

SIMUnet: an open-source tool for the simultaneous fit of PDFs and SMEFT coefficients

Mark N. Costantini

IRN TERASCALE 2023 @ Marseille



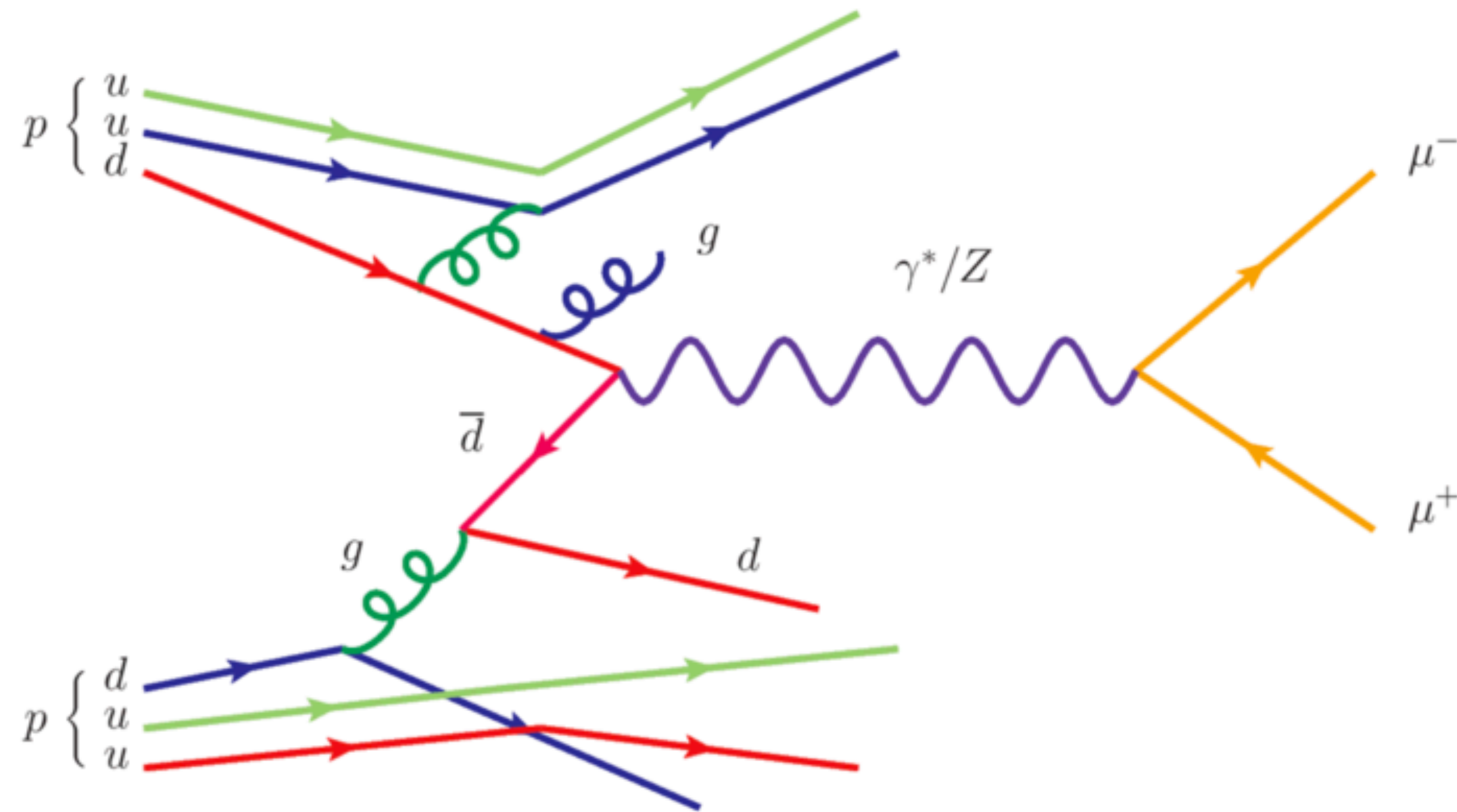
Outline

- Introduction: PDF and SMEFT global fits
- SIMUnet: methodology overview and features
- SIMUnet: Applications
- SIMUnet Release



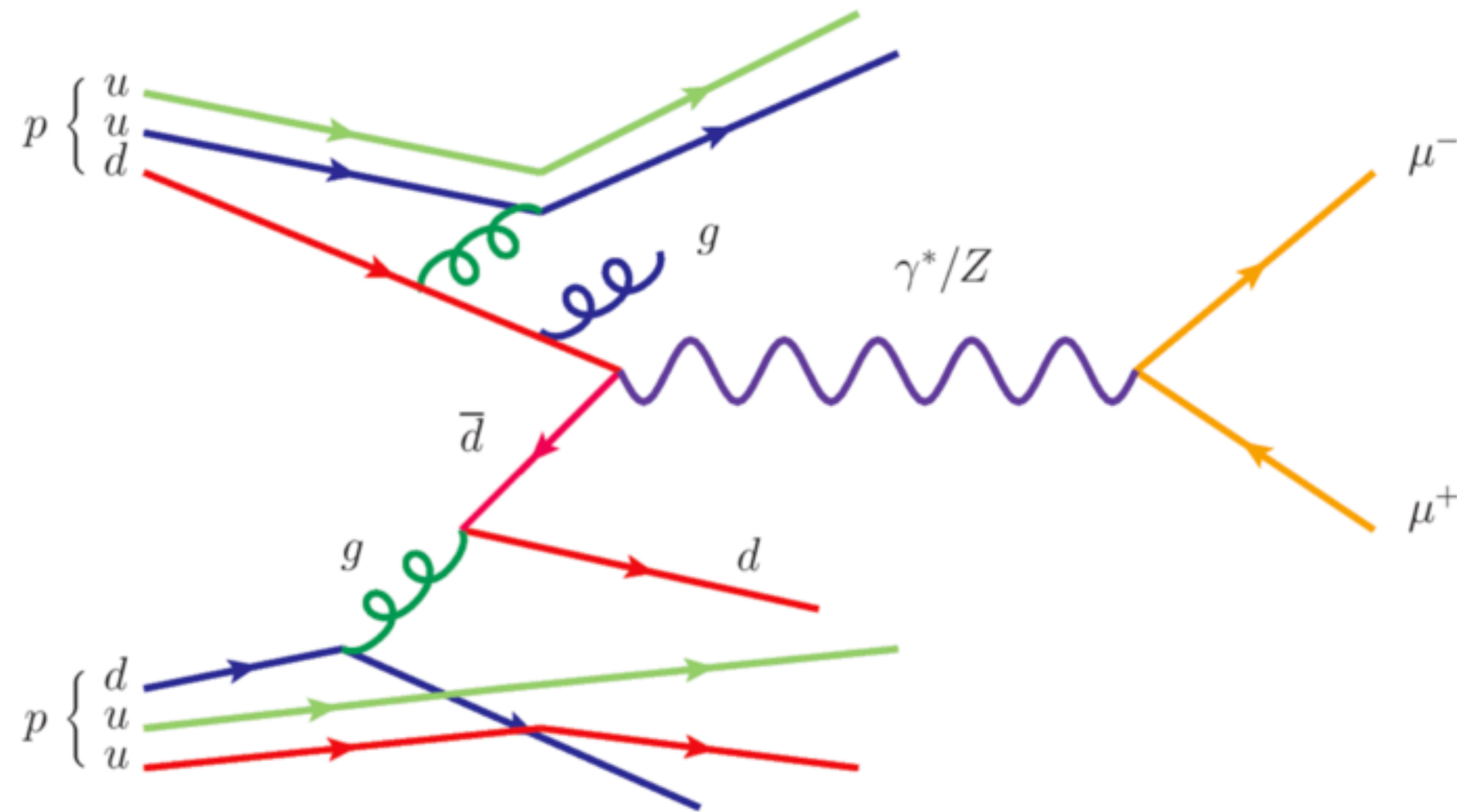
Introduction

Fitting PDFs and Physical Parameters



- **Theory predictions** for collider experiments are obtained from the standard **factorisation formula**

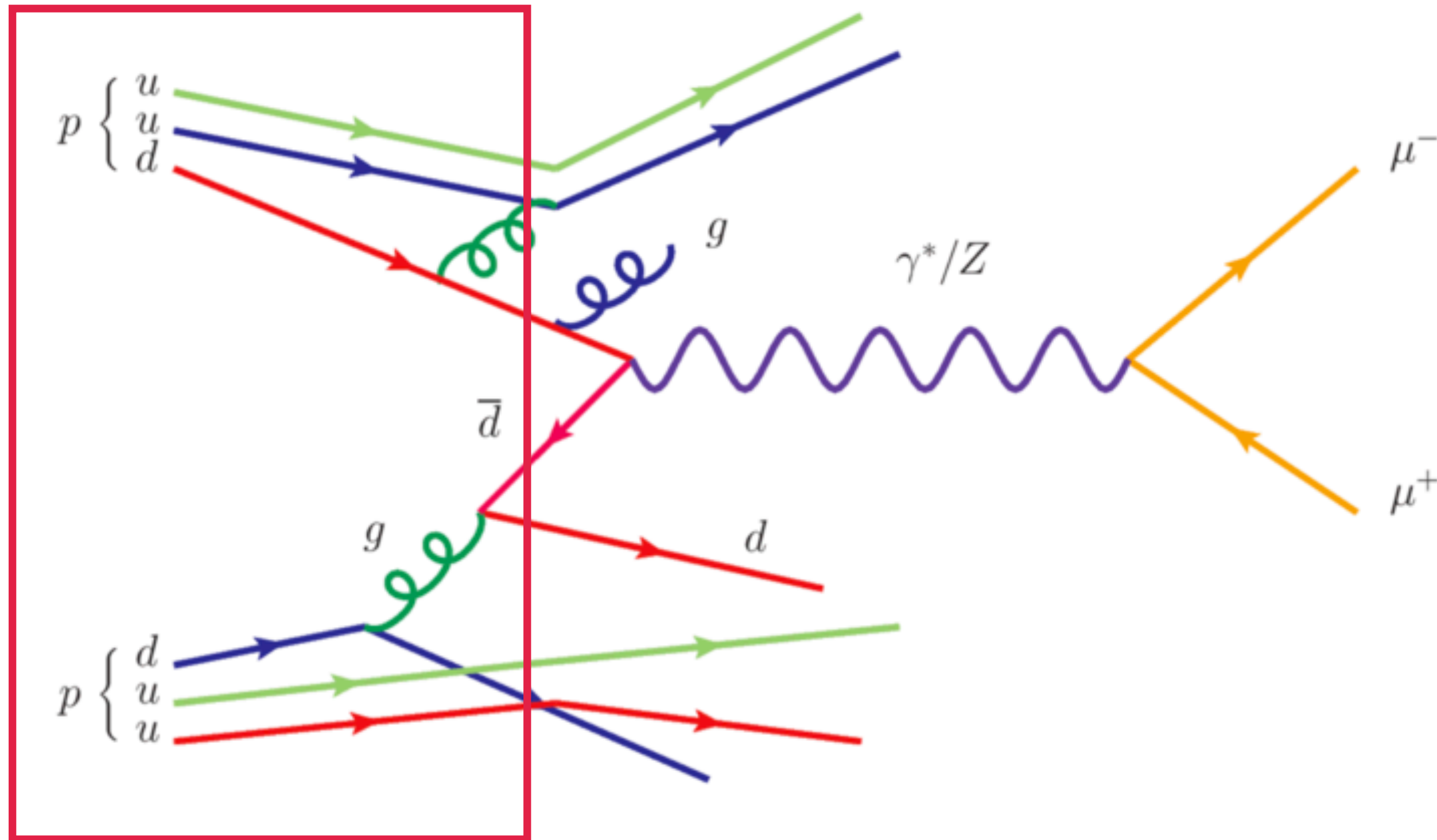
Fitting PDFs and Physical Parameters



- **Theory predictions** for collider experiments are obtained from the standard **factorisation formula**

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

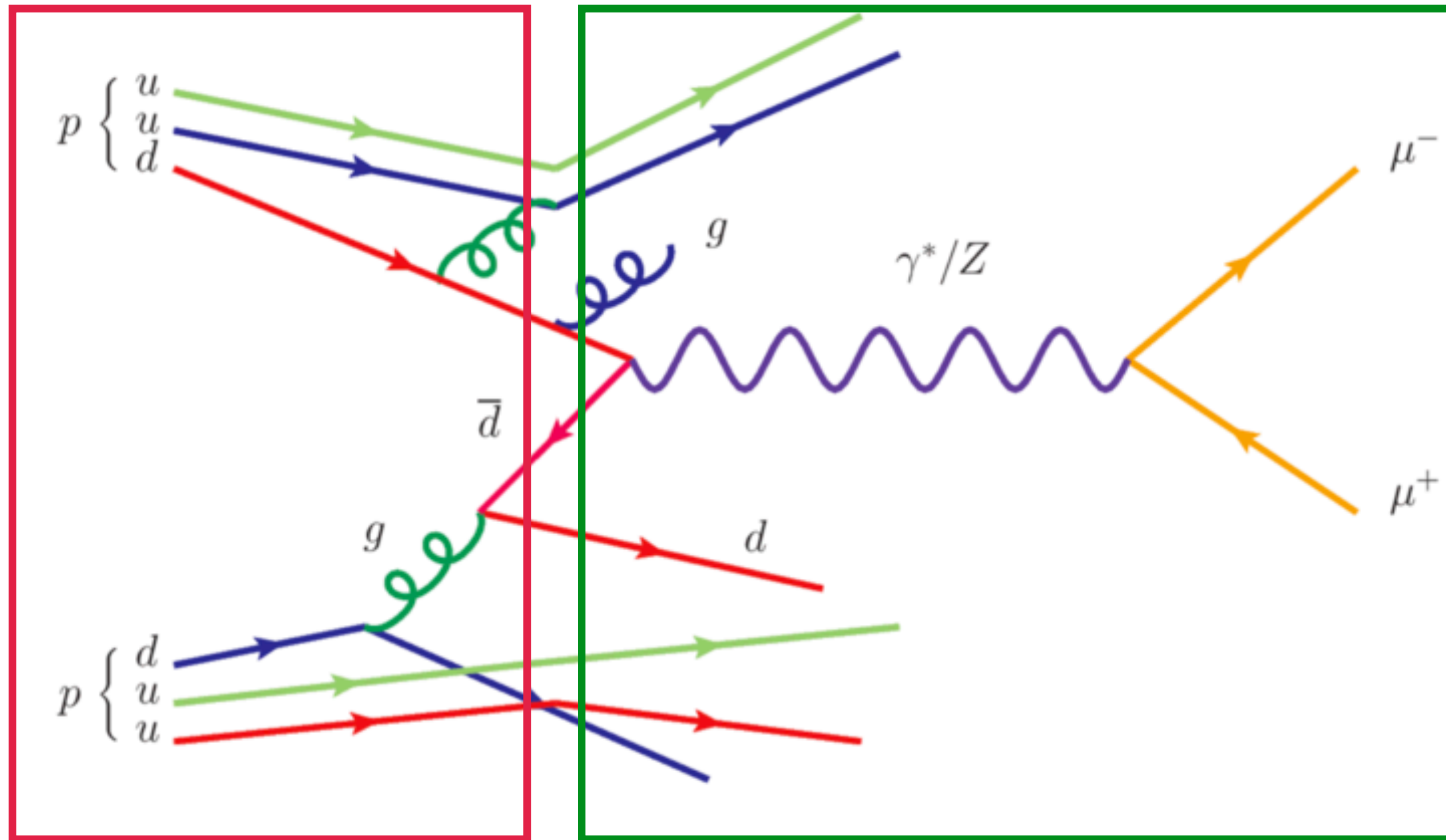
Fitting PDFs and Physical Parameters



- **Theory predictions** for collider experiments are obtained from the standard **factorisation formula**

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

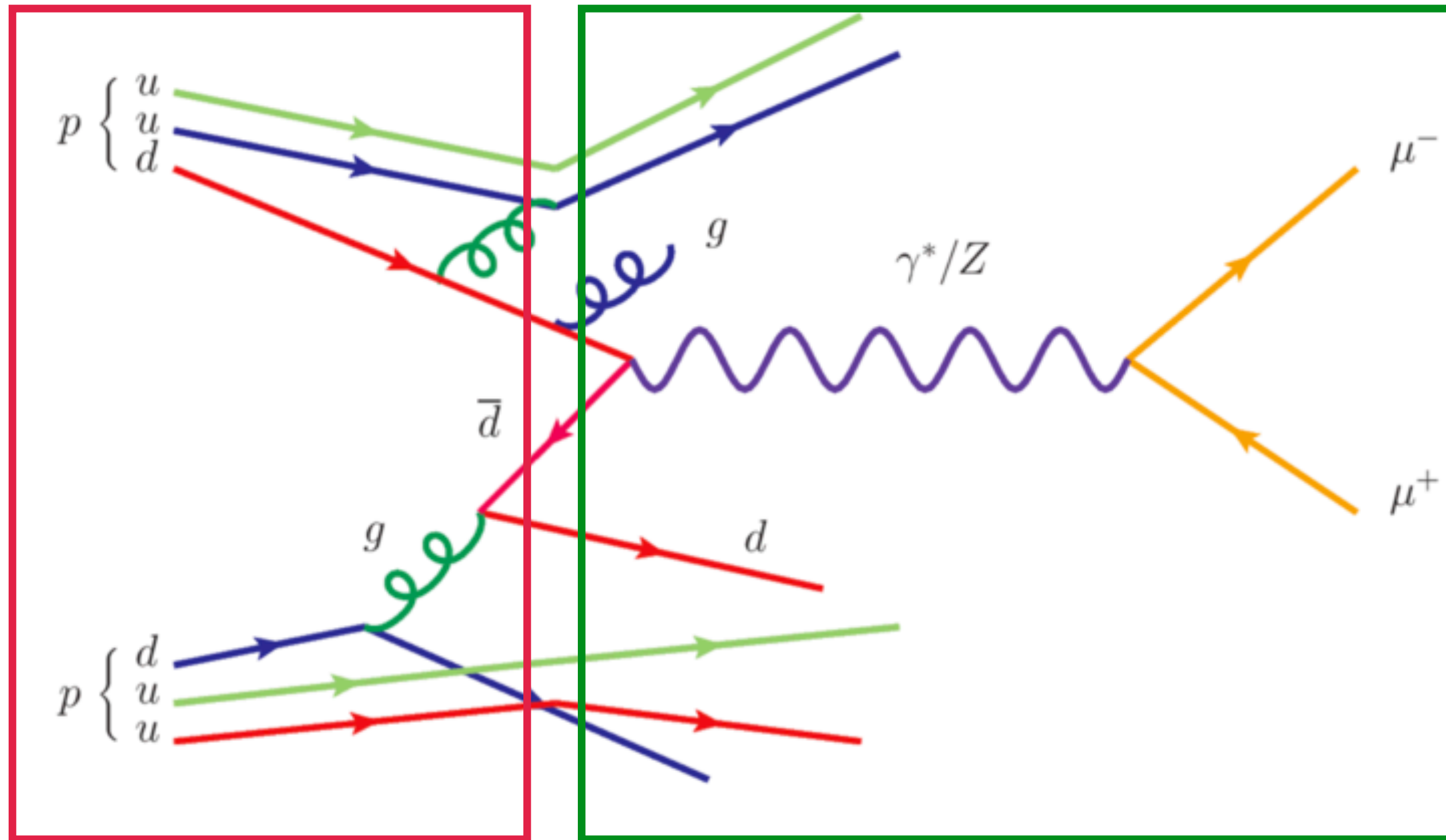
Fitting PDFs and Physical Parameters



- **Theory predictions** for collider experiments are obtained from the standard **factorisation formula**

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

Fitting PDFs and Physical Parameters



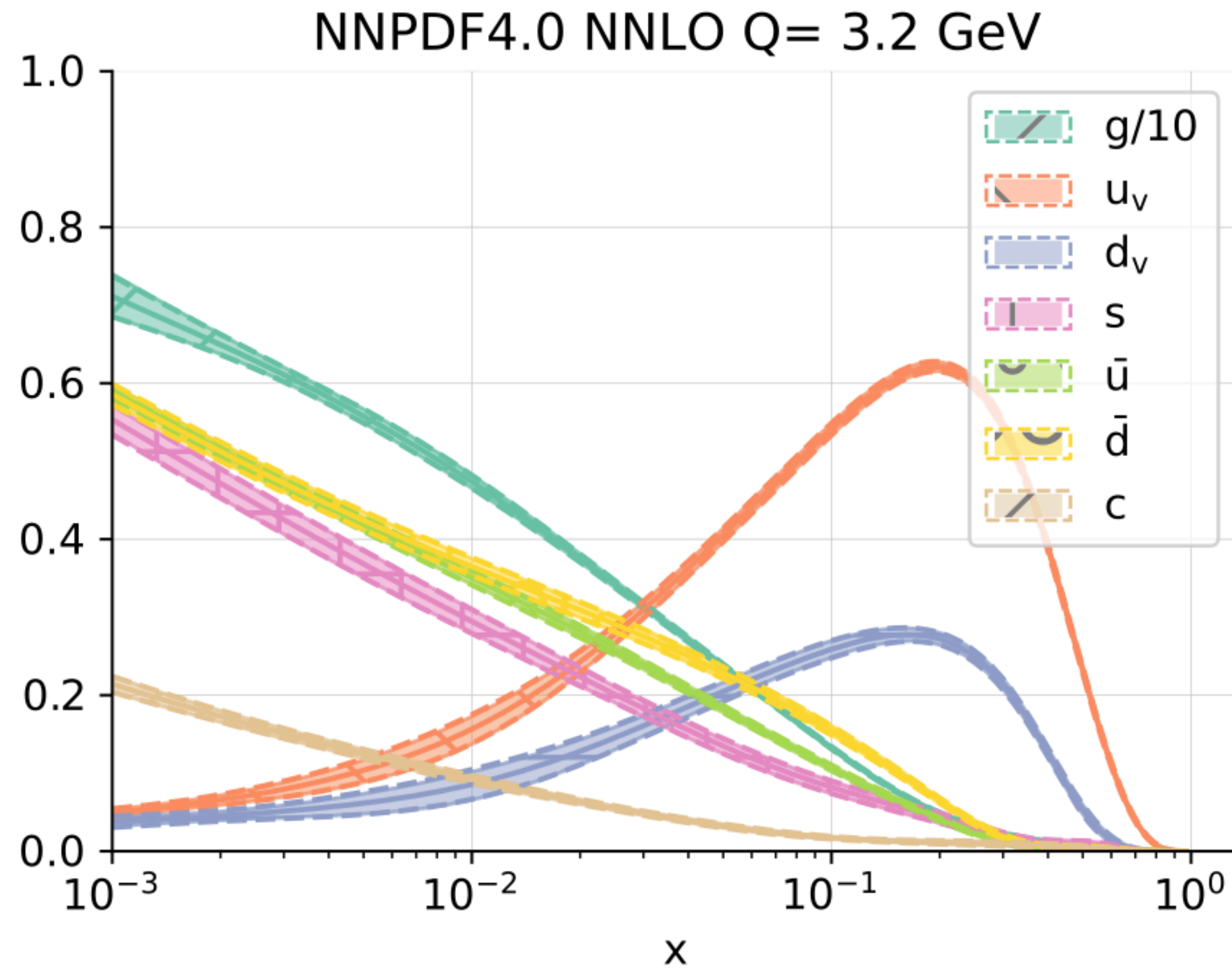
- **Theory predictions** for collider experiments are obtained from the standard **factorisation formula**

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

Theory predictions are functions of :

1. Physics parameters c : e.g. α_S , m_W , **Wilson Coefficients** if we use the **SMEFT**
2. PDF parameters θ e.g. the weights of the neural network parametrising the PDF at the initial scale

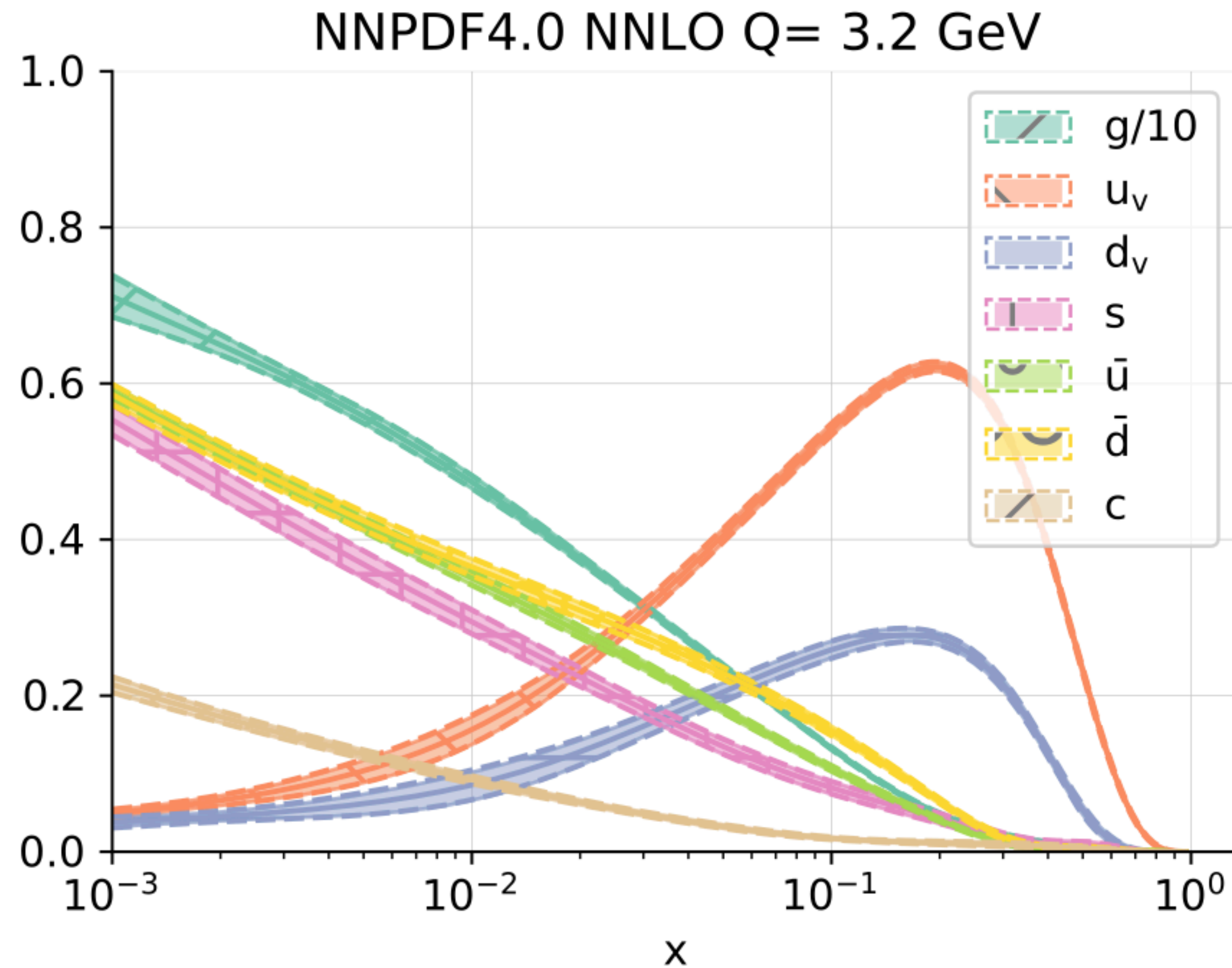
PDF Parameter Fits



- Global fits are used for the extraction of PDFs

NNPDF40, Ball et al, arXiv: 2109.02653

PDF Parameter Fits

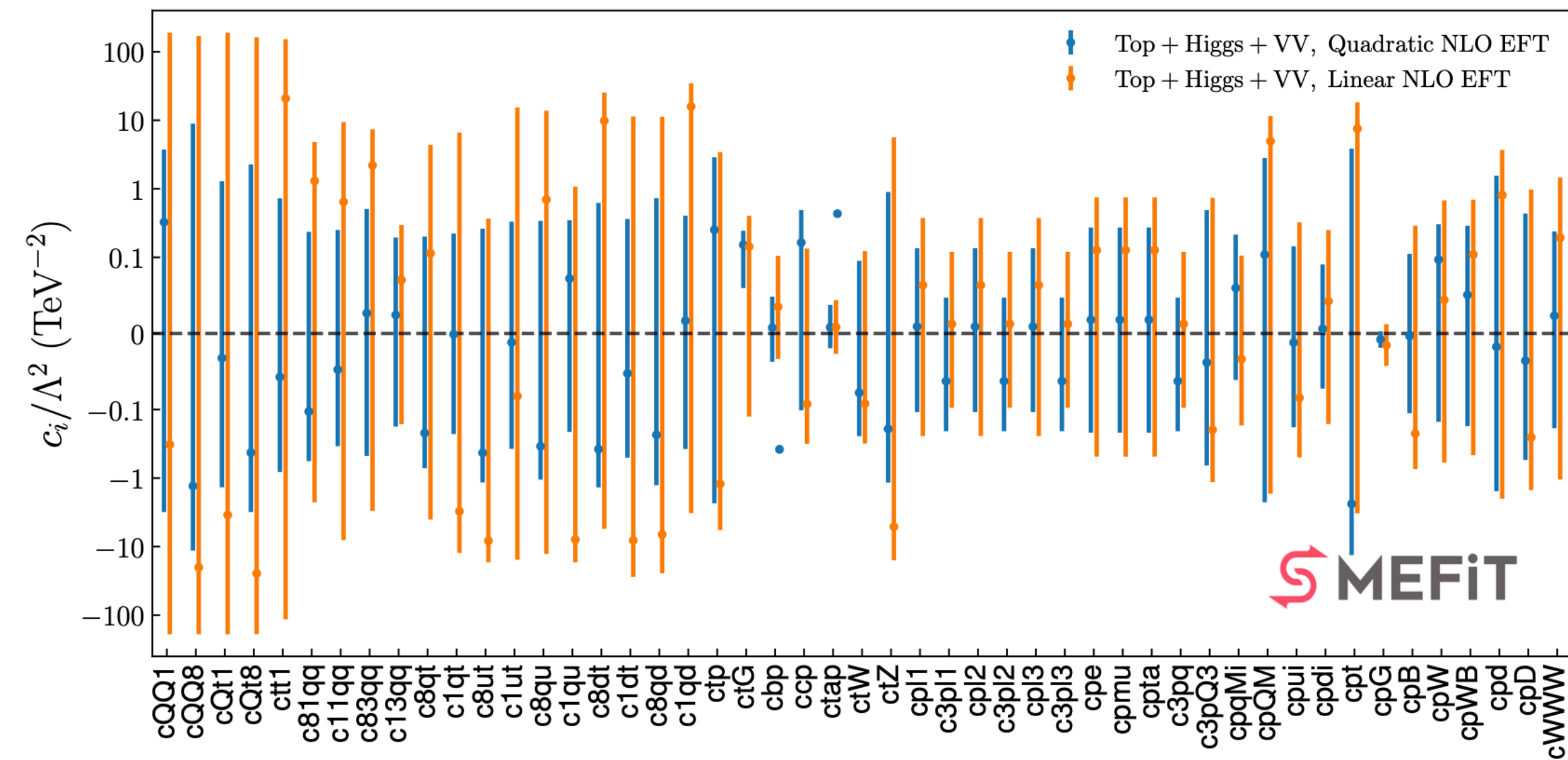


- Global fits are used for the extraction of PDFs
- In a PDF fit the physical parameter c is usually kept fixed

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

NNPDF40, Ball et al, arXiv: 2109.02653

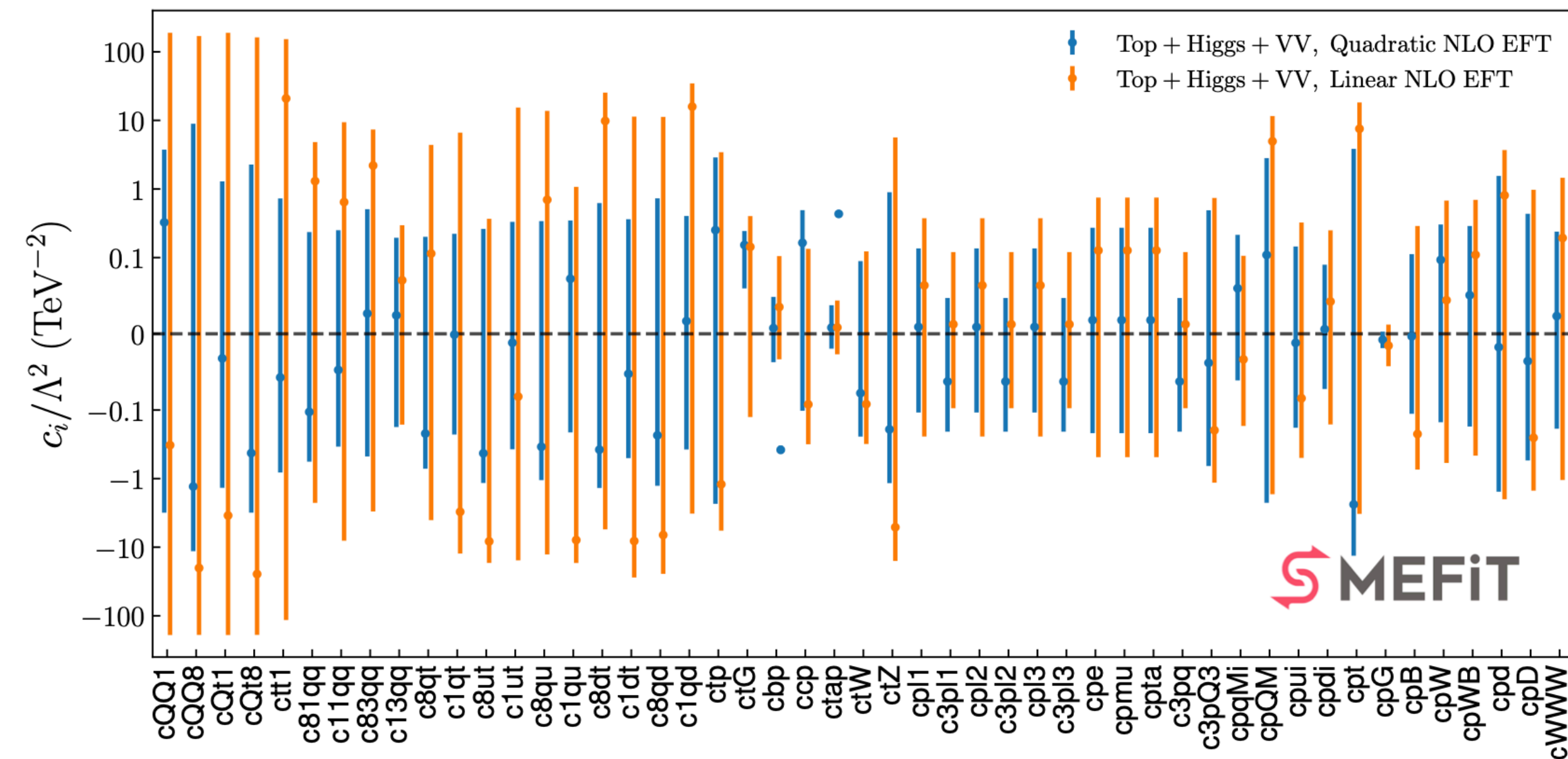
SMEFT Fits



- Global fits are used to fit Wilson Coefficients

SMEFiT, Ethier et al, arXiv: 2105.00006

SMEFT Fits



- Global fits are used to fit Wilson Coefficients
- In an EFT fit the PDF parameters θ are usually kept fixed

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

SMEFiT, Ethier et al, arXiv: 2105.00006

Simultaneous PDF and SMEFT Fits

- Fitting PDF and SMEFT parameters separately can lead to inconsistencies
 - Fitted PDFs can depend implicitly on fixed physical parameters used in the fit.
 - Bounds on physical parameters can depend implicitly on the fixed PDF set used in the fit.

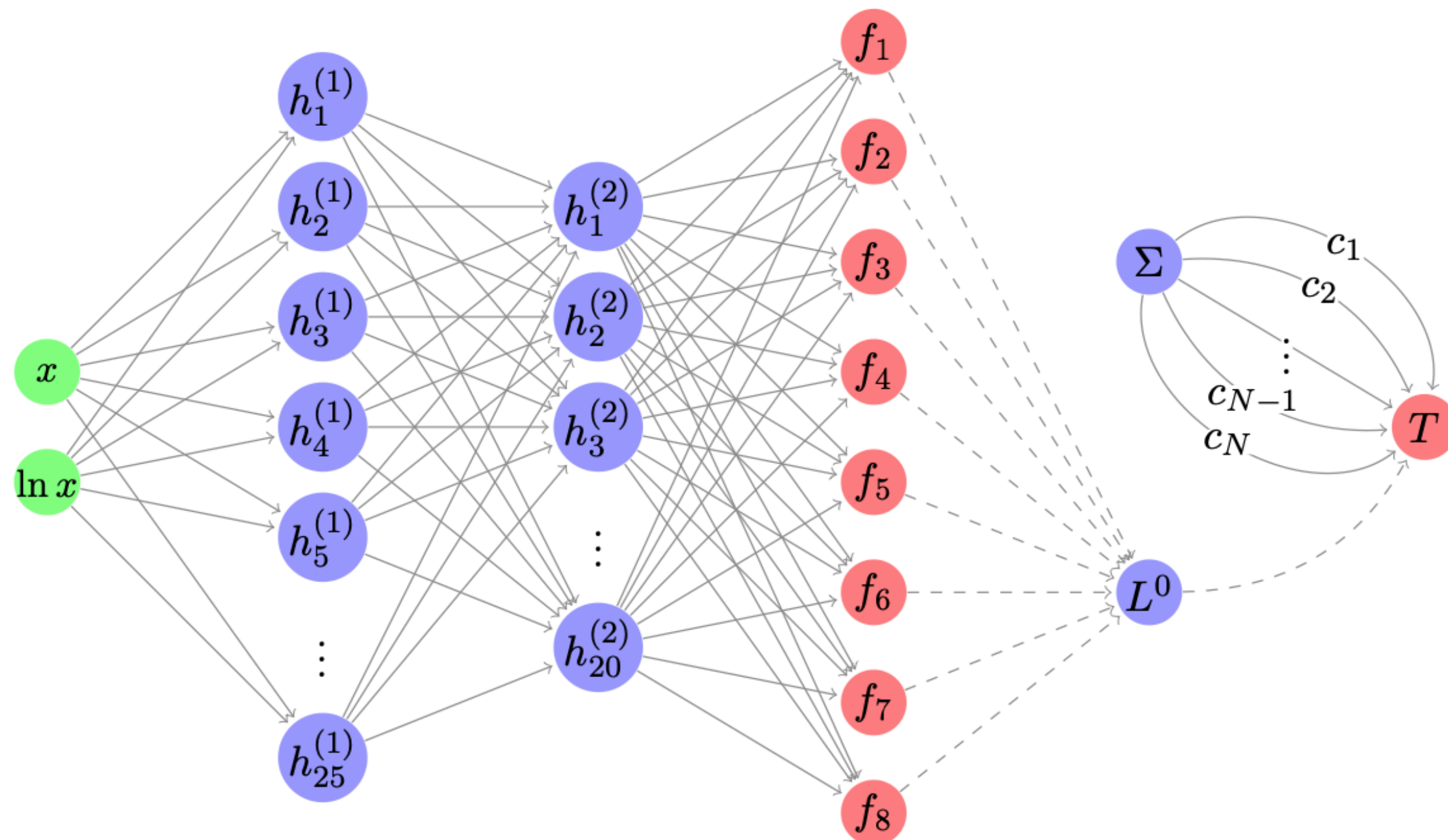
$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$



SIMUnet

SIMUnet: Methodology

Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction

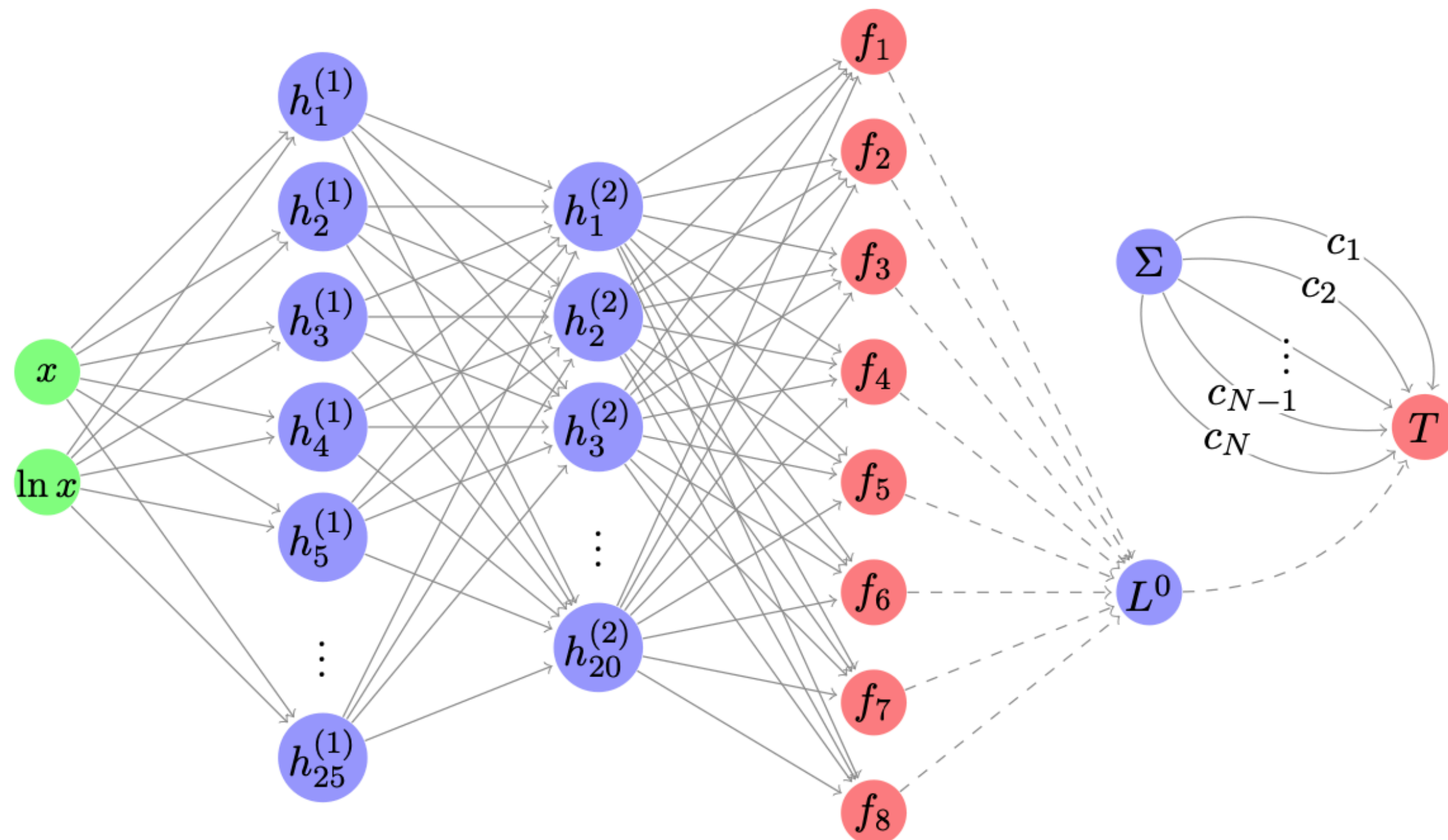


- Simultaneous fit of PDF and theory parameters (e.g. EFT WC)

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

SIMUnet: Methodology

Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction



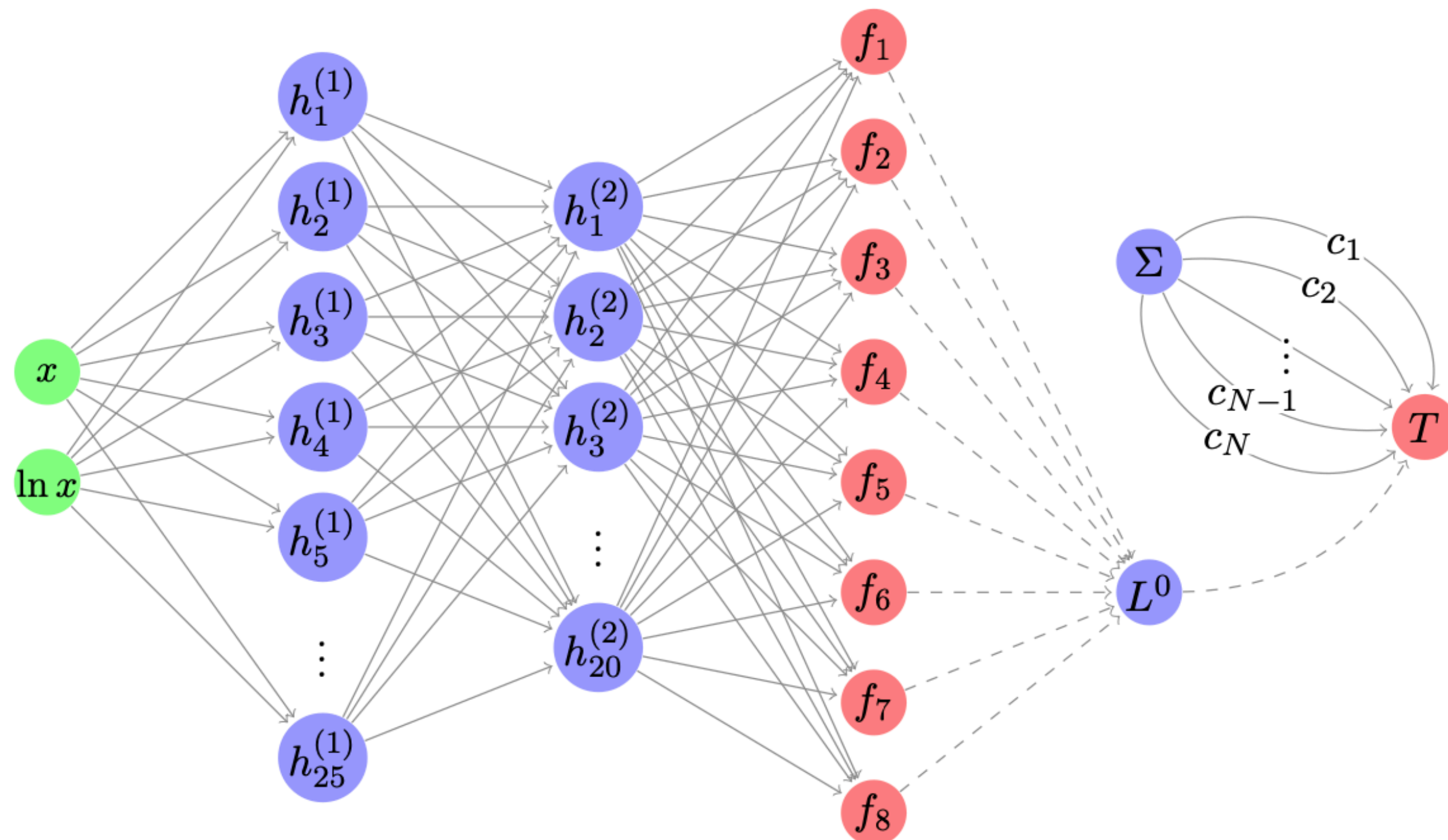
- Simultaneous fit of PDF and theory parameters (e.g. EFT WC)

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

- Extension of NNPDF₄₀ framework to account for theory parameters dependence of partonic cross sections

SIMUnet: Methodology

Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction



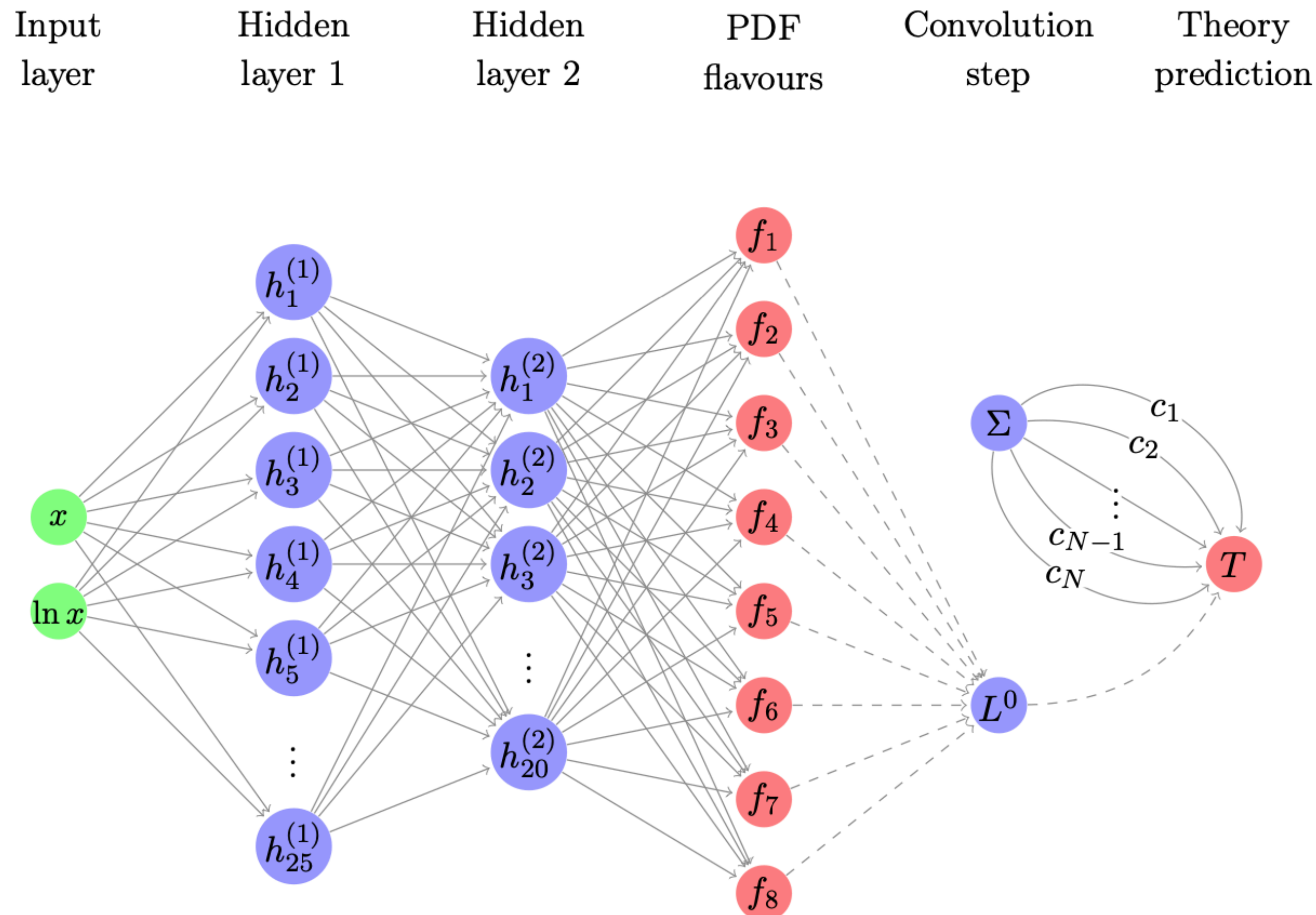
- Simultaneous fit of PDF and theory parameters (e.g. EFT WC)

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

- Extension of NNPDF₄₀ framework to account for theory parameters dependence of partonic cross sections

➔ Extra combination layer to the NNPDF₄₀ Neural Network

SIMUnet: Methodology



- Simultaneous fit of PDF and theory parameters (e.g. EFT WC)

$$T(\{\theta\}, \{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

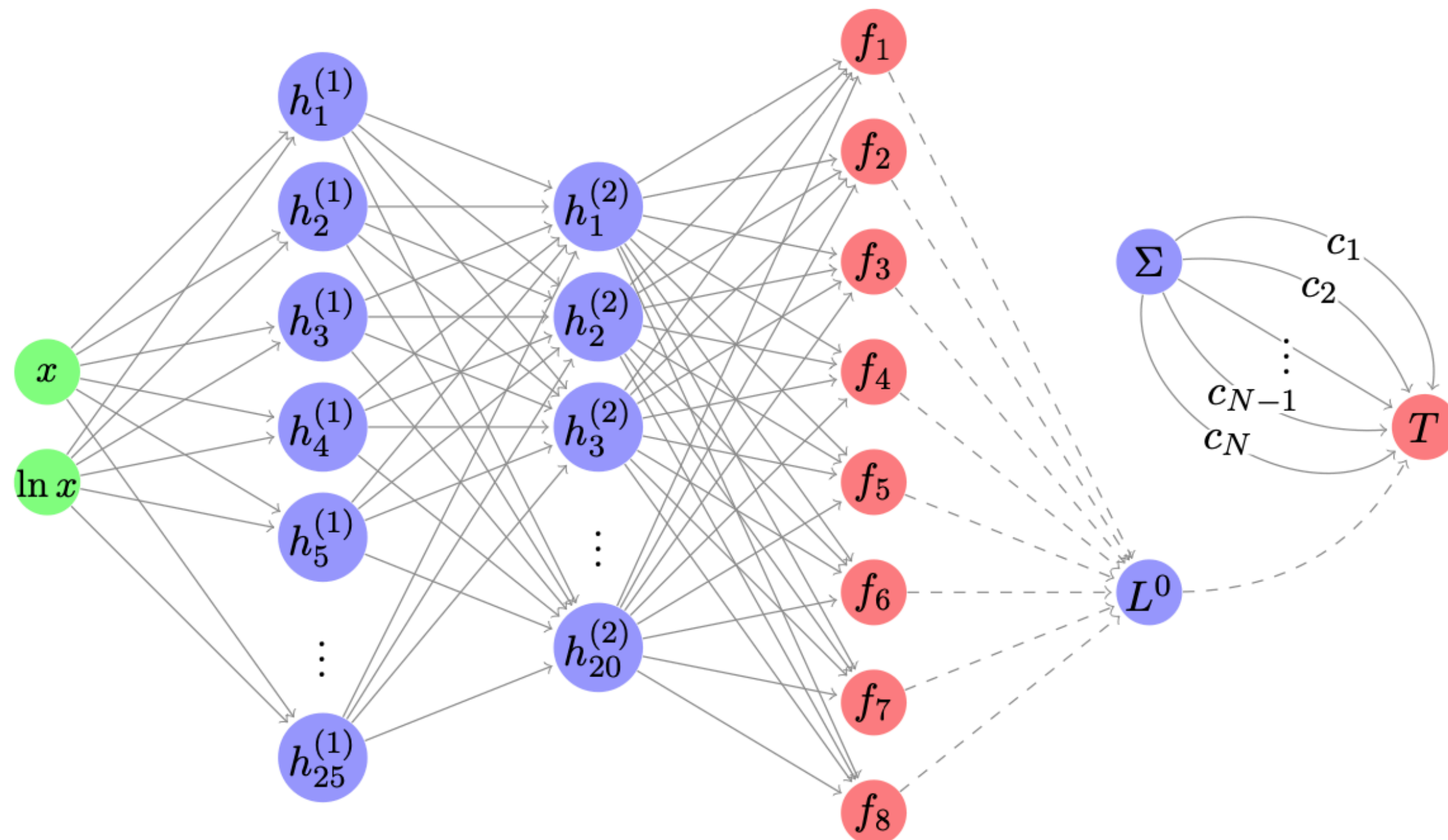
- Extension of NNPDF₄₀ framework to account for theory parameters dependence of partonic cross sections

→ Extra combination layer to the NNPDF₄₀ Neural Network

→ SMEFT corrections added in the form of K-factors

SIMUnet: Methodology

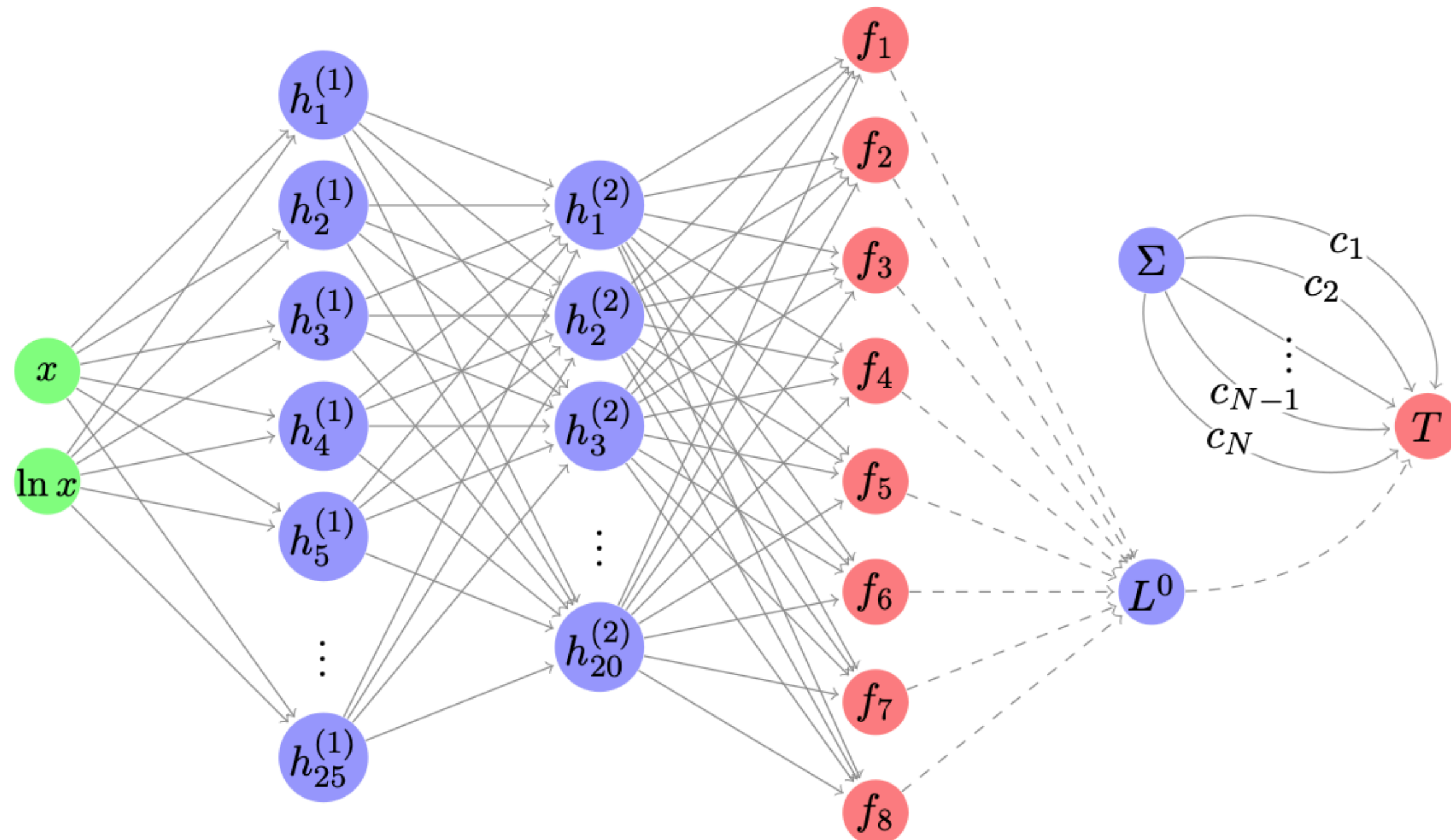
Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction



- **Uncertainty Quantification:** NNPDF Monte Carlo replica method.

SIMUnet: Methodology

Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction

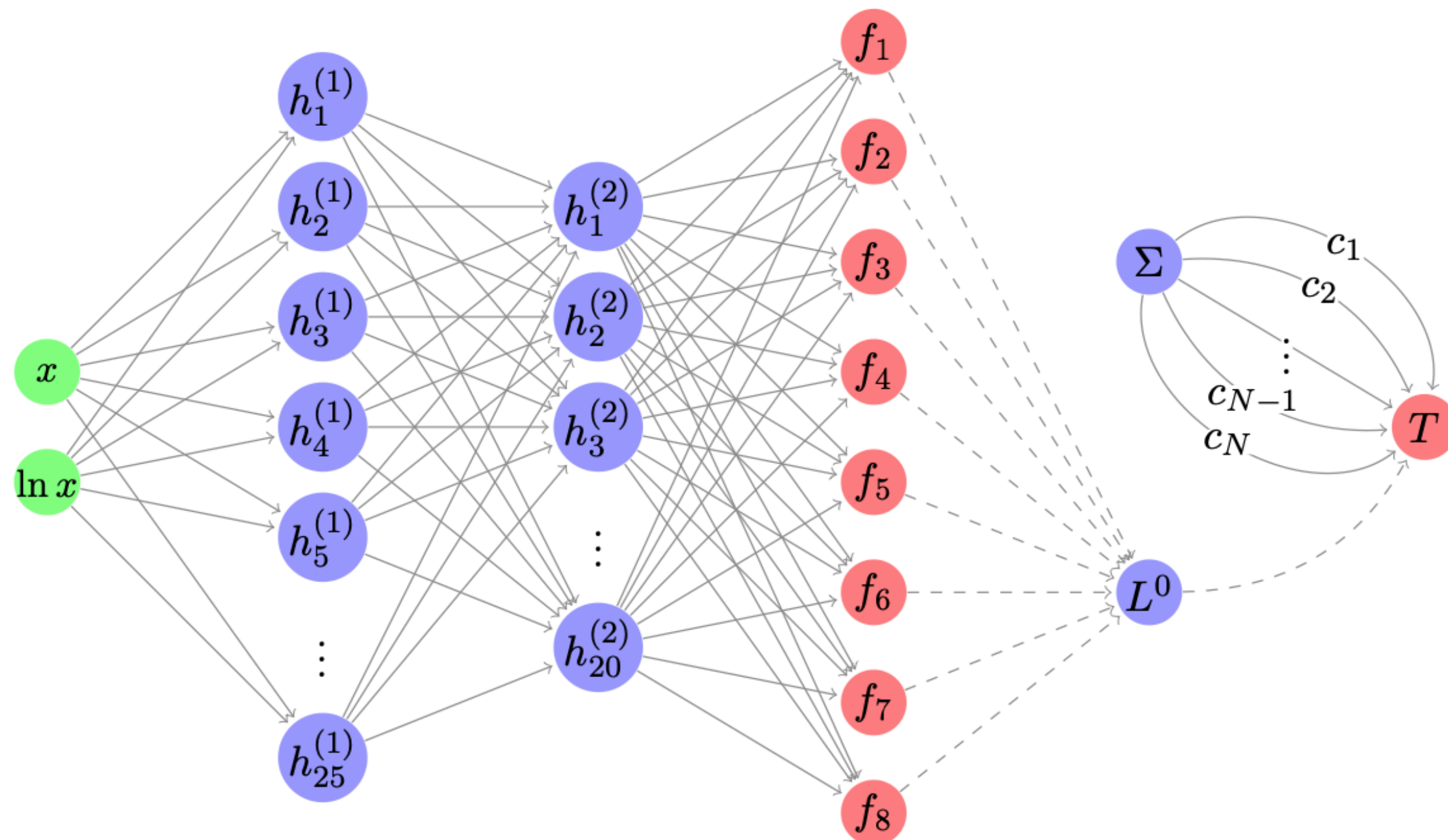


- **Uncertainty Quantification:** NNPDF Monte Carlo replica method.

→ Generate N_{rep} data replica
 $d_i \sim \mathcal{N}(d, C_{\text{exp}}), i \in 1, \dots, N_{\text{rep}}$

SIMUnet: Methodology

Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction

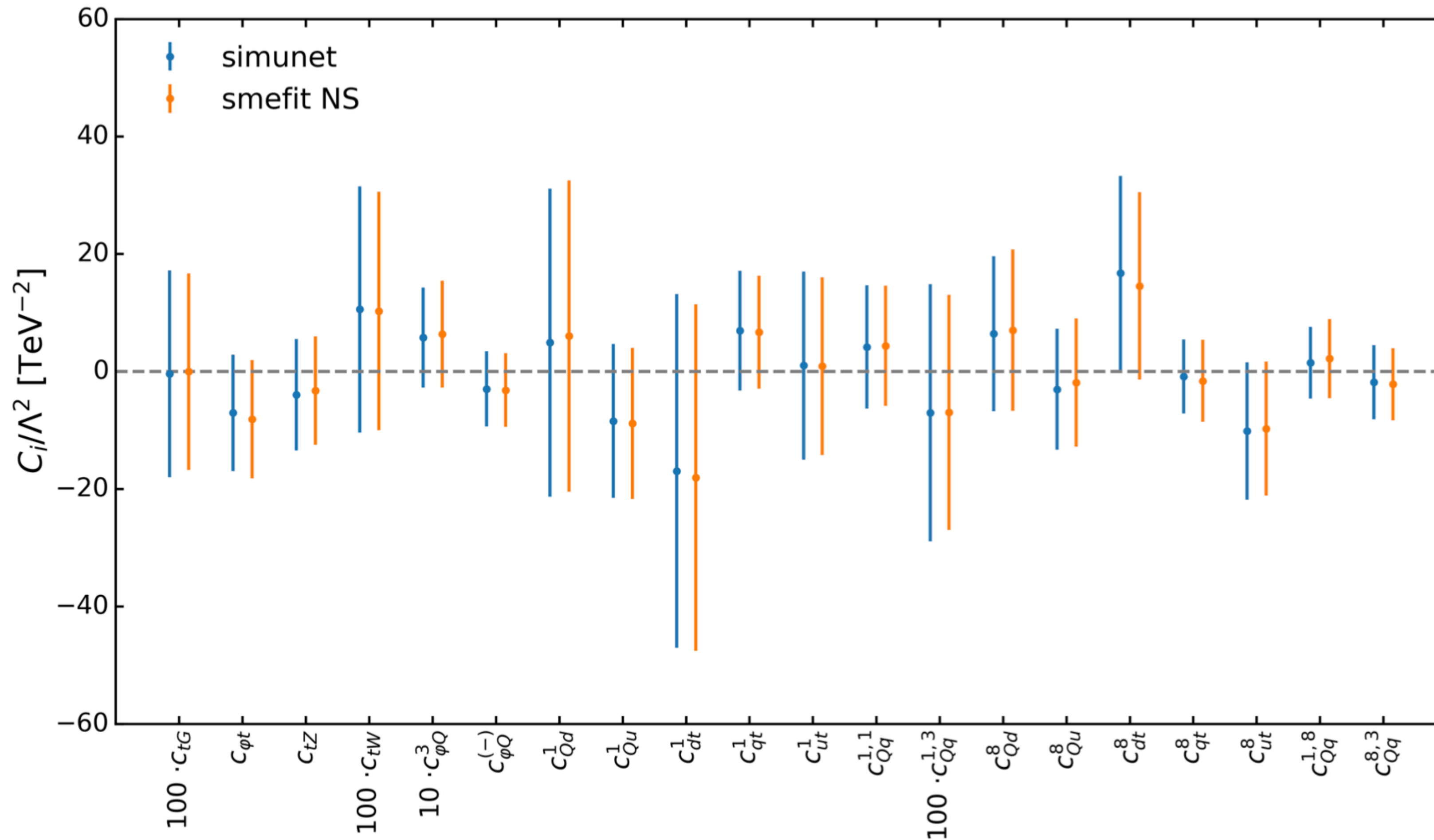


- **Uncertainty Quantification:** NNPDF Monte Carlo replica method.

→ Generate N_{rep} data replica
 $d_i \sim \mathcal{N}(d, C_{\text{exp}}), i \in 1, \dots, N_{\text{rep}}$

→ Fit each replica separately so as to get a distribution in the space of PDFs

SIMUnet: Fixed-PDF EFT fits



- Freeze weights parametrising PDFs ($\theta = \bar{\theta}$) and optimise only Wilson Coefficients c

SIMUnet: Inject New Physics in the data

- Assume knowledge of “true” PDF f_{in} and use it to generate fake-data

$$D_{\text{SM}} = T[f_{\text{in}}] + \eta, \eta \sim \mathcal{N}(0, C_{\text{exp}}^{t0})$$

- Assume a certain BSM scenario and inject new physics in fake-data

$$D_{\text{BSM}}^* = D_{\text{SM}} \left(1 + cK_{\text{lin}} + c^2K_{\text{quad}} \right)$$

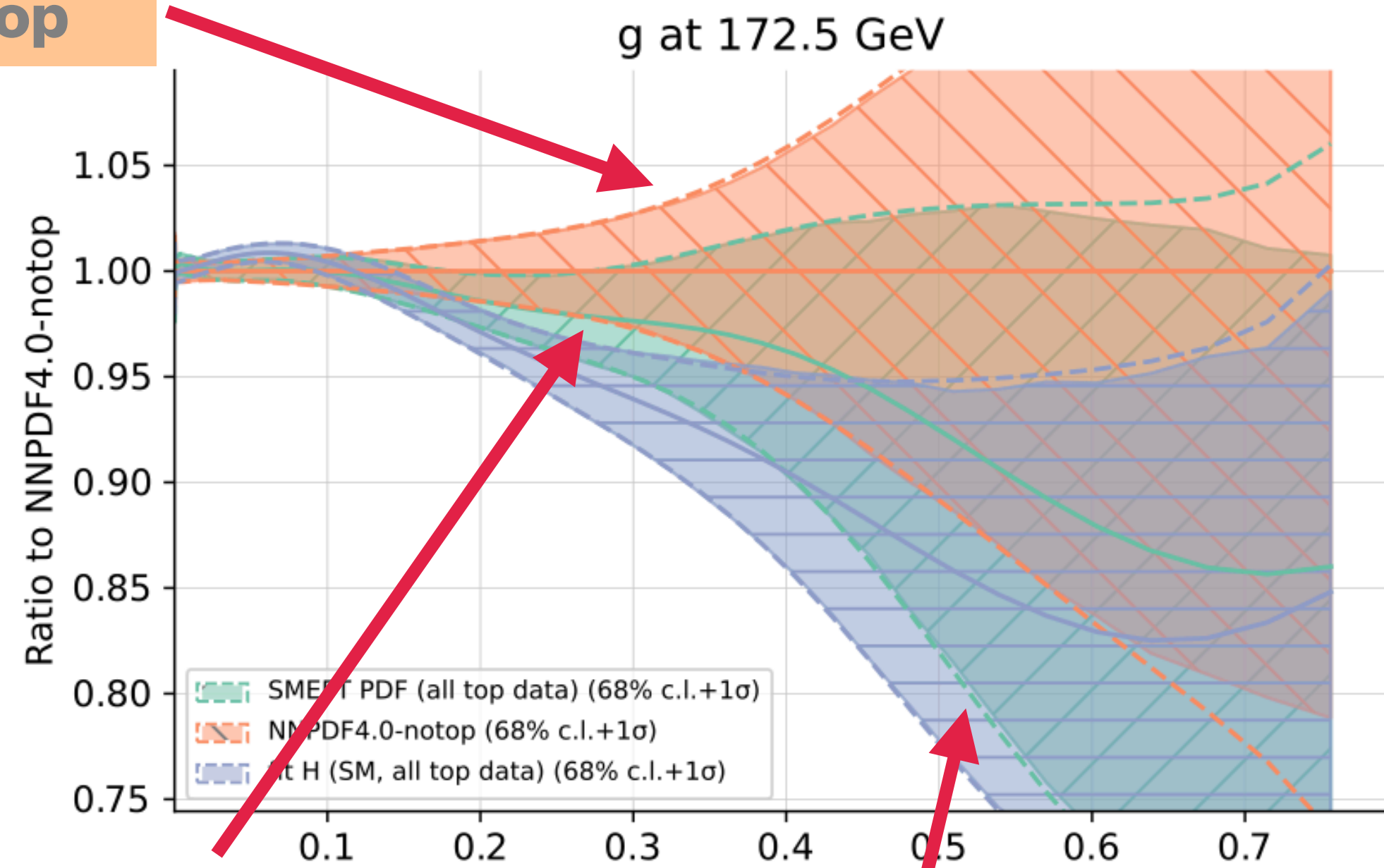
- Check effect of fitting assuming SM: can PDFs absorb new physics?

→ arXiv: 2307.10370

SIMUnet: Applications

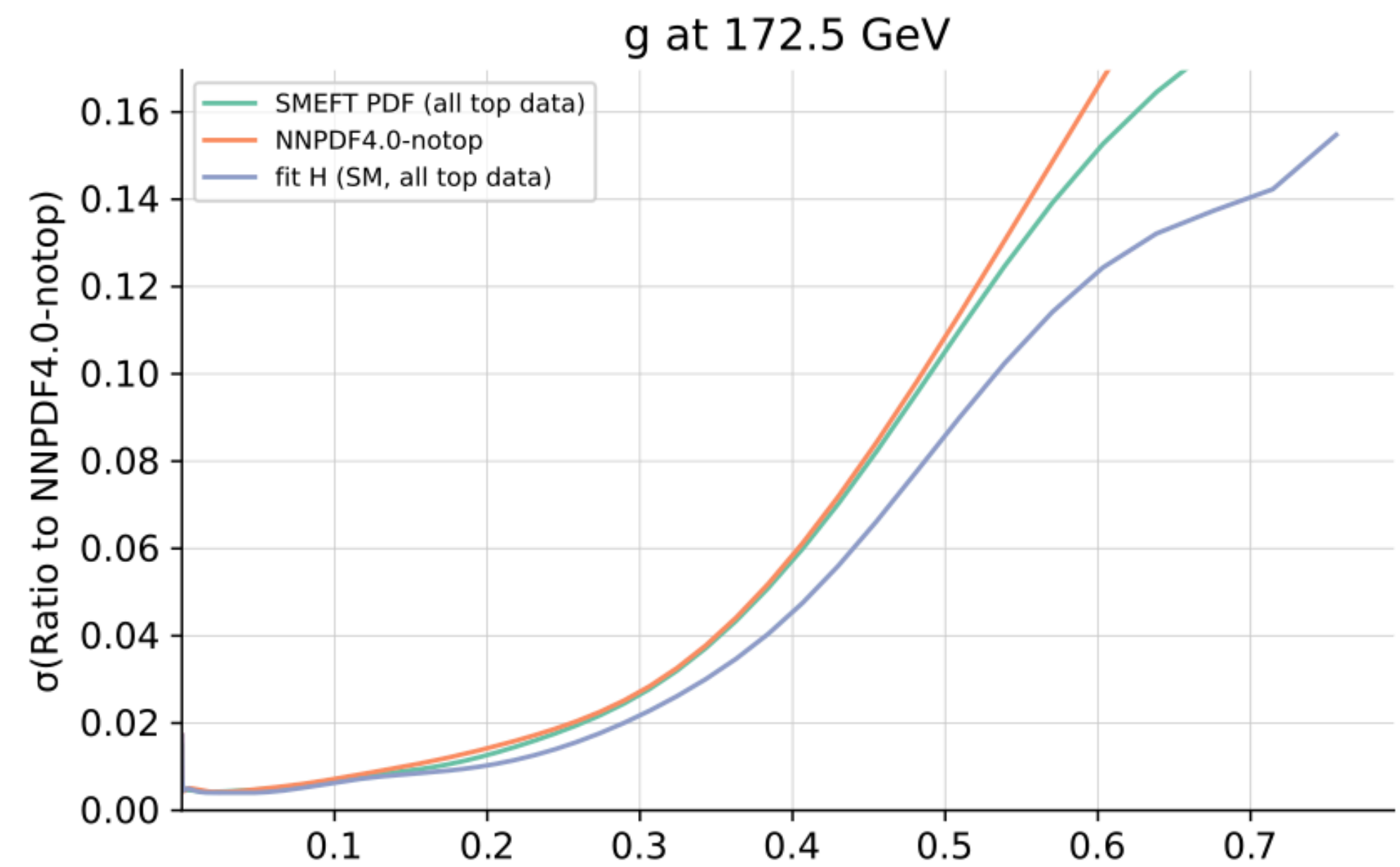
Simultaneous PDF SMEFT Fits: PDFs

NNPDF4.0,
no-top



SMEFT PDF

NNPDF4.0,
all top data



- Marked difference between SM gluon PDF and SMEFT gluon PDF

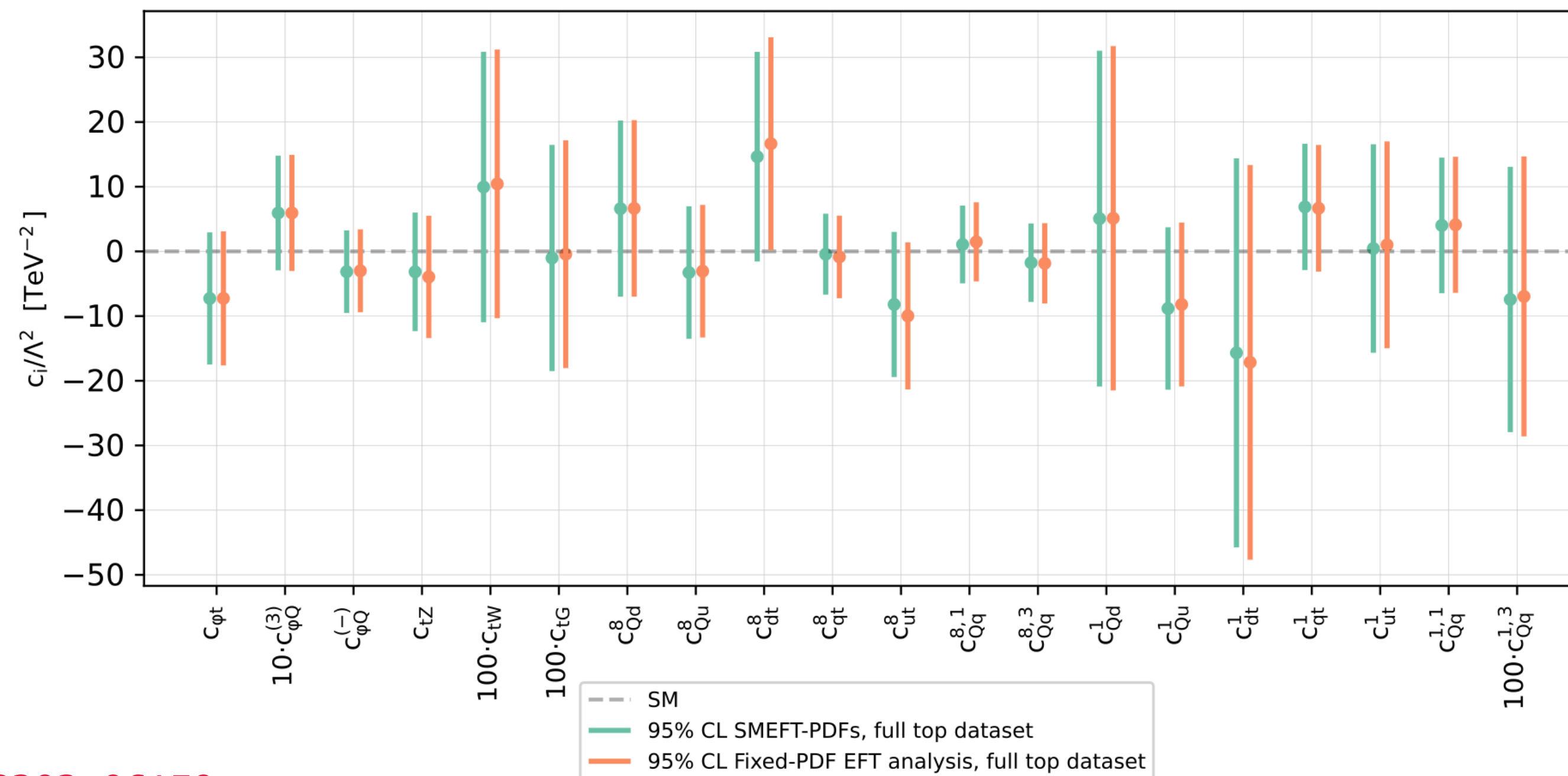
→ Larger PDF uncertainties

Simultaneous PDF SMEFT Fits: EFT

- Moderate impact at the level of the bounds on the EFT coefficients
- Mild shift in the central values and slight broadening of the uncertainties

Fixed PDF,
EFT fits

SMEFT-PDF
fits



PDF-SMEFT interplay questions

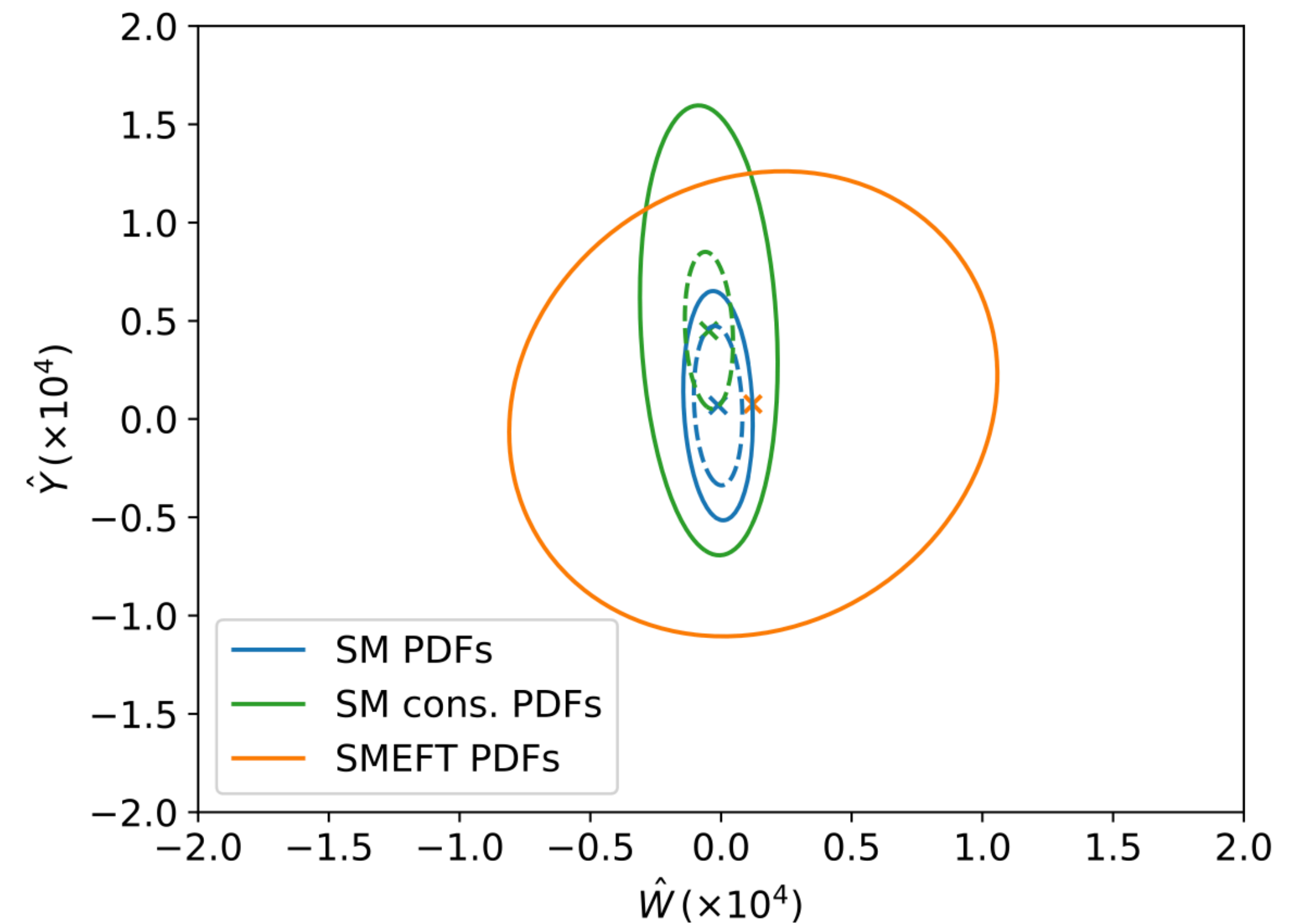
Question: won't the PDF-SMEFT interplay be negligible?

- 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 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 questions

Question: won't the PDF-SMEFT interplay be negligible?

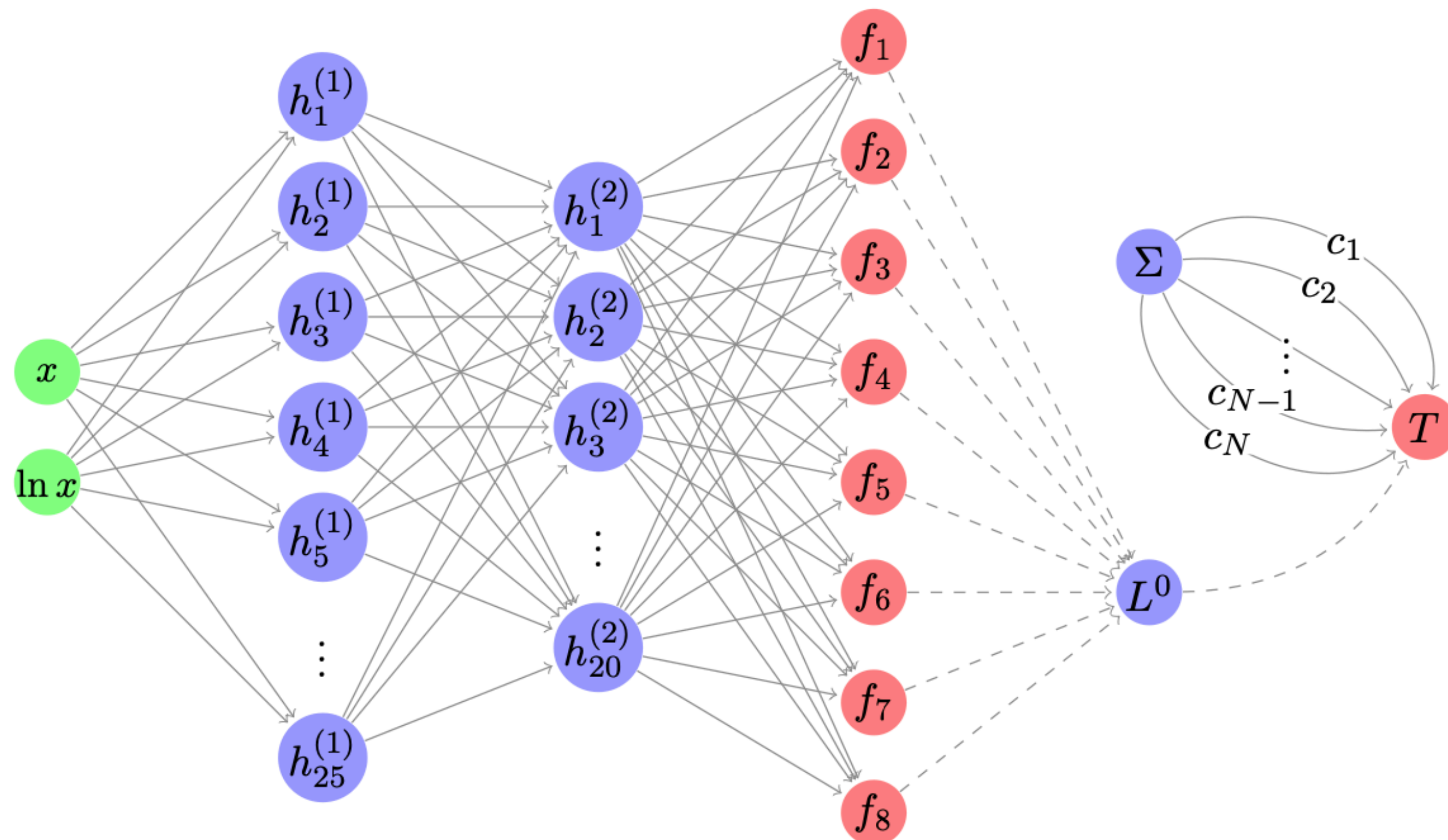
→ However, it was also shown in Greljo et al., 2104.02723, that interplay is very important between the \hat{W} , \hat{Y} operators and PDFs when using project high luminosity DY data.



SIMUnet: ReLease

SIMUnet Release

Input layer Hidden layer 1 Hidden layer 2 PDF flavours Convolution step Theory prediction



- A tool for **simultaneous** PDF and SMEFT Wilson Coefficients fits

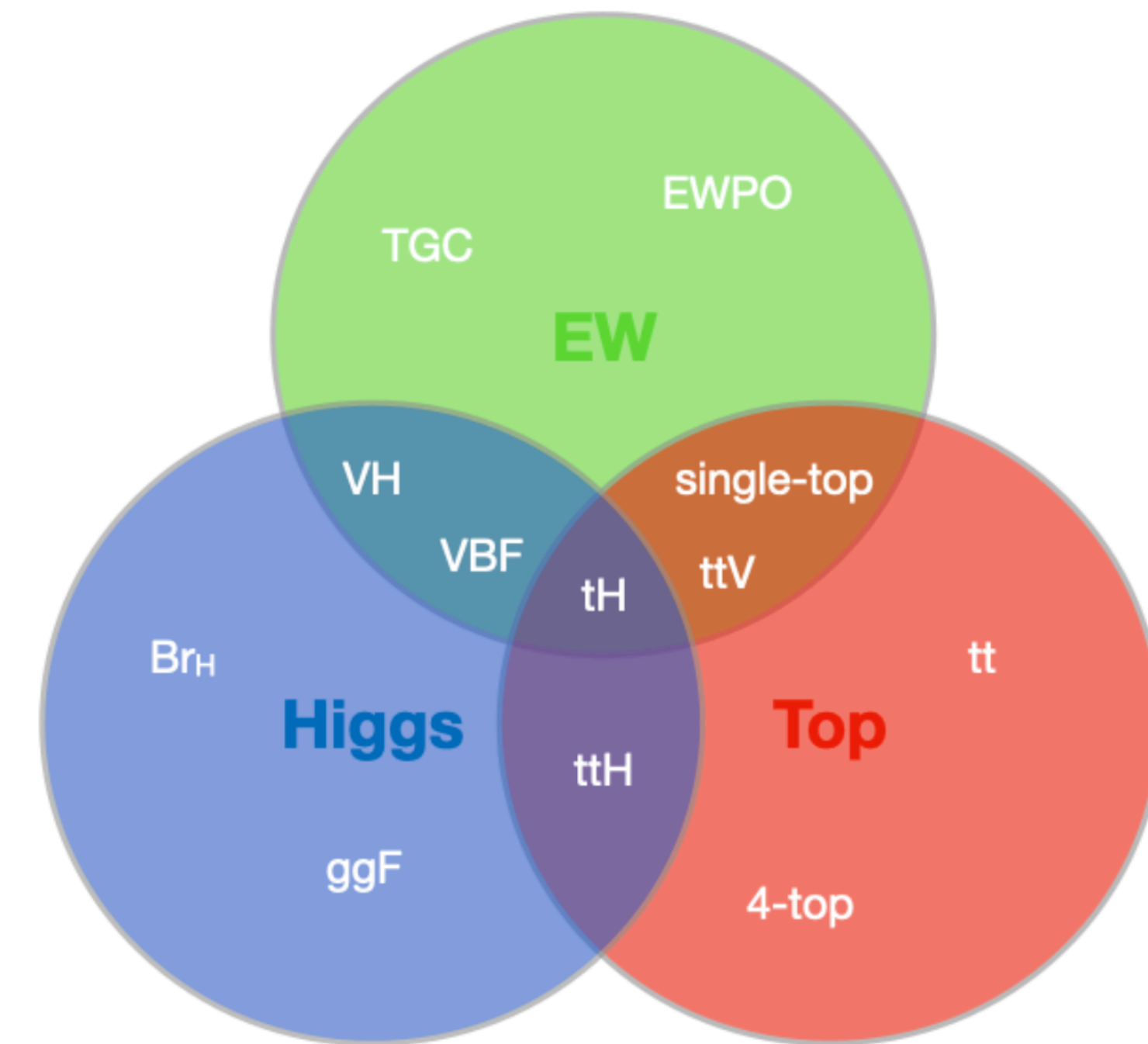
→ **Linear in the SMEFT**

SIMUnet Release: New Data

- **EW precision observables**

→ 44 data points

Dataset Name	N_dat	Reference	Theory Tables
EWPO on Z resonance	19	hep-ex/0509008	SMEFIT
W branching ratio	3	1302.3415	SMEFIT
Bhabha Scattering	21	1302.3415	SMEFIT
Alpha EW	1	PDG	PDG



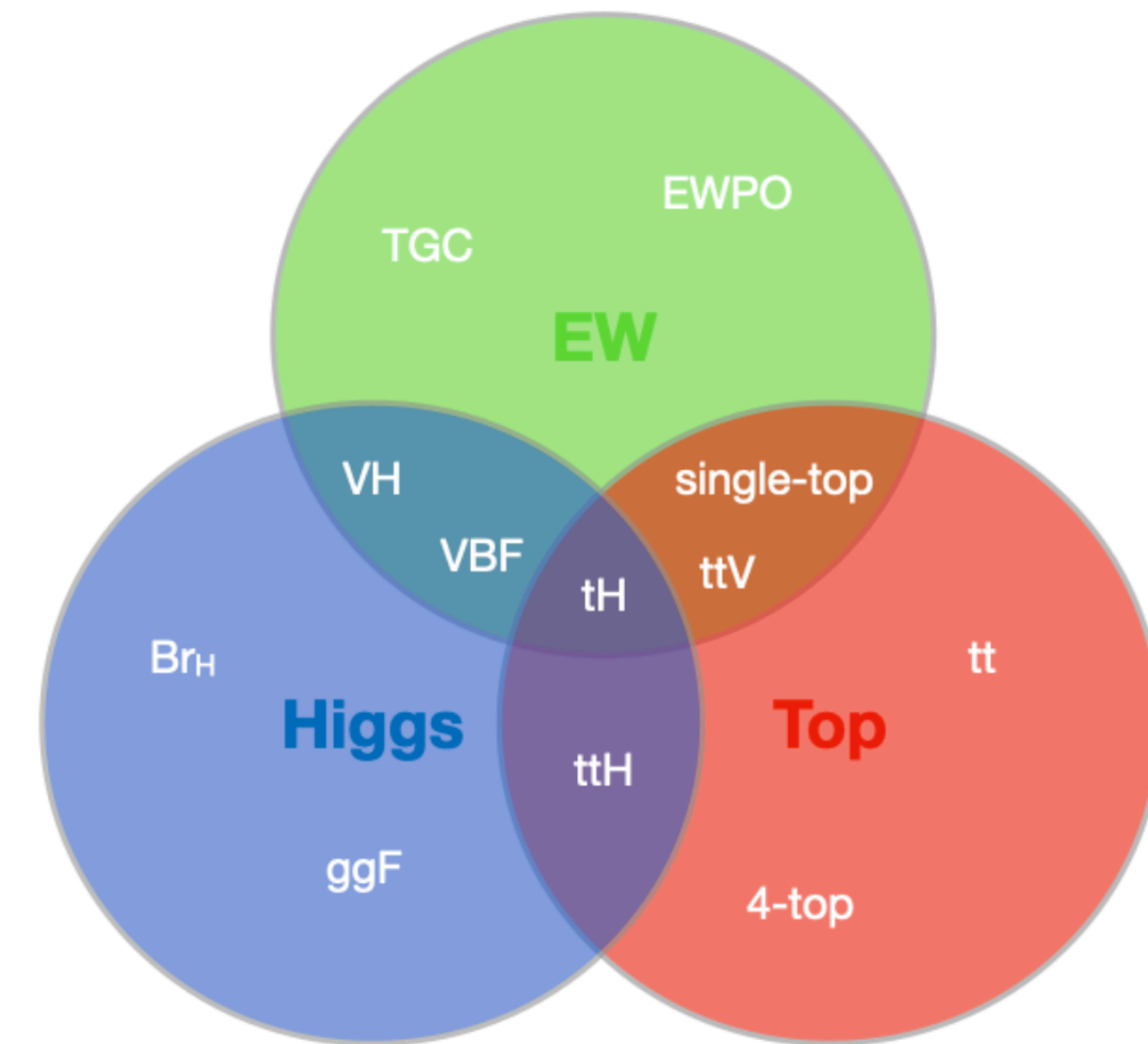
© Ken Mimasu

SIMUnet Release: New Data

- **Diboson production from LEP and LHC**

→ 82 data points

Dataset Name	N_dat	Reference	Theory Tables
LEP W- differential angular xsec	40	1302.3415	SMEFiT
ATLAS, W+W- differential xsec	13	1905.04242	SMEFiT
ATLAS, WZ transverse mass	6	1902.05759	SMEFiT
CMS, WZ transverse momentum	11	1901.03428	SMEFiT
ATLAS, Zjj azimuthal differential xsec	12	2006.15458	FitMaker



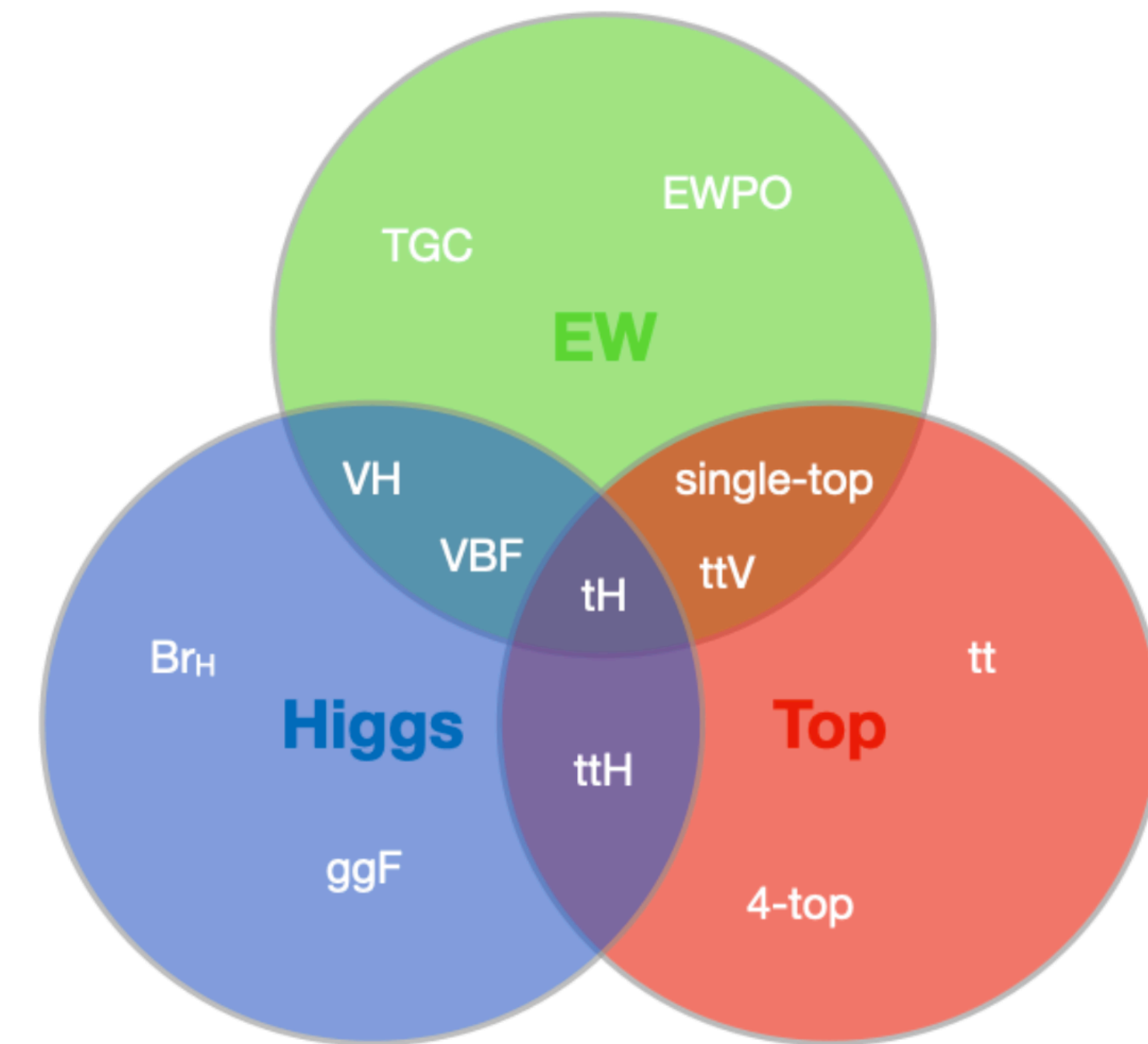
© Ken Mimasu

SIMUnet Release: New Data

- **Higgs production and decay**

→ 73 data points from Run I and II

Dataset Name	N_dat	Reference	Theory Tables
Production and decay rates (8 TeV)	22	1606.02266	SMEFiT
Production and decay rates (13 TeV)	24	1809.10733	SMEFiT
Production xsec and branching fractions	25	1909.02845	FitMaker
Zgamma decay	1	2005.05382	FitMaker
Dimuon decay	1	2007.07830	FitMaker



© Ken Mimasu

SIMUnet Release: Open Source

```
dataset_inputs:
...
- {dataset: ATLAS_TTBAR_8TEV_LJETS_TOTAL, cfac: [QCD], simu_fac: "EFT_NLO"}
...
# # # tW
- {dataset: ATLAS_SINGLETOPW_8TEV_TOTAL, simu_fac: "EFT_NLO", use_fixed_predictions: True}
...

fixed_pdf_fit: False
# load_weights_from_fit: 221103-jmm-no_top_1000_iterated # If this is uncommented, training starts here.

simu_parameters:
- {name: '0tG', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0pt', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0tZ', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0tW', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '03pQ3', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}

...|
```

- **NNPDF4.0 like interface**

SIMUnet Release: Open Source

```
dataset_inputs:
...
- {dataset: ATLAS_TTBAR_8TEV_LJETS_TOTAL, cfac: [QCD], simu_fac: "EFT_NLO"}
...
# # # tW
- {dataset: ATLAS_SINGLETOPW_8TEV_TOTAL, simu_fac: "EFT_NLO", use_fixed_predictions: True}
...

fixed_pdf_fit: False
# load_weights_from_fit: 221103-jmm-no_top_1000_iterated # If this is uncommented, training starts here.

simu_parameters:
- {name: '0tG', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0pt', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0tZ', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0tW', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '03pQ3', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}

...|
```

- **NNPDF4.0 like interface**

→ Similar runcards

SIMUnet Release: Open Source

```
dataset_inputs:
...
- {dataset: ATLAS_TTBAR_8TEV_LJETS_TOTAL, cfac: [QCD], simu_fac: "EFT_NLO"}
...
# # # tW
- {dataset: ATLAS_SINGLETOPW_8TEV_TOTAL, simu_fac: "EFT_NLO", use_fixed_predictions: True}
...

fixed_pdf_fit: False
# load_weights_from_fit: 221103-jmm-no_top_1000_iterated # If this is uncommented, training starts here.

simu_parameters:
- {name: '0tG', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0pt', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0tZ', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '0tW', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: '03pQ3', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
...|
```

- **NNPDF4.0 like interface**

→ Similar runcards

→ Analysis tools e.g. automated reports

SIMUnet Release: Open Source

- SIMUnet website with theory, tutorials and analysis files

The screenshot displays the SIMUnet website interface. On the left is a navigation sidebar with the PBSP logo and a network icon. The sidebar contains the following sections: **Methodology** (Methodology), **Tutorial** (SIMUnet tutorial), **Dataset selection** (Dataset selection), **Results** (SIMUnet), **Simunet analysis files** (simunet analysis), and **Bibliography** (Bibliography). The main content area is titled "SIMUnet" and includes a description: "SIMUnet is a machine learning framework to perform fits of parton distribution functions (PDFs) and SMEFT Wilson coefficients. It is an open-source, flexible methodology that allows the user to perform simultaneous PDF-SMEFT fits, and fixed-PDF SMEFT fits using a global dataset. It also provides analysis tools to evaluate different metrics of these fits, including quality metrics, uncertainties, PDF and SMEFT correlations, etc." Below the text is a section titled "PDF and SMEFT fits" containing two plots. The left plot, titled "g at 172.5 GeV", shows the "Ratio to NNPDF4.0-notop" on the y-axis (ranging from 0.75 to 1.05) against "x" on the x-axis (ranging from 0.1 to 0.7). It features three shaded regions representing uncertainty bands: SMEFT PDF (all top data) (68% c.l.+1 σ) in green, NNPDF4.0-notop (68% c.l.+1 σ) in orange, and fit H (SM, all top data) (68% c.l.+1 σ) in blue. The right plot, also titled "g at 172.5 GeV", shows the "r(Ratio to NNPDF4.0-notop)" on the y-axis (ranging from 0.00 to 0.16) against "x" on the x-axis (ranging from 0.1 to 0.7). It displays three curves: SMEFT PDF (all top data) (green), NNPDF4.0-notop (orange), and fit H (SM, all top data) (blue). A legend in the top right corner lists: Contents, PDF and SMEFT fits, Can PDFs absorb new physics?, The PBSP team, and Citation policy. At the bottom of the plots, the text "example figure :label: fig-sm_smeft_pdf" is visible.

Summary / Outlook

- SIMUnet:
 - Extension of NNPDF_{4.0} to include physics parameter dependence in partonic cross sections
 - NNPDF_{4.0} Monte Carlo replica uncertainty propagation
 - Fixed-PDF, EFT Fits
 - Contaminated Closure Tests
- SIMUnet Applications:
 - PDFs and SMEFT interplay in the Top sector
- Simunet Release:
 - Tool for simultaneous PDF SMEFT fits (Linear in SMEFT Wilson Coefficients)
 - Global Dataset: EW, Higgs and Top sector
 - Open source



Backup



Pitfalls of the Monte-Carlo replica method

Kassabov et al, arXiv: 2303.06159

Assume that the theory prediction is quadratic in the SMEFT Wilson coefficient c

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

Suppose that experimental datapoint d is distributed as $d \sim \mathcal{N}(t(c), \sigma^2)$

Bayesian Credible Intervals can be computed from posterior

$$\mathbb{P}(c | d) \propto \exp\left(-\frac{1}{2\sigma^2}(d - t(c))^2\right)$$

Monte Carlo Confidence Intervals can be computed from the knowledge of how d is distributed

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