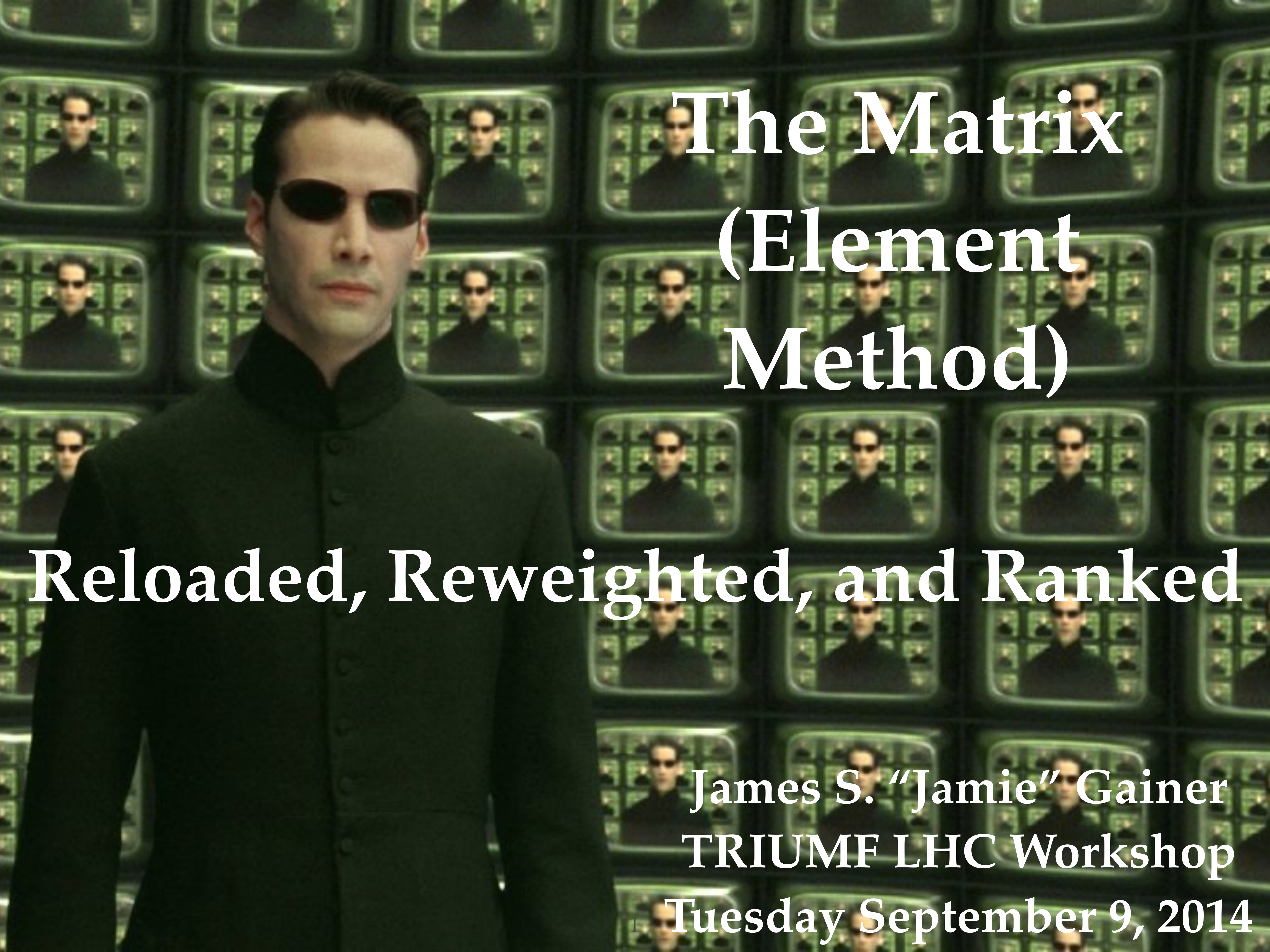


The Matrix (Element Method)

Reloaded, Reweighted, and Ranked

James S. "Jamie" Gainer

TRIUMF LHC Workshop

Tuesday September 9, 2014

What is the **best** we can do in a given analysis?

The Goal of HEP

- To Discover New Physics

The Goal of HEP

- To Discover New Physics

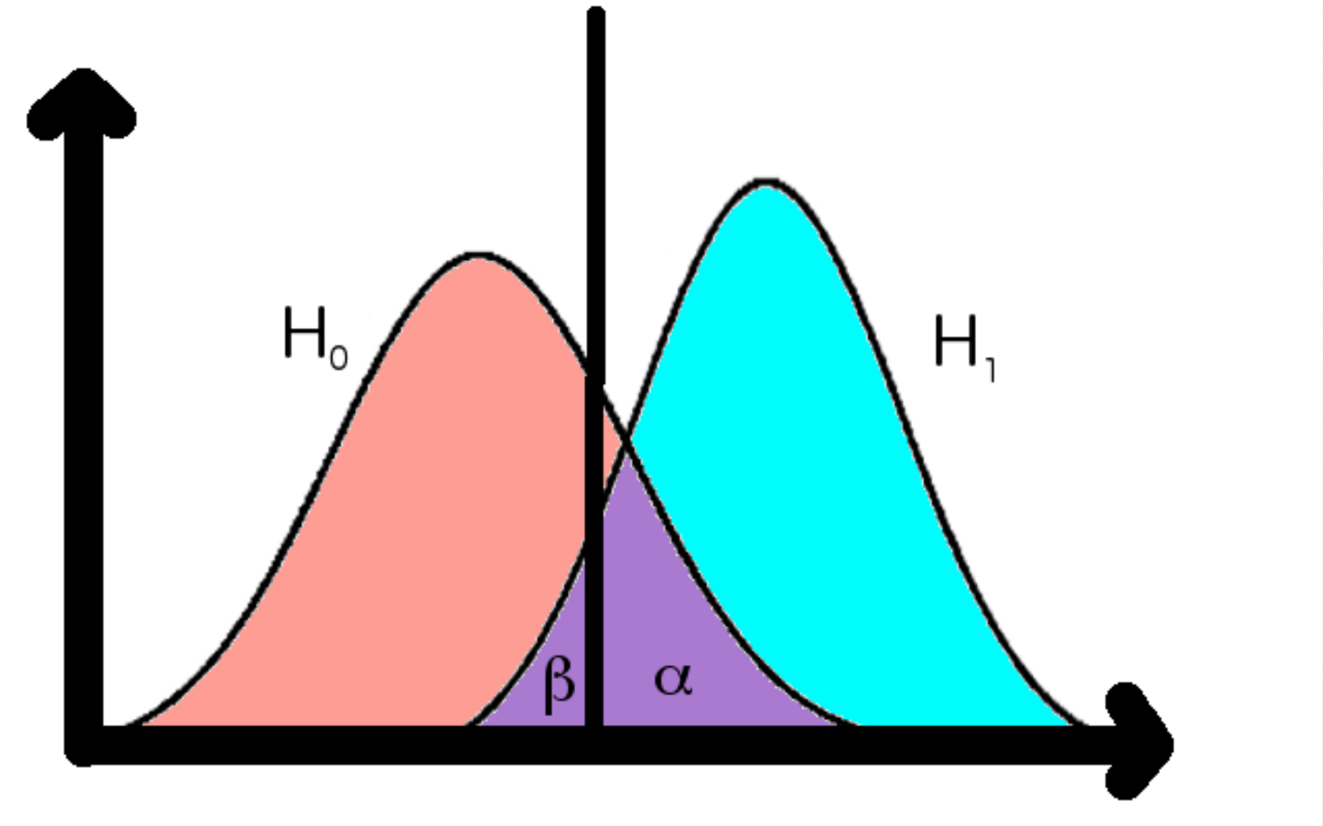
  Especially if it's really there.

The Goal of HEP

- To Discover New Physics
  Especially if it's really there.

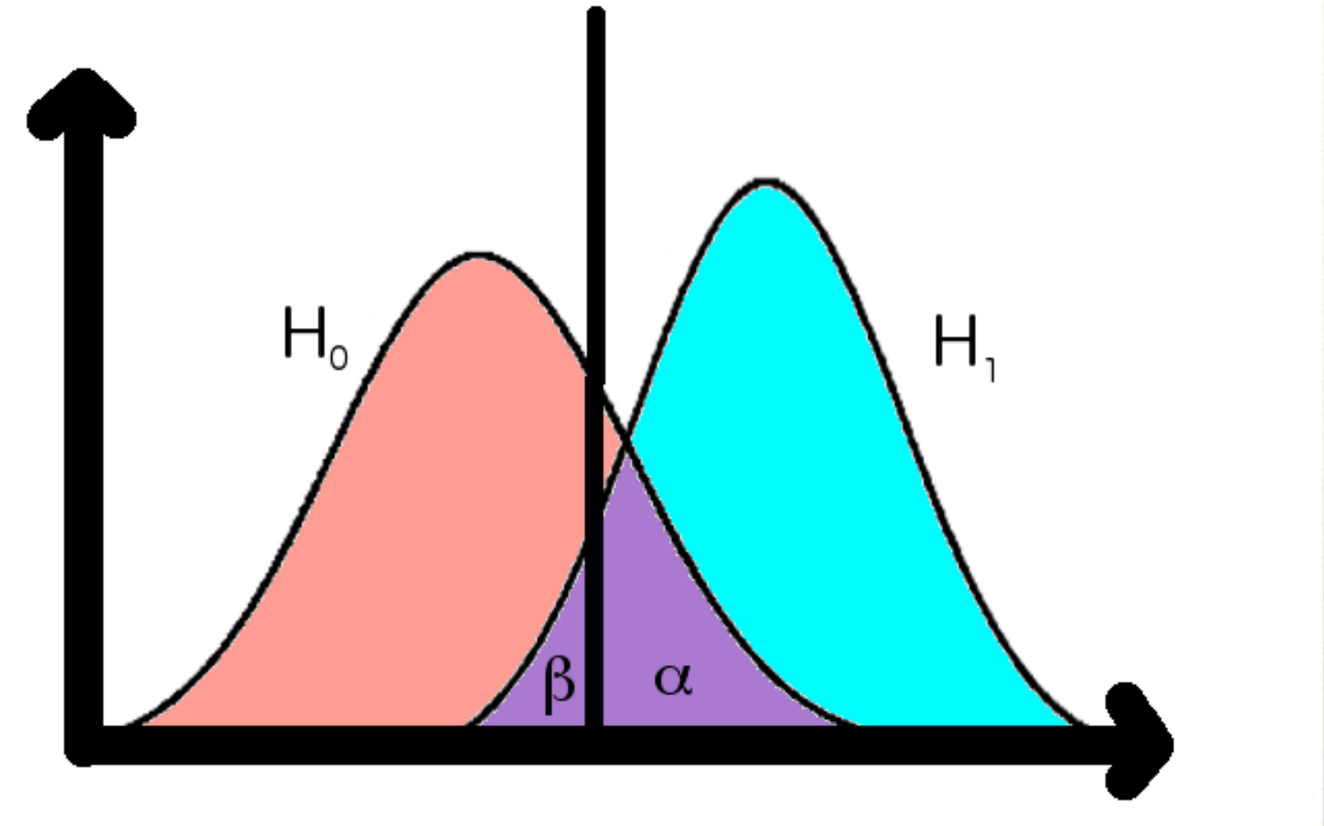
- To set strong limits if it's really not there.

Choosing Between Hypotheses



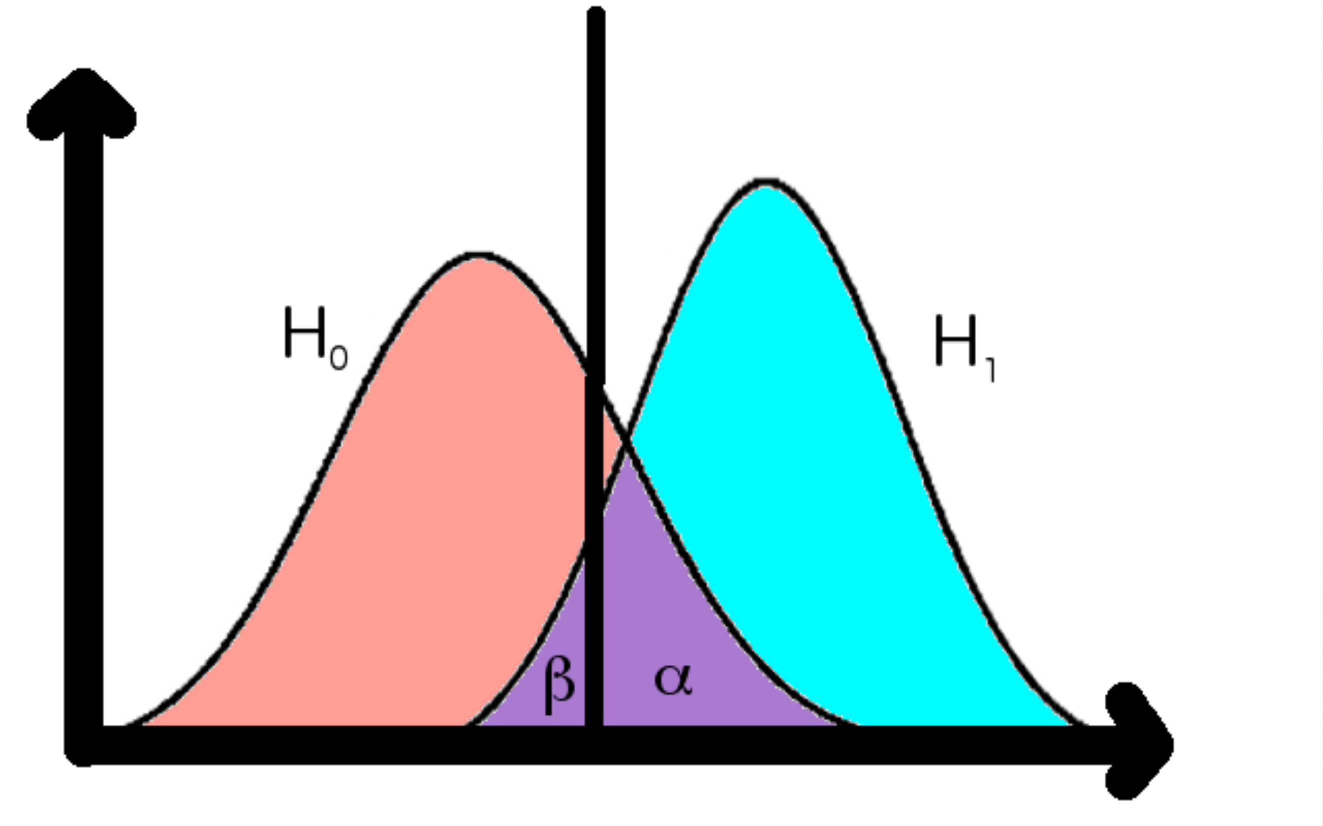
- Both discovery and exclusion are special cases of a very general problem in statistics: how to choose between two hypotheses:
  - H_0: Events are described by a “background” hypothesis (null hypothesis)
  - H_1: Events are described by a “signal plus background” hypothesis

Choosing Between Hypotheses



- We can fail in two distinct ways:
  - Choose H_0 when H_1 is true. This is the false negative rate, β. The **power** of the test is 1-β.
  - Choose H_1 when H_0 is true. This is the false positive rate, α, which is also called the **size**.

Choosing Between Hypotheses

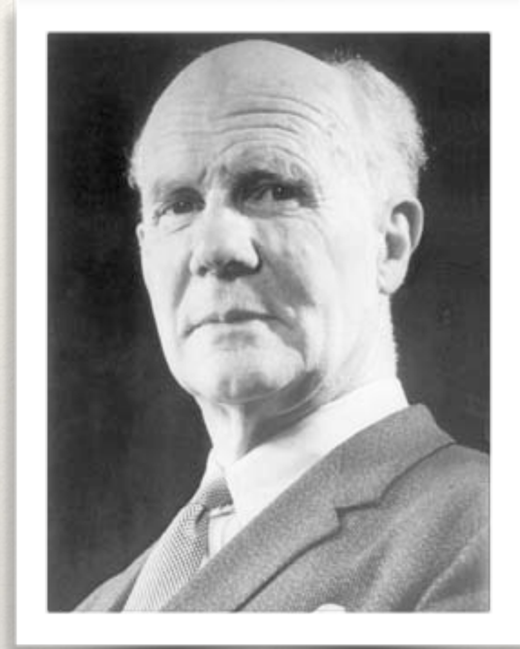


- Trade-off between power and size.
- Choose cut-off accordingly.
- But what we want from our test statistic is maximal power from a given size.
  - Keep in mind that we have plotted a particular choice of test statistic, in general a different choice gives different H_0 and H_1 curves.
- Can an optimal test statistic be obtained?

Neyman-Pearson Lemma

Actually Neyman and Pearson were roughly the same age. Google works in mysterious ways...

$$\Lambda(E) = \frac{L(H_1 \mid E)}{L(H_0 \mid E)}$$



- Yes, by the Neyman-Pearson lemma, the likelihood ratio maximizes power for a fixed choice of size.
- Note: this is the likelihood as calculated using all observables: no information thrown away

Likelihood

- Probability: function of observables for set values of parameters
- Likelihood: function of parameters for set values of observables
- Otherwise same function
- So in particle physics the likelihood is essentially the differential cross section, suitably normalized, after integration over unobserved particles and detector response.

$$\mathcal{P}(\mathbf{p}_i^{\text{vis}}|\alpha) = \frac{1}{\sigma_\alpha}\int dx_1 dx_2 \frac{f_1(x_1)f_2(x_2)}{2sx_1x_2} \times \left[\prod_{i\in\text{final}}\int \frac{d^3p_i}{(2\pi)^3 2E_i}\right] |M_\alpha(p_i)|^2 \prod_{i\in\text{vis}} \delta(\mathbf{p}_i - \mathbf{p}_i^{\text{vis}})$$

Likelihood

- Probability: function of observables for set values of parameters
- Likelihood: function of parameters for set values of observables
- Otherwise same function
- So in particle physics the likelihood is essentially the differential cross section, suitably normalized, after integration over unobserved particles and detector response.

$$\mathcal{P}(\mathbf{p}_i^{\text{vis}}|\alpha) = \frac{1}{\sigma_\alpha}\int dx_1 dx_2 \frac{f_1(x_1)f_2(x_2)}{2sx_1x_2} \times \left[\prod_{i\in\text{final}}\int\frac{d^3p_i}{(2\pi)^3 2E_i}\right] |M_\alpha(p_i)|^2 \prod_{i\in\text{vis}} \delta(\mathbf{p}_i - \mathbf{p}_i^{\text{vis}})$$

Use of this likelihood, containing the squared **matrix element** gives us the name "Matrix Element Method"

- A particularly nice aspect of likelihood-based analyses is that all spin correlations, kinematic features, etc. are taken into account automatically.
- So clearly an analysis that used the likelihood for all the variables that describe events would be optimal.
  This raises minor questions, like
- **Are we using the MEM for every analysis?**
- **If not, why not?**

- The Matrix Element Method **IS** widely used in B-physics, and in a number of top and Higgs analyses

- But there are challenges
  - NLO
  - Detector Resolution
  - Modelling reducible backgrounds
  - Computational demand of calculating likelihoods (especially due to integration over invisible particle momenta)

- Much work has gone into addressing these challenges.
- **NLO/ parton showers**
  (Alwall, Freitas, Mattelaer (2010), Soper, Spannowsky (2011) (2012), Campbell, Giele, Williams (2012)^2, Campbell, Ellis, Giele, Williams (2013))
- **Detector Resolution**
  (e.g., Chen, Di Marco, Lykken, Spiropulu, Vega-Morales, Xie 2014).
- Programs to calculate likelihoods/ weights
  - **MadWeight** (general)
    Artoisenet, Mattelaer (2008), Artoisenet, Lemaitre, Maltoni, Mattelaer (2010)
  - For four-lepton final state:
    JHUGen (**MELA**)
    Gao, Gritsan, Guo, Melnikov, Schulze, Tran (2010),
    Bolognesi, Gao, Gritsan, Melnikov, Schulze, Tran, Whitbeck (2012)
    MEKD
    Avery, Bourilkov, Chen, Cheng, Drozdetskiy, Gainer, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, Snowball (2012)
    Chen, Cheng, Gainer, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, Snowball (2013)
  - **MadGraph**, **CalcHEP,** etc. provide standalone code …
- Still difficulties. In fact you could argue that, in part as a response to these difficulties, there are

Two Matrix Element Methods

Directly Calculate Likelihood

Pros: Optimal, by Neyman-Pearson lemma

Cons: Very challenging to obtain exact likelihood

Use Likelihood-Based Variable

Pros: Much more feasible

Trade-off between sensitivity and ease of calculation

Two Matrix Element Methods

Directly Calculate Likelihood

Pros: Optimal, by Neyman-Pearson lemma

Cons: Very challenging to obtain exact likelihood

Think of MEM-variables like any other kinematic variables.

Use Likelihood-Based Variable

Pros: Much more feasible

Trade-off between sensitivity and ease of calculation

Using approximations does not make one wrong, only less sensitive.

Two Matrix Element Methods

Directly Calculate Likelihood

Pros: Optimal, by Neyman-Pearson lemma

Cons: Very challenging to obtain exact likelihood

Think of MEM-variables like any other kinematic variables.

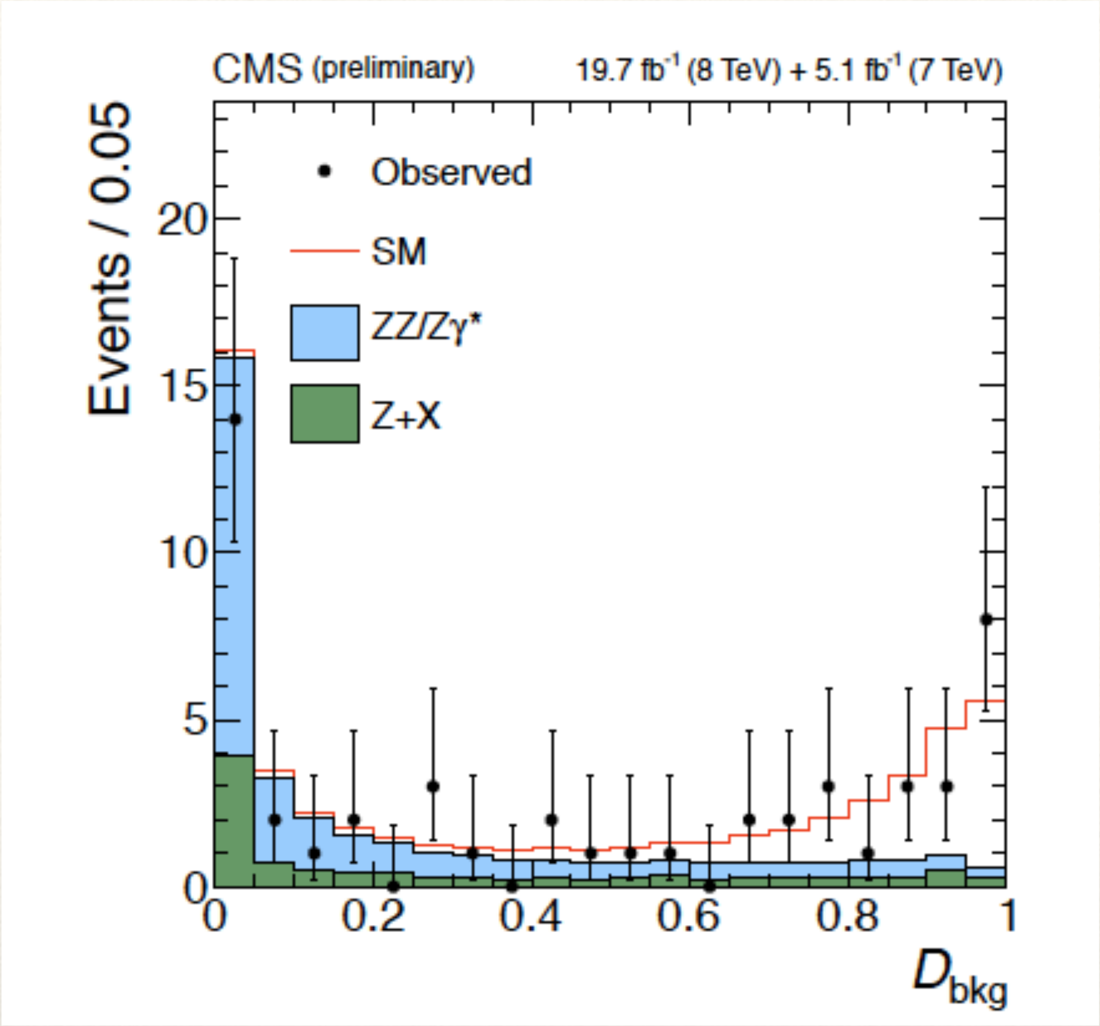
Use Likelihood-Based Variable

Pros: Much more feasible

Trade-off between sensitivity and ease of calculation

Using approximations does not make one wrong, only less sensitive.

The main difference is one of statistical interpretation.
On the left we directly obtain our test statistic.
On the right we use the values that might go into this (e.g. $P_s/(P_s + P_b)$) as kinematic variables, like M_{T2} or M_{2CC}



HIG-14-014

- We see that there is, e.g., a Z + X contribution to D_{bkg}, which is a variable derived from the signal and background likelihoods.
- Effectively the above plot shows a likelihood of likelihood-based variables
- CMS has used templates (essentially 1 or 2-D likelihoods) which employed likelihood-based variables in four-lepton studies
- This is an example of the 2nd approach: “Likelihood-based variable” or “kinematic discriminant”

- What opportunities do we gain by using the MEM-variable approach?
- Are there new questions which we can answer?

Yukon Gold Rush

What if we don’t have a signal hypothesis?

What if we don’t have a signal hypothesis?

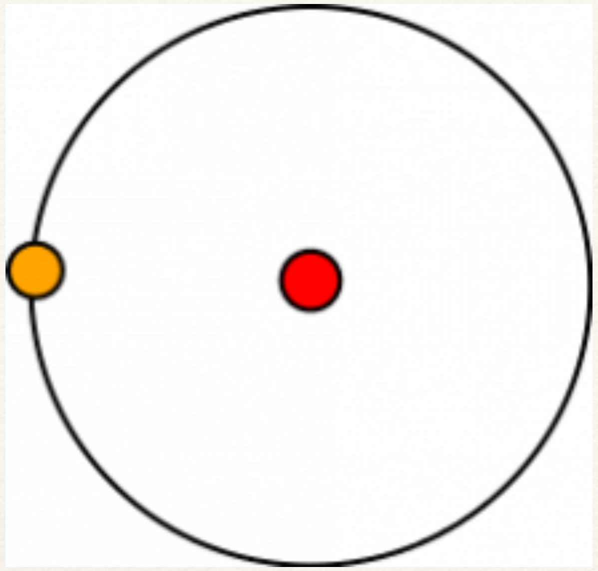
Debnath, Gainer, Matchev
arXiv:1405.5879

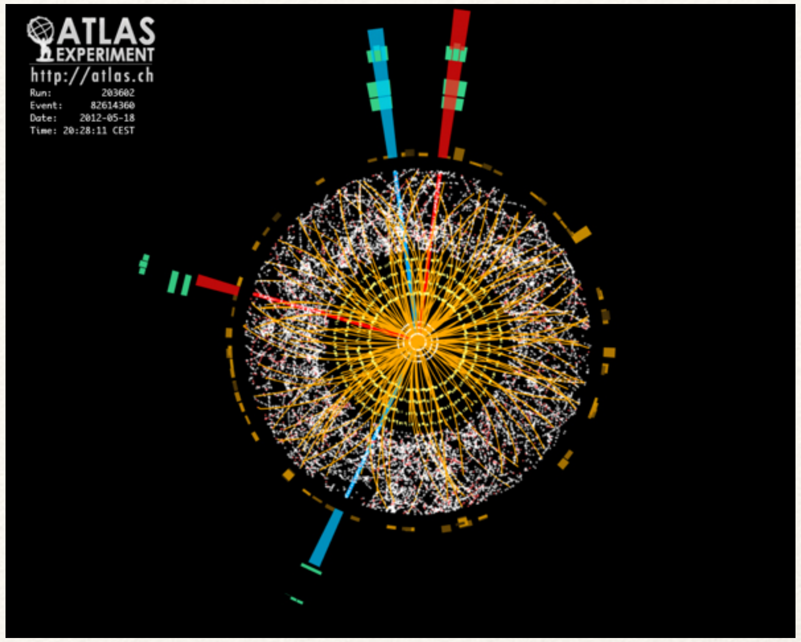
- Whether viewed as calculating likelihood, or as obtaining a kinematic variable, the MEM involves both the signal and background differential cross sections.
- So we need both a signal and a background model.
- What if we don’t have a signal model?

- For example, consider some well-studied channel, like multi-jets plus missing energy.
- Often we would perform an analysis with some SUSY model as the signal.
- But someone, we'll call him Carlos Wagner, might complain:

"How do you know it's SUSY?
What significance can you claim without making this possibly unwarranted assumption?"

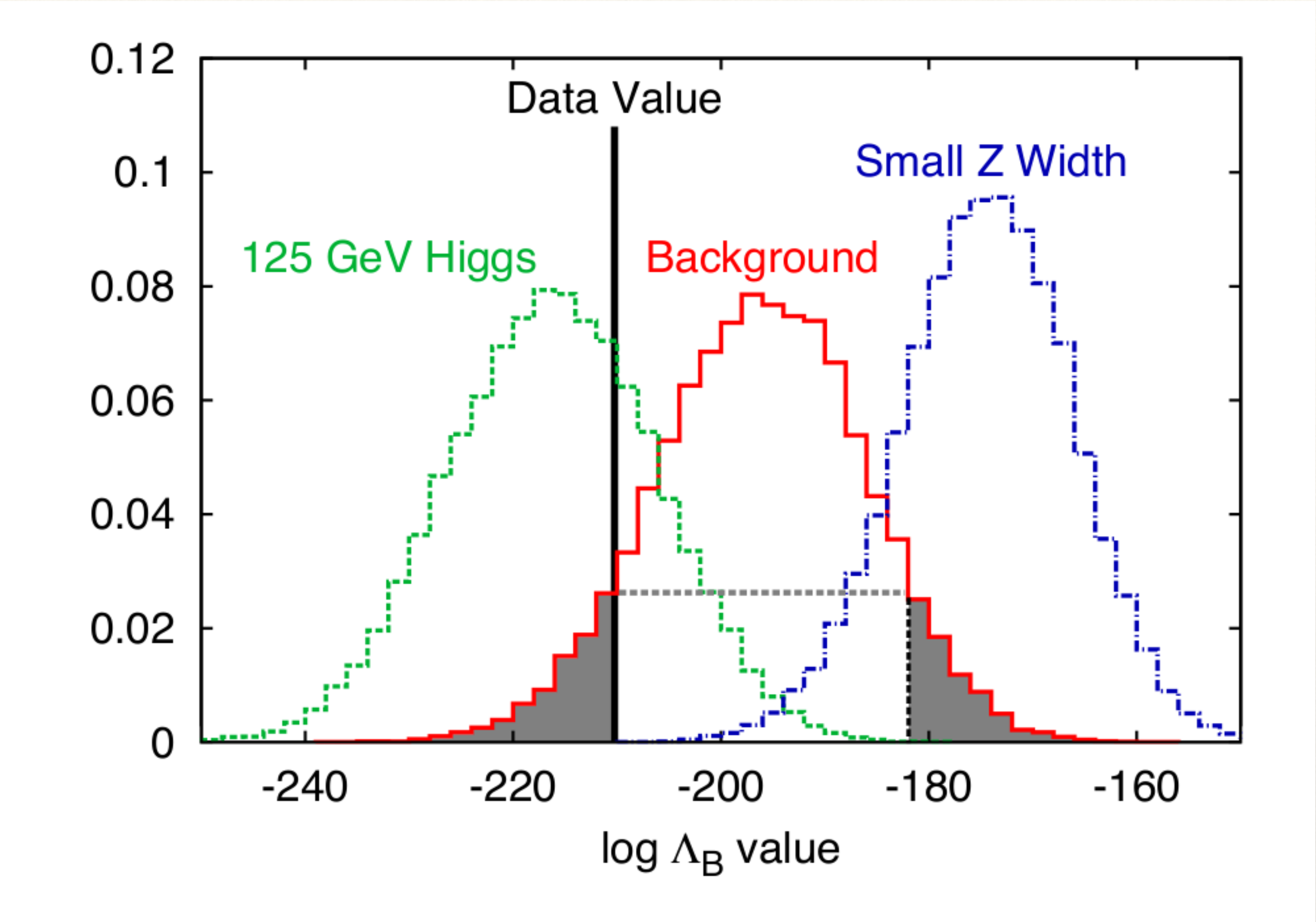
–Carlos Wagner?



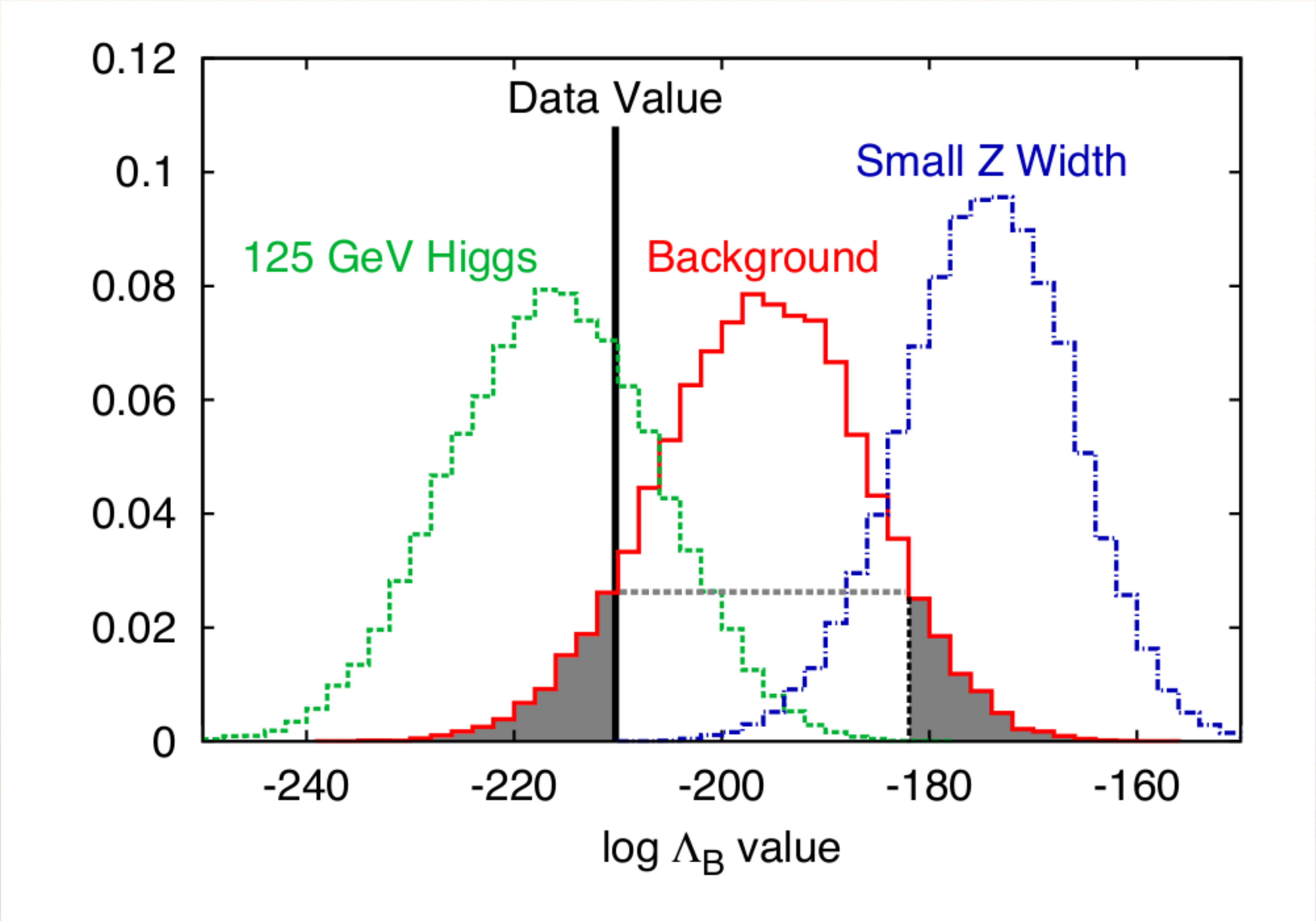


- We will investigate this question using the four-lepton final state at 125 GeV
- "Hydrogen atom of the Matrix Element Method"—- Chris Lester
- Actual applications would be to more complicated final states: we're doing fine in four-leptons already!

- Approach: use a likelihood-based variable
- I.e., calculate the differential cross section (or something similar) for all events using the background hypothesis
- Compare the observed values of this quantity to the background distribution for this quantity.



- Naive approach: do a 20 event pseudoexperiment
- Take the total value of log (background squared matrix element)



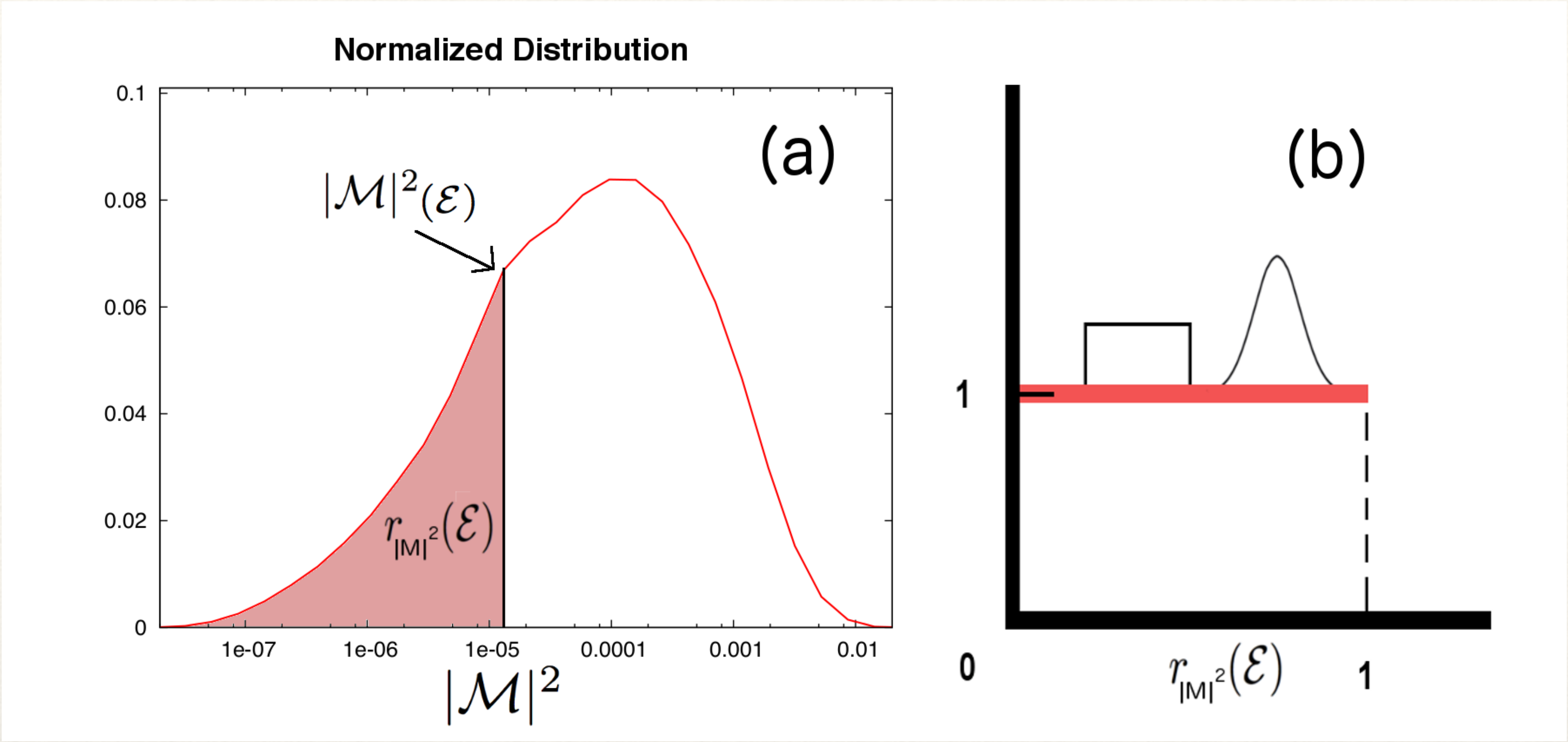
- This is crude, but it works

Variables Should Be Flattened

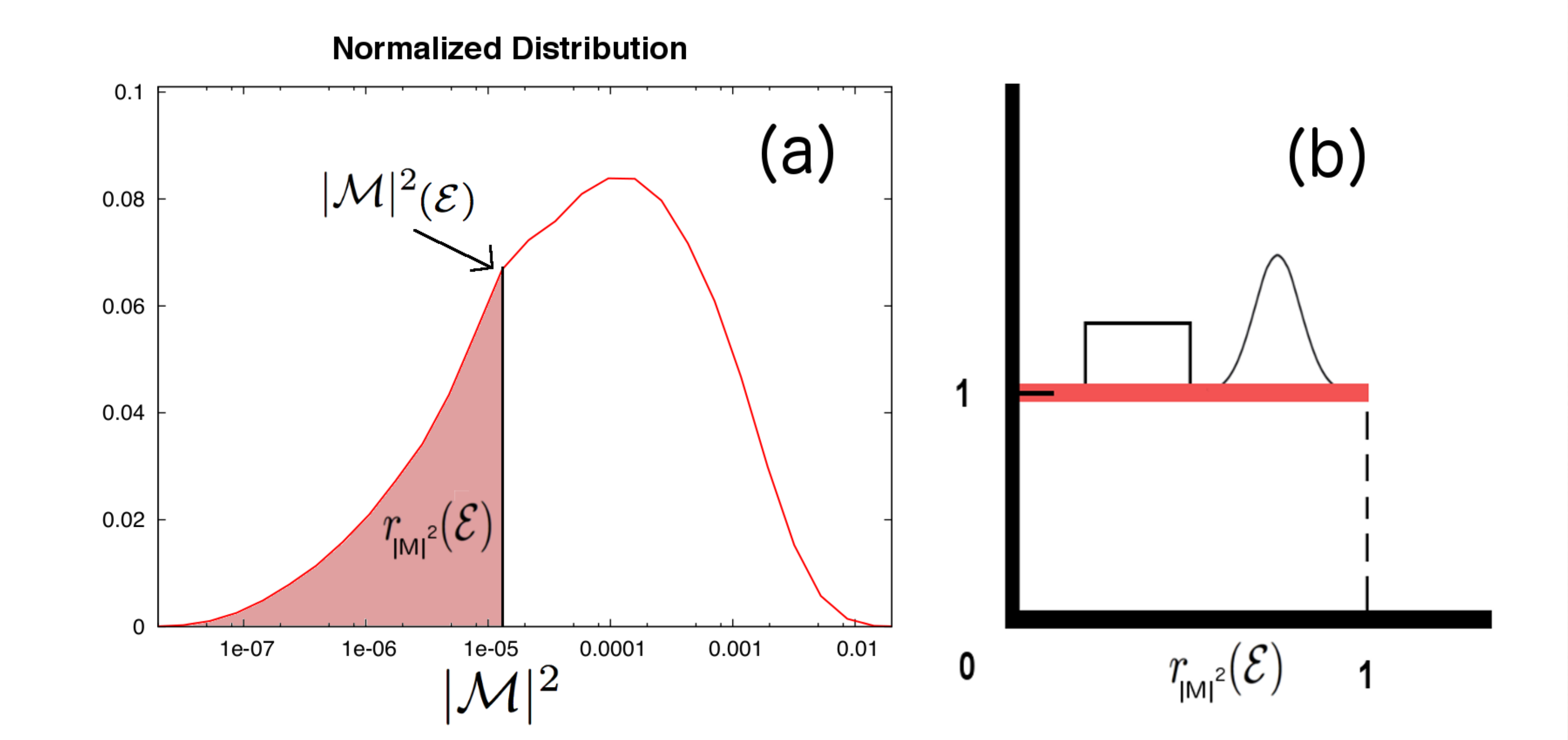
- A more realistic analysis would use e.g. a binned χ^2
    - Would like an equal number of background events in each bin
- Differential cross section values not totally intuitive: relative values more meaningful

Variables Should Be Flattened

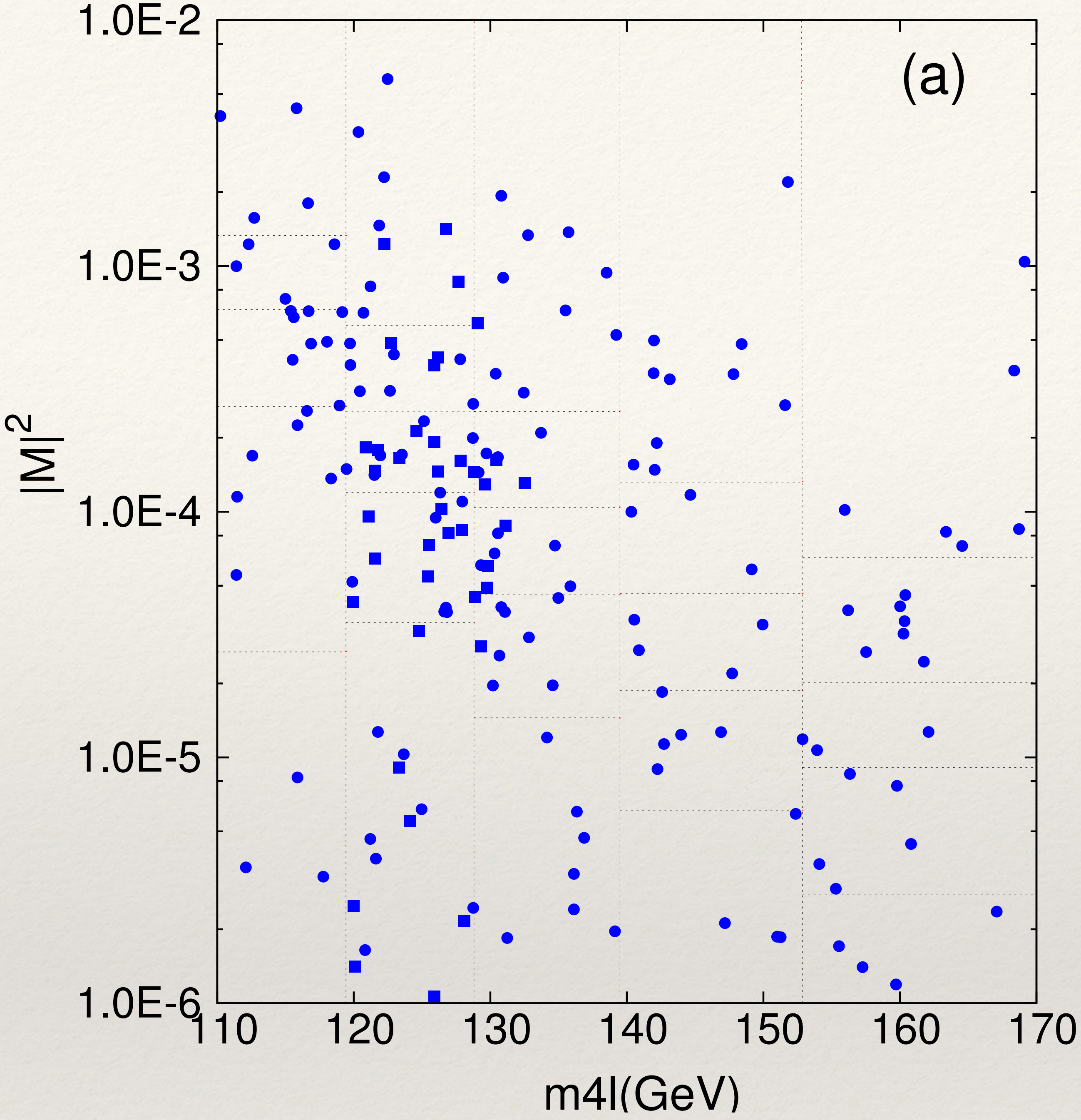
- A more realistic analysis would use e.g. a binned χ^2
  - Would like an equal number of background events in each bin
- Differential cross section values not totally intuitive: relative values more meaningful



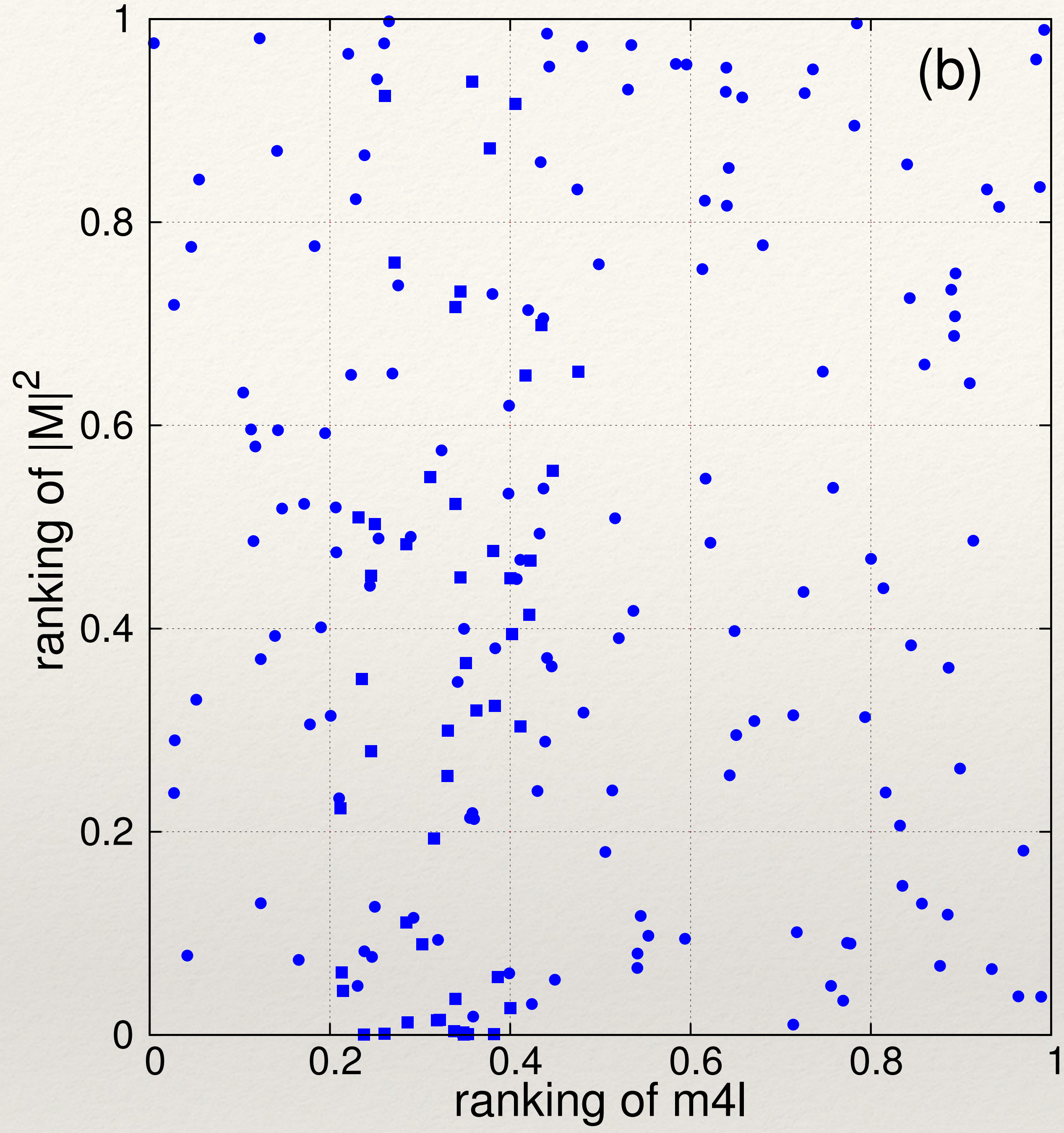
- We achieve this flattening by replacing a given value of our variable, $|\mathcal{M}|^2$, with the integral over the background distribution up to that value.
- I.e., we replace a variable with its cdf.
- This makes the background distribution flat; signals are deviations from flatness.



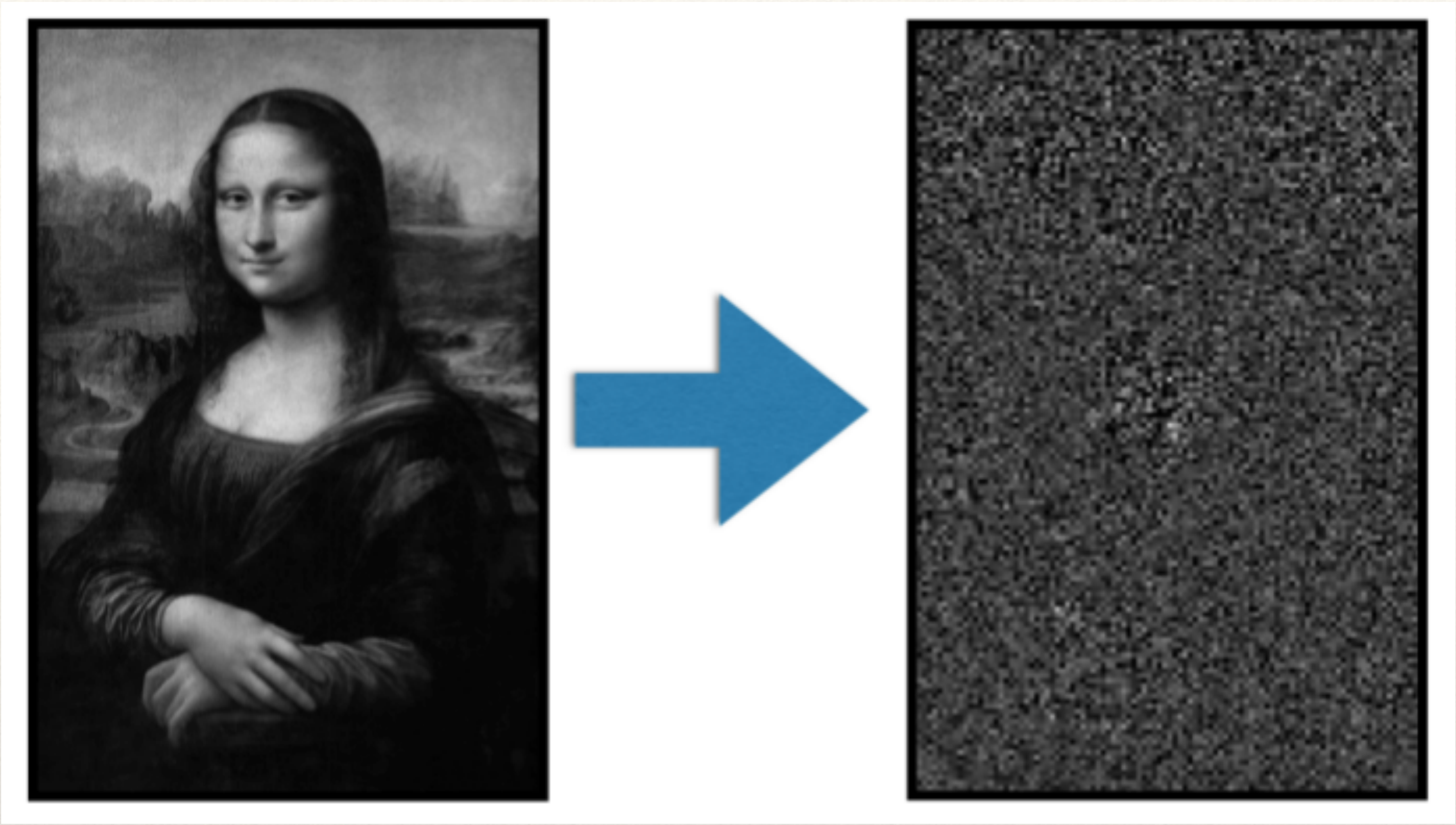
- **Note**: it is straightforward to include detector effects, reducible backgrounds, etc.
- One simply includes their contributions in the background distribution of the MEM-variable.



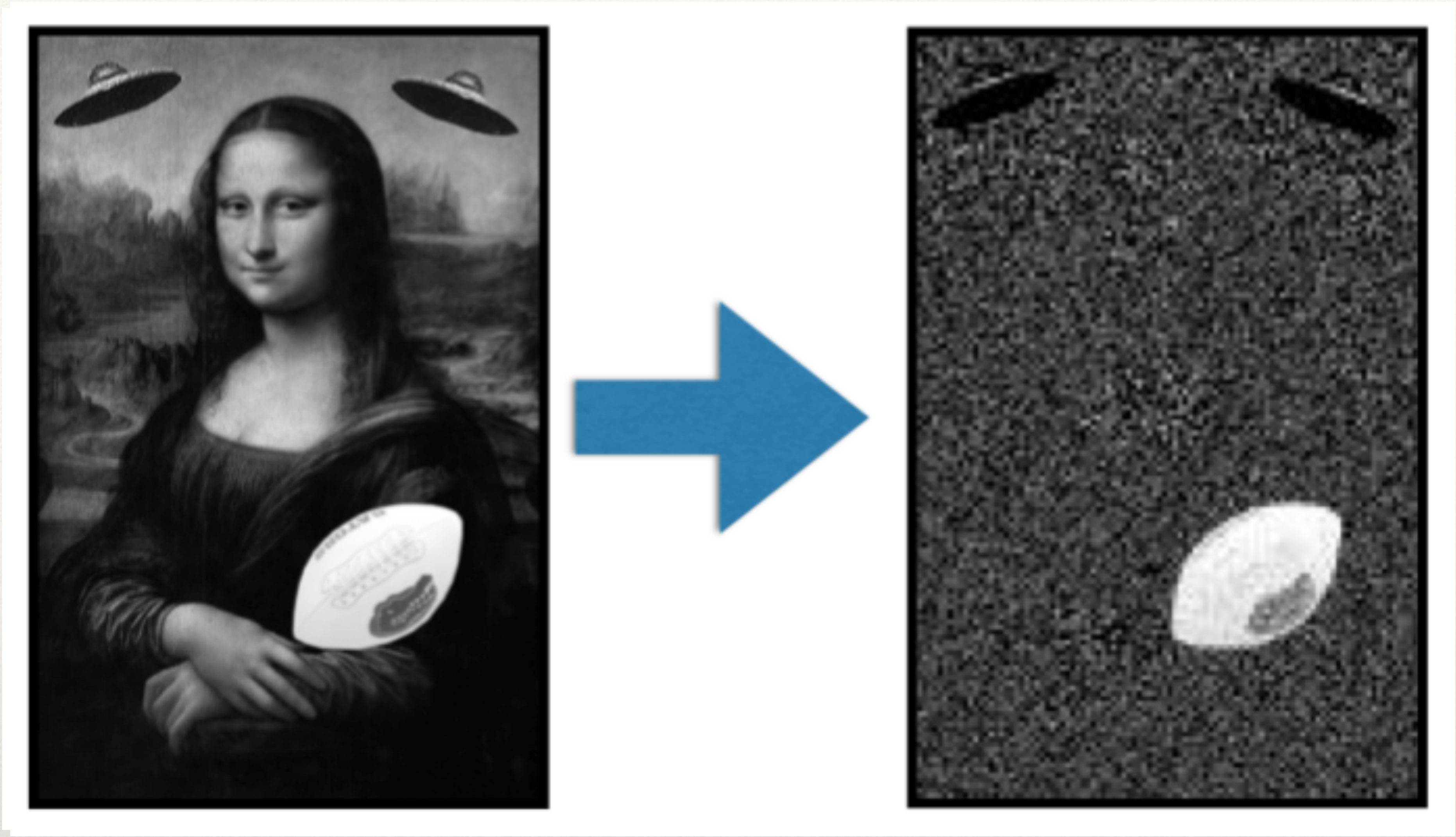
- We extend this to looking at 2-variables
- Here we use squared matrix element and four-lepton invariant mass
- Hard to interpret excesses and deficits



- Flattening both variables: we can tell when excesses and deficits are significant



- If we have a sufficient understanding of our background, we can use more dramatic flattening procedures
- Here the Mona Lisa (ML) is the SM background
- The right panel is MC generated using the ML pdf but weighted by 1/ (ML pdf)



- Add signal: UFOs and football
- Generate "events", reweight histogram by 1/ ML pdf as before
- Signal clearly visible

- We’ve shown that we can take advantage of the “variable” interpretation of MEM variables to perform totally model-independent searches, and that flattening the resulting variables is helpful

- We've shown that we can take advantage of the "variable" interpretation of MEM variables to perform totally model-independent searches, and that flattening the resulting variables is helpful

And now for something very much related to our initial topic, but different from the use of the background MEM for model-independent discoveries

Gainer, Lykken, Matchev, Mrenna, Park
arXiv:1404.7129, JHEP

Too many notes!

parameters
Too many ~~notes!~~

BSM Models with Lots of Parameters

SUSY frameworks like

General Gauge Mediation

pMSSM

NMSSM (more without Z_3 symmetry!)

also

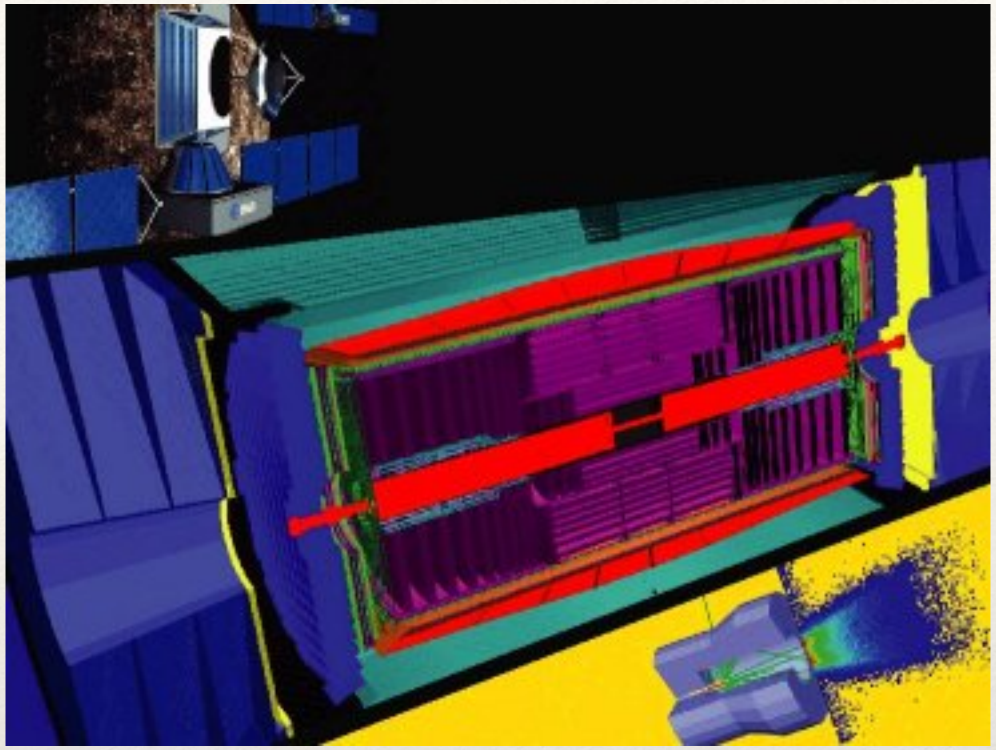
Effective theories

Your favorite model?

The Problem with Parameters

- Need large MC samples to evaluate sensitivity
- Here MC = “fullsim” MC: run through high-powered detector simulator (Geant)
- Time consuming!
  - ~event per minute

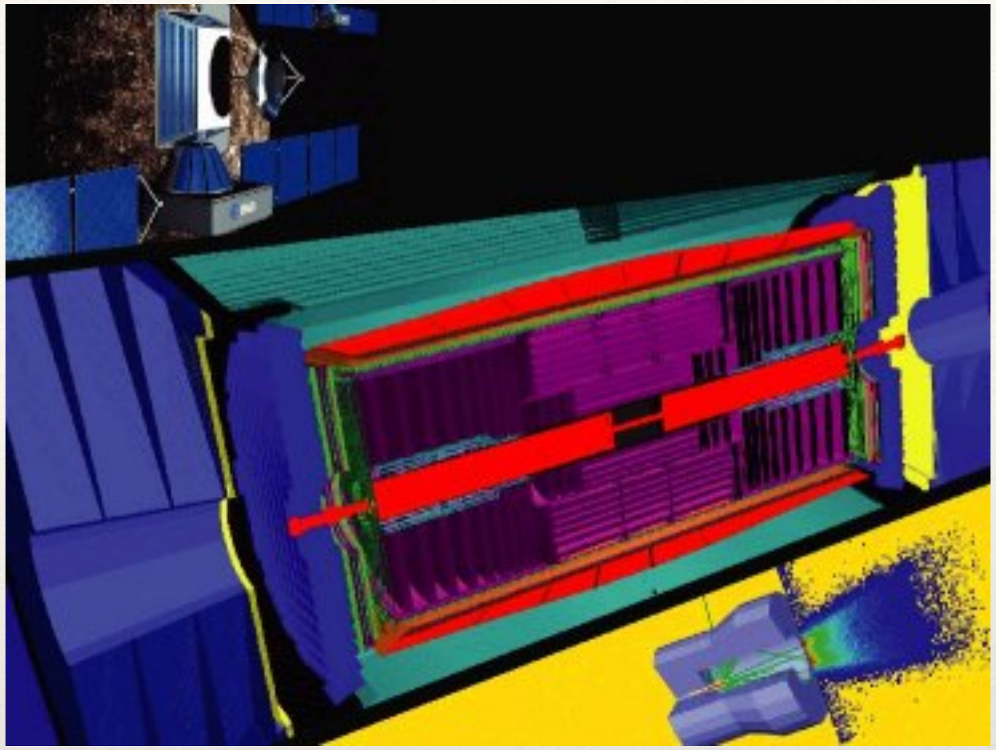
    vs.
  - events per second



- Makes it challenging to use a computationally intensive variable (MEM) to study large parameter spaces in full generality

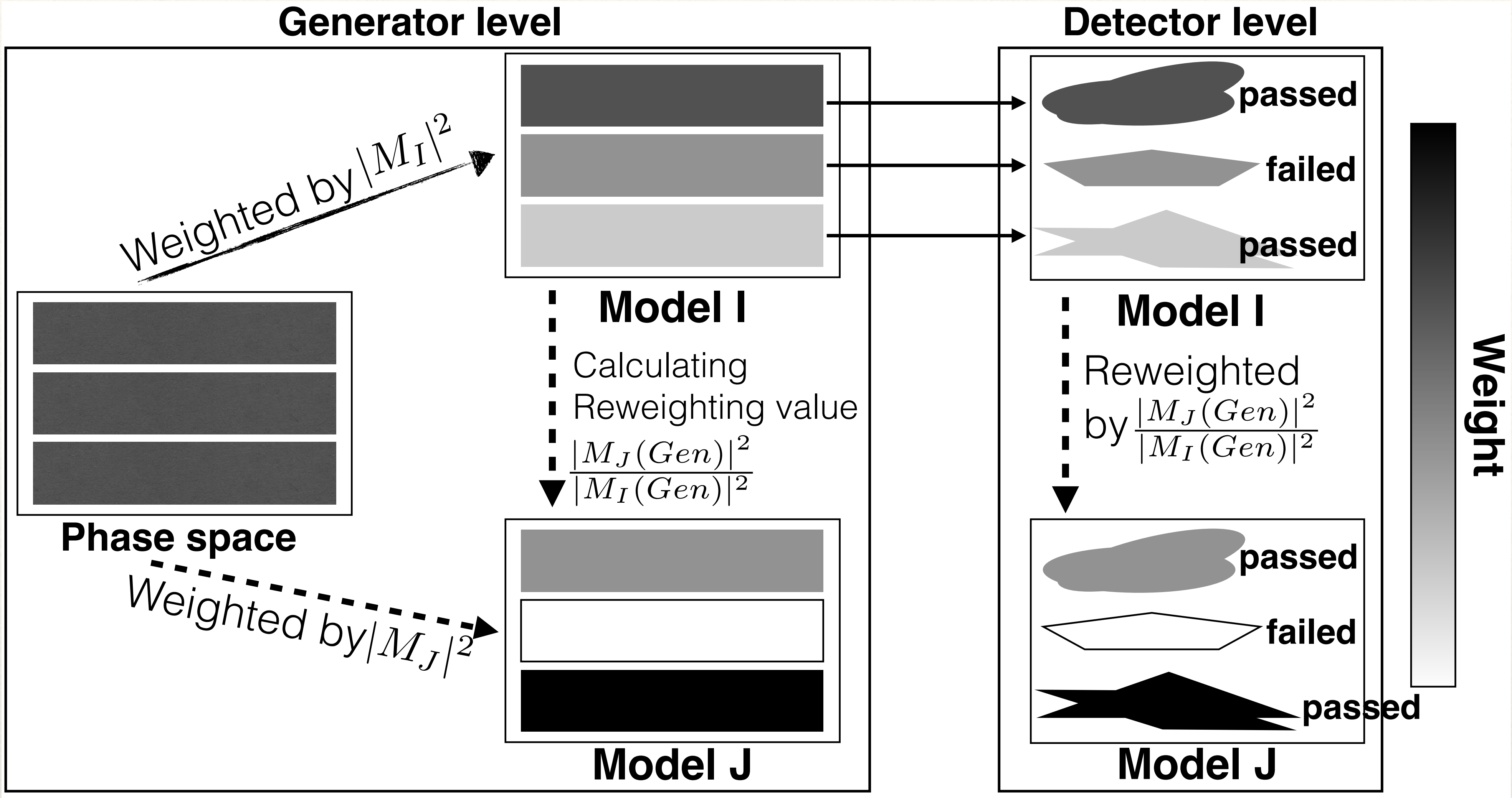
The Problem with Parameters

- Need large MC samples to evaluate sensitivity
- Here MC = "fullsim" MC: run through high-powered detector simulator (Geant)
- Time consuming!
    - ~event per minute
      vs.
    - events per second

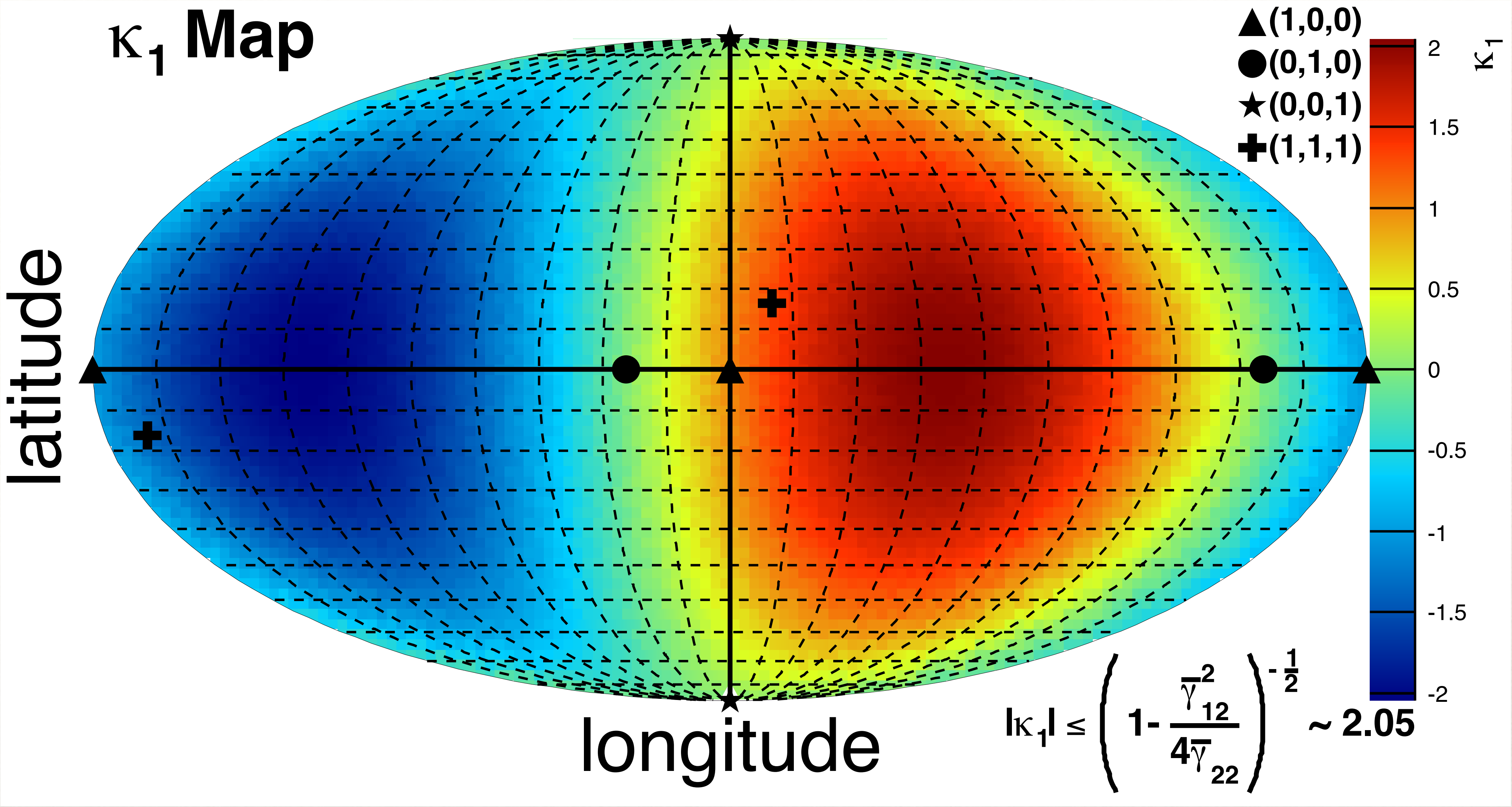


- Makes it challenging to use a computationally intensive variable (MEM) to study large parameter spaces in full generality

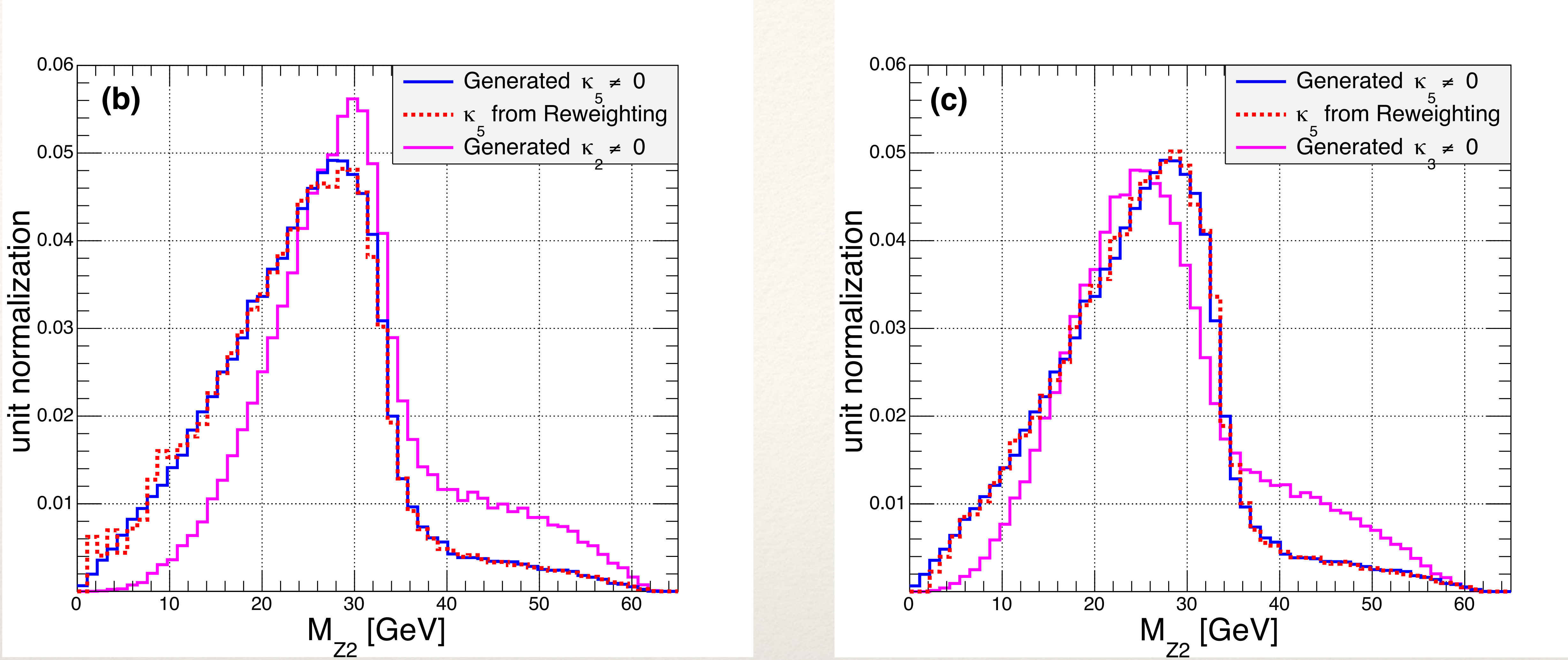
or does it???



- **Goal**: obtain the best possible modeling of a distribution using as few fullsim events as possible
- **Approach**: take unweighted events generated for some model
- **Key point**: the weight is the ratio of parton-level differential cross sections using truth information
(e.g., use neutrino momentum if there are neutrinos in the event)

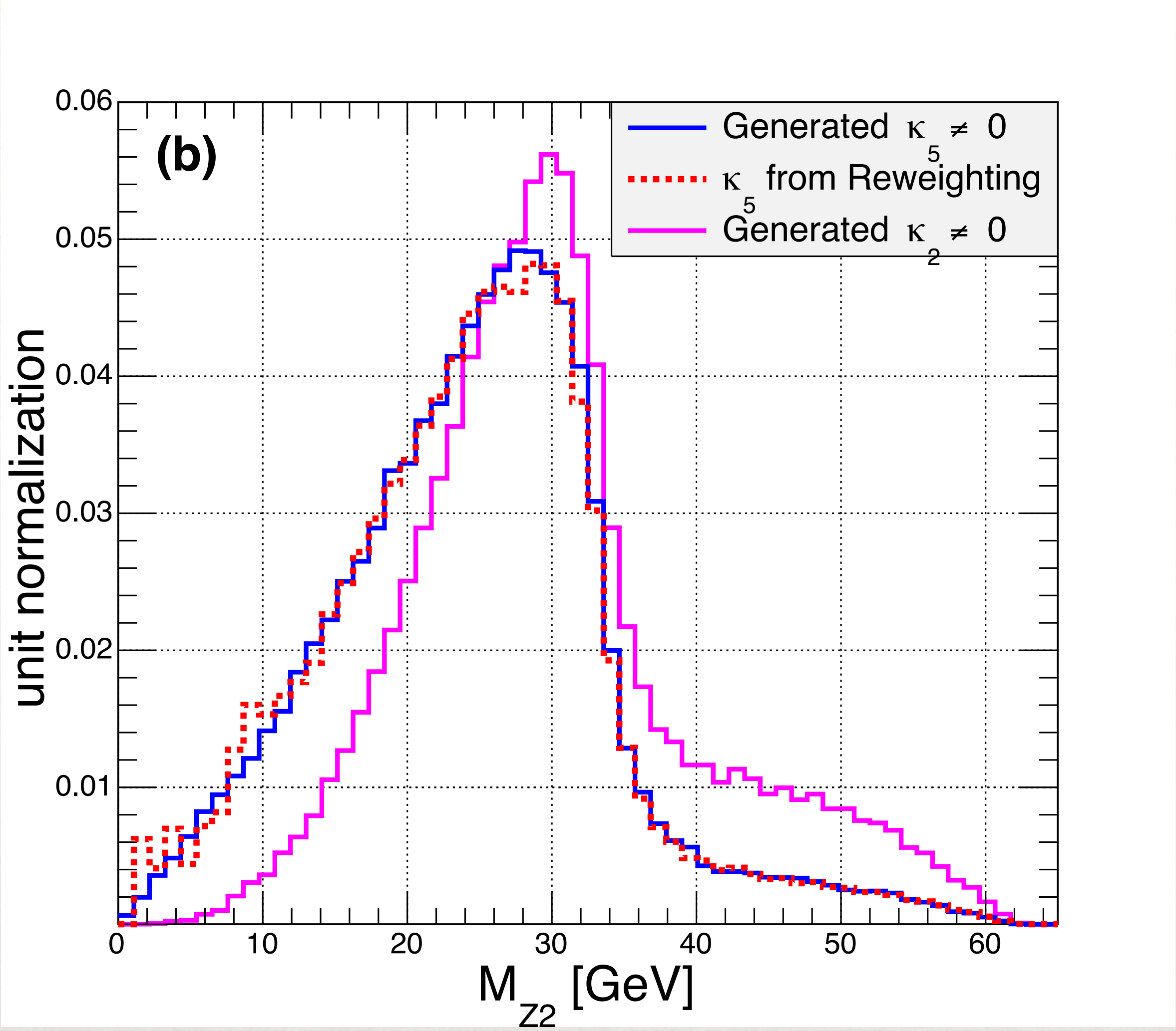


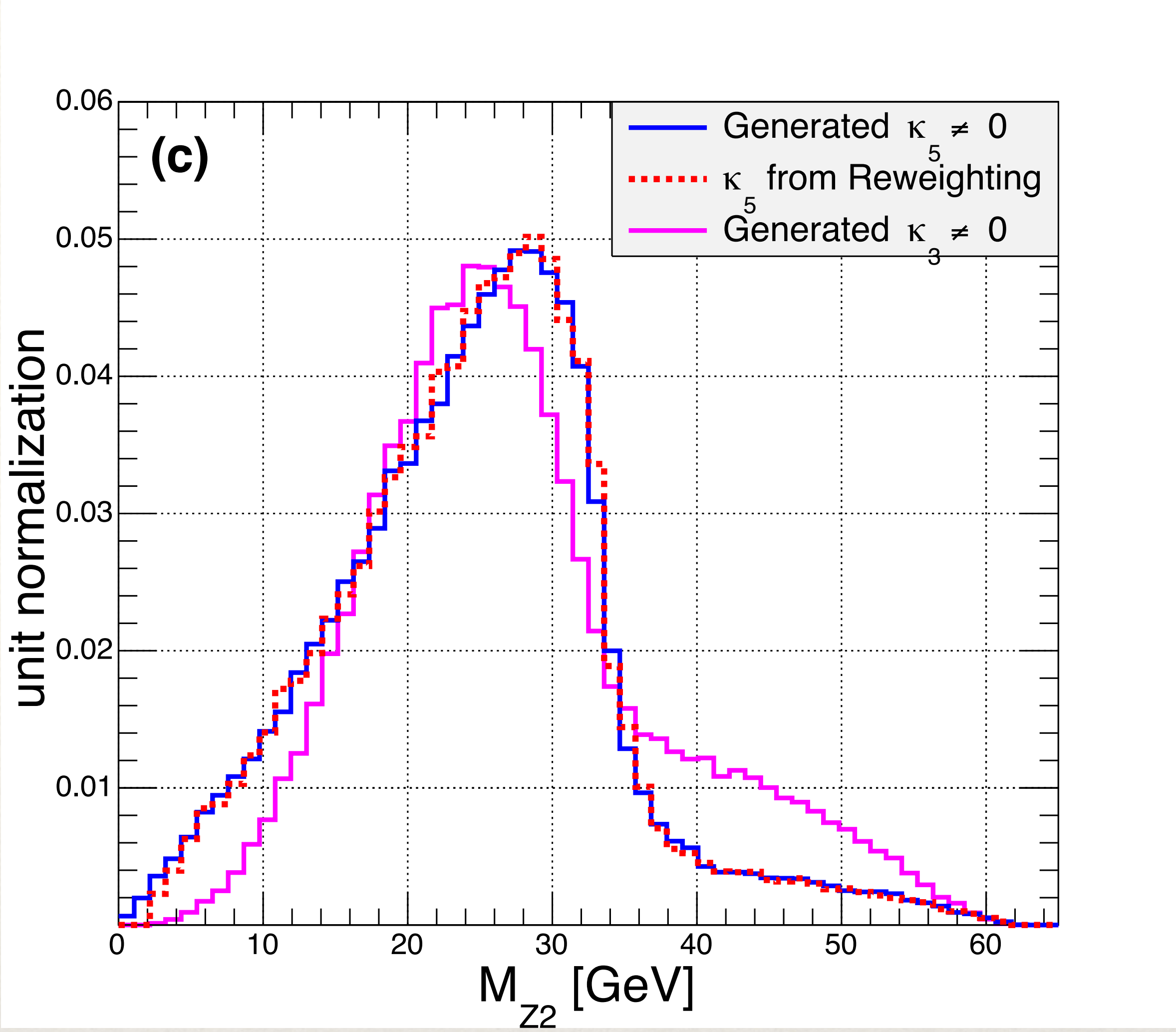
- Reweighting to find the effects of standard cuts on X -> 4ℓ cross sections in a parameter space of general Higgs couplings (from Gainer, Lykken, Matchev, Mrenna, Park, 2013)



- Here we have reweighed events generated from various “pure” couplings to operators in the effective field theory of XZZ couplings to obtain the distribution of the lighter Z* invariant mass when mX = 125 GeV

- Ultimately what we want from our MC samples, when doing a MEM analysis, is the distribution of MEM variables for the background and any point in our signal parameter space





- Clearly we CAN do this with reweighting
- Lots of technicalities:
  - Errors on histograms, tails of distributions, etc. Ask me during coffee break!

Conclusions

- The Matrix Element Method is an optimal(-ish) way to discover new physics
- The "MEM-variable" approach is more feasible,
- and allows for model-independent searches using the background matrix element.
- MEM challenges are being overcome
- For example, large parameter spaces can be tamed using reweighting.
- I look forward to a Matrix (Element Method) Revolution in Run 2 at the LHC!