Metal-insulator transition in the generalized Hubbard model

Metal–insulator transition in the generalized Hubbard model

Anatoly A Ovchinnikov

Institute for Nuclear Research of the Russian Academy of Sciences, Moscow 117312, Russia

Abstract. We present the exact ground-state wave function and energy of the generalized Hubbard model, subject to the condition that the number of doubly occupied sites is conserved, for a wide, physically relevant range of parameters. For one hole and one doubly occupied site the existence of the ferromagnetic ground state is proved which allows one to determine the critical value of the on-site repulsion corresponding to the point of metal–insulator transition. For the one-dimensional model the exact solution for special values of the parameters is obtained.

1. Introduction

The Hubbard model is the generic model to describe correlations in narrow-band systems [1]. The on-site repulsion is due to the matrix elements of the Coulomb interaction corresponding to the on-site Wannier states while the other matrix elements are neglected. However, the on-site repulsion can be sufficiently strong and the values of the matrix elements corresponding to the pair of nearest-neighbour sites can be comparable with the value of the simplest nearest-neighbour hopping amplitude. The Hamiltonian of the model which is often referred to as a generalized Hubbard model contains the interaction terms of the fourth order in the electron creation and annihilation operators corresponding to the nearest-neighbour sites. The Hamiltonian is

$$\hat{H} = -t \sum_{\langle ij \rangle \sigma} (c_{i\sigma}^+ c_{j\sigma} + c_{j\sigma}^+ c_{i\sigma}) + X \sum_{\langle ij \rangle \sigma} (c_{i\sigma}^+ c_{j\sigma} + c_{j\sigma}^+ c_{i\sigma})(n_{i-\sigma} + n_{j-\sigma})$$

$$+ V \sum_{\langle ij \rangle} n_i n_j + U \sum_i n_{1i} n_{2i} \tag{1}$$

where $\sigma = 1, 2$ is the projection of spin, $n_{i\sigma} = c_{i\sigma}^+ c_{i\sigma}$, $n_i = n_{1i} + n_{2i}$ and $\langle ij \rangle$ denotes the pair of nearest-neighbour sites. The generalized model (1) has been studied previously by several authors [2, 3, 4]. The other models with a similar kind of hopping term were considered in [5, 6]. Even for $X, V \ll U$ the presence of interactions which directly couple nearest-neighbour sites should lead to new effects. For example the correlated hopping terms are believed to play an essential role in the formation of high-T_c superconductivity [2]. In the present paper we consider the Hamiltonian (1) at $X = t$. At this particular value the model is simpler than the conventional Hubbard model. Note that the corresponding region in parameter space is quite realistic in view of the estimates of these parameters for different systems (for example, see [3, 4] and references therein). At half filling the ground state of the model at $X = t$ can be found exactly in any dimensions in a wide range of the parameters t, U, V. We also study the metal–insulator transition at the critical value U_c

which can be found exactly in our case in contrast to the Hubbard model where the Mott picture of the metal–insulator transition [7] is not directly applicable at least for a simple square or cubic lattice where presumably (in spite of the predictions based on the Gutzwiller approximation [8]) the system is the antiferromagnetic insulator at arbitrary U. Recently the same model was studied by Strack and Vollhardt [3] with the help of supersymmetric representation. We show that the wave functions proposed are in fact the ground states of the model in a range of parameters which is much wider than the region found in [3].

First, we show that at half filling the exact ground state of the model at $X = t$ and $U > z \max(2t, V)$ (z is the coordination number of the lattice) is a highly degenerate state without the doubly occupied sites and the system is a paramagnetic insulator. For the on-site repulsion $U < U_c$, where U_c is the critical value, which is not necessarily coincident with the obtained bound, the creation of holes and doubly occupied sites is energetically favourable. At this point the transition to the metallic state takes place. In complete analogy with the Nagaoka theorem for the infinite-U Hubbard model [9] we prove that in the sector of Hilbert space with one hole and one doubly occupied site the state with the lowest energy is ferromagnetic. We show that at $V < 2t$ the critical value is determined by the ground-state energy of the problem with one hole and one doubly occupied site and $U_c = 2zt$. At the present time the stability of the Nagaoka state at finite concentration of holes is not proved. However, it is supposed to be the correct ground state at sufficiently small concentration of holes [10]. Assuming the stability of the ferromagnetic state at finite concentration of holes and doubly occupied sites at low concentration of holes we find the density of holes below the point of the metal–insulator transition. For a bipartite lattice we find another region of the parameters where the determination of the ground state is possible. For a square or cubic lattice at $U < 2zV - z\max(2t, V)$ the ground state is given by the state with the electrons occupying only one of the sublattices. In particular, at $V > 2t$ the ground state is known exactly at arbitrary U. The transition between two different ground states take place at $U = zV$. Finally we consider the generalized Hubbard model at $X = t$ in one dimension. The dependence of the energy on the total spin for an arbitrary number of holes is studied. These results may be useful in the context of study of the stability of the ferromagnetic state for the infinite-U Hubbard model in higher dimensions. It is shown that the model is exactly solvable at $V = 0$. We also comment on the behaviour of different generalizations of the model (2) which include the antiferromagnetic coupling. A brief description of the results presented in section 2 was given in [11].

2. Generalized Hubbard model in arbitrary dimensions

Consider the model (1) at half filling ($\bar{n} = \bar{n}_1 + \bar{n}_2 = 1$) at $X = t$. The Hamiltonian

$$\hat{H} = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^+ c_{j\sigma} (1 - n_{i-\sigma} - n_{j-\sigma}) + \text{HC} + V \sum_{\langle ij \rangle} n_i n_j + U \sum_i n_{1i} n_{2i} \qquad (2)$$

where HC stands for Hermitian conjugate, conserves the number of doubly occupied sites $\hat{N} = \sum_i n_{1i} n_{2i}$: $[\hat{H}, \hat{N}] = 0$. The eigenstates of $\hat{H}$ correspond to a definite number of doubly occupied sites N which at the half filling coincides with the number of holes. First, let us prove that at $U > z\max(2t, V)$ the ground state corresponds to $N = 0$. It is convenient to express the Hamiltonian (2) in terms of the fermionic creation and annihilation operators of the holes (c_i^+, c_i) and the doubly occupied sites (d_i^+, d_i) defined starting from the ferromagnetic state $|F\rangle = \sum_i c_{2i}^+ |0\rangle$. The up-spin electrons are described by the Holstein–Primakoff hard-core Bose operators (b_i^+, b_i). To obtain the interaction term $\sim$V one can

make the following substitution:

$$n_{1i} = n_i + n_{di} \qquad n_{2i} = 1 - n_i - n_{ci}.$$

The Hamiltonian is

$$\hat{H} = -t \sum_{\langle ij \rangle} (c_i^+ c_j + d_i^+ d_j)(b_j^+ b_i + (1 - n_i)(1 - n_j)) + \text{HC}$$

$$+ V \sum_{\langle ij \rangle} (n_{ci} - n_{di})(n_{cj} - n_{dj}) + U\hat{N} + zVL/2 \tag{3}$$

where $n_{ci} = c_i^+ c_i$, $n_{di} = d_i^+ d_i$, $n_i = b_i^+ b_i$, and the constraint $n_{ci} + n_{di} + n_i \leqslant 1$ which is equivalent to the infinite on-site repulsion between the particles is implied. $\hat{N} = \sum_i n_{di}$ is the number of doubly occupied sites and the energy NU is due to the last term in equation (2).

2.1. Variational theorem

The upper bound for the ground-state energy $E_0 \leqslant zVL/2$ (L is the number of lattice sites) of the Hamiltonian (2) can be obtained using the variational wave function with singly occupied sites

$$|\phi\rangle = \prod_{i \in \mathcal{L}} c_{1i}^+ \prod_{j \in \mathcal{L}'} c_{2j}^+ |0\rangle \tag{4}$$

where $\mathcal{L}$ and $\mathcal{L}'$ are arbitrary disjoint sets of lattice sites which together build up the total lattice. In the representation (3) $|\phi\rangle$ is the state without the fermions. It is easy to derive the lower bound for the energy E_0. The Hamiltonian (3) can be considered as a matrix $H_{\alpha\beta}$ where the indices α, β enumerate the set of possible configurations of particles on the lattice $\alpha = (i_1 \dots i_N | j_1 \dots j_N | l_1 \dots l_M)$ where i, j, l are the coordinates of holes, doubly occupied sites and hard-core bosons respectively (see the appendix). One can see that due to the Fermi statistics of c and d particles the non-diagonal matrix elements of $H_{\alpha\beta}$ corresponding to the kinetic-energy term of the Hamiltonian (3) are equal to $\pm t$. Clearly, for the Bose statistics these matrix elements would be equal to $-t$. The diagonal matrix elements are determined by the second term of equation (3). The following theorem can be easily proved. For each of the eigenvalues E of any Hermitian matrix $H_{\alpha\beta}$ at least one of the inequalities

$$|E - H_{\alpha\alpha}| \leqslant \sum_{\beta \neq \alpha} |H_{\alpha\beta}| \tag{5}$$

is satisfied. In particular, for the restriction on E_0 one should take the minimal value of $H_{\alpha\alpha}$ $(-zVN)$. The right-hand side of this inequality is determined by the number of hopping processes allowed for a given state α. Since the total number of c and d particles is $2N$ the maximal value of the right-hand side of equation (5) is $2ztN$. In this way we find the following lower bound for the ground-state energy: $E_0(N) > (-2zt - zV + U)N + zVL/2$. One can further improve this estimate in the following way. For example, the hopping of a c particle to the nearest-neighbour site occupied by a d particle is not possible and one should not take into account these terms in the right-hand side of equation (5). However, in this case the contribution $-V$ to the diagonal matrix element $H_{\alpha\alpha}$ does exist. In the opposite case of isolated c and d particles the binding energy is absent while the hopping processes are possible. Thus we obtain the following lower bound for the energy:

$$E_0(N) \geqslant (-z \max(2t, V) + U)N + zVL/2. \tag{6}$$

Taking into account the upper bound $E_0 < zVL/2$ and equation (6) we see that at $U \geqslant z \max(2t, V)$ the ground-state wave function does not contain the holes and the doubly occupied sites, $N = 0$, and the energy is exactly equal to $E_0 = zVL/2$. The ground state is 2^L-fold degenerate and is given by equation (4). For comparison the bound for U found in [3] is $4zt + zV$. The authors use the following method to obtain the lower bound for the energy. With the help of the operators

$$P_{ij\sigma}^+ = c_{i\sigma}(1 - n_{i-\sigma}) + c_{j\sigma}(1 - n_{j-\sigma}) \qquad Q_{ij\sigma}^+ = c_{i\sigma}^+ n_{i-\sigma} + c_{j\sigma}^+ n_{j-\sigma}$$

the Hamiltonian can be represented in the form

$$\hat{H} = t \sum_{\langle ij \rangle \sigma} (P_{ij\sigma}^+ P_{ij\sigma} + Q_{ij\sigma}^+ Q_{ij\sigma}) - 2zt \sum_i (1 - n_i)$$
$$+ (U - 4zt) \sum_i n_{1i} n_{2i} + V \sum_{\langle ij \rangle} n_i n_j. \tag{7}$$

Since the average of the first term in equation (7) over the ground state is positive definite the bound $U > 4zt + zV$ is obtained. Clearly for an arbitrary lattice, at $t, V \ll U$ and a small deviation $t - X \neq 0$ the degeneracy is absent and the model (1) reduces to an effective Heisenberg model with antiferromagnetic coupling constant $4(t - X)^2/U$. For a bipartite lattice one has the antiferromagnetically ordered ground state.

For $U < U_c$ the creation of holes and doubly occupied sites take place. We shall see that at $V < 2t$ the critical value U_c is determined by the value of the lowest energy for the state with one empty and one doubly occupied site ($N = 1$). This problem can be solved exactly in two and three dimensions. The Hamiltonian (3) has the same form as an analogous Hamiltonian for the infinite-U Hubbard model away from half filling. The only difference is the absence of d fermions in the latter case. The hopping term for the bosons is absent as well and their interaction with the fermions has the same form. According to the Nagaoka theorem [9], for one hole the ground state of the infinite-U Hubbard model is ferromagnetic (the total spin is maximal). In our case it is possible to prove that at $N = 1$ i.e. for one hole and one doubly occupied site the ground state is ferromagnetic. Following the original Nagaoka proof consider the hopping process in which the hole and the doubly occupied site start from given positions and come back to the same positions after a number of steps. The configuration of spins (the bosons) can be different in the initial and the final states. For one hole the corresponding energy-dependent self-energy part introduced in [9] is positive. The same would be true for an arbitrary number of holes if the holes obeyed the Bose statistics in the sense of the Hamiltonian (3). Clearly in our case the positivity condition is satisfied since for $N = 1$ the statistics of c and d particles which represent the two different species of fermions is not important. The other steps of the Nagaoka proof can be applied without modification (the presence of the attraction $\sim V$ in equation (3) is not important either). Thus at $N = 1$ the state with $S = S_{\max} - 1$ which is the maximal total spin for a state with one hole is the ground state of the system. In the representation (3) that means that the number of bosons is equal to zero for the projection of spin $S^z = -(S_{\max} - 1)$. The ground-state energy of two particles has the form $E_0(N = 1) = -2zt + U + a/L$, where the last term is due to the interaction of c and d particles. Let us suppose that the interaction part of the energy is positive, $a > 0$ (the corresponding condition for the parameter V will be found later). Then the critical value of U is

$$U_c = 2zt. \tag{8}$$

At $U > U_c$ the ground state is given by equation (4) and at $U < U_c$ the creation of holes is energetically favourable. The critical value (8) is the point of the metal–insulator transition.

Before solving the two-particle problem let us comment on the behaviour of the system at finite density of c and d fermions. At the present time the stability of the Nagaoka state for the infinite-repulsion Hubbard model at finite concentration of holes is not proved. It is supposed that the ground state is ferromagnetic at sufficiently small concentration of holes (for discussion see [10]). In our case we will also assume the stability of the ferromagnetic state $(S = S_{max} - N)$ at sufficiently small density of holes and doubly occupied sites $\rho = N/L$. Then at $U < 2zt$ and $|2zt - U| \to 0$ the density $\rho \to 0$ and the analogue of the Nagaoka state is realized. The system of two species of interacting fermions at equal density ρ should be considered. At low density the ground-state energy can be evaluated as a series in the small parameter ($|\ln \rho|^{-1}$ and $\rho^{1/3}$ respectively in two and three dimensions). At low density the dependence of the energy on ρ has the form

$$E/L = -(2zt - U)\rho + \mathcal{E}_0(\rho) + \tfrac{1}{2}a(\rho)\rho^2 \qquad (9)$$

where the second term is due to the Fermi statistics. In the lowest order $\mathcal{E}_0(\rho) = 4\pi t\rho^2$ in two dimensions and $\mathcal{E}_0(\rho) = (\tfrac{1}{5}6^{5/3}\pi^{4/3})t\rho^{5/3}$ in three dimensions. In the lowest order in ρ the function $a(\rho)$ is determined by the two-particle scattering amplitude at low energy. The scattering amplitude can be expressed through the interaction energy of two particles of different species a/L in the finite volume L. In fact the low-density limit of $a(\rho)$ coincides with the value of the parameter a. The behaviour of the function $a(\rho)$ is different in two and three dimensions. In three dimensions it is constant in the lowest order in ρ while in two dimensions $a(\rho) = 8\pi t/|\ln \rho| + O(t/(\ln \rho)^2)$ [12]. The ground-state density of holes is determined from the condition of the minimum of the expression (9). In the limit $U \to 2zt$ the minimum is determined by the first two terms in equation (9). For example, in two dimensions

$$\rho_0 = \frac{2zt - U}{4\pi} \qquad (10)$$

with accuracy up to the terms of order $\sim 1/\ln \rho_0$. One can also take into account the leading corrections in ρ_0 in this formula. ρ_0 is small at small deviation of U from the critical value (8) in agreement with our assumption about the stability of the ferromagnetic state. Although the description of the system (3) at arbitrary number of bosons is not possible it will be shown that at $2zt - U \to 0$ the minimum of the energy as a function of ρ found from equation (9) is the correct ground-state energy of the system in agreement with equation (8).

2.2. Solution of the two-particle problem.

Let us proceed with the solution of the two-particle problem. The interaction potential (3) contains the infinite on-site repulsion $\tilde{U} \to \infty$ and the attraction of strength V at nearest-neighbour sites. Let us consider the case of the three-dimensional cubic lattice. The ground state corresponds to the total momentum of two particles equal to zero. One can seek the wave function in the form

$$|\psi\rangle = \sum_k F(k)c_k^+ d_{-k}^+|0\rangle \qquad (11)$$

where $c_k^+ = L^{-1/2}\sum_i \epsilon^{iki}c_i^+$. The function (11) is the eigenvector of eigenvalue E if the function $F(k)$ satisfies the Schrödinger equation

$$(E - 2t\epsilon_p)F(p) = \frac{1}{L}\sum_k \left(\tilde{U} - V\epsilon_{k-p}\right)F(k) \qquad (12)$$

where $\epsilon_k = -2\sum_{\alpha=1}^{3}\cos k_\alpha$. Let us define the function $J(k) = (E - 2t\epsilon_k)F(k)$. Equation (12) takes the form

$$J(p) = \frac{1}{L}\sum_k \frac{\tilde{U} - V\epsilon_{k-p}}{E - 2t\epsilon_k}J(k). \tag{13}$$

In order to solve the equation (13) for the ground-state energy $E_0 = -2zt + a/L$ let us extract from the sum the term with $k = 0$, which is most divergent for $L \to \infty$. We can also substitute the value $2t\epsilon_0$ ($\epsilon_0 = -z$) for E in the sum over $k \neq 0$ in equation (13) since in three dimensions the energy difference $\epsilon_k - \epsilon_0$ is at least of order $\sim L^{-2/3}$ and the interaction correction is of order $\sim 1/L$. We get

$$J(p) = \frac{(\tilde{U} - V\epsilon_p)J(0)}{2ta} + \frac{1}{L}\sum_{k\neq 0}\frac{\tilde{U} - V\epsilon_{k-p}}{2t(\epsilon_0 - \epsilon_k)}J(k). \tag{14}$$

Defining the new function $\Gamma(k) = aJ(k)/J(0)$ we obtain from (14) the equation

$$2t\Gamma(p) = \tilde{U} - V\epsilon_p + \frac{1}{L}\sum_{k\neq 0}\frac{\tilde{U} - V\epsilon_{k-p}}{\epsilon_0 - \epsilon_k}\Gamma(k) \tag{15}$$

which is nothing but the equation for the scattering amplitude at zero energy $\Gamma(0) = a$. The sum in equation (15) can be replaced by an integral. The solution of the equation (15) can be represented in the form $\Gamma(k) = \Gamma_0 + \epsilon_k\Gamma_1$. Substituting this function into equation (15) we get two equations for the unknown constants Γ_0, Γ_1. In the limit $\tilde{U} \to \infty$ the result of the calculations for the simple cubic lattice ($z = 6$) has the form

$$a = \frac{z(2t - V)}{Wz - V(Wz - 1)/2t} \qquad {}^{\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots} \tag{16}$$

where $W = 0.2527$ stands for the Watson integral

$$W = \frac{1}{2(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} dk_x\, dk_y\, dk_z \frac{1}{3 - \cos k_x - \cos k_y - \cos k_z}.$$

The expression (16) is valid at $V < 2t$. At $V = 2t$ the amplitude a vanishes which indicates the existence of the two-particle bound state at $V > 2t$. Vanishing of the interaction correction to the energy at $V = 2t$ can be seen from analogy with the two-magnon problem in the ferromagnet where in the ground state the total spin should be maximal. The same conclusion can be made for two dimensions. To calculate the parameter a one should substitute the sum

$$\frac{1}{L}\sum_{k\neq 0}\frac{1}{\epsilon_k - \epsilon_0} \sim \frac{1}{4\pi}\ln L$$

for W in the formula (16). The corrections to the equation (15) are of order $(1/\ln L)^2$ and the terms of that order in equation (16) cannot be fixed. However, the expansion in $1/\ln L$ breaks down only at $2t - V \sim 1/\ln L$ and at these values of V the perturbation theory in $2t - V$ can be used since at $V = 2t$ the ground-state wave function is known: $\hat{P}_G c_0^+ d_0^+|0\rangle$ (the analogue of the $S = S_{max}$ state in the ferromagnet; $\hat{P}_G$ is the Gutzwiller projector). As in 3D the interaction correction changes sign at $V = 2t$ and the bound-state solution appears at $V > 2t$.

2.3. Formal proof of the equation (8)

Finally it is necessary to show that, whether or not the Nagaoka state is realized at a given density ρ, the correction to the energy,

$$E/L = -(2zt - U)\rho + \mathcal{E}(\rho) \tag{17}$$

is strictly positive $\mathcal{E}(\rho) > 0$ and does not vanish in the thermodynamic limit at $\rho \neq 0$. That means that the density is really small in the neighbourhood of the point of metal–insulator transition. We have to obtain the lower bound for $\mathcal{E}(\rho)$ (17). Let us modify the Hamiltonian (3) in such a way as to decrease the corresponding ground-state energy. First, let us replace in equation (3) the repulsion $V \sum (n_{ci}n_{cj} + n_{di}n_{dj})$ at nearest-neighbour sites by the attraction of the same form $(V \rightarrow -V)$ and then make the substitution $V \rightarrow 2t$ so that the resulting interaction takes the form

$$-2t \sum_{\langle ij \rangle}(n_{ci} + n_{di})(n_{cj} + n_{dj}) + U\hat{N}.$$

Second, instead of fermions, consider the particles c, d, obeying the Bose statistics. Since for bosons the Nagaoka state is the ground state at arbitrary density and the Hamiltonian is symmetric with respect to the replacement $c \leftrightarrow d$, the ground-state wave function $\phi(i_1 \ldots i_N | j_1 \ldots j_N)$ which is the totally symmetric function of its arguments coincides with the ground-state wave function of the Heisenberg ferromagnet in the representation of the Holstein–Primakoff bosons with the projection of spin $S^z = S_{\max} - N$ and $S = S_{\max}$. In fact, the corresponding wave function, which is the positive definite and totally symmetric function, is the eigenstate of the modified Hamiltonian. Actually it is the ground state of the modified Hamiltonian since for the wave function which changes sign the substitution

$$\phi(i_1 \ldots i_N | j_1 \ldots j_N) \rightarrow |\phi(i_1 \ldots i_N | j_1 \ldots j_N)|$$

would lower the energy and the positive-definite eigenstate is unique because of the orthogonality condition. Note that these considerations can be a basis of a simple proof of the Nagaoka theorem both in our case and in the case of the infinite-U Hubbard model [9]. Therefore the lower bound for the energy is $(-2zt + U)N$ and we find that $\mathcal{E}(\rho) > 0$. Clearly at finite density $\mathcal{E}(\rho)$ does not vanish in the thermodynamic limit. Thus it is proved that at $2zt - U \rightarrow 0$ the value of ρ minimizing the energy $\rho_0 \rightarrow 0$. Consequently, at $V < 2t$ the point of metal–insulator transition is indeed given by equation (8).

According to the Mott picture for large coupling U the density of states exhibits two bands with the centres separated by U. In the absence of electron correlations the width of each Hubbard band is zt, and the gap between the bands is expected to vanish at $U/2zt = 1$. The Hubbard bands are usually obtained in the framework of a special single-particle Green function decoupling approximation scheme proposed by Hubbard [1], the Hartree–Fock-type approximation which was exact for zero interaction energy or zero bandwidth (which amounts to a specific decomposition of a certain Green function and has no justification; for example see [13] for a critical discussion). However, strictly speaking, in the conventional Hubbard model the very notion of the Hubbard bands is justified only in the limit $U \rightarrow \infty$, since the number of doubly occupied sites is conserved only in this limit. In our model the number of doubly occupied sites is conserved and the notion of the Hubbard bands has a precise meaning beyond the framework of any approximation. Note also that our results can be used to explain the results of the numerical calculations for a small lattice [14].

2.4. *Ground state for a bipartite lattice*

For an arbitrary *bipartite* lattice one can find another region of the parameters where the determination of exact ground state is possible. For example consider the simple square or cubic lattice. The wave function which corresponds to a charge-density wave with maximal order parameter,

$$|\chi\rangle = \prod_{i \in \mathcal{A}} c_{1i}^+ c_{2i}^+ |0\rangle \qquad (18)$$

where $\mathcal{A}$ is one of the sublattices, is an eigenfunction of $\hat{H}$. The wave function (18) can be used to obtain an upper bound for the energy: $E_0 \leqslant UL/2$. To obtain the lower bound it is convenient to define the operators $\tilde{c}_{1i} = (-1)^i c_{1i}$, $\tilde{c}_{2i} = (-1)^i c_{2i}^+$. At half filling the particle number corresponding to the new operators is the same. In terms of the operators $\tilde{c}_{i\sigma}, \tilde{c}_{i\sigma}^+$ the state $|\chi\rangle$ is an antiferromagnetically ordered state with singly occupied sites. Since in terms of $\tilde{c}_{i\sigma}, \tilde{c}_{i\sigma}^+$ the kinetic-energy term has the same form as in equation (2), in this representation the number of doubly occupied sites is conserved and the lower bound for the energy of the state with N holes can be obtained using the representation (3) and the theorem (5). The energy of isolated c and d particles is zV and their interaction at nearest-neighbour sites is $-V$. Thus the lower bound for the energy as a function of N is

$$E_0(N) \geqslant (2zV - z\max(2t, V) - U)N + UL/2. \qquad (19)$$

One can see from equation (19) that the wave function (18) is the ground state at $U < 2zV - z\max(2t, V)$. This ground state is unique apart from a twofold degeneracy due to the two sublattices and describes the non-magnetic insulator. At $V > 2t$ we obtain the condition $U < zV$. Since it was shown that at $U > zV$ the ground state is given by equation (4), for a bipartite lattice at $V > 2t$ the ground state is found for arbitrary U. In the Hubbard model ($X = 0$, $V = 0$) the function (4) is the ground-state wave function only for $U = \infty$. In the generalized Hubbard model with $X = t$ this ground state is already realized at *finite* U with $U > z\max(2t, V)$. Similarly, while in the extended Hubbard model ($X = 0$, $V \neq 0$) the function (18) is the ground-state wave function only for $V = \infty$, in the generalized Hubbard model with $X = t$ this ground state is already realized at *finite* V.

3. Generalized Hubbard model in one dimension

Let us consider the generalized Hubbard model (2) in one dimension. Before considering the model (2) let us study the ground-state multiplicity (the value of the total spin) in the cases when the model is exactly solvable in one dimension. Namely we consider the model (2) at $V = 0$ and at $V \neq 0$ in the sector with no doubly occupied sites.

For the one-dimensional system two different cases can be considered: (1) the case of an open chain; (2) the case of a closed chain of finite length. Both cases are equivalent in the thermodynamic limit ($L \to \infty$). First, consider an open chain or equivalently a chain of infinite length. In this case the Nagaoka theorem is valid at an arbitrary number of holes and doubly occupied sites since the Hamiltonian (2) is invariant under the transformation which changes the statistics of holes. In other words for an open chain the holes can be considered as bosons. That does not mean that the eigenstates with $S < S_{\text{max}}$ cannot be degenerate in energy with the state $S = S_{\text{max}}$. In fact all the eigenstates including the ground state are degenerate in the total spin S. For example let us consider the eigenstates of the Hamiltonian (2) without the doubly occupied sites. At $V = 0$ in the case when the numbers of both c and d particles are not equal to zero we get the additional degeneracy of the

eigenstates due to the two species of particle. Let us seek the ground-state wave function in the following form:

$$\psi(i_1, \ldots, i_N | l_1, \ldots, l_M) = \psi_0(i_1, \ldots, i_N)\phi(\lambda_1, \ldots, \lambda_M) \qquad (20)$$

where i_α are the coordinates of c particles ($N = N_c$) and λ_β are the coordinates of the spin bosons on a 'supperlattice' which consists of $L_1 = L - N$ lattice sites which are not occupied by the holes ($L_1 = N_e$ is the number of electrons)

$$\lambda_\alpha = l_\alpha - \sum_{\beta=1}^{N} \theta(l_\alpha - i_\beta) \qquad \alpha = 1, \ldots, M.$$

If ψ_0 is the eigenstate of the Hamiltonian (2) in the sector $S = S_{\max}$ then the wave function (20) is the eigenfunction of the Hamiltonian for an arbitrary function $\phi(\lambda_1, \ldots, \lambda_M)$. Thus the ground state is degenerate in the total spin S. Let us turn to the case (2) and see how this degeneracy is resolved at finite L. Note that unlike the open chain, for the closed chain the spectrum depends on the statistics of particles. The coordinates on a superlattice can be defined in the same way by fixing the initial and the final sites of the chain ($l_\alpha = 1, \ldots, L$ and $\lambda_\alpha = 1, \ldots, L_1$). After the substitution (20) we consider the functions ψ_0 and ϕ_0 extended to the infinite chain and subjected to the appropriate boundary conditions. In order to satisfy the periodic boundary conditions for the function ψ (20) the functions $\psi_0(i_1, \ldots, i_N)$ and $\phi(\lambda_1, \ldots, \lambda_M)$ should satisfy the following boundary conditions:

$$\phi(\lambda_1, \ldots, \lambda_\alpha + L_1, \ldots, \lambda_M) = \phi(\lambda_1, \ldots, \lambda_\alpha, \ldots, \lambda_M) \cdot \qquad \alpha = 1, \ldots, M \qquad (21)$$

and

$$\psi_0(i_1, \ldots, i_\alpha + L, \ldots, i_N) = \exp(iq)\psi_0(i_1, \ldots, i_\alpha, \ldots, i_N) \qquad (22)$$

where the boundary conditions for the function ψ_0 are determined by the total momentum q corresponding to the function ϕ:

$$\phi(\lambda_1 + 1, \ldots, \lambda_M + 1) = \epsilon^{iq}\phi(\lambda_1, \ldots, \lambda_M). \qquad (23)$$

The periodic function ϕ is the symmetric function of its arguments which vanishes at $\lambda_\alpha = \lambda_\beta$. At arbitrary V the function ψ_0 is determined by the set of the momenta k_α, $\alpha = 1, \ldots, N$. The energy is

$$E = -2t \sum_{\alpha=1}^{N} \cos k_\alpha.$$

For instance at $V = 0$ we have the free fermion determinant

$$\psi_0 = \det_{\alpha\beta} \left[\exp(ik_\alpha i_\beta)\right] \qquad k_\alpha = 2\pi n_\alpha/L$$

where n_α are integers. The periodic function ϕ can be characterized by the set of momenta $q_\alpha = 2\pi m_\alpha/L_1$ where m_α are integers or half integers (see below). An arbitrary number of zeros $q_\alpha = 0$ is possible which corresponds to the value of the difference $S - S^z$. The total momentum $q = \sum_\alpha 2\pi m_\alpha/L_1$. Thus we obtain the equation for the momenta k_α:

$$k_\alpha = \frac{2\pi}{L} \left(n_\alpha + \sum_{\beta=1}^{M} \frac{m_\beta}{L_1}\right). \qquad (24)$$

As a basis in the space of the symmetric functions (21) one can choose the eigenstates of the Hamiltonian of the free hard-core bosons (XX model) on a chain of length L_1. In this case m_α are integers (half integers) for M odd (even), and $m_\alpha \neq m_\beta$ for $\alpha \neq \beta$. This automatically gives the eigenfunctions of the Hamiltonian (2). However, in general these

functions are not the eigenfunctions of the operator of the total spin S. In order to classify
the eigenstates according to their spin one can choose the basis given by the eigenstates of
the Heisenberg ferromagnet. In this case an arbitrary number of the momenta q_α can be
equal to zero and the non-zero momenta are determined by the system of equations

$$\epsilon^{iq_\alpha L_1} = (-1)^{M-1} \exp\left(2i\sum_{\beta=1}^{M} \tan^{-1}(u_\alpha - u_\beta)\right) \qquad u_\alpha = \tfrac{1}{2}\cot(q_\alpha/2). \quad (25)$$

The total spin $S = |L_1/2 - M|$, where M is the number of the non-zero momenta. The total
momentum is $q = \sum_\alpha 2\pi m_\alpha/L_1$ and we obtain the equation (24) where n_α are integers
and m_α are integers (half integers) for M odd (even) (we assume L to be even). The
same formulas could be obtained starting from the problem of N_e spinless fermions and the
hard-core bosons (upturned spins) on a 'superlattice' consisting of the lattice sites occupied
by the electrons. The only difference is the number of the momenta k_α: $\alpha = 1, \ldots, N_e$.
In this representation the same results could be obtained by taking the limit $U \to \infty$ in the
exact Bethe *anzatz* solution of the 1D Hubbard model [15] (for example see [16]). In fact
one can redefine the quantum numbers according to $n_\alpha \to n_\alpha - M/2$, $m_\alpha \to m_\alpha + L_1/2$ to
obtain the equation (24) with n_α integers (half integers) for M even (odd) and m_α integers
(half integers) for $(L_1 - M)$ odd (even) in agreement with the results of [16].

From equation (24) the ground-state energy as a function of the total spin can be found.
As an example consider the splitting of the energy levels with $S = S_{max}$ and $S = S_{max} - 1$ for
an even number of holes (L is assumed to be even). Clearly the ground state corresponds
to the values $q = 0$ and $q = \pi$ respectively and the value $E_0 = E_0(S_{max} - 1)$ is an
absolute ground-state energy. Thus we obtain the energy $E_0(S_{max}) = E_0 \cos(\pi/L)$, where
$E_0 = -2t (\sin(\pi/L))^{-1} \sin(\pi N/L)$. For all $S < S_{max} - 1$ the minimal energy levels are
nearly degenerate (i.e. the energy splitting is of higher order in $1/L$ at large L) with the
ground-state energy E_0. A similar picture can be obtained for N odd. In this case we find
$E(S_{max}) = E_0$ ($M = 0$, $q = 0$) in agreement with the Nagaoka theorem ($N = 1$). Clearly
the same procedure (20)–(25) could be performed at arbitrary V if the doubly occupied
sites are absent, or at $V = 0$ and an arbitrary number of the doubly occupied sites. For
instance in the first case the Bethe *anzatz* equations for the wave function ψ_0 (20) with
twisted boundary conditions (22) should be used.

Let us consider the model (2) at arbitrary U and arbitrary filling fraction $\bar{n}$. It was
proved that in the thermodynamic limit the eigenstates are degenerate. Thus it is sufficient
to consider the eigenstates with $S = S^z = S_{max}$, which corresponds to the absence of the
spin bosons. First, consider the model at $V = 0$. At $V = 0$ the system is equivalent to
the system of free fermions with an extra degeneracy due to the two different species of
particles, (c, d) with an infinite on-site repulsion. The density of doubly occupied sites
$\rho = N/L$ is determined by the minimum of the energy of free fermions with the total
density $n_0 + 2\rho$,

$$E_0(\rho)/L = -\frac{2t}{\pi} \sin[\pi(n_0 + 2\rho)] + U\rho \qquad (26)$$

where $n_0 = 1 - \bar{n}$ is the concentration of holes in the limit of large U. Thus at $U > U_c$
where

$$U_c = 4t \cos(\pi n_0)$$

the transition to the state with no doubly occupied sites, $\rho = 0$, takes place. Of course
at $n_0 \neq 0$ the point $U = U_c$ is not related to the metal–insulator transition. Away from
half filling the analogue of the ground state (18) i.e. the ground state without the singly

occupied sites is realized at $U < -4t$. At $V \neq 0$ the model is not exactly solvable in the sector of Hilbert space with the doubly occupied sites. In the previous section it was shown that at arbitrary $V < 2t$ and $U > 4t$ we have the number of doubly occupied sites $N = 0$ (really $N = 0$ at $U > U_c$ where the critical value $U_c < 4t$). At these values the ground state is equivalent to the ground state of the Heisenberg chain with the anisotropy parameter $V/2t$ in the sector with the projection of the total spin related to the number of holes. The spectrum of the anisotropic Heisenberg chain can be determined exactly with the help of the Bethe *anzatz*. Thus in 1D the ground state and the low-energy excitations of the model (2) can be found exactly at $U > U_c$ and an arbitrary $\bar{n}$. At $V/2t < 1$ and $V = 2t$ the spectrum of charge excitation is gapless and the system is a metal (if the number of holes $n_0 > 0$). The same is true for $V/2t > 1$ and $n_0 < 1/2$. Note that away from half filling the transition between the state with no doubly occupied sites and the state with no singly occupied sites analogous to the states (4) and (18) does not take place at $U = 2V$ under the condition $V/2t > 1$. If the concentration of holes is exactly $n_0 = \frac{1}{2}$ and $V > 2t$ there is a gap in the spectrum of charge excitations [17]. Thus at the filling fraction $\bar{n}_1 = \bar{n}_2 = \frac{1}{4}$ the system undergoes another metal–insulator transition at the point $V = 2t$. In general the model (2) is not exactly solvable in the sector with a non-zero number of doubly occupied sites. .

Although at the point of metal–insulator transition the ground state is ferromagnetic for two- and three-dimensional systems and degenerate in the total spin in one dimension the existence of the transition is not connected with the ferromagnetic order. In fact the metal–insulator transition is a general phenomenon in models with the kinetic-energy term conserving the number of doubly occupied sites (2). For example one can study the metal–insulator transition in the one-dimensional models with an antiferromagnetic coupling which are exactly solvable in the absence of doubly occupied sites [18, 19, 20]. Apart from the term $\sim X$ ($X = t$) and the on-site repulsion $\sim U$ these models include the interaction of the form $J \sum (S_i S_j - \frac{1}{4} n_i n_j)$ or $J \sum (S_i S_j + \frac{3}{4} n_i n_j)$ at $J = 2t$. Although these models are not exactly solvable at $N \neq 0$ at the half filling the existence of a metal–insulator transition can be shown and the critical value of U can be found exactly $U_c = 2zt \ln 2$. The model which is exactly solvable at arbitrary N was proposed in [5]. The Hamiltonian is the sum of the Hamiltonian of the t–J model (modified to include the doubly occupied sites) and the permutation term of the form

$$2t \sum_{\langle ij \rangle} (c_i^+ c_j d_j^+ d_i + c_j^+ c_i d_i^+ d_j).$$

One can also change the sign $J \rightarrow -J$ to obtain the integrable model with the ferromagnetic ground state. The ground state of this model is equivalent to the ground state of the t–J model [18] with the up- and down-spin electrons replaced by c and d particles (which corresponds to the SU(2) 'η-spin' symmetry of [21]). The concentration of d particles should be found from the condition of minimum of the energy.

4. Conclusion

In conclusion, for the model (2) it was shown that at half filling and $U > z \max(2t, V)$ the ground state is given by equation (4). The problem with one hole and one doubly occupied site in the ferromagnetic background was solved. We proved that it is the lowest-energy state in the sector of Hilbert space with one empty and one doubly occupied site. We established that at $V < 2t$ the critical value of U corresponding to the point of metal–insulator transition $U_c = 2zt$. Under the assumption of the stability of the ferromagnetic state at finite

concentration of holes the density of holes was found at $V < 2t$ and $U \to U_c$. Finally, for a *bipartite* lattice with constant number of nearest neighbours at $U < 2zV - z \max(2t, V)$ the exact ground-state wave function is given by equation (18). At $V > 2t$ the ground state is found exactly at arbitrary value of the parameter U. The transition between the states (4) and (18) occurs at $U = zV$. At $V < 2t$ the exact ground state is found at $U > 2zt$ (4) and $U < 2zV - 2zt$ (18). For the one-dimensional model at $V = 0$ the exact solution was presented. We have also studied the dependence of the energy on the total spin. Recently a part of the results obtained in section 3 was independently obtained in [22].

Acknowledgment

Appendix

Following Nagaoka [9], we introduce a set of orthogonal and normalized many-body wave functions which completely span the Hilbert space. We use the representation (3) in terms of the operators of holes $c_i^+ = c_{2i}$, doubly occupied sites $d_i^+ = c_{1i}^+$ and the overturned spins (hard-core bosons) $b_i^+ = c_{1i}^+ c_{2i}$ starting from the ferromagnetic state $\sum_i c_{2i}^+ |0\rangle$. Because of the Fermi statistics we have to be cautious about the order of c and d operators in the definition of the states α, β. Let, for each site i of the lattice, R_i be an integer number $R_i = 1, \ldots, L$. We can define the following order among the lattice sites. Setting up the coordinate system we assign a pair of integer coordinates (i_x, i_y) to each site of the lattice. If for the two sites i, j $i_x < j_x$ then we define $R_i < R_j$. When $i_x = j_x$ the order is determined by their y coordinates: $R_i < R_j$ for $i_y < j_y$. With this order we now introduce the states

$$|\alpha\rangle = c_{i_1}^+ \ldots c_{i_N}^+ d_{j_1}^+ \ldots d_{j_N}^+ b_{l_1}^+ \ldots b_{l_M}^+ |0\rangle$$

where the order of the operators for a given set of the lattice sites $\alpha = (i_1 \ldots i_N | j_1 \ldots j_N | l_1 \ldots l_M)$ is given by the condition

$$R_{i_1} < \ldots < R_{i_N} \qquad R_{j_1} < \ldots < R_{j_N} \qquad R_{l_1} < \ldots < R_{l_M}.$$

With this definition the non-diagonal matrix elements $H_{\alpha\beta} = \langle \alpha | \hat{H} | \beta \rangle$ of the Hamiltonian (3) are equal to $-t$ when one hole (doubly occupied site) changes order with an even number of holes (doubly occupied sites). In the opposite case the matrix elements $H_{\alpha\beta}$ are equal to $+t$. Clearly for Bose statistics the matrix elements would be equal to $-t$. Since c and d particles are two distinct species of fermion one can use the prescription given by the equation for $|\alpha\rangle$. For $N = 1$ the non-diagonal matrix elements are also equal to $-t$ which justifies our statement that the Fermi statistics of c and d particles is not important for the case of one hole and one doubly occupied site. Thus for the model (2) the Nagaoka proof [9] can be used without modifications for one hole and one doubly occupied site.

References

[1] Hubbard J 1963 *Proc. R. Soc.* A **276** 238
     Gutzwiller M C 1963 *Phys. Rev. Lett.* **10** 159
[2] Marsiglio F and Hirch J E 1990 *Phys. Rev.* B **41** 6435
     Hirch J E and Marsiglio F 1990 *Phys. Rev.* B **41** 2049; 1989 *Physica* C **162–164** 591

[3] Strack R and Vollhardt D 1993 *Phys. Rev. Lett.* **70** 2637
[4] Baeriswyl D, Horsch P and Maki K 1988 *Phys. Rev. Lett.* **60** 70
Campbell D K, Gammel J T and Loh E Y 1990 *Phys. Rev.* B **42** 475
Vries P, Michielsen K and Raedt H 1993 *Phys. Rev. Lett.* **70** 2463
Appel J, Grodzicki M and Paulsen F 1993 *Phys. Rev.* B **47** 2812
[5] Essler F H L, Korepin V E and Schoutens K 1992 *Phys. Rev. Lett.* **68** 2960; 1993 *Phys. Rev. Lett.* **70** 73
[6] Montorsi A and Rasetti M 1991 *Phys. Rev. Lett.* **66** 1383
[7] Mott N F 1974 *Metal–Insulator Transitions* (London: Taylor & Francis)
[8] Brinkman W F and Rice T M 1970 *Phys. Rev.* B **2** 4302
Kotliar G and Ruckenstein A E 1986 *Phys. Rev. Lett.* **57** 1362
[9] Nagaoka Y 1966 *Phys. Rev.* **147** 127
[10] Shastry B S, Krishnamurthy H R and Anderson P W 1990 *Phys. Rev.* B **41** 2375
Barbieri A, Riera J A and Young A P 1990 *Phys. Rev.* B **41** 11 697
Kopp T, Ruckenstein A E and Schmitt-Rink S 1990 *Phys. Rev.* B **42** 6850
Richmond R and Rickayzen G 1969 *J. Phys. C: Solid State Phys.* **2** 528
Basile A G and Elser V 1990 *Phys. Rev.* B **41** 4842
[11] Ovchinnikov A A 1993 *Mod. Phys. Lett.* B **7** 1397
[12] Bloom P 1975 *Phys. Rev.* B **12** 125
[13] Cyrot M 1977 *Physica* B **91** 141
[14] de Vries P, Michielsen K and De Raedt H 1993 *Phys. Rev. Lett.* **70** 2463
[15] Lieb E H and Wu F Y 1968 *Phys. Rev. Lett.* **20** 1445
[16] Ogata M and Shiba H 1990 *Phys. Rev.* B **41** 2326
[17] Cloizeaux J and Gaudin M 1966 *J. Math. Phys.* **7** 1384
[18] Schlottmann P 1987 *Phys. Rev.* B **36** 5177
[19] Sutherland B 1975 *Phys. Rev.* B **12** 3795
[20] Bares P A, Blatter G and Ogata M 1991 *Phys. Rev.* B **44** 130
[21] Yang C N 1989 *Phys. Rev. Lett.* **63** 2144
Yang C N and Zhang S 1990 *Mod. Phys. Lett.* B **4** 759
[22] de Boer J, Korepin V E and Schadschneider A *Preprint* ITP-SB-94-18