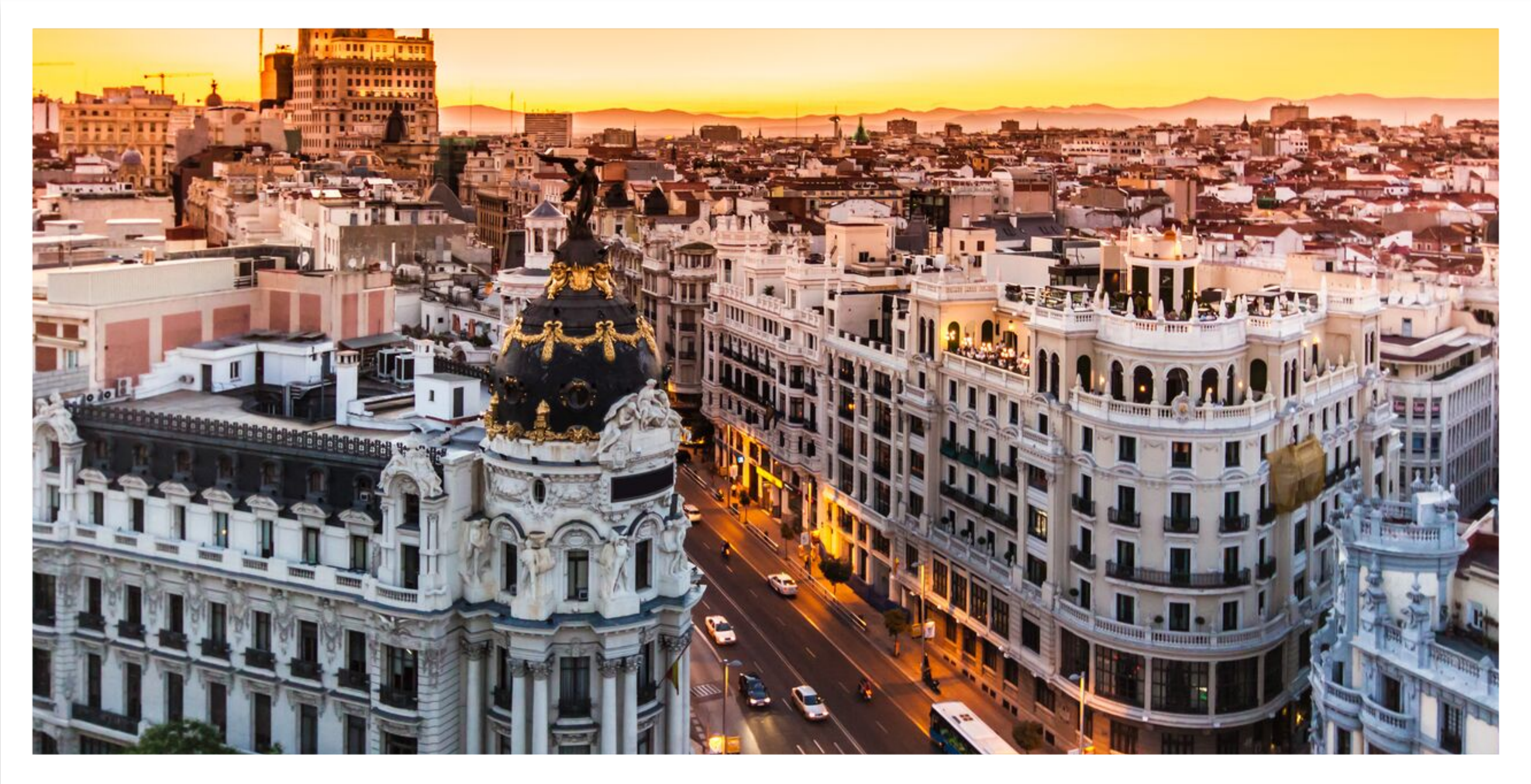


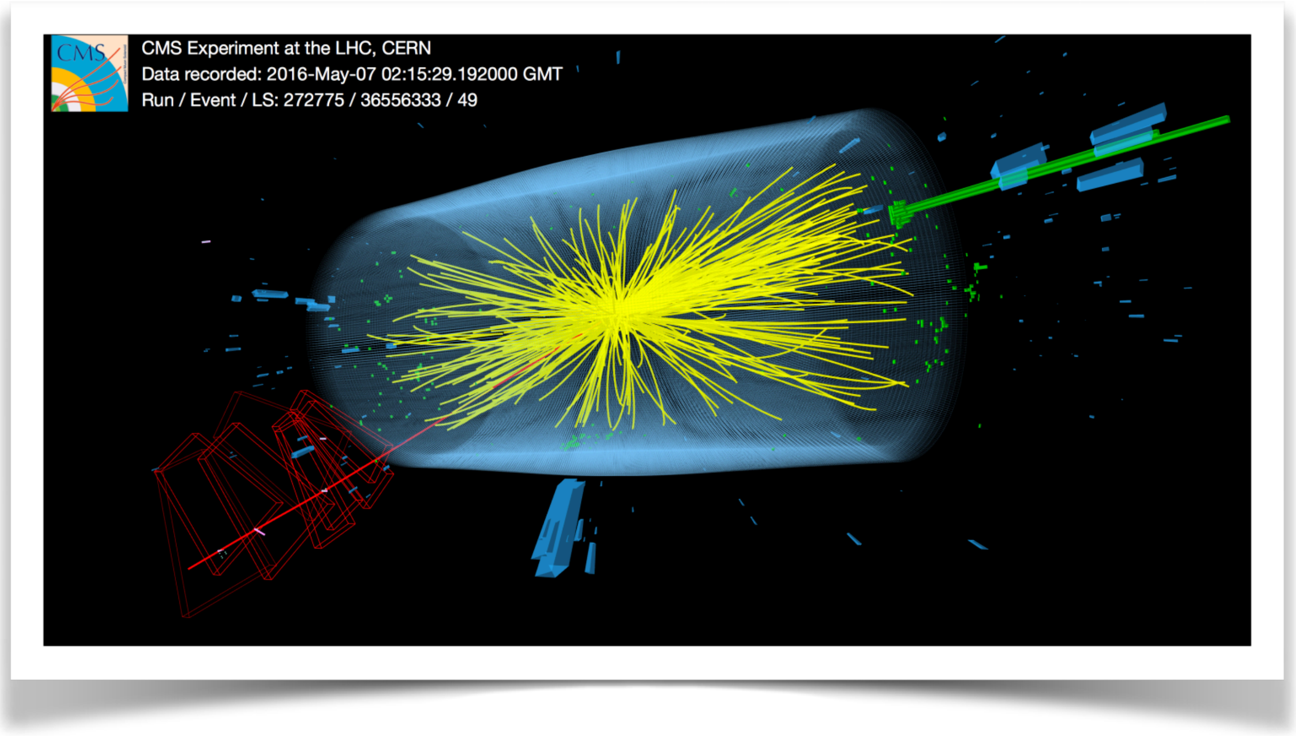
AI goes MAD

Symbolic regression for precision LHC physics

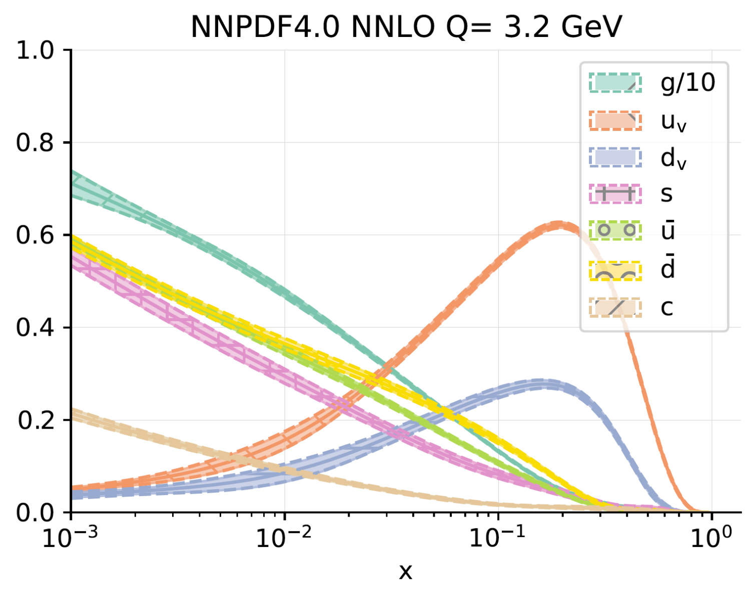
Manuel Morales-Alvarado

INFN, Sezione di Trieste, SISSA

Outline

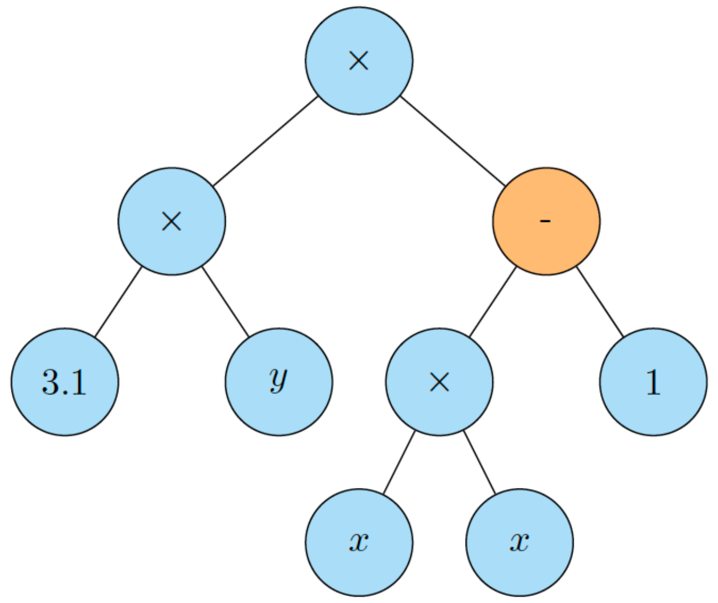


Some high energy physics



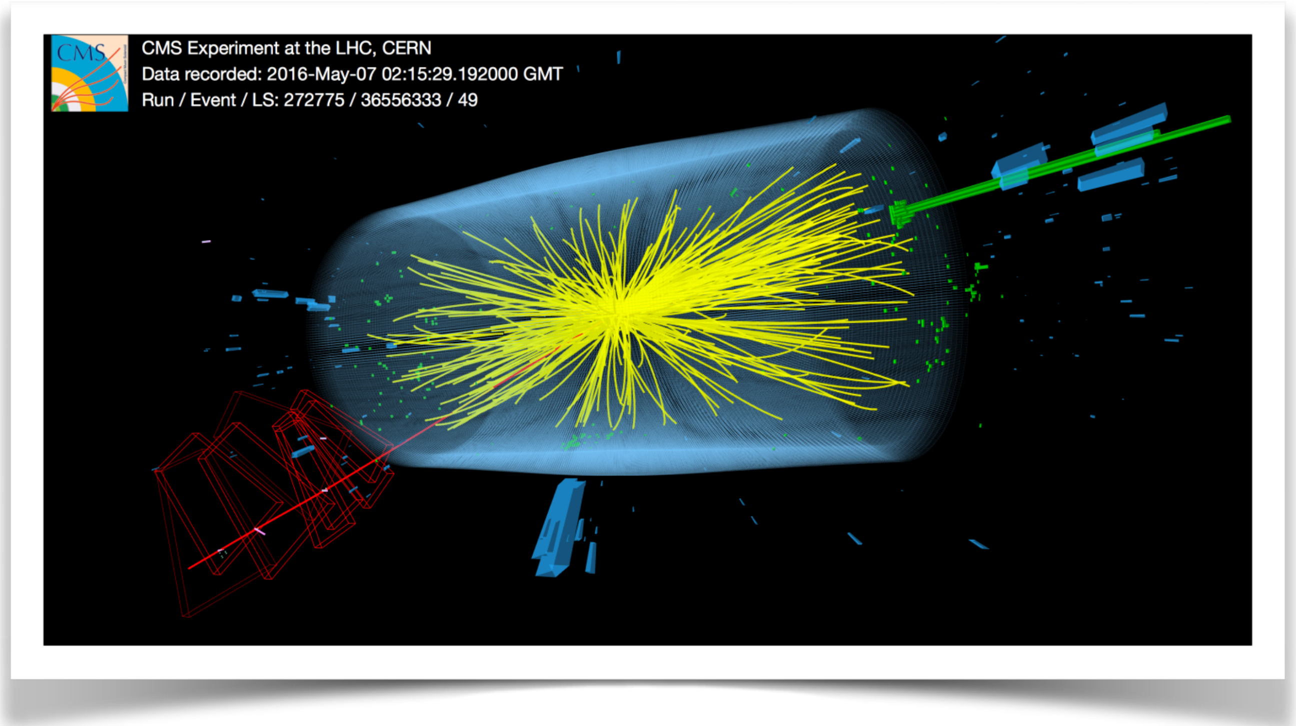
$$+ \quad \frac{c_i}{\Lambda^2}$$

Precision: deep learning the structure of hadrons and new physics



Symbolic regression

Outline



Some high energy physics

$$+ \frac{c_i}{\Lambda^2}$$

Precision: deep learning the structure of hadrons and new physics

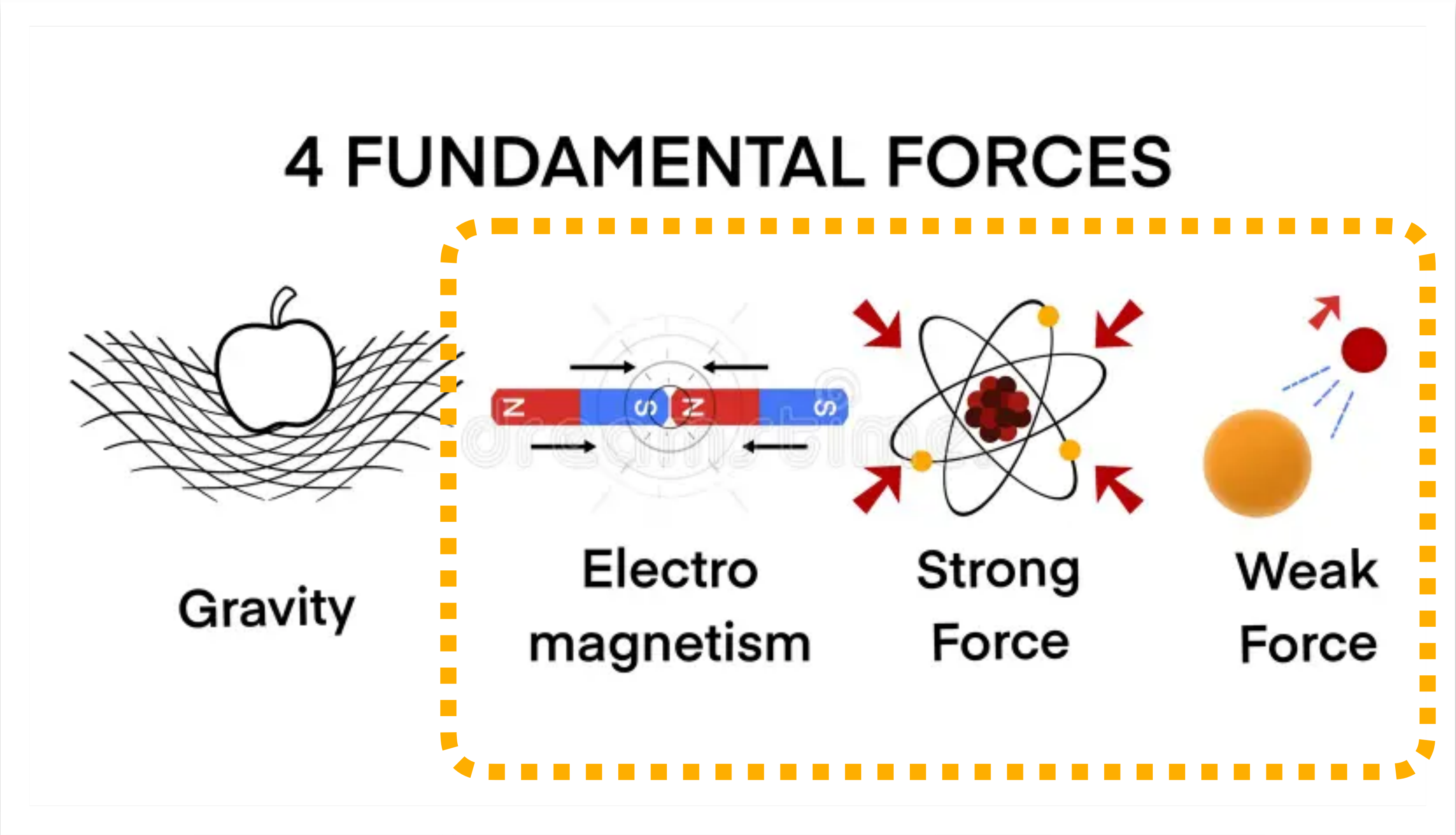
Symbolic regression

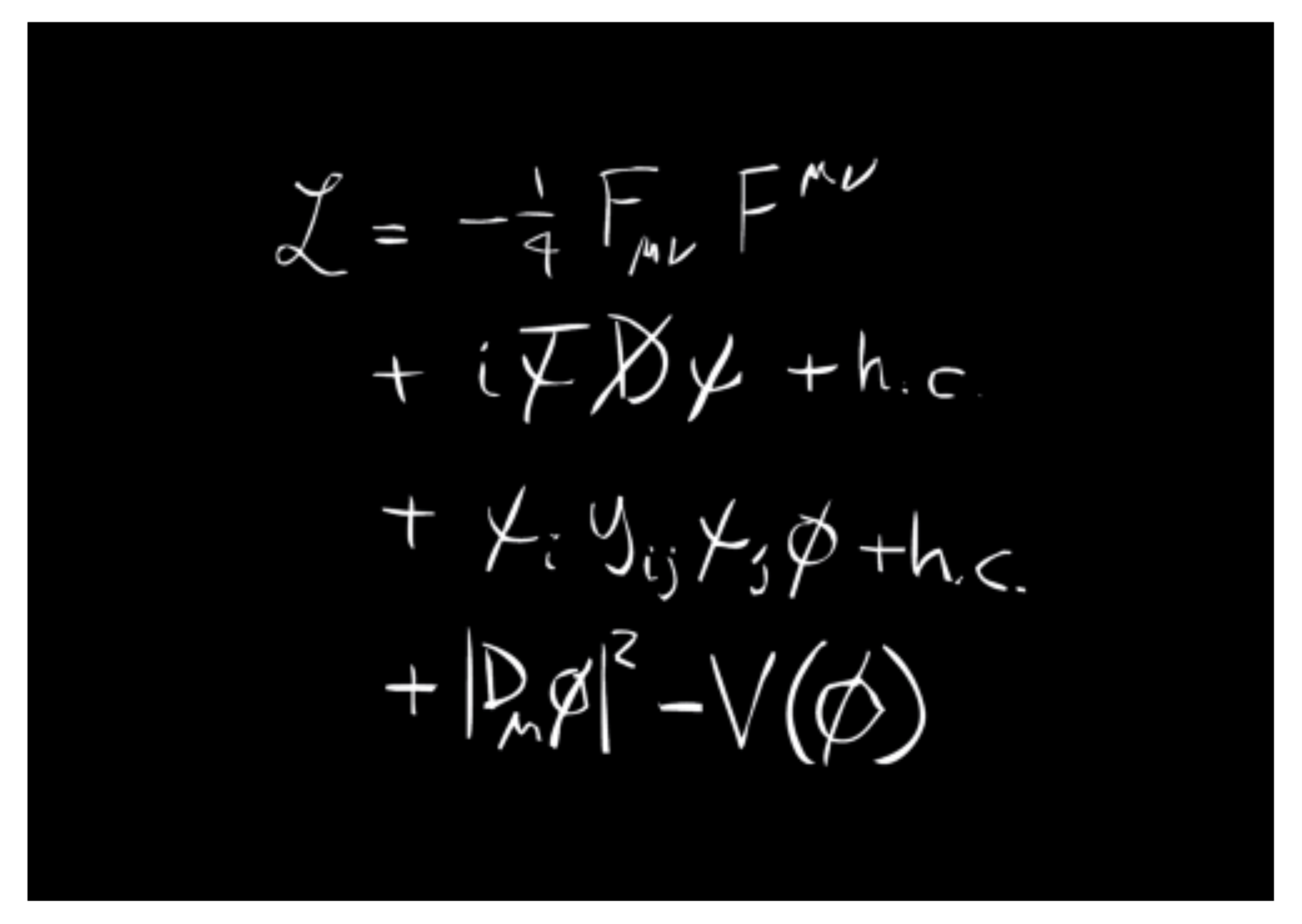
Fundamental interactions and particles

High energy physics: the description of fundamental particles and their interactions. It is (very) well described by the Standard Model (SM).

- The SM is a gauge theory that describes elementary particles and their interactions in terms of relativistic quantum fields.

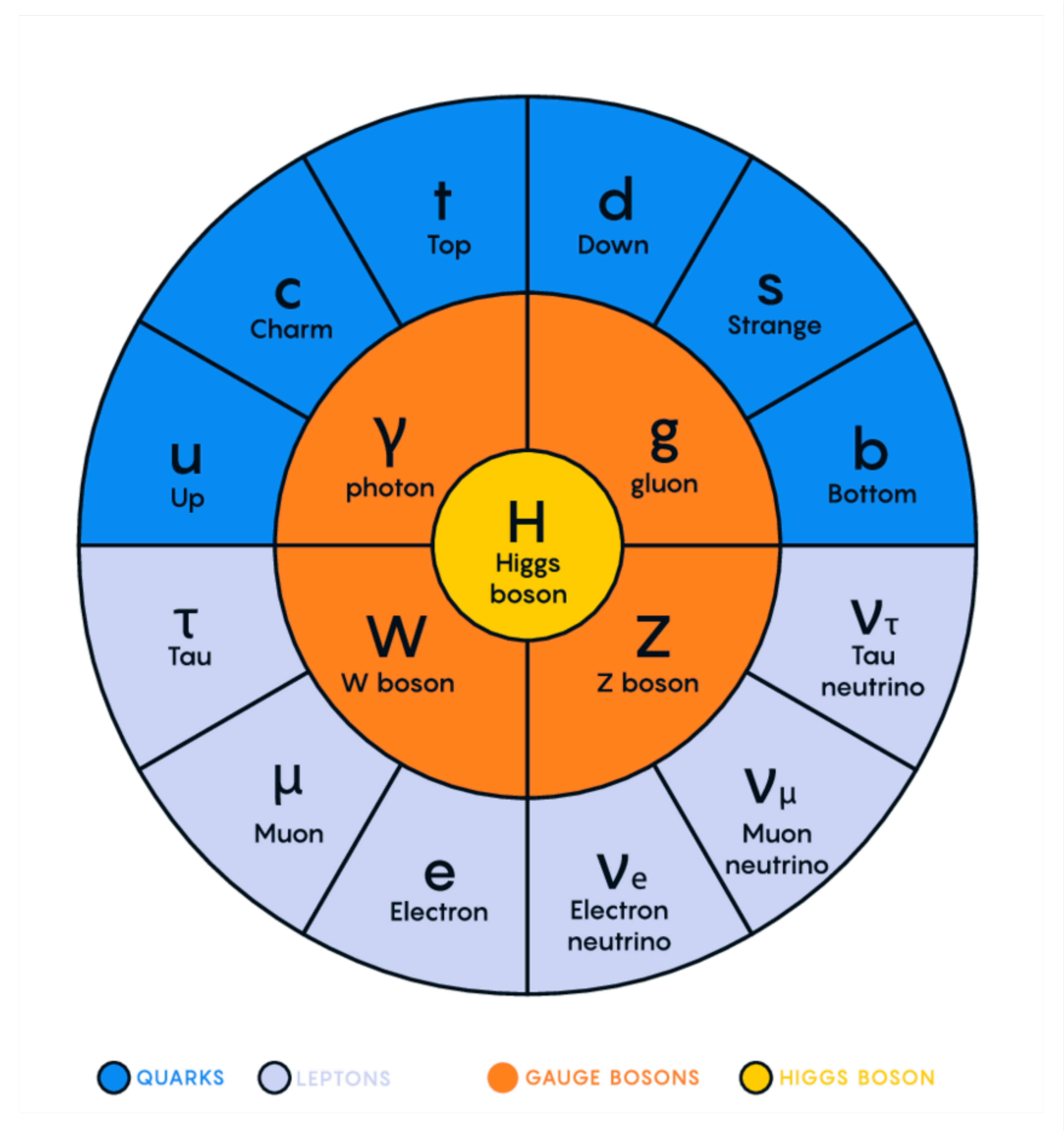
- The SM describes 3 of the 4 fundamental forces through gauge bosons, spin = 1 particles.
  - Electromagnetism (**photon**)
  - The strong force (**gluons**)
  - The weak force (**W** and **Z bosons**)





Fundamental interactions and particles

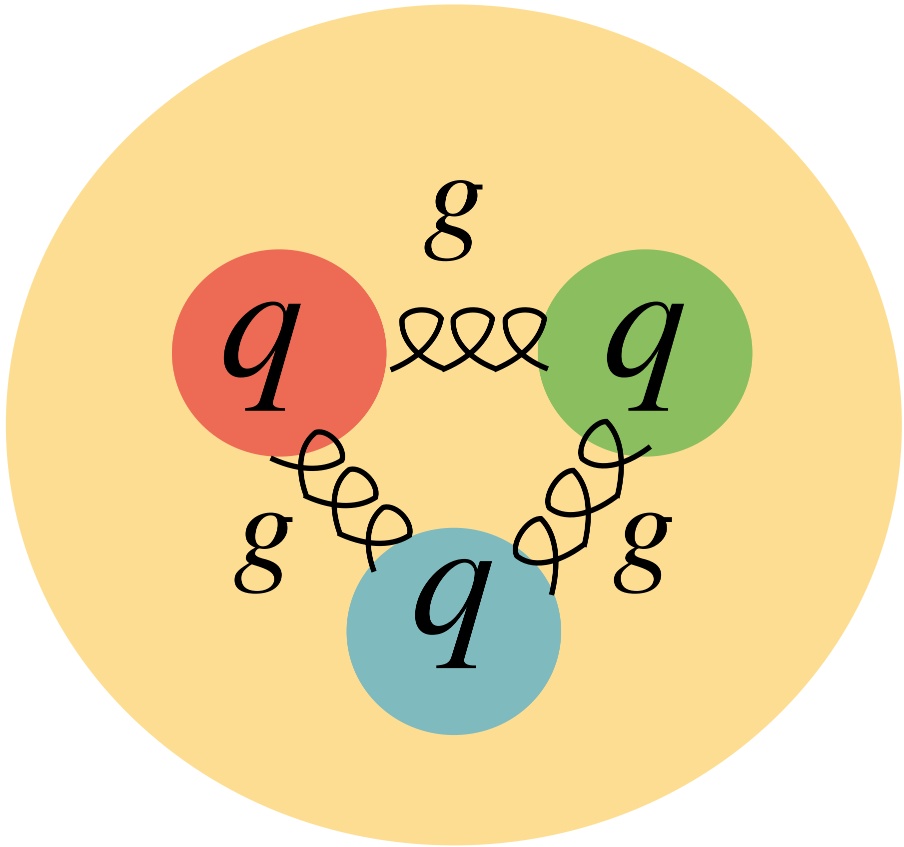
- In the SM there are also **quarks** and **leptons**, spin = 1/2 fermions.
- Quarks carry colour charge and interact via the strong force, whereas leptons do not.
- In the SM we also have the **Higgs** boson, a spin = 0 particle, responsible for giving mass to the elementary particles.



Protons and partons

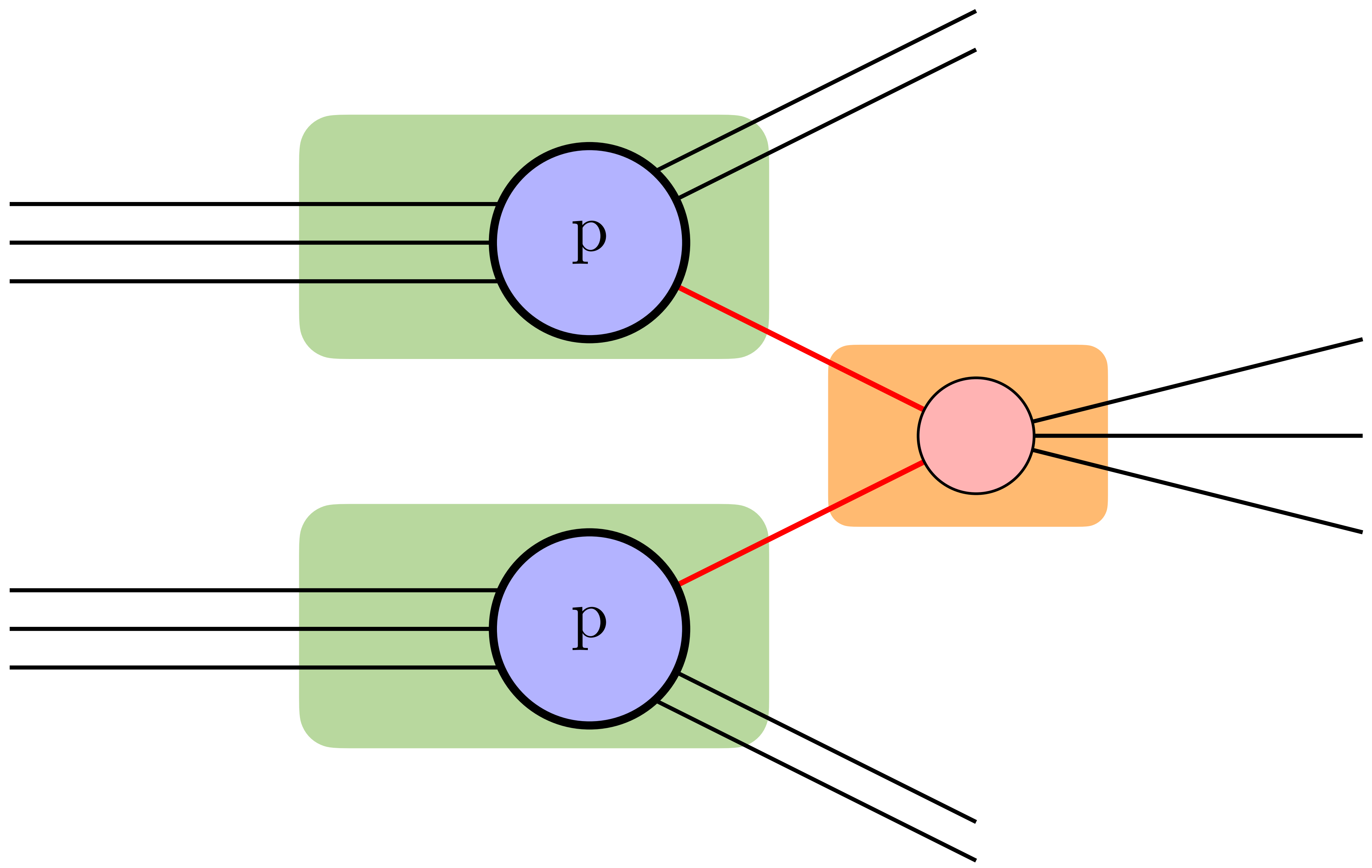
- At the LHC we collide *protons*.
- Proton are not elementary particles. They are QCD bound states of elementary particles called *partons* (quarks, gluons, …).
- The parton model postulates that interactions between hadrons (protons, in particular) are interactions of point-like parton convolved with functions that parametrise the structure of the hadron: *parton distribution functions* (PDFs).

Proton



Factorisation and hadronic observables

Consider a proton-proton collision:



$$\sigma = \sum_{i,j} \int_0^1 dx_1 \int_0^1 dx_2 f_i(x_1) f_j(x_2) \hat{\sigma}_{ij}(x_1, x_2)$$

$x_{1,2}$: fraction of the hadron's momentum that is carried by the hadron

$\hat{\sigma}_{ij}$: partonic cross section

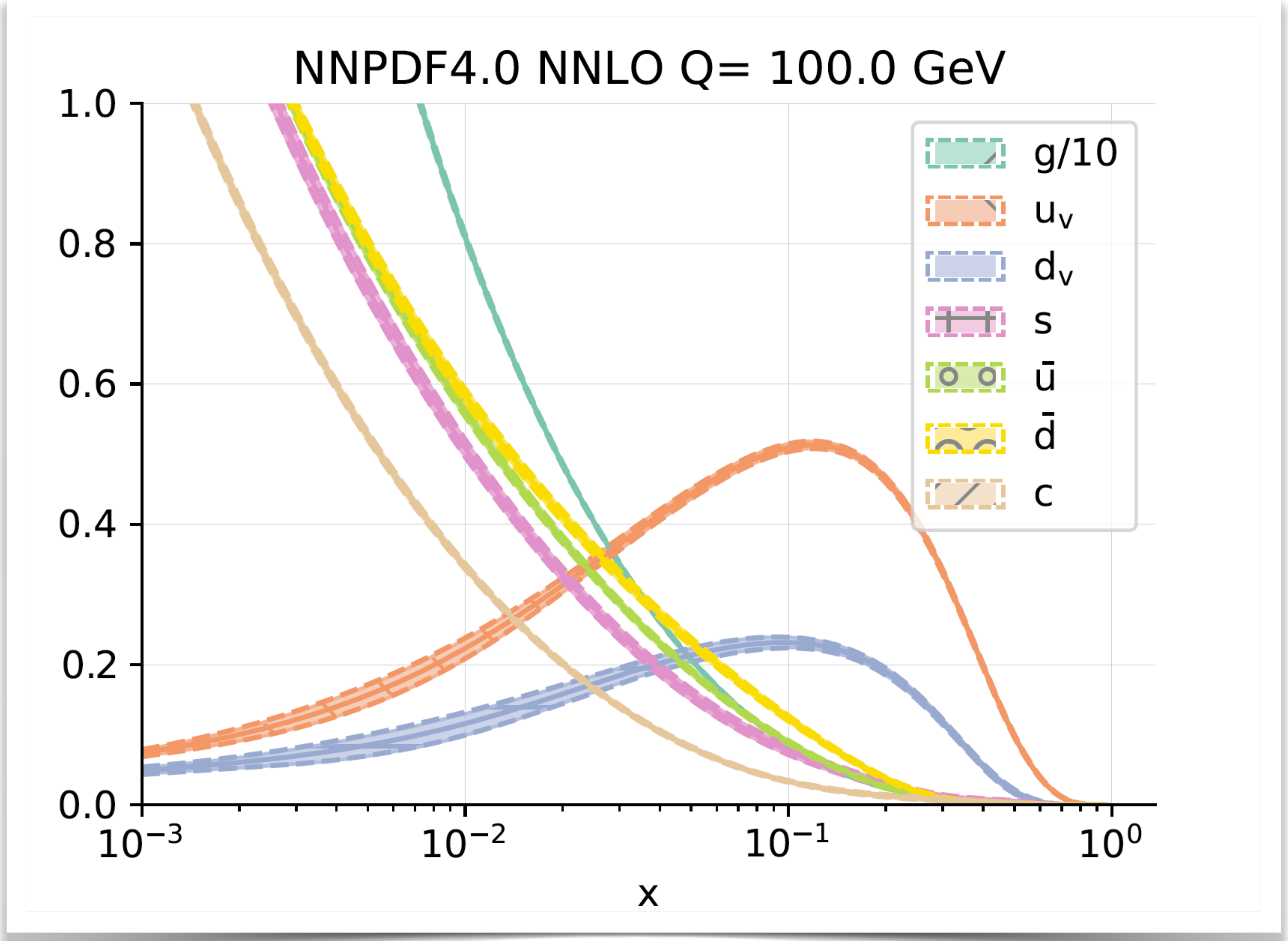
$f_i(x)$: PDF of parton of type i

- Radiative corrections introduce IR singularities that have to be factorised: $f(x) \to f(x, Q^2)$

PDF determination

PDFs **cannot** be calculated from first principles in perturbation theory, they have to be extracted from *data*.

$$f(x, Q^2)$$



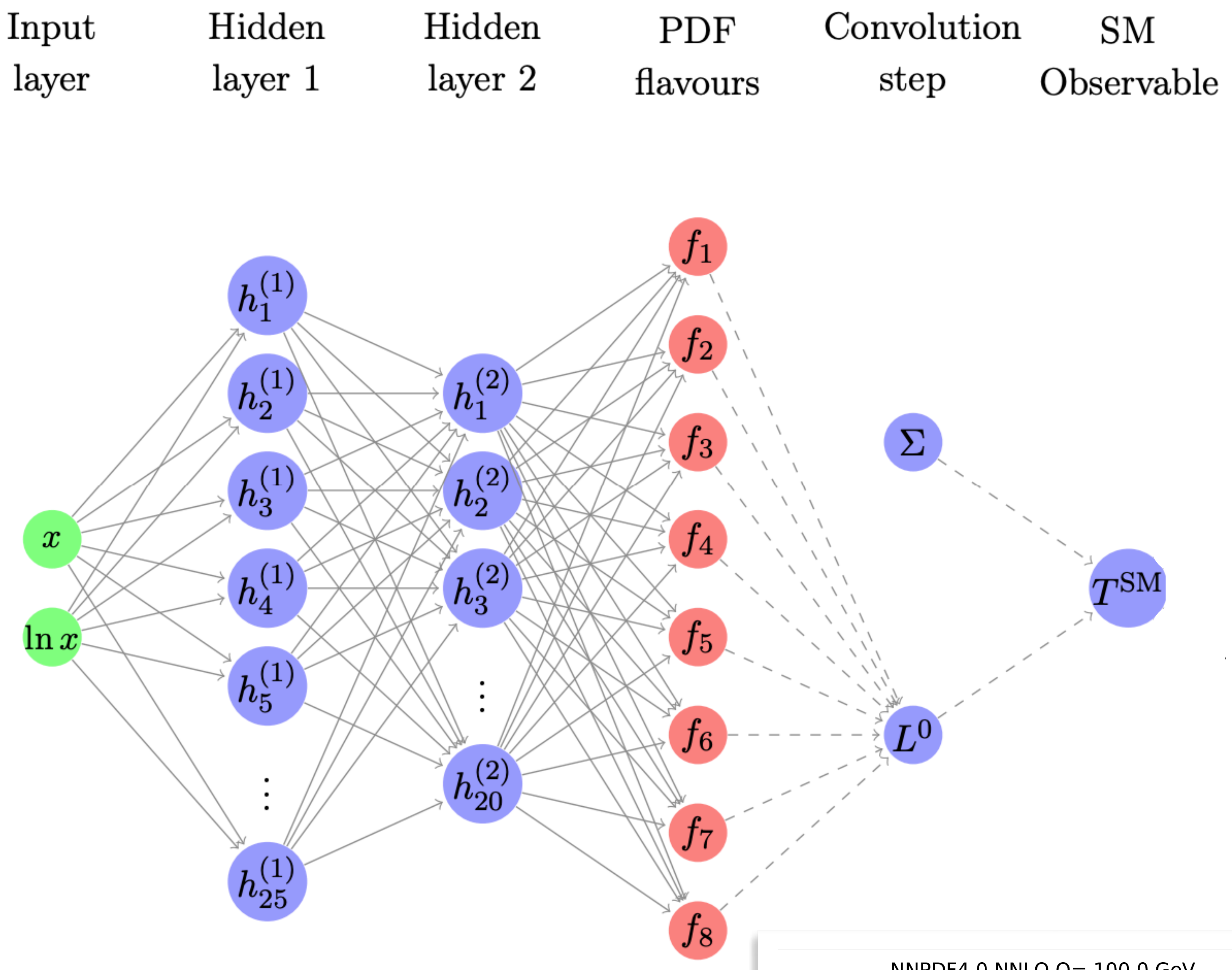
Some recent PDF fits include:

- NNPDF4.0, Ball et al., 2109.02653
- MSHT20, McGowan et al., 2012.04684
- CT18, Hou et al., 1912.10053

NNPDF4.0

- PDFs are parametrised by neural networks (NNs):

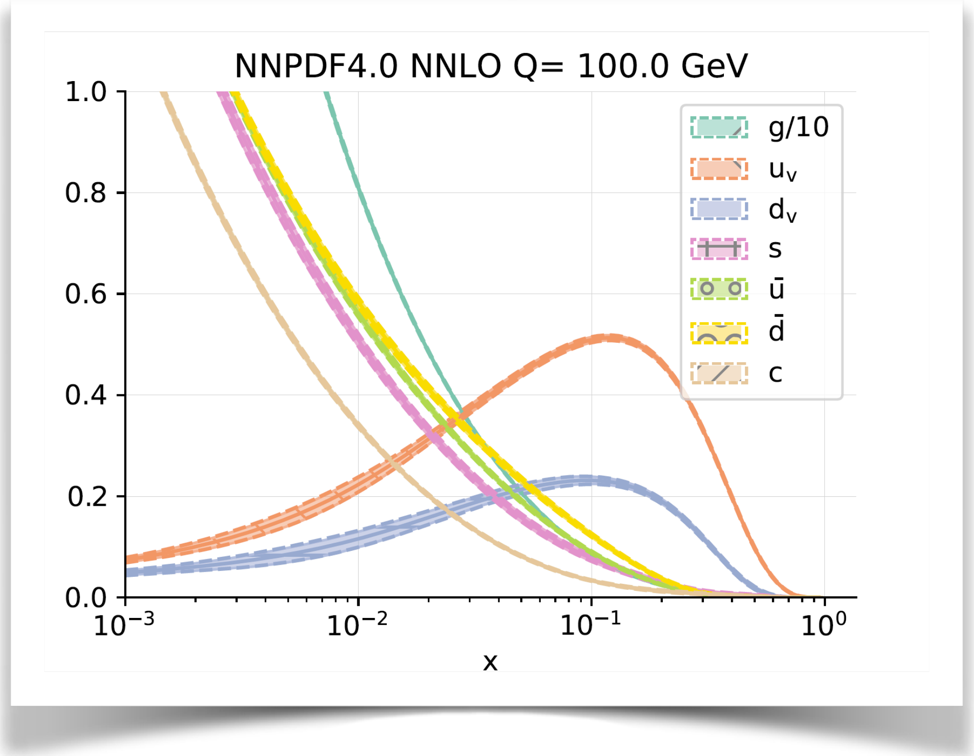
$$f(x, Q_0^2) = \mathrm{NN}(x)$$



- The optimal weights of the NN are found by minimising a loss function:

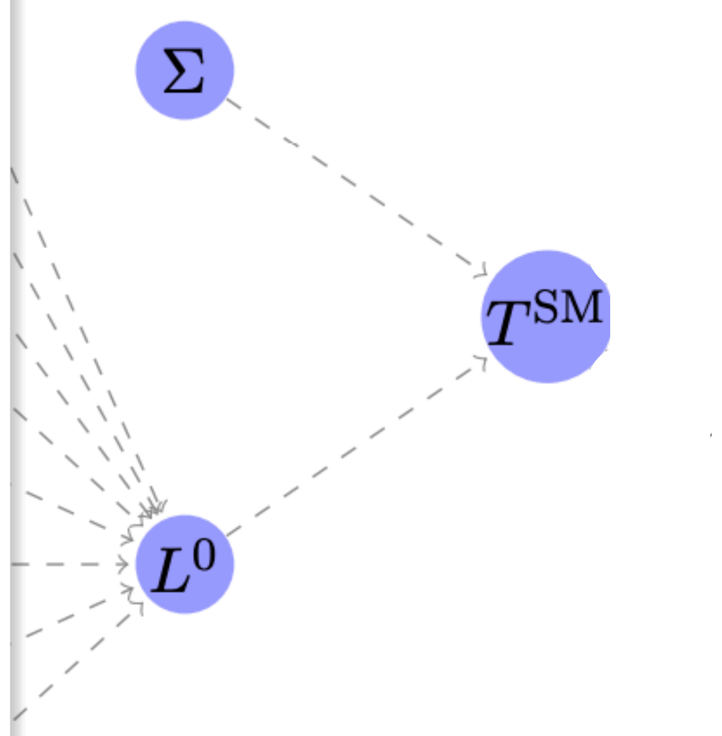
$$\chi^2(\theta) = \frac{1}{N_{\mathrm{dat}}}(\mathbf{D} - \mathbf{T}(\theta))^T(\mathbf{cov})^{-1}(\mathbf{D} - \mathbf{T}(\theta))$$

$$\theta = \theta_{opt}$$



- Uncertainty is propagated via de Monte Carlo replica method. 2109.02653, 2404.10056

NNPDF4.0



- PDFs are par
networks (N

$f(x,$

Other parametrisations are possible. For example, in the case of MSHT20, PDFs are parametrised with *fixed functional forms* of the type:

$$xf(x,Q_0^2) \;=\; A(1-x)^{\eta}x^{\delta}\left(1+\sum_{i=1}^{n}a_iT_i^{\mathrm{Ch}}(y(x))\right)$$

T_i^{Ch} : Chebyshev polynomials

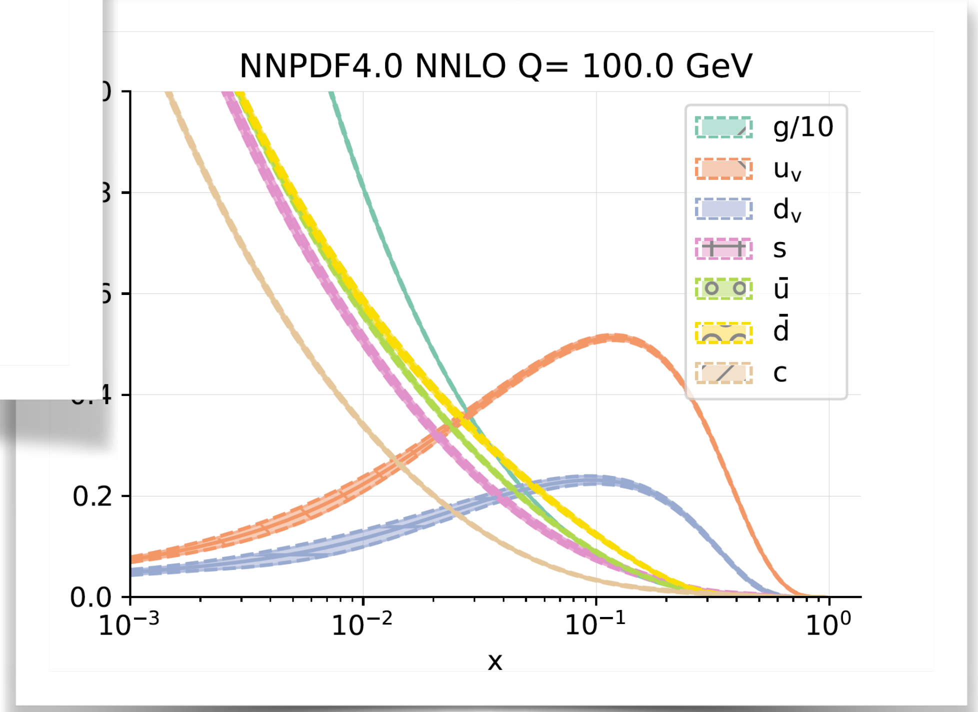
- The optimal
minimising a

$\chi^2(\theta) = \frac{1}{N_{\mathrm{da}}}$

where

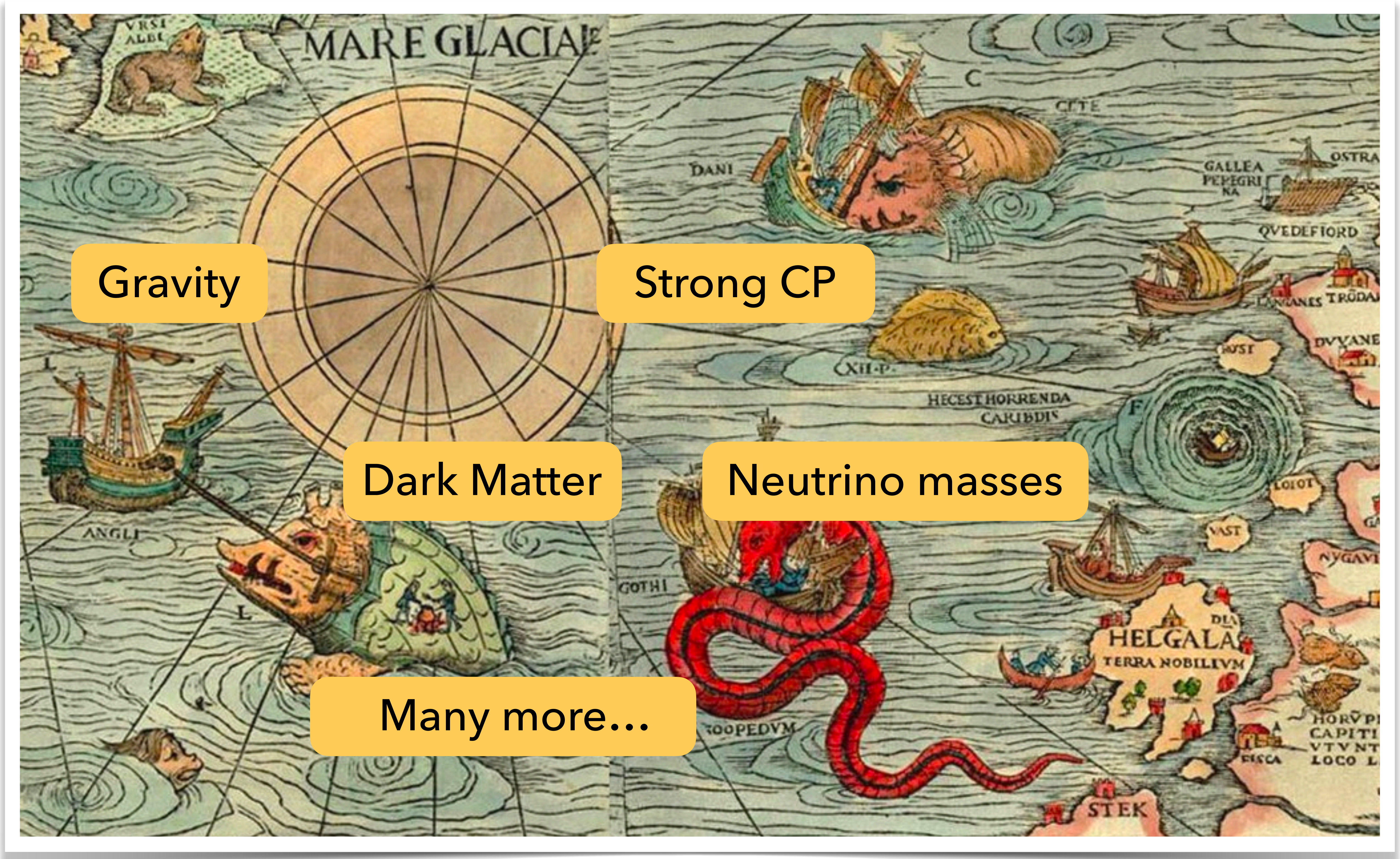
- Uncertainty is propagated via de Monte Carlo replica method.

2109.02653, 2404.10056



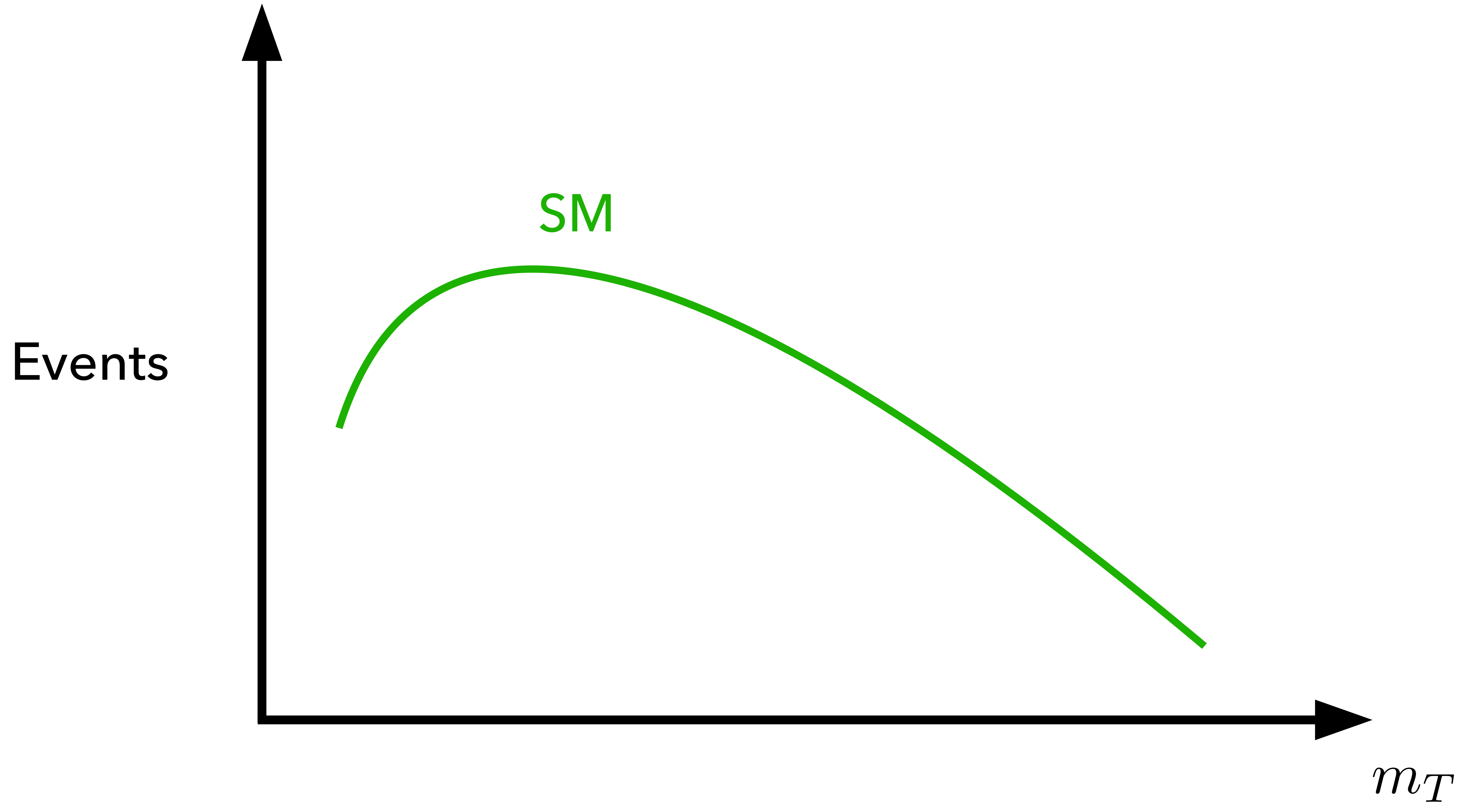
Is that it?

All that we have discussed so far is in the context of the SM, but we know that it cannot be the whole story …



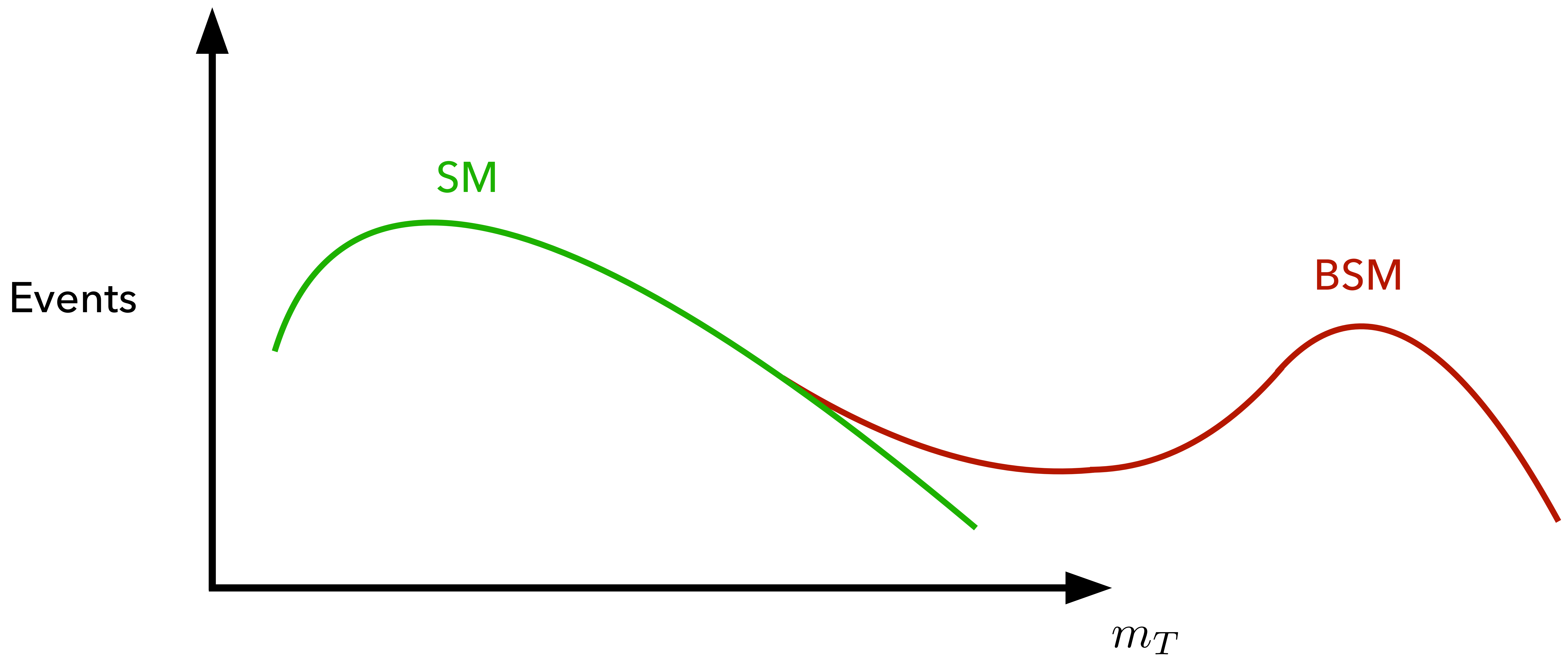
Beyond the SM

We can look for physics beyond the SM (BSM)!



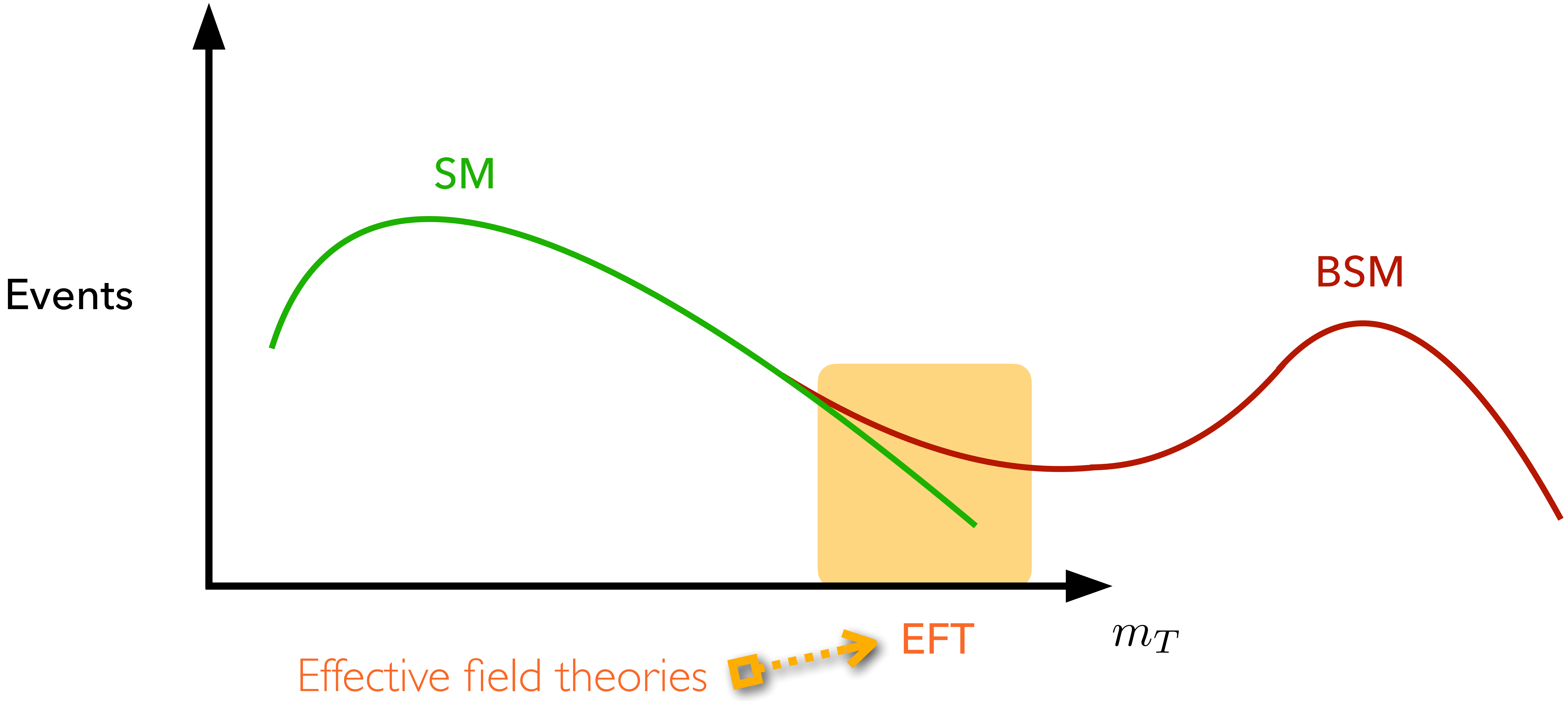
Beyond the SM

We can look for physics beyond the SM (BSM)!



Beyond the SM

We can look for physics beyond the SM (BSM)!



SM effective field theory

- The deviations from the SM predictions can be parametrised in terms of the SM effective field theory (SMEFT).
- The SMEFT supplements the SM Lagrangian with higher dimensional operators whose action is suppressed by a new physics scale:

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i}{\Lambda^2} Q_i^{(6)} + \cdots$$

$\{c_i\}$	: Wilson coefficients
$\{Q_i^{(6)}\}$	: Dimension-6 operators
Λ	: New physics scale

SMEFT fits

Precision is essential!

- The mismatch between the SM and the experiment is:

Linear in the c_i coefficients!

$$\Delta O = O^{\text{exp}} - O^{\text{SM}} = \sum_i \frac{c_i A_i}{\Lambda^2} + \mathcal{O}\left(\frac{1}{\Lambda^3}\right)$$

- For BSM searches we are interested in measuring the c_i coefficients as best as we can:
    - We need a global dataset.
    - We need precise and accurate experimental measurements and theory predictions.

But here is the problem …

PDF-BSM interplay

The problem is that PDF and BSM worlds do not usually talk to each other.

θ : PDF parameters

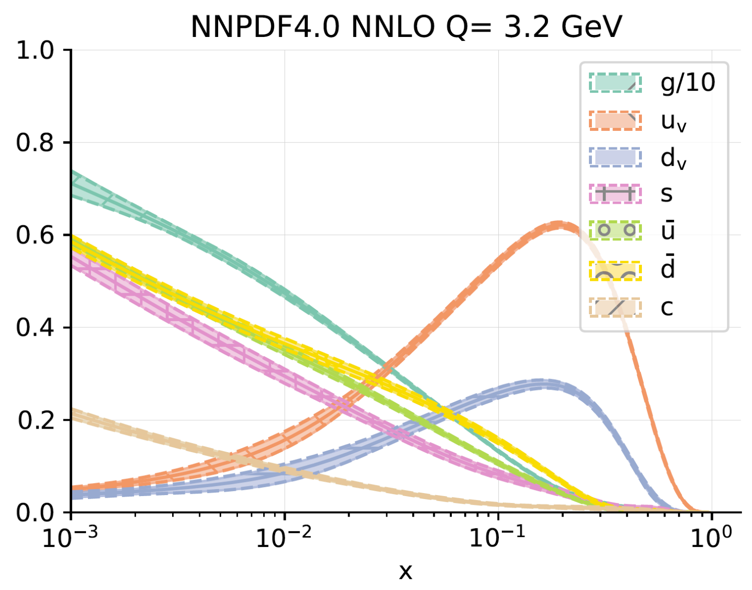
c : BSM parameters

PDF fits	EFT fits
BSM coefficient are kept fixed	PDF coefficients are kept fixed
$c = \bar{c}$	$\theta = \bar{\theta}$
$\sigma(\theta, \bar{c}) = \mathrm{PDF}(\theta) \otimes \hat{\sigma}(\bar{c})$	$\sigma(\bar{\theta}, c) = \mathrm{PDF}(\bar{\theta}) \otimes \hat{\sigma}(c)$

Need a **simultaneous** determination!

Outline

Some high energy physics



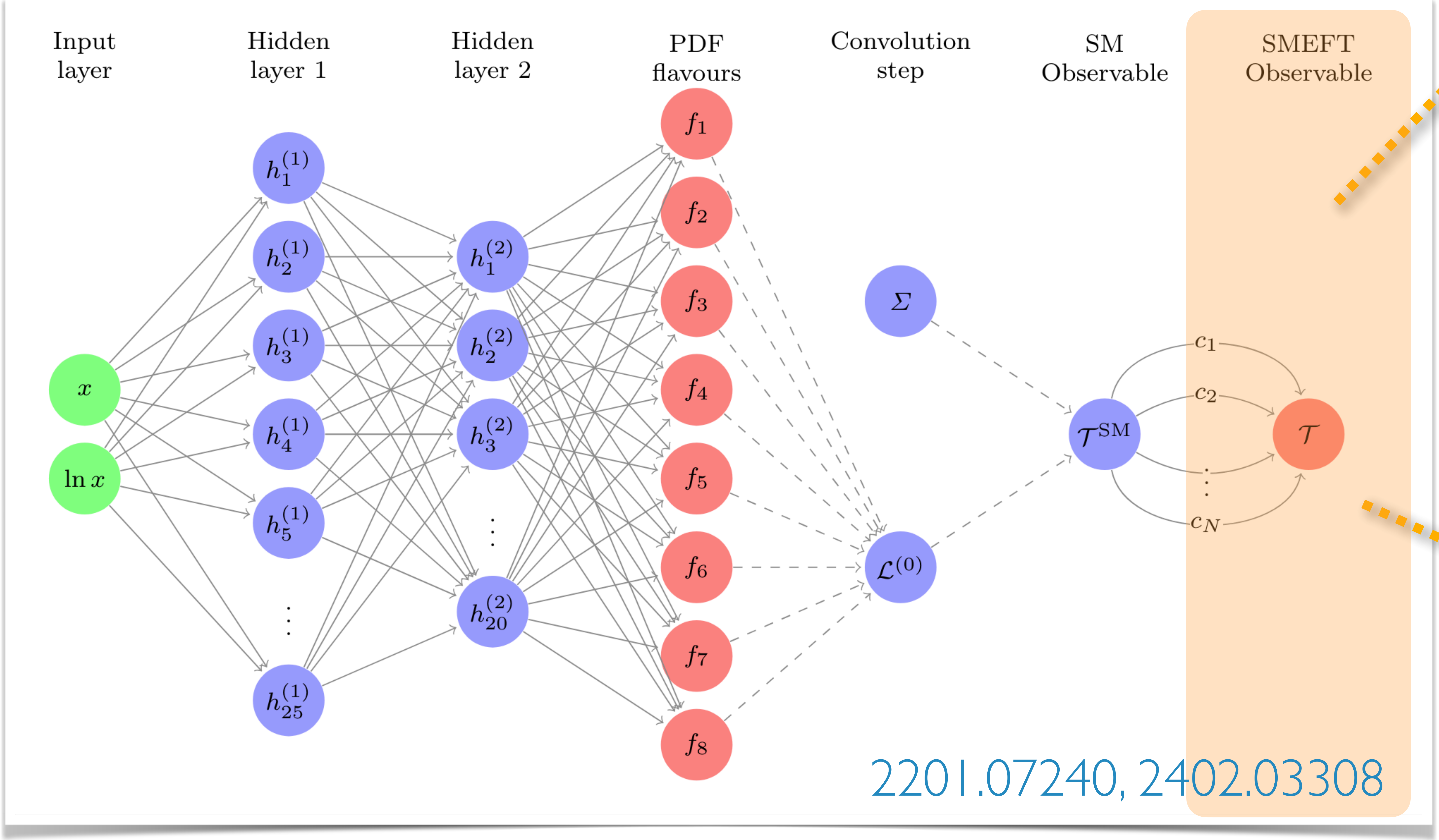
$$+ \quad \frac{c_i}{\Lambda^2}$$

Precision: deep learning the structure of hadrons and new physics

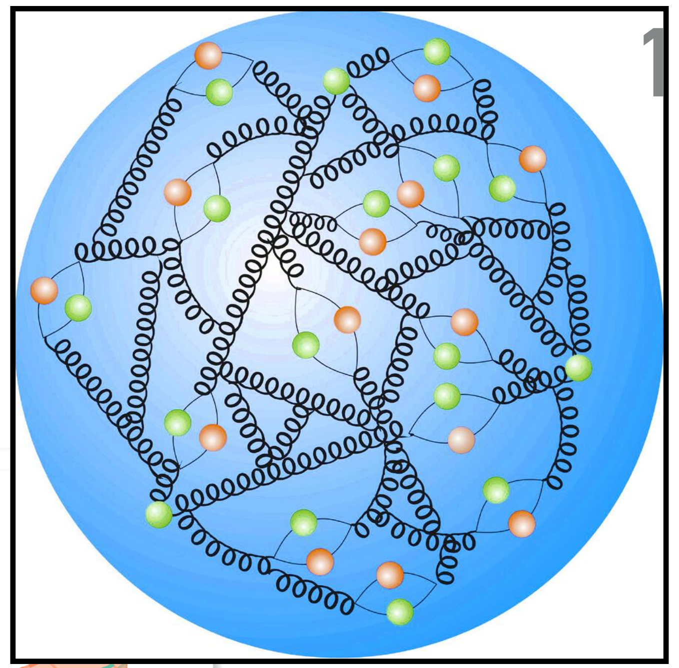
Symbolic regression

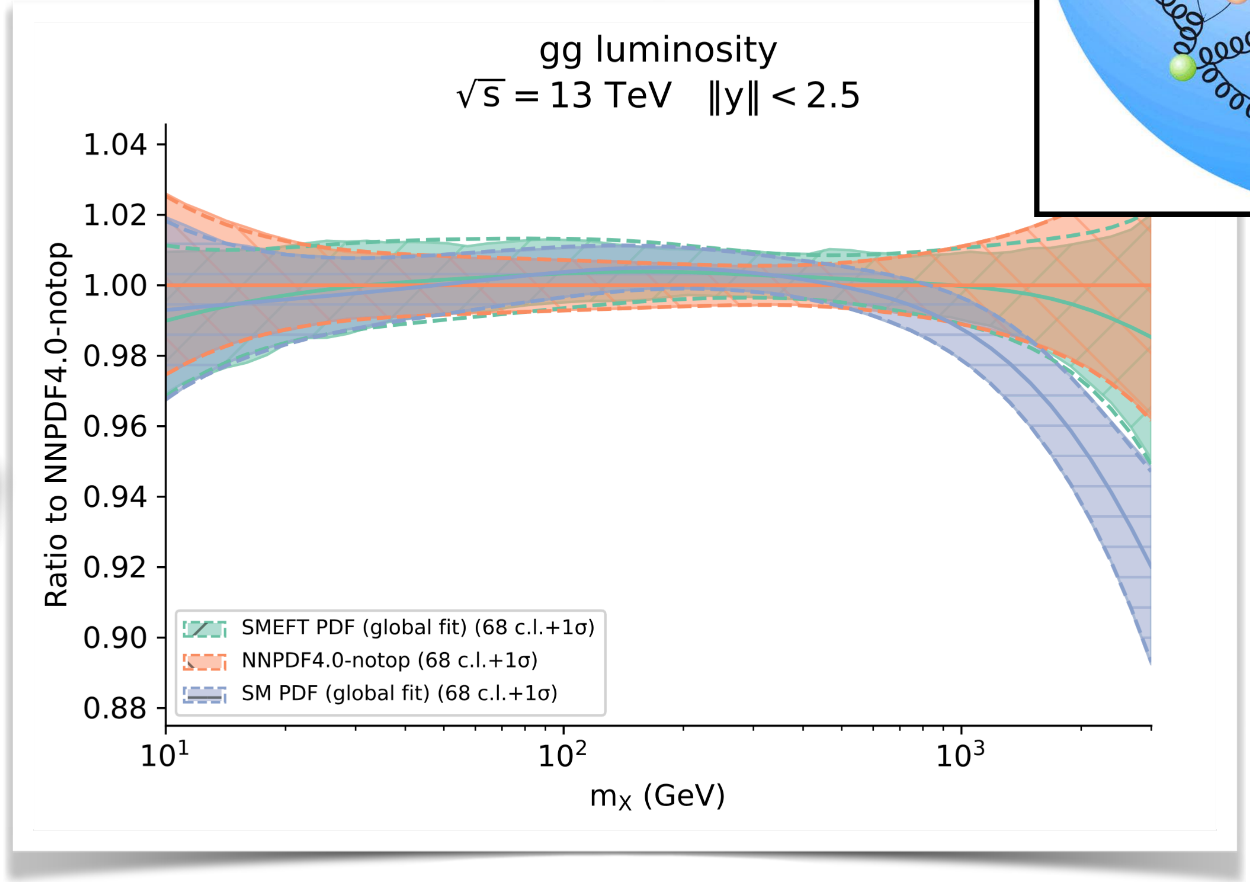
SIMUnet

We can explore the structure of the proton and potential signals of BSM physics *simultaneously* with deep learning

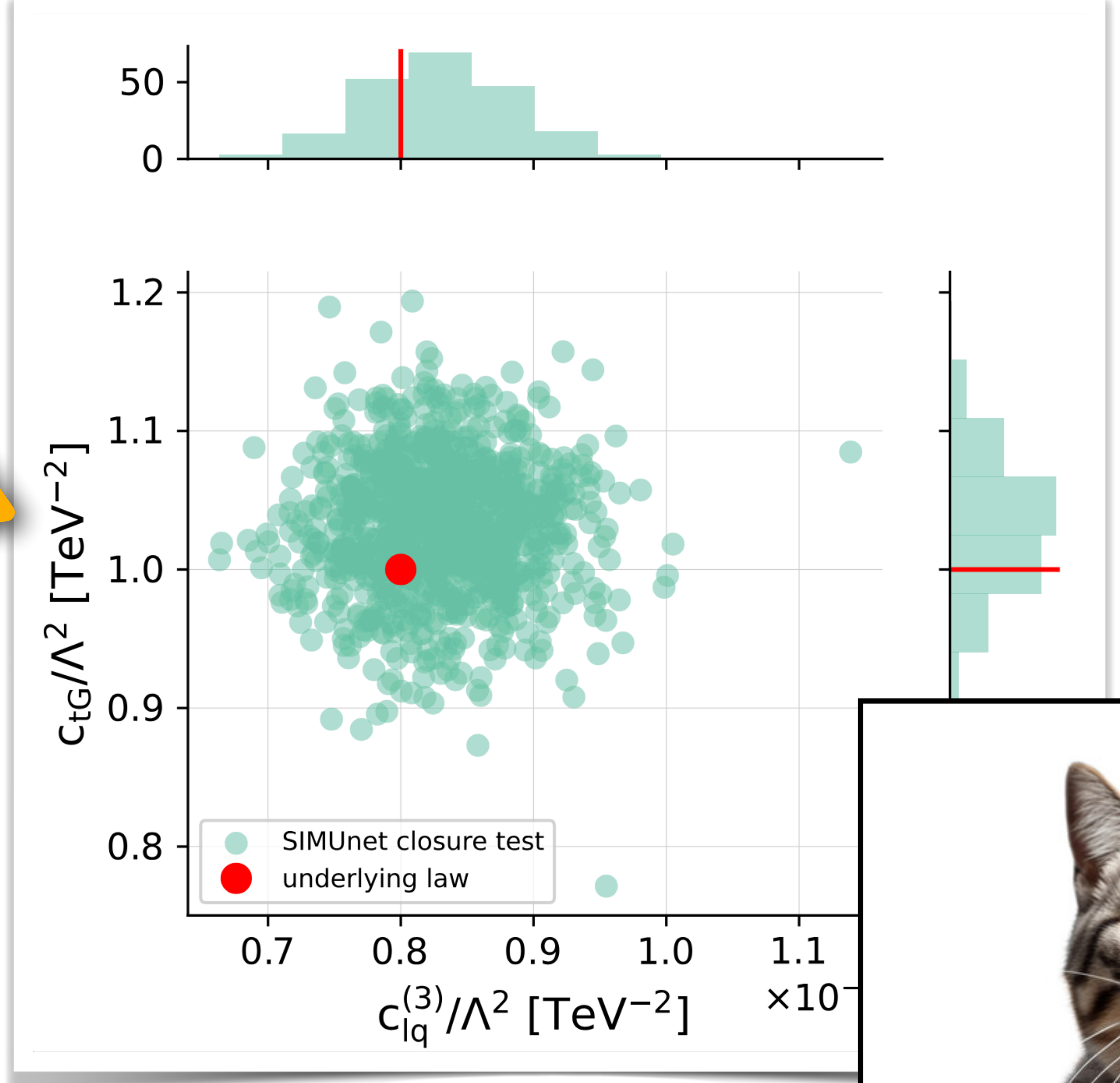


2201.07240, 2402.03308

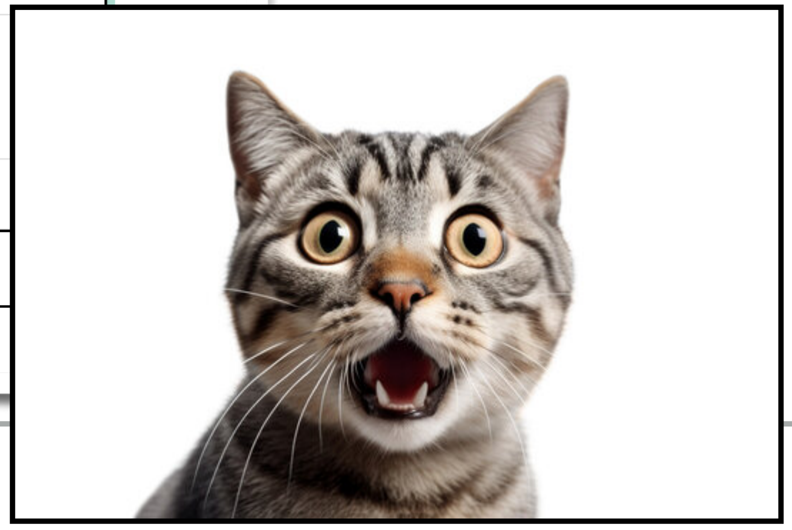




proton structure



new physics

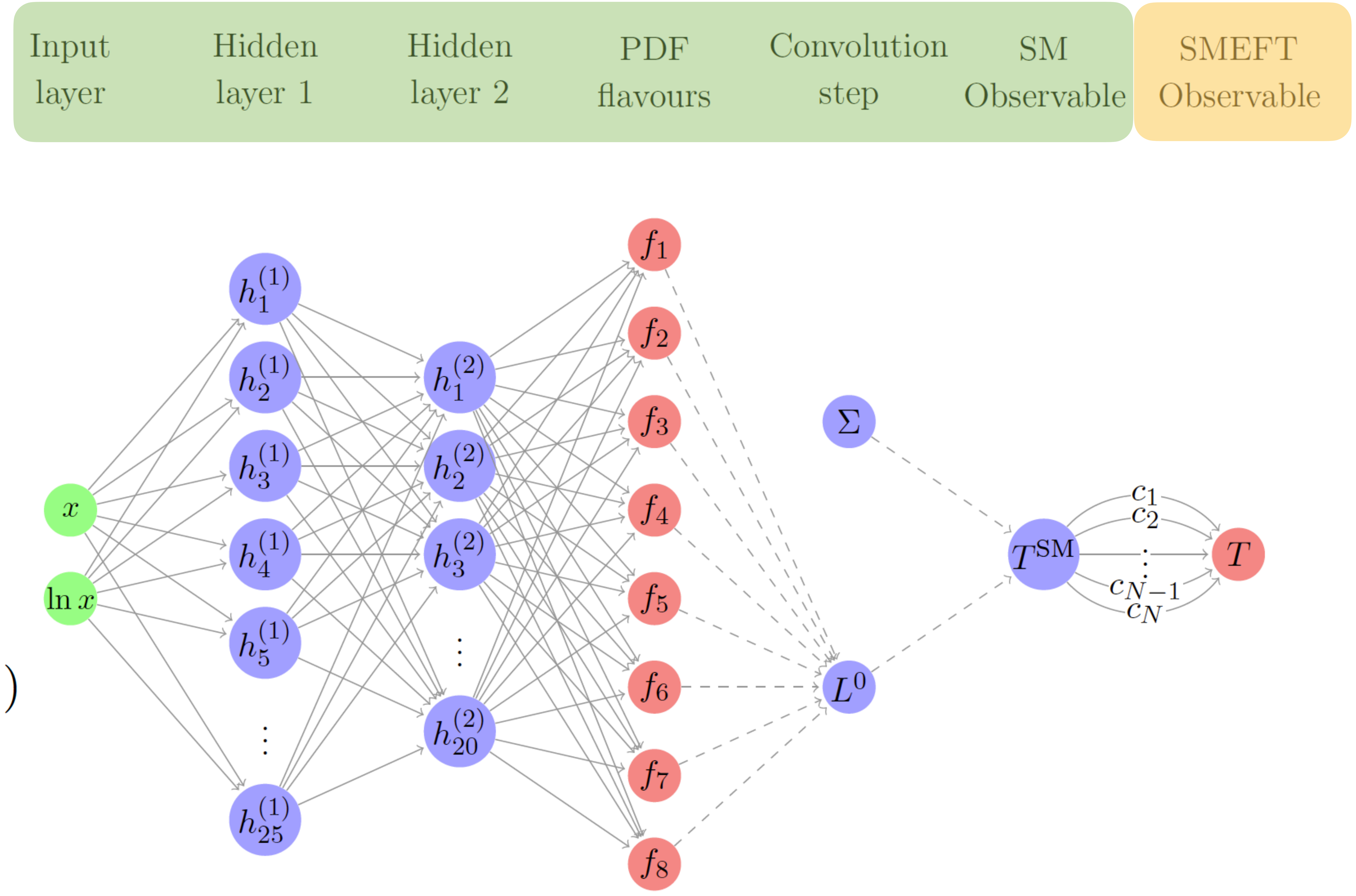


SIMUnet

- SIMUnet: NNPDF architecture supplemented with a SMEFT layer.
- PDF and BSM parameters are optimised simultaneously.

$$\chi^2(\hat{\theta}) = \frac{1}{N_{\rm dat}}(\mathbf{D} - \mathbf{T}(\hat{\theta}))^T(\mathbf{cov})^{-1}(\mathbf{D} - \mathbf{T}(\hat{\theta}))$$

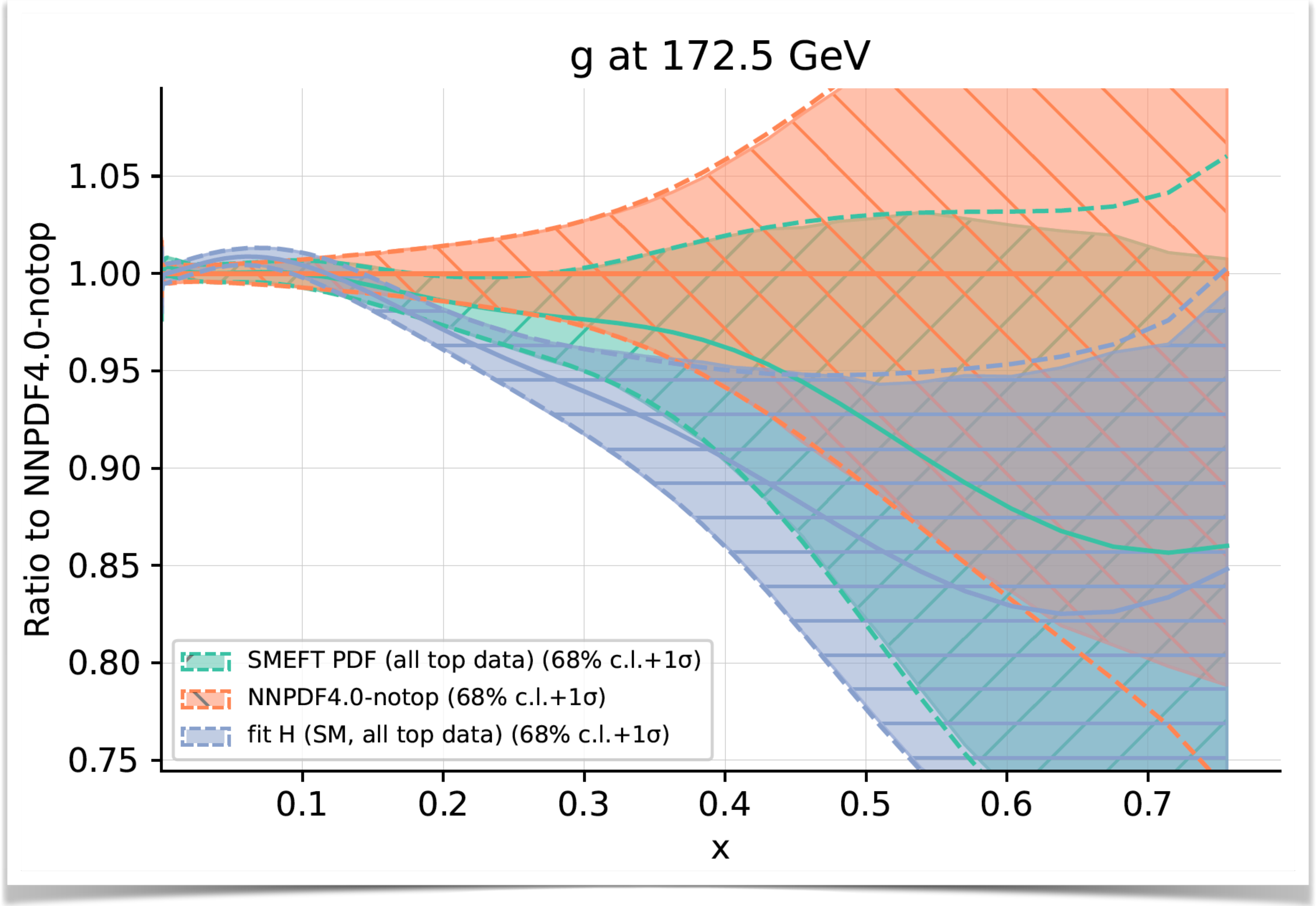
$$\hat{\theta} = \theta \cup \{c_i\}$$



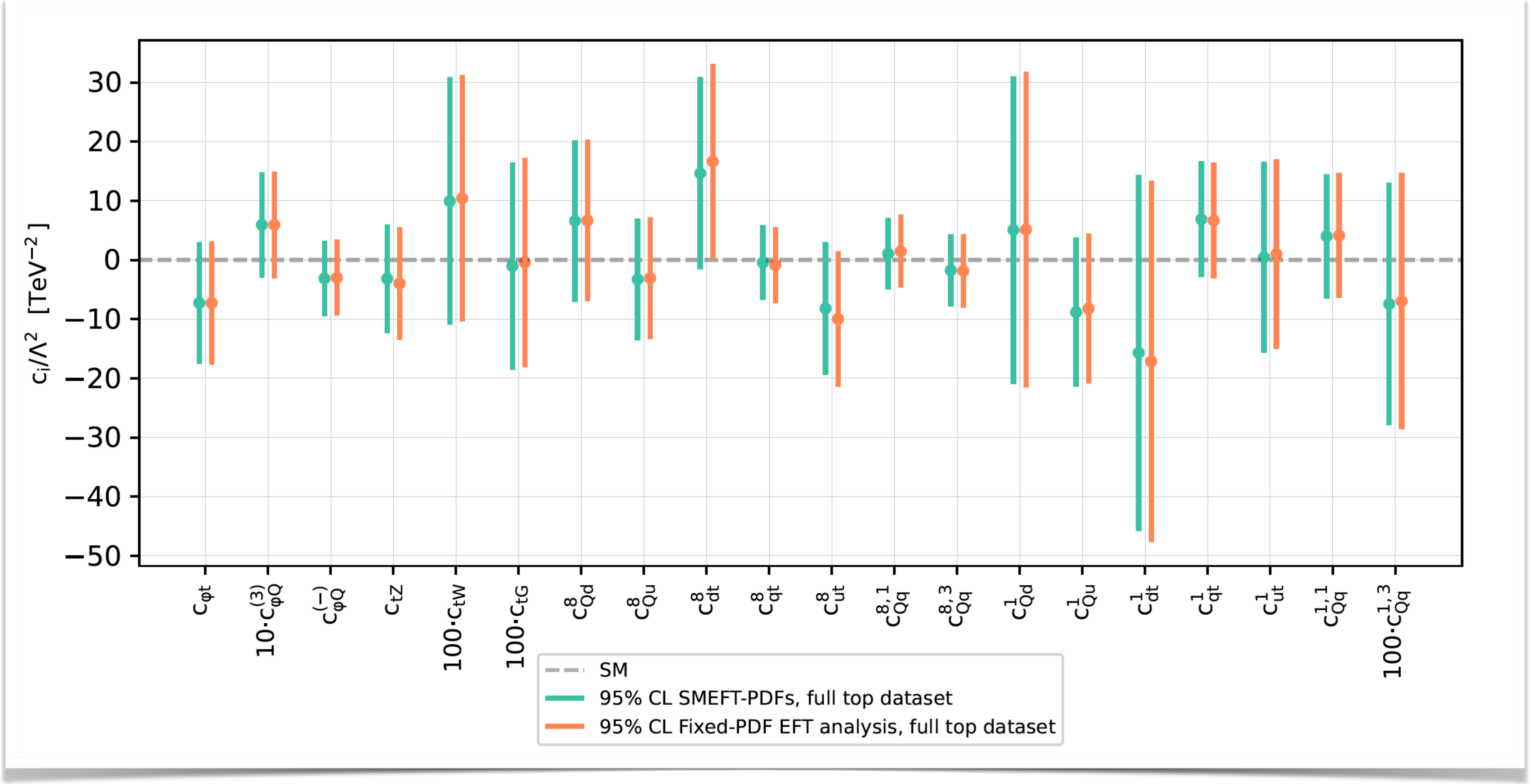
SIMUnet: simultaneous PDF + BSM coefficients

Here we perform a fit in the top-quark sector.

PDFs



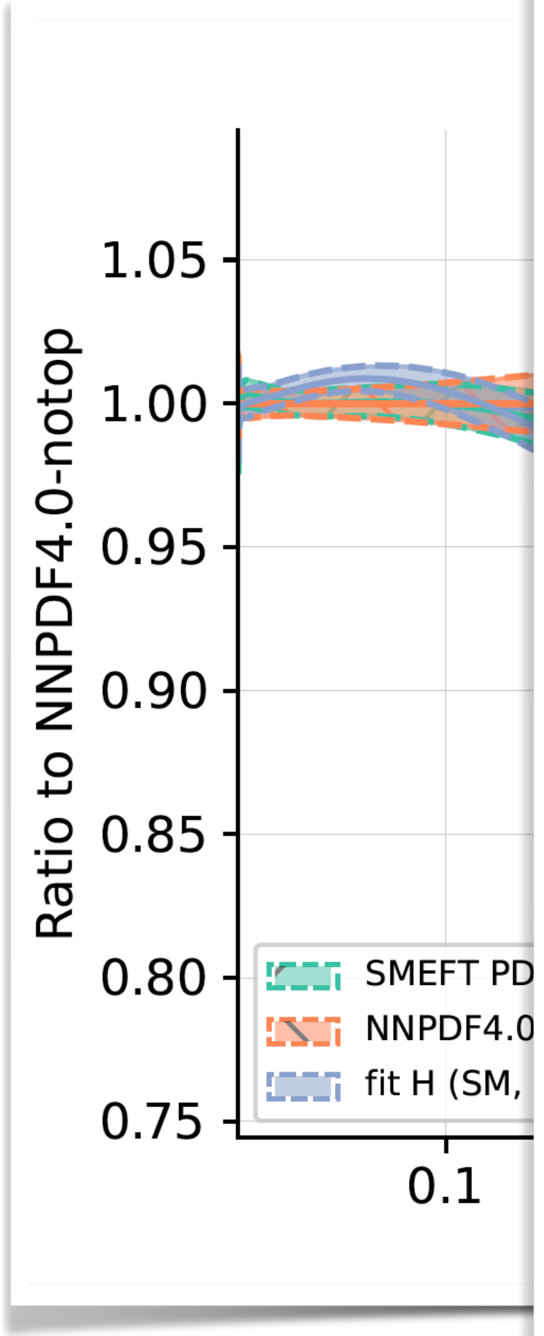
BSM



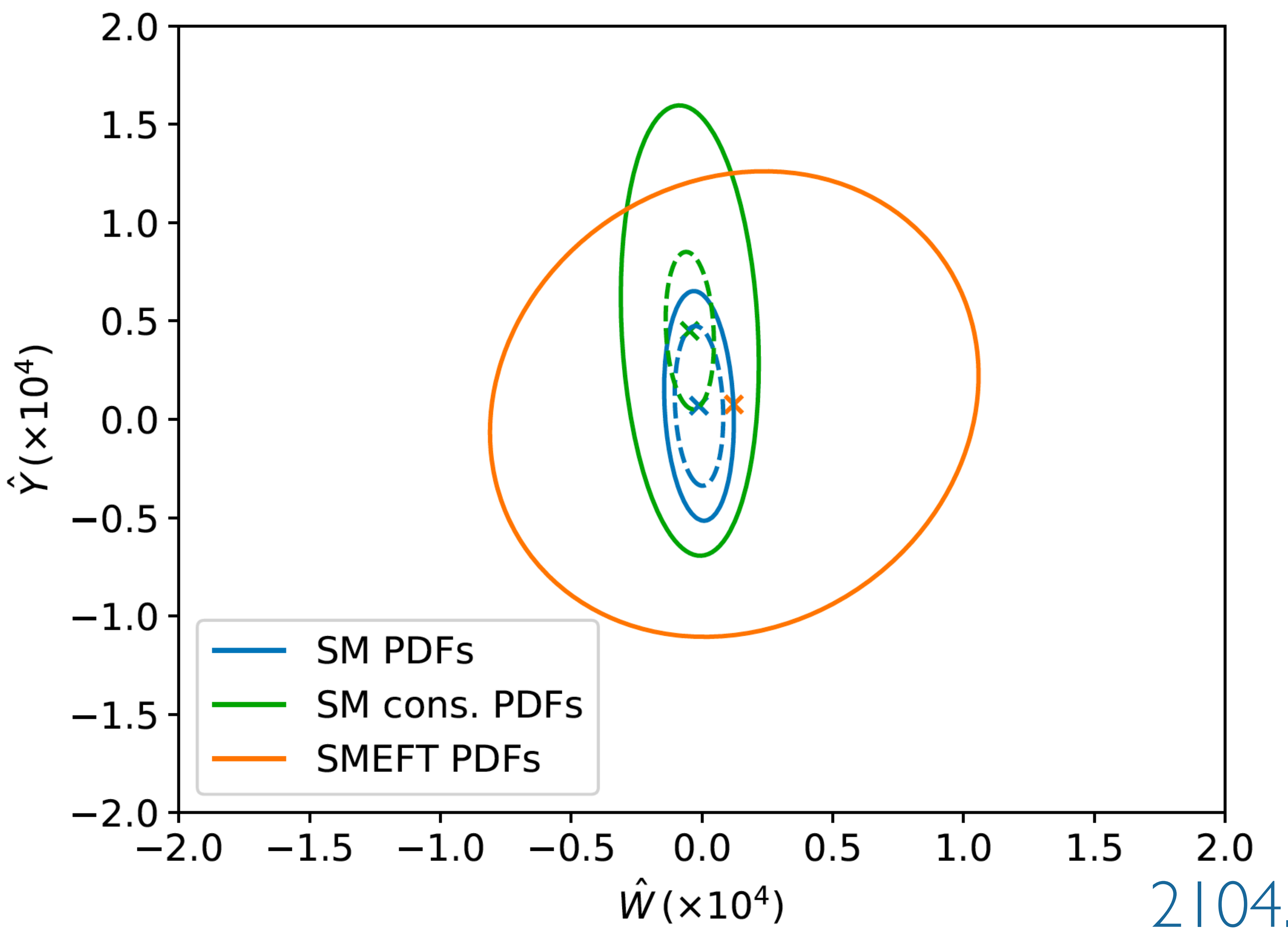
2303.06159

SIMUnet: simultaneous PDF + BSM coefficients

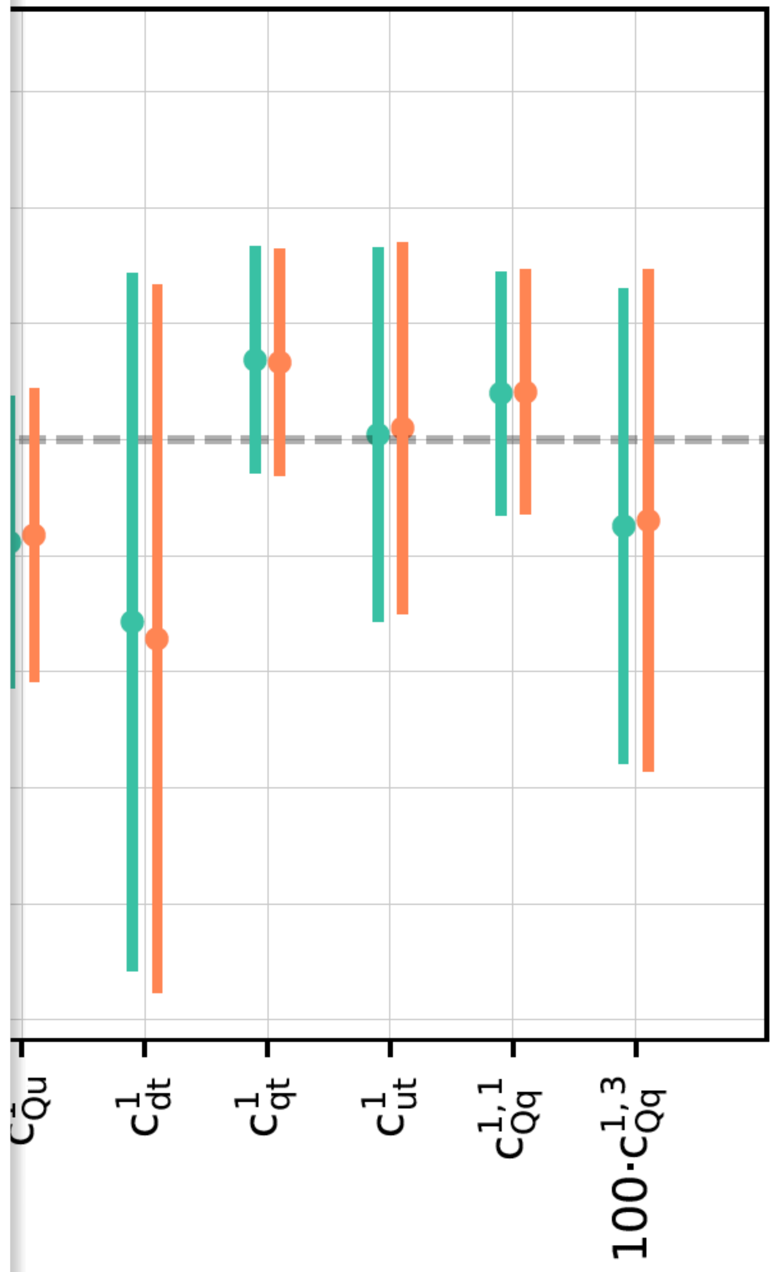
Here we



Neglecting the PDF-BSM interplay can lead to an overestimate of the constraints on BSM physics in other sectors.



2104.02723



2303.06159

SIMUnet: BSM physics absorption

Now, we move on to a different potential problem.

- Both PDFs and potential BSM physics are extracted from fits to data.
- Can PDFs **absorb** potential BSM physics?
  - We will work in a setting where we know the law of Nature: we will generate pseudodata according to our model, perform a fit, and see what comes out.

Let us suppose that the law of Nature is given by the SM plus some new physics (NP) contributions

$$T \equiv T(\theta_{\rm SM}, \theta_{\rm NP})$$

We have

True value of the observable

$$T^* \equiv T(\theta^*_{\rm SM}, \theta^*_{\rm NP})$$

The observed data

$$D_0 = T^* + \eta \qquad \eta \sim \mathcal{N}(0, \Sigma)$$

Pseudodata generation

- We perform two kinds of fits:

Fit name	Nature	Fitted parameters
Baseline	Standard Model: $\theta^*_{\mathrm{NP}} \equiv 0$	Standard Model only: θ_{SM}
Contaminated	SM + new physics: $\theta^*_{\mathrm{NP}} \neq 0$	Standard Model only: θ_{SM}

- NP is injected into the pseudodata via K-factor

$$T \equiv \left(1 + cK_{\mathrm{lin}} + c^2 K_{\mathrm{quad}}\right) \hat{\sigma}^{\mathrm{SM}} \otimes \mathcal{L}$$

* NB: now we are not focusing on simultaneous fits: our goal is to assess the absorption of the NP effects in a purely SM fit

NP absorption

- How can we assess if PDFs have been contaminated by NP?
- If there is a considerable number of data that enter the fit that are not affected NP, and some that data that are affected by NP, they could appear inconsistent and poorly described in the global fit.
- PDFs have **absorbed** NP if the fit quality is good i.e., for each dataset,

$$n_\sigma = \frac{\chi^2 - 1}{\sigma_{\chi^2}} < 2$$

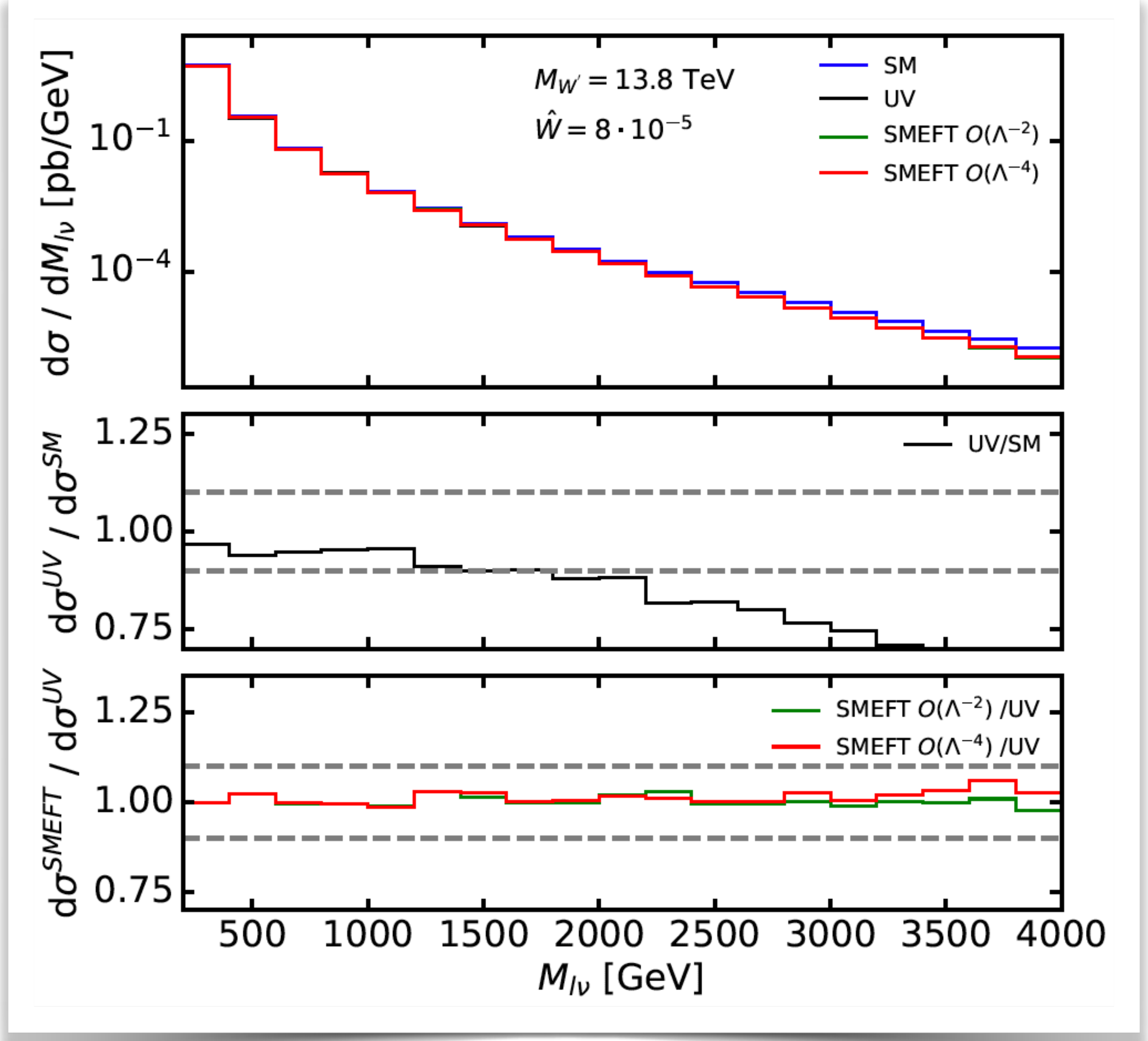
[2109.02653]

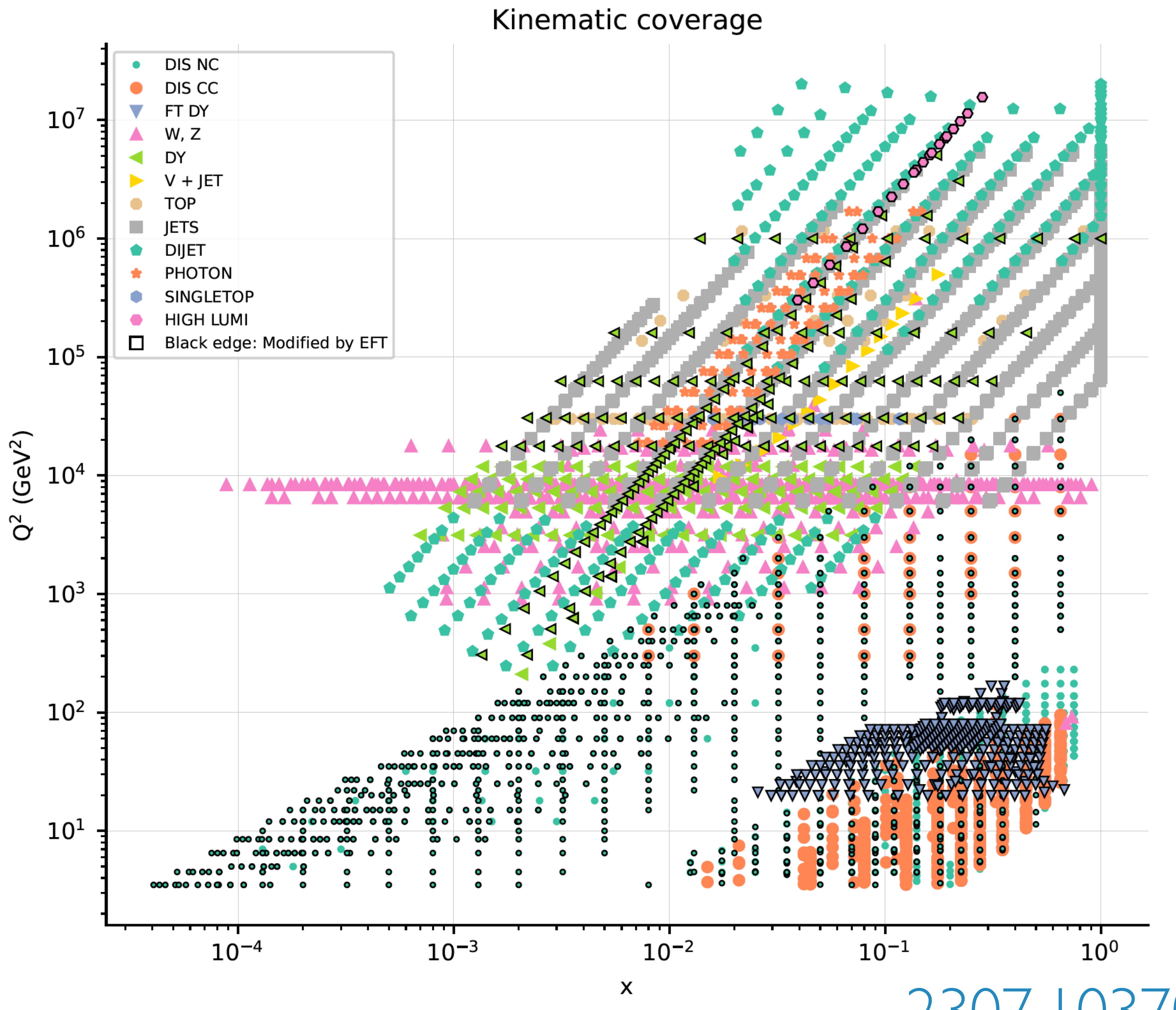
- Alternatively, we say that the PDFs would be "**contaminated**".

New physics model

- We work with a W' NP model and see its effects in LHC processes with respect to the SM.
- This NP model affects Drell-Yan and deep inelastic scattering datasets.

$$\mathcal{L}_{\rm SMEFT}^{W'} = \mathcal{L}_{\rm SM} - \frac{g^2 \hat{W}}{2m_W^2} J_L^{a,\mu} J_{L,\mu}^a$$

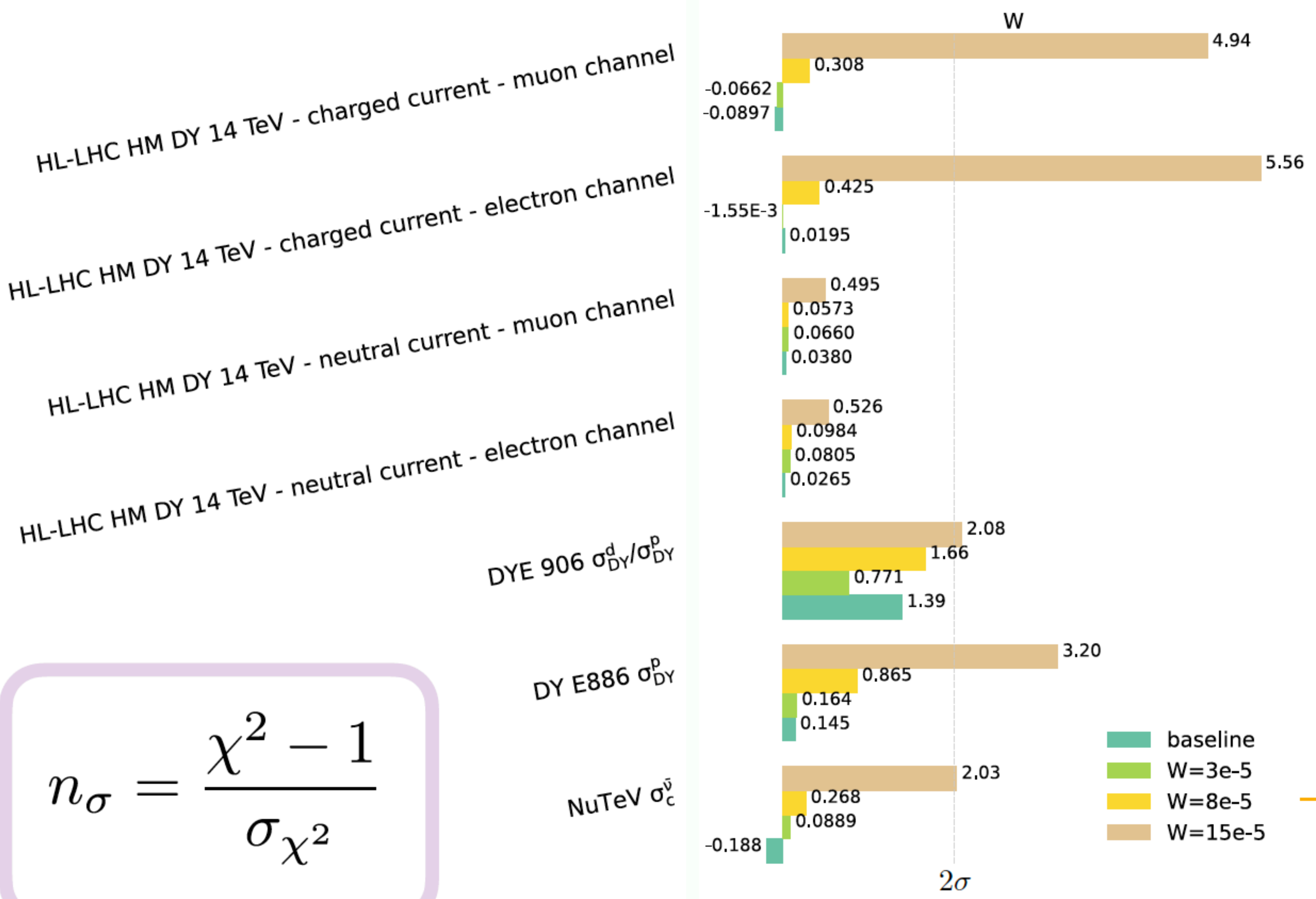




2307.10370

PDF fit quality summary

* Slide by M. Madigan!



$$n_\sigma = \frac{\chi^2 - 1}{\sigma_{\chi^2}}$$

Yes: PDFs absorb new physics

$$\hat{W} = 8 \cdot 10^{-5}, \quad M_{W'} \approx 14 \text{ TeV}$$

Effects of using a contaminated PDF set

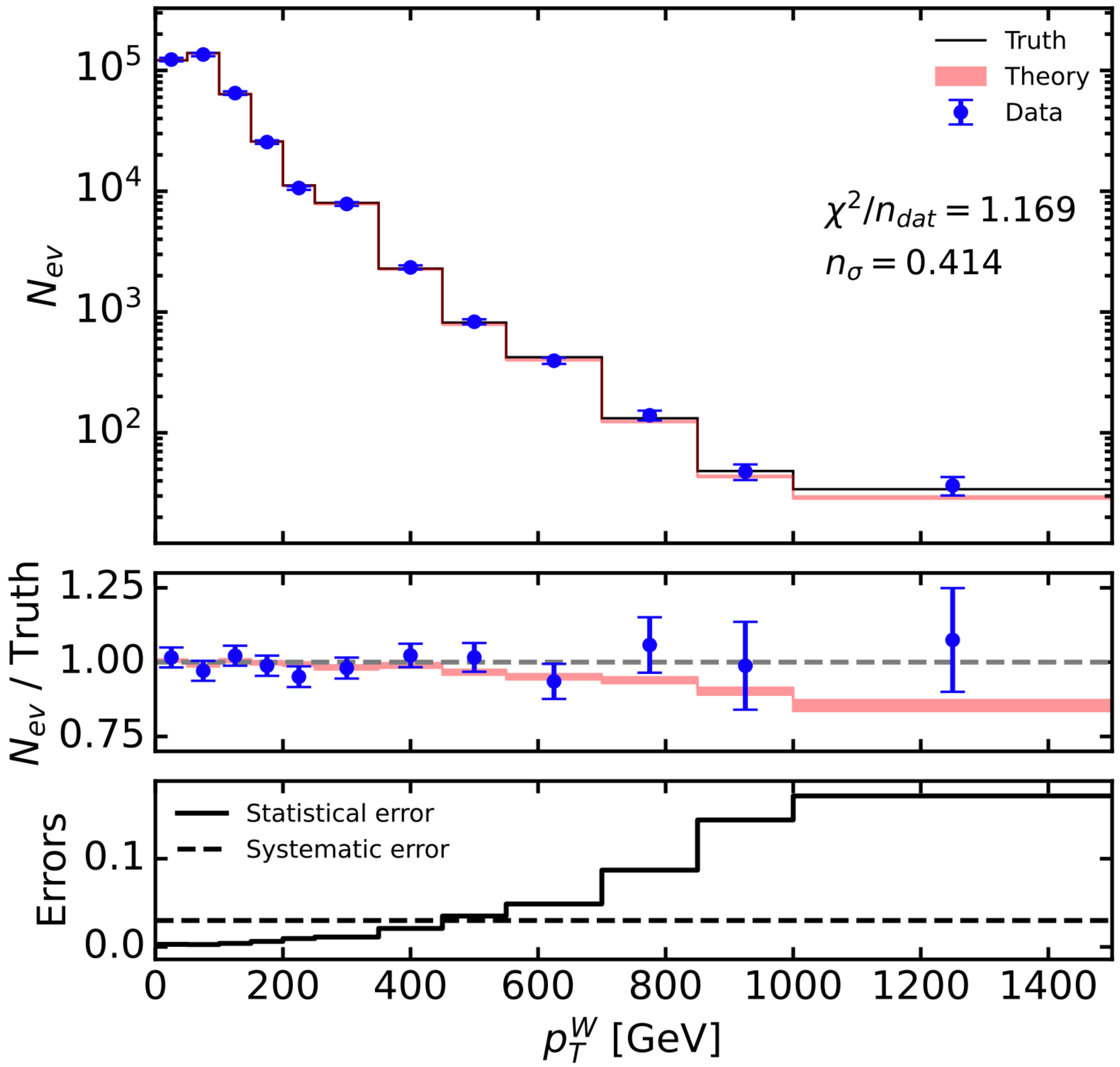
What happens is we unknowingly use contaminated PDFs?

- Consider the results of W+ H production (not affected by NP).
- There is an apparent but spurious tension with the SM.

Data: "True" PDF x SM

Theory: "Cont" PDF x SM

@ HL-LHC



SIMUnet: *summary* and comments

- A open-source deep learning methodology to study precision in the PDF-BSM interplay.
- It allows the simultaneous determination of PDFs and BSM coefficients.
- It allows for the assessment of NP absorption by the PDFs (which can also lead to apparent but spurious tensions).
- There are several directions in which to extend the analysis (quad. SMEFT, running effects, inclusion of new data in the analysis, etc).
- Recently, there has been increasing interest in comparing and stress testing different PDF fitting methodologies, comparing directly NN v/s fixed-form parametrisations, etc.

  Costantini et al., 2404.10056, MSHT, 2407.07944

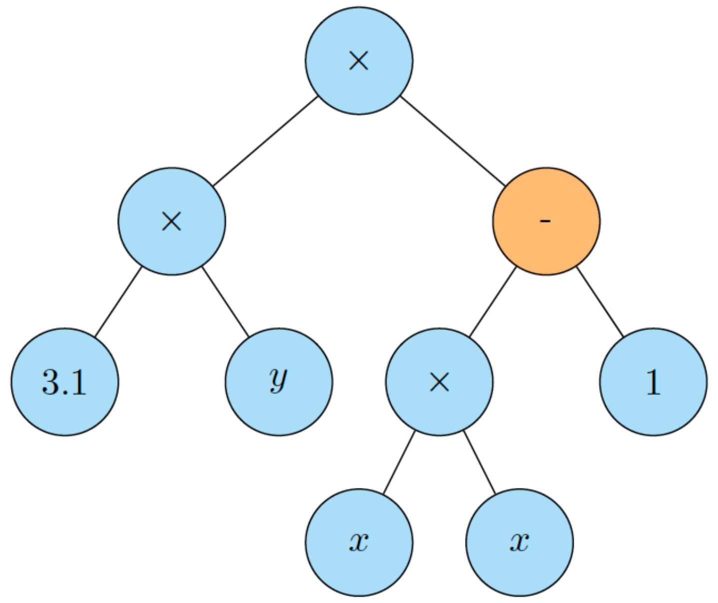
  - **Could ML techniques help in these cases?**

Outline

Some high energy physics

$+ \quad \frac{c_i}{\Lambda^2}$

Precision: deep learning the structure of hadrons and new physics



Symbolic regression

Symbolic regression

- Symbolic regression (SR) is a supervised learning technique.
- It seeks to find closed-form analytical expressions (equations) to map inputs to output without completely predefining the functional form of the mapping.
- It has been used in high energy physics in multiple contexts (e.g. 2109.10414, 2206.08901, 2211.08420).
- It can provide interpretable and simpler results (a domain scientist can find connections between a SR result and existing knowledge)

> Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte.
>
> If I had more time, I would have written a shorter letter.
>
> ---
>
> *from Lettres Provinciales,*
> *by B. Pascal.*

Symbolic regression

- We make use of the PySR library (2305.01582).
- Our aim is to find analytic formulas for quantities that do not have them yet. However, in precision physics it is *essential* that SR is able to recover what we know from first principles so:
  - We will study if well-known QED cross sections can be recovered.
  - We will study how the binning of an observable and different levels of noise affect the result.
  - We will discuss how SR can help with current and future challenges (angular coefficients).

Symbolic regression

- Formulas are described by expression trees (Fig. 1). They have a *complexity* c associated to their number of nodes.
- The fitness of a set (or many sets) of trees is evaluated.
- Mutations and crossover between trees happen through an evolutionary algorithm (Fig. 2).
- The cycle above is iterated.

$3.1y\,(x^2+1) \leftrightarrow$

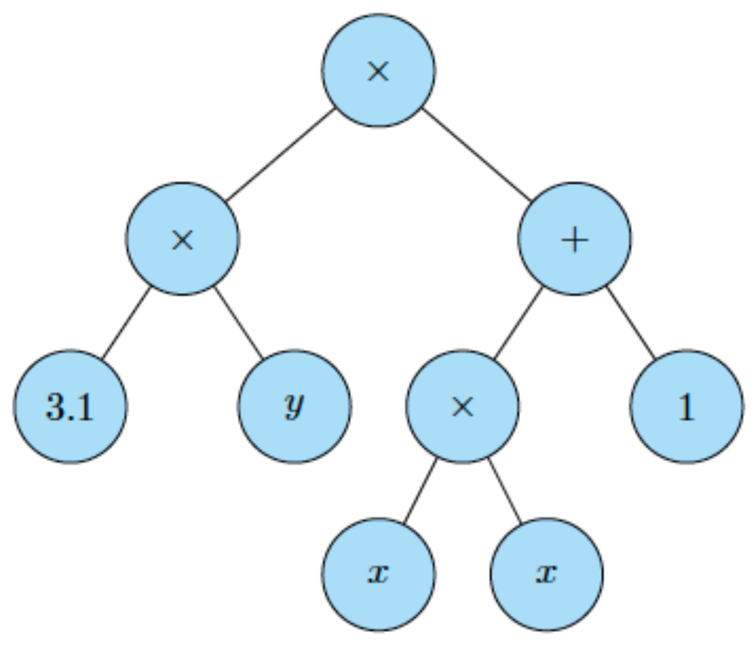


Fig. 1: Expression tree

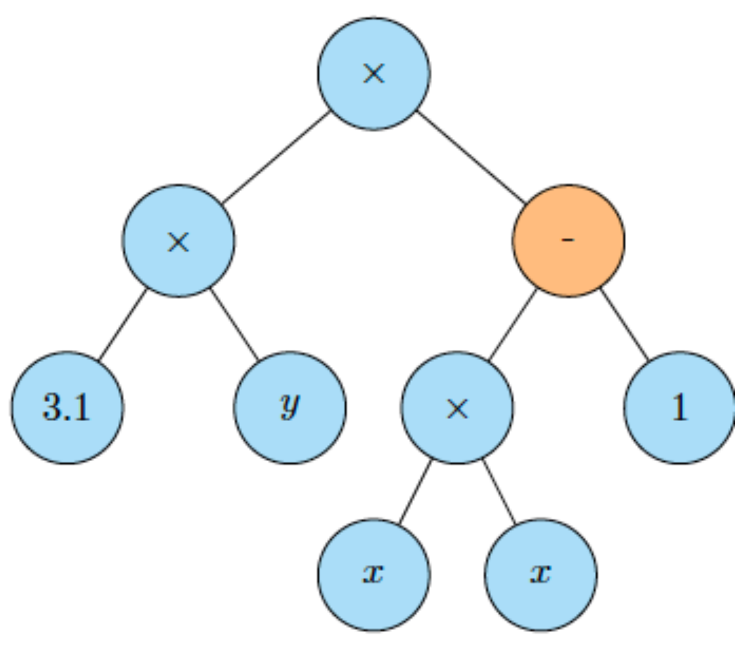


Fig. 2: Mutation

Selection criteria

- Trees can optimise different selection criteria. In PySR these are

Accuracy:

Minimise

$$L = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Score:

Maximise

$$-\frac{\partial \log(L)}{\partial c}$$

Best:

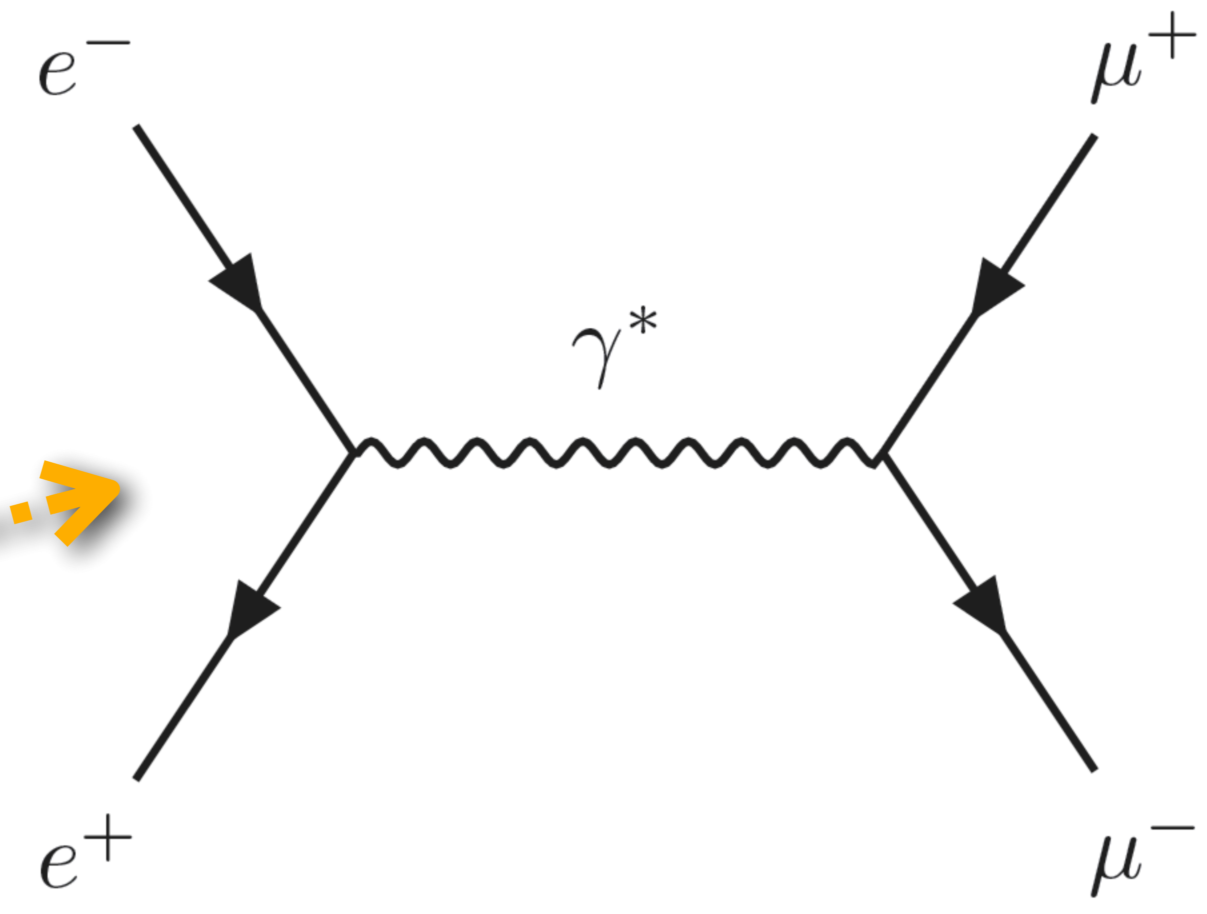
Highest score with

$$L \leq 1.5 \times L_{\min}$$

QED through SR

Can SR rediscover QED cross sections?

We simulate $e^- e^+ \to \mu^- \mu^+$



$$\frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha^2}{2s}(1+\cos^2\theta)$$

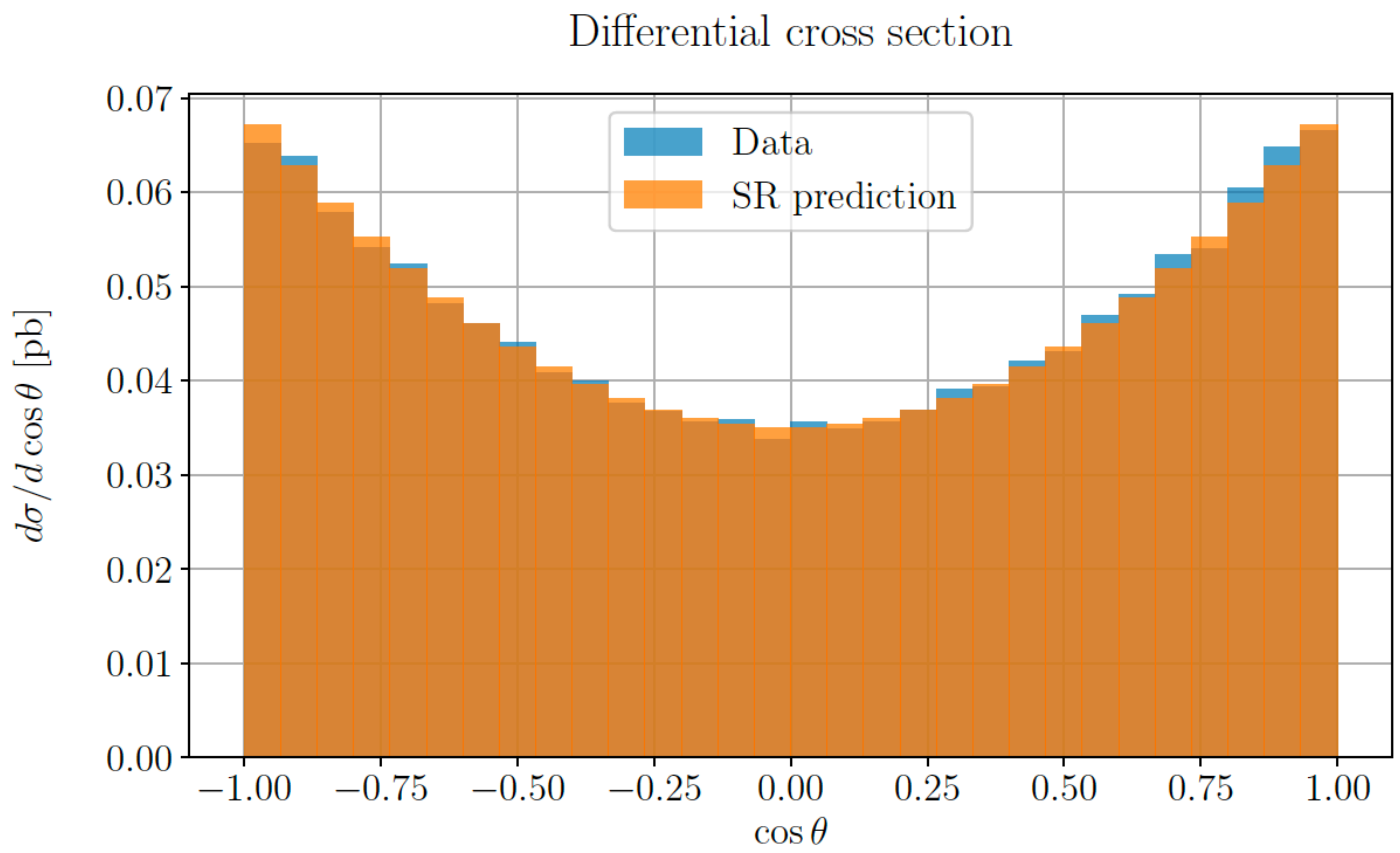


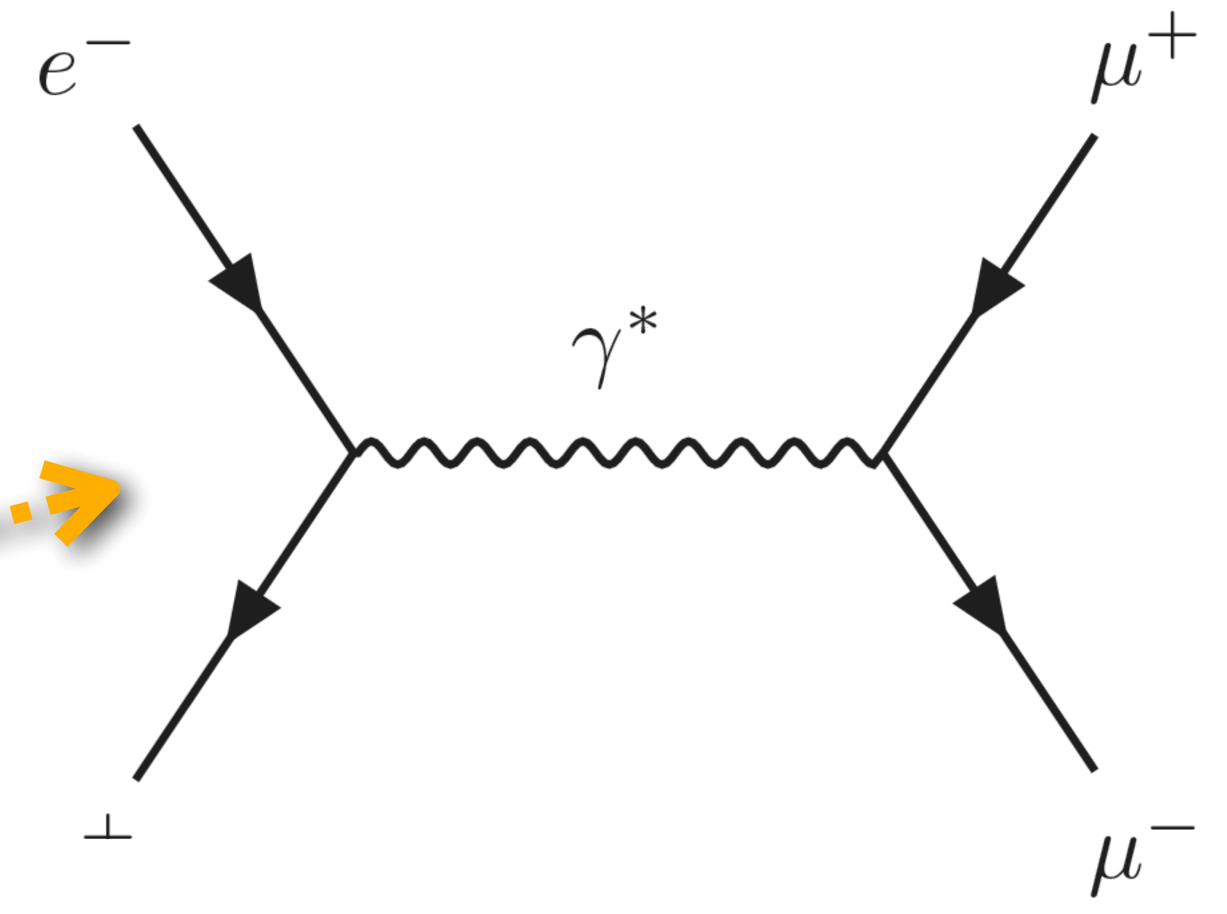
Table 1: Equations according to the three selection criteria for different bin sizes with $x_0 \equiv \cos\theta$. The numbers that appear in these expressions have been approximated to the 5th decimal place.

Bins	Accuracy	Best	Score
10	$x_0(x_0 + 0.00798)(0.00111 \cdot x_0^3 + 0.03459) + 0.03503$	$x_0^2(0.00111 \cdot x_0 + 0.03459) + 0.03503$	$0.03459 \cdot x_0^2 + 0.03503$
20	$x_0(x_0 + 0.01825)(-0.00155 \cdot x_0(x_0 - 0.05138) + 0.03579) + 0.03485$	$x_0(0.03447 \cdot x_0 + 0.00064) + 0.03498$	$0.03447 \cdot x_0^2 + 0.03498$
200	$x_0^2(-0.64647 \cdot x_0(0.00119 \cdot x_0 - 0.00151) + 0.03495) + 0.03495$	$0.03447 \cdot x_0^2 + 0.03495$	$0.03447 \cdot x_0^2 + 0.03495$
1000	$0.03447 \cdot x_0(0.01821 \cdot x_0^2 + x_0 + 0.00723) + 0.03495$	$0.03447 \cdot x_0^2 + 0.03495$	$0.03447 \cdot x_0^2 + 0.03495$

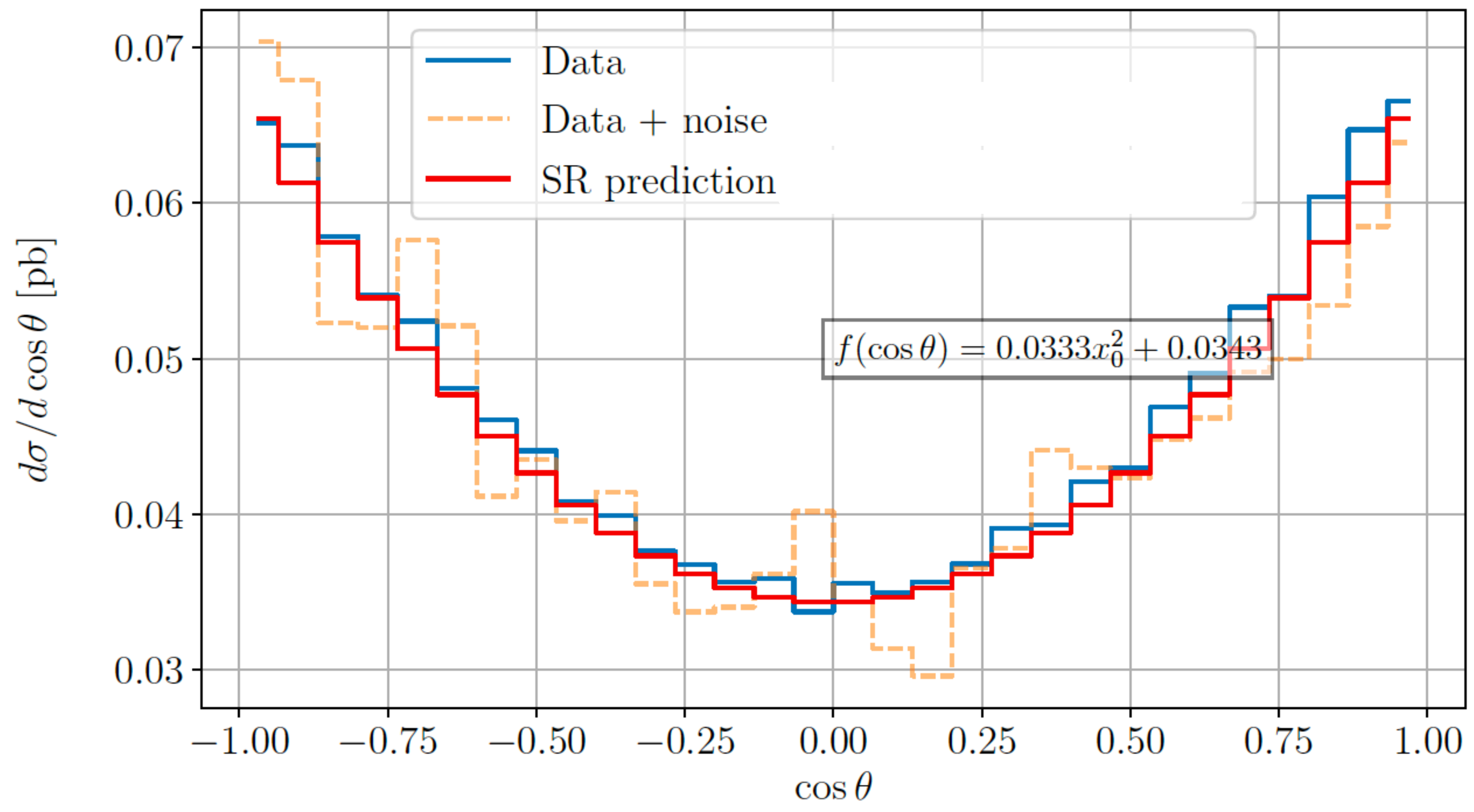
QED through SR

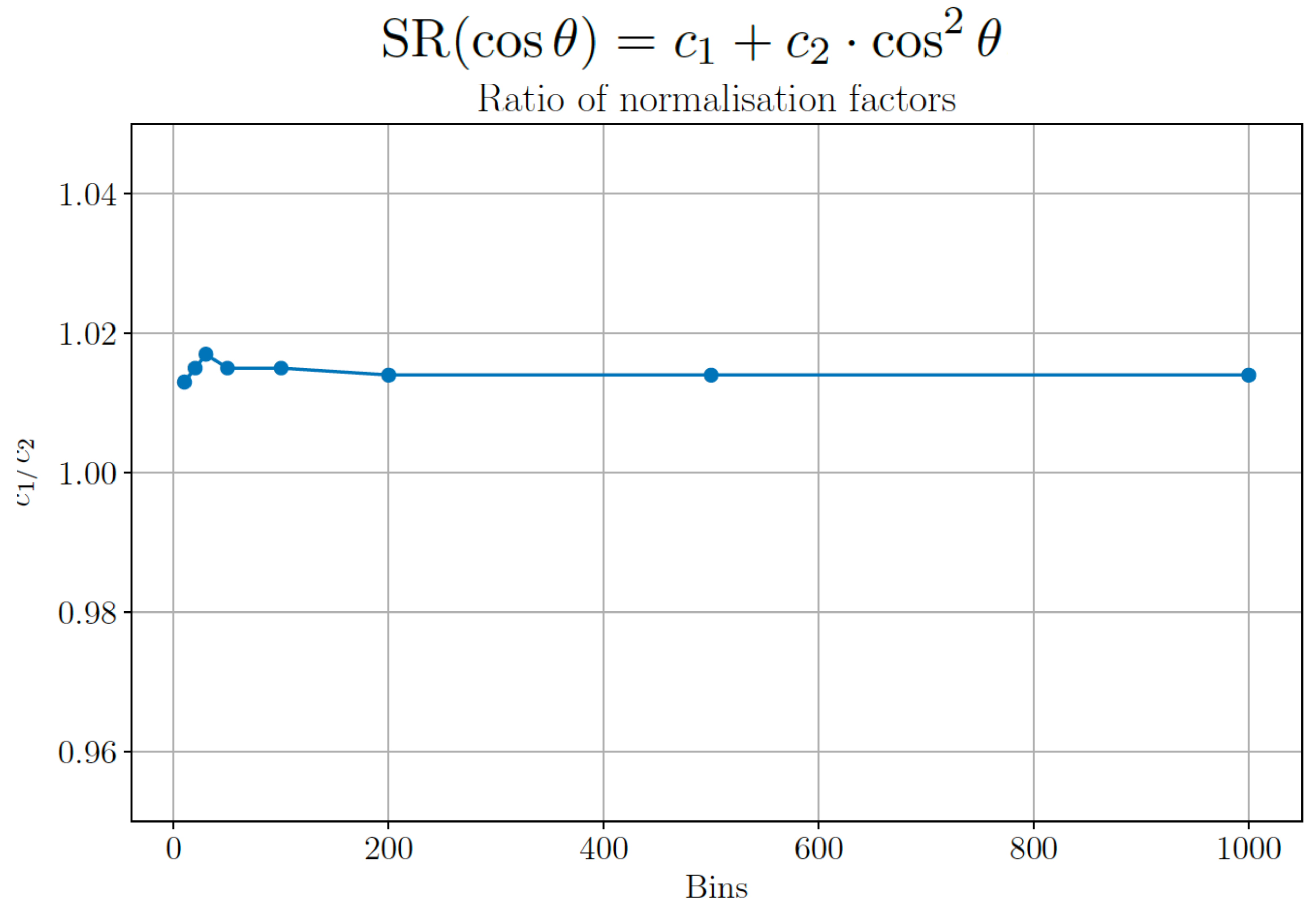
Can SR rediscover QED cross sections?

We simulate $e^-e^+ \to \mu^-\mu^+$



$$\frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha^2}{2s}(1+\cos^2\theta)$$





Angular coefficients through SR

SR can also be applied to more complex process (W/Z production at NLO/NNLO):

$$\frac{d^5\sigma}{dp_T\, dy_{ll}\, dm_{ll}\, d\cos\theta\, d\phi} = \frac{3}{16\pi}\frac{d^3\sigma^{U+L}}{dp_T\, dy_{ll}\, dm_{ll}}\left[(1+\cos^2\theta) + \sum_{i=0}^{7} P_i(\theta,\phi)\, A_i\right],$$

- P_i are spherical harmonics (they encode the angular dependence).
- A_i are *angular coefficients* (they encode the dependence on the other kinematics) for which there is no analytical formula.
- Finding a formula for these coefficients would be useful for experiment and theory.

Angular coefficients through SR

On our way to the full set of angular coefficients, there are intermediate scenarios that have to be studied.

$$\frac{d^5\sigma}{dp_T\, dy_{ll}\, dm_{ll}\, d\cos\theta\, d\phi} = \frac{3}{16\pi}\frac{d^3\sigma^{U+L}}{dp_T\, dy_{ll}\, dm_{ll}}\left[(1+\cos^2\theta) + \sum_{i=0}^{7} P_i(\theta,\phi)\, A_i\right],$$

E.g. leading order Drell-Yan via Z boson (only the A_4 coefficient is non-zero).

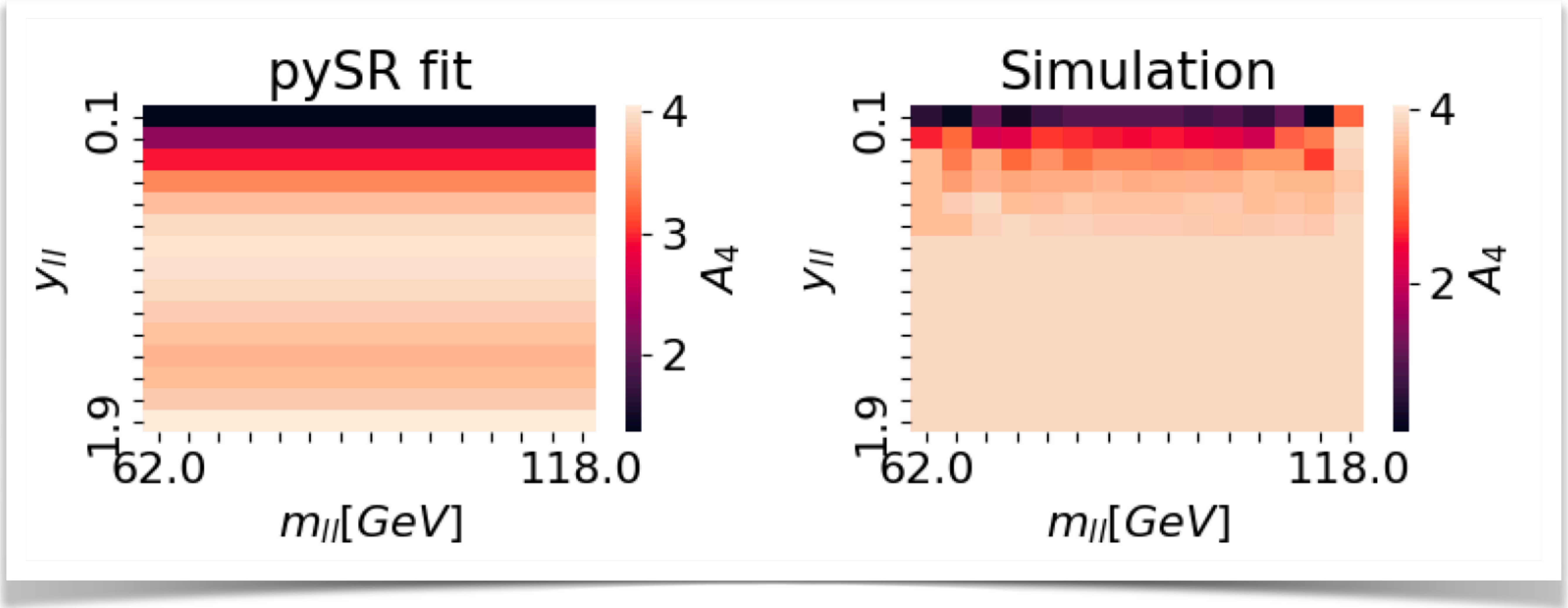


Table 2: Complexity and equations for the A_4 coefficient.

Loss	Complexity	Expression
0.154545	9	$-1.5356994\lvert y_{ll}\rvert^2 + 3.9449594\lvert y_{ll}\rvert + 1.613458$
0.092542	11	$\lvert y_{ll}\rvert^3 - 4.4877133\lvert y_{ll}\rvert^2 + 6.2672553\lvert y_{ll}\rvert + 1.2202095$
0.073751	15	$2\lvert y_{ll}\rvert^3 - 7.4880592\lvert y_{ll}\rvert^2 + 8.6747910371\lvert y_{ll}\rvert + 0.8127258$

PDFs through SR

The scaling in *x* of the PDFs is dictated by *Regge theory*:

$$\lim_{x\to 1} f(x,Q^2) = 0 \qquad\qquad \lim_{x\to 0} f(x,Q^2) \propto x^{\alpha}$$

We can integrate these requirements in the parametrisations we discussed:

$$f(x,Q^2) = x^{\alpha}(1-x)^{\beta}\mathrm{NN}(x,Q^2)$$

$$f(x,Q^2) = x^{\alpha}(1-x)^{\beta} f_{\mathrm{red}}(x,Q^2)$$

Could SR be of help?

$$f(x,Q^2) = x^{\alpha}(1-x)^{\beta}\mathrm{SR}(x,Q^2)$$

Corrections from Regge theory?

NN/analytical comparison?

Simpler functional forms with flexibility?

What about *uncertainty*?

Conclusion

- Precision interpretations and calculations are playing an important role in the search for BSM physics.
- Deep learning is allowing us to explore the interplay between the PDFs (that parametrise the structure of the proton) and BSM physics.
- SR can bring new insights to our studies. We find pure *accuracy* not be a good selection criterion, model complexity is also relevant.
- A careful assessment of the robustness and stability of the SR results is necessary in the era of precision physics.

Thank you for your attention!

Backup slides

Warsaw basis

In the SMEFT, we supplement the SM Lagrangian with towers of higher dimensional operators

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{c_i}{\Lambda^2} Q_i^{(6)} + \cdots$$

Up to dimension 6, these operators can be parametrised in the Warsaw basis

X^3		φ^6 and $\varphi^4 D^2$		$\psi^2\varphi^3$	
Q_G	$f^{ABC} G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	Q_φ	$(\varphi^\dagger\varphi)^3$	$Q_{e\varphi}$	$(\varphi^\dagger\varphi)(\bar l_p e_r \varphi)$
$Q_{\widetilde G}$	$f^{ABC} \widetilde G_\mu^{A\nu} G_\nu^{B\rho} G_\rho^{C\mu}$	$Q_{\varphi\Box}$	$(\varphi^\dagger\varphi)\Box(\varphi^\dagger\varphi)$	$Q_{u\varphi}$	$(\varphi^\dagger\varphi)(\bar q_p u_r \widetilde\varphi)$
Q_W	$\varepsilon^{IJK} W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$	$Q_{\varphi D}$	$\left(\varphi^\dagger D^\mu\varphi\right)^\star \left(\varphi^\dagger D_\mu\varphi\right)$	$Q_{d\varphi}$	$(\varphi^\dagger\varphi)(\bar q_p d_r \varphi)$
$Q_{\widetilde W}$	$\varepsilon^{IJK} \widetilde W_\mu^{I\nu} W_\nu^{J\rho} W_\rho^{K\mu}$				

$X^2\varphi^2$		$\psi^2 X\varphi$		$\psi^2\varphi^2 D$	
$Q_{\varphi G}$	$\varphi^\dagger\varphi\, G_{\mu\nu}^A G^{A\mu\nu}$	Q_{eW}	$(\bar l_p \sigma^{\mu\nu} e_r)\tau^I\varphi W_{\mu\nu}^I$	$Q_{\varphi l}^{(1)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar l_p \gamma^\mu l_r)$
$Q_{\varphi\widetilde G}$	$\varphi^\dagger\varphi\, \widetilde G_{\mu\nu}^A G^{A\mu\nu}$	Q_{eB}	$(\bar l_p \sigma^{\mu\nu} e_r)\varphi B_{\mu\nu}$	$Q_{\varphi l}^{(3)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu \varphi)(\bar l_p \tau^I\gamma^\mu l_r)$
$Q_{\varphi W}$	$\varphi^\dagger\varphi\, W_{\mu\nu}^I W^{I\mu\nu}$	Q_{uG}	$(\bar q_p \sigma^{\mu\nu} T^A u_r)\widetilde\varphi\, G_{\mu\nu}^A$	$Q_{\varphi e}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar e_p \gamma^\mu e_r)$
$Q_{\varphi\widetilde W}$	$\varphi^\dagger\varphi\, \widetilde W_{\mu\nu}^I W^{I\mu\nu}$	Q_{uW}	$(\bar q_p \sigma^{\mu\nu} u_r)\tau^I\widetilde\varphi\, W_{\mu\nu}^I$	$Q_{\varphi q}^{(1)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar q_p \gamma^\mu q_r)$
$Q_{\varphi B}$	$\varphi^\dagger\varphi\, B_{\mu\nu} B^{\mu\nu}$	Q_{uB}	$(\bar q_p \sigma^{\mu\nu} u_r)\widetilde\varphi\, B_{\mu\nu}$	$Q_{\varphi q}^{(3)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu \varphi)(\bar q_p \tau^I\gamma^\mu q_r)$
$Q_{\varphi\widetilde B}$	$\varphi^\dagger\varphi\, \widetilde B_{\mu\nu} B^{\mu\nu}$	Q_{dG}	$(\bar q_p \sigma^{\mu\nu} T^A d_r)\varphi\, G_{\mu\nu}^A$	$Q_{\varphi u}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar u_p \gamma^\mu u_r)$
$Q_{\varphi WB}$	$\varphi^\dagger\tau^I\varphi\, W_{\mu\nu}^I B^{\mu\nu}$	Q_{dW}	$(\bar q_p \sigma^{\mu\nu} d_r)\tau^I\varphi\, W_{\mu\nu}^I$	$Q_{\varphi d}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu \varphi)(\bar d_p \gamma^\mu d_r)$
$Q_{\varphi\widetilde WB}$	$\varphi^\dagger\tau^I\varphi\, \widetilde W_{\mu\nu}^I B^{\mu\nu}$	Q_{dB}	$(\bar q_p \sigma^{\mu\nu} d_r)\varphi\, B_{\mu\nu}$	$Q_{\varphi ud}$	$i(\widetilde\varphi^\dagger D_\mu \varphi)(\bar u_p \gamma^\mu d_r)$

$(\bar LL)(\bar LL)$		$(\bar RR)(\bar RR)$		$(\bar LL)(\bar RR)$	
Q_{ll}	$(\bar l_p\gamma_\mu l_r)(\bar l_s\gamma^\mu l_t)$	Q_{ee}	$(\bar e_p\gamma_\mu e_r)(\bar e_s\gamma^\mu e_t)$	Q_{le}	$(\bar l_p\gamma_\mu l_r)(\bar e_s\gamma^\mu e_t)$
$Q_{qq}^{(1)}$	$(\bar q_p\gamma_\mu q_r)(\bar q_s\gamma^\mu q_t)$	Q_{uu}	$(\bar u_p\gamma_\mu u_r)(\bar u_s\gamma^\mu u_t)$	Q_{lu}	$(\bar l_p\gamma_\mu l_r)(\bar u_s\gamma^\mu u_t)$
$Q_{qq}^{(3)}$	$(\bar q_p\gamma_\mu \tau^I q_r)(\bar q_s\gamma^\mu \tau^I q_t)$	Q_{dd}	$(\bar d_p\gamma_\mu d_r)(\bar d_s\gamma^\mu d_t)$	Q_{ld}	$(\bar l_p\gamma_\mu l_r)(\bar d_s\gamma^\mu d_t)$
$Q_{lq}^{(1)}$	$(\bar l_p\gamma_\mu l_r)(\bar q_s\gamma^\mu q_t)$	Q_{eu}	$(\bar e_p\gamma_\mu e_r)(\bar u_s\gamma^\mu u_t)$	Q_{qe}	$(\bar q_p\gamma_\mu q_r)(\bar e_s\gamma^\mu e_t)$
$Q_{lq}^{(3)}$	$(\bar l_p\gamma_\mu \tau^I l_r)(\bar q_s\gamma^\mu \tau^I q_t)$	Q_{ed}	$(\bar e_p\gamma_\mu e_r)(\bar d_s\gamma^\mu d_t)$	$Q_{qu}^{(1)}$	$(\bar q_p\gamma_\mu q_r)(\bar u_s\gamma^\mu u_t)$
		$Q_{ud}^{(1)}$	$(\bar u_p\gamma_\mu u_r)(\bar d_s\gamma^\mu d_t)$	$Q_{qu}^{(8)}$	$(\bar q_p\gamma_\mu T^A q_r)(\bar u_s\gamma^\mu T^A u_t)$
		$Q_{ud}^{(8)}$	$(\bar u_p\gamma_\mu T^A u_r)(\bar d_s\gamma^\mu T^A d_t)$	$Q_{qd}^{(1)}$	$(\bar q_p\gamma_\mu q_r)(\bar d_s\gamma^\mu d_t)$
				$Q_{qd}^{(8)}$	$(\bar q_p\gamma_\mu T^A q_r)(\bar d_s\gamma^\mu T^A d_t)$

$(\bar LR)(\bar RL)$ and $(\bar LR)(\bar LR)$		B-violating	
Q_{ledq}	$(\bar l_p^j e_r)(\bar d_s q_t^j)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(q_s^{\gamma j})^T C l_t^k\right]$
$Q_{quqd}^{(1)}$	$(\bar q_p^j u_r)\varepsilon_{jk}(\bar q_s^k d_t)$	Q_{qqu}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{quqd}^{(8)}$	$(\bar q_p^j T^A u_r)\varepsilon_{jk}(\bar q_s^k T^A d_t)$	Q_{qqq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jn}\varepsilon_{km}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(q_s^{\gamma m})^T C l_t^n\right]$
$Q_{lequ}^{(1)}$	$(\bar l_p^j e_r)\varepsilon_{jk}(\bar q_s^k u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{lequ}^{(3)}$	$(\bar l_p^j \sigma_{\mu\nu} e_r)\varepsilon_{jk}(\bar q_s^k \sigma^{\mu\nu} u_t)$		

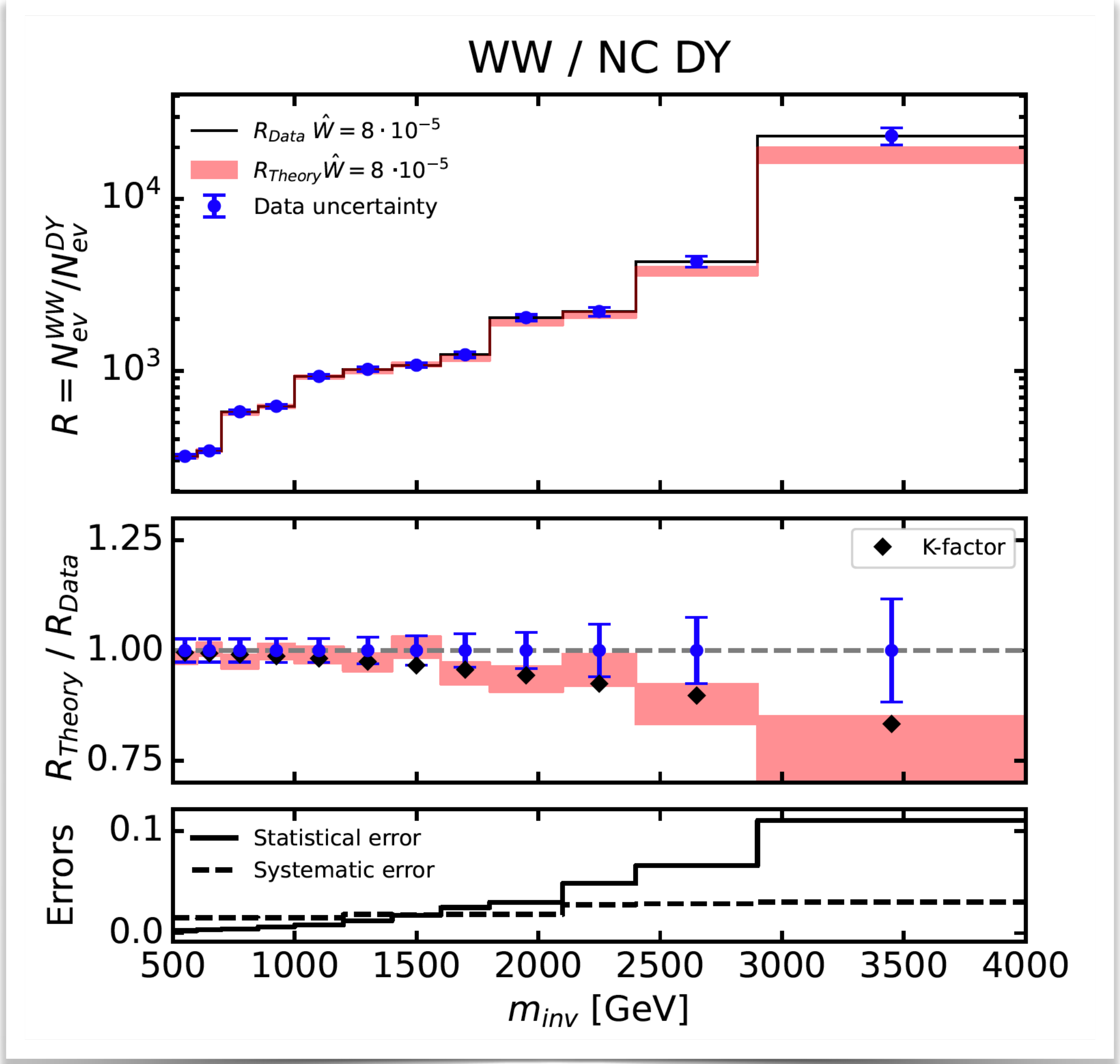
1008.4884

How to disentangle the contamination

Large-x antiquark PDFs were mostly responsible for accommodating NP. More low energy measurements could help to flag inconsistencies.

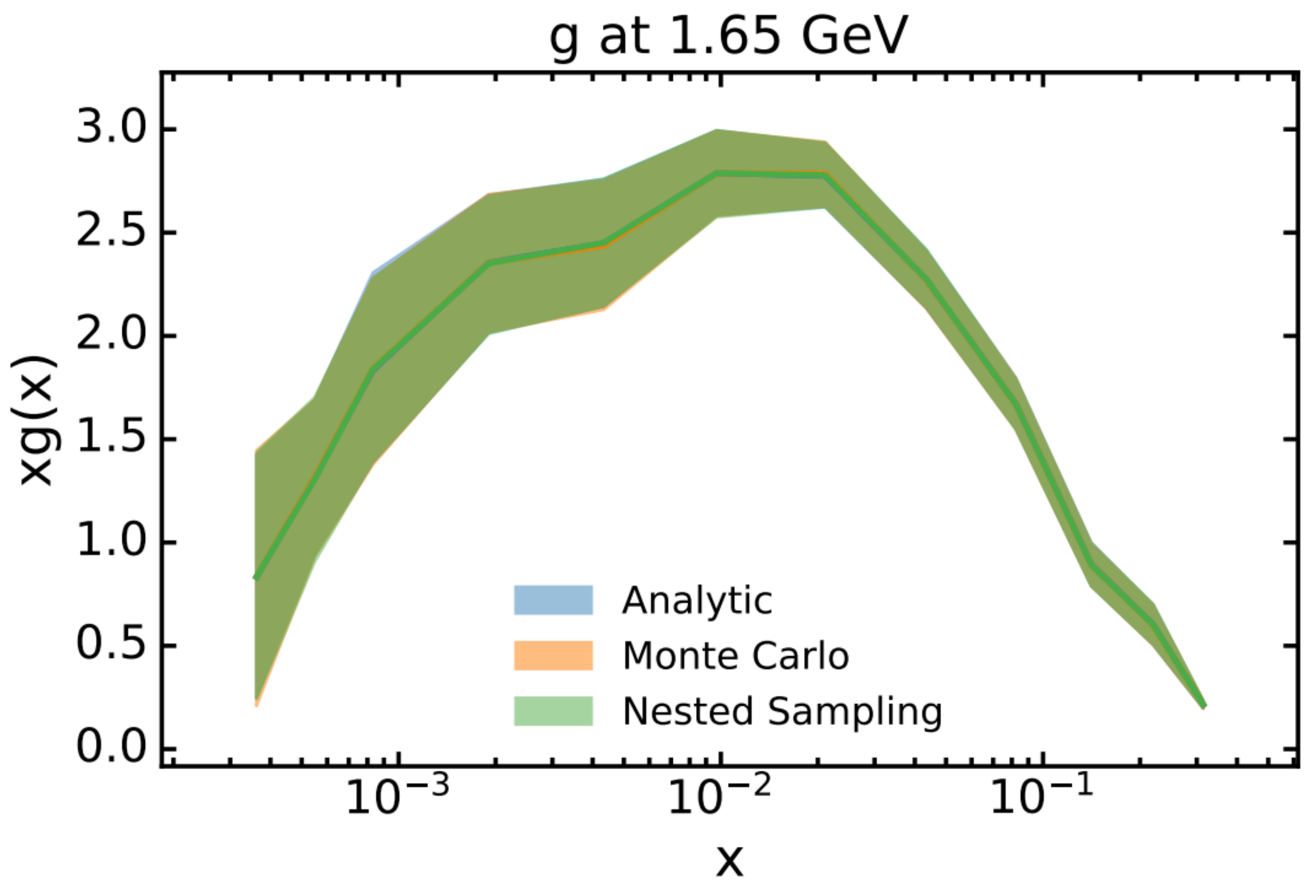
We can disentangle NP effects with ratio observables that probe the same lumi channels.

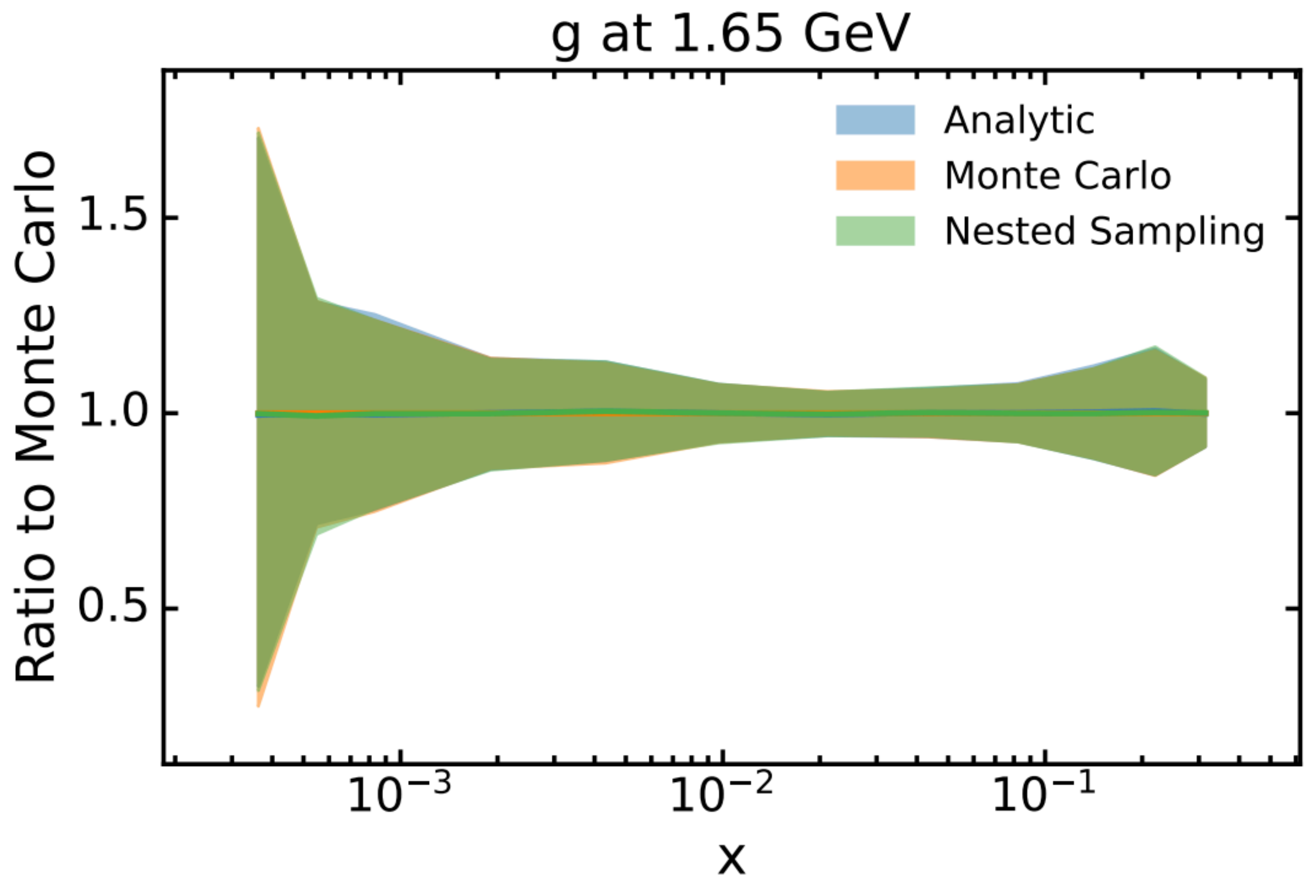
However, one cannot determine if NP is present in DY or diboson datasets.



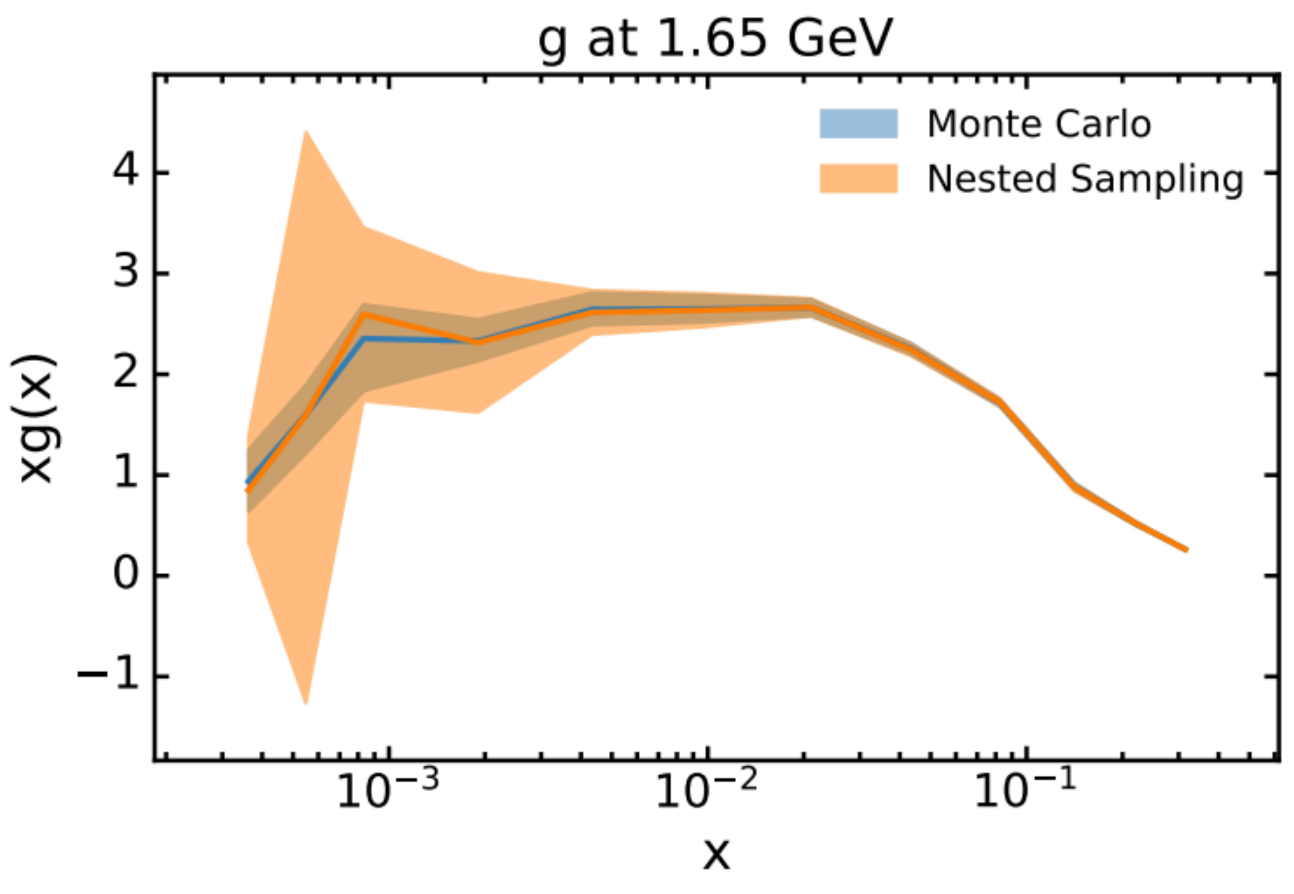
MC Replica vs Bayesian analysis [2404.10056]

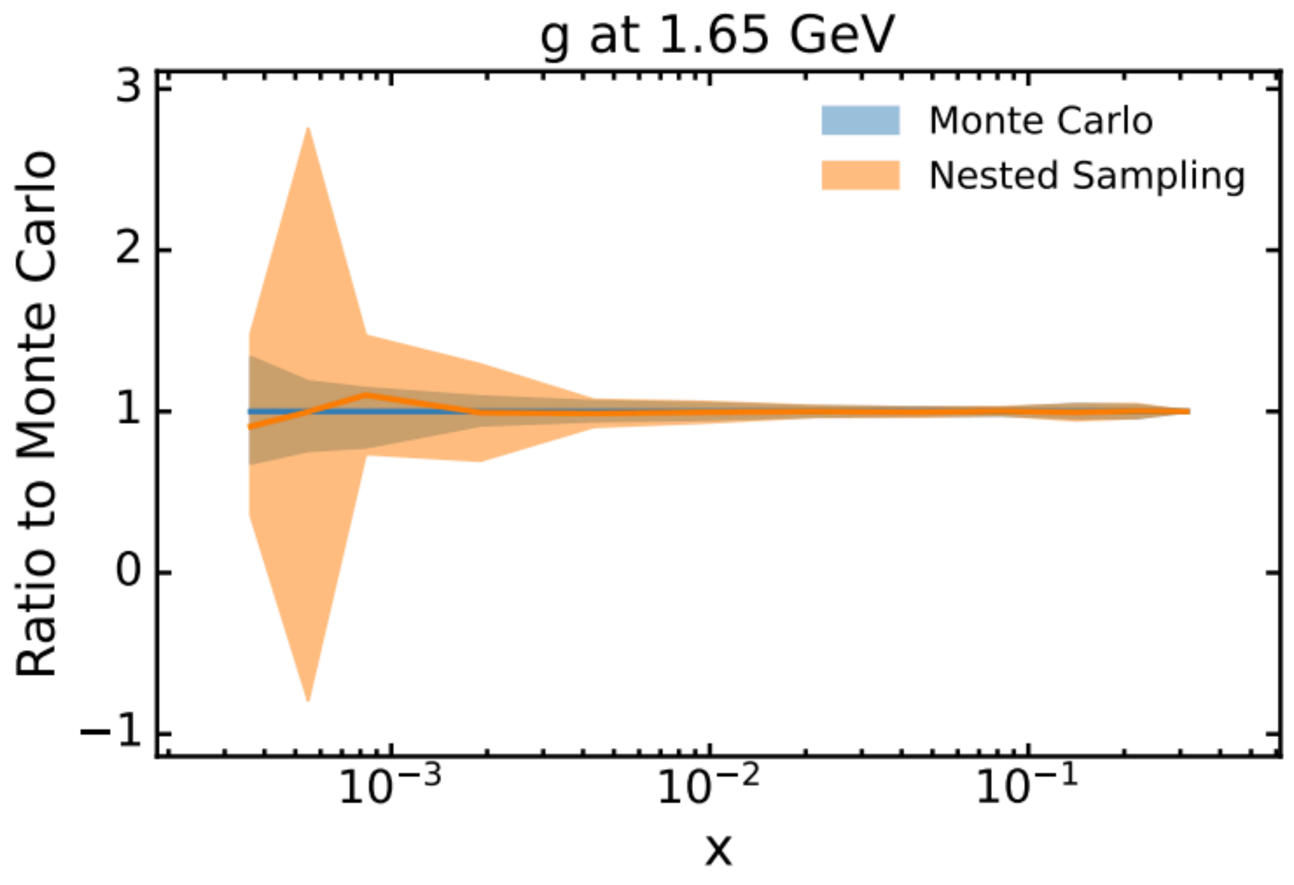
PDF: DIS case





PDF: Hadronic case





MC Replica vs Bayesian analysis

Wilson coefficients