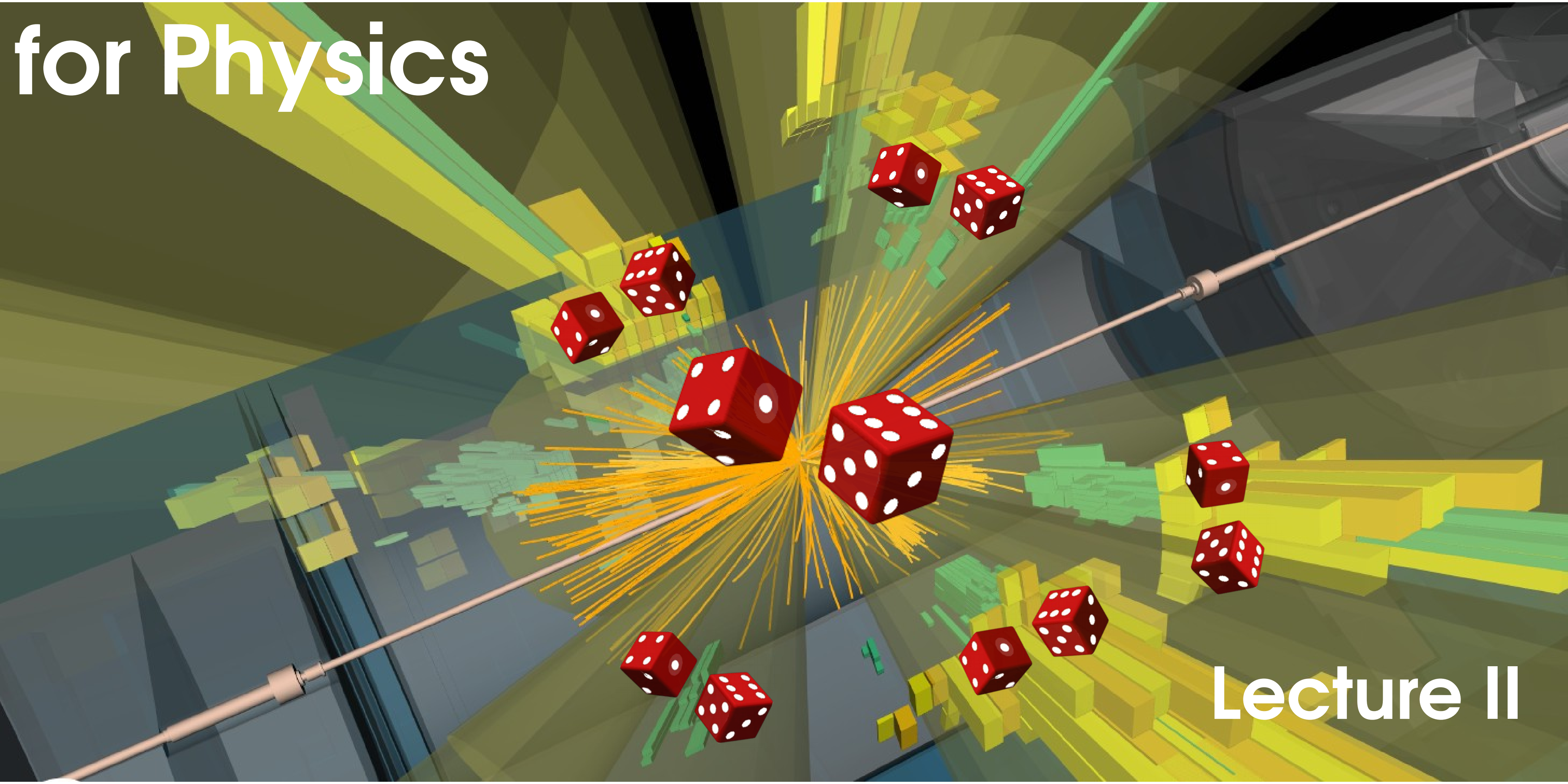
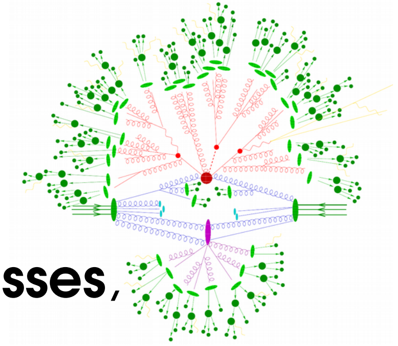
Statistical analysis methods for Physics



Nicolas Berger (LAPP Annecy)

Reminders From Lecture I



Physics measurement data are produced through **random processes**,
Need to be described using a statistical model:

Description	Observable	Likelihood
Counting	n	**Poisson** $P(n;S,B)=e^{-(S+B)}\frac{(S+B)^{n}}{n!}$
Binned shape analysis	n_i, i=1..N_{bins}	**Poisson product** $P(n_i;S,B)=\prod_{i=1}^{n_{\text{bins}}} e^{-(S f_i^{\text{sig}}+B f_i^{\text{bkg}})}\frac{(S f_i^{\text{sig}}+B f_i^{\text{bkg}})^{n_i}}{n_i!}$
Unbinned shape analysis	m_i, i=1..n_{evts}	**Extended Unbinned Likelihood** $P(m_i;S,B)=\frac{e^{-(S+B)}}{n_{\text{evts}}!}\prod_{i=1}^{n_{\text{evts}}} S\,P_{\text{sig}}(m_i)+B\,P_{\text{bkg}}(m_i)$

Model can include multiple **categories**, each with a separate description

Includes **parameters of interest** (POIs) but also **nuisance parameters** (NPs)

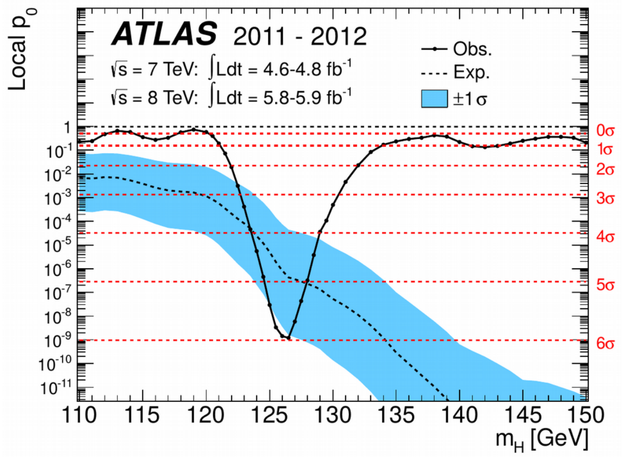
Reminders From Lecture I

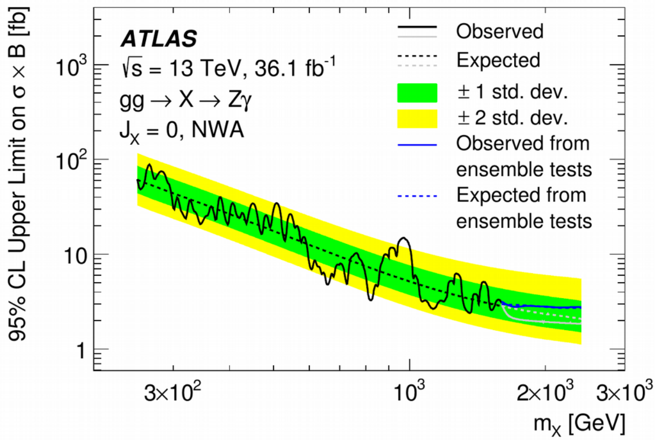
To **estimate a parameter value**, use the Maximum-likelihood estimate (MLE), a.k.a. Best-fit value of the parameter,

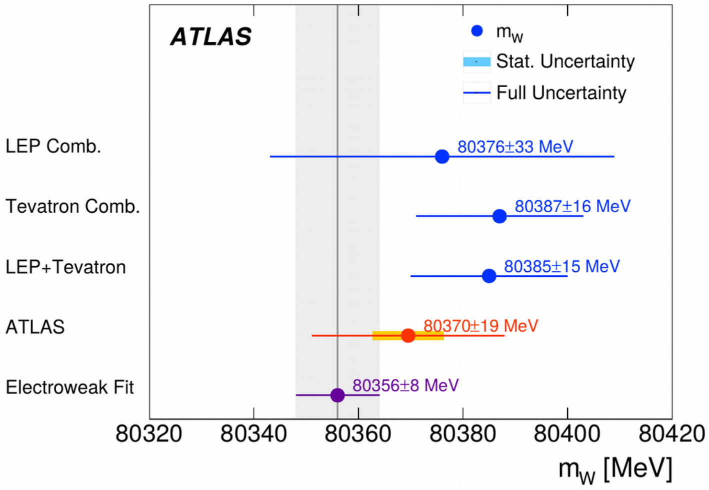
Today, further results:

- **Discovery**: we see an excess – is it a (new) signal, or a background fluctuation ?
- **Upper limits**: we don't see an excess – if there is a signal present, how small must it be ?
- **Parameter measurement:** what is the allowed range ("confidence interval") for a model parameter ?

→ The Statistical Model already contains all the needed information – how to use it ?







Outline

Lecture I:

- Statistics basics
- Describing measurements
- Computing statistics results:

Today:

- **Computing statistics results:**
  - Discovery
  - Limits
  - Confidence intervals
- **Profiling**

Lecture III: Look-Elsewhere Effect, Bayesian methods Practical modeling, BLUE

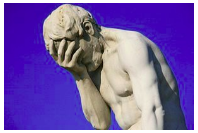
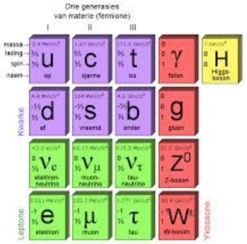
Computing Statistical Results

II. Testing Hypotheses

Hypothesis Testing

Hypothesis: assumption on model parameters, say value of S (e.g. **H_0 : S=0**)

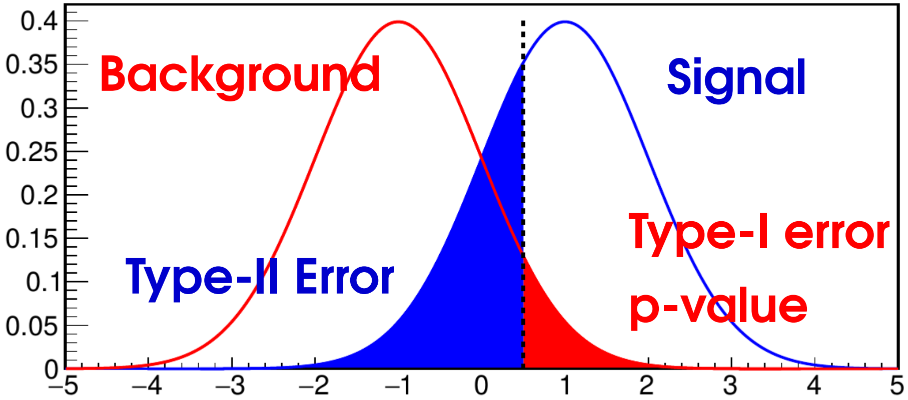
→ Goal : determine if H_0 is true or false using a test based on the data

Possible outcomes:	**Data disfavors H_0** **(Discovery claim)**	**Data favors H_0** **(Nothing found)**
H_0 is false **(New physics!)**	**Discovery!** 	**Missed discovery** **Type-II error** **(1 - Power)** 
H_0 is true **(Nothing new)**	**False discovery claim** **Type-I error** **(→ p-value, significance)** 	**No new physics,** **none found** 

Stringent discovery criteria

⇒ **lower Type-I errors**, **higher Type-II errors**

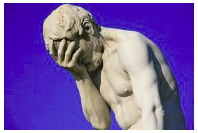
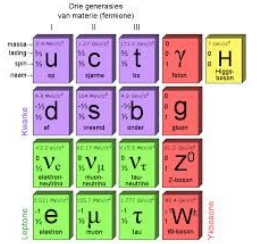
→ **Goal**: test that minimizes Type-II errors **for given level of Type-I error**.



Hypothesis Testing

Hypothesis: assumption on model parameters, say value of S (e.g. **H_0 : S=0**)

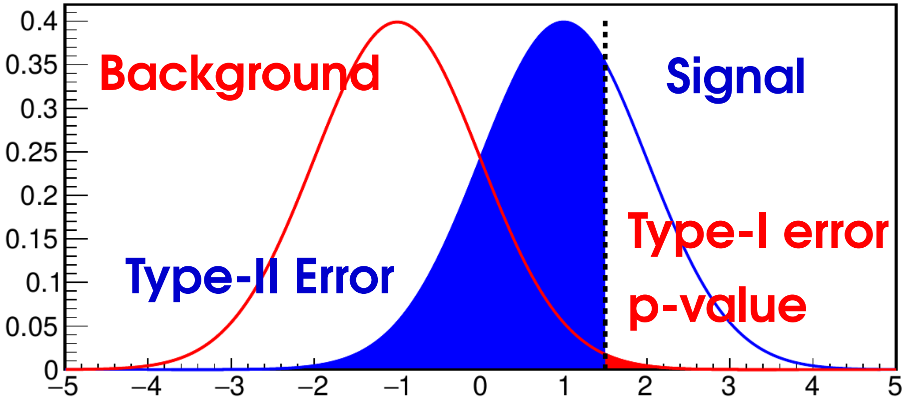
→ Goal : determine if H_0 is true or false using a test based on the data

Possible outcomes:	**Data disfavors H_0** **(Discovery claim)**	**Data favors H_0** **(Nothing found)**
H_0 is false **(New physics!)**	**Discovery!** 	**Missed discovery** **Type-II error** **(1 - Power)** 
H_0 is true **(Nothing new)**	**False discovery claim** **Type-I error** **(→ p-value, significance)** 	**No new physics,** **none found** 

Stringent discovery criteria

⇒ **lower Type-I errors**, **higher Type-II errors**

→ **Goal**: test that minimizes Type-II errors **for given level of Type-I error**.



Hypothesis Testing with Likelihoods

Neyman-Pearson Lemma

When comparing two hypotheses H_0 *and* H_1*, the optimal discriminator is the* ***Likelihood ratio*** *(LR)*

$$\frac{L(H_1;data)}{L(H_0;data)}$$

As for MLE, choose the hypothesis that is more likely **for the data**.

→ **Minimizes Type-II uncertainties** for given level of Type-I uncertainties
→ Always need an **alternate hypothesis** to test against.

Caveat: Strictly true only for *simple hypotheses* (no free parameters)

→ **In the following**: all tests based on LR, will focus on p-values (Type-I errors), trusting that Type-II errors are anyway as small as they can be...

Statistical Results as Hypothesis Tests

Usual HEP results can be recast in terms of **hypothesis testing**:

- **Discovery**: is the data compatible with background-only ?
  → H_0 : only background is present
  → How well can we **reject** H_0 ? → **p-value (significance)**

- **Upper limits**: no excess observed – how small must the signal be ?
  → $H_0(S)$: B + some signal S
  → How small can we make S, and still reject $H_0(S)$ at 95% C.L. (p=5%) ?

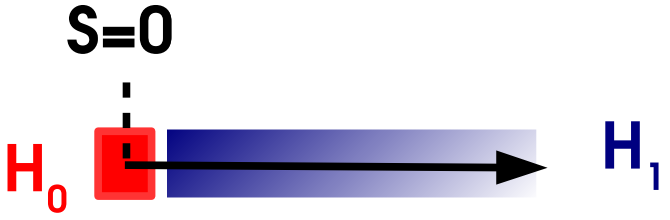
- **Parameter measurement**
  → $H_0(\mu)$: some parameter value μ
  → What values μ are **not** rejected at 68% C.L. (p=32%) ?
  ⇒ **1σ confidence interval on μ**

In all cases, H_0 : ***null hypothesis*** – what we are trying to disprove

Computing Statistical Results
III. Discovery

Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

Discovery: Test Statistic



Discovery :

- **H_0 : background only (S = 0)** against
- **H_1**: presence of a signal (**S ≠ 0**)

→ For H_1, any S≠0 is possible, which to use ? **The one preferred by the data, Ŝ.**

⇒ Use LR $\dfrac{L(S=0)}{L(\hat{S})}$

→ In fact use the **test statistic** $$t_0 = -2\log\frac{L(S=0)}{L(\hat{S})}$$

→ t_0 is computed from the observed data – fit to data to get Ŝ.

→ **t_0 always ≥ 0**, $t_0 = 0$ reached for $\hat{S} = 0$.

→ t_0 measures the relative *likelihood* of H_1 vs. H_0 in data:

Large values of t_0 ⇔ large observed S

Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

Discovery p-value

Large values of $t_0 = -2\log\frac{L(S=0)}{L(\hat{S})}$

⇒ large observed $\hat{S}$

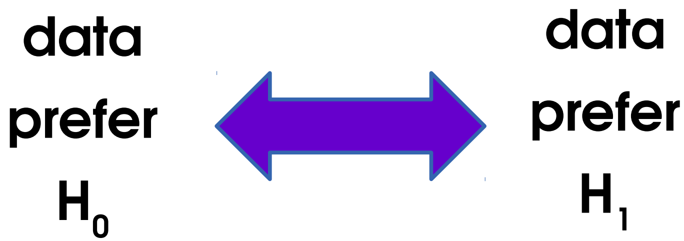
⇒ **$H_0(S=0)$** ***disfavored*** compared to **$H_1(S\neq 0)$**.

How large t_0 before we can exclude H_0 ?
(and claim a discovery!)

p-value **:** *Fraction of outcomes that are* ***at least as H_1-like*** *(signal-like)* ***as data****,*
when ***H_0 is true*** *(no signal present).*

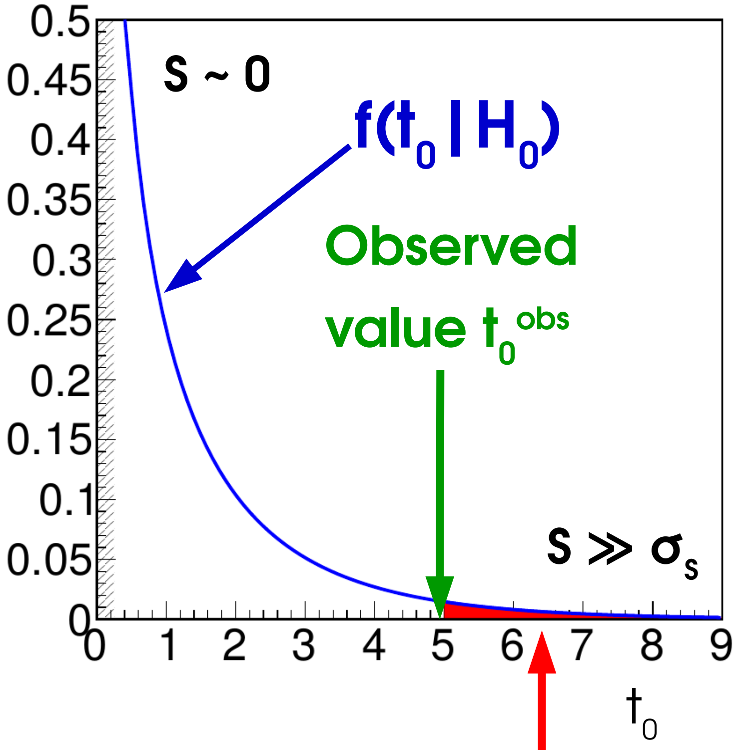
→ Smaller p-value ⇒ Stronger case for discovery

→ Compute from distribution **$f(t_0|H_0)$** of t_0 if H_0 is true:



data prefer H_0

data prefer H_1

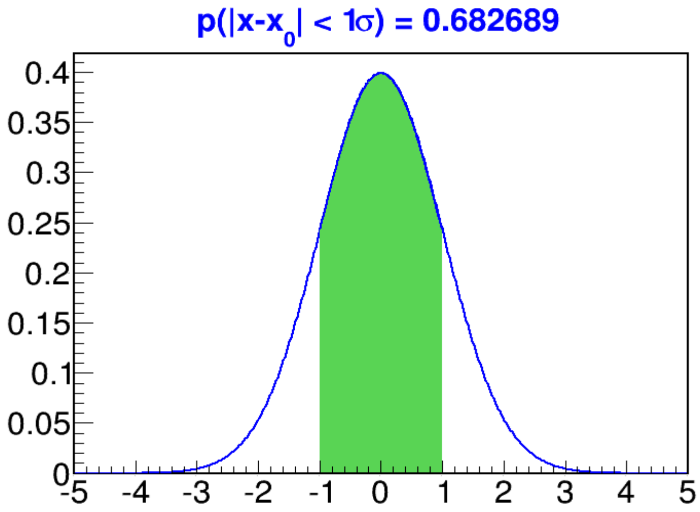


$$p_0 = \int_{t_0^{obs}}^{\infty} f(t_0|H_0)\, dt_0$$

Discovery significance

Interesting p-values are quite small
⇒ express in terms of Gaussian quantiles
→ Significance Z

$$p_0 = 1 - \int_{-Z}^{+Z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$
$$= 1 - 2\,\Phi(Z)$$



$$\Phi(Z) = \int_{-\infty}^{Z} G(u;0,1)\,du$$

Z	p-value
1	0.32
2	0.045
3	0.003
5	6×10^{-7}

In ROOT:
p_0 → Z **(Φ)** : `ROOT::Math::gaussian_quantile_c`
Z → p_0 **(Φ^{-1})** : `ROOT::Math::gaussian_cdf_c`

⇒ How small is small enough ?
→ Conventionally, discovery for **$p_0 = 6\ 10^{-7}$ ⇔ $Z = 5\sigma$**

Asymptotic Approximation: Wilks' Theorem

Cowan, Cranmer, Gross & Vitells
Eur.Phys.J.C71:1554,2011

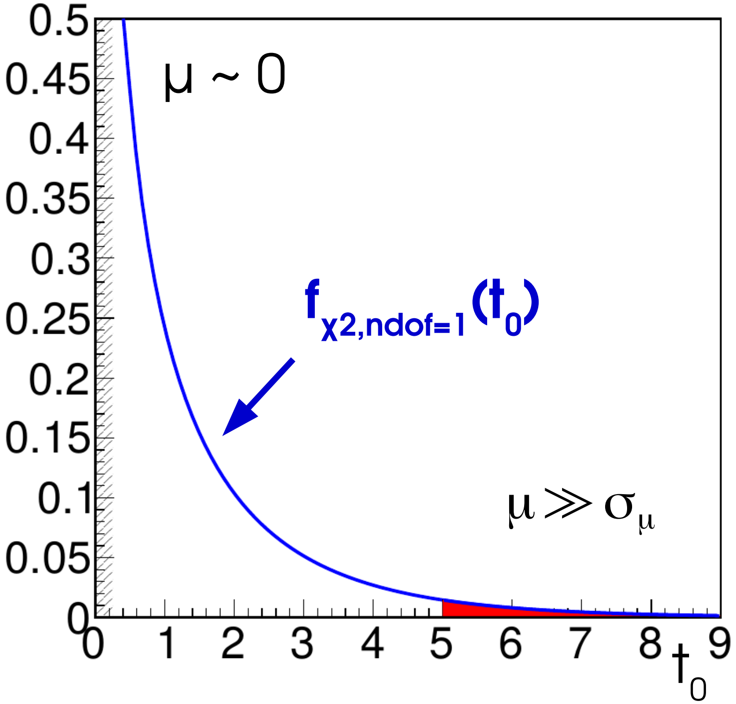
→ Assume **Gaussian regime for $\hat{S}$** (e.g. large n_{evts})

⇒ Central-limit theorem :

t_0 is distributed as a χ^2 under the hypothesis H_0

$$t_0 = -2\log\frac{L(S=0)}{L(\hat{S})}$$

$$f(t_0 \mid H_0) = f_{\chi^2(n_{dof}=1)}(t_0)$$



In particular, significance:

$$Z = \sqrt{t_0}$$

By definition,
$t_0 \sim \chi^2 \Rightarrow \sqrt{t_0} \sim G(0,1)$

Typically works well for for event counts O(5) and above (5 already "large"...)

The 1-line "proof" : asymptotically L and S are Gaussian, so

$$L(S) = \exp\left[-\frac{1}{2}\left(\frac{S-\hat{S}}{\sigma}\right)^2\right] \Rightarrow t_0 = \left(\frac{\hat{S}}{\sigma}\right)^2 \Rightarrow t_0 \sim \chi^2(n_{dof}=1) \text{ since } \hat{S} \sim G(0,\sigma)$$

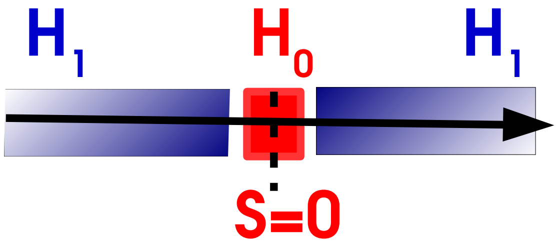
One-sided vs. Two-Sided

If **$\hat{S} < 0$**, is it a *discovery* ? (does reject the S=0 hypothesis…)

Usual assumption : only **$\hat{S} > 0$** is a *bona fide* signal

⇒ Change statistic so that **$\hat{S} < 0$** ⇒ **$t_0 = 0$** (perfect agreement with H_0, as for $\hat{S} = 0$)

Two-sided



$$t_0 = -2\log\frac{L(S=0)}{L(\hat{S})}$$

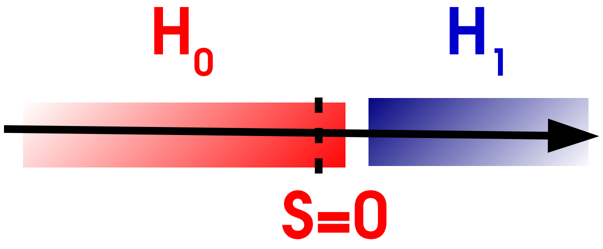
$$Z = \Phi^{-1}\left(1 - \frac{p_0}{2}\right)$$

By convention, factor 2 in p-values for a given Z

Test Statistic

p_0	Z	p_0
0.32	1	0.16
0.003	3	0.0015
6×10^{-7}	5	3×10^{-7}

One-sided



$$q_0 = \begin{cases} -2\log\dfrac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ 0 & \hat{S} < 0 \end{cases}$$

$$Z = \Phi^{-1}(1 - p_0)$$

⇒ Same Z in both cases for a given signal S

One-Sided Asymptotics

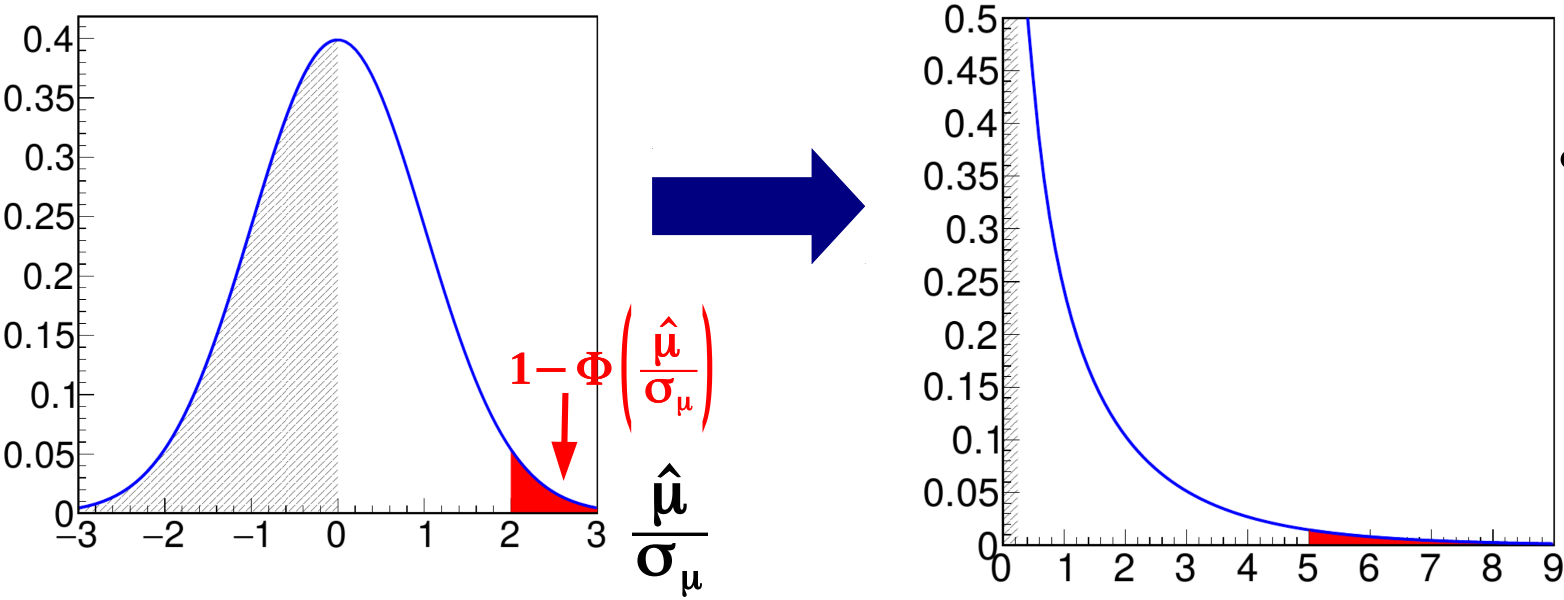
→ **One-sided test**:

$$q_0 = \begin{cases} -2\log\dfrac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ 0 & \hat{S} < 0 \end{cases}$$

Asymptotics: "half-χ^2" distribution:

$$f(q_0 \mid S=0) = \frac{1}{2}\delta(q_0) + \frac{1}{2} f_{\chi^2(n_{dof}=1)}(q_0)$$

Discovery p-value: $p_0 = 1 - \Phi\left(\sqrt{q_0}\right)$ **Significance**: $Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$

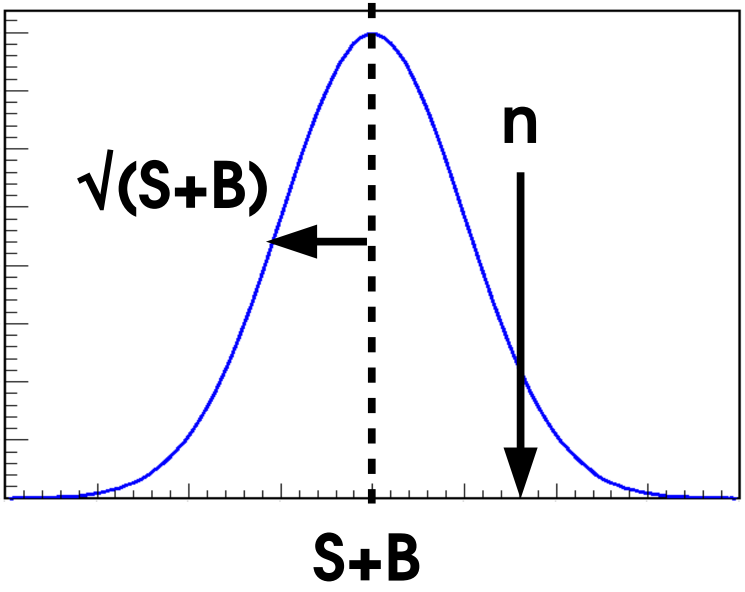


Φ : normal CDF

$$\Phi(z) = \int_{-\infty}^{z} G(u;0,1)\,du$$

$$q_0 = \left(\frac{\hat{\mu}}{\sigma_\mu}\right)^2$$

Example: Gaussian Counting



Count number of events n in data

→ assume n large enough so process is Gaussian

→ assume B is known, measure S

Likelihood :

$$L(S;n) = e^{-\frac{1}{2}\left(\frac{n-(S+B)}{\sqrt{S+B}}\right)^2}$$

$$\lambda(S;n) = \left(\frac{n-(S+B)}{\sqrt{S+B}}\right)^2$$

MLE for S : **Ŝ = n – B**

Test statistic: assume Ŝ > 0,

$$q_0 = -2\log\frac{L(S=0)}{L(\hat{S})} = \lambda(S=0) - \lambda(\hat{S}) = \left(\frac{n-B}{\sqrt{B}}\right)^2 = \left(\frac{\hat{S}}{\sqrt{B}}\right)^2$$

Finally:

$$Z = \sqrt{q_0} = \frac{\hat{S}}{\sqrt{B}}$$

Known formula!

→ Strictly speaking only valid in Gaussian regimge

Example: Poisson Counting

Same problem but now not assuming Gaussianity

$$L(S;n) = e^{-(S+B)}(S+B)^n \qquad \lambda(S;n) = 2(S+B) - 2n\log(S+B)$$

MLE: **Ŝ = n – B**, same as Gaussian

Test statistic (for Ŝ > 0): $q_0 = \lambda(S=0) - \lambda(\hat{S}) = -2\hat{S} - 2(\hat{S}+B)\log\frac{B}{\hat{S}+B}$

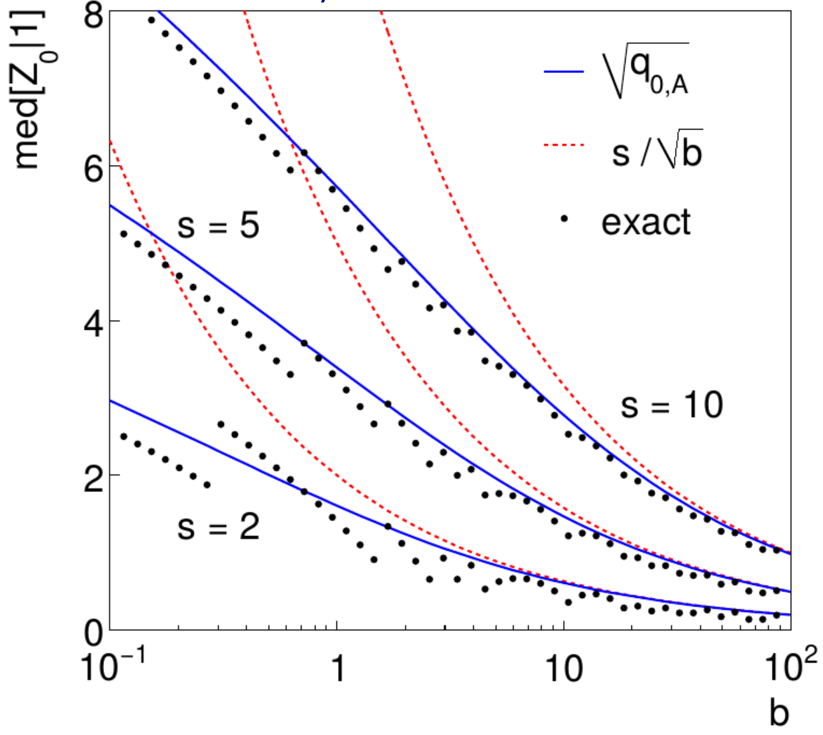
Assuming asymptotic distribution for q_0,

$$Z = \sqrt{2\left[(\hat{S}+B)\log\left(1+\frac{\hat{S}}{B}\right) - \hat{S}\right]}$$

Exact result can be obtained using pseudo-experiments → close to $\sqrt{q_0}$ result

Asymptotic formulas justified by Gaussian regime, but remain valid even for small values of S+B (5!)

Eur.Phys.J.C71:1554,2011

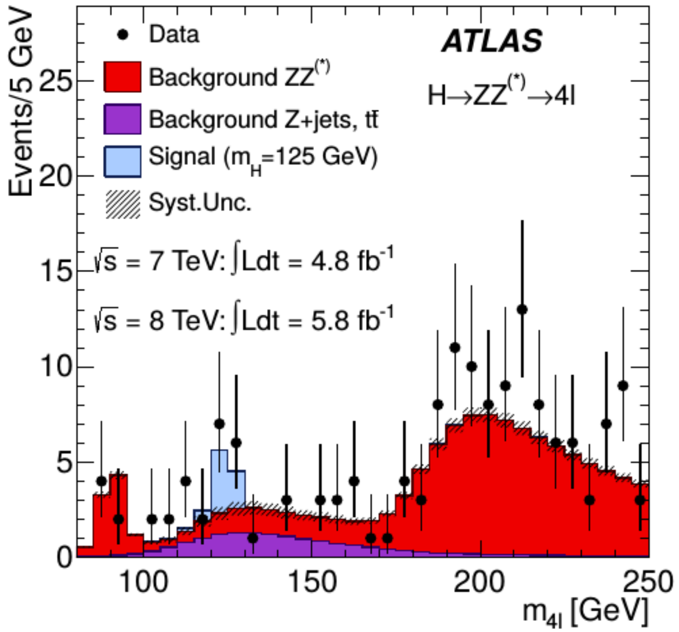


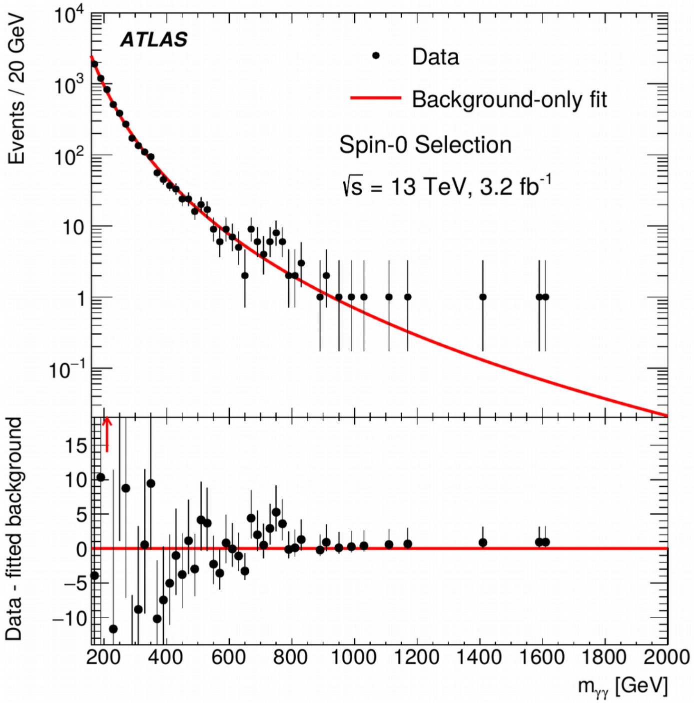
See G. Cowan's slides for case with B uncertainty

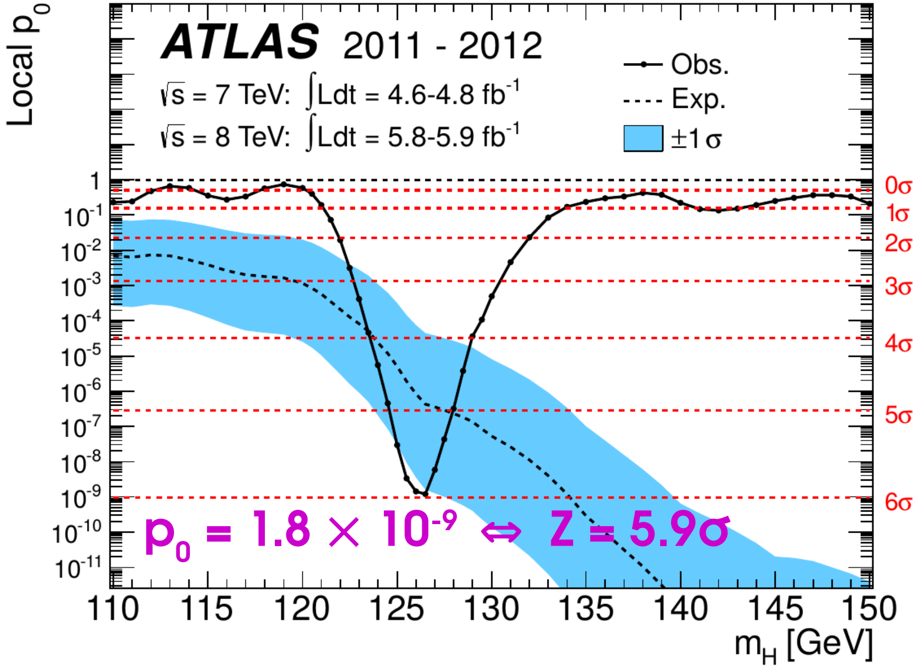
Some Examples

High-mass X→γγ Search: JHEP 09 (2016) 1

Higgs Discovery: Phys. Lett. B 716 (2012) 1-29

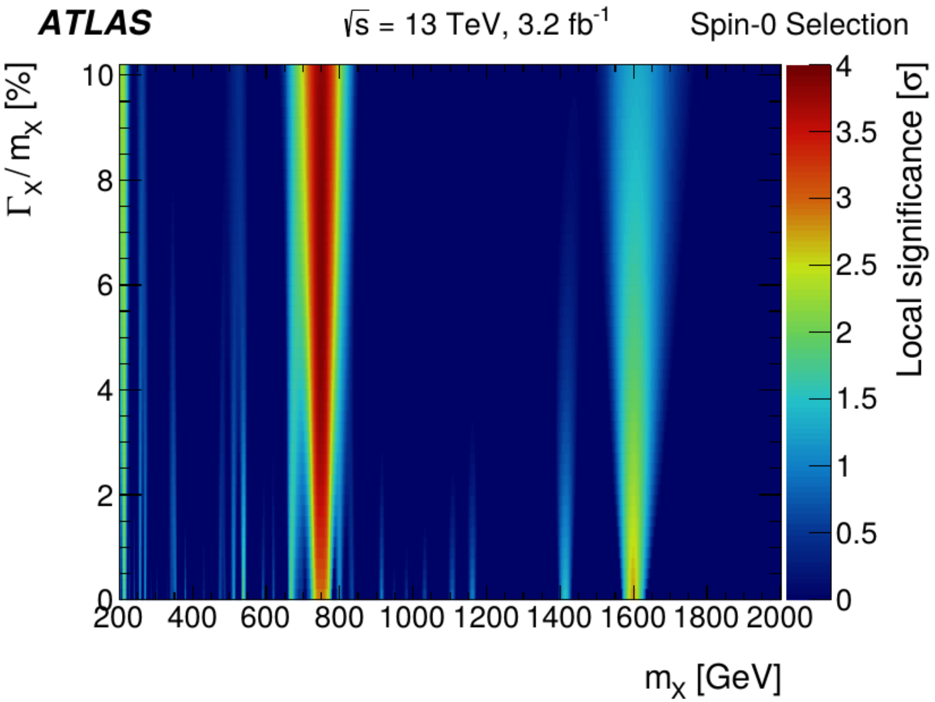






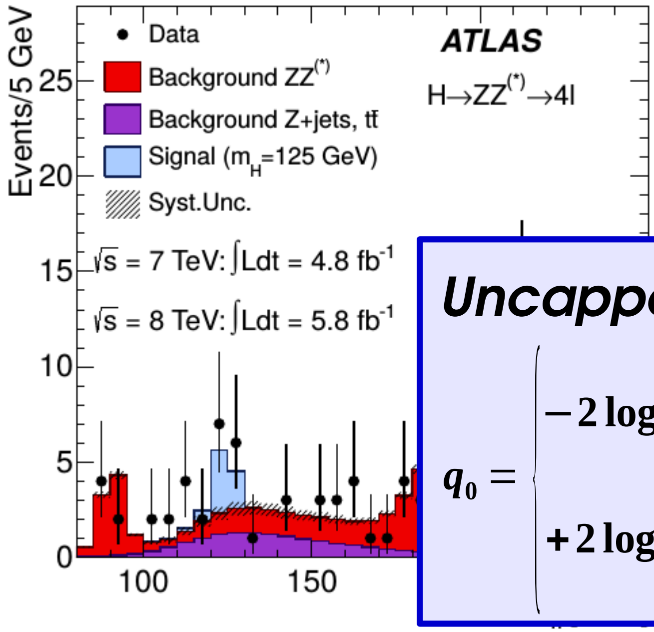
$p_0 = 1.8 \times 10^{-9} \Leftrightarrow Z = 5.9\sigma$

$$Z = \Phi^{-1}(1-p_0) = sgn(q_0)\sqrt{|q_0|}$$



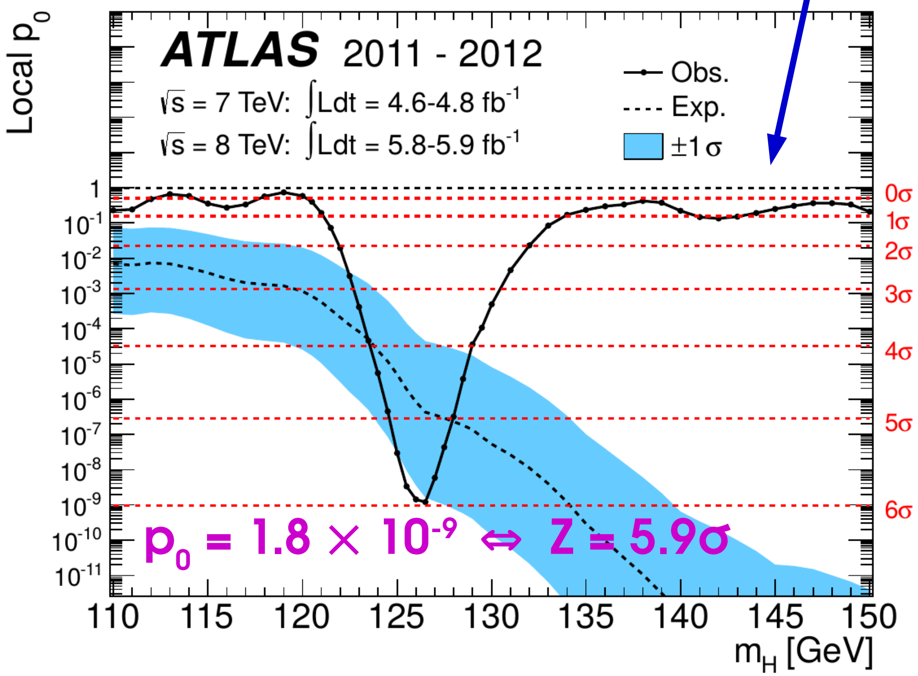
Some Examples

Higgs Discovery: Phys. Lett. B 716 (2012) 1-29



Uncapped q_0:

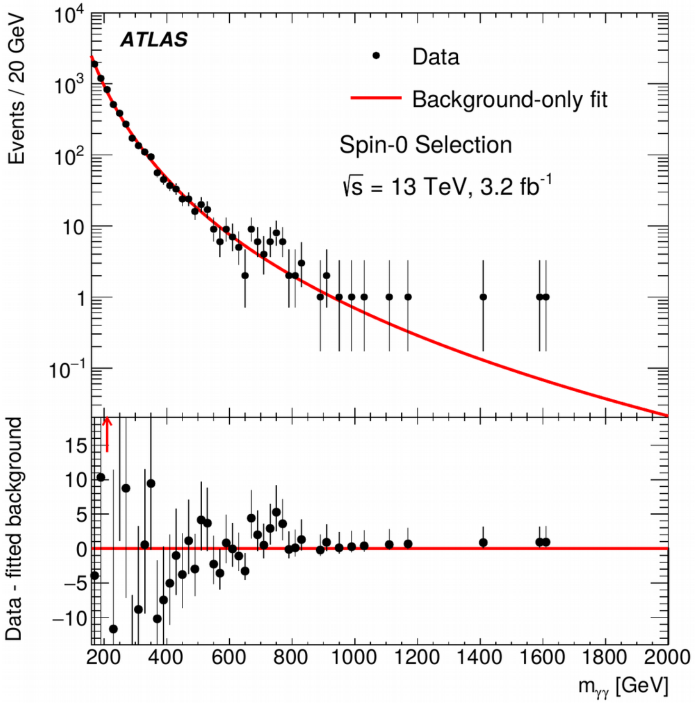
$$q_0 = \begin{cases} -2\log\dfrac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ +2\log\dfrac{L(S=0)}{L(\hat{S})} & \hat{S} < 0 \end{cases}$$

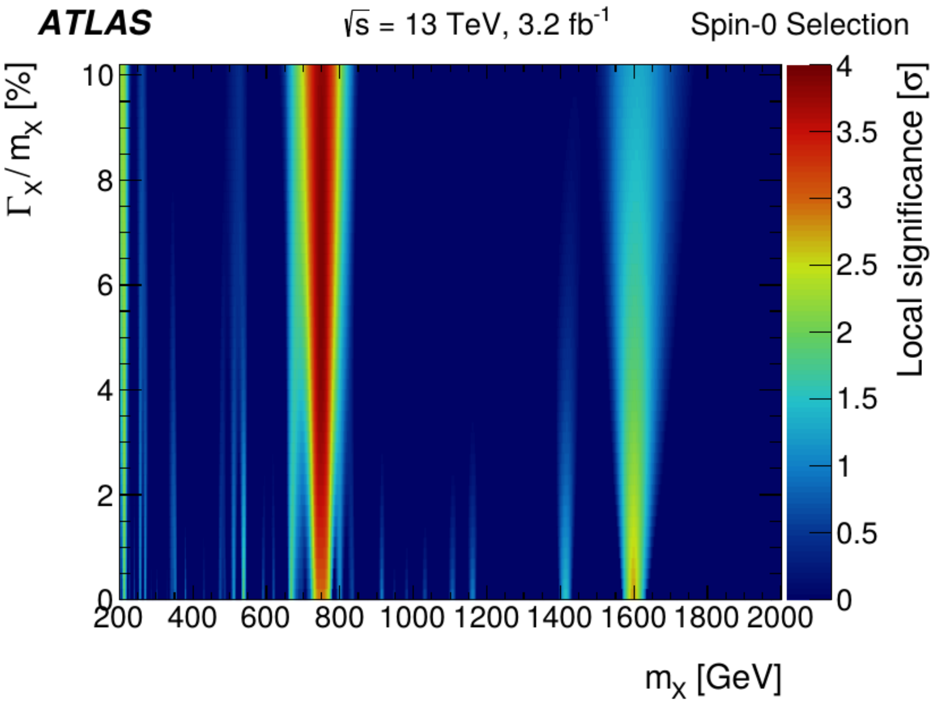


$p_0 = 1.8 \times 10^{-9} \Leftrightarrow Z = 5.9\sigma$

$$Z = \Phi^{-1}(1-p_0) = \mathrm{sgn}(q_0)\sqrt{|q_0|}$$

High-mass X→γγ Search: JHEP 09 (2016) 1





Takeaways

Given a statistical model P(data; μ), define likelihood **L(μ) = P(data; μ)**

To estimate a parameter, use value **μ̂** that maximizes L(μ).

To decide between hypotheses H_0 and H_1, use the **likelihood ratio** $\frac{L(H_0)}{L(H_1)}$

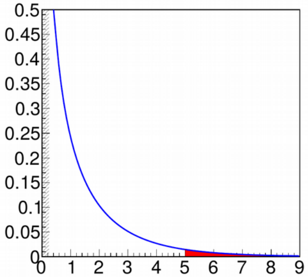
To test for **discovery**, use $q_0 = \begin{cases} -2\log\frac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ +2\log\frac{L(S=0)}{L(\hat{S})} & \hat{S} < 0 \end{cases}$

For large enough datasets (n > 5), $Z = \sqrt{q_0}$

For a **Gaussian** measurement, $Z = \frac{\hat{S}}{\sqrt{B}}$

For a **Poisson** measurement, $Z = \sqrt{2\left[(\hat{S}+B)\log\left(1+\frac{\hat{S}}{B}\right)-\hat{S}\right]}$

What was the question ?



Definition of the p-value:

$$\text{p-value} = \frac{\text{number of signal-like outcomes with only background present}}{\text{all outcomes with only background present}}$$

So **5σ significance (p_0~10^{-7})** ⇔ *Occurs once in 10^7 if only background present*

However this is **NOT** "~~*One chance in 10^7 to be a fluctuation*~~"

The first statement is about **data probabilities – P(data; H_0)**

The second is on **P(H_0)** itself – not addressed in the framework described so far
→ makes sense in a ***Bayesian*** context, more on this later in these lectures.

It's also a different statement (although they sometimes get confused)
→ If a signal outcome is also very unlikely, **we may not want to reject H_0, even with p_0 ~ 10^{-7}.**

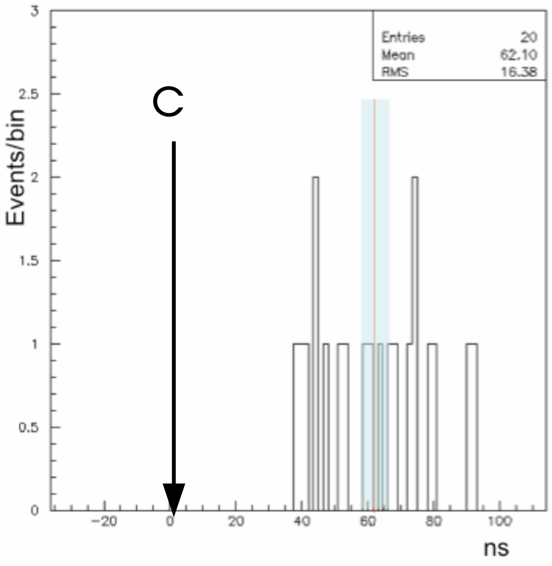
What was the question ?

e.g. Faster-than-light neutrino anomaly

$(v\text{-}c)/c = (2.37 \pm 0.32\ (\text{stat.})\ {}^{+0.34}_{-0.24}\ (\text{sys.}))\times 10^{-5}$ **6.2σ above c**

"***despite the large significance of the measurement*** *reported here and the stability of the analysis,* ***the potentially great impact of the result motivates the continuation of our studies*** *in order to investigate possible still unknown systematic effects that could explain the observed anomaly."*

⇒ Very unlikely to be a background fluctuation, but hard to believe ***since alternative (v>c) is far-fetched***



"Extraordinary claims require extraordinary evidence"

Alternative: $$P(\text{fluctuation}) = \frac{\text{number of signal-like outcomes with only B present}}{\text{number of signal-like outcomes from any source (S or B)}}$$

$$= \frac{P(\text{deviation}|B)\,P(B)}{P(\text{deviation}|S)\,P(S) + P(\text{deviation}|B)\,P(B)}$$

→ Needs ***a priori* P(S) and P(B)** → Bayesian methods, discussed later

→ In frequentist context, only have $\mathbf{p_0 = P(deviation\,|\,B)}$

⇒ **However usually same conclusion, assuming P(S) is not ≪ p_0...**

Outline

Lecture I:

- Statistics basics
- Describing measurements
- Computing statistics results:

Today:

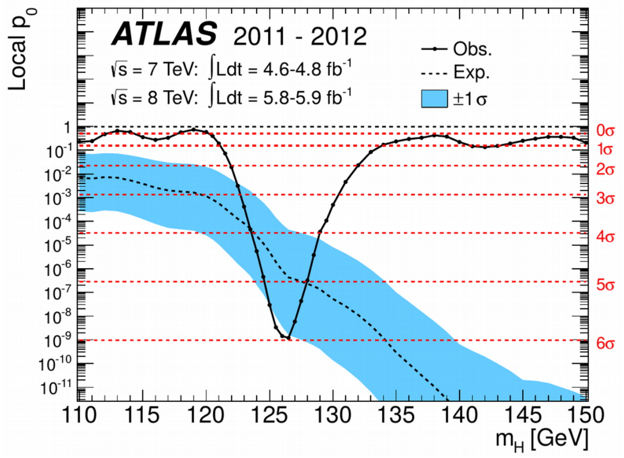
- Computing statistics results:
  - Discovery
  - **Limits**
  - Confidence intervals
- **Profiling**

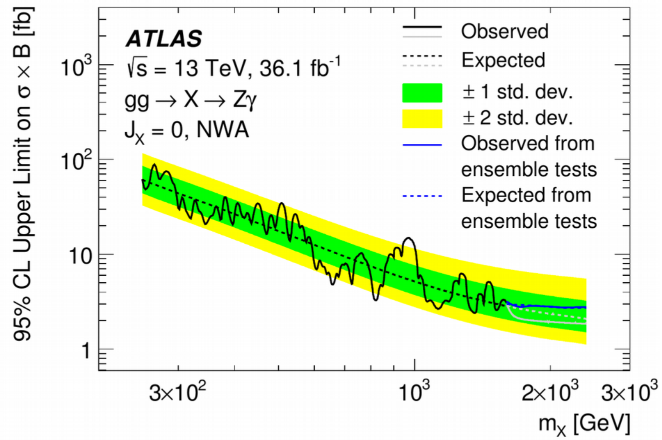
Usual Statistical Results

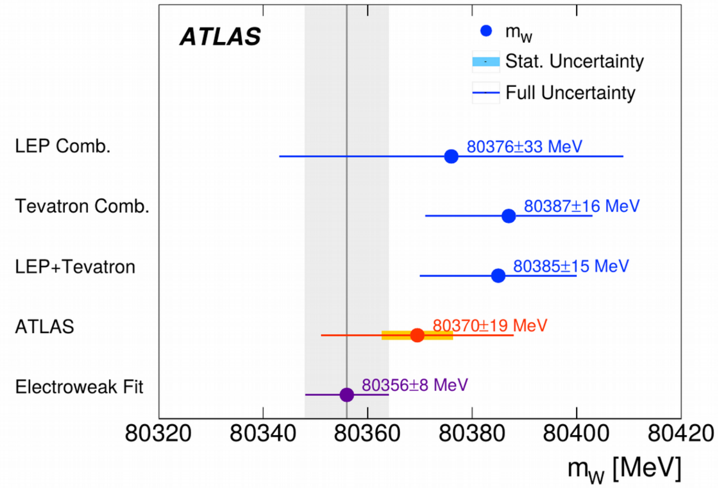
- **Discovery**: we see an excess – is it a (new) signal, or a background fluctuation ?

- **Upper limits**: we don´t see an excess – if there is a signal present, how small must it be ?

- **Parameter measurement:** what is the allowed range ("confidence interval") for a model parameter ?





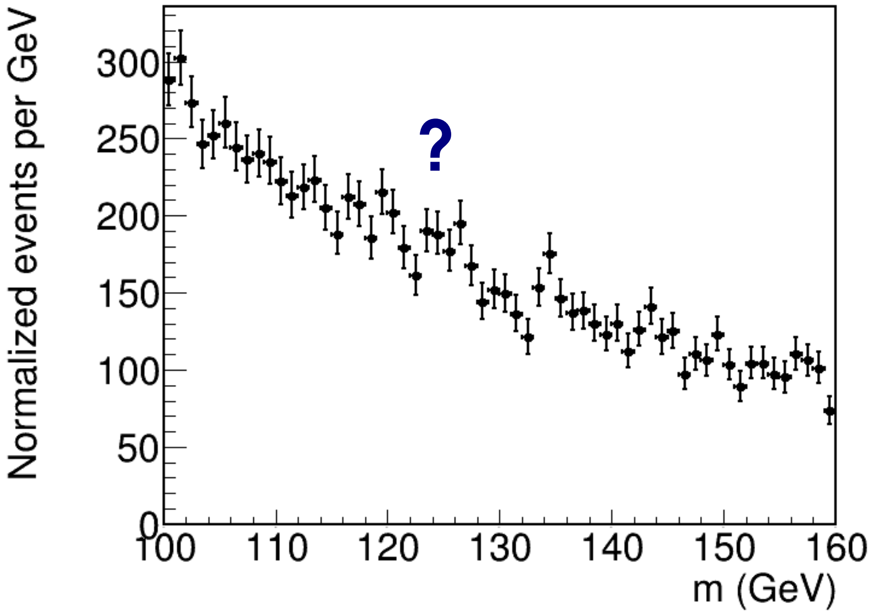


Upper Limits

Hypothesis tests for Limits

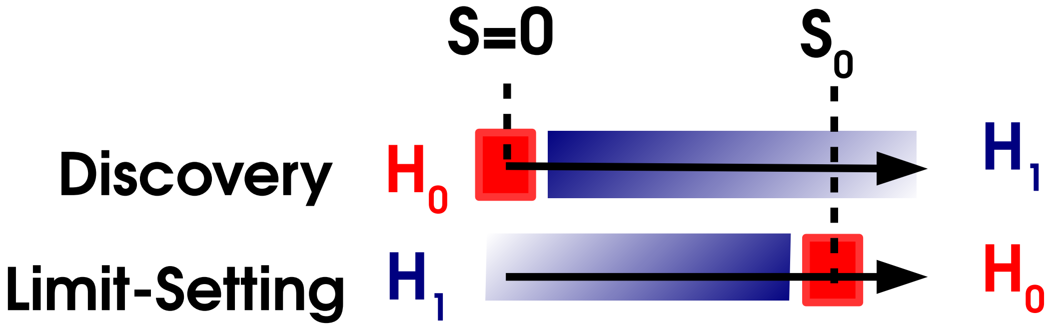
If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

→ More interesting to **exclude large signals** → **Upper limits on signal yield**



For **discovery**

- Try to exclude $\mathbf{H_0 : S=0}$
- Alternative : $\mathbf{H_1 : S > 0}$
- Report p-value for the test (or Z)



S=0 **S_0**

Discovery $\mathbf{H_0}$ $\mathbf{H_1}$

Limit-Setting $\mathbf{H_1}$ $\mathbf{H_0}$

For **limit-setting**:

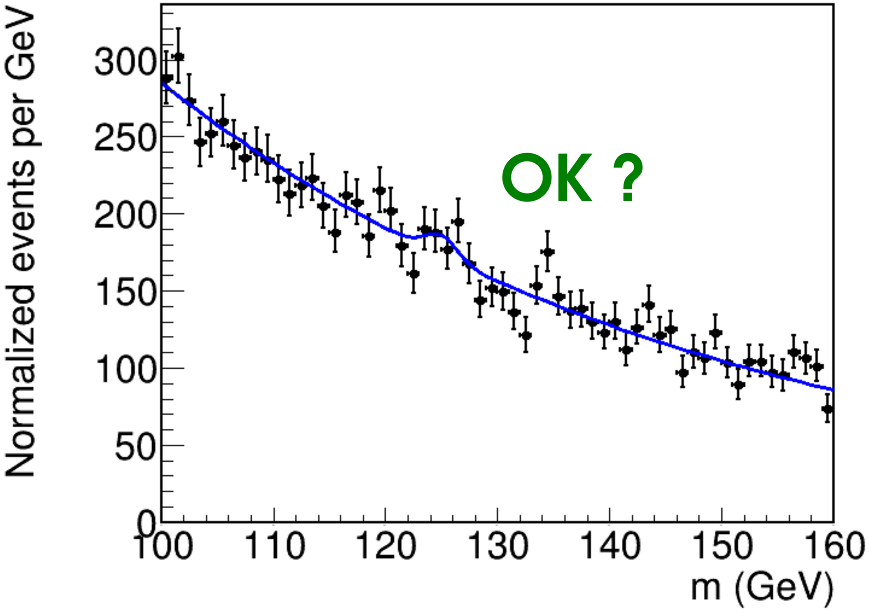
- Try to exclude $\mathbf{H_0 : S=S_0}$
- Alternative : $\mathbf{H_1 : S < S_0}$
- Usually, **adjust S_0 to get a predefined p-value** (typically **5%**)

  → ***Confidence Levels***: **CL = 1 - p** (p = 5% ⇔ 95% CL)

Hypothesis tests for Limits

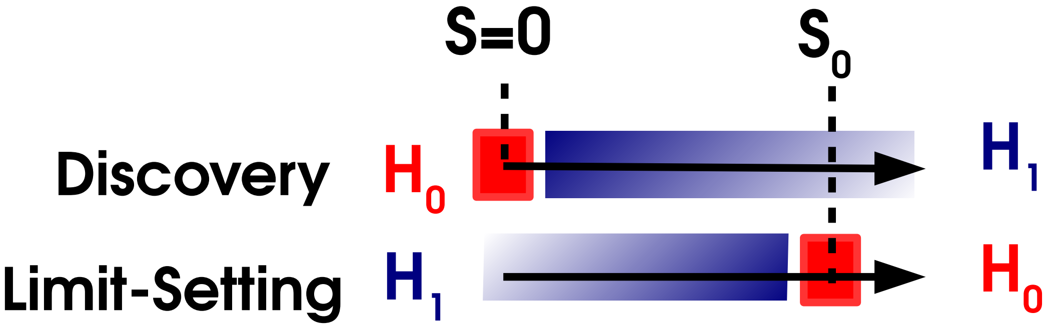
If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

→ More interesting to **exclude large signals** → **Upper limits on signal yield**



For **discovery**

- Try to exclude $H_0 : S=0$
- Alternative : $H_1 : S > 0$
- Report p-value for the test (or Z)



For **limit-setting**:

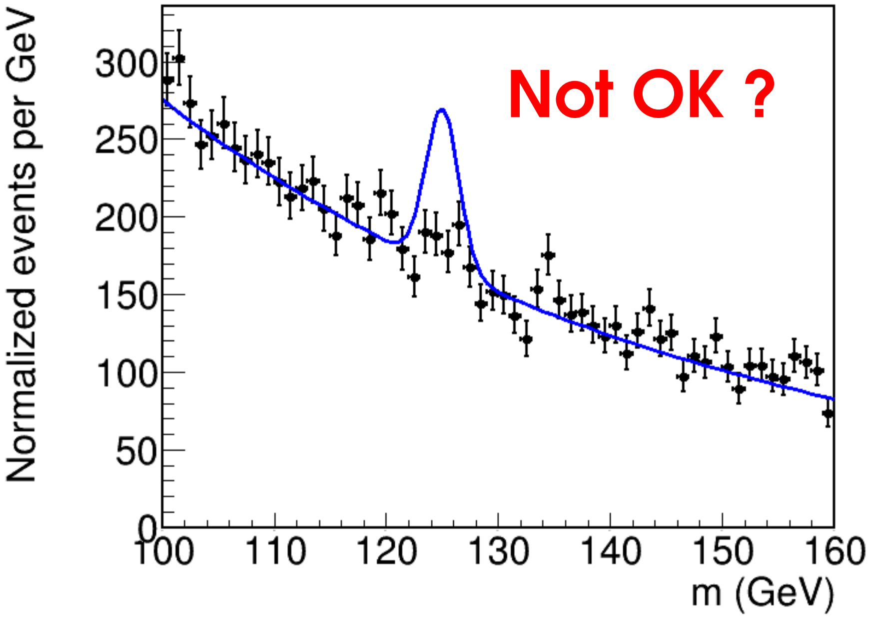
- Try to exclude $H_0 : S=S_0$
- Alternative : $H_1 : S < S_0$
- Usually, **adjust S_0 to get a predefined p-value** (typically **5%**)

→ ***Confidence Levels***: **CL = 1 - p** (p = 5% ⇔ 95% CL)

Hypothesis tests for Limits

If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

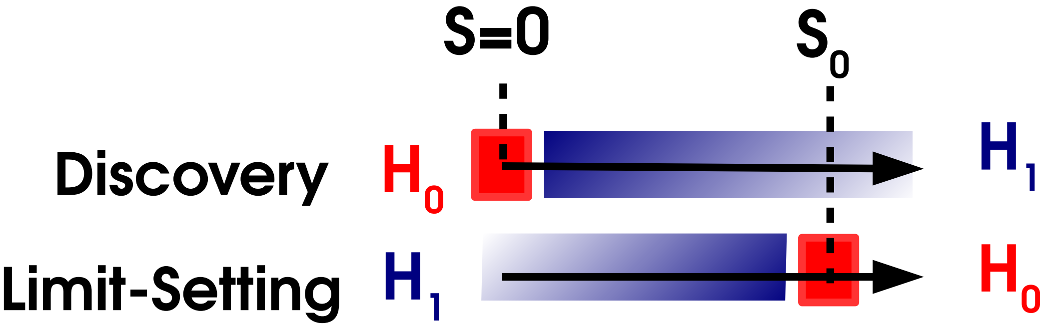
→ More interesting to **exclude large signals** → **Upper limits on signal yield**



Not OK ?

For **discovery**

- Try to exclude **H_0 : S=0**
- Alternative : **H_1 : S > 0**
- Report p-value for the test (or Z)



For **limit-setting**:

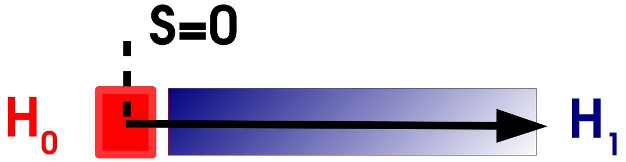
- Try to exclude **H_0 : S=S_0**
- Alternative : **H_1 : S < S_0**
- Usually, **adjust S_0 to get a predefined p-value** (typically **5%**)

  → ***Confidence Levels***: **CL = 1 - p** (p = 5% ⇔ 95% CL)

Test Statistic for Limit-Setting

Discovery :

- $H_0 : S = 0$
- $H_1 : S > 0$



Compare

$$q_0 = -2 \log \frac{L(S=0)}{L(\hat{S})}$$

← Likelihood of H_0

← Likelihood of H_1

Limit-setting

- $H_0 : S = \mu_0$
- $H_1 : S < \mu_0$

Compare

$$q_{S_0} = -2 \log \frac{L(S_0)}{L(\hat{S})}$$

← Likelihood of H_0

← Likelihood of H_1

$\hat{S} \sim S_0$ (no exclusion) : $q_{S0} \sim 0$

$\hat{S} \ll S_0$ (good exclusion) : $q_{S0} \gg 1$

Same as q_0 : large values

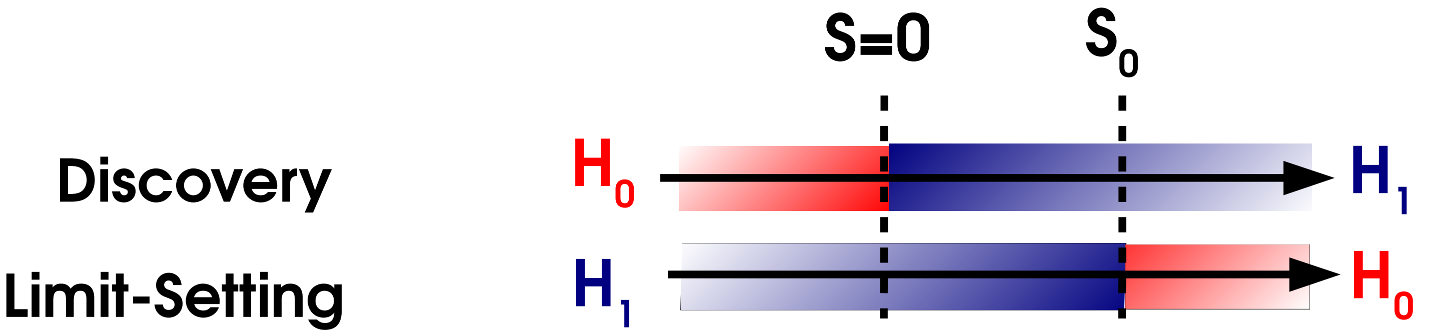
⇒ good rejection of H_0.

One-sided Test Statistic

For upper limits, alternate is **$H_1 : S < \mu_0$** :

→ If **large** signal observed ($\hat{S} \gg S_0$), does not favor H_1 over H_0

→ Only consider **$\hat{S} < S_0$** for **H_1**, and include **$\hat{S} \geq S_0$** in **H_0**.



⇒ Set **q_{S0} = 0 for $\hat{S} > S_0$** – only small signals ($\hat{S} < S_0$) help lower the limit.

→ Also treat separately the case S < 0 to avoid technical issues in -2logL fits.

$$\tilde{q}_{S_0} = \begin{cases} 0 & \hat{S} \geq S_0 \\ -2\log\dfrac{L(S=S_0)}{L(\hat{S})} & 0 \leq \hat{S} \leq S_0 \\ -2\log\dfrac{L(S=S_0)}{L(S=0)} & \hat{S} < 0 \end{cases}$$

Asymptotics:

q_{S0} ~ "½χ^2" under $H_0(S=S_0)$, same as q_0, except for special treatment of $\hat{S} < 0$.

$$p_0 = 1 - \Phi\left(\sqrt{q_{S_0}}\right)$$

Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

Inversion : Getting the limit for a given CL

Procedure

→ Consider **H_0** : $H(S=S_0)$ – alternative **H_1** : $H(\hat{S} < S_0)$

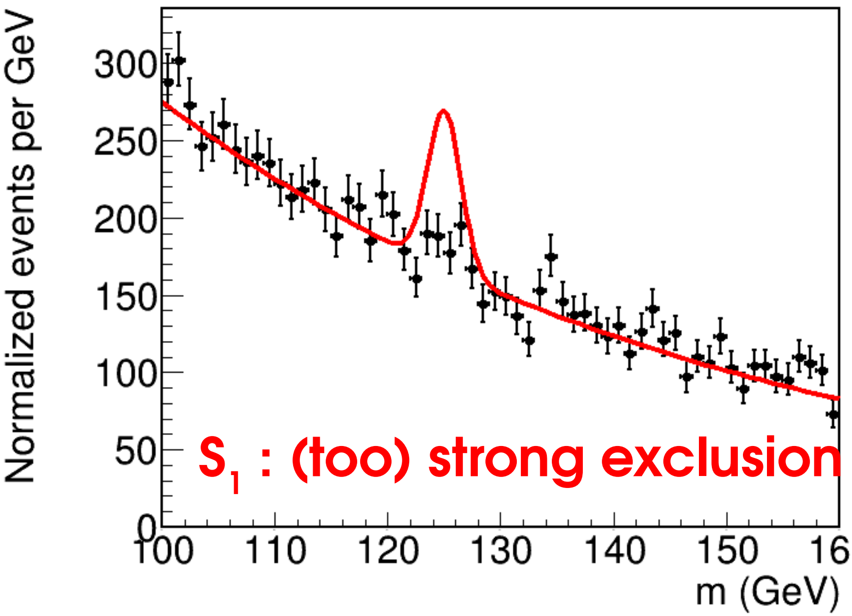
→ Compute q_{S0}, get **exclusion p-value p_{S0}.**

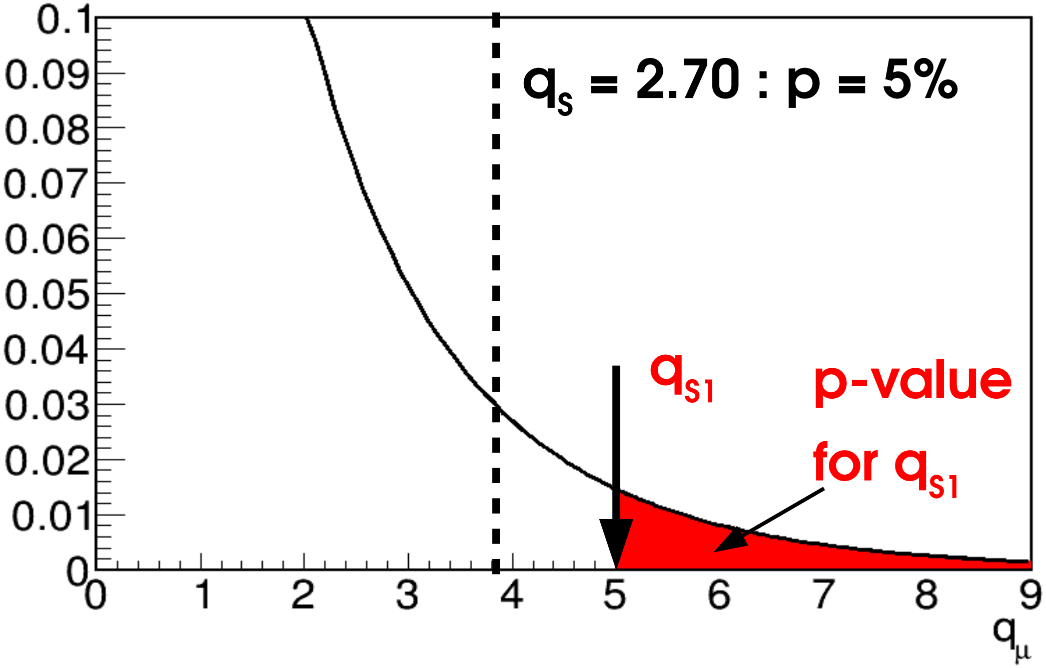
→ Adjust S_0 until 95% CL exclusion (p_{S0} = 5%) is reached

Asymptotics: set target in terms of q_{S0} : $\sqrt{q_{S_0}} = \Phi^{-1}(1 - p_0)$

Asymptotics

CL	Region
90%	$q_S > 1.64$
95%	$q_S > 2.70$
99%	$q_S > 5.41$





Inversion : Getting the limit for a given CL

Procedure

→ Consider **H_0** : H(S=S_0) – alternative **H_1** : H($\hat{S} < S_0$)

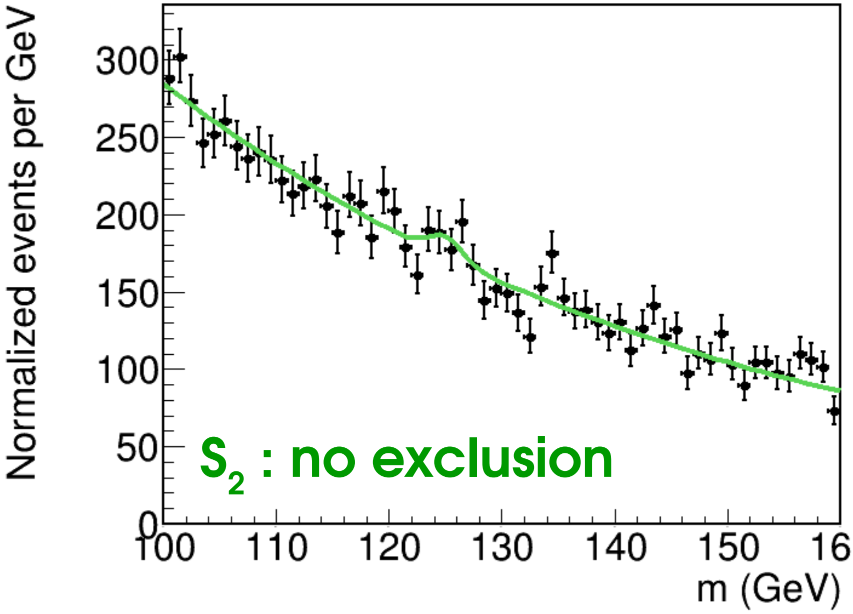
→ Compute q_{S0}, get **exclusion p-value p_{S0}.**

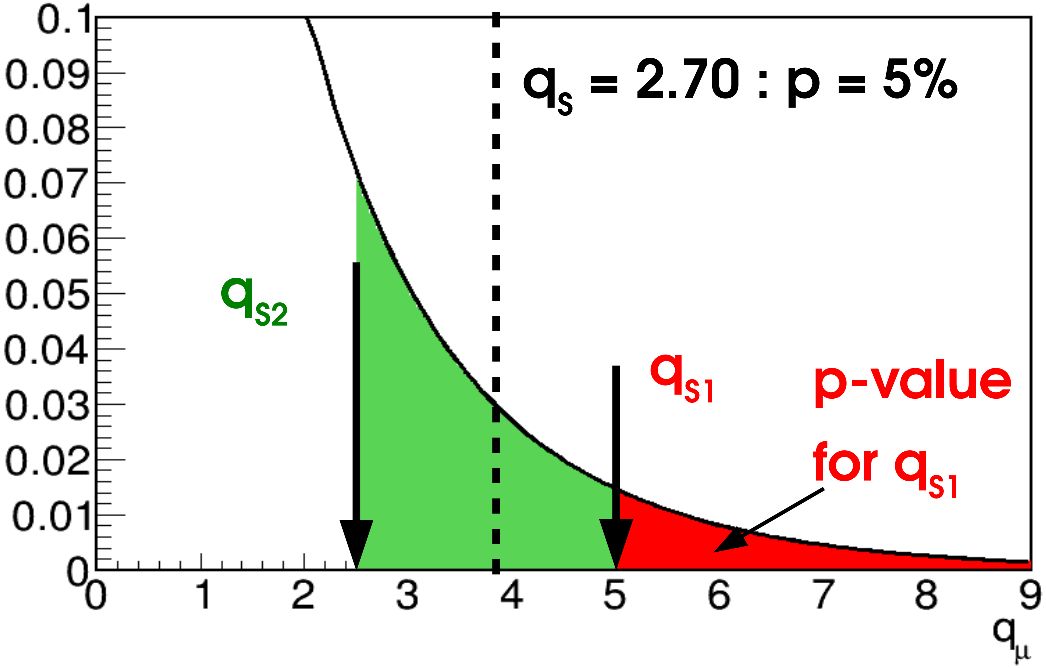
→ Adjust S_0 until 95% CL exclusion (p_{S0} = 5%) is reached

Asymptotics: set target in terms of q_{S0} : $\sqrt{q_{S_0}} = \Phi^{-1}(1 - p_0)$

Asymptotics

CL	Region
90%	$q_S > 1.64$
95%	$q_S > 2.70$
99%	$q_S > 5.41$





Inversion : Getting the limit for a given CL

Procedure

→ Consider **H_0** : $H(S=S_0)$ – alternative **H_1** : $H(\hat{S} < S_0)$

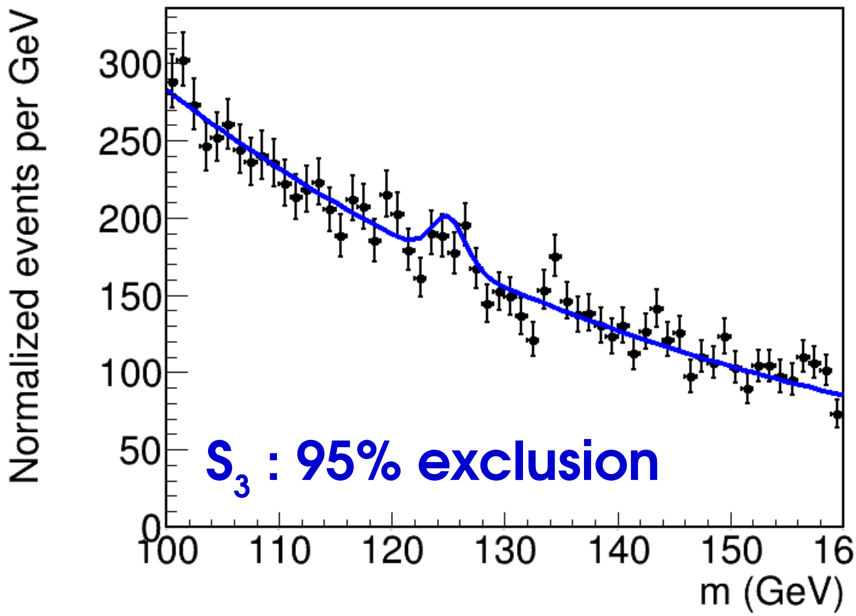
→ Compute q_{S0}, get **exclusion p-value p_{S0}.**

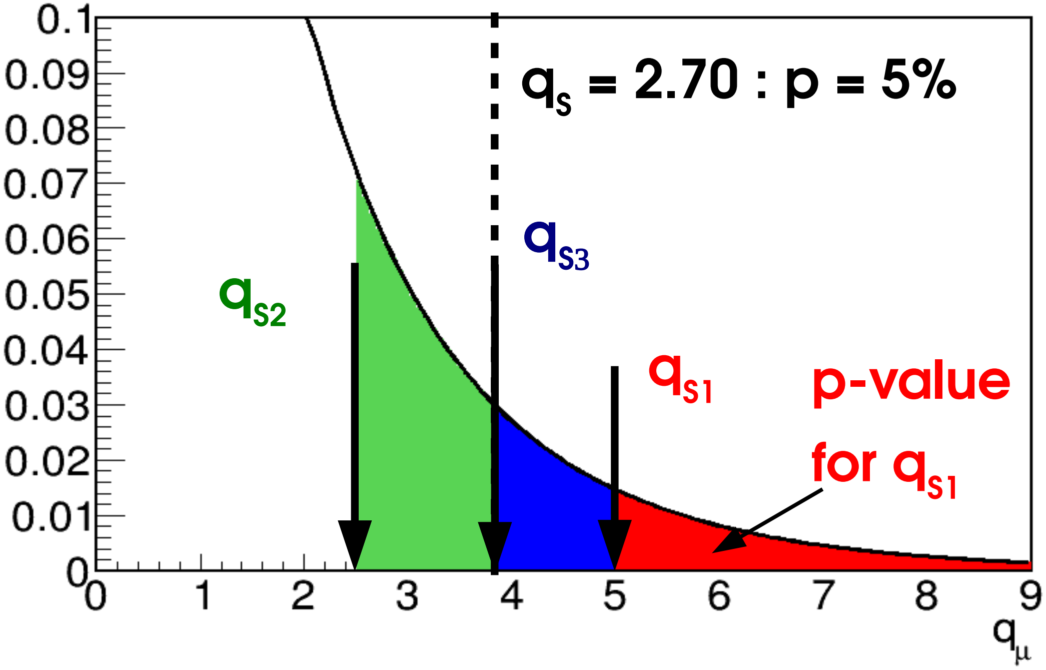
→ Adjust S_0 until 95% CL exclusion (p_{S0} = 5%) is reached

Asymptotics: set target in terms of q_{S0} : $\sqrt{q_{S_0}} = \Phi^{-1}(1-p_0)$

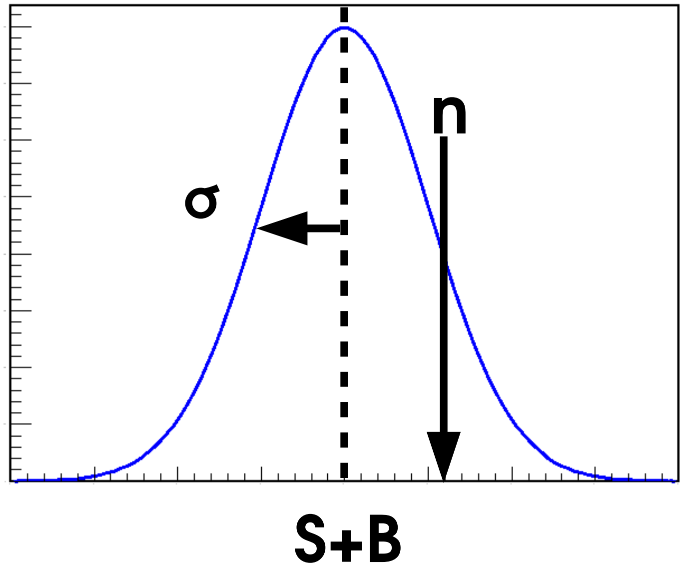
Asymptotics

CL	Region
90%	$q_S > 1.64$
95%	$q_S > 2.70$
99%	$q_S > 5.41$





Upper Limits: Gaussian Example



Usual Gaussian counting example with known B:

$$\lambda(S) = \left(\frac{n-(S+B)}{\sigma_S}\right)^2$$

Reminder:

Best fit signal : $\hat{S} = n - B$

Significance: $Z = \hat{S}/\sqrt{B}$

Compute the 95% CL upper limit on S:

$$q_{S_0} = -2\log\frac{L(S=S_0)}{L(\hat{S})} = \lambda(S_0) - \lambda(\hat{S}) = \left(\frac{n-(S_0+B)}{\sigma_S}\right)^2 = \left(\frac{S_0-\hat{S}}{\sigma_S}\right)^2$$

for $S_0 > \hat{S}$

so $q_{S_0} = 2.70$ for $S_0 = \hat{S} + \sqrt{2.70}\,\sigma_S$

And finally $S_{\mathrm{up}} = \hat{S} + 1.64\,\sigma_S$ at 95 % CL

Upper Limit Pathologies

Upper limit: $S_{up} \sim \hat{S} + 1.64\, \sigma_S$.

Problem: for negative Ŝ, get **very** good observed limit.

→ For Ŝ sufficiently negative, even $S_{up} < 0$!

How can this be ?

→ **Background modeling issue ?…** Or:

→ This is a **95%** limit

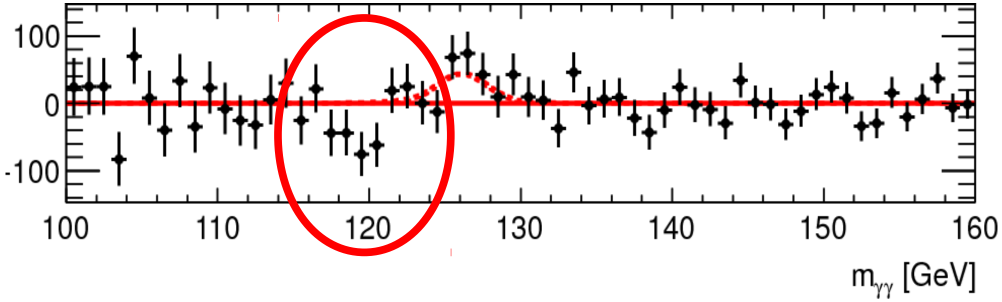
⇒ **5% of the time, the limit wrongly excludes the true value,** *e.g.* S*=0.

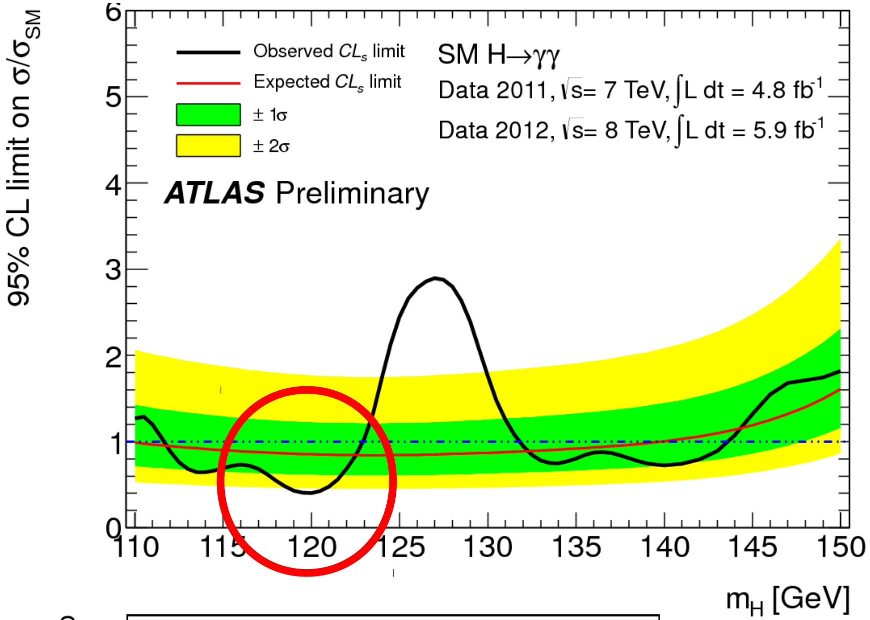
But if we assume S must be >0, we know a priori this is just a fluctuation.

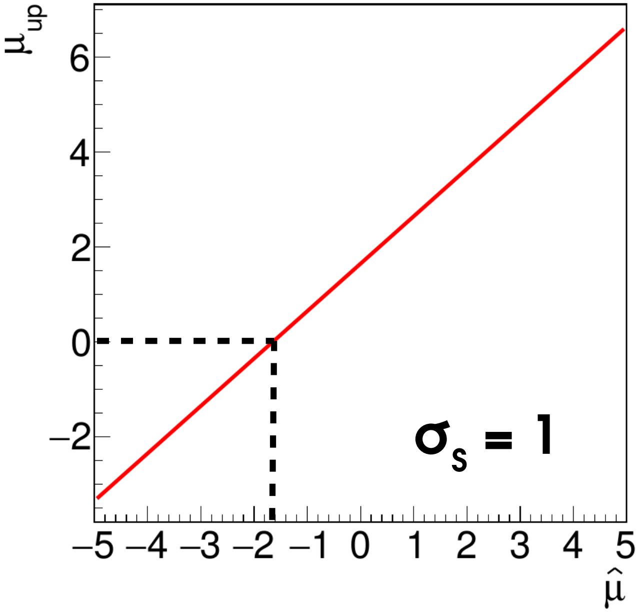
Options

→ **live with it**: sometimes report limit < 0

→ **Special procedure to avoid these cases**

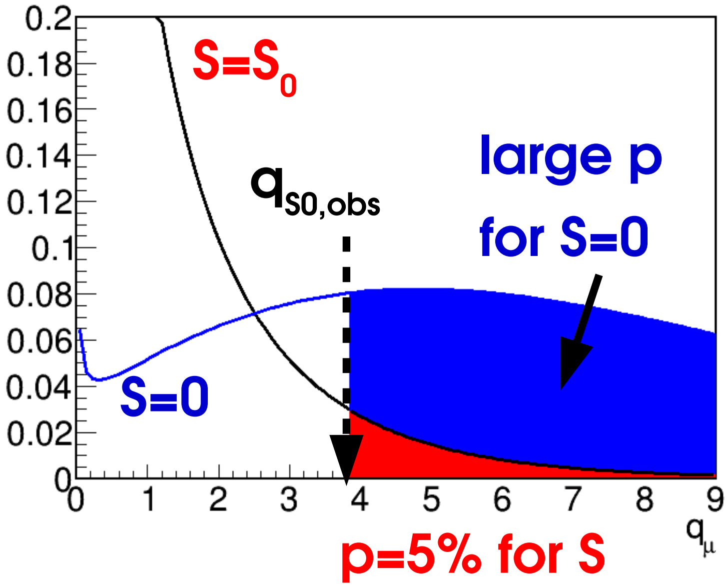






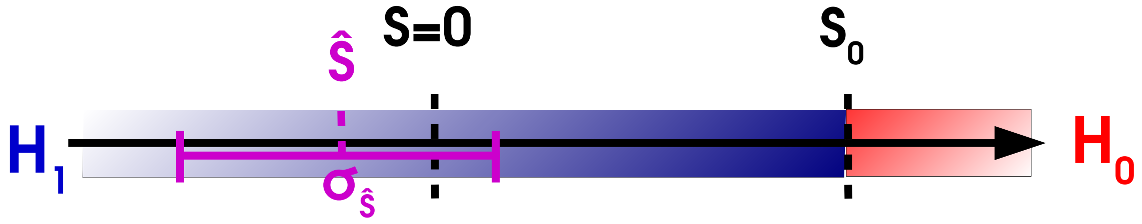
Upper Limit Pathologies

When setting limits, goal is to exclude large S, **to indicate that S~0.** What happens at S=0 ?



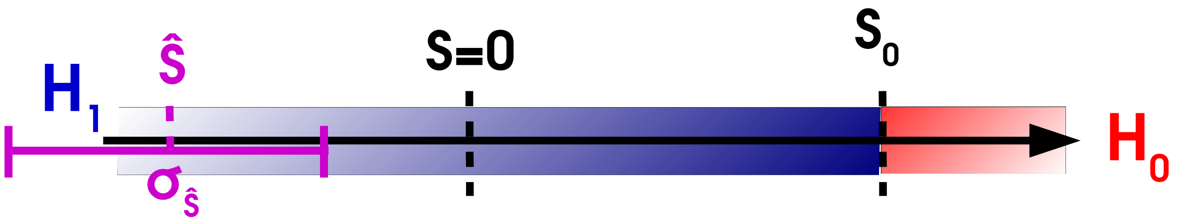
Normal case: $\hat{S} \sim 0$, **S=0 not excluded** :

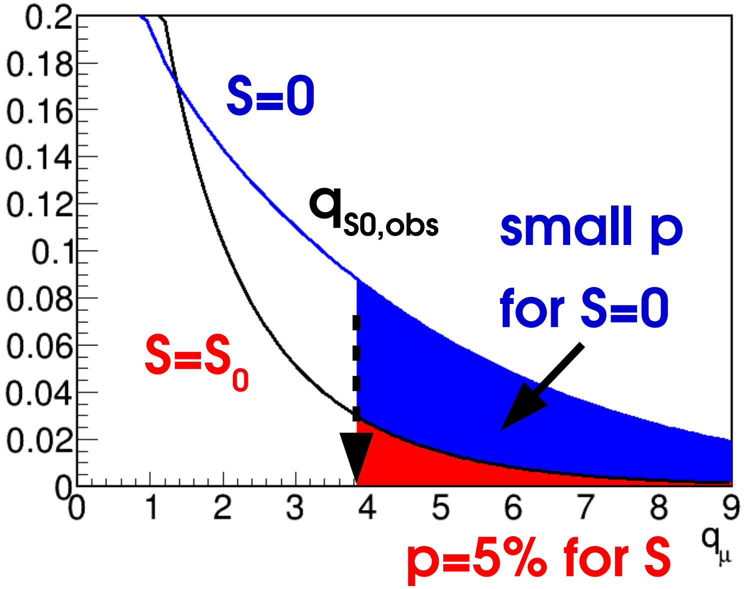
$S_{up} = \hat{S} + 1.64\, \sigma_S > 0$, large p-value for S=0



Pathological case, very negative $\hat{S}$, **S=0 also excluded** :

$S_{up} = \hat{S} + 1.64\, \sigma_S < 0$, p-value for S=0 also small





→ However we know a priori that $S \geq 0$

⇒ Inject this information into the procedure

CL_s

A. Read, J.Phys. G28 (2002) 2693-2704

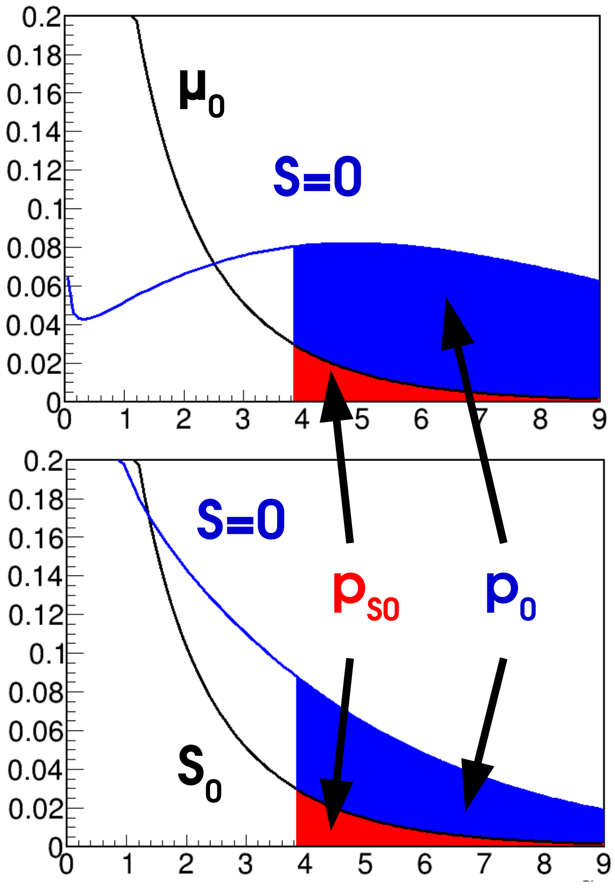
Usual solution in HEP : **CL_s**.

→ Compute modified p-value $$p_{CL_s} = \frac{p_{S_0}}{p_0}$$

- p_{S0} is the usual p-value (5%)
- p_0 is the p-value computed under H(S=0).

⇒ **Rescale** exclusion at S_0 by exclusion at S=0.

→ Somewhat ad-hoc, but good properties…



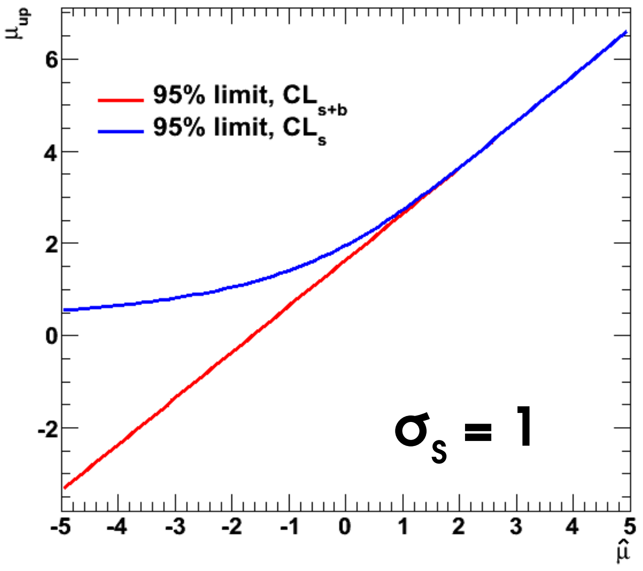
Good case : p_0 ~ O(1)

p_{CLs} **~ p_{S0} ~ 5%, no change**.

Pathological case : $p_0 \ll 1$

$p_{CLs} \sim p_{S0}/p_0 \gg 5\%$

→ no exclusion ⇒ worse limit, usually >0 as desired



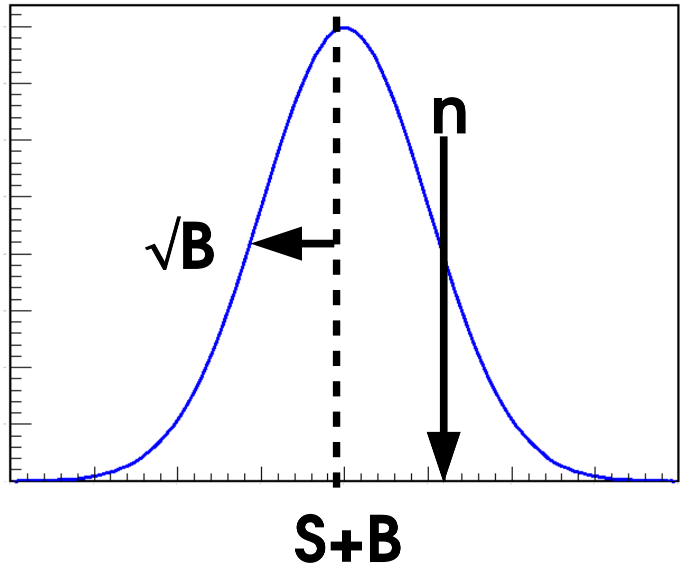
Drawback: ***overcoverage***

→ limit is actually >95% CL for small p_0.

$\mathrm{CL_s}$: Gaussian Example

Usual Gaussian counting example with known B:

$$\lambda(S) = \left(\frac{n-(S+B)}{\sigma_S}\right)^2$$



Reminder

Best fit signal : **Ŝ = n - B**

$\mathrm{CL_{s+b}}$ limit:

$$S_{\mathrm{up}} = \hat{S} + 1.64\,\sigma_S \ \text{at } 95\,\% \ \mathrm{CL}$$

$\mathrm{CL_s}$ upper limit : still have

$$q_{S_0} = \left(\frac{S_0-\hat{S}}{\sigma_S}\right)^2 \quad (\text{for } S_0 > \hat{S})$$

so need to solve

$$p_{CL_s} = \frac{p_{S_0}}{p_0} = \frac{1-\Phi\left(\sqrt{q_{S_0}}\right)}{1-\Phi\left(\sqrt{q_{S_0}} - S_0/\sigma_S\right)} = 5\,\%$$

for Ŝ = 0,

$$S_{\mathrm{up}} = \hat{S} + \left[\Phi^{-1}\left(1 - 0.05\,\Phi\left(\hat{S}/\sigma_S\right)\right)\right]\sigma_S \ \text{at } 95\,\% \ \mathrm{CL}$$

$\Phi(0) = 0.5 \Rightarrow$ at 95% CL, $\mathrm{CL_s}$: $S_{\mathrm{up}} = 1.96\,\sigma_S$ $\quad\mathrm{CL_{s+b}}$: $S_{\mathrm{up}} = 1.64\,\sigma_S$

$\hat{S} \sim G(S, \sigma_S)$ so

Under $H_0(S = S_0)$:

$$\sqrt{q_{S_0}} \sim G(0,1)$$
$$p_{S_0} = 1-\Phi\left(\sqrt{q_{S_0}}\right)$$

Under $H_0(S = 0)$:

$$\sqrt{q_{S_0}} \sim G(S_0/\sigma_S, 1)$$
$$p_0 = 1-\Phi\left(\sqrt{q_{S_0}} - S_0/\sigma_S\right)$$

CL$_s$: Poisson Rule of Thumb

Same exercise, for the Poisson case

Exact computation : sum probabilities of cases “at least as extreme as data” (n)

$$p_{S_0}(n) = \sum_0^n e^{-(S_0+B)} \frac{(S_0+B)^k}{k!}$$ and one should solve $p_{CL_s} = \frac{p_{S_{up}}(n)}{p_0(n)} = 5\%$ for S_{up}

For n = 0: $$p_{CL_s} = \frac{p_{S_{up}}(0)}{p_0(0)} = e^{-S_{up}} = 5\% \Rightarrow S_{up} = \log(20) = 2.996 \approx 3$$

⇒ Rule of thumb: when n$_{obs}$=0, the 95% CL$_s$ limit is 3 events (for any B)

Asymptotics: as before, $$q_{S_0} = \lambda(S_0) - \lambda(\hat{S}) = 2(S_0 + B - n) - 2n \log \frac{S_0+B}{n}$$

For n = 0, $$q_{S_0}(n=0) = 2(S_0+B)$$

$$p_{CL_s} = \frac{p_{S_0}}{p_0} = \frac{1-\Phi\left(\sqrt{q_{S_0}(n=0)}\right)}{1-\Phi\left(\sqrt{q_{S_0}(n=0)} - \sqrt{q_{S_0}(n=B)}\right)} = 5\%$$

⇒ S$_{up}$ ~ 2, exact value depends on B

⇒ Asymptotics not valid in this case (n=0) – need to use exact results, or toys

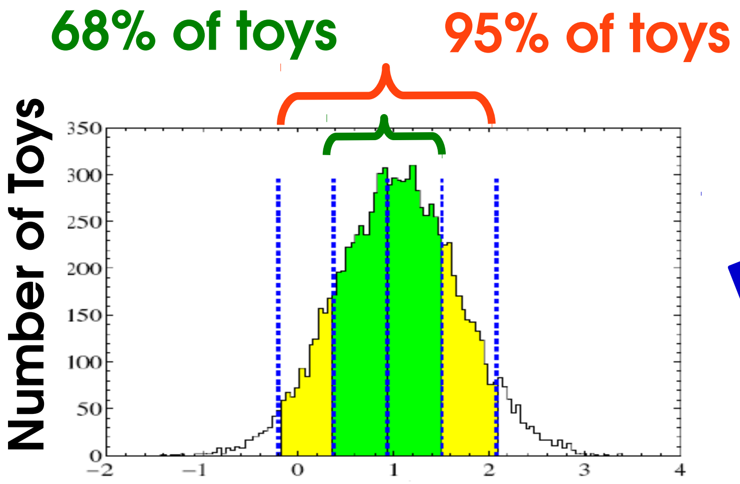
Expected Limits: Toys

***Expected* results**: median outcome under a given hypothesis
→ usually B-only by convention, but other choices possible.

Two main ways to compute:

→ Pseudo-experiments (*toys*):

- Generate pseudo-data in B-only hypothesis
- Compute limit
- Repeat and histogram the results
- Central value = median, bands based on quantiles

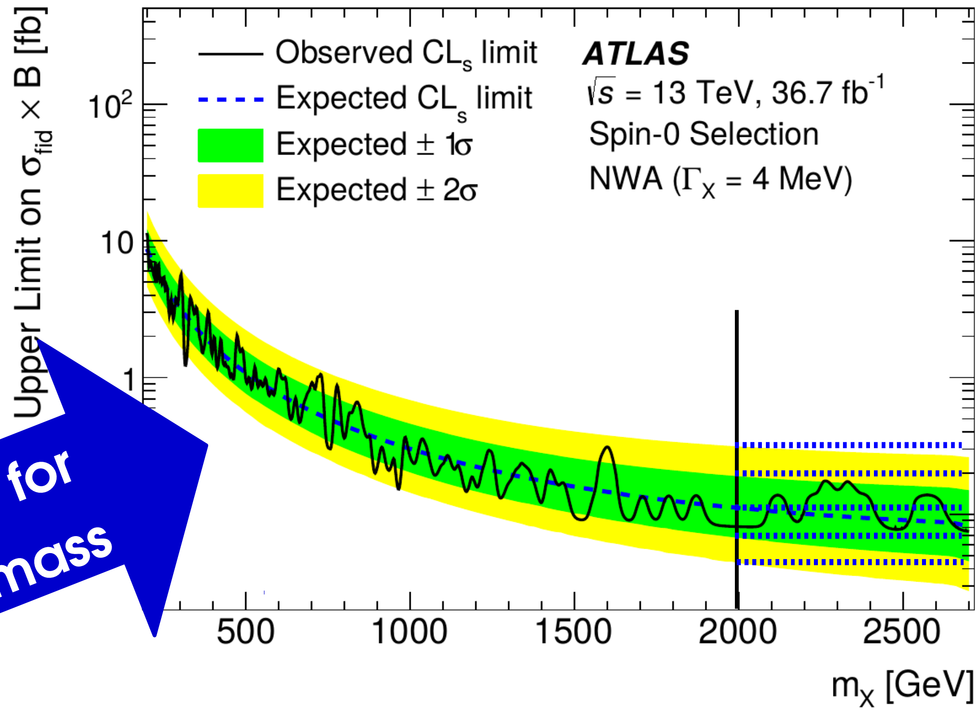


Eur.Phys.J.C71:1554,2011

Computed limit

Repeat for each mass

Phys. Lett. B 775 (2017) 105



Expected Limits: Asimov

***Expected* results**: median outcome under a given hypothesis
→ usually B-only by convention, but other choices possible.

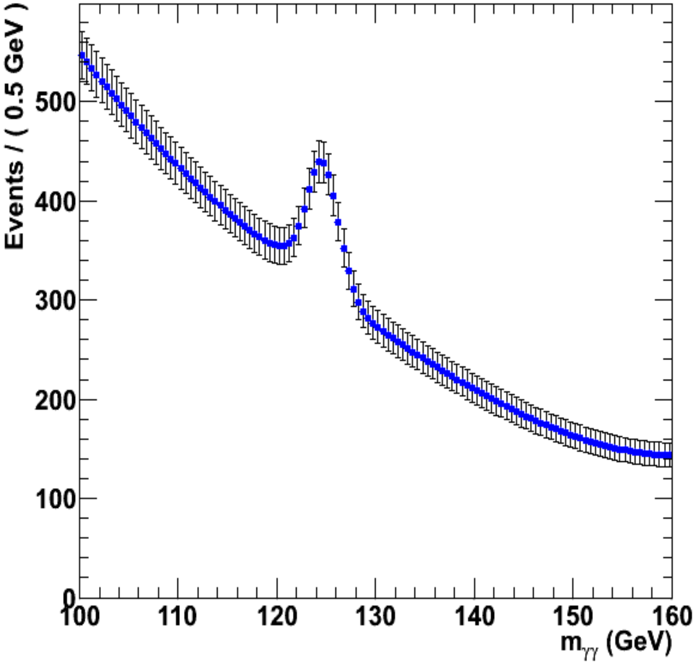
Two main ways to compute:

Strictly speaking, Asimov dataset if
$\hat{X} = X_0$ **for all parameters** X,
where X_0 is the generation value

→ *Asimov Datasets*

- Generate a "perfect dataset" – *e.g.* for binned data, set bin contents carefully, no fluctuations.
- Gives the median result immediately:
  median(toy results) ↔ result(median dataset)
- Get bands from asymptotic formulas:
  Band width

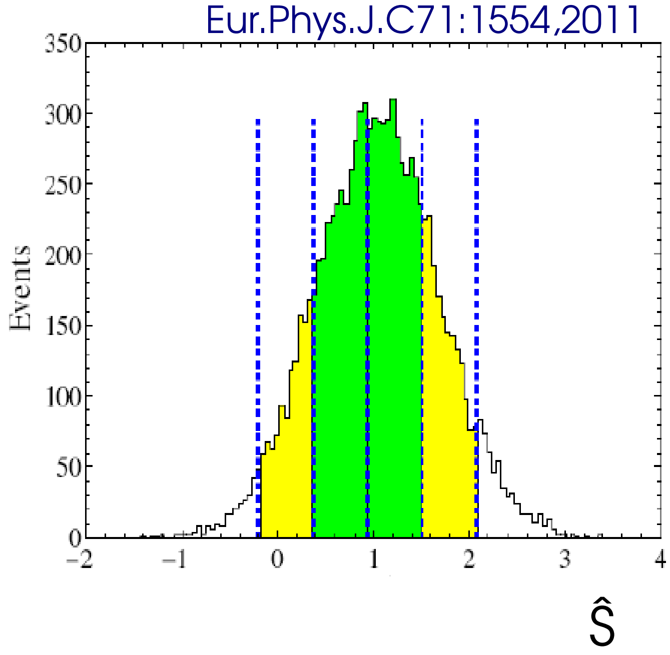
$$\sigma^2_{S_0,A} = \frac{S_0^2}{q_{S_0}(\text{Asimov})}$$



⊕ Much faster (1 "toy")

⊖ Relies on Gaussian approximation

$\mathrm{CL_s}$: Gaussian Bands



Eur.Phys.J.C71:1554,2011

Usual Gaussian counting example with known B:

95% $\mathrm{CL_s}$ upper limit on S:

$$S_{\mathrm{up}} = \hat{S} + \left[\Phi^{-1}\left(1 - 0.05\,\Phi\left(\hat{S}/\sigma_S\right)\right)\right]\sigma_S$$

with $\sigma_S = \sqrt{B}$

Compute expected bands for S=0:

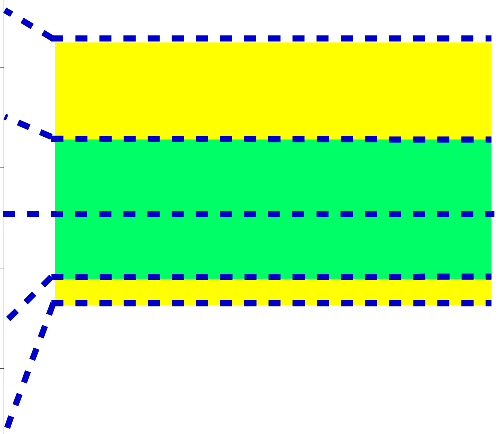
→ **Asimov dataset ⇔ Ŝ = 0** :

$$S^{0}_{\mathrm{up,exp}} = 1.96\,\sigma_S$$

→ **± nσ bands**:

$$S^{\pm n}_{\mathrm{up,exp}} = \left(\pm n + \left[1 - \Phi^{-1}\left(0.05\,\Phi\left(\mp n\right)\right)\right]\right)\sigma_S$$

n	$S_{\mathrm{exp}}^{\pm n}/\sqrt{B}$
+2	3.66
+1	2.72
0	1.96
-1	1.41
-2	1.05

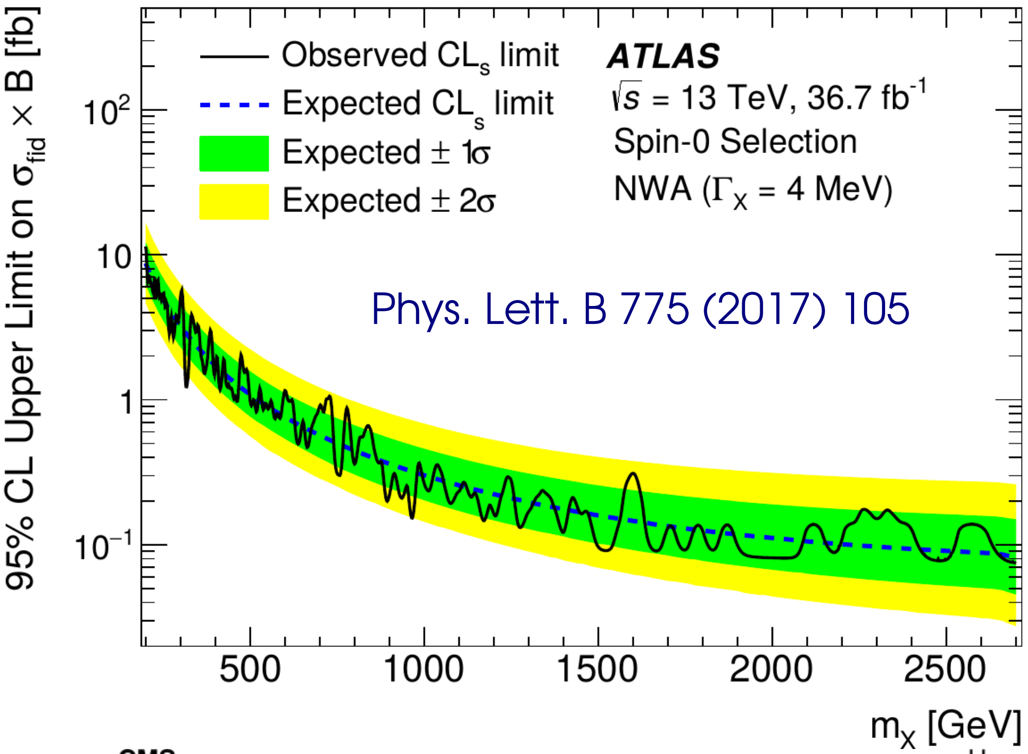


CLs :

- Positive bands somewhat reduced,
- Negative ones more so

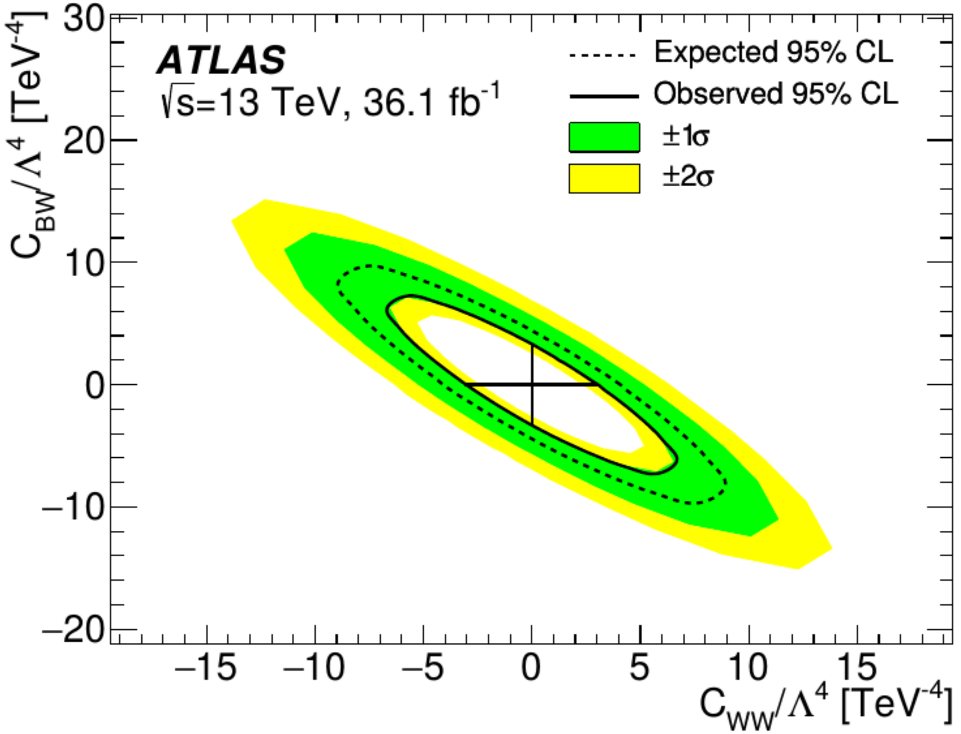
Band width from $\sigma^2_{S,A} = \dfrac{S^2}{q_S(\mathrm{Asimov})}$ depends on S, for non-Gaussian cases,different values for each band...

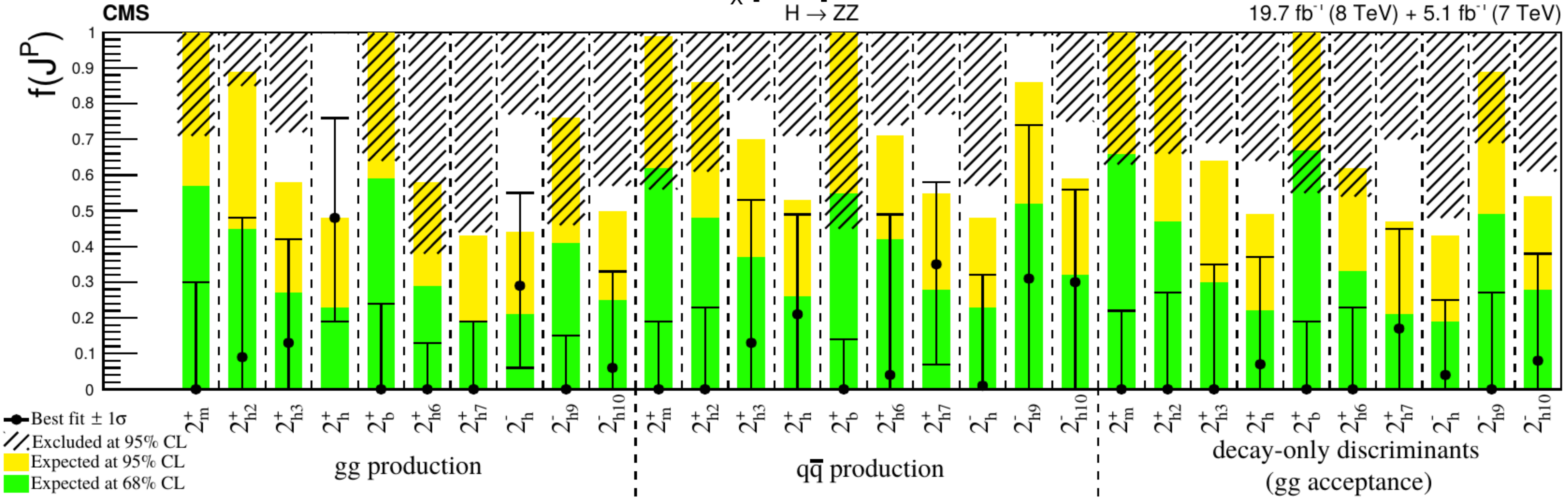
Upper Limit Examples



Phys. Lett. B 775 (2017) 105

ATLAS 2015-2016 4l aTGC Search





Phys. Rev. D 92 (2015) 012004

Outline

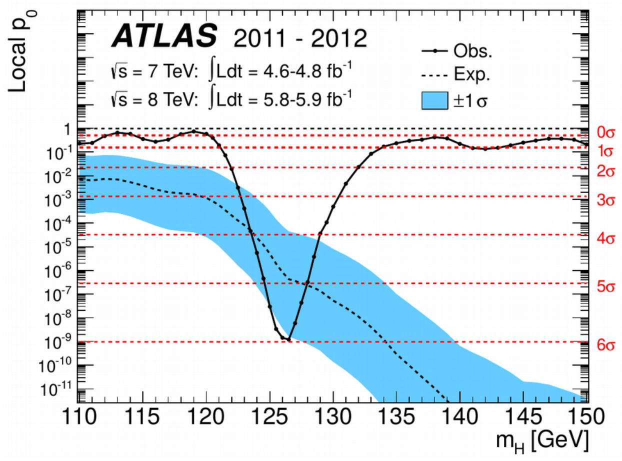
Lecture I:

- Statistics basics
- Describing measurements
- Computing statistics results:

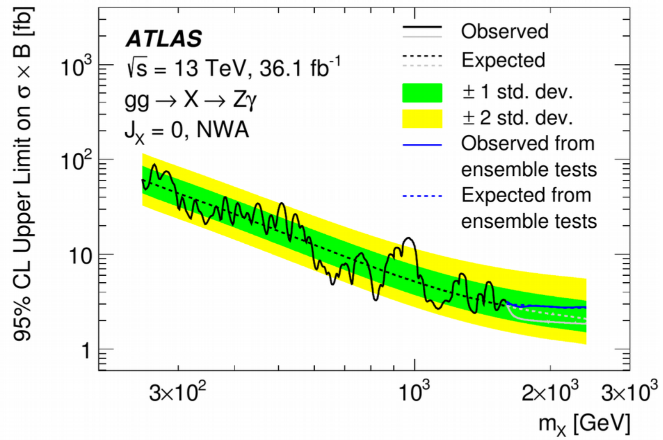
Today:

- Computing statistics results:
  - Discovery
  - Limits
  - **Confidence intervals**
- **Profiling**

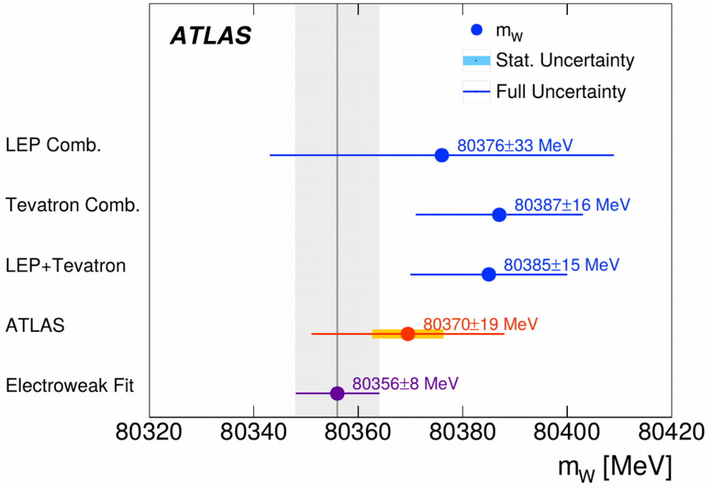
Usual Statistical Results



- **Discovery**: we see an excess – is it a (new) signal, or a background fluctuation ?



- **Upper limits**: we don't see an excess – if there is a signal present, how small must it be ?



- **Parameter measurement:** what is the allowed range ("confidence interval") for a model parameter ?

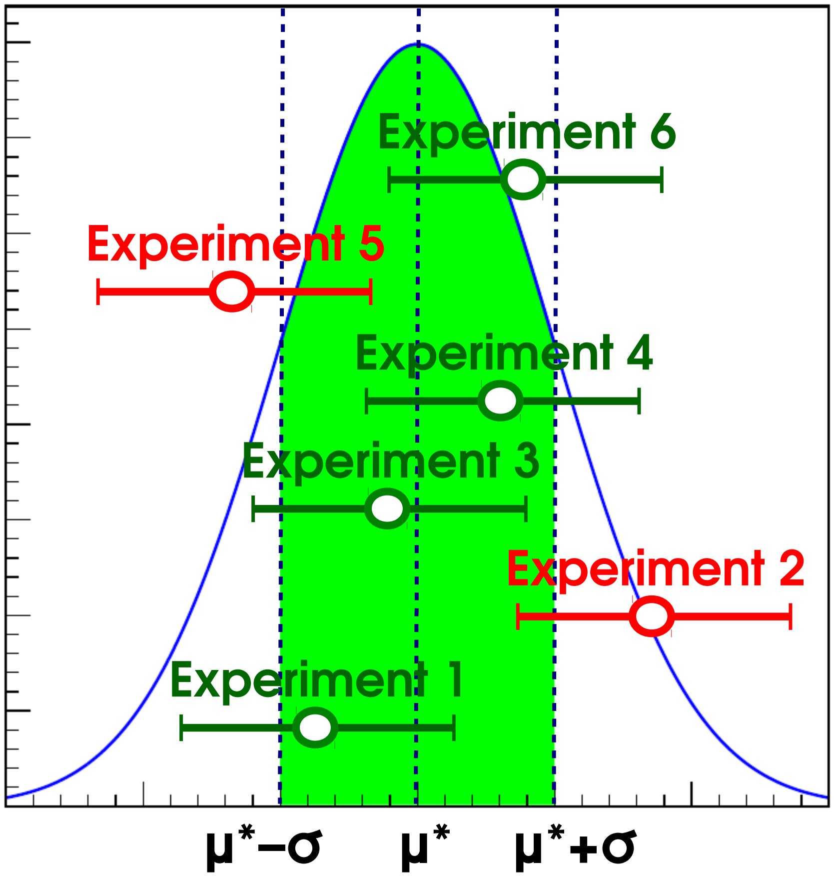
Confidence Intervals

Gaussian Inversion

If $\hat{\mu} \sim G(\mu^*, \sigma)$, known quantiles :

$$P\left(\mu^* - \sigma < \hat{\mu} < \mu^* + \sigma\right) = 68\,\%$$

This is a probability for $\hat{\mu}$, not μ^*!
→ μ^* is a **fixed number**, **not a random variable**



But we can invert the relation:

$$P\left(\mu^* - \sigma < \hat{\mu} < \mu^* + \sigma\right) = 68\,\%$$

$$\Rightarrow\ P\left(|\hat{\mu} - \mu^*| < \sigma\right) = 68\,\%$$

$$\Rightarrow\ P\left(\hat{\mu} - \sigma < \mu^* < \hat{\mu} + \sigma\right) = 68\,\%$$

→ This gives the desired statement on μ^* : *if we repeat the experiment many times, **$[\hat{\mu} - \sigma, \hat{\mu} + \sigma]$** will **contain the true value** 68% of the time:* **$\hat{\mu} = \mu^* \pm \sigma$**

This is a statement ***on the interval* $[\hat{\mu} - \sigma, \hat{\mu} + \sigma]$** obtained for each experiment

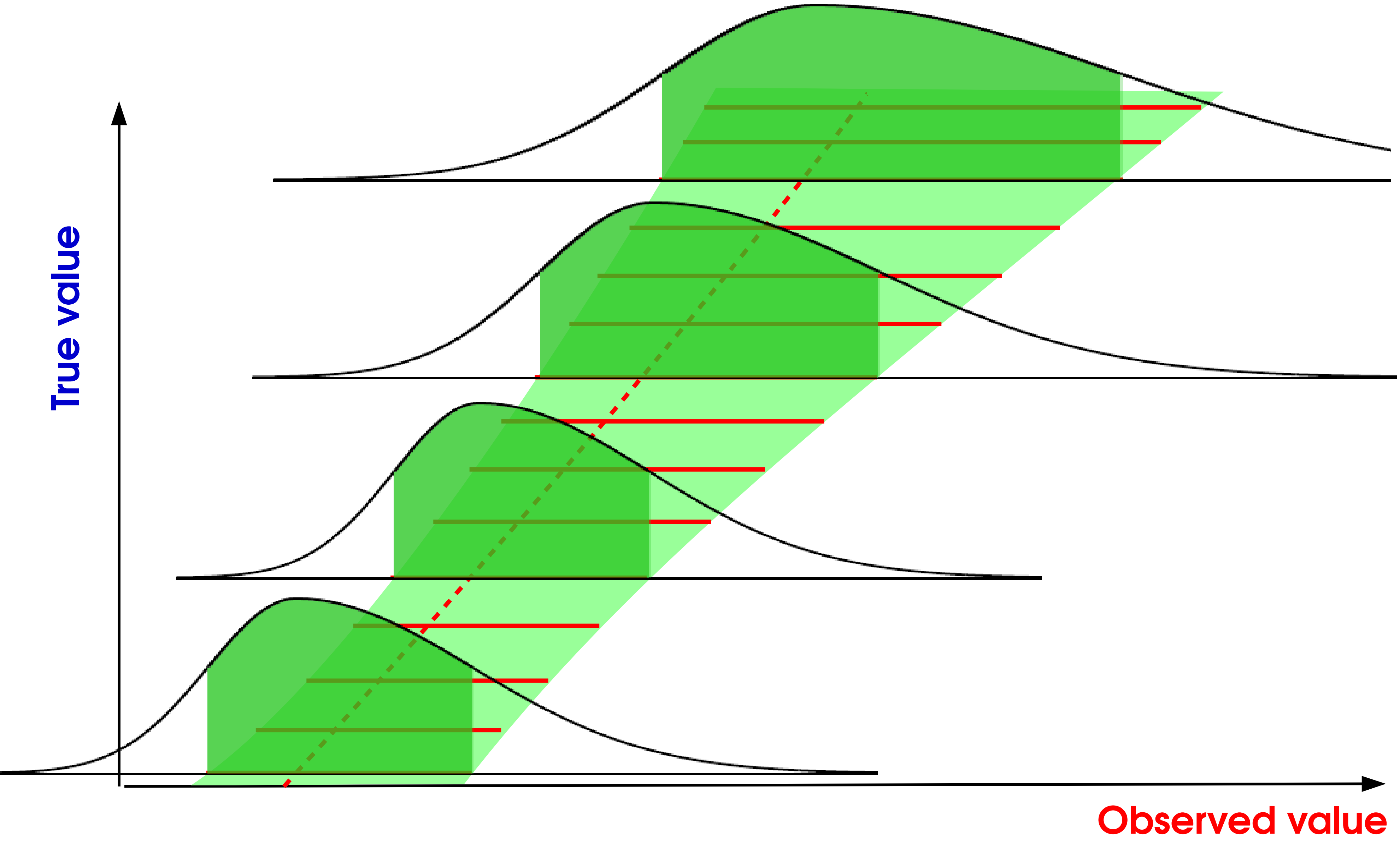
Works in the same way for other interval sizes: **$[\hat{\mu} - Z\sigma, \hat{\mu} + Z\sigma]$** with

Z	1	1.96	2
CL	0.68	0.95	0.955

Neyman Construction

General case: Build 1σ intervals of observed values for each true value

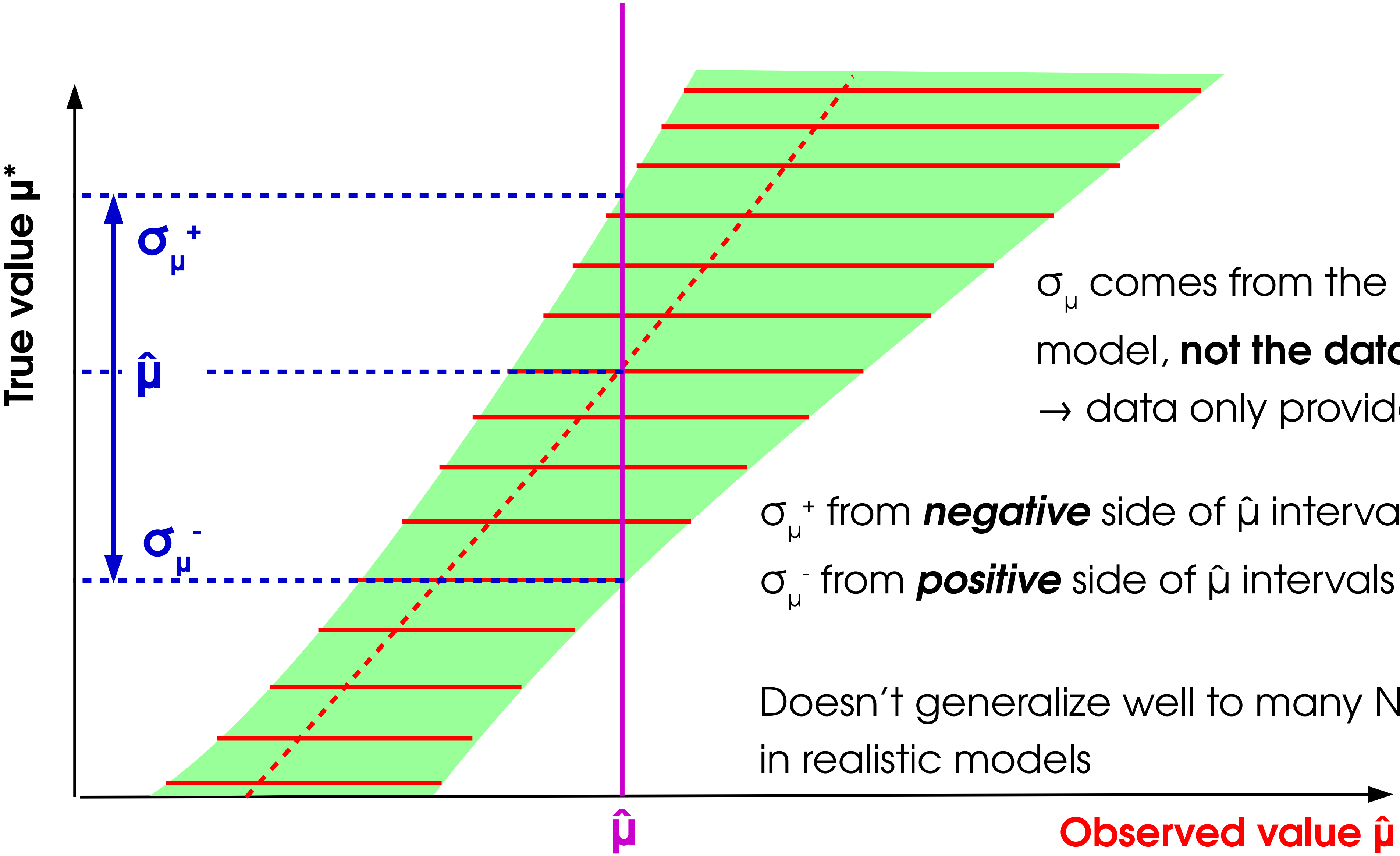
⇒ *Confidence belt*



Inversion using the Confidence Belt

General case: Intersect belt with given $\hat{\mu}$, get $P\left(\hat{\mu} - \sigma_\mu^- < \mu^* < \hat{\mu} + \sigma_\mu^+\right) = 68\,\%$

→ Same as before for Gaussian, works also when $P(\mu^{obs} | \mu)$ varies with μ.



σ_μ comes from the model, **not the data**

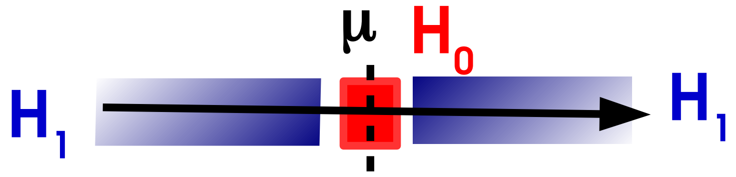
→ data only provides $\hat{\mu}$.

σ_μ^+ from ***negative*** side of $\hat{\mu}$ intervals

σ_μ^- from ***positive*** side of $\hat{\mu}$ intervals

Doesn't generalize well to many NPs in realistic models

Likelihood Intervals



Confidence intervals from L:

- Test H(μ_0) against alternative using $$t_{\mu_0} = -2\log\frac{L(\mu=\mu_0)}{L(\hat{\mu})}$$
- Two-sided test since true value can be higher or lower than observed

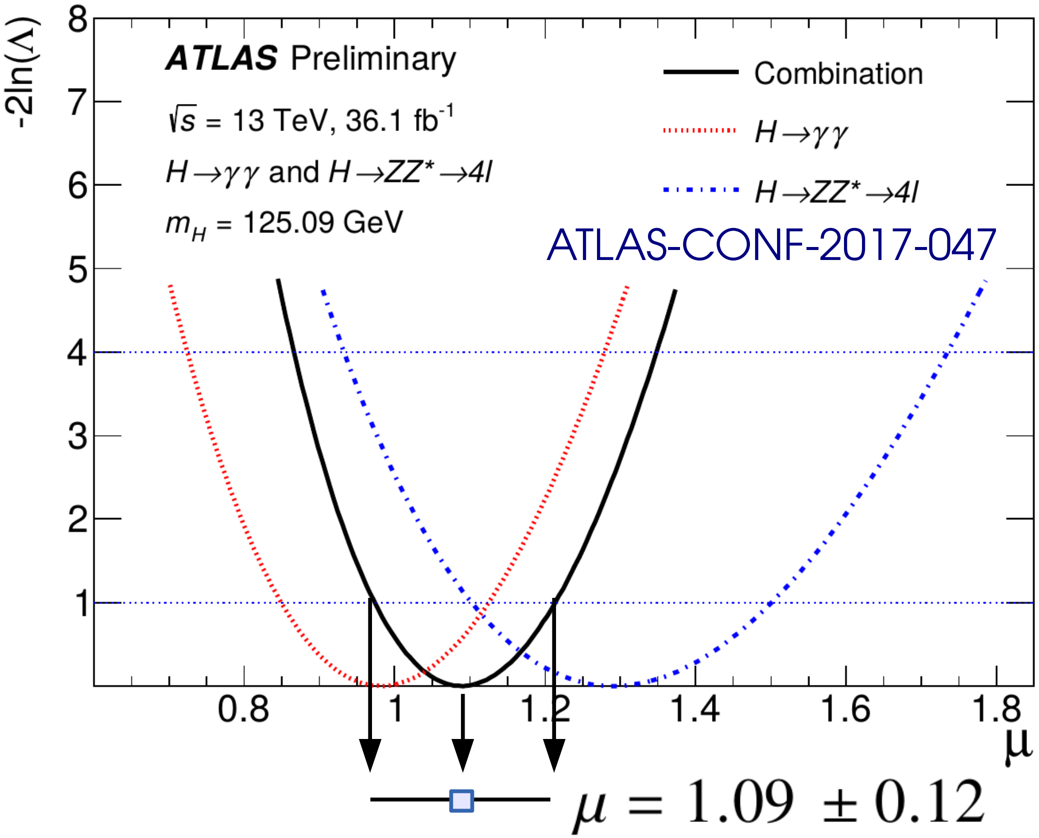
μ can be several POI!

Asymptotics:

- **$t_\mu \sim \chi^2(N_{POI})$ under H(μ_0)**
- **$\sqrt{t_\mu} \sim G(0,1)$** (Gaussian with d=N_{POI})

In practice:

- Plot t_μ vs. μ
- The minimum occurs at **$\mu = \hat{\mu}$**
- Crossings with **$t_\mu = Z^2$** give the **±Zσ uncertainties** (for N_{POI}=1)



ATLAS-CONF-2017-047

→ **Gaussian case**: parabolic profile, $t_\mu = \left(\frac{\mu-\hat{\mu}}{\sigma}\right)^2 \Rightarrow \mu_\pm = \hat{\mu} \pm \sigma$ at $t_\mu = 1$

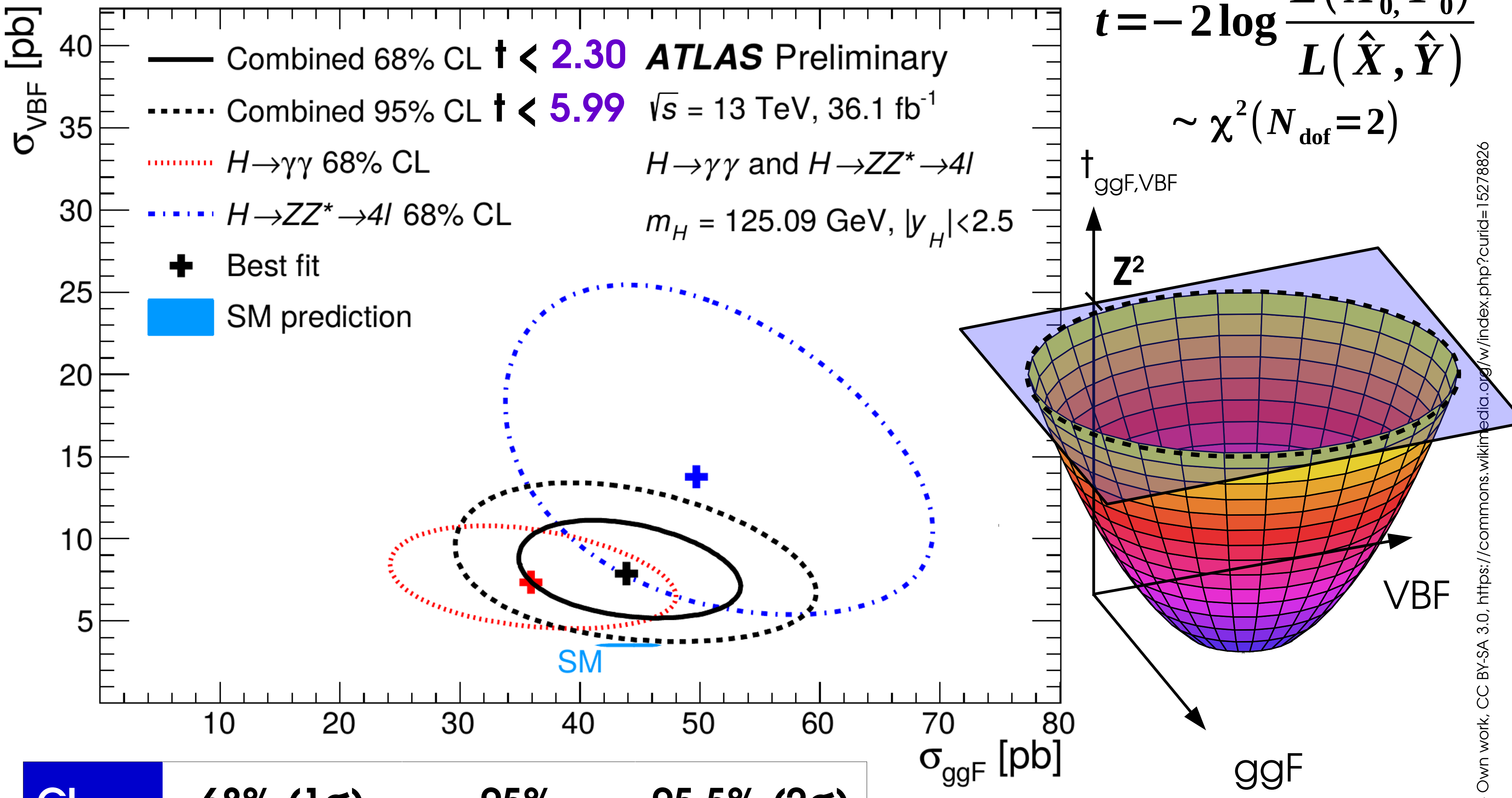
same result as Neyman construction, also robust against non-Gaussian effects.

2D Example: Higgs σ_{VBF} vs. σ_{ggF}

ATLAS-CONF-2017-047

$$t = -2\log\frac{L(X_{0}, Y_{0})}{L(\hat{X}, \hat{Y})}$$

$$\sim \chi^{2}(N_{\text{dof}}=2)$$



By Krishnavedala - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=15278826

CL	68% (1σ)	95%	95.5% (2σ)
1D Z^2	1	3.84	4
2D Z^2	2.30	5.99	6.18

Gaussian case: elliptic paraboloid surface

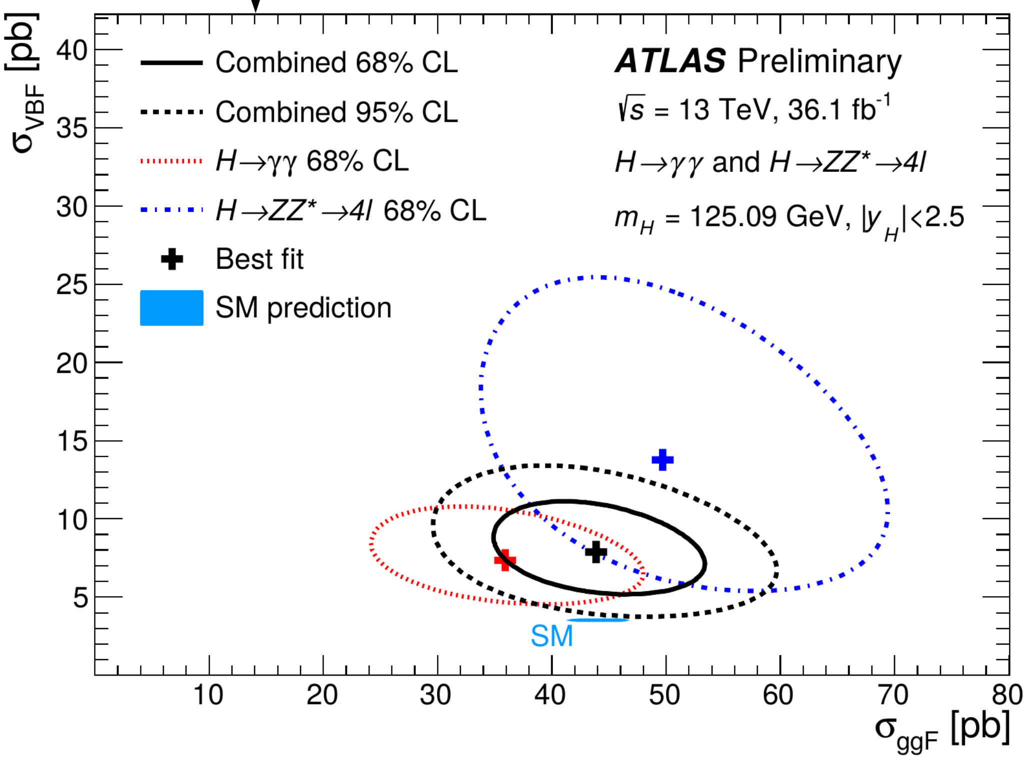
Reparameterization

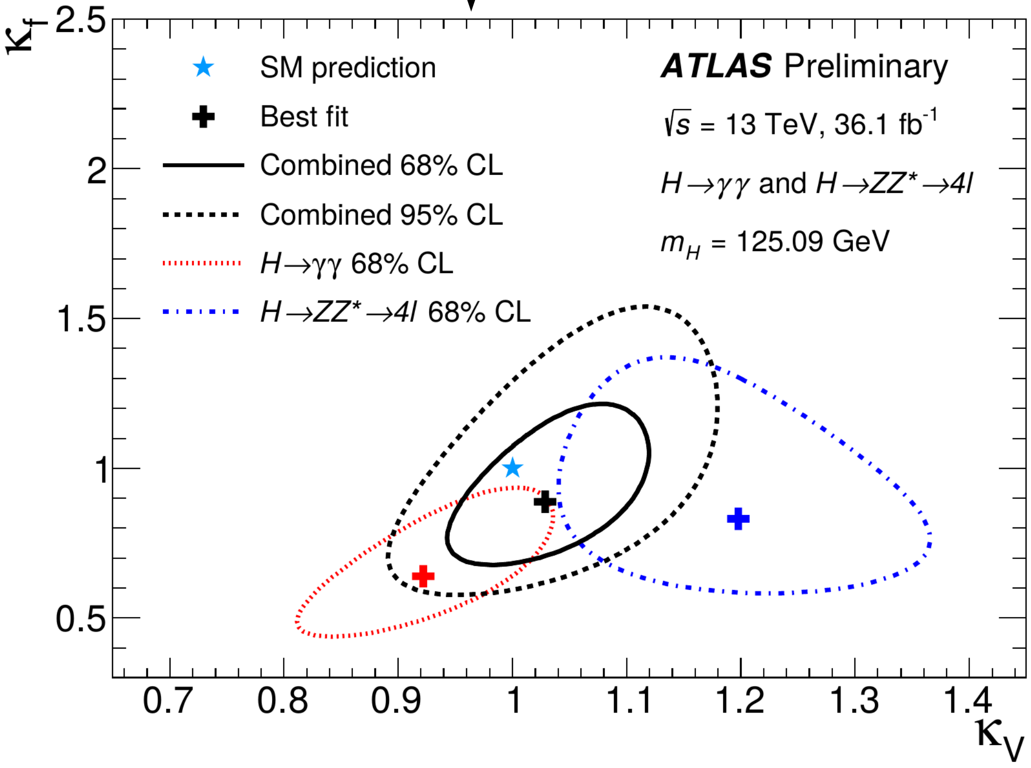
Start with basic measurement in terms of e.g. **σ×B**

→ How to measure derived quantities (couplings, parameters in some theory model, etc.) ? **→ just reparameterize the likelihood:**

e.g. Higgs couplings: σ_{ggF}, σ_{VBF} sensitive to Higgs coupling modifiers κ_V, κ_F.

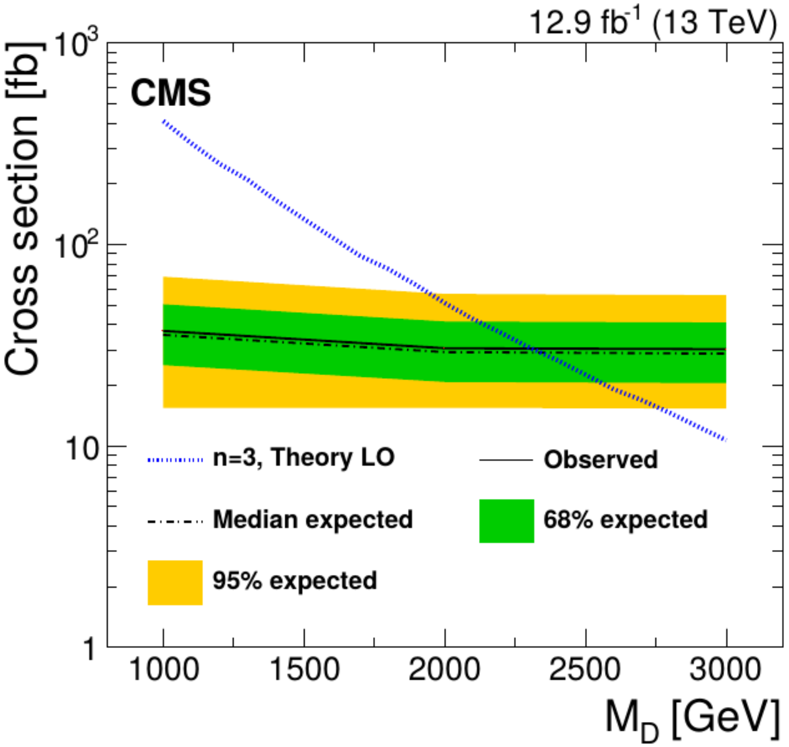
$$L(\sigma_{ggF}, \sigma_{VBF}) \xrightarrow[\sigma_{VBF} \to \sigma_{VBF}(\kappa_V, \kappa_F)]{\sigma_{ggF} \to \sigma_{ggF}(\kappa_V, \kappa_F)} L(\sigma_{ggF}(\kappa_V, \kappa_F), \sigma_{VBF}(\kappa_V, \kappa_F)) \equiv L'(\kappa_V, \kappa_F)$$

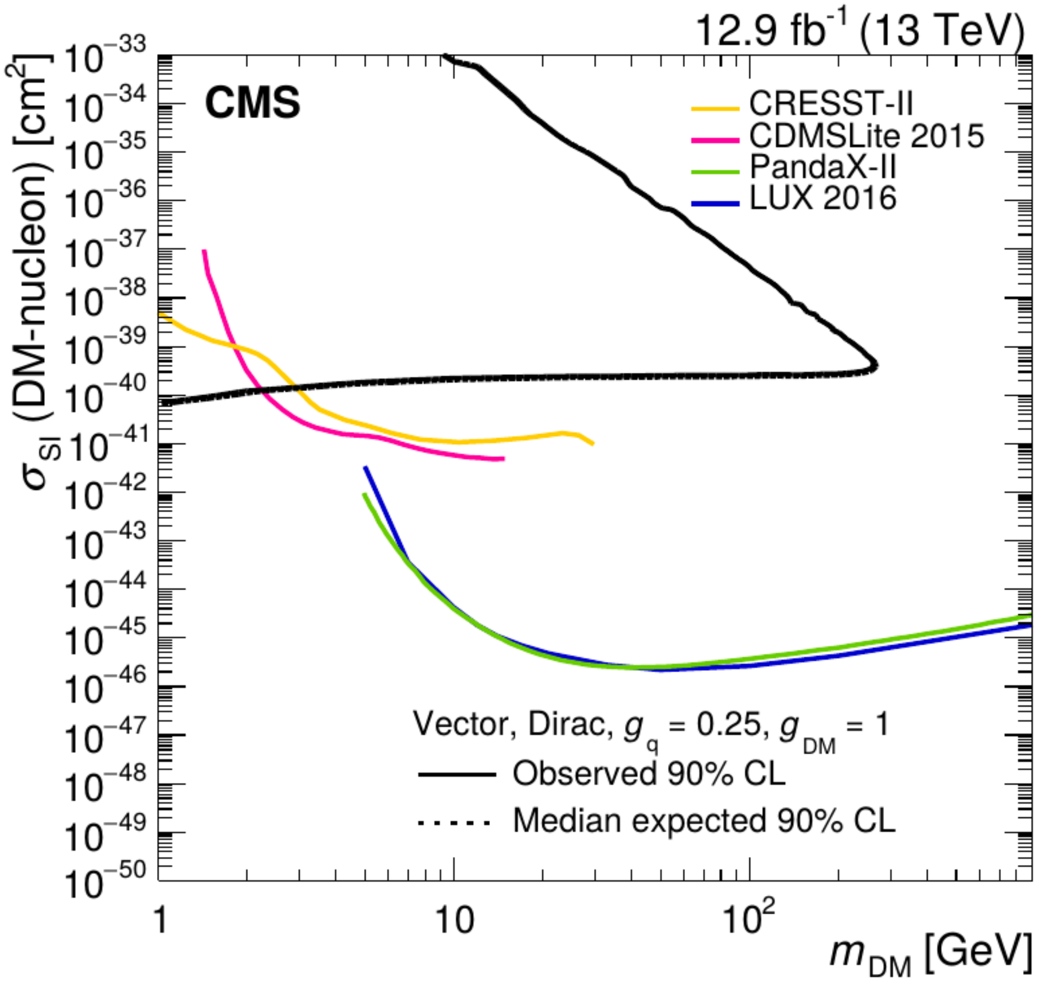


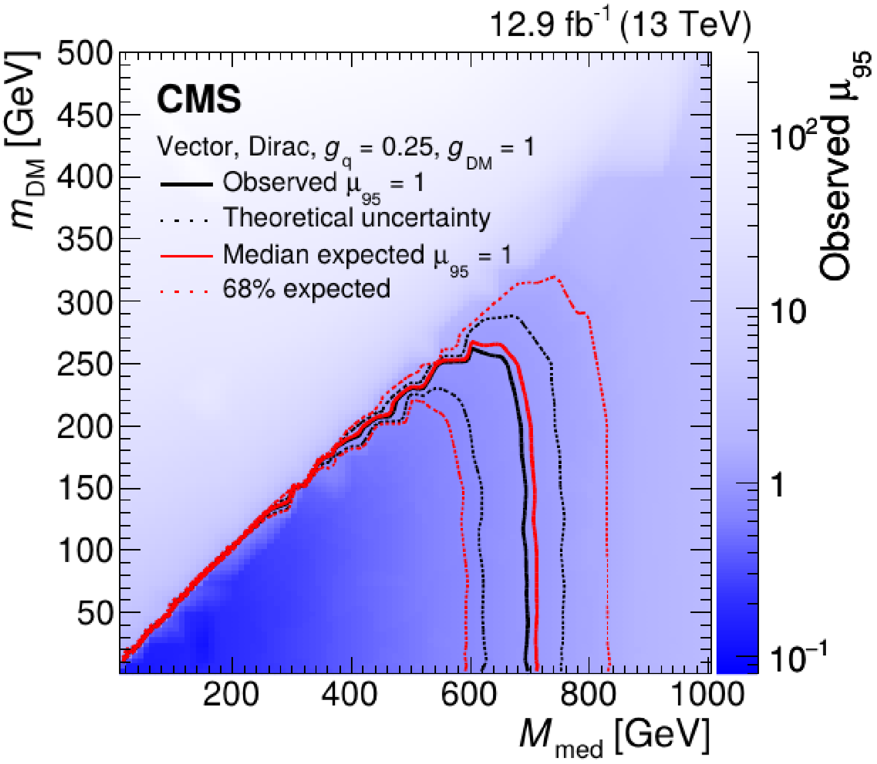


Reparameterization: Limits

CMS Run 2 Monophoton Search: measured **N_s** in a counting experiment reparameterized according to various DM models





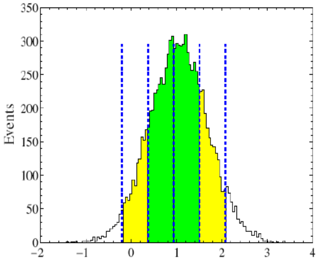


Takeaways

Limits : use LR-based test statistic:

$$\tilde{q}_{\mu_0} = \begin{cases} 0 & \hat{\mu} \geq \mu_0 \\ -2\log\dfrac{L(\mu=\mu_0)}{L(\hat{\mu})} & 0 \leq \hat{\mu} \leq \mu_0 \\ -2\log\dfrac{L(\mu=\mu_0)}{L(\mu=0)} & \hat{\mu} < 0 \end{cases}$$

→ Use **CL$_s$ procedure** to avoid negative limits



Poisson regime, n=0 : **S_{up} = 3 events**

Gaussian regime, n=0 : **S_{up} = 1.96 σ_{Gauss}**

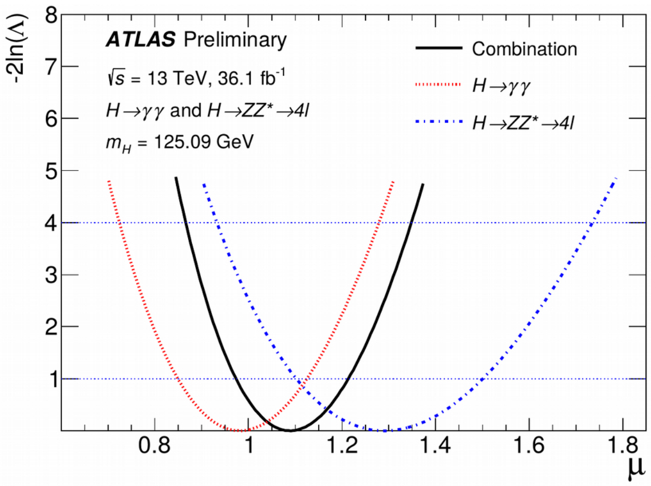
Uncertainty bands: obtain from toys or from Asimov

$$\sigma^2_{S,A} = \frac{S^2}{q_S(\text{Asimov})}$$

Confidence intervals: use $t_{\mu_0} = -2\log\dfrac{L(\mu=\mu_0)}{L(\hat{\mu})}$

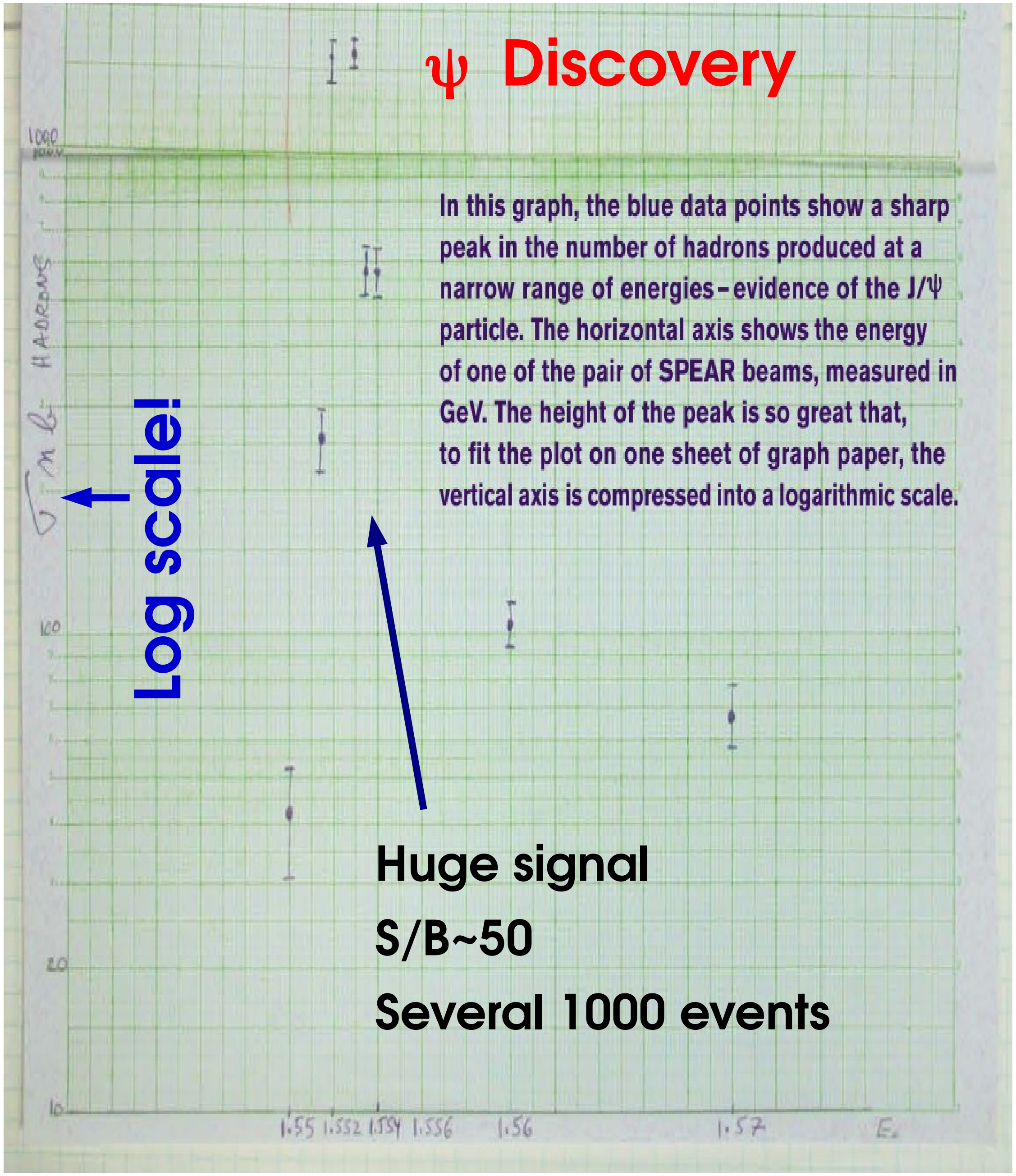
→ 1D: crossings with $t_{\mu 0} = Z^2$ for ±Zσ intervals

Gaussian regime: $\mu = \hat{\mu} \pm \sigma_{Gauss}$ (1σ interval)



Historical Aside

Classic Discoveries (1)



ψ Discovery

In this graph, the blue data points show a sharp peak in the number of hadrons produced at a narrow range of energies – evidence of the J/ψ particle. The horizontal axis shows the energy of one of the pair of SPEAR beams, measured in GeV. The height of the peak is so great that, to fit the plot on one sheet of graph paper, the vertical axis is compressed into a logarithmic scale.

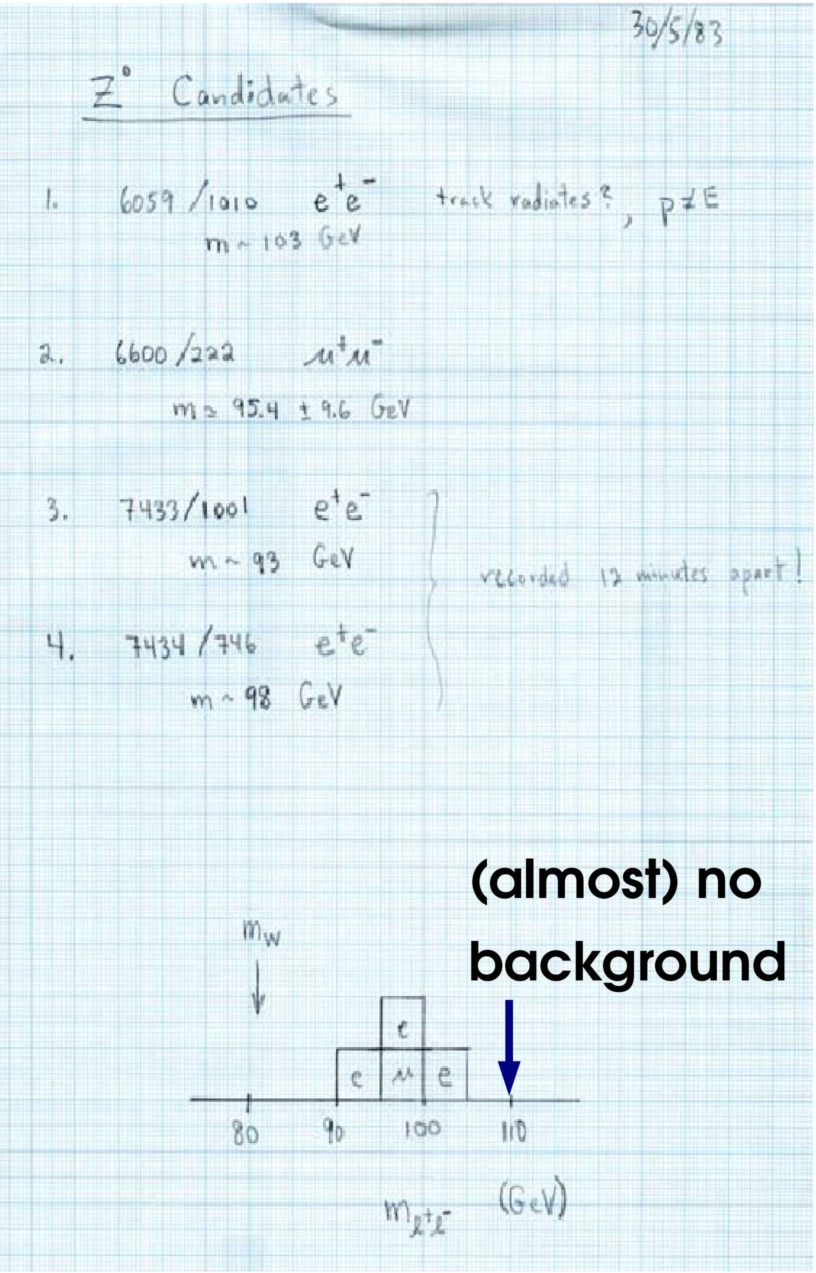
Log scale!

Huge signal

S/B~50

Several 1000 events

Z^0 Discovery



30/5/83

Z° Candidates

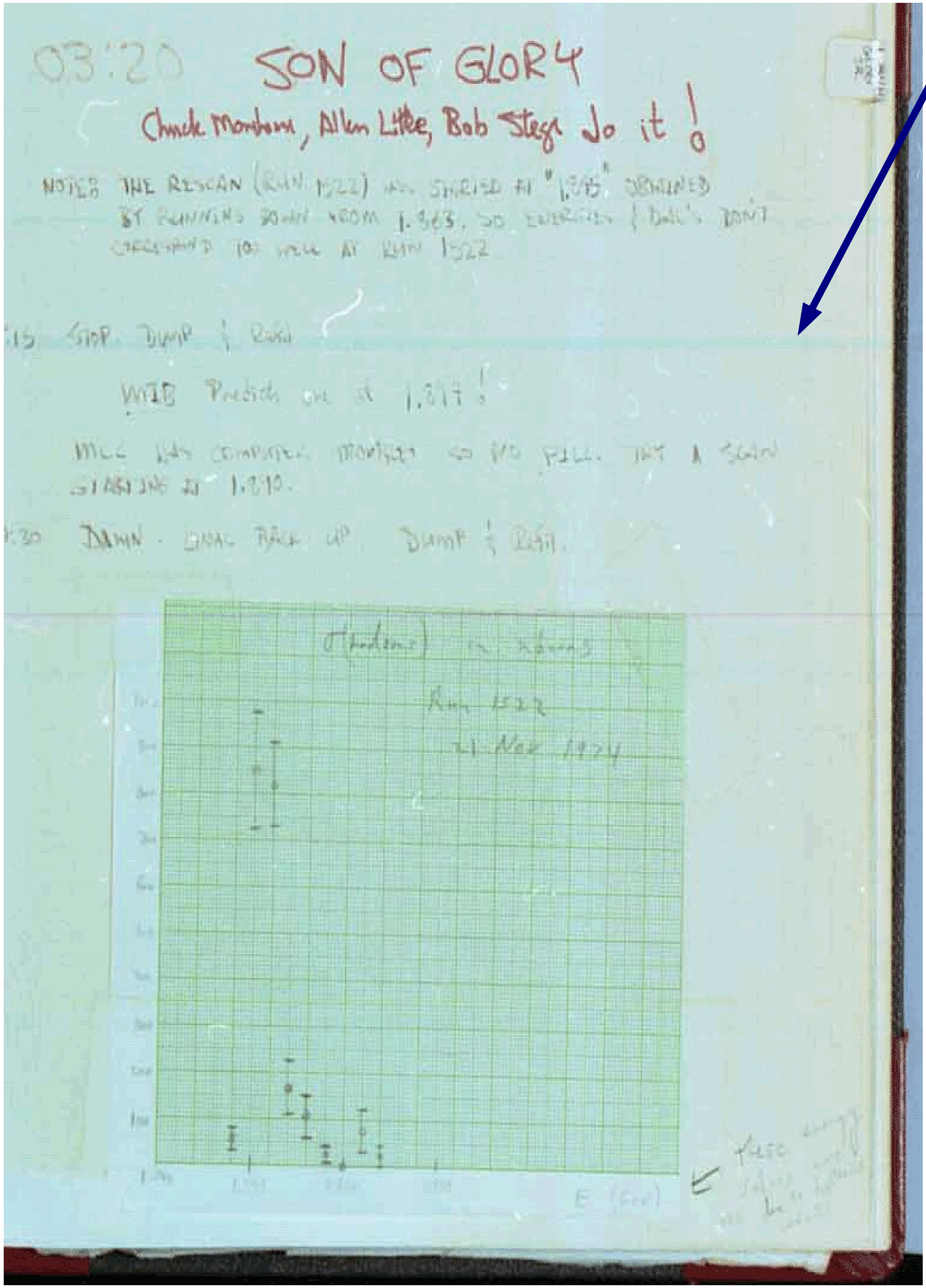
1. 6059 / 1010 e^+e^- track radiates ?, $p \neq E$
   $m \sim 103$ GeV
2. 6600 / 222 $\mu^+\mu^-$
   $m \simeq 95.4 \pm 9.6$ GeV
3. 7433/1001 e^+e^-
   $m \sim 93$ GeV
4. 7434 / 746 e^+e^-
   $m \sim 98$ GeV

recorded 12 minutes apart!

(almost) no background

Logbook of J. Rohlf, 1983-05-30

Classic Discoveries (2)



03:20 SON OF GLORY

Chuck Mondum, Allen Little, Bob Stega do it !

NOTE: THE RESCAN (RUN 1522) WAS STARTED AT "1.845" OBTAINED BY RUNNING DOWN FROM 1.863, SO ENERGIES & DAL'S DON'T CORRESPOND TOO WELL AT RUN 1522

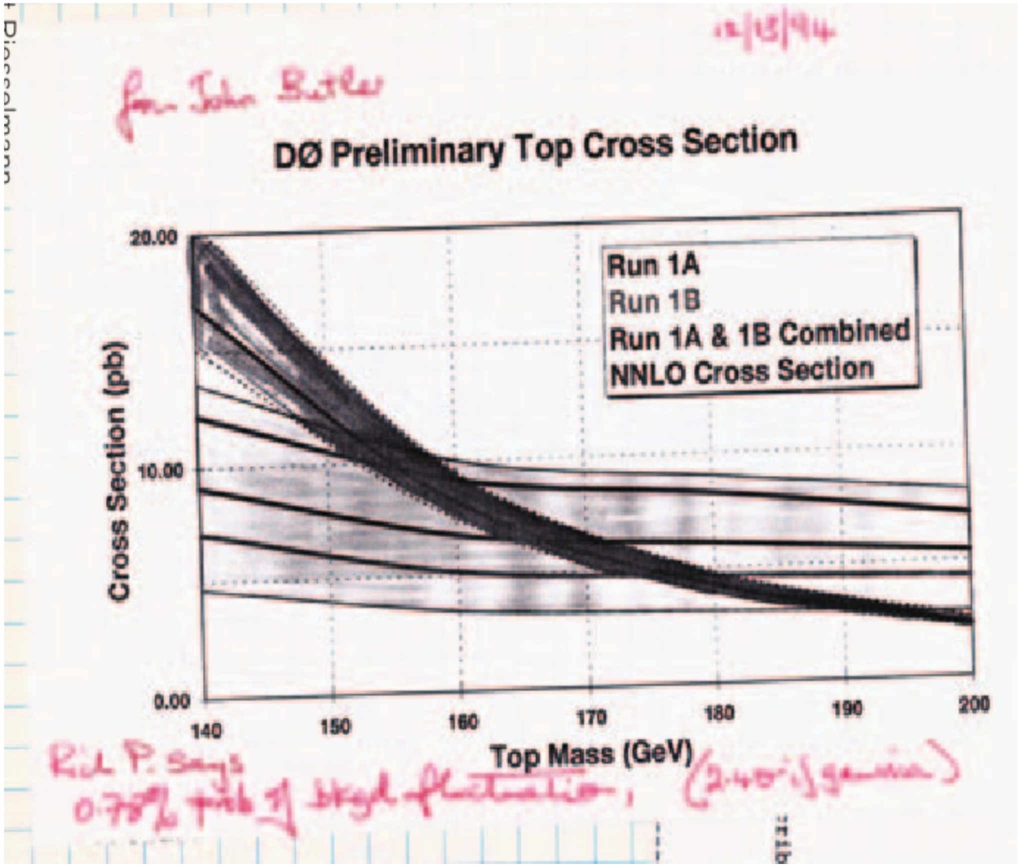
:15 STOP. DUMP & RAMP

WIB Predicts one at 1.817 !

MCC HAS COMPUTER TROUBLES SO NO FILL. TRY A SCAN STARTING AT 1.810.

:30 DAMN - LINAC BACK UP. DUMP & RAMP.

ψ' : discovered online by the (lucky) shifters



12/13/94

for John Butler

Rich P. says 0.78% prob of bkgd fluctuation, (2.4σ if genuine)

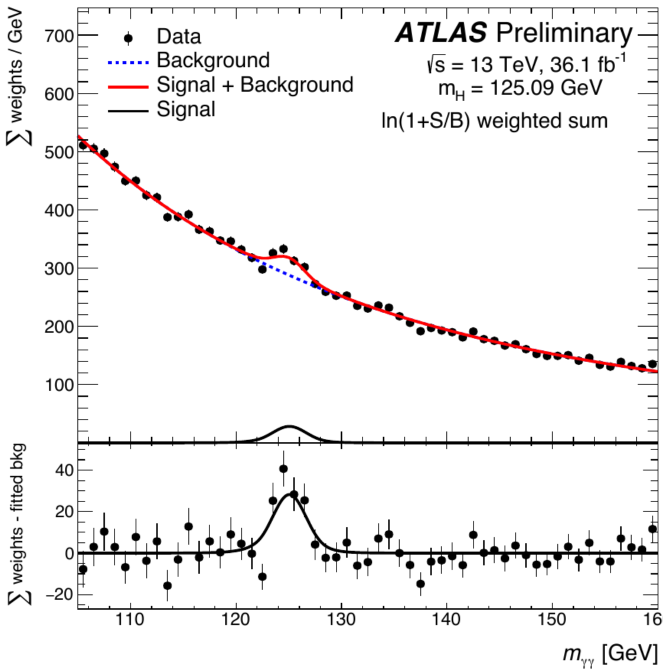
First hints of top at D0:
O(10) signal events,
a few bkg events, 2.4σ

And now ?

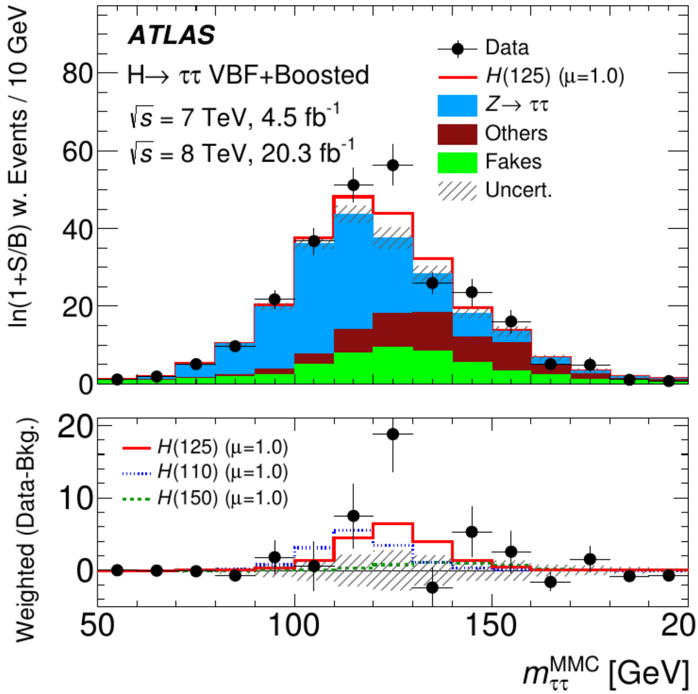
Short answer: The high-signal, low-background experiments have been done already (although a surprise would be welcome...)

e.g. at LHC:

- **High background levels**, need precise modeling
- **Large systematics,** need to be described accurately
- **Small signals**: need optimal use of available information :
  - **Shape analyses** instead of counting
  - **Categories** to isolated signal-enriched regions



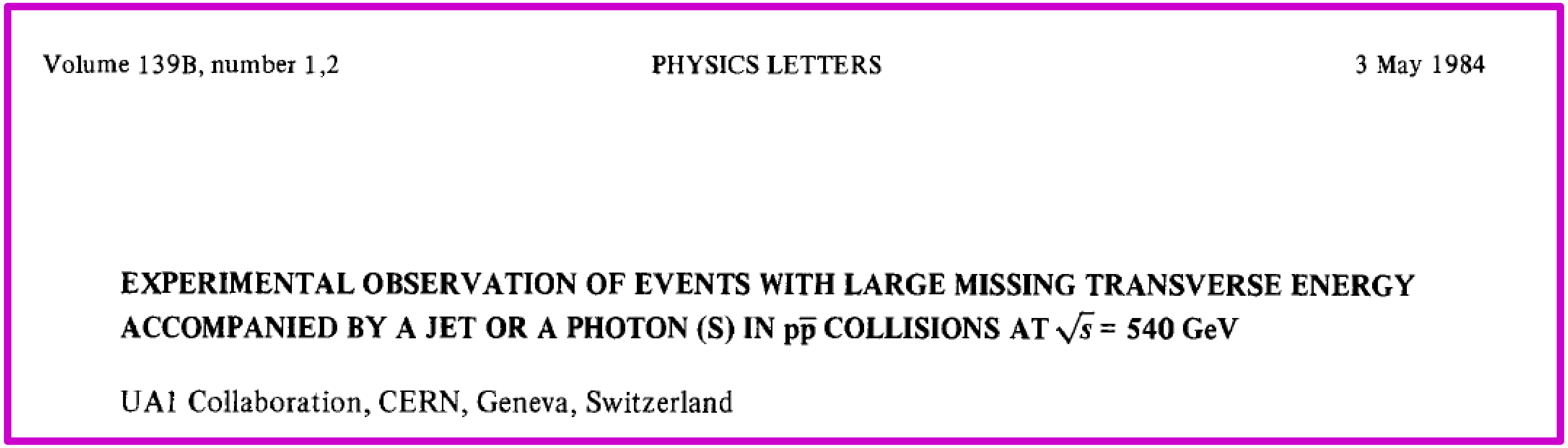
ATLAS-CONF-2017-045



JHEP 12 (2017) 024

Discoveries that weren't

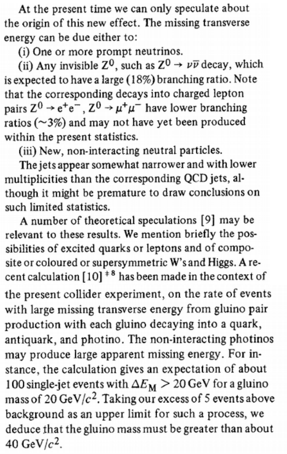
UA1 Monojets (1984)



Volume 139B, number 1,2 PHYSICS LETTERS 3 May 1984

EXPERIMENTAL OBSERVATION OF EVENTS WITH LARGE MISSING TRANSVERSE ENERGY ACCOMPANIED BY A JET OR A PHOTON (S) IN $p\bar{p}$ COLLISIONS AT $\sqrt{s}$ = 540 GeV

UA1 Collaboration, CERN, Geneva, Switzerland



At the present time we can only speculate about the origin of this new effect. The missing transverse energy can be due either to:

(i) One or more prompt neutrinos.

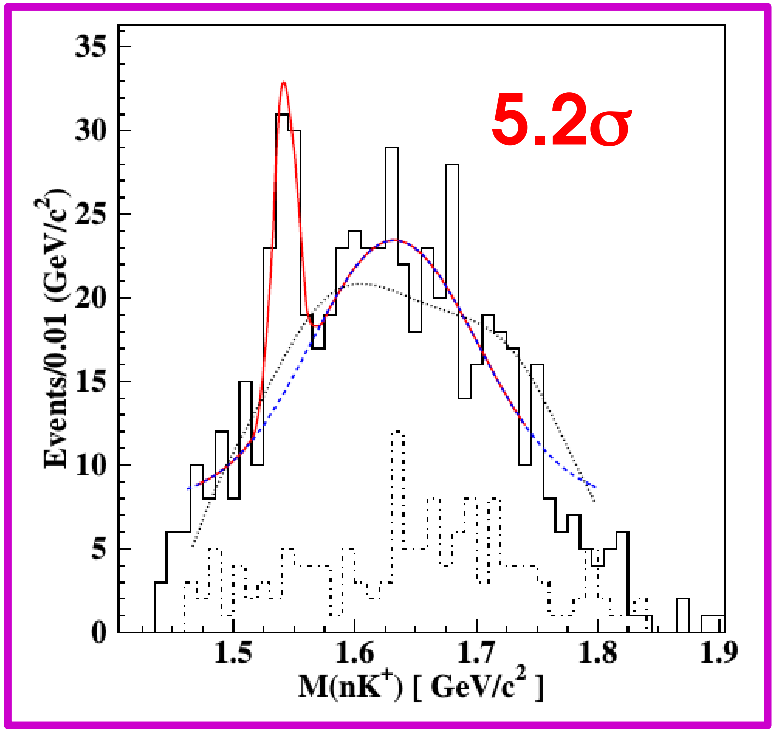
(ii) Any invisible Z^0, such as $Z^0 \to \nu\bar{\nu}$ decay, which is expected to have a large (18%) branching ratio. Note that the corresponding decays into charged lepton pairs $Z^0 \to e^+e^-$, $Z^0 \to \mu^+\mu^-$ have lower branching ratios (~3%) and may not have yet been produced within the present statistics.

(iii) New, non-interacting neutral particles.

The jets appear somewhat narrower and with lower multiplicities than the corresponding QCD jets, although it might be premature to draw conclusions on such limited statistics.

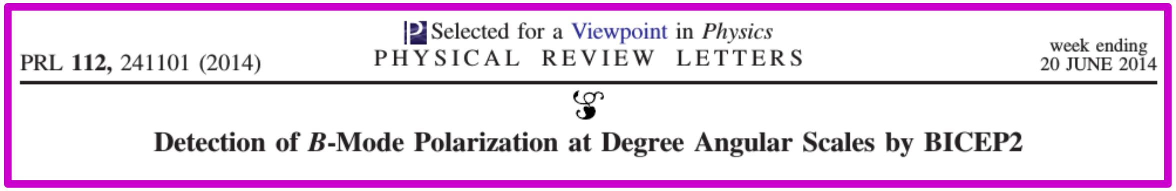
A number of theoretical speculations [9] may be relevant to these results. We mention briefly the possibilities of excited quarks or leptons and of composite or coloured or supersymmetric W's and Higgs. A recent calculation [10] ‡8 has been made in the context of the present collider experiment, on the rate of events with large missing transverse energy from gluino pair production with each gluino decaying into a quark, antiquark, and photino. The non-interacting photinos may produce large apparent missing energy. For instance, the calculation gives an expectation of about 100 single-jet events with $\Delta E_M > 20$ GeV for a gluino mass of 20 GeV/c^2. Taking our excess of 5 events above background as an upper limit for such a process, we deduce that the gluino mass must be greater than about 40 GeV/c^2.

Pentaquarks (2003)



Phys. Rev. Lett. 91, 252001 (2003)

BICEP2 B-mode Polarization (2014)



Selected for a Viewpoint in *Physics*

PRL **112,** 241101 (2014) PHYSICAL REVIEW LETTERS week ending 20 JUNE 2014

Detection of *B*-Mode Polarization at Degree Angular Scales by BICEP2

$$r = 0.20^{+0.07}_{-0.05}, \text{ with } r = 0 \text{ disfavored at } 7.0\sigma.$$

Avoid spurious discoveries!

→ Treatment of modeling uncertainties, systematics in general

Outline

Lecture I:

- Statistics basics
- Describing measurements
- Computing statistics results:

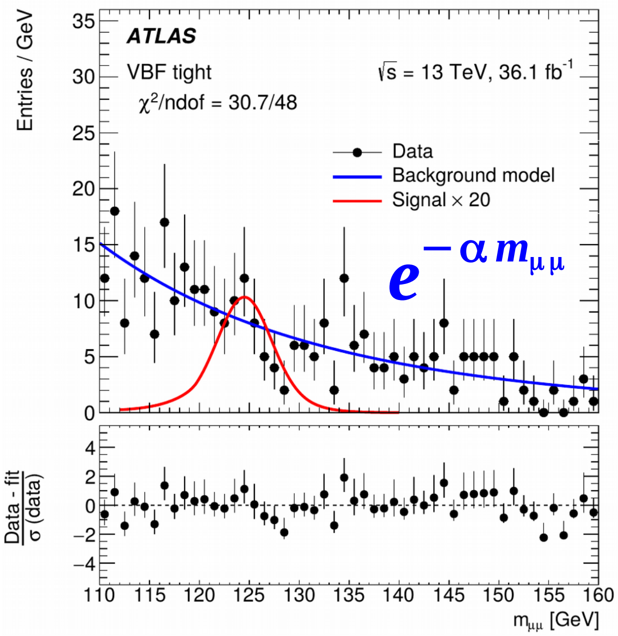
Today:

- Computing statistics results:
  - Discovery
  - Limits
  - Confidence intervals
- **Profiling**

Profiling

Nuisances and Systematics

Phys. Rev. Lett. 119 (2017) 051802



Likelihood typically includes

- **Parameters of interest** (POIs) : **S, σ×B, m_W**, …
- **Nuisance parameters** (NPs) : other parameters needed to define the model

  → Ideally, **constrained by data** like the POI

  e.g. shape of H→μμ continuum bkg

What about systematics ?

= what we don't know about the random processs

⇒ **Parameterize using additional NPs**

→ By definition, **not constrained by the data**

⇒ Cannot be free, or would spoil the measurement (lumi free ⇒ no σ×B measurement!)

⇒ **Introduce a constraint in the likelihood:**

> "Systematic uncertainty is, in any statistical inference procedure, the uncertainty due to the incomplete knowledge of the probability distribution of the observables.
>
> G. Punzi, ***What is systematics ?***

$$L(\mu, \theta; \text{data}) = L_{\text{measurement}}(\mu, \theta; \text{data})\ C(\theta)$$

POI

Systematics NP

Measurement Likelihood

NP Constraint term

⇒ penalty for $\theta \neq \theta^{\text{nominal}}$

Frequentist Constraints

Prototype: NP measured in a separate ***auxiliary* experiment**
e.g. luminosity measurement

→ Build the combined likelihood of the main+auxiliary measurements

$$L(\mu,\theta;\mathbf{data}) = L_{\mathbf{main}}(\mu,\theta;\mathbf{main\ data})\ L_{\mathbf{aux}}(\theta;\mathbf{aux.\ data})$$

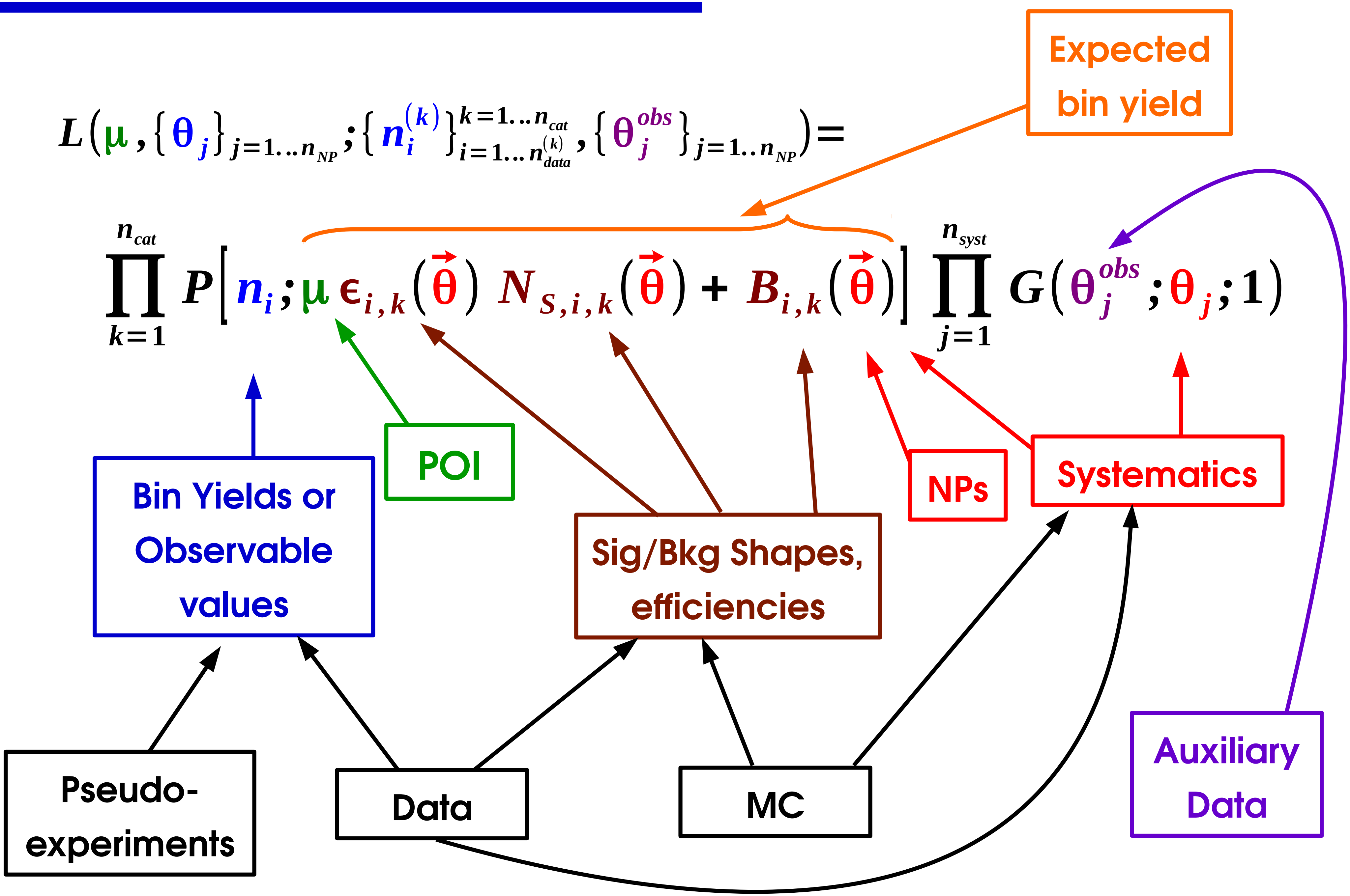
Independent measurements:
⇒ just a product

Gaussian form often used by default: $L_{\mathbf{aux}}(\theta;\mathbf{aux.\ data}) = G(\theta^{\mathbf{obs}};\theta,\sigma_{\mathbf{syst}})$

In the combined likelihood, **systematic NPs are constrained**
→ now same as other NPs: **all uncertainties statistical in nature**

→ Often no clear setup for auxiliary measurements
e.g. theory uncertainties on missing HO terms from scale variations
→ Implemented in the same way nevertheless ("pseudo-measurement")

Likelihood, the full version (binned case)



$$L\left(\mu,\{\theta_j\}_{j=1\ldots n_{NP}};\{n_i^{(k)}\}_{i=1\ldots n_{data}^{(k)}}^{k=1\ldots n_{cat}},\{\theta_j^{obs}\}_{j=1..n_{NP}}\right)=$$

$$\prod_{k=1}^{n_{cat}} P\left[n_i;\overbrace{\mu\,\epsilon_{i,k}(\vec{\theta})\,N_{S,i,k}(\vec{\theta})+B_{i,k}(\vec{\theta})}\right]\prod_{j=1}^{n_{syst}} G(\theta_j^{obs};\theta_j;1)$$

Expected bin yield

POI

NPs

Systematics

Bin Yields or Observable values

Sig/Bkg Shapes, efficiencies

Pseudo-experiments

Data

MC

Auxiliary Data

× number of categories!

Wilks' Theorem, the unabridged version

The likelihood usually has NPs:

- **Systematics**
- Parameters fitted in data

→ **What values to use when defining the hypotheses ?** → H(μ=0, θ=?)

Answer: **let the data choose ⇒ use the best-fit values (*Profiling*)**

⇒ **Profile Likelihood Ratio** (PLR)

$$t_{\mu_0} = -2\log\frac{L(\mu=\mu_0, \hat{\hat{\theta}}_{\mu_0})}{L(\hat{\mu}, \hat{\theta})}$$

$\hat{\hat{\theta}}_{\mu_0}$ best-fit value for $\mu=\mu_0$ (conditional MLE)

$\hat{\theta}$ overall best-fit value (unconditional MLE)

Wilks' Theorem: PLR also follows a χ² !

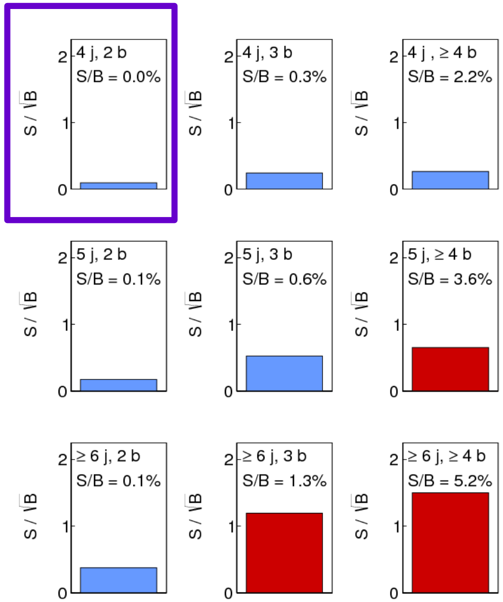
$$f\left(t_{\mu_0} \mid \mu=\mu_0\right) = f_{\chi^2(n_{dof}=1)}\left(t_{\mu_0}\right)$$

also with NPs present

→ Profiling "builds in" the effect of the NPs

⇒ Can treat the PLR as a **function of the POI only**

Profiling Example: ttH→bb

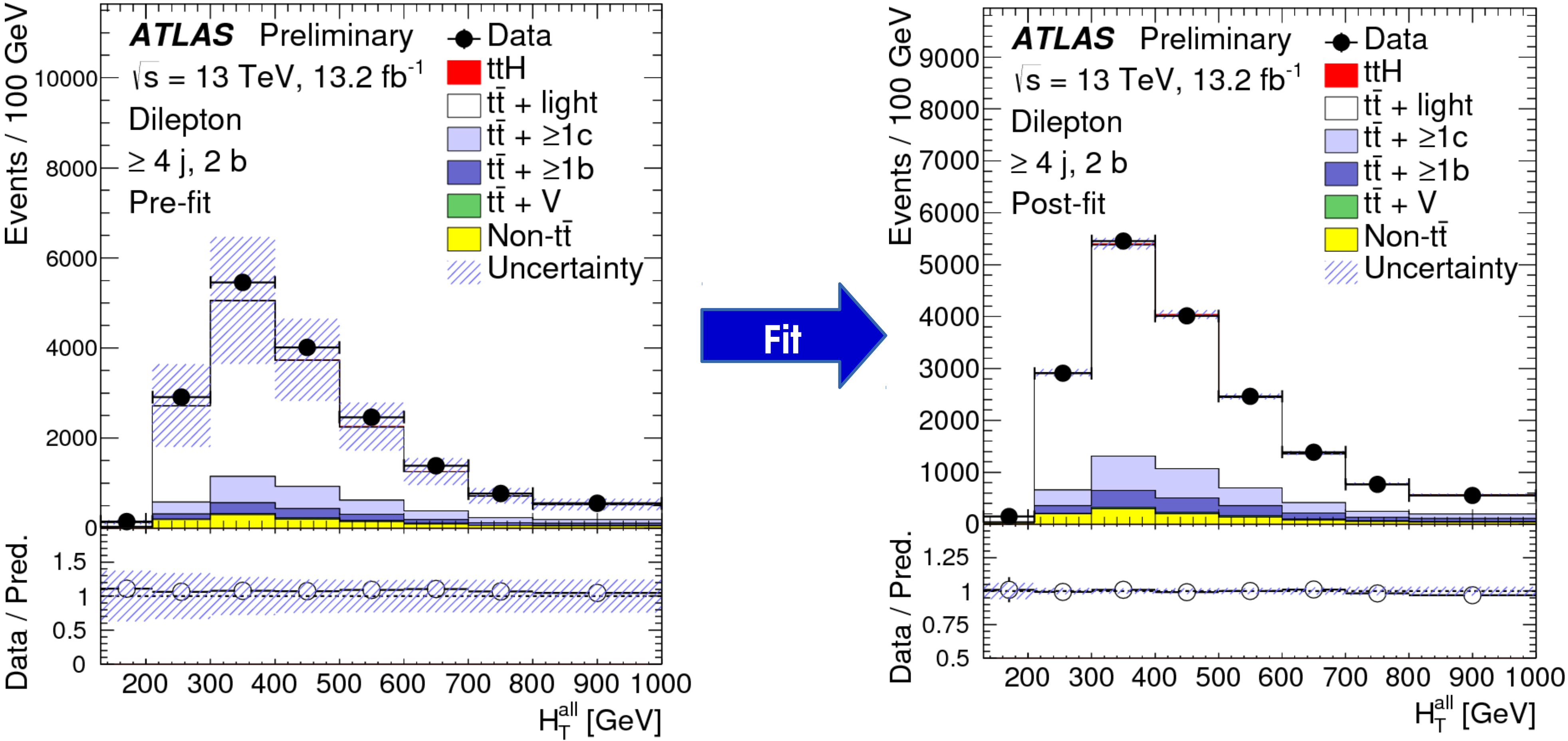


Profiled parameters fixed by aux. meas. + data : here CRs

→ **Reduction in large uncertainties on tt bkg**

→ **Propagates to the high-S/B categories** through the statistical modeling

⇒ **Care needed in the propagation** (e.g. different kinematic regimes)

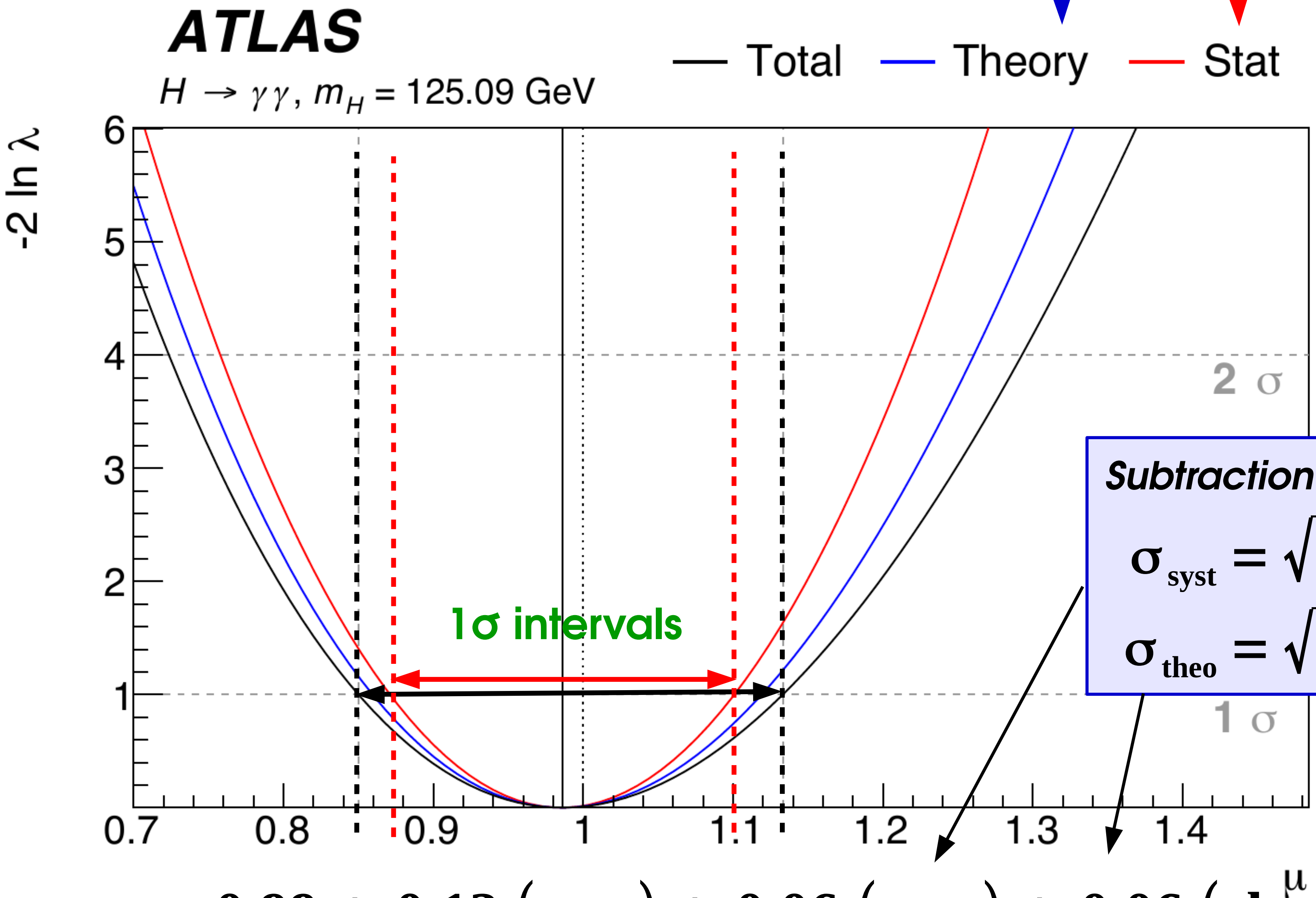


ATLAS-CONF-2016-08

Uncertainty decomposition

All systematics NPs fixed to 0 : statistical uncertainty only

exp. syst. NPs fixed to 0 : stat+theory uncertainty



Subtraction in quadrature

$$\sigma_{\text{syst}} = \sqrt{\sigma^2_{\text{total}} - \sigma^2_{\text{stat}}}$$

$$\sigma_{\text{theo}} = \sqrt{\sigma^2_{\text{stat+theo}} - \sigma^2_{\text{stat}}}$$

$$\mu = 0.99 \pm 0.12\,(\text{stat}) \pm 0.06\,(\text{syst}) \pm 0.06\,(\text{theo})$$

Pull/Impact plots

ATLAS-CONF-2016-058

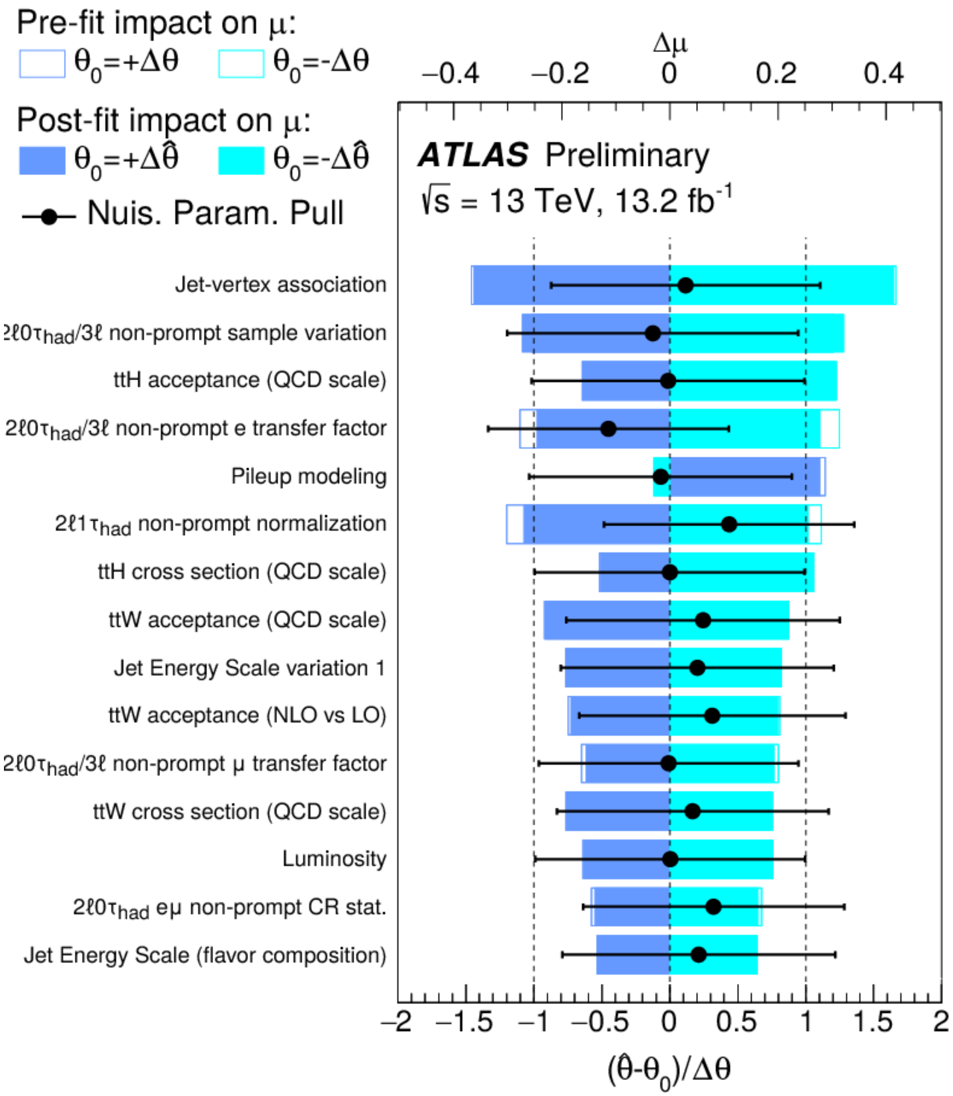
Systematics are described by NPs included in the fit. Nominally:

- **NP central value = 0** : corresponds to the pre-fit expectation (usually MC)
- **NP uncertainty = 1** : since NPs normalized to the value of the syst. :

$$N = N_0 \left(1 + \sigma_{\text{syst}}\ \theta\right),\ \theta \sim G(0,\ 1)$$

Fit results provide information on impact of the systematic on the result:

- **If central value ≠ 0**: some data feature absorbed by nonzero value ⇒ Need investigation if large pull
- **If uncertainty < 1** : systematic is constrained by the data ⇒ Needs checking if this legitimate or a modeling issue
- **Impact on result** of ±1σ shift of NP



Pull/Impact plots

Systematics are described by NPs included in the fit. Nominally:

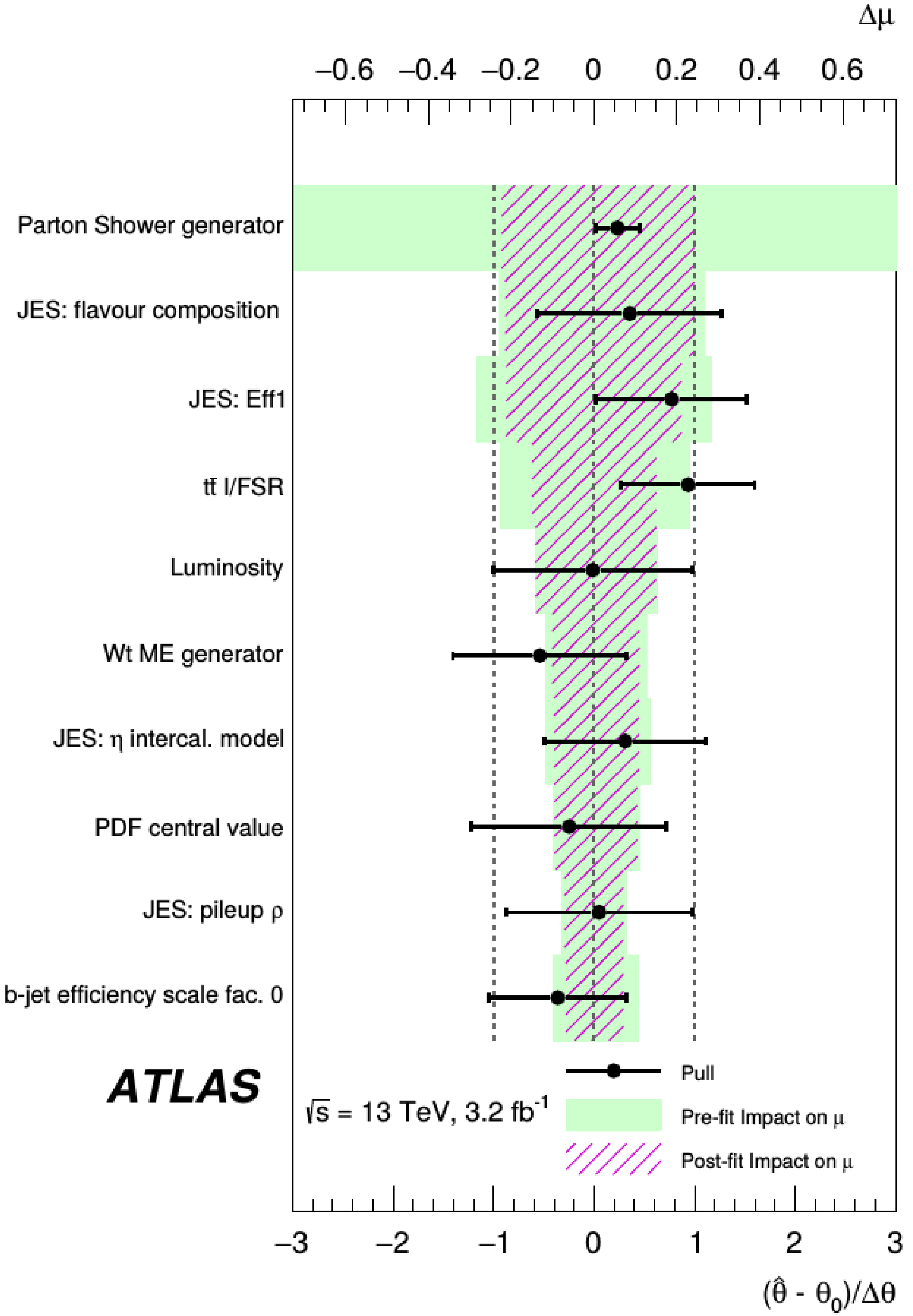
- **NP central value = 0** : corresponds to the pre-fit expectation (usually MC)
- **NP uncertainty = 1** : since NPs normalized to the value of the syst. :

$$N = N_0\left(1 + \sigma_{\text{syst}}\,\theta\right),\ \theta \sim G(0,\,1)$$

Fit results provide information on impact of the systematic on the result:

- **If central value ≠ 0**: some data feature absorbed by nonzero value ⇒ Need investigation if large pull
- **If uncertainty < 1** : systematic is constrained by the data ⇒ Needs checking if this legitimate or a modeling issue
- **Impact on result** of ±1σ shift of NP

13 TeV single-t XS (arXiv:1612.07231)



Takeaways

Systematics: uncertainties on the **form of the statistical model** (as opposed to the uncertainties encoded in the model itself)

→ Implemented using additional nuisance parameters in the model

→ Constrained by adding ***auxiliary measurements*** (sometimes fictitious ones) to the model – usually represented by a single Gaussian for each NP.

$$L(\mu, \theta; \mathrm{data}) = L_{\mathrm{main}}(\mu, \theta; \mathrm{main\ data})\ G(\theta^{\mathrm{obs}}, \theta, 1)$$

⇒ **Systematics treated in the same way as statistical uncertainties**, although we still keep track of ***systematics NPs*** for bookkeeping purposes

Profiling: when testing a hypothesis, use the best-fit values of the nuisance parameters: ***profile likelihood ratio***.

$$\frac{L(\mu=\mu_0, \hat{\hat{\theta}}_{\mu_0})}{L(\hat{\mu}, \hat{\theta})}$$

Wilks' Theorem: the PLR has the same asymptotic properties as the LR without systematics: can profile out NPs and just deal with POIs.

→ NPs still show up in the PLR as increased uncertainties – Gaussian case:

$$\sigma_{\mathrm{total}} = \sqrt{\sigma_{\mathrm{stat}}^2 + \sigma_{\mathrm{syst}}^2}$$

Profiling can have unintended effects – need to carefully check behavior

Summary of Statistical Results Computation

Methods provide:

→ **Optimal use of information from the data under general hypotheses**

→ **Arbitrarily complex/realistic models (up to computing constraints...)**

→ **No Gaussian assumptions in the measurements**

Still often assume Gaussian behavior of PLR – but weaker assumption and can be lifted with toys

Systematics treated as auxiliary measurements – modeling can be tailored as needed

→ **Single PLR-based framework for all usual classes of measurements**

Discovery testing

Upper limits on signal yields

Parameter estimation

Beyond Asymptotics: Toys

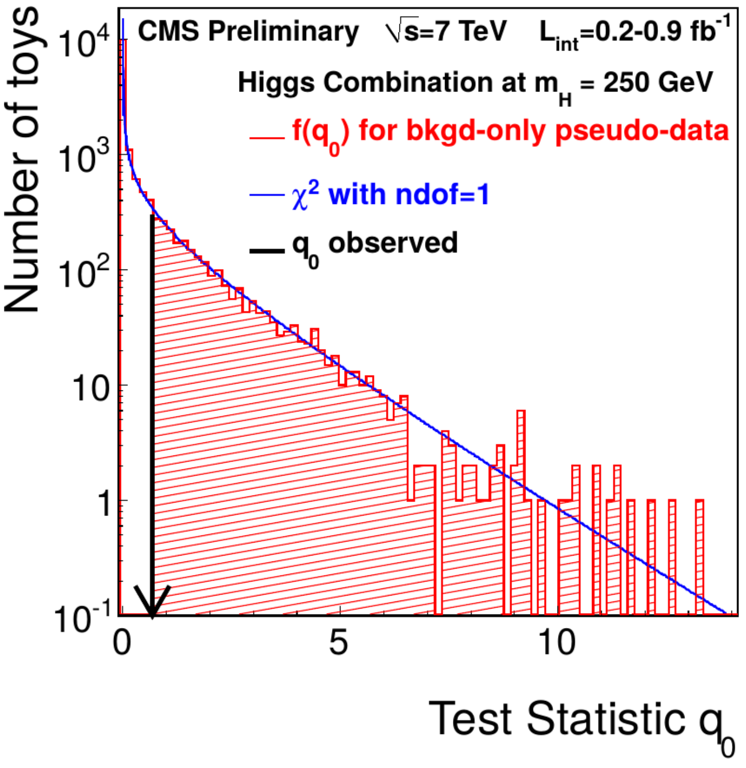
CMS-PAS-HIG-11-022

Asymptotics usually work well, but break down in some cases – e.g. **small event counts.**

Solution: generate ***pseudo data*** (**toys**) using the PDF, under the tested hypothesis

→ Also randomize the observable (θ^{obs}) of each auxiliary experiment: $G(\theta^{obs};\theta,\sigma_{syst})$

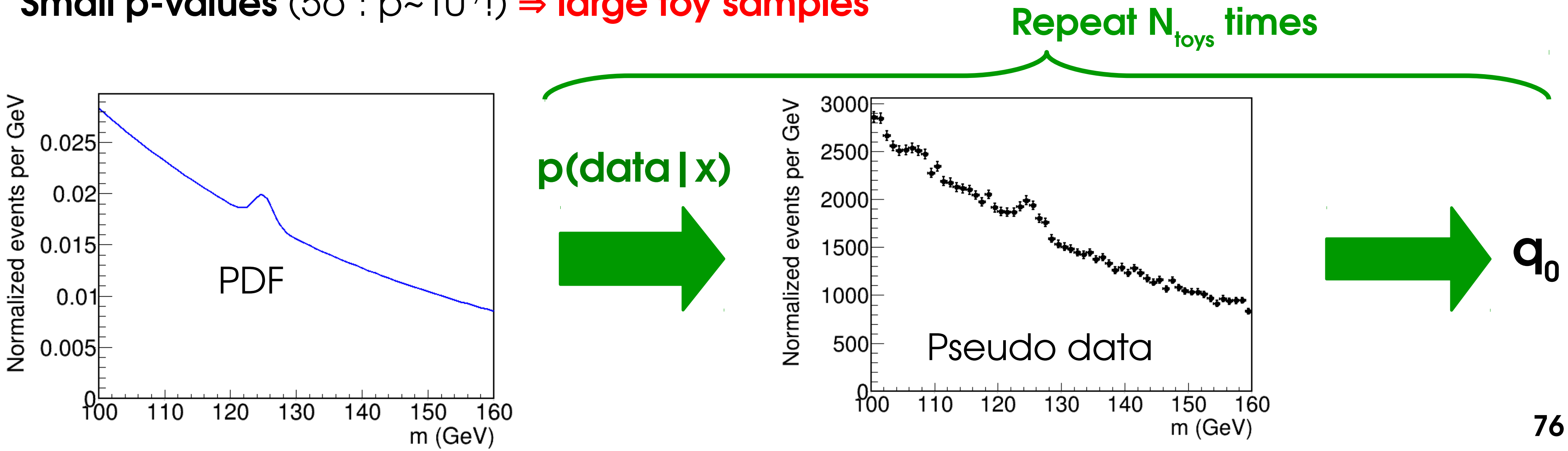
→ Samples the true distribution of the PLR



⇒ Integrate above observed PLR to get the p-value

→ Precision limited by number of generated toys,

Small p-values (5σ : p~10^{-7}!) ⇒ **large toy samples**



Repeat N_{toys} times

p(data | x)

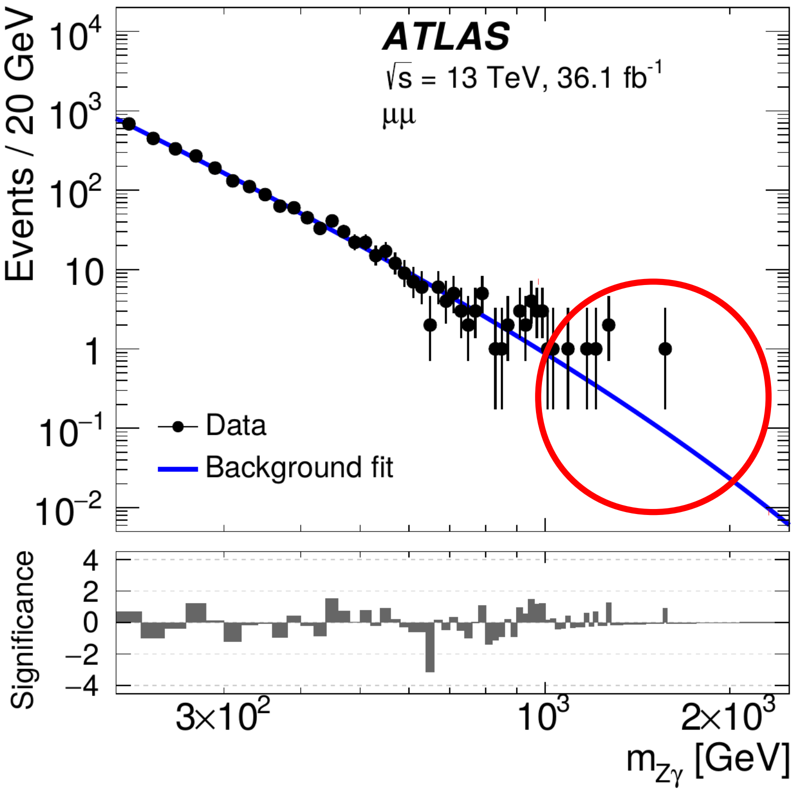
q_0

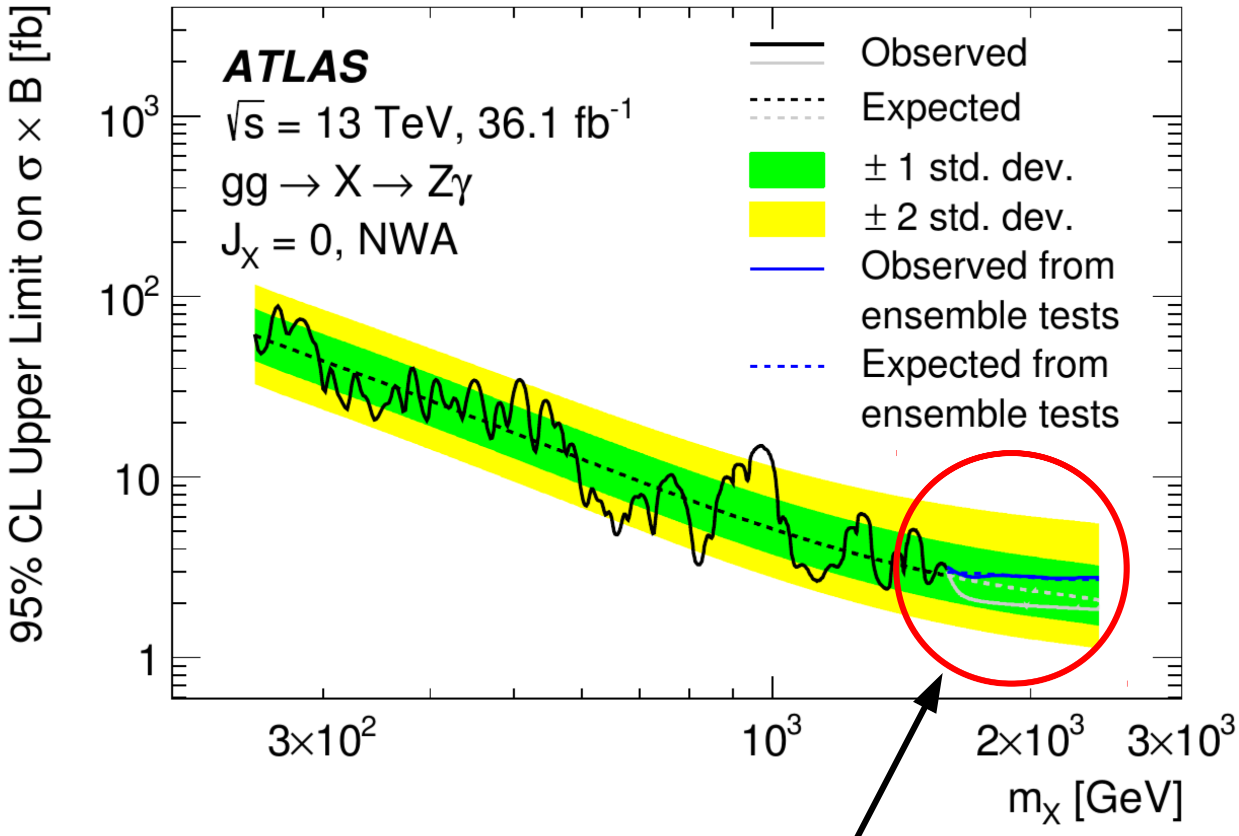
Toys: Example

arXiv:1708.00212

ATLAS X→Zγ Search: covers 200 GeV < m_X < 2.5 TeV

→ for m_X > 1.6 TeV, low event counts ⇒ derive results from toys





Asymptotic results (in gray) give optimistic result compared to toys (**in blue**)

Outline

Lecture I:

- **Statistics basics**
- **Describing measurements**
- **Computing statistics results:**

Today:

- **Computing statistics results:**
  - Discovery
  - Limits
  - Confidence intervals
- **Profiling**

Lecture III: Look-Elsewhere Effect, Bayesian methods, Practical modeling, BLUE

Extra Slides

Gaussian Profiling

Gaussian measurement with 1 POI μ and 1 NP θ:

$$L(\mu, \theta; \underbrace{\hat{\mu}, \hat{\theta}}_{\text{“data”}}) = \exp\left[-\frac{1}{2}\begin{pmatrix}\mu-\hat{\mu}\\ \theta-\hat{\theta}\end{pmatrix}^T C^{-1}\begin{pmatrix}\mu-\hat{\mu}\\ \theta-\hat{\theta}\end{pmatrix}\right] \qquad C = \begin{bmatrix}\sigma_\mu^2 & \gamma\sigma_\mu\sigma_\theta\\ \gamma\sigma_\mu\sigma_\theta & \sigma_\theta^2\end{bmatrix}$$

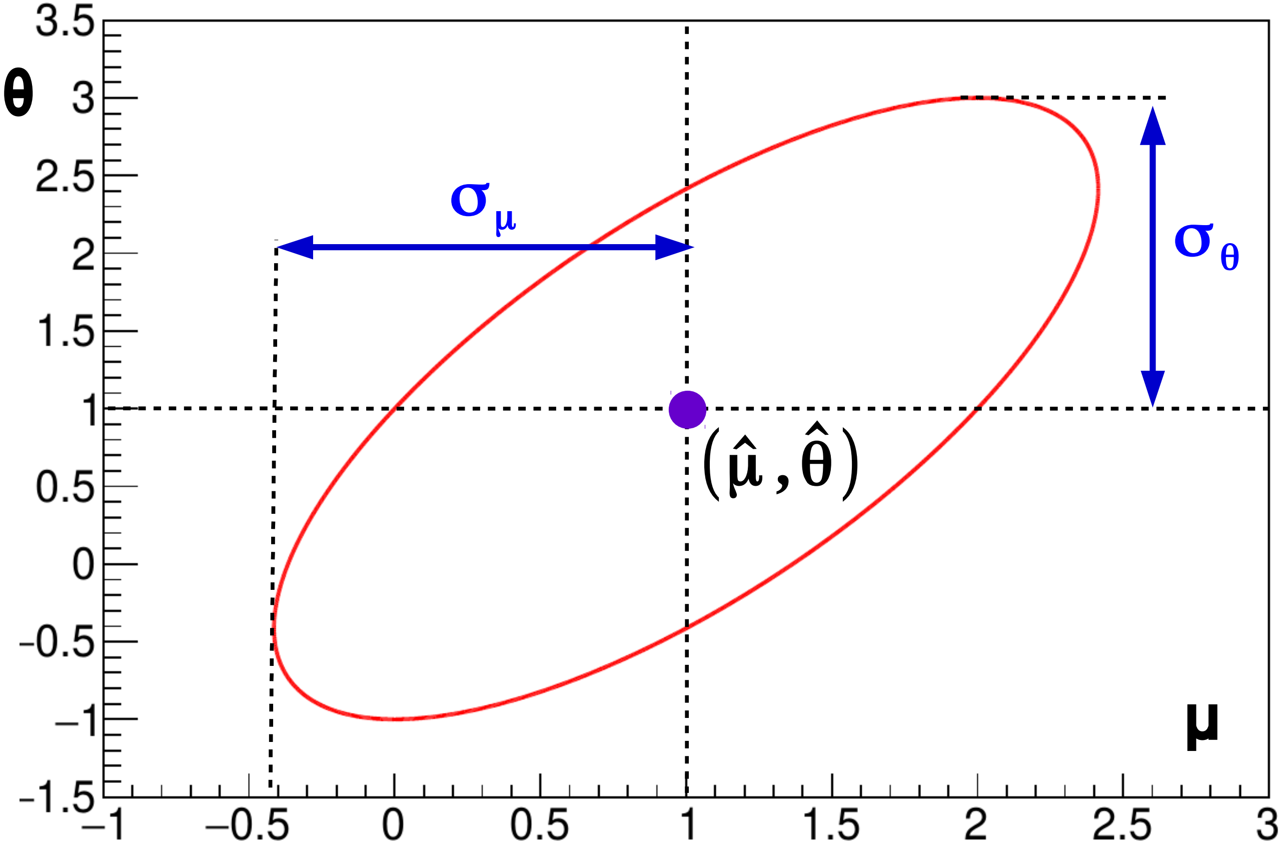
→ λ(μ, θ) defines an **ellipse:**

$$F \equiv C^{-1} = \begin{bmatrix}F_{\mu\mu} & F_{\mu\theta}\\ F_{\mu\theta} & F_{\theta\theta}\end{bmatrix}$$

$$\lambda(\mu, \theta; \hat{\mu}, \hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 + 2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta}) + F_{\theta\theta}(\theta-\hat{\theta})^2$$

Uncertainty on μ:

- **From C, with θ included:** σ_μ



Gaussian Profiling

$$C = \begin{bmatrix} \sigma_\mu^2 & \gamma\sigma_\mu\sigma_\theta \\ \gamma\sigma_\mu\sigma_\theta & \sigma_\theta^2 \end{bmatrix}$$

$$F = \begin{bmatrix} F_{\mu\mu} & F_{\mu\theta} \\ F_{\mu\theta} & F_{\theta\theta} \end{bmatrix}$$

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 + 2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta}) + F_{\theta\theta}(\theta-\hat{\theta})^2$$

Profiled θ (minimize λ at fixed μ) :

$$\hat{\hat{\theta}}(\mu) = \hat{\theta} - F_{\theta\theta}^{-1}F_{\theta\mu}(\mu-\hat{\mu})$$

Profile likelihood ratio:

$$\lambda(\mu,\hat{\hat{\theta}}(\mu);\hat{\mu},\hat{\theta}) = \left(F_{\mu\mu} - F_{\mu\theta}F_{\theta\theta}^{-1}F_{\theta\mu}\right)(\mu-\hat{\mu})^2 = C_{\mu\mu}^{-1}(\mu-\hat{\mu})^2 = \left(\frac{\mu-\hat{\mu}}{\sigma_\mu}\right)^2$$

$$\uparrow$$

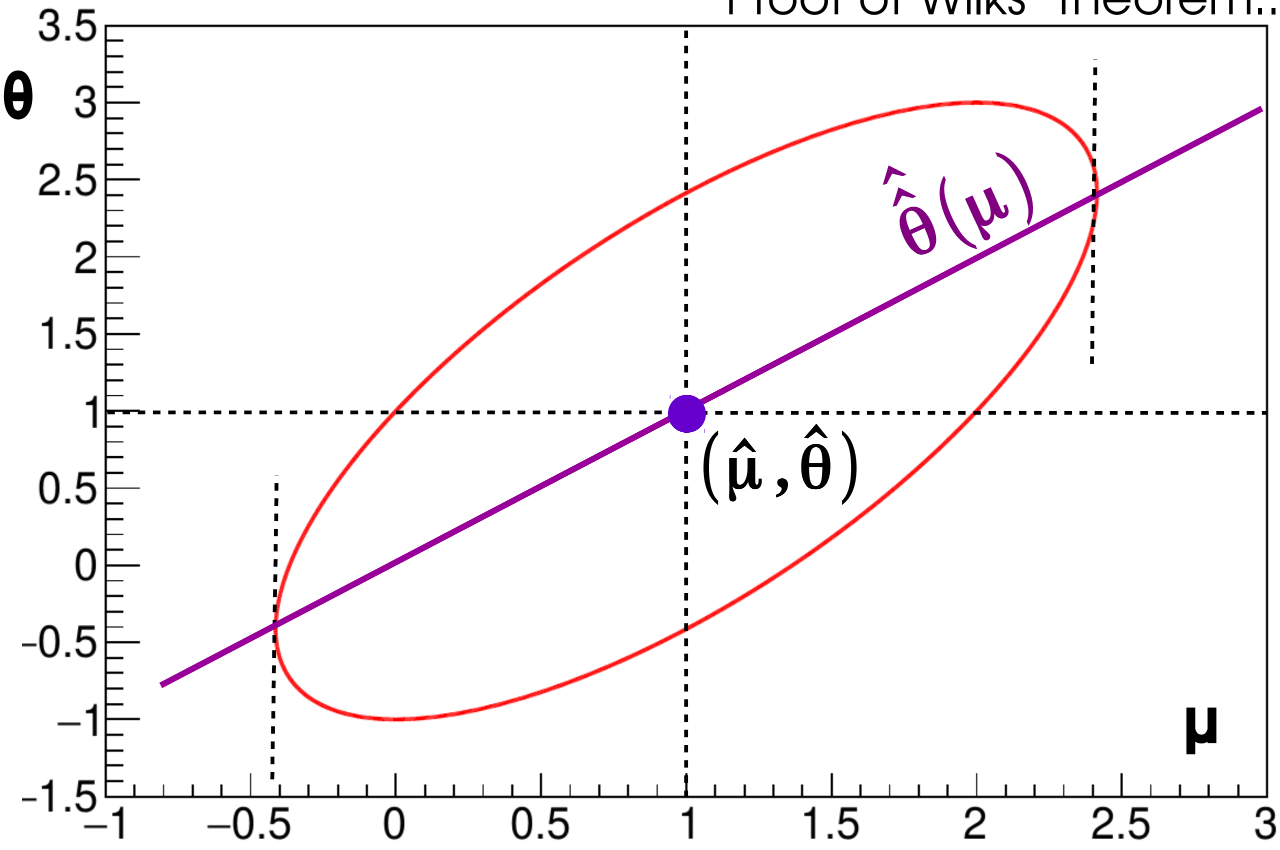
$$F_{\mu\mu} \neq C_{\mu\mu}^{-1} \text{ !!}$$

Uncertainty on μ:

- **From C:** σ_μ
- **From PLR:** σ_μ

Profiled θ **crosses ellipse at vertical tangents** by definition (L is lower at other points on the tangent)

Proof of Wilks' theorem...



Gaussian Profiling

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2+2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta})+F_{\theta\theta}(\theta-\hat{\theta})^2$$

$$F \equiv C^{-1} = \frac{1}{1-\gamma^2}\begin{bmatrix} \frac{1}{\sigma_\mu^2} & \frac{\gamma}{\sigma_\mu\sigma_\theta} \\ \frac{\gamma}{\sigma_\mu\sigma_\theta} & \frac{1}{\sigma_\theta^2} \end{bmatrix}$$

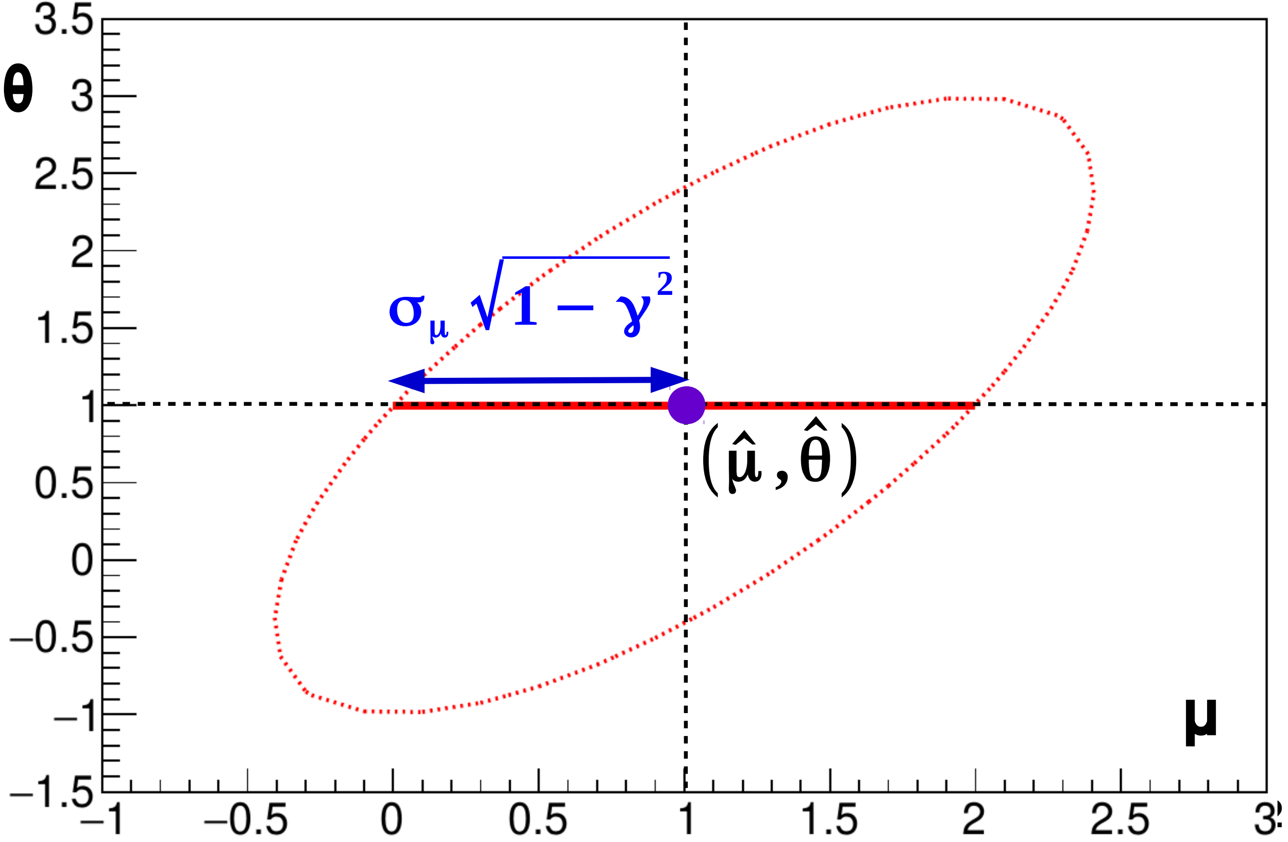
→ For fixed $\theta = \hat{\theta}$, $\lambda(\mu)$ defines an interval:

$$\lambda(\mu,\theta=\hat{\theta};\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 = \left(\frac{\mu-\hat{\mu}}{\sigma_\mu\sqrt{1-\gamma^2}}\right)^2$$

Uncertainty on μ:

- **From C:** σ_μ
- **From PLR:** σ_μ

- **From λ(μ):** $\sigma_\mu\sqrt{1-\gamma^2}$



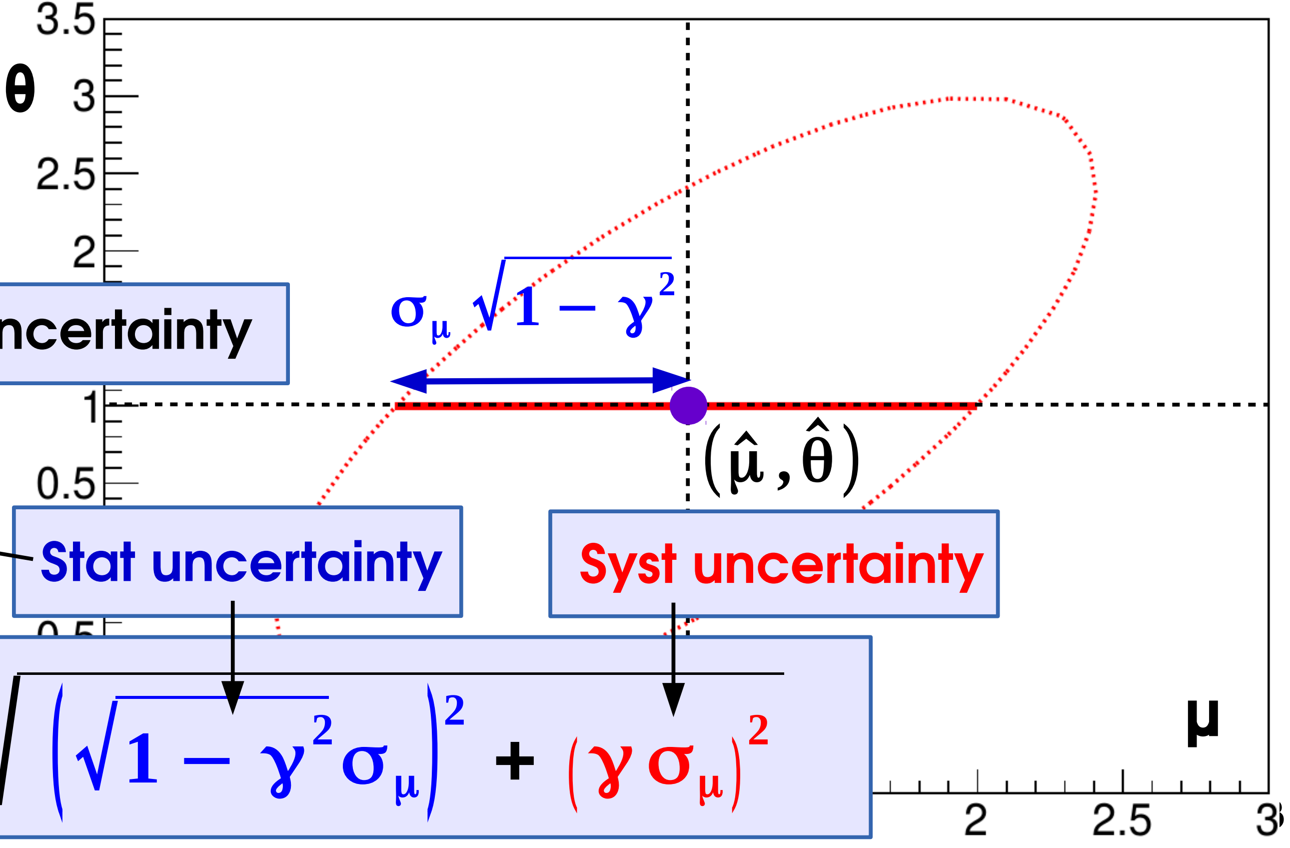
Gaussian Profiling

$$\lambda(\mu,\theta;\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2+2F_{\mu\theta}(\mu-\hat{\mu})(\theta-\hat{\theta})+F_{\theta\theta}(\theta-\hat{\theta})^2$$

$$F \equiv C^{-1} = \frac{1}{1-\gamma^2}\begin{bmatrix}\frac{1}{\sigma_\mu^2} & \frac{\gamma}{\sigma_\mu\sigma_\theta}\\ \frac{\gamma}{\sigma_\mu\sigma_\theta} & \frac{1}{\sigma_\theta^2}\end{bmatrix}$$

→ For fixed θ = $\hat{\theta}$, λ(μ) defines an interval:

$$\lambda(\mu,\theta=\hat{\theta};\hat{\mu},\hat{\theta}) = F_{\mu\mu}(\mu-\hat{\mu})^2 = \left(\frac{\mu-\hat{\mu}}{\sigma_\mu\sqrt{1-\gamma^2}}\right)^2$$



Uncertainty on μ:

- **From C:** σ_μ
- **From PLR:** σ_μ

Total uncertainty

- **From λ(μ):** $\sigma_\mu\sqrt{1-\gamma^2}$

Stat uncertainty

Syst uncertainty

$$\sigma_\mu = \sqrt{\left(\sqrt{1-\gamma^2}\,\sigma_\mu\right)^2 + (\gamma\,\sigma_\mu)^2}$$