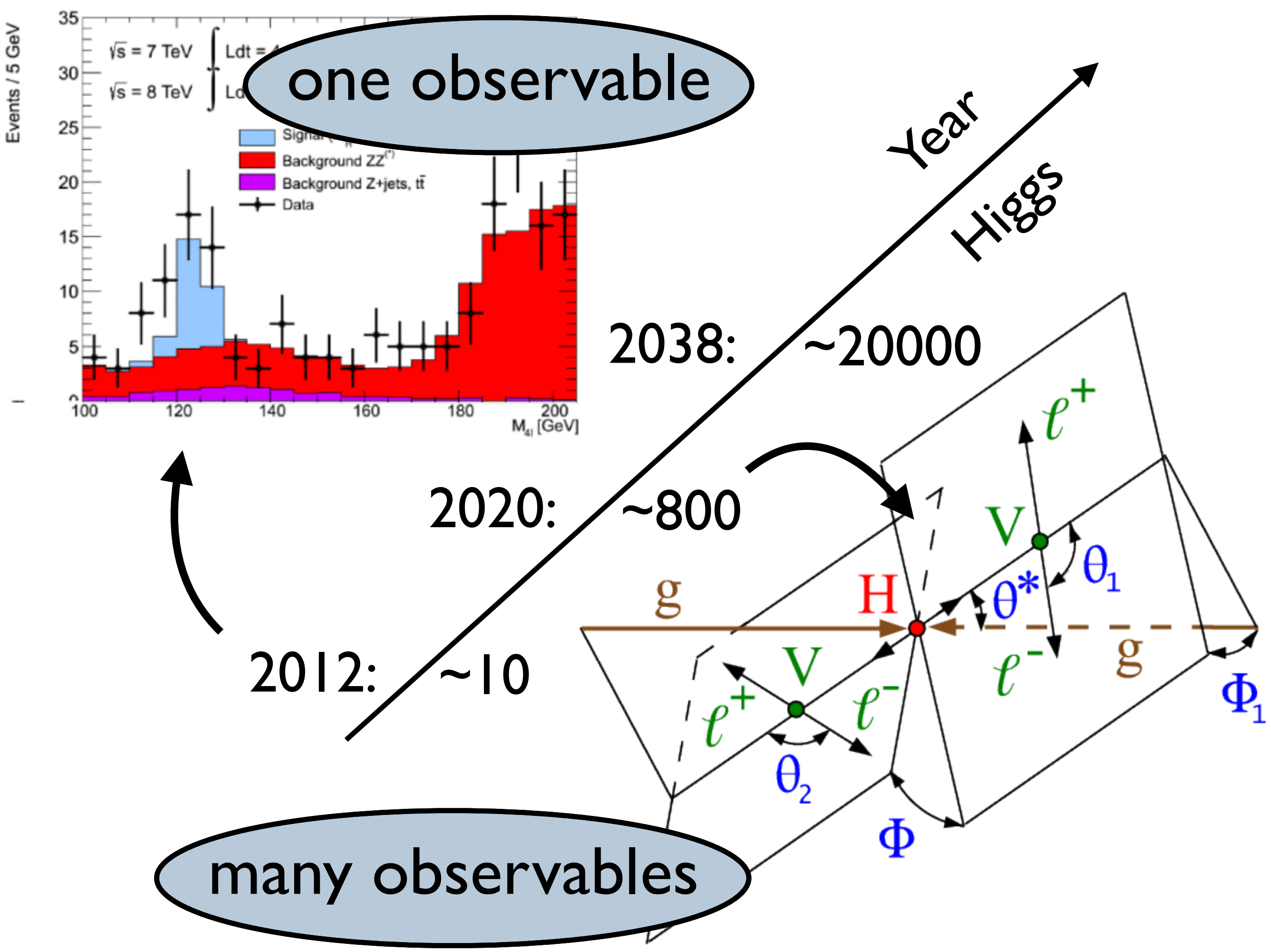
Theory perspective on new analysis and theoretical techniques

Felix Kling
felixk@slac.stanford.edu

SLAC NATIONAL ACCELERATOR LABORATORY

30th October 2020
Higgs 2020

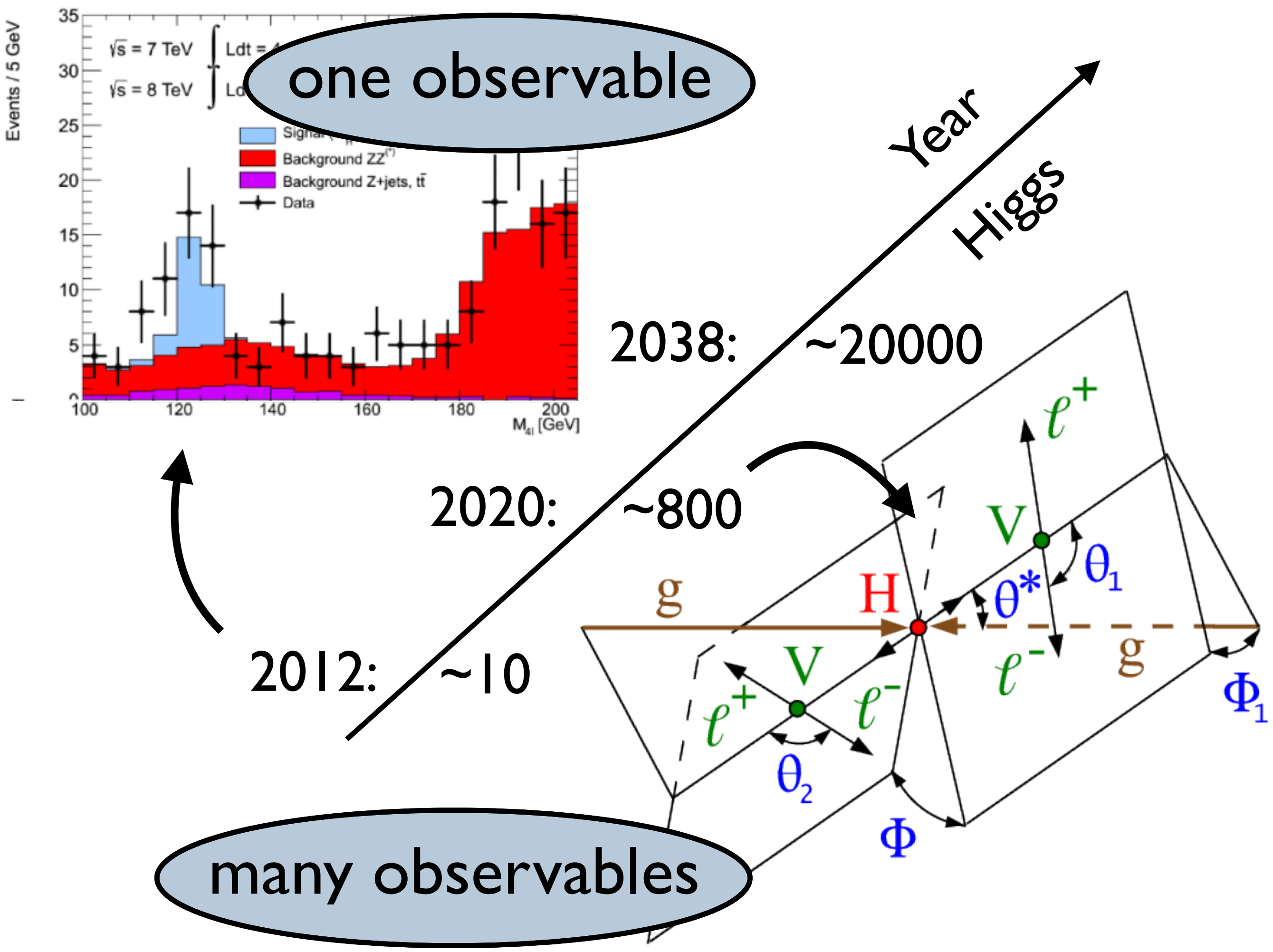
Theory in an Era of Data

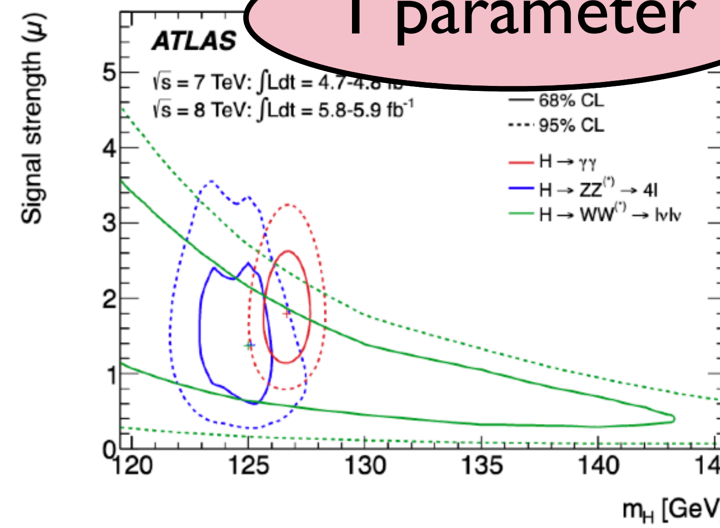


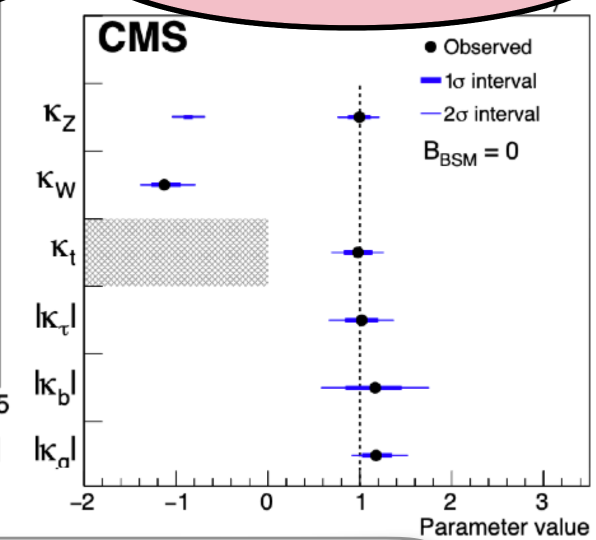
We live in an era of Data

large statistics, many kinematic features

Theory in an Era of Data







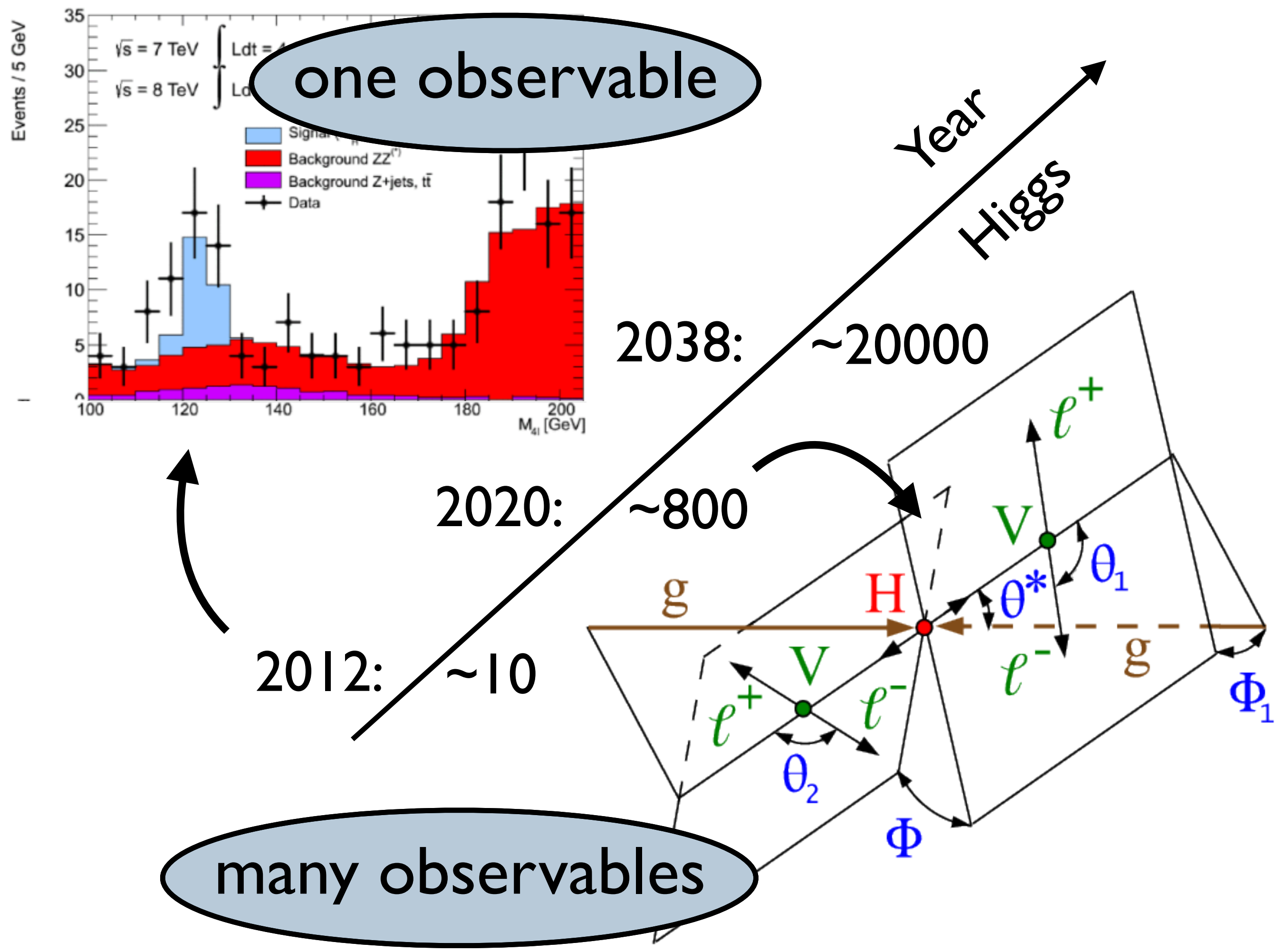
$$\begin{aligned}\Delta\mathcal{L}_{SILH} =\ & \frac{\bar{c}_H}{2v^2}\partial^\mu\left(H^\dagger H\right)\partial_\mu\left(H^\dagger H\right) + \frac{\bar{c}_T}{2v^2}\left(H^\dagger \overleftrightarrow{D^\mu} H\right)\left(H^\dagger \overleftrightarrow{D}_\mu H\right) - \frac{\bar{c}_6\lambda}{v^2}\left(H^\dagger H\right)^3 \\ & + \left(\left(\frac{\bar{c}_u}{v^2} y_u H^\dagger H \bar{q}_L H^c u_R + \frac{\bar{c}_d}{v^2} y_d H^\dagger H \bar{q}_L H d_R + \frac{\bar{c}_l}{v^2} y_l H^\dagger H \bar{L}_L H l_R\right) + \text{h.c.}\right) \\ & + \frac{i\bar{c}_W g}{2m_W^2}\left(H^\dagger \sigma^i \overleftrightarrow{D^\mu} H\right)\left(D^\nu W_{\mu\nu}\right)^i + \frac{i\bar{c}_B g'}{2m_W^2}\left(H^\dagger \overleftrightarrow{D^\mu} H\right)\left(\partial^\nu B_{\mu\nu}\right) \\ & + \frac{i\bar{c}_{HW} g}{m_W^2}\left(D^\mu H\right)^\dagger \sigma^i \left(D^\nu H\right) W^i_{\mu\nu} + \frac{i\bar{c}_{HB} g'}{m_W^2}\left(D^\mu H\right)^\dagger \left(D^\nu H\right) B_{\mu\nu} \\ & + \frac{\bar{c}_\gamma g'^2}{m_W^2} H^\dagger H B_{\mu\nu} B^{\mu\nu} + \frac{\bar{c}_g g_S^2}{m_W^2} H^\dagger H G^a_{\mu\nu} G^{a\mu\nu}\end{aligned}$$

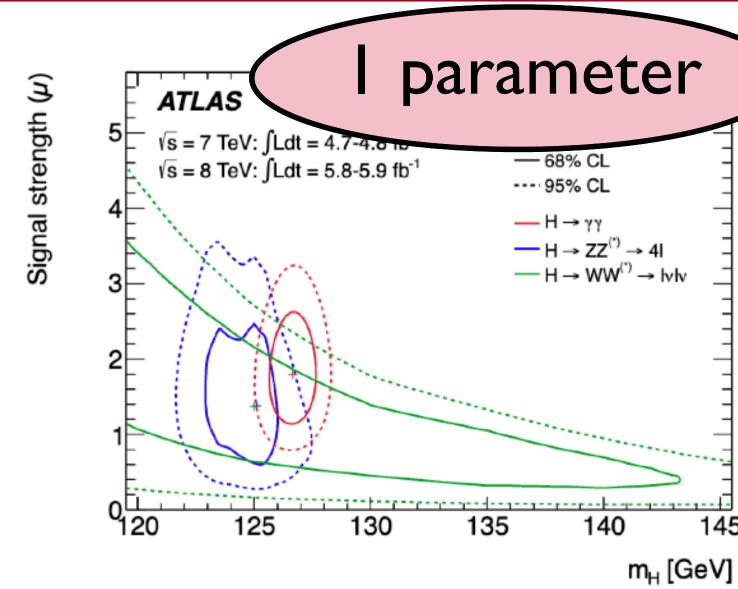
many parameters

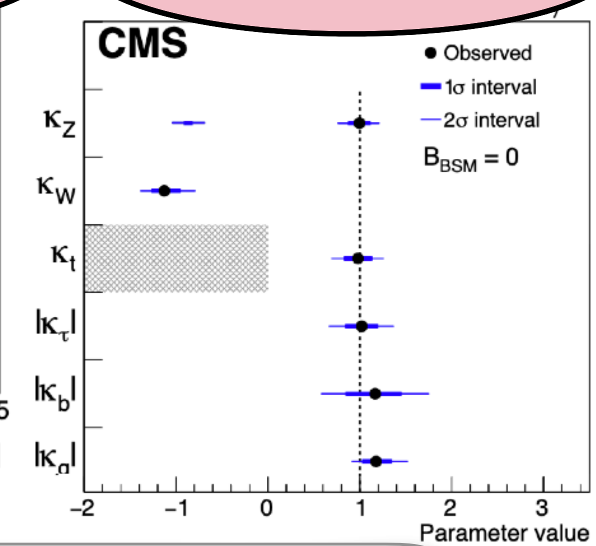
We live in an era of Data
large statistics, many kinematic features

High-dimensional theories
e.g. SM effective field theory

Theory in an Era of Data







$$
\begin{aligned}
\Delta\mathcal{L}_{SILH} =\ & \frac{\bar{c}_H}{2v^2}\partial^\mu\left(H^\dagger H\right)\partial_\mu\left(H^\dagger H\right) + \frac{\bar{c}_T}{2v^2}\left(H^\dagger \overleftrightarrow{D^\mu} H\right)\left(H^\dagger \overleftrightarrow{D}_\mu H\right) - \frac{\bar{c}_6\lambda}{v^2}\left(H^\dagger H\right)^3 \\
& + \left(\left(\frac{\bar{c}_u}{v^2} y_u H^\dagger H \bar{q}_L H^c u_R + \frac{\bar{c}_d}{v^2} y_d H^\dagger H \bar{q}_L H d_R + \frac{\bar{c}_l}{v^2} y_l H^\dagger H \bar{L}_L H l_R\right) + \text{h.c.}\right) \\
& + \frac{i\bar{c}_W g}{2m_W^2}\left(H^\dagger \sigma^i \overleftrightarrow{D^\mu} H\right)\left(D^\nu W_{\mu\nu}\right)^i + \frac{i\bar{c}_B g'}{2m_W^2}\left(H^\dagger \overleftrightarrow{D^\mu} H\right)\left(\partial^\nu B_{\mu\nu}\right) \\
& + \frac{i\bar{c}_{HW} g}{m_W^2}\left(D^\mu H\right)^\dagger \sigma^i \left(D^\nu H\right) W^i_{\mu\nu} + \frac{i\bar{c}_{HB} g'}{m_W^2}\left(D^\mu H\right)^\dagger \left(D^\nu H\right) B_{\mu\nu} \\
& + \frac{\bar{c}_\gamma g'^2}{m_W^2} H^\dagger H B_{\mu\nu} B^{\mu\nu} + \frac{\bar{c}_g g_S^2}{m_W^2} H^\dagger H G^a_{\mu\nu} G^{a\mu\nu}
\end{aligned}
$$

many parameters

We live in an era of Data
large statistics, many kinematic features

High-dimensional theories
e.g. SM effective field theory

Each process is typically sensitive to several Higgs couplings which effect a variety of kinematic distributions. We naturally need multivariate analyses.

How can we extract all the information from the data?

The Likelihood Function

θ → $p(x|\theta)$ simulation → $\mathbf{x}$ → $p(x|\theta)$ inference → $\hat{\theta} \pm \Delta\theta$

θ	$p(x\|\theta)$	$\mathbf{x}$	$p(x\|\theta)$	$\hat{\theta} \pm \Delta\theta$
theory parameters	simulation	observables	inference	estimator/uncertainty

The Likelihood Function

θ —— $p(x|\theta)$ ——→ x —— $p(x|\theta)$ ——→ $\hat{\theta} \pm \Delta\theta$

theory parameters — simulation — observables — inference — estimator/uncertainty

Likelihood Function $p(x|\theta)$

likelihood of an observation x as a function of the theory parameter θ

Likelihood Ratio $r(x|\theta_{ref},\theta)=p(x|\theta)/p(x|\theta_{ref})$

"how much more likely is data x described by theory θ than θ_{ref}"

Neyman-Pearson Lemma

The log-likelihood ratio log $r(x|\theta_{ref},\theta)$ is the most powerful test statistic to discriminate between two hypotheses θ_{ref} and θ.

To optimize Higgs measurements, we need to know $p(x|\theta)$

The Likelihood Function

θ → $p(x|\theta)$ → $\mathbf{x}$ → $p(x|\theta)$ → $\hat{\theta} \pm \Delta\theta$

theory parameters — simulation — observables — inference — estimator/uncertainty

Likelihood Function $p(x|\theta)$

likelihood of an observation x as a function of the theory parameter θ

Likelihood Ratio $r(x|\theta_{ref},\theta)=p(x|\theta)/p(x|\theta_{ref})$

"how much more likely is data x described by theory θ than θ_{ref}"

Neyman-Pearson Lemma

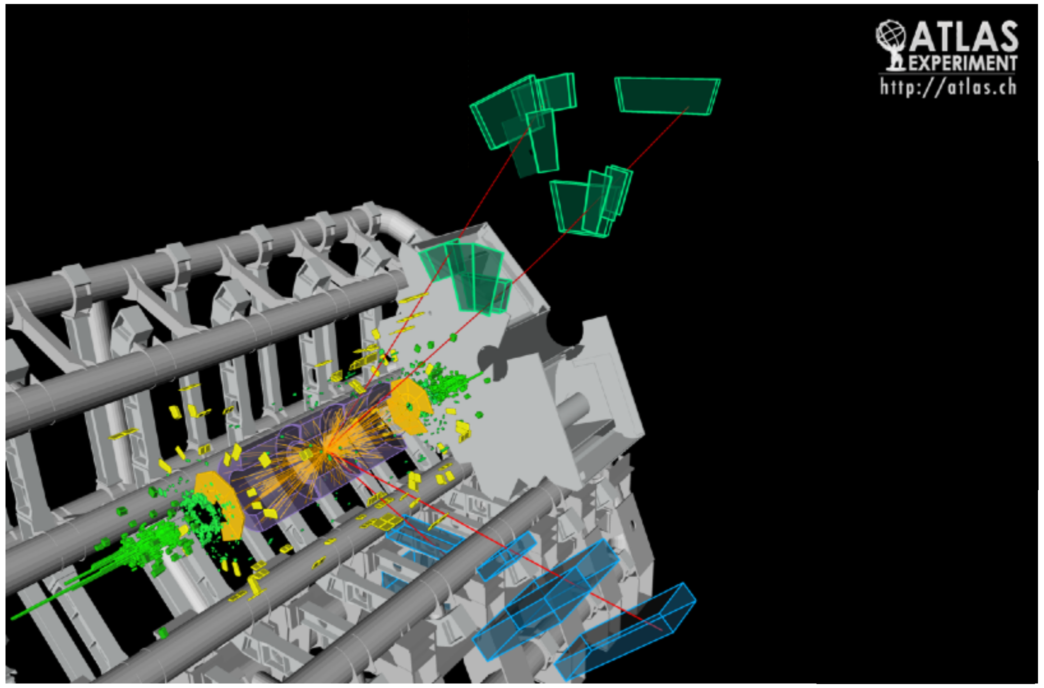
The log-likelihood ratio log $r(x|\theta_{ref},\theta)$ is the most powerful test statistic to discriminate between two hypotheses θ_{ref} and θ.

To optimize Higgs measurements, we need to know $p(x|\theta)$

Our simulation tools obtain events distributed according to $p(x|\theta)$.

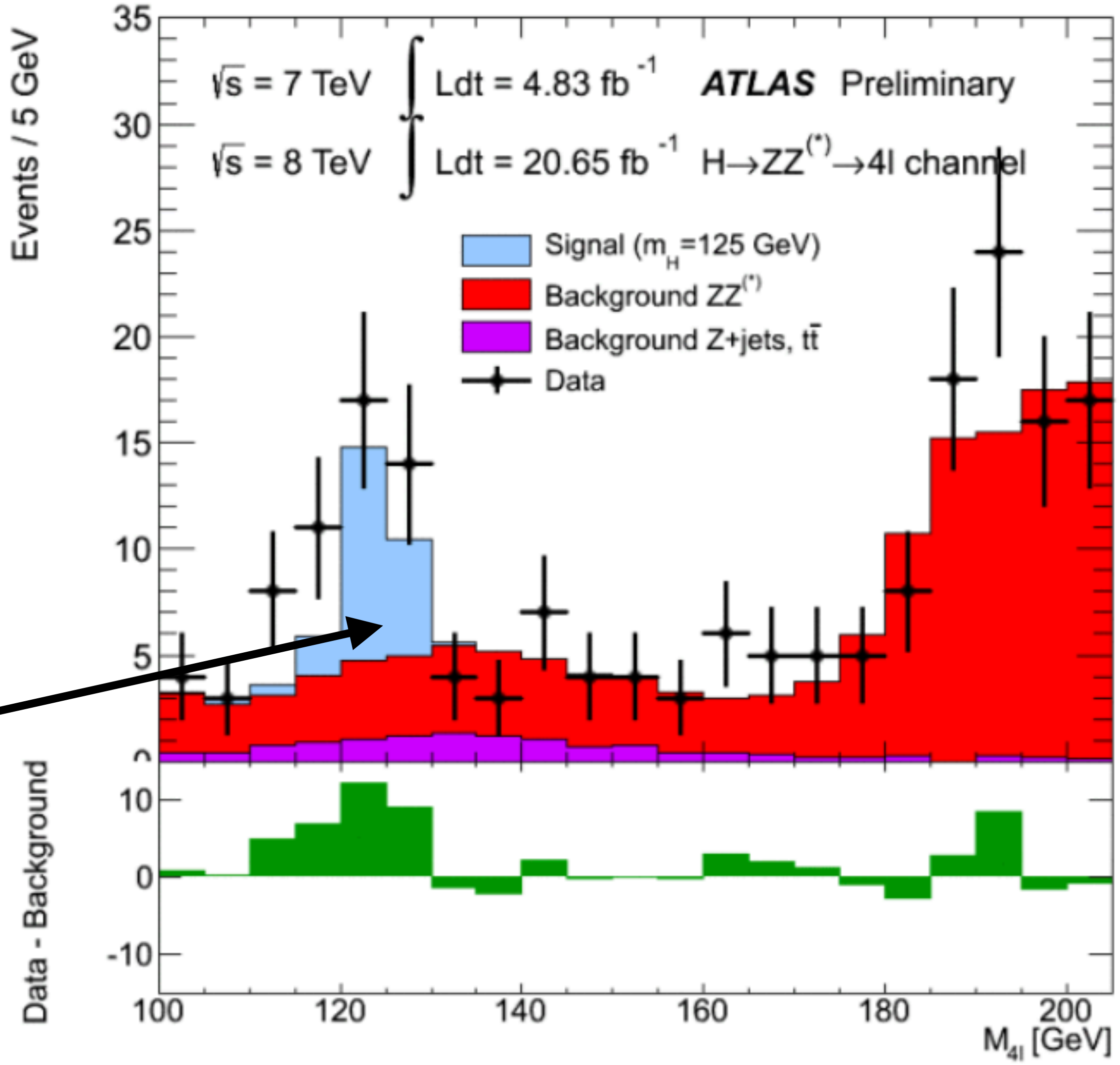
However, we cannot calculate $p(x|\theta)$. To do inference, we need to estimate $p(x|\theta)$.

Traditional Method: Histograms



complex data: 10^8 sensors

analyze one carefully
chosen summary statistics



invariant mass: $m(4\mu)$

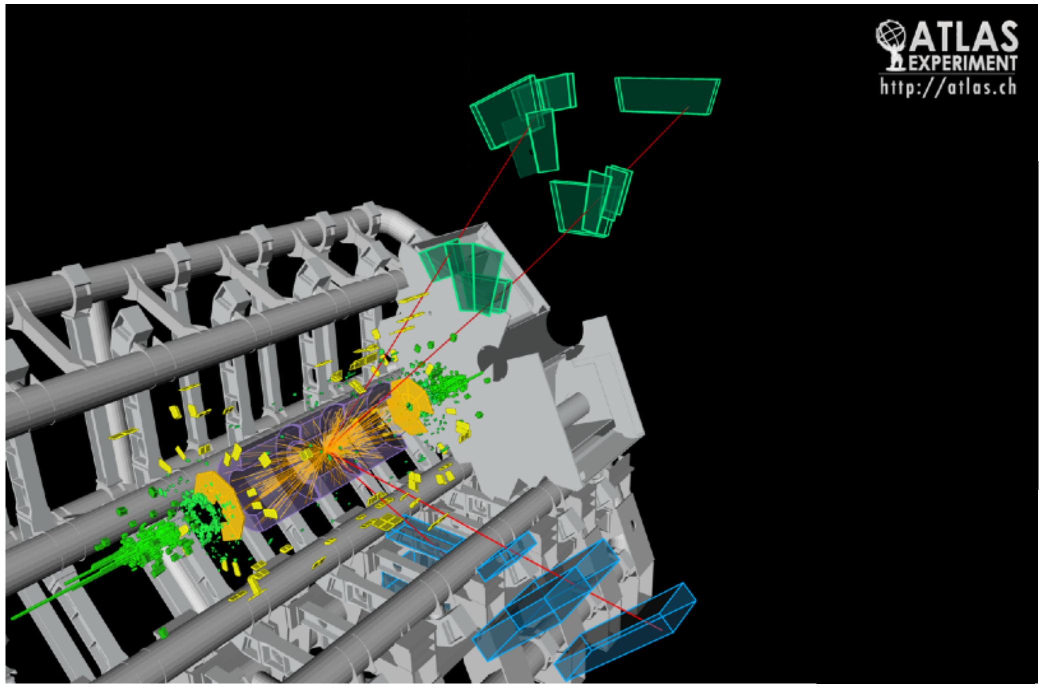
+ very easy

– restricted to view observables

– ignores all the information in rest of the data

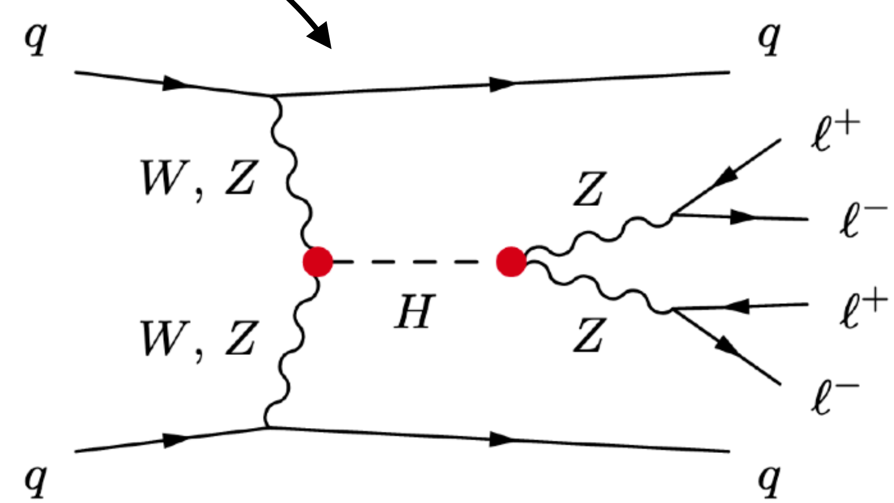
histogram ↔ $p(x|\theta)$

Multivariate Method: Matrix Elements



complex data: 10^8 sensors

match to parton level event



calculate matrix elements

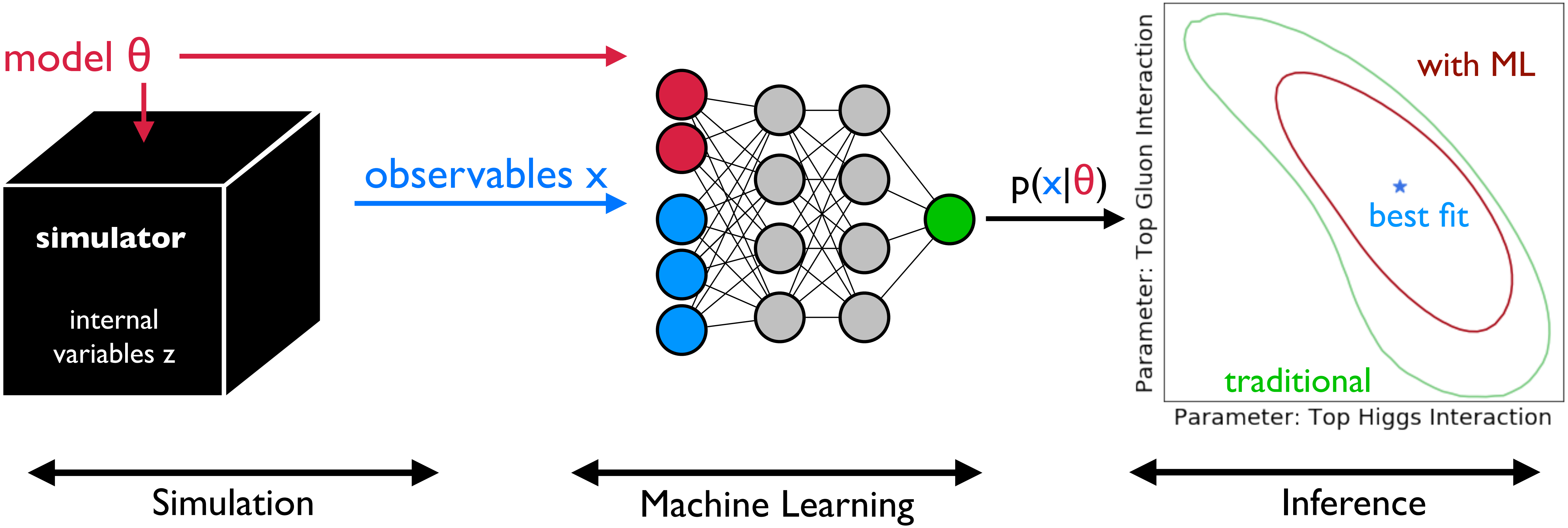
$p(x|\theta) \sim |M(x|\theta)|^2$

+ direct connection to underlying physics
+ works great at parton level
+ S' vs. S is easy

– requires approximations in reality
– can be slow/computationally expensive
– S vs. BG can be hard

Inference Methods

Multivariate Method: Machine Learning

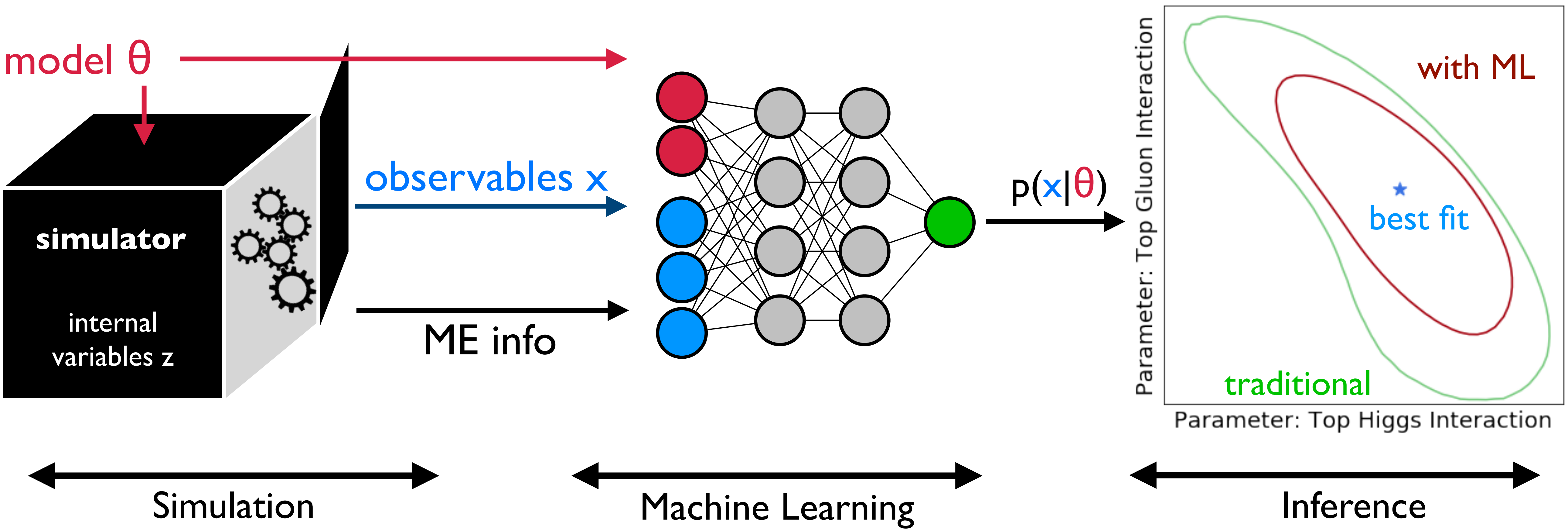


+ works great for S vs. BG
+ can be very powerful and quick

– struggles with S' vs S, especially if S' and S are very similar
– sometimes hard to understand what's going on

Inference Methods

Multivariate Method: The MadMiner Approach



power of machine learning

physics insight of matrix element information

MadMiner

[J. Brehmer, K. Cranmer, G. Louppe, J. Pavez 1805.00013, 1805.00020,1805.12244]

[J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

The Fisher Information

In SMEFT, we expand new physics around the SM using higher dimensional operators.
For practical reasons, we often stop at the lowest order: dimension 6.

$$L = L_{\text{SM}} + \sum (f_i/\Lambda^2) \times \mathcal{O}_i$$

our theory parameters θ

We can also locally expand the expected log-likelihood ratio around the SM (θref)

$$\mathbb{E}[-2 \log r_{\text{full}}(x|\theta)|\theta_{\text{ref}}] = I_{ij}(\theta_{\text{ref}}) \times (\theta - \theta_{\text{ref}})_i(\theta - \theta_{\text{ref}})_j + \cdots$$

The Fisher Information

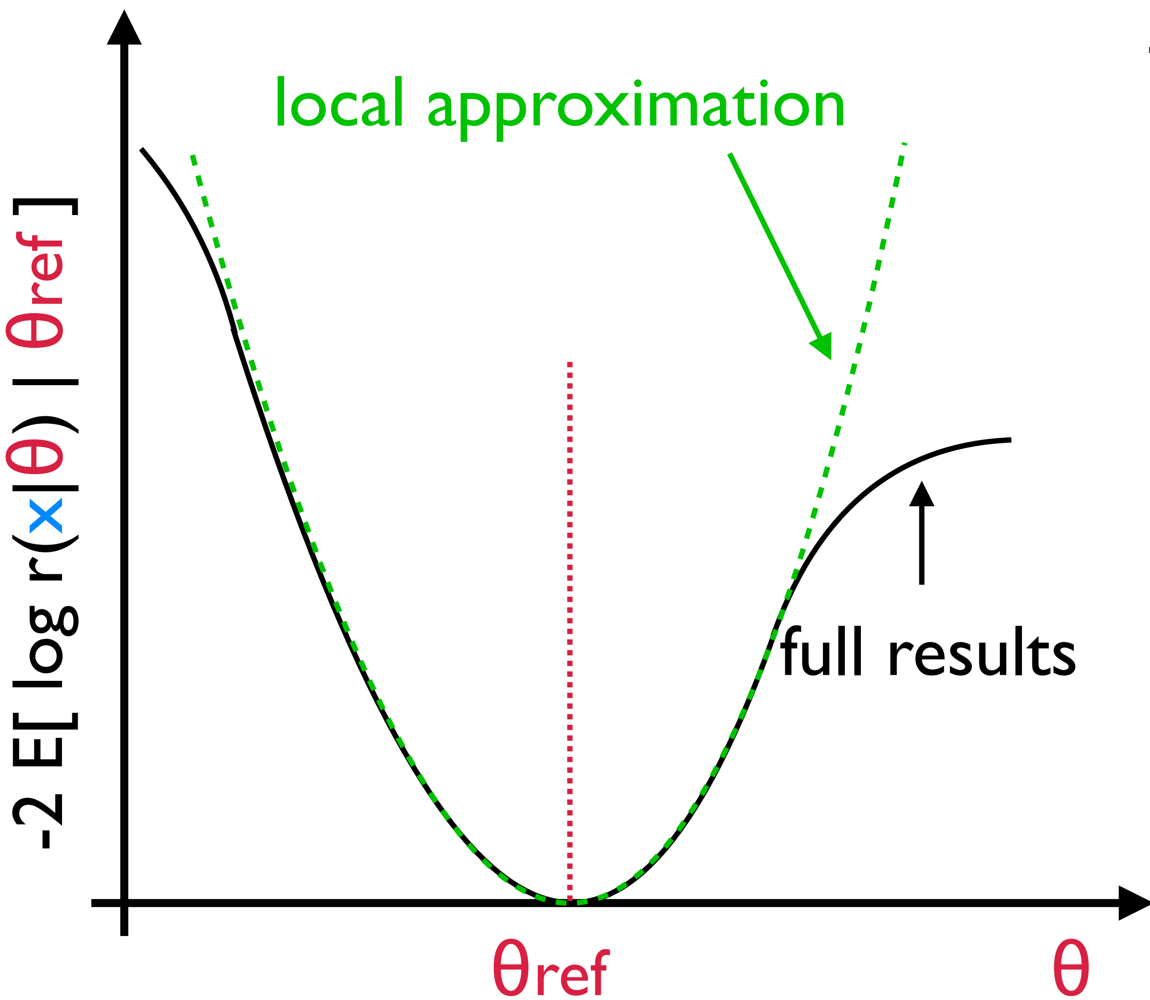
In SMEFT, we expand new physics around the SM using higher dimensional operators.
For practical reasons, we often stop at the lowest order: dimension 6.

$$L = L_{\rm SM} + \sum (f_i/\Lambda^2) \times \mathcal{O}_i$$

our theory parameters θ

We can also locally expand the expected log-likelihood ratio around the SM (θref)

$$\mathbb{E}[-2\log r_{\rm full}(x|\theta)|\theta_{\rm ref}] = I_{ij}(\theta_{\rm ref}) \times (\theta - \theta_{\rm ref})_i(\theta - \theta_{\rm ref})_j + \cdots$$



The expansion parameter is the Fisher Information,

$$I_{ij}(\theta_{\rm ref}) = \mathbb{E}\left[\frac{\partial \log p_{\rm full}(x|\theta)}{\partial\theta_i}\frac{\partial \log p_{\rm full}(x|\theta)}{\partial\theta_j}\middle|\theta_{\rm ref}\right]$$

It is easy to calculate and has many useful properties:

* simple: nxn matrix (for n theory parameters)
* independent of parameterizations of x
* covariant under θ→ θ'
* additive between experiments / phase-space

What can we do with these methods?

Estimate maximum sensitivity for Higgs measurements.

locally optimal observables (SALLY)

$O = d \log p(x|\theta)/d\theta \,|_{\theta_{ref}}$

likelihood ratio test (ALICES)

calculate $E[\log r(x|\theta) \mid \theta_{ref}]$

Example: Fully leptonic ttH production

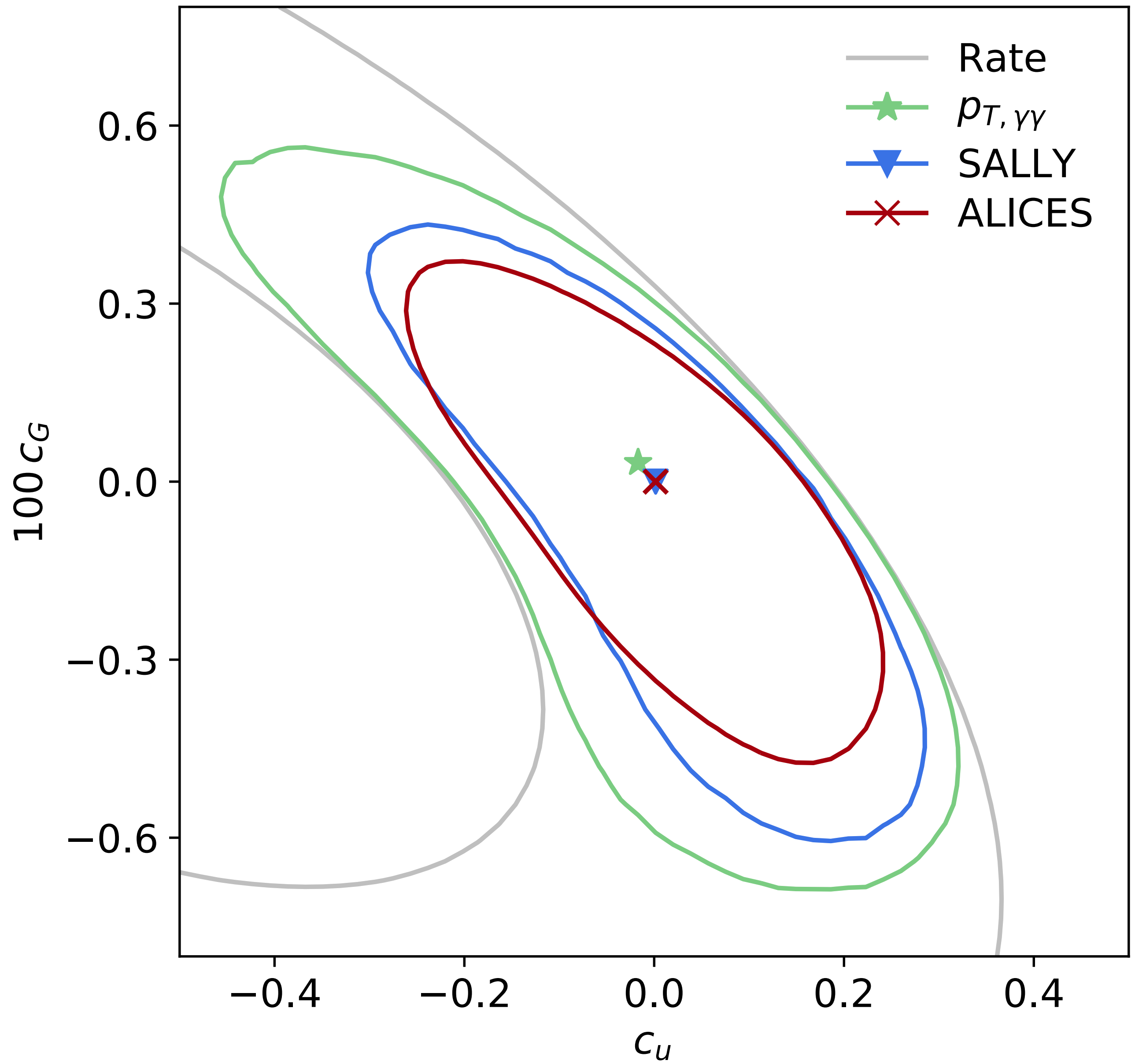
$pp \to ttH \to bb\, \ell\ell\, \gamma\gamma + MET$

to probe SMEFT

$O_u \sim (H^\dagger H)(H^\dagger Q_L)\, u_R$

$O_G \sim (H^\dagger H)\, G_{\mu\nu}\, G^{\mu\nu}$

Result: Multivariate methods can significantly enhance physics sensitivity (comparable to doubling the luminosity)



Example from [J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

What can we do with these methods?

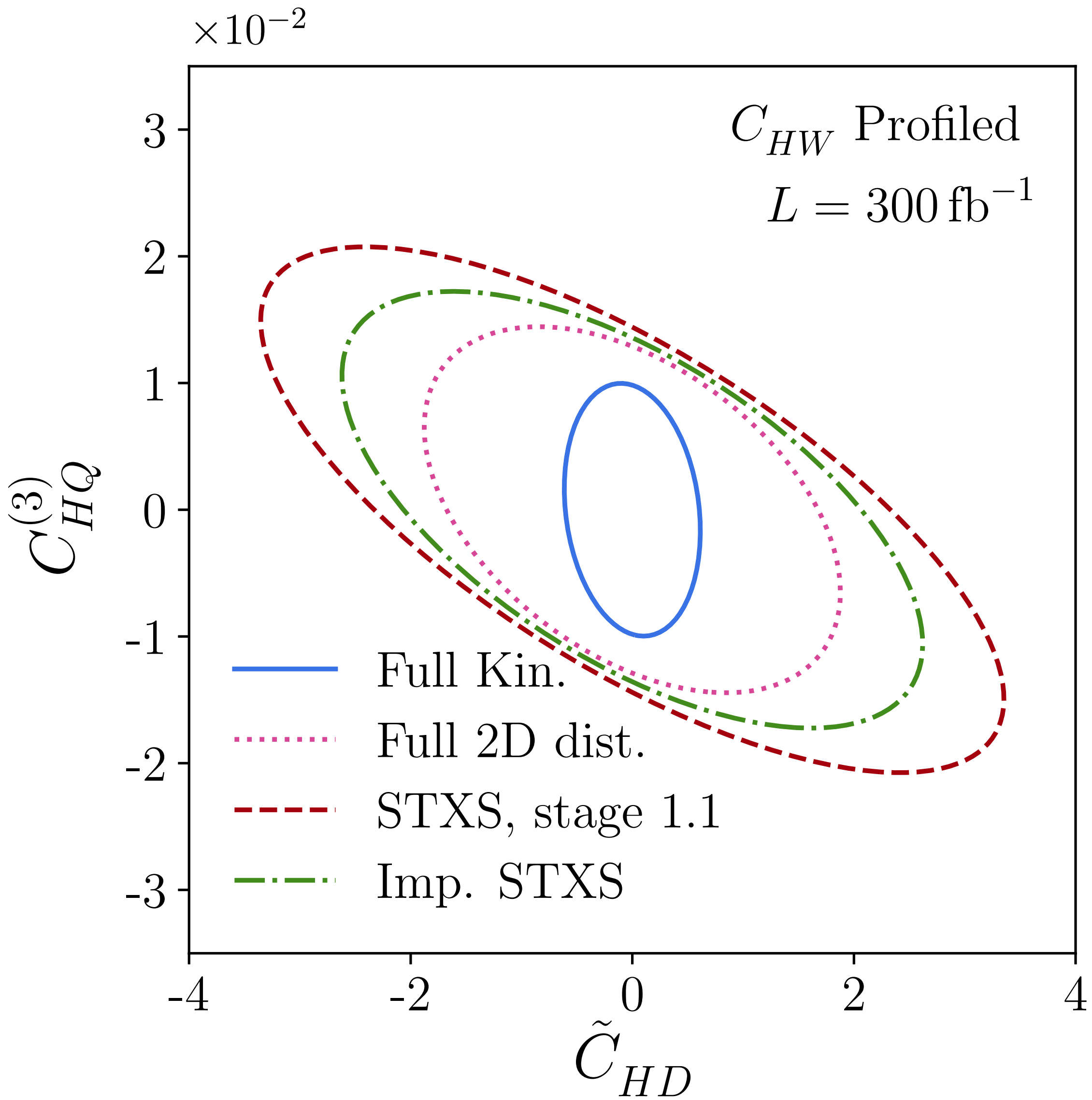
Quantify performance of difference analysis techniques.

The Fisher information encodes the maximum sensitivity of observables to theory parameters for a given experiment.

We can use it to quantify the information in different analyses, and compare their performance.

Example: Compare performance of STXS in WH production with H→bb.

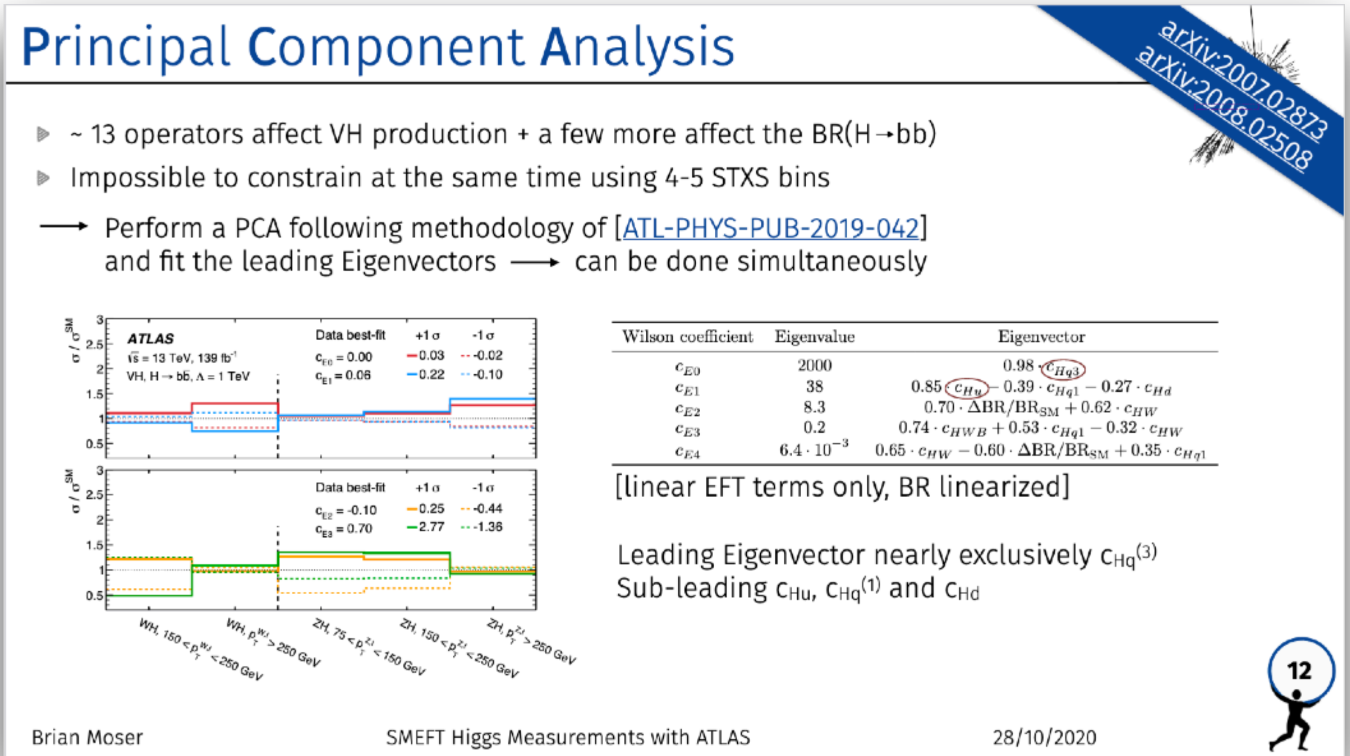
Result: Current STXS (which mostly consists of different pTH bins) lose a lot of information.



Example from [J. Brehmer, S. Dawson, S. Homiller, FK, T. Plehn, 1908.06980]

What can we do with these methods?

Study even high-dimensional parameter spaces fully differentially



Principal Component Analysis

arXiv:2007.02873
arXiv:2008.02508

- ~ 13 operators affect VH production + a few more affect the BR(H→bb)
- Impossible to constrain at the same time using 4-5 STXS bins

⟶ Perform a PCA following methodology of [ATL-PHYS-PUB-2019-042] and fit the leading Eigenvectors ⟶ can be done simultaneously

Wilson coefficient	Eigenvalue	Eigenvector
c_{E0}	2000	$0.98 \cdot c_{Hq3}$
c_{E1}	38	$0.85 \cdot c_{Hu} - 0.39 \cdot c_{Hq1} - 0.27 \cdot c_{Hd}$
c_{E2}	8.3	$0.70 \cdot \Delta\mathrm{BR}/\mathrm{BR}_{\mathrm{SM}} + 0.62 \cdot c_{HW}$
c_{E3}	0.2	$0.74 \cdot c_{HWB} + 0.53 \cdot c_{Hq1} - 0.32 \cdot c_{HW}$
c_{E4}	$6.4 \cdot 10^{-3}$	$0.65 \cdot c_{HW} - 0.60 \cdot \Delta\mathrm{BR}/\mathrm{BR}_{\mathrm{SM}} + 0.35 \cdot c_{Hq1}$

[linear EFT terms only, BR linearized]

Leading Eigenvector nearly exclusively $c_{Hq}^{(3)}$
Sub-leading c_{Hu}, $c_{Hq}^{(1)}$ and c_{Hd}

Brian Moser   SMEFT Higgs Measurements with ATLAS   28/10/2020   12

slide from Brian Moser's talk [link]

Disentangling the effects of many different theory parameters is challenging.

Fisher Information is a convenient and natural description when parameter space is large.

Can also be done fully differentially.

Example: Information on Wilson coefficients in WBF Higgs production (including CP-odd ones) from fully differential distributions.

$$I_{ij} = \begin{array}{c} \begin{matrix} f_{\phi,2} & f_W & f_B & f_{WW} & f_{BB} & f_{W\widetilde{W}} & f_{B\tilde{B}} & \mathrm{Im}\, f_W & \mathrm{Im}\, f_B & \mathrm{Im}\, f_{WW} & \mathrm{Im}\, f_{BB} & \mathrm{Im}\, f_{W\widetilde{W}} & \mathrm{Im}\, f_{B\tilde{B}} \end{matrix} \\ \begin{pmatrix} 4942 & -968 & -50 & 54 & 2 & -7 & 0 & -1 & 0 & 2 & 0 & 36 & 0 \\ -968 & 715 & 35 & -191 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & -55 & -1 \\ -50 & 35 & 6 & -9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 54 & -191 & -9 & 321 & 3 & -1 & 0 & 0 & 0 & 1 & 0 & 72 & 1 \\ 2 & -3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -7 & 1 & 0 & -1 & 0 & 359 & 4 & 41 & 1 & -81 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 41 & 0 & 6 & 0 & -12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & -81 & -1 & -12 & 0 & 23 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 36 & -55 & -2 & 72 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 21 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

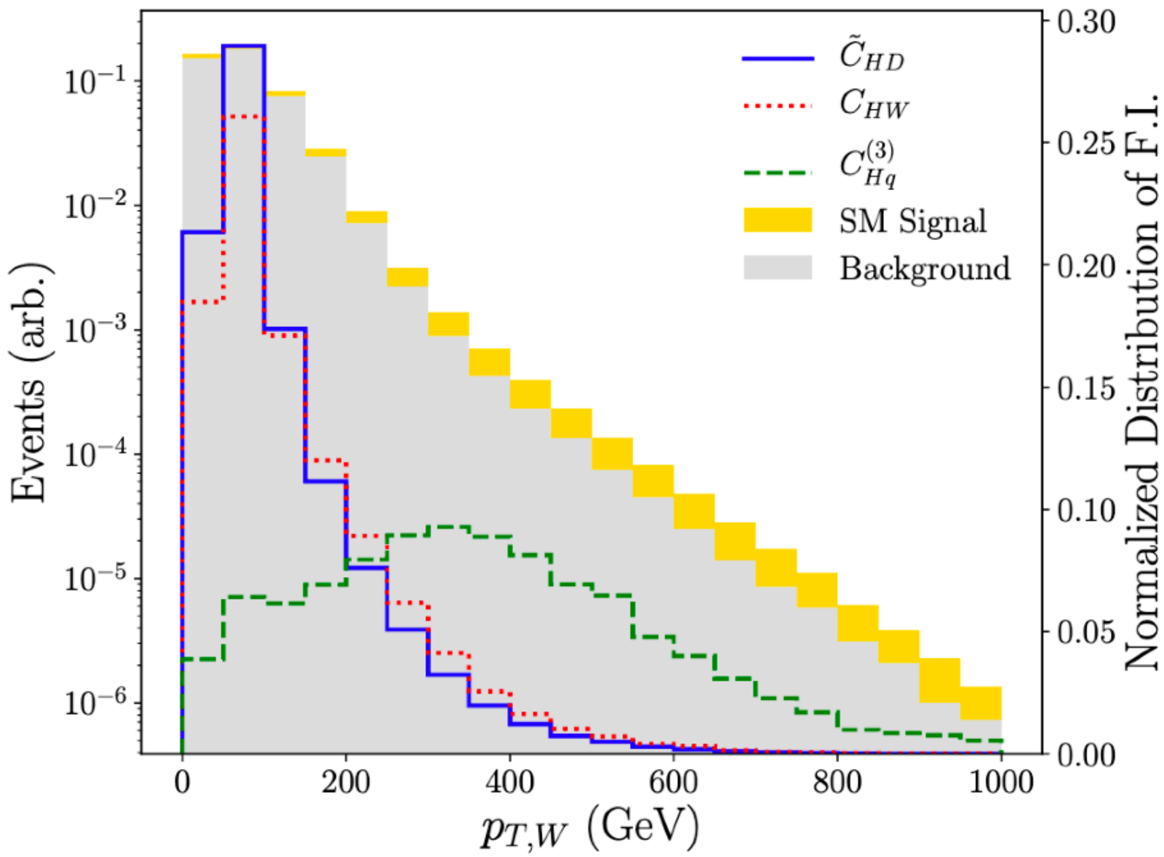
Example from [J. Brehmer, FK, T. Plehn, T. Tait, 1712.02350]

What can we do with these methods?

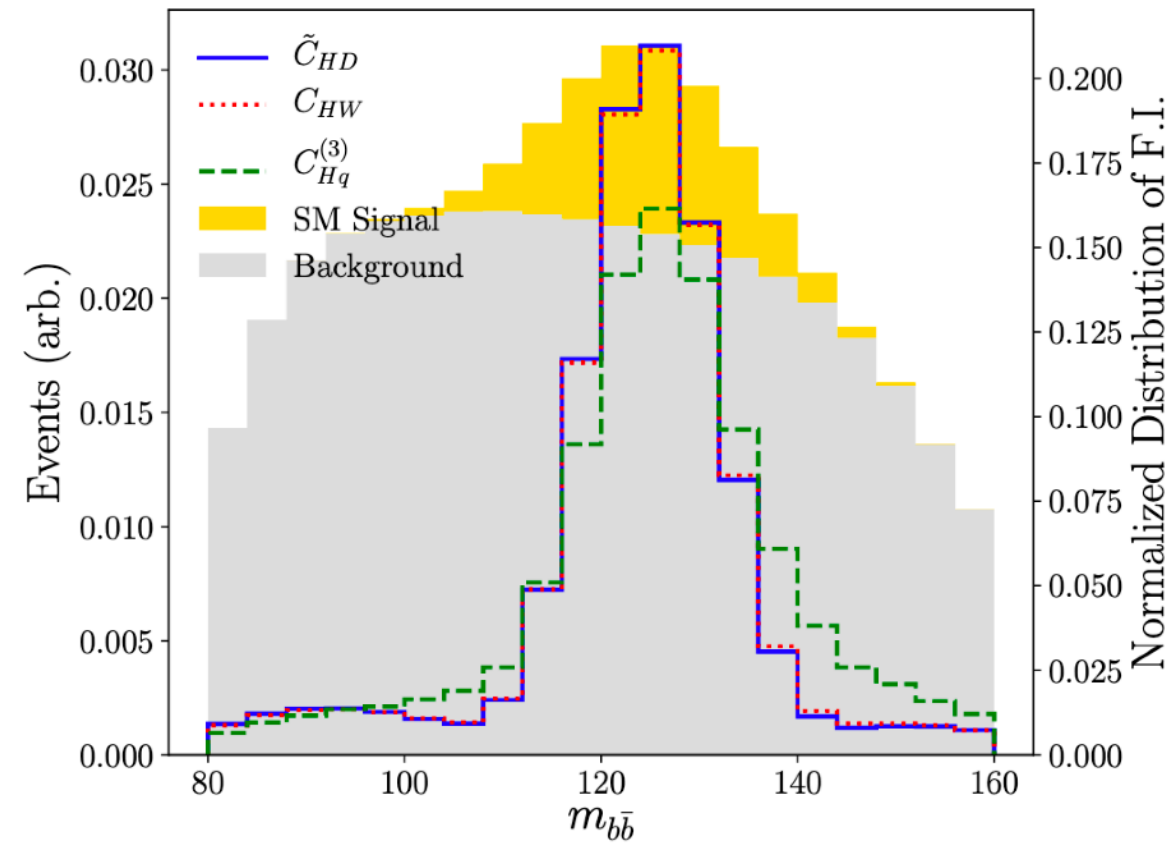
Identify relevant phase-space regions

When designing an analysis, it is helpful to understand which phase space regions carry most in formation. We can do that by looking at the distribution of information.

Example: Distributions in WH production with H→bb



Momentum enhanced operators have Fisher Information peaked in high momentum bins



Including backgrounds can change naive distribution of information

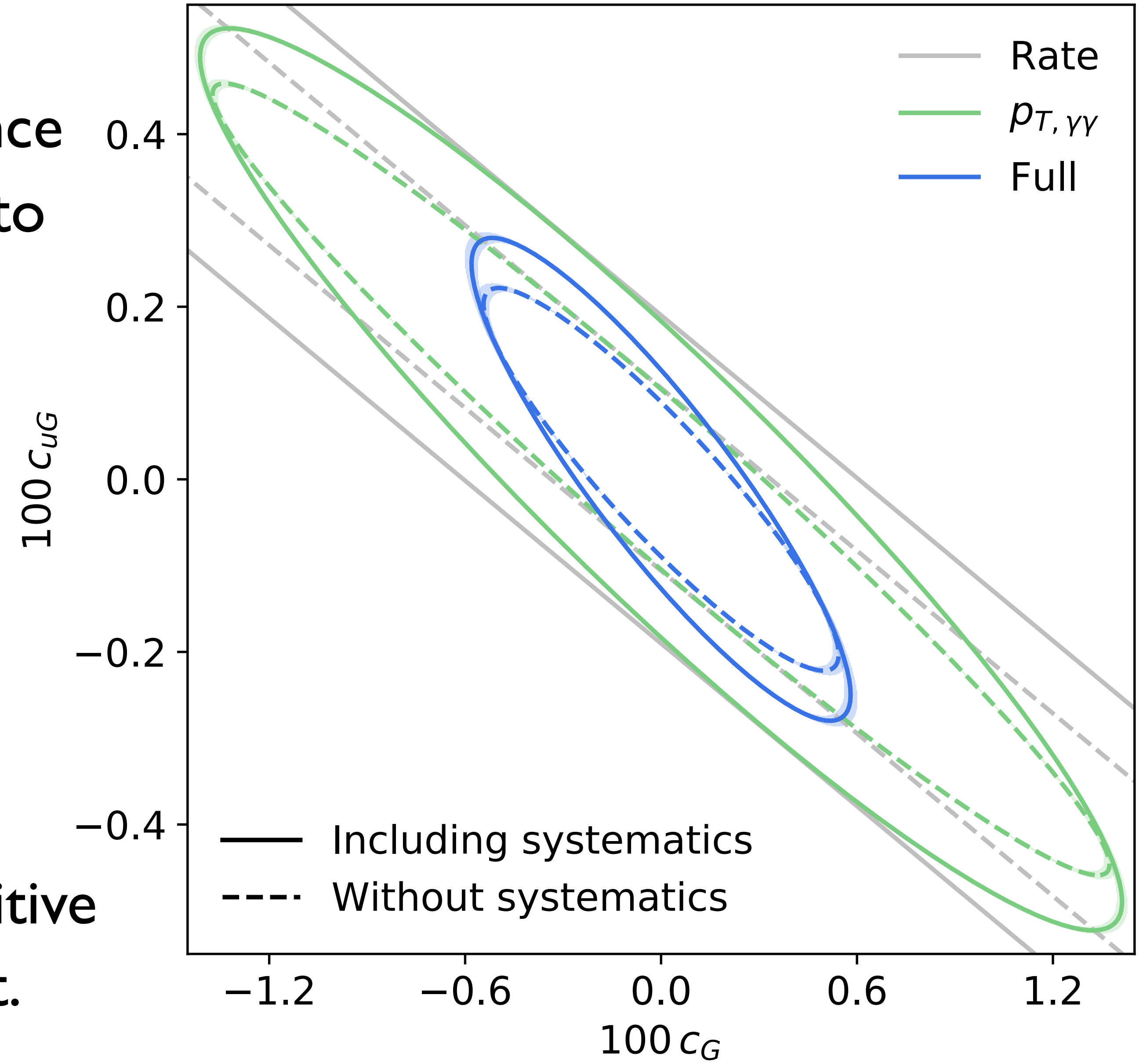
Example from [J. Brehmer, S. Dawson, S. Homiller, FK, T. Plehn, 1908.06980]

What can we do with these methods?

Study the effect of systematics on our physics sensitivity.

We are often looking for subtle kinematic effect. It's therefore important to understand if effects of new physics can be faked by systematics.

Systematics can be parameterized using nuisance parameter ν. We can use the same methods to learn r(x|θ,ν) and then profile over ν.



Example: Fully leptonic ttH production with scale and PDF uncertainties.

Result: Systematic reduce reach in rate-sensitive direction; multivariate analysis is more robust.

Example from [J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

Conclusion / Start of Discussion

Most of the information on Higgs couplings in contained it kinematic distributions. We need multivariate analysis methods to extract this information.

There are many new tools and methods available. Let's use them to better understand and optimize Higgs measurements.

The LHC is approaching an era of precision Higgs measurements. We need to talk more about systematics when studying sensitivities.

MadMiner: The Tool

MadMiner [J. Brehmer, FK, I. Espejo, K. Cranmer 1907.10621]

- automizes these techniques
- straightforward to apply them to LHC problems

- out of the box: Pheno-level analysis
  * MadGraph, Pythia, Delphes
  * backgrounds
  * PDF/scale uncertainties
  * ML uncertainties
  * morphing
  * many inference techniques (SALLY, ALICES …)

- scalable to state-of-the-art experimental tools

- python package
  * modular interface
  * extensive documentation
  * on GitHub

    github.com/diana-hep/madminer
  * easy to install

    `pip install madminer`