Modern st~~atistics~~ analysis tools -- how can we ~~gain some~~ apply our intuition

Daniel Whiteson, UC Irvine
&
Kyle Cranmer, NYU

- State of the art
  - Matrix element techniques for likelihood ratios

- List of issues/advances
  - Issue 1: Overspecific matrix-elements
  - Issue 2: “Stone-age” matrix-elements
  - Issue 3: Reliance on simulation
  - Issue 4: Statistical applications

Want to compare two hypotheses: SM (null) SM+X (new physics)

Want to compare two hypotheses: SM (null) SM+X (new physics)

Best approach is the likelihood ratio:

$$LR(\text{event } z) = \frac{P(\ z \mid SM)}{P(\ z \mid SM+X)}$$

z is vector of measured quantities (leptons, jets, etc)

The probabilities are not trivial to calculate

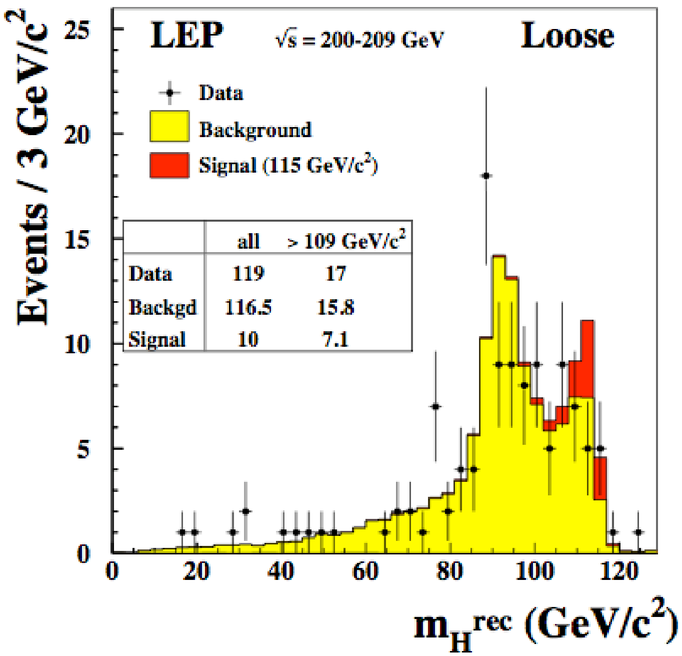
$$\frac{P(\ z \mid SM)}{P(\ z \mid SM+X)}$$

Traditionally,

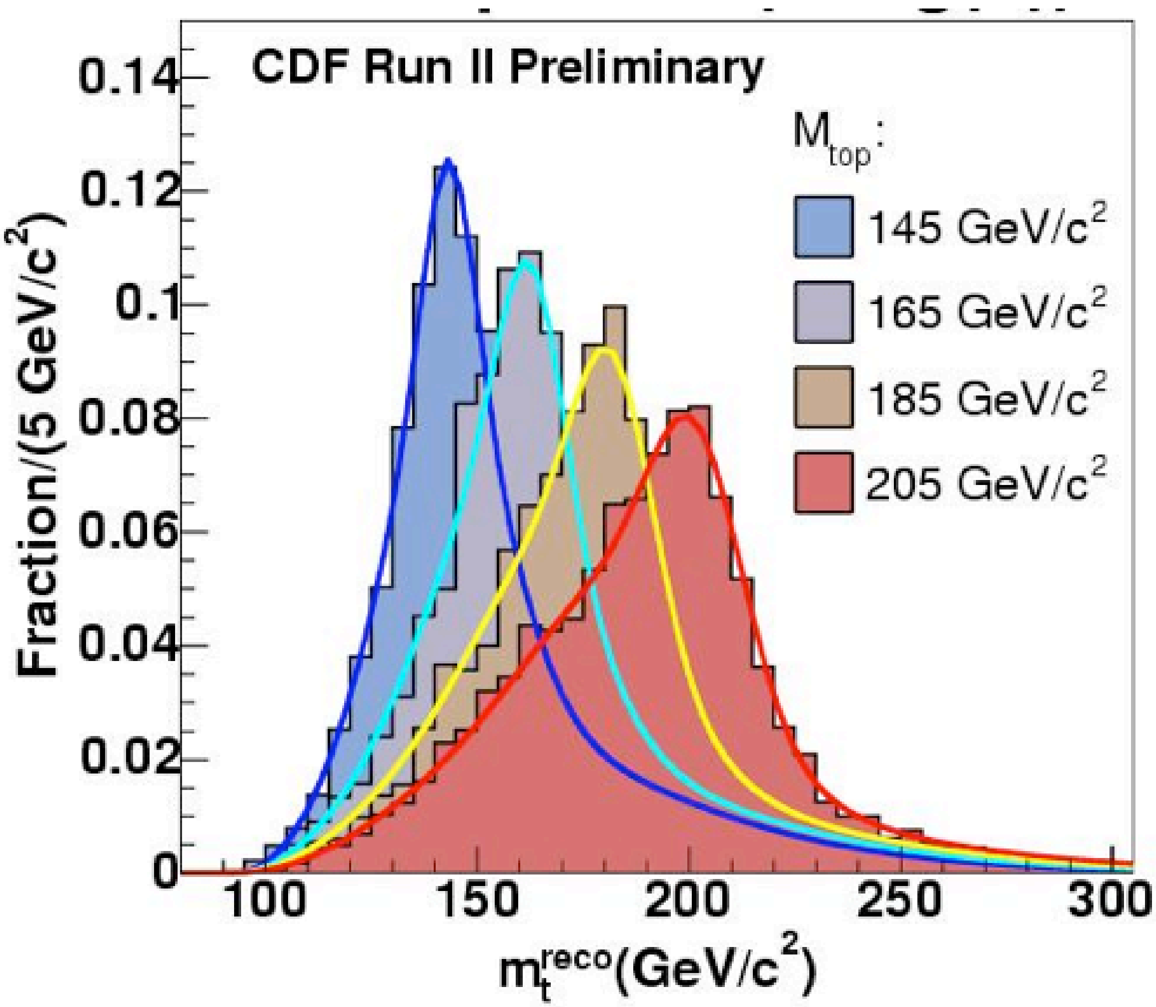
(1) choose some distinguishing variable

(2) Simulate events, fill histograms for both hypotheses

Note:Two events with same m_{reco} have identical effect on analysis



Likelihood parameterization. Curse of dimensionality makes it difficult to parametrize in more than 1 or 2 dimensions



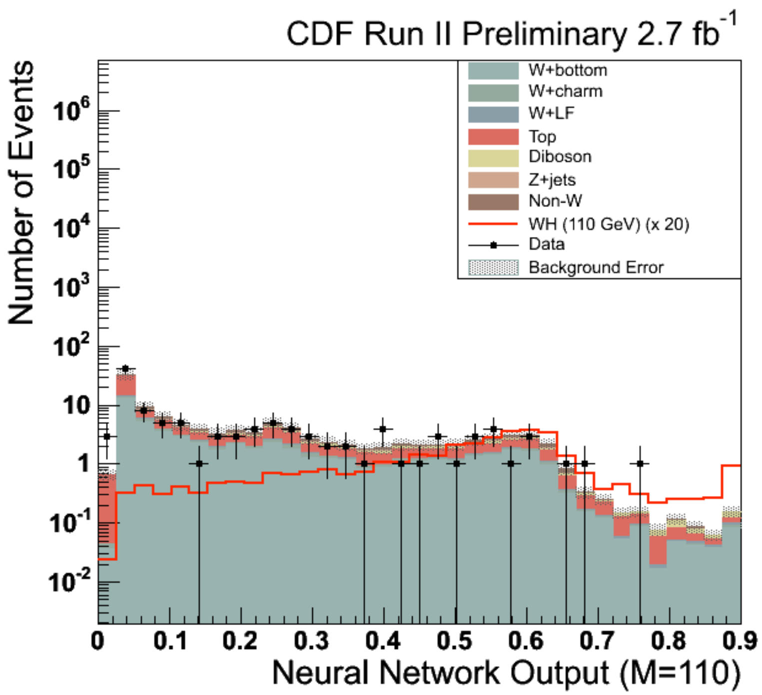
The probabilities are not trivial to calculate

$$\frac{P(\ z\ |\ SM)}{P(\ z\ |\ SM+X)}$$

Advanced,

(1) Use NN/BDT/KDE to reduce many dimensions down to 1

(2) Simulate events, fill histograms for both hypotheses



Matrix-element likelihood:

Calculate probability directly

$$P(\text{event } z \mid SM) = P(z \mid \text{process } A) + P(z \mid \text{process } B) + \ldots.$$

where

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

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

$f_{TF}(y,z)$

The transfer function takes us from
parton(y) to detector-level(z)

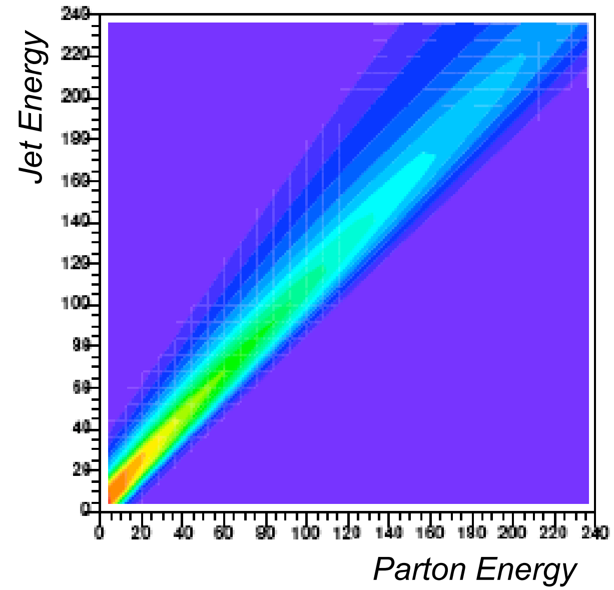
Use a parametrized description

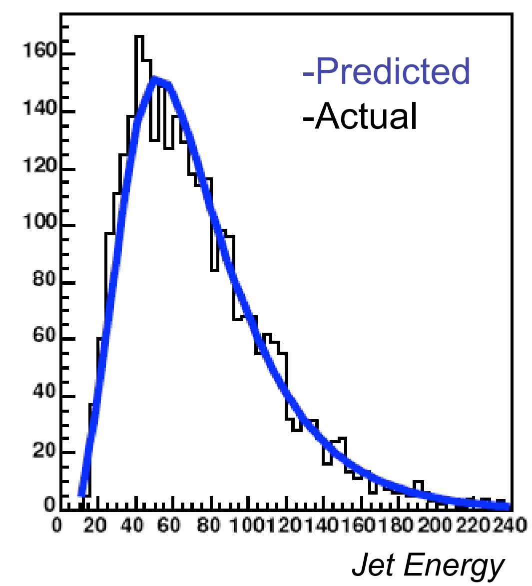
- Angles are perfectly measured (jets, leptons)
- Energy response can be parametrized

Equivalent to parametric detector simulation

-Can never be as detailed as full simulation
-Retains connection to physics intuition

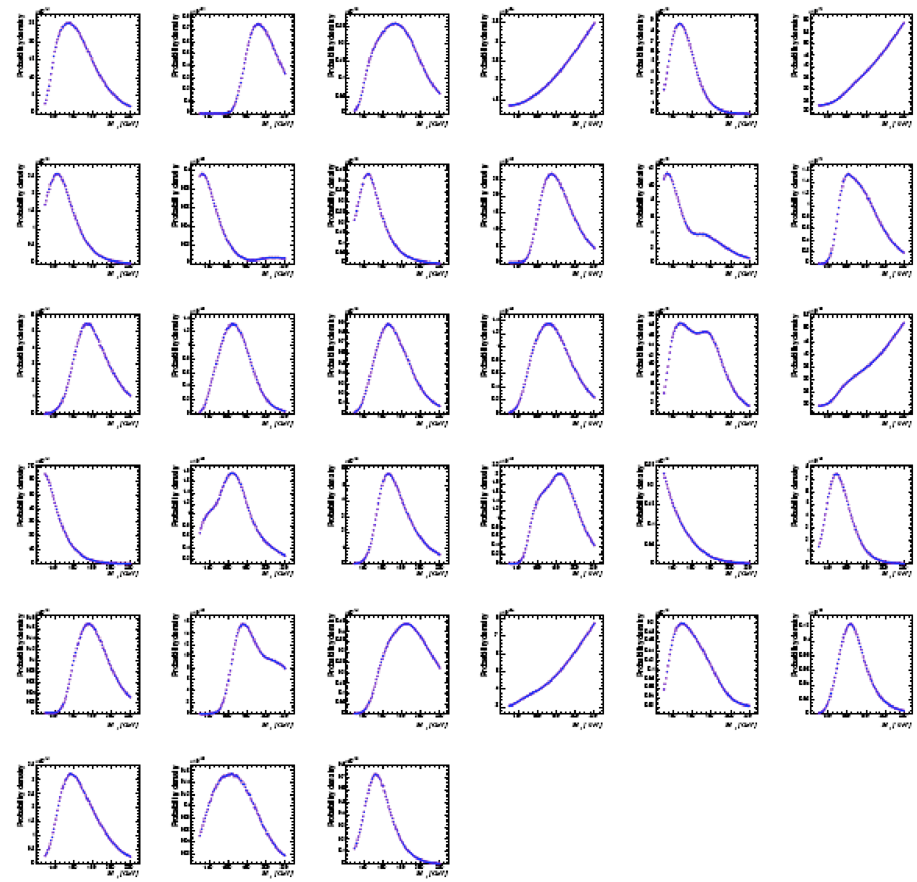
In contrast to reco-mass templates
-Less/no intuition in parametrization
-Disconnect from physics knowledge





Previous applications

Consider continuous parameter of SM (M_{top})

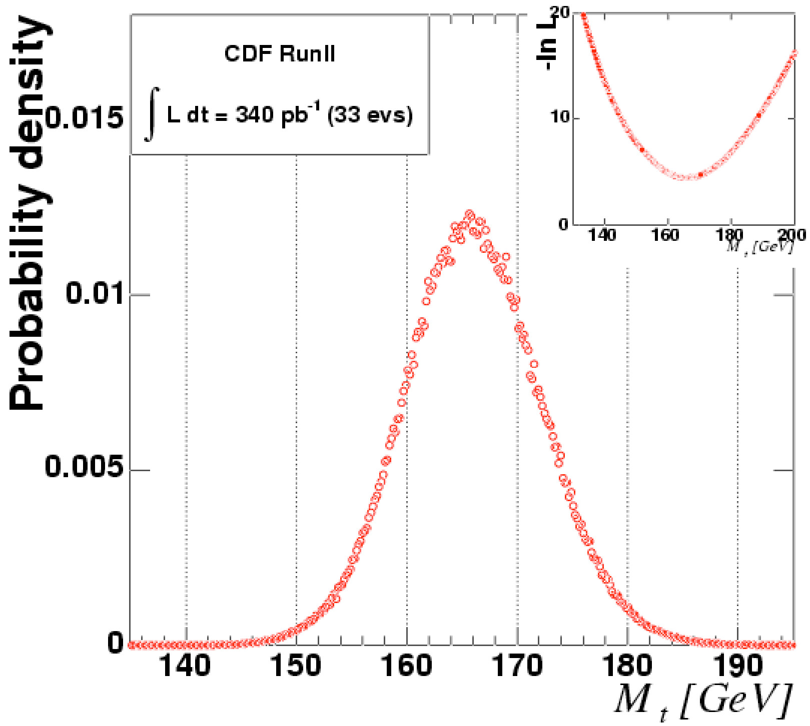


$$P(\text{ CDF } | \; M_t) = P(\text{event } z_1 \; | \; M_t) \;\; x \;\; P(z_2 \; | \; M_t) \;\; x \; P(z_3 \; | \; M_t) \; ....$$

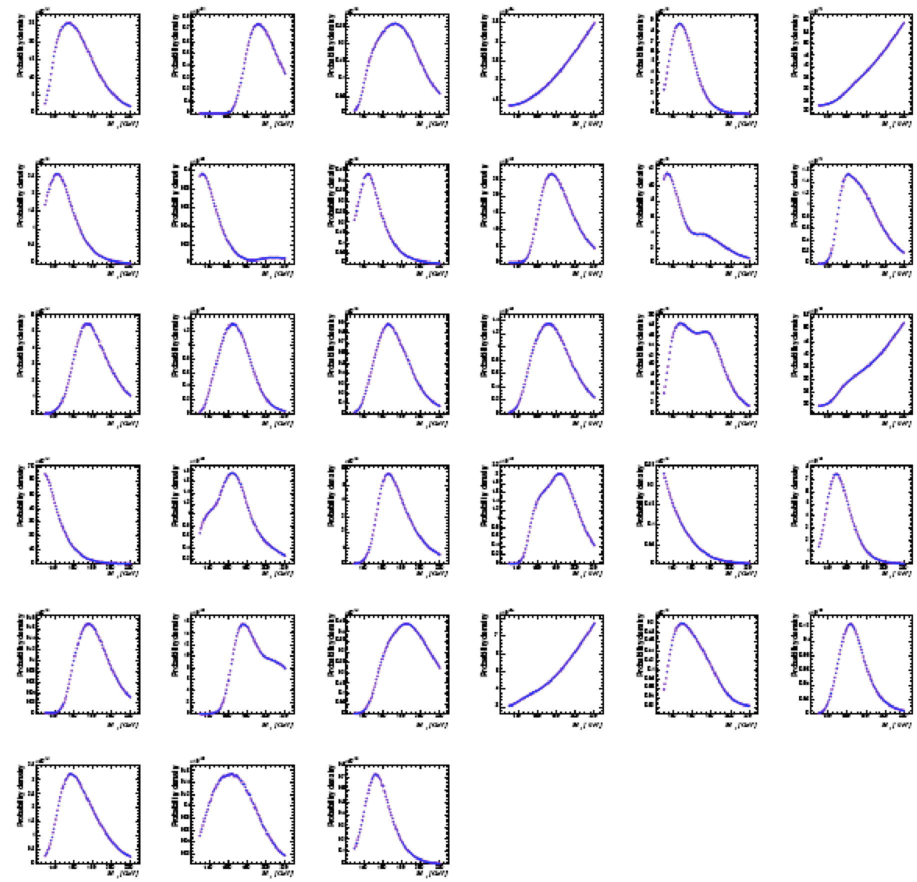
Note that P is a function of M

Previous applications

Consider continuous parameter of SM (M_{top})



=

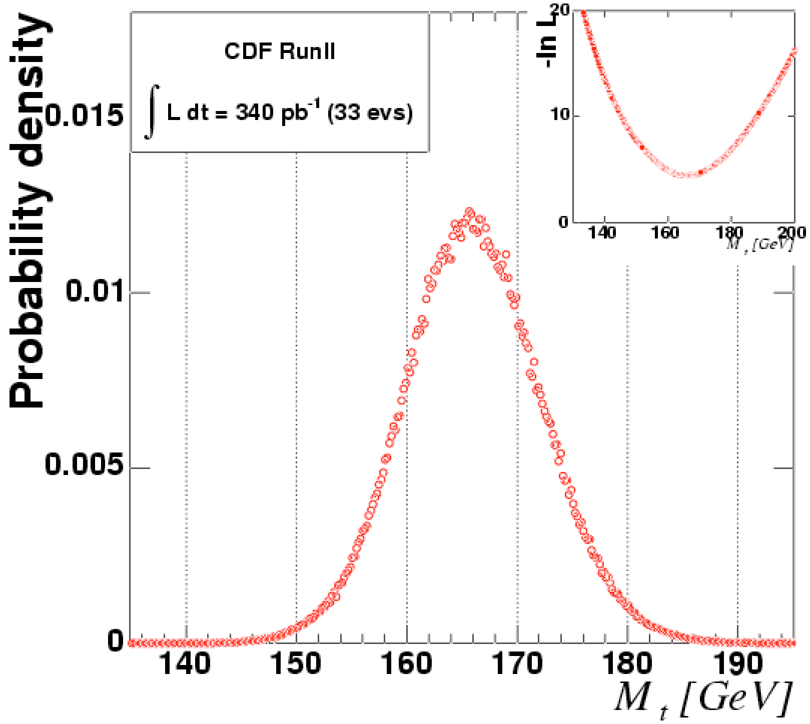


$$P(\text{ CDF } | \; \color{red}{M_t}) = P(\text{event } z_1 \; | \; \color{red}{M_t}) \;\; x \;\; P(z_2 \; | \; \color{red}{M_t}) \;\; x \; P(z_3 \; | \; \color{red}{M_t}) \; ....$$

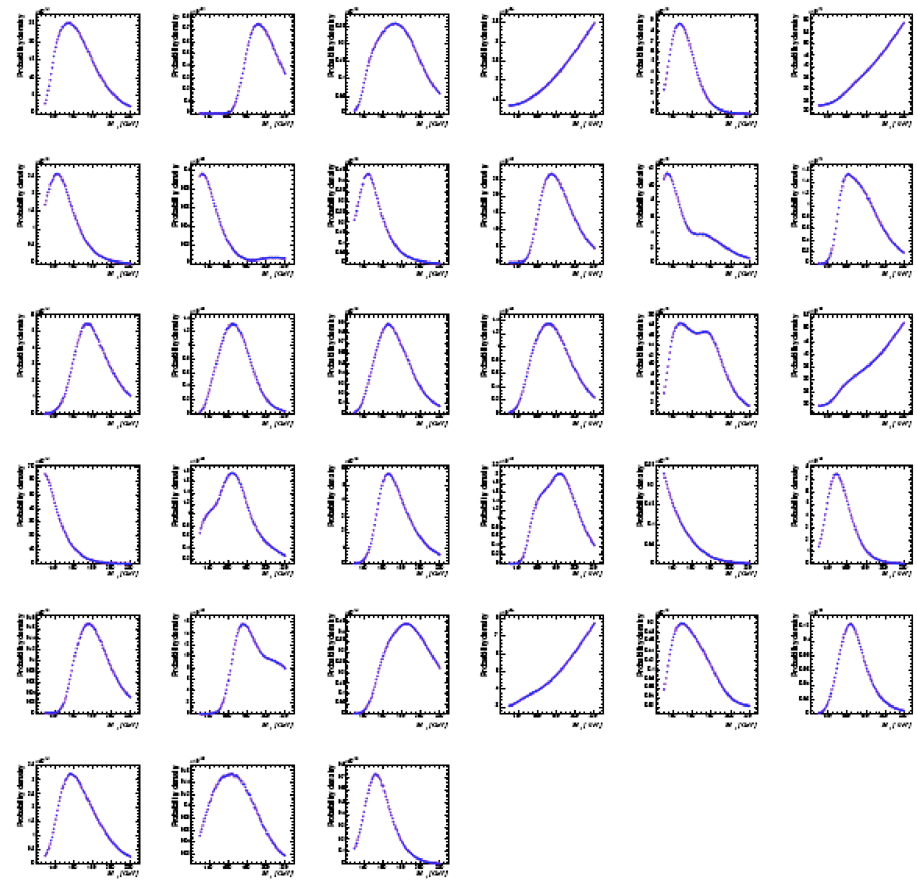
Note that P is a function of M

Previous applications

Consider continuous parameter of SM (M_{top})



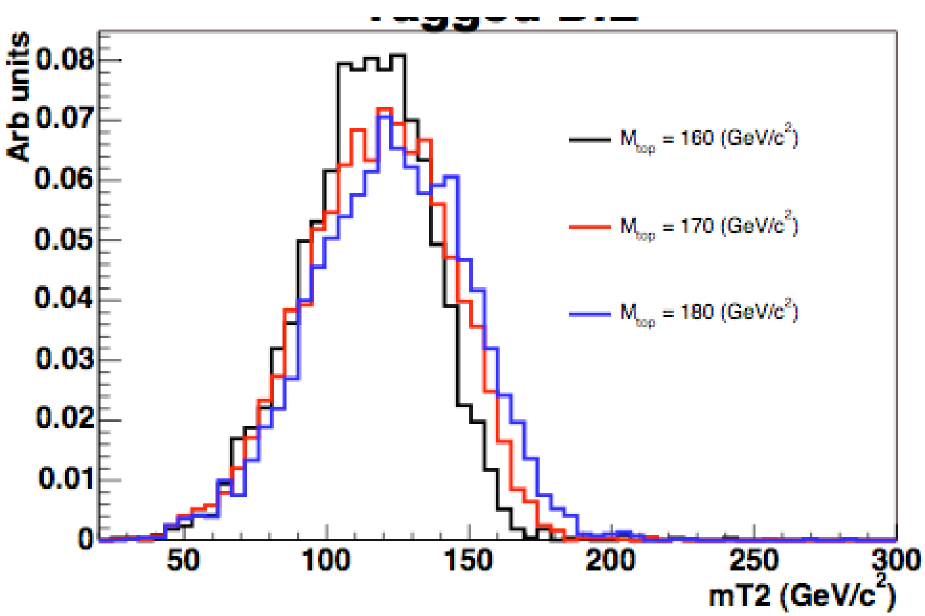
=



Note: each event's likelihood has a different dependence on top mass
events contribute more than just location of peak
allows well-measured events to have stronger impact

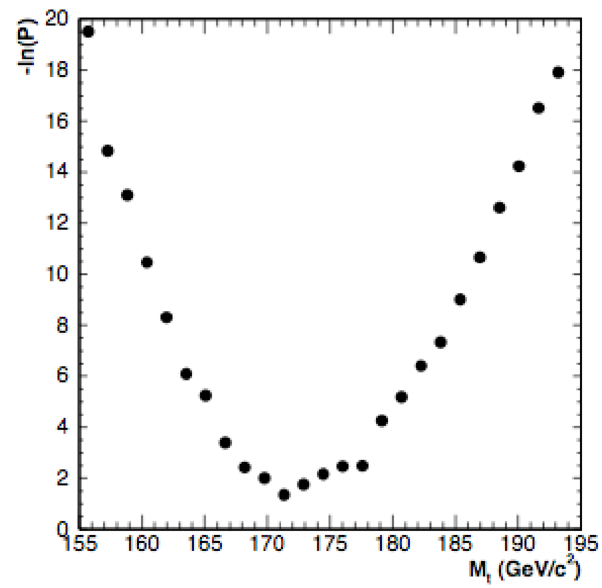
How does this compare to other techniques?

I've heard rumors that the "MT2" technique has been applied to CDF data and gives a 20% improvement



MT2 (3/fb)

167.9 +- 4.5(stat) +- 2.8(syst)



ME method (2/fb)

171.2 +- 2.7(stat) +- 2.9(syst)

Issue #1

The matrix elements are too specific

or

"I prefer to just use kinematics"

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

Method allows any matrix-element from

1) OSET description

2) Effective Lagrangian of your choice

3) SM ttbar

4) SSM (not MSSM), UED, etc

Fair point:

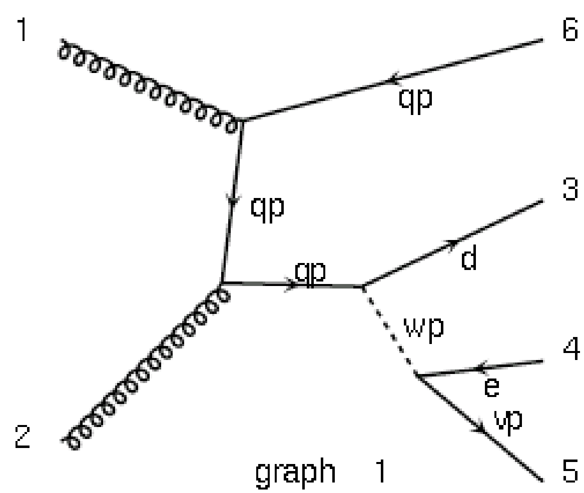
This does not allow for easy generalization of common features across similar processes: i.e. incomplete specification of kinematics.

Squark pairs -> 2 leptons, 2 jets, ME_T

Effective Lagrangian: squark, chargino,sneutrino/LSP. (Choose SUSY-like spins)

Thanks to Johan, Tilman, Frank, Mihoko...

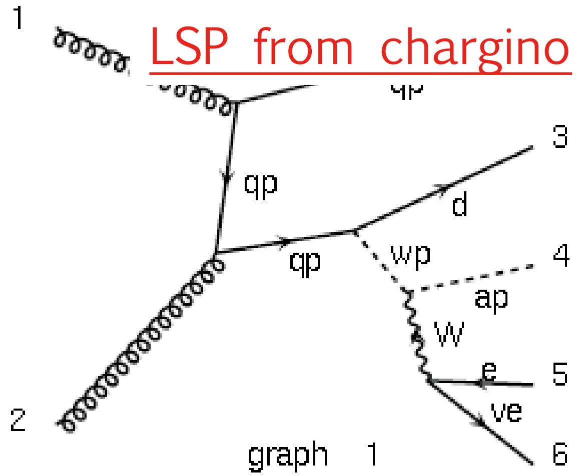
LSP from/as sneutrino



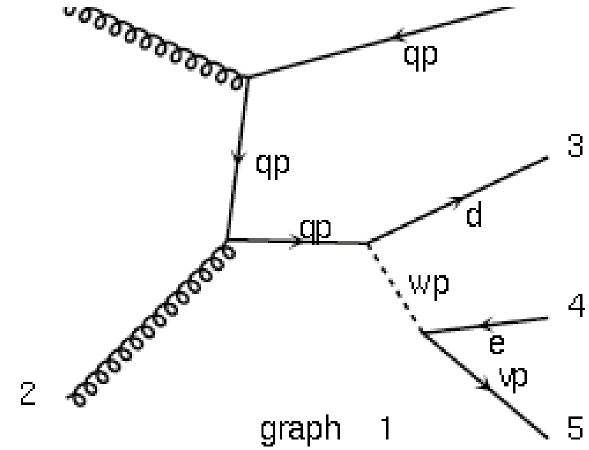
Squark pairs -> 2 leptons, 2 jets, ME_T

Effective Lagrangian: squark, chargino,sneutrino/LSP. (Choose SUSY-like spins)

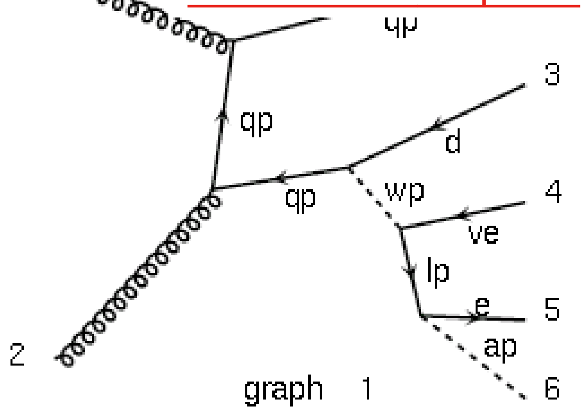
LSP from chargino



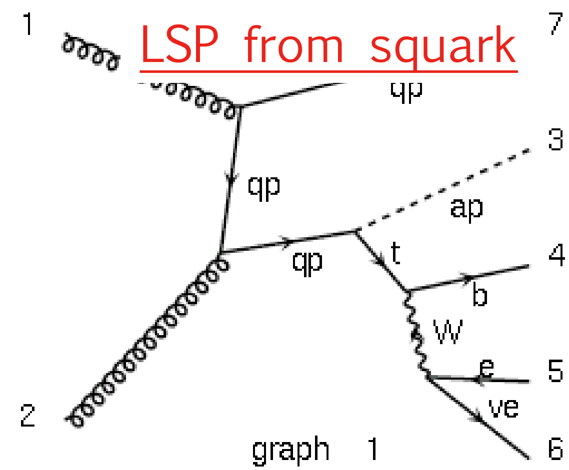
LSP from/as sneutrino



LSP from slepton



LSP from squark

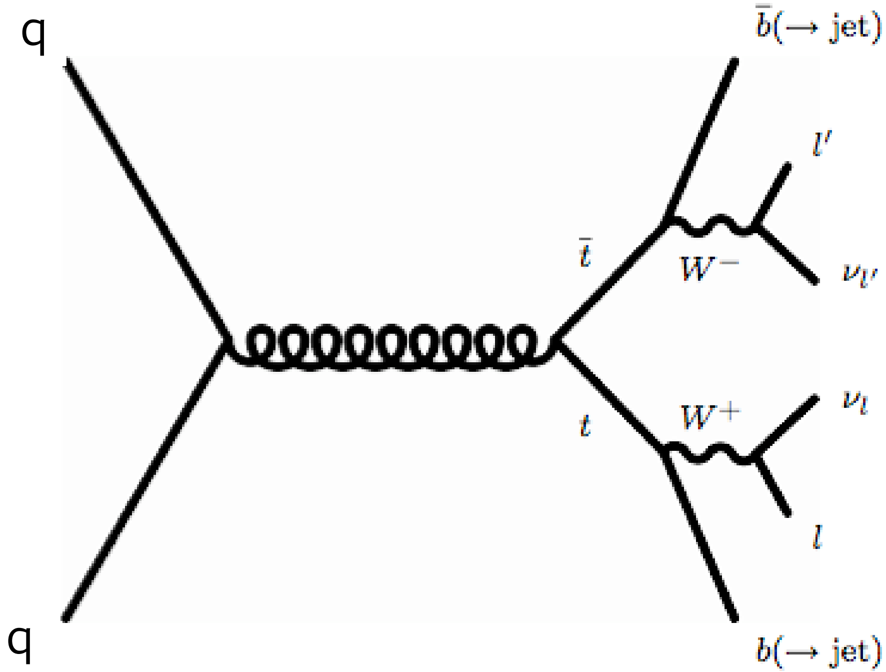


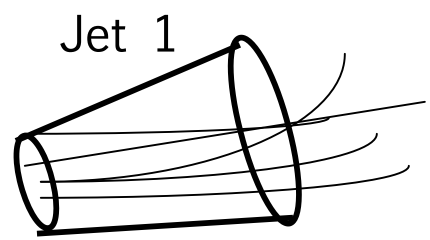
Issue #2

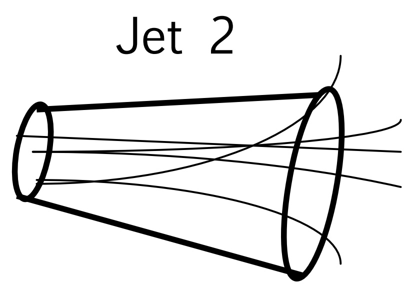
The matrix elements are from the stone age

or

"Why did we spend 10 years developing ME+PS machinery if you're going to just use the Z+2p ME?"





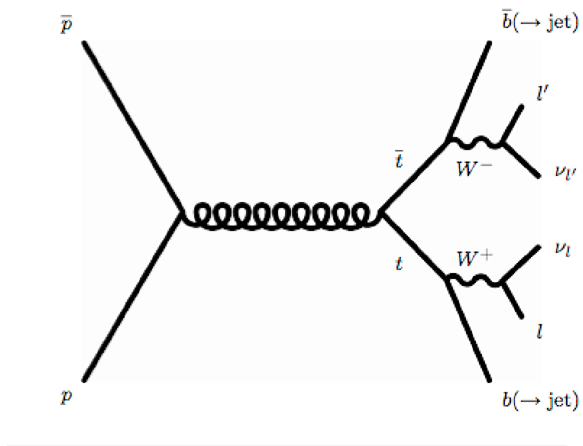


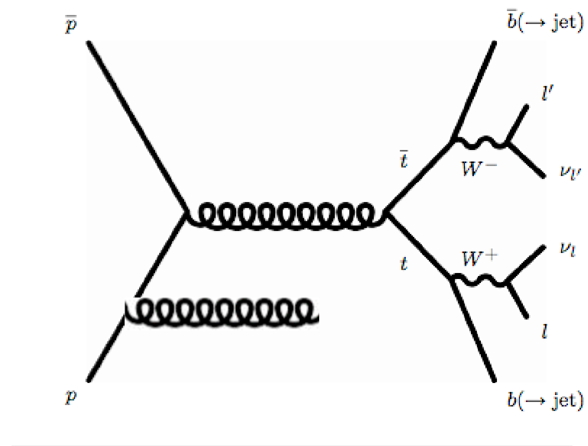
Very simplistic:

Each object connected to a final-state particle

No allowance for radiation*

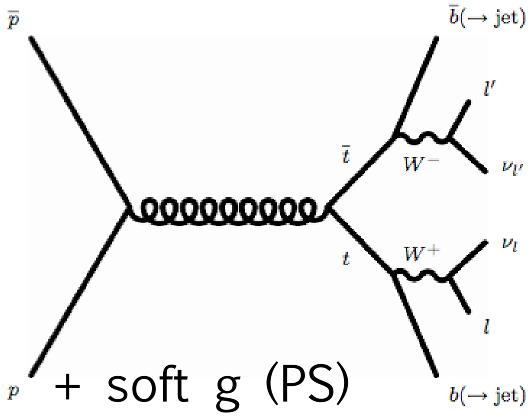
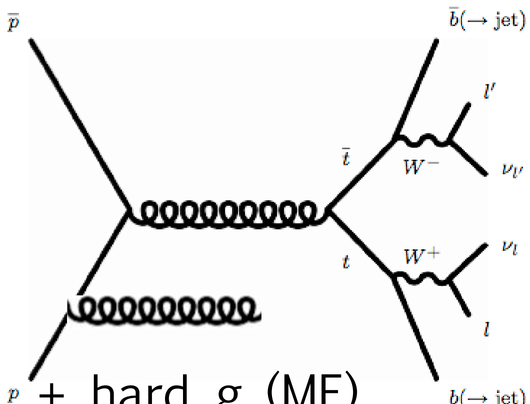
*intregrating over p_T of the hard process has been done, but not rigorously





New prescription

- Use X, X+1p, X+2p MEs
- Run parton-shower code on external legs to generate soft P_T
- Cluster particle jets and match (piggyback on matching technology!)
- Transfer function now only describes detector response

	2 jet event	3 jet event
 + soft g (PS)	PS gives description of p_T of tt system	Deweighted by prob for soft PS jet to be reconstructed over jet threshold TF(10GeV shower→15 GeV jet)
 + hard g (ME) + soft g (PS)	Deweighted by prob to lose hard jet TF(p→no jet) Integrate over 4v of hard gluon	PS gives description of p_T of tt+g system

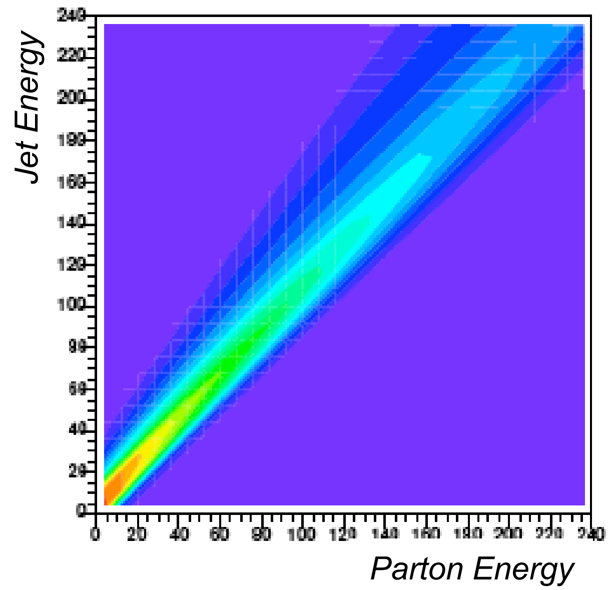
Issue #3

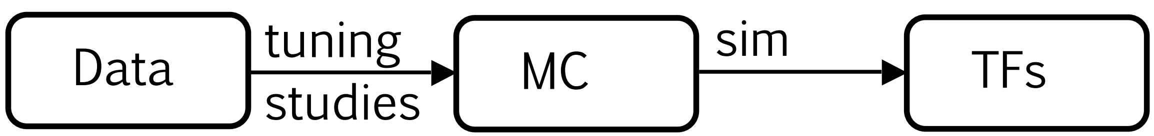
Reliance on simulation for transfer functions

or

"Tevatron experiments only used these at the end of their runs, because they're too dependent on simulation to be used in early data."

Simulation free?



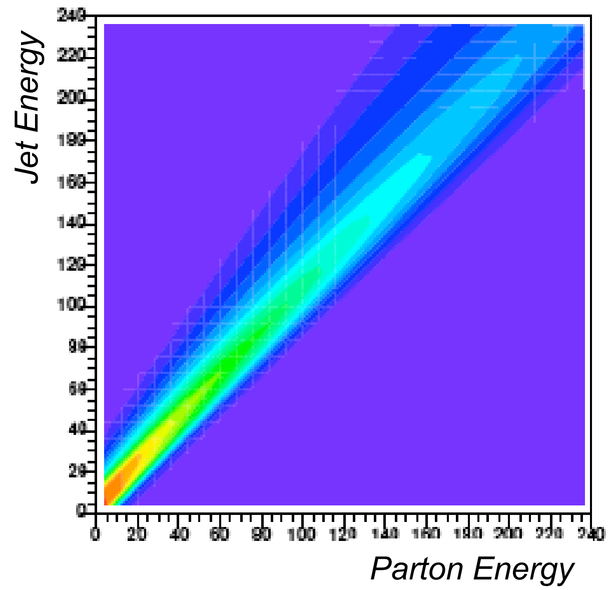


Major weakness

TFs derived from simulation

- Relies on simulation to be tuned
- This will take a long time
- requires large sample

Simulation free?

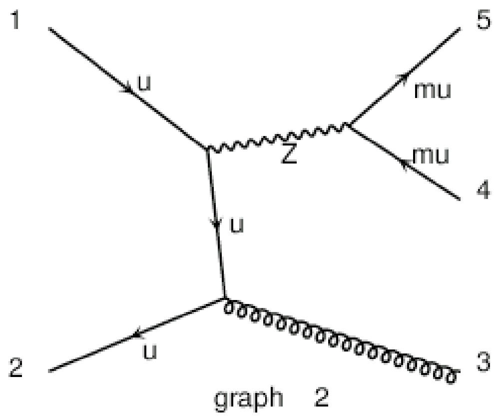


Is this necessary?

Samples used to tune jet response
are also powerful to determine TFs

Use: Z+jets, gamma+jets, semileptonic ttbar

TFs from data



Fit TFs from data samples

Maximize

$$\prod_{x} P(x|TF) = \int dp \; |\mathcal{M}|^2 \; f_{TF}(x,p)$$

w.r.t TF parameters

Advantages

Same strategy as MC tuning

- find sample which is clean, and sensitive to TFs

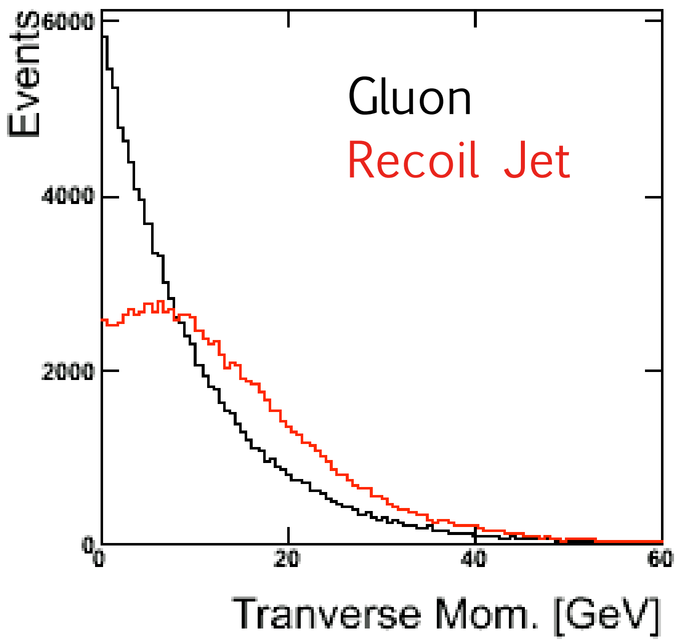
No reliance on simulation

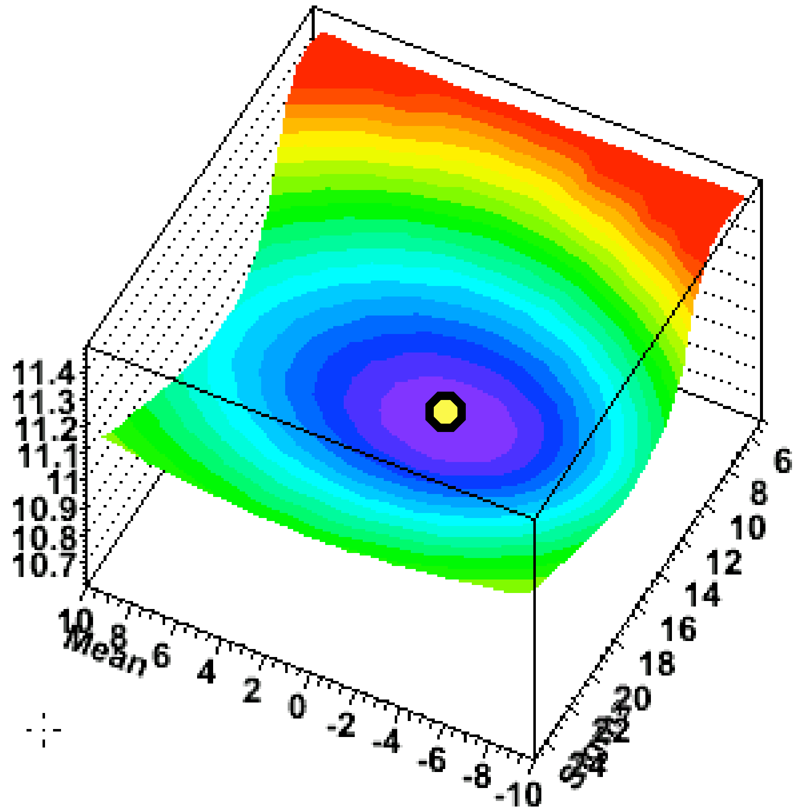
- this integral can be very fast (done analytically)

TFs naturally fit to give best description

- even if model is imperfect
- systematics can be extracted as well

First Try



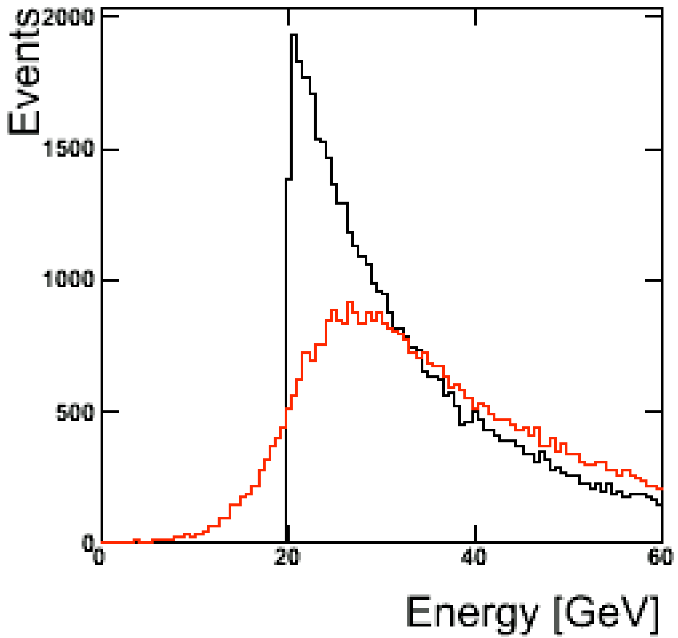


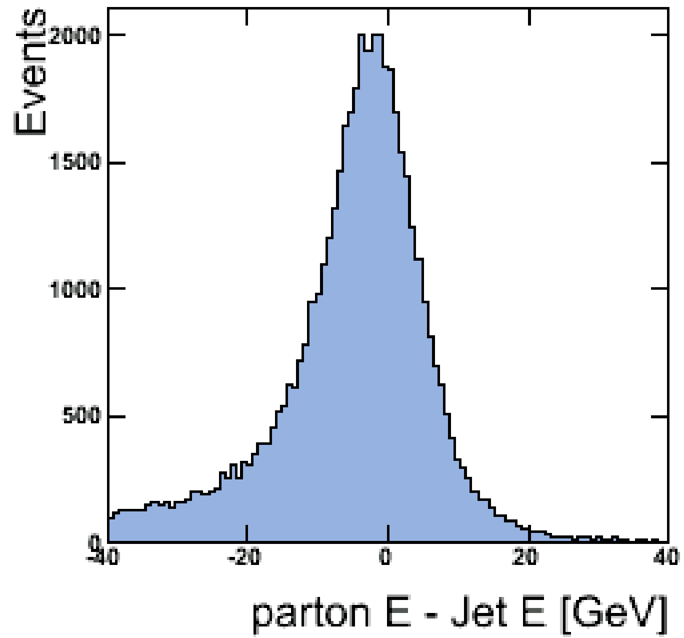
Toy example

Smear with 10-GeV width Gaussian

Extract parameters from smeared events

More realistic Transfer Functions





Smear partons with double Gaussian TF

2x5=10 parameters: $\mu_1,\mu_2,\sigma_1,\sigma_2,f$

[const and energy dep for each]

inspired by CDF transfer functions

Results

Events

Smear partons with TFs

Fit

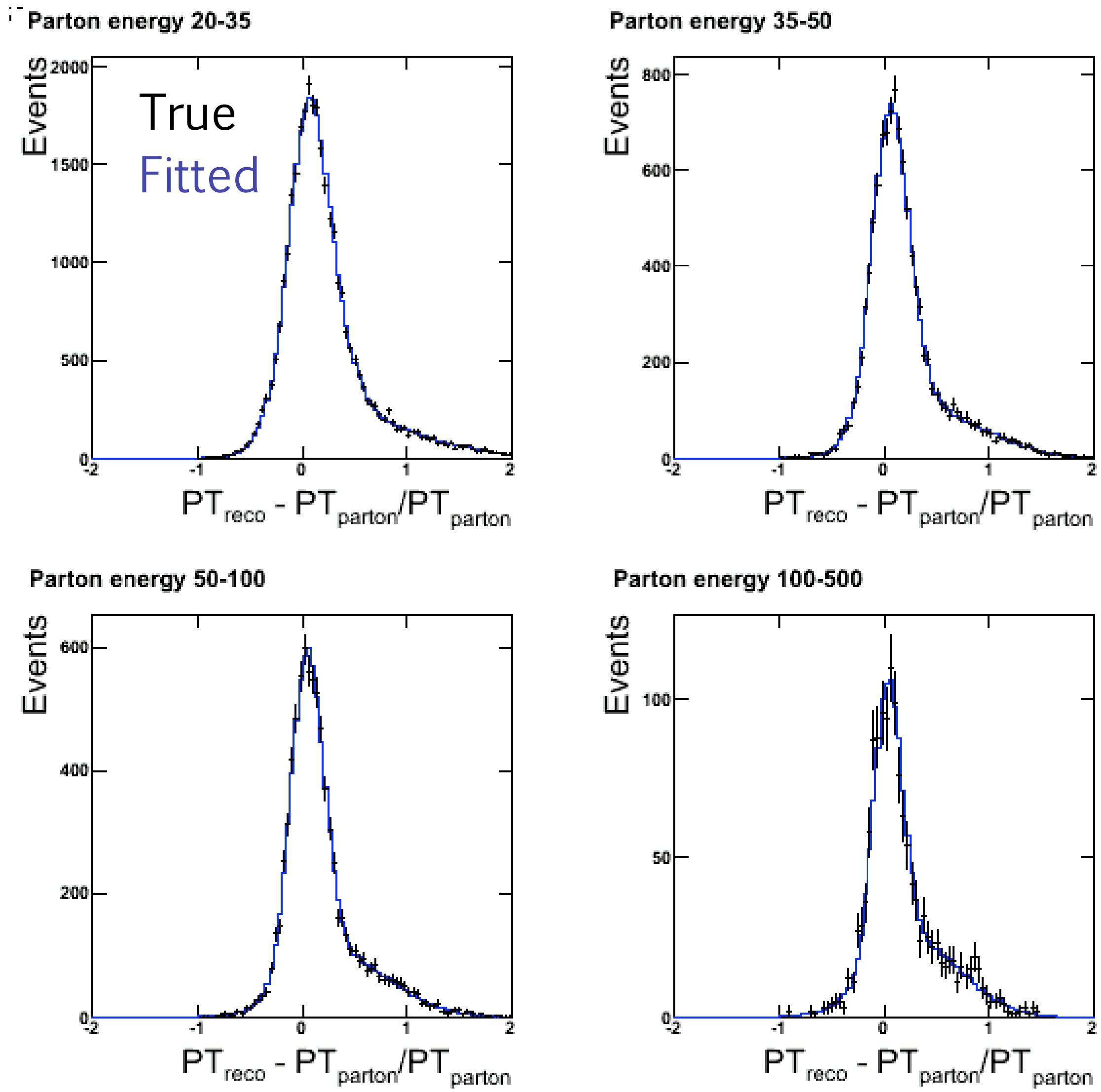
Minimize 10D space with Minuit

Check

Resmear partons with fitted parameters

Caveats

No backgrounds, no ISR, etc



Issue #4

Statistical applications

or

"Your talk was supposed to be on statistics..."

Searches Previous applications for searches

CDF Wh:

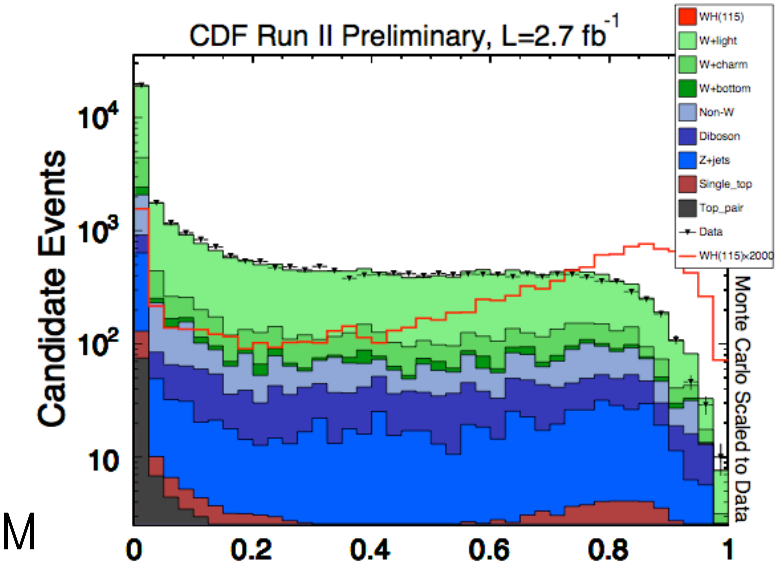
form discriminant for each event

$P_s / P_s + P_b$

Create templates and fit for best signal rate S_{best}

Specific template used for each mass M

P_s not maximized over M



$P_s / P_s + P_b$

New approach for searches

Use mass shape information

Calculate likelihood (L) for each event
as function of mass (M) and signal rate (S)

Define measured mass, signal as
point (M,S) which maximizes joint L

Use Feldman-Cousins to set limits

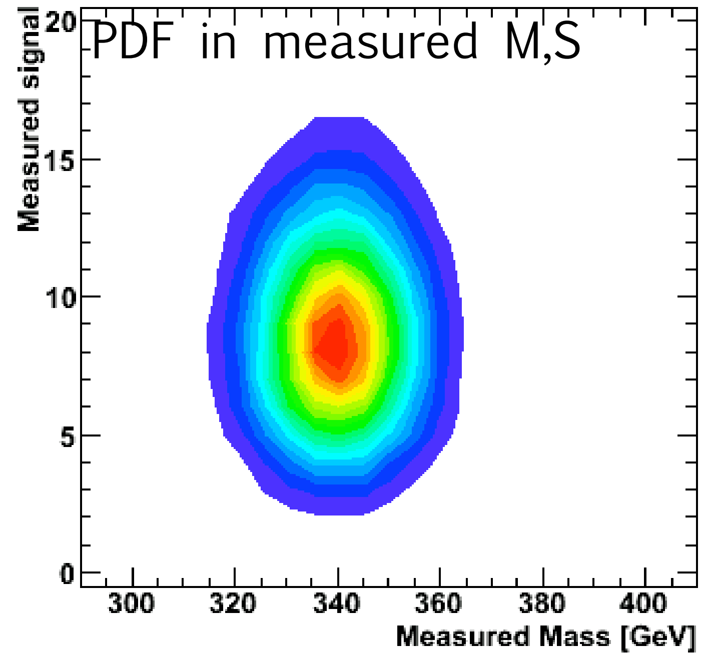
Advantages

P explicitly a function of M

Example: heavy t' search

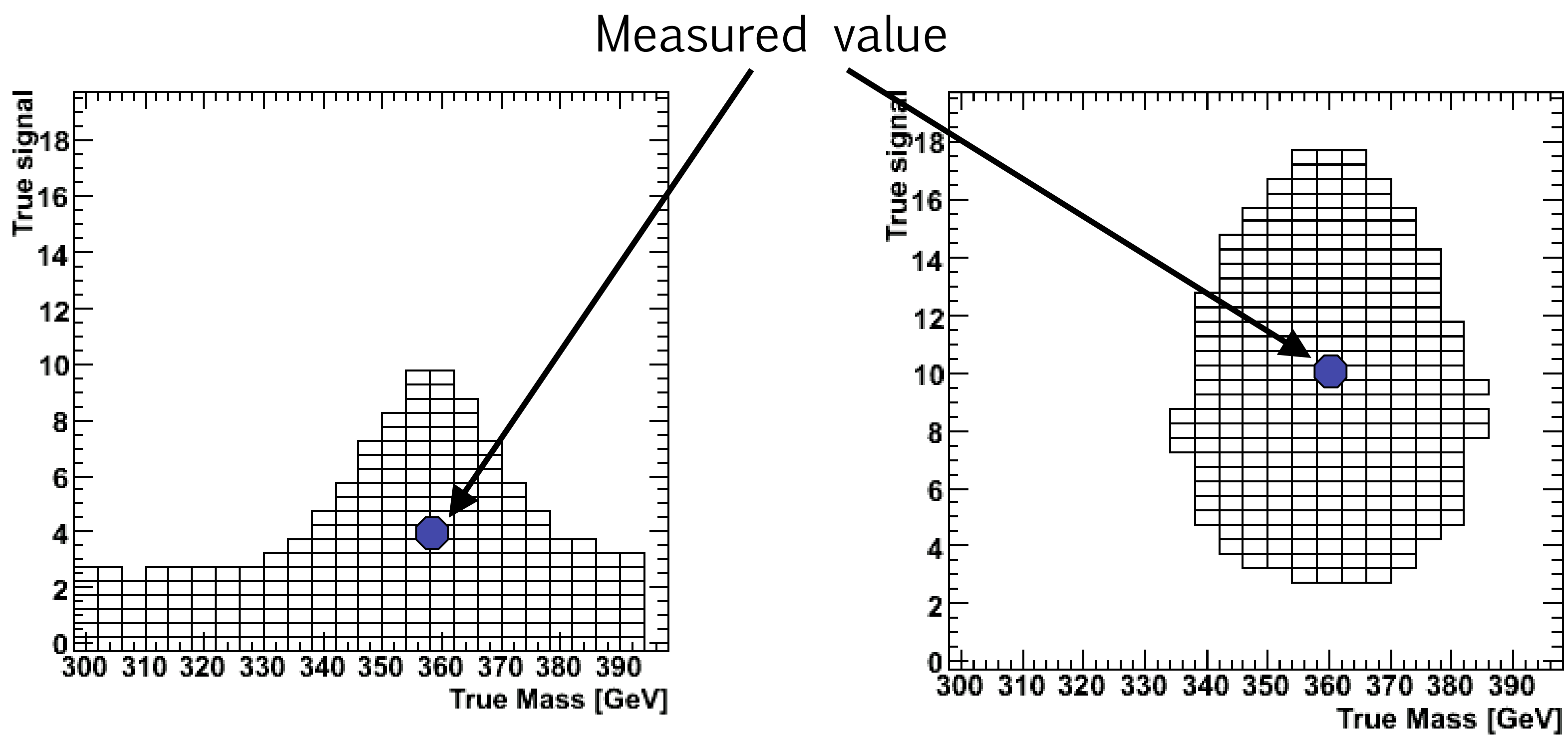
For true mass = 340

For true signal = 8



Contour

Region in true space for a measured value



Our 95% CL band is made from all the true points whose 95% band in measured space includes our measured value.

If region includes N_{signal} = 0, set a limit
If region excludes N_{signal} = 0, claim a discovery

Summary

Matrix-element-based likelihoods

-apply our physics intuition and a-priori knowledge

Technically have been primitive

- limited by CPU resources
- can/should apply same technology to ME-based MC generation as to ME-based likelihood calculations

Current weaknesses can be overcome

- Simulation dependence
- Statistical applications

Previous applications for searches

Non-optimal information use:

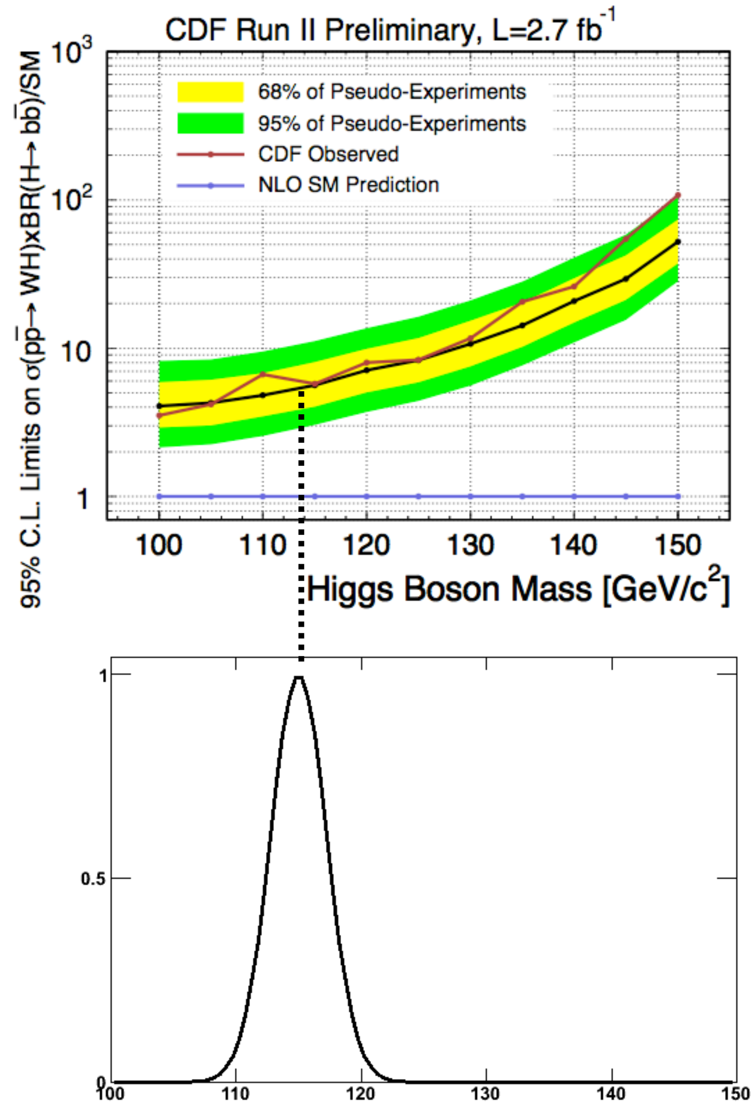
For analysis at M_h only evaluate $P_s(M_h)$ no maximization over M_h

Well measured events contribute as much as poorly measured events

Limit at each point done independently

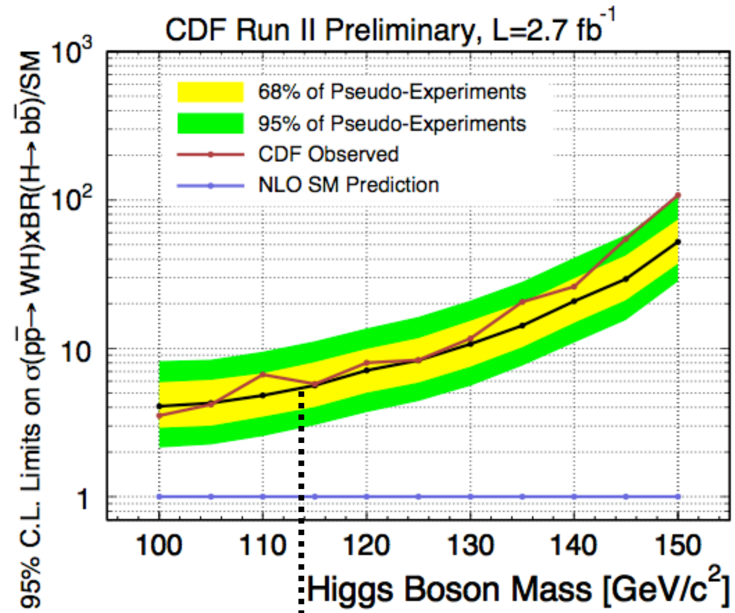
$P_s(110)$ does not affect limit at M_h=115

Wide events have same effect as narrow events



Single event $L(M_h)$

Previous applications for searches



Non-optimal information use:

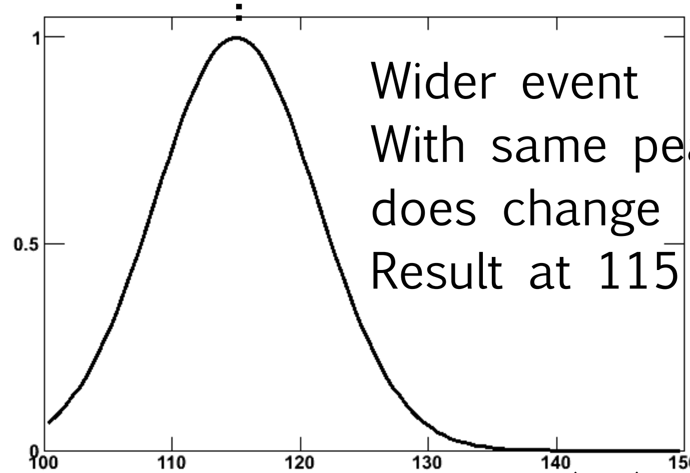
For analysis at M_h only evaluate $P_s(M_h)$
no maximization over M_h

Well measured events contribute as much as poorly measured events

Limit at each point done independently

$P_s(110)$ does not affect limit at M_h=115

Wide events have same effect as narrow events



Wider event
With same peak
does change
Result at 115

Single event $L(M_h)$