Parametrized classifiers for optimal EFT sensitivity

<u>Alfredo Glioti</u> – Siyu Chen – Giuliano Panico – Andrea Wulzer

École Polytechnique Fédérale De Lausanne

2007.10356

Higgs and Effective Field Theory 2021
14/04/2021

EPFL

Motivations

Huge variety of possible measurements at the LHC
(high PT probes, Higgs couplings and distributions, …)

Huge variety of putative New Physics effects in model-independent (**EFT**) approach

⇩

Effective and **systematic** data analysis techniques needed to **maximize sensitivity**

Motivations

We can parametrize New Physics using the SMEFT Lagrangian

Theory ⇒ **Predictions**

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \theta_i O_i^{d=6} + \dots$$

Wilson coefficients

$$\sigma(\theta), \quad p(x|\theta) = \frac{1}{\sigma(\theta)} \frac{d\sigma(\theta)}{dx}$$

Measurable observables

Motivations

We can parametrize New Physics using the SMEFT Lagrangian

Theory

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \theta_i O_i^{d=6} + \dots$$

Wilson coefficients

More common approach

Predictions

$$\sigma(\theta), \quad p(x|\theta) = \frac{1}{\sigma(\theta)} \frac{d\sigma(\theta)}{dx}$$

Measurable observables

Simplified predictions

$$\sigma_i(\theta), \qquad i = 1, \dots N_{bins}$$

Are we losing information? Sometimes yes!

Motivations

We can parametrize New Physics using the SMEFT Lagrangian

Theory

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \theta_i O_i^{d=6} + \ldots$$

Wilson coefficients

Predictions

$$\sigma(\theta), \quad p(x|\theta) = \frac{1}{\sigma(\theta)} \frac{d\sigma(\theta)}{dx}$$

Measurable observables

Extracting the **full information** would require the likelihood p(x|θ) (as function of both x and θ)

How to access p(x|θ)?

Monte Carlo generators work in the "forward mode":

1) sample unobservable "partonic" variables z_{part} from a known $p(z_{part}|\theta)$
2) Transforms z_{part} to x, event by event

Unknown distribution of observables x

$$p(x|\theta) \approx \int dz_{part} p(x|z_{part}) p(z_{part}|\theta)$$

z_{part} is normally very far from x (e.g. invisible particles, NLO effects…)

$$x \neq z_{part}$$

Even worse if we include showering, hadronization, detector…

How to access $p(x|\theta)$?

Monte Carlo generators work in the "forward mode":

1) sample unobservable "partonic" variables z_{part} from a known $p(z_{part}|\theta)$
2) Transforms z_{part} to x, event by event

$$p(x|\theta) \approx \int dz_{det} dz_{had} dz_{show} dz_{part} p(x|z_{det}) p(z_{det}|z_{had}) p(z_{had}|z_{show}) p(z_{show}|z_{part}) p(z_{part}|\theta)$$

z_{part} is normally very far from x (e.g. invisible particles, NLO effects…)

$$x \neq z_{part}$$

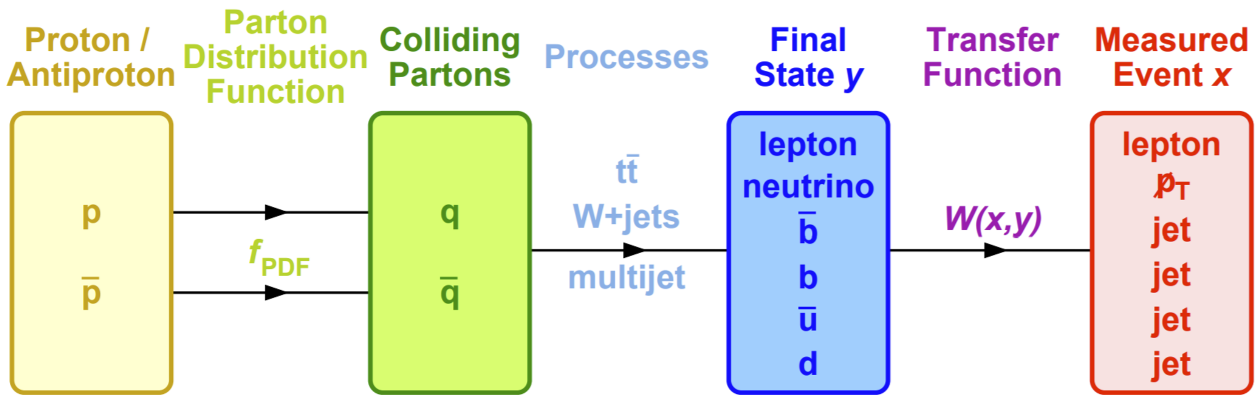
Even worse if we include showering, hadronization, detector…

How to access p(x|θ)?

The Matrix Element Method

See e.g.
Kondo , J. Phys. Soc. Jap. 57 (1988) 4126
Artoisenet & al. 1007.3300
Fiedler & al. 1003.1316
Martini & al. 1506.08798, 1712.04527

Approximate p(x|θ) as **parton level** + transfer function convolutions



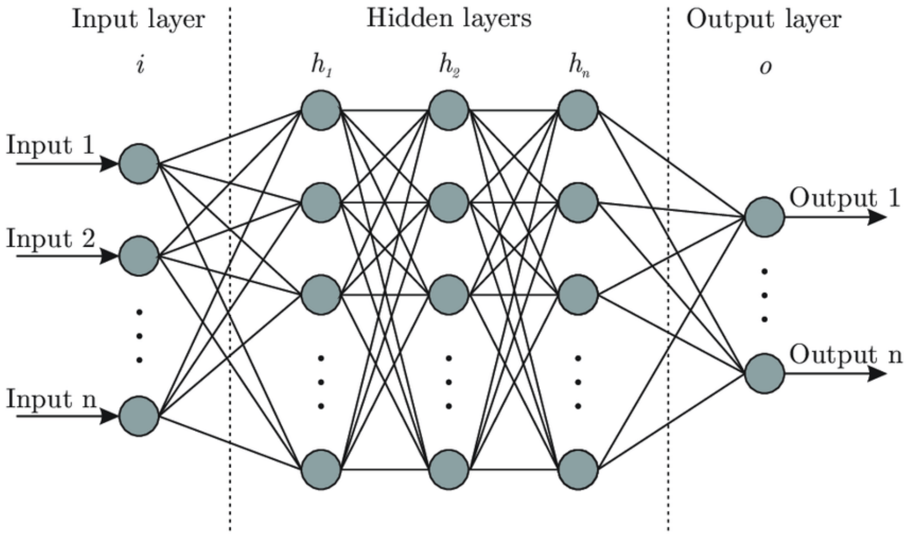
Problems:

- **phenomenological modeling** of transfer function is required
- **case by case design** required, **not** systematically improvable

$p(x|\theta)$ through Machine Learning

Brehmer & al. 1805.00013

Basic idea: approximate $p(x|\theta)$ with **Neural Networks**: $p(x|\theta) \leftrightarrow nn(x;w)$



The result will be **fully differential** on **all observables**, quick to evaluate and it can be obtained with a relatively small amount of Monte Carlo points.

No transfer functions modeling required.

Universal and systematically improvable

Simple Classifier

Consider **fixed** $\theta = \bar{\theta}$. We can **learn** $p(x|\bar{\theta})$ with respect to the SM ($\theta = 0$).

Training sample $\mathcal{T} = \{(x_i \sim p(x|0), y_i = 0), (x_i \sim p(x|\bar{\theta}), y_i = 1)\}$

$$\mathcal{L} = \frac{1}{N_T} \sum (nn(x_i; w) - y_i)^2$$

Loss function.
Minimized by training (wrt w)

Infinite training sample limit

$$\mathcal{L} \to \int dx \, [p(x|0)(nn(x) - 0)^2 + p(x|\bar{\theta})(nn(x) - 1)^2]$$

$$\frac{\delta \mathcal{L}}{\delta nn} = 0 \implies nn(x) = \frac{p(x|\bar{\theta})}{p(x|0) + p(x|\bar{\theta})} \implies \tau(x; \bar{\theta}) \equiv \frac{p(x|\bar{\theta})}{p(x|0)} = \frac{nn(x)}{1 - nn(x)}$$

Quadratic Classifier

Chen, AG, Panico, Wulzer 2007.10356

The simple classifier has some **drawbacks**:
- must be **trained for every value of** $\bar{\theta}$ **.**
- Target is **small θ** where **training is less effective**

NEW IDEA: use **θ as a training input** and train for **large values** and force a **quadratic dependence on θ** for the final output τ

$$\frac{d\sigma_{BSM}}{dx} = \frac{d\sigma_{SM}}{dx} + \theta \frac{d\sigma_1}{dx} + \theta^2 \frac{d\sigma_2}{dx} > 0$$

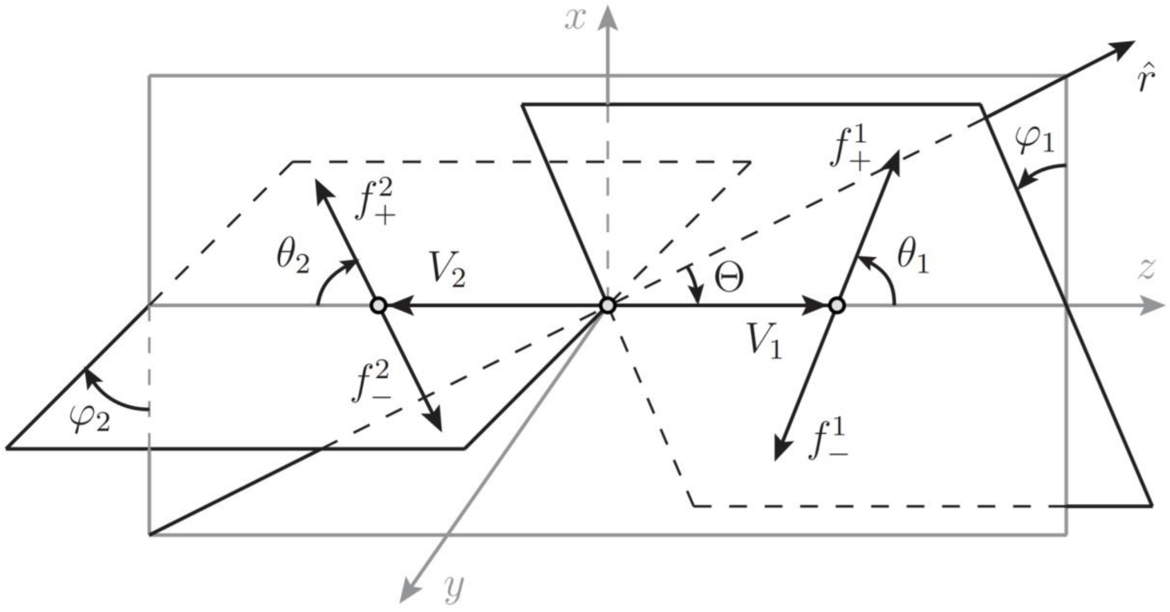
Loss function $\mathcal{L} = \frac{1}{N_T} \sum (nn(x_i) - y_i)^2$ where $nn(x) \to \tau(x;\theta) = (1 + \theta\, nn_\alpha(x))^2 + \theta^2 nn_\beta(x)^2$

We can **generalize** the Simple Classifier Loss function to train **two neural networks** to learn the **linear** and **quadratic** terms separately gaining more discriminating power!

Application: WZ production

Franceschini & al. 1708.07823
Panico & al. 1712.01310

$$p\,p \to W^{\pm}\,Z \to (l^{\pm}\,\nu)\,(l^{+}\,l^{-})$$



BSM contribution growing with collision energy from **two operators**

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

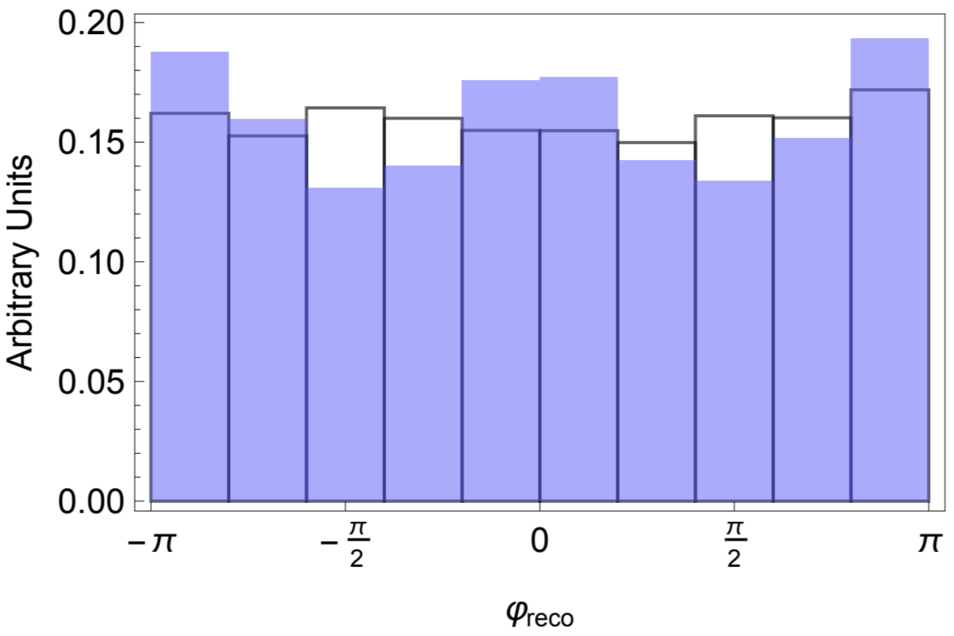
$$\mathcal{O}_W = G_W \varepsilon_{abc} W_\mu^{a\,\nu} W_\nu^{b\,\rho} W_\rho^{c\,\mu}$$

Six independent and discriminating variables: ŝ + 5 angles

Why measure the decay angles?

BSM and SM contribute to **different helicities** of W and Z

$$SM \to 0/0,\ \pm 1/\mp 1$$
$$\mathcal{O}_{\varphi q}^{(3)} \to 0/0$$
$$\mathcal{O}_W \to \pm 1/\pm 1$$

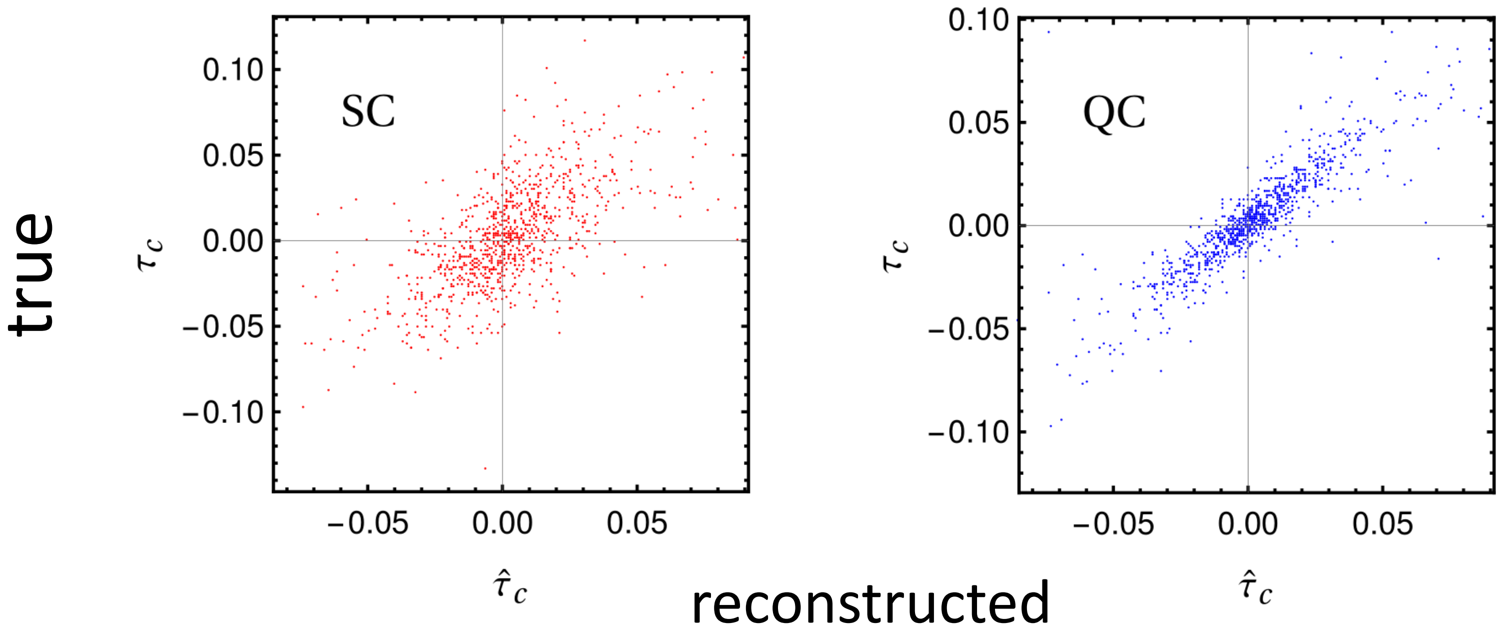


Integrating over the decay angles makes us **lose** this **discriminating information**.
For **Ow** integrating kills **interference with the SM**

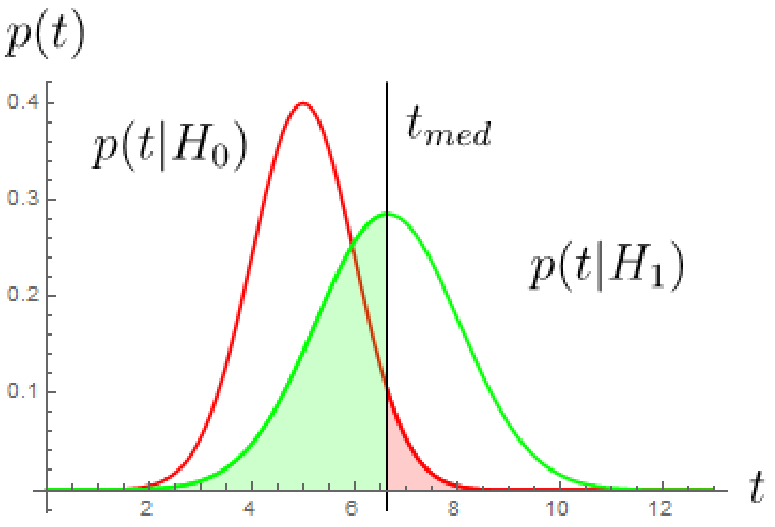
Performance studies

Homemade Monte Carlo implementing **approximate analytic** LO distribution

Simplest (qualitative) test is **true vs reconstructed** pdf ratio



Performance studies



Neyman-Pearson lemma: log-likelihood ratio is the optimal test statistics
(max power at fixed significance)

$$t(\mathcal{D};\theta) = -2\left(\bar{N}_\theta - \bar{N}(0) + \sum_{x_i\in\mathcal{D}} \log\left[\frac{\bar{N}(\theta)}{\bar{N}(0)}\frac{p(x_i|\theta)}{p(x_i|0)}\right]\right)$$

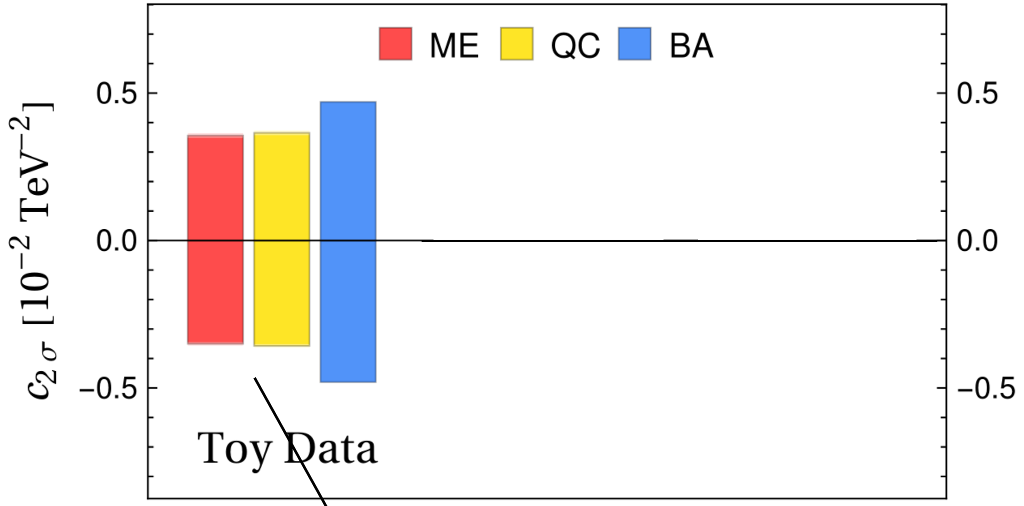
$\tau(x;\theta)$

Neyman-Pearson reach with **true vs reconstructed τ** defines a notion of "**optimality**" to objectively compare the different analysis strategies.

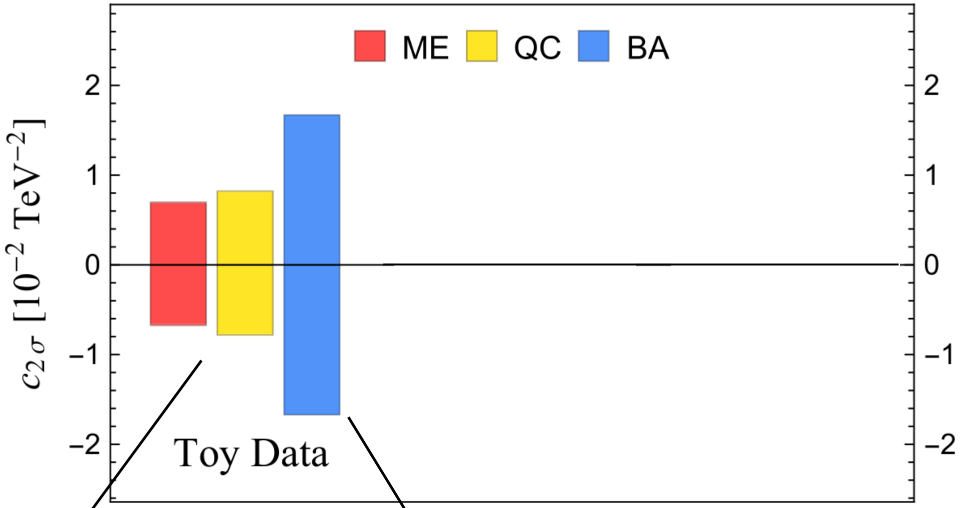
Results

ME = Toy Matrix Element; **QC** = Quadratic Classifier; **BA** = Binned Analysis

$G^{(3)}_{\varphi q}$ – 2σ Exclusion Reach



G_W – 2σ Exclusion Reach

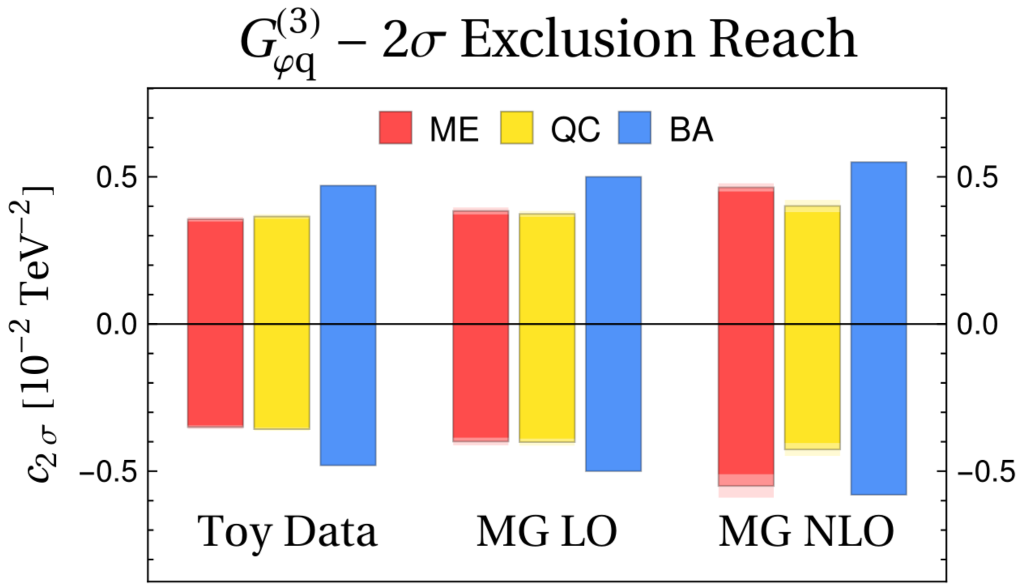


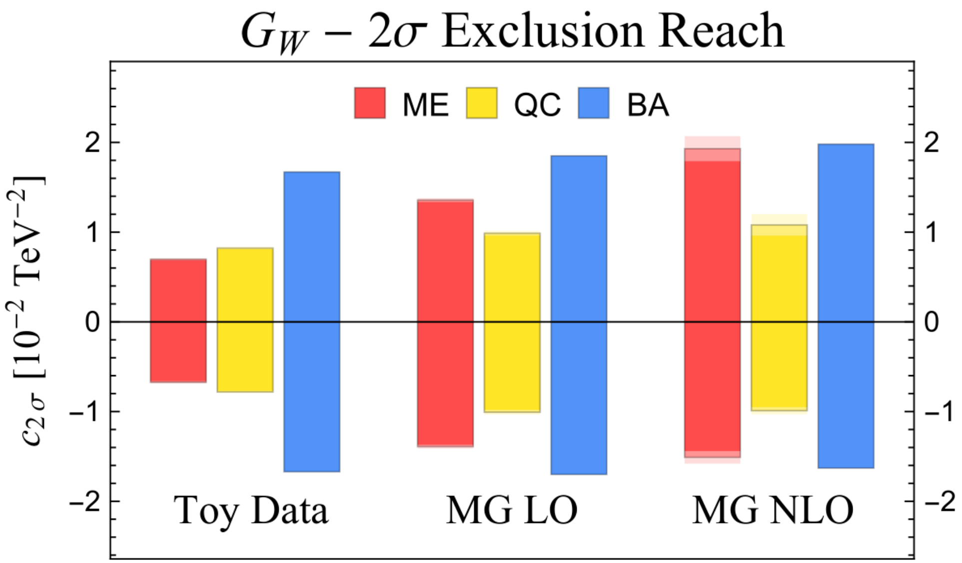
Our method is "nearly optimal". More training points bridge residual gap.

Factor of 2 improvement possible compared to Binned Analysis

Results

ME = Toy Matrix Element; **QC** = Quadratic Classifier; **BA** = Binned Analysis





Neural Network reach is stable (same architecture and same training size)
Toy Matrix Element becomes ineffective already at LO
Reach improvement confirmed also at NLO

Conclusions

Multivariate analysis can greatly **increase the sensitivity** on BSM parameters, especially when new physics enters in multiple observables with a complex interference pattern.

Standard analysis methods **cannot** handle multivariate analysis while also being systematically and easily improvable.

Machine learning methods can overcome these difficulties by **learning the fully differential distribution** accurately from a Monte Carlo sample. Making the approximation more reliable is just a matter of increasing the training sample size and network architecture.

Embedded **quadratic functional dependence** enables more **accurate ratio reconstruction** in the Wilson Coefficient parameter space (generalization to several coefficients trivial).

In progress: including PDF uncertainties, quantifying the impact of detector (Delphes) effects, …