How to optimize Higgs Measurements using Information Geometry

work with **Johann Brehmer**, Kyle Cranmer and Tilman Plehn

see previous talk

arXiv: 1612.05261

Felix Kling

UCIrvine

Pheno 2017

Introduction

What we learned from Johann's talk

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

$$f(x|\theta) \longrightarrow I_{ij}(\theta) = -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial\theta_i \partial\theta_j}\bigg|\theta\right] \longrightarrow \mathrm{cov}[\hat{\theta}|\theta_0] \leq I_{ij}^{-1}(\theta_0)$$

probability distribution — Fisher Information — Cramer-Rao Bound

Introduction

What we learned from Johann's talk

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

$$f(x|\theta) \longrightarrow I_{ij}(\theta) = -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial\theta_i \partial\theta_j}\bigg|\theta\right] \longrightarrow \mathrm{cov}[\hat{\theta}|\theta_0] \leq I_{ij}^{-1}(\theta_0)$$

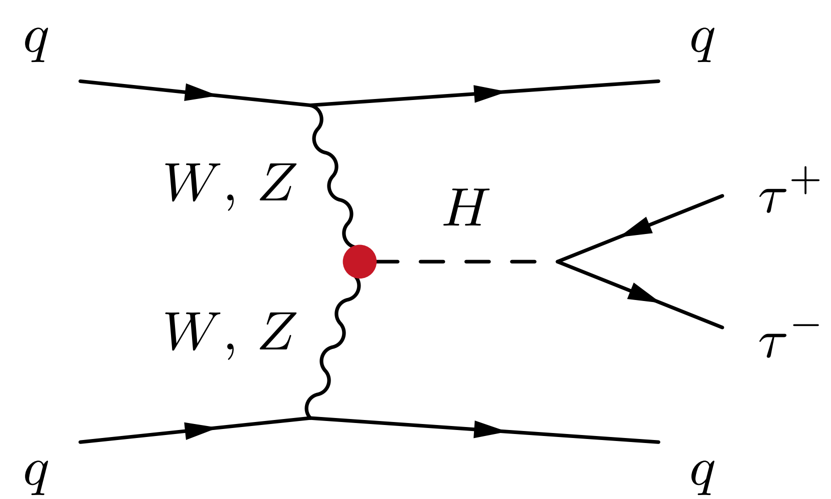
probability distribution — Fisher Information — Cramer-Rao Bound

↓

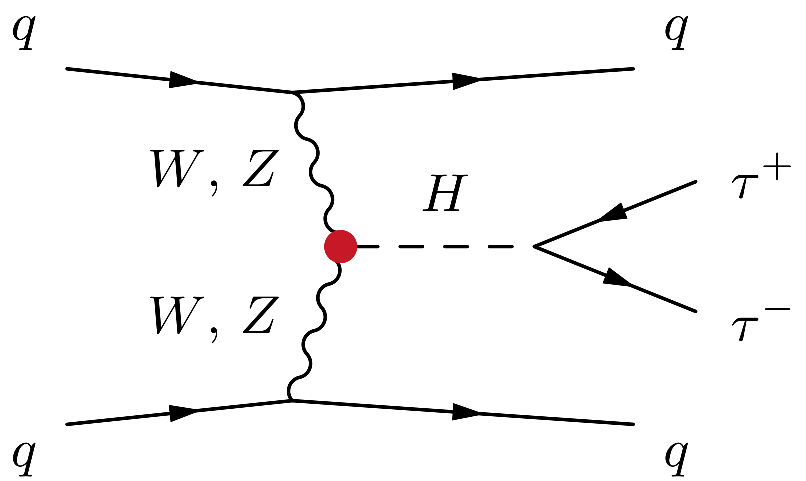
Application:

Constraining Higgs Effective Field Theory in Weak Boson Fusion

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{f_i^{d=6}}{\Lambda^2}\mathcal{O}_i^{d=6} + \cdots$$



Introduction



Weak Boson Fusion:

- well known probe of Higgs-gauge structure
- interesting kinematics of tagging jets

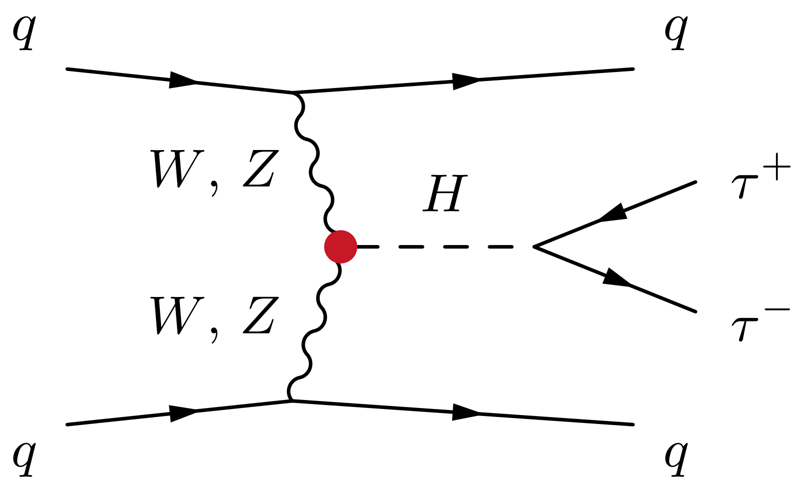
see hep-ph/9808468, hep-ph/0105325, 1212.0843

Introduction

Weak Boson Fusion:

- well known probe of Higgs-gauge structure
- interesting kinematics of tagging jets

see hep-ph/9808468, hep-ph/0105325, 1212.0843



Theory Language: Dim-6-Operators of HEFT: $\mathcal{L} \supset \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i$

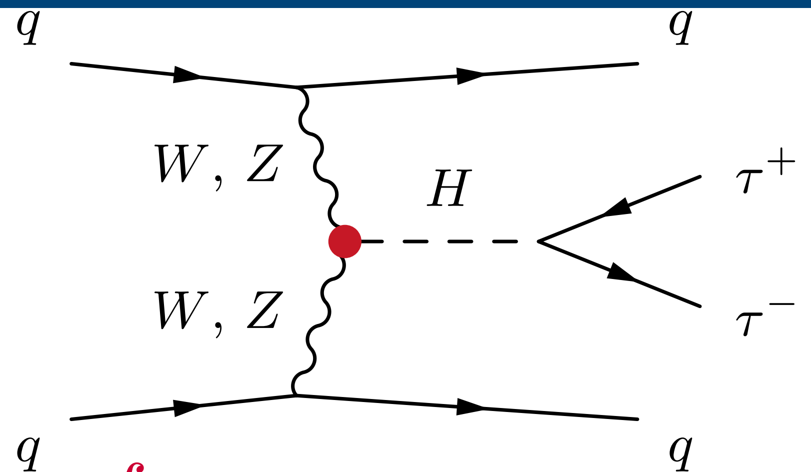
- total rate: $\mathcal{O}_{\phi,2} = \frac{1}{2}\, \partial^\mu (\phi^\dagger \phi)\, \partial_\mu (\phi^\dagger \phi)$
- new kinematic structures:

$$\mathcal{O}_W = i \frac{g}{2} (D^\mu \phi)^\dagger \sigma^k (D^\nu \phi)\, W^k_{\mu\nu} \qquad \mathcal{O}_{WW} = -\frac{g^2}{4} (\phi^\dagger \phi)\, W^k_{\mu\nu}\, W^{\mu\nu\, k}$$

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

- model parameters: wilson coefficients: $\theta = \left(f_{\phi,2},\, f_W,\, f_{WW},\, f_B,\, f_{BB}\right)$

Introduction



Weak Boson Fusion:

- well known probe of Higgs-gauge structure
- interesting kinematics of tagging jets

see hep-ph/9808468, hep-ph/0105325, 1212.0843

Theory Language: Dim-6-Operators of HEFT: $\mathcal{L} \supset \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i$

- total rate: $\mathcal{O}_{\phi,2} = \frac{1}{2}\,\partial^\mu(\phi^\dagger\phi)\,\partial_\mu(\phi^\dagger\phi)$
- new kinematic structures:

$$\mathcal{O}_W = i\frac{g}{2}\,(D^\mu\phi)^\dagger\sigma^k(D^\nu\phi)\,W^k_{\mu\nu} \qquad \mathcal{O}_{WW} = -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}\,W^{\mu\nu\,k}$$

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

- model parameters: wilson coefficients: $\theta = \left(f_{\phi,2}, f_W, f_{WW}, f_B, f_{BB}\right)\frac{v^2}{\Lambda^2}$

Outline:
- 1. How does the tool work?
- 2. How well can we measure Wilson coefficients?
- 3. Where in phase space is the information?
- 4. How good is my analyses?
- 5. Does the EFT series converge?

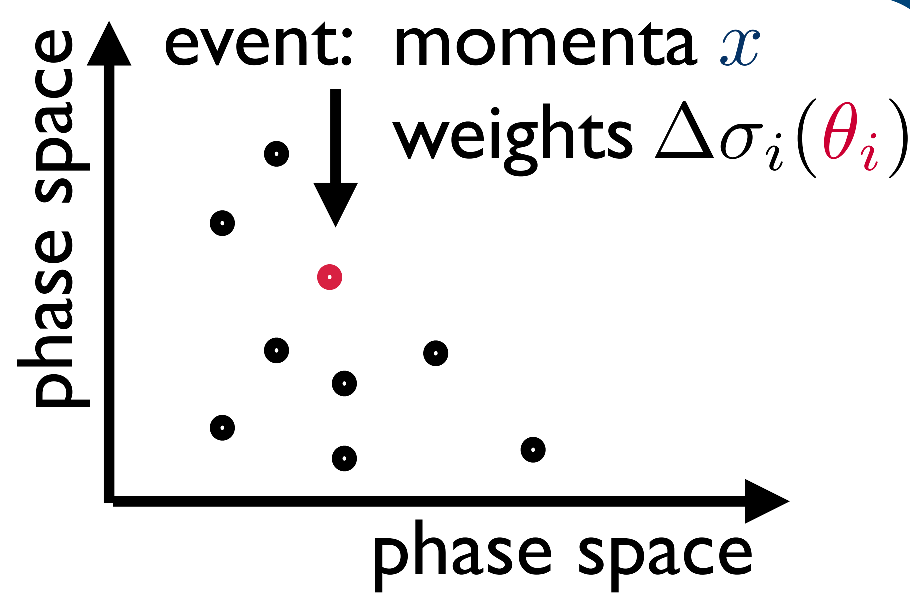
1. How does the tool work?

1. Generate events via Monte Carlo: MadMax

- modified version of MG5

see 1311.2591, 1607.07441

- modeling detector response
  - Smearing
  - Efficiencies
  - Trigger Cuts
  - Backgrounds



2. Morphing: get weights for arbitrary model θ

- $\Delta\sigma(\theta) = \sum_i a_i \Delta\sigma_i$

see ATL-PHYS-PUB-2015-047

3. Calculate Fisher Information

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

4. Cramer Rao Bound defines maximal sensitivity $\mathrm{cov}[\hat{\theta}|\theta_0] \leq I_{ij}^{-1}(\theta_0)$

- reach for $\theta = f v^2/\Lambda^2$

2. How well can we measure Wilson coefficients?

$\sqrt{s} = 13\ \mathrm{TeV}$
$L \cdot \epsilon = 30\ \mathrm{fb}^{-1}$

Fisher Info at the SM:

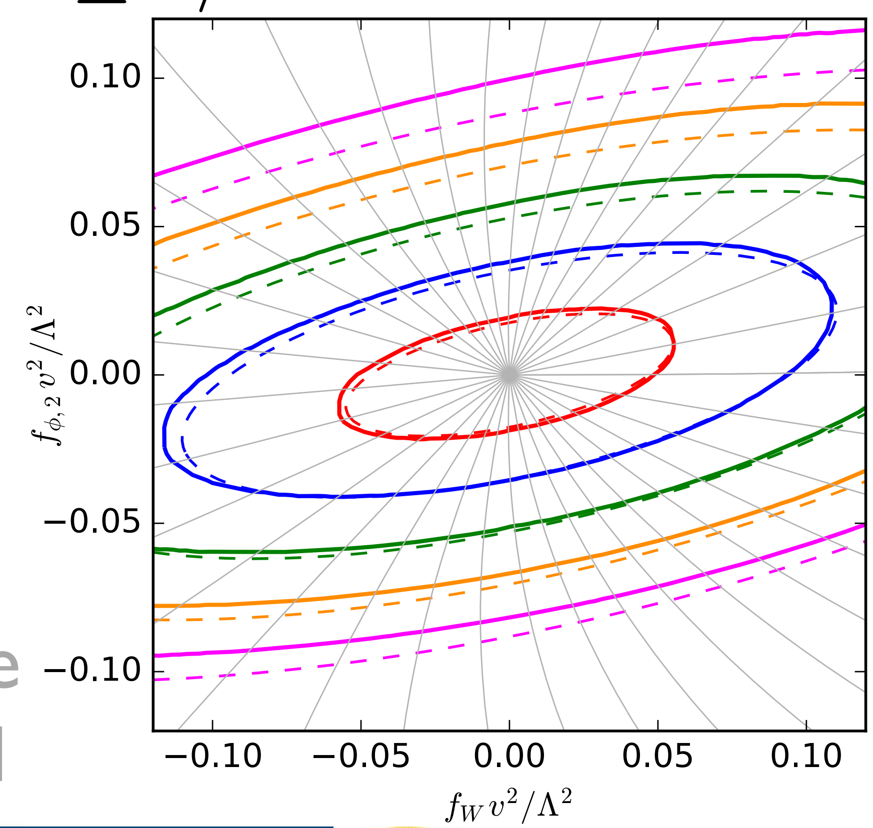
- How well can we constrain new physics, assuming SM is true?

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

- Cramer-Rao bound: minimum measurement error $\Delta\theta \geq 1/\sqrt{I}$
- $\mathcal{O}_{\phi,2}$ direction can be measured best, followed by $\mathcal{O}_W, \mathcal{O}_{WW}$. Large mixing between operators

$$\Delta(f_{\phi,2}v^2/\Lambda^2) \gtrsim 0.02 \qquad \Delta(f_W v^2/\Lambda^2) \gtrsim 0.13$$

- distance measure $d^2 = I_{ij}(\theta_0)(\theta^i - \theta_0^i)(\theta^j - \theta_0^j)$
  ~ unlikeliness to measure θ if θ_0 is true 'in sigmas'

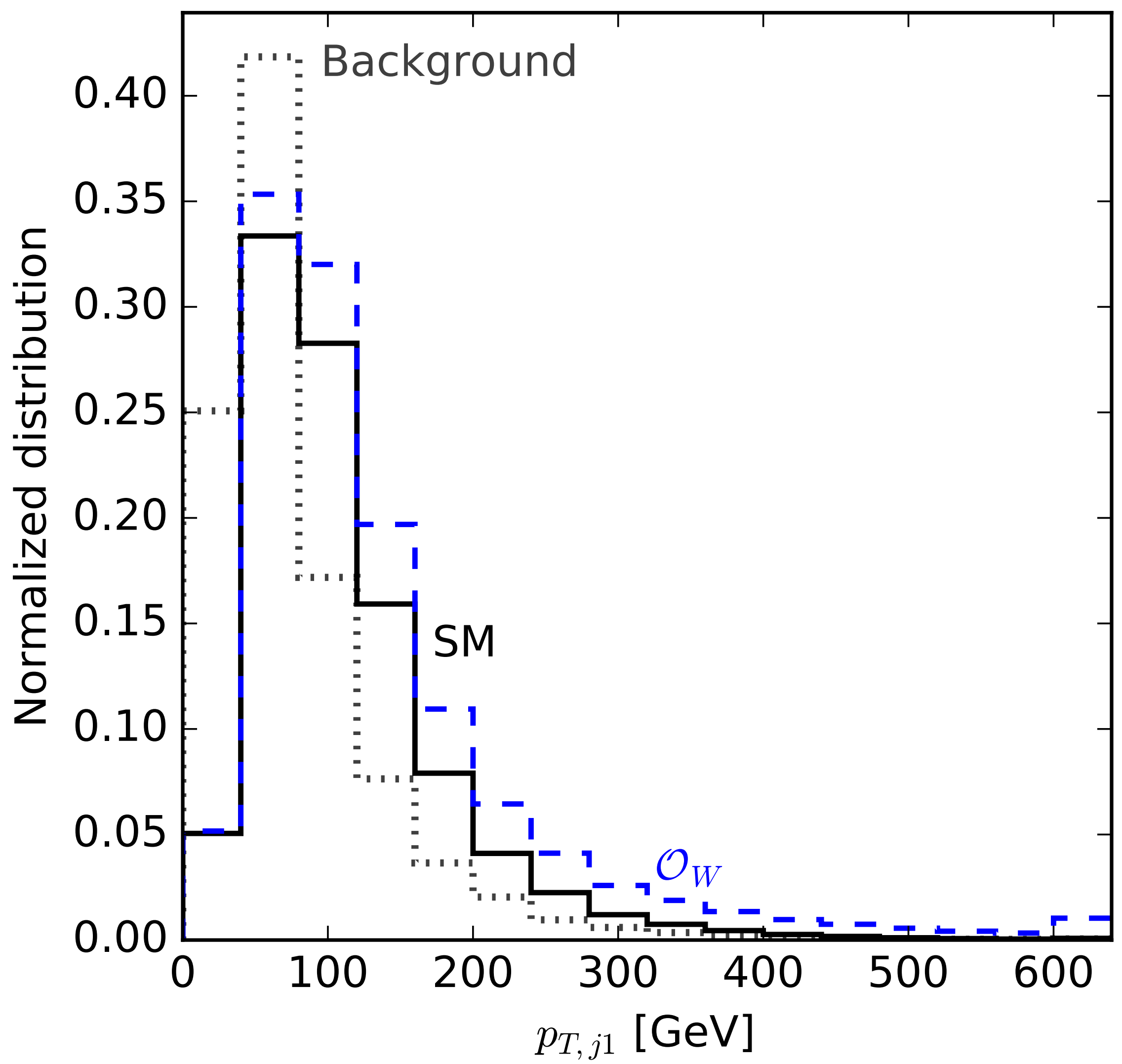


Contours of distance d=1,2,3,4,5 from SM

3. Where in phase space is the information?

pT of the hardest tagging jet

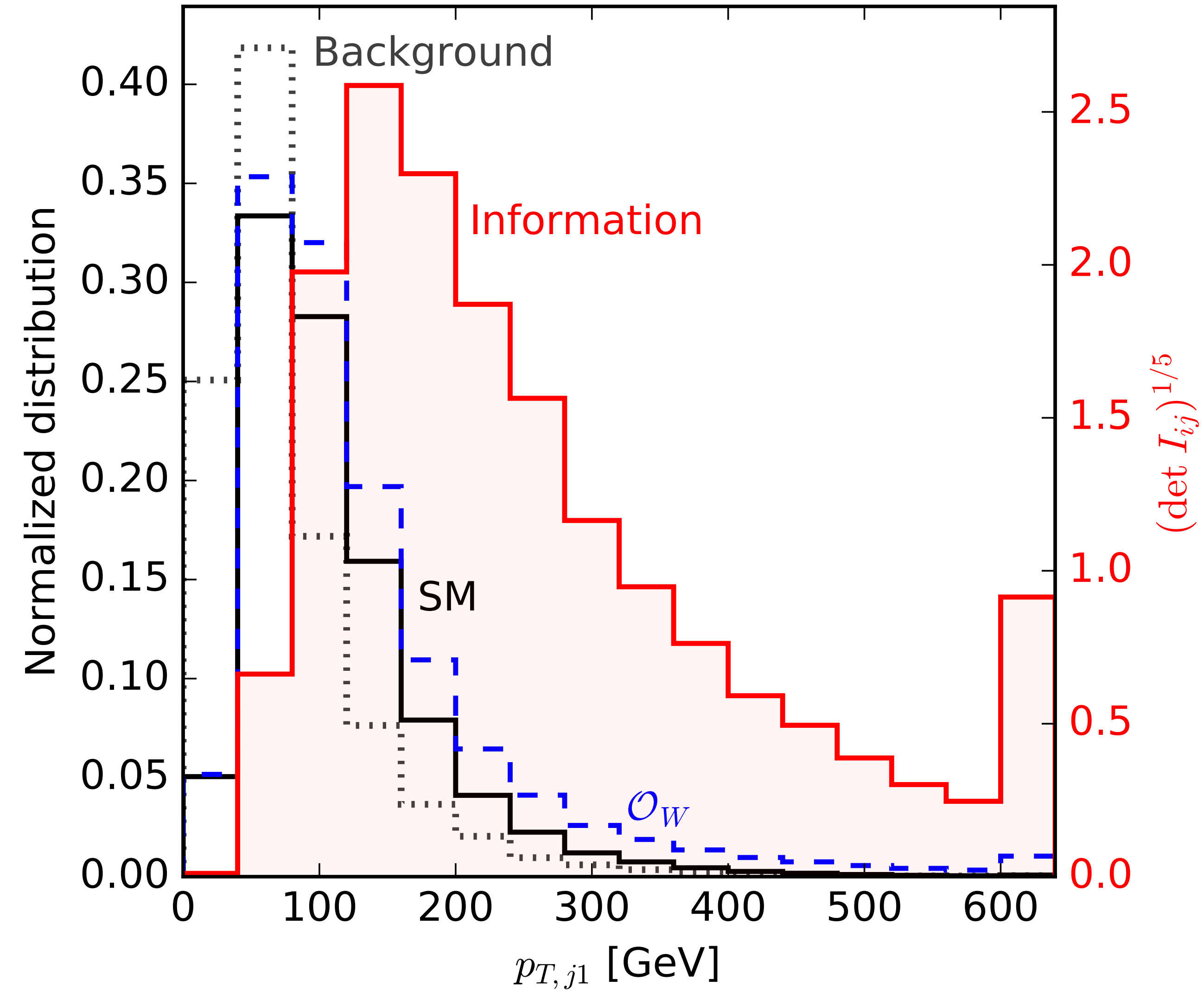
- virtuality measure:
  Strongly correlated with momentum transfer through production vertex
- increasing amplitude of dim-6 operator



3. Where in phase space is the information?

pT of the hardest tagging jet

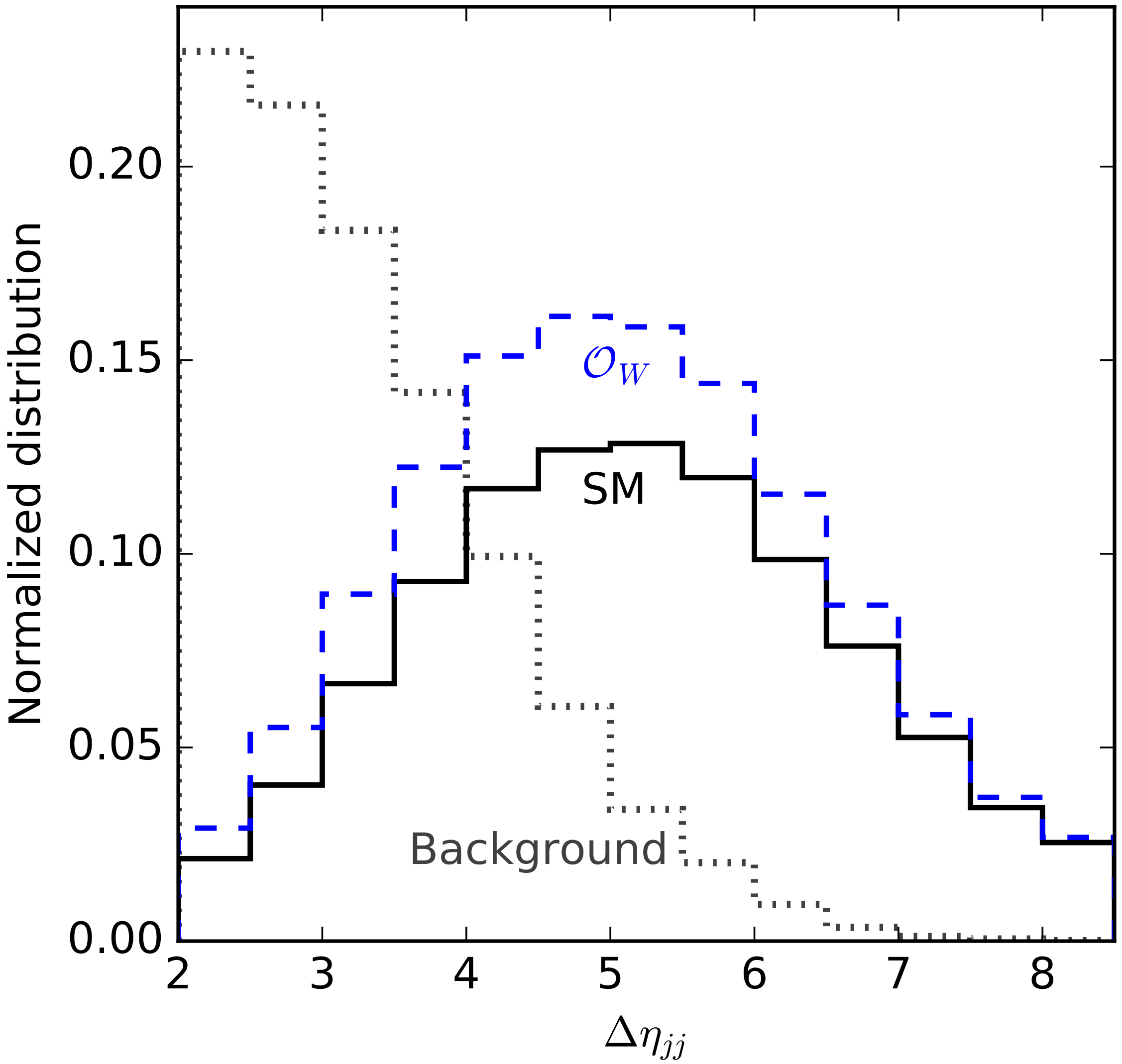
- virtuality measure:
  Strongly correlated with momentum transfer through production vertex
- increasing amplitude of dim-6 operator
- most information in high pT tail



3. Where in phase space is the information?

pT of the hardest tagging jet

- virtuality measure:
  Strongly correlated with momentum transfer through production vertex
- increasing amplitude of dim-6 operator
- most information in high pT tail



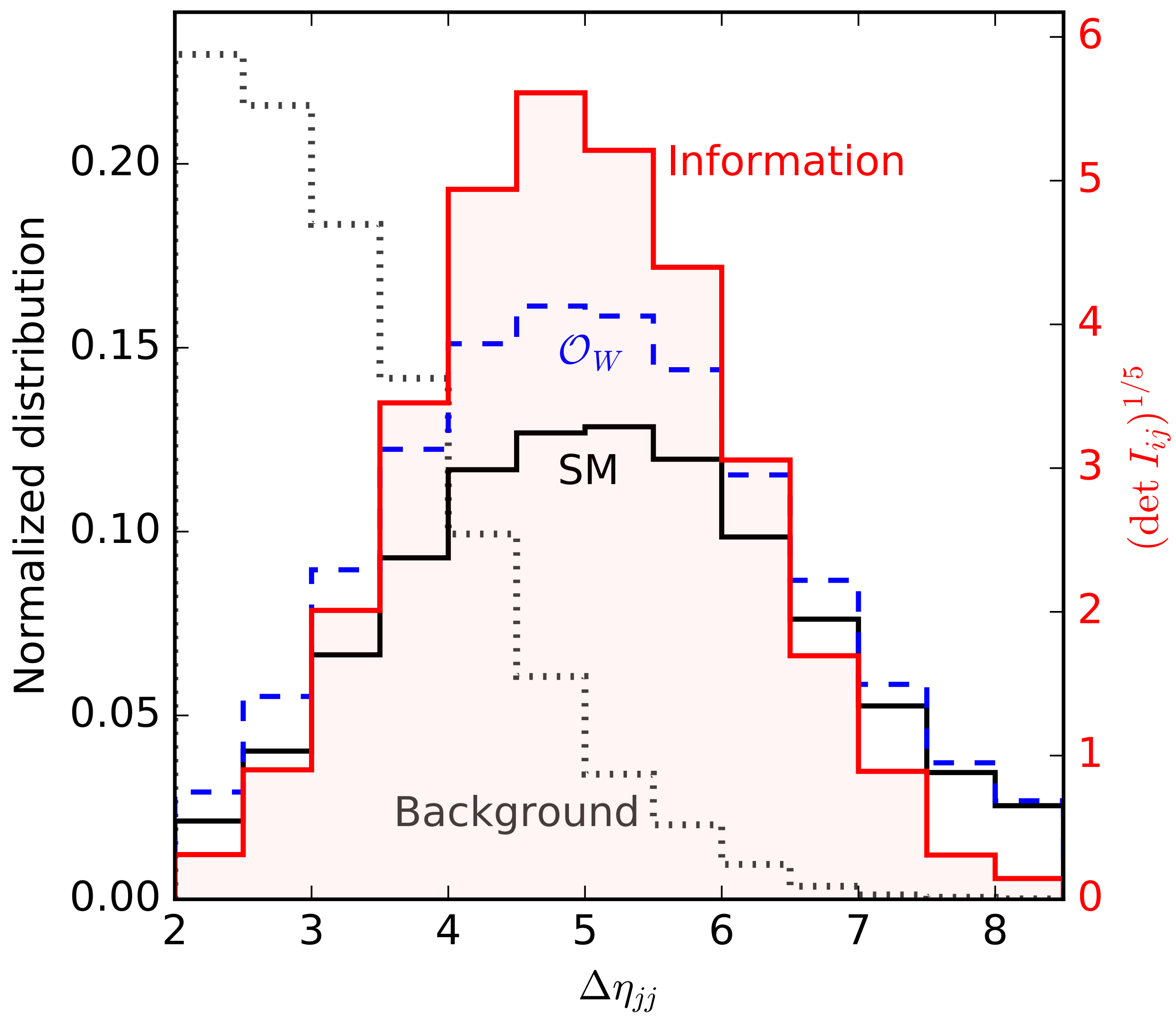
jet rapidity difference

- better background suppression at large $\Delta\eta_{jj}$
- momentum dependent operators have largest effect at medium $\Delta\eta_{jj}$

3. Where in phase space is the information?

pT of the hardest tagging jet

- virtuality measure: Strongly correlated with momentum transfer through production vertex
- increasing amplitude of dim-6 operator
- most information in high pT tail



jet rapidity difference

- better background suppression at large $\Delta\eta_{jj}$
- momentum dependent operators have largest effect at medium $\Delta\eta_{jj}$
- strong WBF cuts ($\Delta\eta_{jj}$> 4.2) lose information of dim-6 operators

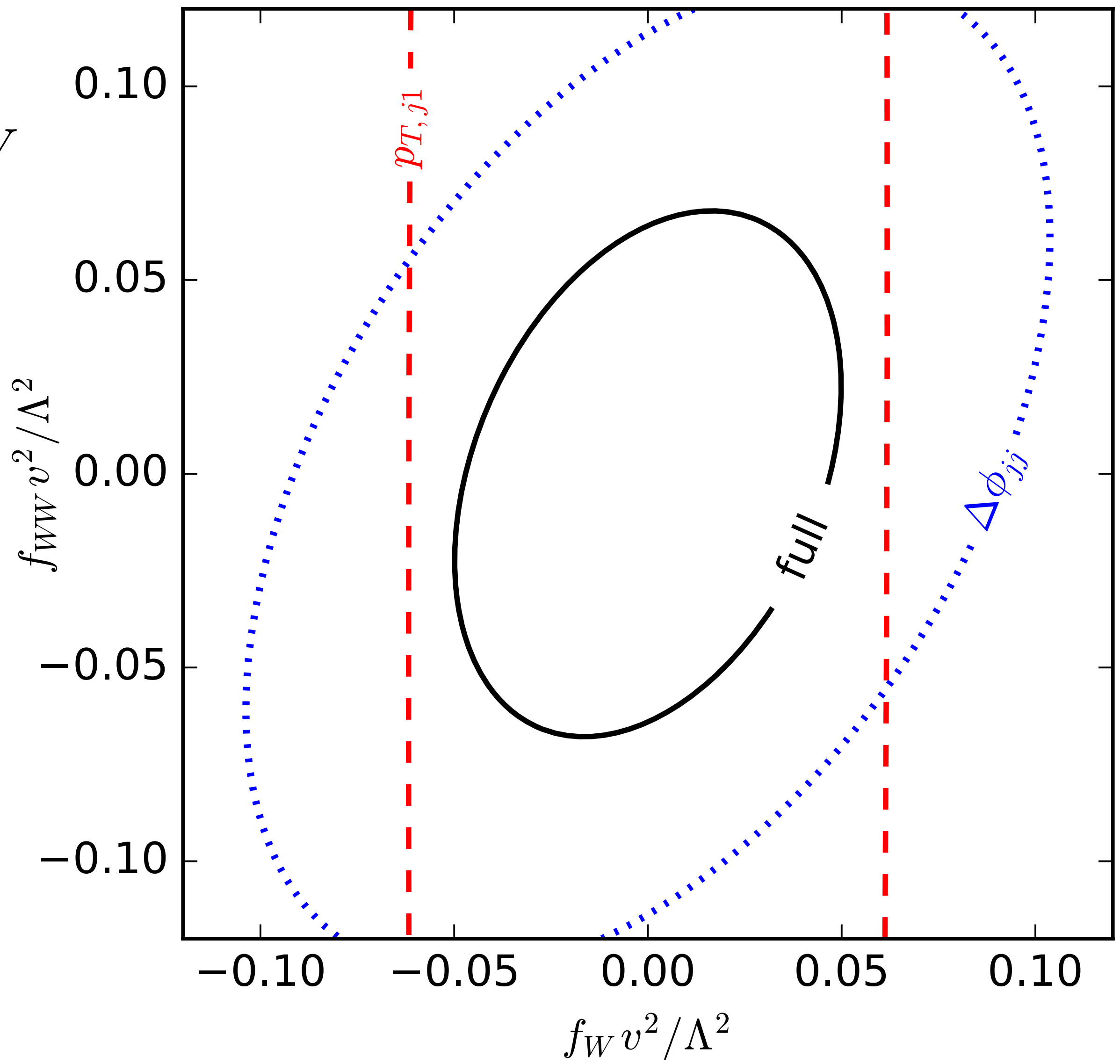
4. How good is my analysis?

Histograms:

- project high-dimensional phase space onto 1 or 2-dimensional histogram
- reduced information
- Allows to compare histogram-based and multivariate analyses

Histograms:

- Virtuality measure ($p_{T,j1}$) probes only $\mathcal{O}_W$
- Angular correlation between tagging jets ($\Delta\phi_{jj}$) sensitive to $\mathcal{O}_{WW}$
- **Full phase space** information



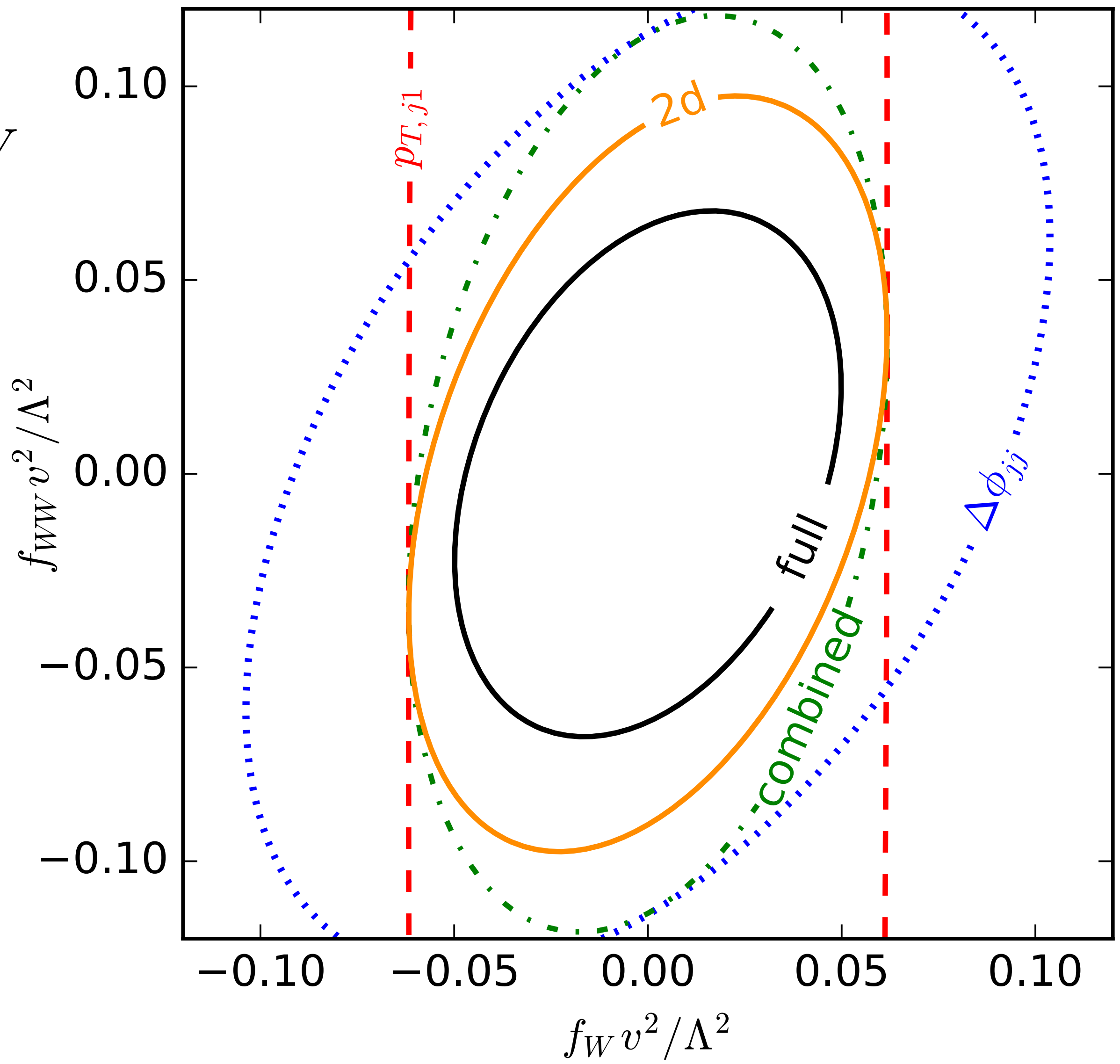
4. How good is my analysis?

Histograms:

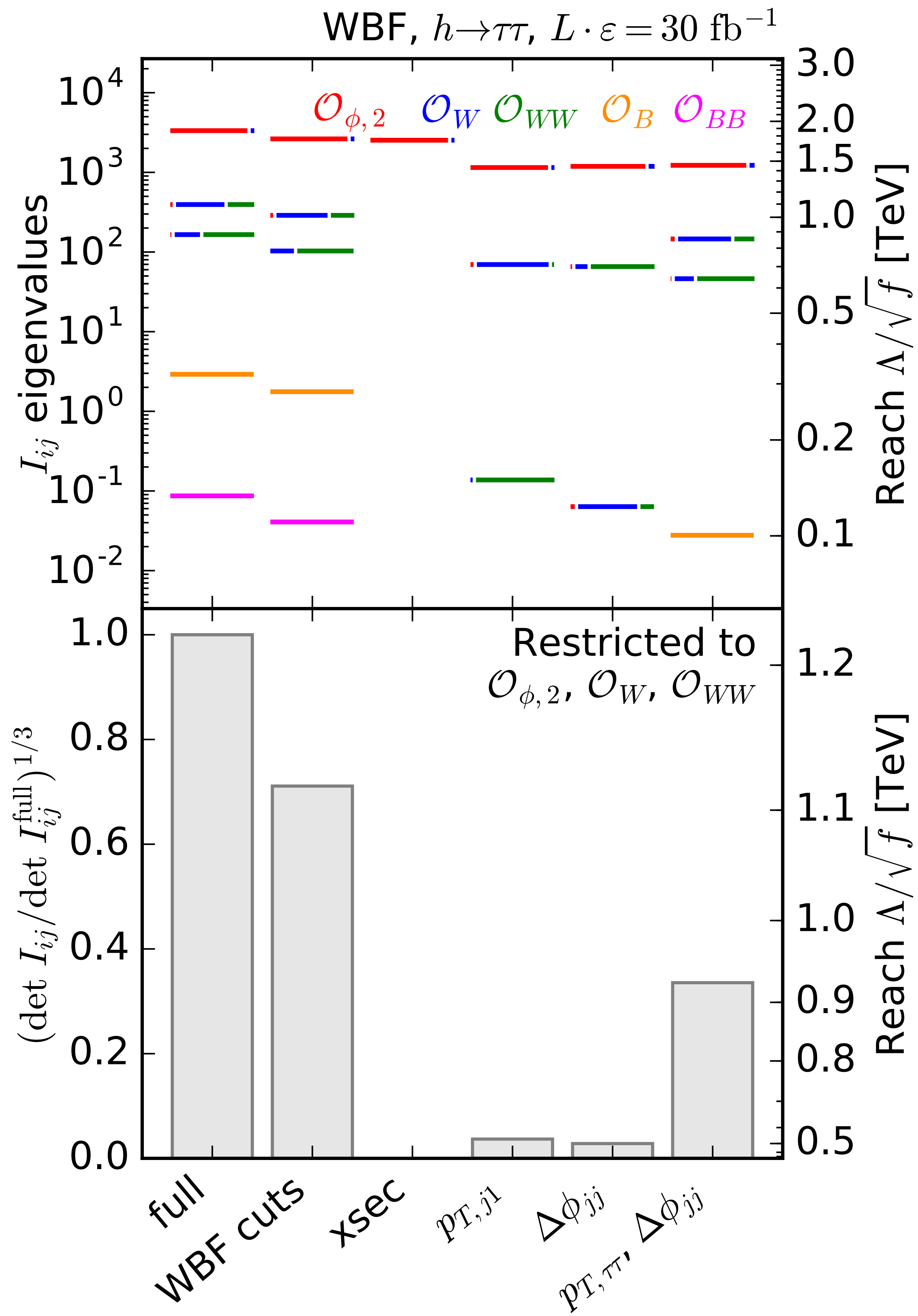
- project high-dimensional phase space onto 1 or 2-dimensional histogram
- reduced information
- Allows to compare histogram-based and multivariate analyses

Histograms:

- Virtuality measure ($p_{T,j1}$) probes only $\mathcal{O}_W$
- Angular correlation between tagging jets ($\Delta\phi_{jj}$) sensitive to $\mathcal{O}_{WW}$
- **Full phase space** information
- 2-dimensional histogram necessary for stringent constraints, but still not close to full information



4. How good is my analysis?



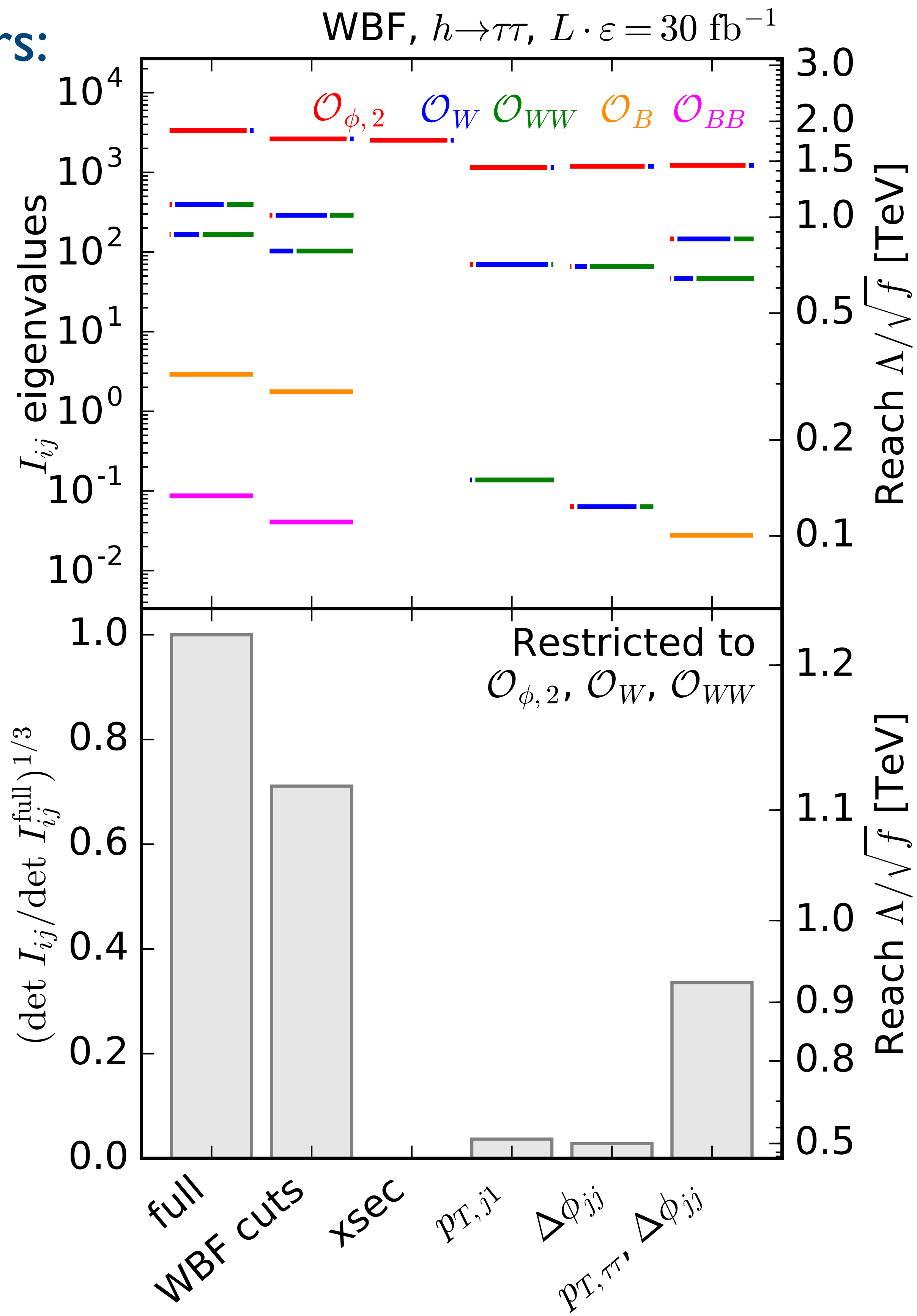
4. How good is my analysis?

Eigenvalues/Eigenvectors:

Which operators can be measured well in which distribution?

Determinant:

Invariant measure of information



Cramer Rao Bound:

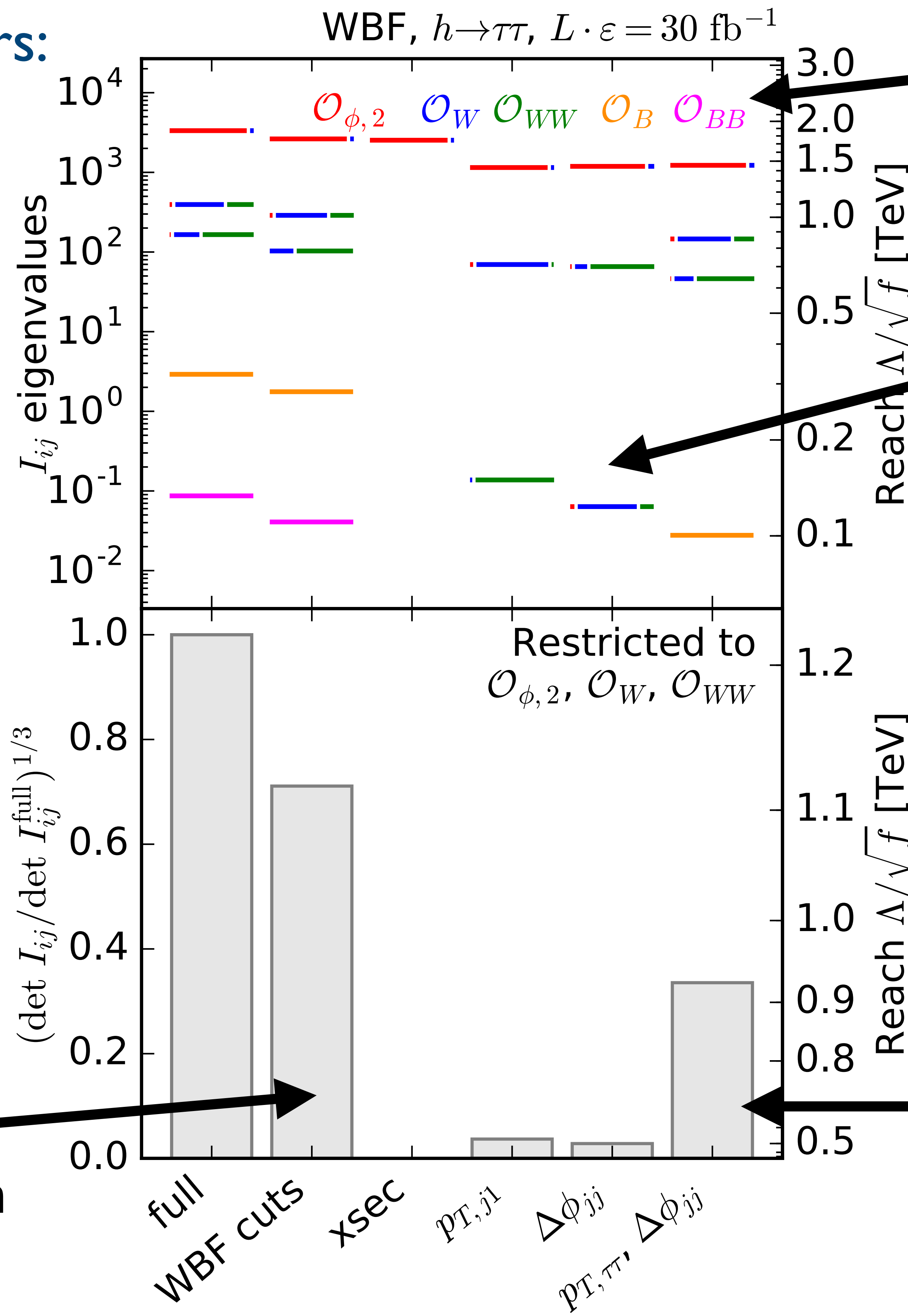
$$\frac{f v^2}{\Lambda^2} \sim \frac{1}{\sqrt{I}}$$

4. How good is my analysis?

Eigenvalues/Eigenvectors:
Which operators can be measured well in which distribution?

Determinant:
Invariant measure of information

WBF cuts lose 30% of information



$\mathcal{O}_B$, $\mathcal{O}_{BB}$ are poorly constrained

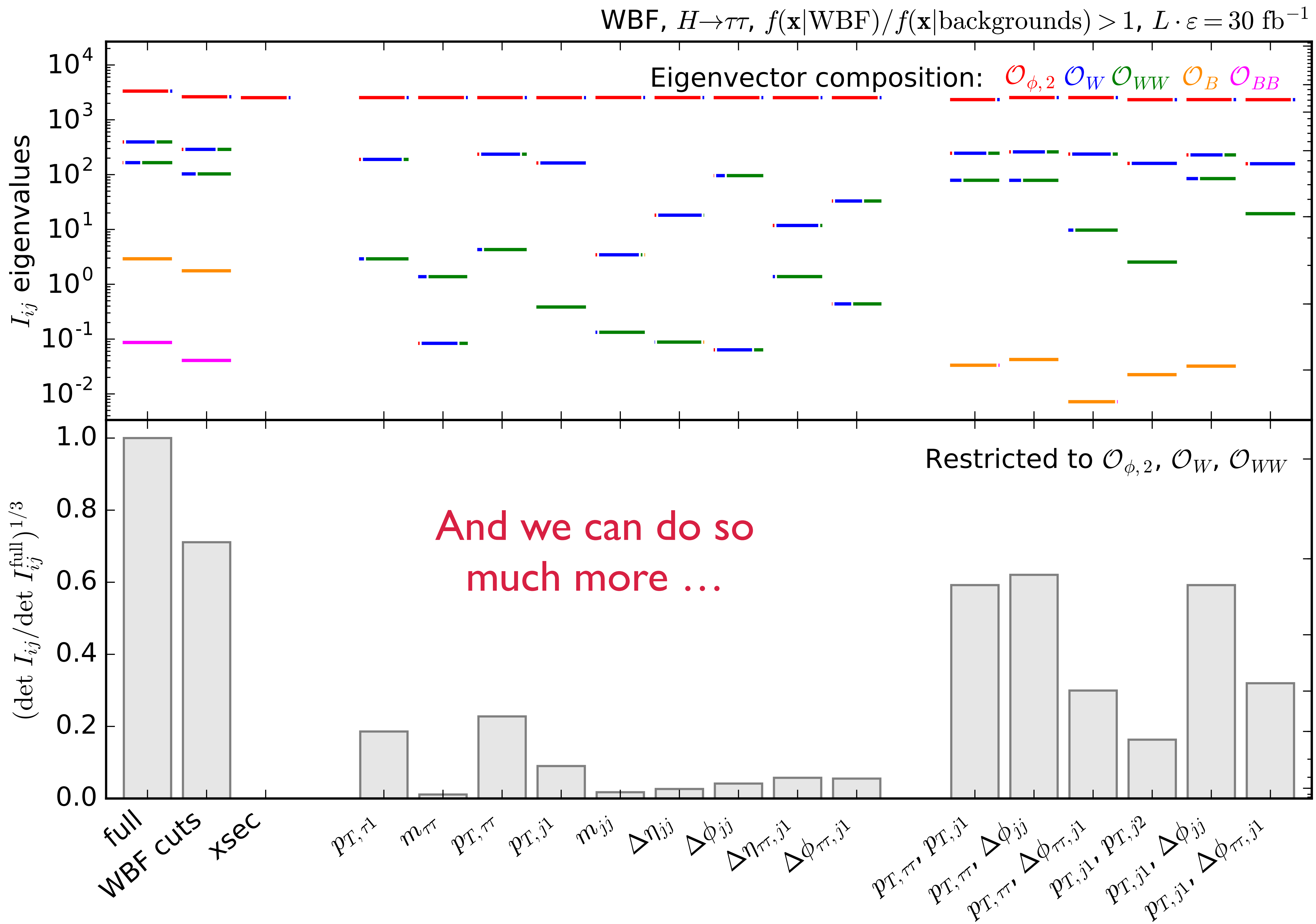
1D histogram suffer from blind directions

Cramer Rao Bound:

$$\frac{f v^2}{\Lambda^2} \sim \frac{1}{\sqrt{I}}$$

2D histogram still loses 65% of information

4. How good is my analysis?



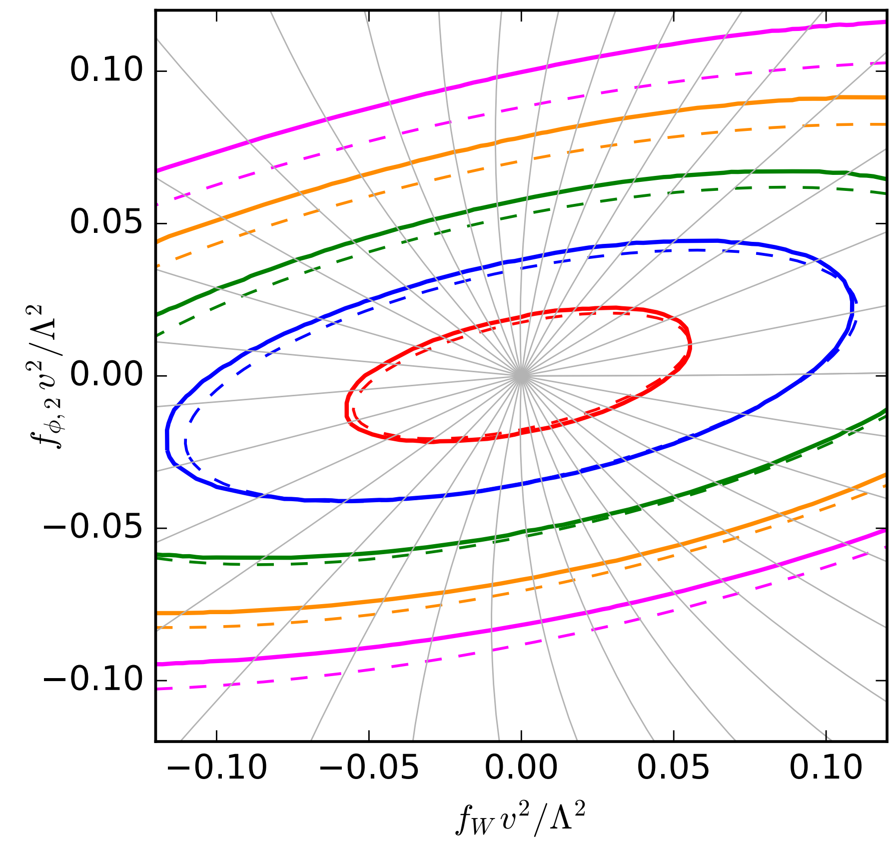
5. Does the EFT series converge?

Distance Measure ~ unlikeliness to measure θ if θ_0 is true 'in sigmas'

local distance: $d^2 = I_{ij}(\theta_0)(\theta^i - \theta_0^i)(\theta^j - \theta_0^j)$
(dashed)

global distance: $d = \min_{\theta(s)} \int_{s_a}^{s_b} ds \sqrt{I_{ij} \frac{d\theta_i}{ds} \frac{d\theta_j}{ds}}$
(solid)

Contours of distance
d=1,2,3,4,5 from SM

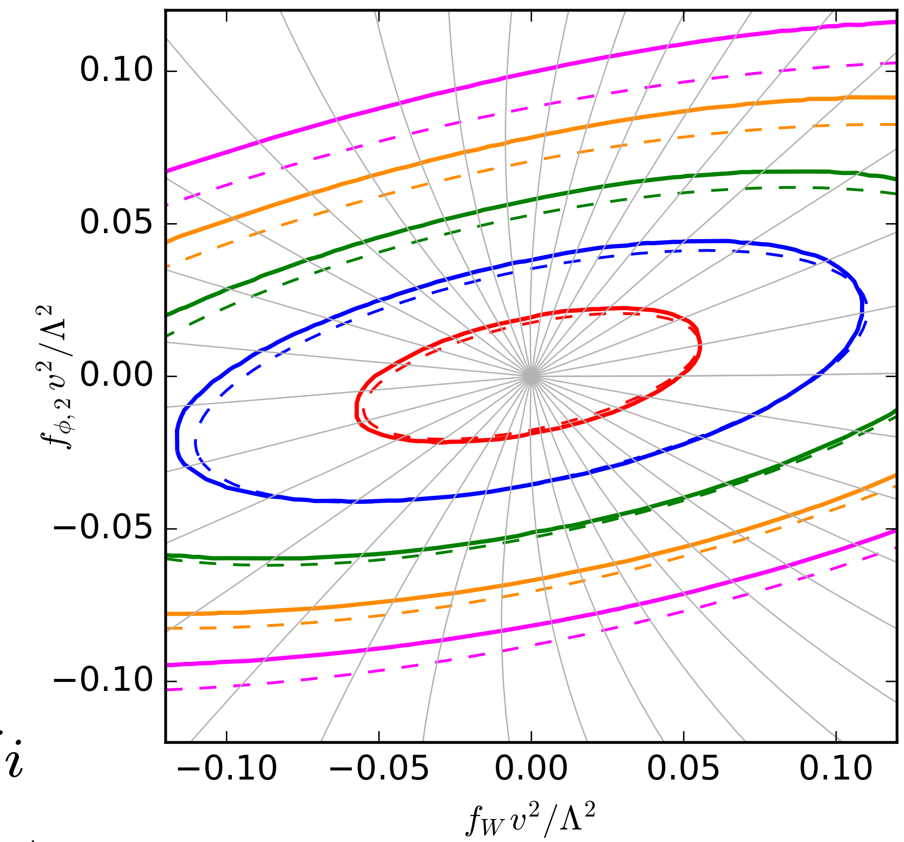


5. Does the EFT series converge?

→ **Distance Measure** ~ unlikeliness to measure θ if θ_0 is true 'in sigmas'

local distance: $d^2 = I_{ij}(\theta_0)(\theta^i - \theta^i_0)(\theta^j - \theta^j_0)$
(dashed)

global distance: $d = \min_{\theta(s)} \int_{s_a}^{s_b} ds \sqrt{I_{ij} \frac{d\theta_i}{ds} \frac{d\theta_j}{ds}}$
(solid)



Contours of distance
d=1,2,3,4,5 from SM

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

→ Information geometry for dim-6 operators $\theta_i = f_i^{d=6} v^2/\Lambda^2$

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

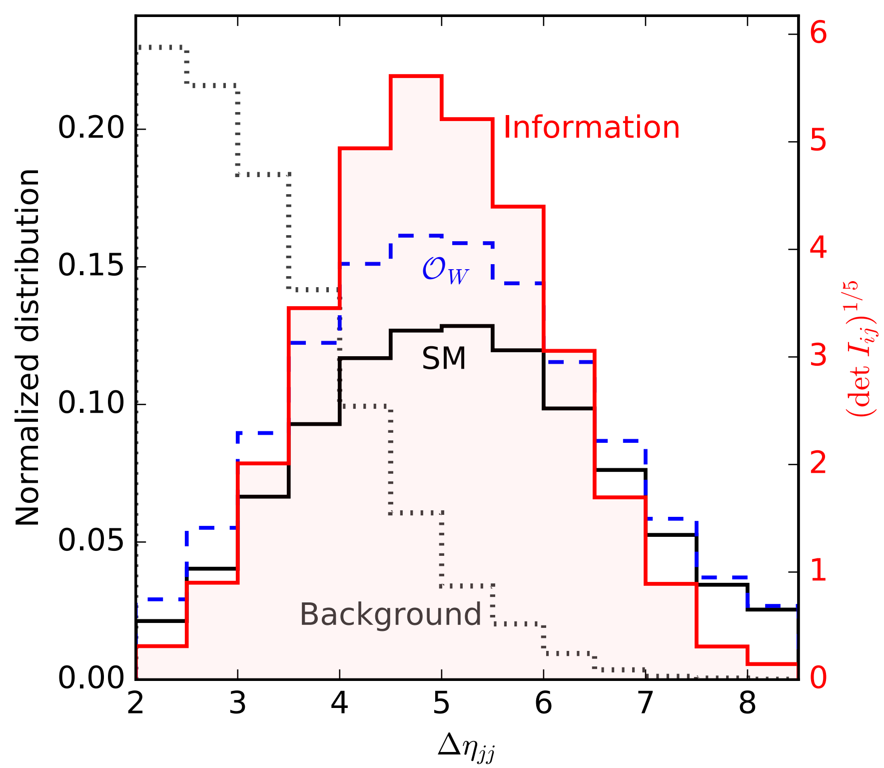
Difference between local/global distance ⟷ size of $\mathcal{O}(\Lambda^{-4})$ effects

Conclusion

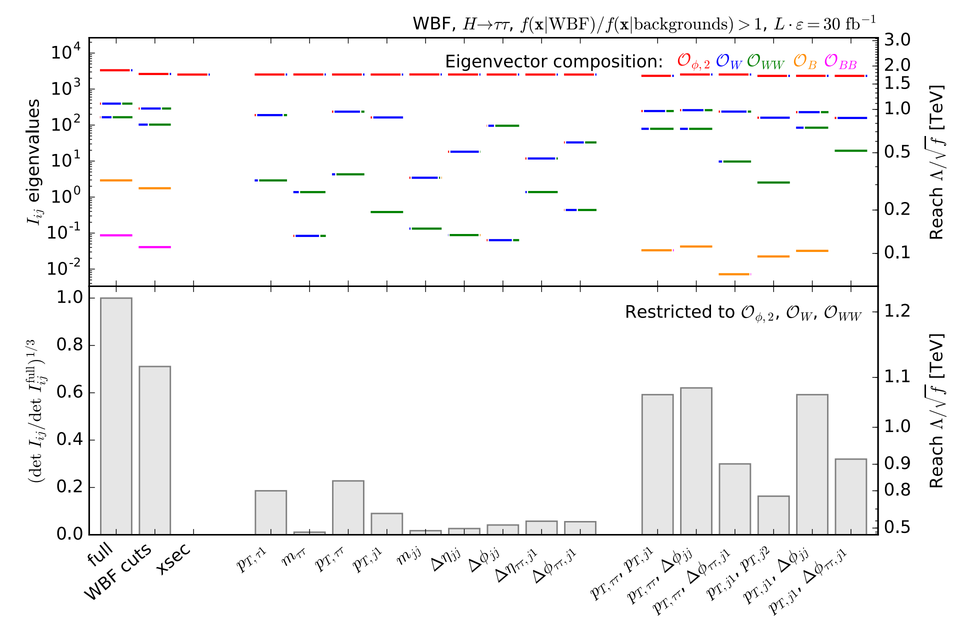
WBF in the HEFT framework as an example for Information Geometry

We can predict maximum precision at which Wilson coefficients can be measured.

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



We can obtain differential distribution of information.



We can quantitatively compare performance of histogram-based and multivariate analyses.