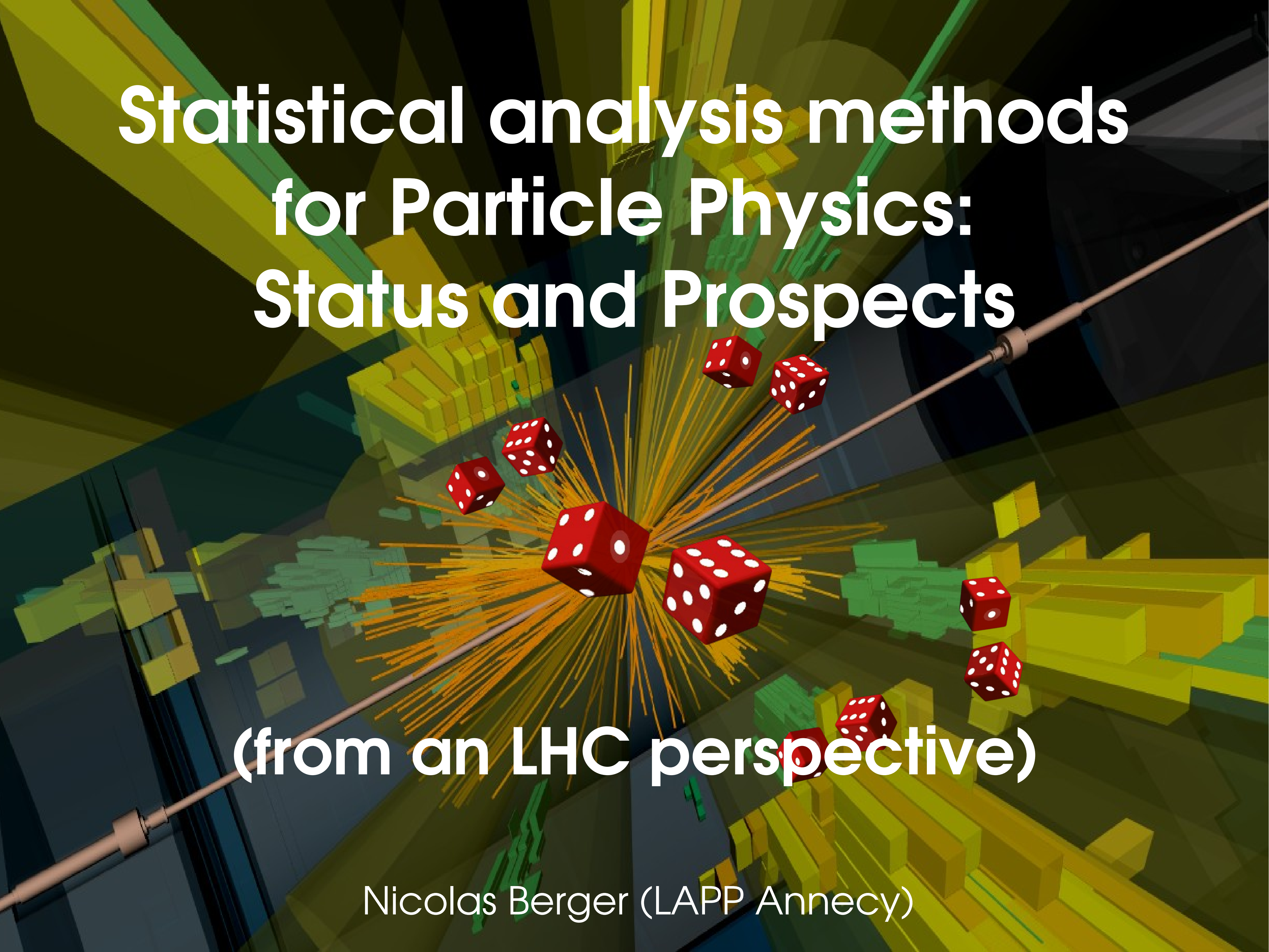
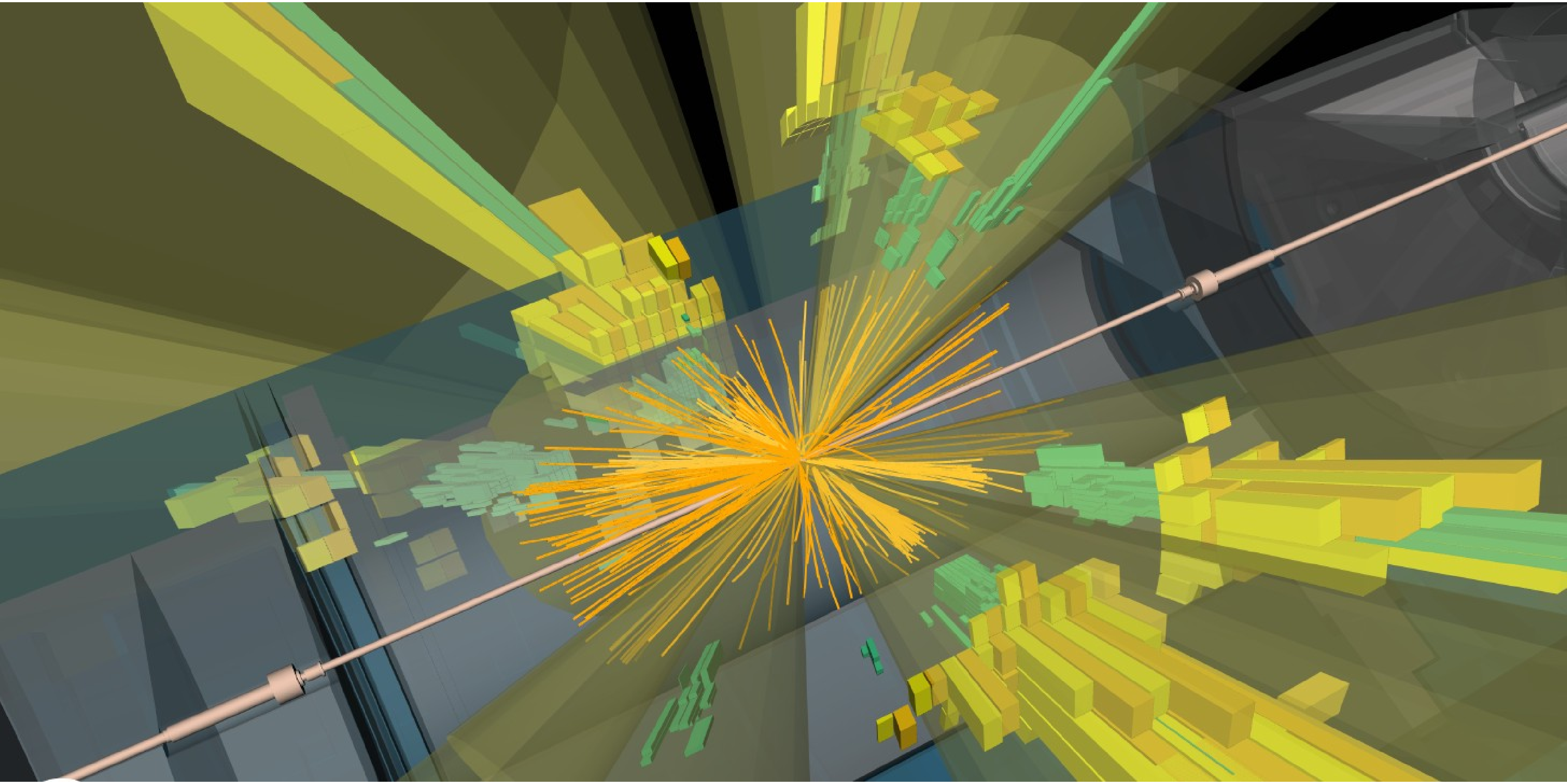


Statistical analysis methods for Particle Physics: Status and Prospects

(from an LHC perspective)

Nicolas Berger (LAPP Annecy)

Randomness in High-Energy Physics



Experimental data is produced by incredibly complex processes!

Randomness in High-Energy Physics

Experimental data is produced by incredibly complex processes

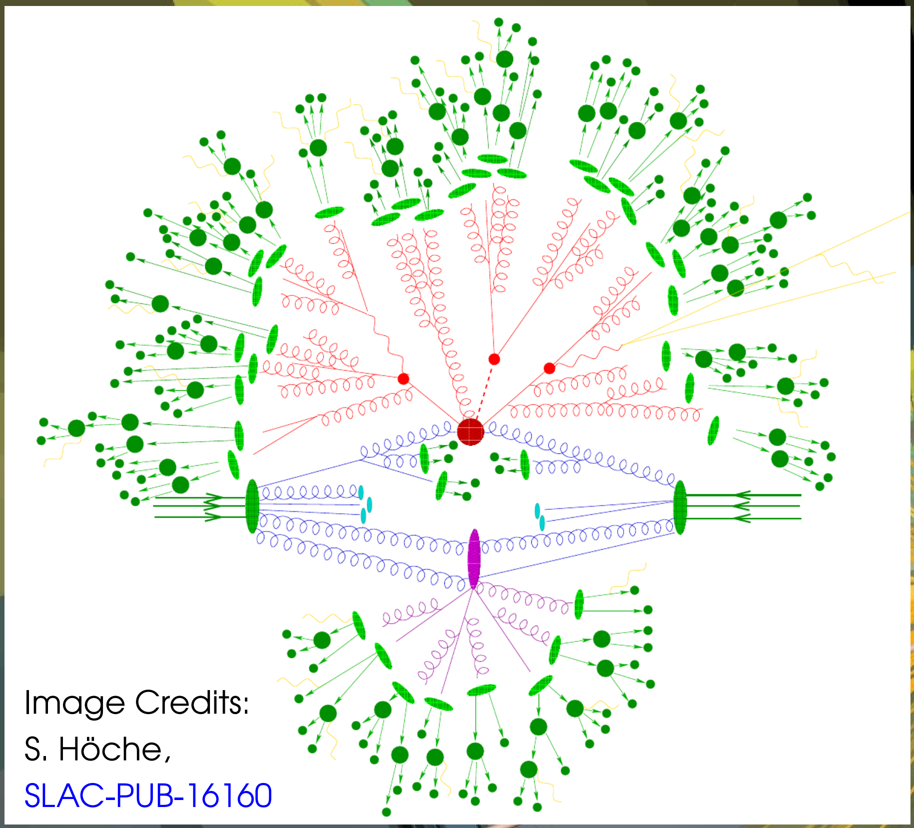
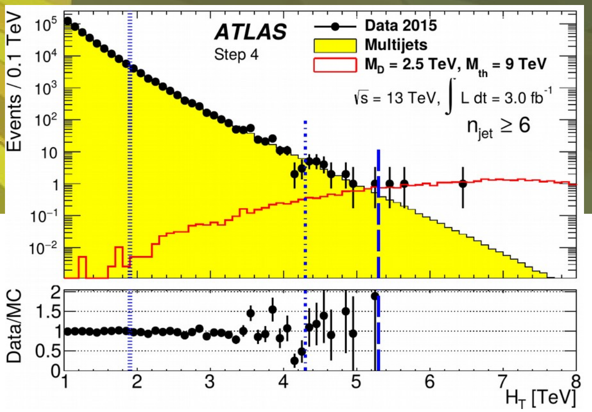


Image Credits:
S. Höche,
SLAC-PUB-16160

- Hard scattering
- PDFs, Parton shower, Pileup
- Decays
- Detector response
- Reconstruction

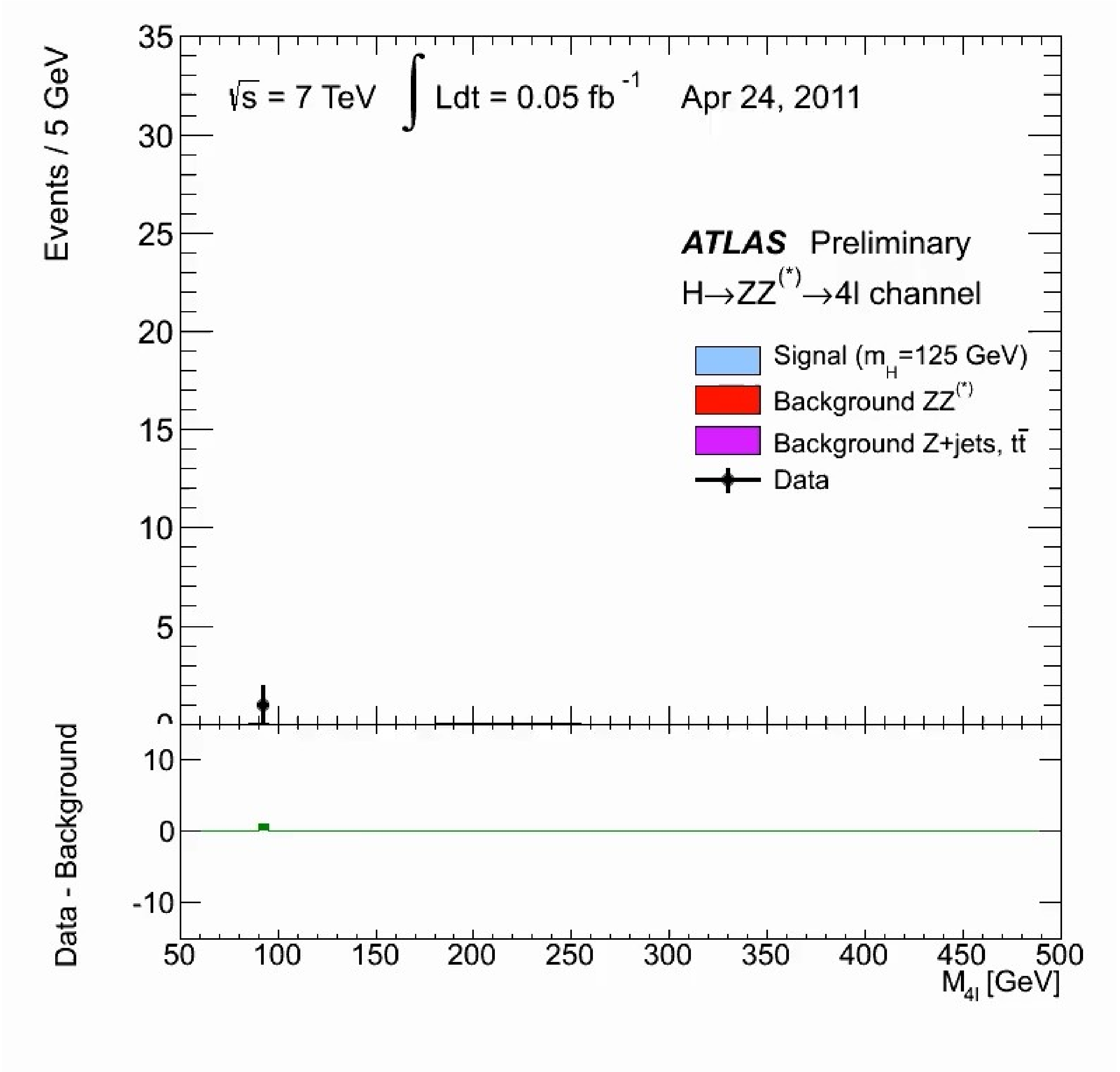


Randomness involved in all stages

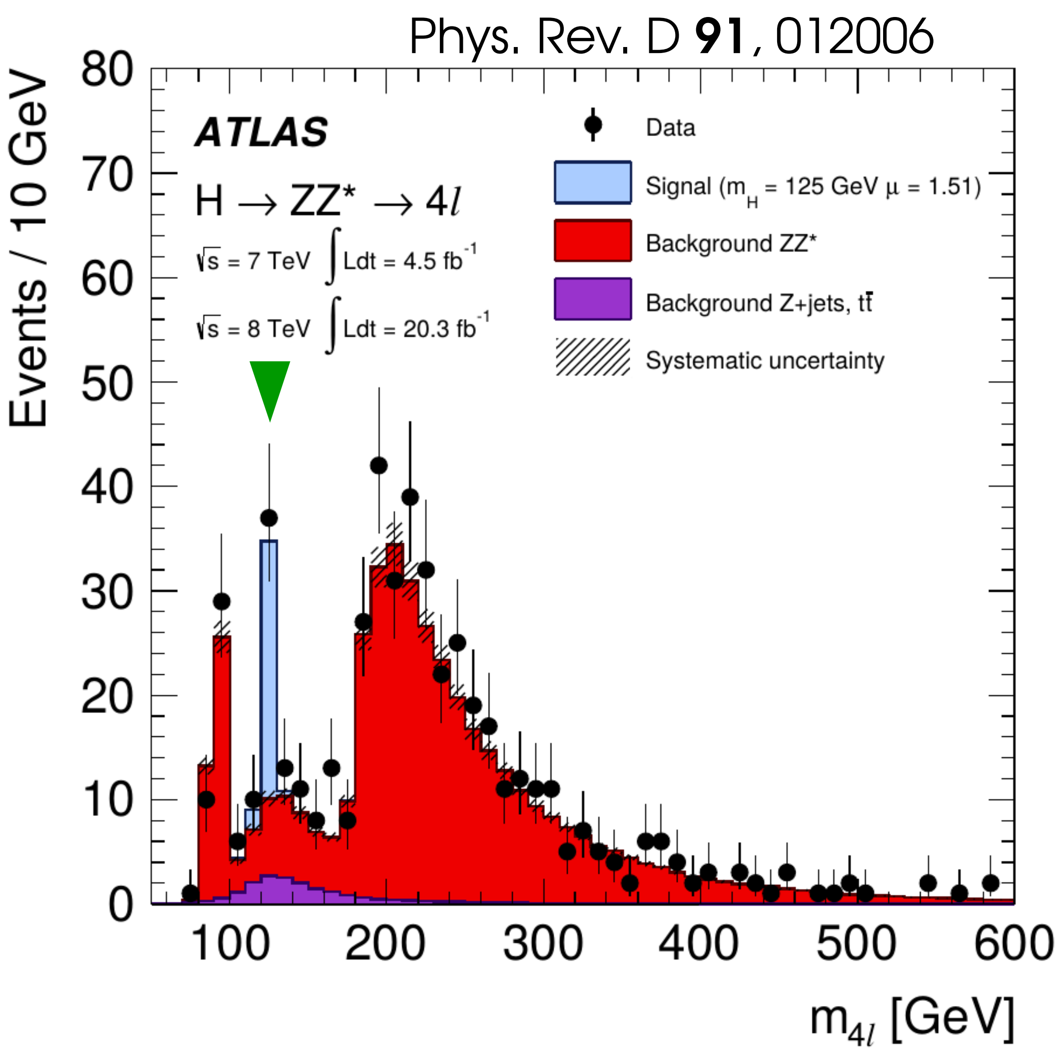
→ **Classical** randomness: detector reponse

→ **Quantum** effects in production, decay

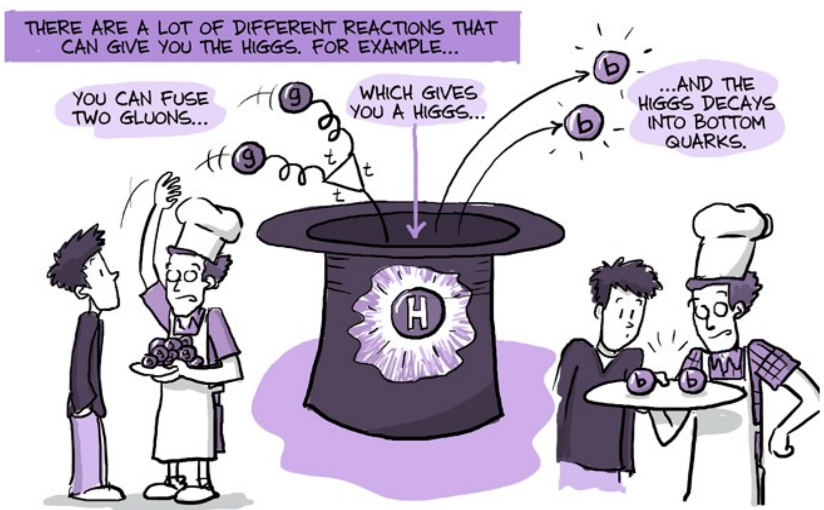
Quantum Randomness: H→ZZ*→4l



Quantum Randomness: H→ZZ*→4l



Rare process: Expect 1 signal event every **~6 days**



http://www.phdcomics.com/comics/archive.php?comicid=1489

Quantum randomness: "Will I get an event today ?" → only **probabilistic** answer

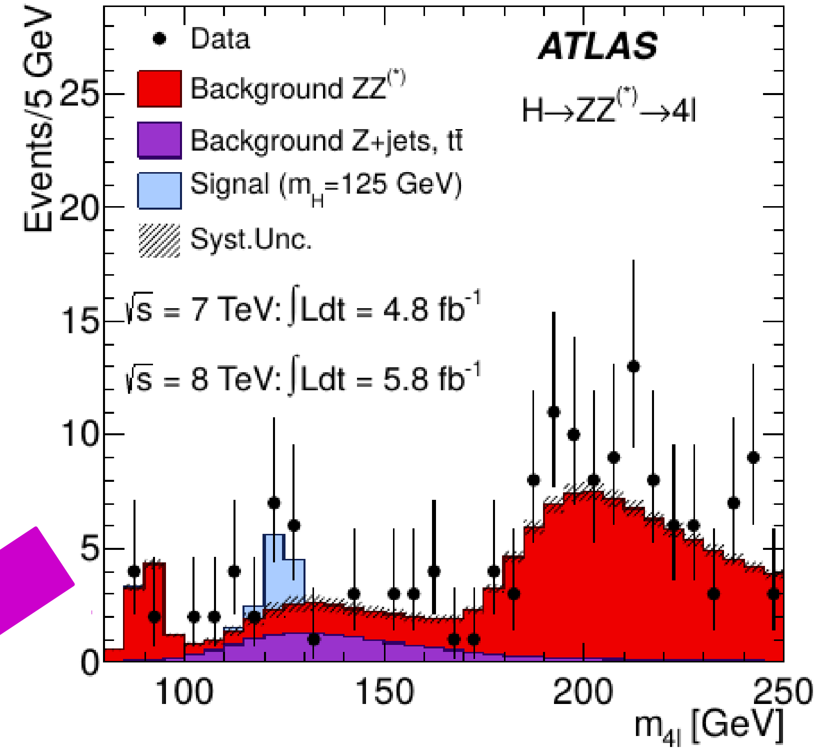
Discoveries

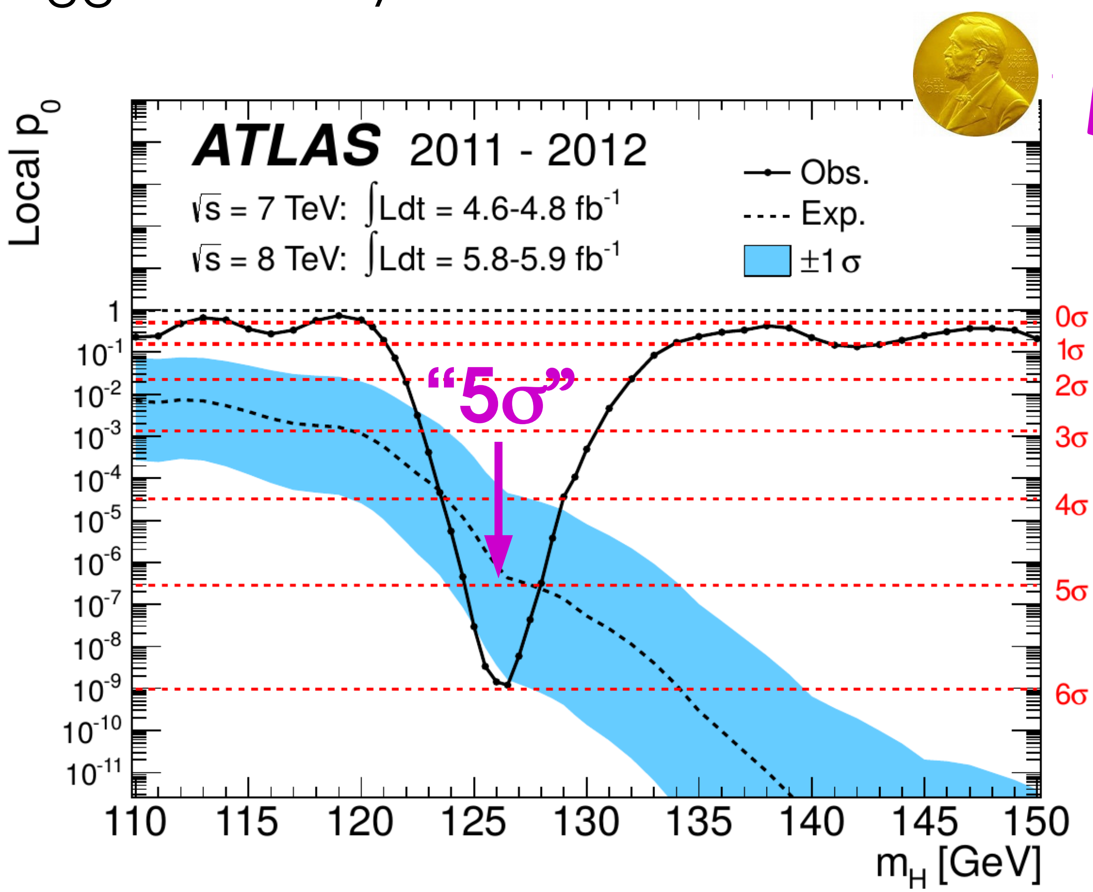
Randomness ⇒ fluctuations.

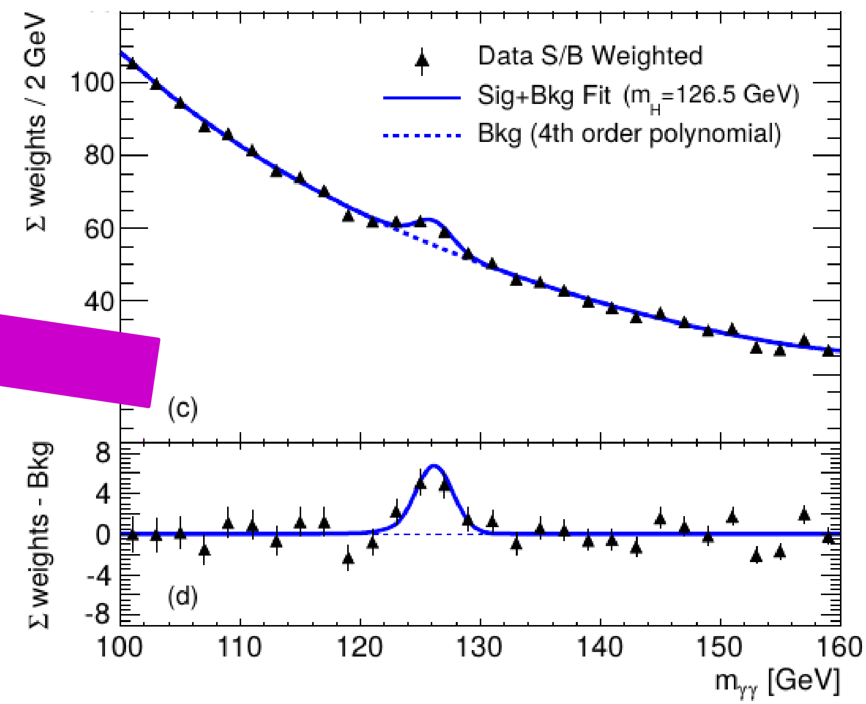
How to distinguish this from New Physics ?

→ Need to quantify confidence in an excess...

Higgs discovery : **"We have 5σ" !**







Phys. Lett. B 716 (2012) 1-29

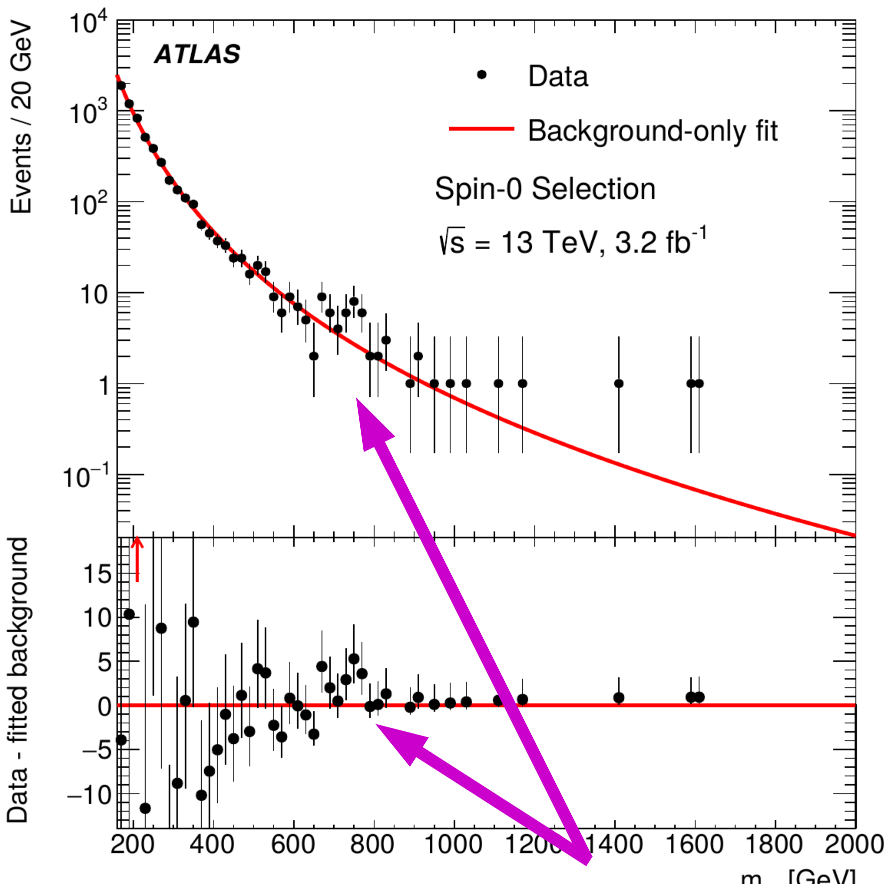
Discoveries ...or not

Randomness ⇒ fluctuations.

How to distinguish this from New Physics ?

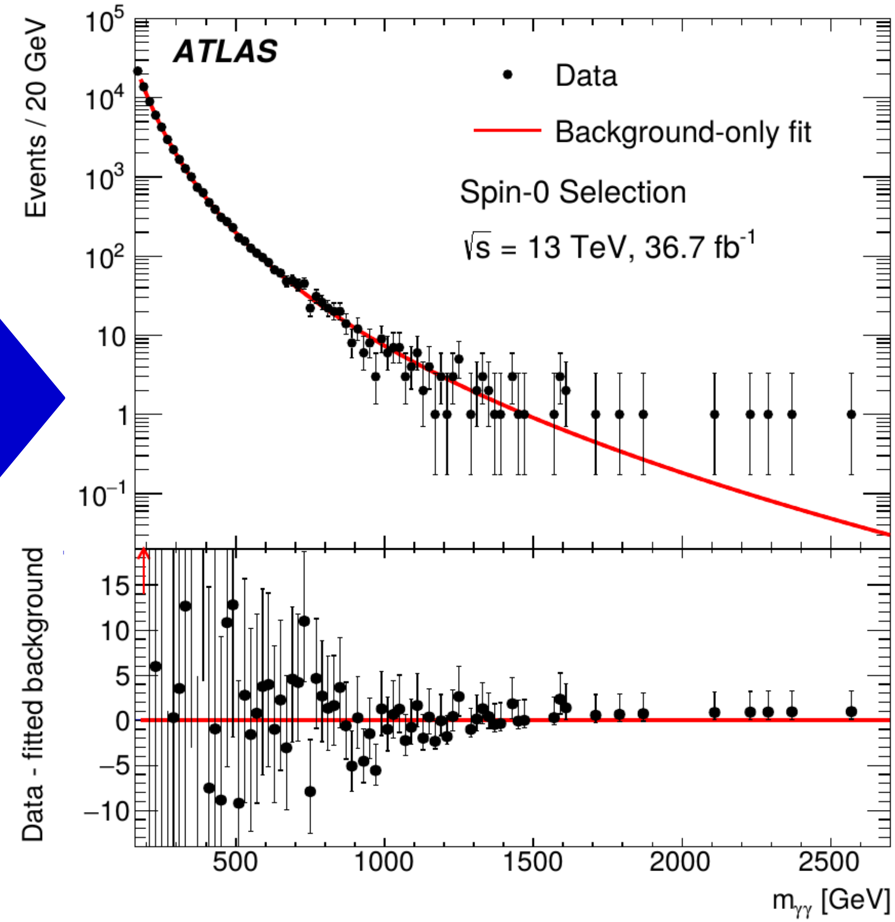
→ … and robust methods to control **spurious "discoveries"**…

JHEP 09 (2016) 1



1 Year Later

Phys. Lett. B 775 (2017) 105



New Physics ?? ~~3.9σ !?~~ ... 2.1σ

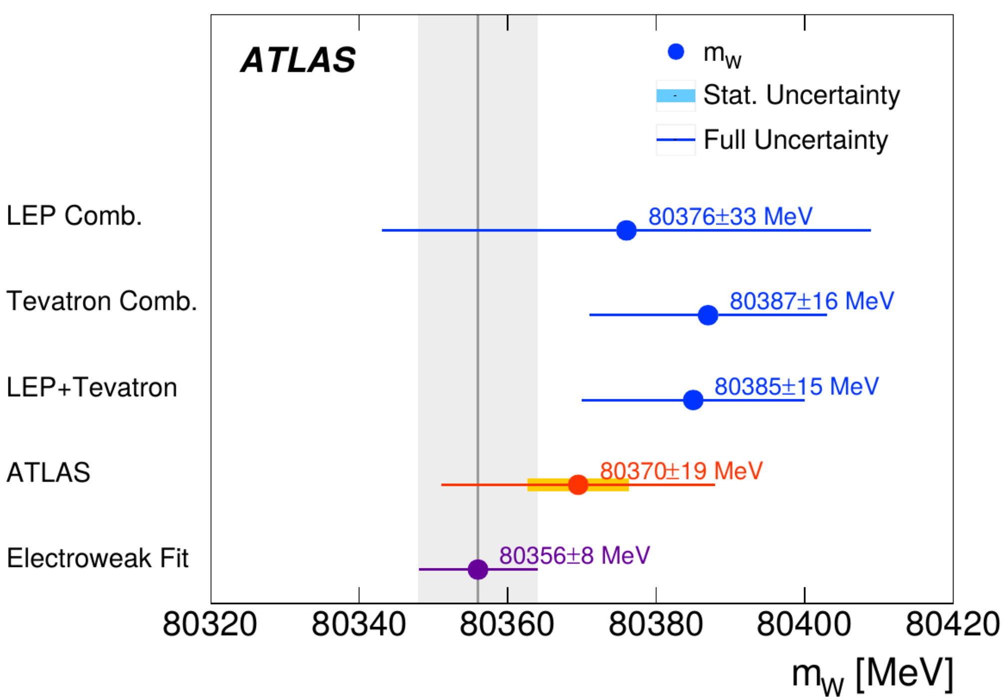
Parameter Measurements

Randomness ⇒ Measurement uncertainties

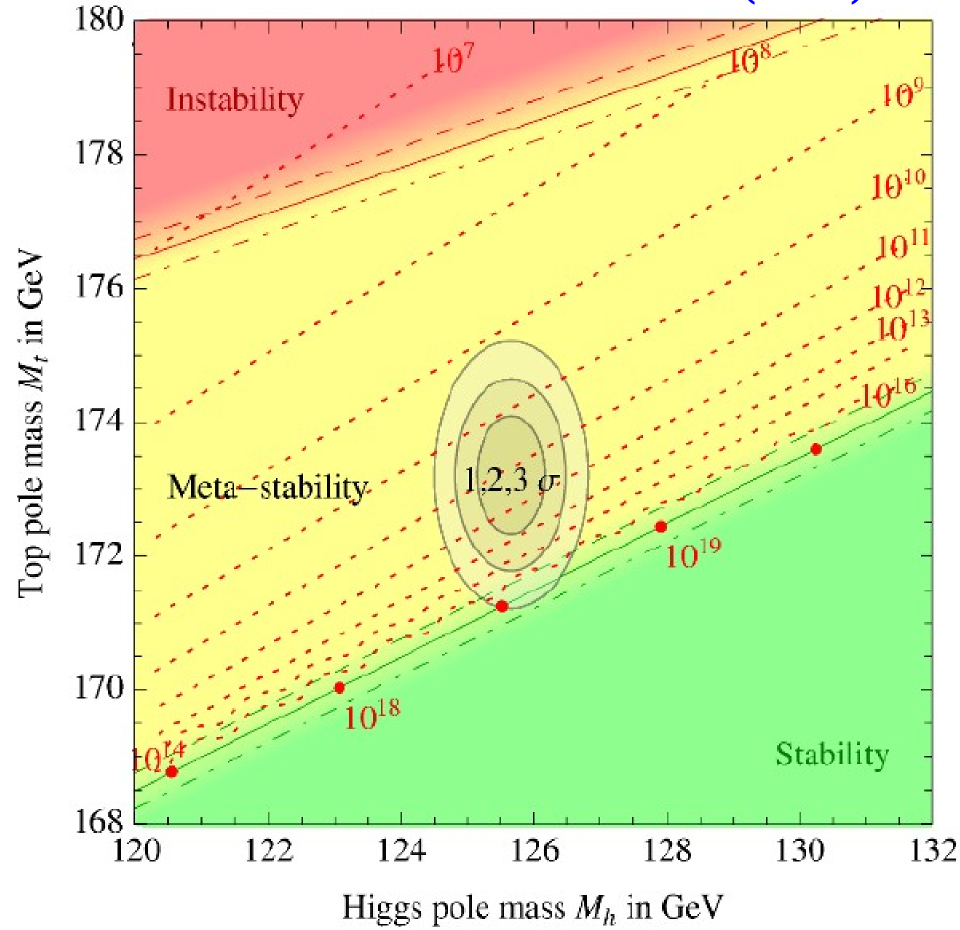
Key point for **precision measurements**,

⇒ Answers to important questions (especially if no new peaks found at high mass…)

arXiv:1701.07240



JHEP12(2013)089

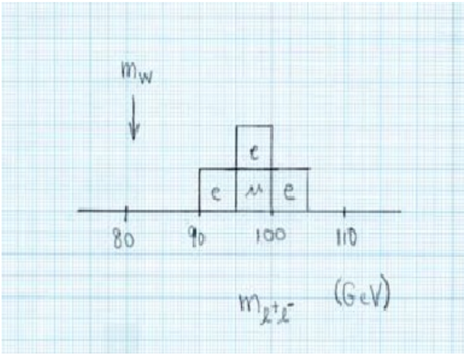


Consistency of the SM...

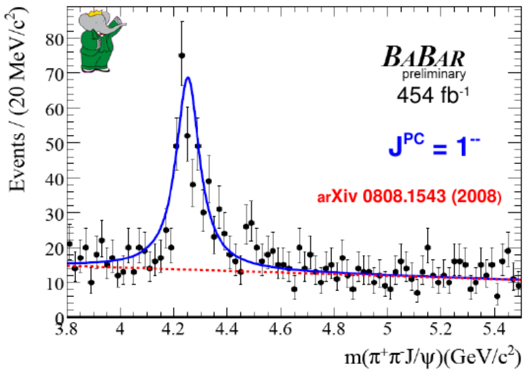
… or the fate of the universe

Statistical Methods

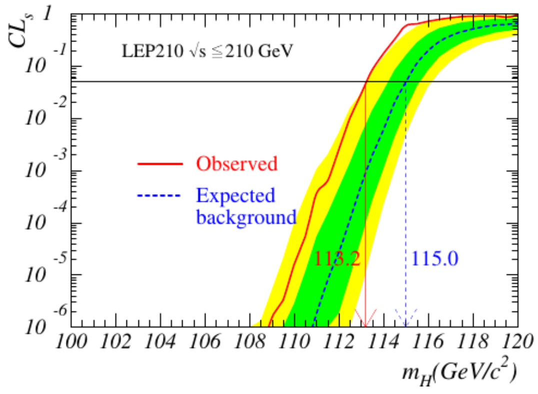
- **Many ways to perform statistical analyses**:
  → types of results
  → modeling assumptions
  → available CPU power



- Large experimental efforts ⇒ new developments: LEP, TeVatron, BaBar/Belle, LHC, etc.



- **Long term trend:**
  → more complex experiments
  → more focus on systematics
  ⇒ more detailed statistical modeling



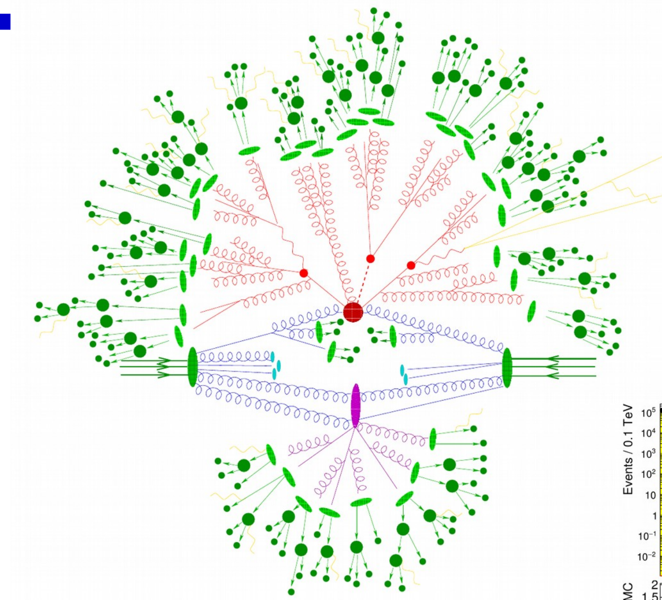
- **This talk: (biased) summary of current practices at LHC** :
  → Focus on frequentist interpretation, profiling of systematics
  → Many aspects relevant also for other methods
  e.g. Statistical modeling

Statistical Modeling

Statistical Model

Goal:
Describe the random process by which the data was obtained.

→ Build a **Statistical Model**

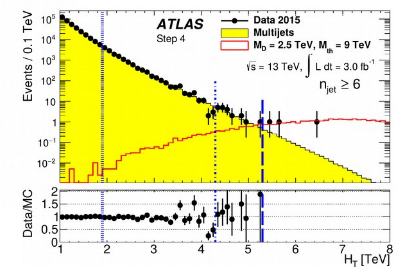


Hard scattering

Decays

Detector response

Reconstruction



Ingredients:

1. **Statistical description** of the random aspects
   ⇒ **Probability distributions**

2. **Assumptions** on the underlying statistical processes (physics, etc.)
   → Uncertainties on the assumptions themselves: **systematic uncertainties**

> "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 ?***

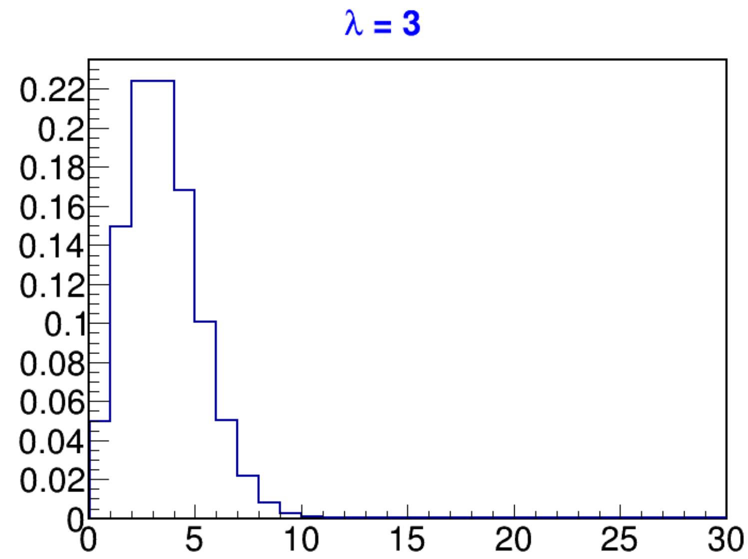
Statistical results can only be as accurate as the model itself !

Modeling Rare Processes: Poisson Counting

Counting experiment:

Observable: a **number of events n**
→ describe by a **Poisson distribution**

$$P(n;\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}$$



Typically both signal and background expected:

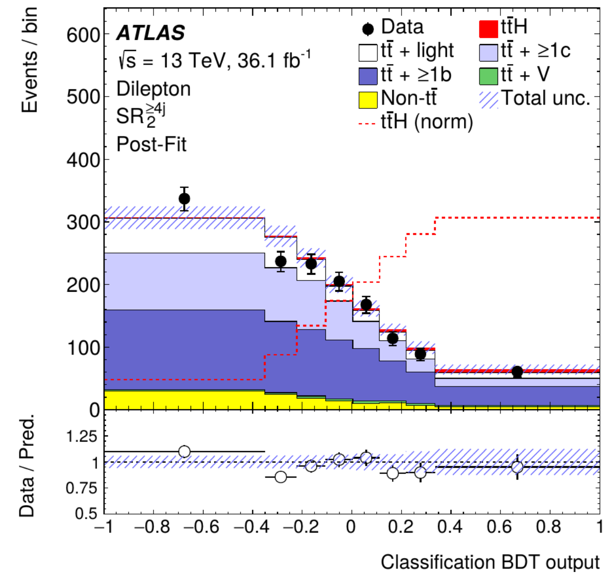
$$P(n;S,B) = e^{-(S+B)}\frac{(S+B)^n}{n!}$$

S : # of events from signal process
B : # of events from bkg. process(es)

→ Example: **assume B is known,** use the **measured n** to find out about the **parameter S**.

usually up to uncertainties → **systematics**

Model 2: Binned Shape Analysis



Count events in N separate regions

⇒ measure a histogram $n_1 \ldots n_N$.

Per-bin fractions (=shapes) of Signal and Background

$$P(\{n_i\}; S, B) = \prod_{i=1}^{N} e^{-(S f_{S,i} + B f_{B,i})} \frac{(S f_{S,i} + B f_{B,i})^{n_i}}{n_i!}$$

Poisson distribution in each bin

N=1: Back to the simple counting analysis

→ Can obtain fractions directly from MC

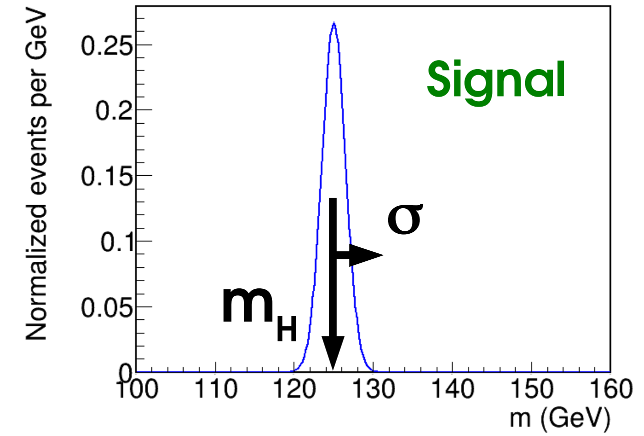
→ **MC stat fluctuations** can create artefacts, especially for S≪B.

Model 3: Unbinned Shape Analysis

Observable: event-by-event $m_1 \dots m_n$

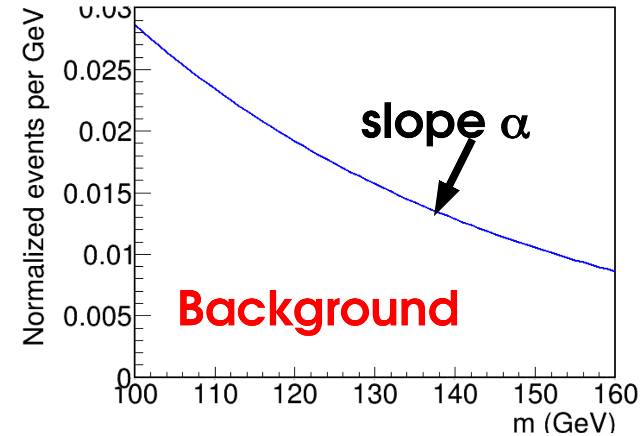
→ Describe shape of the **distribution of m**

→ Deduce the **probability to observe $m_1 \dots m_n$**



H→γγ-inspired example:

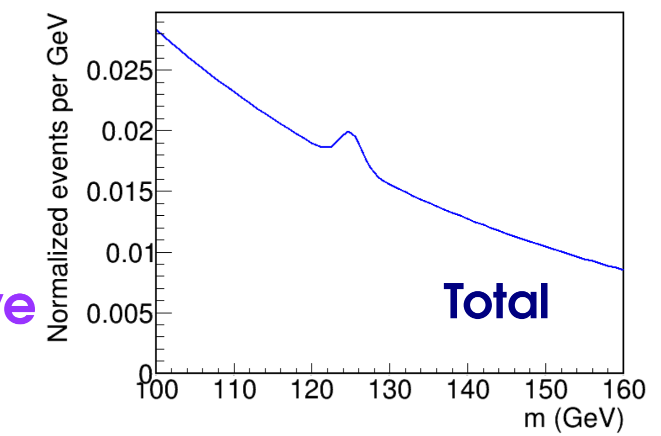
- Gaussian signal $P_{\text{signal}}(m) = G(m; m_H, \sigma)$
- Exponential bkg $P_{\text{bkg}}(m) = \alpha\, e^{-\alpha m}$



Expected yields : **S, B**

⇒ **Total PDF for a single event**:

$$P_{\text{total}}(m) = \frac{S}{S+B} G(m; m_H, \sigma) + \frac{B}{S+B} \alpha\, e^{-\alpha m}$$



⇒ **Total PDF for a dataset**

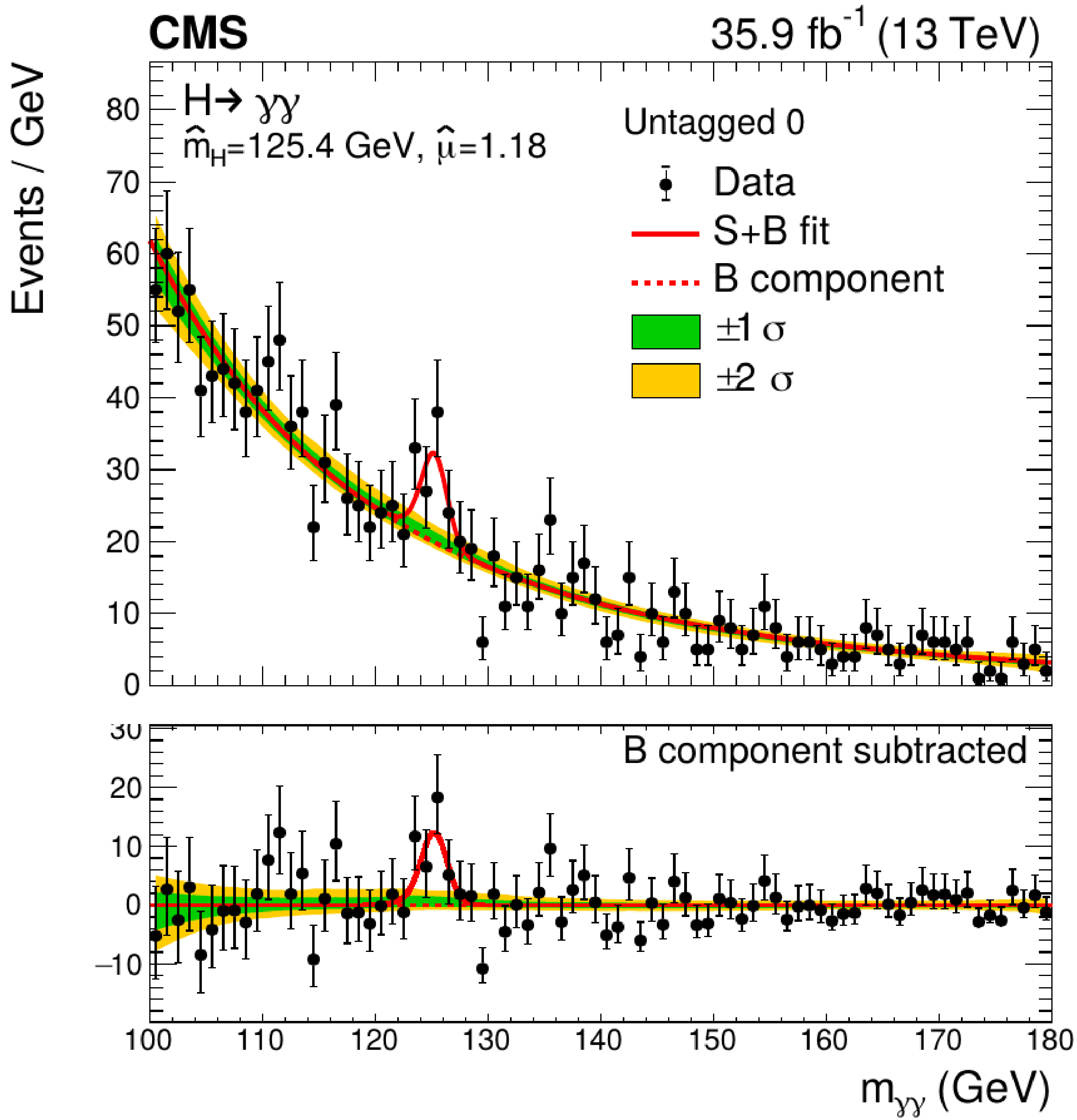
Probability to observe the value m_i

Probability to observe n events

$$P(\{m_i\}_{i=1 \dots n}) = e^{-(S+B)} \frac{(S+B)^n}{n!} \prod_{i=1}^{n} \frac{S}{S+B} G(m_i; m_H, \sigma) + \frac{B}{S+B} \alpha\, e^{-\alpha m_i}$$

H→γγ

JHEP 11 (2018) 185



Categories

Multiple analysis regions often used

→ Useful to model separately if

- Regions with better sensitivity (avoids dilution)
- *Control regions* for backgrounds
- Multiple signal measurements

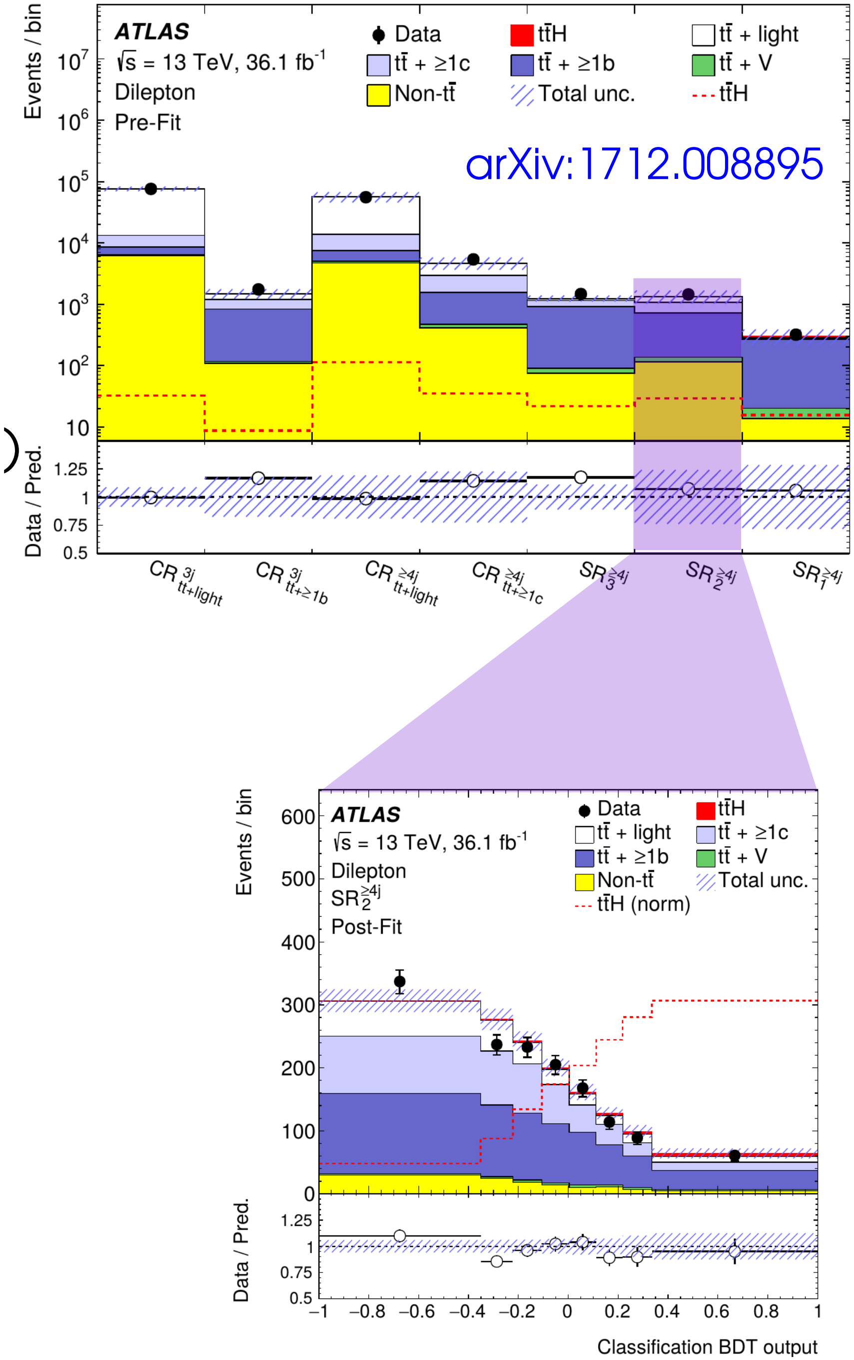
⇒ ***Analysis categories*** :

PDF for category k

$$P\left(S;\{n_i^{(k)}\}_{i=1\ldots n_{\mathrm{evts}}^{(k)}}^{k=1\ldots n_{\mathrm{cats}}}\right)=\prod_{k=1}^{n_{\mathrm{cats}}}P_k\left(S;\{n_i^{(k)}\}_{i=1\ldots n_{\mathrm{evts}}^{(k)}}\right)$$

No overlaps between categories

⇒ No stat. correlations ⇒ product of PDFs.

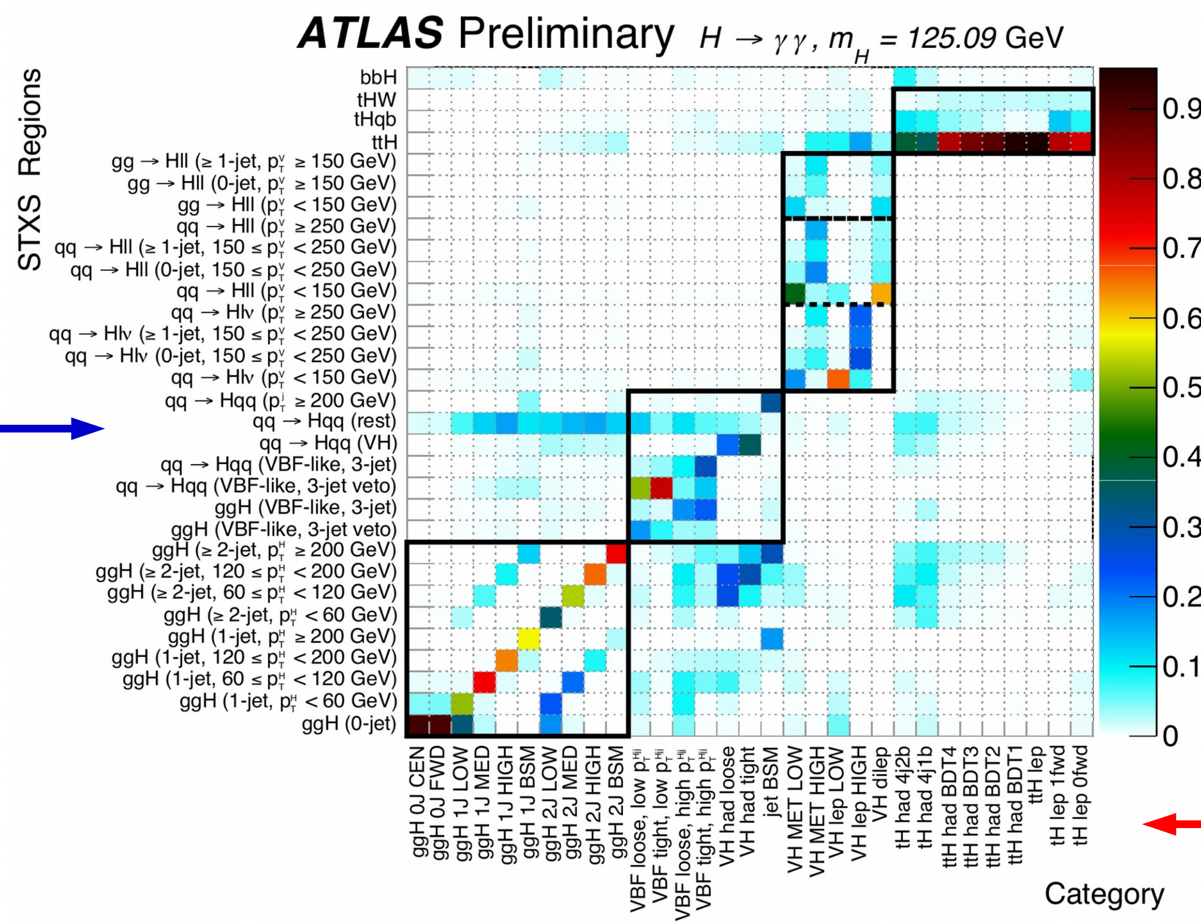


Similar to a-posteriori combination but allows proper handling of correlated parameters (CR scale factors, systematics, etc.)

Categories for H→γγ Property Measurements

Categories also useful to provide measurements of separate kinematic regions
→ e.g. differential cross-section measurements

Targeted truth regions →



← **Analysis Selections**

H→ γγ Properties Measurement (ATLAS-CONF-2017-045)

Many **categories**, combined analysis for optimal use of all information

Systematics

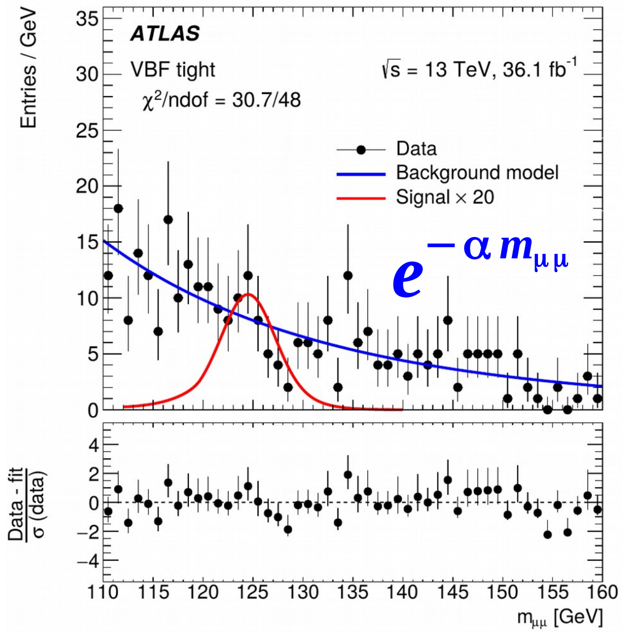
Phys. Rev. Lett. 119 (2017) 051802

Statistical model 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 process

e.g. integrated luminosity L of a data sample for a cross-section measurement

"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 ?***

⇒ Parameterize using additional free parameters (NPs)

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

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

⇒ Need to inject additional information

Frequentist Constraints

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

→ Build the combined PDF of the **main**+**auxiliary** measurements

$$P(\sigma, \theta_{lumi}; \mathbf{data}) = P_{\mathrm{main}}(\sigma, \theta_{lumi}; \mathbf{main\ data})\ P_{\mathrm{aux}}(\theta_{lumi}; \mathbf{lumi\ data})$$

Independent measurements: ⇒ just a product

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

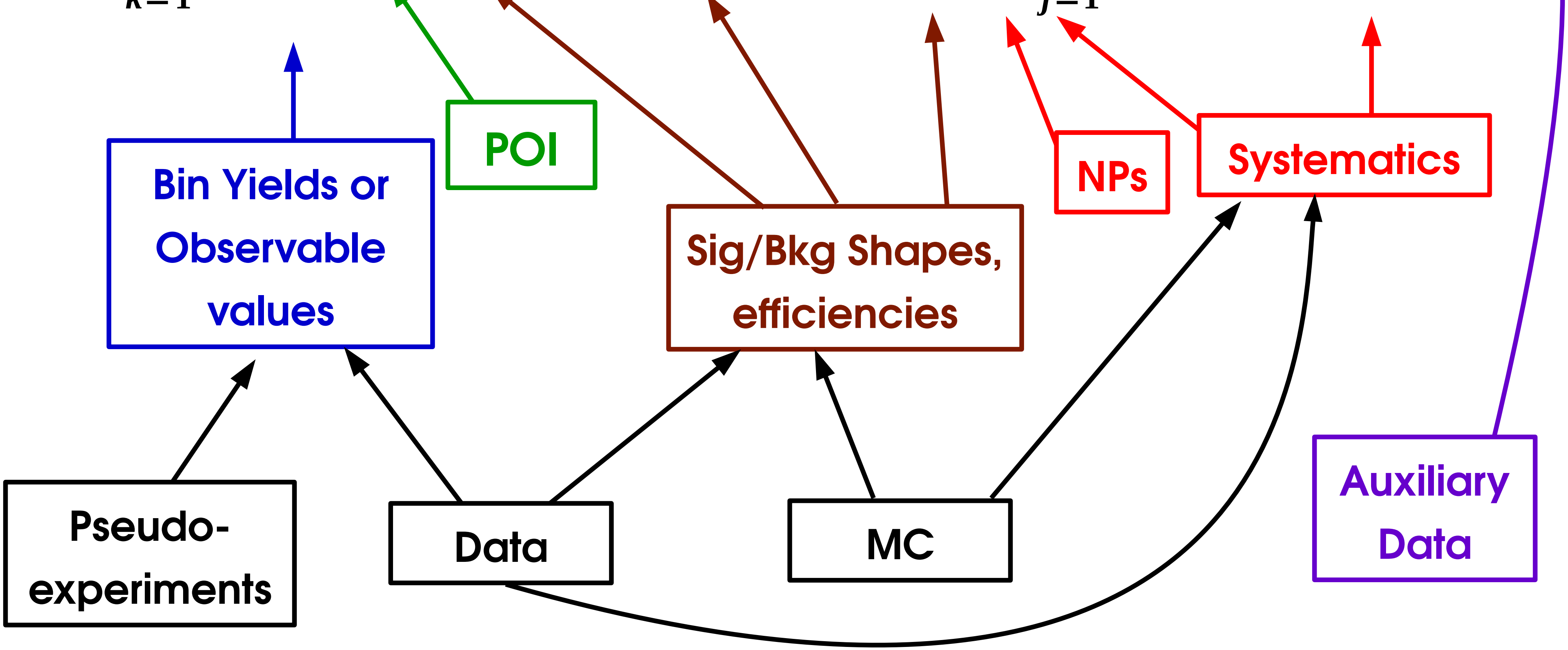
In the combined PDF, **systematic NPs are constrained**
Systematics → just additional NPs

→ 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)

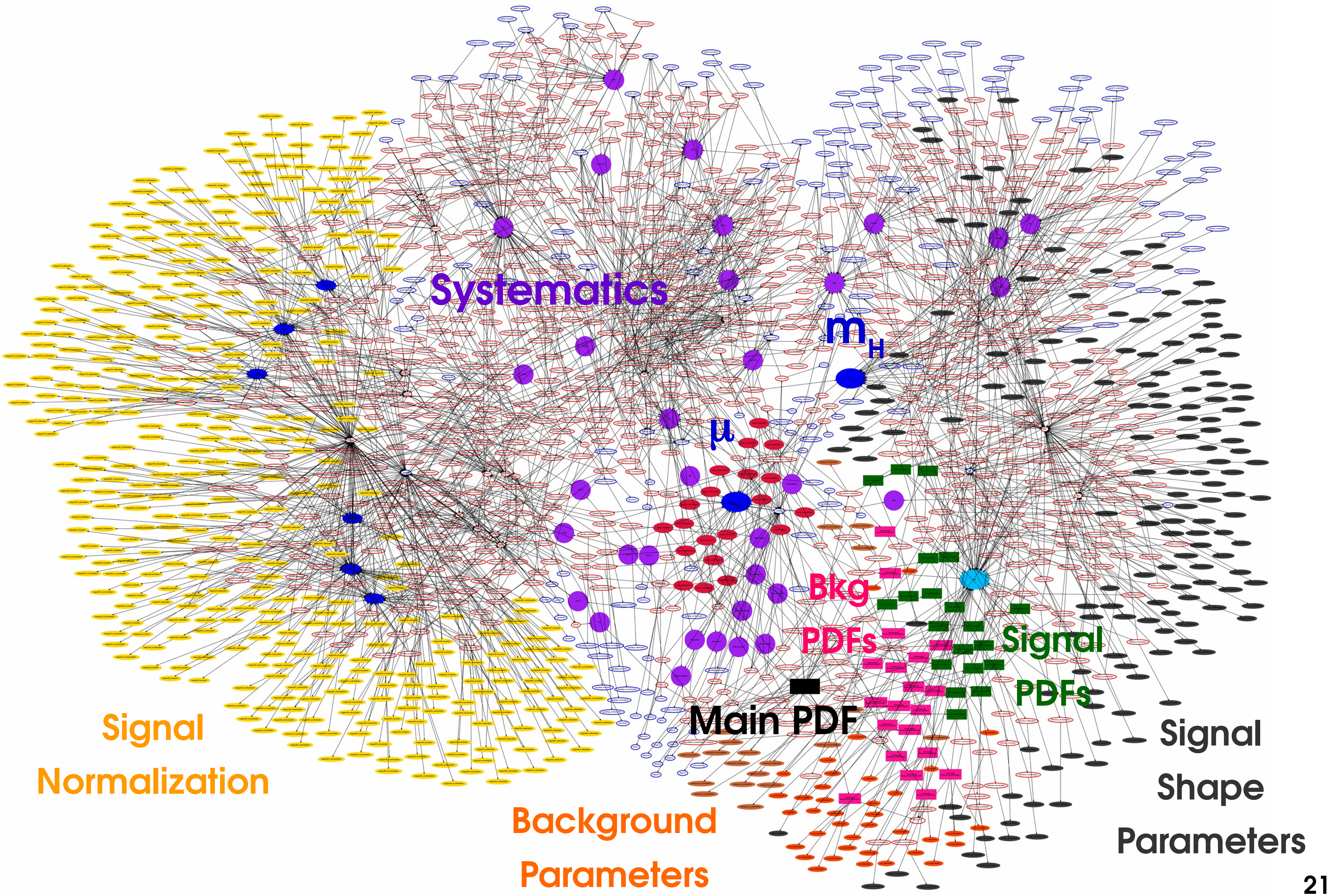
$$L\left(\mu, \{\theta_j\}_{j=1\ldots n_{NP}}; \{n_i^{(k)}\}_{i=1\ldots n_{data}^{(k)}}^{k=1\ldots n_{cat}}, \{\theta_j^{obs}\}_{j=1..n_{NP}}\right) =$$

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



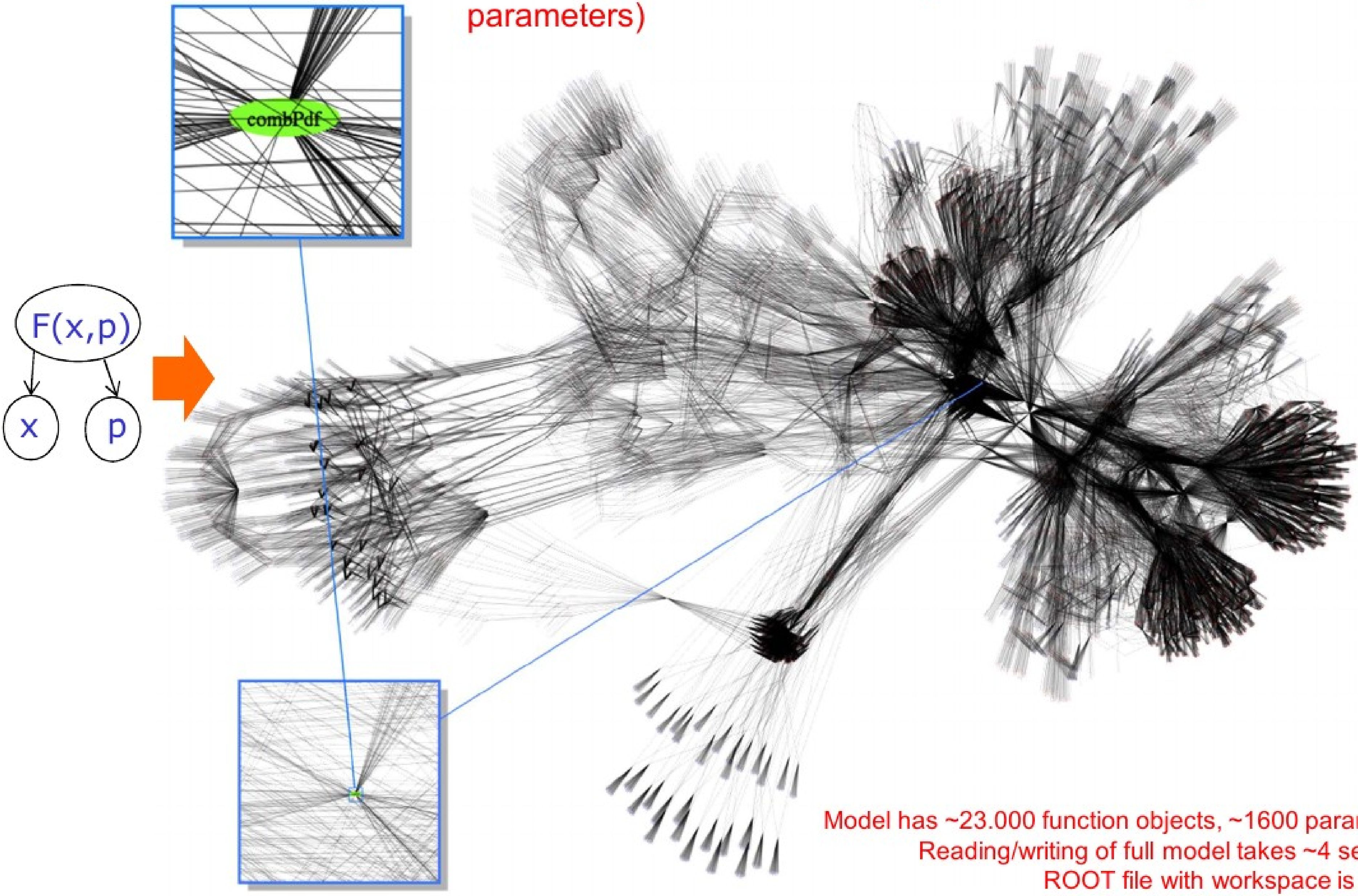
× number of categories!

Model Example: H→γγ Discovery Analysis



ATLAS Higgs Combination Model

Atlas Higgs combination model (23.000 functions, 1600 parameters)



Model has ~23.000 function objects, ~1600 parameters
Reading/writing of full model takes ~4 seconds
ROOT file with workspace is ~6 Mb

Computing Results

Using the PDF

Model describes the distribution of the observable: **P(data; parameters)**

⇒ Possible outcomes of the experiment, for given parameter values

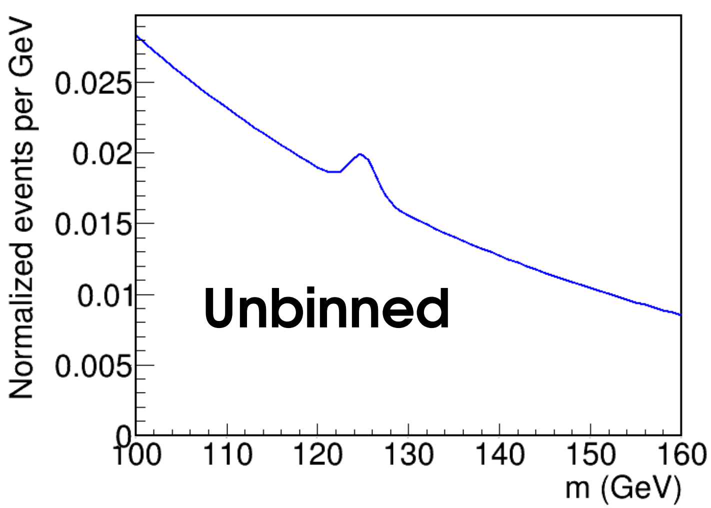
Can draw random events according to PDF : **generate (*pseudo-)data***

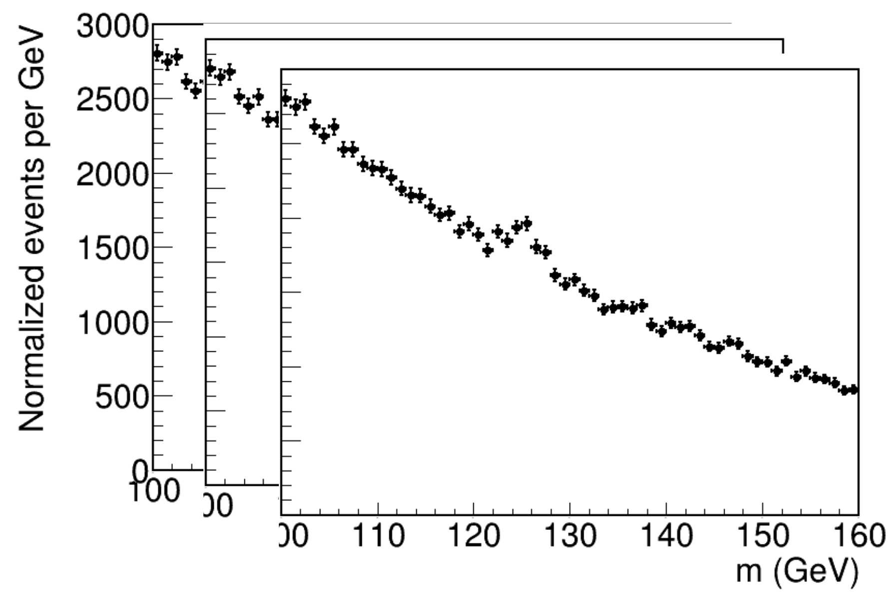
$$P(\lambda=5)$$

2, 5, 3, 7, 4, 9, ...

Each entry = separate "experiment"

Generate





Not a trivial task (huge challenge for HL-LHC!)

but not the main goal here

Likelihood

Model describes the distribution of the observable: **P(n; λ), P(data; parameters)**

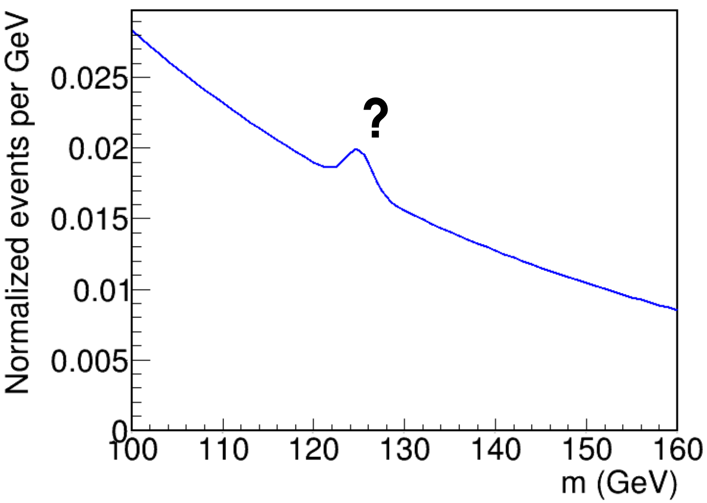
⇒ Possible outcomes of the experiment, for given parameter values

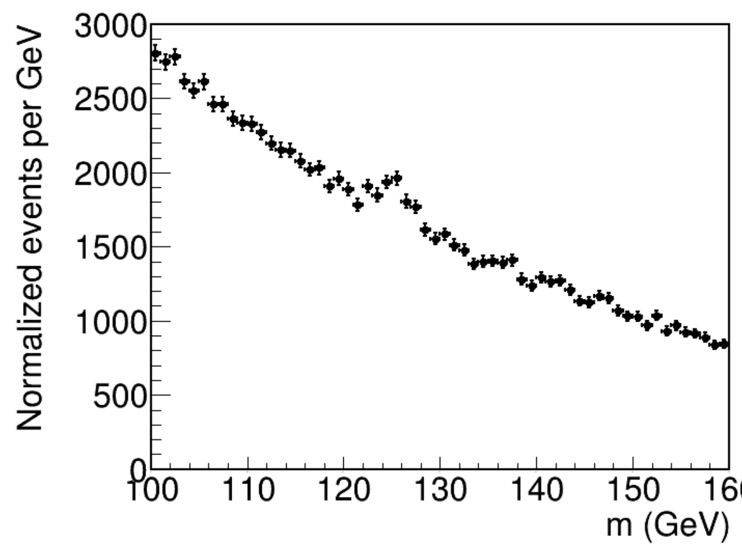
We want the other direction: **use data to get information on parameters**

$$P(\lambda = ?)$$

2

Estimate



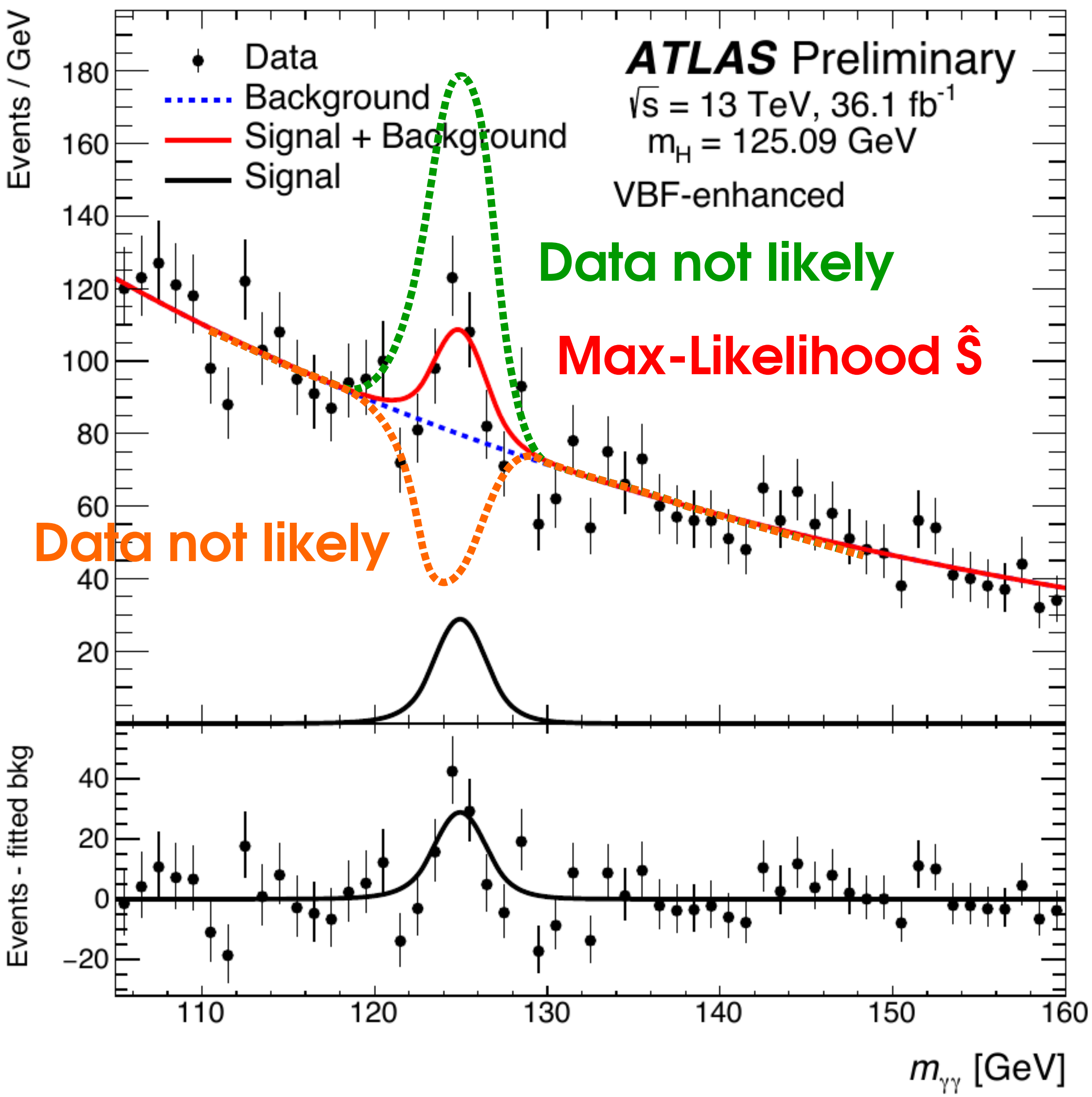


Likelihood: L(parameters) = P(data;parameters)

→ **same as PDF,** but evaluated on data and function of the parameters

Estimating a Parameter: Maximum Likelihood

$$L(S, B; m_i) = e^{-(S+B)} \prod_{i=1}^{n_{\text{evts}}} S\, P_{\text{sig}}(m_i, m_H) + B\, P_{\text{bkg}}(m_i)$$



Maximum Likelihood: *value **Ŝ** of **S** for which the **observed data** is **most likely** ?*

In practice:

Just the **usual best-fit value** from MINUIT, RooFit, etc.

Good properties for large n_{evts}:

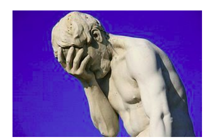
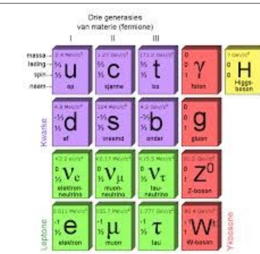
- Converges to true value ("consistent")
- Smallest possible RMS ("efficient")
- Gaussian-distributed

ATLAS-CONF-2017-045

Going further: Hypothesis Testing

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

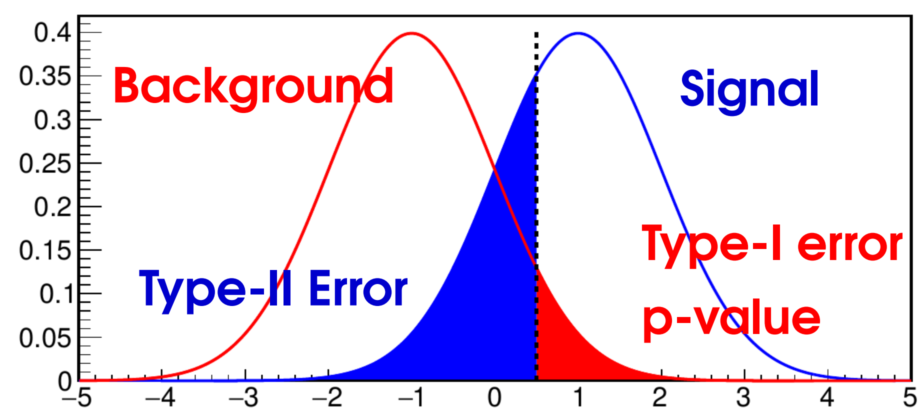
→ Goal : determine if H_0 is true or false using a test based on the data

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**

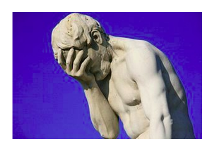
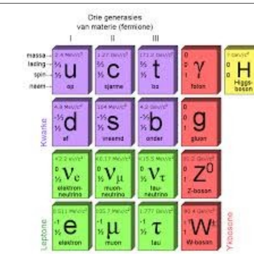
→ **Goal**: test that minimizes Type-II errors **for given level of Type-I error**.



Going further: Hypothesis Testing

Hypothesis: assumption on model parameters, say value of S (e.g. $\mathbf{H_0 : S=0}$)

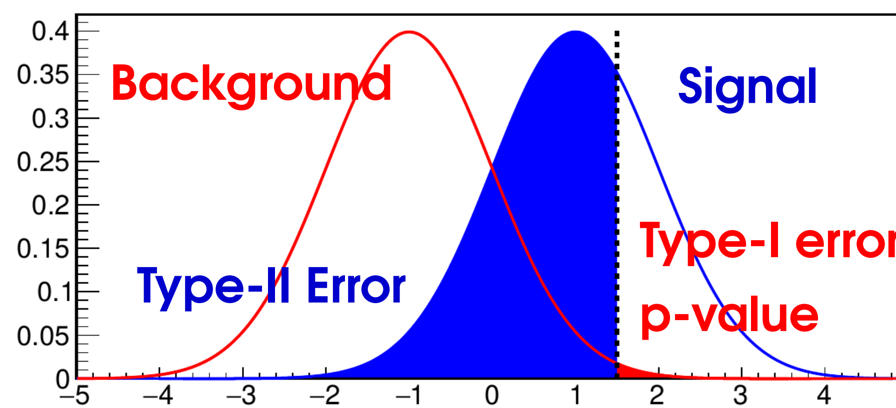
→ Goal : determine if H_0 is true or false using a test based on the data

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, say $S=S_0$ *and* $S=S_1$*, the optimal discriminator is the* ***Likelihood ratio*** *(LR)*

$$\frac{L(S=S_1; data)}{L(S=S_0; data)}$$

As for MLE, choose the hypothesis that is most likely **given the data**.
→ **Minimizes Type-II uncertainties** for given level of Type-I uncertainties

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

What about nuisance parameters ? (systematics, etc.)

Hypothesis Testing with Likelihoods

Profile Likelihood Ratio

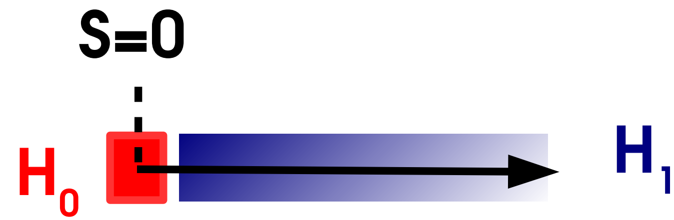
When comparing two hypotheses, say $S=S_0$ *and* $S=S_1$*, define the* ***Profile Likelihood ratio*** *(PLR) :*

$$\frac{L\left(S=S_1,\ \hat{\hat{\theta}}(S_1);\ data\right)}{L\left(S=S_0,\ \hat{\hat{\theta}}(S_0);\ data\right)}$$

Again, use the value of the NP θ that is **most likely given the data** : **Profiling**
Not guaranteed to be optimal, but works extremely well in practice

→ **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...

Discovery



Discovery :

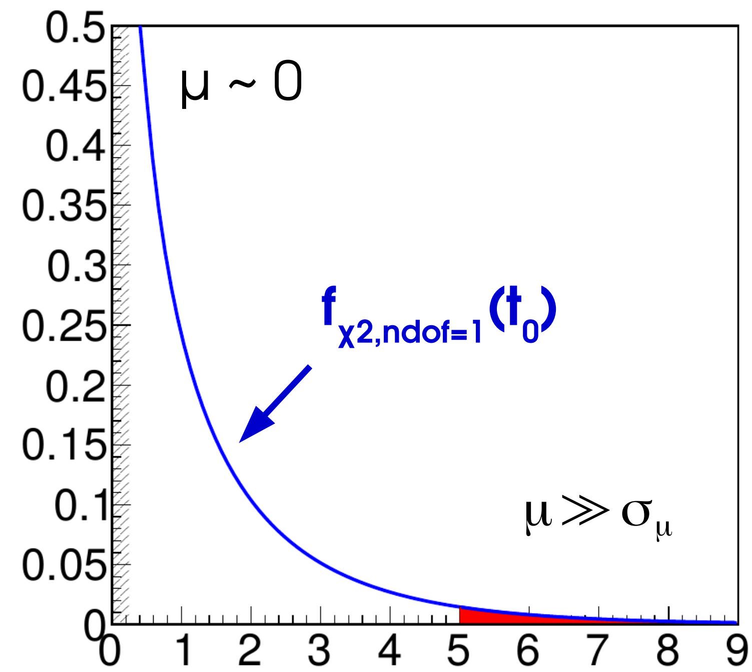
- **H_0 : background only (S = 0)** against
- **H_1**: presence of a signal (**S ≠ 0**)

→ For H_1, any S≠0 is possible, which to use ?

The one preferred by the data, $\hat{S}$.

⇒ Use

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



Why ?

→ Large values of t_0 ⇔ large observed S

→ Gaussian limit (n_{obs} > 5): t_0 follows a **χ^2 with n_{dof}=1**, regardless of NPs!

→ In particular,

$$Z = \sqrt{t_0}$$

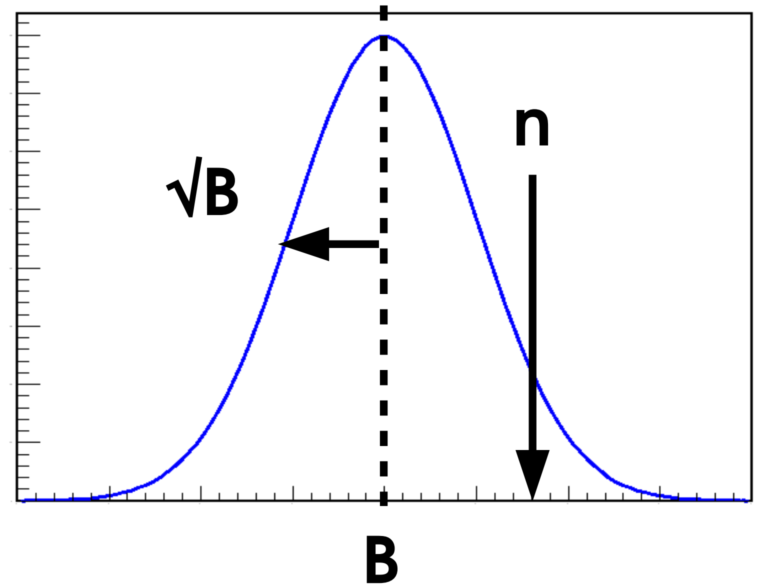
Cowan, Cranmer, Gross & Vitells, Eur.Phys.J.C71:1554,2011

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 : **$\hat{S} = n - B$**

Test statistic: assume Ŝ > 0,

$$t_0 = -2\log\frac{L(S=0)}{L(\hat{S})} = \lambda(S=0) - \lambda(\hat{S}) = \left(\frac{n-B}{\sqrt{B}}\right)^2 = \left(\frac{\hat{S}}{\sqrt{B}}\right)^2$$

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: $\hat{S} = n - B$, same as Gaussian

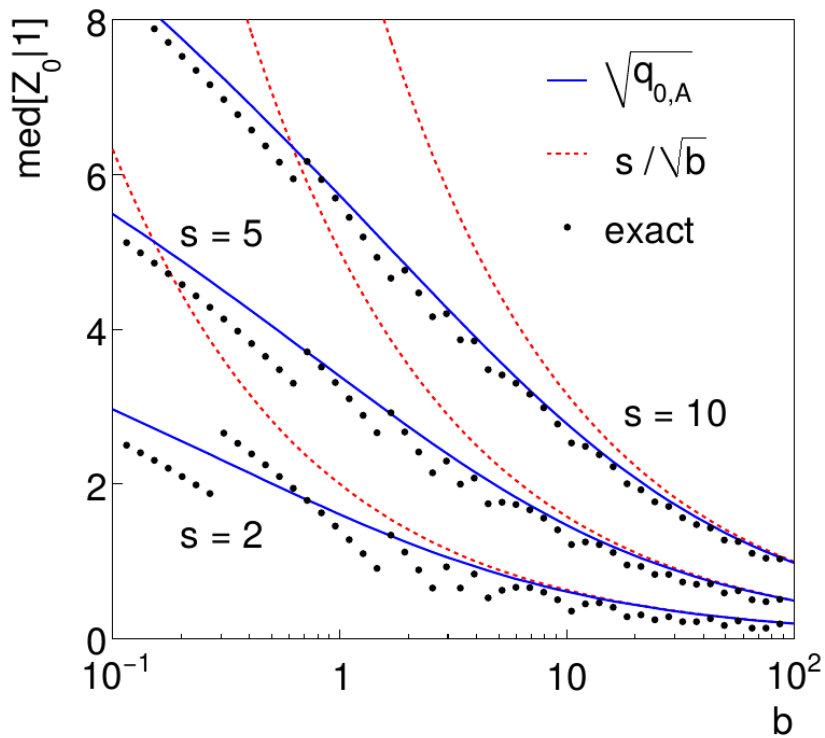
Test statistic (for $\hat{S} > 0$): $$q_0 = \lambda(S=0) - \lambda(\hat{S}) = -2\hat{S} - 2(\hat{S}+B)\log\frac{B}{\hat{S}+B}$$

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

Eur.Phys.J.C71:1554,2011

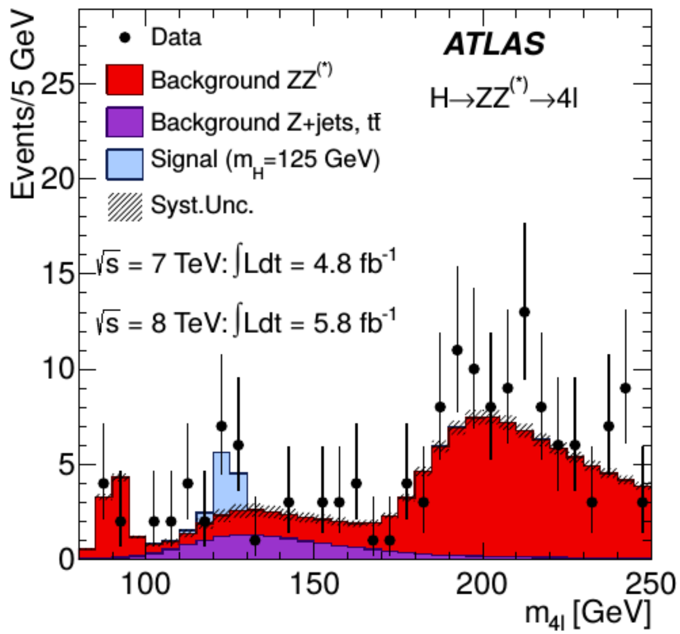


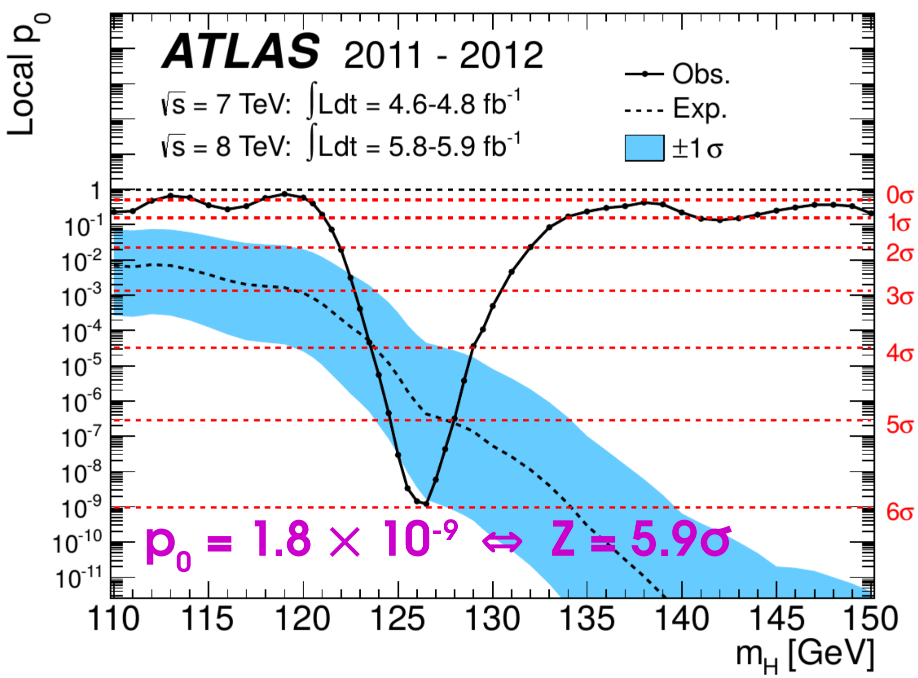
Asymptotic formulas justified by Gaussian regime, but remain valid even for small values of S+B (5!)

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

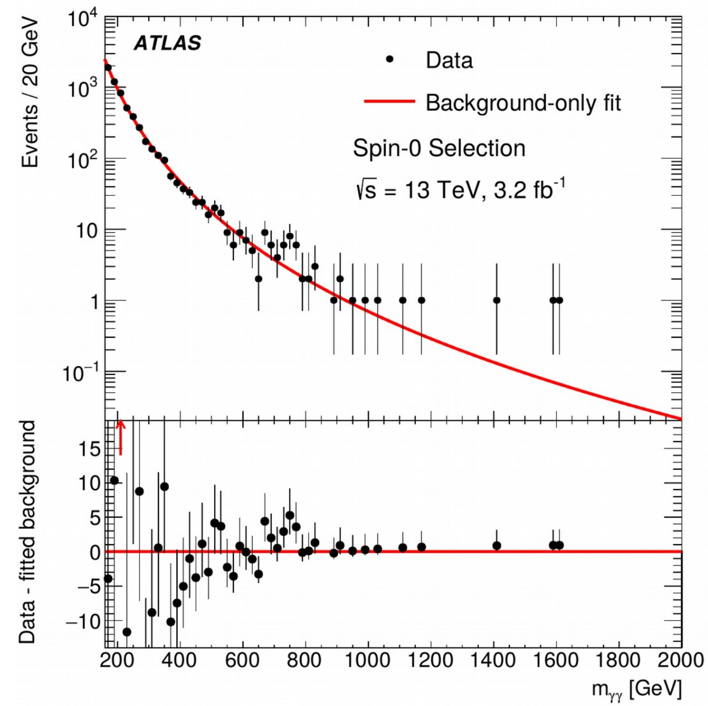
Some Examples

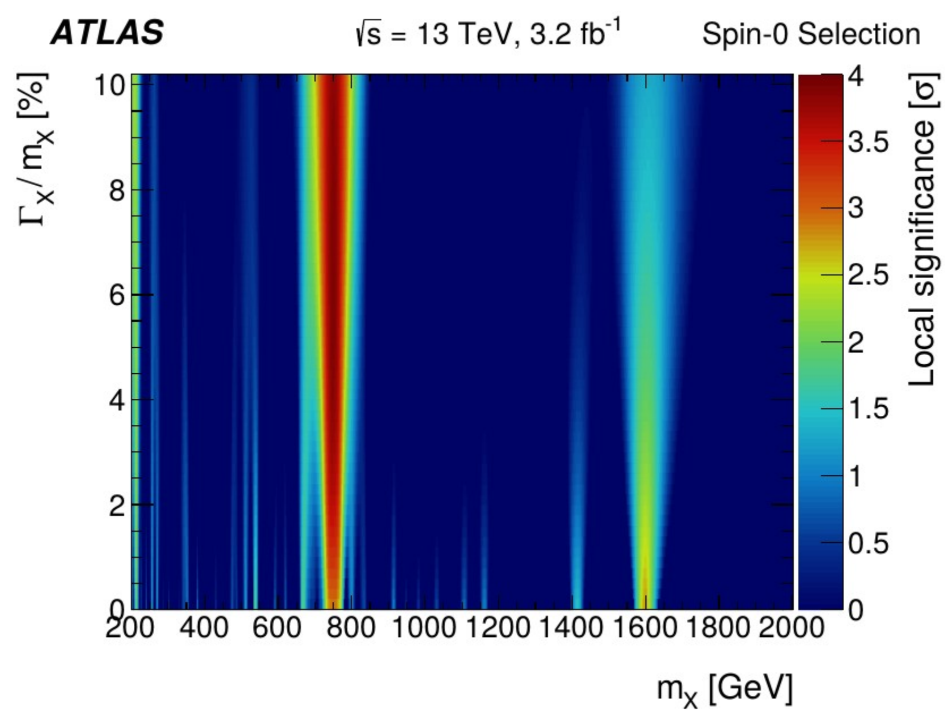
Higgs Discovery: Phys. Lett. B 716 (2012) 1-29



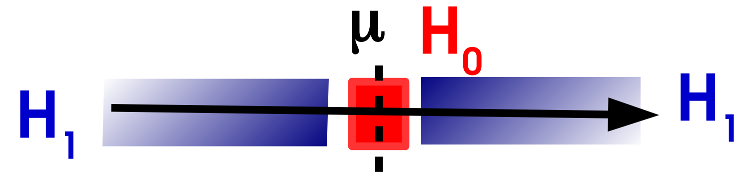


High-mass X→γγ Search: JHEP 09 (2016) 1





Likelihood Intervals



Confidence intervals from L:

- Test $H(\mu_0)$ against alternative using

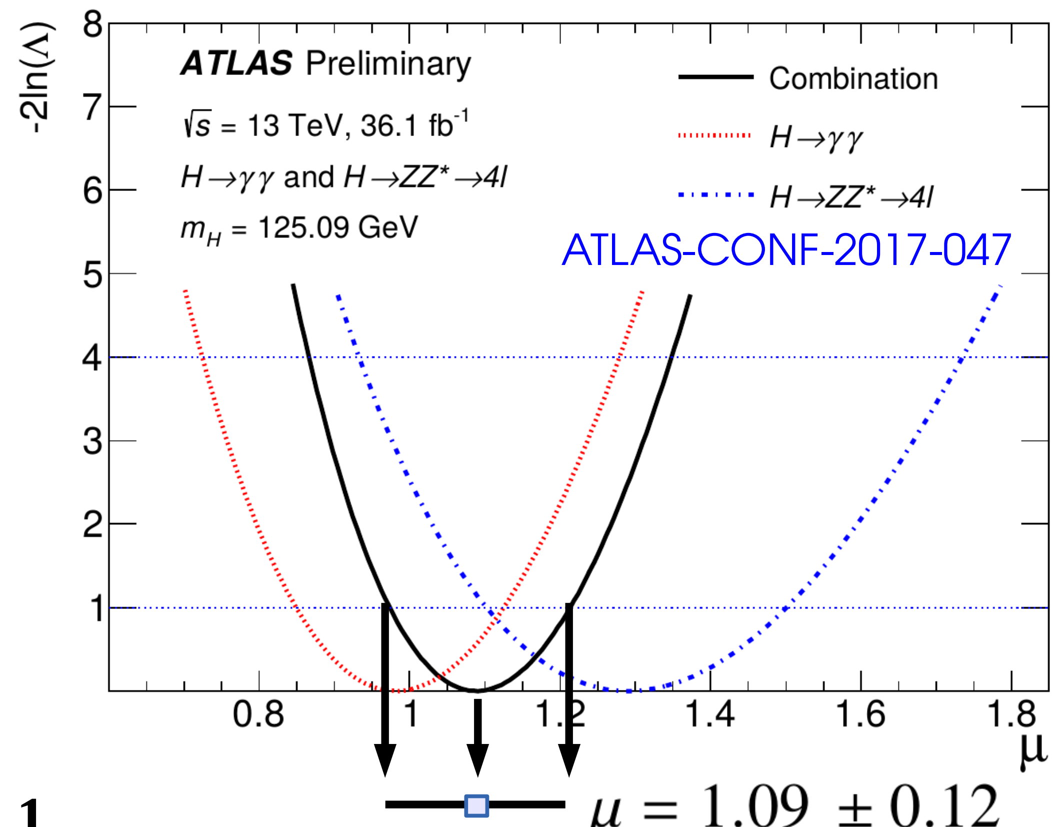
$$t_{\mu_0} = -2\log\frac{L(\mu=\mu_0)}{L(\hat{\mu})}$$

μ can be several POI!

Asymptotics: **$t_\mu \sim \chi^2(N_{POI})$ under $H(\mu_0)$**

In practice: (N_{POI}=1)

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



→ **Gaussian case**: parabolic profile,

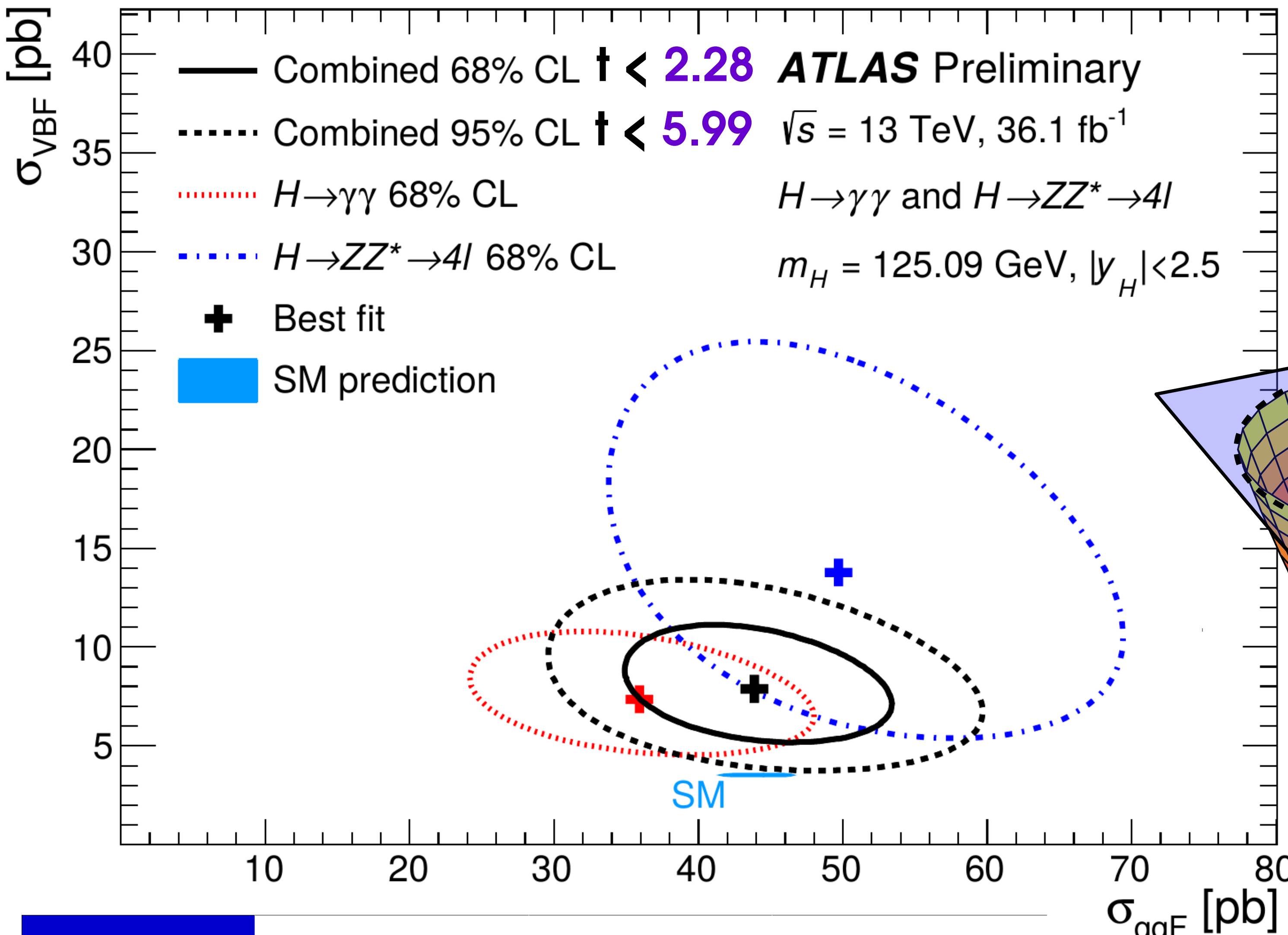
$$t_\mu = \left(\frac{\mu-\hat{\mu}}{\sigma}\right)^2 \Rightarrow \mu_\pm = \hat{\mu} \pm \sigma \text{ at } t_\mu = 1$$

→ robust against non-Gaussian effects.

→ **Can set upper limits on parameters using similar methods**

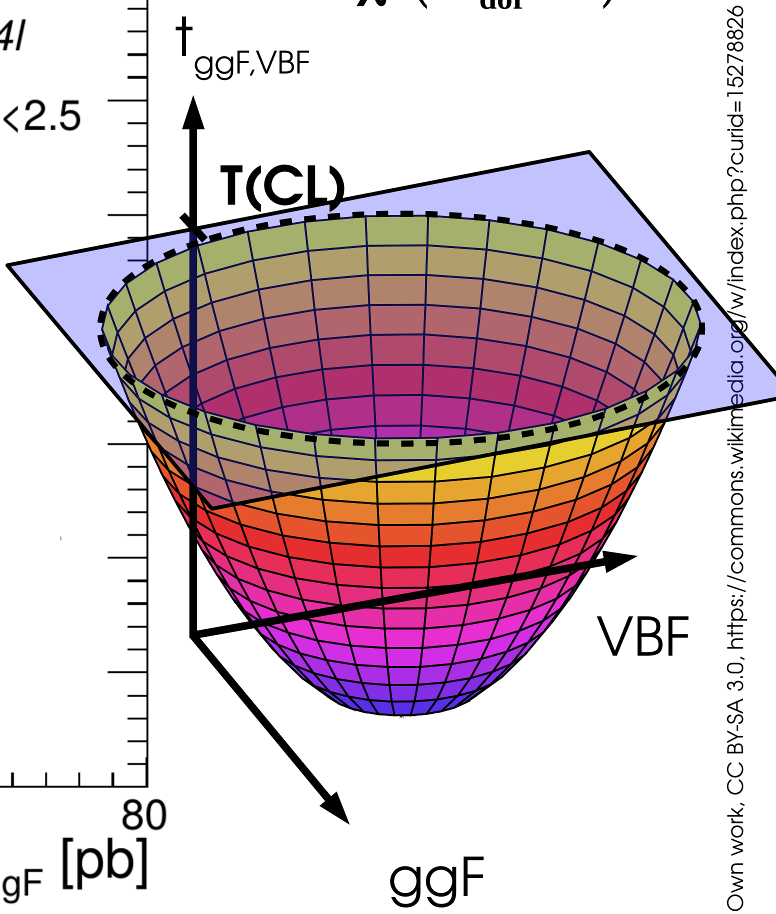
2D Example: Higgs σ_{VBF} vs. σ_{ggF}

ATLAS-CONF-2017-047



$$t = -2\log\frac{L(X_0, Y_0)}{L(\hat{X}, \hat{Y})}$$

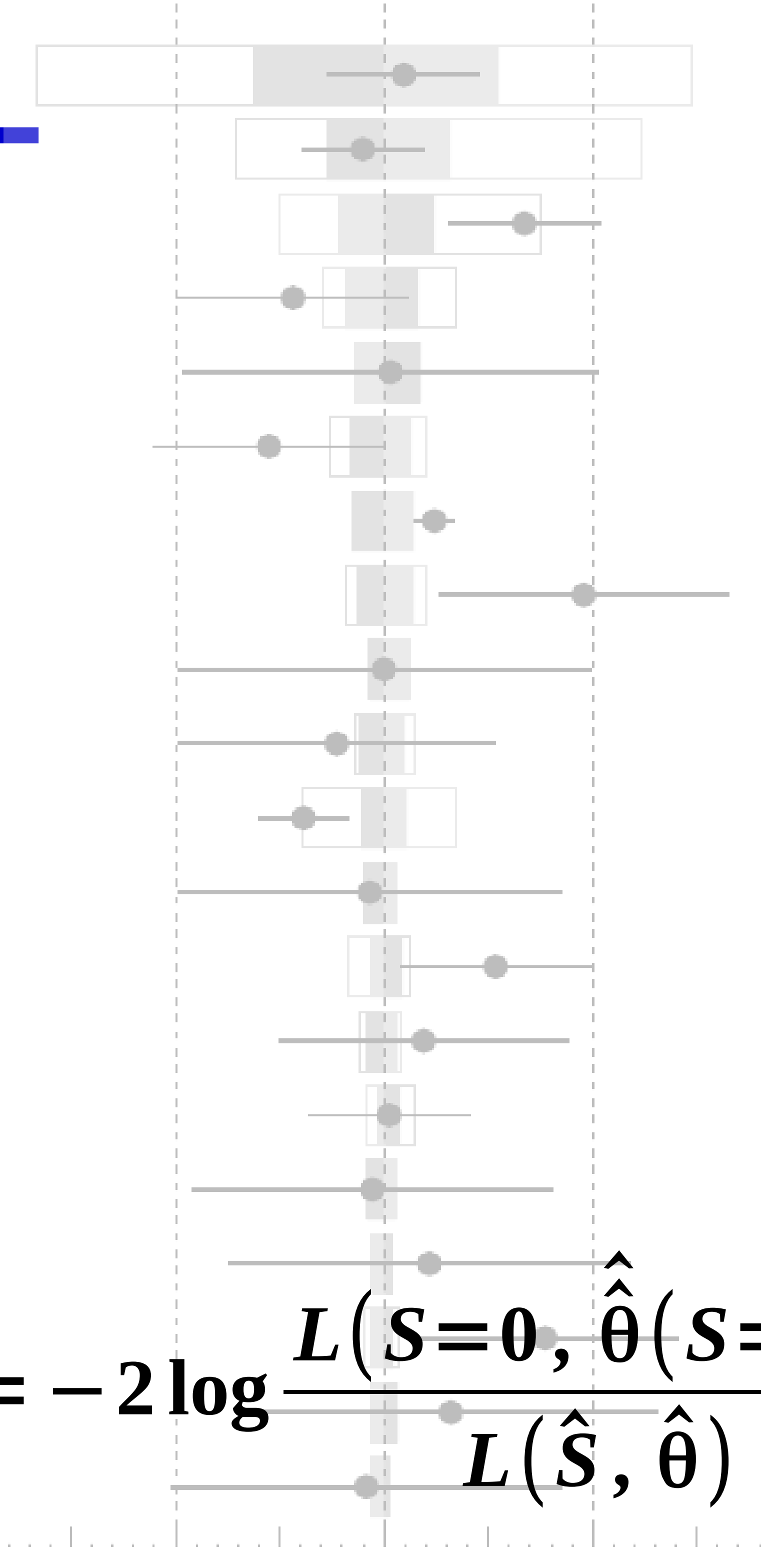
$$\sim \chi^2(N_{dof} = 2)$$



CL	68.3% (1σ)	68%	95%
1D T(CL)	1	0.989	3.84
2D T(CL)	2.30	2.28	5.99

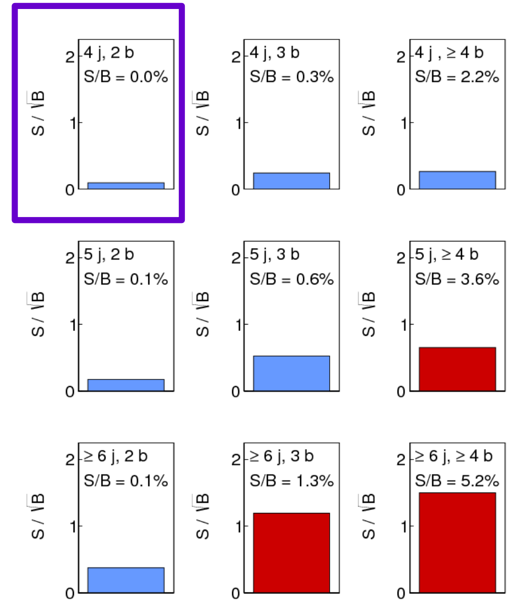
Gaussian case: elliptic paraboloid surface

Profiling



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

Profiling : ttH→bb as an example

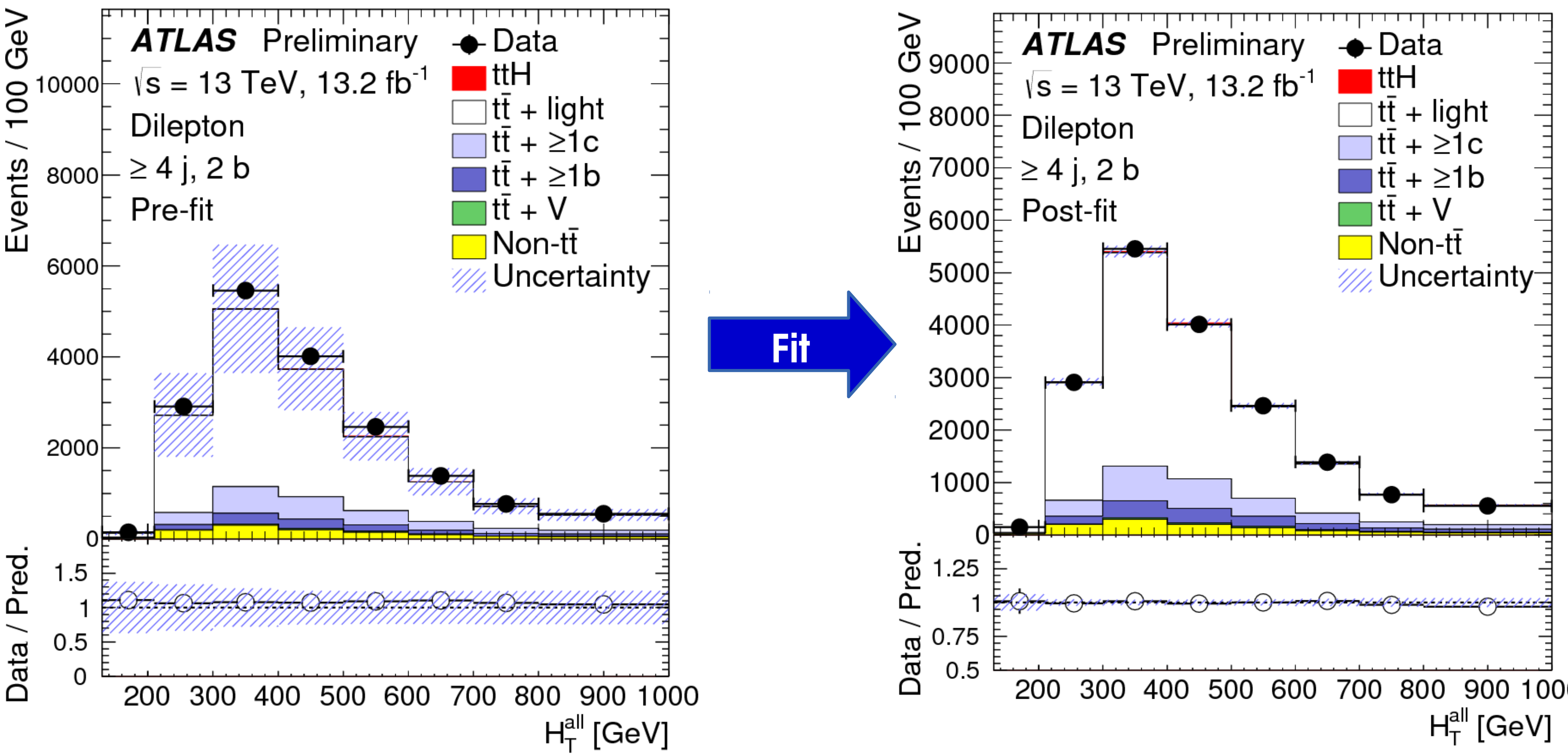


Analysis uses low-S/B categories to constrain backgrounds.

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

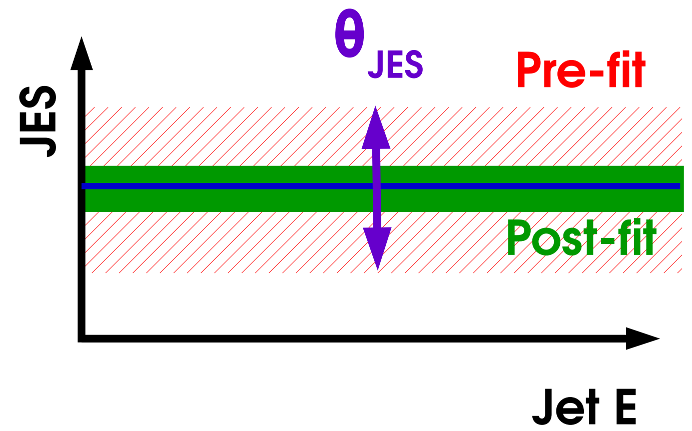
→ **Propagates to the high-S/B categories** through the statistical modeling

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



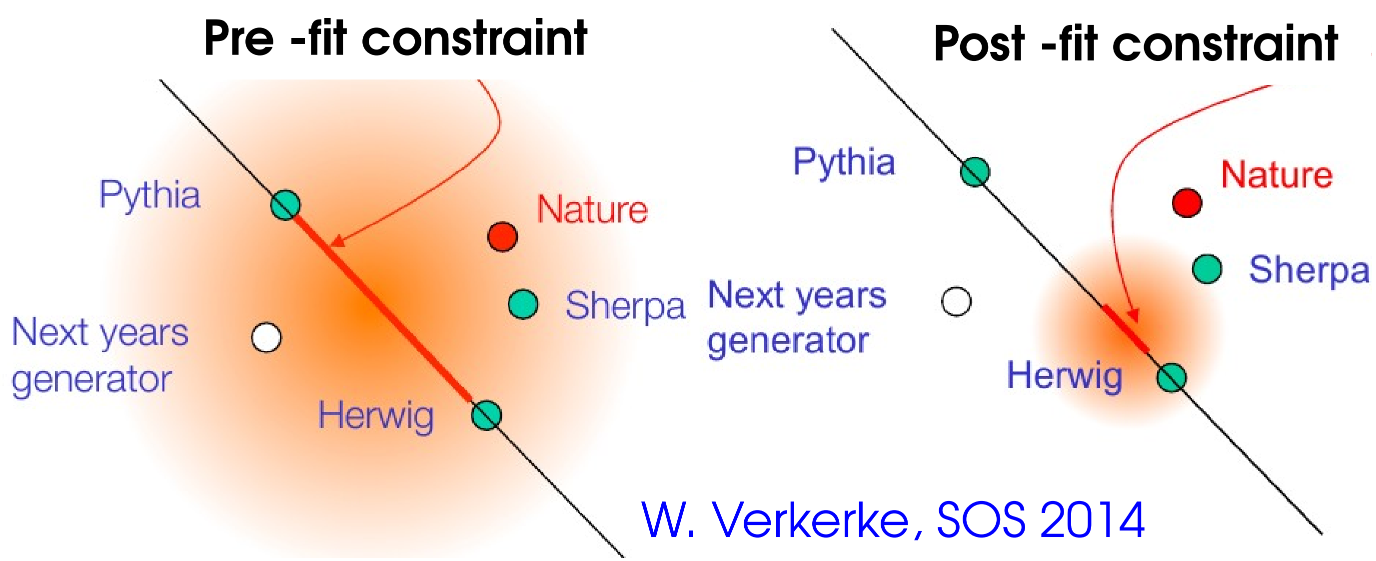
ATLAS-CONF-2016-08

Profiling Issues



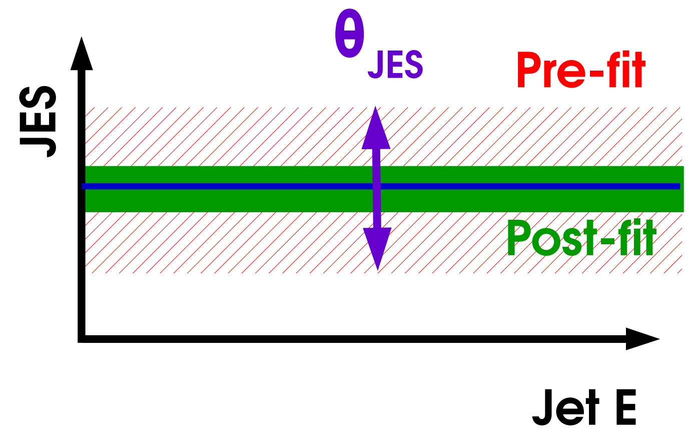
Too simple modeling can have unintended effects
→ e.g. single Jet E scale parameter:
⇒ Low-E jets calibrate high-E jets – intended ?

***Two-point* uncertainties**:
→ Interpolation may not cover full configuration space, can lead to too-strong constraints



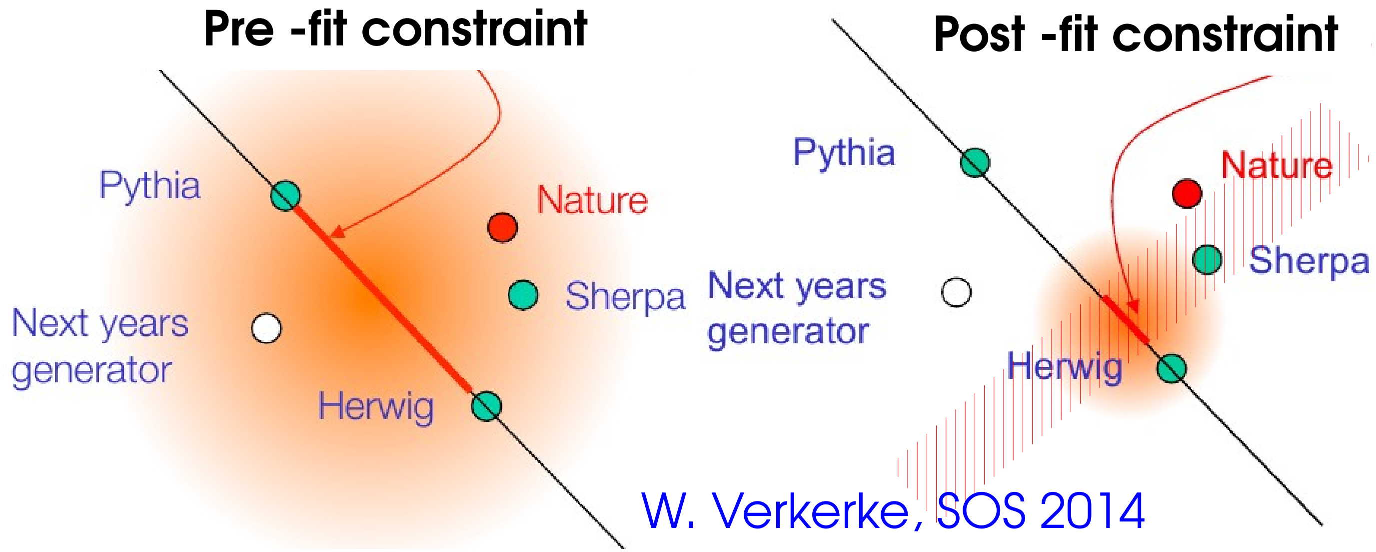
W. Verkerke, SOS 2014

Profiling Issues



Too simple modeling can have unintended effects
→ e.g. single Jet E scale parameter:
⇒ Low-E jets calibrate high-E jets – intended ?

***Two-point* uncertainties**:
→ Interpolation may not cover full configuration space, can lead to too-strong constraints



W. Verkerke, SOS 2014

Pull/Impact plots

ATLAS-CONF-2016-058

Critical to check syst modeling!
→ Ongoing program
→ Getting more important as syst uncertainties start to dominate

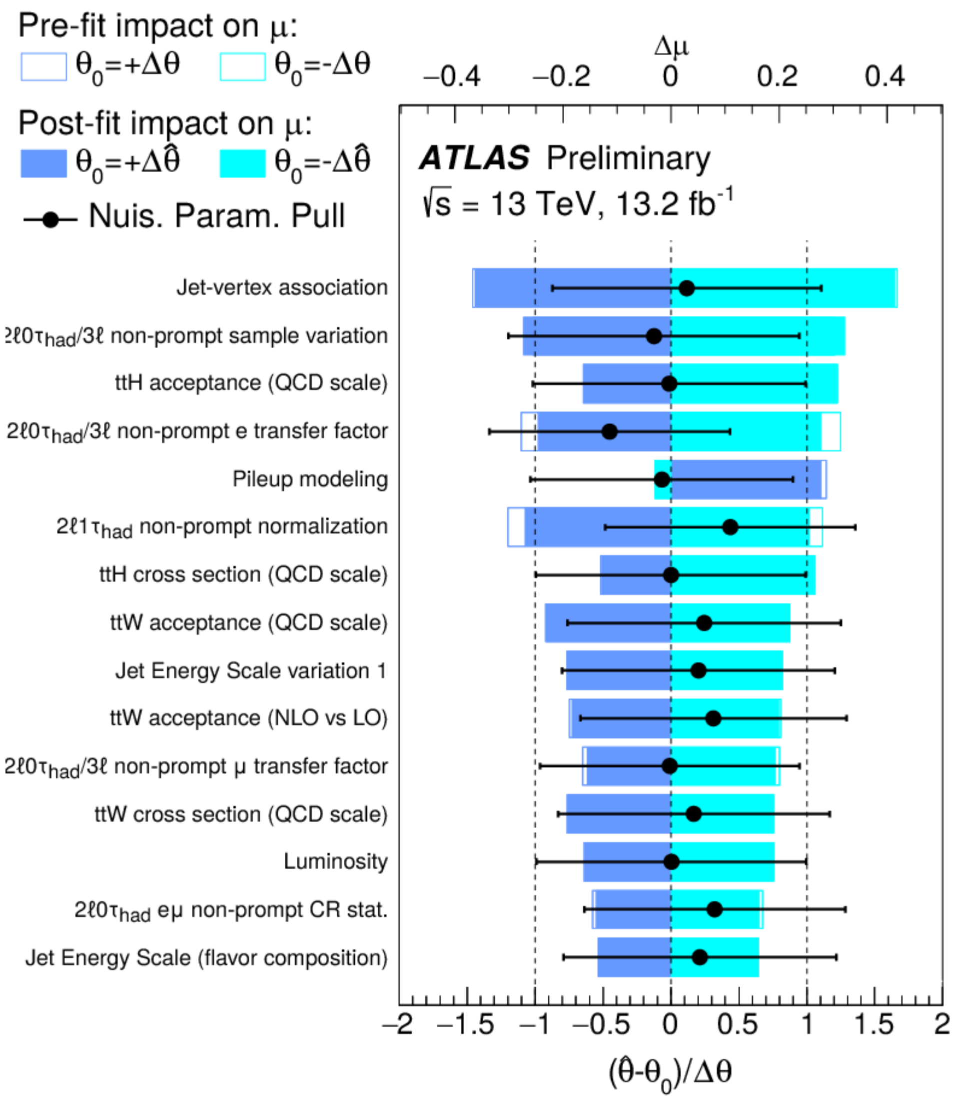
Example: pull/impact plots

Systematics NPs : usually

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

- **central value = 0,** → pre-fit expectation (usually MC)
  → If not: data/MC discrepancy ?
- **uncertainty = 1** (normalized to the magnitude of the systematic)
  → If not: syst NP constrained by data ⇒ legitimate, or modeling issue ?

Impact on result of ±1σ shift of NP



Pull/Impact plots

Critical to check syst modeling!
→ Ongoing program
→ Getting more important as syst uncertainties start to dominate

Example: pull/impact plots

Systematics NPs : usually

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

- **central value = 0,** → pre-fit expectation (usually MC)

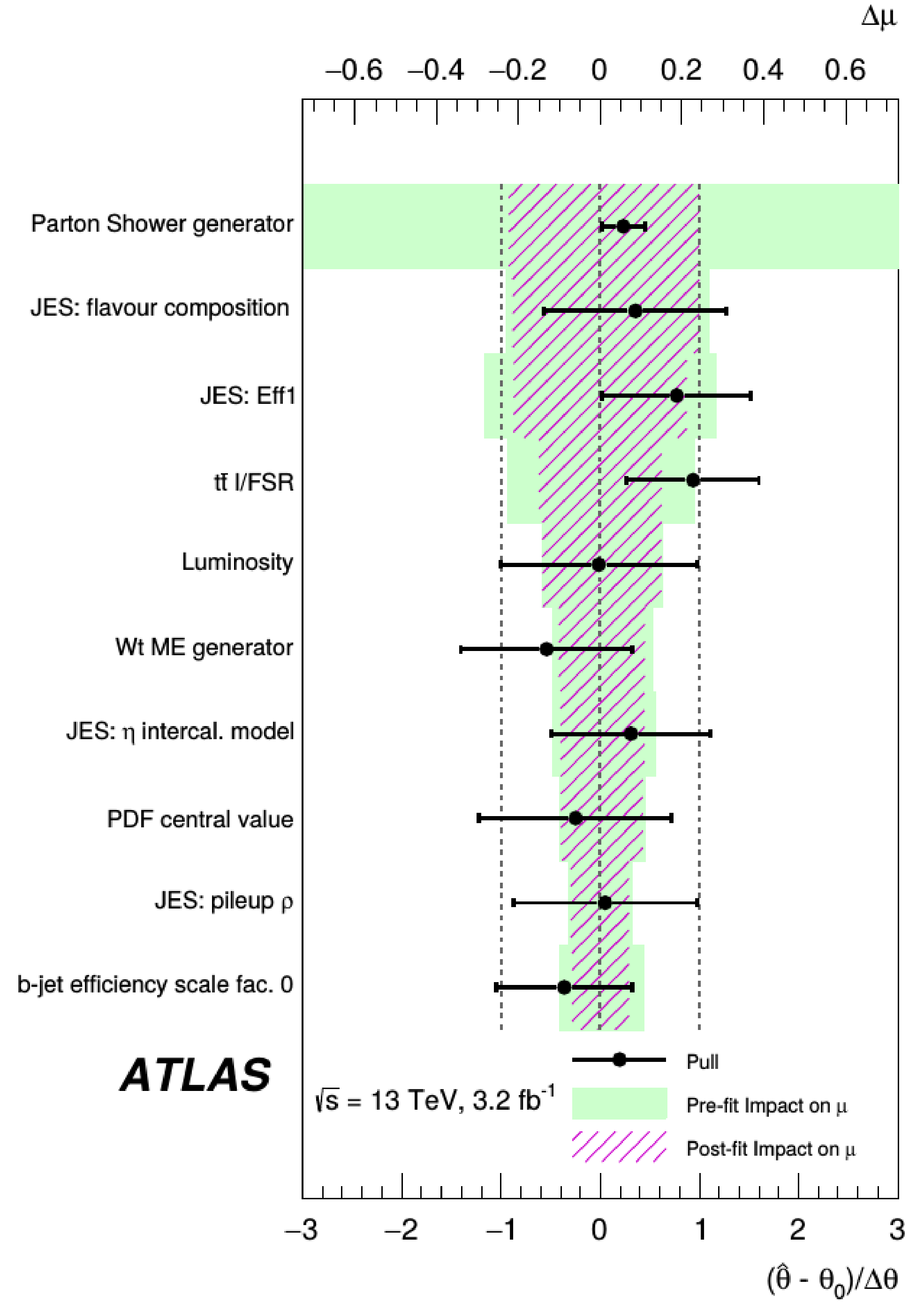
  → If not: data/MC discrepancy ?

- **uncertainty = 1** (normalized to the magnitude of the systematic)

  → If not: syst NP constrained by data ⇒ legitimate, or modeling issue ?

Impact on result of ±1σ shift of NP

13 TeV single-t XS (arXiv:1612.07231)



Pull/Impact plots

13 TeV single-t XS (arXiv:1612.07231)

Critical to check syst modeling!
→ Ongoing program
→ Getting more important as syst uncertainties start to dominate

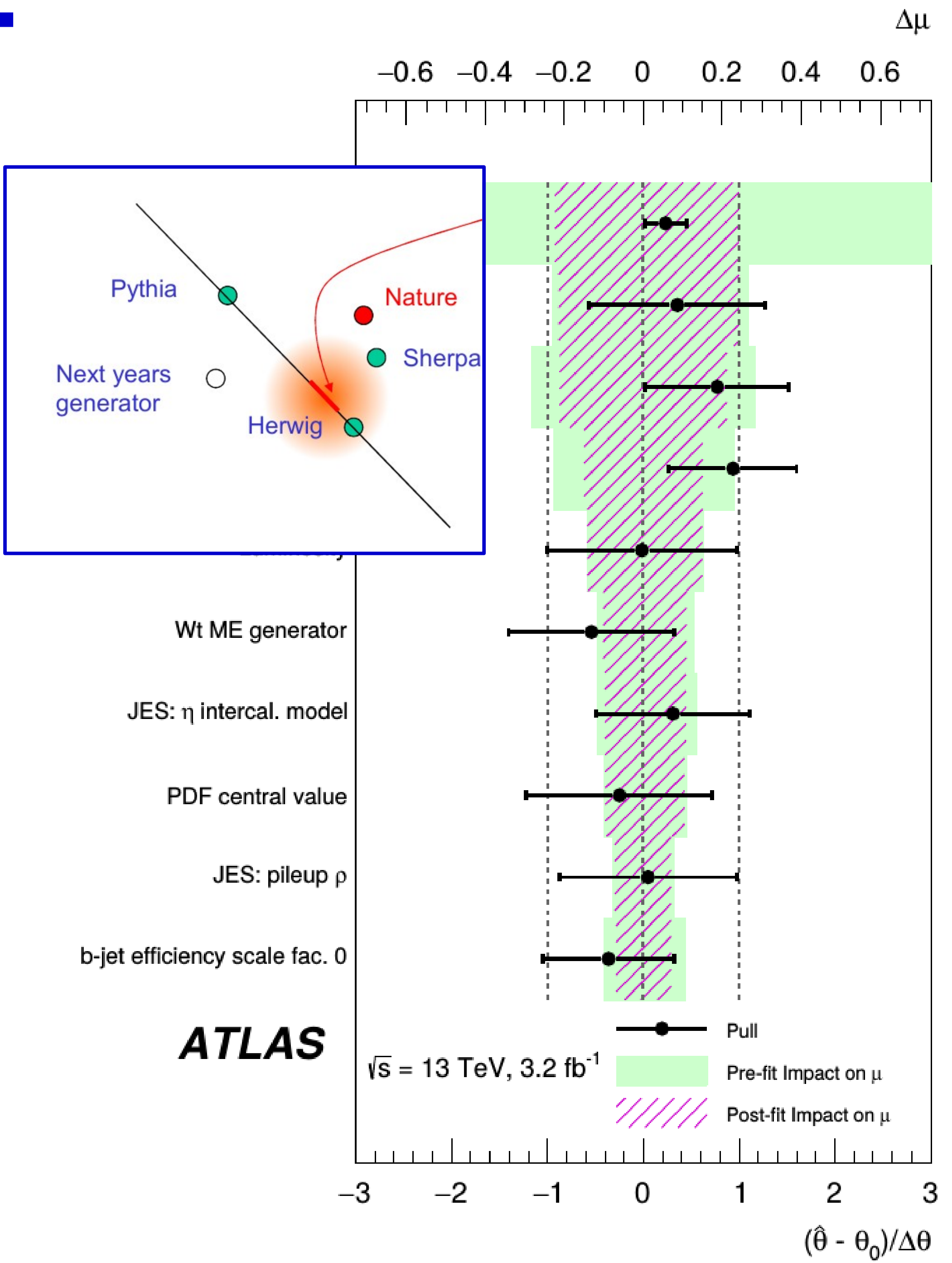
Example: pull/impact plots

Systematics NPs : usually

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

- **central value = 0,** → pre-fit expectation (usually MC)
  → If not: data/MC discrepancy ?
- **uncertainty = 1** (normalized to the magnitude of the systematic)
  → If not: syst NP constrained by data ⇒ legitimate, or modeling issue ?

Impact on result of ±1σ shift of NP



Presentation of Results

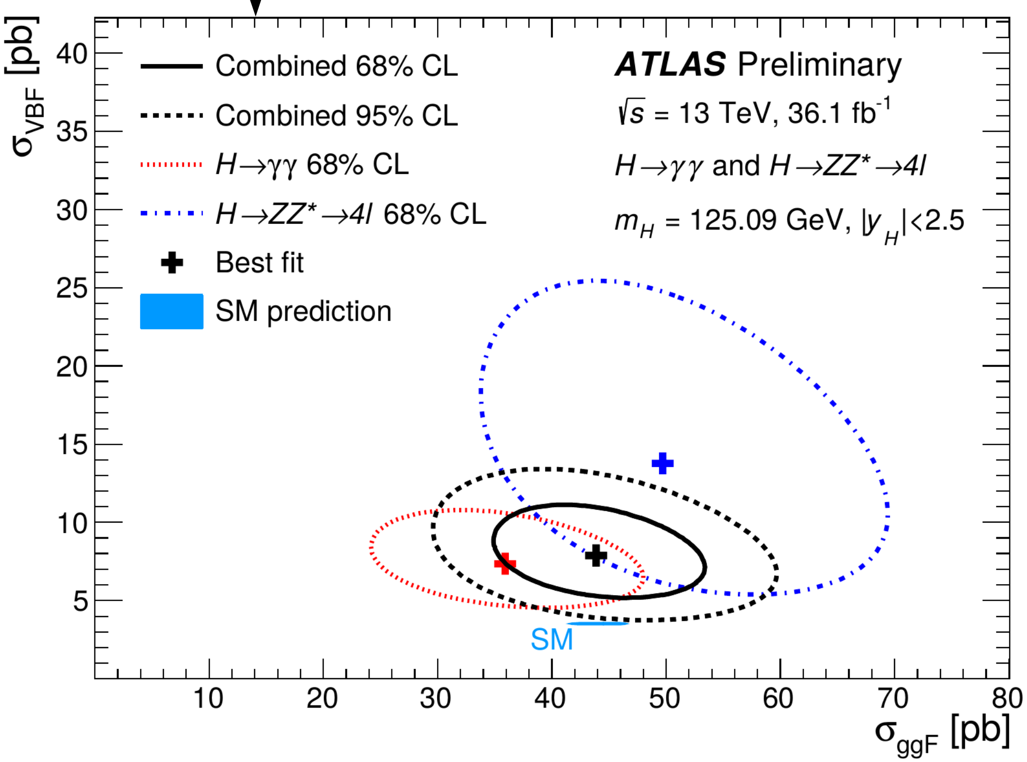
Reparameterization

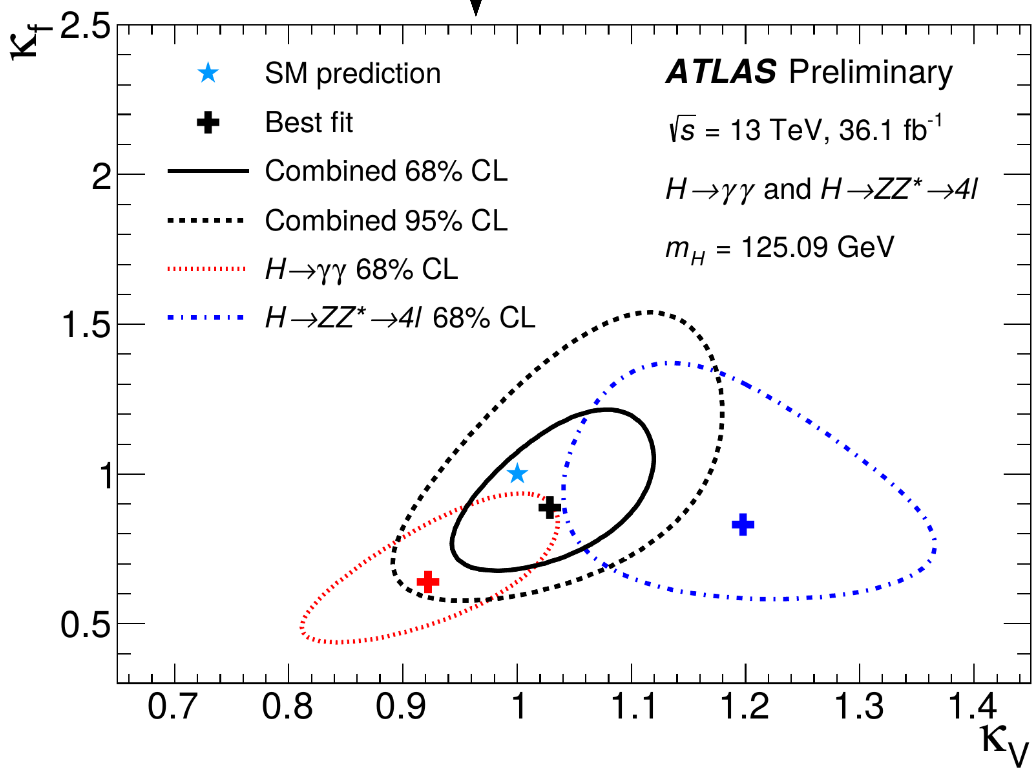
Start with basic measurement in terms of e.g. **$\sigma \times B$**

→ 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.

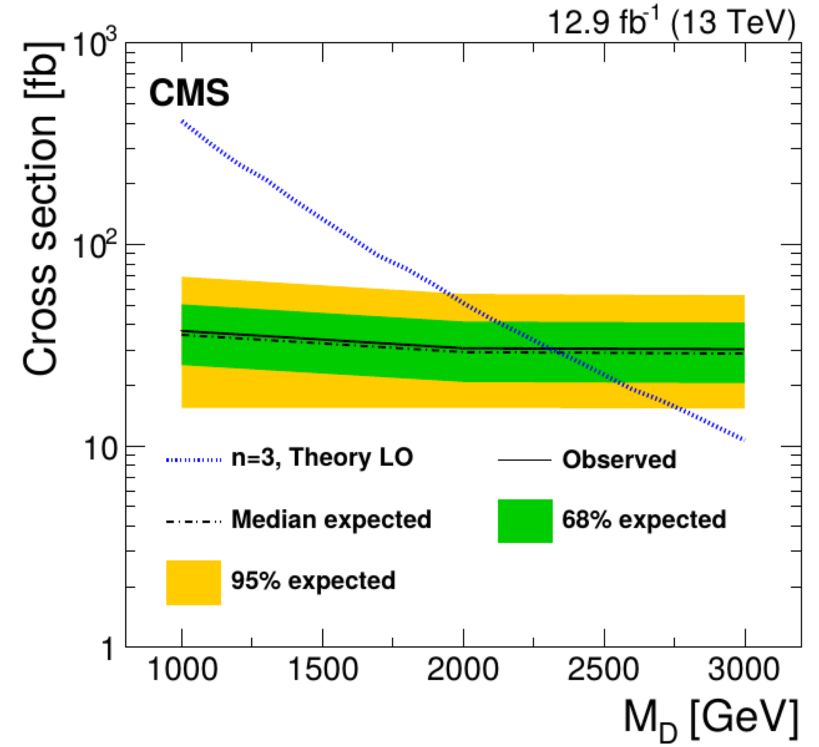
$$L(\sigma_{ggF}, \sigma_{VBF}) \xrightarrow[\sigma_{VBF} \to \sigma_{VBF}(\kappa_V, \kappa_F)]{\sigma_{ggF} \to \sigma_{ggF}(\kappa_V, \kappa_F)} L(\sigma_{ggF}(\kappa_V, \kappa_F), \sigma_{VBF}(\kappa_V, \kappa_F)) \equiv L'(\kappa_V, \kappa_F)$$

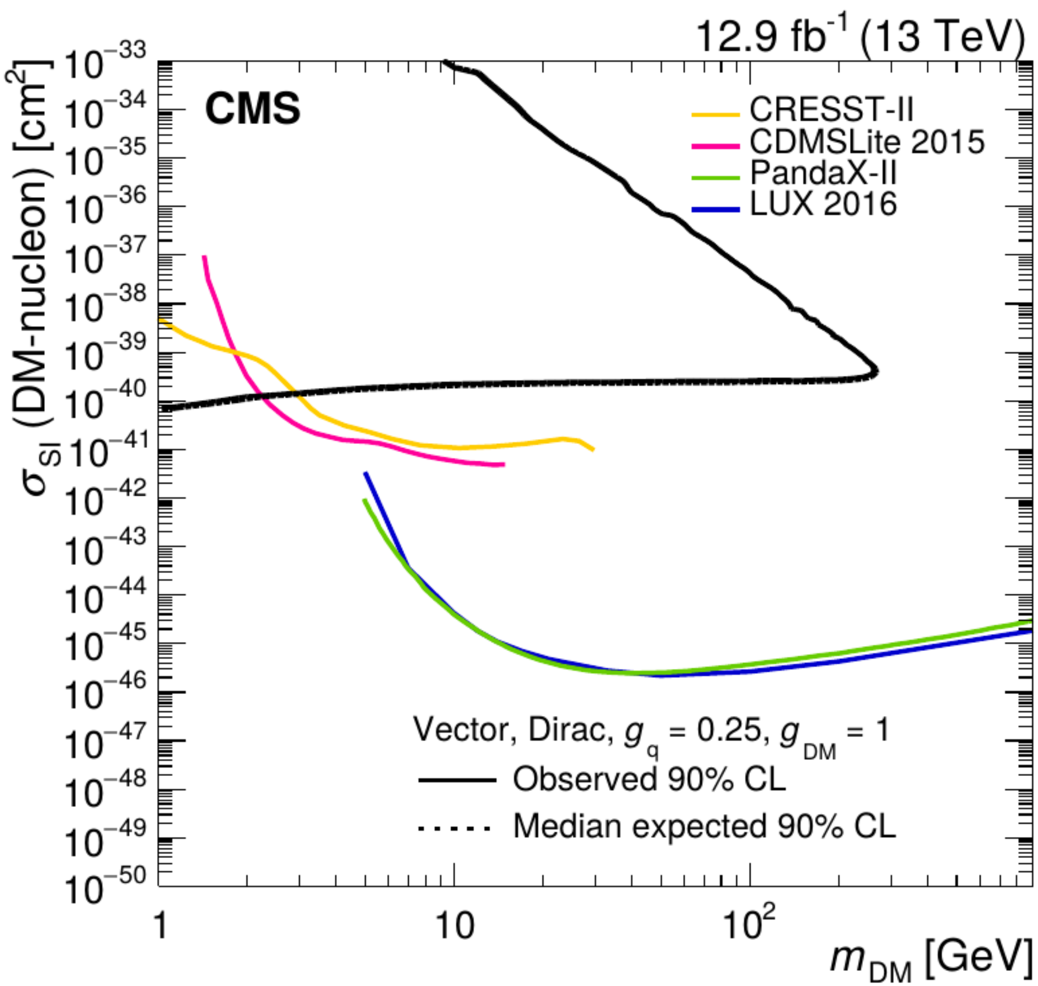


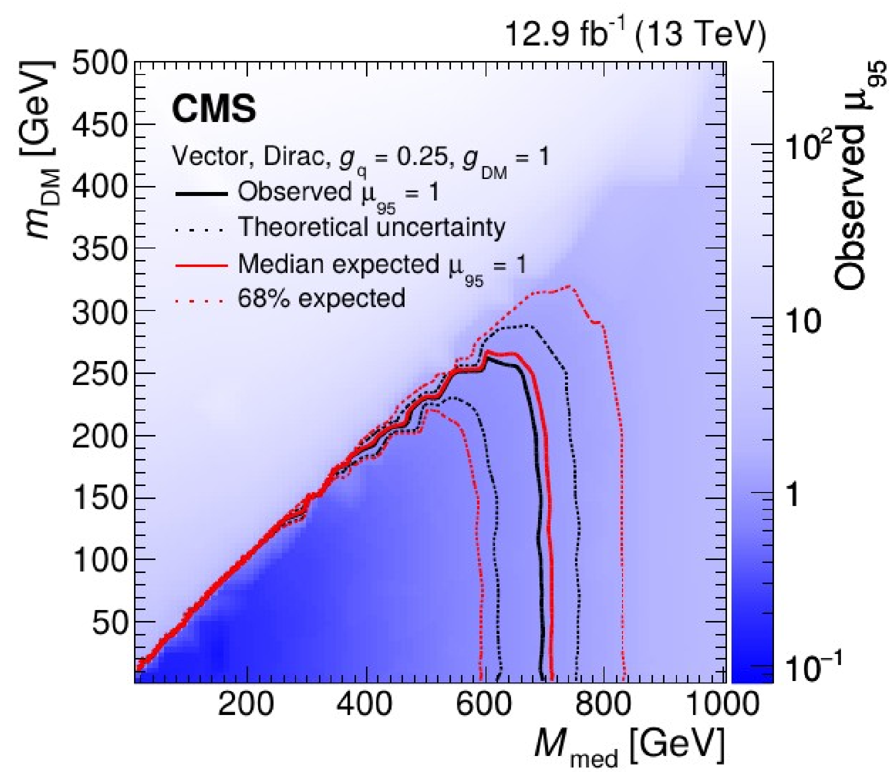


Reparameterization: Limits

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

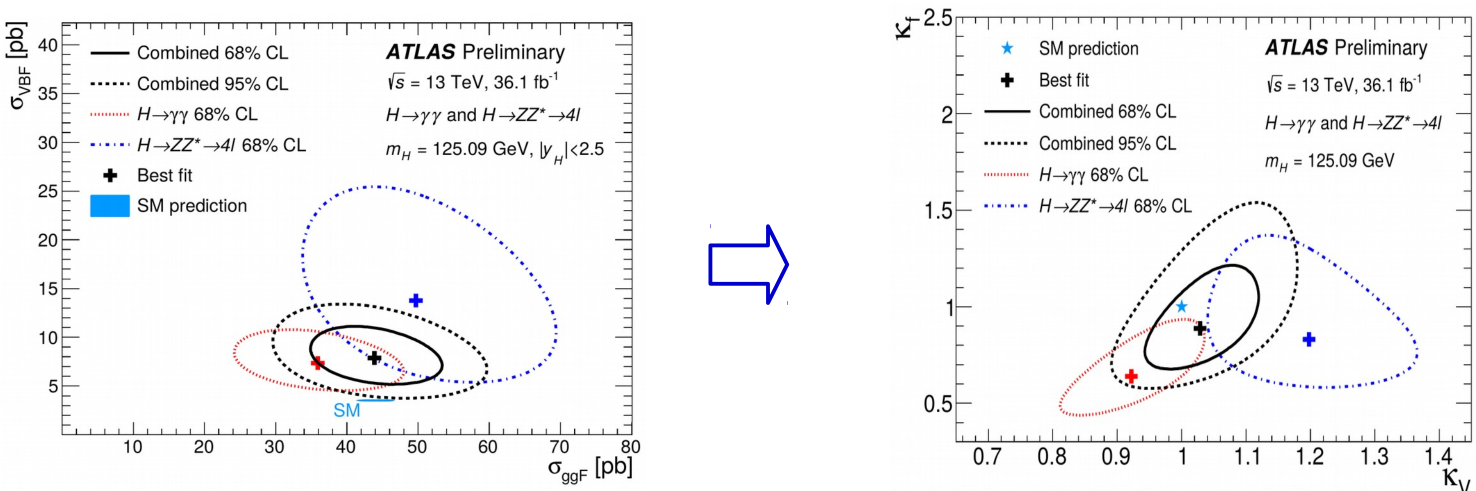






Presentation of Results

Measurements often recast to constrain a particular theory model.
→ Ideally, by **reparameterizing the likelihood** and repeating the measurement



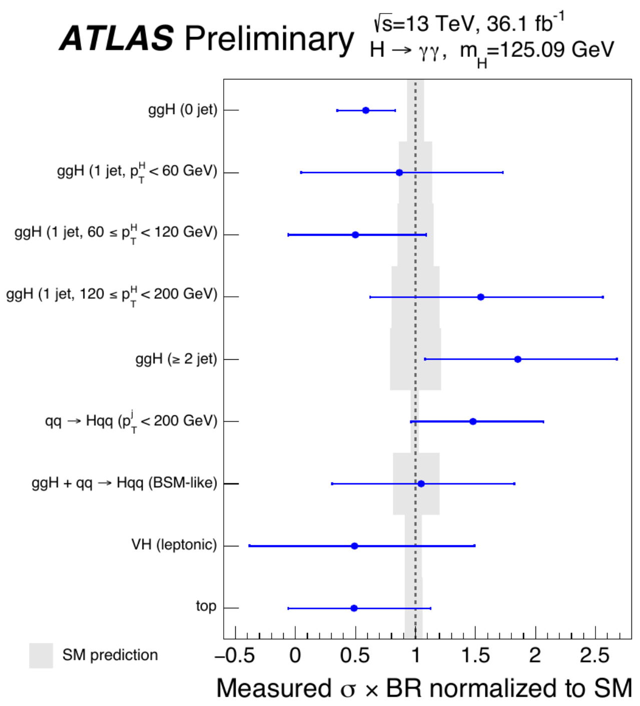
⇒ **Done by experiments for selected benchmark models**
→ However, **usually too complex to implement for many models**

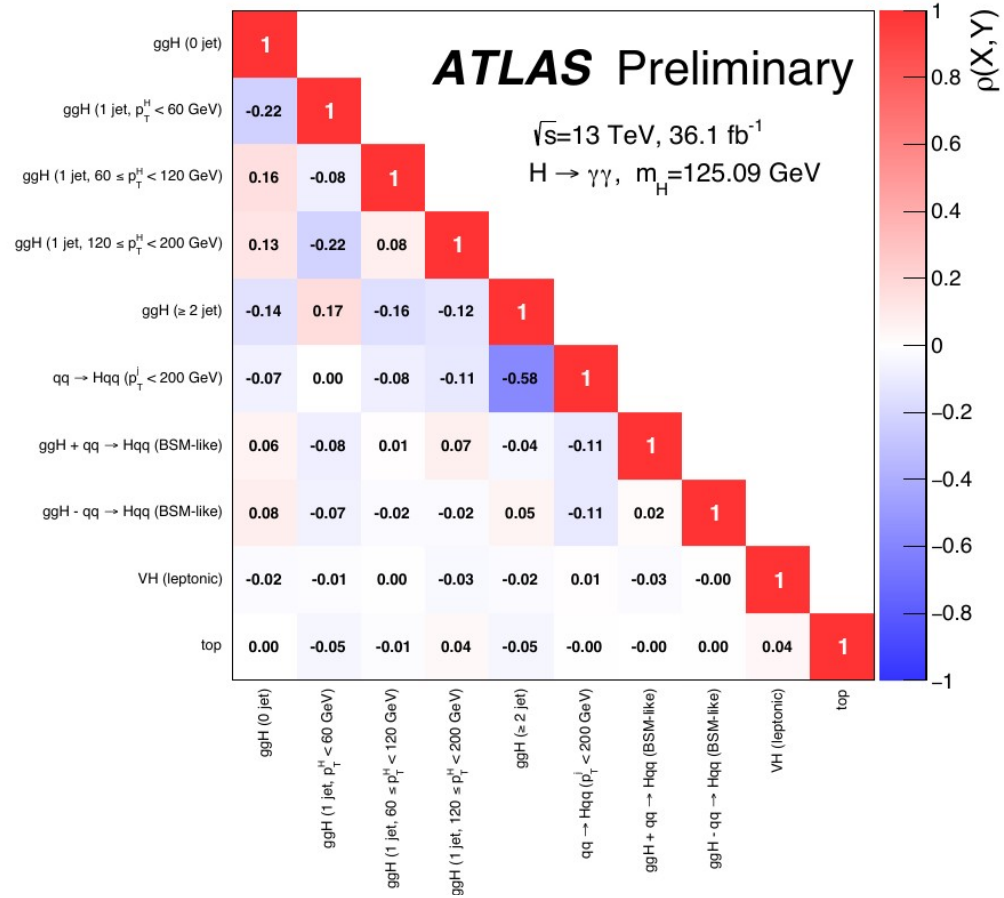
→ Publishing full likelihood typically impractical – most theorists do not want to deal with 4000 NPs...

→ **Other approaches:** e.g. reimplementing the analysis in a public fast-simulation framework (e.g. SUSY searches). However clear accuracy limitations

Presentation of Results

→ **Current solution**: publish covariance matrices in HEPData, together with the individual measurements





	ggH (0 jet)	ggH (1 jet, $p_T^H < 60$ GeV)	ggH (1 jet, $60 \le p_T^H < 120$ GeV)	ggH (1 jet, $120 \le p_T^H < 200$ GeV)	ggH (≥ 2 jet)	qq → Hqq ($p_T^j < 200$ GeV)	ggH + qq → Hqq (BSM-like)	ggH - qq → Hqq (BSM-like)	VH (leptonic)	top
ggH (0 jet)	1									
ggH (1 jet, $p_T^H < 60$ GeV)	-0.22	1								
ggH (1 jet, $60 \le p_T^H < 120$ GeV)	0.16	-0.08	1							
ggH (1 jet, $120 \le p_T^H < 200$ GeV)	0.13	-0.22	0.08	1						
ggH (≥ 2 jet)	-0.14	0.17	-0.16	-0.12	1					
qq → Hqq ($p_T^j < 200$ GeV)	-0.07	0.00	-0.08	-0.11	-0.58	1				
ggH + qq → Hqq (BSM-like)	0.06	-0.08	0.01	0.07	-0.04	-0.11	1			
ggH - qq → Hqq (BSM-like)	0.08	-0.07	-0.02	-0.02	0.05	-0.11	0.02	1		
VH (leptonic)	-0.02	-0.01	0.00	-0.03	-0.02	0.01	-0.03	-0.00	1	
top	0.00	-0.05	-0.01	0.04	-0.05	-0.00	-0.00	0.00	0.04	1

→ **Valid in the Gaussian approximation**

→ To go further, need some form of **simplified likelihoods**

- **Profile likelihood** – function of POI only (NPs profiled out)
- **Additional terms** for non-Gaussian effects

→ Significantly more complex (many dimensions!)

→ Will be needed eventually as measurements become syst-dominated

Other Methods

BLUE

Commonly-used ansatz for combination of measurements:

1. **Build a χ²:** (same as -2logL for Gaussian L)

$$\chi^2(X) = \sum_i \left(X_i^{obs} - X\right) C_{ij}^{-1} \left(X_j^{obs} - X\right)$$

C_{ij} : covariance matrix of measurements:

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

ρ: correlation coefficients

2. **Estimate combined X from minimum of χ²(X)**

$$\lambda = \frac{C^{-1}J}{J^T C^{-1} J}, \quad J = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$$

- In the Gaussian case, **equivalent to ML solution**
  → **"Best" :** minimizes the combined uncertainty
- Solution is a linear combination of the inputs:

λ_i = combination weight of measurement i

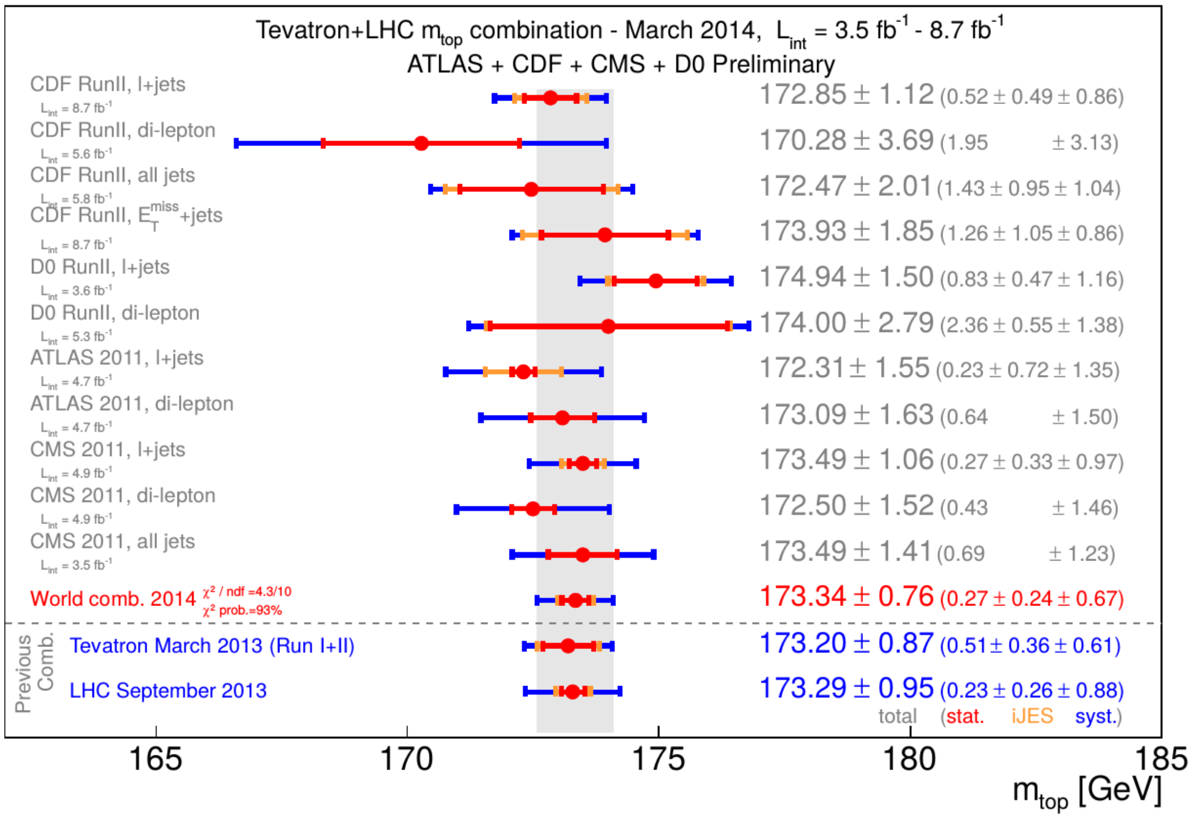
$$\hat{X} = \sum_i \lambda_i X^{obs,i}$$

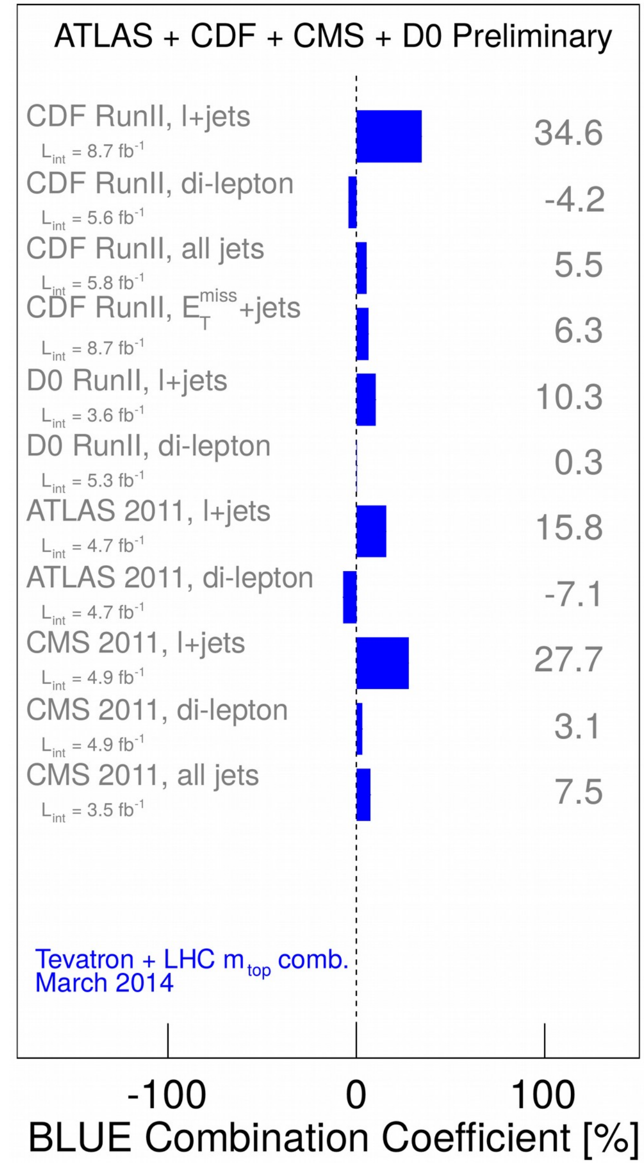
⇒ **"Best Linear Unbiased Estimator"** (**BLUE**)

BLUE Example

ATLAS-CONF-2014-008

Example: World m_{top} combination





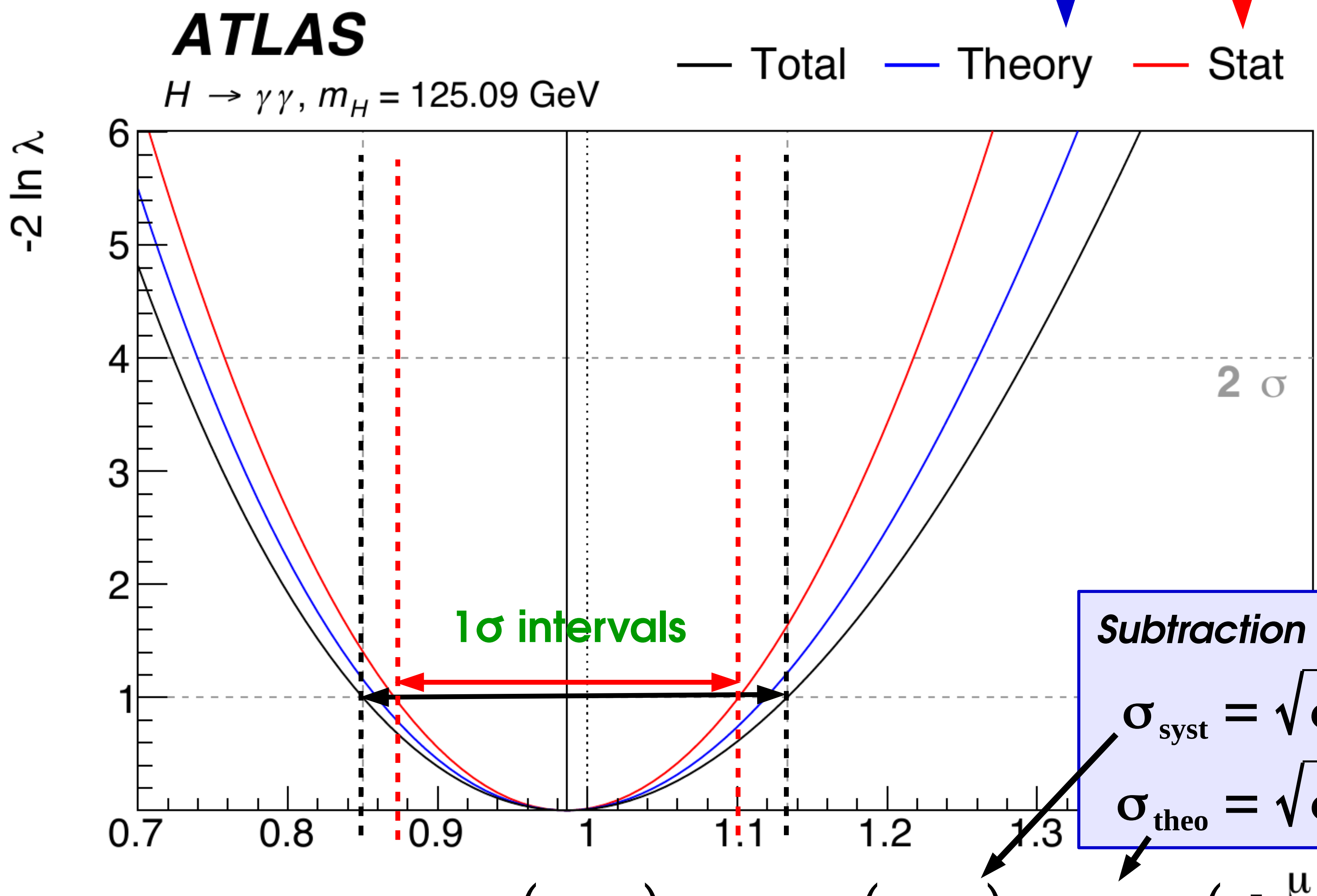
Limitation: relies on **Gaussian assumptions** (satisfied in this case!)

Non-trivial results for strong correlations, see Eur. Phy. J. C 74 (2014), 2717)

Uncertainty decomposition

All systematics NPs fixed to 0 : statistical uncertainty only

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



Subtraction in quadrature

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

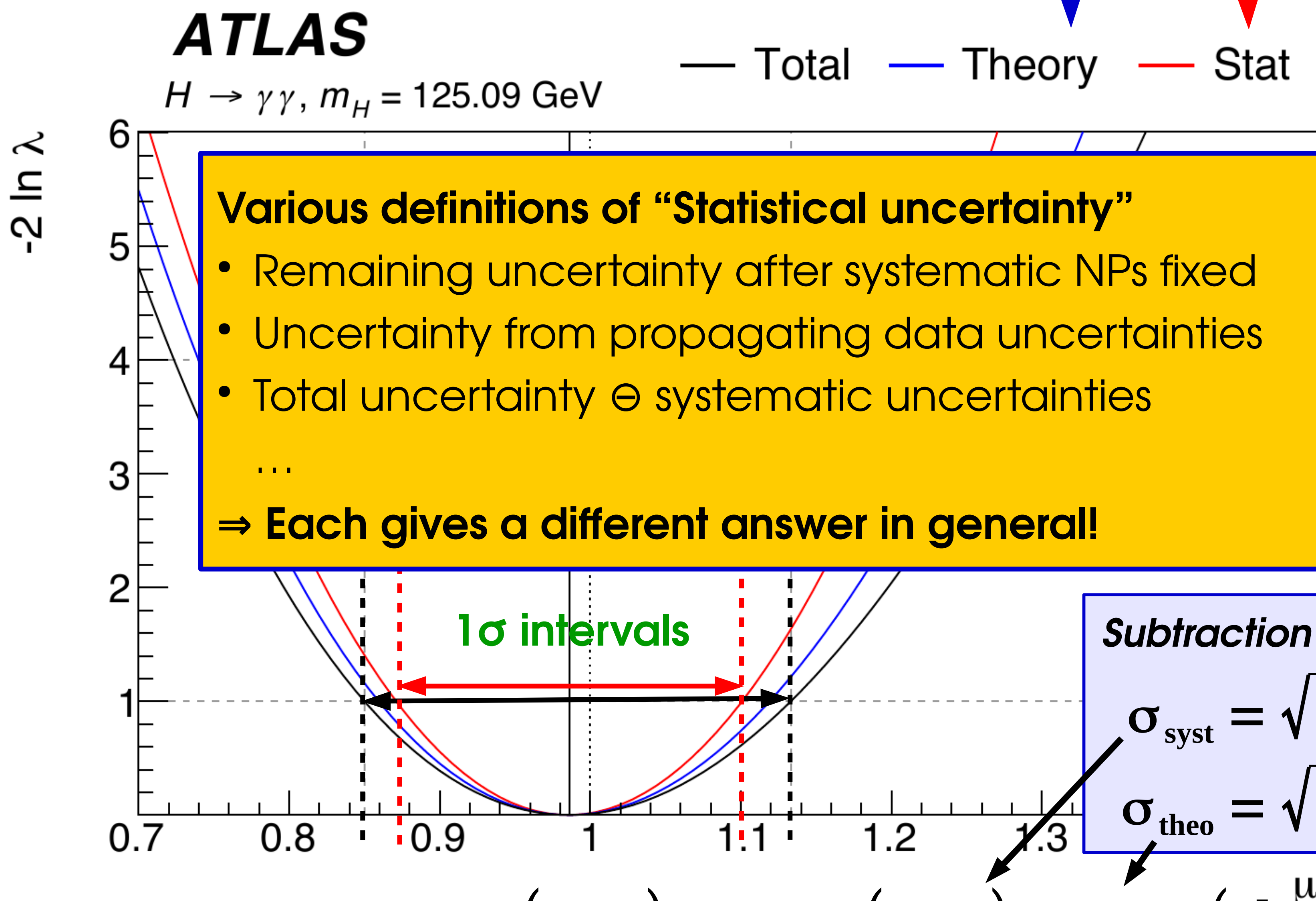
$$\sigma_{theo} = \sqrt{\sigma^2_{stat+theo} - \sigma^2_{stat}}$$

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

Uncertainty decomposition

All systematics NPs fixed to 0 : statistical uncertainty only

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



Various definitions of "Statistical uncertainty"

- Remaining uncertainty after systematic NPs fixed
- Uncertainty from propagating data uncertainties
- Total uncertainty ⊖ systematic uncertainties
- …

⇒ **Each gives a different answer in general!**

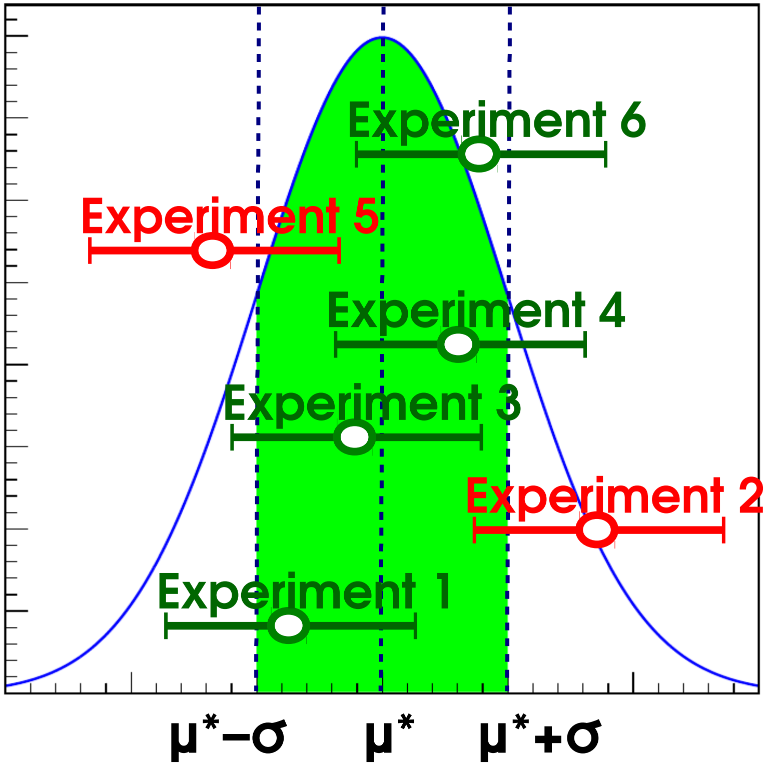
Subtraction in quadrature

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

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

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

Frequentist vs. Bayesian



All methods described so far are **frequentist**

- Probabilities (p-values) refer to outcomes if the experiment were **repeated identically many times**
- Parameters value are **fixed but unknown**

- Probabilities apply to measurements:

→ **"m_H = 125.09 ± 0.24 GeV"** :

→ i.e. $[125.09 - 0.24\ ;\ 125.09 + 0.24\ ]$ GeV has p=68% to contain **the** true m_H.

→ if we repeated the experiment many times, we would get different intervals, 68% of which would contain the true m_H.

→ **"5σ Higgs discovery"**

- if there is really no Higgs, such fluctuations observed in 3.10^{-7} of experiments

Not exactly the crucial question – what we would really like to know is

What is the probability that the excess we see is a fluctuation

→ we want **P(no Higgs | data)** – but all we have is **P(data | no Higgs)**

Frequentist vs. Bayesian

Can use **Bayes' theorem** to address this:

$$P(\mu|data) = \frac{P(data|\mu)}{P(data)} P(\mu)$$

same as in the frequentist formalism (=likelihood)

Prior Probability

irrelevant normalization factor

Can compute P(μ | data), **if we provide P(μ)**

→ Implicitly, we have now made μ into a random variable

- Is m_H, or the presence of H(125), randomly chosen ?
- In fact, different definition of p: ***degree of belief***, not from frequencies.
- P(μ) ***Prior degree of belief*** – critical ingredient in the computation

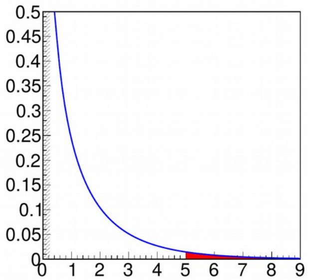
Compared to frequentist PLR:

⊕ answers the "right" question

⊖ answer depends on the prior

"Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentist use impeccable logic to deal with an issue that is of no interest to anyone." - **Louis Lyons**

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*

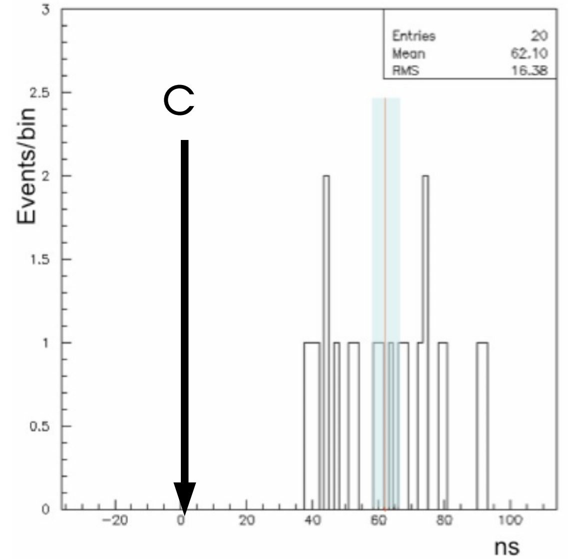
However this is **NOT** "~~*One chance in 10^7 to be a fluctuation*~~"

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.

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

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

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."*

"Extraordinary claims require extraordinary evidence"

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

Alternative:

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

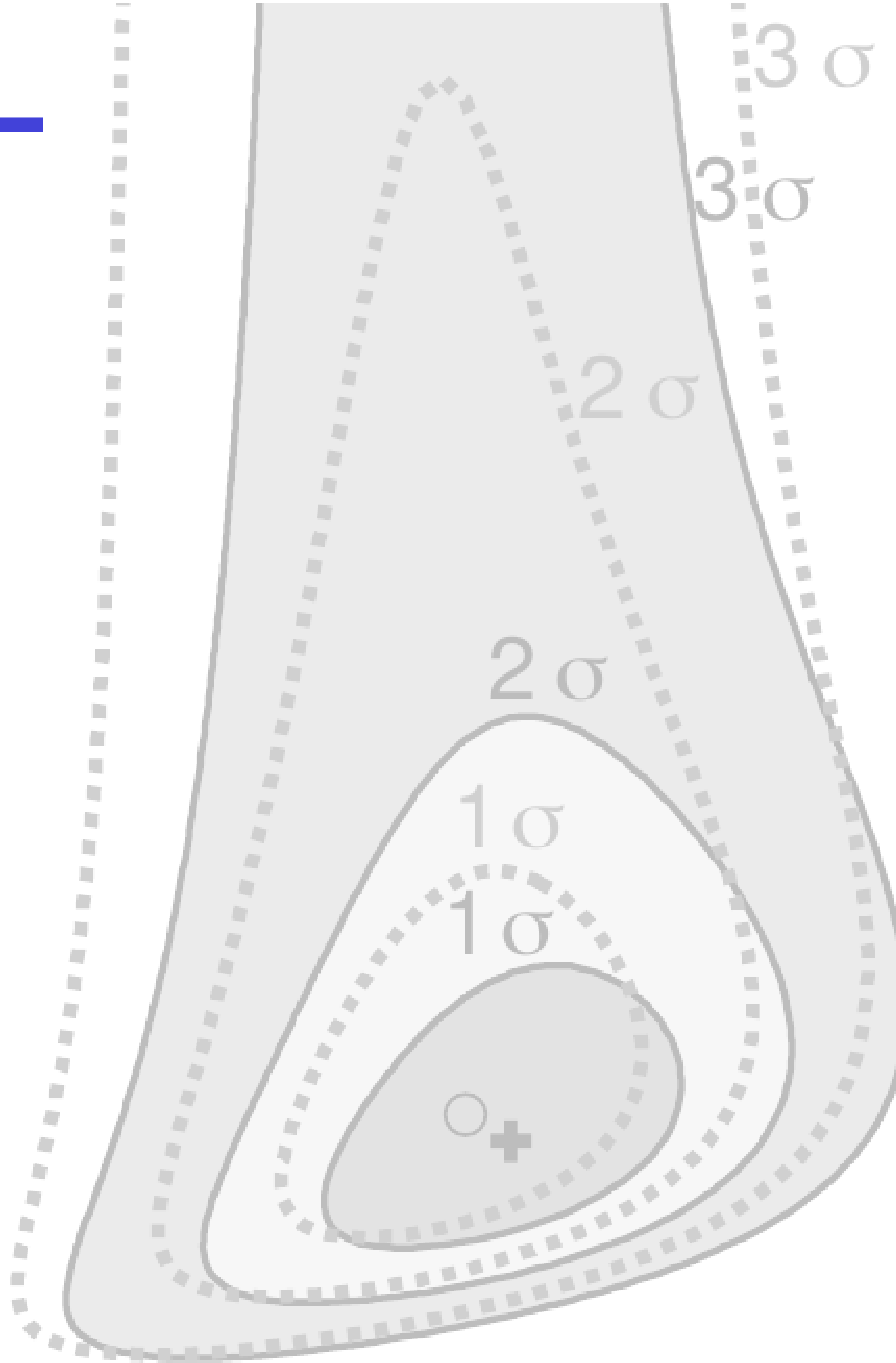
$$= \frac{P(\text{deviation}|B)\,P(B)}{P(\text{deviation}|S)\,P(S) + P(\text{deviation}|B)\,P(B)}$$

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

→ In frequentist context, only have **p_0 = P(deviation | B)**

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

Expected Sensitivity

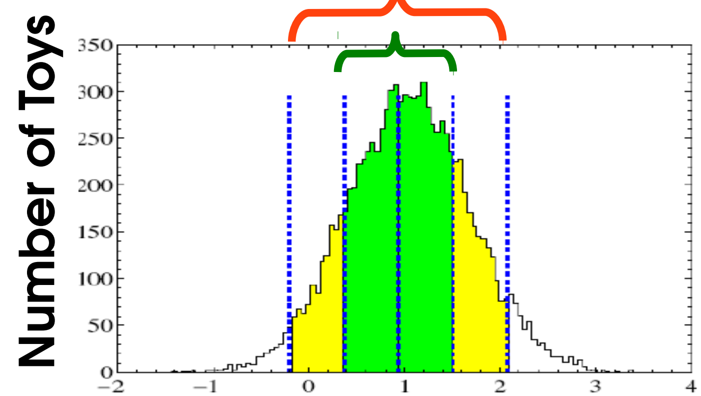


Expected Results

Eur.Phys.J.C71:1554,2011

***Expected* results**: median outcome under a given hypothesis
→ e.g. SM case, background only, etc

68% of toys 95% of toys



Computed result

Two main ways to compute:

→ **Pseudo-experiments**: use statistical model to generate pseudo-data ("toy data"),

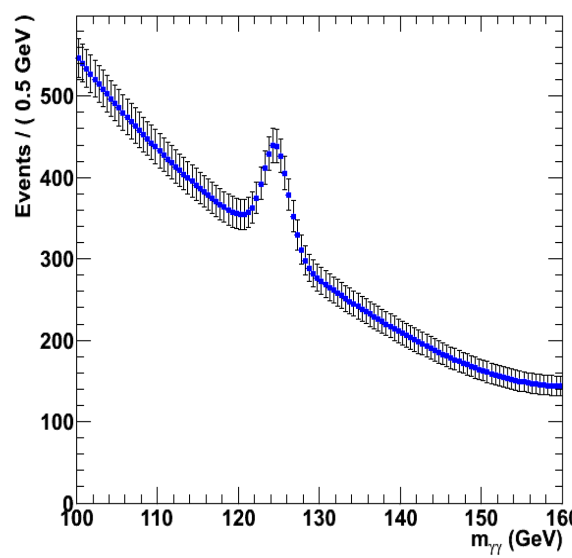
→ ***Asimov Datasets***

- Generate a "perfect dataset"
  e.g. for binned data, each bin set to expectation:
- 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**



Strictly speaking, Asimov for X_0

⇔ $\hat{X} = X_0$ **for all parameters** X,

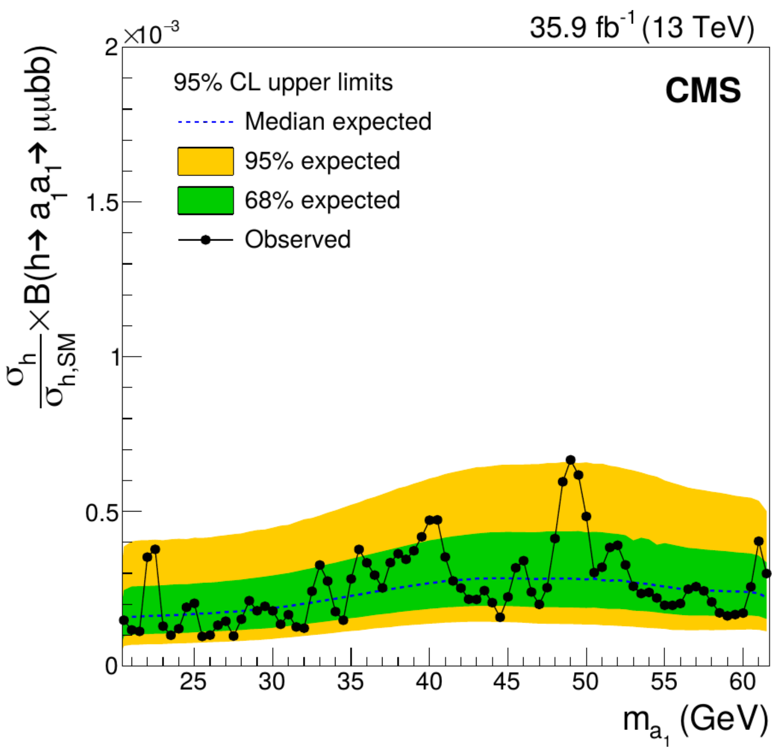
Expected Limits

1D: Asimov & toys give similar results → Asimov used in most cases

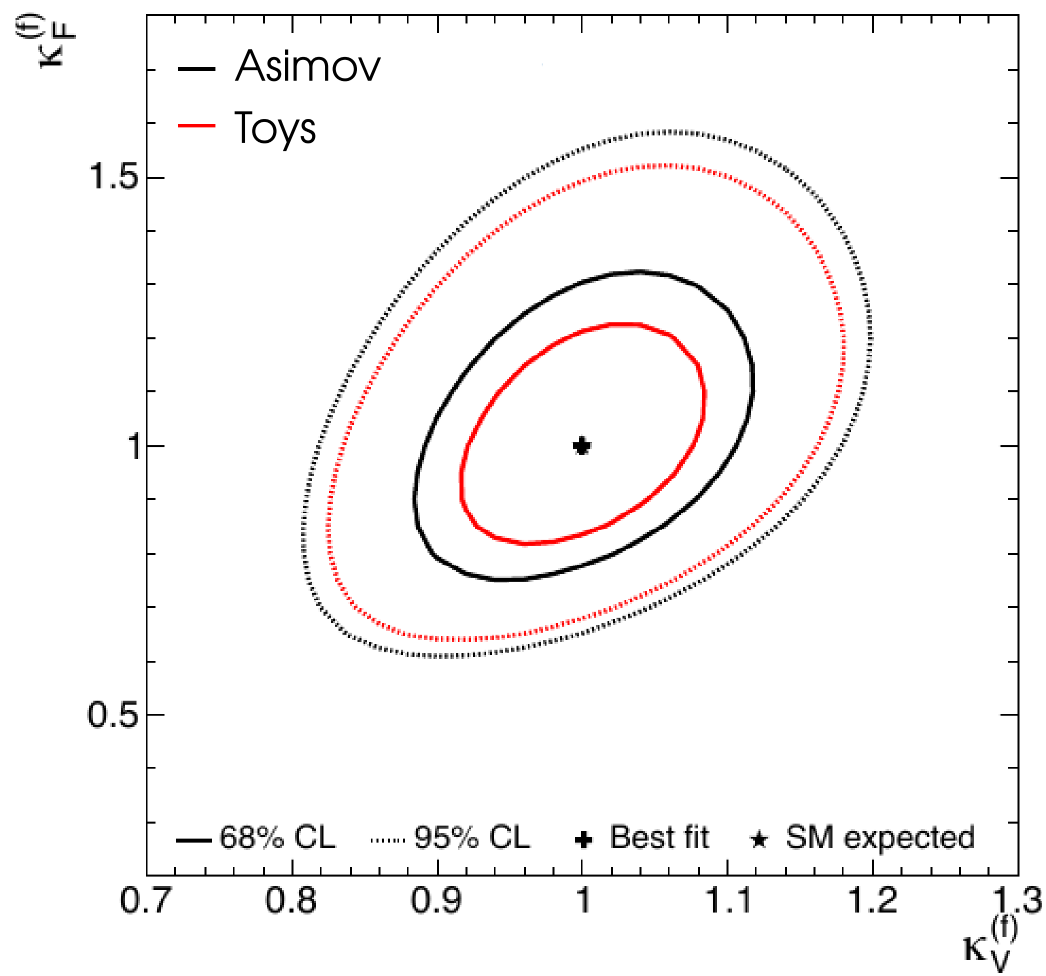
2D: **Different results: "Typical" and "Median" exclusion do not match!**

→Asimov still preferred since "typical" is usually more relevant.

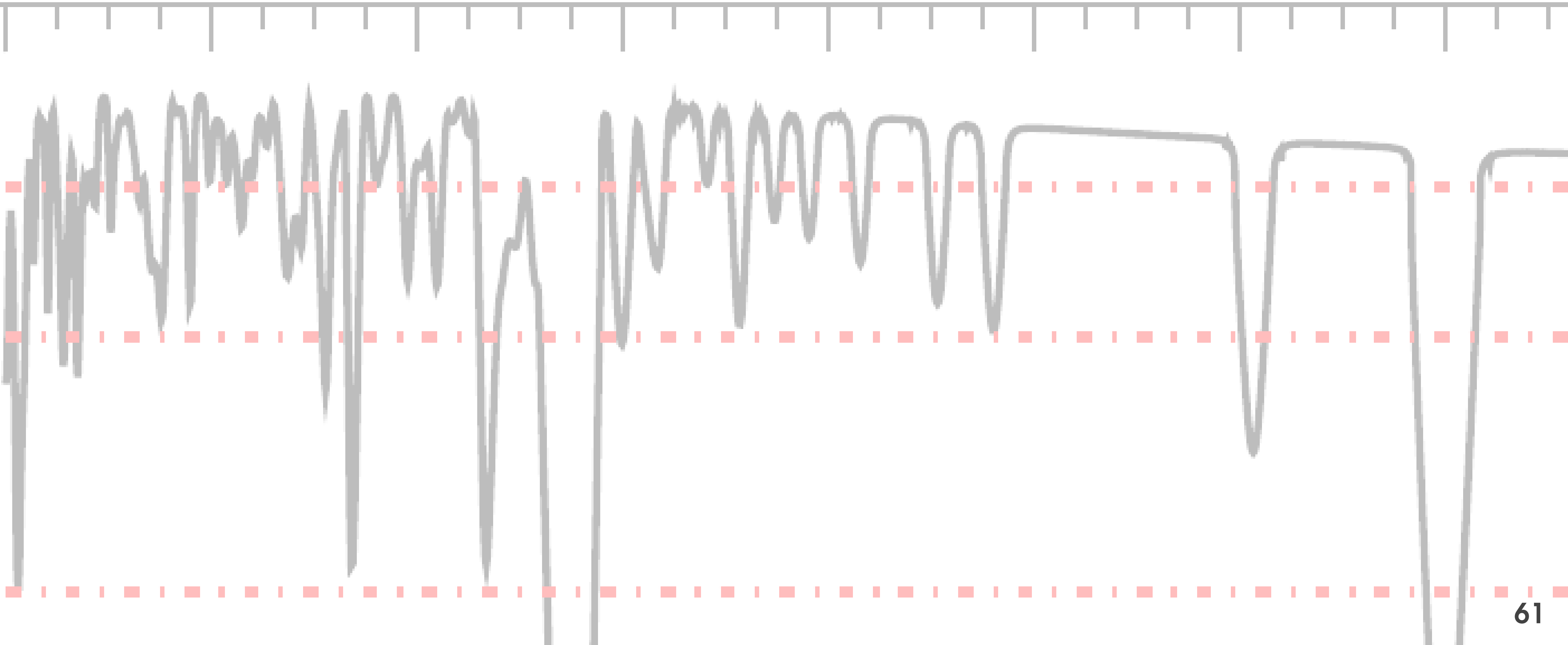
CMS-HIG-18-011 ; CERN-EP-2018-309



Scan courtesy of Stefan Gadatsch



Look-Elsewhere Effect

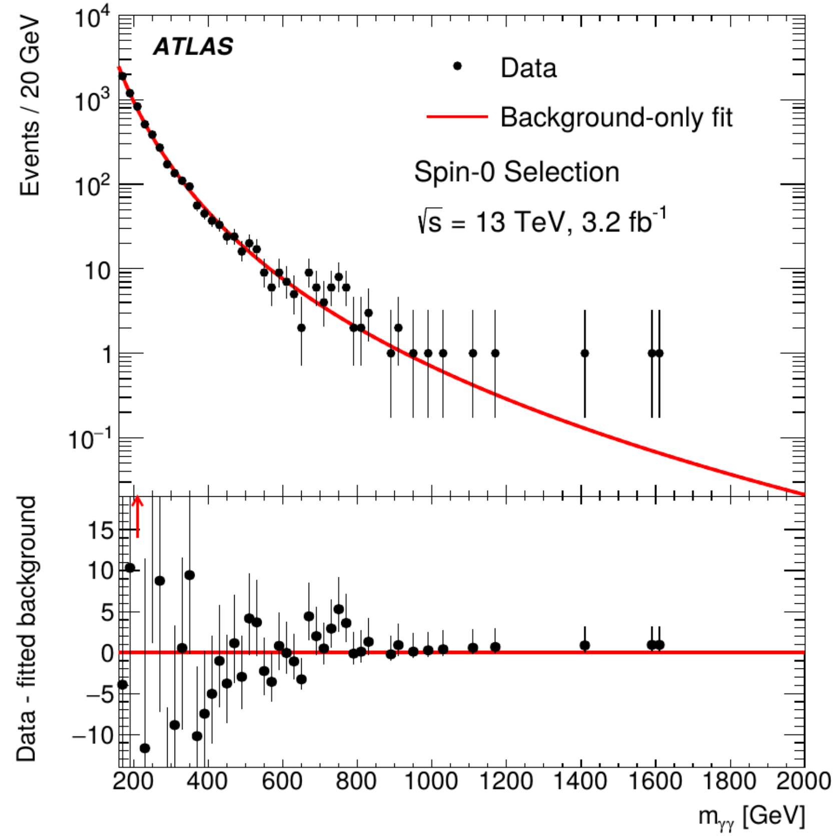


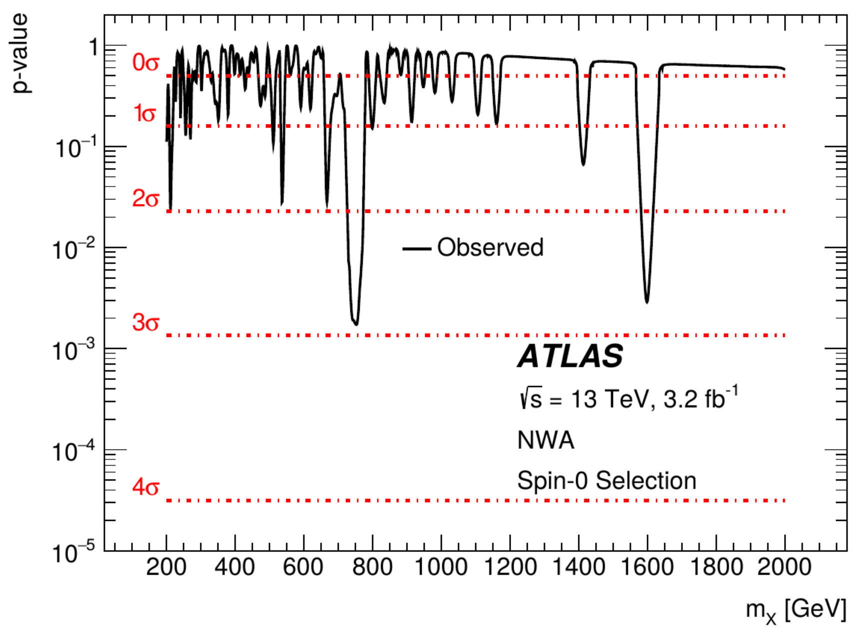
Look-Elsewhere effect

Sometimes, unknown parameters in signal model
e.g. p-values as a function of m_X

⇒ Effectively performing **multiple, simultaneous searches**

→ If e.g. small resolution and large scan range, **many independent experiments**





→ More likely to find an excess **anywhere in the range**, rather than in a **predefined** location

⇒ ***Look-elsewhere effect*** (LEE)

Probability for a fluctuation: ***anywhere*** in the range → **Global** significance

at a given location → **Local** significance

Global Significance

Global p-value → $$p_{global} = 1 - (1 - p_{local})^N \approx N\ p_{local}$$ ← **Local p-value**

N: **Trials factor**

→ $\mathbf{p_{global} > p_{local}}$ ⇒ global fluctuation more likely ⇒ less significant : $\mathbf{Z_{global} < Z_{local}}$

Trials factor : **naively** = # of independent intervals:
However this is usually **wrong**

$$N_{\text{trials}} \overset{??}{=} N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}$$

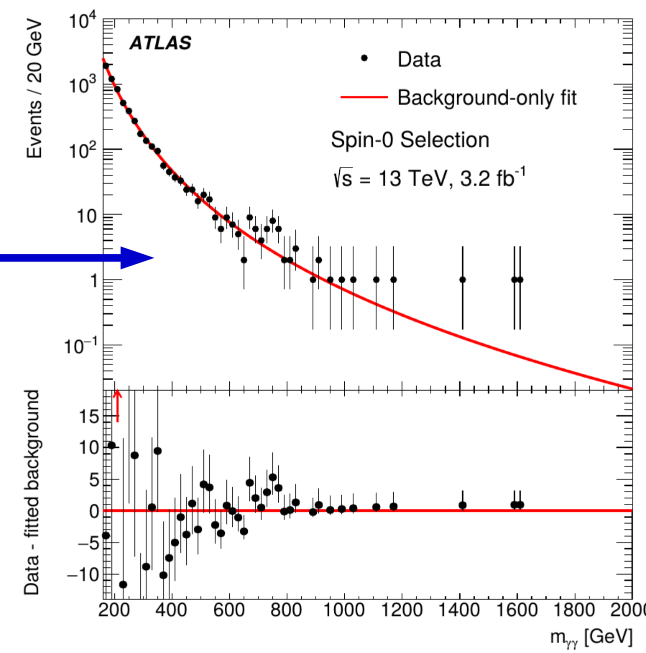
Actually, $$N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}}\, N_{\text{indep}}\, Z_{\text{local}}$$

Gross & Vitells EPJC 70:525-530,2010

(1 POI, asymptotic limit)

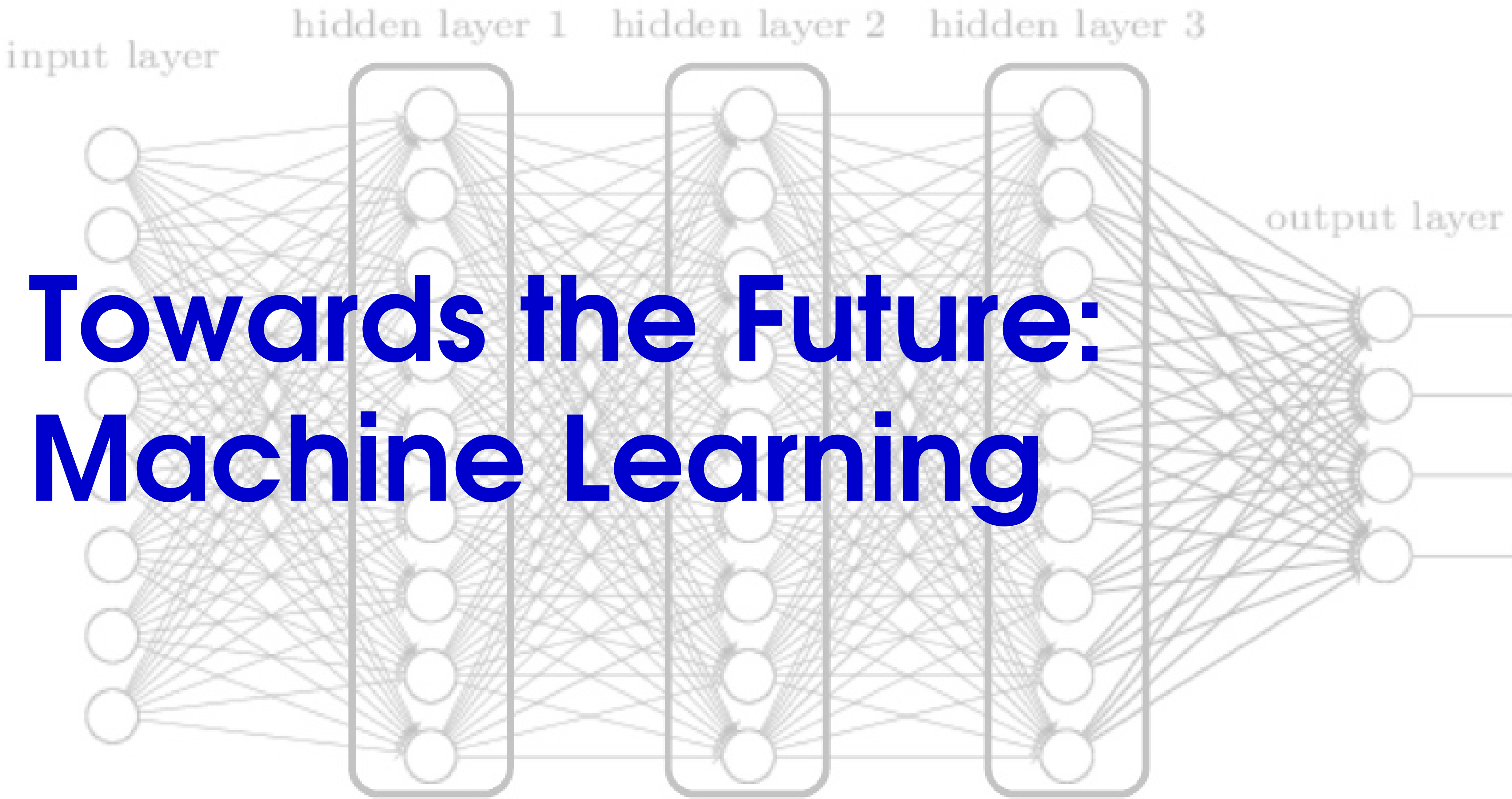
Can also use brute-force toys:

Z_{local} = 3.9σ !
(⇒ p_{local} ~ 5 10^{-5}),

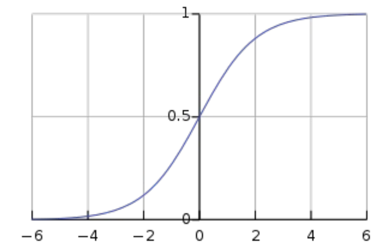


Generate toys ⇒ find such an excess 2% of the time
⇒ p_{global} ~ 2 10^{-2}, **Z_{global} = 2.1σ** Less exciting...

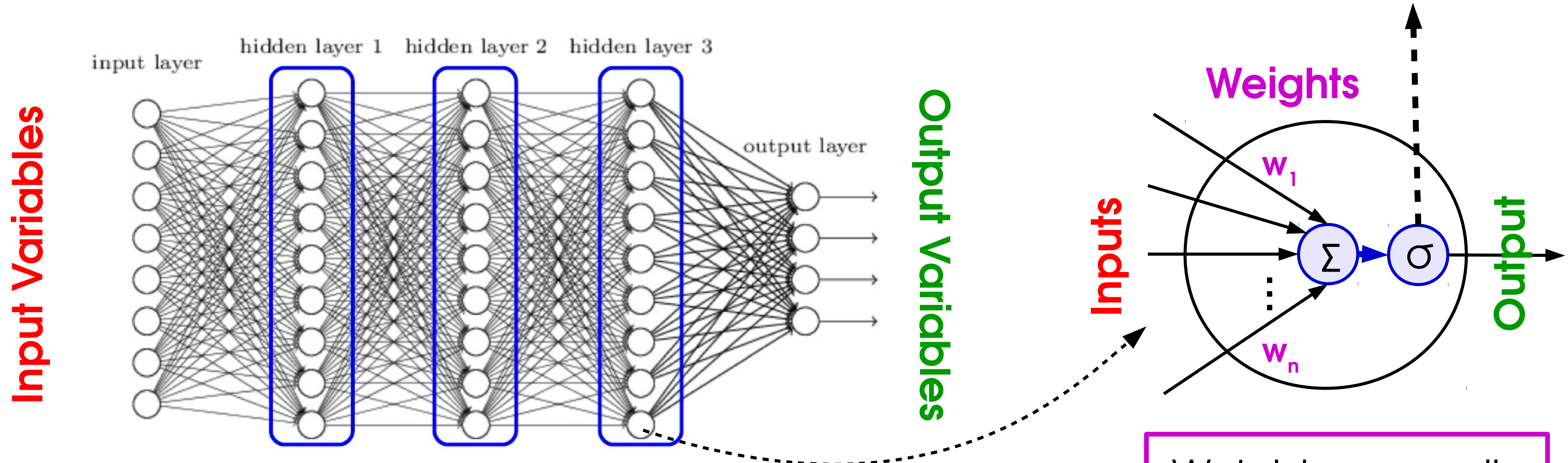
Towards the Future: Machine Learning



Machine Learning



Old idea, now reaching maturity in HEP applications.
Main example is **neural networks**:



Weights w_i usually from test data

Evolution towards **Deep networks**
→ several hidden layers
→ many neurons per layer

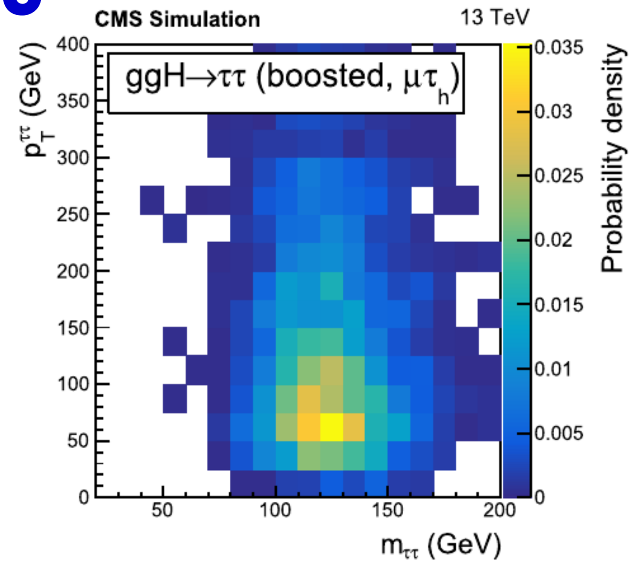
Made possible by

→ **Increased computing power** (e.g. GPUs)

→ **New methods** : Cross-entropy training (same as max. likelihood), dropout, non-sigmoid activation functions, etc.) to improve training performance

Machine Learning Discriminants

PLB 779 (2018) 283–316



Usual statistical methods work well for

→ **Event counting**

→ **1D distributions**

ML : Build discriminant,

⇒ use in 1D shape analyses

Already in common use (e.g. BDTs)

DNNs:

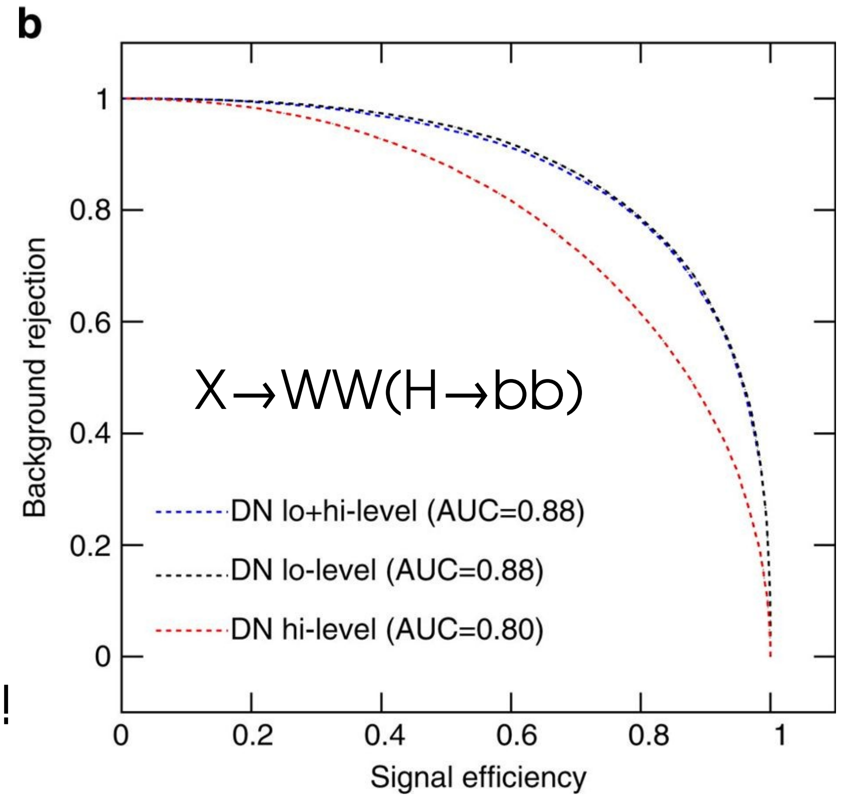
→ Better performance

→ Can work on low-level inputs (4-vectors)

⇒ No need for "hand-crafted" variables

⊖ Still can't do better than Likelihood ratio

⊕ Can provide arbitrarily good approximations!



P. Baldi, P. Sadowski & D. Whiteson, Nature Comm. vol. 5, 4308 (2014)

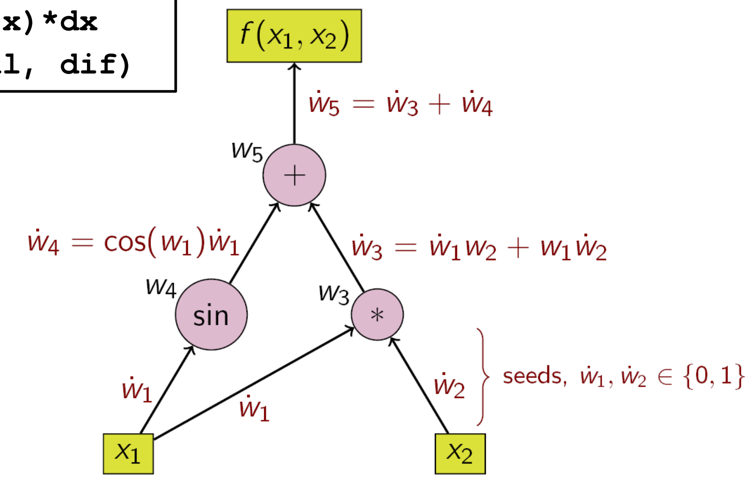
ML Computing Backends

ML computing-intensive ⇒ efficient implementations: e.g. TensorFlow, PyTorch

→ **Parallelization, use of GPU architecture**

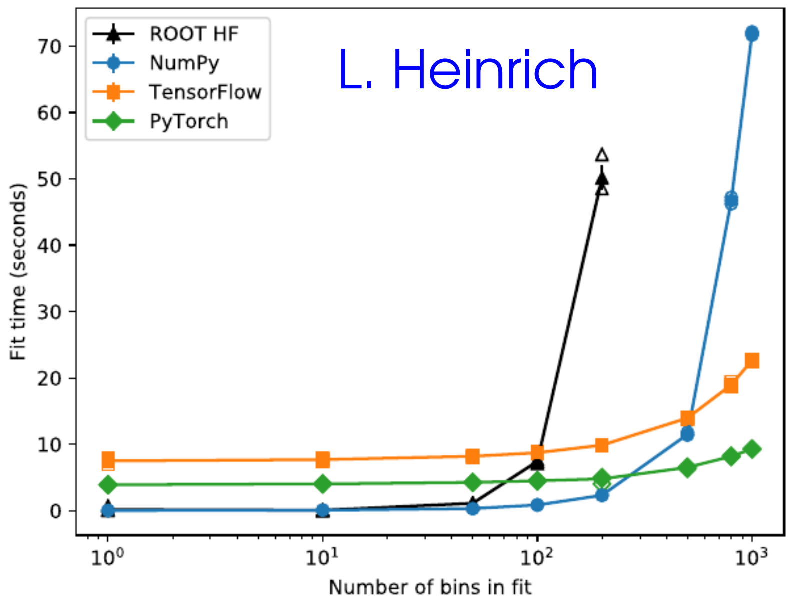
→ **New techniques**: e.g. **automatic gradient** computations

```
def f(x, dx):
  val = sin(x)
  dif = cos(x)*dx
  return (val, dif)
```



ATLAS: pyhf, reimplementation of ROOT-based HistFactory framework

CMS: TF implementation of combine code



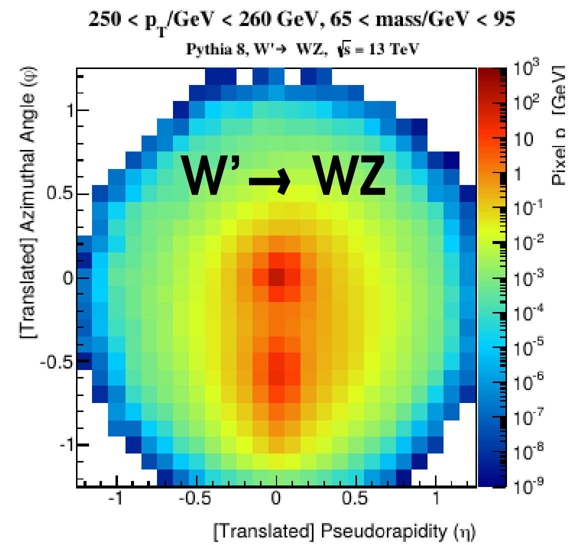
	Likelihood	Likelihood+Gradient	Hessian
Combine, TR1950X 1 Thread	10ms	830ms	-
TF, TR1950X 1 Thread	70ms	430ms	165s
TF, TR1950X 32 Thread	20ms	71ms	32s
TF, 2x Xeon Silver 4110 32 Thread	17ms	54ms	24s
TF, GTX1080	7ms	13ms	10s
TF, V100	4ms	7ms	8s

J. Bendavid

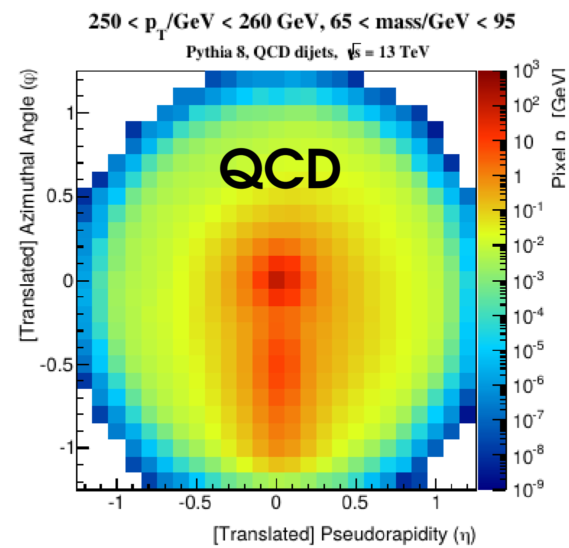
Other Applications

Many other applications:

Convolutive neural networks (CNNs)
→ "computer vision" : treat physics objects as images
⇒ Ideal for future high-granularity detectors



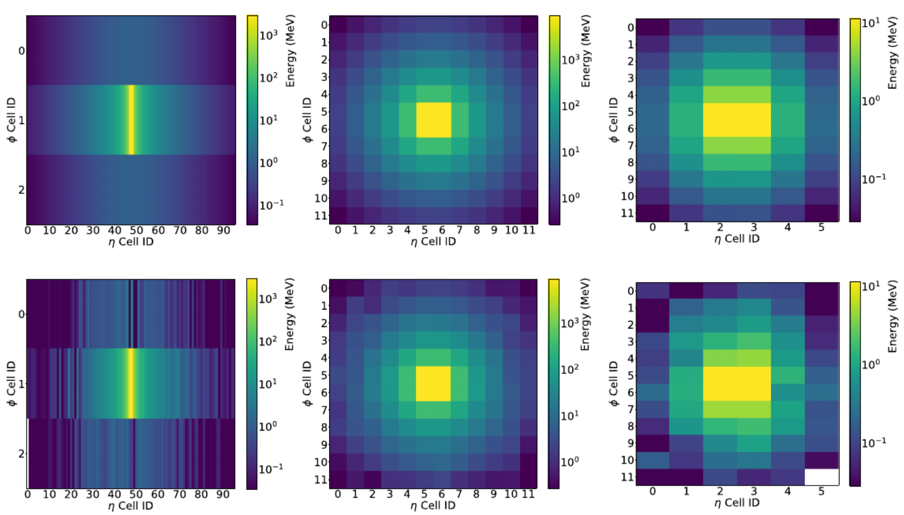
Recurrent NNs (RNNs)
→ language processing : treat collections of objects (tracks, cluster, cells) as sentences



JHEP 07 (2016) 069

Adversarial NNs

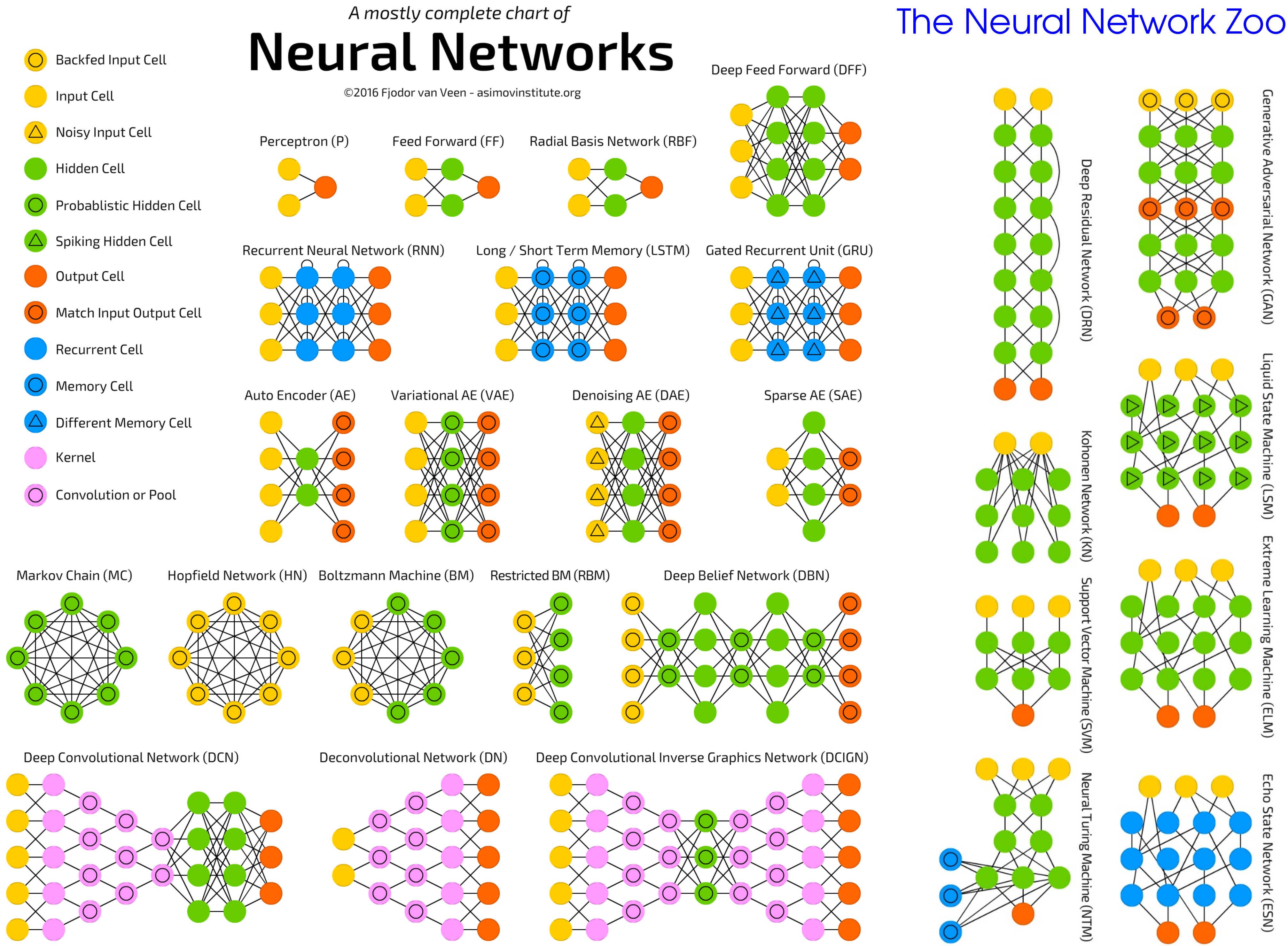
→ trained in pairs to optimize against systematics, or data/MC differences.



PRL. 120, 042003 (2018)

Factors 100-1000 gain in shower simulation time

The Neural Network Zoo



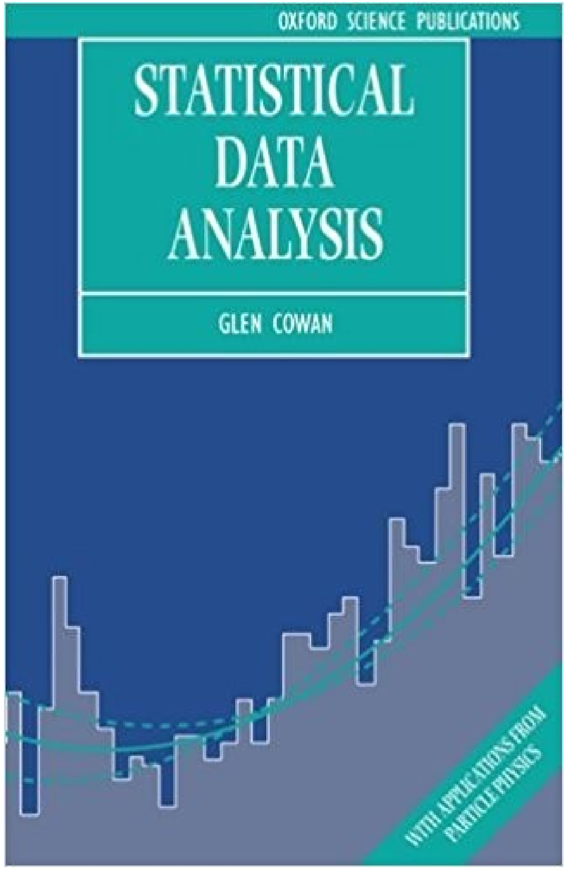
Conclusion

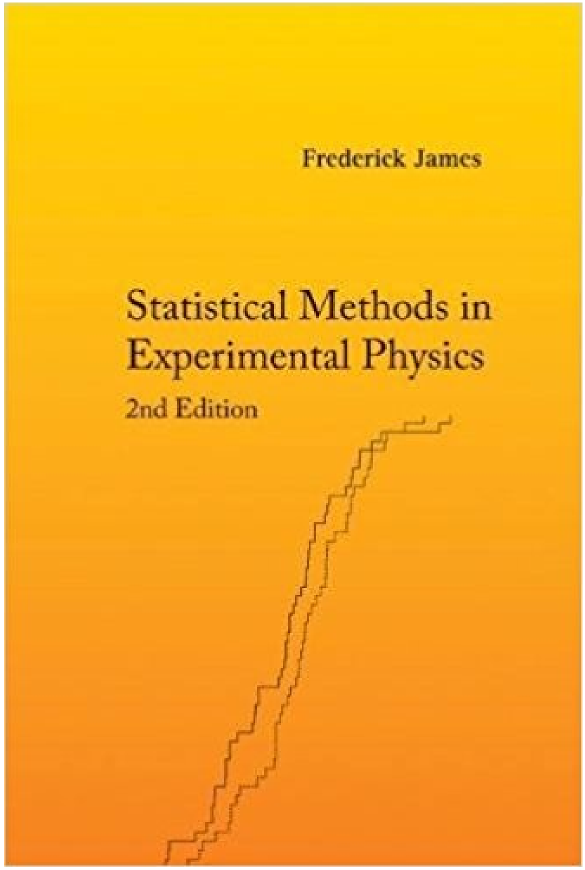
- **New developments in statistical methods in the last decade or so**
  → Baseline methods reaching maturity

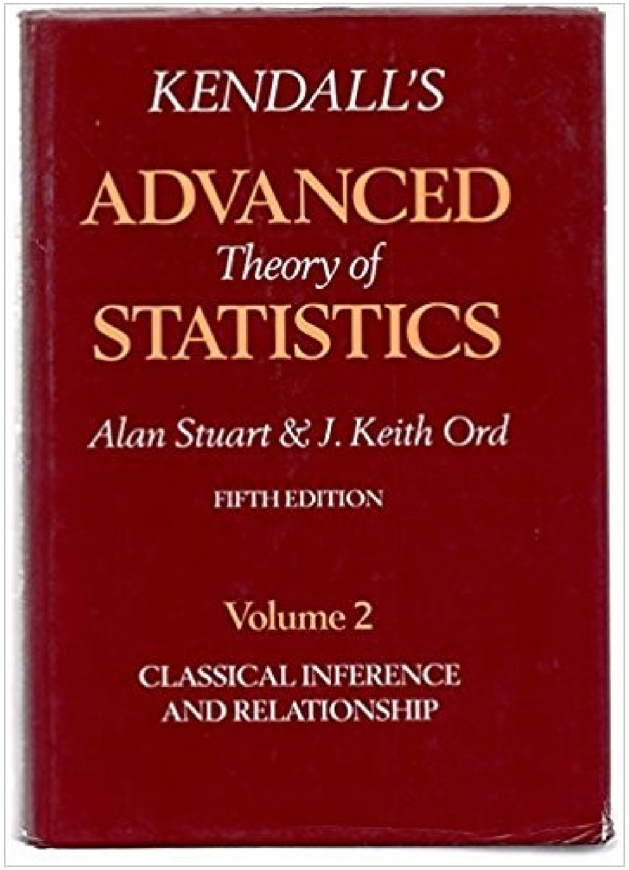
- **Many challenges still to be addressed**
  → Modeling complex experiments (systematics)
  → Pusblishing data to allow efficient re-interpretation
  → …

- **New horizons going towards machine learning**

Books and Courses







Some courses available online:

Glen Cowan's Cours d'Hiver and 2010 CERN Academic Training lectures

Kyle Cranmer's CERN Academic Training lectures

Louis Lyons'and Lorenzo Moneta's CERN Academic Training Lectures

Extra Material

BLUE and PLR

PLR Computation: 2 measurements
+ 1 auxiliary measurement

$$X_1 = X + \Delta_1\theta \sim G(X^*, \sigma_1)$$
$$X_2 = X + \Delta_2\theta \sim G(X^*, \sigma_2)$$
$$\theta \sim G(0,\ 1)$$

Single measurement: $\lambda(X,\theta) = \frac{1}{\sigma_1^2}(X + \Delta_1\theta - X_1^{obs})^2 + (\theta - \theta^{obs})^2$

MLEs:
$$\begin{cases} \hat{\theta} = \theta^{obs} \\ \hat{X} = X_1^{obs} - \Delta_1\theta^{obs} \end{cases}$$

PLR: $\lambda(X) = \frac{(X - \hat{X})^2}{\sigma_{1,tot}^2}$ $\qquad \sigma_{1,tot}^2 = \sigma_1^2 + \Delta_1^2$

Combination: $\lambda(X,\theta) = \frac{1}{\sigma_1^2}(X + \Delta_1\theta - X_1^{obs})^2 + \frac{1}{\sigma_2^2}(X + \Delta_2\theta - X_2^{obs})^2 + (\theta - \theta^{obs})^2$

MLE: $\hat{X} = \lambda_1 X_1^{obs} + \lambda_2 X_2^{obs} + \lambda_\theta \theta^{obs}$ $\qquad \lambda_{1(2)} = \frac{\sigma_{2(1),tot}^2 - \Delta_1\Delta_2}{\sigma_{1,tot}^2 + \sigma_{2,tot}^2 - 2\Delta_1\Delta_2}$

PLR: $\lambda(X) = \frac{(X - \hat{X})^2}{\sigma_{X,tot}^2}$ $\qquad \sigma_{X,tot}^2 = \frac{\sigma_{1,tot}^2\sigma_{2,tot}^2 - \Delta_1^2\Delta_2^2}{\sigma_{1,tot}^2 + \sigma_{2,tot}^2 - 2\Delta_1\Delta_2}$

BLUE and PLR

Eur. Phys. J. C, 74 (2014) 2717

BLUE computation: measurements X_1 and X_2 with uncorrelated statistical uncertainties σ_1 and σ_2, correlated systematics Δ_1 and Δ_2.

Single measurement: stat uncertainty $\boldsymbol{\sigma_1}$, systematic $\boldsymbol{\Delta_1}$

- Uncorrelated uncertainties
- Assume everything is Gaussian

⇒ Uncertainties add in quadrature:

$$\sigma^2_{1,\mathrm{tot}} = \sigma^2_1 + \Delta^2_1$$

Combination:

$$C = \begin{bmatrix} \sigma^2_{1,\mathrm{tot}} & \rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}} \\ \rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}} & \sigma^2_{2,\mathrm{tot}} \end{bmatrix} \qquad \rho = \frac{\Delta_1 \Delta_2}{\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}$$

BLUE weights

$$\hat{X} = \lambda_1 X_1^{\mathrm{obs}} + \lambda_2 X_2^{\mathrm{obs}}$$

$$\lambda_{1(2)} = \frac{\sigma^2_{2(1),\mathrm{tot}} - \rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}{\sigma^2_{1,\mathrm{tot}} + \sigma^2_{2,\mathrm{tot}} - 2\rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}$$

Propagate uncertainties from C:

$$\sigma^2_{X,\mathrm{tot}} = \frac{\sigma^2_{1,\mathrm{tot}}\sigma^2_{2,\mathrm{tot}}\left(1-\rho^2\right)}{\sigma^2_{1,\mathrm{tot}} + \sigma^2_{2,\mathrm{tot}} - 2\rho\,\sigma_{1,\mathrm{tot}}\sigma_{2,\mathrm{tot}}}$$

Beyond Asymptotics: Toys

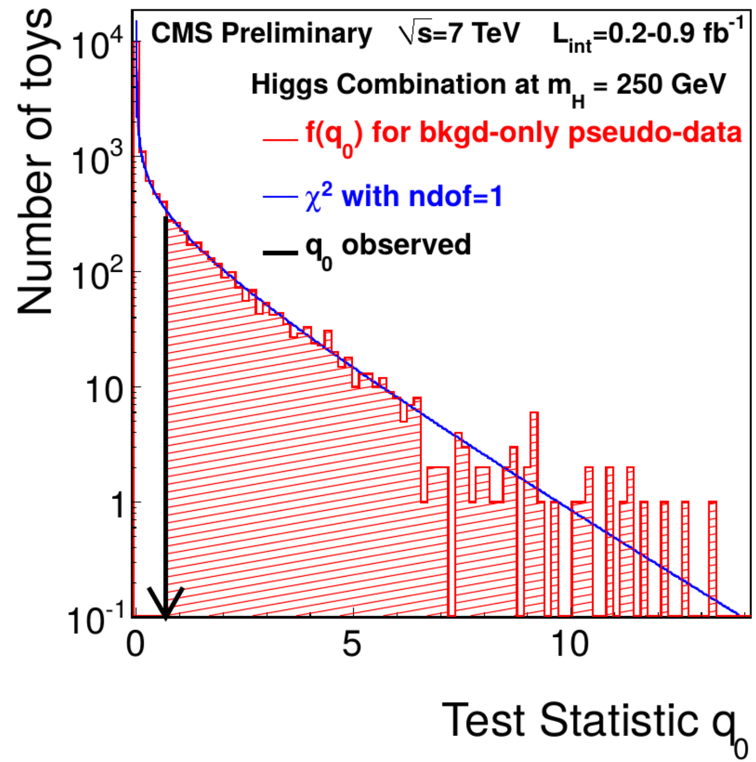
CMS-PAS-HIG-11-022

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

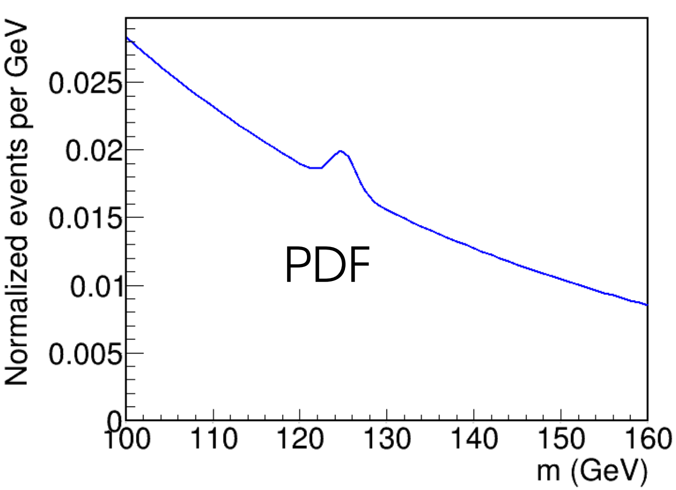


⇒ Integrate above observed PLR to get the p-value

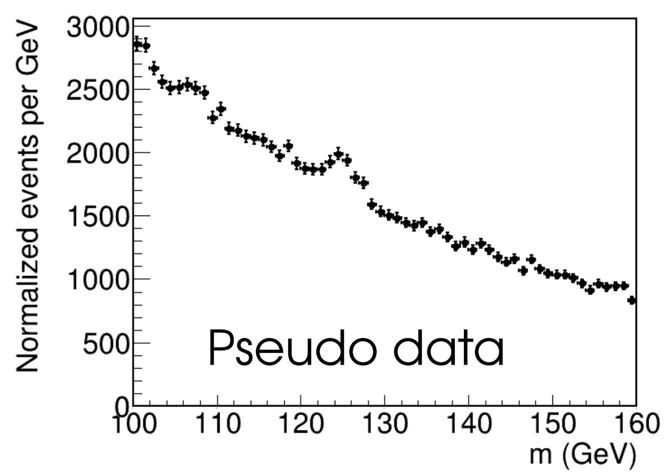
→ Precision limited by number of generated toys,

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

Repeat N_{toys} times



p(data | x) →



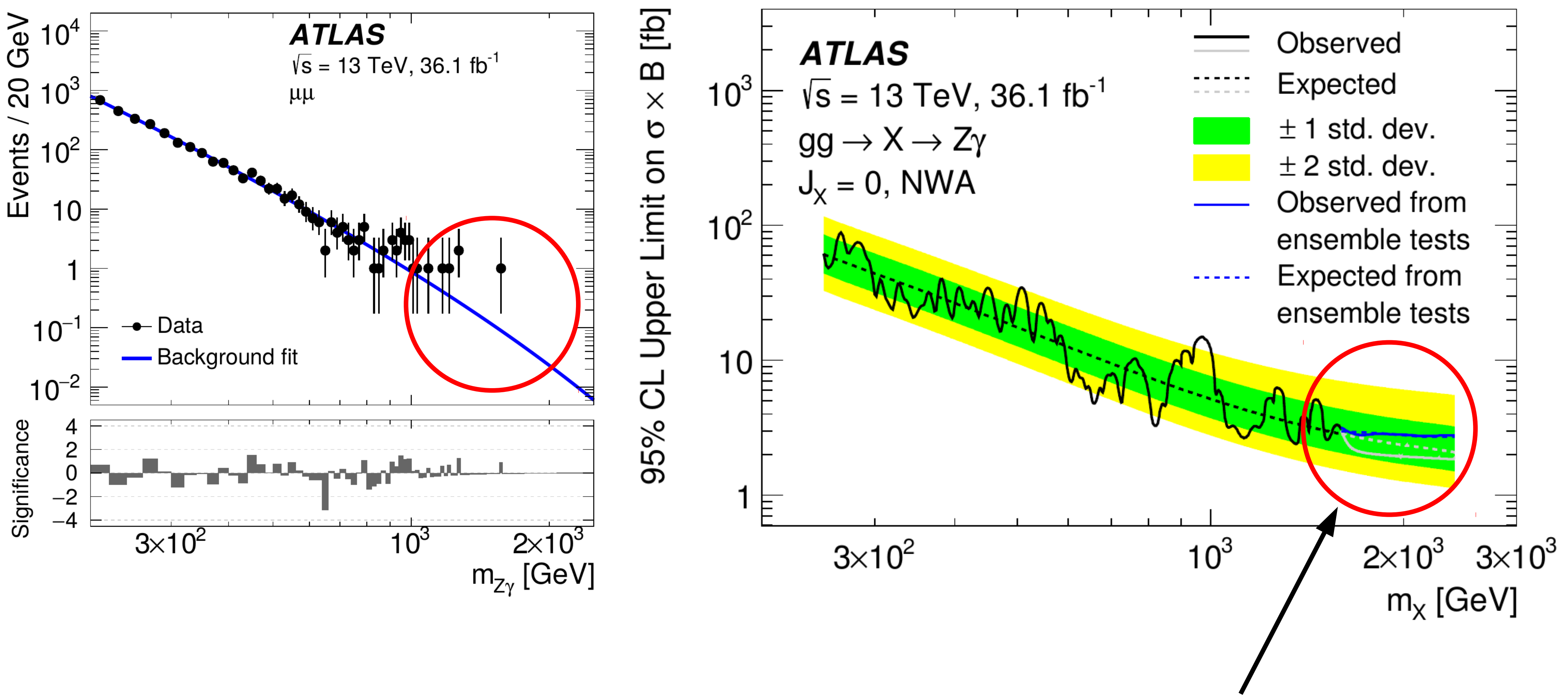
→ q_0

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**)

Rare Processes

HEP : almost always rare processes

ATLAS :

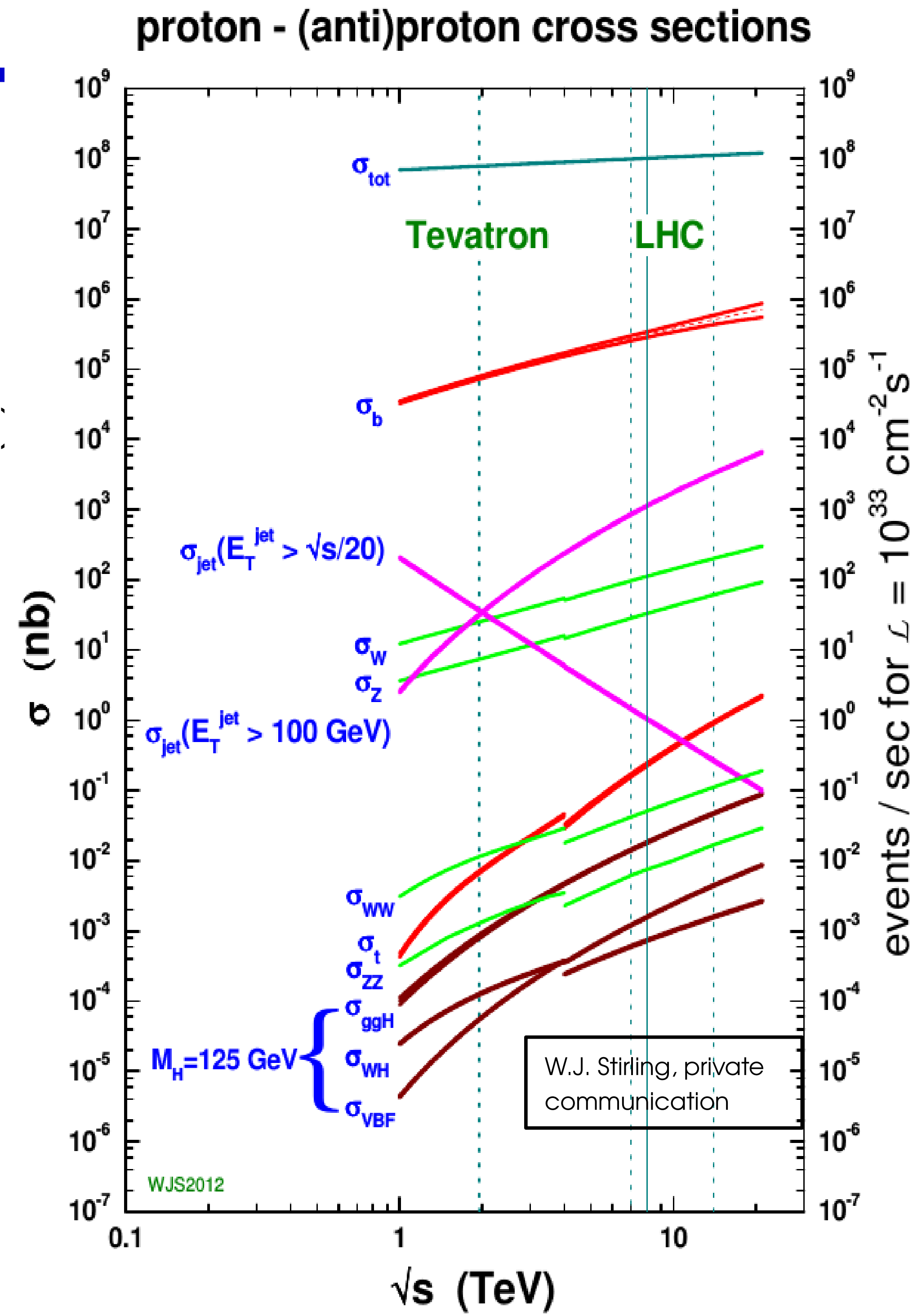
- **Event rate ~ 1 GHz**
  ($L \sim 10^{34}\ cm^{-2}s^{-1} \sim 10\ nb^{-1}/s$, $\sigma_{tot} \sim 10^8$ nb,)
- **Trigger rate ~ 1 kHz**
  (Higgs rate **~ 0.1 Hz**)

⇒ P ~ 10^{-6} ≪ 1 ($P_{H \to \gamma\gamma}$ ~ **10^{-13}**)

A day of data: **N ~ 10^{14} ≫ 1**

Large N, small P ⇒ **Poisson regime!**

(Large N = design requirement,
to get not-too-small λ=NP...)



Asymptotic Approximation: Wilks' Theorem

→ Assume **Gaussian regime for Ŝ** (e.g. large n_{evts})

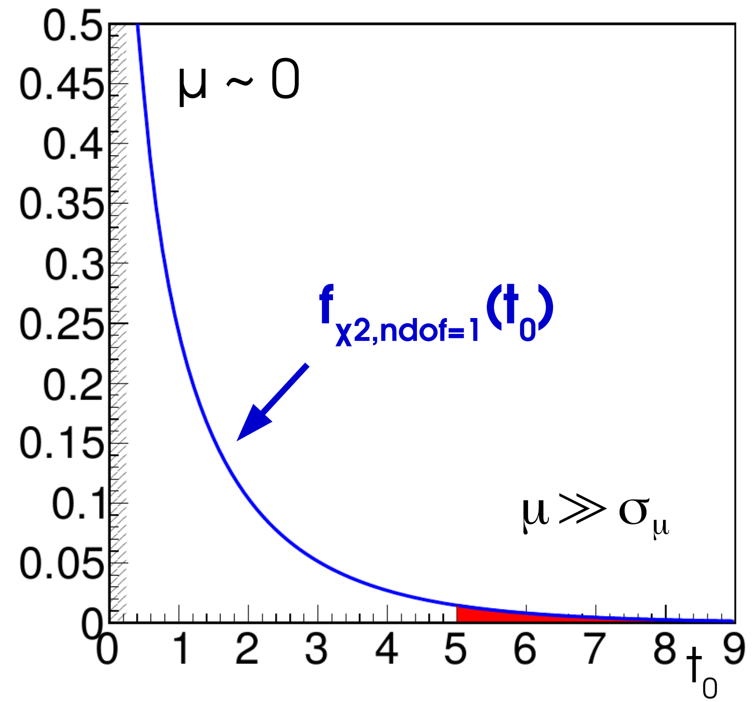
Cowan, Cranmer, Gross & Vitells
Eur.Phys.J.C71:1554,2011

⇒ Central-limit theorem :

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

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

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



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

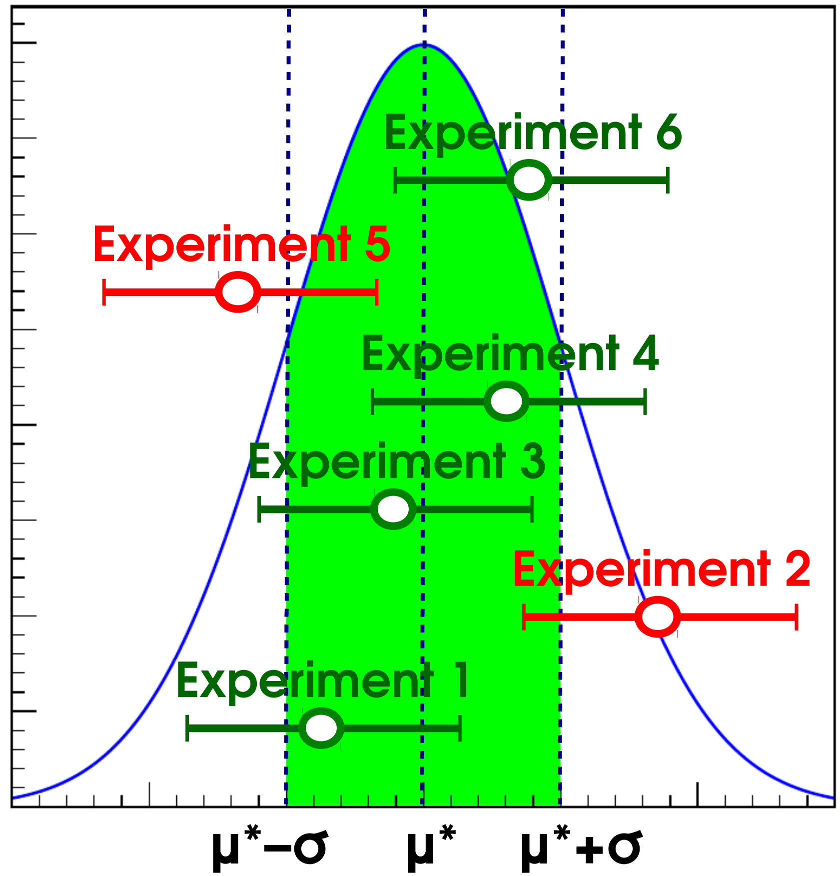
$$L(S) = \exp\left[-\frac{1}{2}\left(\frac{S-\hat{S}}{\sigma}\right)^2\right] \Rightarrow t_0 = \left(\frac{\hat{S}}{\sigma}\right)^2 \Rightarrow t_0 \sim \chi^2(n_{dof}=1) \text{ since } \hat{S} \sim G(0,\sigma)$$

Intervals

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,* ***[$\hat{\mu}$ - σ, $\hat{\mu}$ + σ]*** *will* ***contain the true value*** *68% of the time:* ***$\hat{\mu} = \mu^* \pm \sigma$***

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

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

Z	1	1.96	2
CL	0.68	0.95	0.955

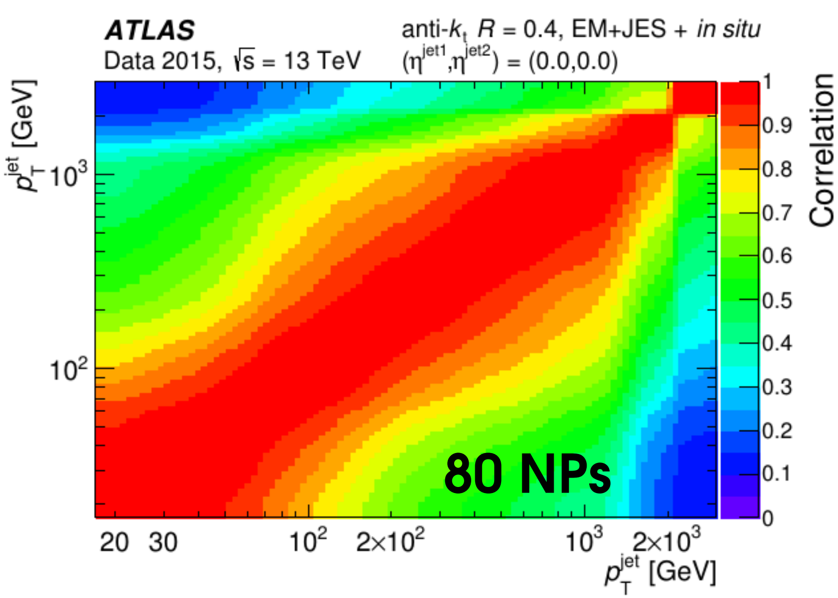
Systematics NPs

Each systematics NP represent **an independent source of uncertainty**
⇒ Usually constrained by a single 1-D PDF (Gaussian, etc.)

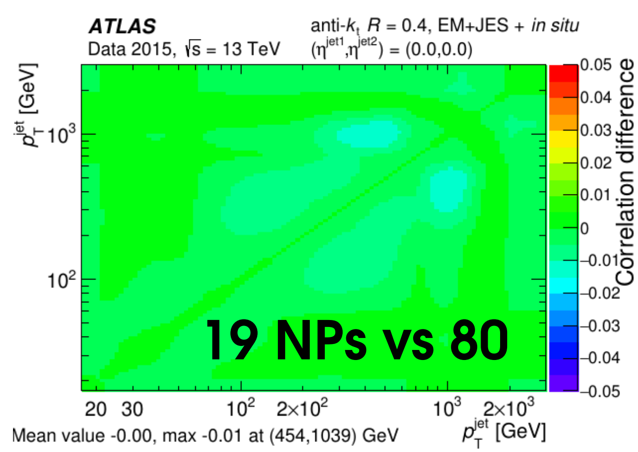
Sometimes multiple parameters **conjointly constrained** by an n-dim. PDF.
→ multiple measurements constraining multiple NPs

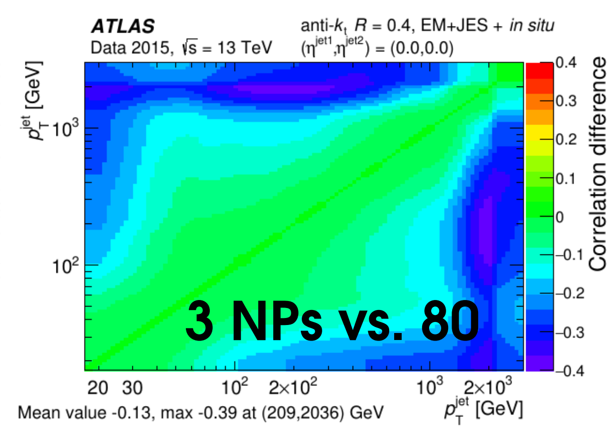
Assume **n-dim Gaussian** PDF: then can **diagonalize the covariance matrix C** and re-express the uncertainties in basis of eigenvector NPs ⇒ **n 1-dim PDFs**

Can also diagonalize to **prune** irrelevant uncertainties: keep NPs with large eigenvalues, combine in quadrature the others

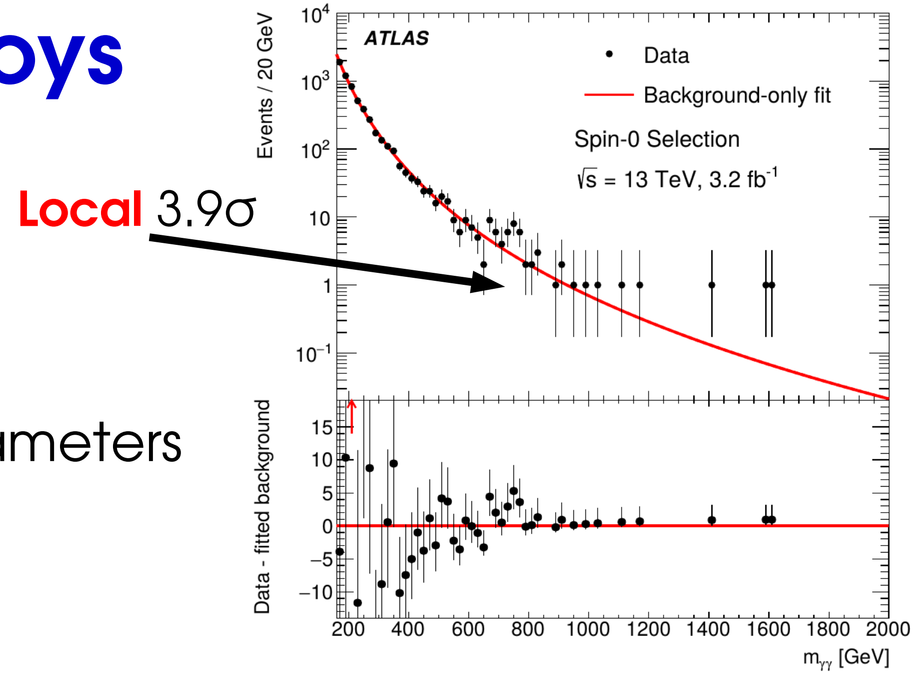


Phys.Rev. D96 (2017) no.7, 072002





Global Significance from Toys



Principle: repeat the analysis in toy data:

→ generate pseudo-dataset

→ perform the search, scanning over parameters as in the data

→ report the largest significance found

→ repeat many times

⇒ The frequency at which a given Z_0 is found ***is*** the global p-value

e.g. **X→γγ Search**: **Z_{local} = 3.9σ** (⇒ $p_{local} \sim 5\ 10^{-5}$),
scanning 200 < m_X< 2000 GeV and 0 < Γ_X < 10% m_X

→ In toys, find such an excess 2% of the time
⇒ $p_{global} \sim 2\ 10^{-2}$, **$Z_{global}$ = 2.1σ** Less exciting...

⊕ **Exact treatment**
⊖ **CPU-intensive** especially for large Z (need ~O(100)/p_{global} toys)

Global Significance from Asymptotics

Principle: approximate the global p-value in the asymptotic limit
→ reference paper: **Gross & Vitells, EPJ.C70:525-530,2010**

$$N_{indep} = \frac{\text{scan range}}{\text{peak width}}$$

Asymptotic trials factor (1 POI):

$$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

→ Trials factor is **not just N_{indep},** also depends on Z_{local} !

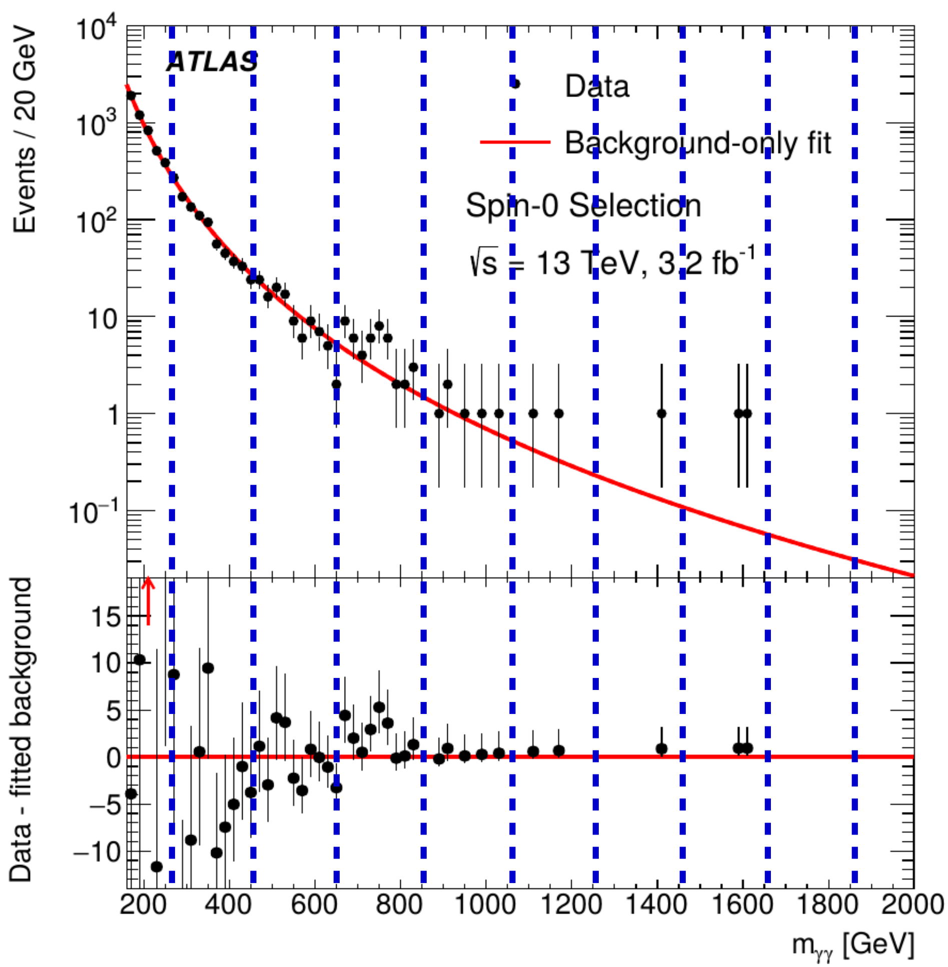
Why ?

→ slice scan range into N_{indep} regions of size ~ peak width
→ search for a peak in each region

⇒ Indeed gives $N_{trials}=N_{indep}$.

However this misses peaks sitting on **edges between regions**

⇒ true N_{trials} is **> N_{indep}**!



Global Significance from Asymptotics

Principle: approximate the global p-value in the asymptotic limit
→ reference paper: **Gross & Vitells, EPJ.C70:525-530,2010**

$$N_{indep} = \frac{\text{scan range}}{\text{peak width}}$$

Asymptotic trials factor (1 POI):

$$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

→ Trials factor is **not just N_{indep},** also depends on Z_{local} !

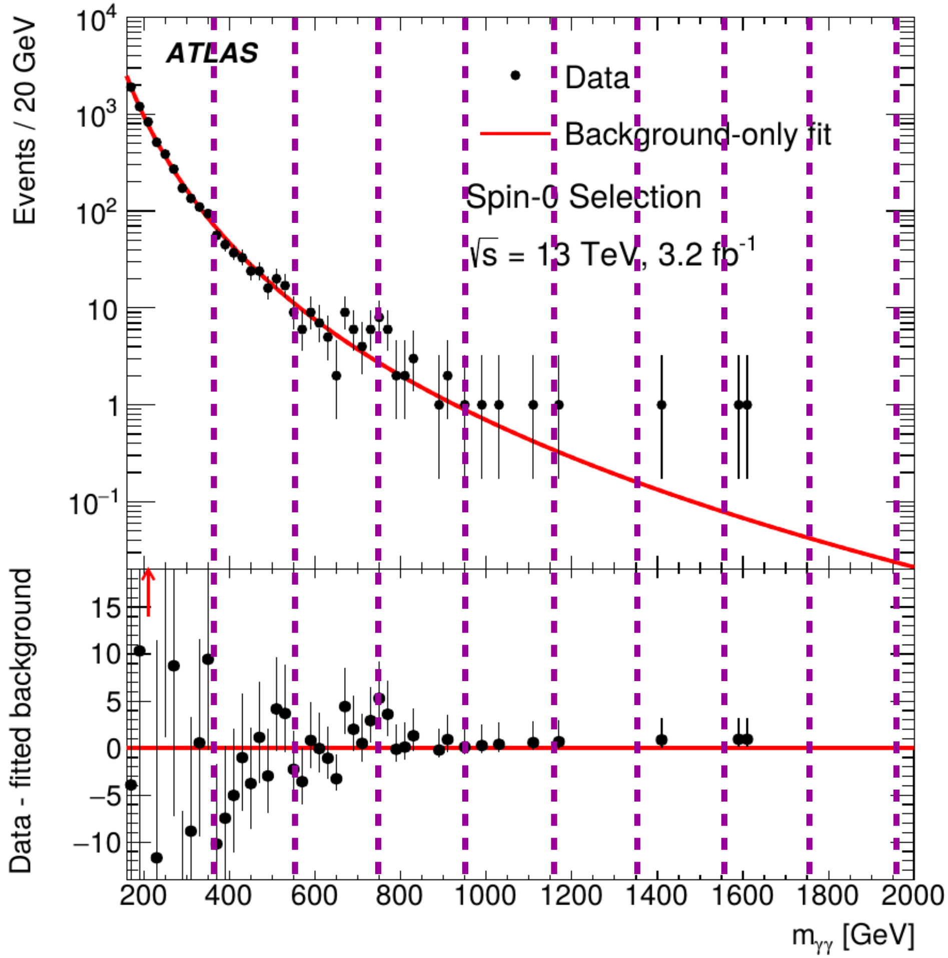
Why ?

→ slice scan range into N_{indep} regions of size ~ peak width
→ search for a peak in each region

⇒ Indeed gives $N_{trials}=N_{indep}$.

However this misses peaks sitting on **edges between regions**

⇒ true N_{trials} is **> N_{indep}**!



Illustrative Example

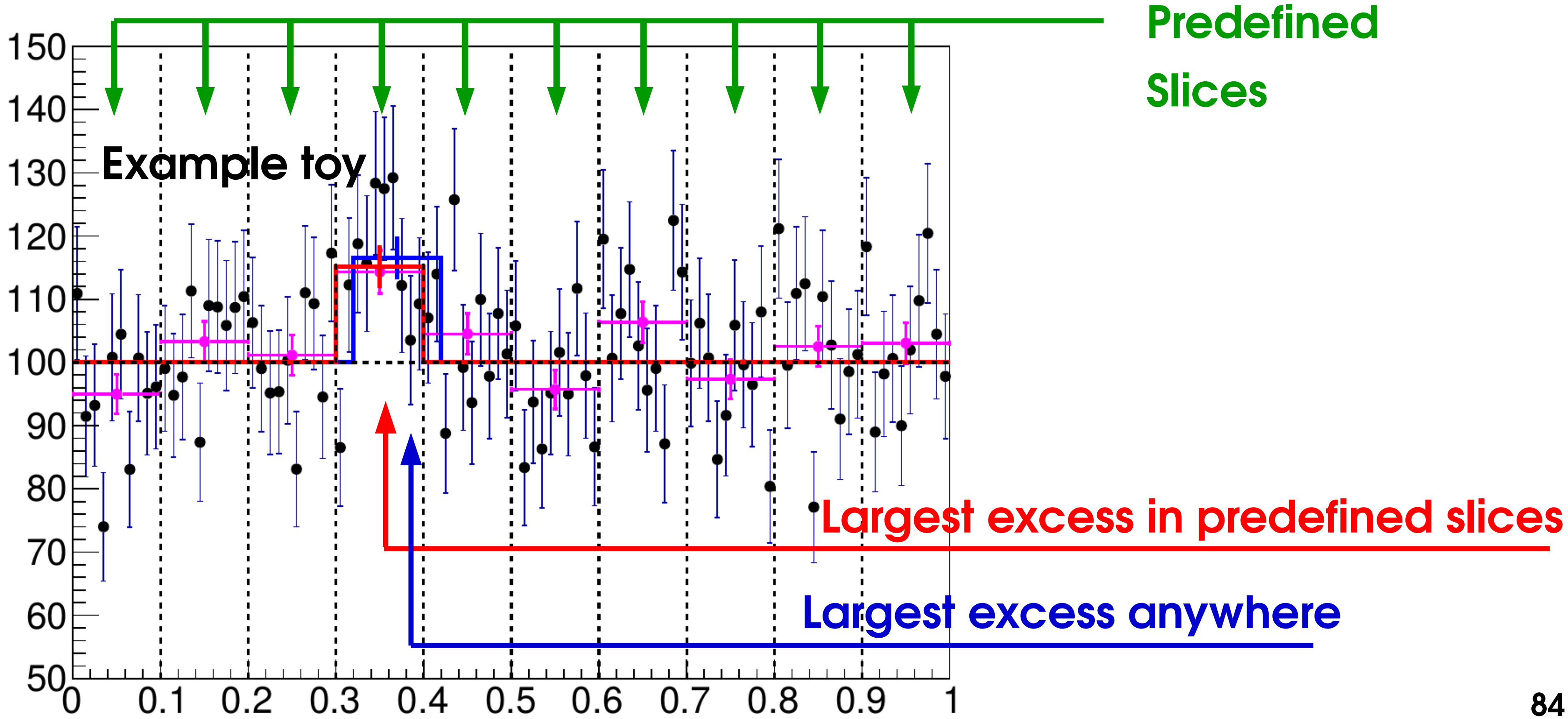
Test on a simple example: generate toys with

→ flat background (100 events/bin)

→ count events in a fixed-size sliding window, look for excesses

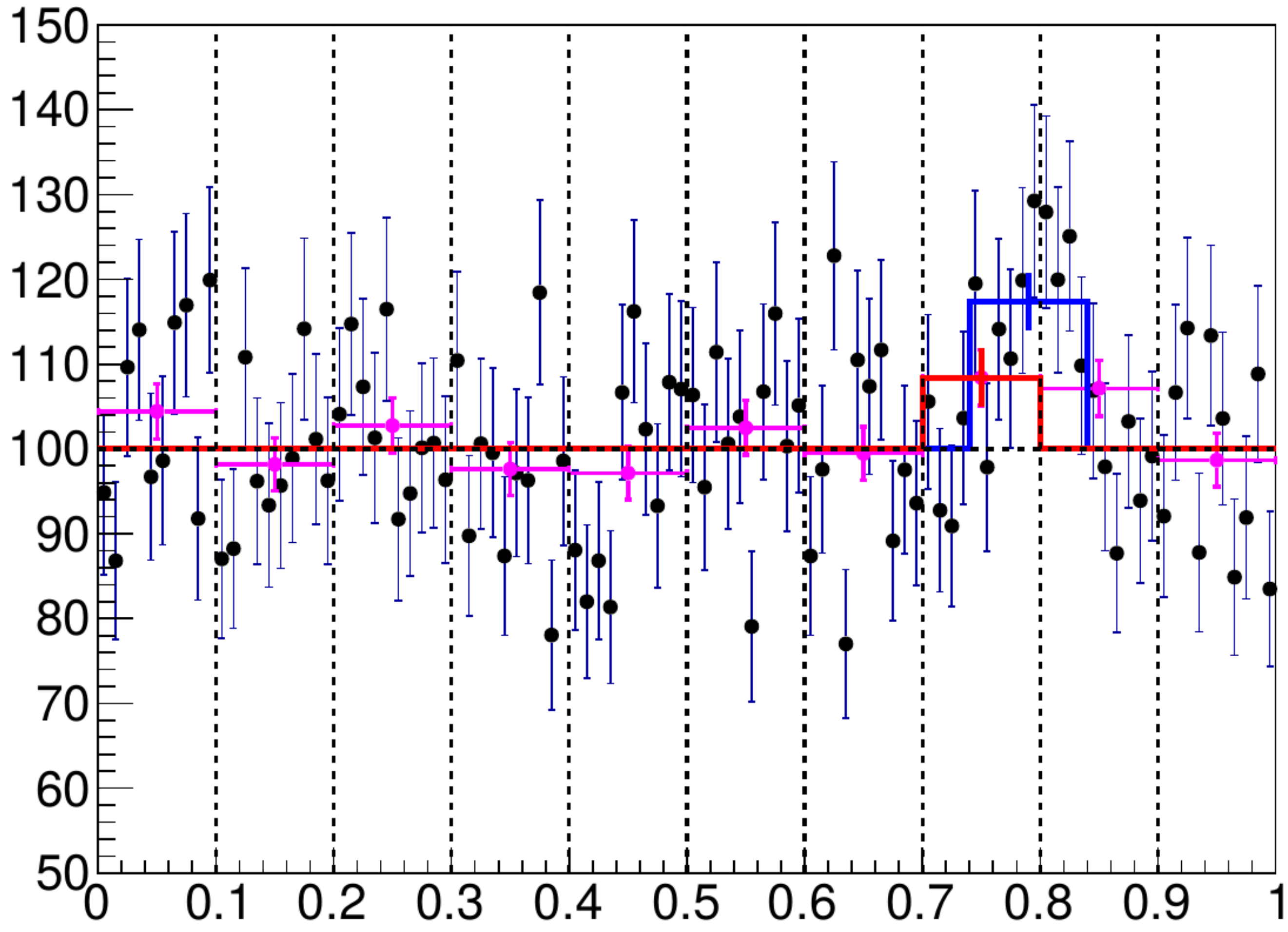
Two configurations:

1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum



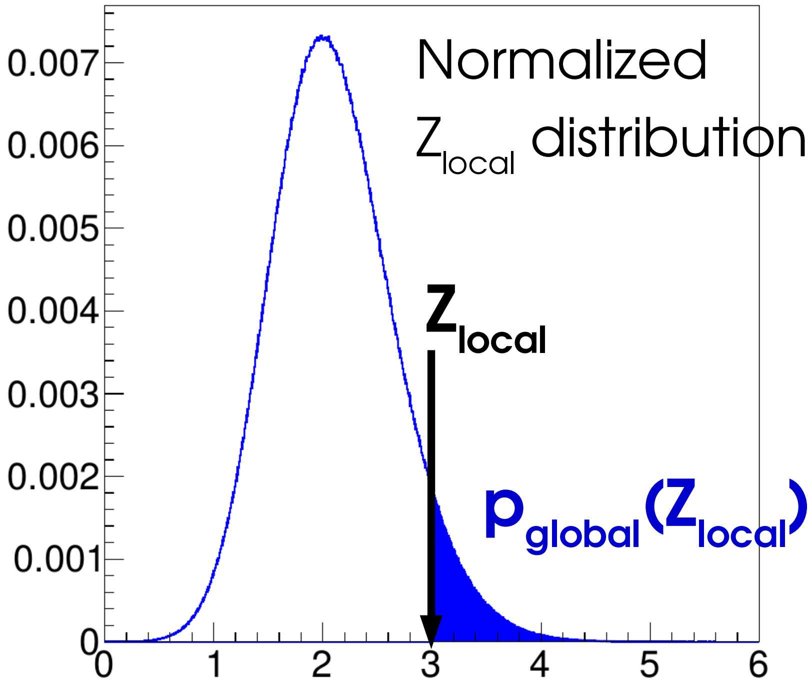
Illustrative Example (2)

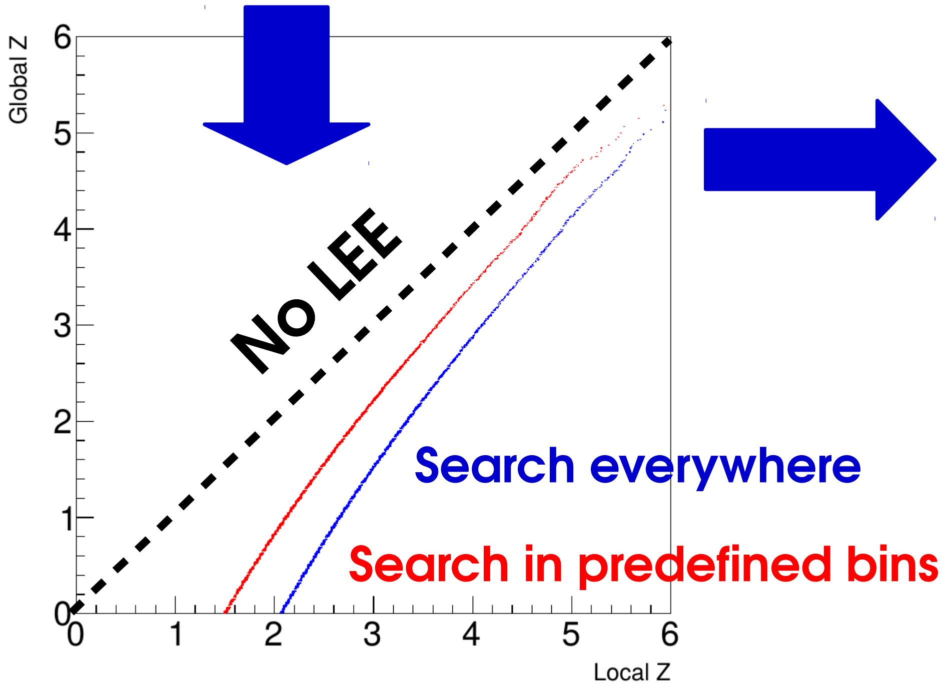
Very different results if the excess is **near a boundary :**

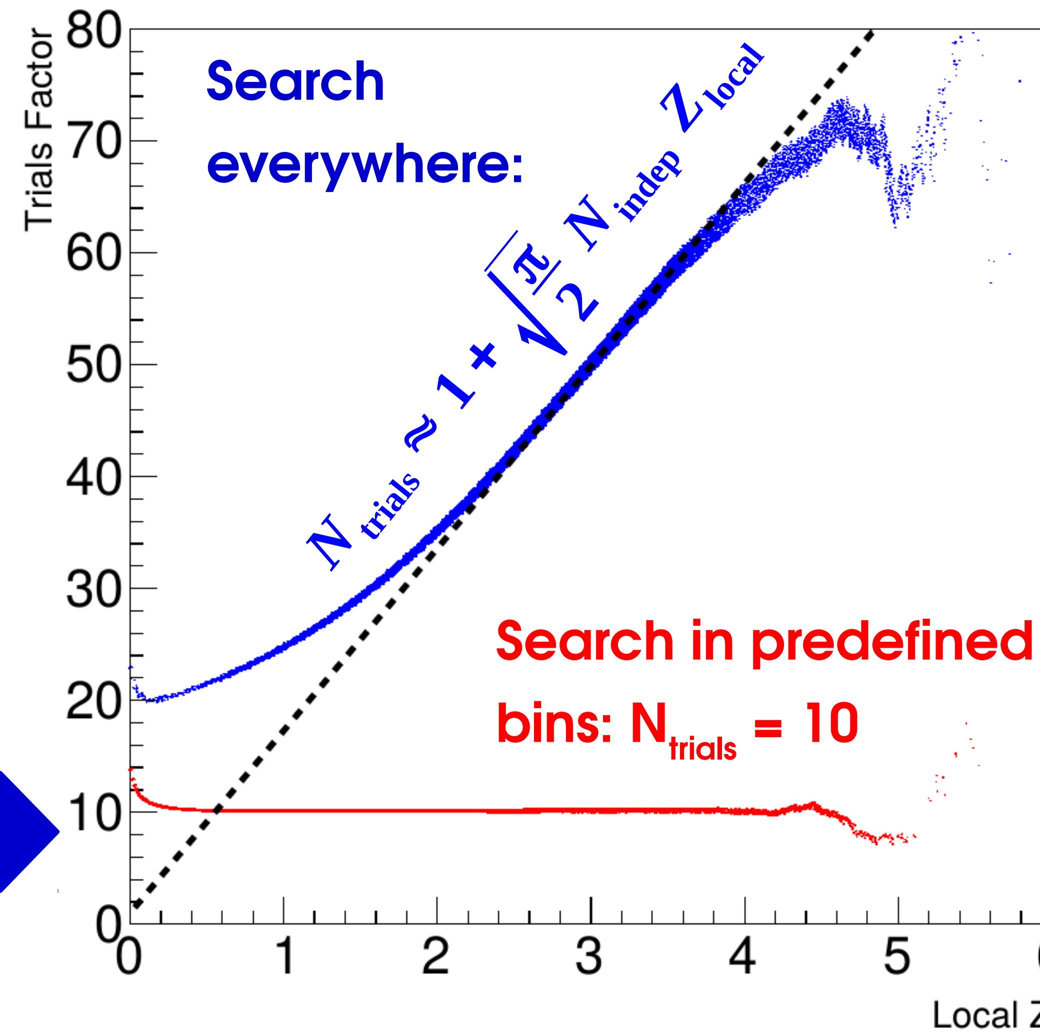


1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum

Illustrative Example (3)

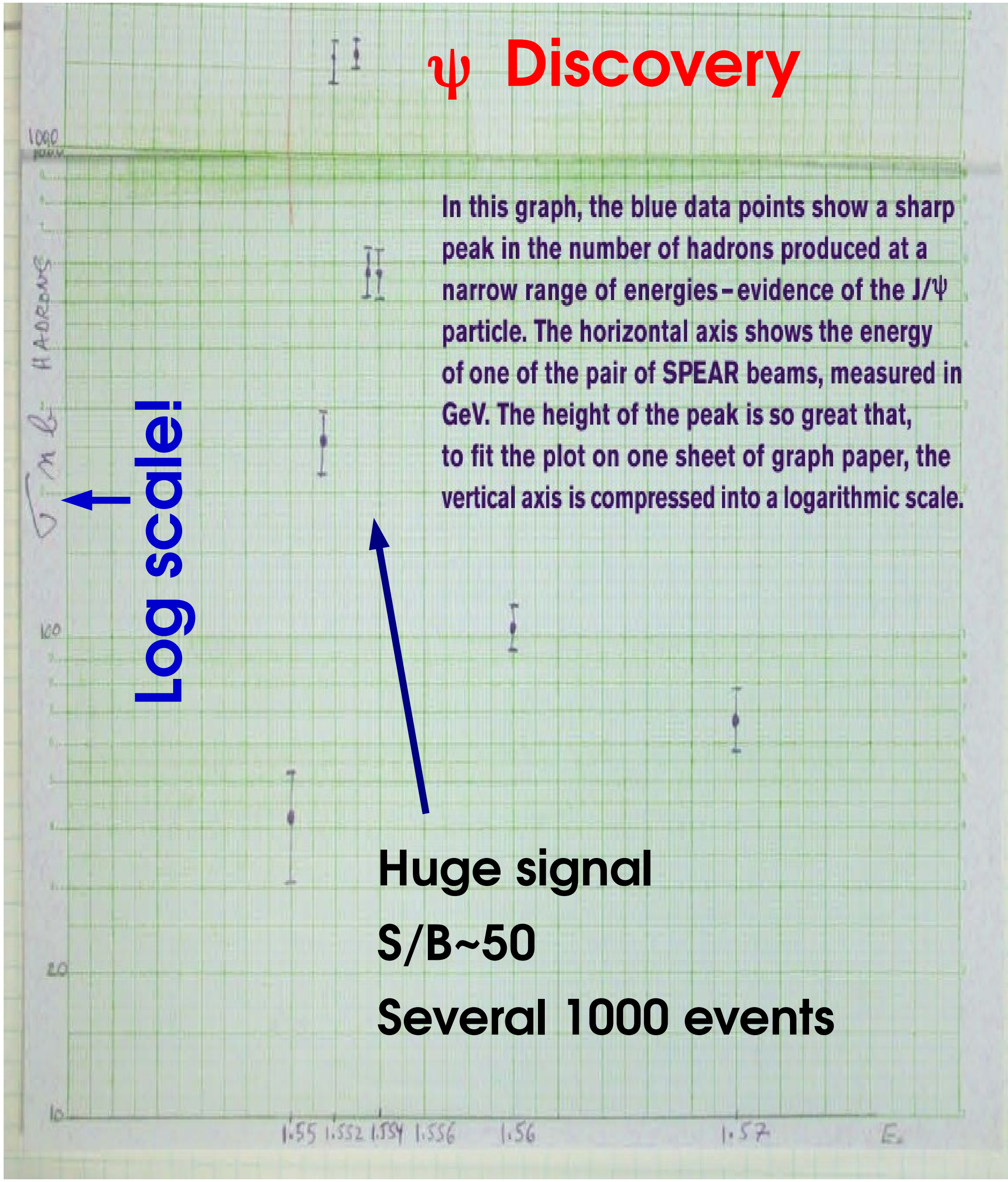




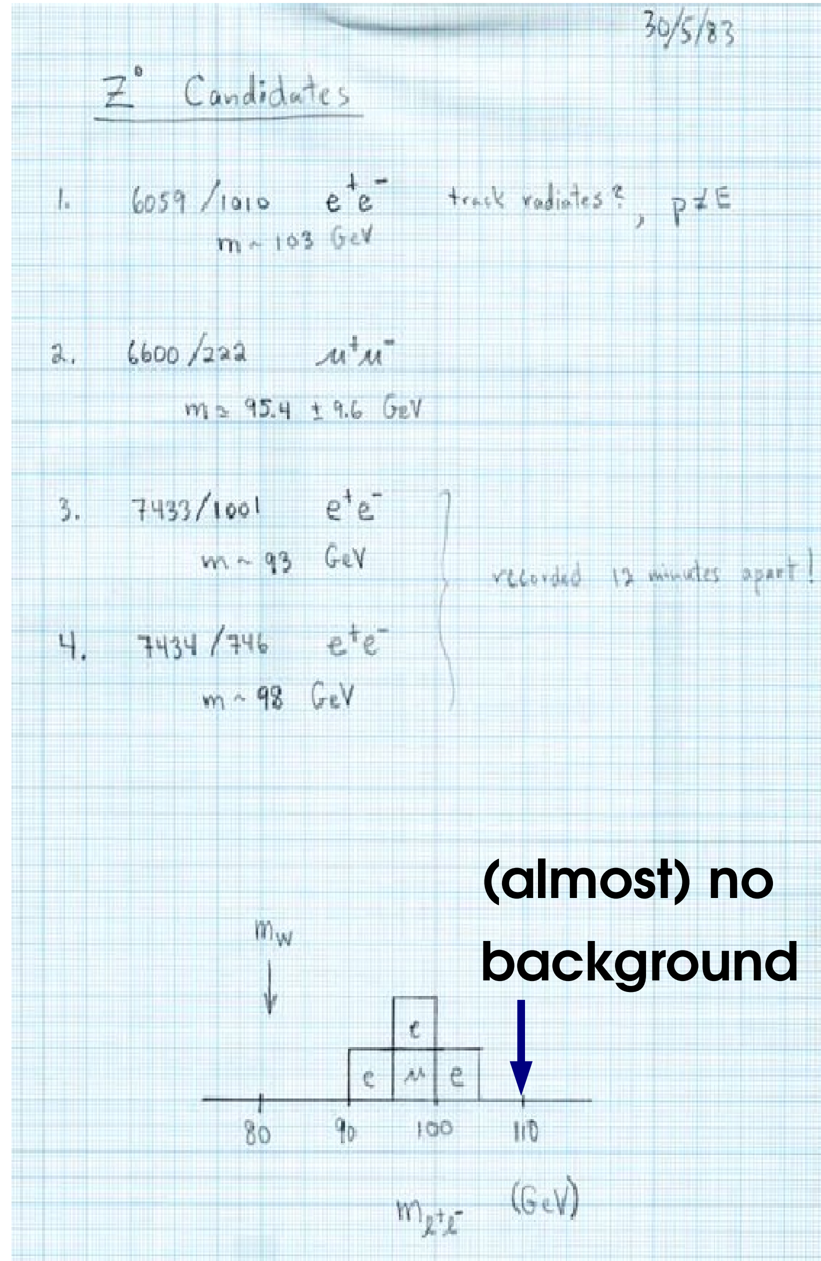


Searching everywhere gives the extra Z_{local} dependence

Classic Discoveries (1)

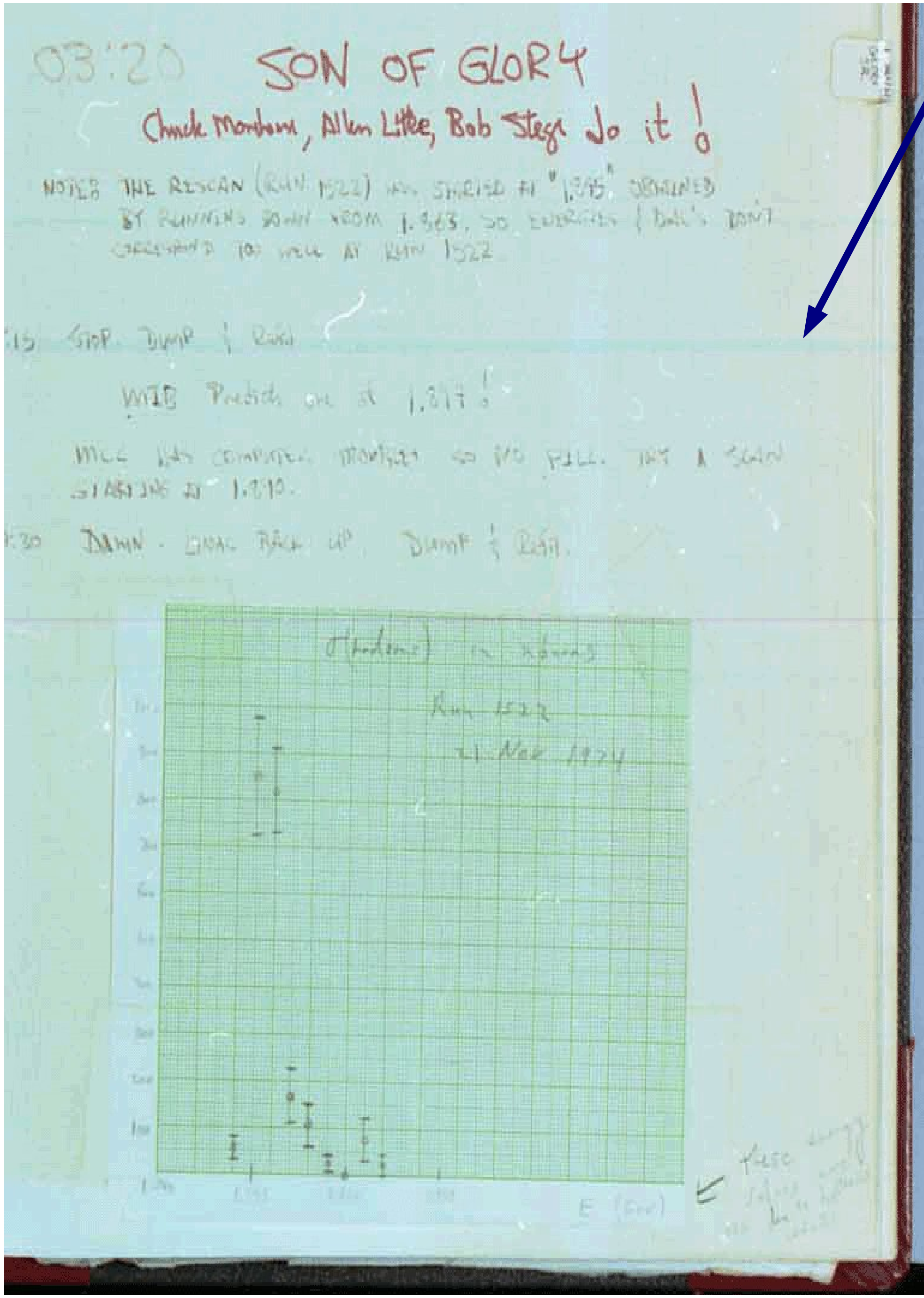


Z⁰ Discovery

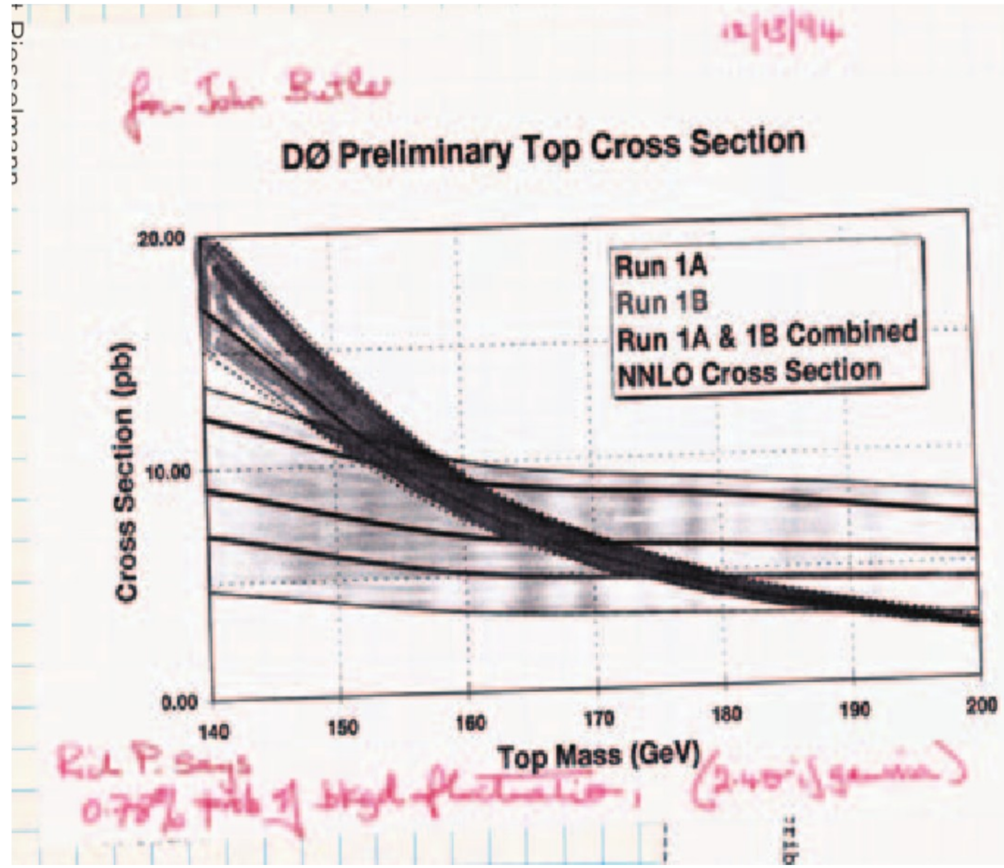


Logbook of J. Rohlf, 1983-05-30

Classic Discoveries (2)



ψ' : discovered online by the (lucky) shifters



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

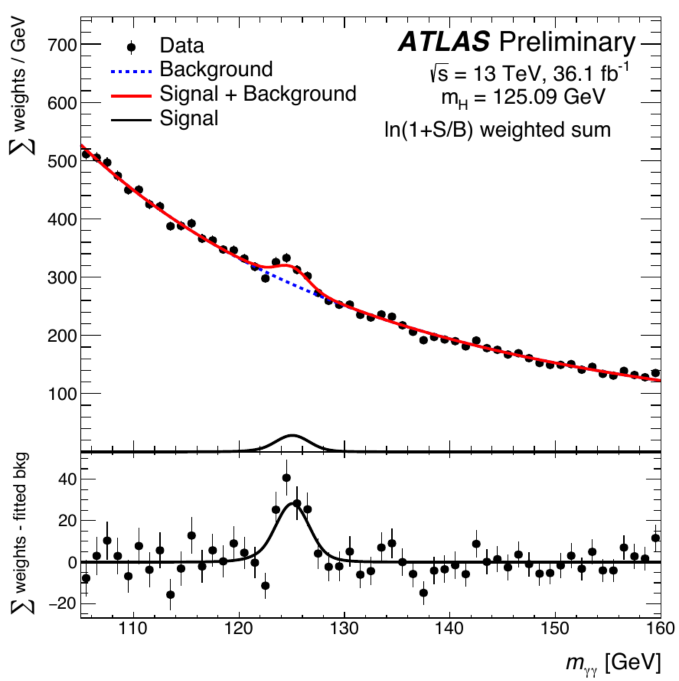
And now ?

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

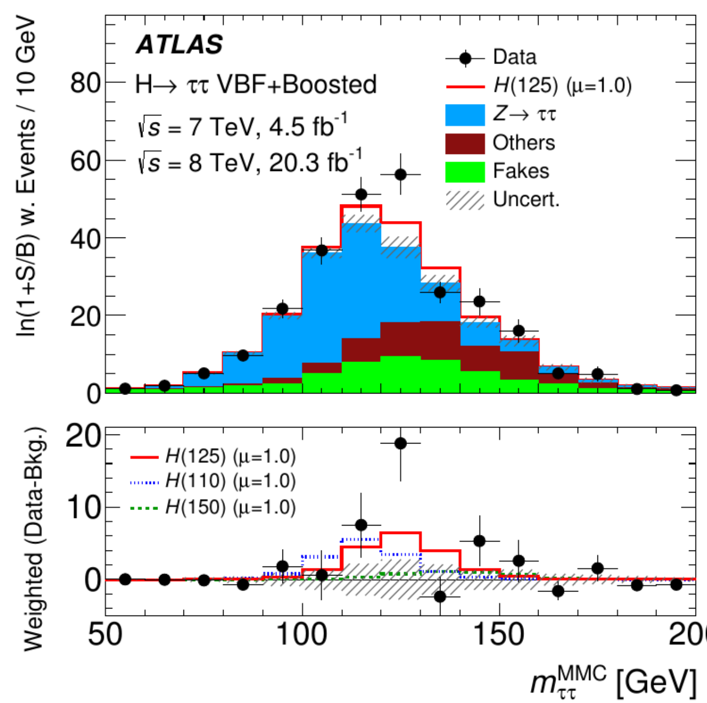
e.g. at LHC:

- **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.

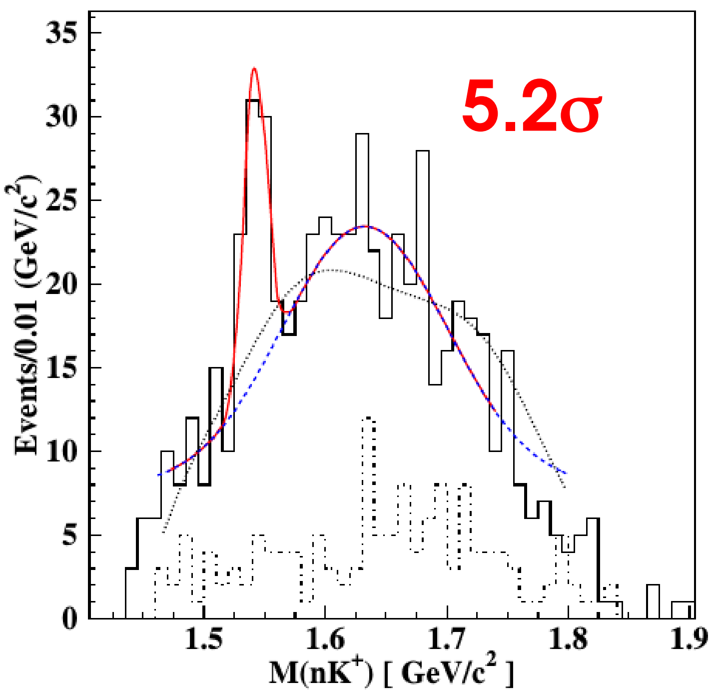
(ii) Any invisible Z^0, such as $Z^0 \to \nu\bar{\nu}$ decay, which is expected to have a large (18%) branching ratio. Note that the corresponding decays into charged lepton pairs $Z^0 \to e^+e^-$, $Z^0 \to \mu^+\mu^-$ have lower branching ratios (~3%) and may not have yet been produced within the present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

CaloGAN

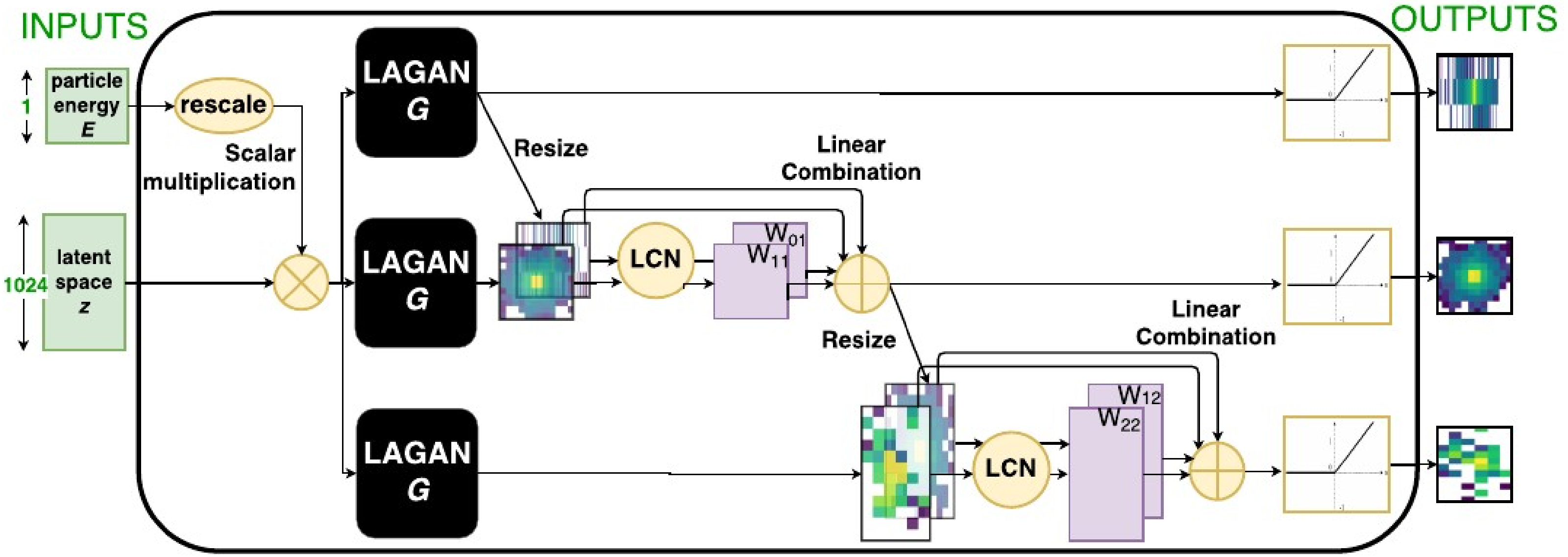


FIG. 4: Composite Generator, illustrating three stream with attentional layer-to-layer dependence.

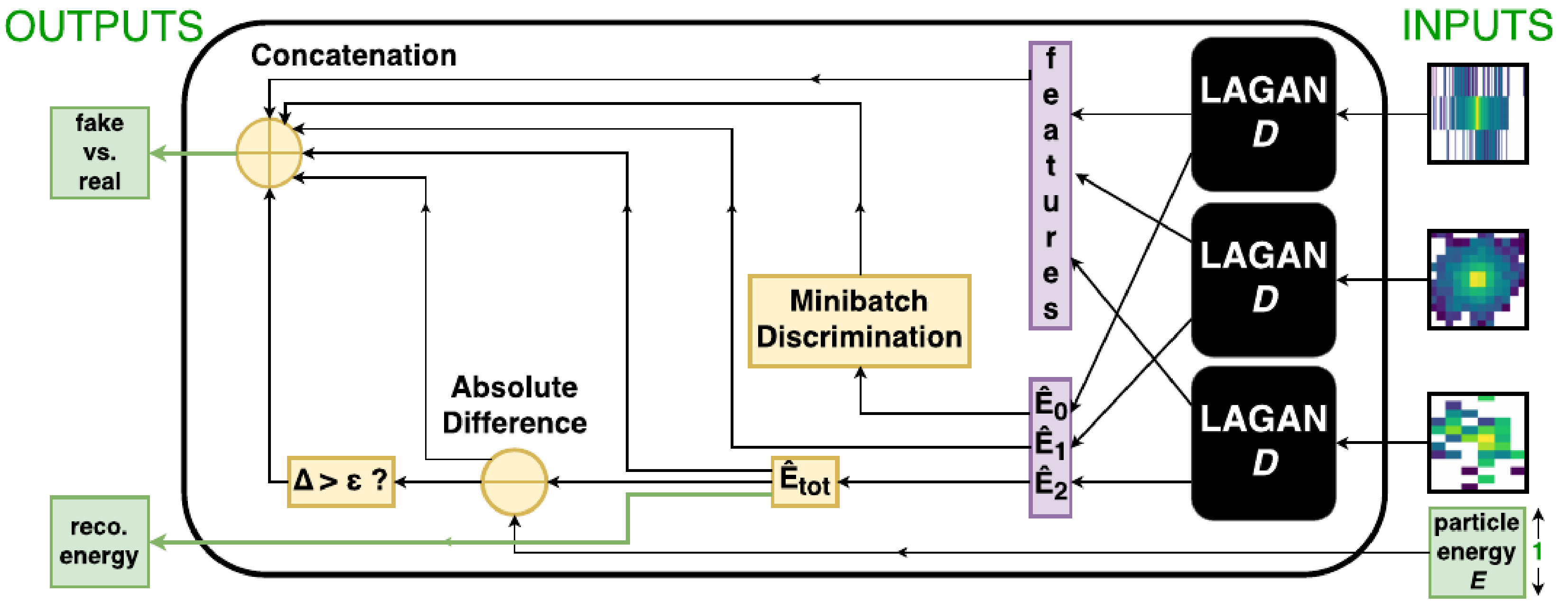


FIG. 5: Composite Discriminator, depicting additional domain specific expressions included in the final feature space.

M. Paganini et al., 1705.02355

Generation Method	Hardware	Batch Size	milliseconds/shower
GEANT4	CPU	N/A	1772 ←
CALOGAN	CPU	1	13.1
		10	5.11
		128	2.19
		1024	2.03
	GPU	1	14.5
		4	3.68
		128	0.021
		512	0.014
		1024	0.012 ←