The LHC Run II top quark legacy on global PDF and SMEFT analyses

for QCD@LHC2022, in Orsay, France

QCD@LHC2022

28 November 2022 to 2 December 2022

IJCLab Orsay, France

James Moore, University of Cambridge

PBSP

erc

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 Alvarado, Cameron Voisey (*former*)

Talk overview

1. The need for joint PDF-SMEFT determinations

2. The SIMUnet methodology for joint PDF-SMEFT fits

3. The impact of Run II top data on joint PDF-SMEFT fits

1. - The need for joint PDF-SMEFT determinations

PDF-SMEFT interplay

- Usually, the PDF world and the SMEFT world **do not talk...**

PDF-SMEFT interplay

- Usually, the PDF world and the SMEFT world **do not talk...**

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

PDF-SMEFT interplay

- Usually, the PDF world and the SMEFT world **do not talk...**

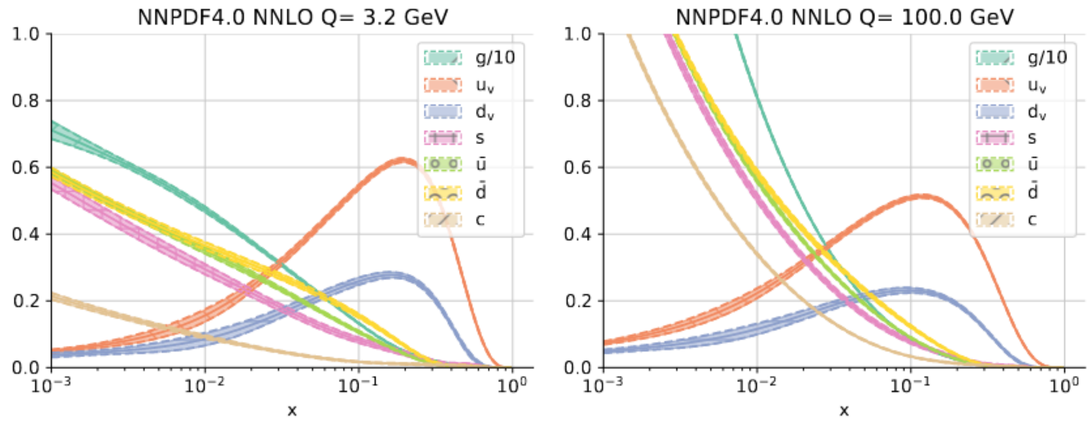
PDF fits

- 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.



PDF-SMEFT interplay

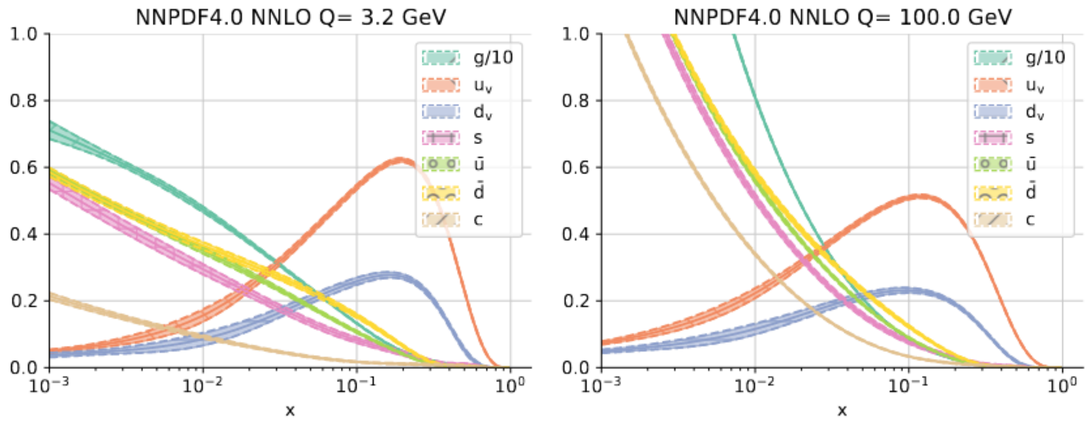
- Usually, the PDF world and the SMEFT world **do not talk...**

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

PDF-SMEFT interplay

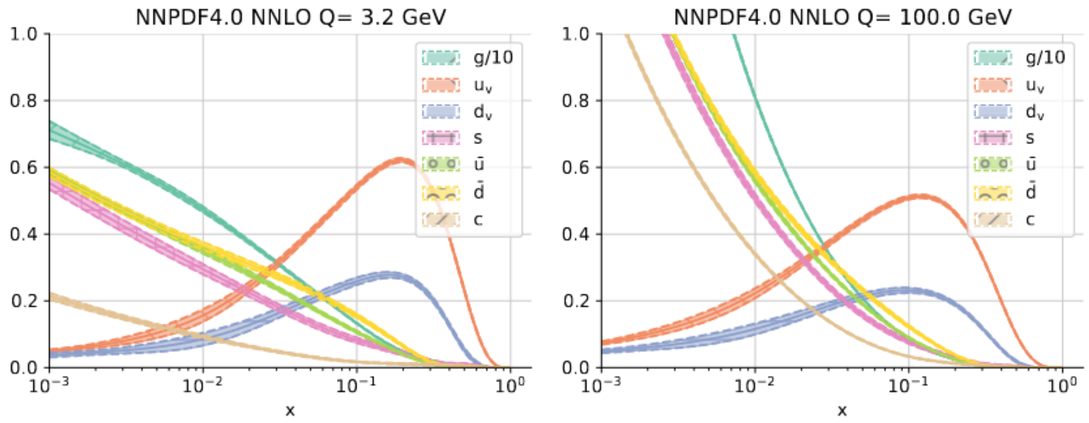
- Usually, the PDF world and the SMEFT world **do not talk...**

PDF 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.

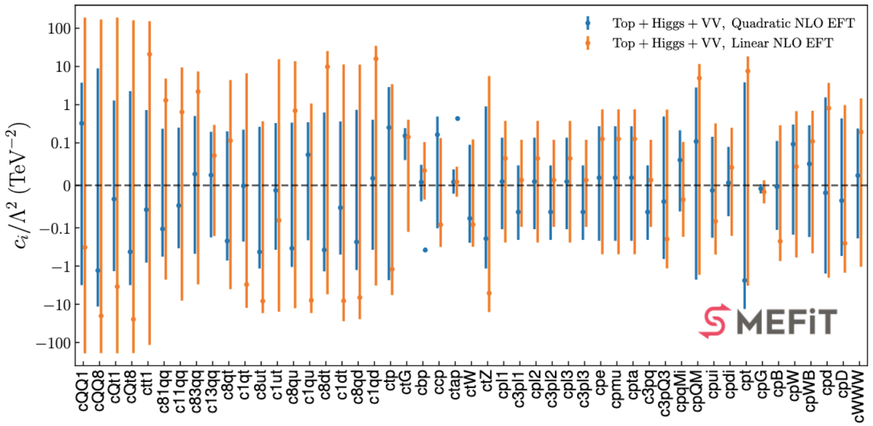


SMEFT fits

- 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.



PDF-SMEFT interplay

- **This could lead to inconsistencies.**

PDF fits

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

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

SMEFT fits

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

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

PDF-SMEFT interplay

- **This could lead to inconsistencies**.

PDF fits

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

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

SMEFT 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).

PDF-SMEFT interplay

- **This could lead to inconsistencies**.

PDF fits

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

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

SMEFT fits

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

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

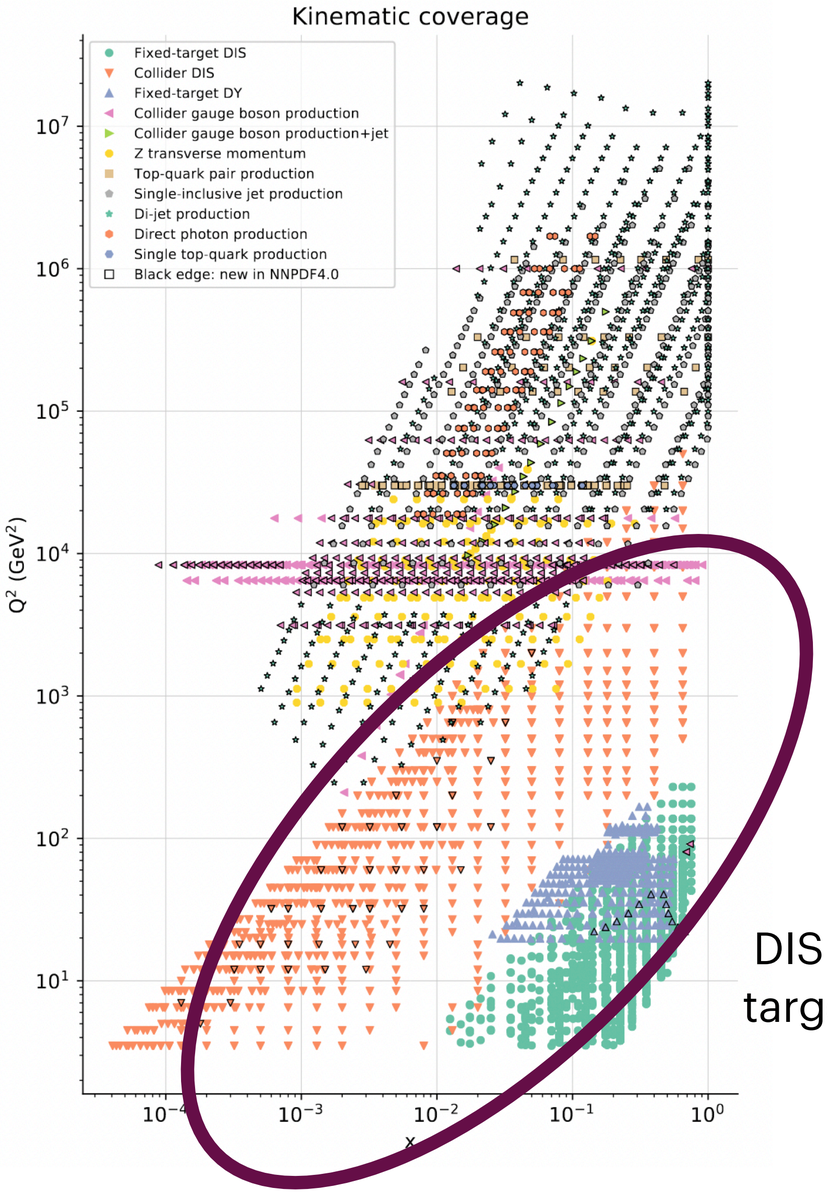
- In particular, if we fit PDFs **assuming all SMEFT couplings are zero**, but then **use those PDFs in a fit of SMEFT couplings**, our resulting bounds **could be misleading** (the same applies to SM parameters).
- We could even **miss New Physics**, or **see New Physics that isn't really there**!

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.

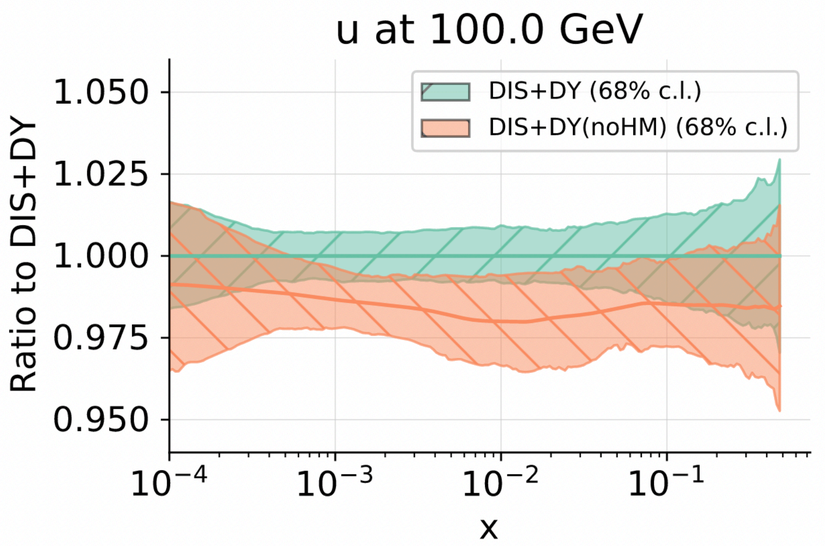


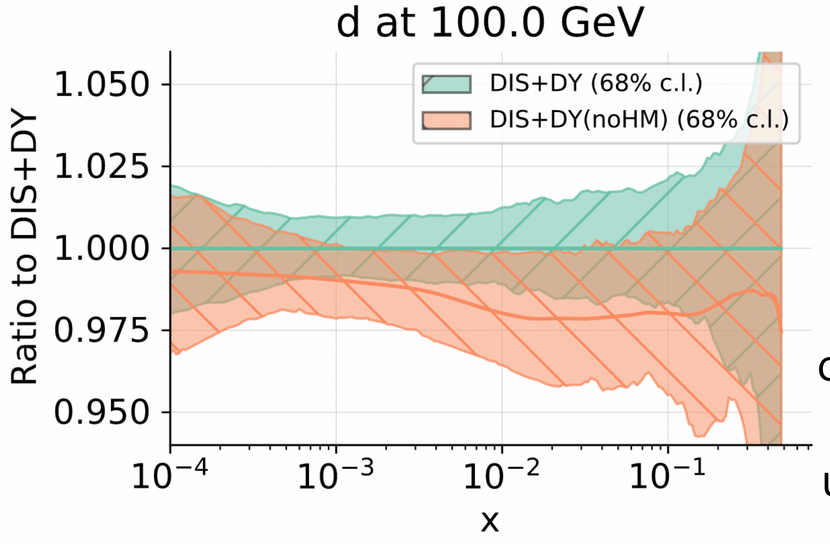
DIS and fixed target DY data

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

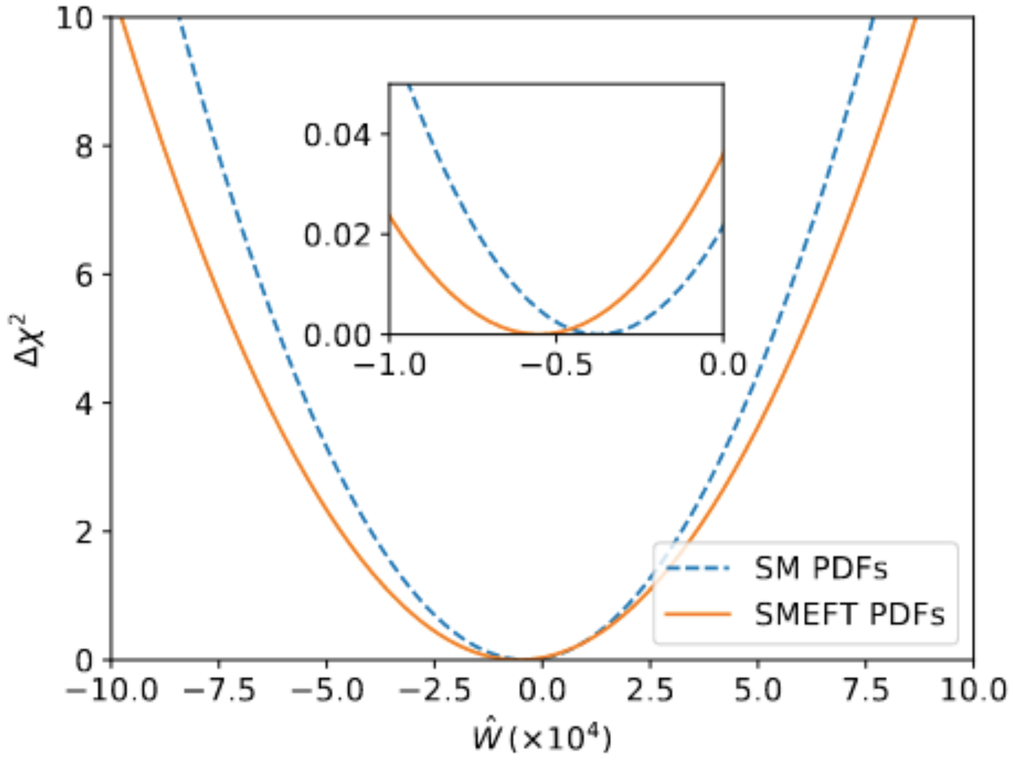
- *Question 2*: **Won’t the PDF-SMEFT interplay be negligible?**
  - It depends on the scenario!

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.

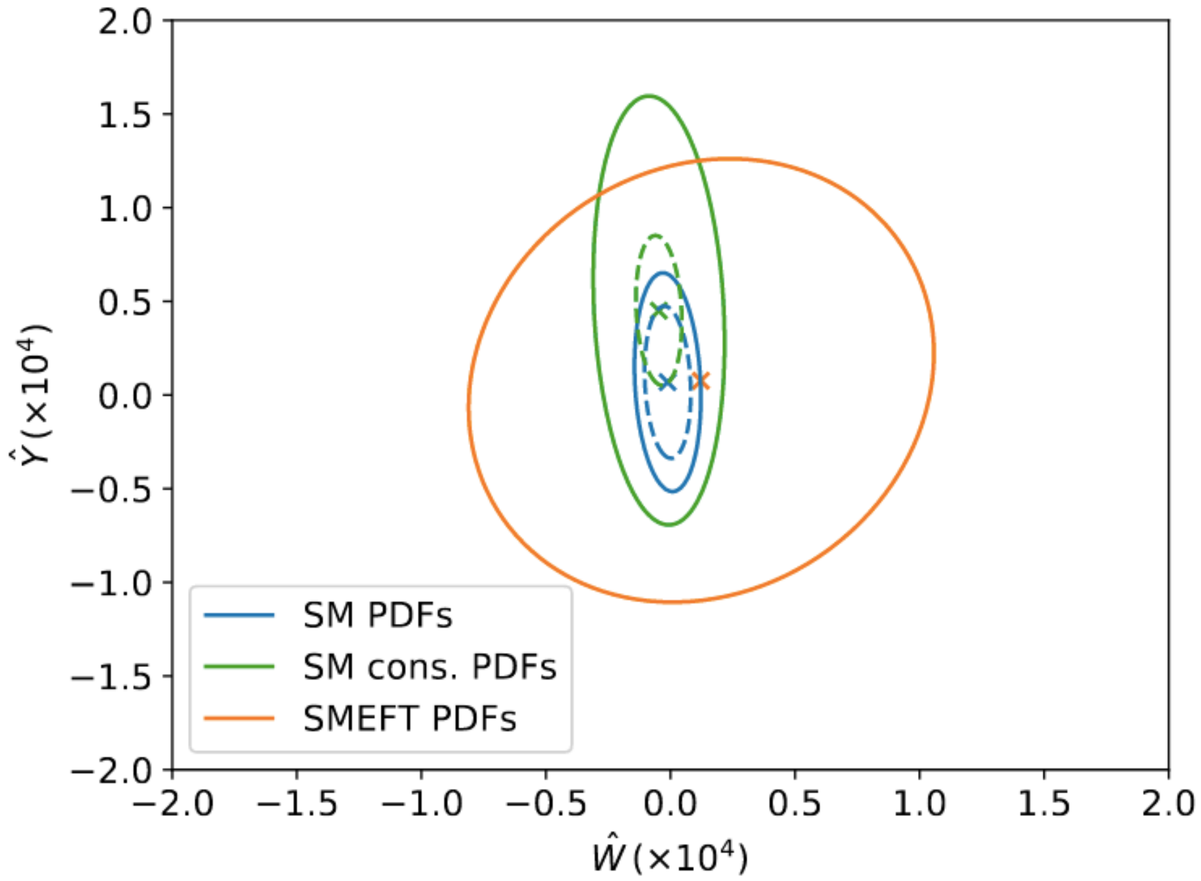
PDF-SMEFT interplay: natural questions

- *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.



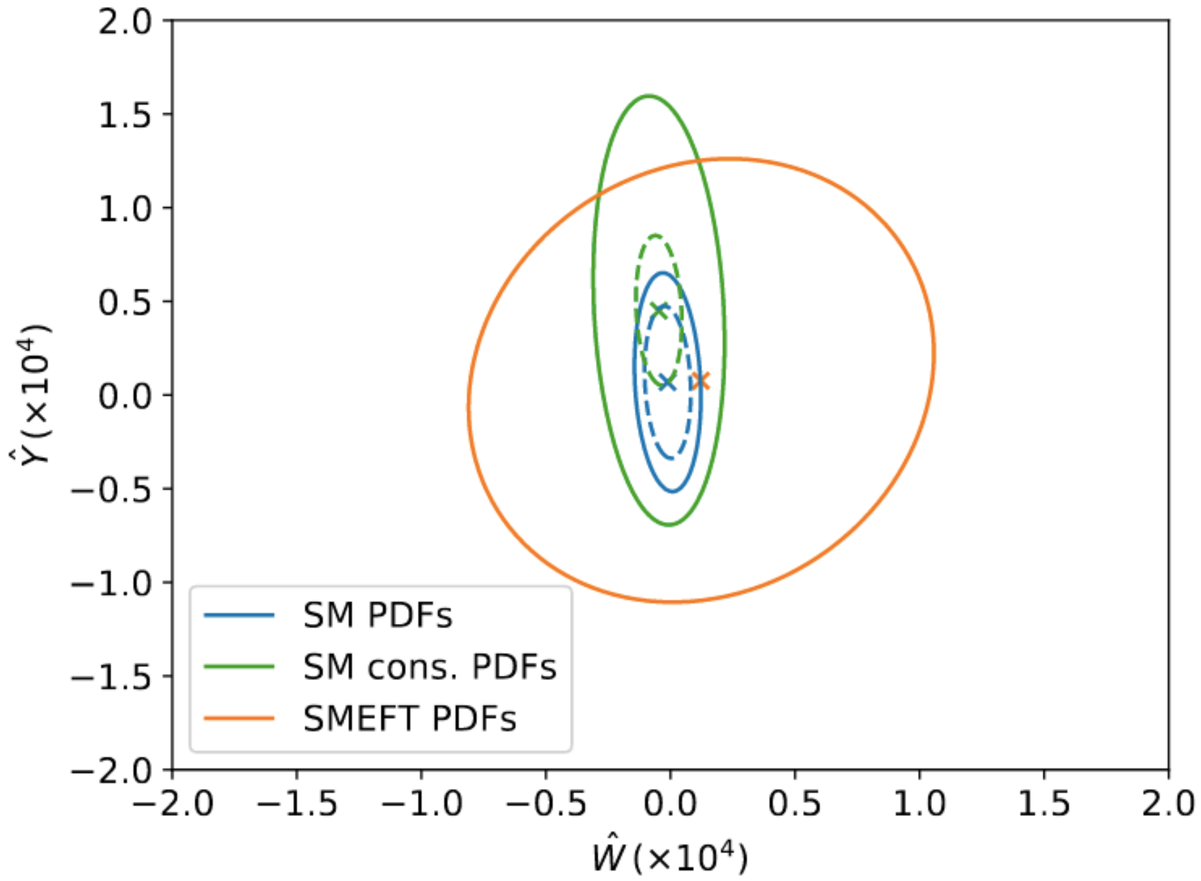
PDF-SMEFT interplay: natural questions

- *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**!



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.
- 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

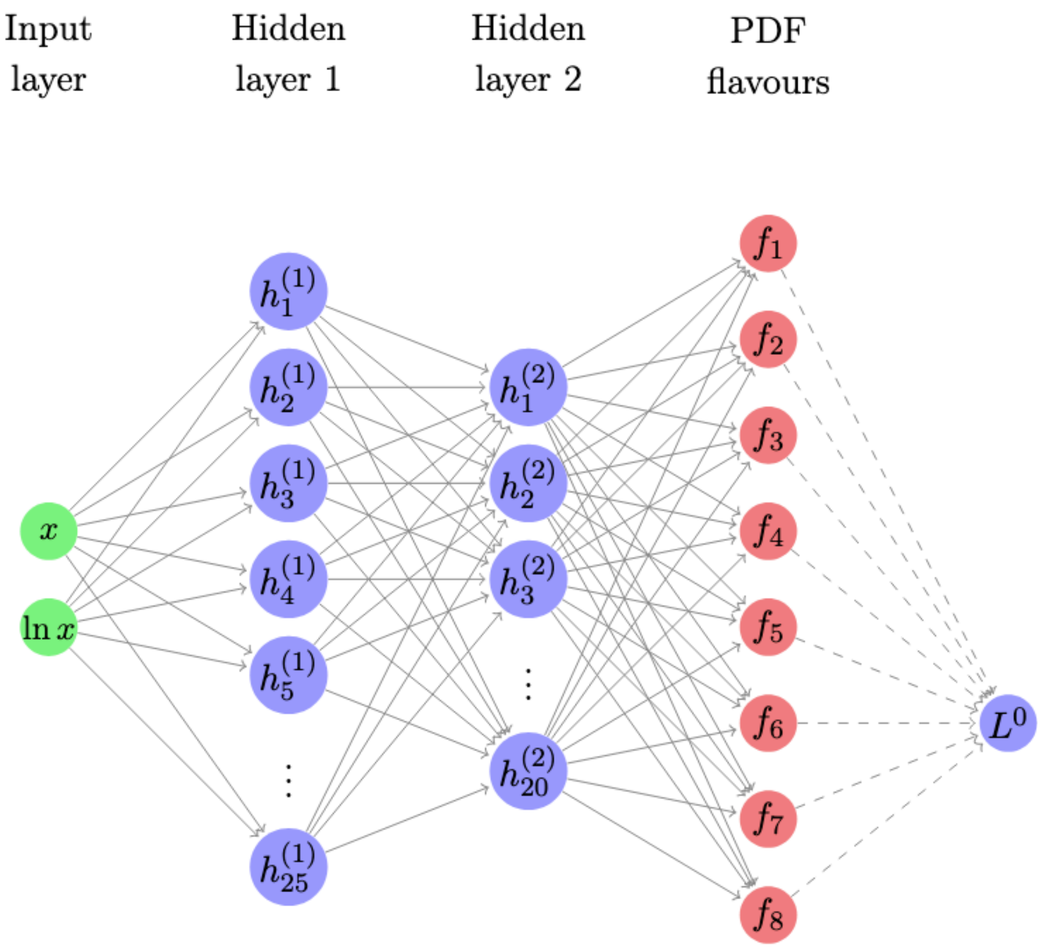
3. SIMUnet methodology

- 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

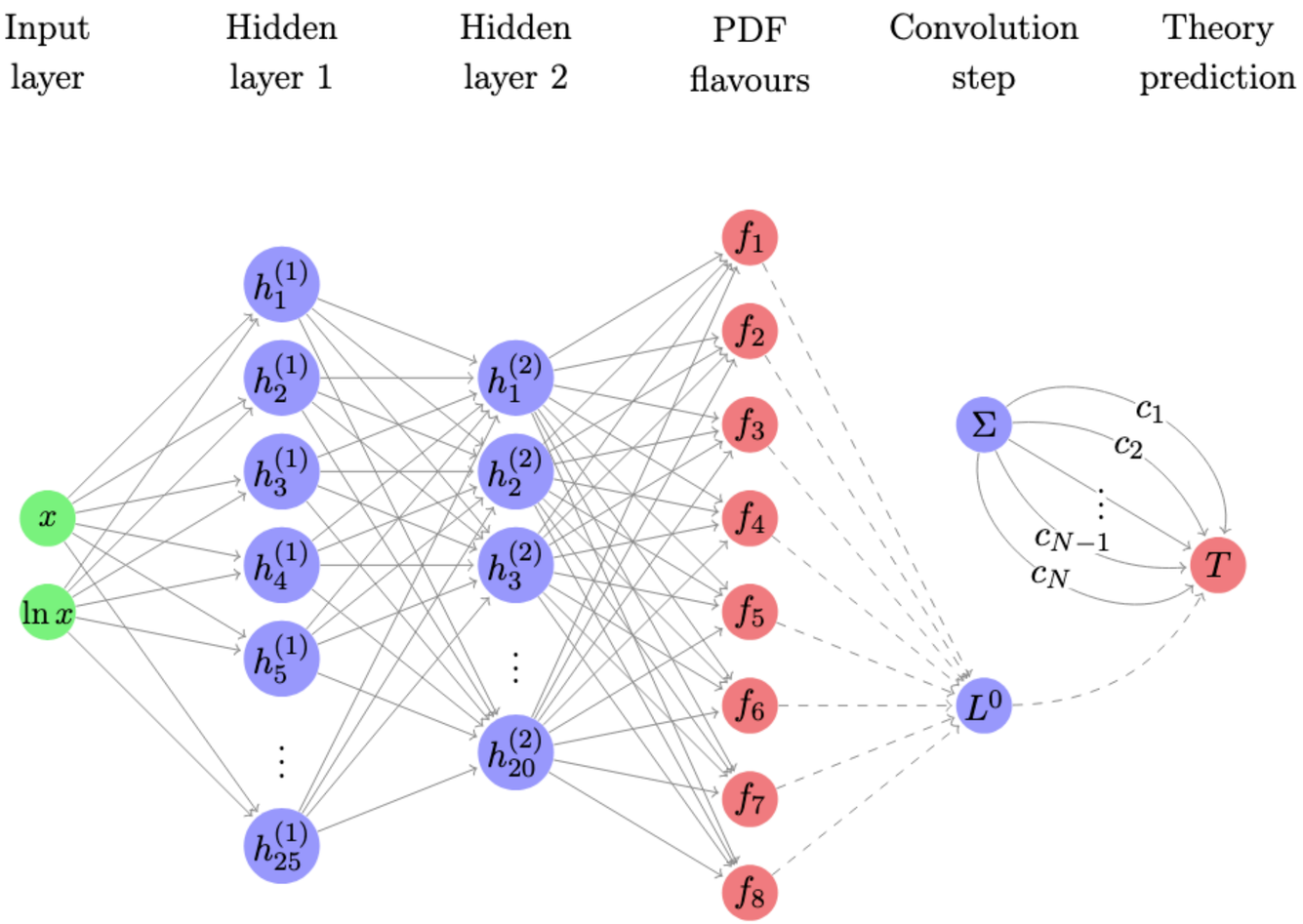
The SIMUnet methodology: details

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



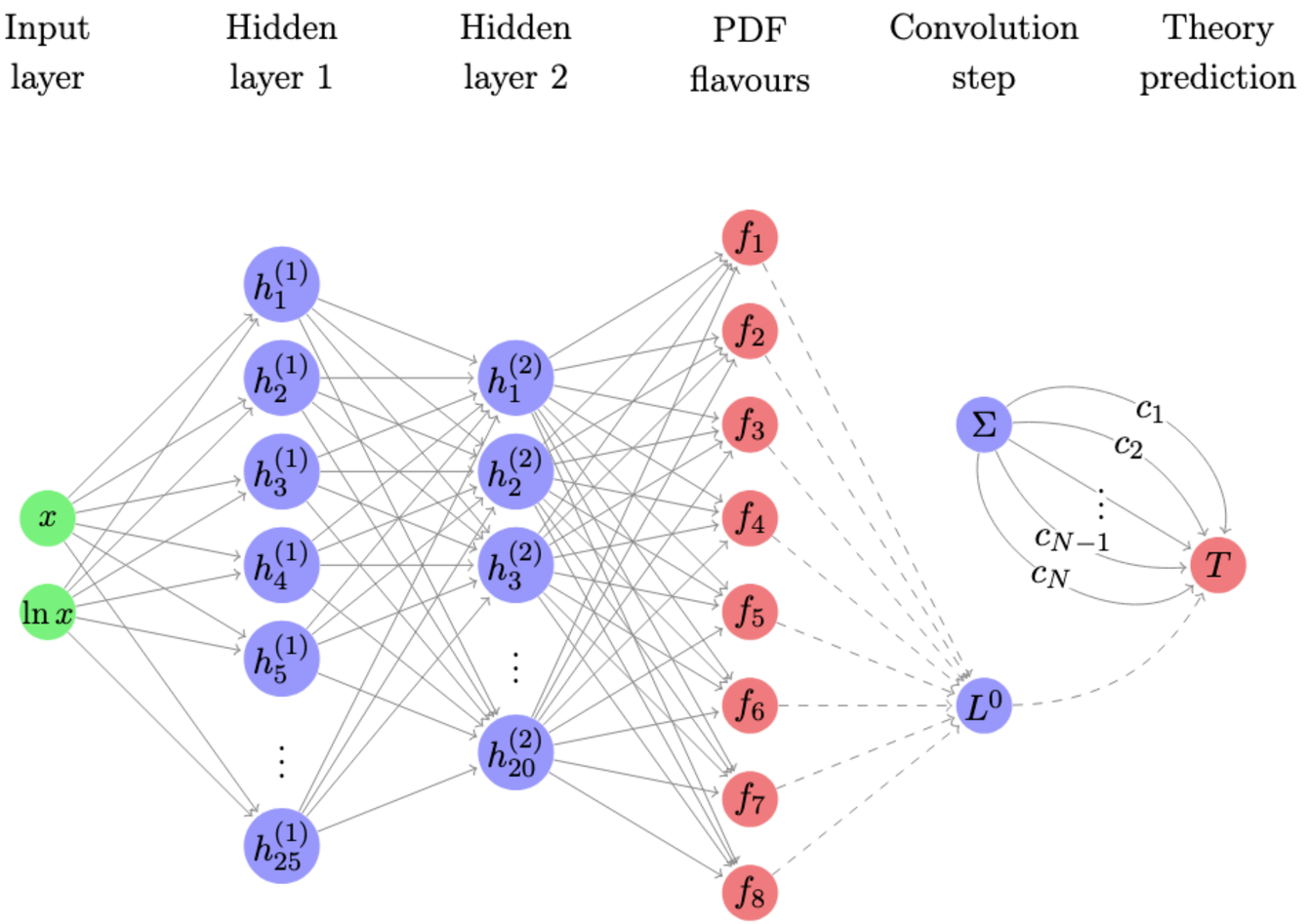
The SIMUnet methodology: details

- 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**.



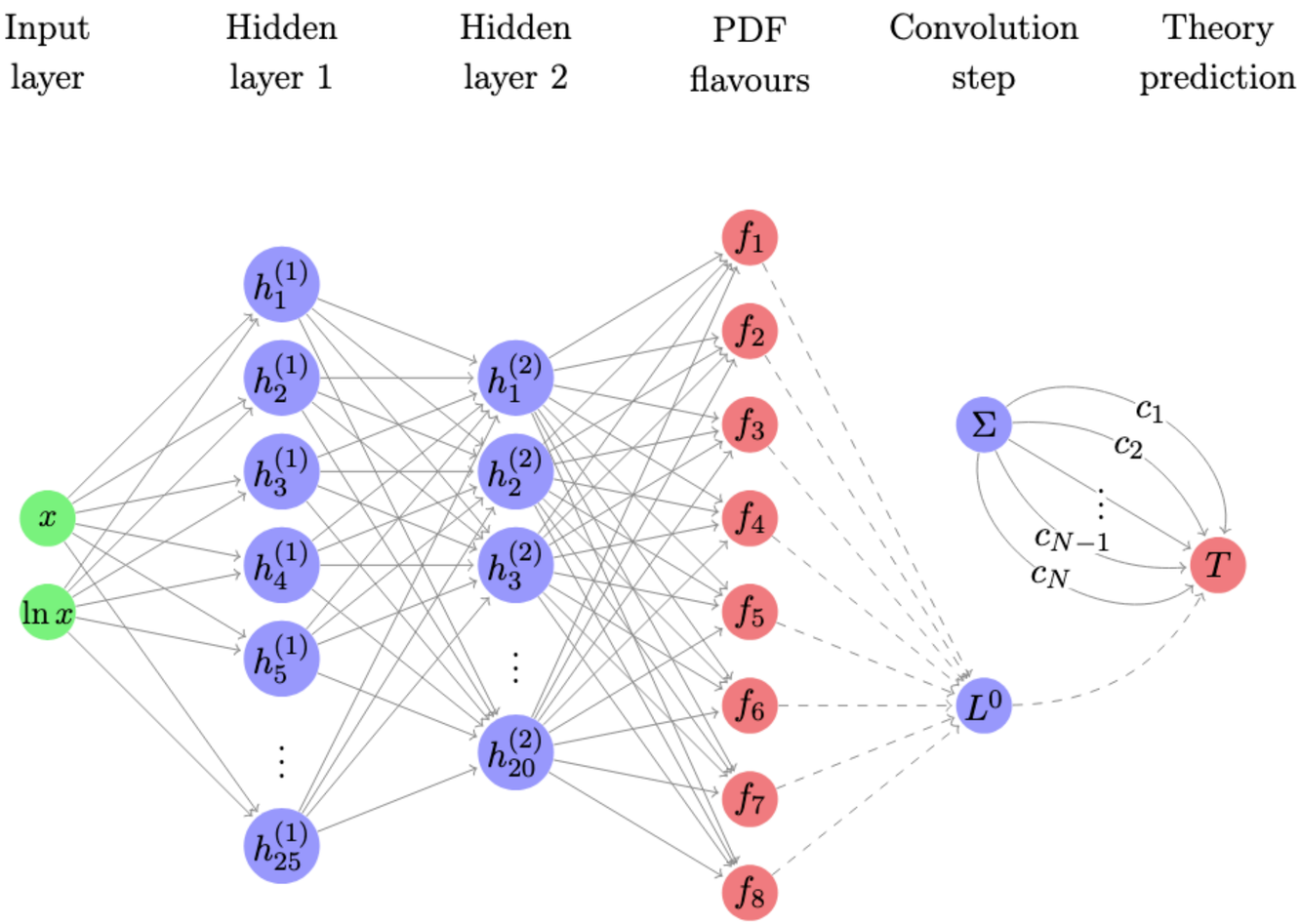
The SIMUnet methodology: details

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



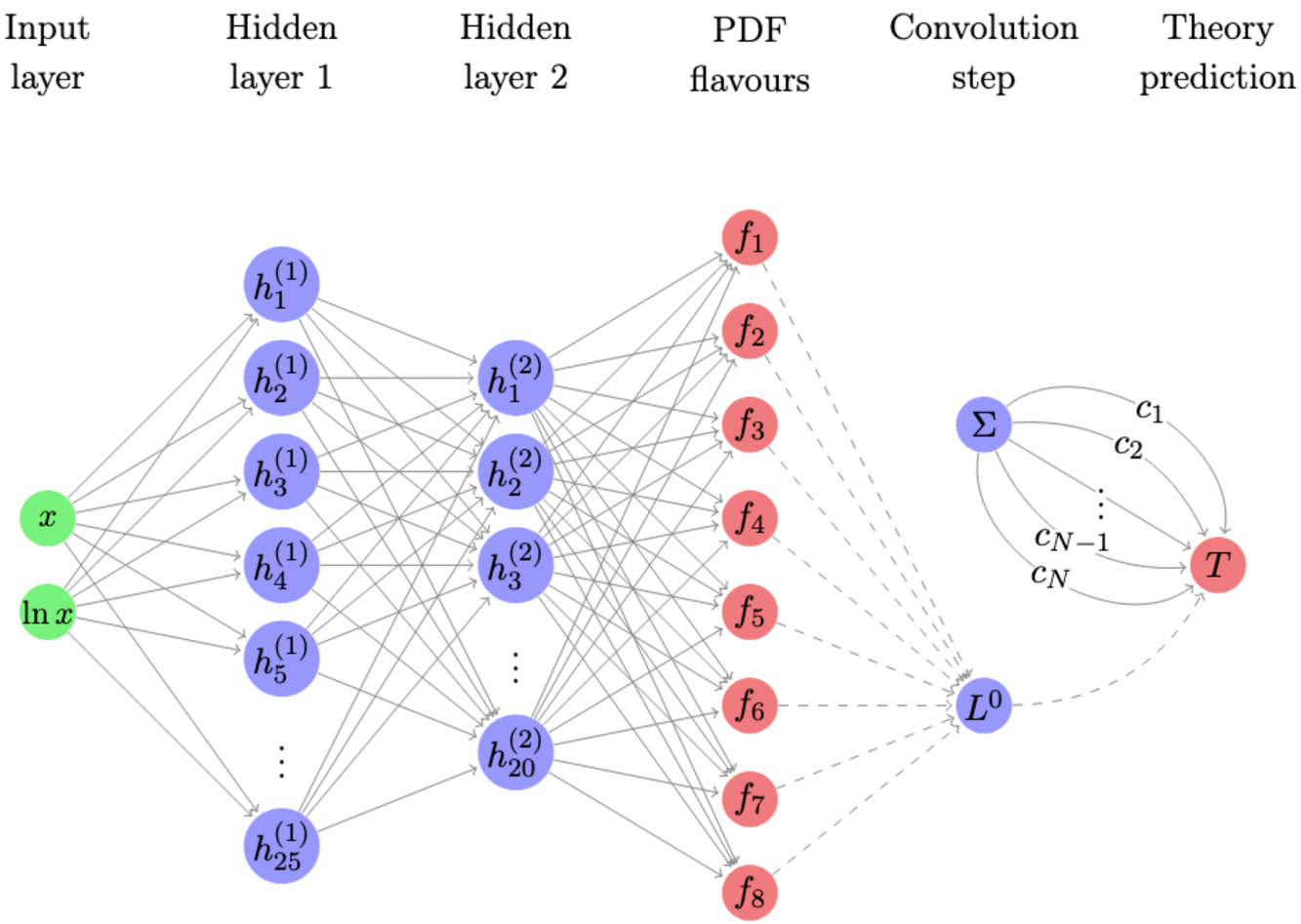
The SIMUnet methodology: details

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



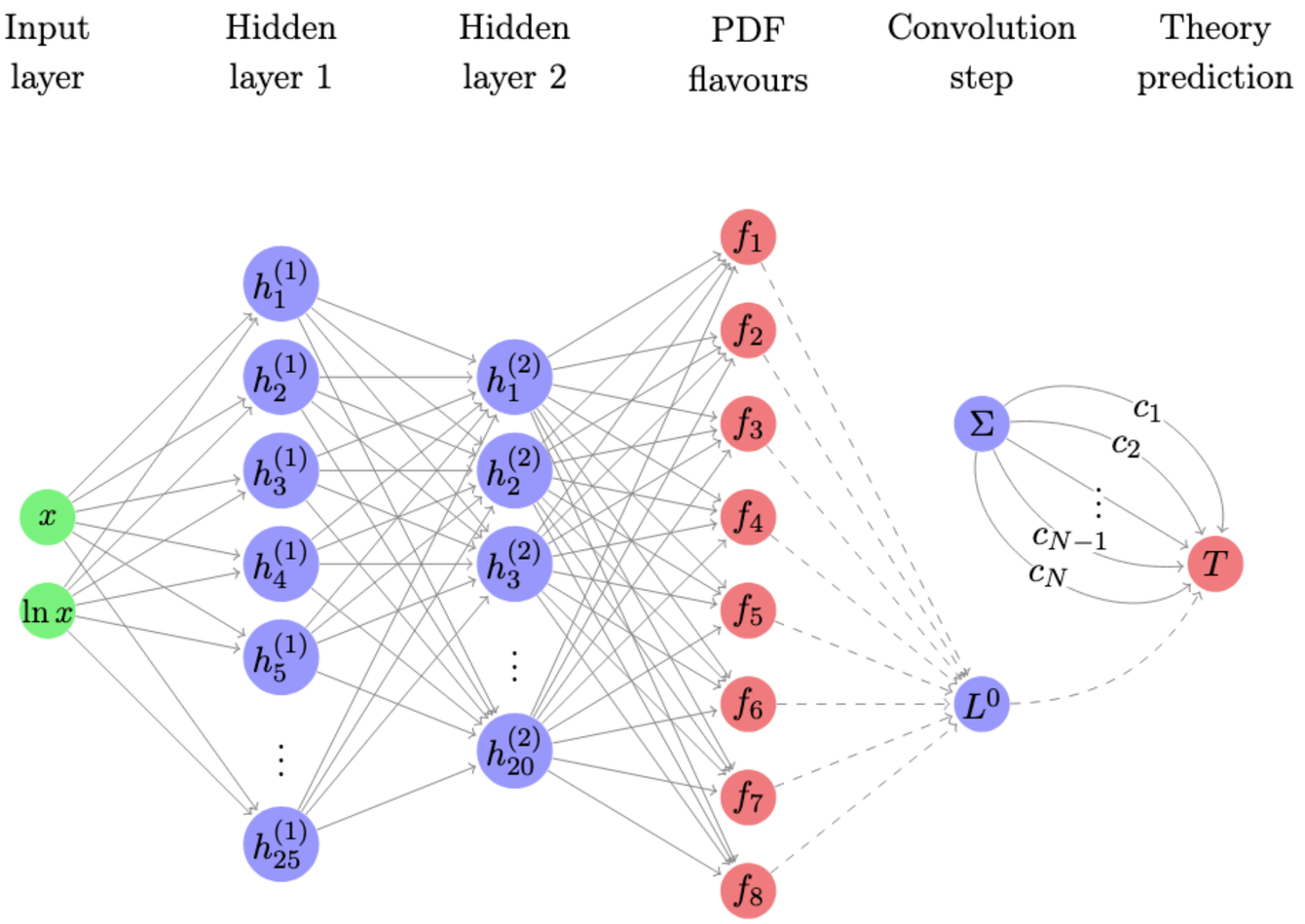
The SIMUnet methodology: details

- 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**.

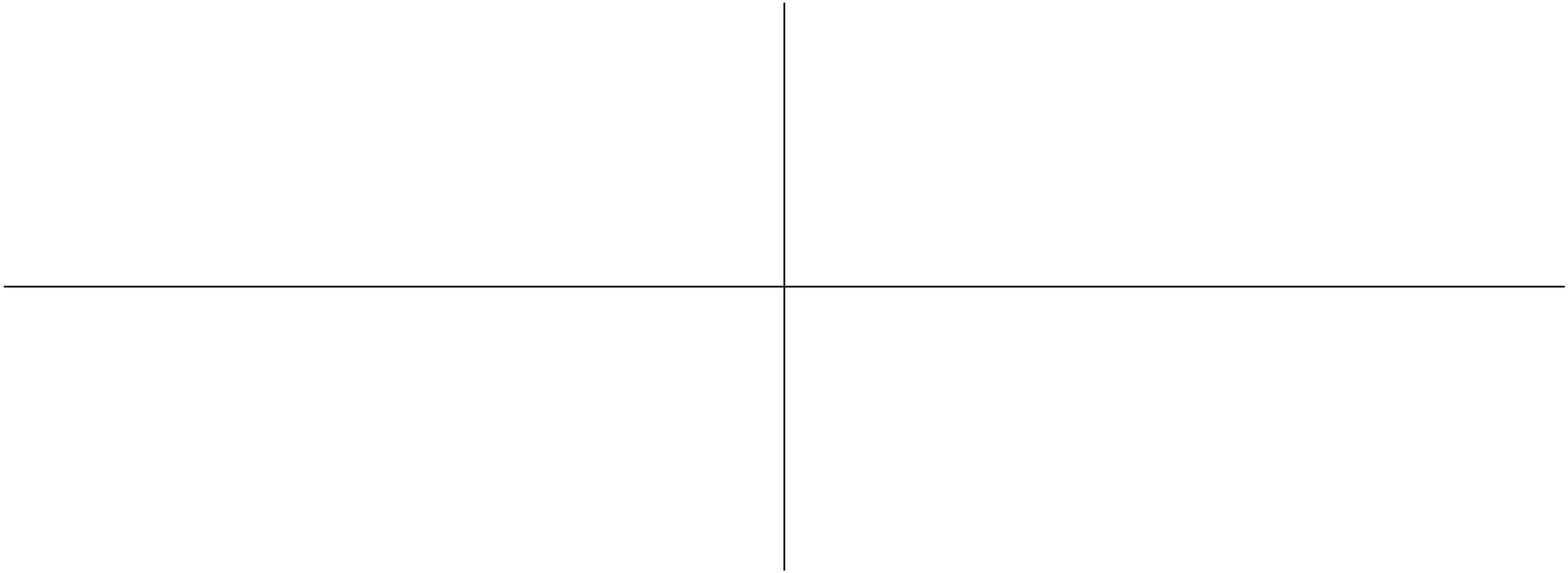


3. - The impact of Run II top data on joint PDF-SMEFT fits

Based on an upcoming publication
by the PBSP team + Juan Rojo

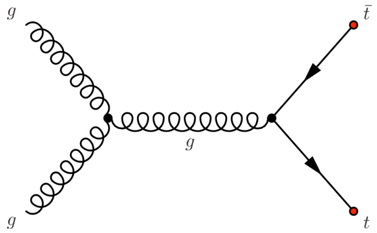
Run II top quark data

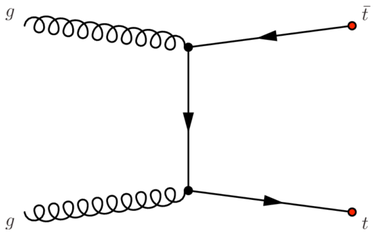
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:

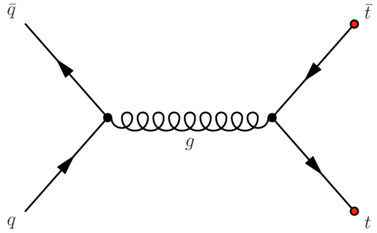


Run II top quark data

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



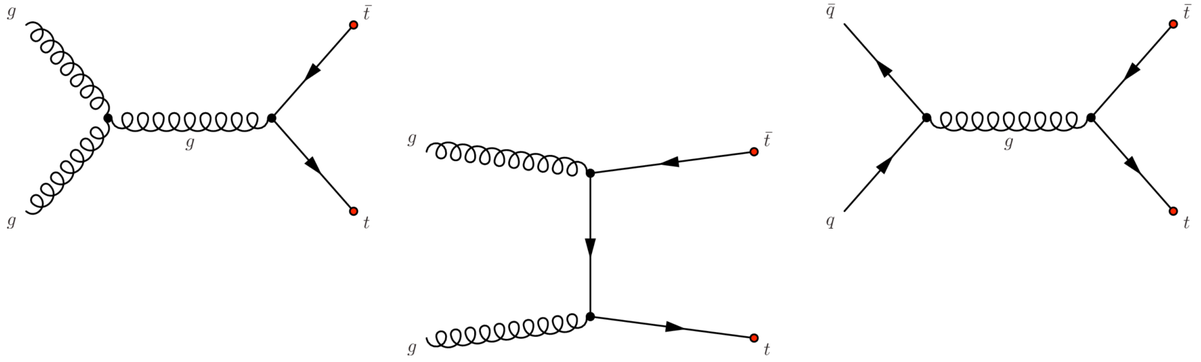




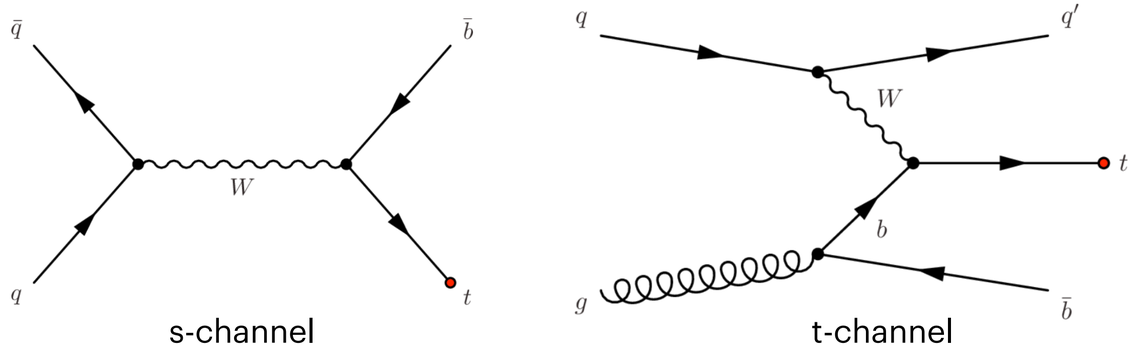
inclusive $t\bar{t}$

Run II top quark data

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



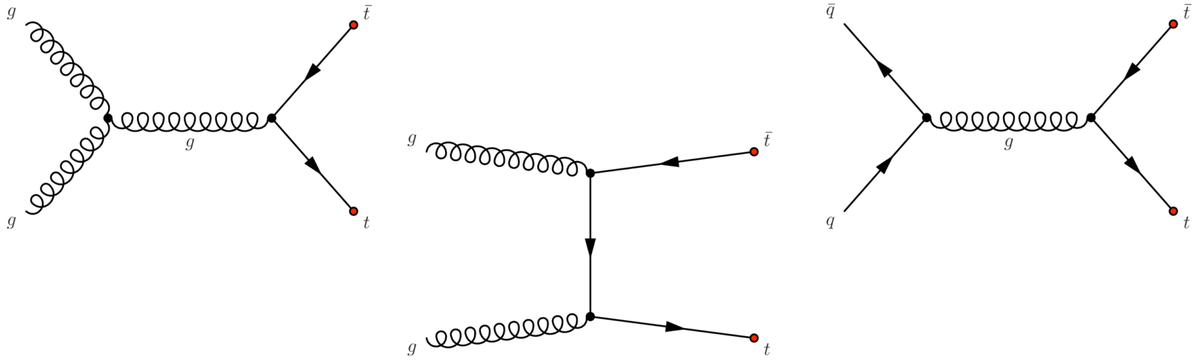
inclusive $t\bar{t}$



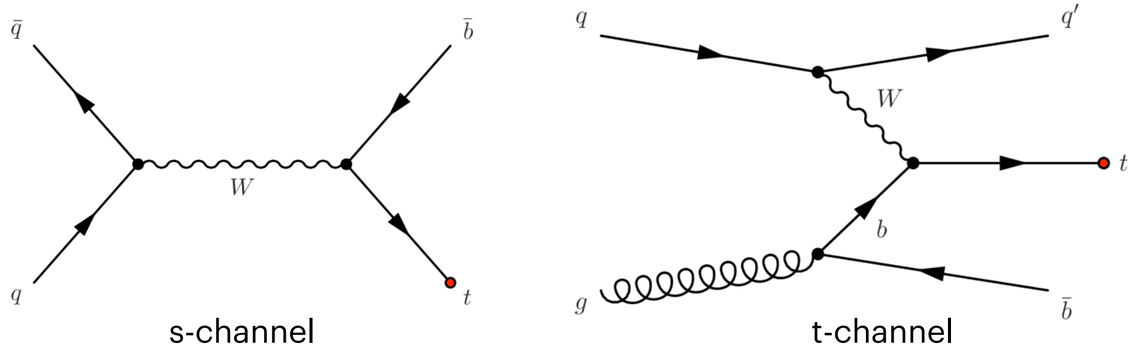
single top

Run II top quark data

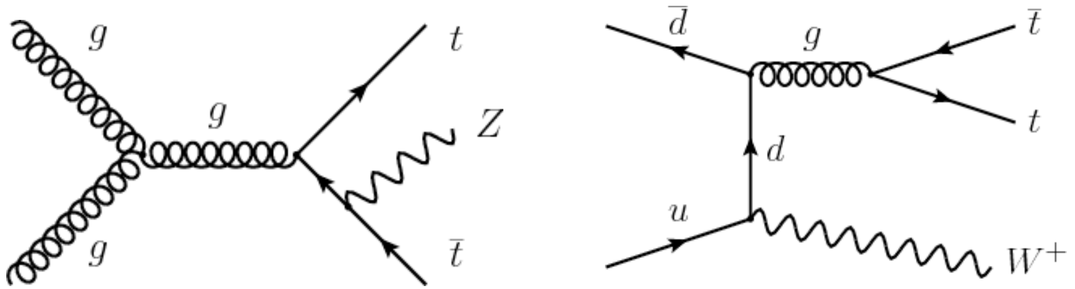
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



inclusive $t\bar{t}$



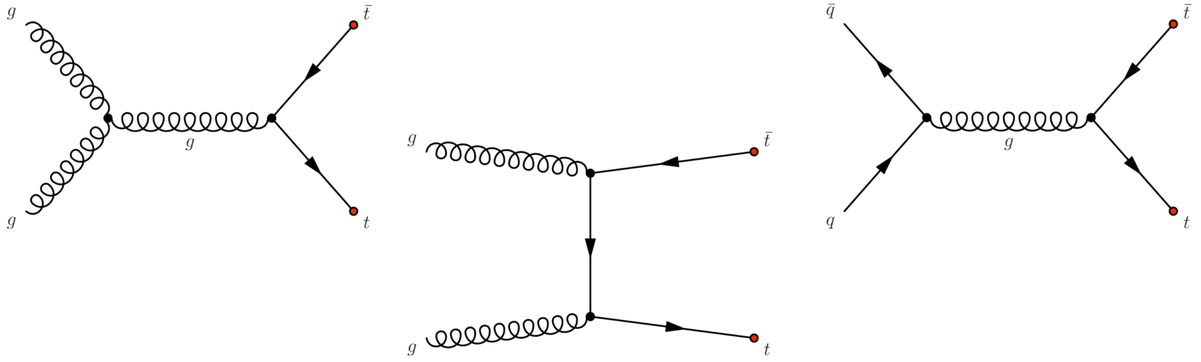
single top



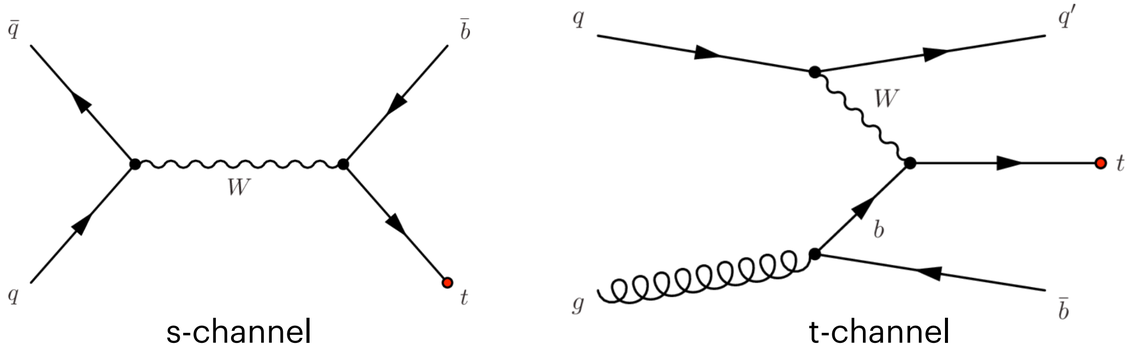
associated $t\bar{t}$

Run II top quark data

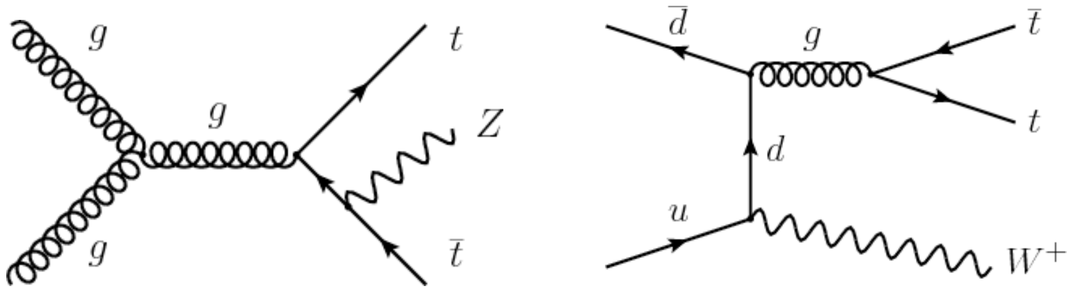
- **Huge amount of Run II top quark data** from ATLAS and CMS. Four basic processes:



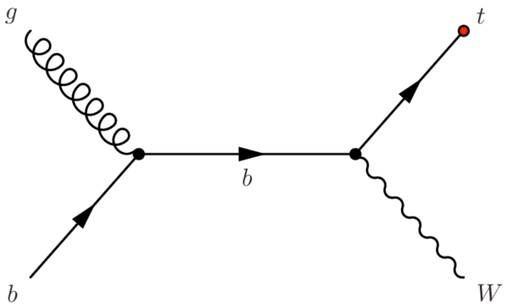
inclusive $t\bar{t}$



single top



associated $t\bar{t}$



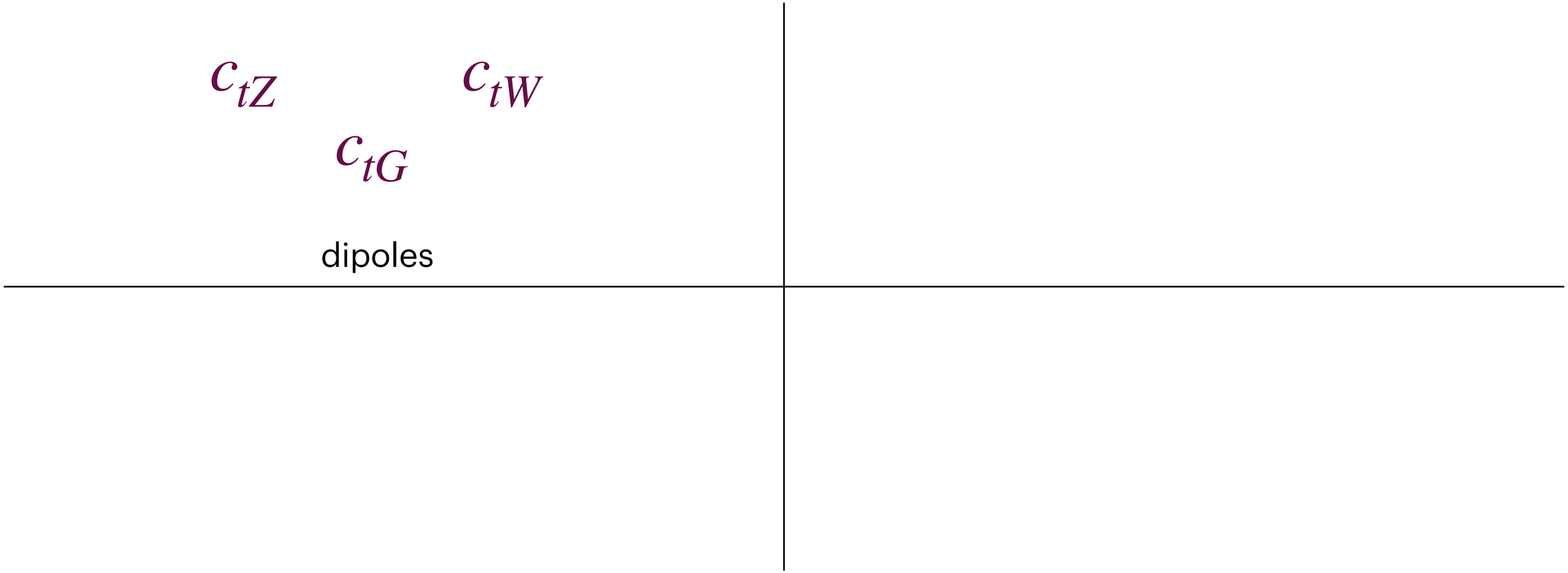
associated single top

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**:

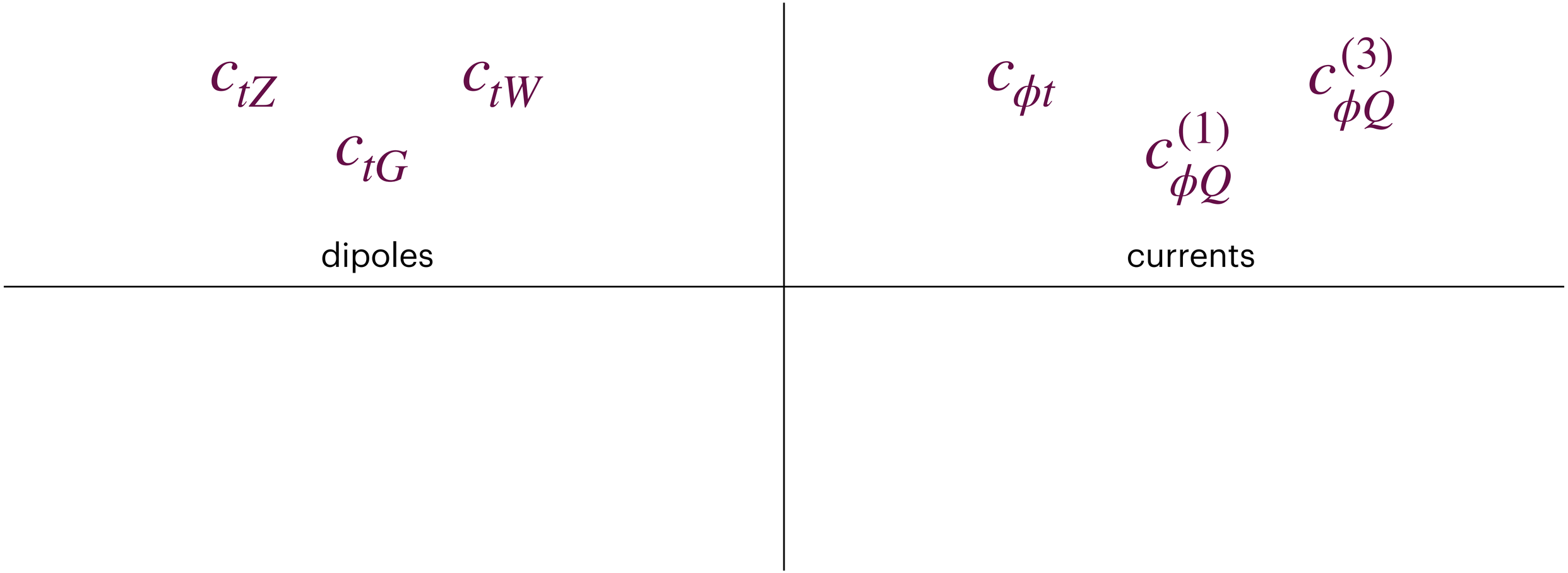
Run II top quark data

- 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**:



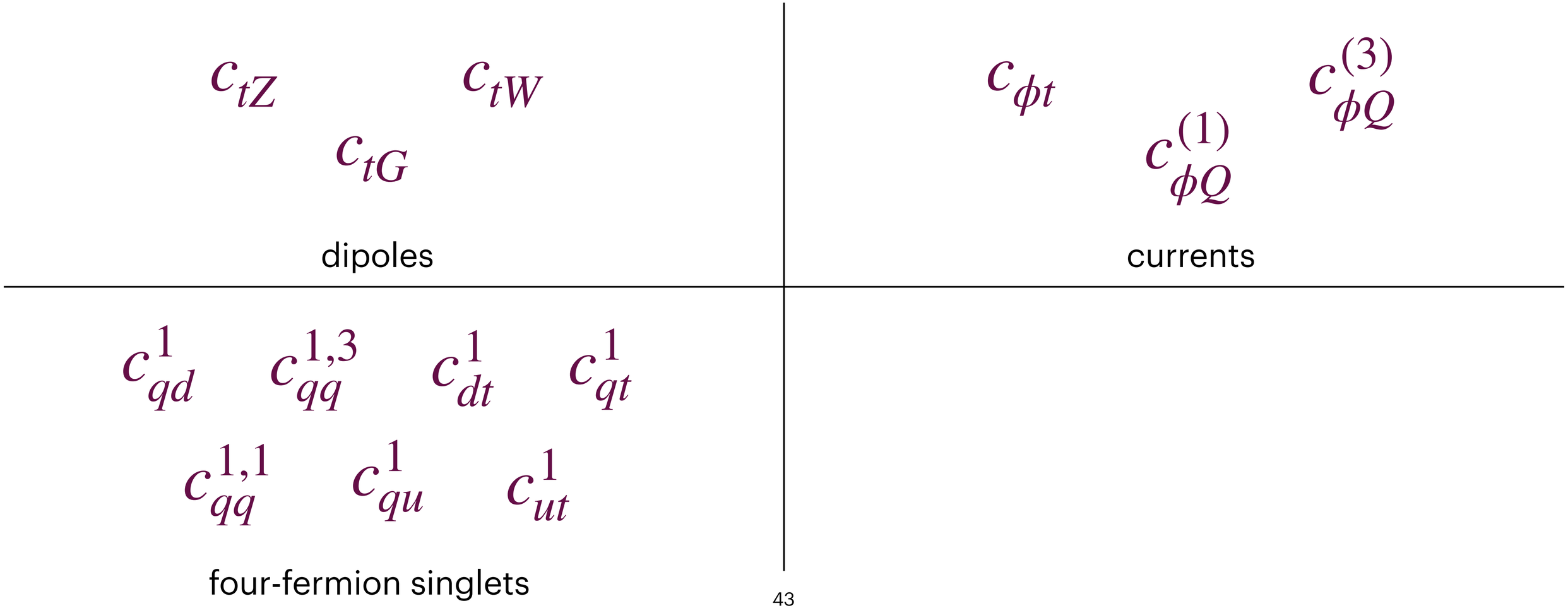
Run II top quark data

- 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**:



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?

Preliminary results: 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.

Preliminary results: 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 **185 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset of the measurements used in:

Preliminary results: 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 **185 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)

Preliminary results: 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 **185 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)

Preliminary results: fit settings

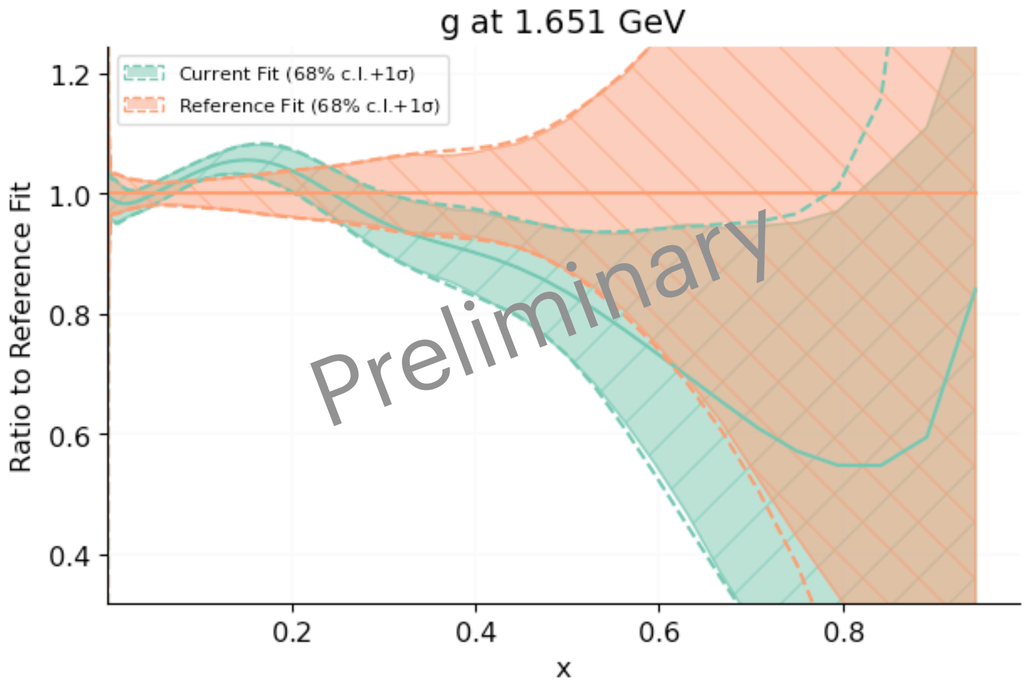
- Using the SIMUnet methodology, we have performed simultaneous determinations of PDFs and top-sector WCs using the **most comprehensive** and **up-to-date** LHC top dataset possible.
- We use **185 top data points** from ATLAS and CMS, for the four top processes described above, which comprise a superset 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)

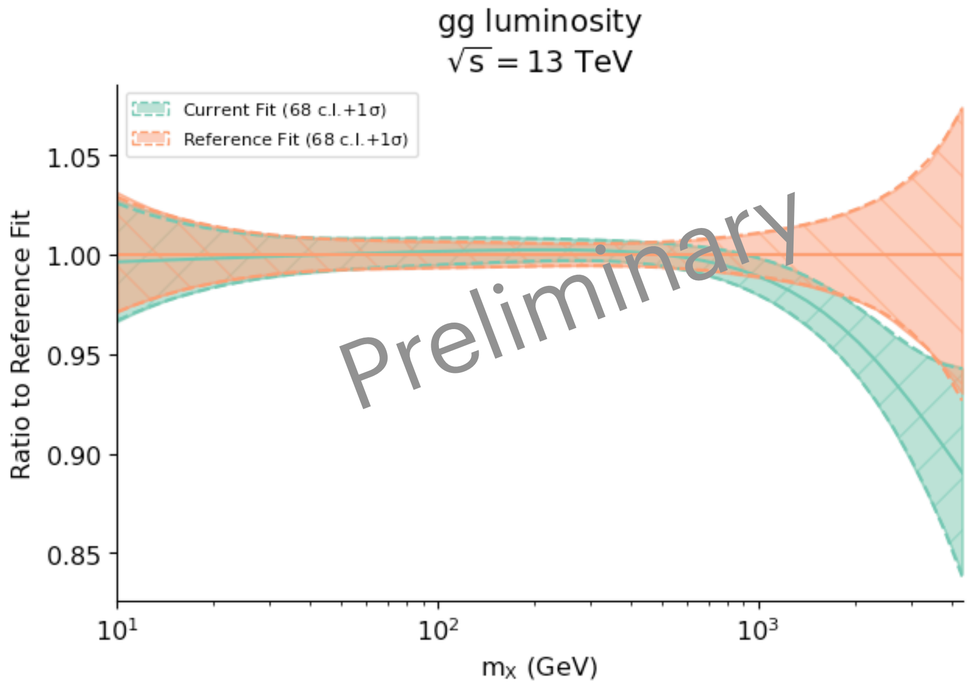
Preliminary results: fit settings

- 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 **185 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 and quadratic corrections in the SMEFT**.

Preliminary results: PDFs and luminosities

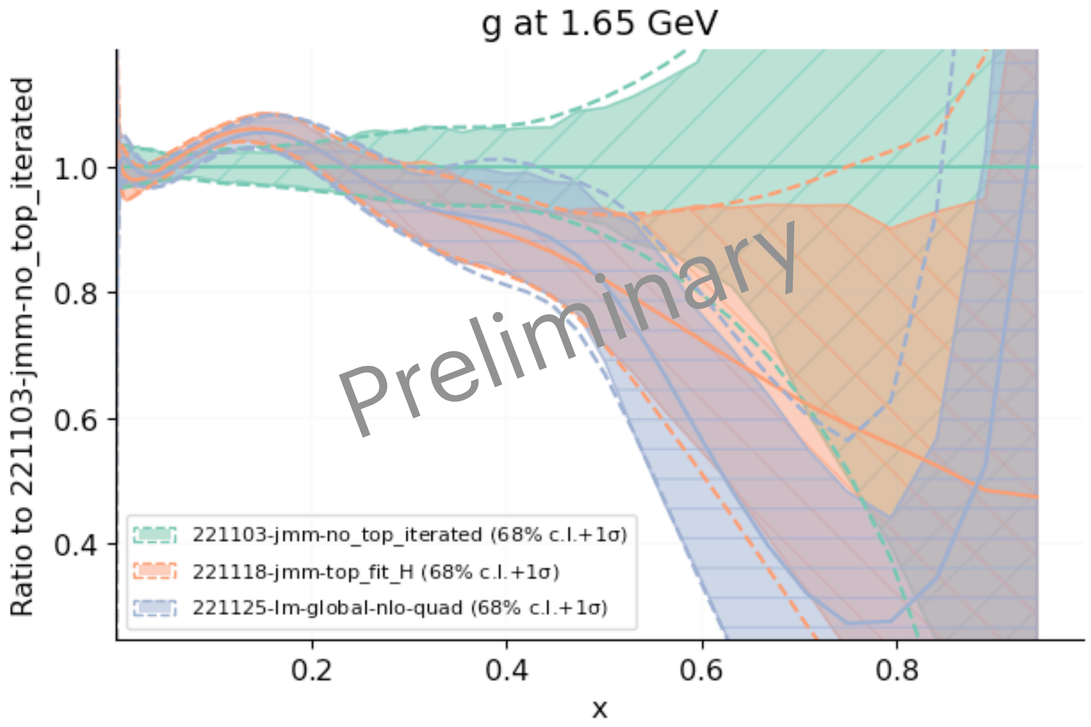
- First, we consider the impact of our dataset on PDFs **in the SM**. As expected from previous PDF studies, the contribution of the top data results in a **significant shift in the gluon PDF** in the **high-x region**.
- Orange: no-top PDF, green: SM PDF including top data:

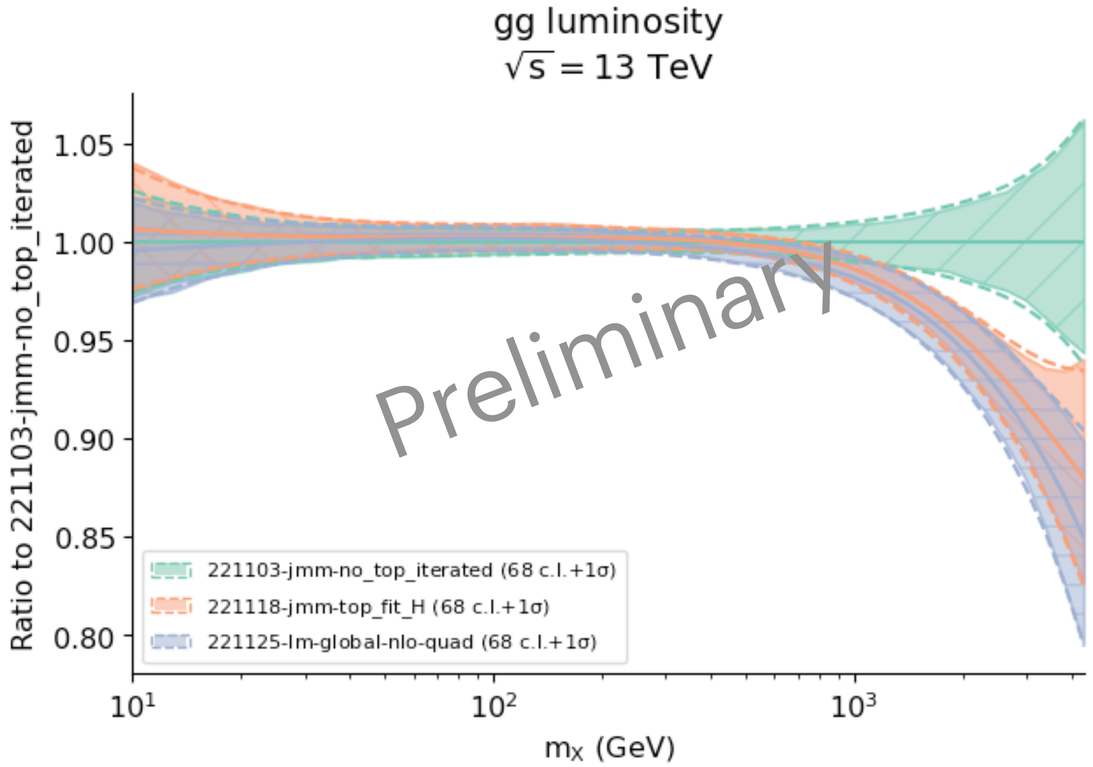




Preliminary results: PDFs and luminosities

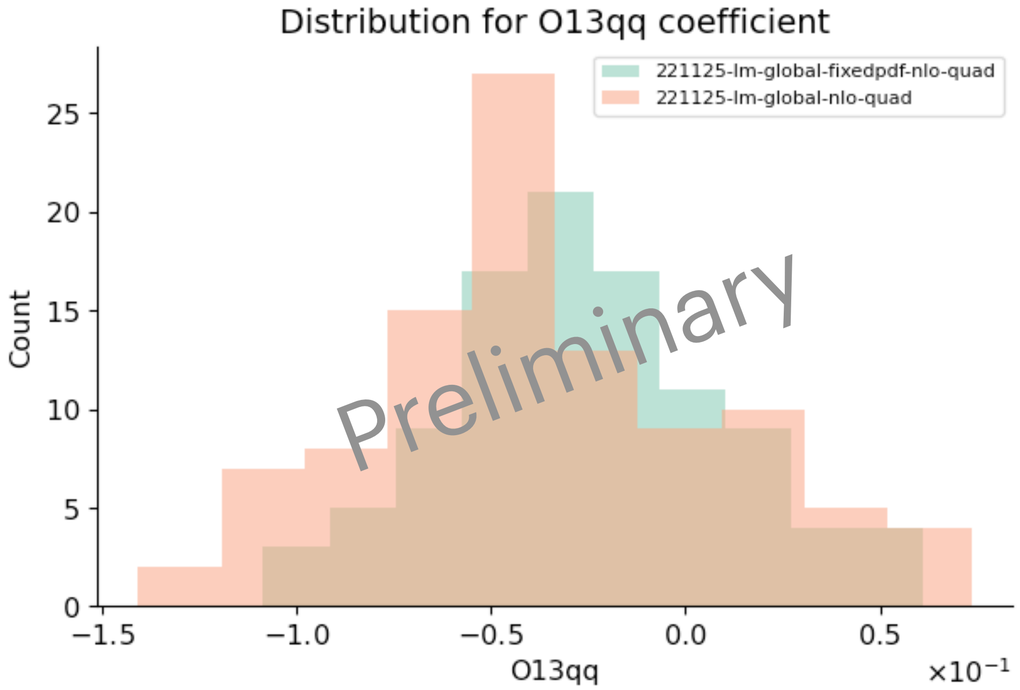
- When we perform a simultaneous fit of PDFs and WCs, the shift in the high-x gluon is **slightly enhanced**.
- Green: no-top PDF, orange: SM fit, blue: simultaneous fit:

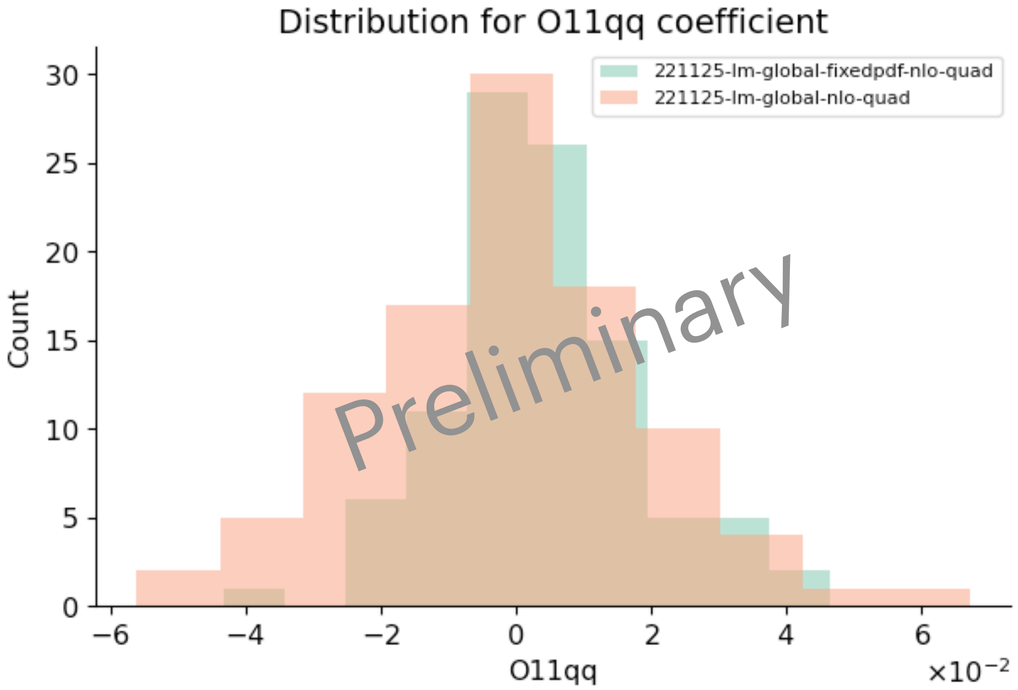




Preliminary results: Wilson coefficients

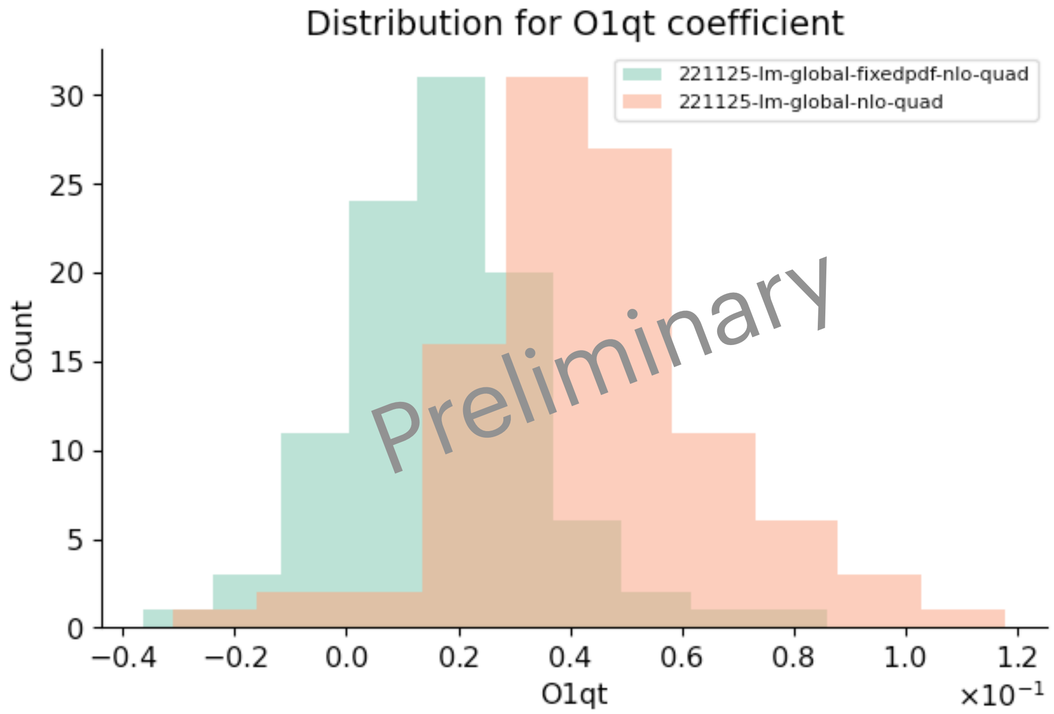
- Using the SIMUnet methodology, we find that bounds on the WCs are **slightly broader** (on average ~20%) compared to using fixed PDFs.
- Example distributions (orange: simultaneous, green: no-top fixed PDF):

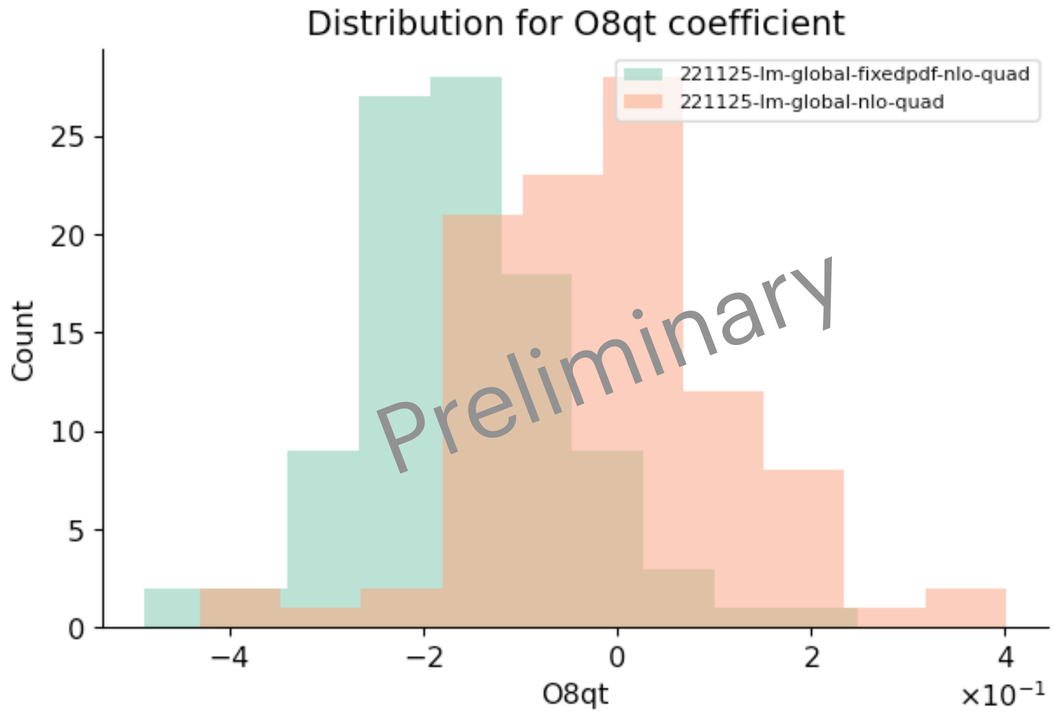




Preliminary results: Wilson coefficients

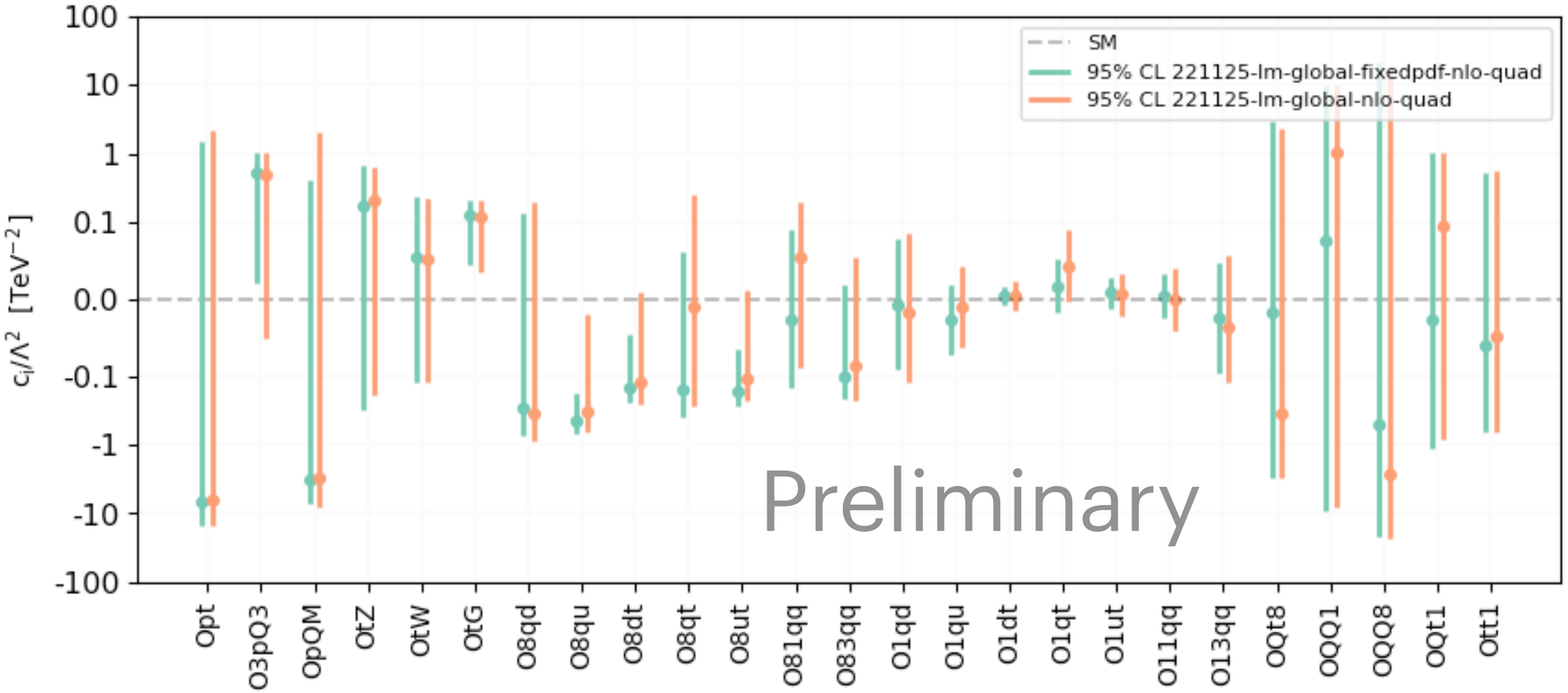
- For some operators, there is additionally a **small shift** in the central value of the distributions.
- Example distributions (orange: simultaneous, green: no-top fixed PDF):





Preliminary results: Wilson coefficients

- Summary of preliminary bounds (note logarithmic scale):
- Orange: simultaneous, green: no-top fixed PDF.



Preliminary results: fit quality

- Finally, using the SIMUnet methodology, we achieve a **significantly better** χ^2 statistics than using a fixed no-top baseline PDF in a SMEFT-only fit:

	SMEFT-only fit, using fixed no-top PDF set	**Simultaneous fit of PDFs and SMEFT**
χ^2 to top-sector data	1.287	1.034
Total χ^2	1.164	1.150

Preliminary results: fit quality

- Finally, using the SIMUnet methodology, we achieve a **significantly better** χ^2 statistics than using a fixed no-top baseline PDF in a SMEFT-only fit:

	SMEFT-only fit, using fixed no-top PDF set	**Simultaneous fit of PDFs and SMEFT**
χ^2 to top-sector data	1.287	1.034
Total χ^2	1.164	1.150

mainly driven by **better PDF fit**

Conclusions

Conclusions

- **Simultaneous determination of PDFs and SMEFT WCs** will be **very important in future analyses.**

Conclusions

- **Simultaneous determination of PDFs and SMEFT WCs** will be **very important in future analyses.**

- A **flexible, fast methodology, SIMUnet,** is available for simultaneous determination of PDFs and WCs.

Conclusions

- **Simultaneous determination of PDFs and SMEFT WCs** will be **very important in future analyses.**

- A **flexible, fast methodology, SIMUnet,** is available for simultaneous determination of PDFs and WCs.

- The PBSP team, and Juan Rojo, have applied this new methodology to PDF-SMEFT fits in the **top-sector**.

Thanks for listening!
Questions?