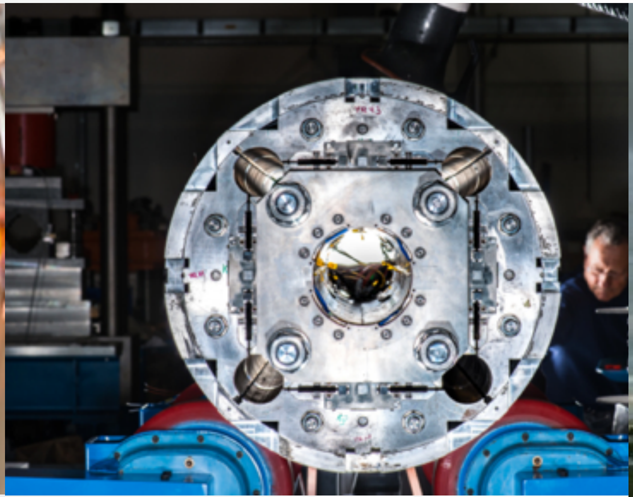
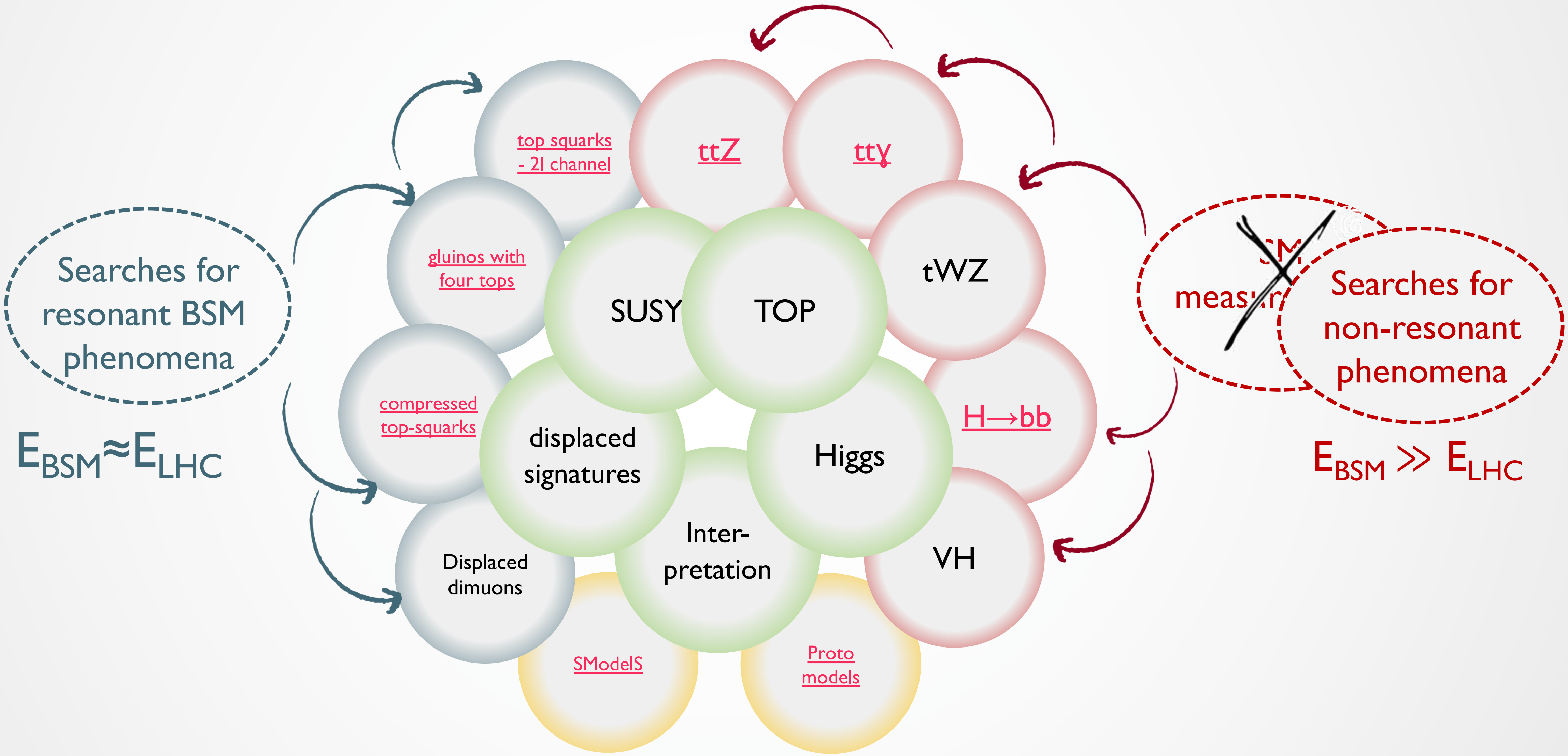
# EXPLORING EFT WITH ML AT THE LHC

R. Schöfbeck (HEPHY Vienna), Sept. 8th, 2022

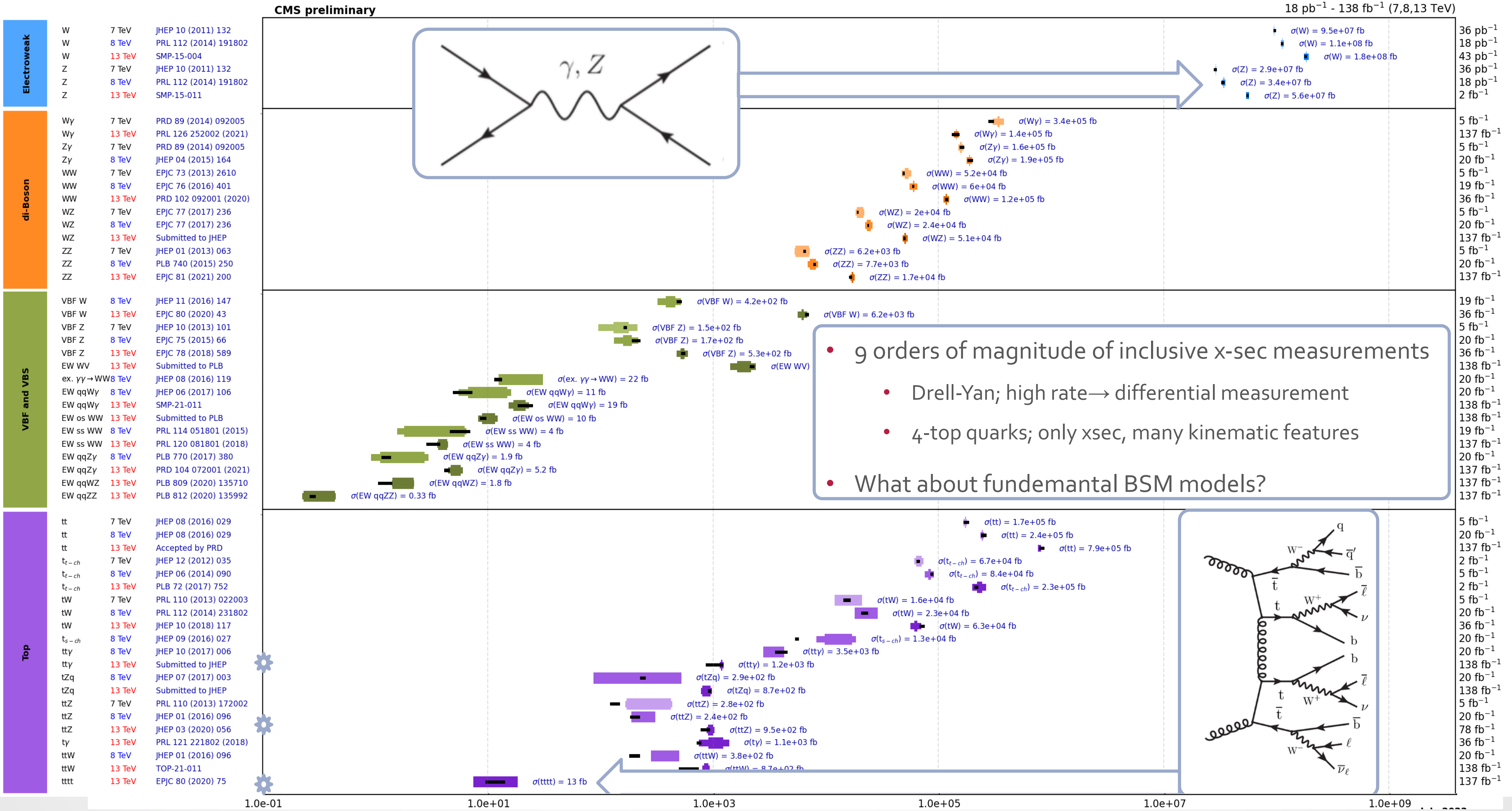



# ACTIVITIES @ HEPHY (CMS DATA ANALYSIS)

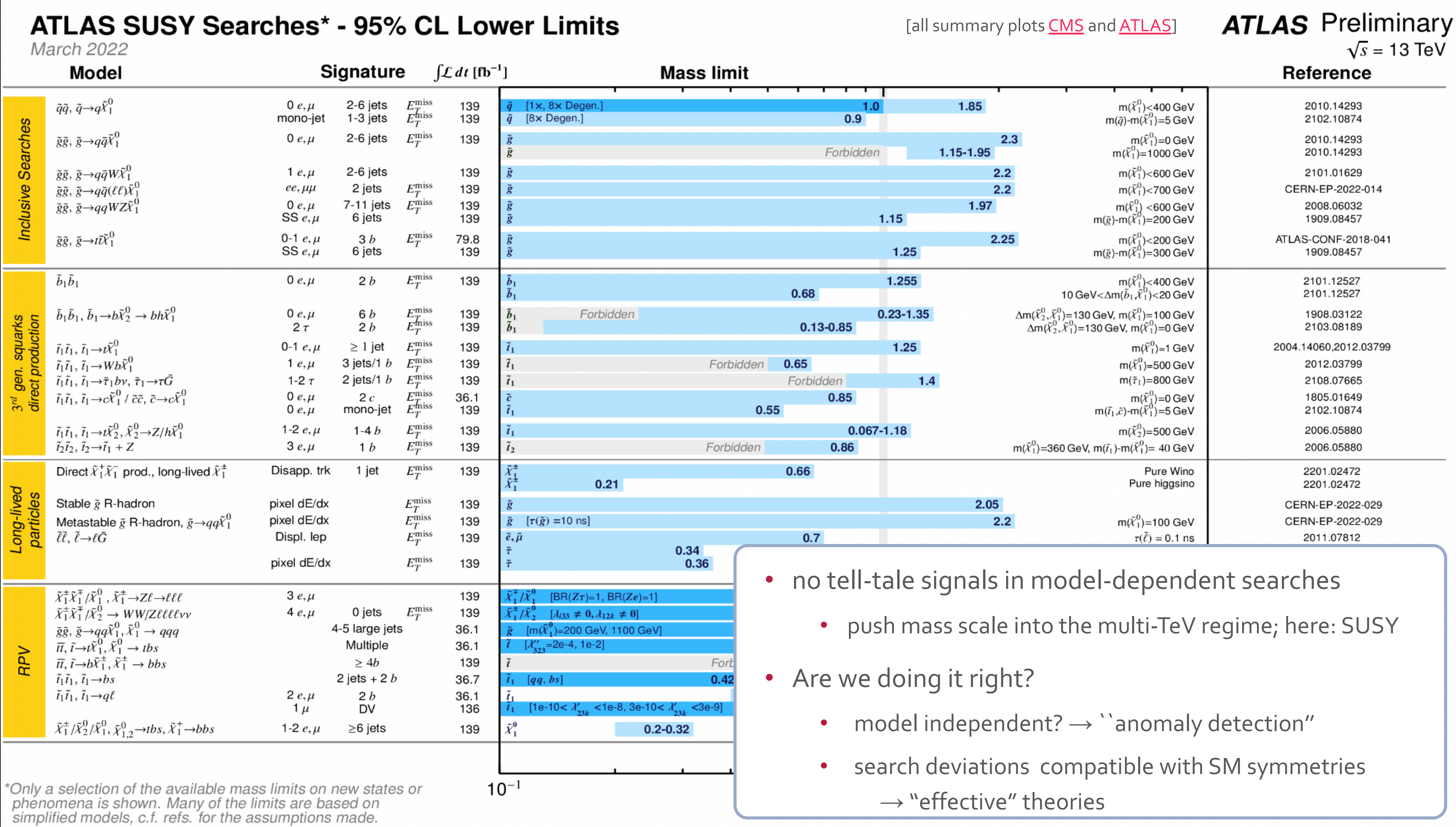


# How well is the Standard Model? – Inclusive cross sections

[all summary plots CMS and ATLAS]




- 9 orders of magnitude of inclusive x-sec measurements
  - Drell-Yan; high rate→ differential measurement
  - 4-top quarks; only xsec, many kinematic features
- What about fundemantal BSM models?




- no tell-tale signals in model-dependent searches
  - push mass scale into the multi-TeV regime; here: SUSY
- Are we doing it right?
  - model independent? → ``anomaly detection"
  - search deviations compatible with SM symmetries
    → "effective" theories

# WHERE TO LOOK AT THE LHC?

- Let's add all terms compatible with the SM symmetries
  - respect SM symmetries: $SU(3)_c \otimes SU(2)_L \otimes U(1)$
  - 59 operators at d=6 [JHEP10(2010)085 ]

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}}^{(4)} + \sum_i \frac{\theta_i}{\Lambda^2} \mathcal{O}_{i,6} + \text{h.c.}$$

$$\Downarrow$$

$$\sigma \sim \left| \mathcal{M}_{\text{SM}} + \theta_i \mathcal{M}_{\text{EFT}}^i \right|^2$$

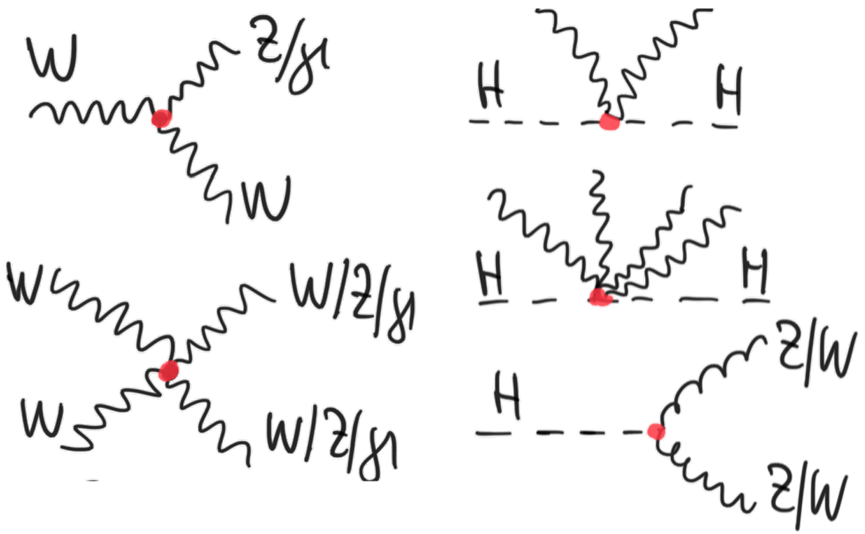
*probability = wave-function²*

$$\Downarrow$$

$$\sigma = \sigma^{\text{SM}} + \boxed{\sum_i \frac{\theta_i}{\Lambda^2} \sigma_i^{\text{int.}}} \boxed{+ \sum_{i,j} \frac{\theta_i\, \theta_j}{\Lambda^4} \sigma_{i,j}^{\text{BSM}}}$$

interference → linear or quadratic effects

$$\mathcal{O}_{\varphi W} \qquad \left(\varphi^\dagger \varphi - \frac{v^2}{2}\right) W_I^{\mu\nu} W_{\mu\nu}^I$$

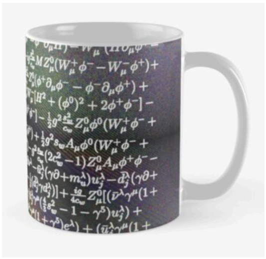


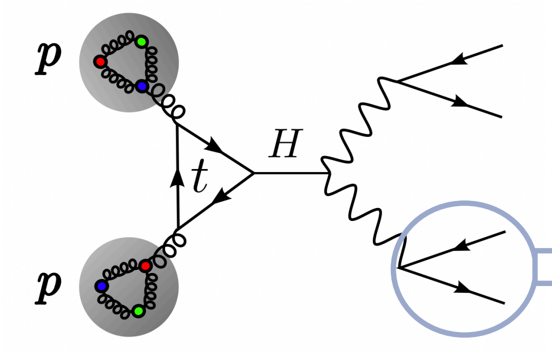

*coupling modifications or new interactions*

- SM-EFT effects are polynomial modifications with varying coefficients over feature space

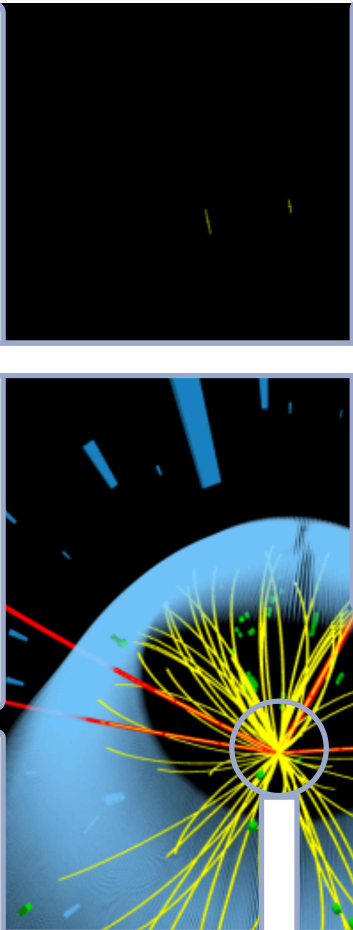
# CONFRONT THE MODEL WITH COLLIDER DATA

## Hard scatter event

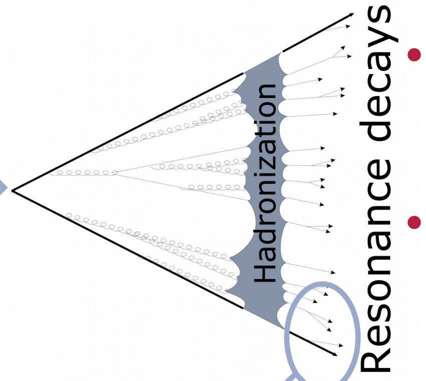





- initial state – standard model
- intermediate (quantum/virtual) particle, "beyond" SM in the language of QFTs

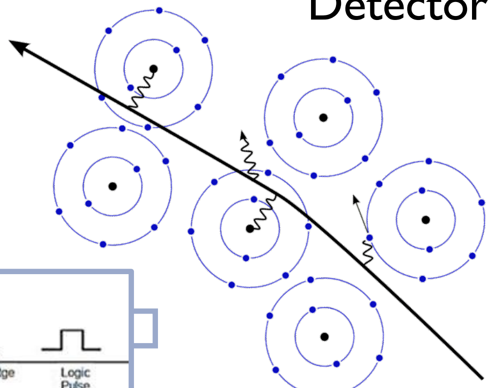


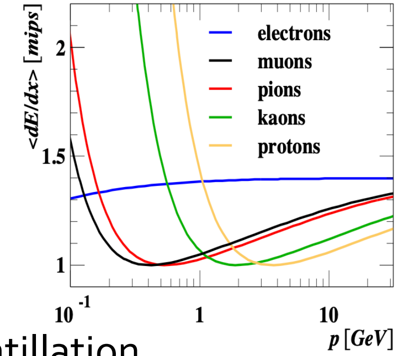
## Shower, hadronization & decays



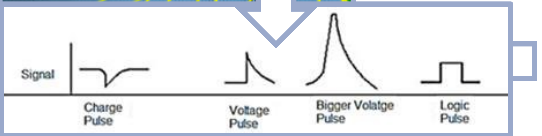

- parton shower
- factorization / hadronization models
- decay branchings, calculated & measured
- *partly empirical*

## Detector interactions





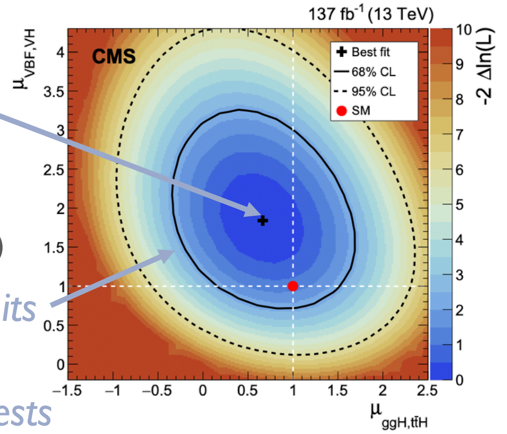

- ionization (Bethe-Bloch), szintillation, brem., transition-radiation … *mostly empirical*

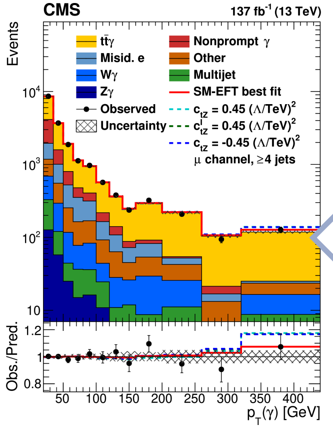



*Maximum-likelihood estimate*

$\theta_{MLE}$ = $\mathrm{argmax}_{\theta}\ L(\mathcal{D},\theta)$

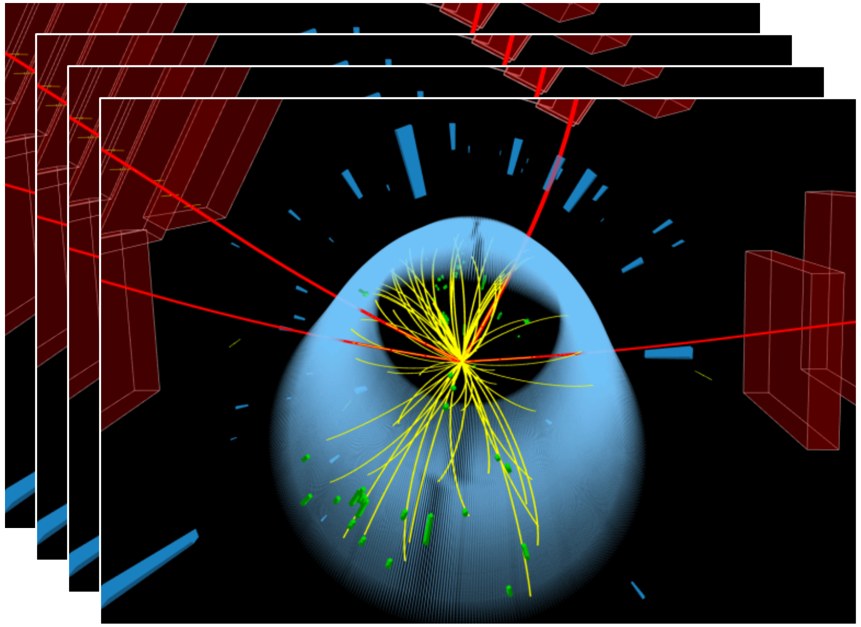
*Confidence limits based on likelihood-ratio tests*



# THE LIKELIHOOD FUNCTION

arxiv:1503.0x7622



Neyman-Pearson Lemma: The *likelihood* ***ratio*** *test statistic is optimal*

data-set with feature vectors **x**

$$q(\mathcal{D}) = \log \frac{L(\mathcal{D}|\boldsymbol{\theta}\ )}{L(\mathcal{D}|\boldsymbol{\theta}_0)} \sim \sum_{i=1}^{N} \log \frac{p(\boldsymbol{x}_i|\boldsymbol{\theta}\ )}{p(\boldsymbol{x}_i|\boldsymbol{\theta}_0)}$$

theory parameters

Likelihood ratio "trick": label two values of **θ** with z=0,1

$$L = \int \mathrm{d}\boldsymbol{x} \underbrace{\sum_{z\in\{0,1\}} p(\boldsymbol{x}, z)}_{\text{training sample}} \left(z - \hat{f}(\boldsymbol{x})\right)^2$$

training sample

truth (supervised)

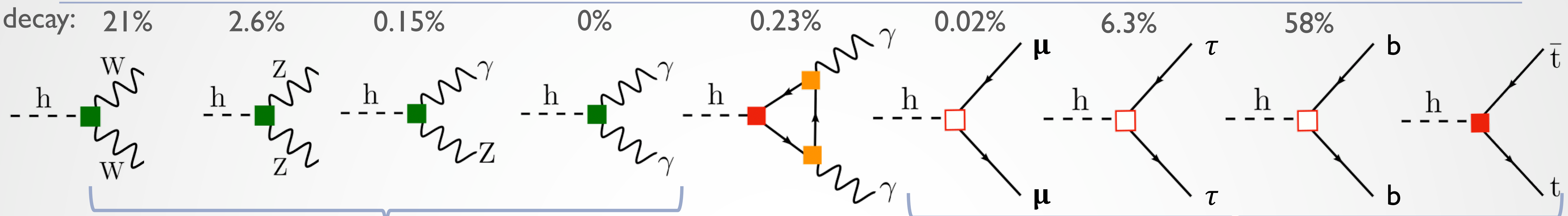
classifier

$$f^*(\boldsymbol{x}) = \frac{p(\boldsymbol{x},1)}{p(\boldsymbol{x},1)+p(\boldsymbol{x},0)} = \frac{1}{1+\boxed{r(\boldsymbol{x})}}$$

provides a close-to optimal test statistic

→ straightforward to obtain from supervised learning

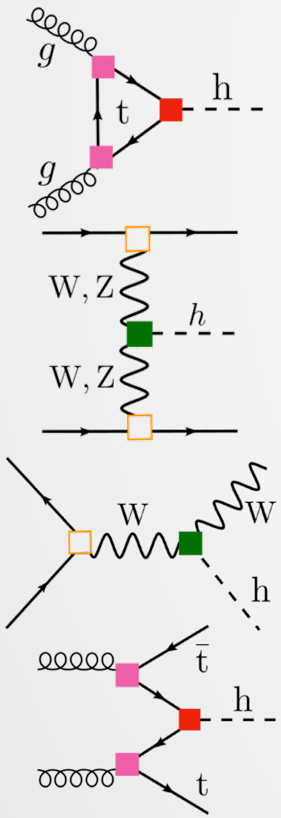
# DISENTANGLING SM-EFT IN THE HIGGS-SECTOR




Higgs – Boson couplings

Higgs – Boson loop induced

Higgs – Fermion couplings




45 pb gluon fusion

3.5 pb vector-boson fusion

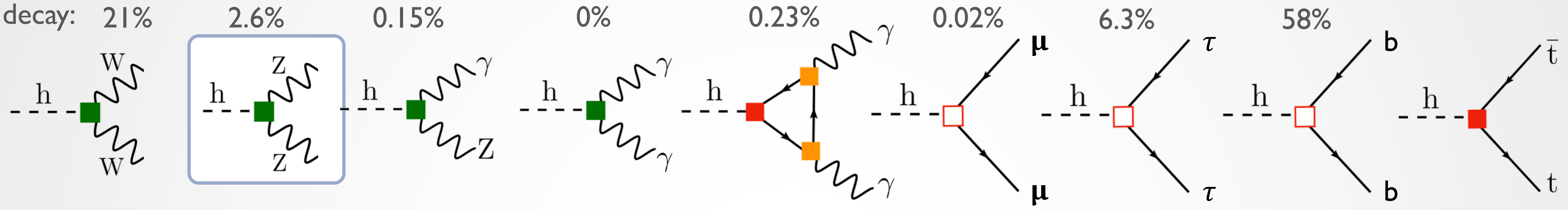
1.2 pb associate W

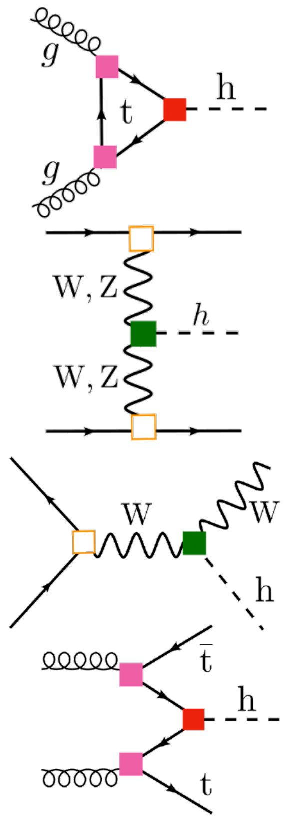
0.5 pb ttH

Higgs production modes with their SM-EFT couplings

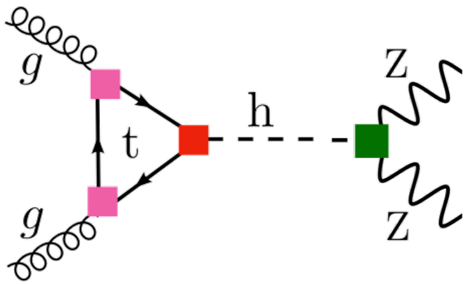
# THE HIGGS IN THE GOLDEN CHANNEL

EPJC 80 (2020) 957




$H \to ZZ^{*} \to 4l$

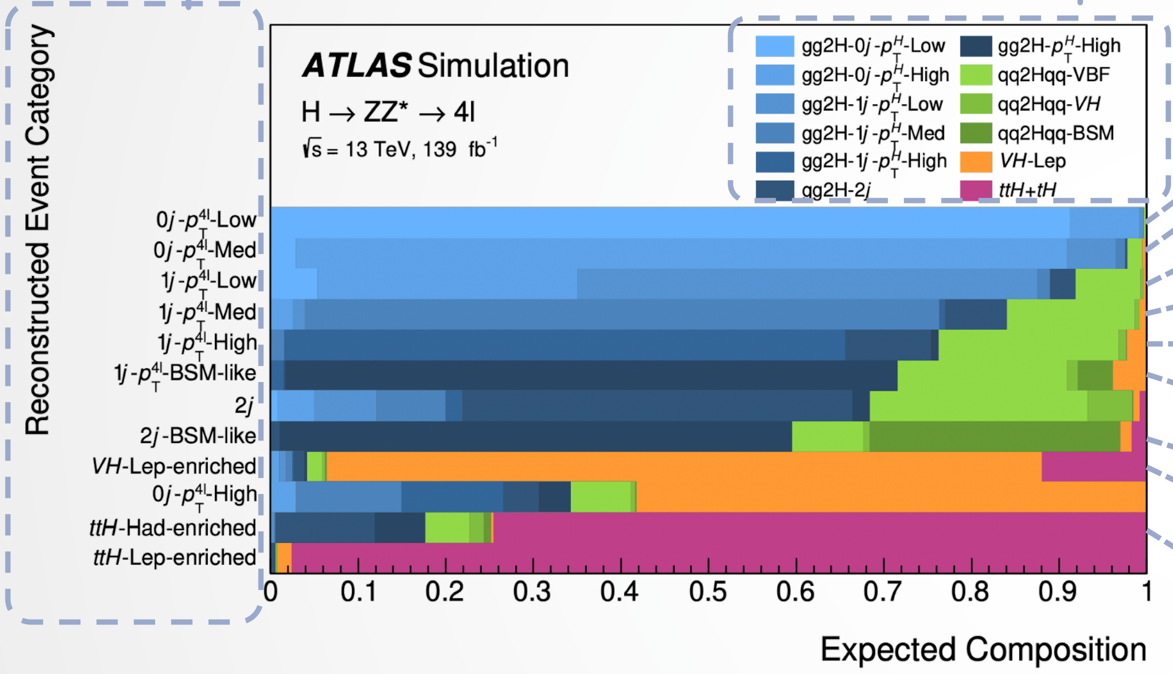



- example #1: $ZZ^{*}$ decay channel in all production modes
- experimentally clean ("golden channel")
- 10 = 5 (+5 CP odd) operators: $c_{HW}$, $c_{HB}$, $c_{HW}$, $c_{uH}$, $c_{HWB}$
- attempt to optimally disentangle production modes

# THE HIGGS IN THE GOLDEN CHANNEL

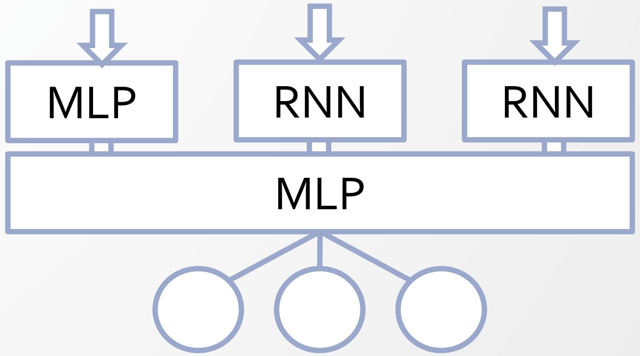
EPJC 80 (2020) 957

- Reconstructed bins contain a mixture of production channels and backgrounds (mostly ZZ*)



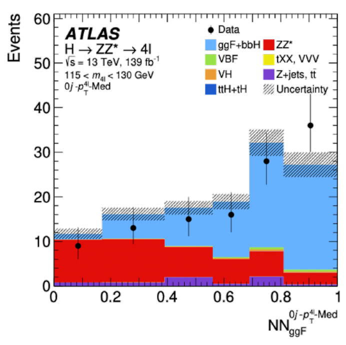

| Processes | MLP | Lepton RNN | Jet RNN |
|---|---|---|---|
| ggF, $ZZ^*$ | $p_T^{4\ell}, D_{ZZ^*}, m_{12}, m_{34}, \lvert\cos\theta^*\rvert, \cos\theta_1, \phi_{ZZ}$ | $p_T^\ell, \eta_\ell$ | - |
| ggF, VBF, $ZZ^*$ | $p_T^{4\ell}, p_T^j, \eta_j, \Delta R_{4\ell j}, D_{ZZ^*}$ | $p_T^\ell, \eta_\ell$ | - |
| ggF, VBF, $ZZ^*$ | $p_T^{4\ell}, p_T^j, \eta_j, E_T^{miss}, \Delta R_{4\ell j}, D_{ZZ^*}, \eta_{4\ell}$ | $p_T^\ell, \eta_\ell$ | - |
| ggF, VBF | $p_T^{4\ell}, p_T^j, \eta_j, E_T^{miss}, \Delta R_{4\ell j}, \eta_{4\ell}$ | $p_T^\ell, \eta_\ell$ | - |
| ggF, VBF, $VH$ | $m_{jj}, p_T^{4\ell jj}$ | $p_T^\ell, \eta_\ell$ | $p_T^j, \eta_j$ |
| ggF, VBF | $\eta_{ZZ}^{Zepp}, p_T^{4\ell jj}$ | $p_T^\ell, \eta_\ell$ | $p_T^j, \eta_j$ |
| $VH$, $ttH$ | $N_{jets}, N_{b\text{-jets},70\%}, E_T^{miss}, H_T$ | $p_T^\ell$ | - |
| ggF, $ttH$, $tXX$ | $p_T^{4\ell}, m_{jj}, \Delta R_{4\ell j}, N_{b\text{-jets},70\%},$ | $p_T^\ell, \eta_\ell$ | $p_T^j, \eta_j$ |

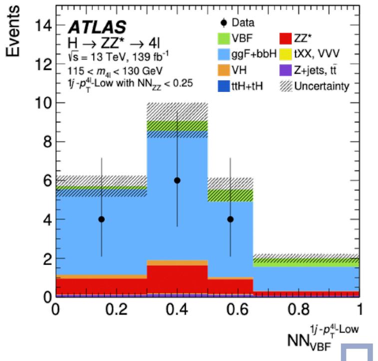
- ML is used to separate production modes in each category
- per reco-channel: NNs trained with 2-7 observables
  - combine with RNNs (LSTMs) using variable-length jets and leptons
  - common network layer for multiclassification in e.g., ggF, VBF, ZZ*



# THE HIGGS IN THE GOLDEN CHANNEL

EPJC 80 (2020) 957




Likelihood = prod. of Poissonians

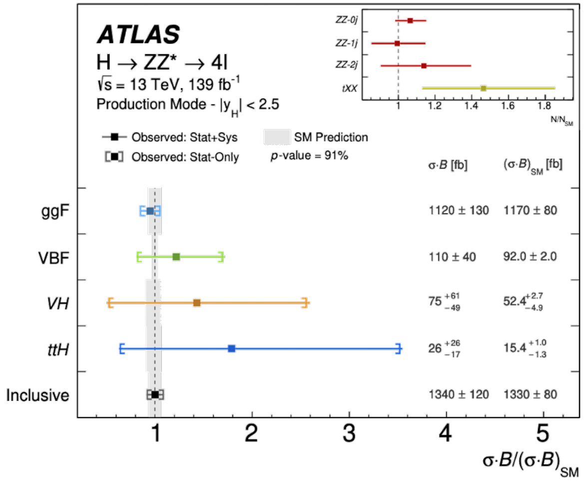
auxiliary measurements

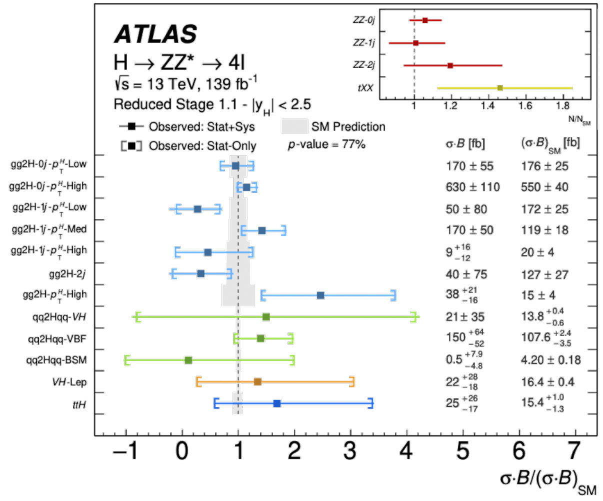
$$L(\mathcal{D}|\boldsymbol{\sigma},\boldsymbol{\nu}) = \prod_{j=1}^{N_{\mathrm{cat.}}} \prod_{i=1}^{N_{\mathrm{bin}}^{(j)}} \mathrm{Pois}\left( N_{i,j} \middle| \mathcal{L} \sum_{p=1}^{N_{\mathrm{prod}}} \sigma_{\mathrm{SM}}^{(p)} \mathcal{B}_{\mathrm{SM}}^{4\ell} A_{i,j}^{(p)}(\boldsymbol{\nu}) + B_{i,j}(\boldsymbol{\nu}) \right) \times \prod_{m=1}^{N_{nuis.}} C_m(\boldsymbol{\nu})$$

Log-likelihood ratio test statistic

$$q(\boldsymbol{\sigma}) = -2\log \frac{L(\mathcal{D}|\boldsymbol{\sigma},\hat{\hat{\boldsymbol{\nu}}}_{\boldsymbol{\sigma}})}{L(\mathcal{D}|\hat{\boldsymbol{\sigma}},\hat{\boldsymbol{\nu}})}$$

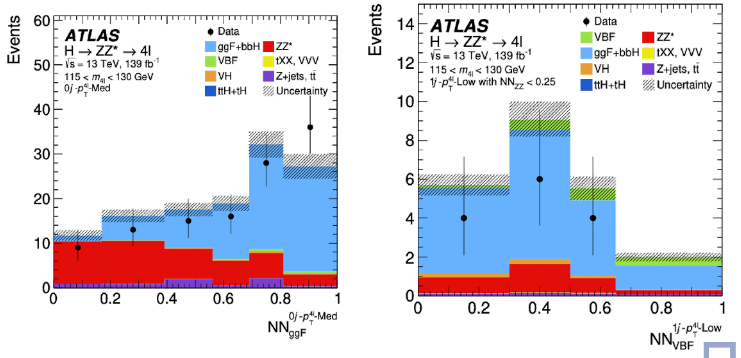
(profiled, to deal with nuisances)



# THE HIGGS IN THE GOLDEN CHANNEL

EPJC 80 (2020) 957



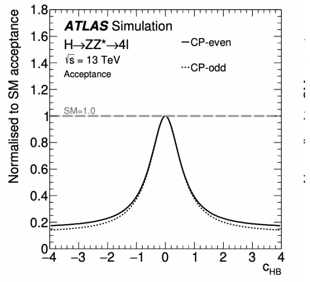

signal-strength modifiers

defined at the fiducial level

production | decay | acceptance

$$\mu^{(p)}(\boldsymbol{\theta}) = \frac{\sigma^{(p)}(\boldsymbol{\theta})}{\sigma^{(p)}_{\mathrm{SM}}} \times \frac{\mathcal{B}^{4\ell}(\boldsymbol{\theta})}{\mathcal{B}^{4\ell}_{\mathrm{SM}}} \times \frac{A(\boldsymbol{\theta})}{A_{\mathrm{SM}}}$$

universal!




Likelihood = prod. of Poissonians

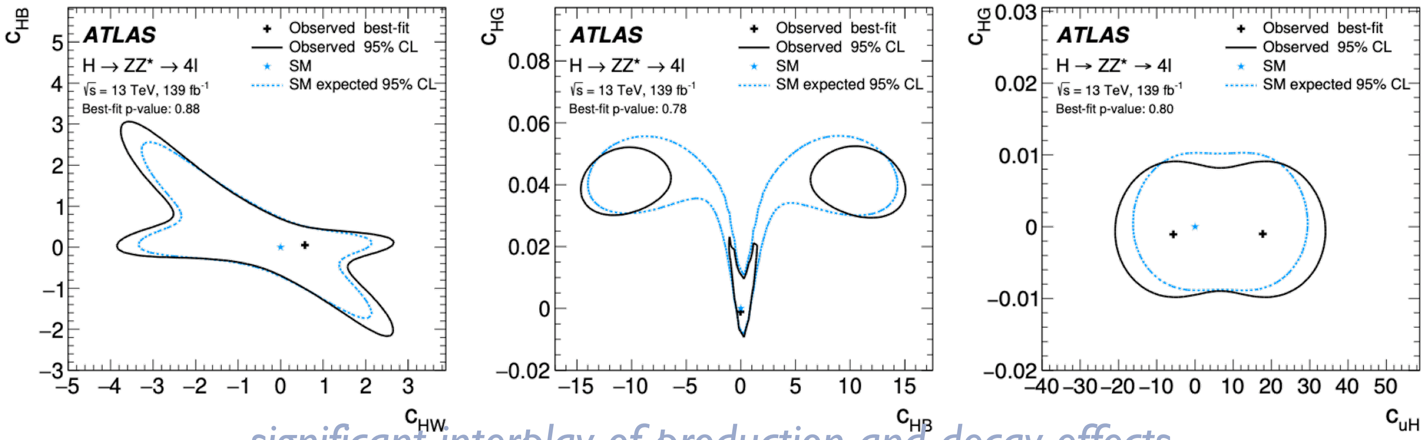
auxiliary measurements

$$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \prod_{j=1}^{N_{\mathrm{cat.}}} \prod_{i=1}^{N^{(j)}_{\mathrm{bin}}} \mathrm{Pois}\left(N_{i,j} \,\middle|\, \mathcal{L} \sum_{p=1}^{N_{\mathrm{prod}}} \mu^{(p)}(\boldsymbol{\theta})\sigma^{(p)}_{\mathrm{SM}}\,\mathcal{B}^{4\ell}_{\mathrm{SM}}\,A^{(p)}_{i,j}(\boldsymbol{\nu}) + B_{i,j}(\boldsymbol{\nu})\right) \times \prod_{m=1}^{N_{nuis.}} C_m(\boldsymbol{\nu})$$

Log-likelihood ratio test statistic

$$q(\boldsymbol{\theta}) = -2\log\frac{L(\mathcal{D}|\boldsymbol{\theta},\hat{\hat{\boldsymbol{\nu}}}_{\boldsymbol{\theta}})}{L(\mathcal{D}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\nu}})}$$

(profiled, to deal with nuisances)



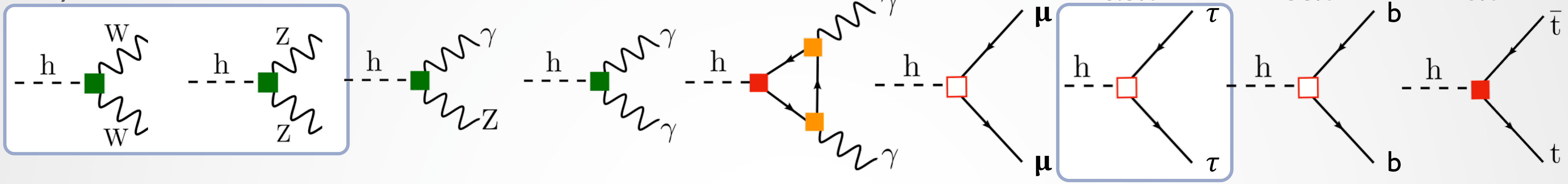

*significant interplay of production and decay effects*

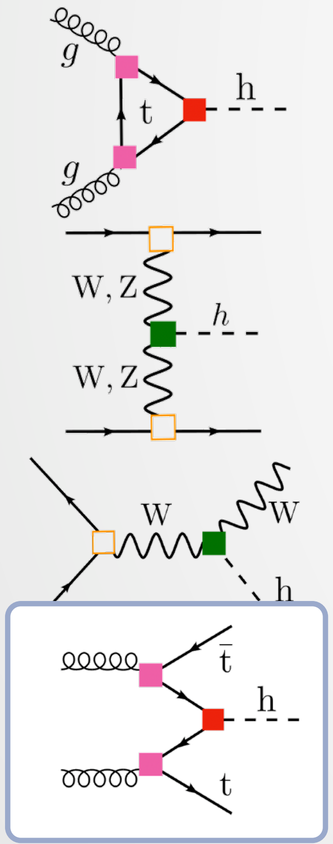
*learn "only" the likelihood ratio of different SM production modes*

# TTH IN THE MULTILEPTON CHANNEL

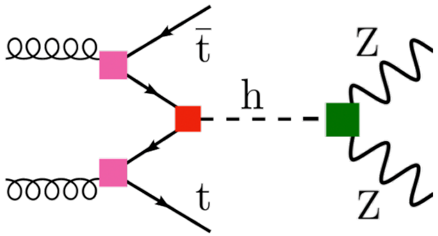
JHEP (submitted)

decay: 21% 2.6% 0.15% 0% 0.23% 0.02% 6.3% 58% 0%



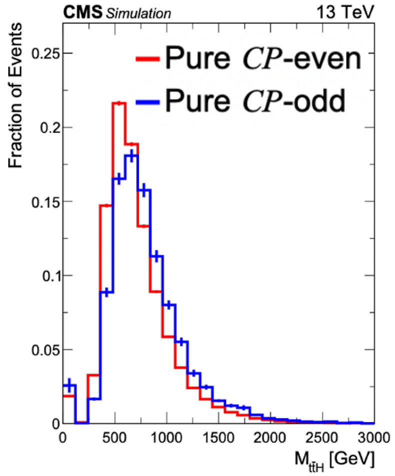




ttH multilepton




- example #2: t(t)H multilepton in 2ℓSS+0$\tau$, 2ℓSS+1$\tau$, 3ℓ final states
- 3 DNNs for signal/background multi-classification
- targets t-t-H Yukawa coupling (■) in $\kappa$-framework
- in SM-EFT: "CP" structure (complex phase) of $HH^\dagger \bar{q}_p u_r \tilde{H}$

- use ML for separating CP-even vs. odd effects
  - gradient-BDT XGBoost
  - 38 input features (kinematic properties)

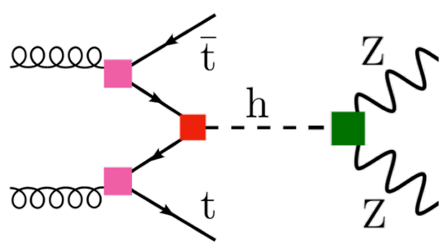


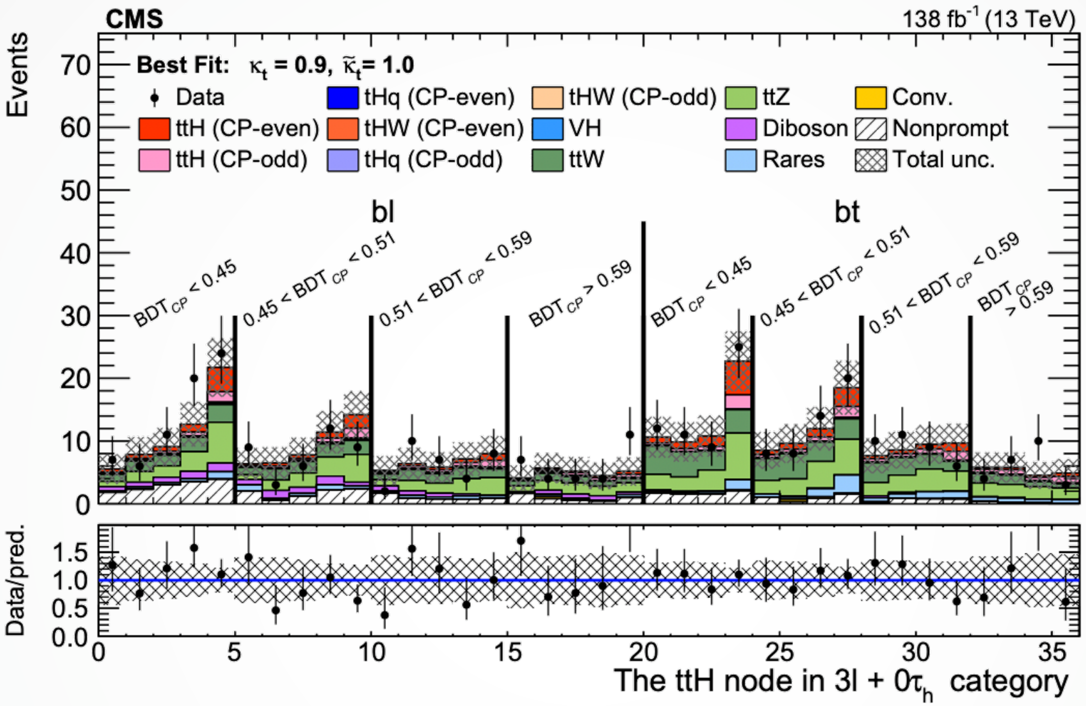

+ 37 other observables

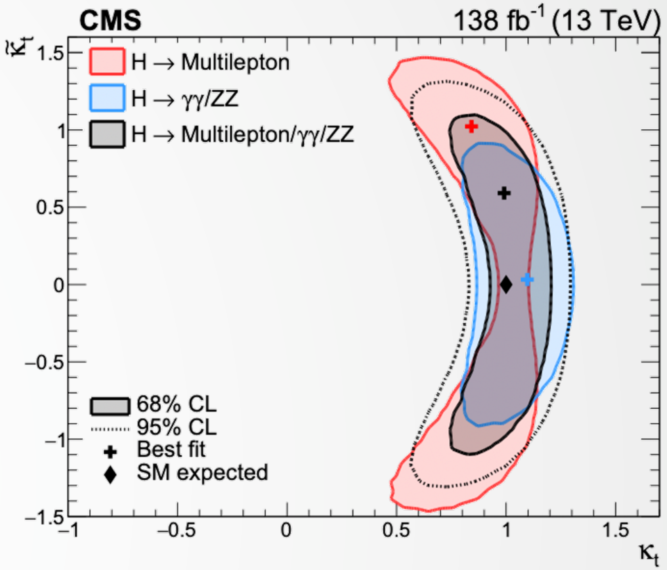
# TTH IN THE MULTILEPTON CHANNEL

JHEP (submitted)

ttH multilepton



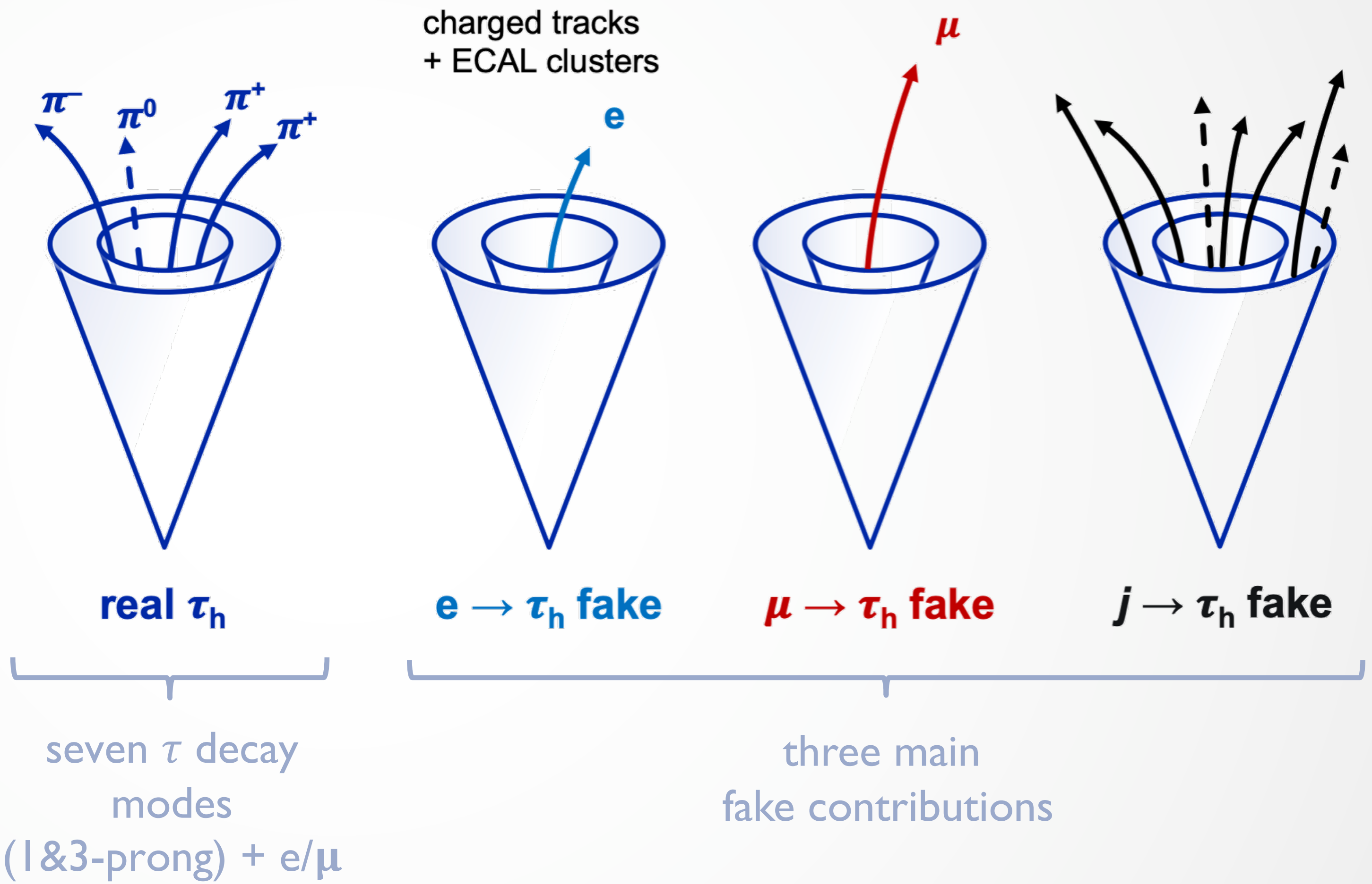




- BDT exploits the likelihood trick to obtain CP even/odd fraction from the data
- limits on deviations of the t-t-H interaction ( $\kappa_t$ , $\widetilde{\kappa}_t$ ) including combinations with other final states
- example of learning "of" SM-EFT effects
- issue: large top backgrounds from ttZ and ttW in all measurement regions → combine sectors!
- $\tau$ lepton ID performance has significant impact

# INTERLUDE: CLASSIFICATION WITH DEEPTAU

JINST 17 (2022) P07023
sketch from Izaak Neuteligns

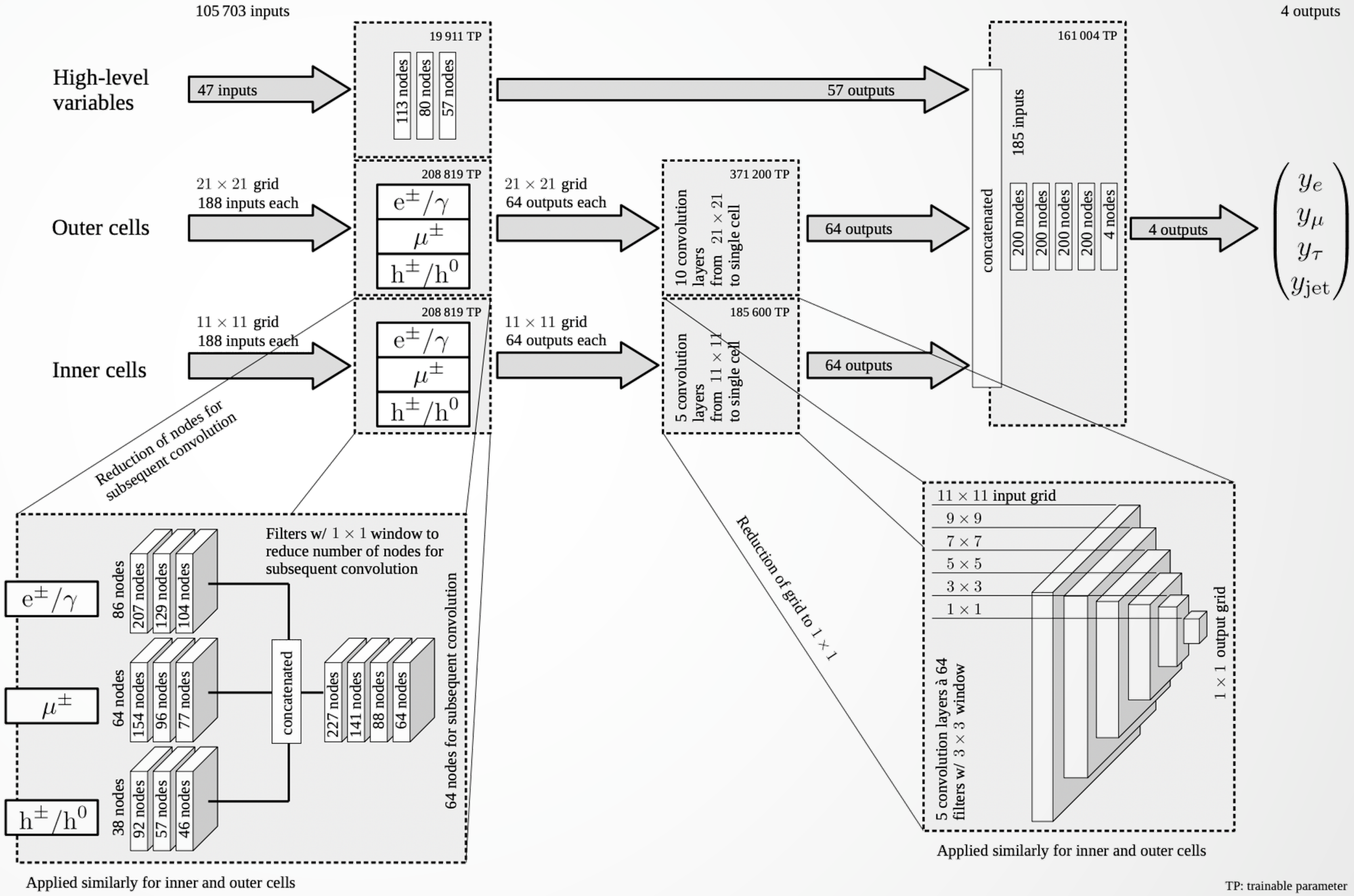
- $\tau$ leptons in the detector
  - 5 hadronic + 2 leptonic decays
  - three main fake contributions
- new [DeepTau] identification algorithm classifies $\tau_h$ modes
- similar to
  - ATLAS $\tau$ ID [using RNNs]
  - [DeepJet] for g/c/b/uds/leptons identification and
  - [DeepAK8] for t/W/Z/H decays
- high level *candidate-features (fully connected)* and *feature maps* on two grids of all particles in the vicinity in *convolutional layers*
- 140M $\tau$ candidates, 690hrs



# INTERLUDE: CLASSIFICATION WITH DEEPTAU

JINST 17 (2022) P07023

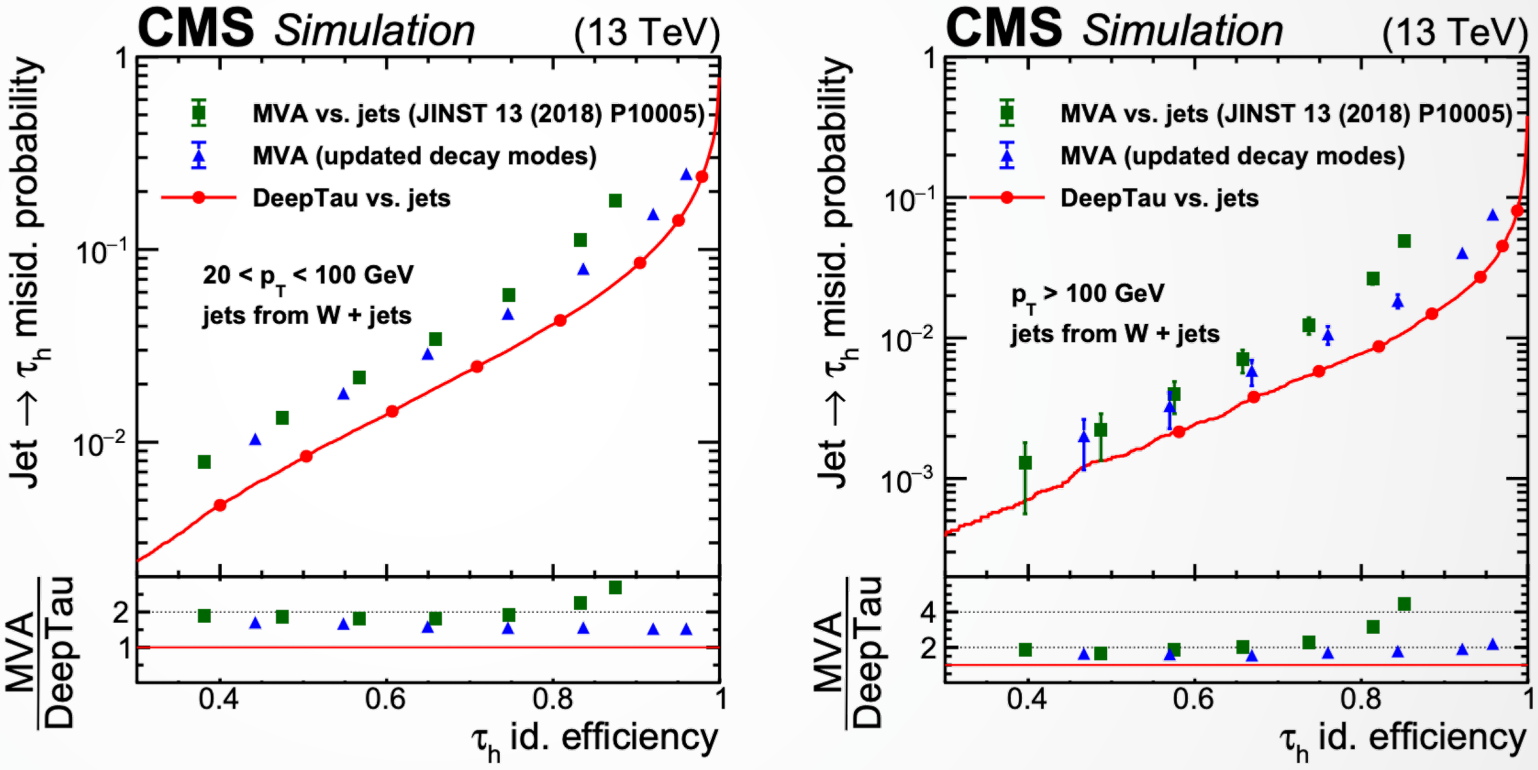
- $\tau$ leptons in the detector
  - 5 hadronic + 2 leptonic decays
  - three main fake contributions
- new [DeepTau] identification algorithm classifies $\tau_h$ modes
- similar to
  - ATLAS $\tau$ ID [using RNNs]
  - [DeepJet] for g/c/b/uds/leptons identification and
  - [DeepAK8] for t/W/Z/H decays
- high level *candidate-features (fully connected)* and *feature maps* on two grids of all particles in the vicinity in *convolutional layers*
- 140M $\tau$ candidates, 690hrs



# INTERLUDE: CLASSIFICATION WITH DEEPTAU

JINST 17 (2022) P07023

- $\tau$ leptons in the detector
  - 5 hadronic + 2 leptonic decays
  - three main fake contributions
- new [DeepTau] identification algorithm classifies $\tau_h$ modes
- similar to
  - ATLAS $\tau$ ID [using RNNs]
  - [DeepJet] for g/c/b/uds/leptons identification and
  - [DeepAK8] for t/W/Z/H decays
- high level *candidate-features (fully connected)* and *feature maps* on two grids of all particles in the vicinity in *convolutional layers*
- 140M $\tau$ candidates, 690hrs



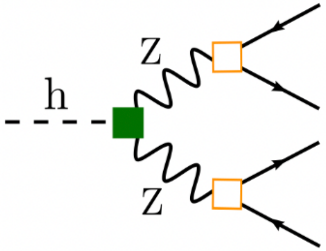

*improvement of background suppression by almost factor 2 when compared to previous $\tau$ MVA not using the convolutional layers*

# RECENT SM-EFT RESULTS (SELECTION!)

4l JHEP 07 (2021) 005

ATL-PHYS-PUB-2021-010

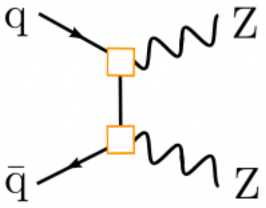
## $H\rightarrow Z^{*}Z$




$H\rightarrow 4l$ EPJC 80 (2020) 957

$H\rightarrow 4l$ PRD 104, 052004 (2021)

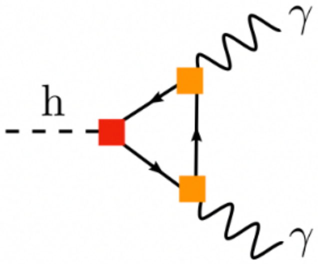
## ZZ




PRD 97 (2018) 032005

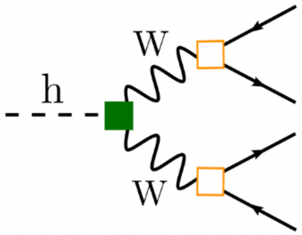
EPJC 81 (2021) 200

## $H\rightarrow\gamma\gamma$




arxiv:2202.00487

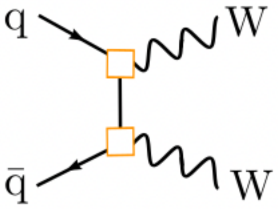
## $H\rightarrow W^{*}W$




$H\rightarrow WW$, eμ

EPJC 82 (2022) 622

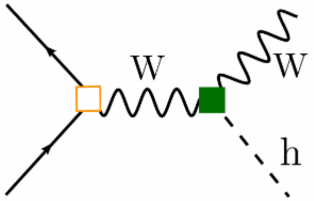
## $W^{\pm}W^{\mp}$




$W^{\pm}W^{\mp}$

PRD 102, 092001 (2020)

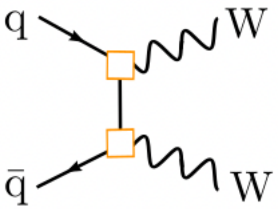
## W/Z+H (H→bb)




resolved EPJC 81 (2021) 178

boosted PLB 816 (2021) 136204

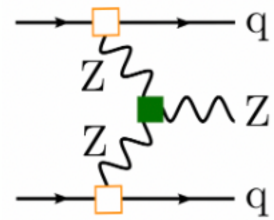
## $W^{\pm}W^{\mp}$ (+ ≥ 1 jet)




JHEP 06 (2021) 003

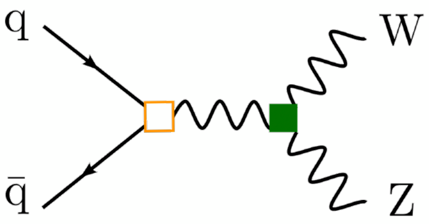
PRD102, 092001 (2020)

## VBF Z + jj



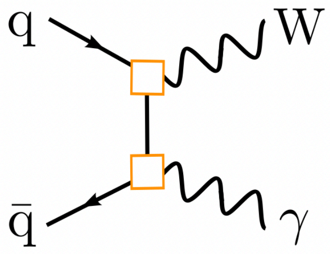

EPJC 81(2021)163

## WZ




CMS-SMP-20-0014

## Wγ




PRD sub.

CMS-SMP-20-005

# RECENT SM-EFT RESULTS (SELECTION!)

4l JHEP 07 (2021) 005

ATL-PHYS-PUB-2021-010

## $H \to Z^*Z$

$$
\begin{array}{ll}
O_{uH} & HH^\dagger \bar{q}_p u_r \tilde{H} \\
O_{HG} & HH^\dagger G^A_{\mu\nu} G^{\mu\nu A} \\
O_{HW} & HH^\dagger W^l_{\mu\nu} W^{\mu\nu l} \\
O_{HB} & HH^\dagger B_{\mu\nu} B^{\mu\nu} \\
O_{HWB} & HH^\dagger \tau^l W^l_{\mu\nu} B^{\mu\nu}
\end{array}
$$

+CP odd

$H \to 4l$ EPJC 80 (2020) 957

$H \to 4l$ PRD 104, 052004 (2021)

## ZZ

aTGC

PRD 97 (2018) 032005

EPJC 81 (2021) 200

## $H \to \gamma\gamma$

$$
\begin{array}{ll}
O_{uH} & HH^\dagger \bar{q}_p u_r \tilde{H} \\
O_{HG} & HH^\dagger G^A_{\mu\nu} G^{\mu\nu A} \\
O_{HW} & HH^\dagger W^l_{\mu\nu} W^{\mu\nu l} \\
O_{HB} & HH^\dagger B_{\mu\nu} B^{\mu\nu} \\
O_{HWB} & HH^\dagger \tau^l W^l_{\mu\nu} B^{\mu\nu}
\end{array}
$$

arxiv:2202.00487

## $H \to W^*W$

HC framework

CP even/odd

($O_{HW}$, $O_{HB}$ + CP odd)

$H \to WW$, eμ

EPJC 82 (2022) 622

## $W^\pm W^\mp$

$$\mathcal{O}_{WWW} = \frac{c_{WWW}}{\Lambda^2} W_{\mu\nu} W^{\nu\rho} W_\rho{}^\mu$$

$$\mathcal{O}_W = \frac{c_W}{\Lambda^2} (D^\mu\Phi)^\dagger W_{\mu\nu} (D^\nu\Phi)$$

$$\mathcal{O}_B = \frac{c_B}{\Lambda^2} (D^\mu\Phi)^\dagger B_{\mu\nu} (D^\nu\Phi)$$

+CP odd

$W^\pm W^\mp$

PRD 102, 092001 (2020)

## W/Z+H (H→bb)

$$
\begin{array}{ll}
c_{HWB} & O_{HWB} = H^\dagger \tau^I H W^I_{\mu\nu} B^{\mu\nu} \\
c_{HW} & O_{HW} = H^\dagger H W^I_{\mu\nu} W^{\mu\nu}_I \\
c_{Hq3} & O^{(3)}_{Hq} = (H^\dagger i \overleftrightarrow{D}^I_\mu H)(\bar{q}_p \tau^I \gamma^\mu q_r) \\
c_{Hq1} & O^{(1)}_{Hq} = (H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{q}_p \gamma^\mu q_r) \\
c_{Hu} & O_{Hu} = (H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{u}_p \gamma^\mu u_r) \\
c_{Hd} & O_{Hd} = (H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{d}_p \gamma^\mu d_r) \\
c_{dH} & O_{dH} = (H^\dagger H)(\bar{q} d H)
\end{array}
$$

resolved EPJC 81 (2021) 178

boosted PLB 816 (2021) 136204

## $W^\pm W^\mp$ (+ ≥ 1 jet)

$$\mathcal{O}_{WWW} = \frac{c_{WWW}}{\Lambda^2} W_{\mu\nu} W^{\nu\rho} W_\rho{}^\mu$$

$$\mathcal{O}_W = \frac{c_W}{\Lambda^2} (D^\mu\Phi)^\dagger W_{\mu\nu} (D^\nu\Phi)$$

$$\mathcal{O}_B = \frac{c_B}{\Lambda^2} (D^\mu\Phi)^\dagger B_{\mu\nu} (D^\nu\Phi)$$

JHEP 06 (2021) 003

PRD102, 092001 (2020)

## VBF Z + jj

$$\mathcal{O}_{WWW} = \frac{c_{WWW}}{\Lambda^2} W_{\mu\nu} W^{\nu\rho} W_\rho{}^\mu$$

$$\mathcal{O}_W = \frac{c_W}{\Lambda^2} (D^\mu\Phi)^\dagger W_{\mu\nu} (D^\nu\Phi)$$

+CP odd

EPJC 81(2021)163

## WZ

$$\mathcal{O}_{WWW} = \frac{c_{WWW}}{\Lambda^2} W_{\mu\nu} W^{\nu\rho} W_\rho{}^\mu$$

$$\mathcal{O}_W = \frac{c_W}{\Lambda^2} (D^\mu\Phi)^\dagger W_{\mu\nu} (D^\nu\Phi)$$

+CP odd

CMS-SMP-20-0014

## Wγ

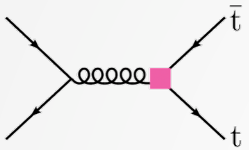
$$\mathcal{O}_{WWW} = \frac{c_{WWW}}{\Lambda^2} W_{\mu\nu} W^{\nu\rho} W_\rho{}^\mu$$

PRD sub.

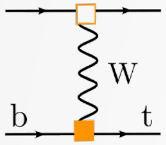
CMS-SMP-20-005

# TOP AND DIBOSON SECTORS

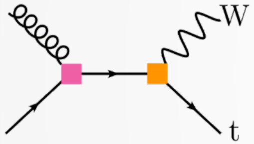
tt
811 pb



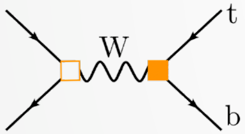

t (t-channel)
217 pb



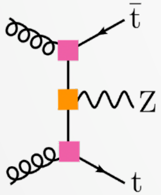

tW
72 pb



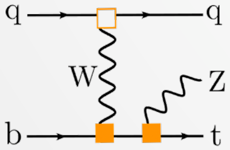

t (s-channel)
10 pb



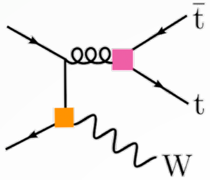

ttZ
1 pb



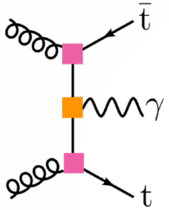

tZq
0.088 pb



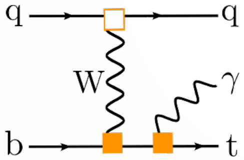

ttW
0.55 pb



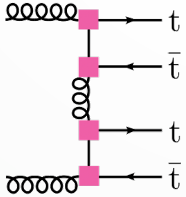

ttɣ
0.80 pb



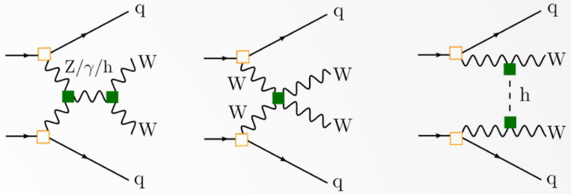

tɣq
0.4 pb



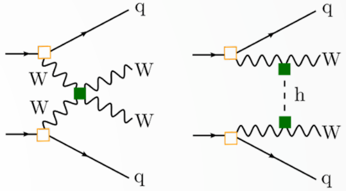

tttt
0.009 pb



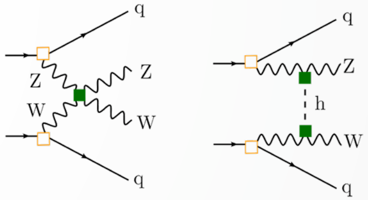

W±W∓
EWK 10 fb



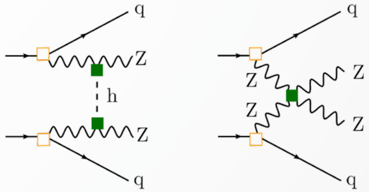

W±W±
EWK 4 fb



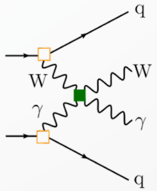

WZ
EWK 2.15 fb



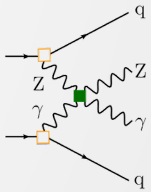

ZZ
EWK 0.82 fb



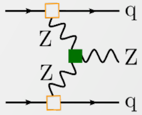

Wɣ
EWK 20.4 fb



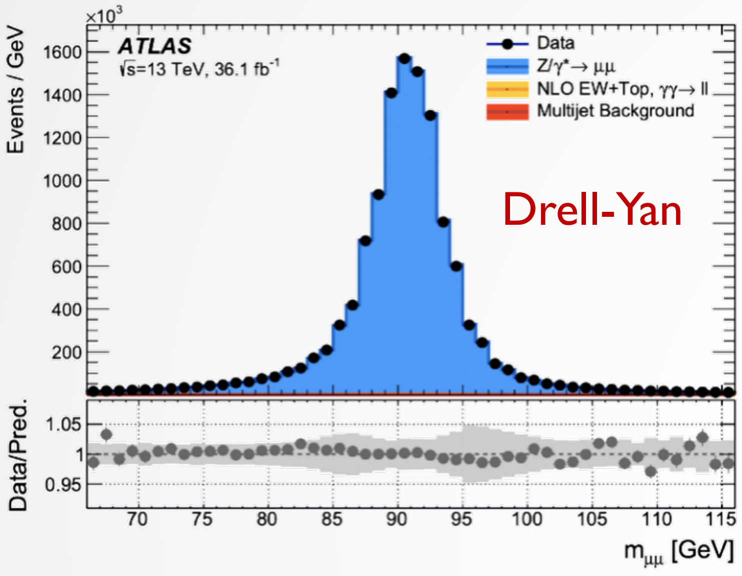

Zɣ
EWK 7.8 fb

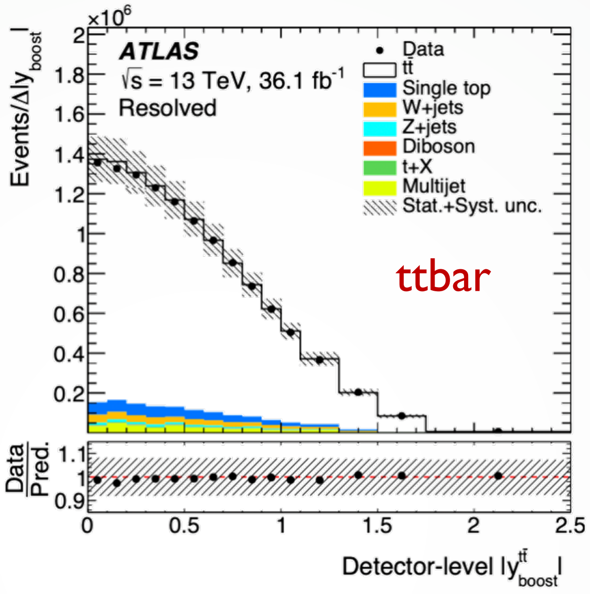


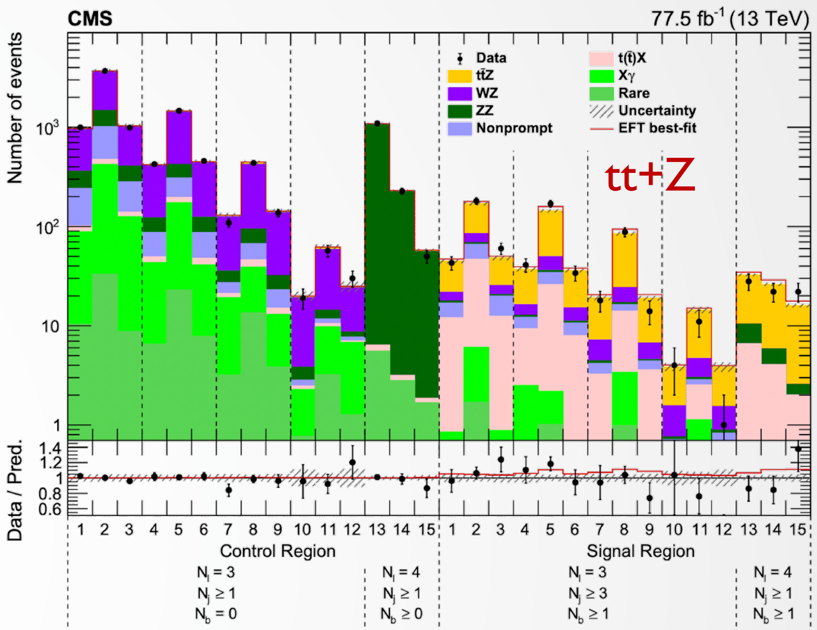

Z
EWK 37.4 fb

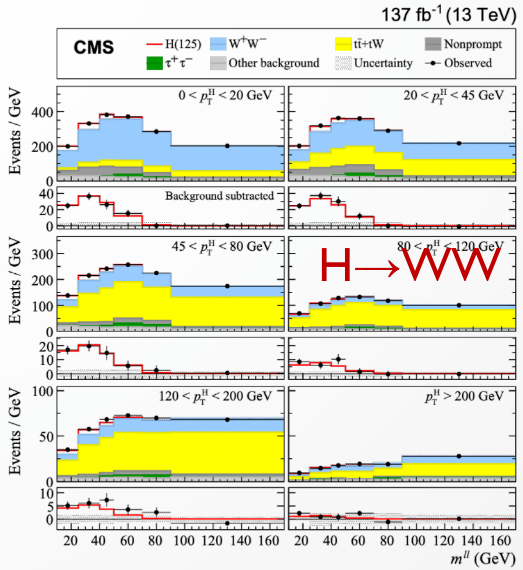


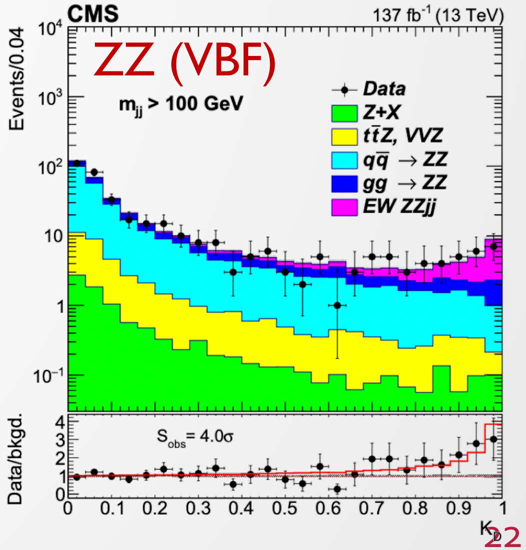
# SM-EFT EFFECTS ARE EVERYWHERE



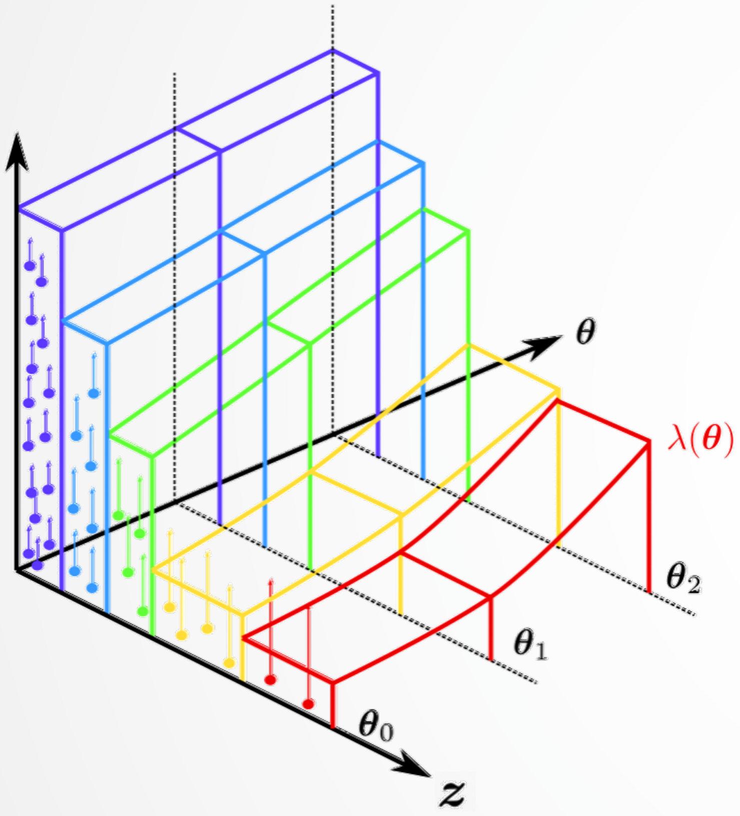




- Solve background correlations like a triangular matrix (i.e. staged):
  - Multi-differential high-dimensional SM-EFT analysis of candles:
    - Drell-Yan, W+Jets, ttbar, single-top (t), etc.
  - Then move to ZH (+ Drell-Yan), WH (+ttbar), H→WW (+WW and ttbar)
- Can go in parallel provided re-interpretation is feasible
  - Needs close-to complete likelihood → a whole separate discussion
- ML versatile tools to optimally extract SM-EFT effects without too much tuning need → parametrized classifiers are an example

# EXPLOITING PARAMETRIZED *SIMULATION* WITH TREE ALGOS




- Quantum field theory: Differential cross section have structure

$$\mathrm{d}\sigma(\boldsymbol{\theta}) \propto |\mathcal{M}_{\mathrm{SM}}(\boldsymbol{z}) + \boldsymbol{\theta}_a \mathcal{M}^a_{\mathrm{BSM}}(\boldsymbol{z})|^2 \mathrm{d}\boldsymbol{z}$$

probability = wave function, squared

  - sampling **z** at a fixed $\boldsymbol{\theta}_0$
  - re-evaluate the likelihood for a few alternative $\boldsymbol{\theta}$
  - fix polynomial coefficients of event weights $w_i(\boldsymbol{\theta})$

$$w_i(\boldsymbol{\theta}) = w_{i,0} + \sum_a w_{i,a}\,\theta_a + \frac{1}{2}\sum_{a,b} w_{ab}\,\theta_a\theta_b = \frac{\sigma(\boldsymbol{\theta}\;)}{\sigma(\boldsymbol{\theta}_0)} \cdot r(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

SM — interference — pure SM-EFT

*interpretation valid at LO*

  - obtain predictions *parametrized* in $\boldsymbol{\theta}$
    from MC simulation run in “forward mode”

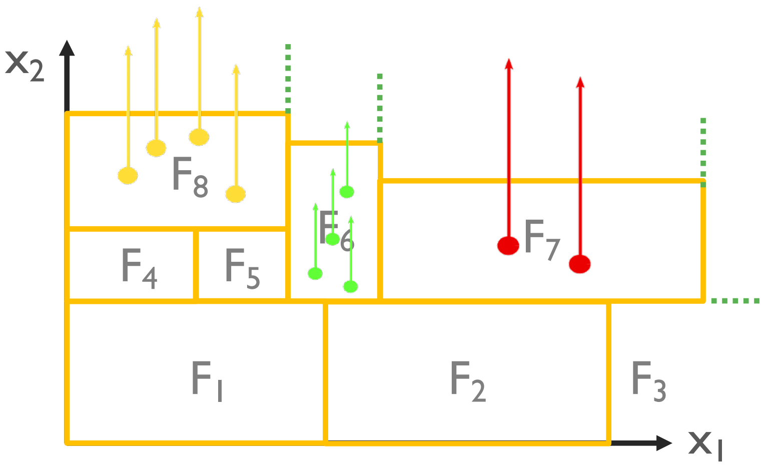
# TREES & BOOSTING

[arXiv:2107.10859, arXiv:2205:12976]

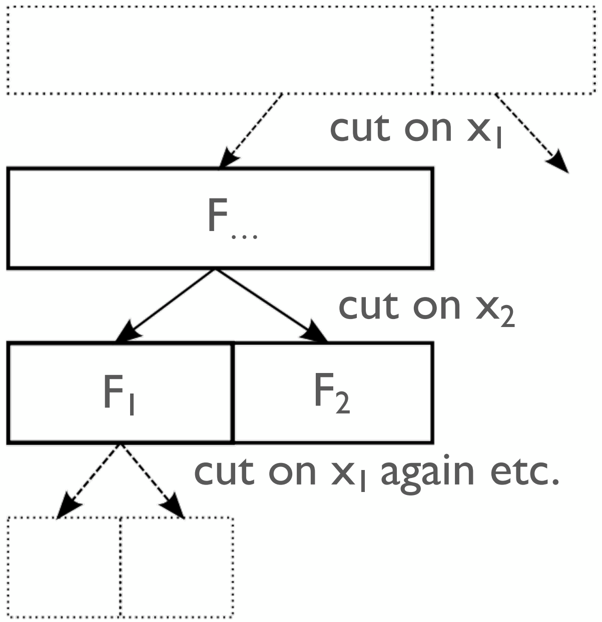
**weak learner**

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

**phase-space partitioning**




**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 $F_j$
- Fitting tree: Optimize "node split positions" on some loss. Trained (e.g. greedily) on the *ensemble*.
  - An axis-aligned tree is limited. Remove the limitation iteratively with "boosting".

# LEARNING MORE WITH TREES

[arXiv:2107.10859, arXiv:2205:12976]

Regress in R, including its the polynomial $\boldsymbol{\theta}$ dependence

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

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

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

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

Tree ansatz for each a, ab:
$F_j(\boldsymbol{\theta})$ polynomial with const. coeff. (per node)

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

Solve for the predictor on the empirical distribution (simulated sample)

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

No trainable parameters in the predictor

Solve for optimal partitioning with greedy CART algorithm

$$L = -\sum_{\boldsymbol{\theta}\in\mathcal{B}} \sum_{j\in\mathcal{J}} \frac{w_j^2(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$$

split only if
$w_j(\boldsymbol{\theta})$ is *positive* $\forall\boldsymbol{\theta}$

We'll find an optimized tree.
→ boost

# CONCRETE SOLUTION: TREE BOOSTING

[arXiv:2107.10859, arXiv:2205:12976]

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

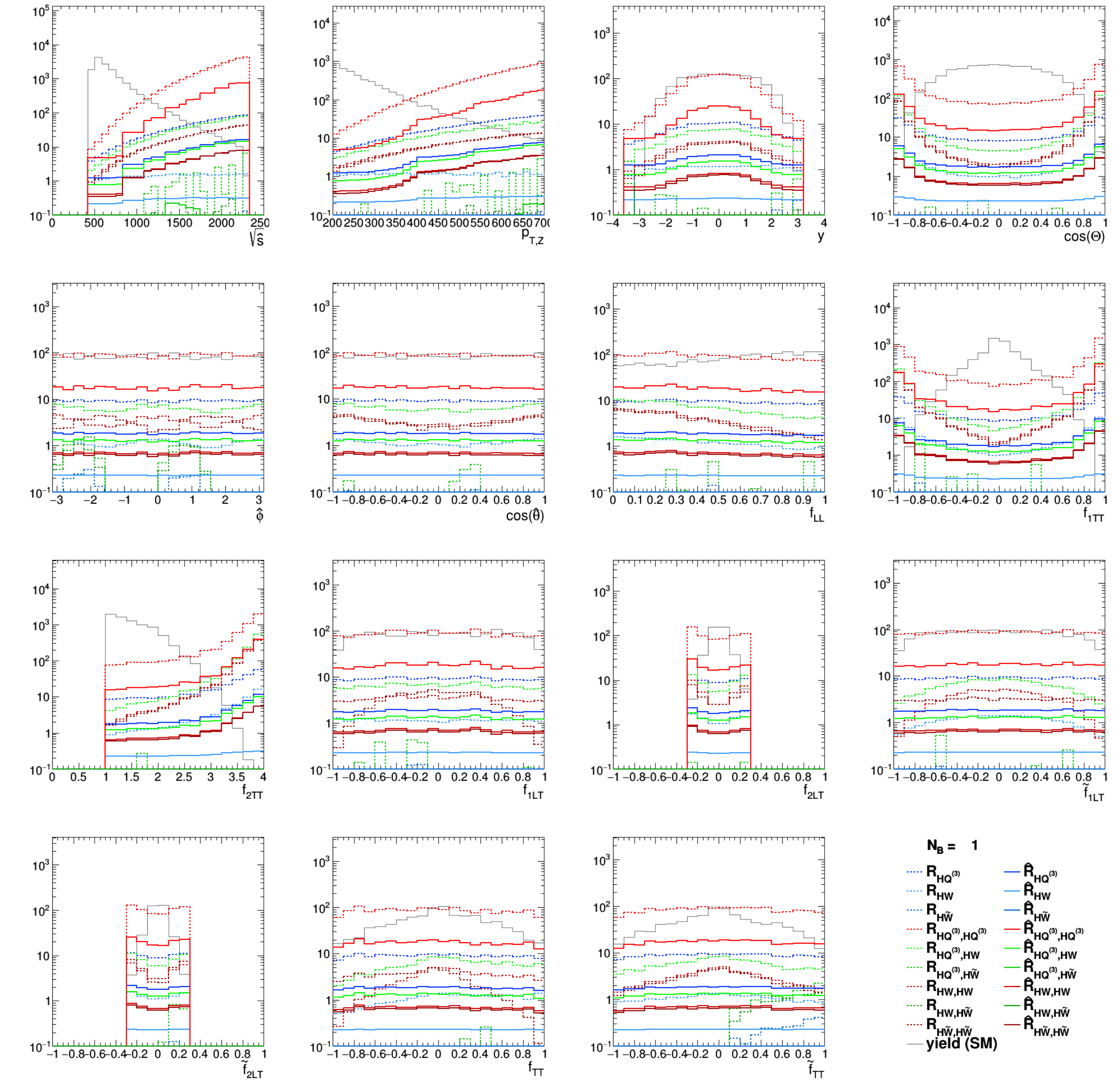
$$\hat{F}^{(b)}(\boldsymbol{x},\boldsymbol{\theta}) = \underbrace{\hat{f}(\boldsymbol{x},\boldsymbol{\theta})}_{\text{current iteration}} + \eta \underbrace{\hat{F}^{(b-1)}(\boldsymbol{x},\boldsymbol{\theta})}_{\text{previous iteration}}$$

- Insert into the loss function:

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

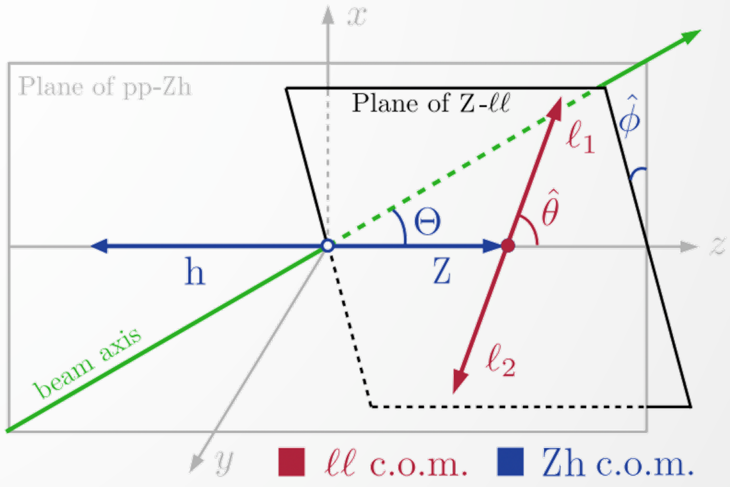
$$w_i^{(b)}(\boldsymbol{\theta}) \quad \rightarrow \quad w_i^{(b-1)}(\boldsymbol{\theta}) - \eta\, w_i^{(b-1)}(\boldsymbol{\theta}_0) \hat{F}^{(b-1)}(\boldsymbol{x}_i,\boldsymbol{\theta})$$

.... perform this iteratively




[arXiv:2107.10859, arXiv:2205:12976]

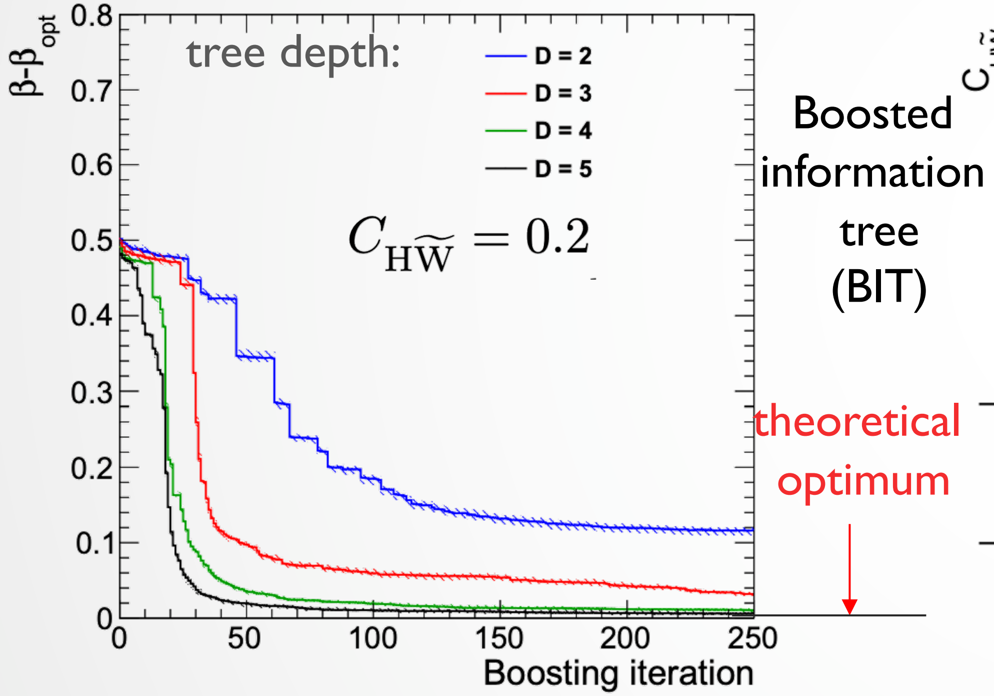
- "Boosted Information tree" (BIT)
- 500 k events, 3 WCs, 9 coefficients
  - 9 minutes training
- Tested in a ZH toy model, and a more realistic Delphes study, including backgrounds

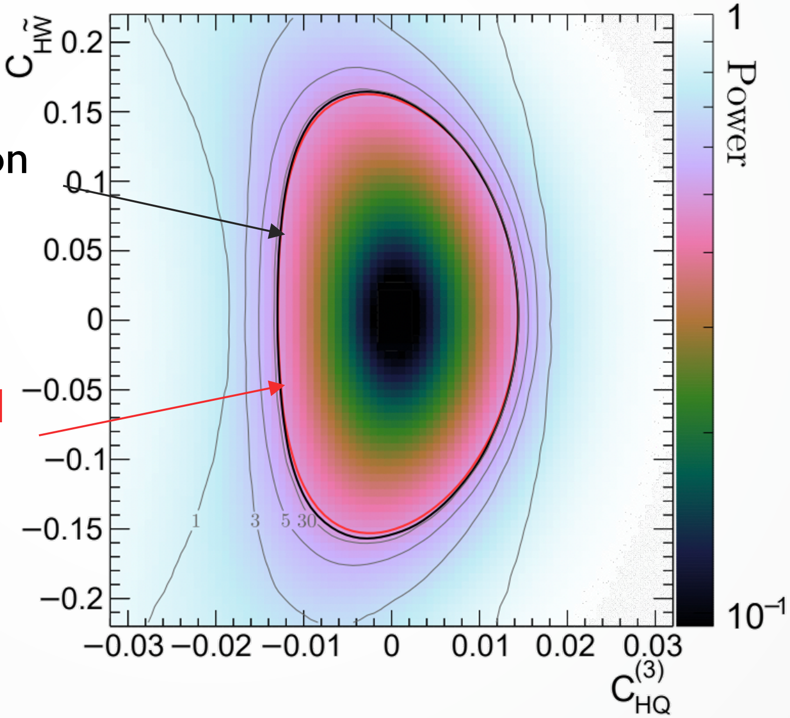


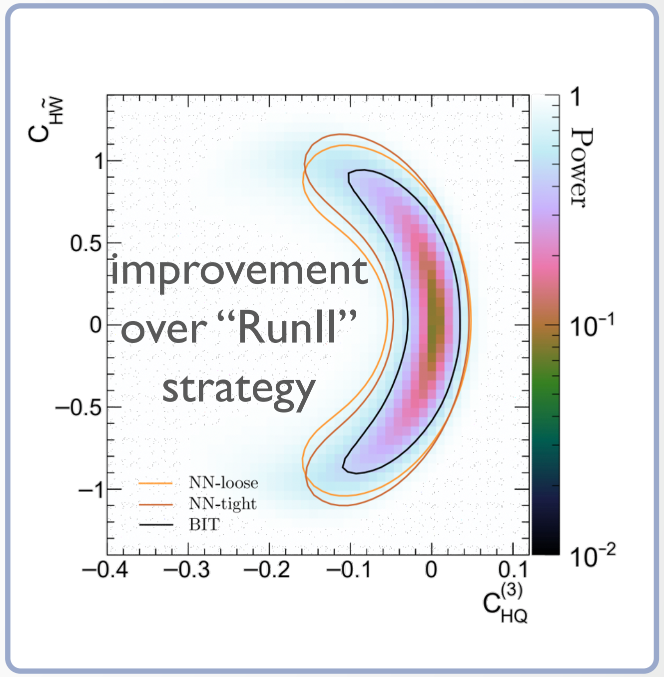
# OPTIMALITY IN TOY DATA

[arXiv:2107.10859, arXiv:2205:12976]

- Test with toy simulation in ZH final state – unbinned likelihood ratio test statistic
- Neyman-Pearson: The LL ratio test statistic has the highest power (1-β) for a given test size (CL=95%)



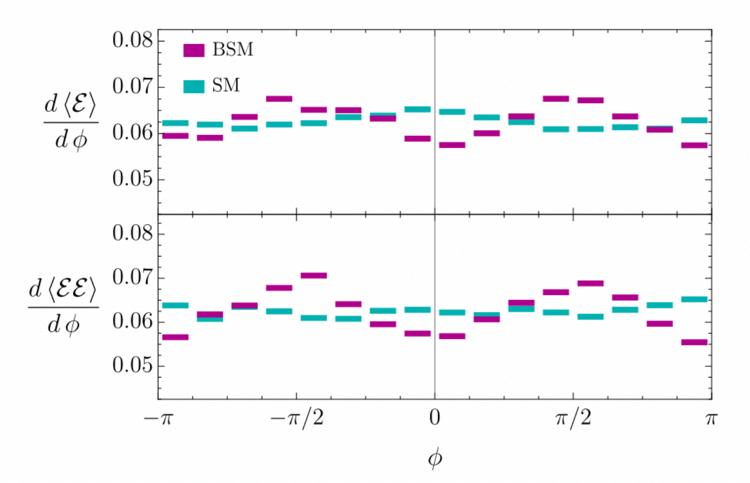

- tree depth D=4 sufficient. Instead of the unbinned case, $N_{bin}$ = 5 already very close to optimum
- significant improvement when including backgrounds and comparing to conventional Run-II strategy

# THE SM-EFT CHALLENGE IS NOT ABOUT ML

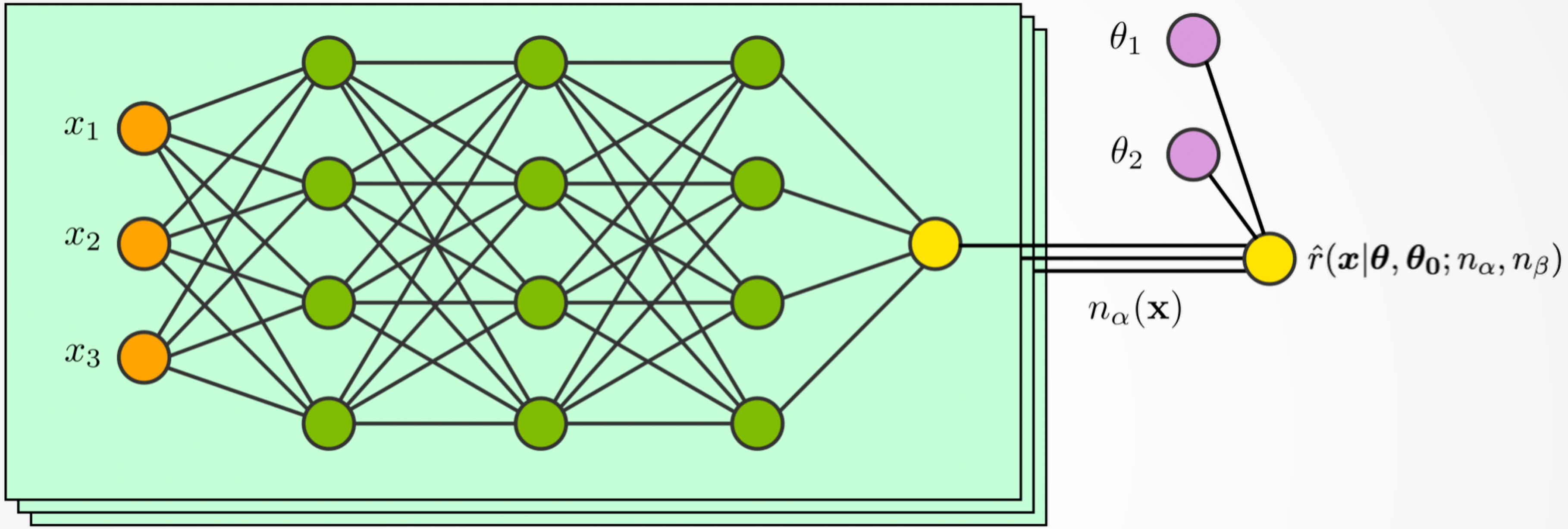
[arxiv:2207.03511]

- Methods to parametrize (close-to) optimal observables for the various final states are established
- The challenge is a 100+ combination (aka global fit) that we can trust
  - across all processes,
  - across all operators,
  - across many years of experimental developments,
  - and while theoretical predictions improve
- The important groundwork is on what we publish, and how complete (uncertainties & correlations), and reproducible it is.
  - ML algorithms help to semi-automatize the analysis design while we can stay receptive to new theory ideas

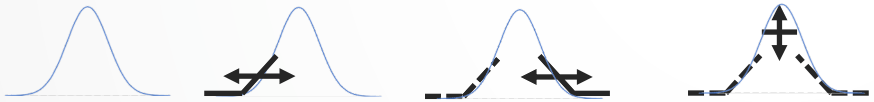
Energy correlators in hadronically decaying 1 TeV W bosons



# NEURAL NETWORKS REGRESS; THE BIT DOESN'T

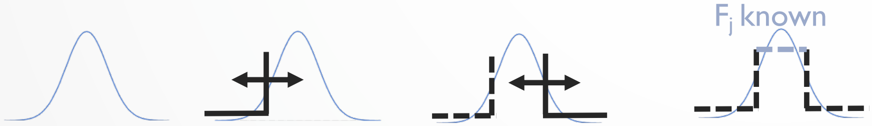



NNs per layer



(do not take too literally)

BIT per depth




(quite litarally what is happening)

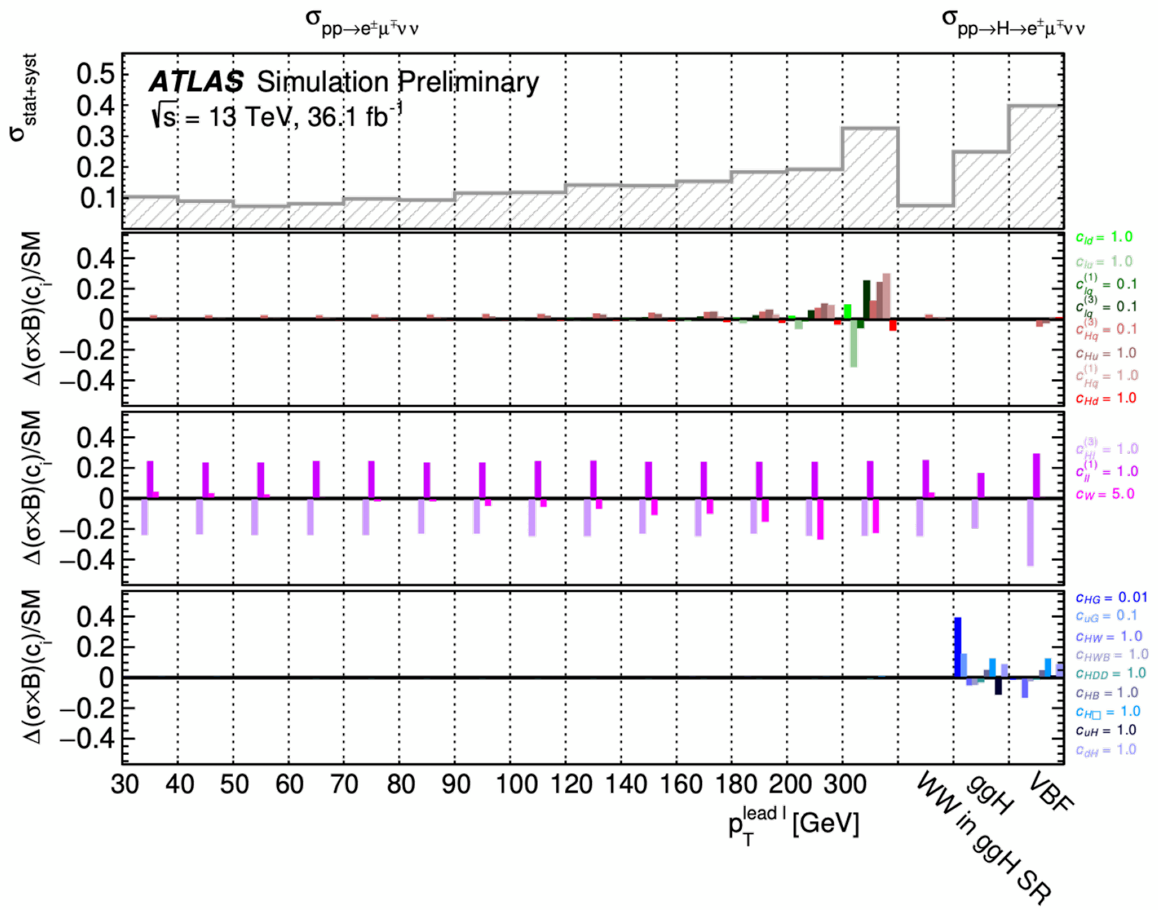
→ fewer DOF need regressing

- Each NN layer maps $L_{n+1} = \sigma(W_{ij}\, L_n + b_i\,)$. These DOF need to select & predict the regressed values.
- In the BIT, we only select. The prediction ($F_j$) is computed from the boxed events. This is possible, because a tree algorithm is (greedely) trained on the *ensemble*. The BITs' DOF are NOT updated event-by-event.

# WW AND H → WW* COMBINATION (LEPTONIC)

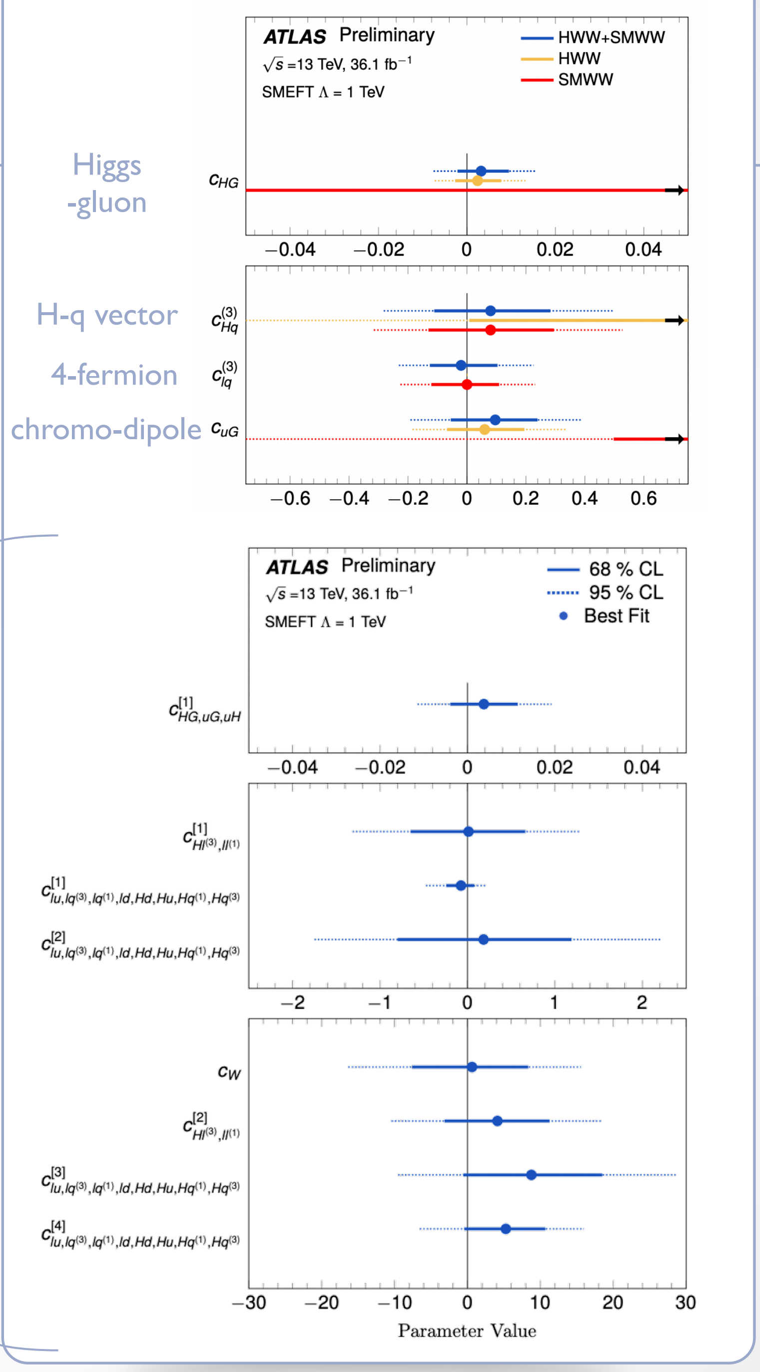
[ATL-PHYS-PUB-2021-010](ATL-PHYS-PUB-2021-010), 36fb⁻¹

- Preformed combined fit of
  1. signal strengths of ggH and VBF in H → WW*
  2. SM WW unfolded differential $p_T$(lead-.l) x-sec
- 20 SM-EFT operators affecting the measurements
- physics-guided eigenbasis probes 8 directions
  - Assume a $U(3)^5$ flavor symmetry
- Stepping stone for more global EFT combinations
- STXS combination: [ATLAS-CONF-2020-053]



# WW AND H → WW* COMBINATION

- Preformed combined fit of
  1. signal strengths of ggH and VBF in H → WW*
  2. SM WW unfolded differential $p_T$(lead-.l) x-sec
- 20 SM-EFT operators affecting the measurements
- physics-guided eigenbasis probes 8 directions
  - Assume a $U(3)^5$ flavor symmetry
- Stepping stone for more global EFT combinations
- STXS combination: [ATLAS-CONF-2020-053]



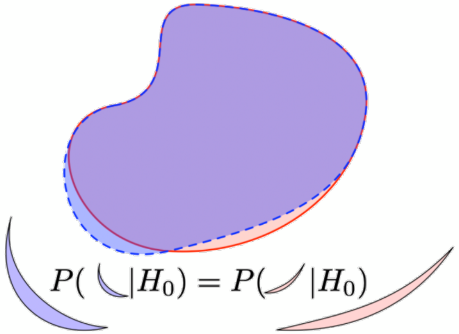
# LIKELIHOOD RATIO TRICK

arxiv:1503.0x7622

Neyman-Pearson Lemma:
The *likelihood* ***ratio***

$$q(\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta}\ )}{L(\mathcal{D}|\boldsymbol{\theta}_0)} \sim \sum_{i=1}^{N} \log \boxed{\frac{p(\boldsymbol{x}_i|\boldsymbol{\theta}\ )}{p(\boldsymbol{x}_i|\boldsymbol{\theta}_0)}}$$

is the optimal test- statstic
in hypothesis tests




Provides the
lowest mis-identification probability
for a given signal efficieny
(No free parameters! ↔ simple hypothesis)

- Train a discriminator to separate signal from background and regress in the truth label

$$L = \underbrace{\int \mathrm{d}\boldsymbol{x} \sum_{z\in\{0,1\}} p(\boldsymbol{x}, z)}_{\text{training sample with mixture of signal (1) and bkg (0)}} \Big(\underset{\uparrow\ \text{truth}}{z} - \underset{\uparrow\ \text{expressive function, e.g, a NN or a decision tree}}{\hat{f}(\boldsymbol{x})}\Big)^2$$

$$= \int \mathrm{d}\boldsymbol{x}\ \left(p(\boldsymbol{x},0)\hat{f}(\boldsymbol{x})^2 + p(\boldsymbol{x},1)(1-\hat{f}(\boldsymbol{x}))^2\right)$$

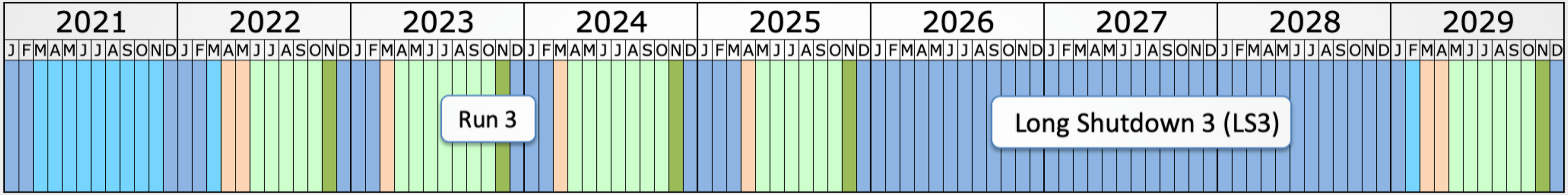
- Training, e.g., with p(z=0) = p(z=1)

$$f^*(\boldsymbol{x}) = \frac{p(\boldsymbol{x},1)}{p(\boldsymbol{x},1)+p(\boldsymbol{x},0)} = \frac{1}{1+\boxed{r(\boldsymbol{x})}}$$

"Likelihood ratio trick" provides a close-to optimal test statistic

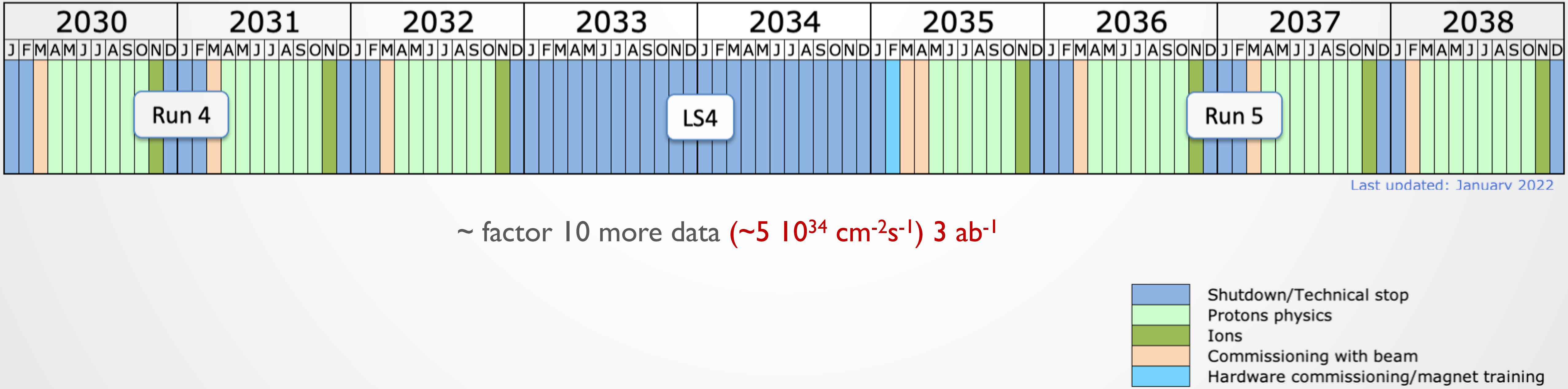
Let's look at SM-EFT applications & issues …

# LHC LONG TERM SCHEDULE




Run 3 dataset will ~ double to 300 $fb^{-1}$ (~$10^{34}$ $cm^{-2}s^{-1}$)

→ HL-LHC




~ factor 10 more data (~5 $10^{34}$ $cm^{-2}s^{-1}$) 3 $ab^{-1}$

# SENDING MIXED SIGNALS TO THE LOSS FUNCTION

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

mixing signals &
case dependent mixes

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

- MSE (& cross-entropy) loss functions average the training data set
  - less-than-ideal for linear effects
- Does not reflect what we know about the $\boldsymbol{\theta}$ dependence
- The real issue is the necessity for a case-dependent training
- The challenge of global SM-EFT searches will require a high degree of automatization
  - Need strategies for learning (approximations) of the log-likelihood suitable for high parameter dimensions

# HOW TO PARAMETRIZE?

- Quantum field theory: Differential cross section predict polynomial SM-EFT dependence:

$$d\sigma(\boldsymbol{\theta}) \propto |\mathcal{M}_{\mathrm{SM}}(\boldsymbol{z}) + \boldsymbol{\theta}_a \mathcal{M}^a_{\mathrm{BSM}}(\boldsymbol{z})|^2 d\boldsymbol{z}$$

probability = wave function, squared

  - additivity of the matrix element → incur a simple (polynomial) dependence in **θ** for fixed configuration z

$$\frac{d\sigma(\boldsymbol{x},\boldsymbol{\theta})}{d\boldsymbol{x}} = \frac{d\sigma_{\mathrm{SM}}(\boldsymbol{x})}{d\boldsymbol{x}} + \sum_a \theta_a \frac{d\sigma^a_{\mathrm{int.}}(\boldsymbol{x})}{d\boldsymbol{x}} + \frac{1}{2}\sum_{a,b}\theta_a\theta_b \frac{d\sigma^{ab}_{\mathrm{BSM}}(\boldsymbol{x})}{d\boldsymbol{x}}$$

- Neyman-Pearson: $q(\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta}\ \ )}{L(\mathcal{D}|\mathrm{SM})}$ where $L(\mathcal{D}|\boldsymbol{\theta}) = \mathrm{P}_{\mathcal{L}\sigma(\boldsymbol{\theta})}(N) \times \prod_{i=1}^{N} p(\boldsymbol{x}_i|\boldsymbol{\theta})$

"normalization" "shape"

$$q_{\boldsymbol{\theta}}(\mathcal{D}) = \underbrace{\mathcal{L}(\sigma_{\boldsymbol{\theta}} - \sigma_{\mathrm{SM}})}_{\mathrm{const.}} - \sum_{\boldsymbol{x}_i \in \mathcal{D}} \log R(\boldsymbol{x}_i|\boldsymbol{\theta}, \mathrm{SM})$$

Optimality can be achieved with cross-section ratio R or its **universal** coefficient functions $R_a$, $R_{ab}$

$$R(\boldsymbol{x}|\boldsymbol{\theta},\mathrm{SM}) = \frac{d\sigma(\boldsymbol{x},\boldsymbol{\theta})/d\boldsymbol{x}}{d\sigma(\boldsymbol{x},\mathrm{SM})/d\boldsymbol{x}} = 1 + \sum_a \theta_a R_a(\boldsymbol{x}) + \frac{1}{2}\sum_{a,b}\theta_a\theta_b R_{ab}(\boldsymbol{x})$$

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

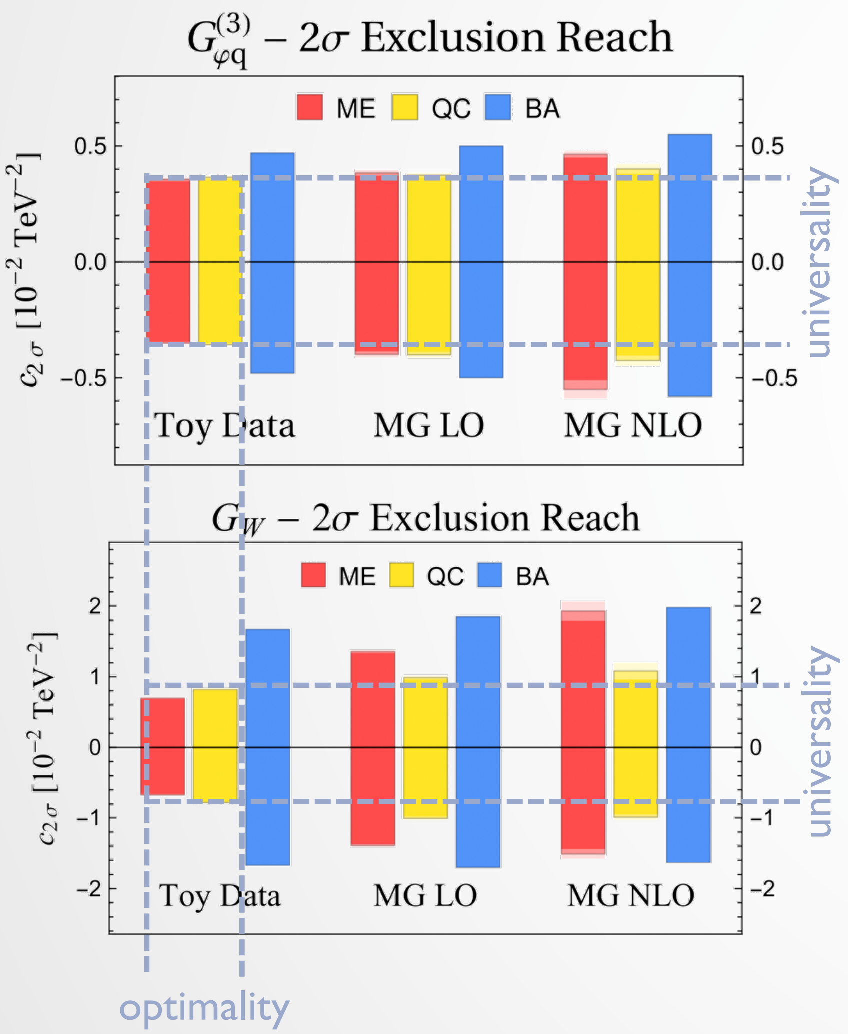
*NB #1 Curse of dimensionality is lifted!!*
*15 operators → 136 coefficients*

*NB #2: R is positive: Fit universal dependence using the most general quadratic polynomial*

# OPTIMAL PARAMETRIZED CLASSIFIERS

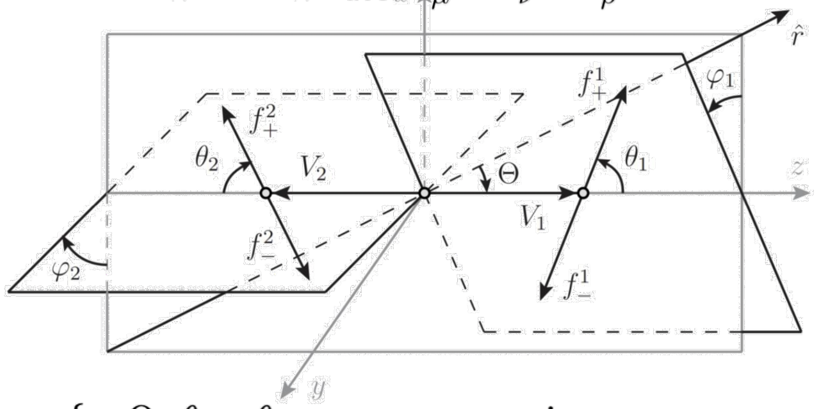
JHEP 05 (2021) 247
ATLAS-CONF-NOTE-2016-043,
JHEP 07 (2022) 032

- studied in the context of $p\,p \to W^{\pm}\,Z \to (l^{\pm}\,\nu)\,(l^{+}\,l^{-})$ for the most important SM-EFT operators

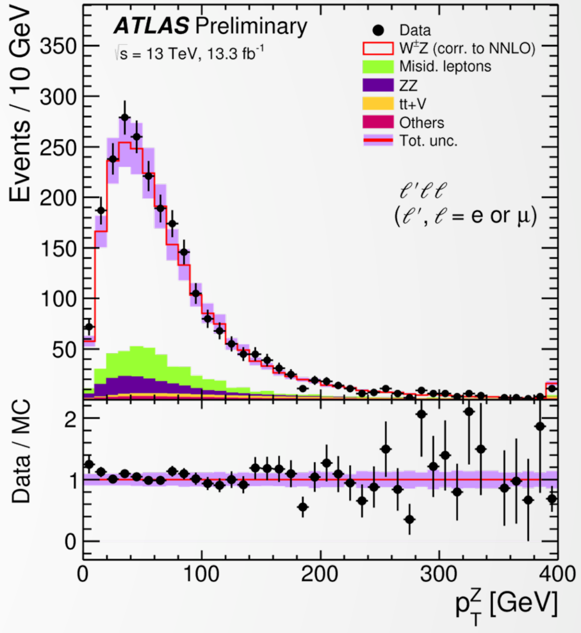



$$\mathcal{O}^{(3)}_{\varphi q} = G^{(3)}_{\varphi q}\,(\overline{Q}_L \sigma^a \gamma^\mu Q_L)\,(iH^\dagger \overleftrightarrow{D}^a_\mu H)$$

$$\mathcal{O}_W = G_W \varepsilon_{abc} W^{a\,\nu}_{\mu} W^{b\,\rho}_{\nu} W^{c\,\mu}_{\rho}$$




$$x = \{s,\, \Theta,\, \theta_W,\, \theta_Z,\, p_{T,\mathrm{ZW}},\, p_{T,\mathrm{Z}},\, \sin\varphi_W,\, \cos\varphi_W,\, \sin\varphi_Z,\, \cos\varphi_Z\}$$

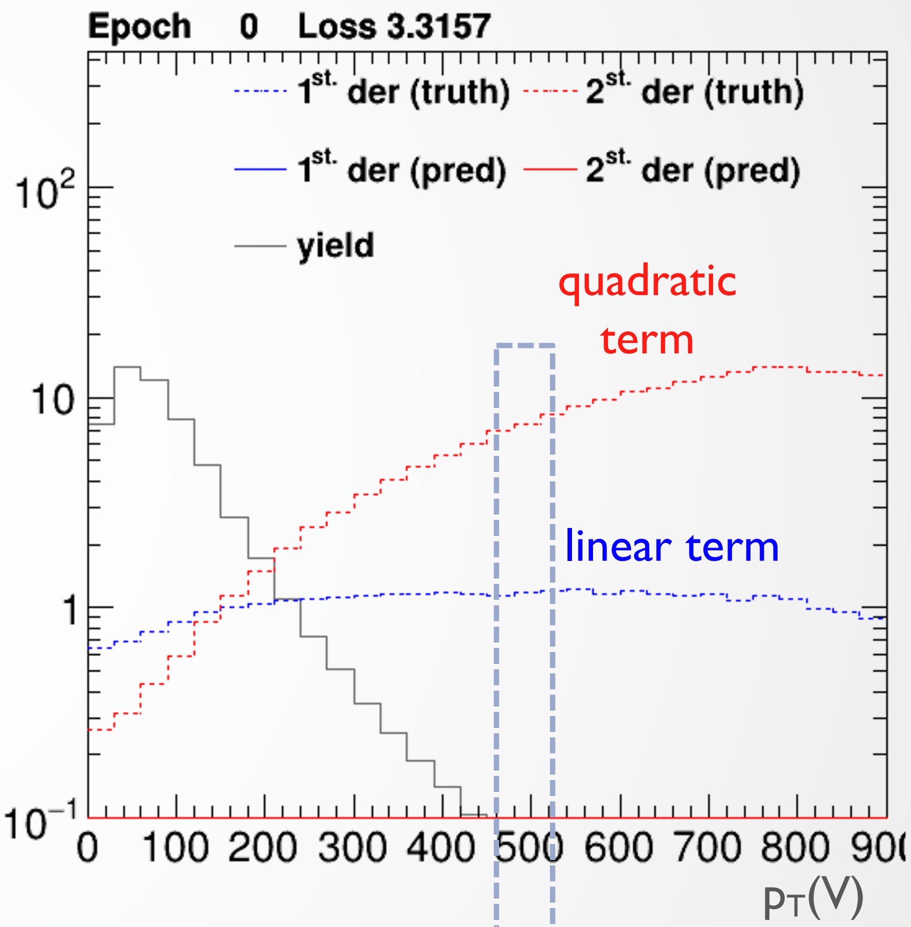



- high purity, ~85%-90% as seen by ATLAS and CMS (with SM-EFT)
- Adam optimizer, pytorch, $10^4$ epochs, learning rate of $10^{-4}$
  - 4 hidden layers á 32 nodes, 2 networks simultaneously trained
  - alternatives configurations studied
- establish optimality with analytic model (Toy), very similar at (N)LO

# PYTORCH IMPLEMENTATION

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

- ZH production, analytic model, 500k events
  - Single coefficient: $c_{HW}$ (tested with up to 3)
  - 4 hidden layers á 32 nodes, 2 networks simultaneously trained
  - $10^4$ epochs, Adam optimizer, LR=$10^{-4}$
- The training is *simultaneous* and it must be!
  - Positivity is a property of the polynomial, not of an individual coefficient.
- several options to emphasise the tails
  - bias loss with function of A(x) or choosing base points
- just a proof of principle implementation




$R(p_T \approx 500) = 1 + \text{blue } \theta + \text{red } \theta^2$

# "PARTICLE PHYSICS STRUCTURE"

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]

- It's somewhat of a miracle that one can regress on the observable-level likelihood ratio

| Observables | Integration over intractable factors | Detector & reconstruction | Parton shower | Parton-level momenta | Theory parameters |
|---|---|---|---|---|---|

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int \mathrm{d}\boldsymbol{z}_\mathrm{d}\, \mathrm{d}\boldsymbol{z}_\mathrm{s}\, \mathrm{d}\boldsymbol{z}_\mathrm{p} \quad p(\boldsymbol{x}|\boldsymbol{z}_\mathrm{d}) \quad p(\boldsymbol{z}_\mathrm{d}|\boldsymbol{z}_\mathrm{s}) \quad p(\boldsymbol{z}_\mathrm{s}|\boldsymbol{z}_\mathrm{p}) \quad p(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta})$$

$$q_{\boldsymbol{\theta}}(\mathcal{D}) = \text{const.} - \sum_{\boldsymbol{x}_i \in \mathcal{D}} \log \frac{\sigma(\boldsymbol{\theta})}{\sigma(\boldsymbol{\theta}_0)} \frac{p(\boldsymbol{x}_i|\boldsymbol{\theta}\ )}{p(\boldsymbol{x}_i|\boldsymbol{\theta}_0)}$$

superpowers

$$\mathrm{d}\sigma(\boldsymbol{\theta}) \propto$$

$$|\mathcal{M}_{\mathrm{SM}}(\boldsymbol{z}) + \boldsymbol{\theta}_a \mathcal{M}^a_{\mathrm{BSM}}(\boldsymbol{z})|^2 \mathrm{d}\boldsymbol{z}$$

calcuable & re-calcuable
(aka tractable) theory prediction

based on this talk: C. Kranmer, J. Brehmer

# "JOINT" DISTRIBUTIONS ARE MUCH SIMPLER

- To understand the power of simulation, look at the simpler "joint" pdf

1. The intractable factors cancel in the joint LR

$$r(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\theta}_0) \equiv \frac{p(\boldsymbol{x}, \boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{s}, \boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta})}{p(\boldsymbol{x}, \boldsymbol{z}_\mathrm{d}, \boldsymbol{z}_\mathrm{s}, \boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta}_0)} = \frac{p(\boldsymbol{x}|\boldsymbol{z}_\mathrm{d})}{p(\boldsymbol{x}|\boldsymbol{z}_\mathrm{d})} \frac{p(\boldsymbol{z}_\mathrm{d}|\boldsymbol{z}_\mathrm{s})}{p(\boldsymbol{z}_\mathrm{d}|\boldsymbol{z}_\mathrm{s})} \frac{p(\boldsymbol{z}_\mathrm{s}|\boldsymbol{z}_\mathrm{p})}{p(\boldsymbol{z}_\mathrm{s}|\boldsymbol{z}_\mathrm{p})} \frac{p(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta})}{p(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta}_0)} \propto \frac{|\mathcal{M}(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta})|^2}{|\mathcal{M}(\boldsymbol{z}_\mathrm{p}|\boldsymbol{\theta}_0)|^2}$$

Change in likelihood of observation x (with history z) going from $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}$

staged simulation: Intractable factors cancel

re-calcuable theory prediction

2. Now fit a general function on the join space with a regressor depending only on the observables:

$$L = \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) \left( f(\boldsymbol{x}, \boldsymbol{z}) - \hat{f}(\boldsymbol{x}) \right)^2 \longrightarrow \min \qquad f^*(\boldsymbol{x}) = \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) f(\boldsymbol{x}, \boldsymbol{z})}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)}$$

Latent space is integrated

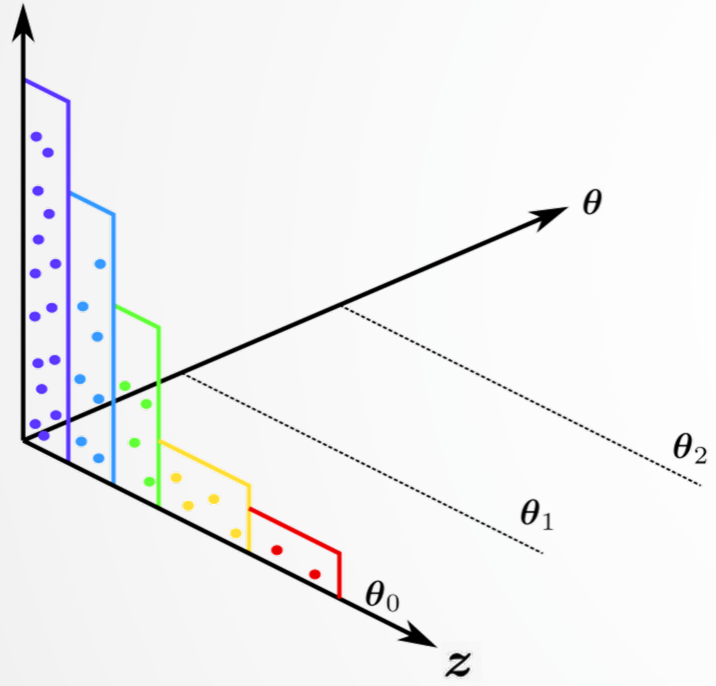
3. Now chose f(x,z) = r(x,z | $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$) which is available in simulation & fit with expressive function:

$$f^*(\boldsymbol{x}) = \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) f(\boldsymbol{x}, \boldsymbol{z})}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) r(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\theta}_0)}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) \frac{p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})}{p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)}}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})}{\int \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta}_0)}$$

Available from simulation $\Longrightarrow$ what we actually want: change in likelihood of a specific observation

*... statistical framework of all the parametrized classifiers*

# EXPLOITING PARAMETRIZED *SIMULATION* WITH TREES




- Quantum field theory: Differential cross section have structure

$$d\sigma(\boldsymbol{\theta}) \propto |\mathcal{M}_{\mathrm{SM}}(\boldsymbol{z}) + \boldsymbol{\theta}_a \mathcal{M}^a_{\mathrm{BSM}}(\boldsymbol{z})|^2 d\boldsymbol{z}$$

probability = wave function, squared

  - sampling **z** at a fixed $\boldsymbol{\theta}_0$
  - evaluate d$\sigma$($\boldsymbol{\theta}$) for sufficient number of base-points $\boldsymbol{\theta}$
  - fix polynomial coefficients of event weights $w_i(\boldsymbol{\theta})$

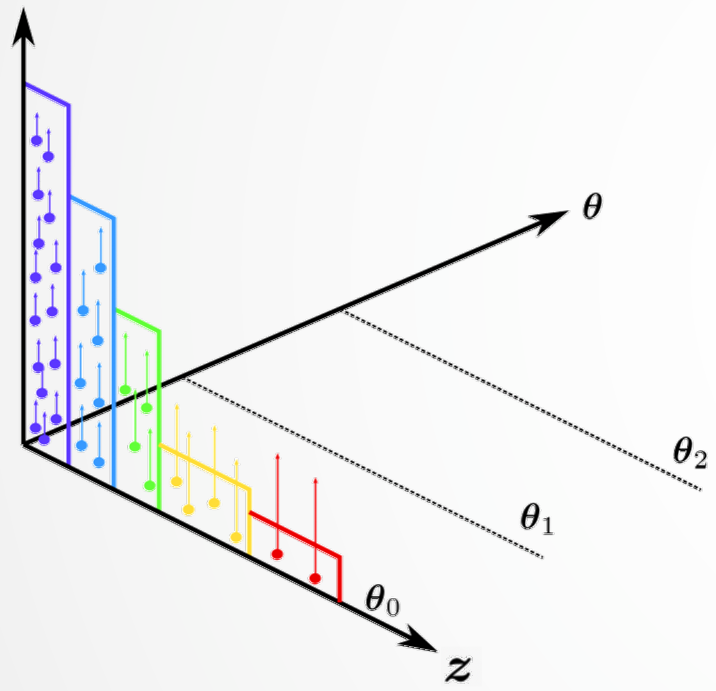
$$w_i(\boldsymbol{\theta}) = w_{i,0} + \sum_a w_{i,a}\,\theta_a + \frac{1}{2}\sum_{a,b} w_{ab}\,\theta_a\theta_b = \frac{\sigma(\boldsymbol{\theta}\ )}{\sigma(\boldsymbol{\theta}_0)} \cdot r(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

SM interference pure SM-EFT

*interpretation valid at LO*

  - obtain predictions for, e.g., yields for all x,z and $\boldsymbol{\theta}$

# EXPLOITING PARAMETRIZED *SIMULATION* WITH TREES




- Quantum field theory: Differential cross section have structure

$$\mathrm{d}\sigma(\boldsymbol{\theta}) \propto |\mathcal{M}_{\mathrm{SM}}(\boldsymbol{z}) + \boldsymbol{\theta}_a \mathcal{M}^a_{\mathrm{BSM}}(\boldsymbol{z})|^2 \mathrm{d}\boldsymbol{z}$$

probability = wave function, squared

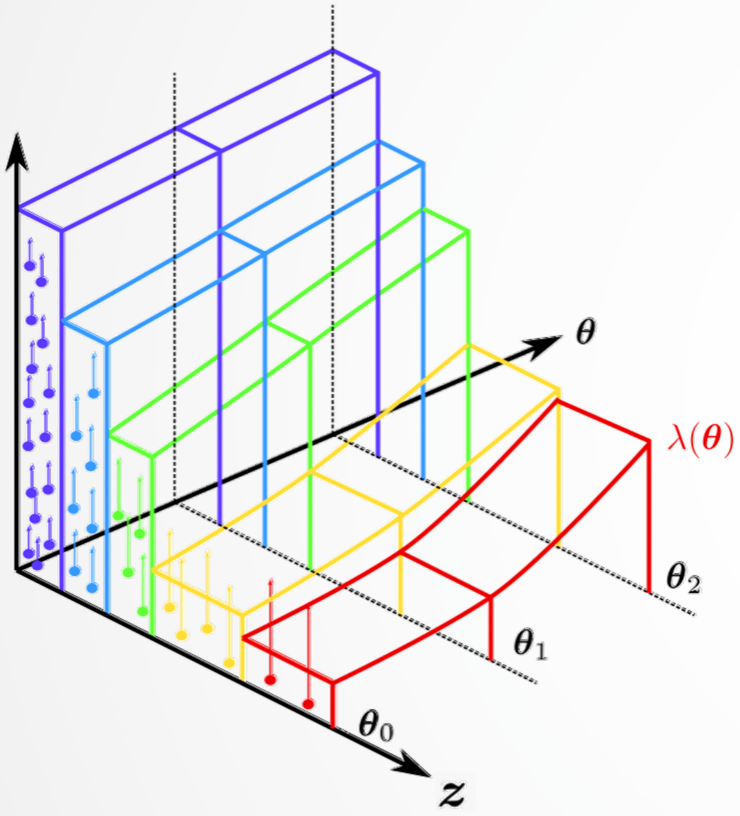
- sampling **z** at a fixed $\boldsymbol{\theta}_o$
- evaluate $\mathrm{d}\sigma(\boldsymbol{\theta})$ for sufficient number of base-points $\boldsymbol{\theta}$
- fix polynomial coefficients of event weights $w_i(\boldsymbol{\theta})$

$$w_i(\boldsymbol{\theta}) = w_{i,0} + \sum_a w_{i,a}\,\theta_a + \frac{1}{2}\sum_{a,b} w_{ab}\,\theta_a\theta_b = \frac{\sigma(\boldsymbol{\theta})}{\sigma(\boldsymbol{\theta}_0)} \cdot r(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

SM | interference | pure SM-EFT | *interpretation valid at LO*

- obtain predictions for, e.g., yields for all x,z and $\boldsymbol{\theta}$

# EXPLOITING PARAMETRIZED *SIMULATION* WITH TREES




- Quantum field theory: Differential cross section have structure

$$d\sigma(\boldsymbol{\theta}) \propto |\mathcal{M}_{\mathrm{SM}}(\boldsymbol{z}) + \boldsymbol{\theta}_a \mathcal{M}^a_{\mathrm{BSM}}(\boldsymbol{z})|^2 d\boldsymbol{z}$$

probability = wave function, squared

  - sampling **z** at a fixed $\boldsymbol{\theta}_0$
  - evaluate d$\sigma$($\boldsymbol{\theta}$) for sufficient number of base-points $\boldsymbol{\theta}$
  - fix polynomial coefficients of event weights $w_i(\boldsymbol{\theta})$

$$w_i(\boldsymbol{\theta}) = w_{i,0} + \sum_a w_{i,a}\,\theta_a + \frac{1}{2}\sum_{a,b} w_{ab}\,\theta_a\theta_b = \frac{\sigma(\boldsymbol{\theta}\ )}{\sigma(\boldsymbol{\theta}_0)} \cdot r(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

SM interference pure SM-EFT

*interpretation valid at LO*

  - obtain predictions for, e.g., yields for all x,z and $\boldsymbol{\theta}$

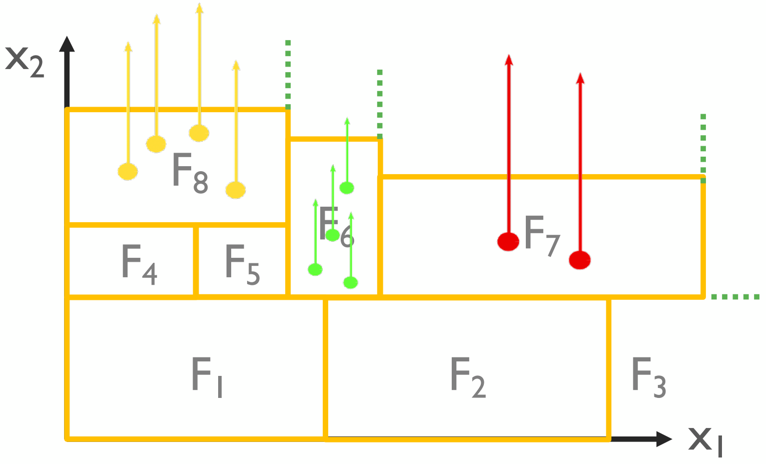
# TREES & BOOSTING

[arXiv:2107.10859, arXiv:2205:12976]

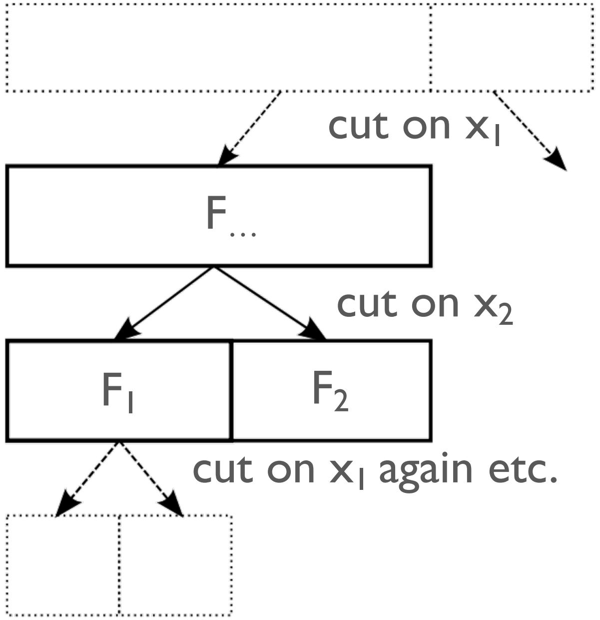
**weak learner**

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

**phase-space partitioning**




**training phase:**
e.g. “CART” algo




- Let us make a tree-based prediction for R or its coefficient function
- Weak learner: Tree ↔ Associates a predictive function $F_j$ (flexible!) with a sub-region j of a partitioning
- Fitting tree: Optimize ”node split positions” on some loss. Trained (e.g. greedily) on the *ensemble*.
  - Rectangular cuts are very limiting. Remove the limitation with “boosting”.

# LEARNING THE SCORE FUNCTION

[arXiv:2107.10859, arXiv:2205:12976]

- Example: Learn a local version of the model, described by the score function (local LLR)

$$\boldsymbol{t}(\boldsymbol{x}|\boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{x}|\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0} = \frac{\nabla_{\boldsymbol{\theta}} p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x}|\boldsymbol{\theta})}\Big|_{\boldsymbol{\theta}_0}$$

- Only the joint score $\boldsymbol{t}(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)$ is available in training. This is enough, though.

$$L = \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) \left( t_a(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0) - \hat{F}_a(\boldsymbol{x}) \right)^2 \longrightarrow \min$$

formal solution: $F_a^*(\boldsymbol{x}) = t_a(\boldsymbol{x}|\boldsymbol{\theta}_0)$

tree ansatz

$$= \sum_{(\boldsymbol{x},\boldsymbol{z},w)_i \in \mathcal{D}} w_{i,0} \left| \frac{w_{i,a}}{w_{i,0}} - \hat{F}_a(\boldsymbol{x}_i) \right|^2 = \sum_{j\in\mathcal{J}} \sum_{i\in j} w_i \left| \frac{w_{i,a}}{w_i} - F_j \right|^2 = \sum_{i=1}^{N_{\mathrm{sim}}} \frac{w_{i,a}^2}{w_i} - 2 \sum_{j\in\mathcal{J}} F_j \sum_{i\in j} w_{i,a} + \sum_{j\in\mathcal{J}} F_j^2 \sum_{i\in j} w_i$$

$$\frac{\partial L}{\partial F_j} = 0 \longrightarrow F_j = \frac{\sum_{i\in j} w_{i,a}}{\sum_{i\in j} w_i}$$

…re-insert $F_j$ into L …

the predictor does NOT have trainable parameters!

$$L = -\sum_{j\in\mathcal{J}} \frac{\left(\sum_{i\in j} w_{i,a}\right)^2}{\sum_{i\in j} w_i} = -\sum_{j\in\mathcal{J}} I^{(\lambda_j)}$$

maximise Fisher information of a Poisson $\theta_a$ measurement

# CONCRETE SOLUTION: TREE BOOSTING

[arXiv:2107.10859, arXiv:2205:12976]

- Boosting: Fit linear model iteratively to pseudo-residuals of the preceding iteration
- Ansatz :

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

- Insert into the loss function:

$$\underbrace{\mathrm{MSE}[\hat{f}_a^{(b)}]}_{\text{current iteration}} = \sum_{(\boldsymbol{x},\boldsymbol{z},w)_i\in\mathcal{D}} w_{i,0}\left|\underbrace{\frac{w_{i,a}}{w_{i,0}} - \eta \underbrace{\hat{F}_a^{(b-1)}(\boldsymbol{x}_i)}_{\text{previous iteration}}}_{\text{pseudo-residual}} - \underbrace{\hat{f}_a^{(b)}(\boldsymbol{x}_i)}_{\text{current iteration}}\right|^2 = \underbrace{\sum_{(\boldsymbol{x},\boldsymbol{z},w)_i\in\mathcal{D}} w_{i,0}\left|\underbrace{\frac{w_{i,a} - \eta w_{i,0}\hat{F}_a^{(b-1)}(\boldsymbol{x}_i)}{w_{i,0}}}_{\text{reweighting}} - \hat{f}_a^{(b)}(\boldsymbol{x}_i)\right|^2}_{\text{MSE structure at iteration b}}$$

.... perform this iteratively

# LEARNING MORE WITH TREES

[arXiv:2107.10859, arXiv:2205:12976]

Regress in one of the coefficient functions of R

$$R(\boldsymbol{x}|\boldsymbol{\theta},\mathrm{SM}) = \frac{\mathrm{d}\sigma(\boldsymbol{x},\boldsymbol{\theta})/\mathrm{d}\boldsymbol{x}}{\mathrm{d}\sigma(\boldsymbol{x},\mathrm{SM})/\mathrm{d}\boldsymbol{x}} = 1 + \sum_a \theta_a R_a(\boldsymbol{x}) + \frac{1}{2}\sum_{a,b}\theta_a\theta_b R_{ab}(\boldsymbol{x})$$

$$L = \int \mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{z}\, p(\boldsymbol{x},\boldsymbol{z}|\mathrm{SM}) \left(R_{a(,b)}(\boldsymbol{x},\boldsymbol{z}) - \hat{F}_{a(,b)}(\boldsymbol{x})\right)^2$$

Tree ansatz for each a, ab: $\hat{F}(\boldsymbol{x}) = \sum_{j\in\mathcal{J}} \mathbb{1}_j(\boldsymbol{x}) F_j$

$F_j$ is a constant (per node)

… Solve for $F_j$ & reinsert … $F_j = \frac{\sum_{i\in j} w_{i,ab}}{\sum_{i\in j} w_i}$

Solve for optimal partitioning with CART algorithm

$$L = -\sum_{j\in\mathcal{J}} \frac{\left(\sum_{i\in j} w_{i,ab}\right)^2}{\sum_{i\in j} w_i}$$

$\rightarrow$ boost

Regress in R, including its the polynomial $\boldsymbol{\theta}$ dependence

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

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

Tree ansatz for each a, ab: $\hat{F}(\boldsymbol{x},\boldsymbol{\theta}) = \sum_{j\in\mathcal{J}} \mathbb{1}_j(\boldsymbol{x}) F_j(\boldsymbol{\theta})$

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

… Solve for $F_j$ & reinsert … $F_j(\boldsymbol{\theta}) = \frac{\sum_{i\in j} w_i(\boldsymbol{\theta})}{\sum_{i\in j} w_i(\boldsymbol{\theta}_0)} \equiv \frac{w_j(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$

Solve for optimal partitioning with CART algorithm

$$L = -\sum_{\boldsymbol{\theta}\in\mathcal{B}}\sum_{j\in\mathcal{J}} \frac{w_j^2(\boldsymbol{\theta})}{w_j(\boldsymbol{\theta}_0)}$$

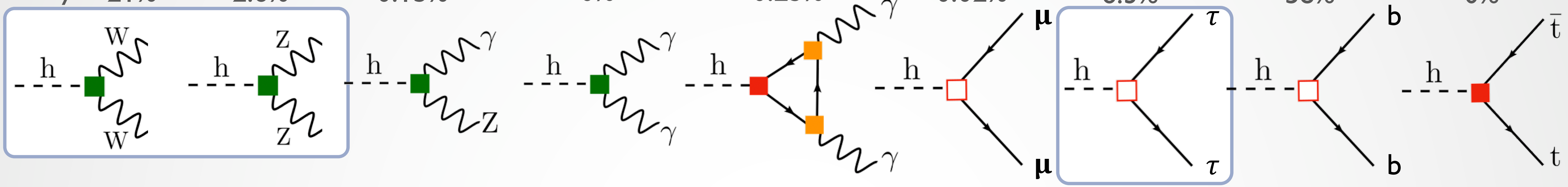
split only if $w_j(\boldsymbol{\theta})$ is *positive* $\forall\boldsymbol{\theta}$

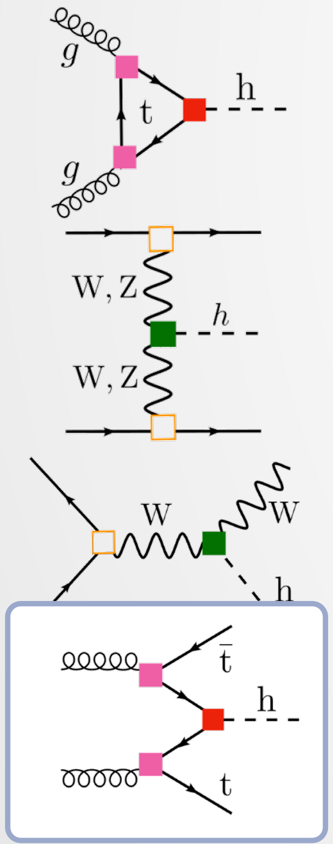
$\rightarrow$ boost

# TTH IN THE MULTILEPTON CHANNEL

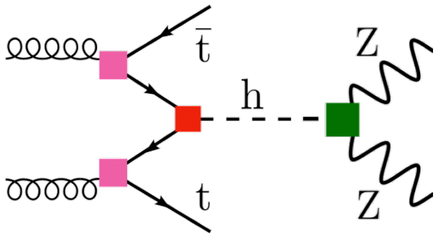
[JHEP (submitted)]

decay: 21% 2.6% 0.15% 0% 0.23% 0.02% 6.3% 58% 0%



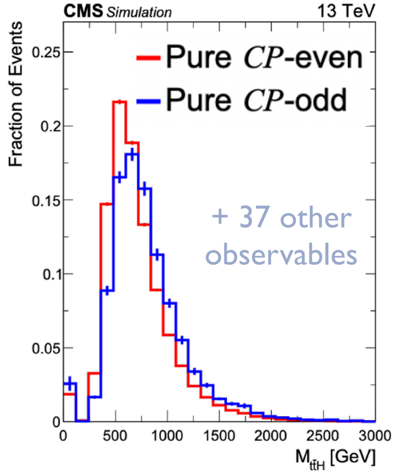




ttH multilepton

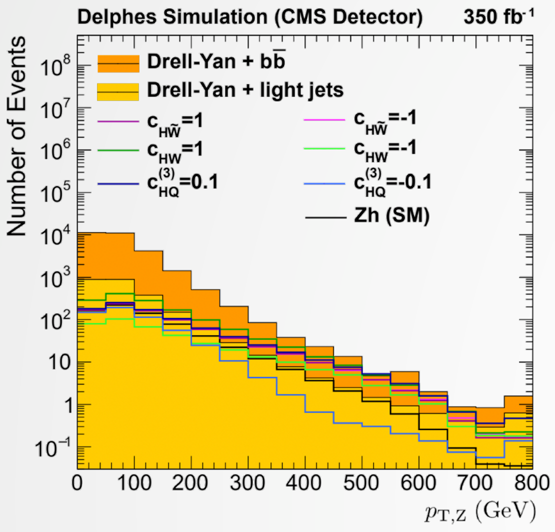


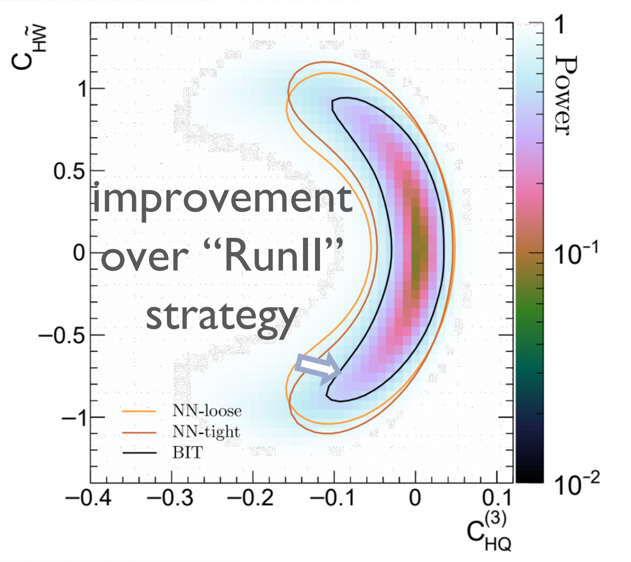

- example #2: t(t)H multilepton in 2ℓSS+0$\tau$, 2ℓSS+1$\tau$, 3ℓ final states
- deep convolutional network [DeepTau] for $\tau$ reconstruction
  - uses tracking, calorimetry, muons via particle-flow collections
- 3 DNNs for signal/background multi-classification

- targets t-t-H Yukawa coupling (■) in $\kappa$-framework
  - in SM-EFT: "CP" structure (complex phase) of $HH^{\dagger}\bar{q}_p u_r \tilde{H}$
  - CP violating effects in couplings to bosons (■) supressed by $\Lambda^4$
- use ML for separating CP-even vs. odd: gradient-BDT XGBoost
  - 38 input features (kinematic properties)

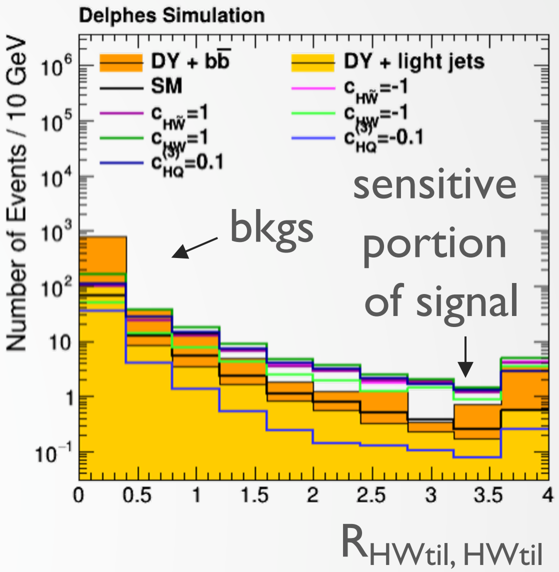


# INCLUDING BACKGROUNDS IN THE TRAINING [arXiv:2107.10859, arXiv:2205:12976]




| Observable |
|---|
| $H_T$ |
| $N_{\mathrm{jet}}$ |
| $p_T(j_1)$, $p_T(j_2)$, $p_T(j_3)$ |
| $\lvert\eta(j_1)\rvert$, $\lvert\eta(j_2)\rvert$, $\lvert\eta(j_3)\rvert$ |
| $p_T(\mathrm{h})$, $\lvert\eta(\mathrm{h})\rvert$ |
| $p_T(\mathrm{Z})$, $\lvert\eta(\mathrm{Z})\rvert$ |
| $\Theta$, $\hat{\theta}$, $\hat{\phi}$ |
| $f_{LL}, \ldots, \tilde{f}_{TT'}$ |
| $p_T(\ell_2)/p_T(\ell_1)$ |
| $\Delta\phi(\ell_1, \ell_2)$, $\lvert\Delta\eta(\ell_1, \ell_2)\rvert$ |
| $\Delta\phi(\text{b-jet}_1, \text{b-jet}_2)$, $\lvert\Delta\eta(\text{b-jet}_1, \text{b-jet}_2)\rvert$ |
| $m(\text{b-jet}_1, \text{b-jet}_2)$ |
| $p_T(\text{b-jet}_2)/p_T(\text{b-jet}_1)$ |
| $\Delta R(\mathrm{Z}, \mathrm{h})$, $\lvert\Delta\eta(\mathrm{Z}, \mathrm{h})\rvert$, $m(\mathrm{Z}, \mathrm{h})$ |
| $\Delta R(\text{non b-jet}, \mathrm{Z})$, $\Delta R(\text{non b-jet}, \mathrm{h})$ |
| Thrust |

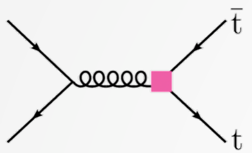



- Include most important background processes
  - Simulate signal using MG5 & SMEFTsim + Delphes
  - learn $R(x|\boldsymbol{\theta})$ for ZH + Drell-Yan (correctly weighted)
- Train on observables that capture EFT dependence and also discriminate between backgrounds
- Compare with the CMS "Run II" strategy: NN to separate background, then fit $p_T(Z)$
  - substantial improvement

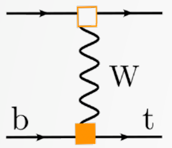
# TOP QUARKS + X IN SM-EFT

JHEP 12 (2021) 083

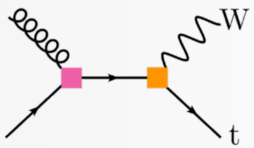
tt
811 pb



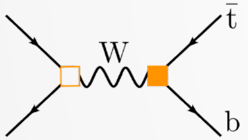

t (t-channel)
217 pb



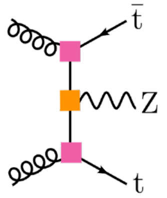

tW
72 pb



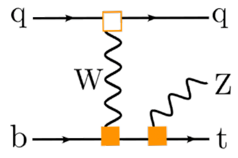

t (s-channel)
10 pb




ttZ
1 pb



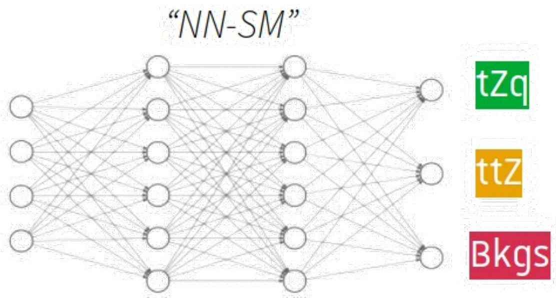

tZq
0.088 pb

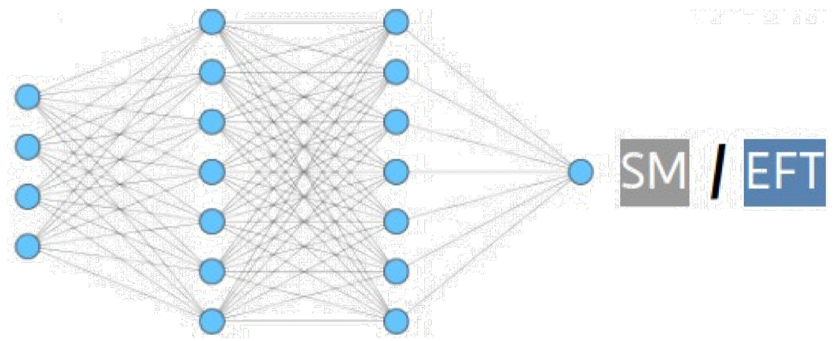



- example #3: top quark – Z boson coupling
- exploit kinematics in ttZ/tZq/tWZ final states
  - low-background final states; bkg for tt+Higgs
  - 5 SM-EFT operators
- Extensive use of MVAs
  - Multiclassifier to discriminate between several SM processes
    - using 33 (mostly kinematic) event properties
  - 8 neural network *binary classifiers to BSM events*

| | |
|---|---|
| Weak top dipole interactions | $\mathcal{O}_{tZ}$, $\mathcal{O}_{tW}$ |
| LH vector couplings | $\mathcal{O}^{3}_{\varphi Q}$, $\mathcal{O}^{-}_{\varphi Q}$ |
| RH vector couplings | $\mathcal{O}_{\varphi t}$ |

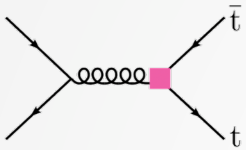
2 or 3 hidden layers
50-100 neurons
ReLU activation,
sigmoid output
LR 0.001 (decaying)
Adam optimizer



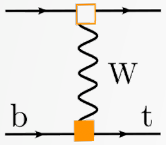



# TOP AND DIBOSON SECTORS

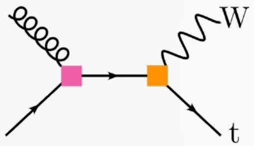
**tt**
811 pb



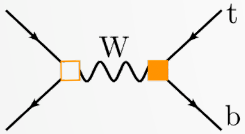

**t (t-channel)**
217 pb



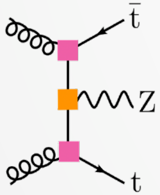

**tW**
72 pb



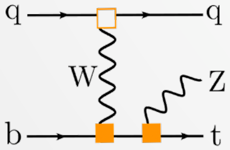

**t (s-channel)**
10 pb



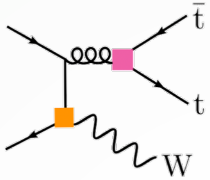

**ttZ**
1 pb



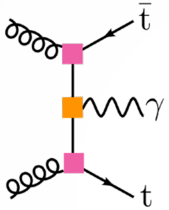

**tZq**
0.088 pb



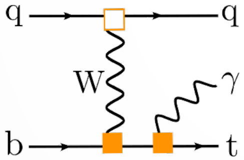

**ttW**
0.55 pb



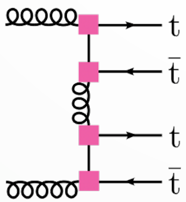

**ttγ**
0.80 pb



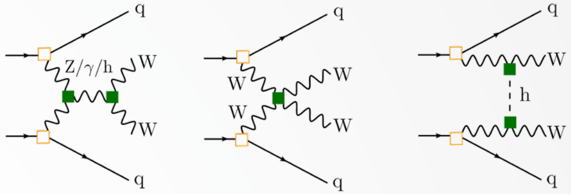

**tγq**
0.4 pb



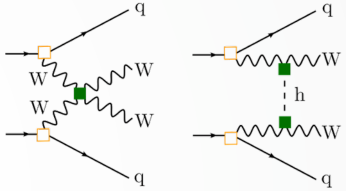

**tttt**
0.009 pb



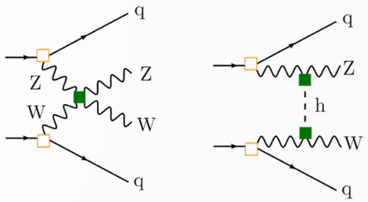

**W±W∓**
EWK 10 fb



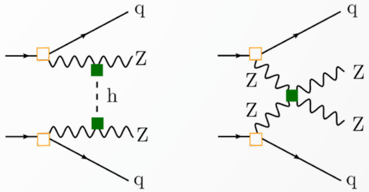

**W±W±**
EWK 4 fb



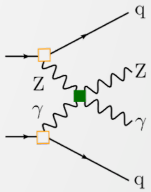
**WZ**
EWK 2.15 fb



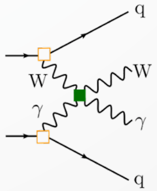
**ZZ**
EWK 0.82 fb



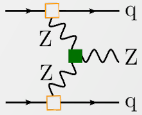
**Zγ**
EWK 7.8 fb



**Wγ**
EWK 20.4 fb



**Z**
EWK 37.4 fb



# DEEP TAU (DECAY MODES)

charge-conjugate decays.

| Decay mode | Resonance | $\mathcal{B}$ (%) | |
|---|---|---|---|
| Leptonic decays | | 35.2 | |
| $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ | | | 17.8 |
| $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ | | | 17.4 |
| Hadronic decays | | 64.8 | |
| $\tau^- \to h^- \nu_\tau$ | | | 11.5 |
| $\tau^- \to h^- \pi^0 \nu_\tau$ | $\rho(770)$ | | 25.9 |
| $\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$ | $a_1(1260)$ | | 9.5 |
| $\tau^- \to h^- h^+ h^- \nu_\tau$ | $a_1(1260)$ | | 9.8 |
| $\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$ | | | 4.8 |
| Other | | | 3.3 |