Differential Measurements for Global SMEFT Analyses

Juan Rojo
VU Amsterdam & Theory group, Nikhef

CMS Physics Days:
Effective Field Theory program
Zoom, 9th May 2023

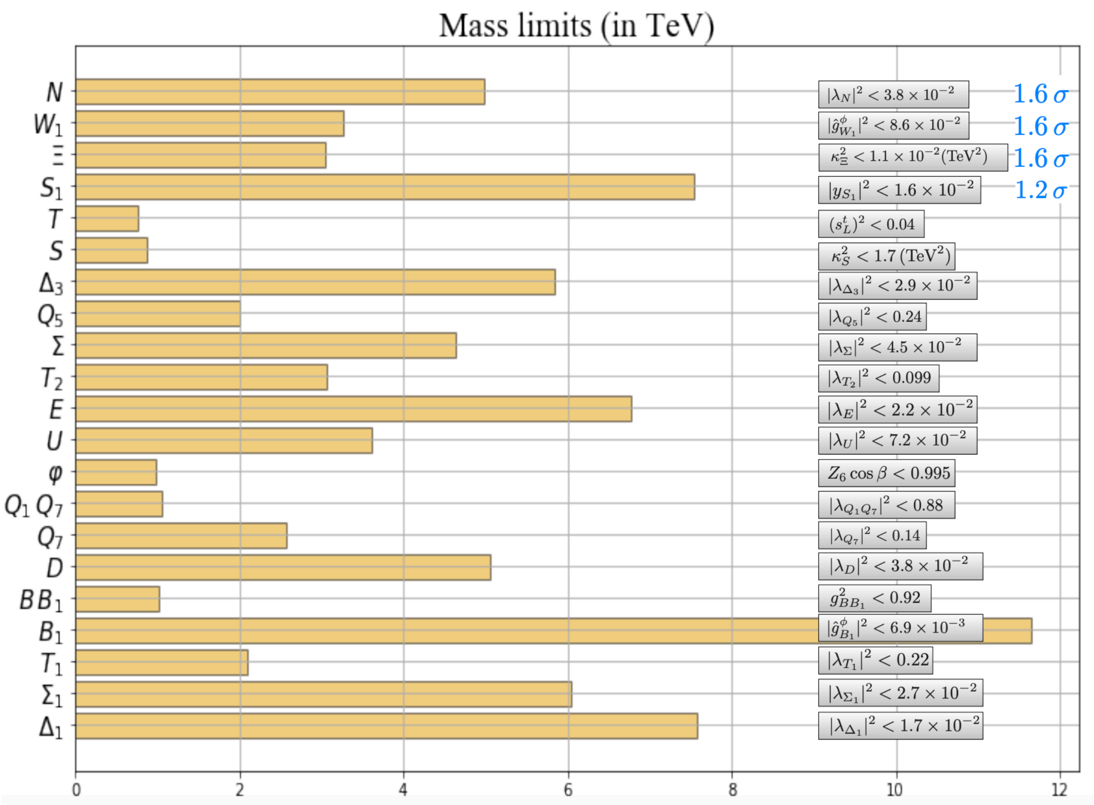
Why the SMEFT?

$$\mathcal{L}_{\text{SMEFT}}(\{c_i\},\Lambda) = \mathcal{L}_{\text{SM}} + \sum_{d=5}^{\infty}\sum_{i=1}^{N_d} c_i^{(d)} \frac{\mathcal{O}_i^{(d)}}{\Lambda^{d-4}}$$

- **Low-energy limit** of generic UV-complete theories with linearly realised EWSB
- **Complete basis** at any given mass-dimension: systematic parametrisation of BSM effects
- **Fully renormalizable**, full-fledged QFT: compute higher orders in QCD and EW

Exploit the full power of **SM measurements** for model-independent BSM searches: constrain large classes of BSM scenarios matched to the SMEFT

tree-level, single-field extensions of the SM
FitMaker, Ellis et al 2020



Why the SMEFT?

☑ EFT operators may induce **growth with the partonic centre-of-mass energy**: increased sensitivity in LHC cross-sections in the TeV region

$$\sigma(E) = \sigma_{\rm SM} \times (E) \left(1 + \sum_i^{N_{d6}} \omega_i \frac{c_i v^2}{\Lambda^2} + \sum_i^{N_{d6}} \widetilde{\omega}_i \frac{c_i E^2}{\Lambda^2} + \mathcal{O}\left(\Lambda^{-4}\right) \right)$$

from inclusive cross-sections

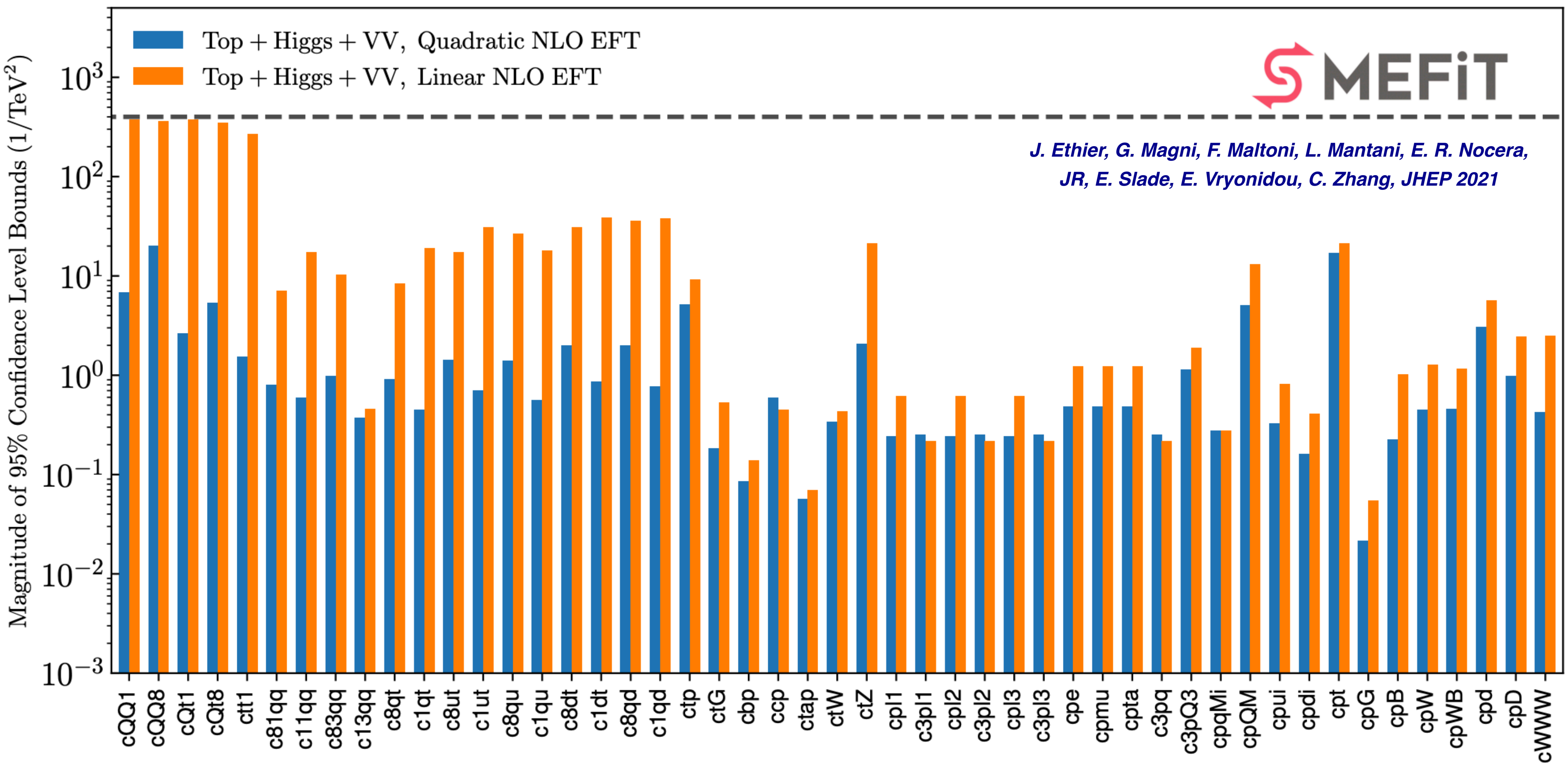
from high-energy differential measurements

e.g. an LHC measurement of a **high-energy tail at 4 TeV with 50% precision** may have comparable impact on the SMEFT coefficients as a **LEP measurement with 0.1% precision**

☑ Differential measurements also provide handles to **disentangle contributions from different EFT operators** since kinematics are different in each bin

flat directions when considering inclusive measurements may go away if the measurement is made differential

State-of-the-art



Global EFT fits include data on **top quark, Higgs, and gauge boson production**, both inclusive and differential measurements & the constraints from LEP EWPOs

some groups also include LHCb flavour data, in connection with flavour anomalies, but fully ``global LHC EFT fit'' still missing

Global SMEFT fits are ``precision physics'': NNLO QCD + NLO EW for SM, NLO QCD for SMEFT, linear and quadratic EFT corrections, …

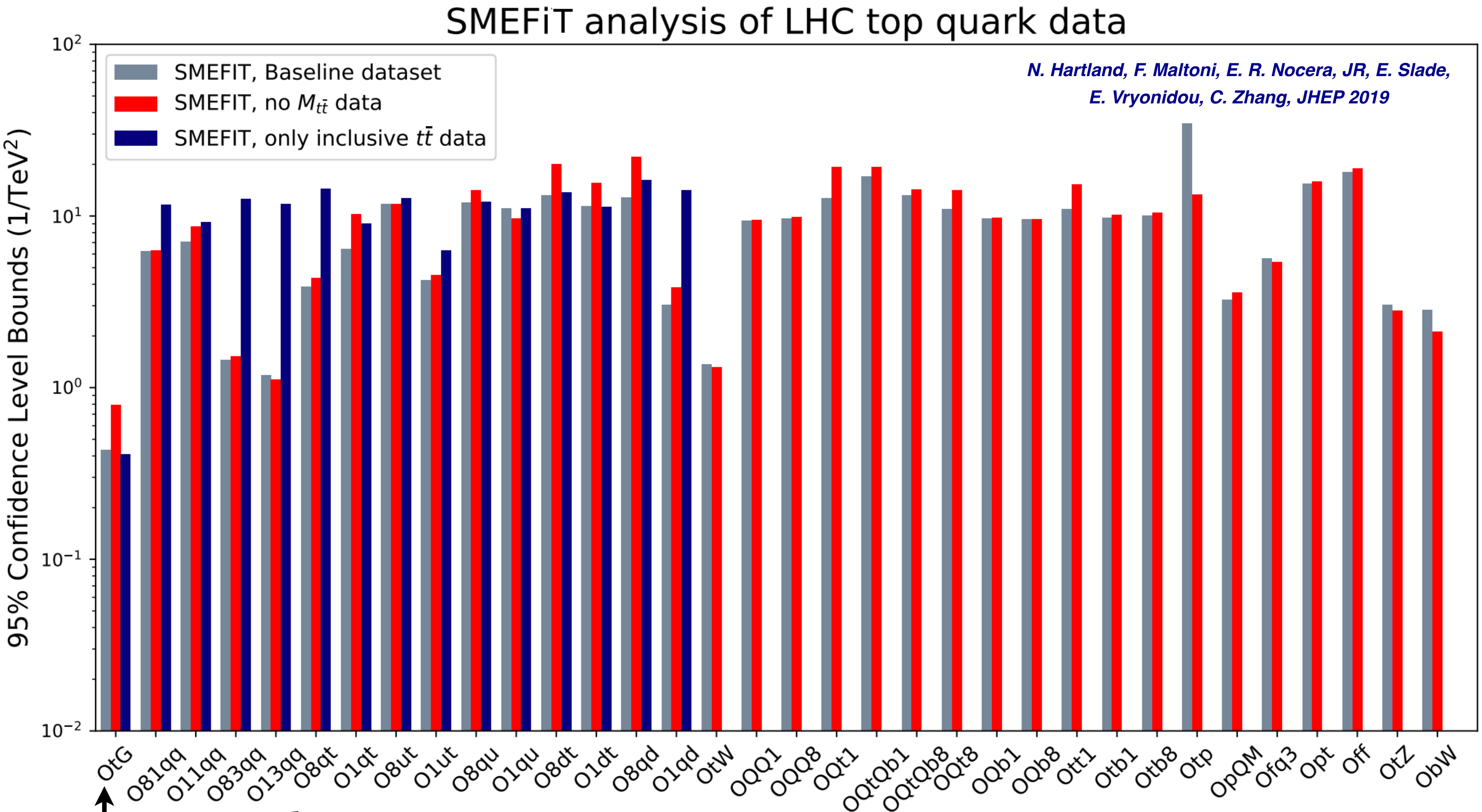
Inclusive vs differential data

Category	Processes	$n_{\rm dat}$	
Top quark production	$t\bar{t}$ (inclusive) **(incl LHC charge asy)**	94	**incl+diff**
	$t\bar{t}Z$, $t\bar{t}W$ **(incl ptZ in ttZ)**	14	**incl+diff**
	single top (inclusive)	27	**incl+diff**
	tZ, tW	9	**incl**
	$t\bar{t}t\bar{t}$, $t\bar{t}b\bar{b}$	6	**incl**
	Total	**150**	
Higgs production and decay	Run I signal strengths	22	**incl**
	Run II signal strengths	40	**incl**
	Run II, differential distributions & STXS	35	**diff**
	Total	**97**	
Diboson production	LEP-2 **(WW)**	40	**diff**
	LHC **(WW & WZ)**	30	**diff**
	Total	**70**	
Baseline dataset	**Total**	**317**	

Some operators mostly constrained by **inclusive data** (including LEP's EWPOs) in the global SMEFT fit, others by **differential measurements**

Inclusive vs differential data

Top-only EFT fit, compare **baseline** with fit **wo differential data**



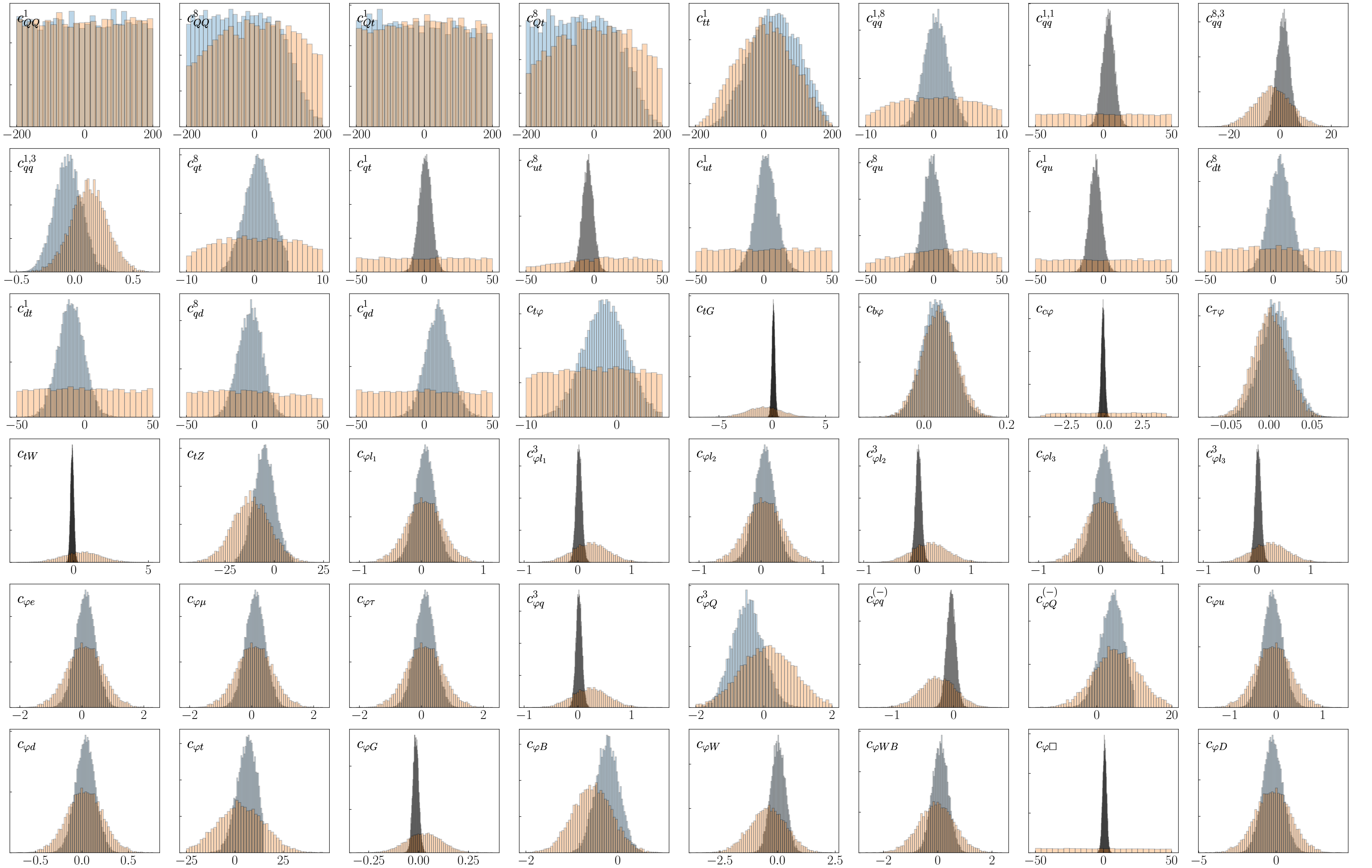
OtG dominated by inclusive cross-sections

2-light-2-heavy dominated by mtt differential distributions

only incl measurements available for 4-heavy operators

Inclusive vs differential data

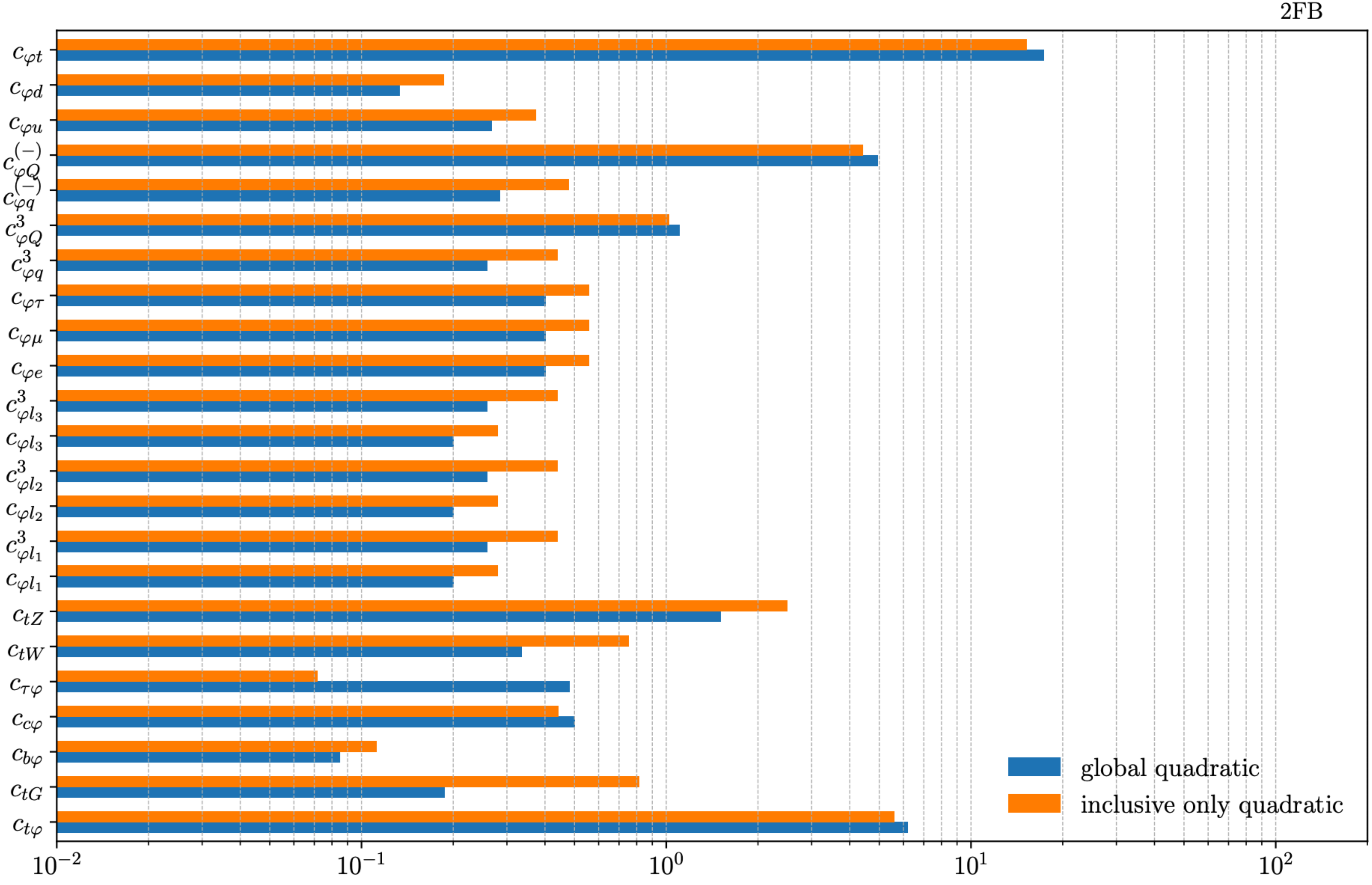
Global EFT fit, compare **baseline** with fit **wo differential data**



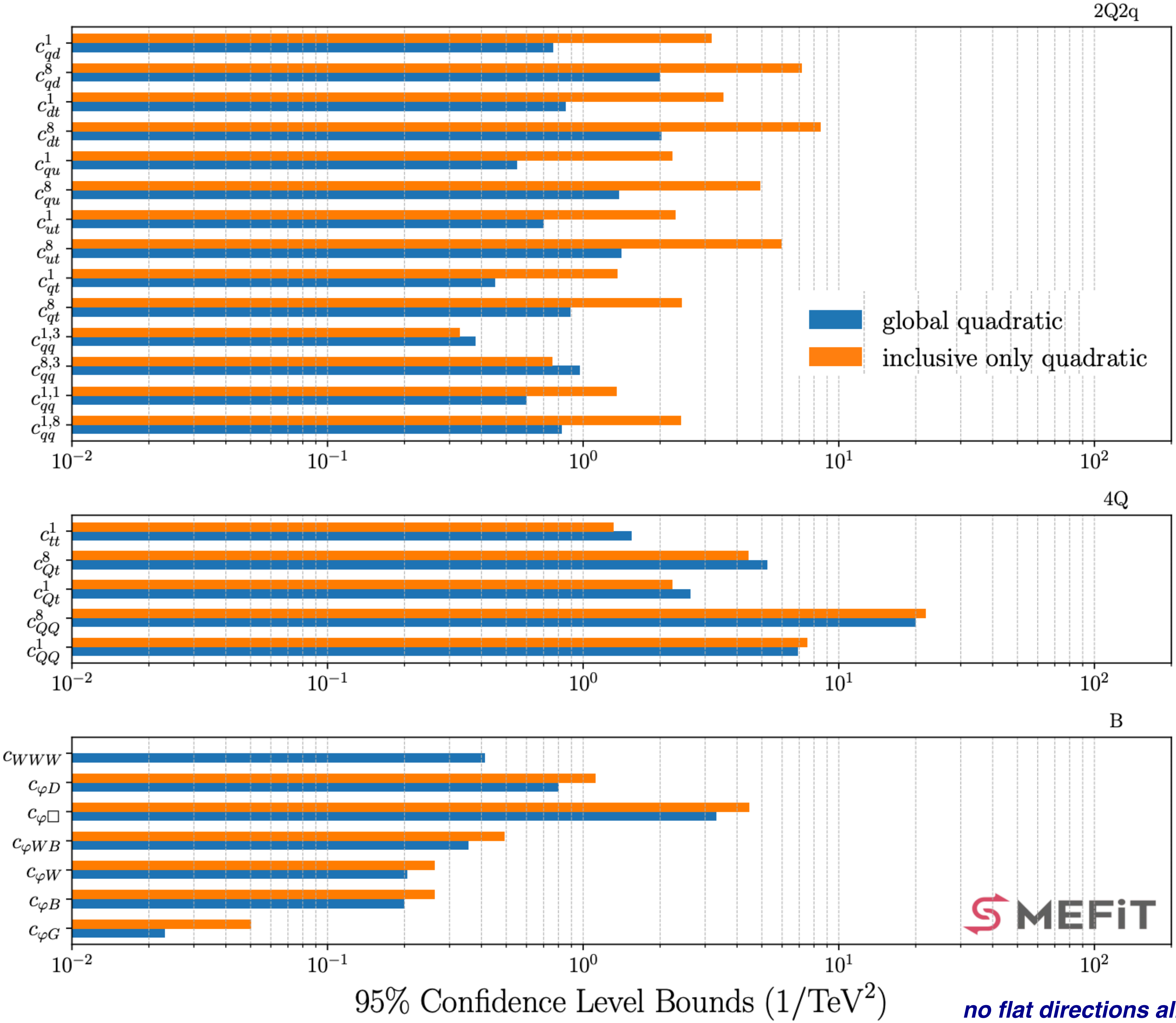
differential data instrumental to close flat directions and strengthen the EFT constraints

Inclusive vs differential data

Global EFT fit, compare **baseline** with fit **wo differential data**



Inclusive vs differential data



no flat directions allowed in quadratic fit, diff measurements still strengthen EFT constraints

Differential Measurements for EFT fits

Differential measurements are clearly important for the LHC EFT program. Which other **physics questions** we can tackle thanks to differential measurements?

- Testing validity of EFT expansion

  with high-energy measurements

- Validate RGE running of EFT operators

  with differential measurements

- Quantify impact of upcoming LHC measurements

  the smefit framework reloaded

- Determine maximal EFT sensitivity of a given process

  by means of machine learning methods

- Assess interplay between EFT and PDF fits

  simultaneous SMEFT-PDFs determination

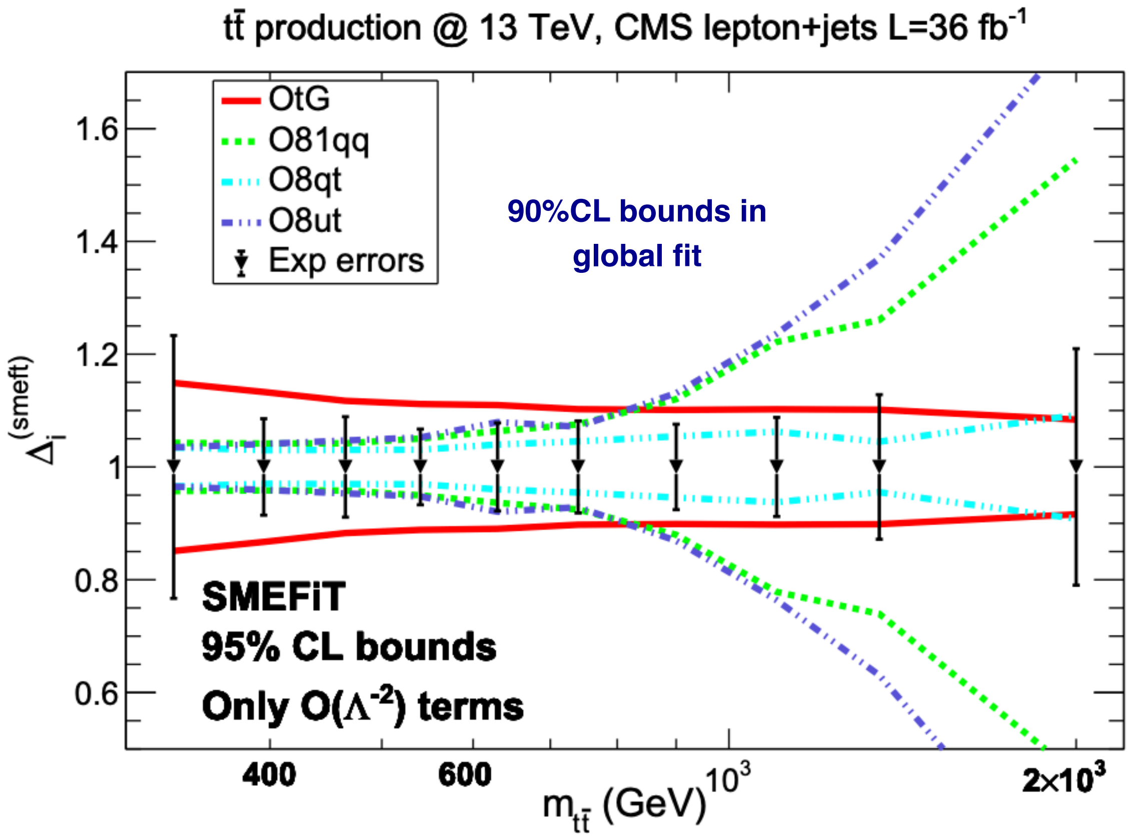
Tests of EFT validity

Theory predictions in the SMEFT depend on the ratio **c/Λ²** .
They also depend on the hard scale of the process $\hat{s}$

$$\sigma_{\rm LHC}\left(\{c_i/\Lambda^2\},\hat{s}\right)$$

A **EFT (model-independent) interpretatio**n of data can only constrain **c/Λ²** .
Separate bounds on **c** need to assume some scale **Λ**

Only consistent whenever $\sqrt{s} \ll \Lambda$ for all the data used in the fit



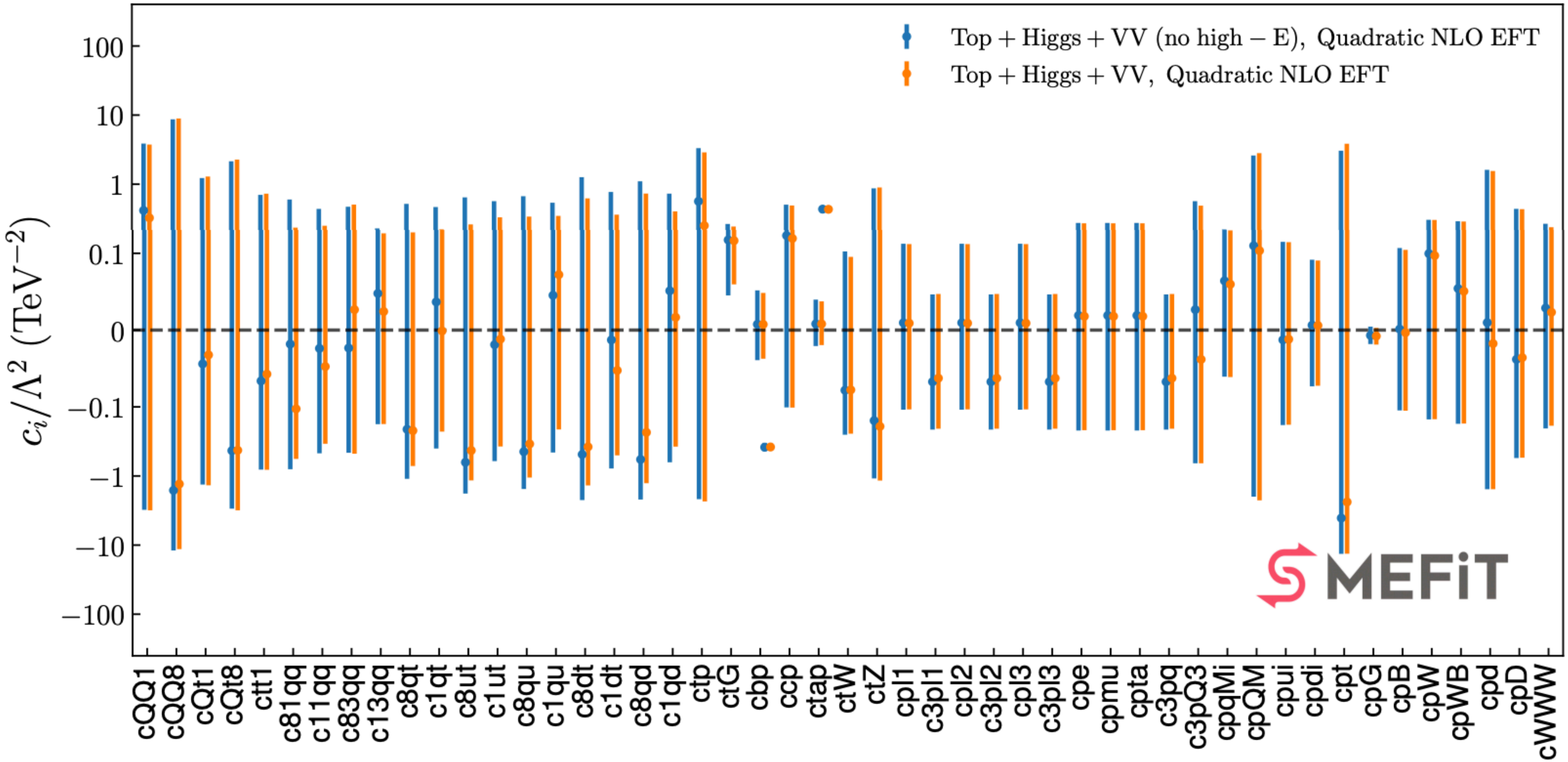
EFT corrections grow with energy for many operators

Differential measurement enable **tests of the EFT validity**, and allow disentangling whether dominant constraints come from high-energy tails

Tests of EFT validity

Assess EFT results stability by removing all bins involving $\hat{s} \geq 1\ \text{TeV}$

2-light-2-heavy operators relevant for high-mass top-quark pair worse constrained, but in general **moderate differences**

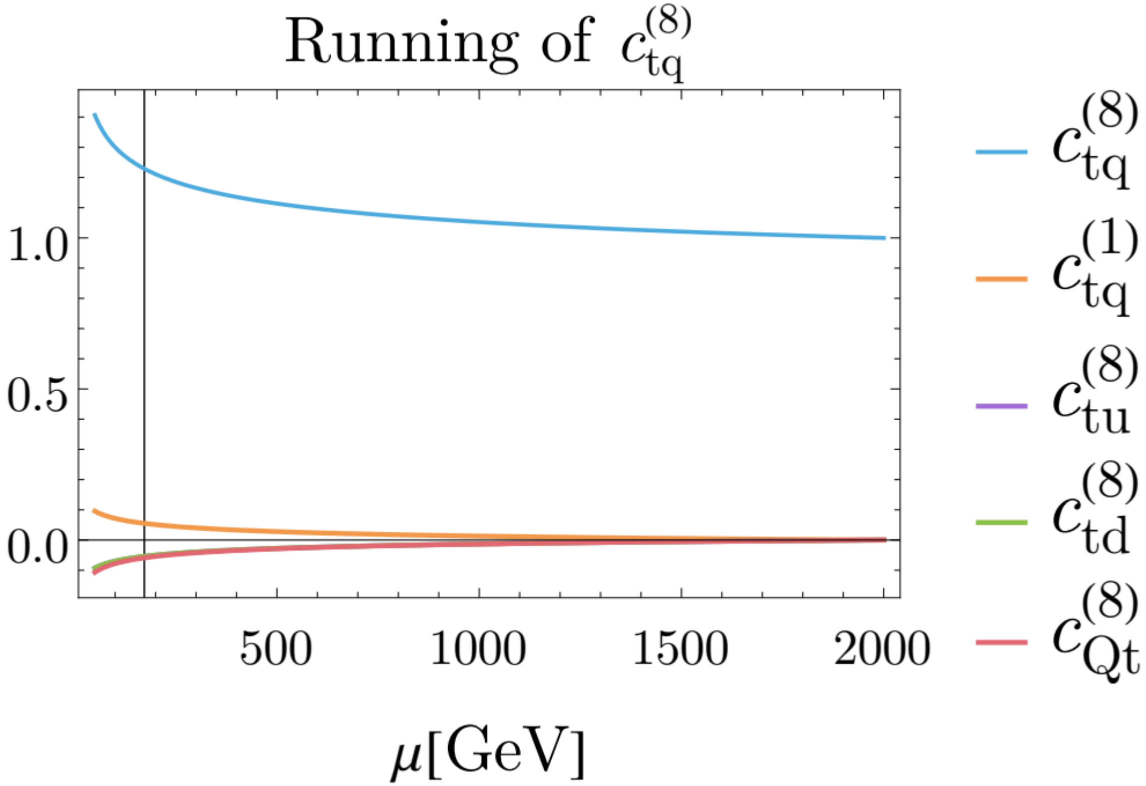


Current data moderately sensitive to high-E tails (low stats), to revisit when updated high-stat measurements available

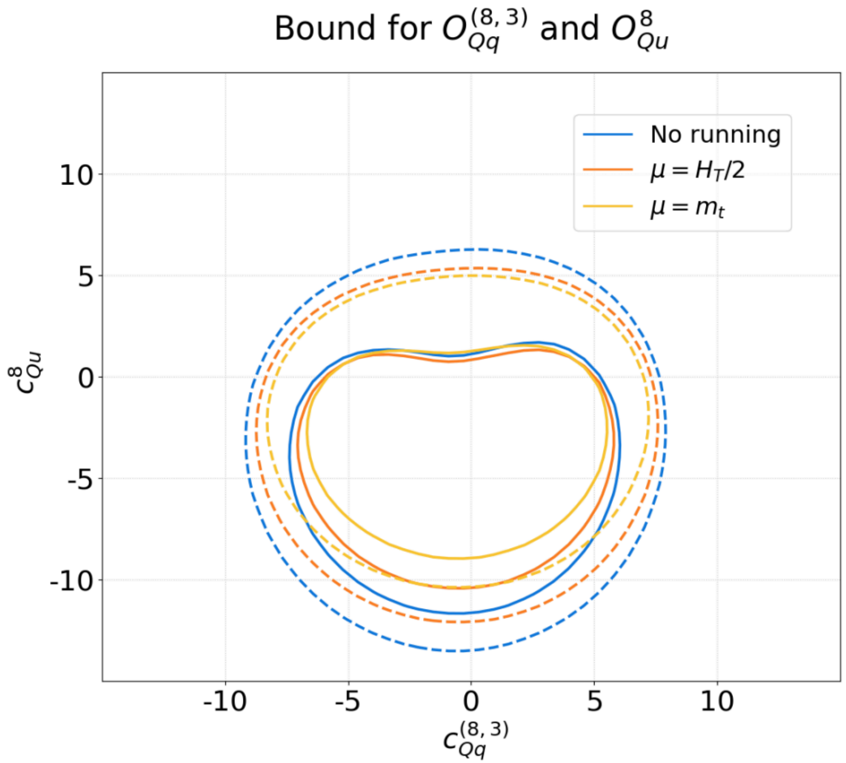
Operator Running and Mixing

EFT Wilson coefficient are only scale-independent at LO. In the presence of NLO QCD corrections they **run with the scale and mix**

even if only c8tq is non-zero at 2 TeV, other coefficients become non-zero at 400 GeV



RGE effects moderate for current data



R. Aoude, F. Maltoni, O. Mattelaer, C. Severi, E. Vryonidou 22

RGE effects partially accounted for is **NLO corrections for EFT cross-sections** are accounted for

Accounting for RGE effects (possibly via NLO corrections) important to relate processes with different scales and to consistently interpret differential measurements

SMEFiT reloaded

- Ultimately, the best way to answer the question ``how I would obtain better SMEFT constraints, doing my measurement a la A, a la B, or a la C?" is by **including them within a global SMEFT fit** and comparing results

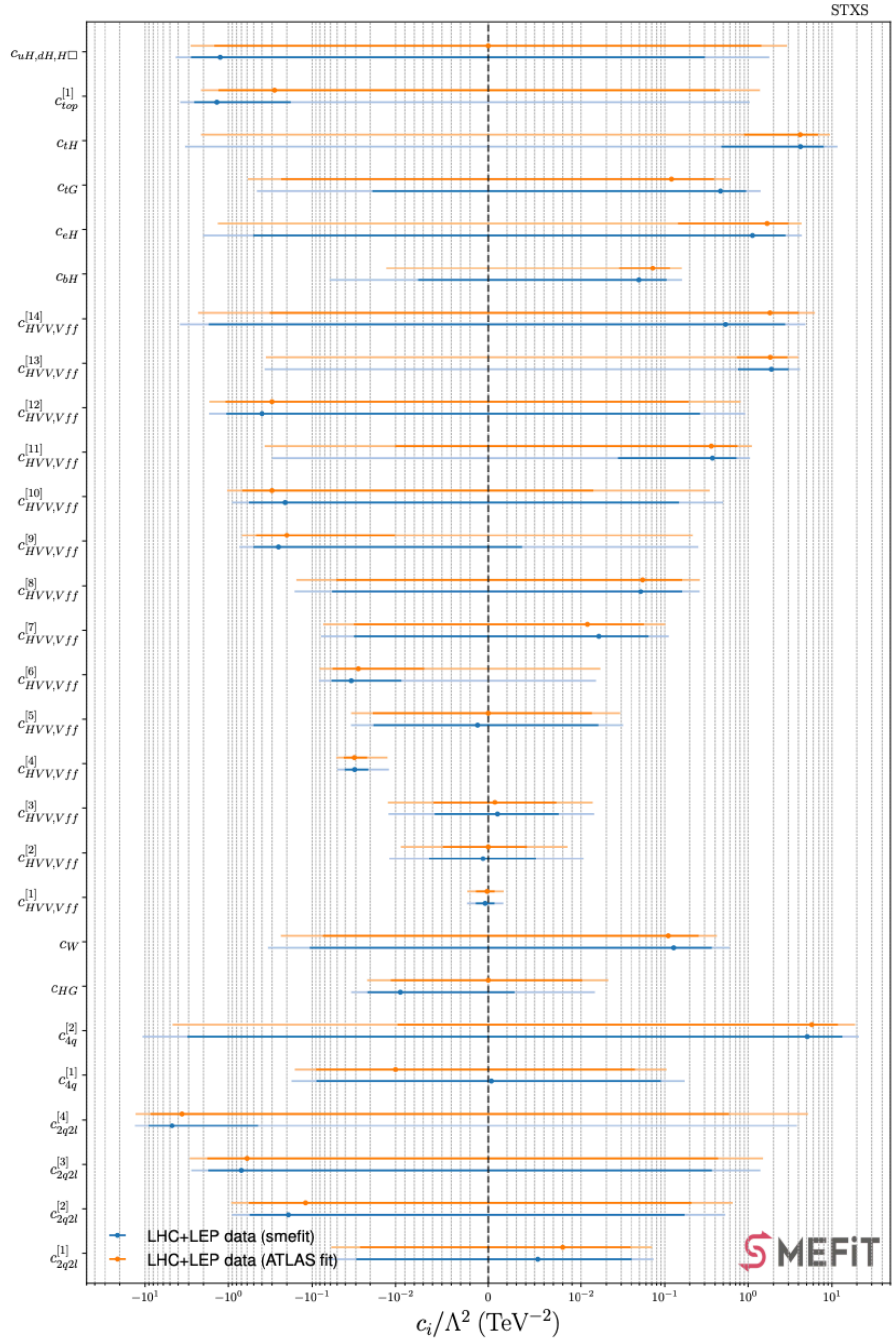
- Several tools available. We recently released **SMEFiT as an open-source framework** reproducing all previous results and extended with various improvements

  https://lhcfitnikhef.github.io/smefit_release/

- As an application example, we reproduce the results of the **ATLAS EFT interpretation of Higgs measurements** based on the full Run II dataset

T. Giani, G. Magni, JR, EPJC in press (2023)

reproduce ATLAS EFT fit results in their same basis …



SMEFiT reloaded

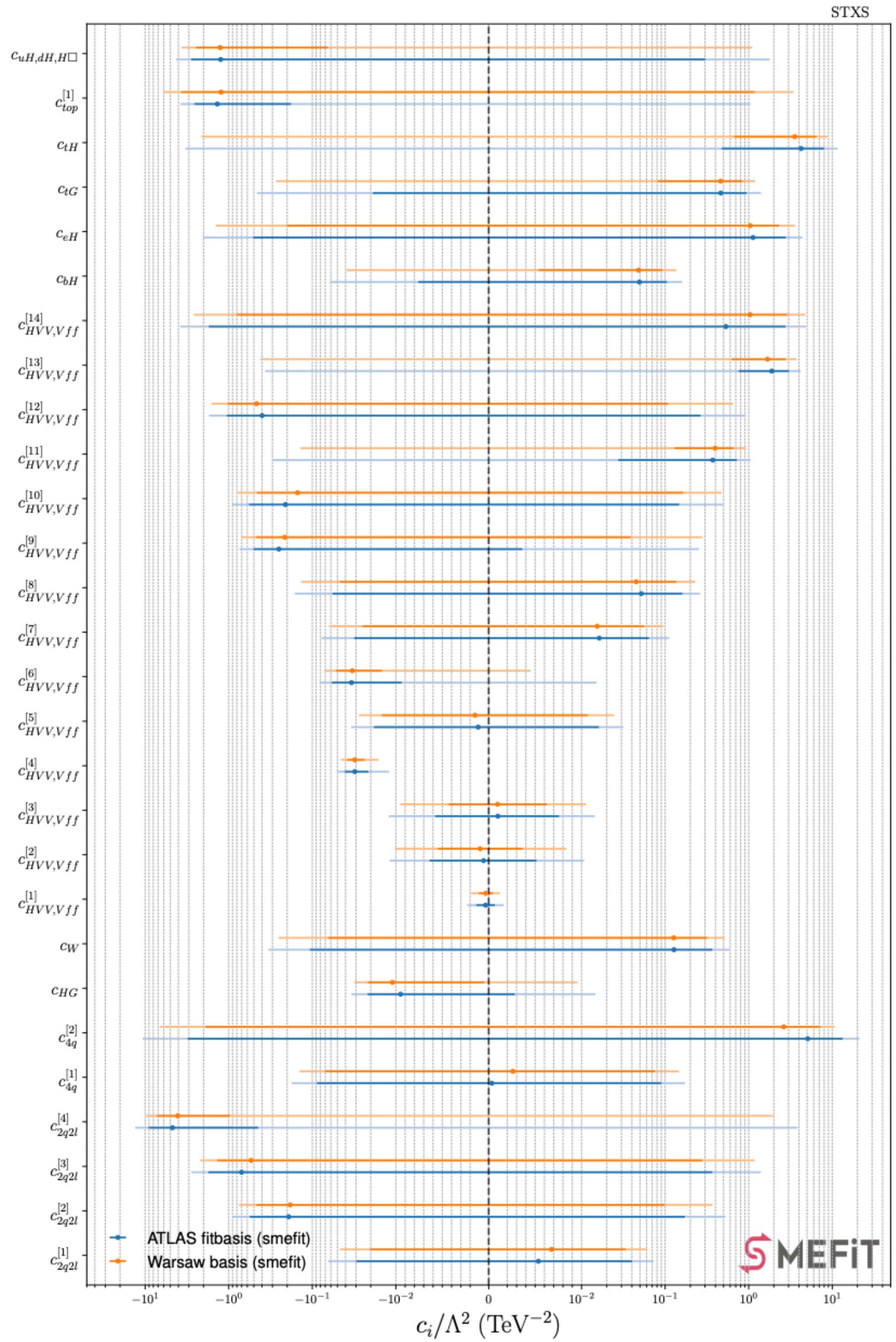
- Ultimately, the best way to answer the question ``how I would obtain better SMEFT constraints, doing my measurement a la A, a la B, or a la C?" is by **including them within a global SMEFT fit** and comparing results

- Several tools available. We recently released **SMEFiT as an open-source framework** reproducing all previous results and extended with various improvements

  https://lhcfitnikhef.github.io/smefit_release/

- As an application example, we reproduce the results of the **ATLAS EFT interpretation of Higgs measurements** based on the full Run II dataset

explicitly verity basis stability: ATLAS vs Warsaw



benchmarking essential to make progress in the ``EFT precision" era

SMEFiT reloaded

What would you (a CMS experimentalist) need to **quantify the impact of your measurement** when added on top of a global SMEFT fit?

The **statistical model of your measurement**, either in a multi-gaussian approximation or with the full likelihood model

Parametrisation of the measured observable in terms of **EFT corrections in the Warsaw basis** (or any other basis related to Warsaw by a rotation)

Translate into **smefit .json format**

Translate into **smefit .json format**

Run **smefit** either for the **standalone** measurement or together with the **global dataset**

plot posteriors,
2d marginalised contours

statistical validation,
SM stress-tests

data vs theory comparison,
predictions for other processes

evaluating the EFT parametrisation is typically the most time consuming part - evaluating the posterior distributions with smefit takes a few hours even with around 60 parameters

Optimal observables for global EFT fits

Which kind of measurement is **most sensitive to the SMEFT parameter space?**

Difficult question to answer in general since SMEFT-sensitive measurements can be:

- Inclusive or (1,2,3, …)-differential, with wide range of possible kinematic variables
- Binned (choice of binning?) or unbinned
- Unfolded at parton level, at particle level, or at detector level

relevant to many other extractions of SM & BSM parameters from data

construct **unbinned multivariate measurements** to determine the **optimal sensitivity** that a given process can have on SMEFT operators by means of **machine learning techniques**

quantify when going (multi)-differential is worth the hassle!

Unbinned multivariate likelihood

sum over events

event probability

$$\mathcal{L}(\boldsymbol{c}) = \frac{\nu_{\rm tot}(\boldsymbol{c})^{N_{\rm ev}}}{N_{\rm ev}!} e^{-\nu_{\rm tot}(\boldsymbol{c})} \prod_{i=1}^{N_{\rm ev}} f_\sigma\left(\boldsymbol{x}_i, \boldsymbol{c}\right) \qquad f_\sigma\left(\boldsymbol{x}, \boldsymbol{c}\right) = \frac{1}{\sigma_{\rm fid}(\boldsymbol{c})} \frac{d\sigma(\boldsymbol{x}, \boldsymbol{c})}{d\boldsymbol{x}} \, ,$$

event kinematics

retains full information on event-by-event kinematics

Many proposals available *(including by CMS colleagues here). Focus for simplicity on results from my group*

Optimal observables for global EFT fits

the dependence of the cross-section on **kinematic variables and all EFT coefficients**

$$r_\sigma(\boldsymbol{x},\boldsymbol{c}) \equiv \frac{f_\sigma(\boldsymbol{x},\boldsymbol{c})}{f_\sigma(\boldsymbol{x},\boldsymbol{0})} = 1 + \sum_{j=1}^{n_{\rm eft}} r_\sigma^{(j)}(\boldsymbol{x})c_j + \sum_{j=1}^{n_{\rm eft}}\sum_{k\geq j}^{n_{\rm eft}} r_\sigma^{(j,k)}(\boldsymbol{x})c_j c_k$$

parametrised with **neural networks** trained to Monte Carlo simulations & benchmarked with exact calculations

$$\hat{r}_\sigma(\boldsymbol{x},\boldsymbol{c}) = 1 + \sum_{j=1}^{n_{\rm eft}} \mathrm{NN}^{(j)}(\boldsymbol{x})c_j + \sum_{j=1}^{n_{\rm eft}}\sum_{k\geq j}^{n_{\rm eft}} \mathrm{NN}^{(j,k)}(\boldsymbol{x})c_j c_k$$

extendable to **arbitrary number** of kinematic variables and EFT coefficients: training can be parallelised

methodological uncertainties (e.g. finite training samples) assess with the **replica method**

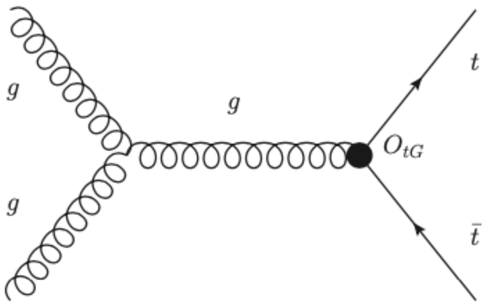
$$\hat{r}_\sigma^{(i)}(\boldsymbol{x},\boldsymbol{c}) \equiv 1 + \sum_{j=1}^{n_{\rm eft}} \mathrm{NN}_i^{(j)}(\boldsymbol{x})c_j + \sum_{j=1}^{n_{\rm eft}}\sum_{k\geq j}^{n_{\rm eft}} \mathrm{NN}_i^{(j,k)}(\boldsymbol{x})c_j c_k\,, \qquad i = 1,\ldots,N_{\rm rep}$$

each replica trained to an independent set of MC events

representation of the probability distribution in the space of ML models

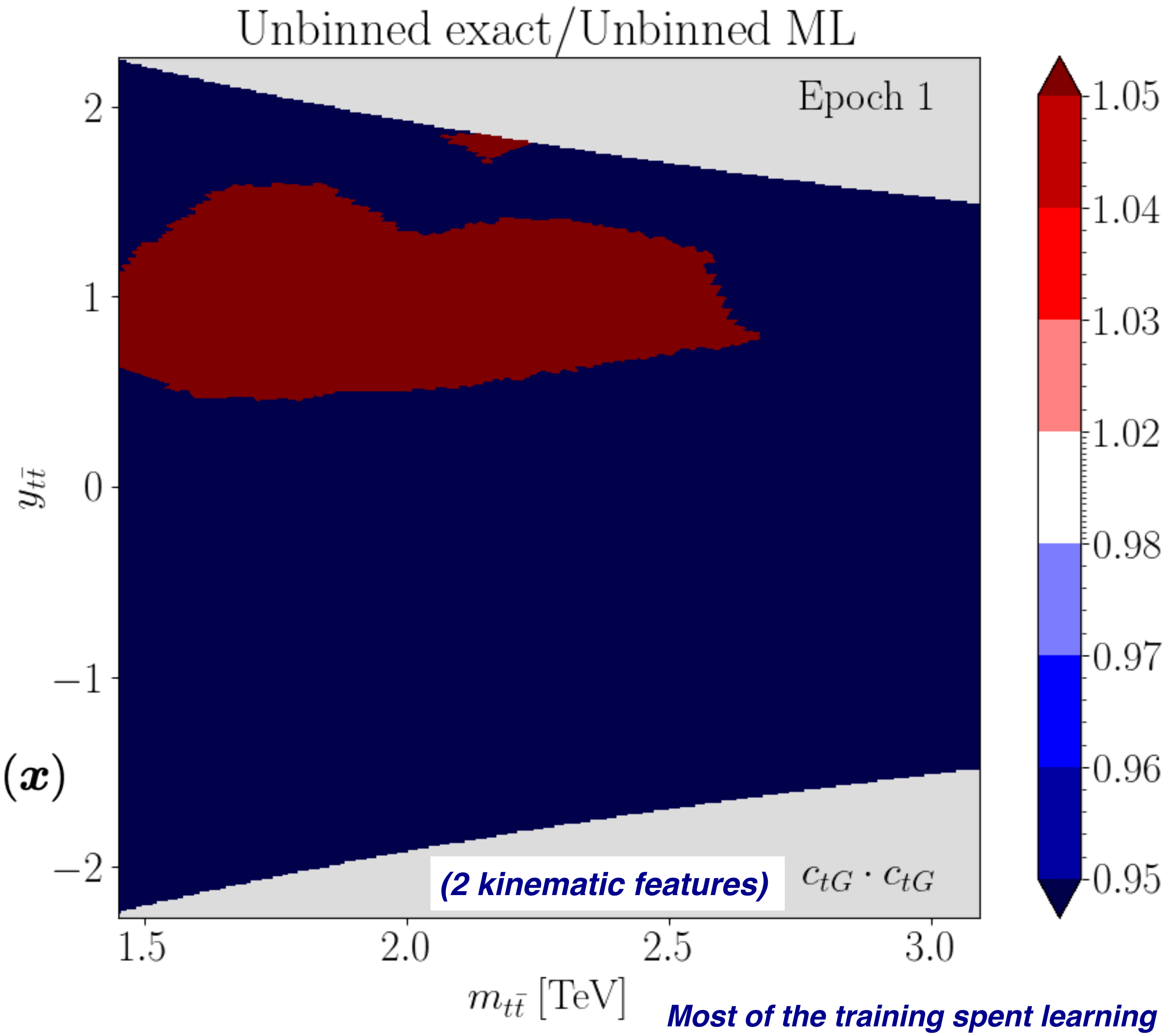
R. Gomez-Ambrosio, J. ter Hoeve, M. Madigan, JR, V. Sanz, JHEP 2023

Optimal observables for global EFT fits



$$r_\sigma(\boldsymbol{x}, c_j^{(\mathrm{tr})}, c_k^{(\mathrm{tr})}) = 1 + c_j^{(\mathrm{tr})} c_k^{(\mathrm{tr})} \mathrm{NN}^{(j,k)}(\boldsymbol{x})$$

$$\boldsymbol{x} = (m_{t\bar{t}}, y_{t\bar{t}})$$



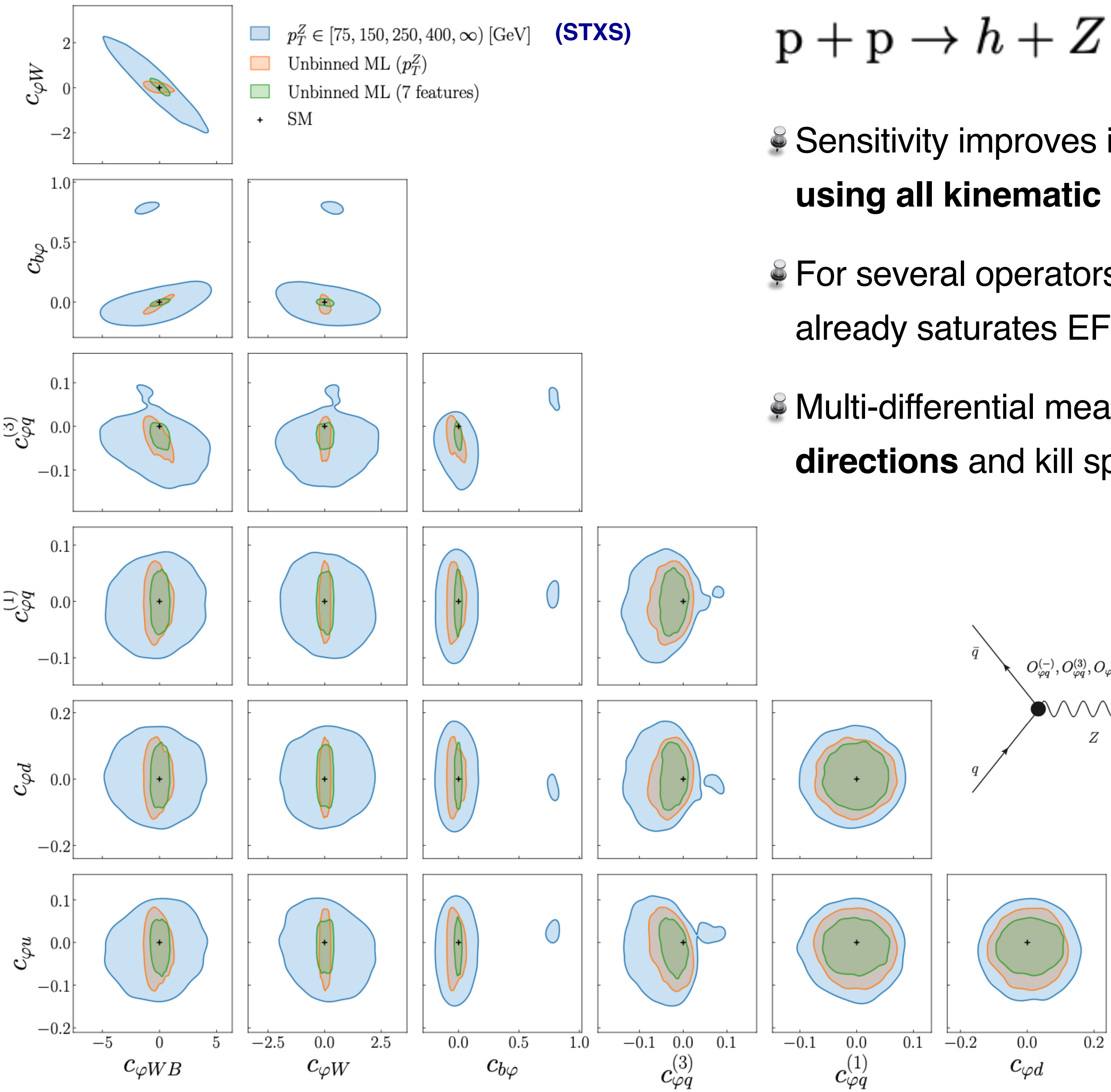
Most of the training spent learning the low-population (e.g. high-energy) tails

NN training by minimising cross-entropy loss function

$$L[g(\boldsymbol{x}, \boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c}) \sum_{i=1}^{N_{\mathrm{ev}}} \log(1 - g(\boldsymbol{x}_i, \boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0}) \sum_{j=1}^{N_{\mathrm{ev}}} \log g(\boldsymbol{x}_j, \boldsymbol{c}) \qquad g = (1 + r_\sigma)^{-1}$$

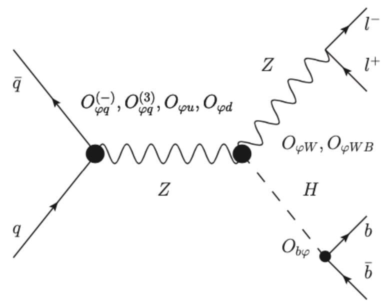
EFT constraints from Higgs+Z production

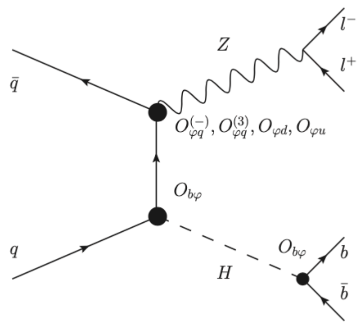
Marginalised 95 % C.L. intervals, $\mathcal{O}(\Lambda^{-4})$ at $\mathcal{L} = 300\ \text{fb}^{-1}$



$$\mathrm{p} + \mathrm{p} \to h + Z \to b + \bar{b} + \ell^+ + \ell^-$$

- Sensitivity improves in unbinned analysis and when **using all kinematic information**
- For several operators, a **traditional STXS analysis** already saturates EFT sensitivity
- Multi-differential measurements **close flat directions** and kill spurious solutions





take-home message: useful to quantify under which conditions the EFT sensitivity is saturated

The SMEFT PDF interplay

``How can one be sure you are not reabsorbing BSM physics into PDF fits?"

Assuming the **SM**, the theory calculations that enter a global PDF fit are:

$$\sigma_{\rm LHC}(\boldsymbol{\theta}) \propto \sum_{ij=u,d,g,\ldots} \int_{M^2}^{s} d\hat{s}\ \underbrace{\mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})}_{\text{SM PDFs}}\ \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s(M))$$

However in the case of BSM physics, here parametrised by the **SMEFT,** the correct expression is:

$$\sigma_{\rm LHC}\left(\boldsymbol{c}, \Lambda, \boldsymbol{\theta}\right) \simeq \left(\int_{M^2}^{s} d\hat{s}\ \overbrace{\mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})}^{\text{SMEFT PDFs}}\ \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s(M)) \right) \times \left(1 + \sum_{m=1}^{N_6} c_m \frac{\kappa_m}{\Lambda^2} + \sum_{m,n=1}^{N_6} c_m c_n \frac{\kappa_{mn}}{\Lambda^4} \right),$$

PDF parameters

SMEFT coefficients

Proof of concept for DIS: S. Carrazza, C. Degrande, S. Iranipour, JR, M. Ubiali, PRL 2019

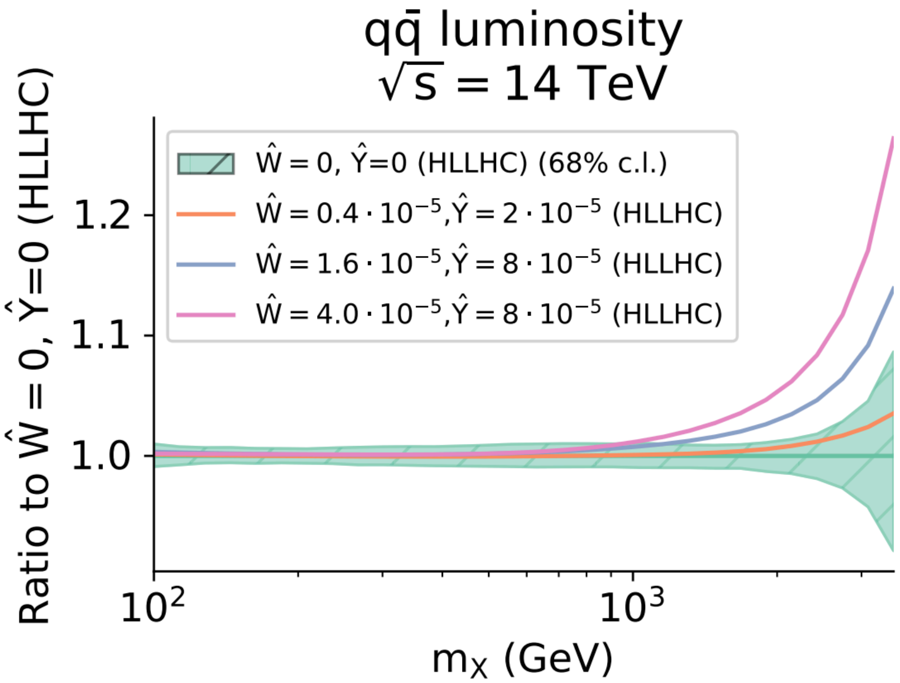
How different are ``SM PDFs" & ``SMEFT PDFs"? Can we quantify the risk of **fitting away BSM** in PDFs?

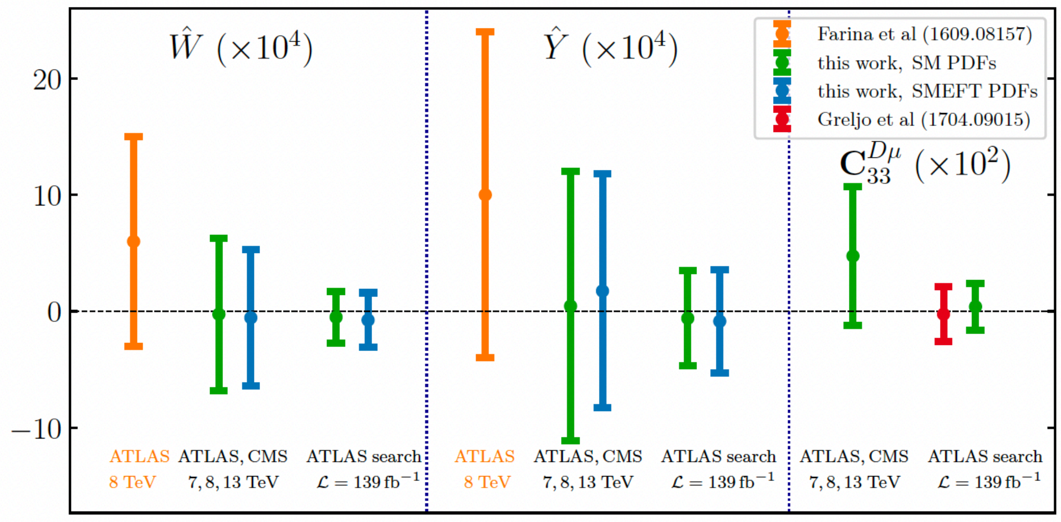
SMEFT PDFs from high-E Drell-Yan

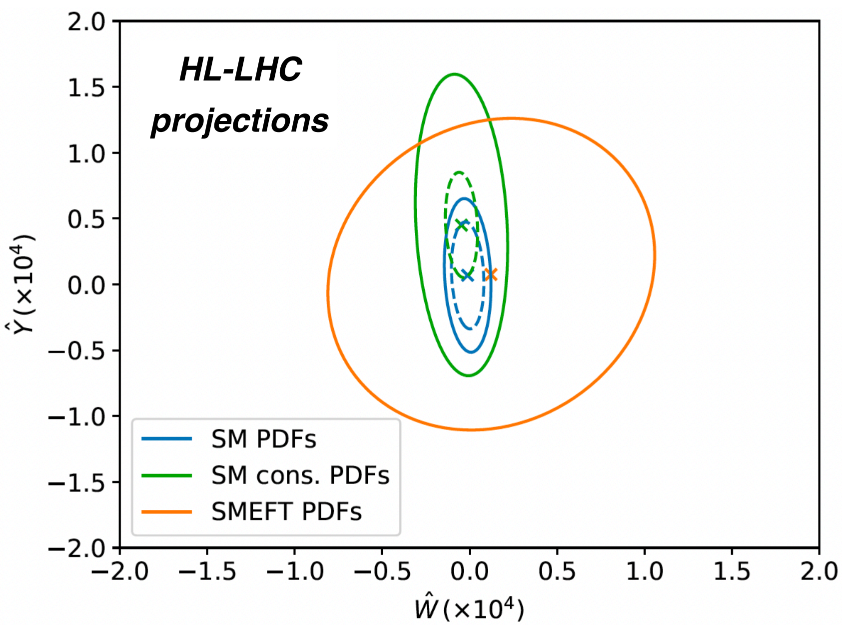
Extract SMEFT- PDFs from global fit where **high-mass DY cross-sections** account for EFT effects in two benchmark scenarios

Available data: limited interplay between PDF and EFT fits, best constraints from **searches**

HL-LHC: EFT effects, if present, would be **reabsorbed into PDFs**





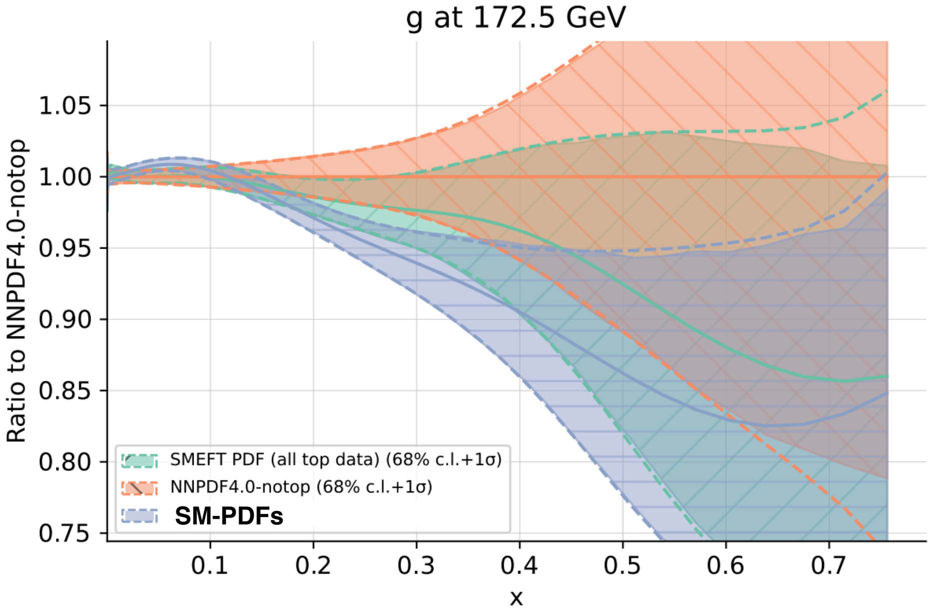


main limitation: ``SM" differential measurements take much longer than searches, preventing implementation in global SMEFT fits

A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, JR, M. Ubiali, C. Voisey, JHEP 2021

SMEFT PDFs from high-E top quarks

Consider all **available LHC top quark data** (including Run II legacy) and interpret them in terms of *i)* **SM-PDFs**, *ii)* **(fixed-PDF) EFT fit**, and *iii)* **SMEFT-PDFs**



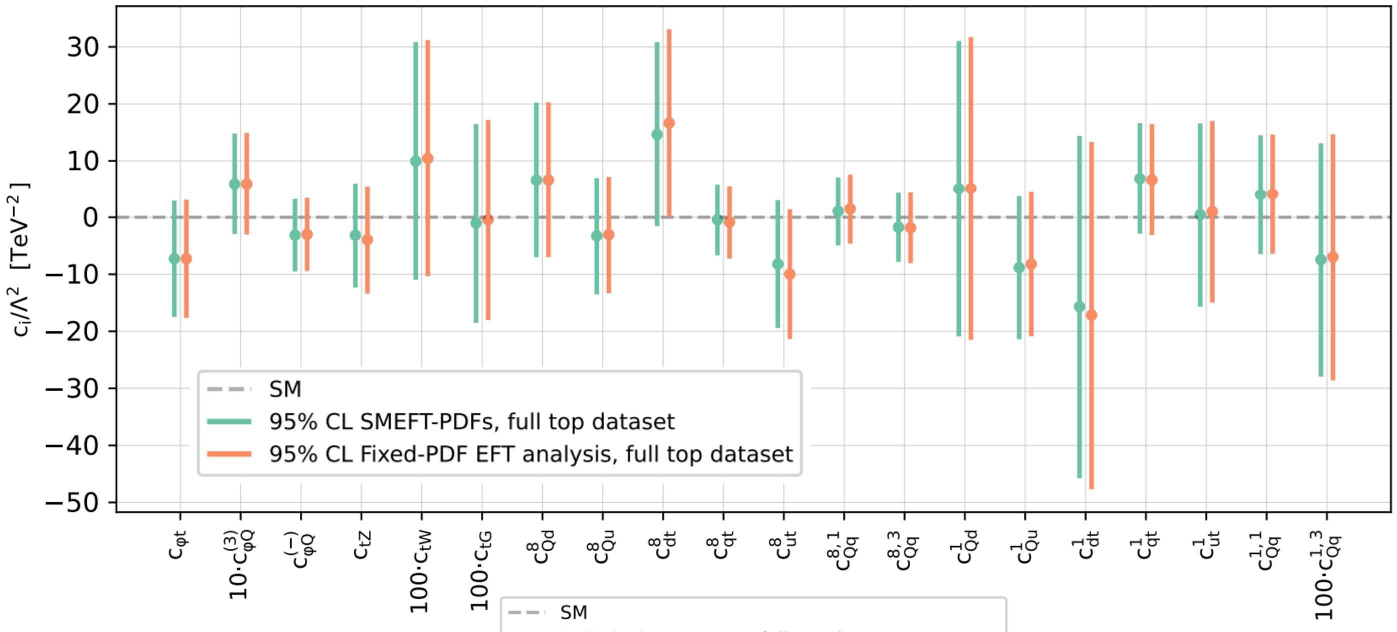
Large-*x* gluon **distorted by EFT effects**, which partially absorb the data pulls

EFT fit results stable when SMEFT-PDFs used (at least for Run II data)

Z. Kassabov, M. Madigan, L. Mantani , J. Moore , M.Morales-Alvarado, JR , M. Ubiali, JHEP under review

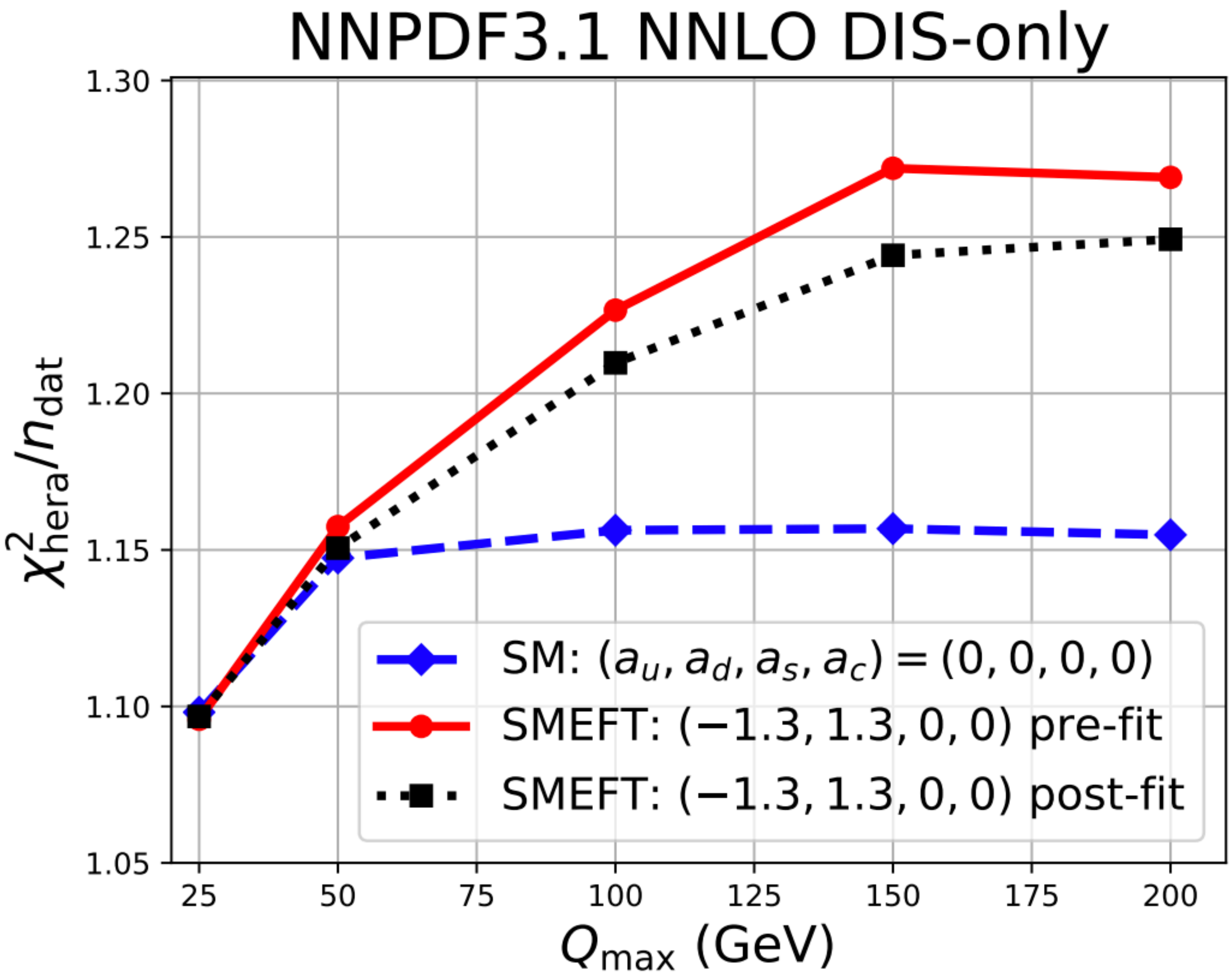
Impact on PDFs dominated by high-*E* top quark differential data

The **SMEFT PDF interplay** will only increase as more data is collected



Fingerprinting EFT effects

Tell-tale sign of SMEFT effects: **rapid variation with *Q*** (with QCD evolution slower)



DIS data, similar tests can be applied with LHC data

Only differential measurements can cleanly discriminate between QCD and BSM effects!

Summary and outlook

- The SMEFT framework provides a robust strategy to **interpret LHC data in terms of new BSM** phenomena while reducing model assumptions

- A single SMEFT analysis constrains a **plethora of UV-complete scenarios** (matched to the SMEFT) at one: bridge between data and BSM models

- **Differential measurements** instrumental for success of the SMEFT program: sensitivity to high-E tails, operator running/mixing, close flat directions & spurious solutions, …

- Differential measurements maye be sensitive to other phenomena beyond the EFT, such as PDFs. Crucial to assess **their interplay quantitatively** and define mitigation strategies

- Machine learning tools enable **quantifying the ``maximum sensitivity"** that can be expected form a given measurement for a specific set of SMEFT operators

- Public tools (like smefit) make it possible to **internally determine the impact of your own measurements,** either stand-alone or within a global EFT fit