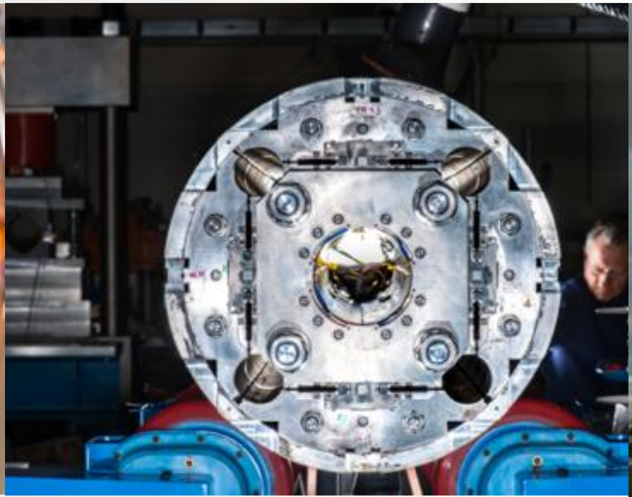
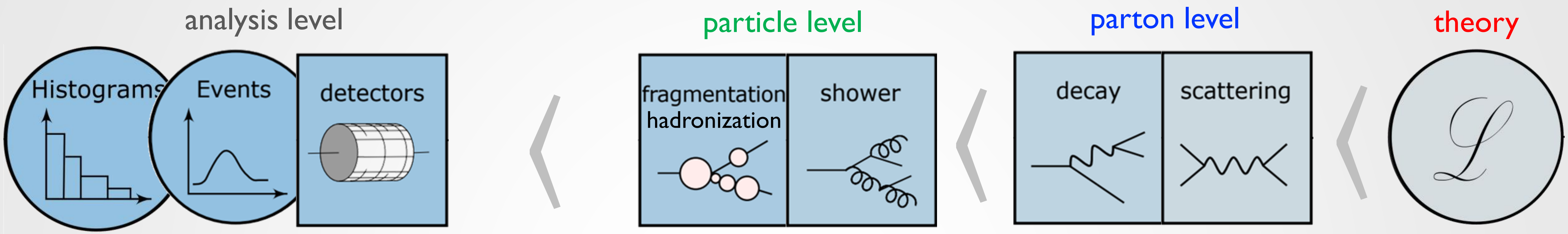
MACHINE LEARNING & UNCONVENTIONAL APPROACHES TO EFT

R. Schöfbeck (HEPHY Vienna, FNAL), July 12th, 2023, on behalf of the ATLAS and CMS collaborations

A CONDITIONAL SEQUENCE

adapted from arXiv:2211.01421



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

ML facilitates this inversion by exploiting that simulation samples

$$x_{\rm det}, z_{\rm ptl}, z_{\rm p} \sim p(x_{\rm det}, z_{\rm ptl}, z_{\rm p}|\theta)$$

"Simulation based inference"
(Divide & Conquer)

1. Generators run in 'forward mode'

2. Pick up uncertainties

$$p(z_{\rm ptl}|z_{\rm p}, \nu_{\rm th.})$$

$$p(x_{\rm det}|z_{\rm ptl}, \nu_{\rm exp.})$$

$$\frac{1}{\sigma_\theta}\frac{d\sigma_\theta}{dz_p} = p(z_{\rm p}|\theta)$$

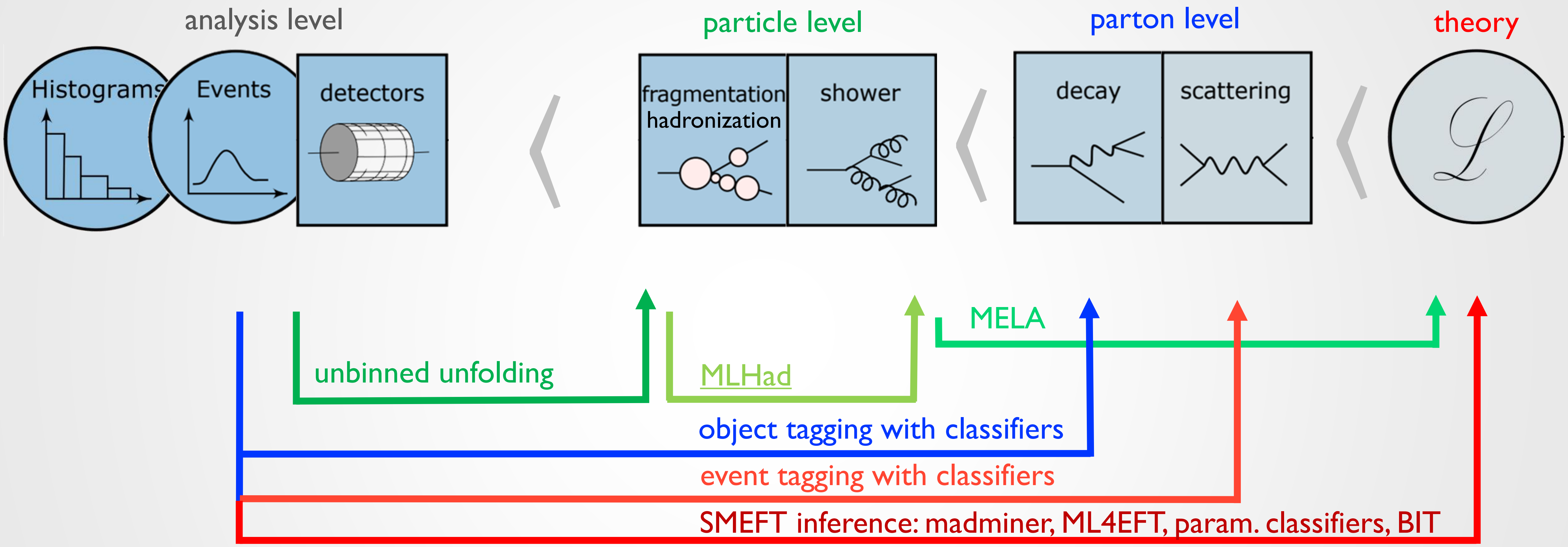
parton-level differential cross section
~ pdf

$\cancel{p(\theta)}$

θ NOT stochastic; Frequentist

DIVIDE & CONQUER

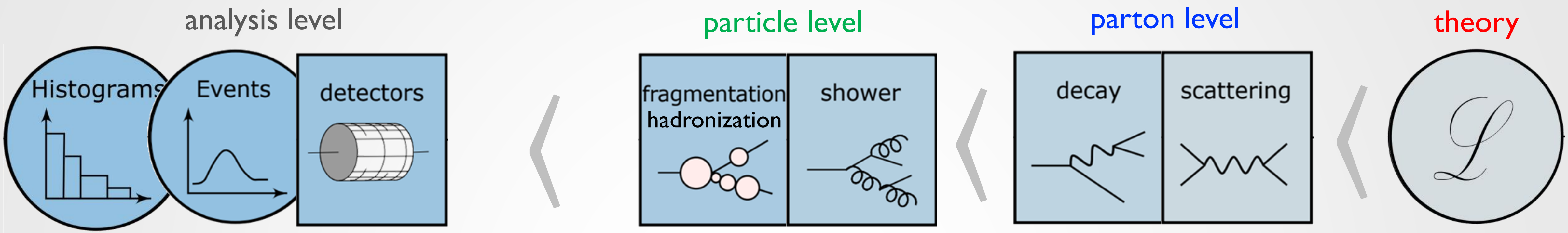
adapted from arXiv:2211.01421



all references in the backup

THE LIKELIHOOD RATIO TRICK

adapted from arXiv:2211.01421



Event classification $\boldsymbol{\theta} \rightarrow \text{isSig} \in \{0,1\}$

$$L = -\langle \log f(x) \rangle_{\text{isSig}=1} + \langle f(x) - 1 \rangle_{\text{isSig}=0} = \sum_{\text{isSig}} \int \mathrm{d}x \cdots$$

$$f^*(x) = \frac{p(x|\text{isSig}=1)}{p(x|\text{isSig}=0)}$$

Learn LR by classification;
“Likelihood ratio trick”
achieve NP optimality (for x-sec)

Outline:

1. How we use ML
2. What ML could do for SMEFT analyses
3. How analysis could look like at HL-LHC

THE STANDARD MODEL EFFECTIVE FIELD THEORY

- organizing principle: mass dimension

$$\mathcal{L}_{eff} = \mathcal{L}_{SM}^{(4)} + \sum \frac{C_x}{\Lambda^2} O_{6,x} + h.c.$$

  - Keep SM symmetries
    - $SU(3)_c \otimes SU(2)_L \otimes U(1)$
  - Keep particle content
  - BSM scale hierarchy
- operators affect all SM predictions
- 59 operators at d=6 [JHEP10(2010)085]

$$\frac{C_{\phi W}}{\Lambda^2} \left(\phi^\dagger \phi\right) W_I^{\mu\nu} W_{\mu\nu}^I$$

known SM particles

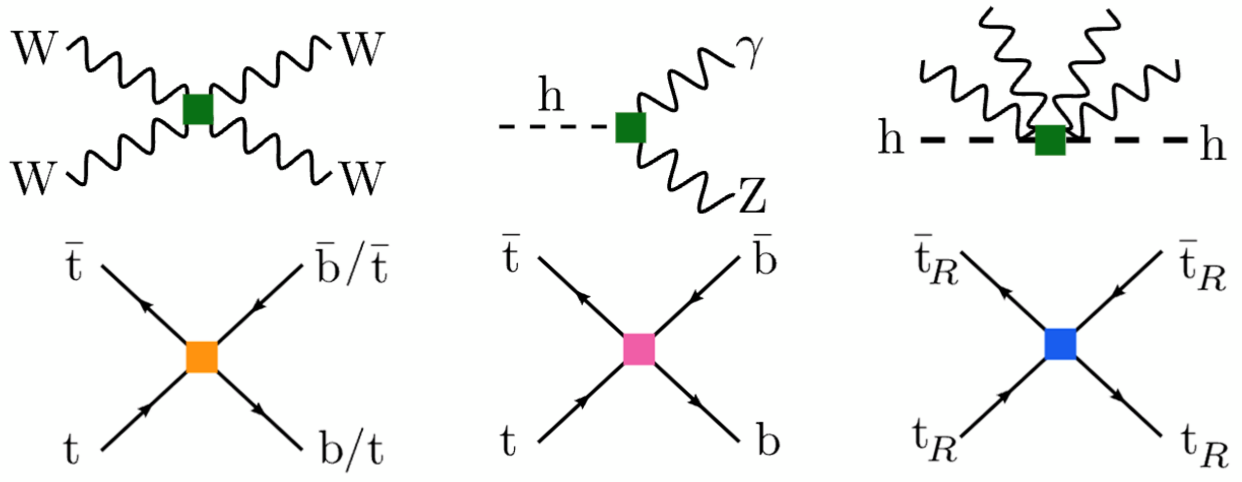
unknown coefficients

$$\frac{C_{qq}^{(8)}}{\Lambda^2} \left(\bar{q}\, \gamma^\mu T^A\, q\right) \left(\bar{q}\, \gamma_\mu T^A\, q\right)$$

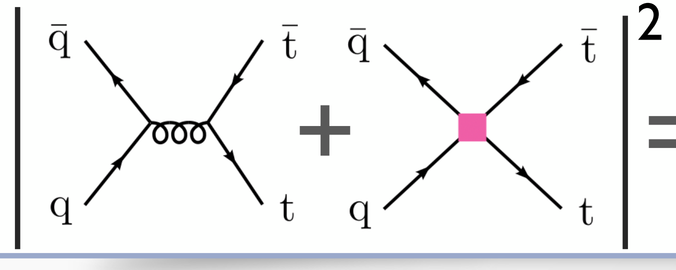
$$\frac{C_{qq}^{(3)}}{\Lambda^2} \left(\bar{q}\, \gamma^\mu \tau^I\, q\right) \left(\bar{q}\, \gamma_\mu \tau^I\, q\right)$$

known SM symmetries

Anomalous couplings & new interactions (tiny selection!)



- Predicting rates from "squared" diagrams:

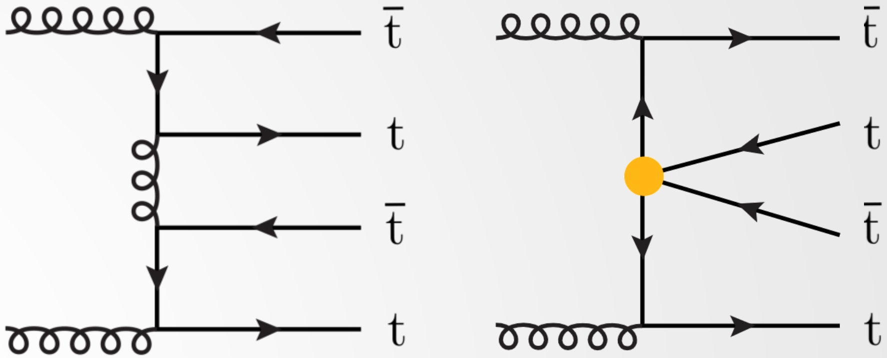


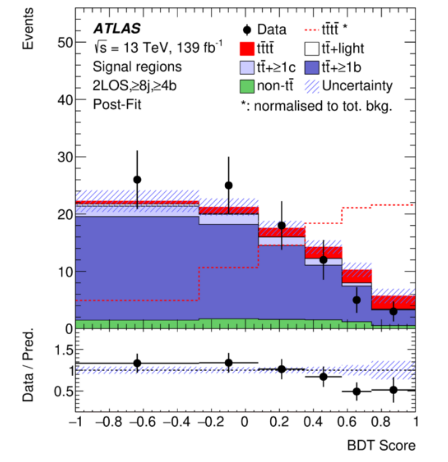
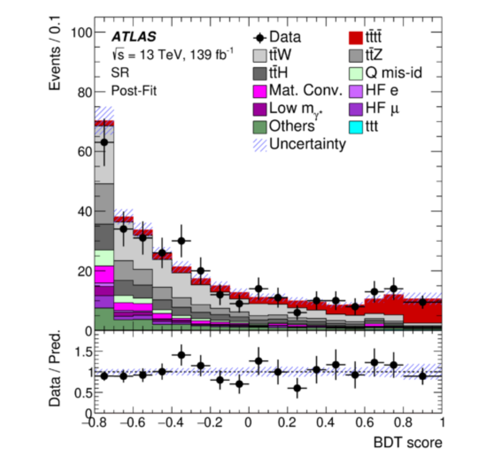
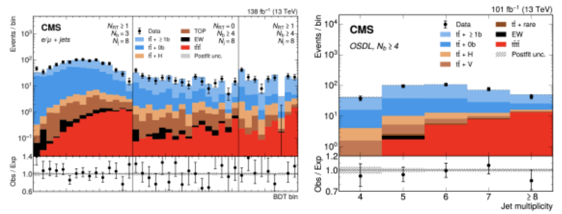
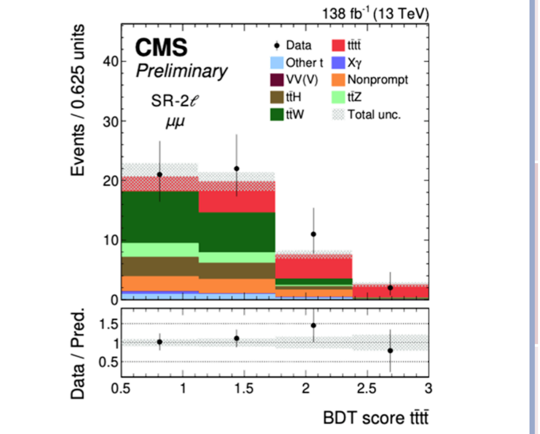
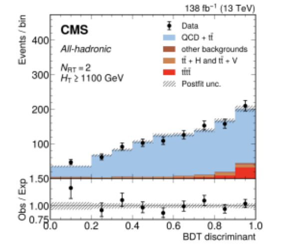
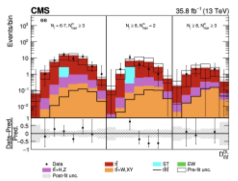
$$= \sigma^{SM} + \frac{C}{\Lambda^2}\sigma^{int} + \frac{C^2}{\Lambda^4}\sigma^{quad}$$

- Quite exceptional simplification! [LHC EFT prediction note]
- Being general & keeping SM symmetries: ask big questions!

FOUR TOP QUARK PRODUCTION

- ATLAS and CMS measure tttt in all decay channels – 0ℓ to 4ℓ
- Statistically limited: σ(SM) = 12.0+2.2-2.5 fb
  - most sensitive channel: 2ℓ with a same charge lepton pair
- Event-level BDTs, so far, are the workhorse classifiers



ATLAS EXPERIMENT	0ℓ	1ℓ	2ℓ(OS)	2ℓ(SS)	3ℓ+	CMS	0ℓ	1ℓ	2ℓ(OS)	2ℓ(SS)	3ℓ+
		JHEP 11 (2021) 118		EPJC 80 (2020) 1085			TOP-21-005			TOP-22-006	
2018						2018					
2017						2017		JHEP 11 (2019) 082 			
2016						2016					

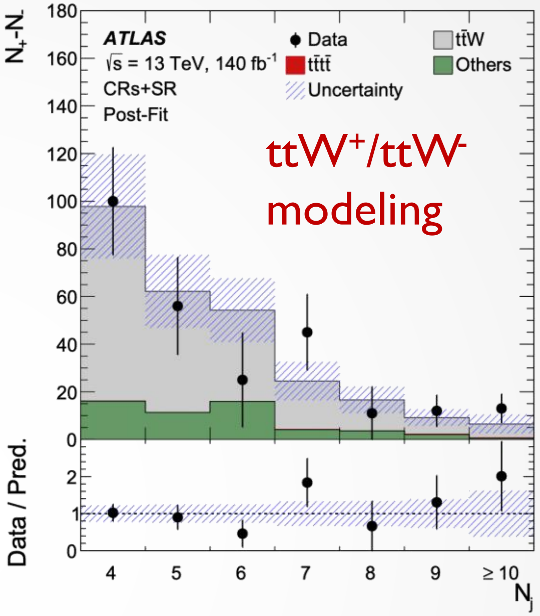
μ(tttt) = 2.0 (+0.8-0.6), 4.7σ

σ(tttt)=17.9 ±3.6 ± 2.5, >5σ

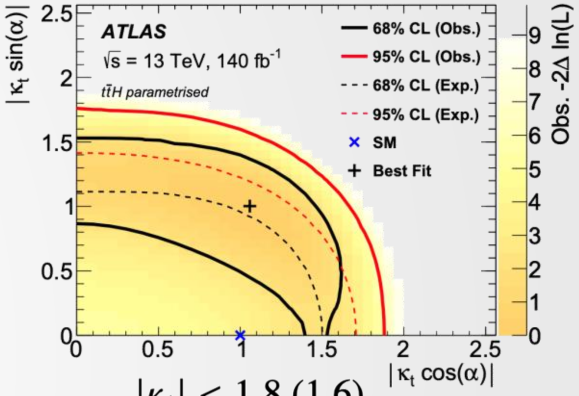
FOUR TOP QUARKS WITH A GNN: ATLAS

[2303.13937]

- New ATLAS result combining 2ℓSS and ≥3ℓ channels
  - Better ttW background estimation procedure based on charge-dependent N_{jet} scaling patterns
  - Separate treatment of 3t, tttW, tttq
  - Lower jet (≥20 GeV) and lepton (≥ 15 GeV) p_T cuts
- Graph-NN discriminant [GRAPH_NETS]
  - Edge-Convolution layers exploit multi-jet correlation
  - Leptons, E_T^{miss}, variable-length jet system
  - *classifier* training for cross-section measurement – LLR trick



Constraints on $\varkappa_t$ and CP mixing angle α



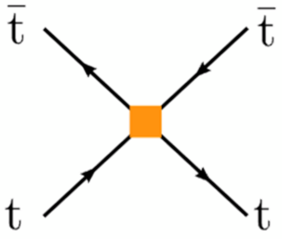
$|\kappa_t| < 1.8\ (1.6)$

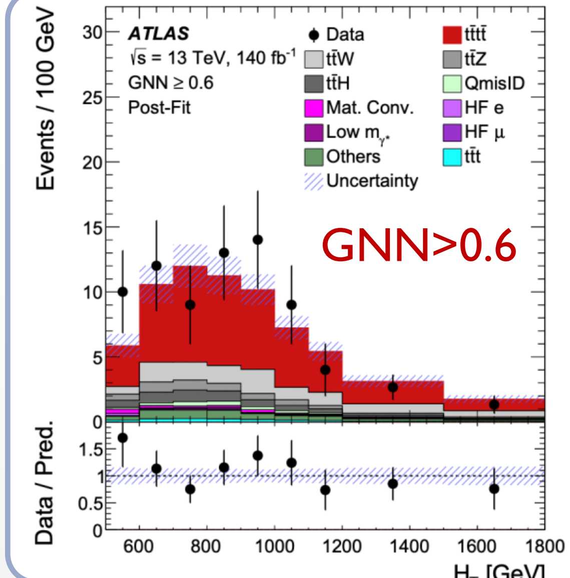
- Result $$\sigma_{t\bar{t}t\bar{t}} = 22.5^{+4.7}_{-4.3}(\text{stat})\ ^{+4.6}_{-3.4}(\text{syst})\ \text{fb} = 22.5\ ^{+6.6}_{-5.5}\ \text{fb}$$

$$\mu = 1.9 \pm 0.4(\text{stat})\ ^{+0.7}_{-0.4}(\text{syst}) = 1.9\ ^{+0.8}_{-0.5}$$

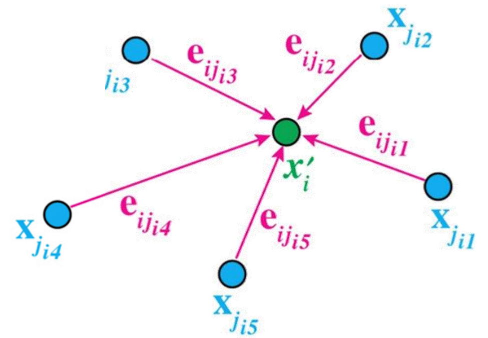
6.1σ (4.3σ) expected, consistent with SM at 1.8σ

Operators	Expected C_i/Λ^2 [TeV $^{-2}$]	Observed C_i/Λ^2 [TeV $^{-2}$]
O^1_{QQ}	[-2.4, 3.0]	[-3.5, 4.1]
O^1_{Qt}	[-2.5, 2.0]	[-3.5, 3.0]
O^1_{tt}	[-1.1, 1.3]	[-1.7, 1.9]
O^8_{Qt}	[-4.2, 4.8]	[-6.2, 6.9]





Edge convolution [1801.07829]



Amenable for you problem's symmetry!!

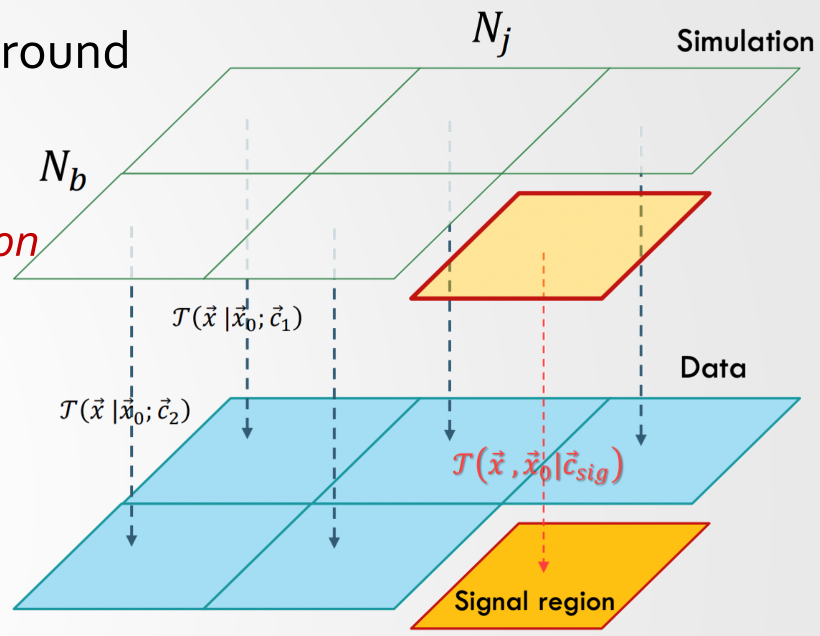
CALIBRATE BDT SHAPE WITH ABCDNN

[illustrations]
[ABCDnn]
[TOP-21-005]

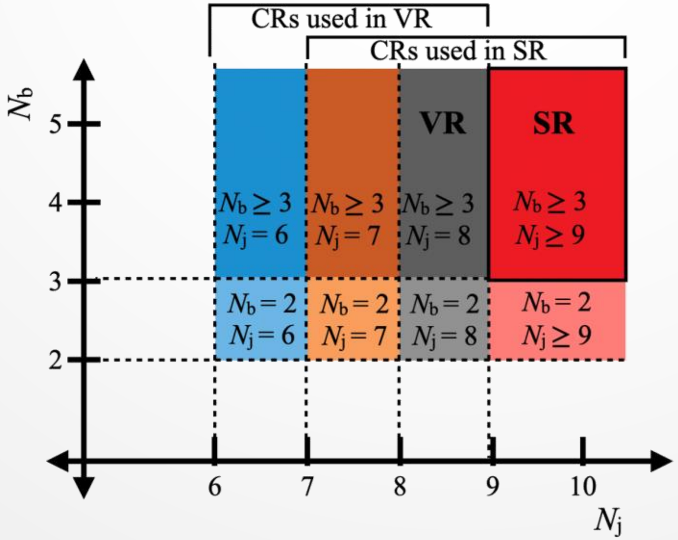
- CMS : BDT classifier from 20 features for all-hadronic four-top background
- Corrects BDT shape using [ABCDnn]: Neural autoregressive flow
  - Learn a invertible transformation of H_T/BDT shape from *data* to *simulation* *conditioned on a region c*

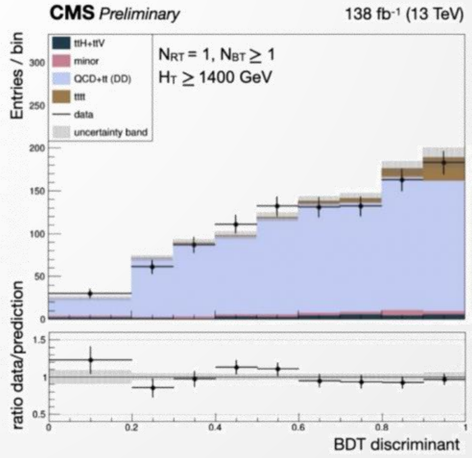
$$\int \mathcal{T}(\vec{x},\vec{x}_0|\vec{c})\, f_{src}(\vec{x}_0)\, d\vec{x}_0 = f_{target}(\vec{x}|\vec{c})$$

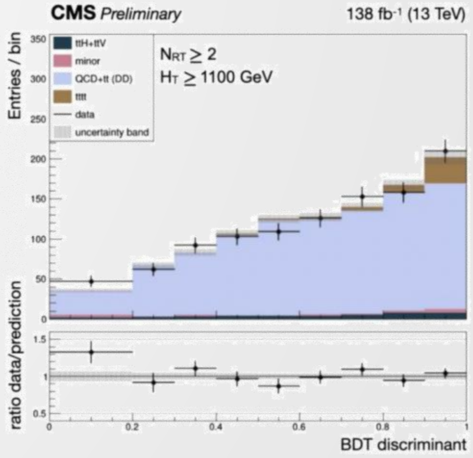
  - Technically, a DNN predicts the parameters of a bijective mapping
    - Encoding of indexed region is DNN input → extrapolate to SR



- NN version of traditional ABCD method
- Validation region between SR and CRs (N_{jet}=8)



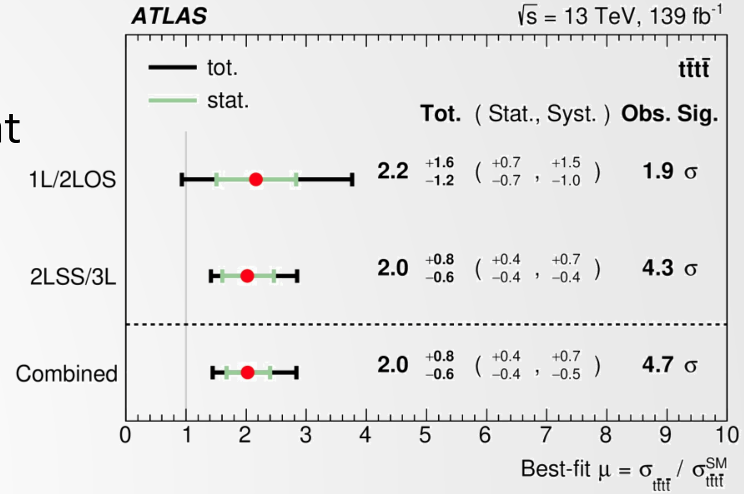




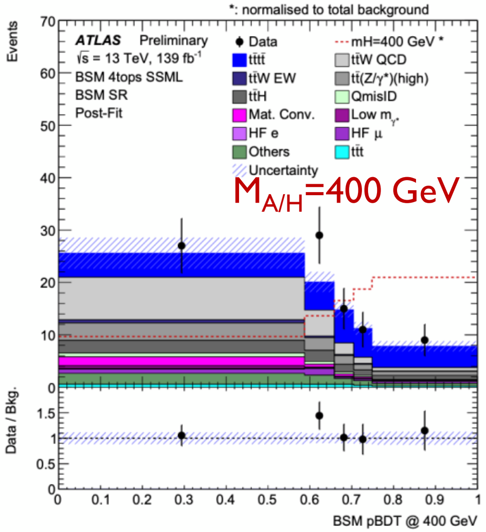
A PARAMETRIZED CLASSIFIER IN TT+(H/A→TT)

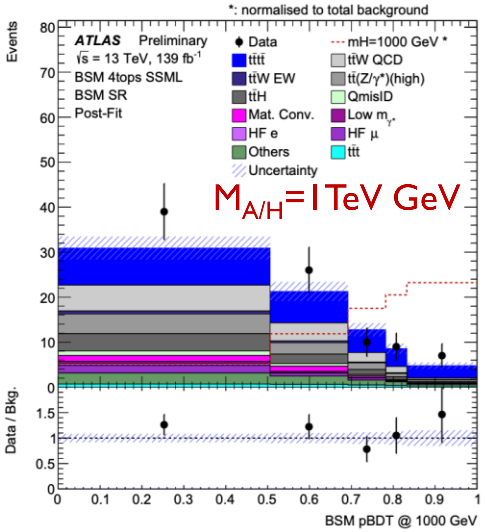
[arXiv:2211.01136]

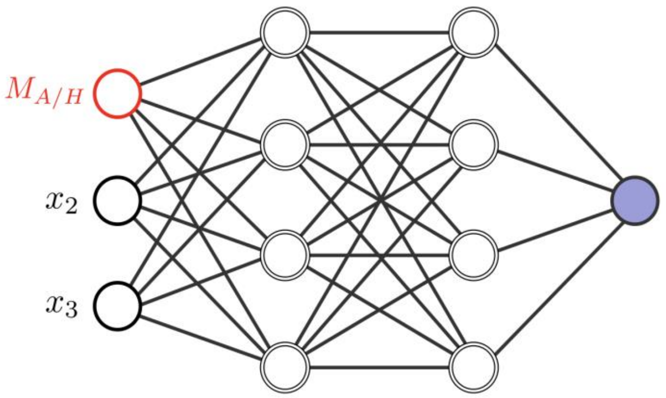
- The 2HDM model as a function of $M_{A/H}$ predicts resonant 4t production
  1. Use the signal region from the ATLAS 2ℓSS /≤3l 4t cross section measurement
  2. Train "parametrized" multi-variate discriminate as a function of $M_{A/H}$
  - example of a one-parameter "parametrized classifier"

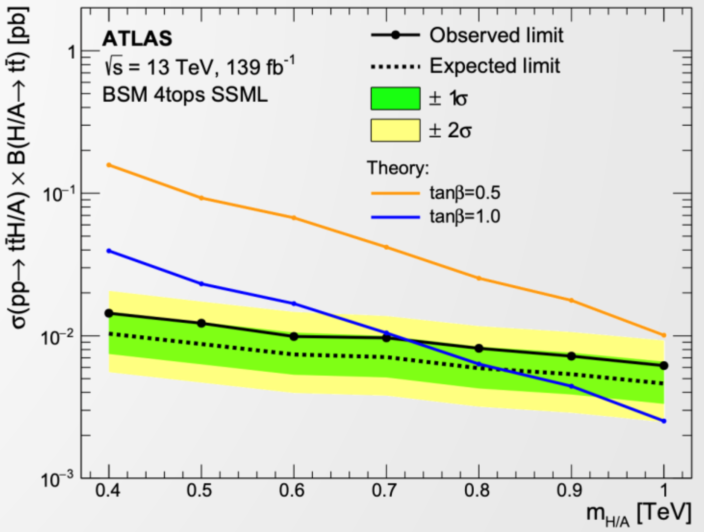


same data, different "test statistic" for each *mass* hypothesis:







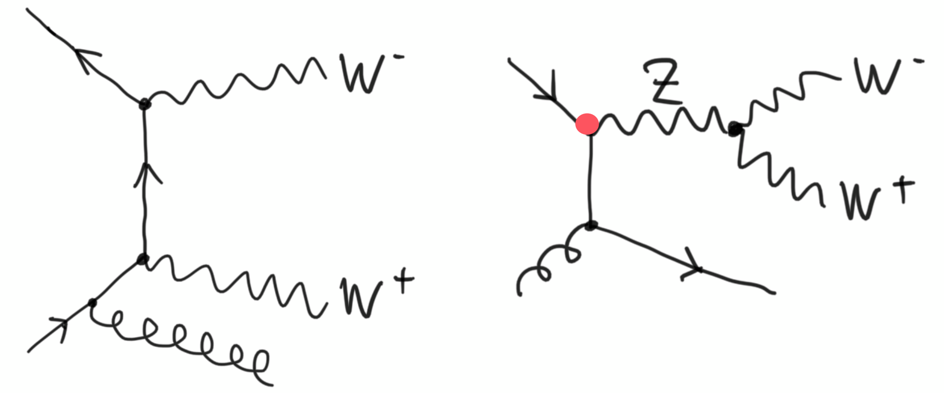


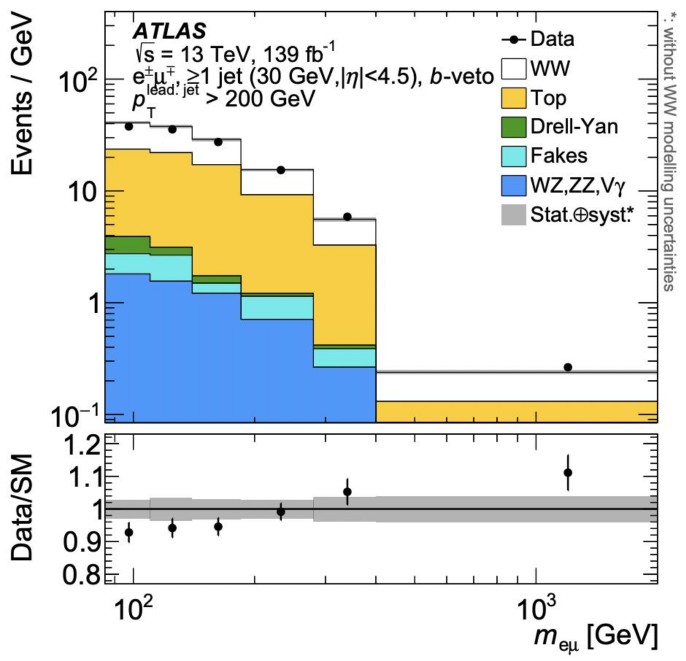
- Can use a similar technique for high-dimensional EFT measurements?

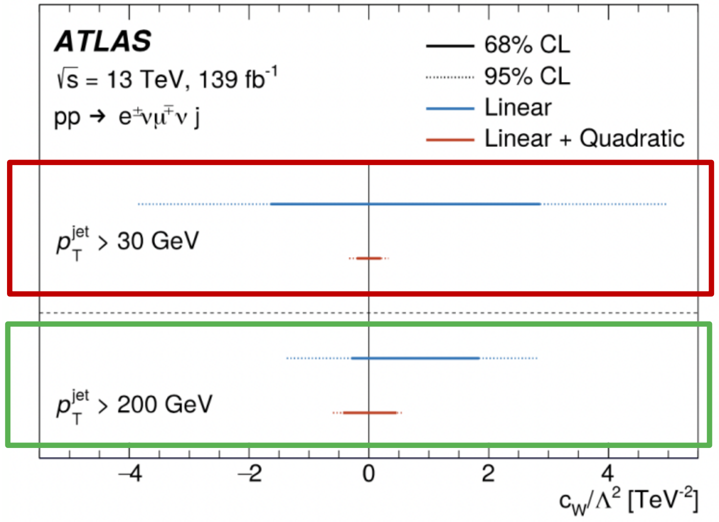
ENHANCING LINEAR SMEFT SENSITIVITY

JHEP 06 (2021) 003

- Linear dim6 term is the only unambigous contribution
- Consider W^+W^- production in association ≥ 1 jet
  - $e\mu$ channel has negligible Drell-Yan background
  - Inclusive and differentiation measurements
  - 12 kinematic variables (lepton, jet, …) are measured







← Helicity suppression

← Recovery

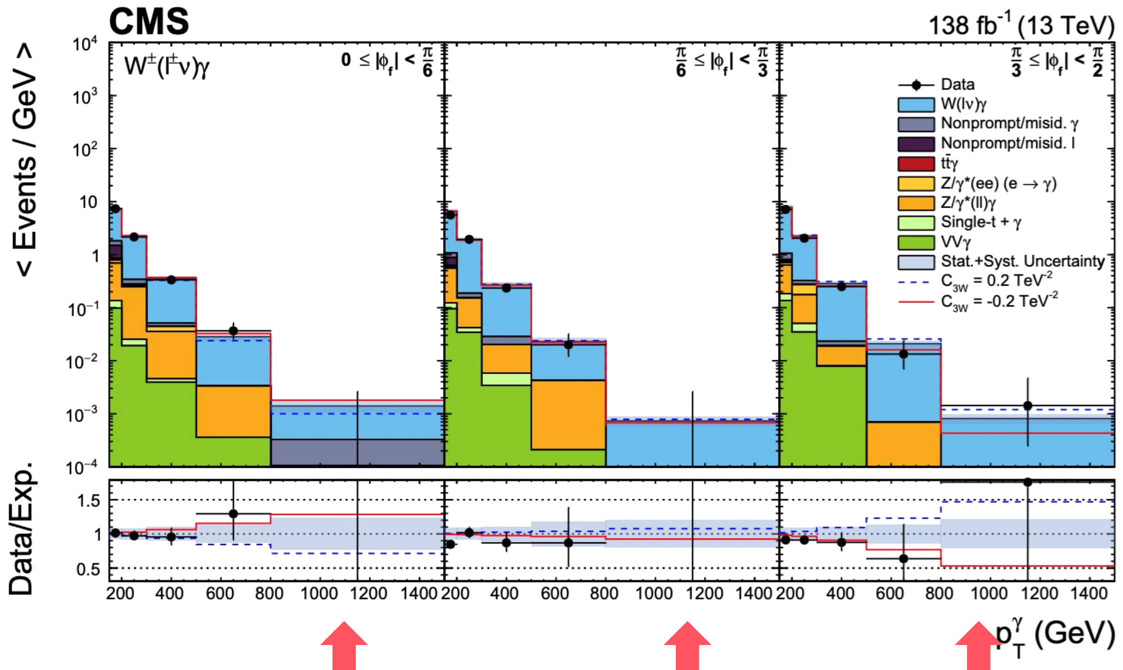
- Why the jet requirement? BSM interference cancels among helicities

$$\sigma = \sigma^{\mathrm{SM}} \underbrace{+ C_{3\mathrm{W}}\sigma^{\mathrm{int.}}}_{\text{Cancellations among helicities}} \underbrace{+ C_{3\mathrm{W}}^2\sigma^{\mathrm{BSM}}}_{\text{Same order as dim. 8}}$$

- hard jet (p_T> 200 GeV) requirement changes helicity composition

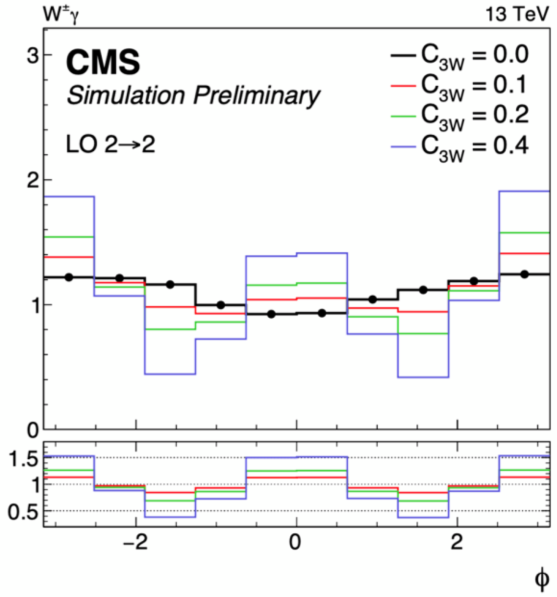
LINEAR SMEFT SENSITIVITY IN Wɣ PRODUCTION

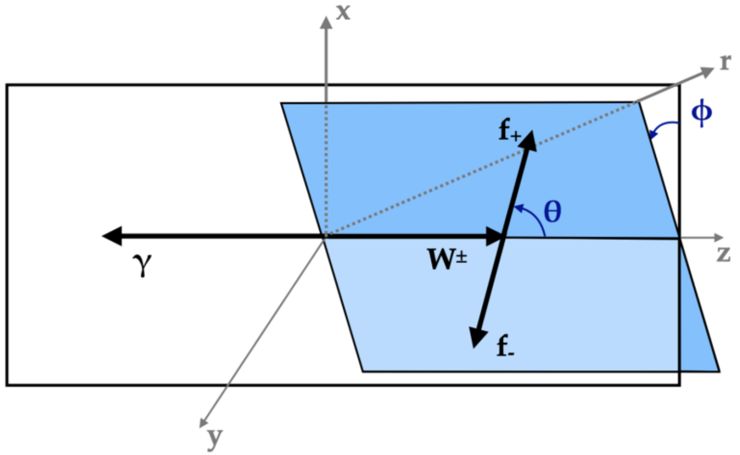
SMP-20-005



CP-even CP-odd CP-even

Interference resurrection arXiv:1708.07823





CP-even: $\cos(2\phi) \leftrightarrow \mathcal{O}_W$

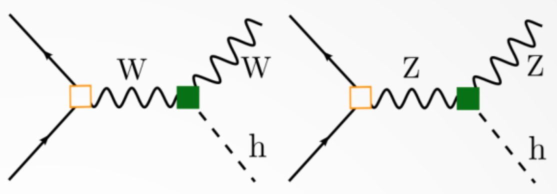
CP-odd : $\sin(2\phi) \leftrightarrow \mathcal{O}_{\widetilde{W}}$

- Boosting to the diboson center-of-mass frame allows to reconstruct decay plan angle ϕ
- It's distribution carries information on BSM effects in the $W_{L/R}$ helicites.
- Binning $p_T(\gamma)$ in ϕ recovers CP structure; facto 5-10: $-0.062 < C_{3W}/\Lambda^2 < 0.053$ TeV^{-2} $\rightarrow \Lambda_{BSM} \sim 5$ TeV

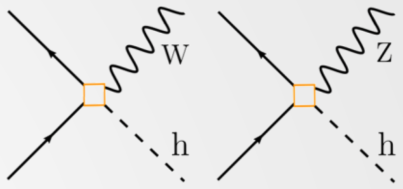
GOALS FOR MACHINE-LEARNING *OF* EFT

[EPJC 81 (2021) 178]

- SMEFT effects can be
  1. in the tails of the distributions because, e.g. 4-point functions grow with energy
  2. in angular observables & correlations, sometimes encoding CP-violating effects
     - "interference resurrection" PLB 2017 11 086
       "method of moments" JHEP 06 (2021) 031
     - Enhance / single out the linear term
       - Up to triple-angular correlations, x5-10 boost in sensitivity
  3. on top of "kinematically complex" backgrounds
     - Def: Usually amenable to classification MVAs
     - Unify the training target with classification
- What happens if we classify SMEFT vs. SM?



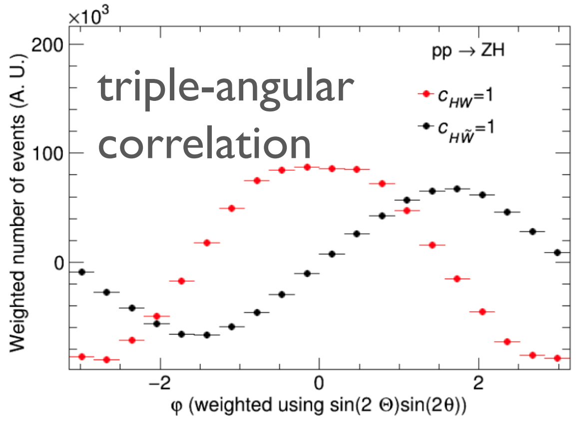
cHW/cHWtil

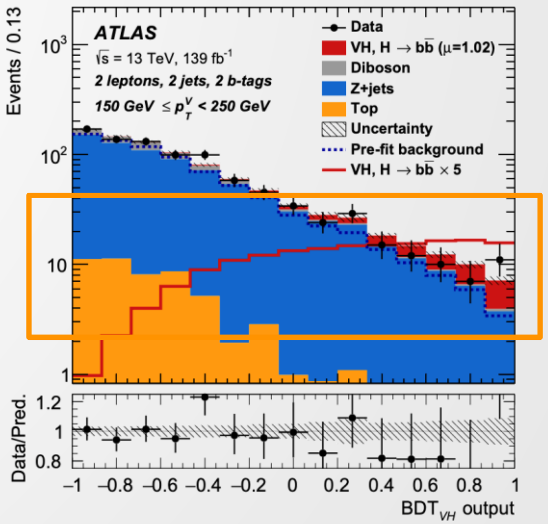


cHQ3 ↔ also in ttZ

Tree-level SMEFT amplitude of ZH (transverse polarisation):

$$\hat{\mathcal{M}}_\sigma^{\lambda=\pm} = g_Z m_Z\sqrt{\hat s}\left[\frac{g_{Z\sigma}}{\hat s - m_Z^2} + c_{\theta_W}\left(1+\frac{\hat s - m_h^2}{m_Z^2}\right)\left(\frac{g_{Z\sigma}c_{\theta_W}}{\hat s - m_Z^2}+\frac{Q_q e s_{\theta_W}}{\hat s}\right)\frac{v^2}{\Lambda^2}C_{HW} - \frac{2i\lambda k\sqrt{\hat s}}{m_Z^2}c_{\theta_W}\left(\frac{g_{Z\sigma}c_{\theta_W}}{\hat s - m_Z^2}+\frac{Q_q e s_{\theta_W}}{\hat s}\right)\frac{v^2}{\Lambda^2}C_{H\widetilde{W}}\right] + g_Z^2\frac{\sqrt{\hat s}}{m_Z}T_q^{(3)}\frac{v^2}{\Lambda^2}C_{HQ^{(3)}},$$

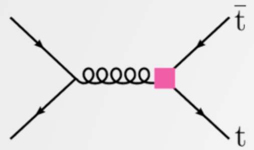




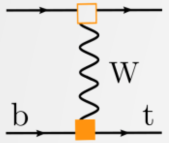
TOP QUARK PAIR + Z BOSON

JHEP 12 (2021) 083

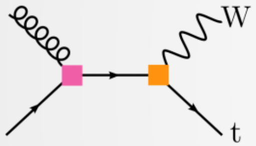
tt
811 pb



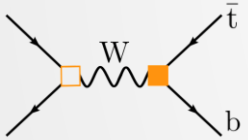
t (t-channel)
217 pb



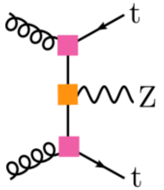
tW
72 pb



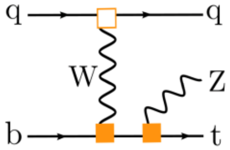
t (s-channel)
10 pb



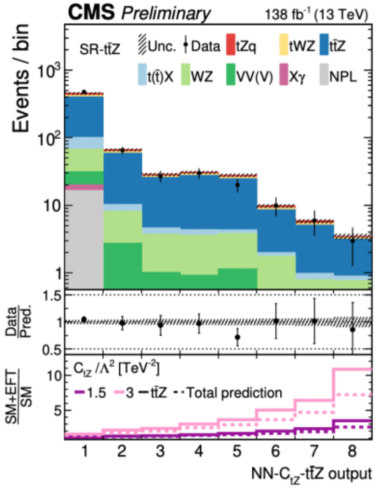
ttZ
1 pb

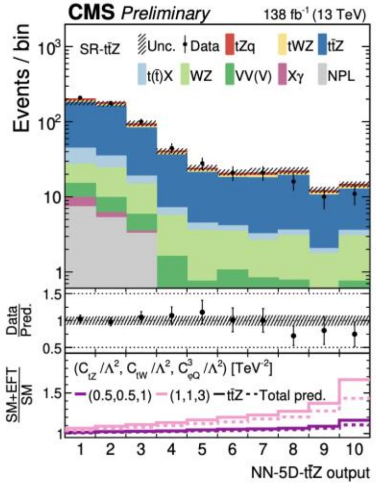


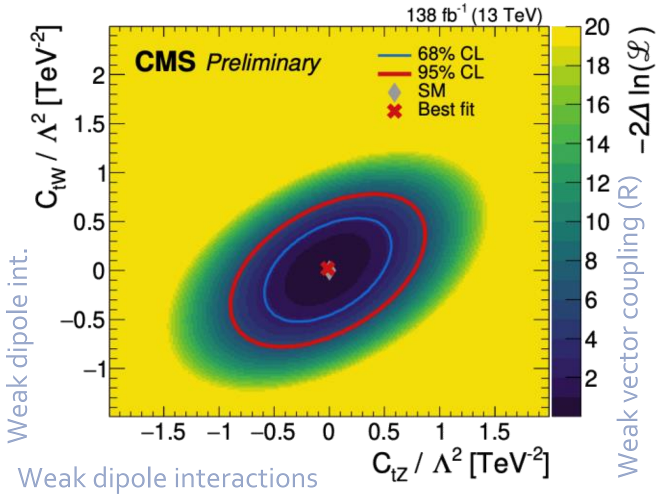
tZq
0.088 pb

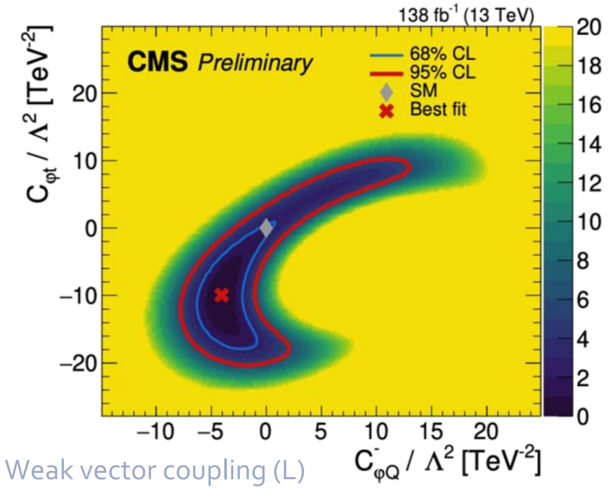


- Measure the top quark – Z boson coupling
- Train separate "SM vs. EFT" classifiers
  - Single operator O_{tZ}, O_{tW}, $O^3_{\phi Q}$
  - different trainings for different limits (!)
  - "likelihood trick" for SMEFT effects
- signal extraction with 1D, 2D, and 5D LL fit
  - Sampling of parameter space in the training
  - Targeted signals differ kinematically, but no parametrized training is used
  - Signal mix
  - no large linear terms → OK
- Best current limits









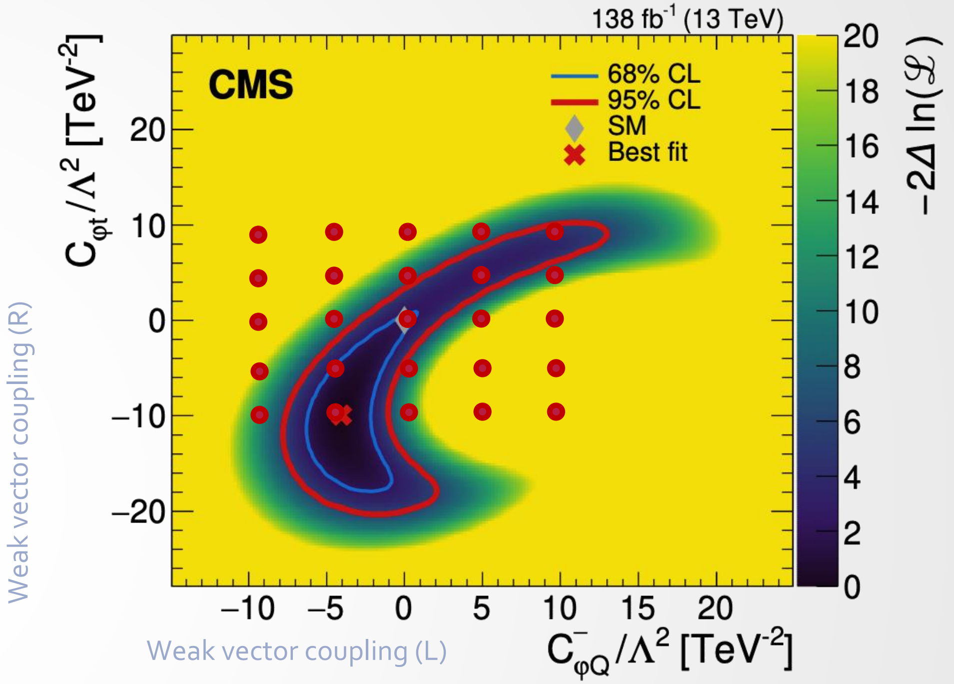
CAN WE JUST LEARN EFT EFFECTS "ON AVERAGE"?

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

$\boldsymbol{\theta}$ - *ignorant*

mixing signals &
case dependent mixes

$$f^*(\boldsymbol{x}) = \frac{1}{1+r_{\mathcal{B}}(\boldsymbol{x})} \qquad r_{\mathcal{B}}(\boldsymbol{x}) = \frac{\frac{1}{|\mathcal{B}|}\sum_{\boldsymbol{\theta}\in\mathcal{B}} p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\mathrm{SM})}$$



[TOP-21-001]

- Sending 'mixed signals' to the loss function
  - Averages the training data set - suboptimal when linear effects dominate
  - Classifier does not reflect knowledge on the $\boldsymbol{\theta}$-dependence
- Solution: Back to the drawing board & inject $\boldsymbol{\theta}$ polynomial SMEFT dependence in estimator.

SIMULATION BASED INFERENCE

[Madminer 1805.00020]
Full list of references in backup

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

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

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

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.

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

Available in simulation!
(MSE loss only for illustration)

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

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

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

Latent space is integrated in numerator and denominator

Wulzer et.al. [JHEP 05 (2021) 247]

RS et. al., [2107.10859]

PARAMETRIZED CLASSIFIERS: NETS & TREES

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

Make loss function aware of analytic SMEFT structure

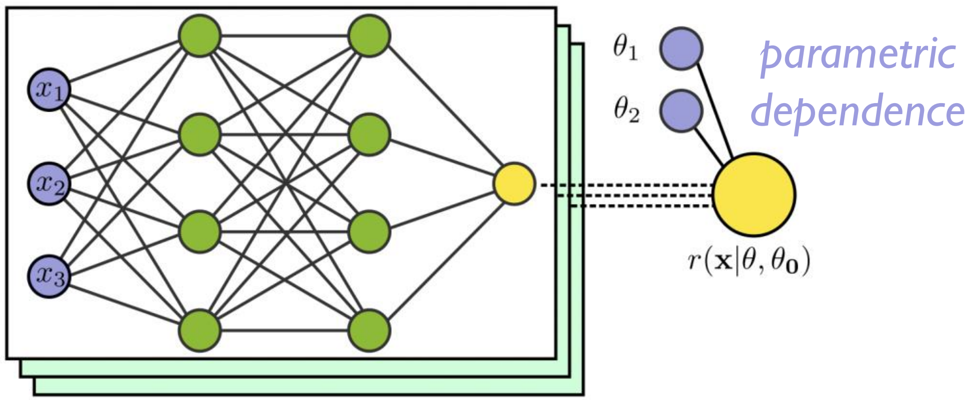
Invert likelihood trick
with positive polynomial of NN -outputs

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

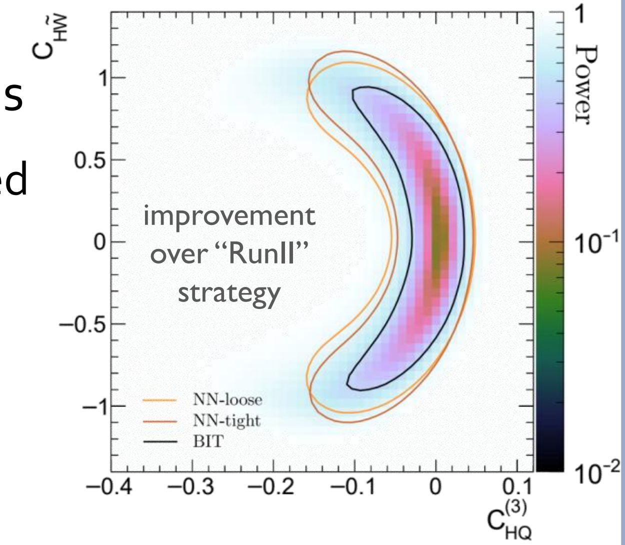
$$\hat{r}(\boldsymbol{x};\boldsymbol{\theta}) = \left(1 + \sum_a \theta_a \hat{n}_a(\boldsymbol{x})\right)^2 + \sum_a \left(\sum_{b \geq a} \theta_b \hat{n}_{ab}(\boldsymbol{x})\right)^2$$

Fit NNs simultaneously

inject new technology here

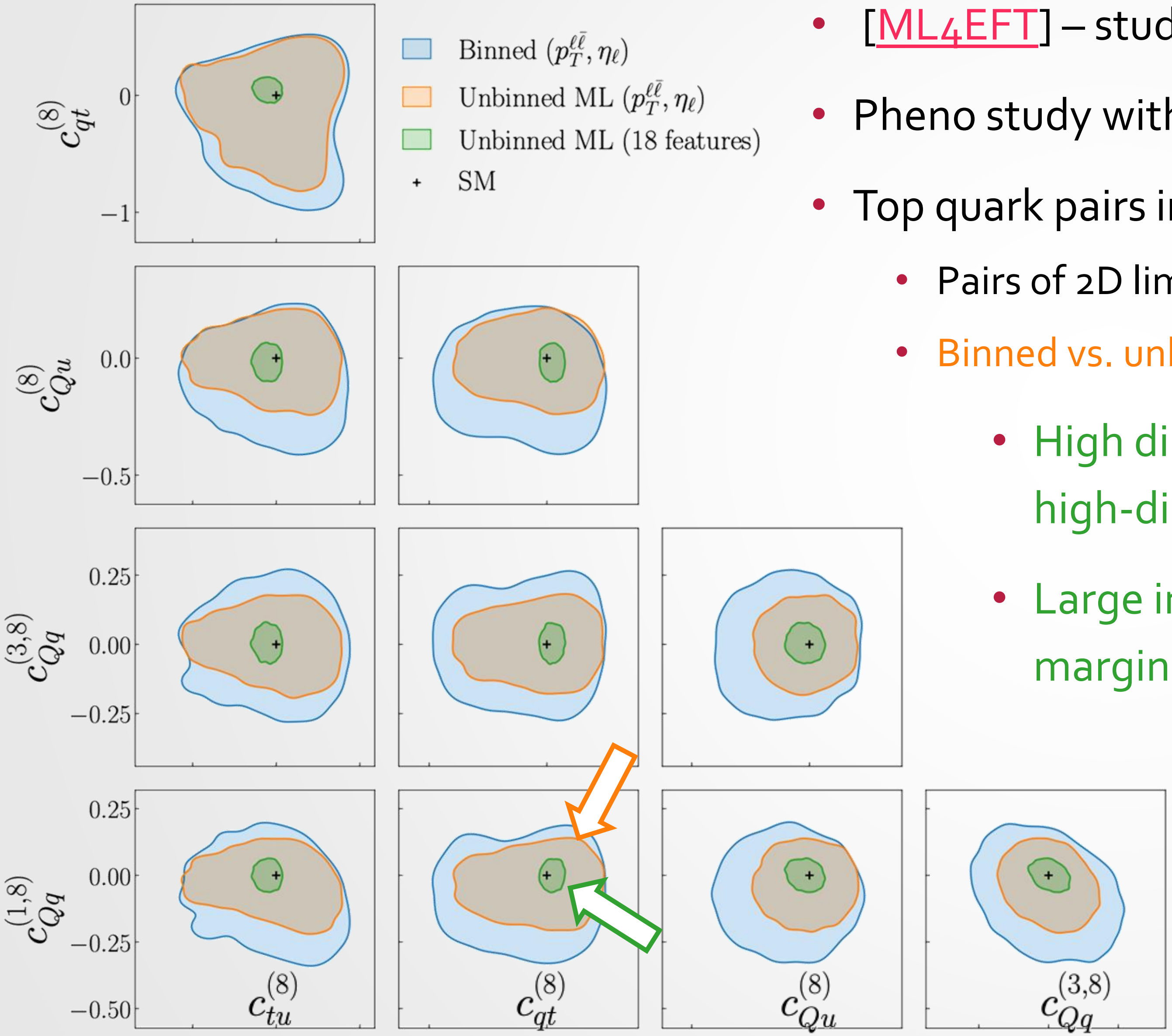


- Parametrized classifiers learn coefficients of the polynomial SMEFT dependence
- Statistical setup established by Madminer [1805.00013] [1805.00020] [1805.12244]
- Many variants, e.g., Boosted Information Tree [2205.12976]
- Used in *ongoing* analyses
  1. Benefits marginalized high-dimensional interpretations
  2. Should be done *unbinned*



- Is it important at all?

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 the marginalized limits

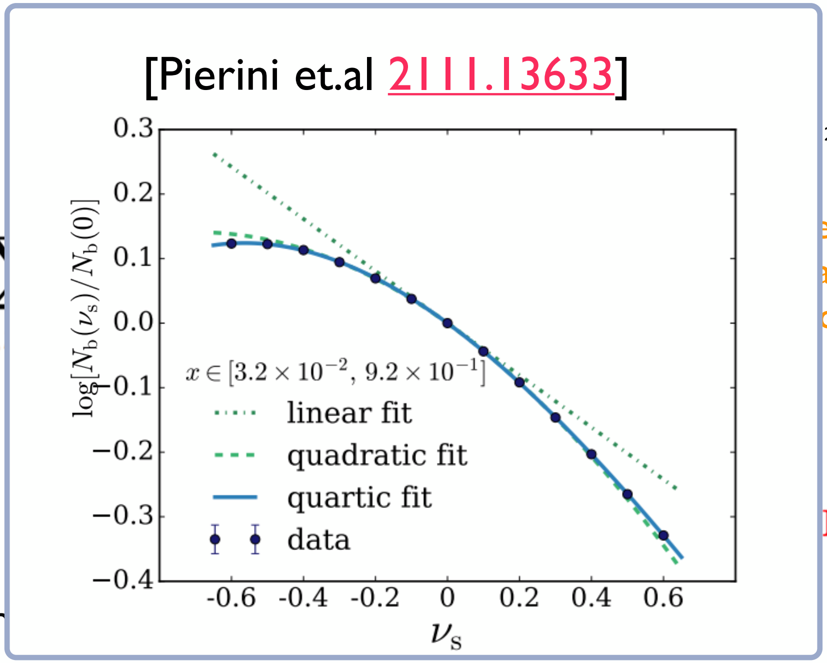
- Take seriously constraining power from SM candle
- Whether the sensitivity gain survives systematics in an unbinned detector-level analysis is an open question

TOWARDS UNBINNED ANALYSIS

- Binned parametrized classifiers are impractical for high SMEFT parameter dimension
- What's missing to go all-in? Systematics.
  -
  - ²

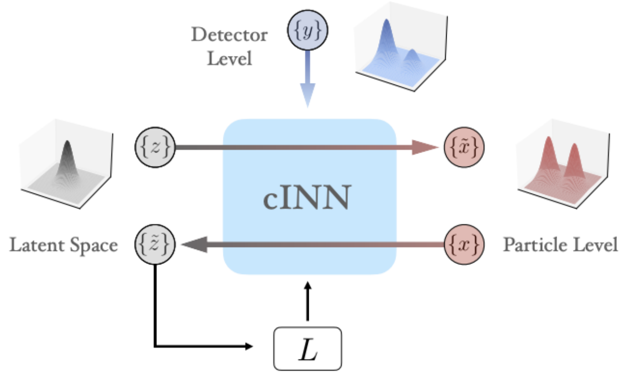
$L = \sum$ ematic
$\theta,$ ations
cheap!

M,ν=0



- A ch h event counts in the profiling
- Divide & conquer #1: Experiments begun machine-learning certain nuisances: h_{damp}, b-fragmentation

- Divide & conquer #2: Unbinned unfolding for high dimensions
- Consider on the conditional pdf $p(x_{det}|z_{ptl})$ which can be evaluated in the forward mode
- Unfolding algorithms use Bayes' theorem
  $$p(x_{det}|z_{ptl})p(z_{ptl}) = p(z_{ptl}|x_{det})p(x_{det})$$
  to learn $p(z_{ptl}|x_{det})$; GAN & other generative versions
  - Mostly iterative, to remove simulated prior



[community paper]
e.g. [OmniFold]
[cINN], [all]

- Report unbinned unfolded data; then SMEFT analysis

SUMMARY

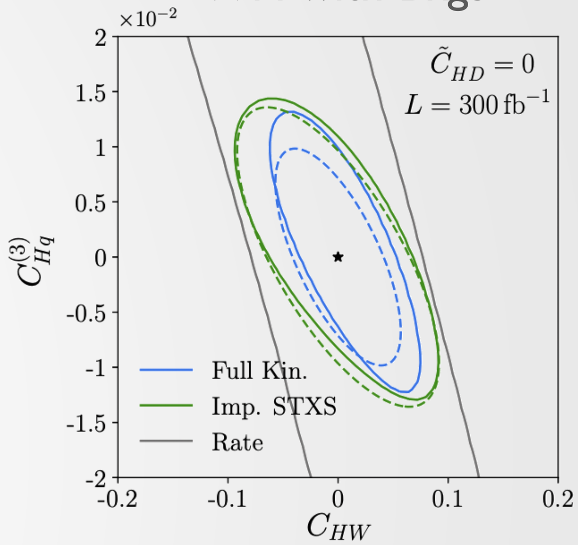
- EFT measurements are particularly high-dimensional: number of WC, nuisances, observables
  - Three assumptions: Keep symmetries, particle content, BSM mass gap
  - Can handle all-operator fits at linear dim6 with low-dimensional Poisson data O(~100-200 bins)
  - Can we / should we go beyond?
    - O(10) observations per final state, including dim6^2 & dim8 , many nuisances
- At HL-LHC we will need more ML-facilitated parametrization to support high-dimensional interpretations
  - HEP now profits from developments in adjacent fields: gNNs, normalizing flows, transformers, …
  - Parametrized SMEFT classifiers can capitalize on this technology and facilitate SMEFT analyses
- Not yet clear whether unbinned analyses will be part of the LHC's legacy
  - An incorrect dismissal is a costly mistake
- "The long term goal is that you believe the uncertainties"
  - Jure Zupan on Monday

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

→ talk later today