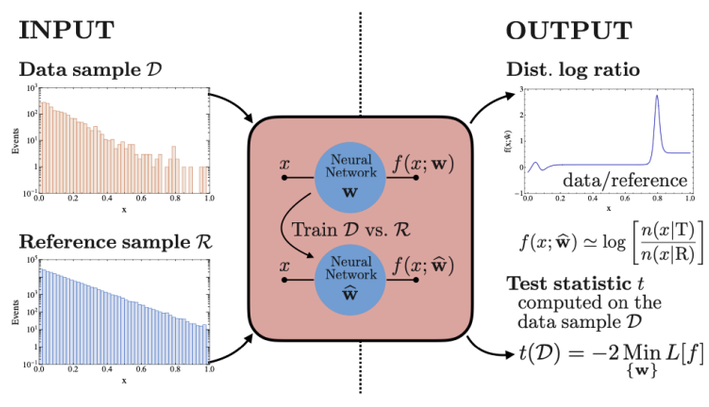
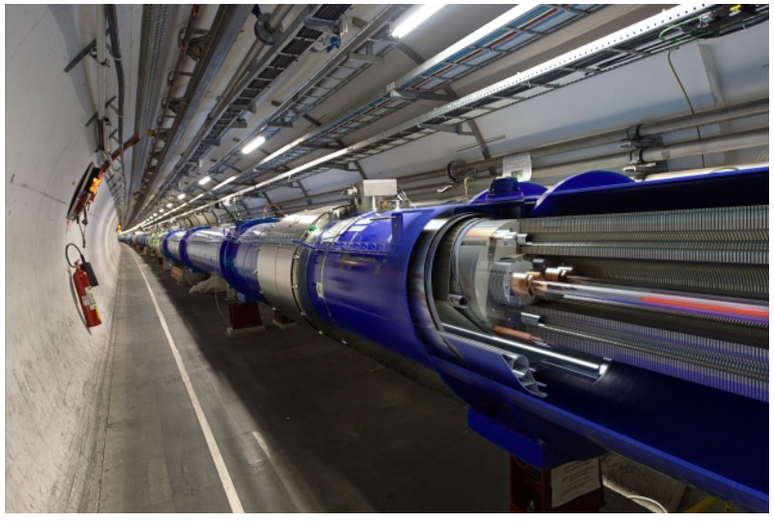
Parametrised Classifiers for Optimal EFT Sensitivity



1806.02350



Siyu Chen, 13 Sep 2022

Based on *JHEP* 05 (2021) 247 or on arxiv: 2007.10356 (S.C., Glioti, Panico, Wulzer)

Motivation

- BSM physics search is **indirect** and requires high sensitivity in statistical hypothesis testing

- Optimal sensitivity can be obtained if the **likelihood** is known in a closed form

  …but we don't have access to that.

- Approximation methods:

  - matrix element method

    Kondo, K. (1988) JPS, 57(12), 4126-4140
    Artoisenet & al. 1007.3300
    Fiedler & al. 1003.1316
    Martini & al. 1506.08789 & 1712.04527

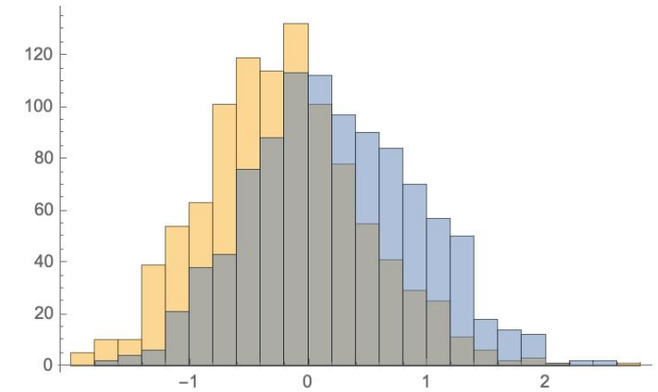
  - binned analysis

  - multivariate binned analysis?

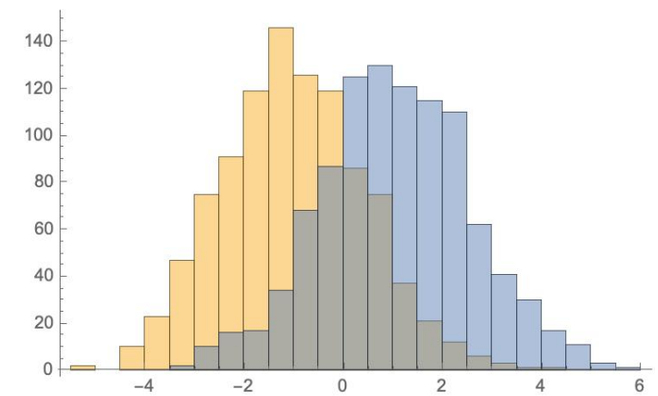
    Cranmer, Pavez & Louppe. 1506.02169
    Baldi et al. 1601.07913
    Brehmer et al. 1805.00013
    Brehmer et al. 1805.12244
    Stoye et al. 1808.00973
    Brehmer et al. 1907.10621

- Selecting one discriminating kinematic variable to do the binned analysis is a waste! We should be able to do better.

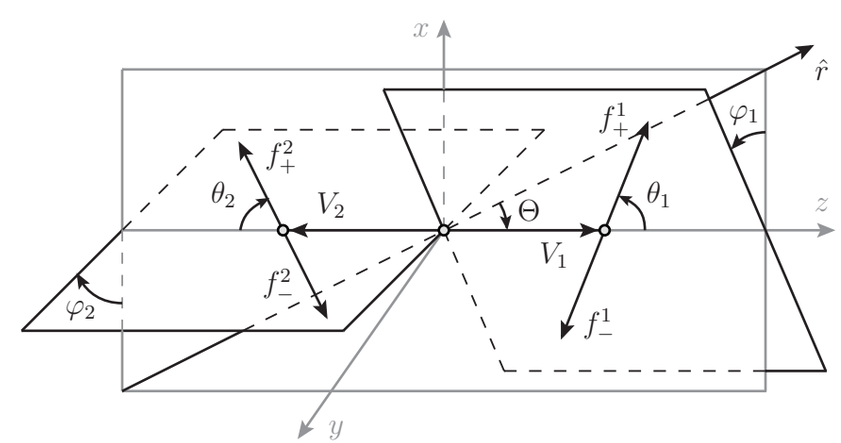
  Franceschini, Panico, Pomarol, Riva & Wulzer, 1712.01310
  Panico, Riva, Wulzer, 1708.07823



Test statistics distribution -binned analysis, 1 variable



Test statistics distribution -binned analysis, 2 variables



diboson production and decay: kinematic variables

Methodology: approximating the likelihood

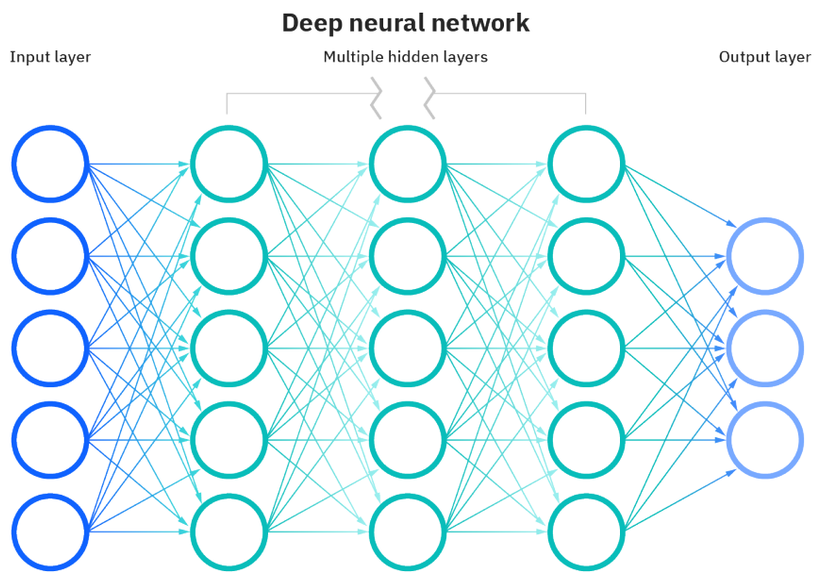
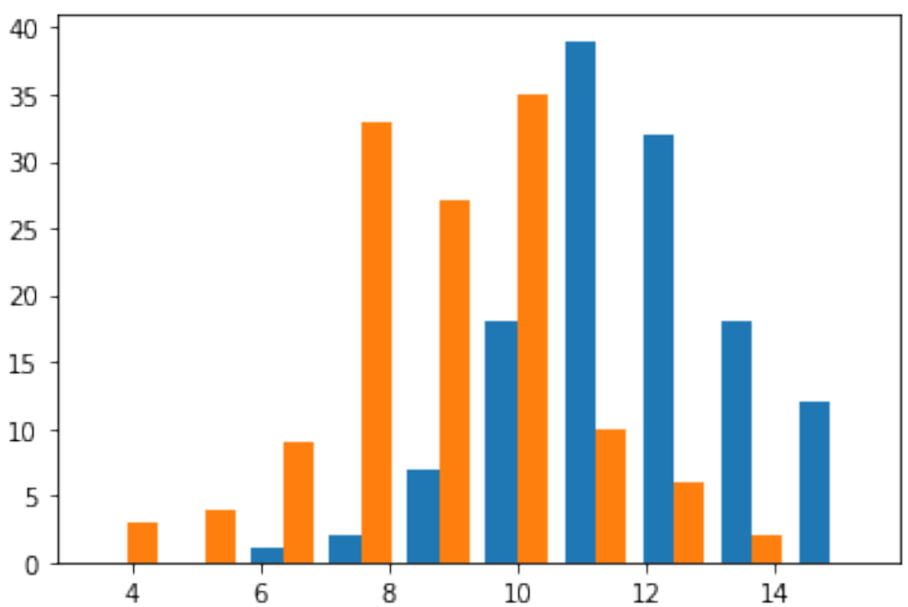


Illustration of a fully connected neural network



test statistics distribution using $f(x)$

- Simulation events involve an intractable process

$d\sigma(z\,|\,H)$ **parton level variables z**

↓ random variables

$d\sigma(x\,|\,H)$? **measurable x**

- We use neural networks to approximate $d\sigma(x\,|\,H)$, or better, $d\sigma(x\,|\,H_1)/d\sigma(x\,|\,H_0)$

- The learning problem

$$\text{Loss} = \int_{e\in H_0} d\sigma_0[f(x)]^2 + \int_{e\in H_1} d\sigma_1[1-f(x)]^2$$

… in the large sample limit, …

$$r(x) \to d\sigma(x\,|\,H_1)/d\sigma(x\,|\,H_0)$$

and we can approximate the likelihood.

Methodology: Parametrisation

- The form of $f(x)$ is too general: in fact, we know that BSM physics plays a part in the form

$$d\sigma(x\,|\,H_1)/d\sigma(x\,|\,H_0) = 1 + c\#_1 + c^2\#_2$$

- We demand this specific form from networks

$$r(x) = [1 + \alpha(x)c]^2 + \beta(x)^2 c^2$$

- restructure of neural network architecture

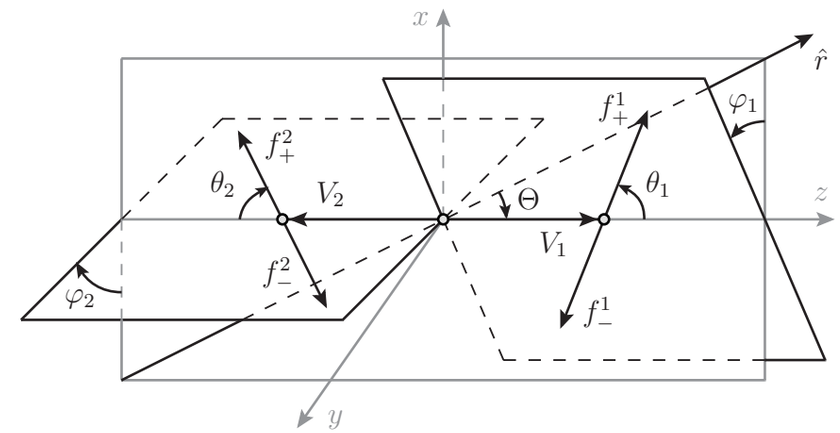
- regularisation term

Tod, G., Both, G. J., & Kusters, R. (2021) a*rXiv preprint arXiv:2109.11939*.

"physics-informed neural networks"

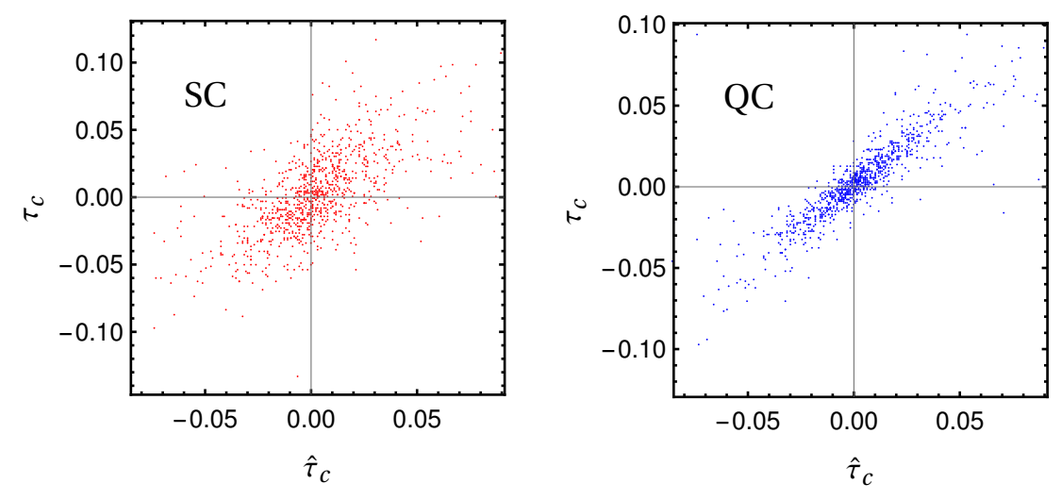
- doubly parametrised likelihood ratio estimator $p(x\,|\,\theta_1)/p(x\,|\,\theta_0)$

Brehmer, J., Kling, F., Espejo, I., & Cranmer, K. (2020). *Computing and Software for Big Science*, *4*(1), 1-25.



$$\mathcal{O}_{\phi q}^{(3)} = G_{\phi q}^{(3)} \left(\bar{Q}_L \sigma^a \gamma^\mu Q_L\right) \left(iH^\dagger \overleftrightarrow{D}_\mu^a H\right)$$

$$\mathcal{O}_W = G_W \epsilon_{abc} W_\mu^{a,\nu} W_\nu^{b,\rho} W_\rho^{c,\mu}$$



Scattering of truth value $\tau(x) = \log[d\sigma(x\,|\,H_1)/d\sigma(x\,|\,H_0)]$ against the reconstructed $\hat{\tau}(x)$

Based on *JHEP* 05 (2021) 247 or on arxiv: 2007.10356 (S.C., Glioti, Panico, Wulzer)

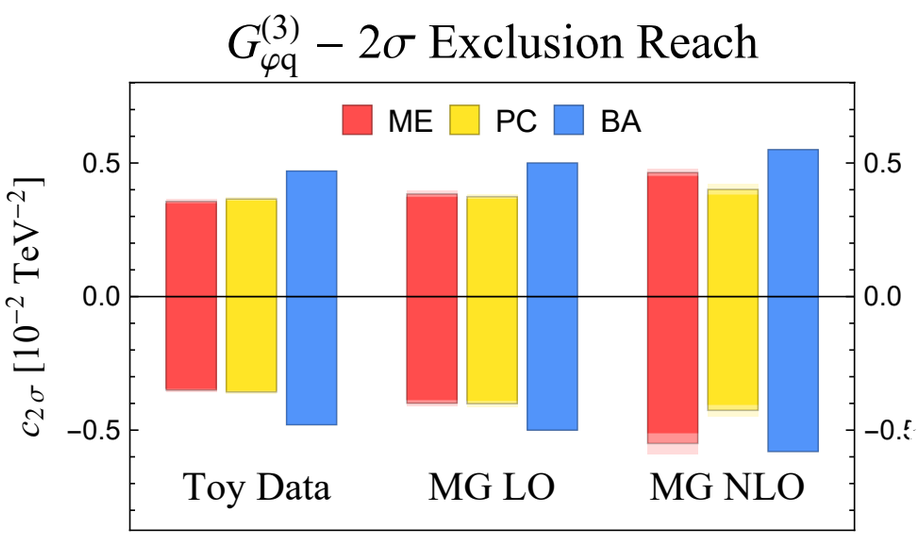
Results:
matrix element, classifier, and parametrised classifier

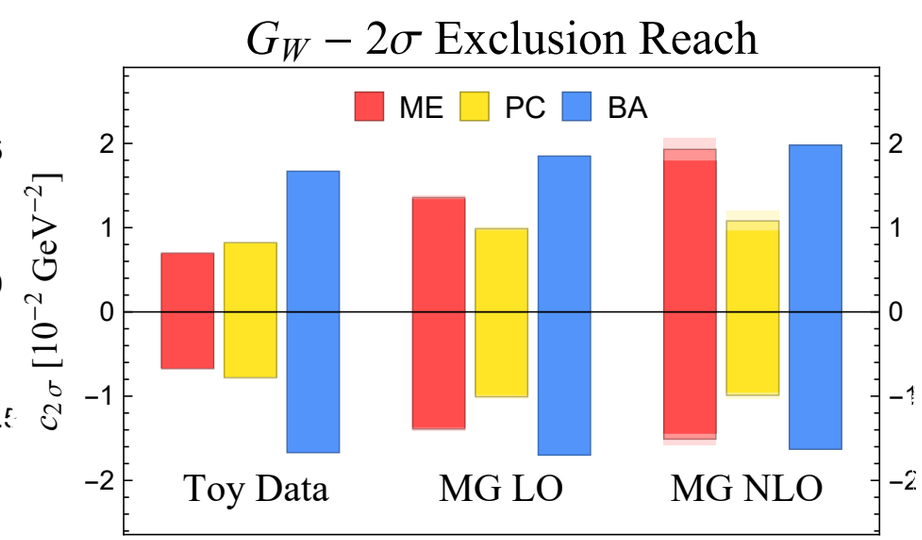
- We test our classifier on 3 types of data…

  toy data | MG LO | MG NLO

- … and compare the performance of 3 different methods

  ME matrix element | **QC** parametrised classifier | **BA** binned analysis





Work in progress: Reweighted events

- Event reweighing is a standard technique in simulation, for M Wilson coefficients $\{c_0, c_1, \ldots c_M\}$

$$z, [w_{c_0}, w_{c_1}, \ldots, w_{c_M}]$$

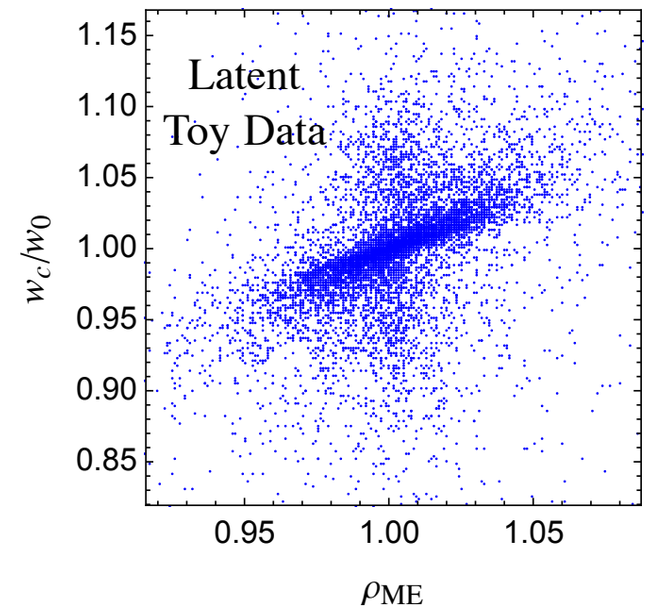
random variables

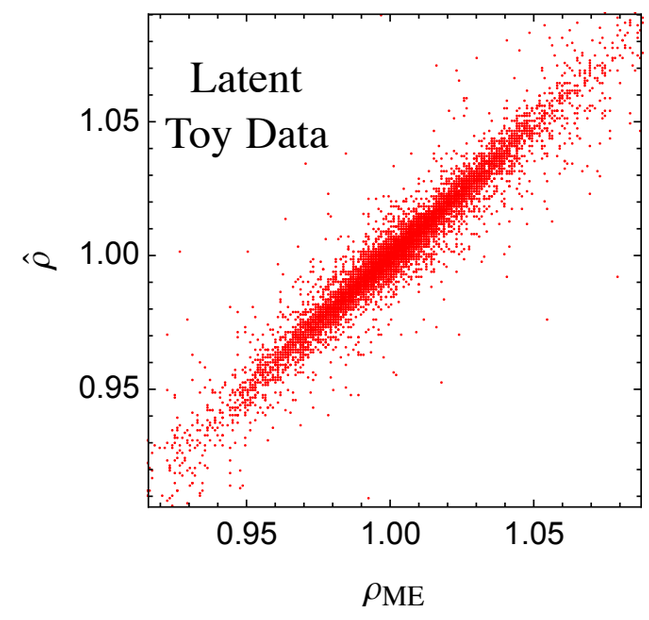
$$x, [w_{c_0}, w_{c_1}, \ldots w_{c_M}]$$

… BSM effects are considered independent from showers, detector effects, etc.

- Instead of generating samples according to multiple hypotheses in full, one sample with multiple reweights is enough.

$$\text{Loss} = \sum_{c_i} \left[\sum_{e \in H_0} w_0 [f(x)]^2 + \sum_{e \in H_1} w_{c_i} [1 - f(x)]^2 \right]$$





Work in progress: Reweighted events

- Event reweighing is a standard technique in simulation, for M Wilson coefficients $\{c_0, c_1, \ldots c_M\}$

$$z, [w_{c_0}, w_{c_1}, \ldots, w_{c_M}]$$

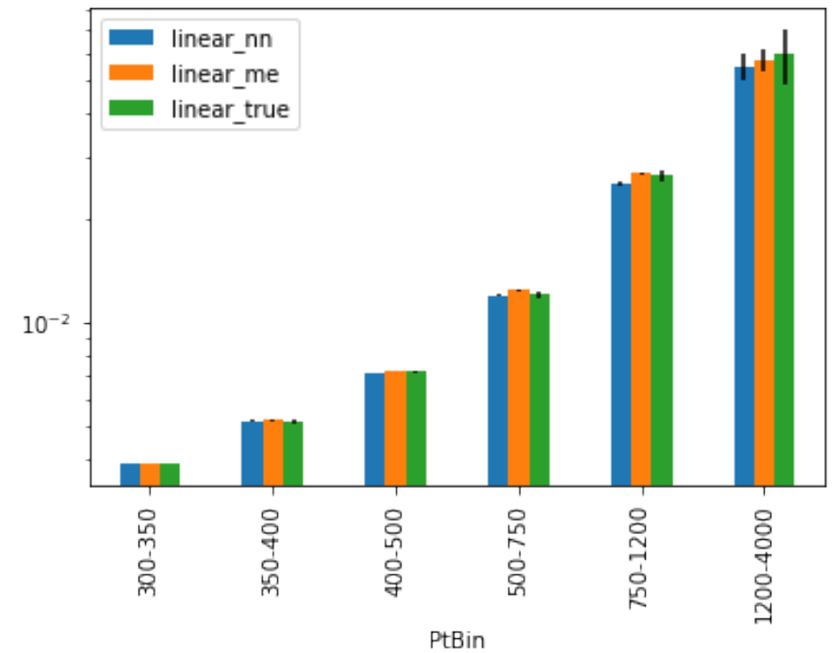
random variables

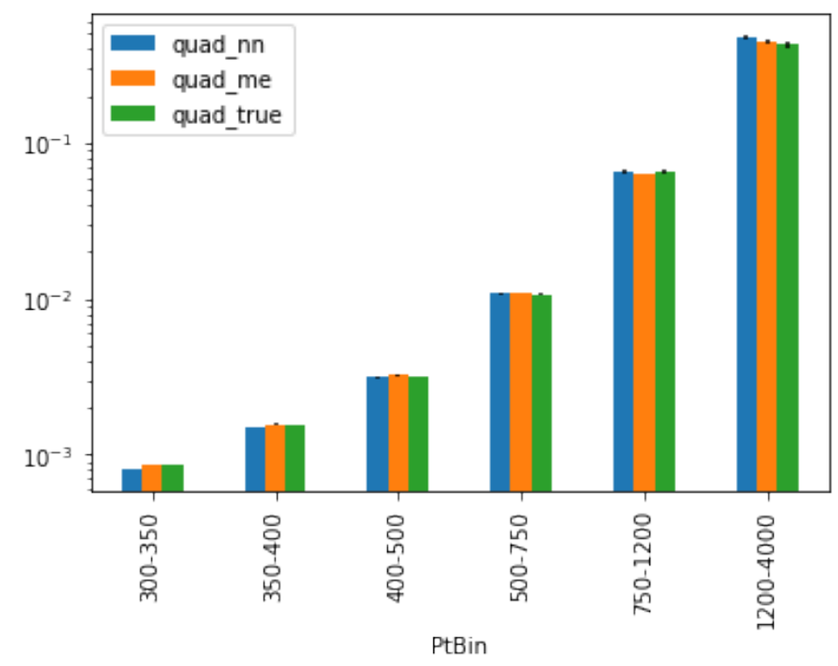
$$x, [w_{c_0}, w_{c_1}, \ldots w_{c_M}]$$

… BSM effects are considered independent from showers, detector effects, etc.

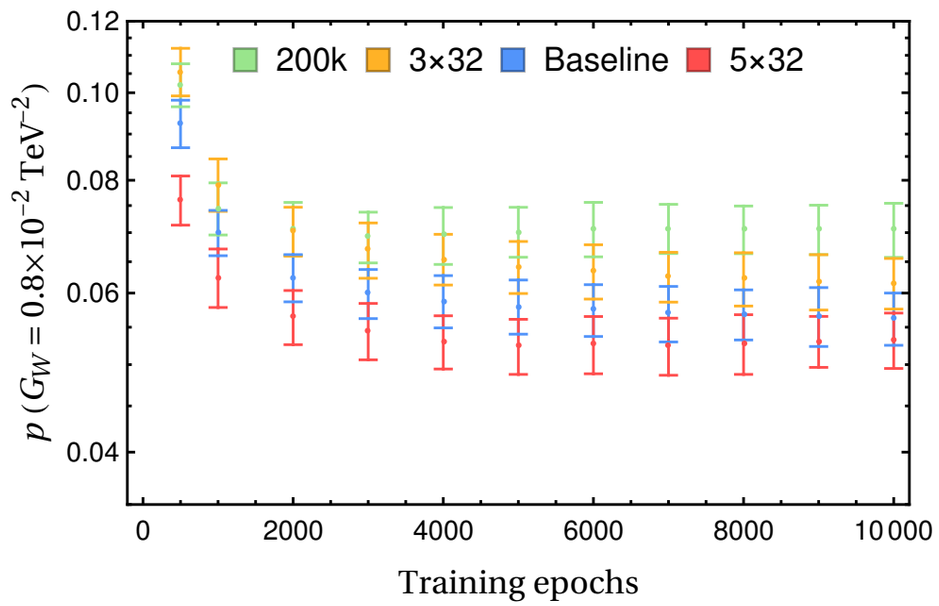
- Instead of generating samples according to multiple hypotheses in full, one sample with multiple reweights is enough.

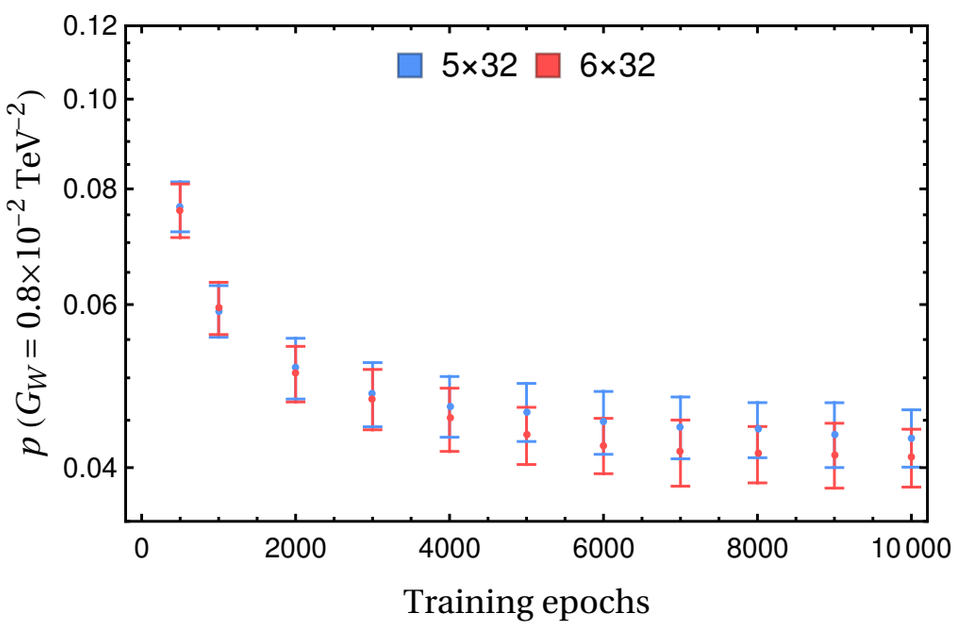
$$\text{Loss} = \sum_{c_i} \left[\sum_{e \in H_0} w_0 [f(x)]^2 + \sum_{e \in H_1} w_{c_i} [1 - f(x)]^2 \right]$$





Conclusions





- Approximating the likelihood is very useful in obtaining a high sensitivity in hypothesis testing

$$t_{\mathrm{NP}} = L(\mathrm{data}\,|\,H_1)/L(\mathrm{data}\,|\,H_0)$$

- Formulating the learning in terms of a classification problem derives the estimator of the likelihood ratio

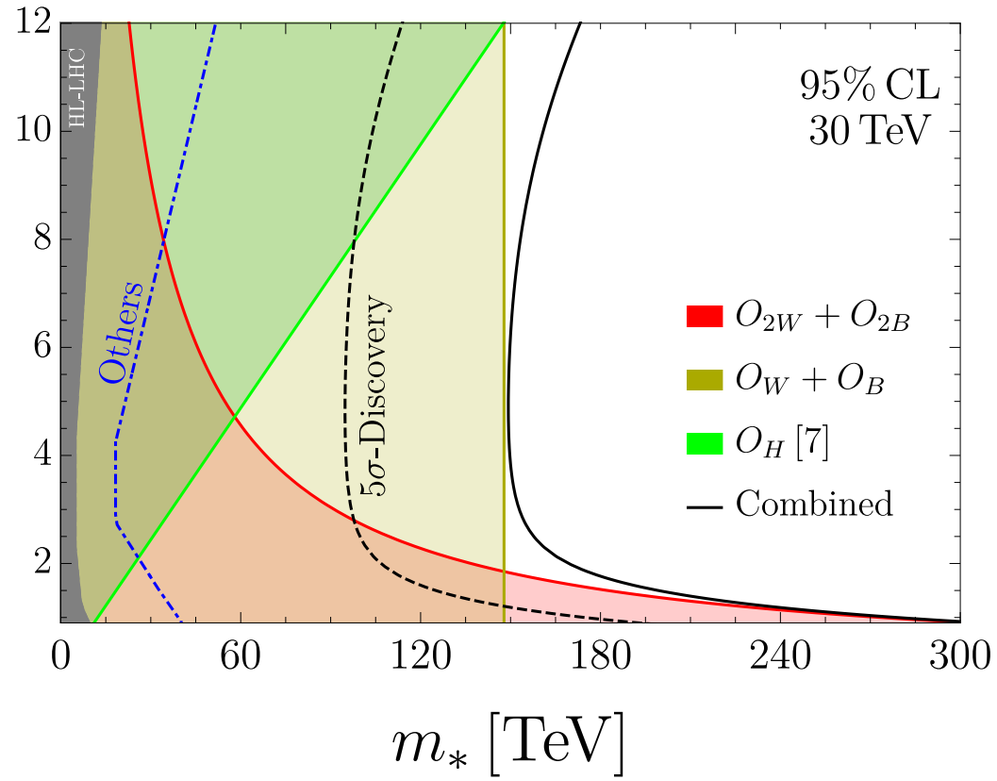
$$f(x) \xrightarrow{\text{large sample limit}} 1/(1 + r(x))$$

- Parametrisation gives better sensitivity because the complexity of the functional space is greatly reduced

$$r(x) = [1 + \alpha(x)c]^2 + \beta(x)^2 c^2$$

- Reweighting is a powerful technique to employ in this algorithm to avoid generating data multiple times

$$\{c_0, c_1, \ldots c_M\} \to \{w_0, w_1, \ldots w_M\}$$



Chen, S., Glioti, A., Rattazzi, R. *et al.* Learning from radiation at a very high energy lepton collider. *J. High Energ. Phys.* **2022**, 180 (2022).

Thank you!