Mapping out symmetry violation in nucleon structure or “What’s wrong with QCD?”

John Arrington
Argonne National Laboratory
9700 S. Cass Ave, #203
Argonne, IL 60439, USA

ABSTRACT

A good introduction to the quark sub-structure of matter (protons and neutrons) provides simple pictures that build on the similarities to other bound states: molecules, atoms and nuclei. The observation of behavior consistent with or deviating from the analogies helps us to design measurements that answer key questions about the internal structure of complex systems.

However, when it comes to the detailed quark-gluon structure of matter, these simple, intuitive pictures are fundamentally and dramatically flawed. I will discuss the unique and often problematic nature of QCD and the reasons why so few of these analogies provide complete or compelling descriptions of hadrons. In spite of this, these simple pictures still play a key role in guiding our thinking and providing a starting ground from which to build the intuition necessary to make progress in the study of hadronic physics.

I will start by discussing the sub-structure of atoms, nuclei, and nucleons, with the aim of highlighting the unique nature of hadronic matter. I will then discuss the impact this has on the field of hadronic physics, and how we can use simple models, and in particular test the symmetries of these models, to better understand the nature of hadrons. I will end by examining the study of nucleon form factors, emphasizing the role that testing these simple models and symmetries plays.

I – INTRODUCTION

The theme of the school this year was “Symmetries in Subatomic Systems” and my topic was “Mapping out symmetry violation in nucleon structure”. While I do not usually think of probing nucleon structure in the context of studying symmetries, measurements aimed at testing a particular feature or assumed symmetry of a simple model are a large part of the historic study of the structure of the proton and neutron. Understanding the symmetries involved is often critical in trying to gain insight into complex systems. This is especially true in hadronic physics where the complex dynamics of bound quark systems leads to the breakdown of many of our simple pictures and our intuitive understanding of more familiar bound systems.

This is a paper in two parts. The first is a discussion of the unique and often vexing nature of QCD, designed to give a feeling of the challenges in understanding the structure of hadronic matter. From there, I will discuss how the study of hadronic physics is shaped by the exotic nature of QCD. The goal here will be to provide a general picture of the issues involved rather than a precise quantitative description.

The second part discusses the measurement of proton and neutron form factors, which I take as a prototypical example of a program that can tell us about the nature of matter in spite of the difficulties in interpretation resulting from our limited practical understanding of QCD. I will provide a broad overview of such measurements, focusing on the remarkable progress made so far.

II – A brief review of the structure of matter: The road to QCD

Early studies of matter assumed that atoms were the fundamental, indivisible elements of matter. As elements were discovered and their properties studied, elements were categorized into the periodic table, guided by patterns in their behavior. That fact that the atom was divisible, with multiple layers of structure yet to be discovered, did not preclude such categorization or make it less relevant when the underlying structure of atoms was discovered. This sub-structure would in fact provide a much simplified underlying picture to explain the somewhat complicated periodic table and the patterns observed in the behavior of its elements.

The discovery of the sub-structure of atoms led to many different pictures of the atom which aimed to explain the quantization of mass and charge of atoms. Studies of cathode rays provided evidence for a light, negatively charged constituent of atoms, which were later observed to be the same particle as those observed in some radioactive decays which would eventually be called electrons. The production of positively charged ions (canal rays or anode rays) from gasses showed that atoms also contained heavy, positively charged constituents. While the electrons were universal, the canal rays from different atoms had difference charge-to-mass ratios. The picture of the nucleus as a heavy, positive core with a number of light, negative ions led to the plum pudding model of the atom (fig. 1), where a large positive core contains several small, negatively charged electrons, and the elements in the periodic table are arranged according to the charge of the positive core.

The Rutherford gold foil experiment demonstrated that this picture was not correct, and that the positive charge was contained in a small, dense core, leading to the Rutherford or planetary model of the atom (fig. 1). It was also proposed that the nuclear charge and the atomic number were not just similar but identical. In scattering alpha particles from nitrogen gas, Rutherford detected hydrogen nuclei, suggesting that the nitrogen nucleus must contain hydrogen nuclei, which were then called protons and taken to be the elementary particles from which nuclei were built, thus explaining the identity between mass number and nuclear charge. With the discovery of the neutron, a picture of the atom containing all of the basic ingredients was complete. As we obtained a better understanding of quantum mechanics and the electronic structure of atoms, we end up with more complete and accurate pictures such as the Rutherford-Bohr and Schrodinger models of the atom (fig. 1). These pictures provide a natural explanation of the structure observed in the periodic table. Rather than "breaking" our understanding of the elements, the construction of atoms from a set of simple constituents provided the relations between mass and charge, while the electronic structure explained the categorization of elements into the families of the periodic table.

While these models included the main ingredients of the atom, the proton, neutron, and electron, the early models treated these as fundamental particles until a measurement of the proton magnetic moment demonstrated that it was not a structureless Dirac particle. This led to investigations of the sub-structure of the proton and neutron, and eventually to models of their structure in terms of quarks and gluons. Early quark models described the proton and neutron in terms of three heavy "constituent quarks", each of which accounts for approximately one-third of the proton's 938 MeV/c^2 mass. In this description, the proton consists of two up quarks with charge +2/3 and one down quark with charge of -1/3.

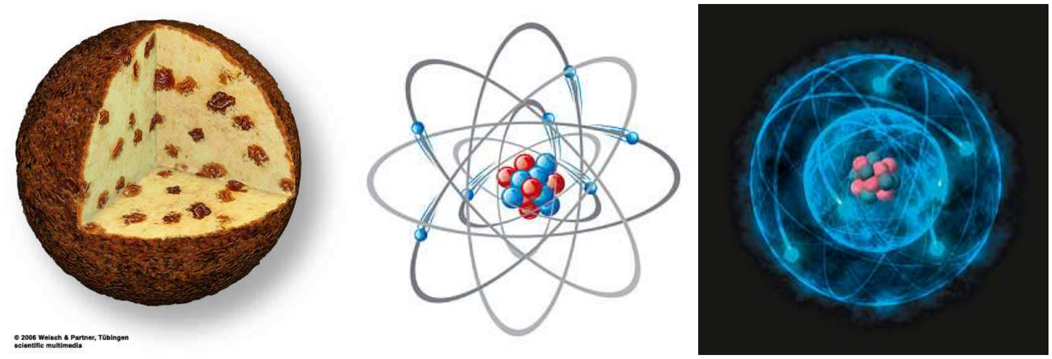


Figure 1: Evolution of atomic models. The J.J. Thomson plum pudding model (left), Rutherford planetary model (center), and quantum mechanical models (right).

The neutron consists of one up and two down quarks. This explains the charge of the proton and neutron. The fact that the constituent quarks are assumed to be nearly identical except for the charge yields the nearly identical proton and neutron masses. However, such simple constituent quark models (CQMs) had difficulty describing the zoo of hadrons that had been discovered, as clear from the masses of the lightest mesons ($M_{\pi,K,\eta,\rho\rangle} \approx 140, 490, 550, 790$ MeV/c^2; $M_{\Lambda,\Sigma,\Delta} \approx 1120, 1190, 1230$ MeV/c^2). Again, while there were some significant limitations with these models, they provided a simple picture of the constituents of hadronic matter which explained some of the observed features (such as the families of baryons and mesons) and which left the overall picture of the nucleus, modeled in terms of protons and neutrons intact while explaining features such as the similarity of the proton and neutron.

In each case, a new picture of matter emerges, but does not replace the previous picture. We have pictures of the nature of matter at several scales and each of these is still in widespread use. While the models at each level evolve, e.g. as we go from the plum pudding model to the modern picture of the atom, we do not replace these models with what we learn as we go to the next layer of sub-structure. The reason for this is the fact that the energy and length scales are very different for each of these pictures. When dealing with a problem where the electronic structure of the atom is important, the size and energy scales involved are almost always such that the internal structure of the nucleus is irrelevant. Thus, the nucleus can be treated as a fundamental constituent, with its complicated dynamics absorbed into the static properties of the nucleus, treated as an effective fundamental degree of freedom in the problem. So while the nucleus is a complex, dense, and highly energetic system, it is treated as a static object with no internal structure or dynamics. This picture is entirely incorrect, but extremely effective, as the nuclear dynamics occur on very short time and distance scales, such that the diffuse and slow moving electrons do not see these details.

This is an incredibly important feature that has made it possible to uncover the nature of matter in a way that would have been nearly impossible if understanding the basic structure and interactions of atoms required a detailed model of the atom in terms of Z protons, N neutrons, and Z electrons (or as 3A constituent quarks and Z electrons). The separation of scale has allowed us to learn about matter one layer at a time, with the discovery and understanding of each subsequent layer providing greater insight into the simple structure seen at the larger scale. At every step, we can describe the system of interest as collection of constituents, bound by some interaction, which can be well modeled without worrying about the internal structure of these constituents. When we understand the next layer of structure, we can make a “map” from that layer back to the previous layer, e.g. build up protons and neutrons from the constituent quarks. If one then wants to deal with the larger system, in this case the nucleus, it becomes more efficient to switch to the model that neglects the sub-structure of the proton and neutron, as this provides more efficient and effective description. This also has the advantage that a simple model of the structure of the proton and neutron can explain many of their features that are important for modeling the nucleus. But one does not require a full or complete description, as the main point is to understand the reason for particular properties of the proton and neutron rather than to quantify these properties.

The next step in this process is the attempt to understand the problems in the naïve constituent quark model. These constituent quarks can be thought of as composite objects made up of nearly-massless current quarks and massless gluons bound together by the color interaction of Quantum Chromodynamics or QCD. This step is often described as being analogous to the previous advances, where what is treated as an elementary constituent at one scale is found to have substructure when probed more deeply. However, the bound states of QCD have several unusual and even unique features that make it fundamentally different.

- Neither the constituents (quarks and gluons) nor the color charge associated with their interactions appear as a relevant degree of freedom in the interactions of hadrons.
- The naïve constituents – the valence quarks – account for only a tiny fraction of the hadron mass.
- Hadrons do not have a fixed, well-defined number of quarks or gluons.

These will be discussed in the following section.

III – The “unnatural” nature of QCD

Of course, there is nothing unnatural about QCD. However, the bound states of QCD have many properties that differ from other bound states, at both a quantitative and qualitative level. As such, the simple, intuitive pictures we may have of bound states of quarks and gluons have been found to be incorrect in many ways. They provide us a simple framework that we can use to guide our thinking, which can be extremely useful in spite of the flaws in these pictures, but it is useful to consider some of the unique properties of QCD and how they impact our approach to understanding hadronic matter. In this section, I will discuss several of the unusual features of QCD.

Bound states such as nuclei, atoms, molecules and even brick walls share several characteristics, a well defined set of constituents held together by a some relatively weak binding. In this context, “weak” binding means that the total mass of the system is very close to the sum of the constituent masses, and any corrections due to the binding energy of the system are very small compared to these masses. The constituent quark model is a simplified picture of quark matter, which is similar to these other systems, and can explain many of the properties of protons, neutrons, and other baryons. While it is common to work in these simplified models, in particular by neglecting sub-structure of the constituents used in a particular model, the constituent quark model is much further from reality than in these other cases.

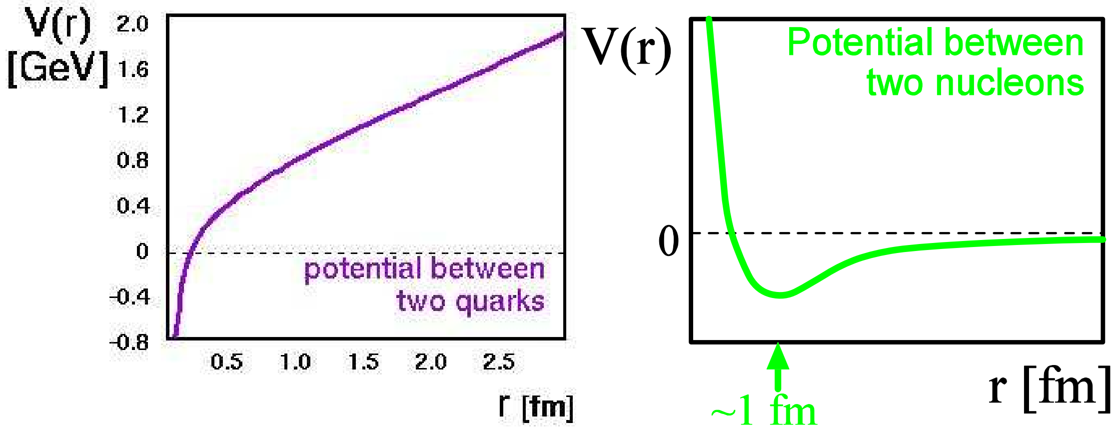


Figure 2: Illustration of the quark—quark and nucleon—nucleon potentials.

Hadrons are built from quarks and gluons, with the color interaction of QCD providing the binding that forms hadrons. However, the individual quarks and gluons are never seen in isolation; they appear only in bound states. The color charge of QCD, which drives the interactions of quarks and gluons and is an essential property in determining the nature of the spectrum of hadrons, is also hidden. The fact that quarks, gluons, and color are hidden in matter is a consequence of the nature of the strong interaction of QCD. The color interaction is extremely strong and does not decrease over distance. While gravitational and coulomb interactions are long range, the force falls as 1/r, while the force between two quarks is nearly constant, roughly 1 GeV/fm, except at very short distances. This is not only an extremely strong interaction compared to the other forces, but also compared to the masses of the quarks which are of the scale 5 MeV for the light quarks. For two quarks in this region of constant force, and thus linear potential energy, the energy required to pull the two quarks apart by just another 0.01 fm is enough to generate quark-antiquark pairs from the vacuum. Thus, even given the energy necessary to overcome the QCD interaction and pull a quark out of a hadron, quark-antiquark pairs will be produced. This allows the formation of new hadrons, e.g. with the produced quark replacing the quark taken from the initial hadron, and the anti-quark binding with the removed quark to form a meson.

So it is not just that the QCD interaction is stronger than the other forces, but also that it does not decrease with range and is strong compared to the masses of the quarks that yields behavior unlike other systems. The only way quarks can be separated is if they are bound together into systems where the QCD color charges cancel, i.e. in color neutral or color-singlet objects. Not only does this explain why individual quarks and gluons (which also carry color charge) are never observed in isolation, it also explains why we do not observe color as a relevant degree of freedom in matter, as only color-singlet objects are stable.

Finally, the fact that the energy in the QCD interactions is large compared to the constituent masses means that hadrons do not have a constant and well defined set of constituents that define the properties and account for most of the mass of the bound state. The quantum numbers of hadrons can be explained by three quarks for hadrons, or a quark-antiquark pair for mesons. However, this is simply the minimum set of constituents that can explain the spectrum of hadrons, and with three current quarks accounting for only ~10 MeV of mass, most of the energy in the system is stored in the fields and particles associated with the QCD interaction; the gluons and the “sea” of virtual quark-antiquark pairs.

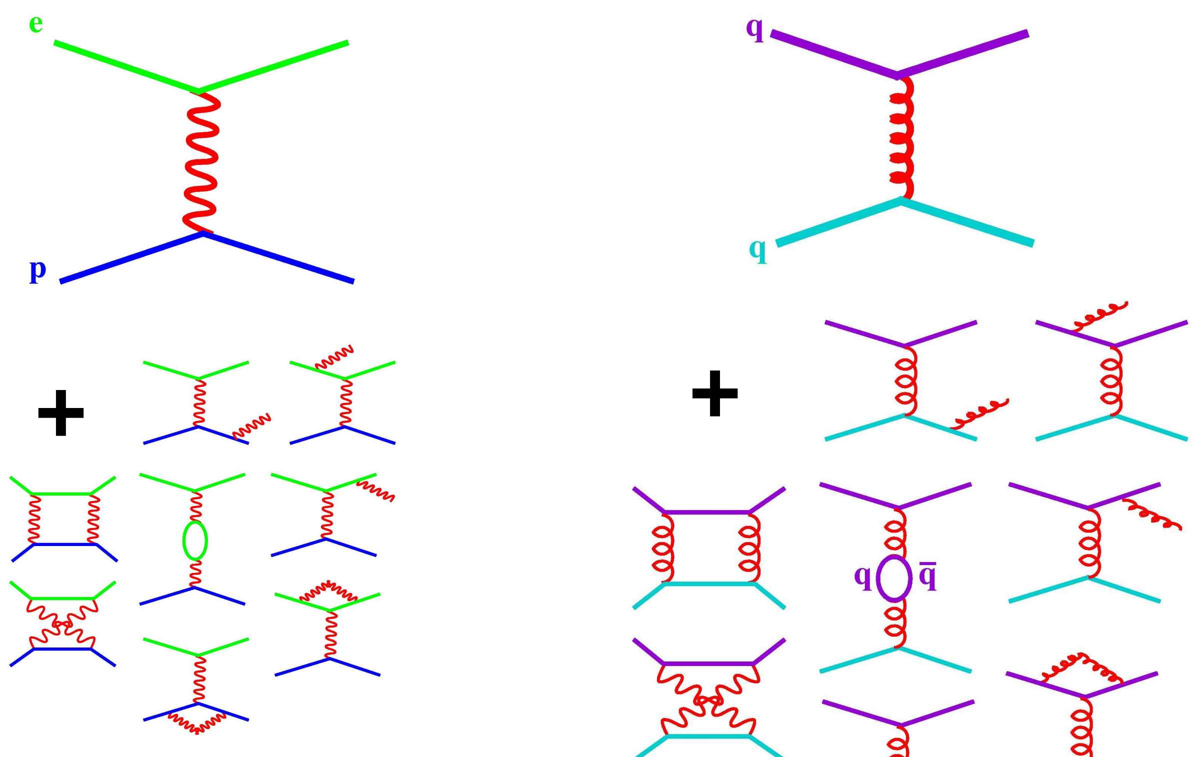
Given this, one can interpret the successes of the constituent quark models in predicting static properties of the hadrons as an indication that the constituent quarks can be thought of as being bound states made up of one valence quark with its own sea of gluons and quark-antiquark pairs. This is a simple and common picture that provides yet another layer of matter, with constituent quarks being the bound state of QCD, and hadrons as bound systems of the constituent quarks. However, lattice QCD studies of heavy quarks suggest that the gluon field does not localize around the valence quarks, but forms so-called "flux-tubes", yielding a strong gluon field in the space between two valence quarks. For light quarks, it is not clear which picture is most correct, and so it is not clear if there can be a particularly meaningful description of matter in terms of these intermediate constituent quarks. In fact, the constituent quark model is based on the fact that the quantum numbers (charge, spin) and some aspects of the spectra of baryons can be reproduced by three up or down quarks, or up, down and strange if one includes hadrons with strange quarks. These aspects can be reproduced simply by having three valence quarks; it does not require three massive quarks. The idea that the constituent quarks have a mass of ~300 MeV comes from the assumption that the three constituent quarks provide the bulk of the mass, and thus roughly one-third of the mass of the proton and neutron, which are the lightest baryons. The fact that the proton and neutron have nearly identical masses suggests that the masses of the up and down constituent quarks are nearly identical. However, this can be explained if the gluon field and quark-antiquark sea, which make up most of the mass, is unchanged when replacing one of the valence up quarks with a down quark. Thus, the characteristics of the constituent quark model can be reproduced in a complex system with a large number of light quarks and antiquarks, as long as there is a net excess of three quarks.

So while one can think of the constituent quark model as yet another layer in the structure of matter, it is not clear if this is correct. Treating protons and neutrons as fundamental particles when modeling nuclei is simply a matter of taking real bound states of matter and treating using them as the effective degrees of freedom in building up more complex states. But it is not clear that constituent quarks, as we think of them, actually exist as a real bound state of quarks and gluons, and thus this may simply be a model that explains some characteristics of matter, but which does not represent a true picture of how quarks and gluons interact. Even if it is accurate to think of constituent quarks as bound states of quarks and gluons, it is still very different from other bound states. There is not a fixed number of constituents bound together, as quark-antiquark pairs appear and disappear, leaving only the requirement of a net excess of a single quark in the case of a constituent quark, or an excess of three quarks in the case of a baryon.

IV – The role of simple models and symmetries

Given that the nature of bound states of quarks is intrinsically different than other bound states we are familiar with, it should not be surprising that QCD also poses significant difficulties in the calculation of hadrons structure and quark-gluon interactions.

One key property of QCD is that the color interaction becomes extremely weak at large energy scales, a property known as asymptotic freedom. This means that the high energy interaction of quarks and gluons is the most straightforward case for making precise calculations in QCD. When the interaction is weak (coupling constant less than one), one can make a diagrammatic expansion, as one does in QED, illustrated in Figure 3.



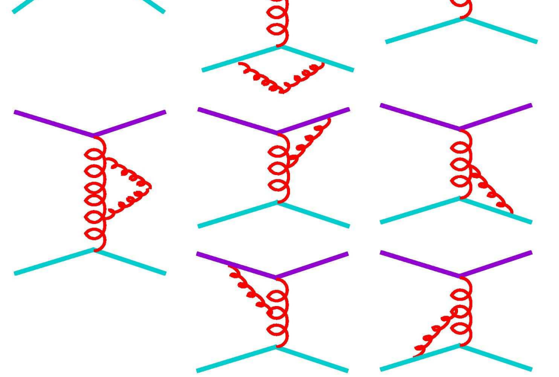


Figure 3: First and second order diagrams for QED (left) and QCD (right). The first order diagram is single photon (gluon) exchange, while higher order diagrams involve additional particle production or exchanges. Because the gluon has color, new diagrams involving gluon-gluon interactions occur in QCD. In QED, the higher order diagrams are suppressed by $\alpha_{EM} \sim 1/137$, while for QCD, they are suppressed by a factor $\alpha_{QCD} \sim 0.1$-0.3 at high energy.

While these diagrammatic expansions can be evaluated very precisely by going to high order in QED, perturbative QCD (pQCD) is significantly more difficult to use for precise calculations. First, because the gluon carries color and therefore can couple directly to other gluons, leading to additional diagrams, and second, because the strong coupling constant is much larger, yielding a significantly slower convergence for these calculations. Worse yet, in the low energy regime, the coupling constant becomes large, and the perturbative approach does not converge. It is the low energy interaction that is relevant for understanding the nature of bound states, and thus one must have a non-perturbative approach to try and understand hadrons.

From this, one can also see that the gluon plays a somewhat different and more significant role in the structure of hadrons than the photon does in the structure of atoms. The combination of the gluon self coupling and the much stronger coupling yields a much more significant virtual gluon field within hadrons. With the strong coupling of these gluons to virtual quark-antiquark pairs, the gluon field also leads to the significant contribution of the sea of quarks and antiquarks in matter. One can compare this to other bound states to see just how significant a role the glue plays in the structure of protons and neutrons. In the hydrogen atom, the binding energy is 13.6 eV, which is ~10^{-8} times the mass of the atom and ~10^{-5} times the mass of the lightest constituent, the electron. In nuclei, the forces are stronger, but the binding of ~10-20 MeV/nucleon is still only 1-2% of the ~1 GeV per nucleon. In the proton, the three constituent quarks have a mass of ~10 MeV, compared to the ~1 GeV mass, meaning that the gluon field not only yields the existence of a bound state, it actually provides 99% of the mass of the system. So not only do gluons play a significant role and the quark-antiquark sea have a significant contribution to the total quark content of the proton, but in fact it is the field rather than the valence quarks that binds the proton together that provides the bulk of the mass. This explains why the proton and neutron masses are essentially the same, unaffected by the change of an up quark into a down quark; it's not that the up and down constituent quarks are identical, it's that the valence quarks have almost no impact on the mass of the nucleon.

While the initial picture of the proton was a bound system of three massive constituent quarks, examining the nature of QCD in more detail shows that the structure is much more complicated and less static than in other bound systems, with a large number of light, virtual constituents inside of the proton, yielding a picture where one can model much of the nucleon's behavior based on the idea of a dominant contribution coming from the field that binds the three valence quarks together, with those quarks providing mainly the quantum numbers of the nucleon, as illustrated in Figure 4.

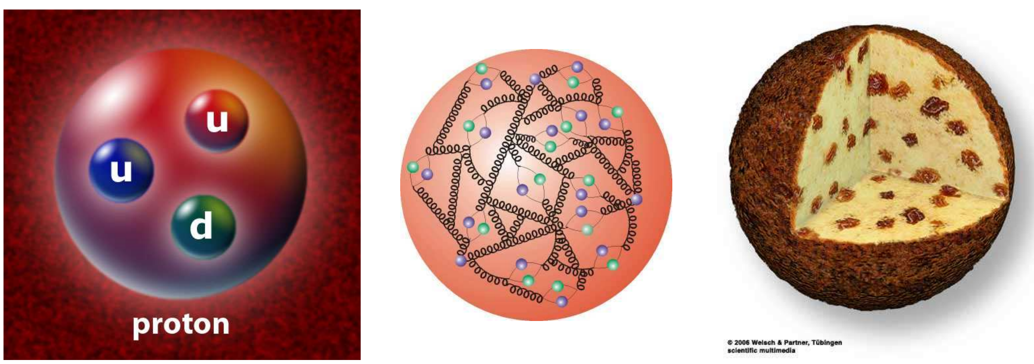


Figure 4: Evolution of hadronic models. The simple constituent quark picture (left), a picture built of many current quarks and gluons (middle), and back to a new plum pudding model, emphasizing the large role of the gluonic field relative to the quarks in the proton.

V – Nucleon Form Factors

The form factor is one of the most basic quantities that describes the structure of a composite object such as a hadron, nucleus, or atom. It is also an example of a case where examining symmetries, or more specifically symmetry-violation, can provide a great deal of insight when dealing with the structure of an extremely complicated bound states such as hadrons, where direct calculations of the bound state structure are not feasible.

In electromagnetic interactions, one can calculate the cross section for elastic scattering between two point-like objects. For scattering of a light spin-1/2 particle from a structureless spin-1/2 target, one obtains the Mott cross section:

$$\left(\frac{d\sigma}{d\Omega}\right) = \frac{\alpha^2\hbar^2c^2}{4E^2\sin^4(\theta/2)}\left(\frac{E'}{E}\right)\left\{\cos^2\left(\frac{\theta}{2}\right) - \frac{q^2}{2M^2c^2}\sin^2\left(\frac{\theta}{2}\right)\right\},$$

Where E and E' are the initial and final electron energies, θ is the electron scattering angle, M is the target mass, and q^2 is the square of the four-momentum transfer to the target: $q^2 = -Q^2 = \nu^2 - \underline{q}^2$, where ν is the energy of the virtual exchange photon and $\underline{q}$ is its momentum.

For elastic scattering from an object with extended structure, the cross section is modified. In non-relativistic scattering from a charge distribution, the elastic cross section is modified by the form factor, which accounts for the extended spatial structure of the charge distribution and is equal to the Fourier transform of the charge density $\rho(r)$.

$$f(\mathbf{Q}) = \int \rho(\mathbf{r}) e^{i\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r}$$

For scattering from a non-point spin-1/2 target, e.g. electron-proton scattering, there are two form factors: the Dirac form factor $F_1(Q^2)$ and the Pauli form factor $F_2(Q^2)$, where a point target would have $F_1(Q^2)=1$ and $F_2(Q^2)=0$. It is common to express the cross section in terms of the Sachs charge and magnetic form factors, $G_E(Q^2)= F_1(Q^2) - \tau F_2(Q^2)$, $G_M(Q^2)= F_1(Q^2) + F_2(Q^2)$, where $\tau = Q^2/(4M^2)$. In this case, the cross section can be written in a simple form:

$$d\sigma/d\Omega = \sigma_{Mott}/(1+\tau) * [\ \tau\ (G_M(Q^2))^2 + \varepsilon(G_E(Q^2))^2\ ],$$

Where σ_{Mott} is the point cross section and $\varepsilon = [1+2(1+\tau)\tan^2(\theta_e/2)]^{-1}$ is the virtual photon polarization parameter. In the non-relativistic limit, the form factors are related to the Fourier transforms of the spatial distributions of charge and magnetization in the proton, so in the limit $Q^2 \rightarrow 0$, reduce to the charge and magnetic moment of the nucleon: $G_{Ep}(0)=1$, $G_{Mp}(0)=\mu_p$, $G_{En}(0)=0$, and $G_{Mn}(0)=\mu_n$.

In a simple non-relativistic constituent quark model, one might expect the charge and magnetization distributions of the proton to simply be the sum of the charge and magnetization carried by the quarks, and thus both distributions would reflect the spatial distribution of the quarks within the proton and neutron, which would also represent the matter distribution. Early measurements of the proton form factors in the 50s and 60s suggested that they were both well approximated by a dipole form factor:

$$G_D(Q^2) = 1 / (1+Q^2/\alpha^2)\ ;\ \ \alpha^2 = 0.71\ \mathrm{GeV}^2$$

Early measurements of the neutron suggested that its magnetic form factors was also consistent with the dipole form, while the neutron charge form factor which was zero due to the cancellation of contributions from up and down quarks.

These measurements confirmed that the form factors were pretty well described by the simple non-relativistic constituent quark model mentioned above, with nearly identical up and down quark spatial distributions. The Fourier transform of the dipole Q^2 dependence is an exponential, suggesting that the distribution of the quarks was an exponential distribution: $\rho_q(r) \propto \exp(-\alpha r)$. Note that accounting for relativistic effects should smooth off the distribution so that there is no longer a non-smooth peak at r=0. Further measurements focused on extending the Q^2 range of the measurements and improving precision on the neutron form factor extraction. Figure 5 shows a summary of the results based on measurements taken through 1995.

While deviations from the dipole form were seen, they were typically around or below 10%, and it was difficult to make precise comparisons of these deviations in the different form factors. However, it was clear that the neutron magnetic form factor was not exactly zero. While $G_E^n(Q^2=0)=0$ because the neutron has no net charge, it is possible for the distribution to have regions of net negative charge and regions of net positive charge, thus yielding a non-zero charge distribution and charge form factor. The fact that G_{En} is positive at low Q^2 implies that there is a positive core of charge and a negative cloud at larger distances. This implies an excess of up quarks near the center and down quarks at greater distances. This breaking of the assumed symmetry in our starting picture shows that there is a clear and important contribution of some mechanism that breaks the symmetry between the interaction of up and down quarks within the nucleon. This in turn has implications for all of the electromagnetic form factors.

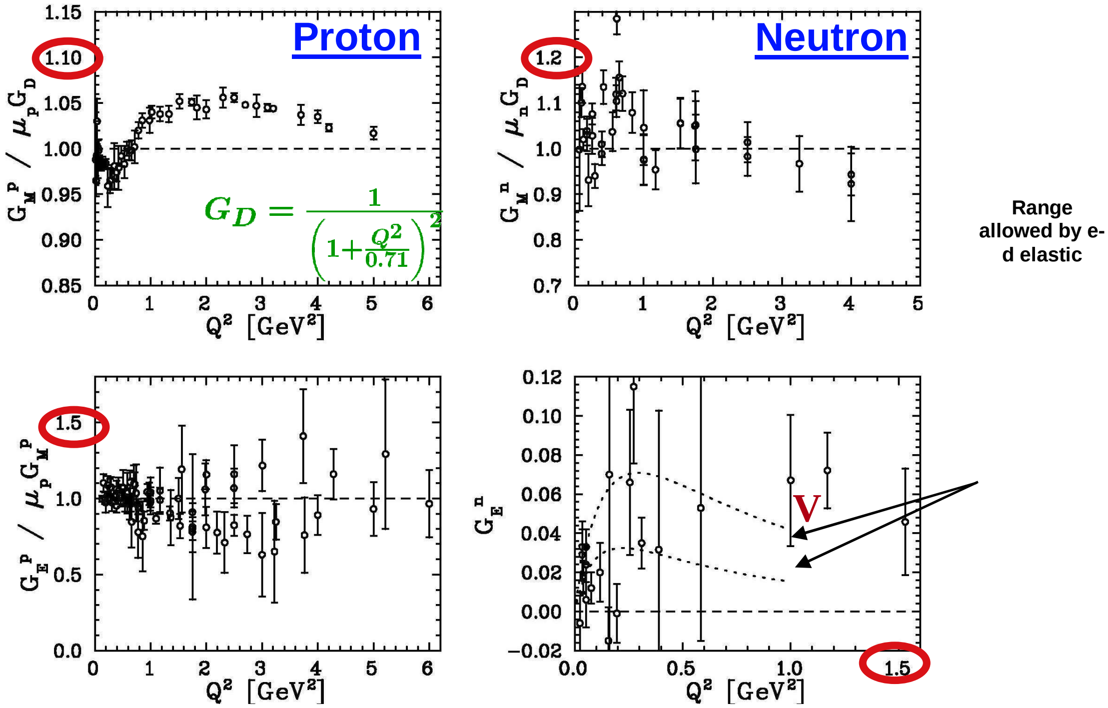
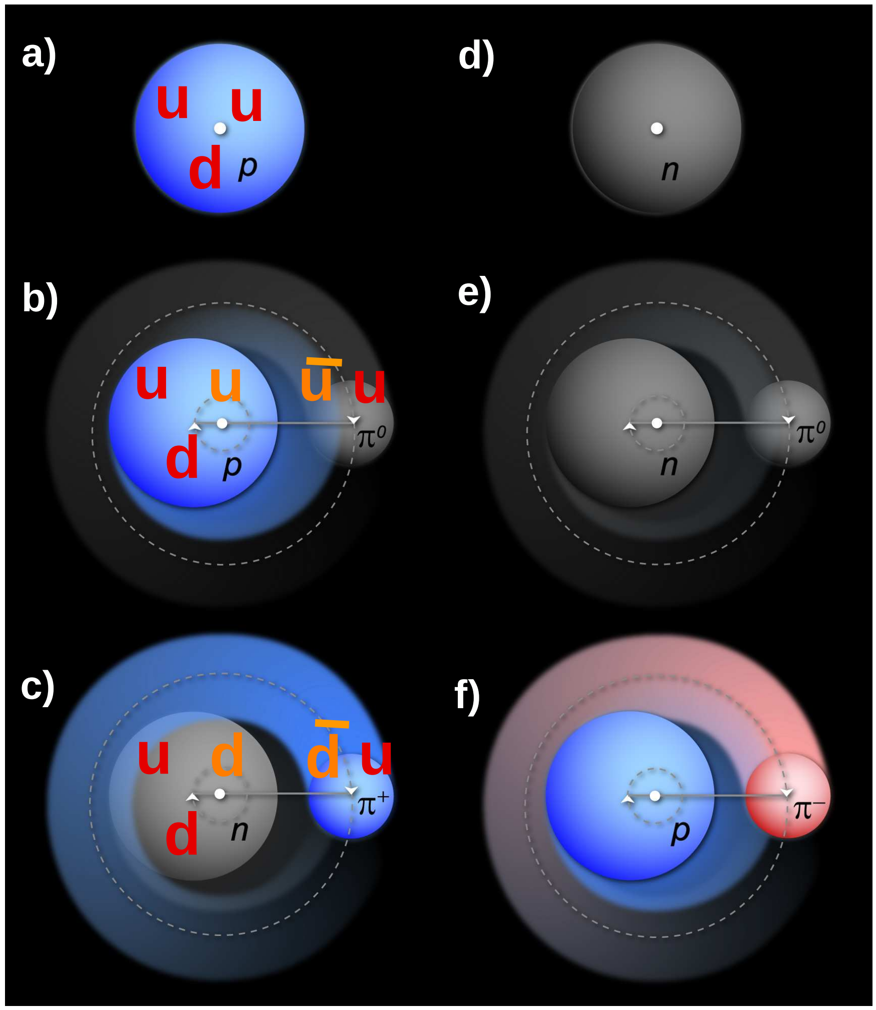


Figure 5: Proton and neutron electromagnetic form factors relative to the dipole form, based on measurements from the 1990s and earlier.

This can actually be more simply understood in terms of a hadronic picture of the nucleon. A proton (Fig 6a) can undergo a quantum fluctuation into a proton plus neutral pion (6b) or a neutron plus a positive pion (6c). The p + π^0 system will lead to a slight increase in the size of the proton's charge distribution, due to the motion of the proton around the center of mass of the system. The n + π^+ system will yield a contribution to the charge distribution at larger distances, as the light pion will be further from the center of mass. The same effect happens in the neutron, but the impact is much greater; the n + π^0 component (6e) will have little effect as both are charge neutral, while the p + π^- contribution yields a positive core of charge and a negative "pion cloud" contribution – exactly what is needed to explain the observed positive value of G_{En}.

The non-zero value of G_{En} was the first clear indication of a deviation from the simple, symmetric picture of the quark structure of the proton and neutron. After the early measurements, the focus turned to making improved measurements at higher Q^2 values, yielding information on the short-distance scale structure of the proton, and improved precision measurements on the neutron. Neutron measurements were especially problematic as there are no free neutron targets. Measurements had to be performed in quasi-elastic (single nucleon knockout) scattering from the deuteron and other light nuclei, which required significant corrections be made to extract the neutron structure. It was necessary to correct for the difference between a bound and free neutron, and to isolate the electron—neutron scattering from the electron—proton scattering. The latter requires that one either detects the struck neutron or subtracts the contribution from electron—proton quasielastic scattering. Because the proton form factors are larger than those of the neutron, e—p scattering dominates the cross section making it difficult to make precise extractions of the e—n cross section. Measurements where the neutron is detected have to deal with the difficulty of detecting a neutral particle in addition to worrying about possible reinteraction of the struck nucleon and the `spectator' proton.

Figure 6: Illustration of the impact of the pion cloud on the charge distributions of the proton and neutron; blue represents positive charge and red represents negative charge. For the proton (a), the neutral pion (b) contribution yields a small increase in the charge radius, due to the motion of the proton about the center-of-mass of the proton-pion system, while the positive pion (c) yields a long-distance contribution to the charge distribution. For the neutron (d), the p+π^- contribution yields a positive core and negative pion cloud. *[Graphic credit: Joshua Rubin, Argonne National Laboratory]*



Even in the case of the proton, there were limitations to the precision with which form factors could be extracted from cross section measurements. A Rosenbluth or Longitudinal-Transverse separation of the form factors involves measuring the cross section as a function of scattering angle at fixed Q^2 and using the simple angular dependence of the reduced cross section:

$$\sigma_R = \tau\,(G_M(Q^2))^2 + \varepsilon(\theta)\,(G_E(Q^2))^2 \;.$$

Note that at fixed Q^2, the form factors are just constant values, and ε is the only quantity that varies with scattering angle. Thus, one can fit the reduced cross section to a linear function of ε, and the intercept will yield G_M^2 and the slope will yield G_E^2. However, at very small or large Q^2 values, the factor $\tau = Q^2/(4M^2)$ will strongly suppress or enhance the contribution from G_M. So for small Q^2, one is only sensitive to G_M at very small scattering angles, where $\varepsilon(\theta) \rightarrow 0$, while at larger Q^2 values, G_E has little contribution to the cross section and so is very difficult to isolate. The problem is again worse for the neutron, as $G_{En} \ll G_{Mn}$ at all Q^2 values, making it extremely difficult to obtain information on G_{En}, even if one can isolate the e—n cross section from quasielastic scattering from a deuteron target. The uncertainties (~1%) associated with correcting for higher-order interactions (shown in Fig. 3) led to a significant limit for the precision with which one could extract the G_M at low Q^2 and G_E at high Q^2.

In spite of the limitations of the extractions based on unpolarized scattering, data covering a wide range of Q^2 values was available for all four electromagnetic form factors, as shown in Fig. 5. These measurements started around 1960, but by the mid-80s, progress had slowed significantly. New measurements were generally making only incremental progress in either the precision or Q^2 range of the data, as improvements were severely limited by the intrinsic limitations of the Rosenbluth separation method. While it had been known for some time that measurements of the spin-dependent cross section, accessible with spin-polarized beams and/or targets, could overcome these limitations, the technical difficulties in making polarized cross section measurements limited such studies, and in the mid-1990s, only a few of proof-of-principle measurements were available. However, these measurements clearly demonstrated the power of polarized scattering techniques, in particular for improved measurements of the neutron structure.

Over the last two decades, remarkable technical progress has been made in the development polarized targets, intense high-polarization electron beams, and detectors to measure nucleon polarization. These have allowed measurements utilizing polarization through the scattering of polarized electrons from polarized nucleons (beam-target asymmetry measurements), or polarized electron scattering from unpolarized nucleons, with detection of the polarization of the final-state nucleon (recoil-polarization measurements), both of which give access to the spin-dependent cross sections. The advantage of such measurements is that they have a very different sensitivity to G_E and G_M than the Rosenbluth extractions from the unpolarized cross sections. For the recoil-polarization measurements, one determines the longitudinal and transverse polarization of the final-state nucleon, P_L and P_T, which relate to the nucleon polarization in the electron nucleon scattering plane, with P_L the component along its momentum and P_T the component perpendicular to its momentum. These polarization components are both sensitive only to the ratio G_E/G_M, and by taking the ratio, the overall beam polarization and analyzing power of the detector (representing the ability of the polarimeter to measure the polarization of the incoming nucleon) cancel, significantly reducing the systematic uncertainties, and giving:

$$G_E/G_M = -(P_T/P_L)\,(E_e+E_{e'})/2M\ ,$$

where E_e and $E_{e'}$ are the electron beam energy and scattered electron energy, respectively, and M is the target nucleon mass. For polarized target measurements, one can form similar combinations of the asymmetry for scattering with parallel and perpendicular beam and target polarization.

In all cases, these measurements are sensitive to the ratio G_E/G_M. Thus, they cannot be used by themselves to extract the individual form factors. They can however be combined with the unpolarized cross section measurements to extract G_E and G_M. For the proton, this is particularly useful at large and small Q^2 values, where the unpolarized cross section is dominated by one form factor. The dominant form factor can be measured in the Rosenbluth separation and combined with polarization measurements of the ratio to make a precise extraction of the other form factor. For the neutron, this allows a dramatically improved measurement of G_{En} which is a small contribution to the cross section at essentially all Q^2 values. Figure 7 shows an updated version of the measurements of the nucleon electromagnetic form factors including a large number of measurements utilizing polarization degrees of freedom as well as some cross section ratio measurements of d(e,e'n)/d(e,e'p) for extraction of the neutron magnetic form factor (hollow squares in upper-right hand panel).

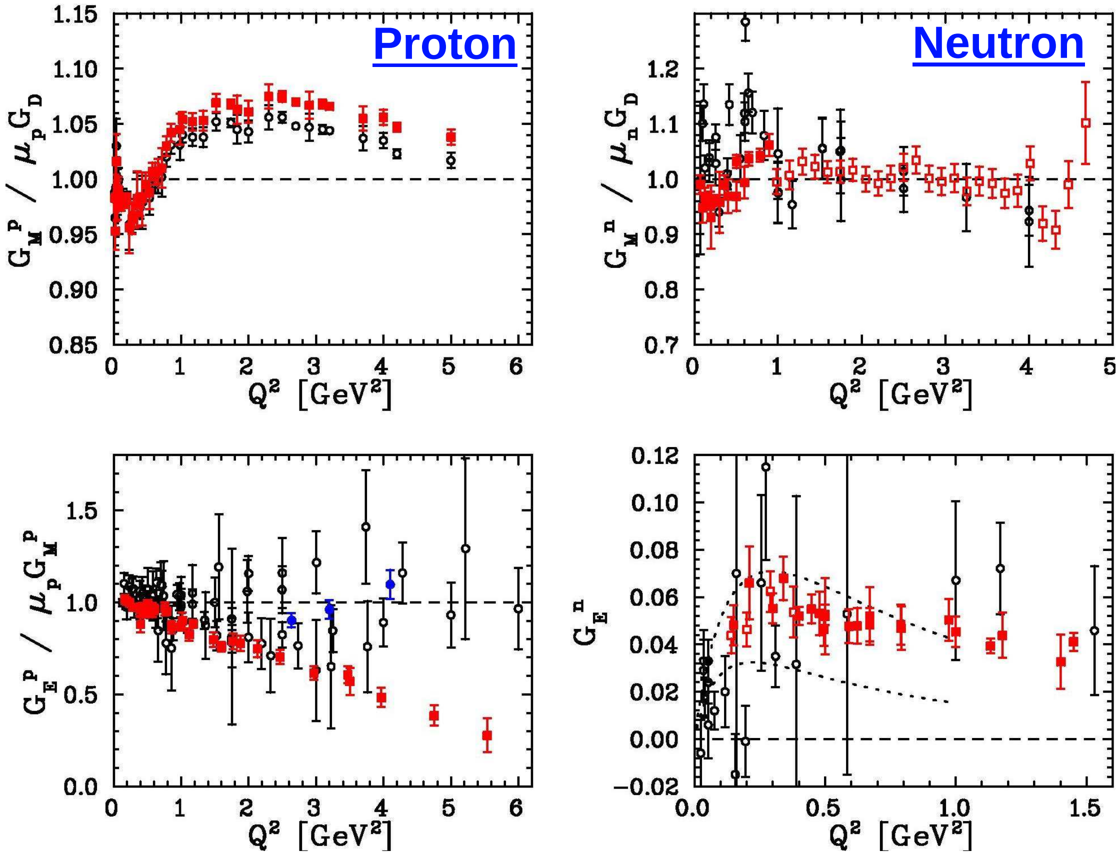


Figure 7: Proton and neutron electromagnetic form factors relative to the dipole form. More recent measurements, mainly utilizing polarization degrees of freedom, are shown in red and blue.

These new data greatly improved both the precision and Q^2 range of the measurements of the neutron form factors. For the proton, the results go beyond simply improving the precision or Q^2 range of the data. The ratio of the proton electric and magnetic form factors shows a strong, nearly linear decrease with Q^2. While the previous measurements did not precisely map out the high Q^2 behavior, the polarization measurements clearly showed a qualitatively different behavior.

One very surprising result of these polarization measurements on the proton came from the extraction of G_E at high Q^2 using recoil polarization at Jefferson Lab, which was at odds with decades of cross section measurements. The high- Q^2 polarization data indicates that the ratio of electric to magnetic form factors is constant at very low Q^2 and then falls almost linearly with Q^2, as shown in Figure 7. This contradicted the simple nonrelativistic quark models and previous Rosenbluth measurements that suggested $G_E=G_M/\mu_p$. The difference between the polarization and cross section extractions is now believed to be the result of two-photon exchange contributions. These have little impact on polarization measurements but significantly affect Rosenbluth extractions of G_E at high Q^2, making it appear as though $G_E=G_M/\mu_p$. Definitive tests of the effect on the cross sections using precise comparisons of positron and electron scattering are under way, but there is significant theoretical support and indirect experimental evidence suggesting that two photon exchange explains why the cross section measurements give incorrect values for G_E at high Q^2. Calculations and phenomenological estimates of the two-photon exchange corrections also yield a modification to the extracted value of G_M, which is the reason for the small increase in G_M.

While this decrease of $\mu_p G_E/G_M$ with Q^2 was largely unexpected, it is perhaps not so surprising that relativistic effects become important at these large momenta, and that the predictions of non-relativistic models break down. The most direct impact of the difference in the Q^2 dependence of the charge and magnetic form factors is the fact that this implies differing charge and magnetization densities. However, while there is a simple connection between the form factors and charge/magnetization densities in the non-relativistic case, relativistic corrections are model-dependent and potentially very large. Figure 8 shows an extraction of the densities for a particular prescription for the relativistic corrections.

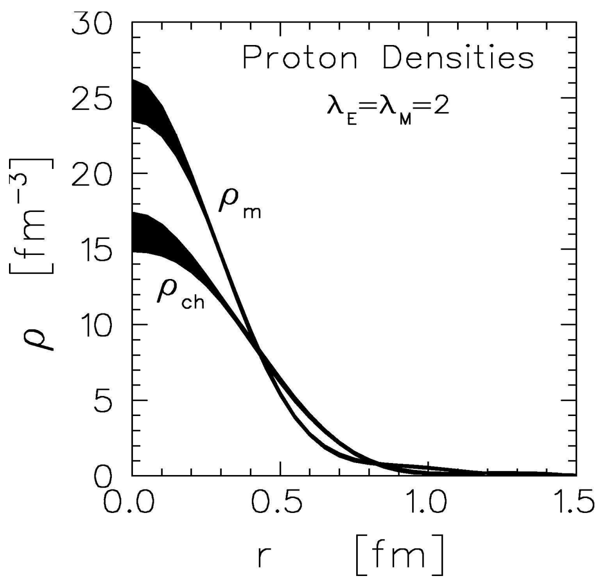


Figure 8: Extraction of the charge and magnetization density of the proton after inclusion of the recoil polarization data that shows significant difference in the Q^2 dependence of the charge and magnetic form factors. ***[Figure from J. Kelly, Phys. Rev. C 66, 065203 (2002)]***

These results led to an explosion of theoretical interest in trying to understand the proton form factors. One very significant conclusion was that the falloff of G_E is due to significant contributions from quark orbital angular momentum, which is extremely difficult to isolate in direct measurements. In addition, these results led to studies of the correlation between the spatial distribution of the quarks and the spin or momentum that they carry, showing that the spherically symmetric distribution of the proton is formed from a rich collection of complex overlapping structures, as illustrated in Figure 9.

Measurements at large momentum transfers have provided information on the details of the proton structure, and the importance of relativity and quark orbital angular momentum in the fine structure of the proton. By going to lower momentum transfer and thereby utilizing longer wavelength probes, we become sensitive to the size of the proton. This is the region where one expects to see the impact of the "pion cloud" of the proton, which arises from brief fluctuations of the proton into a bound system of a proton or neutron and pion, illustrated in Fig. 6. Because of the nucleon-pion mass difference, the nucleon will be located nearer the center of mass, and the pion will contribute at large distances. So the proton–π^0 contribution yields a small "blurring" of the intrinsic proton charge, due to the motion of the proton relative to the proton–π^0 center of mass, while the neutron–π^+ configuration contributes to the charge distribution at larger distances. This effect is most clear in the neutron electric form factor, which would be extremely small in the absence of

such effects, and where the contribution from fluctuations into bound proton–π^- yields a positive core and a negative pion cloud. A recent analysis showed suggestions of a similar structure in the proton form factors.

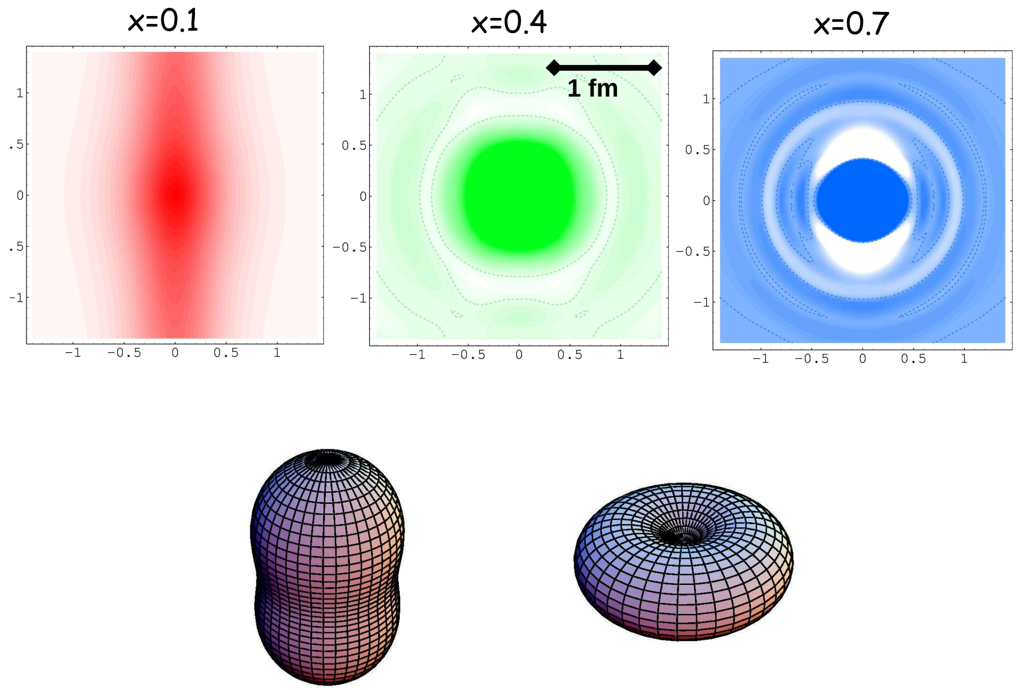


Figure 9: Model-based extractions of the proton sub-structure, accounting for correlations between quark position, momentum, or spin. The top plots show the spatial distribution of the quarks as a function of the quark momentum fraction (x). The bottom figures show the spatial distributions for quarks whose spin is aligned (left) or anti-aligned (right) to the proton's spin. In both cases, while the distributions for quarks of specific momentum or spin can have complicated spatial structure, the overall proton structure is spherically symmetric. ***[Figures from A.Belitsky, X.Ji, F.Yuan, PRD69:074014 (2004); G.Miller, PRC 68:022201 (2003)]***

With precise polarization transfer measurements in the low-Q^2 region, it is possible to better examine the form factors for indications of structure related to the pion cloud, and to improve extractions of the proton charge and magnetic radii. A new high-precision experiment of this work was performed in Hall A at JLab, and achieved ~1% total uncertainties on $\mu_p G_E = G_M$ for Q^2 from 0.3 – 0.7 GeV^2, as seen in Fig. 10. This should be compared to typical uncertainties of 3–5% from the best previous cross section measurements, and ≥2% for the previous low-Q^2 polarization measurements. More recently, a large set of high-precision cross section measurements from Mainz was analyzed to yield a global fit of the form factors in the low Q^2 region. However, the cross section measurements are still sensitive to two-photon exchange corrections, and while they do not have the same dramatic impact as they do for the high-Q^2 data, they are very important when making high precision measurements.

These new data significantly improve the precision of the extracted form factors at low Q^2. They also provide us some direct information on the proton structure. While the high-Q^2 measurements suggested that the form factor ratio fell linearly at high Q^2, they also suggested that the ratio became one at a small but finite Q^2 value (0.2-0.3 GeV^2). From non-relativistic constituent quark models, one would expect $G_E = G_M / \mu_p$, so this behavior

suggested that a non-relativistic picture, with identical spatial distributions of charge, magnetization, and matter (all arising from the quark distributions) was valid up to some scale above which relativistic corrections become important. This new data suggests that the deviations from the non-relativistic picture are important even at low Q^2, and may even extend down to $Q^2=0$. It was often thought that at low Q^2, the non-relativistic model would be a good approximation, and thus it came as a surprise that the form factors deviate at low Q^2. However, Q^2 measures the scale at which the system is probed, not the energy scale of the constituents being probed. If relativistic effects in the proton yield modification to the overall spatial distributions, then these effects can appear at all scales in the proton's structure.

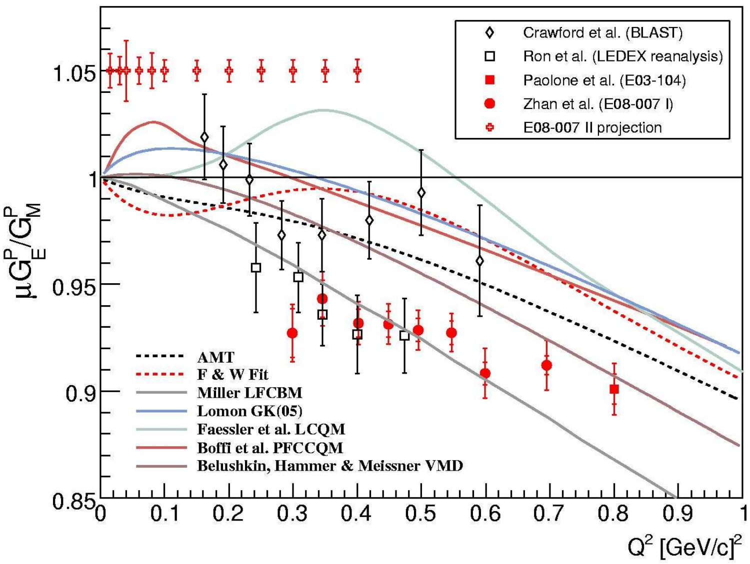


Figure 10: Low-Q^2 polarization measurements of G_E/G_M from the completed recoil polarization measurements (solid red circles) and planned polarized target measurements (hollow diamonds) compared to a set of predictions and earlier fits.

These data also show a generally smooth Q^2 dependence, without the suggestions of low Q^2 structure suggested by the BLAST data or the global analysis of Freidrich and Waltcher ("F&W Fit" in Fig. 10) which were interpreted as a possible signature of pion cloud effects that are important at low Q^2. Of course, the measurement of the form factor ratio is only sensitive to the difference between the pion impact on the charge and magnetization distributions of the proton. The impact on the charge distribution is illustrated in Fig. 6, and is relatively straightforward to visualize. Because the magnetic moment of the pion is zero, the impact on the magnetization distribution will not be as straightforward, arising through more complicated effects such as the motion of the charged pions yielding magnetic currents.

In addition, these low Q^2 measurements isolate the large-scale structure of the proton and in the limit $Q^2 \rightarrow 0$, can be used to extract the charge and magnetization radii of the proton. Non-relativistically, the form factor represents the Fourier transform of the charge density. Using this fact and expanding the form factor in powers of Q^2, we obtain $G_E(Q^2) \approx 1 - \langle r^2 \rangle Q^2/6$, and see that the RMS radius is directly related to the slope of the charge form factor at $Q^2=0$. In reality, relativistic corrections break down this relation and so the slope of the form factor at $Q^2=0$ does not yield the true rest-frame RMS charge distribution of the proton, as there are small but non-trivial model-dependent corrections. However, the non-relativistic definition is the common definition used in extracting the proton charge radius. This is of particular interest at the moment, due to recent measurements of the charge radius from muonic hydrogen.

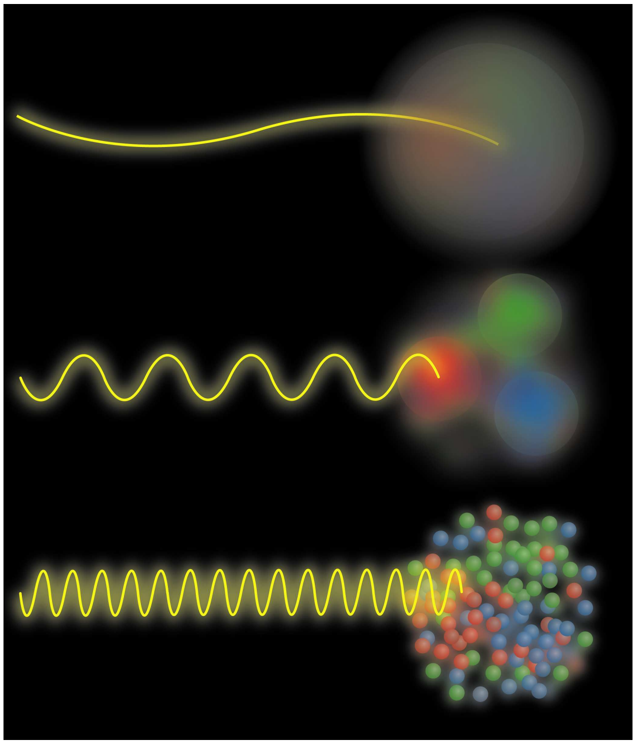


Figure 11: Illustration of the electromagnetic interaction with the proton when varying the momentum transfer and thus the scale of the probe. At low Q^2, the probe is not sensitive to the quark sub-structure of the proton, and yield information on the overall size and long-scale spatial distribution. ***[Graphic credit: Joshua Rubin, Argonne National Laboratory]***

Figure 12 (left panel) illustrates how the energy levels in hydrogen (or muonic hydrogen) can be used to extract the proton radius. The red curve illustrates the coulomb potential for the electron in a hydrogen atom for the case of a point-like proton. For a finite-sized proton, the potential is screened when the electron is inside of the proton's charge distribution, and rather than the 1/r potential, there is a cutoff in the small distance behavior, shown as the blue curve at small distances. This leads to a shift in the lowest lying energy level for the electron (as the s-wave distribution spends the most time inside of the proton), and so measuring this level (or the transition to the higher p state) is sensitive to the proton radius. If the electron is replaced with a muon, its greater mass means that it spends much more time inside of the proton, and the shift is much greater.

Figure 12 (right panel) shows the result of several more-recent extractions of the proton charge radius. The results labeled "Sick", "Bernauer, et al.", and "This work" come from electron scattering analyses, based on the form factor slope at $Q^2=0$ from a fit to low Q^2 form factor data. The CODATA result comes from a global analysis of many atomic physics measurements, with the proton radius extraction being mainly sensitive to the Lamb shift measurements in hydrogen, as discussed above. The final result, "Pohl, et al.", is the result of the muonic hydrogen Lamb shift measurement, with dramatically smaller uncertainties than the atomic hydrogen measurement due to its much great sensitivity to the proton radius. The red dotted lines indicate the average of the four measurements based on the electron—proton interaction. The consistency of these results, including both electron scattering and Lamb shift measurements is important since there are very different experimental and systematic effects in the extraction of the radius in these techniques. The electron scattering measurements require precise data at low Q^2, good knowledge of the absolute normalization of the scattering cross sections, and careful correction for higher-order effects, both standard radiative corrections and two-photon exchange contributions. The Lamb shift measurements require very large corrections for any other effect which modifies the energy level splitting between the s and p states, as well as precise extractions of other quantities (e.g. the Rydberg constant). While these are large corrections, they should be reliably calculable in QED, and thus allow for an extraction of the proton radius with relatively small uncertainties.

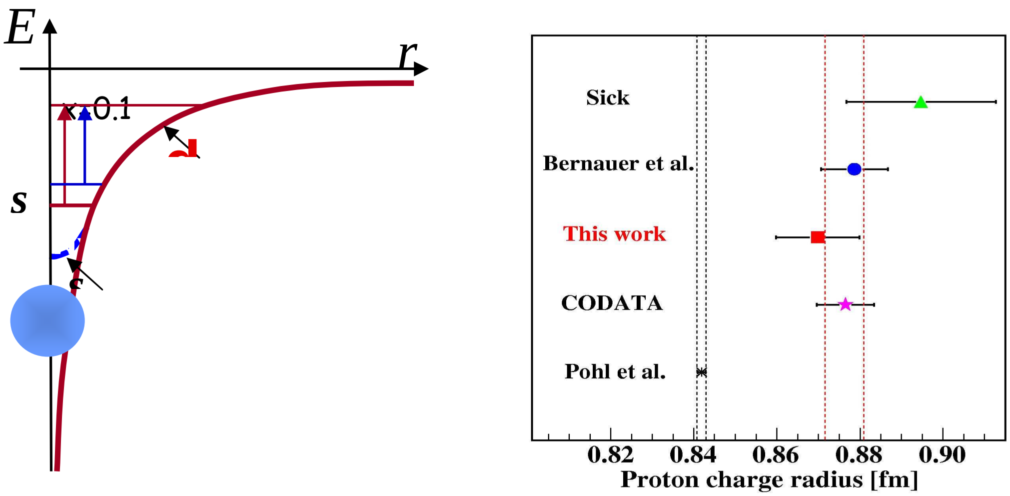


Figure 12: Left: Illustration of the impact of the finite size of the proton on the hydrogen energy levels. Right: Comparison of extractions of the proton radius from muonic hydrogen (Pohl, et al.), and various results based on the electron—proton interaction. ***[Right panel taken from X. Zhan, et al., arXiv:1002.0318 (2011)]***

The large discrepancy between the e—p and μ—p measurements suggests two possibilities: either something is wrong or missing in the extraction of the radius from the muonic hydrogen (experimental effects or corrections to the energy levels), or there is some aspect of the interaction that is different between electron and muon interactions. As yet, there is no clear indication as to the source of the discrepancy, although it is an area of great activity.

Finally, these high-precision low-Q^2 data can also help provide another test of the assumptions that go into the simplified model of the form factors. In addition to looking at the difference between the up and down quark distributions though the neutron electric form factor, we can also look for the contribution of strangeness to the proton and neutron's distributions. Because the nucleon has no net strangeness, the total strange and anti-strange quark contributions, as does the charge from the up and down quarks in the neutron electric form factor. However, if there is a difference in the spatial distribution of the strange quarks and antiquarks, e.g. due to a "Kaon cloud" contribution, then there can be non-zero strange quark contribution to the nucleon's electric form factor. In this case, one needs three observables in order to separate the up, down, and strange quark contributions to the form factor. If we assume charge symmetry, i.e. the distribution of up quarks in the proton is the same as the distribution of down quarks in the neutron, it is straightforward to expand the proton and neutron electric form factors in terms of the up, down, and strange quark contributions in the proton, as they are simply a charge-squared weighted sum of the individual quark distributions. One needs a third observable which yields a different relative weighting of these contributions. Elastic scattering via the weak interaction provides just such an observable, as the interaction is mediated by the exchange of a W or Z boson, and samples the quark distributions weighted by their weak charge. This can be achieved through neutral current neutrino scattering, but it is difficult to precisely isolate the elastic scattering because it is not possible to have a well defined initial beam energy and it is not possible to detect the scattered neutrino.

While neutrino scattering measurements are difficult, it is also possible to measure this in electron scattering. While the electron-proton interaction is dominantly electromagnetic, mediated by the exchange of a virtual photon, it can also occur via the exchange of the neutral Z boson. This interaction is much weaker, and so has a negligible contribution to the overall cross section. Because the weak interaction violates parity, one can look for the parity-violating contribution to the cross section (related to the beam spin-dependent cross section when scattering from an *unpolarized* target). This occurs through the interference of the photon exchange and the Z-boson exchange diagrams, and is generally at the 10^{-6} to 10^{-4} level in the cross section; extremely small but still possible to isolate because it is the dominant contribution in the spin-dependent cross section when scattering from an unpolarized target.

Several measurements have been performed at BATES, Jefferson Lab, and Mainz, which have extracted the parity-violating contribution to the elastic electron—proton scattering cross section. By combining these with the proton and neutron elastic electromagnetic form factors, we can attempt to isolate the strange quark contribution. One complication is caused by the fact that the strange quark contribution to the enters via a combination of its contribution to G_E and G_M, just as the unpolarized cross section is a combination of the proton electric and magnetic form factors. Therefore, data at multiple angles and Q^2 values is required to isolate the strange quark contributions. In addition, data on the deuteron and ^{4}He can improve the separation by constraining corrections to the parity-violating asymmetry or by providing different sensitivity to the two form factors, e.g. the asymmetry for elastic scattering from ^{4}He is sensitive only to the *strange* quark contribution to the electric form factor. Figure 13 shows the status of these measurements. Over a range in Q^2, constraints on a linear combination of the strange contribution to G_E and G_M (left panel) show that the overall strangeness contribution is small, but perhaps non-zero at the larger Q^2 values, although a high-precision measurement at Q^2=0.62 GeV^2, in the region where the indication of non-zero strangeness is the largest. The right panel shows the constraints on the strangeness contributions to G_E and G_M based on the combined results from measurements taken near Q^2 = 0.1 GeV^2. At this low Q^2 values, there is no indication of strange quark contributions.

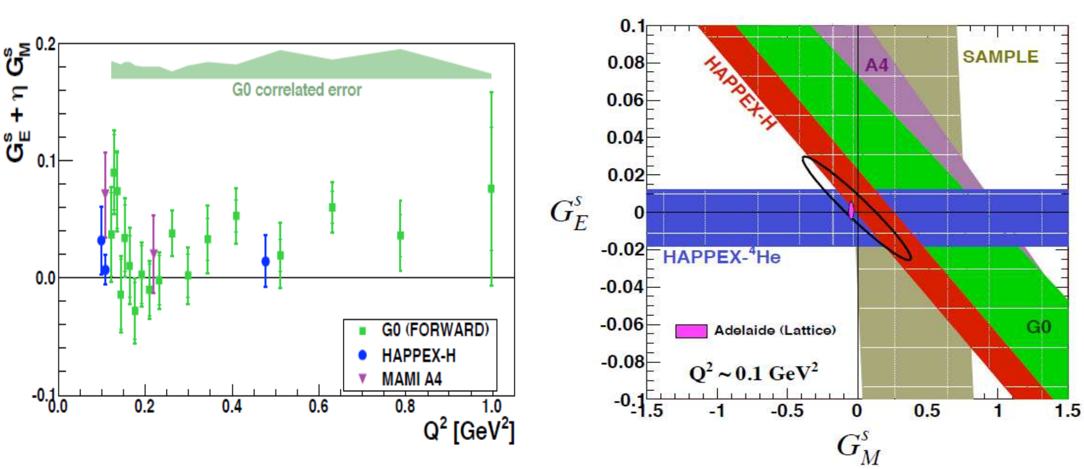


Figure 13: The strange quark contributions to the proton electromagnetic form factor. Left panel shows the linear combination of the electric and magnetic form factor contributions that is probed in the forward angle experiments ($\eta = \tau\, G_{Ep} / \varepsilon\, G_{Mp}$). Right panel shows the combined limits on both contributions for the data near Q^2 = 0.1 GeV^2.
[Figures from K. Paschke, et al., J. Phys. Conf. Ser. 299 012003 (2011).]

V – Conclusions

There has been dramatic progress in the study of proton and neutron electromagnetic form factors over the last decade, due largely to the advent of high current, highly polarized electron beam facilities coupled with high figure-of-merit polarized targets and recoil polarimeters. Together, these have allowed us to take advantage of the unique sensitivity that polarization degrees of freedom have in uncovering the nucleon form factors, and thus the spatial distribution of the change and magnetization of the form factors.

By probing the spatial structure, we can test simple models of nucleon structure which for the most part include maximal symmetry in the proton wavefunction. Early measurements showed a difference between the up and down quark distributions based on the incomplete cancellation which yielded a non-zero value for the neutron electric form factor. Now, we can more precisely probe the assumed symmetries by looking for differences between the distribution of charge and magnetization in the proton, comparisons of the proton and neutron distributions of magnetizations, and by including parity-violating electron scattering measurements which allow us to more fully separate the up, down, and strange quark distributions.

Beyond testing these simple models, these new data have driven significant progress in detailed modeling of the proton structure. In addition, these measurements have fed into the new study of generalized parton distributions (GPDs), which require significant modeling, but which relate to correlations between the spin, space, and momentum of the quarks, required, e.g. to go beyond symmetric spatial distributions and one-dimensional momentum distributions into more complicated quantities. This allows for study of quantities that go beyond what can be probed using only elastic scattering, e.g. orbital angular momentum of the quarks in the proton.

For those interested in a similar (and highly informal) discussion/review of QCD as it applies to nuclei, I recommend "Hadrons in the nuclear medium: Quarks, nucleons, or a bit of both?" [J. Arrington, arXiv:nucl-ex/0602007] – a set of lecture notes on QCD in nuclei from the HUGS lectures at Jefferson Lab in 2005.

For further details on the form factor program, as well as the broader hadron structure program, I recommend "New Insights into the Structure of Matter: The First Decade of Science at Jefferson Lab" [J.Phys.Conf.Ser.299 (2011)]. The 2nd and 3rd contributions cover the form factor program and the parity-violating electron scattering presented here in more detail, but still providing a broad overview of recent progress in the field. It also includes a wide range of contributions.

This work was supported by the U. S. Department of Energy, Office of Nuclear Physics, under Contract DE-AC02-06CH11357. My thanks to Josh Rubin and Ushma Kriplani for their input on the writeup.