Likelihood Methods for Higgs Properties Measurements

James "Jamie" Gainer

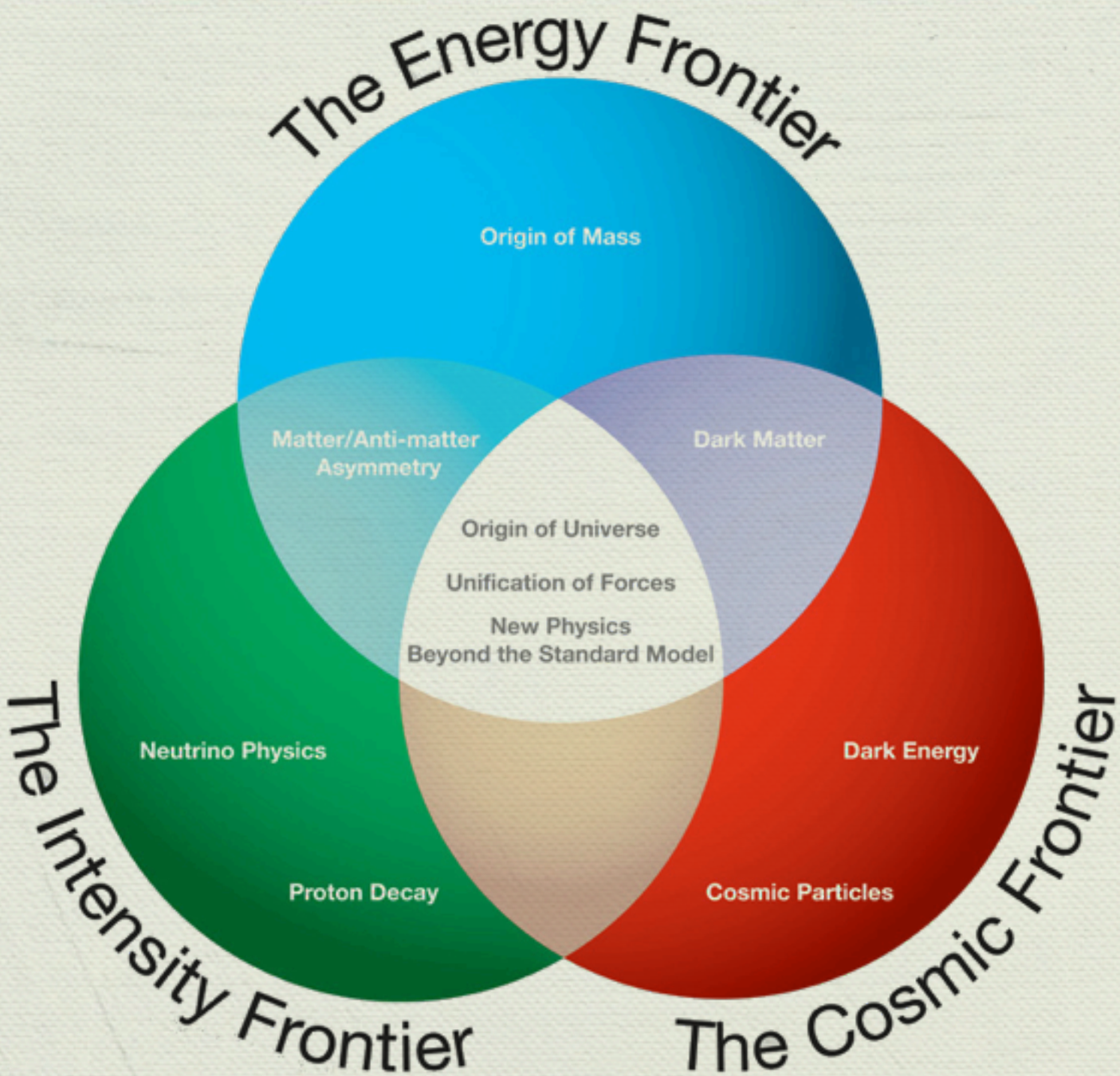
Snowmass Energy Frontier Meeting at BNL

Working Group Session, 8: Higgs Spin, CP Mixtures, and Other Properties (Higgs)
Friday, April 5, 2013

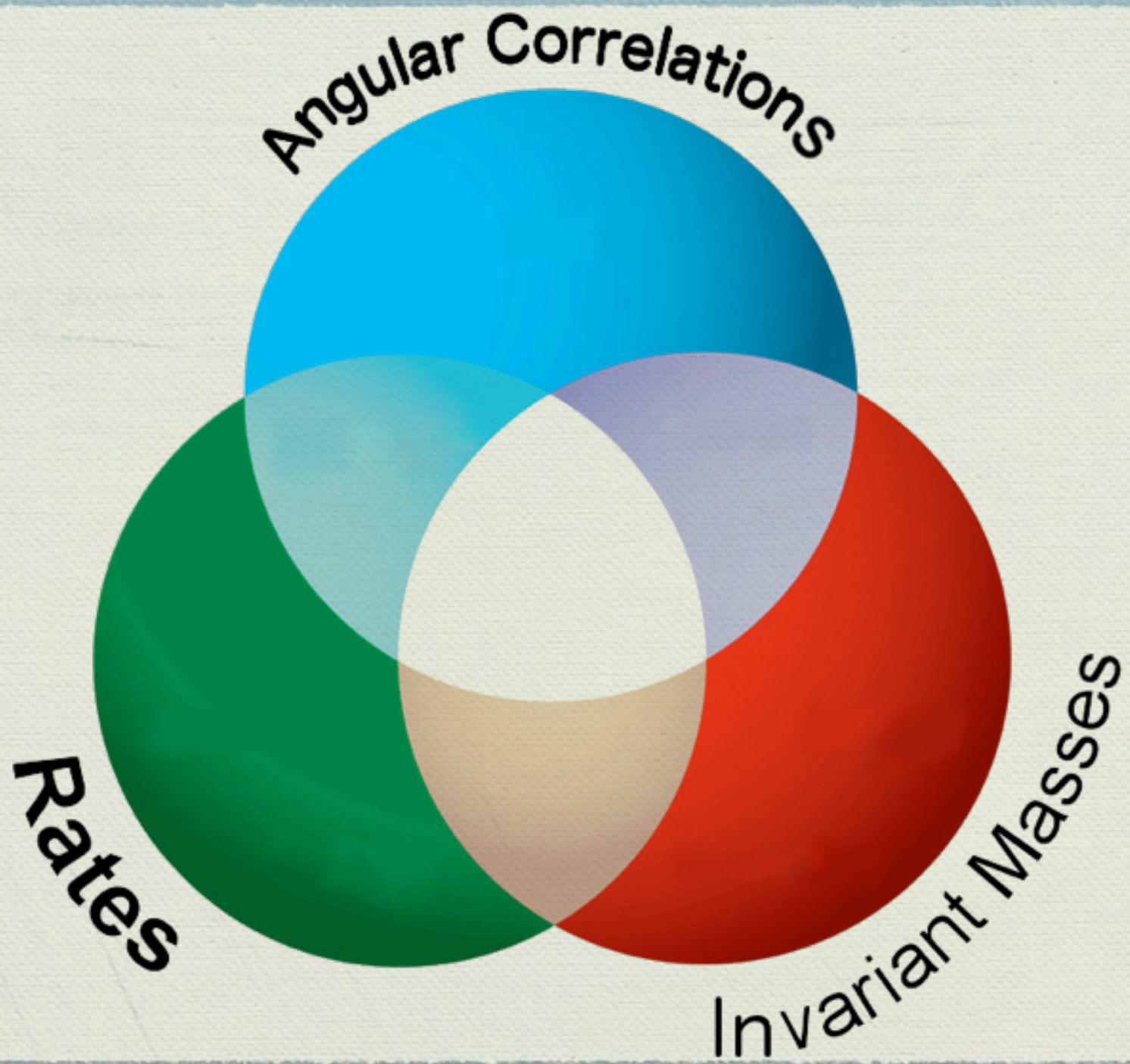
Outline / Apologia

- I was asked to provide a brief review of likelihood methods for Higgs properties measurements
- I'll briefly review the use of methods that utilize likelihood discriminants, focusing on multivariate analyses
- I'll focus on $H \to VV^* \to 4f$ or V 2f: motivated by the assumption that the Higgs is a scalar
- "Theorists perspective":
  After the review, I'll focus on a particular multivariate likelihood analysis
  - which has been vital in measuring Higgs spin and couplings
  - which is even more important for Higgs spin and coupling measurements going forward
  - where theorist input may be useful

It's Snowmass...



It's Snowmass...



Complimentarity

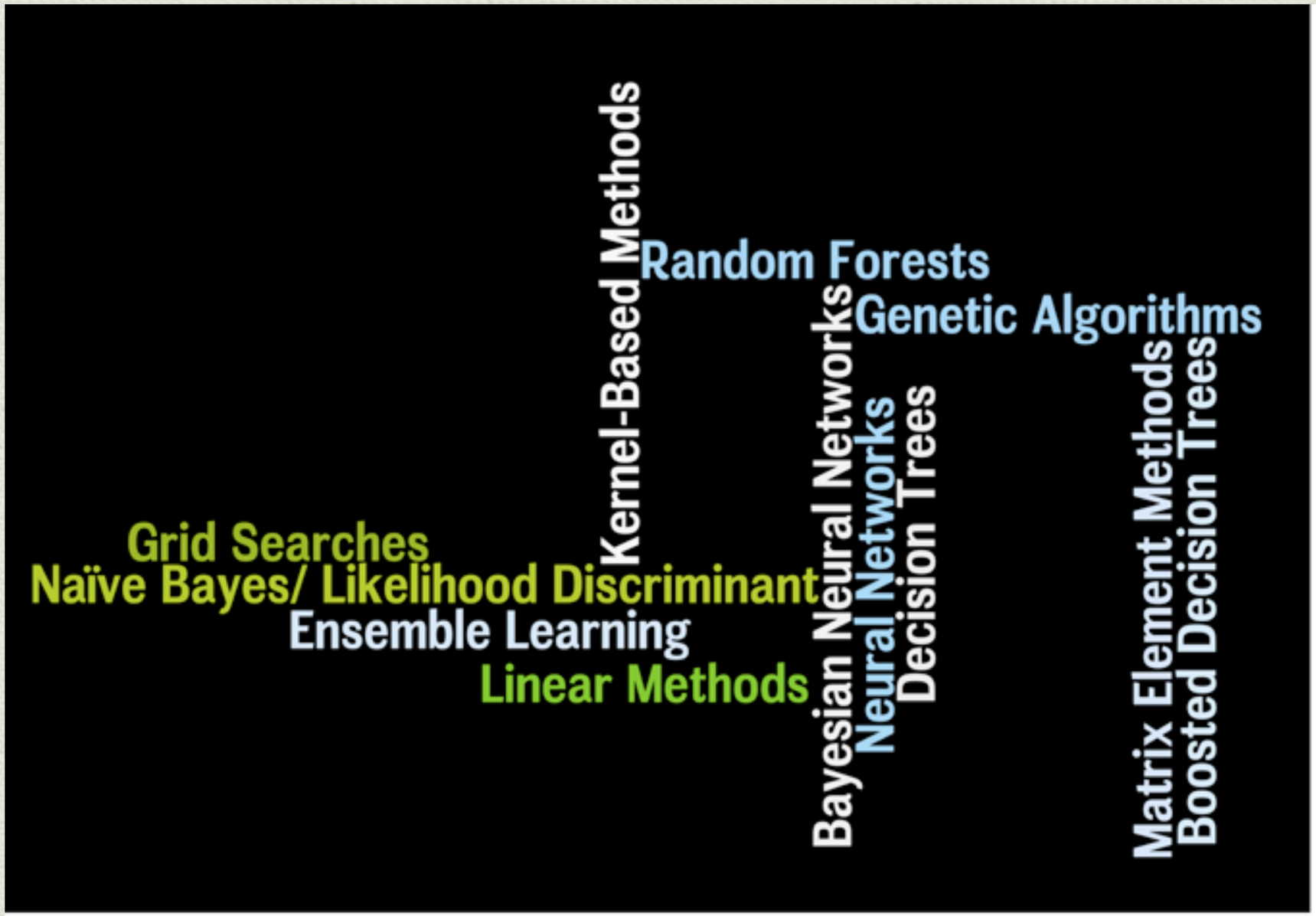
- Overall rate, angular information, and invariant mass distributions provide complimentary approaches for determining the properties of the BEH/ SM Scalar/ God/ Higgs boson

- Invariant mass distributions: either the invariant mass of the putative Higgs, or of intermediate states, (e.g. Zs and Ws)

Complimentarity

- It's conceptually useful to think about the three approaches separately,
- However multivariate analyses can automatically take all three kinds of information into account

Multivariate Analyses

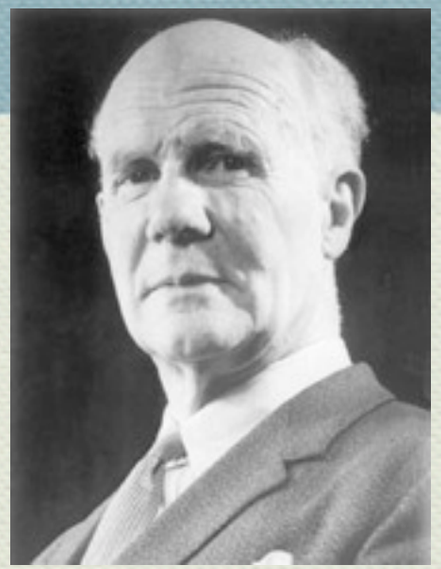


cf e.g. Bhat: Annu. Rev. Nucl. Part. Sci. 2011. 61:281–309

Test Statistics

- Often (but not always) in MVA one uses a likelihood discriminant as a test statistic
- In theory language, a test statistic is a function from experiments to real numbers. $t : \{E\} \quad \rightarrow \quad \mathbb{R}$
- A **good** test statistic lets us determine which hypothesis best explains the data
- **Example**: there are two types of events "signal" and "background" then want would like a test statistic for which higher values corresponds to signal events and low values corresponds to background events

Neyman-Pearson Lemma



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

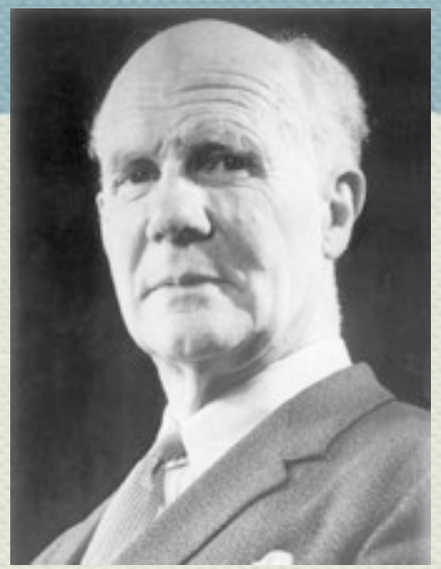
- Why use a likelihood discriminant?

- In a mathematically well-defined sense, the best choice of test statistic for distinguishing between two hypotheses (like "signal" and "background") is the likelihood ratio/ discriminant

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

where H_0 and H_1 are two alternative hypotheses and E is the data in an experiment

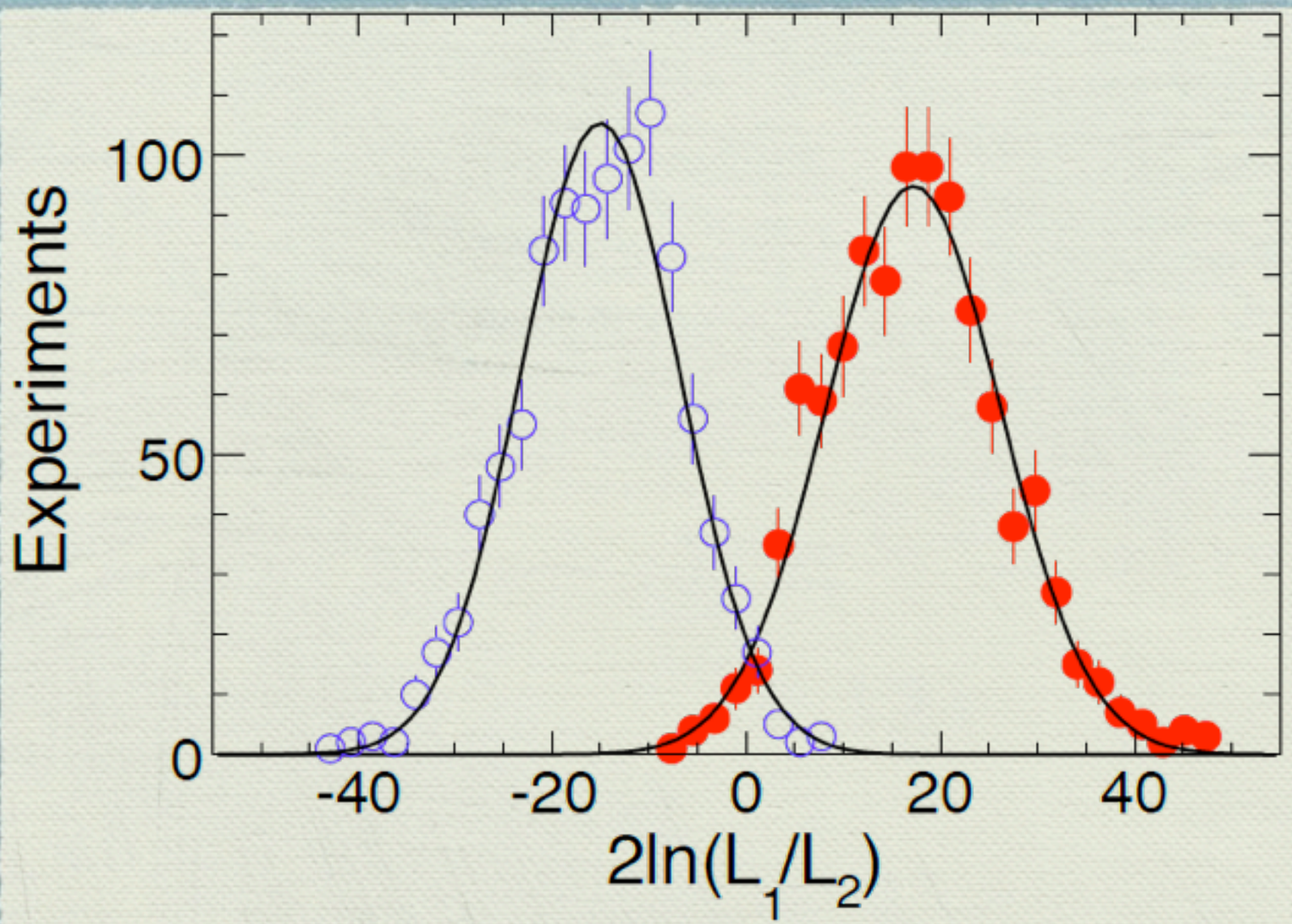
Neyman-Pearson Lemma



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

- Think of H_0 as the "null" hypothesis
- Then the procedure of "assigning" the experiment to the alternative hypothesis H_1 when $\Lambda(E) > r$
  - Minimizes the false negative rate for a given false positive rate (which is a function of r).

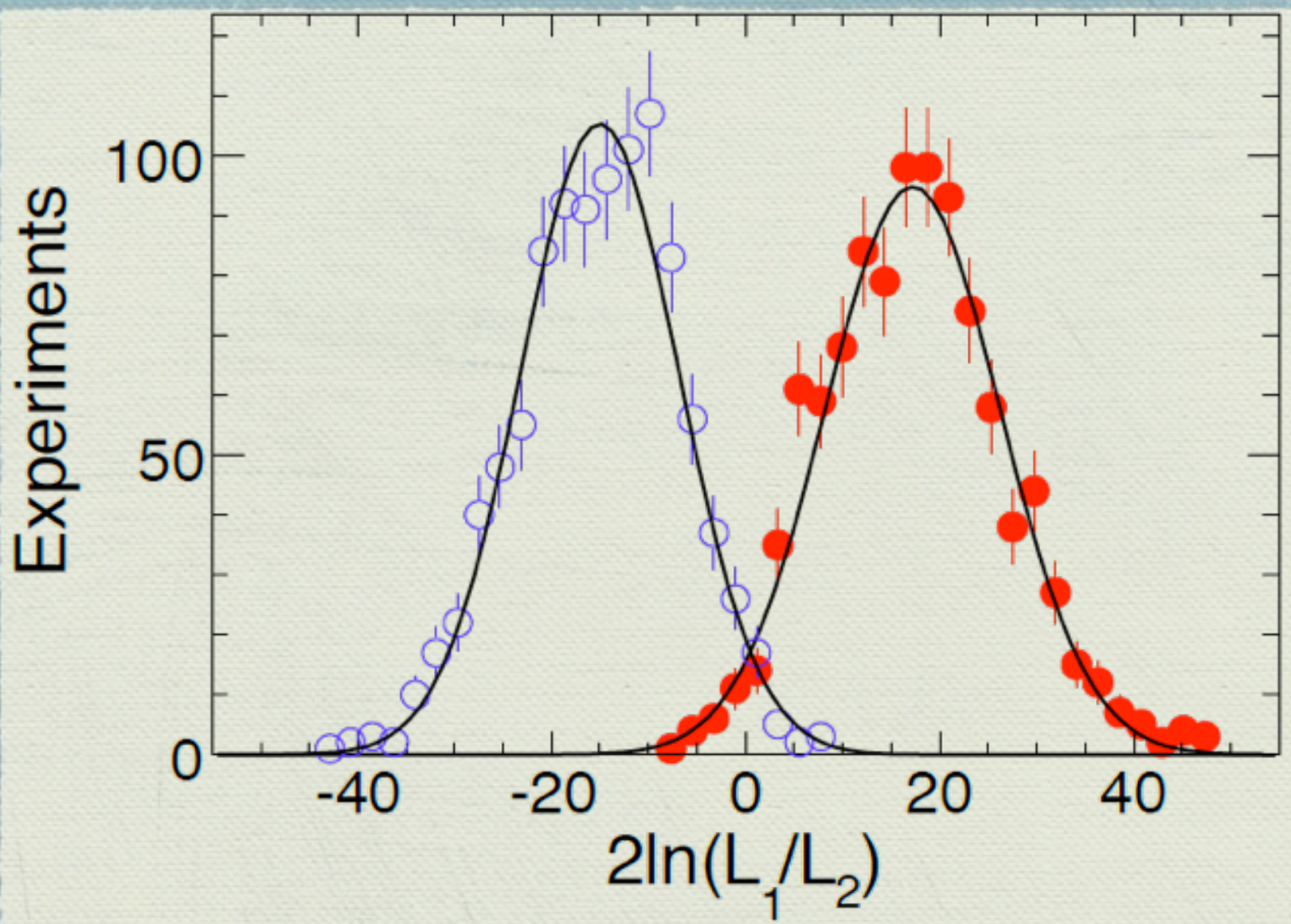
Likelihood Ratios



Gao, Gritsan, Guo, Melnikov, Schulze, Tran (2010)

- L_1 is likelihood for 0^+ Higgs hypothesis
- L_2: 0^- Higgs hypothesis
- 1000 pseudo-experiments using 30 events each
- Generated according to 0^+ hypothesis (red), 0^- hypothesis (blue)

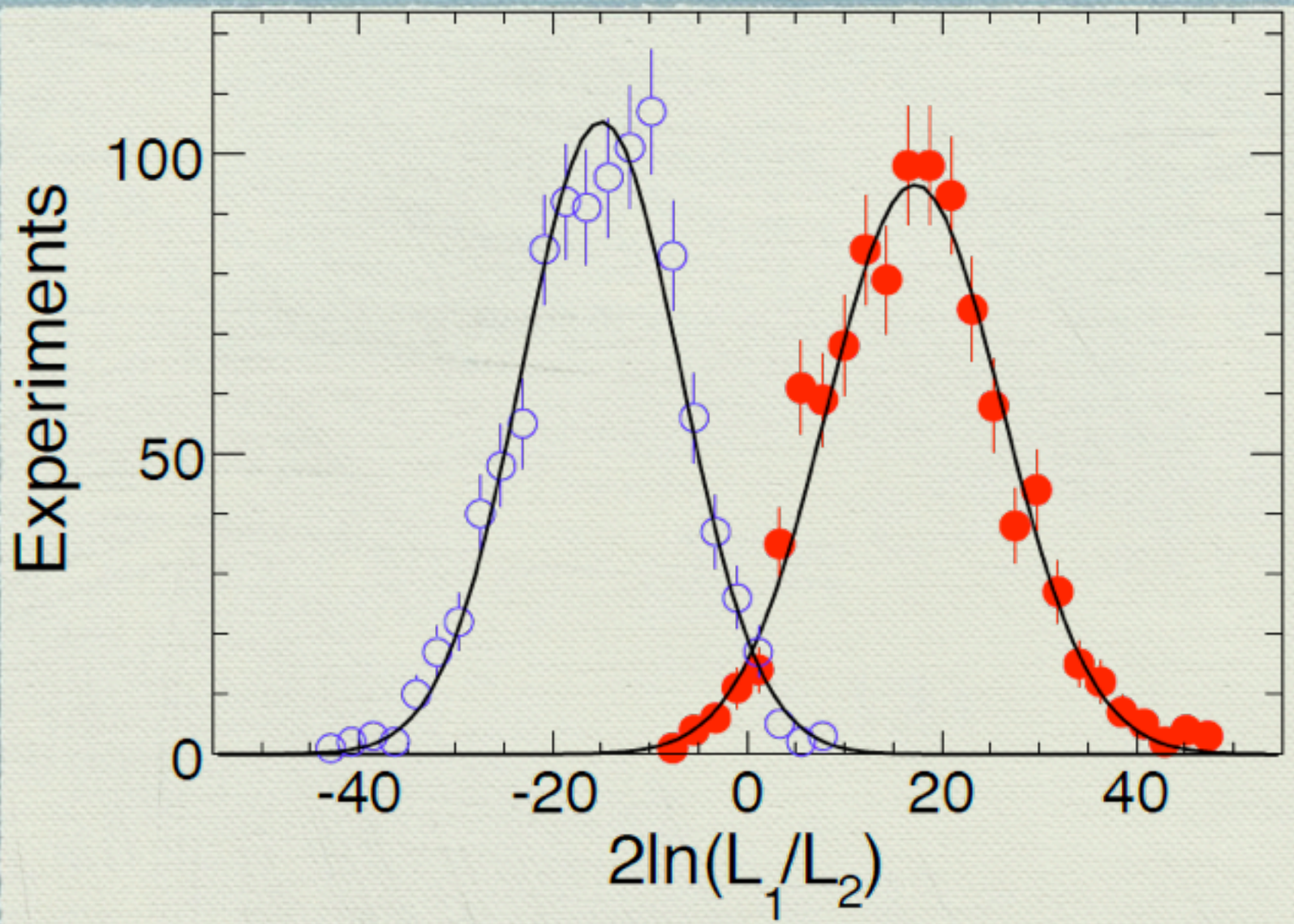
Likelihood Ratios



Gao, Gritsan, Guo, Melnikov, Schulze, Tran (2010)

- If we wanted to "assign" observed events to 0^+ or 0^- we can pick a cutoff value of $2\ln(L_1/L_2)$

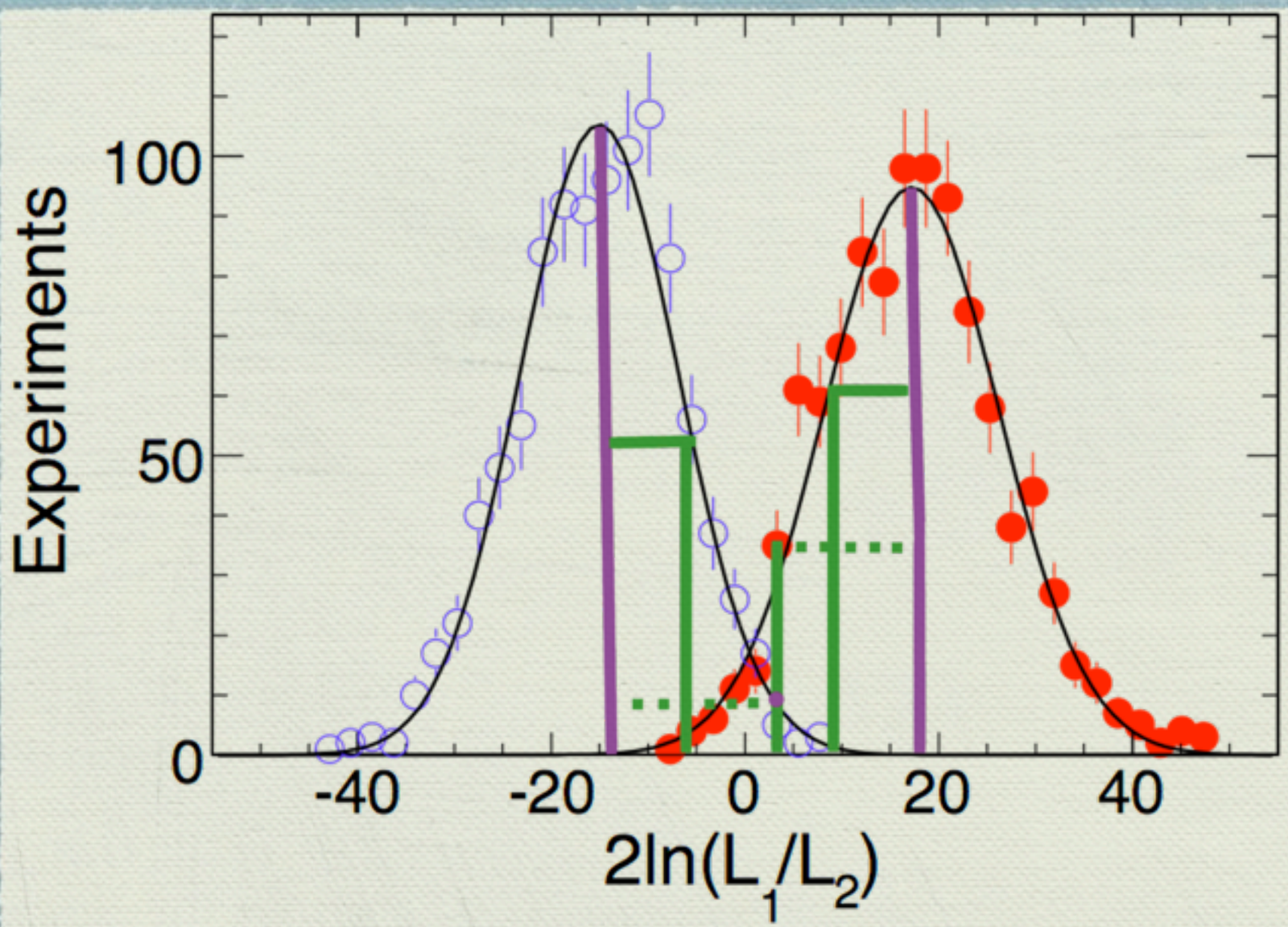
Likelihood Ratios



Gao, Gritsan, Guo, Melnikov, Schulze, Tran (2010)

- In physics we are not just trying to "pick" a hypothesis
- We want to quantify the extent to which one hypothesis is preferred by the data

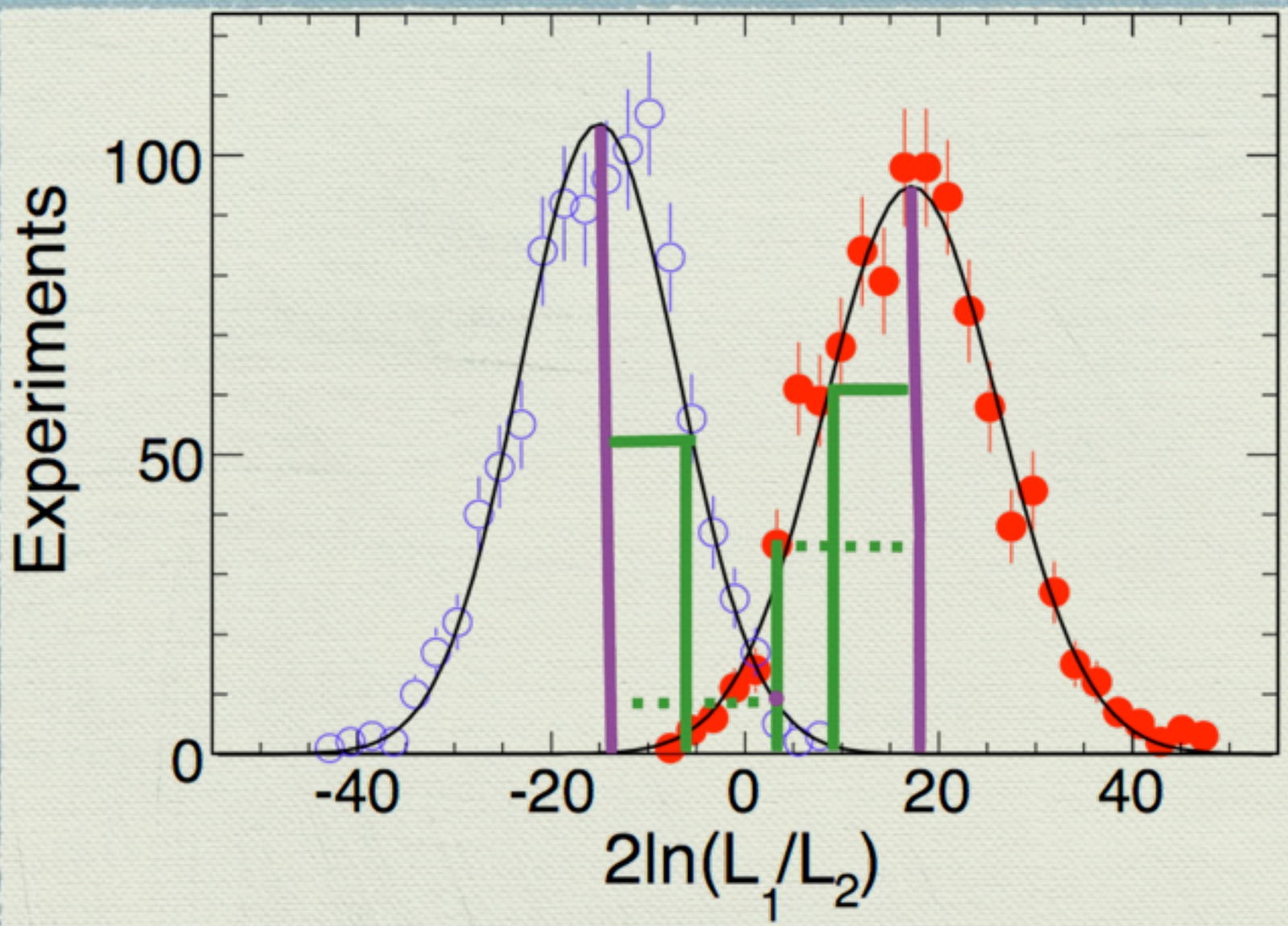
Likelihood Ratios



If the measured value of the test statistic, 2ln(L1/L2), is 3...

- Distributions of the test statistic give us p-values for each hypothesis as well as the relative likelihood that the underlying model is e.g. 0^+ or 0^-

Likelihood Ratios

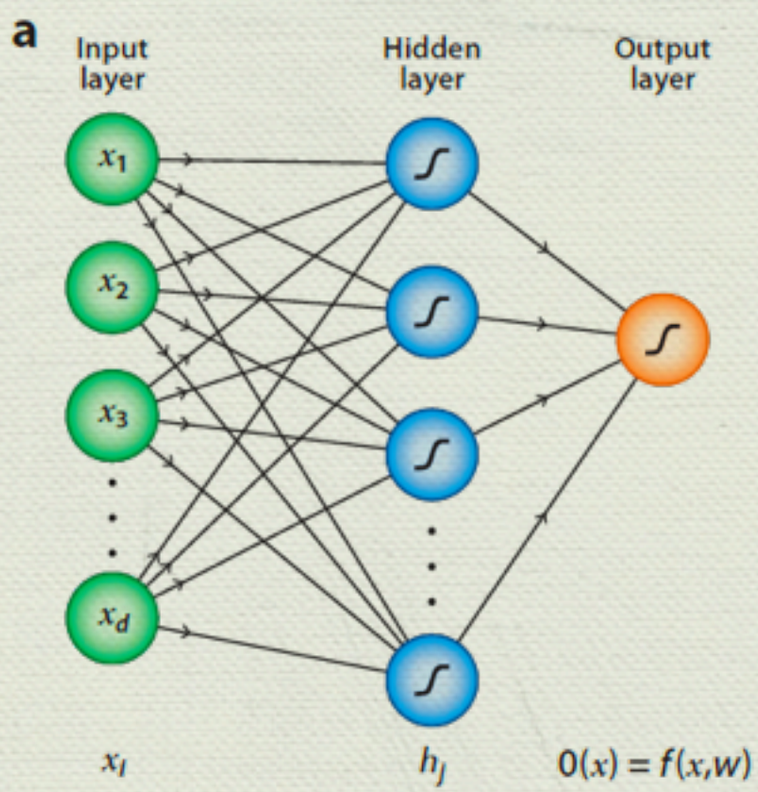


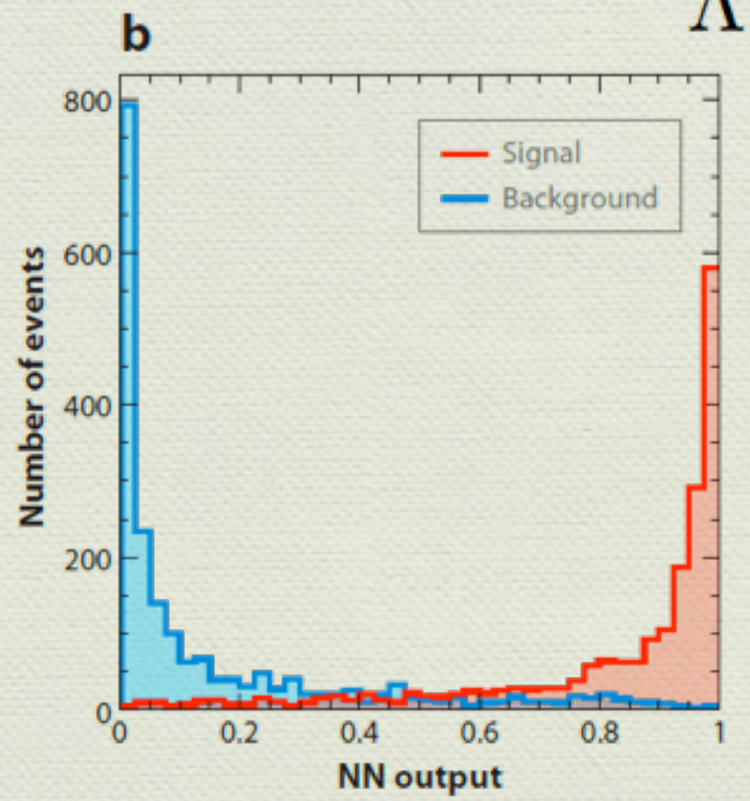
- In principle, we should know the p-value directly from the likelihood
- In practice, using pseudo-experiments to determine the p-value accounts for approximations in the likelihood (higher order, detector effects, backgrounds not included in the likelihood)

MVAs / Likelihood

- Various MVA (Bayesian Neural Nets, Boosted Decision Trees, etc.) are essentially techniques to efficiently approximate the likelihood functions in the discriminant

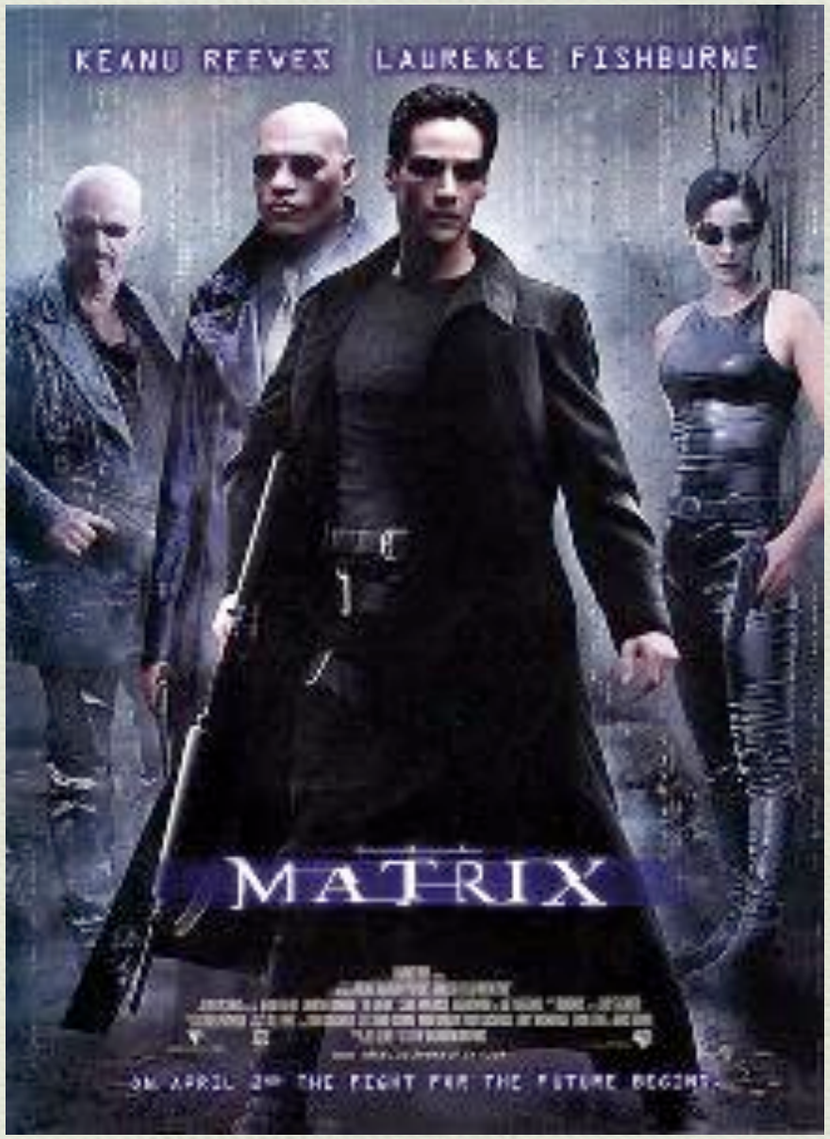
$$\Lambda(E) = \frac{L(H_1 \mid E)}{L(H_0 \mid E)}$$





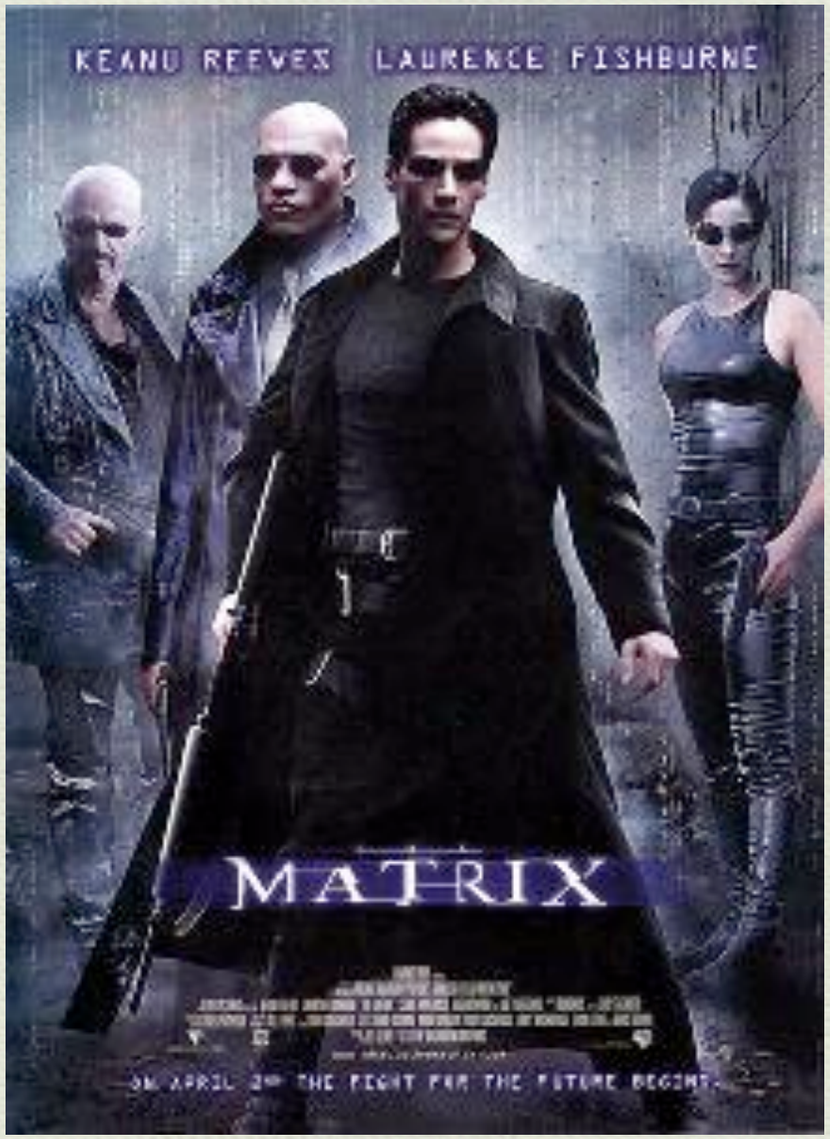
Bhat: Annu. Rev. Nucl. Part. Sci. 2011. 61:281–309

Matrix Element Method



- In the Matrix Element Method (MEM), the likelihoods are calculated analytically
- Likelihood / probability is the differential cross section (up to normalization)

Matrix Element Method



Pros

- Transparency: the likelihood is the differential cross section, not the result of a machine learning process
- Flexibility: Since the likelihood is the differential cross section, one already has the likelihood for any point in the parameter space
- Role for theorists?* We know a thing or two about cross sections.

MEM in Golden Channel

$$L_{\mu\alpha} = X\, g_{\mu\alpha} - (Y + i\,Z)\,\frac{k_\alpha k_\mu}{M_Z^2} + (P + i\,Q)\,\epsilon_{\mu\alpha}\frac{p_1 p_2}{M_Z^2}$$

$$\begin{aligned} d\Gamma[0] \quad &= X^2\; d\Gamma[0^+] + (P^2 + Q^2)\; d\Gamma[0^-] \\ &+ X\,P\,d\Gamma[0,\,\mathrm{Todd}] + X\,Q\,d\Gamma[0,\,\mathrm{Codd}] \\ &+ X\,Y\,d\Gamma_{XY} + (Y^2 + Z^2)\,d\Gamma_{YY}\ . \end{aligned}$$

$$\begin{aligned} \frac{d\Gamma[0^+]}{dc_1\,dc_2\,d\phi} &\propto m_1^2\,m_2^2\,m_H^4\big[1 + c_1^2 c_2^2 + (\gamma_b^2 + c^2)s_1^2 s_2^2 \\ &+ 2\gamma_a\,c\,s_1 s_2\,c_1 c_2 + 2\eta^2(c_1 c_2 + \gamma_a\,c\,s_1 s_2)\big] \end{aligned}$$

$$\begin{aligned} \Delta d\Gamma[0] \;&=\; X\,Z\,d\Gamma_{XZ} + Y\,P\,d\Gamma_{YP} \\ &+\; Y\,Q\,d\Gamma_{YQ} + Z\,P\,d\Gamma_{ZP} + Z\,Q\,d\Gamma_{ZQ}\ ; \end{aligned}$$

• • •

De Rújula, Lykken, Pierini, Rogan, Spiropulu (2010)

$$A(X \to VV) = v^{-1}\epsilon_1^{*\mu}\epsilon_2^{*\nu}\left(a_1 g_{\mu\nu} m_X^2 + a_2\,q_\mu q_\nu + a_3\epsilon_{\mu\nu\alpha\beta}\,q_1^\alpha q_2^\beta\right)$$

$$\begin{aligned} A_{00} &= -\frac{m_X^4}{4v m_V^2}\left(a_1(1+\beta^2) + a_2\beta^2\right)\ , \\ A_{++} &= \frac{m_X^2}{v}\left(a_1 + \frac{i a_3 \beta}{2}\right)\ , \\ A_{--} &= \frac{m_X^2}{v}\left(a_1 - \frac{i a_3 \beta}{2}\right)\ . \end{aligned}$$

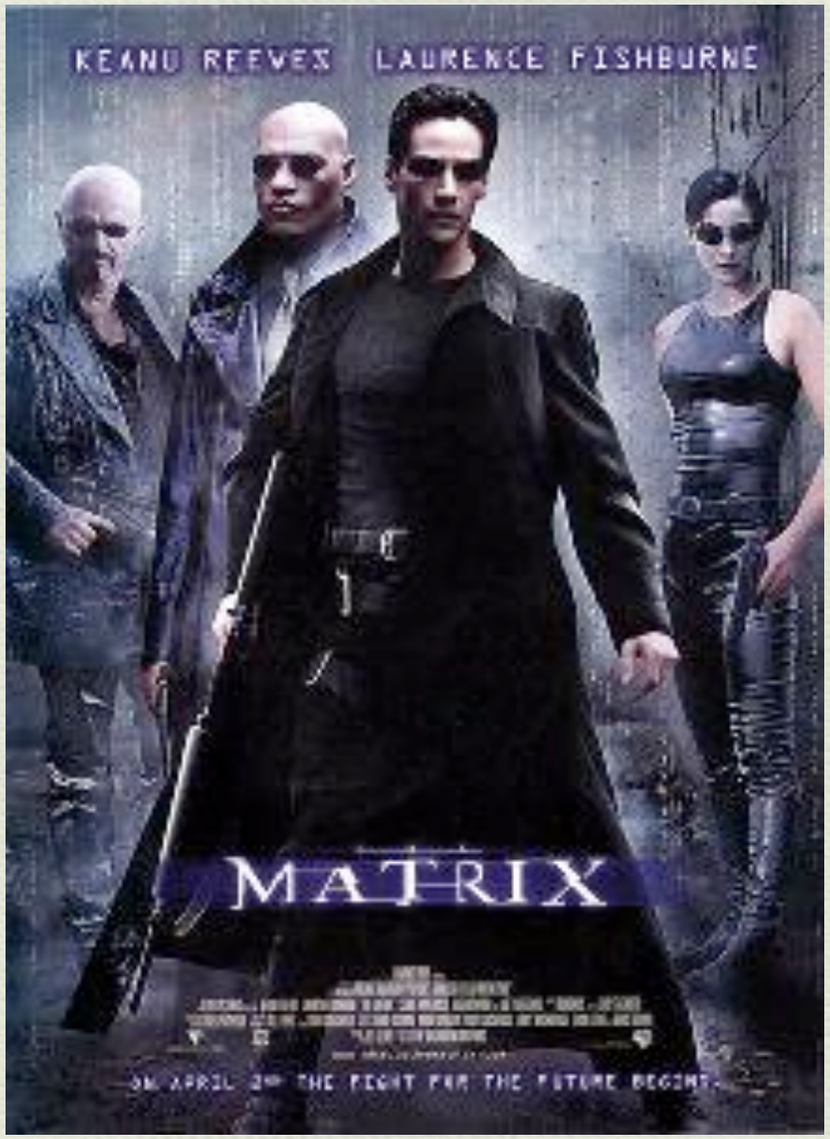
$$\begin{aligned} \frac{128\pi\, d\Gamma_{J=0}}{9\,\Gamma\, d\cos\theta_1\, d\cos\theta_2 d\Phi} \;=\; & 4\,(1 - f_{++} - f_{--})\,\sin^2\theta_1 \sin^2\theta_2 \\ & + (f_{++} + f_{--})\left((1+\cos^2\theta_1)(1+\cos^2\theta_2) + 4R_1 R_2 \cos\theta_1\cos\theta_2\right) \\ & - 2\,(f_{++} - f_{--})\left(R_1\cos\theta_1(1+\cos^2\theta_2) + R_2(1+\cos^2\theta_1)\cos\theta_2\right) \\ & + 4\sqrt{f_{++}(1 - f_{++} - f_{--})}\,(R_1 - \cos\theta_1)\sin\theta_1(R_2 - \cos\theta_2)\sin\theta_2\cos(\Phi + \phi_{++}) \\ & + 4\sqrt{f_{--}(1 - f_{++} - f_{--})}\,(R_1 + \cos\theta_1)\sin\theta_1(R_2 + \cos\theta_2)\sin\theta_2\cos(\Phi - \phi_{--}) \\ & + 2\sqrt{f_{++} f_{--}}\,\sin^2\theta_1\sin^2\theta_2\cos(2\Phi + \phi_{++} - \phi_{--})\ . \end{aligned}$$

• • •

Gao, Gritsan, Guo, Melnikov, Schulze, Tran (2010)

All information necessary to calculate likelihood is found in tractable, analytic formulae.

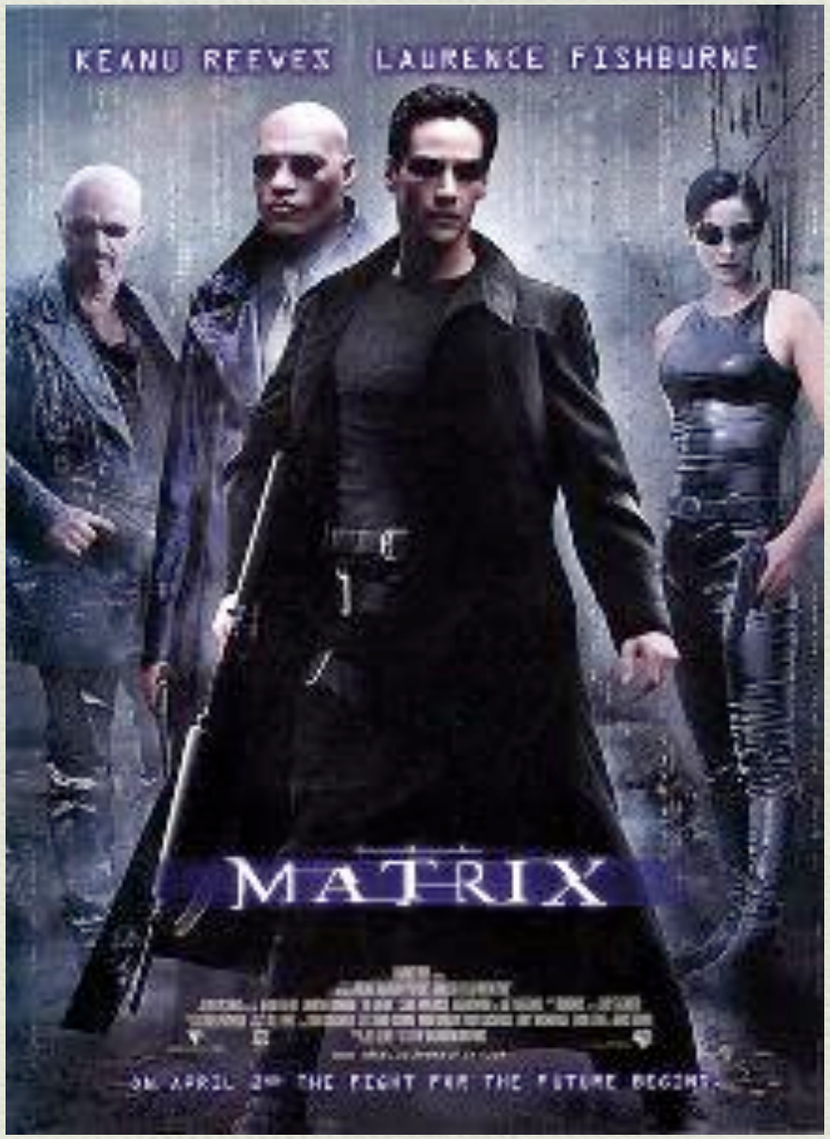
Matrix Element Method



Cons

- Can be very computationally intensive
- Especially when
  - taking into account finite detector resolution
    (through transfer functions)
  - integrating over unobserved particles (neutrinos, etc.)
- Remember: need to perform procedure for zillions of pseudo-experiments, not just for observed signal-like events

Matrix Element Method



Cons

- Downsides minimized
  - for leptonic final states
  - when final state fully reconstructed
  - e.g. $H \to ZZ^* \to 4l$

MEM for Higgs: Theory

MEM for Higgs (couplings or extracting signal)

- ZZ: ZZ: De Rújula, Lykken, Pierini, Rogan, Spiropulu (2010)
- ZZ: JG, Kumar, Low, Vega-Morales (2011)
- Zγ: JG, Keung, Low, Schwaller (2011)
- ZZ (NLO) x 2: Campbell, Giele, Williams (2012)
- ZZ: Bolognesi, Gao, Gritsan, Melnikov, Schulze, Tran, Whitbeck (2012)
- ZZ: Stolarski, Vega-Morales (2012)
- ZZ: Avery, Bourilkov, Chen, Cheng, Drozdetskiy, JG, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, Snowball (2012)
- W+W+jj: Freitas, JG (2012) [Talk Wednesday, Thursday]
- Zγ (NLO): Campbell, Ellis, Giele, Williams (2013)
- . . .

Packages and Frameworks

MELA, JHUGen
Gao, Gritsan, Guo, Melnikov, Schulze, Tran (2010)
Bolognesi, Gao, Gritsan, Melnikov, Schulze, Tran, Whitbeck (2012)
NLOME
Campbell, Giele, Williams (2012)
MEKD*
Avery, Bourilkov, Chen, Cheng, Drozdetskiy, JG, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, Snowball (2012)
MadWeight*
Artoisenet, Mattelaer (2008)
Artoisenet, Lemaitre, Maltoni, Mattelaer (2010)
***MadGraph based**
Alwall, Herquet, Maltoni, Mattelaer, Stelzer (2011)
Alwall, Demin, de Visscher, Frederix, Herquet, Maltoni, Plehn, Rainwater, Stelzer (2007); etc.

MEM for Higgs: Theory

MEM for Higgs (couplings or extracting signal)

- ZZ: ZZ: De Rújula, Lykken, Pierini, Rogan, Spiropulu (2010)
- ZZ: JG, Kumar, Low, Vega-Morales (2011)
- Zγ: JG, Keung, Low, Schwaller (2011)
- ZZ (NLO) x 2: Campbell, Giele, Williams (2012)
- ZZ: Bolognesi, Gao, Gritsan, Melnikov, Schulze, Tran, Whitbeck (2012)
- ZZ: Stolarski, Vega-Morales (2012)
- ZZ: Avery, Bourilkov, Chen, Cheng, Drozdetskiy, JG, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, Snowball (2012)
- W+W+jj: Freitas, JG (2012) [Talk Wednesday Afternoon]
- Zγ (NLO): Campbell, Ellis, Giele, Williams (2013)
- . . .

NLO/ Extra Radiation

Studies of the Matrix Element Method at NLO/ additional radiation/ showering are a good example of how theorists can contribute

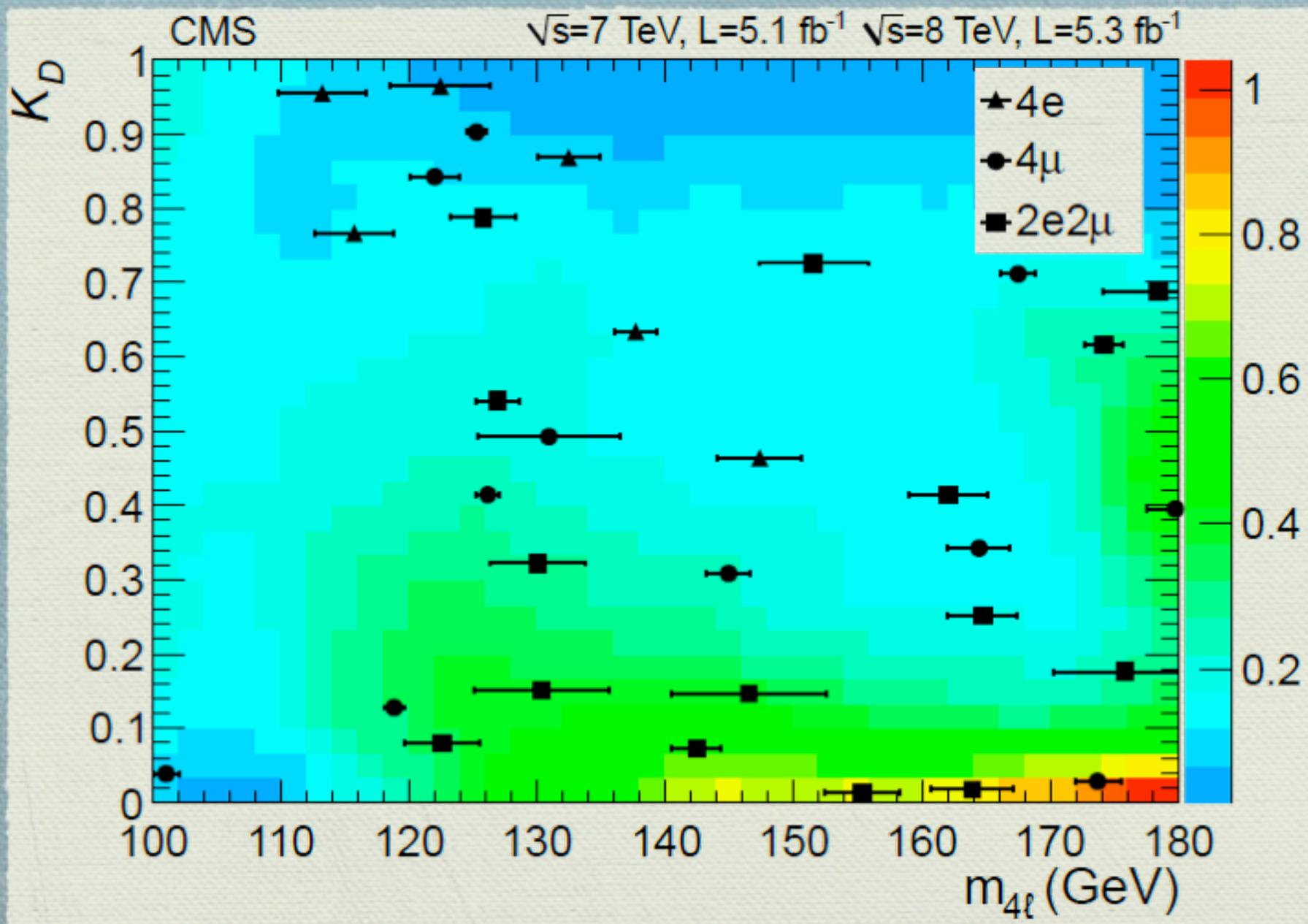
Zγ (NLO): Campbell, Ellis, Giele, Williams (2013)

ZZ (NLO) x 2: Campbell, Giele, Williams (2012)

Alwall, Freitas, Mattelaer (2010)

Soper, Spannowsky (2011)

Discovery!



Phys.Lett. B716 (2012) 30-61, CMS-PAS-HIG-12-016

- CMS used MEM: MELA to separate H→ZZ*→4l signal from background
- One of 2 major channels in July 4 Higgs discovery

Properties Measurements with MVA

$H \to ZZ^* \to 4l$

J^P	production	comment	expect (μ=1)	obs. 0^+	obs. J^P	CL_s
0^-	$gg \to X$	pseudoscalar	2.6σ (2.8σ)	0.5σ	3.3σ	0.16%
0^+_h	$gg \to X$	higher dim operators	1.7σ (1.8σ)	0.0σ	1.7σ	8.1%
2^+_{mgg}	$gg \to X$	minimal couplings	1.8σ (1.9σ)	0.8σ	2.7σ	1.5%
$2^+_{mq\bar{q}}$	$q\bar{q} \to X$	minimal couplings	1.7σ (1.9σ)	1.8σ	4.0σ	<0.1%
1^-	$q\bar{q} \to X$	exotic vector	2.8σ (3.1σ)	1.4σ	$>4.0\sigma$	<0.1%
1^+	$q\bar{q} \to X$	exotic pseudovector	2.3σ (2.6σ)	1.7σ	$>4.0\sigma$	<0.1%

MEM using MELA, JHUGen, MEKD

		BDT analysis				J^P-MELA analysis			
		tested J^P for an assumed 0^+		tested 0^+ for an assumed J^P	CL_S	tested J^P for an assumed 0^+		tested 0^+ for an assumed J^P	CL_S
		expected	observed	observed*		expected	observed	observed*	
0^-	p_0	0.0037	0.015	0.31	0.022	0.0011	0.0022	0.40	0.004
1^+	p_0	0.0016	0.001	0.55	0.002	0.0031	0.0028	0.51	0.006
1^-	p_0	0.0038	0.051	0.15	0.060	0.0010	0.027	0.11	0.031
2^+_m	p_0	0.092	0.079	0.53	0.168	0.064	0.11	0.38	0.182
2^-	p_0	0.0053	0.25	0.034	0.258	0.0032	0.11	0.08	0.116

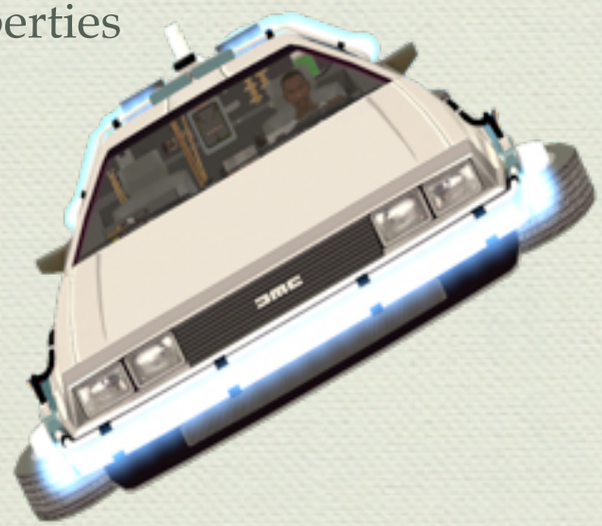
BDT/MEM using MELA/ JHUGen.

- Both ATLAS and CMS also use likelihood-based MVA for spin measurements in $H \to WW^* \to 2l\ 2\nu$ and $H \to \gamma\gamma$

The Future?

$$\mathcal{L} = X\left[\kappa_1 \frac{m_Z^2}{v} Z_\mu Z^\mu + \frac{\kappa_2}{2v} F_{\mu\nu}F^{\mu\nu} + \frac{\kappa_3}{2v} F_{\mu\nu}\tilde{F}^{\mu\nu}\right]$$

- Lowest dimensional HZZ couplings with three different symmetry properties
- Much work to measure or constrain all three couplings simultaneously
  - Very challenging measurement for likely values of parameters
  - CMS: f_{a3}
  - Important challenge for LHC and beyond



- Also interesting to measure or constrain other dimension-5 (non-gauge invariant) operators
- and to constrain other spin hypotheses in greater generality
- Similar story in other channels, though HZZ is very well suited for these sorts of measurements at the LHC
- The Matrix Element Method, among other MVA likelihood methods, will be an important part of this story