The Lesson of Quantum Theory, edited by J. de Boer, E. Dal and O. Ulfbeck

Strange Quantum Numbers in Condensed Matter and Field Theory

J. Robert Schrieffer

University of California,
Santa Barbara, California, USA

Contents

Aside

It is a great honor and pleasure for me to join in celebrating the hundredth anniversary of the birth of Niels Bohr. I had the good fortune of spending the spring of 1958 in the Summer House on Blegdamsvej. There I met many exciting physicists and was moved by the tradition of openness and enthusiasm for science which pervaded the Institute. On several occasions Niels Bohr invited me to discuss with him the theory of superconductivity. We were joined by Aage Bohr and Léon Rosenfeld. The discussions focussed on Bohr's ideas concerning superconductivity, some of which trace back to the late 1920s. We also discussed the theory just proposed by Bardeen, Cooper and myself. These discussions made a very deep impression on me and remain one of the high points of my life.

In addition to benefiting from the science and hospitality at Blegdamsvej, I learned of Weisskopf's theorem, and soon provided another proof of its validity by marrying a lovely Danish girl who continues to provide ties to Denmark. It is in this sense of extended family that I join with you and the Bohr family in this joyful celebration.

1. Introduction

From the early days of quantum theory, a fundamental principle of measurement theory has been the indivisibility of an elementary particle. For example, consider the traditional Stern–Gerlach experiment illustrated in fig. 1, in which a particle of spin $s = \frac{1}{2}$ is prepared in an eigenstate of σ_x. The particle passes through a magnetic field gradient along z, such that the incoming wave splits into two separated beams, + and −, which pass through slits in a shutter and strike a screen which detects the beams.

If the wavefunction $\psi(r, t)$ describes a classical field, one would expect one half a particle to pass through each slit, each time a particle is projected toward the shutter. On the contrary, quantum theory and experiment show that either:

(1) a full particle is observed to pass through the + slit with charge $-e$, and no particle passes through the − slit, or

(2) a full particle passes the − slit and not through the + slit.

The probability for observations (1) and (2) are both $\frac{1}{2}$. Thus, an elementary particle cannot be split into two half particles.

A related question is the following: can stable particles (excitations) of sharp fractional charge exist in systems composed solely of particles of integer charge? Several clarifications are in order. Firstly, by sharp we mean that every time the charge of the particle is accurately measured, the same fractional result will occur, rather than a distribution of values, the average of which is fractional. Secondly, while vacuum polarization effects renormalize all charge, these effects cancel out in the ratio of the dressed fractional charge and the dressed primative charge from which the system is constructed.

The first step in this story occurred in 1976 when R. Jackiw and C. Rebbi [1] discovered that if a spinless Dirac field is coupled to a nonlinear scalar background field which supports kink-like solutions in one space dimension, the negative energy sea in the vicinity of the kink is depleted by one half a fermion. Also, a zero-energy fermion state (zero mode) was found to be bound to the kink, which, if filled, leads to a total excess of one half a fermion associated with the kink.

Independently, W.P. Su, A.J. Heeger and the author [2], studying quasi one-dimensional conductors, discovered that soliton excitations can occur whose charge

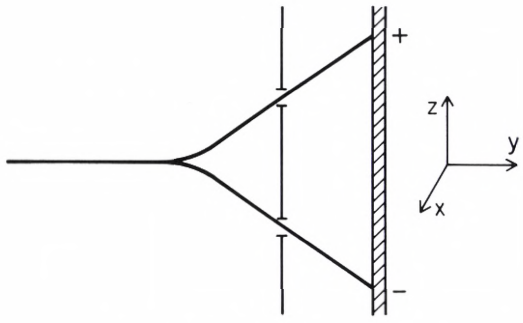

Fig. 1. Stern–Gerlach experiment.

per spin is $\pm e/2$, where $-e$ is the electronic charge in the medium. Since two spin directions are equally populated in these systems, the soliton carries a total charge $Q = -e$, but a spin $s = 0$. This surprising result appears to violate Kramer's theorem relating the electron number and the allowed spin value, i.e. one electron $(-e)$ must correspond to spin $\frac{1}{2}$. Fundamentally, these peculiar effects arise from the same physical effect: vacuum charge flows at domain walls in quantized amounts in systems having discretely degenerate vacua. The sharp quantization of charge is a consequence of this discrete symmetry breaking. In essence, the wave-functions of the negative-energy (occupied) fermion states are distorted by the Bose field without change of the fermion occupation numbers [3].

A third example of fractional charge arises in the fractional quantum Hall effect. In this case, electrons confined to two space dimensions are subject to a strong magnetic field perpendicular to the plane. While semiclassical theory predicts the transverse conductivity σ_{xy} to be a linear function of the electron density ν, experiments by von Klitzing et al. [4] showed that σ_{xy} exhibits plateaus at integer values of ν, corresponding to filled Landau levels. Steps at fractional ν were observed by Tsui et al. [5]. To account for the latter result, Laughlin [6] proposed a fluid-state theory possessing fractionally charged quasi particles. Recently, an alternative theory based on a Wigner-crystal approach has been advanced by Kivelson et al. [7] and this also exhibits fractionally charged excitations. In these theories, as in those mentioned above, only vacuum flow of current is involved in the fractional charge. However, the charge quantization comes about from a somewhat different mechanism in the Hall effect, namely local energy stabilization near the core of the excitation rather than discretely degenerate vacua extending over large regions of space (nonlocal energy stabilization).

Below, we briefly discuss these examples of "charge splitting without violating quantum mechanics".

2. *A classical example*

Consider the infinite line with the integers marked off as in fig. 2(top). We place particles, each of charge q, on the odd sites starting at $-\infty$, leaving the even sites vacant. Having filled a given site, say number 1, we make an error and place the next particle at 2 instead of 3. We continue placing charges on every other site, i.e. the even sites to $+\infty$. Note that we have made two domains: $-\infty$ to 1 with odd sites occupied, even sites empty (termed the A phase); and 2 to $+\infty$ with the reverse occupancy (termed the B phase). A domain wall separates the two phases and is located at the midpoint $x = \frac{3}{2}$ between the two adjacent occupied sites.

Suppose we move the particle initially located on site 2 to site 3, as shown in fig. 2(bottom). Notice that the midpoint between the two adjacent occupied sites has moved to $x' = \frac{3}{2} + 2$, that is, the domain wall moves *two units* even though we moved the particle only *one unit*! Let us now determine the effective charge Q associated with the domain wall, i.e. the charge which will be observed in long-wavelength experiments. We do this by equating the charge in the electric dipole moment Δp of the system calculated in two ways, namely through the particle motion

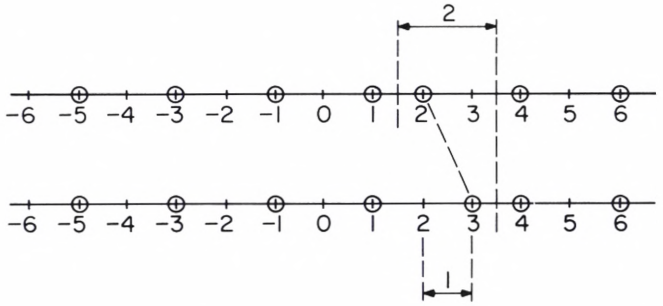

Fig. 2. A classical example of a fractionally charged excitation, illustrating the difference between the particle and domain-wall interpretation of configuration changes.

$q \cdot (+1)$, and through the domain wall motion $Q \cdot (+2)$. Since Δp must be independent of how we describe the system we have $q = 2Q$ or $Q = q/2$. Thus from integer charges q, we have discovered an "excitation" whose charge Q is fractional, $Q/q = \frac{1}{2}$. While this example appears to be trivial, it illustrates one mechanism for the occurrence of fractional charge, namely domain-wall or more generally topological solitons separating degenerate ground states which have different negative-energy fermion wavefunctions. The charge conjugate wall with $Q/q = -\frac{1}{2}$ is given by leaving two vacant spaces rather than no spaces between particles at the wall.

Clearly, we could generalize the model to charge $Q/q = \frac{1}{3}$, by placing charges on every third site, with one rather than two spaces between particles at the wall, etc. [8]. Below we discuss fractionalization in quantum systems.

3. Solitons and chain conductors

The simplest example of charge fractionalization in a quantum system occurs in the linear polymer [9] trans-polyacetylene $(CH)_x$, illustrated in fig. 3. While each carbon has four valence electrons, three of these are dynamically mute, being involved in strong bonds with its three neighboring atoms (two C's and one H). The remaining electron is in a p_z orbital and is free to wander along the chain to form a one-dimensional metal. In band language, this p_z band is half filled, because of two spin states, and should lead to large electrical conductivity. In fact, precisely the reverse is true, as first discussed by R.E. Peierls [10]. He showed that any one-dimensional Fermi gas, when coupled to lattice distortions or phonons, leads to a spontaneous symmetry-breaking. In this case the coupling breaks the combined symmetry operation of translation by one unit along x, and π rotation about x. The broken-symmetry state has a periodicity of 2 units along x and corresponds to a modulation of carbon–carbon bond lengths, long, short, long, short, ... In practice these bond-length changes are very small, less than 5% of the unstretched bond; however, they have a dramatic effect on the electron spectrum, opening up a gap of $2\Delta \simeq 1.4$ eV at the Fermi surface, converting the undistorted metal to a large-gap

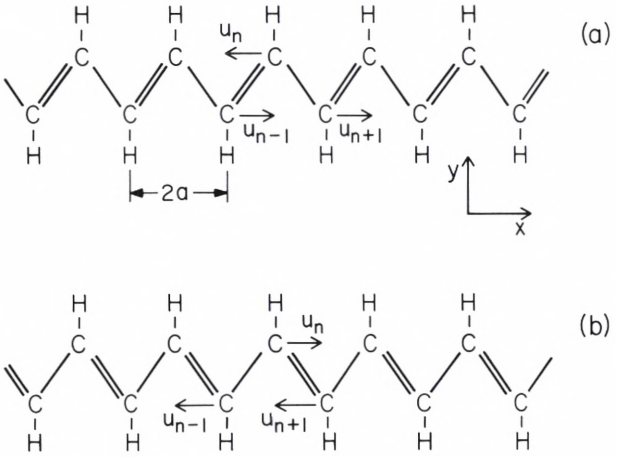

Fig. 3. Trans-polyacetylene $(CH)_x$ in its two degenerate ground states.

semiconductor. The fact that the gap opens at precisely the Fermi surface is no accident since the system energy would increase were it elsewhere in momentum space. Chemists term the Peierls instability dimerization or bond alternation.

To obtain a more quantitative understanding of these effects consider the model Hamiltonian [2] describing the coupled electron–phonon system in trans $(CH)_x$:

$$H = \sum_{n,s} t_n (C^+_{n+1,s} C_{ns} + \text{h.c.}) + \sum_n \left\{ \frac{\Pi_n^2}{2M} + \frac{K}{2}(u_{n+1} - u_n)^2 \right\}, \tag{1}$$

where C^+_{ns} creates a p_z electron of spin s on site n. u_n is the displacement of the nth (CH) group along the x-axis, with the momentum Π_n conjugate to u_n. The electronic hopping matrix element t_n is modulated by the phonons and is well approximated by

$$t_n = t_0 + \alpha(u_n - u_{n+1}), \tag{2}$$

where t_0 generates the bare-electron band structure and α is the electron–phonon coupling constant. In field theoretic terms, H describes a Fermi field linearly coupled (α) to a Bose field on a lattice. The finite lattice spacing provides a natural cutoff for the theory.

Despite the simplicity of H, its ground state and excitations are known only approximately at present. Fortunately, typical electron frequencies are large compared to phonon frequencies so that the electrons can be integrated out within the adiabatic approximation (one-loop level) to obtain an effective potential V_{eff} for the phonon field. It is convenient to introduce the staggered Bose field

$$\phi_n \equiv (-1)^n u_n, \tag{3}$$

to remove the rapid spacial oscillations of u_n which occur near the minimum-energy

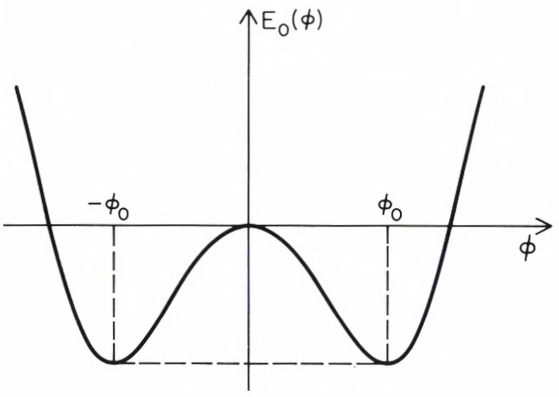

Fig. 4. The effective potential V_{eff} as a function of the amplitude of the staggered displacement (Bose field) ϕ, for ϕ spacially uniform. The peak at $\phi = 0$ exhibits the Peierls instability with degenerate vacua at the broken-symmetry values $\pm \phi_0$.

configuration. For the special case $\phi_n = \phi$, i.e. perfect alternating bond length with amplitude ϕ, V_{eff} is illustrated in fig. 4. The negative curvature of V_{eff} at $\phi = 0$ reflects the Peierls instability, while minima at $\pm\phi_0$ represent the two-fold degenerate broken-symmetry vacua. The physical origin of this dynamical symmetry breaking is clear. Nonzero $\langle\phi\rangle$ leads to a gap 2Δ (or mass $m \equiv \Delta$) in the electron spectrum, where $\Delta = 4\alpha\langle\phi\rangle$, as shown in fig. 5. The energies of the occupied states are lowered by the existence of the gap, leading to the decrease of V_{eff} for small ϕ. For large ϕ, the lattice-strain effects represented by the positive K term in H, dominate the interaction terms. Thus, within the harmonic fluctuation approximation about each ground state, the system appears to be a conventional semiconductor.

However, a closer look reveals the system to be highly unconventional, in that in addition to phonons, the stable excitations are not electrons and holes as in conventional semiconductors but topological solitons with peculiar charge–spin relations. To understand how this comes about, consider the displacement pattern shown in fig. 6 in which a domain wall is located at site 1 (fig. 6a). If the displacements were large, a pair of p_z electrons ($\uparrow\downarrow$) would be localized on each short bond and none would be localized on long bonds, just as in the classical model. When the electron pair localized on the bond joining 1 and 2 moves one unit

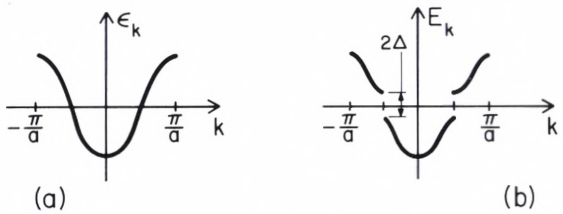

Fig. 5. The π electronic structure: (a) for $\langle\phi\rangle = 0$, and (b) for $\langle\phi\rangle = \pm\phi_0$. The gap 2Δ arises from and drives the spontaneous symmetry-breaking by negative energy states, being lowered when $\langle\phi\rangle \neq 0$.

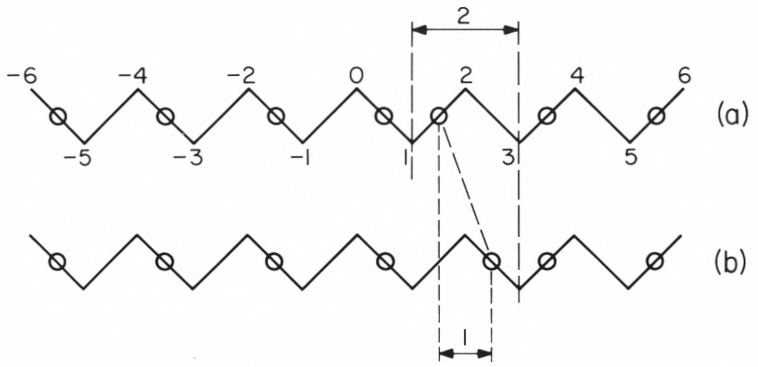

Fig. 6. The motion of a kink in trans-$(CH)_x$ is illustrated in analogy with the classical example shown in fig. 2. In practice, the kink width is on the order of fourteen lattice spacings rather than two spacings as shown, reflecting electrons being in scattering states rather than in the localized states used for this schematic illustration.

to the bond joining 2 and 3 as in fig. 6b, the domain wall moves from site 1 to site 3, i.e. two units. Therefore, we expect the effective charge of the kink to be $-e/2$ per spin direction, or a total charge of $-e$.

Another way of understanding this result is illustrated in fig. 7 in which N (even) (CH) groups form a ring. Initially the system is in the A phase, shown in panel (a) with ϕ plotted radially. The spectrum has $N/2$ levels in the positive and $N/2$ levels in the negative energy regions. In panel (b), ϕ is distorted to form a soliton S and an antisoliton $\bar{S}$. If S and $\bar{S}$ are widely separated, the electronic spectrum exhibits two zero-energy states, each split symmetrically from the positive and negative energy continua. Furthermore, since the spectral sum rule for the local density of

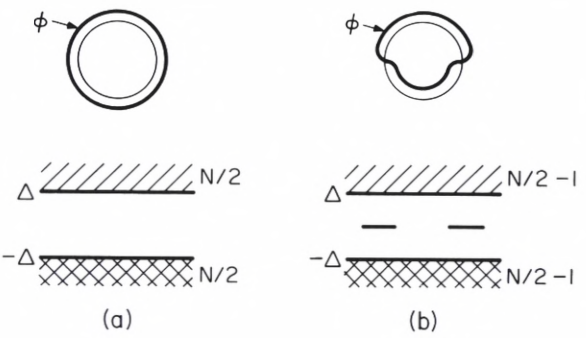

Fig. 7. A trans-$(CH)_x$ chain having N (CH) groups in a ring configuration. In (a) the system is in the A vacuum and in (b) the system is distorted to have a soliton S and a widely spaced antisoliton $\bar{S}$. While the positive and negative continua each have $N/2$ states in (a), two states near zero-energy appear in (b), one localized near S and the other near $\bar{S}$. Since the total number of states is conserved, the positive and negative energy continua are each depleted by one state when S and $\bar{S}$ are created. The essential point is that the state depletion (and hence charge depletion per spin) of the negative-energy states is $\frac{1}{2}$ from the vicinity of S and $\frac{1}{2}$ from $\bar{S}$. This is the origin of the fractional charge of these objects.

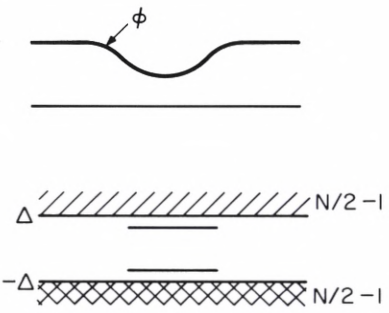

Fig. 8. A polaron P or bag state leading to two split off states in the gap. Note that P can be interpreted as a bound $S\bar{S}$ pair (see fig. 7).

states $\rho_{nn}(E)$ requires

$$\int_{-\infty}^{\infty} \rho_{nn}(E)\, \mathrm{d}E = 1, \tag{4}$$

we see that spectral weight for the zero-mode wavefunction ψ_0 for S (or $\bar{S}$) is stolen from the immediate vicinity of S (or $\bar{S}$). This depletion occurs symmetrically from the + and − energy continua since H is invariant under charge conjugation. Thus, the negative sea is depleted by a $\frac{1}{2}$ state per spin near S and the same holds true near $\bar{S}$. Clearly, this occurs by ϕ_n acting as a scattering potential which distorts or phase shifts the continuum states near S and $\bar{S}$.

In addition to soliton or kink states, trans-$(CH)_x$ supports polaron, P, or bag states [11,12]. In fig. 8, ϕ_n is shown having a dip which splits off two states, at $\pm\epsilon$, from the continuum. Only when the total number of electrons in $\pm\epsilon$ is either 1 or 3 is the bag stable. Three electrons in $\pm\epsilon$ corresponds to an electron polaron which has a charge $-e$ (since vacuum depletion leads to a charge $+2e$ when both $\pm\epsilon$ are empty) while 1 electron in $\pm\epsilon$ corresponds to a hole polaron, of charge $+e$. The charge and spin relations are summarized in table 1.

The doubly charged polarons P^{2+} and P^{2-} decay rapidly to $S^+\bar{S}^+$ or $S^-\bar{S}^-$. Thus, P is essentially a strongly bound $S\bar{S}$ pair.

Table 1
Charge (Q) and spin (s) relations in a conventional semiconductor.

Object	Symbol	Q	s
Electron	e^-	$-e$	$\frac{1}{2}$
Hole	e^+	$+e$	$\frac{1}{2}$
$(CH)_x$ soliton	S^-	$-e$	0
	S^0	0	$\frac{1}{2}$
	S^+	$+e$	0
Polaron	P^-	$-e$	$\frac{1}{2}$
	P^+	$+e$	$\frac{1}{2}$

Detailed calculations [2] predict that for trans-$(CH)_x$ the soliton width 2ξ is approximately 14 lattice sites or about 18 Å, and the soliton rest energy [13] is $E_S \cong (2/\pi)\Delta$, while the soliton mass M_s is approximately six electron mass units. As discussed below, many experiments confirm these general results [14], although electron–electron Coulomb interactions, quantum fluctuations of ϕ, etc., lead to some quantitative changes [9,15,16].

4. *A relativistic model*

Closely related to the $(CH)_x$ model is the relativistic model studied by Jackiw and Rebbi [1]. They considered a neutral scalar ϕ^4 field weakly coupled to a spinless Dirac field in one space dimension,

$$H = \int \mathrm{d}x \left\{ \psi^+(\alpha p + \beta g\phi)\psi + \tfrac{1}{2}\dot{\phi}^2 + \frac{1}{2}\left(\frac{\mathrm{d}\phi}{\mathrm{d}x}\right)^2 + V(\phi) \right\}, \tag{5}$$

where

$$V(\phi) = \frac{-a\phi^2}{2} + \frac{b\phi^4}{4}. \tag{6}$$

As shown in fig. 9, a kink occurs when ϕ moves from the $-\phi_0$ to the ϕ_0 vacuum. Treating ϕ as a background field, Jackiw and Rebbi found a zero mode ψ_0 which if empty leads to a fermion number $n_f = -\frac{1}{2}$ associated with the kink, while $n_f = +\frac{1}{2}$ if ψ_0 is occupied. Again, this is due to the flow of vacuum charge as negative-energy states distort at fixed occupation in response to the distortion of $\phi(x)$.

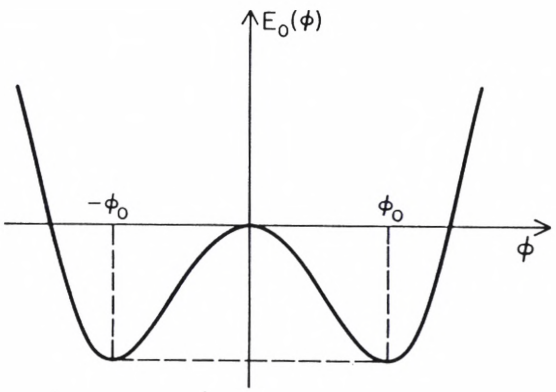

Fig. 9. The potential $V(\phi)$ for a ϕ^4 relativistic model. For $\langle\phi\rangle = \pm\phi_0$, a mass appears in the fermion spectrum. When a kink is introduced a zero mode occurs and the kink is fractionally charged. Precisely the same physical origin of fractional charge occurs here as in the $(CH)_x$ case.

5. Sharpness of fractional charge

The question has been raised if fractional charge is a sharp quantum observable or if only the expected value of Q is fractional. This has been answered by direct calculation [17] as follows. When one defines the charge of an extended object, it is the long-wavelength limit Q_f of the charge form-factor which is relevant. Thus, if $\rho(x)$ is the charge-density operator, we have

$$Q_f = \int f(x)\, \rho(x)\, dx, \tag{7}$$

where for convenience we choose a Gaussian sampling function

$$f(x) = e^{-x^2/L^2} \tag{8}$$

and let $L \to \infty$ for an infinite system. One can prove that the mean square fluctuations of Q_f about its mean value $\langle S | Q_f | S \rangle \equiv Q_s$ vanish exponentially as $L \to \infty$,

$$\langle S | (Q_f - Q_s)^2 | S \rangle - \langle 0 | Q_s^2 | 0 \rangle \to e^{-L/\xi}, \tag{9}$$

where ξ is the half-width of the soliton. Note that the fluctuations of the vacuum are subtracted since these are unrelated to soliton fluctuations. In any event the vacuum fluctuations also vanish as the sampling size $L \to \infty$. Thus, for smooth sampling of the charge, a sharp fractional charge occurs.

6. The experimental situation

While fractional charge of isolated objects (as opposed to bound quarks) has not been observed in particle physics, strong evidence exists that soliton excitations, as discussed above, do exist in quasi one-dimensional conductors. Heeger et al. observed the rapid transport of spin without charge in NMR and ESR experiments. For example, electron spin resonance experiments on prestine trans-$(CH)_x$ show that objects with spin $\frac{1}{2}$ are moving rapidly, near the speed of sound, rather than being localized on a given (CH) group. The solitons' motion is reflected as a motional narrowing of the spin resonance line width. At liquid helium temperatures, the width grows since the soliton presumably comes to rest, however, a residual motional narrowing occurs. This residual width is a direct measurement of the delocalization of the zero mode wavefunction ψ_0 over the size $2\xi \sim 14$ Å of the soliton.

In addition to showing S^0 and $\bar{S}^0$ have spin $\frac{1}{2}$, Heeger et al. showed that $S^{\pm}$ and $\bar{S}^{\pm}$ have spin 0, in agreement with table 1. Finally, polarons $P^{\pm}$ have been observed to have spin $\frac{1}{2}$. The charged soliton and the polarons are created by doping the

material with donor or acceptor impurities such as Na or Cl, etc., with one soliton being created per impurity atom.

Another elegant method of creating solitons in trans-$(CH)_x$ is by photoproduction of $S\bar{S}$ pairs. A photon of energy $\gtrsim 2\Delta$ is absorbed creating a bare electron–hole pair across the gap. These excitations are highly unstable and decay in one phonon period $\sim 10^{-13}$ s into an $S\bar{S}$ pair, some fraction of which separates near the speed of sound. Experiments [14] suggest that only charged solitons $S^{\pm}\bar{S}^{\mp}$ are formed in this way. Optical absorption of these photoproduced $S^{\pm}$ and $\bar{S}^{\mp}$ shows transitions between the top of the negative energy sea and the zero mode (called the gap state) or from this state to the positive energy sea. There is also evidence of direct $S\bar{S}$ production below the 2Δ threshold. In addition, the shape-oscillation mode of the soliton (in essence ξ oscillation) has been observed.

Another intriguing experiment by Dalton et al. [18] is electron nuclear double resonance (ENDOR) on prestine trans-$(CH)_x$, which probes the spacial distribution of spin in a neutral soliton. Their results are consistent with the soliton spin per site oscillating with a period of two lattice sites, as predicted by theory [19]. The reverse (down) spin between up-spin sites is due to Coulomb exchange effects. Remarkably, as the soliton moves, the up-spin sites remain up-spin and vice versa so that this distinction between up and down sites (or down and up sites for $\bar{S}^0$ with spin up) is preserved, precisely as predicted by theory.

There are many other experiments supporting the soliton model of $(CH)_x$, including transport properties, luminescence and photoconductivity, etc. [14] Finally, if the degeneracy of the minima of V_{eff} at $\pm\phi_0$ is split, one would predict a confinement potential $V_c(x) = -c|x|$ binding the $S\bar{S}$ pair. This is observed, for example, in cis-$(CH)_x$ where the A and B phases are not degenerate by symmetry, the chemical structure being square wave rather than zig zag as in the trans material.

7. *Quantum Hall effect*

A second mechanism leading to quantization of non-integer charge occurs in the so-called fractional quantum Hall effect. While the Hall effect historically was observed in three-dimensional materials, such as the semiconductors Ge and Si, recent interest focussed on systems such as MOSFET and heterojunction semiconductor devices, in which a two-dimensional (x, y) electron gas exists on a surface or at an interface. A very strong magnetic field $B_0 \simeq 10^5$ Oe is applied along z, as sketched in fig. 10. If an electric current j_x is made to flow along x, an electric (Hall) field E_y proportional to j_x is observed.

7.1. *One-electron theory*

The classical theory of the Hall effect is extremely simple. Each electron is acted on by the Lorentz force

$$\boldsymbol{F} = -e\left(\boldsymbol{E} + \frac{\boldsymbol{v}\times\boldsymbol{B}}{c}\right), \tag{10}$$

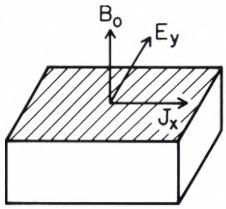

Fig. 10. Configuration for the two-dimensional Hall effect.

which vanishes in steady state. Thus, $v_x = cE_y/B_0$ and the current is given by

$$j_x = -nev_x = \left(\frac{-nec}{B_0}\right)E_y \equiv \sigma_{xy}E_y, \tag{11}$$

where σ_{xy} is the Hall conductivity. For fixed B_0, σ_{xy} is linear in the two-dimensional electron density n.

von Klitzing, Dorda and Pepper [4] discovered that for devices having extremely long electron relaxation times τ, so that $\omega_c\tau \gg 1$, with $\omega_c = eB/mc$ being the cyclotron frequency, σ_{xy} is not a linear function of n but has plateaus at "integer" values of n, corresponding to the complete filling of a Landau level, i.e. $n = \nu B_0/\varphi_0$ where $\nu = 1, 2, 3\ldots$ and $\varphi_0 = hc/e$ is Dirac's flux quantum. Since Planck's constant appears, at least a semiclassical theory is required. If one neglects electron–electron interactions, the system is described by electrons independently filling the Landau levels $E_l = (l + \frac{1}{2})\hbar\omega_c$, $l = 0, 1, 2\ldots$. Using the classical expression for σ_{xy} with n given by $\nu B_0/\phi_0$ for general ν, one finds

$$\sigma_{xy} = -\nu\frac{e^2}{h}. \tag{12}$$

Therefore σ_{xy} is predicted to directly measure e^2/h. At present the relative error in determining e^2/h by this method is on the order of 10^{-7}, comparable with the Josephson-effect measurement of this constant.

What produces the plateaus at $\nu = 1, 2, 3\ldots$? It is thought that for ν near an integer, the extra electrons (or holes) about integer occupation are localized in bound states, presumably due to crystal defects. Only when all of the localized trap states are full, do the extra electrons add to the current. Remarkably, sum rules show that the depletion of the continuum states to form bound states does not reduce the Hall conductivity for $\nu =$ integer [20].

Following the discovery of the integer effect, Tsui, Störmer and Gossard [5] discovered plateaus at fractional $\nu = p/q$, where q is an odd integer and p is integer. These experiments require even greater care than integer ν since the effects are quite subtle. While the gaps between one-electron Landau levels give a natural basis for accounting for the integer effect, it is clear that electron–electron interactions must be involved in the fractional ν effect.

7.2. Liquid-phase theory

Soon after the discovery of Tsui et al., Laughlin [6] proposed a trial wavefunction for the interacting electron gas in the form of a Jastrow-type state, i.e. a product of pair correlation factors f_{ij},

$$\Psi_0 = \left(\prod_{ij} f_{ij}\right) \exp\left(-\sum_k \frac{|Z_k|^2}{4l_0^2}\right), \tag{13}$$

where $Z_k = x_k + \mathrm{i}\, y_k$ is the complex coordinate of electron k, and l_0 is the magnetic length $\pi l_0^2 \equiv \varphi_0/B_0$. For $\hbar\omega_c \gg e^2/\epsilon l_0$ (ϵ = the dielectric constant), only the lowest Landau level $l = 0$ need be included in the many-body basis states. It is seen that Ψ_0 must be a polynomial in Z_j. Laughlin chose $f_{ij} = (Z_i - Z_j)^m$, where $m = 1, 3, 5\ldots$, so that Ψ_0 is properly antisymmetric. For m non integer, Ψ_0 cannot describe the system since higher Landau levels are admixed. The density is $\nu = 1/m$ for the Laughlin state.

To accommodate extra electrons or holes near $\nu = 1/m$, Laughlin proposed including a factor $\prod_{j'}(Z_{j'} - Z_0)$ for a quasi-particle centered at Z_0. This factor acts as a raising operator for the angular momentum of each particle about Z_0. One finds that the charge of the quasi-particle (hole in this case) is $Q = \nu e = e/m = e$, $\frac{1}{3}e$, $\frac{1}{5}e, \ldots,$ while a conjugate factor acting on Ψ_0 produces quasi-particles of negative fractional charge. Note that the quantization of charge arising in this case is not due to the discrete degeneracy of the ground state, but rather from the local quantization implied by the restriction of remaining in the lowest Landau level imposed by large B_0. The raising of all angular momenta about Z_0 by one unit depletes the region surrounding Z_0 by precisely ν electrons. Therefore, since no charge accumulation occurs away from the vicinity of Z_0, we see $Q = e/m$. However, Laughlin has argued that his state describes an incompressible fluid so that the charge νe swept out from the origin cannot be screened by polarization effects away from Z_0. Spin is not considered here since the spin Zeeman energy is also assumed large compared to $e^2/\epsilon l_0$. Another way of phrasing the issue is that the factor $\prod_{j'}(Z_{j'} - Z_0)$ represents a singular gauge transformation which arises from a conceptual point magnetic vortex tube threading the plane at Z_0. Only if the flux ϕ of the line is a multiple of ϕ_0 does the state remain in the lowest Landau level and has low energy.

Finally, we note that higher-order plateaus, $\nu = p/q$ where $p = 2, \ldots$ have been interpreted by Haldane [21] and by Halperin [22] as arising from Laughlin condensation of quasi-particles from a $\nu = 1/m$ state to form a new state which in turn has quasi-particles which condense, etc. This forms a hierarchy of p/q charged particles.

7.3. Quasi-particle statistics

To work with fractionally charged quasi-particles, one must know their statistic. Instinctively, since electrons gain a phase of π on interchange, it seems reasonable

that two quasi-particles i and j of charge νe would gain a phase $\nu\pi$ on interchange in two dimensions, i.e.

$$\Psi(i, j) = e^{i\pi\nu}\, \Psi(j, i). \tag{14}$$

That this is true in two dimensions was proved by Arovas, Wilczek and the author [23] using the extended adiabatic theorem [24,25]. Namely, as one adiabatically interchanges 1 and 2, the wavefunction picks up a phase $\gamma(t)$ given by

$$i\dot{\gamma} = \langle \Psi(t) \mid \dot{\Psi}(t) \rangle. \tag{15}$$

By using Laughlin's state for two quasi-particles i and j one finds the result given in eq. (14).

The free energy of a non-interacting gas of such quasi-particles is peculiar [26].

7.4. Crystalline-phase theory

While the liquid-state theory of Laughlin along with the Haldane–Halperin hierarchy give an explanation of much of the data on the fractional quantum Hall effect, a difficulty occurs. One can prove that the correct state at low electron density, $\nu \ll 1$ is a Wigner crystal in which particles are located on a triangular lattice in two dimensions. An important question is at what density ν_c do the crystal and liquid phases have equal energy? Starting with the Wigner crystal state, Maki and Zotos [27] calculated the vibration spectrum and found the crystal is differentially stable except in the range $0.45 < \nu < 0.55$. Since the Laughlin description of the liquid only holds near $\nu = \frac{1}{3}, \frac{1}{5}, \ldots$, the differential instability of the liquid has not been calculated. Presumably, the lower boundary is between $\nu_c = 0.2$ and 0.45 although this question remains open at present.

Recently, Kivelson, Kallin, Arovas and the author (KKAS) [7] proposed an alternative theory of the fractional Hall effect in terms of a collective ring-exchange mechanism starting from the crystalline phase. Using a path-integral formulation, they find important processes are those in which a ring having L electrons collectively tunnels to a new configuration, such that $Z_j \to Z_{j+1}$. Because flux is enclosed by the tunneling current, each ring enters with a phase given by ϕ/ϕ_0, where ϕ is the flux enclosed by the area A of the ring. For $\nu = p/q$ with p and q integer, these phases are such that exchange energy of large loops on average add in phase. However, for other values of ν, the crystal and magnetic lattices are incommensurate and the exchange energy of large rings is random in sign and cancels. Thus, the ground state energy $E_0(\nu)$ is found to exhibit cusps at $\nu_i = p/q$ with the cusp-like part of E_0 varying as $|\delta\nu| \ln |\delta\nu|$ near each cusp, where $\delta\nu = \nu - \nu_i$. The strength of the cusp is largest for $p = 1$ and q small, i.e. $\nu = \frac{1}{3}, \frac{1}{5}, \ldots$.

To determine $E_0(\nu)$ one writes

$$E_0(\nu) = -\lim_{\beta \to \infty} \frac{\partial}{\partial \beta} \ln Z, \quad Z \equiv \mathrm{Tr}\, e^{-\beta H}, \tag{16}$$

where the trace is over many-body states constructed solely from the lowest Landau level. The effective Hamiltonian is given by

$$H = P_0\Big(\sum_{i\neq j} V_{ij}\Big)P_0, \tag{17}$$

where P_0 projects onto the lowest Landau level. Roughly, V_{ij} is given by

$$V_{ij} = \frac{e^2}{\epsilon\big[\,|Z_i - Z_j|^2 + \lambda^2\big]^{1/2}}, \tag{18}$$

where λ is the extent of the one-electron orbitals in the z direction.

The so-called coherent states $|R\rangle$ form a convenient one-electron basis set for taking the trace in eq. (16). These states have wavefunctions

$$\langle \boldsymbol{r}\,|\,\boldsymbol{R}\rangle = \frac{1}{\sqrt{2\pi l_0^2}}\,\exp\!\left\{-\frac{1}{4l_0^2}\,|\boldsymbol{r}-\boldsymbol{R}|^2 + \frac{\mathrm{i}}{2l_0^2}(\boldsymbol{r}\times\boldsymbol{R})\cdot\hat{z}\right\}, \tag{19}$$

and satisfy the completeness relation

$$\int \frac{\mathrm{d}^2\boldsymbol{R}}{2\pi}\,|\boldsymbol{R}\rangle\langle\boldsymbol{R}| = P_0. \tag{20}$$

By writing $\mathrm{e}^{-\beta H} = (\mathrm{e}^{-\epsilon H})^M$ with $\epsilon = \beta/M$ and inserting between each factor of $\mathrm{e}^{-\epsilon H}$ the product of projectors P_0 from eq. (20) for all particles, one finds, as $M\to\infty$, a path-integral representation of Z:

$$Z = \mathcal{N}\sum_{\mathscr{P}} \operatorname{sign}\,\mathscr{P}\int \prod_{j=1}^{N} \mathscr{D}\boldsymbol{R}_j(\tau)\;\mathrm{e}^{-S[\boldsymbol{R}(\tau)]}, \tag{21}$$

with the boundary conditions $\boldsymbol{R}_j(0) = \boldsymbol{R}_{\mathscr{P}(j)}(\beta)$. Here $\mathscr{P}$ permutes the particle coordinates. The action S is given by

$$S[R] = \tfrac{1}{2}\int_0^\beta \mathrm{d}\tau\left\{-\mathrm{i}\sum_{j=1}^{N}\big(\dot{\boldsymbol{R}}_j\times\boldsymbol{R}_j\big)\cdot\hat{z} + \sum_{i\neq j} V_{\mathrm{c}}\big(\boldsymbol{R}_i - \boldsymbol{R}_j\big)\right\}, \tag{22}$$

where V_{c} is the effective Coulomb interaction eq. (18) in the coherent state basis.

To evaluate Z we consider the saddle point (semiclassical) approximation in which one continues $X_j(\tau)$ and $Y_j(\tau)$ to complex values, where $\boldsymbol{R}_j = (X_j,\, Y_j)$. One finds those paths $\{\boldsymbol{R}_{jc}(\tau)\}$ which minimize the classical action S. These extremal paths satisfy

$$\mathrm{i}\frac{\mathrm{d}X_j}{\mathrm{d}\tau} = \frac{\partial V_j}{\partial Y_j}, \qquad \mathrm{i}\frac{\mathrm{d}Y_j}{\mathrm{d}\tau} = -\frac{\partial V_j}{\partial X_j}. \tag{23}$$

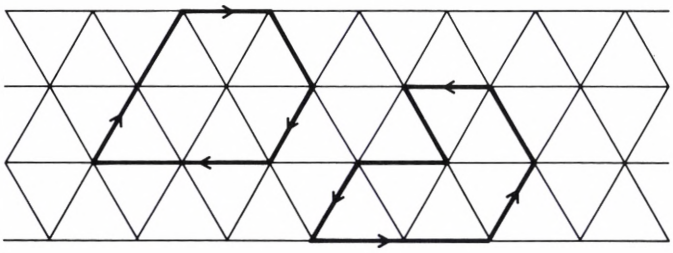

Fig. 11. Collective ring-exchange processes in which a set of L electrons simultaneously rotate so that each moving electron finally occupies the initial position of its moving nearest neighbor. Contributions from different rings add in phase only for the Landau-level filling factor $\nu = p/q$, where p and q are integers.

Then, Z is approximately given by the sum over external classical paths c:

$$Z = \sum_c D[R^c] \exp\{-S[R^c]\}, \tag{24}$$

where D is the fluctuation determinant arising from fluctuations quadratic in $\delta \boldsymbol{R}^c$ about each path. As is well known, tunneling processes are often well treated by such an approximation.

Using the above procedure, KKAS find an important set of paths are those in which a ring of L electrons tunnels from an initial configuration $\boldsymbol{R}^0$ to a final configuration $\boldsymbol{R}^f$ with $\boldsymbol{R}^f_j = \boldsymbol{R}^0_{j+1}$ or $\boldsymbol{R}^0_{j-1}$, corresponding to a clockwise or counterclockwise rotation of the ring, as illustrated in fig. 11. In this path, all tunneling electrons move simultaneously, staying out of each other's way as well as away from non-moving electrons. This path is in contrast to a sequence of conventional pairwise exchange processes in which electrons must surmount high Coulomb barriers produced by non-moving electrons. Thus, while ring exchange via pairwise exchange processes is very weak for interesting values of ν, the above collective ring exchange processes is considerably stronger, e.g. by a factor on the order of 10^4–10^5 even for a ring containing three electrons, for $\nu = \frac{1}{3}$.

For a large number of particles L in a ring of area A, one finds the contribution to Z in time $\mathrm{d}\tau$ is

$$Z_{LA} = \tau_0^{-1}\, \mathrm{d}\tau \exp\left[-\alpha(\nu) L + \mathrm{i} h N_A + O(\ln\, L)\right], \tag{25}$$

where $\alpha(\nu)$ is the tunneling exponent per particle, $h \equiv \pi(\nu^{-1} - 1)$ and $N_A = \nu A/\pi l_0^2$ is the number of Wigner-lattice unit cells contained in area A. The ν^{-1} term in h gives the flux contribution to the phase while the factor -1 accounts for the Pauli principle factor for even and odd L loops, since L and N_A are both even or both odd for a triangular lattice. Calculations show $\alpha\,(\frac{1}{3}) \simeq 0.81$.

While the contribution of large L rings is essential to obtain cusps, it might appear that large L is exponentially suppressed due to the tunneling factor $\mathrm{e}^{-\alpha L}$. However, path counting shows that the number of closed paths of length L varies as $\mathrm{e}^{\kappa L}$ so that aside from the phase question, the sum over rings would become large

for $\alpha(\nu) < \kappa$, i.e. when ν increases to the point where the tunneling is sufficiently probable that "path entropy" dominates "path energy". Thus, we expect nonanalytic structure in $E_0(\nu)$ when rings add in phase on the average and in addition ν is larger than a critical density ν_c, such that $\alpha(\nu_c) = \kappa$.

The ring summation is conveniently carried out by mapping the problem onto the discrete Gaussian spin model [28]:

$$H_{DG} = \alpha \sum_{(\lambda,\gamma)} (S_\lambda - S_\gamma)^2 + ih \sum_\lambda s_\lambda, \tag{26}$$

in an imaginary magnetic field h. As illustrated in fig. 12, the spins S_λ live on the dual lattice, where S_λ is defined as the number of clockwise minus the number of counterclockwise collective exchange-rings encircling the dual lattice point λ during the time interval τ_0 centered at time τ. It is helpful to think of the rings as domain boundaries separating regions having different values of S_λ. For $\nu \ll 1$, α is large and the ring density is low. In this case rings rarely overlap. However for larger ν, α decreases (tunneling increases) and the ring density increases. Since two rings cannot share a common edge during one time slice τ_0, a ring repulsion must be included. The $(S_\lambda - S_\gamma)^2$ factor leads to such a repulsion since by crossing two separated boundaries one would obtain a factor of $1^2 + 1^2 = 2$ in the energy while crossing a double boundary leads to $2^2 = 4$. While a power higher than 2 in eq. (26) might be preferable, the results are likely to be insensitive to this change.

From studies of the discrete Gaussian model [28] and related models [29], it is believed that $E_0(\nu)$ has cusps of the form $|\delta h| \ln(2\pi/|\delta h|)$ or $|\delta \nu| \ln|\delta \nu|$ at all rational $h/2\pi$ for $\alpha(\nu) < \alpha_c[h(\nu)]$. Estimates of α lead to cusps at densities $\nu = \frac{1}{3}$, $\frac{1}{5}$, $\frac{2}{5}$, $\frac{2}{7}$, $\frac{1}{4}$, $\frac{3}{7}$ and $\frac{4}{9}$ although some phases may be unstable with respect to competing phases. By charge-conjugation symmetry, cusps are also expected at the conjugate densities $1 - \nu$. We note that the crystal lattice approach produces the higher-order plateaus, e.g. $\nu = \frac{2}{5}$, $\frac{2}{7}$, $\frac{3}{7}$, $\frac{4}{9}$, without constructing a Haldane–Halperin hierarchy.

Finally, we turn to quasi-particle excitations in the collective ring-exchange scheme. As discussed above, the apparent reason for the plateaus in $\sigma_{xy}(\nu)$ at

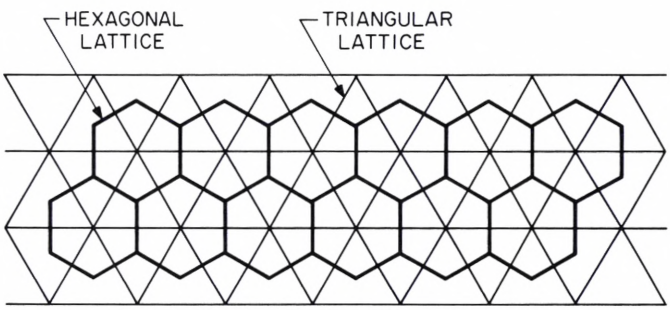

Fig. 12. The dual lattice for the pseudo-spins on to which the cooperative ring-exchange problem is mapped. The walls between domains having different spin orientation are the ring-exchange paths of the original problem.

fractional ν_i is the stability of the condensate for these particular densities. That is, there exists an energy gap for adding or subtracting charge about each fractional ν_i such that $\alpha(\nu) < \alpha_c[h(\nu)]$. To understand how such a gap comes about consider a uniform compression of the density, about a cusp at ν_i. The elastic modulus

$$K = \frac{\partial E_0}{\partial \nu}\bigg|_{\nu_i} \to \infty \tag{27}$$

so that the crystal is incompressible at the cusp. However, suppose we thread a point flux tube φ through the plane, say at a point Z_0 on the dual lattice. As in Laughlin's approach, this singular gauge transformation "blows a hole" in the electron density near Z_0 with charge being transported to the sample boundary. Direct calculation shows that the charge transport to the boundary is νe for $\varphi = \varphi_0$, the Dirac flux quantum. Furthermore, the quasi-particle charge is localized within the magnetic unit cell of size $\sim l_0$ at Z_0. For the opposite sign of φ, the electron density accumulates near Z_0 to form a quasi-particle of charge $-\nu e$.

A question remains: why choose $\varphi = \pm\varphi_0$? It is readily seen that if φ is not an integer multiple of φ_0, cooperative exchange rings encircling Z_0 will have phases which do not add coherently near the cusp densities ν_i. Therefore there will be a discontinuous energy increase as φ deviates from $\pm\varphi_0$. It is this phenomenon which leads to quantization of the fractional charge in this scheme. An estimate [7] of the quasi-particle energy is $\Delta \equiv E_{qp}(\nu) \sim 0.5\ \nu^2 e^2/\epsilon l_0$.

As in other descriptions, the plateaus of $\sigma_{xy}(\nu)$ presumably arise from added density $\delta\nu$, occupying localized defect states inside the gap $-\Delta < E < \Delta$.

8. Conclusion

While quantum mechanics tells us that accurate observations of an isolated elementary particle will always produce integer charge, excitations in systems or fields of integer charge can carry sharp fractional charge Q. The fractional part of Q generally arises from flow of charge in the vacuum, without producing added excitations. Thus, it is the negative-energy states which are deformed, not the occupation numbers which are changed, when fractional charge is formed. The sharp quantization of Q can arise from either a discretely degenerate broken symmetry vacuum, as in one-dimensional conductors or in field theoretic models, or by local energy constraints coupled with incompressible flow of vacuum charge to the boundaries, as in the fractional quantum Hall effect.

While direct observation of an isolated fractional charge in a medium remains for the future, strong evidence exists for the effects discussed above in quasi one-dimensional conductors such as $(CH)_x$.

I hope these ideas would have pleased Niels Bohr.

Acknowledgements

The author would like to acknowledge the hospitality of the Niels Bohr Institute on the occasion of the Niels Bohr Centenary and the support of the National Science Foundation Grant no. DMR82-16285.

References

[1] R. Jackiw and C. Rebbi, Phys. Rev. D13 (1976) 3398.
[2] W.P. Su, J.R. Schrieffer and A.J. Heeger, Phys. Rev. Lett. 42 (1979) 1698; Phys. Rev. B22 (1980) 2099.
[3] R. Jackiw and J.R. Schrieffer, Nucl. Phys. B190 (1981) 253.
[4] K. von Klitzing, G. Dorda and M. Pepper, Phys. Rev. Lett. 45 (1980) 494.
[5] D.C. Tsui, H.L. Störmer and A.C. Gossard, Phys. Rev. B25 (1982) 1405.
[6] R.B. Laughlin, Phys. Rev. Lett. B23 (1981) 5632.
[7] S. Kivelson, C. Kallin, D. Arovas and J.R. Schrieffer, Phys. Rev. Lett. 56 (1986) 873.
[8] W.P. Su and J.R. Schrieffer, Phys. Rev. Lett. 46 (1981) 738.
[9] For a review of solitons and fractional charge in quasi one-dimensional conductors, see Highlights of Condensed Matter Theory, Int. School of Physics, Enrico Fermi, Course LXXXIX eds F. Bassani, F. Fermi and M.P. Tozzi (North-Holland, Amsterdam, 1985) p. 300.
[10] R.E. Peierls, Quantum Theory of Solids (London, 1955) p. 108.
[11] W.P. Su and J.R. Schrieffer, Proc. Nat. Acad. Sci (USA) 77 (1980) 5626.
[12] D.K. Campbell, A.R. Bishop and K. Fesser, Phys. Rev. B26 (1982) 6862.
[13] H. Takayama, Y.R. Lin-Liu and K. Maki, Phys. Rev. B21 (1980) 2388.
[14] A.J. Heeger, Philos. Trans. Soc. London A314 (1985) 17.
[15] S. Kivelson, in: Solitons, Modern Problems in Condensed Matter Sciences, Vol. 17, eds V.E. Zakharov, V.L. Pokrovskii and S.E. Trullinger (North-Holland, Amsterdam, 1986) ch. 6.
[16] W.P. Su, Handbook on Conducting Polymers, ed. T. Skotheim, (Marcel Dekker, New York, 1985).
[17] S. Kivelson and J.R. Schrieffer, Phys. Rev. B24 (1982) 6447;
J.S. Bell and R. Rajaraman, Phys. Lett. 116B (1982) 151.
[18] H. Thomann, L.R. Dalton, Y. Tomkiewicz and N.S. Shiren, Phys. Rev. Lett. 50 (1983) 533.
[19] A.J. Heeger and J.R. Schrieffer, Solid. State Commun. 48 (1983) 207.
[20] T. Ando, Y. Matsumoto and Y. Uemmura, J. Phys. Soc. Jpn 39 (1975) 279.
[21] F.D.M. Haldane, Phys. Rev. Lett. 51 (1983) 605.
[22] B.I. Halperin, Phys. Rev. Lett. 62 (1984) 1583.
[23] D. Arovas, J.R. Schrieffer and F. Wilczek, Phys. Rev. Lett. 53 (1984) 722.
[24] N.V. Berry, Proc. Soc. London Ser. A392 (1984) 45.
[25] B. Simon, Phys. Rev. Lett. 51 (1983) 2167.
[26] D. Arovas, J.R. Schrieffer, F. Wilczek, A. Zee, Nucl. Phys. B250 [FS13] (1985) 117.
[27] K. Maki and X. Zotos, Phys. Rev. B28 (1983) 4349.
[28] S.T. Chui and J.D. Weeks, Phys. Rev. B14 (1976) 4978.
[29] W.Y. Shih and D. Stroud, Phys. Rev. B32 (1985) 158.

Discussion, session chairman W. Kohn

Anderson: Is it obvious to you that the two representations of the fractional quantum Hall effect state, namely that of Laughlin and of KKAS are different? Or do they describe the same state in different ways?

Schrieffer: At present we have not been able to find a direct link between these two candidates for the ground state. Laughlin's is based on a fluid state while KKAS starts with a Wigner-crystal state. Thus, if one calculates the density correlation-function for the two states one would expect to find strong crystalline order, at least for short-range correlations in KKAS but not in the Laughlin fluid phase. Collective ring exchange weakens the crystallinity and long-wavelength fluctuation will smear out long-range crystalline behaviour.

Kohn: In your lecture you did not mention the fact that, associated with the quantum Hall effect, there is the occurrence of essentially vanishing resistance. Can you comment on this aspect of the quantum Hall effect within the framework of your theory?

Schrieffer: We are in the midst of attempting to understand this problem. In essence, the cusp nature of the energy leads to the incompressibility of the system at the densities ν_i. This in turn apparently produces a gap in the collective excitation spectrum, leading to activated resistance, as is observed.

Kohn: What about the effect of imperfections?

Schrieffer: We have not put these effects in at present. We believe it will turn out that, like in superconductors, the flow simply adjusts to the impurity and goes around it.