# Comment on Hubbard's Theory of the Mott Transition*

D. M. EDWARDS, A. C. HEWSON
*Department of Mathematics, Imperial College, London, S.W. 7, England*

It is shown that no well-defined Fermi surface exists on the "metallic" side of the Mott transition in Hubbard's theory. An alternative approach is suggested.

## 1. INTRODUCTION

In a series of papers, Hubbard has considered the problem of electron correlation in narrow energy bands. In papers I[1] and III[2] he deals with a one-band model defined by the Hamiltonian

$$H=\sum_{\mathbf{k},\sigma}\epsilon_{\mathbf{k}}c_{\mathbf{k}\sigma}{}^{+}c_{\mathbf{k}\sigma}+(I/N)\sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{q}}c_{\mathbf{k}_1+\mathbf{q}\uparrow}{}^{+}c_{\mathbf{k}_1\uparrow}c_{\mathbf{k}_2-\mathbf{q}\downarrow}{}^{+}c_{\mathbf{k}_2\downarrow}, \tag{1}$$

where $I$ is the interaction energy of two electrons on the same Wannier function and $N$ is the number of atoms in the crystal. Hubbard investigated the pseudoparticle density of states by calculating the one-particle Green function $G_{\mathbf{k}\sigma}(E)$, using decoupling techniques described by Zubarev.[3] The guiding principle in Hubbard's choice of decoupling procedures is that the theory should be exactly correct in the atomic limit as $\Delta\to 0$, where $\Delta$ is the width of the band of one-electron energies $\epsilon_{\mathbf{k}}$. In paper I, where the decoupling used is the simplest one consistent with this principle, it is found that the pseudoparticle density of states splits into two subbands for any finite interaction $I$. In a case where the unperturbed density of states $p(\epsilon)$ is symmetric about the center of the band, each subband contains exactly $N$ states. Thus, if the number of electrons present is one per atom, the lower subband will be completely full in the ground state and the system will be an insulator for any finite $I$. This is clearly incorrect, and in paper III Hubbard refined the theory so that the splitting of the pseudoparticle density of states into subbands occurs only for $I/\Delta$ greater than a critical value. This splitting then appears to correspond to a metal–nonmetal transition of the type discussed by Mott.[4] A characteristic property of a metal is the existence of a Fermi surface and it is interesting to investigate whether a Fermi surface exists on the metallic side of the transition in Hubbard's theory. In this paper it is shown, by considering a particular band structure, that no well-defined Fermi surface occurs. An alternative approach, which might overcome this difficulty is briefly discussed.

## 2. NONEXISTENCE OF A FERMI SURFACE

A well-defined Fermi surface is characterised by a discontinuity in the occupation number

$$\bar{n}_{\mathbf{k}\sigma}=\langle c_{\mathbf{k}\sigma}{}^{+}c_{\mathbf{k}\sigma}\rangle, \tag{2}$$

the average being calculated for the ground state. The most favorable case for a sharp Fermi surface is in the limit of small $I/\Delta$, and we calculate $\bar{n}_{\mathbf{k}\sigma}$ in this limit for a band structure $\epsilon_{\mathbf{k}}$ corresponding to a density of states

$$p(\epsilon)=(4/\pi\Delta)[1-(2\epsilon/\Delta)^2]^{1/2}, \quad \text{if } |\epsilon|<\tfrac{1}{2}\Delta$$
$$=0, \quad \text{otherwise.} \tag{3}$$

Hubbard[2] obtains the one-particle Green function in the form

$$G_{\mathbf{k}\sigma}(E)=-\{2\pi[F(E)-\epsilon_{\mathbf{k}}]\}^{-1}, \tag{4}$$

where for the band considered, $F(E)$ satisfies a cubic equation. The relevant root of this equation is the one which leads to a positive density of pseudoparticle states, and this may be obtained explicitly to order $I^2$. The quantity required is given by

$$\bar{n}_{\mathbf{k}}=i\lim_{\delta\to 0+}\int_{-\infty}^{0}[G_{\mathbf{k}}(E+i\delta)-G_{\mathbf{k}}(E-i\delta)]dE, \tag{5}$$

where $E$ is measured from the perturbed Fermi energy, and we find

$$\bar{n}_{\mathbf{k}}=\pi^{-1}\int_{-\Delta/2}^{0}\frac{6I^2(9\Delta^2-32E^2)[(\frac{1}{2}\Delta)^2-E^2]^{1/2}dE}{[(E-\epsilon_{\mathbf{k}})(9\Delta^2-32E^2)+2I^2E]^2+36I^4[(\frac{1}{2}\Delta)^2-E^2]}. \tag{6}$$

For $I>0$ the integrand is a continuous function of $E$ and $\epsilon_{\mathbf{k}}$ for $-\Delta/2\leq E\leq 0$ and $\epsilon_{\mathbf{k}}\geq-\Delta/2$. It follows that $\bar{n}_{\mathbf{k}}$ is a continuous function of $\epsilon_{\mathbf{k}}$, and hence of $\mathbf{k}$. For small $\epsilon_{\mathbf{k}}$ and I

$$\bar{n}_{\mathbf{k}}=\tfrac{1}{2}-\pi^{-1}\tan^{-1}(3\Delta\epsilon_{\mathbf{k}}/I^2), \tag{7}$$

* Read by M. H. Cohen.

[1] J. Hubbard, Proc. Roy. Soc. (London) **A276,** 238 (1963).
[2] J. Hubbard, Proc. Roy. Soc. (London) **A281,** 401 (1964).
[3] D. N. Zubarev, Usp. Fiz. Nauk **71,** 71 (1960) [Sov. Phys.— Usp. **3,** 320 (1960)].
[4] N. F. Mott, Proc. Phys. Soc. (London) **62,** 416 (1949); Phil. Mag. **6,** 287 (1961).

so that $\bar{n}_k$ varies from 1 to 0 over a range in $\epsilon_k$ of order $I^2/3\Delta$. It is readily shown, by examining the pole of $G_{k\sigma}$, that the lifetime $\tau$ of a pseudoparticle at the Fermi level is given by $\tau^{-1}=I^2/3\Delta$, so that the form of the second term on the right of (7) is $\pi^{-1}\tan^{-1}\tau\epsilon_k$. Clearly, a discontinuity in $\bar{n}_k$ can arise only if $\tau\to\infty$ as $\epsilon_k\to 0$. Thus Hubbard's theory does not exhibit a sharp Fermi surface, even for very small $I/\Delta$, owing to a finite pseudoparticle lifetime at the Fermi level.

## 3. DISCUSSION

The reason why Hubbard's theory gives a finite pseudoparticle lifetime at the Fermi level is not hard to find. The usual argument for an infinite quasiparticle lifetime at the Fermi surface is a phase-space argument involving conservation of momentum. Hubbard's decoupling procedure, with its emphasis on the atomic limit, results in a self-energy $\Sigma(\mathbf{k}, E)=E-F(E)$ which is independent of $\mathbf{k}$; this is inconsistent with momentum conservation. This difficulty does not arise in a $t$-matrix approach based on repeated two-particle scattering. It is shown below, for a low-density case, that a splitting of the pseudoparticle spectrum can result from such an approach.

An approximate form for $\Sigma_\sigma(\mathbf{k}, E)$ at low density is

$$\Sigma_\sigma(\mathbf{k}, E)=I\sum_p n_{p,-\sigma}\times\left[1+I\sum_q\frac{(1-n_{p+q,-\sigma})(1-n_{k-q,\sigma})}{\epsilon_{p+q}+\epsilon_{k-q}-\epsilon_p-E}\right]^{-1}. \quad (8)$$

For $E=\epsilon_k$ this corresponds to Kanamori's[5] approximation. It is the $E$ dependence, however, that can lead to a splitting of the pseudoparticle spectrum. This can be seen in the limit of very low density and zero bandwidth, where $\epsilon_k=T_0$, a constant; then $\Sigma_\sigma(\mathbf{k}, E)$ reduces to $In_{-\sigma}[1-I(E-T_0)^{-1}]^{-1}$ which is identical to that derived by Hubbard in paper I in the same limit. The pole in $\Sigma_\sigma(E)$ leads to a quadratic equation $E-\epsilon_k-\Sigma_\sigma(E)=0$ for the pseudoparticle energy and a splitting of the spectrum.

## ACKNOWLEDGMENT

We wish to thank A. E. K. Dowson for a stimulating conversation.

### Discussion of Edwards' Paper

R. V. LANGE (Brandeis University): I think this paper is really quite nice and makes the point that in judging the Hubbard model one should judge the model rather than any particular solution to the model. Lots of people confuse Hubbard paper I [Proc. Roy. Soc. (London) **A276**, 238 (1961)] with the Hubbard Hamiltonian and what Edwards is saying is that we shouldn't confuse Hubbard paper III [Proc. Roy. Soc. (London) **A281**, 401 (1964)] with the Hubbard Hamiltonian even though Hubbard's solution there is much much better. I just wanted to point out that this momentum dependence in the mass operator which Edwards says must come back in is the type of term that Esterling and I found. One doesn't have to leave the idea that you'll do perturbation theory in the kinetic energy in order to get this momentum dependence in. I mean, Edwards went all the way to $T$ approximations to get it back in, but if you don't decouple, which is a terribly dangerous thing to do in the narrow-band limit, you can get that same sort of dependence, even with the spirit of Hubbard's kinetic-energy perturbation theory.

[5] J. Kanamori, Progr. Theoret. Phys. (Kyoto) **30**, 275 (1963).