The top quark legacy of the LHC Run II for PDF and SMEFT analyses

for the University of Cambridge Department of Applied Mathematics and Theoretical Physics, May 2023

James Moore, University of Cambridge

PBSP

European Research Council
Established by the European Commission

PBSP: Physics Beyond the Standard Proton

- The **PBSP group** is based at the **University of Cambridge**, and is headed by **Maria Ubiali**; the project is **ERC-funded.**
- The aim is to **investigate interplay between BSM physics and proton structure** - the subject of the rest of this talk!
- The team members are:
  - *Postdocs:* Zahari Kassabov, Maeve Madigan, Luca Mantani
  - *PhD students:* Mark Costantini, Shayan Iranipour (*former*), Elie Hammou, **James Moore**, Manuel Morales, Cameron Voisey (*former*)

PBSP

European Research Council
Established by the European Commission

Talk overview

1. PDFs: a lightning introduction

2. PDF fitting

3. Joint PDF-SMEFT fits

4. The SIMUnet methodology

5. The top quark legacy of the LHC Run II for PDF and SMEFT analyses

1. - PDFs: a lightning introduction

Hadron structure through PDFs

- Hadrons are **QCD bound states** - they are **strongly-coupled, non-perturbative** objects.

$$\mathscr{L} = -\frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu} + \sum_q \overline{q}(i\gamma_\mu D^\mu - m_q)q \longrightarrow \textbf{hadrons?}$$

Hadron structure through PDFs

- Hadrons are **QCD bound states** - they are **strongly-coupled, non-perturbative** objects.

$$\mathscr{L} = -\frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu} + \sum_q \bar{q}(i\gamma_\mu D^\mu - m_q)q \longrightarrow \textbf{hadrons?}$$

- But we still want to make predictions for experiments involving hadrons!

Hadron structure through PDFs

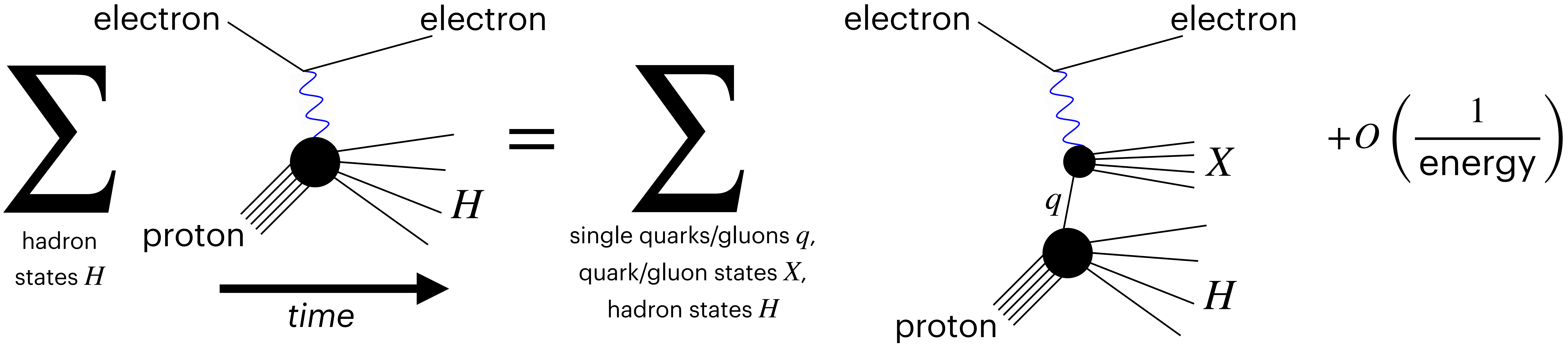
- Hadrons are **QCD bound states** - they are **strongly-coupled, non-perturbative** objects.

$$\mathscr{L} = -\frac{1}{4} G^a_{\mu\nu} G^{a,\mu\nu} + \sum_q \bar{q}(i\gamma_\mu D^\mu - m_q)q \longrightarrow \textbf{hadrons?}$$

- But we still want to make predictions for experiments involving hadrons!
- **Solution:** package all non-perturbative elements into unknown functions, called **parton distribution functions (PDFs)**.

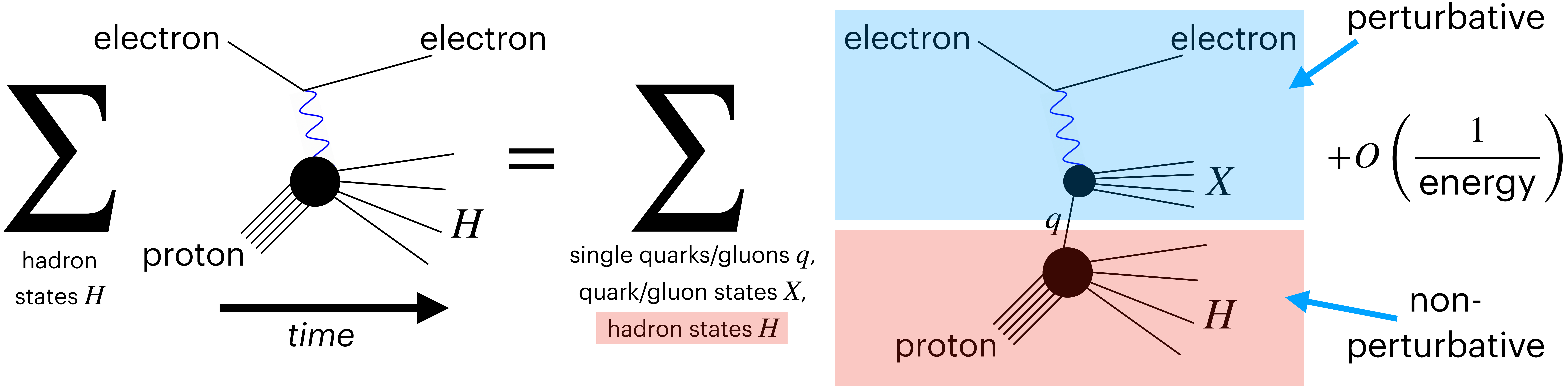
Factorisation theorems

- This is formalised through **factorisation theorems.**
- Model case: **deep inelastic scattering**, $e^- + \text{proton} \to e^- + \text{any hadron.}$



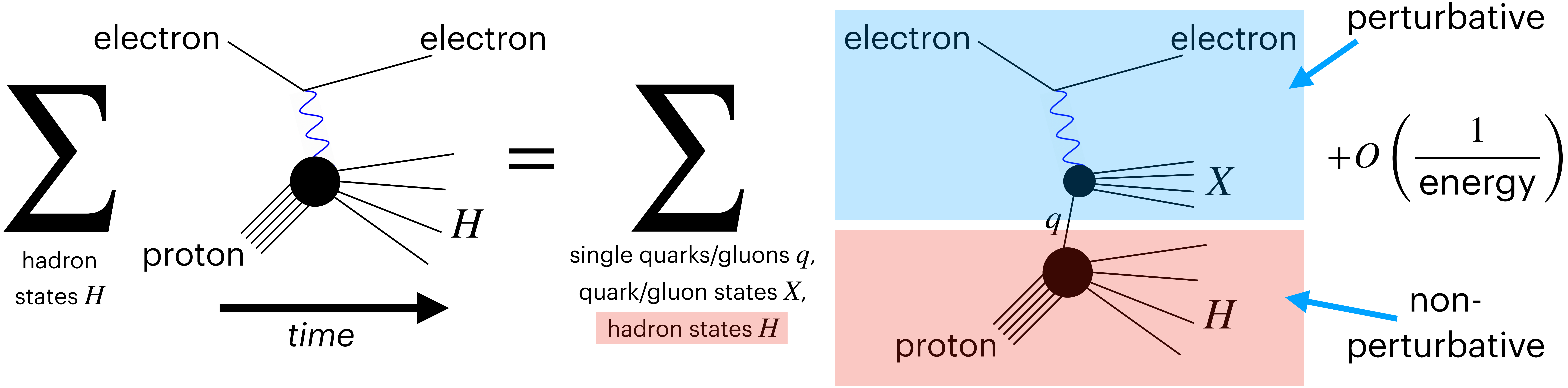
Factorisation theorems

- This is formalised through **factorisation theorems.**
- Model case: **deep inelastic scattering**, $e^- +$ proton $\rightarrow e^- +$ any hadron.



Factorisation theorems

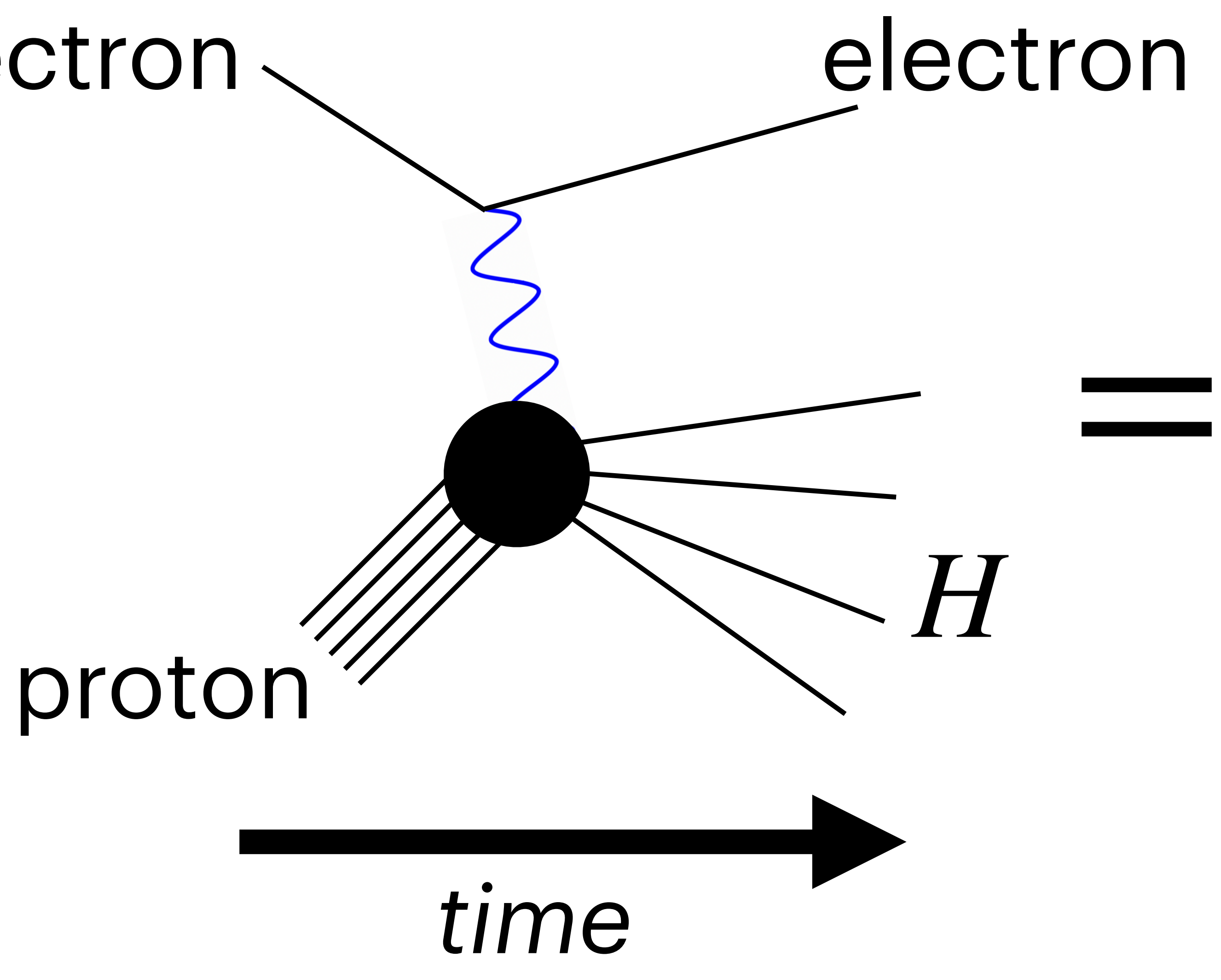
- This is formalised through **factorisation theorems.**
- Model case: **deep inelastic scattering**, $e^- + \text{proton} \to e^- + \text{any hadron}$.



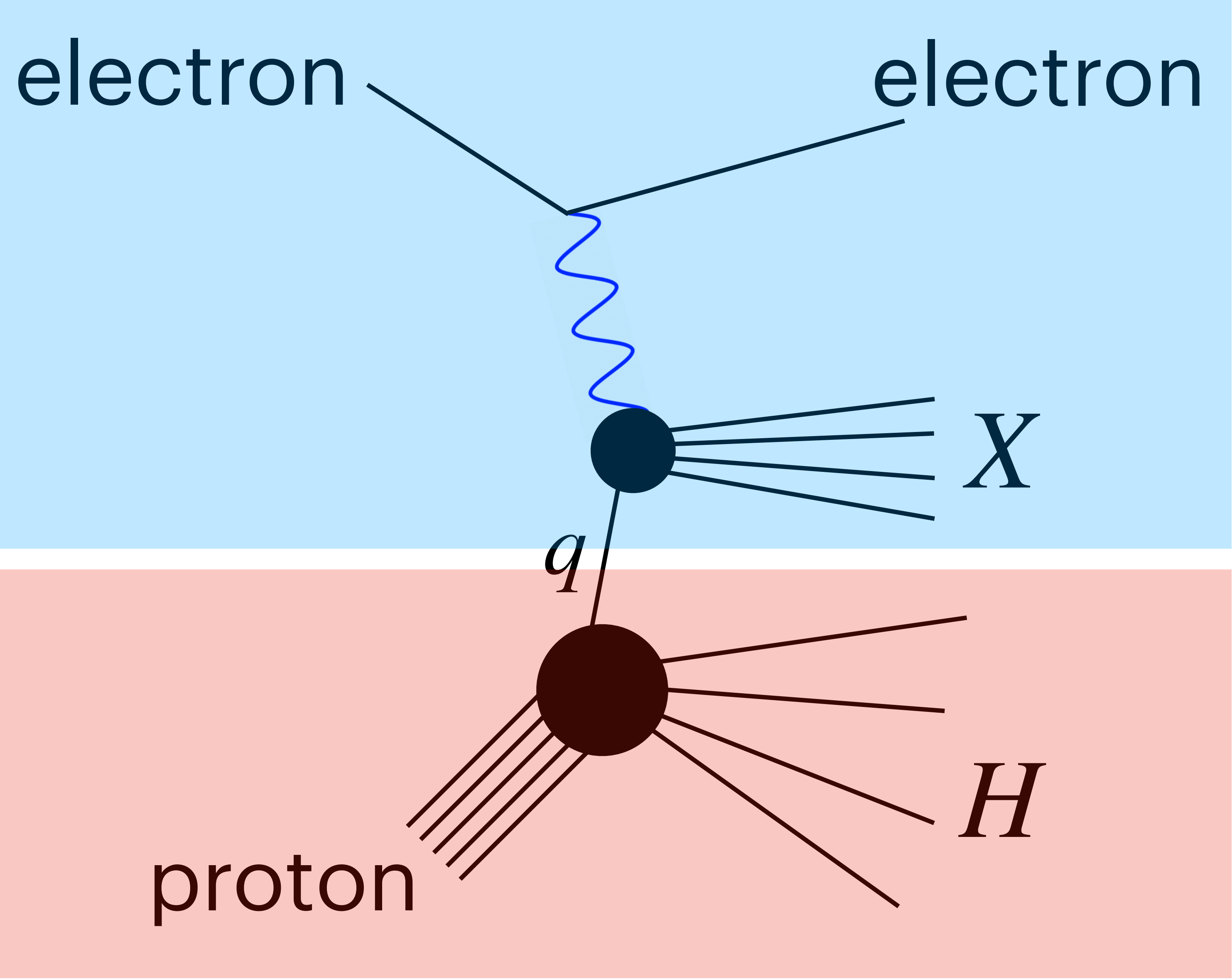
- The calculation is split into a **perturbative process-dependent part,** and a **non-perturbative, BUT universal, parton distribution function**.

Factorisation theorems

$\sum$ (hadron states H)



$=$ $\sum$ (single quarks/gluons q, quark/gluon states X, hadron states H)



$+O\left(\frac{1}{\text{energy}}\right)$

In maths...

$$\sigma(x, Q^2) = \sum_{\substack{\text{single quarks/gluons } q,\\ \text{quark/gluon states } X}} \int_x^1 \frac{dy}{y} \, \hat{\sigma}_{eq\to eX}\left(\frac{x}{y}, Q^2\right) f_q(y, Q^2) + O\left(\frac{1}{\text{energy}}\right)$$

Mellin convolution

Factorisation theorems

In maths… $$\sigma(x, Q^2) = \sum_{\substack{\text{single quarks/gluons } q, \\ \text{quark/gluon states } X}} \int_x^1 \frac{dy}{y} \, \hat{\sigma}_{eq \to eX}\left(\frac{x}{y}, Q^2\right) f_q(y, Q^2) + O\left(\frac{1}{\text{energy}}\right)$$

- **Loosely speaking,** the PDFs $f_q(x, Q^2)$ capture the probability that a certain constituent will be **ejected** in a collision.

Factorisation theorems

In maths... $$\sigma(x, Q^2) = \sum_{\substack{\text{single quarks/gluons } q, \\ \text{quark/gluon states } X}} \int_x^1 \frac{dy}{y} \hat{\sigma}_{eq \to eX}\left(\frac{x}{y}, Q^2\right) f_q(y, Q^2) + O\left(\frac{1}{\text{energy}}\right)$$

- **Loosely speaking,** the PDFs $f_q(x, Q^2)$ capture the probability that a certain constituent will be **ejected** in a collision. They depend on:
  - A **momentum fraction** x - how much of the proton's momentum the ejected constituent carries

Factorisation theorems

In maths... $$\sigma(x, Q^2) = \sum_{\substack{\text{single quarks/gluons } q, \\ \text{quark/gluon states } X}} \int_x^1 \frac{dy}{y} \hat{\sigma}_{eq \to eX}\left(\frac{x}{y}, Q^2\right) f_q(y, Q^2) + O\left(\frac{1}{\text{energy}}\right)$$

- **Loosely speaking,** the PDFs $f_q(x, Q^2)$ capture the probability that a certain constituent will be **ejected** in a collision. They depend on:
  - A **momentum fraction** x - how much of the proton's momentum the ejected constituent carries
  - An **energy scale** Q^2 (comes from **absorbing collinear divergences**)

Factorisation theorems

In maths...
$$\sigma(x, Q^2) = \sum_{\substack{\text{single quarks/gluons } q, \\ \text{quark/gluon states } X}} \int_x^1 \frac{dy}{y} \hat{\sigma}_{eq \to eX}\left(\frac{x}{y}, Q^2\right) f_q(y, Q^2) + O\left(\frac{1}{\text{energy}}\right)$$

- **Loosely speaking,** the PDFs $f_q(x, Q^2)$ capture the probability that a certain constituent will be **ejected** in a collision. They depend on:
  - A **momentum fraction** x - how much of the proton's momentum the ejected constituent carries
  - An **energy scale** Q^2 (comes from **absorbing collinear divergences**)
  - The fact we are colliding **protons** - if we started with a neutron, we would need different PDFs

Universality of PDFs

- Importantly, PDFs are **universal.** The **same** parton distributions can **also** be used in the **Drell-Yan process**: the collision of two protons to make an **electron-positron pair**, plus any hadrons.

Universality of PDFs

- Importantly, PDFs are **universal.** The **same** parton distributions can **also** be used in the **Drell-Yan process**: the collision of two protons to make an **electron-positron pair**, plus any hadrons.

Scaling of PDFs

- Whilst the PDFs are non-perturbative, we can still say something about their Q^2-dependence, which enters the PDFs when we **absorb collinear IR divergences**.
- Just as in **standard UV renormalisation theory**, this leads to a Callan-Symanzik equation for the PDFs called the **DGLAP equation**:

$$Q^2 \frac{\partial f_q(x, Q^2)}{\partial Q^2} = \sum_{\text{quarks/gluons } q'} \int_x^1 \frac{dy}{y} P_{qq'}\left(\frac{x}{y}\right) f_{q'}(x, Q^2)$$

- The functions (technically distributions) $P_{qq'}$ are called **splitting functions** and can be determined perturbatively.

Scaling of PDFs

$$Q^2 \frac{\partial f_q(x, Q^2)}{\partial Q^2} = \sum_{\text{quarks/gluons } q'} \int_x^1 \frac{dy}{y} P_{qq'}\left(\frac{x}{y}\right) f_{q'}(x, Q^2)$$

- This means if we know the PDFs at some **initial energy scale** Q_0, we can compute them at some energy scale $Q > Q_0$ by solving DGLAP.
- In particular, only the x-dependence of the PDFs is truly **unknown**.
- We can obtain this x-dependence by **fits to collider data**, as we shall now describe...

Summary of PDFs

- The **non-perturbative structure** of hadrons can be parametrised by **parton distribution functions** $f_q(x, Q^2)$, which depend only on the **type of hadron** being collided, **not** on the process.

- The PDFs have **known** Q^2**-dependence**, described by a linear system of **integro-differential equations** called the **DGLAP equations**.

- The PDFs have **unknown** x**-dependence**, which must be obtained through fits to experimental data.

2. - PDF fitting

How to make PDFs...

- *TLDRN:* Fitting PDFs using experimental data is an **ill-posed problem**.

How to make PDFs...

- *TLDRN:* Fitting PDFs using experimental data is an **ill-posed problem**.
- In short, you have **finite amounts of data** from experiments, but the space of possible PDFs is **infinite-dimensional**. What do we do?

How to make PDFs...

- *TLDRN*: Fitting PDFs using experimental data is an **ill-posed problem**.
- In short, you have **finite amounts of data** from experiments, but the space of possible PDFs is **infinite-dimensional**. What do we do?
- PDF fitting groups **assume a functional form** for the PDFs at some **initial energy scale**, parametrised by a finite set of parameters. They then obtain the PDF at all energy scales using the **DGLAP equation**.

How to make PDFs...

- *TLDRN*: Fitting PDFs using experimental data is an **ill-posed problem**.
- In short, you have **finite amounts of data** from experiments, but the space of possible PDFs is **infinite-dimensional**. What do we do?
- PDF fitting groups **assume a functional form** for the PDFs at some **initial energy scale**, parametrised by a finite set of parameters. They then obtain the PDF at all energy scales using the **DGLAP equation**.
- Example functional form:

$$f(x, Q_0^2) = Ax^{\alpha}(1-x)^{\beta}\left(1 + ax^{1/2} + bx + cx^{3/2}\right)$$

large and small x behaviour motivated by **Regge theory**

polynomial in $\sqrt{x}$

How to make PDFs…

- The best-fit parameters are found by **minimising the χ^2-statistic**, which measures the **goodness of fit** of our model:

$$\chi^2 = (\text{data} - \text{theory})^T \text{covariance}^{-1} (\text{data} - \text{theory})$$

How to make PDFs...

- The best-fit parameters are found by **minimising the χ^2-statistic**, which measures the **goodness of fit** of our model:

$$\chi^2 = (\text{data} - \text{theory})^T \text{covariance}^{-1} (\text{data} - \text{theory})$$

vector of
data points

How to make PDFs...

- The best-fit parameters are found by **minimising the χ^2-statistic**, which measures the **goodness of fit** of our model:

$$\chi^2 = (\text{data} - \text{theory})^T \text{covariance}^{-1} (\text{data} - \text{theory})$$

vector of data points

vector of theory predictions, from factorisation theorems

How to make PDFs...

- The best-fit parameters are found by **minimising the χ^2-statistic**, which measures the **goodness of fit** of our model:

$$\chi^2 = (\text{data} - \text{theory})^T \text{covariance}^{-1} (\text{data} - \text{theory})$$

vector of data points

vector of theory predictions, from factorisation theorems

experimental* covariance matrix

How to make PDFs...

- The best-fit parameters are found by **minimising the χ^2-statistic**, which measures the **goodness of fit** of our model:

$$\chi^2 = (\text{data} - \text{theory})^T \text{covariance}^{-1} (\text{data} - \text{theory})$$

vector of data points

vector of theory predictions, from factorisation theorems

experimental* covariance matrix

- *General idea*: we want **theory to be close to data**, but if the data is **more uncertain**, we don't require such precise agreement.

How to make PDFs...

- It's not good enough to find the PDF parameters which give just the **central data values** because experimental data comes with **uncertainty**. We must also **propagate errors** properly too.

How to make PDFs...

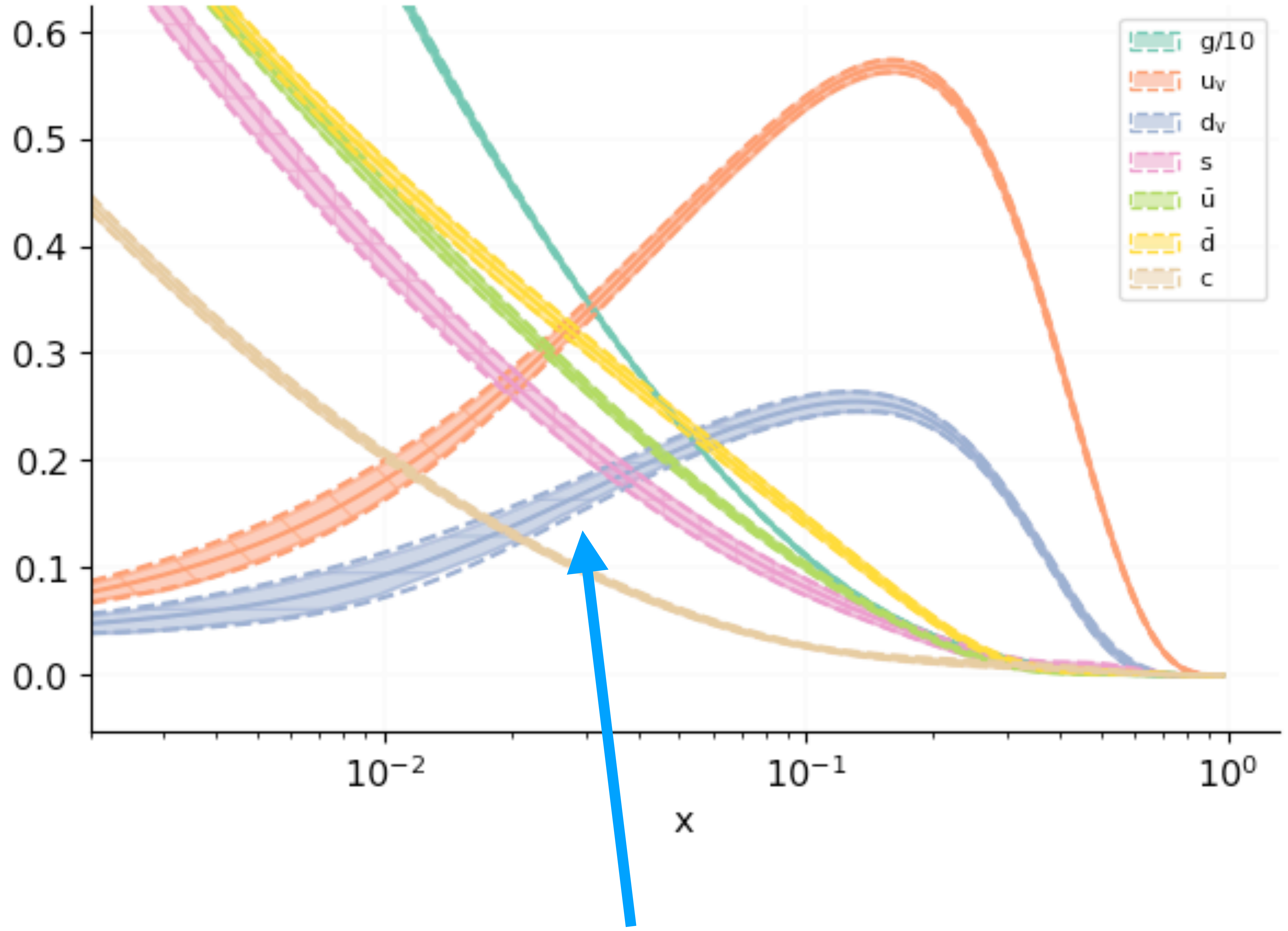
- It's not good enough to find the PDF parameters which give just the **central data values** because experimental data comes with **uncertainty**. We must also **propagate errors** properly too.
- One way to handle this is using **Monte Carlo error propagation**.* We create 100 different copies of **Monte Carlo pseudodata**, generated as a **multivariate Gaussian distribution** around the central data, then find the best-fit PDF parameters for each of the 100 copies.

How to make PDFs...

- It's not good enough to find the PDF parameters which give just the **central data values** because experimental data comes with **uncertainty**. We must also **propagate errors** properly too.
- One way to handle this is using **Monte Carlo error propagation**.* We create 100 different copies of **Monte Carlo pseudodata**, generated as a **multivariate Gaussian distribution** around the central data, then find the best-fit PDF parameters for each of the 100 copies.
- We can then take **envelopes** to get uncertainties from the resulting **PDF ensemble**.

How to make PDFs...

- It's not good enough to find the PDF parameters which give just the **central data values** because experimental data comes with **uncertainty**. We must also **propagate errors** properly too.
- One way to handle this is using **Monte Carlo error propagation**.* We create 100 different copies of **Monte Carlo pseudodata**, generated as a **multivariate Gaussian distribution** around the central data, then find the best-fit PDF parameters for each of the 100 copies.
- We can then take **envelopes** to get uncertainties from the resulting **PDF ensemble**.



PDFs with error bands

The choice of functional form

- The choice of functional form that we have suggested so far is:

$$f(x, Q_0^2) = Ax^{\alpha}(1-x)^{\beta}\left(1 + ax^{1/2} + bx + cx^{3/2}\right)$$

The choice of functional form

- The choice of functional form that we have suggested so far is:

$$f(x, Q_0^2) = Ax^{\alpha}(1-x)^{\beta}\left(1 + ax^{1/2} + bx + cx^{3/2}\right)$$

- This seems a bit arbitrary though! To try to remove as much **bias** as possible, another possible choice is to parametrise the PDFs using a **neural network** instead:

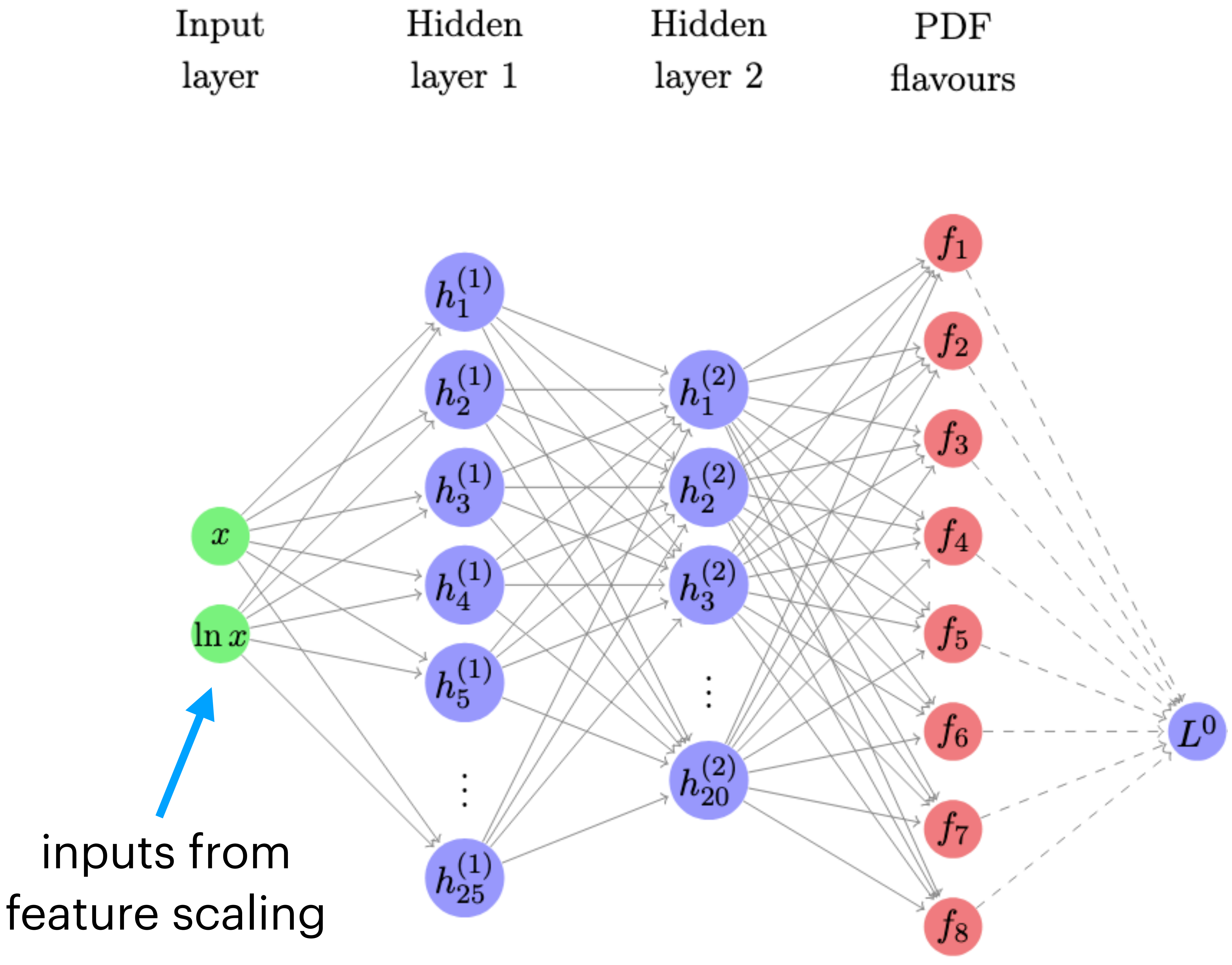
$$f(x, Q_0^2) = Ax^{\alpha}(1-x)^{\beta}\mathrm{NN}(x, \omega)$$

- Here, $\mathrm{NN}(x, \omega)$ is a **neural network** which takes in x as an argument, and has network parameters ω.

The choice of functional form

$$f(x, Q_0^2) = Ax^{\alpha}(1-x)^{\beta}\mathrm{NN}(x, \omega)$$

- The neural network parametrisation is used by the **NNPDF collaboration**, whose fitting code is **publicly available**.
- *See 2109.02653 and 2109.02671 for details.*



3. - Joint PDF-SMEFT fits

The Standard Model is *incomplete*...

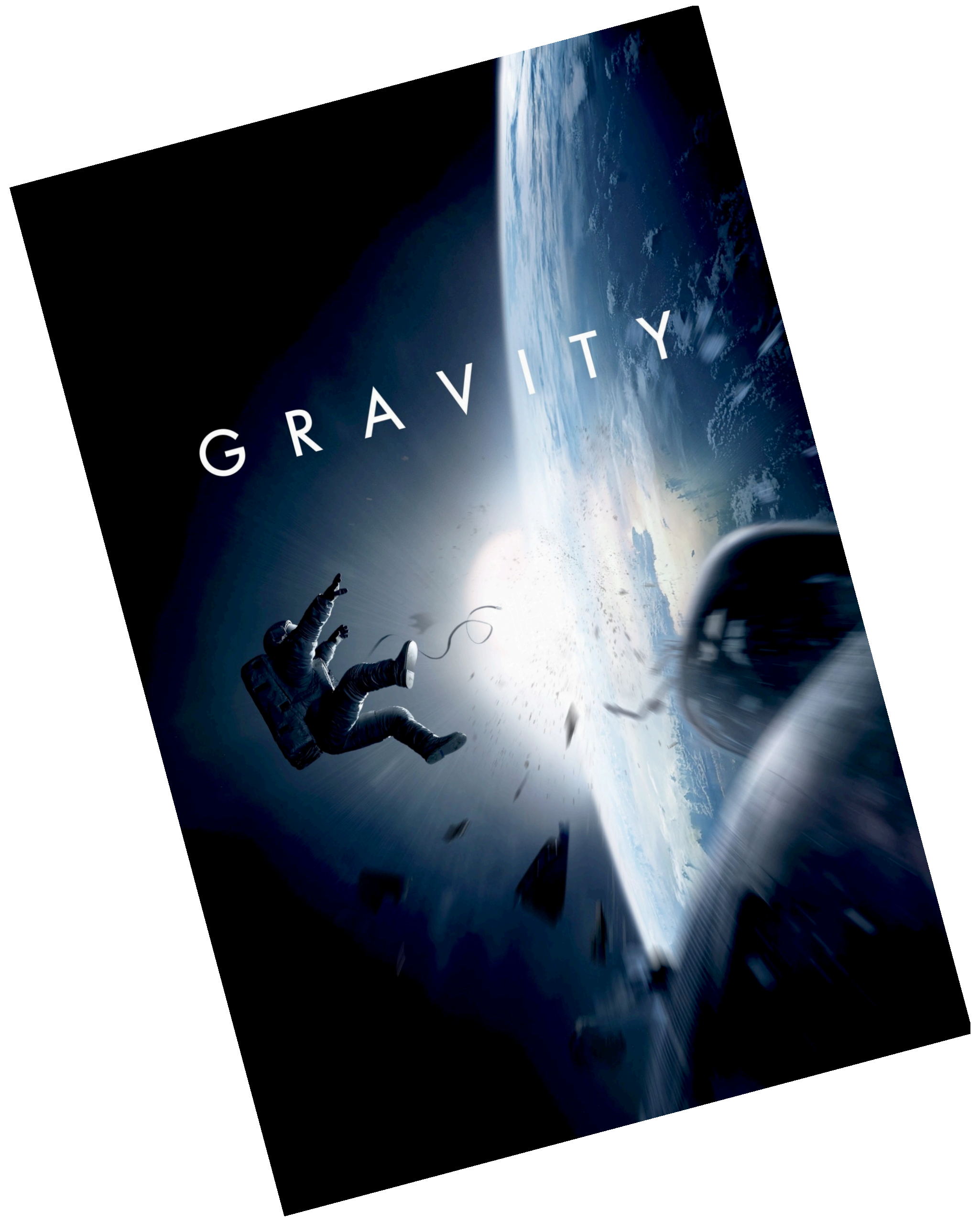
- PDF fitting usually assumes that the **Standard Model is correct**.

The Standard Model is *incomplete*...

- PDF fitting usually assumes that the **Standard Model is correct**.
- However, whilst the Standard Model has been **extremely successful**, it is known to be incomplete. There are lots of things it does not describe:

The Standard Model is *incomplete*...

- PDF fitting usually assumes that the **Standard Model is correct**.
- However, whilst the Standard Model has been **extremely successful**, it is known to be incomplete. There are lots of things it does not describe:
  - *Gravity*



The Standard Model is *incomplete*...

- PDF fitting usually assumes that the **Standard Model is correct**.
- However, whilst the Standard Model has been **extremely successful**, it is known to be incomplete. There are lots of things it does not describe:
  - *Gravity*
  - *Dark matter*

The Standard Model is *incomplete*...

- PDF fitting usually assumes that the **Standard Model is correct**.
- However, whilst the Standard Model has been **extremely successful**, it is known to be incomplete. There are lots of things it does not describe:
  - *Gravity*
  - *Dark matter*
  - *Neutrino masses*

The Standard Model is *incomplete*...

- PDF fitting usually assumes that the **Standard Model is correct**.
- However, whilst the Standard Model has been **extremely successful**, it is known to be incomplete. There are lots of things it does not describe:
  - *Gravity*
  - *Dark matter*
  - *Neutrino masses*
  - *Baryon number asymmetry*

The Standard Model is *incomplete...*

- PDF fitting usually assumes that the **Standard Model is correct**.
- However, whilst the Standard Model has been **extremely successful**, it is known to be incomplete. There are lots of things it does not describe:
  - *Gravity*
  - *Dark matter*
  - *Neutrino masses*
  - *Baryon number asymmetry*
  - *many more...*

So how do we fix the Standard Model?

- For example, to **include dark matter** in the Standard Model, we might **hypothesise new particles** and add them in. The Standard Model Lagrangian density is augmented to:

$$\mathscr{L}_{\text{new}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_{\text{dark matter}}$$

So how do we fix the Standard Model?

- For example, to **include dark matter** in the Standard Model, we might **hypothesise new particles** and add them in. The Standard Model Lagrangian density is augmented to:

$$\mathscr{L}_{\text{new}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_{\text{dark matter}}$$

- We could then **try to produce the new particles directly** (*direct detection*), or **fit existing data using this theory to see if we get a better fit** (*indirect detection*).

So how do we fix the Standard Model?

- For example, to **include dark matter** in the Standard Model, we might **hypothesise new particles** and add them in. The Standard Model Lagrangian density is augmented to:

$$\mathscr{L}_{\text{new}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_{\text{dark matter}}$$

- We could then **try to produce the new particles directly** (*direct detection*), or **fit existing data using this theory to see if we get a better fit** (*indirect detection*).
- However, there are **thousands** of possibilities, so just guessing particles seems a bit like **stabbing in the dark**!

So how do we fix the Standard Model?

- For example, to **include dark matter** in the Standard Model, we might **hypothesise new particles** and add them in. The Standard Model Lagrangian density is augmented to:

$$\mathscr{L}_{\text{new}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_{\text{dark matter}}$$

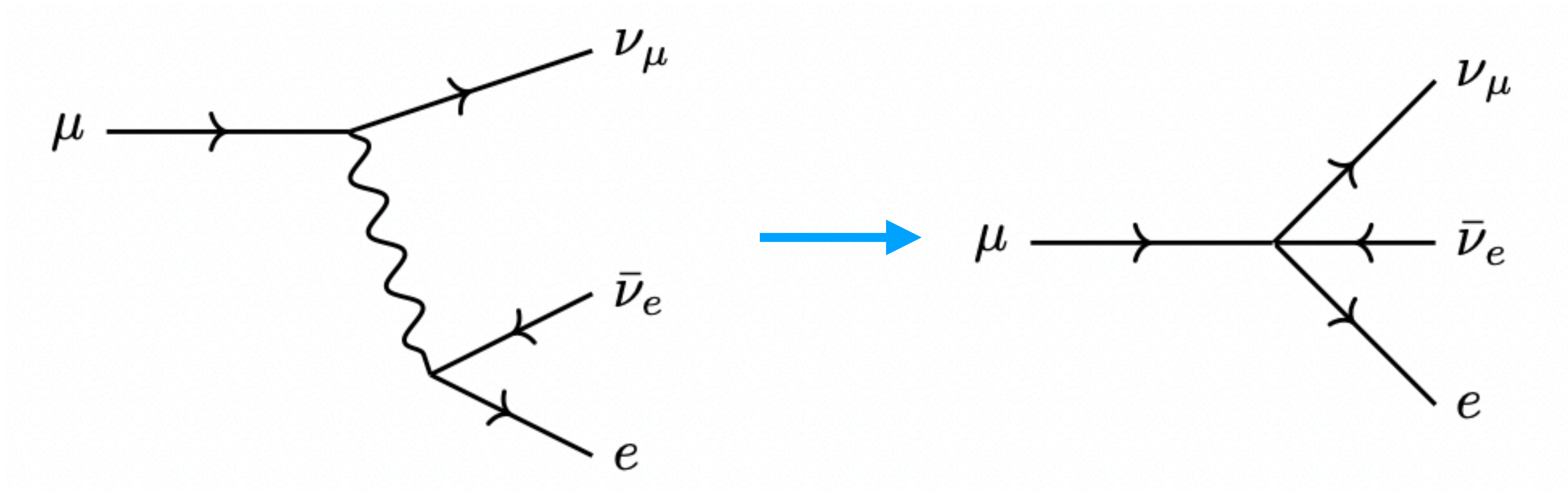
- We could then **try to produce the new particles directly** (*direct detection*), or **fit existing data using this theory to see if we get a better fit** (*indirect detection*).
- However, there are **thousands** of possibilities, so just guessing particles seems a bit like **stabbing in the dark**!
- Some models are **more motivated** than others, but it would be nice to have a more general approach...

Enter the SMEFT...

- Fortunately, the language of **effective field theory** exists to help us tackle this problem.

Enter the SMEFT...

- Fortunately, the language of **effective field theory** exists to help us tackle this problem.
- *Idea*: at **low energies** we can **integrate out heavy particles from a theory**, giving **effective non-renormalisable interactions**:



- Integrating out particles can also yield **shifts in SM couplings**.

Enter the SMEFT...

- Since **any*** heavy particle manifests at low energies as non-renormalisable interactions, if we are hunting for **extensions of the SM**, we can simply **add on all non-renormalisable operators built from the SM fields** (and respecting the SM symmetries):

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

Enter the SMEFT...

- Since **any*** heavy particle manifests at low energies as non-renormalisable interactions, if we are hunting for **extensions of the SM**, we can simply **add on all non-renormalisable operators built from the SM fields** (and respecting the SM symmetries):

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- We can organise the additional non-renormalisable operators by their **mass dimension**, with higher-dimensional operators being **suppressed** by **powers of** $1/\Lambda$, where Λ is a characteristic scale of the New Physics.

SMEFT fits

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- **Fitting collaborations** try to determine the couplings in $\mathscr{L}_5, \mathscr{L}_6 \dots$ via **precise fits to collider data**.

SMEFT fits

$$\mathscr{L}_{\mathrm{SMEFT}} = \mathscr{L}_{\mathrm{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- **Fitting collaborations** try to determine the couplings in $\mathscr{L}_5, \mathscr{L}_6 \ldots$ via **precise fits to collider data**.
- Unfortunately, there are **2499 different operators** in $\mathscr{L}_6$, so this is a lot of work! At the moment, people can only fit subsets of the operators at a time.

SMEFT fits

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- **Fitting collaborations** try to determine the couplings in $\mathscr{L}_5, \mathscr{L}_6 \ldots$ via **precise fits to collider data**.
- Unfortunately, there are **2499 different operators** in $\mathscr{L}_6$, so this is a lot of work! At the moment, people can only fit subsets of the operators at a time.
- However, the number of operators **decreases significantly** if we **assume additional symmetries**, e.g. **no baryon number violation**. There are only **59 operators** if we assume **flavour universality**.

SMEFT fits

$$\mathscr{L}_{\mathrm{SMEFT}} = \mathscr{L}_{\mathrm{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- **Fitting collaborations** try to determine the couplings in $\mathscr{L}_5, \mathscr{L}_6 \ldots$ via **precise fits to collider data**.
- Unfortunately, there are **2499 different operators** in $\mathscr{L}_6$, so this is a lot of work! At the moment, people can only fit subsets of the operators at a time.
- However, the number of operators **decreases significantly** if we **assume additional symmetries**, e.g. **no baryon number violation**. There are only **59 operators** if we assume **flavour universality**.
- The main sectors studied so far are: **top**, **Higgs** and **electroweak** physics.

SMEFT fits

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- Finally, note that various fitting groups **just fit** the SMEFT couplings, for example the **SMEFiT collaboration**, and the **FitMaker collaboration**.

SMEFT fits

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \mathscr{L}_5 + \mathscr{L}_6 + \cdots$$

- Finally, note that various fitting groups **just fit** the SMEFT couplings, for example the **SMEFiT collaboration**, and the **FitMaker collaboration**.
- In particular, SMEFiT and FitMaker both assume a **SM PDF input**. This could be **problematic** because the PDFs were fitted **assuming no New Physics**...

Joint PDF-SMEFT fits?

- In more detail ($\otimes$ is shorthand for the **Mellin convolution**)...

Joint PDF-SMEFT fits?

- In more detail ($\otimes$ is shorthand for the **Mellin convolution**)...

PDF parameter fits

- Fix SMEFT parameters (usually to zero), $c = \bar{c}$:

$$\sigma(\bar{c}, \theta) = \hat{\sigma}(\bar{c}) \otimes \text{PDF}(\theta)$$

- Optimal PDF parameters θ^* then have an **implicit dependence** on initial SMEFT parameter choice: $\text{PDF}(\theta^*) \equiv \text{PDF}(\theta^*(\bar{c}))$.

Joint PDF-SMEFT fits?

- In more detail ($\otimes$ is shorthand for the **Mellin convolution**)...

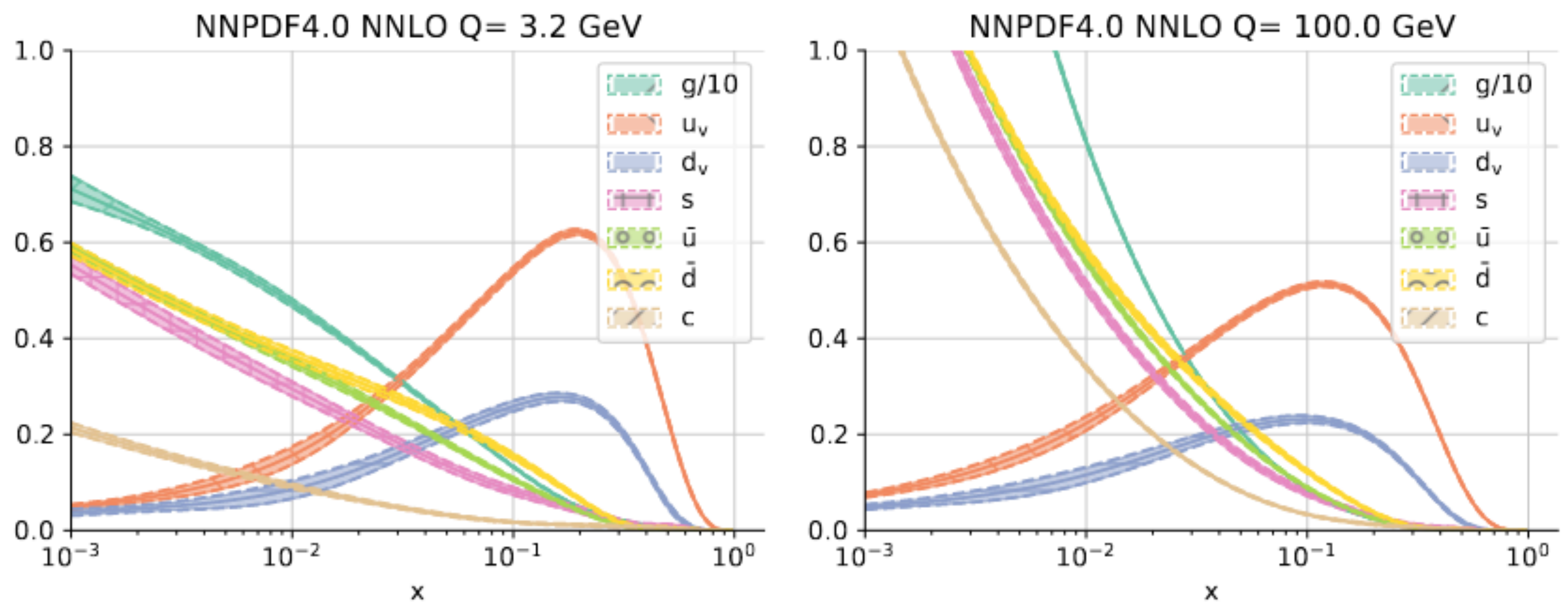
PDF parameter fits

- Fix SMEFT parameters (usually to zero), $c = \bar{c}$:

$$\sigma(\bar{c}, \theta) = \hat{\sigma}(\bar{c}) \otimes \text{PDF}(\theta)$$

- Optimal PDF parameters θ^* then have an **implicit dependence** on initial SMEFT parameter choice: $\text{PDF}(\theta^*) \equiv \text{PDF}(\theta^*(\bar{c}))$.

- E.g. NNPDF4.0 fit, Ball et al., 2109.02653.



Joint PDF-SMEFT fits?

- In more detail ($\otimes$ is shorthand for the **Mellin convolution**)...

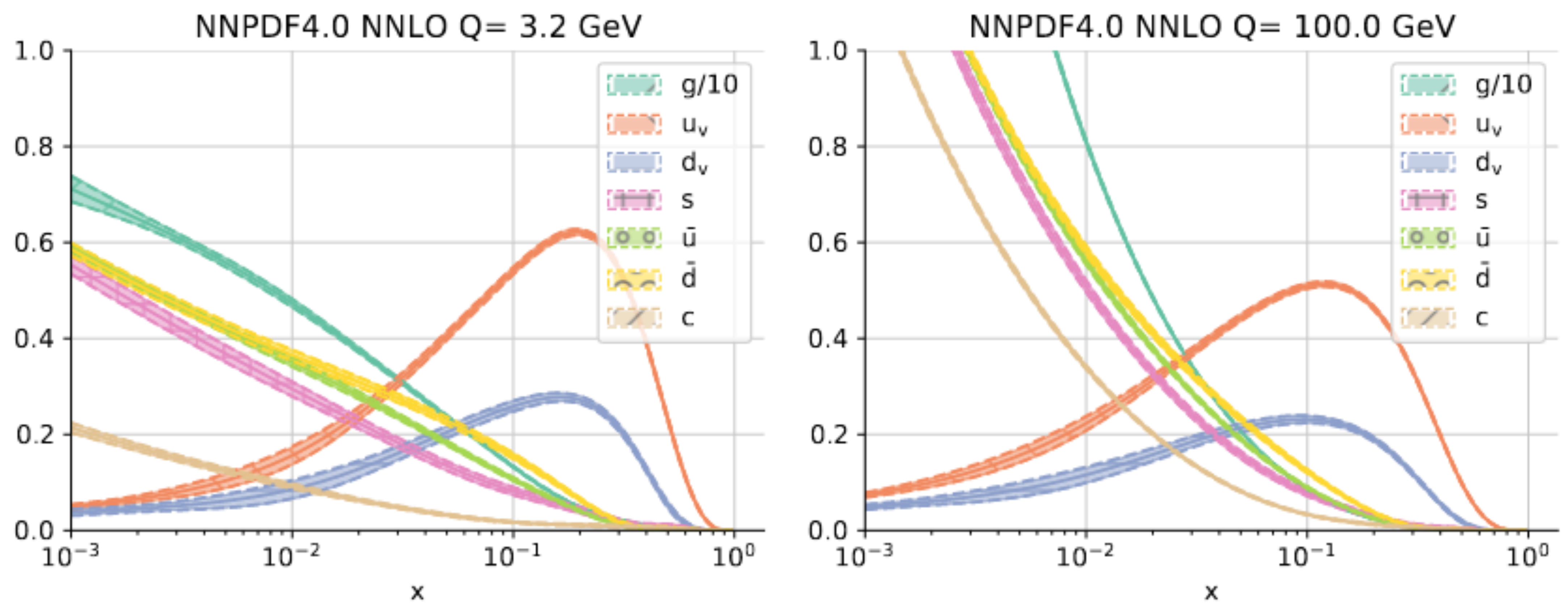
PDF parameter fits

- Fix SMEFT parameters (usually to zero), $c = \bar{c}$:

$$\sigma(\bar{c}, \theta) = \hat{\sigma}(\bar{c}) \otimes \text{PDF}(\theta)$$

- Optimal PDF parameters θ^* then have an **implicit dependence** on initial SMEFT parameter choice: $\text{PDF}(\theta^*) \equiv \text{PDF}(\theta^*(\bar{c}))$.

- E.g. NNPDF4.0 fit, Ball et al., 2109.02653.



SMEFT parameter fits

- Fix PDF parameters $\theta = \bar{\theta}$:

$$\sigma(c, \overline{\theta}) = \hat{\sigma}(c) \otimes \text{PDF}(\overline{\theta})$$

- Optimal SMEFT parameters c^* then have an **implicit dependence** on PDF choice: $c^* = c^*(\overline{\theta})$.

Joint PDF-SMEFT fits?

- In more detail ($\otimes$ is shorthand for the **Mellin convolution**)...

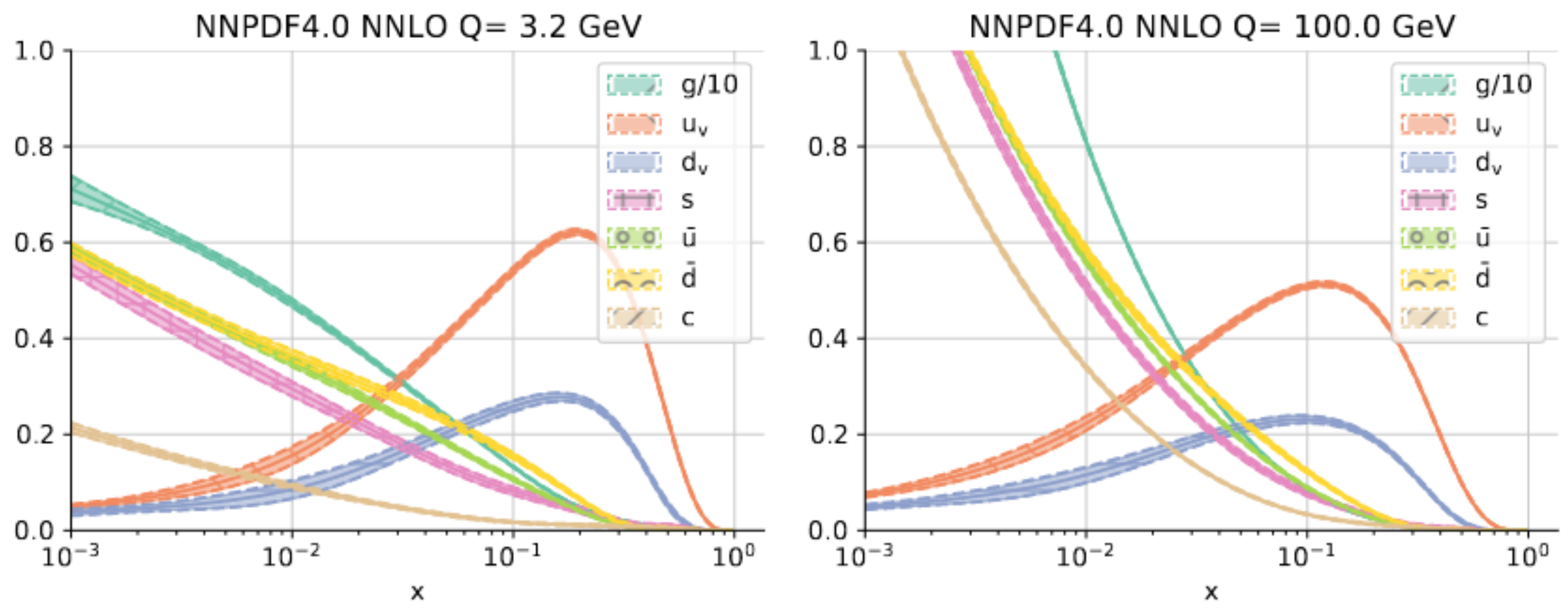
PDF parameter fits

- Fix SMEFT parameters (usually to zero), $c = \bar{c}$:

$$\sigma(\bar{c}, \theta) = \hat{\sigma}(\bar{c}) \otimes \text{PDF}(\theta)$$

- Optimal PDF parameters θ^* then have an **implicit dependence** on initial SMEFT parameter choice: $\text{PDF}(\theta^*) \equiv \text{PDF}(\theta^*(\bar{c}))$.

- E.g. NNPDF4.0 fit, Ball et al., 2109.02653.



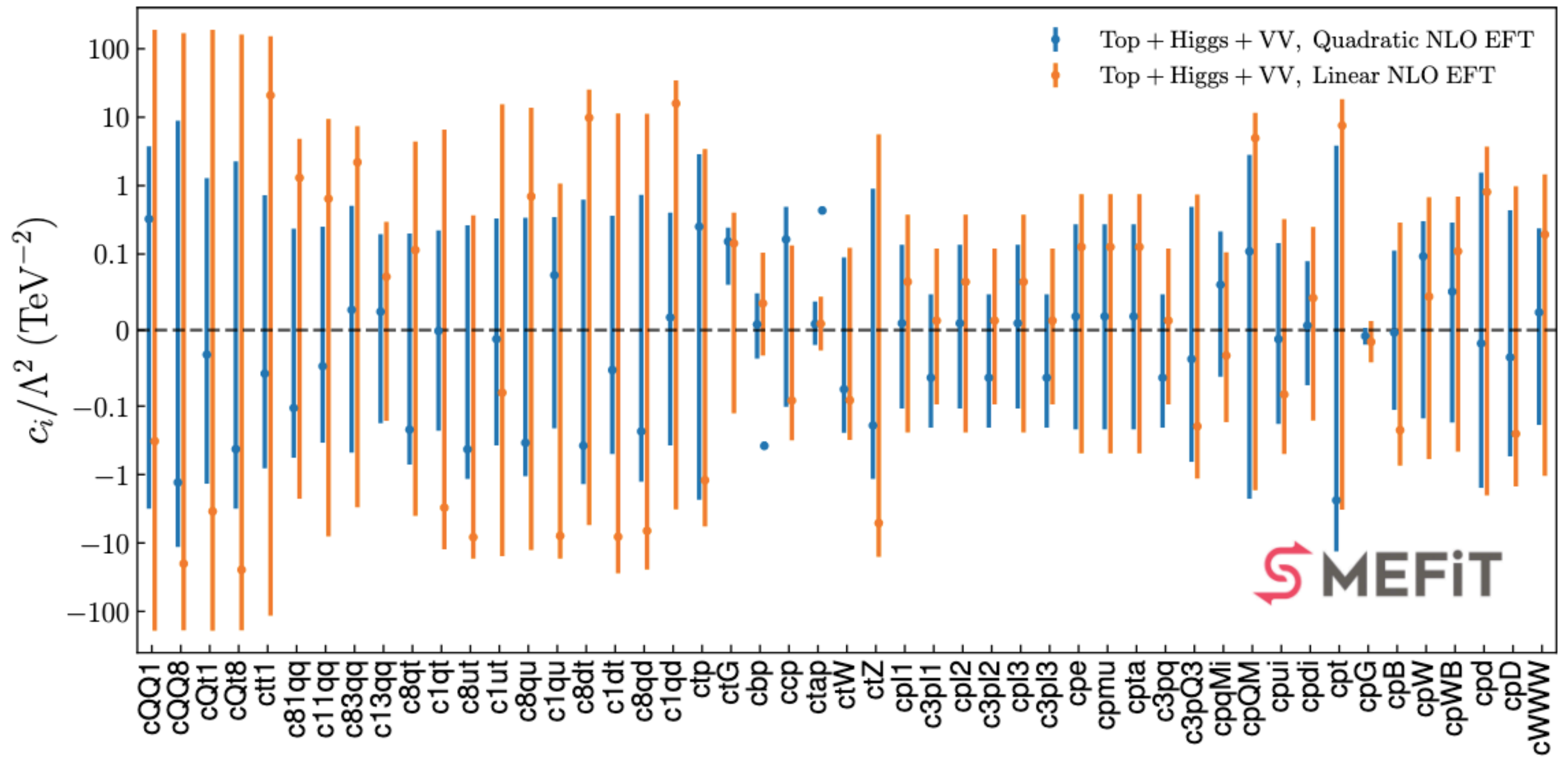
SMEFT parameter fits

- Fix PDF parameters $\theta = \bar{\theta}$:

$$\sigma(c, \bar{\theta}) = \hat{\sigma}(c) \otimes \text{PDF}(\bar{\theta})$$

- Optimal SMEFT parameters c^* then have an **implicit dependence** on PDF choice: $c^* = c^*(\bar{\theta})$.

- E.g. SMEFiT, Ethier et al., 2105.00006.



Joint PDF-SMEFT fits?

- **This could lead to inconsistencies.**

PDF parameter fits

$$\text{PDF}(\theta^*) \equiv \text{PDF}(\theta^*(\bar{c}))$$

- Fitted PDFs can depend implicitly on fixed SMEFT parameters used in the fit.

SMEFT parameter fits

$$c^* \equiv c^*(\bar{\theta})$$

- Bounds on SMEFT parameters can depend implicitly on the fixed PDF set used in the fit.

Joint PDF-SMEFT fits?

- **This could lead to inconsistencies**.

PDF parameter fits

$$\text{PDF}(\theta^*) \equiv \text{PDF}(\theta^*(\bar{c}))$$

- Fitted PDFs can depend implicitly on fixed SMEFT parameters used in the fit.

SMEFT parameter fits

$$c^* \equiv c^*(\bar{\theta})$$

- Bounds on SMEFT parameters can depend implicitly on the fixed PDF set used in the fit.

- In particular, if we fit PDFs **assuming all SMEFT couplings are zero**, but then **use those PDFs in a fit of SMEFT couplings**, our resulting bounds **could be misleading**. The same applies to SM parameters.

Joint PDF-SMEFT fits?

- **This could lead to inconsistencies.**

PDF parameter fits

$$\mathrm{PDF}(\theta^*) \equiv \mathrm{PDF}(\theta^*(\bar{c}))$$

- Fitted PDFs can depend implicitly on fixed SMEFT parameters used in the fit.

SMEFT parameter fits

$$c^* \equiv c^*(\bar{\theta})$$

- Bounds on SMEFT parameters can depend implicitly on the fixed PDF set used in the fit.

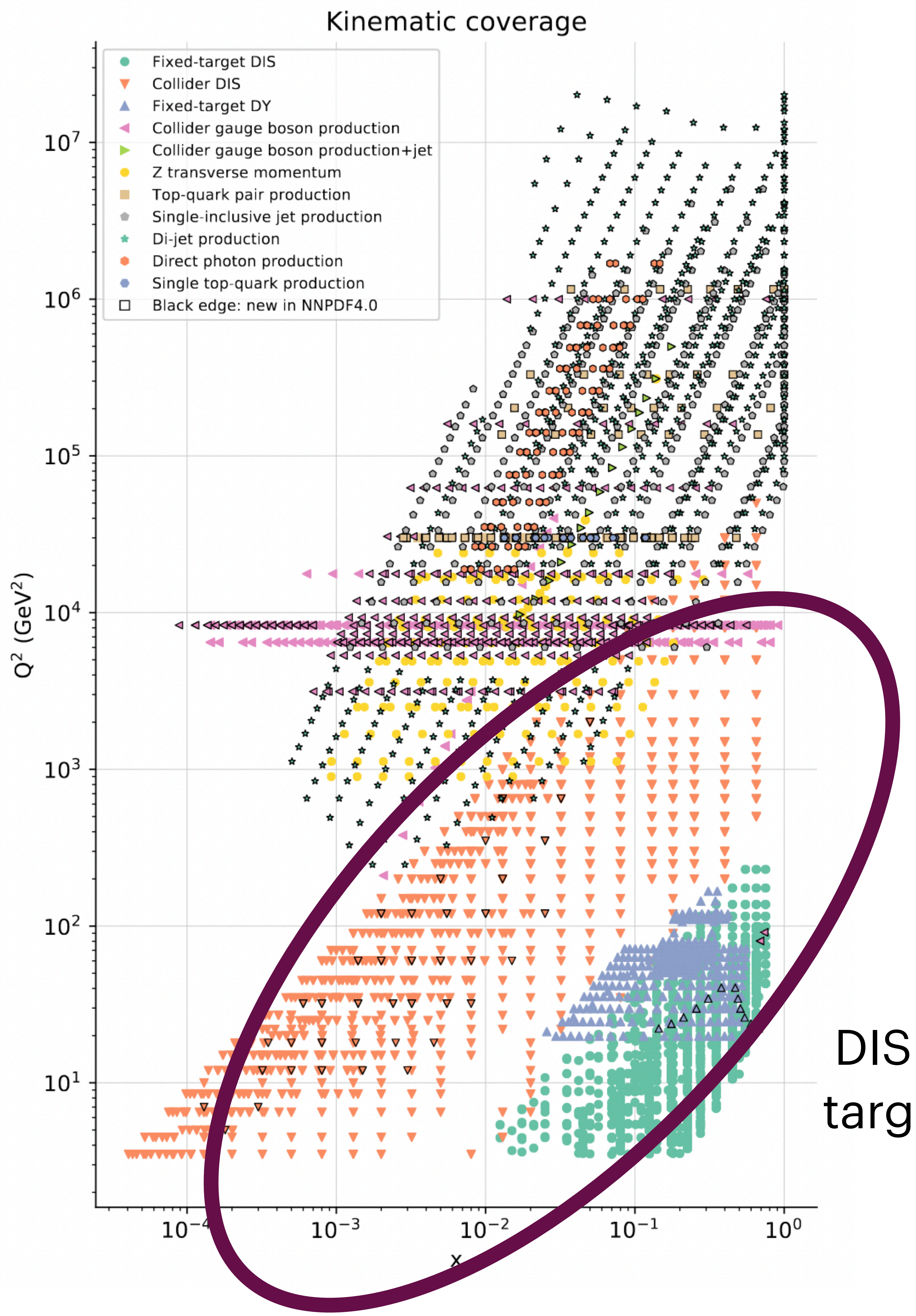
- In particular, if we fit PDFs **assuming all SMEFT couplings are zero**, but then **use those PDFs in a fit of SMEFT couplings**, our resulting bounds **could be misleading**. The same applies to SM parameters.
- We could even **miss New Physics**, or **see New Physics that isn't really there**!

PDF-SMEFT interplay: natural questions

- *Question 1:* **Can't I just use PDF sets which are fitted using data that is not affected by SMEFT operators?**

PDF-SMEFT interplay: natural questions

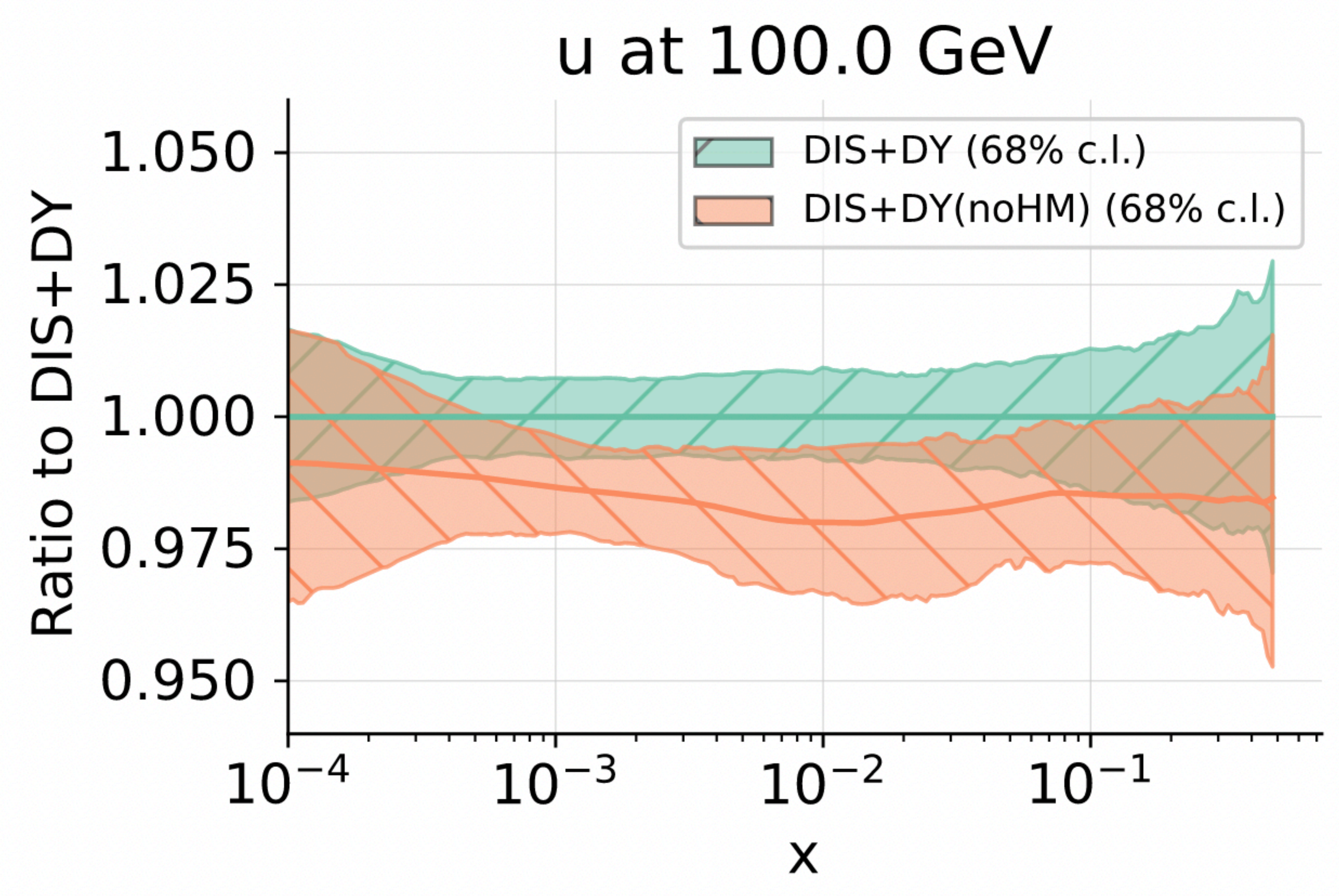
- *Question 1*: **Can't I just use PDF sets which are fitted using data that is not affected by SMEFT operators?**
  - It depends on the SMEFT operators. Some operators (e.g. four-fermion operators) will **contaminate DIS and DY data**, which comprise the majority of the data going into PDF fits. So often '*uncontaminated PDFs*' *don't exist*!
  - Right: kinematic coverage of NNPDF4.0 by dataset.

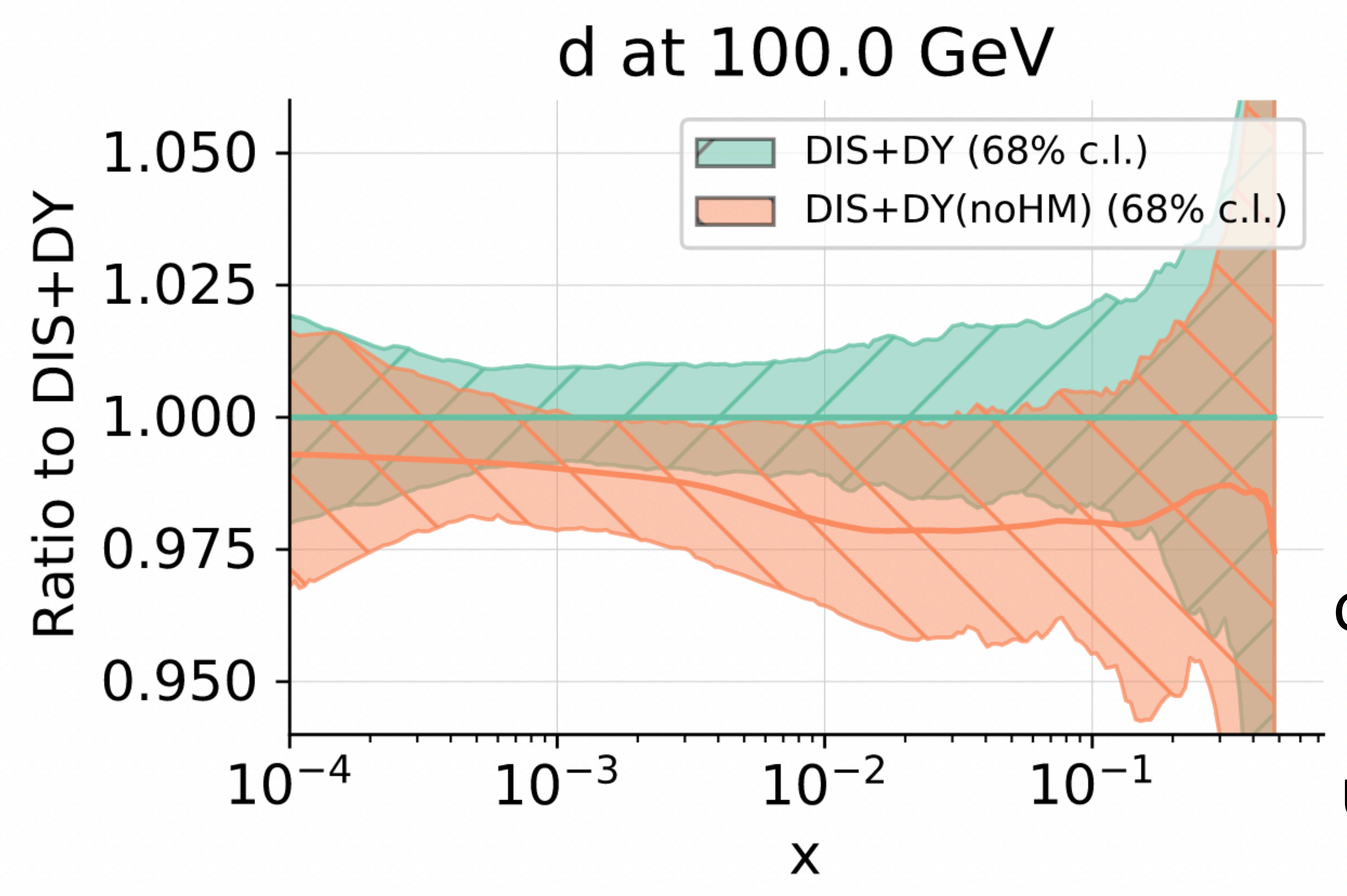


DIS and fixed target DY data

PDF-SMEFT interplay: natural questions

- *Question 1*: **Can't I just use PDF sets which are fitted using data that is not affected by SMEFT operators?**
  - Furthermore, if we include more data in a PDF fit, we obtain **better quality fits**. Therefore, we expect that using 'uncontaminated PDFs' will result in **poorer quality SMEFT fits**; we won't be using the 'best quality' PDFs that are available - this is shown explicitly in *Greljo et al., 2104.02723*, where PDF sets including and excluding high-mass DY data are compared.





without certain DY data points, high-x uncertainty increases

PDF-SMEFT interplay: natural questions

- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**

PDF-SMEFT interplay: natural questions

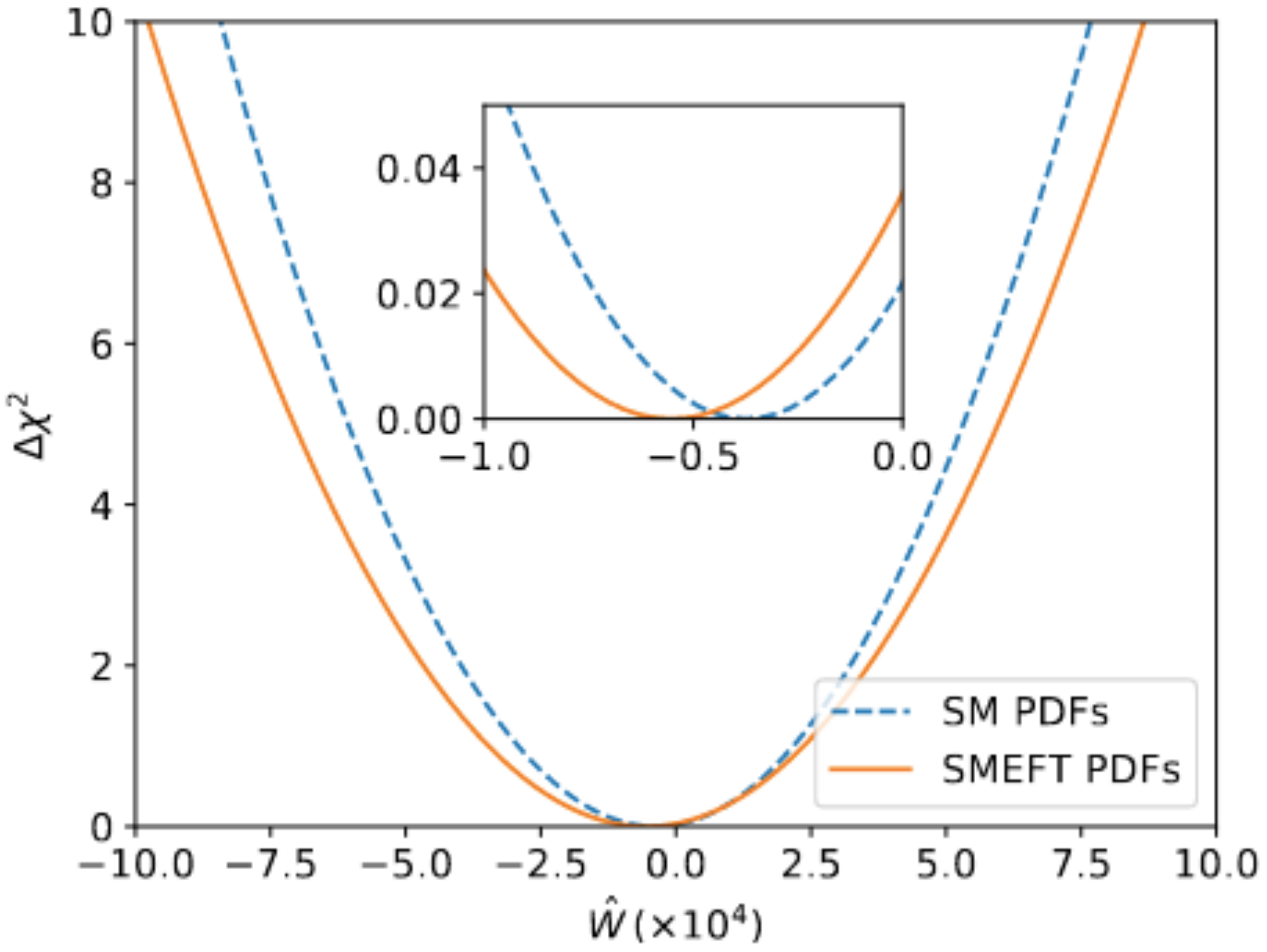
- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**
  - It depends on the scenario!

PDF-SMEFT interplay: natural questions

- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**
  - It depends on the scenario!
  - It was shown in *Carrazza et al., 1905.05215*, that interplay is very mild in the case of simultaneous extractions of four-fermion operators and PDFs using DIS-only data.

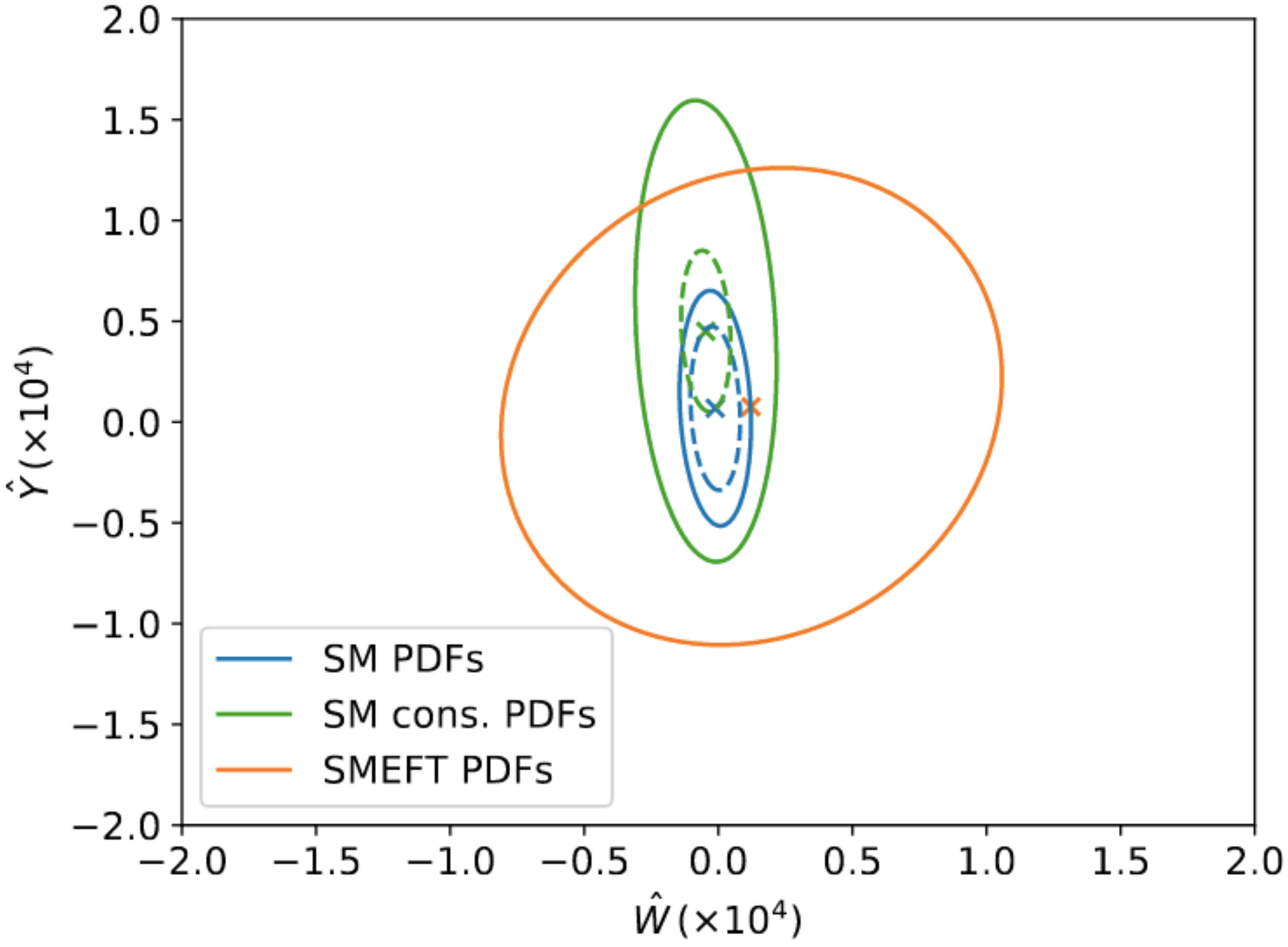
PDF-SMEFT interplay: natural questions

- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**
  - It depends on the scenario!
  - It was shown in *Carrazza et al., 1905.05215*, that interplay is very mild in the case of simultaneous extractions of four-fermion operators and PDFs using DIS-only data.
  - Similarly, it was shown in the PBSP team's earlier study, *Greljo et al., 2104.02723*, that interplay is mild between the $\hat{W}, \hat{Y}$ operators and PDFs using current DIS and DY data.



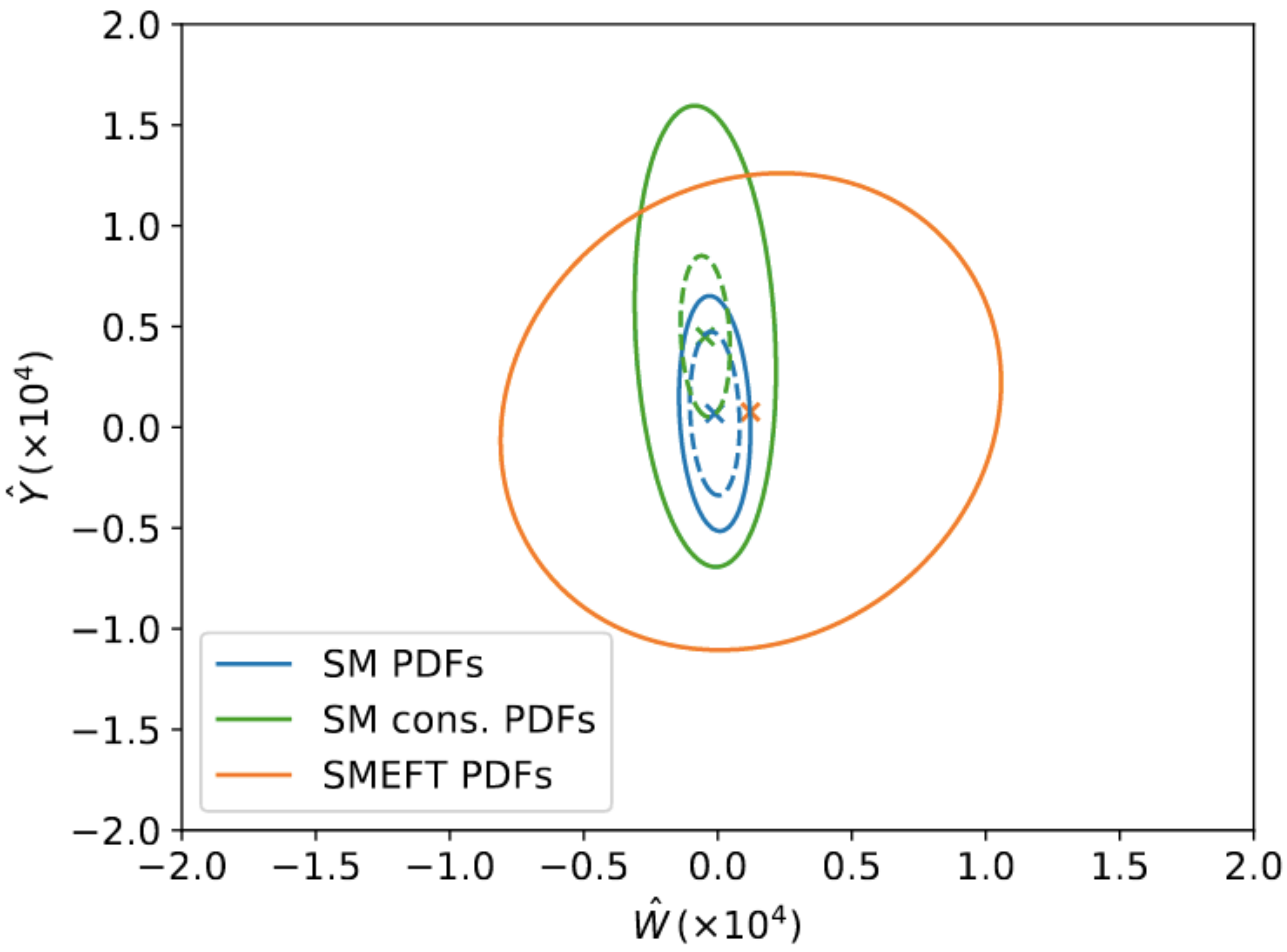
PDF-SMEFT interplay: natural questions

- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**
  - *However,* it was also shown in *Greljo et al., 2104.02723*, that interplay is **very significant** between the $\hat{W}$, $\hat{Y}$ operators and PDFs using **projected high-luminosity DY data**.



PDF-SMEFT interplay: natural questions

- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**
  - *However,* it was also shown in *Greljo et al., 2104.02723*, that interplay is **very significant** between the $\hat{W}$, $\hat{Y}$ operators and PDFs using **projected high-luminosity DY data**.
  - We see that using fixed PDFs results in a **significant underestimation** of uncertainties on the WCs - we might wrongly conclude **New Physics**!



4. - The SIMUnet methodology for joint PDF-SMEFT fits

PDF-SMEFT interplay: methodology

- With the need for simultaneous PDF-SMEFT determinations established, we now need an **efficient methodology** to perform the fits.

PDF-SMEFT interplay: methodology

- With the need for simultaneous PDF-SMEFT determinations established, we now need an **efficient methodology** to perform the fits.
- Three main methodologies available:

PDF-SMEFT interplay: methodology

- With the need for simultaneous PDF-SMEFT determinations established, we now need an **efficient methodology** to perform the fits.
- Three main methodologies available:

1. 'Scan' methodology

- Select a grid of benchmark SMEFT points.
- Perform PDF fits at each benchmark point.
- Construct a χ^2-surface and obtain bounds.

See 1905.05215 and 2104.02723

PDF-SMEFT interplay: methodology

- With the need for simultaneous PDF-SMEFT determinations established, we now need an **efficient methodology** to perform the fits.
- Three main methodologies available:

1. 'Scan' methodology

- Select a grid of benchmark SMEFT points.
- Perform PDF fits at each benchmark point.
- Construct a χ^2-surface and obtain bounds.

See 1905.05215 and 2104.02723

2. CTEQ-TEA methodology

- Model the χ^2-surface as a neural network, with inputs given by PDF parameters and WCs.
- After training the network, use Lagrange multiplier scans to minimise χ^2.

See 2201.06586 and 2211.01094

PDF-SMEFT interplay: methodology

- With the need for simultaneous PDF-SMEFT determinations established, we now need an **efficient methodology** to perform the fits.
- Three main methodologies available:

1. 'Scan' methodology

- Select a grid of benchmark SMEFT points.
- Perform PDF fits at each benchmark point.
- Construct a χ^2-surface and obtain bounds.

See 1905.05215 and 2104.02723

2. CTEQ-TEA methodology

- Model the χ^2-surface as a neural network, with inputs given by PDF parameters and WCs.
- After training the network, use Lagrange multiplier scans to minimise χ^2.

See 2201.06586 and 2211.01094

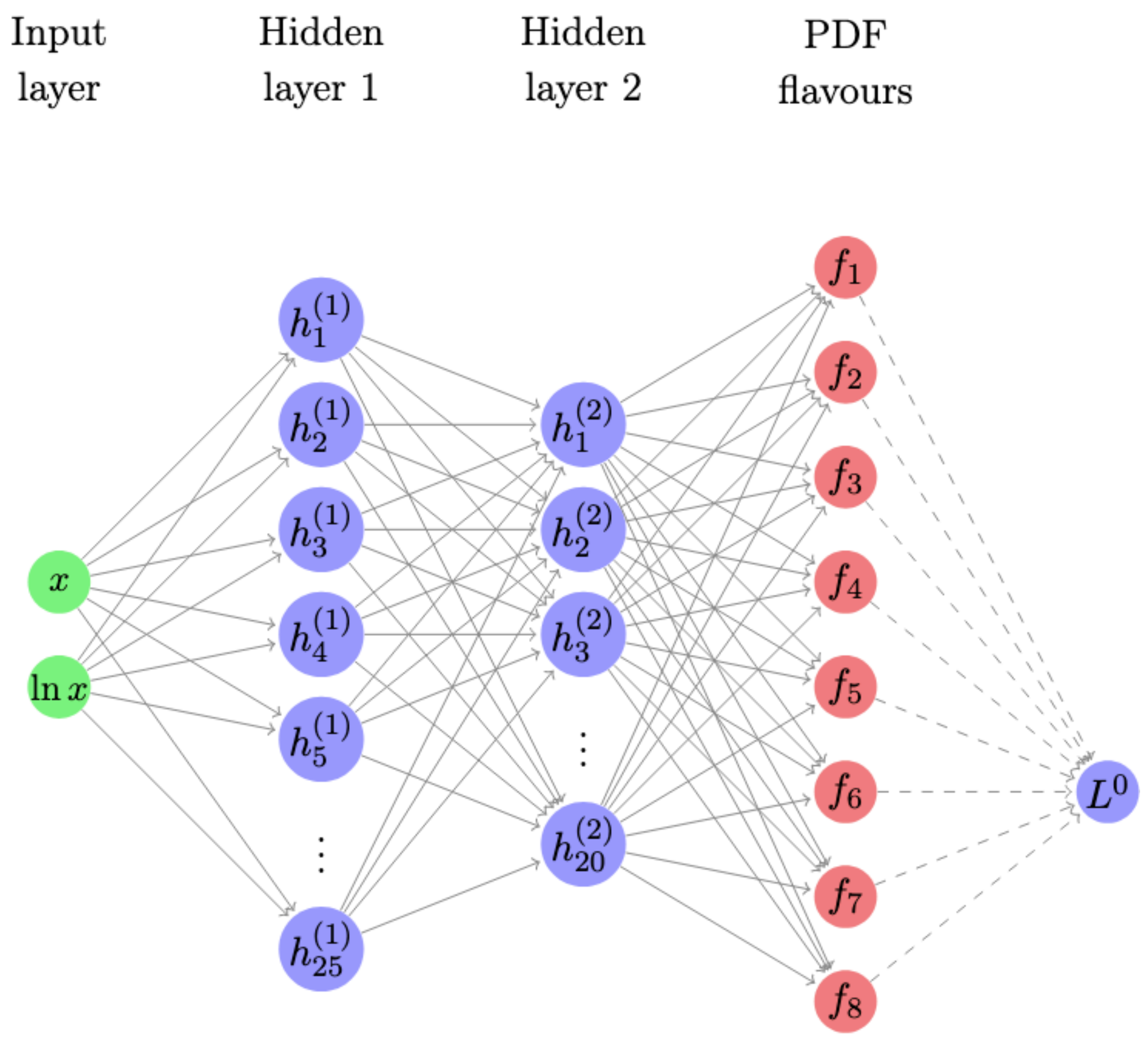
3. SIMUnet methodology

- Extend the NNPDF replica networks with a new layer with edges corresponding to the WCs.
- Train the network as per an NNPDF fit, but also learning the WCs.

See 2201.07240

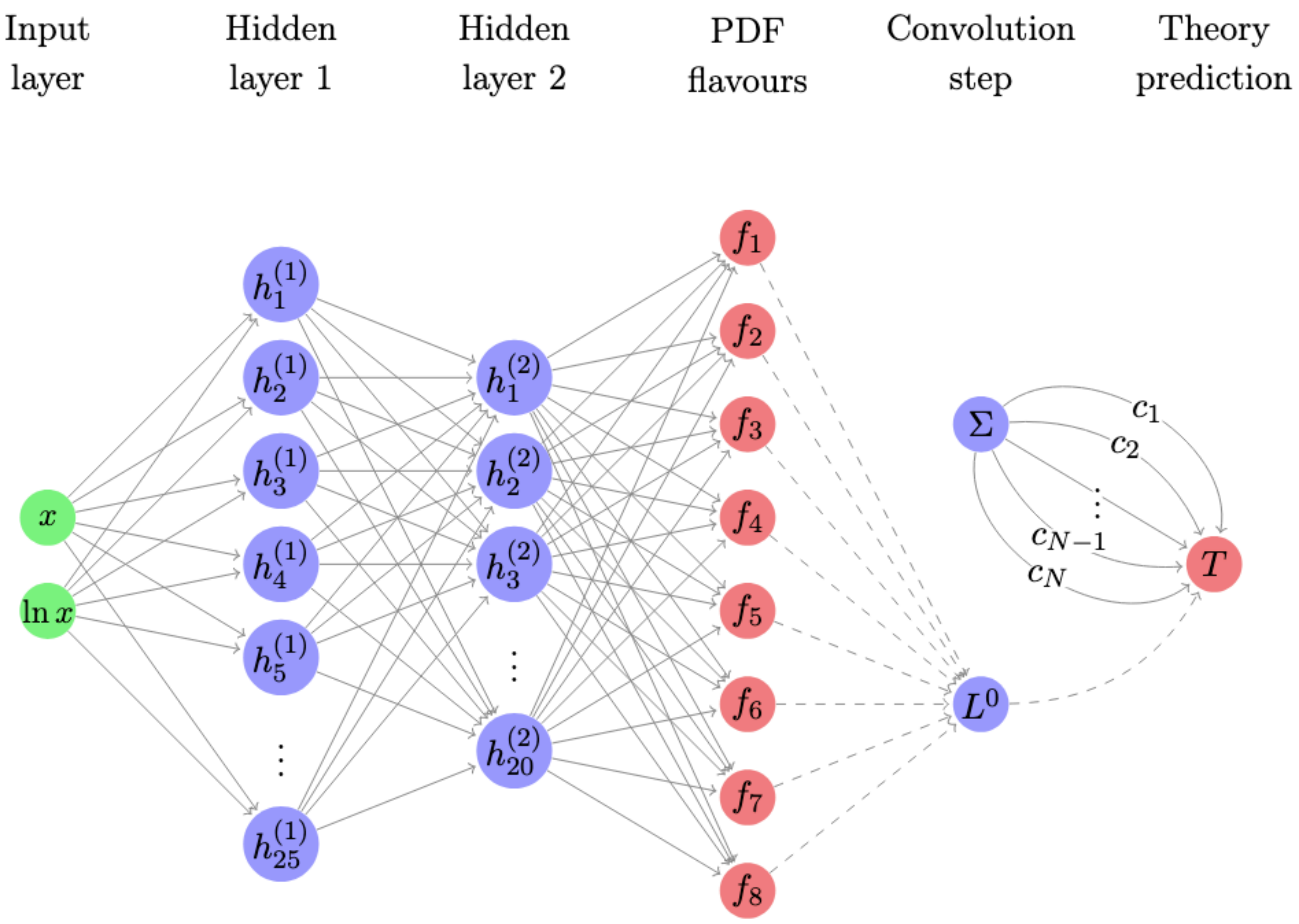
The SIMUnet methodology: details

- The SIMUnet methodology **extends the existing NNPDF neural network** with an additional **convolution layer**.



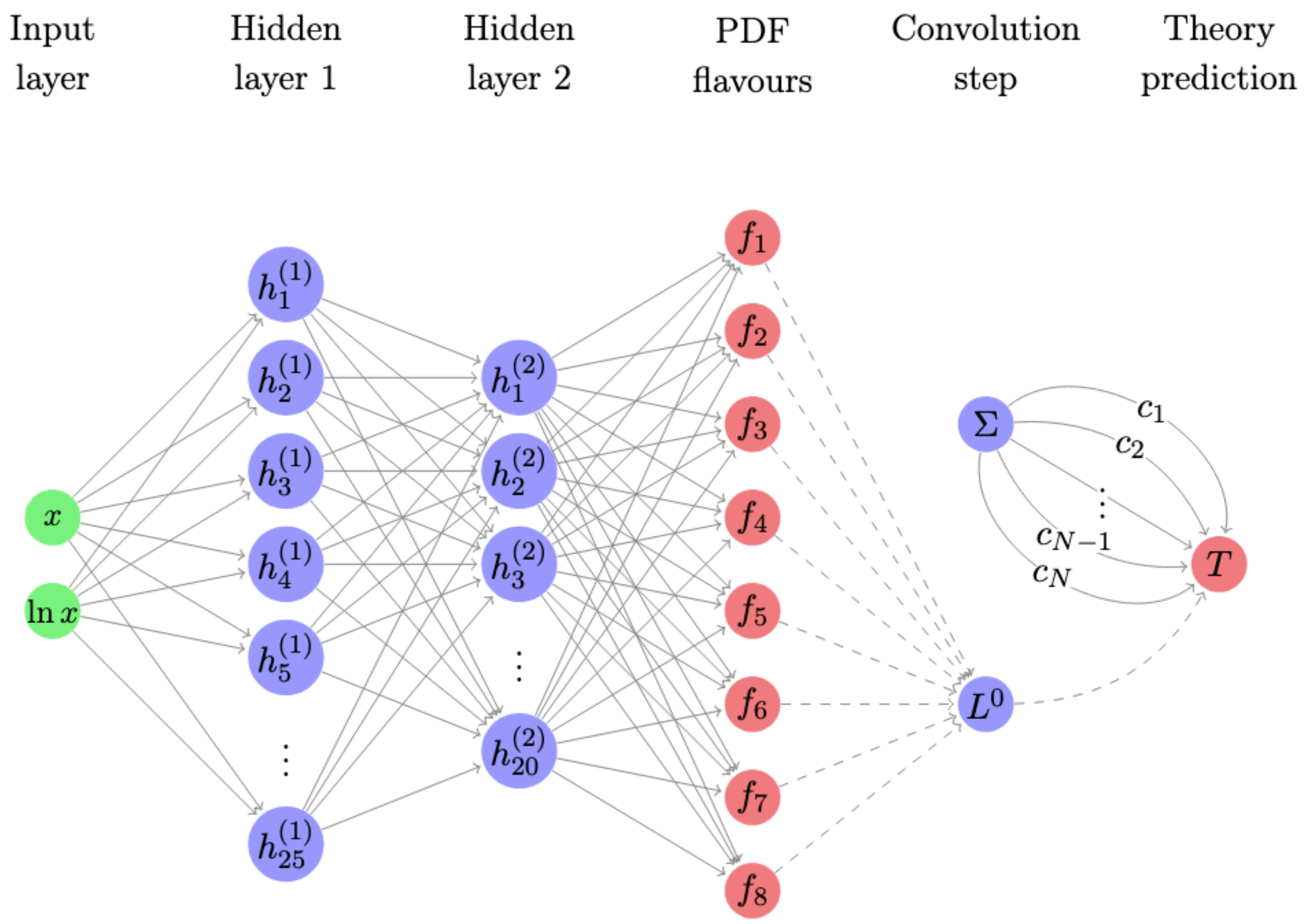
The SIMUnet methodology: details

- The SIMUnet methodology **extends the existing NNPDF neural network** with an additional **convolution layer**.
- The SMEFT couplings are added as **weights of neural network edges**, and are **trained alongside the PDFs**.



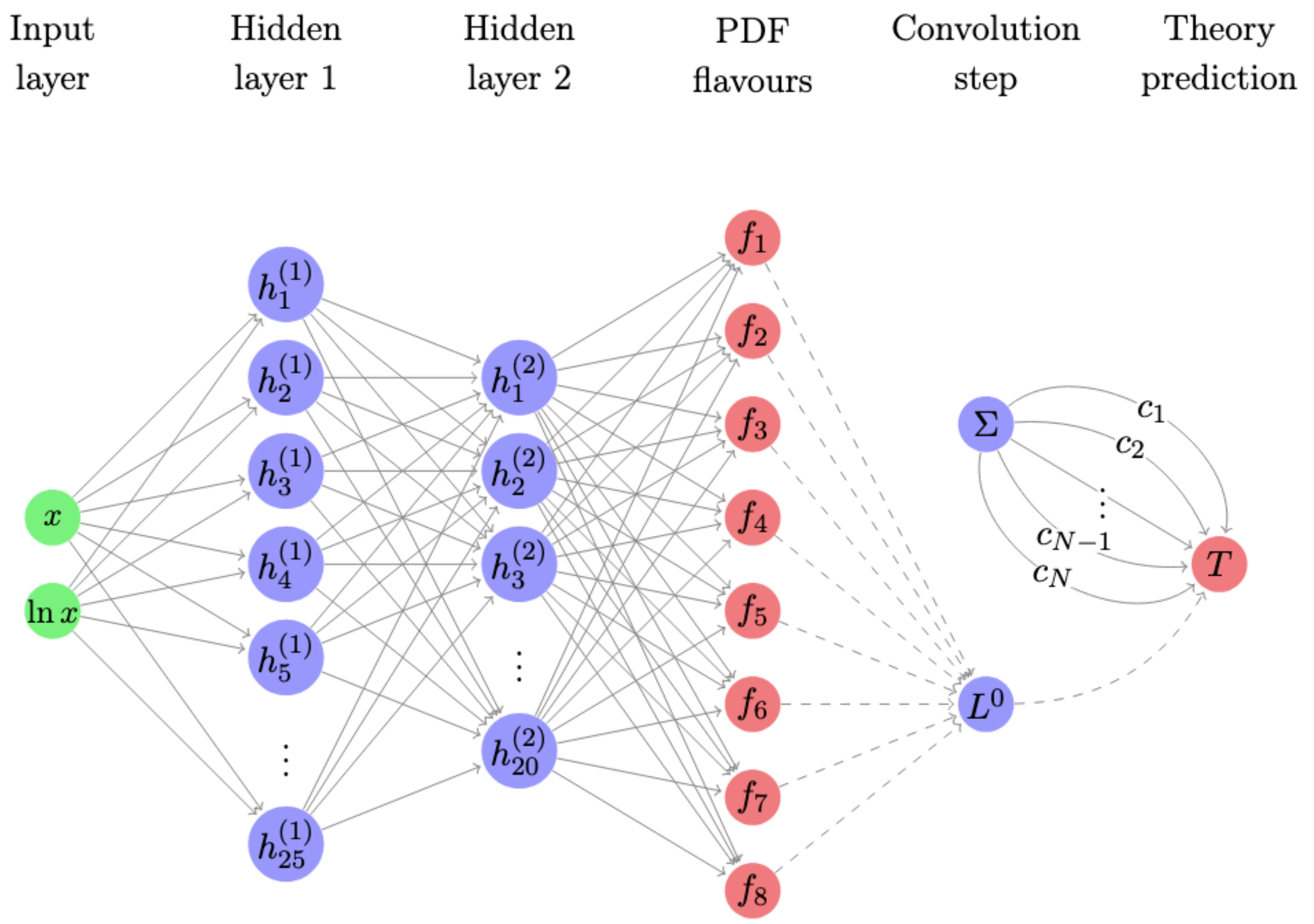
The SIMUnet methodology: details

- The SIMUnet methodology allows for **a lot of flexibility**:



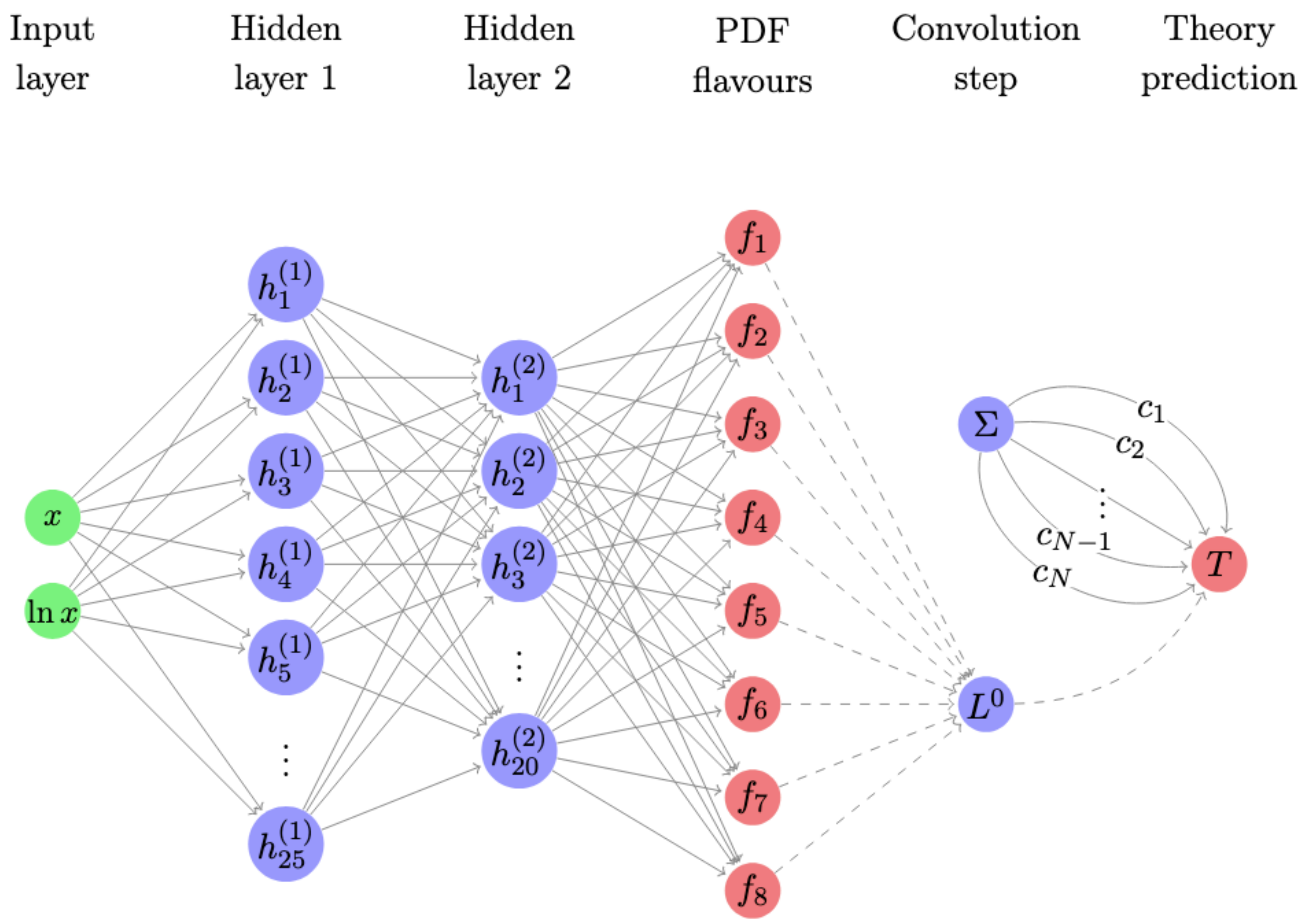
The SIMUnet methodology: details

- The SIMUnet methodology allows for **a lot of flexibility**:
  - Can include **quadratic*** SMEFT corrections through **non-trainable edges**.



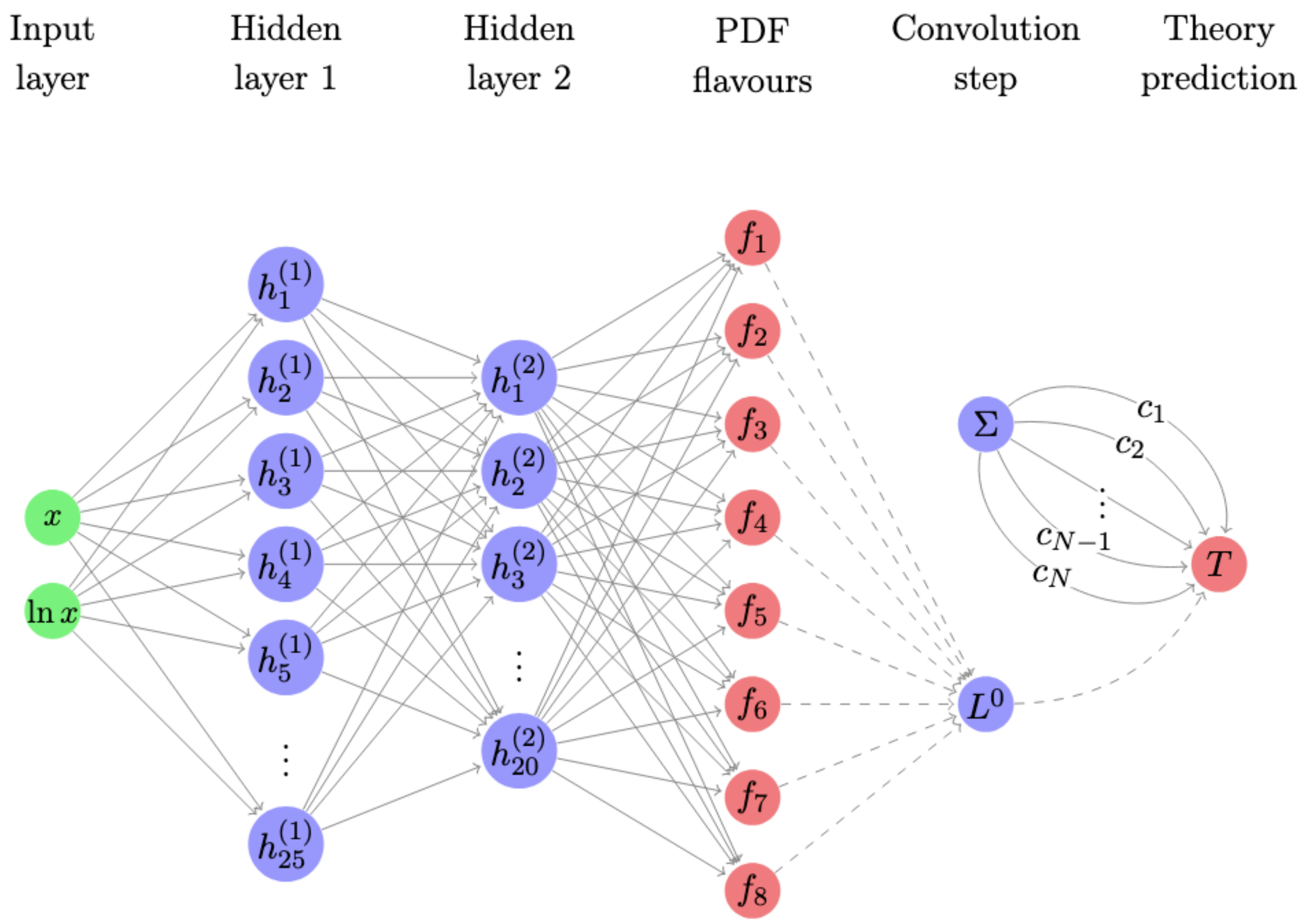
The SIMUnet methodology: details

- The SIMUnet methodology allows for **a lot of flexibility**:
  - Can include **quadratic*** SMEFT corrections through **non-trainable edges**.
  - Can easily include **PDF-independent observables**.



The SIMUnet methodology: details

- The SIMUnet methodology allows for **a lot of flexibility**:
  - Can include **quadratic*** SMEFT corrections through **non-trainable edges**.
  - Can easily include **PDF-independent observables**.
  - Can perform **fixed PDF fits** by **freezing the PDF part of the network**.



5. - The top quark legacy of the LHC Run II for PDF and SMEFT analyses

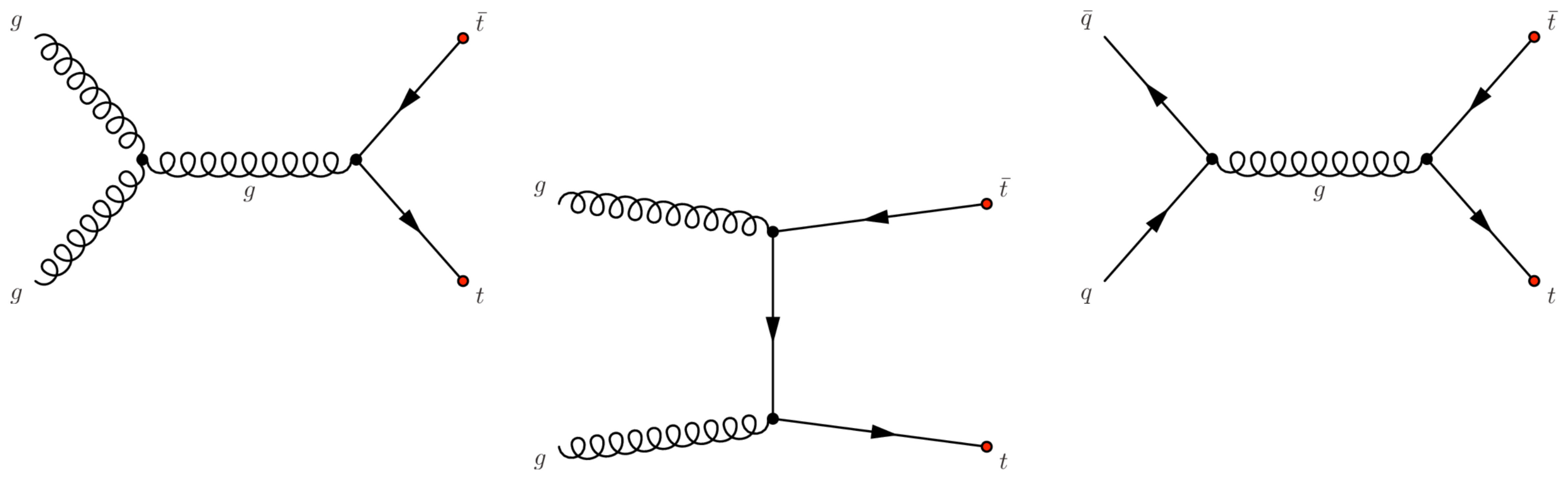
Based on 2303.06159

Run II top quark data

- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:

Run II top quark data

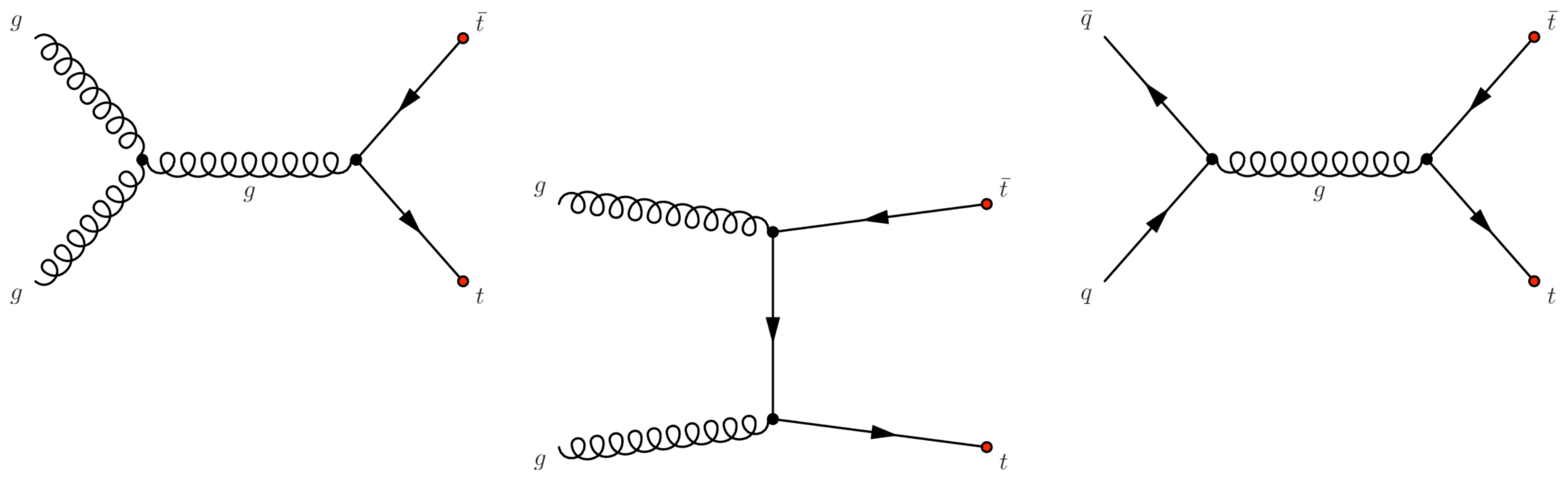
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



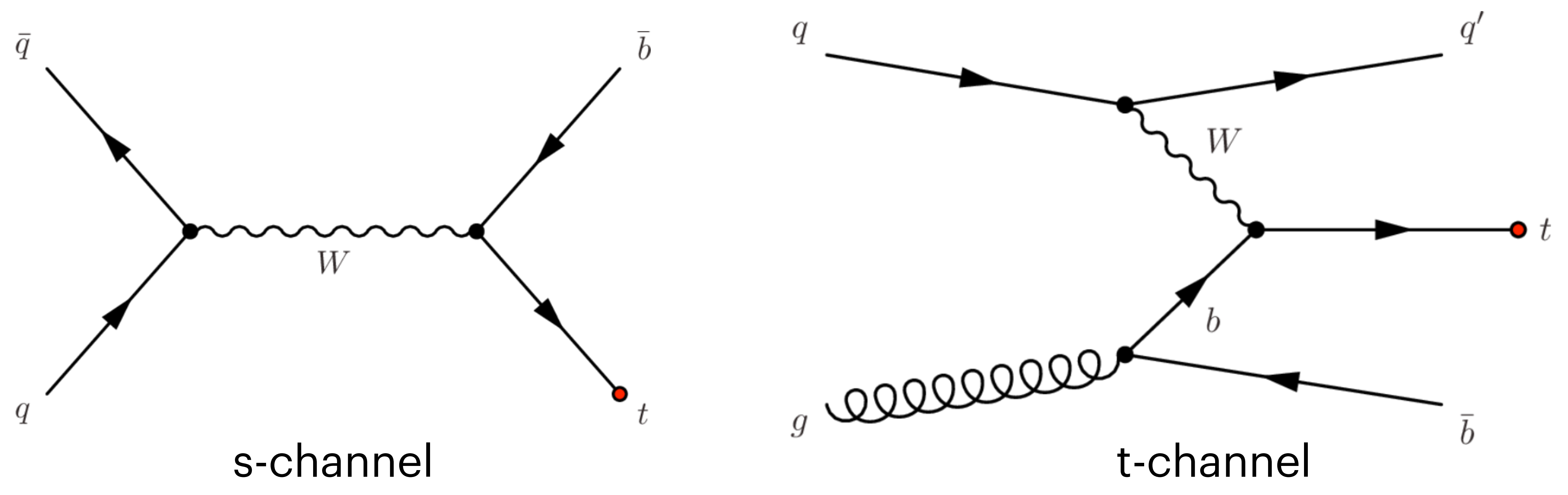
inclusive $t\bar{t}$

Run II top quark data

- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



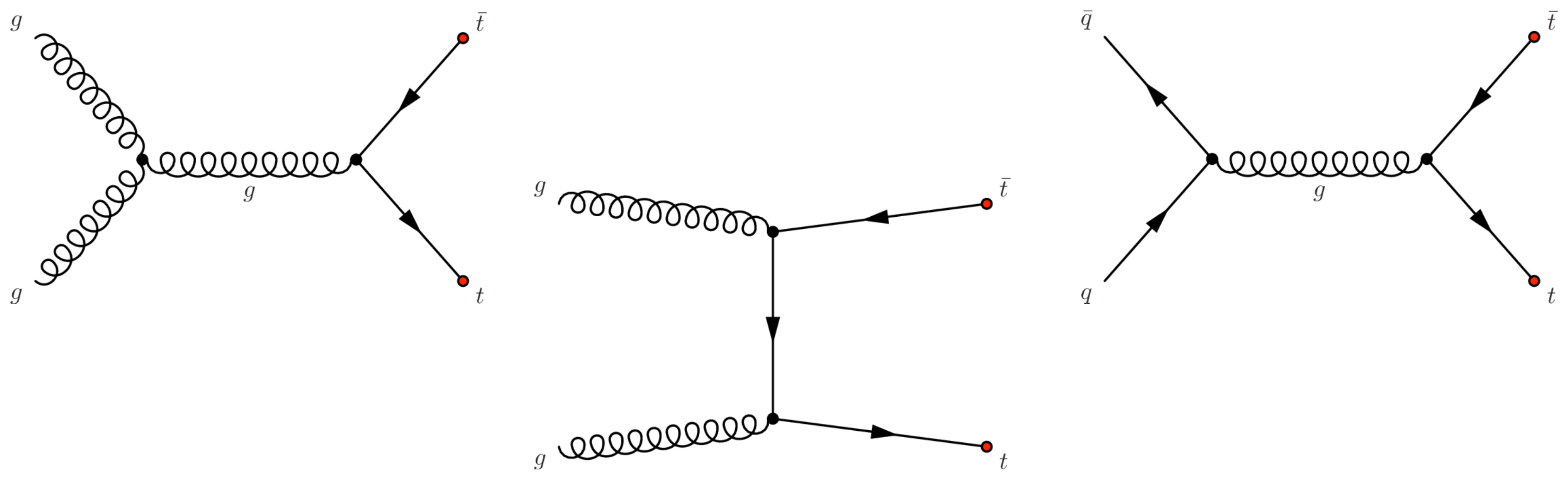
inclusive $t\bar{t}$



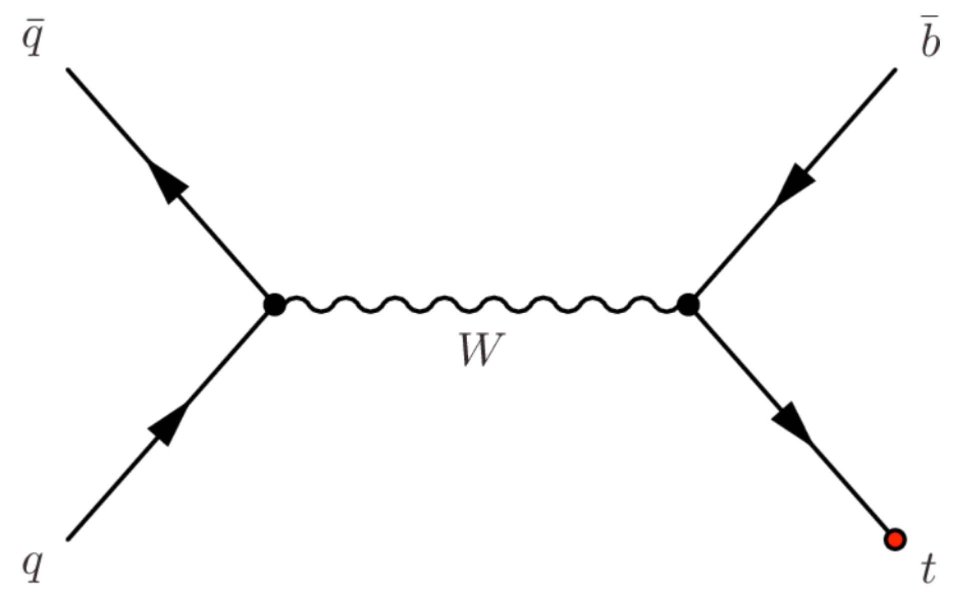
single top

Run II top quark data

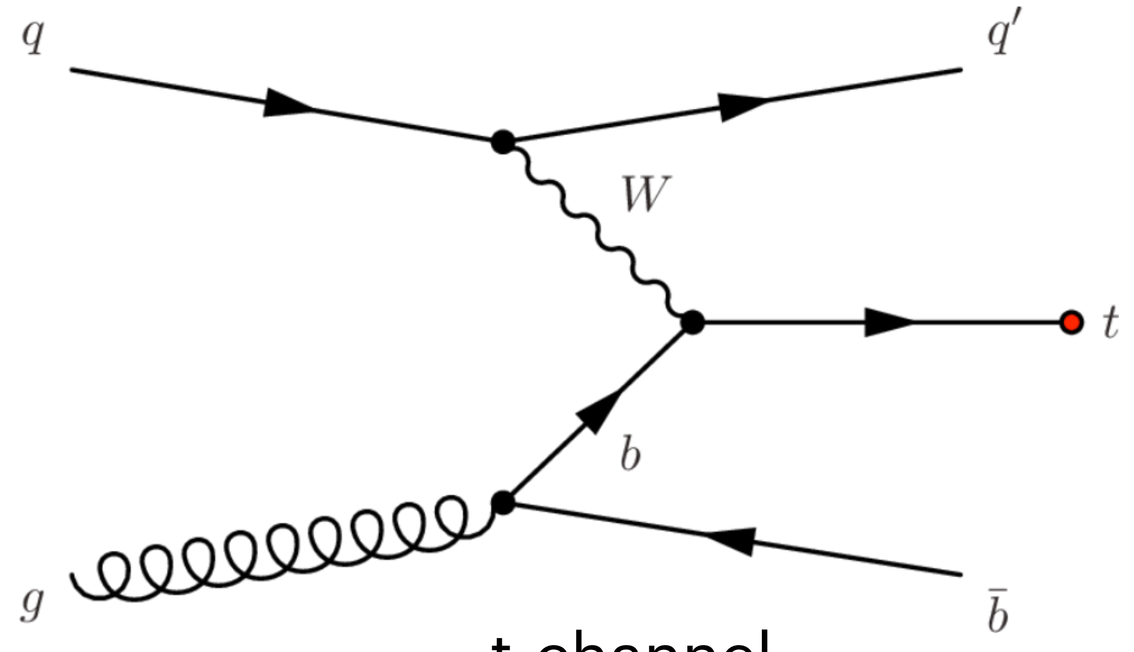
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



inclusive $t\bar{t}$

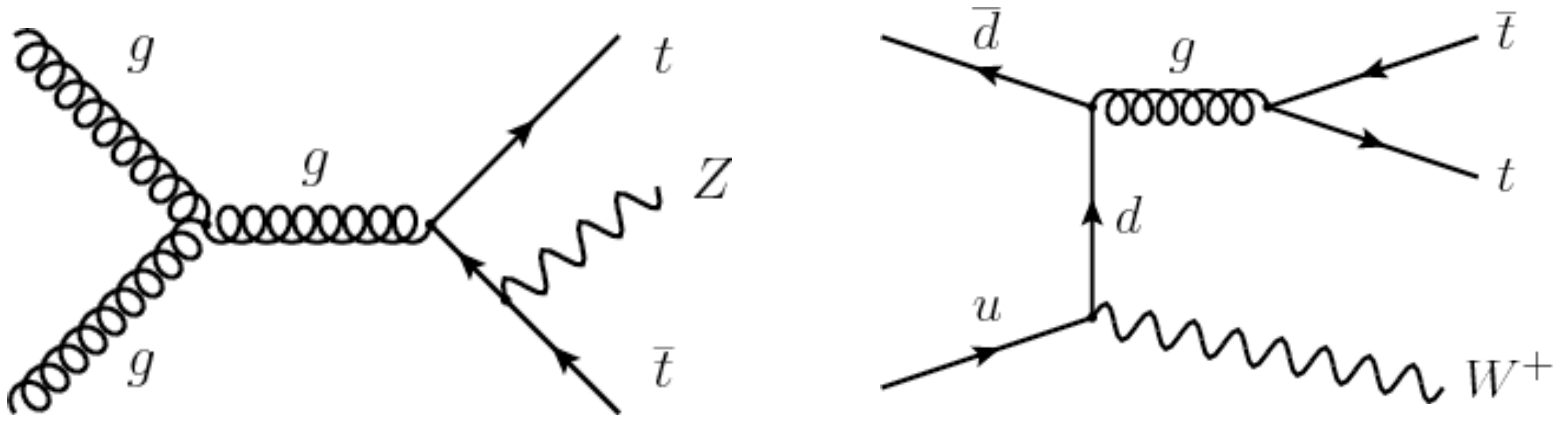


s-channel



t-channel

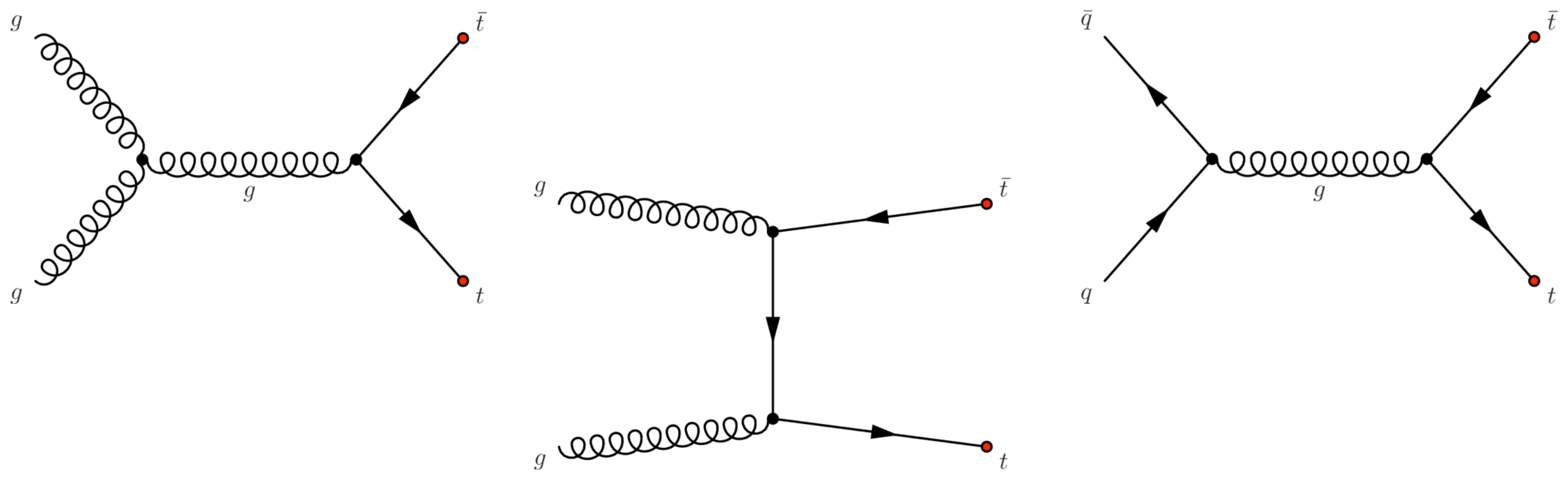
single top



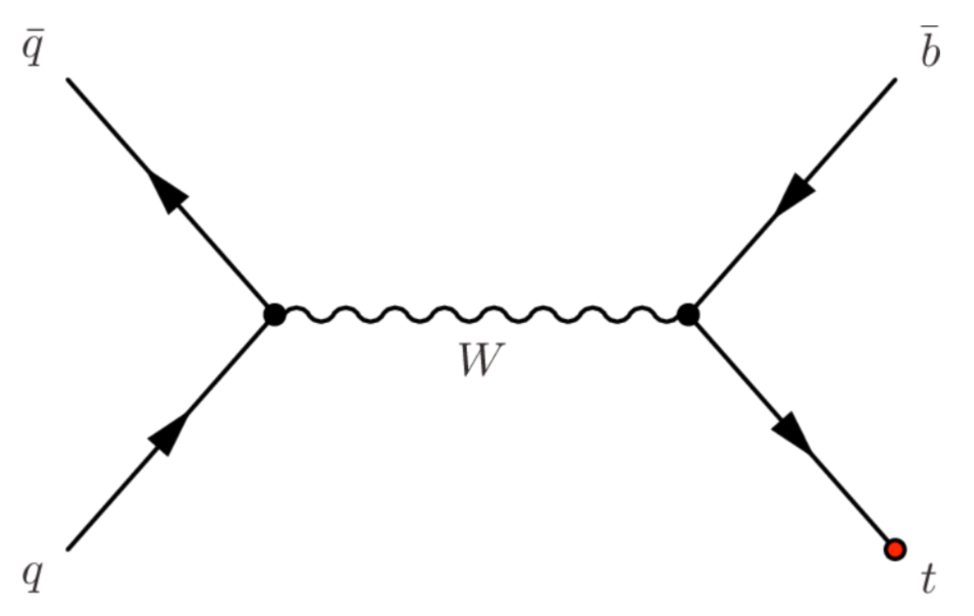
associated $t\bar{t}$

Run II top quark data

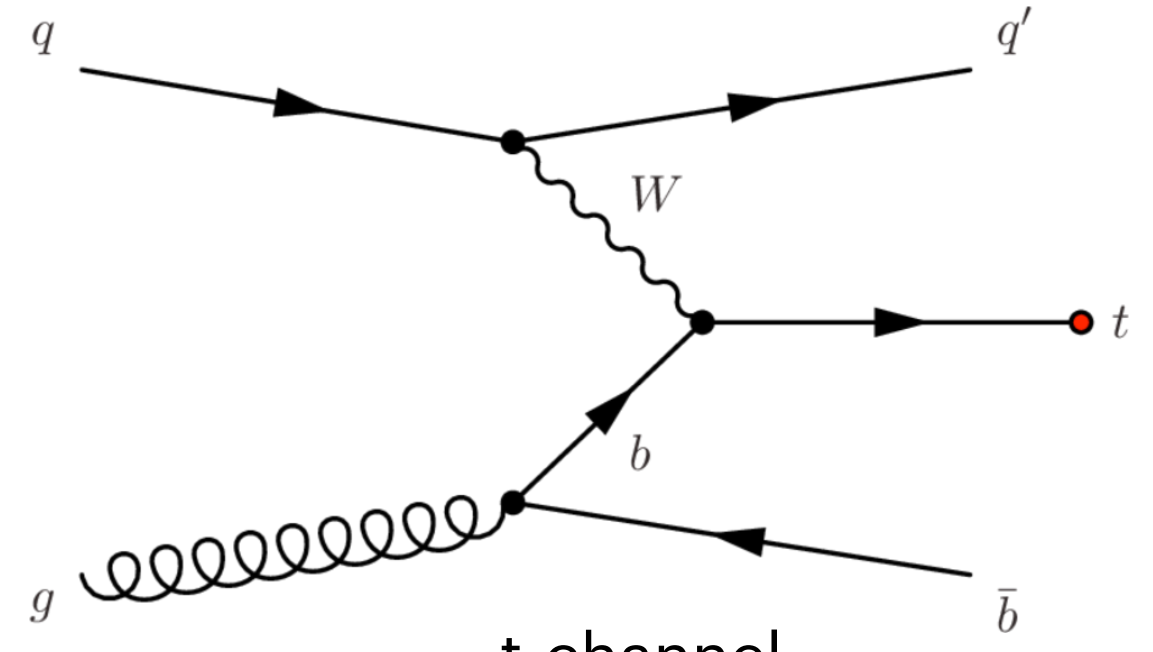
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



inclusive $t\bar{t}$

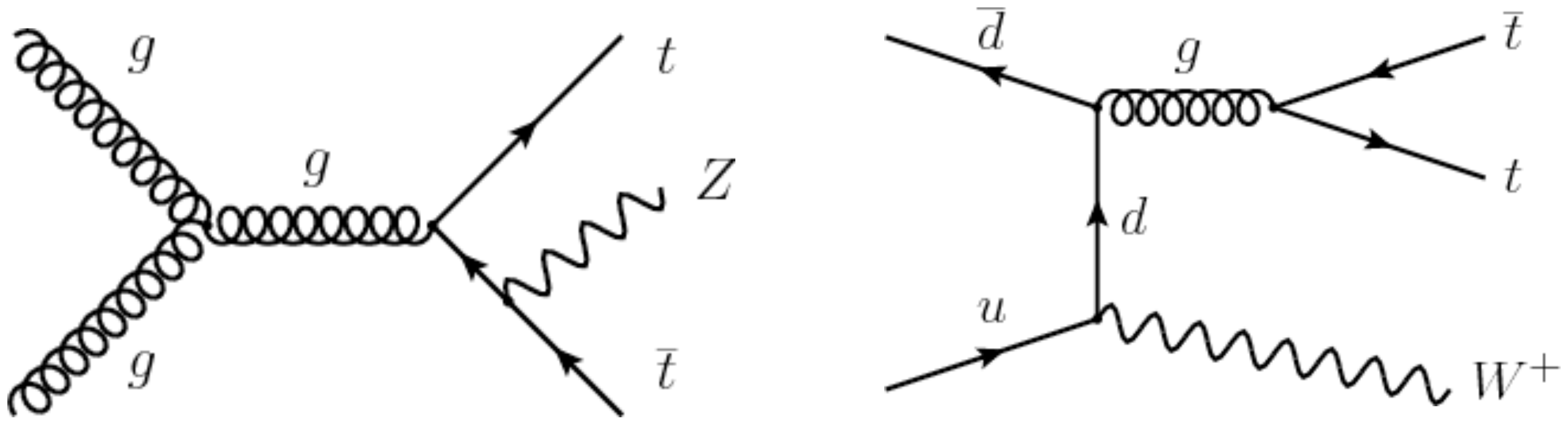


s-channel

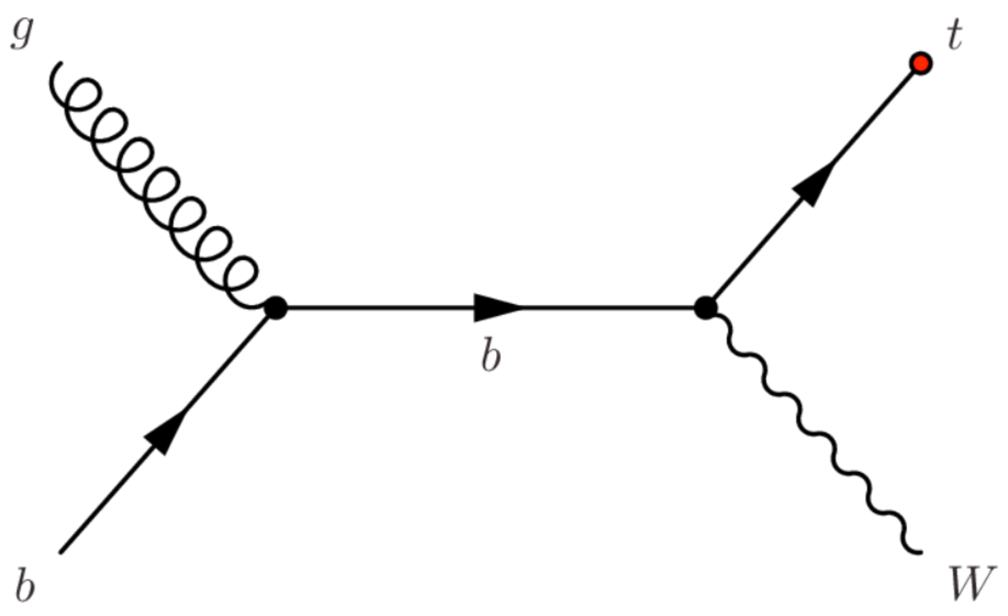


t-channel

single top



associated $t\bar{t}$



associated single top

Run II top quark data

- Currently, both $t\bar{t}$ and single-t data are **included in PDF fits**. But predictions for these processes are **also** impacted by **SMEFT operators**:

Run II top quark data

- Currently, both $t\bar{t}$ and single-t data are **included in PDF fits**. But predictions for these processes are **also** impacted by **SMEFT operators**:

c_{tZ} c_{tW}

c_{tG}

dipoles

Run II top quark data

- Currently, both $t\bar{t}$ and single-t data are **included in PDF fits**. But predictions for these processes are **also** impacted by **SMEFT operators**:

dipoles	currents
c_{tZ}, c_{tW}, c_{tG}	$c_{\phi t}$, $c_{\phi Q}^{(1)}$, $c_{\phi Q}^{(3)}$

Run II top quark data

- Currently, both $t\bar{t}$ and single-t data are **included in PDF fits**. But predictions for these processes are **also** impacted by **SMEFT operators**:

c_{tZ} c_{tW} c_{tG}

dipoles

$c_{\phi t}$ $c_{\phi Q}^{(3)}$ $c_{\phi Q}^{(1)}$

currents

c_{qd}^{1} $c_{qq}^{1,3}$ c_{dt}^{1} c_{qt}^{1} $c_{qq}^{1,1}$ c_{qu}^{1} c_{ut}^{1}

four-fermion singlets

Run II top quark data

- Currently, both $t\bar{t}$ and single-t data are **included in PDF fits**. But predictions for these processes are **also** impacted by **SMEFT operators**:

dipoles	currents
c_{tZ} c_{tW} c_{tG}	$c_{\phi t}$ $c_{\phi Q}^{(1)}$ $c_{\phi Q}^{(3)}$

four-fermion singlets	four-fermion octets
c_{qd}^{1} $c_{qq}^{1,3}$ c_{dt}^{1} c_{qt}^{1} $c_{qq}^{1,1}$ c_{qu}^{1} c_{ut}^{1}	c_{qd}^{8} $c_{qq}^{8,3}$ c_{dt}^{8} c_{qt}^{8} $c_{qq}^{8,1}$ c_{qu}^{8} c_{ut}^{8}

Key questions for the rest of the talk:

1. How do WC bounds compare between fixed PDF EFT-fits and simultaneous fits?

Key questions for the rest of the talk:

1. How do WC bounds compare between fixed PDF EFT-fits and simultaneous fits?

2. How do PDFs compare between SM PDF fits and simultaneous PDF-EFT fits?

Fit settings

- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.

Fit settings

- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **175 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset of the measurements used in:

Fit settings

- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **175 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset of the measurements used in:
  - NNPDF4.0 (84 top data points, inclusive $t\bar{t}$ and single top only)

Fit settings

- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **175 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset of the measurements used in:
  - NNPDF4.0 (84 top data points, inclusive $t\bar{t}$ and single top only)
  - SMEFiT (143 top data points)

Fit settings

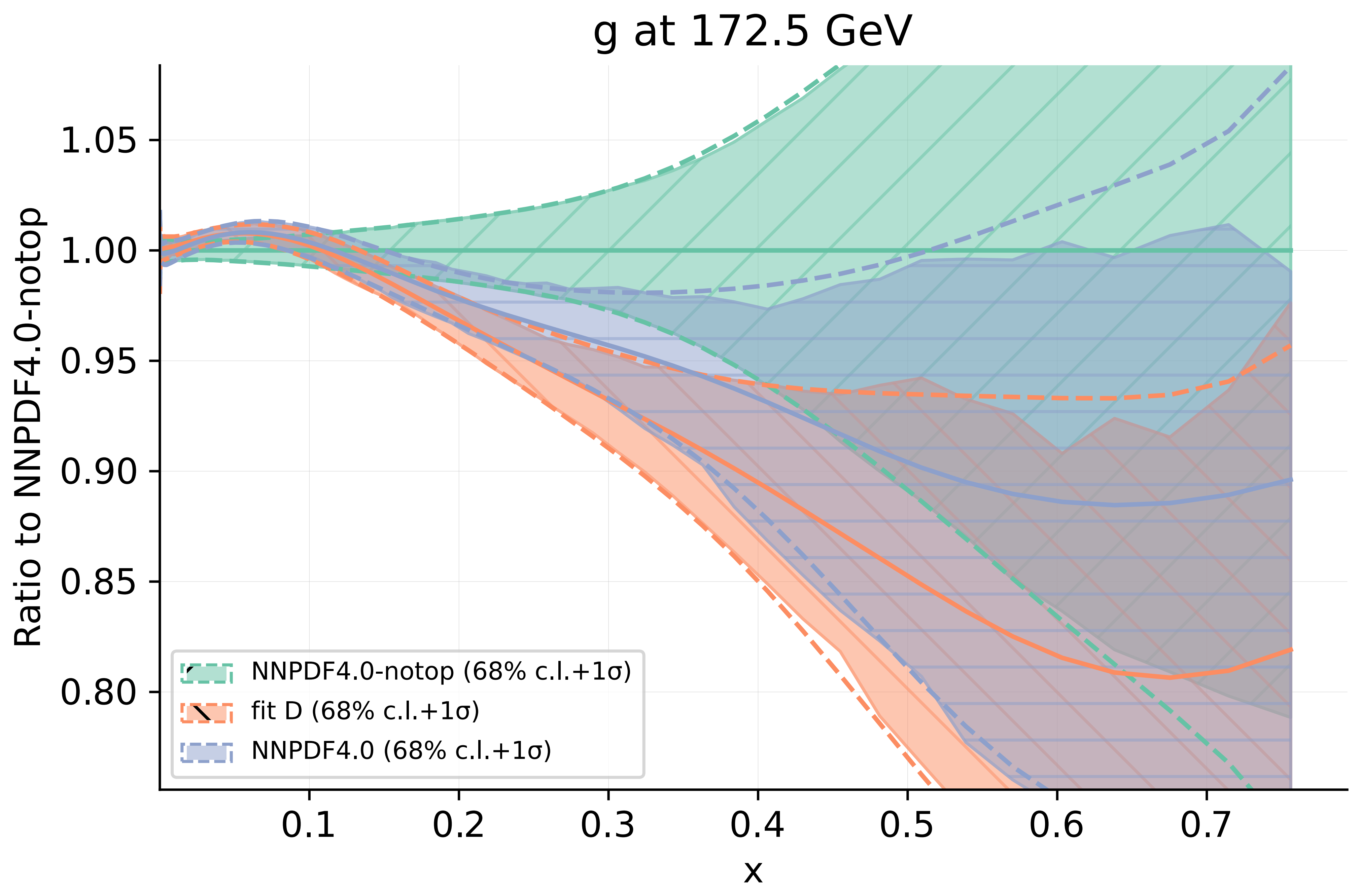
- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **175 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset of the measurements used in:
  - NNPDF4.0 (84 top data points, inclusive $t\bar{t}$ and single top only)
  - SMEFiT (143 top data points)
  - Fitmaker (137 top data points)

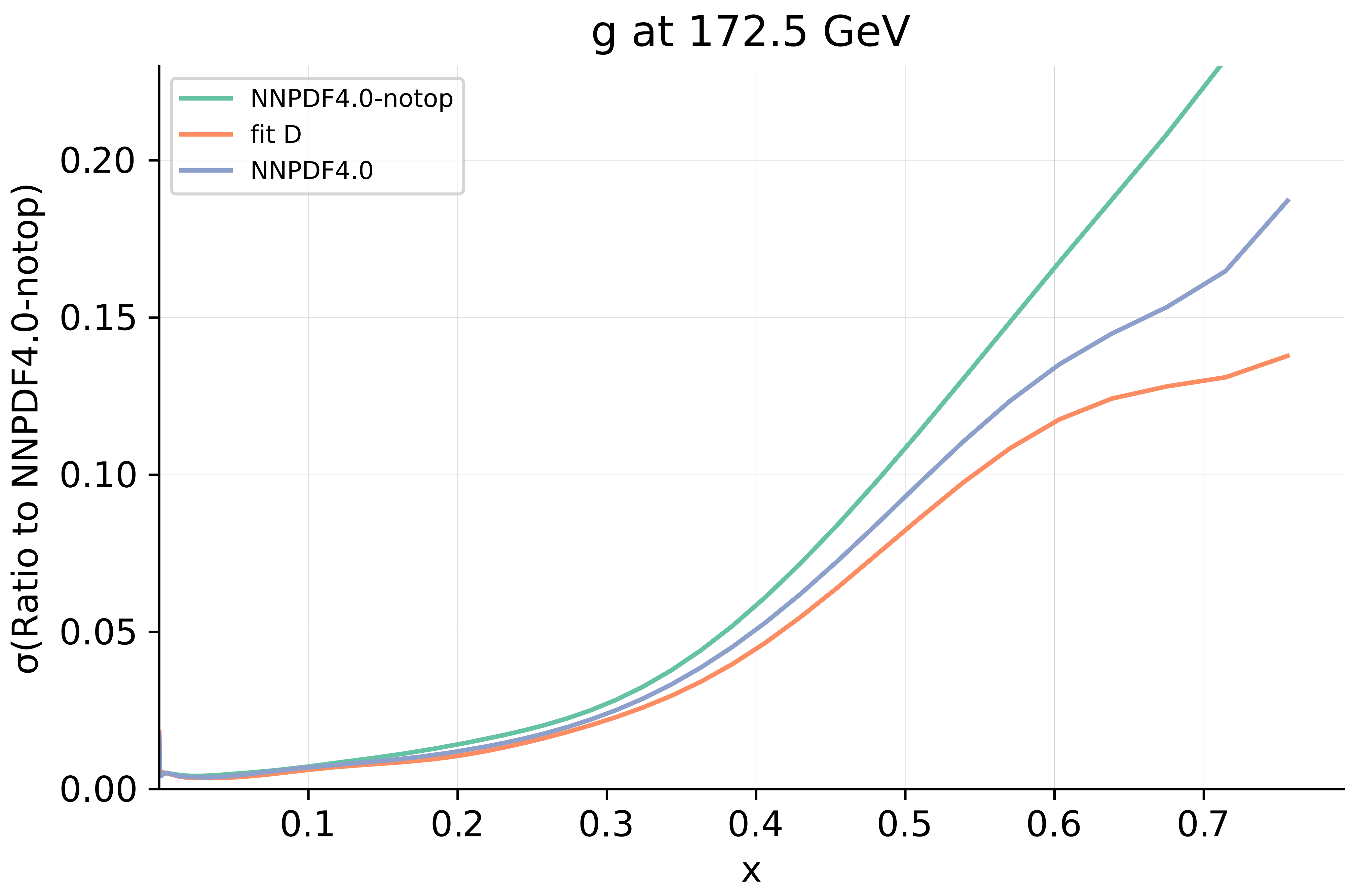
Fit settings

- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **175 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset of the measurements used in:
    - NNPDF4.0 (84 top data points, inclusive $t\bar{t}$ and single top only)
    - SMEFiT (143 top data points)
    - Fitmaker (137 top data points)
- We work with theory predictions accurate to **NNLO in QCD in the SM**, and include **NLO QCD in the SMEFT**. Some fits are **linear in the SMEFT**, some are **quadratic** - a point we will return to.

PDFs in the SM - impact of inclusive $t\bar{t}$ and single-top

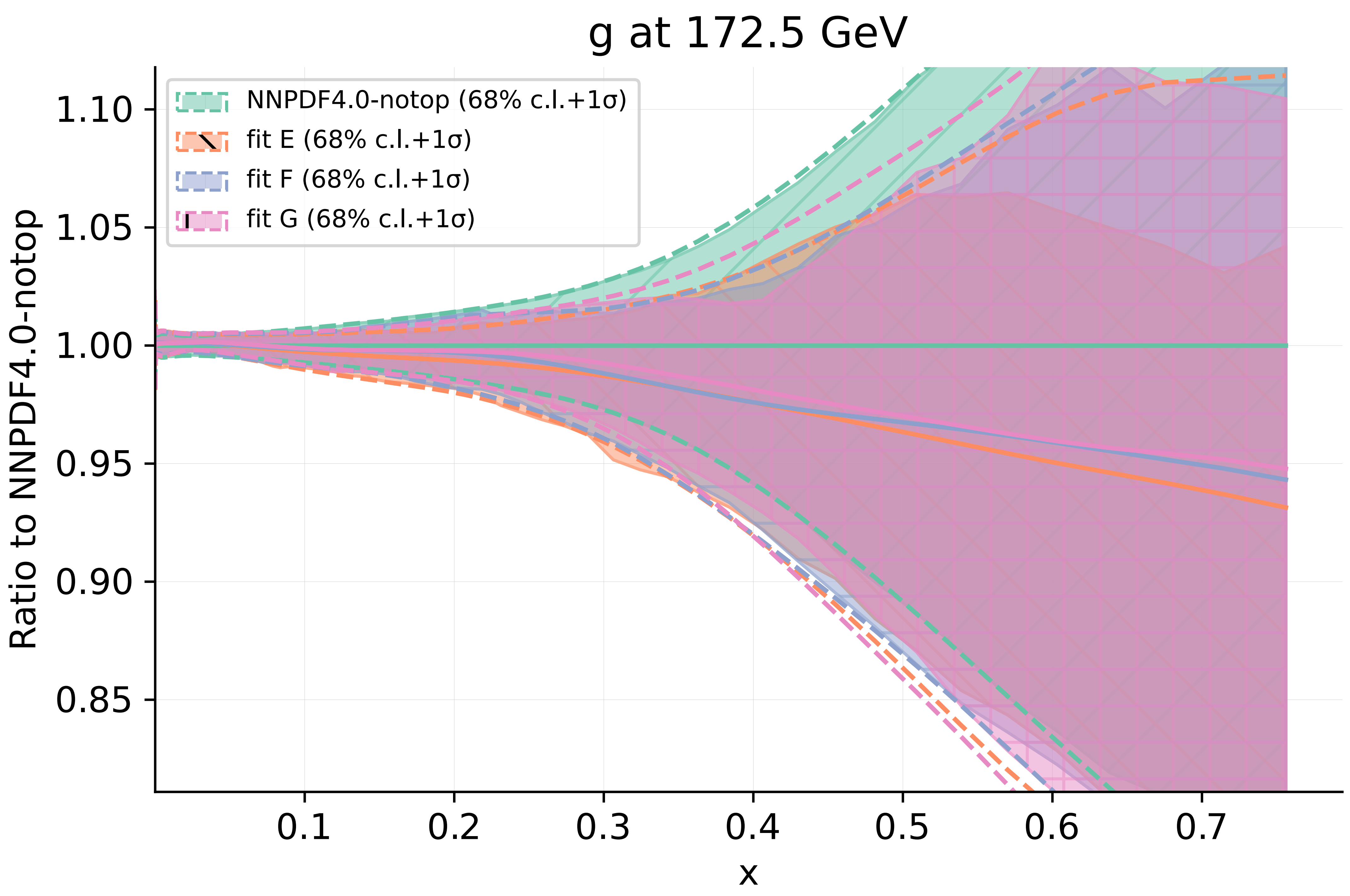
- First, we consider the impact of our dataset on PDFs **in the SM**.
- Begin by considering the updates to the **inclusive** $t\bar{t}$ and **single-top** dataset relative to NNPDF4.0. If we perform a SM PDF fit using only our new inclusive $t\bar{t}$ and single-top data, we see a more pronounced effect on the **large-x gluon** relative to NNPDF4.0. The **uncertainty** is also **further reduced**.

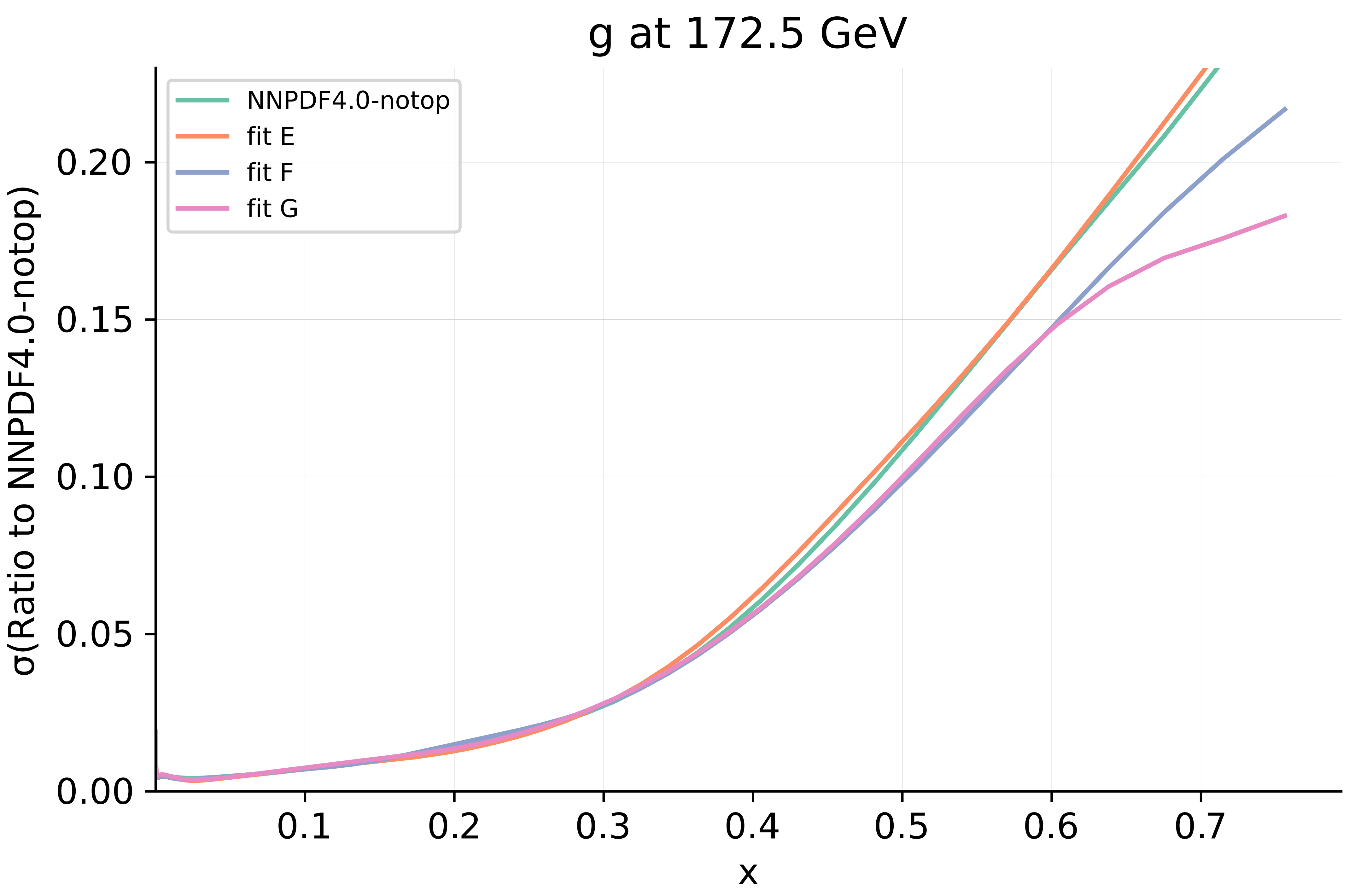




PDFs in the SM - impact of associated top

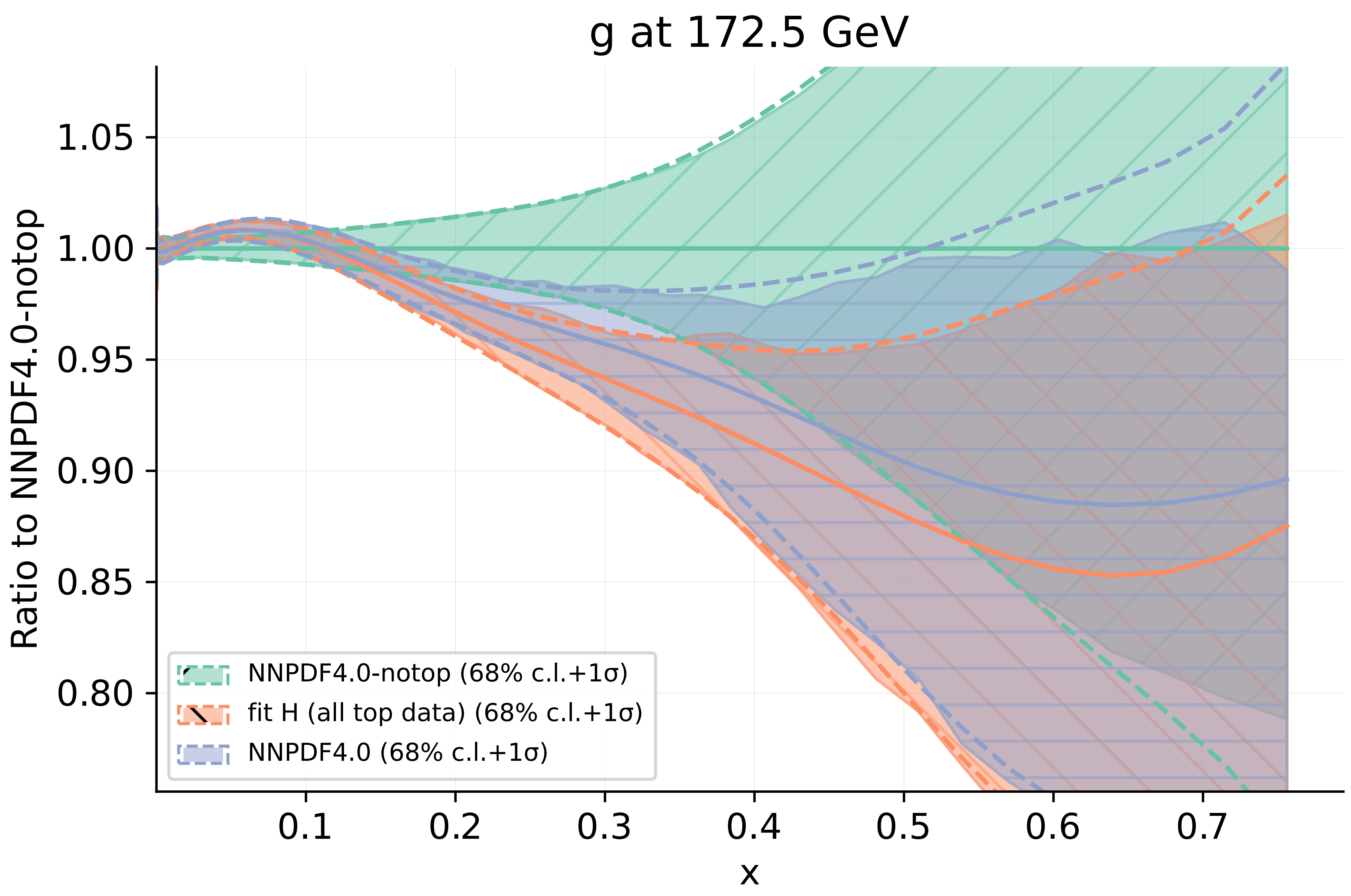
- Next, for the first time we consider the impact of **associated top data** in a PDF fit. There is only a very mild effect on the central value of the gluon, reducing it at large-x, and fractionally reducing uncertainty.

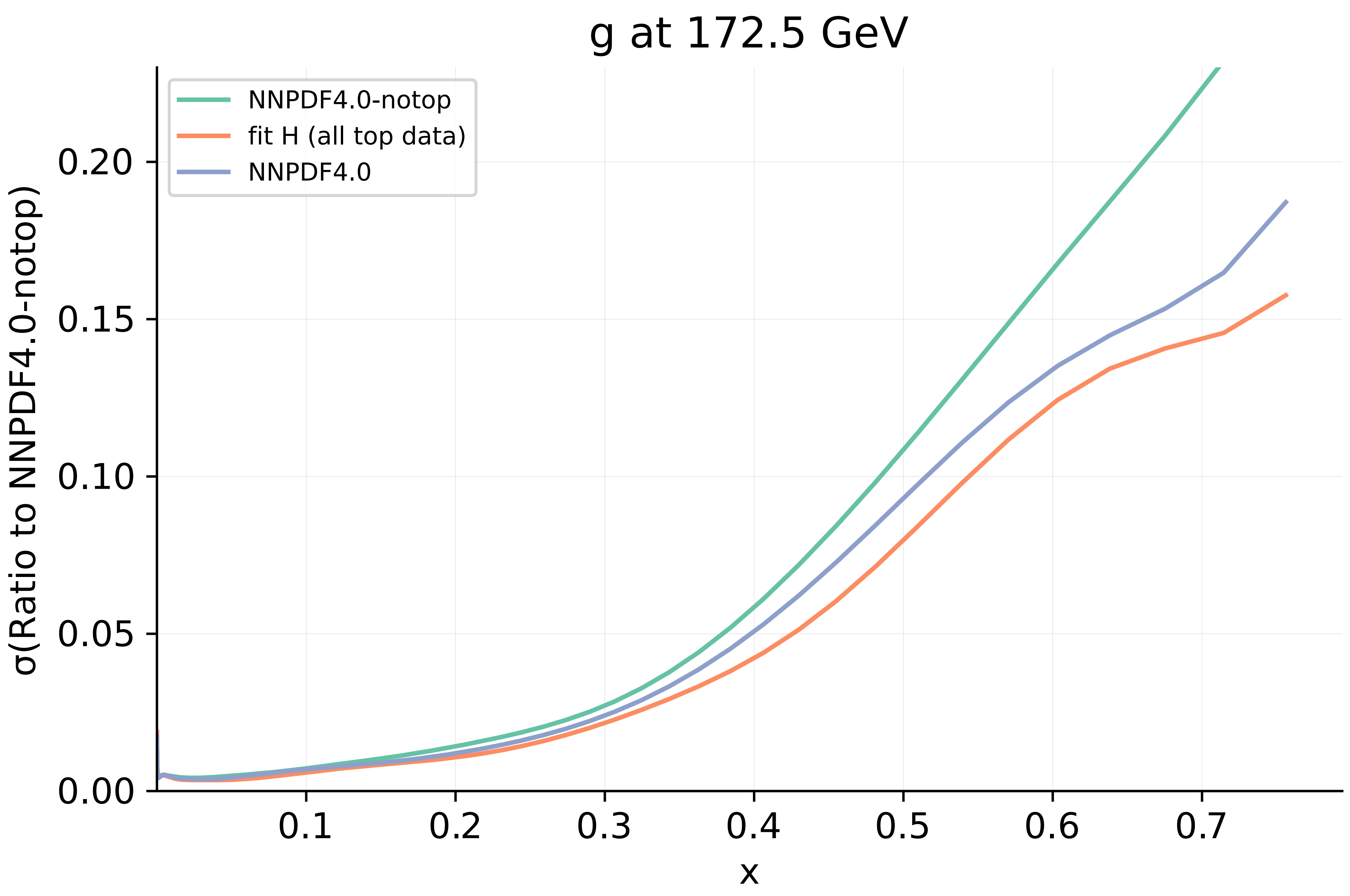




PDFs in the SM - impact of all new top data

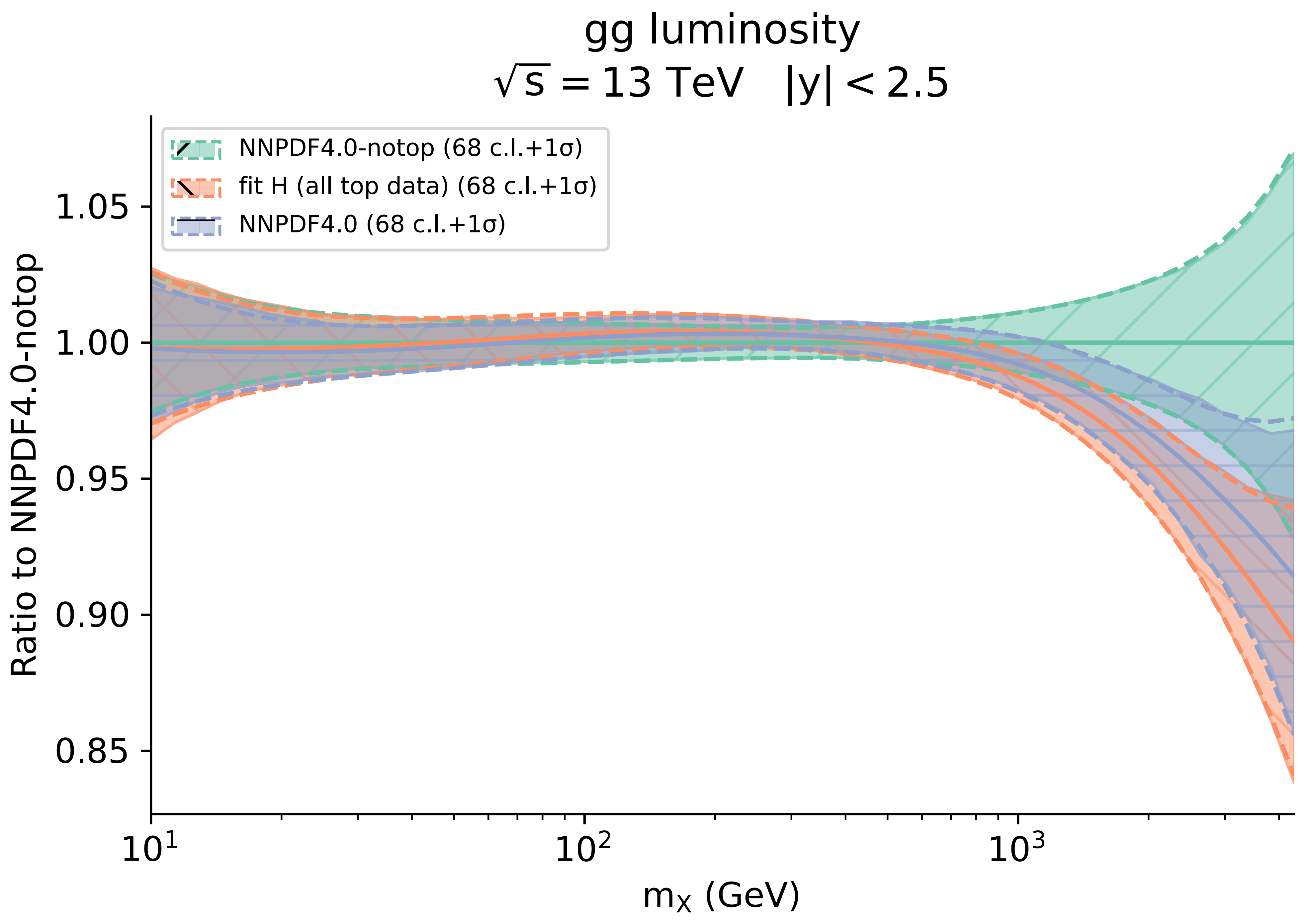
- Finally, we present the results of a **complete PDF fit** including **all our new top data**. As expected, the effect on the large-x gluon is broadly the same as the effect of just including the inclusive $t\bar{t}$ and single-top data, but is mildly tempered by the associated top data.

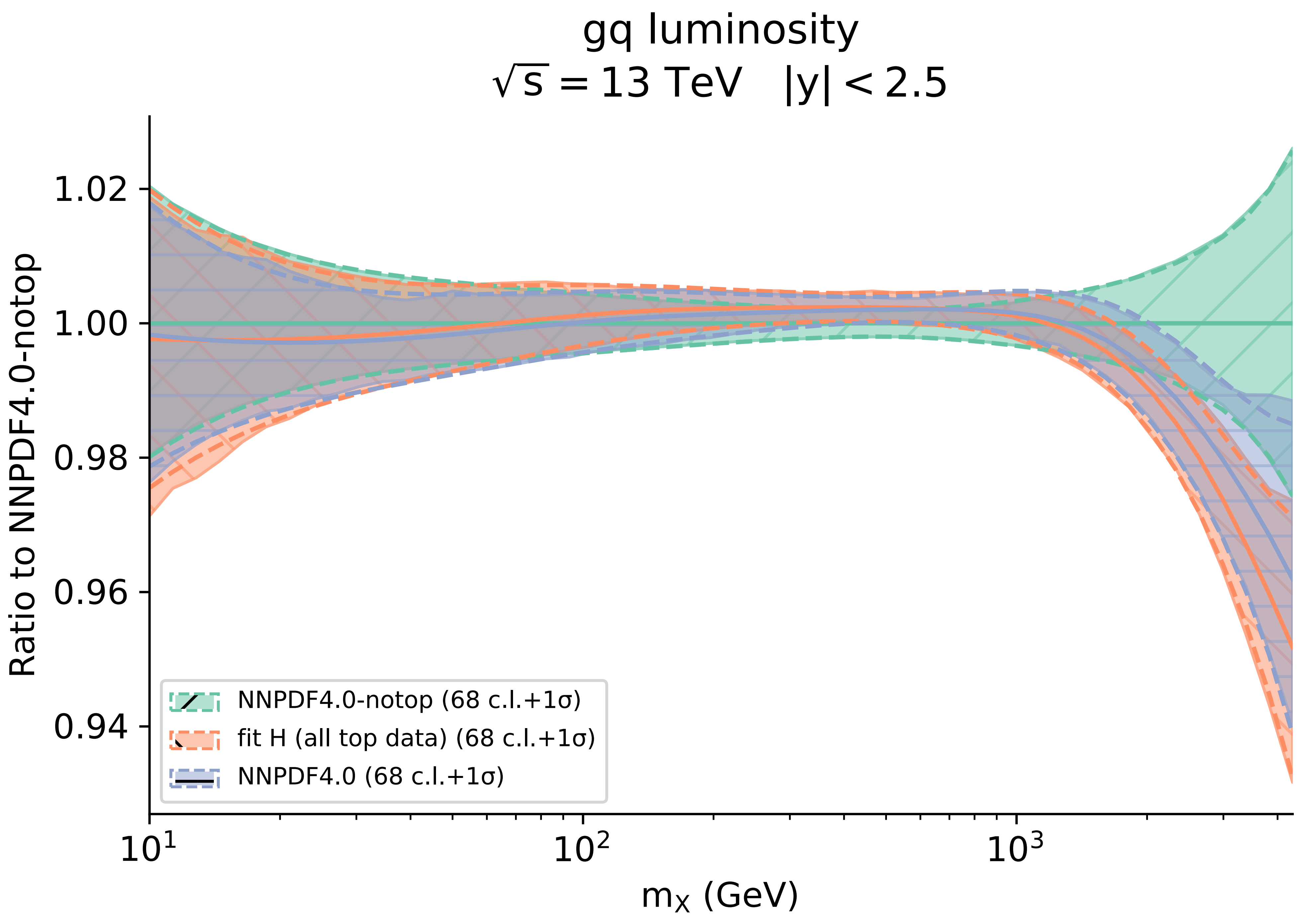




PDFs in the SM - impact of all new top data

- A similar trend holds for the **PDF luminosities**, with our new updated fit compatible with NNPDF4.0, but with the central luminosity reduced relative to NNPDF4.0 at very large invariant mass.

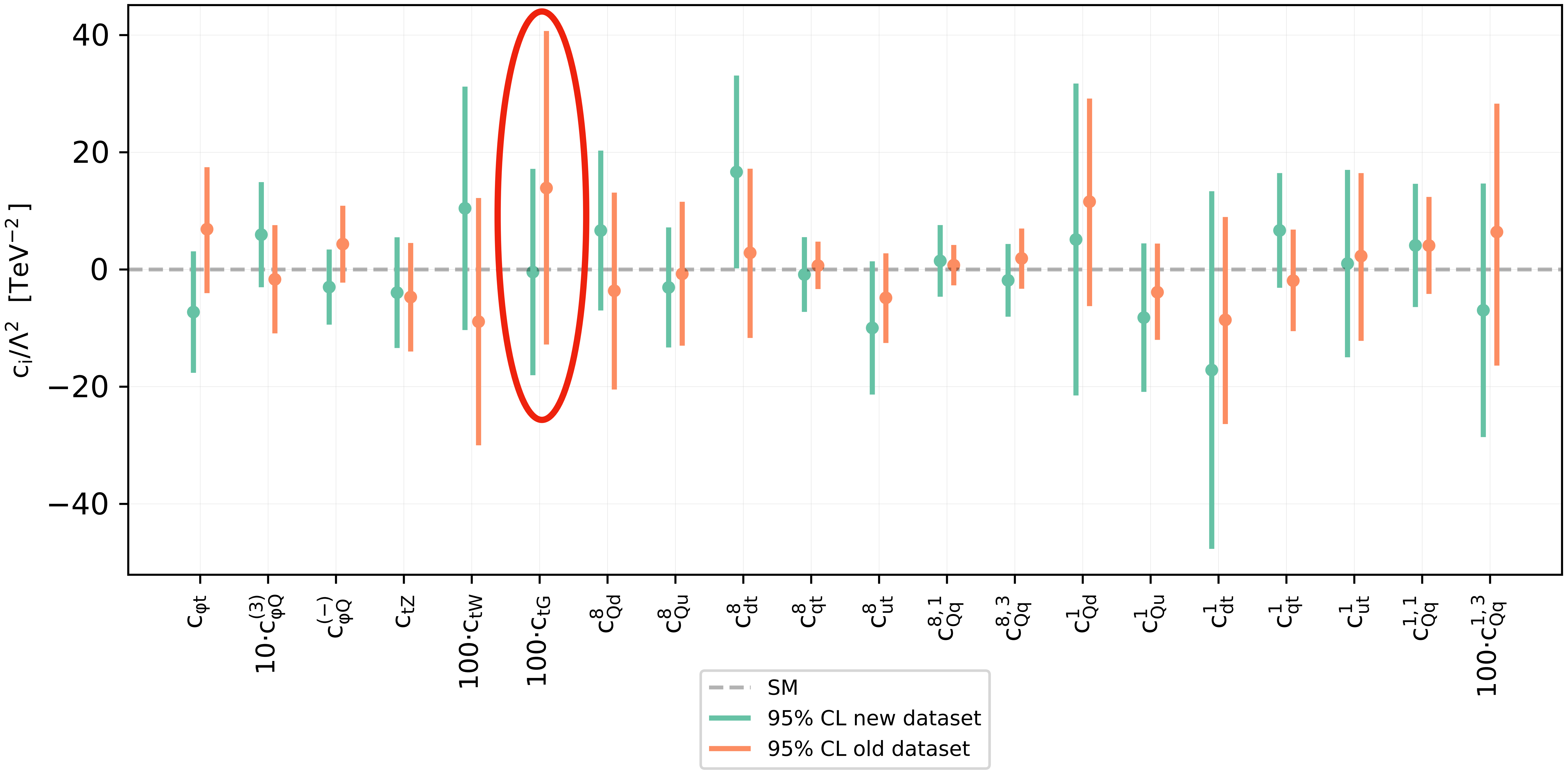




SMEFT-only fits: linear SMEFT

- We have also performed SMEFT-only fits to see the impact of our new dataset relative to previous SMEFT-fits, namely **SMEFiT**.

- At the **linear level** in the SMEFT, best improvement is seen in c_{tG}, whose bound undergoes a 35% tightening - this is traced to more precise total $t\bar{t}$ measurements.

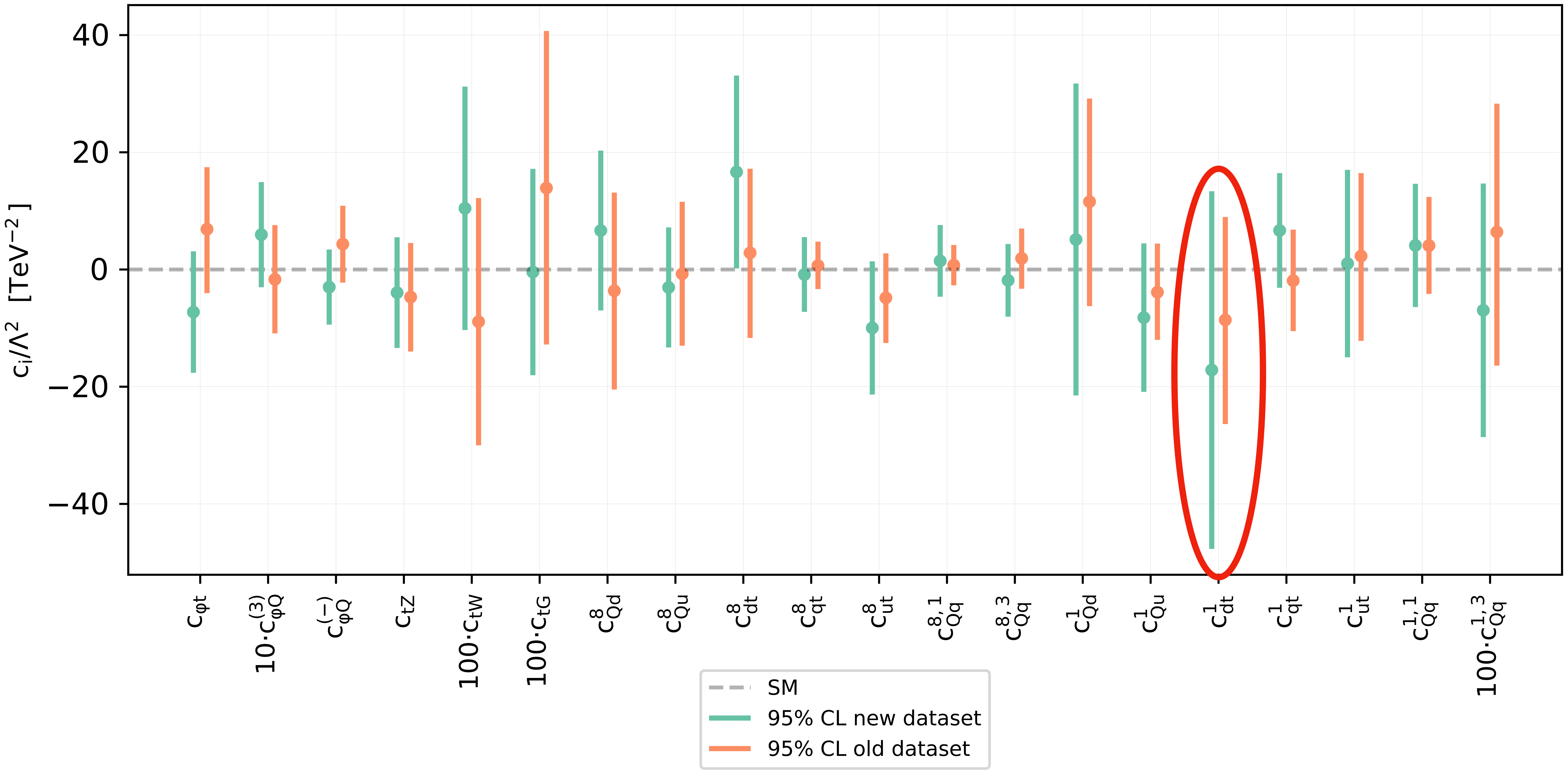


SMEFT-only fits: linear SMEFT

- Some other coefficients undergo a **shift in the central value**, but no tightening or broadening of the constraint.

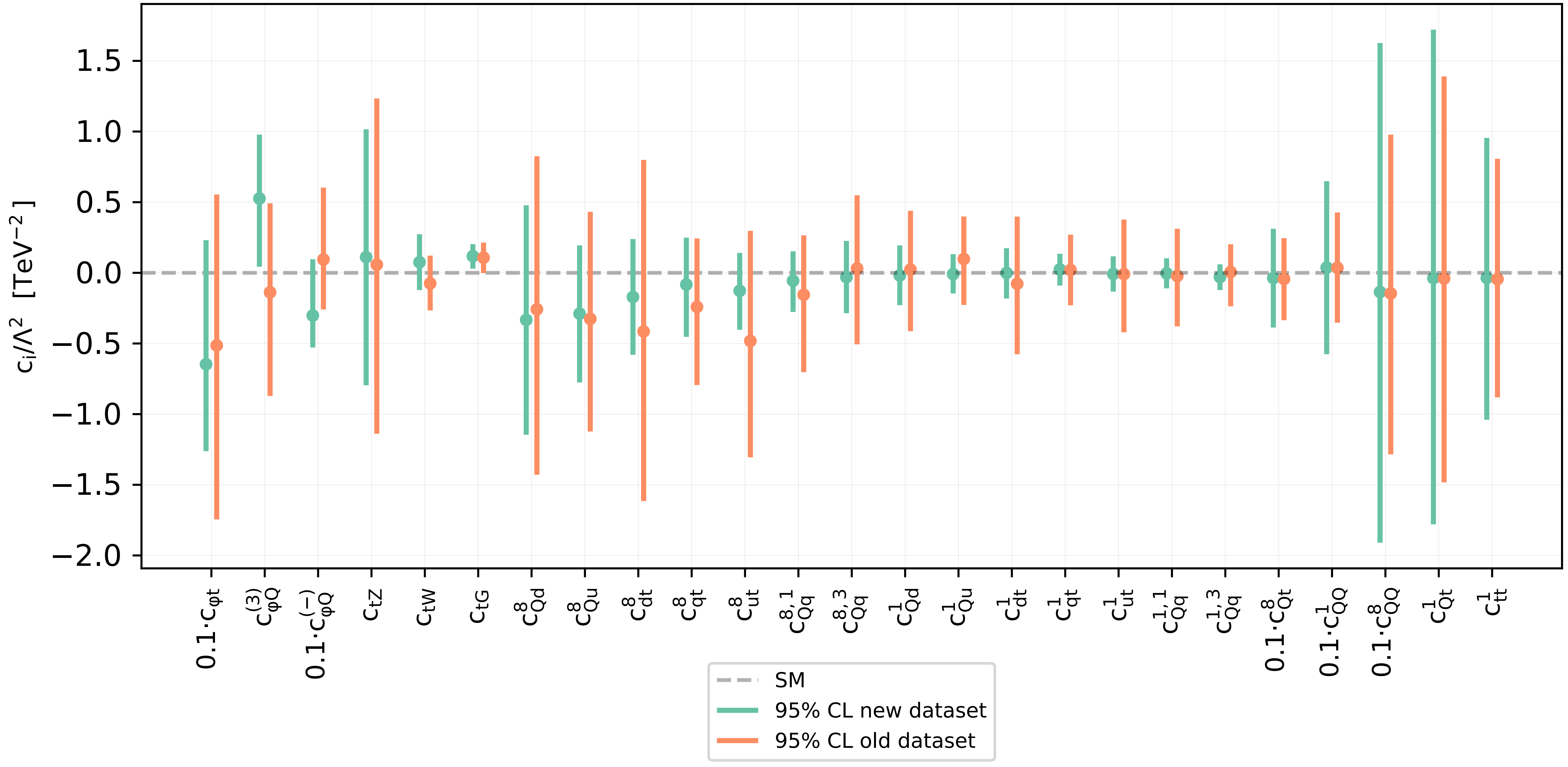
- Some coefficients have **broader bounds** than previously obtained, in particular some of the four-fermion operators.

- However, bounds are very weak here anyway, and likely challenge EFT validity.



SMEFT-only fits: quadratic SMEFT

- Results are **much more promising** when **quadratic SMEFT effects** are included. A **significant tightening** of bounds is seen for most operators.
- Only the five **four-heavy operators** experience broadening relative to the old dataset. This could point to some inconsistency in the $t\bar{t}t\bar{t}$ and $t\bar{t}b\bar{b}$ data, but with such large uncertainties, it is difficult to be precise.



PDF-SMEFT correlation

- We can try to get intuition for the result of the **joint PDF-SMEFT** fit by considering the **PDF-SMEFT correlation** in the SMEFT-only fits.

PDF-SMEFT correlation

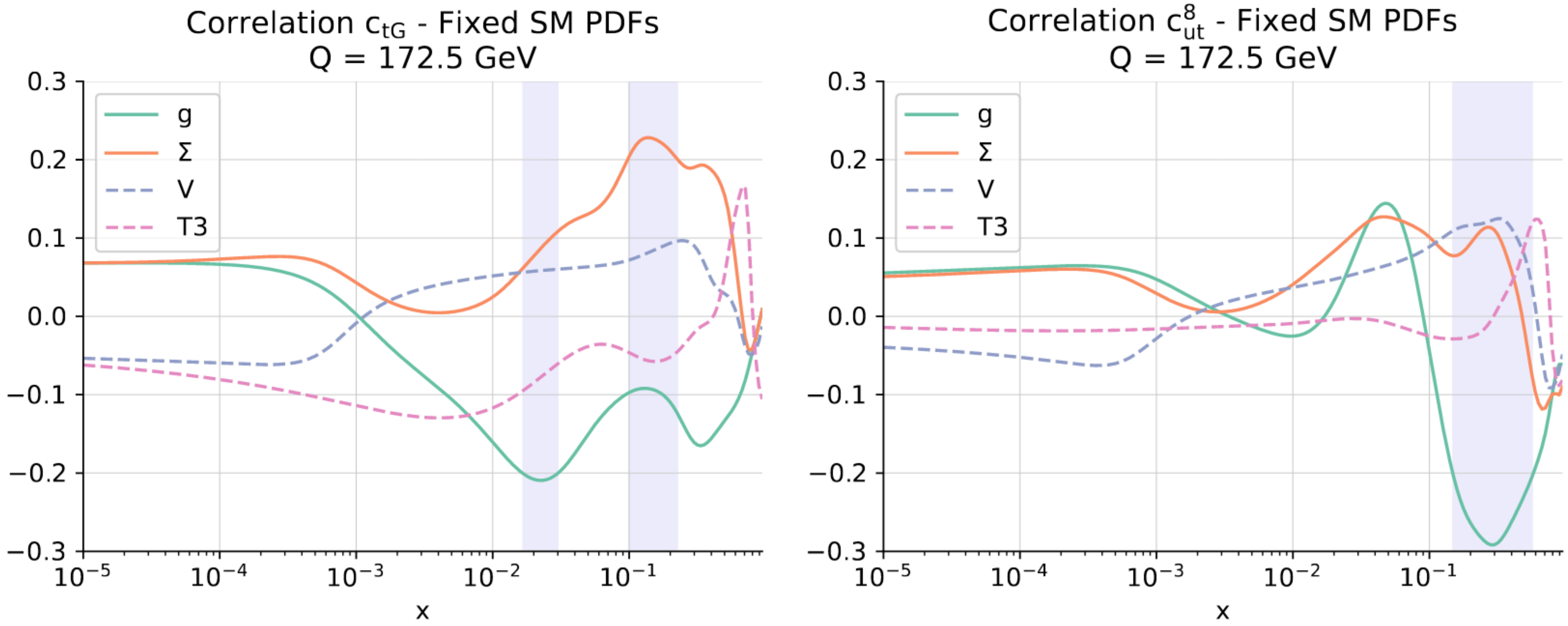
- We can try to get intuition for the result of the **joint PDF-SMEFT** fit by considering the **PDF-SMEFT correlation** in the SMEFT-only fits.

- This is defined for each Wilson coefficient and each PDF flavour by:

$$\rho\left(c, f(x,Q^2)\right) = \frac{\left\langle c^{(k)} f^{(k)}(x,Q^2)\right\rangle_k - \left\langle c^{(k)}\right\rangle_k \left\langle f^{(k)}(x,Q^2)\right\rangle_k}{\sqrt{\left\langle (c^{(k)})^2\right\rangle_k - \left\langle c^{(k)}\right\rangle_k^2}\sqrt{\left\langle \left(f^{(k)}(x,Q^2)\right)^2\right\rangle_k - \left\langle f^{(k)}(x,Q^2)\right\rangle_k^2}}$$

PDF-SMEFT correlation

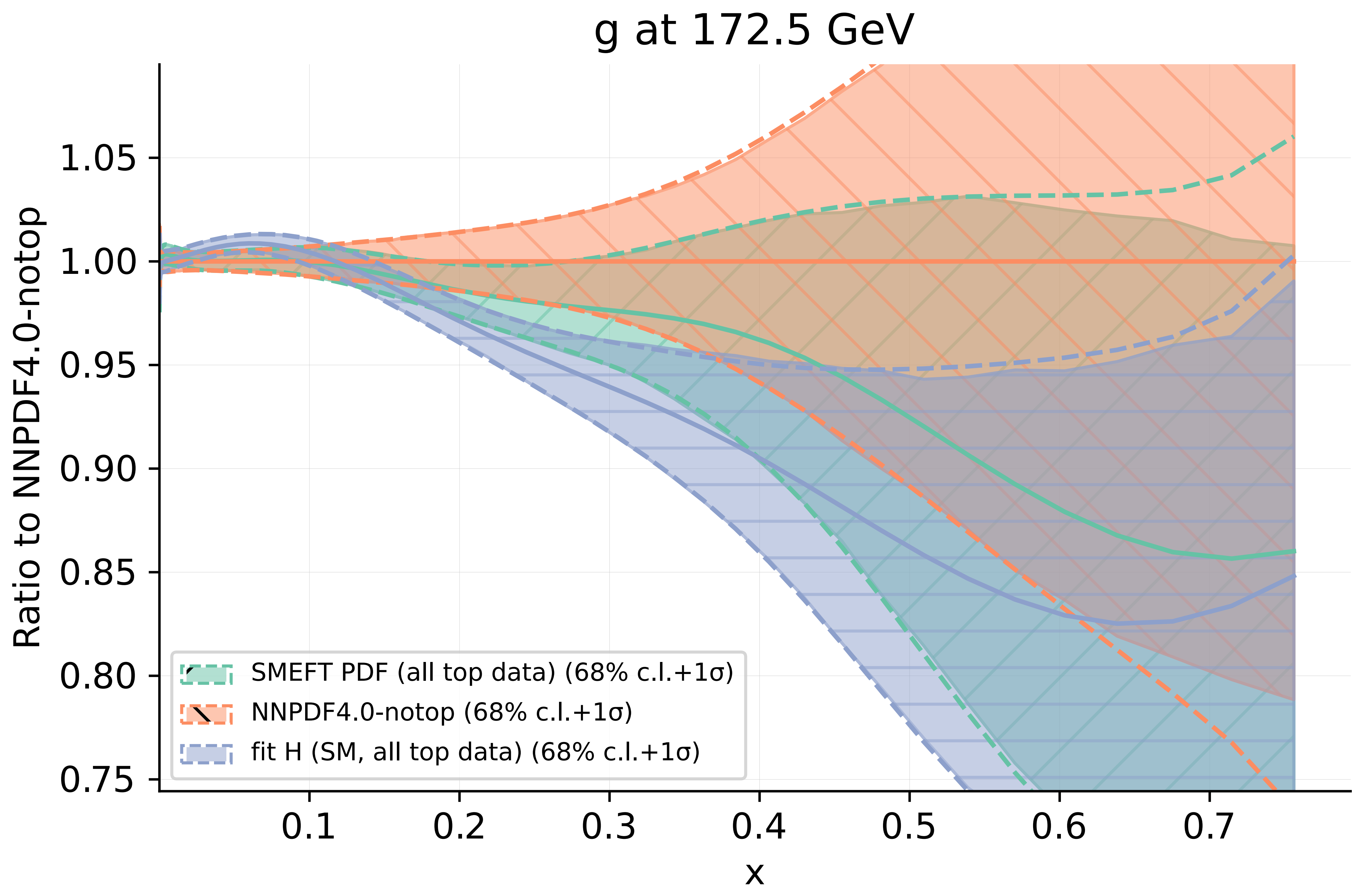
- We see the **strongest correlation** between the Wilson coefficients and the gluon PDF at high-x, as to be expected. The correlation is still **mild** though, suggesting that the interplay will also be **relatively mild**.

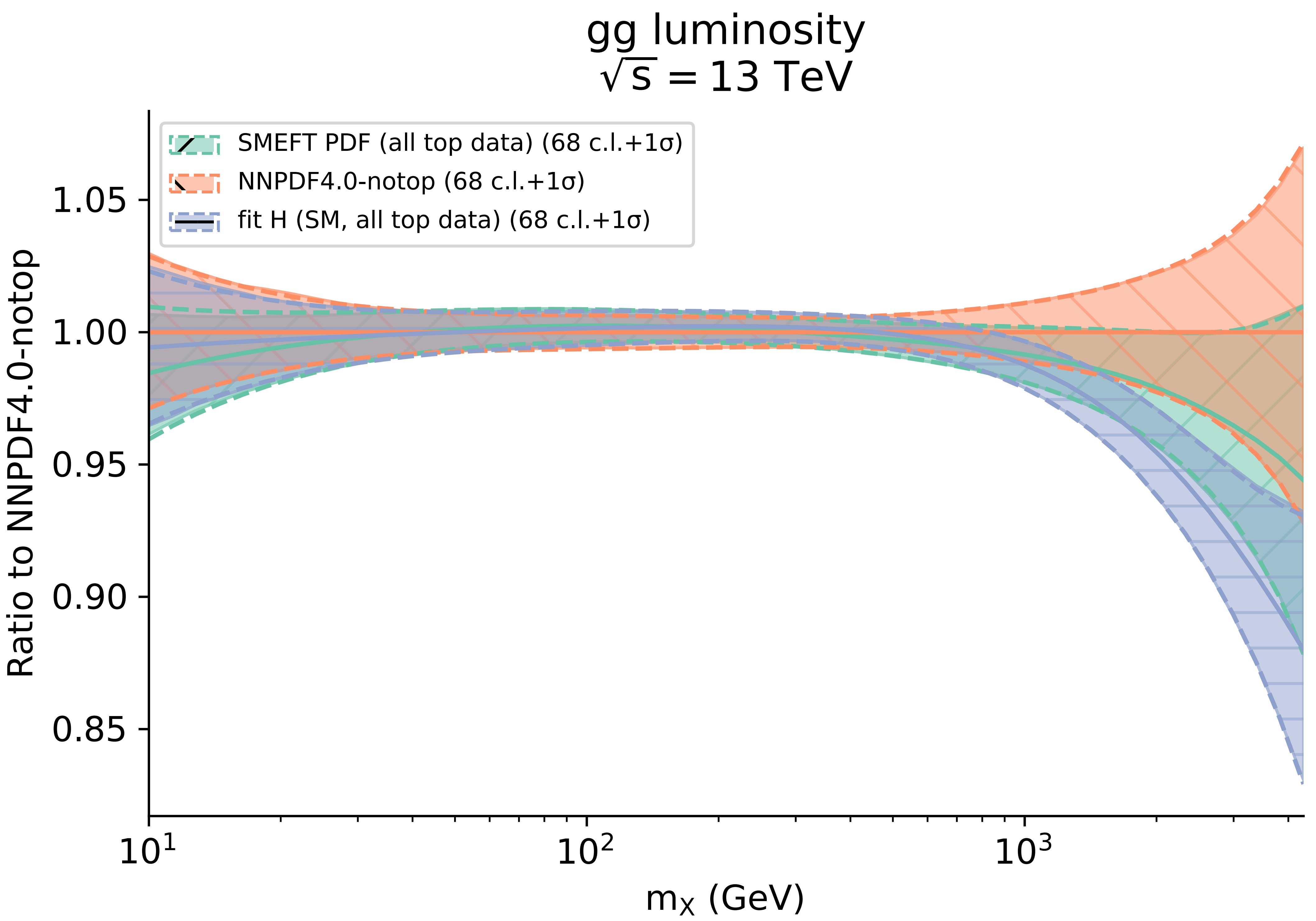


Now, let's do the joint fit...

Joint PDF-SMEFT fits: linear SMEFT

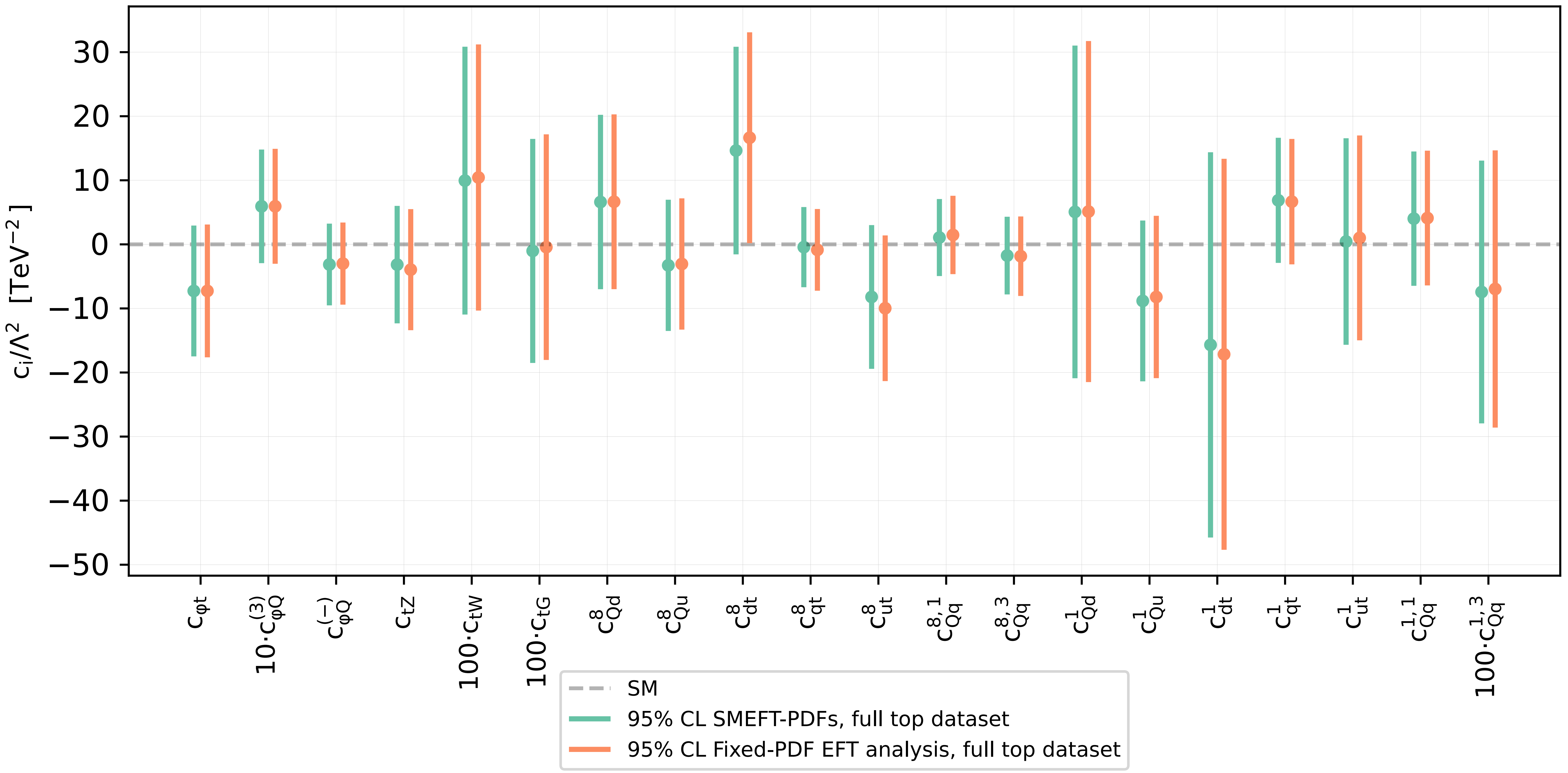
- Finally, we present the key result of the work: a **simultaneous** determination of PDFs and SMEFT Wilson coefficients. We start assuming **linear SMEFT**.
- In terms of the gluon PDFs and luminosities, we find that a simultaneous determination **reduces the pull** of the top data from the **non-top baseline**.





Joint PDF-SMEFT fits: linear SMEFT

- On the other hand, we find that the bounds on the Wilson coefficients are **very stable** between a simultaneous PDF-SMEFT fit and a SMEFT-only fit.



- This indicates that within a **linear EFT interpretation** of the top data, the PDF effects are **currently subdominant**.

Joint PDF-SMEFT fits: quadratic SMEFT

- **Next obvious fit…** joint PDF-SMEFT fit using **quadratic SMEFT contributions?** Could interplay be more pronounced there … ?

Joint PDF-SMEFT fits: quadratic SMEFT

- **Next obvious fit…** joint PDF-SMEFT fit using **quadratic SMEFT contributions?** Could interplay be more pronounced there … ?

- However… during the course of our study, we discovered an important problem with the **Monte Carlo replica method** used to propagate uncertainties in the SIMUnet methodology.

Joint PDF-SMEFT fits: quadratic SMEFT

- **Next obvious fit…** joint PDF-SMEFT fit using **quadratic SMEFT contributions?** Could interplay be more pronounced there … ?

- However… during the course of our study, we discovered an important problem with the **Monte Carlo replica method** used to propagate uncertainties in the SIMUnet methodology.

- The issue is such that quadratic results with the SIMUnet methodology (and indeed with any methodology that uses the Monte Carlo replica method) are **currently unreliable**.

Joint PDF-SMEFT fits: quadratic SMEFT

- **Next obvious fit…** joint PDF-SMEFT fit using **quadratic SMEFT contributions?** Could interplay be more pronounced there … ?

- However… during the course of our study, we discovered an important problem with the **Monte Carlo replica method** used to propagate uncertainties in the SIMUnet methodology.

- The issue is such that quadratic results with the SIMUnet methodology (and indeed with any methodology that uses the Monte Carlo replica method) are **currently unreliable**.

- An **upcoming publication** will describe the issue in more detail; for now, here's the basics…

Pitfalls of the Monte-Carlo replica method

- For simplicity, consider a single data point d with experimental variance σ^2, which we attempt to describe using the **quadratic** theory, involving a single theory parameter c:

$$t(c) = t^{\text{SM}} + t^{\text{lin}} c + t^{\text{quad}} c^2$$

- The Monte-Carlo replica method propagates the uncertainty from the data to the theory parameter by fitting to **pseudodata**. We sample lots of pseudodata replicas from a normal distribution based on the data, $d_p \sim N(d, \sigma^2)$, and define the corresponding **parameter replicas** to be a random function of the pseudodata given by minimising the χ^2-statistic:

$$c_p(d_p) = \text{arg min}_c \left(\frac{(t(c) - d_p)^2}{\sigma^2} \right)$$

Pitfalls of the Monte-Carlo replica method

- In this very simple example, one can compute the distribution function of the parameter replicas analytically; it is given by:

$$P_{c^{(i)}}(c) \propto \delta\left(c + \frac{t^{\text{lin}}}{2t^{\text{quad}}}\right) \int_{-\infty}^{t_{\text{min}}} dx \; \exp\left(-\frac{1}{2\sigma^2}(x-d)^2\right) + \frac{2}{|2ct^{\text{quad}} + t^{\text{lin}}|} \exp\left(-\frac{1}{2\sigma^2}(d - t(c))^2\right)$$

- Here, t_{min} is the minimum value of the theory (which is a parabola).

Pitfalls of the Monte-Carlo replica method

- In this very simple example, one can compute the distribution function of the parameter replicas analytically; it is given by:

$$P_{c^{(i)}}(c) \propto \delta\left(c + \frac{t^{\text{lin}}}{2t^{\text{quad}}}\right) \int_{-\infty}^{t_{\text{min}}} dx \ \exp\left(-\frac{1}{2\sigma^2}(x-d)^2\right) + \frac{2}{|2ct^{\text{quad}} + t^{\text{lin}}|} \exp\left(-\frac{1}{2\sigma^2}(d - t(c))^2\right)$$

- Here, t_{min} is the minimum value of the theory (which is a parabola).

- **Key features to note:**

Pitfalls of the Monte-Carlo replica method

- In this very simple example, one can compute the distribution function of the parameter replicas analytically; it is given by:

$$P_{c^{(i)}}(c) \propto \delta\left(c + \frac{t^{\text{lin}}}{2t^{\text{quad}}}\right) \int_{-\infty}^{t_{\text{min}}} dx \; \exp\left(-\frac{1}{2\sigma^2}(x-d)^2\right) + \frac{2}{|2ct^{\text{quad}} + t^{\text{lin}}|} \exp\left(-\frac{1}{2\sigma^2}(d - t(c))^2\right)$$

- Here, t_{min} is the minimum value of the theory (which is a parabola).

- **Key features to note:**
  - Part of the distribution looks like a **scaled version** of what we would expect from a **Bayesian method with uniform prior**.

Pitfalls of the Monte-Carlo replica method

- In this very simple example, one can compute the distribution function of the parameter replicas analytically; it is given by:

$$P_{c^{(i)}}(c) \propto \delta\left(c + \frac{t^{\text{lin}}}{2t^{\text{quad}}}\right) \int_{-\infty}^{t_{\min}} dx \ \exp\left(-\frac{1}{2\sigma^2}(x-d)^2\right) + \frac{2}{|2ct^{\text{quad}} + t^{\text{lin}}|} \exp\left(-\frac{1}{2\sigma^2}(d - t(c))^2\right)$$

- Here, t_{min} is the minimum value of the theory (which is a parabola).

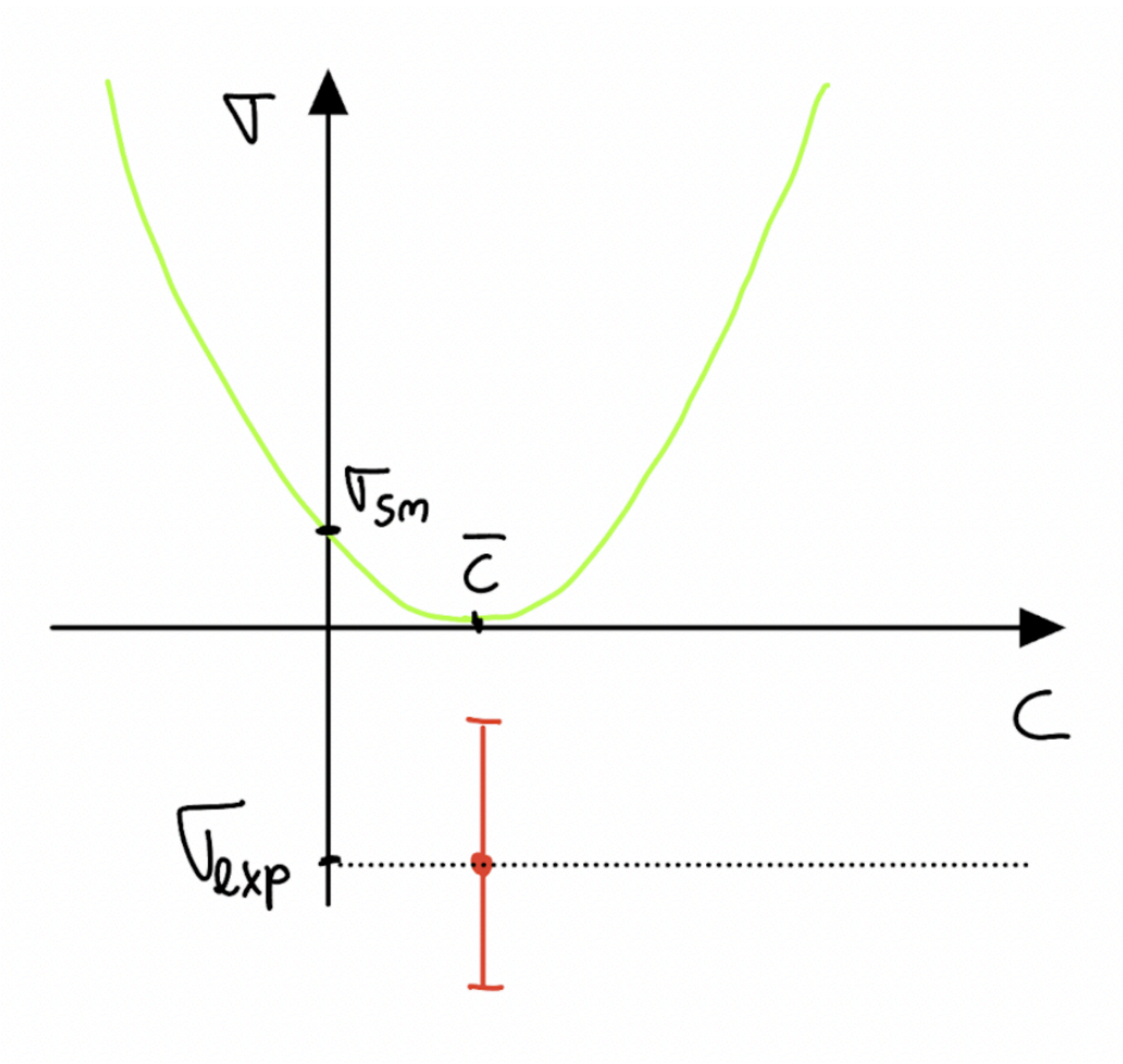
- **Key features to note:**
  - Part of the distribution looks like a **scaled version** of what we would expect from a **Bayesian method with uniform prior**.
  - There is also a **delta function spike** in the distribution - interesting to ask: why...?

Pitfalls of the Monte-Carlo replica method

- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.

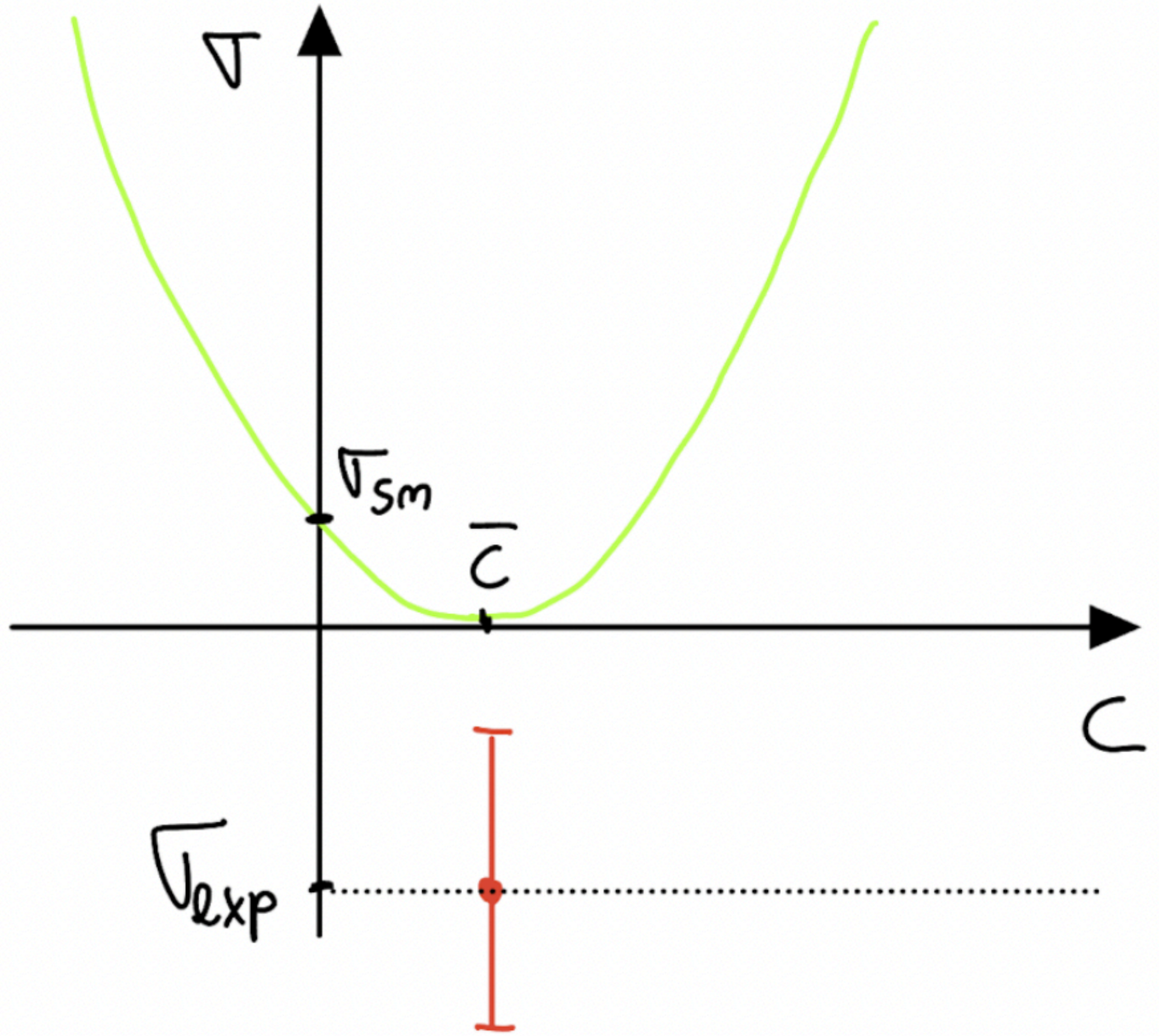
Pitfalls of the Monte-Carlo replica method

- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.



Pitfalls of the Monte-Carlo replica method

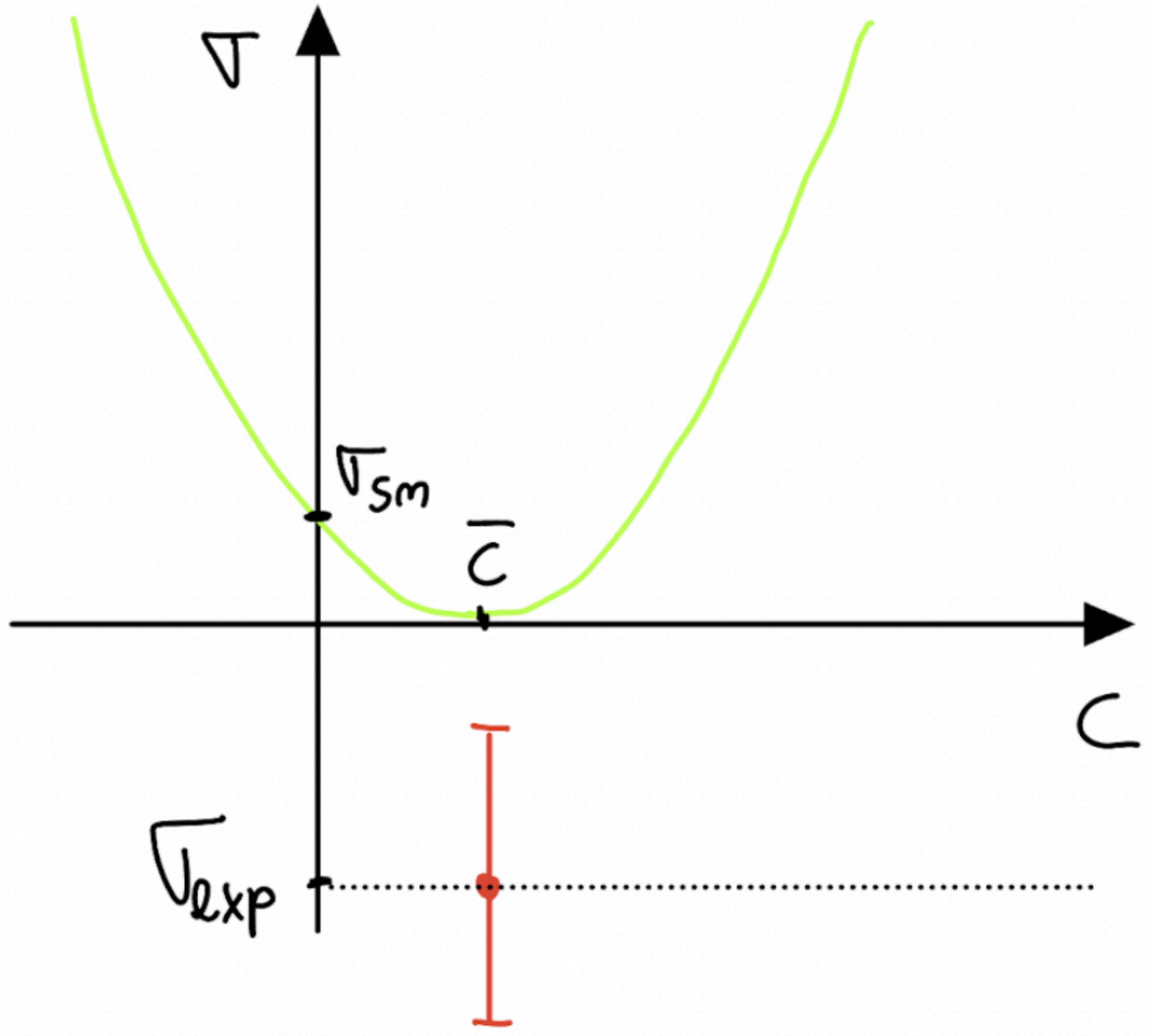
- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.



- This occurs if the experimental data falls **below the minimum** of the theory, or **above but close** to the minimum.

Pitfalls of the Monte-Carlo replica method

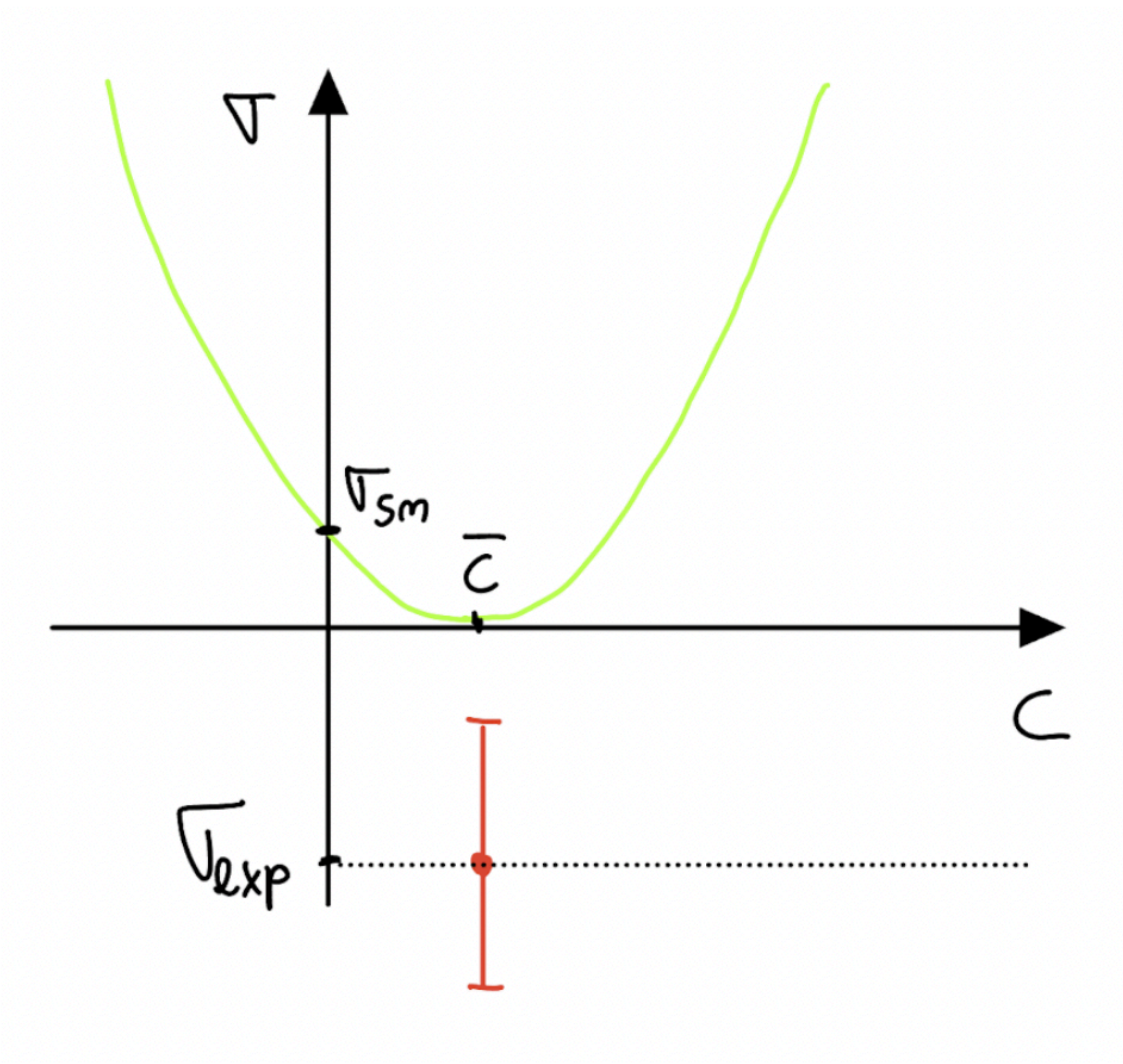
- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.



- This occurs if the experimental data falls **below the minimum** of the theory, or **above but close** to the minimum.
- Any pseudodata replica that falls below the minimum results in the **same parameter replica**, corresponding to the parameter value that gives the minimum.

Pitfalls of the Monte-Carlo replica method

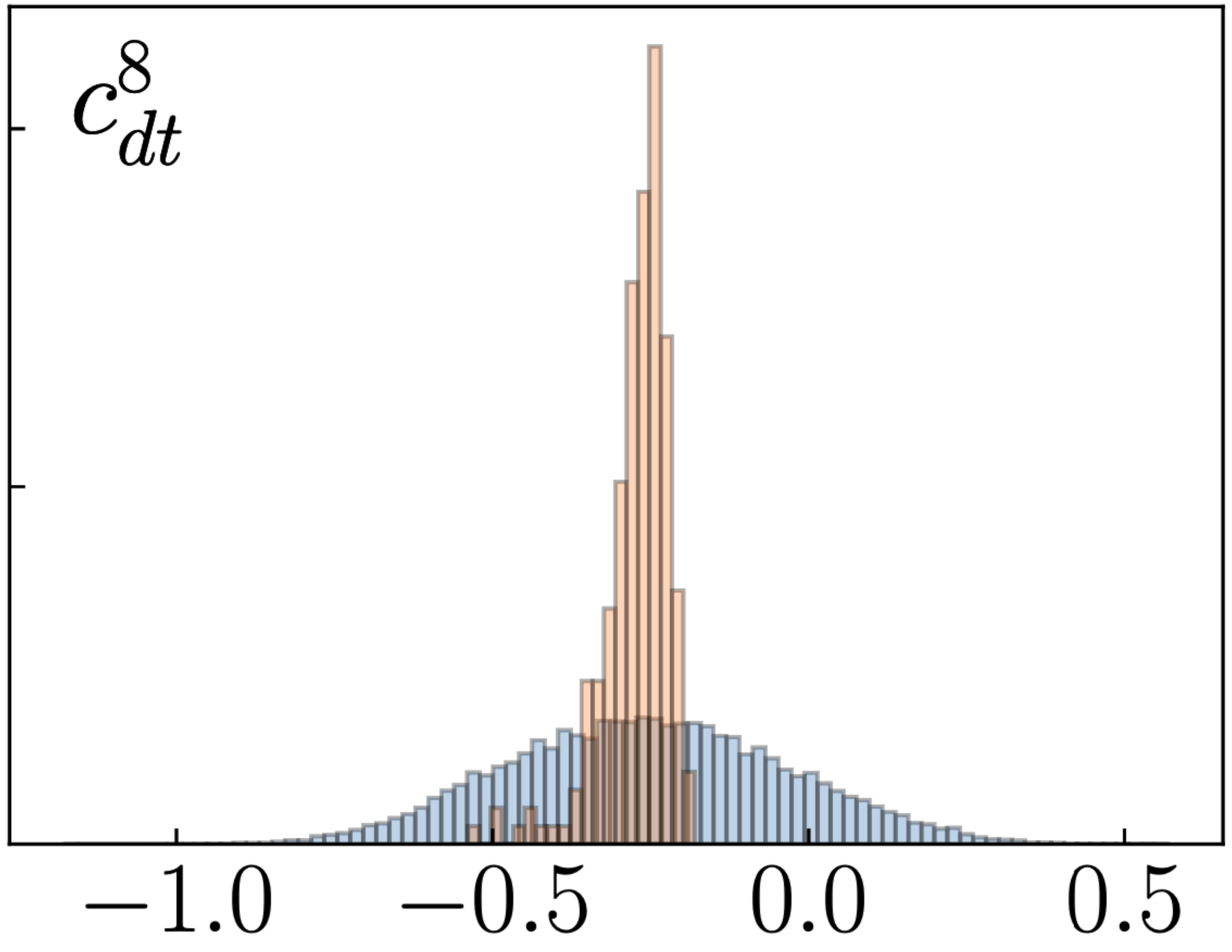
- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.



- This occurs if the experimental data falls **below the minimum** of the theory, or **above but close** to the minimum.
- Any pseudodata replica that falls below the minimum results in the **same parameter replica**, corresponding to the parameter value that gives the minimum.
- This gives rise to the spike in the distribution at $c = -t^{\text{lin}}/2t^{\text{quad}}$.

Pitfalls of the Monte-Carlo replica method

- These problems extend to our top fit… for example in a **realistic quadratic fit** of one operator c_{dt}^{8}, we get the following comparison between the Monte-Carlo method (orange) and a Bayesian method with uniform prior (blue).
- We see that **Monte-Carlo massively underestimates uncertainties.**



Key questions for the future:

Can the MC replica method be modified to agree with Bayesian methods?

To what extent do existing fits (in the SMEFT world, PDF world, and beyond) that use the MC replica method underestimate uncertainties?

Conclusions

Conclusions

- **Simultaneous determination of PDFs and BSM parameters**, will be **very important in future analyses** (especially as we enter Run III).

- Members of the **PBSP team** have already produced three works in the direction of simultaneous PDF-SMEFT fits: (i) a **phenomenological study** 2104.02723 showing the need for simultaneous extraction; (ii) a **methodology** (SimuNET, 2201.07240) capable of **fast simultaneous fitting**; (iii) a **comprehensive simultaneous extraction** of PDFs and SMEFT couplings from the **full LHC Run II top dataset**, 2303.06159.

Thanks for listening!
Questions?