Better LHC measurements through information geometry

Johann Brehmer

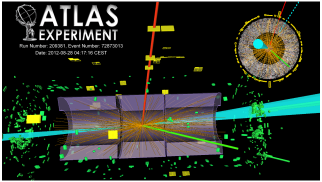
Heidelberg University

Based on 1612.05261
with Kyle Cranmer, Felix Kling, and Tilman Plehn

Pheno 2017

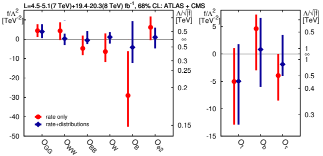
Inference at the LHC

Complex data $\mathbf{x}$



[ATLAS 1501.04943]

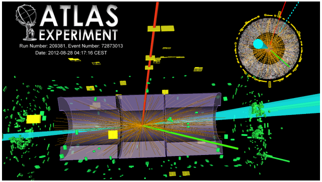
Many parameters $\boldsymbol{\theta}$



[T. Corbett et al 1505.05516]

Inference at the LHC

Complex data $\mathbf{x}$

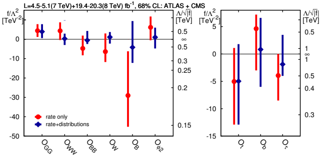


[ATLAS 1501.04943]

- Conventional analyses:
  - standard kinematic observables

  $\Rightarrow$ reproducible and transparent;
  don't scale well with complexity

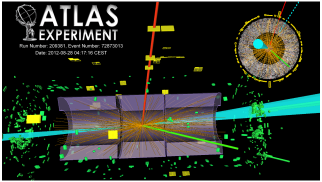
Many parameters $\boldsymbol{\theta}$



[T. Corbett et al 1505.05516]

Inference at the LHC

Complex data $\mathbf{x}$

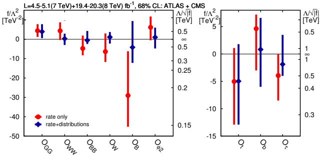


[ATLAS 1501.04943]

- Conventional analyses:
  - standard kinematic observables
  - $\Rightarrow$ reproducible and transparent; don't scale well with complexity

- Multivariate methods:
  - matrix-element-based
  - likelihood-free inference (machine learning)
  - $\Rightarrow$ powerful black boxes

Many parameters $\boldsymbol{\theta}$



[T. Corbett et al 1505.05516]

Efficient measurements need guidelines

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

1. What is the maximum sensitivity of a measurement?
2. Where in phase space is the information?
3. How powerful are different observables?

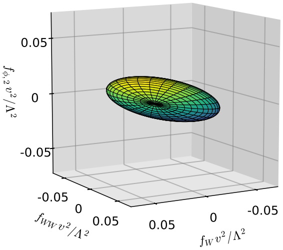
Efficient measurements need guidelines

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

1. What is the maximum sensitivity of a measurement?
2. Where in phase space is the information?
3. How powerful are different observables?

Now:	a statistics tool box based on information geometry
Next talk by Felix Kling:	application to Higgs measurements and SM EFT

1. What is the maximum sensitivity of a measurement?



Cramér-Rao bound

- Measurement process:

$$\underset{\text{true parameters}}{\boldsymbol{\theta}_0} \xrightarrow{f(\mathbf{x}|\boldsymbol{\theta}_0)} \underset{\text{observables}}{\mathbf{x}} \xrightarrow{\hat{\boldsymbol{\theta}}(\mathbf{x})} \underset{\text{estimator}}{\hat{\boldsymbol{\theta}}}$$

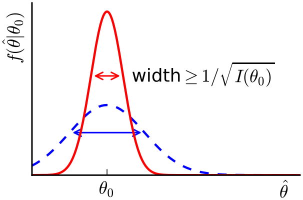
Cramér-Rao bound

- Measurement process:

$$\underset{\text{true parameters}}{\boldsymbol{\theta}_0} \xrightarrow{f(\mathbf{x}|\boldsymbol{\theta}_0)} \underset{\text{observables}}{\mathbf{x}} \xrightarrow{\hat{\boldsymbol{\theta}}(\mathbf{x})} \underset{\text{estimator}}{\hat{\boldsymbol{\theta}}}$$

- Cramér-Rao bound for unbiased estimators:
  [C. R. Rao 1945; H. Cramér 1946]

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\middle|\boldsymbol{\theta}_0\right]_{ij} \geq I^{-1}_{ij}(\boldsymbol{\theta}_0)$$



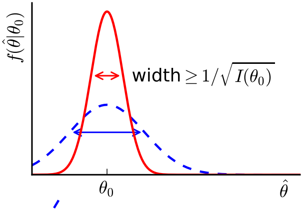
Cramér-Rao bound

- Measurement process:

$$\underset{\text{true parameters}}{\boldsymbol{\theta}_0} \xrightarrow{f(\mathbf{x}|\boldsymbol{\theta}_0)} \underset{\text{observables}}{\mathbf{x}} \xrightarrow{\hat{\boldsymbol{\theta}}(\mathbf{x})} \underset{\text{estimator}}{\hat{\boldsymbol{\theta}}}$$

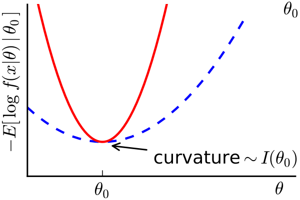
- Cramér-Rao bound for unbiased estimators:
  [C. R. Rao 1945; H. Cramér 1946]

$$\operatorname{cov}\left[\hat{\boldsymbol{\theta}}\middle|\boldsymbol{\theta}_0\right]_{ij} \geq I^{-1}_{ij}(\boldsymbol{\theta}_0)$$



  with Fisher information
  [F. Edgeworth 1908; R. Fisher 1925; ...]

$$I_{ij}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 \log f(\mathbf{x}|\boldsymbol{\theta})}{\partial\theta_i\,\partial\theta_j}\middle|\boldsymbol{\theta}\right]$$



$\Rightarrow I_{ij} \sim$ maximal precision with which $\boldsymbol{\theta}$ can be measured in an experiment

The Fisher information and the LHC

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

- Properties:
  - Describes all directions in theory space
  - Additive between experiments / phase-space regions
  - Independent of parametrization of $\mathbf{x}$
  - Covariant under $\boldsymbol{\theta} \to \boldsymbol{\theta}'$

The Fisher information and the LHC

- Properties:
  - Describes all directions in theory space
  - Additive between experiments / phase-space regions
  - Independent of parametrization of $\mathbf{x}$
  - Covariant under $\boldsymbol{\theta} \to \boldsymbol{\theta}'$

- Fisher information in LHC processes:
  - Extended likelihood ansatz:

$$f(\mathbf{x}|\boldsymbol{\theta}) = \underbrace{\mathrm{Pois}(n|\sigma L)}_{\text{Total event number}} \prod_{i=1}^{n} \underbrace{f^{(1)}(\mathbf{x}_i|\boldsymbol{\theta})}_{\text{Kinematics of each event}}$$

  - MC integration gives

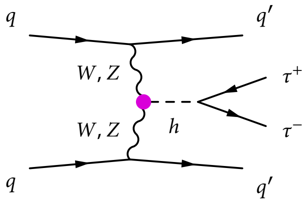
$$I_{ij}(\boldsymbol{\theta}) = L \sum_{\text{events } k} \frac{\partial \Delta\sigma_k}{\partial \theta_i} \frac{1}{\Delta\sigma_k} \frac{\partial \Delta\sigma_k}{\partial \theta_j}$$

$\Rightarrow$ Can calculate all $I_{ij}(\boldsymbol{\theta})$ from a single MC run

Cramér-Rao in practice

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

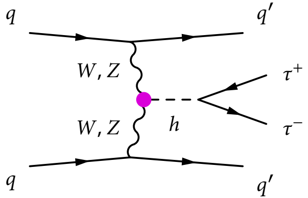
- Toy example:
  - Weak boson fusion, $h \to \tau\tau$
  - NP parameters $f_{\phi,2}$, f_W, f_{WW} characterizing hWW coupling
  - All the cool physics (and all the dirty details) in the next talk!



Cramér-Rao in practice

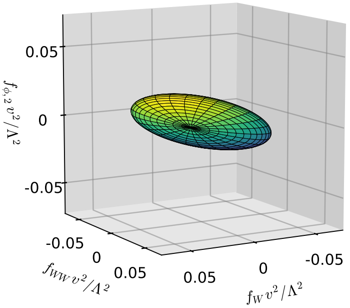
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

- Toy example:
  - Weak boson fusion, $h \to \tau\tau$
  - NP parameters $f_{\phi,2}$, f_W, f_{WW} characterizing hWW coupling
  - All the cool physics (and all the dirty details) in the next talk!



- Fisher information $\leftrightarrow$ minimal error ellipsoids:

$$I_{ij}(\mathbf{0}) = \begin{array}{c} \\ \end{array}\begin{array}{cccc} f_{\phi,2} & f_W & f_{WW} & \\ \left(\begin{matrix} 3202 & -625 & -7 \\ -625 & 451 & -110 \\ -7 & -110 & 244 \end{matrix}\right) & & & \begin{matrix} f_{\phi,2} \\ f_W \\ f_{WW} \end{matrix} \end{array} \longleftrightarrow$$



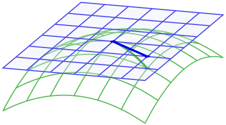
Information geometry

- Geometric interpretation: [C. R. Rao 1945, S. Amari 1968; …]
  - Parameter space of theory $\rightsquigarrow$ manifold
  - Parametrization θ_i $\rightsquigarrow$ map (coordinates)
  - Fisher information I_{ij} $\rightsquigarrow$ Riemannian metric

- Distance measures:
  - Local / tangent space at $\boldsymbol{\theta}_0$:

$$d_{\text{local}}(\boldsymbol{\theta};\boldsymbol{\theta}_0) = \sqrt{I_{ij}(\boldsymbol{\theta}_0)\;(\theta^i - \theta^i_0)\;(\theta^j - \theta^j_0)}$$

  $\sim$ unlikeliness to measure $\boldsymbol{\theta}$ if $\boldsymbol{\theta}_0$ is true, 'in sigmas'

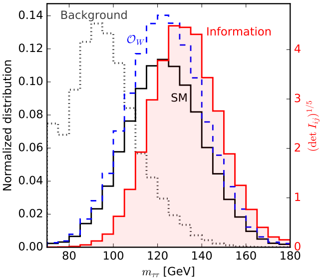


  - Global along geodesics:

$$d(\boldsymbol{\theta}_a, \boldsymbol{\theta}_b) = \min_{\boldsymbol{\theta}(s)} \int_{s_a}^{s_b} \mathrm{d}s \sqrt{I_{ij}(\boldsymbol{\theta})\;\frac{\mathrm{d}\theta_i(s)}{\mathrm{d}s}\;\frac{\mathrm{d}\theta_j(s)}{\mathrm{d}s}}$$

  - Difference between $d_{\text{local}}(\boldsymbol{\theta},\mathbf{0})$ and $d(\boldsymbol{\theta},\mathbf{0})$ $\leftrightarrow$ impact of $\mathcal{O}\left(\boldsymbol{\theta}^2\right)$ contributions

2. Where in phase space is the information?



The differential information

- Differential information with respect to any observable:

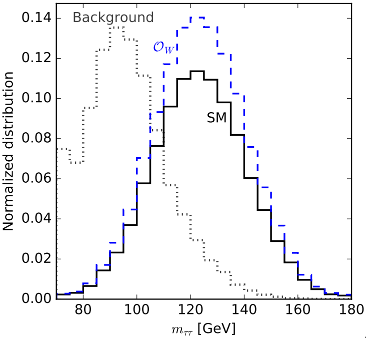
$$I_{ij}(\boldsymbol{\theta}) = \sum_{\text{events}} L \, \frac{\partial \Delta\sigma}{\partial \theta_i} \, \frac{1}{\Delta\sigma} \, \frac{\partial \Delta\sigma}{\partial \theta_j} = \sum_{\text{bins } b} \underbrace{\sum_{\text{events in } b} L \, \frac{\partial \Delta\sigma}{\partial \theta_i} \, \frac{1}{\Delta\sigma} \, \frac{\partial \Delta\sigma}{\partial \theta_j}}_{\text{information in } b}$$

The differential information

- Differential information with respect to any observable:

$$I_{ij}(\boldsymbol{\theta}) = \sum_{\text{events}} L \frac{\partial \Delta\sigma}{\partial \theta_i} \frac{1}{\Delta\sigma} \frac{\partial \Delta\sigma}{\partial \theta_j} = \sum_{\text{bins } b} \underbrace{\sum_{\text{events in } b} L \frac{\partial \Delta\sigma}{\partial \theta_i} \frac{1}{\Delta\sigma} \frac{\partial \Delta\sigma}{\partial \theta_j}}_{\text{information in } b}$$

- Right: $m_{\tau\tau}$ distribution
  - SM Higgs vs NP Higgs vs Z background rates



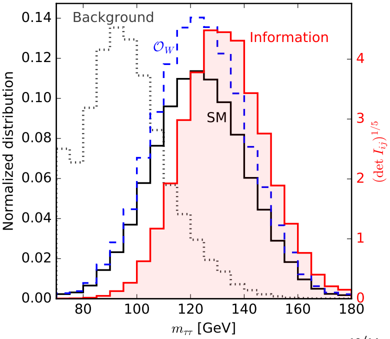
The differential information

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

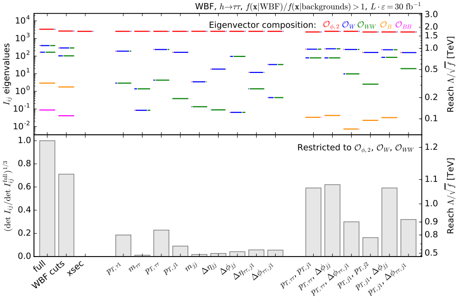
- Differential information with respect to any observable:

$$I_{ij}(\boldsymbol{\theta}) = \sum_{\text{events}} L \frac{\partial \Delta\sigma}{\partial \theta_i} \frac{1}{\Delta\sigma} \frac{\partial \Delta\sigma}{\partial \theta_j} = \sum_{\text{bins } b} \underbrace{\sum_{\text{events in } b} L \frac{\partial \Delta\sigma}{\partial \theta_i} \frac{1}{\Delta\sigma} \frac{\partial \Delta\sigma}{\partial \theta_j}}_{\text{information in } b}$$

- Right: $m_{\tau\tau}$ distribution
  - SM Higgs vs NP Higgs vs Z background rates
  - Distribution of differential information



3. How powerful are different observables?



Information in individual distributions

- Reduced information in histogram (rather than full kinematics):

$$I_{ij}^{\text{distribution}}(\boldsymbol{\theta}) = \sum_{\text{bins } b} L \, \frac{\partial \sigma_b(\boldsymbol{\theta})}{\partial \theta_i} \, \frac{1}{\sigma_b(\boldsymbol{\theta})} \, \frac{\partial \sigma_b(\boldsymbol{\theta})}{\partial \theta_j}$$

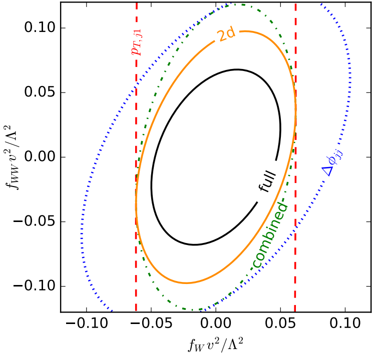
Information in individual distributions

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

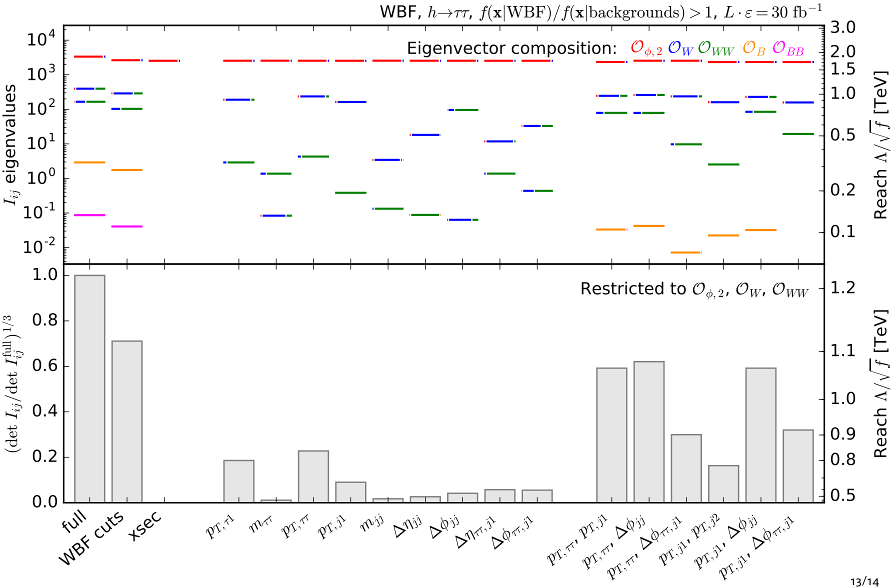
- Reduced information in histogram (rather than full kinematics):

$$I_{ij}^{\text{distribution}}(\boldsymbol{\theta}) = \sum_{\text{bins } b} L \, \frac{\partial \sigma_b(\boldsymbol{\theta})}{\partial \theta_i} \, \frac{1}{\sigma_b(\boldsymbol{\theta})} \, \frac{\partial \sigma_b(\boldsymbol{\theta})}{\partial \theta_j}$$

- Right: constraining power of WBF distributions
  - different observables and their combination
  - full kinematics

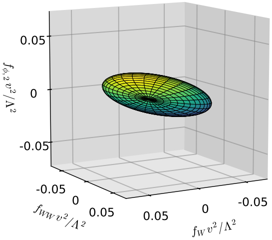


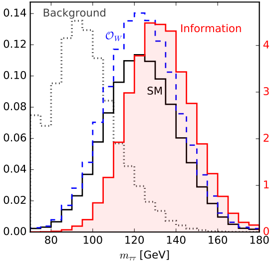
Comparison of observables

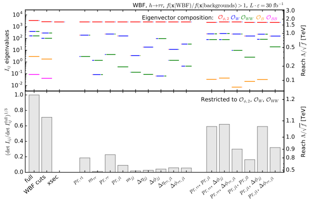


Conclusions

Information geometry lets us…







- calculate the maximum sensitivity of any LHC process
- find the important phase-space regions
- select the most powerful observables, compare them to multivariate methods

Stay tuned for physics applications!

Bonus material

Belated introduction slide

- There's probobably[1] new physics in the Higgs sector
  - Hierarchy problem
  - Fermion masses
  - DM
  - Baryon asymmetry
  - …

with a bit of creativity

- Measurement of Higgs properties most exciting mission for Run 2

until the LHC finds something really cool

- Need model-independent parametrisation of Higgs properties

[1] No warranty, expressed or implied

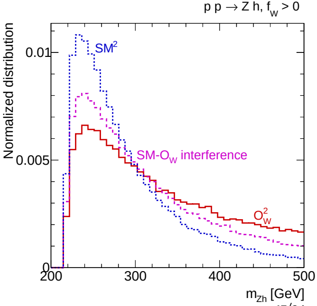
SM effective field theory

[W. Buchmuller, D. Wyler 85; K. Hagiwara, S. Ishihara, S. R. Szalapski, D. Zeppenfeld 93; B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

- New physics at $\Lambda \gg E_{\text{LHC}} \sim m_h$?

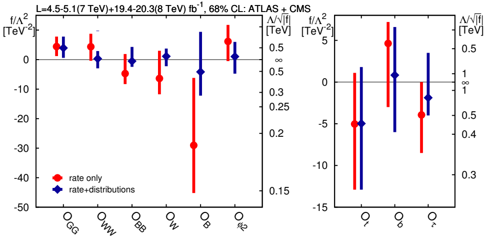
$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \underbrace{\sum_i^{59} \frac{f_i^{d=6}}{\Lambda^2} \mathcal{O}_i^{d=6}}_{\text{e. g. } \mathcal{O}_W = (D^\mu \phi)^\dagger \sigma^k (D^\nu \phi) W^k_{\mu\nu} \dots} + \sum_k \frac{f_k^{d=8}}{\Lambda^4} \mathcal{O}_k^{d=8} + \dots$$

- Dimension-6 operators: perfect language for new physics signatures in Higgs sector?
  - Model independence?
  - Correlations between Higgs, LHC TGC, LEP, …
  - Total rates + distributions



Sensitivity vs validity

- Run I fit: [T. Corbett, O. Eboli, D. Goncalves, J. Gonzalez-Fraile, T. Plehn, M. Rauch 1505.05516]

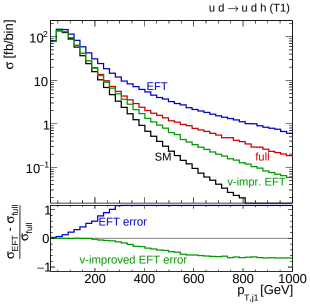


- Kinematic information up to $E \lesssim 400$ GeV crucial
- Sensitive to NP scales $\Lambda \sim \sqrt{f} \cdot 400$ GeV

- Is the dimension-six model still useful?
  - Strongly coupled NP: works fine
  - Weakly coupled NP: no guarantee, but works in many scenarios (with v-improved matching)

    [JB, A. Freitas, D. Lopez-Val, T. Plehn 1510.03443]
  - EFT less reliable in high-energy tails

[See YR4 and references therein ;), note similar questions for DM EFTs]

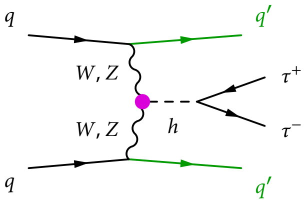


Weak boson fusion (WBF), $h \to \tau\tau$

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

- Kinematics of tagging jets sensitive to Higgs-gauge interaction

  [D. Rainwater, D. Zeppenfeld, K. Hagiwara hep-ph/9808468; T. Plehn, D. Rainwater, D. Zeppenfeld hep-ph/0105325; C. Englert, D. Gonçalves-Netto, K. Mawatari, T. Plehn 1212.0843; …]



- Model: dimension-6 Higgs-gauge operators

$$\boldsymbol{\theta} = \frac{v^2}{\Lambda^2} \begin{pmatrix} f_{\phi,2} \\ f_W \\ f_{WW} \\ f_B \\ f_{BB} \end{pmatrix}$$

$f_{\phi,2} \longrightarrow \mathcal{O}_{\phi,2} = \frac{1}{2}\, \partial^\mu (\phi^\dagger \phi)\, \partial_\mu (\phi^\dagger \phi)$ rescales all h couplings

$f_W \longrightarrow \mathcal{O}_W = \mathrm{i}\frac{g}{2}\, (D^\mu \phi)^\dagger \sigma^k (D^\nu \phi)\, W^k_{\mu\nu}$

$f_{WW} \longrightarrow \mathcal{O}_{WW} = -\frac{g^2}{4}\, (\phi^\dagger \phi)\, W^k_{\mu\nu}\, W^{\mu\nu\, k}$

hWW, hZZ kinematics

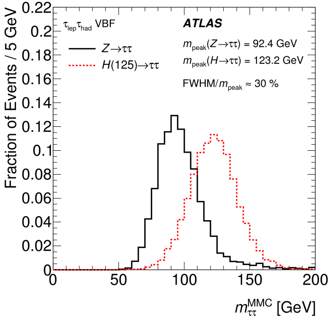
$f_B \longrightarrow \mathcal{O}_B = \mathrm{i}\frac{g}{2}\, (D^\mu \phi^\dagger)(D^\nu \phi)\, B_{\mu\nu}$

$f_{BB} \longrightarrow \mathcal{O}_{BB} = -\frac{g'^2}{4}\, (\phi^\dagger \phi)\, B_{\mu\nu}\, B^{\mu\nu}$

hZZ kinematics

Setup

- Tools: MadGraph 5, MadMax
  [J. Alwall et al 1405.0301;
  K. Cranmer, T. Plehn hep-ph/0605268;
  T. Plehn, P. Schichtel, D. Wiegand 1311.2591;
  F. Kling, T. Plehn, P. Schichtel 1607.07441]
- Backgrounds:
  - QCD and electroweak $Z \to \tau\tau$
  - Gluon-fusion Higgs production



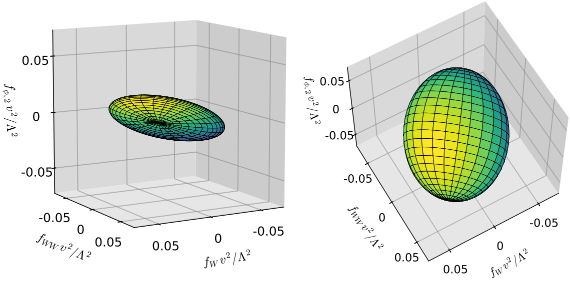
- Approximations:
  - τ decays not simulated — BR for semileptonic $\tau\tau$ mode
  - Parton level — CJV survival probabilities from literature [D. Rainwater, D. Zeppenfeld, K. Hagiwara hep-ph/9808468]
  - No detector simulation — $m_{\tau\tau}$ smeared by single / double Gaussian fitted to ATLAS results [ATLAS 1501.04943, see above]
  - No systematic or theory uncertainties
- $\sqrt{s} = 13$ TeV, $L \cdot \varepsilon = 30$ fb^{-1}
- Cuts: $p_{T,j} > 20$ GeV, $|\eta_j| < 5.0$, $\Delta\eta_{jj} > 2.0$, $\Delta R_{jj} > 0.4$

Maximum sensitivity in WBF, $h \to \tau\tau$

- Fisher information at the SM:

$$
\begin{array}{c@{}c@{}c}
 & \begin{array}{ccccc} \mathcal{O}_{\phi,2} & \mathcal{O}_W & \mathcal{O}_{WW} & \mathcal{O}_B & \mathcal{O}_{BB} \end{array} & \\
I_{ij}(\mathbf{0}) = & \begin{pmatrix} 3202 & -625 & -7 & -35 & 0 \\ -625 & 451 & -110 & 23 & -2 \\ -7 & -110 & 244 & -6 & 3 \\ -35 & 23 & -6 & 4 & 0 \\ 0 & -2 & 3 & 0 & 0 \end{pmatrix} & \begin{array}{c} \mathcal{O}_{\phi,2} \\ \mathcal{O}_W \\ \mathcal{O}_{WW} \\ \mathcal{O}_B \\ \mathcal{O}_{BB} \end{array}
\end{array}
$$

- Minimal errors $\Delta\theta \geq 1/\sqrt{I}$:
  - Largest eigenvalue along $\mathcal{O}_{\phi,2}$:
    $$\Delta(f/\Lambda^2) \gtrsim 0.3\ \text{TeV}^{-2}$$
  - $\mathcal{O}_W$-$\mathcal{O}_{WW}$ plane:
    $$\Delta(f/\Lambda^2) \gtrsim 1.0\ \text{TeV}^{-2}$$
  - Large mixing



A hierarchy of scales?

- EFT approach based on $E^2/\Lambda^2 \ll 1$
- Test this scale hierarchy!
  - Limit momentum flow with cut

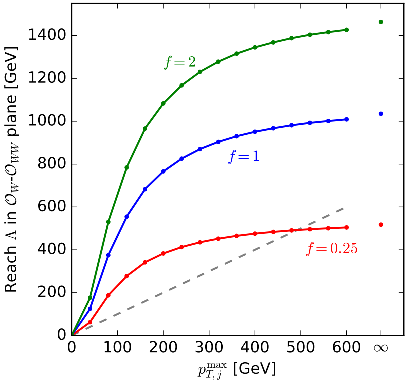
$$E \sim p_{T,j} < p_{T,j}^{\text{max}}$$

  - Precision

$$\Delta(f\, v^2/\Lambda^2) = 1/\sqrt{I}$$

corresponds to new physics reach

$$\Lambda = \sqrt{f}\; v\; I^{1/4}$$



Geometry of effective field theories

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

- Remember

$$I_{ij}(\boldsymbol{\theta}) = L \sum_{\text{events}} \frac{\partial \Delta\sigma(\boldsymbol{\theta})}{\partial \theta_i} \frac{1}{\Delta\sigma(\boldsymbol{\theta})} \frac{\partial \Delta\sigma(\boldsymbol{\theta})}{\partial \theta_j}$$

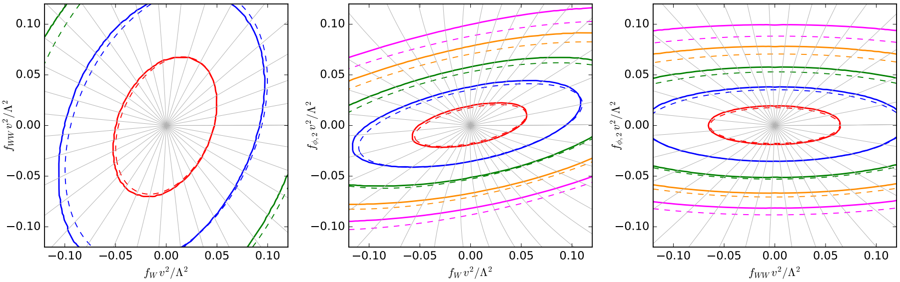
$\Rightarrow$ $I_{ij}(\mathbf{0})$ only sensitive to linear effects $\Delta\sigma \sim \theta_i \Delta\sigma_i$

- Information geometry for dimension-6 operators, $\theta_i = f_i^{d=6} v^2/\Lambda^2$:

$$\Delta\sigma = \underbrace{\underbrace{\Delta\sigma_{SM} + \sum_i \frac{f_i^{d=6}}{\Lambda^2} \Delta\sigma_i}_{I_{ij}(\mathbf{0}),\ \text{local distances at SM}} + \sum_{i,j} \frac{f_i^{d=6} f_j^{d=6}}{\Lambda^4} \Delta\sigma_{ij}}_{I_{ij}(\boldsymbol{\theta} \neq 0),\ \text{global distances}} + \underbrace{\sum_k \frac{f_k^{d=8}}{\Lambda^4} \Delta\sigma_k}_{\text{always missing}} + \mathcal{O}\left(1/\Lambda^6\right)$$

$\Rightarrow$ Difference between local and global distances $\leftrightarrow$ size of $\mathcal{O}\left(1/\Lambda^4\right)$ effects

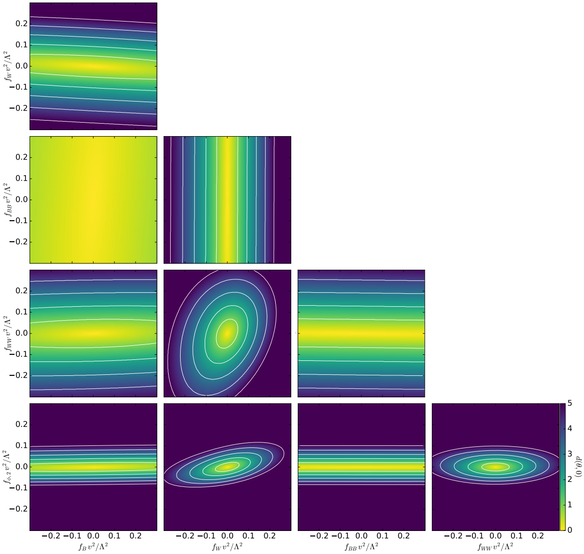
Global vs local distances for WBF



Contours of local (dashed) and global (solid) distances $d = 1, 2, 3, \ldots$ from SM

Other parameters set to zero

WBF distances



Distances from SM $d(\boldsymbol{\theta},\mathbf{0})$

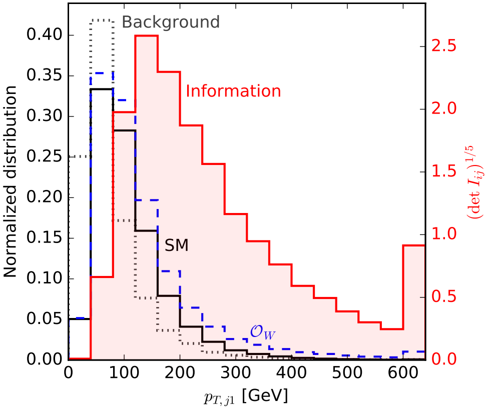
Optimal precision ($d = 1$):

$$\Delta(f_{\phi,2}\,v^2/\Lambda^2) \approx 0.02$$
$$\Delta(f_W\,v^2/\Lambda^2) \approx 0.05$$
$$\Delta(f_{WW}\,v^2/\Lambda^2) \approx 0.05$$

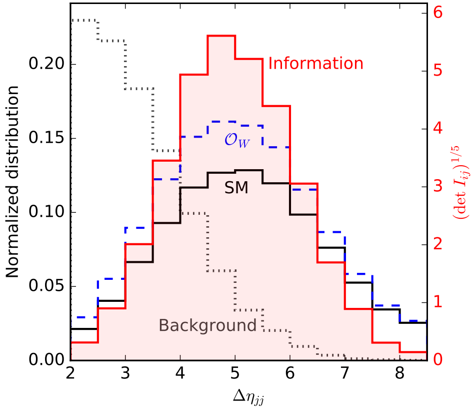
Differential information over $p_{T,j}$

Strongly correlated with momentum transfer E through production vertex:
measures $\mathcal{O} \sim \partial^2/\Lambda^2 \sim E^2/\Lambda^2$

SM
$f_W/\Lambda^2\, v^2 = 0.5$
QCD $Z \to \tau\tau$
$\det I_{ij}(\mathbf{0})$



Differential information over $\Delta\eta_{jj}$

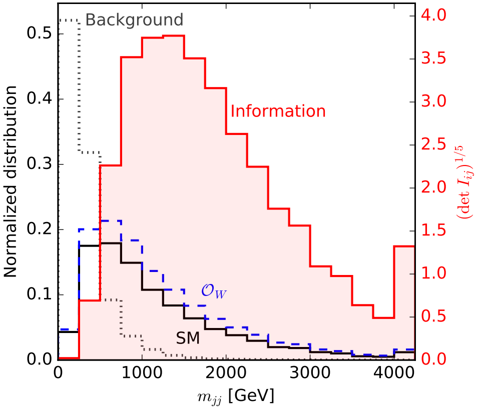


Trade-off:

- Background suppression better at large $\Delta\eta_{jj}$
- Momentum-dependent operators have largest effects at medium $\Delta\eta_{jj}$

SM
$f_W/\Lambda^2 v^2 = 0.5$
QCD $Z \to \tau\tau$
$\det I_{ij}(\mathbf{0})$

Differential information over m_{jj}



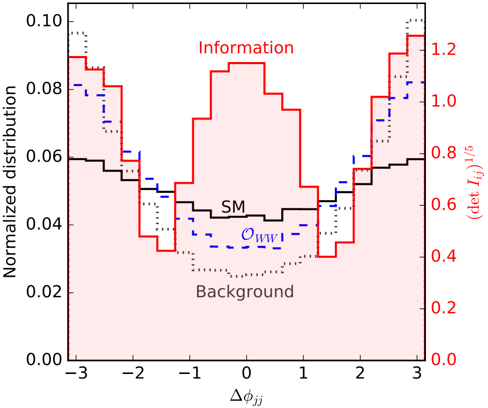
SM

$f_W/\Lambda^2\, v^2 = 0.5$

QCD $Z \to \tau\tau$

$\det I_{ij}(\mathbf{0})$

Differential information over $\Delta\phi_{jj}$



SM

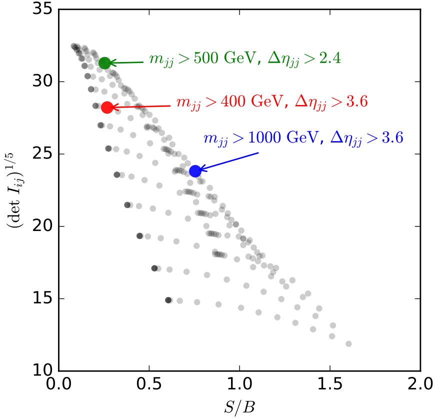
$f_{WW}/\Lambda^2 v^2 = 0.5$

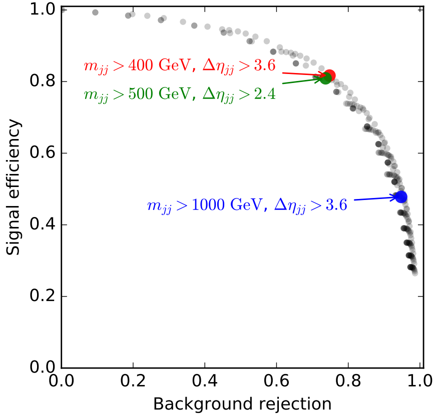
QCD $Z \to \tau\tau$

$\det I_{ij}(\mathbf{0})$

Optimizing cuts

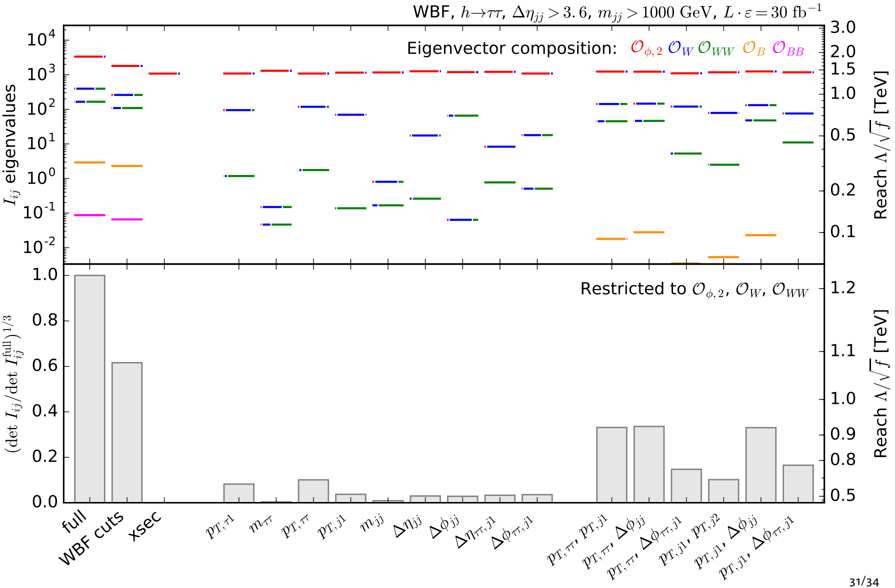
- Scan over m_{jj} and $\Delta\eta_{jj}$ cuts $\rightsquigarrow$ signal and background rate, $I_{ij}(\mathbf{0})$
- Trade-off between information and purity (left)
- Standard ROC curves (right) can be misleading





Common cuts: 105 GeV $< m_{\tau\tau} <$ 165 GeV, $p_{T,j1} >$ 50 GeV

WBF observables after conventional cuts



Adding systematic uncertainties

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Procedure:

- Add nuisance parameter to Fisher information:

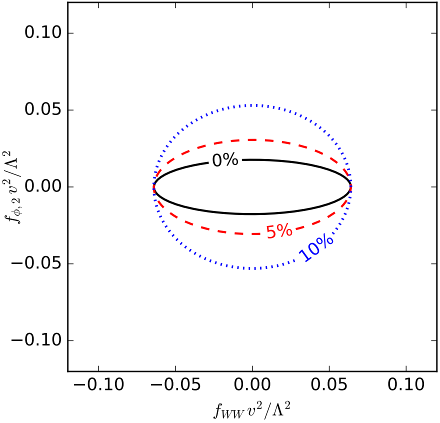
$$I_{ij} = \begin{pmatrix} I_t & I_m^T \\ I_m & I_n \end{pmatrix}$$

- Profiled information:

$$I_{\text{profiled}} = I_t - I_m^T \, I_n^{-1} \, I_m$$

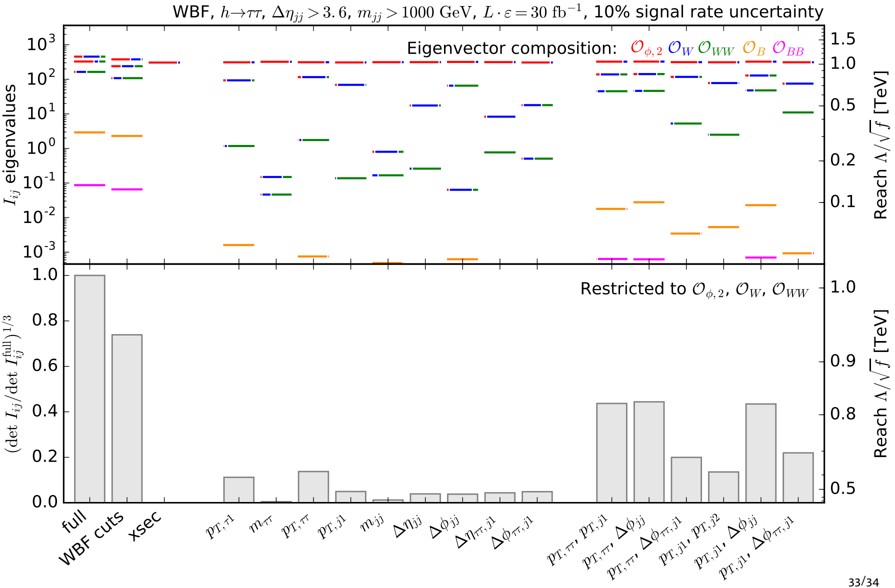
[T. Edwards, C. Weniger 1704.05458]

Local distances from SM, profiled over Gaussian uncertainties of 5% or 10% on signal rate:



Other parameters set to zero

WBF observables with systematics



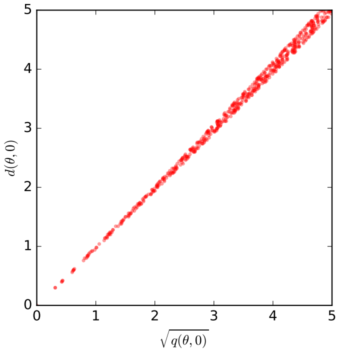
Fisher information vs likelihood ratio

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

- Confidence intervals based on hypothesis tests with likelihood ratio: are the Fisher information results relevant?

- Check!
  - Sample points $\boldsymbol{\theta}$ in $\mathcal{O}_W$-$\mathcal{O}_{WW}$ plane
  - Compare information distance $d(\boldsymbol{\theta}, \mathbf{0})$ to expected log likelihood ratio

$$q(\boldsymbol{\theta}|\mathbf{0}) = E\left[-2\,\log \frac{f(\mathbf{x}|\boldsymbol{\theta})}{f(\mathbf{x}|\mathbf{0})}\middle|\mathbf{0}\right]$$



$\Rightarrow$ Conclusions from information approach should also apply to limit setting