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

*for EPS 2023, in Hamburg*

**James Moore, University of Cambridge**

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

# Related talks

- See also...
  - ✳ **Maeve Madigan: Can PDFs absorb new physics?**
    - On Tuesday, Maeve talked about the consequences of performing a SM PDF fit, when the data going into the fit is actually described by a New Physics model.
  - ✳ **Xiaomin Shen: Simultaneous extraction of PDFs and SMEFT parameters from jet and ttbar data**
    - Today, Xiaomin will talk about a simultaneous PDF-SMEFT extraction using top and jet data, using a different methodology to the one used in the study we present here.

# Talk overview

## 1. Joint PDF-SMEFT fits

## 2. The SIMUnet methodology

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

# 1. - Joint PDF-SMEFT fits

# The problem...

- Fits of **parton distributions** and **SMEFT Wilson coefficients** do not normally talk...

# The problem...

- Fits of **parton distributions** and **SMEFT Wilson coefficients** do not normally talk...

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

# The problem...

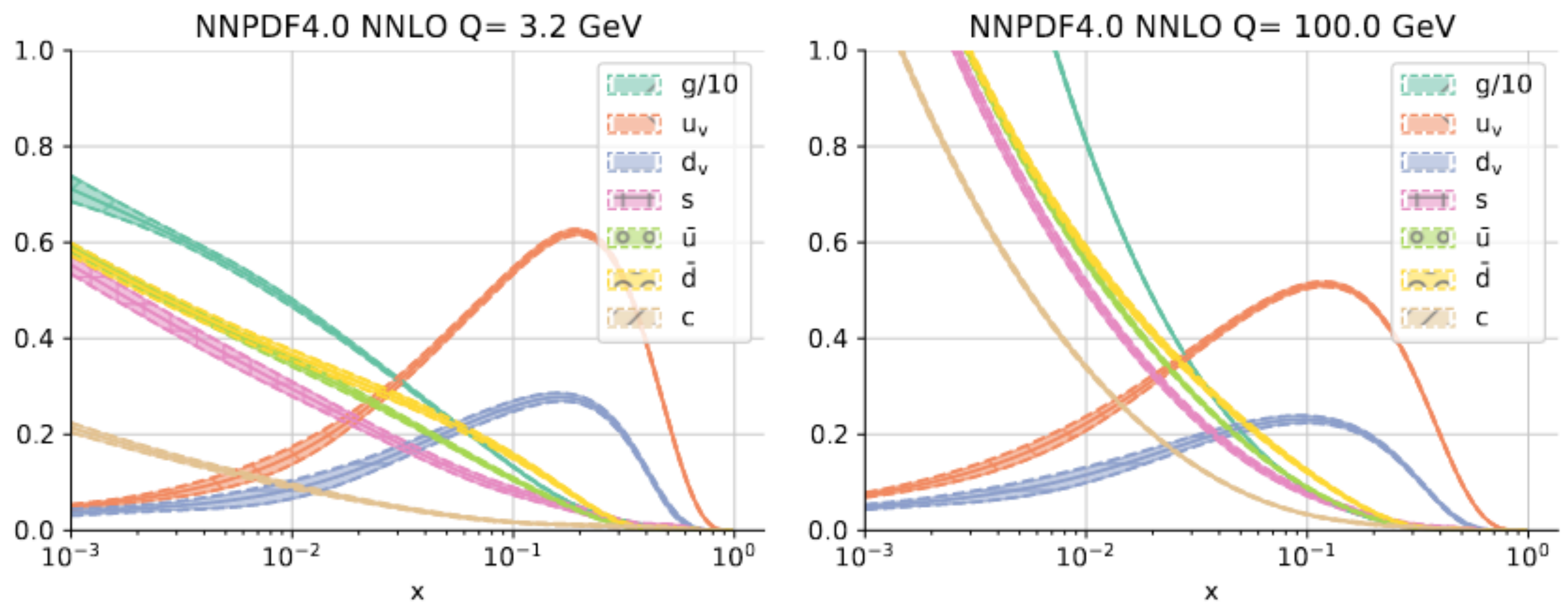
- Fits of **parton distributions** and **SMEFT Wilson coefficients** do not normally talk...

## 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 $\theta^*$ 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.



# The problem...

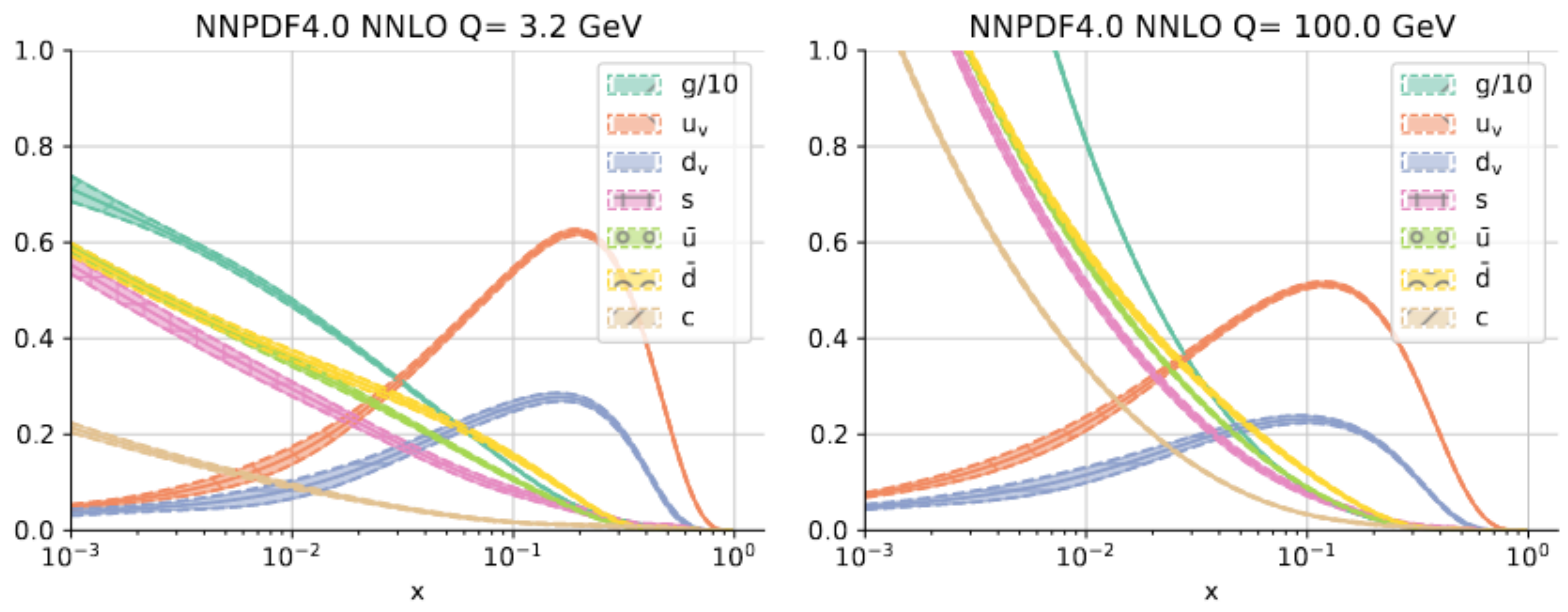
- Fits of **parton distributions** and **SMEFT Wilson coefficients** do not normally talk...

## 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 $\theta^*$ 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: $c^* = c^*(\bar{\theta})$.

# The problem...

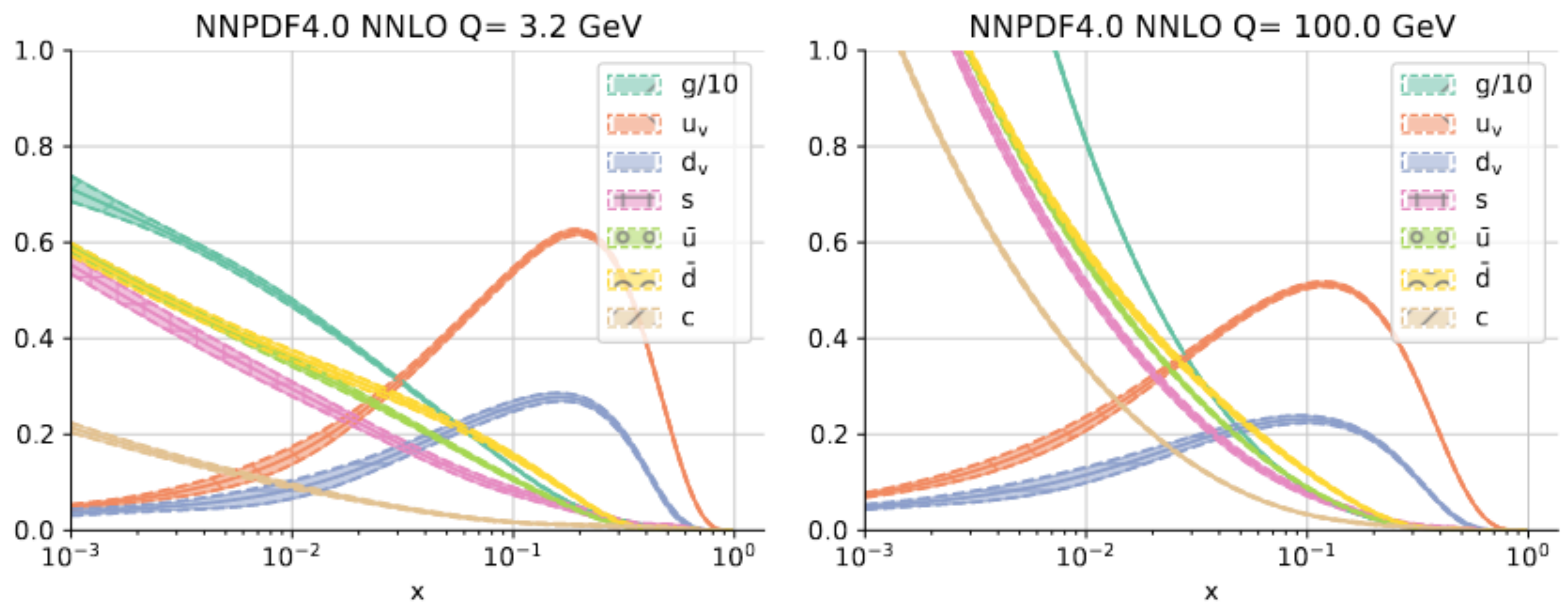
- Fits of **parton distributions** and **SMEFT Wilson coefficients** do not normally talk...

## 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 $\theta^*$ 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.

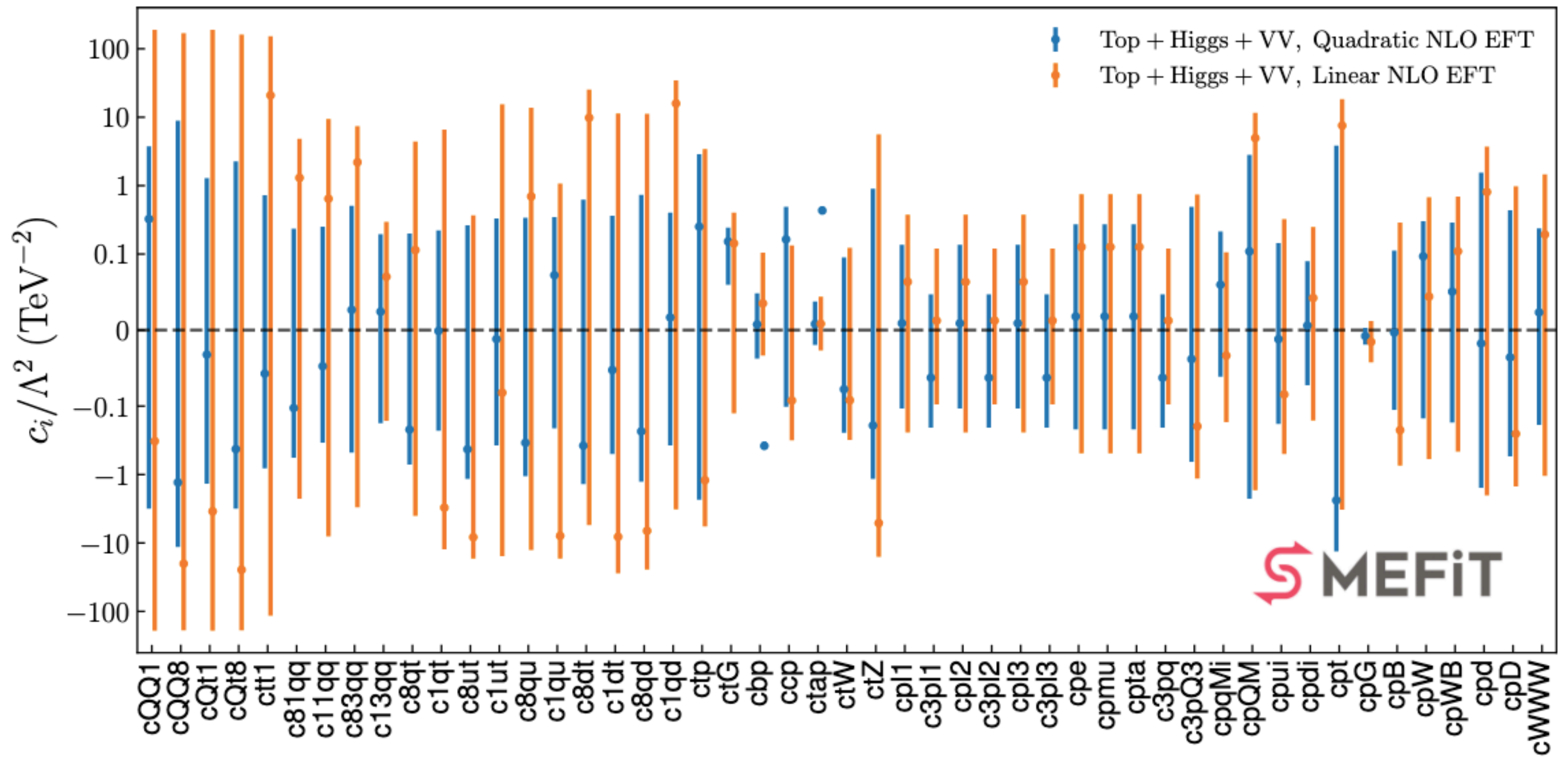



## SMEFT parameter fits

- 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.



# The problem...

- **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.

# The problem...

- **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.

# The problem...

- **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.

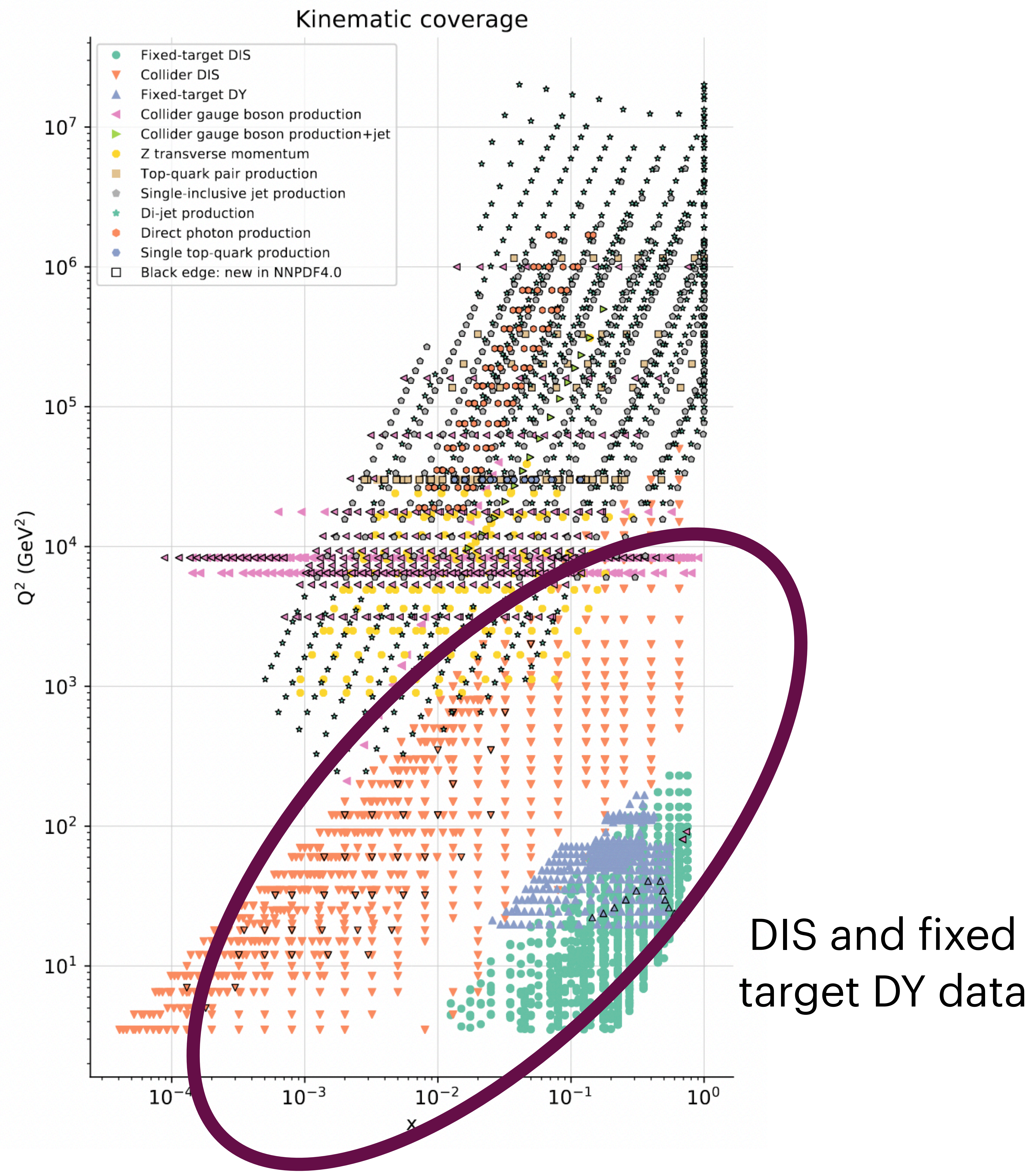
- In particular, if we fit PDFs **assuming all SMEFT couplings are zero**, but then **use those PDFs in a fit of SMEFT couplings**, our resulting bounds **could be misleading**. The same applies to SM parameters.
- We could even **miss New Physics**, or **see New Physics that isn't really there**!

# 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?**

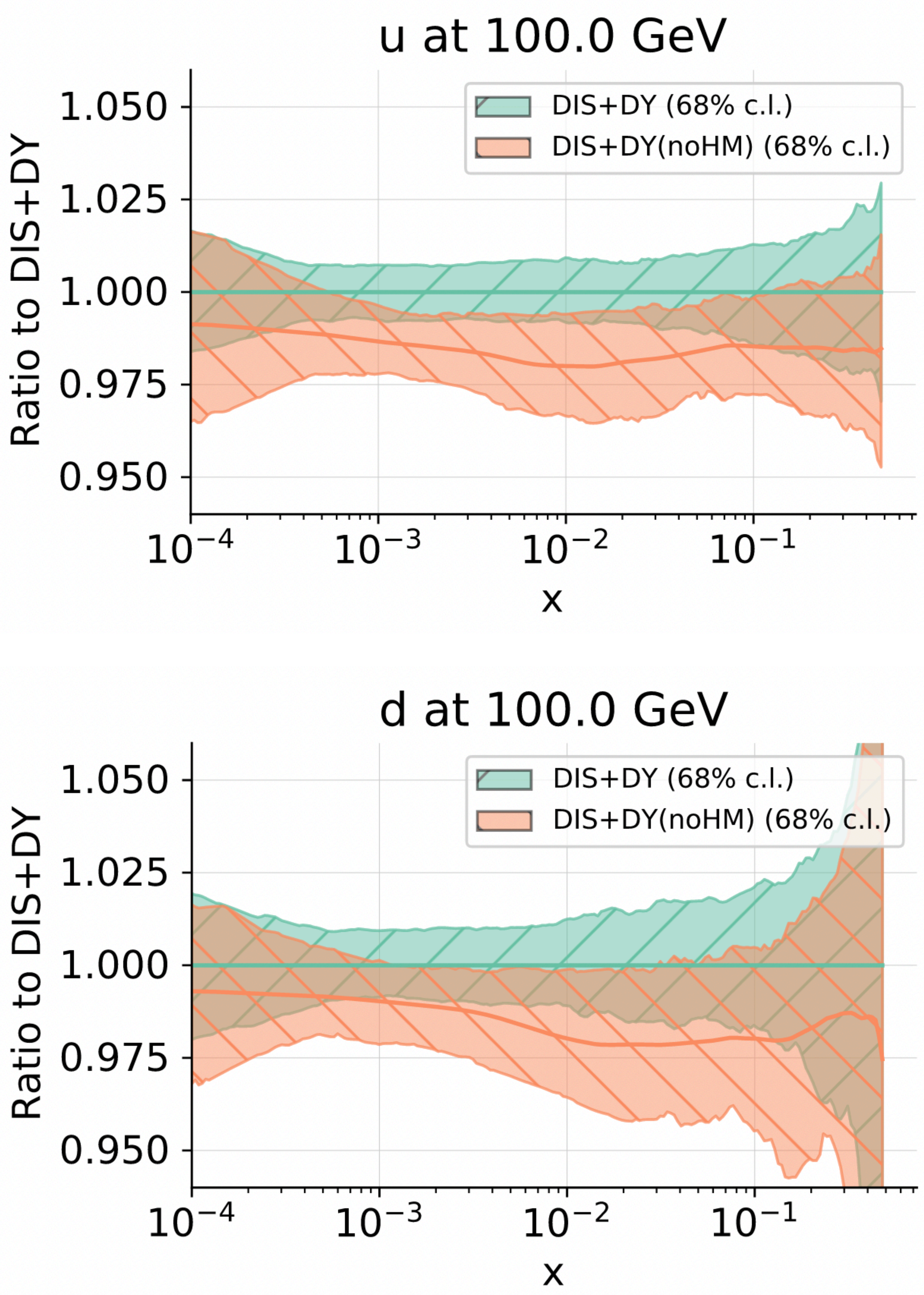
# PDF-SMEFT interplay: natural questions

- *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.



# 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

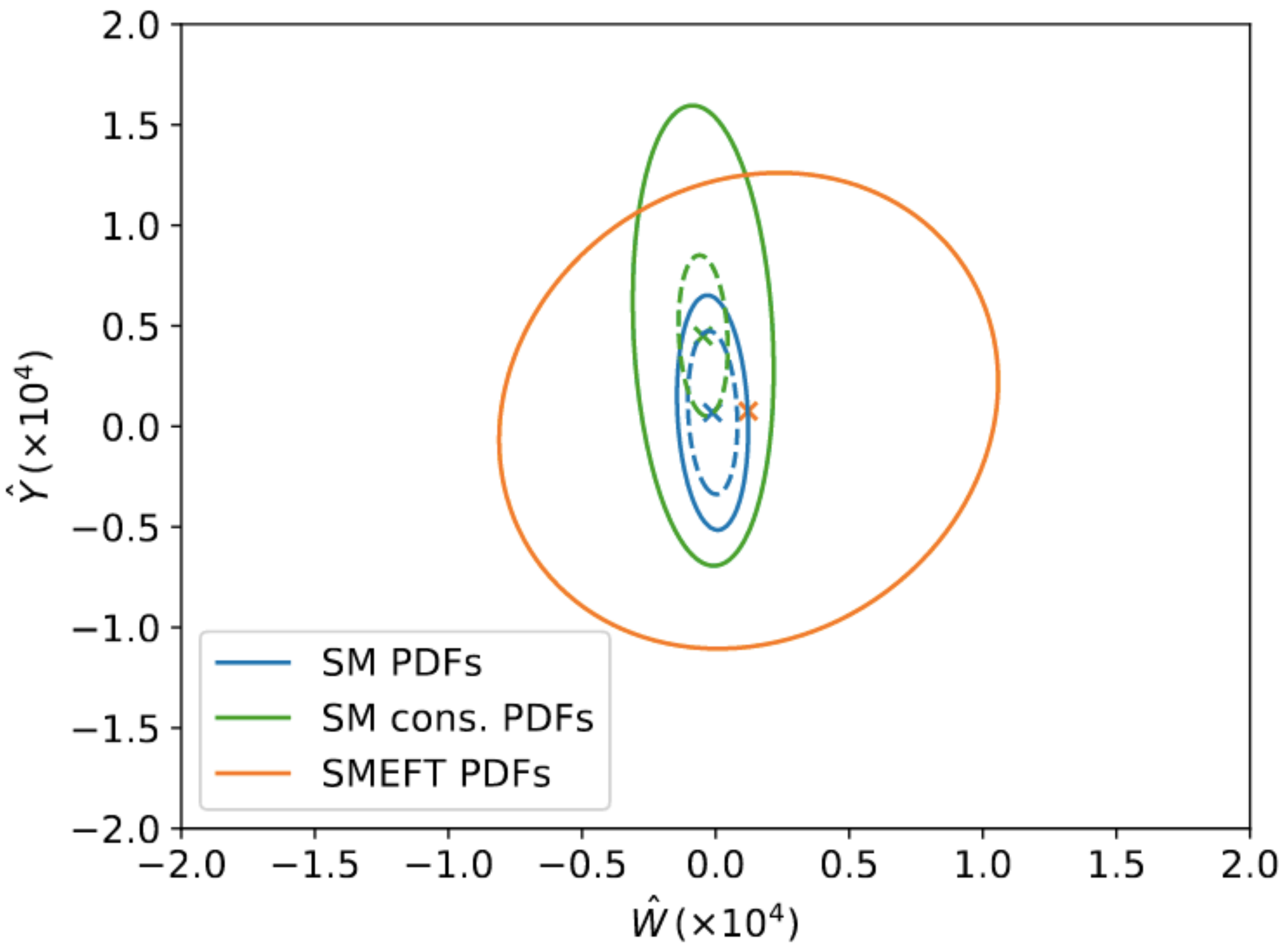
- *Question 2:* **Won't the PDF-SMEFT interplay be negligible?**
  - It depends on the scenario!

# 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.
  - 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**.



# 2. - 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.

- There are three main methodologies available: (i) a **scan** in Wilson coefficient space (see 1905.05215 and 2104.02723); (ii) the **CTEQ-TEA** methodology (see 2201.06586 and 2211.01094); (iii) the **SIMUnet** methodology (see 2201.07240).

# PDF-SMEFT interplay: methodology

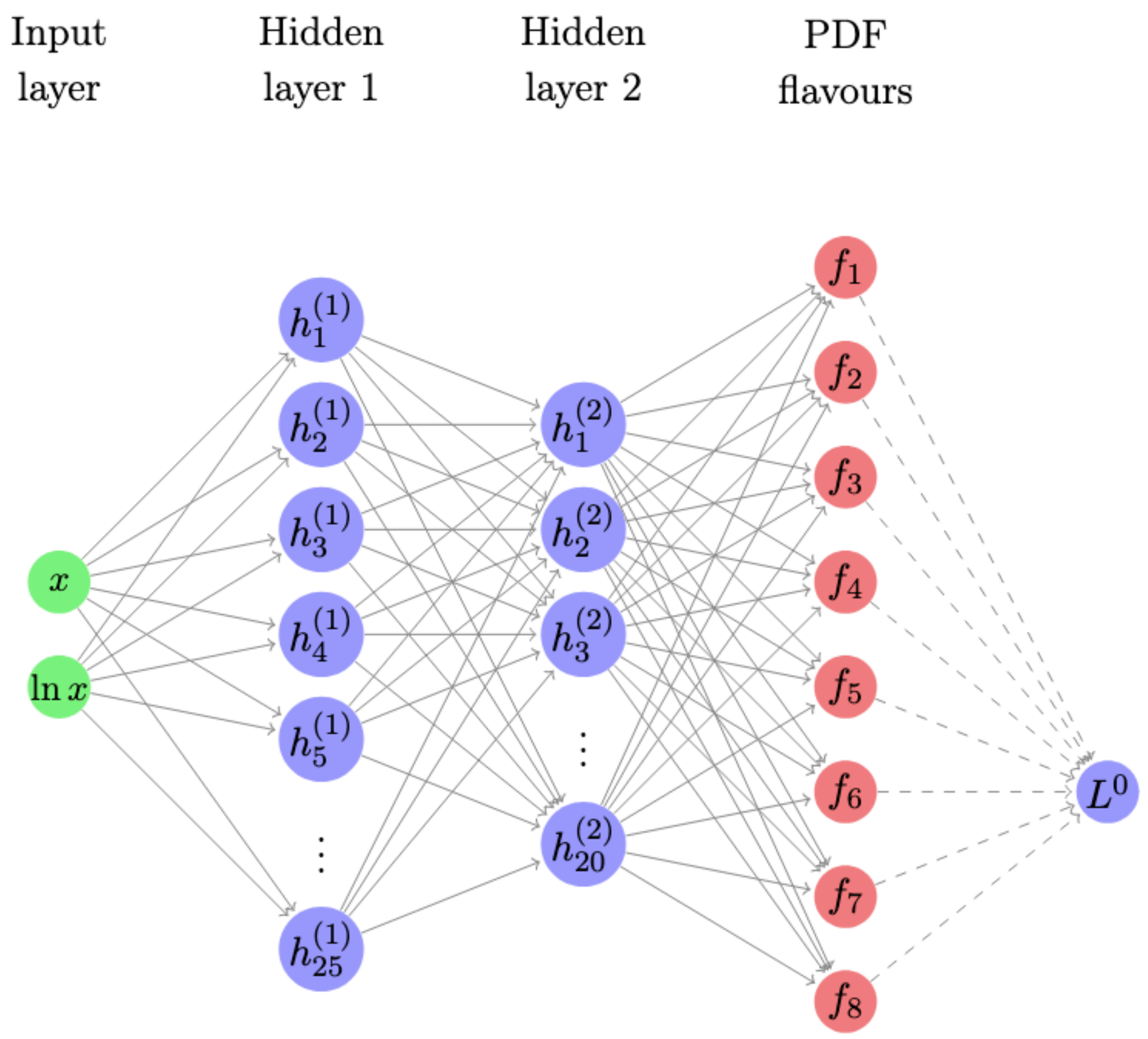
- With the need for simultaneous PDF-SMEFT determinations established, we now need an **efficient methodology** to perform the fits.

- There are three main methodologies available: (i) a **scan** in Wilson coefficient space (see 1905.05215 and 2104.02723); (ii) the **CTEQ-TEA** methodology (see 2201.06586 and 2211.01094); (iii) the **SIMUnet** methodology (see 2201.07240).

- We will focus only on **SIMUnet**.

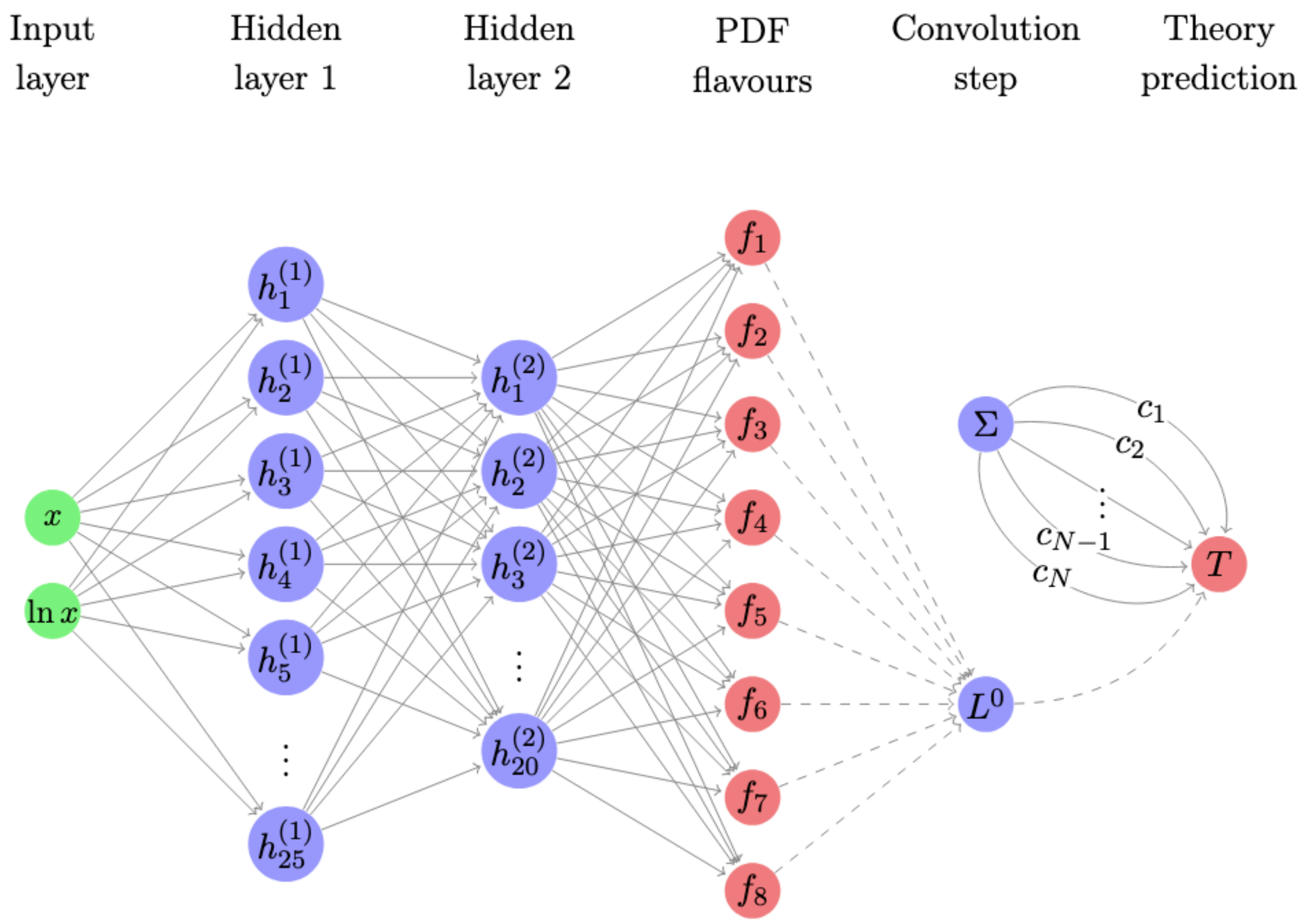
# The SIMUnet methodology: details

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



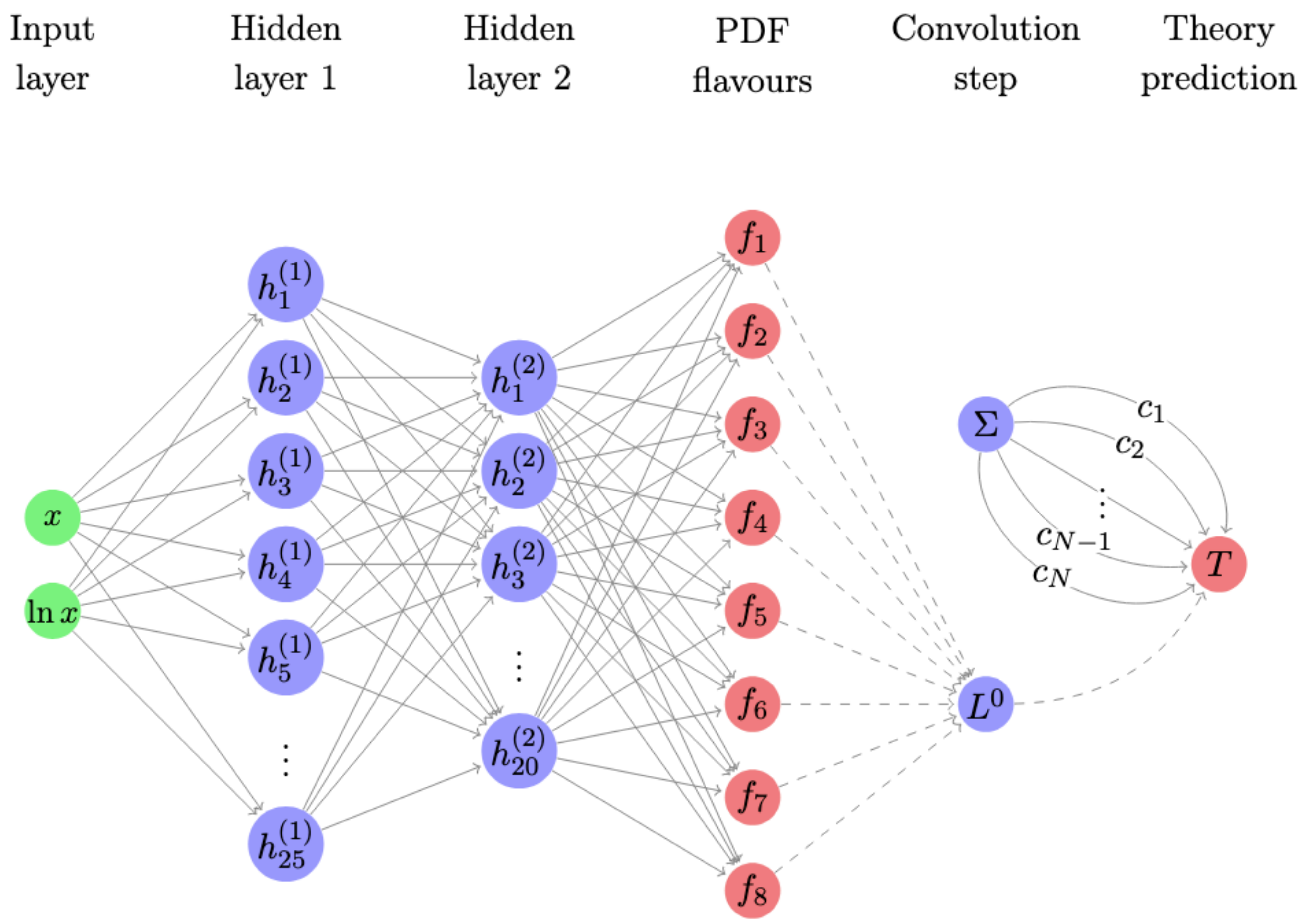
# The SIMUnet methodology: details

- 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**.



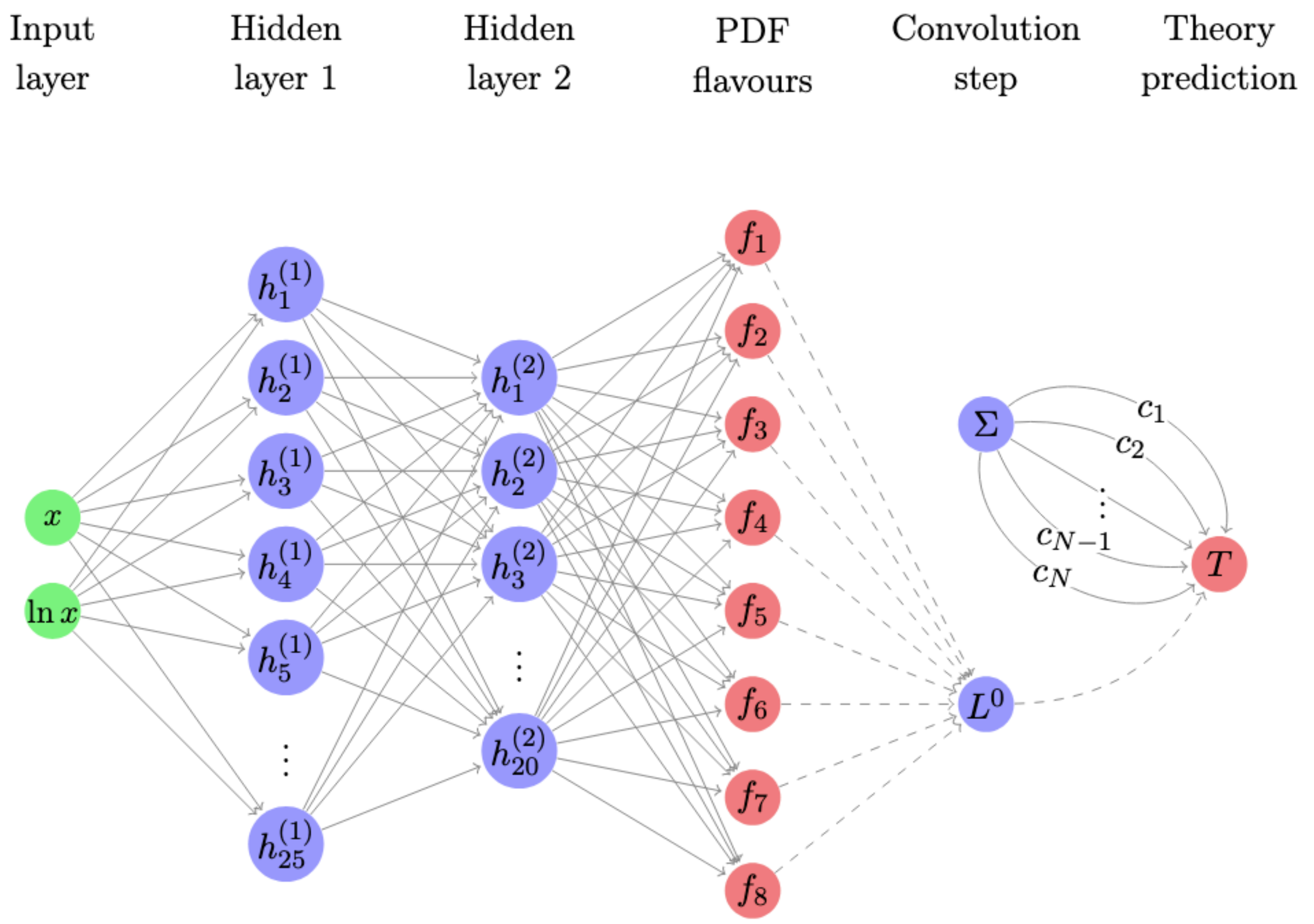
# The SIMUnet methodology: details

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



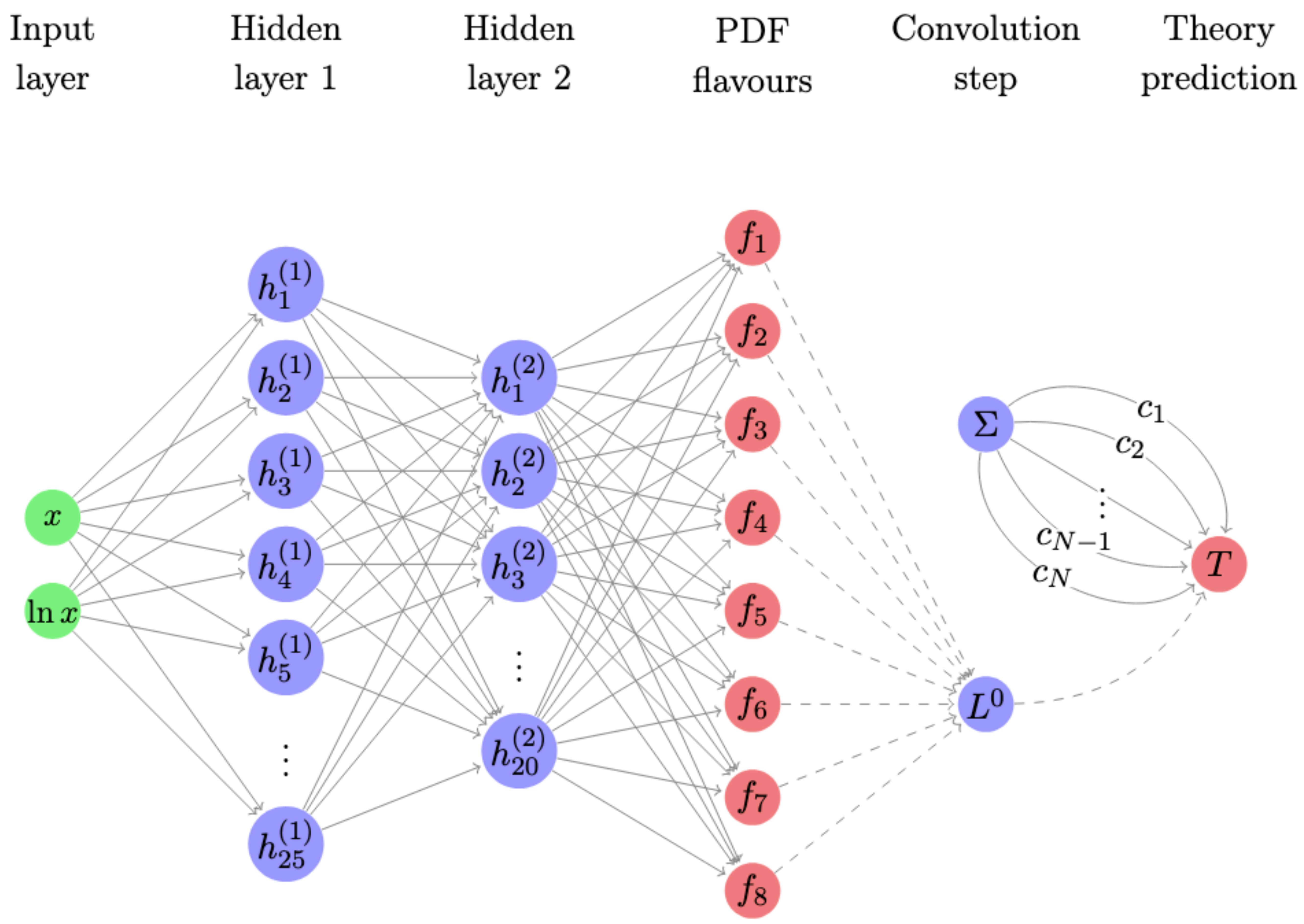
# The SIMUnet methodology: details

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



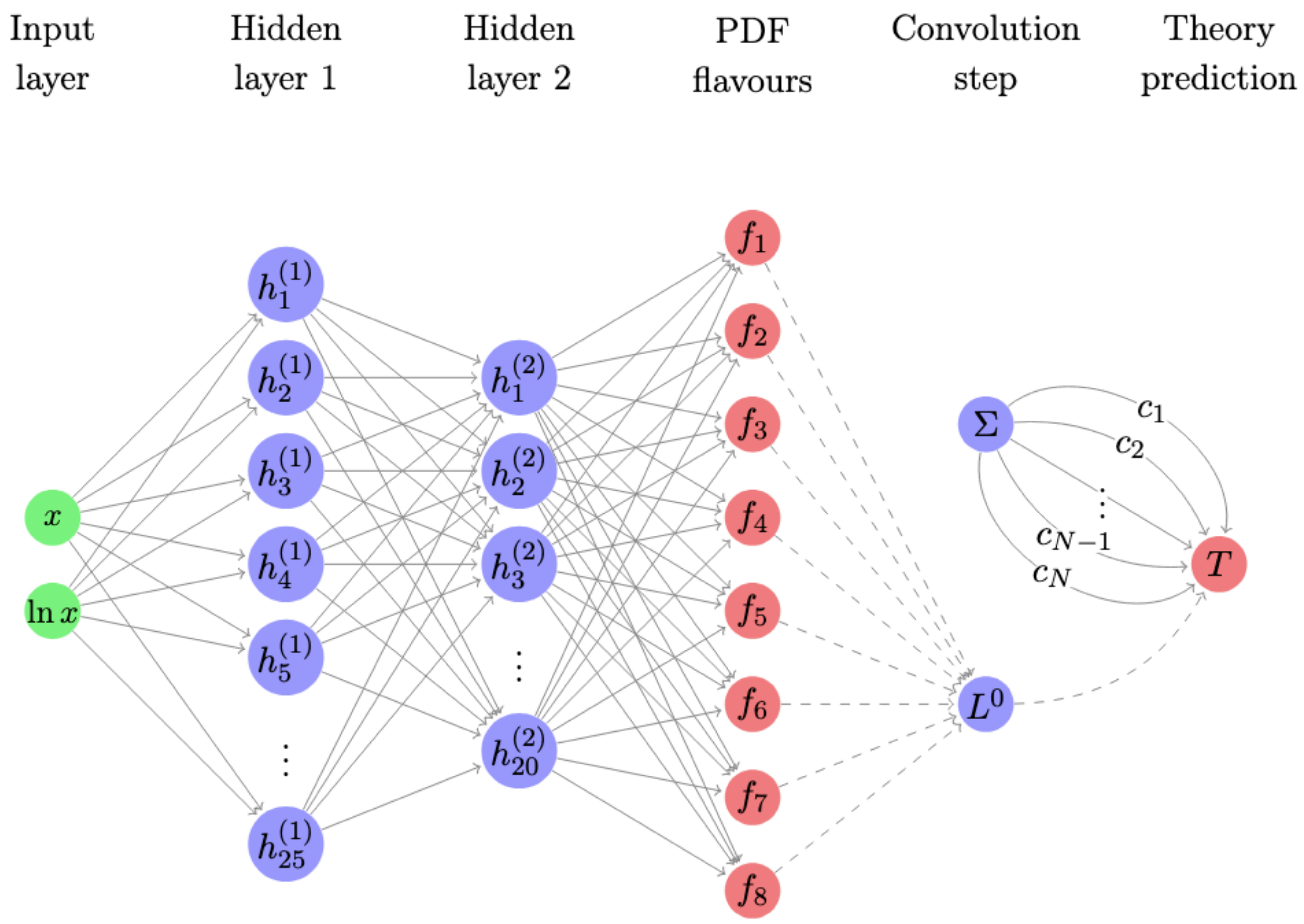
# The SIMUnet methodology: details

- 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**.



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

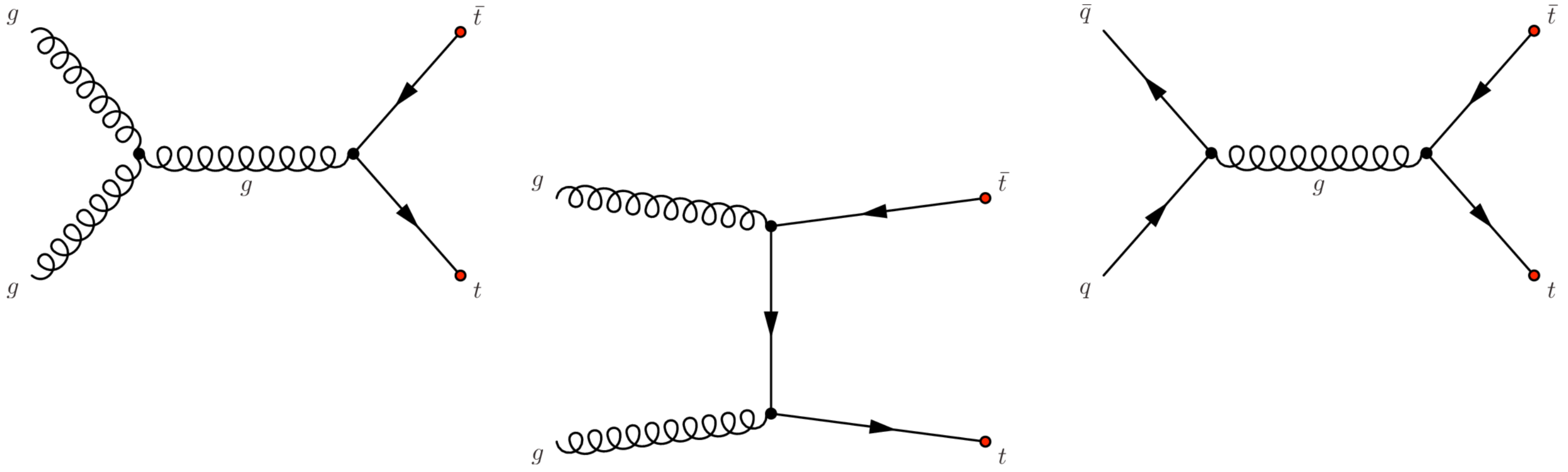
***Based on 2303.06159***

# 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

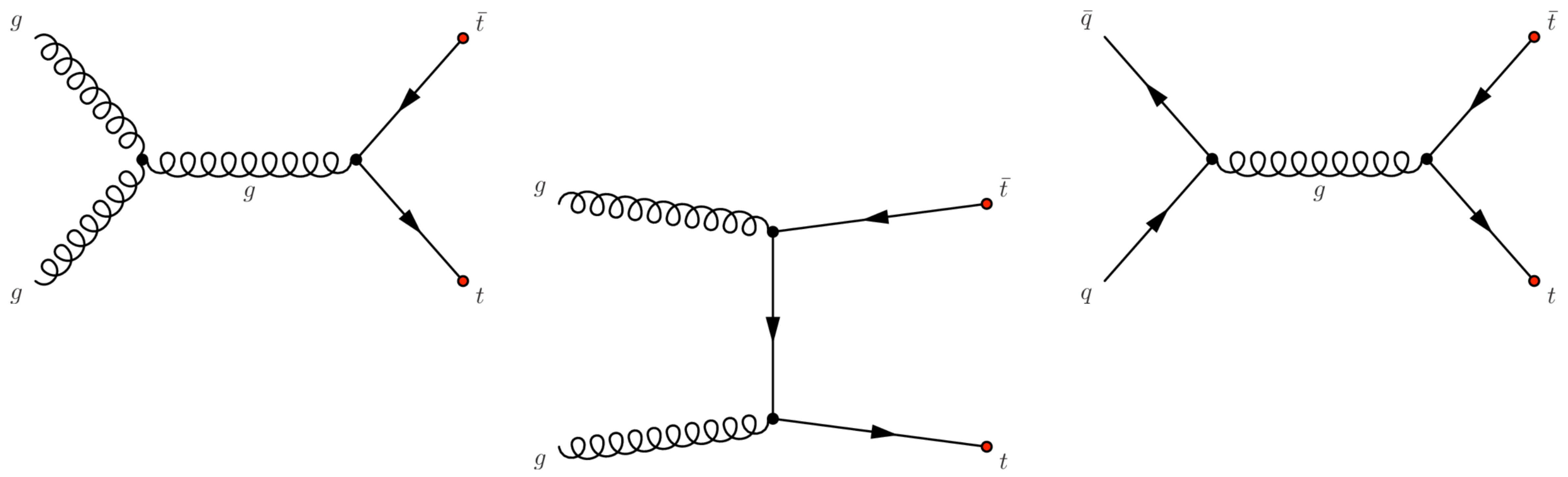
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



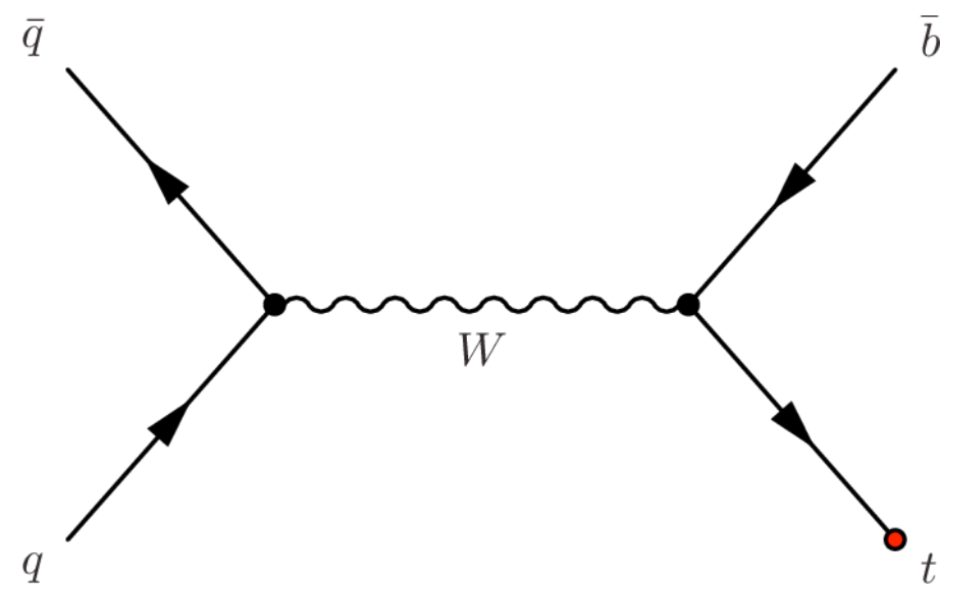

inclusive $t\bar{t}$

# Run II top quark data

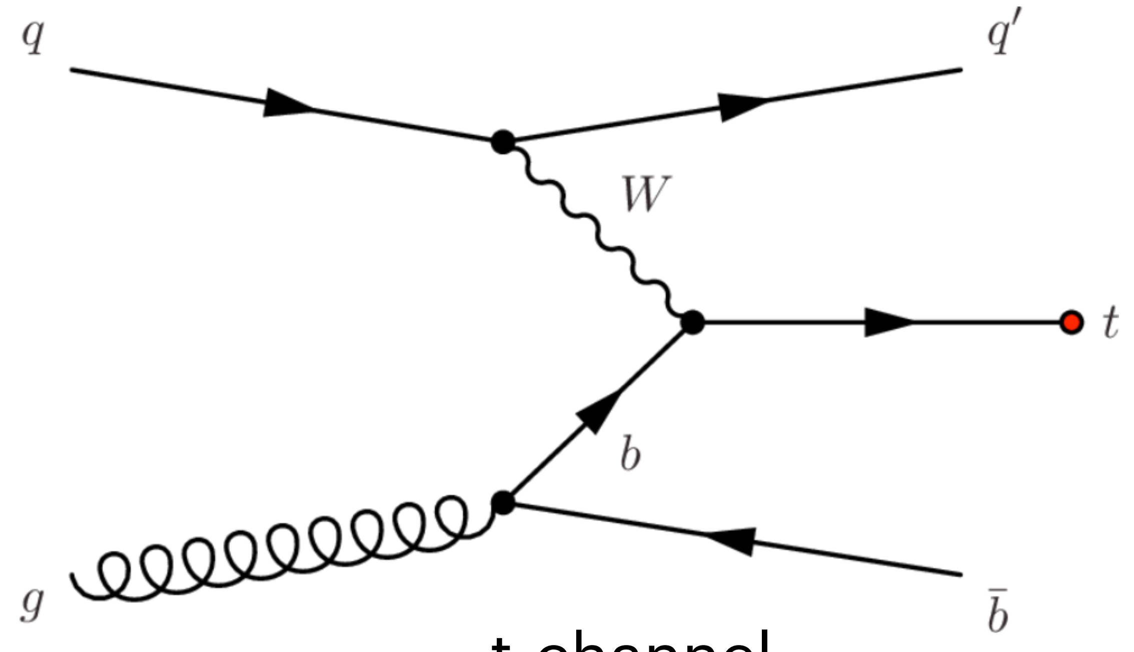
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:




inclusive $t\bar{t}$




s-channel

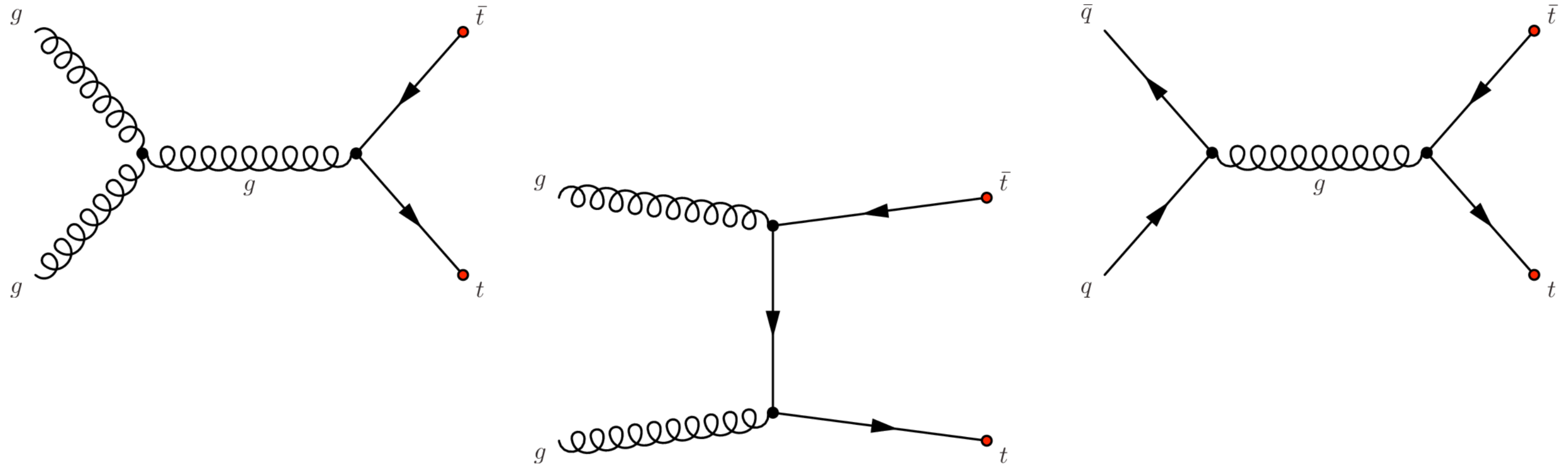



t-channel

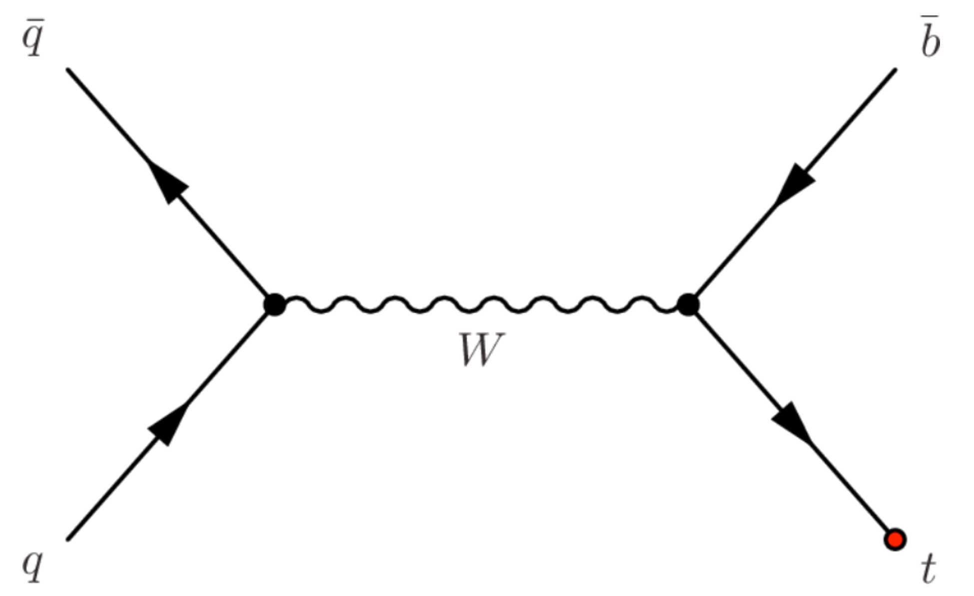
single top

# Run II top quark data

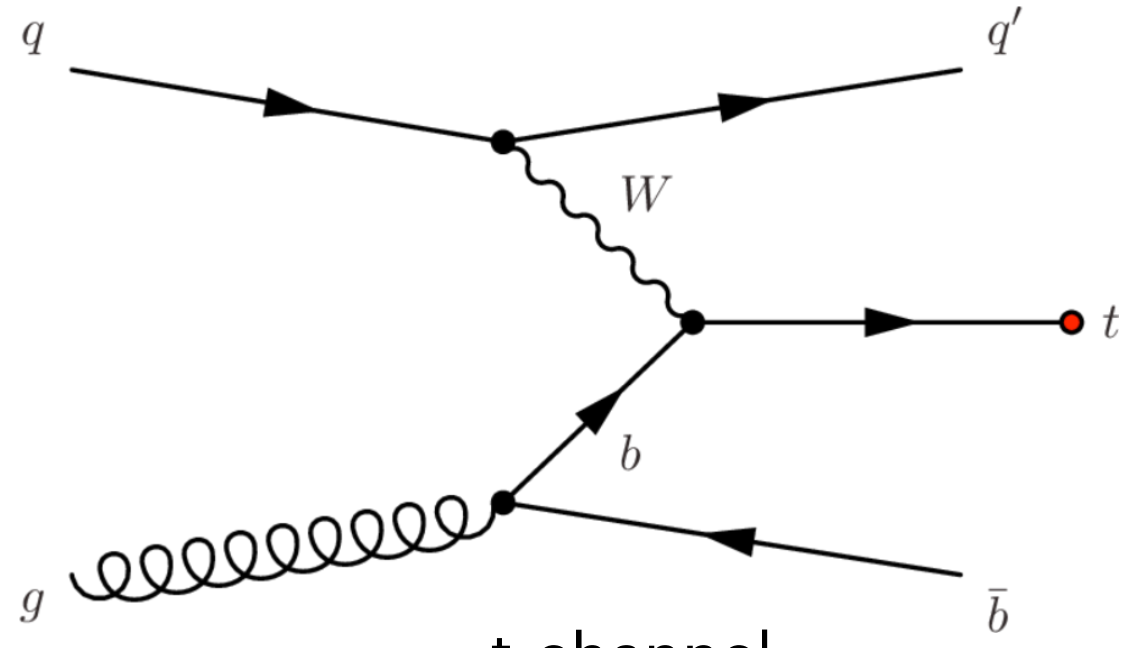
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:




inclusive $t\bar{t}$

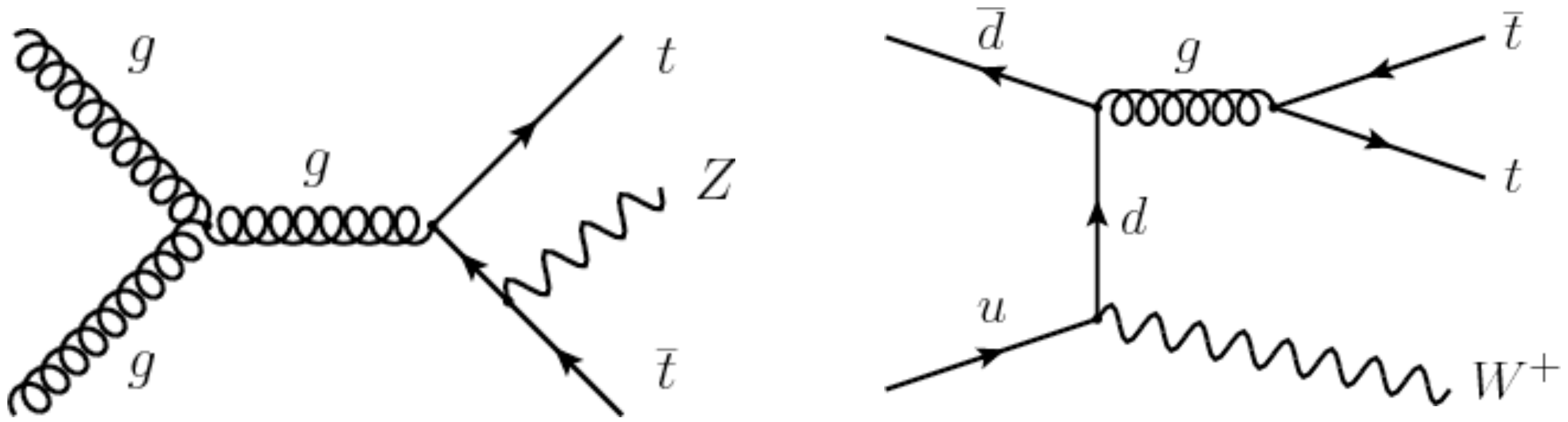



s-channel




t-channel

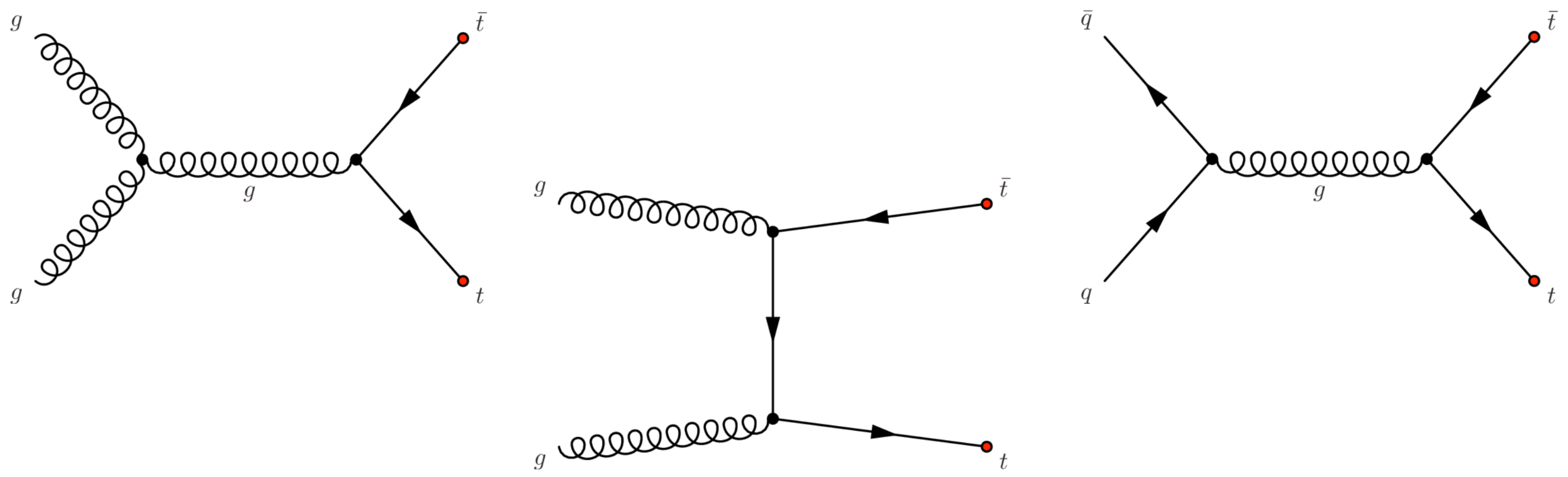
single top



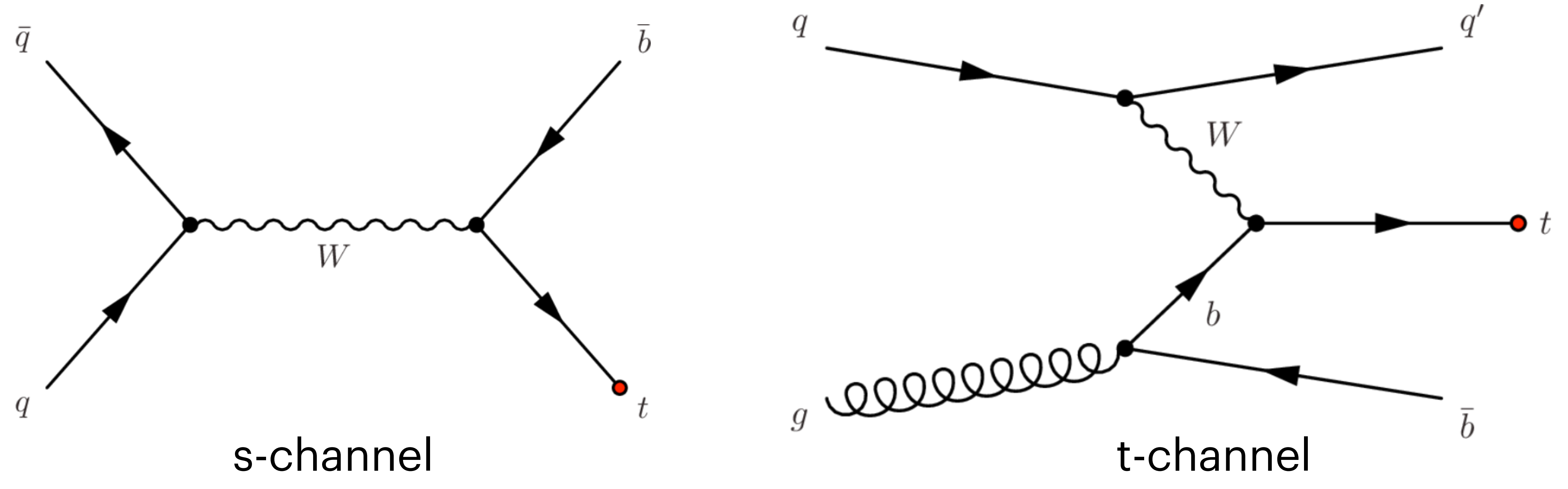

associated $t\bar{t}$

# Run II top quark data

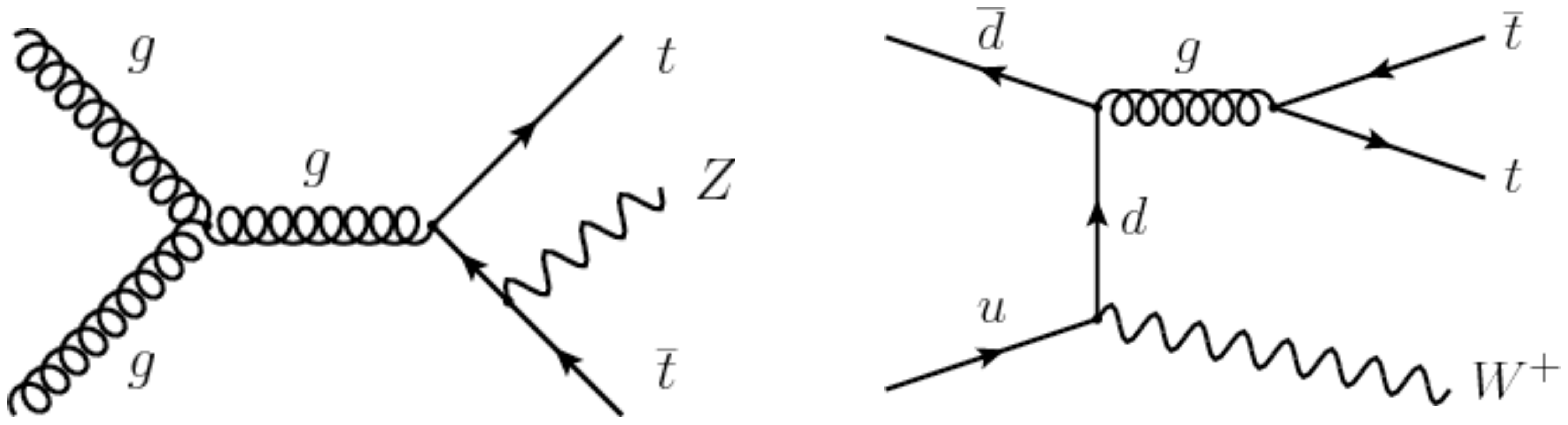
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



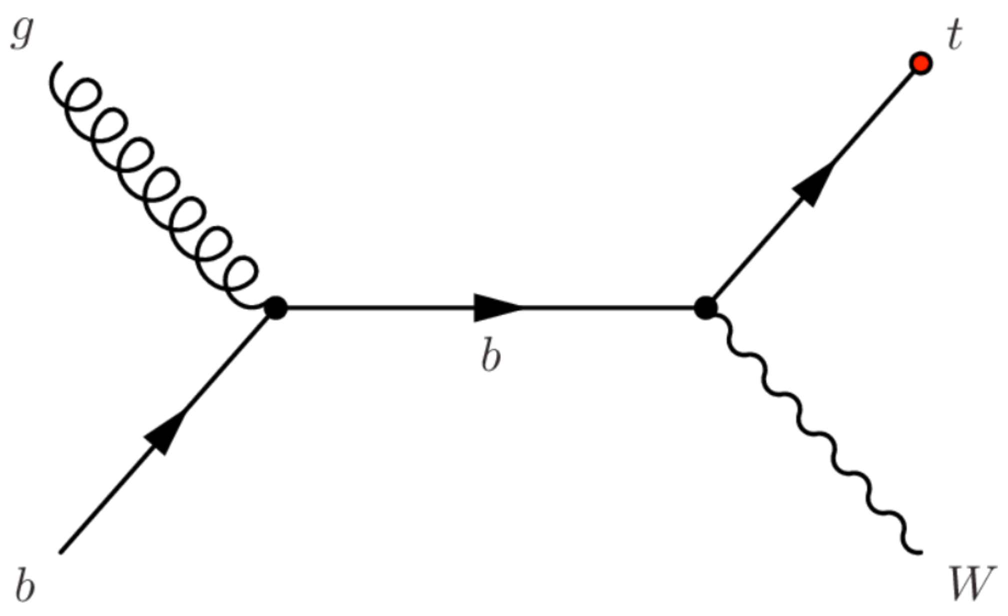
inclusive $t\bar{t}$




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**:

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

dipoles

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

currents

# 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

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

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**:

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

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

$c_{qd}^{8}$ $c_{qq}^{8,3}$ $c_{dt}^{8}$ $c_{qt}^{8}$ $c_{qq}^{8,1}$ $c_{qu}^{8}$ $c_{ut}^{8}$

four-fermion octets

# 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)

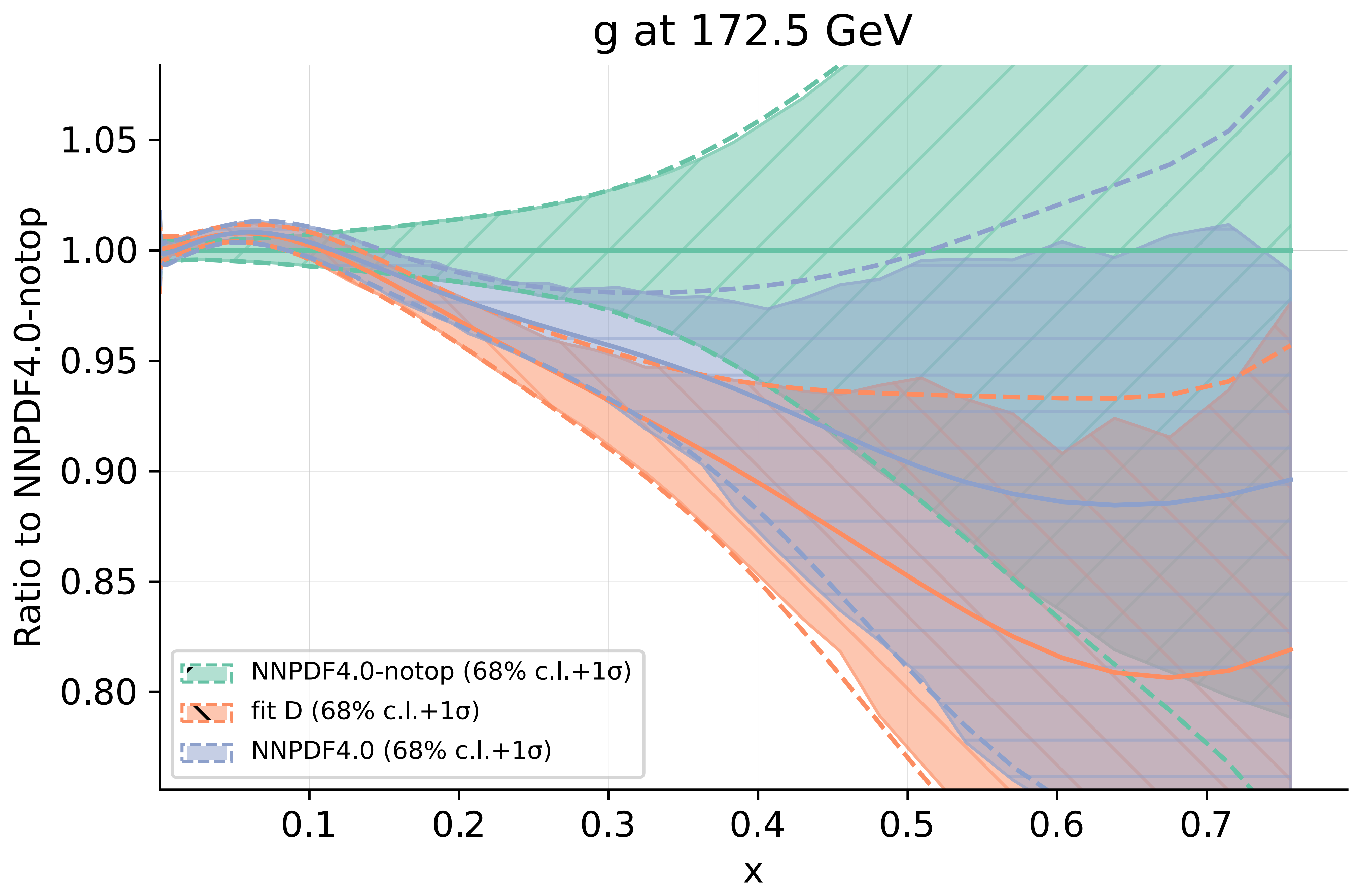
# Fit settings

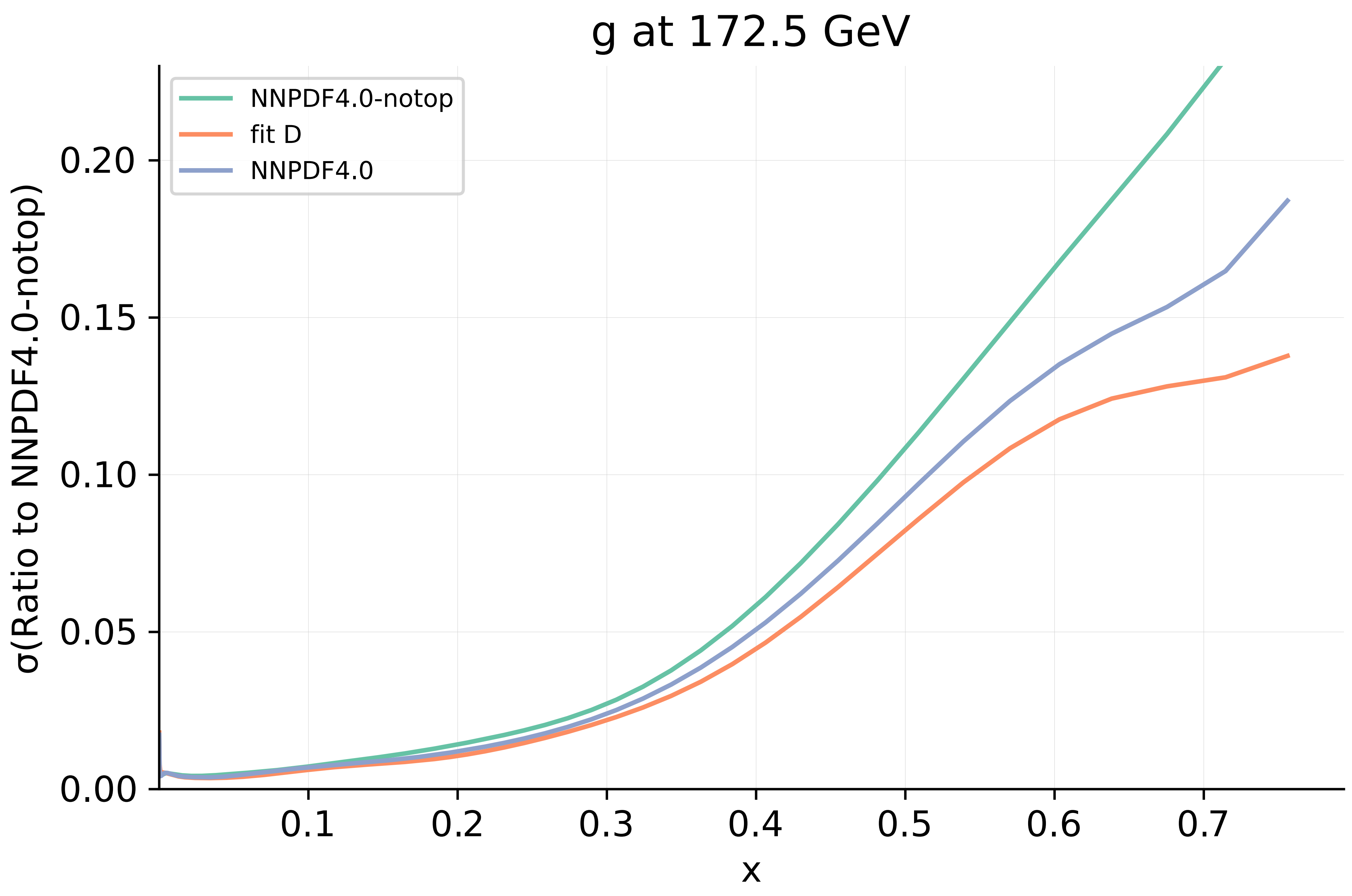
- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **175 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset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

# Let’s start the results with the PDF-only fits...

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

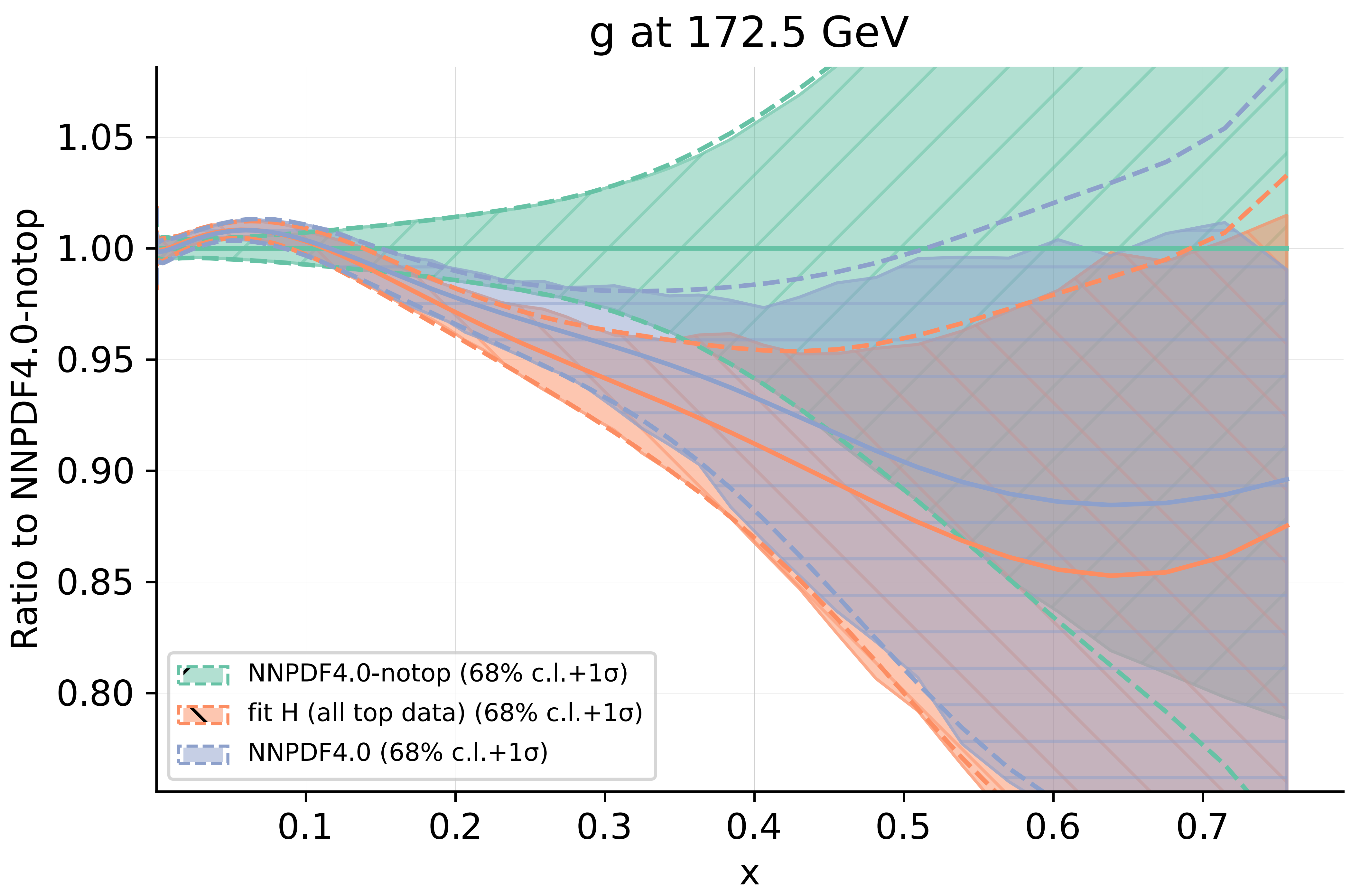
- First, we consider the impact of our dataset on PDFs **in the SM**.
- Begin by considering the updates to the **inclusive** $t\bar{t}$ and **single-top** dataset relative to NNPDF4.0. If we perform a SM PDF fit using only our new inclusive $t\bar{t}$ and single-top data, we see a more pronounced effect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

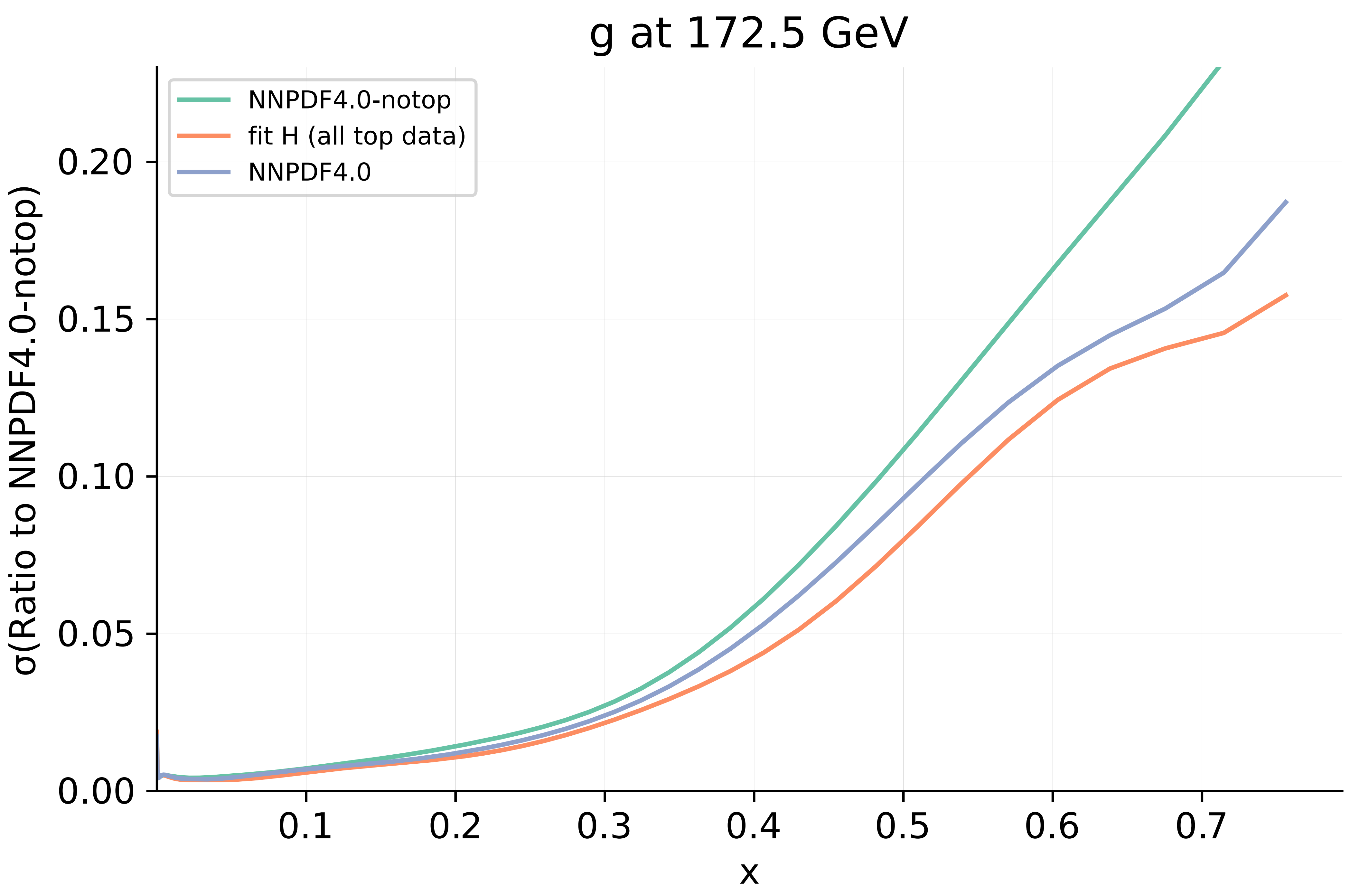


# PDFs in the SM - impact of all new top data

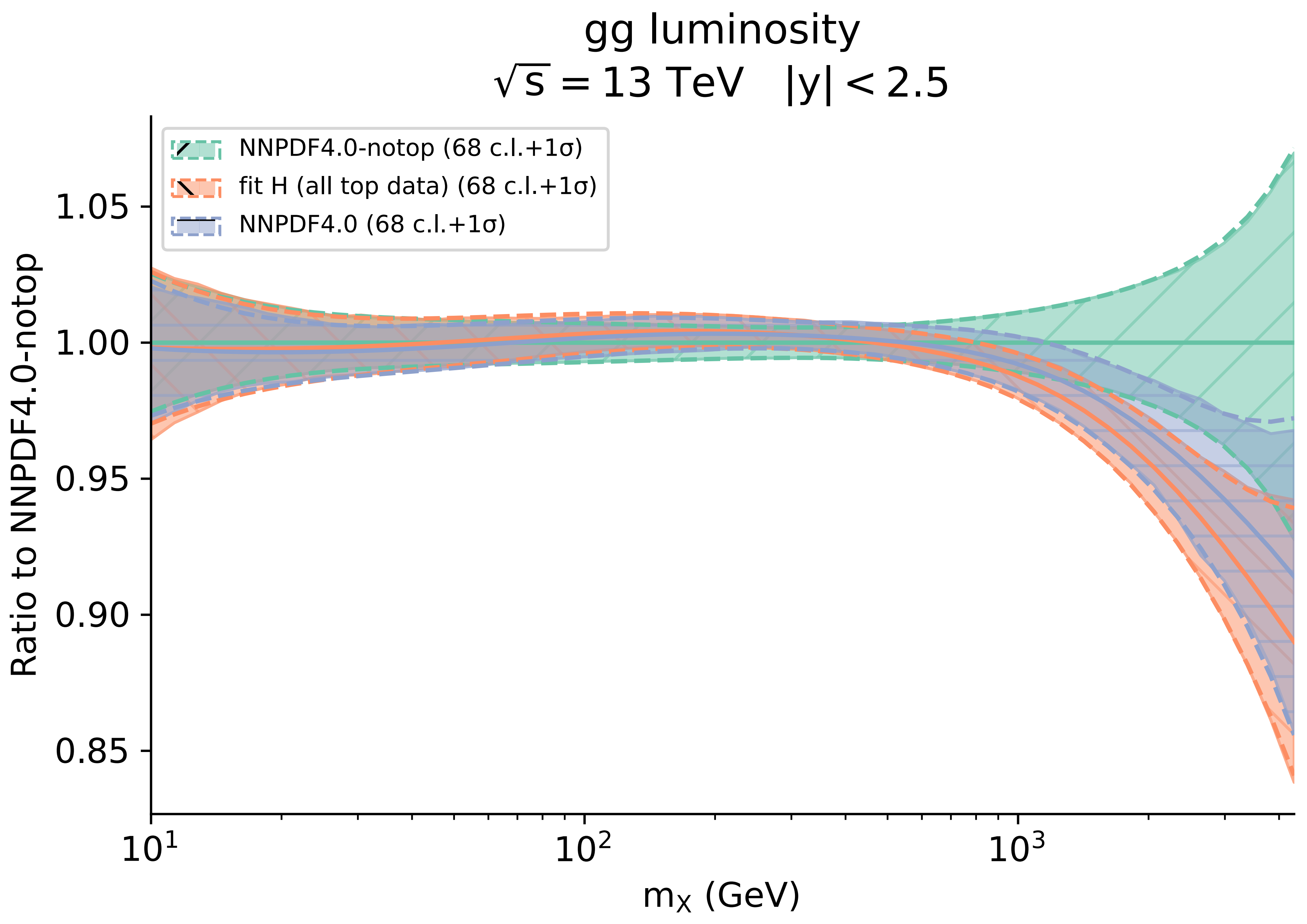
- Finally, we present the results of a **complete PDF fit** including **all our new top data**. As expected, the effect on the large-$x$ gluon is broadly the same as the effect of just including the inclusive $t\bar{t}$ and single-top data, but is mildly tempered by the associated top data.

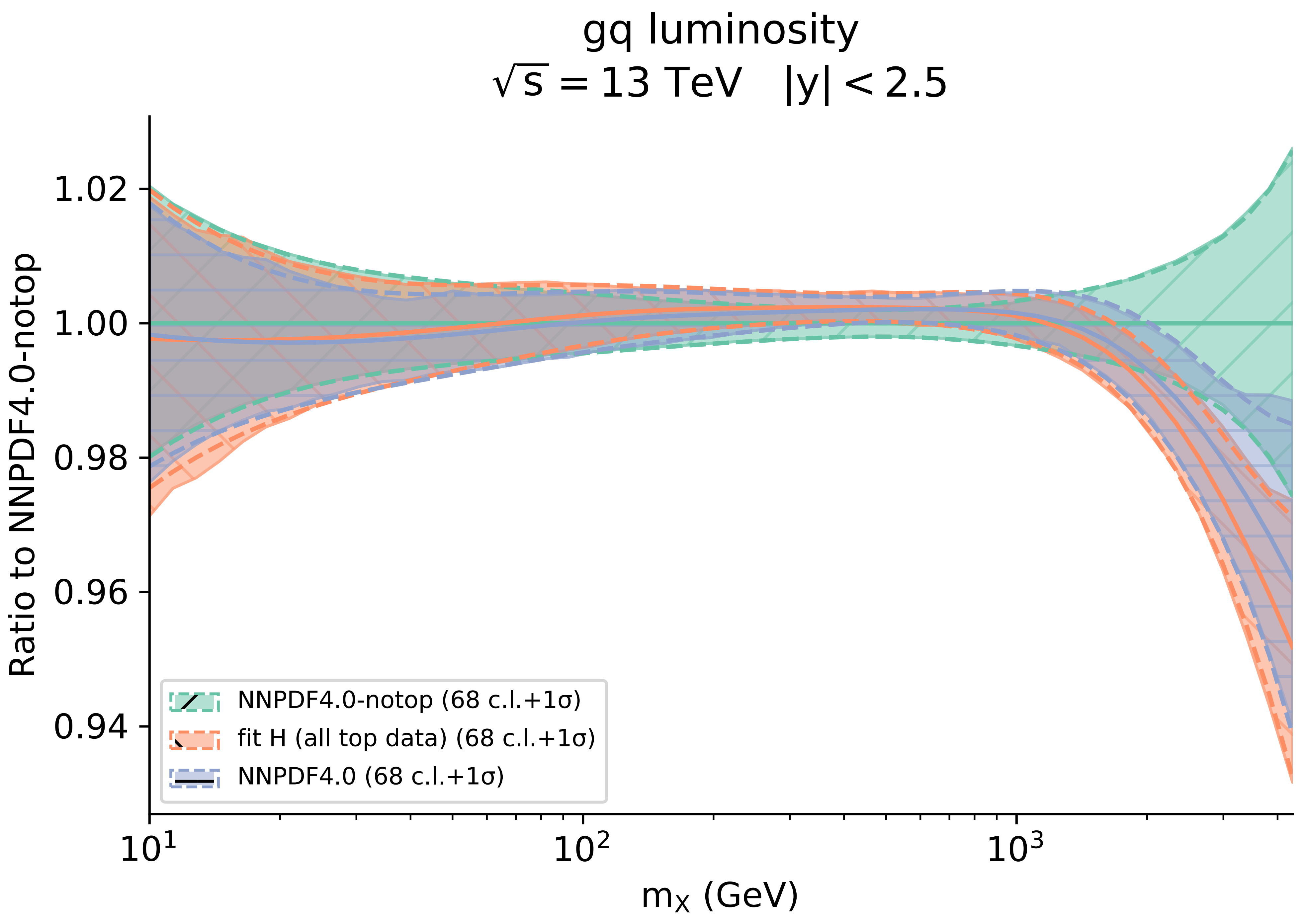


# PDFs in the SM - impact of all new top data

- 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.



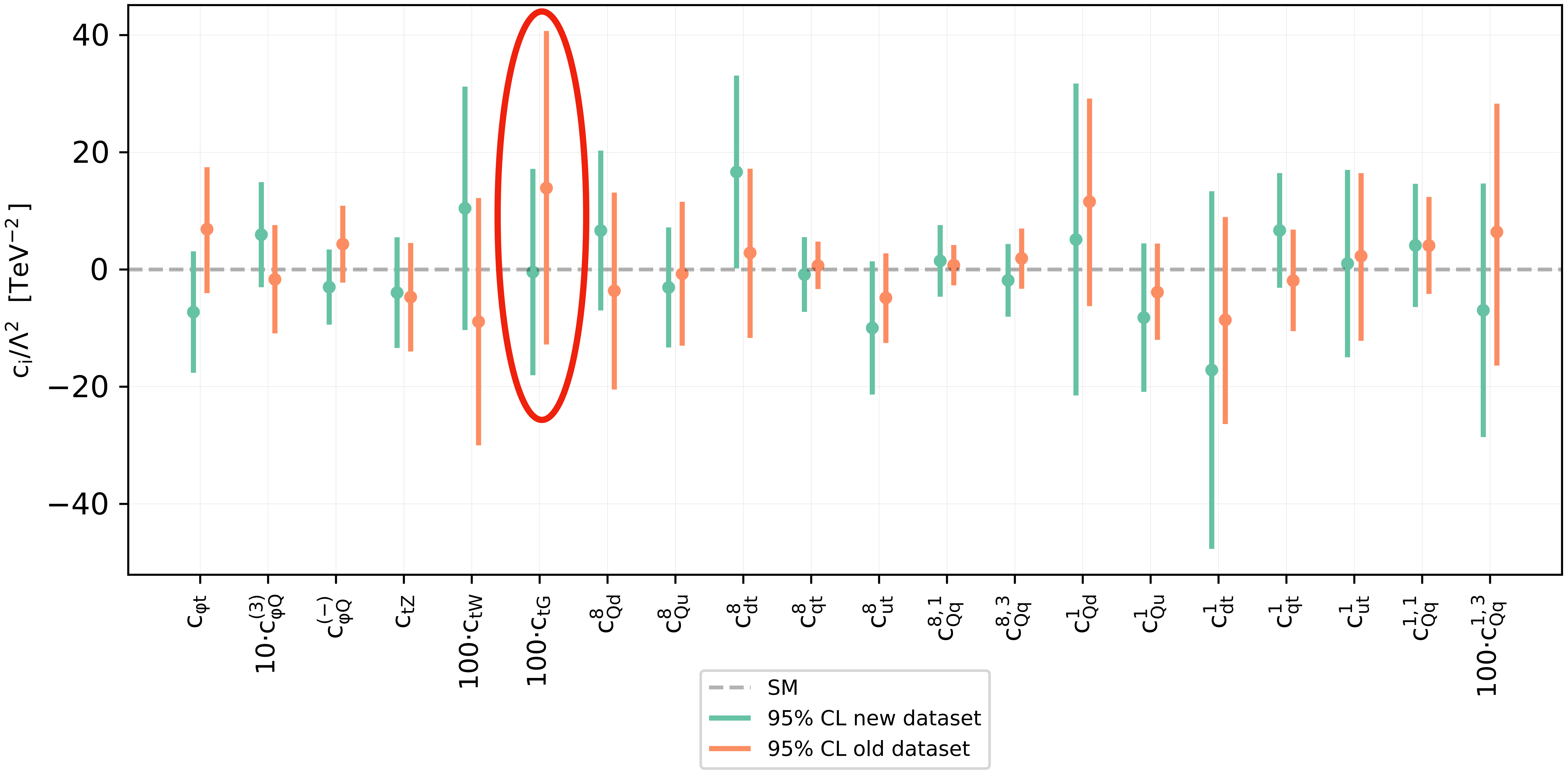



# Now let’s see the SMEFT-only fits…

# 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.

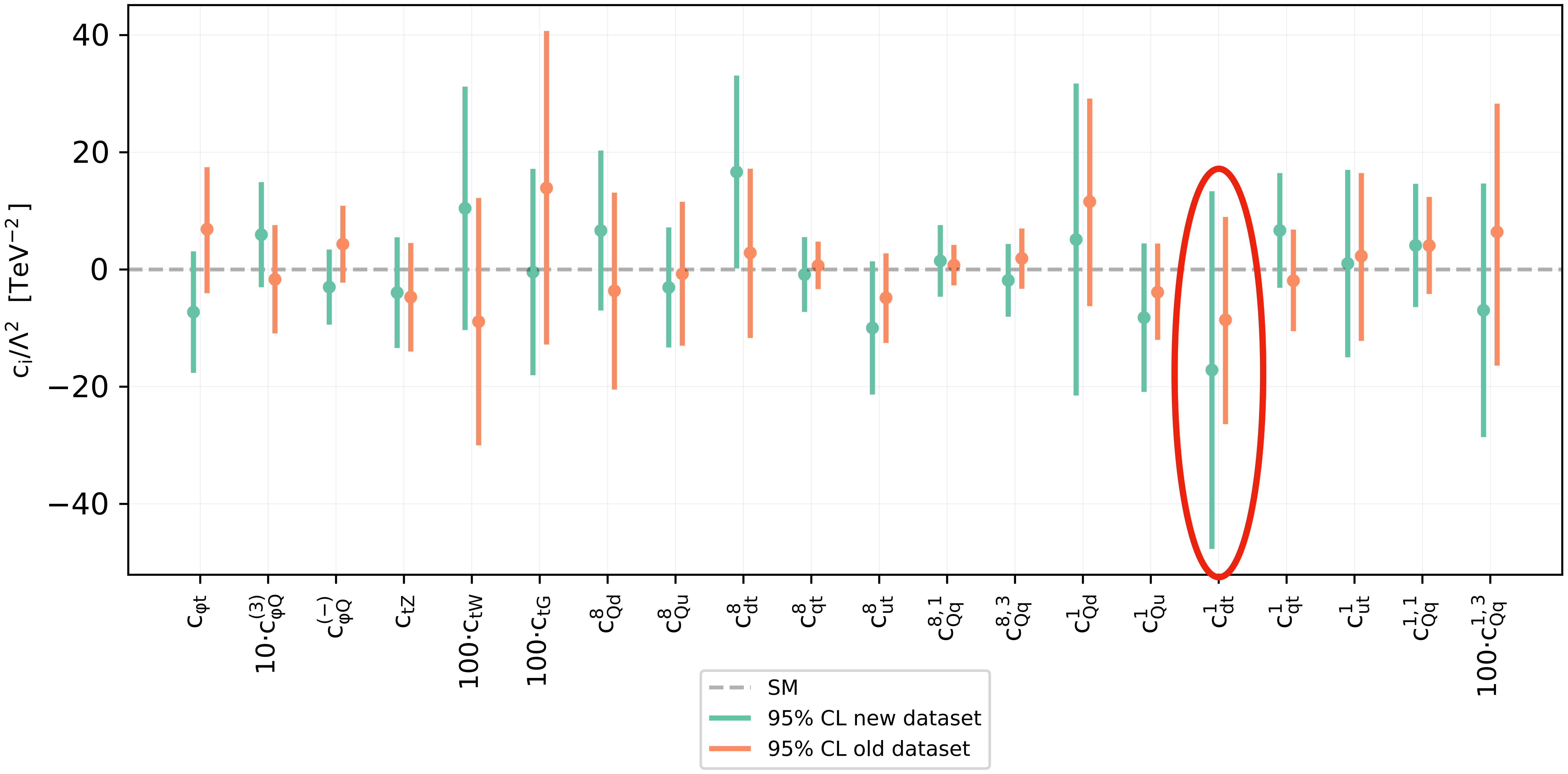


# SMEFT-only fits: linear SMEFT

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

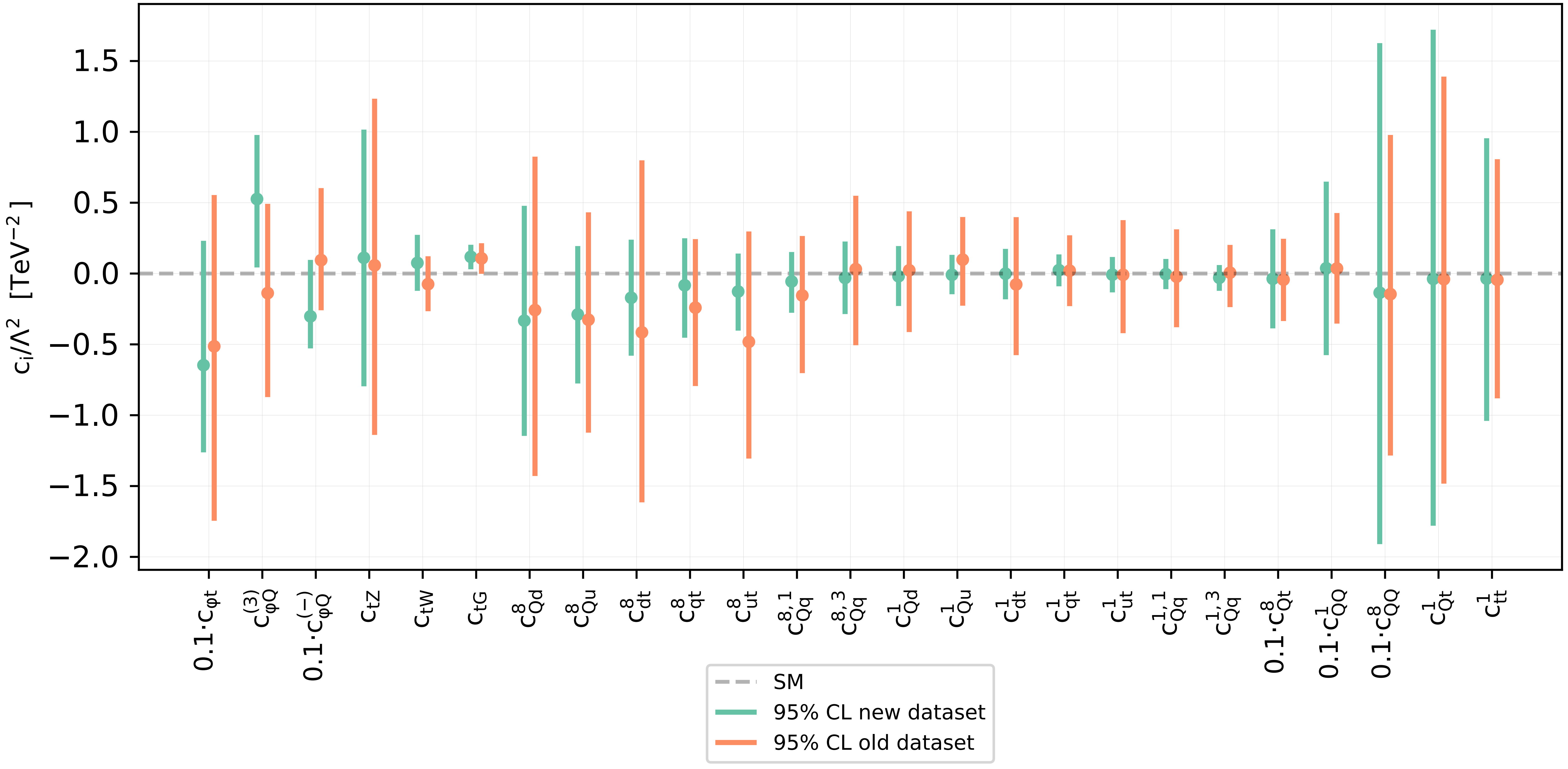
- Some coefficients have **broader bounds** than previously obtained, in particular some of the four-fermion operators.

- 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

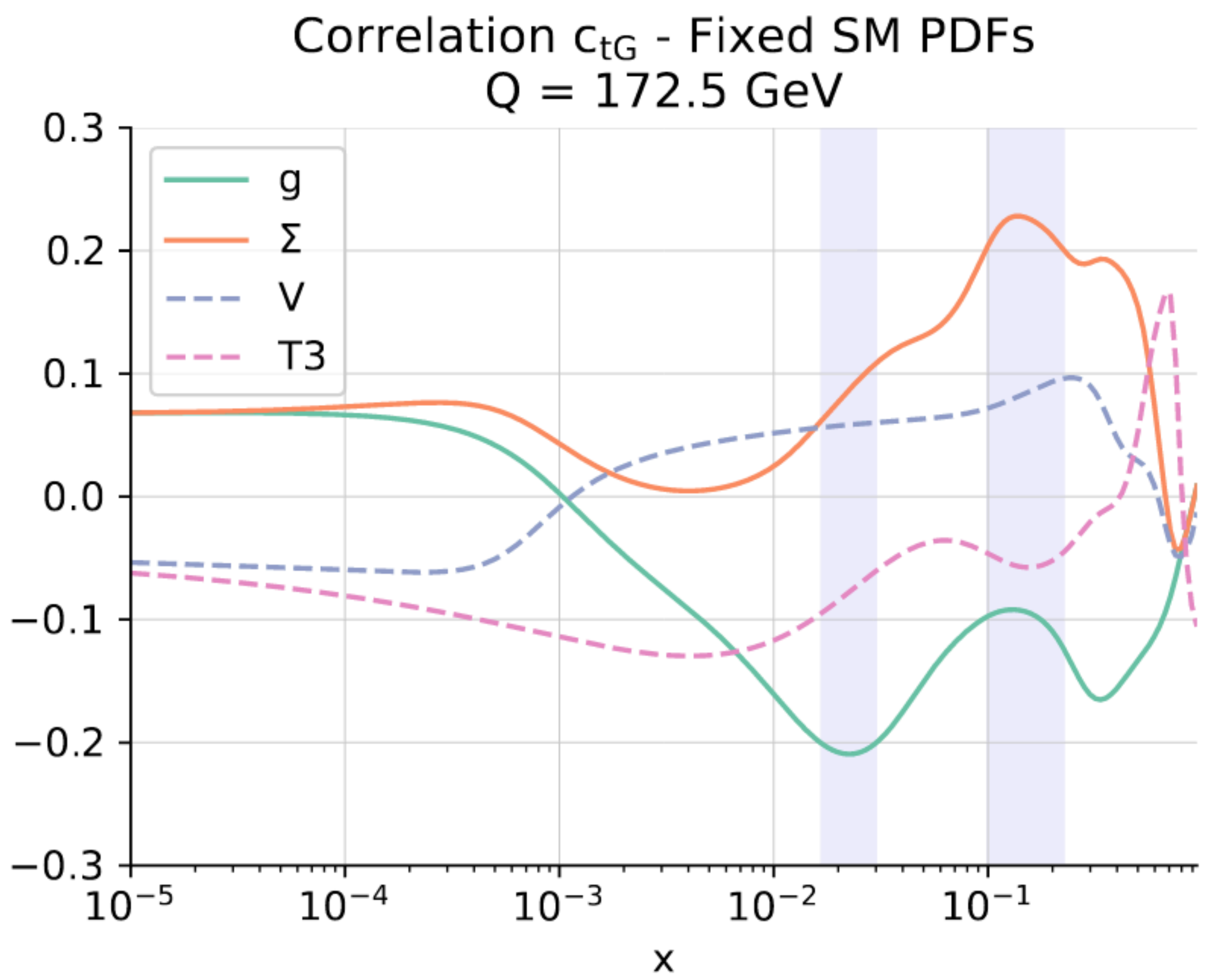
- We can try to get intuition for the result of the **joint PDF-SMEFT** fit by considering the **PDF-SMEFT correlation** in the SMEFT-only fits.

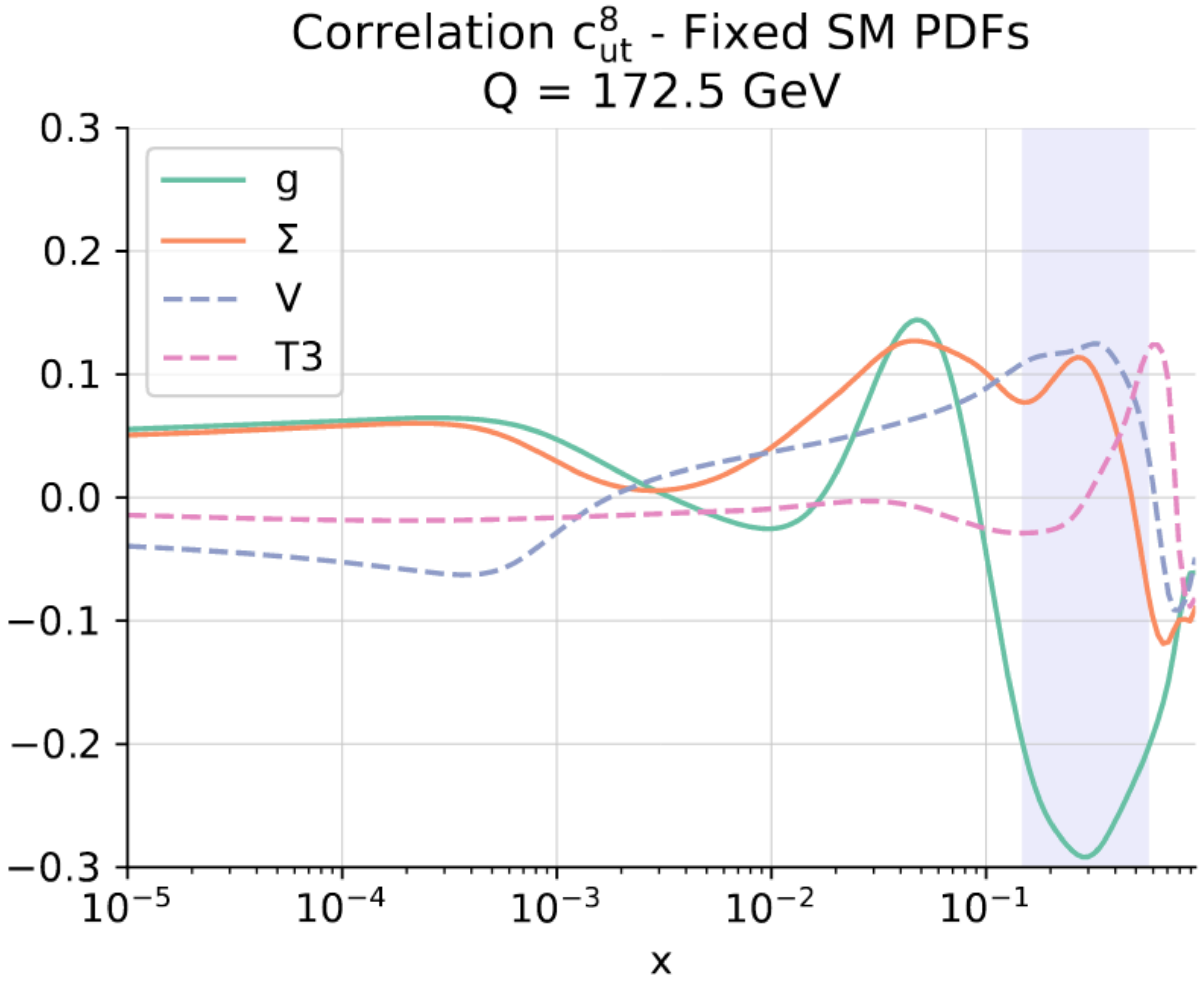
- This is defined for each Wilson coefficient and each PDF flavour by:

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

# 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**.

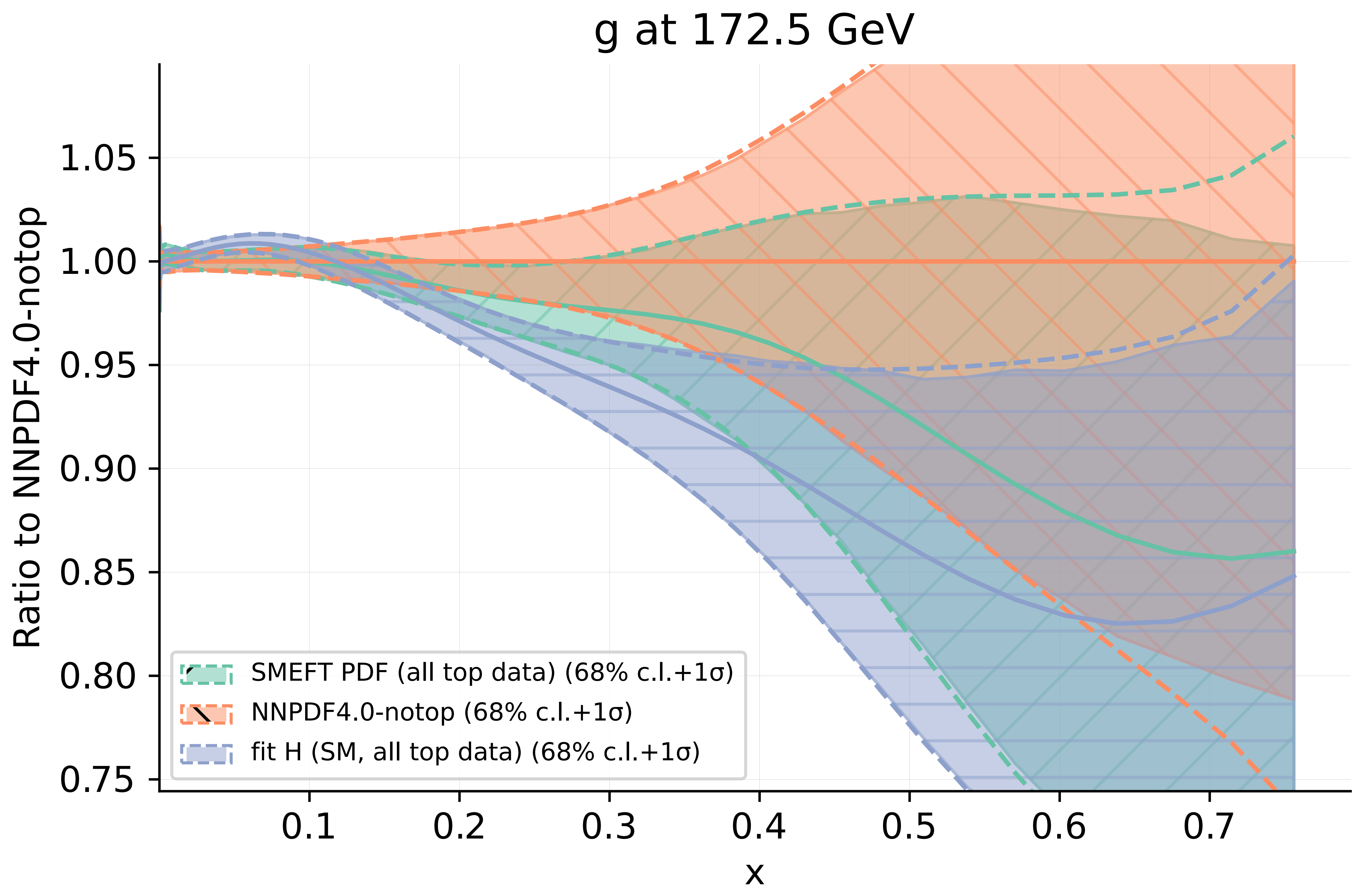


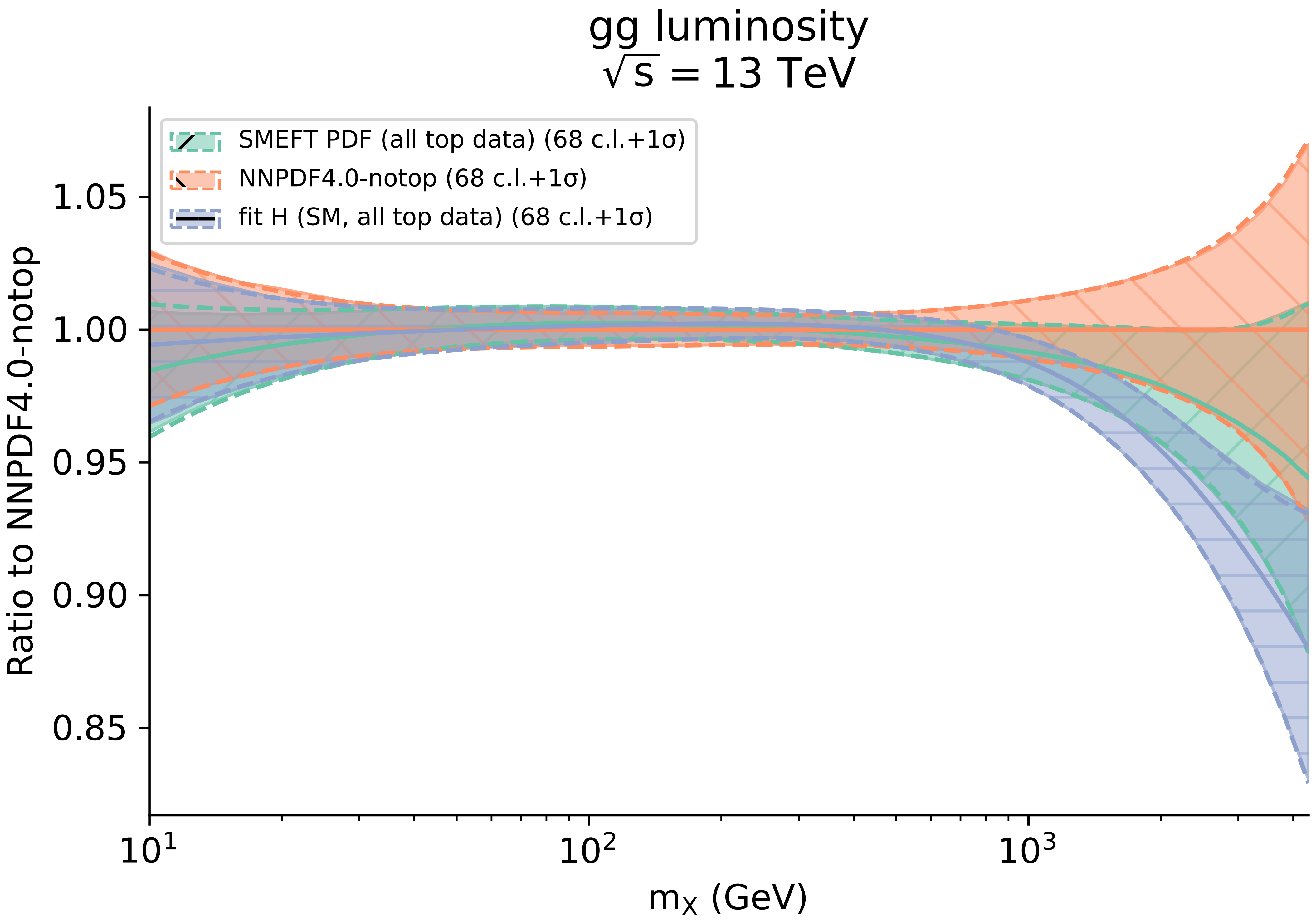



# Now, let's do the joint fit...

# Joint PDF-SMEFT fits: linear SMEFT

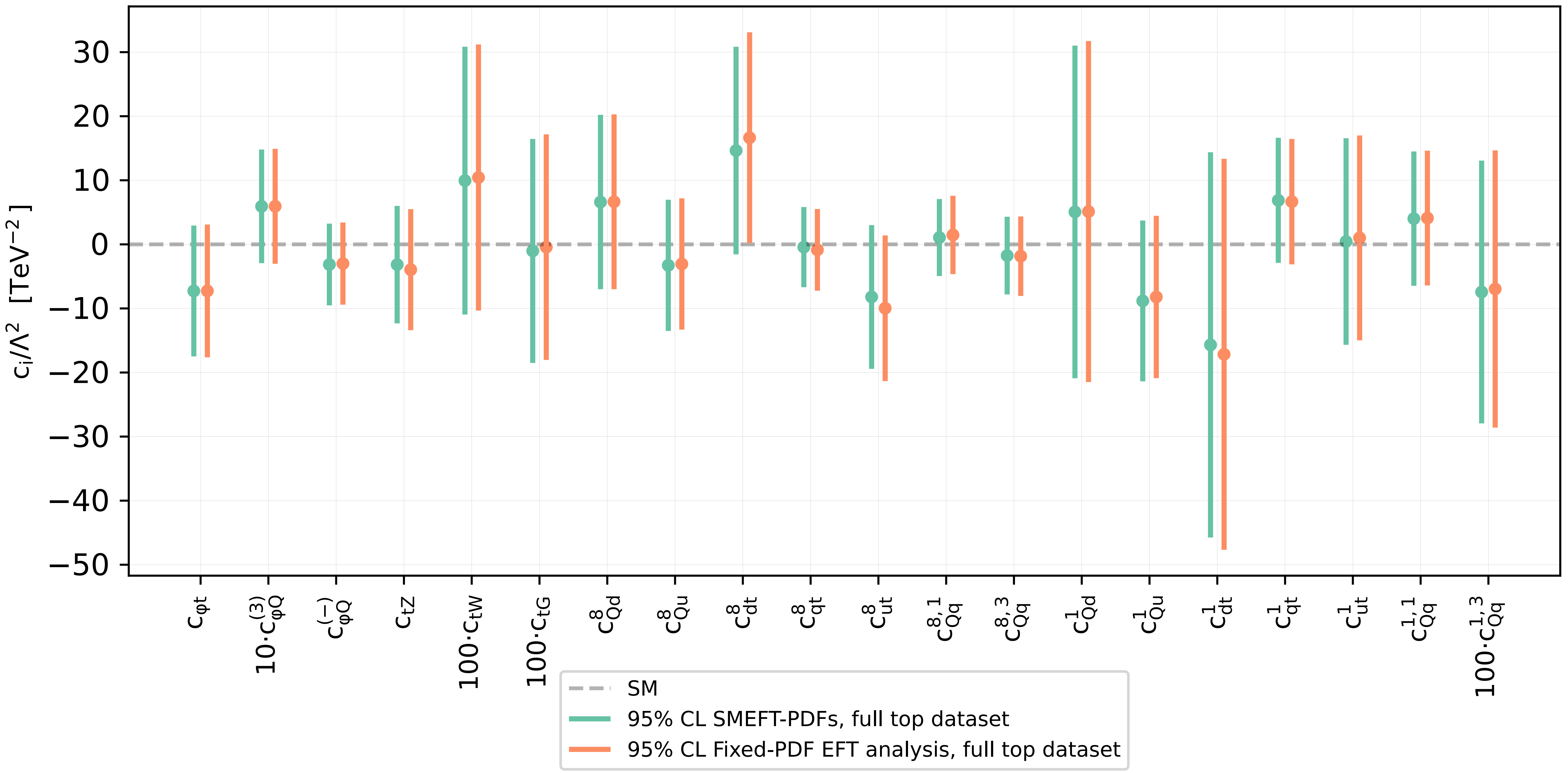
- Finally, we present the key result of the work: a **simultaneous** determination of PDFs and SMEFT Wilson coefficients. We start assuming **linear SMEFT**.
- In terms of the gluon PDFs and luminosities, we find that a simultaneous determination **reduces the pull** of the top data from the **non-top 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 $\sigma^2$, 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 $\chi^2$-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, $t_{\text{min}}$ 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, $t_{\text{min}}$ 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, $t_{\text{min}}$ 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:

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

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

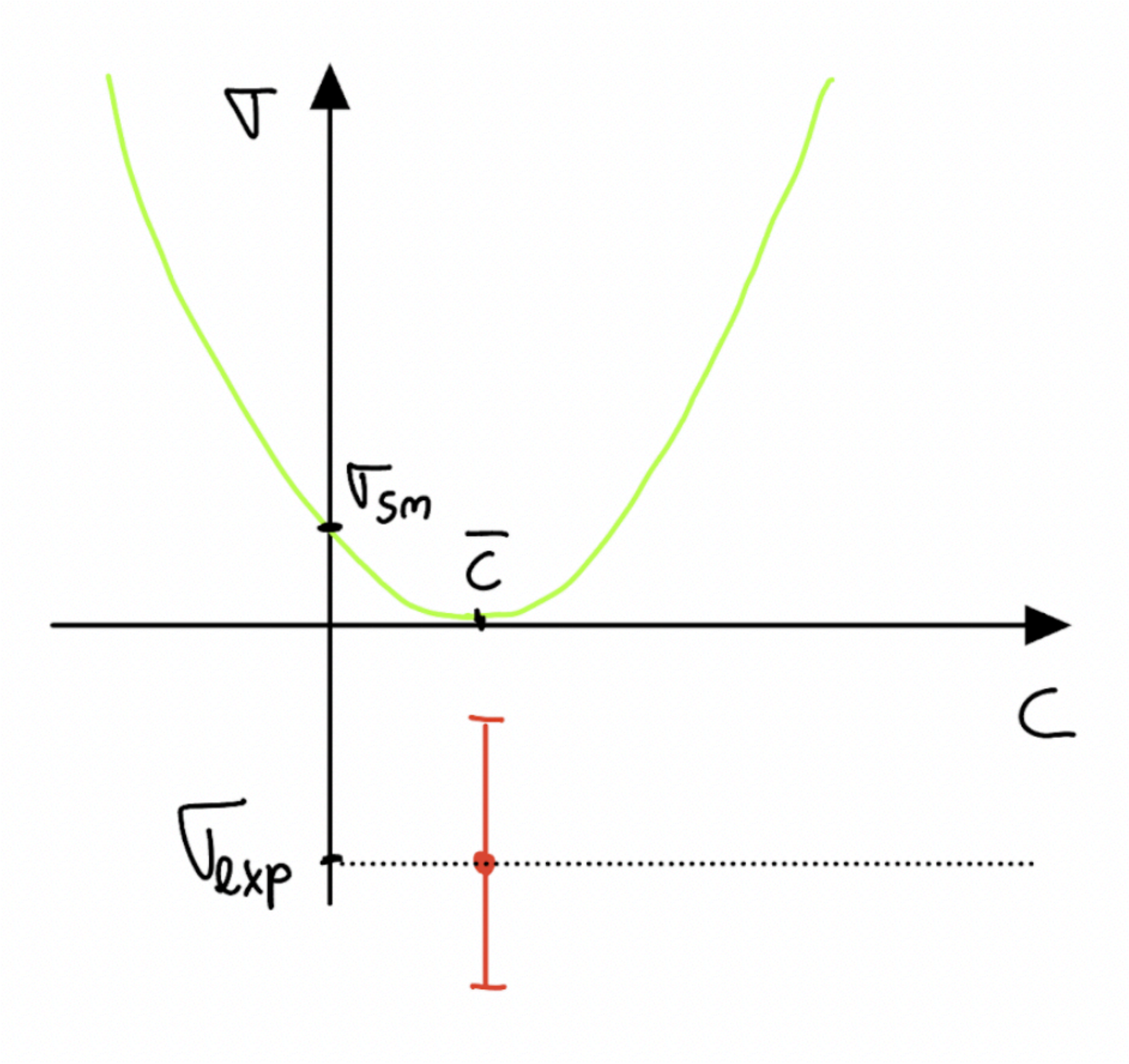
- **Key features to note:**
  - Part of the distribution looks like a **scaled version** of what we would expect from a **Bayesian method with uniform prior**.
  - There is also a **delta function spike** in the distribution - interesting to ask: why...?

# 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**.

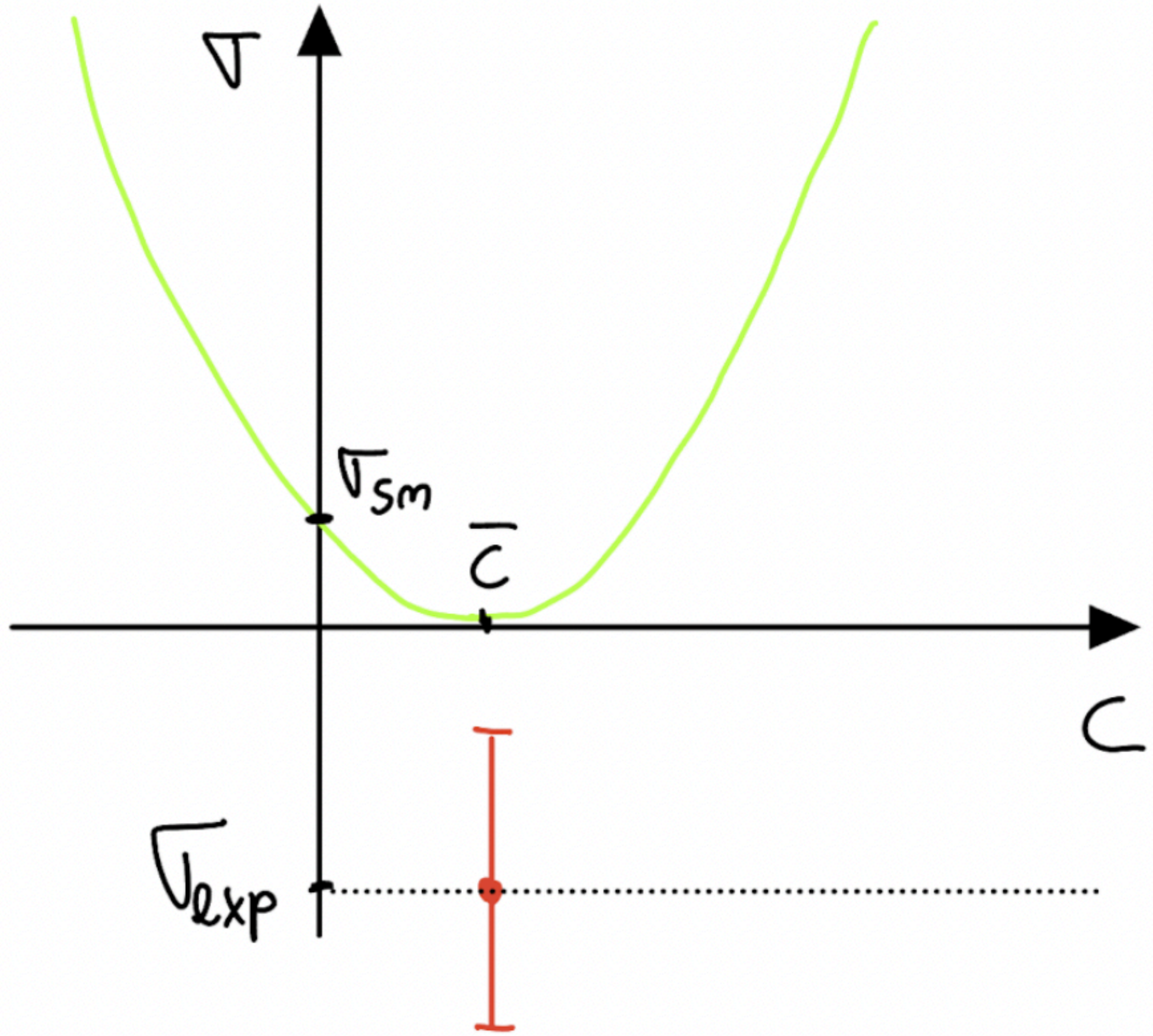
# Pitfalls of the Monte-Carlo replica method

- 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

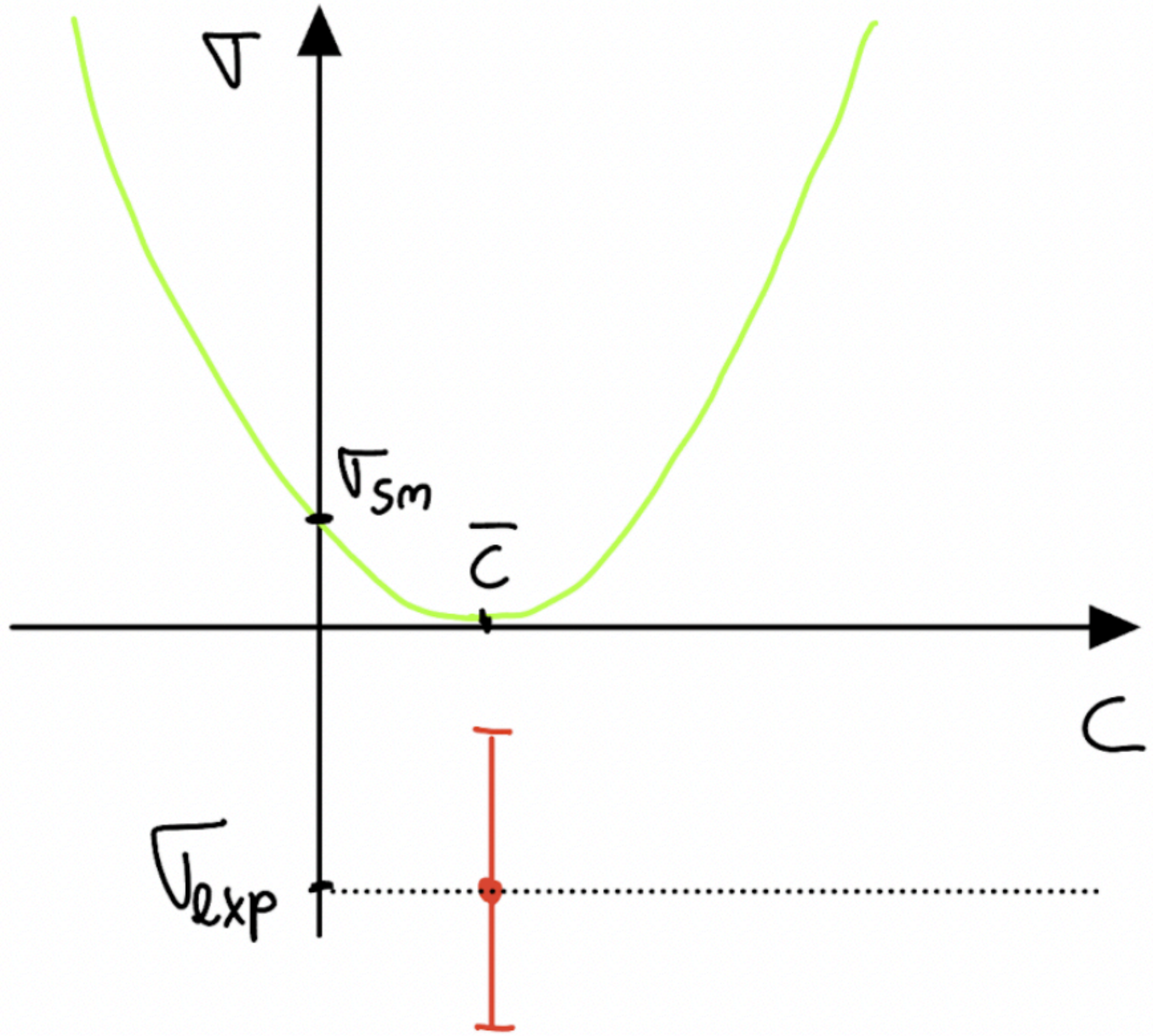
- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.




- 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

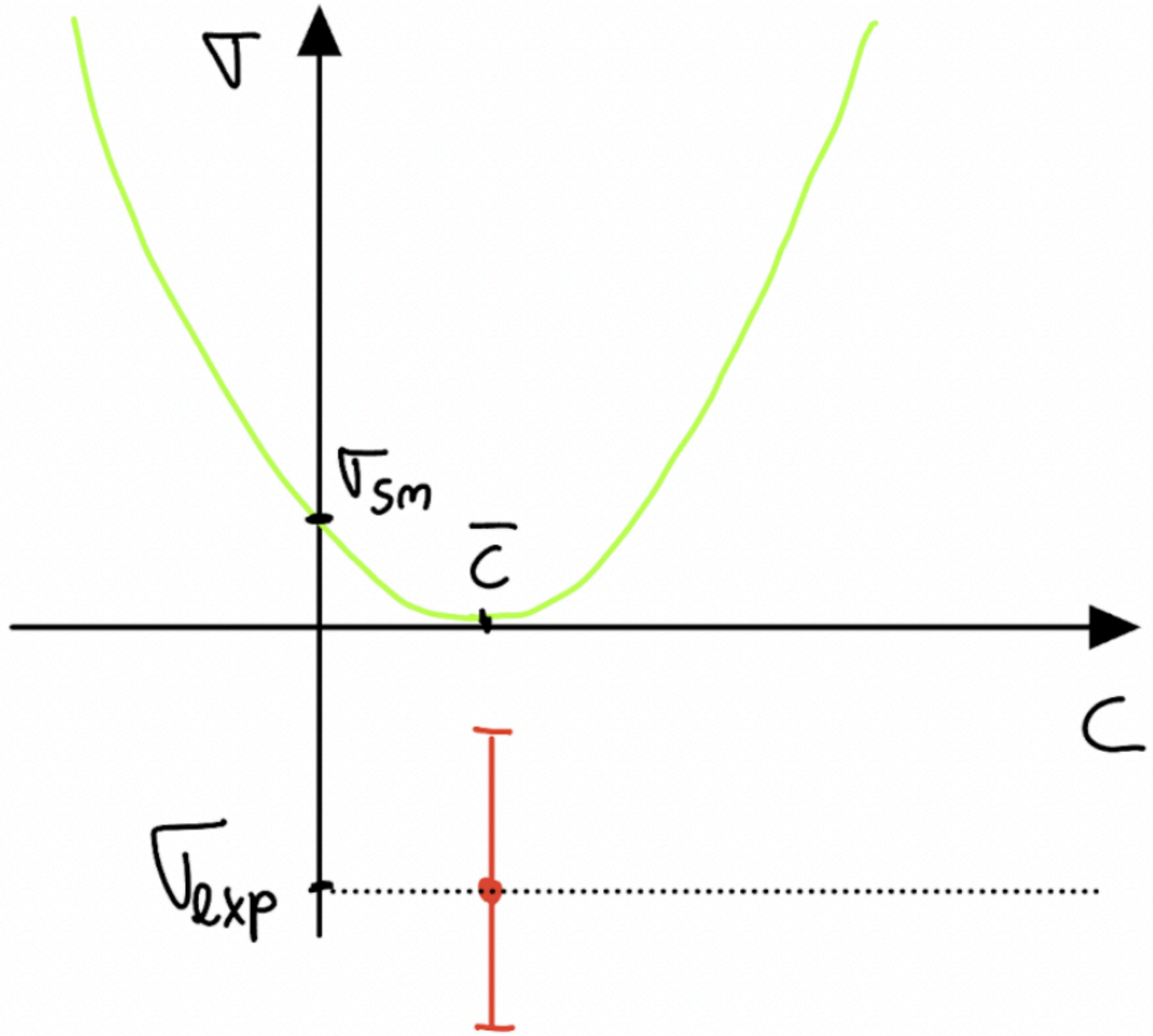
- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.




- 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

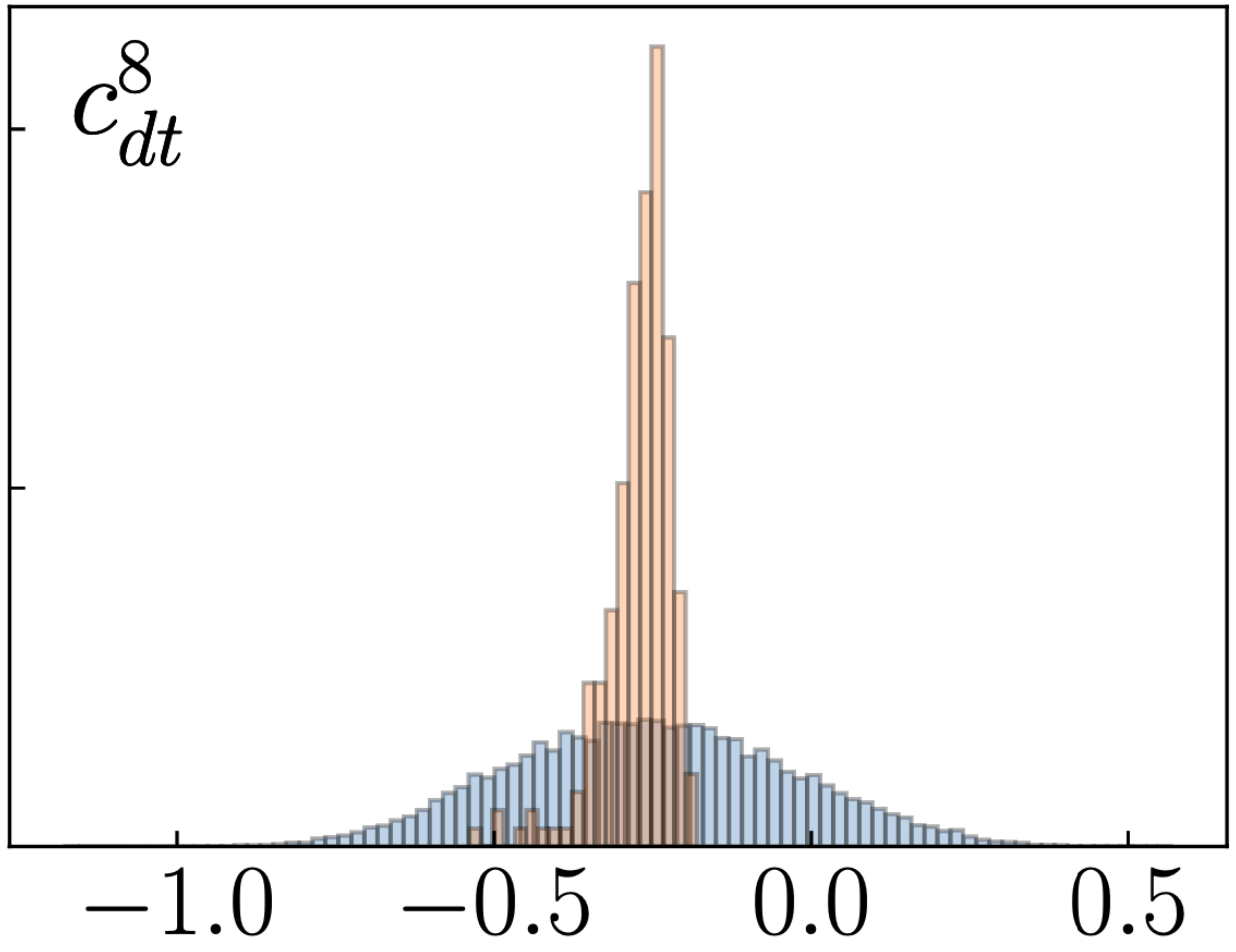
- The **minimum of the theory** can result in many pseudodata replicas falling **below the range of the theory**.




- 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?