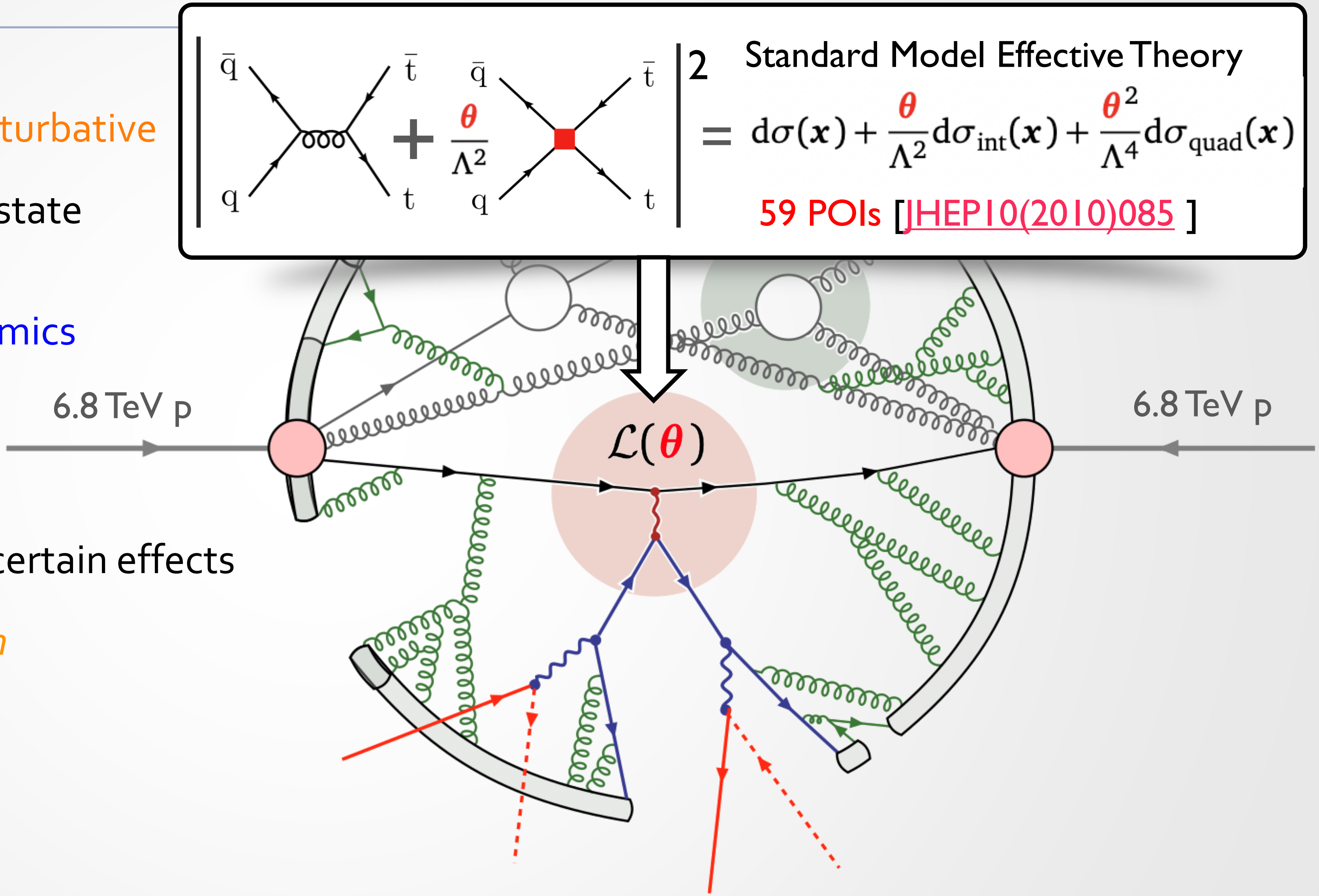
REFINABLE
~~PRECISION~~ SYSTEMATICS FOR UNBINNED HYPOTHESIS TESTS

R. Schöfbeck (HEPHY Vienna), May 16th, 2024

THE LHC AS A MICROSCOPE

- At low energies, QCD is non-perturbative
- LHC elevates the proton bound state to the perturbative regime, exposing the constituents' dynamics
  - Access to POIs - **θ**
- After the hard scatter, many uncertain effects
- Model in *staged event simulation*



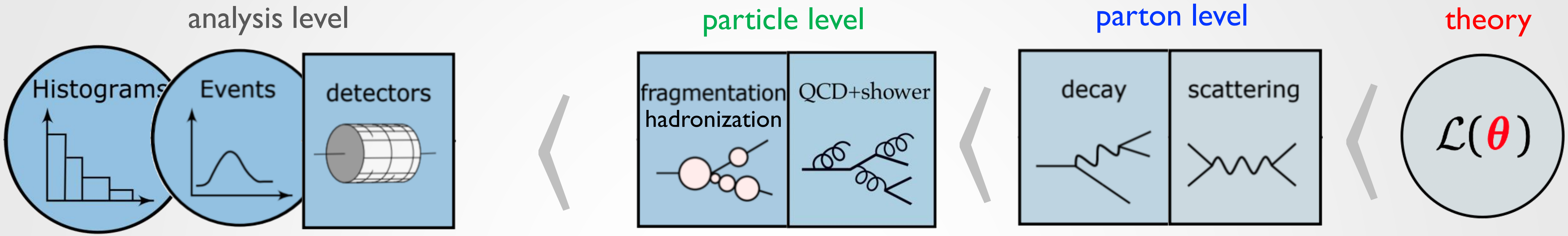
Standard Model Effective Theory

$$= d\sigma(\boldsymbol{x}) + \frac{\boldsymbol{\theta}}{\Lambda^2} d\sigma_{\text{int}}(\boldsymbol{x}) + \frac{\boldsymbol{\theta}^2}{\Lambda^4} d\sigma_{\text{quad}}(\boldsymbol{x})$$

59 POIs [JHEP10(2010)085]

A CONDITIONAL SEQUENCE

adapted from arXiv:2211.01421



$$p(x_{\rm det}|\theta) = \int dz_{\rm ptl} \int dz_{\rm p}[\cdots]\ p(x_{\rm det}|z_{\rm ptl})$$

$$p(z_{\rm ptl}|z_{\rm p})$$

$$p(z_{\rm p}|\theta)$$

Likelihood ratio test statistic

$$\mathrm{LR}(x_{\rm det}|\mathrm{H}_1,\mathrm{H}_2) \equiv \frac{p(x_{\rm det}|\mathrm{H}_1=\boldsymbol{\theta},\boldsymbol{\nu})}{p(x_{\rm det}|\mathrm{H}_2=\mathrm{SM},\boldsymbol{\nu}=0)}$$

Use (un)binned parametrizations in ***θ*, *ν*.**

1. Generators run in 'forward mode'
2. More uncertainties $p(z_{\rm ptl}|z_{\rm p},\nu_{\rm th.})$ $p(x_{\rm det}|z_{\rm ptl},\nu_{\rm exp.})$

$$\frac{1}{\sigma_{\theta,\nu_{\rm th.}}}\frac{d\sigma_{\theta,\nu_{\rm th.}}}{d\boldsymbol{z}_p} = p(\boldsymbol{z}_p|\theta,\nu_{\rm th.})$$

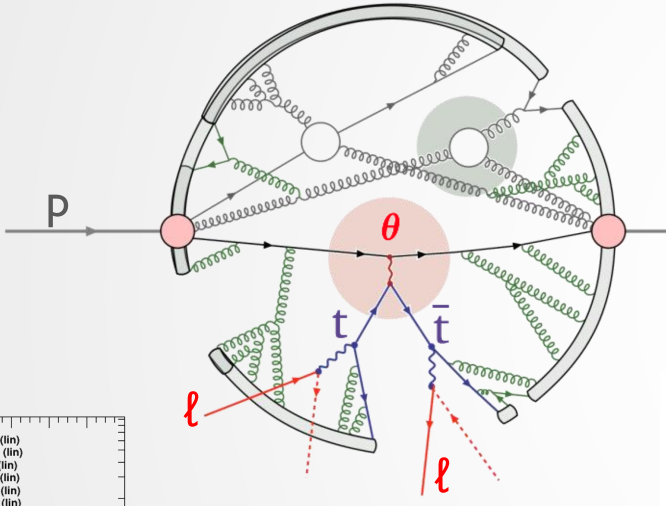
parton-level
differential cross section
$\boldsymbol{\nu}_{\rm th}$ … scale unc., PDF

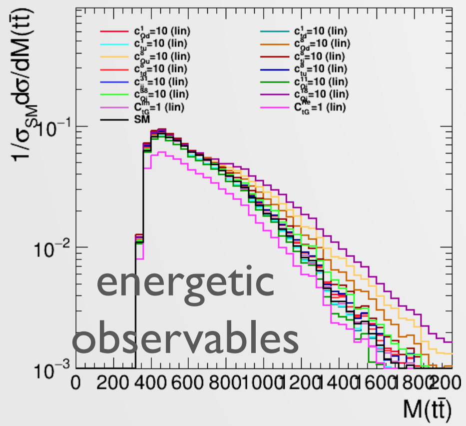
$\cancel{p(\theta)}$

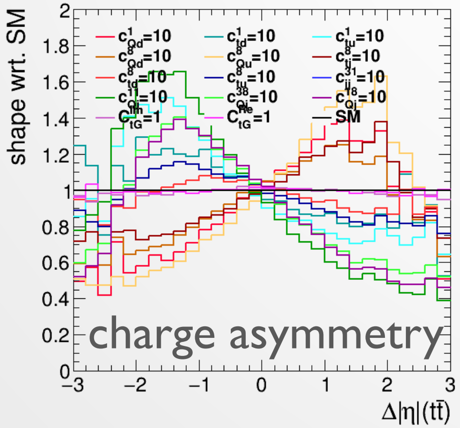
θ NOT stochastic; Frequentist

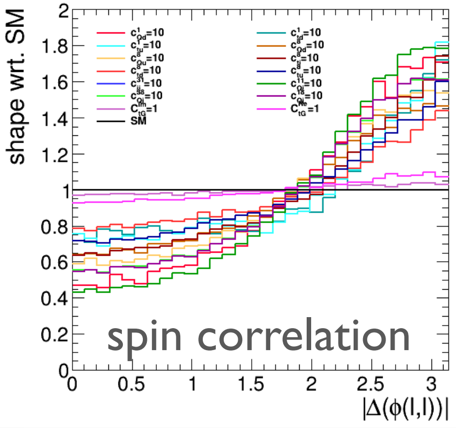
TOP QUARK PAIRS IN THE 2ℓ CHANNEL

- Top quark pair production with 2ℓ:
  - Clean probes of new physics in a messy environment



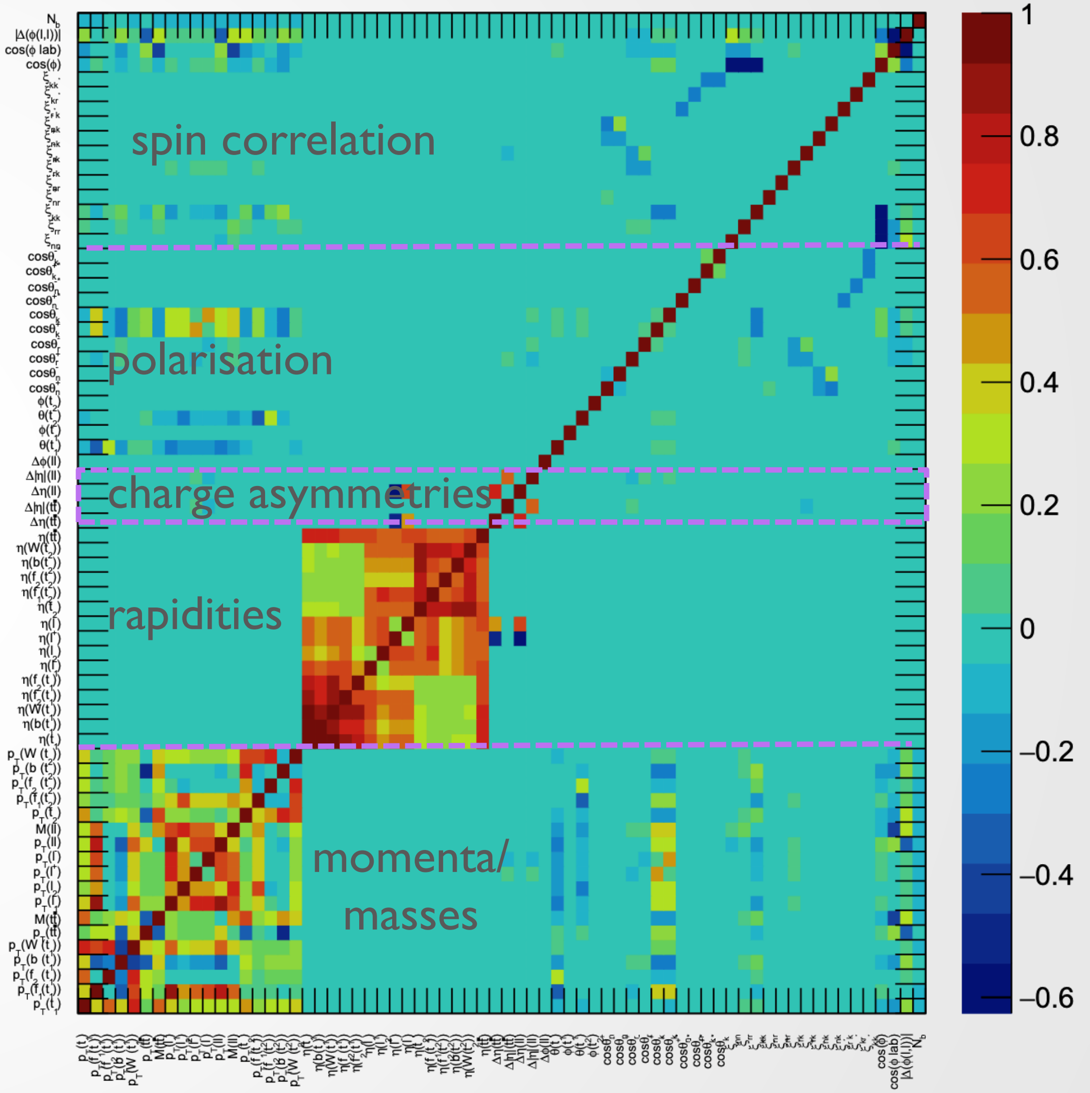






≳ 72 features

≳ 15 SMEFT POIs



linear feature correlation in tt(2ℓ)

Typically use only 1 or 2 features!

HARD-SCATTER MODELING

Madminer [1506.02169] [1805.00013] [1805.00020] [1805.12244] [1907.10621] [1908.06980] [2109.10414]

1. Analytic predictions for the SMEFT predictions at the parton level - easily recalcuable

$$\mathrm{d}\sigma_{\mathrm{SMEFT}}(\boldsymbol{z}_p|\boldsymbol{\theta}, \nu_R, \nu_F, \boldsymbol{\nu}_{\mathrm{PDF}}) \propto \sum_{f_1,f_2} \left|\mathcal{M}_{\mathrm{SMEFT}}(\boldsymbol{z}_p|\boldsymbol{\theta}, \mu_R(\nu_R), \mu_F(\nu_F))\right|^2$$
$$\times \, \mathrm{PDF}(f_1, x_{\mathrm{Bjorken},1}, \mu_F(\nu_F), \boldsymbol{\nu}_{\mathrm{PDF}})\mathrm{PDF}(f_2, x_{\mathrm{Bjorken},2}, \mu_F(\nu_F), \boldsymbol{\nu}_{\mathrm{PDF}})\mathrm{d}\boldsymbol{z}_p$$

2. Phenomena at lower energy scales largely factorize:

→ Conditional propabilities factor out [*Madminer*, full Refs. in backup]

→ Access to the 'joint likelihood' ratio for POIs and some systematic effects.

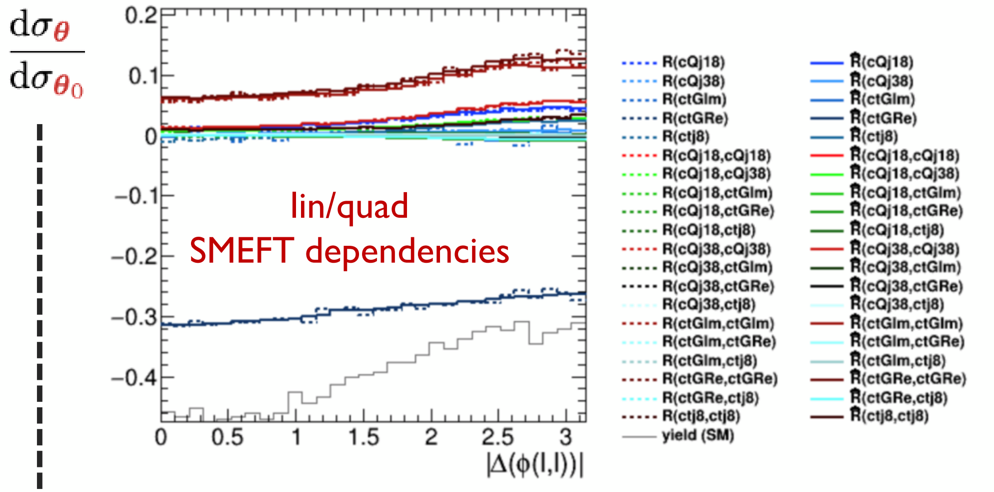
$$r(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{\nu}) = \frac{\sigma(\boldsymbol{\theta}, \boldsymbol{\nu})}{\sigma(\mathrm{SM})} \frac{p(\boldsymbol{x}_i, \boldsymbol{z}_{\mathrm{reco},i}, \boldsymbol{z}_{\mathrm{ptl},i}, \boldsymbol{z}_{\mathrm{p},i}|\boldsymbol{\theta}, \boldsymbol{\nu})}{p(\boldsymbol{x}_i, \boldsymbol{z}_{\mathrm{reco},i}, \boldsymbol{z}_{\mathrm{ptl},i}, \boldsymbol{z}_{\mathrm{p},i}|\mathrm{SM})} = \frac{\sigma(\boldsymbol{\theta}, \boldsymbol{\nu})}{\sigma(\mathrm{SM})} \frac{p(\boldsymbol{x}|\boldsymbol{z}_{\mathrm{reco}})}{p(\boldsymbol{x}|\boldsymbol{z}_{\mathrm{reco}})} \frac{p(\boldsymbol{z}_{\mathrm{reco}}|\boldsymbol{z}_{\mathrm{ptl}})}{p(\boldsymbol{z}_{\mathrm{reco}}|\boldsymbol{z}_{\mathrm{ptl}})} \frac{p(\boldsymbol{z}_{\mathrm{ptl}}|\boldsymbol{z}_{\mathrm{p}})}{p(\boldsymbol{z}_{\mathrm{ptl}}|\boldsymbol{z}_{\mathrm{p}})} \frac{p(\boldsymbol{z}_{\mathrm{p},i}|\boldsymbol{\theta}, \boldsymbol{\nu})}{p(\boldsymbol{z}_{\mathrm{p},i}|\mathrm{SM})} \sim \frac{\left|\mathcal{M}(\boldsymbol{z}_{\mathrm{p},i}, \boldsymbol{\theta}, \boldsymbol{\nu})\right|^2}{\left|\mathcal{M}(\boldsymbol{z}_{\mathrm{p},i}, \mathrm{SM})\right|^2}$$

LEARNING PARAMETER DEPENDENCIES

$$L = \left\langle \left(r(x_{\mathrm{det}}, z_{\mathrm{ptl}}, \cdots, z_{\mathrm{p}}|\theta) - \hat{f}_{\theta}(x_{\mathrm{det}}) \right)^{2} \right\rangle_{\mathrm{SM}}$$

Only nominal simulation

- ML fit of SMEFT dependence: A solved problem
  [Madminer / P. Classifiers / ML4EFT / Boosted Information Tree → Refs in backup]



SMEFT dependence: $\mathbb{R}^{72} \longrightarrow \mathbb{R}^{136}$
(learned with Boosted Information Tree)

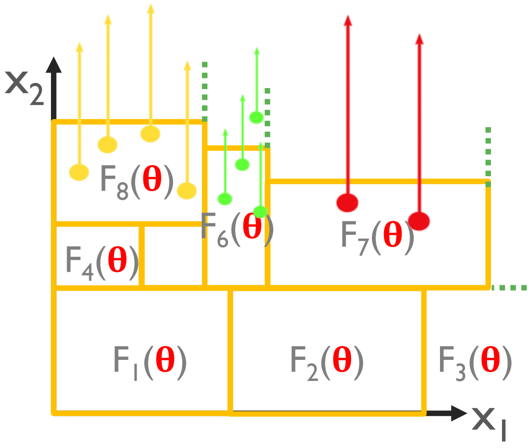
- Can now compute parametrized (unbinned) test statistic

$$q_{\boldsymbol{\theta},\boldsymbol{\nu}}(\mathcal{D}) \equiv \log \frac{L(\mathcal{D}|\boldsymbol{\theta}\ \ ,\boldsymbol{\nu}\ \ )}{L(\mathcal{D}|\boldsymbol{\theta}_0,\boldsymbol{\nu}_0)} = -\mathcal{L}(\boldsymbol{\nu})\sigma(\boldsymbol{\theta},\boldsymbol{\nu}) + \mathcal{L}(\boldsymbol{\nu}_0)\sigma(\boldsymbol{\theta}_0,\boldsymbol{\nu}_0) + \sum_{i=1}^{N(\mathcal{D})} \log\left(\frac{\mathcal{L}(\boldsymbol{\nu}\ \ )}{\mathcal{L}(\boldsymbol{\nu}_0)} \frac{d\sigma_{\boldsymbol{\theta}\ \ ,\boldsymbol{\nu}}}{d\sigma_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0}}(\boldsymbol{x}_i) \right) + \sum_{k=1}^{N_{nuis.}} \log \frac{C_k(\nu_k)}{C_k(\nu_{0,k})}$$

TREE ALGORITHM FOR SMEFT LEARNING

[arXiv:2107.10859, arXiv:2205:12976]

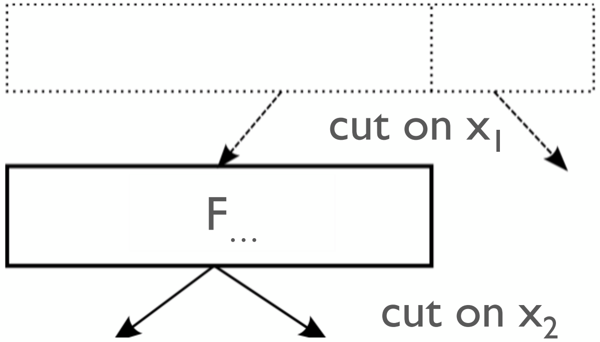
Phase-space partitioning



A simple tree

$$\hat{F}(\boldsymbol{x},\boldsymbol{\theta}) = \underbrace{\sum_{j\in\mathcal{J}} \overbrace{\mathbb{1}_j(\boldsymbol{x})}^{\text{non-linearity)}}}_{\text{phase space partitioning J}} \underbrace{F_j(\boldsymbol{\theta})}_{\text{prediction } F_j}$$

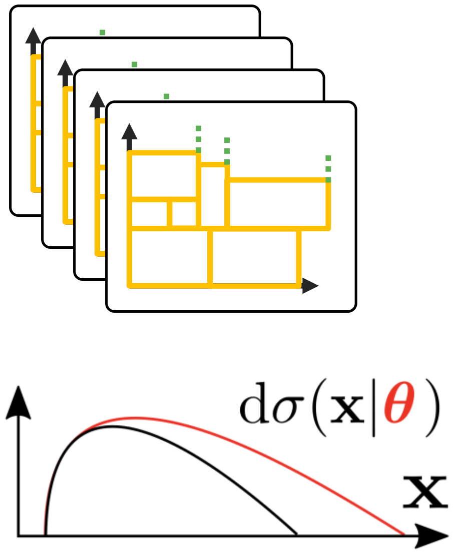
Cart algorithm



$$F_j(\boldsymbol{\theta}) = \frac{\sum_{\boldsymbol{x}_i\in j} w_j(\mathrm{SM})\, r_i(\boldsymbol{\theta})}{\sum_{\boldsymbol{x}_i\in j} w_j(\mathrm{SM})} = \frac{\int \mathrm{d}z \frac{\mathrm{d}\sigma_{\boldsymbol{\theta}}}{\mathrm{d}(x,z)}}{\int \mathrm{d}z \frac{\mathrm{d}\sigma_{\mathrm{SM}}}{\mathrm{d}(x,z)}}$$

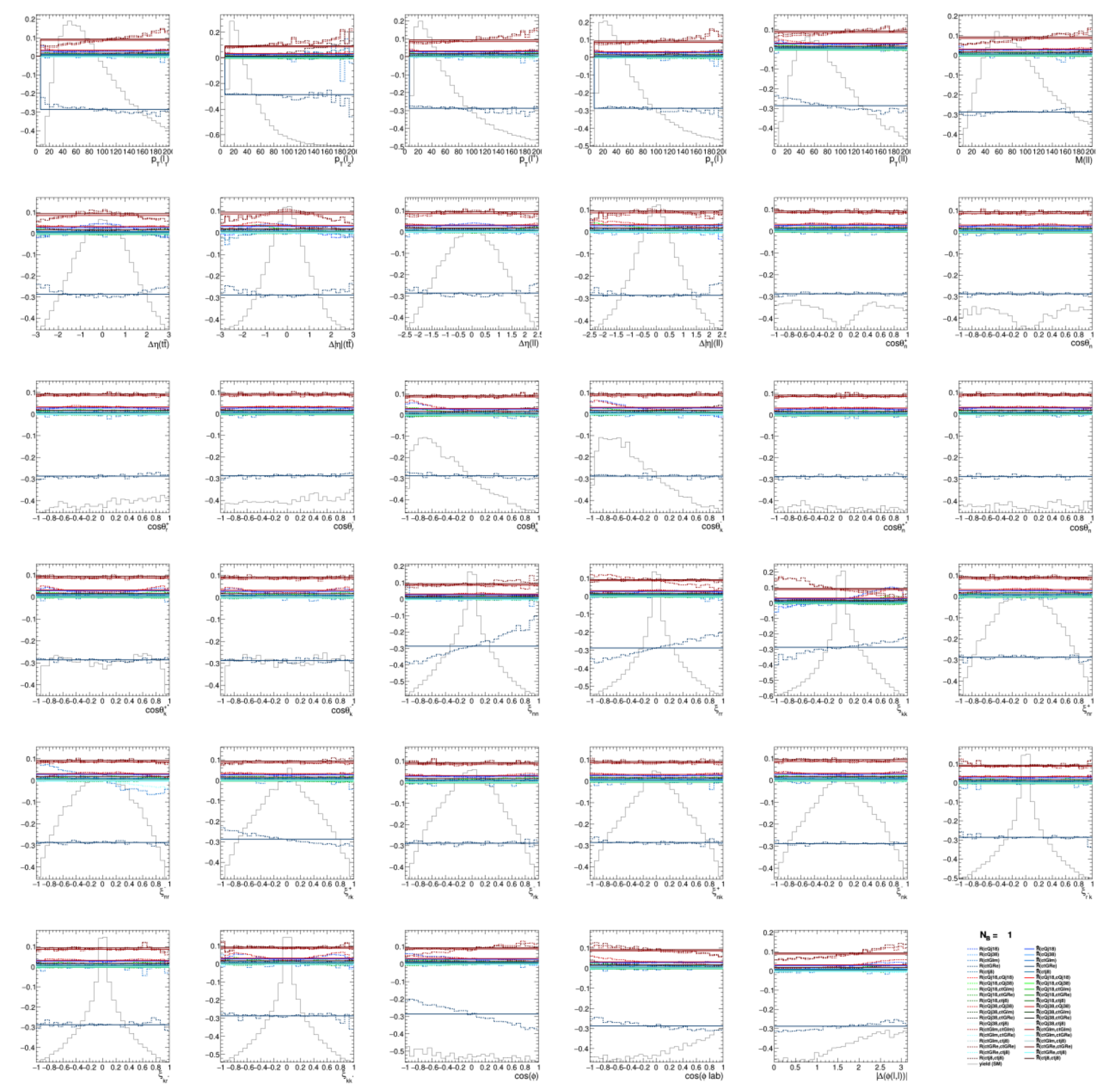
"magic" (Kyle C.)

Boosting

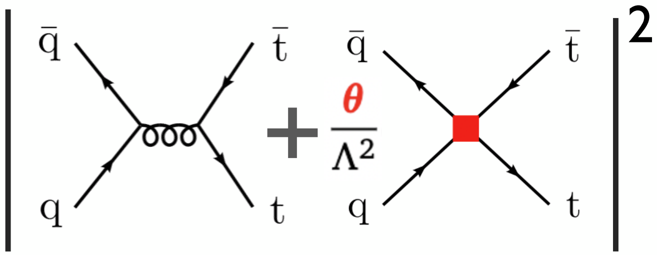


- A tree is a hierarchical phase-space partitioning
  - Boosted Information Tree: Associate each region j with a polynomial $F_j(\boldsymbol{\theta})$
  - The non-linearity is in the change across node positions
  - Fitting tree: Optimize "node split positions" on some loss. Can compute $F_j(\boldsymbol{\theta})$ from events in node.
- Boosting elevates tree to an arbitrarily expressive regressor for $\mathrm{d}\sigma(\mathbf{x}|\boldsymbol{\theta})$ - ratios

LEARNING SMEFT IN TTBAR



- 35 features top quark pairs (2ℓ)
- “Boosted Information Tree (BIT)”
  - NN are equivalent
  - 5 POIs, 20 functions simultaneously learned
  - 300 trees, D=5, ~9 hrs of training
  - also more realistic study, including backgrounds [2107.10859], [2205.12976]
- Learning coefficient functions to compute parametrized optimal oberables



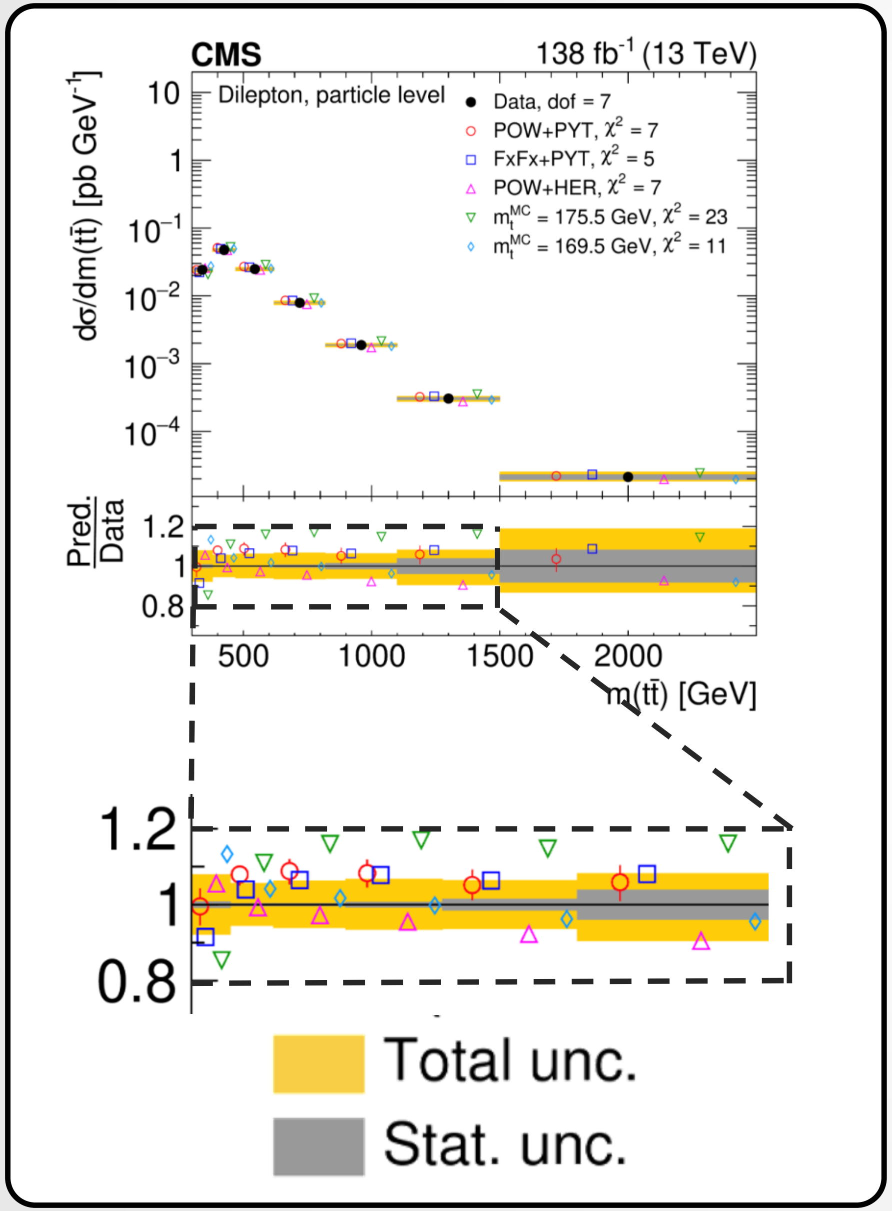
$$=$$

$$d\sigma(\boldsymbol{x}) + \frac{\boldsymbol{\theta}}{\Lambda^2} d\sigma_{\text{int}}(\boldsymbol{x}) + \frac{\boldsymbol{\theta}^2}{\Lambda^4} d\sigma_{\text{quad}}(\boldsymbol{x}) =: R(\boldsymbol{x}|\boldsymbol{\theta}) d\sigma(\boldsymbol{x})$$

→ parametrized $q(\boldsymbol{\theta}|\mathcal{D})$

BACK TO REALITY!

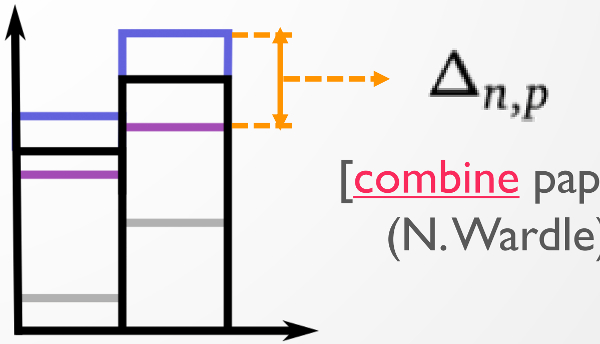
[CMS-TOP-20-006]



- Systematics dominate in many/most applications
- Binned analyses? Use additive model with exponentials

$$\text{prediction}(\boldsymbol{\theta}, \boldsymbol{\nu}) = \sum_{p=1}^{N_p} R_{n,p}(\boldsymbol{\theta}) \exp\left(\boldsymbol{\nu}^{\mathsf{T}} \Delta_{n,p,1} + \boldsymbol{\nu}^{\mathsf{T}} \Delta_{n,p,2} \boldsymbol{\nu}\right) \sigma_{n,p}(\mathrm{SM})$$

- How to find the parameters Δ?
  - "Vary simulation" ⟷ Generate synthetic datasets
  - shift JEC, scale b-tagging efficiencies, PS weights, hDamp



[combine paper]
(N.Wardle)

- Decades of experience with modeling choices

REFINABLE MODELING IN 3-STEPS

1. Let's write an unbinned *additive* model

$$\mathrm{d}\Sigma(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\nu}) = \sum_{p=1}^{N_p} \underbrace{R_p(\boldsymbol{x}|\boldsymbol{\theta})}_{\text{SMEFT}}\, \underbrace{\alpha_p^{\nu_p}}_{\substack{\text{normalisation}\\ \text{(``k-factors'')}}} \exp\underbrace{\left\{\boldsymbol{\nu}^{\intercal}\Delta_{n,p,1}(\boldsymbol{x}) + \boldsymbol{\nu}^{\intercal}\Delta_{n,p,2}(\boldsymbol{x})\boldsymbol{\nu} + \ldots\right\}}_{\text{systematics} \quad S_p(\boldsymbol{x}|\boldsymbol{\nu})}\, \mathrm{d}\sigma_p(\boldsymbol{x}|\mathrm{SM})$$

2. The experimentalist (not the framework) decides on further specification

TT(2ℓ) has 90% purity: We have a single EFT process and a number of small backgrounds (DY, non-prompt,…)

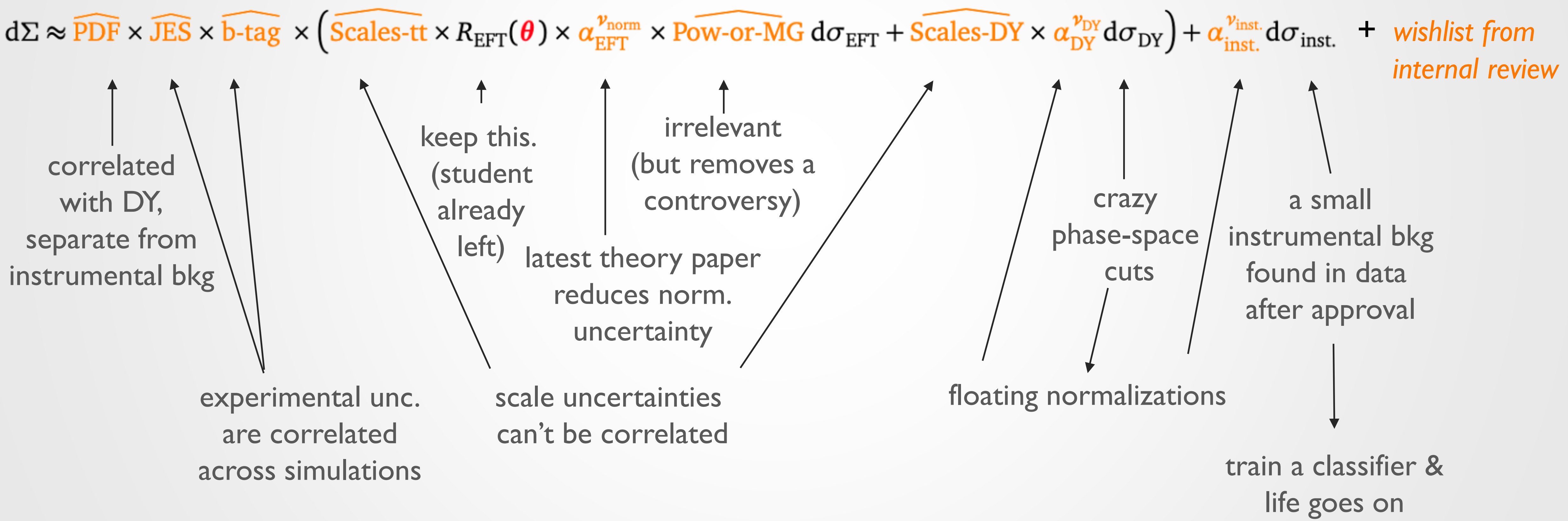
$$\mathrm{d}\Sigma(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\nu}) = R_{\mathrm{EFT}}(\boldsymbol{x}|\boldsymbol{\theta})\, S_{\mathrm{EFT}}(\boldsymbol{x}|\boldsymbol{\nu})\, \mathrm{d}\sigma_{\mathrm{EFT}}(\boldsymbol{x}|\mathrm{SM}) + \sum_{p=1}^{N_{\mathrm{bkgs}}} \alpha_p^{\nu_p} S_p(\boldsymbol{x}|\boldsymbol{\nu})\mathrm{d}\sigma_p(\boldsymbol{x}|\mathrm{SM})$$

3. Form the ratio & learn the factors!

$$\frac{\mathrm{d}\Sigma(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\nu})}{\mathrm{d}\Sigma(\boldsymbol{x}|\mathrm{SM})} = \frac{R_{\mathrm{EFT}}(\boldsymbol{x}|\boldsymbol{\theta})\, S_{\mathrm{EFT}}(\boldsymbol{x}|\boldsymbol{\nu}) + \sum_{p=1}^{N_{\mathrm{bkgs}}} \alpha_p^{\nu_p} S_p(\boldsymbol{x}|\boldsymbol{\nu}) \boxed{\frac{\mathrm{d}\sigma_p(\boldsymbol{x}|\mathrm{SM})}{\mathrm{d}\sigma_{\mathrm{EFT}}(\boldsymbol{x}|\mathrm{SM})}}}{1 + \sum_{p=1}^{N_{\mathrm{bkgs}}} \boxed{\frac{\mathrm{d}\sigma_p(\boldsymbol{x}|\mathrm{SM})}{\mathrm{d}\sigma_{\mathrm{EFT}}(\boldsymbol{x}|\mathrm{SM})}}} \approx \frac{\overset{\substack{\text{1) SMEFT learning}\\ \downarrow}}{\hat{R}_{\mathrm{EFT}}(\boldsymbol{x}|\boldsymbol{\theta})}\, \overset{\substack{\text{2) systematics learning}\\ \downarrow}}{\hat{S}_{\mathrm{EFT}}(\boldsymbol{x}|\boldsymbol{\nu})} + \sum_{p=1}^{N_{\mathrm{bkgs}}} \alpha_p^{\nu_p} \hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}) \hat{g}_p(\boldsymbol{x})}{1 + \sum_{p=1}^{N_{\mathrm{bkgs}}} \underset{\text{3) classifiers}}{\hat{g}_p(\boldsymbol{x})}}$$

Adding systematics or processes *doesn't invalidate* partial training!

REFINING THE INDUCITIVE BIAS



$$d\Sigma \approx \widehat{\text{PDF}} \times \widehat{\text{JES}} \times \widehat{\text{b-tag}} \times \left(\widehat{\text{Scales-tt}} \times R_{\text{EFT}}(\boldsymbol{\theta}) \times \alpha_{\text{EFT}}^{\nu_{\text{norm}}} \times \widehat{\text{Pow-or-MG}}\, d\sigma_{\text{EFT}} + \widehat{\text{Scales-DY}} \times \alpha_{\text{DY}}^{\nu_{\text{DY}}}\, d\sigma_{\text{DY}}\right) + \alpha_{\text{inst.}}^{\nu_{\text{inst.}}}\, d\sigma_{\text{inst.}}$$

+ *wishlist from internal review*

correlated with DY, separate from instrumental bkg

keep this. (student already left)

irrelevant (but removes a controversy)

latest theory paper reduces norm. uncertainty

crazy phase-space cuts

a small instrumental bkg found in data after approval

experimental unc. are correlated across simulations

scale uncertainties can't be correlated

floating normalizations

train a classifier & life goes on

- Stay flexible. Reduce model mis-specification.

LEARNING SYSTEMATICS


Autalops Die Antilope

lepton
efficiencies

jet energy
& resolution


b-tagging


renormalization
and factorization
scales

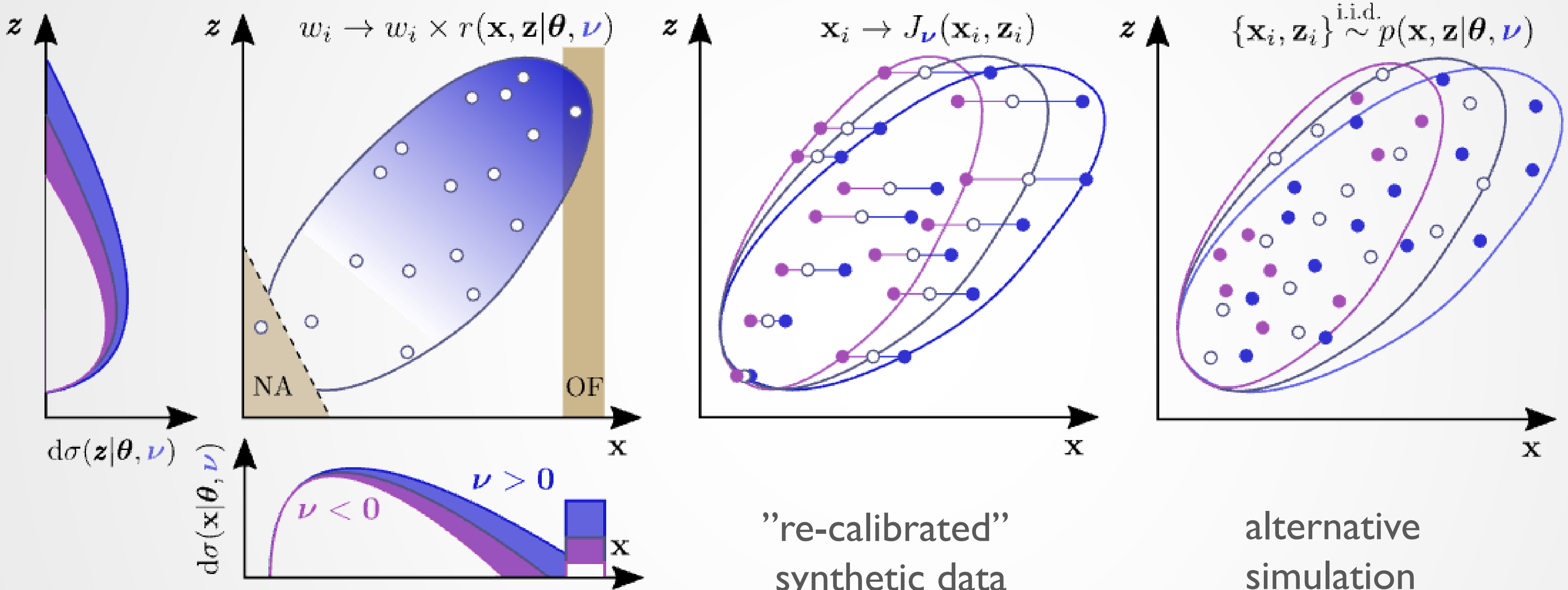
BESTIARY OF SYSTEMATIC UNCERTAINTY MODELING

Uncertainty prescription	Weights?	Calibration?	Per-object?	Parametric?	Latent dependent?	To be learned?
Trigger	☑	⊖	☑	⊖	⊖	(☑)
Pileup/MinBias	☑	⊖	⊖	⊖	⊖	(☑)
Lepton efficiency	☑	⊖	☑	⊖	⊖	(☑)
Lepton scale	⊖	☑	☑	⊖	⊖	☑
Jet energy scale	⊖	☑	☑	☑	⊖	☑
Jet energy resolution	⊖	☑	☑	☑	☑	☑
B-tagging	☑	(⊖)	☑	☑	☑	☑
tW/DY/diboson normalization	⊖	⊖	⊖	⊖	☑	☑
QCD scales	☑	⊖	⊖	☑	☑	☑
TT generator modeling	⊖	⊖	⊖	⊖	☑	☑
Top pt	☑	⊖	⊖	⊖	☑	☑
Color rec.	⊖	⊖	⊖	(⊖)	☑	☑
α_S in PS	☑	⊖	⊖	⊖	☑	☑
hdamp	⊖	⊖	⊖	⊖/☑	☑	☑
PDF	☑	⊖	⊖	(☑)	☑	☑
Luminosity						⊖

ATLAS & CMS systematic uncertainties from the differential xsec-combination

SYNTHETIC DATA SETS

- Let's view simulation as a sampling of the "joint space", spanned by latent and observed features



"reweighted" synthetic data

"re-calibrated" synthetic data

alternative simulation

- Well established procedures for "synthetic" data sets that model systematic variations
- From now on: assume we have simulated data for a limited set of nuisance parameter values $\boldsymbol{\nu}$

LEARNING PARAMETERIZATIONS

- “Likelihood ratio trick”

$$L_{\boldsymbol{\nu},\mathrm{CE}}[\hat f] = -\langle \log \hat f(\boldsymbol{x})\rangle_{\boldsymbol{x},\boldsymbol{z},\mathrm{SM}} - \langle \log(1-\hat f(\boldsymbol{x}))\rangle_{\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\nu}}$$

$$f^{*}_{\mathrm{CE}}(\boldsymbol{x}) = \frac{1}{1+\frac{\mathrm{d}\sigma(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|\mathrm{SM})}}$$

- Parametric ansatz:

$$\hat f_{\boldsymbol{\nu}}(\boldsymbol{x}) = \frac{1}{1+\exp(\hat g_{\boldsymbol{\nu}}(\boldsymbol{x}))}$$

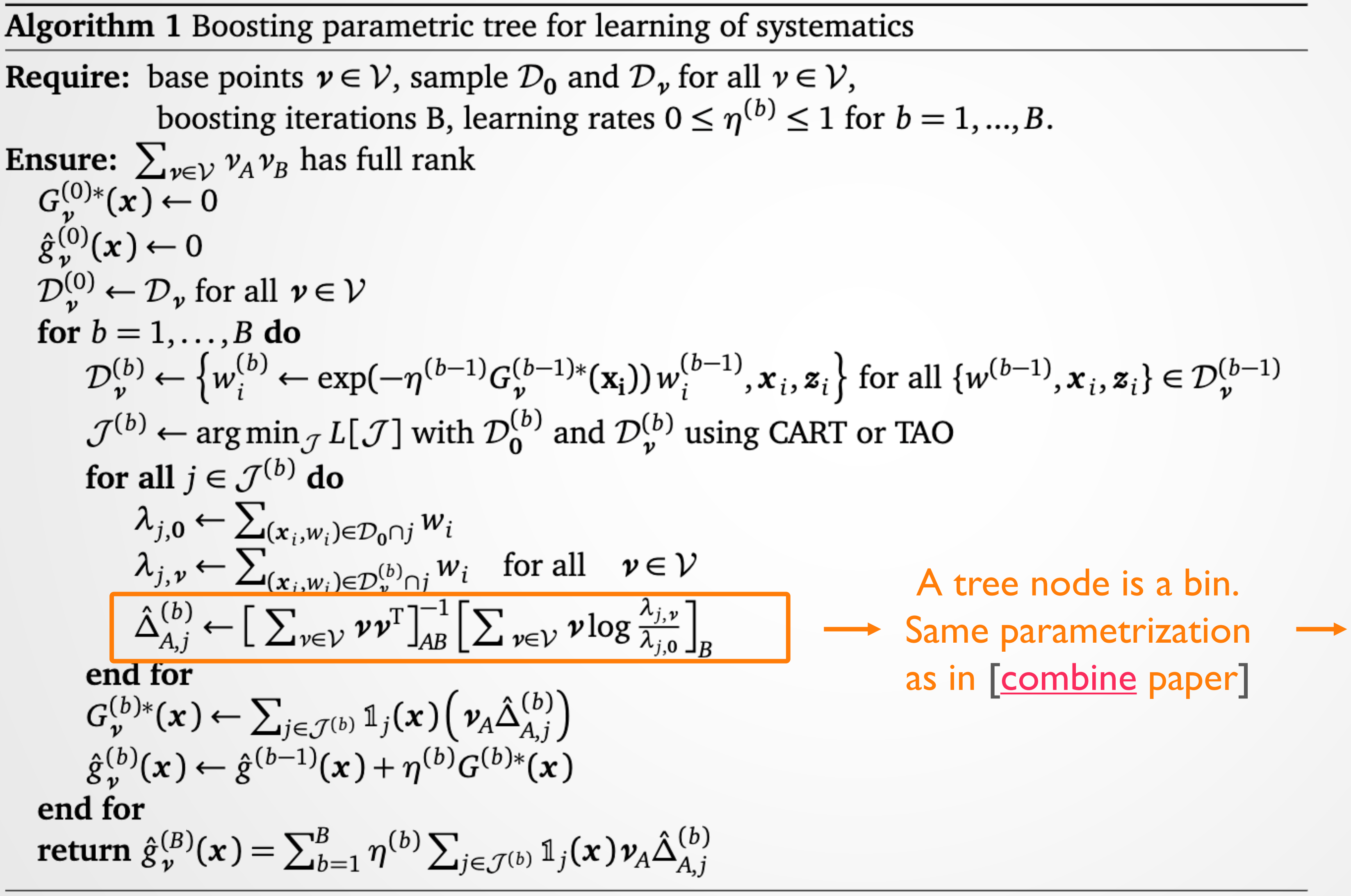
$$\hat g_{\boldsymbol{\nu}}(\boldsymbol{x};\hat{\boldsymbol{\Delta}}_1,\hat{\boldsymbol{\Delta}}_2,\ldots) = \boldsymbol{\nu}^{\mathsf{T}}\hat{\boldsymbol{\Delta}}_1(\boldsymbol{x}) + \boldsymbol{\nu}^{\mathsf{T}}\hat{\boldsymbol{\Delta}}_2(\boldsymbol{x})\boldsymbol{\nu} + \ldots$$

Implement coefficient functions Δ(x) as NNs or trees: $\hat{\boldsymbol{\Delta}}_{1,2,\ldots}(\boldsymbol{x}) = \sum_{j\in\mathcal{J}} \mathbb{1}_j(\boldsymbol{x})\hat{\boldsymbol{\Delta}}_{j;1,2,\ldots}$

- Sufficiently many synthetic data sets

$$L[\hat{\boldsymbol{\Delta}}] = \sum_{\boldsymbol{\nu}\in\mathcal{V}} L_{\boldsymbol{\nu},\mathrm{CE}}[\hat f_{\boldsymbol{\nu}}] \longrightarrow \exp(g^{*}_{\boldsymbol{\nu}}(\boldsymbol{x})) \approx \frac{\mathrm{d}\sigma(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|\mathrm{SM})}$$

WHY DOES NOBODY LIKE TREE BOOSTING?

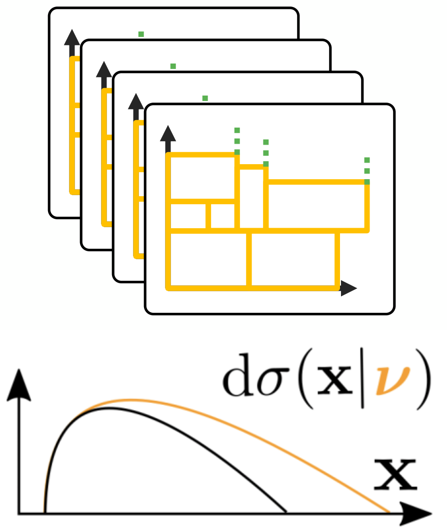


Algorithm 1 Boosting parametric tree for learning of systematics

```
Require: base points $\boldsymbol{\nu} \in \mathcal{V}$, sample $\mathcal{D}_0$ and $\mathcal{D}_{\boldsymbol{\nu}}$ for all $\boldsymbol{\nu} \in \mathcal{V}$,
         boosting iterations B, learning rates $0 \le \eta^{(b)} \le 1$ for $b = 1, ..., B$.
Ensure: $\sum_{\boldsymbol{\nu}\in\mathcal{V}} \nu_A \nu_B$ has full rank
  $G_{\boldsymbol{\nu}}^{(0)*}(\boldsymbol{x}) \leftarrow 0$
  $\hat{g}_{\boldsymbol{\nu}}^{(0)}(\boldsymbol{x}) \leftarrow 0$
  $\mathcal{D}_{\boldsymbol{\nu}}^{(0)} \leftarrow \mathcal{D}_{\boldsymbol{\nu}}$ for all $\boldsymbol{\nu} \in \mathcal{V}$
  for $b = 1, \ldots, B$ do
      $\mathcal{D}_{\boldsymbol{\nu}}^{(b)} \leftarrow \left\{ w_i^{(b)} \leftarrow \exp(-\eta^{(b-1)} G_{\boldsymbol{\nu}}^{(b-1)*}(\mathbf{x_i}))\, w_i^{(b-1)}, \boldsymbol{x}_i, \boldsymbol{z}_i \right\}$ for all $\{w^{(b-1)}, \boldsymbol{x}_i, \boldsymbol{z}_i\} \in \mathcal{D}_{\boldsymbol{\nu}}^{(b-1)}$
      $\mathcal{J}^{(b)} \leftarrow \arg\min_{\mathcal{J}} L[\mathcal{J}]$ with $\mathcal{D}_{\mathbf{0}}^{(b)}$ and $\mathcal{D}_{\boldsymbol{\nu}}^{(b)}$ using CART or TAO
      for all $j \in \mathcal{J}^{(b)}$ do
          $\lambda_{j,0} \leftarrow \sum_{(\boldsymbol{x}_i, w_i) \in \mathcal{D}_0 \cap j} w_i$
          $\lambda_{j,\boldsymbol{\nu}} \leftarrow \sum_{(\boldsymbol{x}_i, w_i) \in \mathcal{D}_{\boldsymbol{\nu}}^{(b)} \cap j} w_i$   for all   $\boldsymbol{\nu} \in \mathcal{V}$
          $\boxed{\hat{\Delta}_{A,j}^{(b)} \leftarrow \left[ \sum_{\boldsymbol{\nu}\in\mathcal{V}} \boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{T}} \right]_{AB}^{-1} \left[ \sum_{\boldsymbol{\nu}\in\mathcal{V}} \boldsymbol{\nu} \log \frac{\lambda_{j,\boldsymbol{\nu}}}{\lambda_{j,0}} \right]_B}$
      end for
      $G_{\boldsymbol{\nu}}^{(b)*}(\boldsymbol{x}) \leftarrow \sum_{j\in\mathcal{J}^{(b)}} \mathbb{1}_j(\boldsymbol{x}) \left( \nu_A \hat{\Delta}_{A,j}^{(b)} \right)$
      $\hat{g}_{\boldsymbol{\nu}}^{(b)}(\boldsymbol{x}) \leftarrow \hat{g}^{(b-1)}(\boldsymbol{x}) + \eta^{(b)} G^{(b)*}(\boldsymbol{x})$
  end for
  return $\hat{g}_{\boldsymbol{\nu}}^{(B)}(\boldsymbol{x}) = \sum_{b=1}^{B} \eta^{(b)} \sum_{j\in\mathcal{J}^{(b)}} \mathbb{1}_j(\boldsymbol{x})\, \nu_A \hat{\Delta}_{A,j}^{(b)}$
```

→ A tree node is a bin.
Same parametrization
as in [combine paper] →

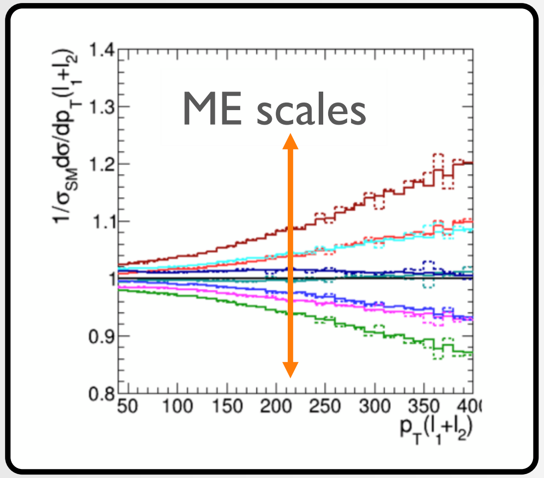
Boosted *Parametric* Tree

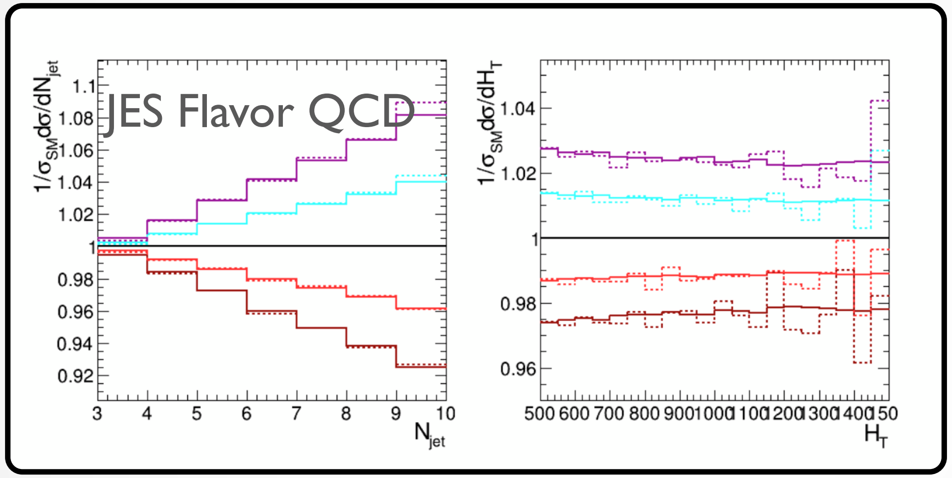


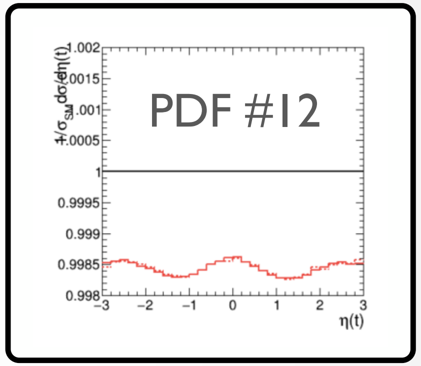
implements
x-dependent model
of cross section ratio

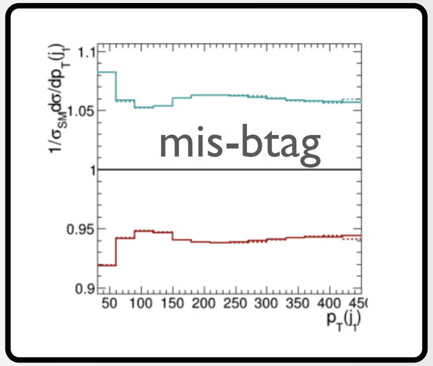
EXAMPLES!

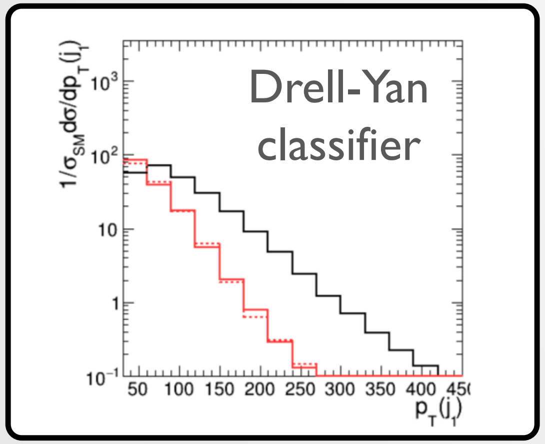
- 1D projections for a few values of $\boldsymbol{\nu}$. Available parametrically as $g_{\nu}(\mathbf{x})$.
- Semi-realistic toy study (Delphes) with publicly available information

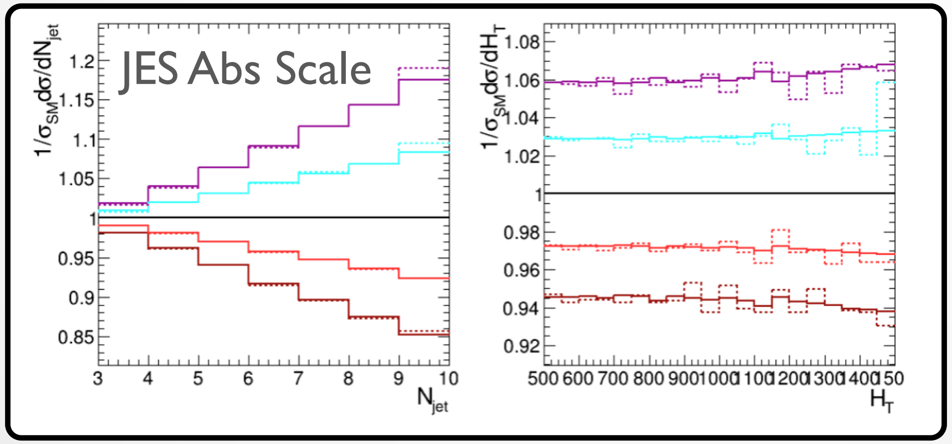


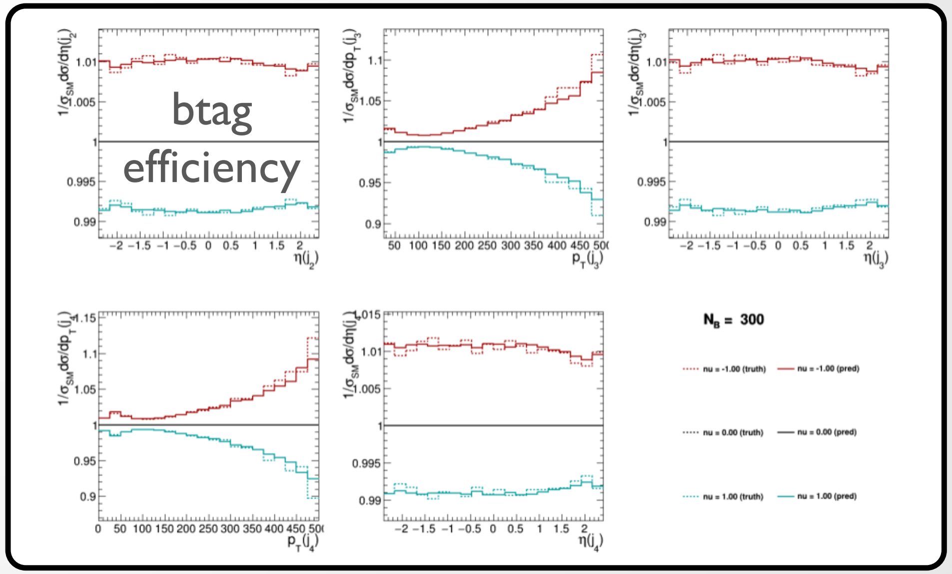










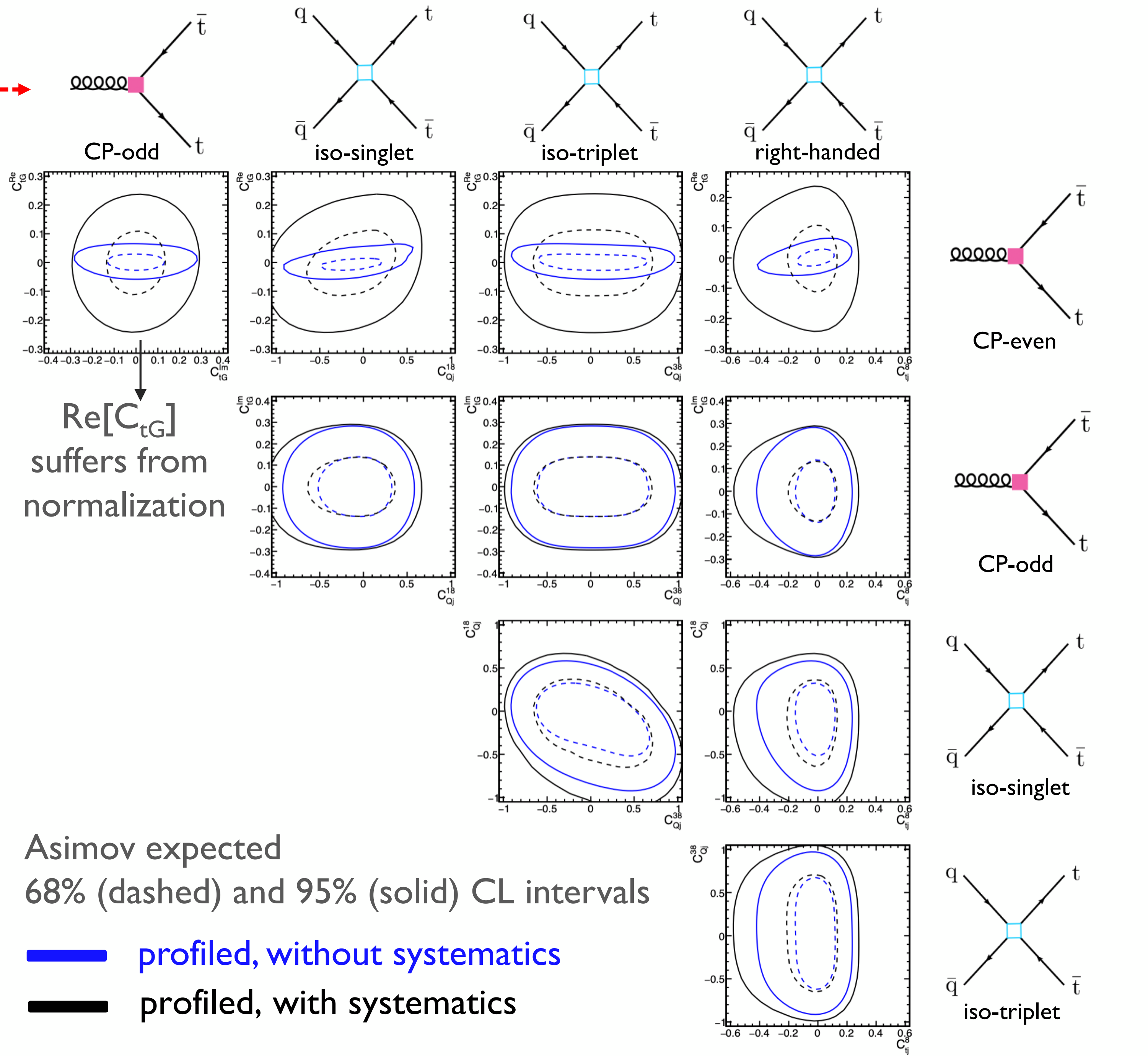


RESULTS

- TT(2ℓ) + DY background

$$d\Sigma(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\nu}) = \boxed{S_{\text{common}}(\boldsymbol{x}|\boldsymbol{\nu})}\Big(R_{t\bar{t}(2\ell)}(\boldsymbol{x}|\boldsymbol{\theta}) + \boxed{S_{\text{DY}}(\boldsymbol{x}|\boldsymbol{\nu})}\,\alpha_{\text{DY}}^{\nu_{\text{DY}}}\,g_{\text{DY}}(\boldsymbol{x})\Big)d\sigma_{t\bar{t}(2\ell)}(\boldsymbol{x}|\text{SM})$$

Uncertainty	Nuisances	size
Lepton efficiency	1	<1%
Jet energy scale	6	~2-15%
Jet energy resolution	1	~2-3%
B-tagging	5	1-10%
DY normalization	1	floating
TT(2ℓ) normalization	1	20%
Powheg vs. MG	1	~5%
α_S in PS	2	<1%
ME scales	2	2-20%
PDF	100	< 0.1%
Luminosity	1	5%



Re[C_{tG}] suffers from normalization

Asimov expected
68% (dashed) and 95% (solid) CL intervals

▬ profiled, without systematics

▬ profiled, with systematics

SUMMARY

- We're getting ready for *systematically dominated* unbinned analyses.

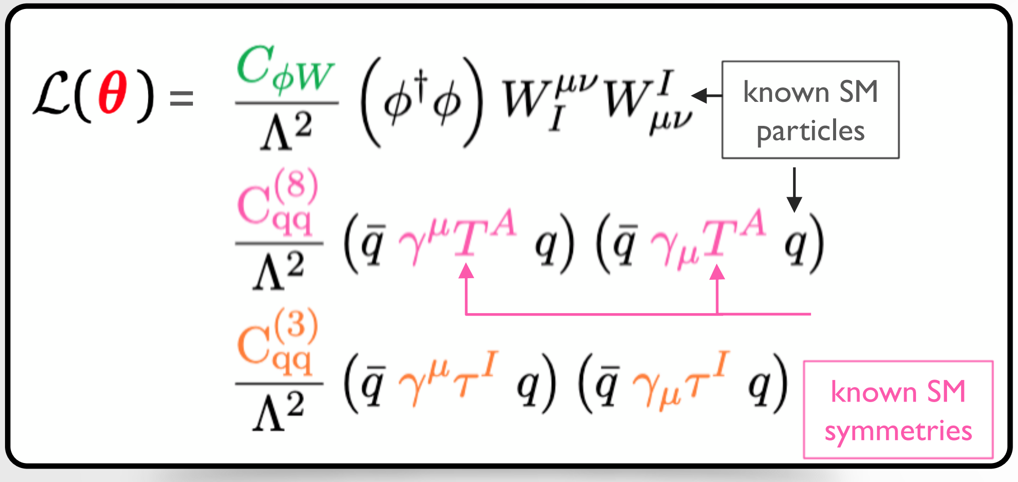
- SMEFT learning is "done" – many options.

- Trees are great.
  - Latent-space integration at the level of the terminal node, large training speed-up
  - "Boosted Information Tree" – learn quadratic SMEFT polynomial [arXiv:2107.10859, arXiv:2205:12976]
  - "Boosted Parametric Tree" – a fully paramertric tree-based regressor for uncertainty-learning

- Doing physics is even better.
  - Retain our decades of modeling-experience for constructing incrementally improvable likelihood(-ratios).

WHAT PARAMETERS ARE OF INTEREST?

- No signals beyond SM so far
  - Parametrize possible deviations

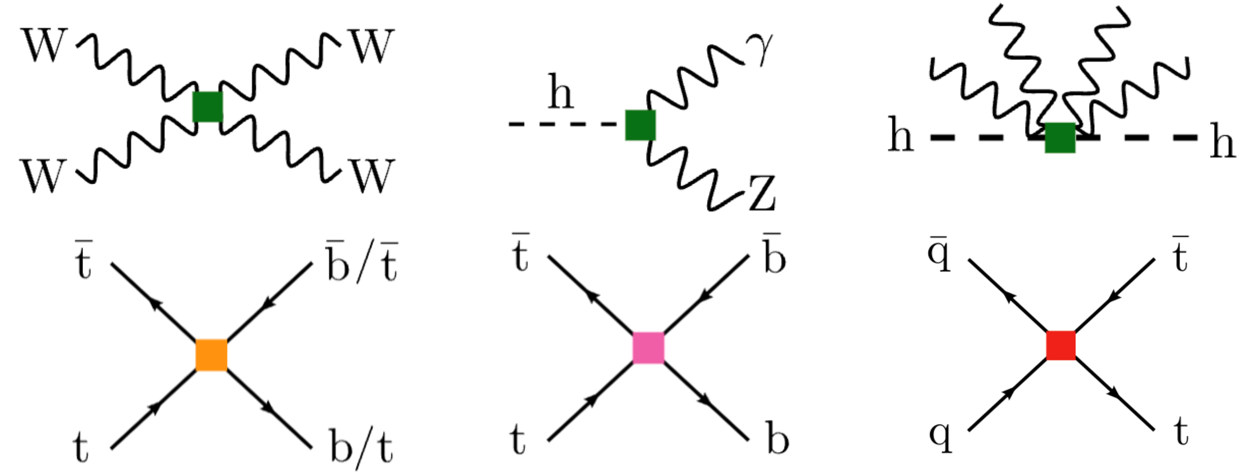


$$\mathcal{L}(\boldsymbol{\theta}) = \frac{C_{\phi W}}{\Lambda^2}\left(\phi^\dagger \phi\right) W_I^{\mu\nu} W^I_{\mu\nu}$$
$$\frac{C^{(8)}_{qq}}{\Lambda^2}\left(\bar q\, \gamma^\mu T^A\, q\right)\left(\bar q\, \gamma_\mu T^A\, q\right)$$
$$\frac{C^{(3)}_{qq}}{\Lambda^2}\left(\bar q\, \gamma^\mu \tau^I\, q\right)\left(\bar q\, \gamma_\mu \tau^I\, q\right)$$

known SM particles

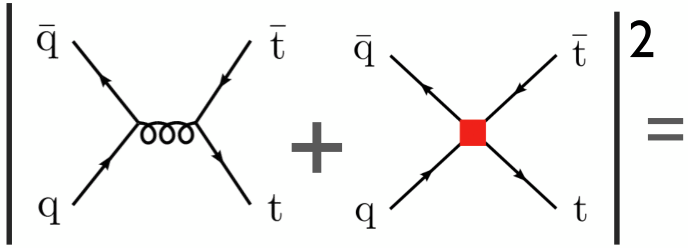
known SM symmetries

Anomalous couplings & new interactions (tiny selection!)



- "Standard model effective field theory"
  - 59 operators at d=6 [JHEP10(2010)085]
  - Modifications affect all processes

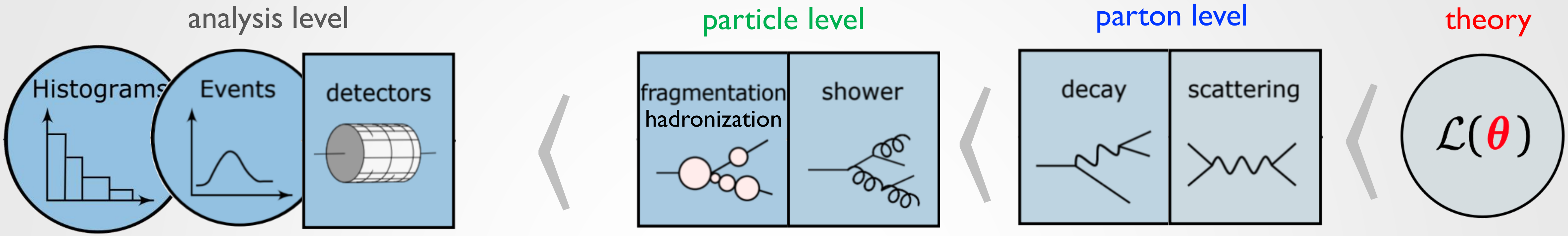
- We predict rates from "squared" diagrams:



$$= d\sigma(\boldsymbol{x}) + \frac{\boldsymbol{\theta}}{\Lambda^2} d\sigma_{\text{int}}(\boldsymbol{x}) + \frac{\boldsymbol{\theta}^2}{\Lambda^2} d\sigma_{\text{quad}}(\boldsymbol{x})$$

A CONDITIONAL SEQUENCE

adapted from arXiv:2211.01421



$$p(x_{\rm det}|\theta) = \int {\rm d}z_{\rm ptl} \int {\rm d}z_{\rm p} [\cdots]\; p(x_{\rm det}|z_{\rm ptl})$$

$$p(z_{\rm ptl}|z_{\rm p})$$

$$p(z_{\rm p}|\theta)$$

Likelihood ratio is the optimal statistic (Neyman-Pearson Lemma)

$$\mathrm{LR}(x_{\rm det}|\mathrm{H}_1,\mathrm{H}_2) \equiv \frac{p(x_{\rm det}|\mathrm{H}_1 = \boldsymbol{\theta}, \boldsymbol{\nu})}{p(x_{\rm det}|\mathrm{H}_2 = \mathrm{SM}, \boldsymbol{\nu} = 0)}$$

Use (un)binned parametrizations in $\boldsymbol{\theta}$, $\boldsymbol{\nu}$.

1. Generators run in ‘forward mode’
2. More uncertainties
   $p(z_{\rm ptl}|z_{\rm p}, \nu_{\rm th.})$
   $p(x_{\rm det}|z_{\rm ptl}, \nu_{\rm exp.})$

$$\frac{1}{\sigma_{\theta,\nu_{\rm th.}}} \frac{{\rm d}\sigma_{\theta,\nu_{\rm th.}}}{{\rm d}\boldsymbol{z}_p} = p(\boldsymbol{z}_p|\theta, \nu_{\rm th.})$$

parton-level differential cross section

$\boldsymbol{\nu}_{\rm th}$ … scale unc., PDF

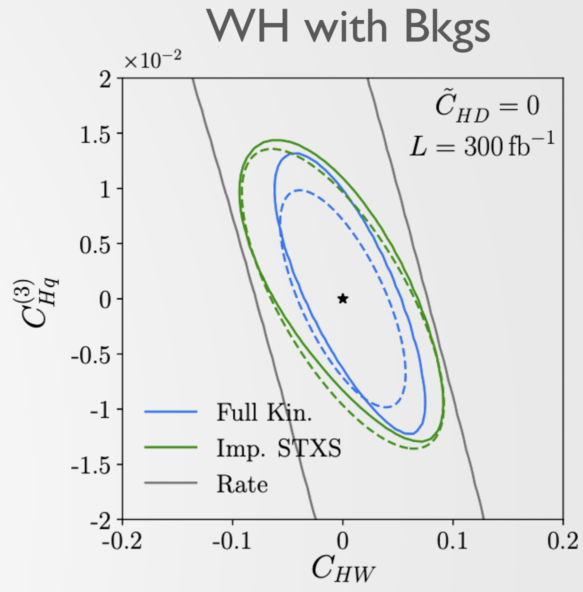
$$\cancel{p(\theta)}$$

θ NOT stochastic; Frequentist

REFERENCES

- Madminer: Neural networks based likelihood-free inference & related techniques
  - K. Cranmer , J. Pavez , and G. Louppe [1506.02169]
    J. Brehmer, K. Cranmer, G. Louppe, J. Pavez [1805.00013] [1805.00020] [1805.12244]
    J. Brehmer, F. Kling, I. Espejo, K. Cranmer [1907.10621]
  - J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn [1908.06980]
  - A. Butter, T. Plehn, N. Soybelman, J. Brehmer [2109.10414]
  - established many of the *main ideas* & *statistical interpretation* in various *NN applications*
- Weight derivative regression (A.Valassi) [2003.12853]
- Parametrized classifiers for SM-EFT: NN with quadratic structure
  - S. Chen, A. Glioti, G. Panico, A. Wulzer [*JHEP 05 (2021) 247*]
- Boosted Information Trees: Tree algorithms & boosting
  - S. Chatterjee, S. Rohshap, N. Frohner, R.S., D. Schwarz [2107.10859], [2205.12976]
- ML4EFT R. Ambrosio, J. Hoeve, M. Madigan, J. Rojo, V. Sanz [2211.02058]

- All approaches are "SMEFT-specific ML" with differences mostly on the practical side



my practical experience

→ talk later today

PARAMETRIC TREES FOR SMEFT

[arXiv:2107.10859, arXiv:2205:12976]

Want to regress in r, exploiting its the polynomial θ dependence

$$r(x, z|\theta) = \frac{\mathrm{d}\sigma(\boldsymbol{x},\boldsymbol{\theta})/\mathrm{d}\boldsymbol{x}}{\mathrm{d}\sigma(\boldsymbol{x},\mathrm{SM})/\mathrm{d}\boldsymbol{x}}$$

→ will allow to compute the optimal LLR test statistic q($\mathcal{D}$)

$$L = \sum_{\boldsymbol{\theta}\in\mathcal{B}} \int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x},\boldsymbol{z}|\mathrm{SM}) \left(r(x,z|\theta) - \hat{F}(\boldsymbol{x},\boldsymbol{\theta}) \right)^2$$

Tree ansatz

$$\hat{F}(\boldsymbol{x},\boldsymbol{\theta}) = \underbrace{\sum_{j\in\mathcal{J}}}_{\text{find optimal partitioning}} \mathbb{1}_j(\boldsymbol{x}) \underbrace{F_j(\boldsymbol{\theta})}_{\text{find optimal predictor}}$$

$F_j(\boldsymbol{\theta})$ polynomial with const. coeff. (per node)

Eliminate the predictive function

$$F_j(\boldsymbol{\theta}) = \frac{\sum_{i\in j} w_i(\boldsymbol{\theta})}{\sum_{i\in j} w_i(\boldsymbol{\theta}_0)} = \frac{\int \mathrm{d}z \frac{\mathrm{d}\sigma_{\boldsymbol{\theta}}}{\mathrm{d}(x,z)}}{\int \mathrm{d}z \frac{\mathrm{d}\sigma_{\mathrm{SM}}}{\mathrm{d}(x,z)}}$$

NP!

The latent space integration happens at the node-level and removes learnable parameters

Solve for optimal partitioning with greedy CART algorithm

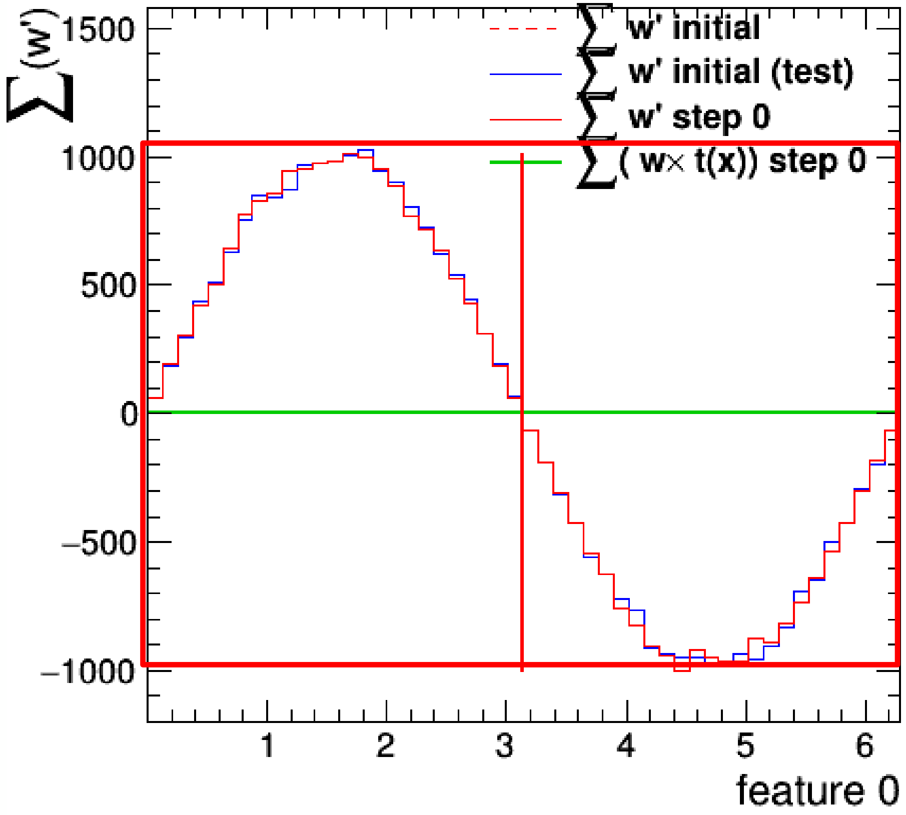
$$L = -\sum_{\boldsymbol{\theta}\in\mathcal{B}} \sum_{j\in\mathcal{J}} \frac{w_j^2(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)} = -\sum_{j\in\mathcal{J}} \sum_{\boldsymbol{\theta}\in\mathcal{B}} \theta^a \theta^b I_{ab}^{(j)} + \mathcal{O}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^3$$

We're optimizing the Fisher information!

We'll find an optimized tree.
→ boost

1D TOY EXAMPLE

- $\mathrm{pdf}(x|\boldsymbol{\theta}) = 1/(2\boldsymbol{\pi}) - \boldsymbol{\theta}\,\mathrm{Sin}(x)$
- We predict the first derivative of the pdf wrt. to the parameter



GIF animation not showing in pdf