Evidence for intrinsic charm quarks in the proton \[\text{[Nature608.483]}\]

Felix Hekhorn
Jefferson Lab Remote Theory Seminar June 2023

Acknowledgement: This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement number 740006.
INTRINSIC CEVEHOLES AT THE SSC

Stanley J. Brodsky
Stanford Linear Accelerator Center, Stanford University, Stanford CA 94305

John C. Collins
Department of Physics, Illinois Institute of Technology, Chicago IL 60616
and
High Energy Physics Division, Argonne National Laboratory, Argonne IL 60439

Stephen D. Ellis
Department of Physics, FM-15, University of Washington, Seattle WA 98195

John F. Gunion
Department of Physics, University of California, Davis CA 95616

Alfred H. Mueller
Department of Physics, Columbia University, New York NY 10027

Summary
The possibility of the production at high energy of heavy quarks, supersymmetric particles and other large mass colored systems via the intrinsic twist-six components in the proton wave function is discussed. While the existing data do not rule out the possible relevance of intrinsic charm production at present energies, the extrapolation of such intrinsic contributions to very high masses and energies suggests that they will not play an important role at the SSC.

sufficiently large. The data from the EMC collaboration on deep-inelastic muon scattering could also be interpreted on deep-inelastic muon scattering could also be interpreted as suggesting an unexpectedly large charm structure function in the region $x > 0.3$.

The possible existence of such a new production mechanism is of great importance for design considerations at the SSC. An example of the importance of this issue is that, if intrinsic large $x$ production is dominant, experiments and, perhaps, even the machine should be designed to focus on the forward "dissociative" regime. The ques-
Evidence for intrinsic charm quarks in the proton

The NNPDF Collaboration

Abstract

The theory of the strong force, quantum chromodynamics, describes the proton in terms of quarks and gluons. The proton is a state of two up quarks and one down quark bound by gluons, but quantum theory predicts that in addition there is an infinite number of quark–antiquark pairs. Both light and heavy quarks, whose mass is respectively smaller or bigger than the mass of the proton, are revealed inside the proton in high-energy collisions. However, it is unclear whether heavy quarks also exist as a part of the proton wavefunction, which is determined by non-perturbative dynamics and accordingly unknown: so-called intrinsic heavy quarks\(^1\). It has been argued for a long time that the proton could have a sizable intrinsic component of the lightest heavy quark, the charm quark. Innumerable efforts to
Nature 2022 in the media

**Physics World**

Protons contain intrinsic charm quarks, machine-learning analysis suggests
23 Aug 2022

**De Volkskrant**

Proton bevat een wonderlijk extra deeltje: de ‘charm quark’
Protonen, fundamentele bouwstenen van alle materie, blijken een nieuw ingrediënt te bevatten: de ‘charm quark’. Natuurkundigen reageren opgetogen: ‘Verbazingwekkend dat er nog iets nieuws valt te leren over een oude bekende als het proton.’
Frank Roessen 17 Augustus 2022, 21:05

**New Scientist**

Physicists surprised to discover the proton contains a charm quark
The textbook description of a proton says it contains three smaller particles - two up quarks and a down quark - but a new analysis has found strong evidence that it also holds a charm quark

**Deutschlandfunk**

Protonen mit Charm: Auch schwere Quarks finden sich in Kernbauteinen
Ostfildern, Frk 18 August 2022, 00:41Uhr

**Science**

Une étude confirme que le proton possède un quark charm intrinsèque
Pier Brossette · 19 août 2022
0. Introduction

1. NNPDF4.0 [EPJC82.428]

2. Intrinsic Charm [Nature608.483]

3. Summary
1. NNPDF4.0 [EPJC82.428]
Fitting PDFs

Theory T(f) \quad Methodology \quad Data D

Fit \rightarrow PDF f
Data: History

![Graph showing the history of data in NNPDF. The x-axis represents years from 2008 to 2022, and the y-axis represents the number of data points and datasets. There are separate sections for Tevatron, LHC Run I, and LHC Run II. The number of data points and datasets continues to increase over time.](image-url)
Methodology: Replicas

- Data is given by central values and covariance matrix
- generate Monte Carlo data replicas which as an ensemble represent the experiment
- fit one PDF replica to each data replica
- ⇒ ensemble of PDF replica
Theory: pineline [2302.12124]

https://nnpdf.github.io/pineline

- "Industrialization of High-Energy Theory Predictions":
  - collect diverse generators in an "assembly line"
  - NNPDF4.0: > 4.5k datapoints + > 10 generators
- be reproducible (i.e. track data and metadata) and open source!
- not yet in NNPDF4.0 but any future release

⇒ please provide new calculations in an "interfaceable" way
PineAPPL is a fast interpolation grid library that

- extends to arbitrary orders in QCD and EW coupling
- provides a very good Command Line Interface
- provides several interfaces: C, C++, Fortran, Rust, Python
- can convert APPLgrid [EPJC66.503] and FastNLO [DIS12.217]
- interfaces to Mg5 [JHEP07.079], yadism, Vrap [PRD69.094008] - soon
  MATRIX [EPJC78.537]

Github: https://github.com/NNPDF/pineappl
Website: https://nnpdf.github.io/pineappl/
Theory: **EKO** [EPJC82.976]

DGLAP:

\[
\mu^2_F \frac{d}{d\mu^2_F} \mathbf{E}(\mu^2_F \leftarrow \mu^2_{F,0}) = \mathbf{P}(a_s(\mu^2_R), \mu^2_F) \otimes \mathbf{E}(\mu^2_F \leftarrow \mu^2_{F,0})
\]

with

\[
f(\mu^2_F) = \mathbf{E}(\mu^2_F \leftarrow \mu^2_{F,0}) \otimes f(\mu^2_{F,0})
\]

- compute in Mellin space, but deliver in x space
- correct treatment of intrinsic PDFs
Theory: EKO Project Management

- Fully open source: 🌐 https://github.com/NNPDF/eko
- Written in Python
- Fully documented: 📚 https://eko.readthedocs.io/
**Theory: yadism**

Yet Another DIS Module

- [https://github.com/NNPDF/yadism](https://github.com/NNPDF/yadism)
- [https://yadism.readthedocs.io](https://yadism.readthedocs.io)

- DIS coefficient function database:

<table>
<thead>
<tr>
<th></th>
<th>light</th>
<th>heavy</th>
<th>intrinsic</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>$O(a_s^2)$ [VVM05]</td>
<td>$O(a_s^2)$ [Hek19]</td>
<td>$O(a_s)$ [KS98]</td>
</tr>
<tr>
<td>CC</td>
<td>$O(a_s^2)$ [MRV08]</td>
<td>$O(a_s)$ [GKR96]</td>
<td>$O(a_s)$ [in prep.]</td>
</tr>
</tbody>
</table>

- implemented flavor number schemes: ZM-VFNS, FFNS, FONLL
Go to the past and look into the (back then) future!

\[
\chi^2 / N \text{ (only exp. covmat)}
\]

<table>
<thead>
<tr>
<th>(dataset)</th>
<th>NNPDF4.0</th>
<th>pre-LHC</th>
<th>pre-Hera</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-HERA</td>
<td>1.09</td>
<td>1.01</td>
<td>0.90</td>
</tr>
<tr>
<td>pre-LHC</td>
<td>1.21</td>
<td>1.20</td>
<td>23.1</td>
</tr>
<tr>
<td>NNPDF4.0</td>
<td>1.29</td>
<td>3.30</td>
<td>23.1</td>
</tr>
</tbody>
</table>
Go to the past and look into the (back then) future!

\[ \chi^2 / N \text{ (exp. and PDF covmat)} \]

<table>
<thead>
<tr>
<th>(dataset)</th>
<th>NNPDF4.0</th>
<th>pre-LHC</th>
<th>pre-Hera</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-HERA</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-LHC</td>
<td>1.17</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>NNPDF4.0</td>
<td>1.12</td>
<td>1.30</td>
<td>1.38</td>
</tr>
</tbody>
</table>

- without data PDF errors have to be big
- with PDF errors the total uncertainty increases, and accommodates for difference between predictions and new data
Fake a universe with known input assumptions

1. Assume a “true” underlying PDF (e.g. a single PDF replica)
2. Produce fake data distributed accordingly
3. Perform a fit to this fake data

Observe statistical estimators (e.g. bias and variance)
→ Is the truth within one sigma in 68% of cases?

\[
\begin{array}{c|c}
\sqrt{\text{bias/variance}} & \xi^{(\text{data})}_{1\sigma} \\
\hline
1.03 \pm 0.05 & 0.68 \pm 0.02
\end{array}
\]
NNPDF4.0: PDF plot

NNPDF4.0 at $Q = 3.2$ GeV

The plot shows the PDFs for different quark flavors and antiquark flavors as a function of $x$. The PDFs are labeled with $g/10$, $u$, $d$, $v$, $s$, $\bar{u}$, $\bar{d}$, and $c$.
• typical uncertainties in data region: singlet $\sim 1\%$, nonsinglet $\sim 2 - 3\%$

• data region: $10 \lesssim M_X \lesssim 3 \times 10^3$ GeV, $-4 \lesssim y \lesssim 4$
2. Intrinsic Charm  [Nature608.483]
Nomenclature

- **perturbative charm**
  - is fully perturbative, i.e., predictable at all scales
  - generated by matching conditions and evolution
  - always present for $\mu_F > \mu_h = m_h$
  - it is $g \rightarrow c\bar{c}$, so (mostly) no asymmetry possible ($c \neq \bar{c}$)
  - default for CTEQ and MSHT

- **intrinsic charm**
  - is non-perturbative
  - charm in 3 light flavor scheme
  - CTEQ: use a model (e.g. [BHPS])

- **fitted charm**
  - default for NNPDF
  - don’t assume anything - just fit charm in 4 flavor scheme!
  - is an arbitrary mixture of intrinsic and perturbative charm
Mass Dependency on PDFs

![Graphs showing mass dependency on charm PDFs]

**default charm PDF**
- $Q = 1.65$ GeV
- $m_c = 1.38$ GeV
- $m_c = 1.51$ GeV
- $m_c = 1.64$ GeV

**perturbative charm PDF**
- $Q = 1.65$ GeV

**default charm PDF**
- $Q = 100$ GeV

**perturbative charm PDF**
- $Q = 100$ GeV
slow perturbative convergence of OME: NNLO and N3LO differ significantly
Matching Conditions and Backward Evolution

For (forward) evolution across a matching scale $\mu_h^2$:

$$\tilde{f}^{(n_f+1)}(\mu_{F1}^2) = \tilde{E}^{(n_f+1)} \left( \mu_{F1}^2 \leftarrow \mu_h^2 \right) R^{(n_f)} \tilde{A}^{(n_f)}(\mu_h^2) \tilde{E}^{(n_f)} \left( \mu_h^2 \leftarrow \mu_{F0}^2 \right) \tilde{f}^{(n_f)}(\mu_{F0}^2)$$

with $R^{(n_f)}$ a flavor rotation matrix and $\tilde{A}^{(n_f)}(\mu_h^2)$ the operator matrix elements (partially known up to $\mathcal{N}^3\text{LO}$)
For (forward) evolution across a matching scale $\mu_h^2$:

$$\tilde{f}^{(n_f+1)}(\mu_{F,1}^2) = \tilde{E}^{(n_f+1)}(\mu_{F,1}^2 \leftarrow \mu_h^2)R^{(n_f)}(\mu_h^2)\tilde{A}^{(n_f)}(\mu_h^2)\tilde{E}^{(n_f)}(\mu_h^2 \leftarrow \mu_{F,0}^2)\tilde{f}^{(n_f)}(\mu_{F,0}^2)$$

with $R^{(n_f)}$ a flavor rotation matrix and $\tilde{A}^{(n_f)}(\mu_h^2)$ the operator matrix elements (partially known up to $N^3$LO) for backward evolution:

- invert $\tilde{E}^{(n_f)}$: simple (invert RGE flow) ✓
- invert $R^{(n_f)}$: simple (static matrix) ✓
- invert $\tilde{A}^{(n_f)}$: expanded or exact
Strategy

Based on NNPDF4.0 [EPJC82.428]

- 4FNS Charm PDF constrained by experimental data for $Q > Q_0$
  - NNPDF4.0 dataset
  - NNLO QCD calculations

- 4FNS Charm PDF parametrised at $Q_0$
  - Deep-learning parametrisation
  - Monte Carlo representation of uncertainties

- 4FNS to 3FNS transformation
  - NNLO or $N^3$LO matching conditions

- Intrinsic (3FNS) Charm
  - Scale-independent
  - PDF and MHO uncertainties

QCD evolution
in **3FNS** a valence-like peak is present

- for $x \leq 0.2$ the perturbative uncertainties are quite large
- the carried momentum fraction is within 1%
The PDF Plot with Model Comparison

- In [3FNS] a valence-like peak is present
- For $x \leq 0.2$ the perturbative uncertainties are quite large
- The carried momentum fraction is within 1%
- predict better recent measurement
- reweighting is consistent
• direct measurement of $F_2^c$
• evidence for intrinsic charm claimed, but experiment disputed
• adding EMC data is consistent
- direct measurement of $F_2^c$
- can distinguish intrinsic charm scenarios
• we find a 3σ evidence of intrinsic charm
• result is stable with mass variation, dataset variation
Charm Asymmetry (PRELIMINARY!)

- also parametrize $c^- = c - \bar{c} \Rightarrow$ intrinsic!
- significance for baseline now $> 3\sigma$
- $\sim 1.5\sigma$ evidence for $c^- \neq 0$
3. Summary
Summary

We fit the charm PDF in order to get

- realistic error estimate
- no strong dependence on charm mass
- no sensitivity to MHOU in matching condition

We find

- large uncertainties and charm compatible with zero at small $x$
- $3\sigma$ evidence for an intrinsic charm valence-like peak

The road ahead:

- more data $\rightarrow 5\sigma$ evidence
- $c - \bar{c}$ asymmetry phenomenology
Thank you!
4. Backup slides
- about 50 new datasets & 400 extra datapoints
- DIS/FTDY: dataset as in NNPDF3.1 + nomad neutrino + SeaQuest DY
- LHC: full 7 TeV and 8 TeV dataset & extensive use of 13 TeV data
- several new processes: prompt photon; single top; dijets; Hera jets
Methodology: Neural Network and Hyperoptimization

\[ f_j(x, Q_0) = x^{-\alpha_j}(1 - x)^{\beta_j} \text{NN}_j(x) \]

- functional form: neural network (corresponds to many “effective parametrizations”)
- choose model parameters? hyperoptimization! (i.e. scan parameter space)
- prevent overfitting!
The workhorse in the background: PineAPPL
Intrinsic Charm, NNLO match (PDFU)
4FNS Charm, Q=1.51 GeV (PDFU)
Intrinsic Charm, N^3LO match (PDFU)
IC - dataset variation

Baseline dataset
+ EMC $F_2$

DIS dataset

Collider dataset

no LHCb W, Z data
\[ x_c^+ (x) \]

- \( m_c = 1.38 \text{ GeV (PDF+MHOU)} \)
- \( m_c = 1.51 \text{ GeV (PDF+MHOU)} \)
- \( m_c = 1.64 \text{ GeV (PDF+MHOU)} \)