\mathcal{O}$(10nb)



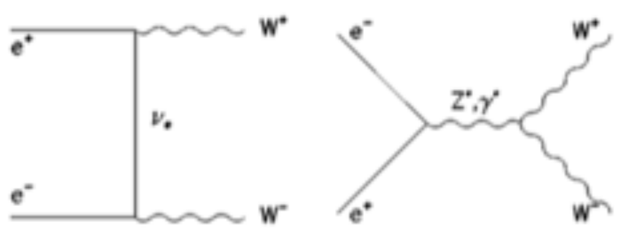
W→tau nu

$\mathcal{O}$(1pb)



$\mathcal{O}$(100pb)

$\mathcal{O}$(7pb)



$\mathcal{O}$(17pb)

$\mathcal{O}$(10fb)

$\mathcal{O}$(3pb)

And reducible backgrounds
e.g light jets mis-reconstructed as b-jets / taus

Analyses

SM backgrounds at LEP were well modelled (generators) and simulated (detectors).
Often use Monte Carlo to model the backgrounds !

Calibrate the detectors at the Z peak

Confront this situation with hadron colliders

Measurements development: get started by studying SM rare processes WW/ZZ production before attacking the Higgs

4 jets channel:
the most sensitive at LEP2:

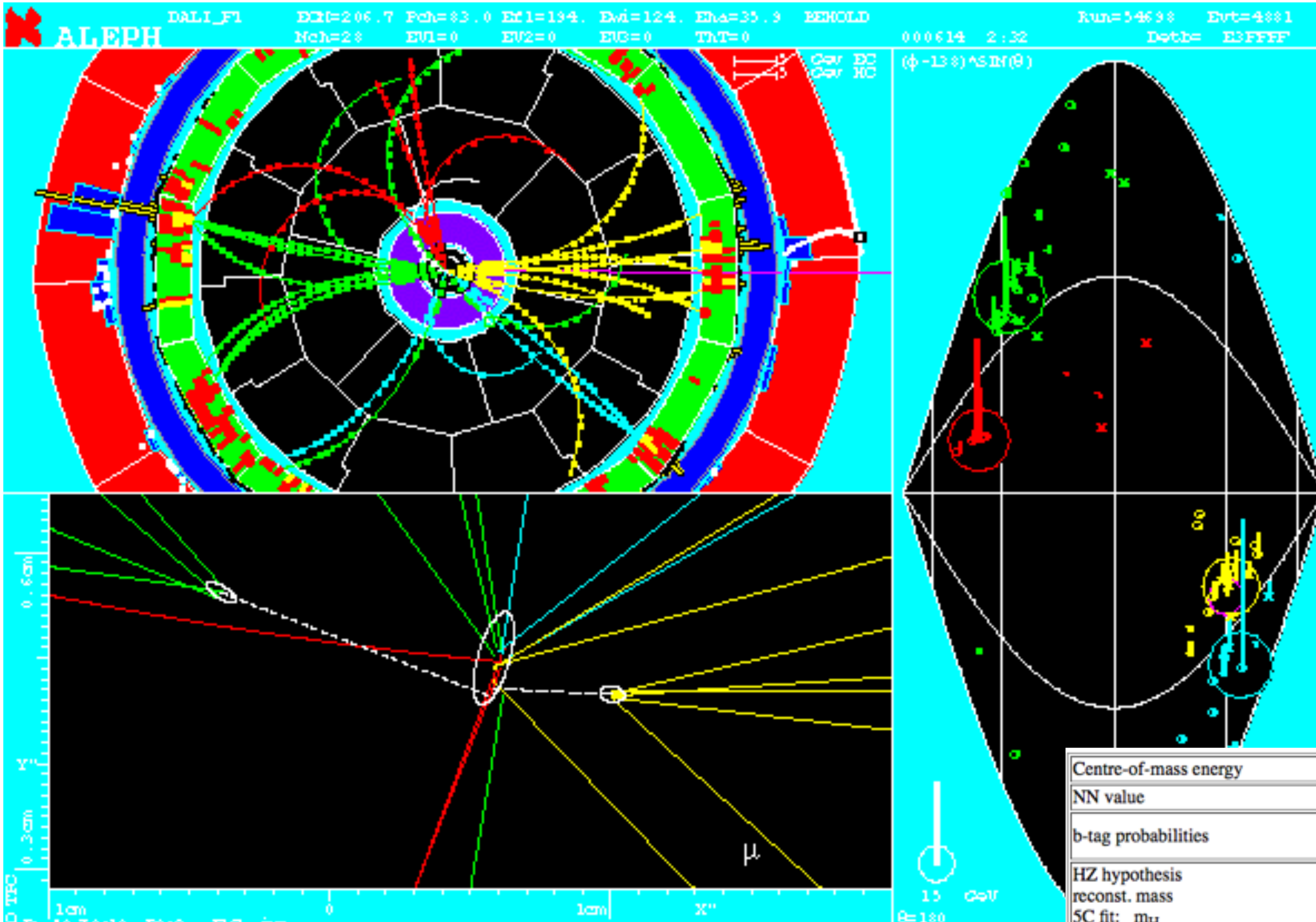
- at kinematic threshold Z and H are produced at rest: 4 jets in one plane
  - main backgrounds $e^+e^- \rightarrow ZZ$, $e^+e^- \rightarrow WW$

- $Z \rightarrow bb$: 4b case high purity but pairing ambiguities
  - main background ZZ

- typical mass resolution ~3 GeV

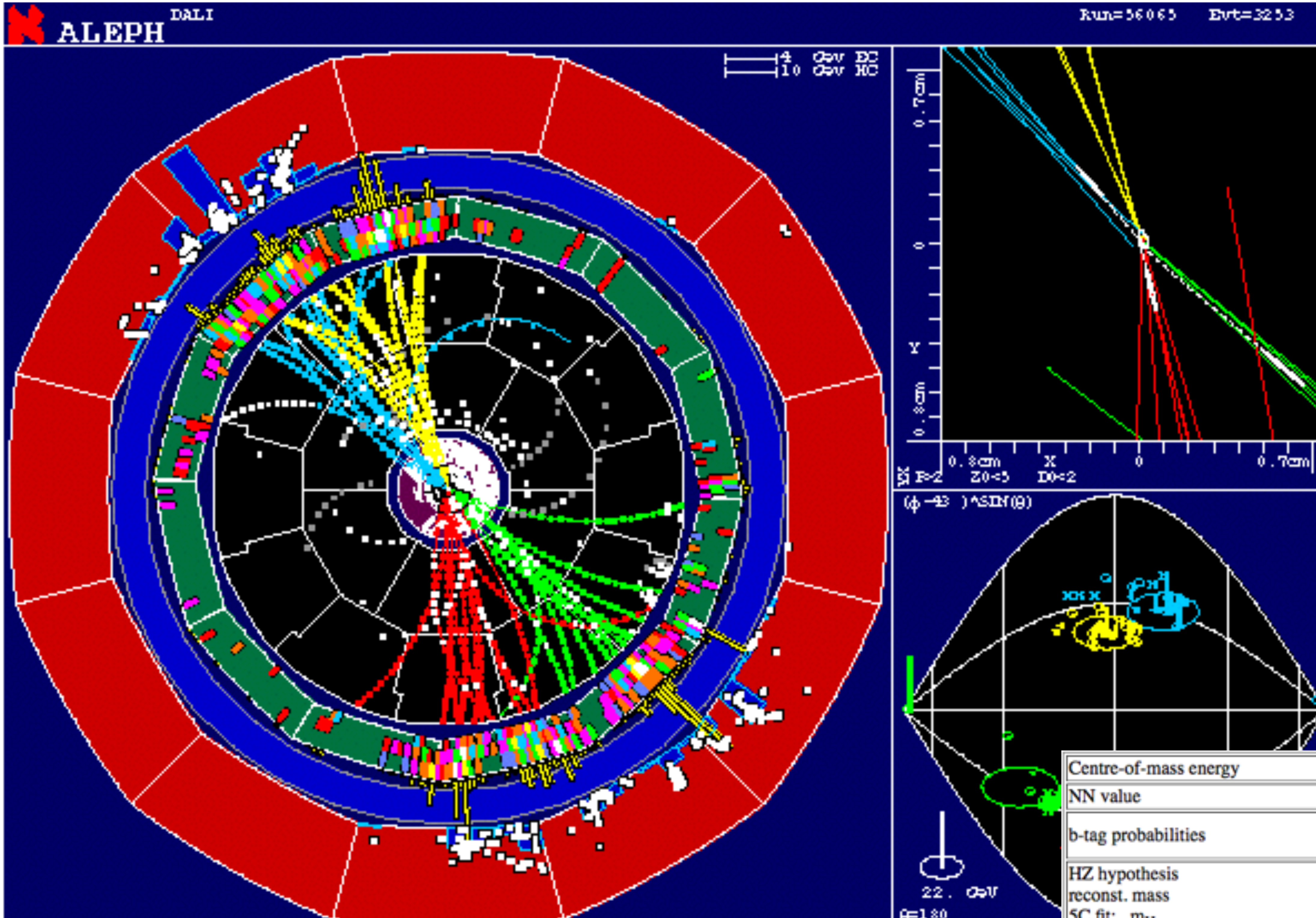
Channels sensitivity is different at LEP and at the hadron colliders: why ?
what happens to the 4 jets channel ?

qqbb



Centre-of-mass energy	206.7 GeV
NN value	0.996
b-tag probabilities	0.99 0.99 0.14 0.01
HZ hypothesis reconst. mass 5C fit: m_H m_Z	 114 ± 3 GeV 112.4 GeV 93.3 GeV
ZZ hypothesis m_Z m_Z	 102 GeV 91.7 GeV

bbbb



Centre-of-mass energy	206.7 GeV
NN value	0.997
b-tag probabilities	0.99 0.99 0.99 0.78
HZ hypothesis reconst. mass 5C fit: m_H m_Z	 114 ± 3 GeV 113.8 GeV 91.9 GeV
ZZ hypothesis m_Z m_Z	 97 GeV 94 GeV

3 reconstructed secondary vtx

Analyses

Missing Energy: Why at an hadron collider we use the missing *transverse* energy ?

- $H \rightarrow bb$ and $Z \rightarrow \nu\nu$
- Main background : ZZ (irreducible)

- typical mass resolution as in the 4jets (!) ~3 GeV

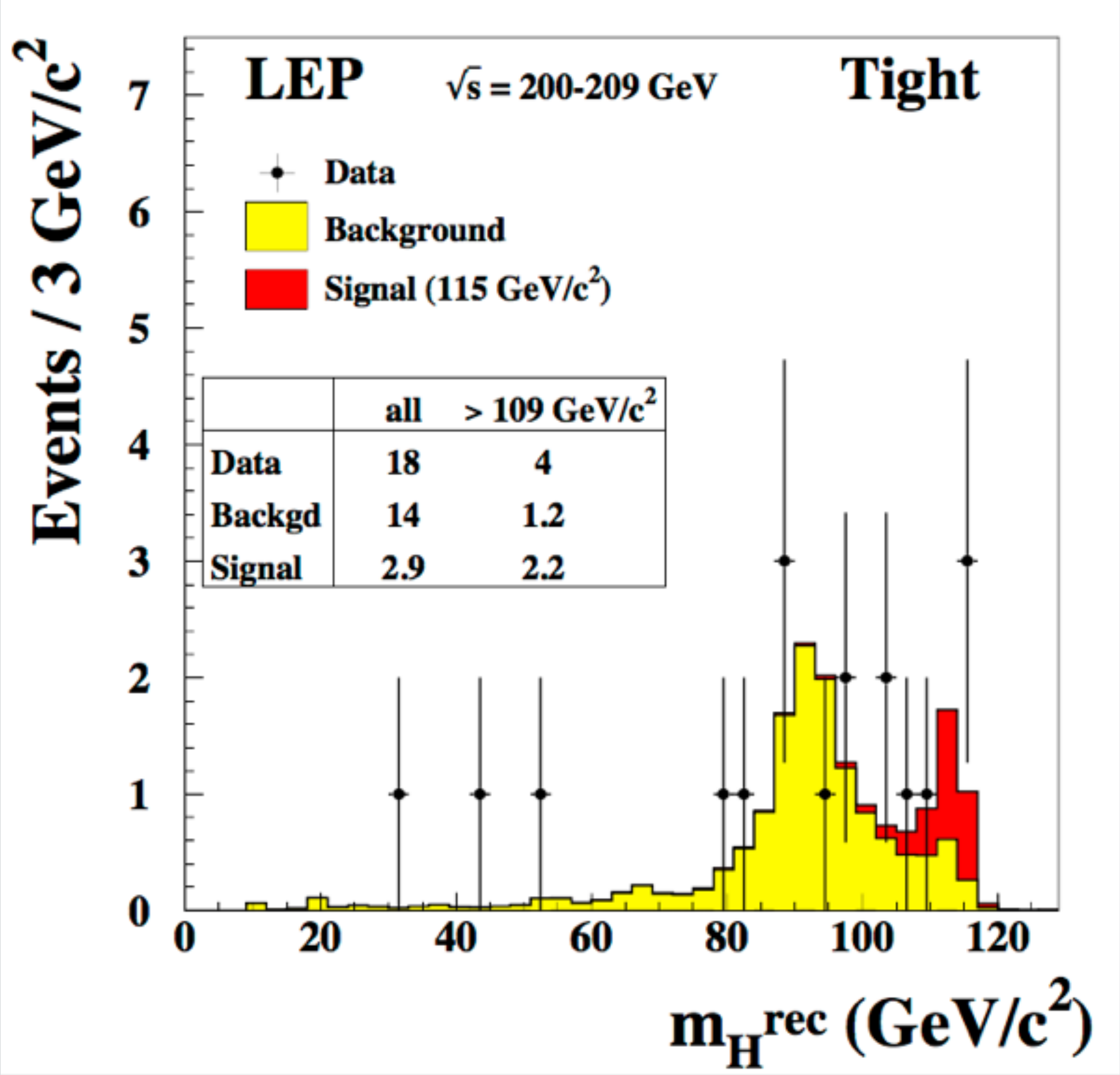
l^+l^- channel:

- $H \rightarrow bb$ and $Z \rightarrow l^+l^-$
- very small branching ratio (3% $Z \rightarrow ll$)
- Main background : ZZ (irreducible)

$\tau^+\tau^-$ channel:

- $H \rightarrow bb$ and $Z \rightarrow \tau^+\tau^-$ (neutrino in the final state)
- very small branching ratio (3% $Z \rightarrow ll$)
- Main background : ZZ (irreducible) and $Z \rightarrow$ bb (mis-reconstructed as τ)

Statistical inference at LEP



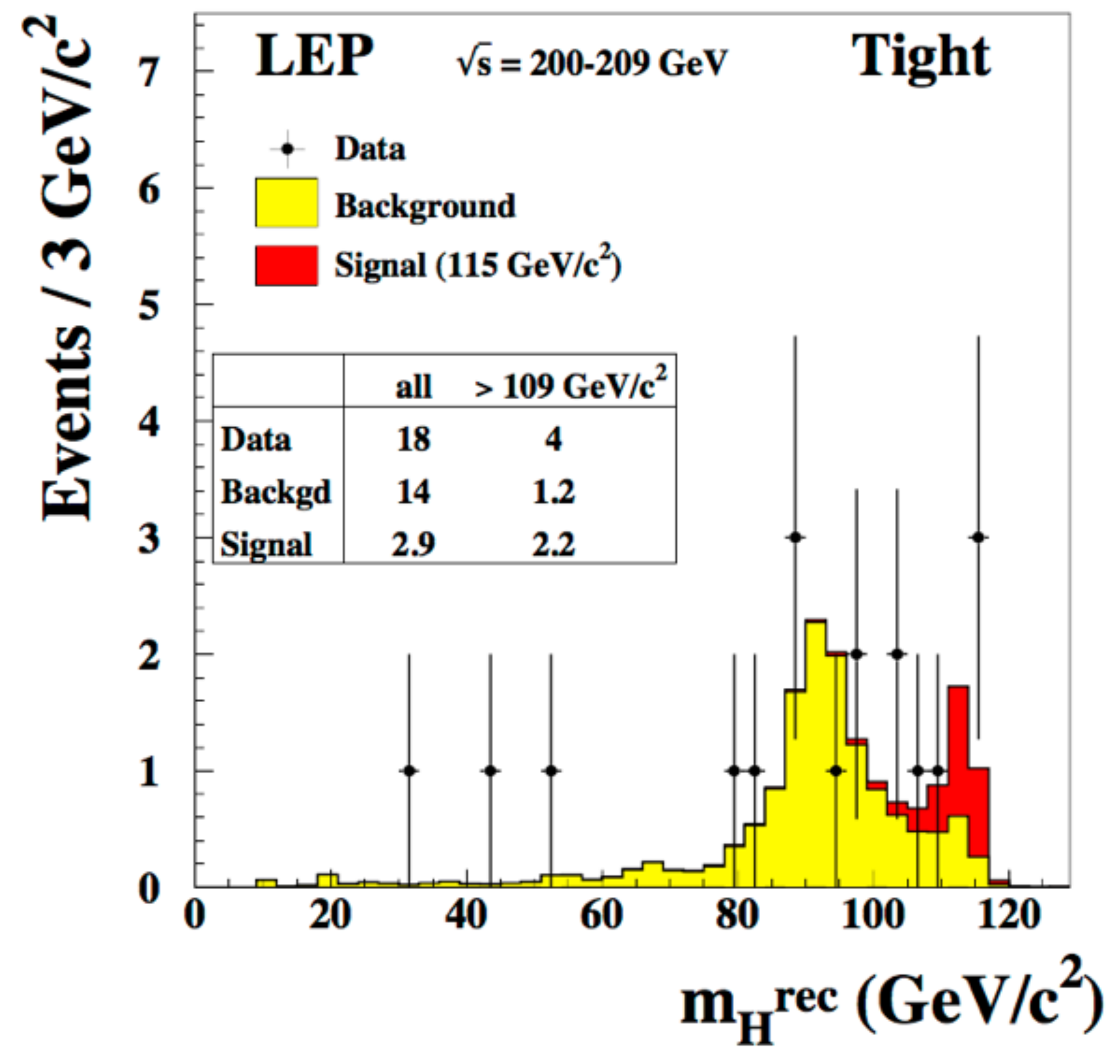
	all	> 109 GeV/c²
Data	18	4
Backgd	14	1.2
Signal	2.9	2.2

Use data to take decisions

Formulate an hypothesis (precisely), collect data, test the data against the hypothesis then accept or reject.

The way the hypothesis are defined is “reversed”, i.e. you always check that a hypothesis is NOT consistent with data.

In statistics/physics one cannot meaningfully accept a hypothesis: one can ONLY reject them.



Definitions:

H_0: null hypothesis defined to be the hypothesis under consideration.

H_1: Alternative hypothesis

Typically H_0 is the background only hypothesis while H_1 adds the presence of some signal.

Simple hp: the expected PDF of the random variable (data) is completely fixed/specified

Composite hp: not all parameters are fixed, but they lie within a range

Test statistics

To quantify the agreement between the observed data and a given hypothesis one constructs a function of the measured data (x) and the given hypothesis Hp

$$\text{test statistics} := t(x|Hp)$$

Typical test statistics:
- number of events
- a function of the observables (invariant mass from a 4 vector)
- a likelihood
- a ratio of likelihoods
-…

The choice of the test statistic $t(x|Hp)$ depends on the particular case,
there is no general rule !

Different test statistics will give different "results": PHYSICS judgement is important !

Test statistics

Once you defined your test statistics you want to know how it is distributed in the H_0 (bkg only) and H_1 (sig+bkg) hypotheses

$$P(t|Hp)$$

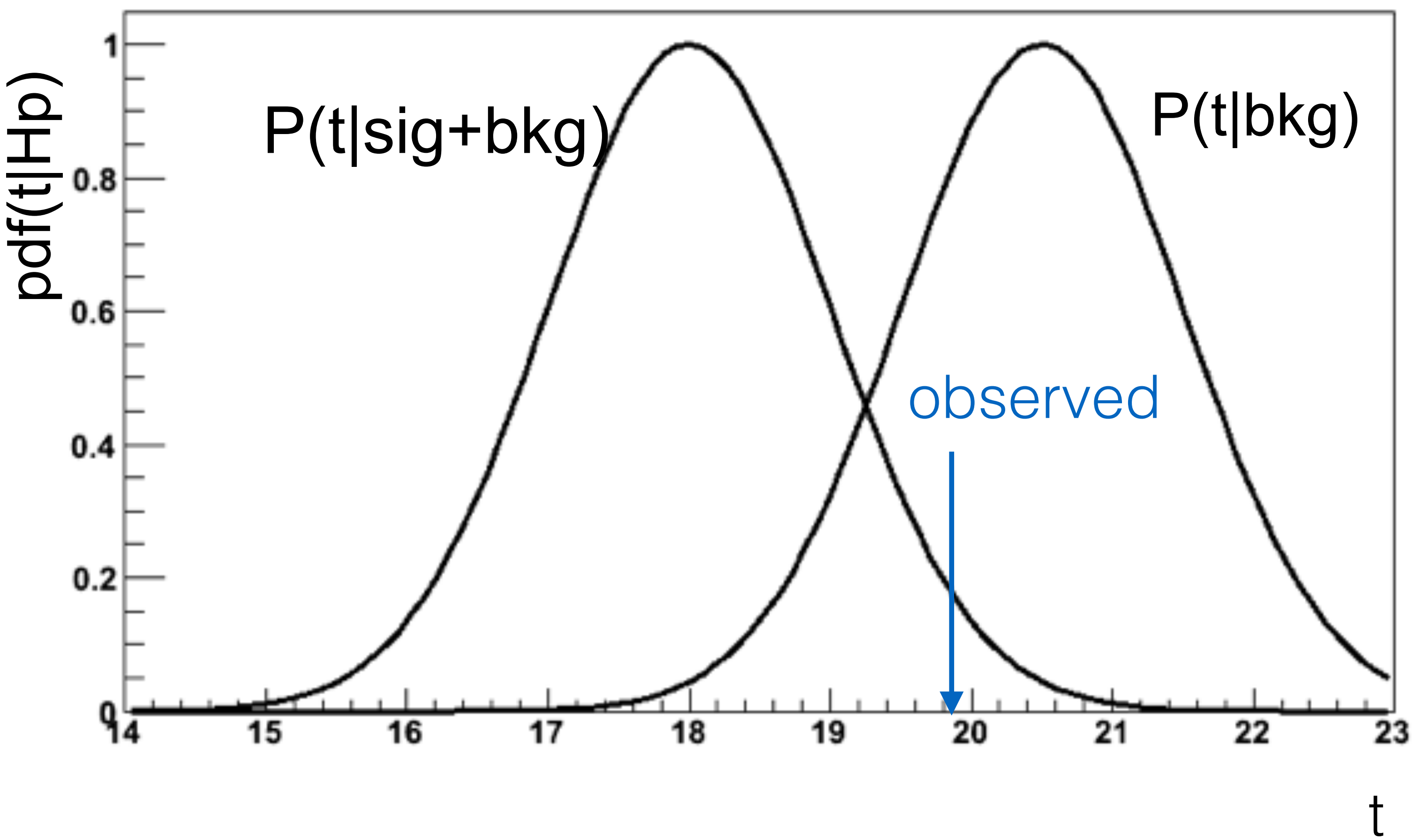
Naively to produce the distribution of the test statistics you would use a pure sample of background to get $P(t|H_0)$ and a sample of signal+background to get $P(t|H_1)$.

In general we don't have a labelled sample of signal and background samples from data.
Sometimes we can use data in "signal-free" control regions to build the pdf for the background, but often (and by definition in case of searches) we don't have a clean sample of "background-free" signal to build its pdf.

Use Monte Carlo and in particular "toy samples"
Toy means you don't run the full generator+detector simulation. You generate pseudo-data sampling some high level distributions (e.g. the reconstructed mass of the higgs candidate)

Hypothesis testing

Finally use the expected distributions of the test statistics to decide if a candidate is signal or background.



The name of the game will be to use the test statistics to quantitatively say if your data contains signal

Error types

Different ways of mistakenly interpret the data:
Type I: reject a true hypothesis (false negative)
Type II: accept a false hypothesis (false positive)

eg: “Law court” The accused proclaims himself as innocent (H_0).
Type I: he’s really innocent and the jury rejects the hypothesis and convict him
Type II: he’s really guilty (and a lier) and the court accept the hypothesis and let him off.

	H_0 is true **Truly not guilty**	**H_1 is true** **Truly guilty**
Accept Null Hypothesis Acquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

twiki

eg: “Bump hunting” You analyze a mass spectrum. The hypothesis is bkg only (H_0).
Type I : there’s really no resonance, you reject the H_0 and you publish rubbish
Type II: there is a real resonance, you accept H_0 and you miss the Nobel

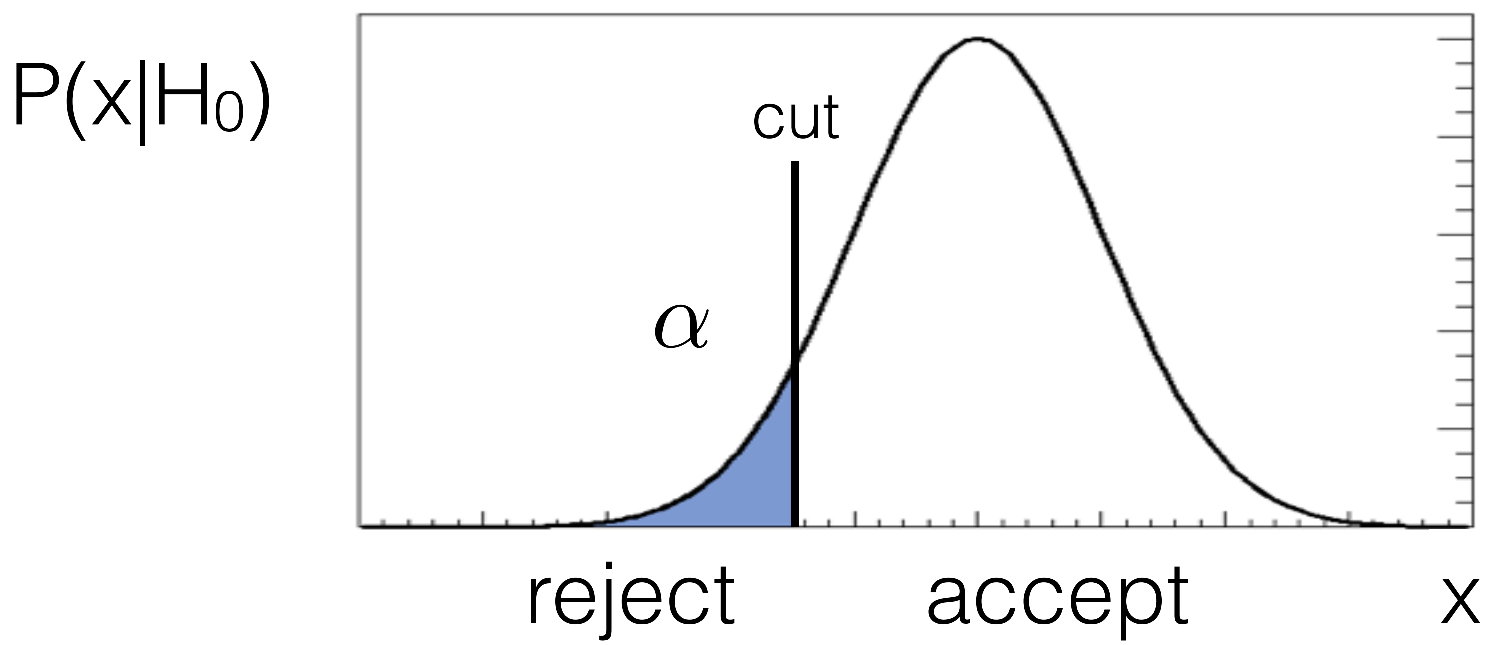
Test statistics properties

1) Significance (or Size)

Type I errors can be controlled pretty well:

Suppose you have a test statistics x (the data itself) and the null hypothesis H_0: $P(x|H_0)$

Partition the range of x in 2 regions. Define: acceptance / rejection



The probability for a type I error is the integral of $P(x|H_0)$ over the rejection region. This is called significance of the test

$$\alpha = \int_{-\infty}^{\text{cut}} P(x|H_0)dx$$

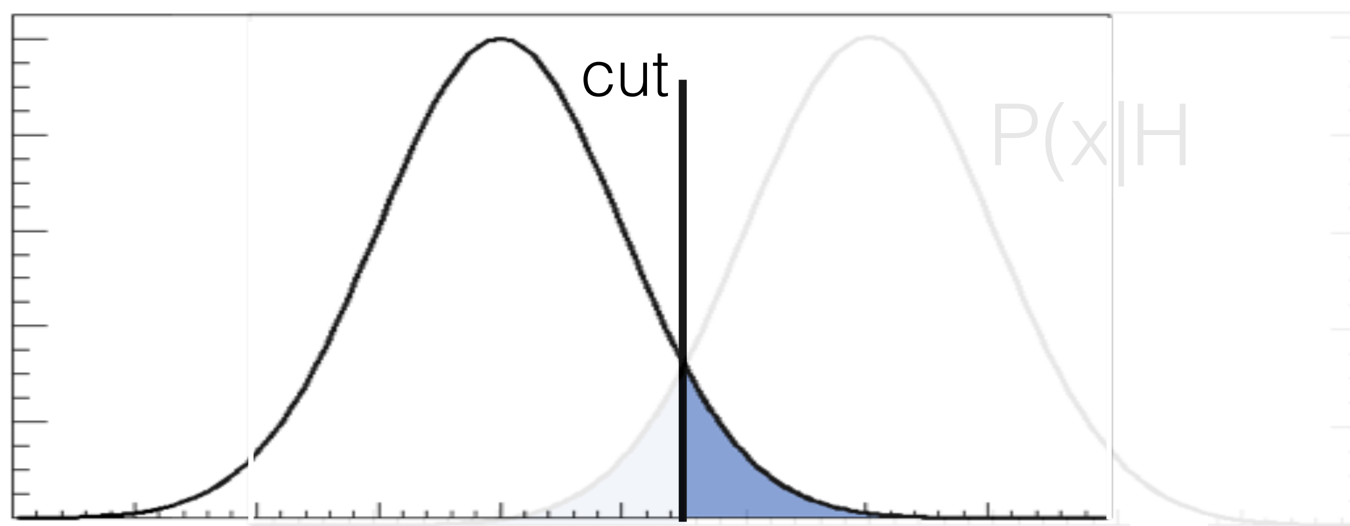
Typical values of α are 5%, 1%.

A test has a significant level of 1-α if the integrated probability to reject a true hypothesis is less of equal to α

Test statistics properties

2) Power

Suppose you have an alternative simple hypothesis H_1 and $P(x|H_1)$ is known.



$$\beta = \int_{\text{cut}}^{\infty} P(x|H_1)dx$$

$1 - \beta$ is called the power of the test

A good test is the one with both α and β small,
i.e. high significance and high power
(i.e. H_0 and H_1 very different; large separation)

Likelihood ratio

Remember: the "best" test is the one that makes both α and β as small as possible. If both H_0 and H_1 are simple we can use the:

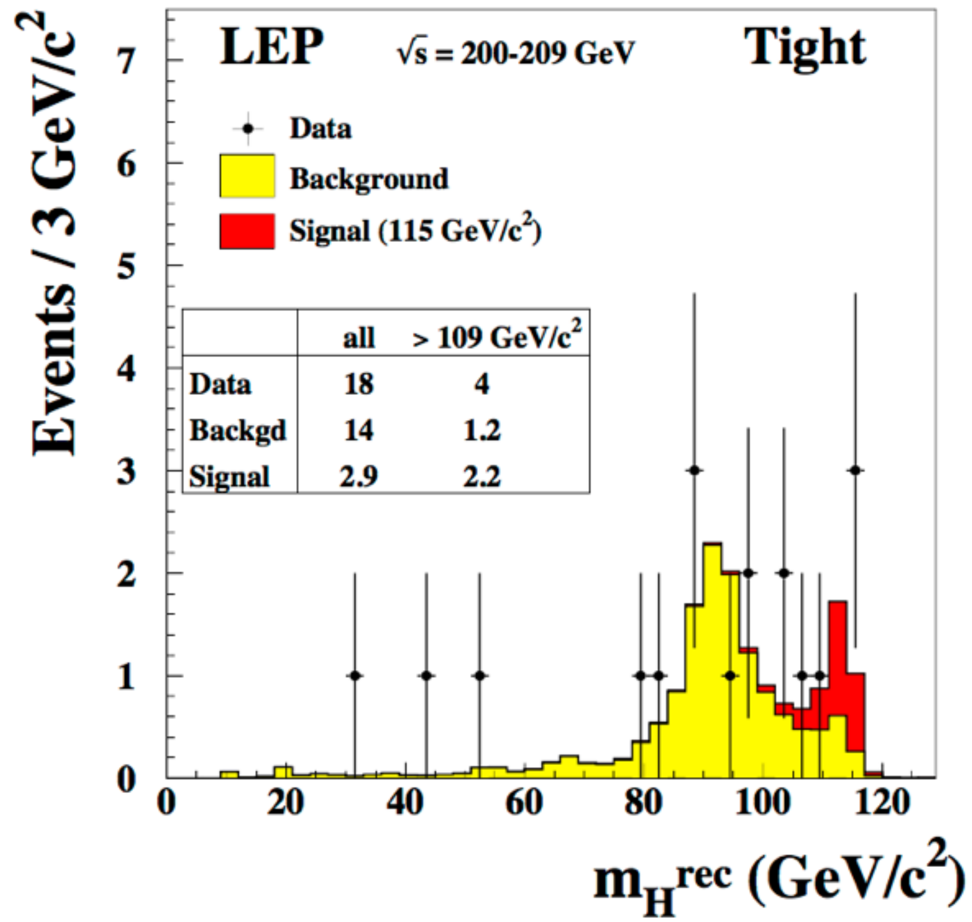
Neyman-Pearson lemma:
the acceptance region giving the highest power (i.e. the highest purity) for a given significance level α (or efficiency 1-α) is the region of the space such that

$$\frac{g(\mathbf{t}|H_0)}{g(\mathbf{t}|H_1)} > c.$$

c is determined by the desired efficiency.

The one dimensional statistics $r = \frac{g(\mathbf{t}|H_0)}{g(\mathbf{t}|H_1)}$ is called "likelihood ratio"

LEP test statistics



Take as an example variable the reconstructed mass of the Higgs candidate. The variable is binned (histograms).

For each bin we know the expected number of events from Signal and Background.

The probability to observe a number of events n with v expected is given by: $P(n,\nu) = \frac{\nu^n}{n!} e^{-\nu}$ (Poisson)

The test statistics at LEP was chosen to be the likelihood ratio: $Q = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}$

$$\mathcal{L}_{s+b} = \frac{(s+b)^n}{n!} e^{-(s+b)} \cdot \prod_{j=1}^{n} \frac{sS(x_j) + bB(x_j)}{s+b}$$

$$\mathcal{L}_b = \frac{b^n}{n!} e^{-b} \cdot \prod_{j=1}^{n} B(x_j)$$

binned

where:
s = # expected sig events, function of m_H
b = # expected bkg events
n = # observed events
x_j = value of the discriminating var j
$S(x_j)$ = signal pdf for the vars x, function of m_H
$B(x_j)$ = bkg pdf for the vars x

LEP test statistics

Generalization to N measurements (k runs over the measurements)

$$\mathcal{L}(\eta) = \prod_{k=1}^{N} \frac{\exp[-(\eta s_k(m_{\rm H}) + b_k)]\ (\eta s_k(m_{\rm H}) + b_k)^{n_k}}{n_k!} \times \prod_{j=1}^{n_k} \frac{\eta s_k(m_{\rm H}) S_k(\vec{x}_{jk}; m_{\rm H}) + b_k B_k(\vec{x}_{jk})}{\eta s_k(m_{\rm H}) + b_k}$$

where: $\eta = 1$ gives $\mathcal{L}_{s+b}$
$\eta = 0$ gives $\mathcal{L}_{b}$
k runs over the N channels (different decay, different data periods, etc…)
s_k = # expected sig events
b_k = # expected bkg events
n_k = # observed events in channel k
x_{jk} = value of the discriminating var j in channel k
$S(x_{jk})$ = signal pdf for the vars x in channel k
$B(x_{jk})$ = bkg pdf for the vars x in channel k

$$Q = \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}$$

$$q = -2\ln Q(m_{\rm H}) = 2\sum_{k=1}^{N}\left[s_k(m_{\rm H}) - \sum_{j=1}^{n_k} \ln\left(1 + \frac{s_k(m_{\rm H}) S_k(\vec{x}_{jk}; m_{\rm H})}{b_k B_k(\vec{x}_{jk})}\right)\right]$$

Each event contributes with a weight to the test statistics

To avoid numerical precision issues in treating very small numbers (we're multiplying several small probabilities, i.e. numbers 0<=p<=1) we usually work with the logarithm of Q

Systematic uncertainties

Systematics uncertainties are included through nuisance parameters.

e.g. the background is known with an uncertainty, so: $b_k \rightarrow b_k \cdot f(\theta_k, \sigma_k)$

best value of b_k

uncertainty on ϑ_k

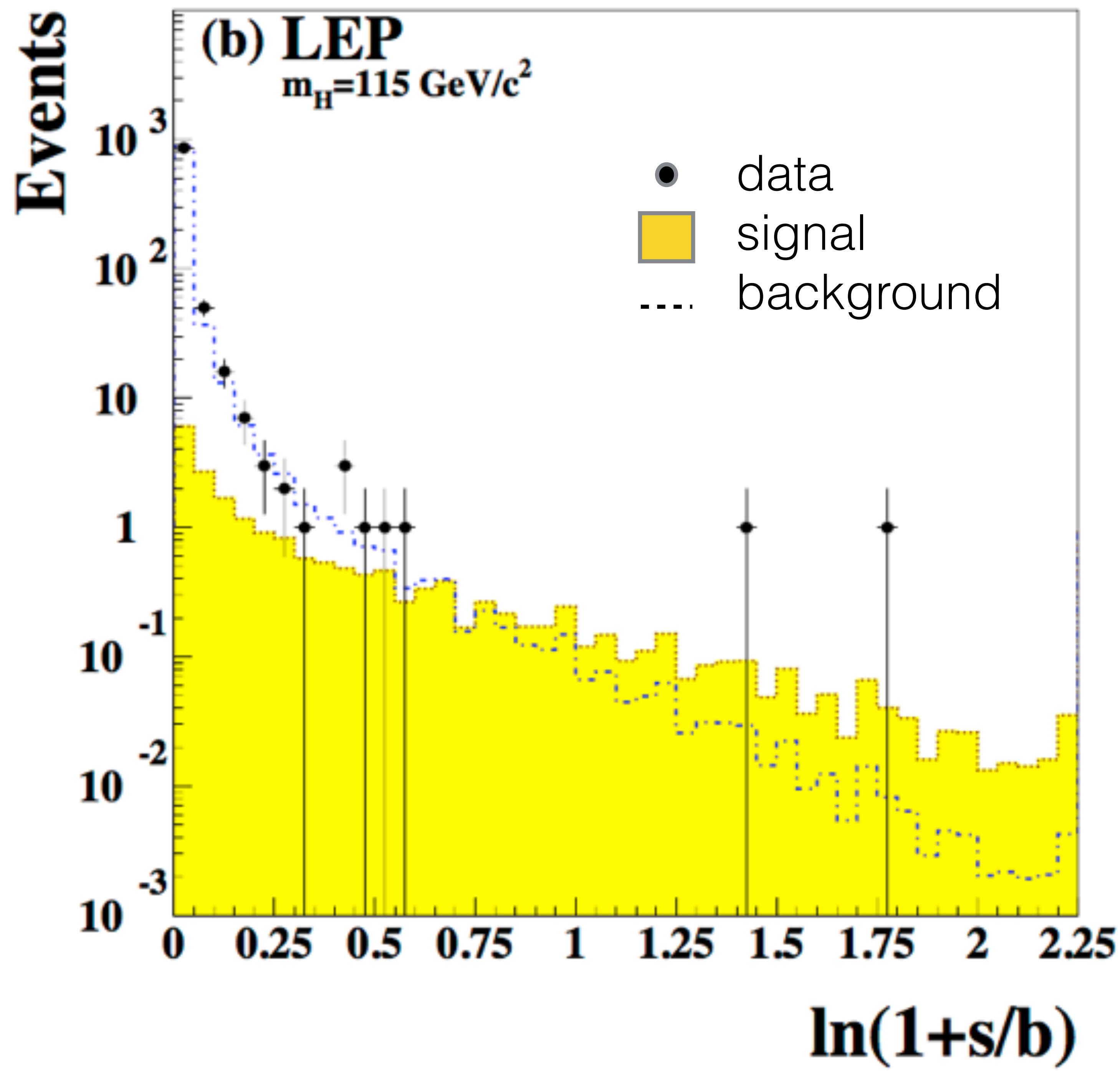
The function *f* represents a constraint on the parameter b_k

Each term of the likelihood affected by a systematic uncertainty gets multiplied by a constraint term (the functional form of the constraint depends on the variable and it can be a gaussian, log-norm, etc…)

LEP results

$$-2\ln Q(m_{\rm H}) = 2\sum_{k=1}^{N}\left[s_k(m_{\rm H}) - \sum_{j=1}^{n_k}\ln\left(1+\frac{s_k(m_{\rm H})S_k(\vec{x}_{jk};m_{\rm H})}{b_kB_k(\vec{x}_{jk})}\right)\right]$$

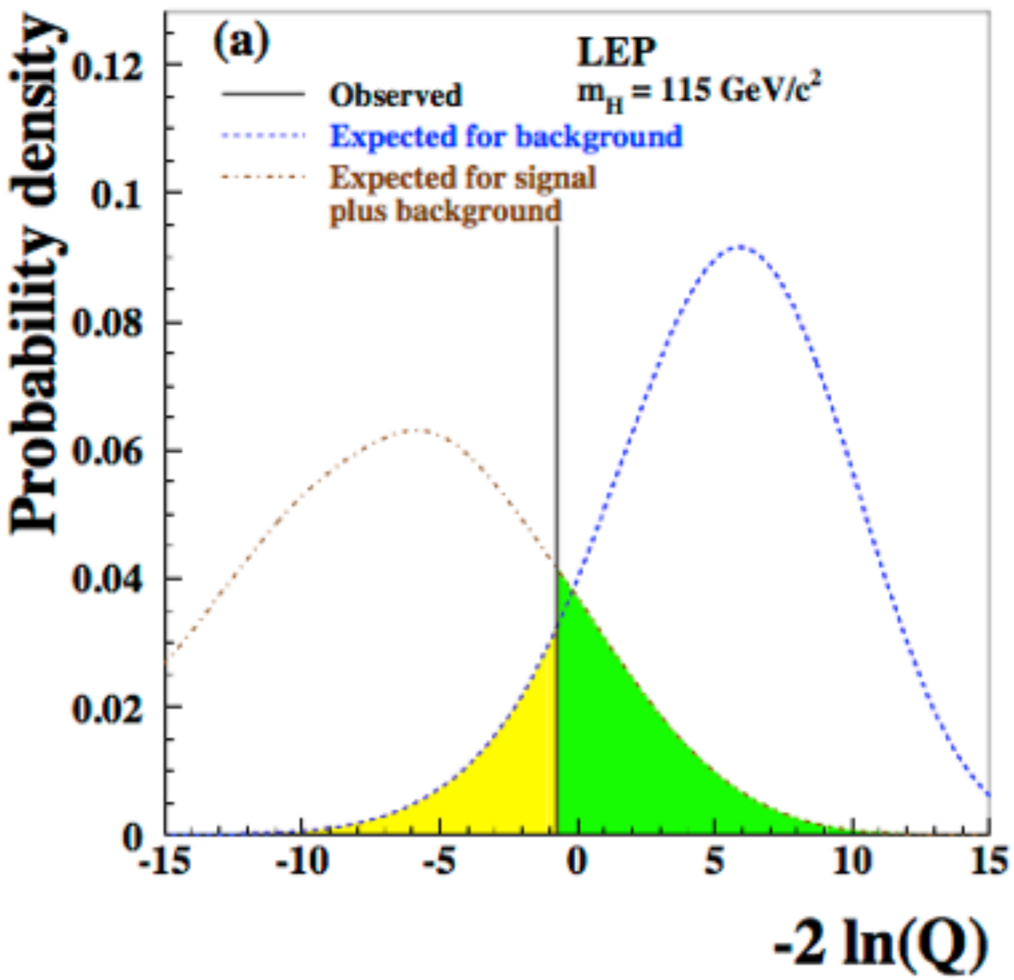
Because the separation between signal and background depends logarithmically on S/B, the right tail of the log(1+s/b) distribution shows the important region for signal search



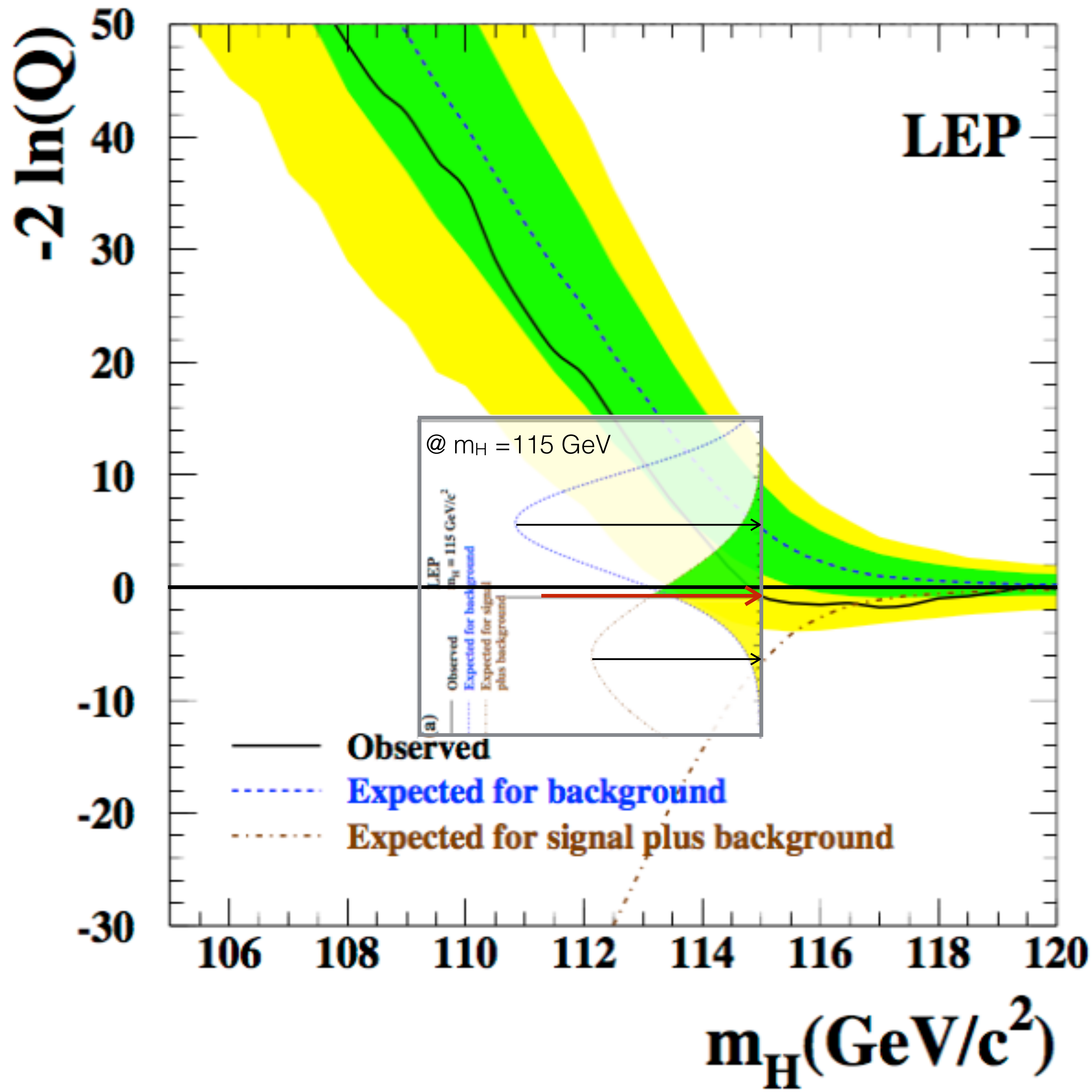
-2 lnQ for different hypotheses

$$-2\ln Q(m_{\mathrm{H}}) = 2\sum_{k=1}^{N}\left[s_k(m_{\mathrm{H}}) - \sum_{j=1}^{n_k}\ln\left(1+\frac{s_k(m_{\mathrm{H}})S_k(\vec{x}_{jk};m_{\mathrm{H}})}{b_k B_k(\vec{x}_{jk})}\right)\right]$$

Notes:
- q depends on the test mass
- q on data is computed with the nuisance at their best fit value

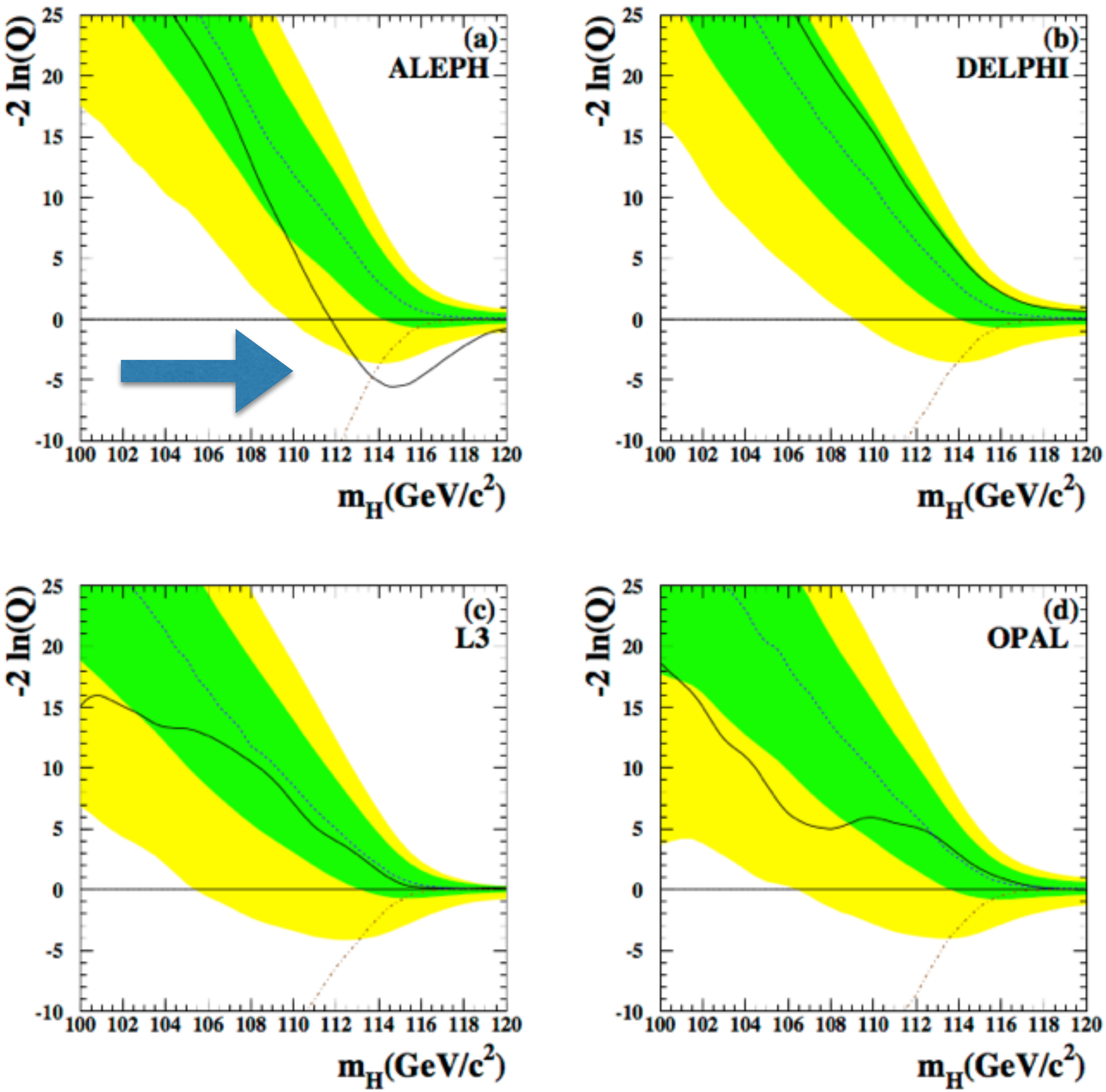


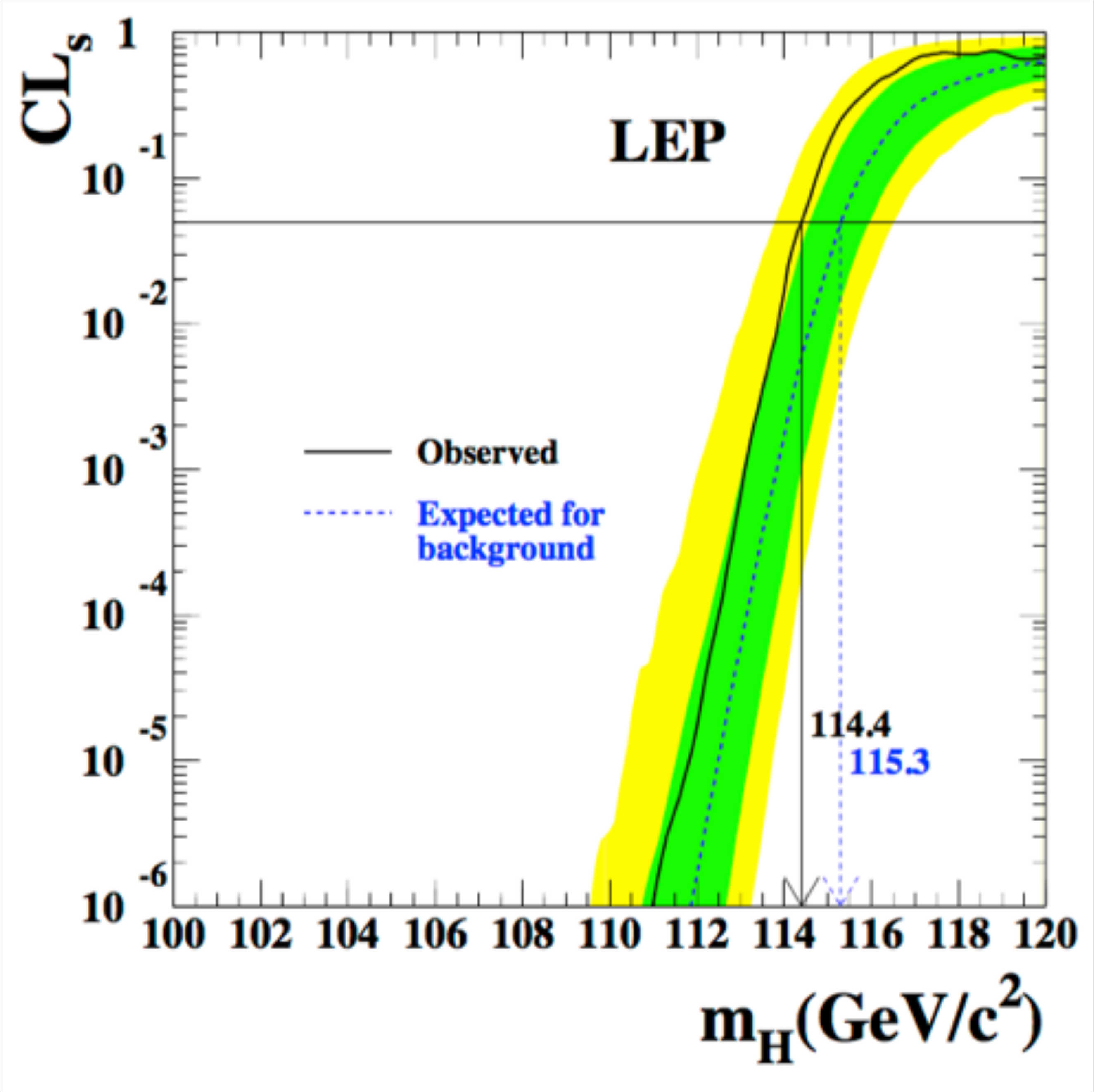
LEP results



-2 ln(Q)
LEP
@ m_H =115 GeV
Observed
Expected for background
Expected for signal plus background
$m_H(GeV/c^2)$

ADLO results



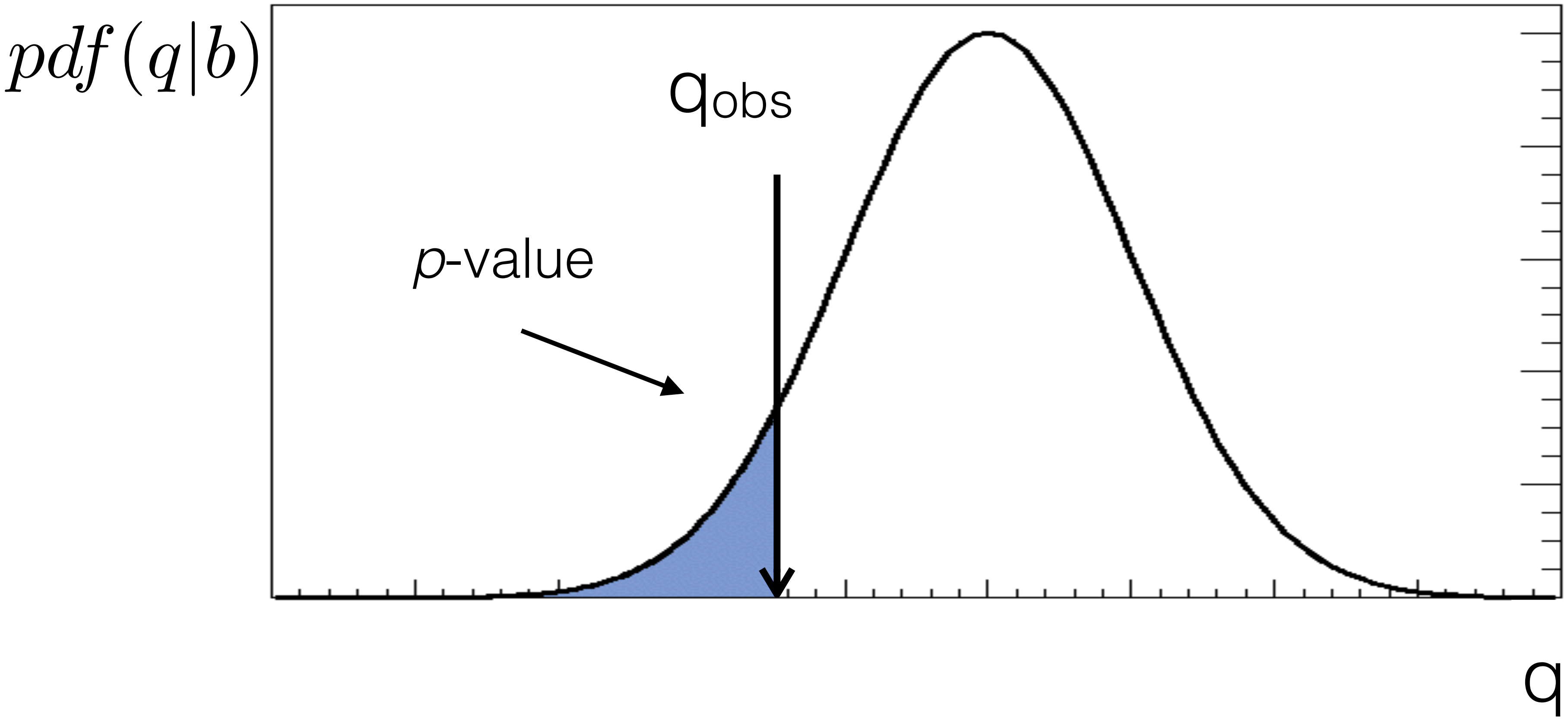


CL_s
LEP
Observed
Expected for background
114.4
115.3
$m_H(GeV/c^2)$

CLs @ LEP

p-values

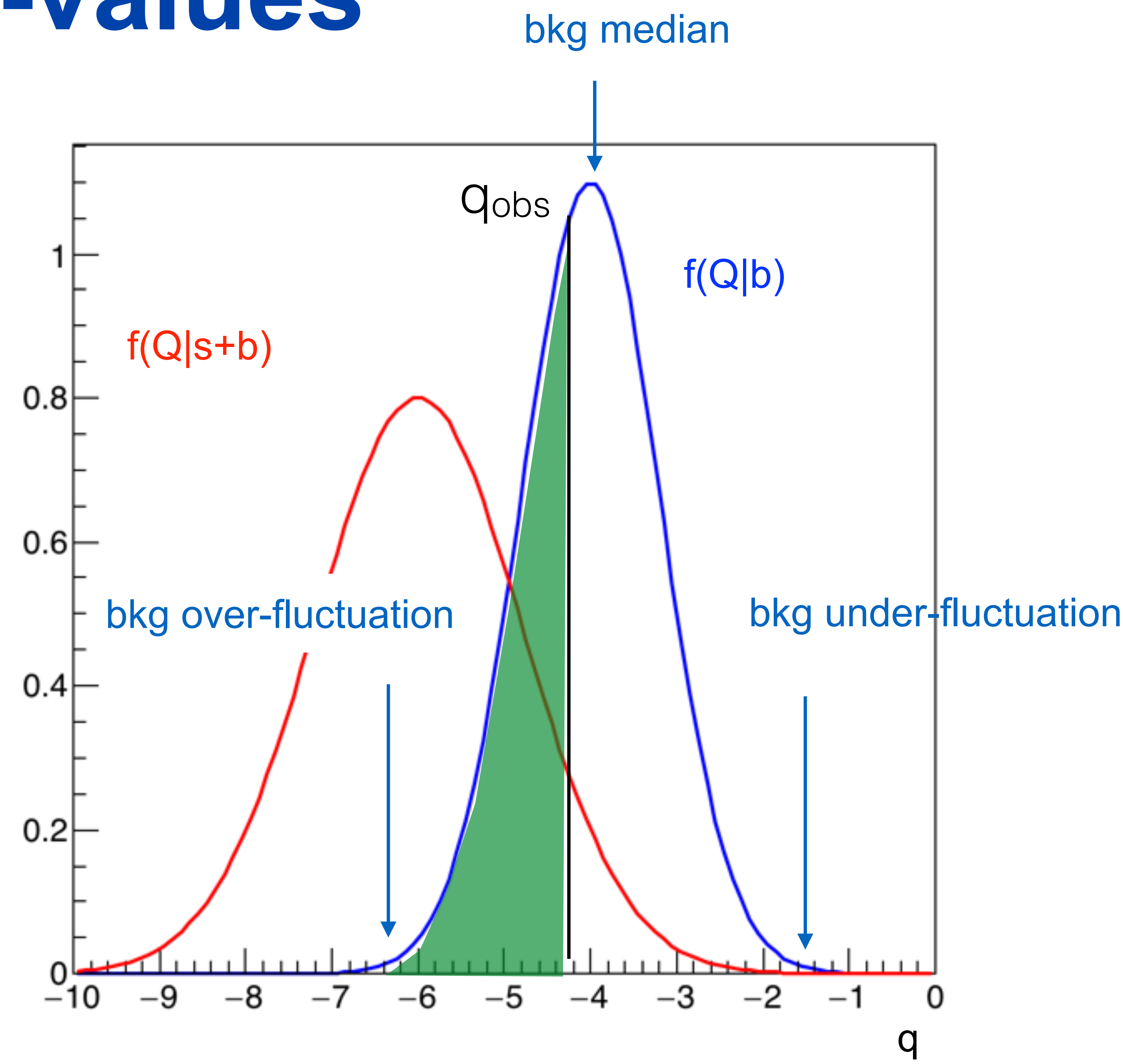
p-value = *probability, under the assumption of H, to observe data with equal or lesser compatibility with H relative to the data we got*



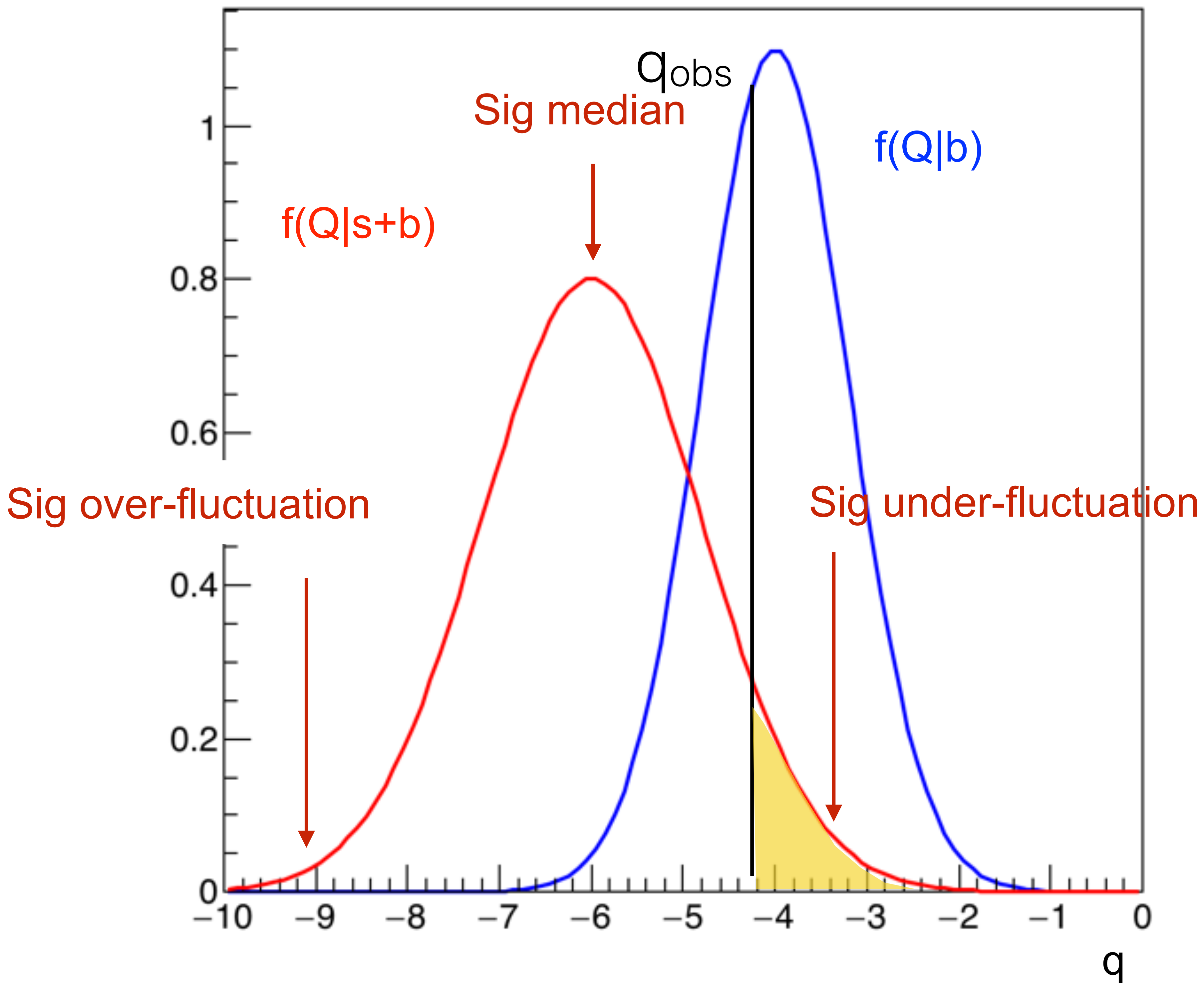
In the HEP folklore we claim (on the background-only hypothesis):

- observation if the p-value < $1.4 \; 10^{-3}$ (3σ)
- discovery if the p-value < $2.9 \; 10^{-7}$ (5σ)

more *p*-values



more *p*-values

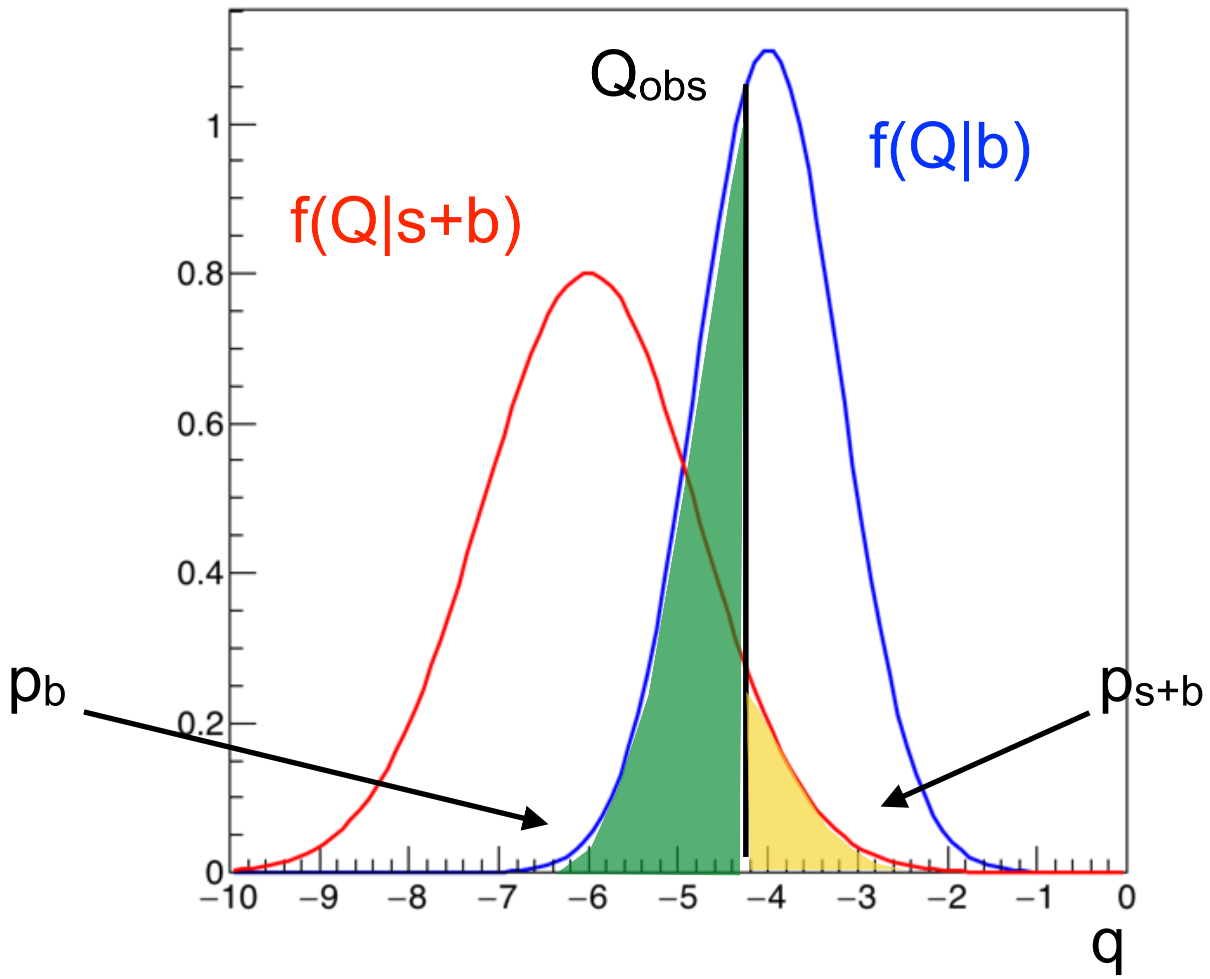


more *p*-values

The p-value is a general way to quantify a deviation.
We usually define :

p_b deviation defined on the background hypothesis

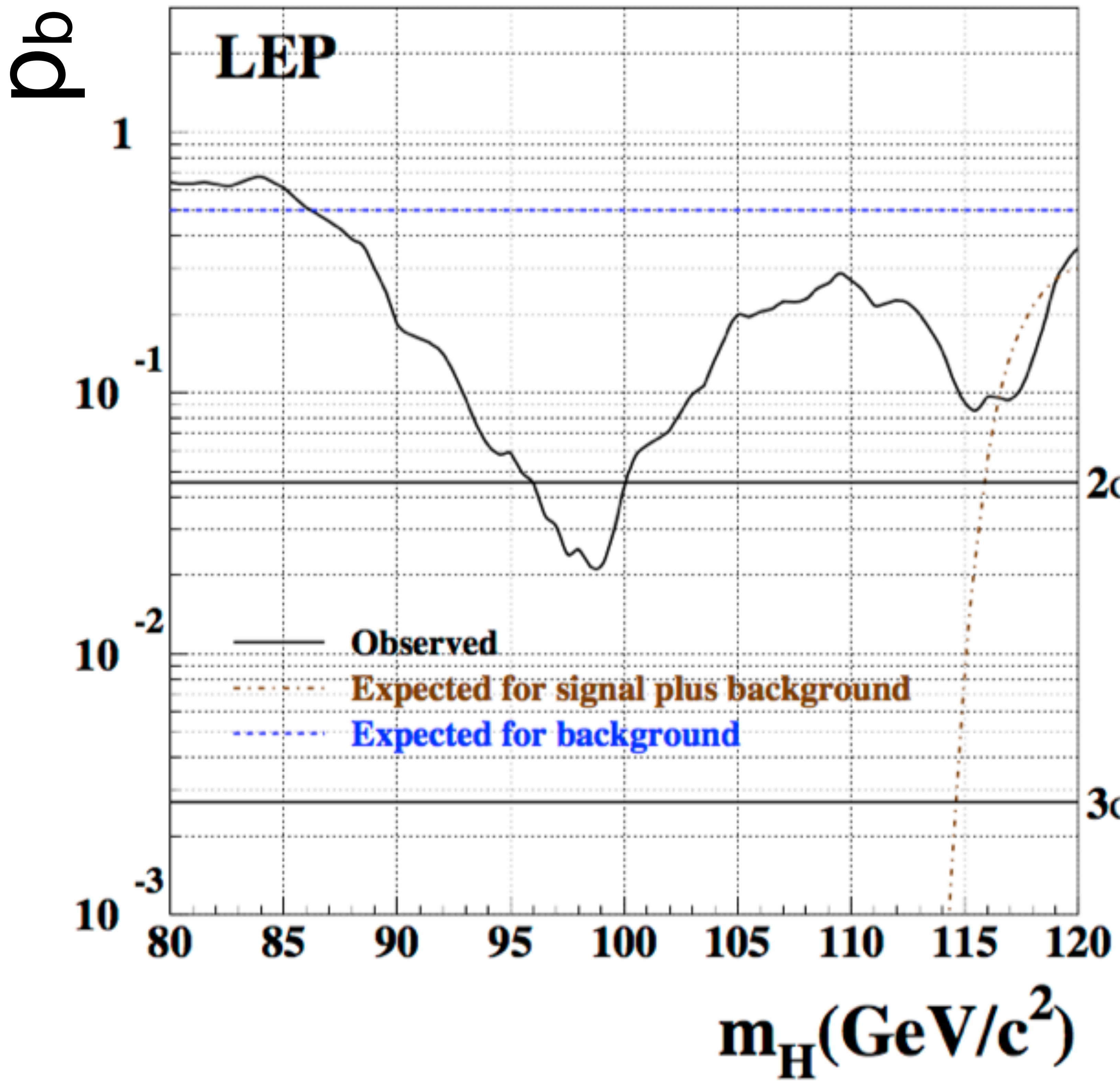
p_{s+b} deviation defined on the signal + background hypothesis



If $p_{s+b} < \alpha$ (computed on f(Q|s+b)) we reject the s+b hp at 1-α CL —> 5 sigmas

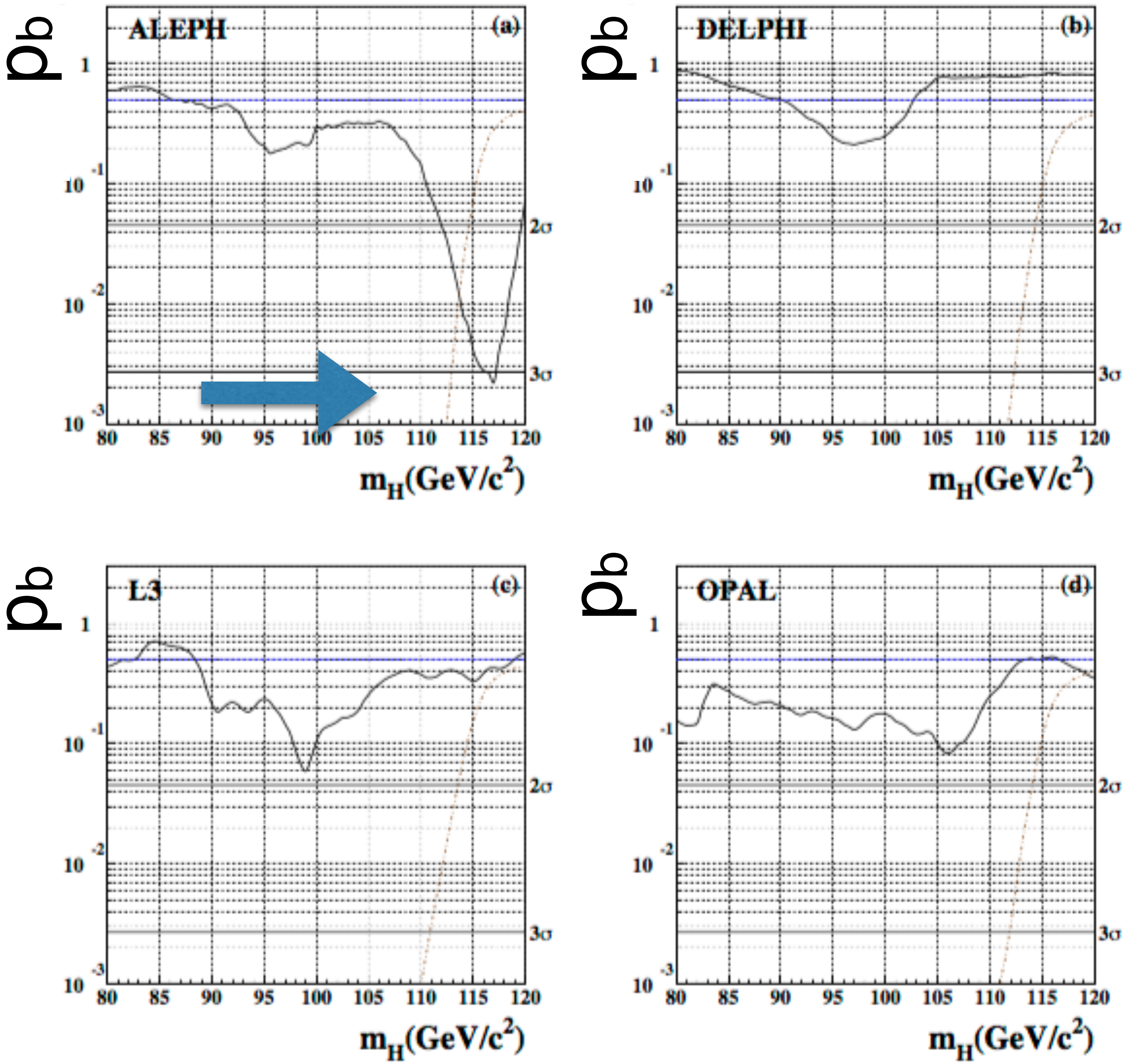
If $p_b < \alpha$ (computed on f(Q|b)) we reject the b hp at 1-α CL —> 2 sigmas

LEP results



50%
background only

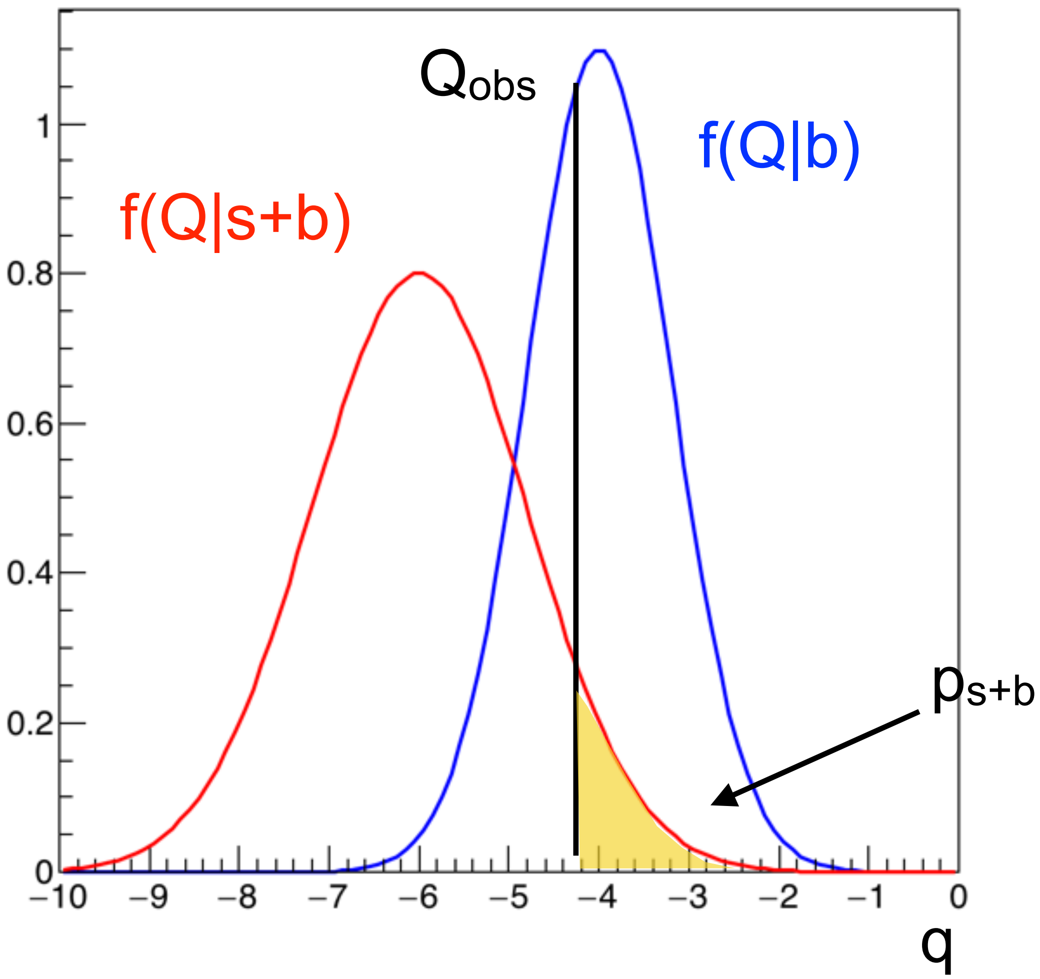
LEP results



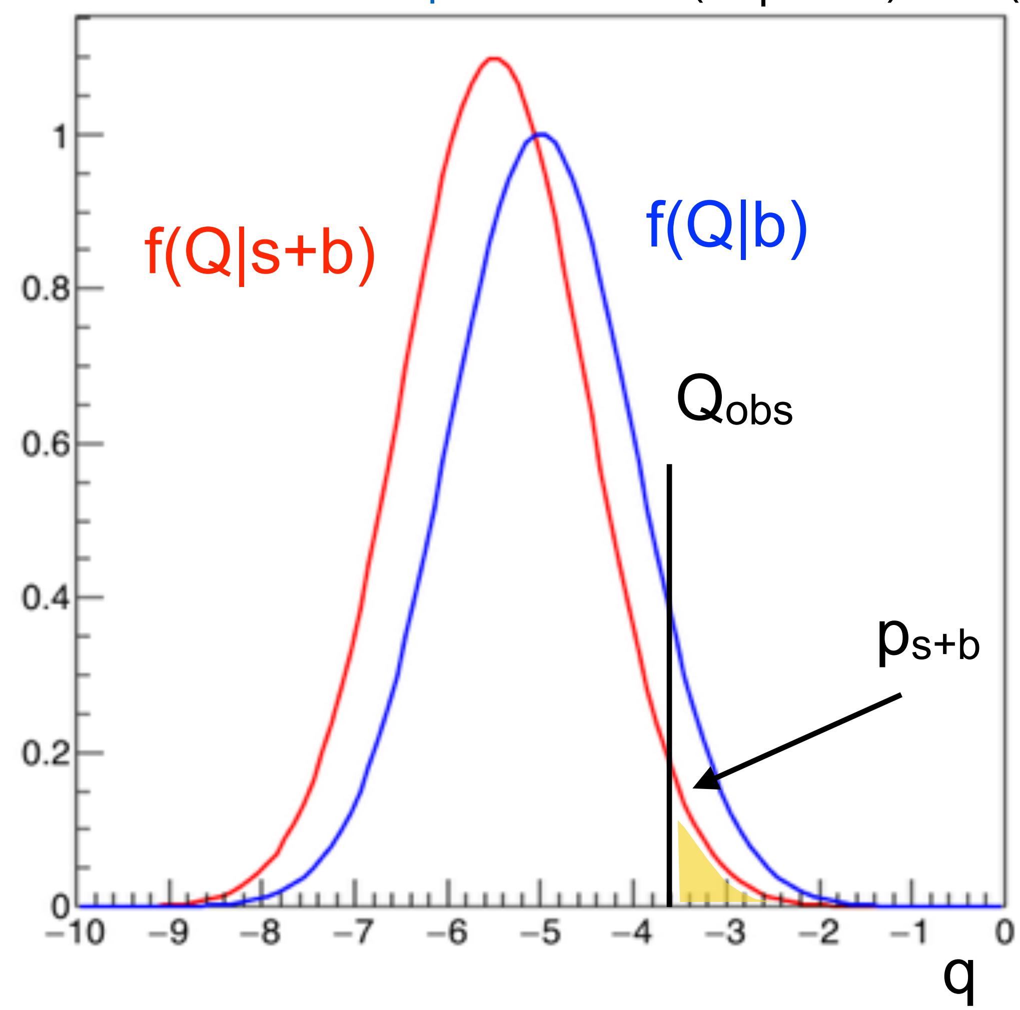
ALEPH
(a)
DELPHI
(b)
L3
(c)
OPAL
(d)
p_b
1
10^{-1}
10^{-2}
10^{-3}
2σ
3σ
80 85 90 95 100 105 110 115 120
$m_H(GeV/c^2)$

The issue about sensitivity

Well separated



NOT well separated f(Q|s+b) ~ f(Q|b)



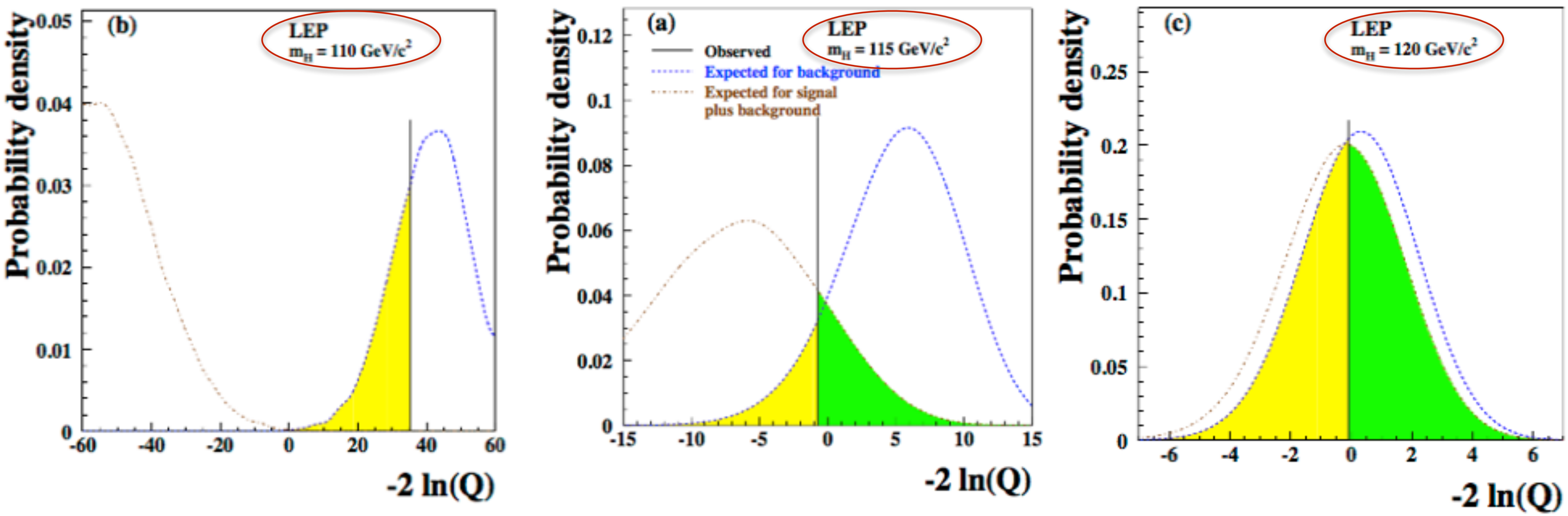
The small separation power means low sensitivity to the signal:
Suppose you look at LEP for a Higgs at 125 GeV (signal not kinematically accessible): whatever test statistics you use to discriminate s+b from b will look practically identical for both cases.

If the number of events fluctuates below the expected bkg,both s+b and b are disfavoured. But, given the low p_{s+b} (e.g. 5%) one can exclude the s+b at 95% CL, which is wrong because you don't have sensitivity !

We don't want to exclude a signal that we are not sensitive to !
We have to include some information about p_b

The issue about sensitivity

Here we use the same data and build test hypothesis changing the test mass m_H.



The higher the mass the lower the power of the test the lower the sensitivity

CLs method

$$CL_s \equiv CL_{s+b}/CL_b$$

Vocabulary A. Read: “confidence level = p-value”

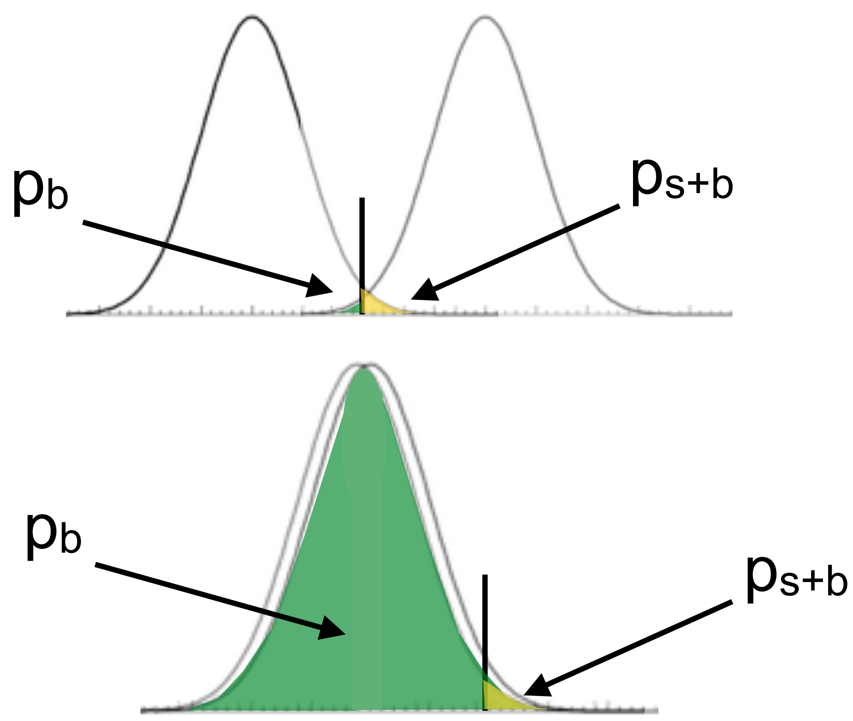
The CLs is NOT a p-value (it’s a ratio of p-values) NEVERTHELESS we will say that

a signal will be considered excluded at the confidence level CL if $1\text{-}CL_s \leq CL$

Consequences:

1) the false exclusion rate is less than the nominal 1-CL

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1-p_b}$$



In case of clear background exclusion

$p_b \rightarrow 0$ then $CL_s \rightarrow p_{s+b}$

In case of no separation

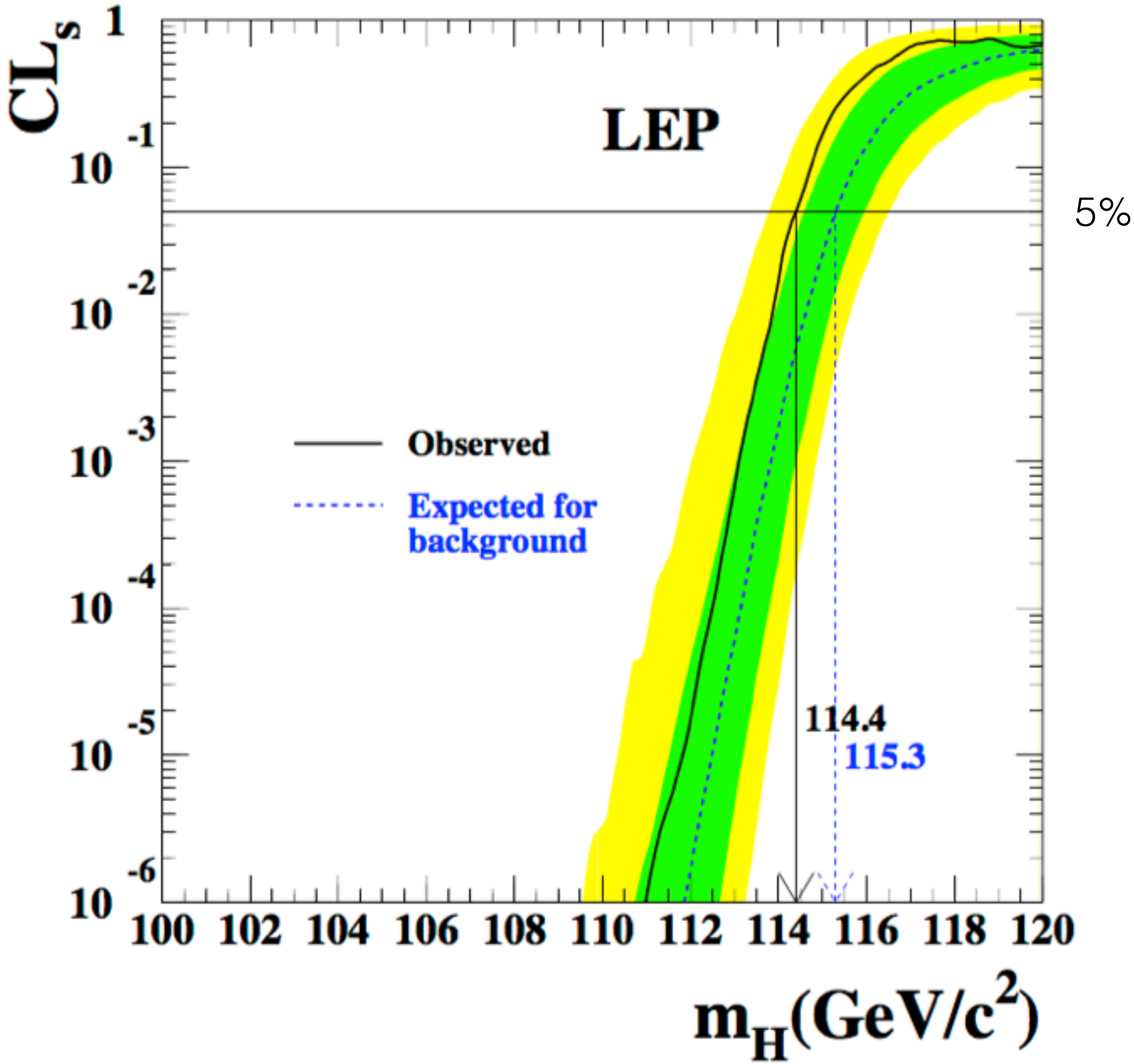
$p_{s+b} \rightarrow 1\text{-}p_b$ $CL_s \rightarrow 1$

i.e. the difference between CL and CLs will increase the lower the sig/bkg separation

2) the use of CL_s increases the “coverage” of the analysis

(i.e. at a given CL you exclude a smaller region of the space of parameters)

Final LEP exclusion



Bibliography

The searches for Higgs Bosons at LEP
M. Kado and C. Tully, Annu. Rev. Nucl. Part. Sci. 2002. 52:65-113

LEP Working group for Higgs Boson Searches: ADLO collaborations
Phys. Lett. B565, 61-75 (2003). arXiv:hep-ex/0306033

Designing and building LEP
K. Hubner, Physics reports 403-404 (2004) 177-188

The Higgs Hunter's Guide
by J.F. Gunion, H. Haber, G. Kane, S.Dawson