SLAC-PUB-10483
YITP-SB-03-22
MIT-CTP-3379

Fractional Charge Definitions and Conditions

Alfred Scharff Goldhaber*
C.N.Yang Institute for Theoretical Physics, State University of New York, Stony Brook, NY 11794-3840 USA†
Stanford Linear Accelerator Center, Menlo Park, CA 94025
Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

Fractional charge is known through theoretical and experimental discoveries of isolable objects carrying fractions of familiar charge units – electric charge Q, spin S, and the difference of baryon and lepton numbers B-L. With a few simple assumptions all these effects may be described using a generalized version of *charge renormalization* for locally conserved charges, in which medium correlations yield familiar adiabatic, continuous renormalization, or sometimes nonadiabatic, discrete renormalization. Fractional charges may be carried by *fundamental particles* or *fundamental solitons*. Either picture works for the simplest fractional-quantum-Hall-effect quasiholes, though the particle description is far more general. The only known fundamental solitons in three or fewer space dimensions d are the kink ($d = 1$), the vortex ($d = 2$), and the magnetic monopole ($d = 3$). Further, for a charge not intrinsically coupled to the topological charge of a soliton, only the kink and the monopole may carry fractional values. The same reasoning enforces fractional values of B-L for electrically charged elementary particles. [J. Math. Phys. **44**, 3607 (2003), corrected and sharpened]

1. INTRODUCTION

The ascendancy of fundamental particles in thinking about microscopic physics began with atoms and molecules, followed by electrons, photons, and nuclei, then nucleons and neutrinos, quarks, gluons, and W and Z bosons. The pattern of a hierarchy of length scales, with the particles of one scale being compounds of new fundamental particles at a shorter scale, has replayed itself several times over. There is no direct evidence indicating whether this pattern terminates eventually. However, string theory and its developments raise the prospect that at sufficiently short scales the fundamental objects are not particles, but rather extended entities, so that the pattern might indeed come to an end.

A reason to worry about the universal validity of the particle description at currently accessible scales has come from theoretical and experimental discoveries of fractional charge. If at the beginning of microscopic physics all kinds of different charges had been observed, with no rational relation among them, progress in understanding would have been impeded seriously indeed. We know that did not happen, but in principle the recent discoveries might herald an era where precisely such chaos in the pattern of charges could emerge. The purpose of this work is to present some simple definitions and theorems, based on minimal assumptions, which imply that fractional charge fits into the familiar framework of charge renormalization, and consequently is so tightly constrained that any threat of 'charge chaos' is precluded.

In the standard model of strong and electroweak interactions there are three isolable, quantized charges observable in vacuum at zero temperature that may be locally conserved, electric charge Q, measured in units of the electron charge e, spin $\vec{S}$, whose projection onto a fixed axis has integer or half-integer values in units of $\hbar$, and the difference between baryon and lepton charge B-L, usually assumed to have the value 1 for a proton or neutron, and -1 for an electron or neutrino. Despite an exponentially small effect associated with electroweak instantons, which 't Hooft [1] recognized would produce baryon decay through the Adler-Bell-Jackiw chiral anomaly [2], and possible additional interactions associated with very high energy scales, B+L also may be treated as conserved in many contexts. If electroweak effects and the difference of light quark masses may be ignored, so that attention is focused exclusively on the strong interactions, then also isospin $\vec{I}$ may be treated as conserved. Even if electromagnetic interactions are included, still I_3, as well as flavor charges of higher generations of quarks – strangeness, charm, bottom, and top – are conserved.

Thus, many isolable, quantized charges observable in the laboratory are exactly or at least quite accurately conserved. It should be noted that besides these charges there are the continuous (i.e., nonquantized), locally conserved charges corresponding to energy and momentum, which precisely because they have a continuum of allowed values need not concern us further here.

At this point it may be worthwhile to discuss a bit more what is meant by the concept 'charge'. Of course, the prototype example is electric charge, whose conservation follows from the Maxwell equations. Already in classical physics it was understood that the existence of a conserved charge could be deduced from a symmetry of the dynamics. In quantum physics, this is expressed in terms of a unitary (or in the case of time reversal symmetry, antiunitary) operator which commutes with the Hamiltonian. If that symmetry be continuous, then the generators of the symmetry must be self-adjoint operators. In certain cases, such as the generators of the rotation group in three space dimensions, which are identified as the angular momentum or spin of a system under study, the commutation relations among the generators lead directly to the quantization of the allowed values of the charge. However, for electric charge such a deduc-

*goldhab@insti.physics.sunysb.edu
†permanent address

Work supported by Department of Energy contract DE-AC03-76SF00515.

tion has not yet been possible, and for B-L there is not even a known framework to discuss the symmetry beyond the statement that phenomenologically established couplings conserve the charge, i.e., the phenomenological Lagrangian is invariant under the unitary transformation generated by the operator B-L. Thus, in terms of current knowledge, symmetries imply conservation and sometimes even quantization of observable charges, but there is at this point no assurance that all cases of apparently conserved and quantized charges are consequent to symmetries which can be identified in any way other than by recognizing that the charges seem to be conserved.

In relativistic physics, a conserved charge whose local density is defined must be locally conserved. That is, if the charge in some volume changes, the immediately surrounding volume must experience an equal and opposite change in charge, or expressed differently, the rate of change of the charge in a volume is equal and opposite to the net current flowing out of the volume. The reason is easily found by assuming the contradiction to this assertion: If in one inertial frame a conserved charge disappeared in one place and reappeared instantly at a distant place, then in different frames of reference boosted by velocity shifts from the original frame the total charge at certain times either would vanish or would be double the original value, evidently violating charge conservation.

How may charges be measured or observed? Again, the prototype is electric charge, which may be measured through the electromagnetic interaction, either by determining the influence on a particle of a specified electromagnetic field (thus measuring the Lorentz-force charge), or by using test particles with known charges to measure the electric flux coming out of the particle (thus measuring the Gauss-law charge). In the quantum context, these two may be called the Aharonov-Bohm charge and the local charge, respectively [3]. For other types of charge (i.e., not coupled to gauge fields), one must use more indirect methods, such as counting different spin states of an object. To help understand the conceptual structure, one may introduce hypothetical abelian gauge fields weakly coupled to any conserved, localizable charge one wishes to measure.

Are there any circumstances in which fractional values of charges may be found, and if so, what are the precise conditions for this to occur? Let us turn to that issue.

2. DEFINITIONS AND THEOREMS

Def. 1 (Fundamental particle): A particle is called *fundamental* if it has no discernible internal structure. This means that at the shortest distance scales such particles correspond to local fields appearing in the Lagrangian, with only perturbative interactions whose effects can be estimated accurately.

Remarks: Under this criterion, the established fundamental particles are the photon, the gluon, the W^+ and Z bosons, along with two leptons, the electron and the neutrino, and two quarks, the up and the down quark, plus their antiparticles (the photon and Z certainly, and the neutrino possibly, are their own antiparticles). In the standard model, copies of the leptons and the quarks appear again in two more families.

While quarks and gluons are not isolable in vacuum, the fact that quantum chromodynamics (QCD) exhibits asymptotic freedom [4], meaning weak coupling at short distance scales, implies that at those scales they may be detected and (for quarks) their electric charge measured. As the term "fundamental" suggests, the charges of all known particles in vacuum can be constructed from those of the fundamental particles. In addition to these known fundamental particles there may be other, more massive ones, and also fundamental solitons as discussed below.

Def. 2 (Fundamental soliton): A *fundamental soliton* may be described (with well-controlled quantum corrections) as a classical field configuration with localized energy, where the long-range field pattern implies that no process with finite action could dissolve it.

Description of fundamental solitons: The *kink* in $d = 1$ may be described by a classical scalar field which has equal potential energy density minima for two or more distinct values of the field. Thus the field can go asymptotically to one value as $x \to +\infty$ and another value as $x \to -\infty$. No finite-action process could destroy this structure, which nevertheless possesses a finite, localized energy in the region between the two asymptotic zones. Consequently the kink has a conserved *topological charge*, and so there would be no contradiction if it carried fractional values of other charges.

In $d = 2$, a *vortex* can be described by a complex scalar field which rotates in phase by $2n\pi$ at large distances from its center as the radial direction rotates by 2π, while the field magnitude approaches a fixed value with increasing radius, again to minimize the field's potential energy density. In the Abelian Higgs model for such a vortex, gradient energy is kept finite through coupling of the scalar field to a gauge field, which makes the covariant azimuthal gradient negligible at large distances, and implies a magnetic flux stored near the center. There is an alternate description of the asymptotic fields, in which the scalar field goes to a fixed constant and the gauge field (pure gauge) corresponds to an Aharonov-Bohm phase factor $e^{2\pi i nq/Q}$ relating the phase of a charged particle wave function at, e.g., azimuthal angles $\phi = 0$ and $\phi = 2\pi$, where Q is the charge of the scalar field and q is the charge of a particle which experiences a nontrivial Aharonov-Bohm effect upon diffraction around the flux. Again, no finite-action process could destroy the vortex.

In $d = 3$, a configuration again containing a gauge field along with one or more other classical fields generates a long-range *magnetic monopole* field, which cannot be destroyed with finite action, regardless of the precise details of the monopole interior.

Def. 3 (Elementary particle): In a given medium, an isolable particle is called *elementary* if it is fundamental or if any fundamental constituents could not be isolated in ordinary vacuum. If the particle carries conserved charges, it can be destroyed only by processes conserving all those charges, e.g., annihilation with its antiparticle.

Remarks: Because quarks and gluons cannot be isolated, they are represented in the list of elementary particles by composite objects, of which the lightest are the (B=0) π mesons and the (B=1) nucleons (proton and neutron). It is significant that protons, and hadrons in general, exhibit strong short-range interactions, so that they really are isolable only at long distance scales. Still, in nuclear matter it possible to make an accurate description of the dynamics including quasiparticles with the same charges as neutrons and protons, suitably redefined so that the "nucleon quasiparticles" interact weakly. It is a useful perspective to consider these quasiparticles as nucleons whose effective interactions are renormalized by the strong, short-range interactions of the nuclear medium: Each quasinucleon may be viewed as having a nucleon kernel surrounded by a cloud of medium polarization. A very similar approach has been most successful in describing electron quasiparticles in many different condensed-matter systems. Thus, while the definition of an elementary particle may be medium-dependent, there often is a simple correspondence between sets of elementary particles in different media.

Because all elementary particles either are fundamental or have charges which may be constructed from those of the fundamental particles and solitons, these objects may be considered the building blocks for everything else.

Def. 4 (Fractional charge): An object carrying only a portion of the charges of finite combinations of elementary particles may be said to carry *fractional charge*.

Remarks: In principle, such a particle might have, for example, the same spin as the electron, but an electric charge which is an irrational fraction of e. There is no known instance like this for particles in vacuum, but in any insulating medium exactly such a phenomenon is found, and in the accepted interpretation this fractional value is treated as a *renormalization* of the electron charge from its value in vacuum.

By this definition, the electric charges of quarks are examples of fractional charge, but the fractional values arise trivially, because these fundamental particles carry a smaller unit of charge than any (isolable) elementary particle.

Some thought experiments may illuminate the definition. First, imagine a massive particle such as a proton slowly entering an insulating medium. As it enters, its local charge is reduced and, by the time it has penetrated far inside, the extra charge is found on the surface of the insulator. Thus, total charge is locally conserved throughout the process, but it ends up fractionated between the charge localized on the particle and the charge on the surface. A second experiment, even in vacuum, invokes a slow increase from zero in the value of α, the electromagnetic coupling. As this occurs, the electric field measurable at some distance from a charged particle increases in strength, but not quite proportionally to α, because vacuum polarization increasingly screens the field. Again, charge is locally conserved, because as the coupling increases there is an outward flow of current, with the current density proportional to the electric field at each point.

As mentioned already, the charge which is fractionated is the local charge, while the AB charge remains invariant. This is illustrated by another thought experiment. Imagine a capacitor stuffed with dielectric and set at voltage V. Then an electron of charge $-e$ passing through the capacitor will acquire from the electric field a net energy $-eV$, regardless of the magnitude of the dielectric response. Of course, the presence of the dielectric implies an increase in the amount of surface charge on the capacitor plates to achieve the same V as for the plates in vacuum, but once this V is established the effect on the electron is not further modified by the dielectric.

Def. 5 (Breakup): On passing from one medium to another, a particle may undergo *charge breakup*, meaning that on the other side there are several particles instead of one, each with only a portion of the set of charges carried by the one particle in the original medium.

Remarks: Breakup evidently is an intrinsically nonadiabatic process, as the (integer) number of mobile degrees of freedom changes discontinuously. Note that charge breakup may occur when, for example, a fast electron enters a conventional insulator, knocking loose a number of electrons each of which penetrates far into the medium. As a result, very little charge may be left on the surface. Nevertheless, there is a big distinction between such a case and one where the surface simply cannot take up charge at all, as occurs when an electron enters from above a two-dimensional layer exhibiting the fractional quantum Hall effect.

Now we are ready for the first theorem.

Theorem 1 (Conservation of particles with fractional charge): An isolable particle that has part of the conserved charge(s) of previously identified elementary particles in its medium must itself be an elementary or fundamental particle (or soliton) of that medium. (As indicated above, a particle may carry fractional charge with respect to particles in *different* media, as a consequence of medium-dependent charge renormalization.)

Proof: This statement follows directly from the definition of an elementary particle, because the conserved fractional charge(s) cannot be reproduced by any finite assembly of particles carrying integer values of the same charge(s).

Remarks: Any of the three fundamental solitons potentially would be able to carry charges which are pieces or fractions of those carried by other elementary particles in the same medium. From the above argument, no other solitons can carry fractional charge, a fact already understood for nontopological solitons [5]. A type of nontopological soliton much discussed recently is the 'Q ball' [6], [7], a configuration of a charged scalar field which carries a very large electric charge, stabilized by the attractive self-interaction of the scalar field. Such a system evidently can be made to disintegrate, and therefore could not carry a fraction of elementary charges.

Perhaps the most famous topological soliton which according to the above criterion is not fundamental is the skyrmion, described in the nonlinear sigma model by a 4-component scalar field with fixed magnitude. The soliton corresponds to a map from R_3 (with spatial infinity treated as a single point) to S_3. The winding number of the map is an exactly conserved integer, which Skyrme proposed should be identified with baryon number [8]. However, in its coupling to light fermions, the skyrmion must have effects equivalent to those of a similar object in the linear sigma model, where a fourth-order polynomial potential density is minimized for a specified magnitude of the Skyrme field. For this structure the topological quantum number could be destroyed with finite action by temporarily creating a zero in the Skyrme field at the center of the skyrmion, and then allowing the topological charge to flow into the zero and disappear.

MacKenzie and Wilczek [9] and D'Hoker and Goldstone [10] each computed flows of baryon current involved in the adiabatic creation of a skyrmion, and found integer baryon number, but implicitly left open the possibility that some exotic circumstance might produce a fractional result instead. Their construction works exactly the same way for the original skyrmions or for the almost-stable objects in the linear sigma model, and therefore the argument here shows that one never could obtain fractional charge for this system. There is one apparent way out: Insist that the high-energy behavior indeed is governed by the strict nonlinear sigma model, with fixed magnitude of the Skyrme field. The trouble now is that this theory is well-known to be nonrenormalizable, so that this option is undefined – one only may use the theory with an energy cutoff, which is equivalent to replacing the model with a linear sigma model. Stated differently, using the nonlinear sigma model implies particular behavior at arbitrarily high energies of what should be only an effective field theory. This would be tantamount to introducing new particles associated with those high energies.

Theorem 2 (Fractional charge from conventional renormalization): For a fractional charge to be associated with conventional renormalization, the particle must be an electric charge or a magnetic monopole in $d = 3$, or a kink in $d = 1$. Otherwise (in particular for the vortex in $d = 2$), the fractional charge must be intrinsic to the structure of the particle or fundamental soliton.

Proof: For conventional renormalization, as some coupling parameter changes adiabatically, there must be a current flow of the relevant charge into or out of the particle in question. Far away from the $d = 3$ electric charge, or the magnetic monopole, there is a radial $1/r^2$ field. Thus, a local current density proportional to that field and to the time rate of change of some scalar or pseudoscalar parameter would provide a steady net current into the particle. Clearly the long-range field is necessary to give direction as well as the correct radial dependence to the current density. In one space dimension, the different asymptotic behaviors of the field to right and left of the kink can determine locally the sign and magnitude of a current, again proportional to the time rate of change for some suitable coupling parameter.

Remarks: Thus, in these cases the possibility of creating fractional charge by continuous variation of a suitable parameter is open, while for other isolable objects it is not. Note that electrically charged particles in $d = 1$ or $d = 2$ are not isolable, because the energy for a particle-antiparticle pair diverges with separation. In all cases except the two in $d = 3$ and the one in $d = 1$ just described, the renormalization cannot be accomplished by a flow from infinity, and therefore must be intrinsic to the structure of the particle. For description of such intrinsic fractional charge as due to discrete renormalization to be meaningful, it must be possible to identify some 'core' of the particle, with its characteristic conserved charge, to which the medium polarization (leading to net fractional charge) is attached. That turns out to be possible, and so at least a useful perspective, for all known cases.

We have seen that among solitons only the fundamental ones (stabilized by long-range physics) can nucleate fractional charge. The obvious corollary is that the only other possible 'kernels' are those elementary particles which can be taken as given (i.e., determined by short-range physics). Blankenbecler and Boyanovsky [11] have presented another perspective which leads to the same conclusion as the one here for the case of fractional fermion charge induced by topology. They argue that the high-energy coupling of fermions carrying integer values of such a charge is influenced by the asymptotic field of the soliton, and this directly determines the fractional part of the charge localized on the soliton.

3. ILLUSTRATIONS AND COMMENTS

Before going on, it is important to emphasize that the above theorems give necessary conditions for fractional charge – they do not demonstrate that it occurs. Such demonstrations were the important content of works to be cited below.

It has been shown here that the only solitons whose topological charges could have a conventional renormalization to produce fractional values of certain other charges are kinks and monopoles, the two types of object first found by Jackiw and Rebbi [JR] [12] to carry fractional charge – to be precise, fermion number $F = 1/2$. Their results were verified elegantly using adiabatic flow methods by Goldstone and Wilczek [13] and by Seiberg and Witten [14]. These methods can be implemented in such a way that the soliton remains intact, while variation of certain couplings 'decorates' the object with fractional F and perhaps also fractional electric charge.

Fractional local charges are significant only if they are eigenvalues rather than expectation values; they must be locally conserved sharp quantum observables. For charges to be sharp, in one space dimension spatial smearing of the corresponding charge density operator is required [15], while in higher dimensions temporal smoothing is needed as well [16], [3].

Perhaps the most dramatic observation of fractional charge is associated with the fractional quantum Hall effect discovered by Tsui, Störmer, and Gossard [17]. Here Laughlin [18] concluded that the quasiparticles carry a simple fraction of an electron charge, so that an electron entering the medium could break up into several quasiparticles (something dramatically different from the breakup into reduced quasiparticle charge and remnant surface charge when an electron enters a conventional insulator). This result was vindicated in several experiments, by Goldman and Su, de Picciotto et al., Seminadayar et al., and successive works [19].

Jain's composite-fermion description [20] of odd-denominator FQHE states identifies the composite fermions as electrons whose strong repulsive mutual correlations renormalize their charges to the observed fractional values associated with the quasiparticles, so that it is natural to identify the quasiparticles as electrons dressed by the medium. In that perspective, the AB charge of a quasiparticle should be the same as that of an electron, but it is accepted that the force on the particle due to a Maxwell electric field $\vec{E}$ parallel to the Hall plane may be obtained by using precisely the fractional local charge already mentioned. In the dressed-electron picture, this is understood as resulting from an induced Chern-Simons field in the Hall plane which partly comensates the effect of the Maxwell field [21]. The large conceptual advantage of this perspective, embodied in composite-fermion theory [20], is that it not only unifies the description of different FQHE regimes, but also provides a close correspondence with familiar condensed-matter systems, and their quasielectron excitations.

The original description of a quasihole for simple Hall fractions was given by Laughlin [18] in terms of a fundamental soliton in the form of a special type of vortex – a vortex in the many-body ground-state wave function. This vortex requires neither Higgs nor gauge field, but only has been realized under the special conditions for which the Laughlin ansatz gives the ground state wave function. Now one describes the quasihole as a soliton with local charge and AB charge both equal to the same fraction. The ability to describe a quasiparticle either as a fundamental soliton or a dressed particle is not unique. Possibly the earliest example of such dual descriptions of a particle comes in the ($d = 1$) duality between the Thirring model with its fermionic excitations and the sine-Gordon model with its solitons [22]. However, the two descriptions of FQHE quasiparticles may be the first or even the only case where such duality applies to a particle with fractional charge. Whether one chooses the vortex or the dressed-electron description, the local charge is the same fraction, but the AB charge either has fractional or unit value. In these terms, the experiment of Goldman and Su mentioned above was a measurement of fractional effective AB charge, while the other experiments measuring shot noise were sensitive to fractional local charge. Of course, all observables are identical for the case where the two descriptions coincide, namely, quasiholes in simple Laughlin states, a truly remarkable duality. However. for quasiparticles in these states, and both holes and particles of all other (odd-denominator) Hall fractions, there is no successful soliton prescription, while the composite-fermion description has been worked out in full detail, exhibiting excellent quantitative agreement with all known experimental and theoretical tests [23]. This dressed-electron picture differs from familiar renormalization because the correlations producing it are quantized by the requirement of single-valuedness of the many-body wave function, so cannot be achieved by continuous wave-function deformation.

After JR, and following the independent work of Su, Schrieffer, and Heeger [SSH] on kinks in polyacetylene [24], there were a number of studies confirming and elaborating on the original finding that solitons can carry fractional charge. Shankar and Witten [25] used bosonization to put fermions and bosons on the same footing in the kink system, taking account of possible back-reaction by fermion on boson degrees of freedom, and confirming the JR result. Su and Schrieffer [26] found examples of kinks in condensed matter models with other rational fractions. Earlier, Witten [27] showed that magnetic monopoles could have fractional electric charge determined by the vacuum angle (or equally well by a crossed electric-magnetic susceptibility like that for a medium with dipolar molecules carrying both electric and magnetic moments). Sikivie [28] put this in the context of conventional insulator behavior, showing that if a monopole passes through a domain wall between different values of the vacuum angle θ, then the change in electric charge of the monopole is exactly compensated by a change in the surface charge spread over the wall.

The work of Witten [27] had two parts. First was a demonstration that the very gauge transformation which has a compact U(1) action on conventional charged particles with quantized electric charge can have a noncompact action on a magnetic monopole, allowing fractional electric charge on the monopole. [For an extended treatment of this point see [29].] This discussion complements the method of Theorem 3 above, based on the long-range monopole magnetic field. The noncompact action of the electric-charge gauge transformation on the dyon is consistent with gauge invariance because in the mutual interaction of two dyons there is besides the normal gauge interaction an extra term, gauge-variant but not contributing to the equations of motion for the pair [30], [31].

The second part of Witten's work was a direct construction of the fractional charge from the vacuum angle. Thus he found both necessary and sufficient conditions for the monopole to carry fractional electric local and AB charge. Note that the latter clearly is not conserved when the monopole passes through a domain wall, because AB charge is not defined for the (effectively immobile) wall.

As mentioned, Goldstone and Wilczek [GW] [13] developed an adiabatic flow analysis showing that F for the JR monopole must change by $\Delta F = \frac{1}{2}$ as the fermion mass is chirally rotated from isoscalar to isovector, a shift which preserves the electric charge of the monopole, as well as reflection symmetry between positive and negative electric charge, but violates fermion charge-conjugation symmetry. Callan [32] considered fermions light even compared to the Coulomb energy required to confine a unit of electric charge within a monopole radius, finding $F = \frac{1}{2}$ appears in a natural way. Over a long period, the notions of electric-magnetic duality, supersymmetry, and JR fermion zero modes were locked together, mutually reinforcing all three [33], [14]. In particular, Seiberg and Witten [SW] [14] used an adiabatic flow analysis complementary to that of GW, which allowed them to follow the change of F and electric charge Q as the isoscalar part of the fermion mass is lowered from infinity, where the fermion is totally decoupled (meaning any fractional part of F must vanish), to a point where the net mass of one member of the fermion isodoublet goes through zero. This gives $F = 1/2$, with electric charge also changing by $1/2$ as vacuum angle changes by $\pi/2$, preserving fermion charge-conjugation symmetry and the double degeneracy of the monopole ground state (but violating electric-charge reflection symmetry), and putting into clear perspective the case in [32]. The GW and SW analyses together confirm remarks in [12],[34] that not all classical symmetries of the fermion-monopole system can be preserved under quantization. $F = 1/2$ is a robust result, but vacuum angle and hence fractional electric charge are affected by the way in which the fermion Yukawa coupling to the Higgs field is generated.

This subject of adiabatic flow brings attention to the beautiful thought experiment of Laughlin [18], who imagined piercing a Hall layer with an infinitely thin tube of flux, gradually increased from zero to one flux quantum. Through the Faraday effect and the Hall effect this assures the localization of a fractional charge, immediately showing that the quasiparticles of this system must have fractional charge, with respect to electrons in a different medium – free space. Note that the fractional value is not a surprise, in view of the behavior of normal insulators. However, if we imagine inserting an electron into the layer from above, there is no surface in which to leave part of its locally conserved electric charge, so that the charge instead must be deposited on several quasiparticles. This breakup into many particles was a new and theoretically unanticipated phenomenon.

The concept of fractional soliton charge was at least implicit in the work of Skyrme [8], who argued that his classical field configuration could be quantized with half-integer isospin and spin (a possibility shown consistent with the usual spin-statistics connection in [35]). Half-integer values would allow the skyrmion to be identified with the nucleon, but by Theorem 2 are impossible without elementary or fundamental isospinor fermions: Modification of the short-distance, high-energy part of the Skyrme action (e.g., replacing the nonlinear constraint in his sigma model with a quartic action in the chiral field) could destabilize the skyrmion, so that its 'topological' charge would be 'unwound' in a process with finite action, and therefore in principle not absolutely conserved. Microscopic analyses agree, indicating that the spin and isospin of the skyrmion will be integer or half-integer as the number of colors of up and down quarks is even or odd [36], and therefore integer if there are no quarks.

These considerations may be put more dramatically. The fact that the skyrmion is in a class of objects some of which are not conserved immediately implies that there must be some underlying structure to account for spin and isospin charge values that are fractional with respect to the meson charges of the sigma model. Thus at best the skyrmion could be a useful description for reasonably low-energy and long-distance properties of the nucleon. That indeed is the case, but this simple reasoning could have been made at any time after Skyrme's original work. Perhaps an intuitive appreciation of this point contributed to initial resistance to his model. Paradoxically, if the model had been embraced, it might have slowed down the approach to quarks and QCD which now gives an intellectual basis for the skyrmion's success in the appropriate domain.

Skyrme's model describes the nucleon entirely in terms of an $SU(2)$ matrix function $U(\mathbf{r}, t)$, while in a 'hybrid' model the U function is used outside a chosen 'bag' radius R, and inside are free quarks with boundary condition at the bag wall parameterized by the chiral angle associated with $U(|\mathbf{r}| = R)$ [37]. Goldstone and Jaffe [38] showed that the simple boundary condition guessed in [37] meets the requirement of net integer baryon number B. Thus, for the nucleon it becomes possible to interpolate smoothly between (nonfundamental) soliton and (fundamental) particle (quark) descriptions, and therefore neither can involve intrinsically fractional charges.

An analogous interpolation has been found for FQHE quasiparticles, which for $\nu = 1/(2n+1)$ can blow up into arbitrarily spread-out 'baby skyrmions' when the Zeeman splitting between the two possible electron spin orientations becomes negligible. Thus, adjusting the Zeeman splitting allows interpolation between microscopic quasiparticles and macroscopic solitons. Of course, the charges of the soliton and the quasiparticle are the same. This theoretical result for FQHE follows well-established results for skyrmions of the integer quantum Hall effect with small Zeeman splitting [39].

The SSH kink analysis [24] shows that in one space dimension 'spinons' with spin $\frac{1}{2}$ but no charge and 'holons' with charge $\pm e$ but no spin can travel independently. Kivelson et al. [40] proposed that such objects might play a role in the planar dynamics which appears to be critical in high T_C superconductivity. Detailed studies suggested that if so, either these fractional objects are connected by strings [41] or are able to move only along certain lines in the plane [42]. However, Senthil and Fisher [43] observed that a dynamics leading to vortices in some effective gauge field could make the string connecting a spinon and a holon simply a gauge artifact, exhibiting zero tension. Hence it becomes imaginable that the two types of particle could move freely in the plane.

The uniqueness of magnetic monopoles among solitons in $d = 3$ as possible carriers of fractional particle charge is connected with other special properties, such as the ability to convert the dynamics of the lowest fermion partial wave into a one-dimensional problem on a half-line. This is an example of the fact that the chiral anomaly for electrodynamics with $d = 3$ may be written as the product of a magnetic-field contribution which reduces the problem to $d = 1$, and an electric-field contribution just like that for QED in $d = 1$ [44]. The same long-range magnetic field is responsible for the unique possibility of creating an object with half-integer spin [45] and Fermi-Dirac statistics [46] from bosons in a world with no fundamental fermions, a possibility *not* available to the skyrmion, contrary to some statements in the literature.

An interesting example of fractional charge is the Higgs-Chern-Simons vortex in abelian 2+1 D gauge theory, a soliton which carries a conserved topological charge, the quantized magnetic flux. There is a locally conserved Gauss-law electric charge $Q = \kappa\Phi + q_H$, with κ the Chern-Simons [CS] coupling [47], Φ the quantized magnetic flux, and q_H the Noether charge of the Higgs field. Evidently Q vanishes by the Gauss law, but of course q_H does not, and is not even conserved if κ varies. Indeed, with the gauge kinetic term F^2 omitted, the resulting 'self-dual' vortex [48] has vanishing Q density everywhere! This system manifestly violates electric charge conjugation symmetry, and generates fractional values for q_H. The fact that q_H would vary if κ changed implies that κ must be constant if one is to interpret q_H as a conserved fractional charge. With this assumption, one sees that a given value of κ, crucial to the soliton dynamics, indeed enforces an intrinsic relation between the charge and the soliton structure, maintaining consistency with Theorem 3. For nonabelian CS theory, there must be quantization of κ [49], but for the abelian case relevant here there may be some flexibility in the allowed values. The self-dual vortex has been proposed, though without explicit identification as such, to be the charge-carrying quasiparticle of the $\nu = 5/2$ FQHE state.

As in all FQHE phenomena one has here an interesting dimensional hybridization: The dynamics of the Hall layer, including a long-range 'statistical' or pure-gauge interaction with for the 5/2 case both nonabelian [50] and abelian [51] contributions, is in $d = 2$. Meanwhile, the dynamics of the Maxwell field remains in $d = 3$, so that electric charge is isolable even though it would not be so in a fully $d = 2$ system. Such dimensional hybridization has been discussed in string-inspired brane theory, with, for example, gauge fields propagating in the brane and gravity propagating in the 'bulk', i.e., the entire space [52].

Up to now we have not considered fractional spin in this discussion. Of course, in $d = 1$ spin is not defined, while in $d = 3$ the nonabelian character of the rotation group assures spin quantization. However, in $d = 2$ the logical possibility of fractional spin is open. Paranjape [53] realized that such spin indeed could be induced by magnetic flux; Boyanovsky and Blankenbecler [54] gave a simple exposition of the mechanism. That the familiar connection between spin and statistics holds in $d = 2$ for fractional spin and fractional statistics follows from elementary conservation laws [55],[21]: To be precise, if one defines $s_e = (s + \bar{s})/2$, the part of the spin symmetric under particle$\leftrightarrow$antiparticle, with $0 < |s_e| < \frac{1}{2}$, then there is a contribution to the phase factor on exchange of two indistinguishable particles given by $e^{-2\pi i s_e}$.

The one example of fractional charge by conventional renormalization that remains to be discussed is that of the familiar electrically charged elementary particles mentioned at the beginning, the proton and the electron. Suppose we say that the neutron has $\mathrm{B-L} = 1$, and the neutrino has $\mathrm{B-L} = -1$. What is the value of $\mathrm{B-L}$ for the charged particles? In first approximation, one may neglect all contributions to QED vacuum polarization except that of electron loops. As a result, the proton is accompanied by a tenuous cloud of electron vacuum polarization, and so has still $\mathrm{B} = 1$, but $\mathrm{L} = \epsilon$. Meanwhile, charge renormalization of the electron implies that it has $\mathrm{L} = 1 - \epsilon$. Thus, $\mathrm{B-L}$ for the proton is $1 - \epsilon$, and $\mathrm{B-L}$ for the electron is $\epsilon - 1$. For neutron decay, this gives initial $\mathrm{B-L} = 1$, and final $\mathrm{B-L} = (1-\epsilon)_{\mathrm{p}} + (\epsilon - 1)_{\mathrm{e}} + 1_{\bar{\nu}} = 1$. In principle, it would be consistent to introduce a new gauge field, weakly coupled to $\mathrm{B-L}$. This would allow direct observation of the different values for the neutral and the electrically charged particles. However, the main point to make here is that if there is some locally conserved charge carried by some particle, then it can consistently generate a fractional shift in another charge carried by that particle. The fact that in some cases this fractional shift is quantized, while in others it can be varied continuously, is important, but the parallelism between the two types of case may be even more important.

A complementary perspective emerges from considering the $B - L$ gauge-field coupling between neutron and proton. What was discussed already shows that the neutron would see a somewhat weaker field due to electron vacuum polarization. Alternatively, if one considers the influence of the neutron on the proton, the neutron's $B - L$ field mixes with the electromagnetic field due to the same vacuum polarization, so that it couples to the combination of AB charges $B - \epsilon Q$. The AB charges do not change, but the field acting on them is modified in a nontrivial way. Of course, in line with the principle of reciprocity, either way of calculating the interaction between neutron and proton gives the same answer.

4. CONCLUSION

Results of this codification and development of the literature on fractional charge include:

(1) Gauge charge, which in the case of a charge coupled to a local gauge field is the Lorentz-force charge of a particle in classical mechanics or the Aharonov-Bohm charge in quantum mechanics, is a fundamental, quantized attribute of a fundamental particle or an elementary particle built of fundamental particles, never renormalized or fractionated by change in medium, scale, or coupling strength. Gauge charge is not necessarily conserved; for example, there may be processes conserving local charge while allowing excitation of the vacuum to mobilize previously latent particles carrying gauge charge. For a fundamental soliton, there may be a fractional gauge charge compared to the unit found on elementary particles in the same medium, if so with the fraction equal to the corresponding value for local or Gauss-law charge.
(2) The (locally conserved) local charge of a particle with specified gauge charge may have a fractional value in one medium or at one scale, with respect to its value in a different medium or at a different scale.
(3) In a given medium an isolable particle may carry a fraction of the local charge(s) of other, elementary particles only if the first one is itself elementary, which includes the possibility that it is a fundamental soliton.
(4) Fractional charge may result from conventional, continuous flow if the particle is a $d = 1$ kink, or a $d = 3$ electric charge or magnetic monopole. For a $d = 2$ vortex fractional charge can only arise from intrinsic structure.

The considerations about elementary particles and fundamental solitons show a striking (and typical) duality. The particles and their constituents are established in the dynamics of the smallest distance and highest energy scales, while the stability of the solitons is assured by the dynamics of the largest distance and lowest energy scales.

It is interesting to wonder about possible extensions of this analysis. There may well be novel charges for $d > 3$, and in certain circumstances fractional values of these charges. Even within the domain $d \leq 3$, there is a class of issues that remain a matter of art rather than systematic deduction, namely, the determination of the discrete renormalizations which by definition are not directly susceptible to the well-developed techniques used for conventional, continuous renormalization. Perhaps consistency relations of the type developed by Su [56] for FQHE would help.

This article has almost no equations, and deals as well with solitons which do and ones which do not exhibit exact integrability, the latter evidently violating the original definition of the term by Zabusky and Kruskal [57]. Mathematical physics uses the notion of exactness in different ways. Emphasized in this journal issue are exactly soluble systems, used to infer generic properties that may be compared with real systems whose dynamics generally are not exactly soluble. On the other side are systems which may not be fully soluble, but because of local conservation laws still may show some exact properties. This work manifestly is intended as a contribution to the latter category, determining, from mild assumptions, stringent requirements for the occurrence of fractional charge.

Acknowledgments: A sceptical remark by Jeffrey Goldstone about skyrmions with fractional B was an early stimulus for this study, supported in part by the National Science Foundation under grant PHY-0140192. Michael Dine, Roman Jackiw, Jainendra Jain, Vladimir Korepin, Barry McCoy, and Mikhail Stephanov gave instructive comments and criticisms. The Theoretical Physics Group at the Stanford Linear Accelerator Center (Department of Energy Contract No. DE-AC03-768F00515) and the Center for Theoretical Physics at the Massachusetts Institute of Technology (Department of Energy Cooperative Agreement No. DF-FC02-94ER40818) provided generous hospitality during a sabbatical leave.

References

[1] G. 't Hooft, Phys. Rev. D **14**, 3432 (1976); Erratum-ibid. D **18**, 2199 (1978).
[2] S.L. Adler, Phys. Rev. **177**, 2426(1969); J.S. Bell and R. Jackiw, Nuovo Cim. **A60**, 47 (1969).
[3] A.S. Goldhaber and S.A. Kivelson, Phys. Lett. **255B**, 445 (1991).
[4] D.J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973); H.D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
[5] T.D. Lee and Y. Pang, Phys. Rept. **221**, 251 (1992).
[6] S. Coleman, Nucl. Phys. **B262**, 263 (1985).
[7] A. Kusenko and M.E. Shaposhnikov, Phys. Lett. **B418**, 46 (1998).
[8] T.H.R. Skyrme, Proc. R. Soc. (London) **A260**, 127 (1961).
[9] R. MacKenzie and F. Wilczek, Phys. Rev. D **30**, 2194,2260 (1984).
[10] E. D'Hoker and J. Goldstone, Phys. Lett. **B158**, 429 (1985).
[11] R. Blankenbecler and D. Boyanovsky, Phys. Rev.D **31**, 2089 (1985).
[12] R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).
[13] J. Goldstone and F. Wilczek, Phys. Rev. Lett. **47**, 986 (1981).
[14] N. Seiberg and E. Witten, Nucl. Phys. **B431**, 484 (1994).
[15] S. Kivelson and J.R. Schrieffer, Phys. Rev. B **25**, 6447 (1982); R. Rajaraman and J.S. Bell, Phys. Lett. **116B**, 151 (1982); S.A. Kivelson, Phys. Rev. B **26**, 4269 (1982); J.S. Bell and R. Rajaraman, Nucl. Phys. **B220**, 1 (1983); Y. Frishman and B. Horovitz, Phys. Rev. B **27**, 2565 (1983); R. Jackiw, A.K. Kerman, I. Klebanov, and G. Semenoff, Nucl. Phys. **B225**, 233 (1983).
[16] M. Requardt, Commun. Math. Phys. **50**, 259 (1976).

[17] D.C. Tsui, H.L. Störmer, and A.C. Gossard, Phys. Rev. Lett. **48**, 1559 1982.
[18] R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).
[19] V.J. Goldman and B. Su, Science **267**, 1010 (1995); R. de Picciotto, M. Reznikov, M. Heiblum, V. Umansky, G. Bunin, and D. Mahalu, Nature **389**, 162 (1997); L. Saminadayar, D.C. Glattli, Y. Jin, and B. Etienne, Phys. Rev. Lett. **79**, 2526 (1997); M. Reznikov, R. de Picciotto, T.G. Griffiths, M. Heiblum, and V. Umansky, Nature **399**, 238 (1999); T.G. Griffiths, E. Comforti, M. Heiblum, A. Stern, and V. Umansky, Phys. Rev. Lett. **85**, 3918 (2000); Y.C. Chung, M. Heiblum, V. Umansky, Phys. Rev. Lett. **91**, 216804 (2003).
[20] J.K. Jain, Phys. Rev. Lett. **63**, 199 (1989).
[21] A.S. Goldhaber and J.K. Jain, Phys. Lett. **199A**, 267 (1995).
[22] W.E. Thirring, Ann. Phys. **3**, 91 (1958); J.K. Perring and T.H.R. Skyrme, Nucl. Phys. **31**, 550 (1962); S.R. Coleman, Phys. Rev. D **11**, 2088 (1975).
[23] S. Das Sarma and J. Pinczuk (Eds.), *Perspectives in quantum Hall effects* (Wiley, New York) 1997.
[24] W.P. Su, J.R. Schrieffer and A.J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).
[25] R. Shankar and E. Witten, Nucl. Phys. **B141**, 349 (1978); **B148**, 538 (E) (1979); E. Witten *ibid* **B142**, 285 (1978).
[26] W.P. Su and J.R. Schrieffer, Phys. Rev. Lett. **46**, 738 (1981).
[27] E. Witten, Phys. Lett. **86B**, 282 (1979).
[28] P. Sikivie, Phys. Lett. **B137**, 353 (1984).
[29] D. Giulini, Mod. Phys. Lett. **A10**, 2059 (1995).
[30] F. Wilczek, Phys. Rev. Lett. **49**, 957 (1982).
[31] A.S. Goldhaber, R. MacKenzie, and F. Wilczek, Mod. Phys. Lett. **A4**, :21 (1989).
[32] C.G. Callan Jr., Phys. Rev. D **26**, 2058 (1982).
[33] C. Montonen and D. Olive, Phys. Lett. **72B**, 117 (1977); P. Goddard, J. Nuyts, and D. Olive, Nucl. Phys. **B125**, 1 (1977); E. Witten and D. Olive, Phys. Lett. **78B**, 97 (1978); H. Osborn, Phys. Lett. **83B**, 321 (1979); A. Sen, Int. J. Mod. Phys. **A9**, 3707 (1994); N. Seiberg and E. Witten, Nucl. Phys. **B426**, 19 (1994); (Erratum) *ibid* **B430**, 485 (1994).
[34] A.S. Goldhaber, Phys. Rev. D **16**, 1815 ((1977).
[35] D. Finkelstein and J. Rubinstein, J. Math. Phys. **9**, 1762 (1968).
[36] E. Witten, Nucl. Phys. **B223**, 433 (1983); E. D'Hoker and E. Farhi, Phys. Lett. **134B**, 86 (1984).
[37] M. Rho, A.S. Goldhaber, and G.E. Brown, Phys. Rev. Lett. **51**, 747 (1983).
[38] J. Goldstone and R.L. Jaffe, Phys. Rev. Lett. **51**, 1518 (1983).
[39] R.K. Kamilla, X.G. Wu, and J.K. Jain, Solid State Commun. **99**, 289 (1996). See references therein for the IQHE case.
[40] S.A. Kivelson, D.S. Rokhsar, and J.P. Sethna, Phys. Rev. B **35**, 8865 (1987).
[41] R.B. Laughlin, J. Low Temp. Phys. **99**, 443 (1995).
[42] V.J. Emery, S.A. Kivelson, and O. Zachar, Phys Rev B **56**, 6120 (1997).
[43] T. Senthil and M.P.A. Fisher, J..Phys. A **34**, L119 (2001).
[44] J. Ambjørn, J. Greensite, and C. Peterson, Nucl. Phys. **B221**, 381 (1983).
[45] P.A.M. Dirac, Proc. R. Soc. London. Ser. A **133**, 60 (1931); M.N. Saha, Indian J. Phys. **10**, 145 (1936) and Phys. Rev. **75**, 1968 (1949); H.A. Wilson, Phys. Rev. **75**, 309 (1949); A. S. Goldhaber. *ibid.*, **140**, B1407 (1965): R. Jackiw and C. Rebbi, Phys. Rev. Lett. **36** 1116 (1976); P. Hasenfratz and G. 't Hooft, *ibid.* **36**, 1119 (1976).
[46] A.S. Goldhaber, Phys. Rev. Lett. **36**, 1122 (1976).
[47] W. Siegel, Nucl. Phys. **B156**, 135 (1979); R. Jackiw and S. Templeton, Phys. Rev. D **23**, 2291 (1981); J.F. Schonfeld, Nucl. Phys. **B185**, 157 (1981).
[48] J. Hong, Y. Kim, and P. Y. Pac, Phys. Rev. Lett. **64**, 2230 (1990); R. Jackiw and E. Weinberg, *ibid.*, 2234 (1990); R. Jackiw, K. Lee, and E.J. Weinberg, Phys. Rev. D **42**, 3488 (1990). J.Phys.A **34**, L119 (2001).
[49] S. Deser, R. Jackiw, S. Templeton, Phys. Rev. Lett. **48**, 975 (1982), Annals Phys. **140**, 372 (1982), Erratum-ibid.**185** 406 (1988), Annals Phys. **281**, 409 (2000).
[50] G.W. Moore and N. Read, Nucl. Phys. **B360**, 362 (1991).
[51] M. Greiter, X-G. Wen, and F. Wilczek, Phys. Rev. Lett. **66**, 3205 (1991).
[52] W. Siegel,⟨http://insti.physics.sunysb.edu/∼siegel/parodies/sgs.html⟩ (1985), J. Polchinski, hep-th/9611050.
[53] M.B. Paranjape, Phys. Rev. Lett. **55**, 2390 (1985); Phys. Rev. D **36**, 3766 (1987).
[54] R. Blankenbecler and D. Boyanovsky, Phys.Rev. D **34**, 612 (1986).
[55] D.J. Thouless and Y-S. Wu, Phys. Rev. B **31**, 1191 (1985); A.S. Goldhaber and R. MacKenzie, Phys. Lett. **B214**, 471 (1988); T.H. Hansson, M. Roček, I. Zahed, and S.C. Zhang, *ibid.*, **214**, 475 (1988); T. Einarsson, S.I. Sondhi, S.M. Girvin, and D.P. Arovas, Nucl. Phys. **B441**, 515 1995; J.M. Leinaas, cond-mat/9903329. .
[56] W.P. Su, Phys. Rev. B **34**, 1031 (1986).
[57] N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).