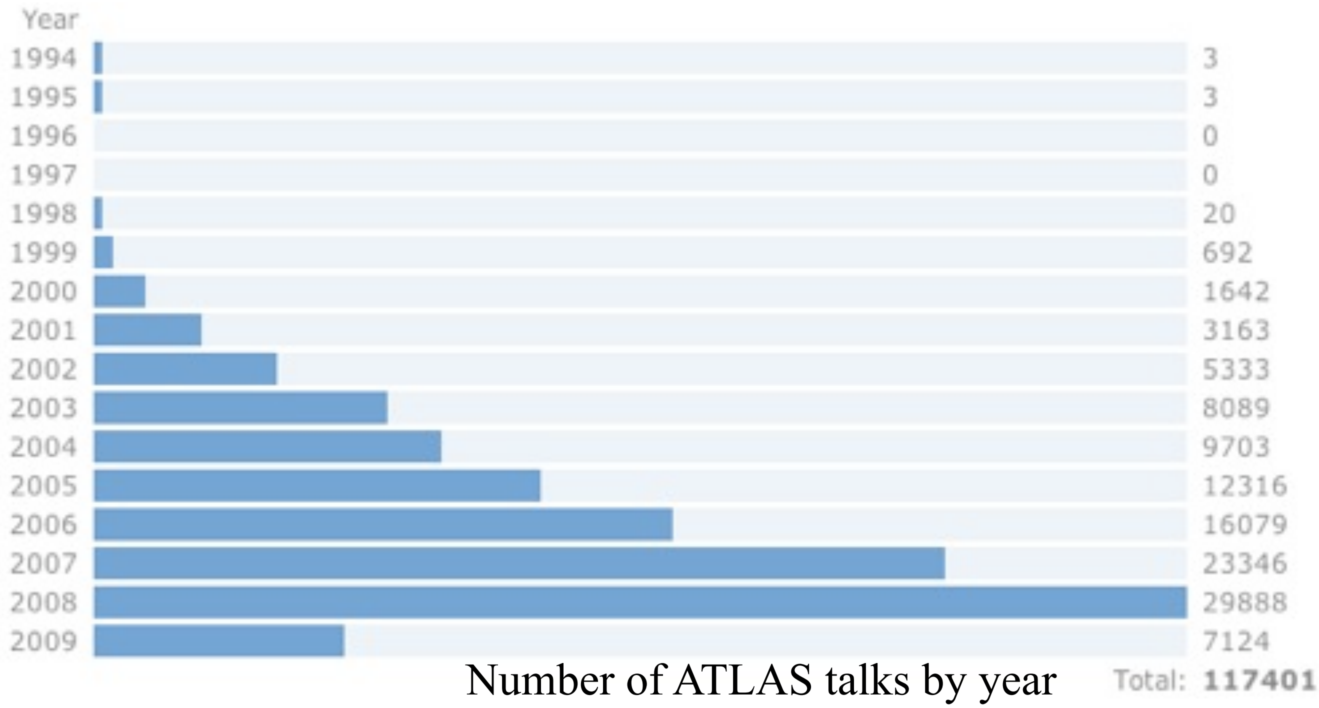
~~Modern statistics -- how can we gain some intuition?~~

Including Systematics and New ways to Publishing - are they useful?

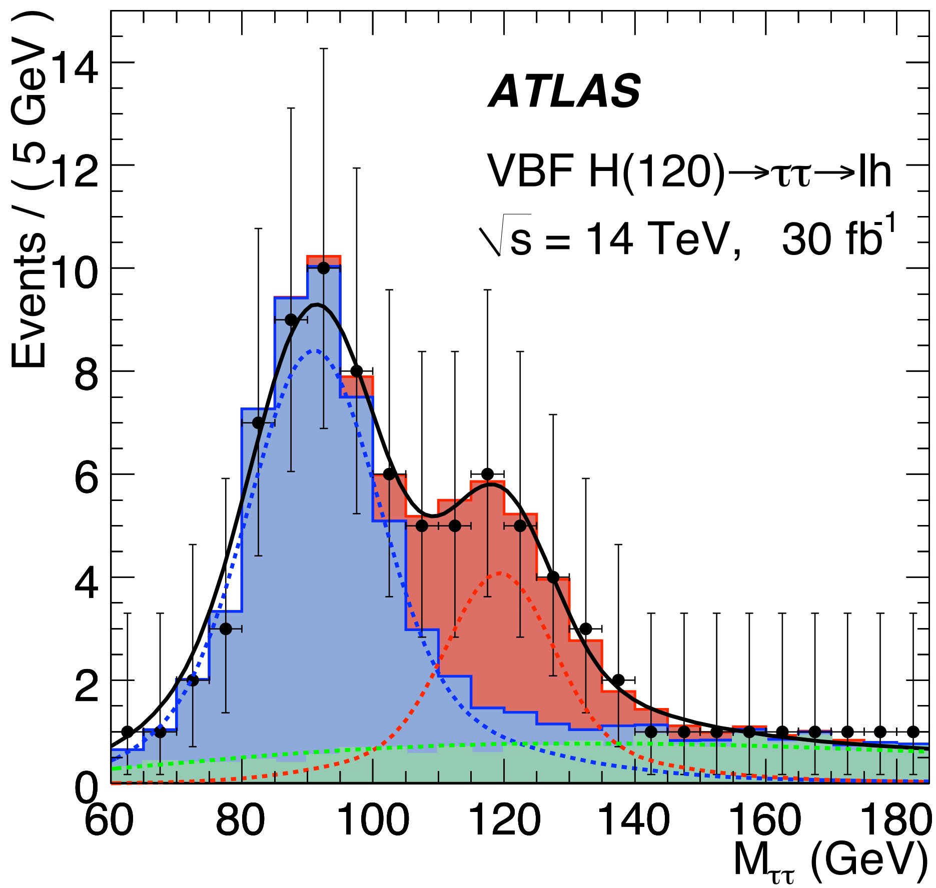


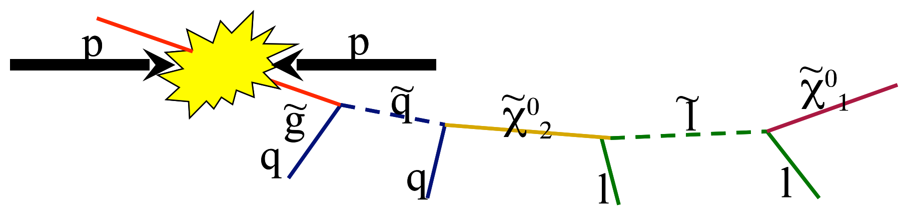
Number of ATLAS talks by year

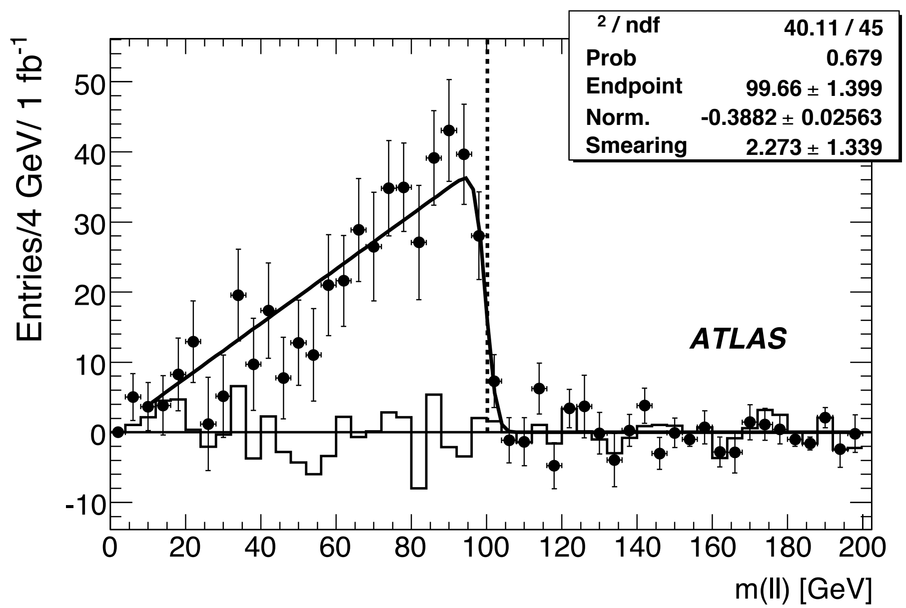
What's the Model

Beyond just providing histograms, any inference that we do will bring in some type of model for the data.

- What are the models in these two specific examples?



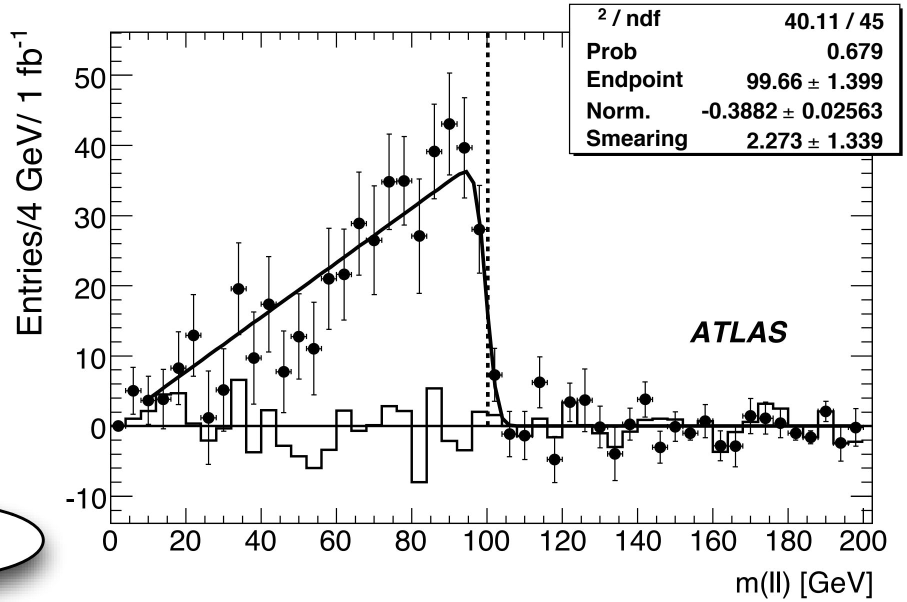




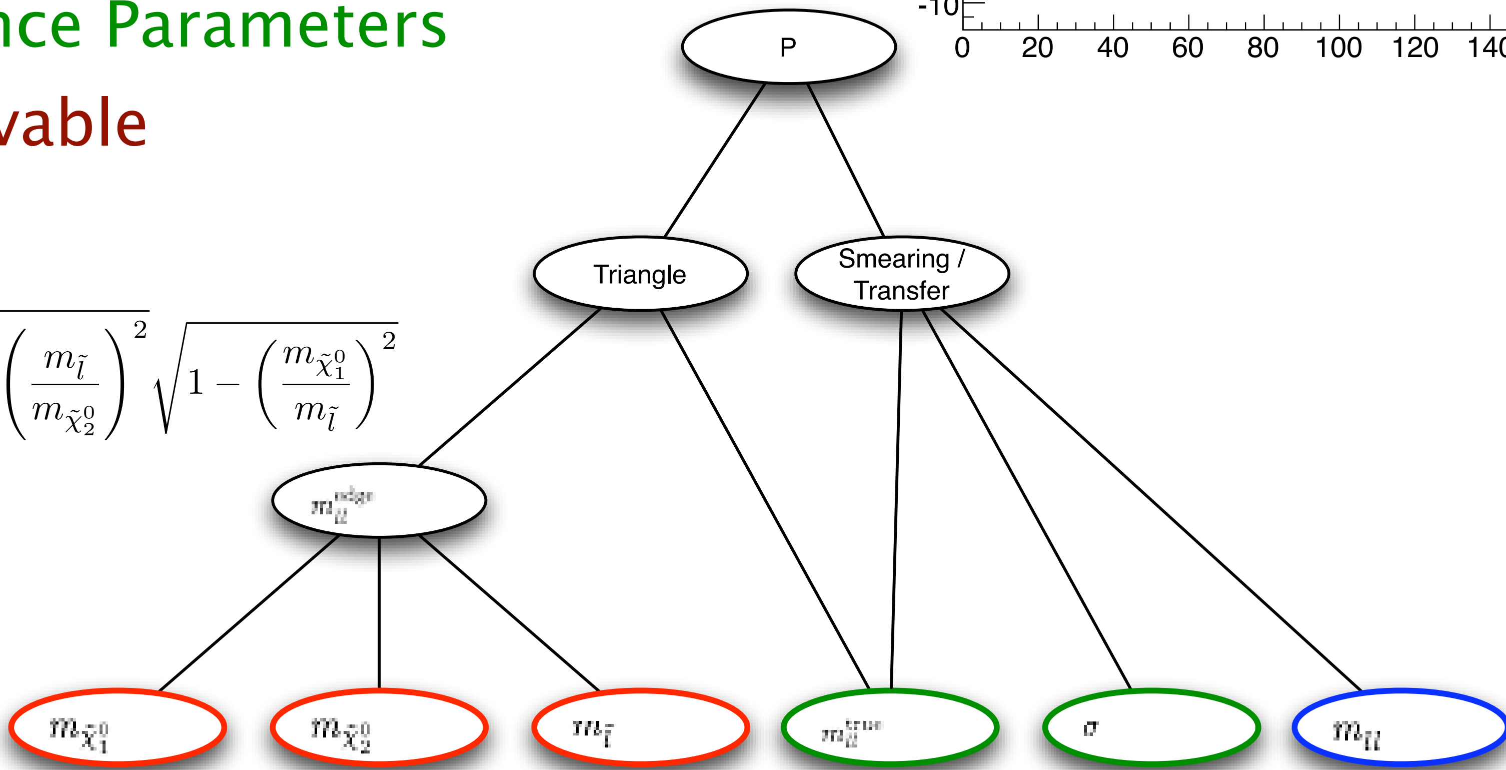
A graphical representation

Here is a graphical representation of this model that outlines its structural relationships

- Functions
- Parameters of Interest
- Nuisance Parameters
- Observable



$$m_{ll}^{\rm edge} = m_{\tilde\chi_2^0}\sqrt{1-\left(\frac{m_{\tilde l}}{m_{\tilde\chi_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde\chi_1^0}}{m_{\tilde l}}\right)^2}$$



$$P(m_{ll}|m_{\tilde\chi_1^0}, m_{\tilde\chi_2^0}, m_{\tilde l}, \sigma) = {\rm Triangle}(m_{ll}^{true}, m_{ll}^{\rm edge}(m_{\tilde\chi_1^0}, m_{\tilde\chi_2^0}, m)) \oplus {\rm Smearing}(m_{ll}^{true}, m_{ll})$$

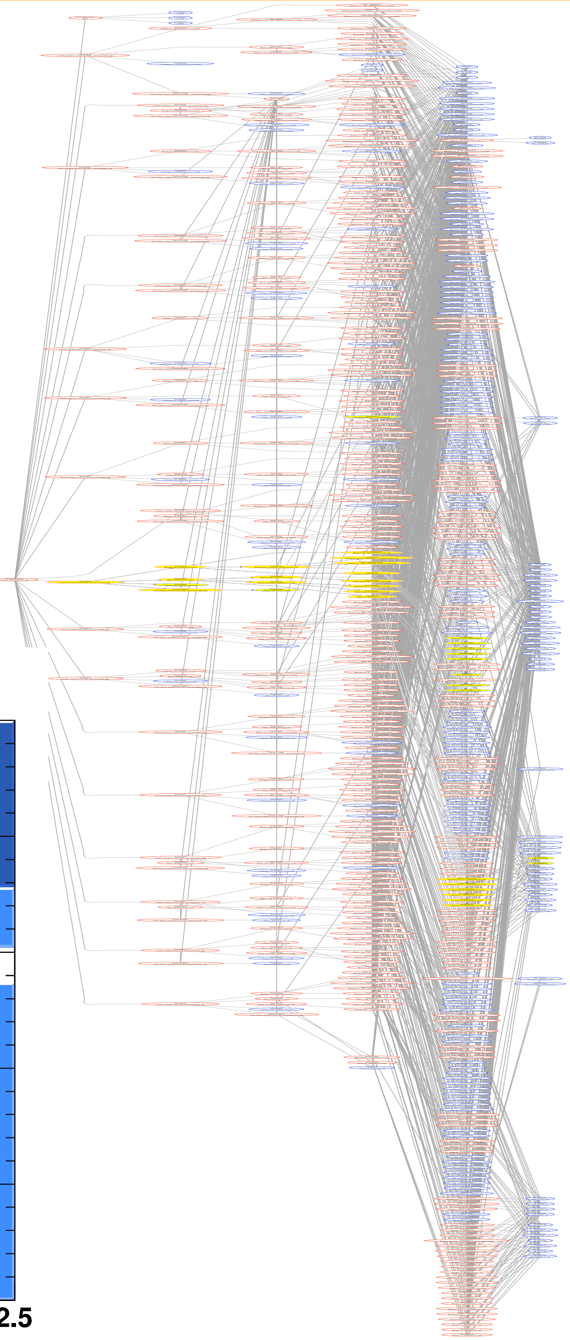
Thursday, February 26, 2009

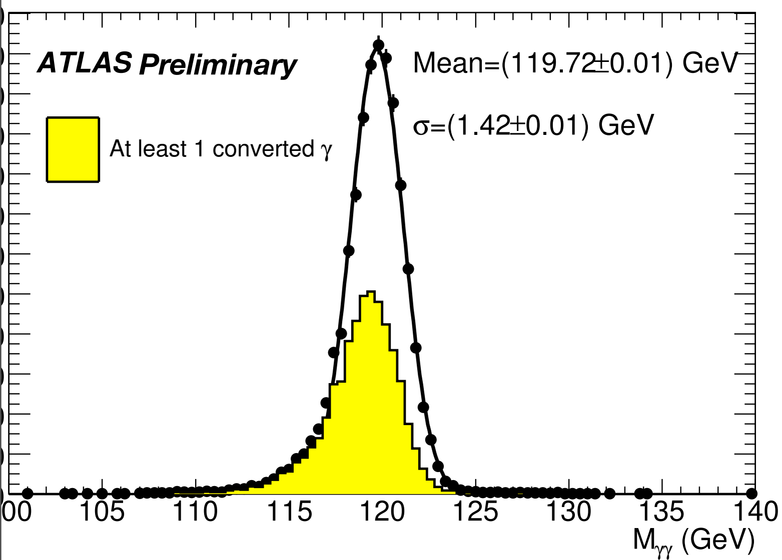
What's the Model

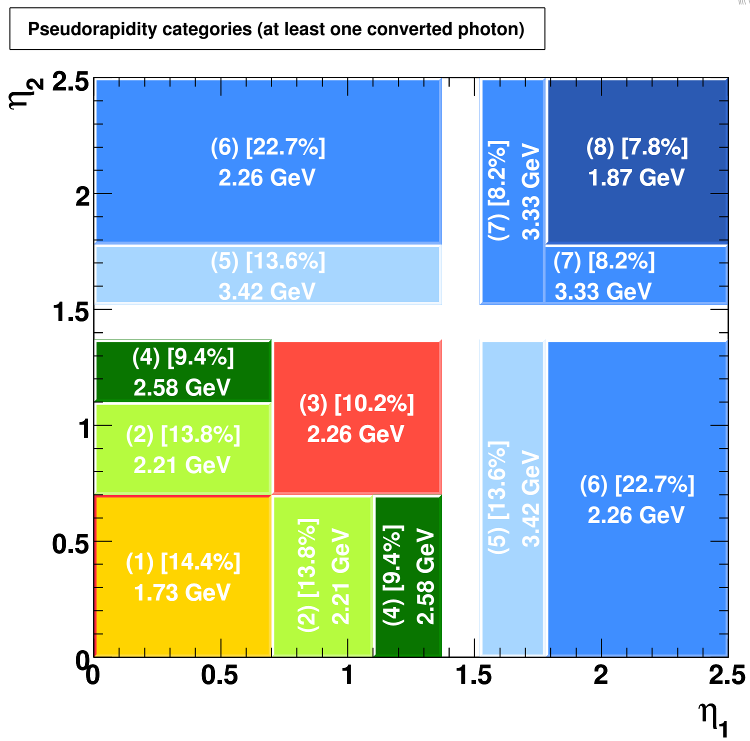
ATLAS fits to the $M_{\gamma\gamma}$ spectrum are categorized by the number of jets x conversion status x rapidity region

Similar sensitivity as CMS result

Complex model represented by this graph
(automatically generated by RooFit/RooStats)







Matrix Element Method as a Parametrization

With the same di-lepton mass distribution, we can either:

- relate edge according to:

$$m_{ll}^{\text{edge}} = m_{\tilde{\chi}_2^0}\sqrt{1-\left(\frac{m_{\tilde{l}}}{m_{\tilde{\chi}_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{l}}}\right)^2}$$

- incorporate matrix element techniques

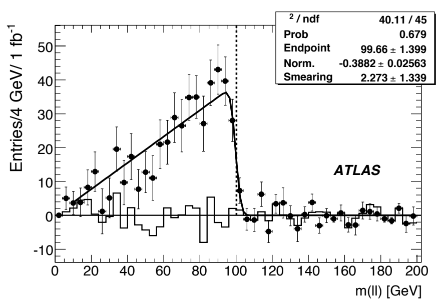
Matrix-element likelihood:
Calculate probability directly

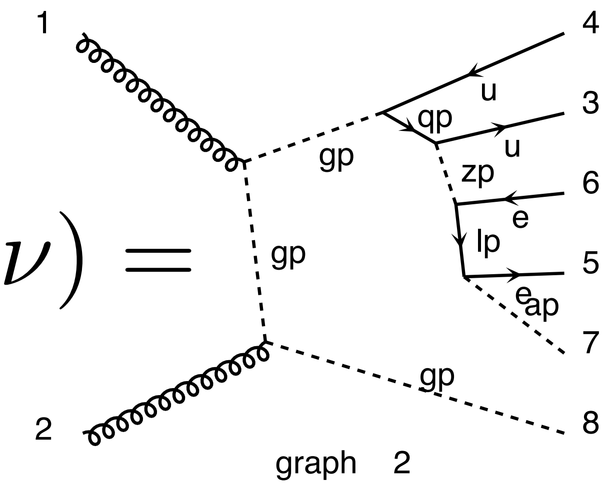
P(event z | SM) = P (z | process A) + P(z | process B) + ...

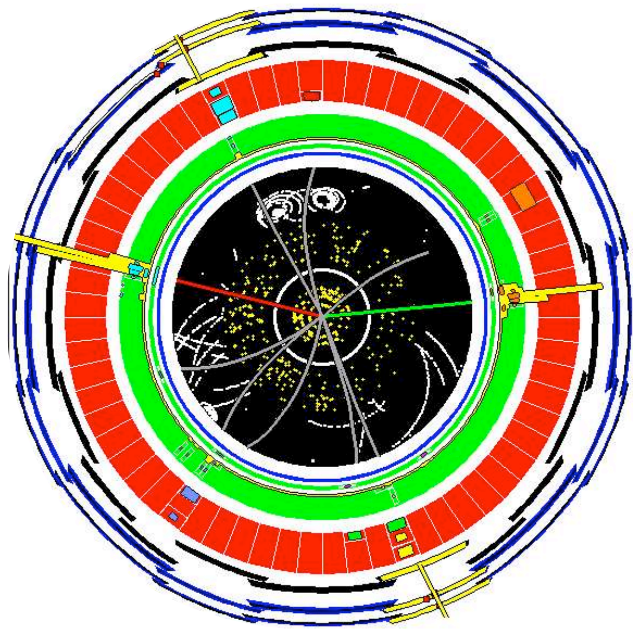
where

$P(\, z \mid A\,) = \int dy\ |\mathcal{M}_A|^2 f_p f_{\bar{p}}\ f_{TF}(y,z) = d\sigma_A/dz$

- dy: Integration over parton-level quantities
- $|\mathcal{M}_A|^2 f_p f_{\bar{p}}$: Matrix-element*PDFs for process A *(Issue 2)*
- $f_{TF}(y,z)$: Parton(y) to detector(z) transfer function (TF) describes parton-shower and detector response in parametrized form *(Issue 2)*

$P($ 

$|\mu,\nu) =$ 

$\oplus$ 

Matrix Element Method as a Parametrization

With the same di-lepton mass distribution, we can either:

- relate edge according to:

$$m_{ll}^{\text{edge}} = m_{\tilde{\chi}_2^0}\sqrt{1-\left(\frac{m_{\tilde{l}}}{m_{\tilde{\chi}_2^0}}\right)^2}\sqrt{1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{l}}}\right)^2}$$

- incorporate matrix element techniques
  - naturally, could include more kinematic info -> more power.

Matrix-element likelihood:
Calculate probability directly

P(event z | SM) = P (z | process A) + P(z | process B) + ...

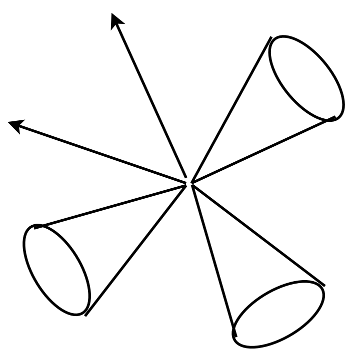
where

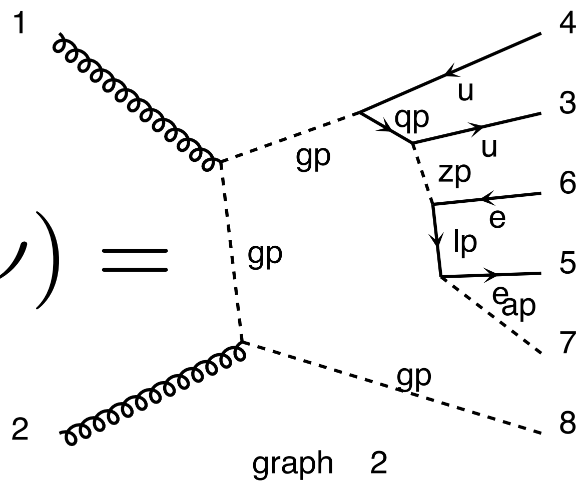
$P(\ z\ |\ A\ ) = \int dy\ |\mathcal{M}_A|^2 f_p f_{\bar{p}}\ f_{TF}(y,z) = d\sigma_A/dz$

Parton(y) to detector(z) transfer function (TF) describes parton-shower and detector response in parametrized form *(Issue 2)*

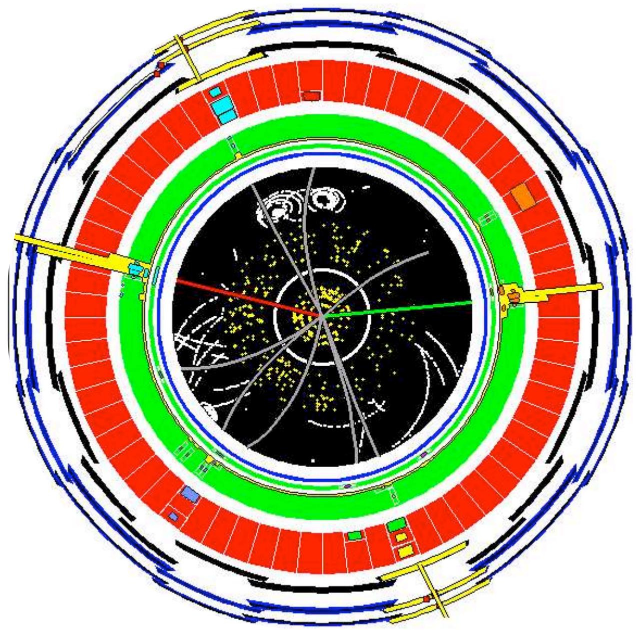
Matrix-element*PDFs for process A *(Issue 2)*

Integration over parton-level quantities

$P($  $|\mu,\nu) =$



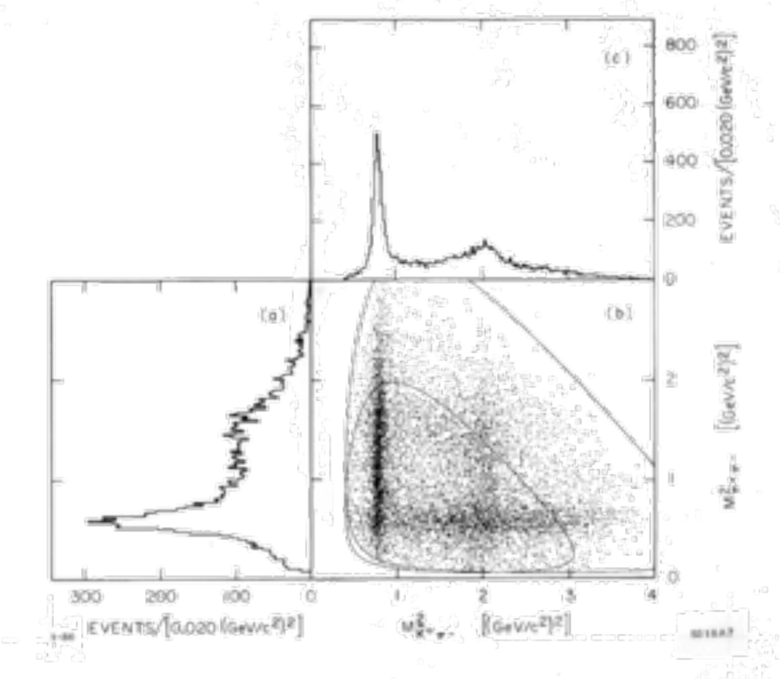
$\oplus$



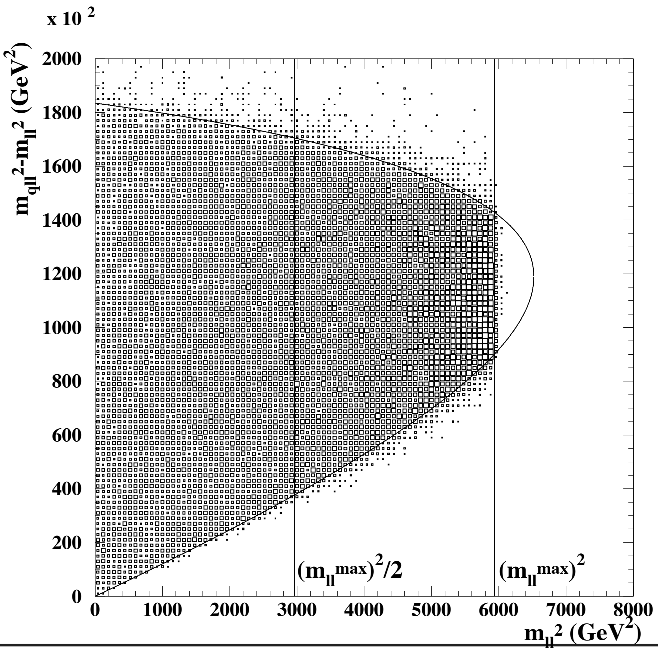
Matrix Element vs. Purely Kinematic Features

Tuesday, we had a nice discussion about unifying kinematic features (eg. edges, MT2, hybrid method, etc.) in terms of likelihood-ratio inspired observables restricted to boundaries in phase space (or other well-defined features)

- less dependence on $|\mathcal{M}|^2$ across phase space
- basically generalizations of Dalitz plots



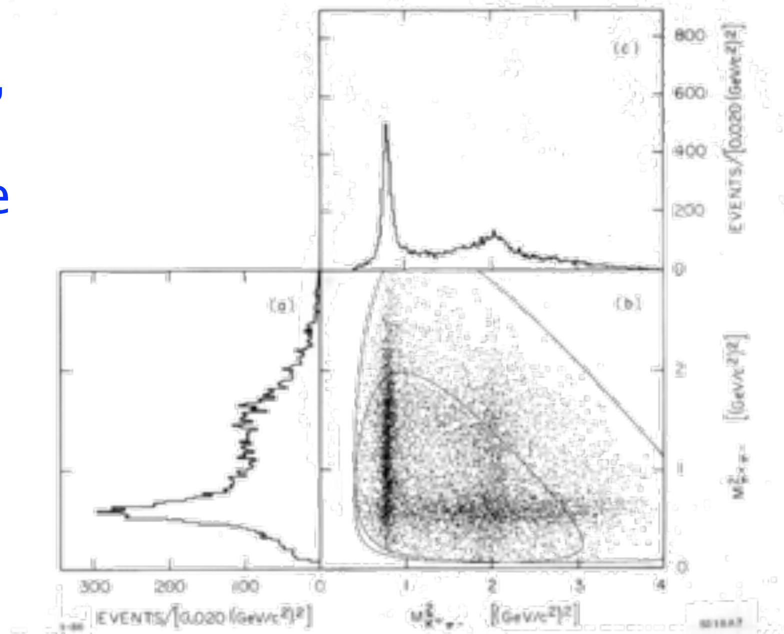
Tovey, Costanzo [arXiv:0902.2331]



Matrix Element vs. Purely Kinematic Features

Tuesday, we had a nice discussion about unifying kinematic features (eg. edges, MT2, hybrid method, etc.) in terms of likelihood-ratio inspired observables restricted to boundaries in phase space (or other well-defined features)

- less dependence on $|\mathcal{M}|^2$ across phase space
- basically generalizations of Dalitz plots
- ... I remembered that J^P of parent can lead to vanishing densities near kinematic endpoints.



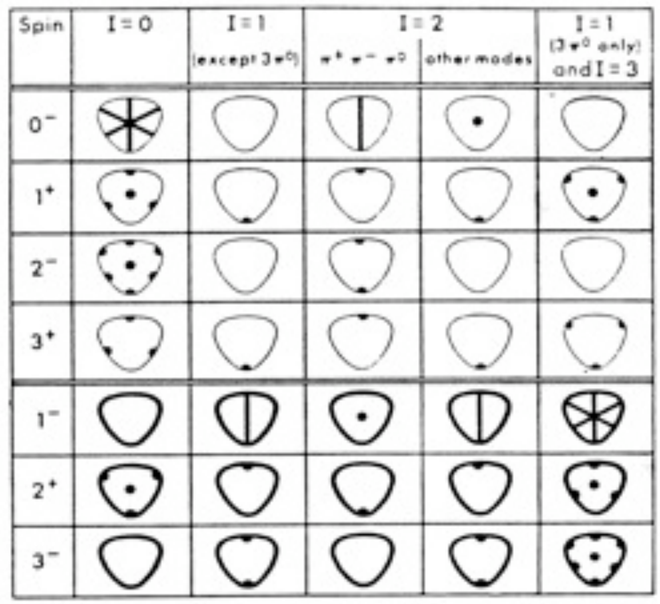
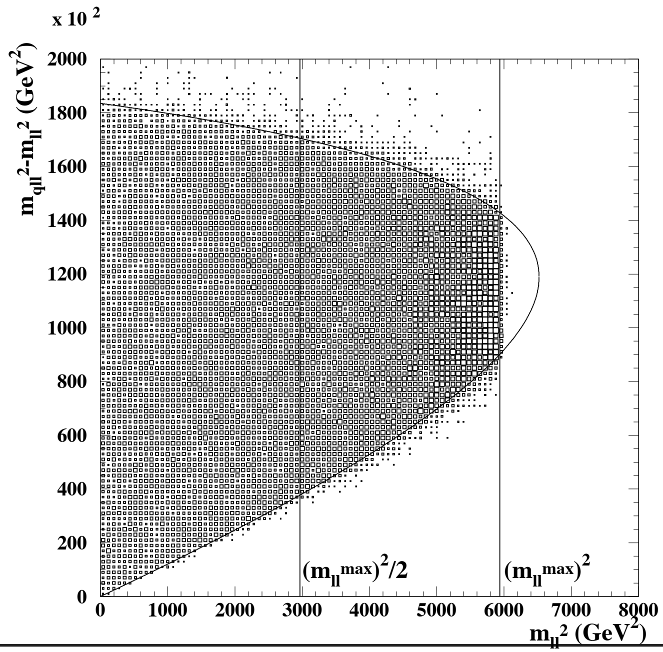


FIG. 2. Regions of the 3π Dalitz plot where the density must vanish because of symmetry requirements are shown in black. The vanishing is of higher order (stronger) where black lines and dots overlap. In each isospin and parity state, the pattern for a spin of $J+$ even integer is identical to the pattern for spin J, provided $J \geq 2$. (Exception: vanishing at the center is not required for $J \geq 4$.)

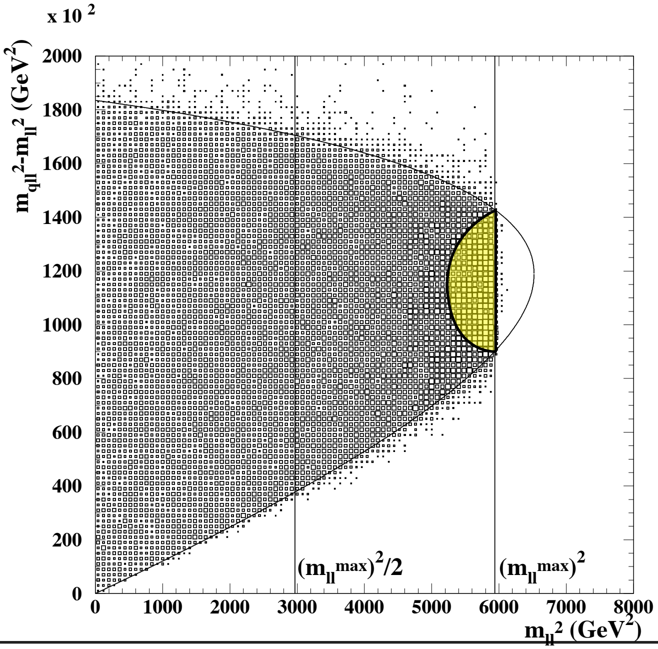
Tovey, Costanzo [arXiv:0902.2331]



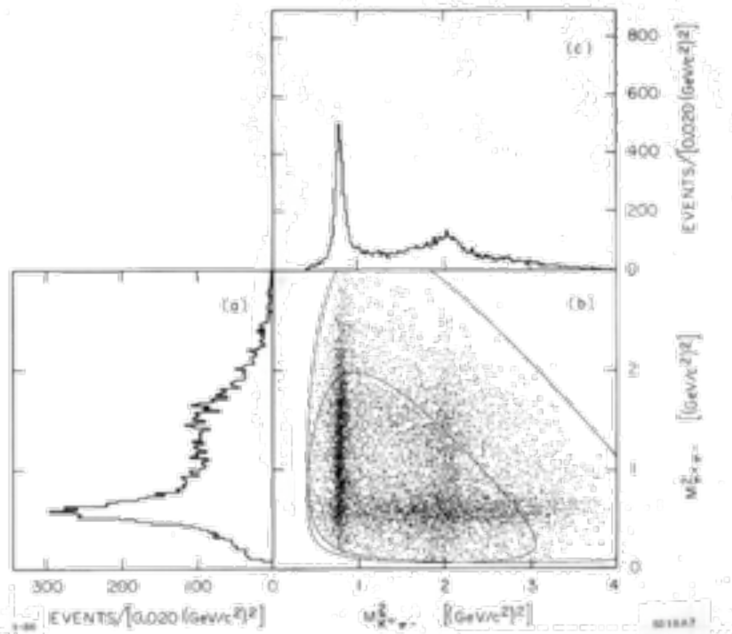
Matrix Element vs. Purely Kinematic Features

Tuesday, we had a nice discussion about unifying kinematic features (eg. edges, MT2, hybrid method, etc.) in terms of likelihood-ratio inspired observables restricted to boundaries in phase space (or other well-defined features)

- less dependence on $|\mathcal{M}|^2$ across phase space
- basically generalizations of Dalitz plots
- ... I remembered that J^P of parent can lead to vanishing densities near kinematic endpoints.
- Don't see holes without spin. Could this lead to "false endpoints" due to M.E. suppression?



Tovey, Costanzo [arXiv:0902.2331]



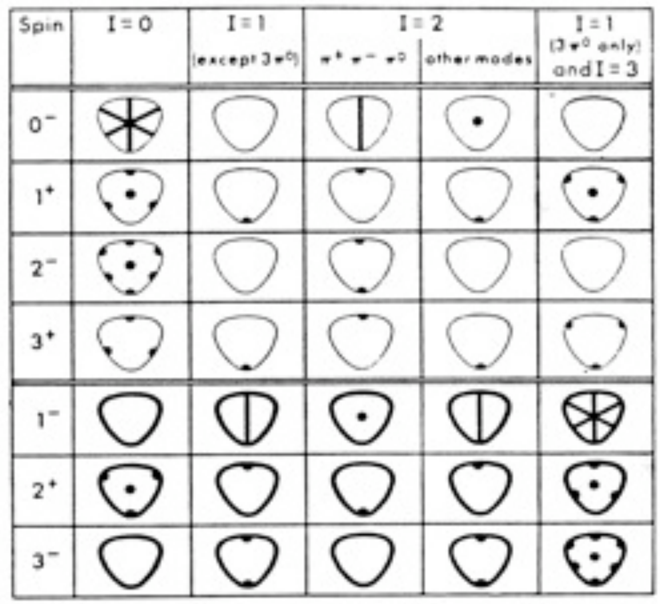
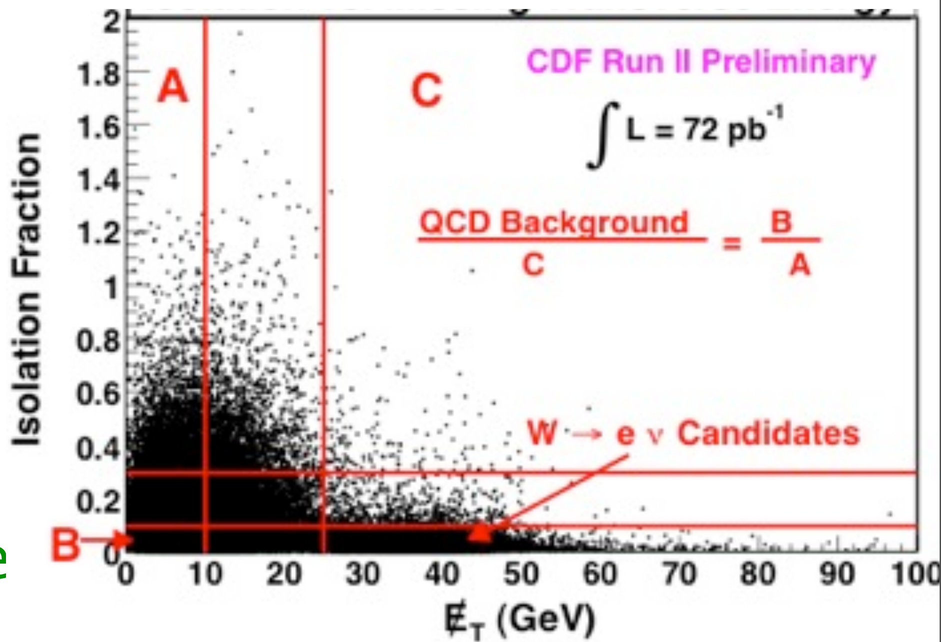


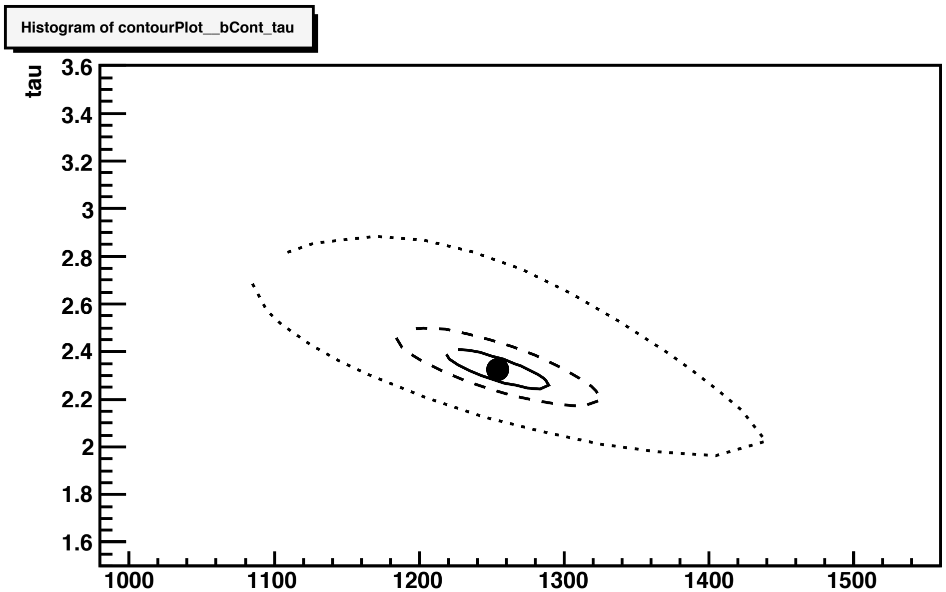
FIG. 2. Regions of the 3π Dalitz plot where the density must vanish because of symmetry requirements are shown in black. The vanishing is of higher order (stronger) where black lines and dots overlap. In each isospin and parity state, the pattern for a spin of $J+$ even integer is identical to the pattern for spin J, provided $J \geq 2$. (Exception: vanishing at the center is not required for $J \geq 4$.)

Data Driven SUSY Background Estimation

Even simple number-counting type analyses have a model.

- Background is estimated as B/A*C
  - simple propagation of errors misses across Poisson gives non-Gaussian correlations
  - Given measurements in A,B,C,D we can see contours with clear non-Gaussian correlations



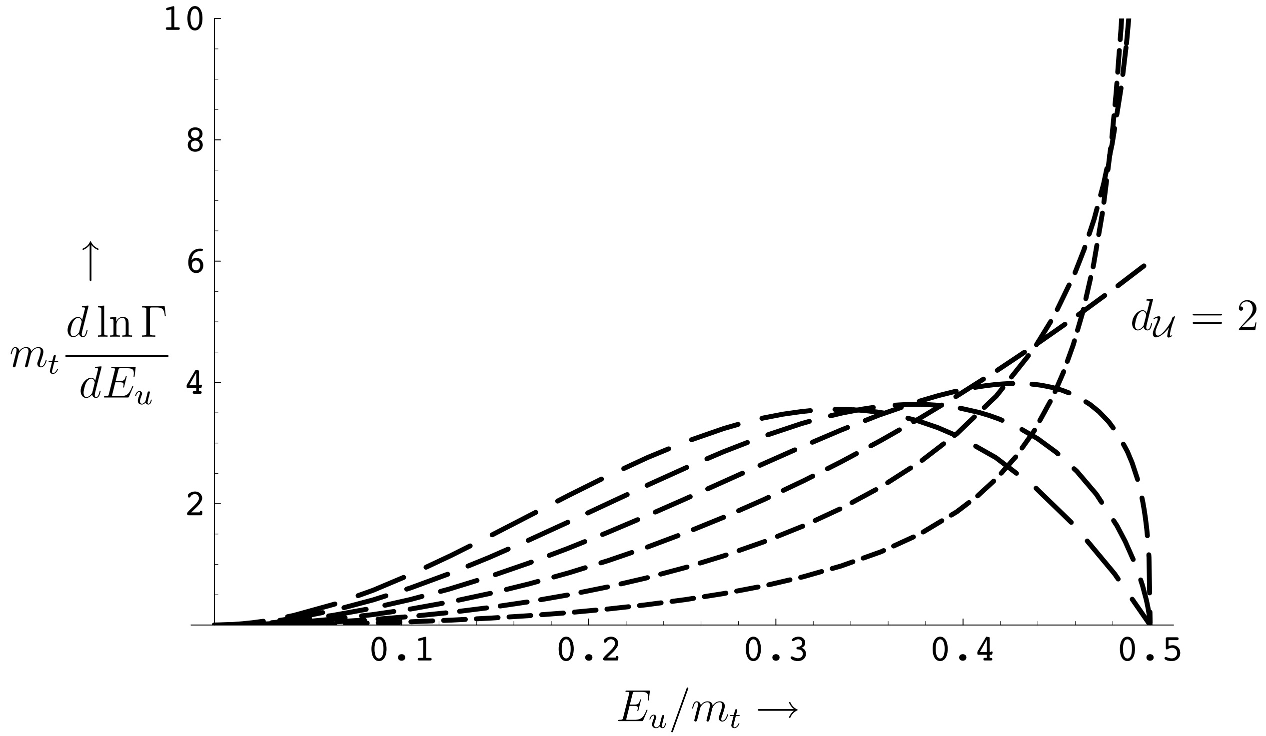



```
RooRealVar   s("s","s",_x-_y, 0., 200.);
RooRealVar   b("b","b",_y*_xCont/_yCont,  0., 500.);
RooRealVar   tau("tau","tau",_yCont/_xCont,0,5);
RooRealVar   bCont("bCont","b control",_xCont,  0., 2000.);

RooFormulaVar  splusb("splusb","s+b",RooArgSet(s,b));
RooProduct   bTau("bTau","b*tau",RooArgSet(b, tau));
RooProduct   bContTau("bContTau","bCont*tau",RooArgSet(bCont, tau));
RooRealVar   x("x","x", _x, 0., 1000.);
RooRealVar   y("y","y", _y,  0., 1000.);
RooRealVar   yCont("yCont","y Control", _yCont,  0., 3000.);
RooRealVar   xCont("xCont","x Control", _xCont,  0., 3000.);

RooPoisson sigRegion("sigRegion","sigRegion",x,splusb);
RooPoisson sideband("sideband","sideband",y,bTau);
RooPoisson xContDist("xContDist","xCont dist",xCont,bCont);
RooPoisson yContDist("yContDist","yCont dist",yCont,bContTau);

RooProdPdf joint("joint","joint", RooArgSet(sigRegion,sideband,
                                          xContDist, yContDist) );
```

Unparticle have models too

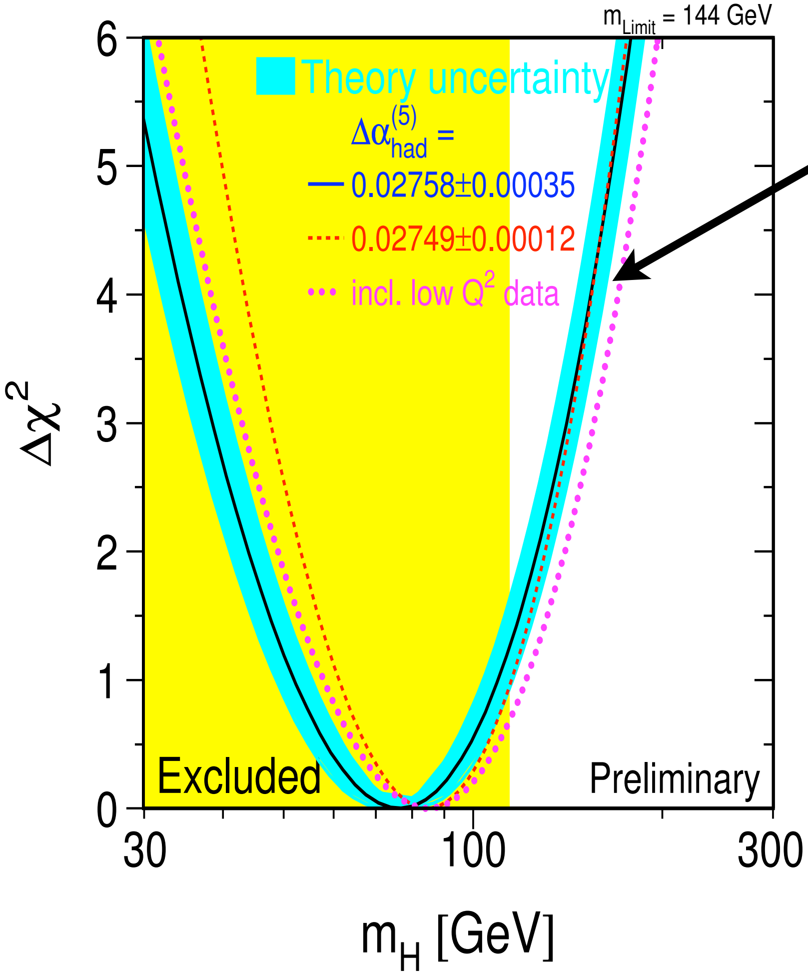
While we’ve been talking about Matrix Element methods and Feynman-diagram like models for new physics...

... unparticles have models too.



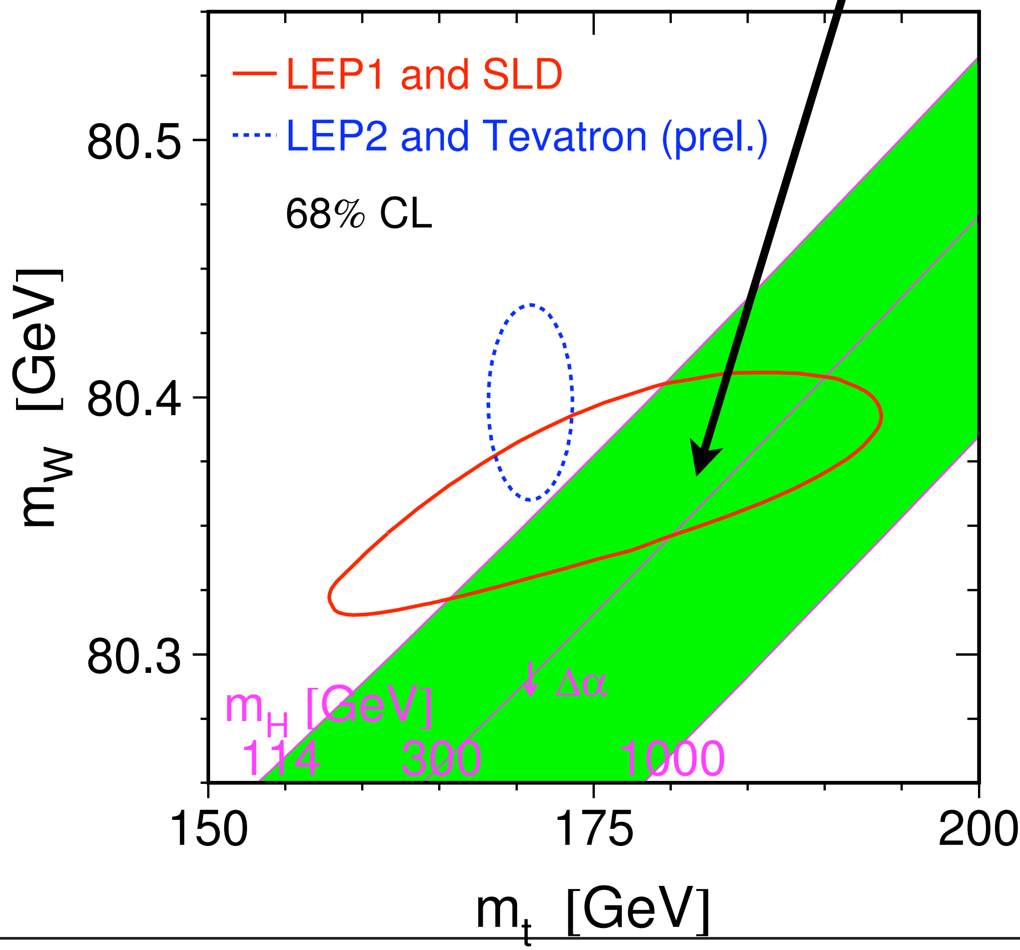
Examples of Published Likelihoods

Eventually, we will aim to publish some likelihood functions of measured parameters.

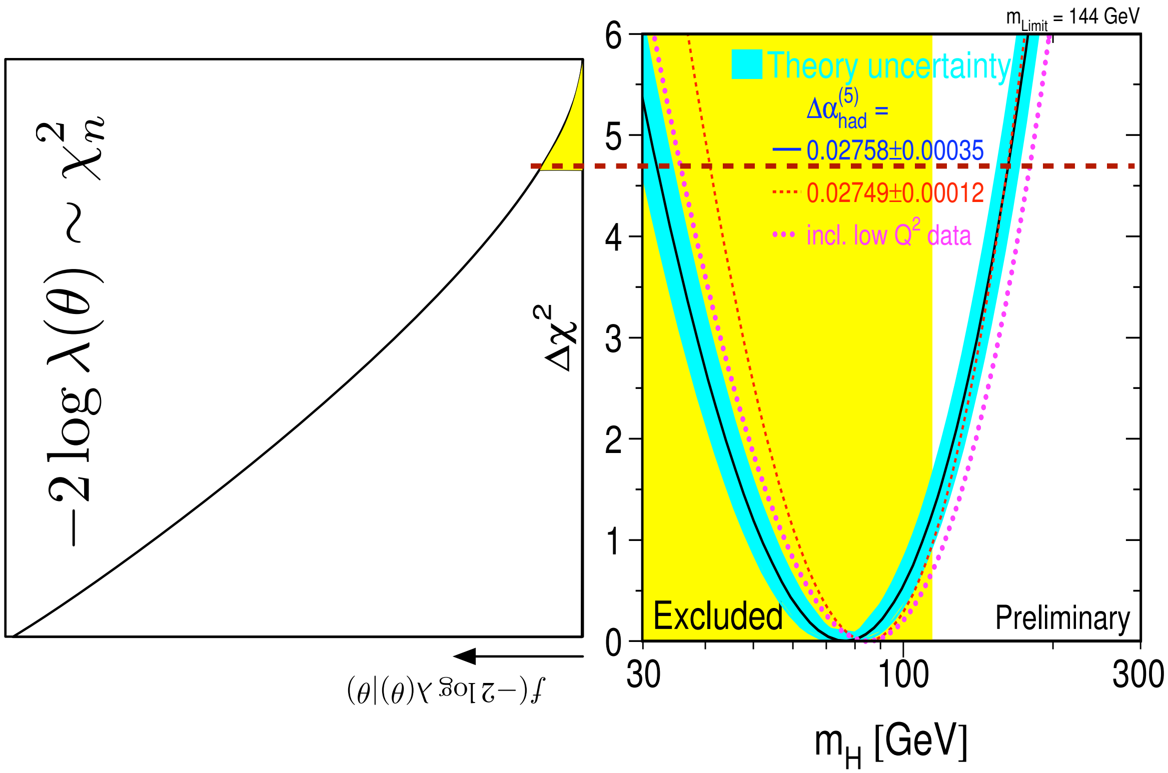


You can find examples of published likelihoods in 1D

In 2-D you just get the contours



Likelihood-based Intervals



Let's recall quickly the logic of limits derived from these types of likelihood curves.

- A certain value of $-2\log\lambda(\mu)$ corresponds to a χ^2
- We know the χ^2 distribution, so we can translate to a 95% limit

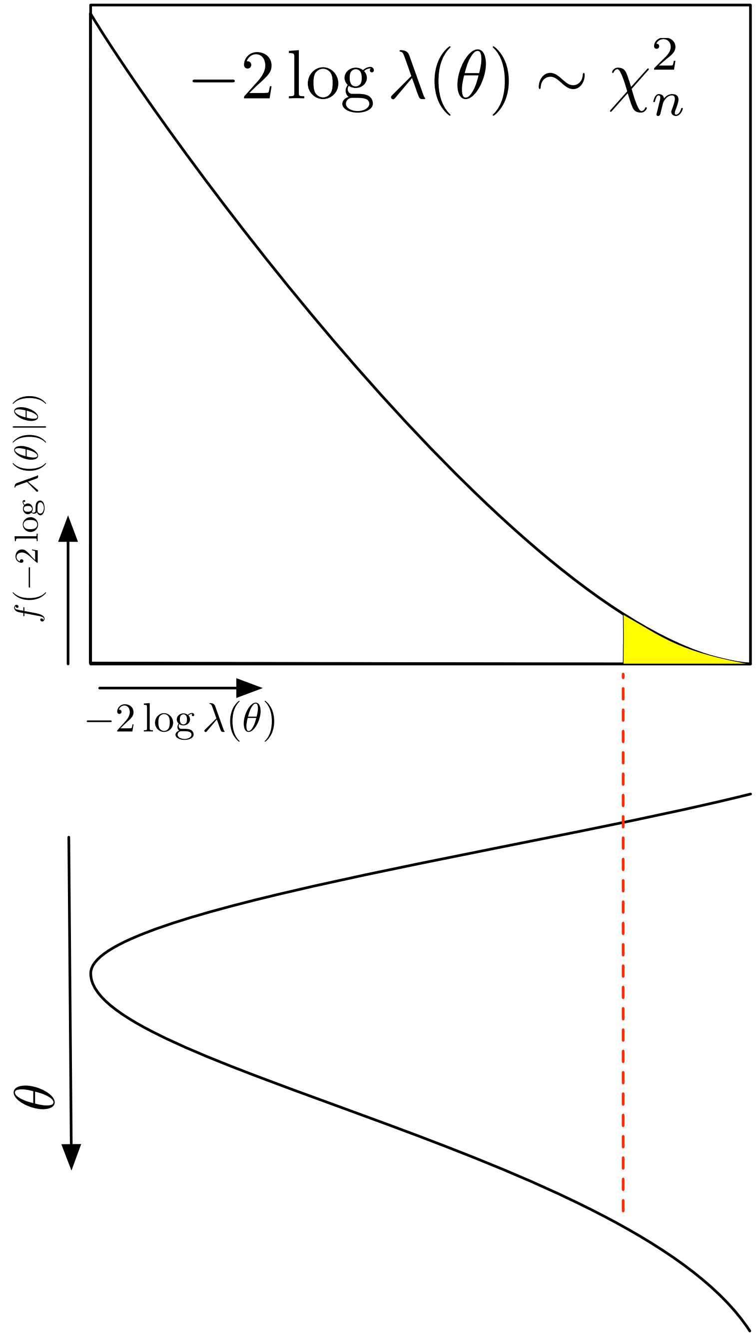
Likelihood-based Intervals

Wilks's theorem tells us how the profile likelihood ratio is distributed

- asymptotically it is χ^2
- there are some restrictions on the parametrized family of models
  - eg. boundaries & look-elsewhere

For these limits to be right, the model needs to include the 'true distribution':

- eg. what is actually producing the data
- eg. nature + our real detectors



Systematics, Systematics, Systematics

Incorporating Systematics by "Profiling"

Let μ be the parameters we are interested in

Let ν be the rest of them (nuisance parameters)

Let $\hat{\nu}$ be the best fit to the data

Let $\hat{\hat{\nu}}(\mu)$ be the best fit to the data with μ constrained

Consider the profile likelihood ratio

$$\lambda(\mu) = \frac{P(data|\mu, \hat{\hat{b}}(\mu), \hat{\hat{\nu}}(\mu))}{P(data|\hat{\mu}, \hat{b}, \hat{\nu})}$$

one can see the **function** is independent of true values of

- though its **distribution** might depend indirectly

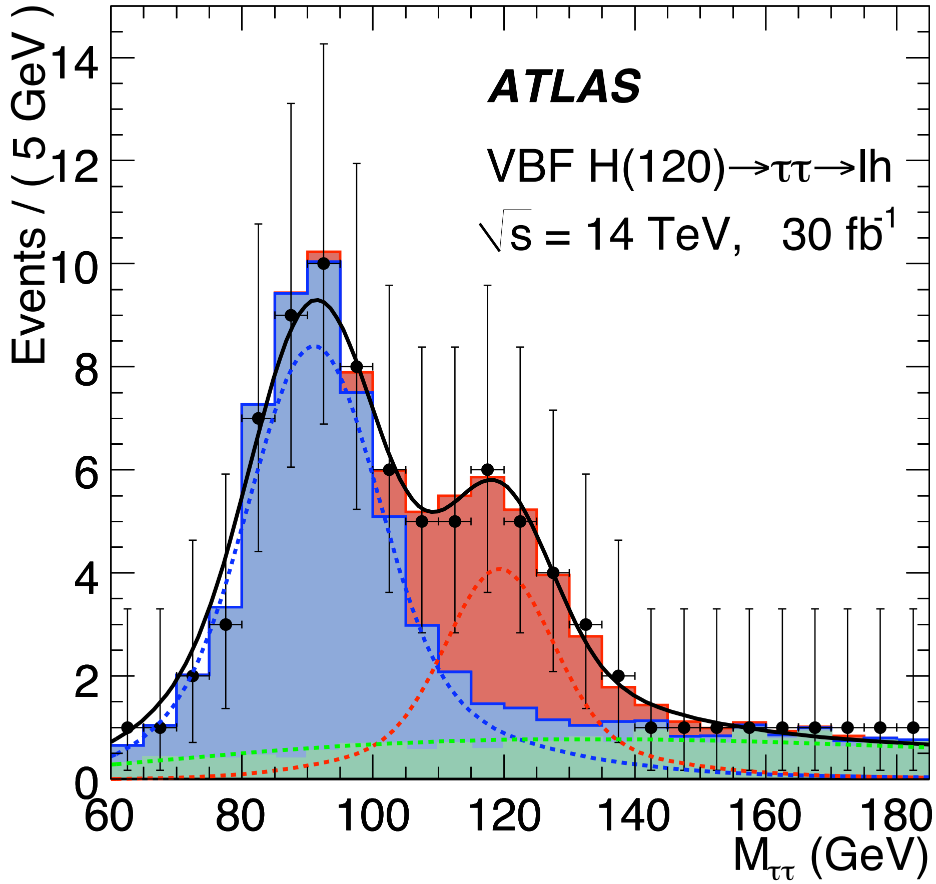
Again, if the model includes the true distribution, then we know its distribution: $-2\log\lambda(\mu) \sim \chi^2_N$

An example

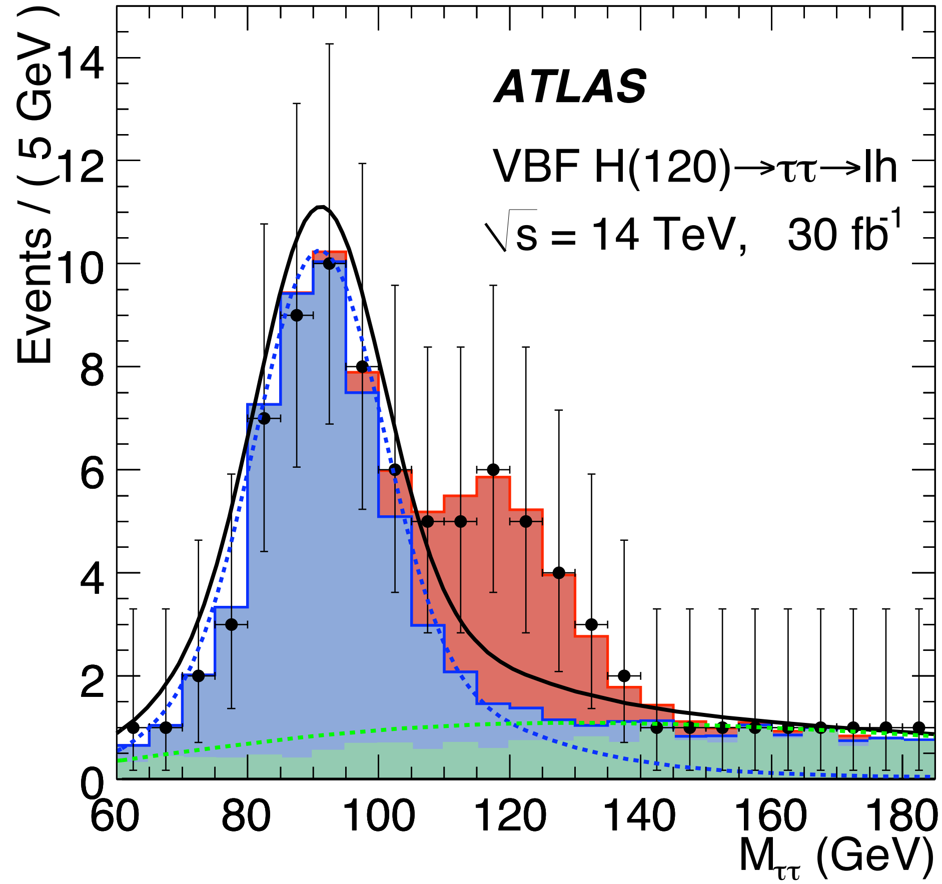
Essentially, you need to fit your model to the data twice: once with everything floating, and once with signal fixed to 0

$$\lambda(\mu=0) = \frac{L(data|\mu=0,\hat{\hat{b}}(\mu=0),\hat{\hat{\nu}}(\mu=0))}{L(data|\hat{\mu},\hat{b},\hat{\nu})},$$

$L(data|\hat{\mu},\hat{b},\hat{\nu})$



$L(data|\mu=0,\hat{\hat{b}},\hat{\hat{\nu}})$



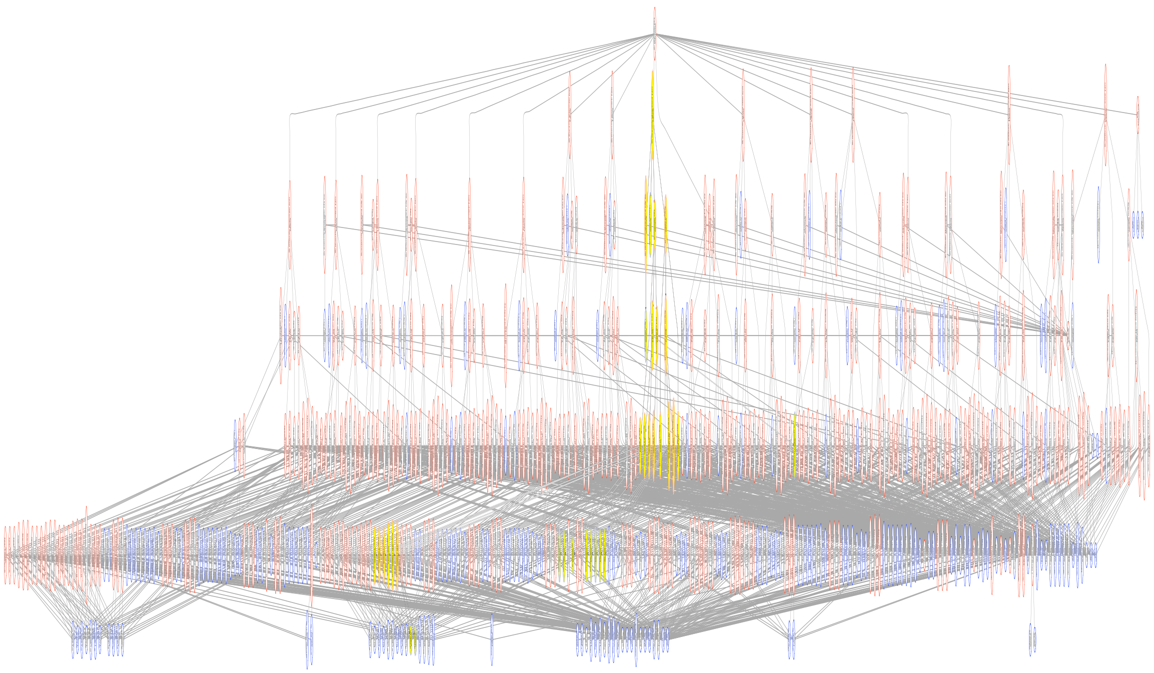
Checking Wilks

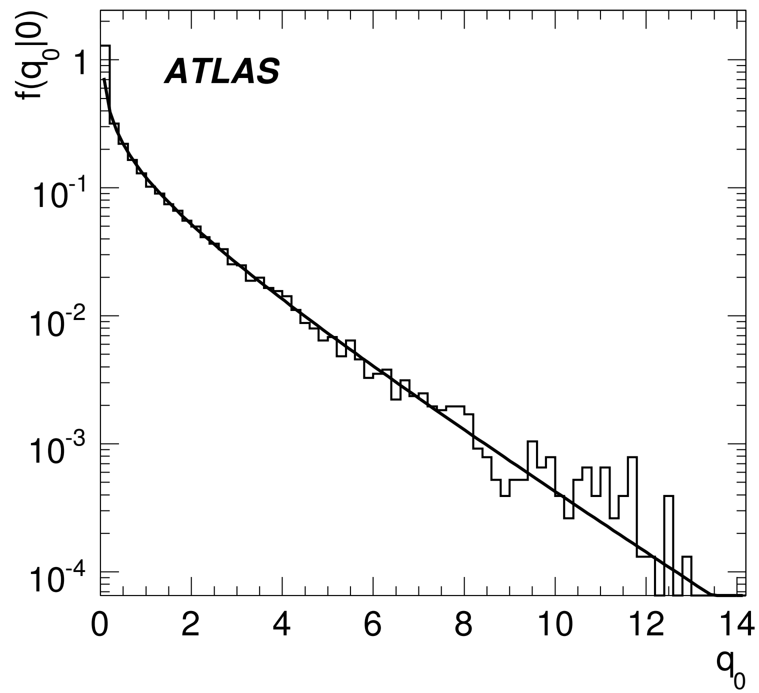
An recent ATLAS Higgs example:

- even with very complicated pdf, distribution looks χ^2

$$\lambda(\mu=0) = \frac{L(data|\mu=0, \hat{\hat{b}}(\mu=0), \hat{\hat{\nu}}(\mu=0))}{L(data|\hat{\mu}, \hat{b}, \hat{\nu})},$$

$$q_0 = -2\log\lambda(\mu=0)$$



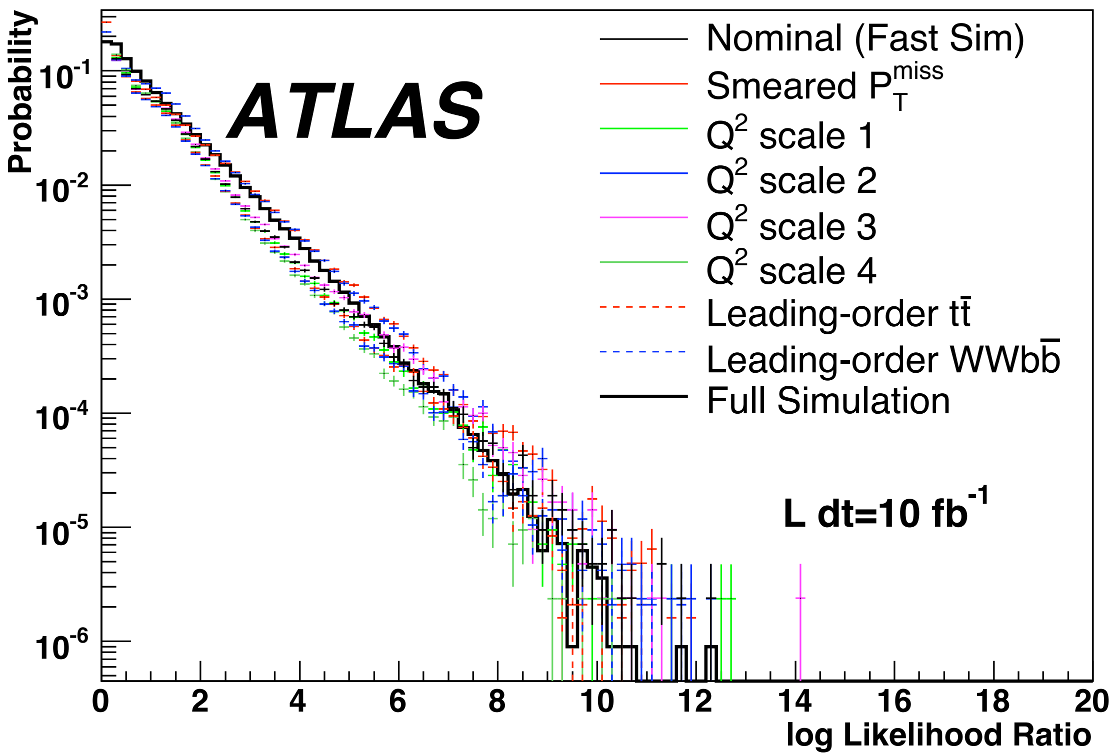


Experimentalist Justification

So far this looks a bit like magic. How can you claim that you incorporated your systematic just by fitting the best value of your uncertain parameters and making a ratio?

It won't unless the the parametrization is sufficiently flexible.

So check by varying the settings of your simulation, and see if the profile likelihood ratio is still distributed as a chi-square



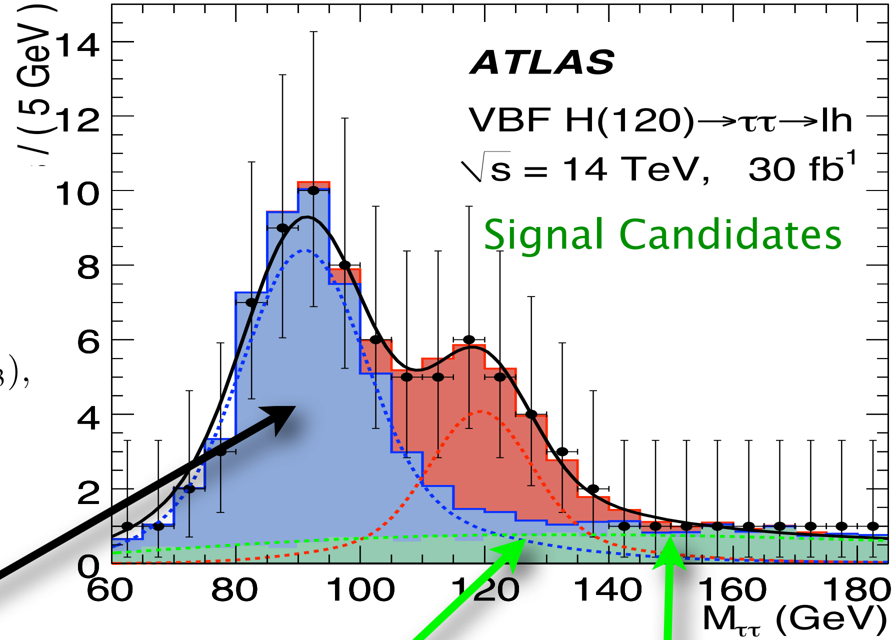
Here it is pretty stable, but it's not perfect (and this is a log plot, so it hides some pretty big discrepancies)

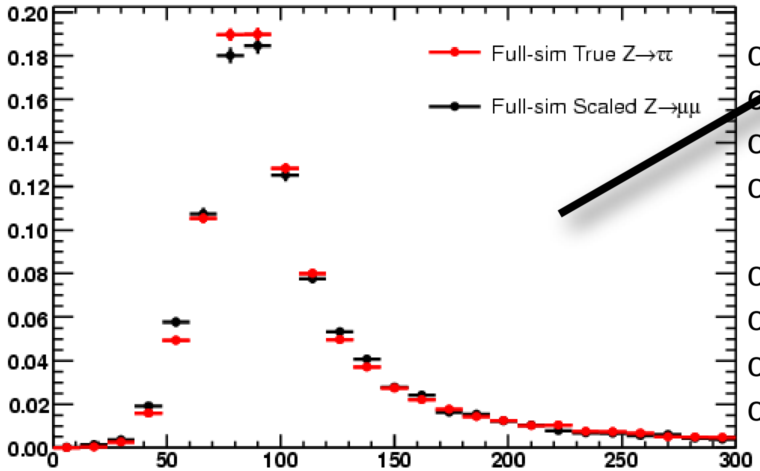
The profile approach works asymptotically and only if your parametrization is sufficiently flexible.

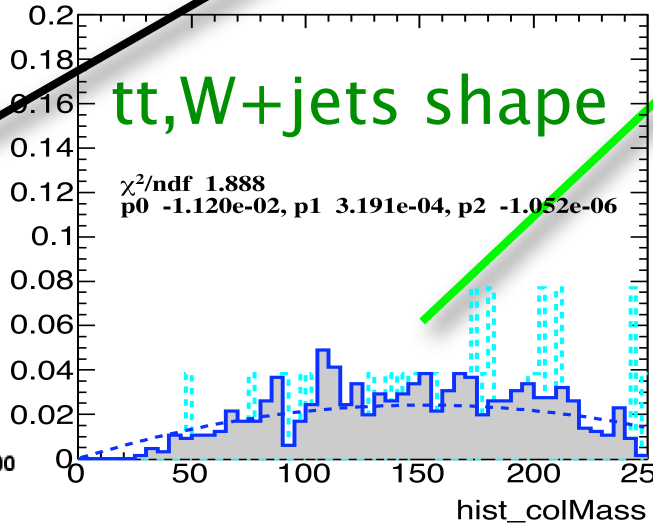
Simultaneous Fitting Strategy

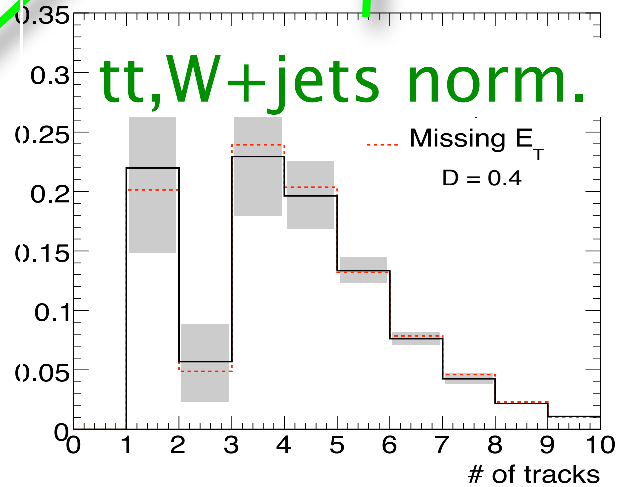
Simultaneous fit to signal and control samples control background shape and normalization from data

$$
\begin{aligned}
L(data|\mu, M_H, \nu) &= L_{track}(\text{track multiplicity}|r_{QCD}) \\
&\times L_Z(\text{Z+jets control}|\sigma_Z) \\
&\times L_{QCD}(\text{QCD control}|a_0, a_1, a_2, a_3) \\
&\times L_{s+b}(\text{signal candidates}|\mu, M_H, \sigma_H, \sigma_Z, r_{QCD}, a_0, a_1, a_2, a_3),
\end{aligned}
$$







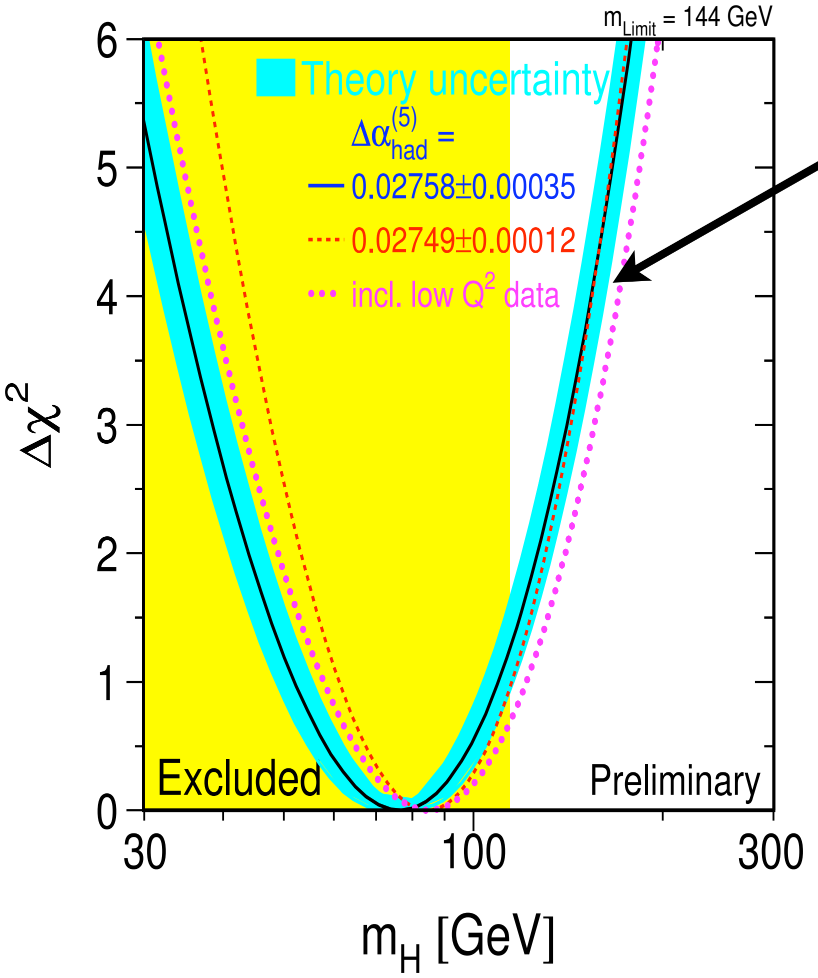


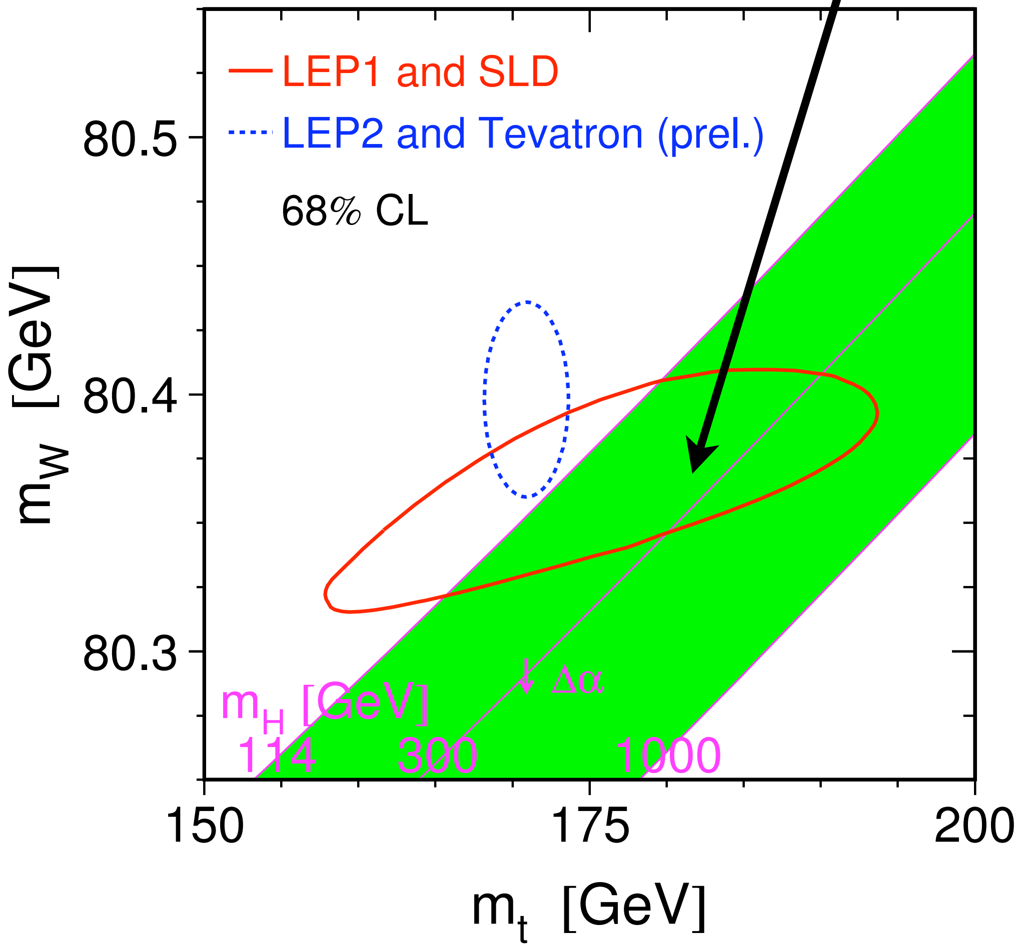
Examples of Published Likelihoods

In statistical meetings for experiments, we agreed to publish likelihood functions

You can find examples of published likelihoods in 1D

In 2-D you just get the contours





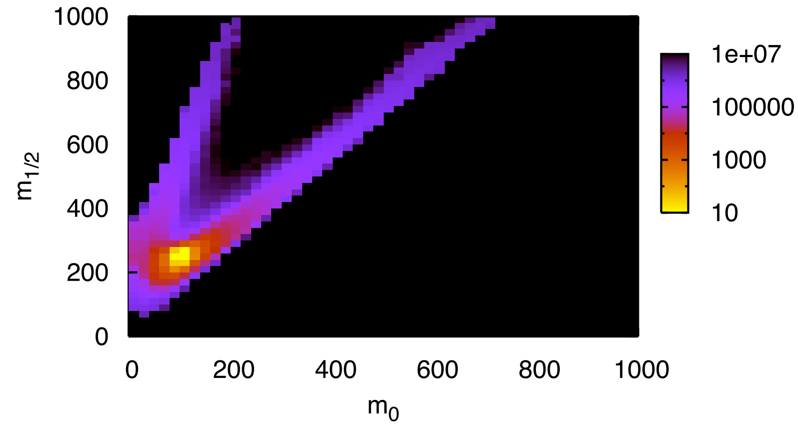
Surely we can do better!

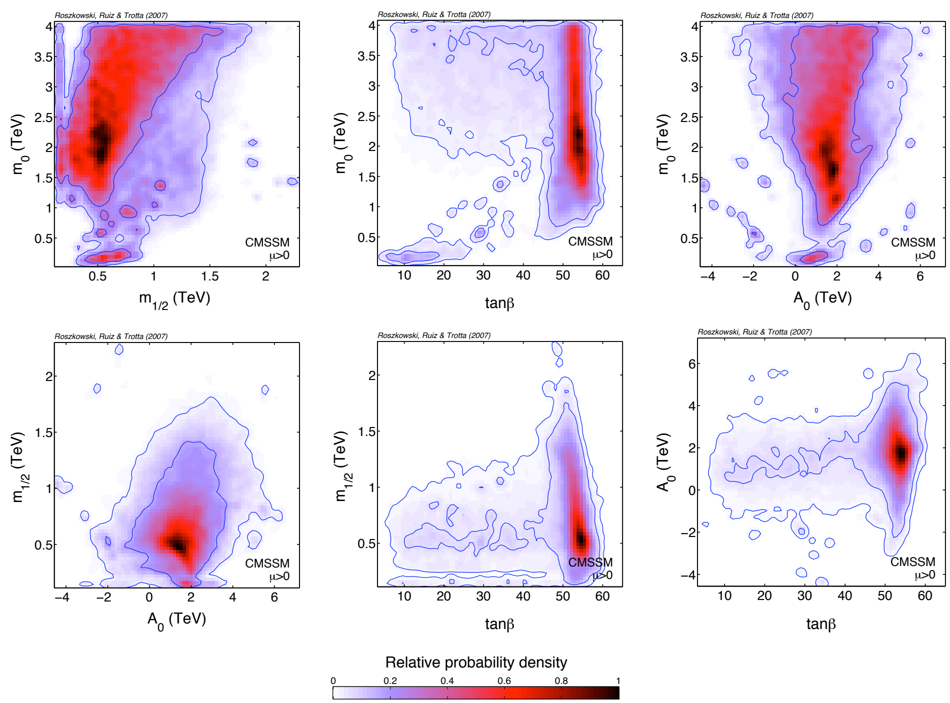
Publishing Likelihood Maps

From existing measurements, several groups have done weather forecasting for LHC:

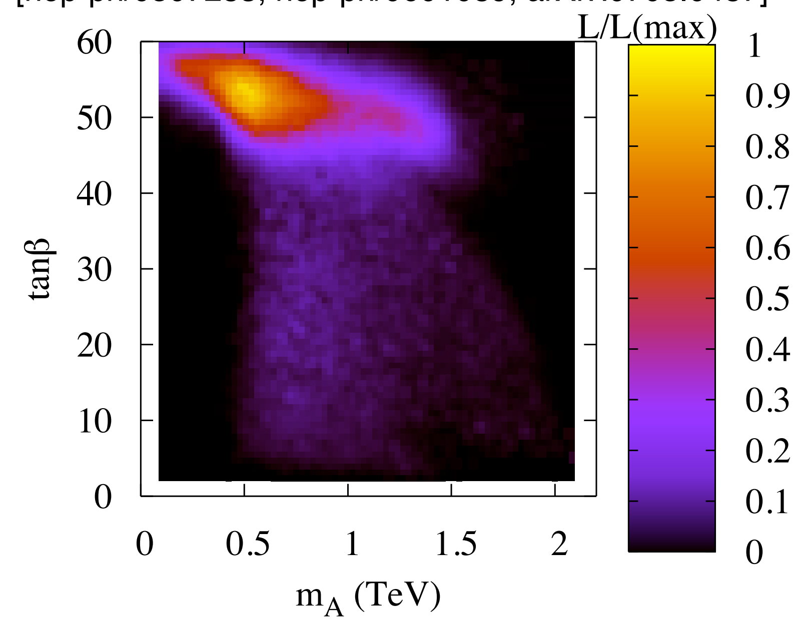
- KISMET publishing full likelihood maps
  http://users.hepforge.org/~allanach/benchmarks/kismet.html
- Ideally, inputs to FITTINO and SFITTER would be a true likelihood map

[arXiv:0710.2822]

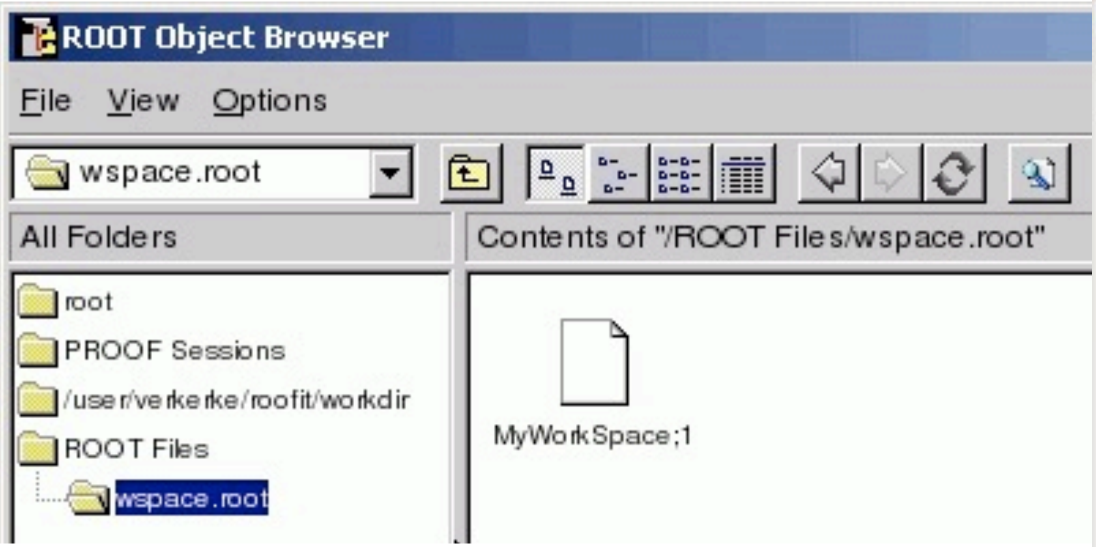




[hep-ph/0507283, hep-ph/0601089, arXiv:0705.0487]



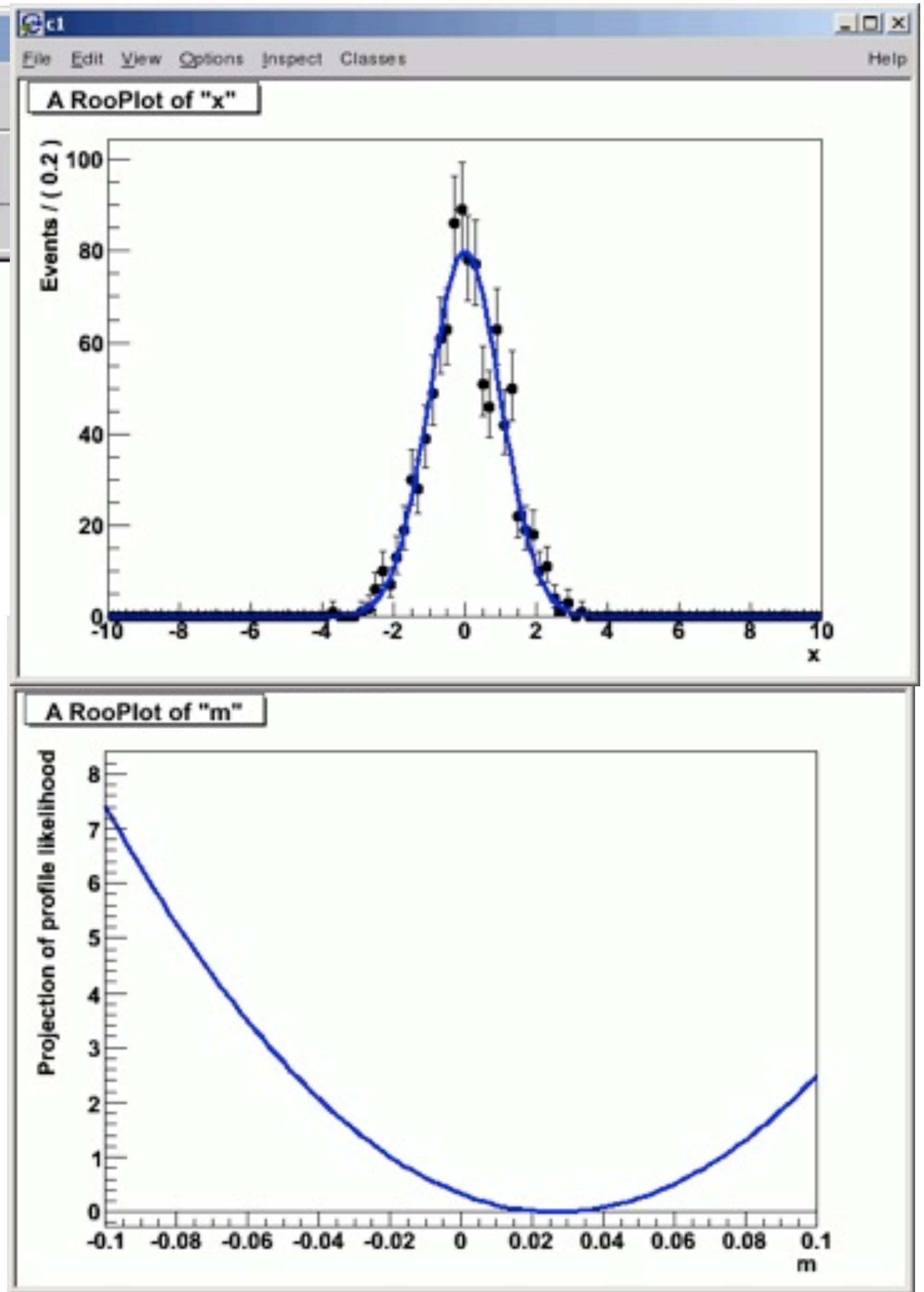
Example of Digital Publishing in RooFit/RooStats



RooFit's Workspace now provides the ability to save in a ROOT file the full likelihood model and the minimal data necessary to reproduce likelihood function.

It's generic, we decide how to parametrize the model.

Need this for combinations, we should publish them to some repository!



Combining Results: An Example

A combination example

- Combining 'ATLAS' and 'CMS' result from persisted workspaces

Read ATLAS workspace
```
TFile* f = new TFile("atlas.root") ;
RooWorkspace *atlas = f->Get("atlas") ;
```

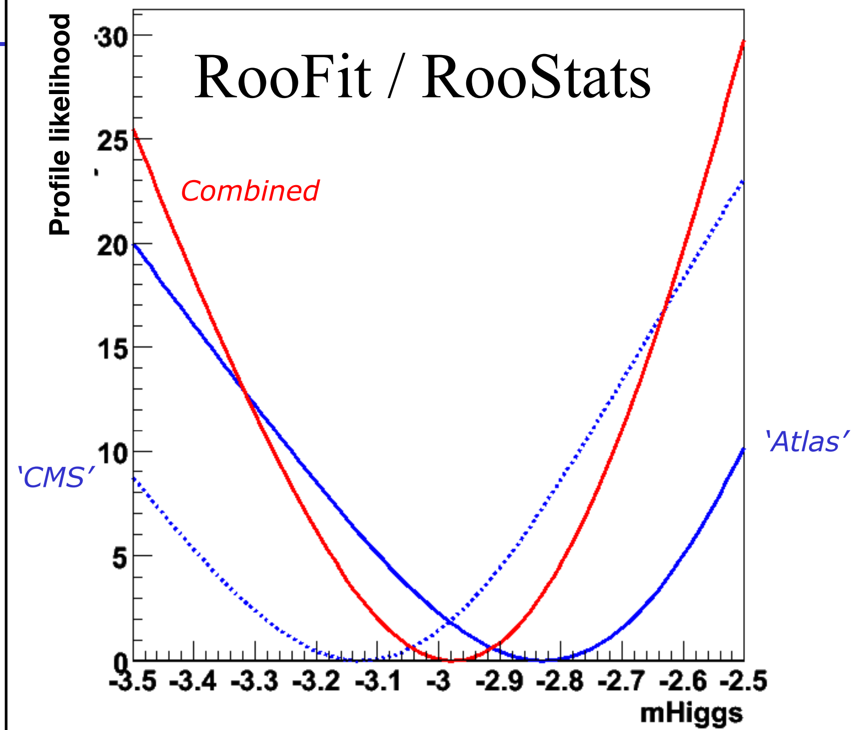
Read CMS workspace
```
TFile* f = new TFile("cms.root") ;
RooWorkspace *cms = f->Get("cms") ;
```

Construct combined LH
```
RooAddition nllCombi("nllCombi","nll CMS&ATLAS",
                     RooArgSet(*cms->function("nll"),*atlas->function("nll"))) ;
```

Construct profile LH in mHiggs
```
RooProfileLL pllCombi("pllCombi","pll",nllCombi,*atlas->var("mHiggs")) ;
```

Plot Atlas,CMS, combined profile LH
```
RooPlot* mframe = atlas->var("mHiggs")->frame(-3.5,-2.5) ;
atlas->function("nll")->plotOn(mframe)) ;
cms->function("nll")->plotOn(mframe),LineStyle(kDashed)) ;
pllCombi.plotOn(mframe,LineColor(kRed)) ;
mframe->Draw() ; // result on next slide
```

Wouter Verkerke, NIKHEF

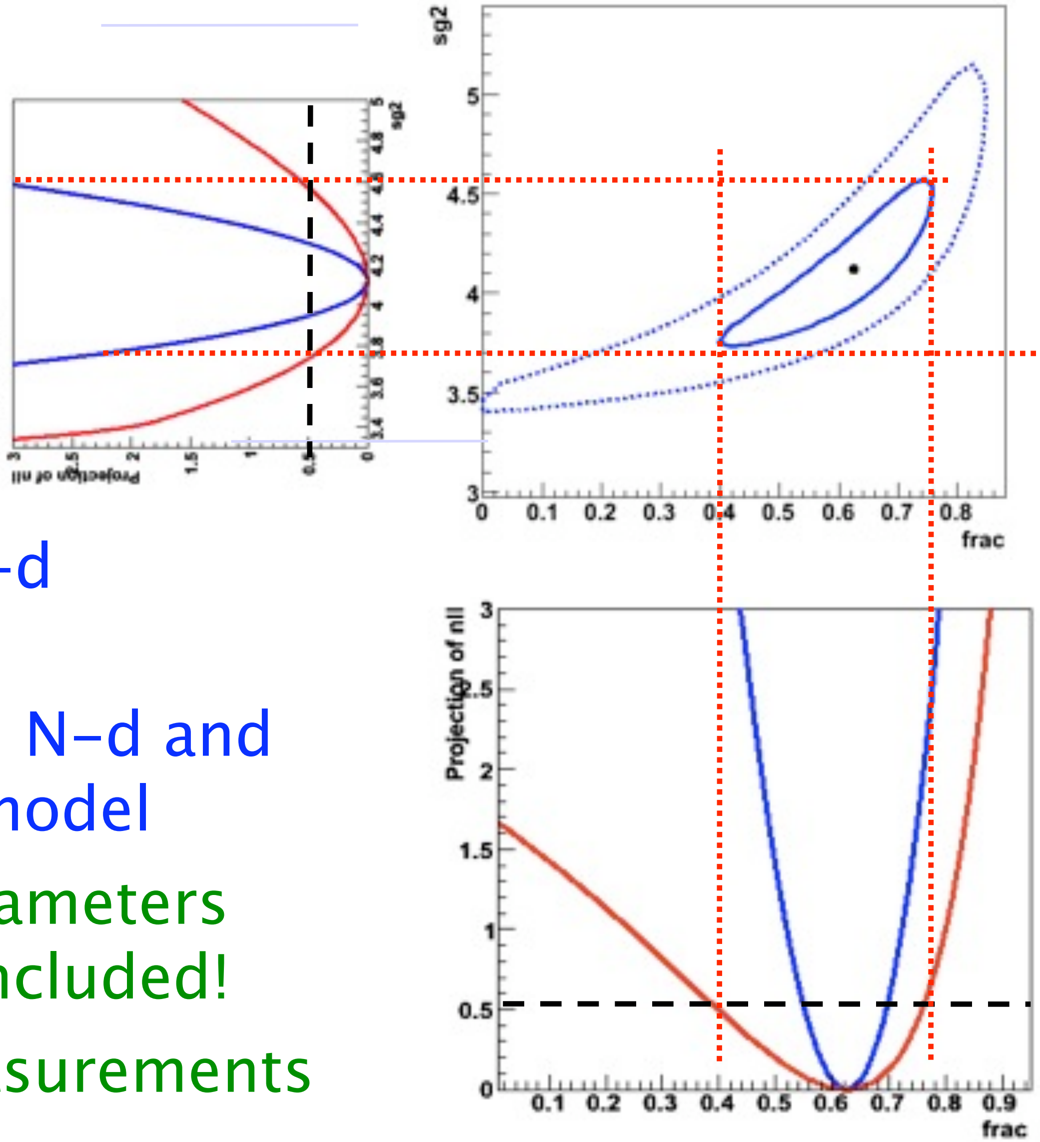


By using the workspace, it is easy to share results, ideal for combinations.

Example above shows opening an 'atlas' and 'cms' workspace, and performing a combined fit to a common parameter with profile likelihood.

Extracting Contours from these results

The workspace can represent arbitrary models with many parameters of interest and many nuisance parameters



- One can plot 2-d contours, 1-d likelihood functions
- One can evaluate likelihood in N-d and use to evaluate a theoretical model
- If the model has nuisance parameters for systematics, they will be included!
- Easy to combine multiple measurements

Taken from Wouter Verkerke, NIKHEF