Nucleon Structure Functions: Experiments and Models

S.E. Kuhn
Old Dominion University, Norfolk, VA 23529, USA
E-mail: kuhn@jlab.org

In these lectures, I describe measurements of the polarized and unpolarized structure functions of the proton and the neutron. The main emphasis is on a clear, straightforward description of what is being measured and how it can be interpreted in a simple quark model of the nucleon. The intended audience are graduate students who are beginning their work in Nuclear or Particle Physics. More advanced students are referred to the extensive literature.

1 Introduction

Deep inelastic structure functions of the nucleon have been studied for three decades. The initial measurements at the Stanford Linear Accelerator Center (SLAC) confirmed the quark-parton picture of the nucleon. Since then, more and more precise measurements have been conducted at several accelerators (SLAC, CERN, Fermilab, DESY), and there seems to be no end to new and surprising results, from the original "EMC-Effect" (the modification of nucleon structure functions in nuclei) over the violation of the Gottfried sum rule (the apparent breakdown of flavor symmetry in the quark-antiquark sea of the proton) to the most recent indications of quark substructure at HERA/DESY.

Polarized structure functions have been measured since the late 1970's, and these measurements have also contributed their share of surprising results, in particular the so-called "spin crisis". The study of both polarized and unpolarized structure functions remains very much an active field, with new theoretical papers appearing at a rate of several per week. Experimental results are published at only slightly longer intervals (typically each of the active experimental collaborations publishes several new papers every year), and even new comprehensive overviews of the field are written every few months (see, e.g., the reviews by Cooper-Sarkar et al. [1] for unpolarized structure functions and Ramsey or Böttcher [2] for polarized ones). There are no signs of a slow-down, with several ongoing and new experiments underway and new initiatives planned.

In this context, yet another survey or review article would not make much sense. Instead, the purpose of this article is to give a simple, intuitive introduction into the physics and the experimental techniques of structure function measurements, and to offer a starting point for more in-depth study of the vast literature. It is based on six lectures on the subject given by the author during the 1997 Hampton University Graduate Summer School (HUGS '97). The intended audience includes graduate students who are relatively new to the field of Particle Physics. A suitable introductory textbook would be that by Halzen [3], by Povh et al. [4] or by Wong [5]. Additional useful background material can be found in the Particle Data Group Collection [6] and the monograph by Roberts [7], as well as the conference proceedings of many recent conferences [8,9].

In the following section, I introduce a simple "toy model" for the quark structure of the nucleon, and derive some relevant properties of spin-dependent observables in this model. In Section 3, I give a coherent description of electron scattering in the elastic, resonance and deep inelastic regime. The emphasis is again on a simple(–minded?) interpretation of the measured quantities. In Section 4, I discuss modifications of this simple picture at finite momentum transfer, including scaling violations and higher-twist effects. The next two sections (5 and 6) treat spin structure functions, both in the scaling limit and at lower momentum transfer, and in the final section (7) I give a brief outlook on future planned experiments.

2 A Simple Quark Model of Nucleons and other Baryons

One of the original successes of the quark model was the explanation of the "light" ground-state baryons ($p, n, \Sigma^-, \Sigma^0, \Sigma^+, \Lambda^0, \Xi^-$ and Ξ^0) and their excited states (Δ's, Σ^*'s and Ξ^*'s) as bound states of just three different types of quarks, U (up), D (down) and S (strange). These 3 quarks could be understood as members of a 3-dimensional (triplet) representation of the flavor symmetry group $SU(3)$ corresponding to the Isospin and Strangeness quantum numbers in hadrons, assigning them quantum numbers as shown in Table 1. In this picture, the proton (with Charge $Q = 1$ and Baryon Number $B = 1$) can be described as a combination of 2 up and 1 down quark ($|UUD\rangle$), while the neutron has 1 up and 2 down quarks ($|UUD\rangle$). Assigning each quark a spin of $S_q = 1/2$ could also accommodate the overall $S = 1/2$ spin of the ground state baryons above as well as the overall $S = 3/2$ spin of excited baryons like the Δ; however, only after introduction of the additional Color quantum number (r, g and b) could one understand the internal wave function of these baryons. By assuming that the quark wave function in each case is totally antisymmetric in the Color quantum number, one can constrain the combination of the remaining quantum numbers (Flavor, Spin and Angular Momentum) to be totally symmetric. In the ground state, total angular momentum is assumed to be zero (therefore symmetric), which leaves us with the task to construct fully symmetric wave functions in spin$\times$flavor space.

This is easy in the case of the excited baryon states, where the total spin

Table 1: Quantum numbers of the three lightest quarks.

Flavor	Isospin I	I_3	Strangeness S	Charge Q	Baryon Number B
U	$1/2$	$+1/2$	0	$+2/3$	$1/3$
D	$1/2$	$-1/2$	0	$-1/3$	$1/3$
S	0	0	-1	$-1/3$	$1/3$

of $S = 3/2$ requires a wave function which is separately symmetric in spin and flavor. In a (hopefully) intuitive short hand notation, we can therefore write the wave function of the Δ^{++} as $|\Delta^{++}\uparrow\rangle = |U\uparrow U\uparrow U\uparrow\rangle$ and that of the Δ^- as $|\Delta^-\uparrow\rangle = |D\uparrow D\uparrow D\uparrow\rangle$. [a] States with more than one type of quark are only slightly more complicated, e.g., the Δ^+ can be written as

$$|\Delta^+\uparrow\rangle = 1/\sqrt{3}\left(|U\uparrow U\uparrow D\uparrow\rangle + |U\uparrow D\uparrow U\uparrow\rangle + |D\uparrow U\uparrow U\uparrow\rangle\right). \quad (1)$$

However, from now on we will use the more simple form, for instance $|\Delta^+\uparrow\rangle = |U\uparrow U\uparrow D\uparrow\rangle$, where a symmetrization over all flavors is understood implicitly.

The case of the proton is a bit more complicated, since the wave function cannot be symmetric in spin and flavor separately. The most intuitive way to derive the proton wave function is by observing that 2 of the 3 quarks are equal (U), and therefore their relative spin wave function should be symmetric also. This leads to the conclusion that the two U–quarks couple their spins to a total spin of one. Let's denote the case where this spin has a z-projection of $+1$ as $(UU\Uparrow) := |U\uparrow U\uparrow\rangle$, while the projection with $S_z = 0$ will be indicated by $(UU\Rightarrow) := 1/\sqrt{2}\left(|U\uparrow U\downarrow\rangle + |U\downarrow U\uparrow\rangle\right)$. We can now combine the spin $1/2$ of the remaining D quark with the spin 1 of the UU pair in two ways to get total spin and projection $1/2$; the proper way follows simply from insertion of the correct Clebsch-Gordon coefficients:

$$|P\uparrow\rangle = 1/\sqrt{3}\left(\sqrt{2}|(UU\Uparrow)D\downarrow\rangle - |(UU\Rightarrow)D\uparrow\rangle\right). \quad (2)$$

The neutron wave function can be gotten from Eq. 2 by replacing all U's with D's and vice versa (and inserting an overall minus sign).

Once in hand, one can use these wave functions to try and explain some of the other well-known properties of the nucleons, for instance their anomalous magnetic moments. Relativistic quantum mechanics predicts that the magnetic moment for a pointlike particle with charge Z, spin S and mass M_N should be $\mu = Z\mu_N 2\mathbf{S}$, where $\mu_N = e/2M_N$ is the (nuclear) magneton. For the proton and the neutron one finds experimentally $\mu = (Z+\kappa_N)\mu_N 2\mathbf{S}$ ($Z = 1$ for

[a] These wave functions are for the case $S_z = S$; wave functions with different spin projections can be derived from this form by using the spin lowering operator σ^-.

the proton and $Z = 0$ for the neutron) with the anomalous magnetic moments for proton and neutron, $\kappa_p = 1.79$ and $\kappa_n = -1.91$. This clearly points to the composite character of these particles.

If we want to explain these magnetic momenta in terms of the quark ones, we need to know what to use for their masses (assuming the quarks themselves follow the rule for pointlike objects). An "obvious" choice would be $M_q =$ 330 MeV for both types of quarks ($q = U$ and $q = D$), since the mass of the nucleon could be explained in this way as simply the sum of the masses of its three constituent quarks, minus a small binding energy (also, this choice gives us the right answer – see below). The predicted magnetic moment of the up and down quarks in this picture would be $\mu_q = z_q(M_N/m_q)\mu_N$, which yields $\mu_U = \frac{2}{3}(939 \text{ MeV}/330 \text{ MeV})\mu_N = 1.897\mu_N$ for the up quark and $\mu_D = -\frac{1}{3}(939 \text{ MeV}/330 \text{ MeV})\mu_N = -0.949\mu_N$ for the down quark.

From Eq. 2, we take the probability of finding an up quark in the proton with its spin pointing up, minus the probability for the spin to point down, to be $\Delta U = \frac{2}{3}2 + \frac{1}{3}0 = 4/3$, while the corresponding probability difference for the down quark is $\Delta D = \frac{2}{3}(-1) + \frac{1}{3}1 = -1/3$. Multiplying this with the magnetic moments from above, we get $\mu_p = \frac{4}{3}1.897\mu_N - \frac{1}{3}(-0.949\mu_N) = 2.85\mu_N$, corresponding to a "predicted" $\kappa_p = 1.85$. For the neutron, the roles of ΔU and ΔD are simply interchanged (this follows from Isospin symmetry and is true even if the actual quark wave functions are much more complicated, as surely they are). This yields $\mu_n = \frac{4}{3}(-0.949\mu_N) - \frac{1}{3}1.897\mu_N = -1.90\mu_N$, in nearly perfect agreement with the measured κ_n.

Obviously, this good agreement in both cases can be "fine-tuned" by the choice of quark masses (one can get exactly the correct values for κ_p and κ_n if one allows for a small mass difference between U and D quarks). However, the remarkable result is that the required quark masses are perfectly reasonable in the overall frame of the Constituent Quark Model (CQM), since they also explain (approximately) the mass of the proton. This agreement is actually very surprising, as we will soon see, since constituent quarks really aren't elementary objects at all (they are more like "quasi-particles" with a rich internal structure of their own). The CQM is much less successful in explaining other static properties of the nucleon, for example the axial vector coupling constant g_A measured in β–decay. We now turn our attention to this quantity.

The hadronic part of the matrix element for weak charged current transitions like β–decay or electron capture contains two parts, one which transforms like a vector (V) and one which transforms like an axial vector (A). The combination $V - A$ of these two parts, which behave in opposite ways under the parity operation (coordinate inversion), is responsible for the fact that the weak interaction violates parity. The vector part of the hadronic current can be written as $V = g_V \langle n|\tau^-|p\rangle$ [b], where the isospin lowering operator τ^- turns the proton into a neutron (which means that the matrix element $\langle n|\tau^-|p\rangle$ simply equals to one). The factor g_V accounts for the fact that the weak interaction acts on the level of quarks, not nucleons. In our example, electron capture on the proton, the elementary process is the conversion of an U quark into a D quark, with simultaneous emission of an electron–neutrino. Therefore, we can write $V = \langle n|\sum_q \tau_q^-|p\rangle$, where the sum goes over all quarks in the nucleon and $\tau_q^-|U\rangle = |D\rangle$ and $\tau_q^-|D\rangle = 0$. This expression can be evaluated directly within our model, using Eq. 2 (the reader is strongly encouraged to try this). However, it is more instructive to make use of an important result from (Isospin–$SU(2)$) group theory: The matrix element of any isospin-lowering operator τ^- between two members of an iso–doublet (e.g., the proton and the neutron or the U and D quarks) yields the same result as the matrix element of the operator τ^3 sandwiched between two $I_3 = +1/2$ states:

$$\langle n|\sum_q \tau_q^-|p\rangle = \langle p|\sum_q \tau_q^3|p\rangle. \tag{3}$$

The operator τ_q^3 simply "measures" twice the third component of the isospin for each quark q ($\tau^3|U\rangle = +1$, $\tau^3|D\rangle = -1$), and the right hand side of Eq. 3 therefore yields simply twice the I_3 quantum number of the proton ($2(1/2 + 1/2 - 1/2) = 1$). This must be equal to $g_V \langle n|\tau^-|p\rangle = g_V$, and we can conclude that g_V must be equal to 1. Since the third component of the isospin is directly related to the charge of a particle ($Q = B/2 + I_3$), our conclusion follows simply from the fact that charge is conserved (no matter how many quarks there are in a proton, their total charge must add up to 1 and their total baryon number to 1 also). This is the reason why the simple fact that $g_V = 1$ is often referred to as "Charged Vector Current Conservation" (CVC).

Things are a bit more complicated in the case of the axial vector, **A**. The three spatial components of this axial 4–vector [c] are given by $A_i = g_A\langle n|\sigma^i\tau^-|p\rangle$, where the Pauli spin matrices σ^i act on the *spin* wave function in the same way the τ's act on isospin. Again, the fundamental matrix element is to be taken at the quark level, $A_i = \langle n|\sum_q \tau_q^- {\sigma^i}_q|p\rangle$. For our purpose, it suffices to look at only one component of the axial vector, the z-component. [d] In addition, we can also use the Isospin symmetry argument from before to

[b] This is actually the zero-component of a 4-vector, which is the only relevant contribution at very low energies, as in the case of nuclear β–decay.
[c] Again, this is the dominant term at low energies.
[d] The others can be related to A_z using spin operators.

finally write

$$g_A \langle p \uparrow | \sigma^z \tau^3 | p \uparrow \rangle = A_z = \langle p \uparrow | \sum_q \tau_q^3 {\sigma^z}_q | p \uparrow \rangle. \tag{4}$$

The left hand side simply gives g_A, but the right hand side is *not* equal to one. The operator combination $\tau_q^3 {\sigma^z}_q$, applied to a given quark, yields the product of twice its isospin projection times twice its spin projection:

$$\begin{aligned} &\tau_q^3 {\sigma^z}_q |U \uparrow\rangle = +|U \uparrow\rangle, \; \tau_q^3 {\sigma^z}_q |U \downarrow\rangle = -|U \downarrow\rangle, \\ &\tau_q^3 {\sigma^z}_q |D \uparrow\rangle = -|D \uparrow\rangle, \; \tau_q^3 {\sigma^z}_q |D \downarrow\rangle = +|D \downarrow\rangle. \end{aligned} \tag{5}$$

In other words, the result for the U quarks in the proton alone gives the same quantity ΔU that we defined before, and for the D quarks we get $-\Delta D$. Our final result is therefore $g_A = \Delta U - \Delta D = 4/3 - (-1/3) = 5/3$, indeed not unity (the reader is again urged to calculate this number "by hand", starting directly from Eq. 4). It is important to note that the relationship $g_A = \Delta U - \Delta D$ is true regardless of the "true" wave function of the proton (it follows from Isospin symmetry and the assumption that the axial vector is given by quark degrees of freedom, as in Eq. 4). Only the numerical result (5/3) depends on our explicit assumption about the make–up of the proton, and it is this number that is unfortunately a bit off. Indeed, the experimental value is $g_A = 1.26$, which is only 75% of our result.

This "failure" of the CQM (or at least its most naive version used here) has been attributed to several possible causes. For one, even a quark with a mass of 330 MeV is at least semi–relativistic if enclosed in the rather confined space of a proton (with its radius of less than 1 fm), according to the Heisenberg uncertainty principle. But in relativistic quantum mechanics, spin and orbital angular momentum are not separately conserved quantum numbers, so that our assumption $L = 0$ is no longer exactly true. This means that some of the total angular momentum of the proton is due to orbital motion of the quarks, and therefore less of its spin is carried by the quark spins (only about 75%, if we believe our result). A refinement of our model along these lines, the Relativistic CQM has been developed by several authors [10,11] and is generally better able to describe this and other properties of the nucleon.

However, there is a more fundamental reason why we shouldn't expect g_A to come out right in our simple model. As already mentioned before, constituent quarks are not really fundamental particles, but rather quasiparticles which themselves can have a complicated "internal" structure. As we now know (see Section 4), the real fundamental building blocks of all hadrons are the so–called "current" quarks, which carry the same quantum numbers but are much lighter (few MeV's) and truly pointlike (until further notice). Each constituent quark must contain at least one "valence" current quark with the same quantum numbers, but can also contain a number of quark–antiquark pairs of all kinds of different flavors (up, down, strange,...). In the following, we will denote these fundamental quarks with lower case letters u, d, s and so on, to distinguish them from the constituent quarks. The corresponding antiquarks are $\overline{u}$, $\overline{d}$, and $\overline{s}$. These fundamental quarks are bound inside the nucleon by Color forces transmitted by the exchange of gluons, which can also carry spin and orbital angular momentum. It is these fundamental quarks that the weak interaction really couples to (since its range is essentially zero and certainly much smaller than even the size of constituent quarks). It is therefore perhaps not surprising that the CQM is better at explaining large-range properties of the nucleon (like its magnetic moment) than short-range ones. We will see additional examples of this later on.

Nevertheless, our result that $g_A = \Delta u - \Delta d$ is still true, once we apply it to the current quarks inside the nucleon instead. We have to amend our definition of Δf, though, to include both quarks and antiquarks of the same flavor f. This works out since the antiquarks that take part in the weak interaction have the *opposite* isospin than the corresponding quarks, but also opposite helicity (the weak interaction couples to lefthanded particles and righthanded antiparticles), which yields overall the same sign.

One can expand our arguments to the case of strangeness-changing weak interactions, for instance the β–decay of a Λ into a proton, electron and electron–antineutrino. The basic process here is the change of an s quark into an u quark, and one can again define a vector and an axial vector part for this transition. By expanding our group theory arguments to the (approximate) $SU(3)$ symmetry group of Isospin and Strangeness, one can show that in this case the axial vector coupling g_A can be related to the combination $\Delta u + \Delta d - 2\Delta s$, again taken for the proton (the Λ and the proton belong to the same $SU(3)$ octet). This quantity is often denoted as $3F - D$ (for historical reasons, I assume). In our CQM, we would simply get back the total spin of the proton (if we ignore relativistic effects and orbital motion), since there are no constituent strange quarks in the proton. We therefore would predict $g_A = 1$ in this case, while the experimental result is $g_A = 0.72$, which is again a $\approx 75\%$ reduction. A global fit to *all* hyperon β–decays yields $3F - D = 0.58$ (even smaller), which could be interpreted as a breakdown of the underlying $SU(3)$ symmetry or as an indication that Δs is not zero after all (there are other possible explanations).

In spite of these shortcomings, the CQM has many successes explaining hadron structure (especially its refined versions like the relativistic CQM).

For instance, one can describe the transition from the nucleon to the Delta resonance as a spin flip transition, where a quark with its spin opposite to the overall proton spin changes its spin direction, leading to an $S = 3/2$ state. One can also introduce a phenomenological potential for the three constituent quarks, and predict the masses and other properties for excited states. Indeed, many of these excited states have been found at about the right masses, and can be explained as orbital angular momentum excitations (with $L > 0$) or even radial excitations of the 3–quark state. For example, the S_{11} resonance (at about 1535 MeV) has total spin $J = 1/2$, but opposite parity to the nucleon, which can be interpreted as an $L = 1$, $S = 1/2$ state where L and S couple to a total J of 1/2. Conversely, the D_{13} resonance (at nearly the same mass, 1520 MeV) has also negative parity, but $J = 3/2$, which means that here $L = 1$ and $S = 1/2$ couple additively. As a final example, the F_{15} resonance at 1680 MeV has $J^P = 5/2^+$, which indicates an $L = 2$ state. Many more resonances are predicted and observed (mostly in pion scattering), but the three listed above are the most prominent in electromagnetic transitions (including electron scattering).

3 Lepton Scattering and Structure Functions

Lepton scattering is a very powerful method for unraveling the internal structure of composite systems (like nuclei, nucleons, or even atoms). There are two reasons for this: the underlying electroweak interaction is both well understood, and it is sufficiently weak to allow perturbative treatment. Indeed, it is often sufficient to calculate cross sections in first order of perturbation theory, which corresponds to the exchange of a single virtual photon γ^* or a Z or W boson. [e]

Before we proceed to write down the cross section for inclusive lepton scattering in this (so–called "Born") approximation, we will first describe the measurement process from an experimental point of view. In all cases, one needs first an abundant source of incident leptons (electrons, positrons, muons or neutrinos) at a well-defined energy E and with a well-defined initial direction. In the case of an electron beam, one uses a source of free electrons (like a thermionic gun or a photo-emission source) which are then accelerated to the desired energy. This can be done either in a linear accelerator (SLAC [12], CEBAF [13]) or in a storage ring (HERA at the DESY accelerator center [14]). In the case of muon scattering, one uses a high energy proton beam impinging on

[e] In fact, higher order corrections (called "radiative corrections") **are** important for high-precision experiments; however, these are typically left for the experimentalists to calculate, which could be attributed either to a complete trust of the theorists in our ability to handle the equations of QED, or, less flattering, to their unwillingness to do the work themselves.

a production target (CERN [15], Fermilab [16]). Some of the secondary particles produced (mostly π^+) decay into muons which carry a large fraction of the initial beam energy. These muons can be selected according to their energy and transported to the target under study. Conversely, one can also use the muon (anti-)neutrinos produced in the decay as probes.

The incident lepton beam is directed at a target made of the nuclear species under investigation (hydrogen for protons, deuterium or ^{3}He for neutrons, or heavier nuclei in some cases). Fixed solid state or cryogenic liquid targets are most commonly used, but some of the more recent results have been obtained with a counterrotating proton beam of 800 GeV at the HERA [f] accelerator in Hamburg (Germany).

The outgoing scattered leptons are momentum-analyzed by magnetic fields and detected by a system of particle detectors, like drift chambers, scintillator counters (hodoscopes), Čerenkov or transition radiation (TRC) counters, and electromagnetic calorimeters (shower counters) or muon counters. (See Fig. 5 for an example.) The information from these counters is collected and stored by a data acquisition computer.

During offline analysis, the detector data for each event are analyzed to determine whether the detected particle is indeed the scattered lepton of interest, and to calculate its momentum and direction. The events are then accumulated in bins of the kinematic variables, e.g. the final energy E', polar scattering angle θ and azimuthal angle ϕ. The total number of events recorded in a given kinematic bin will be proportional to the following quantities:

1. The total number of incident leptons, N_i. This can be expressed as the current of incoming leptons per unit time, I_i, integrated over time.

2. The density of target atoms per unit area, $t = N_T/area = \rho[g/cm^3] \times L[cm] \times 6.022 \cdot 10^{23}/A$, where A is the atomic weight of the nuclear species in the target. The product of t with I_i is called the luminosity $\mathcal{L}$ and has units of $1/cm^2/s$.

3. The size of the kinematic bin, e.g. $\Delta E' \Delta\Omega = \Delta E' sin(\theta) \Delta\theta \Delta\phi$.

The ratio of the detected number of events in a given bin over the product $N_i t$ has dimensions of area and is called the partial cross section $\Delta\sigma$ for this bin. In the limit where the size of the kinematic bin becomes very small, the ratio $\Delta\sigma/\Delta E' \Delta\Omega$ approaches a finite limit value which is called the (double) *differential* cross section $d\sigma/dE' d\Omega$.

[f] HERA stands for Hadron Electron Ring Anlage or Accelerator. However, it was found that positrons are preferable to electrons as probes - for purely technical reasons - and therefore the "E" is somewhat unjustified at present.

Theoretically, the partial cross section for inclusive lepton scattering can be written to lowest order following Fermi's Golden Rule:

$$\Delta\sigma = \frac{1}{j_{in}}\frac{2\pi}{\hbar}|\mathcal{M}_{fi}|^2\Delta\Phi. \tag{6}$$

Here, j_{in} is the current density of the incoming lepton and $\Delta\Phi$ is the phase space spanned by the kinematic bin. In the following, we want to concentrate on elastic electromagnetic scattering first.

In this case the transition matrix element $\mathcal{M}_{fi} = \langle\psi_f|\mathcal{H}_{int}|\psi_i\rangle$ can be written as $\mathcal{M}_{fi} = ej_\mu^{lepton}(1/Q^2)ezj^\mu_{target}$ where z is the charge of the target in units of e and $Q^2 = -q_\mu q^\mu$ is the (negative) square of the four momentum $q = (\nu, \mathbf{q}) = (E - E', \mathbf{k}_i - \mathbf{k}_f)$ transferred to the target. (Q^2 can be calculated in the target rest system as $Q^2 = 4EE'sin^2(\theta/2)$.)

For the (hypothetical) case of spinless leptons scattering off a spin- and structureless target, the result is

$$\Delta\sigma = \frac{4z^2\alpha^2(\hbar c)^2}{Q^4}E'^2\Delta\Omega \tag{7}$$

(α is the fine structure constant). The magnetic interaction due to the lepton spin adds an additional factor $(1 - \beta^2 sin^2(\theta/2))$ which simplifies to $cos^2(\theta/2)$ for high energy leptons. Similarly, a spin-1/2 (point) target yields another factor $(1 + 2\nu^2/Q^2tan^2(\theta/2))$. Finally, the target recoil results in a factor E'/E, so that the differential cross section has the form

$$\frac{\Delta\sigma}{\Delta\Omega} = \frac{4z^2\alpha^2(\hbar c)^2E'^2}{Q^4}\frac{E'}{E}cos^2(\theta/2)(1 + 2\nu^2/Q^2tan^2(\theta/2)). \tag{8}$$

This cross section does not contain a term $\Delta E'$ since for elastic scattering, E' is uniquely determined by kinematics alone.

The form of the cross section, Eq. 8, is only valid for a structureless point target ("Dirac" particle), where the wave function (and therefore the electromagnetic current, j^μ_{target}) can be written down exactly. In the case of a finite size target with internal structure (like the proton), the cross section gets modified by the insertion of so-called form factors (G_E and G_M, the electric and magnetic form factors, in case of a spin-1/2 target). These form factors depend on the momentum transfer squared, Q^2, and parameterize the distribution of charges and currents inside the target. In fact, the charge form factor $G_E(Q^2)$ can be interpreted as the Fourier transform of the charge distribution inside the target, and $G_M(Q^2)$ as the Fourier transform of the distribution of electric current and magnetic moment throughout the target. For a pointlike Dirac target, $G_E(Q^2) = G_M(Q^2) \equiv z$. In the case of a finite size target, the electric form factor $G_E(Q^2)$ converges to z for $Q^2 \to 0$, since the wavelength of the exchanged photon becomes too large to resolve the inner structure of the target. However, $G_M(Q^2)$ does not necessarily have the same limit; if the target has an anomalous magnetic moment ($\mu \neq \mu_N$), $G_M(Q^2)$ converges towards μ/μ_N instead. The complete elastic cross section for targets with internal structure becomes

$$\frac{\Delta\sigma}{\Delta\Omega} = \frac{4\alpha^2(\hbar c)^2 E'^2 cos^2\frac{\theta}{2}}{Q^4}\frac{E'}{E}\left(\frac{G_E^2(Q^2) + \tau G_M^2(Q^2)}{1+\tau} + 2\tau\, tan^2\frac{\theta}{2}\, G_M^2(Q^2)\right), \tag{9}$$

where $\tau = \nu^2/Q^2$.

As an example, both the magnetic and electric form factor of the proton have been found to follow approximately a so-called dipole form: $G_E(Q^2) = G_M(Q^2)/(\mu/\mu_N) = (1 + Q^2/0.71\ \mathrm{GeV}^2)^{-2}$ which can be interpreted as the Fourier transform of an exponentially falling charge distribution.

In the following, we want to discuss how the cross section Eq. 9 changes in the case of inelastic scattering. For this purpose, it is useful to introduce some more kinematic variables. Of great importance is the invariant mass of the unobserved final state (the sum of all energies of the target fragments in their center-of-mass system), W. Since in the target rest system, the final state has four momentum $P'^\mu = (M + \nu, \mathbf{q})$, we can calculate $W^2 = P'^\mu P'_\mu = M^2 + 2M\nu + \nu^2 - \mathbf{q}^2 = M^2 + 2M\nu - Q^2$. In the case of elastic scattering, we must have $W^2 = M^2$ and therefore $2M\nu = Q^2$, or $x \equiv 1$ where $x \equiv x_{Bj} = Q^2/2M\nu$. If we transfer more energy to the target, we can excite higher mass resonant states, with $W = M_{res}$. At even higher energy transfer (deep inelastic region), we can create a continuous spectrum of multi-particle final states. In these cases, the cross section becomes a function of θ and E'. Figure 1 shows an example for the cross section for electron scattering off protons. The marks indicate the positions of some well-known nucleon resonances, some of which can be clearly seen as peaks in the cross section (these correspond to the Δ^+, S_{11} and D_{13}, and F_{15} resonances discussed in Section 1.)

Clearly, we have to replace the form factors in Eq. 9 with functions of both Q^2 and ν to describe the inelastic cross section. Specifically, the expression $\frac{G_E^2(Q^2)+\tau G_M^2(Q^2)}{1+\tau}$ is replaced by the structure function $W_2(Q^2, \nu)$ and $\tau G_M^2(Q^2)$ is replaced by a second structure function, $W_1(Q^2, \nu)$. From this substitution, it is clear that $W_1(Q^2, \nu)$ parameterizes the *transverse* part of the electromagnetic transition matrix element $|\mathcal{M}_{fi}|^2$, now taken between the target initial ground state i and an unbound (continuum) final state f. The structure function W_2, on the other hand, contains both transverse and longitudinal

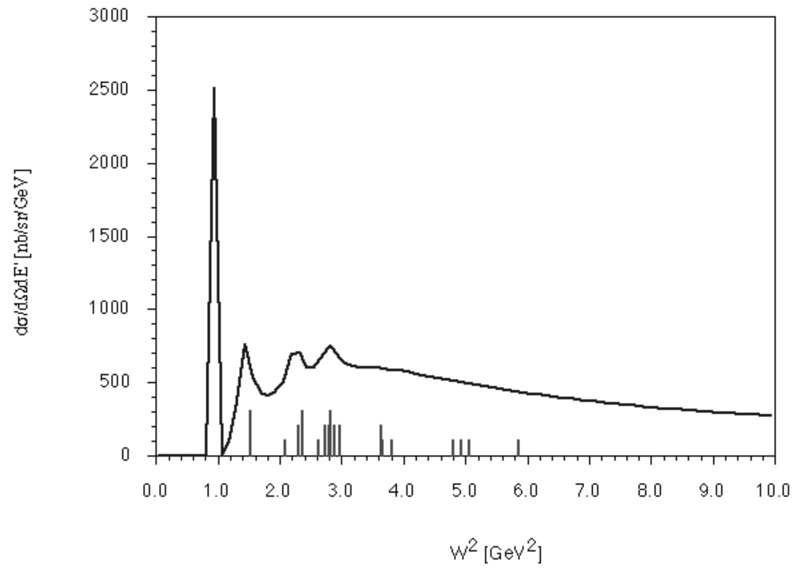

Figure 1: Cross section $d\sigma/d\Omega dE'$ in nb/sr/GeV for electron-proton scattering at 9.71 GeV electron energy and 7 degree scattering angle. The large peak below $W^2 = 1$ GeV^2 is the elastic peak, which is smeared out due to resolution and radiative effects included in this calculation.

("charge") transition matrix elements. Alternatively, one often introduces a longitudinal structure function $W_L(Q^2,\nu)$ in analogy with $G_E(Q^2)$, so that $W_2(Q^2,\nu) = \frac{W_L(Q^2,\nu)+W_1(Q^2,\nu)}{1+\tau}$. The ratio $R = W_L(Q^2,\nu)/W_1(Q^2,\nu)$ indicates the relative importance of longitudinal and transverse transition strength.

Using these newly defined structure functions, one can write the inelastic inclusive cross section as

$$\frac{\Delta\sigma}{\Delta Q^2 \Delta\nu} = \frac{4\pi\alpha^2(\hbar c)^2 E' cos^2(\theta/2)}{Q^4 E}(W_2(Q^2,\nu) + 2\,tan^2(\theta/2)\,W_1(Q^2,\nu)), \quad (10)$$

where we have replaced the kinematic bin $\Delta\Omega\Delta E'$ with the kinematic bin $\Delta Q^2 \Delta\nu = (EE'/\pi)\Delta\Omega\Delta E'$. Using our alternative set of structure functions, Eq. 10 can also be written (after some lengthy but elementary algebra) as

$$\frac{\Delta\sigma}{\Delta Q^2 \Delta\nu} = \frac{4\pi\alpha^2(\hbar c)^2 E' cos^2(\theta/2)}{Q^4 E}\frac{W_1(Q^2,\nu)}{\epsilon(1+\tau)}(1 + \epsilon R(Q^2,\nu)), \quad (11)$$

with $\epsilon = (1 + 2(1+\tau)tan^2(\theta/2))^{-1}$.

In the case of a transition to a resonant state, $W_1(Q^2, \nu)$ is peaked around the energy transfer needed to excite the resonance, $\nu_R = (W_{Res}^2 - M^2 + Q^2)/2m$, but it is not a delta function $\delta(\nu - \nu_R)$. This is due to the fact that all nucleon resonances decay with very short life time, leading to a finite width $\Gamma \propto 1/T$ of their mass peaks. The Q^2–dependence of the structure functions once again reflects the finite spatial extent of both the nucleon ground state and the excited resonant state. Therefore, for any given resonance, $W_1(Q^2, \nu)$ falls off rapidly with high Q^2, just like the nucleon form factors (often with the same "dipole"–like behavior). However, if the wave length $\lambda \propto 1/Q$ of the exchanged virtual photon becomes sufficiently small ($Q^2 \gg 1$ GeV^2) and the energy transfer is high enough ($W > 2 - 3$ GeV - the deep inelastic region), one can think of the scattering process as *elastic* scattering off the quarks inside the nucleon target, and the structure functions exhibit *scaling* behavior. We will now discuss this limit of deeply inelastic scattering (DIS) in detail.

Since (elementary) quarks are Dirac particles, we would expect that the cross section simplifies to Eq.8, and the structure functions would become delta functions $\delta(\nu - \nu_q(el))$ with $\nu_q(el) = Q^2/2m_q$ (this implies $x_q(el) = m_q/M_N$). However, since quarks are bound inside nucleons, they are not at rest, which leads to a "Doppler" smearing of the possible energy transfers ν, as well as x. To understand the meaning of the variable x in that case, it is useful to consider a different coordinate system, the so-called Breit frame.

We start by noting that since $Q^2 > 0$ in electron scattering, and since $\mathbf{q}^2 = Q^2 + \nu^2$, we have always $|\mathbf{q}| > \nu$, so we can find a coordinate system (using a Lorentz boost along the direction of the virtual photon) in which $\nu = 0$ (and obviously $|\mathbf{q}| = \sqrt{Q^2} = Q$), the Breit frame. The necessary boost parameters are $\Gamma = |\mathbf{q}|/Q$ and $\Gamma\beta = \nu/Q$. The nucleon's four momentum in this system is $P^\mu = (M\,|\mathbf{q}|/Q, 0, 0, -M\,\nu/Q)$. Now let's assume that the virtual photon strikes a quark with initial longitudinal momentum $-p_z^q(i)$ in the Breit frame. Since the photon carries no energy, it cannot change the overall energy or momentum of the quark, if we neglect binding effects (this is justified by the property of asymptotic freedom, which is one of the hallmarks of QCD). Therefore, we must require that the longitudinal momentum after scattering $p_z^q(f) = -p_z^q(i) + Q$ has the same magnitude, $p_z^q(i)$, as the initial one. This requires that $p_z^q(i) = Q/2$. In other words, the virtual photon can *only* scatter off a virtual quark that carries just the right longitudinal momentum, $Q/2$, in the Breit frame. By taking the ratio between this momentum and the longitudinal momentum of the whole nucleon, we get $-p_z^q(i)/(-M\nu/Q) = Q^2/(2M\nu) \equiv x$. We thus realize that the quantity x gets a new meaning in deep inelastic scattering: it is the fraction of the nucleon's longitudinal momentum carried by the struck quark, calculated in the Breit frame. If we call the probability to find a quark with momentum fraction x $q(x)$, we can incorporate this probability function in Eq. 8 to get the partial cross section for deep inelastic scattering:

$$\Delta\sigma = \frac{4\pi z_q^2 \alpha^2 (\hbar c)^2 E' cos^2(\theta/2)}{Q^4 E} (q(x)\Delta x + 2\nu^2/Q^2 tan^2(\theta/2) q(x)\Delta x)\Delta Q^2. \tag{12}$$

We can use the relation $\Delta x = -Q^2/(2M\nu^2)\Delta\nu = -x\Delta\nu/\nu$ to rewrite this as

$$\frac{\Delta\sigma}{\Delta Q^2 \Delta\nu} = \frac{4\pi\alpha^2(\hbar c)^2 E' cos^2(\theta/2)}{Q^4 E} (\frac{x}{\nu} z_q^2 q(x) + \frac{1}{M} tan^2(\theta/2) z_q^2 q(x)). \tag{13}$$

Finally, we have to include contributions from *all* different quark flavors f, each with its own probability distribution $f(x)$ and charge z_f. If we define new structure functions $F_1(x) = \frac{1}{2}\sum_f z_f^2 f(x)$ and $F_2(x) = x\sum_f z_f^2 f(x)$, we can write down the final form for the deep inelastic cross section as

$$\frac{\Delta\sigma}{\Delta Q^2 \Delta\nu} = \frac{4\pi\alpha^2(\hbar c)^2 E' cos^2(\theta/2)}{Q^4 E} (\frac{1}{\nu} F_2(x) + 2\, tan^2(\theta/2) \frac{1}{M} F_1(x)). \tag{14}$$

Comparison with Eq. 10 immediately shows that $F_1(x) = MW_1(Q^2,\nu)$ and $F_2(x) = \nu W_2(Q^2,\nu)$. This means that in this kinematic region, the structure functions become functions of one variable alone (x), while the dependence on Q^2 vanishes — they "scale". Furthermore, we expect the relationship $F_2(x) = 2xF_1(x)$ to hold, which follows directly from our expressions for F_1 and F_2 above in terms of the quark distribution functions $f(x)$.

4 Unpolarized Structure Functions F_1 and F_2

From our result in the previous section, it is clear that one can learn a lot about the internal (quark-) structure of the nucleon by studying the structure functions F_1 and F_2. The value of $F_1(x)$ at a given x can be directly interpreted as (one–half of) the likelihood of finding a quark with longitudinal momentum fraction x, summed over all quark flavors weighted with the corresponding quark charges squared. While this interpretation is, strictly speaking, frame dependent (the way we introduced it, it refers to the Breit frame), one can see that in the limit $Q^2 \to \infty$ but x fixed the Breit frame coincides with the infinite momentum frame (IMF) in which the third component of the nucleon momentum, $P^3_{\mathrm{Breit}} = -M\nu/Q = -Q/2x$, approaches infinity, $P^3 \to -\infty$. In this (scaling) limit, x measures the momentum fraction of the quarks in the IMF, which is independent of the other kinematic variables of the reaction.

One of the important results from early deep inelastic scattering (DIS) experiments was the (approximate) confirmation of the Callan-Gross relationship [17] $F_2(x) = 2xF_1(x)$. [g] Our derivation of this relationship at the end of the previous Section depends crucially on the assumption that the elastic cross section on a single quark can be described by Eq. 8, i.e. the cross section for a pointlike spin-1/2 (Dirac) particle. If quarks had no spin, we would have $F_1(x) = 0$ instead. The confirmation of scaling and the Callan-Gross relationship therefore show that nucleons are indeed made of (nearly) massless elementary spin-1/2 particles which become asymptotically free at large momentum transfers.

More information can be obtained by writing down the quark decomposition of the structure functions explicitly:

$$F_1(x) = \frac{1}{2}\left(\frac{4}{9}\left[u(x) + \bar{u}(x)\right] + \frac{1}{9}\left[d(x) + \bar{d}(x) + s(x) + \bar{s}(x)\right] + ...\right), \qquad (15)$$

if we neglect the heavier c, b and t quarks. Measuring $F_1(x)$ over a wide range of x should give us information on the quark distribution functions $q(x)$. In practice, one often measures $F_2(x)$ instead, since this quantity has no (typically small) factor $tan^2(\theta/2)$ in front of it. Either way, measuring one of these structure functions alone will not be enough to unravel the contributions of all different quark flavors (6 unknown functions of x in Eq. 15).

One possibility is to assume that the "sea" quarks $s, \bar{u}, \bar{d}, \bar{s}$ do not contribute significantly, and that $u(x) \approx 2d(x)$ in the proton. These assumptions are in line with our naive CQM, and cannot be expected to be a realistic description of the current quark distributions. However, they hold approximately "on average", so that we can at least estimate the integral

$$\int_0^1 F_2(x)dx = \int_0^1 x\left(\frac{4}{9}u(x) + \frac{1}{9}d(x)\right)dx = \int_0^1 xd(x)dx. \qquad (16)$$

Using again our assumption, $u(x) \approx 2d(x)$, one sees that this integral should equal 1/3 of the overall quark momentum distribution, weighed by the momentum fraction x. In other words, from this integral we can determine the longitudinal momentum fraction carried by all quarks in the proton together, as $3\int_0^1 F_2(x)dx = x_{total}$. From our CQM, we would of course expect the

[g] One can measure both F_1 and F_2 independently by using a method called "Rosenbluth separation". Basically, one varies the scattering angle θ while simultaneously changing the beam energy to keep x and Q^2 constant. Since F_1 has an extra factor $tan^2(\theta/2)$ in front of it, it's contribution will be different for these different kinematics and by a linear fit both F_1 and F_2 can be extracted.

number to be 1 (since there is nothing else but quarks inside a proton in this model). However, the experimental result is about 0.5, which means that only 1/2 of the proton momentum is carried by quarks. The other half is carried by the exchange bosons of the strong force, the gluons, which are responsible for binding the quarks inside the proton. Since these gluons carry no electric charge, they are invisible to lepton scattering experiments (except for indirect effects like the "missing momentum" described above).

A second integral we could calculate is $3\int_0^1 [F_2(x)/x]dx$, which should simply yield the total number of quarks in the proton. From the CQM, we would expect the number to be 3 (it certainly can not be less than that in *any* quark-based model). However, experimentally it turns out that the integral does not even seem to converge (it grows without bounds towards smaller and smaller x). The reason for this is that there are an infinite number of virtual quark-antiquark pairs constantly being created and annihilated in the proton, most of which carry only a tiny fraction x of the proton momentum. We will discuss this picture in more detail later in this section.

If we want to go beyond the simplifying assumptions of the CQM, we need additional information to unravel the individual contributions of the different quark flavors to F_1 and F_2. One possible angle comes from the insight that the quark distributions of protons and neutrons must be related via Isospin symmetry. If we take a proton with quark distribution functions $u_p(x), \bar{u}_p(x), d_p(x), \bar{d}_p(x), s_p(x)$ and $\bar{s}_p(x)$ and rotate it by 180 degrees in Isospin space, it becomes a neutron, while the quark flavors become transformed like $u \to d, d \to u, s \to s$, and equivalently for the antiquarks. Since Isospin is a very good symmetry of the strong interaction, we expect that the corresponding quark distributions are the same: $u_p(x) = d_n(x), \bar{u}_p(x) = \bar{d}_n(x), d_p(x) = u_n(x), \bar{d}_p(x) = \bar{u}_n(x)$, while the s and $\bar{s}$ distributions are the same in both nucleons (s quarks are iso-singlets). This relationship has become so widely ingrained that most people don't bother to make the distinction, and simply refer to, e.g., $u(x) \equiv u_p(x) = d_n(x)$ (we will follow that convention from now on).

This means that Eq. 15 describes the structure function $F_1^p(x)$ for the proton, while the neutron one becomes

$$F_1^n(x) = \frac{1}{2}\left(\frac{4}{9}\left[d(x) + \bar{d}(x)\right] + \frac{1}{9}\left[u(x) + \bar{u}(x) + s(x) + \bar{s}(x)\right] + ...\right), \quad (17)$$

all expressed in the "universal" isospin-symmetric distribution functions. Combining Eq. 15 and Eq. 17 allows us to gain additional information on the quark distributions. Of particular interest is the (isovector) difference

$$F_1^p(x) - F_1^n(x) = \frac{1}{2}\left(\frac{3}{9}(u(x)+\bar{u}(x)) - \frac{3}{9}(d(x)+\bar{d}(x))\right), \tag{18}$$

which has no contribution from strange quarks. We can further separate the distributions $u(x)$ and $d(x)$ into a valence part $u_V(x)$ and $d_V(x)$ and a "sea" part $u_S(x) = u(x) - u_V(x)$ and $d_S(x) = d(x) - d_V(x)$, where the valence quarks are the carriers of the nucleon quantum numbers (therefore, their total must be exactly $\int u_V(x)dx = 2$ and $\int d_V(x)dx = 1$). This means that the integral over x of the difference (Eq. 18) should be equal to

$$\begin{aligned} 2\int_0^1 \left[F_1^p(x) - F_1^n(x)\right] dx &= \frac{1}{3}\left(2 - 1 + (N_{u_S} + N_{\bar{u}}) - (N_{d_S} + N_{\bar{d}})\right) \\ &= \frac{1}{3} + \frac{2}{3}(N_{\bar{u}} - N_{\bar{d}}). \end{aligned} \tag{19}$$

In the second line, we made use of the fact that sea quarks can only be created in quark-antiquark pairs, and therefore the sum of sea quarks and antiquarks of a given flavor simply equals twice the number of antiquarks. If we assume that the quark distributions are the same for $\bar{u}$ and $\bar{d}$, we should expect the result to be simply equal to 1/3 (this is known as the Gottfried sum rule [18]).

Experimentally, it is not quite straightforward to evaluate the integral Eq. 19, since there are no free neutron targets. The next best thing are targets made of deuterium, which can be interpreted as an equal number of protons and neutrons. Figure 2 shows a parameterization [19] of the world's data on both the proton and deuterium which can be used for that purpose. However, one has to worry about binding effects and Fermi-smearing (due to the motion of the nucleons in deuterium) as well as possible modifications of the free nucleon structure functions for nucleons bound in nuclei (like deuterium). These deviations from the free nucleon structure functions are collectively known as the "EMC–effect", named after the collaboration at CERN that first discovered them experimentally [20]. Nevertheless, the best present estimate of the integral Eq. 19 from data taken on protons and deuterons yields 0.235 ± 0.026 [21], which is considerably different from the prediction 1/3. At face value, this result can be interpreted to mean that on average there are 0.15 more $\bar{d}$ quarks than $\bar{u}$ quarks. This has been related to the fact that the proton can dissociate into a neutron and a π^+ meson (with the CQM composition $|U\bar{D}\rangle$), while π^- mesons are less easy to produce.

Another quantity of interest is the ratio $F_1^n(x)/F_1^p(x)$ in the limit of large x, where one expects valence quarks to dominate. In that case, we can write

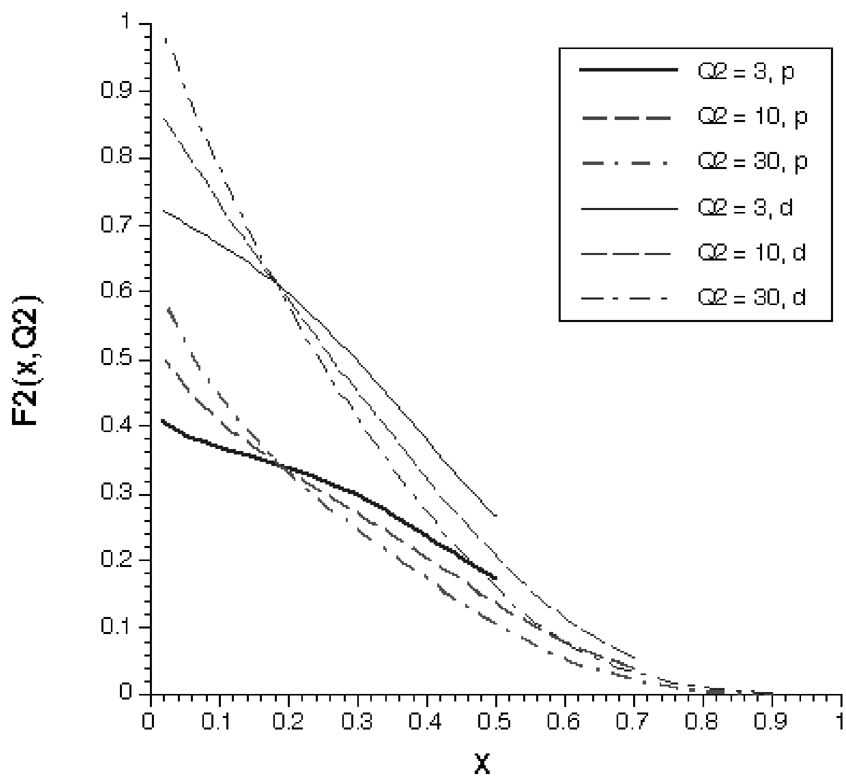

Figure 2: Parameterization of the world's data on the structure function $F_2^p(x, Q^2)$ of the proton and $F_2^d(x, Q^2)$ of the deuteron. The neutron structure functions $F_2^n(x, Q^2)$ is approximately equal to the difference between these two sets of curves. (Note that in most publications, $F_2^d(x, Q^2)$ is defined *per nucleon*, meaning the values given in the literature will be half those shown in this figure). Notice that the ratio $F_2^n(x, Q^2)/F_2^p(x, Q^2)$ is clearly decreasing towards higher x. The evolution of $F_2(x, Q^2)$ can also be clearly seen; strength is "moved" from high x towards small x. The curve for each Q^2 is shown up to the boundary of the scaling region $W^2 > 4.0$ GeV^2.

this ratio as

$$F_1^n(x)/F_1^p(x) = \frac{\left(\frac{4}{9}d_V(x) + \frac{1}{9}u_V(x)\right)}{\left(\frac{1}{9}d_V(x) + \frac{4}{9}u_V(x)\right)}. \tag{20}$$

If one assumes that the shape of the u_V and d_V distributions are the same at high x (and therefore $u_V(x) = 2d_V(x)$), the result should equal 2/3. On the other hand, the data [21] seem to favor a ratio closer to 1/4, which would imply that $u_V(x) \gg d_V(x)$ at large x. This means that u quarks are much more likely to carry nearly all of the proton momentum than d quarks; however, the difficulties stemming from the EMC effect discussed above are especially pronounced in the high-x region, so that this result is not uncontroversial [22]. There are some intriguing theoretical arguments why $u_V(x) \gg d_V(x)$ should be true at large x, though. It is thought that the strong interaction between two quarks favors the state where the quark spins point in opposite direction, leading to a total $S = 0$ state. This means that one needs more energy for a $S = 1$ qq state (hence the higher mass for the Δ versus the nucleon). From this it can be concluded that the two valence u quarks in the proton which do couple to $S = 1$ tend to carry a larger fraction of the total proton mass than the valence d quark, which translates to a higher likelihood for a large momentum fraction x in the Breit frame.

If we want to find additional information on the contribution of the different quark flavors to the proton structure functions, we have to turn to neutrino scattering. There are important differences in the cross sections for the reactions (i) $N(\nu_\mu, \mu^-)$ and (ii) $N(\bar{\nu}_\mu, \mu^+)$ relative to charged lepton (electron or muon) scattering, all due to the fact that these reactions proceed through the Weak Interaction (charged W boson exchange) instead of the electromagnetic one:

1. Reaction (i) involves a W^+ boson which must be absorbed by the struck quark. Therefore, it can only occur on quarks which have negative charge initially, i.e., d, s and $\bar{u}$. Conversely, reaction (ii) involves a W^- boson and can only occur on positively charged quarks, namely $u, \bar{d}$ and $\bar{s}$. This allows us to separate the contribution from these two different groups of quarks by comparing ν_μ and $\bar{\nu}_\mu$ – induced reactions.

2. The coupling strength to each type of quark is the same for all flavors, i.e. the structure functions contain no weighing factors z_q^2. In the case of u and d type quarks, this follows simply from the structure of the weak interaction that we have already discussed in Section 2. For s quarks, the statement above is true only if we have enough energy to create not only u quarks but also c quarks in the final state. Otherwise, the

contribution from s quarks would be suppressed by the Cabbibo factor $sin(\theta_{Cabibbo}) \approx 0.22$. [h]

3. Due to the $V-A$ structure of the Weak Interaction, all leptons in reaction (i) must be left-handed, with spins pointing opposite to their momentum. If the reaction occurs on a d or s quark, these have to be left-handed as well. In the center-of-mass system, it follows that the total spin in the direction of the relative momentum is zero. This means that there are no restrictions on the scattering angle from angular momentum conservation. On the other hand, if the struck quark is a $\bar{u}$, it must be *right*-handed, leading to an overall spin of 1 in its initial direction. The final state contains again a left-handed lepton and a right-handed antiquark, with overall spin 1. Conservation of angular momentum requires that this spin points in the same direction as the initial one. This leads to a suppression of events where the lepton scatters by a large angle in the center of mass, equivalent to a large energy loss ν in the lab. By measuring the angular (or ν) dependence of the structure functions, one can separate the parts coming from quarks and antiquarks. This works in full analogy for reaction (ii) as well.

4. As one would expect, the cross section is vastly smaller than for electromagnetic scattering, since it is proportional to $G_F^2 M_W^4/(M_W^2+Q^2)^2$ instead of α^2/Q^4 (G_F is the Fermi coupling constant for weak charged current interactions and $M_W \approx 80$ GeV is the mass of the W boson).

Combining these features, one can write the cross section (in the scaling limit, and assuming the Callan-Gross relationship) for reaction (i) as

$$\frac{\Delta\sigma}{\Delta Q^2 \Delta\nu} = \frac{G_F^2 M_W^4}{\pi(M_W^2+Q^2)^2}\frac{x}{\nu}(d(x)+s(x)+(1-\frac{\nu}{E})^2\bar{u}(x)), \qquad (21)$$

where we have assumed that θ is small and all momenta involved are large relative to the nucleon mass (more accurate formulae can be found in the Particle Data Group collection [6]). Similarly, the cross section for reaction (ii) is

$$\frac{\Delta\sigma}{\Delta Q^2 \Delta\nu} = \frac{G_F^2 M_W^4}{\pi(M_W^2+Q^2)^2}\frac{x}{\nu}(\bar{d}(x)+\bar{s}(x)+(1-\frac{\nu}{E})^2 u(x)). \qquad (22)$$

Equations 21 and 22 allow us to extract 4 of our 6 unknowns. One more unknown can be gotten from electromagnetic scattering on the proton and the

[h] Strictly speaking, the contribution from d quarks would be likewise suppressed by $cos(\theta_{Cabibbo})$, but this is numerically close to 1.

neutron, but the resulting six equations are not linearly independent, which still leaves us short by one. However, this linear relationship between the structure functions measured in electromagnetic and weak processes can be tested experimentally and the agreement found is a strong confirmation of the quark-parton picture of deep inelastic scattering and the electro-weak interaction.

For an unambiguous solution, one could again do experiments on protons as well as neutrons, determining 2 more equations analog to Eqs. 21 and 22, with u and d interchanged. This would indeed overdetermine the solution. However, the experimental problems of neutrino-induced reactions are too formidable to achieve high statistical precision that way. This is partially due to the small size of the cross section, which requires immensely massive targets for finite count rates. In addition, the initial lepton (neutrino) is not observed, so its energy is not known and must be inferred from the observed reaction products (including the hadronic debris of the struck quark). This requires that the massive target must be either itself a detector, or interspersed with a large number of sampling detectors.

The smallest statistical errors were achieved by the CCFR experiment at Fermilab [23] which used a 18 m long stack of iron plates alternating with scintillator detectors (corresponding to a target mass of hundreds of tons). The CDHS experiment [24] at CERN used a similar setup. Some lower statistics data were also taken with a hydrogen and deuterium bubble chamber at CERN (BEBC [25]). In practice, one attempts an overall fit to these as well as the electron and muon scattering data to find the best constraints on the individual quark distribution functions.

In the remainder of this Section, we will outline some of the deviations from our simple (scaling) picture of deep inelastic scattering. It was soon observed that the structure functions F_1 and F_2 do not scale exactly, i.e., they depend not only on x but also on the momentum transferred. At high Q^2, there are 2 principal sources of these so-called scaling violations: perturbative QCD corrections to the cross section for point-like quarks, and the evolution of the quark distribution functions with Q^2. Both effects lead to logarithmic corrections depending on $ln(Q^2)$. They are described in detail for example in Ref. [1]. Here, we can only give an intuitive outline.

The perturbative QCD (pQCD) corrections have their origin in the fact that quarks are strongly interacting particles, which can radiate extra gluons in the reaction process. These pQCD corrections can be understood as the QCD analogue of the (QED) radiative corrections briefly discussed in Section 3, except that they are governed by the much larger strength of the strong coupling constant, $\alpha_s/2\pi$. One of the peculiarities of the theory of strong interactions, QCD, is that this coupling constant depends quite strongly on the mass scale (Q^2 in our case). One speaks of the "running" of the coupling constant $\alpha_s(Q^2)$. [i] For instance, the value for α_s at moderate scales ($Q^2 \approx 3$ GeV^2) is nearly 3 times larger than at the scale of the Z mass M_Z^2. This fall-off of α_s at high energies leads to the phenomenon of "asymptotic freedom", which means that at very high Q^2 the quarks behave as if they were free objects. The flip side of asymptotic freedom is the large value of the coupling at low energy scales which is believed to lead to the confinement of quarks inside hadrons (this is the reason why free quarks have never been observed). At sufficiently low energy scales, the coupling reaches order unity and pQCD can no longer be applied.

The evolution of the quark distribution functions, on the other hand, depends on the resolving power of our scattering probe. In the Breit frame, the wave length of the exchanged boson (photon or W/Z vector boson) is directly given by $2\pi/Q$. At very high Q^2, finer details of the proton structure can be resolved than at lower values of Q^2. This finer resolution can reveal that what looked like an ordinary quark carrying momentum fraction x is actually composed of a quark, maybe additional gluons, and sea quark-antiquark pairs, each carrying a smaller fraction $x' < x$ than the original quark. This leads to an increase of the quark distribution functions at low x and a corresponding decrease at higher x as Q^2 increases. This is indeed born out by the data, as shown in Fig. 2. This figure shows a parameterization [19] of the medium-high Q^2 data on proton and deuteron structure functions. Formally, the evolution of the distribution functions is described by the DGLAP evolution equations [26]. It should be noted that to first order (and *only* to first order) perturbation theory, both pQCD effects and the DGLAP evolution can be incorporated into scale-dependent quark distribution functions $q(x, Q^2)$.

One consequence of the Q^2–dependent evolution of the distribution functions is that one gets a handle on the gluon distribution function $g(x, Q^2)$ which cannot be observed directly in lepton-nucleon scattering, since gluons do not interact electromagnetically or weakly. We have already observed earlier in this section that one can infer that gluons must account for roughly 1/2 of the nucleon momentum. By studying the Q^2 dependence of the structure functions one can extract the gluon distribution functions in more detail. This works because a gluon can split up into a quark-antiquark pair perturbatively – this process is part of the the pQCD corrections of the scattering process. The lepton probe can interact with this quark or antiquark, leading to information on the original gluon. This process is termed "photon-gluon fusion" (in the case of electromagnetic scattering). The overall program of DIS then requires

[i] Strictly speaking, the electromagnetic coupling constant also depends on Q^2; however, this dependence is rather weak in the region below $Q^2 \approx 10,000$ GeV$^2 \approx M_Z^2$.

a global fit to *all* measured structure functions, over a wide range of x and Q^2, with a model distribution function for both quarks ($q(x, Q_0^2)$) and gluons ($g(x, Q_0^2)$) at some fixed scale Q_0^2 as input. Tabulations and graphs of the measured structure functions can be found in Ref. [6] and Ref. [1]. Fits to these measurements have been done (and constantly refined) by several groups; see for instance the so-called MRS [27] and CTEQ [28] parameterizations. The overall picture of the distribution of quarks and gluons in the nucleon is remarkably detailed and consistent. However, a theoretical prediction of these distribution functions based on first (QCD) principles seems to be still out of reach. The reason for this difficulty lies in the non-perturbative nature of QCD at the mass scale of the nucleon.

Further deviations from the scaling behavior of the structure functions and the Callan-Gross relation occur at even lower Q^2. These are often subsumed (not quite correctly) under the heading of "higher twist effects". [j] Basically, these effects are due to the fact that quarks *are* after all bound inside the nucleon, have a finite mass and can have a momentum component perpendicular to the virtual photon. In addition, the Breit frame no longer coincides with the infinite momentum frame at lower Q^2. All of these deviations from the scaling behavior are typically proportional to powers of $1/Q$, with a leading term proportional to $1/Q^2$ in unpolarized scattering.

The part of this $1/Q^2$ term coming from binding effects is often referred to as "dynamical higher twist" (or "twist–4"; the terms that survive as $Q^2 \rightarrow \infty$ are "twist–2" or leading twist) and can in principle be calculated using a technique called "operator product expansion" (OPE). This technique involves determining matrix elements of higher twist operators that include correlations between quarks and gluons inside the nucleon.

Of particular interest is the behavior of the ratio $R = W_L/W_1$ between the longitudinal and transverse structure functions we defined in the previous section. The Callan-Gross relationship implies that $R \approx 0$ at high Q^2. However, at lower Q^2 it can become quite sizable due to higher–twist effects. R can be determined experimentally using the technique of Rosenbluth separation discussed at the beginning of this Section. Recent precision data have been obtained at SLAC [29] and by NMC [21]. Figure 3 shows a parameterization of R from a re-analysis of SLAC data [19].

In the region of low $Q^2 < 3$ GeV^2 and moderate x (corresponding to $W < 2$ GeV), the binding of the quarks becomes especially important in the final state. In fact, the structure functions are mostly determined by these

[j] A discussion of the twist expansion of QCD is outside the scope of this paper. Twist is defined as the dimension minus the spin of an operator contributing to the matrix element $\mathcal{M}_{fi}$.

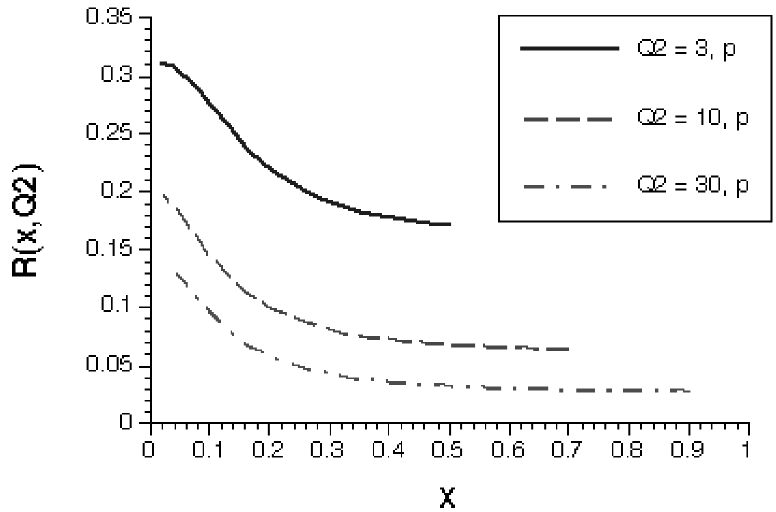

Figure 3: Parameterization of the world's data on the structure function $R^p(x, Q^2)$ of the proton. One can clearly see that R becomes small at high Q^2, in agreement with the Callan-Gross relationship, but large at smaller Q^2 due to higher twist effects. The curve for each Q^2 is shown up to the boundary of the scaling region $W^2 > 4.0$ GeV^2.

final state interactions, which prevent the struck quark from being knocked out of the nucleon. Instead, the transferred energy is shared with the other constituents, leading to a resonant excitation of the nucleon of the whole – these are again the nucleon resonances we discussed before. The energy scale at which the resonant final states become important coincides (but surely not by coincidence) with a resolution $1/Q$ appropriate for the size of the constituent quarks in our naive CQM. At face value, this means we cannot directly measure the quark distribution functions of constituent quarks without taking final state interactions into account. On the other hand, it turns out that if one *averages* over several resonant final states, the measured structure functions still agree with the ones measured in DIS, extrapolated appropriately into the resonance region. Of course, for this extrapolation one has to account for the $1/Q^2$ dependent effects listed above. This can be done by modifying the scaling variable x appropriately (e.g., using the so-called "Nachtmann variable" ξ [30]). This agreement between averaged resonance structure functions and extrapolated DIS structure functions is known as "duality".

5 Introduction to Polarized Structure Functions

Up to this point, we have discussed the information on the quark structure of nucleons that can be gleaned from measurements of the two unpolarized structure functions F_1 and F_2. It can be shown that for unpolarized (spin-1/2) targets and beam, there are only these two independent quantities one can extract (this is based on general properties of the Hamiltonian, like Lorentz and Gauge invariance). If one allows for parity violating terms (as surely one must in the case of neutrino interactions), a third structure function F_3 arises (this corresponds to the extra ν-dependent term in Eqs. 21 and 22).

In the case where both the leptonic probe and the target are polarized, two more structure functions, g_1 and g_2, can be measured in inclusive lepton scattering. In the following, we will turn our attention to these structure functions. We will restrict ourselves to the case of charged lepton scattering (electromagnetic interaction) only.

It is often instructive to think of electromagnetic scattering as a two-step process: The electron (or muon) gives off a virtual photon γ^* which is then absorbed by the target. One can indeed factorize the cross section Eqs. 10 – 11 into two parts: an absorption cross section for the virtual photon $\sigma(\gamma^*)$ and the likelihood for the emission of a virtual photon with kinematics given by ν and Q^2, the so-called virtual photon flux Γ. However, to do that, one has to consider the different possible polarization states of the virtual photon.

Virtual photons can have all 3 possible spin orientations for spin-1 particles, namely $m = -1, 0, 1$, while real photons are massless and can only have $m = \pm 1$. Virtual photons with $m = 0$ are called "longitudinal", since their electric field points along the direction of propagation. This is also the situation one has in classic electrostatic (Coulomb) fields, and therefore it makes sense that longitudinal photons couple to the electric charge of the target. On the other hand, photons with $m = \pm 1$ are called transverse, in analogy to real photons ("ordinary light") which have their electric and magnetic fields pointing perpendicular to the direction of propagation.

In the case of inclusive unpolarized lepton scattering, the cross section is the sum of a transverse photon part and a longitudinal photon part:

$$\frac{d\sigma}{d\Omega dE'} = \Gamma_T \sigma_T(\gamma^*) + \Gamma_L \sigma_L(\gamma^*) \tag{23}$$

The relative strength ("flux") of these two photon polarization states is given by the kinematics of the scattered lepton alone and corresponds to the parameter ϵ we introduced in Section 3 (see Eq. 11): $\Gamma_L = \epsilon\Gamma_T$. The transverse and longitudinal photo-absorption cross sections $\sigma_T(\gamma^*)$ and $\sigma_L(\gamma^*)$ are related to the transverse and longitudinal structure functions W_1 and W_L, also defined in Section 3. The exact relationship is somewhat a matter of convention; we use the so-called "Hand convention" with $\sigma_T(\gamma^*) = K^{-1}4\pi^2\alpha W_1$ and $\sigma_L(\gamma^*) = K^{-1}4\pi^2\alpha W_L$, where $K = (W^2 - M^2)/2M = \nu - Q^2/2M$ is the equivalent energy for a real photon to reach the same final state invariant mass W. Note that the quantity R introduced in Section 3 gives the ratio of longitudinal to transverse photoabsorption cross section: $R = W_L/W_1 = \sigma_L/\sigma_T$. Therefore, we can write Eq. 23 also as

$$\frac{d\sigma}{d\Omega dE'} = \Gamma_T \sigma_T(\gamma^*)(1 + \epsilon R). \tag{24}$$

The virtual photon polarization states $m = 1$ and $m = -1$ contribute equally to the transverse part of the cross section Eq. 23, and their relative strength cannot be controlled in unpolarized scattering. But if the incident lepton is polarized along its direction of motion (i.e., it has helicity $h_e = \hat{s}_e \cdot \hat{k}_e = +1$ or $h_e = -1$), some of this helicity is transferred to the virtual photon: $h_{\gamma^*} = \sqrt{(1-\epsilon^2)}h_e$. Thus, one can think of polarized lepton scattering simply as a way to select the polarization state (in particular the helicity) of the exchanged virtual photon.

If the target is itself polarized along the direction of the virtual photon ($s_z = \pm 1/2$), there will be a difference between the absorption strength for virtual photons of opposite helicity, which is determined by angular momentum conservation. Depending on the relative orientation of target and photon helicity, the final state (after the photon has been absorbed) can either have total spin along the photon direction of $S_z = \pm 1/2$ (if target and photon spin point in opposite directions) or $S_z = \pm 3/2$ (if both point in the same direction). The corresponding virtual photon absorption cross sections are $\sigma_T^{1/2}(\gamma^*)$ and $\sigma_T^{3/2}(\gamma^*)$, with $\sigma_T^{1/2}(\gamma^*) + \sigma_T^{3/2}(\gamma^*) = 2\,\sigma_T(\gamma^*)$.

One can define a (virtual [k]) photon asymmetry A_1 in terms of these cross sections:

$$A_1(\gamma^*) = \frac{\sigma_T^{1/2}(\gamma^*) - \sigma_T^{3/2}(\gamma^*)}{\sigma_T^{1/2}(\gamma^*) + \sigma_T^{3/2}(\gamma^*)}. \tag{25}$$

If the final state is a spin-1/2 object (as is the case in elastic electron-nucleon or electron-quark scattering), obviously $|S_z| = 3/2$ is excluded and therefore $\sigma_T^{3/2}(\gamma^*) = 0$. Clearly, in this case $A_1 \equiv 1$ by definition. On the other hand, consider the excitation of a Δ in (virtual) photon absorption. Since the Δ has spin 3/2, both $|S_z| = 3/2$ and $|S_z| = 1/2$ final states are possible.

[k] All the quantities defined in this paragraph can be equally applied to *real* photons, as well, since those are transverse by definition.

The Clebsch-Gordon coefficient for a $|s = 1/2,\ s_z = 1/2>$ state and a $|s(\gamma^*) = 1,\ s_z(\gamma^*) = 1>$ state coupling to a total of $|S = 3/2,\ S_z = 3/2>$ is, of course, equal to 1. The corresponding Clebsch-Gordon coefficient for coupling a $|s = 1/2,\ s_z = -1/2>$ state and $|s(\gamma^*) = 1,\ s_z(\gamma^*) = 1>$ to $|S = 3/2\ , S_z = 1/2>$ is only $\sqrt{1/3}$, however. It follows that the cross section $\sigma_T^{1/2}(\gamma^*)$ is only 1/3 the size of $\sigma_T^{3/2}(\gamma^*)$, which yields $A_1 = -1/2$ in this case. (We implicitly assumed that the nucleon-Δ transition is purely spin-flip and does not involve orbital angular momentum). In general, A_1 can assume any value between -1 and 1 (this is called the positivity limit).

The asymmetry A_1 can be directly measured for real photon absorption (I'm not saying that this is an easy experiment, though). In lepton scattering, however, we have to account for the virtual photon helicity and the fact that the overall unpolarized cross section has a longitudinal part, as well. This yields for the cross section asymmetry for lepton scattering with helicity $+1$ versus helicity -1 on a target with spin 1/2 aligned opposite to the virtual photon direction

$$A(el) = \frac{d\sigma^+ - d\sigma^-}{d\sigma^+ + d\sigma^-} = \sqrt{(1-\epsilon^2)}\,\frac{A_1(\gamma^*)}{1+\epsilon R}. \qquad (26)$$

Things become a bit more complicated if the target spin is not aligned with the virtual photon direction, but at an angle θ^* relative to that direction. For one, the asymmetry $A(el)$ (Eq. 26) must be multiplied by another factor $cos(\theta^*)$. In addition, the transverse and the longitudinal part of the cross section can interfere in this case, and we must introduce another virtual photon cross section, the interference cross section σ_{LT}. We can define a second virtual photon asymmetry $A_2(\gamma^*) = \sigma_{LT}/\sigma_T$. Since σ_{LT} is bound by $|\sigma_{LT}| \leq \sqrt{\sigma_L\ \sigma_T}$, we have the constraint $|A_2| \leq \sqrt{R}$. In elastic scattering, $\sigma_{LT} \propto \sqrt{\tau}\ G_M\ G_E$ and $\sigma_T \propto \tau\ G_M^2$ (where $\tau = \nu^2/Q^2$ as defined in Section 3), so that $A_2 = G_E/(\sqrt{\tau}\ G_M)$ which is always *equal* to $\sqrt{R}$. Measuring A_2 in elastic scattering is thus a very powerful method to determine the ratio G_E/G_M, which is especially useful if G_E is small as in the case of the neutron. As in the case of A_1, the asymmetry A_2 enters the measured electron asymmetry $A(el)$ with a factor describing the spin state of the virtual photon (in this case, this factor is $\sqrt{2\epsilon(1-\epsilon)}$), a second factor $sin(\theta^*)$ describing the orientation of the target spin [l] and an overall normalization $1+\epsilon R$. The complete expression for the electron asymmetry becomes then

$$A(el) = \sqrt{(1-\epsilon^2)}\,cos(\theta^*)\frac{A_1(\gamma^*)}{1+\epsilon R} + \sqrt{2\epsilon(1-\epsilon)}\,sin(\theta^*)\frac{A_2(\gamma^*)}{1+\epsilon R}. \qquad (27)$$

[l] We assume that the target spin lies in the electron scattering plane, and positive θ^* means that it points towards the scattered electron

In most (deep inelastic) experiments, one chooses the target spin direction either along the *incoming* electron beam direction, or perpendicular to that (again pointing towards the scattered electron). This means that $\theta^* = \theta_q$ (the virtual photon direction) in the first case and $\theta^* = \theta_q + 90^\circ$ in the second. The corresponding electron asymmetries can be written as

$$A_{||} = \frac{d\sigma^{\downarrow\uparrow} - d\sigma^{\uparrow\uparrow}}{d\sigma^{\downarrow\uparrow} + d\sigma^{\uparrow\uparrow}} = D(A_1 + \eta A_2) \tag{28}$$

and

$$A_{\perp} = \frac{d\sigma^{\downarrow\rightarrow} - d\sigma^{\uparrow\rightarrow}}{d\sigma^{\downarrow\rightarrow} + d\sigma^{\uparrow\rightarrow}} = d(A_2 - \zeta A_1). \tag{29}$$

The arrows indicate the relative orientation of electron and target spin. The factor $D = \sqrt{(1-\epsilon^2)}\, cos(\theta_q)/(1+\epsilon R)$ in front of the virtual photon asymmetry A_1 simplifies to

$$D = \frac{1 - \epsilon E'/E}{1 + \epsilon R} \tag{30}$$

and similarly

$$\eta = \frac{\epsilon\sqrt{Q^2}/E}{1 - \epsilon E'/E}\, ,\; d = \sqrt{\frac{2\epsilon}{1+\epsilon}} D \text{ and } \zeta = \frac{1+\epsilon}{2\epsilon}\eta. \tag{31}$$

Equations 28 – 31 show how one can extract A_1 and A_2 from measurements of $A_{||}$ and $A_{\perp}$.

After this somewhat lengthy preamble, we are now ready to introduce the spin structure functions g_1 and g_2. They can be defined in terms of the asymmetries A_1 and A_2 as

$$g_1(x, Q^2) = \frac{\tau}{1+\tau}\left(A_1(x, Q^2) + \frac{1}{\sqrt{\tau}} A_2(x, Q^2)\right) F_1(x, Q^2) \tag{32}$$

and

$$g_2(x, Q^2) = \frac{\tau}{1+\tau}\left(\sqrt{\tau} A_2(x, Q^2) - A_1(x, Q^2)\right) F_1(x, Q^2). \tag{33}$$

In the scaling limit ($\tau \to \infty$), g_1 converges to $g_1(x) = A_1(x)F_1(x)$ (becoming a function of x alone) since $|A_2|$ is bound by $\sqrt{R}$ which disappears in this limit. This allows us to interpret g_1 in terms of quark distribution functions again, in analogy with the form of $F_1(x)$ given in Eq. 15. The effect of the factor $A_1(x)$ is to give us a plus sign for each quark that is polarized in the same direction as the host nucleon, and a minus sign for each quark with opposite polarization (see Fig. 4). This follows from our assumption that we scatter elastically from asymptotically free quarks, and so only quarks which have

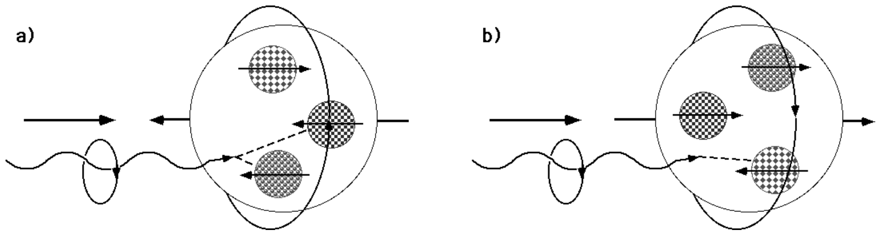

Figure 4: Polarized virtual photon absorption on quarks inside nucleons. One can see that only quarks with their spin parallel to the overall nucleon spin can contribute to $\sigma_T^{1/2}(\gamma^*)$ (a), while quarks with their spin opposite to the nucleon spin contribute to $\sigma_T^{3/2}(\gamma^*)$ (b). The virtual photon has positive helicity ($s_z(\gamma^*) = +1$) in these sketches.

their spin anti-aligned with the virtual photon helicity can absorb that virtual photon. Let us call $\Delta f(x)$ the probability of finding a quark of flavor f with momentum fraction x that has its spin parallel to the nucleon spin, minus the corresponding probability for antiparallel spin. Then we can write

$$g_1(x) = \frac{1}{2}\left(\frac{4}{9}\left[\Delta u(x) + \Delta\bar{u}(x)\right] + \frac{1}{9}\left[\Delta d(x) + \Delta\bar{d}(x) + \Delta s(x) + \Delta\bar{s}(x)\right]\right) \tag{34}$$

in complete analogy with Eq. 15. We can thus interpret $g_1(x)$ as the weighed sum of the quark momentum distributions times their helicity. (Formally, g_1 measures the "axial charge" of the quarks, not their helicity, which is different because of relativistic effects. However, we will ignore this distinction for the time being.)

The second spin structure function, g_2, also becomes a function of x alone in the scaling limit. However, it is harder to come up with an intuitive interpretation of this structure function. It is maybe better to look at the combination $g_T(x) = g_1(x) + g_2(x)$. From Eqs. 32, 33 one can see that this structure function is directly proportional to A_2, which means that g_T measures the interference between longitudinal and transverse photon-quark interaction in the case where the nucleon spin is perpendicular to the virtual photon direction: $g_T = \sqrt{\tau}A_2F_1 = \pm\, G_E/G_M\, F_1$, where the sign depends on the relative orientation of the quark spin and the virtual photon direction. Naively one would assume that the result is simply $g_T = g_1$, since quarks with spins parallel and antiparallel to the transverse polarization of the nucleon enter with opposite signs and $G_E/G_M = 1$ for the point-like quarks. However, relativistic effects play a big role here (a quark which has its spin pointing at 90 degrees relative to the virtual photon in its rest system will have a different transverse polarization in the Breit Frame, in general). This means that g_T is smaller than g_1 at large x in the scaling limit. The leading-twist prediction for g_T is the Wandzura-Wilczek[31] form

$$g_T^{WW}(x, Q^2) = \int_x^1 \frac{g_1(y, Q^2)}{y} \, dy, \tag{35}$$

implying $g_2^{WW} = -g_1 + g_T^{WW}$. Higher twist effects can contribute strongly to $g_2(x, Q^2)$, as well.

How does one determine the spin structure functions g_1, g_2 experimentally? There are in principle two different methods: One possibility is to measure the electron asymmetries $A_{||}$ and $A_\perp$ and to invert Eqs. 28, 29 to extract A_1 and A_2. These can be inserted in Eqs. 32, 33 to get the desired structure functions. This requires of course that F_1 is known with the necessary precision (from unpolarized DIS experiments and their parameterizations). Alternatively, one can write down the cross section difference directly in terms of g_1 and g_2, for instance

$$\frac{d\sigma^{\downarrow\uparrow}}{d\Omega dE'} - \frac{d\sigma^{\uparrow\uparrow}}{d\Omega dE'} = \frac{16\alpha^2(\hbar c)^2 E'^2 \, sin^2(\theta/2)}{Q^4} \left(\frac{E + E' \, cos\theta}{M\nu} g_1 - \frac{1}{\tau M} g_2 \right). \tag{36}$$

However, since absolute cross sections (and cross section differences) are much harder to measure precisely than asymmetries, the first method is usually chosen.

In any case, in addition to similar spectrometers as in the unpolarized case, one needs a polarized electron beam and a polarized nucleon target. This poses a formidable technical challenge, which is the main reason why polarized structure function measurements have a considerably shorter history than unpolarized ones (and the precision achieved so far is still inferior). The first experiments to measure spin structure functions began in the second half of the seventies[32,33], nearly a decade after the first unpolarized DIS experiments. These experiments were conducted at SLAC with polarized electrons created by photoionization of polarized ^{6}Li atoms. The targets were made from butanol where the protons in the bound hydrogen atoms can be polarized.

The first experiment to explore the region of small x was conducted at CERN[34], once again scattering the secondary muon beam from nucleon targets. This experiment (by the EMC collaboration) and its successors (by the SMC collaboration[35,36]) made use of the fact that the muons produced in pion decay are "automatically" polarized, since the decays are due to the Weak Interaction and therefore violate parity. The targets were again butanol (both with polarized protons and with polarized deuterons) as well as ammonia (see below), but of course truly massive amounts were needed to compensate for the low beam rate.

The highest *statistical* precision so far has been reached in a series of experiments carried out at SLAC in the years 1992 through 1997 (E142 [37], E143 [38,39,40,41], E154 [42] and E155). These experiments used a high intensity polarized electron beam (with increasing energy over the years) and polarized proton, deuteron and ^{3}He targets. The polarized electrons came from a laser-driven photo-emission source with a gallium arsenide cathode (after E142, the cathode was strained gallium arsenide which can produce electrons with polarization well over 80%). Three different targets were employed: a polarized ^{3}He target (E142, E154), an ammonia target (NH_3) with polarized protons (E143, E55) and deuterated ammonia (ND_3) with polarized deuterons (E143; E155 used Lithium Deuteride ^{6}LiD instead). The ^{3}He target was gaseous (10 atm pressure, contained in a glass cell up to 30 cm long) and was polarized by spin-exchange with optically (laser-) pumped Rubidium [37]. Since ^{3}He consists of one unpaired neutron and a (mostly) spin-0 proton pair, it can be considered an effective polarized neutron target, albeit with substantial background ("dilution") from the unpolarized protons as well as the glass walls and other target constituents.

The ammonia targets used both at SLAC and CERN use the principle of dynamic nuclear polarization to polarize the hydrogen and deuterium nuclei in NH_3 and ND_3. The ammonia is frozen in small beads and cooled to temperatures around 1 K in the presence of a 5 T magnetic field. Microwave power is radiated to the target to induce hyperfine transitions to the desired nuclear polarization state. Measurements on both ammonia and deuterated ammonia (or ^{6}LiD) yield the spin structure functions of protons and deuterons, which can be used to extract neutron results in a similar way as for F_1 and F_2 (see previous Section). Again, the dilution of the asymmetry by the unpolarized target constituents (nitrogen, liquid helium, and wall material) has to be taken into account carefully.

The scattered electrons are detected and analyzed in spectrometers which are similar to the ones used in unpolarized scattering. The SLAC experiments recorded data from two separate spectrometers (three for E155) simultaneously, each with large angular and momentum acceptance, to maximize the count rate. Figure 5 shows the spectrometer setup for E142 and E143. E154 and E155 employed similar spectrometers.

A novel approach to spin structure function measurements has been developed by the HERMES [43] collaboration at DESY. They use the circulating positron beam in the HERA collider and an internal (stationary) polarized gas target. The positrons are polarized transversely to the ring plane (this hap-

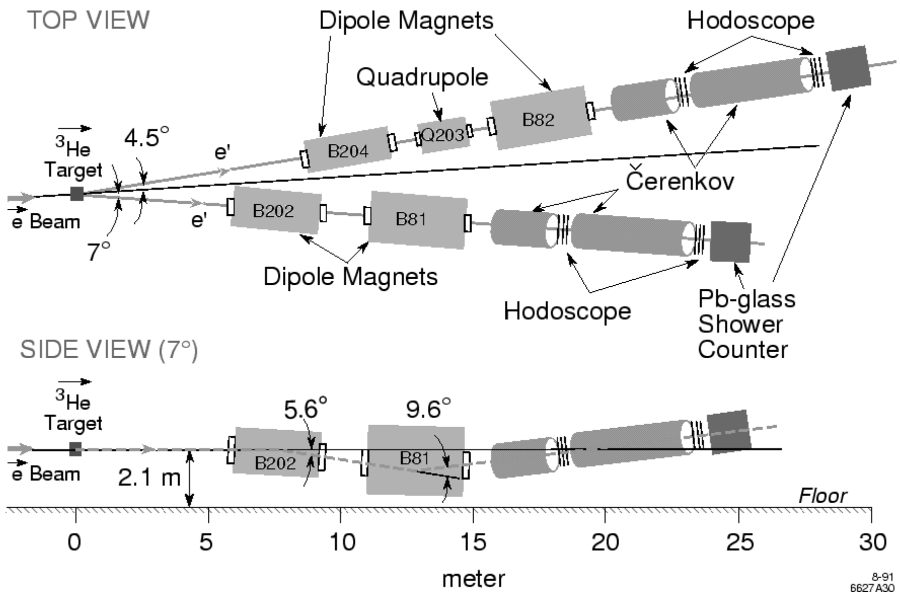

Figure 5: Spectrometers used for E142 and E143. Scattered electrons follow a curved trajectory through the momentum-analyzing magnets and are identified by Čerenkov counters. These are essentially large tanks of dilute gas which emits a flash of light - analog to a supersonic boom - when traversed by a particle moving faster than the speed of light, c/n, in the gas. The hodoscopes are made of a large number of scintillator "fingers" (long, narrow rectangular bars of organic material with attached phototubes to detect the scintillation light). They are arranged in vertical and horizontal arrays so they can determine the trajectory of the electrons through the spectrometer, which allows the experimenters to reconstruct their initial direction and momentum. The shower counters are made out of bars of lead glass which absorb the electrons and convert their energy to light. This light signal, which is proportional to the total electron energy, is also read out by phototubes.

pens "automatically" through synchrotron radiation over time - the so-called Sokolov-Ternov effect). This polarization is turned into the beam direction with a complicated arrangement of magnets in the target region (a "Siberian Snake"). The targets are windowless and contain pure nuclear species (polarized hydrogen, deuterium or ^{3}He). This avoids the dilution of the asymmetry by unpolarized target constituents. On the other hand, the achievable target densities are extremely low, which is partially offset by the high circulating current in a storage ring.

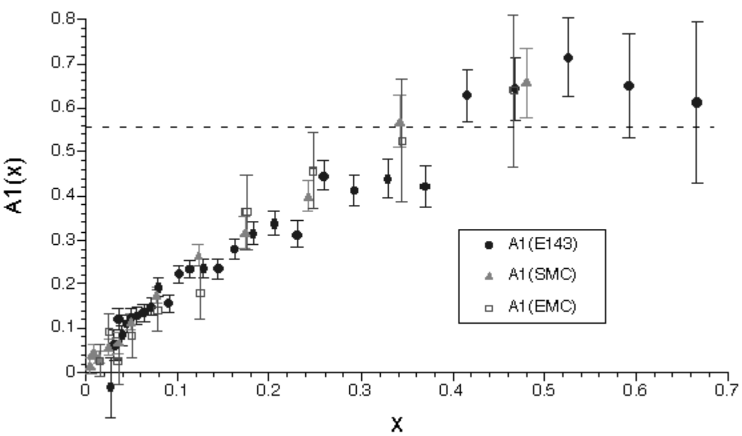

Figure 6: World's data on the asymmetry $A_1^p(x)$ of the proton in the deep inelastic region. The dashed line indicates the naive CQM prediction $A_1^p = 5/9$.

6 Polarized Structure Functions at high and low Q^2

Let us once again look at the information about the internal quark structure of the nucleon that can be gained from the measured spin structure functions, especially the structure function $g_1(x)$ and the asymmetry $A_1(x)$. From our simple CQM (Eq. 2 in Section 2) we can make definite predictions for the asymmetry A_1: for the proton, we expect

$$A_1 = \frac{\frac{4}{9}\Delta U + \frac{1}{9}\Delta D}{\frac{4}{9}U + \frac{1}{9}D} = \frac{\frac{16}{27} - \frac{1}{27}}{\frac{8}{9} + \frac{1}{9}} = \frac{5}{9} \tag{37}$$

This expectation is in reasonable agreement with the existing data at high x, as can be seen in Fig. 6, but of course our static model cannot account for the variation with x observed in the data. A similar calculation for the neutron (again using isospin to relate quark distributions in the neutron to equivalent ones in the proton) yields a predicted asymmetry of $A_1 = 0$, which is also quite close to the data (Fig. 7).

As in the case of unpolarized DIS, more fundamental information can be obtained from integrals over structure functions. In particular the integral over the spin structure function $g_1(x)$ has a very straightforward quark-parton interpretation (compare with Eq. 34):

$$\Gamma_1^p = \int_0^1 g_1^p(x)\, dx = \frac{1}{2}\left(\frac{4}{9}\Delta u + \frac{1}{9}\left[\Delta d + \Delta s\right]\right) \tag{38}$$

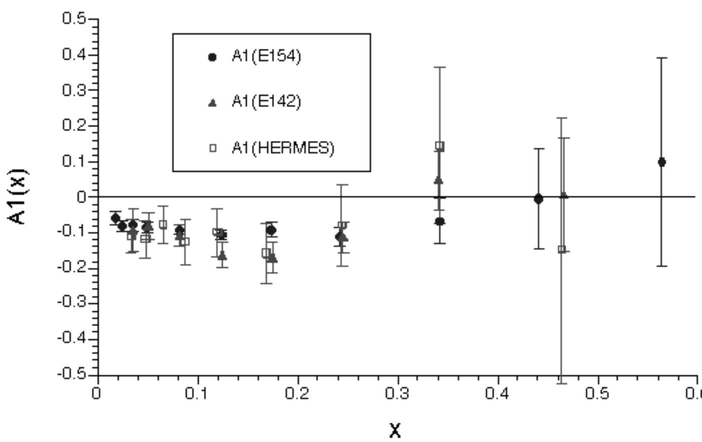

Figure 7: World's data on the asymmetry $A_1^n(x)$ of the neutron in the deep inelastic region. Not shown are data extracted from measurements on deuterium (SMC, E143). The naive CQM prediction is $A_1^n = 0$.

yields the spin probabilities defined in Section 2, weighed with the squared quark charges. The naive CQM prediction for the proton would be $\Gamma_1^p = 1/2 \times 5/9 = 0.278$ for this integral, and $\Gamma_1^n = 0$ for the neutron.

For a somewhat less model-dependent prediction, one can use the values for $\Delta u - \Delta d = 1.26$ and $\Delta u + \Delta d - 2\Delta s = 0.58$ extracted from beta-decays (see Section 2). However, one needs at least one additional assumption (since there are 3 unknowns). If we assume $\Delta s = 0$, which seems reasonable at first glance, we can predict a value of $\Gamma_1^p = 1/2(4/9 \times 0.92 + 1/9 \times (-0.34)) = 0.186$ for the proton and $\Gamma_1^n = -0.024$ for the neutron. This prediction was first derived by Ellis and Jaffe [44] and is referred to as the Ellis-Jaffe sum rule.

Surprisingly, the data lie quite a bit lower than both predictions, around 0.143 for the proton and -0.062 for the neutron. [m] This was first seen in the EMC experiment [34] and was dubbed the "spin crisis". One can use the experimental result as a third constraint on the three unknown quark polarizations and solve the resulting three linear equations. The results are that $\Delta u \approx 0.81$, $\Delta d \approx -0.46$ and $\Delta s \approx -0.12$. The contribution from strange quarks to the proton spin turns out to be nonzero (even negative), which

[m] The exact values depend quite strongly on both the QCD evolution of these integrals (see below) and the assumptions one makes about the unmeasured part of the integral (especially in the small-x region). The latter is a rather complicated issue, since traditional models for low-x extrapolation ("Regge fits") may not be accurate [42]. However, this problem is beyond the scope of this paper.

could explain why the Ellis-Jaffe sum rule is violated. Each of these values individually are not too far from the quark model expectations (once we take a $\sim 75\%$ relativistic reduction into account). However, the trouble is that all deviations are in the same direction (negative), which yields an overall contribution $\Delta\Sigma = \Delta u + \Delta d + \Delta s \approx .23$ of the quark spins to the overall proton spin (our naive QCM expectation is 1!).

Countless theoretical explanations for this discrepancy have been offered since the EMC results first appeared (see, e.g., Ref. [2] for an overview). Some arguments are based on non-trivial properties of QCD, for instance, the so-called "axial ($U(1)$) anomaly" or even the influence of instantons [45]. Within the quark model, a straightforward approach is to assume that some part of the proton spin is carried by orbital angular momentum of the quarks (maybe due to the same "meson cloud" already invoked in the explanation of the Gottfried sum rule violation in Section 4), and an additional contribution comes from the gluonic field (just like gluons make up 50% of the nucleon momentum, they could also account for 1/2 of the nucleon spin) [46]. More precise measurements of different observables will be needed to narrow down the range of possibilities (see Section 7).

A more fundamental prediction can be made for the difference between the proton and the neutron integral. In this case, the contribution from strange quarks cancels, and one gets

$$\Gamma_1^p - \Gamma_1^n = \frac{1}{2}\left(\frac{3}{9}\Delta u - \frac{3}{9}\Delta d\right) = \frac{1}{6}(\Delta u - \Delta d) = \frac{1}{6}g_A \tag{39}$$

where we used our result from Section 2 relating the quark polarizations to the axial coupling g_A. This result is the famous Bjorken sum rule [47], and its derivation requires no *ad hoc* assumptions about quark distributions. The Bjorken sum rule is nowadays considered a crucial test of QCD. The numerical result is 0.21 and has been confirmed by all experiments up to now (after PQCD corrections), to within one standard deviation and with a relative precision of about 7%. It may sound surprising that a quantity measured at very high energies, in DIS, can be expressed in terms of a "low-energy" parameter like g_A measured in neutron beta decay. However, as pointed out before, the weak interaction is essentially pointlike, and the relevant scale is not the neutron decay energy, but the mass of the exchanged W boson, which is also large.

For completeness, we should mention that there is also a sum rule for the integral over the structure function g_2, the Burkhardt-Cottingham sum rule [48]. This sum rule predicts that the integral yields zero for both the proton and the neutron (this result can be inferred from Eq. 35, but is more general than that equation). The (limited) existing data (mostly from E143 [41]) agree with this prediction. At least on average it *is* therefore true that $g_T(x) \approx g_1(x)$, which was our first ("naive") assumption.

All of these sum rules, as well as the structure functions and asymmetries themselves, exhibit some degree of scaling violation (Q^2-dependence), just like the unpolarized structure functions, and for the same reasons. For instance, a "constituent" quark with spin "up" can be resolved at higher Q^2 into a quark with opposite spin and a gluon whose spin 1 accounts for the difference. The logarithmic Q^2 dependence of g_1 and g_2 is governed by similar evolution equations as in the unpolarized case. Several theoretical groups [49,50] have conducted "next-to-leading order" (NLO) analyses of the measured structure functions based on these DGLAP evolution equations.

Since both polarized (g_1, g_2) and unpolarized (F_1, F_2) structure functions depend logarithmically on Q^2, one could guess that the ratio g_1/F_1 or the asymmetry A_1 (which is close to that ratio) might scale much better (depend much less on Q^2) than each of the structure functions. Experimentally, the precision reached is not quite good enough yet to decide this question unambiguously. Again, the most extensive data set is from E143 [40]. These data show rather strong Q^2 dependence for g_1/F_1 below $Q^2 = 1$ GeV^2, but no statistically significant dependence above that value (see also Fig. 6 which shows measurements of A_1^p at vastly different incident energies and therefore different Q^2). Because of our limited knowledge of scaling violations in polarized structure functions, we can not yet pin down the polarized gluon distribution $\Delta g(x, Q^2)$ with sufficient precision, since it has to be inferred from the Q^2-dependence of the quark distributions.

The sum rules do not depend directly on the resolution (momentum transfer) at which the measurement is done, but they exhibit Q^2-dependence due to QCD radiative effects. The Bjorken sum rule, for instance, changes from its asymptotic value of 0.21 to a value of 0.18 (14% less) at $Q^2 \approx 5$ GeV^2. One has to apply these corrections (up to third order pQCD) to achieve agreement with the data. This agreement is considered a strong (successful) test of QCD.

At even lower momentum transfers, higher twist corrections (and other $1/Q^2$ terms like target mass corrections) appear in the polarized structure functions as well. It is quite possible that the significant Q^2-dependence of the ratio g_1/F_1 observed at $Q^2 \leq 1$ GeV^2 (see above) is due to such higher twist effects - it can be fit rather well with an additional $1/Q^2$ term in the parametrization of the world data [40]. Higher twist terms are especially important in the g_2 structure function, leading to a deviation from the Wandzura-Wilczek form Eq. 35. At present, the data on g_2 are not precise enough to reveal any such deviation; however, it is expected that future measurements (especially the extension of E155, which is scheduled for 1999) will reach the necessary precision.

The contribution of the additional $1/Q^2$ terms discussed above can also become important for the sum rules at low Q^2. As an example, the Ellis-Jaffe integral for the proton Γ_1^p receives an extra term of the form [51]

$$\Gamma_1^p(Q^2) = \Gamma_1^p(\infty) + \frac{m_n^2}{9Q^2}(a^{(2)} + 4d^{(2)} + 4f^{(2)}). \tag{40}$$

Some of the coefficients $a^{(2)}, d^{(2)}$ and $f^{(2)}$ can be related to higher moments of the spin structure functions via the Operator Product Expansion (OPE) method. For instance, the twist-2 coefficient $a^{(2)}$ is equal to twice the third moment of g_1:

$$a^{(2)} = 2\int_0^1 x^2 g_1(x)\, dx, \tag{41}$$

while the twist-3 term $d^{(2)}$ can be calculated from g_2:

$$d^{(2)} = 3\int_0^1 x^2 (g_2(x) - g_2^{WW}(x))\, dx. \tag{42}$$

Equation 42 shows that the deviation of g_2 from the Wandzura-Wilczek form is indeed a measure of higher-twist effects, as advertised.

The twist-4 term $f^{(2)}$ cannot be measured directly, but has been estimated (together with $d^{(2)}$) using several different approaches, including the bag model [52] and QCD sum rules [53]. These predictions, together with the data [41], indicate that higher twist effects may change the proton integral by as much as $3\% - 5\%$ at $Q^2 = 3$ GeV^2.

There is another important aspect to consider when one studies the Q^2-dependence of the integrals Γ_1^p and Γ_1^n. As Q^2 decreases, a larger and larger part of the integration extends into the resonance region $W^2 \leq 4$ GeV^2. As an example, at $Q^2 = 3$ GeV^2 already half of the integration range, $x \geq 0.5$, lies in that region (although the dominant contribution to the integral still comes from smaller x). As the resonances become more important, final state interactions will profoundly change the shape of the structure functions and asymmetries (of course, these final state interactions are indeed one of the major sources of higher-twist effects). For example, while A_1^p is a relatively large, positive number at large x in the deep inelastic region (see Fig. 6), it will become *negative* in that same x region once Q^2 is small enough to bring it into the range of the Δ resonance (see our discussion in the previous Section). Correspondingly, the integral Γ_1^p will rapidly decrease in size and eventually become negative overall as Q^2 approaches zero.

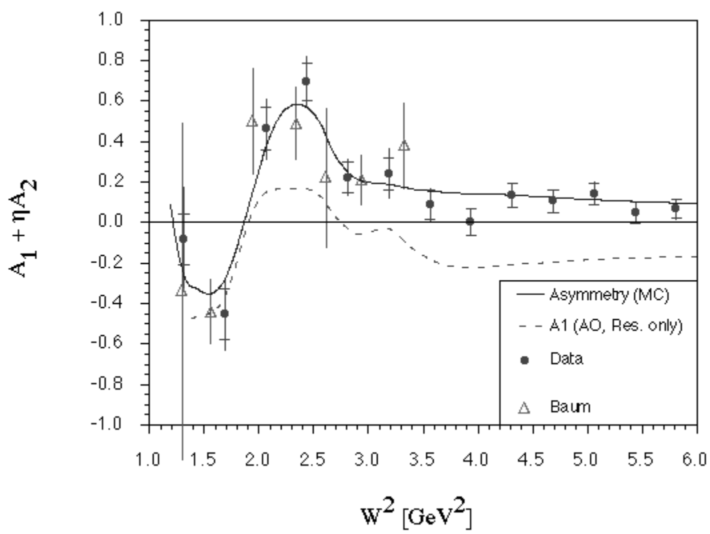

Figure 8: Existing data on the asymmetry A_1 (with some contamination from A_2) in the resonance region, at an average Q^2 of 0.5 GeV^2. The circles are the E143 data, while the triangles come from the E130 measurement. The solid line is the result of a parametrization based on pion production data, adjusted to represent the present data. The dashed line is the contribution to A_1 from the nucleon resonances only.

There is another important sum rule that constrains this behavior of $\Gamma_1^p(Q^2)$ and $\Gamma_1^n(Q^2)$ as $Q^2 \to 0$, the famous Gerasimov-Drell-Hearn (GDH) sum rule [54]. Originally, this sum rule made a prediction about the *real* photon absorption cross sections (at $Q^2 = 0$) defined in the previous Section:

$$\int_{\nu_{thr}}^{\infty} \sigma_T^{1/2}(\nu) - \sigma_T^{3/2}(\nu)\, \frac{d\nu}{\nu} = -\frac{2\pi^2\alpha}{M^2}\kappa_N^2, \tag{43}$$

where ν_{thr} is the threshold energy for pion photoproduction and κ_N is once again the anomalous magnetic moment of the nucleon under study. If one assumes that these cross sections connect smoothly to the virtual photon ones (at $Q^2 > 0$), one can use Eqs. 25-32 to predict the behavior of Γ_1 close to the photon point:

$$\Gamma_1^N(Q^2 \to 0) \to \frac{Q^2}{16\pi^2\alpha} \int_{\nu_{thr}}^{\infty} \sigma_T^{1/2}(\nu) - \sigma_T^{3/2}(\nu)\, \frac{d\nu}{\nu} = -\frac{Q^2}{8M^2}\kappa_N^2. \tag{44}$$

In other words, the GDH sum rule predicts that $\Gamma_1^p(Q^2)$ must go to zero for $Q^2 \to 0$ and must have a *negative* slope before turning over to the positive DIS

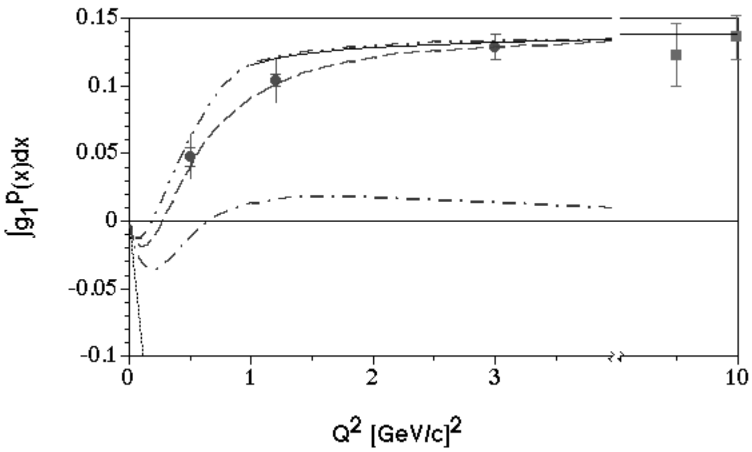

Figure 9: The Q^2-dependence of the integral Γ_1^p. The data shown are from SLAC (E143, at $Q^2 \leq 3$ GeV^2) and CERN (EMC, SMC). The solid line shows the Q^2-evolution of the integral in the DIS domain, while the dotted line indicates the slope predicted by the GDH sum rule. The remaining lines are from models explained in the text.

value. This rather complicated behavior is indeed dominated by the contribution from the Δ resonance.

Inclusive data in the resonance region (at low to moderate Q^2) have been collected by the E130 collaboration [55] (during the first round of experiments on the proton at SLAC) and later again by the E143 collaboration [56] (on both protons and deuterons, with higher statistics). These data exhibit (for the most part) the expected behavior (see Fig. 8): The asymmetry in the region of the Δ resonance is indeed negative and compatible with $A_1 = -0.5$. However, the higher resonances exhibit rather large positive asymmetries, even at low Q^2, which means that the integral Γ_1^p stays positive down to rather small Q^2.

The Q^2-dependence of the integral is shown in Fig. 9. Together with the logarithmic Q^2-evolution (solid line) and the slope predicted by the GDH sum rule (dotted line), two models are shown that attempt to interpolate between the deep inelastic region and the photon point. The model by Soffer and Teryaev [57] (dash-double-dot curve) begins with a smooth interpolation of the integral over the structure function $g_T(x, Q^2)$ (defined earlier in this Section) which converges to Γ_1 at high Q^2 and remains positive down to the photon point. They subtract the contribution from the integral over g_2 (which can be evaluated using the Burkhardt-Cottingham sum rule) to obtain the integral $\Gamma_1(Q^2)$ alone. The strong variation seen in $\Gamma_1(Q^2)$ is due to the variation in the integral over g_2 in their model. Burkert and Ioffe [58] use a parametrization of measured pion electro- and photoproduction amplitudes (encoded in the program "AO") to calculate the contribution from the nucleon resonances to $\Gamma_1(Q^2)$ (dot-dashed curve). They add to this a term that depends smoothly on Q^2 and interpolates between the part that is "missing" at $Q^2 = 0$ to saturate the GDH sum rule and the full value of Γ_1 in the high-Q^2 limit. The resulting curve (dashed) describes all existing data best. The deviation of these data from the deep inelastic evolution can be used to extract information about higher twist contributions to the integral [59]. The resulting correction for Γ_1^p is of order 1% at $Q^2 = 3$ GeV^2, somewhat smaller than the theoretical predictions discussed earlier.

7 Future Measurements

A truly vast amount of data on the inelastic structure of the nucleon has been accumulated over the past 30 years, from both fixed target and colliding beam experiments, with polarized and unpolarized incident electrons, muons and (anti-)neutrinos, on a variety of targets from (polarized and unpolarized) hydrogen through iron. Yet there are still some very fundamental questions which await a definite answer, and correspondingly an impressive number of future experiments (both planned and underway) to address these questions.

In the case of the spin-averaged quark structure of the nucleon, as expressed in the structure functions $F_1(x)$ and $F_2(x)$, much of the recent progress has come from experiments at the proton-positron collider HERA. Using a 800 GeV countercirculating proton beam instead of a fixed gas or liquid target has allowed the experimenters to reach hundred times smaller x or several hundred times larger Q^2 than even the highest energy fixed target experiments (NMC at CERN and E665 at Fermilab). Not only has this new kinematic domain opened up a much more precise determination of gluon and sea quark distributions (via scaling violations and low-x data), but novel phenomena (like "pomeron exchange" in diffractive scattering) may have been observed. A much steeper rise of F_2 than expected was observed for ever smaller x and increasing Q^2. One of the most intriguing discoveries so far is an observed "excess" of events at extremely high Q^2 beyond $15,000$ GeV^2. [60] One possible interpretation is that these events herald the first glimpse of an even deeper layer of matter, with the prospect of future measurements of the "quark structure functions" in analogy to the nucleon ones.

These and other directions of inquiry will keep HERA productive and (hopefully) exciting for many years to come. Complementary information can be expected from existing (Fermilab) and future (LHC) hadron colliders, for instance by studying the so-called Drell-Yan process, in which a lepton pair is created in nucleon- (anti-)nucleon collisions, opening a different window on the quark structure of nucleons. Ultimately, one could conceive colliding the multi-TeV proton beam of LHC with the 80 GeV electron beam of LEP (both will be housed in the same tunnel at CERN), thereby increasing the kinematic range over that of HERA by another large factor.

In the area of polarized structure functions, a lot less is known at present, and correspondingly an even wider array of additional experiments are being prepared, planned or at least discussed. Some of the unsolved questions and puzzles that hopefully will be addressed within the next decade follow below:

1. We still haven't completely unraveled the contributions of the different quark flavors to the spin structure function $g_1(x)$. More high-precision data for proton and deuteron spin structure functions are forthcoming from SLAC experiment E155, which is presently in the analysis stage. However, with only 2 linearly independent measurements (protons and neutrons), one cannot hope to pin down the separate contributions of the 2 valence quark flavors and the sea quarks of all flavors (up, down, strange and maybe even charmed). The solution employed in the case of unpolarized structure functions, measurements with (anti-) neutrino beams, seems out of the question at least at the present, since polarized targets of the necessary thickness haven't been conceived yet.

   A possible solution may lie in the measurement of semi-inclusive reactions, where a meson created in the deep inelastic scattering process is detected in coincidence with the scattered lepton. If one concentrates on mesons which carry a large fraction z of the transferred energy ν, one can expect that the struck quark should be embedded in those mesons. Therefore, observing a "leading" kaon would yield information on the strange quark contribution to the nucleon spin, while the comparison of positively and negatively charged pions constrains u and d quark distributions. First experimental results along these lines have been obtained at HERMES and by SMC. [62] HERMES is upgrading their detector to better separate kaons and pions and will collect more data in the near future. A new proposal at CERN named COMPASS (Common Muon and Proton Apparatus for Structure and Spectroscopy) will extend these studies to higher energies.

2. Our knowledge of the polarized gluon distribution is rather limited at present. The existing data can not even completely rule out the possibility that gluons contribute little or nothing to the spin of the nucleon. In view of the realization that quark spins also contribute only a small

amount, this issue is of great importance. Improved statistics data at both moderate and larger Q^2 obtained by E155 (using a third spectrometer at the relatively large scattering angle of 10°) will help extract the gluon distribution via the observation of scaling violations. However, the precision achieved (and the kinematic range covered) is still far inferior to the unpolarized data.

One obvious solution is to use the same device that already expanded our knowledge of spin averaged gluon distributions, HERA, by colliding polarized protons at 800 GeV with polarized electrons or positrons. This has indeed been proposed [61], but the necessary investment is formidable. Other possibilities include a direct observation of the photon-gluon fusion process (with the detection of charmed mesons as the tell-tale sign), which could be undertaken at SLAC (Proposal E156), COMPASS and HERA. Finally, the Drell-Yan process with polarized proton-proton collisions would open another possible window, which is being considered for the new heavy ion collider RHIC in Brookhaven.

Most of these experiments (at least the ones at HERA and RHIC) would also improve our knowledge of the small-x behavior of g_1, which is presently the largest uncertainty in the evaluation of the integrals Γ_1. Only if we really can extrapolate to $x = 0$ with some confidence will we finally know with certainty whether the Bjorken sum rule is fulfilled and how much of the nucleon spin is accounted for by the sum of the quark spins.

3. Once we know how much angular momentum is carried by gluon and quark spins, we should be able to account for the remainder in the form of orbital angular momentum. At present, practically nothing is known about this issue experimentally, since it depends on the transverse motion of quarks which is not measured in DIS. One approach would be to arrive at a consistent description of the nucleon in terms of constituent quarks and mesons, where the latter carry most of the orbital angular momentum. Experimentally, this would require a much more refined study of the resonance region which constitutes a major part of the scientific program at CEBAF/Jefferson Lab (mostly in Hall B).

   A novel method to study orbital angular motion has been proposed that uses the process of real photon emission after absorption of a (large Q^2) virtual photon, dubbed "Deeply Virtual Compton Scattering" (DVCS).[63] Plans for experiments at an upgraded CEBAF (with 8 GeV or higher beam energy) are presently being formulated.

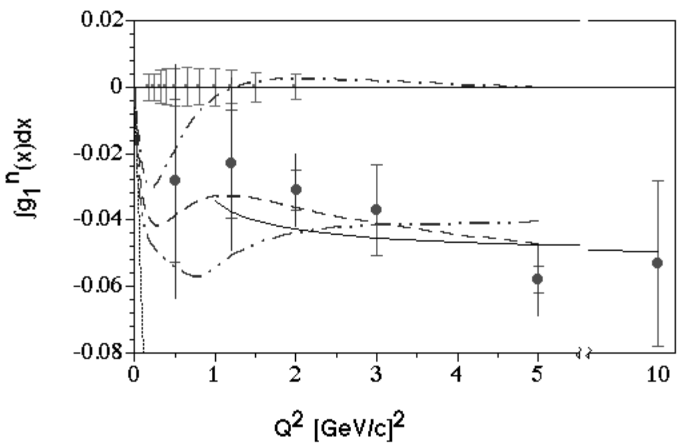

Figure 10: The Q^2-dependence of the integral Γ_1^n. The data shown are from SLAC (E142, E143 and E154, at $Q^2 \leq 5$ GeV^2) and CERN (SMC). The curves come from the same theoretical models as the correspondent ones in Fig. 9. The series of error bars shown at zero values are predicted statistical errors for experiment 93-009 in Hall B.

4. The transverse spin structure function g_T also deserves more attention than it has received so far. It contains complementary information on the nucleon spin, and in particular on higher twist effects. An extension of E155 (called E155x) with mostly transverse target polarization is planned for 1999. Other proposals are being considered at CEBAF/Jefferson Lab.

5. Both higher twist effects and the transition from the DIS picture of quasi-free current quarks to the constituent quarks in the non-perturbative regime of QCD can be studied by measuring spin structure functions and the integrals Γ_1 at low Q^2 down to the photon point. In particular the GDH sum rule has not been rigorously tested yet, but a large number of experiments at low (MAMI, LEGS) to medium (ELSA, GRAAL, CEBAF) energy accelerators have been proposed and are in various stages of preparation. The connection between this sum rule and the deep inelastic region will be studied by several experiments at Jefferson Lab, in all three experimental Halls. Similar targets (p, d and ^{3}He) as in the SLAC experiments will be used. The most comprehensive program has been approved for Hall B, encompassing a complete mapping of the spin structure functions of the proton (NH_3 target) and the deuteron (ND_3 target) from small ($Q^2 \simeq 0.1$ GeV^2) to moderate ($Q^2 \simeq 2.0$ GeV^2) mo-

mentum transfer, and from the elastic peak over the resonance region up to the highest energies available at CEBAF (well into the deep inelastic region). Figure 10 shows the expected density of data points and their estimated statistical errors for the integral $\Gamma_1^n(Q^2)$ from the ND_3 part of this program. The first part of this so-called EG1 running period (which contains both experiments with NH_3 and ND_3 targets, as well as a partial measurement of the GDH sum rule) is scheduled for the second half of 1998.

In summary, there is a truly impressive amount of work still ahead of us, and plenty of opportunities for surprises. Maybe 10 years from now we will finally have a complete picture of the quark and gluon composition of the nucleon. Both experimental progress as well as theoretical advances on the problem of non-perturbative QCD will have contributed to that picture.

Acknowledgments

I would like to express my thanks to my many collaborators on the E142, E143, E154 and E155 collaborations at SLAC, as well as my colleagues in the EG1 running group in CEBAF's Hall B. My research is supported by DOE under grant DE-FG02-96ER40960.

References

1. A.M. Cooper-Sarkar, R.C.E. Devenish, and A. de Roeck, hep-ph/9712301 (Dec. 1997).
2. G.P. Ramsey, hep-ph/9702227 (Feb. 1997);
   H. Böttcher, hep-ph/9712458 (Dec. 1997).
3. Halzen and Martin, "Quarks and Leptons: An Introductory Course in Modern Particle Physics", John Wiley —& Sons (1984).
4. Povh, Rith, Scholz and Zetsche, "Particles and Nuclei", Springer Verlag (1995).
5. S.M. Wong, "Introductory Nuclear Physics", Prentice Hall (1990).
6. Particle Data Group, "The Review of Particle Physics", *Phys. Rev.* D **54**, 1 (1996).
7. R.G. Roberts, "The Structure of the Proton", Cambridge University Press (1990).
8. Carlson and Domingo (Eds.), Proceedings of the 14th International Conference on Particles and Nuclei, World Scientific (1997).
9. De Jager, Ketel, Mulders, Oberski and Oskam-Tamboezer, Proceedings of the 12th International Symposium on High-Energy Spin Physics,

World Scientific (1997).
10. S. Capstick and N. Isgur, *Phys. Rev.* D **34**, 2704 (1986).
11. F.E. Close and Zh-P. Li, *Phys. Rev.* D **42**, 2194 and 2207 (1990).
12. See the SLAC Website at http://www.slac.stanford.edu:80/
13. See the Jefferson Lab Website at http://www.cebaf.gov/#general
14. See the DESY Website at http://www.desy.de/
15. See the CERN Website at http://www.cern.ch/
16. See the Fermilab Website at http://www.fnal.gov/fermilab_home.html
17. C.G. Callen and D. Gross, *Phys. Rev. Lett.* **22**, 156 (1969).
18. K. Gottfried, *Phys. Rev. Lett.* **18**, 1154 (1967).
19. L.W. Whitlow *et al.*, *Phys. Lett.* B **250**, 193 (1990).
20. EMC, J.J. Aubert *et al.*, *Phys. Lett.* B **123**, 275 (1983).
21. NMC, M. Arneodo *et al.*, *Phys. Rev.* D **50**, R1 (1994).
22. A.W. Thomas and W. Melnitchouk, hep-ph/9708484.
23. P.S. Auchinloss *et al.*, *Z. Phys.* C **48**, 411 (1990).
24. J.P. Berge *et al.*, *Z. Phys.* C **49**, 187 (1991).
25. G.T. Jones *et al.*, *Z. Phys.* C **62**, 575 (1994).
26. Yu. Dokshitzer, *Sov. Phys. JETP* **46**, 641 (1977);
V.N. Gribov and L.N. Lipatov, *Sov. J. Nucl. Phys.* **15**, 438,675 (1972);
G. Altarelli and G. Parisi, *Nucl. Phys.* B **126**, 298 (1977).
27. A.D. Martin, R.G. Roberts and W.J. Stirling, *Phys. Lett.* B **354**, 155 (1995).
28. H. Lai *et al.*, *Phys. Rev.* D **55**, 1280 (1997).
29. E140x, L.H. Tao *et al.*, *Z. Phys.* C **70**, 387 (1996).
30. O. Nachtmann, *Nucl. Phys.* B **63**, 237 (1973).
31. S. Wandzura and F. Wilczek, *Phys. Lett.* B **72**, 195 (1977).
32. E80, M.J. Alguard *et al.*, *Phys. Rev. Lett.* **37**, 1261 (1976) and *Phys. Rev. Lett.* **41**, 70 (1978).
33. E130, G. Baum *et al.*, *Phys. Rev. Lett.* **51**, 1135 (1983).
34. EMC, J. Ashman *et al.*, *Phys. Lett.* B **206**, 364 (1988) and *Nucl. Phys.* B **328**, 1 (1989).
35. SMC, D. Adams *et al.*, *Phys. Lett.* B **329**, 399 (1994); SMC, D. Adams *et al.*, *Phys. Rev.* D **56**, 5330 (1997) and SMC, B. Adeva *et al.*, CERN-PPE/97-118 (1997).
36. SMC, B. Adeva *et al.*, *Phys. Lett.* B **302**, 533 (1993) and SMC, D. Adams *et al.*, *Phys. Lett.* B **396**, 338 (1997).
37. E142, P.L. Anthony *et al.*, *Phys. Rev. Lett.* **71**, 959 (1993) and *Phys. Rev.* D **54**, 6620 (1996).
38. E143, K. Abe *et al.*, *Phys. Rev. Lett.* **74**, 346 (1995).
39. E143, K. Abe *et al.*, *Phys. Rev. Lett.* **75**, 25 (1995).

40. E143, K. Abe *et al.*, *Phys. Lett.* B **364**, 61 (1995).
41. E143, K. Abe *et al.*, *Phys. Rev. Lett.* **76**, 587 (1996).
42. E154, K. Abe *et al.*, *Phys. Rev. Lett.* **79**, 26 (1997); *Phys. Lett.* B **404**, 377 (1997) and *Phys. Lett.* B **405**, 180 (1997).
43. HERMES, K. Ackerstaff *et al.*, *Phys. Lett.* B **404**, 383 (1997).
44. J. Ellis and R.L. Jaffe, *Phys. Rev.* D **9**, 1444 (1974).
45. N.I. Kochelev, hep-ph/9710540 (Oct. 1997).
46. V. Barone, T. Calarco, and A. Drago, hep-ph/9801281 (Jan. 1998).
47. J.D. Bjorken, *Phys. Rev.* **148**, 1467 (1966) and *Phys. Rev.* D **1**, 1376 (1970).
48. H. Burkhardt and W.N. Cottingham, *Ann. Phys.* **56**, 453 (1970).
49. M. Stratmann, hep-ph/9710379 (Oct. 1997).
50. D. de Florian, hep-ph/9710378 (Oct. 1997).
51. B. Ehrnsperger, L. Mankiewicz, and A. Schäfer, *Phys. Lett.* B **323**, 439 (1994).
52. X. Ji and P. Unrau, *Phys. Lett.* B **333**, 228 (1994); X. Song, *Phys. Rev.* D **54**, 1955 (1996).
53. I.I. Balitsky, V.M. Braun, and A.V. Kolesnichenko, *Phys. Lett.* B **242**, 245 (1990) and *Phys. Lett.* B **318**, 648 (1993).
54. S. Gerasimov, *Sov. J. Nucl. Phys.* **2**, 430 (1966); S.D. Drell and A.C. Hearn, *Phys. Rev. Lett.* **16**, 908 (1966).
55. E130, G. Baum *et al.*, *Phys. Rev. Lett.* **45**, 2000 (1980).
56. E143, K. Abe *et al.*, *Phys. Rev. Lett.* **78**, 815 (1997).
57. J. Soffer and O.V. Teryaev, *Phys. Rev.* D **51**, 25 (1995).
58. V.D. Burkert and B.L. Ioffe, *Phys. Lett.* B **296**, 223 (1992).
59. X. Ji and W. Melnitchouk, *Phys. Rev.* D **56**, R1 (1997).
60. C. Adloff *et al.*, *Z. Phys.* C **74**, 191 (1997); J. Breitweg *et al.*, *Z. Phys.* C **74**, 207 (1997).
61. A. De Roeck *et al.*, , hep-ph/9801300 (Jan. 1998).
62. SMC, B. Adeva *et al.*, hep-ex/97110080 (Nov. 1997).
63. C.E. Hyde-Wright, *Nucl. Phys.* A **622**, 268c (1997).