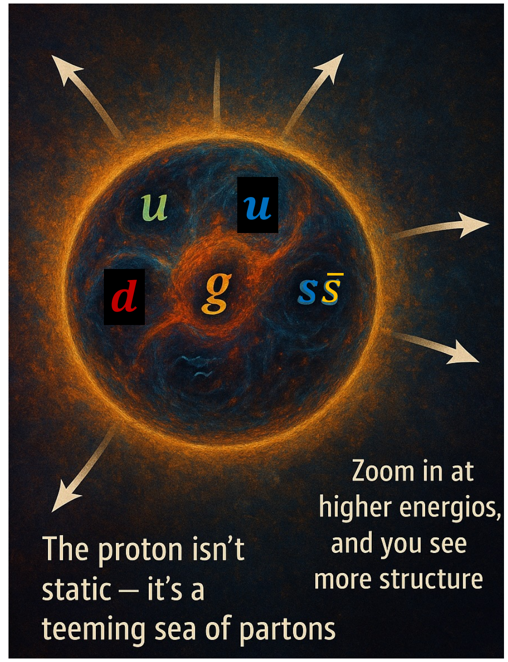
unlocking the proton: the EIC as a Discovery Machine

Lecture 1: **PDFs** and precision at the **LHC**

Tim Hobbs, Argonne National Lab

9 June 2025

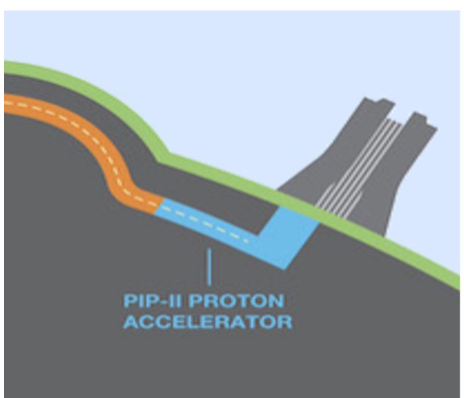


thanks to CTEQ colleagues

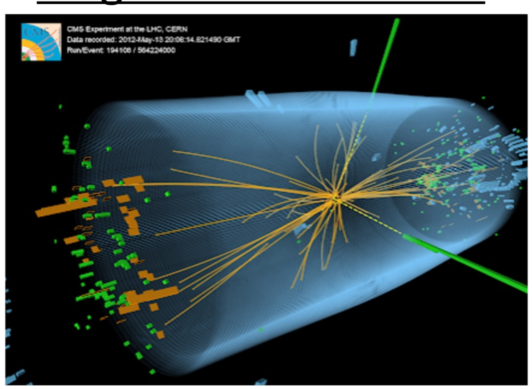
motivation question:
you have now heard about PDFs ---

how can they be measured???

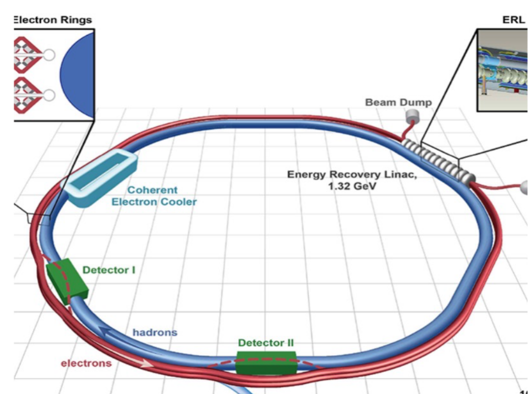
Fermilab



Large Hadron Collider



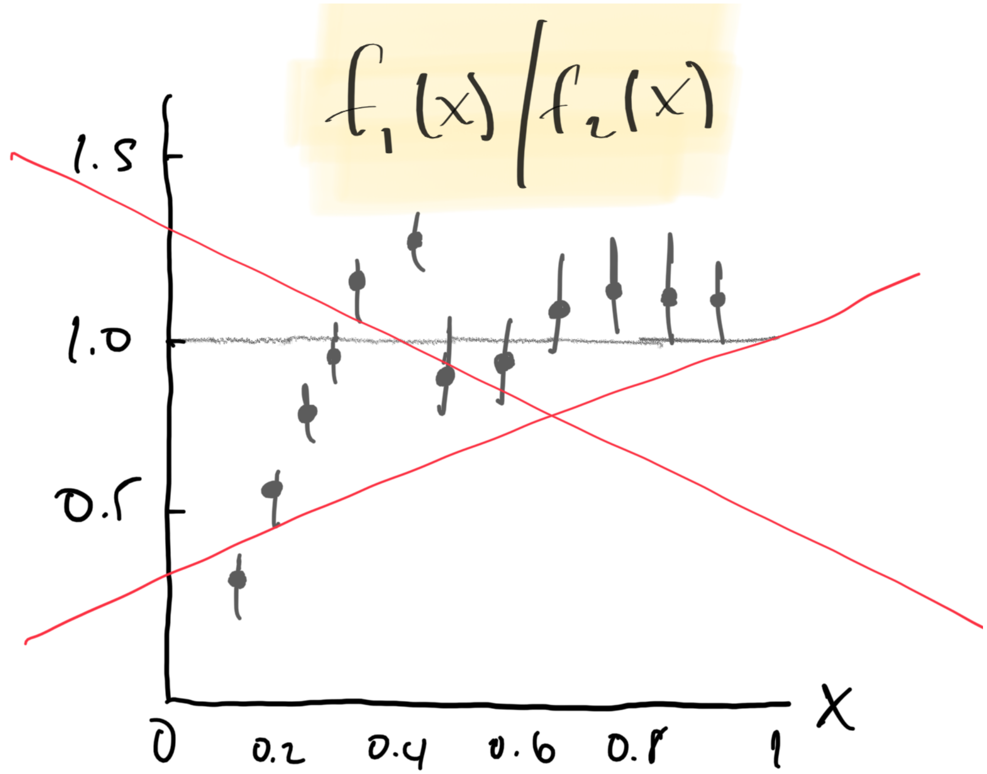
Electron-Ion Collider



apologies! I have begun with a trick question…

- PDFs are **<u>never</u>** "measured", though such abusive language commonplace

→ whatever you do, never create a plot like this:



→ QCD is a *confining theory*; PDFs are inherently theoretical objects

- must be inferred from high-energy data; represent grand <u>inverse problem</u>

→ developments allowing this have a long history predating QCD…

these lectures: PDFs from EIC → LHC and back again

talks will hopefully give you some *feel*; the subject is massive, however!

- **Lecture 1**: PDF fundamentals
  - historical context; definitions
  - EIC motivations and reach
  - statistical underpinnings; uncertainty determination

- **Lecture 2**: PDF applications to particle phenomenology
  - (inter)connections with LHC phenomenology
  - novel analysis methods (AI/ML, …)

about me: Argonne particle theory (will be brief, I promise!) 😉

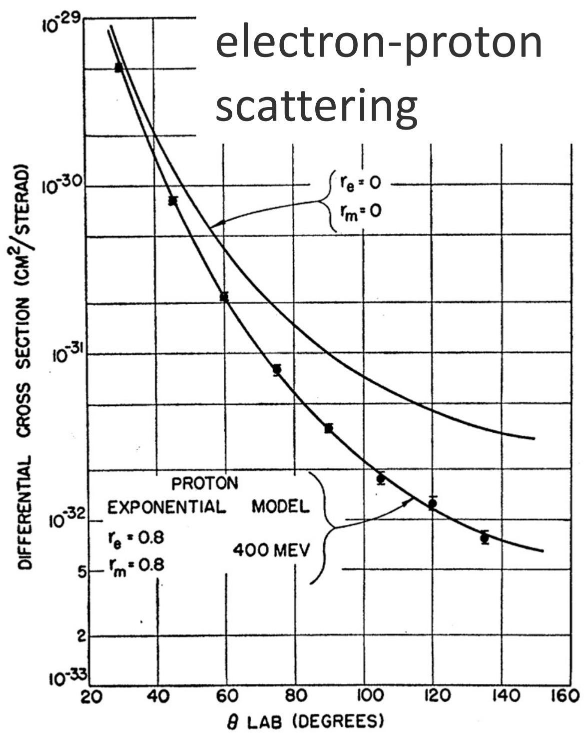
- Staff Scientist in High Energy Physics (HEP) Division Theory Group
  - precision theory at colliders (e.g., LHC)
  - QCD and global analyses (PDFs)
  - beyond standard model (BSM) searches
  - (particle) theory applications of AI/ML and related methods

- Argonne: ‘multi-purpose’ DOE National Laboratory, outside Chicago
  - HEP and nuclear physics; computation; materials science; oh my..
  - please visit us (tim@anl.gov)
  - also, we are hiring a new staff scientist! (apply, 30 June)

evidence of hadronic substructure, ~1950s

$< 10^{-15}\ \mathrm{m}$

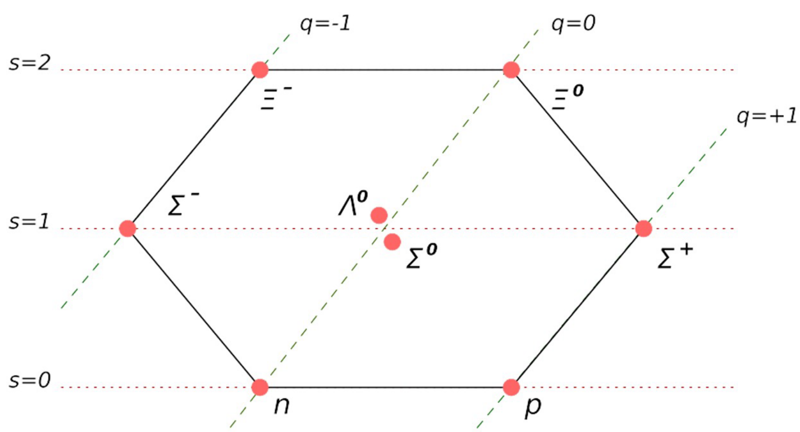
Hofstadter

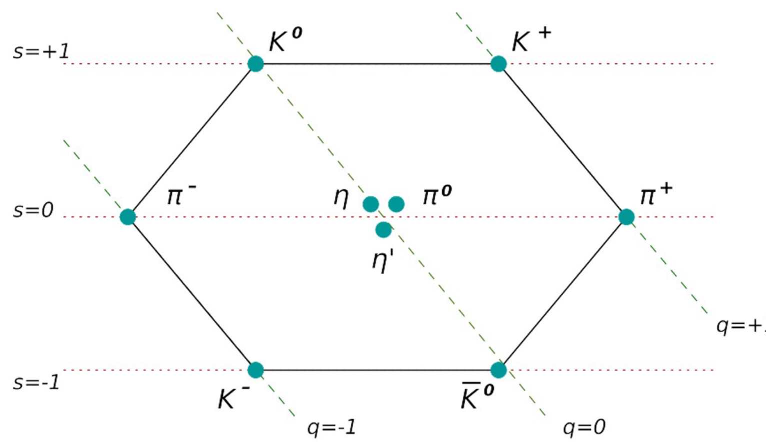


<u>elastic proton form factor</u> (probability for state to remain itself after being externally probed) found to have rapid falloff in energy

$$ep \to e'p'$$

contemporaneously, a ‘particle zoo’ of hadrons was rapidly discovered, with discrete orderings in charge and other degrees of freedom





1960s: inelastic experiments revealed proton's internal landscape

SLAC-MIT experiments: explore *inelastic scattering* of electrons on protons

$$ep \to e' + X$$

at higher energies, the inelastic cross section exhibited "scaling"

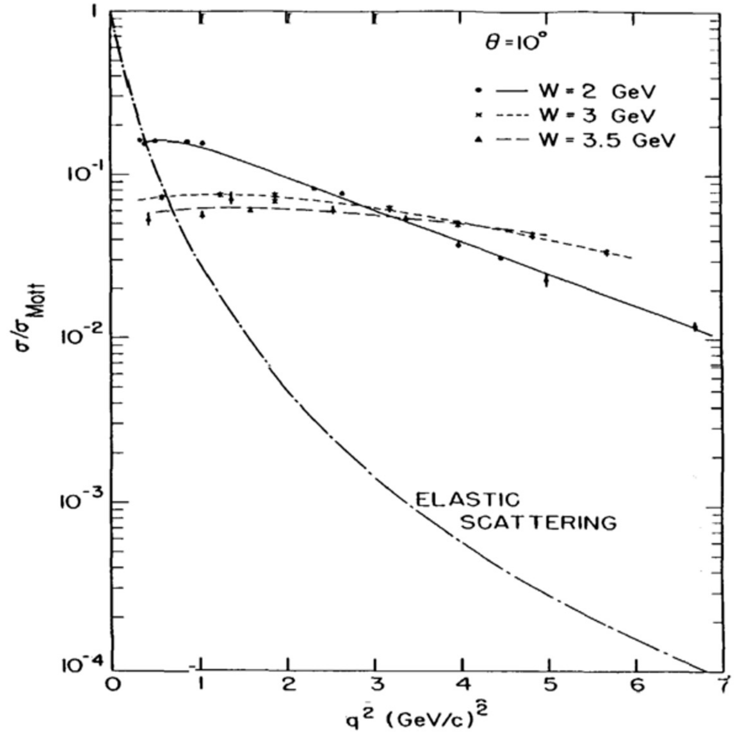
$\to$ *i.e.*, it became *approximately* independent of scattering energy

the scattering appeared to be from dense, pointlike charge cores (Feynman: "*partons*"; Gell-Mann: "***quarks***")

Bjorken

Feynman

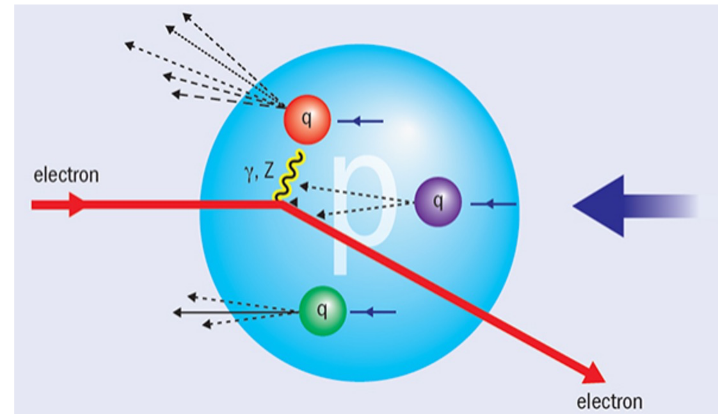
evidence for quarks, Bjorken scaling

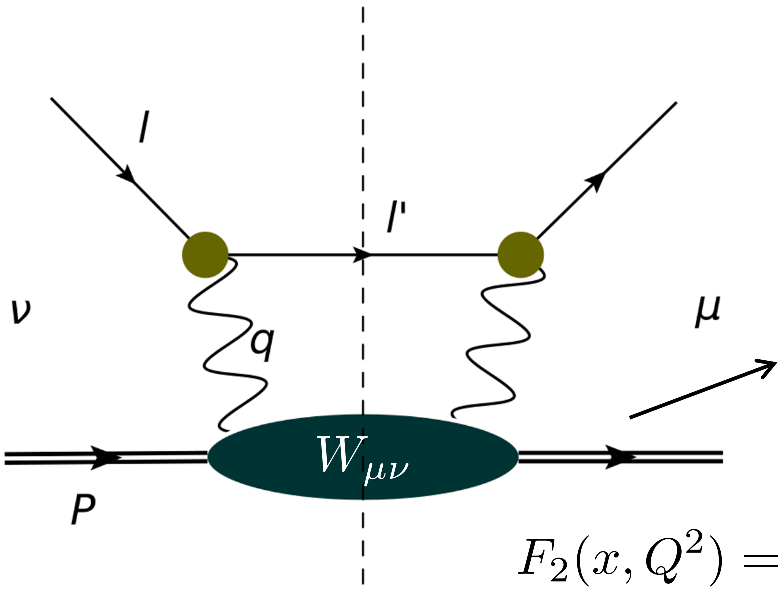


a closer look at how this is done: Deeply-Inelastic Scattering (DIS)

clean access to proton's **quark** and **gluon** (force carrier) constituents

→ electroweak probe sees spatial correlations of quark fields





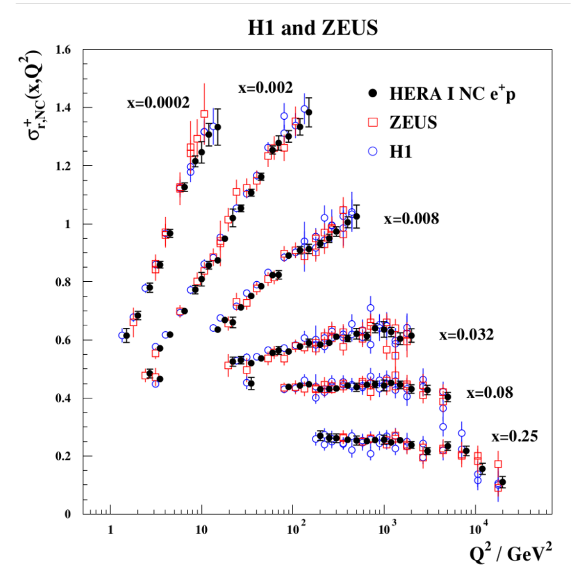
$$W_{\mu\nu} = \cdots + \frac{\hat{P}_\mu \hat{P}_\nu}{P \cdot q} F_2(x, Q^2) + \ldots$$

parton distribution functions (PDFs)

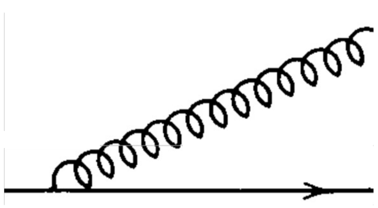
↓

$$F_2(x, Q^2) = x \sum_q e_q^2 \, \underline{(f_q + f_{\bar{q}})[x, Q^2]}$$

f_q : probability for quark q with mom. fraction x, energy Q



DIS data at different momentum scales constrain gluon

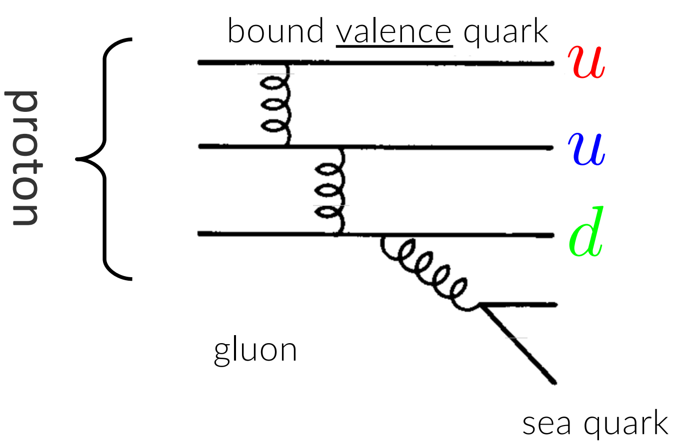


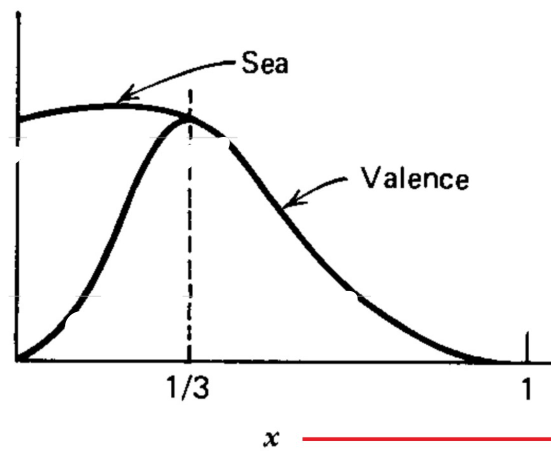
(quark fields exchange momentum through gluon radiation)

we have learned that hadrons and nuclei have an incredibly rich structure

PDF: f_q (x, Q): probability for quark q with mom. fraction x of proton, energy Q

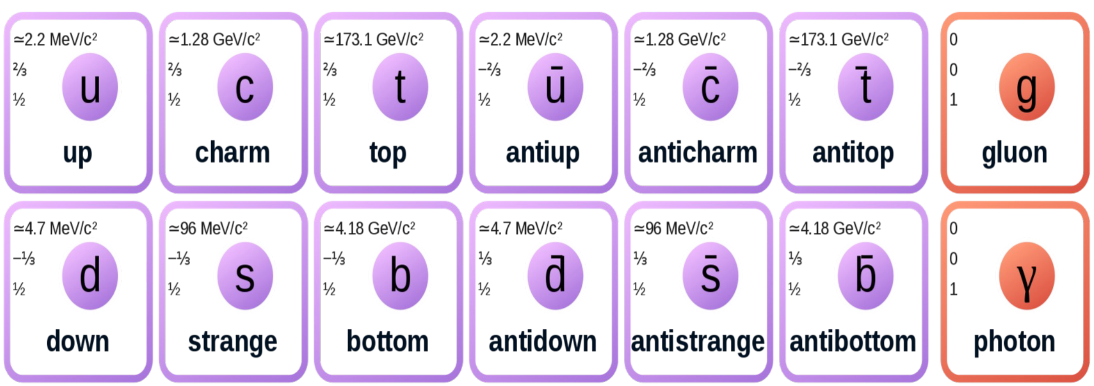
$$2e_u(= 2 \times 2/3) + e_d(= -1/3) = +1$$



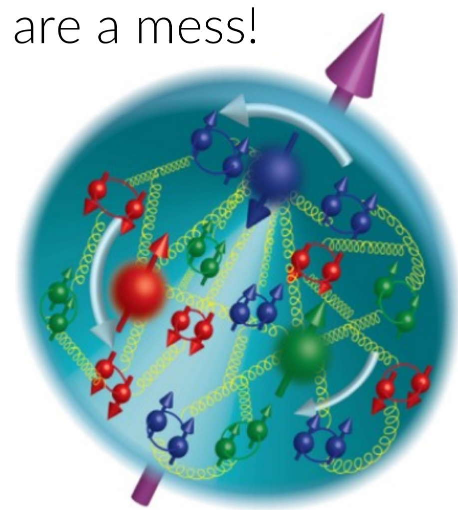


$$x = \frac{\text{quark mom.}}{\text{proton mom.}}$$

a wide array of Standard model particle contribute to proton structure at varying levels:



hadrons are a mess!



how is this information accessed? let QCD work for you!

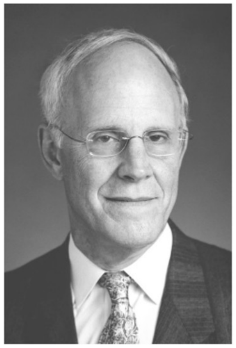


Photo from the Nobel Foundation archive.
David J. Gross
Prize share: 1/3

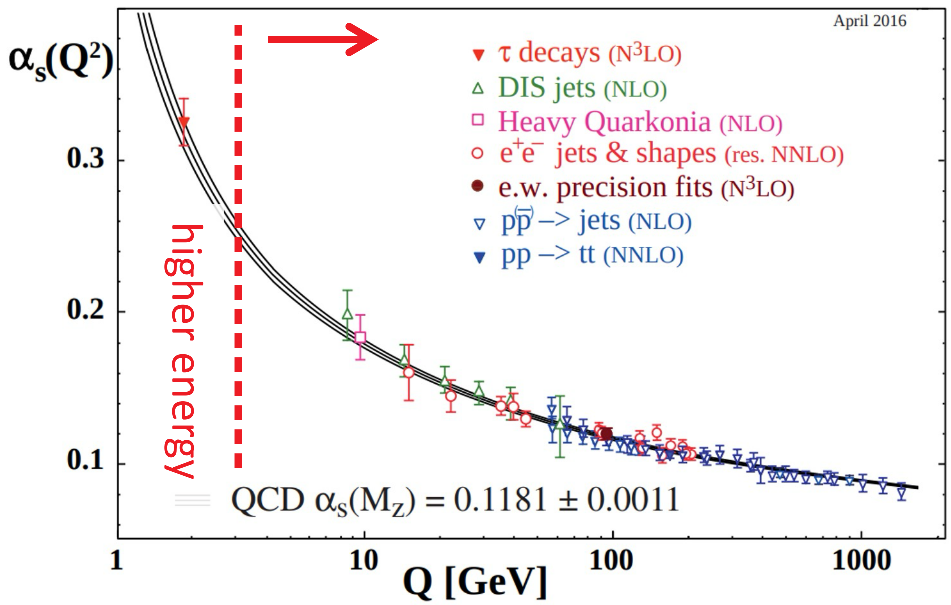
Photo from the Nobel Foundation archive.
H. David Politzer
Prize share: 1/3

Photo from the Nobel Foundation archive.
Frank Wilczek
Prize share: 1/3

the **β**-function of QCD is negative-definite,

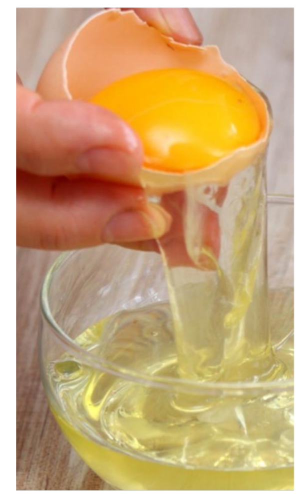
$$\beta(\alpha_s) = \mu_R^2 \frac{d\alpha_s}{d\mu_R^2} = -(b_0 \alpha_s^2 + \dots) < 0$$

→ quark-gluon interactions weak at high energies; **use perturbation theory**



↕ **QCD factorization**

at low energies interactions strong (nonperturbative); **extract PDFs** (other dists.)



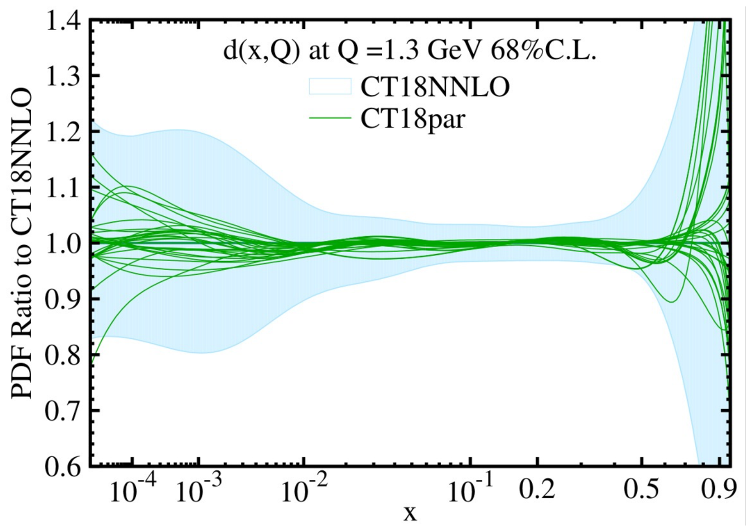
<u>fundamental question</u>: how does QCD, which so successfully describes high-energy processes, give rise to the emergent properties of low-energy bound states?

→ *a chief motivation* for QCD as a field… **answer partly related to PDF behavior**

QCD analyses operationalize factorization into PDF fits

PDFs (other correlation functions): nonperturbative hadronic matrix elements,

$$f_{q/p}(x,\mu^2) = \int \frac{d\xi^-}{4\pi} e^{-i\xi^- k^+} \langle p \,|\overline{\psi}(\xi^-)\gamma^+ \mathcal{U}(\xi^-,0)\psi(0)\,|p\rangle$$



challenging to compute from QCD!

...although there are lattice developments (talk by Lin)

philosophy: lacking first-principles calculation, fit flexible parametrization at suitable boundary condition for QCD evolution:

$$f_{q/p}(x,\mu^2 = Q_0^2) = a_{q_0} x^{a_{q_1}} (1-x)^{a_{q_2}}\, \mathrm{P}[x;\{a_{q_{n-3}}\}]$$

→ perturbatively-calculable evolution specifies dependence on $\mu^2 > Q_0^2$

→ fit the world's data from a diverse range of scales and processes

where can one find the PDFs?

→ the LHAPDF repository of course (Hepforge)

convenient interfaces available (including wrappers for Python/Julia)

…no phenomenologist should leave home without it.

```
tim@CSI364368 DIS % which lhapdf
/opt/homebrew/bin/lhapdf
tim@CSI364368 DIS % lhapdf --version
6.5.3
tim@CSI364368 DIS % lhapdf list "CT18NNLO*"
CT18NNLO
CT18NNLO_as_0110
CT18NNLO_as_0111
CT18NNLO_as_0112
CT18NNLO_as_0113
CT18NNLO_as_0114
CT18NNLO_as_0115
CT18NNLO_as_0116
CT18NNLO_as_0117
CT18NNLO_as_0118
CT18NNLO_as_0119
CT18NNLO_as_0120
CT18NNLO_as_0121
CT18NNLO_as_0122
CT18NNLO_as_0123
CT18NNLO_as_0124
tim@CSI364368 DIS % lhapdf install CT18NNLO_as_0118
CT18NNLO_as_0118.tar.gz:   310.7 KB[100.0%]
```

Get PDFs

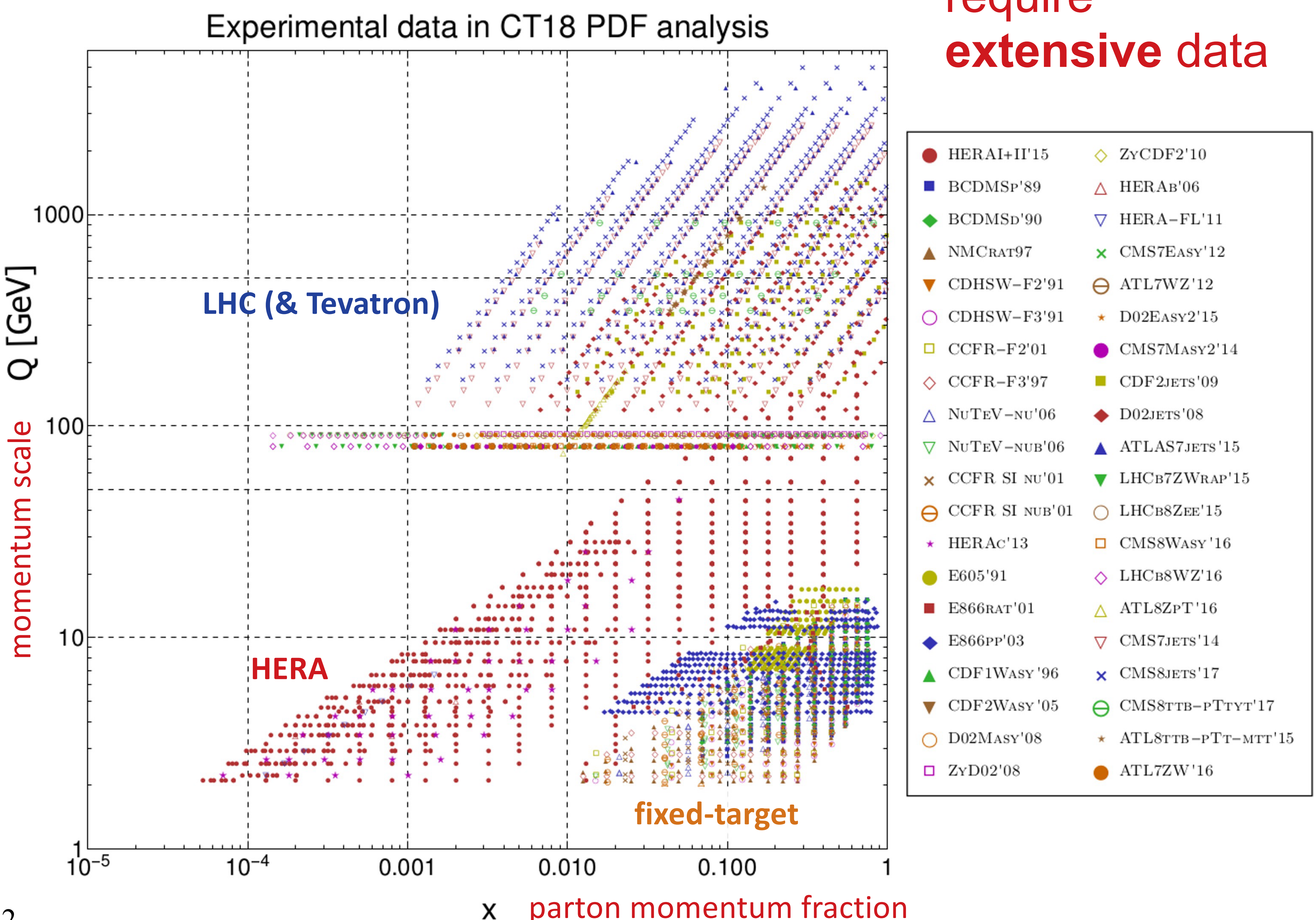
```
In [7]: pdf = LoadData.GetSet("CT18NNLO");

LHAPDF 6.5.3 loading all 59 PDFs in set CT18NNLO
CT18NNLO, version 1; 59 PDF members
```

Sum Rules

```
In [8]: hquadrature((x) -> (pdf["UP"](x, 1.75) - pdf["AUP"](x, 1.75)) / x, 1e-9, 1, abstol = 1e-8)
Out[8]: (1.9999955060750039, 1.9975442513203155e-8)
```

…more fundamentally, how are the PDFs obtained?

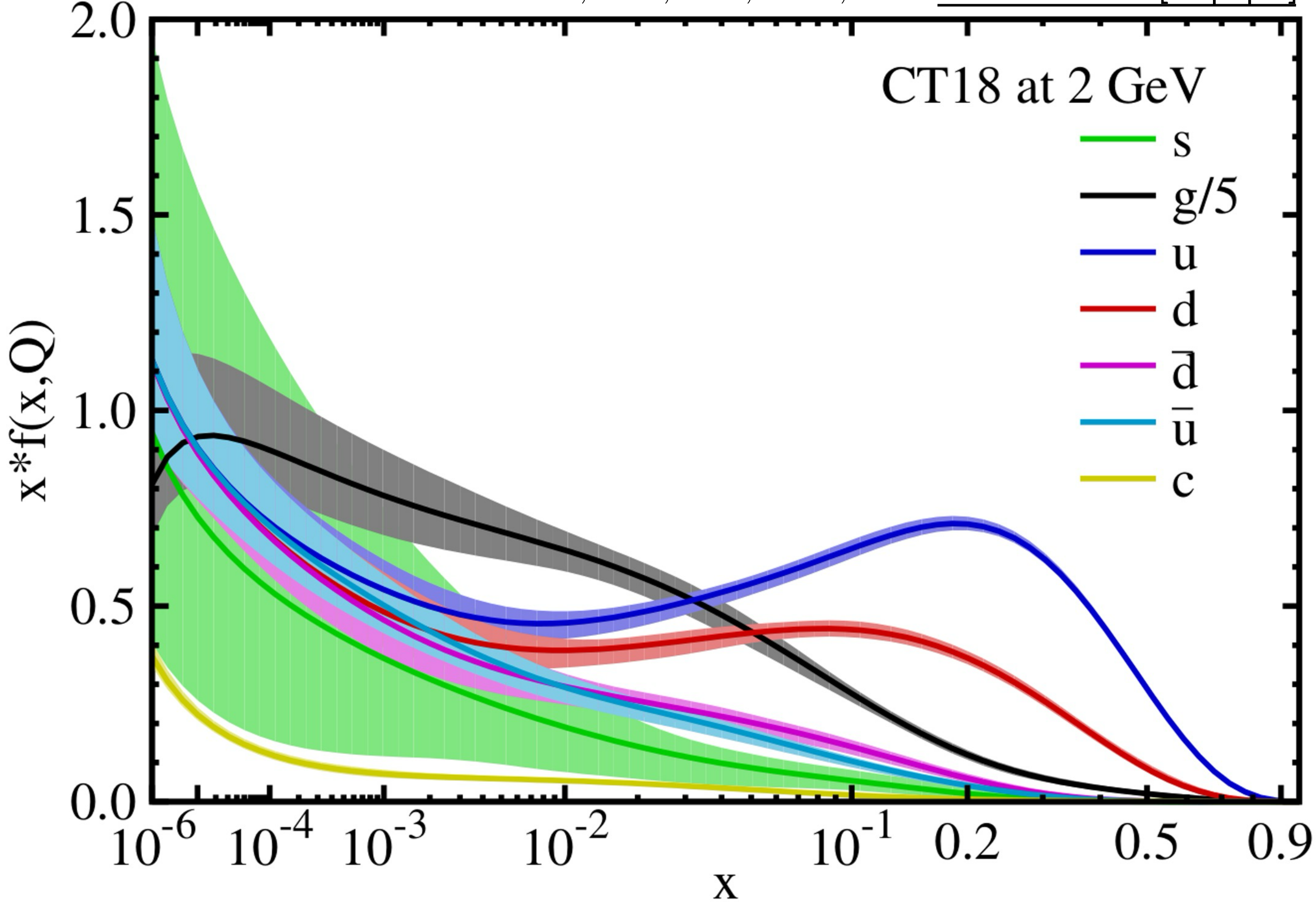
require
extensive data



a modern NNLO global analysis: CT18

→ **determining PDFs is now a sophisticated endeavor with a complicated interplay of theory, computation, and statistical methods**

CT18 → Hou, Gao, TJH, et al., arxiv:1912.10053 [hep-ph].

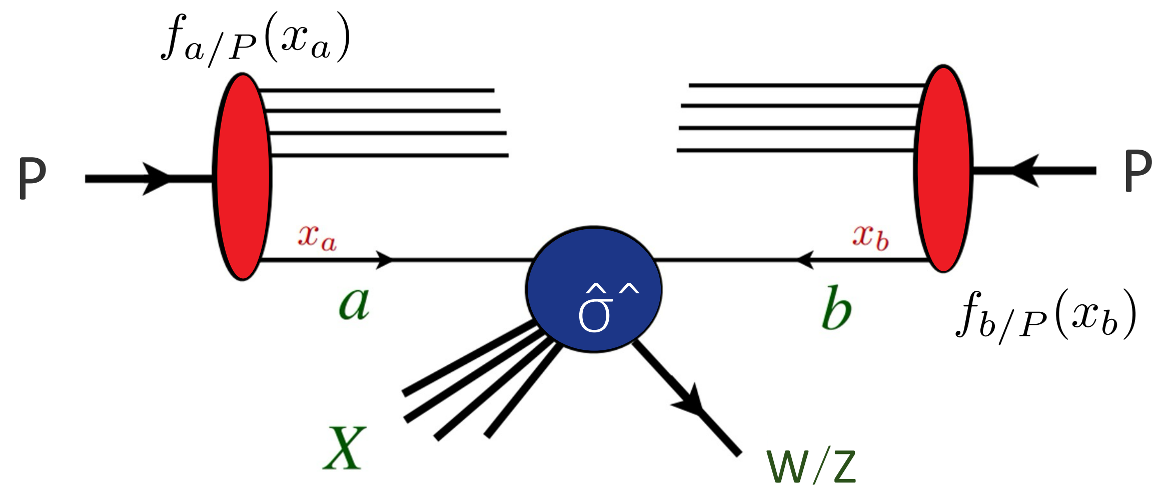


fitted PDFs give picture of proton's 1-dimensional (collinear) structure

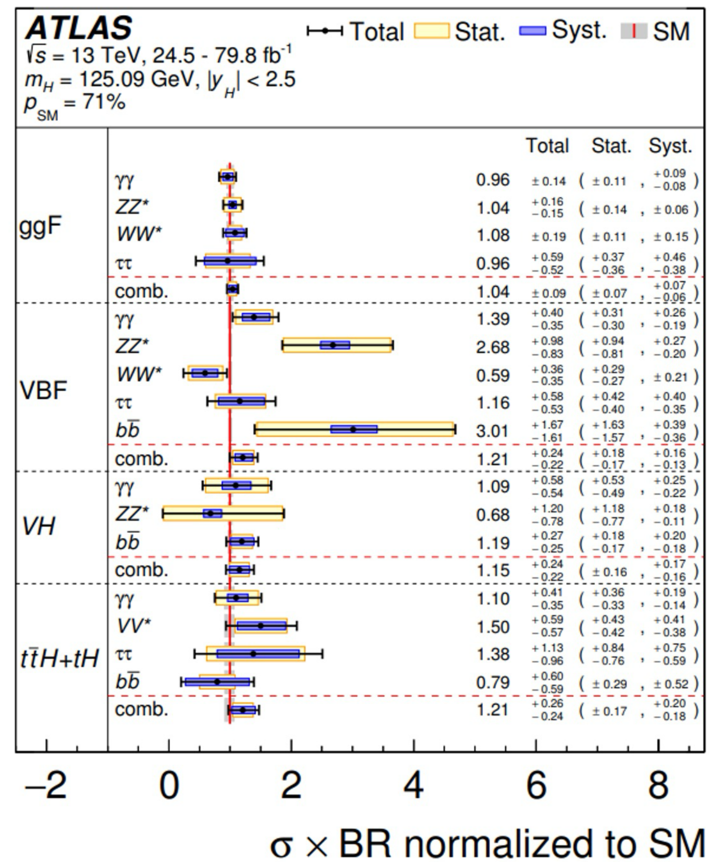
→ prob. to find gluon/quark with **x**-fraction of proton's momentum

HEP measurements at the LHC depend on PDFs!

→ theory predictions for pp collisions require knowledge of parton densities



limit to precision in Higgs; electroweak; QCD (top, jet production); BSM searches; …



more tomorrow…

PDFs are fundamental to our understanding of QCD

→ in addition to practical importance to HEP phenomenology, PDFs quantify aspects of (non)perturbative QCD

I will briefly survey a sampling of these different issues:

→ origin of mass

→ disentangling the nucleon's flavor structure

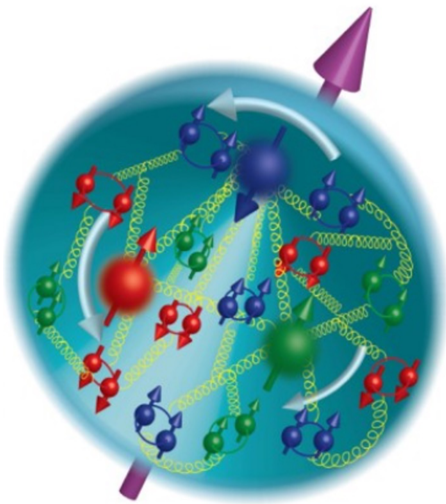
flavor-symmetry violation

strange content

nonperturbative charm

→ origin of spin

→ nature of the nuclear medium



...dig into the guts of this picture

n.b., these are all banner motivations for the EIC

the Electron-Ion Collider (EIC): a QCD machine; more tomorrow

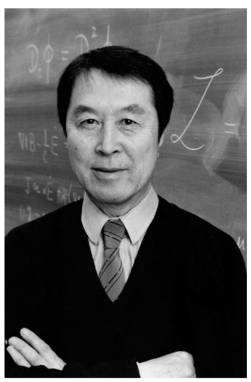
CD-3A: estimated construction cost: $1.7-2.8B; this decade

the only new collider planned in the US for the next half-century

a centerpiece of nuclear-particle physics for the next few decades

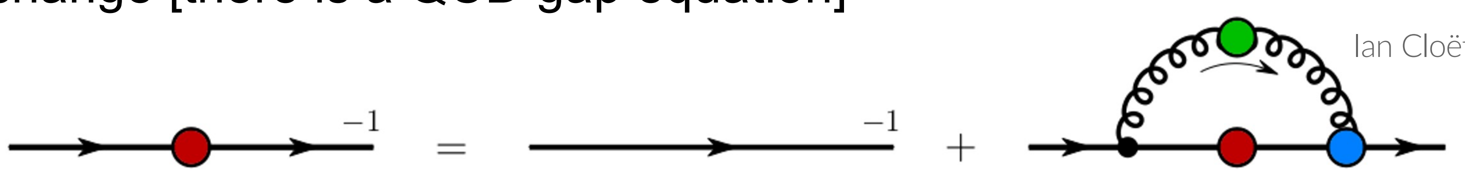
why do you weigh something, rather than (almost) nothing?

→ Higgs mechanism accounts for little of the mass of the visible universe



Y. Nambu

bound quarks develop an effective mass from the energy of gluon exchange [there is a QCD gap equation]



→ nucleon mass comes from quark-gluon energy/motion:

$$M_p = E_q + E_g + \chi_{m_q} + T_g$$

related to moments of quark PDFs, $\langle x \rangle_{q,g} = \int_0^1 dx\, x f_{q,g}(x)$

quark sea flavor-symmetry breaking

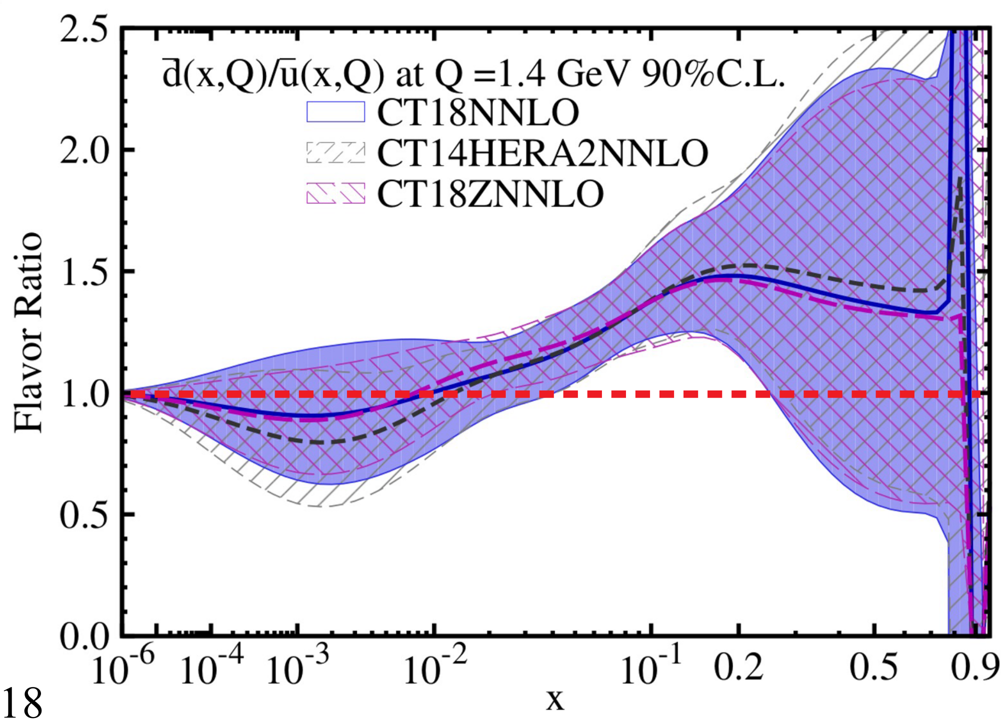
high-energy QCD: $\bar{d} = \bar{u}$ [symmetric sea]

nonperturbative dynamics at low energy (*e.g.*, pion exchange) can break this symmetry

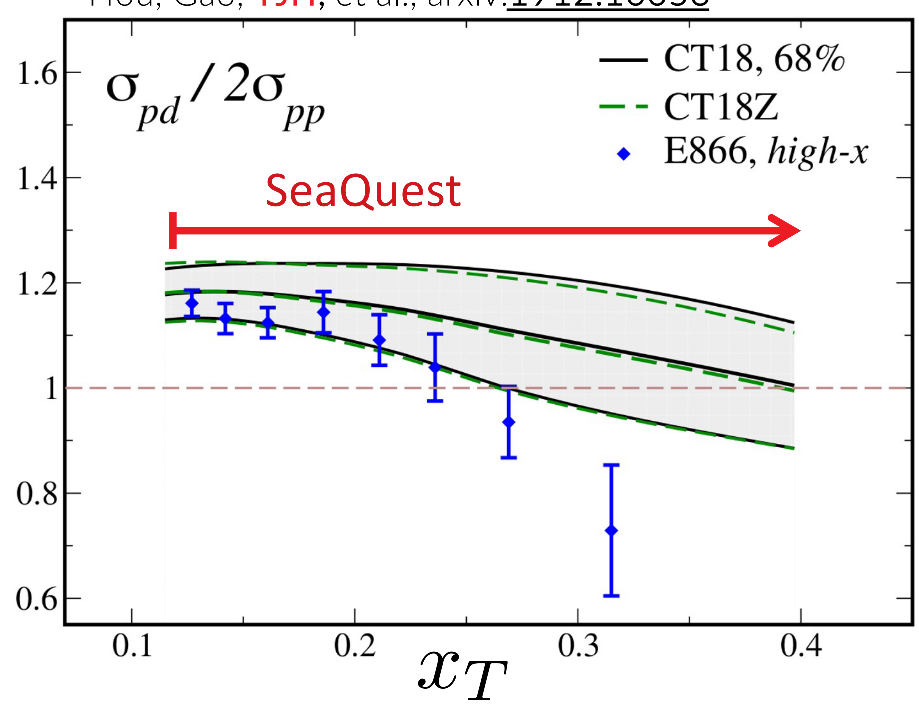
$$p[uud] \to \pi^+[u\bar{d}] + n[udd] \longrightarrow \bar{d} > \bar{u}$$

E866/SeaQuest: measure fixed-target proton-deuteron Drell-Yan

$$\frac{\sigma^{pd}}{2\sigma^{pp}}\Big|_{\mathrm{LO},x_p \gg x_T} \sim \frac{1}{2}\left\{1 + \frac{\bar{d}(x_T)}{\bar{u}(x_T)}\right\}$$



Hou, Gao, TJH, et al., arxiv:1912.10053



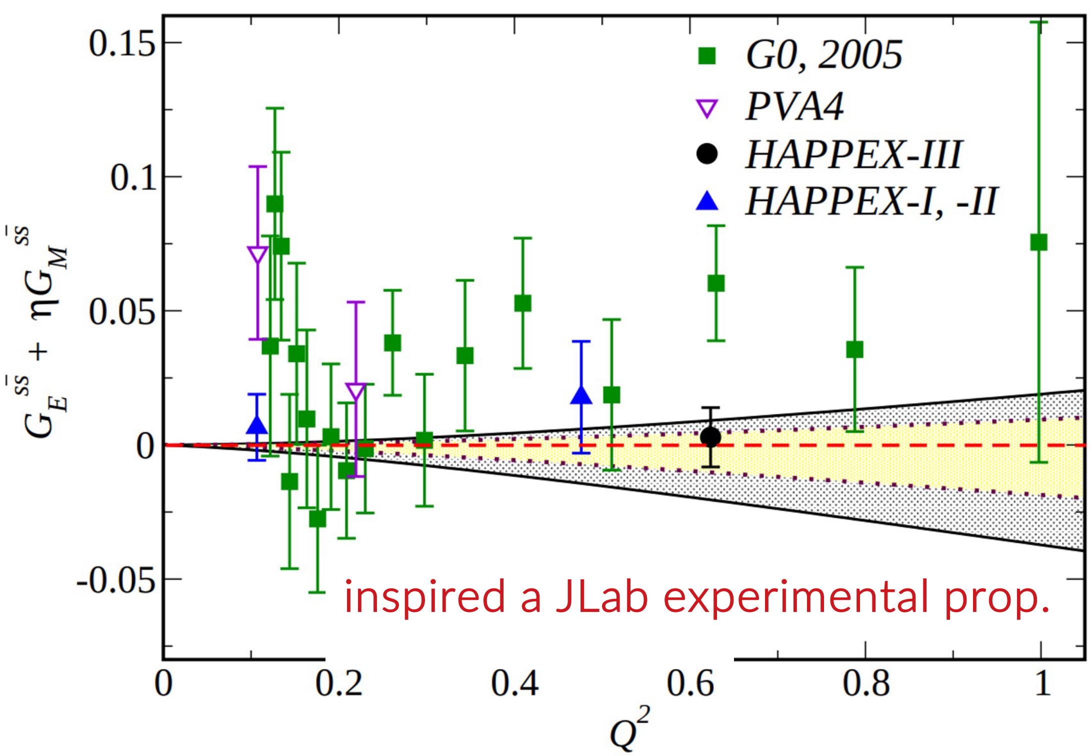
the proton’s quark sea remains poorly determined

→ *e.g.*, the size/shape of strange quark PDF (or, <u>form factor</u>)?

theoretical models can compute strange portion of a related quantity:
the electromagnetic form factor

$$|\Psi_P^\lambda(P^+,\mathbf{P}_\perp)\rangle = \frac{1}{16\pi^3}\sum_{q=s,\bar{s}}\int\frac{dx\,d^2\mathbf{k}_\perp}{\sqrt{x(1-x)}}\,\psi^\lambda_{q\lambda_q}(x,\mathbf{k}_\perp)\;|q;xP^+,x\mathbf{P}_\perp+\mathbf{k}_\perp\rangle$$

TJH, Alberg, Miller; PRC91 (2015) 3, 035205.



...data is not yet precise enough; **PDF-sensitive experiments (e.g., DIS)?**

is there "intrinsic" charm in the proton (?)

→ must resolve for percent-level EW precision; vital for charge-current processes

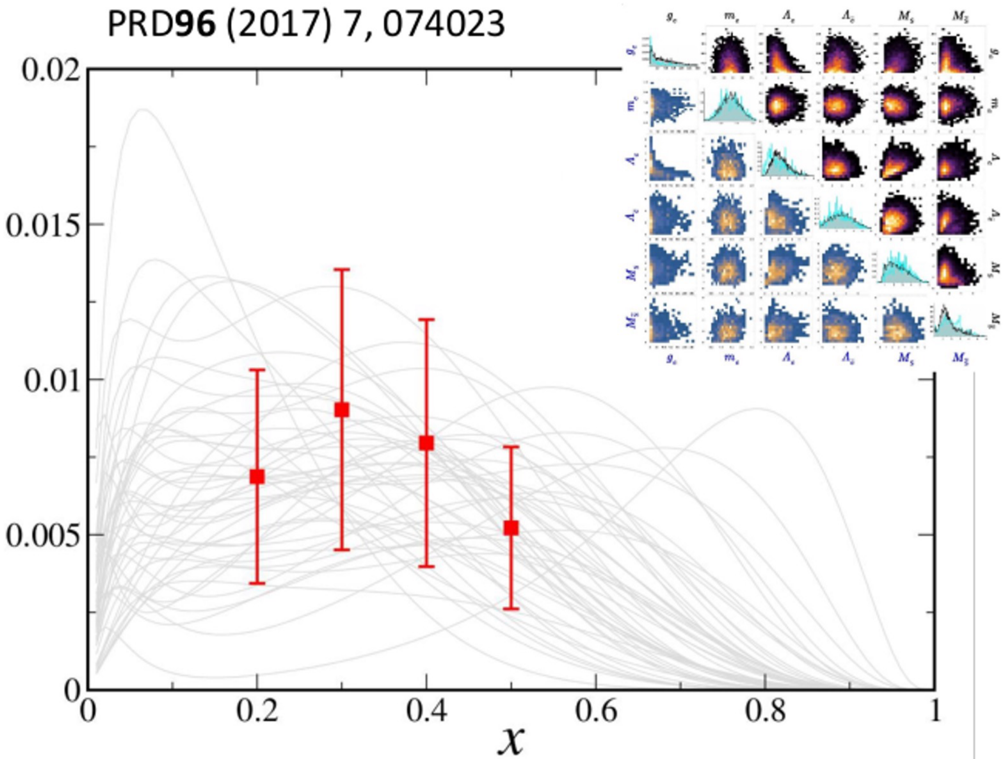
ultimately, question of **model inference, selection**

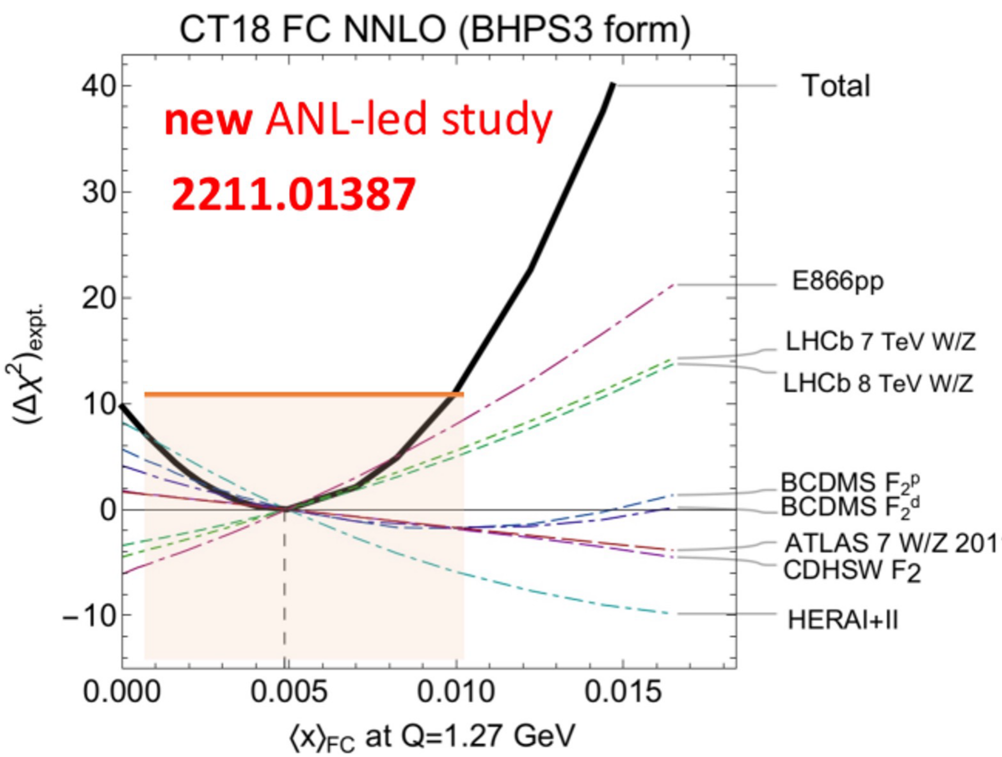
no clear signal: errors remain large; uncertainty quantification essential

$$0 \lesssim \langle x \rangle_{\mathrm{FC}} \lesssim 1\%$$

- HEP data have opposing pulls on 'fitted charm' (FC) PDF

MCMC model exploration: TJH et al., PRD**96** (2017) 7, 074023



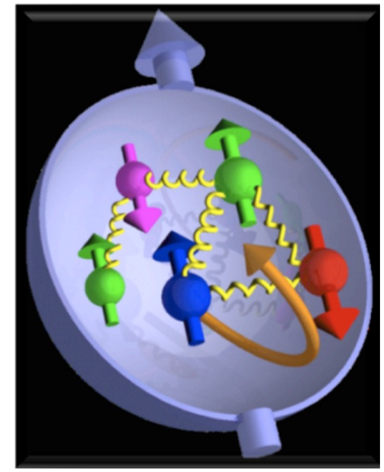


the origin of nucleon spin

→ can we come up with the spin of the proton from quarks/gluons?

$$\text{spin}_{\text{quark}} = \frac{1}{2} \qquad \text{spin}_{\text{gluon}} = 1$$

TJH et al., PRC95 (2017), 3 035205



→ must somehow combine for total spin-½ of proton:

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + (L_q + L_g)$$

quark spin | gluon spin | orbital angular momentum

activity at RHIC, JLab, ...

TJH et al., PRC93 (2016) 5, 052801

$$\langle P|\gamma^+|P\rangle \;\rightarrow\; \langle P|\gamma^+\gamma_5|P\rangle$$

unpolarized PDF | spin PDF

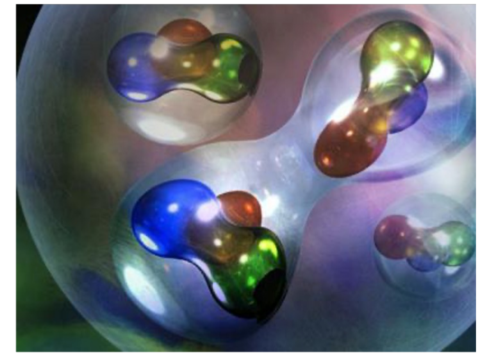
→ surprising connection:

how strange is the proton spin? Affects core-collapse supernovae...

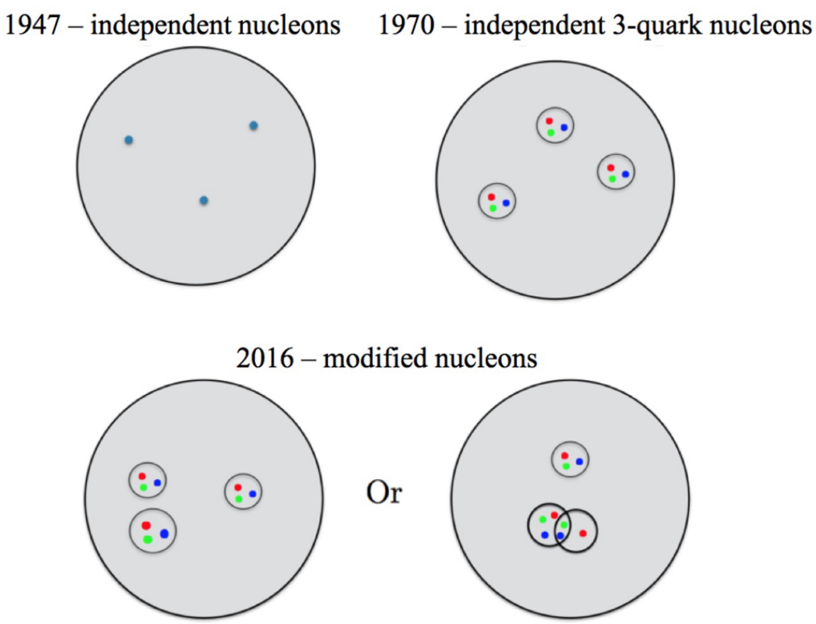
how does QCD generate nuclear structure?

or: what happens to a nucleon once embedded inside a nucleus? does it...

- → do nothing? [this we can now exclude...]
- → 'dissolve' (uniform mean field)?
- → form short-lived correlations?
- → remain distinct, but with modified structure?

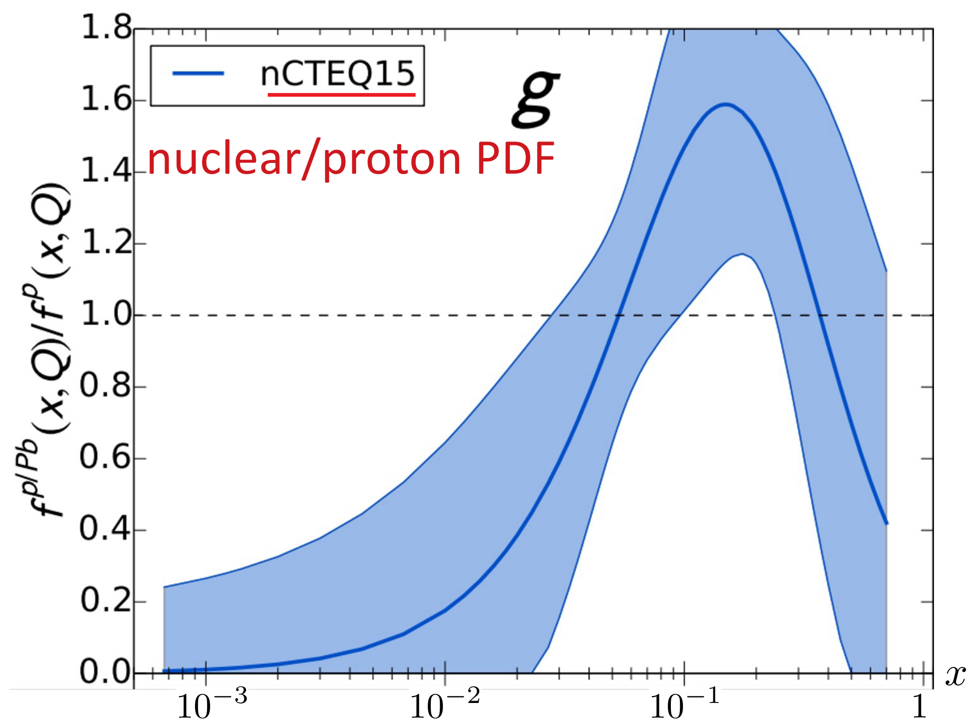


studied in nuclear CTEQ (nCTEQ) effort:



Hen et al., Rev. Mod. Phys. **89** (2017) no.4, 045002

most recently: Kusina, ..., TJH, et al., 2007.09100

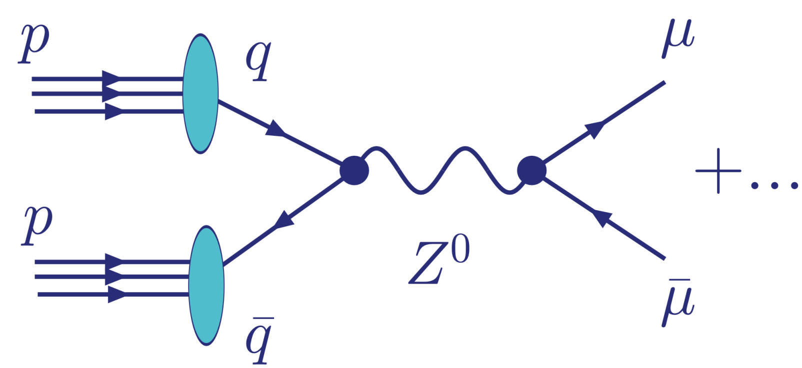


we return to ask: how can PDFs be accessed systematically?

PDFs encode long-distance dynamics; must be separated from pQCD physics

$$f_{q/p}(x,\mu^2) = \int \frac{d\xi^-}{4\pi} e^{-i\xi^- k^+} \langle p \left| \overline{\psi}(\xi^-) \gamma^+ \mathcal{U}(\xi^-,0) \psi(0) \right| p \rangle$$

→ QCD factorization theorem; here, for Drell-Yan processes (e.g., LHC):



$$\sigma_{pp\to\ell\bar{\ell}X} = \sum_{a,b=q,\bar{q},g} \int_0^1 d\xi_1 \int_0^1 d\xi_2\, \widehat{\sigma}_{ab\to Z\to\ell\bar{\ell}} \left(\frac{x_1}{\xi_1}, \frac{x_2}{\xi_2}; \frac{Q}{\mu}\right) f_{a/p}(\xi_1,\mu^2) f_{b/p}(\xi_2,\mu^2)$$
$$+ \mathcal{O}\left(\Lambda^2_{QCD}/Q^2\right)$$

⟶ QCD fact./reg. subtracts large, $\sim \alpha_s^n \ln^k(Q^2/m_q^2)$, logarithms from $\hat{\sigma}$

resums these into the PDFs

kinematically matching theory to experiment

- theoretical predictions evaluated according to leading-order ("Born-level") matchings with external scales in measurements

DIS: $\mu_i \approx Q|_i\,,\ x_i \approx x_B|_i$

x_i : parton mom. fraction

μ_i : factorization scale

hadron-hadron collisions:

$$AB \to CX \qquad \mu_i \approx Q|_i\,,\ x_i^{\pm} \approx \frac{Q}{\sqrt{s}}\exp(\pm y_C)\Big|_i$$

single-inclusive jet production: $Q = 2p_{Tj},\ y_C = y_j$

$t\bar{t}$ pair production: $Q = m_{t\bar{t}},\ y_C = y_{t\bar{t}}$

$d\sigma/dp_T^Z$ measurements: $Q = \sqrt{(p_T^Z)^2 + (M_Z)^2},\ y_C = y_Z$

- higher-order corrections responsible for inherent scale uncertainty

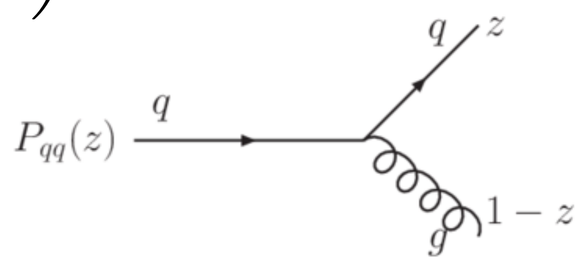
PDFs and scale dependence: Altarelli-Parisi evolution

- high-energy experiments measure at various kinematic *scales*

  → solve self-coupled integrodifferential equations for momentum dependence

$$\mu \frac{df_{i/p}(x,\mu)}{d\mu} = \sum_{j=g,u,\bar{u},d,\bar{d},\ldots} \int_x^1 \frac{dy}{y} P_{i/j}\left(\frac{x}{y}, \alpha_s(\mu)\right) f_{j/p}(y,\mu)$$

- evolution governed by QCD radiation



→ $P_{i/j}$ are probabilities for $j \to ik$ collinear splittings;

... *e.g.*, at leading-order, $P_{i/j}(z) = \delta_{ij} C_F \left(\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z)\right)$

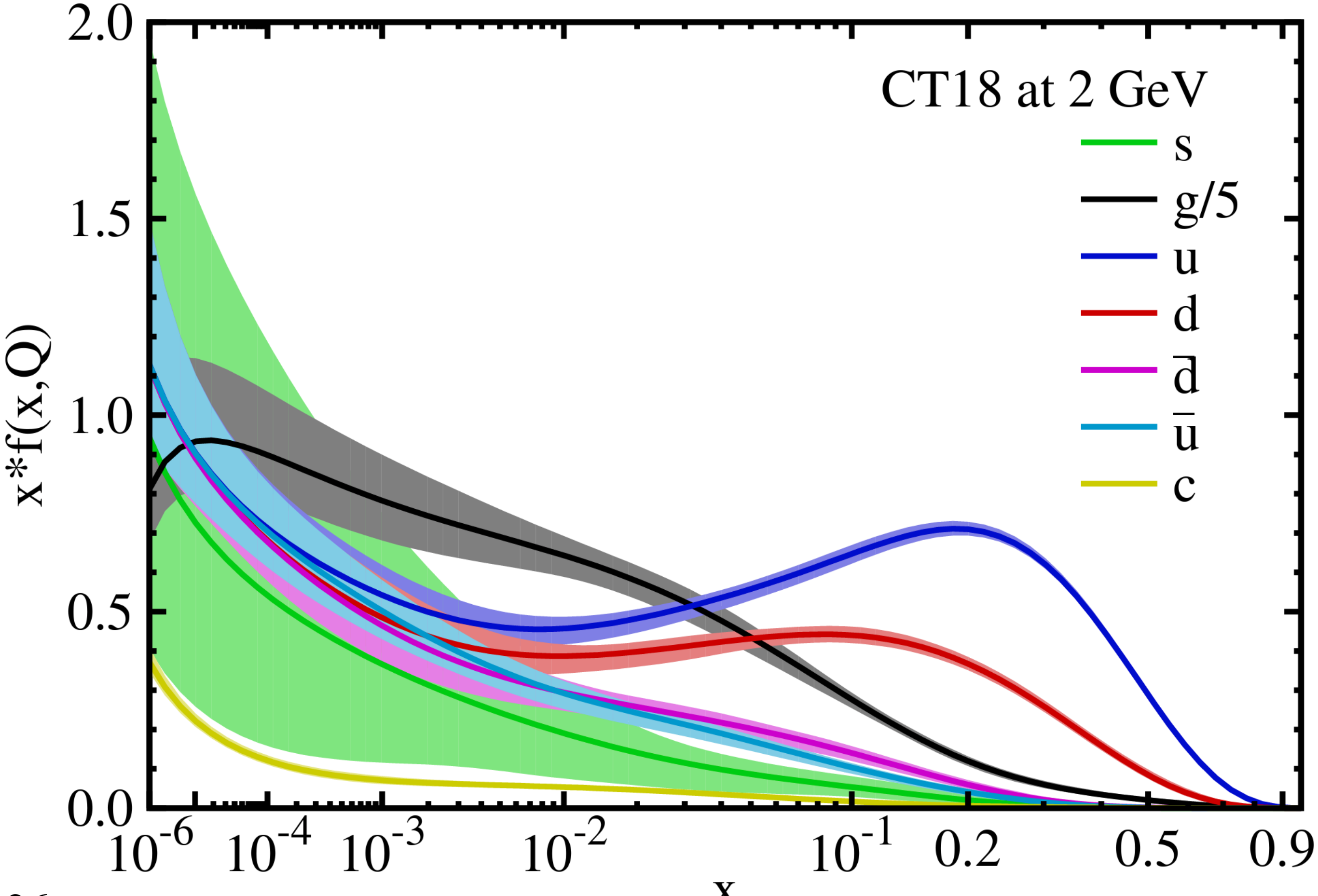
known to $\mathcal{O}(\alpha_s^3)$ [NNLO]:

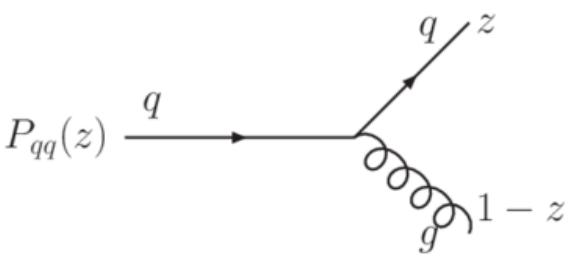
$$P_{i/j}(z,\alpha_s) = \alpha_s P^{(1)}_{i/j}(z) + \alpha_s^2 P^{(2)}_{i/j}(z) + \alpha_s^3 P^{(3)}_{i/j}(z) + \ldots$$

QCD evolution: strong scale dependence in PDF shapes

- at low scale(s), u- and d-PDF valence-like with a high-x peak around x ~ 1/3

  - nonperturbative shapes evident for light-quark sea and gluon distributions



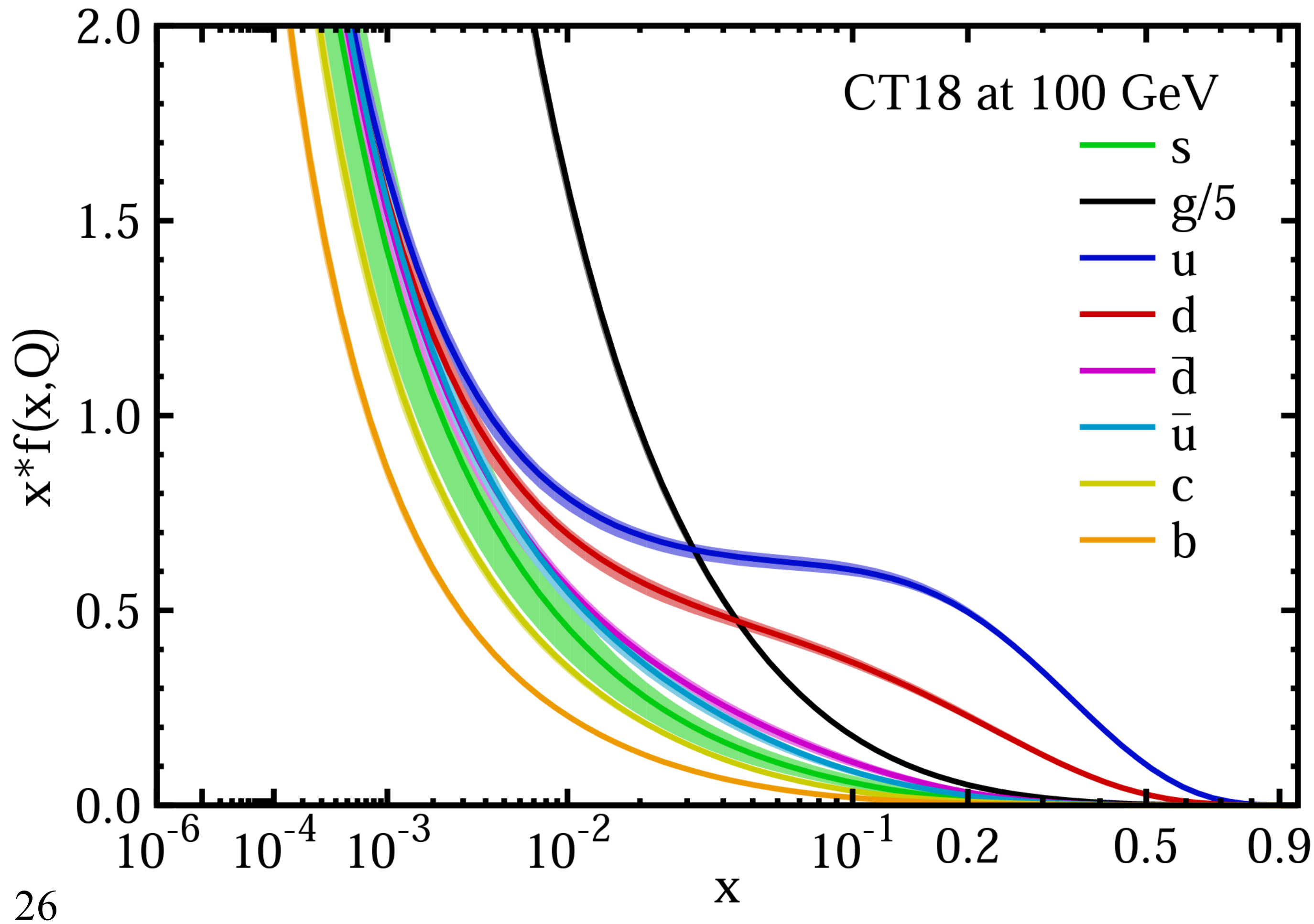


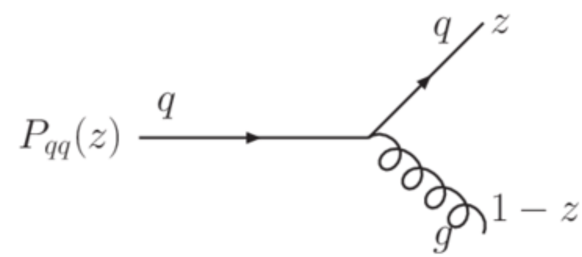
- perturbative evolution generally shifts high-x features in PDFs to low x with increasing scale

QCD evolution: strong scale dependence in PDF shapes

- evolution to higher scale(s); DGLAP coupling to gluon drives singular low-x behavior

  - charm, bottom PDFs generated entirely radiatively; closely track the gluon!





  - note the challenge of separating the light-quark sea

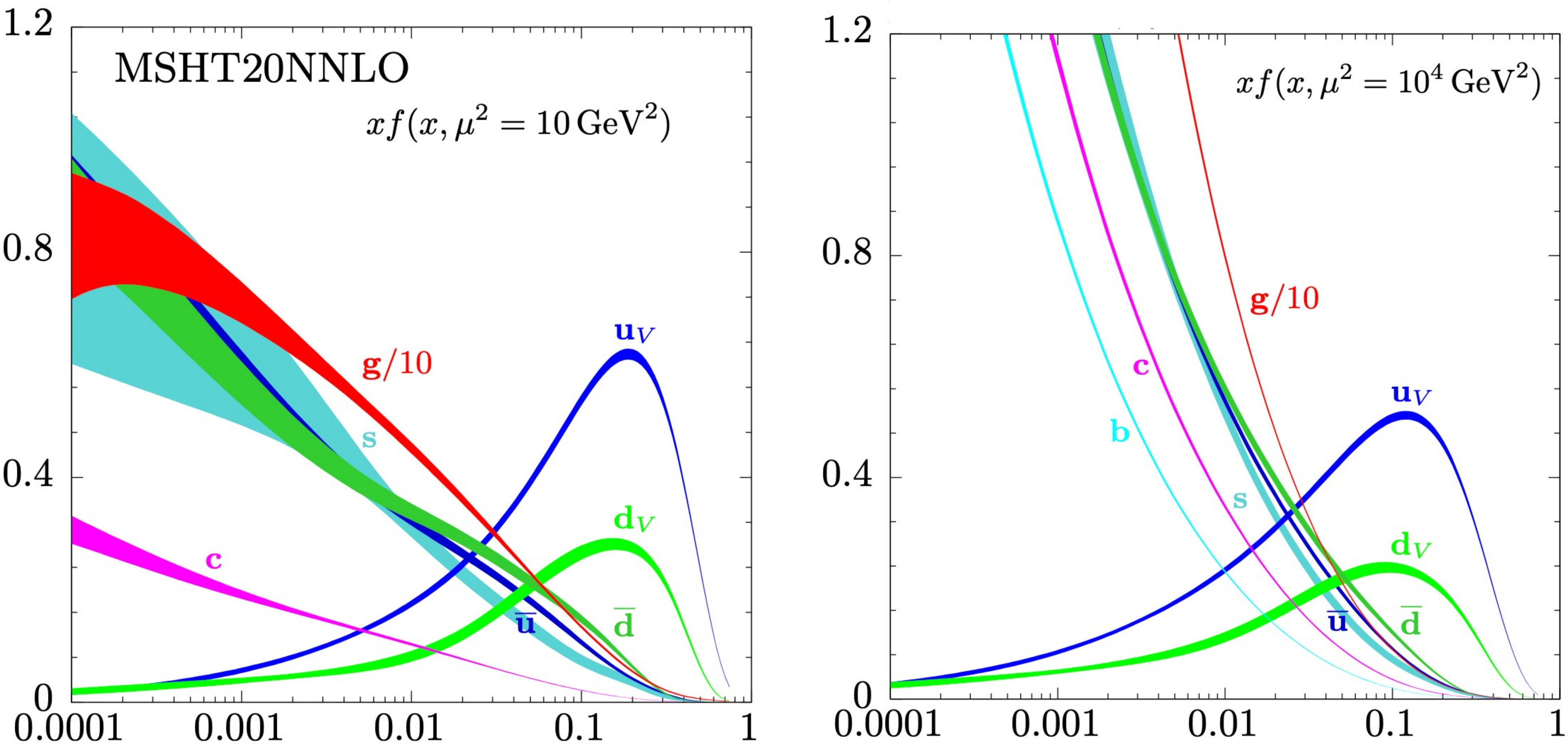
scale dependence: interplay between valence and sea

$$q = q_V + 2\, q_{sea}$$

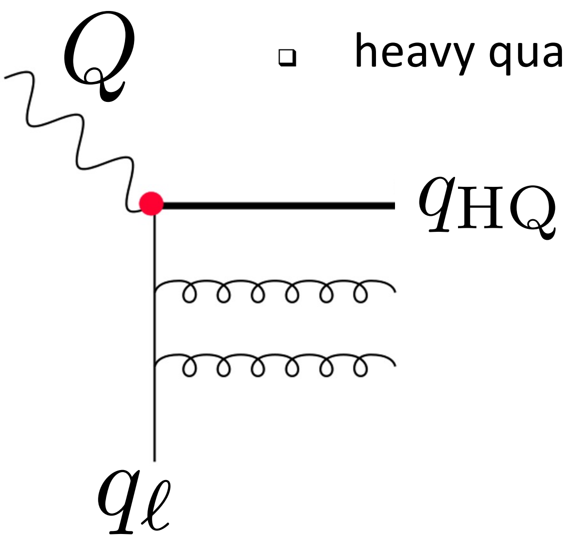
$$q_V = q^-(x) = q(x) - \bar{q}(x)$$

$$q_{sea} = \bar{q}(x)$$

- low-*x* PDFs sea-dominated
- valence evolves slowly; ultimately swamped by sea at large scales

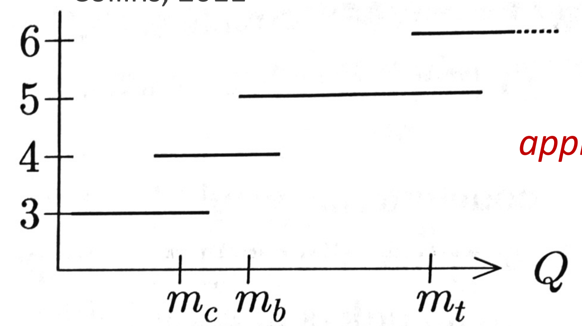


QCD analyses traverse many scales, Q; variable flavor

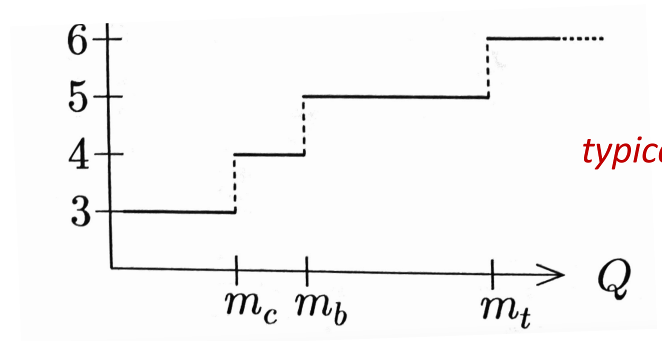


- heavy quarks (HQ) introduce additional effective scale into hard process

Collins, 2011



appropriate transition scales



typical choices (ACOT)

$$Q \gg m_{\rm HQ}$$

- perturbatively-generated HQ PDFs resum large logs,

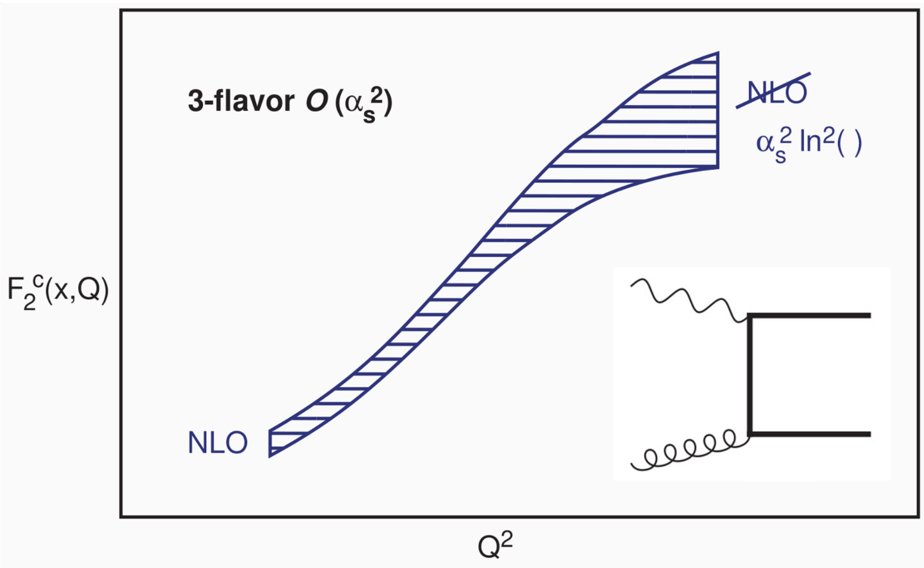
$$\sim \ln(Q^2/m^2_{\rm HQ})$$

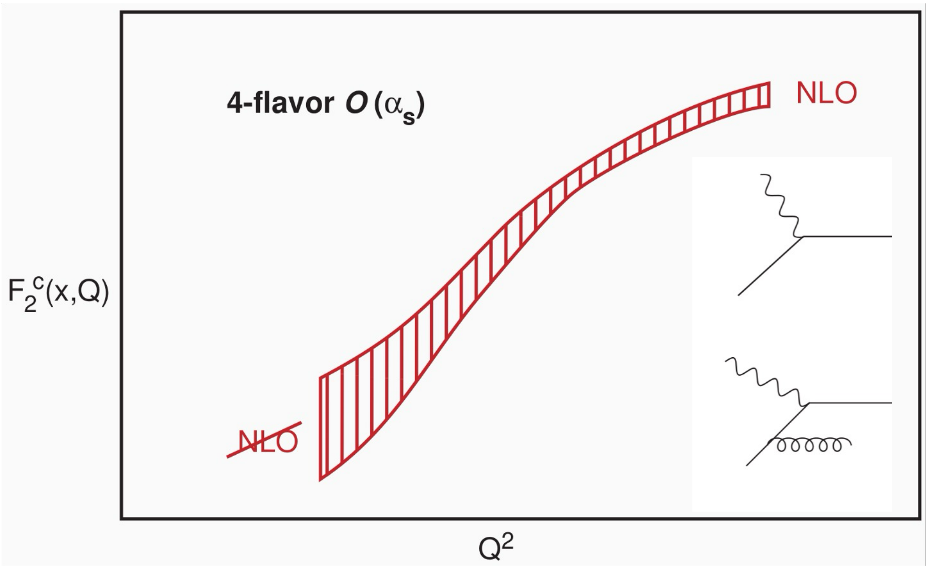
→ number of active parton flavors is a scheme-dependent choice

- PDF fits typically assume a *variable flavor number* scheme with assumed number of active flavors; usually $n_F = 5$

heavy quarks in evolution schemes and higher-order QCD

- higher order(s): improved accuracy in Wilson coeff., control over scale dependence
- at given fixed-order, nontrivial relationship with chosen HQ scheme





→ fixed flavor-number (FFN): $Q \gtrsim M_Q$; flavor-creation (FC) processes with $n_f = 3$

→ zero-mass (ZM) variable flavor-number: $Q \gg M_Q$; flavor-excitation (FE) processes with $n_f = 4$

2 paradigms for different regimes wrt HQ mass scale; **∃ interpolation scheme**?

parametrizing the PDFs: flexibility

- noted earlier: fit *flexible* parametrization at evolution boundary

$$f_{q/p}(x,\mu^2=Q_0^2) = a_{q_0} x^{a_{q_1}} (1-x)^{a_{q_2}}\, \mathrm{P}[x;\{a_{q_{n-3}}\}]$$

  - how much flexibility is needed?

    modern data are sensitive to:

    $d,\ \bar{u},\ g,\ u,\ d,\ s(\neq \bar{s})$; perhaps c

  - in CTEQ family, ~30 parameters

- alternative approach(es): parametrize *x* dependence with neural net

$$f_{q/p}(x,\mu^2=Q_0^2) = a_{q_0} x^{a_{q_1}} (1-x)^{a_{q_2}}\, \mathrm{NN}(x)$$

  - NNPDF: multi-layer perceptron
  - ~1000-member MC replica sets

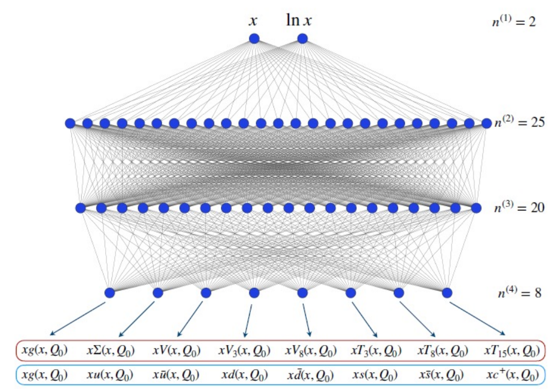


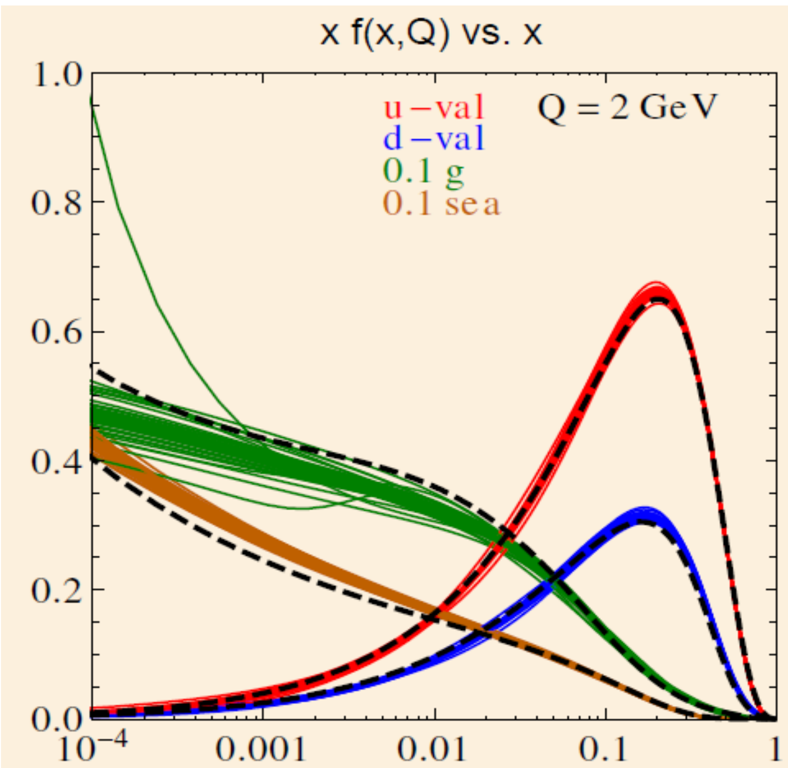
Figure 3.9. The neural network architecture adopted for NNPDF4.0. A single network is used, whose eight output values are the PDFs in the evolution (red) or the flavor basis (blue box). The architecture displayed corresponds to the optimal choice in the evolution basis; the optimal architecture in the flavor basis is different as indicated by Table 3.3).

modern approaches to PDF uncertainty quantification

- Two powerful, complementary representations

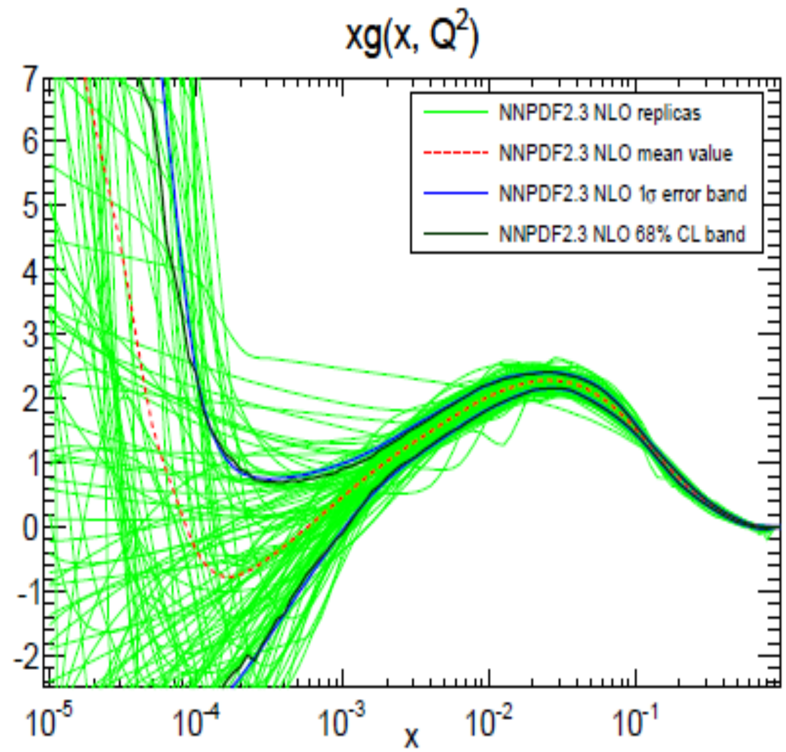
Analytic parametrizations +

Hessian PDF eigenvector sets
(ABM, CTEQ, HERA, MMHT,…)



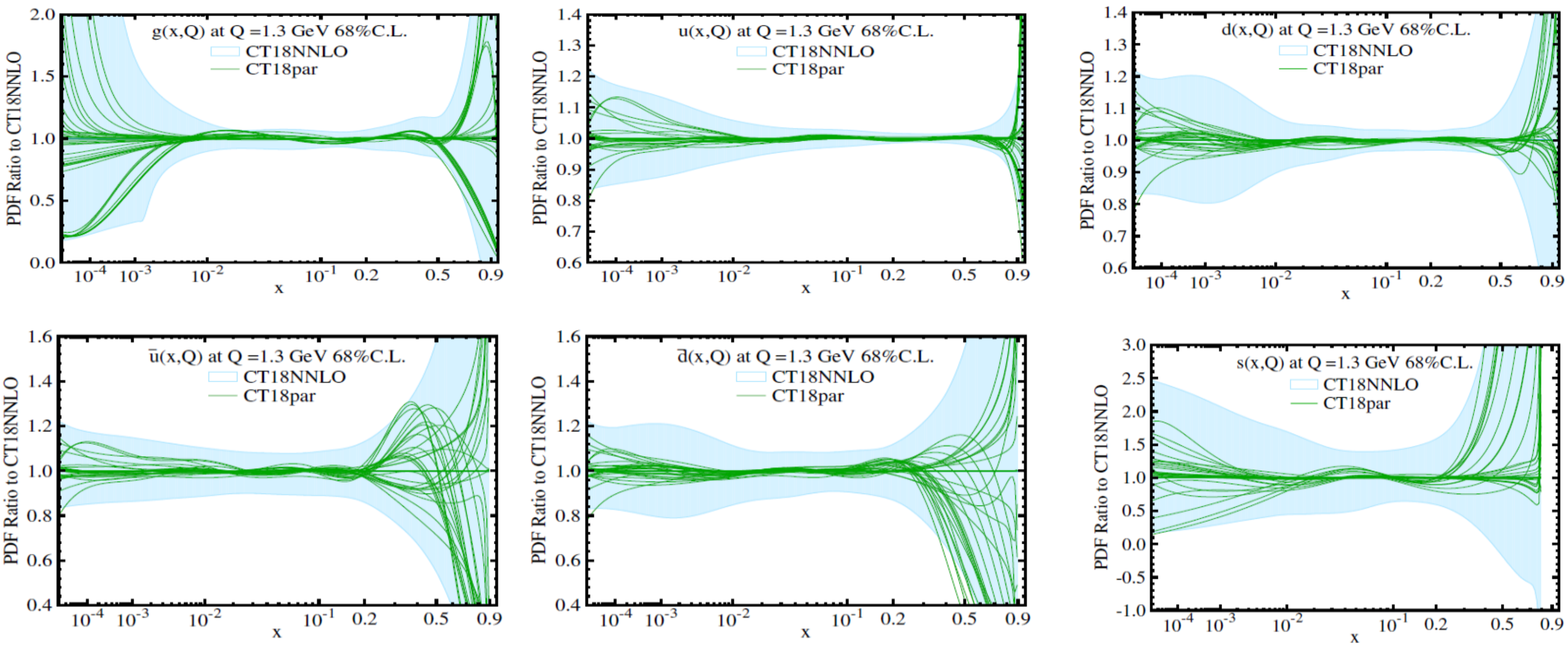
Neural network parameterizations +

Monte Carlo PDF replicas
(NNPDF)



nb: Hessian PDFs can be converted into MC ones, and *vice versa*

250+ candidate nonperturbative parametrization forms of CT18 PDFs



CT18par – sample of **some** nonperturbative forms tried in CT18

No data constrain very large-x or very small-x regions

sources of PDF uncertainty

Kovarik et al., arXiv: 1905.06957

1. **Experimental uncertainties**, e.g., statistical, correlated and uncorrelated systematic uncertainties of each experimental data set;

2. **Theoretical uncertainties** due to the absent radiative contributions, approximations in parton showering simulations

3. **Parameterization uncertainties** associated with the choice of the PDF functional form or AI/ML replica training algorithm
   - contribute at least a half of the CT18 total PDF uncertainty

★ [previous slide]

4. **Methodological uncertainties** associated with the selection of experimental data sets, fitting procedures, and goodness-of-fit criteria.

The uncertainty of published CT18 PDFs estimates the sum of four contributions

→ **in practice, each of these may contribute to tensions among fitted data sets**

statistical interlude

- PDF uncertainties encapsulate behavior of likelihood function about global minimum

$$\chi^2_E(\vec{a}) = \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}) + \sum_{\alpha=1}^{N_\lambda} \overline{\lambda}_\alpha^2(\vec{a})$$

nuisance parameters to handle correlated errors

$$r_i(\vec{a}) = \frac{1}{s_i}\left(T_i(\vec{a}) - D_{i,sh}(\vec{a})\right)$$

- these result in systematic shifts to data central values: $D_i \to D_{i,sh}(\vec{a}) = D_i - \sum_{\alpha=1}^{N_\lambda} \beta_{i\alpha}\overline{\lambda}_\alpha(\vec{a})$

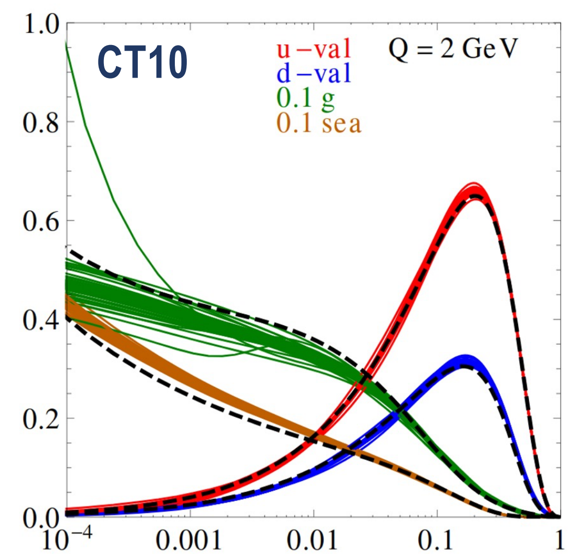
56-dimensional parametric basis $\vec{a}$ obtained by diagonalizing the **Hessian** matrix H determined from χ^2

(following a 28-parameter fit)

use this basis to compute 56-component "normalized" residuals :

$$\delta^{\pm}_{i,l} \equiv \left(r_i(\vec{a}^{\pm}_l) - r_i(\vec{a}_0)\right) / \langle r_0 \rangle_E$$

where $\langle r_0 \rangle_E \equiv \sqrt{\frac{1}{N_{pt}} \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}_0)}$

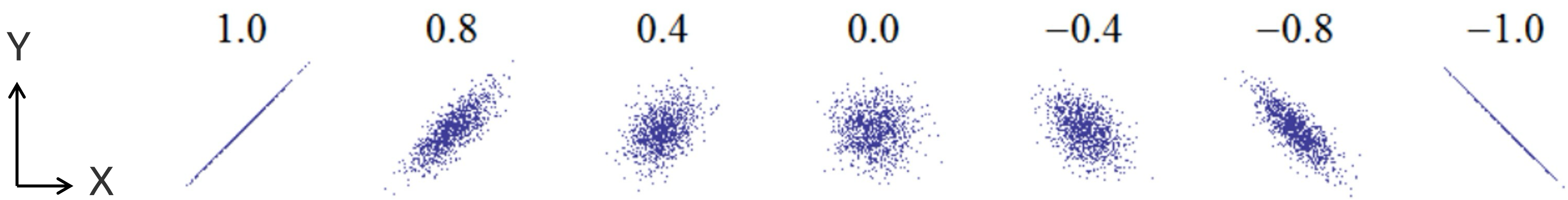


PDF correlations (Hessian formalism)

how does behavior of residuals relate to fitted PDFs and uncertainties?

- e.g., how does PDF uncertainty (at specific x, μ) correlate with residual associated with a theoretical prediction at similar x, μ?

examine Pearson correlation over 56-member Hessian set between PDF of given flavor and (another PDF; residual; cross section; …)



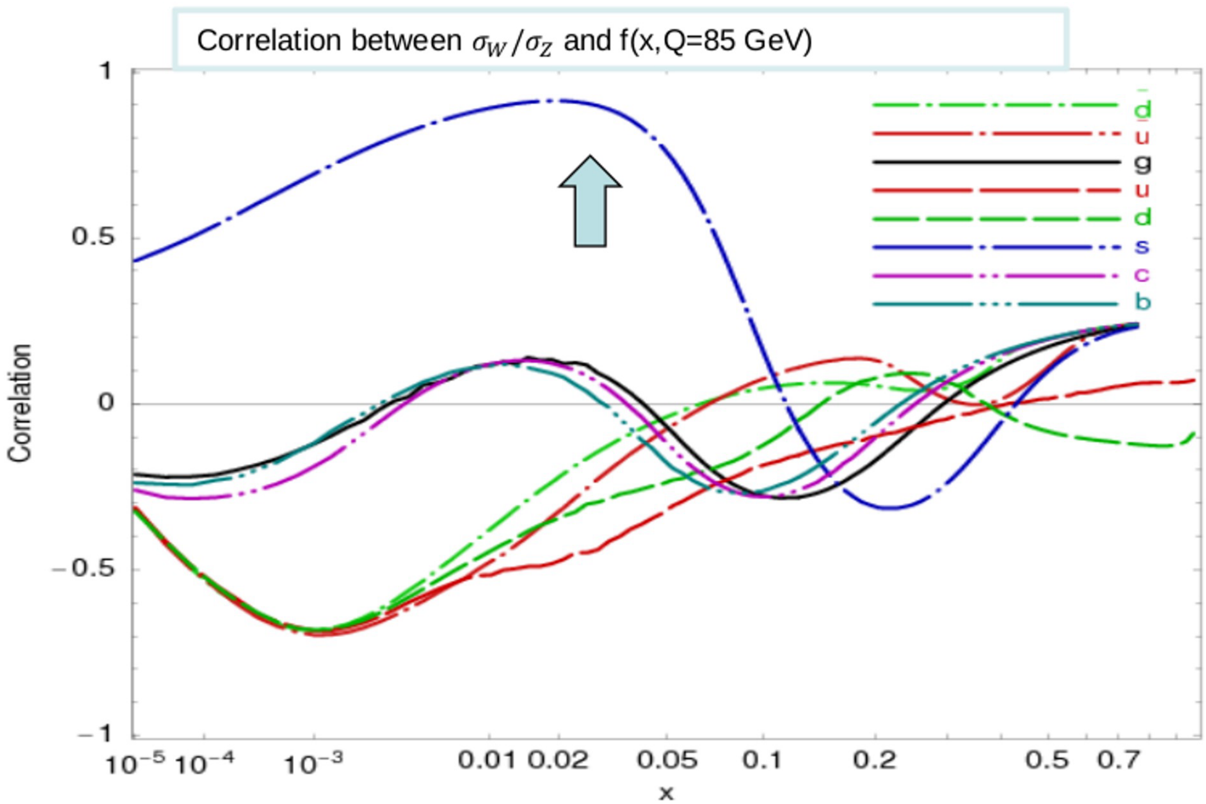
[X,Y] are exactly (anti-)correlated at the far (right) left above.

$$\mathrm{Corr}[X,Y] = \frac{1}{4\Delta X \Delta Y} \sum_{j=1}^{N} (X_j^+ - X_j^-)(Y_j^+ - Y_j^-) \qquad \Delta X = \frac{1}{2}\sqrt{\sum_{j=1}^{N} (X_j^+ - X_j^-)^2}$$

evaluate over n Hessian eigendirections

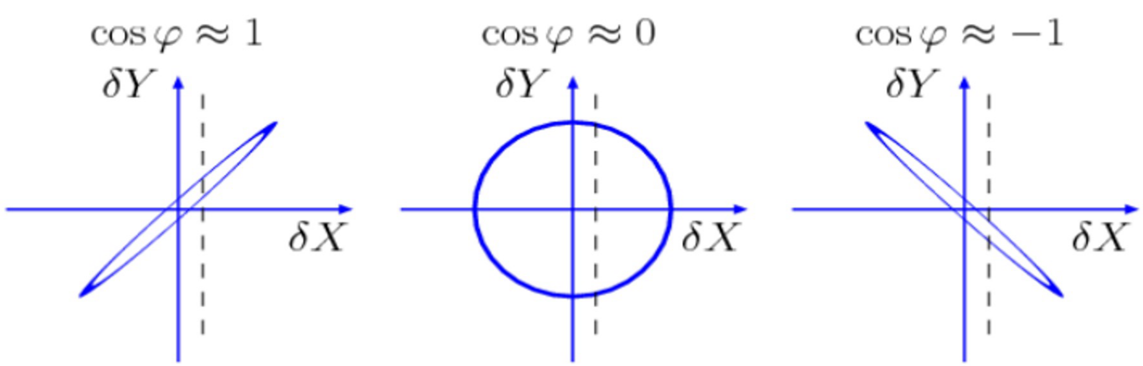
analogous for Monte Carlo…

correlations carry useful information



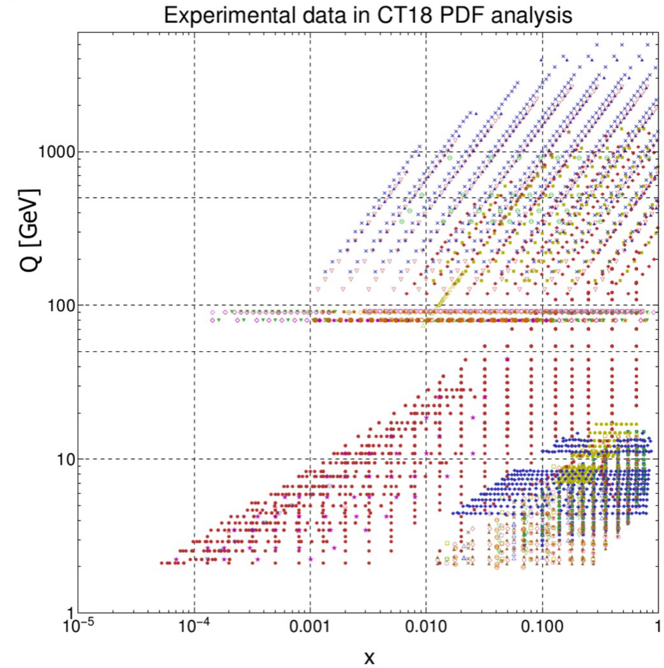
CTEQ6.6 [arXiv:0802.0007]:
$\cos\varphi > 0.7$ shows that the ratio σ_W/σ_Z at the LHC must be sensitive to the strange PDF $s(x,Q)$

$\cos\varphi \approx \pm 1$ suggests that a measurement of X **may** impose tight constraints on Y



But, Corr[X,Y] between **theory** cross sections X and Y does not tell us about **experimental** uncertainties

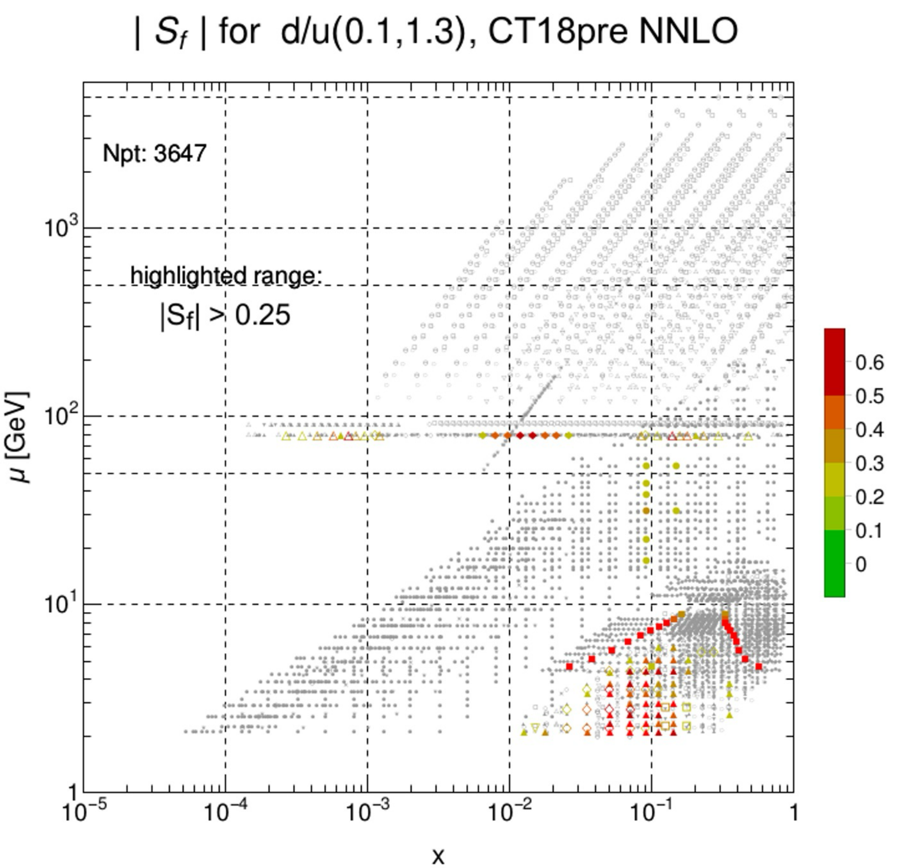
Process	Subprocess	Partons	x range
$\ell^\pm \{p,n\} \to \ell^\pm X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
$\ell^\pm n/p \to \ell^\pm X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
$pp \to \mu^+\mu^- X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
$pn/pp \to \mu^+\mu^- X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
$\nu(\bar{\nu}) N \to \mu^-(\mu^+) X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
$\nu N \to \mu^-\mu^+ X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
$\bar{\nu} N \to \mu^+\mu^- X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
$e^\pm p \to e^\pm X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$10^{-4} \lesssim x \lesssim 0.1$
$e^+ p \to \bar{\nu} X$	$W^+ \{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
$e^\pm p \to e^\pm c\bar{c}X, e^\pm b\bar{b}X$	$\gamma^* c \to c,\ \gamma^* g \to c\bar{c}$	c, b, g	$10^{-4} \lesssim x \lesssim 0.01$
$e^\pm p \to$ jet$+X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
$p\bar{p}, pp \to$ jet(dijet)$+X$	$gg, qg, qq \to 2j$	g, q	$0.00005 \lesssim x \lesssim 0.5$
$p\bar{p} \to (W^\pm \to \ell^\pm \nu) X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, s, \bar{u}, \bar{d}, \bar{s}$	$x \gtrsim 0.05$
$pp \to (W^\pm \to \ell^\pm \nu) X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, s, \bar{u}, \bar{d}, \bar{s}, g$	$x \gtrsim 0.001$
$p\bar{p}(pp) \to (Z \to \ell^+\ell^-)X$	$uu, dd, ..(u\bar{u}, ..) \to Z$	$u, d, s, ..(g)$	$x \gtrsim 0.001$
$pp \to W^- c,\ W^+ \bar{c}$	$gs \to W^- c$	$s, \bar{s}$	$x \sim 0.01$
$pp \to (\gamma^* \to \ell^+\ell^-)X$	$u\bar{u}, d\bar{d}, .. \to \gamma^*$	$\bar{q}, g$	$x \gtrsim 10^{-5}$
$pp \to (\gamma^* \to \ell^+\ell^-)X$	$u\gamma, d\gamma, .. \to \gamma^*$	γ	$x \gtrsim 10^{-2}$
$pp \to b\bar{b} X,\ \ t\bar{t}X$	$gg \to b\bar{b},\ \ t\bar{t}$	g	$x \gtrsim 10^{-5}, 10^{-2}$
$pp \to t(\bar{t}) X,$	$bu(\bar{b}d) \to td(\bar{t}u)$	$b, d/u$	$x \gtrsim 10^{-2}$
$pp \to$ exclusive $J/\psi, \Upsilon$	$\gamma^*(gg) \to J/\psi, \Upsilon$	g	$x \gtrsim 10^{-5}, 10^{-4}$
$pp \to \gamma X$	$gq \to \gamma q,\ g\bar{q} \to \gamma\bar{q}$	g	$x \gtrsim 0.005$

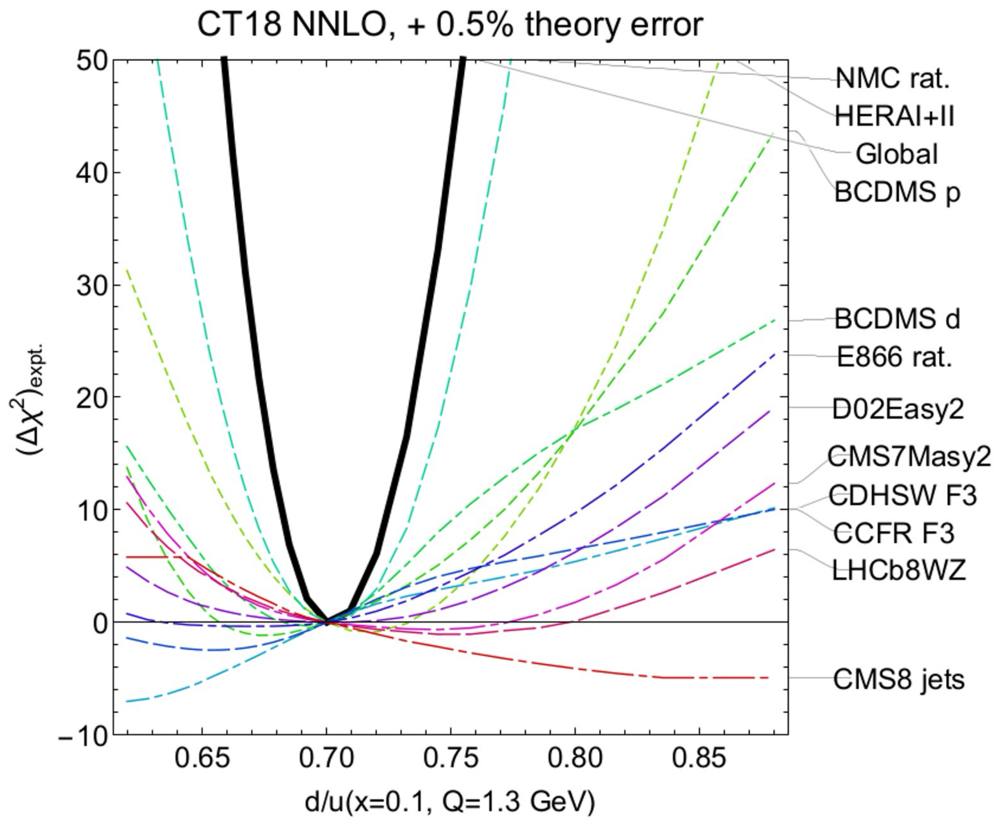


again, many complementary data required to:

- ❑ unravel PDF x, flavor dependence
- ❑ test PDF **universality**
- ❑ test QCD factorization theorem(s)

PDF sensitivity of individual experiments





we can often predict the most sensitive data sets *before* fitting, as shown in this illustration for the large-x PDF ratio d/u

HERA and fixed-target (BCDMS, NMC) data lead

data-driven pulls on PDFs often dominated by few sensitive expts!

Sensitivity ranking tables

experimental pulls have
~power-law falloffs

No.	Expt.	N_{pt}	Rankings, **CT14 HERA2 NNLO PDFs** $\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
1	HERAI+II'15	1120.	620.	0.0922	B		**A**	3	**A**	3	**A**	3	B		C	
2	CCFR-F3'97	86	218.	0.423	C	**1**	C	**1**		3	B	**1**	C	2		
3	BCDMSp'89	337	184.	0.0908			C		C		B	3	C			
4	NMCrat'97	123	169.	0.229	C	2					C	2	B	2		
5	BCDMSd'90	250	141.	0.0939	C				C	3	C	3	C	3		
6	CDHSW-F3'91	96	115.	0.199	C	2	C	2		3	C	2	C	3		
7	E605'91	119	113.	0.158	C	2	C	2				3				
8	E866pp'03	184	103.	0.0935		3	C	3			C	3				
9	CCFR-F2'01	69	89.1	0.215		3		3	C	2		3		2		3
10	**CMS8jets'17**	185	87.6	0.0789					C	3						
11	CDHSW-F2'91	85	82.4	0.162		3		3		3		3	C	3		
12	CMS7jets'13	133	63.8	0.0799					C	3						
13	NuTeV-nu'06	38	58.9	0.259		3		3				3		3	C	**1**
14	**CMS7jets'14**	158	57.5	0.0606					C	3						
15	CCFR SI nub'01	38	49.4	0.217		3		3				3		3	C	**1**
16	**ATLAS7jets'15**	140	48.2	0.0574						3						
17	CCFR SI nu'01	40	48.	0.2		3		3				3		3	C	**1**

Experiments are listed in the descending order of the summed sensitivities to $\bar{d}, \bar{u}, g, u, d, s$

For each flavor, A and 1 indicate the strongest total sensitivity and strongest sensitivity per point

C and 3 indicate marginal sensitivities; low sensitivities are not shown

PDF fitting extendable beyond proton: nuclear PDFs

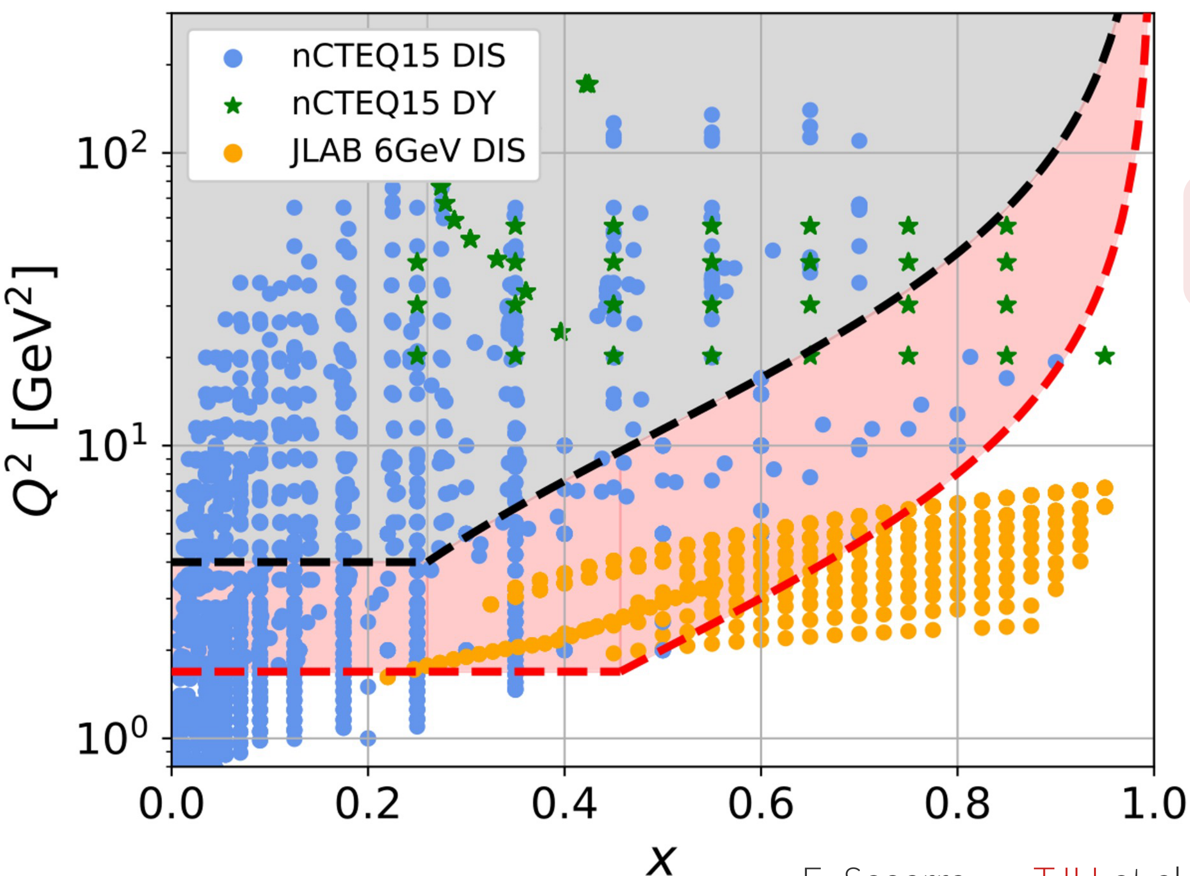
nCTEQ: parametrize and fit nuclear PDFs directly

$$f^A = \frac{Z}{A}\, f^{p/A} + \frac{(A-Z)}{A}\, f^{n/A}$$

$$x f_i^{p/A}(x, Q_0) = c_0 x^{c_1} (1-x)^{c_2} e^{c_3 x} (1 + e^{c_4} x)^{c_5}$$
$$c_k \longrightarrow c_k(A) \equiv p_k + a_k (1 - A^{-b_k})$$

fit range of nuclear data; relax *W, Q* cuts

include TMC, HT prescriptions; deuteron correction

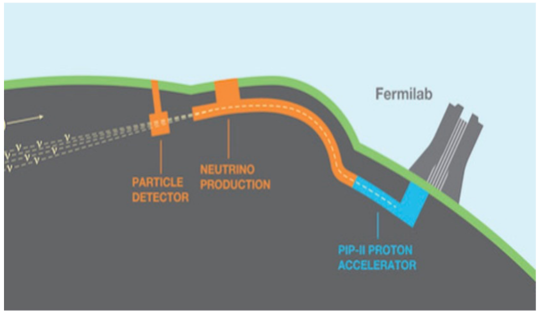


(dependence on proton baseline PDF)

nCTEQ15HiX

E. Segarra, ..., TJH et al., PRD **103** (2021) 11, 114015

nuclear DIS/PDFs impact νA predictions and experiments



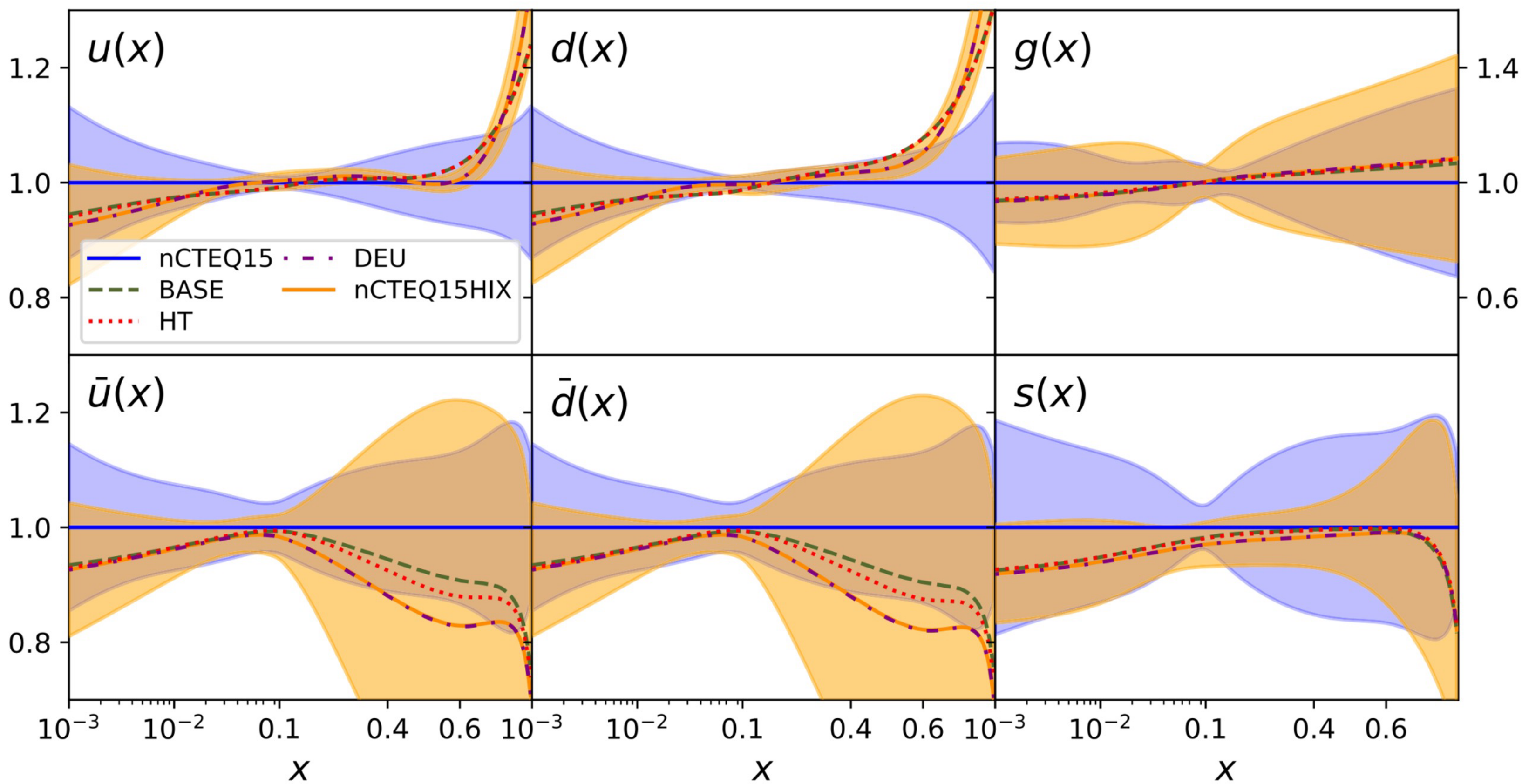
parametrization of *A* dependence allows predictions for ^{56}Fe

(cf. NuTeV, CCFR, ...)

→ low-*Q*, *W* effects sizable in nuclei; need simultaneous treatment with free-nucleon degrees-of-freedom

Iron PDF Ratios to nCTEQ15 ($Q = 2$ GeV)

...also for ^{40}Ar/DUNE...



E. Segarra, ..., TJH et al., PRD **103** (2021) 11, 114015

→ **nuclear and proton PDF fits often interdependent → <u>need simultaneous fits</u>**

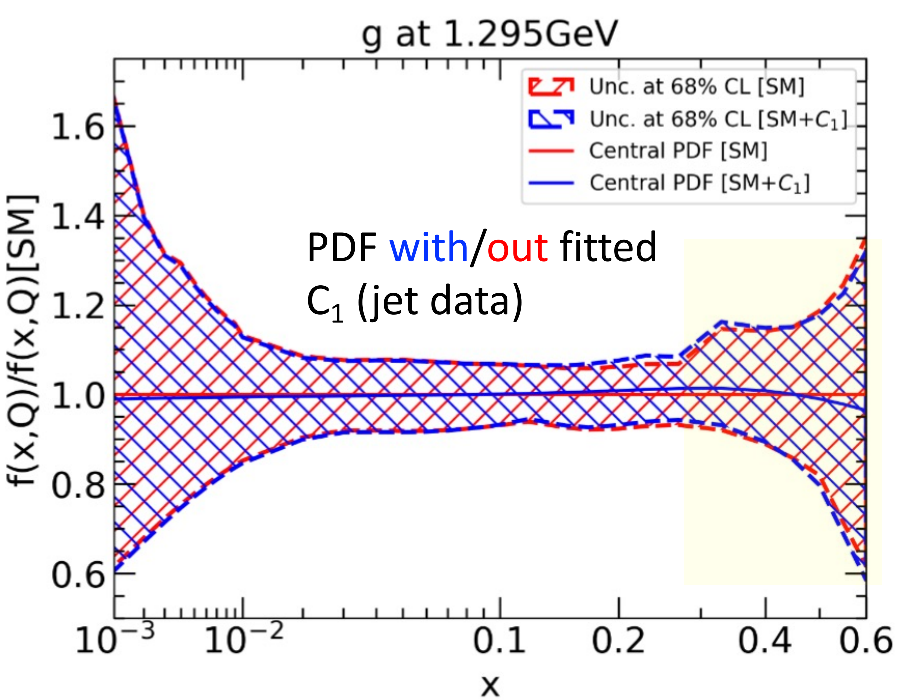
still more universal fits: PDFs with BSM ingredients

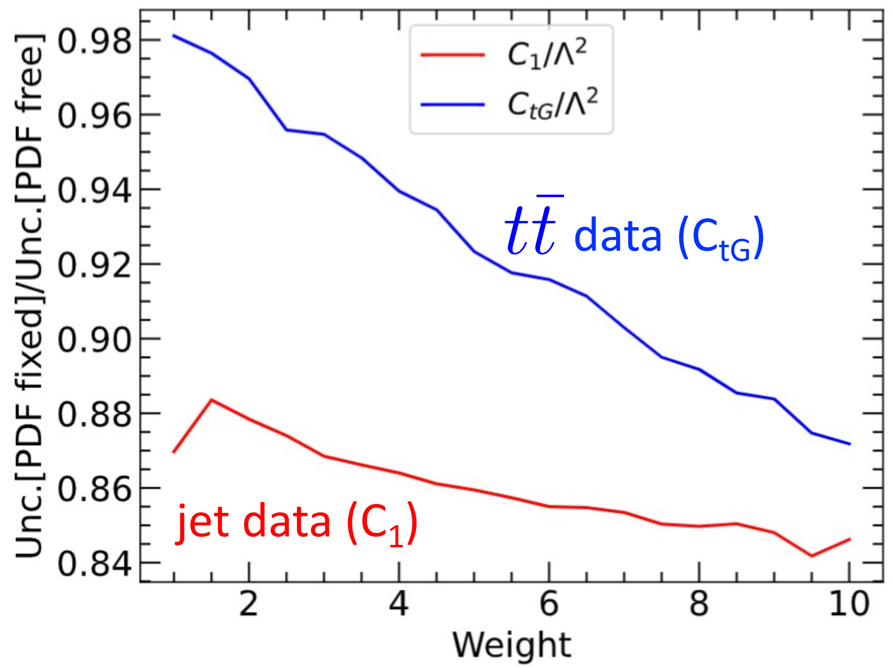
- ongoing effort to constrain BSM model independently via EFT (SMEFT) global fits

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \dots$$

→ to minimize bias: jointly fit PDFs, SMEFT; examine PDF-SMEFT correlations

arXiv:2211.01094





- PDF-SMEFT correlations (*e.g.*, with high-*x* gluon) are mild for jet, $t\bar{t}$ data

→ will likely be more severe with higher precision (HL-LHC); important future effort

conclusions for today

<u>collinear PDFs are central to HEP theory and experiment</u>

→ PDF encode aspects of (non)perturbative QCD

→ complicated interplay in QCD, EW theory; statistics; phenomenology

→ crucial to confront wide array of hadronic experiments

<u>**tomorrow**: PDFs from the EIC to the LHC</u>

→ EIC is a QCD machine; will refine PDF accuracy

→ numerous arrays of Energy and Intensity Frontiers will be impacted

→ essential in quest for more universal PDF-based QCD

Argonne NATIONAL LABORATORY Fermilab BROOKHAVEN NATIONAL LABORATORY CERN

conclusions for today

<u>collinear PDFs are central to HEP theory and experiment</u>

→ PDF encode aspects of (non)perturbative QCD

→ complicated … phenomenology

→ crucial to confront wide array of hadronic experiments

Thanks very much!

<u>tomorrow: PDFs from the EIC to the LHC</u>

→ EIC is a QCD machine; will refine PDF accuracy

→ numerous arrays of Energy and Intensity Frontiers will be impacted