J. Phys. A: Math. Gen. **34** (2001) L119–L125 www.iop.org/Journals/ja PII: S0305-4470(01)20978-0

LETTER TO THE EDITOR

Fractionalization and confinement in the $U(1)$ and Z_2 gauge theories of strongly correlated systems

T Senthil and Matthew P A Fisher

Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106-4030, USA

Received 10 January 2001

Abstract

Recently, we have elucidated the physics of electron fractionalization in strongly interacting electron systems using a Z_2 gauge theory formulation. Here we discuss the connection with the earlier $U(1)$ gauge theory approaches based on the slave boson mean field theory. In particular, we identify the relationship between the holons and Spinons of the slave-boson theory and the true physical excitations of the fractionalized phases that are readily described in the Z_2 approach.

PACS numbers: 7510N, 0550, 7510J

The tantalizing possibility that the electron is fractionalized (i.e. broken apart) is a crucial ingredient of several qualitative or semi-quantitative pictures [1–4] of the physics of the cuprate high-T_c materials. It is extremely important to have a firm theoretical understanding of the physics of electron fractionalization before a complete theory based on such ideas can be developed for the cuprates. This Letter addresses and resolves some theoretical issues that arise in this context.

A popular model Hamiltonian that is often used as the starting point of discussions of the physics of the cuprate materials is the t–J model:

$$H = H_0 + H_J \tag{1}$$

$$H_0 = -t \sum_{<rr'>} \left(c^\dagger_{r\alpha} c_{r'\alpha} + \text{h.c.} \right) - \mu \sum_r N_r \tag{2}$$

$$H_J = J \sum_{<rr'>} \left(\vec{S}_r \cdot \vec{S}_{r'} - \frac{1}{4} N_r N_{r'} \right) \tag{3}$$

with the constraint of no double occupancy of any site, i.e. $N_r = \sum_\alpha c^\dagger_{r\alpha} c_{r\alpha} \leqslant 1$. Here the $c_{r\alpha}$ are electron operators at site r with spin polarization α, and $\vec{S}_r = \frac{1}{2} c^\dagger_r \vec{\sigma} c_r$ is the usual spin operator.

Equally popular, though not very succesful, are attempts [3, 5–7] to describe the possibility of electron fractionalization in the phases of the t–J model using the 'slave' boson representation

$$c_{r\alpha} = h^\dagger_r s_{r\alpha} \tag{4}$$

where h_r creates a hole at site r and is bosonic (dubbed the 'holon'), and $s_{r\alpha}^{\dagger}$ creates a spinful fermion at site r (dubbed the 'Spinon' - note the uppercase 'S'). It is assumed that the holon carries all the electrical charge while the Spinon carries all the spin of the electron. If the 'slave' particles are liberated from each other, then fractionalization is achieved. There are several problems with this program, however, and despite more than a decade of efforts it has not led to a clear description of fractionalized phases. In particular, as is well known, the slave boson representation induces a local internal $U(1)$ gauge symmetry under which

$$h_r \to h_r \mathrm{e}^{\mathrm{i}\theta_r} \qquad s_{r\alpha} \to s_{r\alpha} \mathrm{e}^{\mathrm{i}\theta_r}. \tag{5}$$

As a result there are strong interactions between the slave particles mediated by a compact $U(1)$ gauge field. The properties of the resultant strongly coupled gauge theory are extremely difficult to reliably analyse. It is generally believed [8][1], however, that fluctuations of the gauge field inevitably leads to a confinement of the slave particles —effectively recombining them to form the electron. This precludes any hopes of describing fractionalized phases, within which the holons and Spinons can propagate as independent excitations.

In an alternative approach [10], we have recently demonstrated that a general class of strongly interacting electron models can be recast in the form of a discrete Z_2 gauge theory. This new formulation enables a definitive theoretical characterization of electron fractionalization. Specifically, we demonstrated the possibility of obtaining fractionalized phases in two or higher spatial dimensions. In such a phase, the electron splits into two independant excitations—the spin of the electron is carried by a neutral fermionic excitation (the 'spinon'—note the lowercase 's') and the charge is carried by a bosonic excitation (the 'chargon'). There is a third distinct excitation, namely the flux of the Z_2 gauge field (dubbed the 'vison'). The vison is gapped in the fractionalized phase. The Z_2 gauge theory approach is closely related, and is indeed mathematically equivalent, to the ideas on vortex pairing [11] as a means to achieve fractionalization in two spatial dimensions.

In view of the popularity of the slave boson $U(1)$ gauge theory approach, it seems worthwhile to understand its connections, if any, with the physics of the fractionalized phases discussed in [10]. In particular, a natural question is how the holon and Spinon operators introduced above are related to the physical excitations (chargons, spinons, and visons) of the fractionalized phases. In this Letter, we will expose this connection.

To this end, consider first the slave boson representation of the t–J model. The Hamiltonian (1) may readily be rewritten in terms of the operators h_r and s_r, provided it is supplemented with the local constraints

$$h_r^{\dagger} h_r + \sum_{\alpha} s_{r\alpha}^{\dagger} s_{r\alpha} = 1 \tag{6}$$

at each and every site of the lattice. The Hamiltonian is then invariant under (i) the global (electromagnetic) $U(1)$ transformation $h_r \to h_r \mathrm{e}^{\mathrm{i}\phi}$, (ii) global $SU(2)$ spin rotations $s_r \to U s_r$ with $U \in SU(2)$, and (iii) the local internal gauge transformation equation (5). Note that the constraint (6) implies that the generator of the internal $U(1)$ gauge symmetry is fixed to be one at each lattice site.

After expressing the partition function as a functional integral, the theory proceeds by decoupling terms quartic in h and s with complex Hubbard–Stratonovich fields. The resulting action takes the form

$$S[h, s, a_0, \chi, \eta] = S_\tau + S_{\chi\eta} + S_h + S_s \tag{7}$$

[1] See [9] for a disputed argument with the same conclusions.

where

$$S_\tau = \int_\tau \sum_r \bar{h}_r(\partial_\tau - \mathrm{i}a_{0r})h_r + \mu\bar{h}_r h_r + \bar{s}_{r\alpha}(\partial_\tau - \mathrm{i}a_{0r})s_{r\alpha}$$

$$S_{\chi\eta} = 2J \int_\tau \sum_{<rr'>} \left(|\chi_{rr'}|^2 + |\eta_{rr'}|^2\right)$$

$$S_h = -t \int_\tau \sum_{<rr'>} \chi_{rr'}\bar{h}_r h_{r'} + \mathrm{c.c.} + O(h^4)$$

$$S_s = -J \int_\tau \sum_{<rr'>} \chi_{rr'}\bar{s}_r s_{r'} - \eta_{rr'}(s_{r\uparrow}s_{r'\downarrow} - s_{r\downarrow}s_{r'\uparrow}) + \mathrm{c.c.}.$$

Here $\chi_{rr'}(\tau)$, $\eta_{rr'}(\tau)$ are complex decoupling fields living on the links of the spatial lattice, and the field a_{0r} is a Lagrange multiplier imposing the on-site constraint in (6).

The properties of this action have been examined [12] within a mean-field approximation. Of particular interest to us is the mean-field solution found when the deviation x from half-filling is not too large: $<\chi_{rr'}>=\chi_0$, $<\eta_{rr'}>=\Delta_0\alpha_{rr'}$ with χ_0 and Δ_0 being real constants, and $\alpha_{rr'} = +1$ on horizontal bonds and -1 on vertical bonds. Within this mean field solution, the holons propagate freely with hopping $t_h = t\chi_0$. The Spinons also propagate freely but are paired into a $d_{x^2-y^2}$ condensate. It is hoped that this mean field solution correctly describes the 'd-wave RVB' state proposed pictorially by Anderson [1] and Kivelson *et al* [2], provided the holons are not condensed. This mean field state also has, at first sight, several properties in common with the nodal liquid state [10, 11, 14]. The electron has apparently been broken apart into the holon and the Spinon.

The crucial conceptual question is whether the fluctuations about the mean field solution invalidate its qualitative features. To discuss the fluctuations, note that the mean field solution breaks the internal $U(1)$ gauge symmetry. It is important therefore to keep two kinds of fluctuations: (i) fluctuations in the phase of χ—these become the spatial components of a $U(1)$ gauge field [8], and (ii) fluctuations in the phase of the Spinon pair condensate [13]. To capture the latter, we introduce a d-wave Spinon pair field $\mathrm{e}^{\mathrm{i}\varphi_r^{sp}}$ that couples to a $d_{x^2-y^2}$ pair of Spinons centred at site r. Upon returning to a Hamiltonian description, the fluctuations about the mean field state can then be described in terms of a simple effective Hamiltonian:

$$H_{eff} = H_{hol} + H_s + H_{pair} \tag{8}$$

$$H_{hol} = -t_h \sum_{<rr'>} \mathrm{e}^{\mathrm{i}a_{rr'}} h_r^\dagger h_{r'} + \mathrm{h.c.} + \mu \sum_r n_{hr} \tag{9}$$

$$H_s = -t_s \sum_{<rr'>} \mathrm{e}^{\mathrm{i}a_{rr'}} s_{r\alpha}^\dagger s_{r'\alpha} + \mathrm{h.c.} \tag{10}$$

$$H_{pair} = \Delta_0 \sum_r \left[\mathrm{e}^{\mathrm{i}\varphi_r^{sp}} p_r + \mathrm{h.c.}\right] \tag{11}$$

$$p_r = \sum_{r'}{}' \alpha_{rr'}\mathrm{e}^{\mathrm{i}a_{rr'}} \left(s_{r\uparrow}s_{r'\downarrow} - s_{r\downarrow}s_{r'\uparrow}\right). \tag{12}$$

Here in the last equation r' is nearest neighbour to r, and $t_s = J\chi_0$. For technical reasons we have used a number-phase representation of the holon operator: $h_r = \mathrm{e}^{\mathrm{i}\phi_{hr}}$, with

$$[\phi_{hr}, n_{hr'}] = \mathrm{i}\delta_{rr'} \tag{13}$$

where n_{hr} is the holon number operator, corresponding physically to the hole density.

As required, H_{eff} is invariant under the internal $U(1)$ gauge transformation

$$h_r \to \mathrm{e}^{\mathrm{i}\theta_r} h_r \qquad s_{r\alpha} \to \mathrm{e}^{\mathrm{i}\theta_r} s_{r\alpha} \tag{14}$$

$$a_{rr'} \to a_{rr'} + (\theta_r - \theta_{r'}) \tag{15}$$
$$\mathrm{e}^{\mathrm{i}\varphi_r^{sp}} \to \mathrm{e}^{\mathrm{i}(\varphi_r^{sp}+2\theta_r)}. \tag{16}$$

As expected the Spinon pair field operator $\mathrm{e}^{\mathrm{i}\varphi_r^{sp}}$ creates an excitation with two units of internal $U(1)$ gauge charge. The operators $s_{r\alpha}$ create unpaired Spinons—the analog of BCS quasiparticles for this Spinon pair condensate. Note that the number of unpaired Spinons $s_{r\alpha}^\dagger s_{r\alpha}$ is *not* conserved.

The effective Hamiltonian above must be supplemented with the constraint in equation (6) that the generator of the internal $U(1)$ gauge transformation equal one at each lattice site. The total internal $U(1)$ gauge charge is given by

$$n_{int}(r) = n_{hr} + 2n_r^{sp} + \sum_\alpha s_{r\alpha}^\dagger s_{r\alpha}. \tag{17}$$

Here we have defined a number operator for the Spinon pairs which satisfies $[\varphi_r^{sp}, n_{r'}^{sp}] = \mathrm{i}\delta_{rr'}$, and commutes with the h, s operators. The constraint is therefore

$$n_{int}(r) = 1. \tag{18}$$

Our goal is to recast this Hamiltonian in terms of (physical) charge and spin operators which are invariant under the internal $U(1)$ gauge transformation—the physical motivation for doing so is that it is expected that only particles which do not carry this internal $U(1)$ charge are expected to survive the strong confining effects of the interaction with the $U(1)$ gauge field. To that end, we follow closely the procedure introduced in [10] to deal with *electron* Hamiltonians with structure similar to that of H_{eff} above. We first split the Spinon pair creation operator into two equal pieces:

$$\left(b_{sp,r}^\dagger\right)^2 \equiv \mathrm{e}^{\mathrm{i}\varphi_r^{sp}} \tag{19}$$
$$b_{sp,r}^\dagger = e_r \mathrm{e}^{\frac{1}{2}\mathrm{i}\varphi_r^{sp}} \equiv \mathrm{e}^{\mathrm{i}\phi_r^{sp}}. \tag{20}$$

Here $e_r = \pm 1$, and $b_{sp,r}^\dagger$ creates *half* a Spinon pair. It is readily seen that the phase ϕ_r^{sp} is conjugate to $n_{int}(r)$:

$$[\phi_r^{sp}, n_{int}(r')] = \mathrm{i}\delta_{rr'}. \tag{21}$$

Note that the constraint $n_{int}(r) = 1$ implies that the conjugate phase ϕ_r^{sp} fluctuates wildly. In particular, it precludes any breaking of the internal $U(1)$ gauge symmetry.

We next define new operators invariant under the internal $U(1)$ gauge transformation by binding half the Spinon pair to the holon and the Spinon:

$$h_r = b_r^{sp} b_r^\dagger \qquad s_r = b_r^{sp} f_r. \tag{22}$$

We will denote the operator b_r the chargon[2], and the operator f_r the spinon (note the lowercase 's'). The reason for this terminology is that, as we show below, these correspond precisely to the operators with the same names introduced in [10], and are indeed the physical excitations in the fractionalized phase. Writing the chargon operator as $b_r = \mathrm{e}^{\mathrm{i}\phi_r}$, we see that the phase ϕ_r is conjugate to the hole density n_{hr}: $[\phi_r, n_{hr}] = -i$ [Note 3]. Further the spinon number $f_{r\alpha}^\dagger f_{r\alpha}$ is equal to the number of unpaired Spinons $s_{r\alpha}^\dagger s_{r\alpha}$.

Our plan is to now make a change of variables in H_{eff}, trading in the holon (h), Spinon (s_α) and Spinon pair operators ($\mathrm{e}^{\mathrm{i}\varphi^{sp}}$) in favour of the chargon (b), the spinon (f_α) and the half-Spinon pair operators ($\mathrm{e}^{\mathrm{i}\phi^{sp}}$). It is important to emphasize that of these three new operators, both the chargon and the spinon are invariant under the internal $U(1)$ gauge symmetry—and

[2] A similar definition of $U(1)$ gauge invariant chargon operators was also employed in [15].

[3] The minus sign in the commutator is a consequence of the particular definition of the chargon operator. It actually ensures that the commutator of the chargon phase with the physical electrical charge $N_r = 1 - n_{hr}$ has the usual sign.

all the internal $U(1)$ charge is carried by the operator b_r^{sp}. However, as elucidated in our earlier work [10], the very process of splitting the Spinon pair operator into two pieces introduces a *new* gauge symmetry—a Z_2 gauge symmetry. Specifically, the Hamiltonian H_{eff}, when re-expressed in terms of the three new operators, is invariant under the local transformation:

$$b_r^{sp} \to -b_r^{sp} \qquad b_r \to -b_r \qquad f_r \to -f_r \tag{23}$$

at any given site r. Moreover, it is necessary to impose a (new) constraint [10] on the Hilbert space of these three new operators, so that there is a one-to-one correspondence with the Hilbert space of the original three operators. The precise form of the new constraint in this case is

$$n_{int}(r) - n_{hr} - f_{r\alpha}^{\dagger} f_{r\alpha} = \text{even}. \tag{24}$$

Note also that the hole density $n_{hr} = 1 - N_r$ where N_r, the total electrical charge at site r, is conjugate to the chargon phase: $[\phi_r, N_{r'}] = \mathrm{i}\delta_{rr'}$. Thus, upon using the earlier constraint condition $n_{int}(r) = 1$, equation (24) reduces to

$$N_r - f_{r\alpha}^{\dagger} f_{r\alpha} = \text{even}. \tag{25}$$

This is exactly the same constraint on the chargon and spinon numbers as in [10].

Thus, we may now readily obtain a path-integral expression for the partition function of H_{eff} (in terms of b^{sp}, b and f_α) exactly as in [10]. The Z_2 constraint above can be implemented by means of a projection operator $\mathcal{P}_r = \frac{1}{2}[1 + (-1)^{N_r - f_r^{\dagger} f_r}]$ at each lattice site. The $U(1)$ constraint $n_{int}(r) = 1$ is more conveniently implemented by adding to H_{eff} the term $U \sum_r (n_{int}(r) - 1)^2$ and letting U go to infinity. As detailed in [10], the final result may essentially be written down on symmetry grounds—which in the present context are the $U_{int}(1) \times Z_2$ gauge symmetries, in addition to the global charge $U(1)$ and spin $SU(2)$ symmetries. The final action for the path-integral (for large finite U) takes the form

$$S = S_{z2g} + S_{sp} \tag{26}$$

$$S_{z2g} = S_c + S_s + S_B \tag{27}$$

$$S_{sp} = -\sum_{<ij>} t_{sp}^{ij} \sigma_{ij} \cos(\phi_i^{sp} - \phi_j^{sp} + a_{ij}). \tag{28}$$

Here S_{z2g} is exactly the Z_2 gauge theory action in [10], and describes chargons and spinons minimally coupled to a fluctuating gauge field σ_{ij}. (The indices i, j label the sites of a space-time lattice). Note that all the coupling to the internal $U(1)$ gauge field is through the field ϕ_i^{sp} as expected. The Z_2 gauge field σ_{ij} couples together the ϕ^{sp} with the chargons and the spinons. But note that we may absorb the σ_{ij} into the a_{ij} by shifting:

$$a_{ij} \to a_{ij} + \frac{\pi}{2}(1 - \sigma_{ij}). \tag{29}$$

Then the action for ϕ^{sp} simply becomes,

$$S_{sp} = -\sum_{<ij>} t_{sp}^{ij} \cos(\phi_i^{sp} - \phi_j^{sp} + a_{ij}) \tag{30}$$

and is then completely de-coupled from the Z_2 gauge theory. We may then integrate out the ϕ_{sp} and a_{ij} fields without affecting the rest of the action, and this may be done for any U including the limit $U \to \infty$. We thereby obtain the Z_2 gauge theory of [10], consisting of chargon and spinons minimally coupled to the Z_2 gauge field.

Having derived the Z_2 gauge action from the theory of fluctuations about the slave boson mean field theory, we may now directly take over the discussion of fractionalization from [10]. In particular, it is the chargon and spinon (lower case) fields (b and f_α) which create the physical excitations in a fractionalized phase, and not the holons and Spinons of the slave boson theory. The latter carry an internal $U(1)$ gauge charge, and so are susceptible to the confining effects of the compact $U(1)$ gauge field. The chargons (spinons) are obtained from these by binding half a Spinon pair to the holon (Spinon), thereby neutralizing their internal $U(1)$ charge. Instead the chargons and spinons are coupled to a Z_2 gauge field which allows them to be deconfined in two or higher spatial dimensions [10].

The discussion in this Letter can also be directly taken over to clarify some cryptic (though correct) remarks in the literature on the possibility of deconfined spin-$\frac{1}{2}$ excitations in Heisenberg spin models [16, 17]. These works start with, for instance, the Schwinger boson representation of the Heisenberg spins (which introduces a $U(1)$ gauge symmetry), and propose obtaining fractionalized phases by condensing pairs of Spinons, thereby reducing the gauge symmetry down to Z_2. However, with the constraint that the number of Schwinger bosons at each site is fixed, it would seem, at first sight, that breaking of the $U(1)$ gauge symmetry is prohibited. Nevertheless, the construction given in this Letter shows how one might get fractionalization without actually breaking the $U(1)$ gauge symmetry. The resulting fractionalized phases are then identical to those obtained by imagining that the Spinon pair field has condensed [18].

Before concluding, we emphasize some important differences between the physical pictures for the under-doped cuprates suggested by the $U(1)$ [3, 5] and Z_2 [4] gauge theory approaches. In the $U(1)$ approach, the pseudogap line is associated with the pairing of Spinons. The $U(1)$ theory is, at present, not powerful enough to unambiguously identify the true physical excitations of the system below this temperature scale as a result of the strong interactions with the $U(1)$ gauge field. By contrast, in the Z_2 approach the spinons (which *are* physical excitations in a fractionalized phase) are *always* paired. The pseudogap line is associated with the gapping out of the vison excitations [4]—the vortex-like excitations in the Z_2 gauge field. Once the visons are gapped out, the spinons and chargons are liberated from each other and are the legitimate excitations of the system.

We thank Chetan Nayak for provoking us to think through the contents of this Letter. We particularly thank Yong-Baek Kim for his insistence that we publish these results, and for several useful comments, and S Sachdev for his very constructive criticism and clarifying discussions on [16]. This research was generously supported by the NSF under Grants DMR-97-04005, DMR95-28578 and PHY94-07194.

References

[1] Anderson P W 1987 *Science* **235** 1196
[2] Kivelson S, Rokhsar D S and Sethna J. 1987 *Phys. Rev.* B **35** 8865
[3] Lee P A and Nagaosa N 1992 *Phys. Rev.* **45** 966
[4] Senthil T and Fisher M P A 1999 *Preprint* cond-mat/9912380
[5] Baskaran G, Zou Z and Anderson P W 1987 *Solid State Commun.* **63** 873
[6] Baskaran G and Anderson P W 1988 *Phys. Rev.* B **37** 580
[7] Affleck I and Marston J B 1988 *Phys. Rev.* B **37** 3774
Ioffe L and Larkin A 1989 *Phys. Rev.* B **39** 8988
Lee P A , Nagaosa N, Ng T-K and Wen X G 1998 *Phys. Rev.* B **57** 6003
[8] Read N and Sachdev S 1989 *Phys. Rev. Lett.* **62** 1694
(Read N and Sachdev S 1990 *Phys. Rev.* B **42** 4568)
[9] Nayak C 1999 *Preprint* cond-mat/9912270 (*Phys. Rev. Lett.* at press)
[10] Senthil T and Fisher M P A 2000 *Phys. Rev.* B **62** 7850
(Senthil T and Fisher M P A 1999 *Preprint* cond-mat/9910224)
[11] Balents L, Fisher M P A and Nayak C 1999 *Phys. Rev.* B **60** 1654
Balents L, Fisher and Nayak C 1999 *Preprint* cond-mat/9903294
[12] G. Kotliar G and Liu J 1988 *Phys. Rev.* B **38** 5142

[13] Lee D H 2000 *Phys. Rev. Lett.* **84** 2694
[14] Balents L, Fisher M P L and Nayak C 1998 *Int. J. Mod. Phys.* B **12** 1033
[15] Ye J and S Sachdev S 1991 *Phys. Rev.* B **44** 10173
[16] Read N and Sachdev S 1991 *Phys. Rev. Lett.* **66** 1773
Read N and Sachdev S 1991 *Int. J. Mod. Phys.* B **5** 219
[17] Wen X G *Phys. Rev.* B **44** 2664 1991
See also related work by
Mudry C and Fradkin E 1994 *Phys. Rev.* B **49** 5200
[18] Fradkin E and Shenker S 1979 *Phys. Rev.* D **19** 3682