The Flavor Asymmetry of the Light Quark Sea
from Semi-inclusive Deep-inelastic Scattering

The flavor asymmetry of the light quark sea of the nucleon is determined in the kinematic range $0.02 < x < 0.3$ and 1 GeV$^2 < Q^2 < 10$ GeV$^2$, for the first time from semi-inclusive deep-inelastic scattering. The quantity $(\bar{d}(x) - \bar{u}(x))/((u(x) - d(x)))$ is derived from a relationship between the yields of positive and negative pions from unpolarized hydrogen and deuterium targets. The flavor asymmetry $\bar{d} - \bar{u}$ is found to be non-zero and $x$ dependent, showing an excess of $\bar{d}$ over $\bar{u}$ quarks in the proton.

The flavor content of the nucleon sea has come to be recognized as an important domain for testing models of nucleon structure [1]. Until recently there has been little experimental constraint on the flavor asymmetry in the quark distributions of the light sea quarks in the nucleon. The first was an integral test, based on a comparison of inclusive deep-inelastic scattering on the proton and the deuteron. A flavor symmetric sea, $\int_0^1 dx \left( F^p_q(x) - F^n_q(x) \right)$, where $F^p_q(x)$ and $F^n_q(x)$ are the structure functions of the proton and neutron, respectively. The assumption of isospin symmetry in the quark-parton model allows the Sum to be written as

$$ S_G = \frac{1}{3} \int_0^1 dx \left( u_v(x) - d_v(x) \right) dx - \frac{2}{3} \int_0^1 (\bar{d}(x) - \bar{u}(x)) dx, $$

where $u_v(x)$ and $d_v(x)$ are the density functions of the valence quarks and $\bar{u}(x)$ and $\bar{d}(x)$ those of the antiquarks in the proton. A flavor symmetric sea, $\bar{u}(x) = \bar{d}(x)$, leads to the Gottfried Sum $S_G = 1/3$. A measurement of the Gottfried Sum by NMC [3] resulted in $S_G = 0.235 \pm 0.026$. If isospin symmetry holds, a global flavor asymmetry $\int_0^1 (\bar{d}(x) - \bar{u}(x)) dx \approx 0.15$ would account for the NMC result. Many ideas such as Pauli blocking and pion clouds have been proposed to explain such a flavor asymmetry in the sea [1]. Two methods have been proposed to measure its $x$ dependence: the Drell-Yan process [4] and semi-inclusive deep inelastic scattering [5].

Results are presented here for the $x$ dependence of $(\bar{d} - \bar{u})/(u - d)$ from the analysis of charged pion yields in semi-inclusive deep-inelastic scattering of positrons on unpolarized hydrogen and deuterium targets. The data cover the kinematic range $0.02 < x < 0.3$ and 1 GeV$^2 < Q^2 < 10$ GeV$^2$, where $-Q^2$ is the four-momentum transfer squared of the exchanged virtual photon and $x = Q^2/(2M\nu)$ is the Bjorken variable with $M$ being the proton mass, $\nu = E - E'$ the virtual photon energy, and $E(E')$ the energy of the incident (scattered) positron.

The ratio $(\bar{d} - \bar{u})/(u - d)$ is determined from a quantity that combines good sensitivity to the sea asymmetry with minimal sensitivity to instrumental effects. This is the ratio of the differences between charged pion yields for proton and neutron targets [5]:

$$ r(x, z) = \frac{N_p^\pi^-(x, z) - N_n^\pi^-(x, z)}{N_p^\pi^+(x, z) - N_n^\pi^+(x, z)}, $$

where $z = E^\pi/\nu$ is the fraction of the virtual photon energy carried by the pion and $N^\pi(x, z)$ is the yield of pions coming from deep-inelastic scattering off nucleons. Factorization between the hard scattering process and the hadronization of the struck quark implies

$$ N^\pi^{\pm}(x, z) \propto \sum_i e_i^2 \left[ q_i(x) D_i^{\pi^\pm}(z) + \bar{q}_i(x) D_{\bar{i}}^{\pi^\pm}(z) \right], $$

where $e_i$ is the quark charge in units of the elementary charge, $q_i(x)$ and $\bar{q}_i(x)$ are the density distributions of quarks and antiquarks of flavor $i$, and the fragmentation functions $D_i^{\pi^\pm}(z)$ represent the probability that the quark of flavor $i$ fragments to a charged pion. Assuming isospin symmetry between protons and neutrons as well as charge conjugation invariance, the number of light quark fragmentation functions is reduced to two: the favored and disfavored fragmentation functions [6], e.g. $D_u^{\pi^+}$ and $D_n^{\pi^-}$. Using Eq. 4 to express $r(x, z)$ in terms of quark distributions and fragmentation functions results in

$$ \frac{1 + r(x, z)}{1 - r(x, z)} = \frac{u(x) - d(x) + \bar{u}(x) - \bar{d}(x)}{u_v(x) - d_v(x)} J(z), $$

where $J(z) = \frac{2}{z} \left( 1 + D'(z) \right)$ and $D'(z) = D_u^{\pi^+}(z)/D_n^{\pi^-}(z)$. It should be noted that the right-hand side of Eq. 5 factorizes into two independent functions of $x$ and $z$ respectively. Thus, the equation may be rearranged to isolate a quantity sensitive to the flavor asymmetry:
\[ \frac{d(x) - \pi(x)}{u(x) - d(x)} = \frac{J(z)[1 - r(x, z)] - [1 + r(x, z)]}{J(z)[1 - r(x, z)] + [1 + r(x, z)]} \]  

(6)

The assumptions leading to Eq. 5 imply that the expression on the right side of Eq. 6 is independent of \( z \).

In the HERMES experiment, 27.5 GeV positrons circulating in the HERA storage ring at DESY are scattered on internal hydrogen (H\(_2\)), deuterium (D\(_2\)), and helium-3 (\(^3\)He) targets. The target gas is fed into the center of a 40-cm-long open-ended storage cell. Systematic uncertainties are minimized by cycling the three gases through various permutations. This cycling was performed as often as six times per HERA fill, which lasts typically 8 to 12 hours. Using unpolarized gases, luminosities of \((1 - 5) \times 10^{32}\) cm\(^{-2}\)s\(^{-1}\) are obtained, which are an order of magnitude larger than available with the polarized targets normally employed at HERMES. The luminosity is monitored by a pair of NaBi(WO\(_4\))\(_2\) electromagnetic calorimeters that detect Bhabha-scattered target leptons.

A detailed description of the HERMES spectrometer is provided elsewhere [7]. It is a forward spectrometer that identifies the scattered positron as well as the associated hadrons, in the scattering angle range of 0\(^\circ\) to 22 rad. Positron-hadron discrimination is based on information from four particle identification detectors: a threshold gas Čerenkov counter, a transition-radiation detector, a lead-glass electromagnetic calorimeter, and a preshower detector located directly before the calorimeter. Positrons are distinguished from hadrons with a likelihood method that combines the response from all four particle identification detectors. This particle identification system provides an average positron identification efficiency of 99%, with a hadron contamination that is dependent on the kinematics of the positron but is always less than 1%. One of the features of the spectrometer central to this work is its threshold Čerenkov counter, which distinguishes pions from heavier hadrons for particles with momenta greater than 3.8 GeV while maintaining better than 95% pion identification efficiency at momenta larger than 6 GeV.

The kinematic requirements used in this analysis are \( Q^2 > 1 \) GeV\(^2\), invariant mass \( W \) of the initial photon-nucleon system greater than 2 GeV, and fractional energy \( y = \nu/E \) of the incident lepton transferred to the virtual photon less than 0.85. Events are required to originate from the target-beam interaction region. For the determination of \( r(x, z) \), the data are partitioned in five bins of \( x \) in the range 0.02 < \( x \) < 0.3 and up to six bins in \( z \) in the range 0.2 < \( z \) < 0.8. The pions are selected to be in the current fragmentation region (i.e. derived from the struck quark and not target fragments) by requiring \( x_F = 2P_t/W \) to be greater than 0.25, where \( P_t \) is the momentum component of the pion in the longitudinal direction with respect to the virtual photon in the photon-nucleon center-of-mass frame. Data at Bjorken \( x > 0.3 \) are excluded because they are dominated by scattering from the valence quarks. The yields for the neutron target appearing in Eq. 3 can be inferred from yields obtained from deuterium and hydrogen targets, \( N^{-d}_{\pi^+} - N^{-d}_{\pi^-} \). Thus the determination of \( (d(x) - \pi(x))/(u(x) - d(x)) \) is made by taking ratios of differences between yields from hydrogen and deuterium. The pion yields extracted from the data are \( N^{-d}_{\pi^+}(x, z)/L \) where \( N^{-d}_{\pi^+}(x, z) \) is the number of detected pions associated with deep-inelastic scattering events, and \( L \) is the integrated luminosity (nuclei/cm\(^2\)s) corrected for instrumental dead-time and inefficiencies; the possible effect of different acceptances for \( \pi^+ \) and \( \pi^- \) is included in the systematic uncertainties discussed below. The fragmentation function ratio \( D^z(x) \) for pions was determined from each of the unpolarised targets: H\(_2\), D\(_2\) and \(^3\)He [8]. No statistically significant differences were found. To avoid statistical correlations between the fragmentation functions and the pion ratios, only \( D^z(x) \) extracted from \(^3\)He is used in this analysis. The fragmentation function ratio \( D^z(x) \) determined within the HERMES acceptance is found to vary with \( x \) and \( Q^2 \) typically by less than 10%, except in the low-yield region near \( z = 0.8 \) where as much as 40% may be seen. However, extraction of the flavor asymmetry in bins of \( x \) and \( Q^2 \) revealed the averaged result to be unaffected by this variation. The dependence of the left-right instrumental asymmetry on pion charge is significant only at the lowest momenta and is taken into account as a correction. For all \( x \) bins, the resulting change in the sea asymmetry is less than 1.5%.

The results for the quantity \( (d - \pi)/(u - d) \) are shown in Fig. 1 as a function of \( z \) for five bins in \( x \). The data show no \( z \) dependence and are consistent with the form of factorization shown in Eq. 5. The values of \( (d - \pi)/(u - d) \) averaged over \( z \) are plotted in Fig. 2(a) as a function of \( x \), and are presented in Table I. Within the measured \( x \) region, the values are non-zero and positive everywhere, clearly showing an excess of \( d \) quarks over \( \pi \) quarks in the proton. Also included in this figure are the GRV 94 LO [9], CTEQ 4L [10], MRS (A) [11] low \( Q^2 \), and MRST (98) [12] parameterizations of \( (d - \pi)/(u - d) \), calculated at the average \( Q^2 \) of the events in each \( x \) bin. The HERMES results are determined without making higher order QCD corrections and so only leading order parameterizations should be compared to these results. At low \( Q^2 \), only GRV 94 is available at leading order. The distribution \( d - \pi \) as a function of \( x \) is thus derived from \( (d - \pi)/(u - d) \) using the GRV 94 LO parameterization of \( u_{\nu}(x) - d_{\nu}(x) \), again calculated at the appropriate \( Q^2 \) for each \( x \) bin. The results are presented in Fig. 2(b) and Table I, and give an integral of \( d - \pi \) over the measured \( x \) region of \( \int_{0.02}^{0.3} (d(x) - \pi(x))dx = 0.103 \pm 0.017{\text{(stat)} \pm 0.015{\text{(sys)}}} \). When the data are evolved to a \( Q^2 \) of 2.5 GeV\(^2\), the integral over the measured \( x \) region is \( \int_{0.02}^{0.3} (d(x) - \pi(x))dx = 0.107 \pm 0.018{\text{(stat)} \pm 0.015{\text{(sys)}}} \). The GRV 94 LO param-
eratorization is used to estimate the portion of the integral in the unmeasured range. In the region $x < 0.02$, the contribution to the integral $\int (|d(x) - \bar{d}(x)|)dx$ is significant: $0.048$ (almost 50% of that over the measured $x$ range), while the high $x$ region $x > 0.3$ contributes only $0.006$. Including an extrapolation uncertainty estimate based on differences among the available parameterizations, the total integral over all $x$ is then approximately $0.16 \pm 0.03$, which is consistent with the results first seen by NMC [3].

An alternate approach to the measurement of the flavor asymmetry is the comparison of the Drell-Yan process on protons and deuterons, which is sensitive to the quan-

tity $NMC$ [3].

rameterization of the CTEQ $4M$ [10] parameterization of $\bar{d}/d$ to extract $\bar{d} - \bar{\pi}/\bar{\pi}$ from the measured quantity $d/\pi$. These results given at a $Q^2$ of $54$ GeV$^2$ are included in Fig. 2(b). This figure demonstrates that the sea asymmetry measured in deep-inelastic scattering and in Drell-Yan experiments agree, even though the $Q^2$ of the two experiments differ by a factor of about 20. Thus, the $x$ dependence of $(\bar{d} - \bar{\pi})/(u - d)$ presented here will serve as a further constraint on global parameterizations of parton density functions.

The total systematic uncertainties associated with the determination of the quantity $(\bar{d} - \bar{\pi})/(u - d)$ in the present work are given in Table I. These uncertainties include various effects as follows. After the correction at low momenta, any residual charge-dependent instrumental acceptance asymmetry is estimated to be less than 1.5%. Effects of the finite resolution of the spectrometer and bremsstrahlung have been investigated with Monte Carlo studies. These effects are small and are simply included in the systematic uncertainty. Radiative corrections to the pion yields have been calculated following the prescription of reference [15]. Their effect on the pion ratio in Eq. 3 is estimated to be less than 1% and is included in the systematic uncertainty. Lepton contamination in the pion sample is found to be at most of order one percent, and the differences between the two targets were no more than 0.2%. The effects of hadrons misidentified as pions have been estimated from the Monte Carlo simulations and are also included. Finally, the uncertainty in the fragmentation function ratio $D'(z)$ as measured at HERMES has been accounted for. The statistical uncertainty of $D'(z)$ has been included in quadrature to the statistical uncertainty of the pion yields. The systematic component of this ratio determined from systematic studies of the HERMES data has been included in quadrature to the total uncertainty.

The only additional systematic uncertainty associated with $\bar{d} - \bar{\pi}$ is the choice of parameterization to extract $\bar{d} - \bar{\pi}$ from $(\bar{d} - \bar{\pi})/(u - d)$. The GRV 94 LO [9] parameterization of $u_0(x) - d_0(x)$ is used. The MRS (A) [11] low $Q^2$ parameterization is used as an alternative. The resulting difference in the value of $\bar{d} - \bar{\pi}$ is used as an estimate of this uncertainty.

In summary, the flavor asymmetry of the light quark sea is extracted for the first time from pion yields in semi-inclusive deep-inelastic scattering. The data are recorded at the HERMES experiment in the kinematic range of $0.02 < x < 0.3$ and $1$ GeV$^2 < Q^2 < 10$ GeV$^2$. The quantity $\bar{d} - \bar{\pi}$ is found to be non-zero over the entire $x$ range measured, clearly showing an excess of $\bar{d}$ quarks over $\bar{\pi}$ quarks in the proton sea. The integral over the measured region is $\int_{0.02}^{0.3}(\bar{d}(x) - \bar{\pi}(x))dx = 0.107 \pm 0.018$(stat) $\pm 0.015$(sys) at a $Q^2_o$ of $2.5$ GeV$^2$, which accounts for two-thirds of the Gottfried Sum Rule deficit.

We gratefully acknowledge the DESY management for its support and the DESY staff and the staffs of the collaborating institutions. This work was supported by the FWO-Flanders, Belgium; the Natural Sciences and Engineering Research Council of Canada; the INTAS, HCM, and TMR network contributions from the European Community; the German Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie; the Deutscher Akademischer Austauschdienst (DAAD); the Italian Istituto Nazionale di Fisica Nucleare (INFN); Monbusho, JSPS, and Toray Science Foundation of Japan; the Dutch Foundation for Fundamenteel Onderzoek der Materie (FOM); the U.K. Particle Physics and Astronomy Research Council; and the U.S. Department of Energy and National Science Foundation.

* Deceased.


<table>
<thead>
<tr>
<th>$x$ range</th>
<th>$&lt; x &gt;$</th>
<th>$&lt; Q^2 &gt;$ (GeV$^2$)</th>
<th>$\frac{\bar{d} - \bar{u}}{u - d}$</th>
<th>Stat. error</th>
<th>Sys. error</th>
<th>$\frac{\bar{d} - \bar{u}}{u - d}$</th>
<th>Stat. error</th>
<th>Sys. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.020-0.050</td>
<td>0.038</td>
<td>1.33</td>
<td>0.20</td>
<td>0.08</td>
<td>0.07</td>
<td>0.59</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>0.050-0.075</td>
<td>0.062</td>
<td>1.82</td>
<td>0.42</td>
<td>0.08</td>
<td>0.06</td>
<td>0.85</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>0.075-0.115</td>
<td>0.092</td>
<td>2.52</td>
<td>0.15</td>
<td>0.07</td>
<td>0.06</td>
<td>0.33</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>0.115-0.150</td>
<td>0.131</td>
<td>3.38</td>
<td>0.20</td>
<td>0.08</td>
<td>0.05</td>
<td>0.36</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>0.150-0.300</td>
<td>0.198</td>
<td>4.88</td>
<td>0.17</td>
<td>0.07</td>
<td>0.05</td>
<td>0.25</td>
<td>0.09</td>
<td>0.08</td>
</tr>
</tbody>
</table>
FIG. 1. The distribution $(\bar{d} - u)/(u - d)$ as a function of $z$ in five bins of $x$. The points are fit to a constant for each $x$ bin. The error bars represent statistical and systematic uncertainties added in quadrature. Some points are omitted due to limited statistics.
(a) \((\bar{d} - \bar{u})/(u - d)\) as a function of \(x\). Also included are the GRV 94 LO [9], CTEQ 4lq [10], MRS (A) [11] low \(Q^2\), and MRST (98) [12] parameterizations calculated at the appropriate \(Q^2\) for each \(x\)-bin. (b) \(\bar{d} - \bar{u}\) as a function of \(x\). The curve is the GRV 94 LO [9] parameterization. The open circles represent the E866 [14] determination of \(\bar{d} - \bar{u}\). The inner error bars represent the statistical uncertainties while for all data including those of E866, the total error bars represent statistical and systematic uncertainties added in quadrature.