The asymmetry of antimatter in the proton

The fundamental building blocks of the proton—quarks and gluons—have been known for decades. However, we still have an incomplete theoretical and experimental understanding of how these particles and their dynamics give rise to the quantum bound state of the proton and its physical properties, such as its spin. The two up quarks and the single down quark that comprise the proton in the simplest picture account only for a few per cent of the proton mass, the bulk of which is in the form of quark kinetic and potential energy and gluon energy from the strong force. An essential feature of this force, as described by quantum chromodynamics, is its ability to create matter–antimatter quark pairs inside the proton that exist only for a very short time. Their fleeting existence makes the antimatter quarks within protons difficult to study, but their existence is discernible in reactions in which a matter–antimatter quark pair annihilates. In this picture of quark–antiquark creation by the strong force, the probability distributions as a function of momentum for the antimatter quark pair annihilation are considerably different, with more abundant down antimatter quarks than up antimatter quarks over a wide range of momenta. These results are expected to revive interest in several proposed mechanisms for the origin of this antimatter asymmetry in the proton that had been disfavoured by previous results, and point to future measurements that can distinguish between these mechanisms.

The structure of the proton is a prototypical example of a strongly coupled and correlated system with quarks and gluons interacting according to quantum chromodynamics (QCD). At large energy and momentum scales, the interaction is relatively weak, whereas at lower energy scales the picture is clouded by the increasingly strong interaction. The original quark model, in which the proton consists of two up quarks (u) and one down (d) quark, has an appealing simplicity, but experiments that measure the distributions of quarks as a function of the fractional momentum (x) of the proton that these quarks carry have revealed a rich structure with additional quarks, antimatter quarks

1Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. 2Department of Physics, School of Science, Tokyo Institute of Technology, Tokyo, Japan. 3Randall Laboratory of Physics, University of Michigan, Ann Arbor, MI, USA. 4Institute of Physics, Academia Sinica, Taipei, Taiwan. 5Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. 6Physics Division, Los Alamos National Laboratory, Los Alamos, NM, USA. 7Physics Division, Argonne National Laboratory, Lemont, IL, USA. 8Department of Engineering and Physics, Abilene Christian University, Abilene, TX, USA. 9Accelerator Division, Fermi National Accelerator Laboratory, Batavia, IL, USA. 10Department of Physics and Astronomy, Mississippi State University, Mississippi State, MS, USA. 11Department of Physics, National Kaohsiung Normal University, Kaohsiung, Taiwan. 12Department of Physics, Yamagata University, Yamagata, Japan. 13Department of Physics, Washington University, St. Louis, MO, USA. 14Theoretical Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. 15Department of Physics and Astronomy, University of Colorado, Boulder, CO, USA. 16Department of Physics, University of Maryland, College Park, MD, USA. 17Institute of Particle and Nuclear Studies, KEK, High Energy Accelerator Research Organization, Tsukuba, Japan. 18Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. 19Present address: Accelerator Systems Division, Argonne National Laboratory, Lemont, IL, USA. 20K. Nakano, K. Nakano, E. A. Strickland, J. Wu, Z. Xi, Z. Ye. 21K. Nakano, K. Nakano, E. A. Strickland, J. Wu, Z. Xi, Z. Ye. 22K. Nakano, K. Nakano, E. A. Strickland, J. Wu, Z. Xi, Z. Ye.

J. Dove, B. Korns, R. E. McClellan, S. Miyasaka, D. H. Morton, K. Nagai, S. Prasad, F. Sanft, M. B. C. Scott, A. S. Tadepalli, C. A. Aidala, J. Arrington, A. C. Ayuso, C. L. Barker, C. N. Brown, W. C. Chang, A. Chen, D. C. Christian, B. P. Dannowitz, M. Daehler, M. Diefenthaler, E. El Fassi, D. F. Geesaman, R. Gilmur, Y. Goto, L. Guo, T. J. Hague, R. J. Holt, D. Isenhour, E. R. Kinney, N. Kits, A. Klein, D. W. Kleinjan, Y. Kudo, C. Leung, P. J. Lin, K. Liu, M. X. Liu, W. Lorenzon, N. C. R. Makins, M. Mesquita de Medeiros, P. L. McGaughey, Y. Miyachi, I. Mooney, K. Nakahara, K. Nakano, S. Naras, J.-C. Peng, A. J. Puckett, B. J. Ramson, P. E. Reimer, J. G. Rubin, S. Sawada, T. Sawada, T.-A. Shibata, D. Su, M. Teo, B. G. Tice, R. S. Towell, S. Uemura, S. Watson, S. G. Wang, A. B. Wickes, J. Wu, Z. Xi & Z. Ye.
Antiquarks and gluons, beyond the minimal three-quark Fock state. These additional quarks and antiquarks are referred to as sea quarks. Collectively quarks and gluons are referred to as partons. It is not possible to identify any individual up or down quark as a sea or valence quark, but antiquarks and strange quarks must belong to the sea, so their study promises to reveal information about the structure of the proton. Even before QCD, hadronic models emphasized the importance of the presence of mesons (for example, ref. 5), and therefore (as was realized later) antiquarks, in the physical state of a proton or neutron. Nevertheless, the initial naive expectation was that the sea was formed predominantly by gluons splitting into quark–antiquark pairs. Indeed, several authors assumed that at some low momentum scale the sea quarks and gluons vanish, and at high momentum scales they are all generated by gluon radiation and then gluon splitting. These assumptions were used to describe successfully the existing data in the late 1970s (for example, refs. 6, 7).

In the early 1990s the New Muon Collaboration (NMC) reported measurements of the deep inelastic structure functions ($F_2$) of hydrogen (H) and deuterium (D) at 0.004 < $x$ < 0.8. The cross-section for deep inelastic scattering measures the charge-squared-weighted sum of the quark and antiquark distributions, in this case at an average scale of 4 GeV$^2$/c$^2$ (c, speed of light in vacuum). The integrals of the parton distributions of the proton ($p$) and neutron ($n$) were assumed to have charge symmetry,

$$\int_0^1 u_p(x)\,dx = \int_0^1 d_n(x)\,dx,$$

where $u_p(x)$ is the probability distribution of up quarks in the proton and $d_n(x)$ is the distribution of down quarks in the neutron, with similar integrals for the other quark flavours, and nuclear effects in deuterium were assumed to be small. In that case, their measurements and their estimate of the unmeasured region led NMC to conclude

$$\int_0^1 \frac{dx}{x}[F_2^H(x) - F_2^D(x)] = \frac{1}{3} + \frac{2}{3} \int_0^1 dx[\bar{n}(x) - \bar{d}(x)] = 0.235 \pm 0.026,$$

and thus the integral of $\bar{d}(x)$ is greater than that of $\bar{n}(x)$,

$$\int_0^1 dx[\bar{d}(x) - \bar{u}(x)] = 0.147 \pm 0.039,$$

where $\bar{n}(x)$ and $\bar{d}(x)$ are the distributions of up and down antiquarks in the proton, respectively.

---

**Fig. 1** Ratios $\sigma_D/(2\sigma_H)$. (a) Ratios $\sigma(x_t)/[2\sigma_H(x_t)]$ (red filled circles) with their statistical (vertical bars) and systematic (yellow boxes) uncertainties as a function of $x_t$ (a) and of transverse momentum, $P_T$ (b), and mass, $M$ (c), of the virtual photon. The cross-section ratios are defined as the ratio of luminosity-corrected yields from the hydrogen and deuterium targets. Also shown (open black squares) are the results of the NuSea experiment for the $x_t$ dependence, with statistical uncertainties only. NuSea also reports an overall 1% common systematic uncertainty. The mass scale of the NuSea data is up to 50% larger than that of the SeaQuest data, and the distributions in the other kinematic variable, $x_t$, are slightly different owing to the differing beam energies and acceptances of the experiments. These differences imply that the cross-section ratios do not need to be identical. This is demonstrated by the red solid and violet dashed curves representing NLO calculations of the cross-section ratio with SeaQuest and NuSea kinematics using CT18 parton distributions. The horizontal bars on the data points indicate the width of the bins.
The Drell–Yan process in hadron–hadron collisions is a reaction in which a quark and an antiquark annihilate into a virtual photon, and that virtual photon decays into a lepton–antilepton pair. One can isolate the antiquark distributions from the Drell–Yan cross-section by making use of this property. At lowest order, the Drell–Yan cross-section is given by

$$\frac{d^2\sigma}{dx_0 dx_t} = \frac{4\pi^2}{9\pi} \sum_q e_q^2 q(x_0) q(x_t) + \frac{4\pi\alpha}{3\pi} \frac{d}{dx_0} q(x_0) q(x_t). \quad (3)$$

where $x_0$ and $x_t$ are the momentum fractions of the beam and target partons participating in the reaction, respectively, $e_q$ is the electrical charge of quark flavour $q$, $q(x_0)$ and $q(x_t)$ are the probability distributions for quarks and antiquarks of flavour $q$ in the proton, $a$ is the fine-structure constant, and is the square of the centre of mass energy of the beam and target. In a Drell–Yan measurement at CERN, the NA51 collaboration confirmed that $\bar{d}(x)/\bar{u}(x)$ is larger than $\bar{u}(x)$ at an average scale of 54 GeV$^2$/c$^2$.

There are various mechanisms that may account for the antiquark flavour asymmetry of the proton: recent reviews include refs. 19, 20, 21. Pauli blocking might lead to a flavour asymmetry in the extra $u$–$\bar{u}$ pairs from forming, but the $x$ dependence and even the sign of this mechanism are debated in the literature 22. A related approach involves statistical models of the mesonic degrees of freedom in the proton structure. These latter models (statistical, chiral soliton and meson–baryon) attempt to describe the entire non-perturbative composition of the proton, and a common feature of these models is a rise in the $\bar{d}/\bar{u}$ flavour asymmetry at low $x$. Although at low $x$ this behaviour reproduces the NuSea data, none of these models is able to reproduce the fall-off at higher $x$ observed by NuSea. One feature of the NuSea results, with admittedly limited statistics, is the suggestion that the ratio of $\bar{d}/\bar{u}$ begins to decrease for $x > 0.2$, reaching a value of $\bar{d}(x)/\bar{u}(x) = 0.35 \pm 0.40$ at $x = 0.31$, as seen in Fig. 2.

The SeaQuest experiment at Fermilab (E906) at Fermi National Accelerator Laboratory (Fermilab) was designed to investigate the flavour asymmetry at higher $x$ values than NuSea with the newly constructed experimental apparatus that is described in detail in ref. 23. With a proton beam at an energy of 120 GeV, liquid hydrogen and deuterium targets and a focusing magnet of 10 T m after the target region, the experiment was optimized for the study of the production of quark–antiquark pairs in the intermediate region, with $x_t$ around 0.3. Approximately 0 to 80,000 protons at 53 MHz repetition rate. About
The dependence of the ratio \( \frac{\langle Y_x I_x \rangle}{\langle Y_h I_h \rangle} \) using different functional forms, such as \( \sigma(2\alpha_o) \), \( \sigma(D/(2\alpha_o)) \), and \( \sigma(H(2\alpha_o)) \), was studied. The cross-section ratio at zero intensity, \( \langle Y_x I_x \rangle \), was calculated at next-to-leading order (NLO) with the CT18 parton distribution. The NuSea results are at a different scale, than the SeaQuest results, 22−40 GeV\(^2\)/c\(^2\), than the SeaQuest results, \( \sigma(D/(2\alpha_o)) \), \( \sigma(H(2\alpha_o)) \), and \( \sigma(a_o(2\alpha_o)) \), defined as the ratio of cross-section to hydrogen and deuterium targets. Because parameters \( a \) and \( b \) are fitted to the entire range of \( x \), the systematic uncertainty is fully correlated among all \( x \) bins. The observation that both \( M \) and \( P \) (for all but the highest \( P \) bins) distributions for deuteron and hydrogen have the same shapes helps to confirm that the acceptances are very similar for each target. Also, shown in Fig. 1 are the results from NuSea as a function of \( x \), and the cross-section ratio calculated at next-to-leading order (NLO) with the CT18 parton distribution. The NuSea results are at a different scale, 54 GeV\(^2\)/c\(^2\), than the SeaQuest results, 22−40 GeV\(^2\)/c\(^2\). The cross-section ratios depend on both \( x_o \) and \( x \) and, owing to the differing beam energies and acceptances, the \( x_o \) distributions are slightly different for SeaQuest and NuSea, the effects of which are shown in Fig. 1 and Extended Data Fig. 1.

To extract \( \bar{d}(x)/\bar{t}(x) \), NLO calculations of \( \sigma(2\alpha_o) \) were carried out, starting from several NLO global fits to the parton distributions (CT10\(^{10} \), CT14\(^{14} \), CT18,\(^{16} \)MNHT2014\(^{18} \)). No nuclear correction for deuterium was applied, but a systematic uncertainty of \( (0.3 + 0.3 \%) \) was included according to the range of deviation from unity found in refs. \( 13,14 \). Holding all other parton distributions fixed, including the sum \( \bar{d}(x) + \bar{t}(x) \), the ratio \( \bar{d}(x)/\bar{t}(x) \) for each \( x \) bin was varied until the results converged on the measured cross-section ratios. The correlations of the statistical uncertainties of \( \sigma(2\alpha_o) \) were propagated through the extraction of \( \bar{d}/\bar{t} \). The dependence of the ratio \( \bar{d}(x)/\bar{t}(x) \) above the measured \( x \) region on the value of \( \sigma(2\alpha_o) \) was estimated by varying this value from 1.0 to 0.5 and 2.0. The spread of the results due to the choice of initial parton distributions was always less than half of the statistical error. Each \( x \) bin was subdivided into multiple \( x_o \) sub-bins. The cross-sections for hydrogen and deuterium were calculated separately for each sub-bin at \( \langle Y_x I_x \rangle \) and \( \langle Y_h I_h \rangle \) of that sub-bin, and an acceptance-weighted sum was used to determine the final cross-section. These distributions are given in Extended Data Table 3. Calculations using only one average \( x_o \) and \( x \) for each \( x \) bin were less reliable. It was also found that a leading-order extraction of \( \bar{d}(x)/\bar{t}(x) \) using leading-order parton distributions and cross-section calculations gave very similar results for the ratios as the NLO results.

The resulting \( \bar{d}(x)/\bar{t}(x) \) ratios obtained when starting with the CT18 distributions are given in Table 2 at the scale of each \( x \) bin and are displayed in Fig. 2 in comparison with the NuSea results. The trends between the two experiments at higher \( x \) are quite different. No explanation has been found yet for these differences, even though there is a small overlap between the members of the NuSea and SeaQuest

### Table 1 | Ratios \( \sigma(2\alpha_o)/\sigma(D/(2\alpha_o)) \)

| \( x \) bin | \( (x_1) \) | \( (x_2) \) | \( \langle M \rangle (\text{GeV}/c^2) \) | \( \langle P \rangle (\text{GeV}/c) \) | \( \sigma(D/(2\alpha_o)) \pm \text{stat.} \pm \text{syst.} \) | \( \delta x \) |
|------------|-----------|-----------|----------------|----------------|---------------------------------|-------|
| 0.130−0.160 | 0.147 | 0.888 | 4.71 | 0.651 | 1.211 ± 0.052 ± 0.053 | 0.031 |
| 0.160−0.195 | 0.179 | 0.611 | 4.88 | 0.717 | 1.141 ± 0.043 ± 0.025 | 0.016 |
| 0.195−0.240 | 0.216 | 0.554 | 5.11 | 0.757 | 1.196 ± 0.042 ± 0.044 | 0.019 |
| 0.240−0.290 | 0.263 | 0.519 | 5.46 | 0.786 | 1.165 ± 0.046 ± 0.032 | 0.022 |
| 0.290−0.350 | 0.315 | 0.498 | 5.87 | 0.785 | 1.193 ± 0.050 ± 0.034 | 0.026 |
| 0.300−0.450 | 0.385 | 0.477 | 6.36 | 0.776 | 1.133 ± 0.064 ± 0.039 | 0.030 |

Cross-section ratios \( \sigma(D/(2\alpha_o)) \) binned in \( x \), with their statistical (stat.) and systematic (syst.) uncertainties and the average values for the kinematic variables in each \( x \) bin. The cross-section ratios are defined as the ratio of luminosity-corrected yields from the hydrogen and deuterium targets. The final column is the experimental resolution in \( x \), as determined by Monte Carlo simulations.

### Table 2 | Ratios \( \bar{d}(x)/\bar{t}(x) \)

| \( x \) | \( \bar{d}(x)/\bar{t}(x) \) ± \text{stat.} ± \text{syst.} |
|------|-----------------|
| 0.147 | 1.423 ± 0.093 ± 0.104 |
| 0.179 | 1.338 ± 0.085 ± 0.095 |
| 0.216 | 1.487 ± 0.092 ± 0.110 |
| 0.263 | 1.482 ± 0.094 ± 0.105 |
| 0.315 | 1.645 ± 0.104 ± 0.125 |
| 0.385 | 1.578 ± 0.114 ± 0.153 |

Ratios \( \bar{d}(x)/\bar{t}(x) \) with their upper and lower statistical and systematic uncertainties. The analysis was based on the present cross-section ratio data and NLO calculations of the Drell-Yan cross-sections using CT18 parton distributions for all calculations except \( \bar{d}(x)/\bar{t}(x) \). The systematic uncertainty is fully correlated among all \( x \) bins. The systematic uncertainty does not include a contribution from the choice of the base (CT18) parton distributions, which is small if added in quadrature to the other systematic uncertainties.
collaborations. The present data are reasonably described by the predictions of the statistical parton distributions of Basso et al.\(^2\) and by the chiral effective perturbation theory of Alberg & Miller\(^25\), also shown in Fig. 2. These two calculations emphasize rather different non-perturbative mechanisms that lead to the differences in \(\bar{\Delta}(x)\) and \(\bar{\Delta}(x)\). The present data show that \(\bar{\Delta}\) is greater than \(\bar{\Delta}\) for the entire range measured in this experiment. This provides important support for these and other non-perturbative mechanisms of the QCD structure of the proton that were disfavoured by the NuSea results.

The next major step to help us to assess the various models is to measure how much the spin and angular momentum of the antiquarks contribute to the total spin of the proton. It has long been realized that these models make rather different predictions for the contribution of the total spin of the proton carried by the antiquarks\(^1\). For example, meson–nucleon models predict that little spin is carried by the antiquarks, the statistical model finds a difference in spin of \(\Delta\bar{\Delta}(x) = -\frac{1}{2}\bar{\Delta}(x)\). Experiments are planned or are underway at Fermilab, the Thomas Jefferson National Accelerator Facility, Brookhaven National Laboratory and the European Organization for Nuclear Research to pursue this goal\(^9\).

These results affect the reach of a \(p\rightarrow p\) collider, such as the Large Hadron Collider, for new physics. For example, production of the high-mass \(Z^0\) and \(W^\pm\) particles has been shown to be dominated by light-quark fusion\(^3\). Calculations with contrasting statistical distributions (CTEQ6 distributions), which mimic the present (NuSea) data, show that the ratio \(\bar{\Delta}/\bar{\Delta}\) is relatively insensitive to scale in each case. At mass scales of \(\sim 5\) TeV, just above current limits on \(Z\) production\(^34\), cross-sections driven by the fusion of \(u(x_1, \tau)\) and \(\bar{\Delta}(x_2)\) with \(x_1 = x_2 = 0.3\)–0.4 will be enhanced according to the present results, and cross-sections driven by \(u(x_1, \tau)\) will be diminished, compared to those calculated with the central values of previous parton distributions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03282-z.

1. Ji, X., Yuan, F. & Zhao, Y. Proton spin after 30 years: what we know and what we don’t? Nat. Rev. Phys. 3, 27–38 (2021).
2. Yang, Y.-B. et al. Proton mass decomposition from the QCD energy momentum tensor. Phys. Rev. Lett. 121, 212001 (2018).
3. Ross, D. & Sachrajda, C. Flavour symmetry breaking in antiquark distributions. Nucl. Phys. B 149, 497–516 (1979).
4. Towell, R. S. et al. Improved measurement of the \(\bar{\Delta}/\bar{\Delta}\) asymmetry in the nucleon sea. Phys. Rev. D 64, 052002 (2001).
5. Fermi, E. & Marshall, L. On the interaction between neutrons and electrons. Phys. Rev. 72, 1139–1146 (1947).

6. Glück, M. & Reya, E. Dynamical determination of parton and gluon distributions in quantum chromodynamics. Nucl. Phys. B 130, 76–92 (1977).
7. Parisi, G. & Petronzio, R. On the breaking of Bjorken scaling. Phys. Lett. B 62, 331–334 (1976).
8. Amadozú, P. et al. Gottfried sum from the ratio \(F_2^p/F_2^\bar{p}\). Phys. Rev. Lett. 66, 2712–2715 (1991).
9. Arneodo, M. et al. Reevaluation of the Gottfried sum. Phys. Rev. D 50, R1–R3 (1994).
10. Drell, S. D. & Yan, T.-M. Massive lepton pair production in hadron-hadron collisions at high-energies. Phys. Rev. Lett. 25, 516–520 (1970); erratum 25, 902 (1970).
11. Baldit, A. et al. Study of the isospin symmetry breaking in the light quark sea of the nucleon from the Drell-Yan process. Phys. Lett. B 332, 244–250 (1994).}

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. © This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021, corrected publication 2022.
Methods

For the measurement of the pd-to-pp Drell–Yan cross-section ratios, 50.8-cm-long liquid hydrogen (0.069 interaction lengths) and liquid deuterium (0.116 interaction lengths) targets and an empty target flask were used. The targets were interchanged every few minutes to substantially reduce time-dependent systematic effects.

The SeaQuest spectrometer was constructed for the measurement of muon tracks in the forward region (laboratory angles less than 0.1 rad). It is composed of two magnets and four detector stations, each consisting of fast trigger detectors and drift chambers, distributed over 25 m along the beam direction, with a 1-mm-thick iron muon identification wall before the final detector station. The first magnet provided a 3.07 GeV/c transverse momentum kick between the target and the first detector station to enhance the acceptance for muon pairs resulting from the decay of high-mass virtual photons and to reduce the acceptance for the large background of low-mass (less than 4 GeV/c^2) virtual photon events. It was filled with solid iron to absorb the proton beam and all other hadrons and electrons produced either in the target or this beam dump. A second magnet (with a 0.41 GeV/c transverse momentum kick) located between the first and second detector stations provided for charge and momentum measurements of the muons. The iron hadron absorber between the third and fourth stations was used to establish the identification of muons.

Opposite-sign muon pairs were combined into di-muon candidates. The muons of each candidate were tracked back through the spectrometer to find whether they emerged from a common vertex along the beam path and near the target. The resolution of the vertex location was about 30 cm along the beam direction, compared with the 170 cm separation between the target centre and the average interaction point of protons in the solid iron magnet. Events identified as coming from the target were refitted using the target-centre vertex location, and the di-muon mass, M, the longitudinal momentum in the laboratory frame, p_L, and the transverse momentum, p_T, were determined. With this information, the fractional momentum of the beam and target quarks participating in the reaction were calculated as

\[
x_b = \frac{p_{\text{target}}}{p_{\text{beam}} + p_{\text{target}}} \times \frac{p_{\text{sum}}}{p_{\text{beam}} + p_{\text{target}}},
\]

\[
x_c = \frac{p_{\text{beam}}}{p_{\text{beam}} + p_{\text{target}}} \times \frac{p_{\text{sum}}}{p_{\text{beam}} + p_{\text{target}}},
\]

where p_target and p_beam are the four-momenta of the target and beam, respectively, and p_sum is the sum of the four-momenta of the positive and negative muons. The prominent J/ψ peak (resolution of 0.21 GeV/c^2) and the requirement that events come from the target or beam dump were used to calibrate the field integral of the solid iron magnet. The mass spectra are shown in Extended Data Fig. 3. Detailed field maps coupled with Hall probe measurements were used to calibrate the second magnet.

Various kinematic constraints were placed on the accepted events, the most important ones being the requirement for the virtual photon mass to be greater than 4.5 GeV/c^2 and for the primary vertex to be in the target region. The yields for each target were corrected by subtracting the appropriately weighted yield of the empty target flux. For much of the data sample, the deuterium target had a HD contamination of 8.4% ± 0.4% per molecule and the yields were corrected accordingly. The beam normalization and the uncertainty in the rate-dependence corrections were the dominant systematic errors. Other smaller contributions include the uncertainty in the purity of the deuterium target, uncertainties in the target density, and the contribution of the tail of the J/ψ and ψ' peaks to the di-muon mass spectrum above 4.5 GeV/c^2.

Instantaneous fluctuations in the beam intensity during data collection presented the main challenge in the data analysis. These fluctuations occurred at the accelerator frequency of 53 MHz and led to a luminosity-normalized rate-dependent variation in the yield of events from the deuterium and hydrogen targets for a number of different sources. Several approaches were considered to account for this variation. Generally, the simplest approach would be to reject any event produced when the accelerator was above a certain, arbitrary (fairly low) threshold, and absorb the remainder of the effect into a systematic uncertainty. This would have a substantial impact on the statistical significance of the data. A second approach would be to model each individual effect in Monte Carlo, then parameterize individual effects, and finally apply the combined parameterizations to the data. The systematic uncertainty would need to account for the accuracy of the model and for any still unknown effects. For the present data, a third method was chosen, which allowed the full statistical power of the data to be maintained without requiring each and every effect of the intensity variation to be modelled.

This method considers only the final result—the ratio of event yields between the two targets—as a function of intensity. For each x_i bin, the cross-section ratio was plotted as a function of the instantaneous beam intensity when that event occurred, as illustrated in Extended Data Fig. 2. The effect of the intensity dependence on the final result could then be parameterized from the measured data and then extrapolated to zero intensity. The simplicity of this method is that only the data are used to measure and correct for the intensity dependence. Because the x_i bin boundaries are arbitrary relative to the beam intensity, a smooth, common parameterization for the intensity dependence is to be expected. A variety of parametric forms were compared to the data. One such form is

\[
\frac{Y_d(x_i, I)}{2Y_d(x_i, I)} = R_x + aI + bI^2,
\]

where Y_d(I) is the luminosity-normalized, empty-target-subtracted yield of events from the deuterium (hydrogen) target. In this form, a and b are parameters of the fit that are common to all x_i bins describing the intensity, I, dependence, and R_x is the zero-intensity intercept for that bin. The intercepts resulting from the simultaneous fit of all x_i bins gives the cross-section ratio \(\sigma_d(2\sigma_h)\) for each bin. The common intensity parameters, a and b, correlate \(\sigma_d(2\sigma_h)\) for all bins and are also determined in the simultaneous fit. Other forms were also studied, including, for example,

\[
\frac{Y_d(x_i, I)}{2Y_d(x_i, I)} = R_x + (a_0 + a_1 x_i + (b_0 + b_1 x_i)^2,
\]

which allows for an x_i-correlated intensity dependence. An example of a less conventional extrapolation form that was considered is

\[
\frac{Y_d(x_i, I)}{2Y_d(x_i, I)} = R_x \cos \left( \frac{I}{a_0 + a_1 x_i} \right).
\]

In addition, constraining either a or b to zero and thus eliminating the I for dependence was explored. Using the Akaike information criterion to avoid over-parameterization, the form given in equation (6) was chosen for the extrapolation. The resulting fits from three representative x_i bins are shown in Extended Data Fig. 2. A comparison with a fit to the parameterization in equation (7) was used to estimate the systematic uncertainties. The covariance matrix for the intercepts R_x, resulting from the fit to equation (6) is

\[
\begin{pmatrix}
2.70 & 1.19 & 1.15 & 1.20 & 1.09 & 1.16 \\
1.19 & 1.87 & 1.25 & 1.31 & 1.19 & 1.26 \\
1.15 & 1.25 & 1.79 & 1.25 & 1.15 & 1.21 \\
1.20 & 1.31 & 1.25 & 2.14 & 1.20 & 1.27 \\
1.09 & 1.19 & 1.15 & 1.20 & 2.49 & 1.16 \\
1.16 & 1.26 & 1.21 & 1.27 & 1.16 & 4.06
\end{pmatrix} \times 10^{-3}.
\]
The same technique was independently applied to the data binned in transverse momenta, $P_T$, and mass, $M$.

The cross-section ratios shown in Fig. 1 and listed in Table 1 are not corrected for acceptance. To compare any calculation with the present data, it is necessary to consider the SeaQuest spectrometer's acceptance in $x_b$. The appropriate theoretical cross-section ratio may be calculated for an $x_b$ bin as

$$\left(\frac{\sigma_D}{2\sigma_H}\right)_{\text{calc}} = \frac{\sum A_j \sigma_{Dj}^\text{calc}(x_t, x_b, M)}{2 \sum A_j \sigma_{Hj}^\text{calc}(x_t, x_b, M)},$$

where the subscript $j$ denotes the $j$th $x_b$ sub-bin of the $i$th $x_t$ bin, and $A_j$ is the acceptance for that bin, tabulated in Extended Data Table 3. Finally, $\sigma_{Dj}^\text{calc}(x_t, x_b, M)$ is the calculated cross-section, where the dependence on $x_t, x_b$ and $M$ has been made explicit. The code used by SeaQuest for the NLO calculation of $\sigma_D(x, M)$ was provided by W. K. Tung of CTEQ.

Data availability

Raw data were generated at the Fermi National Accelerator Laboratory. Derived data supporting the findings of this study are available from the corresponding author upon request.

Acknowledgements

We thank G. T. Garvey for contributions to the early stages of this experiment. We also thank the Fermilab Accelerator Division and Particle Physics Division for their support of this experiment. This work was performed by the SeaQuest Collaboration, whose work was supported in part by the US Department of Energy under grants DE-AC02-06CH11357, DE-FG02-07ER41528, DE-SC0006963, the US National Science Foundation under grants PHY 0969239, PHY 1306126, PHY 1452636, PHY 1505458, PHY 1614456, the DP&A and ORED at Mississippi State University; the JSPS (Japan) KAKENHI through grant numbers 21244028, 25247037, 25800133, the Tokyo Tech Global COE Program, Japan; the Yamada Science Foundation of Japan; and the Ministry of Science and Technology (MOST), Taiwan. Fermilab is operated by Fermi Research Alliance, LLC, under contract number DE-AC02-07CH11359 with the US Department of Energy.

Author contributions

P.E.R. and D.F.G. are the co-spokespersons for the experiment. The entire SeaQuest Collaboration constructed the experiment and participated in the data collection and analysis. Substantial contributions to the cross-section ratio analysis were made by graduate students J.D., B.K., R.E.M., S.M., D.H.M., K. Nagai, S.P., F.S., M.B.C.S. and A.S.T. The development of the technique of extrapolation to zero intensity greatly benefited from the work of A.S.T. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.E.R.

Peer review information

Nature thanks Gerald Miller, Gunar Schnell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information

is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Comparison of NuSea and SeaQuest data with NLO calculations. a, b, Comparison of the data from the present work and the NuSea measurements with NLO calculations made at the integrated kinematics of SeaQuest (a) and average kinematics of NuSea (b) based on the CT18 and CTEQ6m parton distributions. Events in the SeaQuest data were produced by a 120-GeV proton beam, whereas in the NuSea data were from an 800-GeV beam. In addition, the spectrometers, although similar in concept, had different acceptances. As a consequence, the cross-section ratios, which convolve $x_t$ with $x_b$, are expected to differ because of their distinct distributions in accepted $x_b$. These kinematic effects can clearly be seen by the difference between the curves. Because an acceptance table analogous to Extended Data Table 3 was not available for NuSea, these calculations used $\langle x_t \rangle$, $\langle x_b \rangle$ and $\langle M \rangle$ of the NuSea data. Both CTQ6m and CT18 have included the NuSea data in their global analysis, so calculations based on those probability distribution functions are expected to agree better with the NuSea data. The red (violet) curve in a (b) is the same as that in Fig. 1a and is repeated here for comparison.
Extended Data Fig. 2 | Extrapolation to zero intensity. Extrapolation to zero intensity fits for representative $x_{\text{t}}$ bins (0.13 ≤ $x_{\text{t}}$ < 0.16 (a), 0.195 ≤ $x_{\text{t}}$ < 0.240 (b) and 0.290 ≤ $x_{\text{t}}$ < 0.350 (c)). The $I$ (intensity) and $I^2$ coefficients are common to all bins. $\chi^2$/d.o.f. = 38.7/40 for the simultaneous fit of all $x_{\text{t}}$ bins (d.o.f., degrees of freedom).
Extended Data Fig. 3 | Reconstructed invariant mass spectra.

**a, b.** Reconstructed muon-pair invariant mass spectra for the liquid hydrogen (a) and liquid deuterium (b) targets. In the lower mass region, the predominant signal is produced by $J/\psi \rightarrow \mu^+\mu^-$ decay, followed by $\mu^+\mu^-$ decay of $\psi'$. The prominence of the $J/\psi$ peak provides a calibration point for the absolute field of the solid iron magnet. At invariant masses above 4.5 GeV/c$^2$, the Drell–Yan process becomes the dominant feature. The data are shown as red points. Additionally, Monte Carlo (MC) simulated distributions of Drell–Yan, $J/\psi$ and $\psi'$, along with measured random-coincidence and empty-target backgrounds, are shown. The sum of these is shown in the blue solid curve labelled 'MC sum'. The normalizations of the Monte Carlo and the random background were from a fit to the data.
### Extended Data Table 1 | Ratios $\sigma_d/(2\sigma_H)$ as a function of $P_T$

| $P_T$ (GeV/c) | $\langle P_T \rangle$ (GeV/c) | $\sigma_D/(2\sigma_H)$ ± stat. ± syst. | $\delta P_T$ (GeV/c) |
|---------------|-------------------------------|---------------------------------|------------------|
| 0.0 – 0.3     | 0.198                         | 1.137 ± 0.049 ± 0.061           | 0.161            |
| 0.3 – 0.5     | 0.405                         | 1.174 ± 0.045 ± 0.052           | 0.177            |
| 0.5 – 0.7     | 0.599                         | 1.209 ± 0.046 ± 0.038           | 0.188            |
| 0.7 – 0.9     | 0.797                         | 1.210 ± 0.046 ± 0.045           | 0.194            |
| 0.9 – 1.2     | 1.035                         | 1.130 ± 0.043 ± 0.037           | 0.198            |
| 1.2 – 1.5     | 1.330                         | 1.287 ± 0.061 ± 0.094           | 0.201            |
| 1.5 – 1.8     | 1.625                         | 1.087 ± 0.078 ± 0.099           | 0.206            |
| 1.8 – 2.3     | 1.915                         | 0.838 ± 0.095 ± 0.162           | 0.204            |

Ratios $\sigma_d/(2\sigma_H)$ with their statistical and systematic uncertainties as a function of transverse momentum, $P_T$. The cross-section ratios are defined as the ratio of luminosity-corrected yields from the hydrogen and deuterium targets. The final column, $\delta P_T$, is the experimental resolution in $P_T$ as determined by Monte Carlo simulation.
Extended Data Table 2 | Ratios $\sigma_D/(2\sigma_H)$ as a function of $M$

| $M$ (GeV/$c^2$) | $\langle M \rangle$ (GeV/$c^2$) | $\sigma_D/(2\sigma_H)$ ± stat. ± syst. | $\delta M$ (GeV/$c^2$) |
|-----------------|-----------------------------|-------------------------------|---------------------|
| 4.4 – 4.6       | 4.55                        | 1.170 ± 0.053 ± 0.059         | 0.24                |
| 4.6 – 4.8       | 4.70                        | 1.204 ± 0.047 ± 0.039         | 0.24                |
| 4.8 – 5.0       | 4.90                        | 1.202 ± 0.048 ± 0.039         | 0.25                |
| 5.0 – 5.2       | 5.10                        | 1.163 ± 0.050 ± 0.039         | 0.26                |
| 5.2 – 5.5       | 5.34                        | 1.123 ± 0.046 ± 0.037         | 0.26                |
| 5.5 – 6.5       | 5.89                        | 1.183 ± 0.043 ± 0.042         | 0.28                |
| 6.5 – 8.8       | 6.91                        | 1.167 ± 0.068 ± 0.096         | 0.30                |

Ratios $\sigma_D/(2\sigma_H)$ with their statistical and systematic uncertainties as a function of mass, $M$. The cross-section ratios are defined as the ratio of luminosity-corrected yields from the hydrogen and deuterium targets. The final column, $\delta M$ is the experimental resolution in $M$ as determined by Monte Carlo simulation.
Extended Data Table 3 | Spectrometer acceptance

| $x_b$ | 0.30–0.35 | 0.40–0.45 | 0.50–0.55 | 0.60–0.65 | 0.70–0.75 | 0.80–0.85 |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0.130–0.160 | 0.0007 | 0.0071 | 0.0188 | 0.0299 | 0.0366 | 0.0432 |
|        | 0.589   | 0.628   | 0.675   | 0.723   | 0.772   | 0.820   |
|        | 0.158   | 0.153   | 0.148   | 0.144   | 0.138   | 0.134   |
|        | 4.54    | 4.60    | 4.68    | 4.77    | 4.92    | 5.07    |
| 0.160–0.195 | 0.0001 | 0.0023 | 0.0105 | 0.0205 | 0.0298 | 0.0384 |
|        | 0.394   | 0.433   | 0.477   | 0.524   | 0.574   | 0.623   |
|        | 0.191   | 0.184   | 0.178   | 0.176   | 0.176   | 0.175   |
|        | 4.56    | 4.63    | 4.74    | 4.91    | 5.09    | 5.27    |
| 0.195–0.240 | 0.0015 | 0.0078 | 0.0176 | 0.0270 | 0.0364 | 0.0436 |
|        | 0.383   | 0.427   | 0.475   | 0.524   | 0.574   | 0.623   |
|        | 0.235   | 0.225   | 0.217   | 0.216   | 0.215   | 0.214   |
|        | 4.55    | 4.65    | 4.78    | 4.99    | 5.21    | 5.43    |
| 0.240–0.290 | 0.0035 | 0.0120 | 0.0207 | 0.0298 | 0.0379 | 0.0455 |
|        | 0.341   | 0.379   | 0.426   | 0.475   | 0.524   | 0.574   |
|        | 0.324   | 0.319   | 0.316   | 0.316   | 0.315   | 0.315   |
|        | 4.95    | 5.18    | 5.46    | 5.76    | 6.05    | 6.33    |
| 0.290–0.350 | 0.0062 | 0.0125 | 0.0203 | 0.0268 | 0.0336 | 0.0405 |
|        | 0.339   | 0.377   | 0.425   | 0.474   | 0.524   | 0.573   |
|        | 0.384   | 0.390   | 0.386   | 0.386   | 0.384   | 0.384   |
|        | 5.38    | 5.72    | 6.04    | 6.38    | 6.69    | 7.00    |
| 0.350–0.450 | 0.0052 | 0.0125 | 0.0203 | 0.0268 | 0.0336 | 0.0405 |
|        | 0.339   | 0.377   | 0.425   | 0.474   | 0.524   | 0.573   |
|        | 0.384   | 0.390   | 0.386   | 0.386   | 0.384   | 0.384   |
|        | 5.38    | 5.72    | 6.04    | 6.38    | 6.69    | 7.00    |

The acceptance relative to a 4π detector and average kinematic values for bins in $x_t$ and $x_b$. Each cell shows, from top to bottom, the acceptance, $\langle x_b \rangle$, $\langle x_t \rangle$ and average mass for each sub-bin.