

6 Introduction to the Hubbard Model

Robert Eder

Institut für Festkörperphysik

Karlsruhe Institute of Technology

Contents

1	The Hubbard model	2
2	Some notation and review of Green functions	3
3	The Hubbard dimer: solution by exact diagonalization	5
4	Spin density wave theory	8
5	The Hubbard-I approximation	14
6	The Gutzwiller wave function	20
7	Summary and discussion	27

1 The Hubbard model

The Hubbard model was proposed in the 1960s to describe electrons in $3d$ transition metals. In these elements, the radial wave function of the $3d$ -electrons has a very small spatial extent. Therefore, when the $3d$ shell is occupied by several electrons, these are forced to be close to one another on the average so that the electrostatic energy is large. The energy of a given transition metal ion therefore varies strongly with the number of electrons it contains. To study the motion of conduction electrons under the influence of this strong Coulomb repulsion Hubbard [1], Kanamori [2], and Gutzwiller [3] proposed a simplified model. Thereby both, the five-fold degeneracy of the $3d$ -orbital and the presence of other bands in the solid are neglected. Rather, one considers a lattice of sites – whereby the geometry of the lattice is not really specified – with one s -like orbital at each site. Orbitals on different sites are assumed to be orthogonal, but for not too distant sites i and j there are nonvanishing matrix elements $t_{i,j}$ of the Hamiltonian between the orbitals centered on these sites. The Coulomb interaction between electrons in orbitals on different sites is neglected, but if two electrons – which then necessarily have opposite spin – occupy the same orbital the energy is assumed to increase by the large amount U to simulate the strong dependence of the energy on the occupation number. If we denote the creation operator for an electron of spin σ in the orbital at lattice site i by $c_{i,\sigma}^\dagger$ the model thus can be written as

$$H = \sum_{i,j} \sum_{\sigma} t_{i,j} c_{i,\sigma}^\dagger c_{j,\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow} = H_t + H_U. \quad (1)$$

Here $n_{i,\sigma} = c_{i,\sigma}^\dagger c_{i,\sigma}$ counts the number of electrons with spin σ in the orbital at site i .

After the discovery of the cuprate superconductors in 1987 and after Zhang and Rice demonstrated [4] that the CuO_2 planes in these compounds can be described by the so-called t - J model – which is equivalent to the Hubbard model in the limit $U/t \gg 1$ – there was renewed interest in the 2-dimensional Hubbard model. However, the lightly doped Mott-insulator – which most probably is the system to be understood in order to solve the many puzzles posed by the cuprate superconductors – is still far from being solved. Accordingly, the purpose of this lecture is to present some of the basic approximations developed for this model.

We consider (1) for a two-dimensional square lattice with N sites, lattice constant $a = 1$, and periodic boundary conditions. For hopping integrals $-t$ between nearest $((1, 0)$ -like) neighbors and $-t'$ between 2^{nd} nearest $((1, 1)$ -like) neighbors the dispersion relation is

$$\varepsilon_{\mathbf{k}} = -2t (\cos(k_x) + \cos(k_y)) - 4t' \cos(k_x) \cos(k_y). \quad (2)$$

The number of electrons with spin σ in the system is denoted by N_σ – whereby we are mostly interested in the nonmagnetic case $N_\uparrow = N_\downarrow$ – so that the number of electrons is $N_e = N_\uparrow + N_\downarrow$. In the following, densities per site will be denoted n , e.g., $n_\uparrow = N_\uparrow/N$. For $n_e = 1$ we have $N_\uparrow = N_\downarrow = N/2$ so that precisely half of the \mathbf{k} -points for each spin direction are occupied and we have a half-filled band, i.e., a metal in conventional band theory. Instead it will be shown below that for sufficiently large U/t the Hubbard model describes an insulator, the so-called Mott-insulator. The region of primary interest for cuprate superconductors is $n_e \geq 0.8$, i.e., the lightly doped Mott-insulator, and $U/t \approx 10$.

2 Some notation and review of Green functions

We first introduce some notation that will be used frequently and give a brief review of imaginary time Green functions. There are numerous excellent textbooks on the use of field theory for condensed matter physics where more details can be found [5–7].

The thermal average of any operator \hat{O} is given by

$$\langle \hat{O} \rangle_{\text{th}} = \frac{1}{Z} \text{Tr} \left(e^{-\beta(H-\mu N)} \hat{O} \right) \quad (3)$$

with the grand partition function

$$Z = \text{Tr} \left(e^{-\beta(H-\mu N)} \right). \quad (4)$$

Introducing the imaginary-time Heisenberg operator (with $K = H - \mu N$)

$$c_\alpha(\tau) = e^{\tau K} c_\alpha e^{-\tau K} \quad \Rightarrow \quad -\frac{\partial c_\alpha(\tau)}{\partial \tau} = [c_\alpha(\tau), K] \quad (5)$$

the *imaginary time Green function* is defined as

$$G_{\alpha,\beta}(\tau) = -\left\langle T c_\alpha(\tau) c_\beta^\dagger \right\rangle_{\text{th}} = -\Theta(\tau) \left\langle c_\alpha(\tau) c_\beta^\dagger \right\rangle_{\text{th}} + \Theta(-\tau) \left\langle c_\beta^\dagger c_\alpha(\tau) \right\rangle_{\text{th}} \quad (6)$$

$$= \frac{1}{Z} \left(-\Theta(\tau) \sum_{i,j} e^{-\beta(E_i - \mu N_i)} e^{\tau(E_i - E_j + \mu)} \langle i | c_\alpha | j \rangle \langle j | c_\beta^\dagger | i \rangle \right. \\ \left. + \Theta(-\tau) \sum_{i,j} e^{-\beta(E_i - \mu N_i)} e^{\tau(E_j - E_i + \mu)} \langle i | c_\beta^\dagger | j \rangle \langle j | c_\alpha | i \rangle \right), \quad (7)$$

where α, β denote some set of quantum numbers, $|i\rangle$ are the exact eigenstates of the system with energies E_i and particle number N_i . Using $\frac{\partial \Theta(\tau)}{\partial \tau} = \delta(\tau)$ it follows from (5) and (6) that the Green function obeys the equation of motion

$$-\frac{\partial}{\partial \tau} G_{\alpha,\beta}(\tau) = \delta(\tau) \left\langle \left\{ c_\alpha, c_\beta^\dagger \right\} \right\rangle_{\text{th}} - \left\langle T [c_\alpha(\tau), K] c_\beta^\dagger \right\rangle_{\text{th}}. \quad (8)$$

It follows from (7) that G is well-defined only if $\tau \in [-\beta, \beta]$ (the reason is that the $E_i - \mu N_i$ are unbounded from above [8]) and that for $\tau \in [-\beta, 0]$ one has $G(\tau + \beta) = -G(\tau)$. It follows that $G(\tau)$ has the Fourier transform (see equation (25.10) in [6] with $\hbar = 1$)

$$G(\tau) = \frac{1}{\beta} \sum_{\nu=-\infty}^{\infty} e^{-i\omega_\nu \tau} G(i\omega_\nu), \quad G(i\omega_\nu) = \int_0^\beta d\tau e^{i\omega_\nu \tau} G(\tau), \quad \omega_\nu = \frac{(2\nu + 1)\pi}{\beta}. \quad (9)$$

The $i\omega_\nu$ are the (Fermionic) Matsubara frequencies. Inserting (7) into (9) one obtains

$$G_{\alpha\beta}(i\omega_\nu) = \frac{1}{Z} \sum_{i,j} \frac{e^{-\beta(E_i - \mu N_i)} + e^{-\beta(E_j - \mu N_j)}}{i\omega_\nu + \mu - (E_j - E_i)} \langle i | c_\alpha | j \rangle \langle j | c_\beta^\dagger | i \rangle. \quad (10)$$

We specialize to a single band and assume that the z -component of the spin is a good quantum number so that the Green function is a scalar and $\alpha = \beta = (\mathbf{k}, \sigma)$. The function $G(\mathbf{k}, i\omega_\nu)$

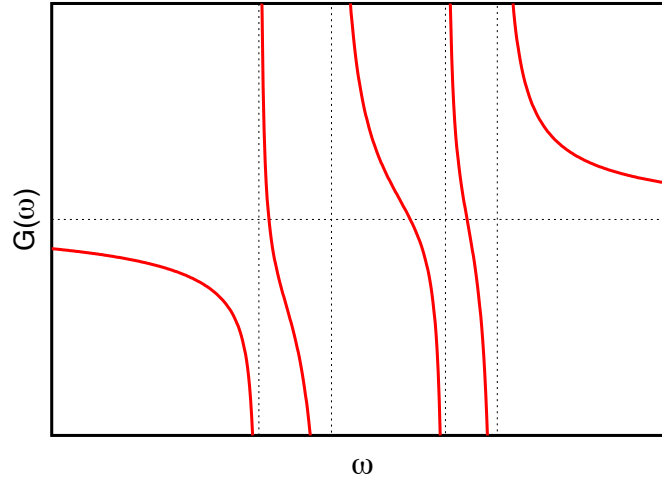


Fig. 1: The Green function $G(\omega)$ for real ω . The dashed vertical lines give the poles, ω_i .

can be analytically continued to the whole complex frequency plane by replacing $i\omega_\nu \rightarrow z$ where z is any complex number. As can be seen from (10) the function $G(z)$ is analytic in the complex z -plane except for the real axis where it has simple poles at $z = E_j - E_i - \mu$. It is this property on which the usefulness of the imaginary-time Green function is based: the analytic continuation of its Fourier transform $G(z)$ can be evaluated along a line infinitesimally above the real axis and then gives the Fourier transform of the retarded real-time Green function – from which the single-particle spectral function $A(\mathbf{k}, \omega)$, i.e., the combined photoemission and inverse photoemission spectrum of a system can be obtained:

$$A(\mathbf{k}, \omega) = -\frac{1}{\pi} \text{Im} G(\mathbf{k}, \omega + i0^+).$$

For this reason the Fourier transform (10) is often called ‘the’ Green function. Equation (10) is the *Lehmann representation* of the Green function.

It is shown in textbooks of field theory [5–7] that the imaginary-time Green function can be expanded in Feynman diagrams (whereas such an expansion is *not* possible for the real-time Green function at finite temperature) and the self-energy $\Sigma(\mathbf{k}, z)$ can be introduced as usual:

$$(z + \mu - \varepsilon_{\mathbf{k}} - \Sigma(\mathbf{k}, z)) G(\mathbf{k}, z) = 1. \quad (11)$$

Next we discuss the analytical structure of the Green function and the self energy. It can be seen from (10) that the Fourier transform of the Green function has the general form

$$G(z) = \sum_i \frac{\alpha_i}{z - \omega_i},$$

where the α_i and ω_i are real numbers. Along the real axis $G(\omega)$ therefore looks like in Fig. 1. This shows that in between any two successive poles ω_i and ω_{i+1} the Green function crosses zero with a negative slope

$$G(\omega) \approx -\beta_i(\omega - \zeta_i).$$

Near the crossing point ζ_i we thus have

$$\Sigma(\omega) = -G^{-1}(\omega) + \omega + \mu - \varepsilon_{\mathbf{k}} = \frac{\sigma_i}{\omega - \zeta_i} + \dots$$

where $\sigma_i = 1/\beta_i$. The self-energy thus has poles on the real axis as well, and these poles are ‘sandwiched’ in between the poles of the Green function. Luttinger has shown [9] that $\Sigma(\omega)$ is essentially determined by these poles and their residua in that it can be written in the whole complex frequency plane as

$$\Sigma(z) = V_{HF} + \sum_i \frac{\sigma_i}{z - \zeta_i} \quad (12)$$

where V_{HF} is equal to the Hartree-Fock potential (or rather its Fourier coefficient with the proper momentum).

3 The Hubbard dimer: solution by exact diagonalization

As a first example we consider the Hubbard model on a dimer

$$H = -t \sum_{\sigma} \left(c_{1,\sigma}^{\dagger} c_{2,\sigma} + c_{2,\sigma}^{\dagger} c_{1,\sigma} \right) + \sum_{i=1}^2 n_{i,\uparrow} n_{i,\downarrow}. \quad (13)$$

This can be solved by *exact diagonalization*, i.e., by constructing a basis of the entire Hilbert space, setting up the Hamilton matrix in this basis and diagonalizing it. The Hamiltonian (13) is invariant under the exchange of the site indices, $1 \leftrightarrow 2$, so that we can classify eigenstates by their parity P under this operation. Alternatively, we might view (13) as describing a ‘2-site ring with periodic boundary conditions’ and hopping integral $t/2$. Such a 2-site ring has two allowed momenta, $k = 0$ and $k = \pi$. The exchange of sites is equivalent to a translation by one lattice site, and since the definition of a Bloch state ψ_k with momentum k is $T_R \psi_k = e^{ikR} \psi_k$ it follows that $P = 1$ is equivalent to $k = 0$, whereas $P = -1$ means $k = \pi$. In the following we will always consider parity and momentum as interchangeable.

The Hilbert space can be decomposed into sectors with fixed electron number N_e , z -component of spin S_z , and parity P . We first consider the sector with $N_e = 2$ and $S_z = 0$, i.e., two electrons with opposite spin, which is equivalent to ‘half filling’. The basis states with $P = \pm 1$ are:

$$\begin{aligned} |1_{\pm}\rangle &= \frac{1}{\sqrt{2}} \left(c_{1,\uparrow}^{\dagger} c_{2,\downarrow}^{\dagger} \pm c_{2,\uparrow}^{\dagger} c_{1,\downarrow}^{\dagger} \right) |0\rangle, \\ |2_{\pm}\rangle &= \frac{1}{\sqrt{2}} \left(c_{1,\uparrow}^{\dagger} c_{1,\downarrow}^{\dagger} \pm c_{2,\uparrow}^{\dagger} c_{2,\downarrow}^{\dagger} \right) |0\rangle. \end{aligned} \quad (14)$$

For the ground state with $P = +1$ we make the ansatz $|\psi_0\rangle = u|1_+\rangle + v|2_+\rangle$. The relevant matrix elements are $\langle 1_+ | H | 2_+ \rangle = -2t$, $\langle 1_+ | H | 1_+ \rangle = 0$ and $\langle 2_+ | H | 2_+ \rangle = U$, so that the ground state energy E_0 and the coefficients u and v can be obtained by solving the eigenvalue problem

$$\begin{pmatrix} 0 & -2t \\ -2t & U \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = E_0 \begin{pmatrix} u \\ v \end{pmatrix}.$$

We find $E_0 = \frac{U}{2} - W$ with $W = \sqrt{\left(\frac{U}{2}\right)^2 + 4t^2}$ and

$$u = -\sqrt{\frac{W + \frac{U}{2}}{2W}}, \quad v = \frac{-2t}{\sqrt{2W(W + \frac{U}{2})}}. \quad (15)$$

Both $|1_+\rangle$ and $|2_+\rangle$ are spin singlets so that $|\psi_0\rangle$ is a singlet as well. For the states with $P = -1$ we have $\langle 1_-|H|2_-\rangle = 0$, whence the eigenstates are $|1_-\rangle$ with energy 0 and $|2_-\rangle$ with energy U . $|1_-\rangle$ is a spin triplet and the remaining two members of the triplet are $c_{1,\uparrow}^\dagger c_{2,\uparrow}^\dagger|0\rangle$ and $c_{1,\downarrow}^\dagger c_{2,\downarrow}^\dagger|0\rangle$. Both states are in fact the only basis states in the sectors $N_e = 2$, $S_z = \pm 1$ and $P = -1$ and hence are eigenstates by construction with energy 0. In the limit $U/t \rightarrow \infty$ we have $W \rightarrow \frac{U}{2} + \frac{4t^2}{U}$. The low energy spectrum of the dimer with two electrons then consists of four states: the spin singlet with energy $-\frac{4t^2}{U}$ and the three members of the triplet with energy 0. We can thus describe the dimer by an effective low energy Hamiltonian of the form

$$H = J \left(\mathbf{S}_1 \cdot \mathbf{S}_2 - \frac{n_1 n_2}{4} \right)$$

where $n_i = \sum_\sigma c_{i,\sigma}^\dagger c_{i,\sigma}$ is the operator of electron number and $\mathbf{S}_i = \frac{1}{2} \sum_{\sigma,\sigma'} c_{i,\sigma}^\dagger \boldsymbol{\tau}_{\sigma,\sigma'} c_{i,\sigma}$ the operator of electron spin at site i ($\boldsymbol{\tau}$ denotes the vector of Pauli matrices) and $J = \frac{4t^2}{U}$ – this is nothing but the *Heisenberg antiferromagnet*.

We can also construct states with $N_e = 1, 3$ and given parity and z -spin $\sigma = \pm \frac{1}{2}$:

$$\begin{aligned} |3_\pm, \sigma\rangle &= \frac{1}{\sqrt{2}} \left(c_{1,\sigma}^\dagger \pm c_{2,\sigma}^\dagger \right) |0\rangle, \\ |4_\pm, \sigma\rangle &= \frac{1}{\sqrt{2}} \left(c_{1,\sigma}^\dagger c_{2,\uparrow}^\dagger c_{2,\downarrow}^\dagger \pm c_{2,\sigma}^\dagger c_{1,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right) |0\rangle. \end{aligned}$$

Since these states are the only ones in their respective (N_e, S_z, P) sector they are again eigenstates by construction and indeed $H|3_\pm, \sigma\rangle = \mp t|3_\pm, \sigma\rangle$ and $H|4_\pm, \sigma\rangle = (U \pm t)|4_\pm, \sigma\rangle$. The energies of the single-electron states can be written in the familiar form $\varepsilon_k = 2 \frac{t}{2} \cos(k)$ expected for a two-site ring with periodic boundary conditions and hopping integral $\frac{t}{2}$. Having found all eigenstates $|j\rangle$ and their energies E_j we can write down the Green function $G(k, \omega)$ using the Lehmann representation (10). Thereby we simplify matters by taking the limit of low temperature and assuming that the chemical potential μ has been chosen such that the thermal occupation factor $e^{-\beta E_j}/Z$ is unity for $|j\rangle = |\psi_0\rangle$ and zero for all other states (this can be achieved by setting, e.g., $\mu = \frac{U}{2}$). Defining electron operators with definite parity by $c_{\pm, \sigma} = \frac{1}{\sqrt{2}}(c_{1,\sigma} \pm c_{2,\sigma})$ we find

$$\begin{aligned} c_{\pm, \sigma} |\psi_0\rangle &= \frac{1}{2} \left(c_{1,\sigma}^\dagger \pm c_{2,\sigma}^\dagger \right) \left(u \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right) + v \left(c_{1,\uparrow}^\dagger c_{1,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{2,\downarrow}^\dagger \right) \right) |0\rangle \\ &= \frac{\pm u + v}{\sqrt{2}} |3_\pm, \bar{\sigma}\rangle, \\ c_{\pm, \sigma}^\dagger |\psi_0\rangle &= \frac{\mp u + v}{\sqrt{2}} |4_\pm, \sigma\rangle, \end{aligned}$$

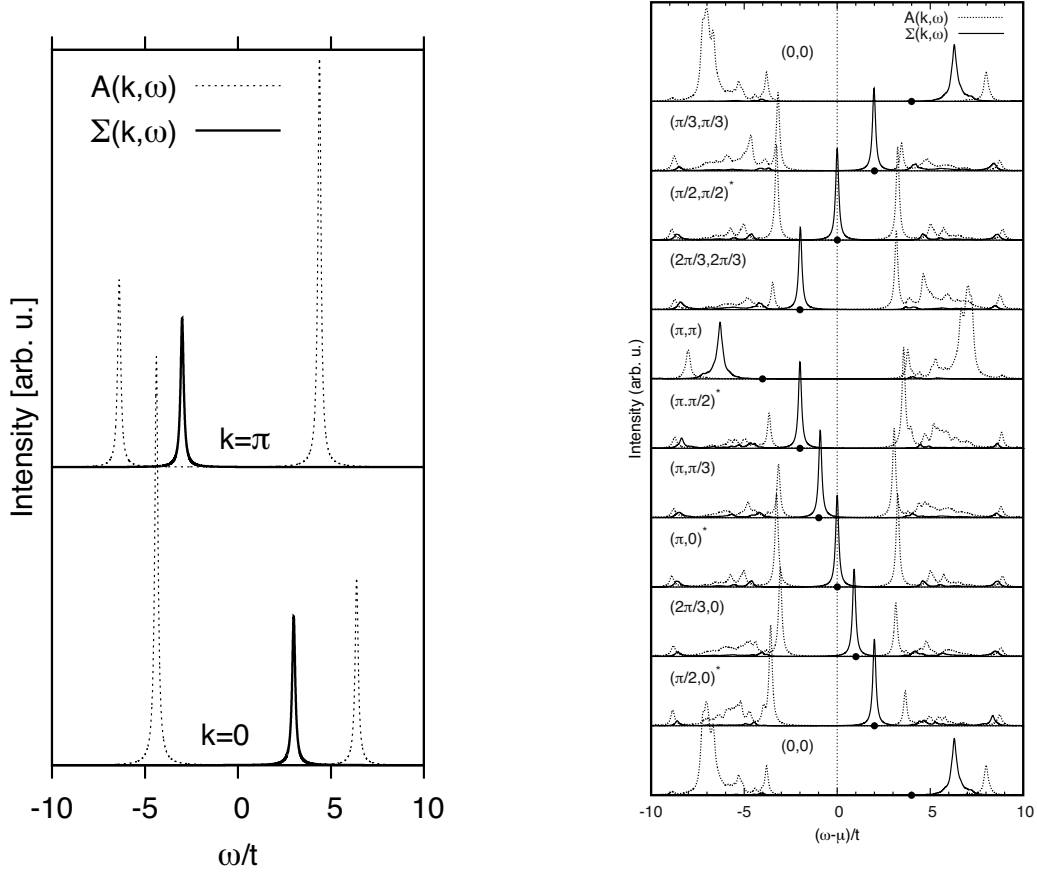


Fig. 2: Single particle spectral function and imaginary part of the self-energy for the Hubbard model with $U/t = 10$, δ -functions have been replaced by Lorentzians. Left: for the dimer, right for the 16 and 18-site cluster with periodic boundary conditions (reprinted with permission from [10], Copyright 2011 by the American Physical society). The dots correspond to $\zeta_{\mathbf{k}} = \frac{U}{2} - \mu - \varepsilon_{\mathbf{k}}$.

and using (15) for u and v we find the squared matrix elements

$$|\langle 3_{\pm}, \bar{\sigma} | c_{\pm, \sigma} | \psi_0 \rangle|^2 = \frac{1}{2} \pm \frac{t}{W}, \quad |\langle 4_{\pm}, \sigma | c_{\pm, \sigma}^{\dagger} | \psi_0 \rangle|^2 = \frac{1}{2} \mp \frac{t}{W}.$$

Using (10) the Fourier transform of the Green function $G_{\pm}(\omega) = -\langle T c_{\pm}(\tau) c_{\pm}^{\dagger} \rangle$ is

$$G_{\pm}(\omega) = \frac{\frac{1}{2} \pm \frac{t}{W}}{\omega + \mu - (E_0 \pm t)} + \frac{\frac{1}{2} \mp \frac{t}{W}}{\omega + \mu - (U \pm t - E_0)}.$$

Here the first term corresponds to $|j\rangle = |\psi_0\rangle$, whereas the second term corresponds to $|i\rangle = |\psi_0\rangle$ in (10). Equating $G_{\pm}^{-1}(\omega) = \omega + \mu \pm t - \Sigma_{\pm}(\omega)$ we can now solve for the self energy:

$$\Sigma_{\pm}(\omega) = \frac{U}{2} + \frac{\left(\frac{U}{2}\right)^2}{\omega + \mu \mp 3t - \frac{U}{2}}.$$

$\Sigma_{\pm}(\omega)$ is indeed consistent with the general form (12) and the additive constant $U/2$ is indeed the Hartree-Fock potential. For the dimer $\Sigma_{\pm}(\omega)$ has only a single pole of strength $(U/2)^2$ which does have a substantial dispersion, in that its positions for $k = 0$ and $k = \pi$ differ by a full $6t$. Plots of $A(k, \omega)$ and the imaginary part of $\Sigma(k, \omega)$ are given in Figure 2. The part

of $A(k, \omega)$ for $\omega < 0$ gives the photoemission spectrum (PES) whereas the part for $\omega > 0$ is the inverse photoemission spectrum (IPES). PES and IPES form two ‘bands’ separated by an energy gap of $\approx U$ – these are the lower and upper Hubbard band. The pole of the self energy is within the Hubbard gap and its dispersion is downward as one goes from $k = 0$ to $k = \pi$, that means ‘inverted’ as compared to that of the energy ε_k . Interestingly, some of the features observed in the dimer generalize to larger systems. Figure 2 also shows $A(k, \omega)$ and the imaginary part of $\Sigma(k, \omega)$ for the allowed momenta of the 16- and 18-site cluster which were obtained by Lanczos diagonalization [10]. In $A(k, \omega)$ one can again recognize the lower and upper Hubbard band – for the larger clusters these ‘bands’ consist not only of sharp peaks but contain extended incoherent continua. As was the case for the dimer, $\Sigma(k, \omega)$ has an isolated peak of strength $\propto (U/2)^2$ within the Hubbard gap. Over a large part of the Brillouin zone this peak has a smooth dispersion which closely follows the ‘inverted’ band dispersion, $\zeta_k = \frac{U}{2} - \varepsilon_k$.

4 Spin density wave theory

Next, we discuss spin density wave theory which is a mean-field theory and which is reasonably successful in describing the antiferromagnetic phase of the Hubbard model. By comparing the result of applying both sides to the four possible states of a single atom – $|0\rangle, |\uparrow\rangle, |\downarrow\rangle$ and $|\uparrow\downarrow\rangle$ – it is easy to verify that

$$U n_{i,\uparrow} n_{i,\downarrow} = \frac{n_i U}{2} - \frac{2U}{3} \mathbf{S}_i^2.$$

This equation shows that the system can lower its Coulomb energy by forming *magnetic moments*, $\langle \mathbf{S}_i^2 \rangle \neq 0$, where $\langle \dots \rangle$ denotes the expectation value. These magnetic moments may be either static – so that $\langle \mathbf{S}_i \rangle \neq 0$ – or fluctuating, i.e., $\langle \mathbf{S}_i \rangle = 0$. An example for fluctuating moments is provided by the Hubbard dimer discussed above. Namely for the ground state $|\psi_0\rangle$ of the dimer

$$\begin{aligned} S_1^z |\psi_0\rangle &= S_1^z \left(\frac{u}{\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right) + \frac{v}{\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{1,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{2,\downarrow}^\dagger \right) \right) \\ &= \frac{u}{2\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger - c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right) \\ \mathbf{S}_1^2 |\psi_0\rangle &= \frac{3u}{4\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right), \end{aligned}$$

so that indeed $\langle \psi_0 | S_1^z | \psi_0 \rangle = 0$ but $\langle \psi_0 | \mathbf{S}_1^2 | \psi_0 \rangle = \frac{3u^2}{4}$ (it is easy to show that $\langle \psi_0 | S_1^x | \psi_0 \rangle = \langle \psi_0 | S_1^y | \psi_0 \rangle = 0$ as well and everything also holds true for site 2).

Spin density wave theory, on the other hand, assumes static moments, $\langle \mathbf{S}_i \rangle \neq 0$. Based on the results for the Hubbard dimer, which showed that the spins have antiferromagnetic correlations, we expect that these static magnetic moments prefer to be antiparallel on neighboring sites. For a 2D square lattice this requirement defines the Néel state: we choose

$$\begin{aligned} \langle n_{i,\uparrow} \rangle &= \frac{n_e}{2} + \frac{m}{2} e^{i\mathbf{Q} \cdot \mathbf{R}_i} \\ \langle n_{i,\downarrow} \rangle &= \frac{n_e}{2} - \frac{m}{2} e^{i\mathbf{Q} \cdot \mathbf{R}_i} \end{aligned} \quad \Rightarrow \quad \langle n_i \rangle = n_e \quad \text{and} \quad \langle S_i^z \rangle = \frac{m}{2} e^{i\mathbf{Q} \cdot \mathbf{R}_i}.$$

Here $\mathbf{Q} = (\pi, \pi)$ and the exponential $e^{i\mathbf{Q}\cdot\mathbf{R}_i}$ is $+1$ (-1) if the sum of x - and y -component of the site i is even (odd) which defines the sublattices A (with $e^{i\mathbf{Q}\cdot\mathbf{R}_i} = 1$) and B (with $e^{i\mathbf{Q}\cdot\mathbf{R}_i} = -1$). We consider the trivial decomposition $n_{i,\sigma} = \langle n_{i,\sigma} \rangle + \delta n_{i,\sigma}$, where the operator $\delta n_{i,\sigma} = n_{i,\sigma} - \langle n_{i,\sigma} \rangle$ describes fluctuations of n_i around its mean value. The basic assumption of all mean-field theories is that these fluctuations are ‘small’. Accordingly, when forming the product $U n_{i,\uparrow} n_{i,\downarrow}$ the ‘2nd order term’ $\delta n_{i,\uparrow} \cdot \delta n_{i,\downarrow}$ is discarded, so that the interaction term becomes

$$\begin{aligned} U \sum_i n_{i,\uparrow} n_{i,\downarrow} &\rightarrow U \sum_i \left(n_{i,\uparrow} \langle n_{i,\downarrow} \rangle + n_{i,\downarrow} \langle n_{i,\uparrow} \rangle - \langle n_{i,\uparrow} \rangle \langle n_{i,\downarrow} \rangle \right) \\ &= \frac{n_e U}{2} \sum_{i,\sigma} n_{i,\sigma} - \frac{mU}{2} \sum_i e^{i\mathbf{Q}\cdot\mathbf{R}_i} (n_{i,\uparrow} - n_{i,\downarrow}) - NU \frac{n_e^2 - m^2}{4}. \end{aligned}$$

Switching to Fourier transformed operators the mean-field Hamiltonian $K_{MF} = H_{MF} - \mu N$ becomes

$$K_{MF} = \sum_{\mathbf{k},\sigma} \tilde{\varepsilon}_{\mathbf{k}} c_{\mathbf{k},\sigma}^\dagger c_{\mathbf{k},\sigma} - \Delta \sum_{\mathbf{k}} \left(c_{\mathbf{k}+\mathbf{Q},\uparrow}^\dagger c_{\mathbf{k},\uparrow} - c_{\mathbf{k}+\mathbf{Q},\downarrow}^\dagger c_{\mathbf{k},\downarrow} \right) - NU \frac{n_e^2 - m^2}{4}, \quad (16)$$

where $\tilde{\varepsilon}_{\mathbf{k}} = \varepsilon_{\mathbf{k}} + \frac{n_e U}{2} - \mu$ and $\Delta = \frac{mU}{2}$. The term $\propto \Delta$ appears to be non-Hermitian at first glance. However, by shifting the summation variable $\mathbf{k} \rightarrow \mathbf{k} + \mathbf{Q}$ and noting that $\mathbf{k} + 2\mathbf{Q} = \mathbf{k}$ one can see that the term is in fact its own Hermitian conjugate.

The Hamiltonian (16) is quadratic in the Fermion operators and thus can be diagonalized by a unitary transformation. In a second step, the value of m is determined by recomputing $\langle n_{i,\sigma} \rangle$ from this solution and demanding self-consistency. We will now carry out this program thereby following Gorkov’s re-derivation of BCS theory [11] in terms of the imaginary time Green functions introduced in the first section.

We define the Green function $G_\sigma(\mathbf{k}, \tau) = -\langle T c_{\mathbf{k},\sigma}(\tau) c_{\mathbf{k},\sigma}^\dagger \rangle$ and with the Hamiltonian (16) its equation of motion is

$$\begin{aligned} -\frac{\partial}{\partial \tau} G_\sigma(\mathbf{k}, \tau) &= \delta(\tau) \langle \{c_{\mathbf{k},\sigma}^\dagger, c_{\mathbf{k},\sigma}\} \rangle - \langle T [c_{\mathbf{k},\sigma}(\tau), K_{MF}] c_{\mathbf{k},\sigma}^\dagger \rangle \\ &= \delta(\tau) + \tilde{\varepsilon}_{\mathbf{k}} G_\sigma(\mathbf{k}, \tau) \mp \Delta \tilde{G}_\sigma(\mathbf{k}, \tau), \end{aligned}$$

where the upper sign holds for $\sigma = \uparrow$ and we have introduced the *anomalous Green function* $\tilde{G}_\sigma(\mathbf{k}, \tau) = -\langle T c_{\mathbf{k}+\mathbf{Q},\sigma}(\tau) c_{\mathbf{k},\sigma}^\dagger \rangle$. In a nonmagnetic system $\tilde{G}_\sigma(\mathbf{k}, \tau)$ would be zero due to momentum conservation but in a magnetic system it can be different from zero. Its equation of motion is

$$\begin{aligned} -\frac{\partial}{\partial \tau} \tilde{G}_\sigma(\mathbf{k}, \tau) &= \delta(\tau) \langle \{c_{\mathbf{k},\sigma}^\dagger, c_{\mathbf{k}+\mathbf{Q},\sigma}\} \rangle - \langle T [c_{\mathbf{k}+\mathbf{Q},\sigma}(\tau), K_{MF}] c_{\mathbf{k},\sigma}^\dagger \rangle \\ &= \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}} \tilde{G}_\sigma(\mathbf{k}, \tau) \mp \Delta G_\sigma(\mathbf{k}, \tau). \end{aligned}$$

The system of equations of motion therefore closes and upon Fourier transformation with respect to τ – whereby $-\partial_\tau \rightarrow i\omega_\nu$ – and becomes

$$\begin{pmatrix} i\omega_\nu - \tilde{\varepsilon}_{\mathbf{k}} & \pm \Delta \\ \pm \Delta & i\omega_\nu - \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}} \end{pmatrix} \begin{pmatrix} G_\sigma(\mathbf{k}, i\omega_\nu) \\ \tilde{G}_\sigma(\mathbf{k}, i\omega_\nu) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (17)$$

Next we recall [12] that for any complex number a and any complex vector \mathbf{b} :

$$(a\mathbf{1} + \mathbf{b} \cdot \boldsymbol{\tau})(a\mathbf{1} - \mathbf{b} \cdot \boldsymbol{\tau}) = a^2 - \mathbf{b}^2 \quad \Rightarrow \quad (a\mathbf{1} + \mathbf{b} \cdot \boldsymbol{\tau})^{-1} = \frac{a\mathbf{1} - \mathbf{b} \cdot \boldsymbol{\tau}}{a^2 - \mathbf{b}^2}.$$

Defining

$$\zeta_{\mathbf{k}} = \frac{1}{2} (\tilde{\varepsilon}_{\mathbf{k}} + \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}), \quad \eta_{\mathbf{k}} = \frac{1}{2} (\tilde{\varepsilon}_{\mathbf{k}} - \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}), \quad (18)$$

the 2×2 coefficient matrix H on the left hand side of (17) can be brought to the form $a\mathbf{1} + \mathbf{b} \cdot \boldsymbol{\tau}$ with $a = i\omega_{\nu} - \zeta_{\mathbf{k}}$ and $\mathbf{b} = (\pm\Delta, 0, -\eta_{\mathbf{k}})$. Multiplying both sides of (17) by H^{-1} we obtain

$$\begin{aligned} G_{\sigma}(\mathbf{k}, i\omega_{\nu}) &= \frac{i\omega_{\nu} - \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}}{(i\omega_{\nu} - \zeta_{\mathbf{k}})^2 - W_{\mathbf{k}}^2} = \frac{Z_{\mathbf{k}}^{(-)}}{i\omega_{\nu} - E_{\mathbf{k}}^{(-)}} + \frac{Z_{\mathbf{k}}^{(+)}}{i\omega_{\nu} - E_{\mathbf{k}}^{(+)}} \\ \tilde{G}_{\sigma}(\mathbf{k}, i\omega_{\nu}) &= \frac{\mp\Delta}{(i\omega_{\nu} - \zeta_{\mathbf{k}})^2 - W_{\mathbf{k}}^2} = \frac{\mp\Delta}{(i\omega_{\nu} - E_{\mathbf{k}}^{(-)})(i\omega_{\nu} - E_{\mathbf{k}}^{(+)})}, \end{aligned} \quad (19)$$

where $W_{\mathbf{k}}^2 = \eta_{\mathbf{k}}^2 + \Delta^2$ and we have introduced the quasiparticle energies and weights

$$E_{\mathbf{k}}^{(\pm)} = \zeta_{\mathbf{k}} \pm W_{\mathbf{k}} \quad Z_{\mathbf{k}}^{(\pm)} = \frac{1}{2} \left(1 \pm \frac{\eta_{\mathbf{k}}}{W_{\mathbf{k}}} \right). \quad (20)$$

From the expression for G we obtain the single particle spectral function

$$A(\mathbf{k}, \omega) = Z_{\mathbf{k}}^{(-)} \delta(\omega - E_{\mathbf{k}}^{(-)}) + Z_{\mathbf{k}}^{(+)} \delta(\omega - E_{\mathbf{k}}^{(+)}).$$

Rather than a single band with dispersion $\varepsilon_{\mathbf{k}}$ and weight 1, SDW theory thus predicts two bands with a reduced and \mathbf{k} -dependent spectral weight. Figure (3) shows $A(\mathbf{k}, \omega)$ obtained by evaluating (19) for different values of Δ . Increasing Δ opens a gap in the original band. For small Δ the spectral weight of the bands has a substantial \mathbf{k} -dependence and for the photoemission spectrum ($\omega < 0$) drops sharply upon crossing the noninteracting Fermi surface at $(\frac{\pi}{2}, \frac{\pi}{2})$. So far we have carried out the first step, the solution of the mean-field Hamiltonian. In the second step we have to recompute $\langle \mathbf{S}_i \rangle$ from this solution. We note that for *any* site i

$$n_e = \langle n_{i,\uparrow} + n_{i,\downarrow} \rangle \quad \text{and} \quad m = \langle n_{i,\uparrow} - n_{i,\downarrow} \rangle e^{i\mathbf{Q} \cdot \mathbf{R}_i}.$$

We sum this over i , divide by N , and switch to Fourier transformed c -operators:

$$\begin{aligned} n_e &= \frac{1}{N} \sum_{\mathbf{k}} \langle c_{\mathbf{k},\uparrow}^{\dagger} c_{\mathbf{k},\uparrow} + c_{\mathbf{k},\downarrow}^{\dagger} c_{\mathbf{k},\downarrow} \rangle = \frac{1}{N} \sum_{\mathbf{k}} \left(G_{\uparrow}(\mathbf{k}, \tau = 0^-) + G_{\downarrow}(\mathbf{k}, \tau = 0^-) \right) \\ &= \frac{2}{N} \sum_{\mathbf{k}} G_{\uparrow}(\mathbf{k}, \tau = 0^-) \\ m &= \frac{1}{N} \sum_{\mathbf{k}} \langle c_{\mathbf{k}+\mathbf{Q},\uparrow}^{\dagger} c_{\mathbf{k},\uparrow} - c_{\mathbf{k}+\mathbf{Q},\downarrow}^{\dagger} c_{\mathbf{k},\downarrow} \rangle = \frac{1}{N} \sum_{\mathbf{k}} \left(\tilde{G}_{\uparrow}(\mathbf{k}, \tau = 0^-) - \tilde{G}_{\downarrow}(\mathbf{k}, \tau = 0^-) \right) \\ &= \frac{2}{N} \sum_{\mathbf{k}} \tilde{G}_{\uparrow}(\mathbf{k}, \tau = 0^-), \end{aligned} \quad (21)$$

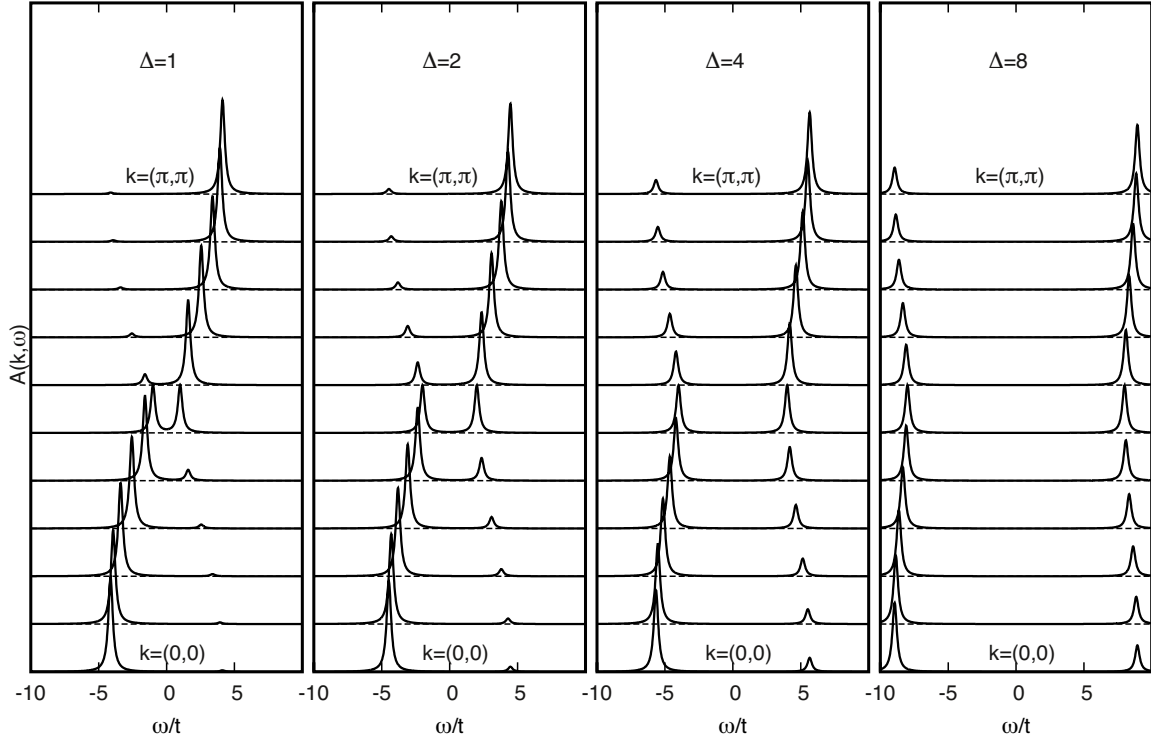


Fig. 3: Single particle spectral function $A(\mathbf{k}, \omega)$ along $(0, 0) \rightarrow (\pi, \pi)$ as obtained by spin density wave mean-field theory for different values of $\Delta = \frac{mU}{2}$ and half-filling. The original band dispersion is that of eqn. (2) with $t = 1$ and $t' = 0$. The part for $\omega < 0$ ($\omega > 0$) gives the photoemission (inverse photoemission) spectrum.

where we shifted the summation variable $\mathbf{k} \rightarrow \mathbf{k} + \mathbf{Q}$ and used (19) in the equation for m . Inserting the Fourier expansion (9) and using (19) for the Green functions we obtain

$$n_e = \frac{2}{N} \sum_{\mathbf{k}} \left(-\frac{1}{\beta} \right) \sum_{\nu} e^{i\omega_{\nu} 0^+} \left(\frac{-Z_{\mathbf{k}}^{(-)}}{i\omega_{\nu} - E_{\mathbf{k}}^{(-)}} + \frac{-Z_{\mathbf{k}}^{(+)}}{i\omega_{\nu} - E_{\mathbf{k}}^{(+)}} \right),$$

$$m = \frac{2}{N} \sum_{\mathbf{k}} \left(-\frac{1}{\beta} \right) \sum_{\nu} e^{i\omega_{\nu} 0^+} \frac{\Delta}{(i\omega_{\nu} - E_{\mathbf{k}}^{(-)})(i\omega_{\nu} - E_{\mathbf{k}}^{(+)})}.$$

We now replace the sum over Matsubara frequencies by a contour integration, which is a standard technique in field theory (see, e.g., section 25 of [6]). Namely for any function $F(\omega)$ which is analytic along the imaginary axis we have

$$-\frac{1}{\beta} \sum_{\nu} F(i\omega_{\nu}) = \frac{1}{2\pi i} \oint_{\mathcal{C}} d\omega f(\omega) F(\omega)$$

where

$$f(\omega) = \frac{1}{e^{\beta\omega} + 1},$$

is the Fermi function and the contour \mathcal{C} encircles the imaginary axis in counterclockwise fashion, see Fig. 4(a). This replacement makes use of the theorem of residues and the easily verified fact that the Fermi function $f(\omega)$ has simple poles with residuum $-1/\beta$ at all Matsubara frequencies $i\omega_{\nu}$. Next we note that the integrals along the two clover-shaped contours in Fig. 4(b)

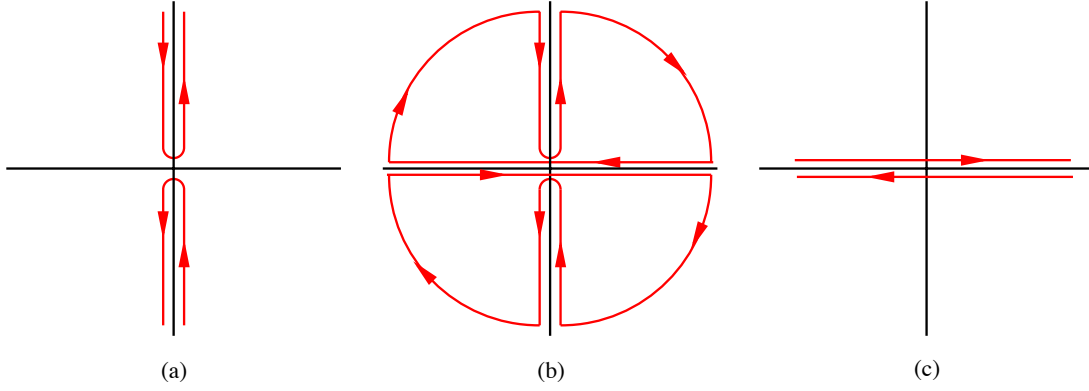


Fig. 4: Since the integrals along the two contours in (b) are zero and the contributions from the circular arcs vanish, the integral along the contour in (a) is equal to that over contour (c).

are zero, *provided* the integrand is analytic in the interior of the two curves. Since the Fermi function $f(\omega)$ has all of its poles along the imaginary axis, whereas the Green function has its poles on the real axis, both of which are outside of the curves in Fig. 4(b), this is certainly true. Next, Jordan's lemma can be invoked to establish that the integrals along the large semicircles vanish. Here the Fermi function $f(\omega)$ guarantees that the contribution from the semicircle with $\text{Re}(\omega) > 0$ vanishes, whereas the factor $e^{\omega 0^+}$ does the same for the semicircle with $\text{Re}(\omega) < 0$ (which also shows that it was necessary to keep this factor). It follows that the integral along the contour \mathcal{C} in Fig. 4(a) is equal to that along the contour \mathcal{C}' in Fig. 4(c) (note the inverted direction of the curves in (c) as compared to (b)!). The contour in Fig. 4(c) encircles the real axis clockwise – it follows from the theorem of residues that the contour integral is $(-2\pi i)$ times the sum of the residues of the respective $f(\omega)F(\omega)$ at its two poles at $E_{\mathbf{k}}^{(\pm)}$. Using the definition $\Delta = \frac{mU}{2}$ we thus obtain

$$n_e = \frac{2}{N} \sum_{\mathbf{k}} \left(Z_{\mathbf{k}}^{(-)} f(E_{\mathbf{k}}^{(-)}) + Z_{\mathbf{k}}^{(+)} f(E_{\mathbf{k}}^{(+)}) \right)$$

$$1 = \frac{U}{N} \sum_{\mathbf{k}} \frac{1}{2W_{\mathbf{k}}} \left(f(E_{\mathbf{k}}^{-}) - f(E_{\mathbf{k}}^{+}) \right).$$

It follows from (2) that $\zeta_{\mathbf{k}} = 4t' \cos(k_x) \cos(k_y) + \frac{U}{2} - \mu$ and $\eta_{\mathbf{k}} = -2t (\cos(k_x) + \cos(k_y))$. Moreover, $\zeta_{\mathbf{k}+\mathbf{Q}} = \zeta_{\mathbf{k}}$ and $\eta_{\mathbf{k}+\mathbf{Q}} = -\eta_{\mathbf{k}}$, so that $W_{\mathbf{k}+\mathbf{Q}} = W_{\mathbf{k}}$. Next, (20) shows that $Z_{\mathbf{k}+\mathbf{Q}}^{(\pm)} = Z_{\mathbf{k}}^{(\mp)}$, $Z_{\mathbf{k}+\mathbf{Q}}^{(\pm)} + Z_{\mathbf{k}}^{(\pm)} = 1$ and $E_{\mathbf{k}+\mathbf{Q}}^{(\pm)} = E_{\mathbf{k}}^{(\pm)}$. These relations makes it possible to restrict the sum over momenta in (22) and (23) to the antiferromagnetic Brillouin zone (AFBZ) and we obtain

$$n_e = \frac{2}{N} \sum_{\mathbf{k} \in \text{AFBZ}} \left(f(E_{\mathbf{k}}^{(-)}) + f(E_{\mathbf{k}}^{(+)}) \right), \quad (22)$$

$$1 = \frac{U}{N} \sum_{\mathbf{k} \in \text{AFBZ}} \frac{1}{W_{\mathbf{k}}} \left(f(E_{\mathbf{k}}^{-}) - f(E_{\mathbf{k}}^{+}) \right). \quad (23)$$

For given $\varepsilon_{\mathbf{k}}$, U and n_e we now have a complete description of the system. For a qualitative discussion, let us assume that the lower band is completely occupied (i.e. $f(E_{\mathbf{k}}^{(-)}) = 1$) and the

upper one completely empty (i.e. $f(E_{\mathbf{k}}^{(+)}) = 0$). Since the number of momenta in the AFBZ is $\frac{N}{2}$ this implies $n_e = 1$ – or ‘half-filling’. In the paramagnetic phase $n_e = 1$ would correspond to a half-filled band whereas SDW theory gives an antiferromagnetic insulator. Since $\Delta = \frac{mU}{2}$ we expect $\Delta/t \rightarrow \infty$ as $U/t \rightarrow \infty$. It follows that $W_{\mathbf{k}} \rightarrow \Delta$ so from (23) we obtain $m \rightarrow 1$ in this limit, which means the fully polarized Néel state. For an electron density $n_e = 1 - \delta$ and in the limit $T \rightarrow 0$ the fraction of momenta in the AFBZ which are occupied is $1 - \delta$ as well – put another way, we have hole pockets which cover a fraction of δ of the AFBZ, and consequently, in the original zone scheme, also a fraction of δ of the full Brillouin zone. Interestingly, in a doped semiconductor – with electron density $n_e = 2 - \delta$ – the hole pockets would cover only a fraction of $\delta/2$ of the Brillouin zone.

The equation for the temperature where m starts to deviate from zero – the so-called Néel temperature T_N – can be obtained by taking the limit $m \rightarrow 0$ whence $W_{\mathbf{k}} \rightarrow |\eta_{\mathbf{k}}|$. Inserting this one finds that T_N is determined by the equation

$$1 = -\frac{U}{N} \sum_{\mathbf{k}} \frac{f(\varepsilon_{\mathbf{k}+\mathbf{Q}}) - f(\varepsilon_{\mathbf{k}})}{\varepsilon_{\mathbf{k}+\mathbf{Q}} - \varepsilon_{\mathbf{k}}}.$$

Note that the left hand side is guaranteed to be positive because $f(\omega)$ is a monotonously decreasing function of its argument and that the temperature appears implicitly in the Fermi functions. Let us assume that we keep the temperature constant and increase U starting from zero. We ask under what conditions ordering sets in for an as low U as possible. In order for the \mathbf{k} -sum to be large, there must be many pairs $(\mathbf{k}, \mathbf{k} + \mathbf{Q})$ such that their energies are close to one another and such that both momenta are close to the Fermi surface – because only then the difference of Fermi functions can be appreciable. This gives us the condition of *Fermi surface nesting*: the ordering vector \mathbf{Q} must connect as long sections of the Fermi surface as possible.

For the general case the system of equations (22) and (23) can be solved by starting with some value m_{in} , determining μ such that (22) gives the correct electron density, and then evaluating m_{out} from (23). By scanning m_{in} one can then determine the value where $m_{out} = m_{in}$. Fig. 5 shows some results obtained in this way. A detailed discussion of the resulting phase diagram of the Hubbard model for various band fillings and values of the hopping integrals t and t' can be found, e.g., in Refs. [14] – here we do not discuss this in detail.

As already mentioned the above derivation was originally invented by Gorkov to re-derive the BCS theory of superconductivity [11] and can be easily generalized to any mean-field theory. The formulation in terms of Green functions makes it easy to include the effects of disorder or spatial variations and the various techniques applied above – such as the use of equations of motion, expressing the order parameter in terms of the anomalous Green function and the evaluation of sums by contour integration – are applied again and again in many papers on advanced problems in superconductivity.

SDW theory describes some features of undoped cuprate superconductors and related compounds correctly, but fails even qualitatively in many aspects. For example, the ‘parent compound’ compounds La_2CuO_4 indeed is an antiferromagnetic insulator with a Néel temperature of around 300 K. SDW theory also qualitatively reproduces the phenomenon of the ‘remnant

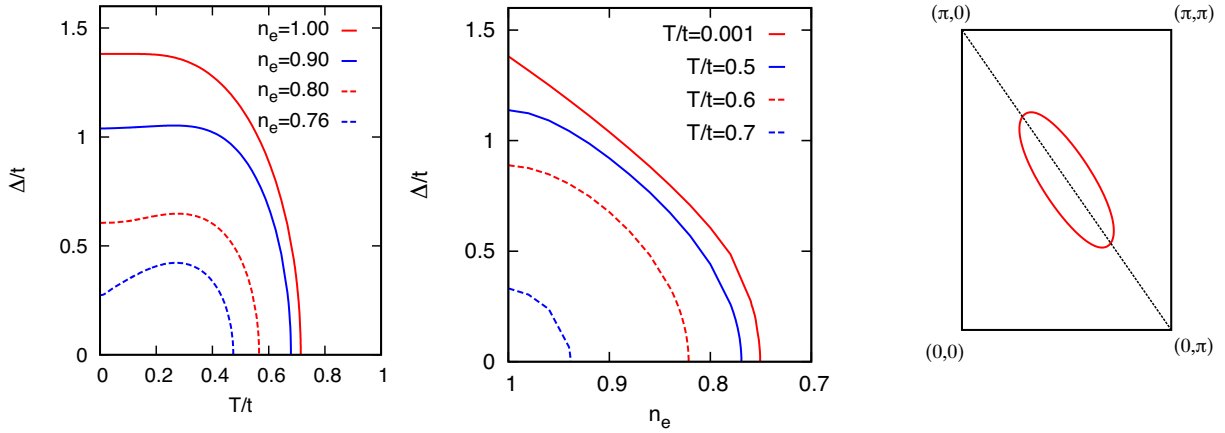


Fig. 5: Dependence of the SDW gap on temperature T for different electron densities n_e (left) and on electron density n_e for different temperatures T (center). Right: Fermi surface for an electron density $n_e = 0.9$. Parameter values are $U/t = 4$ and $t'/t = 0.2$.

Fermi surface' [13]. The phrase describes the experimental observation that the intensity of the quasiparticle band observed in ARPES experiments on undoped CuO-compounds such as $\text{Sr}_2\text{CuO}_2\text{Cl}_2$ drops sharply when crossing the noninteracting Fermi surface, precisely as seen in the spectra in Fig. 3. However, to reproduce the sharpness of the drop seen in experiment would require values of Δ which would be much too small to reproduce the magnitude of the insulating gap.

Moreover, contrary to the predictions of SDW theory the insulating gap does not go to zero at the Néel temperature but is essentially temperature independent - which is the hallmark of a true Mott-insulator. Moreover, for doped compounds such as $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ static antiferromagnetic order disappears already for hole concentrations of a few percent whereby again the Hubbard gap observed in the insulator persists with practically unchanged magnitude. What we therefore really need to describe is a *paramagnetic system* with a Hubbard gap and this will be the objective of the next section.

5 The Hubbard-I approximation

This is the ‘defining approximation’ of the Mott-insulator by which Hubbard for the first time introduced central concepts of strongly correlated electron systems such as the two Hubbard bands [1]. In the following we first give a sloppy re-derivation which is meant to clarify the physical content of the Hubbard-I approximation and then present Hubbard’s rigorous derivation in terms of Green functions.

We consider the limit of finite U and $t_{i,j} = 0$, $N_\uparrow = N_\downarrow = N/2$ so that $N_e = N$. The ground state has one electron per lattice site and the energy is $E = 0$. Since the spin of the electron at any given site is arbitrary this ground state is highly degenerate. We ignore this degeneracy and assume that there is a unique state $|\Psi_0\rangle$ which may be thought of as a suitable superposition of all these degenerate states and which we assume to be ‘disordered’ – it will become clear in a moment what this means.

Next we assume that a small but finite $t_{i,j}$ is switched on. Then, an electron of spin σ can be transferred from a site j to another site i resulting in an empty site at j and a double occupancy at site i . The energy thereby increases by U . The hopping process is possible only if the electron which was originally at the site i has the spin $-\sigma$ and since our initial state $|\Psi_0\rangle$ is ‘disordered’ the probability for this to be the case is $1/2$ – which is the definition of ‘disordered’. We now interpret the original state $|\Psi_0\rangle$ as the vacuum – denoted by $|0\rangle$ – of our theory and the state created by the hopping process as containing a Fermionic hole-like particle at j and a Fermionic double-occupancy-like particle at site i : $d_{i,\sigma}^\dagger h_{j,-\sigma}^\dagger |0\rangle$. The order of the Fermionic operators in this state is due to the fact that in the original hopping term the annihilation operator $c_{j,\sigma}$ which creates the hole stands to the right of the creation operator $c_{i,\sigma}^\dagger$ which creates the double occupancy. Moreover we assign the negative spin to the operator which creates the hole because replacement of, e.g., an \uparrow -electron by a hole decreases the z -spin by $1/2$. We obtain the following Hamiltonian to describe the holes and double occupancies:

$$H_{\text{eff},1} = \frac{1}{2} \sum_{i,j} \sum_{\sigma} \left(t_{i,j} d_{i,\sigma}^\dagger h_{j,-\sigma}^\dagger + H.c. \right) + U \sum_{i,\sigma} d_{i,\sigma}^\dagger d_{i,\sigma}. \quad (24)$$

Once a hole and a double occupancy have been created, each of these particles may be transported further by the hopping term. If we assume that the surplus or missing electron retains its spin – which means that the double occupancies and holes propagate without ‘leaving a trace’ of inverted spins – for example a surplus \uparrow -electron can hop from site i to site j only if the spin at site j is \downarrow – we again have the probability $1/2$ for this process. We therefore can write down the second term terms for the effective Hamiltonian

$$H_{\text{eff},2} = \frac{1}{2} \sum_{i,j} \sum_{\sigma} t_{i,j} \left(d_{i,\sigma}^\dagger d_{j,\sigma} - h_{i,-\sigma}^\dagger h_{j,-\sigma} \right). \quad (25)$$

The negative sign of the hopping term for holes is due to the fact that the original hopping term has to be rewritten as $-t_{i,j} c_{j,\sigma} c_{i,\sigma}^\dagger$ to describe the propagation of a hole. Addition of (24) and (25) and Fourier transformation gives

$$H_{\text{eff}} = \sum_{\mathbf{k},\sigma} \left(\left(\frac{\varepsilon_{\mathbf{k}}}{2} + U \right) d_{\mathbf{k},\sigma}^\dagger d_{\mathbf{k},\sigma} - \frac{\varepsilon_{\mathbf{k}}}{2} h_{\mathbf{k},\sigma}^\dagger h_{\mathbf{k},\sigma} \right) + \sum_{\mathbf{k},\sigma} \frac{\varepsilon_{\mathbf{k}}}{2} \left(d_{\mathbf{k},\sigma}^\dagger h_{-\mathbf{k},-\sigma}^\dagger + H.c. \right), \quad (26)$$

where $\varepsilon_{\mathbf{k}}$ is the Fourier transform of $t_{i,j}$. Note that this now is a quadratic form where the Coulomb interaction is described by the extra energy of U for the double-occupancy-like ‘particle’. Via the unitary transformation

$$\begin{aligned} \gamma_{-, \mathbf{k}, \sigma} &= u_{\mathbf{k}} d_{\mathbf{k}, \sigma} + v_{\mathbf{k}} h_{-\mathbf{k}, -\sigma}^\dagger \\ \gamma_{+, \mathbf{k}, \sigma} &= -v_{\mathbf{k}} d_{\mathbf{k}, \sigma} + u_{\mathbf{k}} h_{-\mathbf{k}, -\sigma}^\dagger \end{aligned} \quad (27)$$

this can be solved, resulting in the dispersion relations for the lower and upper Hubbard band

$$E_{\mathbf{k}}^{(\pm)} = \frac{1}{2} \left(\varepsilon_{\mathbf{k}} + U \pm \sqrt{\varepsilon_{\mathbf{k}}^2 + U^2} \right). \quad (28)$$

In the limit $U/t \gg 1$ this simplifies to $E_{\mathbf{k},-} = \frac{\varepsilon_{\mathbf{k}}}{2}$, $E_{\mathbf{k},+} = \frac{\varepsilon_{\mathbf{k}}}{2} + U$ so that the original band with dispersion $\varepsilon_{\mathbf{k}}$ is split into two bands, separated by a gap of U and each having half of the original width. Qualitatively this is similar to the exact result for the Hubbard dimer. For the case of particle-hole symmetry the chemical potential is $U/2$ so that the lower band is completely filled, the upper one completely empty. Rather than being a metal – as expected for the situation of a half-filled band – the presence of the Coulomb interaction turns the system into an insulator. From the above we can see that this is the consequence of ‘expanding around’ the hypothetical ‘vacuum state’ $|\Psi_0\rangle$ with one electron per site so that we obtain a dilute gas of hole-like and double-occupancy-like particles which are created in pairs and propagate, whereby the double-occupancies have a large ‘energy of formation’ of U .

Next, we derive these results in a more rigorous fashion thereby following Hubbard’s original paper [1]. We split the electron operator into two components

$$c_{i,\sigma} = c_{i,\sigma} n_{i,-\sigma} + c_{i,\sigma} (1 - n_{i,-\sigma}) = \hat{d}_{i,\sigma} + \hat{c}_{i,\sigma}, \quad (29)$$

which obey $[\hat{d}_{i,\sigma}, H_U] = U \hat{d}_{i,\sigma}$ and $[\hat{c}_{i,\sigma}, H_U] = 0$. Next we define the four Green functions

$$G_{\alpha,\beta}(\mathbf{k}, \tau) = - \left\langle T \alpha_{\mathbf{k},\sigma}(\tau) \beta_{\mathbf{k},\sigma}^\dagger \right\rangle, \quad (30)$$

where $\alpha, \beta \in \{\hat{c}, \hat{d}\}$. These obey the equations of motion

$$-\frac{\partial}{\partial \tau} G_{\alpha,\beta}(\vec{k}, \tau) = \delta(\tau) \left\langle \left\{ \beta_{\mathbf{k},\sigma}^\dagger, \alpha_{\mathbf{k},\sigma} \right\} \right\rangle - \left\langle T [\alpha_{\mathbf{k},\sigma}(\tau), H] \beta_{\mathbf{k},\sigma}^\dagger \right\rangle.$$

The commutators $[\alpha_{\mathbf{k},\sigma}, H_U]$ are trivial but the commutators with the kinetic term H_t are involved. After some algebra – thereby using the identity $n_{i,\sigma} = \frac{n_i}{2} + \sigma S_i^z$ – we find:

$$\begin{aligned} [\hat{c}_{i,\uparrow}, H_t] &= \sum_j t_{ij} \left[\left(1 - \frac{n_e}{2}\right) c_{j,\uparrow} + (c_{j,\uparrow} S_i^z + c_{j,\downarrow} S_i^-) - \frac{1}{2} c_{j,\uparrow} (n_i - n_e) + c_{j,\downarrow}^\dagger c_{i,\downarrow} c_{i,\uparrow} \right], \\ [\hat{d}_{i,\uparrow}, H_t] &= \sum_j t_{ij} \left[\frac{n_e}{2} c_{j,\uparrow} - (c_{j,\uparrow} S_i^z + c_{j,\downarrow} S_i^-) + \frac{1}{2} c_{j,\uparrow} (n_i - n_e) - c_{j,\downarrow}^\dagger c_{i,\downarrow} c_{i,\uparrow} \right]. \end{aligned} \quad (31)$$

The first term on the r.h.s describes the ‘simple’ propagation of the hole. The second term is the contraction of the spin-1 operator S_i and the spinor $c_{j,\sigma}$ into a spin-1/2 object. It describes how a hole moves to site j but leaves behind a spin-excitation at site the i . Similarly, the third term describes hopping combined with the creation of a density excitation at site j whereas the last term describes the coupling of a pair-excitation (this would be important for negative U). The Hubbard-I approximation is obtained by keeping only the first term in each of the square brackets on the respective right hand side – obviously a rather crude approximation. After Fourier transformation we obtain

$$\begin{aligned} [\hat{c}_{\mathbf{k},\uparrow}, H] &\approx \left(1 - \frac{n_e}{2}\right) \varepsilon_{\mathbf{k}} (\hat{c}_{\mathbf{k},\uparrow} + \hat{d}_{\mathbf{k},\uparrow}) - \mu \hat{c}_{\mathbf{k},\uparrow} \\ [\hat{d}_{\mathbf{k},\uparrow}, H] &\approx \frac{n_e}{2} \varepsilon_{\mathbf{k}} (\hat{c}_{\mathbf{k},\uparrow} + \hat{d}_{\mathbf{k},\uparrow}) + U \hat{d}_{\mathbf{k},\uparrow} - \mu \hat{d}_{\mathbf{k},\uparrow}. \end{aligned}$$

Using the anticommutation relations $\{\hat{d}_{i,\sigma}^\dagger, \hat{d}_{i,\sigma}\} = n_{i-\sigma}$, $\{\hat{c}_{i,\sigma}^\dagger, \hat{c}_{i,\sigma}\} = (1 - n_{i-\sigma})$, $\{\hat{d}_{i,\sigma}^\dagger, \hat{c}_{i,\sigma}\} = \{\hat{c}_{i,\sigma}^\dagger, \hat{d}_{i,\sigma}\} = 0$ and putting $\langle n_{i,\sigma} \rangle = \frac{n_e}{2}$ we obtain the Fourier transformed equations of motion:

$$\begin{pmatrix} i\omega_\nu + \mu - (1 - \frac{n_e}{2})\varepsilon_{\mathbf{k}} & -(1 - \frac{n_e}{2})\varepsilon_{\mathbf{k}} \\ -\frac{n_e}{2}\varepsilon_{\mathbf{k}} & i\omega_\nu + \mu - \frac{n_e}{2}\varepsilon_{\mathbf{k}} - U \end{pmatrix} \begin{pmatrix} G_{\hat{c},\hat{c}} & G_{\hat{c},\hat{d}} \\ G_{\hat{d},\hat{c}} & G_{\hat{d},\hat{d}} \end{pmatrix} = \begin{pmatrix} 1 - \frac{n_e}{2} & 0 \\ 0 & \frac{n_e}{2} \end{pmatrix}.$$

We now use the identity (which holds for any 2×2 matrix)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

to solve for the Green function matrix $\mathbf{G}(\mathbf{k}, \omega)$. Since $c_{\mathbf{k},\sigma} = \hat{c}_{\mathbf{k},\sigma} + \hat{d}_{\mathbf{k},\sigma}$ the usual electron Green function $G(\mathbf{k}, \omega)$ is $G = G_{\hat{c},\hat{c}} + G_{\hat{c},\hat{d}} + G_{\hat{d},\hat{c}} + G_{\hat{d},\hat{d}}$, which can be brought to the form

$$G(\mathbf{k}, \omega) = \frac{Z_{\mathbf{k}}^{(-)}}{\omega - E_{\mathbf{k}}^{(-)}} + \frac{Z_{\mathbf{k}}^{(+)}}{\omega - E_{\mathbf{k}}^{(+)}} = \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} - \Sigma(\omega)} \quad (32)$$

whereby (with $W_{\mathbf{k}} = \sqrt{U^2 + \varepsilon_{\mathbf{k}}^2 - 2(1 - n_e)\varepsilon_{\mathbf{k}}U}$):

$$E_{\mathbf{k}}^{(\pm)} = \frac{1}{2}(U + \varepsilon_{\mathbf{k}} \pm W_{\mathbf{k}}) - \mu, \quad Z_{\mathbf{k}}^{(\pm)} = \frac{1}{2} \left(1 \pm \frac{\varepsilon_{\mathbf{k}} - (1 - n_e)U}{W} \right),$$

and the self-energy in Hubbard-I approximation is given by

$$\Sigma(\omega) = \frac{n_e}{2}U + \frac{n_e}{2} \left(1 - \frac{n_e}{2} \right) \frac{U^2}{\omega + \mu - (1 - \frac{n_e}{2})U}.$$

The additive constant $\frac{n_e}{2}U$ indeed is the Hartree-Fock potential and there is a single dispersionless pole whose strength is $(\frac{U}{2})^2$ at half-filling, $n_e = \frac{1}{2}$. The pole strength thus agrees with the exact solution for the dimer but the lack of dispersion shows that the Hubbard-I approximation would fail already for the Hubbard dimer.

If we specialize to half-filling and put $n_e = 1$ the quasiparticle energies become

$$E_{\mathbf{k}}^{(\pm)} = \frac{1}{2} \left(U + \varepsilon_{\mathbf{k}} \pm \sqrt{\varepsilon_{\mathbf{k}}^2 + U^2} \right) - \mu.$$

This is (up to the term μ) consistent with the result (28) obtained from the heuristic Hamiltonian (26), thus demonstrating the equivalence of this simple picture with the Hubbard-I approximation. In order to fix the chemical potential for electron densities $n_e < 1$ we write the operator of electron number as

$$\hat{N}_e = 2 \sum_i n_{i,\uparrow} n_{i,\downarrow} + \sum_i (n_{i,\uparrow}(1 - n_{i,\downarrow}) + n_{i,\downarrow}(1 - n_{i,\uparrow})) = \sum_{i,\sigma} (\hat{d}_{i,\sigma}^\dagger \hat{d}_{i,\sigma} + \hat{c}_{i,\sigma}^\dagger \hat{c}_{i,\sigma}).$$

The electron density n_e thus can be expressed in terms of the Green functions (30) as

$$n_e = \frac{2}{N} \sum_{\mathbf{k}} (G_{\hat{d},\hat{d}}(\mathbf{k}, \tau = 0^-) + G_{\hat{c},\hat{c}}(\mathbf{k}, \tau = 0^-)), \quad (33)$$

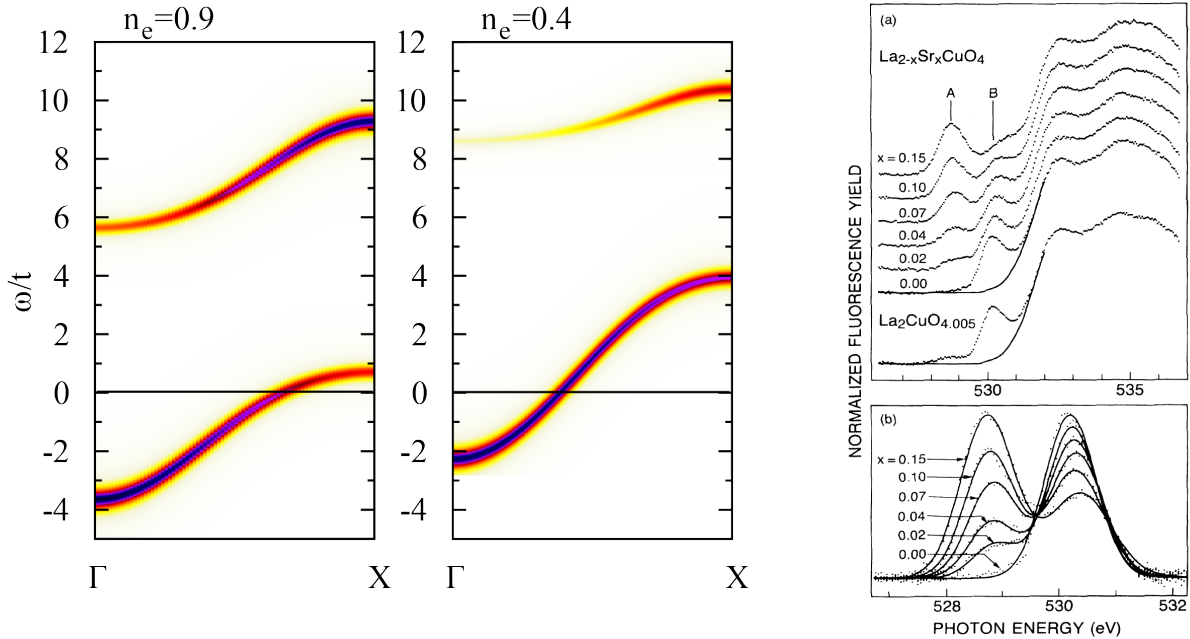


Fig. 6: *Left: Single particle spectral function obtained from the Greens function (32) for two different electron densities. The Fermi energy is zero. Right: X-ray luminescence spectra for $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ show the unoccupied part of the lower Hubbard-band (A) and the upper Hubbard-band (B). With decreasing $n_e \approx 1 - x$ the upper Hubbard-band rapidly loses intensity. Reprinted with permission from [15], Copyright 1991 by the American Physical society.*

which is analogous to the expression (21) and can be evaluated by the same procedure. To simplify matters we specialize to the limit $t/U \ll 1$ and close to half-filling, $n_e = 1 - \delta$ with $\delta \ll 1$. The dispersion relation of the lower band then is

$$E_{\mathbf{k}}^{(-)} \approx \frac{1 + \delta}{2} \varepsilon_{\mathbf{k}} - \frac{\varepsilon_{\mathbf{k}}^2}{4U} + \dots$$

where \dots denotes terms of higher order in the small quantities. For not too large t' this has its maximum at (π, π) . Assuming that the upper band is high above the chemical potential, the electron density becomes

$$n_e = \frac{2}{N} \sum_{\mathbf{k}} f(E_{\mathbf{k}}^{(-)}) \left(\frac{1 + \delta}{2} - \frac{\delta \varepsilon_{\mathbf{k}}^2}{4U^2} + \dots \right).$$

To discuss the Fermi surface we consider the limit $T \rightarrow 0$. For half-filling $n_e = 1$ so that $\delta = 0$ and the lower band must be completely filled – as one would expect for an insulator. Neglecting the term $\propto U^{-2}$ we find that for $\delta > 0$ the fraction of the Brillouin zone where the lower Hubbard band is occupied is $1 - 2\delta$, or, put another way, there are hole pockets around (π, π) whose volume is a fraction 2δ of the Brillouin zone. As a quantitative example, Figure 6 shows the spectral density obtained from the Green function (32) for $U/t = 8$ and different band fillings, whereas Figure 7 shows the resulting Fermi surfaces and the dependence of the Fermi surface volume on electron density. In Fig. 6 one can recognize the two Hubbard bands separated by an appreciable energy gap. For $n_e = 0.9$ – i.e. close to half-filling – the Fermi energy intersects the lower Hubbard band close to (π, π) resulting in the hole pocket around

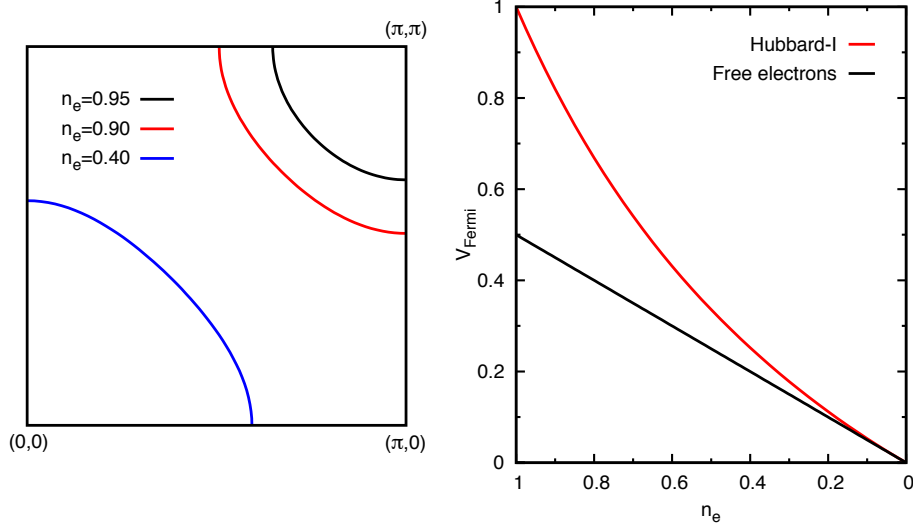


Fig. 7: Left: Fermi surface for different electron densities. Right: Fermi surface volume as a function of electron density n_e .

$X = (\pi, \pi)$ discussed above. An interesting feature seen in Fig. 6 is the transfer of spectral weight from the upper to the lower Hubbard band upon hole doping: as the electron density n_e decreases the upper Hubbard band persists but loses weight, whereas the lower Hubbard band becomes more intense. To understand this we note first that for $n_e \leq 1$ the upper Hubbard band always belongs to the inverse photoemission or electron addition spectrum. Also, we have seen in the simplified derivation that the upper band mainly has double-occupancy character.

As electrons are removed from the system, however, the probability that an added electron is placed at an occupied site to create a double occupancy becomes smaller and consequently the weight of the upper band diminishes. This doping dependent intensity of what would be the conduction band in an ordinary semiconductor or insulator is one of the fingerprints of strong correlations and can be observed experimentally in cuprate superconductors – an example is shown in Fig. 6 [15]. It should be noted, however, that the Hubbard-I approximation considerably underestimates the decrease of the intensity of the upper Hubbard band with doping.

Figure 7 also shows the dependence of the Fermi surface volume V_{Fermi} on electron density n_e . More precisely, this is the fraction of the Brillouin zone where the lower Hubbard band is below E_F , i.e. ‘occupied’. Also shown is V_{Fermi} for free electrons – where $V_{\text{Fermi}} = n_e/2$. The Hubbard-I approximation gives $V_{\text{Fermi}} \rightarrow 1$ – that means a completely filled band – as $n_e \rightarrow 1$, predicts $V_{\text{Fermi}} = 1 - 2\delta = 2n_e - 1$ for n_e slightly less than 1 and approaches the free electron behavior for small n_e . This leads to a peculiar nonlinear dependence $V_{\text{Fermi}}(n_e)$ which most probably is unphysical.

Let us now compare the Hubbard-I approximation to numerical simulations. As we saw in our simplified derivation, an important assumption of the Hubbard-I approximation is the ‘disordered’ ground state. This is best realized at high temperatures, more precisely at a temperature much higher than the characteristic energy of spin excitations, $J = 4t^2/U$. Figure 8 shows the result of a Quantum Monte-Carlo calculation of the spectral density for an 8×8 cluster at the

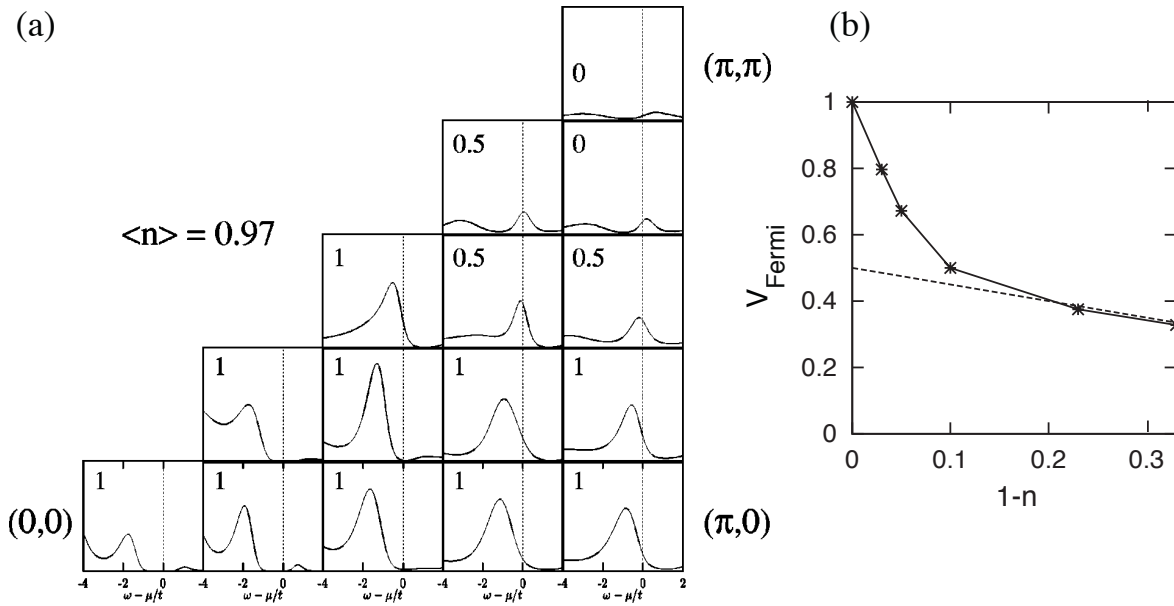


Fig. 8: Right: Single particle spectral function $A(\mathbf{k}, \omega)$ obtained by Quantum Monte-Carlo simulations on an 8×8 cluster at $k_B T = t$. Left: Fermi surface volume (34) deduced from $A(\mathbf{k}, \omega)$ versus electron density. The dashed line corresponds to free electrons. Reprinted with permission from [16], Copyright 2000 by the American Physical society.

rather high temperature $k_B T = t$. The 8×8 cluster has the allowed momenta $(\frac{n\pi}{4}, \frac{m\pi}{4})$ with integer m and n and Fig. 8 shows the part of the spectral density near the chemical potential μ for all allowed momenta in the irreducible wedge of the Brillouin zone for electron densities close to $n_e = 1$. Close to (π, π) a relatively well-defined peak passes through μ as (π, π) is approached and forms a relatively small hole pocket around (π, π) – similar to the prediction of the Hubbard-I approximation in Fig. 6 for $n_e = 0.9$. To study V_{Fermi} an ‘occupation number’ $n_{\mathbf{k}}$ of 1, 0.5 or 0 was assigned to each momentum \mathbf{k} , depending of whether the dispersive peak is below, more or less on, or above the chemical potential at \mathbf{k} . The fractional Fermi surface volume then is

$$V_{\text{Fermi}} = \frac{1}{64} \sum_{\mathbf{k}} n_{\mathbf{k}}, \quad (34)$$

where 64 is the number of momenta in the 8×8 cluster. The obtained estimate for V_{Fermi} is also shown in Figure 8 as a function of electron density and indeed has a rough similarity to the result for the Hubbard-I approximation.

6 The Gutzwiller wave function

This is the second ‘classic’ approximation for the Hubbard model. It starts from the Fermi sea $|FS\rangle$ – i.e. the ground state for $U = 0$ – and reduces the number of double occupancies by acting with a suitable projecting operator. More precisely the Gutzwiller wave function reads [3]

$$|\Phi_G\rangle = \prod_{i=1}^N (1 - \lambda n_{i,\uparrow} n_{i,\downarrow}) |FS\rangle,$$

where λ is a variational parameter to be determined by minimizing the energy E_G . To see what this means, let us go back to the Hubbard dimer. There, the ground state for $U = 0$ is

$$\langle FS \rangle' = c_{+, \uparrow}^\dagger c_{+, \downarrow}^\dagger |0\rangle = \frac{1}{2} \left[\left(c_{1, \uparrow}^\dagger c_{2, \downarrow}^\dagger + c_{2, \uparrow}^\dagger c_{1, \downarrow}^\dagger \right) + \left(c_{1, \uparrow}^\dagger c_{1, \downarrow}^\dagger + c_{2, \uparrow}^\dagger c_{2, \downarrow}^\dagger \right) \right] |0\rangle.$$

The first term in the square bracket does not contain any double occupancy, so that this term is annihilated by acting with $n_{i, \uparrow} n_{i, \downarrow}$ – this term therefore is simply reproduced by the Gutzwiller projector. On the other hand, each contribution in the second term has one double occupancy and thus gets a factor of $1 \cdot (1 - \lambda)$. The Gutzwiller wave function therefore is

$$|\Phi_G\rangle = \frac{1}{\sqrt{2}} |1_+\rangle + \frac{1 - \lambda}{\sqrt{2}} |2_+\rangle,$$

with $|1_+\rangle$ and $|2_+\rangle$ as defined in (14). Therefore, if we choose (with u, v in (15))

$$1 - \lambda = \frac{v}{u} \quad \Rightarrow \quad \lambda = 1 - \frac{2t}{W + \frac{U}{2}},$$

the Gutzwiller wave function is – up to a normalization factor – identical to the exact ground state wave function of the dimer! From the above it also becomes apparent that the Gutzwiller projector reduces the weight of states containing double occupancies.

We proceed to the infinite lattice and first rewrite the Fermi sea as a superposition of real space configurations. Suppressing the spin index (and denoting permutations by σ) we have

$$\begin{aligned} \prod_{j=1}^M c_{\mathbf{k}_j}^\dagger |0\rangle &= \frac{1}{\sqrt{N^M}} \sum_{i_1, i_2, i_3, \dots, i_M} \exp \left(i \sum_{j=1}^M \mathbf{k}_j \cdot \mathbf{R}_{i_j} \right) \prod_{j=1}^M c_{i_j}^\dagger |0\rangle \\ &= \frac{1}{\sqrt{N^M}} \sum_{i_1 > i_2 > i_3 \dots > i_M} \sum_{\sigma} \exp \left(i \sum_{j=1}^M \mathbf{k}_j \cdot \mathbf{R}_{i_{\sigma(j)}} \right) \prod_{j=1}^M c_{i_{\sigma(j)}}^\dagger |0\rangle \end{aligned}$$

In the second line we used the fact that instead of summing over all M -tuples of indices we may as well sum only over ordered M -tuples of indices and then sum over all $M!$ permutations σ of the M indices.

Next, in each of the products $\prod_{j=1}^M c_{i_{\sigma(j)}}^\dagger$ we permute the c_i^\dagger operators back to the ordered sequence $c_{i_1}^\dagger c_{i_2}^\dagger \dots c_{i_M}^\dagger$. The permutation which brings $\sigma(i) \rightarrow i$ obviously is σ^{-1} and since the Fermi sign of σ^{-1} is equal to that of σ we obtain

$$\begin{aligned} &\frac{1}{\sqrt{N^M}} \sum_{i_1 > i_2 > i_3 \dots > i_M} \sum_{\sigma} (-1)^\sigma \exp \left(i \sum_{j=1}^M \mathbf{k}_j \cdot \mathbf{R}_{i_{\sigma(j)}} \right) c_{i_1}^\dagger c_{i_2}^\dagger \dots c_{i_M}^\dagger |0\rangle \\ &= \frac{1}{\sqrt{N^M}} \sum_{i_1 > i_2 > i_3 \dots > i_M} D(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M | i_1, i_2, \dots, i_M) c_{i_1}^\dagger c_{i_2}^\dagger \dots c_{i_M}^\dagger |0\rangle, \end{aligned}$$

where the second line is the definition of the symbol $D(\mathbf{k}_j | i_j)$. From the above we see that the Fermi sea may be thought of as a superposition of real space configurations

$$c_{i_1, \uparrow}^\dagger c_{i_2, \uparrow}^\dagger c_{i_3, \uparrow}^\dagger \dots c_{i_{N_\uparrow}, \uparrow}^\dagger c_{j_1, \downarrow}^\dagger c_{j_2, \downarrow}^\dagger c_{j_3, \downarrow}^\dagger \dots c_{j_{N_\downarrow}, \downarrow}^\dagger |0\rangle$$

which are multiplied by two determinants D , one for each spin direction. Each of these real space configurations has a certain number N_d of doubly occupied sites and therefore gets an additional factor of $(1 - \lambda)^{N_d} < 1$ in the Gutzwiller wave function so that states with a larger number of double occupancies have a smaller weight as compared to the original Fermi sea. Next, the Gutzwiller function can be decomposed into components with fixed N_d :

$$|\Phi_G\rangle = \sum_{N_d} |\Phi(N_d)\rangle$$

where $|\Phi(N_d)\rangle$ is the sum over all real-space configurations with N_d double occupancies, each multiplied by its proper prefactor. The total norm $\langle\Phi_G|\Phi_G\rangle$ can be rewritten as the sum over N_d terms of the form $W(N_d) = \langle\Phi(N_d)|\Phi(N_d)\rangle$ and we now consider which N_d gives the largest contribution in this sum. To compute norms, we need to evaluate expressions such as

$$\begin{aligned} D^*(\mathbf{k}_j|i_j) D(\mathbf{k}_j|i_j) &= \sum_{\sigma, \sigma'} (-1)^\sigma (-1)^{\sigma'} \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot (\mathbf{R}_{i_{\sigma(j)}} - \mathbf{R}_{i_{\sigma'(j)}})\right) \\ &= M! + \sum_{\sigma \neq \sigma'} (-1)^\sigma (-1)^{\sigma'} \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot (\mathbf{R}_{i_{\sigma(j)}} - \mathbf{R}_{i_{\sigma'(j)}})\right). \end{aligned} \quad (35)$$

where in the first term we have collected the $M!$ terms with $\sigma = \sigma'$. At this point, we make an important approximation: (35) still has to be summed over $i_1, i_2, i_3 \dots i_M$. The terms for $\sigma \neq \sigma'$ thereby have a rapidly oscillating phase and a large degree of cancellation will occur in the summation. Accordingly we retain only the first term, i.e., we replace

$$D^*(\mathbf{k}_j|i_j) D(\mathbf{k}_j|i_j) \approx M!.$$

With this approximation the contribution of the states with N_d double occupancies becomes

$$W(N_d) = \frac{N_\uparrow! N_\downarrow!}{N^{N_\uparrow + N_\downarrow}} (1 - \lambda)^{2N_d} C(N_\uparrow, N_\downarrow, N_d),$$

where $C(N_\uparrow, N_\downarrow, N_d)$ is the number of ways in which N_\uparrow electrons with spin \uparrow and N_\downarrow electrons with spin \downarrow can be distributed over the N lattice sites such as to generate N_d double occupancies. This is a straightforward combinatorial problem with the result

$$C(N_\uparrow, N_\downarrow, N_d) = \frac{N!}{N_d! (N_\uparrow - N_d)! (N_\downarrow - N_d)! (N - N_\uparrow - N_\downarrow + N_d)!}.$$

Next, we take the logarithm of $W(N_d)$, use the Stirling formula $\log(N!) \approx N \log(N) - N$ and differentiate with respect to N_d . Introducing the densities $n_d = N_d/N$ etc. we obtain

$$\begin{aligned} \frac{d}{dN_d} \log(W(N_d)) &= \log\left((1 - \lambda)^2 \frac{(n_\uparrow - n_d)(n_\downarrow - n_d)}{n_d(1 - n_\uparrow - n_\downarrow + n_d)}\right), \\ \frac{d^2}{dN_d^2} \log(W(N_d)) &= -\frac{1}{N} \left(\frac{1}{n_d} + \frac{1}{n_\uparrow - n_d} + \frac{1}{n_\downarrow - n_d} + \frac{1}{1 - n_\uparrow - n_\downarrow + n_d}\right) = -\frac{c}{N}, \end{aligned}$$

where $c > 0$ in the last line is of order unity. The first of these equations gives us the n_d where the contribution to the norm, $W(N_d)$ is a maximum. For the general case the formula is somewhat involved, so we specialize to the case $n_\uparrow = n_\downarrow = \frac{1}{2}$ where

$$n_{d,0} = \frac{1 - \lambda}{2(2 - \lambda)}. \quad (36)$$

For the noninteracting case $\lambda \rightarrow 0$ this gives $n_{d,0} = 1/4$ as it has to be. From the second equation we find

$$\begin{aligned} \log(W(N_d)) &= \log(W(N_{d,0})) - \frac{c}{2N} (N - N_{d,0})^2 + \dots \\ W(N_d) &= W(N_{d,0}) \exp\left(-\frac{c}{2N}(N_d - N_{d,0})^2\right) = W(N_{d,0}) \exp\left(-\frac{Nc}{2}(n_d - n_{d,0})^2\right), \end{aligned}$$

which shows that as a function of n_d the weight $W(N_d)$ is a Gaussian of width $\propto N^{-1/2}$. This means, however, that in the thermodynamical limit only states with $n_d = n_{d,0}$ have an appreciable weight in the Gutzwiller wave function and variation of λ simply shifts this sharp peak of $W(N_d)$ to a different $n_{d,0}$. An immediate consequence is that the computation of the expectation value of the interaction Hamiltonian becomes trivial, namely $\langle H_U \rangle = N U n_{d,0}$.

The expectation value of the kinetic energy and is more involved. The above discussion showed that the Gutzwiller wave function is composed of real-space configurations for which the number of double occupancies is close to a certain value $N_{d,0}$ which is smaller than for the noninteracting Fermi sea. This means, however, that the expectation value of the kinetic energy is smaller as well. Namely using again the operators \hat{d} and \hat{c} we have

$$c_{i,\sigma}^\dagger c_{j,\sigma} = \hat{d}_{i,\sigma}^\dagger \hat{d}_{j,\sigma} + \hat{c}_{i,\sigma}^\dagger \hat{d}_{j,\sigma} + \hat{d}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} + \hat{c}_{i,\sigma}^\dagger \hat{c}_{j,\sigma}.$$

If the number of double occupancies is decreased, the expectation value of the first term on the r.h.s. clearly must decrease. Second, since the number of electrons is constant, reducing the number of double occupancies necessarily results in a reduction of the number of empty sites by the same number so that the expectation value of the last term on the r.h.s. also must decrease. The Gutzwiller approximation assumes, that these effects can be taken into account by reducing the expectation value of the kinetic energy of the uncorrelated Fermi sea by suitable renormalization factors η :

$$\frac{\langle \Phi_G | H_t | \Phi_G \rangle}{\langle \Phi_G | \Phi_G \rangle} = \sum_{\sigma} \eta_{\sigma} \langle FS, \sigma | H_t | FS, \sigma \rangle$$

where $|FS, \sigma\rangle$ is the Fermi sea for σ -electrons (if $N_\uparrow = N_\downarrow$ the two terms are of course identical). These renormalization factors η_{σ} thereby are evaluated for an ‘auxiliary wave function’ in which the determinants $D(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M | i_1, i_2, \dots, i_M)$ are replaced by a constant (which would have to be $\sqrt{M!}$ if the auxiliary wave function is supposed to have the same norm as the Gutzwiller wave function) and where the Fermi sign is ignored in the calculation of all matrix elements of the hopping term (this is because the Fermi sign is supposed to be taken care of already by the filling of the uncorrelated Fermi sea according to the Pauli principle!). The

evaluation of the η by combinatorical considerations is discussed in a very transparent way by Ogawa, Kanda, and Matsubara [17]. Here we use an even simpler way of calculating η by introducing four ‘book-keeping kets’ for every site i : $|i, 0\rangle$, $|i, \uparrow\rangle$, $|i, \downarrow\rangle$ and $|i, \uparrow\downarrow\rangle$. They represent in an obvious way the four possible configurations of the site i . Then we define

$$B_i = \frac{|i, 0\rangle + \alpha_\uparrow|i, \uparrow\rangle + \alpha_\downarrow|i, \downarrow\rangle + \beta|i, \uparrow\downarrow\rangle}{\sqrt{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2}} \quad \text{and} \quad |\Psi\rangle = \prod_i B_i |0\rangle$$

with real α_σ and β . The state $|\Psi\rangle$ has norm 1 and if it were translated into a true state of electrons, the numbers of electrons and double occupancies would be

$$\begin{aligned} \langle N_\sigma \rangle &= N \frac{\alpha_\sigma^2 + \beta^2}{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2}, \\ \langle N_d \rangle &= N \frac{\beta^2}{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2}. \end{aligned} \quad (37)$$

These equations can be reverted to give

$$\alpha_\sigma = \sqrt{\frac{n_\sigma - n_d}{1 - n_\uparrow - n_\downarrow + n_d}} \quad \text{and} \quad \beta = \sqrt{\frac{n_d}{1 - n_\uparrow - n_\downarrow + n_d}}. \quad (38)$$

On the other hand $|\Psi\rangle$ does not correspond to a state with a fixed number of electrons so we introduce

$$|\Psi'\rangle = \mathcal{P}(N_\uparrow, N_\downarrow, N_d) |\Psi\rangle,$$

where the projection operator \mathcal{P} projects onto the component of $|\Psi\rangle$ which has precisely $\langle N_\uparrow \rangle$ \uparrow -electrons etc. Next, the representation of the electron annihilation operator $c_{i,\sigma}$ is

$$\tilde{c}_{i,\sigma} = |i, 0\rangle \langle i, \sigma| + |i, -\sigma\rangle \langle i, \uparrow\downarrow|.$$

Here a subtle detail should be noted: in the expression on the r.h.s. it is assumed that a double occupancy always is converted into the state $|i, -\sigma\rangle$ with a positive sign. This would not be the case for the true Fermion operator, where the sign would depend on the sequence of the two electron creation operators on the doubly occupied site. This is precisely the neglect of the Fermi sign that was mentioned above. Then, to estimate the reduction of the kinetic energy due to the reduction of the number of doubly occupied and empty sites we evaluate

$$r(\sigma, n_\uparrow, n_\downarrow, n_d) = \frac{\langle \Psi' | \tilde{c}_{i,\sigma}^\dagger \tilde{c}_{j,\sigma} | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle}. \quad (39)$$

So far our auxiliary wave function has not brought about much simplification because the presence of the projection operator \mathcal{P} makes the computation of r very tedious. It is straightforward to see, however, that if $|\Psi\rangle$ is decomposed into components of fixed N_\uparrow , N_\downarrow and N_d only those components with values of N_\uparrow , N_\downarrow and N_d which deviate by at most $N^{-1/2}$ from the average values (37) have an appreciable weight. This means, however, that \mathcal{P} simply can be dropped so

that we replace $|\Psi'\rangle \rightarrow |\Psi\rangle$ in (39). Then, since $|\Psi\rangle$ is normalized the denominator in (39) can be dropped. Moreover, since $|\Psi\rangle$ is a product state the expectation value of the two operators factorizes and since all sites are equivalent and the coefficients α_σ and β are real the expectation values of $\tilde{c}_{i,\sigma}^\dagger$ and $\tilde{c}_{j,\sigma}$ are identical so that

$$\begin{aligned} r(\sigma, n_\uparrow, n_\downarrow, n_d) &= \langle \Psi | \tilde{c}_{i,\sigma}^\dagger | \Psi \rangle^2 = \left(\frac{\alpha_\sigma + \alpha_{-\sigma}\beta}{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2} \right)^2 \\ &= \left(\frac{\sqrt{n_\sigma - n_d} \sqrt{1 - n_\uparrow - n_\downarrow + n_d} + \sqrt{n_d} \sqrt{n_{-\sigma} - n_d}}{\sqrt{n_\sigma(1 - n_\sigma)}} \right)^2, \end{aligned}$$

where the second line has been obtained by inserting (38). In this way we have expressed $r(n_\sigma, n_d)$ in terms of n_d which in turn is given as a function of λ by (36). Lastly, we divide r by its value for $U \rightarrow 0$, where $n_d = n_\uparrow \cdot n_\downarrow$, so as to obtain the proper limiting value of $\eta = 1$ for $U = 0$, and finally get

$$\eta(\sigma, n_\uparrow, n_\downarrow, n_d) = \left(\frac{\sqrt{n_\sigma - n_d} \sqrt{1 - n_\uparrow - n_\downarrow + n_d} + \sqrt{n_d} \sqrt{n_{-\sigma} - n_d}}{\sqrt{n_\sigma(1 - n_\sigma)}} \right)^2. \quad (40)$$

In varying the energy it is actually easier to switch from λ to n_d as variational parameter. Specializing to the paramagnetic case $n_\uparrow = n_\downarrow$ the energy per site thus becomes

$$e_G = \eta(n_\sigma, n_d) t_0 + n_d U, \quad (41)$$

where $e_G = E_G/N$ and t_0 is the (kinetic) energy of the Fermi sea per site. Using (40) this is now readily minimized with respect to n_d .

The Gutzwiller wave function gives us, strictly speaking, only the ground state energies and some ground state expectation values, but not a band structure. However, we may consider states like

$$|\Phi_G(\mathbf{k})\rangle = \prod_i (1 - \lambda n_{i,\uparrow} n_{i,\downarrow}) c_{\mathbf{k},\uparrow} |FS\rangle,$$

i.e., a state with one hole in the Fermi sea (it is understood that \mathbf{k} is an occupied momentum). The Fermi sea with a hole has energy $E_{FS} - \varepsilon_{\mathbf{k}}$. It thus might seem plausible that the energy of $|\Phi_G(\mathbf{k})\rangle$ is $E_G - \tilde{\varepsilon}_{\mathbf{k}}$ i.e., the energy of the Gutzwiller wave function minus the ‘quasiparticle energy’. Performing the variational procedure for $|\Phi_G(\mathbf{k})\rangle$ amounts to replacing $e_G \rightarrow e_G - \frac{1}{N} \tilde{\varepsilon}_{\mathbf{k}}$, $t_0 \rightarrow t_0 - \frac{1}{N} \varepsilon_{\mathbf{k}}$, $n_\uparrow \rightarrow n_\uparrow - \frac{1}{N}$, and $n_d \rightarrow n_d + \frac{1}{N} \delta n_d$ where δn_d is the as yet unknown shift of n_d . Inserting into (41) and expanding we find

$$e - \frac{1}{N} \tilde{\varepsilon}_{\mathbf{k}} = \left(\eta - \frac{1}{N} \frac{\partial \eta}{\partial n_\uparrow} + \frac{1}{N} \frac{\partial \eta}{\partial n_d} \delta n_d \right) \left(t_0 - \frac{1}{N} \varepsilon_{\mathbf{k}} \right) + n_d U + \frac{1}{N} \delta n_d U.$$

The terms of zeroth order in $1/N$ cancel due to (41) and collecting the first order terms gives

$$\tilde{\varepsilon}_{\mathbf{k}} = \eta \varepsilon_{\mathbf{k}} + t_0 \frac{\partial \eta}{\partial n_\uparrow} - \left(\frac{\partial \eta}{\partial n_d} t_0 + U \right) \delta n_d.$$

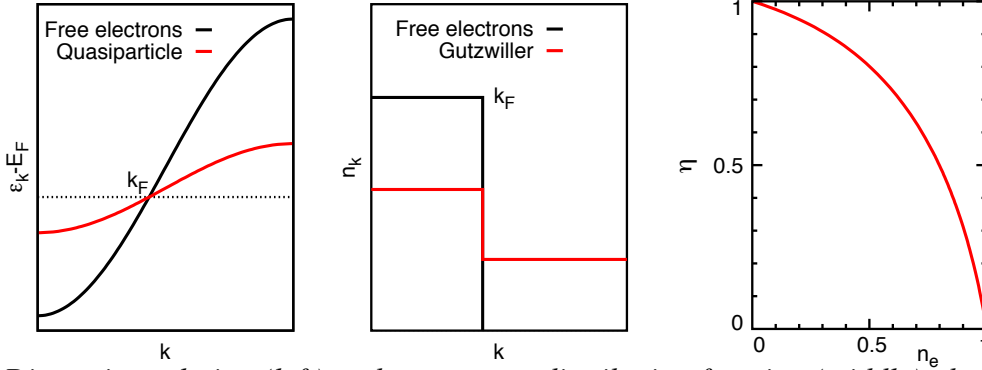


Fig. 9: Dispersion relation (left) and momentum distribution function (middle) obtained from the Gutzwiller wave function compared to the free electron case. The right part shows the dependence $\eta(n_e)$ for the half-filled two-dimensional Hubbard model with $U/t = 16$.

The last term on the r.h.s. vanishes because the expression in the bracket is $\frac{de_G}{dn_d}$. The second term on the r.h.s. gives the change of the kinetic energy of all other electrons due to removal of the single electron. This is a k -independent shift which can be absorbed into a shift of E_F . The quasiparticle dispersion $\tilde{\varepsilon}_k$ therefore follows the original dispersion, but renormalized by the factor $\eta < 1$. This is an effect known as ‘correlation narrowing’.

Next we consider the ground state momentum distribution function, i.e., the ground state expectation value $n_{\mathbf{k}} = 2\langle c_{\mathbf{k},\uparrow}^\dagger c_{\mathbf{k},\uparrow} \rangle$. This can be obtained from the ground state energy by means of the Hellmann-Feynman theorem: $n_{\mathbf{k}} = \frac{\partial E_G}{\partial \varepsilon_{\mathbf{k}}}$. More generally, it is the functional derivative of the ground state energy with respect to $\varepsilon_{\mathbf{k}}$, that means under a change $t_{ij} \rightarrow t_{ij} + \delta t_{ij}$ so that $\varepsilon_{\mathbf{k}} \rightarrow \varepsilon_{\mathbf{k}} + \delta \varepsilon_{\mathbf{k}}$ the change of the ground state energy is

$$e_G \rightarrow e_G + 2 \sum_{\mathbf{k}} n_{\mathbf{k}} \delta \varepsilon_{\mathbf{k}}.$$

From (41) we obtain the variation of e_G as

$$\delta e_G = 2\eta \sum_{\mathbf{k}} n_{\mathbf{k}}^{(0)} \delta \varepsilon_{\mathbf{k}} + \delta n_d \left(\frac{\partial \eta}{\partial n_d} t_0 + U \right).$$

where $n_{\mathbf{k}}^{(0)} = \Theta(E_F - \varepsilon_{\mathbf{k}})$ is the momentum distribution of the Fermi sea. Again, the second term on the r.h.s. vanishes due to the extremum condition for n_d so that $n_{\mathbf{k}} = \eta n_{\mathbf{k}}^{(0)}$. This cannot be entirely correct, however, because we have the sum-rule $2 \sum_{\mathbf{k}} n_{\mathbf{k}} = N_e$ and since this is fulfilled by $n_{\mathbf{k}}^{(0)}$ and $\eta < 1$ it cannot be fulfilled for $n_{\mathbf{k}}$. The solution is that the ‘missing $n_{\mathbf{k}}$ ’ takes the form of a k -independent additive constant which then has to be $(1 - \eta) \frac{n_e}{2}$. In fact, for any $\varepsilon_{\mathbf{k}}$ which can be represented by hopping integrals $t_{i,j}$ one has $\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} = 0$ so that such a k -independent additive constant would not contribute to the variation of e_G . The momentum distribution obtained by the Gutzwiller approximation thus has a step of magnitude η at the position of the original Fermi surface. Let us now consider in some more detail the case $n_\sigma = \frac{1}{2}$ where the Mott-insulator should be realized for large U/t . We find from (40)

$$\eta(n_d) = 16n_d \left(\frac{1}{2} - n_d \right).$$

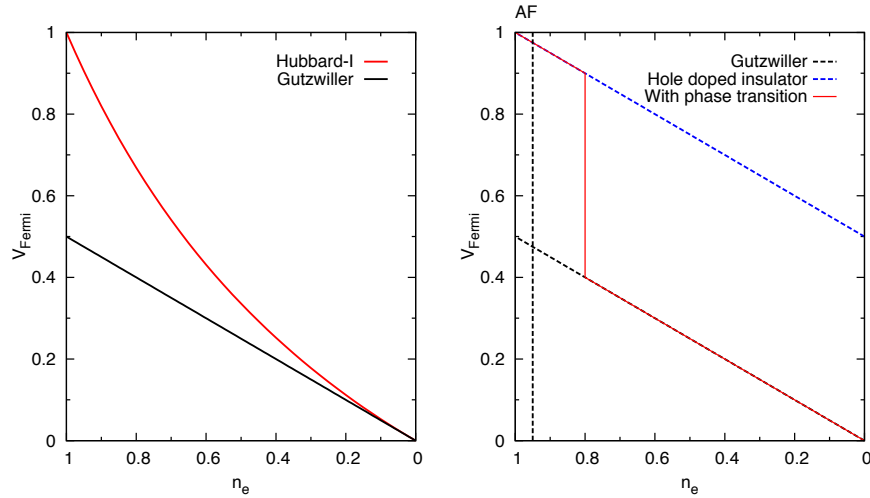


Fig. 10: *Left: Comparison of the Fermi surface volume as obtained from the Hubbard-I approximation and the Gutzwiller wave function. Right: a possible compromise with a phase transition between two phases with different $V_{\text{Fermi}}(n_e)$.*

Minimizing (41) this gives

$$n_d = \frac{1}{4} - \frac{U}{32|t_0|},$$

whereby we have taken into account that $t_0 < 0$ for a half-filled band. Starting from the noninteracting value $\frac{1}{4}$, n_d decreases linearly with U and reaches zero at the critical value $U_c = 8|t_0|$. For $n_d = 0$ we have $\eta = 0$, so that the bandwidth of the quasiparticles becomes zero, i.e., the band mass diverges, and the step in the momentum distribution vanishes as well. This is commonly interpreted as a metal-to-insulator transition as a function of increasing U , the so-called Brinkman-Rice transition [18]. Brinkman and Rice also could show that the magnetic susceptibility diverges at the transition as one would expect it for a diverging effective mass.

Let us now consider the two-dimensional model with nearest neighbor hopping $-t$. Then, $t_0 = -1.621 t$ so that the critical $U_c = 12.969 t$. Figure 9 then shows the dependence of η on n_e for $U/t = 16$ i.e., for $U > U_c$. As $n_e \rightarrow 1$ the renormalization factor $\eta \rightarrow 0$ so that both, the bandwidth and the step in $n_{\mathbf{k}}$ vanish for the half-filled band. The Hubbard-I approximation and the Gutzwiller wave function thus give completely different predictions about what happens when the half-filled band case is approached by increasing the electron density for constant U/t : whereas the Hubbard-I approximation predicts the lower Hubbard band with (almost) constant bandwidth and a hole-pocket-like Fermi surface with a volume $\propto (1 - n_e)$ so that the Fermi surface vanishes at $n_e \rightarrow 1$, the Gutzwiller wave function predicts a Fermi surface with a volume equal to that obtained for free electrons, but with a vanishing bandwidth and spectral weight as $n_e \rightarrow 1$.

7 Summary and discussion

Since a generally accepted theory of the lightly doped 2-dimensional Hubbard model does not seem to exist so far so that also the cuprate superconductors are not really understood as yet, maybe the best one can do at present is to outline the problems that would have to be resolved.

The first one of these is the Fermi surface close to half-filling. As we have seen the Hubbard-I approximation and the Gutzwiller wave function predict completely different behavior close to half-filling: a hole-like Fermi surface with a volume $\propto n_h = 1 - n_e$ in the lower Hubbard-band whose volume tends to zero as $n_e \rightarrow 1$ versus a free-electron-like Fermi surface with volume $\frac{n_e}{2}$ formed by a band whose mass diverges as $n_e \rightarrow 1$. A possible compromise between the two approximations could be as shown in Figure 10: near $n_e = 1$ there are hole pockets with a volume that is strictly proportional to the hole number $n_h = 1 - n_e$ – i.e. the doped Hubbard-band – and then at some critical density a phase transition occurs to a phase where the Fermi surface volume is $n_e/2$. This might be the true behavior which Hubbard-I approximation ‘tries to reproduce’. Viewed this way, the ‘pseudogap phase’ of cuprate superconductors actually should be identified with the hole-pocket phase and the quantum critical point which is surrounded by the superconducting dome corresponds to the transition to the free-electron-like Fermi surface. A theory which is supposed to describe this, first of all, must reproduce the two Hubbard bands – otherwise the hole-doped lower Hubbard band cannot be captured. Next, the two different phases would have to be reproduced which is a considerable problem because there is no obvious order parameter for the transition between a paramagnetic small Fermi surface and a paramagnetic large Fermi surface.

References

- [1] J. Hubbard, Proc. Roy. Soc. London A **276** 238 (1963)
- [2] A. Kanamori, Progr. Theor. Phys. **30**, 275 (1963)
- [3] M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)
- [4] F.C. Zhang and T.M. Rice, Phys. Rev. B **37**, 3759 (1988)
- [5] A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinski:
Methods of Quantum Field Theory in Statistical Physics (Prentice-Hall, New Jersey 1964)
- [6] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems*
(McGraw-Hill, San Francisco, 1971)
- [7] J.W. Negele and H. Orland: *Quantum Many-Particle Systems*
(Addison-Wesley, Redwood, 1988)
- [8] G. Baym and N.D. Mermin, J. Math. Phys. **2**,232 (1961)
- [9] J.M. Luttinger, Phys. Rev. **121**, 942 (1961)
- [10] R. Eder, K. Seki, and Y. Ohta Phys. Rev. B **83**, 205137 (2011)
- [11] L.P. Gorkov, Soviet Phys. JETP **7** 505 (1958)
- [12] L.D. Landau and E.M. Lifshitz: *Lehrbuch der Theoretischen Physik*
(Akademie Verlag Berlin, 1988)
- [13] A. Damascelli, Z. Hussain, and Z.-X. Shen, Rev. Mod. Phys. **75** 473 (2003)
- [14] J.E. Hirsch Phys. Rev. B **31**, 4403 (1985); H.Q. Lin and J.E. Hirsch,
Phys. Rev. B **35**, 3359 (1987)
- [15] C.T. Chen, F. Sette, Y. Ma, M.S. Hybertsen, E.B. Stechel, W.M.C. Foulkes, M. Schuler,
S-W. Cheong, A.S. Cooper, L.W. Rupp, B. Batlogg, Y.L. Soo, Z.H. Ming, A. Krol, and
Y.H. Kao, Phys. Rev. Lett. **66**, **104** (1991)
- [16] C. Gröber, R. Eder, and W. Hanke, Phys. Rev. B **62** 4336 (2000)
- [17] T. Ogawa, K. Kanda, and T. Matsubara, Progr. Theor. Phys. **53**, 614 (1975)
- [18] W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 4302 (1970)

