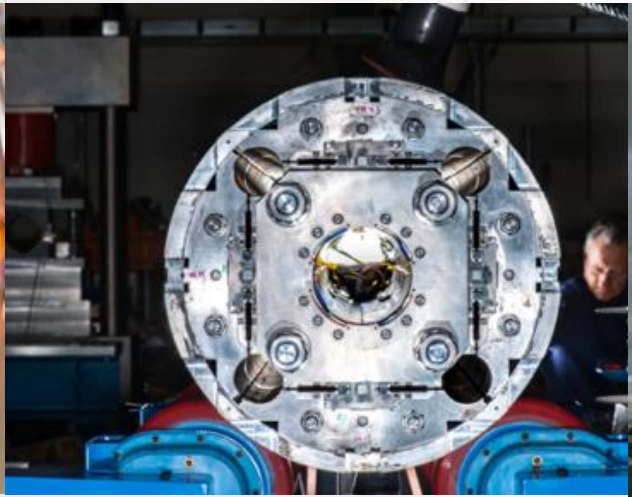
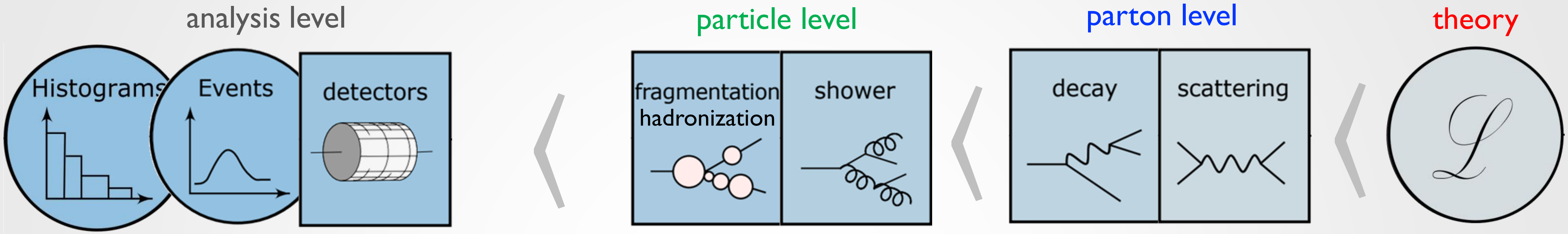
METHODOLOGY FOR SM-EFT SEARCHES IN TTBAR

R. Schöfbeck (HEPHY Vienna, FNAL), Feb 2nd, 2024, TOP workshop



A CONDITIONAL SEQUENCE

adapted from arXiv:2211.01421



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

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

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

We use parametrizations in **θ**, **ν**.
Unbinned analysis: machine-learned

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

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

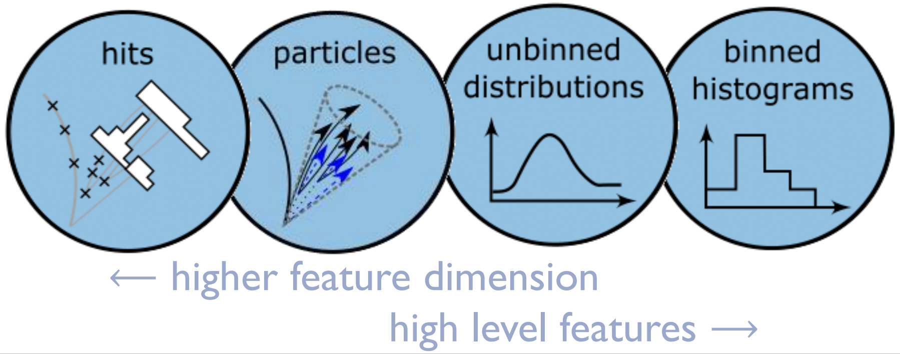
parton-level differential cross section; including decay
$\boldsymbol{\nu}_{\rm th}$ … scale unc.

$\cancel{p(\theta)}$

θ NOT stochastic; Frequentist

QUESTIONS, QUESTIONS, ...

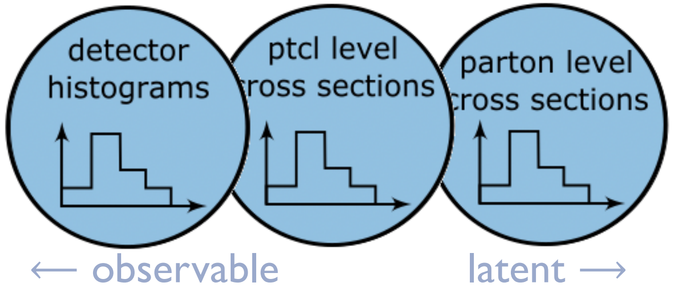
1. How accurately/low-level should the data be represented?



Generic answer: Stop when *the level is sufficient for* θ: $p(x_{\text{lower}}|x_{\text{higher}}, \cancel{\theta})$

Particle-level SMEFT [exception]

2. Take an intermediate step towards "latent" fiducial regions/gen-level?



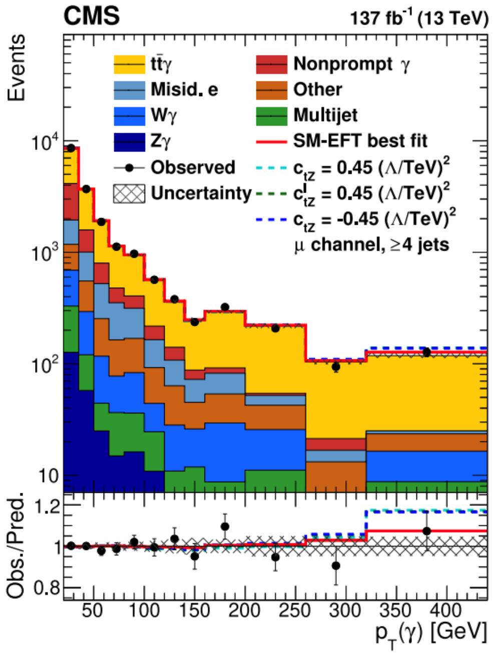
No best answer; make sure systematics modeling allows *combinations*

3. Where to stop, exactly? Publish full likelihood, a Gaussian approximation, CL contours, etc.

DETECTOR LEVEL ANALYSES

- ttγ differential

  TOP-18-010/TOP-21-004

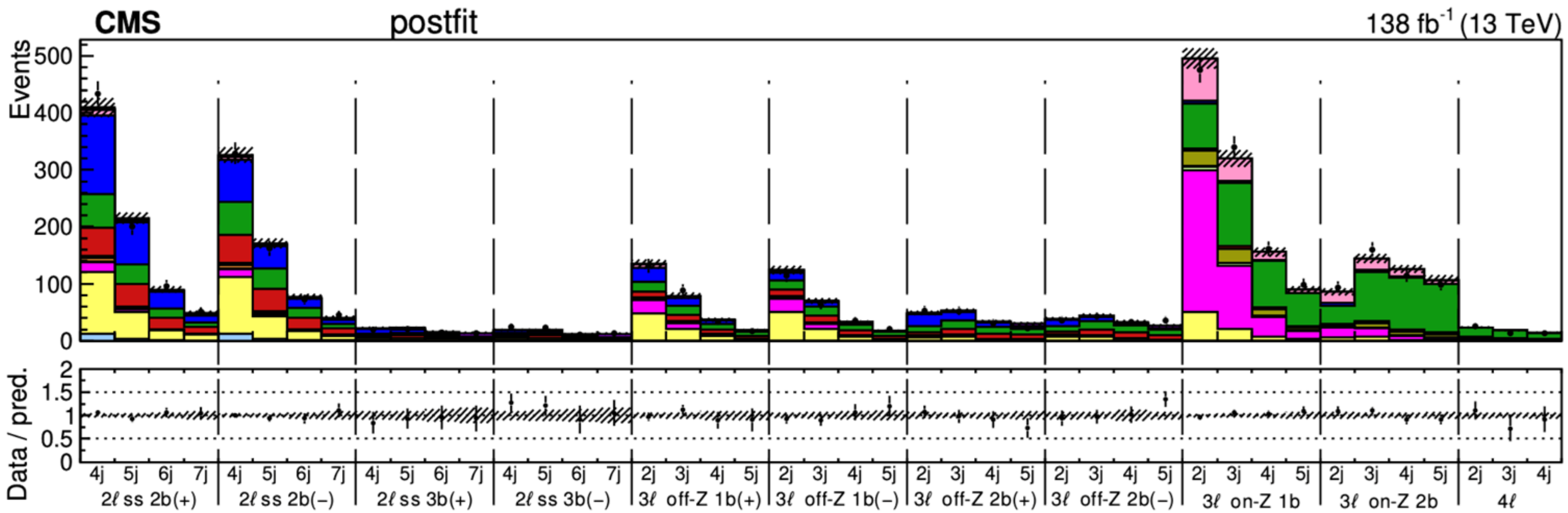


$p(x = p_{\rm T}(\gamma)|C^{(I)}_{\rm tZ})$ held the key to c_{tZ}

Everything else irrelevant!

$p(\sim 50 \text{ different 1D features}|p_{\rm T}(\gamma), \cancel{C^{(I)}_{\rm tZ}})$

optimal observable, best limit



- No optimal observable for many Wilson coefficients.
- tt+X TOP-22-006 captures leading SMEFT dependence with $p_T(\ell,j)$ or $p_T(Z)$
  - D=26 dimensional limits using 178 measurements

DETECTOR LEVEL ANALYSES

- Test statistic: Profiled likelihood ratio (unbinned) θ … SMEFT POI, $\boldsymbol{\nu}_k$… nuisances
- Distribution $q_{\boldsymbol{\theta}}(\mathcal{D}) = -2\log\frac{\max_{\boldsymbol{\nu}} \quad L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu})}{\max_{\boldsymbol{\nu},\boldsymbol{\theta}} L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu})}$ is asymptotically χ^2 *independent of* $\boldsymbol{\nu}$
  - Solve $q_{\boldsymbol{\theta}} = F^{-1}_{\chi^2_{N_{\boldsymbol{\theta}}}}(1-\alpha)$ with α=5% for θ to obtain confidence regions.

$$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \mathrm{Pois}_{\mathcal{L}\sigma(\boldsymbol{\theta},\boldsymbol{\nu})}(N) \times \prod_{i=1}^{N} p(\boldsymbol{x}_i|\boldsymbol{\theta},\boldsymbol{\nu}) = \mathrm{Pois}_{\mathcal{L}\sigma(\boldsymbol{\theta},\boldsymbol{\nu})}(N) \times \prod_{i=1}^{N} \frac{1}{\sigma_{\boldsymbol{\theta},\boldsymbol{\nu}}} \frac{\mathrm{d}\sigma_{\boldsymbol{\theta},\boldsymbol{\nu}}(\boldsymbol{x}_i)}{\mathrm{d}\boldsymbol{x}}$$

- Binned approximation for generic detector-level SMEFT analyses

$$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \prod_{j=1}^{N_{\mathrm{bins}}} \mathrm{Pois}\left(n_j \middle| \underbrace{\lambda_j(\mathrm{SM})}_{\text{SM yield}} \cdot \underbrace{(1 + \boldsymbol{\theta}^{\intercal}\mathbf{R}_{\mathrm{lin},j} + \boldsymbol{\theta}^{\intercal}\mathbf{R}_{\mathrm{quad},j}\boldsymbol{\theta})}_{\text{SMEFT dependence}} \cdot \underbrace{\prod_k \alpha_{j,k}^{\nu_k}}_{\text{prefit uncertainties}} + b_j(\boldsymbol{\nu})\right) \underbrace{\prod_{k=1}^{N_{\mathrm{nuis.}}} C_k(\nu_k)}_{\text{penalties}}$$

FACTORIZING SMEFT AND SYSTEMATICS

- We approximate the yield in a bin j as

$$\lambda_j(\boldsymbol{\theta},\boldsymbol{\nu}) \approx \lambda_j(\mathrm{SM}) \cdot (1 + \boldsymbol{\theta}^\intercal \mathbf{R}_{\mathrm{lin},j} + \boldsymbol{\theta}^\intercal \mathbf{R}_{\mathrm{quad},j}\boldsymbol{\theta}) \cdot \prod_k \alpha_{j,k}^{\nu_k}$$

- How to get the SMEFT dependence from simulation?

$$\lambda_j(\boldsymbol{\theta},\boldsymbol{\nu}) = \mathcal{L}(\boldsymbol{\nu}) \int_{\Delta\mathbf{x}(j)} \mathrm{d}\mathbf{x} \frac{\mathrm{d}\sigma_{\boldsymbol{\theta},\boldsymbol{\nu}}}{\mathrm{d}\mathbf{x}} = \mathcal{L}(\boldsymbol{\nu}) \boxed{\int_{\Delta\mathbf{x}(j)} \mathrm{d}\mathbf{x} \int \mathrm{d}\mathbf{z}\, \sigma(\boldsymbol{\theta},\boldsymbol{\nu}) p(\mathbf{x},\mathbf{z}|\boldsymbol{\theta},\boldsymbol{\nu})}$$

1) simulation samples the joint space (x,z)

$$= \mathcal{L}(\boldsymbol{\nu}) \int_{\Delta\mathbf{x}(j)} \mathrm{d}\mathbf{x} \int \mathrm{d}\mathbf{z}\, \sigma(\boldsymbol{\theta},\boldsymbol{\nu}) p(\mathbf{x},\mathbf{z}|\mathrm{SM},\boldsymbol{\nu}=\mathbf{0}) r(\mathbf{x},\mathbf{z}|\boldsymbol{\theta},\boldsymbol{\nu})$$

2) sampling at a fixed reference point

$$\underbrace{r = \frac{p(x_{\mathrm{det}},\cdots,z_{\mathrm{ptl}},\cdots,z_{\mathrm{p}}|\theta)}{p(x_{\mathrm{det}},\cdots,z_{\mathrm{ptl}},\cdots,z_{\mathrm{p}}|\mathrm{SM})} = \frac{p(x_{\mathrm{det}}|z_{\mathrm{ptl}})\cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}})\cdots p(z_{\mathrm{p}}|\theta)}{p(x_{\mathrm{det}}|z_{\mathrm{ptl}})\cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}})\cdots p(z_{\mathrm{p}}|\mathrm{SM})}}_{\text{assume systematics to factor out – SMEFT is at a higher energy scale than modelling / detector}} = \underbrace{\frac{p(z_{\mathrm{p}}|\theta)}{p(z_{\mathrm{p}}|\mathrm{SM})} \sim \frac{|\mathcal{M}(z_{\mathrm{p}},\theta)|^2}{|\mathcal{M}(z_{\mathrm{p}},\mathrm{SM})|^2}}_{\text{possible “post mortem” (in CMS EFT combination)}}$$

Change in likelihood of simulated observation x with latent “history” z going from “SM” to θ

staged simulation in forward mode: Intractable factors cancel

re-calcuable theory prediction

latent-space re-weighted simulation

assume systematics to factor out – SMEFT is at a higher energy scale than modelling / detector

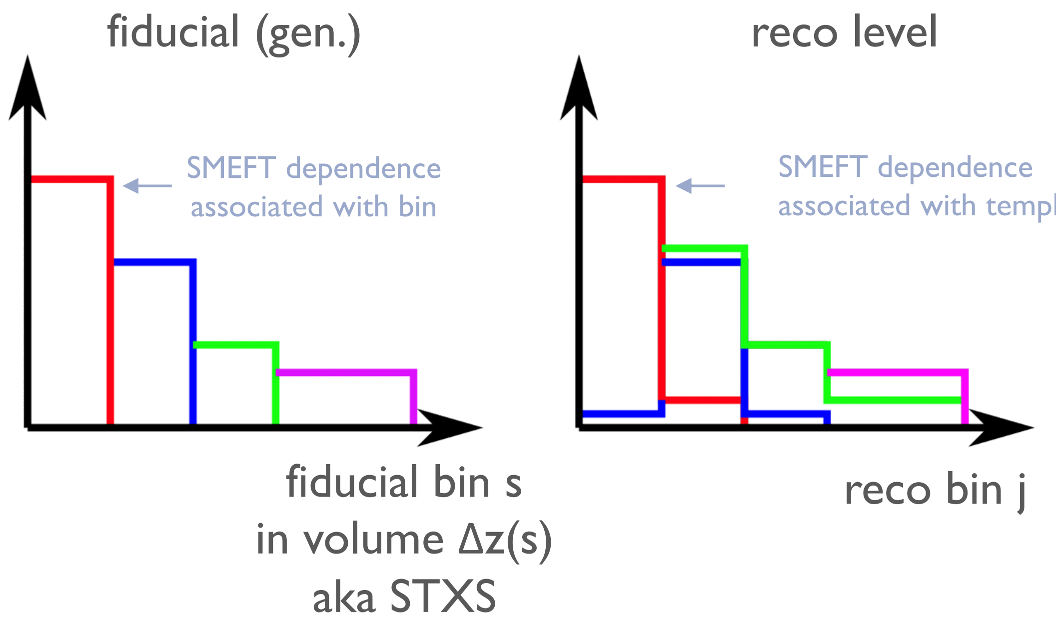
possible “post mortem” (in CMS EFT combination)

UNFOLDED MEASUREMENTS

- A detector level analysis *fully integrates the latent space* in each bin.
- Unfolding: Also split in fiducial (gen-level) bins

$$\int_{\Delta \boldsymbol{x}(j)} \mathrm{d}\boldsymbol{x} \int \mathrm{d}\mathbf{z} \rightarrow \int_{\Delta \boldsymbol{x}(j)} \mathrm{d}\boldsymbol{x} \sum_f \int_{\Delta \mathbf{z}(f)} \mathrm{d}\mathbf{z} \rightarrow \sum_f R_f(\boldsymbol{\theta}) \lambda_{j,f}$$

  - SMEFT dependence on the fiducial bin
  - A good approximation if p(x|z) is well localized and independent of θ.
  - Otherwise, SMEFT "acceptance" effects.
    - TTbar amenable to unfolding
    - Even STXS acceptance effects are under control
  - Well defined data representation for the outside.



unfolded: SMEFT effects on the fiducial bin

$$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \prod_{j=1}^{N_{\text{bins}}} \text{Pois}\left(n_j \middle| \sum_f \lambda_{j,f}(\boldsymbol{\nu}) R_f(\boldsymbol{\theta}) + b_j(\boldsymbol{\nu})\right) \prod_{k=1}^{N_{\text{nuis.}}} C_k(\nu_k)$$

χ^2 APPROXIMATION + CMS EFT COMBINATION

- All cases enter in the CMS EFT combination
  - → Higgs:
    - CMS-HIG-19-015, STXS $H \to \gamma\gamma$
  - → Top:
    - CMS-TOP-17-023, single top, t-channel
    - CMS-TOP-17-002, $t\bar{t}$
    - CMS-TOP-22-006, $t\bar{t}+X$, $t+X$
  - → Electroweak:
    - CMS-SMP-20-005, $W\gamma$
    - CMS-SMP-18-004, WW
    - CMS-SMP-18-003, $Z \to \nu\bar{\nu}$
    - hep-ex/0509008, EWPO from LEP+SLC
  - → QCD:
    - CMS-SMP-20-011, inclusive jets
- Highlight #1: PCA
- Highlight #2: "post mortem reweighting"
  - Evaluate $\dfrac{|\mathcal{M}(z_p,\theta)|^2}{|\mathcal{M}(z_p,\mathrm{SM})|^2}$ on existing sample!!

Gaussian approximate the likelihood ratio
using the total covariance of the unfolded measurement

$$\chi^2(\boldsymbol{\theta}) = \frac{\exp\left\{-\frac{1}{2}\left((\hat{\boldsymbol{R}} - \boldsymbol{R}(\boldsymbol{\theta}))V^{-1}(\hat{\boldsymbol{R}} - \boldsymbol{R}(\boldsymbol{\theta}))\right)\right\}}{(2\pi)^{M/2}\det(V)}$$

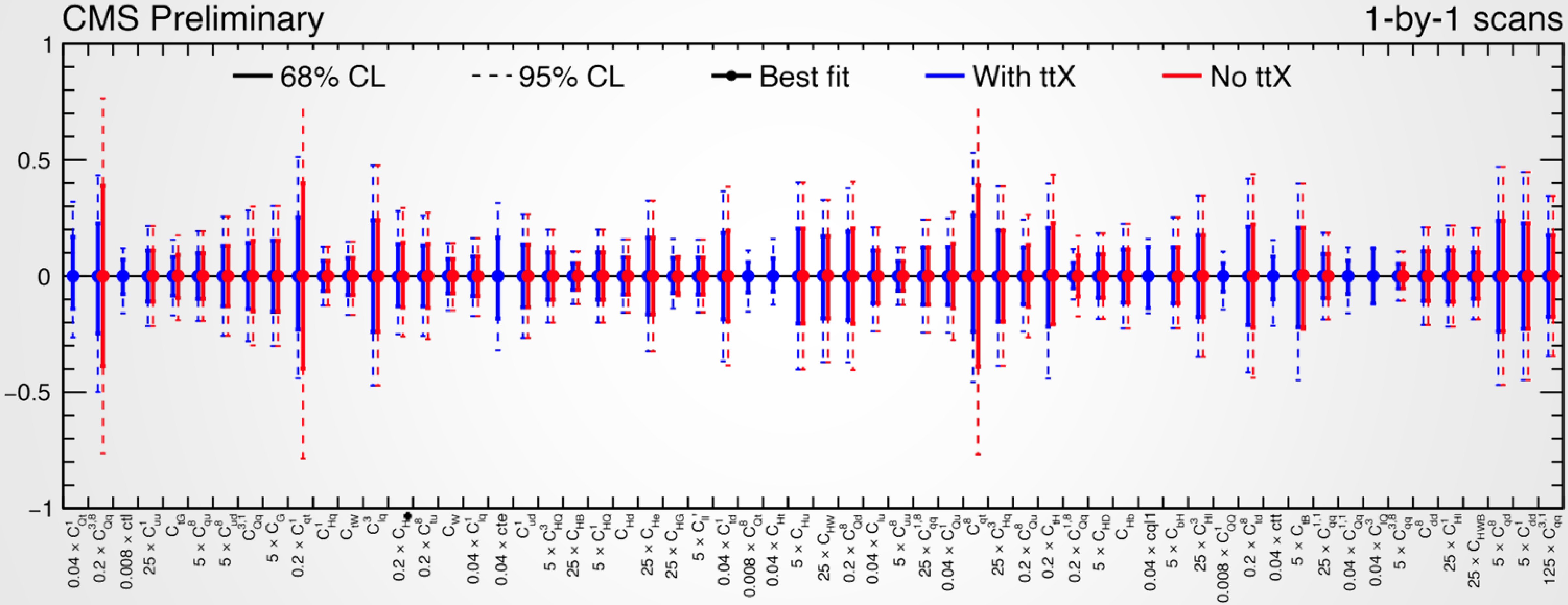
unfolded: SMEFT effects on the fiducial bin

$$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \prod_{j=1}^{N_{\text{bins}}} \text{Pois}\left(n_j \middle| \sum_f \lambda_{j,f}(\boldsymbol{\nu}) R_f(\boldsymbol{\theta}) + b_j(\boldsymbol{\nu})\right) \prod_{k=1}^{N_{\text{nuis.}}} C_k(\nu_k)$$

SMEFT effects on the detector-level bin

$$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \prod_{j=1}^{N_{\text{bins}}} \text{Pois}\left(n_j \middle| R_j(\boldsymbol{\theta})\lambda_j(\boldsymbol{\nu}) + b_j(\boldsymbol{\nu})\right) \prod_{k=1}^{N_{\text{nuis.}}} C_k(\nu_k)$$

CMS EFT COMBINATION

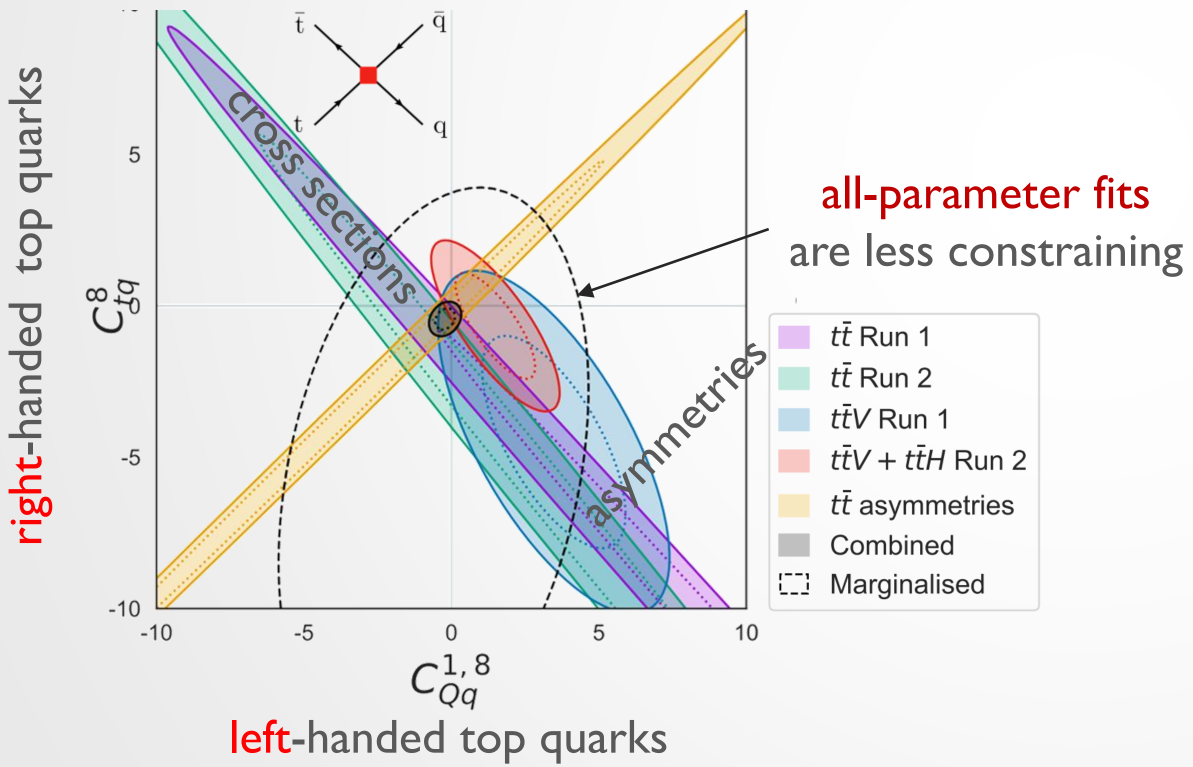


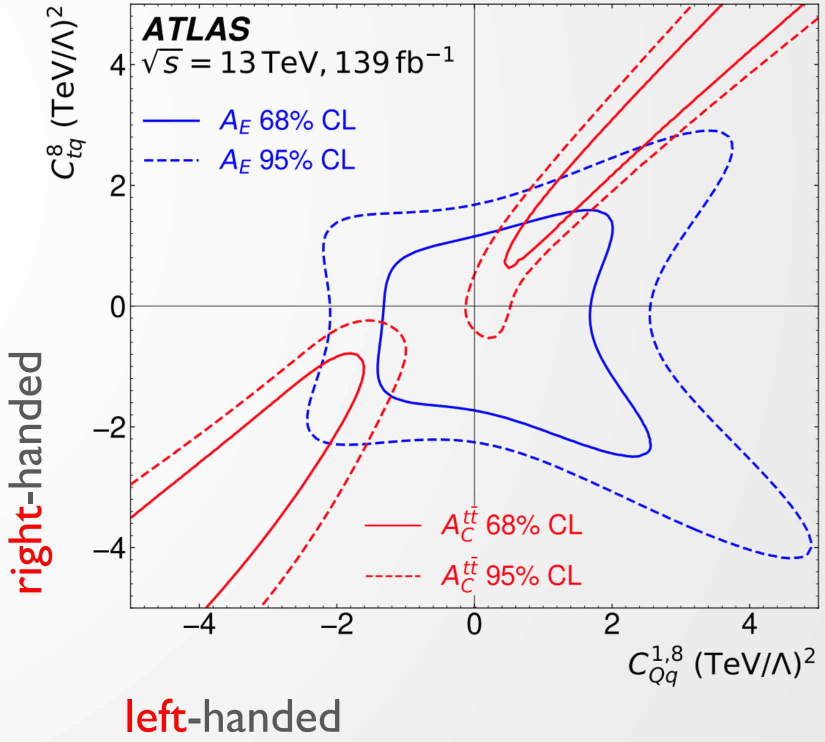
EXTERNAL GLOBAL FITS

[Ellis, Sanz, et.al. FitMaker JHEP04(2021)279]

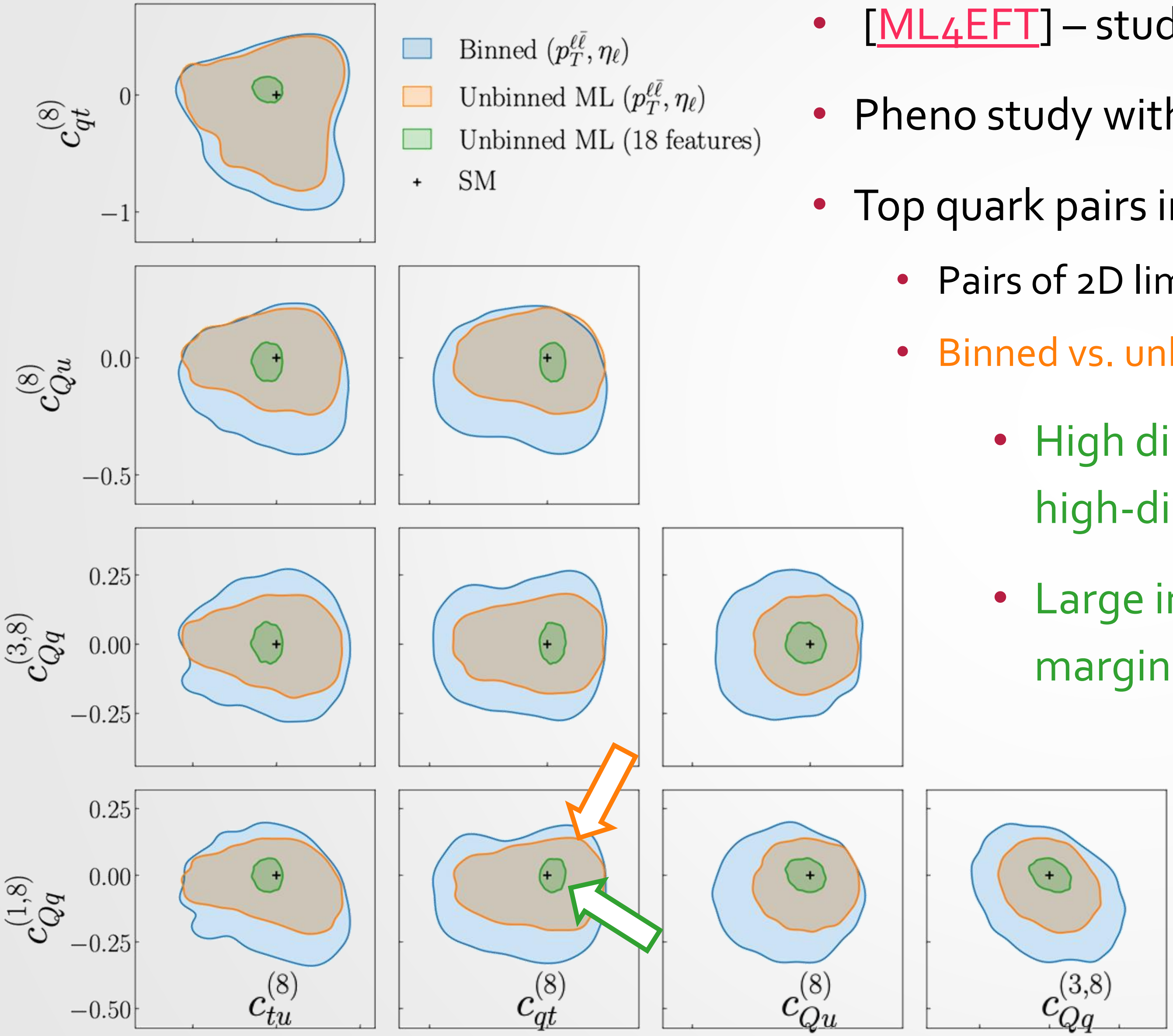
- Early example: Left- and right-handed 4-fermion operators
  - two-at-atime: tight constraint from combined measurement
  - Factor ~10 less powerful marginalized

- ATLAS charge asymmetry vs. energy asymmetry (two-at-a-time) shows comparable same pattern



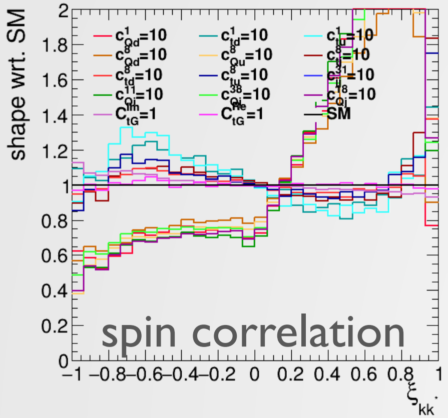


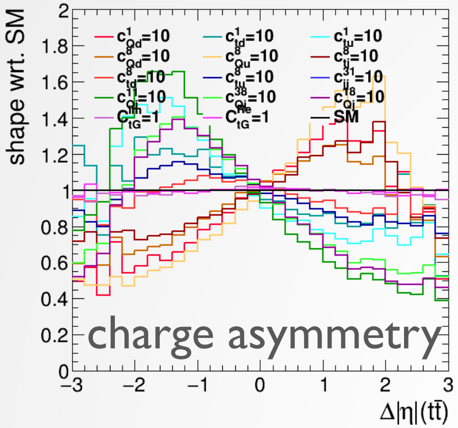
IMPROVING HIGH DIMENSIONAL LIMITS

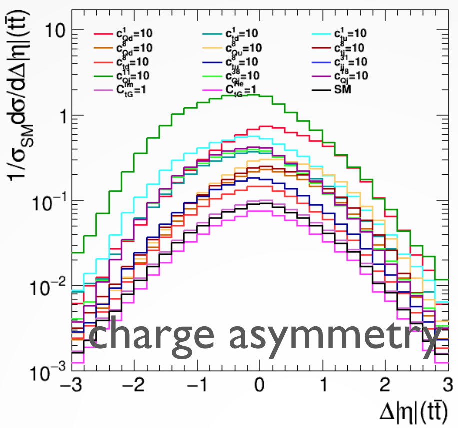


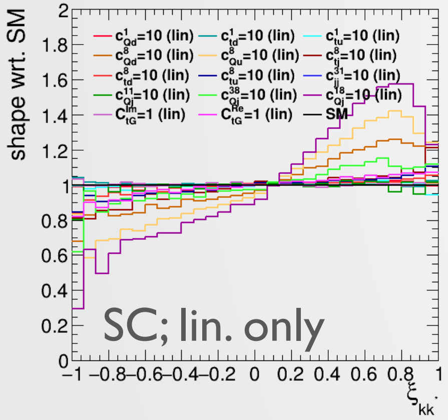
- [ML4EFT] – study ZH and top quark pairs
- Pheno study with parametrized NN classifiers
- Top quark pairs in low (N_f=2) and high feature dimension N_f=18
  - Pairs of 2D limits with 6 more ops marginalized
  - Binned vs. unbinned: Some gain w/ unbinned when using 2 features
    - High dimensional observation (N_f=18) constraining a high-dimensional (N_{coef}=8) model using an SM candle
    - Large improvement for N_f=18– mostly in (only) the marginalized limits
- Whether the sensitivity gain survives systematics in an unbinned detector-level analysis is an open question

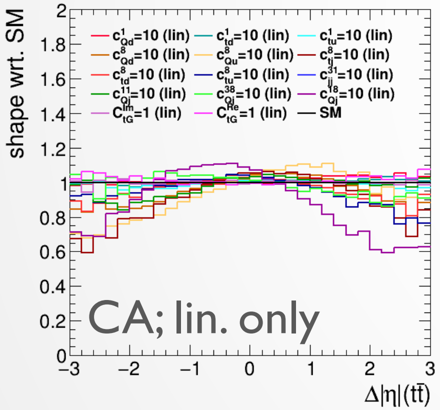
SMEFT EFFECTS & FEATURE CORRELATION

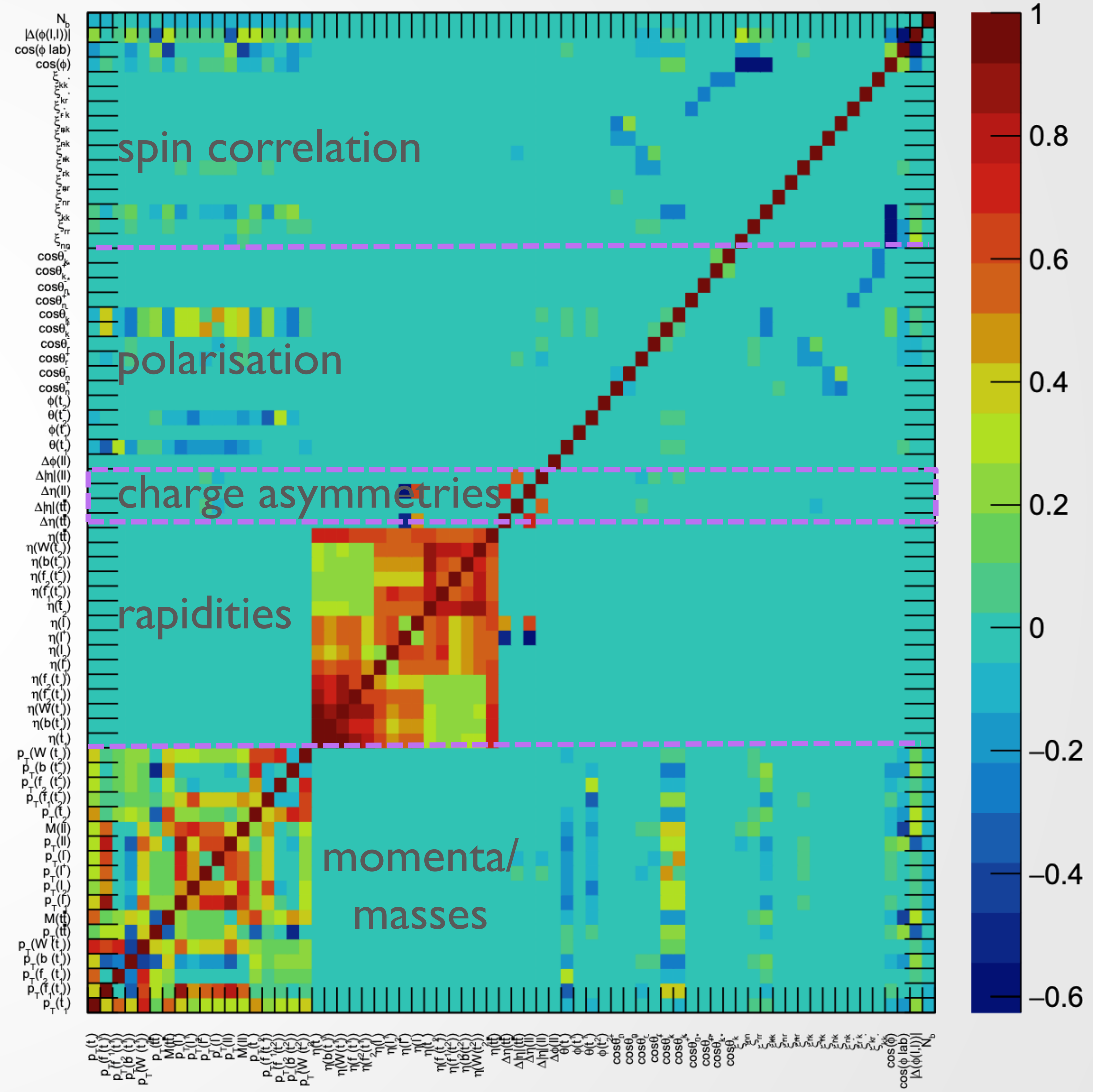












linear feature correlation in tt(2ℓ) for H_T>500

- TT(2ℓ) parton-level kinematics, charge asymmetry, spin-correlation, all have some information on some of the WC
- 16 operators, 72 features: TTbar is a SM candle with EFT sensitivity: Ideal case to develop methodology for an unbinned analyses

TOWARDS AN UNBINNED ANALYSIS

- We must be able to efficiently evaluate

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

- Parametrize the differential cross section ratio in terms of POIs and nuisances.

- ML fit of SMEFT dependence: A solved problem

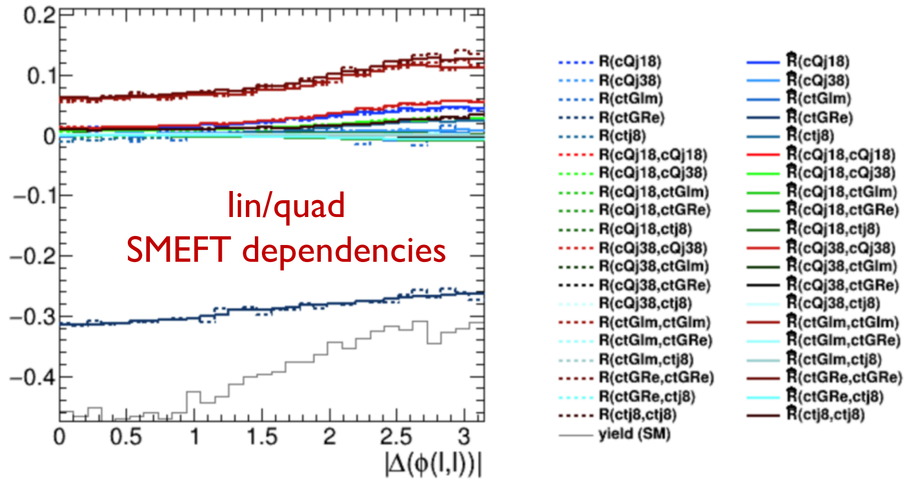
$$L = \left\langle \left(r(x_{\text{det}}, z_{\text{ptl}}, \cdots, z_{\text{p}}|\theta) - \hat{f}_\theta(x_{\text{det}}) \right)^2 \right\rangle_{\text{SM}}$$

Available in simulation!

Only nominal simulation

$$\text{argmin}_{\hat{f}(x)} L = \frac{d\sigma_{\boldsymbol{\theta},\boldsymbol{\nu}=0}}{d\sigma_{\text{SM},\boldsymbol{\nu}=0}}$$

Fit with neural networks or with trees



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

PARAMETRIZING THE LIKELIHOOD RATIO

- What about the systematic dependence?

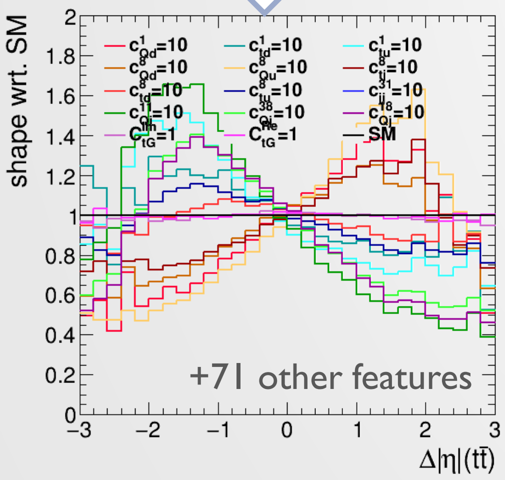
$$\frac{d\sigma_{\theta,\nu}}{d\sigma_{\theta_0,\nu_0}}(x) = \underbrace{\left[\prod_{l=1}^{N_\ell}(1+\Delta\mathrm{SF}(\ell_l))\right]^{\nu_\ell}}_{\text{weight-based systematics}} \underbrace{\times\cdots\times}_{?} \underbrace{\exp\left(\nu_{\mathrm{JEC}}^\top \hat{\delta}_{\mathrm{JEC}}(x) + \nu_{\mathrm{JEC}}^\top \hat{\Delta}_{\mathrm{JEC}}(x)\nu_{\mathrm{JEC}}\right)}_{\text{parametrized systematics}} \times \underbrace{\left(1+\theta^\top \hat{R}_{\mathrm{lin}}(x) + \theta^\top \hat{R}_{\mathrm{quad}}(x)\theta\right)}_{\text{SMEFT dependence}}$$

weight-based systematics

?

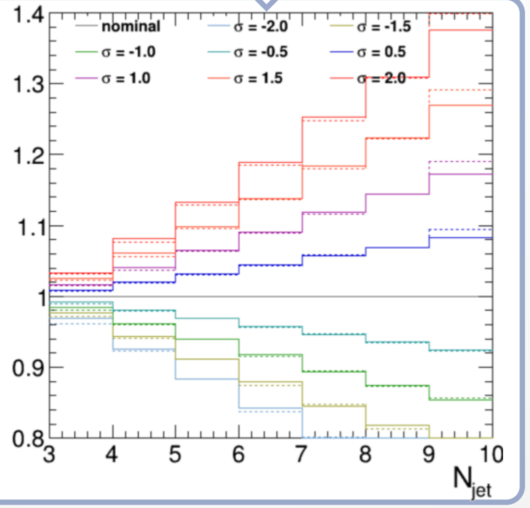
parametrized systematics

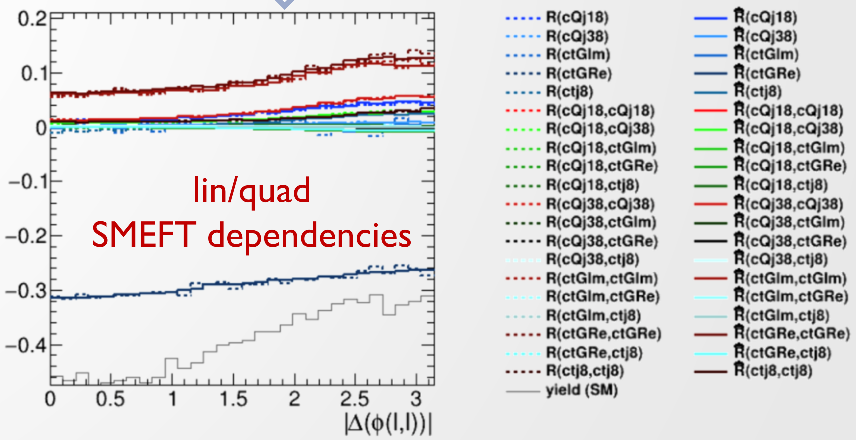
SMEFT dependence: $\mathbb{R}^{72} \to \mathbb{R}^{136}$
(learned with Boosted Information Tree)



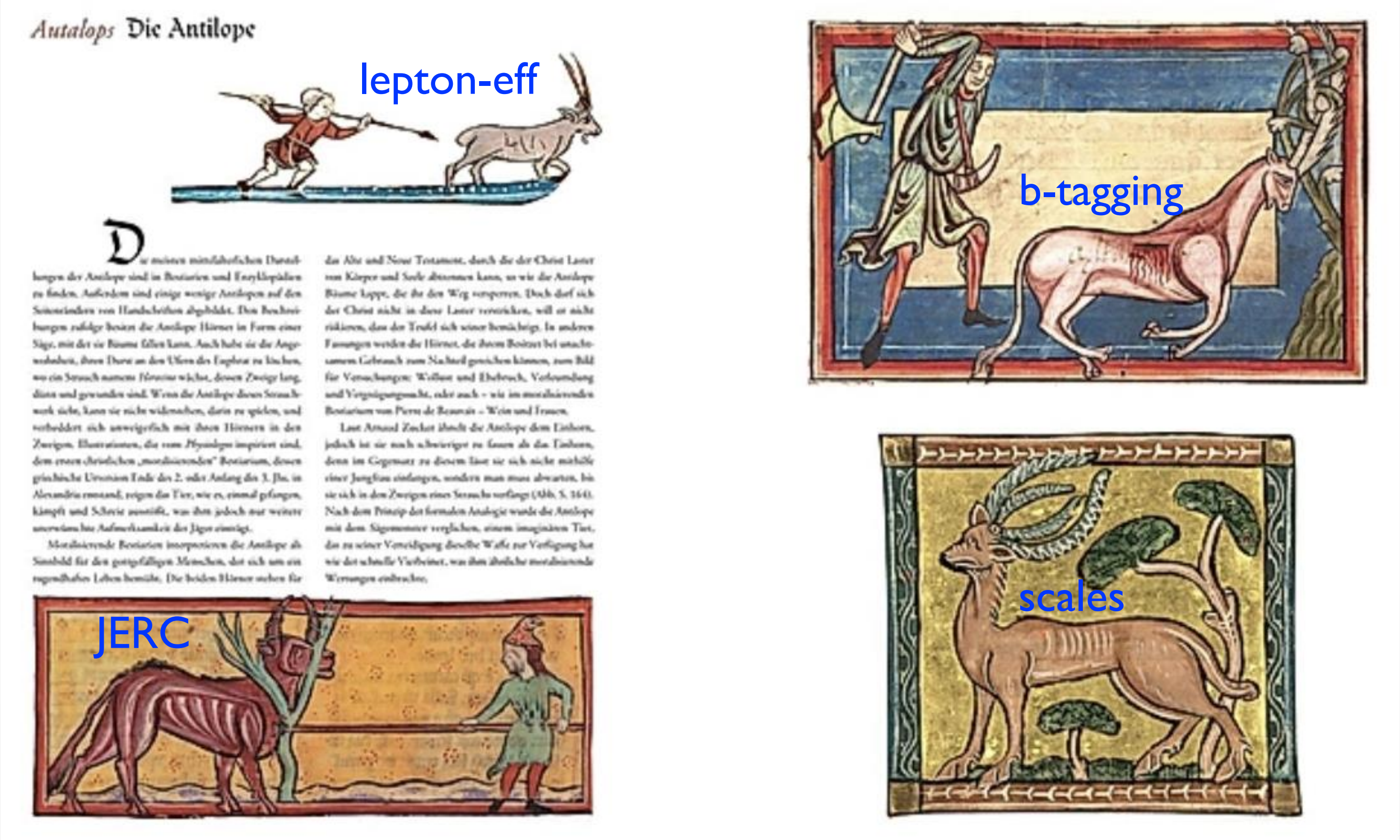
Need to parametrize nuisances, e.g., JEC (here: total)

$$L = -\sum_{\nu\in\mathcal{V}} \left\{ \left\langle \mathrm{Soft}^+(\nu^\top\hat{\delta}(x) + \nu^\top\hat{\Delta}(x)\nu) \right\rangle_{\mathrm{SM}} + \left\langle \mathrm{Soft}^+(-\nu^\top\hat{\delta}(x) - \nu^\top\hat{\Delta}(x)\nu) \right\rangle_{\nu} \right\}$$





BESTIARY OF SYSTEMATICS (RECIPIES)



CALIBRATION SYSTEMATICS

- Need: Likelihood ratio for the same observatio **x** and different **ν**: $r(\mathbf{x}|\boldsymbol{\nu},\mathbf{0}) = \frac{p(\mathbf{x}|\boldsymbol{\nu})}{p(\mathbf{x}|\mathbf{0})}$
- Have: JME prescription "J" that varies **x** and **ν** such that the likleihood const.: $p(\mathbf{x}_i,\mathbf{z}_i|\mathbf{0}) = p\left(J_{\boldsymbol{\nu}}(\mathbf{x}_i,\mathbf{z}_i),\mathbf{z}_i|\boldsymbol{\nu}\right)$
- Solution: Train a parametric regressor a parametrisation

$$\frac{p(\boldsymbol{x}|\boldsymbol{\nu})}{p(\boldsymbol{x}|\mathbf{0})} = \exp\underbrace{\left(\boldsymbol{\nu}_{\mathrm{JEC}}^{\top}\hat{\boldsymbol{\delta}}_{\mathrm{JEC}}(\boldsymbol{x}) + \boldsymbol{\nu}_{\mathrm{JEC}}^{\top}\hat{\boldsymbol{\Delta}}_{\mathrm{JEC}}(\boldsymbol{x})\boldsymbol{\nu}_{\mathrm{JEC}}\right)}_{\text{a factor in the LLR parametric in } \nu_{\mathrm{JEC}}}$$

$$L = -\sum_{\boldsymbol{\nu}\in\mathcal{V}}\left\{\left\langle \mathrm{Soft}^{+}\left(\boldsymbol{\nu}^{\top}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) + \boldsymbol{\nu}^{\top}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}\right)\right\rangle_{\mathrm{SM}} + \left\langle \mathrm{Soft}^{+}\left(-\boldsymbol{\nu}^{\top}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) - \boldsymbol{\nu}^{\top}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}\right)\right\rangle_{\boldsymbol{\nu}}\right\}$$

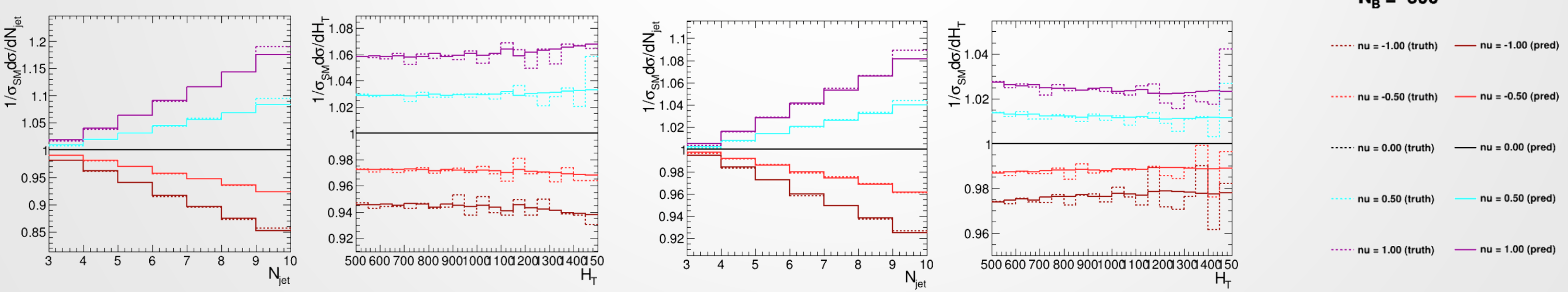
JEC variied samples
Use many!

Tree or NN

Tree or NN

TT(2ℓ), H_T>500 (8 features), **ν**: JES total

ν: JESFlavorQCD



OBJECT LEVEL: B-TAGGING

- Must we learn b-tagging? After all, it produces event weights that can be interpreted as likleihood ratio

$$F(\nu_k, \text{jets}) = \prod_{\text{tagged jets}} \varepsilon_f(p_\mathrm{T}, \eta)\,(\mathrm{SF}_f(p_\mathrm{T}, \eta) + \nu_k \Delta\mathrm{SF}_{f,k}(p_\mathrm{T}, \eta))$$
$$\times \prod_{\text{untagged jets}} (1 - \varepsilon_f(p_\mathrm{T}, \eta)(\mathrm{SF}_f(p_\mathrm{T}, \eta) + \nu_k \Delta\mathrm{SF}_{f,k}(p_\mathrm{T}, \eta)))$$
$$\omega_{\boldsymbol{\nu},i} = \omega_i \prod_{k=1}^{K} \frac{F(\nu_k, \text{jets in event } i)}{F(0, \text{jets in event } i)}.$$

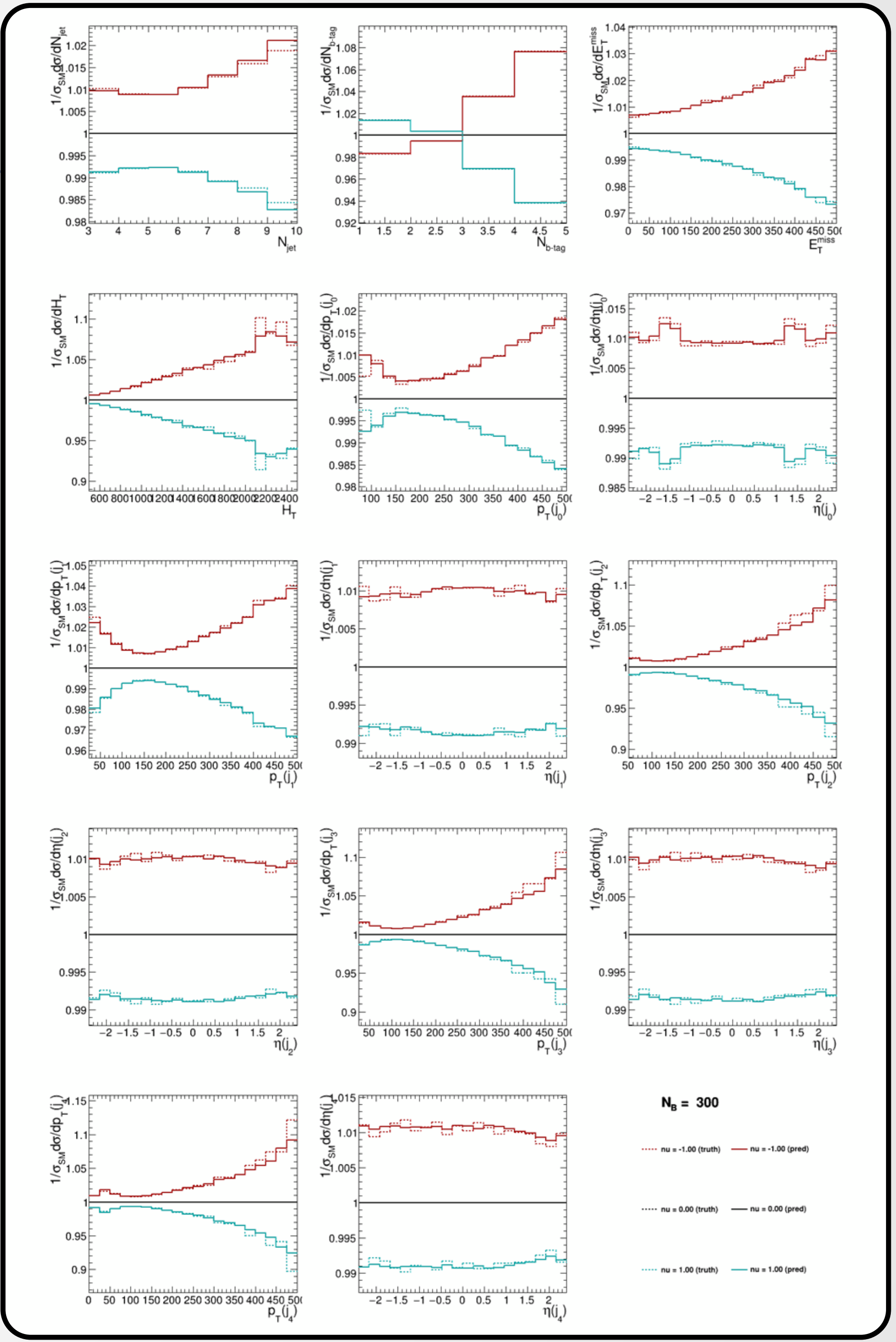
- However f, the true jet flavor, is latent. Jet p_T, η are observed.

$$\frac{\omega_{\boldsymbol{\nu},i}}{\omega_i} \approx \frac{p(\mathbf{x}_i, \mathbf{z}_i|\boldsymbol{\nu})}{p(\mathbf{x}_i, \mathbf{z}_i|\mathbf{0})} = r(\mathbf{x}_i, \mathbf{z}_i|\boldsymbol{\nu}, \mathbf{0})$$

B-tag reweighting is latent

  - Can not compute the LLR($\boldsymbol{\nu}_{HF}$) test statistic from an observation

- B-tagging must be learned.
- Pro: Covers efficiency & udsg/c-fakes



LEPTON UNCERTAINTIES & NORMALIZATIONS

- Must we learn lepton-efficiency SF ?

$$\omega_{\nu,i} = \omega_i \prod_{k=1}^{K} \prod_{\ell=1}^{N_\ell(i)} \left(1 + \nu_k \frac{\Delta_k \mathrm{SF}(\ell)}{\mathrm{SF}(\ell)}\right)$$

  - Parametrized in terms of measured quantities. Need not be learned.

- Normalization uncertainties
  - There is a catch: non-prompt/fakes are sourced by jets, hence we do not use scale factors.
- The non-prompt background component is on the same footing as tW, diboson, ttV backgrounds
- Train a multiclassifier to scale the cross sections within uncertainties

SUMMARY

LEARNING SYSTEMATICS

- Let's start with Cross-Entropy and a single variation $\boldsymbol{\nu}$

$$L = -\int \mathrm{d}\boldsymbol{x}\, \left\{ p(\boldsymbol{x}|\mathrm{SM}) \log \hat{f}_{\boldsymbol{\nu}}(\boldsymbol{x}) + p(\boldsymbol{x}|\boldsymbol{\nu}) \log(1 - \hat{f}_{\boldsymbol{\nu}}(\boldsymbol{x})) \right\}$$

- L converges to $f^{*}_{\boldsymbol{\nu}}(\boldsymbol{x}) \sim \dfrac{1}{1 + \frac{p(\boldsymbol{x}|\boldsymbol{\nu})}{p(\boldsymbol{x}|\mathrm{SM})}}$
- To learn a more general dependence, let's inject our exponential ansatz for NN outputs $\boldsymbol{\delta}$, Δ

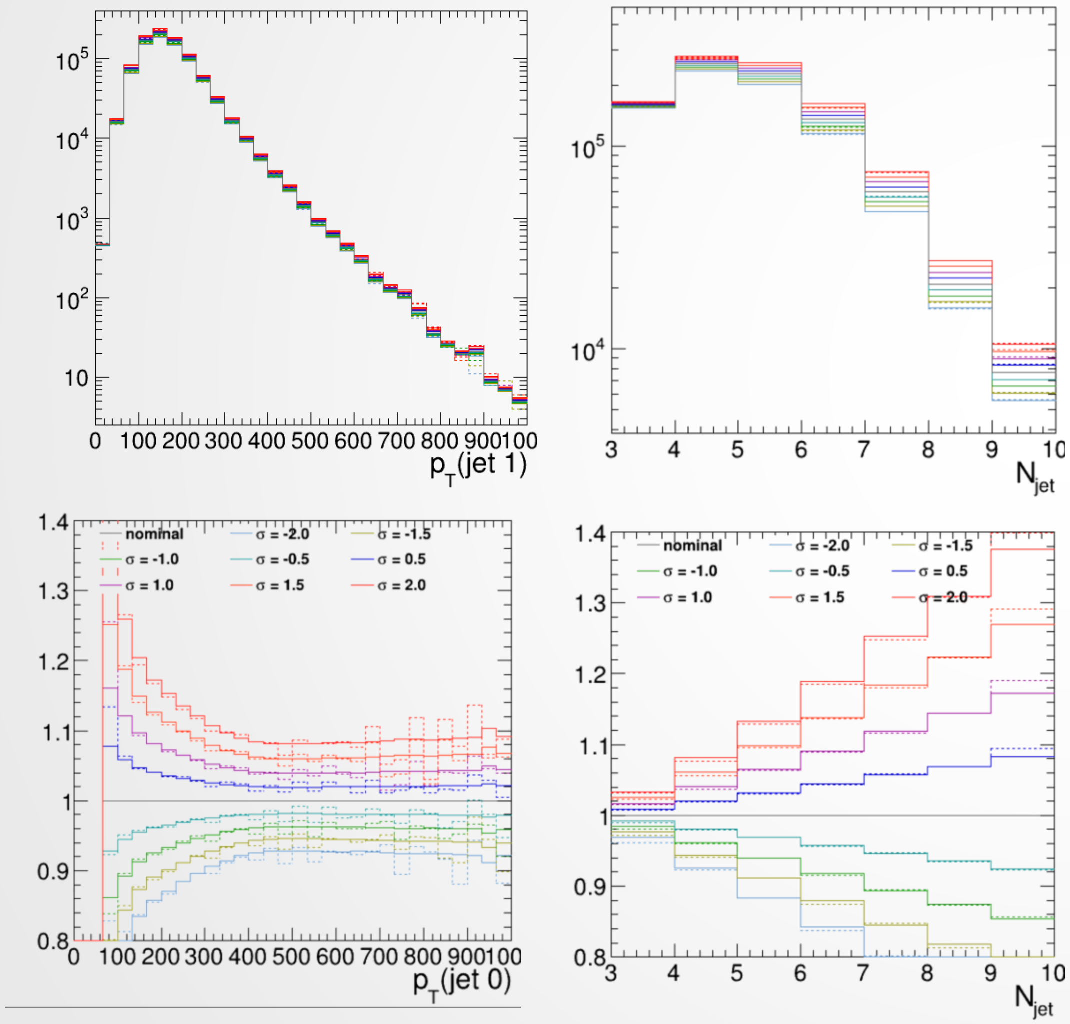
$$\hat{f}_{\boldsymbol{\nu}}(\boldsymbol{x}) = \frac{1}{1 + \exp\left(\boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) + \boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}\right)}$$

- This gives: $$L = \sum_{\boldsymbol{\nu}\in\mathcal{V}} \int \mathrm{d}\boldsymbol{x} \Bigg\{ \sigma(\mathrm{SM})\, p(\boldsymbol{x}|\mathrm{SM}) \log\left(1 + \exp\left(\boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) + \boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}\right)\right)$$ $$+ \sigma(\boldsymbol{\nu})\, p(\boldsymbol{x}|\boldsymbol{\nu}) \log\left(1 + \exp\left(-\boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) - \boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}\right)\right) \Bigg\}$$

which we can write as $$\boxed{L = -\sum_{\boldsymbol{\nu}\in\mathcal{V}} \left\{ \left\langle \mathrm{Soft}^{+}(\boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) + \boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}) \right\rangle_{\mathrm{SM}} + \left\langle \mathrm{Soft}^{+}(-\boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\delta}}(\boldsymbol{x}) - \boldsymbol{\nu}^{\intercal}\hat{\boldsymbol{\Delta}}(\boldsymbol{x})\boldsymbol{\nu}) \right\rangle_{\boldsymbol{\nu}} \right\}}$$

LEARNING JEC & OTHERS

- As a start, fit 7 features to JEC-total [all plots]



Probably more precise than binned 2-point variations

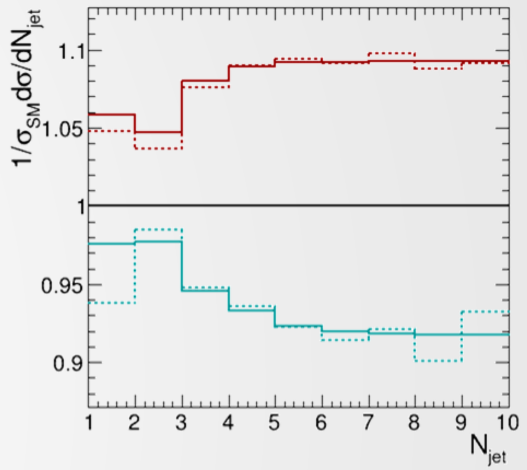
Expect no issues with fitting split JEC nuisance effects

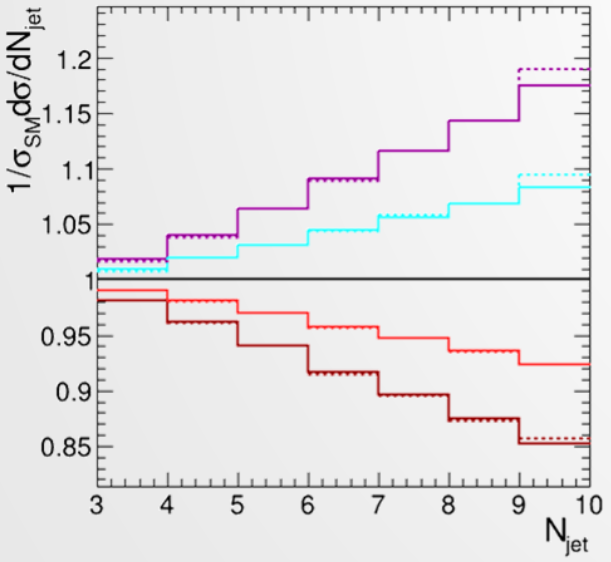
(Famous last words.)

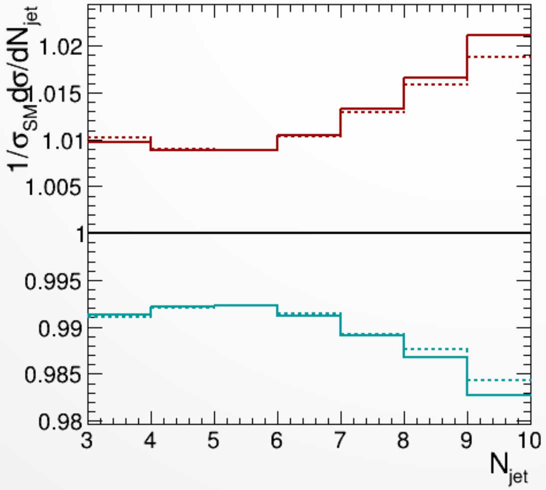
BOOSTED PARAMETRIC TREE

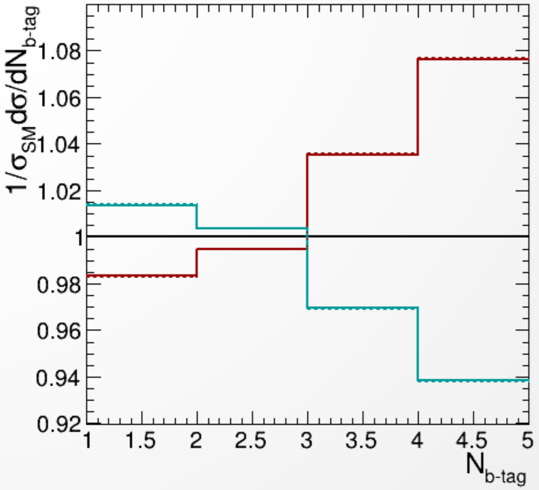
- Long story short: We made a parametric tree-boosting algorithm that learns the Taylor series of the log-likelihood ratio $\exp\left(\boldsymbol{\nu}_{\mathrm{JEC}}^{\intercal}\hat{\boldsymbol{\delta}}_{\mathrm{JEC}}(\boldsymbol{x}) + \boldsymbol{\nu}_{\mathrm{JEC}}^{\intercal}\hat{\boldsymbol{\Delta}}_{\mathrm{JEC}}(\boldsymbol{x})\boldsymbol{\nu}_{\mathrm{JEC}}\right)$
- It has the usual interpretation as a Fisher-Information optimum

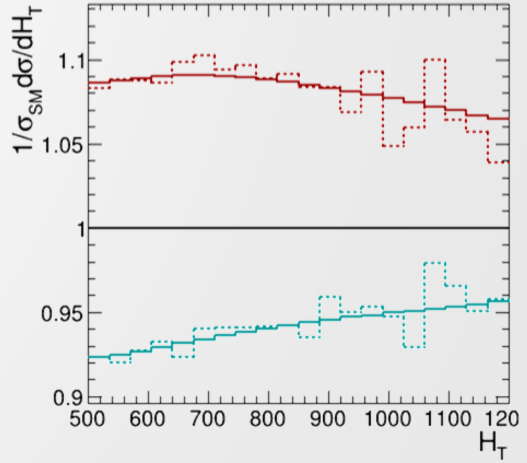
$$L[\mathcal{J}] = \sum_{j\in\mathcal{J}}\sum_{\boldsymbol{\nu}\in\mathcal{V}}\left[\lambda_{j,\mathbf{0}}\log\left(1+\frac{\lambda_{j,\boldsymbol{\nu}}}{\lambda_{j,\mathbf{0}}}\right)+\lambda_{j,\boldsymbol{\nu}}\log\left(1+\frac{\lambda_{j,\mathbf{0}}}{\lambda_{j,\boldsymbol{\nu}}}\right)\right] = -\frac{1}{4}\sum_{j\in\mathcal{J}}\sum_{\boldsymbol{\nu}\in\mathcal{V}}\nu_a\nu_b I_{(ab),j}+\ldots$$











TOOLS FOR R&D: UNBINNED ASIMOV DATASET

- [ML4EFT] derives the non-centrality term of the asymptotic χ^2 distribution of the test statistic as

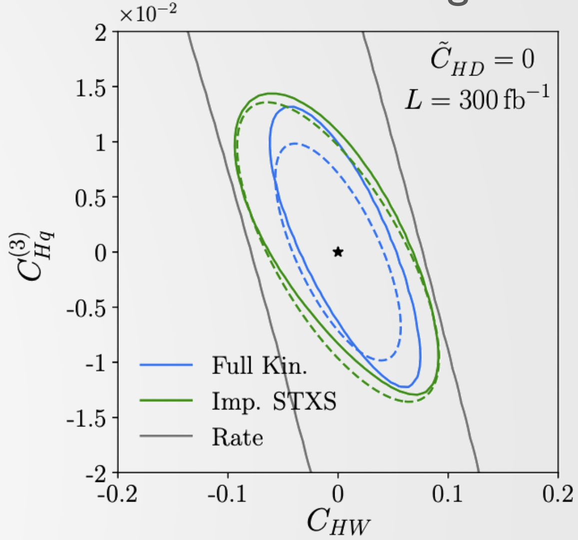
$$-\frac{1}{2}\Lambda(\mathcal{D}) = -\mathcal{L}(\boldsymbol{\nu})\sigma(\boldsymbol{\theta},\boldsymbol{\nu}) + \mathcal{L}(\boldsymbol{\nu}_0)\sigma(\boldsymbol{\theta}_0,\boldsymbol{\nu}_0) + \left\langle \log\left(\frac{\mathcal{L}(\boldsymbol{\nu}\ \ )}{\mathcal{L}(\boldsymbol{\nu}_0)}\frac{\mathrm{d}\sigma_{\boldsymbol{\theta}\ \ ,\boldsymbol{\nu}}}{\mathrm{d}\sigma_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0}}(\boldsymbol{x}_i)\right)\right\rangle_{\boldsymbol{\theta}_0,\boldsymbol{\nu}_0} + \sum_{k=1}^{N_{nuis.}} \log\frac{C_k(\nu_k)}{C_k(\nu_{0,k})}$$

- It is quite marvelous that, after minimizing the LLR, we can compute the asymptotic distribution of the test statistic for arbitrary parameter points without toys

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

WH with Bkgs



my practical experience

EXPLOITING SMEFT REWEIGHTING

$$L = \sum_{\theta\in\mathcal{B}} \left(\langle \hat{f}(x;\theta)^2 \rangle_{\theta} + \langle (1-\hat{f}(x;\theta))^2 \rangle_{\mathrm{SM}} \right)$$

θ *- aware* *EFT sample* *SM sample*

We start with SM and BSM samples

$$= \sum_{\theta\in\mathcal{B}} \int \mathrm{d}x\,\mathrm{dz}\, \left(p(x,z|\theta)\hat{f}(x;\theta)^2 + p(x,z|\mathrm{SM})(1-\hat{f}(x;\theta))^2 \right)$$

Let's write this under one integral
z … latent space

SM sample

$$= \sum_{\theta\in\mathcal{B}} \int \mathrm{d}x\,\mathrm{dz}\, p(x,z|\mathrm{SM}) \left(r(x,z|\theta)\hat{f}(x;\theta)^2 + (1-\hat{f}(x;\theta))^2 \right)$$

"joint" likelihood ratio

… and use just one sample
& joint likelihood ratio

$$r = \frac{p(x_{\mathrm{det}},\cdots,z_{\mathrm{ptl}},\cdots,z_{\mathrm{p}}|\theta)}{p(x_{\mathrm{det}},\cdots,z_{\mathrm{ptl}},\cdots,z_{\mathrm{p}}|\mathrm{SM})} = \frac{p(x_{\mathrm{det}}|z_{\mathrm{ptl}})\cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}})\cdots p(z_{\mathrm{p}}|\theta)}{p(x_{\mathrm{det}}|z_{\mathrm{ptl}})\cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}})\cdots p(z_{\mathrm{p}}|\mathrm{SM})} = \frac{p(z_{\mathrm{p}}|\theta)}{p(z_{\mathrm{p}}|\mathrm{SM})} \sim \frac{|\mathcal{M}(z_{\mathrm{p}},\theta)|^2}{|\mathcal{M}(z_{\mathrm{p}},\mathrm{SM})|^2}$$

Change in likelihood of simulated observation x with latent "history" z going from "SM" to **θ**

staged simulation in forward mode: Intractable factors cancel

re-calcuable theory prediction

weighted simulation

PARAMETRIZED CLASSIFIERS

K. Cranmer , J. Pavez , and G. Louppe [1506.02169]
J. Brehmer, K. Cranmer, G. Louppe, J. Pavez [1805.00013]
"Madminer" [1805.00020]
[1805.12244]
J. Brehmer, F. Kling, I. Espejo, K. Cranmer [1907.10621]

$$L = \sum_{\theta \in \mathcal{B}} \int \mathrm{d}x\, \mathrm{d}z\, p(x, z|\mathrm{SM}) \left(r(x, z|\theta) \hat{f}(x;\theta)^2 + (1 - \hat{f}(x;\theta))^2 \right)$$

MSE or cross entropy

$$\hat{f}(x;\theta) = \frac{1}{1 + \hat{R}(x;\theta)}$$

invert likelihood trick

Similar to
S. Chen, A. Glioti,
G. Panico, A. Wulzer
JHEP 05 (2021) 247

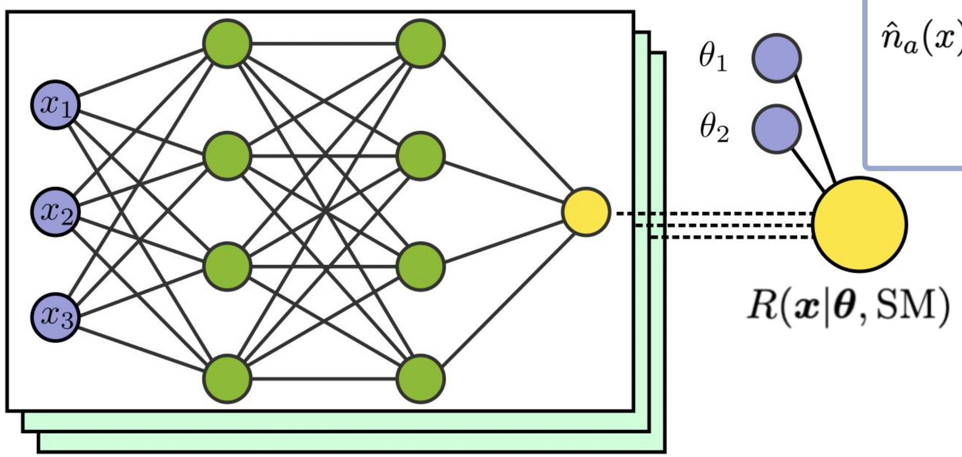
arXiv:2308.05704

insert model knowledge:
fit universal
coefficient functions

$$\hat{R}(x;\theta) = \Big(1 + \sum_a \theta_a \hat{n}_a(x)\Big)^2 + \sum_a \Big(\sum_{b \geq a} \theta_b \hat{n}_{ab}(x)\Big)^2$$

$$\hat{n}_a(x) \to \left.\frac{\partial_a p(x|\ \theta\ )}{p(x|\mathrm{SM})}\right|_{\theta=\mathrm{SM}} = \left.\frac{\partial_a \int \mathrm{d}z\ p(x, z|\theta)}{\int \mathrm{d}z\ p(x, z|\mathrm{SM})}\right|_{\theta=\mathrm{SM}}$$

Integrates latent space!

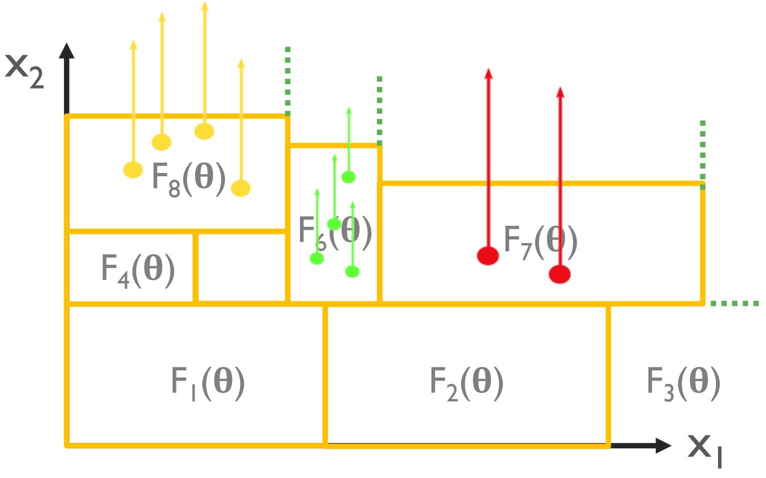


Why would you
want to use trees instead?

A SIMPLE TREE ALGORITHM

[arXiv:2107.10859, arXiv:2205:12976]

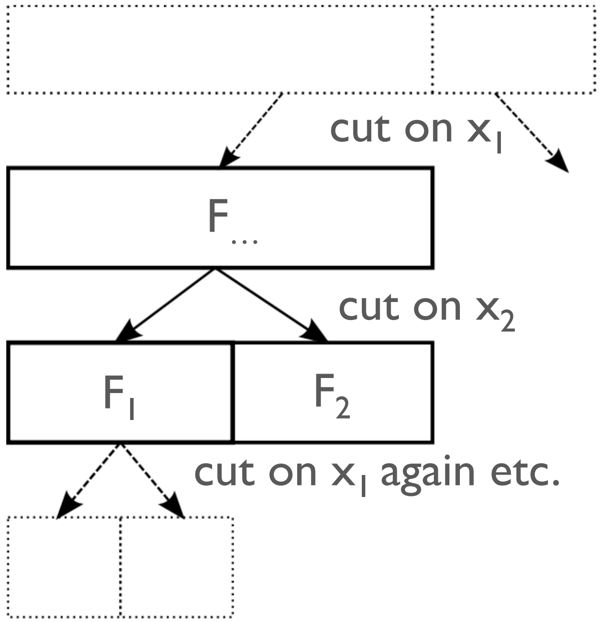
phase-space partitioning



A simple tree

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

training phase:
e.g. "CART" algo



- Let us make a tree-based ansatz for the differential cross-section ratio R
- The "weak learner" is a tree associating a sub-region (j) of a partitioning $\mathcal{J}$ with a predictive function F_j
  - Note: A tree algorithm can have an arbitrarily complicated predictive function; here it is a SMEFT polynomial
  - Fitting tree: Optimize "node split positions" on some loss. Trained (e.g. greedily) on the *ensemble*.

PARAMETRIZED TREES

[arXiv:2107.10859, arXiv:2205:12976]

Regress in r, including 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 $q(\mathcal{D})$

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

$$F^*(\boldsymbol{x}, \boldsymbol{\theta}) = R(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

Tree ansatz

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

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

Remove DOF from predictor:

$$F_j(\boldsymbol{\theta}) = \frac{\sum_{i \in j} w_i(\boldsymbol{\theta})}{\sum_{i \in j} w_i(\boldsymbol{\theta}_0)}$$

sum up event weights within node & divide

No trainable parameters in the predictor

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

CONCRETE SOLUTION: TREE BOOSTING

[arXiv:2107.10859, arXiv:2205:12976]

- Boosting: Fit model iteratively to pseudo-residuals of the preceding iteration with learning rate η
- Ansatz :

$$\hat{F}^{(b)}(\boldsymbol{x},\boldsymbol{\theta}) = \underbrace{\hat{f}(\boldsymbol{x},\boldsymbol{\theta})}_{\text{current iteration}} + \eta\underbrace{\hat{F}^{(b-1)}(\boldsymbol{x},\boldsymbol{\theta})}_{\text{previous iteration}}$$

- Insert into the loss function:

$$\underbrace{L[\hat{f}^{(b)}]}_{\text{current iteration}} = \sum_{\theta\in\mathcal{B}} \int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x},\boldsymbol{z}|\mathrm{SM}) \Big(\underbrace{r(x,z|\theta) - \eta\underbrace{\hat{F}^{(b-1)}(\boldsymbol{x},\boldsymbol{\theta})}_{\text{previous iteration}}}_{\text{pseudo-residual, amounting to event-leve reweighting}} - \underbrace{\hat{f}^{(b)}(\boldsymbol{x},\boldsymbol{\theta})}_{\text{current iteration}}\Big)^2$$

$$w_i^{(b)}(\boldsymbol{\theta}) \quad\rightarrow\quad w_i^{(b-1)}(\boldsymbol{\theta}) - \eta\, w_i^{(b-1)}(\boldsymbol{\theta}_0)\hat{F}^{(b-1)}(\boldsymbol{x}_i,\boldsymbol{\theta})$$

.... perform this iteratively
"Boosted Information Tree"

SIMULATION BASED INFERENCE

[Madminer 1805.00020]
Full list of references in backup

1. Simulation: $p(x_{\mathrm{det}}, \cdots, z_{\mathrm{ptl}}, \cdots, z_{\mathrm{p}}|\theta)$ Needed: $p(x_{\mathrm{det}}|\theta) = \int dz_{\mathrm{ptl}} \cdots \int dz_{\mathrm{p}}\, p(x_{\mathrm{det}}|z_{\mathrm{ptl}}) \cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}}) \cdots p(z_{\mathrm{p}}|\theta)$

2. Exploit simplicity of the joint space: Intractable factors cancel in the joint likelihood ratio

$$r = \frac{p(x_{\mathrm{det}}, \cdots, z_{\mathrm{ptl}}, \cdots, z_{\mathrm{p}}|\theta)}{p(x_{\mathrm{det}}, \cdots, z_{\mathrm{ptl}}, \cdots, z_{\mathrm{p}}|\mathrm{SM})} = \frac{p(x_{\mathrm{det}}|z_{\mathrm{ptl}}) \cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}}) \cdots p(z_{\mathrm{p}}|\theta)}{p(x_{\mathrm{det}}|z_{\mathrm{ptl}}) \cdots p(z_{\mathrm{ptl}}|z_{\mathrm{p}}) \cdots p(z_{\mathrm{p}}|\mathrm{SM})} = \frac{p(z_{\mathrm{p}}|\theta)}{p(z_{\mathrm{p}}|\mathrm{SM})} \sim \frac{|\mathcal{M}(z_{\mathrm{p}}, \theta)|^2}{|\mathcal{M}(z_{\mathrm{p}}, \mathrm{SM})|^2}$$

Change in likelihood of simulated observation x with latent "history" z going from "SM" to θ

staged simulation in forward mode: Intractable factors cancel

re-calcuable theory prediction

weighted simulation

3. Regress (e.g.) in the joint likelihood ratio, ignoring the latent space.

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

Available in simulation!
(MSE loss only for illustration)

4. Obtain change of likelihood for a specific observation, suitably integrating latent histories. NP optimal!

$$\mathrm{argmin}_{\hat{f}(x)} L = \frac{p(x|\theta\quad)}{p(x|\mathrm{SM})} = \text{ ratio of integrals}$$

what we actually want: change in likelihood of a specific observation

Latent space is integrated in numerator and denominator