Thermodynamic Analysis of Metal–Insulator Transitions. Effect of Changes in the Density of States Function*

R. Hoehn and J.M. Honig†

Department of Chemistry, Purdue University, West Lafayette IN, 47907, USA

(Received January 23, 2009)

Prior theories of metal–insulator transitions by Spałek et al. were extended to include quartic terms in the temperature and by introducing two different density of state functions. The effects of these extensions on low-temperature metal–insulator transitions and on reentrant metallic behavior in solids have been investigated.

PACS numbers: 71.10.Hf, 71.20.−b, 71.10.Ay, 71.30.+h, 71.45.Gm

1. Preliminaries

There has been continuing interest in the proper treatment of metal–insulator transitions in general, and in the special situation where one encounters reentrant metallic behavior. This is manifested, for example, in the electrical properties of paramagnetic V_2O_3 doped with small amounts of Cr. With rising temperature T above 180 K, one first observes metallic characteristics; followed by a first-order transition to an insulating state; followed at still higher temperatures by another transition back to the metallic regime [1]. Also, there are instances, such as the $NiS_{2-y}Se_y$ system, that exhibit a metal– insulator transition with rising temperature close to absolute zero [2]; the latter runs contrary to normal observations of an insulator–metal transition with rising T.

These unusual situations have been rationalized lately by invoking the dynamic mean field theoretical approach, as championed by a number of research groups [3]. Their principal aim was to generate partial density-of-states curves which match experimental photoemission results, and to note changes concomitant to the occurrence of the metal–insulator transition. While their success is certainly impressive, the theoretical analysis requires considerable computational expertise, such that the fundamentals operative in forcing the electronic changes are not readily discerned. This invites a retrospection to earlier work, beginning with the discussions by Mott [4] and by Hubbard [5] whose insights provided the impetus to subsequent treatments of electron correlation effects. This included a procedure originally proposed by Spałek and co-workers [6–8] who developed more elementary approach on which the subsequent discussion of this article is based. They considered an assembly of interacting electrons in a solid as a collection of independent quasiparticles that are subject to the framework of solid state physics. In the simplest case, one deals with nondegenerate band that is exactly half-filled, and for which — in the insulating state at the zero of temperature — each of the N lattice sites is occupied by one electron. At nonzero temperatures any given lattice site may be empty, accommodate one electron of either spin, or two electrons with paired spins. Double occupancies are energetically disfavored by the Coulomb repulsion U between the two electrons on the same site; electron interactions between more distant neighbors are ignored. Let η represent the probability of encountering a given lattice site in a doubly occupied configuration. Then the total repulsive energy is specified by $NU\eta$. Relative to this state, the mobile electrons have an average negative kinetic energy $\bar{\varepsilon}$ as they hop between available adjacent sites. However, this process is counteracted by the electronic repulsions. The resulting hindered motion is simulated by introducing a band narrowing factor $\Phi(\eta)$, specified below. The modified kinetic energy is then given by $-N|\bar{\varepsilon}|\,\Phi(\eta)$. If the repulsion effect dominates the solid tends to be an insulator; if electron motion dominates the material is a relatively poor conductor. This model thus supersedes the more primitive conventional analysis in which electron interactions are ignored*. The electrical state of a system is thus determined by the balance between the negative kinetic and positive potential energies; this balance is temperature-dependent, and — under appropriate con-

* Editorial Note: This paper has been priority-edited before the expected shortly ceremony of awarding the Honorary Doctoral Degree (Dr. H. c.) to Professor Jürgen (George) Honig on the occasion of 90-th Anniversary of founding the AGH University of Science and Technology in Kraków.

† corresponding author; e-mail: jmh@purdue.edu

* The more elementary model presents a reasonable approximation for the case of good conductors with a high density of charge carriers. In those cases, the Coulomb interaction for an electron assembly is screened and drops off almost exponentially with separation distance between electrons. However, this approximation fails precisely for the present case under study.

ditions — gives rise to one or more electrical transitions. It's verified by detailed calculations based on the theory expounded in Ref. [6–9].

In the theory which describes this effect Spałek et al. [4] did not systematically or numerically explore all the consequences of their theoretical work. As an extension of their work Datta et al. [10] treated the effects of introducing densities of state (DOS) other than the rectangular DOS considered in the earlier work. Subsequently, the original low-temperature approximation was extended [11] to include terms in T^4 in addition to those in T^2. Here we attempt to investigate the effects of introducing several other DOS functions, in the low temperature approximation that includes term of order T^4. Also, we hope to identify within the framework of this analysis the general feature that drives metal–insulator transitions.

The present paper is based on the expression for the free energy of the electron assembly in a nondegenerate band of the form [11]:

$$\frac{F_i}{N} = -|\bar{\varepsilon}|\,\Phi + U\eta - \frac{\rho_0\lambda^2}{\Phi} - \frac{\Psi}{4\Phi^3}\lambda^4, \tag{1.1}$$

in which

$$\Psi \equiv \frac{7}{5}\rho_0'' - \frac{(\rho_0')^2}{\rho_0}, \quad \lambda^2 \equiv \frac{\pi^2(k_{\mathrm{B}}T)^2}{3}. \tag{1.2}$$

Here k_{B} is Boltzmann's constant, T is the temperature, and ρ_0 represents the density of states (DOS) function per site and per spin at the Fermi level; ρ_0', ρ_0'' represent the first and second derivatives with respect to ε, evaluated at the Fermi level. Without loss of generality the zero of energy may be chosen to coincide with the Fermi energy as long as the DOS for the half–filled band is symmetric about its midpoint.

Since $\eta \leq 1/4$, the quantity Φ may be expanded in powers of η as

$$\Phi(\eta) = f_0 + f_1\eta + f_2\eta^2 + \ldots, \tag{1.3}$$

in which the various f_i are constants that are determined by imposing restrictions on the hopping of electrons between various sites [6]. The resulting quantity F_i, Eq. (1.1), is actually a functional of the unknown η and becomes a thermodynamic free energy upon imposing equilibrium conditions by minimization, setting $\partial F/\partial\eta = 0$. Since Φ appears both in the first term, as well as in the denominator of the temperature-dependent terms, the mathematical operations become unwieldy. It is therefore convenient to replace the quantity Φ by its low-temperature approximation in the denominators, as shown below. The resulting simplification is justified because the use of Eq. (1.1) itself is restricted to low temperatures.

The optimized double occupancy η_0 for a rectangular DOS is given by [6]:

$$\eta_0 = \frac{1}{4}\left(1 - \frac{U}{2W}\right), \tag{1.4}$$

where W is the width of the bare band. This quantity enters in the original derivation because it is related by the expression $W = 2z|q|$ to the transfer integral q which specifies the probability of an electron transfer to available adjacent sites, z being the number of nearest neighbors. Let us note that η_0 must be nonnegative; this introduces the restriction $U/(2W) \leq 1$, so that $2W \equiv U_{\mathrm{c}}$ may be considered as a cut-off energy for the Coulomb interactions; for larger U the material remains an insulator. Furthermore, the optimized band narrowing factor is given by [4]:

$$\Phi_0 = 8\eta_0(1 - 2\eta_0). \tag{1.5}$$

Equations (1.4) and (1.5) specify η_0 and Φ_0 in terms of the ratio U/U_{c} as the independent quantity.

2. Fundamental derivation

We begin by optimizing Eq. (1.1), which leads to

$$\frac{\partial(F_i/N)}{\partial\eta} = 0 = \left(-|\bar{\varepsilon}| + \frac{\rho_0\lambda^2}{\Phi_0^2} + \frac{3\,\Psi\lambda^4}{4\,\Phi_0^4}\right)\frac{\partial\Phi}{\partial\eta} + U. \tag{2.1}$$

On the assumption that the terms in λ will be "small", we expand η in terms of its value η_0 at $\lambda = 0$:

$$\eta = \eta_0 + h\lambda^2 + j\lambda^4, \tag{2.2}$$

and we similarly expand Φ as a Taylor series in λ^2 and λ^4 about its value at $\lambda = 0$:

$$\Phi \equiv \Phi_{\mathrm{A}} = \Phi_0 + (h\lambda^2 + j\lambda^4)\left(\frac{\partial\Phi}{\partial\eta}\right)_0 + \frac{1}{2}(h^2\lambda^4)\left(\frac{\partial^2\Phi}{\partial\eta^2}\right)_0. \tag{2.3}$$

The coefficients h and j are assumed to be independent of η_0.

Use of (2.2) in (2.3) in the current approximation yields

$$\Phi_{\mathrm{A}} = \Phi_0\left(1 + a\lambda^2/\Phi_0 + b\lambda^4/\Phi_0\right), \tag{2.4a}$$

with

$$a \equiv 8(1 - 4\eta_0)h, \quad b \equiv 8(1 - 4\eta_0)j - 16h. \tag{2.4b}$$

Now return to Eq. (2.1), and use the expansion (2.4a) while retaining only terms up to order λ^4; this yields

$$\frac{\partial(F_i/N)}{\partial\eta} = 0 = \left(-|\bar{\varepsilon}| + \frac{\rho_0\lambda^2}{\Phi_0^2} - 2\frac{\rho_0 a\lambda^4}{\Phi_0^3} + \frac{3\,\Psi\lambda^4}{4\,\Phi_0^4}\right) \times \frac{\partial\Phi_{\mathrm{A}}}{\partial\eta} + U. \tag{2.5}$$

In addition, we use Eqs. (2.2)–(2.4) to set

$$\frac{\partial\Phi_{\mathrm{A}}}{\partial\eta} = 8(1 - 4\eta_0) - 32(h\lambda^2 + j\lambda^4). \tag{2.6}$$

Next, let us introduce (2.6) in (2.5) and rearrange the result to obtain a sum of terms $A\lambda^0 + B\lambda^2 + C\lambda^4 = 0$. If this latter relation is to vanish for any arbitrary value of λ, as required by (2.5), we must demand that the coefficient multipliers of λ^0, λ^2 and λ^4 vanish separately. On imposing this requirement on the coefficient A of $\lambda^0 \equiv 1$, we find that

$$8(1 - 4\eta_0) = U/|\bar{\varepsilon}|, \quad \eta_0 = (1/4)(1 - U/U_{\mathrm{c}}), \tag{2.7a}$$

which reproduces Eq. (1.4) and justifies the assertion that η_0 specifies the double occupancy at $\lambda = 0$. Thus, for any $U < U_{\mathrm{c}}$ double occupancies and empty lattice sites exist even at $T = 0$; however, the corresponding transition

probabilities for electron transfers are also very low, so that such materials are exceedingly poor conductors. We also note for future use that at the critical value $U = U_c$, $\eta_0 = 0$, whence

$$U_c = 8|\bar{\varepsilon}|. \tag{2.7b}$$

Proceeding similarly with the coefficient B of λ^2 and by use of (2.7a) we may solve for

$$h = \frac{-\rho_0 U}{32|\bar{\varepsilon}|^2 \Phi_0^2}, \tag{2.7c}$$

and from the coefficient C of λ^4, using (2.7a) and (2.7b), we obtain

$$j = \frac{-\rho_0^2 U}{32|\bar{\varepsilon}|^2 \Phi_0^4} - \frac{\rho_0^2 U^3}{512|\bar{\varepsilon}|^5 \Phi_0^5} - \frac{3U\Psi}{128|\bar{\varepsilon}|^2 \Phi_0^4}. \tag{2.7d}$$

It remains to determine the free energy per site via Eq. (1.1), using Eqs. (2.2) and (2.3). We find that

$$\frac{F_i}{N} = -|\bar{\varepsilon}| \left[\Phi_0 + 8(h\lambda^2 + j\lambda^4)(1 - 4\eta_0) - 16h^2\lambda^4\right]$$

$$+U(\eta_0 + h\lambda^2 + j\lambda^4) - \frac{\rho_0\lambda^2}{\Phi_A} - \frac{\Psi\lambda^4}{4\Phi_A^3}. \tag{2.8}$$

On introducing (2.4b) and (2.7a) we encounter considerable cancellation, leading to the simplified relation

$$\frac{F_i}{N} = -|\bar{\varepsilon}|\,\Phi_0 + U\eta_0 - \frac{\rho_0\lambda^2}{\Phi_A} - \frac{\Psi\lambda^4}{4\Phi_A^3} + 16|\bar{\varepsilon}|h^2\lambda^4. \tag{2.9}$$

Next, expand Φ_A in the denominators using (2.3) while retaining only terms of order λ^2 and λ^4; in this process quantities involving j drop out. Then eliminate h via (2.7c) and a, via (2.4b). After some tedious algebra one ends up with the final expression for the Helmholtz free energy per site as

$$\frac{F_i}{N} = -|\bar{\varepsilon}|\,\Phi_0 + U\eta_0 - \frac{\rho_0\lambda^2}{\Phi_0} - \frac{U^2\rho_0^2\lambda^4}{64|\bar{\varepsilon}|^3\Phi_0^4}$$

$$-\frac{\lambda^4}{4\Phi_0^3}\left[\frac{7}{5}\rho_0'' - \frac{(\rho_0')^2}{\rho_0}\right]. \tag{2.10}$$

For future use it is highly desirable to put everything in dimensionless form. We thus divide both sides by W; note that by (2.7b), $U_c = 8|\bar{\varepsilon}|$; and set $\lambda \equiv Wt$, yielding

$$\frac{F_i}{NW} = -\frac{1}{8}\frac{U_c}{W}\left(1 - \frac{U/W}{U_c/W}\right)^2 - \frac{(\rho_0 W)t^2}{\left[1 - (U/U_c)^2\right]}$$

$$-\frac{8\,(U/U_c)^2\,(\rho_0 W)^2 t^4}{(U_c/W)\left[1 - (U/U_c)^2\right]^4}$$

$$-\frac{W^3 t^4}{4\left[1 - (U/U_c)^2\right]^3}\left[\frac{7}{5}\rho_0'' - \frac{(\rho_0')^2}{\rho_0}\right]. \tag{2.11}$$

In the above the ratio U_c/W, the product $\rho_0 W$, and the multiplier of t^4 in the numerator of the last term are determined via the DOS function $\rho(\varepsilon)$, as shown below. Thus, F_i/NW depends functionally on $t^2 \equiv (\pi^2/3)(k_B T/W)^2$ and parametrically on the choice for U/U_c. Equation (2.11) is the end result of interest; this formulation is a slightly generalized version of a previously published result [11]. The above relations determine the thermodynamic characteristics of an interactive electron assembly in the present approximation.

We now focus attention on the transition from the itinerant to the insulating state. In the latter case every electron remains immobile and each occupies one site in either the "spin up" or the "spin down" configuration; hence, no unoccupied sites remain. The entropy per site for such localized carriers is then given by $S_l/N = (k_B/N)\ln 2$, while their kinetic energies vanishes. Hence the reduced Helmholtz free energy for the totally localized configuration is specified as

$$F_l/NW = -(k_B T/W)\ln 2. \tag{2.12}$$

As is well established, a transition from the itinerant to the localized state occurs under conditions where the free energies, as specified by (2.11) and (2.12), exactly match. This requirement fixes the conditions under which metal–insulator transitions may occur; it forms the basis of much of the subsequent discussion.

3. Special cases

We now consider a variety of special cases.

3.1. Rectangular density of states

The detailed discussion that follows is also applicable, with appropriate modifications, to the remainder of the cases considered below. For the rectangular density of states (RDOS) case we take $\rho = \rho_0 = 1/W$ as constant in the energy range $-W/2 \le \varepsilon \le W/2$ and zero outside. For a half-filled band $|\bar{\varepsilon}| = W/4$; this includes the factor of two required to take account of spin degeneracy; also, $U_c = 8|\bar{\varepsilon}| = 2W$, and the last term in (2.11) drops out. We are left with

$$\frac{F_i}{NW} = -\frac{1}{4}\left(1 - \frac{U}{2W}\right)^2 - \frac{t^2}{1 - (U/2W)^2}$$

$$-\frac{4(U/2W)^2 t^4}{\left[1 - (U/2W)^2\right]^4}. \tag{3.1}$$

Consider first only the quadratic dependence and compare (3.1) with (2.12): Eq. (3.1) gives rise to a set of curves on a plot of F/NW vs. $k_B T/W \equiv \tau$ that depend on the choice for $U/U_c = U/(2W)$. By contrast, Eq. (2.12) specifies a single straight line, as shown in Fig. 1. One notes that for $U/U_c > 0.72$ the curves intersect the straight line at two points, while for smaller values all curves lie below the straight line. In a standard interpretation of free energy plots, the equilibrium configuration is determined by the smaller of the two F/NW values. Accordingly, where intersections occur, the itinerant state prevails at the lowest values of $\tau \equiv k_B T/W$; then there exists a mid range of reduced temperatures where the localized configuration is stable; followed by a higher temperature range where one again encounters the delocalized state. This is the explanation for the reentrant metallic scenario. Where no intersections occur, only the itinerant state is stable over the whole reduced

temperature range. Let us note further that the range of stability of the localized state widens and shifts to higher τ values as U/U_c diminishes towards its critical value near 0.74. Such a widening feature is unphysical since a decrease in U should engender a shrinkage in the domain of localization. This finding signals a failure of the low T approximation to hold at larger values of the upper crossover points; these particular intersections must therefore be disregarded. However, the lower crossover points always fall within the applicable range of the approximation, as was verified by detailed numerical calculations. Similar considerations apply to all later calculations.

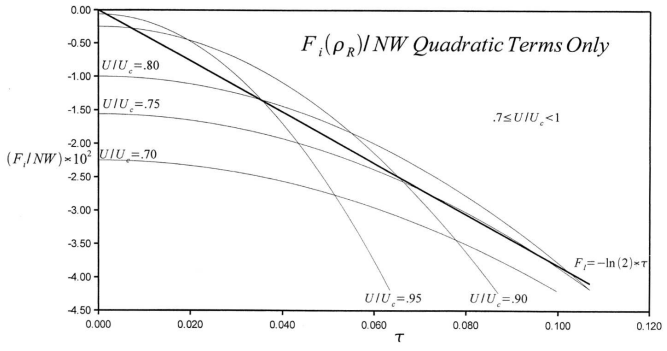

Fig. 1. Reduced Helmholtz free energies for itinerant and localized electrons as a function of reduced temperature τ for various U/U_c ratios with $U_c = 2W$; calculations based on the rectangular DOS specified by Eq. (3.1), with omission of the last term. Note the double intersections of the parabolas with the straight line for $U/U_c > 0.72$. Calculations restricted to parabolic τ dependence.

Comparable results based on the full expression, Eq. (3.4) are shown in Fig. 2; relative to Fig. 1, all transitions are shifted to lower τ values.

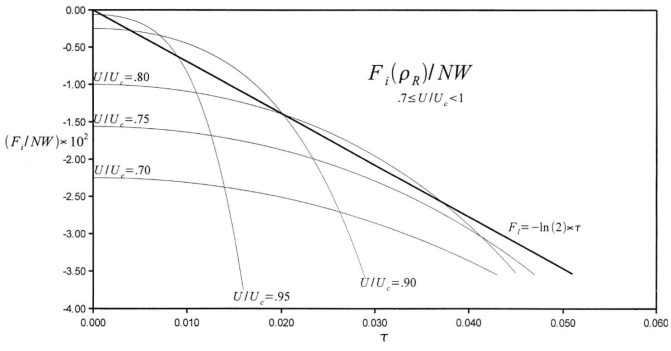

Fig. 2. Plots similar to those of Fig. 1 for the RDOS in which terms of order τ^4 are included.

As a second way of interpreting the above findings, we invoke the match of the two reduced free energies at the transition point. This permits one to determine the phase boundary separating the localized and the itinerant configurations. One sets $F_i = F_l$ to solve the resulting quartic equation by finding the value of τ for a given U/U_c at which (3.1) matches (2.12), for which purpose we employed the Ferrari algorithm [12]. This procedure generated the lower retrograde curve in Fig. 3 which shows the reduced transition temperature as a function of the reduced interaction energy†. For comparison we also show in the upper curve the phase boundary when only the term in t^2 of Eq. (3.1) is retained. The range of reduced temperatures over which the localized configuration is stable is much larger compared to the case where the term in t^4 is included. The difference arises because the last term in (3.1) increases the negative slope of F_i/NW and thereby diminishes the possibility of having these curves intersect with the straight line of Eq. (2.12).

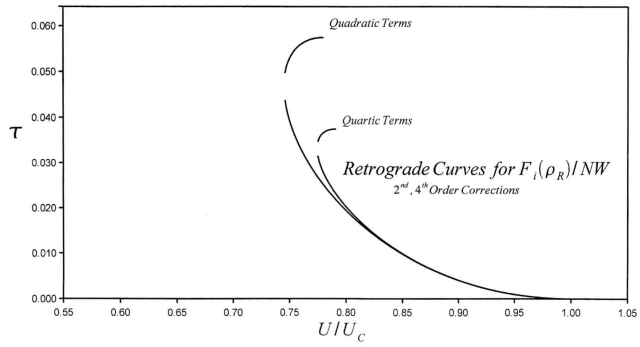

Fig. 3. Retrograde curves representing the phase boundaries which separate the itinerant from the localized regime when F_i/NW is specified for reduced temperatures τ limited to quadratic and/or quartic expansion terms. Calculations based on a rectangular DOS function. The curves were cut off artificially at their maximum values, beyond which the low temperature expansion fails. $U_c = 2W$.

Consistent with the elimination of many upper crossover points of Figs. 1 and 2, we were forced to cut off the retrograde curves at their maximum values, beyond which the curve assumes a negative slope. That the low temperature approximation does fail beyond that point was again verified by calculation. Also, we terminated the curves for values $U/(2W) \equiv U/U_c > 1$. Similar statements apply to all retrograde curves presented later.

Several additional remarks are in order. According to the retrograde curves, for $U/U_c < 0.74$ or 0.76 for quadratic or quartic terms in t, the electron interactions are sufficiently small so as to suppress any transition. In the range $0.74 < U/U_c < 0.78$ or $0.77 < U/U_c < 0.79$ in Fig. 3, raising the reduced temperature at a fixed value of

† The break between the lower and upper part of the curve (as well as in all later retrograde curves) is an artifact of the calculation: the computer program had difficulty matching T values in this interval. However, the reader may readily fill in the gap with a smooth interpolating curve.

U/U_c generates a transition first from the metallic to the insulating state and then a second change from the insulating state back to the metallic region — the regime of reentrant metallic behavior. Lastly, for $0.78 < U/U_c \leq 1$ or $0.79 < U/U_c \leq 1$ and with rising temperature only a single transformation from the metallic to the insulating state is encountered. Here the electronic repulsions become sufficiently strong that the metallic phase is confined to very low reduced temperatures. A second intersection may occur at very high temperatures where the material may no longer be stable, and where the low temperature approximation fails. Similar interpretations obtain for all other retrograde curves presented below.

Lastly, the dependence of the transitions on the degree of occupancy of the sites was investigated. It turns out that for just very tiny departures from a one-to-one match between electron numbers and sites in the crystal, the insulating state is suppressed. This may be understood on the basis that when sites remain unoccupied at $T = 0$, electrons may in principle tunnel from their positions to neighboring, unoccupied sites; the truly insulating state is thereby suppressed. Double occupancies at elevated temperatures simply produce more empty sites, thereby adding to the electron transfers that prevail at $T = 0$. Coincident with intuition, it requires only very slight departures from complete occupation to destabilize the insulating state. This is verified by detailed calculations based on the theory expounded in Ref. [6].

3.2. Semielliptical density of states

We next consider a somewhat more realistic case, namely a normalized semielliptical density of states (SEDOS), which peaks at the center of a symmetrically disposed energy range; it is specified by

$$\rho(\varepsilon) = \frac{4}{\pi W}\sqrt{1 - \left(\frac{2\varepsilon}{W}\right)^2}, \tag{3.2}$$

where the kinetic energy is restricted to $-W/2 \leq \varepsilon \leq W/2$, and is zero outside the range. The density of states function exhibits a maximum at the Fermi level at $\mu = \varepsilon = 0$. Thus,

$$\rho_0 = 4/\pi W,\ \rho_0' = 0,\ \rho_0'' = -16/\pi W^3,\ W > 0. \tag{3.3a}$$

Also, for a half-filled band, while taking account of spin degeneracy, the average kinetic energy is given by

$$\bar{\varepsilon} = 2\int_{-W/2}^{0} \varepsilon\rho(\varepsilon)\mathrm{d}\varepsilon = -2W/3\pi,$$

$$U_c = 8|\bar{\varepsilon}| = 16W/3\pi \approx 1.698W. \tag{3.3b}$$

When these results are inserted in (2.11) one obtains

$$\frac{F_i}{NW} = -\frac{2}{3\pi}\left(1 - \frac{U}{U_c}\right)^2 - \frac{(4/\pi)\,t^2}{1 - (U/U_c)^2} - \frac{24}{\pi}\frac{(U/U_c)^2 t^4}{\left[1 - (U/U_c)^2\right]^4} + \frac{28}{5\pi}\frac{t^4}{\left[1 - (U/U_c)^2\right]^3}. \tag{3.4}$$

The above result may be compared with the corresponding expression for the localized configuration, Eq. (2.12).

In Fig. 4 we show plots of F_i/NW vs. $\tau \equiv k_B T/W$, for a variety of choices of U/U_c, based on Eq. (3.4) and compare these with the straight line relationship based on Eq. (2.12). Note again the double crossover of the free energy curves for the itinerant configuration across that of the localized configuration (straight line), for larger values of $U/U_c > 0.81$, with the same interpretation as was given earlier. For values below this critical ratio only the itinerant state is stable.

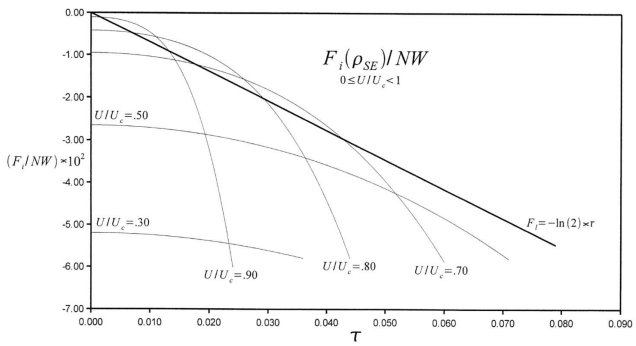

Fig. 4. Reduced Helmholtz free energies for itinerant and localized electrons as a function of reduced temperature τ for various U/U_c ratios; calculations based on Eq. (3.4) for the semielliptical DOS. Note the double intersections of the curves with the straight line for $U/U_c > 0.8$. Expansions restricted to terms in τ^2 and τ^4.

The boundary line between the itinerant and localized states is found as before by equating Eqs. (3.4) and (2.12) and numerically determining [12] the temperatures τ that correspond to various assumed U/U_c values. The resulting retrograde curve is shown in Fig. 5, which is interpreted in a manner similar to that provided above. Results obtained by use of Eq. (3.4) with the t^4 terms missing were published elsewhere [10]; the earlier curve is similar in aspect to that of Fig. 5 and differs only in having a larger maximum τ value. Comparison with Fig. 3 shows that the stability region of the insulating state (inside the retrograde domain) extends over a somewhat larger U/U_c range than was the case for the RDOS situation; also, the entire curve is shifted slightly downward on the τ scale relative to Fig. 3.

3.3. Central peak density of states

This situation is illustrated by use of the function

$$\rho(\varepsilon) = b(\varepsilon^2 - \varepsilon_0)^2, \tag{3.5}$$

whose profiles are presented in Fig. 6 for several values of the parameter ε_0 at a fixed bandwidth W, which is representative of a DOS that peaks weakly or strongly in the central energy range, and has two adjacent dips that vanish for $\varepsilon = \pm\sqrt{\varepsilon_0}$. Normalization is achieved by setting

$$1 = b\int_{-W/2}^{W/2} (\varepsilon^2 - \varepsilon_0)^2 \mathrm{d}\varepsilon. \tag{3.6}$$

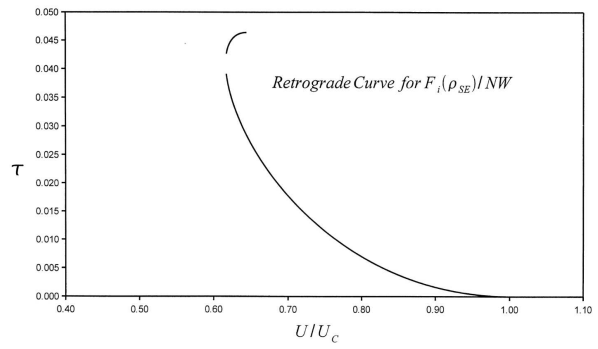

Fig. 5. Retrograde curve for the phase boundary separating the itinerant from the localized regime when F_i/NW is specified for reduced temperatures involving τ limited to quadratic and quartic expansion terms. Calculations based on semiellipticalDOS functions. The curves were cut off artificially at their maximum values, beyond which the low temperature expansion fails.

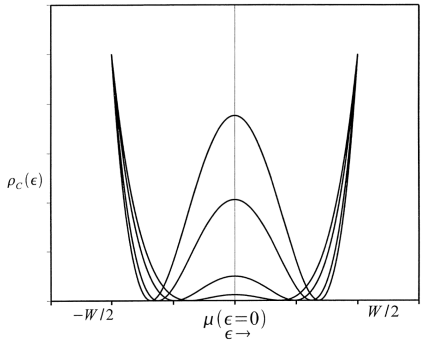

Fig. 6. Plot of the CPDOS function (3.5), (3.7a), (3.7b), with $\varepsilon_0 = g_0(W/2)$, showing the placement of the minima relative to the band edges. The g_0 values increase with increasing size of the central peak.

For ease in future mathematical operations we define a new parameter by

$$\varepsilon_0 \equiv g_0(W/2)^2, \quad 0 < g_0 < 1, \tag{3.7a}$$

where g_0 adjusts the location of the minima in Fig. 6 relative to the end points at $\pm W/2$, thus introducing a new degree of freedom in deciding on the shape of the DOS. Straightforward integration of (3.6) leads to the result

$$b = \frac{8}{W^5}\left(\frac{1}{10} - \frac{1}{3}g_0 + \frac{1}{2}g_0^2\right)^{-1}, \tag{3.7b}$$

where W is the total width of the bare band. We also require the average band energy for the half filled band, symmetrically disposed about the Fermi energy at $\varepsilon = 0$:

$$\bar{\varepsilon} = 2b\int_{-W/2}^{0} \varepsilon(\varepsilon^2 - \varepsilon_0)^2 \mathrm{d}\varepsilon = -\frac{W}{24}\frac{1 - 3g_0 + 3g_0^2}{[1/10 - (1/3)g_0 + (1/2)g_0^2]}, \tag{3.8}$$

from which one can obtain $U_c/W = 8|\bar{\varepsilon}|/W$.

Next, we determine

$$\rho(0) = b\bar{\varepsilon}_0^2 = \frac{g_0^2}{2W[1/10 - (1/3)g_0 + (1/2)g_0^2]}, \tag{3.9a}$$

$$\rho'(0) = 0, \tag{3.9b}$$

$$\rho''(0) = \frac{-8g_0}{W^3[1/10 - (1/3)g_0 + (1/2)g_0^2]}. \tag{3.9c}$$

Equation (2.11) may then be used to determine the reduced Helmholtz free energy for the present central peak density of states (CPDOS) as

$$\frac{F_i}{NW} = \frac{-U_c}{8W}\left[1 - \left(\frac{U}{U_c}\right)\right]^2 - \frac{(1/2)g_0^2t^2}{[1/10 - (1/3)g_0 + (1/2)g_0^2]}\frac{1}{1 - (U/U_c)^2} - 4\left(\frac{U}{U_c}\right)^2\frac{g_0^4t^4}{(U_c/W)\,[1 - (U/U_c)^2]^4} \times \frac{1}{[1/10 - (1/3)g_0 + (1/2)g_0^2]^2} + \frac{14g_0t^4}{5\,[1 - (U/U_c)^2]^3\,[1/10 - (1/3)g_0 + (1/2)g_0^2]}. \tag{3.10}$$

Attention is directed to the obvious requirement $0 < g_0 < 1$. This is also sufficient to guarantee that b in Eq. (3.7b) and $\bar{\varepsilon}$ in Eq. (3.8) remain positive and negative respectively. F_i/NW now depends functionally on t and parametrically on U/U_c as well as on g_0. The introduction of this second parameter provides greater flexibility. U_c/W is specified by Eq. (3.8).

Figure 7 provides representative plots of F_i/NW vs. τ for a variety of U/U_c values (in a half-filled, nondegenerate band) at the fixed value of $g_0 = 0.7055$, relative to the same quantity for the localized configuration, with the usual interpretation of the crossover phenomenon. Again, falling values of U/U_c ultimately unduly broaden the region of stability of the insulating phase, due to the failure of the approximation scheme. However, the low τ intersections fall within the bounds of the approximation. For $U/U_c < 0.80$ only the itinerant state prevails. Figure 8 represents a similar plot in which $U/U_c = 0.95$ is fixed and g_0 is varied as shown. The intercepts tend to bunch into the range $-0.04 > F_i/NW > -0.08$. More extensive calculations indicate that curves for which g_0 decreases from unity to roughly $g_0 = 0.706$ are shifted downward on the τ scale, with a reversal to higher τ values for smaller g_0. Figure 9 illustrates the retrograde nature of the phase boundaries separating the itinerant from the localized state for several g_0 values. While the general shape of the retrograde curves is not very sensitive to the choice for g_0, the τ values at the cutoff do vary with g_0. The range of U/U_c values for these curves is comparable to that of Fig. 3, but the τ range is much lower.

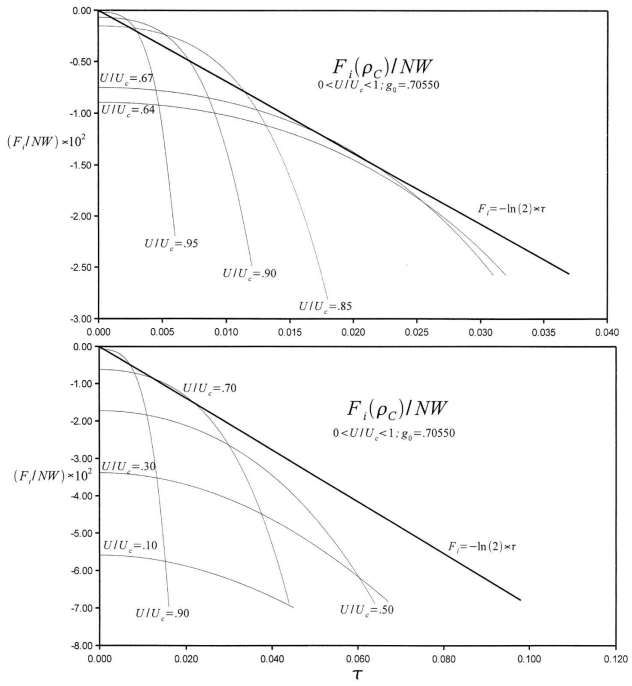

Fig. 7. Reduced Helmholtz free energies for itinerant and localized electrons as a function of reduced temperature τ for two sets of U/U_c ratios at the fixed value $g_0 = 0.7055$; calculations are based on the DOS of Fig. 6. Let us note the double intersections of the curves with the straight line for $U/U_c > 0.8$. Expansions restricted to terms in τ^2 and τ^4.

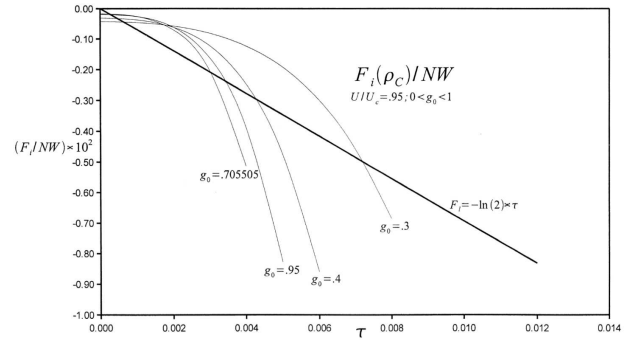

Fig. 8. Reduced Helmholtz free energies for itinerant and localized electrons as a function of reduced temperature τ for various values of g_0 at the fixed ratio $U/U_c = 0.95$; calculations based on the DOS of Fig. 6. Note the various double intersections with the straight line. Expansions restricted to terms in τ^2 and τ^4.

3.4. Pseudogap density of states

Here we consider the pseudogap density of states (PGDOS), DOS which has a minimum in the density of states in the central energy range, namely

$$\rho = -c\{(\varepsilon^2-\varepsilon)^2 - [(W/2)^2-\varepsilon_0]^2\}, \tag{3.11}$$

whose profile is shown in Fig. 10; this is the previous DOS in inverted form and mimics a band structure displaying

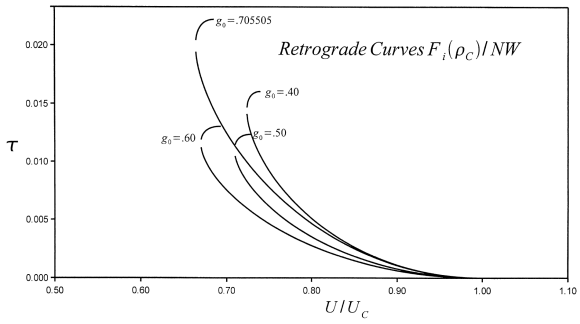

Fig. 9. Retrograde curves representing the phase boundaries which separate the itinerant from the localized regime when F_i/NW is specified for reduced temperatures involving τ limited to quadratic and quartic expansion terms. Calculations based on the DOS function (3.5) to (3.7b). The curves were cut off artificially at their maximum values, beyond which the low temperature expansion fails.

a pseudogap. The conventional normalization procedure leads to the result

$$c = \left[\int_{-W/2}^{W/2}\{(\varepsilon^2-\varepsilon)^2 - [(W/2)^2-\varepsilon_0]^2\}\mathrm{d}\varepsilon\right]^{-1}$$
$$= \frac{4}{W^5}\frac{1}{(1/5)-(1/3)g_0}, \tag{3.12}$$

where now

$$\rho = \frac{4}{W^5(1/5-g_0/3)}$$
$$\times\left[(1-2g_0)\left(\frac{W}{2}\right)^4 + 2g_0\left(\frac{W}{2}\right)^2\varepsilon^2 - \varepsilon^4\right]. \tag{3.13}$$

It follows that

$$\rho(0) = \frac{1-2g_0}{4W(1/5-g_0/3)}, \quad \rho'(0) = 0,$$
$$\rho''(0) = \frac{4g_0}{W^3(1/5-g_0/3)}. \tag{3.14}$$

We also require for a half-filled band

$$\bar{\varepsilon} = 2c\int_{-W/2}^{0}\varepsilon\rho(\varepsilon)\mathrm{d}\varepsilon = -\frac{W}{8}\frac{1/3-g_0/2}{1/5-g_0/3}. \tag{3.15}$$

The reduced Helmholtz free energy for this model is then obtained from Eq. (2.11) as

$$\frac{F_i}{NW} = -\frac{U_c}{8W}\left[1-\left(\frac{U}{U_c}\right)\right]^2$$
$$-\frac{1}{4}\frac{(1-2g_0)t^2}{(1/5-g_0/3)[1-(U/U_c)^2]}$$
$$-\frac{1}{2}\left(\frac{U}{U_c}\right)^2\frac{(1-2g_0)^2t^4}{(U_c/W)(1/5-g_0/3)^2[1-(U/U_c)^2]^4}$$
$$-\frac{7}{5}\frac{g_0t^4}{[1/5-g_0/3][1-(U/U_c)^2]^3}. \tag{3.16}$$

To keep the coefficient of the t^2 term positive, as re-

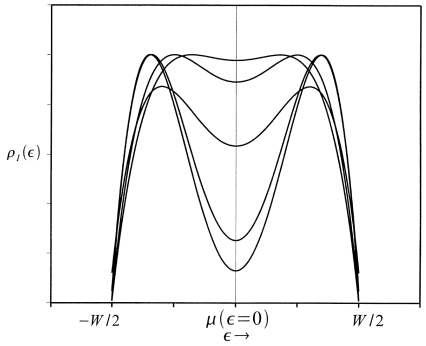

Fig. 10. Sketch of the PGDOS function (3.11) and (3.12), with $\varepsilon_0 = g_0(W/2)$, showing the placement of the maxima relative to the band edges. The g_0 values increase with increasing size of the central minima.

quired on physical grounds, we require that $g_0 < 1/2$; then the quantities $(1/5 - g_0/3)$ and $(1/3 - g_0/2)$ remain positive as well. The resulting calculations are shown in Fig. 11 and Fig. 12 as plots of Eq. (3.16) for a variety of g_0 and U/U_c values; these are compared with the relation for localized carriers. Corresponding retrograde

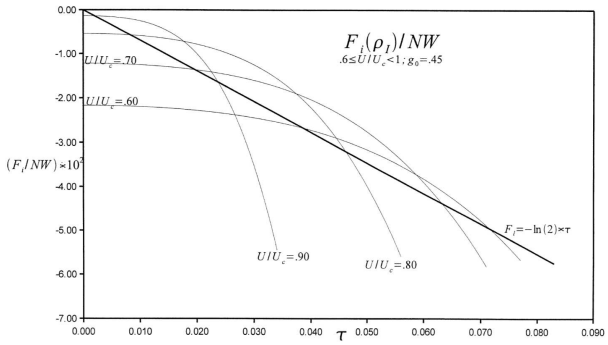

Fig. 11. Reduced Helmholtz free energies for itinerant and localized electrons as a function of reduced temperature τ for various U/U_c ratios at the fixed value $g_0 = 0.45$; calculations based on the DOS of Fig. 10. Note the double intersections of the curves with the straight line for $U/U_c > 0.75$. Expansions restricted to terms in τ^2 and τ^4.

curves are shown in Fig. 13 for several g_0 values. The features encountered here differ from those for the CPDOS case in being much less diffuse, with τ and U/U_c values comparable to those of Fig. 3 for the RDOS.

4. General discussion

Aside from the earlier commentary some other observations are of interest. All retrograde curves displayed above are similar in shape and differ principally in the maximum values of τ beyond which the approximation scheme fails. The set of acceptable τ values where the "low temperature approximation" holds may be determined by noting the cutoffs where the retrograde curves

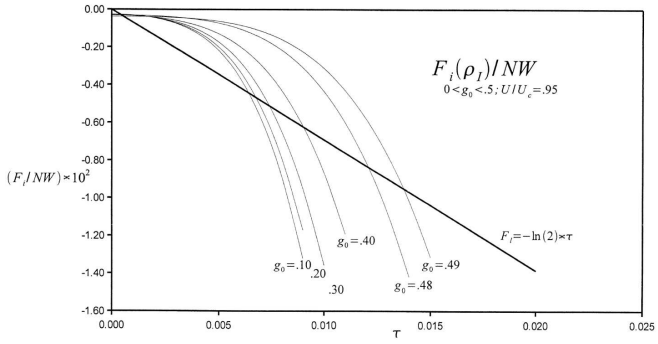

Fig. 12. Reduced Helmholtz free energies for itinerant and localized electrons as a function of reduced temperature τ for various g_0 values at the fixed ratio $U/U_c = 0.95$; calculations based on the PGDOS of Fig. 10. Let us note the double intersections of the curves with the straight line for $U/U_c > 0.75$. Expansions restricted to terms in τ^2 and τ^4.

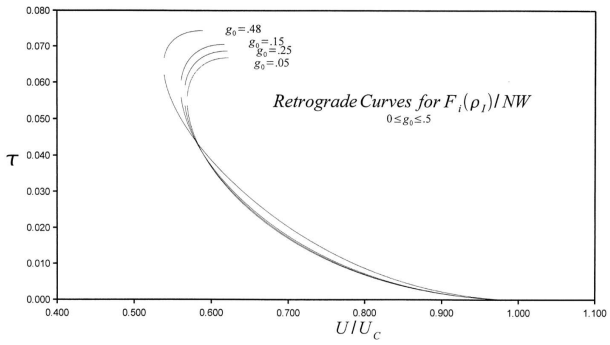

Fig. 13. Retrograde curves representing the phase boundaries which separate the itinerant from the localized regime when F_i/NW is specified for reduced temperatures involving τ limited to quadratic and quartic expansion terms. Calculations based on the PGDOS function (3.16) for various g_0 values. The curves were cut off artificially at their maximum values, beyond which the low temperature expansion fails.

achieve their maximum values. These τ ranges depend on the assumed DOS functions. As a rough approximation, for the rectangular, semielliptical, and pseudogap functions, where $|\bar{\varepsilon}|$ falls roughly halfway in the range $-W/2 \leq \varepsilon \leq 0$ of occupied electron energy states, the upper limit is $\tau \leq 0.046$. By contrast, for the central peak functions, where $|\bar{\varepsilon}|$ is closer to the upper limit of $\varepsilon = 0$, the upper limit falls close to the values $\tau \leq 0.01$ to $\tau \leq 0.013$. This reflects the very different shapes in the DOS curves and the corresponding mathematical relationships involved in the various reduced Helmholtz free energy functions.

The cutoff values also provide an indication whether the values of τ shown in the abscissae of the free energy curves are physically reasonable. For the two ranges discussed above, select $\tau = 0.03$ and $\tau = 0.01$ as representative values. The corresponding reduced tempera-

tures $k_{\rm B}T/W$ lie in the 3×10^{-2} to 10^{-2} range, which is a reasonable ratio for thermal energies relative to bare bandwidths.

From the reduced free energy curves one may also easily determine the reduced heat capacities at constant volume, C_V, using the thermodynamic relation

$$\frac{C_V(t)}{NW} = t\frac{{\rm d}[S_i(t)/NW]}{{\rm d}t}. \tag{4.1}$$

The entropy S may be read off from (2.11), based on the relation $F_i(t)/NW = E_i/NW - tS_i(t)/NW$. This requires us: to single out the temperature-dependent terms in (2.11); reduce their powers of t by one unit; carry out the indicated differentiation; and multiply the resultant by t. One thereby generates the electronic contribution to the total heat capacity in terms of a factor that multiplies t and a factor multiplying t^3. The first term represents a generalized Sommerfeld formulation for the electronic heat capacity, while the second term represents its extension to the next higher power in reduced temperature. Details of these derivations, and specialization to the various DOS function considered above, are left to the reader.

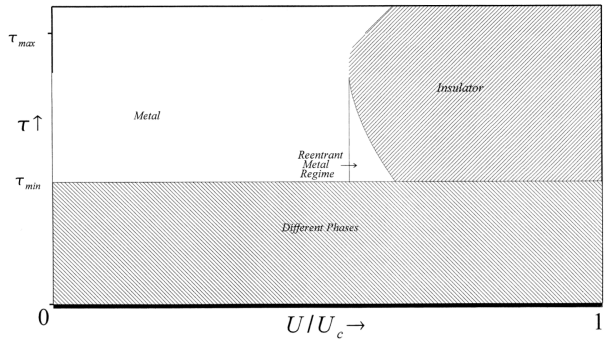

Fig. 14. Low τ phase transitions. Sketch of a composite phase diagram for τ vs. $U/U_{\rm c}$ showing: the domain of low temperatures where most materials undergo phase transitions not accounted for by the present theory; the regime of itinerant and localized electron configurations; and the limited range where reentrant metallic behavior is encountered. The upper sloping boundary of the insulating regime represents a rough extrapolation beyond the low-temperature approximation.

In confronting these calculations with experiment, one should note that in almost all experimental studies, solids undergo some type of phase transition at low temperature such as: magnetic ordering; crystallographic transitions to lower symmetry; order–disorder transformations; and the like. These engender concomitant changes in the DOS that are not considered in the present theory. Hence, the extremely low temperature metallic regime specified by the above theory is not ordinarily experimentally realized; one exception to this was noted earlier. In Fig. 14 we provide a cartoon in which the horizontal line represents the temperature below which, as indicated by the shaded region, the present theory does not apply to most experimental observations. We further delineate on the upper right the approximate location of the phase boundary that separates the insulating from the metallic phase beyond the point where the low-temperature approximation fails. The corresponding transition may be either first order or may consist of a crossover regime. Alternatively, the upper crossover point may require such high temperatures that the solid is no longer stable, and may therefore be unobservable. Pending the construction of a generalized theory, the only guide here is the requirement that the domain of the localized configuration should increase extensively as U approaches its cutoff value $U_{\rm c}$. Attention is also directed to the very limited region of the phase diagram for which reentrant metallic behavior is anticipated.

To summarize: We have examined the effects of changes in the DOS and have extended the range of temperatures for specifying the Helmholtz free energy of itinerant charge carriers in a half-filled, nondegenerate band. By comparison with the Helmholtz free energy of localized carriers we have also delineated the conditions under which metal–insulator transitions are anticipated in the low temperature approximation. This provides a general framework for subsequent calculations of interest to readers, either by extending the temperature range of applicability and/or by examining the effect of introducing other types of DOS functions. In light of the above the present theory is useful primarily in rationalizing observations of both reentrant metallic behavior and of any metal–insulator transition with rising temperatures close to absolute zero.

Acknowledgments

The authors wish to acknowledge their indebtedness to Professor Józef Spałek of the Jagiellonian University in Kraków, Poland, for his pioneering insights that established the theoretical methodology on which the present paper is based.

This research was supported by award SI 06 021 of the Faculty Senior Scientist Mentor Program of the Dreyfus Foundation.

References

[1] H. Kuwamoto, J.M. Honig, J. Appel, *Phys. Rev. B* **22**, 2626 (1980).

[2] X. Yao, J.M. Honig, T. Hogan, C. Kannewurf, J. Spałek, *Phys. Rev. B* **54**, 17469 (1996).

[3] See e.g., M.S. Laad, L. Craco, E. Müller-Hartmann, *Phys. Rev. Lett.* **91**, 156402 (2003); A.I. Potayaev, J.M. Tomczak, S. Biermann, A. Georges, A.I. Lichtenstein, A.N. Rubtsov, T. Saha-Dasgupta, O.K. Andersen, *Phys. Rev. B* **76**, 085127 (2007); G. Keller, K. Held, V. Eyert, D. Vollhardt, V.I. Anisimov, *Phys. Rev. B* **70**, 205116 (2004).

[4] N.F. Mott, *Proc. Roy. Soc. London A* **62**, 416 (1949).

[5] J. Hubbard, *Proc. Roy. Soc. A* **281**, 401 (1964).

[6] J. Spałek, A.M. Oleś, J.M. Honig, *Phys. Rev. B* **28**, 6802 (1983).

[7] J. Spałek, A. Datta, J.M. Honig, *Phys. Rev. B* **33**, 4891 (1986).

[8] J. Spałek, J.M. Honig, M. Acquarone, A. Datta, *J. Magn. Magn. Mat.* **54–57**, 1047 (1986).

[9] J. Spałek, A. Datta, J.M. Honig, *Phys. Rev. Lett.* **59**, 728 (1987).

[10] A. Datta, J.M. Honig, J. Spałek, *Phys. Rev. B* **44**, 8459 (1991).

[11] J. Spałek, M. Kokowski, J.M. Honig, *Phys. Rev. B* **39**, 4175 (1989).

[12] The algorithm is detailed in http://en.wikipedia.org/wiki/Quartic_equation#The_general_case.2C_along_Ferrari.27s_lines This method was adapted for our use with the Excel program.