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

Mark N. Costantini

IRN TERASCALE 2023 @ Marseille

PBSP

UNIVERSITY OF CAMBRIDGE

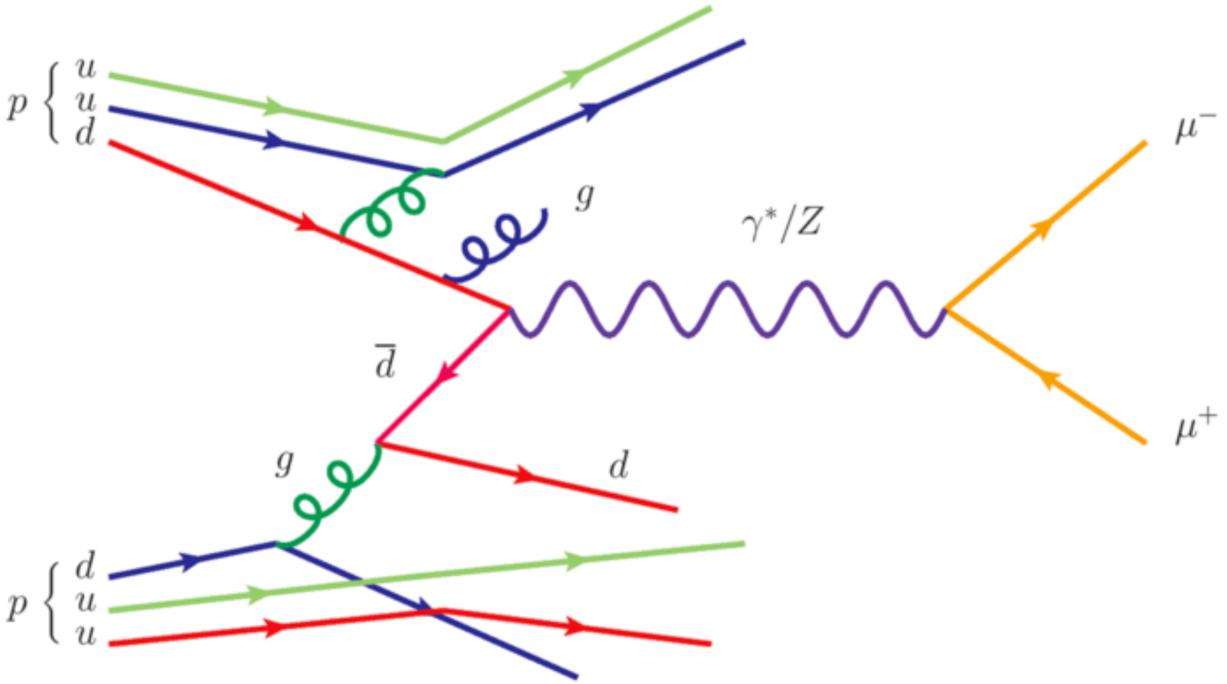
erc European Research Council

Outline

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

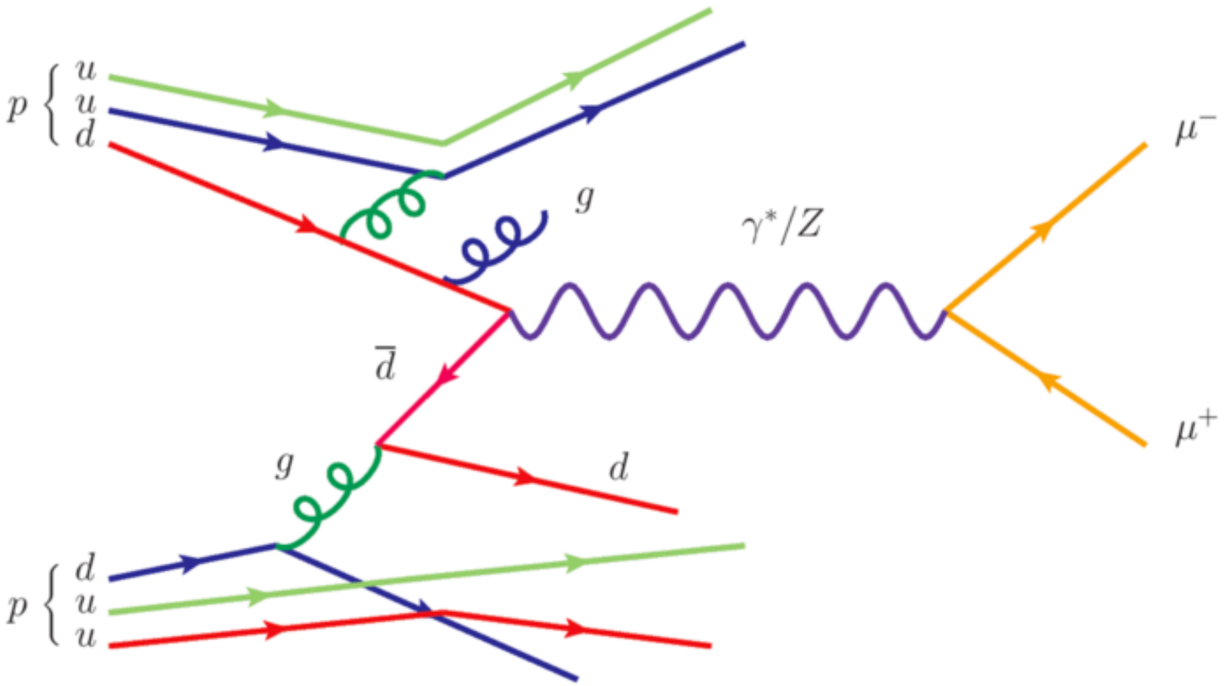
Introduction

Fitting PDFs and Physical Parameters



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

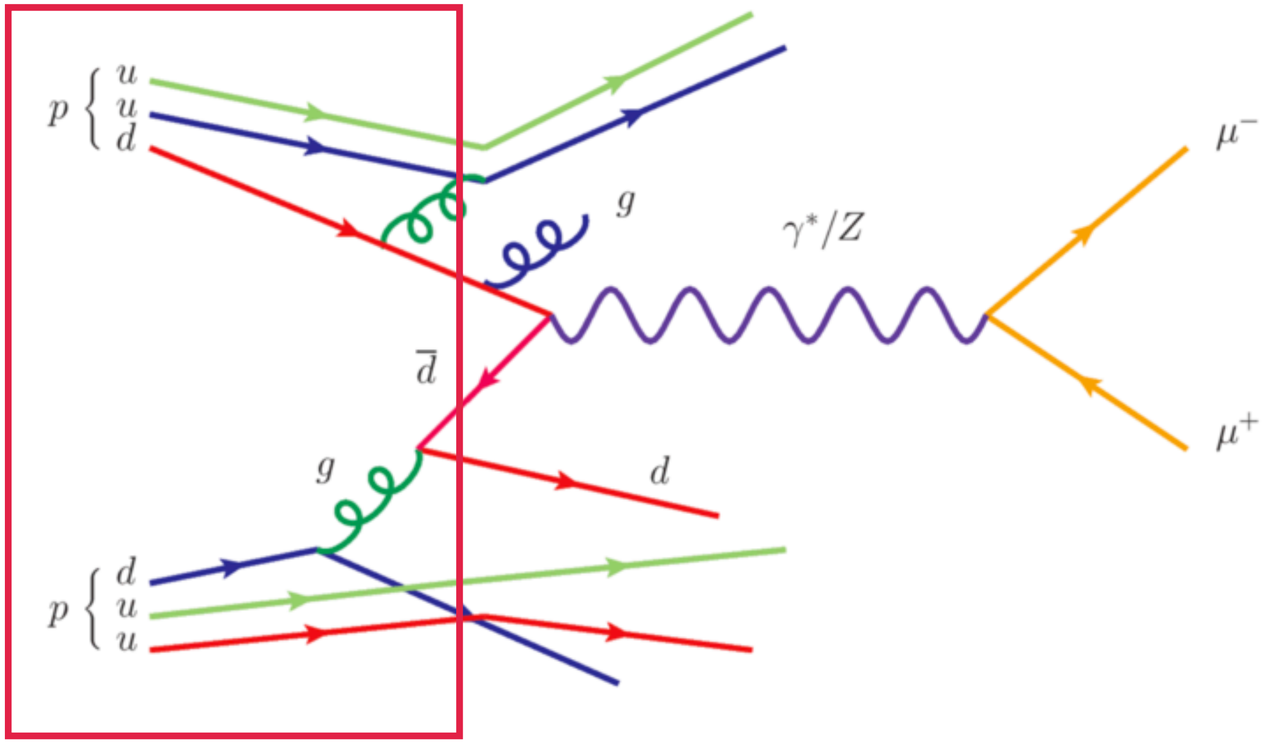
Fitting PDFs and Physical Parameters



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

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

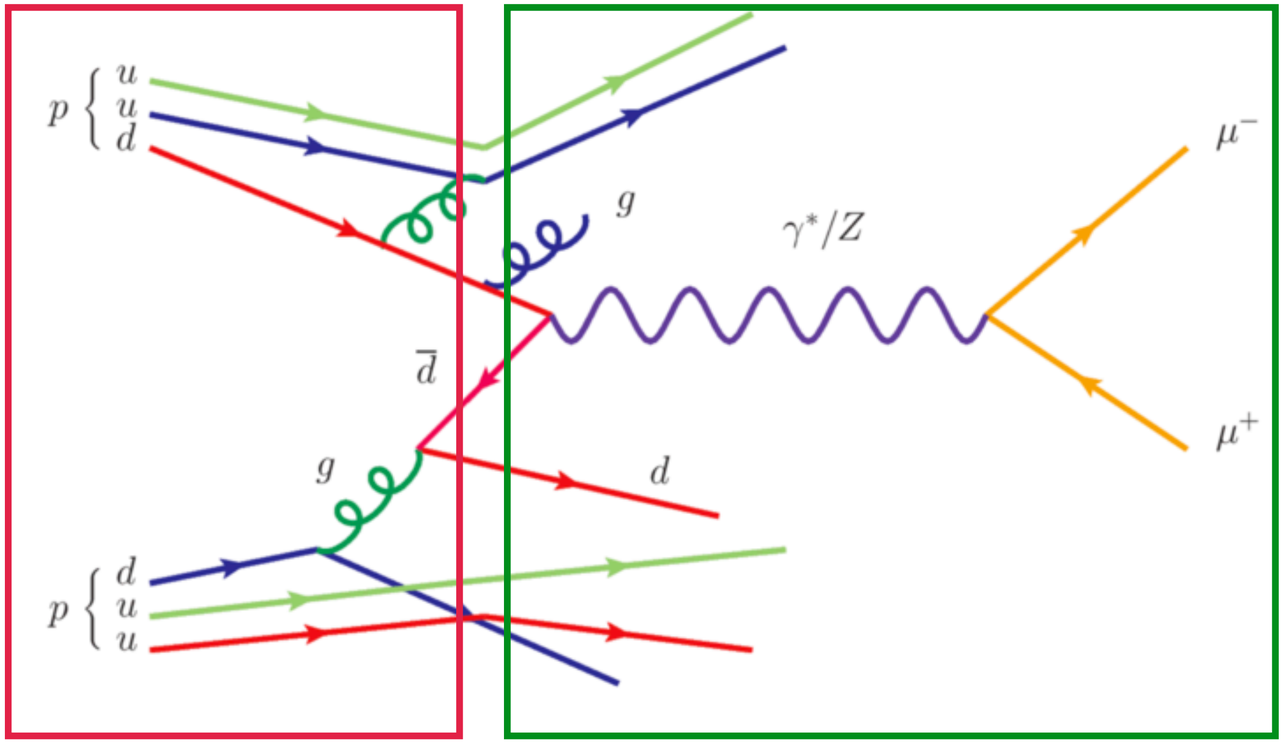
Fitting PDFs and Physical Parameters



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

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

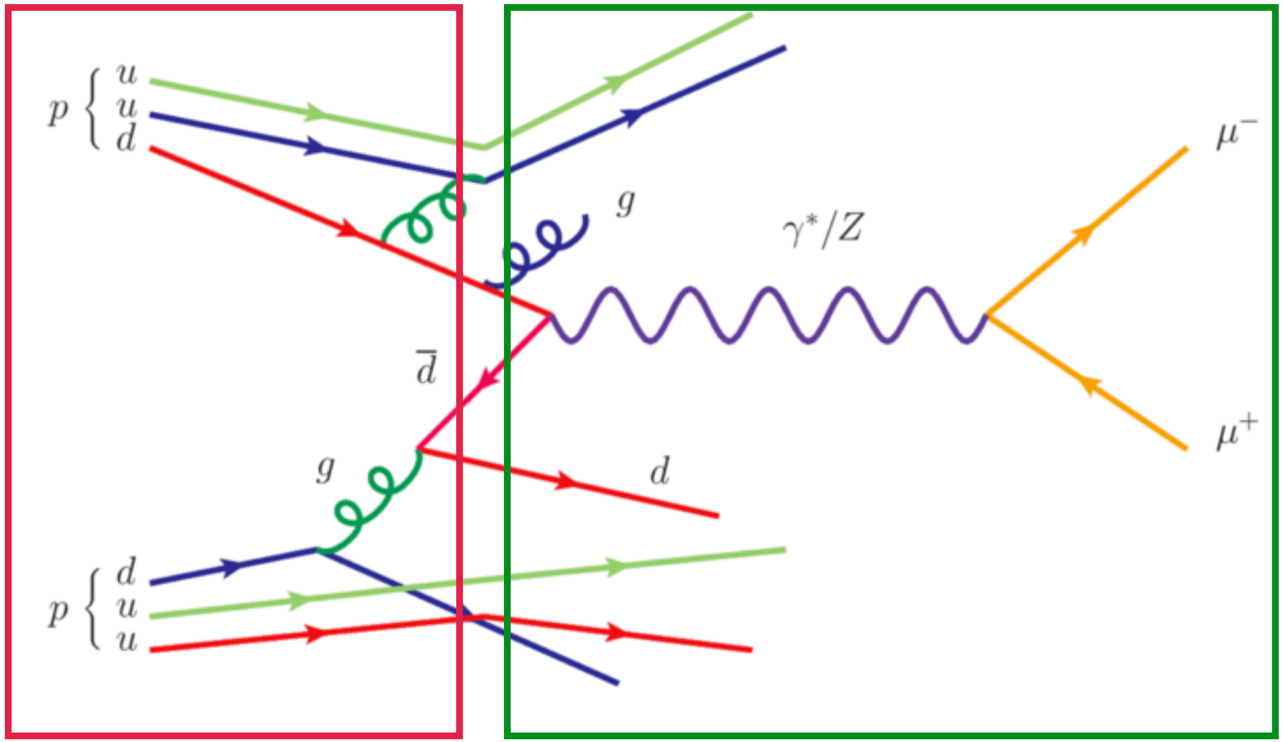
Fitting PDFs and Physical Parameters



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

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

Fitting PDFs and Physical Parameters



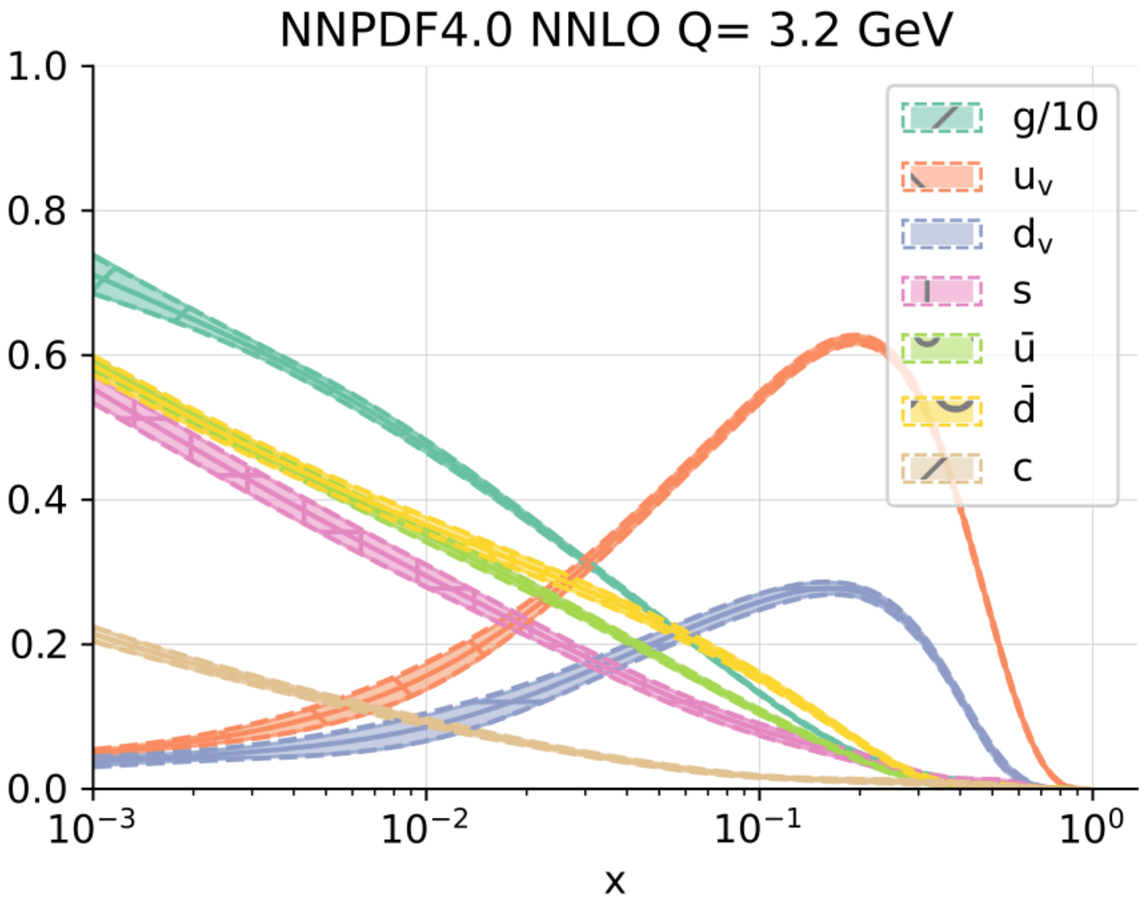
- **Theory predictions** for collider experiments are obtained from the standard **factorisation formula**

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

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

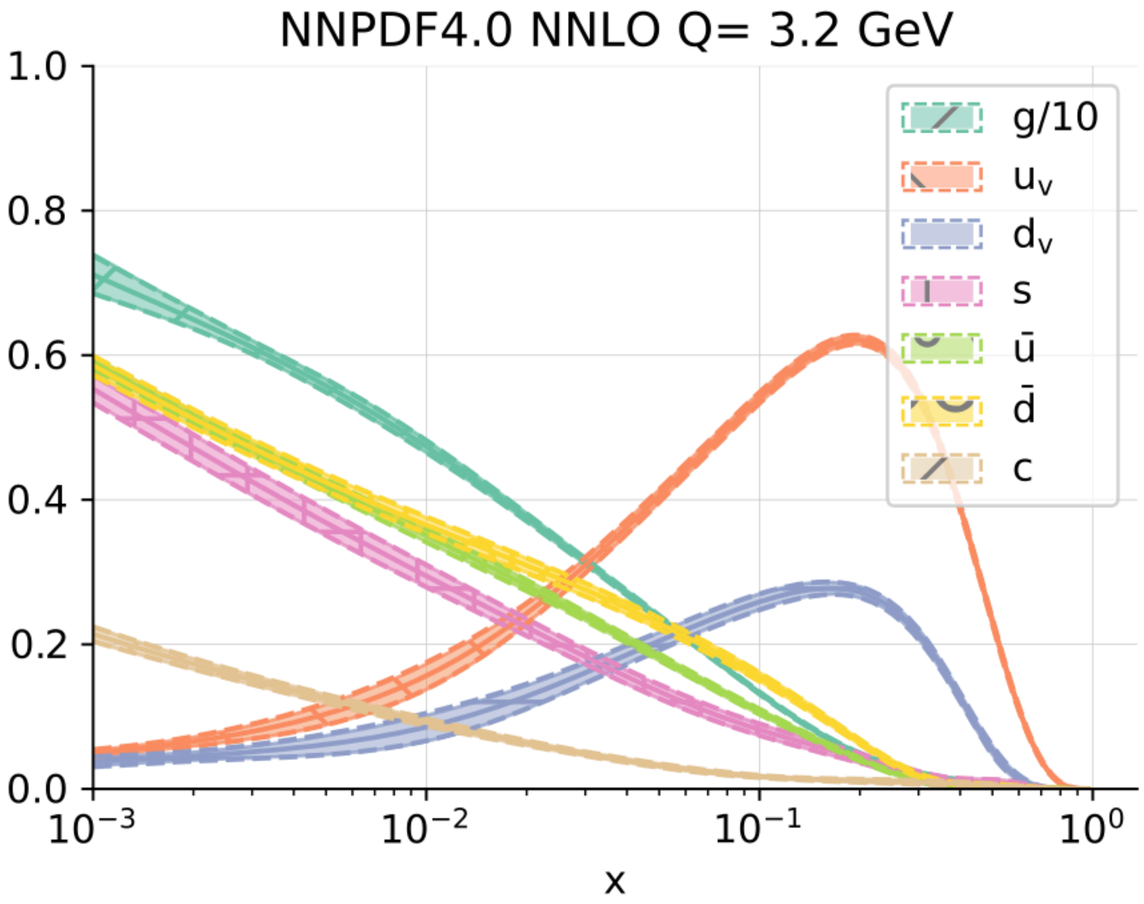
PDF Parameter Fits



- Global fits are used for the extraction of PDFs

NNPDF40, Ball et al, arXiv: 2109.02653

PDF Parameter Fits

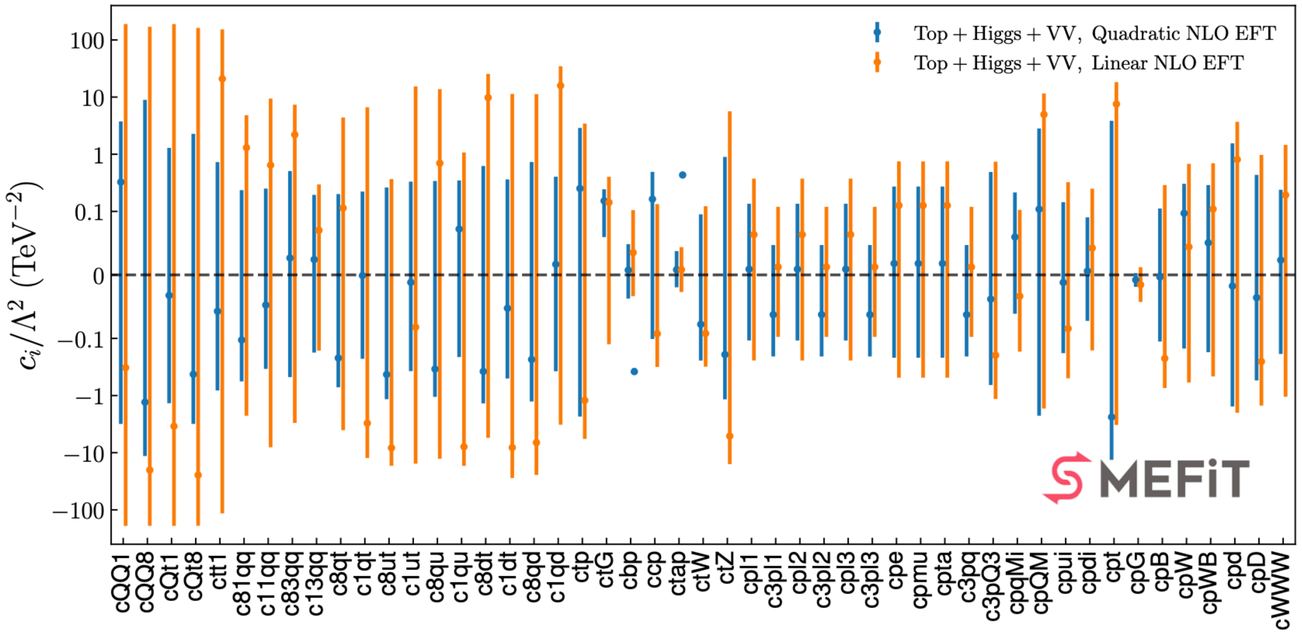


NNPDF40, Ball et al, arXiv: 2109.02653

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

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

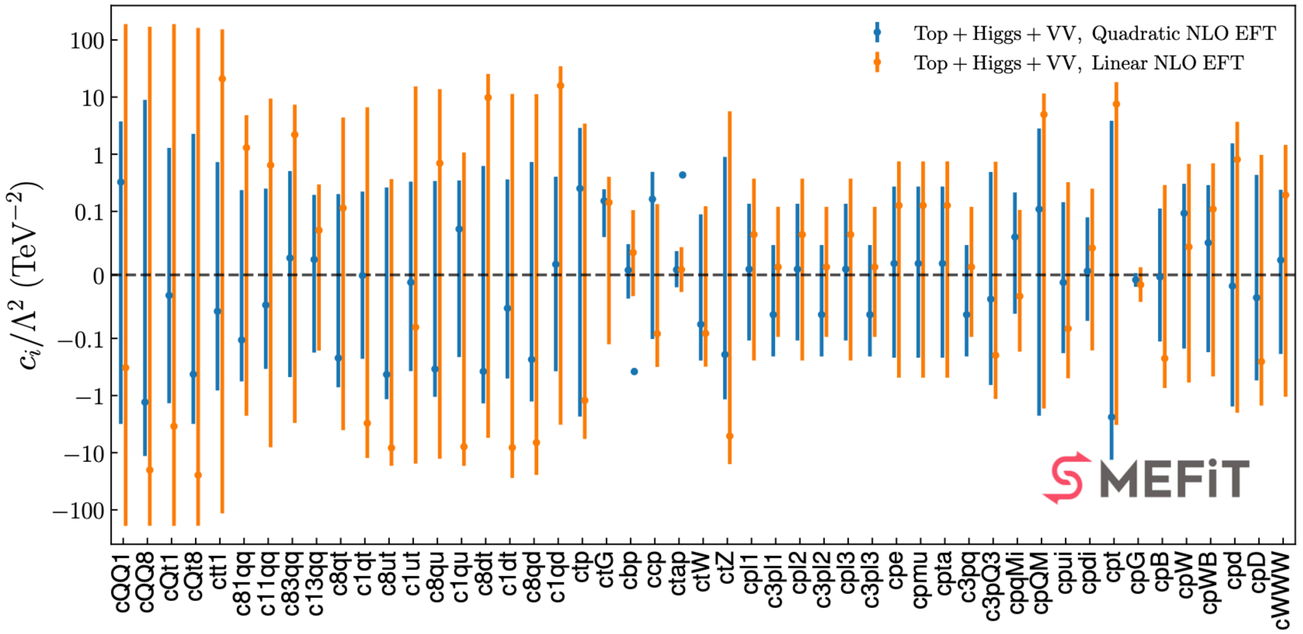
SMEFT Fits



SMEFiT, Ethier et al, arXiv: 2105.00006

- Global fits are used to fit Wilson Coefficients

SMEFT Fits



SMEFiT, Ethier et al, arXiv: 2105.00006

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

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

Simultaneous PDF and SMEFT Fits

- Fitting PDF and SMEFT parameters separately can lead to inconsistencies

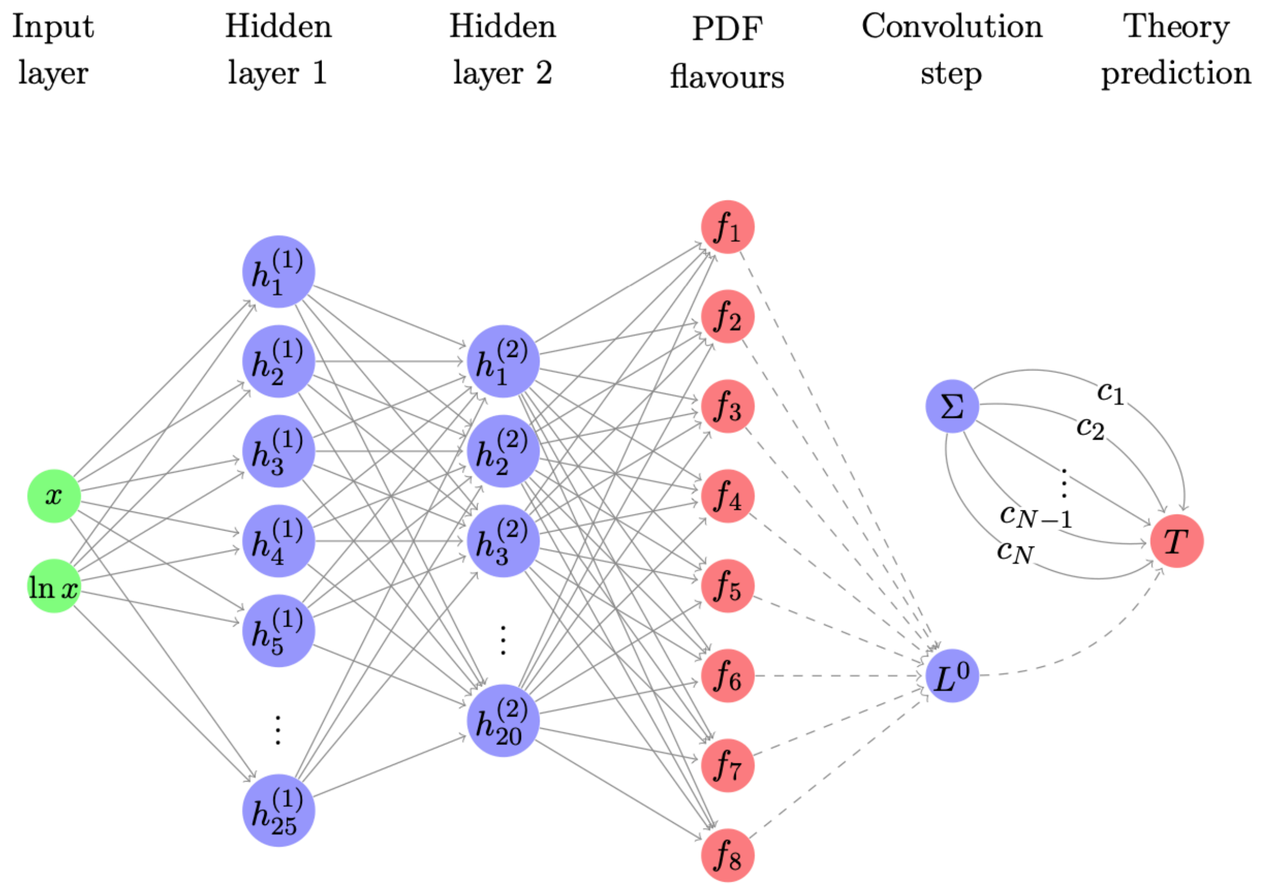
→ Fitted PDFs can depend implicitly on fixed physical parameters used in the fit.

→ 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

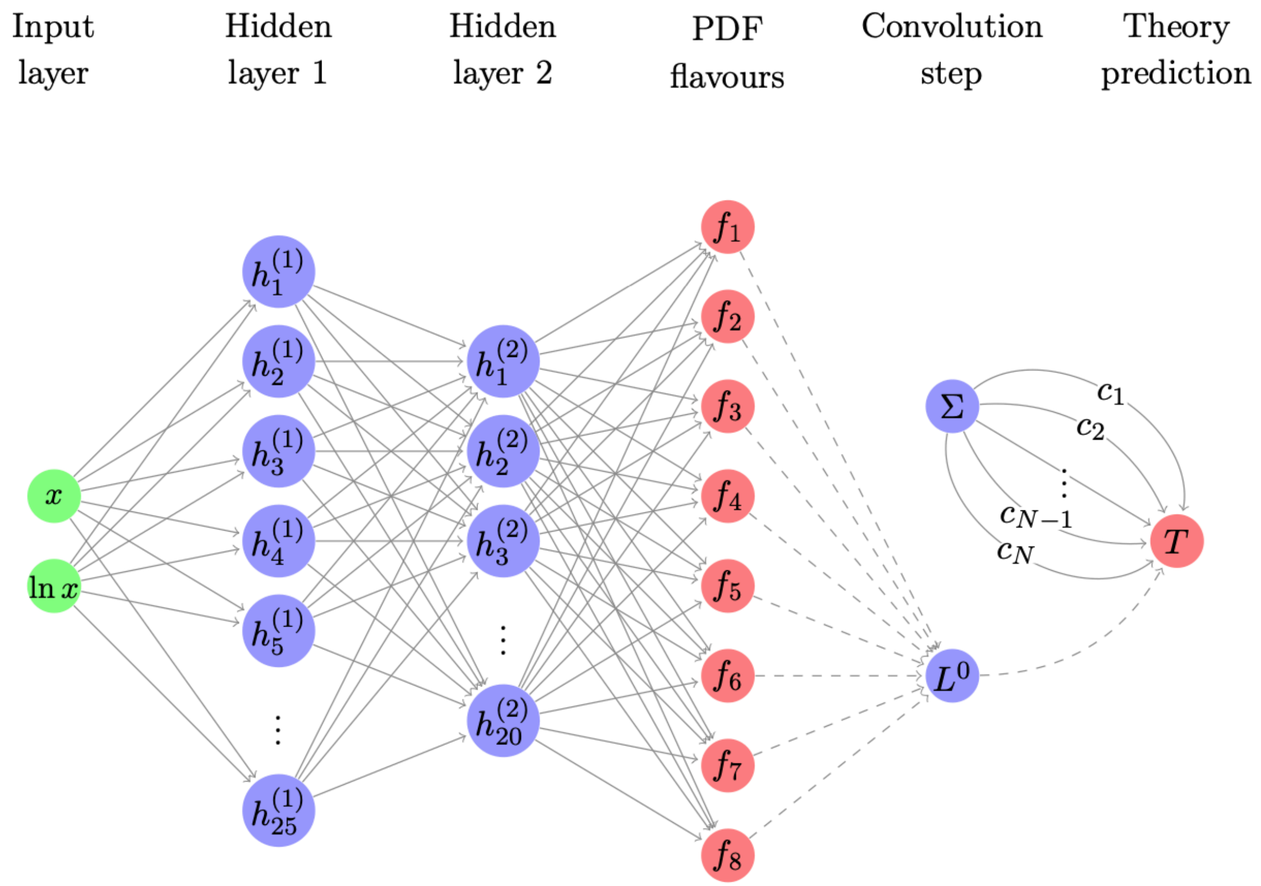


- Simultaneous fit of PDF and theory parameters (e.g. EFT WC)

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

S. Iranipour and M. Ubiali, arXiv: 2201.07240

SIMUnet: Methodology



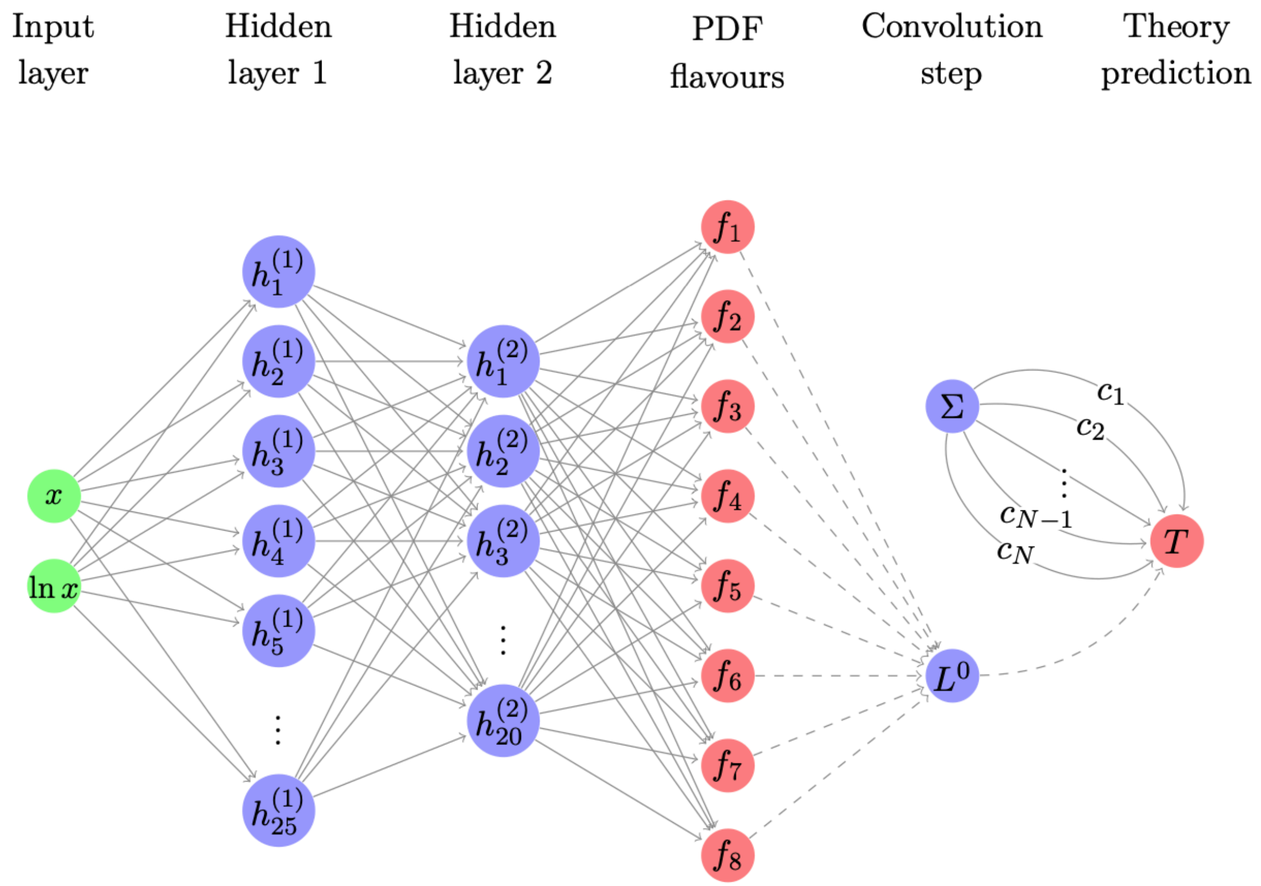
- Simultaneous fit of PDF and theory parameters (e.g. EFT WC)

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

- Extension of NNPDF4.0 framework to account for theory parameters dependence of partonic cross sections

S. Iranipour and M. Ubiali, arXiv: 2201.07240

SIMUnet: Methodology



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

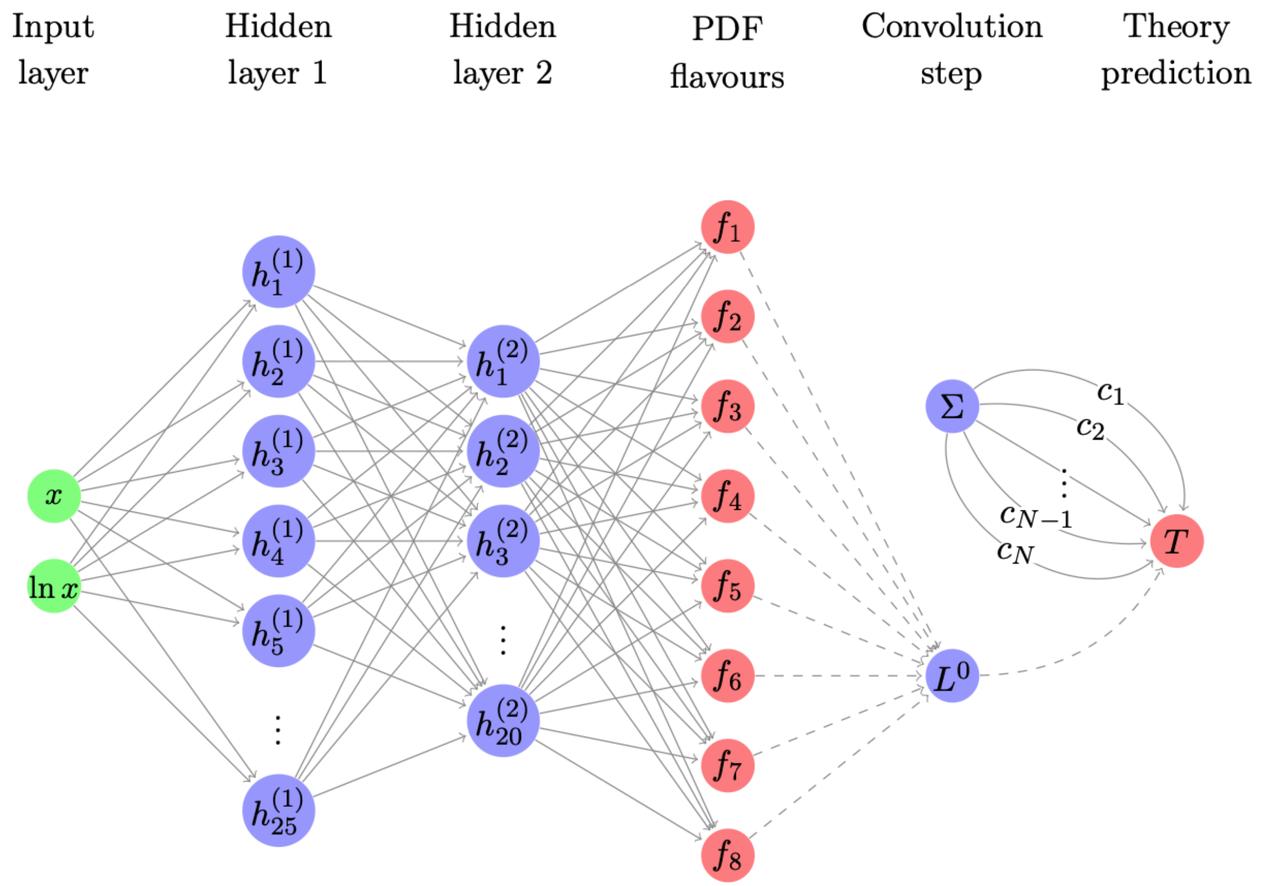
$$T(\{\theta\},\{c\}) = \text{PDFs}(\{\theta\}) \otimes \hat{\sigma}(\{c\})$$

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

→ Extra combination layer to the NNPDF40 Neural Network

S. Iranipour and M. Ubiali, arXiv: 2201.07240

SIMUnet: Methodology



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

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

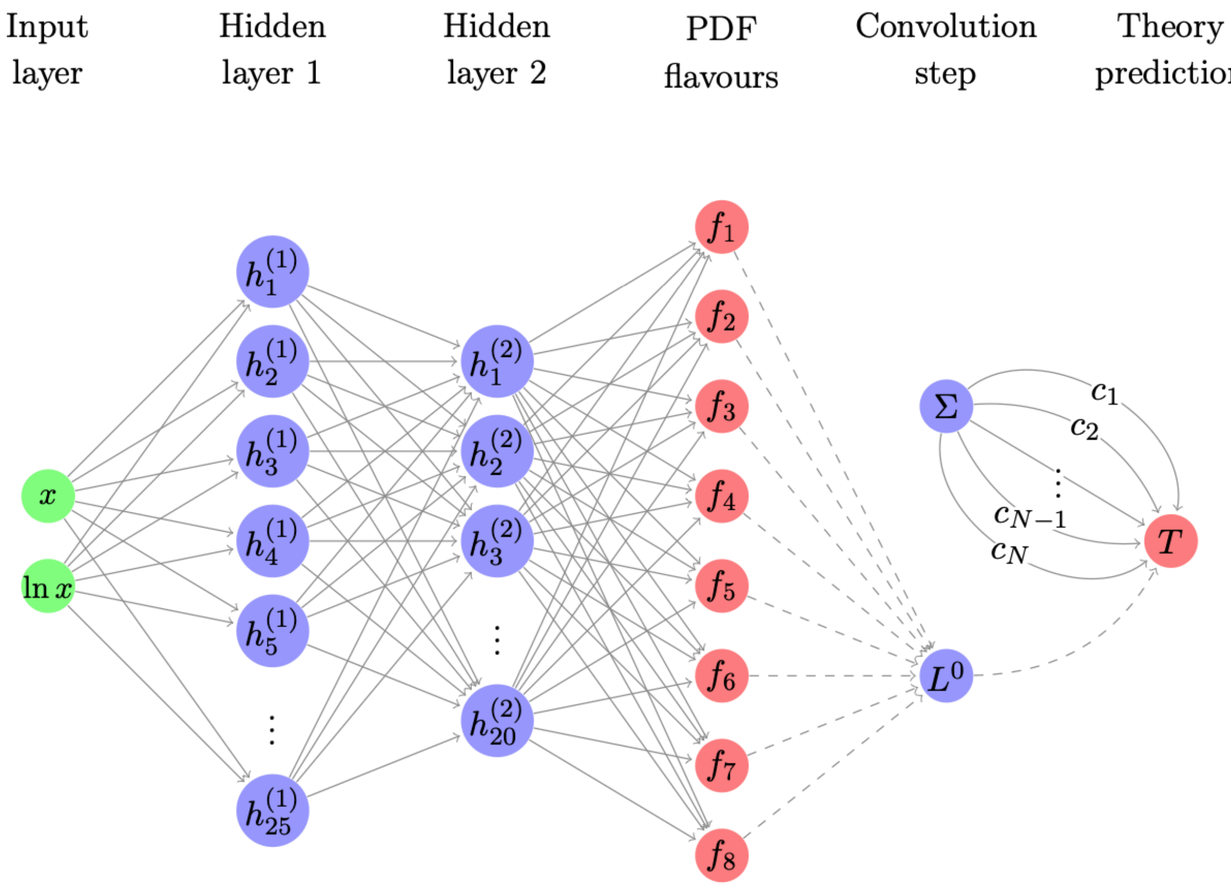
- Extension of NNPDF4.0 framework to account for theory parameters dependence of partonic cross sections

→ Extra combination layer to the NNPDF4.0 Neural Network

→ SMEFT corrections added in the form of K-factors

S. Iranipour and M. Ubiali, arXiv: 2201.07240

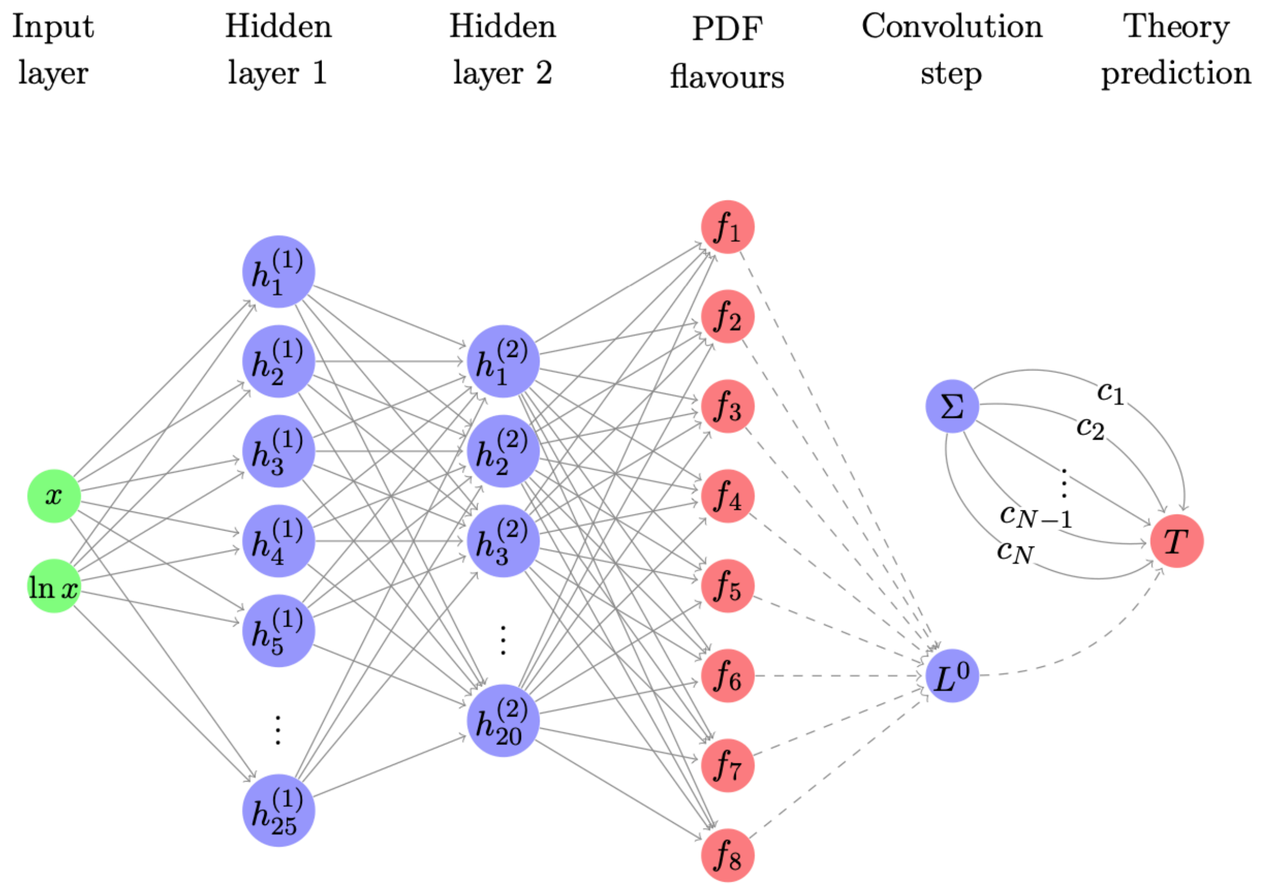
SIMUnet: Methodology



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

S. Iranipour and M. Ubiali, arXiv: 2201.07240

SIMUnet: Methodology



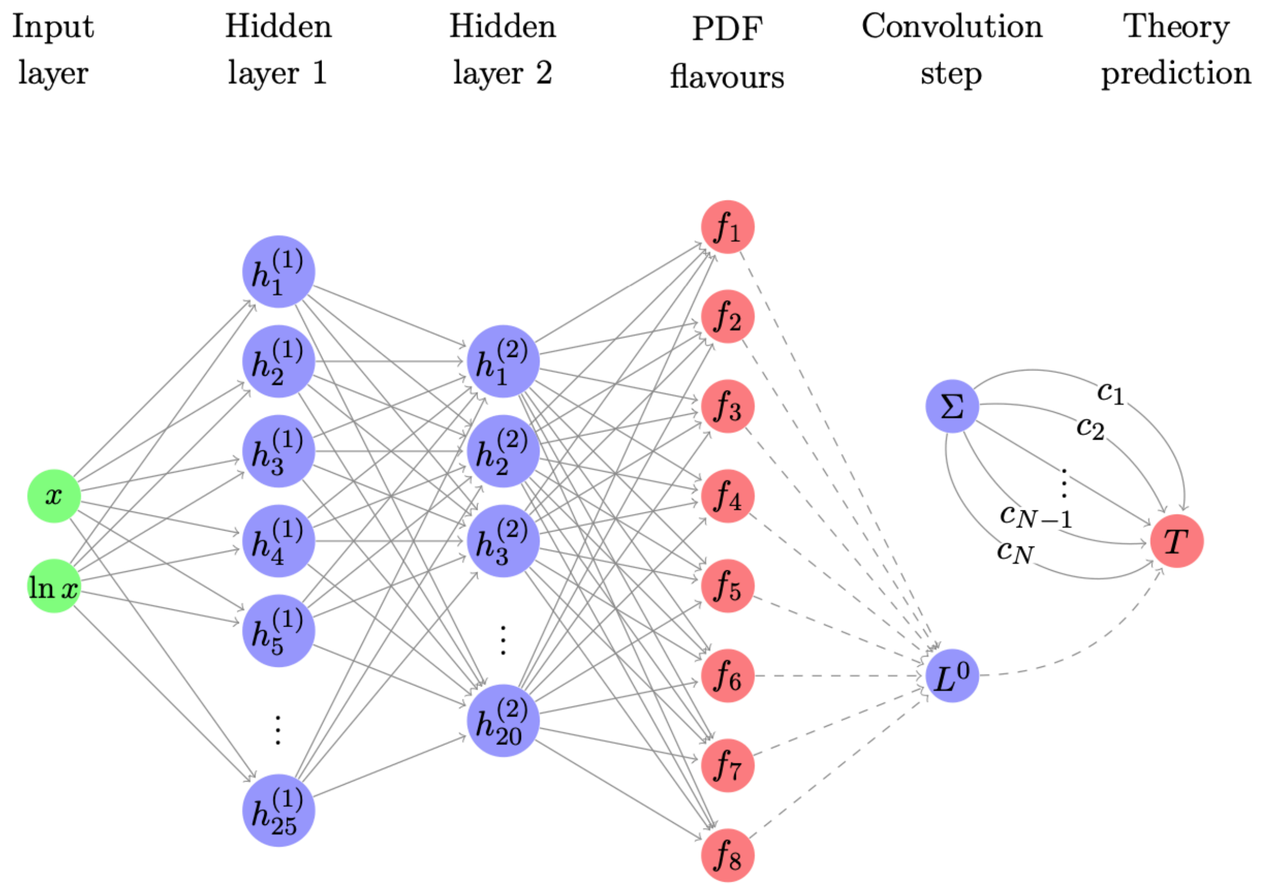
- **Uncertainty Quantification:** NNPDF Monte Carlo replica method.

→ Generate N_{rep} data replica

$$d_i \sim \mathcal{N}(d, C_{\text{exp}}), \quad i \in 1, \ldots, N_{\text{rep}}$$

S. Iranipour and M. Ubiali, arXiv: 2201.07240

SIMUnet: Methodology



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

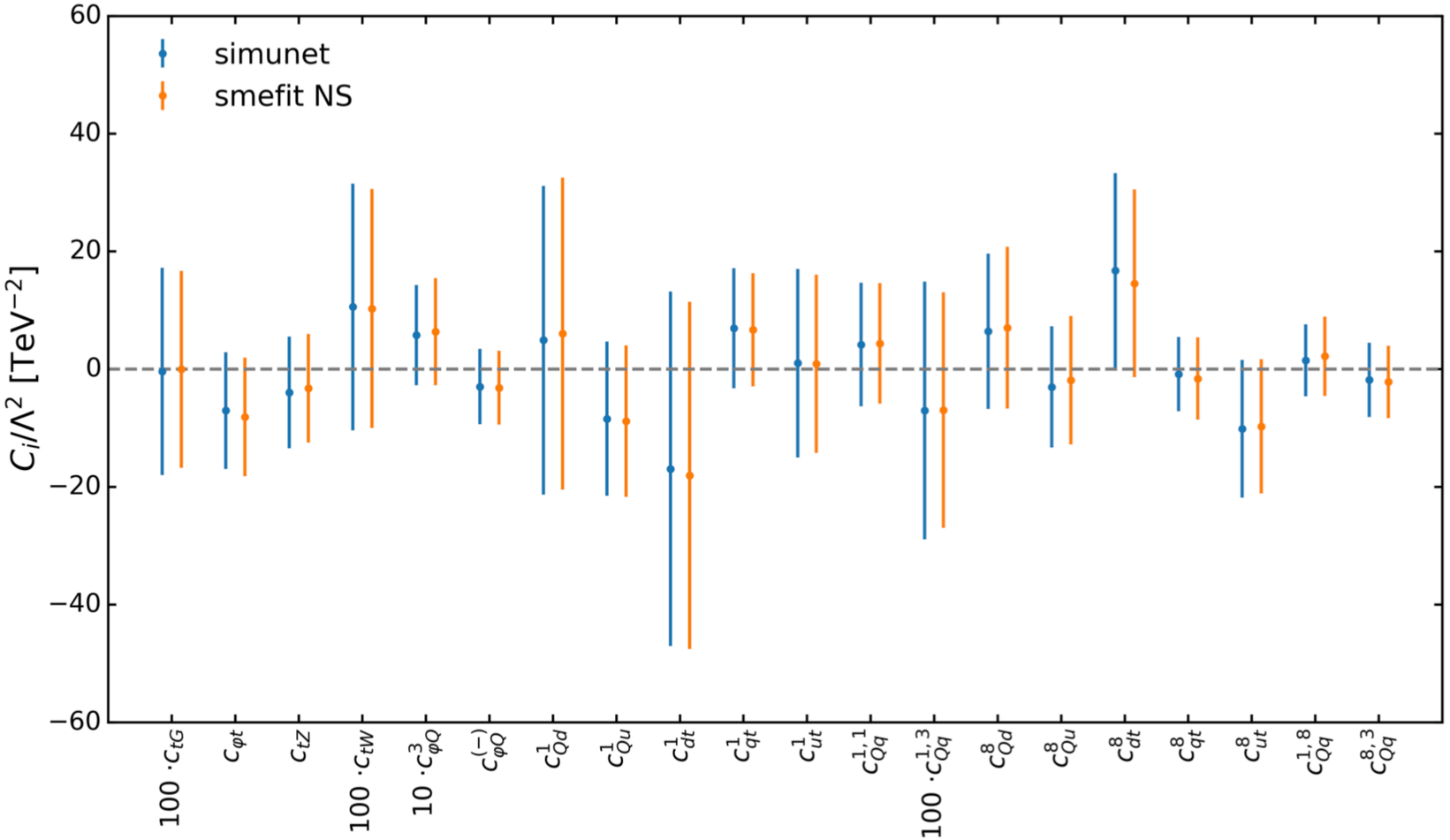
→ Generate N_{rep} data replica

$$d_i \sim \mathcal{N}(d, C_{\text{exp}}), \quad i \in 1, \ldots, N_{\text{rep}}$$

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

S. Iranipour and M. Ubiali, arXiv: 2201.07240

SIMUnet: Fixed-PDF EFT fits



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

Kassabov et al, arXiv: 2303.06159

SIMUnet: Inject New Physics in the data

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

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

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

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

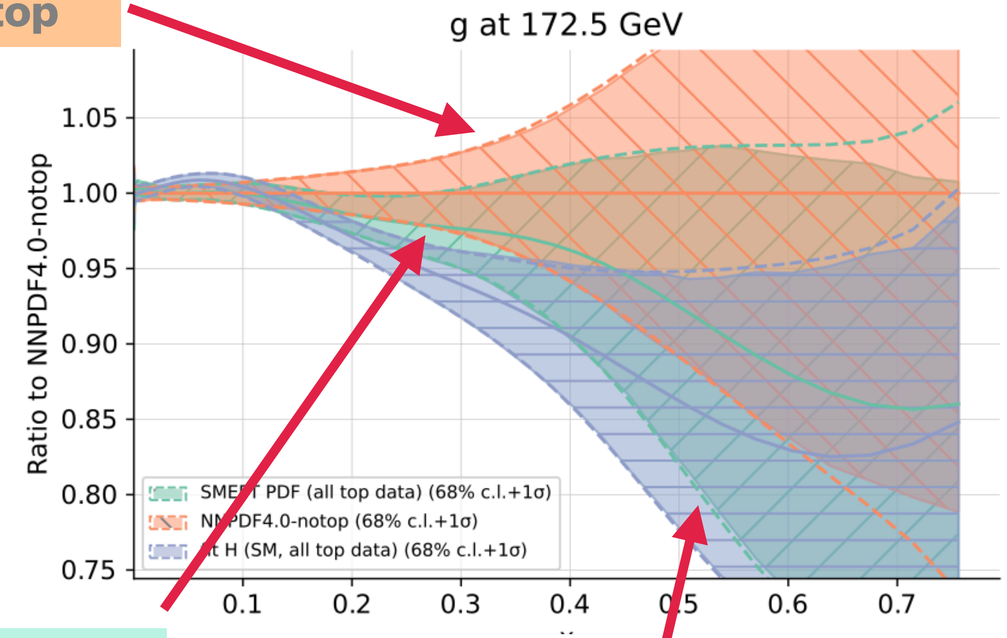
- Check effect of fitting assuming SM: can PDFs absorb new physics?

→ arXiv: 2307.10370

SIMUnet:Applications

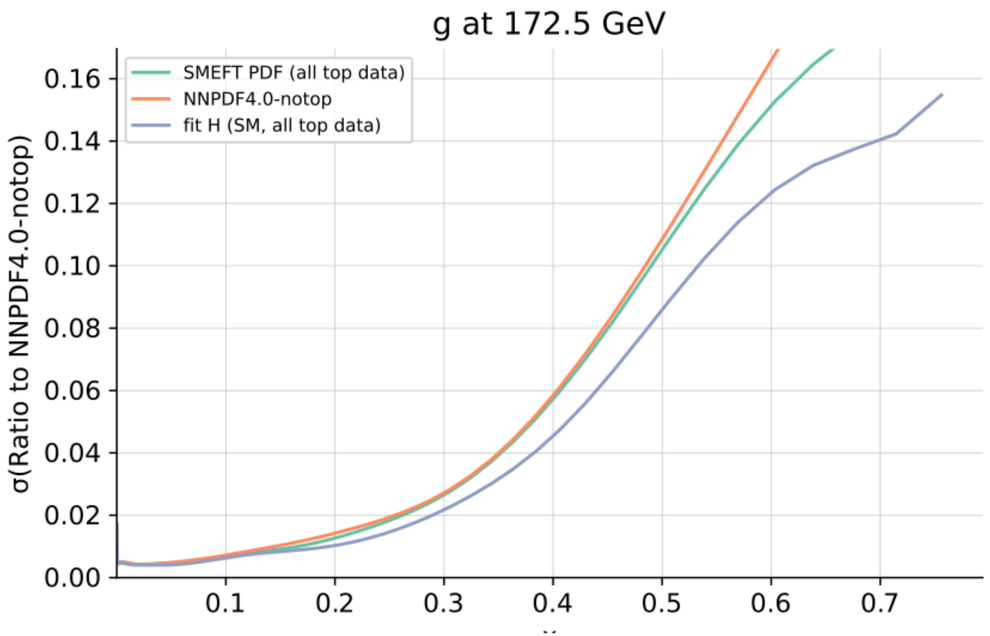
Simultaneous PDF SMEFT Fits: PDFs

NNPDF40, no-top



SMEFT PDF

NNPDF40, all top data



- Marked difference between SM gluon PDF and SMEFT gluon PDF

→ Larger PDF uncertainties

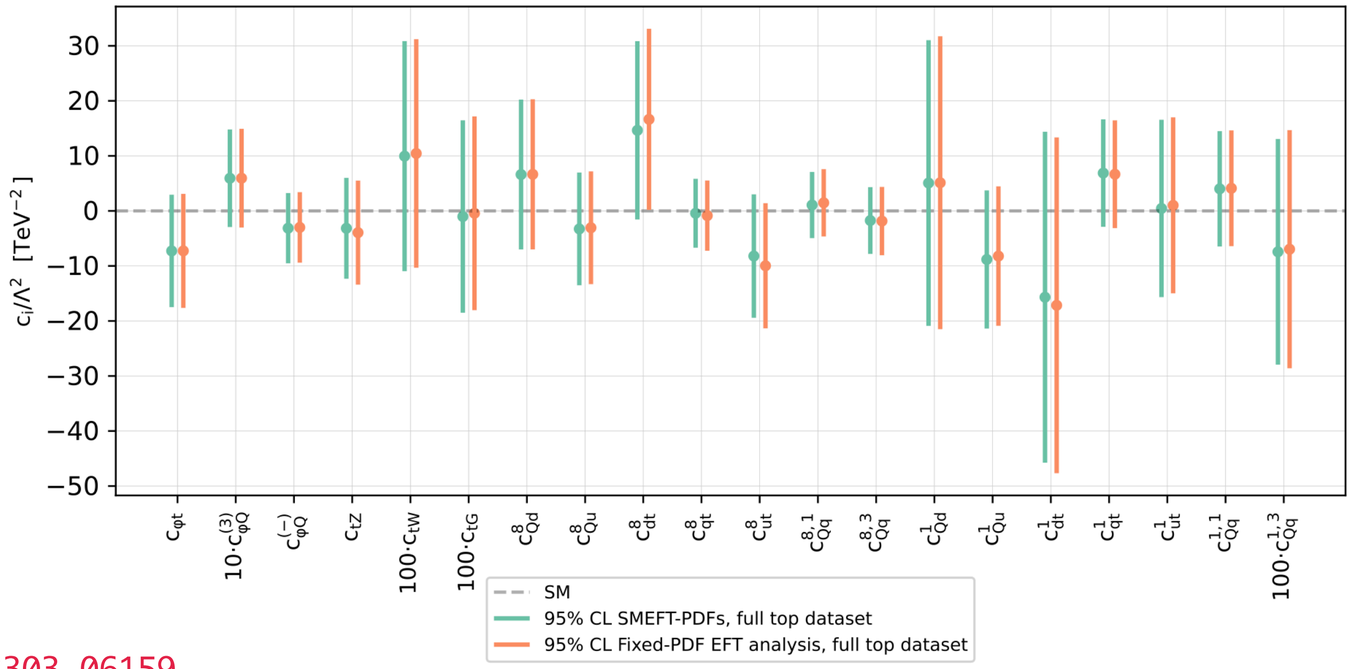
Kassabov et al, arXiv: 2303.06159

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



Kassabov et al, arXiv: 2303.06159

PDF-SMEFT interplay questions

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

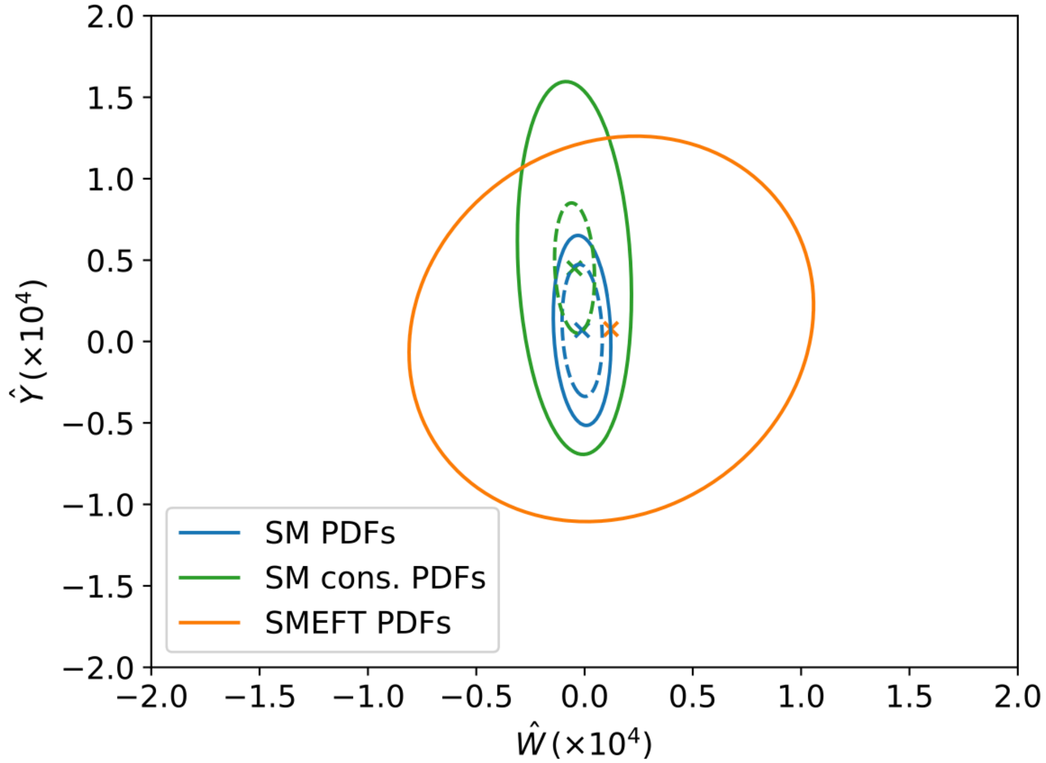
→ It was shown in Carrazza et al., 1905.05215, that interplay is very mild in the case of simultaneous extractions of four-fermion operators and PDFs using DIS-only data.

→ 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

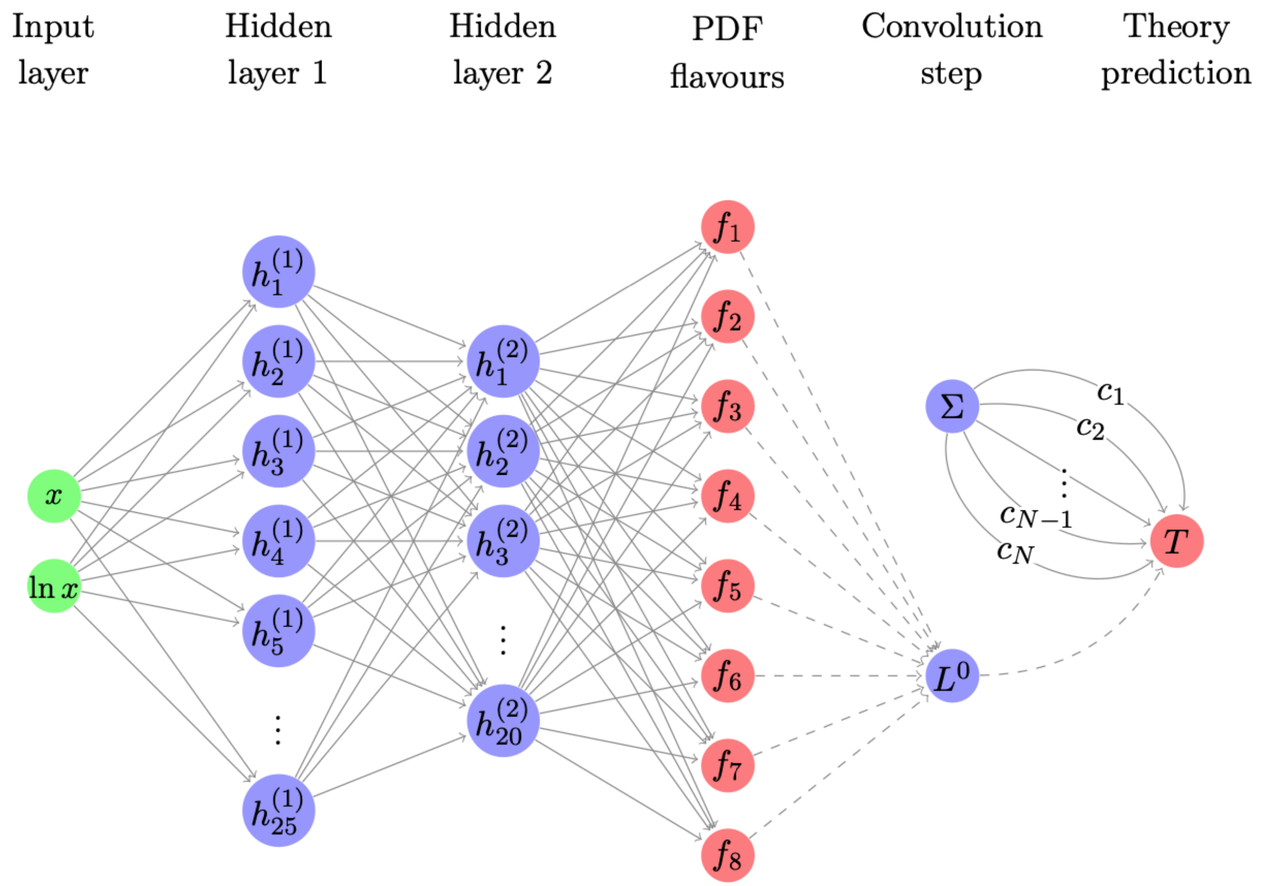
Question: won't the PDF-SMEFT interplay be negligible?

→ 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



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

→ **Linear in the SMEFT**

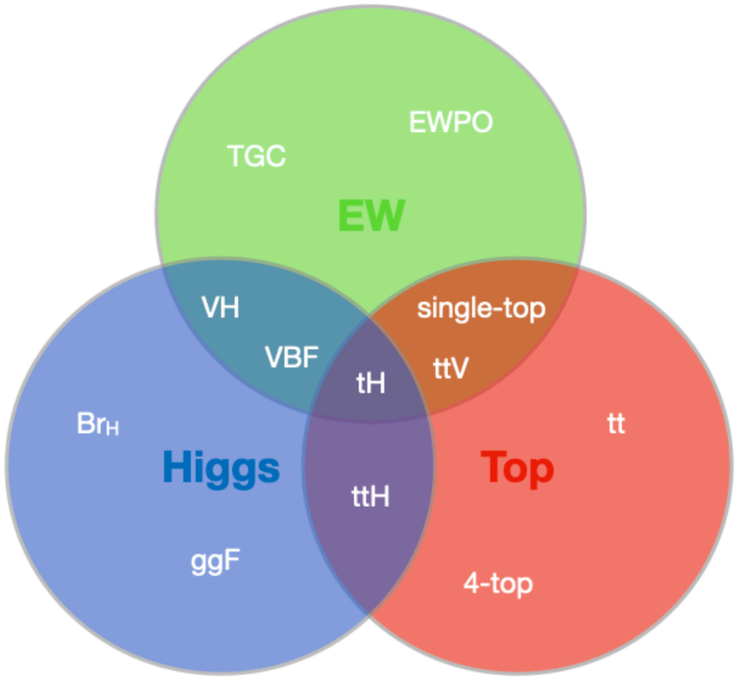
S. Iranipour and M. Ubiali, arXiv: 2201.07240

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



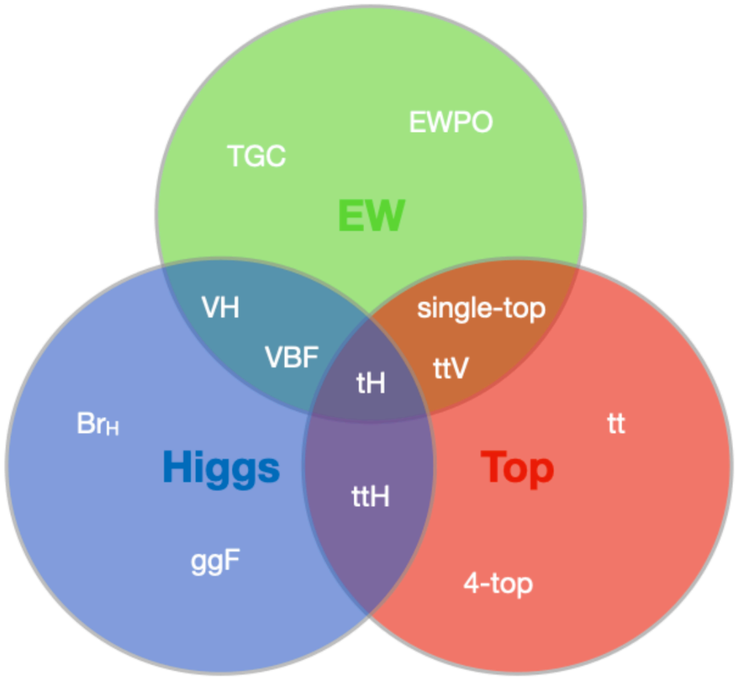
© Ken Mimasu

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



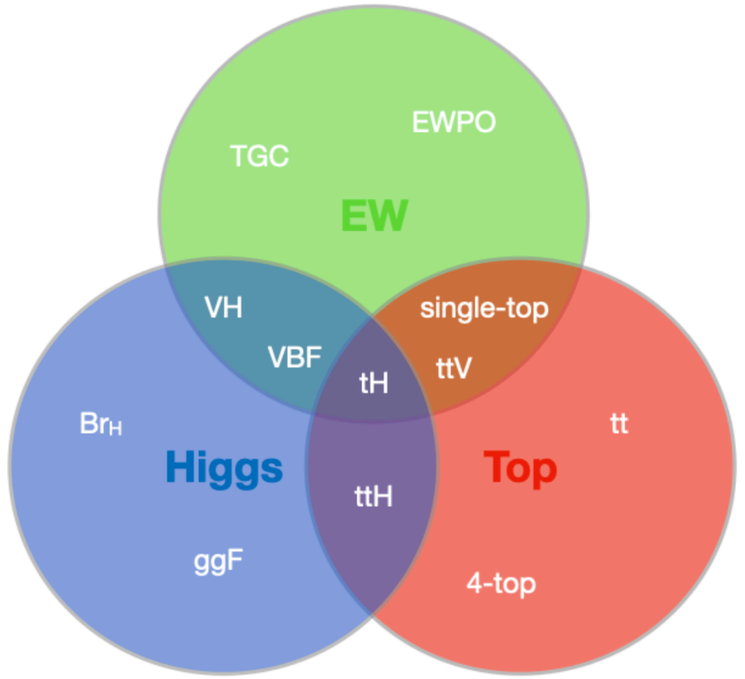
© Ken Mimasu

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

```yaml
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: 'OtG', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'Opt', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'OtZ', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'OtW', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'O3pQ3', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}

...
```

- **NNPDF4.0 like interface**

SIMUnet Release: Open Source

```yaml
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: 'OtG', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'Opt', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'OtZ', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'OtW', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'O3pQ3', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}

...
```

- **NNPDF4.0 like interface**

→ Similar runcards

SIMUnet Release: Open Source

```yaml
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: 'OtG', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'Opt', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'OtZ', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'OtW', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}
- {name: 'O3pQ3', scale: 0.01, initialisation: {type: uniform, minval: -1, maxval: 1}}

...
```

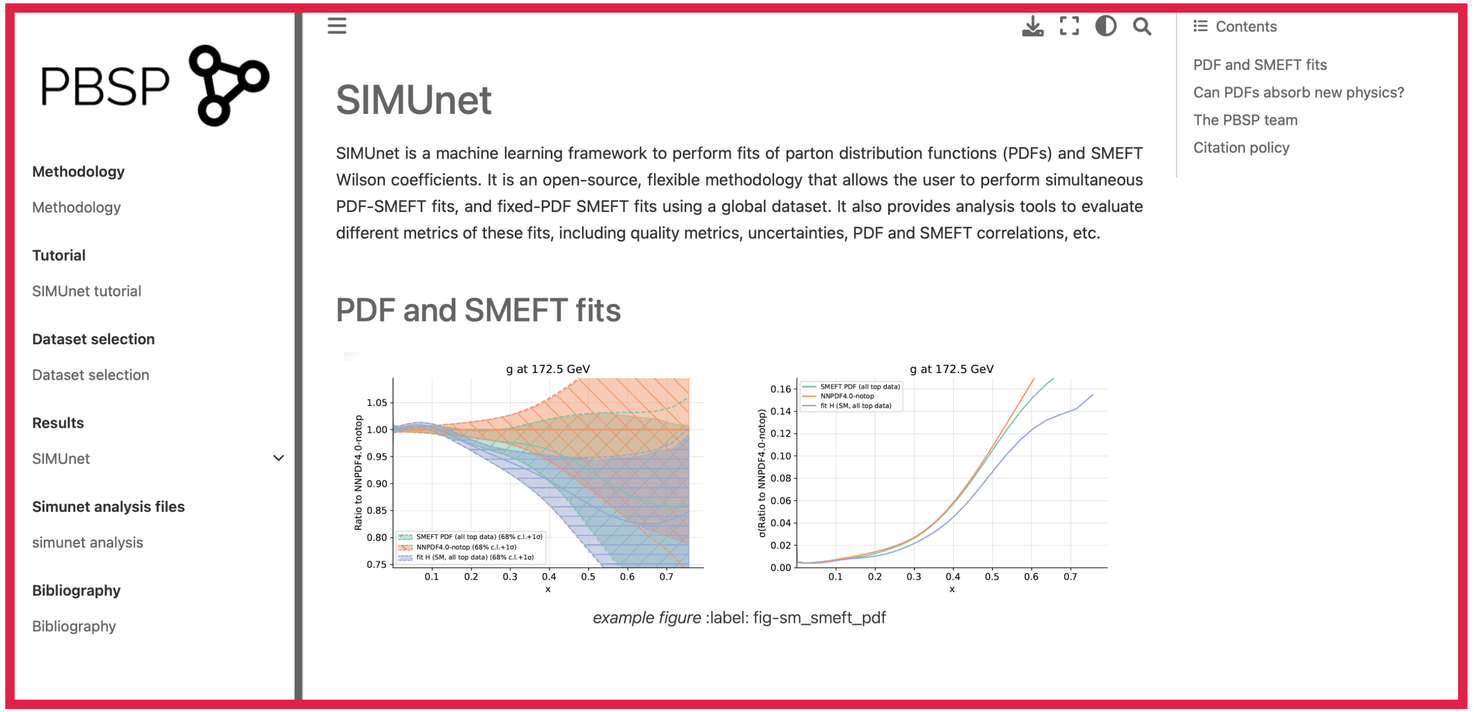
- **NNPDF4.0 like interface**

→ Similar runcards

→ Analysis tools e.g. automated reports

SIMUnet Release: Open Source

- SIMUnet website with theory, tutorials and analysis files



Summary / Outlook

- SIMUnet:
  - Extension of NNPDF40 to include physics parameter dependence in partonic cross sections
  - NNPDF40 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

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

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

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