# Simulation-based inference in the search for CP violation in leptonic WH production

LHC Higgs Working Group WG2+WG3 meeting

Ricardo Barrué (IST, LIP), Patricia Muiño (IST, LIP), Rui Santos (CFTC, ISEL), Valerio Dao (CERN)

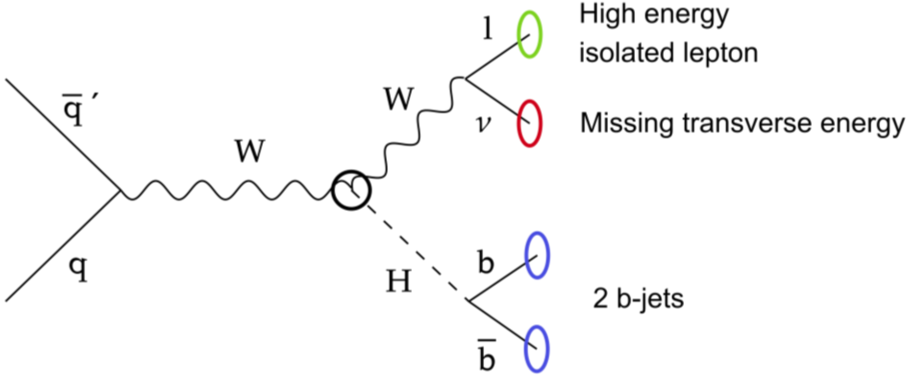
TÉCNICO LISBOA
LIP
FCT Fundação para a Ciência e a Tecnologia
SFRH/BD/150792/2020
REPÚBLICA PORTUGUESA
CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

# Outline

1. CP violation in HWW interaction via WH production
2. Optimal observables
3. Simulation-based inference and optimal observables
4. Analysis introduction
5. Some distributions
6. Results
7. Conclusions

# CP violation in HWW interaction

Goal: optimize search for CP violation in the HWW interaction via WH production



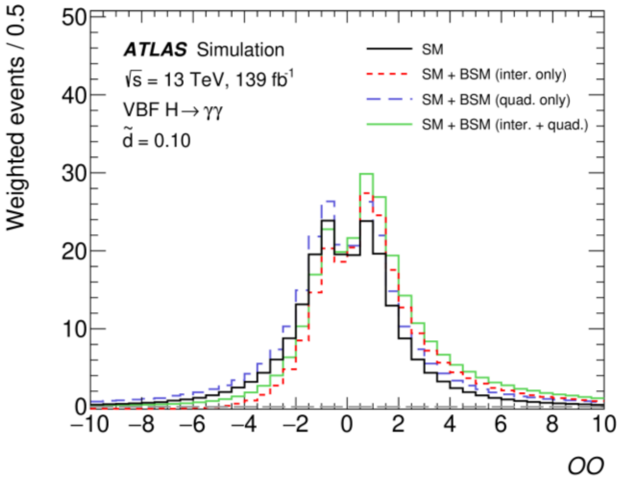

SMEFT, Warsaw basis, **1** dimension-6 CP-odd operator

$$O_{H\widetilde{W}} = \frac{c_{H\widetilde{W}}}{\Lambda^2} H^\dagger H \, \epsilon_{\mu\nu\rho\sigma} W^{I\mu\nu} W^{I\rho\sigma}$$

# Optimal observables

Built from matrix elements, sensitive to interference between SM and BSM CP-odd components

$$OO = \frac{2\Re(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\text{SM}}|^2}$$




<u>Issues:</u>

- Neglect or approximate everything between parton-shower and reconstructed final state
- Require full reconstruction of final state

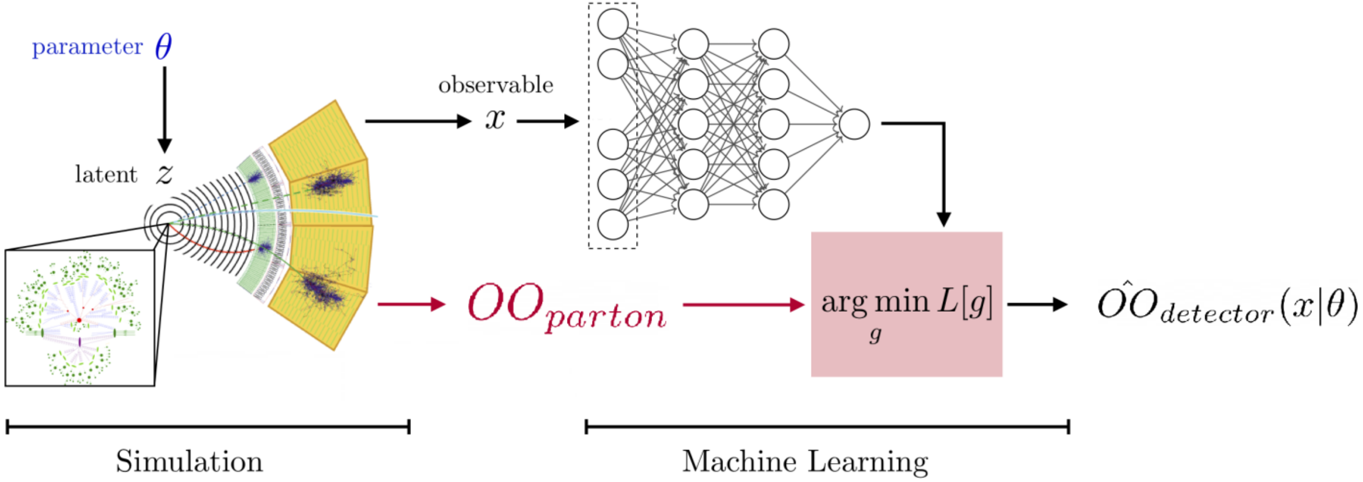
**Can we build an observable optimally sensitive to $c_{H\widetilde{W}}$ using standard detector-level variables ?**

**Can we go without the need to fully reconstruct the neutrino 4-vector ?**

**How does it compare with other observables in the literature ?**

# Simulation-based-inference/SALLY

**SALLY** (Score Aproximates the Likelihood LocallY) estimates detector-level optimal observable exploiting simulation information [1]




[1]: J. Brehmer et al, *MadMiner: Machine learning–based inference for particle physics*, arXiv:1907.10621 (figure largely adapted from figure in paper)

# Analysis introduction

Signal: SMEFTsim3, Λ=1 TeV, **full matrix element**

- $c_{H\widetilde{W}} = 0$ reweighted to 2 BSM benchmarks, interpolated using morphing.

Backgrounds: semileptonic $t\bar{t}$, W+(b)-jets, s-channel single top

Selection cuts applied at generator level [2]:

- $p_{T,\ell} > 10\ GeV, E_T^{miss} > 25\ GeV$
- $p_{T,b} > 35\ GeV$
- $|\eta_{\ell,b}| < 2.5$
- $\Delta R_{bj,\ell j} > 0.4, \Delta R_{bb,\ell b} > 0.4$
- $80\ GeV < m_{bb} < 160\ GeV$
- $p_{T,j} < 30\ GeV$

[2]: J. Brehmer et al, *Benchmarking simplified template cross-sections in WH production*, arXiv:1908.06980

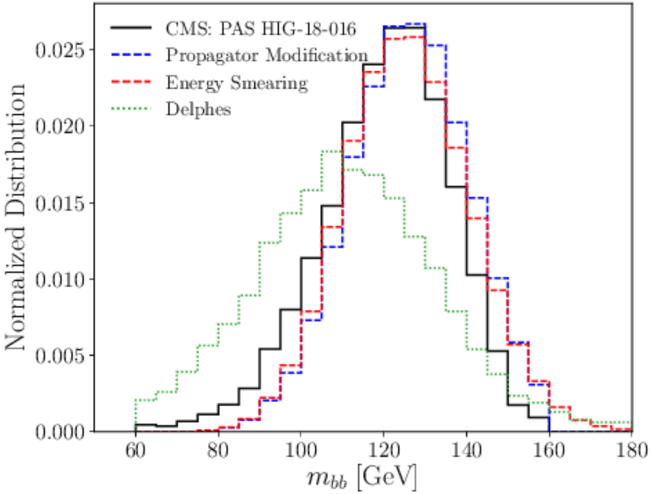
# PS, had. detector response approximation

Pythia+Delphes shown to have a large mismodelling of $m_{bb}$ and $E_T^{miss}$ [2]

Approximated by Gaussian smearing of particle-level quantities:

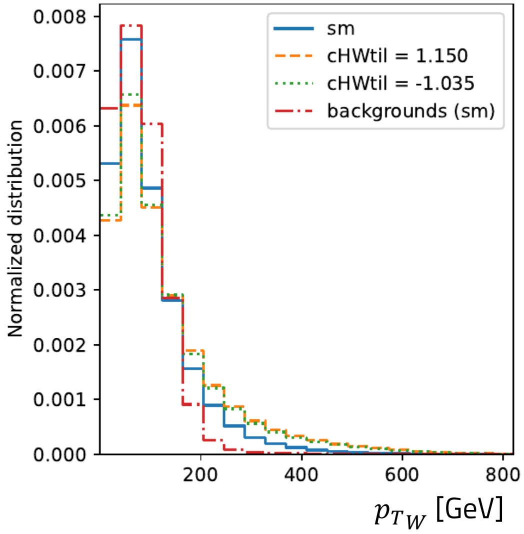
- Neutrino energy/ $E_T^{miss}$: $\sigma_E$ =12.5 GeV
- b-quark energies: $\frac{\sigma_E}{E}$ = 0.1

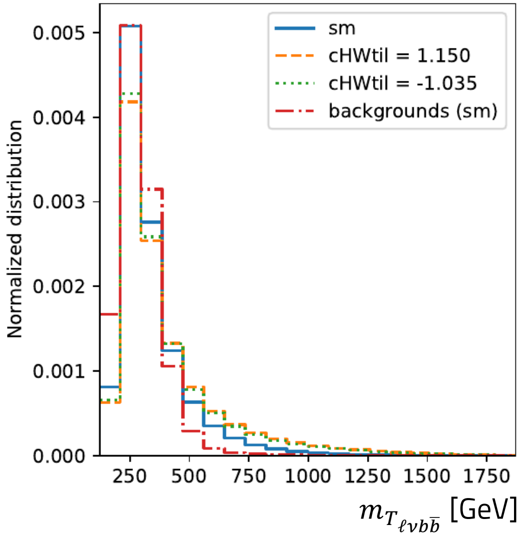
No systematics applied.




[2]: J. Brehmer et al, *Benchmarking simplified template cross-sections in WH production*, arXiv:1908.06980 (figure from paper)

# Energy-related observables



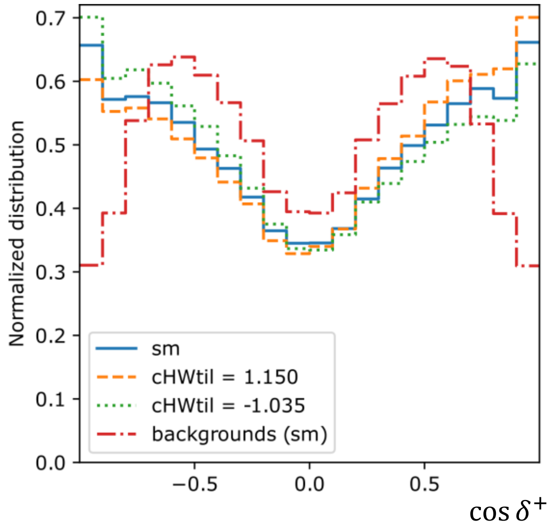



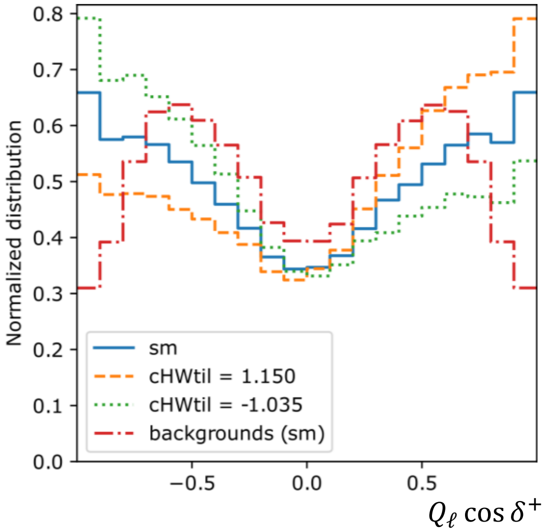

- Sensitivity to non-zero $c_{H\widetilde{W}}$ - S/B increased in high $p_{T_W}$ and $m_{T_{tot}}$ regions w.r.t. SM
- Not sensitive to sign of $c_{H\widetilde{W}}$ - changes in observable come mainly from EFT$^2$ terms

$$\cos\delta^+ = \frac{\vec{p}_\ell^{(W)} . (\vec{p}_H \times \vec{p}_W)}{|\vec{p}_\ell^{(W)}||\vec{p}_H \times \vec{p}_W|} \text{ [3]}$$

# Angular observables

$\vec{p}_\ell^{(W)}$: momentum of lepton in W boson rest frame




- Symmetric for SM signal and backgrounds, asymmetric for $c_{H\widetilde{W}} \neq 0$
- Can extract sign of $c_{H\widetilde{W}}$, weighting by lepton charge increases asymmetry


[3]: R. Godbole et al, *"Jet substructure and probes of CP violation in Vh production"*, arXiv:1409.5449

# SALLY training

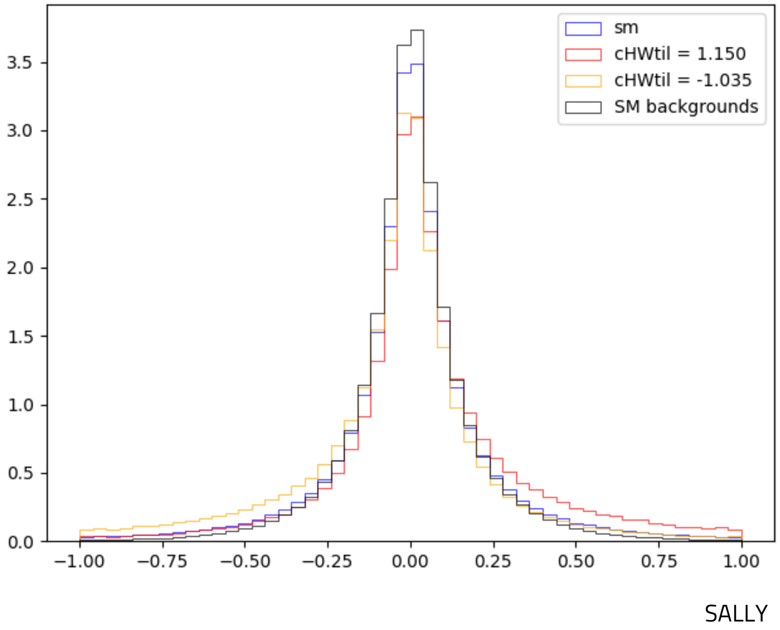
SALLY: ensemble of 5 NNs, 1 hidden layer, 50 epochs, early stopping applied

Basic training input variables (48):

- 4-vector of two b-quarks, lepton
- ${E_T^{miss}}_x, {E_T^{miss}}_y, |E_T^{miss}|$
- $p_T, \theta, \eta, \phi$ of two b-quarks and Higgs
- $m_{bb}$
- $p_{T_W}, \phi_W$
- $\Delta\phi_{bb,b_1\ell,b_2\ell}, \Delta R_{bb,b_1\ell,b_2\ell}$
- $\Delta\phi_{b_1 E_T^{miss},b_2 E_T^{miss},\ell E_T^{miss}}$
- $m_{T_{\ell\nu}}, m_{T_{\ell\nu bb}}$ $(m_{T_{tot}})$

[3]: R. Godbole et al, *"Jet substructure and probes of CP violation in Vh production"*, arXiv:1409.5449

# SALLY observable




Left: distribution of SALLY trained at $c_{H\widetilde{W}} = 0$

Symmetric for SM signal and backgrounds

- asymmetric for $c_{H\widetilde{W}} \neq 0$, can extract sign of $c_{H\widetilde{W}}$ (asymmetry)

# Fisher Information and (linearized) limits

Ranked different observables using Fisher Information at $c_{H\widetilde{W}} = 0$

- Extracting limits with **Local Fisher distance -** likelihood ratio linearized in $c_{H\widetilde{W}}$

| **Observable** | $c_{H\widetilde{W}}$ **S+B 95% CL (L= 300 fb$^{-1}$)** |
|---|---|
| 1D: $p_{T_W}$ | [-1.62,1.62] |
| 2D: $p_{T_W} \times m_{T_{\ell\nu b\bar{b}}}$ | [-1.4,1.4] |
| 1D: $\mathrm{Q}_\ell \cos\delta^+$ | [-0.227,0.227] |
| 2D: $p_{T_W} \times \mathrm{Q}_\ell \cos\delta^+$ | [-0.088, 0.088] |
| MVA: SALLY, 48 input variables | [-0.067, 0.067] |
| MVA: SALLY, 48 input variables + $p_{z_\nu}$, $\mathrm{Q}_\ell \cos\delta^+$, $\mathrm{Q}_\ell \cos\delta^-$, $\cos\theta^*$ | [-0.062, 0.062] |

(Linearized) limits with SALLY tighter than with $\mathrm{Q}_\ell \cos\delta^+$ (factor 3)

- Tighter than with combination of $p_{T_W}$ and $\mathrm{Q}_\ell \cos\delta^+$ (~25%)

# Full limits

Determined expected limits w/ full likelihood ratio **(shape-only)**

- Properly takes into account the effect of terms $\propto c_{\widetilde{HW}}^2$ in the likelihood ratio

| **Observable** | $c_{\widetilde{HW}}$ **S+B 95% CL (L= 300 fb$^{-1}$)** |
|---|---|
| 1D: $p_{T_W}$ | [-0.192, 0.216] |
| 2D: $p_{T_W} \times m_{T_{\ell\nu b\bar{b}}}$ | [-0.36, 0.384] |
| 1D: $\mathrm{Q}_\ell \cos\delta^+$ | [-0.264, 0.216] |
| 2D: $p_{T_W} \times \mathrm{Q}_\ell \cos\delta^+$ | [-0.096, 0.072] |
| MVA: SALLY, 48 input variables | [-0.144, 0.12] |
| MVA: SALLY, 48 input variables + $p_{z_\nu}$, $\mathrm{Q}_\ell \cos\delta^+$, $\mathrm{Q}_\ell \cos\delta^-$, $\cos\theta^*$ | [-0.168, 0.096] |

2D combination of $p_{T_W}$ and $\mathrm{Q}_\ell \cos\delta^+$ yields the best limits

- SALLY no longer optimal when quadratic effects included

# Conclusions

Goal: optimize search for CP violation in the HWW interaction via WH production ($c_{H\widetilde{W}}$).

Studied method to estimate **detector-level optimal observable** (SALLY).

Compared expected 95% CL limits on ($c_{H\widetilde{W}}$) obtained with SALLY observable vs. others.

- SALLY observable more sensitive to linear term than angular observable alone.
- Overall more stringent limits with 2D histogram of $p_{T_W}$ and $\mathrm{Q}_\ell \cos\delta^+$.
    - SALLY observable only optimal when linear effects dominate.

Future work: introduce systematics, dominant CP-even operators in WH.

# Backup

# Optimal observables

The **optimal observable** around a reference parameter point $\theta_{ref}$ is given by $\nabla_\theta \log p(x|\theta)|_{\theta_{ref}}$ [1]

- not calculable at detector-level ($x$), calculable at parton-level ($z_p$)

$$\nabla_\theta \log p(z_p|\theta) \propto \nabla_\theta \log \frac{|\mathcal{M}(z_p|\theta)|^2}{\sigma(\theta)} = \frac{\nabla_\theta |\mathcal{M}(z_p|\theta)|^2}{|\mathcal{M}(z_p|\theta)|^2} - \frac{\nabla_\theta \sigma(\theta)}{\sigma(\theta)}$$

CP-odd Optimal Observables ≡ parton-level Optimal Observable around $\theta = 0$ for SM+interference

- Neglect everything between parton- and detector-level ($z_p = x$)

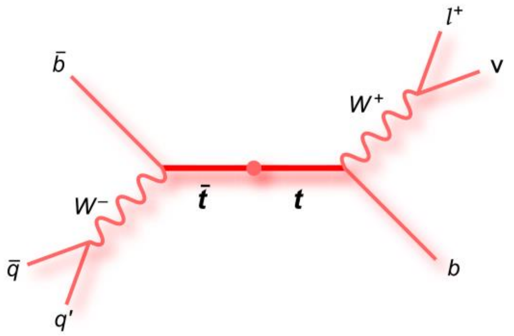
**Can we build a observable optimally sensitive to $c_{H\widetilde{W}}$ using standard detector-level variables ?**

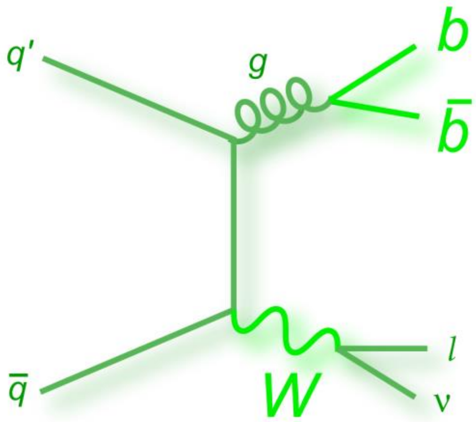
**Can we go without the need to fully reconstruct the neutrino 4-vector ?**

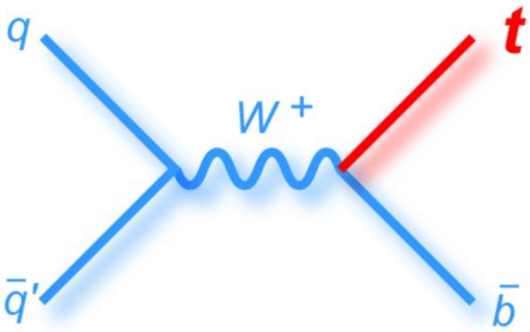
[1]: J. Brehmer et al, *MadMiner: Machine learning–based inference for particle physics*, arXiv:1907.10621

# Backgrounds

Main backgrounds: semileptonic $t\bar{t}$, W+(b)-jets, s-channel single top



# Energy-dependent observable binning

Binning for the limits with kinematic observables

- $p_{T_W} \in [0-75, 75-150, 150-250, 250-\infty]$ GeV
- $p_{T_W} \in [0-75, 75-150, 150-250, 250-400, 400-\infty]$ GeV
- $p_{T_W} \in [0-75, 75-150, 150-250, 250-400, 400-600, 600-\infty]$ GeV
- $p_{T_W} \in [0-75, 75-150, 150-250, 250-400, 400-\infty]$ GeV $\otimes$
  $m_{T_{tot}} \in [0-400, 400-800, 800-\infty]$ GeV
- $p_{T_W} \in [0-75, 75-150, 150-250, 250-400, 400-600, 600-\infty]$ GeV $\otimes$
  $m_{T_{tot}} \in [0-400, 400-800, 800-\infty]$ GeV

SALLY limits are done with 25 equally spaced bins between -1.0 and 1.0.

# Neutrino $p_z$ reconstruction

The neutrino longitudinal momentum, $p_{z_\nu}$ is necessary to calculate angular observables

- Identify $\vec{p}_{T_\nu} \equiv \vec{E}_T^{miss}$ and solve the equation $p_{W_\mu} p_W^\mu = m_W^2$

Quadratic equation leading to two solutions, neglect imaginary parts

- Studied different methods to select the solution

Selecting the solution that has $\min. |\beta_z^W - \beta_z^H|, \beta_z = p_z / \sqrt{p_z^2 + m^2}$

- Minimum of $\Delta R$(parton-level W, reconstructed W)

# Angular observable binning

Binning for the limits with angular and angular+kinematic observables:

- $Q_\ell \cos\delta^+ \in [-1.0 - 0., 0. - 1.0]$
- $Q_\ell \cos\delta^+ \in [-1.0 - 0.5, -0.5 - 0., 0. - 0.5, 0.5 - 1.0]$
- $Q_\ell \cos\delta^+ \in [-1.0 - -2/3, -2/3 - -1/3, -1/3 - 0., 0. - 1/3, 1/3 - 2/3, 2/3 - 1.0]$
- $p_{T_W} \in [0 - 75, 75 - 150, 150 - 250, 250 - 400, 400 - \infty]$ GeV $\otimes$
  $Q_\ell \cos\delta^+ \in [-1.0 - -2/3, -2/3 - -1/3, -1/3 - 0., 0. - 1/3, 1/3 - 2/3, 2/3 - 1.0]$
- $m_{T_{\ell\nu}} \in [0 - 250, 250 - 500, 500 - \infty]$ GeV $\otimes$
  $Q_\ell \cos\delta^+ \in [-1.0 - -2/3, -2/3 - -1/3, -1/3 - 0., 0. - 1/3, 1/3 - 2/3, 2/3 - 1.0]$
- $m_{T_{\ell\nu b\bar{b}}} \in [0 - 400, 400 - 800, 800 - \infty]$ GeV $\otimes$
  $Q_\ell \cos\delta^+ \in [-1.0 - -2/3, -2/3 - -1/3, -1/3 - 0., 0. - 1/3, 1/3 - 2/3, 2/3 - 1.0]$

# The likelihood

The likelihood, $p_{full}$, is the central statistical object in any physics analysis

$$p_{\text{full}} = \text{Pois}(n|\mathcal{L}\sigma(\theta)) \prod_i p(x_i|\theta)$$

The kinematic likelihood, $p(x|\theta)$, can be factorized

- $z_p$: parton-level variables, $z_s$: parton-shower+hadronization variables, $z_d$: detector variables, $x$: reconstructed observables

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \, p(x|z_d)\, p(z_d|z_s)\, p(z_s|z_p)\, p(z_p|\theta)$$

**Can't be calculated analitically** ($z_d$ alone can have >1M variables for Geant 4)

# Morphing

Using morphing to interpolate event weights and distributions from a limited set of benchmarks

- In our case, obtained from MG reweighting of SM sample

$$p(z|\theta) = \sum_c w_c(\theta)\, p(z|\theta_c)$$

MadMiner chooses the optimal benchmark points such that the $\sum w_c^2$ is the minimum

- Avoid numerical instabilities

Gradient of weights and cross-sections can also be derived from the morphing matrix

# Fisher Information

The Fisher Information matrix $I_{ij}(\theta)$ quantifies the sensitivity of a measurement

$$I_{ij}(\theta) \equiv -E\left[\frac{\partial^2 \log p_{\text{full}}(x|\theta)}{\partial\theta_i \partial\theta_j}\right]|\theta$$

- Can be used to benchmark observables $v$ by using $p_{full}(v|\theta)$
- Its differential distribution $dI_{ij}(x,\theta)/dv$ allows defining optimal phase space cuts

We can use the score to extract the full detector-level information:

$$I_{ij}(\theta) = \frac{\mathcal{L}}{\sigma}\frac{\partial\sigma}{\partial\theta_i}\frac{\partial\sigma}{\partial\theta_j} + \frac{\mathcal{L}\sigma}{N}\sum_{x\sim p(x|\theta)} t_i(x)t_j(x)$$

# Local Fisher distance and limits

Fisher Information matrix shows up in Taylor expansion of the log-likelihood ratio

$$-2\,\mathbb{E}\left[\log\frac{p_{\text{full}}(x|\theta)}{p_{\text{full}}(x|0)}\right] = \underbrace{-\mathbb{E}\left[\frac{\partial^2 \log p_{\text{full}}(x|\theta)}{\partial\theta_i\,\partial\theta_j}\right]}_{\equiv I_{ij}}\,\theta_i\,\theta_j + \mathcal{O}(\theta^3)$$

For small deviations around a reference point, $\theta_0$, one can extract limits with

**Local Fisher Distance**

$$d(\theta_1,\theta_0)^2 = I_{ij}(\theta_0)(\theta_1-\theta_0)_i(\theta_1-\theta_0)_j$$

- These are, by definition, **linearized in the parameters of interest**
- Not accurate when terms quadratic in the parameters of interest dominate

# Angular observable distributions (truth)

$$\cos\delta^{+} = \frac{p_{\ell}^{(W)} \cdot (p_H \times p_W)}{|p_{\ell}^{(W)}||p_H \times p_W|}$$

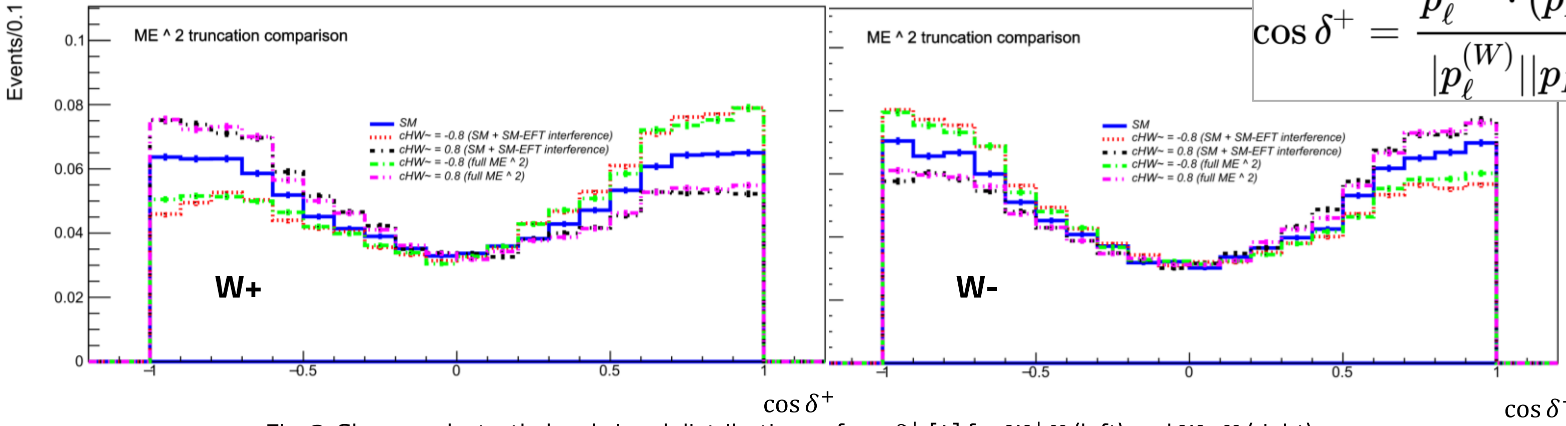



Fig. 3: Shape-only, truth-level signal distributions of $\cos\delta^{+}$ [4] for $W^{+}H$ (left) and $W^{-}H$ (right).

Sensitive to interference (CP-odd) component, (mostly) unaffected by quadratic (CP-even)

For same coupling, asymmetry has opposite signs for opposite charge W bosons

- **Weighting by lepton charge** - $Q_{\ell}\cos\delta^{+}$ - increases asymmetry and sensitivity to sign of $c_{H\widetilde{W}}$

4: arXiv:1409.5449