Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 321 (2003) 1–27
PHYSICA A
www.elsevier.com/locate/physa

The one-dimensional Hubbard model: a reminiscence☆

Elliott H. Lieb[a], F.Y. Wu[b,*]

[a] *Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08544, USA*
[b] *Department of Physics, Northeastern University, Boston, MA 02115, USA*

Abstract

In 1968 we published the solution of the ground state energy and wave function of the one-dimensional Hubbard model, and we also showed that there is no Mott transition in this model. Details of the analysis have never been published, however. As the Hubbard model has become increasingly important in condensed matter physics, relating to topics such as the theory of high-T_c superconductivity, it is appropriate to revisit the one-dimensional model and to recall here some details of the solution.

PACS: 71.10.Fd; 71.27.+a; 75.10.Lp

Keywords: Hubbard model; One dimension; Exact solution

1. Introduction

In a previous paper [1] we reported the solution of the one-dimensional (1D) Hubbard model, showing the absence of the Mott transition in its ground state, but the letter format of the paper did not permit the presentation of all the details of the analysis. Over the years the Hubbard model [2,3] has become more important, for it plays an essential role in several topics in condensed matter physics, including 1D conductors and high-T_c superconductivity. It also plays a role in the chemistry of aromatic compounds (e.g., Benzene [4,5]). Several books [6–9] now exist in which the 1D Hubbard

☆ This article may be reproduced in its entirety for non-commercial purposes.

* Corresponding author.
E-mail address: fywu@neu.edu (F.Y. Wu).

0378-4371/03/$ - see front matter
doi:10.1016/S0378-4371(02)01785-5

model is analyzed, and numerous papers have been written on properties of the model.[1] Almost invariably these publications are based upon results of [1], including the absence of a Mott transition, but without derivation. While other rigorous results on higher dimensional Hubbard models exist, and some of these are reviewed in Refs. [10,11], the ID model stands as the only Hubbard model whose ground state can be found exactly. It has been brought to our attention that it would be useful to students and researchers if some details of the solution could be made available. Here, taking the opportunity of the symposium, StatPhys-Taiwan 2002, which takes place in the year when both of the authors turn 70, we revisit the 1D Hubbard model and present some details of the 34-year old solution.

While our paper [1] contained significant results about the excitation spectrum, it was mainly concerned with the integral equations for the ground state and we concentrate on those equations here. The new, unpublished results are contained in Sections 5–7.

Consider a crystal of N_a lattice sites with a total of N itinerant electrons hopping between the Wannier states of neighboring lattice sites, and that each site is capable of accommodating two electrons of opposite spin, with an interaction energy $U > 0$, which mimics a screened Coulomb repulsion among electrons. The Hubbard model [3] is described by the Hamiltonian

$$\mathscr{H} = T \sum_{\langle ij \rangle} \sum_{\sigma} c_{i\sigma}^{\dagger} c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow} \,, \tag{1}$$

where $c_{i\sigma}^{\dagger}$ and $c_{i\sigma}$ are, respectively, the creation and annihilation operators for an electron of spin σ in the Wannier state at the ith lattice site and $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$ is the occupation number operator. The summation $\langle ij \rangle$ is over nearest neighbors, and one often considers (as we do here) periodic boundary conditions, which means that $\langle ij \rangle$ includes a term coupling opposite edges of the lattice. We are interested in the ground state solution of the Schrödinger equation $\mathscr{H}|\psi\rangle = E|\psi\rangle$.

For bipartite lattices (i.e., lattices in which the set of sites can be divided into two subsets, A and B, such that there is no hopping between A sites or between B sites), such as the 1D chain, the unitary transformation $V^{\dagger}\mathscr{H}V$ leaves $\mathscr{H}$ unchanged except for the replacement of T by $-T$. Here $V = \exp[\mathrm{i}\pi \sum_{i \in A} (n_{i\uparrow} + n_{i\downarrow})]$, with A being one of the two sublattices. Without loss of generality we can, therefore, take $T = -1$. In any event, bipartite or not, we can renormalize U by redefining U to be $U/|T|$. Henceforth, the value of T in (1) is -1 and U is positive and fixed. The dependence of the Hamiltonian and the energy on U will not be noted explicitly.

The commutation relations

$$\left[\sum_{i} n_{i\uparrow}, \mathscr{H} \right] = \left[\sum_{i} n_{i\downarrow}, \mathscr{H} \right] = 0$$

[1] For example, there have been over 500 citations to Ref. [1] in papers published in the *Physical Review* and in *Physical Review Letters* alone from 1968 to 2002. Most of the papers on the one-dimensional Hubbard model can be traced from the *PROLA* link of the American Physical Society web page.

imply that the numbers of down-spin electrons M and up-spin electrons M' are good quantum numbers. Therefore we characterize the eigenstates by M and M', and write the Schrödinger equation as

$$\mathscr{H}|M,M'\rangle = E(M,M')\,|M,M'\rangle\ . \tag{2}$$

Naturally, for any fixed choice of M,M' there will generally be many solutions to (2), so that $|M,M'\rangle$ and $E(M,M')$ denote only generic eigenvectors and eigenvalues. Furthermore, by considering particles as holes, and vice versa, namely, introducing fermion operators

$$d_{i\sigma} = c^{\dagger}_{i\sigma}, \quad d^{\dagger}_{i\sigma} = c_{i\sigma}$$

and the relation $n_{i\sigma} = 1 - d^{\dagger}_{i\sigma}d_{i\sigma}$, we obtain the identity

$$E(M,M') = -(N_a - N)U + E(N_a - M, N_a - M')\,, \tag{3}$$

where

$$N = M + M'$$

is the total number of electrons. Since $N \geqslant N_a$ if, and only if, $(N_a - M) + (N_a - M') \leqslant N_a$, we can restrict our considerations to

$$N \leqslant N_a\ ,$$

namely, the case of at most a "half-filled band". In addition, owing to the spin-up and spin-down symmetry, we need only consider

$$M \leqslant M'\ .$$

2. The 1D model

We now consider the 1D model, and write $|M,M'\rangle$ as a linear combination of states with electrons at specific sites. Number the lattice sites by $1,2,\ldots,N_a$ and, since we want to use periodic boundary conditions, we require N_a to be an even integer in order to retain the bipartite structure. For later use it is convenient also to require that $N_a = 2\times$ (odd integer) in order to be able to have $M = M' = N_a/2$ with M odd. For the 1D model the sum in (1) over $\langle ij\rangle$ is really a sum over $1 \leqslant i \leqslant N_a,\ j = i+1$, plus $1 \leqslant j \leqslant N_a,\ i = j+1$ with $N_a + 1 \equiv 1$.

Let $|x_1,\ldots,x_N\rangle$ denote the state in which the down-spin electrons are located at sites $x_1,x_2,\ldots,x_M$ and the up-spin electrons are at sites $x_{M+1},\ldots,x_N$. The eigenstate is now written as

$$|M,M'\rangle = \sum_{1\leqslant x_i\leqslant N_a} f(x_1,\ldots,x_N)|x_1,\ldots,x_N\rangle\ , \tag{4}$$

where the summation is over all $x_1,\ldots,x_N$ from 1 to N_a, and $f(x_1,\ldots,x_N)$ is the amplitude of the state $|x_1,\ldots,x_N\rangle$.

It is convenient to denote the N-tuple $x_1, x_2, \ldots, x_N$ simply by X.

By substituting (4) into the Schrödinger equation (2), we obtain (recall $T = -1$)

$$-\sum_{i=1}^{N} [f(x_1, \ldots, x_i + 1, \ldots, x_N) + f(x_1, \ldots, x_i - 1, \ldots, x_N)] + U \left[\sum_{i<j} \delta(x_i - x_j)\right] f(x_1, \ldots, x_N) = E f(x_1, \ldots, x_N) , \tag{5}$$

where δ is the Kronecker delta function. We must solve (5) for f and E, with the understanding that site 0 is the same as site N_a and site N_a+1 is the same as site 1 (the periodic boundary condition). Eq. (5) is the 'first quantized' version of the Schrödinger equation (2). It must be satisfied for all $1 \leqslant x_i \leqslant N_a$, with $1 \leqslant i \leqslant N$.

As electrons are governed by Fermi–Dirac statistics, we require that $f(X)$ be antisymmetric in its first M and last M' variables *separately*. This antisymmetry also ensures that $f = 0$ if any two x's in the same set are equal, which implies that the only delta-function term in (5) that are relevant are the ones with $i \leqslant M$ and $j > M$. This is consistent with the definition of $\mathscr{H}$ in (1), in which the only interaction is between up- and down-spin electrons.

The antisymmetry allows us to reinterpret (5) in the following alternative way. Define the region R to be the following subset of all possible values of X (note the $<$ signs):

$$R = \left\{ X : \begin{pmatrix} 1 \leqslant x_1 < x_2 < \cdots < x_M \leqslant N_a \\ 1 \leqslant x_{M+1} < x_{M+2} < \cdots < x_N \leqslant N_a \end{pmatrix} \right\} . \tag{6}$$

In R any of the first M x_i's can be equal to any of the last M', with an interaction energy nU, where n is the number of overlaps of the first set with the second.

The antisymmetry of f tells us that f is completely determined by its values in the subset R, together with the requirement that $f = 0$ if any two x's in the same set are equal (e.g., $x_1 = x_2$). Therefore, it suffices to satisfy the Schrödinger equation (5) when X on the right-hand side of (5) is only in R, together with the additional fact that we set $f = 0$ on the left-hand side of (5) if $x_i \pm 1$ takes us out of R, e.g., if $x_1 + 1 = x_2$. (*Warning:* With this interpretation, Eq. (5) then becomes a self-contained equation in R alone and one should not ask it to be valid if $X \notin R$.)

There is one annoying point about restricting attention to R in (5). When $x_1 = 1$ the left-hand side of (5) asks for the value of f for $x_1 = N_a$, which takes us outside R. Using the antisymmetry we conclude that

$$f(N_a, x_2, \ldots, x_N) = (-1)^{M-1} f(x_2, \ldots, x_M, N_a, x_{M+1}, \ldots, x_N) \tag{7}$$

with similar relations holding for $x_M = 1$, $x_{M+1} = N_a$ or $x_N = 1$. Eq. (7) and its three analogues reflect the "periodic boundary conditions" and, with its use, (5) becomes a self-contained equation on R alone.

We now come to the main reason for introducing R. *Let us assume that M and M' are both odd integers.* Then $(-1)^{M-1}=(-1)^{M'-1}=1$ and we claim that: *For all U, the ground state of our Hamiltonian satisfies*

(1) There is only one ground state and
(2) $f(X)$ is a strictly positive function in R.

To prove (2) we think of (5) as an equation in R, as explained before. We note that all the off-diagonal terms in the Hamiltonian (thought of as a matrix $\widehat{\mathscr{H}}$ from $\mathrm{L}^2(R)$ to $\mathrm{L}^2(R)$) are non-positive (this is where we use the fact that $(-1)^{M-1}=(-1)^{M'-1}=1$). If E_0 is the ground state energy and if $f(X)$ is a ground state eigenfunction (in R), which can be assumed to be real, then, by the variational principle, the function $g(X)=|f(X)|$ (in R) has an energy at least as low as that of f, i.e., $\langle g\,|\,g\rangle=\langle f\,|\,f\rangle$ and $\langle g\,|\widehat{\mathscr{H}}|\,g\rangle\leqslant\langle f\,|\widehat{\mathscr{H}}|\,f\rangle$ since $|f(X)|\,\widehat{\mathscr{H}}(X,Y)\,|f(Y)|\leqslant f(X)\widehat{\mathscr{H}}(X,Y)f(Y)$ for every X,Y. Hence g must be a ground state as well (since it cannot have a lower energy than E_0, by the definition of E_0). Therefore, $g(X)$ must satisfy (5) with the same E_0. Moreover, we see from (5) that $g(X)$ is strictly positive for *every* $X\in R$ (because if $g(Y)=0$ for some $Y\in R$ then $g(Z)=0$ for every Z that differs from Y by one 'hop'; tracing this backward, $g(X)=0$ for every $X\in R$).

Returning now to f, let us assume the contrary of (2), namely, $f(X)>0$ for some $X\in R$, and $f(Y)\leqslant 0$ for some $Y\in R$. We observe that since $h=g-f$ must also be a ground state (because sums of ground states are ground states, although not necessarily normalized), we have a ground state (namely h) that is non-negative and non-zero, but not strictly positive; this contradicts the fact, which we have just proved, that every non-negative ground state must be strictly positive. Thus, (2) is proved.

A similar argument proves (1). If f and f' are two linearly independent ground states then the state given by $k(X)=f(X)+cf'(X)$ is also a ground state and, for suitable c, $k(X)=0$ for some $X\in R$, but k cannot be identically zero. Then $|k|$ will be a non-negative ground state that is not strictly positive, and this contradicts statement (2).

The uniqueness statement (1) is important for the following reason. Suppose that we know the ground state for some particular value of U (e.g., $U=\infty$) and suppose we have a U-dependent solution to (5) in some interval of U values (e.g., $(0,\infty)$) with an energy $E(U)$ such that: (a) $E(\infty)$ is the known ground state energy and (b) $E(U)$ is continuous on the interval. Then $E(U)$ is necessarily the ground state energy in that interval. If not, the curve $E(U)$ would have to cross the ground state curve (which is always continuous), at which point there would be a degeneracy—which is impossible according to (1).

Items (1) and (2) can be used in two main applications. The first is the proof of the fact that when M and M' are odd the ground state belongs to total spin S equal to $|M-M'|/2$ and not to some higher S value. The proof is the same as in Ref. [12]. In Ref. [12] this property was shown to hold for *all* values of M and M', but for an open chain instead of a closed chain. In the thermodynamic limit this distinction is not important.

The second main application of these items (1) and (2) is a proof that the state we construct below using the Bethe Ansatz really is the ground state. This possibility

is addressed at the end of Section 3 where we outline a strategy for such a proof. Unfortunately, we are unable to carry it out and we leave it as an open problem.

We also mention a theorem [13], which states that *the ground state is unique for* $M = M' = N_a/2$ *and* $N_a =$ *even* (*the half-filled band*). There is no requirement for $M = M'$ to be odd.

3. The Bethe Ansatz

The Bethe Ansatz was invented [14] to solve the Heisenberg spin model, which is essentially a model of lattice bosons. The boson gas in the continuum with a positive delta function interaction and with positive density in the thermodynamic limit was first treated in Ref. [15]. McGuire [16] was the first to realize that the method could be extended to continuum fermions with a delta function interaction for $M = 1$. (The case $M = 0$ is trivial.) The first real mathematical difficulty comes with $M = 2$ and this was finally solved in Ref. [17]. The solution was inelegant and not transparent, but was a precursor to the full solution for general M by Gaudin [18] and Yang [19].

We now forget about the region R and focus, instead, on the fundamental regions (note the $\leqslant$ signs)

$$R_Q = \{X\colon\ 1 \leqslant x_{Q1} \leqslant \cdots \leqslant x_{QN} \leqslant N_a\}\ . \tag{8}$$

Here $Q = \{Q1, Q2, \ldots, QN\}$ is the permutation that maps the ordered set $\{1, 2, \ldots, N\}$ into $\{Q1, Q2, \ldots, QN\}$. There are $N!$ permutations and corresponding regions R_Q. The union of these regions is the full configuration space. These regions are disjoint *except* for their boundaries (i.e., points where $x_{Qi} = x_{Q(i+1)}$).

Let $k_1 < k_2 < \cdots < k_N$ be a set of *unequal*, *ordered* and *real* numbers in the interval $-\pi < k \leqslant \pi$, and let $[Q,P]$ be a set of $N! \times N!$ coefficients indexed by a pair of permutations Q, P, all yet to be determined.

When $X \in R_Q$ we write the function $f(X)$ as (the Bethe Ansatz)

$$f(X) = f_Q(x_1, \ldots, x_N) = \sum_P [Q,P] \exp[\mathrm{i}(k_{P1}x_{Q1} + \cdots + k_{PN}x_{QN})]\ . \tag{9}$$

In order for (9) to represent a function on the whole configuration space it is essential that the definitions (9) agree on the intersections of different R_Q's. This will impose conditions on the $[Q,P]$'s.

Choose some integer $1 \leqslant i < N$ and let $j = i+1$. Let P, P' be two permutations such that $Pi = P'j$ and $Pj = P'i$, but otherwise $Pm = P'm$ for $m \neq i, j$. Similarly, let Q, Q' be a pair with the same property (for this same choice of i) but otherwise P, P' and Q, Q' are unrelated.

The common boundary between R_Q and $R_{Q'}$ is the set in which $x_{Qi} = x_{Qj}$. In order to have $f_Q = f_{Q'}$ on this boundary it is sufficient to require that

$$[Q,P] + [Q,P'] = [Q',P] + [Q',P']\ . \tag{10}$$

The reason that this suffices is that on this boundary we have $x_{Qi} = x_{Qj}$ and $k_{Pi} + k_{Pj} = k_{P'i} + k_{P'j}$. Thus, (10) expresses the fact that the exponential factor $\exp[\mathrm{i}(k_{Pi}x_{Qi} + k_{Pj}x_{Qj})] = \exp[\mathrm{i}(k_{P'i}x_{Qi} + k_{P'j}x_{Qj})]$ is the same for Q and Q', and for all values of the other x_m's.

Next we substitute the Ansatz (9) into (5). If $|x_i - x_j| > 1$ for all i, j then, clearly, we have

$$E = E(M, M') = -2\sum_{j=1}^{N} \cos k_j \,. \tag{11}$$

We next choose the coefficients $[Q, P]$ to make (11) hold generally—even if it is not possible to have $|x_i - x_j| > 1$ for all i, j when, for example, the number of electrons exceeds $N_a/2$. The requirement that (11) holds will impose further conditions on $[Q, P]$ similar to (10).

Sufficient conditions are obtained by setting $x_{Qi} = x_{Qj}$ on the right-hand side of (5) and requiring the exponential factors with x_{Qi} and x_{Qj} alone to satisfy (5). In other words, we require that

$$\begin{aligned}
&[Q,P]\mathrm{e}^{-\mathrm{i}k_{Pj}} + [Q,P']\mathrm{e}^{-\mathrm{i}k_{P'j}} + [Q,P]\mathrm{e}^{+\mathrm{i}k_{Pi}} + [Q,P']\mathrm{e}^{+\mathrm{i}k_{P'i}} \\
&\quad = [Q',P]\mathrm{e}^{-\mathrm{i}k_{Pi}} + [Q',P']\mathrm{e}^{-\mathrm{i}k_{P'i}} + [Q',P]\mathrm{e}^{+\mathrm{i}k_{Pj}} + [Q',P']\mathrm{e}^{+\mathrm{i}k_{P'j}} \\
&\qquad + U([Q,P] + [Q,P']) \,.
\end{aligned} \tag{12}$$

If we combine (12) with (10) and recall that $k_{Pj} = k_{P'i}$, etc., we obtain

$$\begin{aligned}
[Q,P] = &\frac{-\mathrm{i}U/2}{\sin k_{Pi} - \sin k_{Pj} + \mathrm{i}U/2}\,[Q,P'] \\
&+ \frac{\sin k_{Pi} - \sin k_{Pj}}{\sin k_{Pi} - \sin k_{Pj} + \mathrm{i}U/2}[Q',P'] \,.
\end{aligned} \tag{13}$$

It would seem that we have to solve both (13) *and* (10) for the $(N!)^2$ coefficients $[Q, P]$, and for each $1 \leqslant i \leqslant N-1$. Nevertheless, (13) alone is sufficient because it implies (10). To see this, add (13), as given, to (13) with $[Q', P]$ on the left side. Since $Q'' = Q$, the result is (10). Our goal, then, is to solve (13) for the coefficients $[Q, P]$ such that the amplitude f has the required symmetry.

These equations have been solved in Refs. [18,19], as we stated before, and we shall not repeat the derivation here. In these papers the function $\sin k$ appearing in (13) is replaced by k, which reflects the fact that Refs. [18,19] deal with the continuum and we are working on a lattice. This makes no difference as far as the algebra leading to Eqs. (14) below is concerned, but it makes a big difference for constructing a proof that these equations have a solution (the reason being that the sine function is not one-to-one).

The algebraic analysis in Refs. [18,19] leads to the following set of $N+M$ equations for the N ordered, real, unequal k's. (Recall that $M \leqslant M'$.) They involve an additional

set of M ordered, unequal real numbers $\Lambda_1 < \Lambda_2 < \cdots < \Lambda_M$.

$$e^{ik_jN_a} = \prod_{\beta=1}^{M} \frac{i\sin k_j - i\Lambda_\beta - U/4}{i\sin k_j - i\Lambda_\beta + U/4}, \quad j = 1,2,\ldots,N$$

$$\prod_{j=1}^{N} \frac{i\sin k_j - i\Lambda_\alpha - U/4}{i\sin k_j - i\Lambda_\alpha + U/4} = -\prod_{\beta=1}^{M} \frac{-i\Lambda_\beta + i\Lambda_\alpha + U/2}{-i\Lambda_\beta + i\Lambda_\alpha - U/2}, \quad \alpha = 1,2,\ldots,M. \tag{14}$$

We remark that an explicit expression for the wave function $f(X)$ has been given by Woynarovich [20, part 1, Eqs. (2.5)–(2.9)].

These equations can be cast in a more transparent form (in which we now really make use of the fact that the k's and Λ's are ordered) by defining

$$\theta(p) = -2\tan^{-1}\left(\frac{2p}{U}\right), \quad -\pi \leqslant \theta \leqslant \pi.$$

Then, taking the logarithm of (14), we obtain two sets of equations

$$N_a k_j = 2\pi I_j + \sum_{\beta=1}^{M} \theta(2\sin k_j - 2\Lambda_\beta), \quad j = 1,2,\ldots,N, \tag{15}$$

$$\sum_{j=1}^{N} \theta(2\sin k_j - 2\Lambda_\alpha) = 2\pi J_\alpha - \sum_{\beta=1}^{M} \theta(\Lambda_\alpha - \Lambda_\beta), \quad \alpha = 1,2,\ldots,M, \tag{16}$$

where I_j is an integer (half-odd integer) if M is even (odd), while J_α is an integer (half-odd integer) if M' is odd (even).

It is noteworthy that in the $U \to \infty$ limit the two sets of equations essentially decouple. The Λ's are proportional to U in this limit, but the sum in (15) becomes independent of j. In particular, when the Λ's are balanced (i.e., for every Λ there is a $-\Lambda$) as in our case, then this sum equals zero.

From (15) and (16) we have the identity

$$k_{total} \equiv \sum_{j=1}^{N} k_j = \frac{2\pi}{N_a}\left(\sum_{j=1}^{N} I_j + \sum_{\alpha=1}^{M} J_\alpha\right). \tag{17}$$

For the ground state, with $N = 2\times$ (odd integer) and $M = N/2 =$ odd, we make the choice of the I_j and J_α that agrees with the correct values in the case $U = \infty$, namely

$$I_j = j - (N+1)/2, \quad J_\alpha = \alpha - (M+1)/2. \tag{18}$$

We are not able to prove the existence of solutions to (15) and (16) that are real and increasing in the index j and α. In the next section, however, we show that the $N \to \infty$ limit of (15) and (16) has a solution, and in Section 6 we obtain the solution explicitly for $N/2M = N/N_a = 1$. This leaves little doubt that (15) and (16) can be solved as well, at least for large N.

Assuming that $M = M' = N/2$ is odd, the solution is presumably unique with the given values of I_j and J_α and belongs to total spin $S = 0$.

Assuming that the solution exists, we would still need a few more facts (which we have not proved) in order to prove that the Bethe Ansatz gives the ground state:

(a) prove that the wave function (9) is not identically zero,
(b) prove that the wave function (9) is a continuous function of U.

From the uniqueness of the ground state proved in Section 2, and the fact that solution (9) coincides with the exact solution for $U = \infty$ (in which case $f_Q(x)$ is a Slater determinant of plane waves with wavenumbers $k_j = 2\pi I_j/N_a$), (a) and (b) now establish that wave function (9) must be the ground state for all U.

Remark. Assuming that the Bethe Ansatz gives the ground state for a given $M \leqslant M'$ then, as remarked at the end of Section 2 (and assuming M and M' to be odd) the value of the total spin in this state is $S = (M' - M)/2$. Thus, the solution to the Bethe Ansatz we have been looking at is a highest weight state of $SU(2)$, i.e., a state annihilated by spin raising operators.

4. The ground state

For the ground state $I_j = I(k_j)$ and $J_\alpha = J(\Lambda_\alpha)$ are consecutive integers or half-odd integers centered around the origin. As stated in Section 3, each k_j lies in $[-\pi, \pi]$ (since $k_j \to k_j + 2\pi n$ defines the same wave function). In the limit of N_a, N, M, $M' \to \infty$ with their ratios kept fixed, the real numbers k and Λ are distributed between $-Q$ and $Q \leqslant \pi$ and $-B$ and $B \leqslant \infty$ for some $0 < Q \leqslant \pi$ and $0 < B \leqslant \infty$. In a small interval $\mathrm{d}k$ the number of k values, and hence the number of j values in (15), is $N_a\rho(k)\,\mathrm{d}k$, where ρ is a density function to be determined. Likewise, in a small interval $\mathrm{d}\Lambda$ the number of Λ values and α values in (16) is $N_a\sigma(\Lambda)\,\mathrm{d}\Lambda$. An alternative point of view is to think of $I(k)$ as a function of the variable k. Then $I(k + \mathrm{d}k) - I(k)$ counts the number of k values between k and $k + \mathrm{d}k$ so we have $\mathrm{d}I(k)/\mathrm{d}k = N_a\rho(k)$. A similar remark holds for $J(\Lambda)$.

The density functions $\rho(k)$ and $\sigma(\Lambda)$ satisfy the obvious normalization

$$\int_{-Q}^{Q} \rho(k)\,\mathrm{d}k = N/N_a, \qquad \int_{-B}^{B} \sigma(\Lambda)\,\mathrm{d}\Lambda = M/N_a\ . \tag{19}$$

By subtracting (15) with j from (15) with $j + N_a\rho(k)\,\mathrm{d}k$, and taking the limit $N_a \to \infty$ we obtain (20) below. Likewise, subtracting (16) with α from (16) with $\alpha + N_a\sigma(\Lambda)\,\mathrm{d}\Lambda$, and taking the limit $N_a \to \infty$ we obtain (21). An alternative point of view is to take the derivatives of (15) and (16) with respect to k_j and Λ_α, respectively, set $\mathrm{d}I/\mathrm{d}k = N_a\rho(k)$, $\mathrm{d}J/\mathrm{d}\Lambda = N_a\sigma(\Lambda)$, and take the $N_a \to \infty$ limit. In either case we obtain

$$1 = 2\pi\rho(k) + 2\cos k \int_{-B}^{B} \mathrm{d}\Lambda\sigma(\Lambda)\theta'(2\sin k - 2\Lambda)\ , \tag{20}$$

$$-2\int_{-Q}^{Q} \mathrm{d}k\, \rho(k)\theta'(2\sin k - 2\Lambda)$$

$$= 2\pi\sigma(\Lambda) - \int_{-B}^{B} \mathrm{d}\Lambda'\sigma(\Lambda')\,\theta'(\Lambda - \Lambda') \tag{21}$$

or, equivalently,

$$\boxed{\rho(k) = \frac{1}{2\pi} + \cos k \int_{-B}^{B} K(\sin k - \Lambda)\sigma(\Lambda)\,\mathrm{d}\Lambda\,,} \tag{22}$$

$$\boxed{\sigma(\Lambda) = \int_{-Q}^{Q} K(\sin k - \Lambda)\rho(k)\,\mathrm{d}k - \int_{-B}^{B} K^2(\Lambda - \Lambda')\sigma(\Lambda')\,\mathrm{d}\Lambda'\,,} \tag{23}$$

where

$$K(\Lambda - \Lambda') = -\frac{1}{\pi}\,\theta'(2\Lambda - 2\Lambda') = \frac{1}{2\pi}\left[\frac{8U}{U^2 + 16(\Lambda - \Lambda')^2}\right]\,,$$

$$K^2(\Lambda - \Lambda') = -\frac{1}{2\pi}\,\theta'(\Lambda - \Lambda') = \frac{1}{2\pi}\left[\frac{4U}{U^2 + 4(\Lambda - \Lambda')^2}\right]$$

$$= \int_{-\infty}^{\infty} K(\Lambda - x)K(x - \Lambda')\,\mathrm{d}x\,.$$

Note that K^2 is the square of K in the sense of operator products. Note also that (22) and (23) are to be satisfied only for $|k| \leqslant Q$ and $|\Lambda| \leqslant B$. Outside these intervals ρ and σ are not uniquely defined, but we *can* and will define them by the right-hand sides of (22) and (23).

The following Fourier transforms will be used in later discussions:

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega\Lambda}K(\Lambda)\,\mathrm{d}\Lambda = \mathrm{e}^{-U|\omega|/4}, \qquad \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega\Lambda}K^2(\Lambda)\,\mathrm{d}\Lambda = \mathrm{e}^{-U|\omega|/2}\,. \tag{24}$$

The ground state energy (11) now reads

$$E(M, M') = -2N_a \int_{-Q}^{Q} \rho(k)\cos k\,\mathrm{d}k\,, \tag{25}$$

where $\rho(k)$ is to be determined together with $\sigma(\Lambda)$ from the coupled integral equations (22) and (23) subject to the normalizations (19).

5. Analysis of the integral equations

In this section, we shall prove that Eqs. (22) and (23) have unique solutions for each given $0 < Q \leqslant \pi$ and $0 < B \leqslant \infty$ and that the solutions are positive and have certain monotonicity properties. These properties guarantee that the normalization

conditions (19) uniquely determine values of Q, B for each given value of N when $M = M' = N/2$ (in this case we have $B = \infty$). However, we have not proved uniqueness of Q, B when $M \neq M'$ (although we believe there is uniqueness). But this does not matter for the absolute ground state since, as remarked earlier, the ground state has $S = 0$ (in the thermodynamic limit) and so we are allowed to take $S^z = 0$. For $M \neq M'$, we have remarked earlier that the solution probably has $S = |M' - M|/2$ and is the ground state for $S = |M' - M|/2$.

An important first step is to overcome the annoying fact (which is relevant for $Q > \pi/2$) that $\sin k$ is not a monotonic function of k in $[-\pi, \pi]$. To do this we note that $(\cos k)K(\sin k - \Lambda)$ is an odd function of $k - \pi/2$ (for each Λ) and hence $\rho(k) - 1/2\pi$ also has this property. On the other hand, $K(\sin k - \Lambda)$ appearing in (23) is an even function of $k - \pi/2$. As a result $\rho(k)$ appearing in the first term on the right-hand side of (23) can be replaced by $1/2\pi$ in the intervals $Q' < k < Q$ and $-Q < k < -Q'$, where $Q' = \pi - Q$. Thus, when $Q > \pi/2$, we can rewrite the $[Q', Q]$ portion of the first integral in (23) as

$$\int_{Q'}^{Q} K(\sin k - \Lambda)\rho(k)\,\mathrm{d}k = \int_{Q'}^{Q} K(\sin k - \Lambda)\,\frac{1}{2\pi}\,\mathrm{d}k$$

$$= \frac{2}{2\pi}\int_{Q'}^{\pi/2} K(\sin k - \Lambda)\,\mathrm{d}k\ .$$

A similar thing can be done for the $[-Q, -Q']$ portion and for the corresponding portions of (19).

The integrals over k now extend at most over the interval $[-\pi/2, \pi/2]$, in which $\sin k$ is monotonic.

We are now in a position to change variables as follows. For $-1 \leqslant x \leqslant 1$ let

$$t(x) = \frac{1}{2\pi}(1 - x^2)^{-1/2}, \quad f(x) = (1 - x^2)^{-1/2}\rho(\sin^{-1}x)\ . \tag{26}$$

In case $Q < \pi/2$, $\rho(\sin^{-1}x)$ is defined only for $\sin x \leqslant Q$, but we shall soon see (after (28)) how to extend the definition of f in this case. We define the step functions for all real x by

$$\begin{aligned} &B(x) = 1, \quad |x| < B, && = 0, \quad \text{otherwise} \\ &A(x) = 1, \quad |x| < a, && = 0, \quad \text{otherwise} \\ &D(x) = H(Q), \quad a < |x| < 1, && = 0, \quad \text{otherwise}\ , \end{aligned} \tag{27}$$

where $a = \sin Q = \sin Q'$ and where

$$H(Q) = 0 \quad \text{if } Q \leqslant \pi/2, \qquad H(Q) = 2 \quad \text{if } Q > \pi/2\ .$$

The integral equations (22) and (23) become

$$f(x) = t(x) + \int_{-\infty}^{\infty} K(x - x')B(x')\sigma(x')\,\mathrm{d}x', \quad |x| \leqslant a\ , \tag{28}$$

$$\sigma(x) = \int_{-\infty}^{\infty} K(x-x')A(x')f(x')\,\mathrm{d}x' + \int_{-\infty}^{\infty} K(x-x')D(x')t(x')\,\mathrm{d}x'$$
$$- \int_{-\infty}^{\infty} K^2(x-x')B(x')\sigma(x')\,\mathrm{d}x', \quad |x| < B\,. \tag{29}$$

Although these equations have to be solved in the stated intervals we can use their right-hand sides to define their left-hand sides for *all* real x. We define $t(x) \equiv 0$ for $|x| > 1$. It is obvious that the extended equations have (unique) solutions if and only if the original ones do. Henceforth, we shall understand the functions f and σ to be defined for all real x.

These equations read, in operator form

$$f = t + \hat{K}\hat{B}\,\sigma\,, \tag{30}$$

$$\sigma = \hat{K}\hat{A}\,f + \hat{K}\hat{D}\,t - \hat{K}^2\hat{B}\,\sigma\,, \tag{31}$$

where $\hat{K}$ is convolution with K and $\hat{A}, \hat{B}, \hat{D}$ are the multiplication operators corresponding to A, B, D (and which are also projections since $\hat{A}^2 = \hat{A}$, etc.).

In view of the normalization requirements (19), the space of functions to be considered is, obviously, $L^1([-a,a])$ for f and $L^1([-B,B])$ for σ. (L^p is the pth power integrable functions and L^∞ is the bounded functions.) Since $K(x)$ is in $L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, it is a simple consequence of Young's inequality that the four integrals in (28) and (29) are automatically in $L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ when $f \in L^1([-a,a])$ and $\sigma \in L^1([-B,B])$. In particular, the integrals are in $L^2(\mathbb{R})$, which allows us to define the operators in (30), (31) as bounded operators on $L^2(\mathbb{R})$. In addition, t is in $L^1(\mathbb{R})$, but not in $L^2(\mathbb{R})$. To summarize, solutions in which f and σ are in $L^1(\mathbb{R})$ automatically have the property that $f - t$ and σ are both in $L^p(\mathbb{R})$ for all $1 \leqslant p \leqslant \infty$.

Theorem 1 (Uniqueness). *The solutions $f(x)$ and $\sigma(x)$ are unique and positive for all real x.*

Remark. The uniqueness implies that f and σ are even functions of x (because the pair $f(-x), \sigma(-x)$ is also a solution). The theorem implies (from the definition (26)) that $\sigma(\Lambda) > 0$ for all real Λ and it implies that $\rho(k) > 0$ for all $|k| \leqslant \pi/2$. It does *not* imply that $\rho(k)$, defined by the right-hand side of (22), is non-negative for all $|k| > \pi/2$. We shall prove this positivity, however, in Lemma 3. Note that the positivity of ρ is equivalent to the statement that $f(x) < 2t(x)$ for all $|x| \leqslant 1$ because, from (22) and the evenness or σ, $\rho(\pi - k) = (1/\pi) - \rho(k)$.

Proof. By substituting (30) into (31) and rearranging slightly we obtain

$$(1 + \hat{K}^2)\sigma = \hat{K}(\hat{A} + \hat{D})t + \hat{K}^2(1 - \hat{B})\sigma + \hat{K}\hat{A}\hat{K}\hat{B}\sigma\,. \tag{32}$$

Since $\hat{K}^2$ is positive definite, $1 + \hat{K}^2$ has an inverse $1/(1 + \hat{K}^2)$, which we can apply to both sides of (32). The convolution operator

$$\hat{R} = \hat{K}(1 + \hat{K}^2)^{-1} \tag{33}$$

has a Fourier transform $\frac{1}{2}\mathrm{sech}(\omega/4)$. The inverse Fourier transform is proportional to $\mathrm{sech}(2\pi x)$ (see (55) below), which is positive. In other words, $\hat{R}$ is not only a positive operator, it also has a positive integral kernel.

We can rewrite (32) as

$$(1-\hat{W})\sigma = \hat{R}(\hat{A}+\hat{D})t \equiv \xi \tag{34}$$

with

$$\hat{W} = \hat{R}\hat{K}(1-\hat{B}) + \hat{R}\hat{A}\hat{K}\hat{B} = \hat{R}[\hat{K} - (1-\hat{A})\hat{K}\hat{B}]\,. \tag{35}$$

The middle expression shows that the integral kernel of $\hat{W}$ is positive.

Clearly, $\xi > 0$ as a function and $\xi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Also, $\hat{W}$ has a positive integral kernel. We note that $\|\hat{R}\| = 1/2$ on $L^2(\mathbb{R})$ since $y/(1+y) \leqslant 1/2$ for $y \geqslant 0$. Also, $\|\hat{K}\| = 1$, and $\|1-\hat{B}\| = 1$, $\|\hat{A}\| = 1$, $\|\hat{B}\| = 1$. In fact, it is easy to check that $\|\hat{R}\hat{A}\hat{K}\hat{B}\| < 1/2$. From this we conclude that $\|\hat{W}\| < 1$ on $L^2(\mathbb{R})$ and thus $1-\hat{W}$ has an inverse (as a map from $L^2(\mathbb{R}) \to L^2(\mathbb{R})$).

Therefore, we can solve (32) by iteration:

$$\sigma = (1 + \hat{W} + \hat{W}^2 + \hat{W}^3 + \cdots)\xi\,. \tag{36}$$

This is a strongly convergent series in $L^2(\mathbb{R})$ and hence (36) solves (32) in $L^2(\mathbb{R})$. It is the unique solution because the homogeneous equation $(1-\hat{W})\phi = 0$ has no solution. Moreover, since each term is a positive function, we conclude that σ is a positive function as well. □

Lemma 1 (Monotonicity in B). *When B increases with Q fixed, $\sigma(x)$ decreases pointwise for all $x \in \mathbb{R}$.*

Proof. Since $1-\hat{A}$ is fixed and positive, we see from the right-hand side of (35) that the integral kernel of $\hat{W}$ is monotone decreasing in $\hat{B}$. The lemma then follows from the representation (36). □

Lemma 2 (Monotonicity in B). *When B increases with Q fixed, $f(x)$ increases pointwise for all $x \in \mathbb{R}$. This implies, in particular, that $\rho(k)$ increases for all $|k| \leqslant \pi/2$ and decreases for all $\pi/2 \leqslant |k| \leqslant \pi$.*

Proof. Consider Eq. (32) for the case $A = 0$. Theorem 1 and Lemma 1 hold in this case, of course. We also note that their proofs do not depend on any particular fact about the function Dt, other than the fact that it is a non-negative function. From these observations we learn that the solution to the equation

$$(1 + \hat{K}^2\hat{B})S = \hat{K}g \tag{37}$$

has the property, for all $x \in R$, that $S(x) \geqslant 0$ and that $S(x)$ is a non-increasing function of B, provided only that $g(x) \geqslant 0$ for all $x \in \mathbb{R}$.

Another way to say this is that the integral kernel of $\hat{V} = (1 + \hat{K}^2\hat{B})^{-1}\hat{K}$ is positive and is a pointwise monotone decreasing function of B.

Now let us rewrite (37) as

$$\hat{K}^2(1 + \hat{K}^2)^{-1}g = \hat{U}g + \hat{K}(1 + \hat{K}^2)^{-1}(1 - \hat{B})S \tag{38}$$

with

$$\hat{U} = \hat{K}\hat{B}\hat{V} = \hat{K}\hat{B}(1 + \hat{K}^2\hat{B})^{-1}\hat{K}\ . \tag{39}$$

The operator $\hat{U}$ has a positive integral kernel since $\hat{V}$, $\hat{K}$, and $\hat{B}$ have one. As B increases the second term on the right-hand side of (38) decreases pointwise (because $(1 - \hat{B})$ decreases as a kernel and S decreases, as we have just proved). The left-hand side of (38) is independent of B and, therefore, the first term on the right-hand side of (38) must increase pointwise. Since this holds for arbitrary positive g, we conclude that the integral kernel of $\hat{U}$, in contrast to that of $\hat{V}$, is a pointwise *increasing* function of B.

Having established the monotonicity property of $\hat{U}$ let us return to f, which we can write (from (30)) as

$$f = (1 + \hat{U}\hat{D})t + \hat{U}\hat{A}f \tag{40}$$

$$= [1 + \hat{U}\hat{A} + (\hat{U}\hat{A})^2 + \cdots](1 + \hat{U}\hat{D})t\ . \tag{41}$$

The series in (41) is strongly convergent (since $\|\hat{A}\| = 1$ and $\|\hat{U}\| \leqslant 1/2$) and thus defines the solution to (40). Since $\hat{U}$ is monotone in $\hat{B}$, (41) tells us that f is also pointwise monotone, as claimed.

Eq. (26) tells us that $\rho(k)$ is increasing in B for $|k| \leqslant \pi/2$ and is decreasing in B for $\pi/2 \leqslant |k| \leqslant \pi$. □

Theorem 2 (Monotonicity in B). *When B increases with Q fixed, N/N_a and M/N increase. When $B = \infty$, we have $2M = N$, and when $B < \infty$ we have $2M < N$ (for all Q).*

Proof. The integral for N/N_a in (19) can be written as $\int_{-\infty}^{\infty}[A\rho + (1/2\pi)D]$, and this is monotone increasing in B since ρ is monotone for $|k| \leqslant \pi/2$ and $A(k) = 0$ for $|k| > \pi/2$.

If we integrate (23) from $\Lambda = -\infty$ to ∞, and use the fact that $\int K = 1$ from (24), we obtain

$$\frac{N}{N_a} = \int_{-Q}^{Q}\rho(k)\,\mathrm{d}k = \int_{-\infty}^{\infty}\sigma(\Lambda)\,\mathrm{d}\Lambda + \int_{-B}^{B}\sigma(\Lambda)\,\mathrm{d}\Lambda \tag{42}$$

which becomes, after making use of the normalization (19)

$$1 = 2\frac{M}{N} + \frac{N_a}{N}\left[\int_{-\infty}^{-B} + \int_{B}^{\infty}\right]\sigma(\Lambda)\,\mathrm{d}\Lambda\ . \tag{43}$$

Now the integrals in (43) decrease as B increases by Lemma 1 and converge to 0 as $B \to \infty$, while N/N_a increases, as we have just proved. It follows that M/N increases monotonically with B, reaching $M/N = 1/2$ at $B = \infty$. If $B < \infty$ then $M/N < 1/2$ since σ is a strictly positive function. □

We turn now to the dependence of σ, ρ on Q, with fixed B. First, Lemma 3 (which was promised in the remark after Theorem 1) is needed.

Lemma 3 (Positivity of ρ). *For all $B \leqslant \infty$, all $Q \leqslant \pi$, and all $|k| \leqslant \pi$, we have $\rho(k) > 0$.*

Proof. As mentioned in the Remark after Theorem 1, the positivity of ρ is equivalent to the statement that $f(x) < 2\ t(x)$ for all $|x| \leqslant 1$. We shall prove $f(x) < 2t(x)$ here.

Owing to the monotonicity in B of f (Lemma 2) it suffices to prove the lemma for $B = \infty$, which we assume now. We see from (41) that for any given value of a the worst case is $Q > \pi/2$, whence $H(Q) = 2$ and $D > 0$. We assume this also.

For the purpose of this proof (only) we denote the dependence of $f(x)$ on a by $f_a(x)$.

We first consider the case $a=0$, corresponding to $Q=\pi$. Let us borrow some information from the next section, where we actually solve the equations for $B = \infty$, $Q = \pi$ and discover (Lemma 5) that $f(x) < 2t(x)$ for $|x| \leqslant 1$ (for $U > 0$).

We see from (40) or (41) that f_a is continuous in a and differentiable in a for $0 < a < 1$ (indeed, it is real analytic). Also, since the kernel $K(x - y)$ is smooth in (x, y) and $t(x)$ is smooth in $x \in (-1, 1)$, it is easy to see that f_a is smooth, too, for $x \in (-1, 1)$. Eq. (28) defines $f_a(x)$ pointwise for all x and $f_a(x)$ is jointly continuous in a, x.

In detail, (40) reads

$$f_a(x) = t(x) + 2\left[\int_{-1}^{-a} + \int_{a}^{1}\right] U(x,x')t(x')\,\mathrm{d}x' + \int_{-a}^{a} U(x,x')f_a(x')\,\mathrm{d}x' \,. \tag{44}$$

Take the derivative with respect to a and set $h_a(x) = \partial f_a(x)/\partial a$. Observe that $\hat{U}$ does not depend on a. We obtain

$$h_a(x) = U(x,a)[f_a(a) - 2t(a)] + U(x,-a)[f_a(-a) - 2t(-a)] + \int_{-a}^{a} U(x,x')h_a(x')\,\mathrm{d}x' \,. \tag{45}$$

(This equation makes sense because $f_a(x)$ is jointly continuous in x, a and $t(x)$ is continuous for $|x| < 1$. Recall that f and t are even functions of x. Note that U here is the kernel of (39) with $B = \infty$, i.e., $\hat{U} = \hat{K}^2(1 + \hat{K}^2)^{-1}$, which is self-adjoint and positive as an operator and positive as a kernel.)

Eq. (45) can be iterated in the same manner as (41) (since $\|\hat{U}\| = 1/2$)

$$h_a(x) = [\hat{U} + \hat{U}\hat{A}\hat{U} + \hat{U}\hat{A}\hat{U}\hat{A}\hat{U} + \cdots](x,a)F(a) \equiv T(x,a)F(a) \,, \tag{46}$$

where $[\cdot](x,a)$ denotes the integral kernel of $\hat{T}=[\cdot]$, and where $F(a)=f_a(a)-2t(a)$ is a number. As an operator. $\hat{T}$ is self-adjoint and positive.

Now $\hat{U}$ has a positive kernel and thus $T(x,a)\geqslant 0$, so $h_a(x)<0$ for all x if and only if $F(a)<0$. We have already noted that $F(0)<0$.

We can integrate (46) to obtain

$$f_a(x)=f_0(x)+\int_0^a h_{a'}(x)\,\mathrm{d}a'=f_0(x)+\int_0^a T(x,a')[f_{a'}(x)-2t(a')]\,\mathrm{d}a'\ . \quad (47)$$

If we subtract $2t(a)$ from this and set $x=a$, we obtain

$$F(a)=G(a)+\int_0^a T(a,a')F(a')\,\mathrm{d}a'\ , \quad (48)$$

where $G(a)=f_0(a)-2t(a)<0$. Another way to state (48) is $F=G+\hat{T}\hat{A}F$.

Eq. (48) implies that $F(a)<0$ for all a, as desired. There are two ways to see this. One way is to note that $\hat{T}$ is monotone increasing in a (as an operator and as a kernel), so $\hat{T}\leqslant\hat{U}+\hat{U}^2+\cdots=\hat{K}^2<1$. Therefore, (48) can be iterated as $F=[1+\hat{T}\hat{A}+\hat{T}\hat{A}\hat{T}\hat{A}+\cdots]G$, and this is negative. The second way is to note that $f_a(a)$ (and hence $F(a)$) is continuous in a. Let a^* be the smallest a for which $F(a)=0$. Then, from (48), $0=F(a^*)=G(a^*)+\int_0^{a^*}T(a^*,a')F(a')\,\mathrm{d}a'<0$, which is a contradiction.

From $F(a)<0$ we can deduce that $f_a(x)-2t(x)<0$ for all $|x|\leqslant 1$. Simply subtract $2t(x)$ from both sides of (47). Then $f_a(x)-2t(x)=\{f_0(x)-2t(x)\}+(\hat{T}\hat{A}F)(x)$. The first term $\{\}<0$ by Lemma 5, which we prove in Section 6 below, and the second term is <0 (since $F<0$). □

Lemma 4 (Monotonicity in Q). *Consider the dependence of the solution to* (30), (31) *on the parameter* $0\leqslant a\leqslant 1$ *for fixed* $B\leqslant\infty$. *For* $Q\leqslant\pi/2$ (*i.e.*, $H(Q)=D=0$), *both* f *and* σ *increase pointwise as* a *increases. For* $Q>\pi/2$ (*i.e.*, $H(Q)=2, Dt=2(1-A)t$), *both* f *and* σ *decrease pointwise as* a *increases.*

If, instead of the dependence on a, *we consider the dependence on* $0\leqslant Q\leqslant\pi$ *of* $\rho(k)$ (*which is defined by* (22) *for all* $|k|\leqslant\pi$) *and of* $\sigma(\Lambda)$ (*which is defined by* (23) *for all real* Λ), *then, as* Q *increases*

$$\rho(k)\ \textit{increases for}\ 0\leqslant|k|<\pi/2\ \textit{and decreases for}\ \pi/2\leqslant|k|\leqslant\pi$$

$$\sigma(\Lambda)\ \textit{increases for all real}\ \Lambda\ . \quad (49)$$

Proof. Concerning the monotonicities stated in the second part of the lemma, (49), we note that as Q goes from 0 to $\pi/2$, *a increases* from 0 to 1, but when Q goes from $\pi/2$ to 0, *a decreases* from 1 to 0. Moreover, $H(Q)=0$ in the first case and $H(Q)=2$ in the second case. This observation shows that the first part of the lemma implies the statement about σ in (49). The statement about ρ in (49) also follows, if we take note of the $\cos k$ factor in (49).

We now turn to the first part of the lemma. The easy case is $Q \leqslant \pi/2$ or $H(Q)=0$. Then (41) does not have the $\hat{U}\hat{D}t$ term and, since $\hat{U}$ has a positive kernel and since $\hat{A}$ has a kernel that increases with a, we see immediately that f increases with a. Likewise, from (34), (35), we see that $\hat{W}$ and ξ increase with a and, from (36), we see that σ increases.

For $Q > \pi/2$ or $H(Q) = 2$, we proceed as in the proof of Lemma 3 by defining $h_a(x) = \partial f_a(x)/\partial a$ and proceeding to (46) (but with $\hat{U}$ given by (39)). This time we know that $F(a) < 0$ (by Lemma 3) and hence $h_a(x) < 0$, as claimed. The monotonicity of $\sigma(x)$ follows by differentiating (29) with respect to a. Then $(\partial\sigma(x)/\partial a)=(\hat{V}\hat{A}h_a)(x)+V(x,a)F(a)$, where $V(x,y)$ is the kernel of $\hat{V}$, which is positive, as noted in the proof of Lemma 2. □

Theorem 3 (Monotonicity in Q). *When Q increases with fixed B, N/N_a and M/N_a increase. When $Q=\pi$, $N/N_a=1$ (for all B), while $N/N_a<1$ if $Q<\pi$.*

Proof. From (42), $N/N_a = 2\int_{-B}^{B}\sigma + 2\int_{B}^{\infty}\sigma$ and this increases with Q by (49). Also, by (42), $N/N_a = \int_{-Q}^{Q}\rho$. When $Q=\pi$, we see from (22) that $\int_{-Q}^{Q}\rho = \int_{-\pi}^{\pi}(1/2\pi) = 1$, so $N/N_a = 1$. To show that $N/N_a < 1$ when $Q<\pi$ we use the monotonicity of σ with respect to B (Lemma 2) and Q (Lemma 3) (with $\sigma_0(\Lambda)$= the value of $\sigma(\Lambda)$ for $B=\infty$, $Q=\pi$) to conclude that $N/N_a \leqslant 2\int_{-B}^{B}\sigma_0 + 2\int_{B}^{\infty}\sigma_0 = \int_{-\infty}^{\infty}\sigma_0 = 1-2\int_{B}^{\infty}\sigma_0 < 1$, since σ_0 is a strictly positive function.

Finally, from (42) we have that $M/N_a=\int_{-B}^{B}\sigma$, and this increases with Q by (49). □

6. Solution for the half-filled band

In the case of a half-filled band, we have $N=N_a$, $M=M'=N/2$ and, from Theorems 2 and 3, $Q=\pi$, $B=\infty$. In this case the integral equations (22) and (23) can be solved. We use the notation $\rho_0(k)$ and $\sigma_0(\Lambda)$ for these solutions.

Substituting (22) into (23) where, as explained earlier, we use $\rho_0(k)=1/2\pi$ in the first term on the right-hand side of (23). Then the integral equation (23) involves only $\sigma_0(\Lambda)$ and can be solved by Fourier transform. Using equations (24) it is straightforward to obtain the solution for σ_0 and its Fourier transform $\hat{\hat{\sigma}}_0$ as

$$\hat{\hat{\sigma}}_0(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega\Lambda}\sigma_0(\Lambda)\,\mathrm{d}\Lambda = \frac{J_0(\omega)}{2\cosh(U\omega/4)}\,, \tag{50}$$

$$\boxed{\sigma_0(\Lambda) = \frac{1}{2\pi}\int_0^{\infty}\frac{J_0(\omega)\cos(\omega\Lambda)}{\cosh(\omega U/4)}\,\mathrm{d}\omega\,,} \tag{51}$$

where

$$J_0(\omega) = \frac{2}{\pi}\int_0^{\pi/2}\cos(\omega\cos\theta)\,\mathrm{d}\theta = \frac{1}{\pi}\int_0^{\pi}\cos(\omega\sin\theta)\,\mathrm{d}\theta \tag{52}$$

is the zeroth order Bessel function.

Next we substitute (51) into (22) and this leads (with (24)) to

$$\rho_0(k) = \frac{1}{2\pi} + \frac{\cos k}{\pi} \int_0^\infty \frac{\cos(\omega \sin k) J_0(\omega)}{1 + e^{\omega U/2}} \, d\omega \, . \tag{53}$$

The substitution of (53) into (25) finally yields the ground state energy, E_0, of the half-filled band as

$$E_0\left(\frac{N_a}{2}, \frac{N_a}{2}\right) = -4N_a \int_0^\infty \frac{J_0(\omega) J_1(\omega)}{\omega(1 + e^{\omega U/2})} \, d\omega \, , \tag{54}$$

where $J_1(\omega) = \pi^{-1} \int_0^\pi \sin(\omega \sin p) \sin p \, dp = \omega \pi^{-1} \int_0^\pi \cos(\omega \sin p) \cos^2 p \, dp$ is the Bessel function of order one.

Remarks. (A) When there is no interaction ($U = 0$), $\hat{K}$ is a δ-function; we can evaluate (51) and (53) as

$$\sigma_0(\Lambda) = \frac{1}{2\pi\sqrt{1 - \Lambda^2}}, \quad |\Lambda| \leqslant 1; \quad = 0, \text{ otherwise} \, ,$$

$$\rho_0(k) = \frac{1}{\pi}, \quad |k| \leqslant \frac{\pi}{2}; \quad = 0, \text{ otherwise} \, .$$

This formula for $\rho_0(k)$ agrees with what is expected for an ideal Fermi gas.

(B) The $U \to \infty$ limit is peculiar. From (50) we see that $\hat{\hat{\sigma}}_0(0) = \frac{1}{2}$, so $\int \sigma_0 = \frac{1}{2}$, but from (51) we see that $\sigma_0(\Lambda) \to 0$ in this limit, uniformly in Λ. On the other hand $\rho_0(k) \to \frac{1}{2}\pi$, for all $|k| \leqslant \pi$, which is what one would expect on the basis of the fact that this 'hard core' gas becomes, in effect, a one-component ideal Fermi gas of $N = N_a$ particles.

We now derive alternative, more revealing expressions for σ_0, ρ_0.

For $\sigma_0(\Lambda)$ we substitute the integral representation (52) for J_0 into (51) and recall the Fourier cosine transform (for $\alpha > 0$)

$$\int_0^\infty \frac{\cos(\omega x)}{\cosh(\omega \alpha)} \, d\omega = \left(\frac{\pi}{2\alpha}\right) \frac{1}{\cosh(\pi x / 2\alpha)} \, . \tag{55}$$

Then, using $2 \cos a \cos b = \cos(a - b) + \cos(a + b)$ we obtain

$$\sigma_0(\Lambda) = \frac{1}{\pi U} \int_0^\pi \frac{d\theta}{\cosh[2\pi(\Lambda + \cos\theta)/U]} > 0 \, . \tag{56}$$

An alternate integral representation can be derived similarly for $\rho_0(k)$, but the derivation and the result is more complicated. We substitute $(1 + e^x)^{-1} = \sum_{n=1}^\infty (-1)^n \exp[-nx]$, with $x = \omega U/2$, into (53) and make use of the identity (with $\alpha = -is \pm c$ in the notation

of Gradshteyn and Ryzhik [21, 6.611.1])

$$2\int_0^\infty e^{-c\omega}J_0(\omega)\cos(\omega s)\,d\omega = [(-c-is)^2+1]^{-1/2}+[(c-is)^2+1]^{-1/2} \quad (57)$$

for $c>0$, and where the square roots $[\quad]^{-1/2}$ in (57) are taken to have a positive real part. This leads to

$$\rho_0(k)=\frac{1}{2\pi}+\frac{\cos k}{2\pi}\sum_{n=1}^{\infty}(-1)^{n+1}\{[(-nU/2-i\sin k)^2+1]^{-1/2} + [(nU/2-i\sin k)^2+1]^{-1/2}\}\,. \quad (58)$$

We can rewrite the sum of the two terms in (58) as a single sum from $n=-\infty$ to ∞, after making a correction for the $n=0$ term (which equals $\cos k/|\cos k|$ for $k\neq\pi/2$). We obtain

$$\rho_0(k)=\frac{1}{2\pi}\left[1+\frac{\cos k}{|\cos k|}\right]-\frac{\cos k}{2\pi}\sum_{n=-\infty}^{\infty}(-1)^n[(nU/2-i\sin k)^2+1]^{-1/2} = \frac{1}{2\pi}\left[1+\frac{\cos k}{|\cos k|}\right]-\frac{\cos k}{2\pi}\,\frac{1}{2\pi i}\int_C \frac{dz}{\sqrt{(zU/2-i\sin k)^2+1}}\,\frac{\pi}{\sin(\pi z)}\,. \quad (59)$$

The contour C encompasses the real axis, i.e., it runs to the right just below the real axis and to the left just above the real axis.

The integrand has two branch points $y_\pm$ on the imaginary axis, where $y_\pm=(2i/U)\times(\sin k\pm 1)$. In order to have the correct sign of the square root in the integrand we define the branch cuts of the square root to extend along the imaginary axis from y_+ to $+\infty$ and from y_- to $-\infty$. We then deform the upper half of the contour C into a contour that runs along both sides of the upper branch cut and in two quarter circles of large radius down to the real axis. In a similar fashion we deform the lower half of C along the lower cut. As the radius of the quarter circles goes to ∞ this gives rise to the following expression:

$$\boxed{\rho_0(k)=\frac{1}{2\pi}\left[1+\frac{\cos k}{|\cos k|}\right]-\frac{\cos k}{2\pi U}[I_-(k)+I_+(k)]>0\,,} \quad (60)$$

where

$$I_\pm(k)=\int_{1\pm\sin k}^{\infty}\frac{d\alpha}{\sinh(2\pi\alpha/U)\sqrt{(\alpha\mp\sin k)^2-1}}\,. \quad (61)$$

By introducing the variable $\alpha=\cosh x\pm\sin k$ we finally obtain the simple expression

$$I_\pm(k)=\int_0^\infty\frac{dx}{\sinh\{(2\pi/U)(\cosh x\pm\sin k)\}}\,. \quad (62)$$

As a consequence of expressions (60) and (62) for ρ_0, we have the crucial bound needed as input at the end of the proof of Lemma 3:

Lemma 5 (ρ bounds). *When* $B=\infty$, $Q=\pi$, *and* $U>0$

$$1/2\pi < \rho_0(k) < 1/\pi \quad \textit{for } 0 \leqslant |k| < \pi/2\,,$$

$$0 < \rho_0(k) < 1/2\pi \quad \textit{for } \pi/2 < |k| \leqslant \pi\,. \tag{63}$$

Equivalently, $f_0(x) < 2t(x)$ *for all* $|x| \leqslant 1$.

Proof. When $\pi/2 < |k| \leqslant \pi$ and $\cos k < 0$ the first term [ ] in (60) is zero while the second term is positive (since $I_{\pm}(k) > 0$). On the other hand, when $0 \leqslant |k| \leqslant \pi/2$, Theorem 1 shows that $\rho_0(k) > 0$. Thus, we conclude that $\rho_0(k) > 0$ for all $|k| \leqslant \pi$.

From (22) and the positivity of σ_0 we conclude that $\rho_0(k) < 1/2\pi$ when $\pi/2 < |k| \leqslant \pi$. From the positivity of $\rho_0(k)$ when $\pi/2 \leqslant |k| \leqslant \pi$ we conclude that the integral in (22) is less than $1/2\pi$ for all values of $0 \leqslant \sin k < 1$ and, therefore, $1/2\pi < \rho_0(k) < 1/\pi$ for $0 \leqslant |k| < \pi/2$. □

7. Absence of a Mott transition

A system of itinerant electrons exhibits a Mott transition if it undergoes a conducting–insulating transition when an interaction parameter is varied. In the Hubbard model one inquires whether a Mott transition occurs at some critical $U_c > 0$. Here we show that there exists no Mott transition in the 1D Hubbard model for all $U > 0$.

Our strategy is to compute the chemical potential μ_+ (resp. μ_-) for adding (resp. removing) one electron. The system is conducting if $\mu_+ = \mu_-$ and insulating if $\mu_+ > \mu_-$.

In the thermodynamic limit we can define μ by $\mu = \mathrm{d}E(N)/\mathrm{d}N$, where $E(N)$ denotes the ground state energy with $M = M' = N/2$. As we already remarked, this choice gives the ground state energy for all U, at least in the thermodynamic limit.

The thermodynamic limit is given by the solution of the integral equations, which we analyzed in Section 5. In this limit one cannot distinguish the odd and even cases (i.e., $M = M' = N/2$ if N is even or $M = M' - 1 = (N-1)/2$ if N is odd.) and one simply has $M/N = 1/2$ in the limit $N_a \to \infty$. In this case Theorem 2 says that we must have $B = \infty$. Then only Q is a variable and Theorem 3 says that Q is uniquely determined by N provided $N \leqslant N_a$.

In the thermodynamic limit we know, by general arguments, that $E(N)$ has the form $E(N) = N_a e(N/N_a)$ and e is a convex function of N/N_a. It is contained in (25) when $N/N_a \leqslant 1$. A convex function has right and left derivatives at every point and, therefore, $\mu_+ =$ right derivative and $\mu_- =$ left derivative are well defined. Convexity implies that $\mu_- \leqslant \mu_+$.

For less than a half-filled band it is clear that $\mu_+ = \mu_-$ since $E(M,M)$ is smooth in $M = N/2$ for $N \leqslant N_a$. The chemical potential cannot make any jumps in this region. But, for $N > N_a$ we have to use hole–particle symmetry as discussed in Section 1 to calculate $E(N)$. The derivatives of $E(N)$, namely μ_+ and μ_-, can now be different

above and below the half-filling point $N = N_a$ and this gives rise to the possibility of having an insulator. We learn from (3) that

$$\mu_+ + \mu_- = U \tag{64}$$

and hence $\mu_+ > \mu_-$ if $\mu_- < U/2$.

We calculate μ_- in two ways, and arrive at the same conclusion

$$\boxed{\mu_-(U) = 2 - 4\int_0^\infty \frac{J_1(\omega)}{\omega[1+\exp(\omega U/2)]}\,\mathrm{d}\omega\,.} \tag{65}$$

The first way is to calculate μ_- from the integral equations by doing perturbation theory at the half-filling point analyzed in Section 6. This is a 'thermodynamic' or 'macroscopic' definition of μ_- and it is given in Section 7.1 below. (From now on μ_- means the value at the half-filling point.)

In Section 7.2 we calculate μ_- 'microscopically' by analyzing the Bethe Ansatz directly with $N = N_a - 4$ electrons. Not surprisingly, we find the same value of μ_-. This was the method we originally employed to arrive at [1, Eq. (23)].

Before proceeding to the derivations of (65), we first show that (65) implies $\mu_- < U/2$ for every $U > 0$. To see that $\mu_- < U/2$ we observe that

$$\mu_-(0) = 2 - 2\int_0^\infty \frac{J_1(\omega)}{\omega}\,\mathrm{d}\omega = 0\,, \tag{66}$$

$$u'_-(0) = \frac{1}{2}\int_0^\infty J_1(\omega)\,\mathrm{d}\omega = \frac{1}{2}\,. \tag{67}$$

Then $\mu_- < U/2$ holds if $\mu''_-(U) < 0$, which we turn to next. Here, (66) is in [21, 6.561.17] and (67) is in [21, 6.511.1].

Expanding the denominator in the integrand of (88) and integrating term by term, we obtain

$$\mu_-(U) = 2 - 4\sum_1^\infty (-1)^n \left[\sqrt{1+\frac{n^2U^2}{4}} - \frac{nU}{2}\right]$$

using which one obtains

$$\begin{aligned}\mu''_-(U) &= 2\sum_{-\infty}^{\infty}(-1)^n \frac{n^2}{(1+n^2U^2/4)^{3/2}} \\ &= \frac{2}{2\pi\mathrm{i}}\int_C \frac{z^2}{(1+U^2z^2/4)^{3/2}}\frac{\pi}{\sin\pi z}\,\mathrm{d}z\,,\end{aligned} \tag{68}$$

where we have again replaced the summation by a contour integral with the contour C encompassing the real axis. The integrand in (68) is analytic except at the poles on the real axis and along two branch cuts on the imaginary axis. This allows us to

deform the path to coincide the imaginary axis, thereby picking up contributions from the cuts. This yields

$$\mu_-''(U) = -\frac{32}{U^3}\int_1^\infty \frac{y^2}{(y^2-1)^{3/2}}\frac{1}{\sinh(2\pi y/U)}\,\mathrm{d}y < 0 \quad \text{for all } U>0\,. \tag{69}$$

Thus, we have established $\mu_+(U) > \mu_-(U)$, and hence the 1D Hubbard model is insulating for all $U>0$. *There is no conducting–insulating transition in the ground state of the 1D Hubbard model* (except at $U=0$).

7.1. Chemical potential from the integral equations

As noted, we take $B=\infty$ and $Q<\pi$. In fact we take $Q=\pi-a$ with a small. (In the notation of Section 5, $a=\sin Q$, but to leading order in a, $\sin Q=\pi-Q$ and we need not distinguish the two numbers.) Our goal is to calculate δE, the change in E using (25) and δN, the change in N using (19); μ_- is the quotient of the two numbers.

As before, we use the notation $\rho(k)$ for the density at $Q=\pi-a$ and $\rho_0(k)$ for the density at $Q=\pi$, as given in (53), (60).

We start with N. As explained earlier, $\rho-1/2\pi$ is odd around $\pi/2$ so, from (19),

$$\begin{aligned}\frac{N}{N_a} &= \int_{-Q}^{Q}\rho = 2\int_0^Q\rho = 2\int_0^a\rho + 2\int_a^{\pi-a}\frac{1}{2\pi}\\ &= 2\int_0^a\rho + \frac{1}{\pi}(\pi-2a) \approx 1+2a\left(\rho_0(0)-\frac{1}{\pi}\right)\,.\end{aligned} \tag{70}$$

In the last expression we used the fact (and will use it again) that ρ is continuous in k and a (as we see from (41)); therefore, we can replace $\int_0^a\rho$ by $a\rho_0(0)$ to leading order in a. We learn from (70) that $\delta N/N_a = 2a(\rho_0(0)-1/\pi)<0$.

The calculation of δE is harder. From (25)

$$\begin{aligned}\frac{E}{N_a} &= -4\int_0^Q\rho\cos k = -4\int_0^a\rho\cos k - 4\int_a^{\pi-a}\rho\cos k\\ &\approx -4a\rho_0(0) - 4\int_a^{\pi-a}\left(\rho-\frac{1}{2\pi}\right)\cos k - \frac{2}{\pi}\int_a^{\pi-a}\cos k\\ &= -4a\,\rho_0(0) - 8\int_a^{\pi/2}\left(\rho-\frac{1}{2\pi}\right)\cos k\\ &= -4a\rho_0(0) + \frac{8}{2\pi}(1-\sin a) - 8\int_0^{\pi/2}\rho\cos k + 8\int_0^a\rho\cos k\\ &\approx +4a\rho_0(0) - \frac{4a}{\pi} - 8\int_0^{\pi/2}\delta\,\rho\cos k + \frac{4}{\pi} - 8\int_0^{\pi/2}\rho_0\cos k\,,\end{aligned} \tag{71}$$

where $\delta\rho \equiv \rho - \rho_0$. The last two terms in (71) are the energy of the half-filled band, $N = N_a$.

Our next task is to compute $\delta\rho$ to leading order in a. It is more convenient to deal with the function $\delta f \equiv f - f_0$ and to note (from (26)) that $\int_0^{\pi/2} \rho(k)\cos k\,\mathrm{d}k = \int_0^1 f(x)\sqrt{1-x^2}\,\mathrm{d}x$. We turn to (41) and find, to leading order, that

$$f \approx (1+2\hat{U})\,t - \hat{U}\hat{A}t + 2\hat{U}\hat{A}\hat{U}\,t = f_0 + \hat{U}\hat{A}f_0 - 2\hat{U}\hat{A}\,t \tag{72}$$

with $f_0 = (1+2\hat{U})\,t$. We note that $\hat{U} = \hat{K}^2(1+\hat{K}^2)^{-1}$ since $B = \infty$ (see (39)) and has a kernel which we will call $u(x-y)$. If g is continuous near 0 (in our case $g = f_0$ or $g = t$) then $(\hat{U}\hat{A}\,g)(x) = \int_{-a}^{a} u(x-y)g(y)\,\mathrm{d}y \approx 2au(x-0)g(0)$ to leading order in a. We also note from (26) that $f_0(0) = \rho_0(0)$. Therefore,

$$\int_0^{\pi/2} \delta\rho\cos k \approx a\left[\rho_0(0) - \frac{1}{\pi}\right]\int_{-1}^{1}\sqrt{1-x^2}\,u(x)\,\mathrm{d}x\,. \tag{73}$$

The integral in (73) is most easily evaluated using Fourier transforms and Plancherel's theorem,

$$\begin{aligned}\int_{-1}^{1}\sqrt{1-x^2}\,\mathrm{e}^{\mathrm{i}\omega x}\,\mathrm{d}x &= 2\int_0^1 \cos(\omega x)\sqrt{1-x^2}\,\mathrm{d}x \\ &= 2\int_0^{\pi/2}\cos(\omega\sin\theta)\cos^2\theta\,\mathrm{d}\theta = \frac{\pi}{\omega}J_1(\omega)\end{aligned} \tag{74}$$

and from (24)

$$\int_{-\infty}^{\infty} u(x)\mathrm{e}^{\mathrm{i}\omega x}\,\mathrm{d}x = [1+\mathrm{e}^{|\omega U/2|}]^{-1}\,. \tag{75}$$

By combining these transforms we can evaluate δE from (71).

$$\frac{\delta E}{N_a} = 2a\left[\rho_0(0) - \frac{1}{\pi}\right]\left[2 - 4\int_0^{\infty}\frac{J_1(\omega)}{\omega[1+\exp(\omega U/2)]}\,\mathrm{d}\omega\right]\,. \tag{76}$$

By dividing (76) by (70) we obtain (65).

7.2. *Chemical potential from the Bethe Ansatz*

The evaluation of the chemical potentials from the Bethe Ansatz is reminiscent of the calculation of the excitation spectrum of the 1D delta-function Bose gas solved by one of us [15]. We consider the case of a half-filled band. To use our results in the previous sections, which hold for M, M' odd, we calculate μ_- by removing 4 electrons, 2 with spin up and 2 with spin down, from a half-filled band. This induces the changes

$$N \to N-4 = N_a - 4, \quad M \to M - 2 = N_a/2 - 2\,. \tag{77}$$

Eqs. (15) and (16) determining the new k' and Λ' now read

$$N_a k'_j = 2\pi I'_j + \sum_{\beta=2}^{M-1} \theta(2\sin k'_j - 2\Lambda'_\beta), \quad j = 3,4,\ldots,N-2\,, \tag{78}$$

$$\sum_{j=3}^{N-2} \theta(2\sin k'_j - 2\Lambda'_\alpha) = 2\pi J'_\alpha - \sum_{\beta=2}^{M-1} \theta(\Lambda'_\alpha - \Lambda'_\beta), \quad \alpha = 2,\ldots,M-1\,. \tag{79}$$

Under the changes (77), the values of I' and J' are the same as those of I and J, namely,

$$I'_j = I_j, \quad j = 3,4,\ldots,N-2\,,$$

$$J'_\alpha = J_\alpha, \quad \alpha = 2,3,\ldots.M-1\,,$$

so they are centered around the origin with $k'_{total} = k_{total}$.

The removal of four electrons causes the values of k and Λ to shift by small amounts, and we write

$$k'_j = k_j + \frac{1}{N_a} w(k_j), \quad \Lambda'_\alpha = \Lambda_a + \frac{1}{N_a} u(\Lambda_a)\,.$$

By taking the differences of (78) and (15), and (79) and (16), and keeping the leading terms, one obtains

$$w(k_j) = \frac{1}{N_a} \sum_{\beta=2}^{M-1} \theta'(2\sin k_j - 2\Lambda_\beta)[2\cos k_j w(k_j) - 2u(\Lambda_\beta)]\,, \tag{80}$$

$$4\theta(2\Lambda_\alpha) - \frac{1}{N_a} \sum_{j=3}^{N-2} \theta'(2\Lambda_\alpha - 2\sin k_j)[2u(\Lambda_\alpha) - 2\cos k_j w(k_j)]$$

$$= -\frac{1}{N_a} \sum_{\beta=2}^{M-1} \theta'(\Lambda_\alpha - \Lambda_\beta)[u(\Lambda_\alpha) - u(\Lambda_\beta)]\,. \tag{81}$$

In deriving these equations we have used facts from our analysis of the integral equations, namely that when $M = M'$, $-\Lambda_1 = \Lambda_M \approx \infty$ (i.e., $=\infty$ in the limit $N_a \to \infty$) and that when $N = N_a$, $-k_1 = k_N \approx -k_2 = k_{N-1} \approx \pi$ as $N_a \to \infty$. Without using these facts there would be extra terms in (80) and (81), e.g., $\theta(2\sin k_j - 2\Lambda_1) + \theta(2\sin k_j - 2\Lambda_M)$, which is ≈ 0 because $-\Lambda_1 = \Lambda_M \approx \infty$.

By replacing the sums by integrals and making use of (20) and (21), we are led to the coupled integral equations

$$r(k) = \int_{-\infty}^{\infty} K(\sin k - \Lambda) s(\Lambda)\,\mathrm{d}\Lambda \tag{82}$$

$$4\theta(2\Lambda) + 2\pi s(\Lambda) - \int_{-\pi}^{\pi} K(\sin k - \Lambda) r(k) \cos k \, \mathrm{d}k$$

$$= -\int_{-\infty}^{\infty} K^2(\Lambda - \Lambda') s(\Lambda') \, \mathrm{d}\Lambda' \,, \tag{83}$$

where

$$r(k) = w(k)\rho_0(k), \quad s(\Lambda) = u(\Lambda)\sigma_0(\Lambda) \,. \tag{84}$$

Eqs. (82) and (83) can be solved as follows. Note that the third term on the left-hand side of (83) vanishes identically after substituting (82) for $r(k)$. Next introduce the Fourier transforms (24) and

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega\Lambda} \, \theta(2\Lambda) \, \mathrm{d}\Lambda = -\left(\frac{2\pi\mathrm{i}}{\omega}\right) \mathrm{e}^{-|\omega|U/4} \,, \tag{85}$$

and we obtain from (83)

$$s(\Lambda) = \frac{2}{\pi} \int_0^{\infty} \frac{\sin \omega\Lambda}{\omega \cosh(\omega U/4)} \, \mathrm{d}\omega \,. \tag{86}$$

Thus, from (82)

$$r(k) = \frac{4}{\pi} \int_0^{\infty} \frac{\sin(\omega \sin k) \, \mathrm{d}\omega}{\omega(1 + \mathrm{e}^{\omega U/2})} \,. \tag{87}$$

Note that we have $r(-k) = -r(k)$ and $s(-\Lambda) = -s(\Lambda)$.

The chemical potential μ_- for a half-filled band is now computed to be

$$\mu_-(U) = \frac{1}{4}\left[E\left(\frac{N_a}{2}, \frac{N_a}{2}\right) - E\left(\frac{N_a}{2} - 2, \frac{N_a}{2} - 2\right)\right]$$

$$= \frac{1}{4}\left[-2\sum_{j=1}^{N} \cos k_j + 2\sum_{j=3}^{N-2} \cos k'_j\right]$$

$$= \frac{1}{4}\left[-2(-1-1-1-1) + 2\sum_{j=3}^{N-2} (\cos k'_j - \cos k_j)\right]$$

$$= 2 - \frac{1}{2}\int_{-\pi}^{\pi} r(k) \sin k \, \mathrm{d}k$$

$$= 2 - 4\int_0^{\infty} \frac{J_1(\omega)}{\omega(1 + \mathrm{e}^{\omega U/2})} \, \mathrm{d}\omega \tag{88}$$

which agrees with (65).

8. Conclusion

We have presented details of the analysis of ground state properties of the 1D Hubbard model previously reported in Ref. [1]. Particularly, the analyses of the integral equations and of the absence of a Mott transition presented here have not heretofore appeared in print.

It is important to note that in order to establish that our solution is indeed the true ground state of the 1D Hubbard model, it is necessary to establish the existence of ordered real solutions to the Bethe Ansatz equations (14) and, assuming the solution exists, proofs of (a) and (b) as listed at the end of Section 3. The fulfillment of these steps remains as an open problem.

Acknowledgements

We are indebted to Daniel Mattis for encouraging us to investigate the jump in the chemical potential as an indicator of the insulator–conductor transition. We also thank Helen Au-Yang and Jacques Perk for helpful discussions. FYW would like to thank Dung-Hai Lee for the hospitality at the University of California at Berkeley and Ting-Kuo Lee for the hospitality at the National Center for Theoretical Sciences, Taiwan, where part of this work was carried out. Work has been supported in part by NSF grants PHY-0139984, DMR-9980440 and DMR-9971503.

References

[1] E.H. Lieb, F.Y. Wu, Phys. Rev. Lett. 20 (1968) 1445–1448, Erratum;
E.H. Lieb, F.Y. Wu, Phys. Rev. Lett. 21 (1968) 192.
[2] M. Gutzwiller, Phys. Rev. Lett. 10 (1963) 159–162.
[3] J. Hubbard, Proc. R. Soc. London A 276 (1963) 238–257;
J. Hubbard, Proc. R. Soc. London A 277 (1964) 237–259.
[4] O.J. Heilman, E.H. Lieb, Trans. N.Y. Acad. Sci. 33 (1970) 116–149;
O.J. Heilman, E.H. Lieb, Ann. N.Y. Acad. Sci. 172 (1971) 583–617.
[5] E.H. Lieb, B. Nachtergaele, Phys. Rev. B 51 (1995) 4777–4791.
[6] M. Gaudin, La Fonction d'onde de Bethe, Masson, Paris, 1983.
[7] Z.N.C. Ha, Quantum Many-Body Systems in One Dimension, World Scientific, Singapore, 1996.
[8] A. Montorsi, The Hubbard Model, World Scientific, Singapore, 1992.
[9] M. Takahashi, Thermodynamics of One-dimensional Solvable Models, Cambridge University Press, London, 1999.
[10] E.H. Lieb, in: D. Iagolnitzer (Ed.), Proceedings of the XIth International Congress of Mathematical Physics, Paris, 1994, International Press, 1995, pp. 392–412.
[11] H. Tasaki, J. Phys. Cond. Matt. 10 (1998) 4353–4378.
[12] E.H. Lieb, D.C. Mattis, Phys. Rev. 125 (1962) 164–172.
[13] E.H. Lieb, Phys. Rev. Lett. 62 (1989) 1201–1204, Errata;
E.H. Lieb, Phys. Rev. Lett. 62 (1989) 1927.
[14] H.A. Bethe, Zeits. f. Physik 71 (1931) 205–226 (Engl. trans. in D.C. Mattis, The Many-Body Problem, World Scientific, Singapore, 1993).
[15] E.H. Lieb, W. Liniger, Phys. Rev. 130 (1963) 1605–1616;
E.H. Lieb, Exact analysis of an interacting bose gas, II. The excitation spectrum, Phys. Rev. 130 (1963) 1616–1624.

[16] J.B. McGuire, J. Math. Phys. 6 (1965) 432;
J.B. McGuire, Attractive potential, J. Math. Phys. 7 (1966) 123.
[17] E.H. Lieb, M. Flicker, Phys. Rev. 161 (1967) 179–188.
[18] M. Gaudin, Phys. Lett. 24 A (1967) 55–56. See also Thesis, University of Paris, 1967, which is now in book form as Travaux de Michel Gaudin: Modèles exactement résolus, Les Editions de Physique, Paris, Cambridge, USA, 1995.
[19] C.N. Yang, Phys. Rev. Lett. 19 (1967) 1312–1314.
[20] F. Woynarovich, J. Phys. C 15 (1982) 85–96, See also;
F. Woynarovich, J. Phys. C 15 (1982) 97–109;
F. Woynarovich, J. Phys. C 16 (1983) 5293–5304;
F. Woynarovich, J. Phys. C 16 (1983) 6593–6604.
[21] I.S. Gradshteyn, I.M. Ryzhik, Tables of Integrals, Series and Products, Academic Press, San Diego, 2000.