J. Phys. A: Math. Gen. 16 (1983) 2049–2062. Printed in Great Britain

Some exact results for a model of Peierls dimerisation and solitonic excitations with non-integral charge

B Sriram Shastry

Tate Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Received 29 November 1982

Abstract. Nonlinear difference equations are derived for a model of Peierls dimerisation and non-integral charges proposed by Su, Schrieffer and Heeger. The dimerised state is shown to satisfy these exactly. A continuum model is obtained from the discrete model by approximating the fermion spectrum and imposing a cut-off to retain the correct number of degrees of freedom. The continuum model with periodic boundary conditions is solved exactly for an arbitrary number of solitonic excitations. The possibility of recovering solutions of the discrete model are discussed, using the solution to study completeness properties.

1. Introduction

There has been considerable interest recently in the properties of linear chain Peierls systems. In the context of polyacetylene it has been proposed (Su *et al* 1979, 1980, Rice 1979) that the low-lying excitations of interest in the system are topological solitons which interpolate between distinctly dimerised ground states. In the context of relativistic field theories, similar excitations have been found with fractional charges (for a review see Jackiw and Schrieffer 1981).

The work of Su *et al* (1979, 1980) is based on a numerical solution of finite sized discrete chains. Subsequently Takayama *et al* (1980) and Brazovski (1980) studied a continuum version of the model and succeeded in solving the latter problem analytically in the presence of a single soliton. A generalisation to the case of many solitons has appeared recently† (Horovitz 1981) which uses an analogy with the quantum sine-Gordon field theory.

In the present work we obtain exact nonlinear difference equations for the fermion amplitudes. These are shown to be satisfied by the Peierls dimerised state. We have not seen a demonstration of this widely accepted result in the literature and hence have provided one in the appendix. These equations are very hard to solve analytically for other excitations, and hence we take a continuum limit appropriate to the half filled case. This procedure introduces an infinite number of states not originally present. The state space of the discrete model. is then embedded in the larger space. It is important to understand the nature of this embedding in comparing the results of discrete and continuum models.

† The paper of Horovitz (1981) came to our attention after we had completed the major part of this work. The model is similar but the motivation and technical details seem different.

We have found it possible to solve the nonlinear equations in the continuum limit exactly. The solution is based on an ansatz for the 'order parameter', which in turn is guided by an approximate solution of the nonlinear equation following an idea of Sutherland (unpublished). The ansatz is shown to be self-consistent. The situation has a striking resemblance to the recent exact solution of a nonlinear eigenvalue problem (Shastry 1983). The present solution is for the case of periodic boundary conditions, which are particularly useful in counting the number of states for a large but finite system. The counting problem is crucial to an understanding of the embedding referred to above.

In § 2 we derive the nonlinear discrete equations for the model proposed by Su *et al* (1979). A continuum limit is taken which is appropriate for the half filled case and the continuum model is stated carefully.

In § 3 this model is solved within an approximate scheme which yields a functional form for the order parameter. In § 4 the ansatz is made and is shown to be self-consistent. Thus the model is solved exactly. In § 5 we highlight the salient features of the solution and comment on these in § 6. For simplicity we retrict to a low density of solitons in §§ 5 and 6 although our general solution is valid for an arbitrary density.

2. Nonlinear equations

We consider a model of N_e spinless fermions on a chain of N_a sites interacting with the lattice displacements u_i in a model proposed by Su *et al* (1980: referred to as SSH). The Hamiltonian is

$$\mathcal{H}' = -\sum_i [t_0 - \alpha(u_{i+1} - u_i)](c_{i+1}^\dagger c_i + \text{HC}) + \tfrac{1}{2}K \sum_i (u_{i+1} - u_i)^2 \tag{2.1}$$

where the c_i's are fermion destruction operators, t_0 the hopping parameter, α the electron–phonon coupling constant and K the elastic stiffness. The lattice is treated within the adiabatic approximation in (2.1), and periodic boundary conditions are imposed ($c_{i+N_a} \equiv c_i$, $u_{i+N_a} \equiv u_i$). We transform (2.1) into a more convenient form by dividing by t_0 and introducing the variables

$$v_i = \alpha/t_0(u_{i+1} - u_i), \qquad \gamma = \alpha^2/Kt_0. \tag{2.2}$$

Thus

$$\mathcal{H} = \sum_i [v_i - 1](c_{i+1}^\dagger c_i + \text{HC}) + (2\gamma)^{-1} \sum_i v_i^2. \tag{2.3}$$

The total energy is obtained from

$$E_{\text{total}} = \sum_{\alpha \in \text{Occ}} \varepsilon_\alpha + (2\gamma)^{-1} \sum_i v_i^2. \tag{2.4}$$

Here ε_α are eigenvalues of the fermion part and the summation is over N_e 'occupied' eigenvalues. The wavefunctions are obtained by expanding as follows:

$$|\psi_\alpha\rangle = \sum_i \psi_i^\alpha c_i^\dagger |\text{vacuum}\rangle. \tag{2.5}$$

The amplitudes obey difference equations obtained from (2.5) and $H|\psi_\alpha\rangle = \varepsilon_\alpha|\psi_\alpha\rangle$ which yields

$$\varepsilon_\alpha \psi_i^\alpha = (v_i - 1)\psi_{i+1}^\alpha + (v_{i-1} - 1)\psi_{i-1}^\alpha. \tag{2.6}$$

The amplitudes obey orthonormality and completeness relations

$$\sum_i \psi_i^{\alpha^*}\psi_i^{\beta} = \delta_{\alpha,\beta}, \qquad \sum_\alpha \psi_i^{\alpha^*}\psi_j^{\alpha} = \delta_{ij}. \tag{2.7}$$

In these relations α's are appropriate labels for the N_a eigenvectors of $\mathcal{H}$.

Equation (2.4) expresses the total energy as a functional of $\{v_i\}$ which requires the explicit solution of the difference equations (2.6). This is awkward if one wishes to explore various forms for $\{v_i\}$. Therefore it is convenient to reformulate the problem in terms of a variational principle where both $\{\psi_i\}$ and $\{v_i\}$ are treated as variables. Consider a functional

$$W_{\text{total}} = \sum_i (v_i - 1) \sum_{\alpha \in \text{Occ}} (\psi_i^{\alpha^*}\psi_{i+1}^{\alpha} + \text{cc}) + (2\gamma)^{-1} \sum_i v_i^2. \tag{2.8}$$

Extremising with respect to $\psi_i^{\alpha^*}$ (subject to orthonormality (2.7)) we recover (2.6). Extremising with respect to v_i (subject to $\Sigma_i v_i = 0$) we find

$$v_i = \gamma\left(J - \sum_{\alpha \in \text{Occ}} (\psi_i^{\alpha^*}\psi_{i+1}^{\alpha} + \text{cc})\right) \tag{2.9}$$

with the Lagrange multiplier

$$J = N_a^{-1} \sum_{\alpha \in \text{Occ}} \sum_i (\psi_i^{\alpha^*}\psi_{i+1}^{\alpha} + \text{cc}). \tag{2.10}$$

Furthermore, at a solution of (2.6) and (2.9), W_{total} equals E_{total}. Thus the stationary points of W_{total} may be regarded as possible excitations of the system.

Equations (2.6) and (2.9) constitute a nonlinear set of equations. The ground state for the half filled band ($N_e = N_a/2$) is believed to be of the Peierls dimerised kind. We show in the appendix that the Peierls state indeed obeys the nonlinear equations exactly for all values of γ. For the other extreme case $N_e \ll N_a$ (i.e. low density), equations (2.6) and (2.9) can be treated in a continuum limit. These equations have been solved exactly (Shastry 1983) and it is found that the Peierls state is stable for arbitrary coupling constants.

We now turn to the study of the excitation spectrum and confine ourselves to the half filled case ($N_e = \frac{1}{2}N_a$). It is convenient to remove a rapidly varying phase factor from the amplitudes of states close to the Fermi momentum. Thus define

$$a_n \equiv (-1)^n \psi_{2n}, \qquad b_n \equiv (-1)^n \psi_{2n-1}, \qquad 1 \leq n \leq L, \tag{2.11}$$

where $L \equiv N_a/2$. Also

$$\omega_n \equiv v_{2n}, \qquad \bar{\omega}_n \equiv -v_{2n-1}. \tag{2.12}$$

The equations of motion (2.6) and (2.9) are transformed into

$$\varepsilon_\alpha a_n^\alpha = (b_{n+1}^\alpha - b_n^\alpha) - (\omega_n b_{n+1}^\alpha + \bar{\omega}_n b_n^\alpha), \tag{2.13}$$

$$\varepsilon_\alpha b_n^\alpha = (a_{n-1}^\alpha - a_n^\alpha) - (\omega_{n-1} a_{n-1}^\alpha + \bar{\omega}_n a_n^\alpha), \tag{2.14}$$

$$\omega_n = \gamma \sum_{\alpha \in \text{Occ}} (a_n^{\alpha^*} b_{n+1}^\alpha + \text{cc}) + \gamma A, \tag{2.15}$$

$$\tilde{\omega}_n = \gamma \sum_{\alpha \in \mathrm{Occ}} (a_n^{\alpha *} b_n^{\alpha} + \mathrm{CC}) - \gamma A. \tag{2.16}$$

The Lagrange multiplier

$$A = (2L)^{-1} \sum_{1 \leq n \leq L} \sum_{\alpha \in \mathrm{Occ}} [a_n^{\alpha *} (b_n^{\alpha} - b_{n+1}^{\alpha}) + \mathrm{CC}]. \tag{2.17}$$

Orthonormality of the amplitudes and completeness (2.7) read

$$\sum_{1 \leq n \leq L} (a_n^{\alpha *} a_n^{\beta} + b_n^{\alpha *} b_n^{\beta}) = \delta_{\alpha,\beta}, \tag{2.18}$$

$$\sum_{\alpha} a_n^{\alpha *} a_m^{\alpha} = \delta_{n,m} = \sum_{\alpha} b_n^{\alpha *} b_m^{\alpha}. \tag{2.19}$$

It is difficult to analyse the above difference equations analytically and so we will replace them by the simplest possible differential equations. The continuum model is then similar to that considered by Takayama *et al* (1980). The smallness of γ (weak coupling) is often cited as a criterion to justify the 'continuisation', although we are not aware of a rigorous discussion in the literature. In the present work we regard it as an assumption of convenience motivated by the hope that the problem would be analytically tractable. Further, one expects that by restricting x to an integer, it may be possible to recover most of the features of the spectrum of the discrete problem. We will comment on this later (in § 6). For the present we *define* the model by the equations (obtained from (2.13)–(2.16))

$$\varepsilon_\alpha a^\alpha(x) = [\mathrm{d}/\mathrm{d}x - 2\omega(x)] b^\alpha(x), \tag{2.20}$$

$$\varepsilon_\alpha b^\alpha(x) = [-\mathrm{d}/\mathrm{d}x - 2\omega(x)] a^\alpha(x), \tag{2.21}$$

$$\omega(x) = \gamma \sum_{\alpha \in \mathrm{Occ}} [a^{\alpha *}(x) b^\alpha(x) + \mathrm{CC}]. \tag{2.22}$$

The constant A in (2.4) can be shown to be $\mathrm{O}(\mathrm{e}^{-L})$ and hence negligible within the continuum solutions; also note that both ω_n and $\tilde{\omega}_n$ have been replaced by $\omega(x)$. The orthonormality of solutions follows from (2.18) as

$$\int_0^1 [a^{\alpha *}(x) a^\beta(x) + b^{\alpha *}(x) b^\beta(x)] \, \mathrm{d}x = \delta_{\alpha,\beta}. \tag{2.23}$$

We impose periodic boundary conditions $a(x) = a(x+L)$, $b(x) = b(x+L)$. An enumeration of the levels from which the N_e occupied levels in (2.22) are chosen would complete the statement of the model. To this end we note that the spectrum of (2.20) and (2.21) may be expected to be unbounded from below as well as above since the equations are analogous to the Dirac equation. Also the continuum model should respect the particle–hole symmetry of the discrete equations (i.e. if ε_α is an eigenvalue, then so is $-\varepsilon_\alpha$). Therefore it is natural to require that one retains precisely $N_\mathrm{a}/2$ states for $\varepsilon > 0$ and an equal number for $\varepsilon < 0$. The occupied levels then correspond to the lowest $N_\mathrm{a}/2$ states. It could be tempting to regard this group of N_a ($=2N_\mathrm{e}$) states as corresponding to the N_a-dimensional state space of the original (discrete) problem. In the general case ($\gamma \neq 0$) the above prescription has the virtue of being natural, and completes the statement of our model.

We note that equations (2.20)–(2.22) can be obtained from extremising a continuum version of the functional (2.8)

$$W^{v}_{\text{total}} = \sum_{\alpha \in \text{Occ}} \int_0^L \mathrm{d}x \left(a^{\alpha *}(x) \frac{\mathrm{d}}{\mathrm{d}x} b^{\alpha}(x) - b^{\alpha *}(x) \frac{\mathrm{d}}{\mathrm{d}x} a^{\alpha}(x) \right.$$
$$\left. -2\omega(x)(a^{\alpha *}(x) b^{\alpha}(x) + \text{cc}) \right) + \frac{1}{\gamma} \int_0^L \omega^2(x)\, \mathrm{d}x. \tag{2.24}$$

At a stationary point (i.e. solution of (2.20)–(2.22)) this reduces to

$$E_{\text{total}} = \sum_{\alpha \in \text{Occ}} \varepsilon_{\alpha} + \frac{1}{\gamma} \int_0^L \omega^2(x)\, \mathrm{d}x. \tag{2.25}$$

3. Approximate solution

In this section we solve equations (2.20)–(2.22) in an approximate scheme. This scheme is similar in spirit to the one devised by Sutherland (unpublished) in a somewhat different context†. We expect the level at the top of the valence band continuum to be the most important one. The strategy is to replace (2.20)–(2.22) by a single-component nonlinear equation for the top level, and a linear equation for the remaining states. This may be achieved by replacing the summation in (2.24) by N_e times the contribution from the top level, a procedure which clearly generates an upper bound to the true energy. This approximation is expected to become exact for $\gamma \gg 1$, i.e. when the band width is negligible. Therefore we find (2.22) replaced by

$$\omega(x) = (\gamma N_e)[2a_0(x)b_0(x)] \tag{3.1}$$

where a_0 and b_0 refer to the wavefunction of the top level of the continuum (assumed real without loss of generality). The equations for a_0 and b_0 may be written as

$$\mathrm{d}a_0(x)/\mathrm{d}x = -b_0(x)[\varepsilon_0 + 4\gamma L a_0^2(x)], \tag{3.2}$$

$$\mathrm{d}b_0(x)/\mathrm{d}x = a_0(x)[\varepsilon_0 + 4\gamma L b_0^2(x)]. \tag{3.3}$$

The remaining fermions obey (2.20) and (2.21) with $\omega(x)$ determined from (3.1). Equations (3.2) and (3.3) possess a constant of motion which renders them exactly integrable. Multiplying (3.2) by a_0 and (3.3) by b_0 and rearranging somewhat, we find

$$\mathrm{d}b_0^2(x)/(\varepsilon_0 + 4\gamma L b_0^2(x)) = -\mathrm{d}a_0^2(x)/(\varepsilon_0 + 4\gamma L a_0^2(x)). \tag{3.4}$$

Therefore

$$[\varepsilon_0 + 4\gamma L b_0^2(x)][\varepsilon_0 + 4\gamma L a_0^2(x)] = \text{constant}. \tag{3.5}$$

Eliminating b_0 from (3.2) using (3.5), we find that the RHS of (3.2) is the square root of a quartic in a_0. Hence the solutions for a_0 and b_0 are Jacobian elliptic functions‡ (compare $(\mathrm{d}/\mathrm{d}u)\,\mathrm{sn}(u) = [(1-\mathrm{sn}^2(u))(1-m\,\mathrm{sn}^2(u))]^{1/2}$). It is straightforward to

† This was reported (Sutherland and Shastry unpublished) at the Rocky Mountain Theoretical Physics Conference, Salt Lake City, Utah, USA in March 1982.

‡ We use the standard notation for Jacobian elliptic functions (e.g. Whittaker and Watson 1963) except that the parameter is displayed rather than the modulus. Wherever the parameter is not displayed, its value is understood to be m.

integrate the equations and we find

$$a_0(x) = A_0 \,\mathrm{sn}(x/\Lambda \,|\, m), \qquad b_0(x) = A_0 \,\mathrm{sn}(x/\Lambda + K \,|\, m). \tag{3.6}$$

The parameter of the elliptic functions is $m \equiv 4\gamma L A_0^2 \Lambda$, and the scaling parameter $\Lambda = 1/|\varepsilon_0|$ ($\varepsilon_0 = -|\varepsilon_0|$). Periodic boundary conditions ($a_0(x) = a_0(x+L)$) yield

$$L/\Lambda = 4Kn. \tag{3.7}$$

Here n is a quantum number which equals the number of oscillations $a_0(x)$ undergoes in the domain $0 \leq x \leq L$. From (3.1), (3.6) and the definition of m, we find

$$\omega(x) = (m/2\Lambda)\,\mathrm{sn}(x/\Lambda)\,\mathrm{sn}(x/\Lambda + K). \tag{3.8}$$

Equation (3.7) implies that the 'order parameter' $\omega(x)$ goes through $4n$ zeros for $0 \leq x \leq L$ and hence we expect $4n$ to equal the number of solitons. Normalisation of (3.6) gives

$$A_0 = [m/8n\Lambda(K-E)]^{1/2}. \tag{3.9}$$

From (3.9) and the definition of m, we obtain

$$K - E = \tfrac{1}{2}\gamma(L/n). \tag{3.10}$$

Equation (3.10) determines m, the parameter of the elliptic functions, as a function of γ and n/L, and hence all the parameters are expressed as functions of γ and n/L. Equations (2.20) and (2.21) may be solved for the remaining fermions, but we shall not pursue this here since essentially the same problem is solved in the next section.

4. Exact solution

We now make the ansatz that (2.20)–(2.21) are solved by the functional form

$$\omega(x) = (m/2\Lambda)\,\mathrm{sn}(u)\,\mathrm{sn}(u+K), \qquad u \equiv x/\Lambda. \tag{4.1}$$

This is of course just (3.8); however, the parameter m and the length scale Λ are undetermined as yet. The ansatz is based upon the success of an earlier similar scheme (Shastry 1983). By squaring (2.20) and (2.21) we obtain Klein–Gordon-like equations

$$(-\mathrm{d}^2/\mathrm{d}x^2 + V_a(x))a^\alpha(x) = \varepsilon_\alpha^2 a^\alpha(x), \tag{4.2}$$

$$(-\mathrm{d}^2/\mathrm{d}x^2 + V_b(x))b^\alpha(x) = \varepsilon_\alpha^2 b^\alpha(x), \tag{4.3}$$

where

$$V_{a,b}(x) = 4\omega^2(x) \mp 2\,\mathrm{d}\omega(x)/\mathrm{d}x. \tag{4.4}$$

Scaling x by Λ and using (4.1) we find

$$\Lambda^2 V_{a,b} = m^2\,\mathrm{sn}^2 u\,\mathrm{sn}^2(u+K) \mp m(\mathrm{d}/\mathrm{d}u)(\mathrm{sn}\,u\,\mathrm{sn}(u+K)). \tag{4.5}$$

Using the identities

$$\mathrm{sn}^2(u)\,\mathrm{sn}^2(u+K) = m^{-1}(1 - \mathrm{cn}^2 u - \mathrm{cn}^2 u + K),$$

$$(\mathrm{d}/\mathrm{d}u)\,\mathrm{sn}\,u\,\mathrm{sn}(u+K) = \mathrm{cn}^2 u - \mathrm{cn}^2(u+K),$$

we find

$$\Lambda^2 V_a = m(2\,\mathrm{sn}^2 u - 1), \tag{4.6}$$

$$\Lambda^2 V_b = m[2\,\mathrm{sn}^2(u+K) - 1]. \tag{4.7}$$

Inserting into (4.2) we obtain

$$(-\mathrm{d}^2/\mathrm{d}u^2 + 2m\,\mathrm{sn}^2 u)a^\alpha(u) = \lambda^\alpha a^\alpha(u), \tag{4.8}$$

$$[-\mathrm{d}^2/\mathrm{d}u^2 + 2m\,\mathrm{sn}^2(u+K)]b^\alpha(u) = \lambda^\alpha b^\alpha(u), \tag{4.9}$$

where

$$\lambda^\alpha = m + \varepsilon_\alpha^2 \Lambda^2. \tag{4.10}$$

Equations (4.9) and (4.10) are in the standard form for Lamé's equation (with $n = 1$ in the notation of Sutherland (1973)). The eigenfunction and eigenvalues are explicitly known (Whittaker and Watson 1963, chap XXIII) and are conveniently expressed in the language of band theory. It is sufficient to consider (4.8) since the corresponding 'partner' b^α can be obtained from (2.21).

4.1. Midgap band

This band corresponds to the first band of Lamé's equation. The eigenfunctions are given by

$$a_q(u) = \text{constant}(\theta_1(u+\mathrm{i}q)/\theta(u))\,\mathrm{e}^{\mathrm{i}u\Phi(q)} \tag{4.11}$$

where $\Phi(q) \equiv \pi q/2KK' + Z(q\,|\,m_1)$. Here θ and Z are Jacobi's theta and zeta functions and $\theta_1(u) = \theta(u+K)$ and $m_1 = 1-m$. The parameter q is necessarily real to ensure bounded (Block-like) solutions and $-K' < q \leq K'$ (since q and $q+2K'$ lead to the same solution). The eigenvalue $\lambda_q = 1 - m_1\,\mathrm{cn}^2(q\,|\,m_1)$, which together with (4.10) gives

$$\varepsilon_q^{\pm} = \pm(\sqrt{m_1}/\Lambda)|\mathrm{sn}(q\,|\,m_1)|. \tag{4.12}$$

Periodic boundary conditions imply

$$L/\Lambda = 4Kn \tag{4.13}$$

and

$$\Phi(q) = \Lambda(2\pi/L) \times \text{integer}. \tag{4.14}$$

From (4.14) we see that q's form a continum in general with a level density $(L/2\pi\Lambda)\mu(q)\,\mathrm{d}q$ where

$$\mu(q) \equiv \mathrm{d}\Phi(q)/\mathrm{d}q = \mathrm{dn}^2(q\,|\,m_1) - 1 + E/K. \tag{4.15}$$

For a given value of n, equation (4.14) implies that there are exactly $2n$ distinct values of q (since for $q = K'$, its maximum value, $\Phi(K') = \pi/2K$ and the RHS equals that for 'integer' $= n$). The integer runs between $-n+1$ and n. Therefore we have a total of $4n$ midgap states (remembering the two signs in (4.12)).

The partial density associated with (4.11) can be found by using Jacobi's addition formula

$$\theta_1(u+\mathrm{i}q)\theta_1(u-\mathrm{i}q) = (\pi/2K\sqrt{m_1})[\theta^2(u)\theta_1^2(\mathrm{i}q) - H^2(u)H_1^2(\mathrm{i}q)]$$

and the definitions $H/\theta = m^{1/4}\,\mathrm{sn}$, $H_1/\theta_1 = m^{1/4}\,\mathrm{cd}$ and Jacobi's imaginary transformation $\mathrm{cd}(iq\,|\,m) = \mathrm{nd}(q\,|\,m_1)$. Thus

$$|a_q(u)|^2 = n_q^m[\mathrm{dn}^2(q\,|\,m_1) - m\,\mathrm{sn}^2(u\,|\,m)]. \tag{4.16}$$

Normalising according to (2.23) we get

$$n_q^m = 1/2L\mu(q). \tag{4.17}$$

Finally let us note that the solution of (4.9) for b_q can be inferred from (2.21) directly. It is readily established that $|b_q(u)|^2 = |a_q(u+K)|^2$ and, moreover, using (2.21)

$$\begin{aligned}(a_q^* b_q^{\pm} + \mathrm{CC}) &= -(\varepsilon_q^{\pm}\Lambda)^{-1}[\mathrm{d}/\mathrm{d}u + 2m\,\mathrm{sn}(u)\,\mathrm{sn}(u+K)]|a_q(u)|^2 \\ &= \pm[2mm_1^{1/2}|\mathrm{sn}(q\,|\,m_1)|n_q^m]\,\mathrm{sn}(u)\,\mathrm{sn}(u+K).\end{aligned} \tag{4.18}$$

(We have used the identity $[\mathrm{d}/\mathrm{d}u + 2m\,\mathrm{sn}(u)\,\mathrm{sn}(u+K)]\,\mathrm{sn}^2(u) = 2\,\mathrm{sn}(u)\,\mathrm{sn}(u+K)$ in (4.18).)

4.2. Valence and conduction bands

These bands correspond to the second band of Lamé's equation. The eigenfunctions of (4.8) are

$$a_p(u) = \mathrm{constant}(H(u - \mathrm{i}p)/\theta(u))\exp[\mathrm{i}u\Psi(p)] \tag{4.19}$$

where

$$\Psi(p) \equiv \mathrm{i}Z(-\mathrm{i}p\,|\,m) = \mathrm{dn}(p\,|\,m_1)\,\mathrm{sc}(p\,|\,m_1) - Z(p\,|\,m_1) - \pi p/2KK'. \tag{4.20}$$

The quantum number p is real and $-K' < p \leqslant K'$ since p and $p + 2K'$ lead to the same state. The corresponding eigenvalue $\lambda_p = 1 + m/\mathrm{cn}^2(p\,|\,m_1)$ and hence from (4.10) we get

$$\varepsilon_p^{\pm} = \pm\Lambda^{-1}\,\mathrm{dc}(p\,|\,m_1). \tag{4.21}$$

The $+(-)$ sign is associated with the conduction (valence) band. Applying periodic boundary conditions we find

$$\Psi(p) = \Lambda(2\pi/L) \times \mathrm{integer}. \tag{4.22}$$

Thus Ψ's form a continuum for large L with a level density $(L/2\pi\Lambda)\sigma(p)\,\mathrm{d}p$ where

$$\sigma(p) \equiv \mathrm{d}\Psi(p)/\mathrm{d}p = \mathrm{dc}^2(p\,|\,m_1) - E/K. \tag{4.23}$$

From (4.23) it is clear that the number of states in this band is infinite (since $\mathrm{dc}(p|m_1) \to \infty$ as $p \to K'$). We must now impose the cut-off procedure discussed in § 2. Since the midgap band contains $4n$ states, the valence and conduction bands must be chosen to contain $L - 2n$ states in each $(L = N_a/2)$. Therefore we retain quantum numbers p in the interval $-p_m < p \leqslant p_m$ where p_m is determined from

$$\Psi(p_m) = (\pi/2K)(L/2n - 1) \tag{4.24}$$

(which follows from 4.22 with integer $= L/2 - n$).

The partial density associated with (4.19) is obtained by using an addition theorem of Jacobi

$$H(u - \mathrm{i}p)H(u + \mathrm{i}p) = (\pi/2Km_1^{1/2})[H^2(u)\theta^2(\mathrm{i}p) - \theta^2(u)H^2(\mathrm{i}p)],$$

the definition $H/\theta = m^{1/4}\,\mathrm{sn}$ and $\mathrm{sn}(\mathrm{i}p\,|\,m) = \mathrm{i}\,\mathrm{sc}(p\,|\,m_1)$. Thus

$$|a_p(u)|^2 = n_p^c[\mathrm{sc}^2(p\,|\,m_1) + \mathrm{sn}^2(u\,|\,m)]. \tag{4.25}$$

Normalisation yields

$$n_p^c = m/2L\sigma(p). \tag{4.26}$$

We note that $b_p(u)$ can be obtained from (2.21) and $|b_p(u)|^2 = |a_p(u+K)|^2$. Let us note also that

$$[a_p^*(u)b_p^{\pm}(u) + \mathrm{CC}] = \mp[2n_p^c\,\mathrm{dc}(p\,|\,m_1)]\,\mathrm{sn}(u)\,\mathrm{sn}(u+K). \tag{4.27}$$

We now consider the self-consistency of our ansatz (4.1). From (2.22) and (4.1) we should have

$$\gamma \sum_{p\in \mathrm{VB}} (a_p^* b_p^- + \mathrm{CC}) + \gamma \sum_{q\in \mathrm{MB}} (a_q^* b_q^- + \mathrm{CC}) \overset{(?)}{=} (m/2\Lambda)\,\mathrm{sn}(u)\,\mathrm{sn}(u+K). \tag{4.28}$$

On the LHS of (4.28) we have chosen the lowest N_e states in the spectrum. From (4.18) and (4.27) we see that the ansatz is functionally satisfied by *every term* in the summation on the LHS identically! Numerical self-consistency is achieved by requiring

$$m/4\gamma\Lambda = \sum_{|p|\leqslant p_m} n_p^c\,\mathrm{dc}(p\,|\,m_1) - m m_1^{1/2} \sum_{|q|\leqslant K'} n_q^m |\mathrm{sn}(q\,|\,m_1)|. \tag{4.29}$$

The summations may be replaced by integrations and on using (4.15), (4.23), (4.17) and (4.26), we obtain after some integrations

$$\tanh^{-1}[\mathrm{sn}(p_m|\,m_1)] - \tanh^{-1}(m_1^{1/2}) = \pi/2\gamma. \tag{4.30}$$

Equations (4.24) and (4.30) constitute two simultaneous (transcendental) equations for p_m and the parameter m in terms of the interaction constant γ and the soliton quantum number n. These must, in general be solved numerically. The total energy of the system may be obtained from (2.25) and after some further integrations we find

$$\begin{aligned} E_{\mathrm{total}}(n) &= -(mL/2\pi\Lambda^2)\,\mathrm{sc}(p_m|\,m_1)\,\mathrm{nc}(p_m|\,m_1) \\ &= -(L/2\pi\Lambda^2)[m_1^{1/2}\cosh(\pi/\gamma) + \tfrac{1}{2}(1+m_1)\sinh(\pi/\gamma)]. \end{aligned} \tag{4.31}$$

These equations are particularly easy to solve in the limit $n/L \ll 1$ (i.e. a low density of solitons). From (4.13) we expect K to be very large in this limit, and on using the asymptotic formula $K(m) \to_{m\to 1} \ln(4/m_1^{1/2})$ we find

$$m_1 \sim 16\exp(-L/2n\Lambda). \tag{4.32}$$

Neglecting terms of $O(m_1)$, (4.24) and (4.30) become

$$\tan p_m - p_m/K = (\pi/2K)(L/2n - 1) + O(m_1), \tag{4.33}$$

$$\tan p_m = \sinh(\pi/2\gamma) + O(m_1) \tag{4.34}$$

(we have lumped together terms like $m_1^{1/2}$, $m_1 \ln m_1$ in the symbol $O(m_1)$). From (4.13) and (4.33) we find

$$\Lambda^{-1} = \Lambda_0^{-1}[1 - (2n/L)(1 - 2p_m/\pi)] + O(m_1) \tag{4.35}$$

where

$$\Lambda_0 \equiv \pi^{-1}\tan p_m = \pi^{-1}\sinh(\pi/2\gamma) + O(m_1). \tag{4.36}$$

The total energy from (4.31) is

$$E_{\text{total}}(n) = -(L/2\pi\Lambda_0^2)\tan p_m \sec p_m[1-(2n/L)(1-2p_m/\pi)]^2 + O(m_1). \tag{4.37}$$

In the limit $n/L \ll 1$, we note that the midgap band width $(2/\Lambda)m_1^{1/2}$ is negligible and these states may be essentially viewed as $4n$ bound states.

5. Features of the solution

(a) Firstly we note that a given n corresponds to $4n$ midgap (soliton) states. It is interesting that the number of solitons is always a multiple of 4, a fact which can be traced back to (2.22) where the 'order parameter' is expressed as a bilinear in fermion amplitudes, each of which undergoes a change in phase in even multiples of π. The case $n = 0$ (zero soliton sector) corresponds to $\omega(x) =$ constant, i.e. a Peierls dimerised state with fermion eigenvalue spectrum

$$\varepsilon_k = \pm(k^2 + \Lambda_0^{-2})^{1/2}. \tag{5.1}$$

For $n \neq 0$, the midgap states may indeed be viewed as interpolating between topologically distinct Peierls dimerised states, as suggested by SSH, and hence termed solitonic.

(b) We compute the soliton creation energy from (4.37) and the definition

$$e_s = (4n)^{-1}[E_{\text{total}}(n) - E_{\text{total}}(0)]. \tag{5.2}$$

Thus

$$e_s = (2\Lambda_0)^{-1}(1-2p_m/\pi)\sec p_m[1-(1-2p_m/\pi)n/L] + O(m_1). \tag{5.3}$$

In the weak coupling limit $\gamma \to 0$, we have (from (4.34)) $p_m \to \pi/2$ and hence $e_s \to 1/\pi\Lambda_0$.

(c) The solution obtained can be generalised to the case of spin-$\frac{1}{2}$ fermions as follows. For this purpose consider $N_e = N_a$, and (2.25) generalised to

$$\tilde{E}_{\text{total}} = 2\sum_{\alpha \in \text{Occ}} \varepsilon_\alpha + \tilde{\gamma}^{-1}\int_0^L \omega^2(x)\,dx.$$

The solution has exactly the same form as in the above case. The self-consistency condition (4.34) is modified to

$$p_m = \tan^{-1}[\sinh(\pi/4\tilde{\gamma})]. \tag{5.4}$$

The soliton energy is changed to

$$e_s = \Lambda_0^{-1}(1-2p_m/\pi)\sec p_m + O(m_1) \tag{5.5}$$

and in the limit $\gamma \to 0$, reduces to $2/\pi\Lambda_0$ in agreement with Takayama *et al* (1980) and Brazovski (1980).

(d) We shall consider next the expectation value of 'charge'. Firstly one has a trivial fractionalisation arising from the fact that the midgap states (4.12) are normalised to unity and have $2n$ peaks, each with an area $1/2n$. This is easily discarded since one may form linear combinations of the degenerate states, and form a set of functions each having area unity under its only peak. Non-trivial fractionalisation occurs, in the picture of SSH, due to the net charge of all the continuum states in the

valence band, distorted by the presence of the solitons. To study this we form partial densities

$$\rho^{A}_{VB}(u) = \sum_{p\in VB} |a_p(u)|^2, \tag{5.6}$$

$$\rho^{B}_{VB}(u) = \sum_{p\in VB} |b^{-}_{p}(u)|^2. \tag{5.7}$$

Combining (4.23), (4.25) and (5.6) and with $\mathrm{sc}(p\,|\,m_1)\to\tan p$, $\mathrm{dc}(p\,|\,m_1)\to\sec(p)$, we get

$$\rho^{A}_{VB}(u) = \tfrac{1}{2}(1-2n/L) + 2np_m/\pi L - (2\pi\Lambda)^{-1}\,\mathrm{cn}^2(x/\Lambda) \tag{5.8}$$

and $\rho^{B}_{VB}(u) = \rho^{A}_{VB}(u+K)$. The total 'charge' density is then

$$\begin{aligned}\rho^{T}_{VB}(u) &= \rho^{A}_{VB}(u) + \rho^{B}_{VB}(u)\\ &= \left(1-\frac{2n}{L}\right) + \frac{4n}{\pi L}p_m - \frac{1}{\pi}p_m\left(\frac{1}{2\Lambda}\,\mathrm{cn}^2(u) + \frac{1}{2\Lambda}\,\mathrm{cn}^2(u+K)\right).\end{aligned} \tag{5.9}$$

It is easy to verify that ρ^{T}_{VB} has $4n$ minima, located exactly at the peaks in the density of the midgap states. The latter may thus be regarded as being formed by depleting the former. Moreover, the integral of ρ^{T}_{VB} is $(L-2n)$ and hence the valence band has the correct number of fermions.

The possibility of a non-integral charge emerges if we measure the area of ρ^{T}_{VB} under any one peak†. We find (for example near the soliton located at zero) for $L \gg l_0 \gg \Lambda$,

$$\int_{-l_0}^{l_0} [\rho^{T}_{VB}(x) - 1]\,\mathrm{d}x = -\pi^{-1}p_m + O(2l_0/L). \tag{5.10}$$

Therefore *fractionalisation* occurs only in the limit $\gamma\to 0$ where $p_m = \pi/2$, otherwise the 'charge' is irrational in general.

This behaviour is closely related to that of the completeness relation‡ (2.19). Setting $n = m$ in (2.19), we define

$$f(x) \equiv \sum_{\alpha\in VB} |a^{\alpha}(x)|^2 + \sum_{\alpha\in CB} |a^{\alpha}(x)|^2 + \sum_{\alpha\in MB} |a^{\alpha}(x)|^2 - 1. \tag{5.11}$$

The indicated summations can be carried out and we find

$$\begin{aligned} f(x) &= 2\rho^{A}_{VB}(x) + (2\Lambda)^{-1}\,\mathrm{dn}^2(x/\Lambda) - 1\\ &= (1-2p_m/\pi)[(2\Lambda)^{-1}\,\mathrm{cn}^2(x/\Lambda) - 2n/L] + O(m_1).\end{aligned} \tag{5.12}$$

Note that the integral of $f(x)$ vanishes but for $x = 0, 2K\Lambda, \ldots, f(x)$ has $2n$ maxima with area $(1-2p_m/\pi)$ under each peak. Then function $f(x)$ is expected to vanish identically if the N_a states retained in our model correspond to the N_a states in the

† This corresponds to fractionalisation of the expectation value of the charge operator. It is believed that the fluctuations of the non-integral charge are vanishingly small in the limit $L\to\infty$. These are some questions as to whether the eigenvalues can themselves be fractional within the context of relativistic field theory (see Rajaraman and Bell 1982). For a review with a solid state emphasis see Prange (1982).

‡ By completeness, we refer to the lattice version in (2.19). The complete spectrum of Lamé's equation is of course complete in the sense of (2.19) with the Dirac delta function on the RHS rather than the Kronecker delta function.

state space of the discrete problem. The only cases for which f vanishes identically are (a) $n = 0$, i.e. the soliton-less Peierls dimerised state for arbitrary γ and (b) $\gamma \to 0$, i.e. the extreme weak coupling limit for any n.

6. Concluding remarks

It is remarkable that the ansatz in § 4 solves the stated problem exactly. Its success is closely related to the similar solvability of a related problem (Shastry 1983). The present solution agrees with the result of Takayama *et al* (1980) and Brazovski (1980) in the weak coupling limit since the midgap states have a density $\mathrm{dn}^2(x/\Lambda)$, which for x close to $2nK\Lambda$ is like $\mathrm{sech}^2(x/\Lambda - 2nK)$. Also the order parameter profile $\mathrm{sn}(u)\,\mathrm{sn}(u+K)$ is locally like the hyperbolic tangent. The soliton energy is also in agreement. Horovitz (1981) also finds that Lamé's equation is relevant to the n-soliton solution for a similar model.

An interesting aspect of our solution is the failure of $f(x)$ (5.12) to vanish locally in general. In a related phenomenon, the expectation value of charge in the proximity of a soliton is $-\pi^{-1}\tan^{-1}[\sinh(\pi/4\tilde{\gamma})]$ (from (5.10) and (5.4)), an irrational number in general. (For the typical numbers quoted by SSH for polyacetylene we obtain a 'charge' $-0.445\,357\,964\ldots$!) It is only in the limit of weak coupling that one obtains a fractional answer. For the lattice model, the arguments of SSH (see especially the discussion in Jackiw and Schrieffer 1981, after equation (2.14)) are quite general, and show that if a soliton-like solution exists then the 'charge' must be exactly $-\frac{1}{2}$. The result obtained for the present model for non-zero γ must then be an artifact of the continuum theory. The origin of the difficulty is seen to be the failure of the completeness relation $f(x) \neq 0$ (equation (5.12)) locally, although globally $\int f\,\mathrm{d}x = 0$. What happens is that in building the midgap states in the functional form (4.11) one 'uses up' momenta which are not contained in the discrete state space, and thus the correspondence between the discrete and continuum models is lost. Thus the embedding of the state space of the discrete problem in the Hilbert space of the continuum problem is not 'well ordered' in energy. In the extreme weak coupling, however, the solution found here is in accord with the results of SSH for the discrete model and provides support for their physical picture for the excitations of the system.

It is worth mentioning that the remarkable factorisation of the solution (commented upon after (4.28)) can be exploited to construct 'weighted' solvable models which satisfy the (discrete) completeness properties. For simplicity consider the case of small n. Here the sum over valence states $\Sigma_{p\in \mathrm{VB}}$ is replaced by $\int (L/2\pi\Lambda)\sigma(p)W(p)\,\mathrm{d}p$, where $W(p)$ is a weight factor which enables us to enforce the completeness relation $f(x) = 0$. Requiring completeness, $\Sigma_{p\in \mathrm{VB}}\, 1 = L - 2n$ and the self-consistency (4.28), we find

$$\delta_n \equiv \pi^{-1}\int_{-\pi/2}^{\pi/2} W(p)\sec^n(p)\,\mathrm{d}p, \qquad \delta_0 = 1, \qquad \delta_1 = 1/\gamma, \qquad \delta_2 = 2\Lambda.$$

Therefore it is clear that every choice of the weight function which obeys the above three relations generates an exactly solvable model obeying the completeness relation. However, since these involve $W(p) > 1$ in general (from $\delta_0 = 1$), it is difficult to assign a physical meaning to the models, and hence to identify a $W(p)$ which would be 'best' in some sense. (A model with $W(p) \leq 1$ can clearly be interpreted as consisting of selective retention of states in the continuum.)

In conclusion we remark that the *discrete* nonlinear equations ((2.6), (2.9)) may be expected to have solutions which are qualitatively similar to the ones found here. It should be possible to compute the exact soliton profiles and energies numerically.

Acknowledgment

It is a pleasure to thank John Bruno and Bill Sutherland for stimulating discussions which led to our interest in this problem, and members of the Theory Group at TIFR for valuable discussions.

Appendix. Stability of Peierls dimerisation

We show in the following that the nonlinear difference equations (2.6) and (2.9) are solved by the Peierls dimerised state

$$V_{2n} = V, \qquad V_{2n+1} = -V. \tag{A1}$$

It is convenient to rewrite (2.6) and (2.9) in the form

$$\begin{aligned} \varepsilon_\alpha \psi^\alpha_{2n} &= (V-1)\psi^\alpha_{2n+1} - (V+1)\psi^\alpha_{2n-1}, \\ \varepsilon_\alpha \psi^\alpha_{2n+1} &= -(V+1)\psi^\alpha_{2n+2} + (V-1)\psi^\alpha_{2n}, \end{aligned} \tag{A2}$$

and

$$V_{2n} = \gamma(J - X_{2n}), \qquad V_{2n+1} = \gamma(J - X_{2n+1}). \tag{A3}$$

We have defined

$$X_m \equiv \sum_{\alpha \in O_{CC}} (\psi^{\alpha *}_m \psi^\alpha_{m+1} + \text{CC}) \tag{A4}$$

so

$$J = N_a^{-1} \sum_{1 \le n \le N_a/2} (X_{2n} + X_{2n-1}). \tag{A5}$$

The sum in (A4) is over the lowest $N_e (= N_a/2)$ eigenstates. The difference equations (A2) are solved by

$$\psi_{2n} = a_k \exp(\mathrm{i}k2n), \qquad \psi_{2n+1} = b_k \exp(\mathrm{i}k2n). \tag{A6}$$

Substituting (A6) into (A2) we find

$$\varepsilon_k^{\pm} = \pm 2(\cos^2 k + V^2 \sin^2 k)^{1/2}, \tag{A7}$$

$$\frac{b_k}{a_k} = \exp(\mathrm{i}\phi_k), \qquad \phi_k \equiv \tan^{-1}\left(\frac{(1+V)\tan k}{1 - V\tan^2 k}\right). \tag{A8}$$

The normalised amplitudes may be chosen in the form

$$a_k = N_a^{-1/2} \exp(-\mathrm{i}\phi_k/2), \qquad b_k = a_k^*. \tag{A9}$$

The $-ve$ $(+ve)$ solutions in (A7) are associated with valence (conduction) bands. Periodic boundary conditions yield

$$k = (2\pi/N_a) \times \text{integer}.$$

Clearly $-\pi/2 < k \leq \pi/2$ since k and $k+\pi$ lead to the same solution. Thus we have $N_a/2$ distinct states in the valence and conduction bands. From (A4), (A5), (A6) and (A9) we obtain

$$X_E = X_{2n} = \int_{-\pi/2}^{\pi/2} \frac{dk}{\pi} \cos \phi_k, \tag{A10}$$

$$X_O = X_{2n+1} = \int_{-\pi/2}^{\pi/2} \frac{dk}{\pi} \cos(\phi_k - 2k), \tag{A11}$$

$$J = \tfrac{1}{2}(X_E + X_O). \tag{A12}$$

Comparing (A3) and (A12) with (A1), we see that the assumption (A1) is self-consistent provided

$$V = \frac{\gamma}{2} \int_{-\pi/2}^{\pi/2} \frac{dk}{\pi} [\cos(\phi_k - 2k) - \cos \phi_k]. \tag{A13}$$

Using (A8) this condition can be written in the form

$$V = [2\gamma V/\pi(1-V^2)][K(1-V^2) - E(1-V^2)]. \tag{A14}$$

This condition can of course, be derived by computing the total energy from (A1), (A7) and (2.4) and minimising with respect to V. The above derivation shows that the Peierls argument is better than variational; it is, in fact, an exact solution of the Euler–Lagrange equations. The problem of showing that this state is the absolute minimum is open.

References

Brazovski S A 1980 *Sov. Phys.-JETP* **51** 342
Horovitz B 1981 *Phys. Rev. Lett.* **46** 742
Jackiw R and Schrieffer J R 1981 *Nucl. Phys.* B **190** 253
Prange R E 1982 *Phys. Rev.* B **26** 991
Rajaraman R and Bell J S 1982 *CERN Preprint TH-3331*
Rice M J 1979 *Phys. Lett.* **71A** 152
Shastry B S 1983 *Phys. Rev. Lett.* **50** 633
Su W P, Schrieffer J R and Heeger A J 1979 *Phys. Rev. Lett.* **42** 1698
—— 1980 *Phys. Rev.* B **22** 2099
Sutherland B 1973 *Phys. Rev.* A **8** 2514
Takayama H, Lin-Liu Y R and Maki K 1980 *Phys. Rev.* B **21** 2388
Whittaker E T and Watson G N 1963 *A Course of Modern Analysis* (Cambridge: CUP)