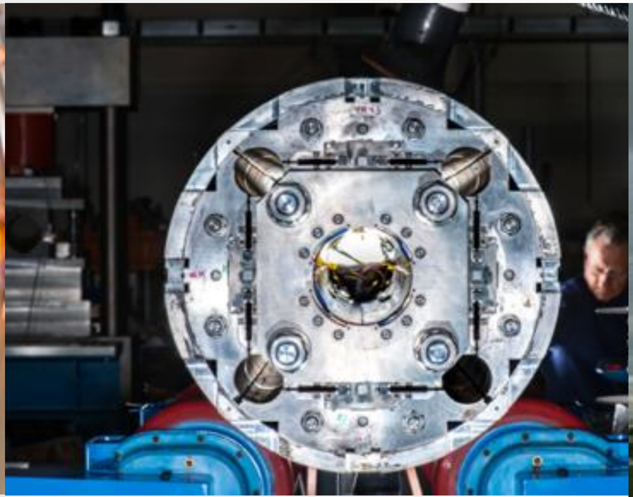
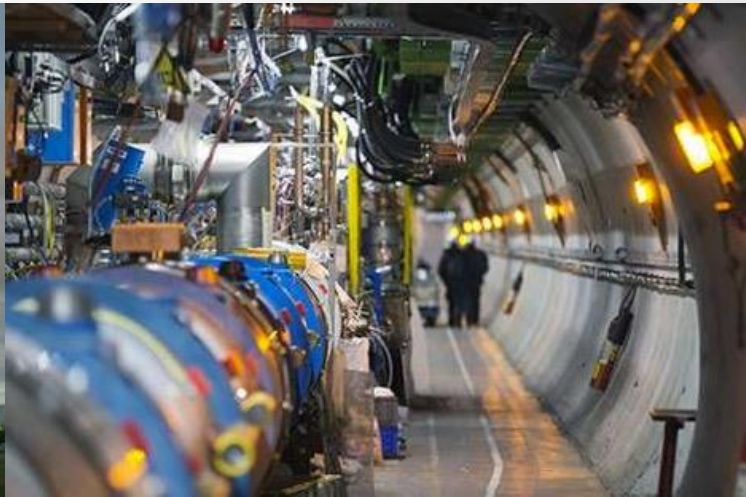
MACHINE-LEARNING & UNCONVENTIONAL APPROACHES TO EFFECTIVE FIELD THEORIES AT THE LHC

R. Schöfbeck (HEPHY Vienna, FNAL), Sept. 8th, 2023, Fermilab Wine and Cheese Seminar

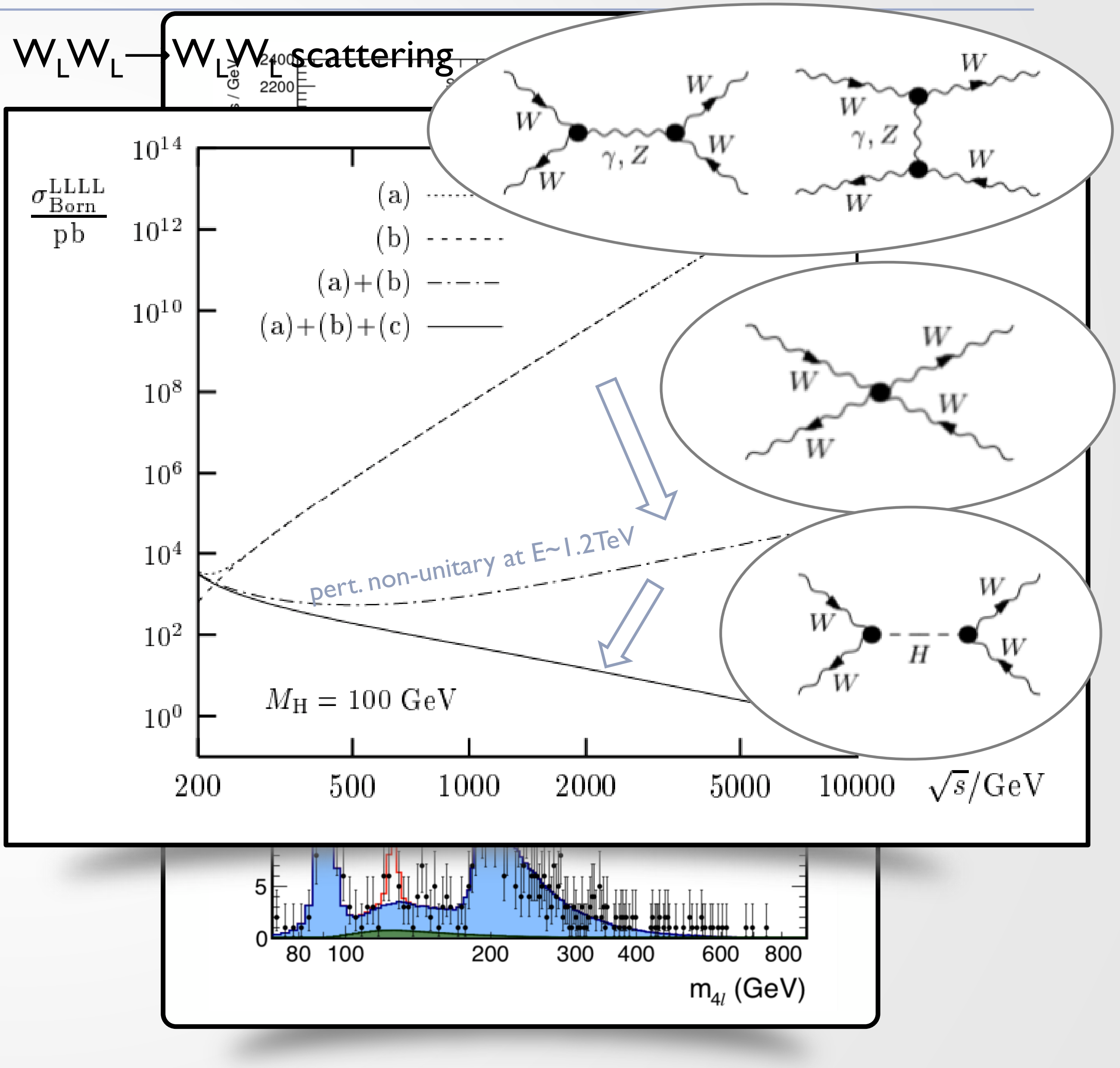


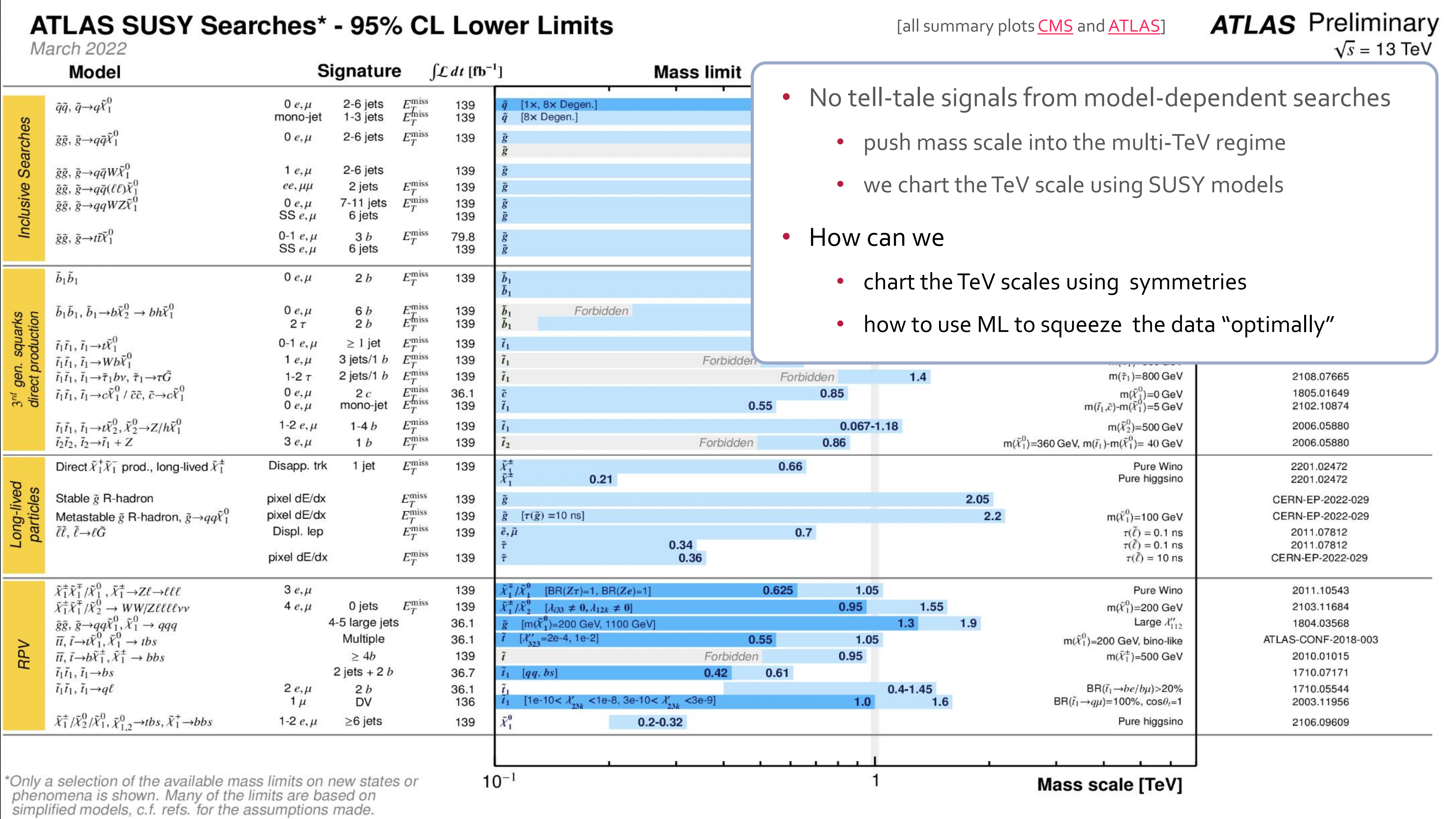
[Denner, Hahn ’97]

[Phil. Trans. R. Soc. A 373: 20140037]

A LONG ROAD

- Early LHC physics goals (1984!)
  - origin of electroweak symmetry breaking & SUSY
  - discover Higgs boson via gluon-induced top loop
    - top quark discovery (1995) in resonant production
- LHC no-loose theorem: $W_L W_L \longrightarrow W_L W_L$
  - resonant Higgs $\lesssim$ 1.2 TeV or strongly interacting BSM
- ATLAS/CMS (2012) Higgs boson discovery (125 GeV)
  - We understand *something* about the TeV scale
- What about other resonances?



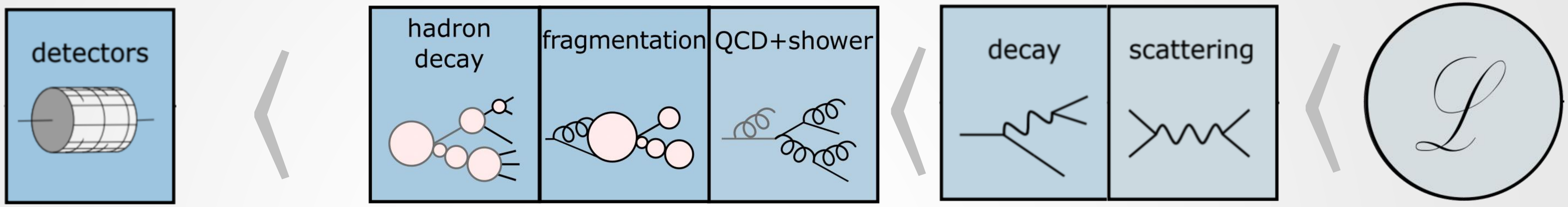


- No tell-tale signals from model-dependent searches
  - push mass scale into the multi-TeV regime
  - we chart the TeV scale using SUSY models
- How can we
  - chart the TeV scales using symmetries
  - how to use ML to squeeze the data “optimally”

A CONDITIONAL SEQUENCE

adapted from arXiv:2211.01421

analysis level | particle level | parton level | theory



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

Squeezing with Neyman-Pearson Lemma
Likelihood ratio is the optimal statistic

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

Would like to evaluate for varying **θ**, **ν**
→ *"Simulation based inference"*
(fit of a function)

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

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

parton-level differential cross section
~ pdf

$\cancel{p(\theta)}$

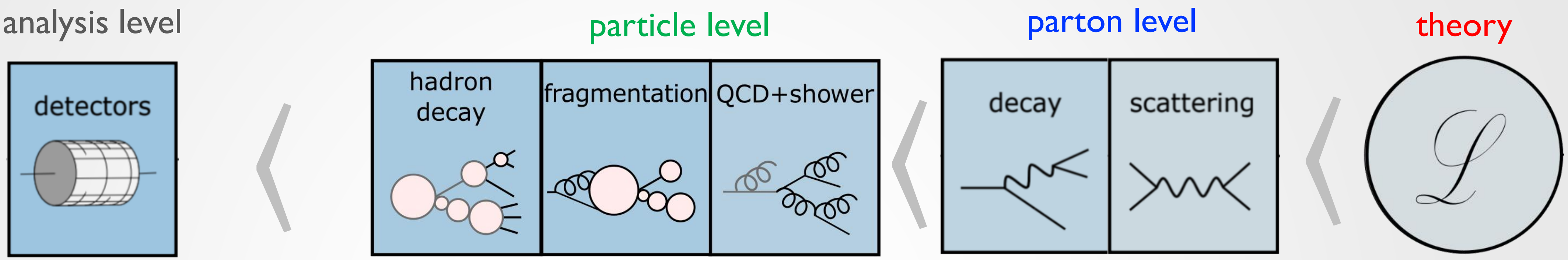
θ NOT stochastic; Frequentist

tagging: $z_p \sim \{0,1\}$

EFT: $\sim \mathbb{R}^{20}$ - $\mathbb{R}^{50}$

A CONDITIONAL SEQUENCE

adapted from arXiv:2211.01421



$$p(x|\mathrm{is}\tau = 1) = \int \mathrm{d}z_{\mathrm{ptl}}[\cdots]\; p(x_{\mathrm{det}}|z_{\mathrm{ptl}}) \qquad p(z_{\mathrm{ptl}}|z_{\mathrm{p}})$$

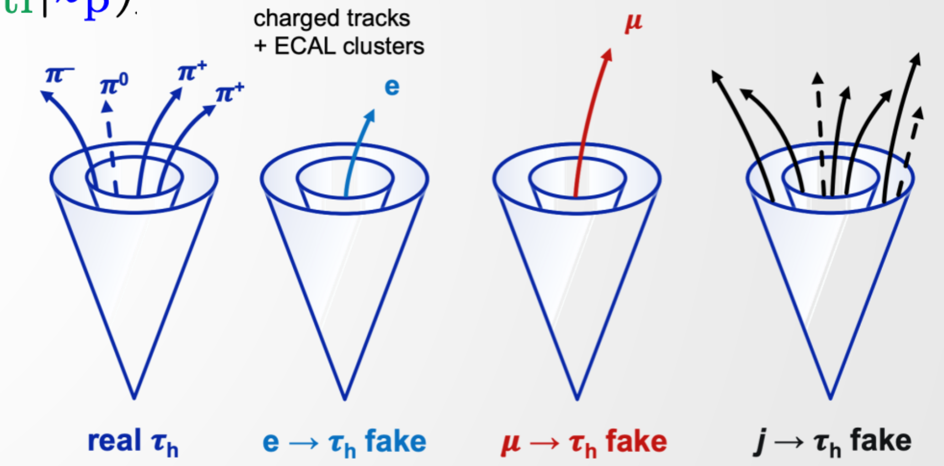
Divide & conquer: *Assume* some degree of factorization

Identify τ leptons: $z_{\mathrm{p}} \to \mathrm{is}\tau \in \{0,1\}$, $x_{\mathrm{det}} \to x \in \tau$ candidate

$$L = -\langle \log f(x) \rangle_{\mathrm{is}\tau=1} + \langle f(x) - 1 \rangle_{\mathrm{is}\tau \neq 1} = \sum_{\mathrm{is}\tau} \int \mathrm{d}x \cdots$$

$$f^*(x) = \frac{p(x|\mathrm{is}\tau = 1)}{p(x|\mathrm{is}\tau \neq 1)}$$

Learn LR by classification;
"Likelihood ratio trick"
achieve NP optimality;



seven τ decay modes

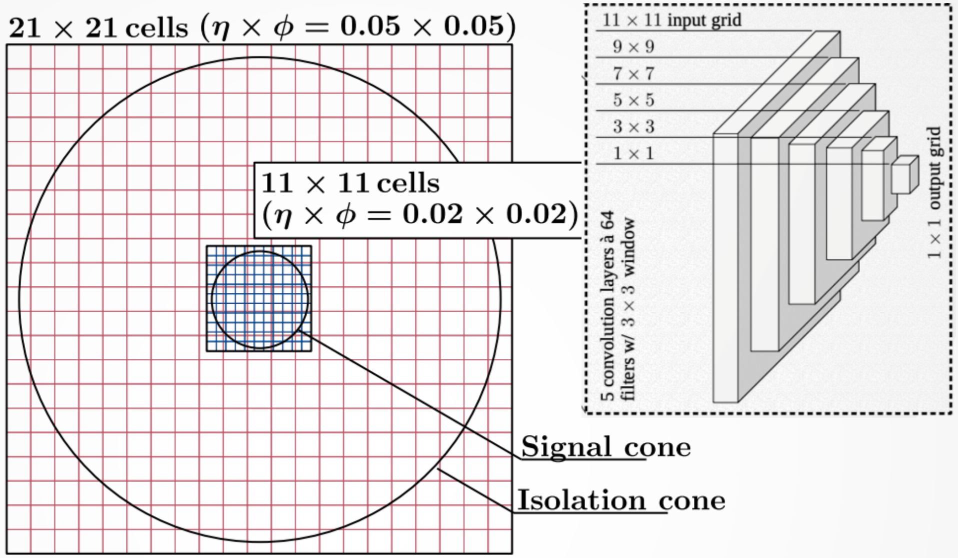
three main fake contributions

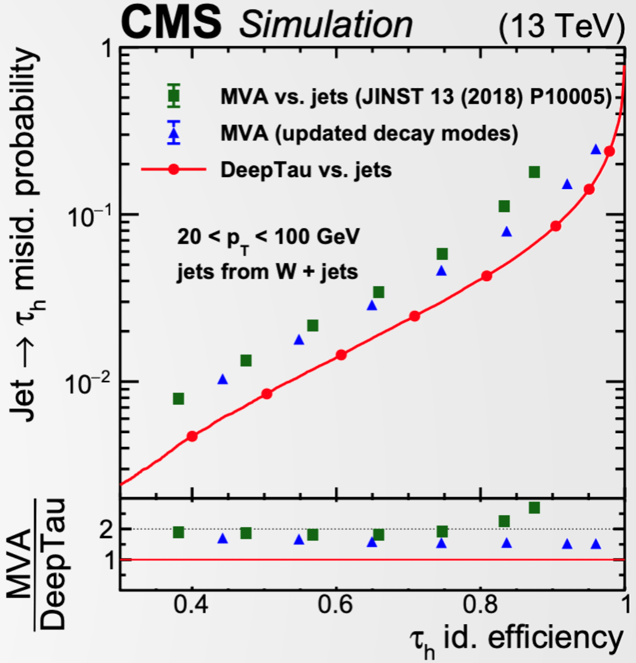
CMS [DeepTau] algorithm: identifies τ_h

DEEPTAU: τ IDENTIFICATION IN CMS

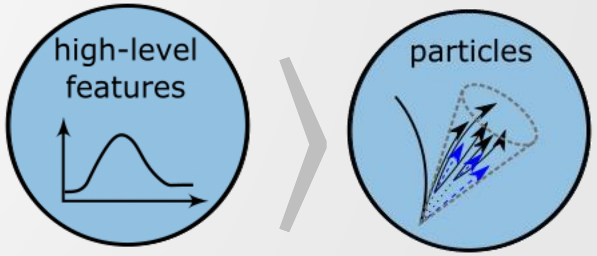
JINST 17 (2022) P07023

- Inputs:
  - High-level *cand-features* (DNN)
  - *Feature maps of constituents* on two grids, 106k inputs
- Convolutional layers
  - Read out particles on grid
  - Exploiting translation symmetry
  - fed into subsequent DNNs
- training: 1.1M parameters 140M τ, 690 hrs
- Result: Factor of two in background suppression from using *constituent-level inputs*



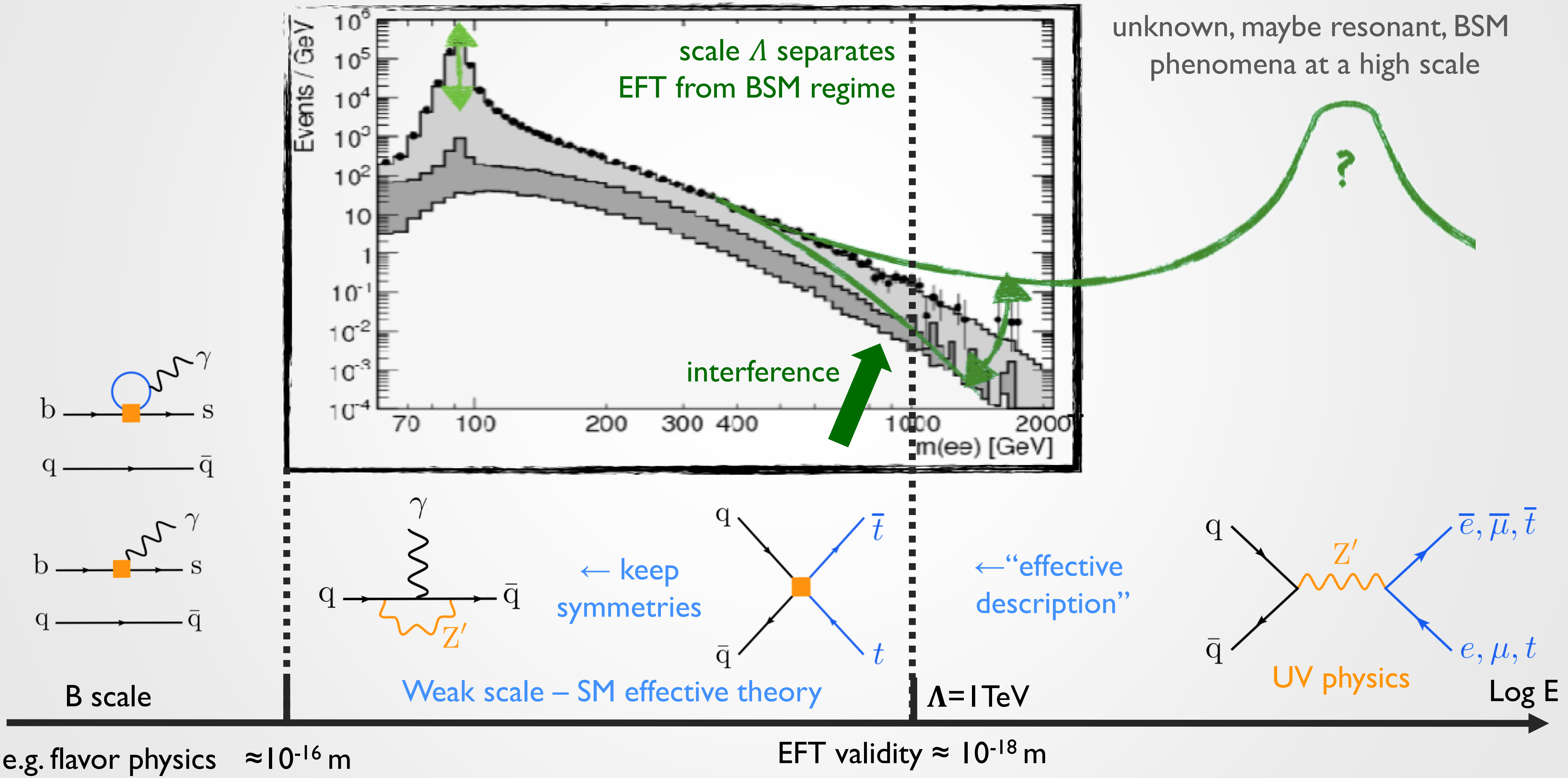


- Make low-level inputs approachable via symmetries
- Similarly, all other taggers: RNNs, graphNNs, transformers, ...
  - ATLAS τ ID [using RNNs], and constituent based [top taggers] using PF flow networks
  - CMS: [DeepJet] for g/c/b/uds/leptons identification, and [DeepAK8] for t/W/Z/H
- What about *theory landscape*?



GOING "LOW-LEVEL" IN THEORY LANDSCAPE

Sketch from F. Riva



THE STANDARD MODEL EFFECTIVE FIELD THEORY

- Organize the pieces in terms of mass dimension:

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

1. Keep SM symmetries
   - $SU(3)_c \otimes SU(2)_L \otimes U(1)$
2. Keep particle content
3. scale hierarchy

- 59 operators affect all SM predictions

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

known SM particles

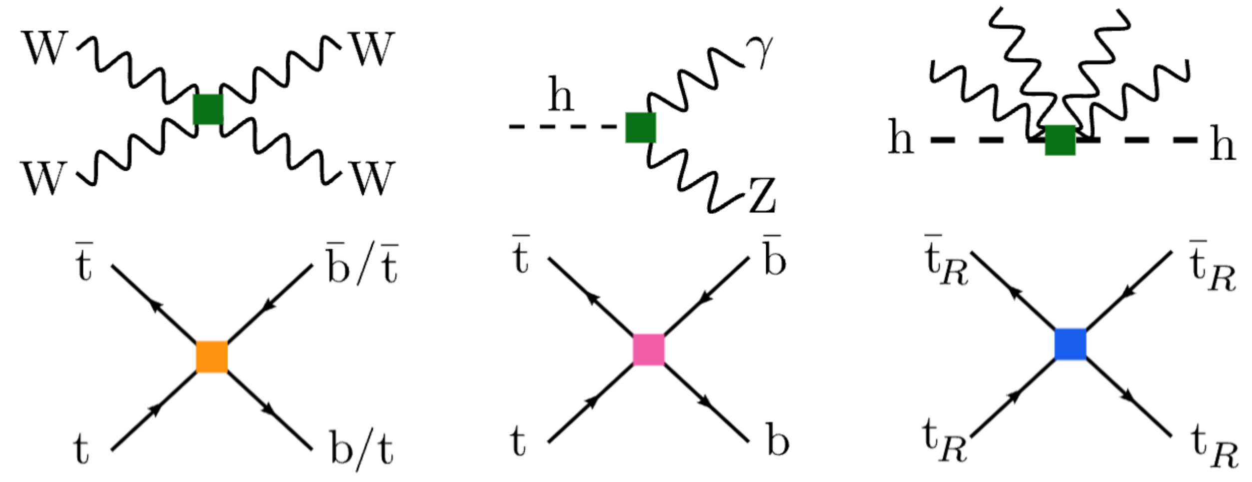
unknown coefficients

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

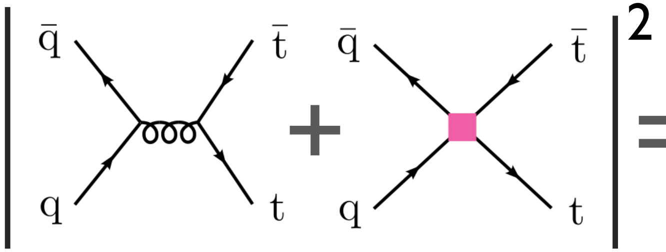
$$\frac{C_{qq}^{(3)}}{\Lambda^2} \left(\bar{q}\, \gamma^\mu \tau^I\, q\right) \left(\bar{q}\, \gamma_\mu \tau^I\, q\right)$$

known SM symmetries

Anomalous couplings & new interactions (tiny selection!)



- Predicting rates from "squared" diagrams:

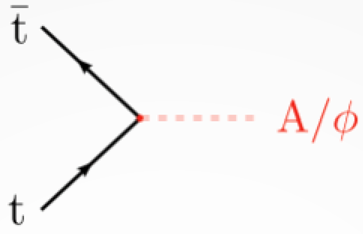


$$\left| \ldots + \ldots \right|^2 = \sigma^{SM} + \frac{C}{\Lambda^2}\sigma^{int} + \frac{C^2}{\Lambda^4}\sigma^{quad}$$

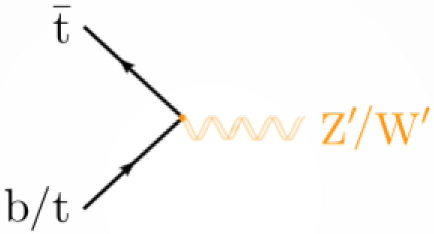
- Quite exceptional simplification!
- Being general & keeping SM symmetries: ask big questions!

NEW FORCES INVOLVING TOP QUARKS?

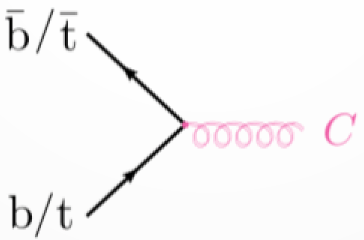
- Extended scalar sectors "two Higgs doublet models" from SUSY or other BSM physics [review]



- High-mass force carriers similar to the W and Z bosons : Z' and W' bosons [review]



- Massive "chiral" colored force carriers, otherwise similar to the gluon: axigluons [Mimasu et.al.]

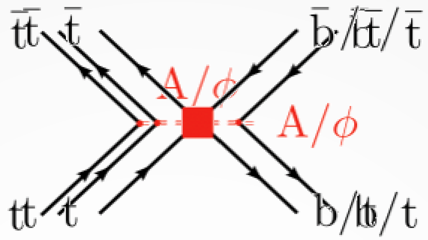


- Composite sector whose bound states mix with the SM particles: (right-handed) top-quark and/or Higgs compositeness [review]

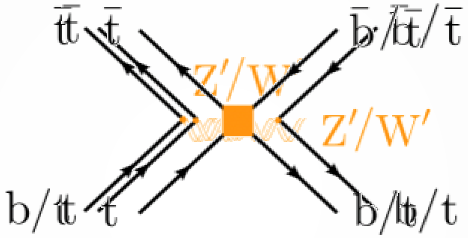
- Hypothetical UV models

NEW FORCES INVOLVING TOP QUARKS?

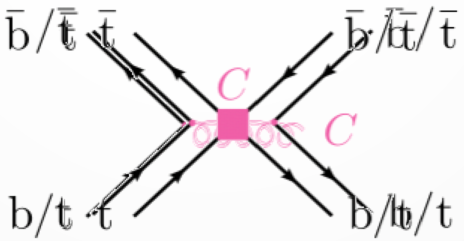
- Extended scalar sectors “two Higgs doublet models” from SUSY or other BSM physics [review]



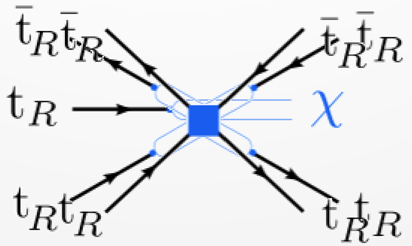
- High-mass force carriers similar to the W and Z bosons : Z′ and W′ bosons [review]



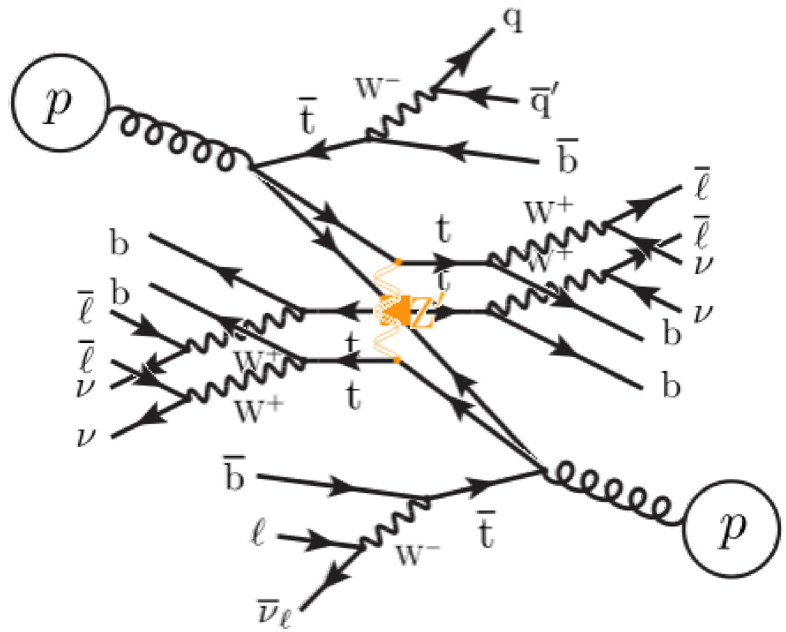
- Massive ”chiral” colored force carriers, otherwise similar to the gluon: axigluons [Mimasu et.al.]



- Composite sector whose bound states mix with the SM particles: (right-handed) top-quark and/or Higgs compositness [review]



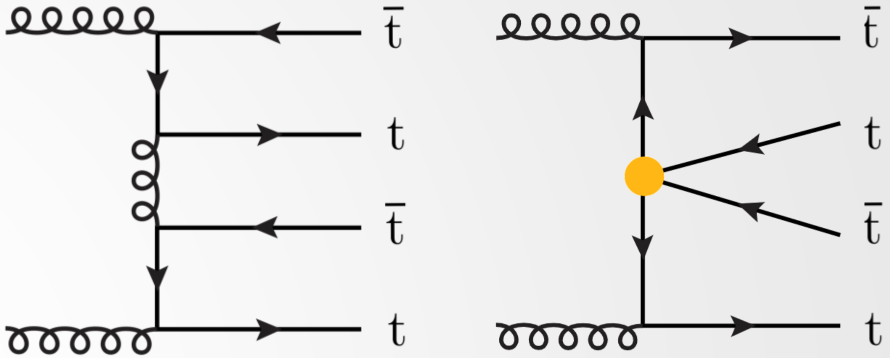
- Hypothetical UV models
  - predict force-carrier exchange
  - modify predictions for LHC processes
  - described by “effective theory”



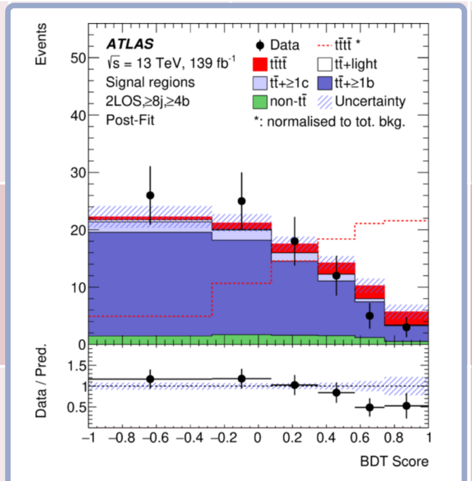
- Search for in LHC data!
- Combine t vs. t & t vs. b & t vs. light quarks

FOUR TOP QUARK PRODUCTION

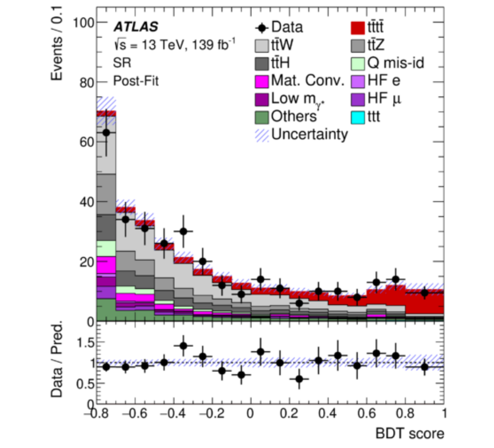
- ATLAS and CMS measure tttt in all decay channels – 0ℓ to 4ℓ
- Statistically limited: σ(SM) = 13.4 + 1 - 2.5 fb
  - most sensitive channel: 2ℓ with a same charge lepton pair
- Event-level BDTs, so far, are the workhorse classifiers

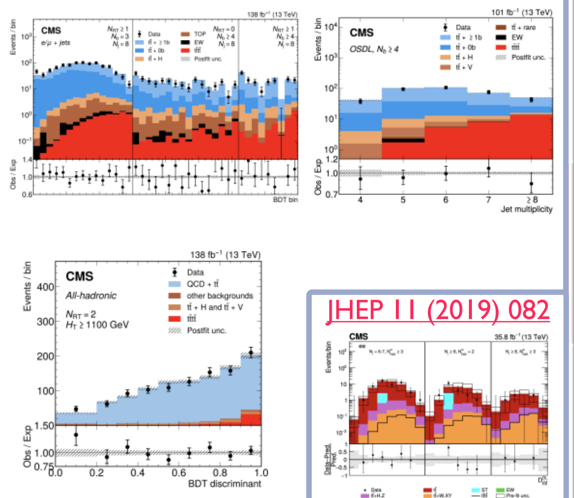


ATLAS EXPERIMENT	0ℓ	1ℓ	2ℓ(OS)	2ℓ(SS)	3ℓ+	CMS	0ℓ	1ℓ	2ℓ(OS)	2ℓ(SS)	3ℓ+
		JHEP 11 (2021) 118		EPJC 80 (2020) 1085			TOP-21-005			TOP-22-006	
2018						2018					
2017						2017					
2016						2016					

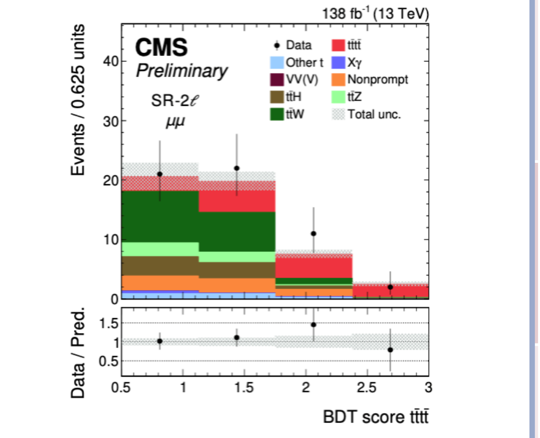


μ(tttt) = 2.0 (+0.8-0.6), 4.7σ





JHEP 11 (2019) 082

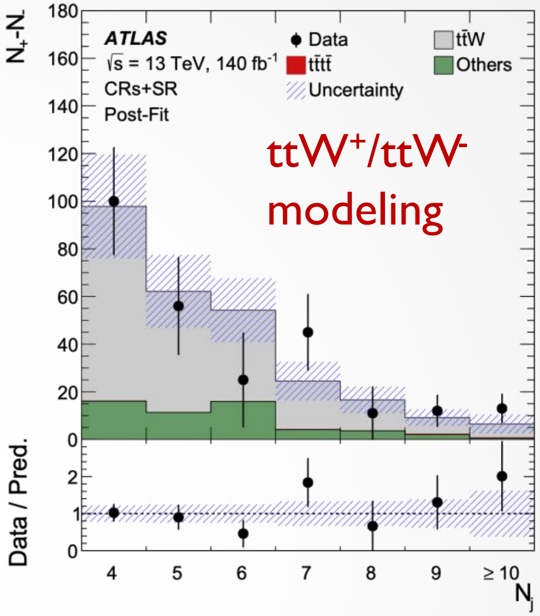


σ(tttt)=17.9 ±3.6 ± 2.5, >5σ

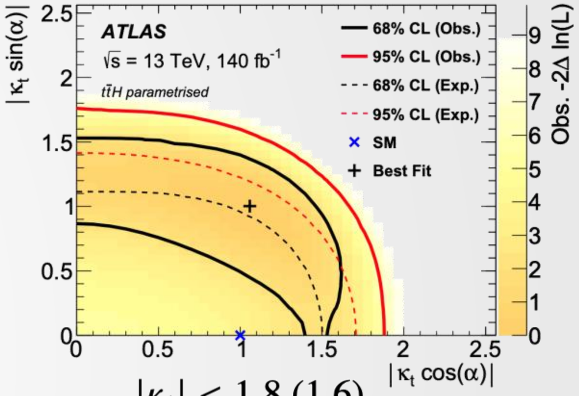
FOUR TOP QUARKS WITH A GNN: ATLAS

[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
- Now using a Graph-NN discriminant [GRAPH_NETS]
  - *Classifier* trained for cross-section measurement – LLR trick
  - Edge-Convolution layers exploit multi-jet correlation
  - Leptons, E_T^{miss}, variable-length jet system



Constraints on $\varkappa_t$ and CP mixing angle α



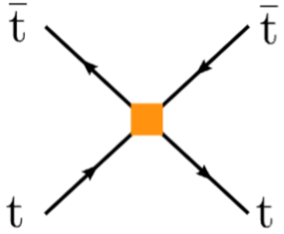
$|\kappa_t| < 1.8\ (1.6)$

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

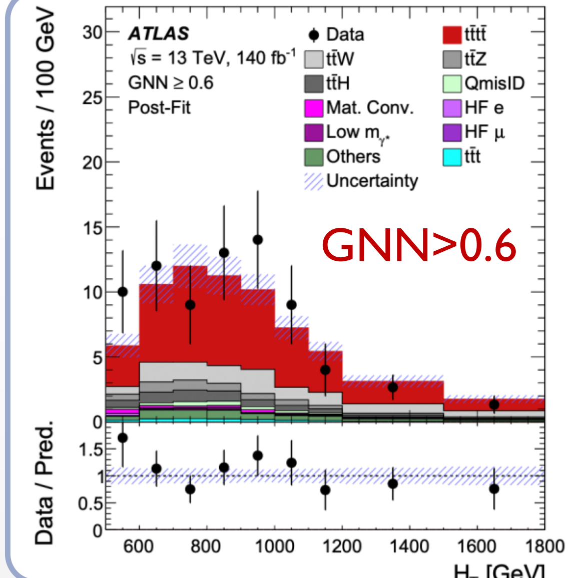
$$\mu = 1.9 \pm 0.4(\text{stat})\ ^{+0.7}_{-0.4}(\text{syst}) = 1.9\ ^{+0.8}_{-0.5}$$

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

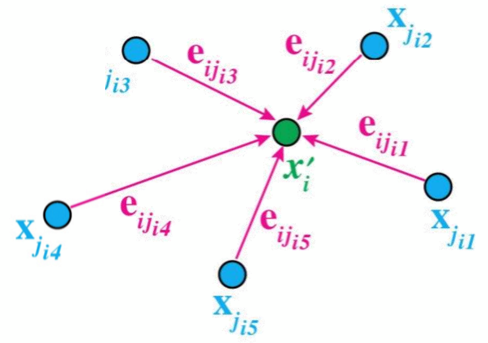
Operators	Expected C_i/Λ^2 [TeV $^{-2}$]	Observed C_i/Λ^2 [TeV $^{-2}$]
O^1_{QQ}	[-2.4, 3.0]	[-3.5, 4.1]
O^1_{Qt}	[-2.5, 2.0]	[-3.5, 3.0]
O^1_{tt}	[-1.1, 1.3]	[-1.7, 1.9]
O^8_{Qt}	[-4.2, 4.8]	[-6.2, 6.9]



Need more measurements & more operators



Edge convolution [1801.07829]



Amenable for your problem's symmetry!!

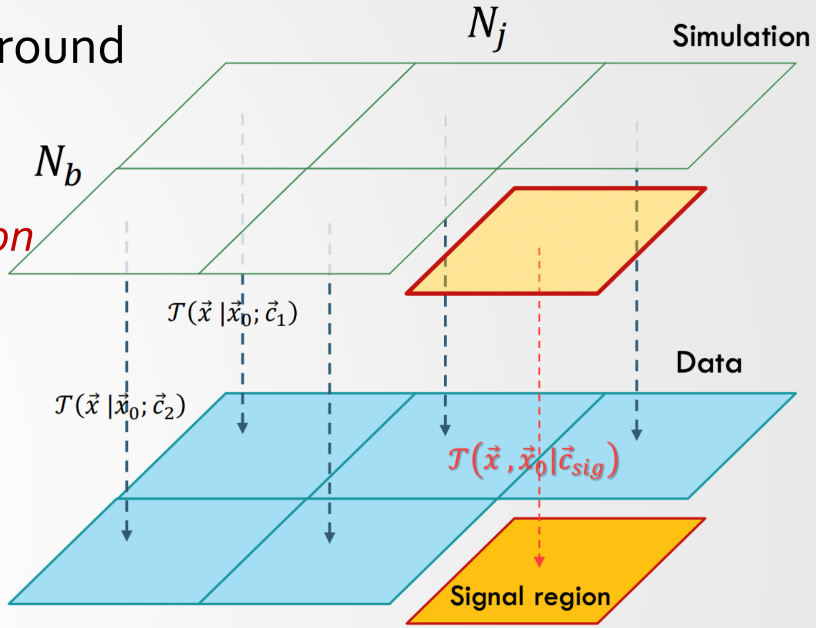
SHAPE CALIBRATION WITH ABCDNN

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

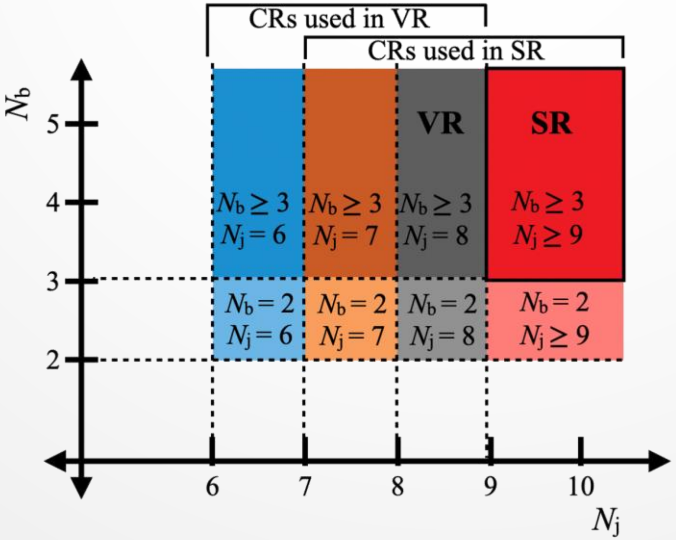
- CMS : BDT classifier from 20 features for all-hadronic four-top background
- Corrects BDT shape using [ABCDnn]: Neural autoregressive flow
  - Learn a invertible transformation of H_T/BDT shape from *data* to *simulation* *conditioned on a region c*

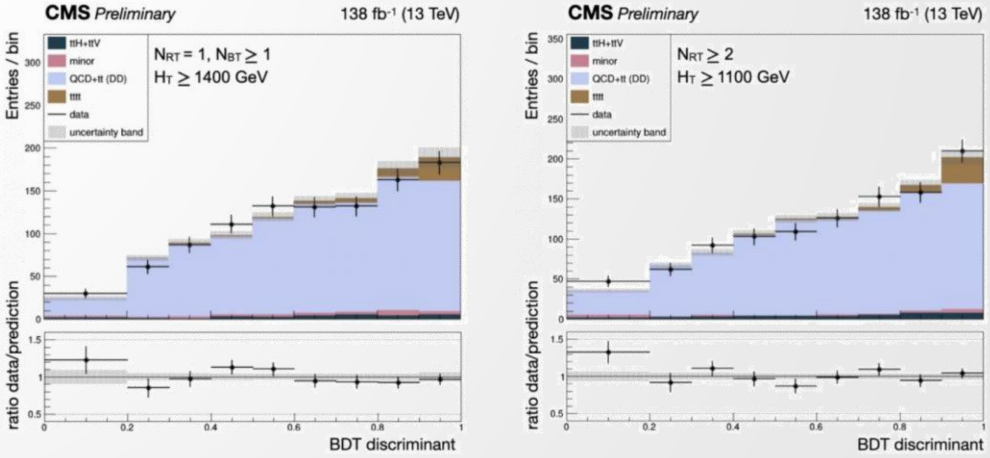
$$\int \mathcal{T}(\vec{x},\vec{x}_0|\vec{c})\, f_{src}(\vec{x}_0)\, d\vec{x}_0 = f_{target}(\vec{x}|\vec{c})$$

  - 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)

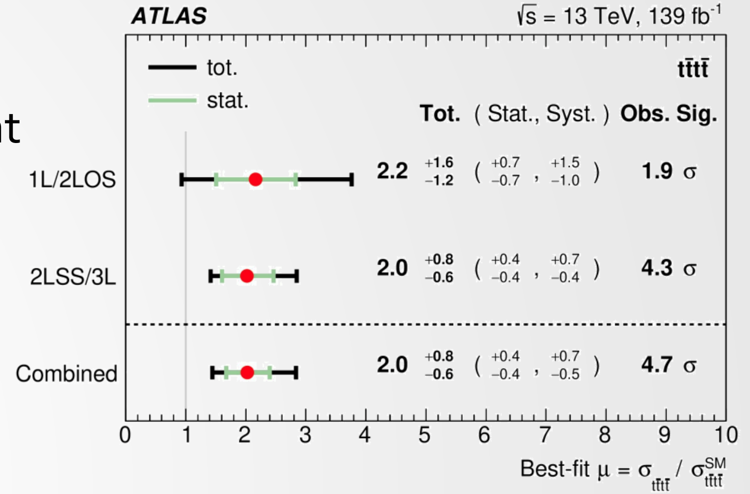




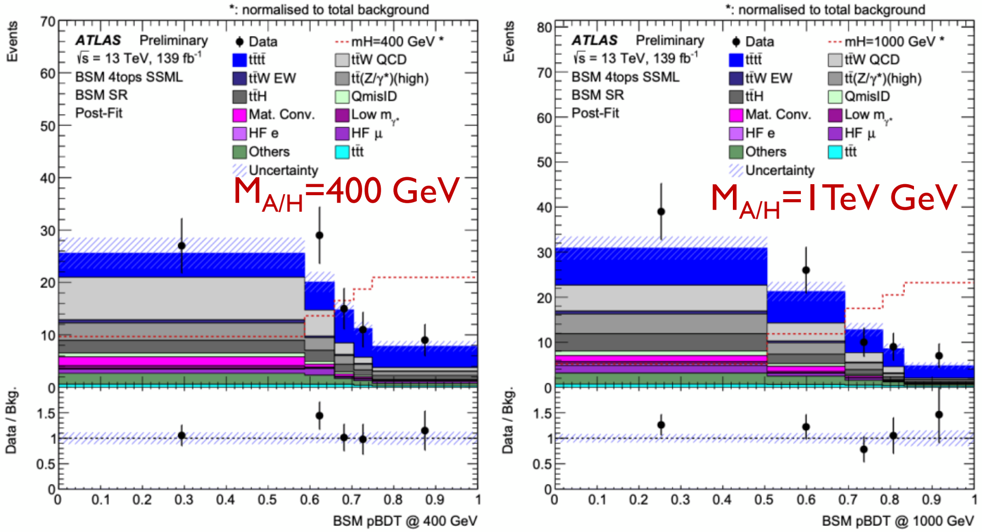
A PARAMETRIZED CLASSIFIER IN TT+(H/A→TT)

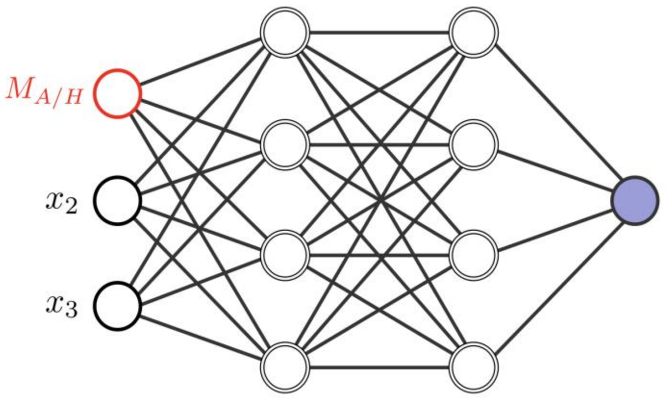
[arXiv:2211.01136]

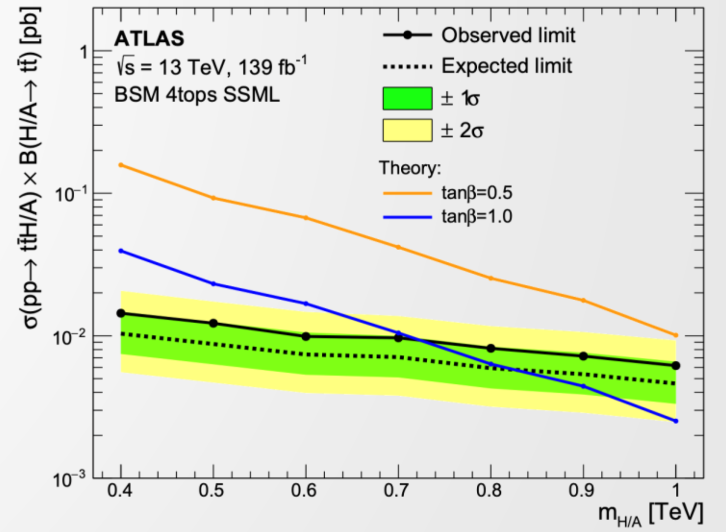
- The 2HDM model as a function of $M_{A/H}$ predicts resonant 4t production
  1. Use the signal region from the ATLAS 2ℓSS /≤3l 4t cross section measurement
  2. Train "parametrized" multi-variate discriminate as a function of $M_{A/H}$
  - example of a one-parameter "parametrized classifier"



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

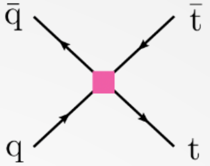






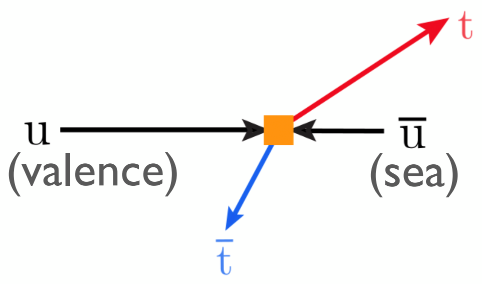
- Can use a similar technique for high-dimensional EFT measurements?

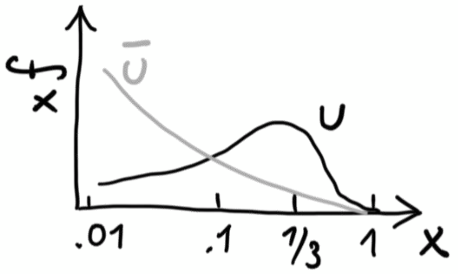
TOP QUARK CHARGE ASYMMETRY

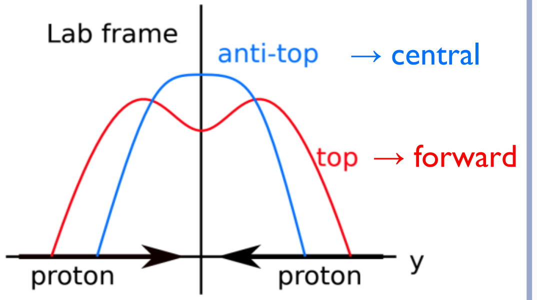


[TOP-21-014] [arxiv:2208:12095]

- Use subtle kinematic effects to target interactions with light quarks
- The "valence" light-quark carries, on average, a larger fraction of the protons momentum compared to anti-quarks



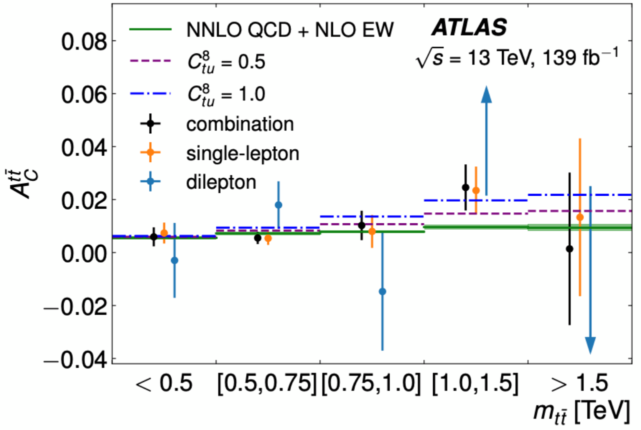


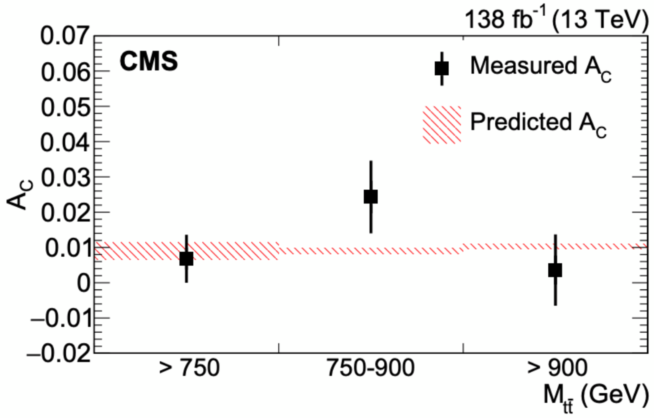


- The +t quark in pair production is more forward
- Charge asymmetry cancels overwhelming gluon-initiated background
  - Permille effect
- CMS (1ℓ) and ATLAS (1ℓ/ 2ℓ, resolved/boosted) have measured A_C(tt)
  - ATLAS A_C(tt) = 0.0068±0.0015 ↔ 4.7σ evidence

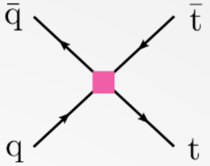
Exploit "forward-backward" symmetry:

$$A_C^{t\bar{t}} = \frac{N(\Delta|y| > 0) - N(\Delta|y| < 0)}{N(\Delta|y| > 0) + N(\Delta|y| < 0)}$$



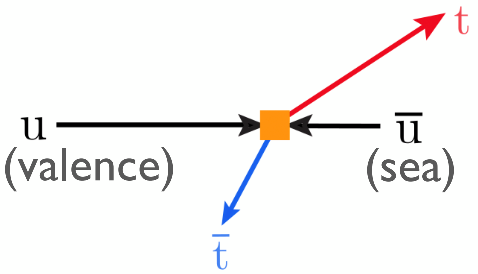


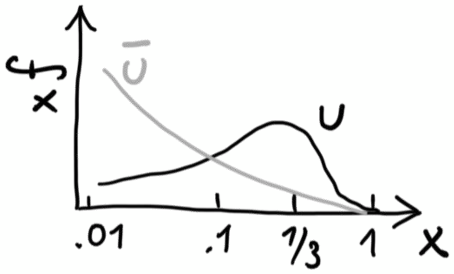
TOP QUARK CHARGE ASYMMETRY

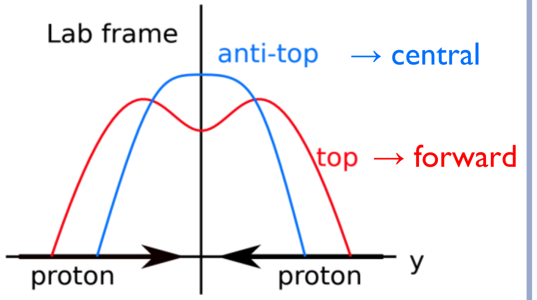


[TOP-21-014] [arxiv:2208:12095]

- Use subtle kinematic effects to target interactions with light quarks
- The "valence" light-quark carries, on average, a larger fraction of the protons momentum compared to anti-quarks

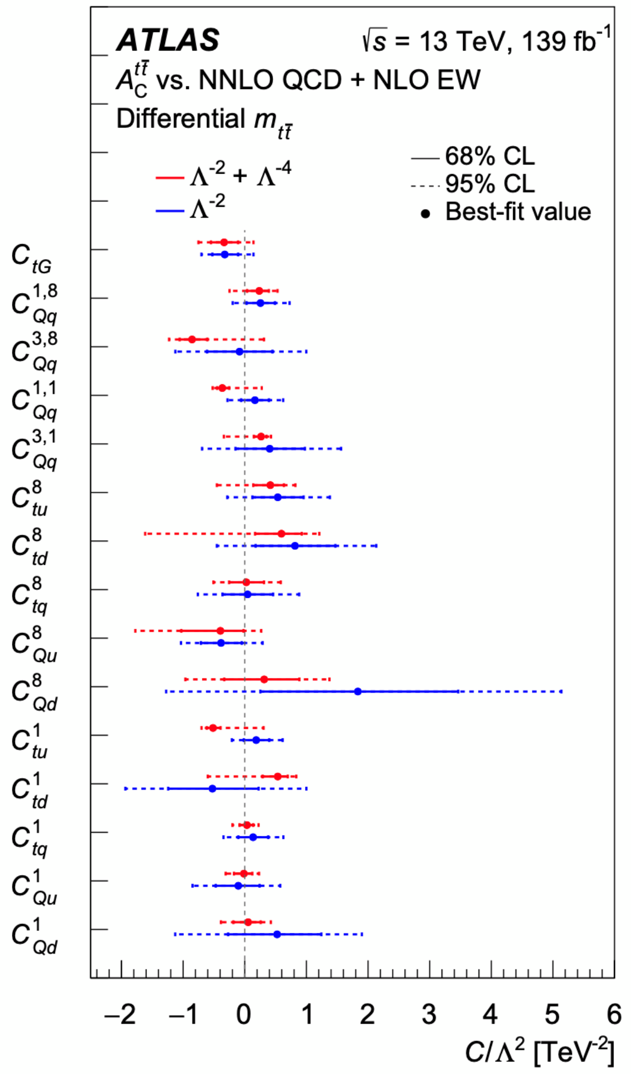




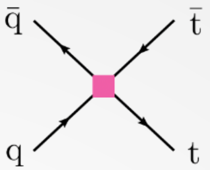


- The +t quark in pair production is more forward
- Charge asymmetry cancels overwhelming gluon-initiated background
  - Permille effect
- CMS (1ℓ) and ATLAS (1ℓ/ 2ℓ, resolved/boosted) have measured $A_C(tt)$
  - ATLAS $A_C(tt) = 0.0068 \pm 0.0015 \leftrightarrow 4.7\sigma$ evidence

Comprehenseive EFT interpretation

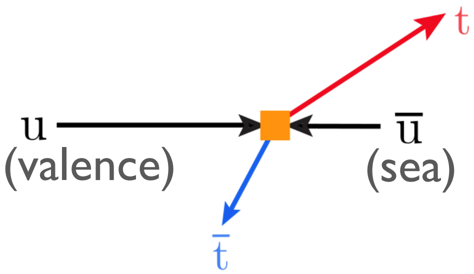


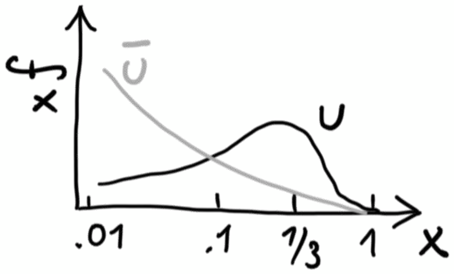
TOP QUARK CHARGE ASYMMETRY

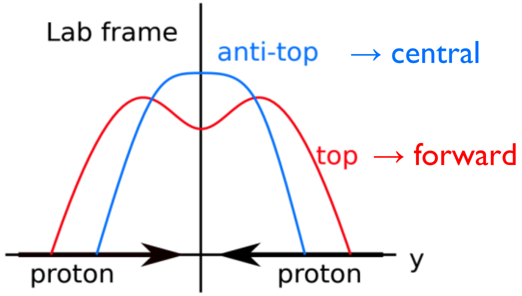


[TOP-21-014] [arxiv:2208:12095]

- Use subtle kinematic effects to target interactions with light quarks
- The "valence" light-quark carries, on average, a larger fraction of the protons momentum compared to anti-quarks

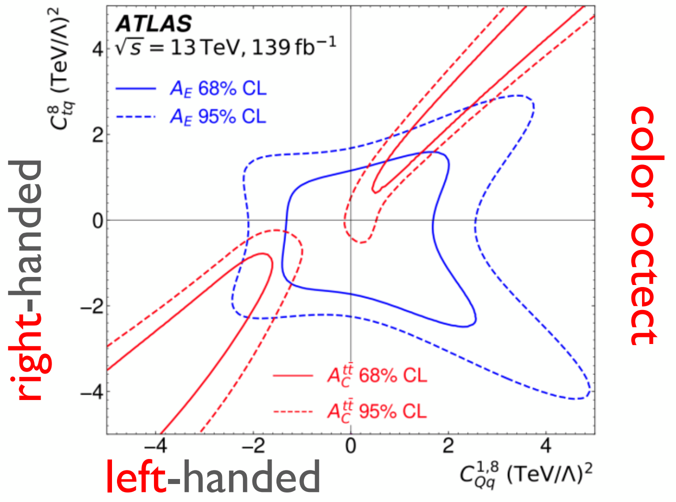






- The +t quark in pair production is more forward
- Charge asymmetry cancels overwhelming gluon-initiated background
  - Permille effect
- CMS (1ℓ) and ATLAS (1ℓ/ 2ℓ, resolved/boosted) have measured A_C(tt)
  - ATLAS A_C(tt) = 0.0068±0.0015 ↔ 4.7σ evidence

"Left-handed" vs. "Right-handed"
→ flat direction



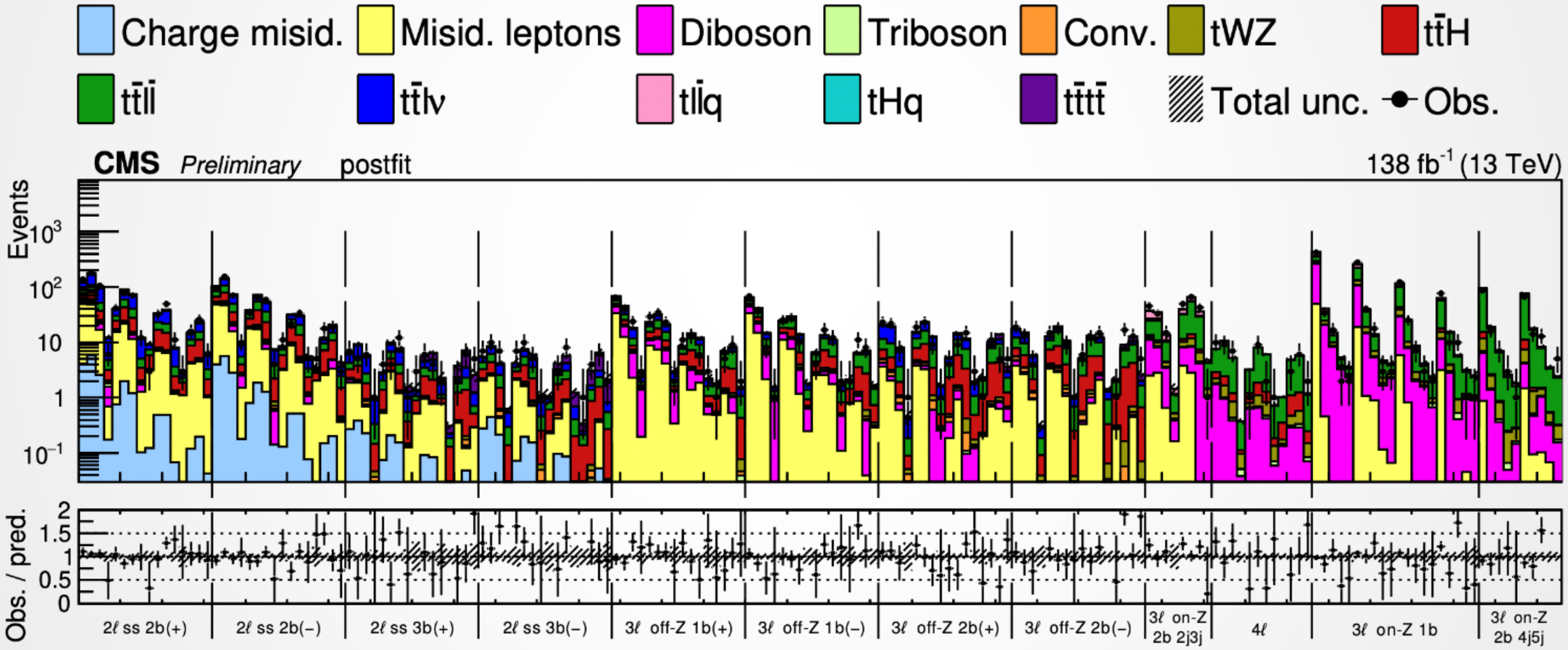
→ resolved with Energy asymmetry

→ need to combine many measurements for unambigous results

TOP QUARKS WITH ADDITIONAL LEPTONS

[TOP-22-006]

- Targets top quark pair production in association with Z/W/H [TOP-22-006]
  - 2ℓSS/3ℓ/4ℓ categories with different b-tag multiplicities and with/without on Z requirement



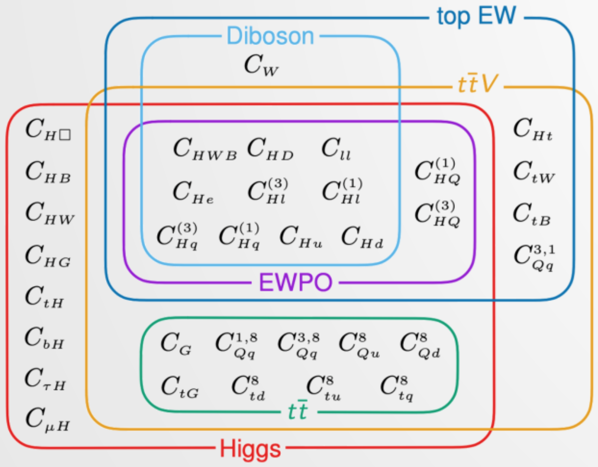
  - 178 measurements with full uncertainty correlation, constraining 22 operators
  - most recent CMS step towards global in-experiment fit

GLOBAL FITS (WITHIN EXPERIMENTS)

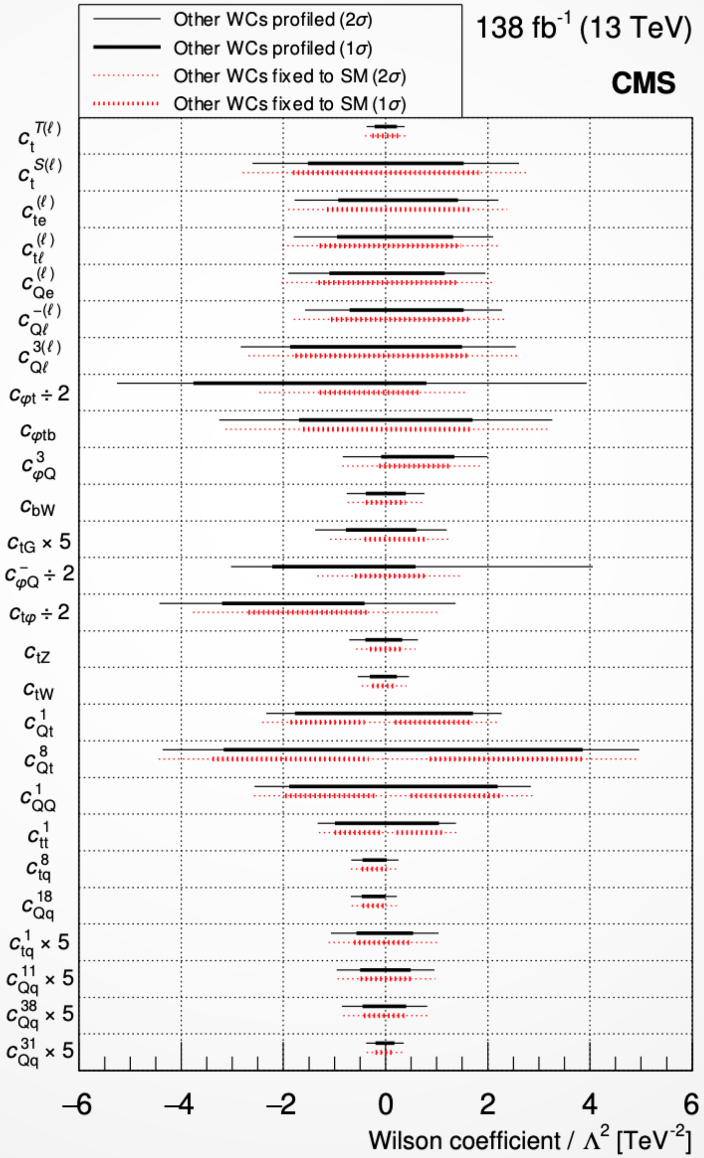
[ATLAS-PHYS-PUB-2022-037]

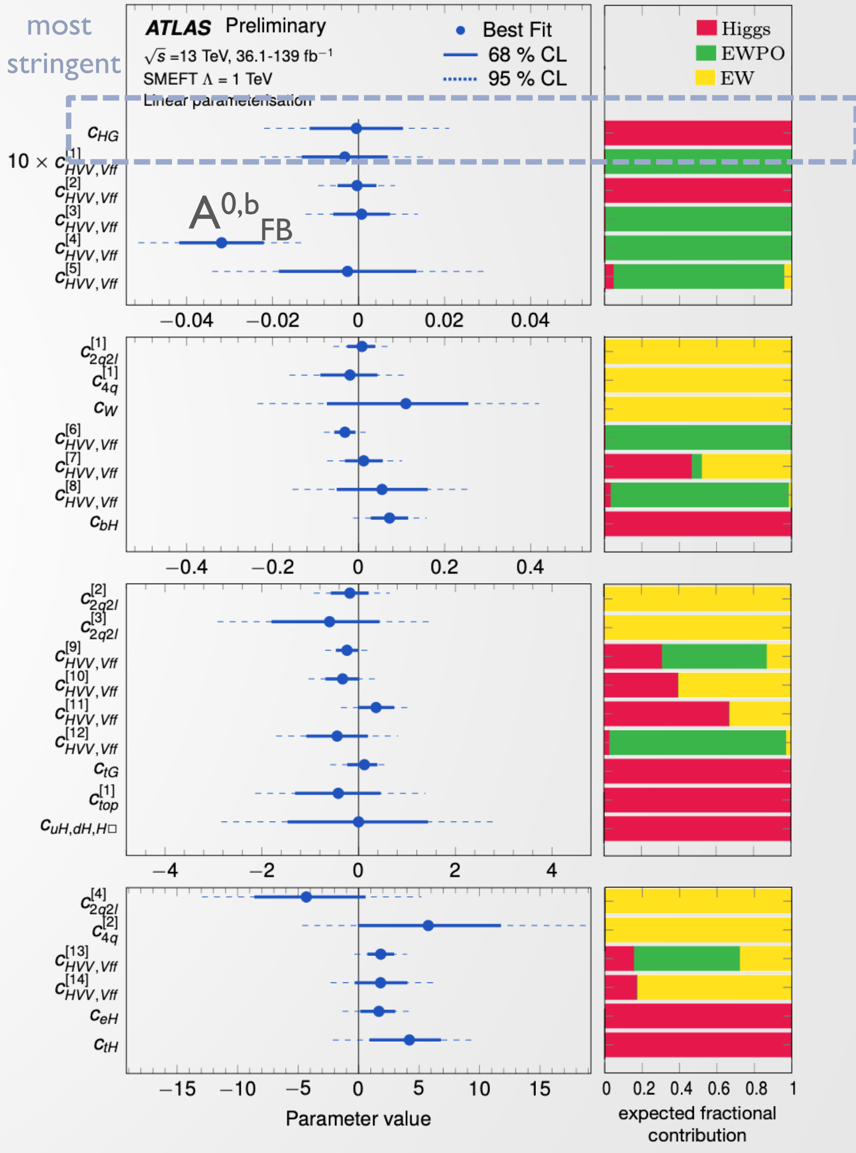
[TOP-22-006]

- CMS "top quark pair + Z/W/H"
  - full 22D uncertainty correlation
  - 22 operators, 178 measurements
- No signal of new forces down to 10^{-18} m
- ATLAS: Higgs+EWK+EWPO
  - LEP & SLC EW precision data
  - 6 coeff. + 22 lin. comb
  - mostly conistent with SM



- Need to combine all sectors

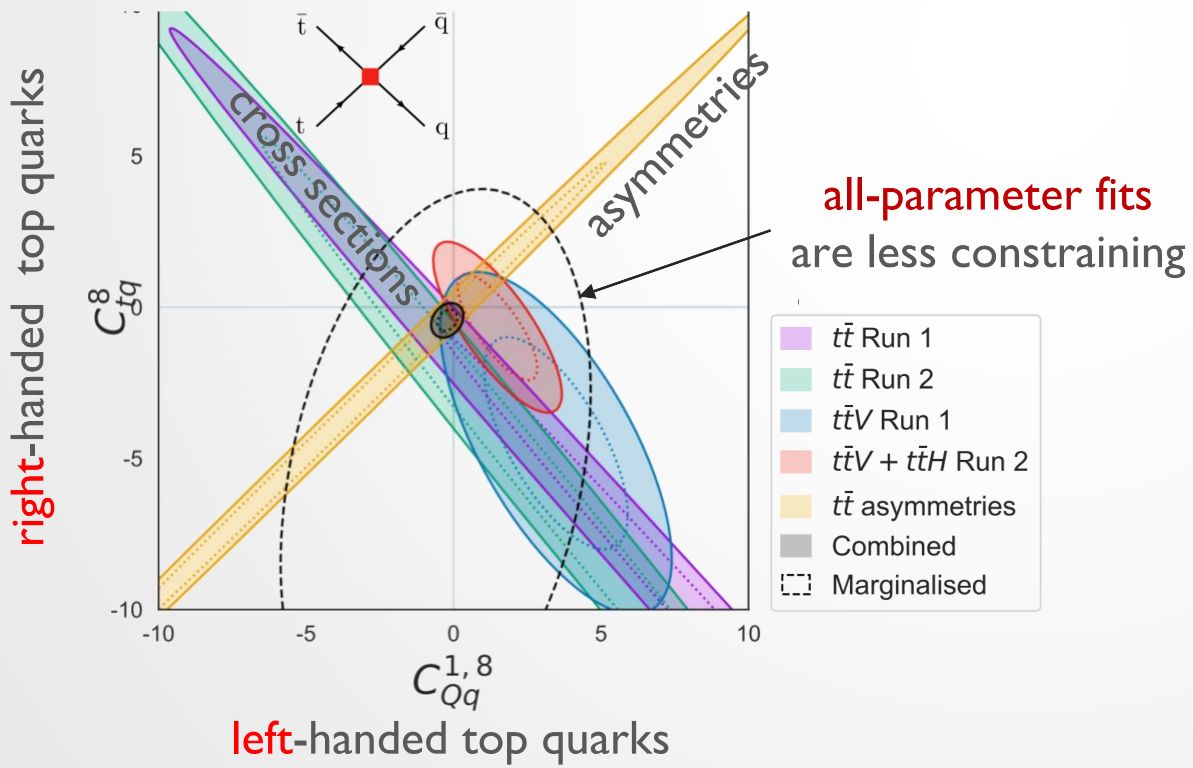




LOOKING INTO MANY DIRECTIONS AT ONCE

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

- Global fits: Combine all available individual measurements – outside the collaborations
  - For single or few operators: tight constraint from combined measurement
- Our earlier example: forces of left- and right-handed top quarks – two operators
- However(!) including all EFT operators leads to much less powerful constraints
  - Physics question: Can we use the kinematic information in the events to resolve the ambiguities?

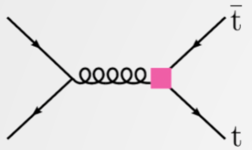


- Can machine-learning help to improve the analysis strategy
- Can we parametrize an EFT classifier?

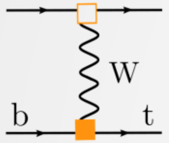
TOP QUARK PAIR + Z BOSON

[JHEP 12 (2021) 083]

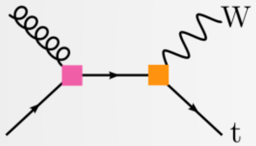
tt
811 pb



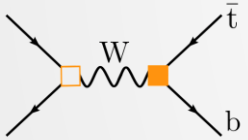
t (t-channel)
217 pb



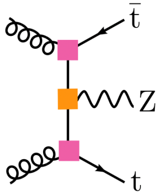
tW
72 pb



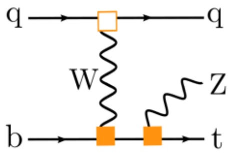
t (s-channel)
10 pb



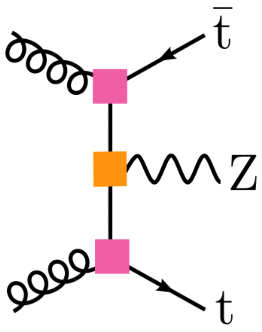
ttZ
1 pb

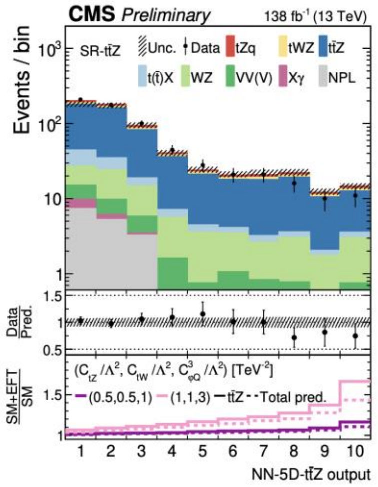


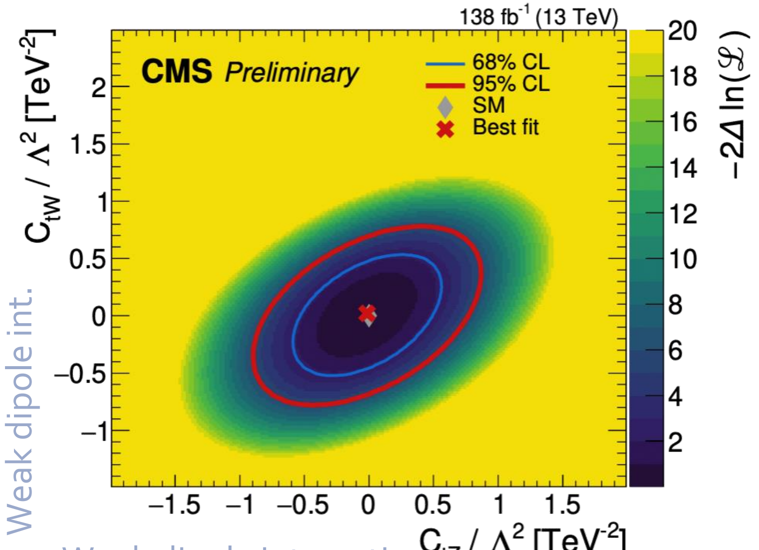
tZq
0.088 pb



- Measure the top quark – Z boson coupling
- Train separate "SM vs. EFT" classifiers
  - "likelihood trick" for SMEFT effects
  - trained on a signal mix, mixing different kinematics
- set limits on weak diple interactions & vector coupling interactions

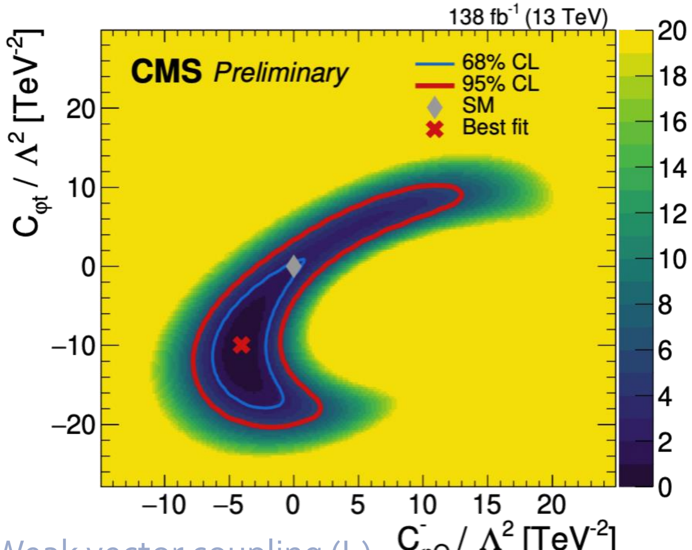






Weak dipole int.

Weak dipole interactions



Weak vector coupling (R)

Weak vector coupling (L)

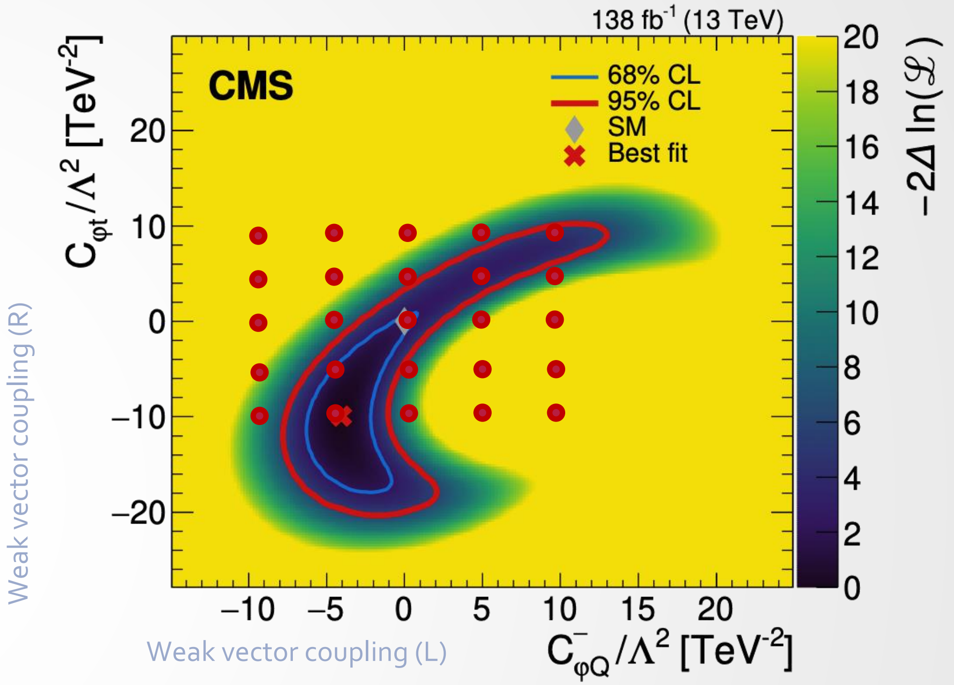
CAN WE JUST LEARN EFT EFFECTS "ON AVERAGE"?

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

$\boldsymbol{\theta}$ - ignorant

mixing signals & case dependent mixes

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



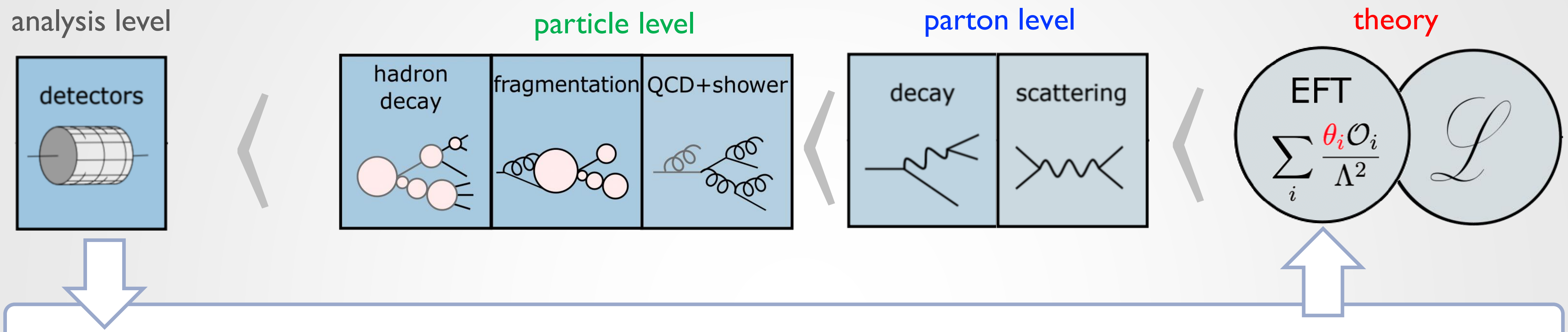
[TOP-21-001]

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

A CONDITIONAL SEQUENCE

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

RS et. al., [2107.10859], [2205.12976]



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

Make loss function aware of analytic SMEFT structure

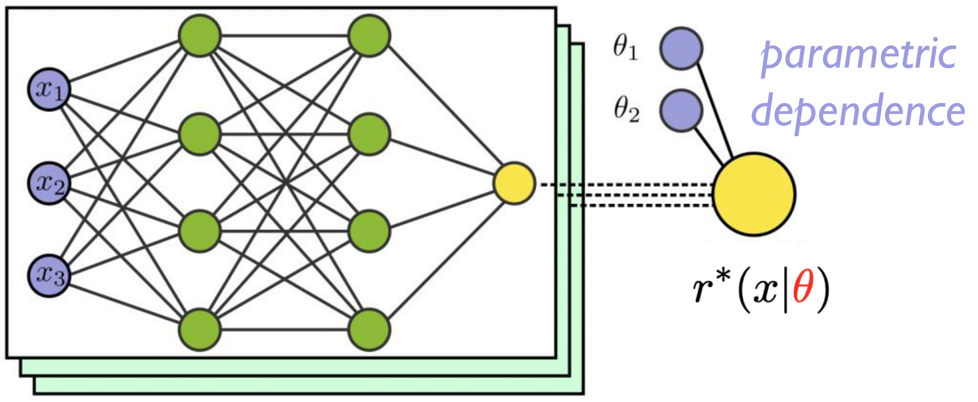
Invert likelihood trick $\quad \hat{f}(\boldsymbol{x};\boldsymbol{\theta}) = \dfrac{1}{1+\hat{r}(\boldsymbol{x};\boldsymbol{\theta})}$

"Domain knowledge": positive quadratic polynomial

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

Minimize: $\quad r^*(x|\theta) = \dfrac{p(x|\theta)}{p(x|\mathrm{SM})}$

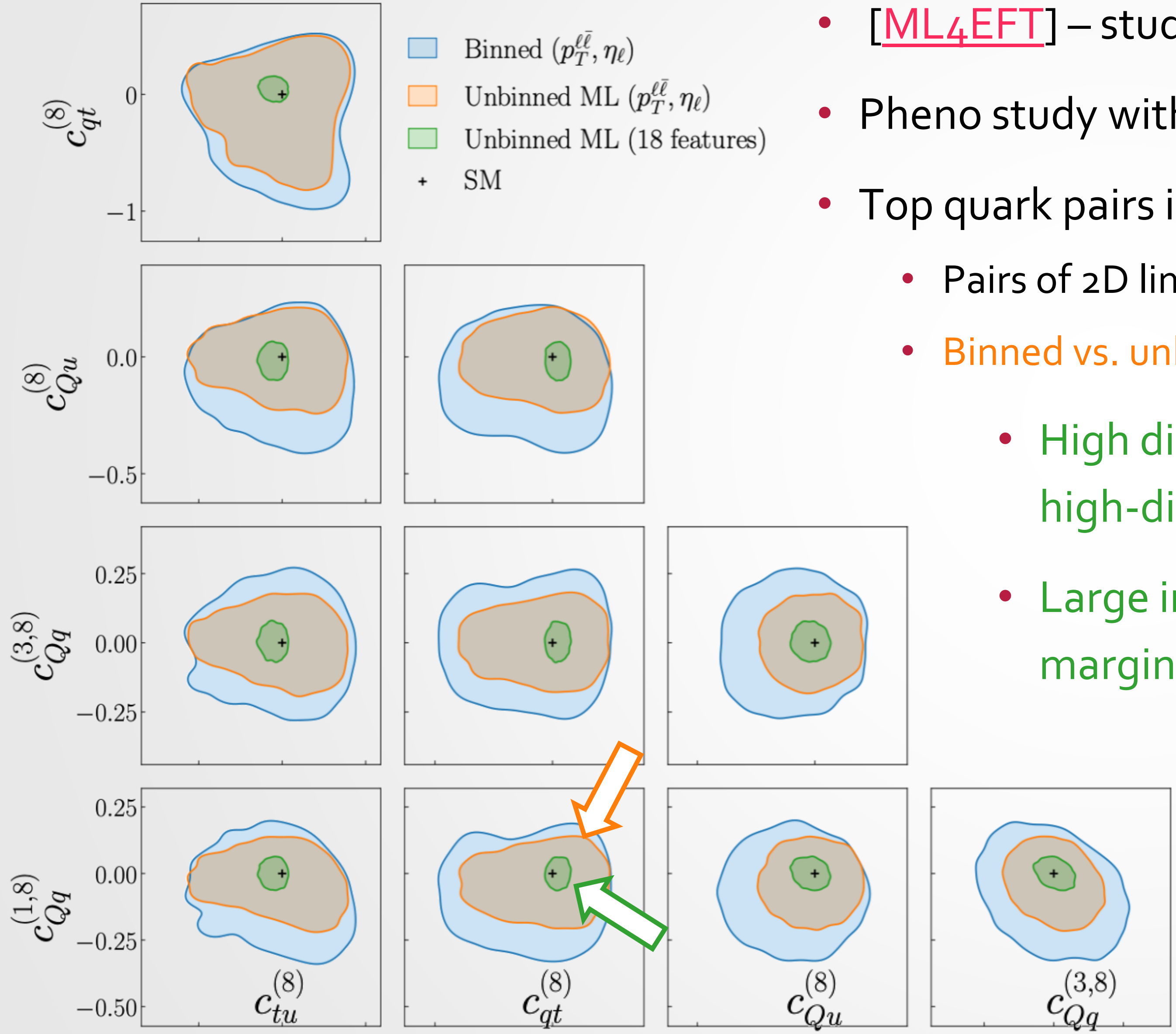
inject new technology here ↴



ML4EFT R. Ambrosio, J. Hoeve, M. Madigan, J. Rojo, V. Sanz [2211.02058]

[CMS-TOP-PAS-20-006]

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

- What's missing to go all-in? Systematics.
  - Systematic effects are not polynomial.
  - However, can be learned with NNs

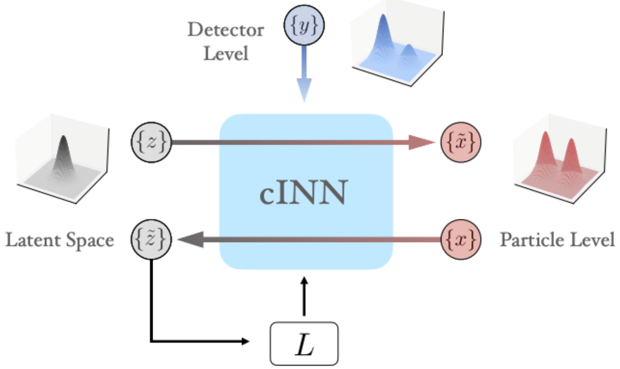
$$L = \sum_{\theta,\nu=0} \left\langle \Big(r(x_{\mathrm{det}}, z_{\mathrm{ptl}}, \cdots, z_{\mathrm{p}} | \theta, \nu) - \hat{f}_{\theta}(x_{\mathrm{det}}) e^{\nu \hat{\delta}_1(x_{\mathrm{det}}) + \nu^2 \hat{\delta}_2(x_{\mathrm{det}})} \Big)^2 \right\rangle_{\mathrm{SM},\nu=0}$$

Systematic variations are cheap!

Improve modeling here

- A challenge: dim(**ν**) , dim(**θ**) ~ 20 – 50, and high event counts in the profiling
- Divide & conquer #1: Experiments begun machine-learning certain nuisances: h_{damp}, b-fragmentation

- Divide & conquer #2: Defer SM-EFT interpretation "Unbinned unfolding in high dimensions" [paper]
- Only $p(x_{\mathrm{det}}|z_{\mathrm{ptl}})$ is available in in forward mode.
- ML-Unfolding algorithms use Bayes' theorem $p(x_{\mathrm{det}}|z_{\mathrm{ptl}})p(z_{\mathrm{ptl}}) = p(z_{\mathrm{ptl}}|x_{\mathrm{det}})p(x_{\mathrm{det}})$ to learn $p(z_{\mathrm{ptl}}|x_{\mathrm{det}})$



[cINN]

- [OmniFold] reweights the observation to the ptl-level
- Report unbinned unfolded data; then SMEFT analysis

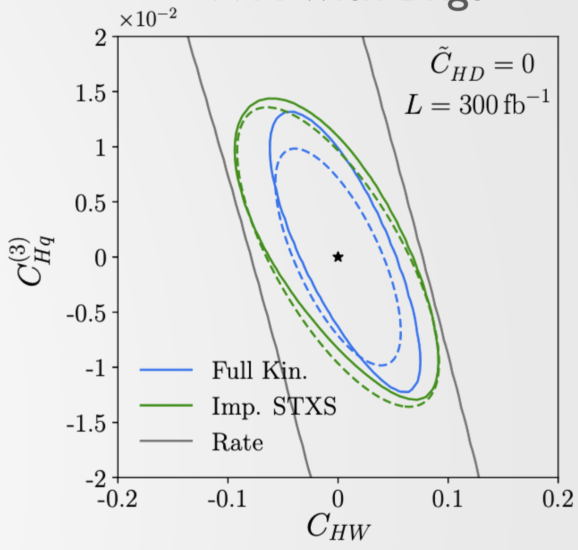
SUMMARY

- The LHC is pushing ever deeper into the TeV regime
  - There are no signs of resonant physics beyond the standard model
- Taking seriously what we know – symmetries & particle content
  - EFT has become the language of choice
  - We can phrase largly model-independent questions
    - E.g.: Forces between heavy quarks on length scales beyond 10^{-18} m
- Imply the need for a global view
  - High-dimensional analyses leave room for ambiguities
  - ML tools can significantly help – particularly in all-operator fits
- For sure, we'll see more global analyses of the LHC data, tackling more of the "big questions"
- Outlook: At higher mass dimension, the number of operators grows exponentially
  - If we loose track of the operators physics meaning, we're just re-representing the dataset
  - A better representation could then be an unbinned unfolded dataset

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

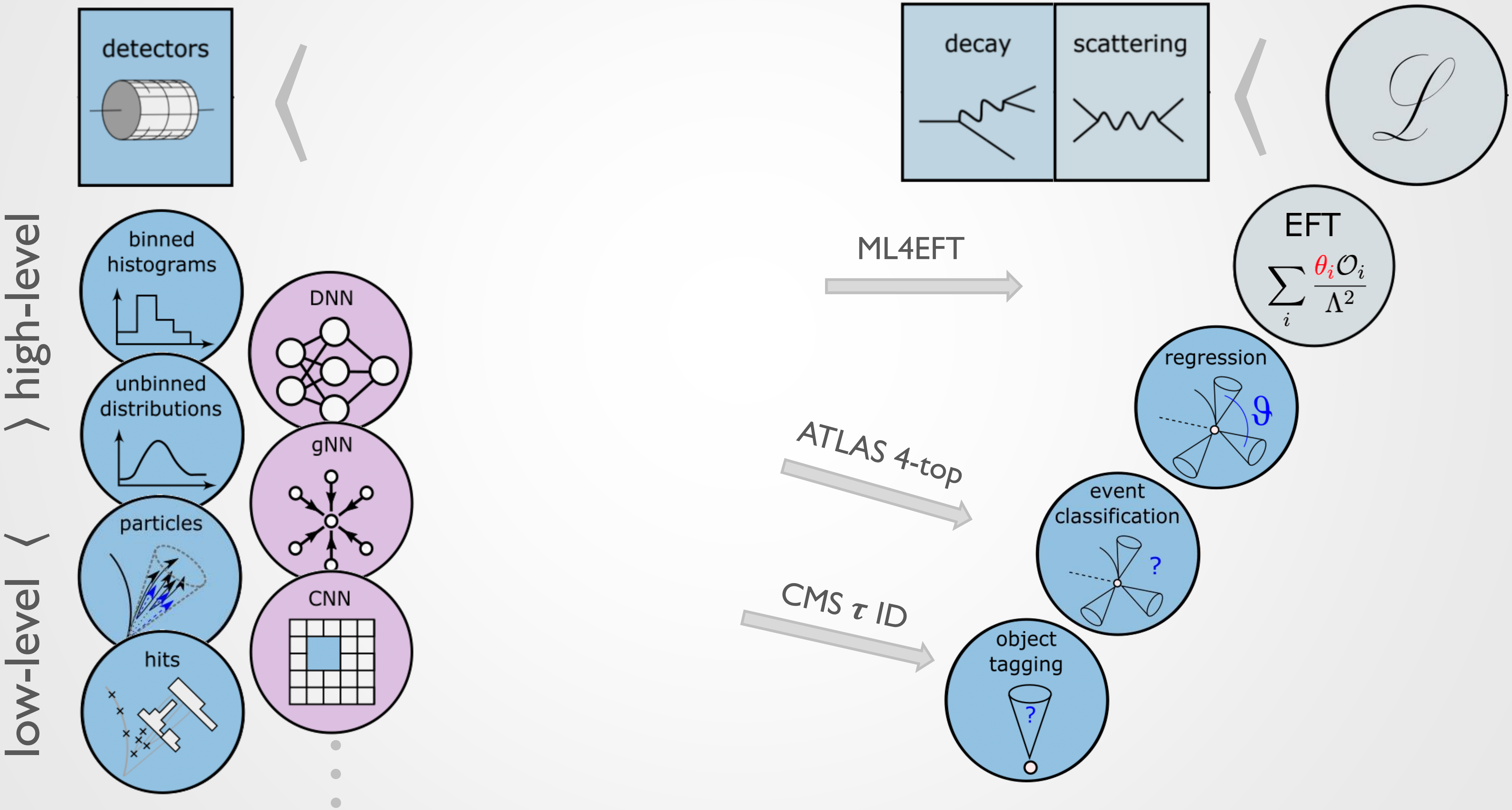


my practical experience

DATA REPRESENTATION

[JINST 13 (2018) P10005]

[JINST 17 (2022) P07023]



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

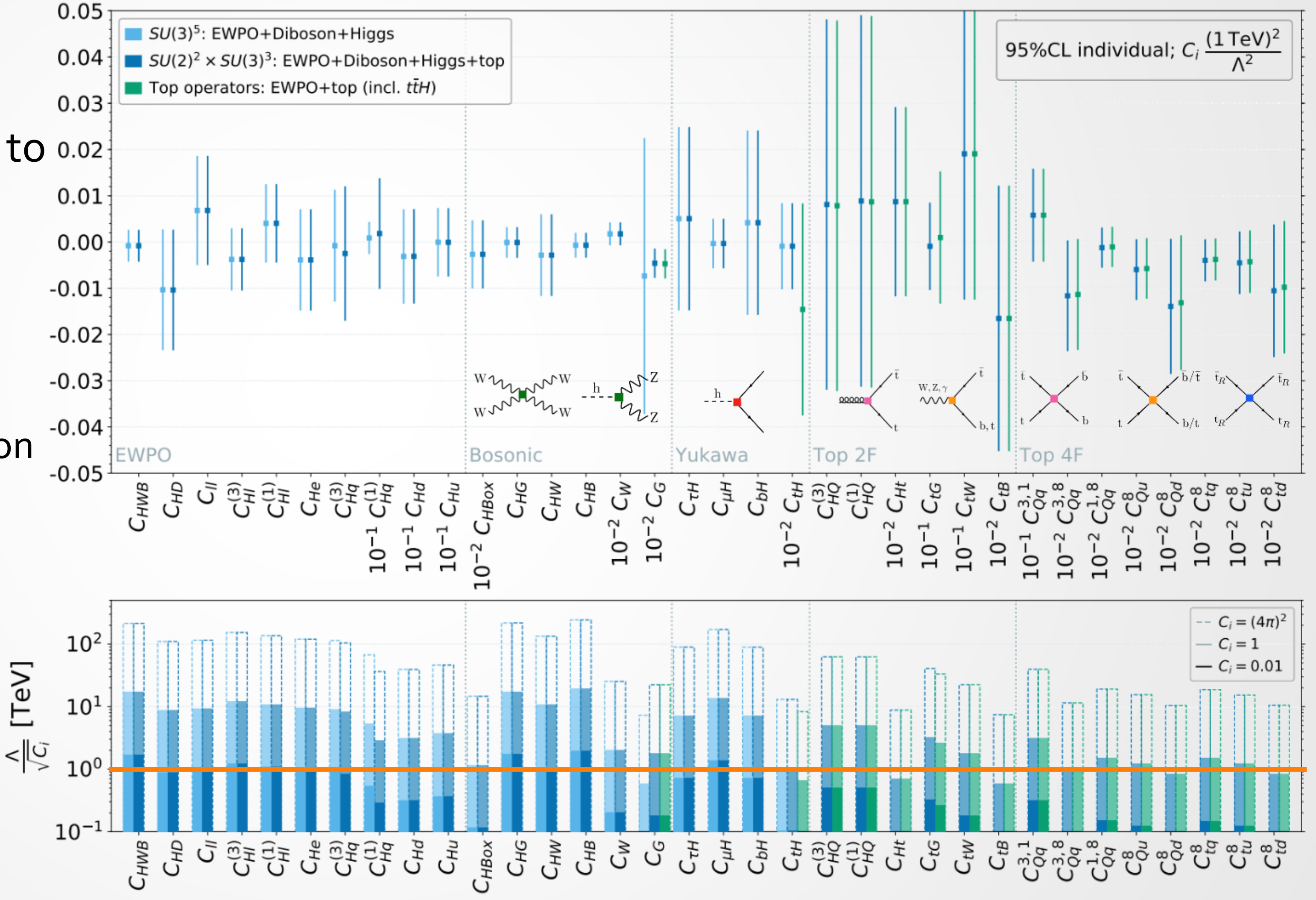
[Ellis, Sanz, et.al. FitMaker JHEP04(2021)279]

[Rojo, Maltoni et.al. SMEFiT JHEP11(2021)089]

GLOBAL FITS

- First global interpratations combining experimental results
- Individual operators constrained to ~ 1TeV regime: 10^{-18} m
- Caveats
  - background-subtracted inputs
  - simplified uncertainty correlation
- All-operator (marginalized) fits significantly less constraining
  - adding more processes → resolve ambiguities
- Experiments move towards more global fits

Fit one operator at a time



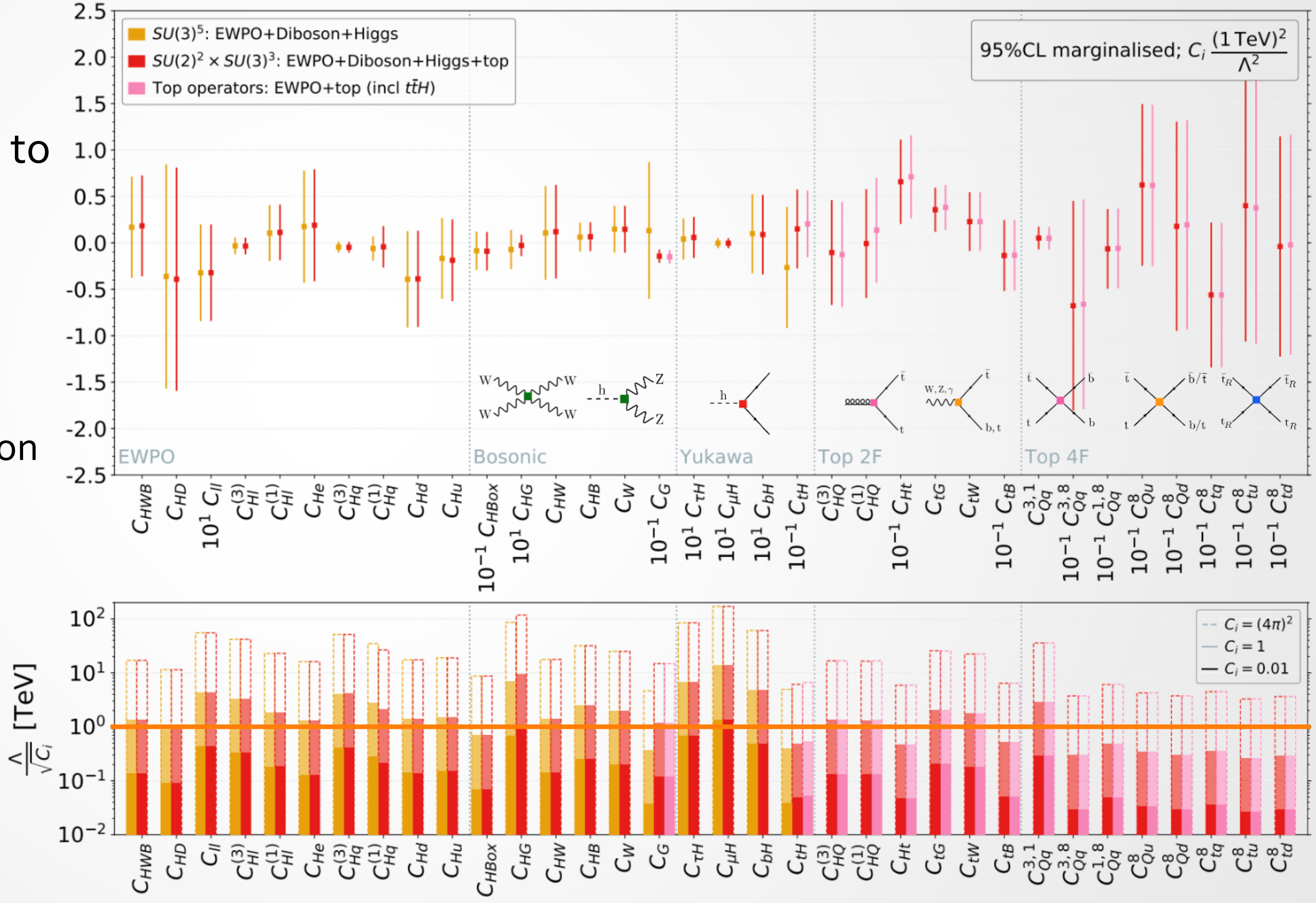
[Ellis, Sanz, et.al. FitMaker JHEP04(2021)279]

[Rojo, Maltoni et.al. SMEFiT JHEP11(2021)089]

GLOBAL FITS

- First global interpratations combining experimental results
- Individual operators constrained to ~ 1TeV regime: 10^{-18} m
- Caveats
  - background-subtracted inputs
  - simplified uncertainty correlation
- All-operator (marginalized) fits significantly less constraining
  - adding more processes → resolve ambiguities
- Experiments move towards more global fits

Fit of all operators



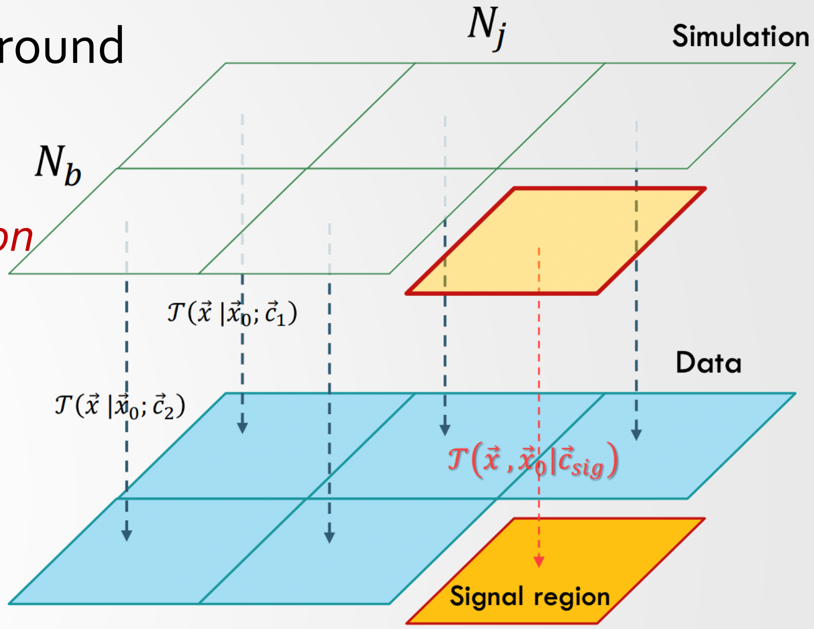
CALIBRATE BDT SHAPE WITH ABCDNN

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

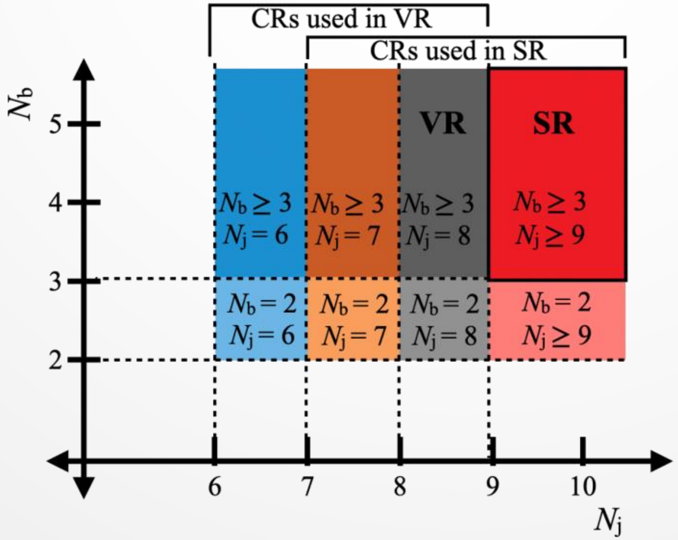
- CMS : BDT classifier from 20 features for all-hadronic four-top background
- Corrects BDT shape using [ABCDnn]: Neural autoregressive flow
  - Learn a invertible transformation of H_T/BDT shape from *data* to *simulation* *conditioned on a region c*

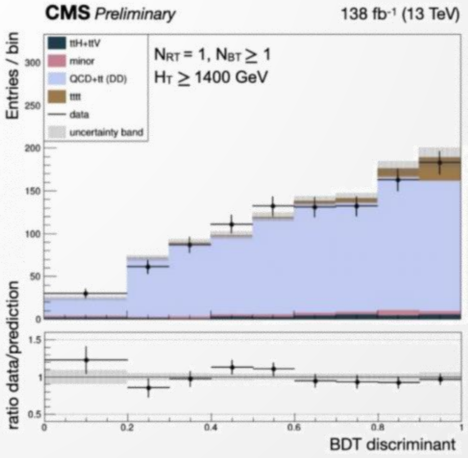
$$\int \mathcal{T}(\vec{x}, \vec{x}_0|\vec{c}) f_{src}(\vec{x}_0)\, d\vec{x}_0 = f_{target}(\vec{x}|\vec{c})$$

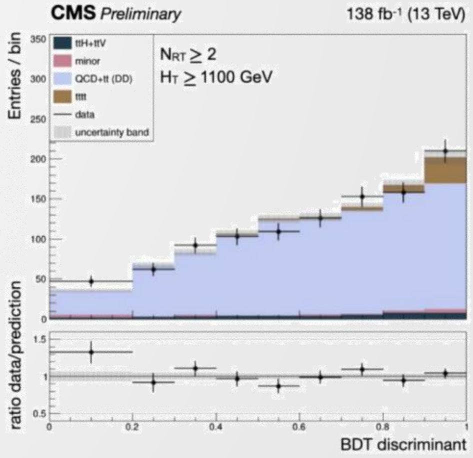
  - Technically, a DNN predicts the parameters of a bijective mapping
    - Encoding of indexed region is DNN input → extrapolate to SR



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



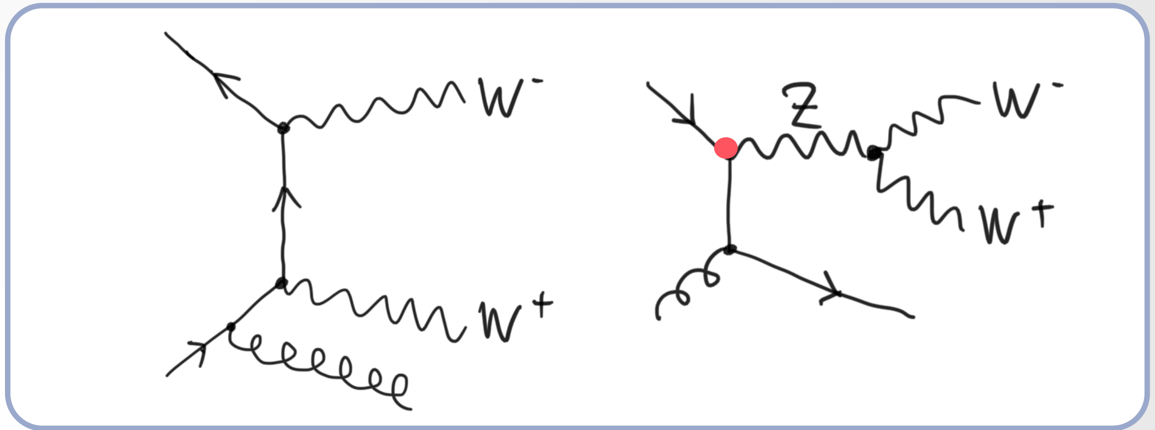


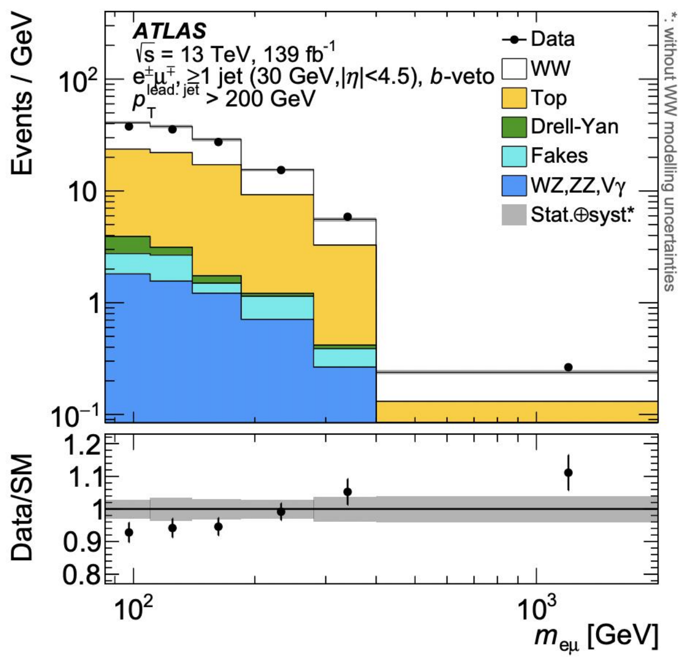


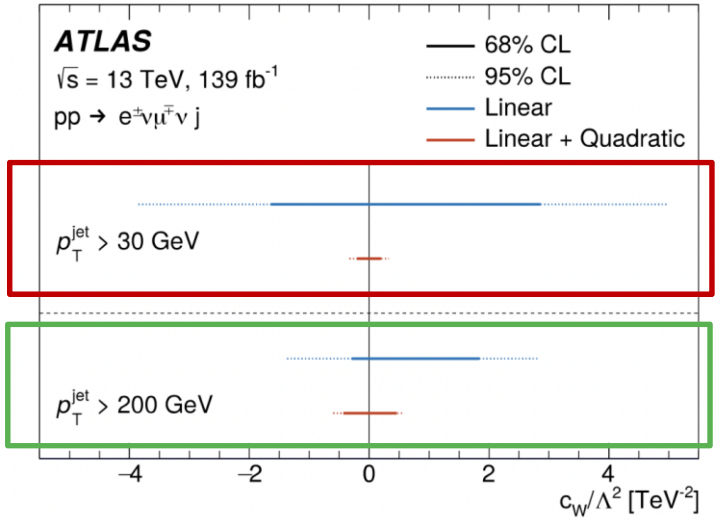
ENHANCING LINEAR SMEFT SENSITIVITY

[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

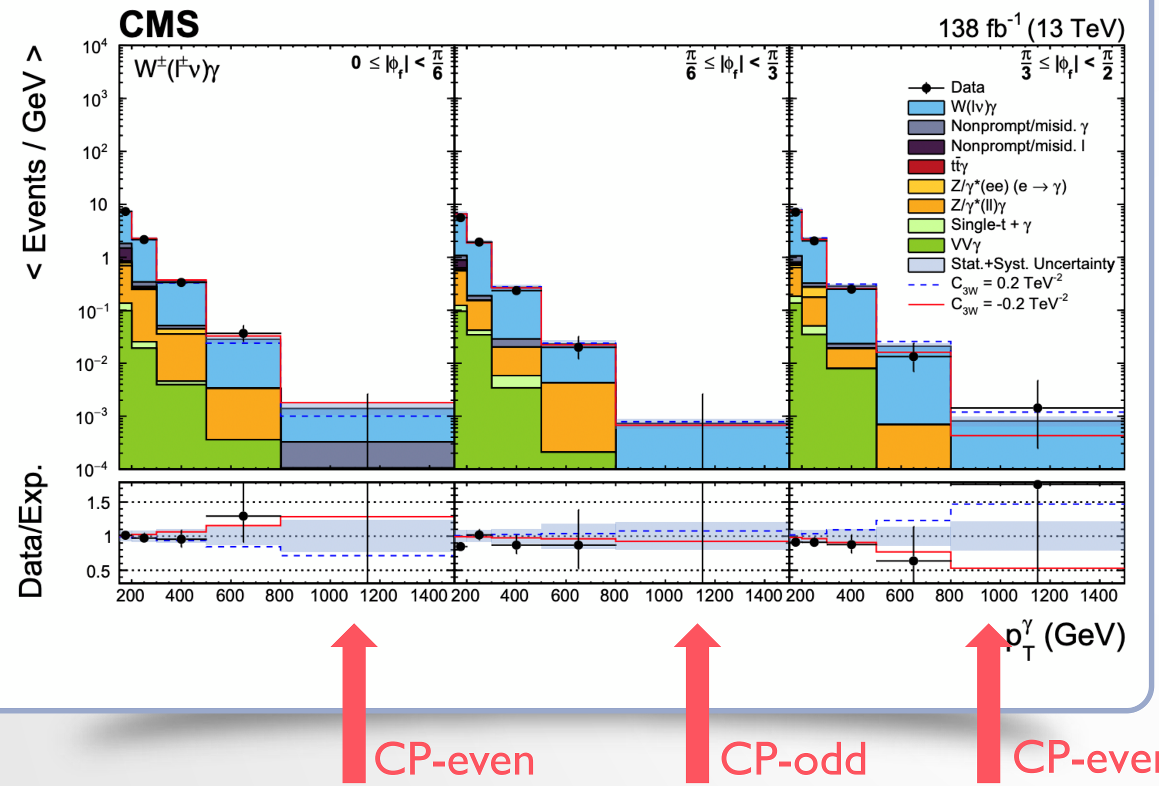
- Why the jet requirement? BSM interference cancels among helicities

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

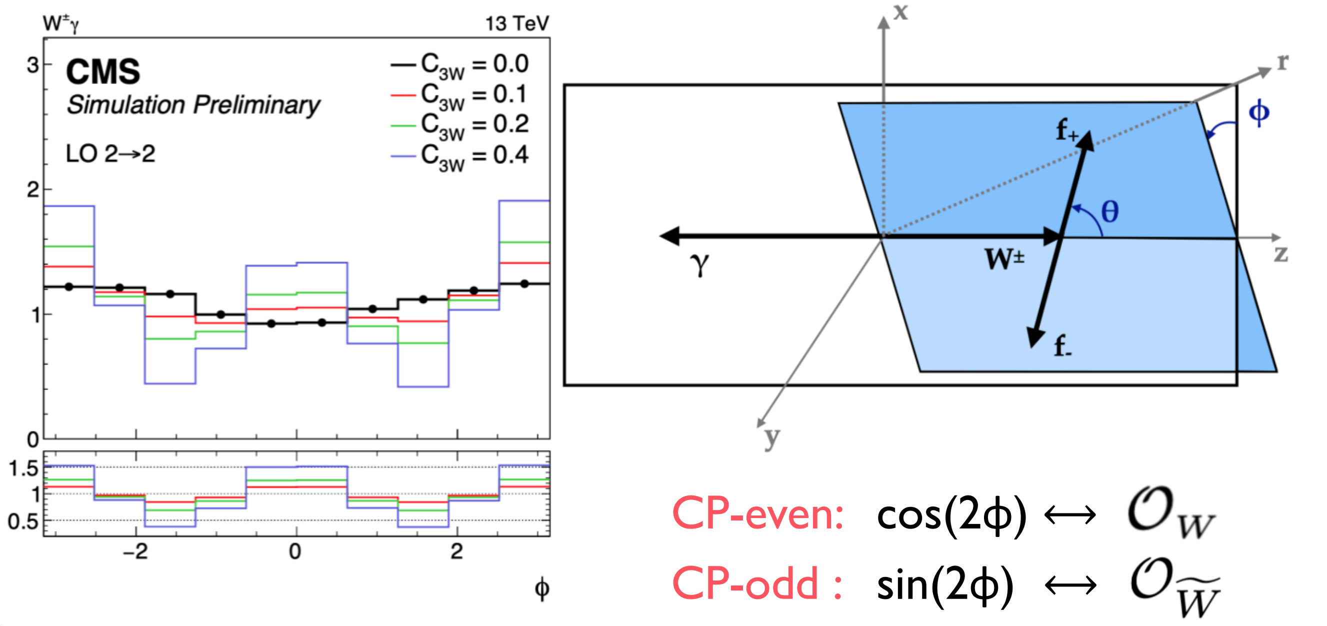
- hard jet (p_T> 200 GeV) requirement changes helicity composition

LINEAR SMEFT SENSITIVITY IN Wɣ PRODUCTION

SMP-20-005



Interference resurrection arXiv:1708.07823

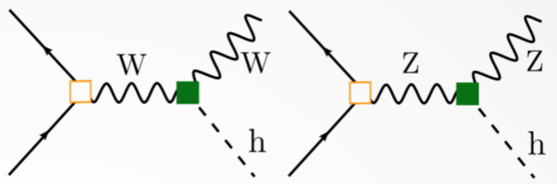


- Boosting to the diboson center-of-mass frame allows to reconstruct decay plan angle ϕ
- It's distribution carries information on BSM effects in the $W_{L/R}$ helicites.
- Binning $p_T(\gamma)$ in ϕ recovers CP structure; facto 5-10: $-0.062 < C_{3W}/\Lambda^2 < 0.053$ TeV^{-2} → Λ_{BSM} ~5 TeV

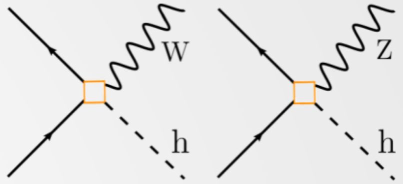
GOALS FOR MACHINE-LEARNING *OF* EFT

[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):

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