Many-body effects at metal–semiconductor junctions:
I. Surface plasmons and the electron–electron screened interaction

J C INKSON

Cavendish Laboratory, Cambridge, UK

Abstract. The electron–electron screened interaction near the metal–semiconductor junction is calculated analytically by the introduction of model bulk dielectric functions which contain all of the essential energy and wavelength dependent features of the real dielectric functions of the metal and semiconductor. The poles of the interaction give the dispersion relations for the surface plasmon excitations at the junction and these are studied for various parameters of the two bulk media. It is shown that there are two possible plasmon modes characteristic of bimetal and metal–classical dielectric interfaces. The metal–dielectric plasmon has a long wavelength energy which is always within the direct band gap energy of the semiconductor and so could be important for some tunnelling experiments. The high energy 'metal–metal' plasmon is damped by the ability to decay into a bulk plasmon of the metal and in addition shows a minimum in the plasmon dispersion relation which is shown to be due to the attractive form of the interaction in that energy range. Finally the rôle of surface plasmons in Schottky barriers is discussed.

1. Introduction

In a recent paper (Inkson 1971b) we derived a new approximation to the full nonlocal energy dependent electron–electron screened interaction near an interface. We showed how the use of such a function would put many-body calculations of surface properties on a firmer footing and in addition discussed the exchange and correlation potential and the surface plasmon dispersion relation for the metal–vacuum system. In this, the first of two connected papers, we derive the screened interaction for the metal–semiconductor junction and examine the properties of the surface (or rather interface) plasmons. This interaction will be used in the second paper to derive the electron self energy and hence an effective nonlocal exchange and correlation potential. The resulting effects upon the band structure in the vicinity of the interface and its application to Schottky barrier formation will also be considered.

Plasmons, collective excitations of the electrons, follow directly from the work of Bohn and Pines (1952, 1953) on the exchange and correlation energy of the electron gas and have been identified many times in electron transmission and reflection experiments (Pines 1955). A surface plasmon mode was first suggested by Ritchie (1957); later Stern and Ferrel (1960) showed that it could be considered as an oscillating charge distribution at the plasma–vacuum interface. They derived a simple relation for the long wavelength energy limit which was confirmed by the elegant experiments of Powell and Swan (1960).

The actual dispersion relation for the surface plasmon (ie how the energy varies with wavelength) is difficult to obtain. Theoretically it should be given by the zeros of the dielectric function for the complete system as in the bulk (Pines 1963). There have been a number of detailed calculations using RPA or the time dependent Hartree method for the metal–vacuum system with varying results (Newns 1970, Fedders 1967, Feibleman 1968). Harris and Griffin (1971) have shown that within RPA the dispersion relation is critically dependent upon the model used for the wavefunctions but they conclude that one should get a linear term in the dispersion relation.

In discussing the metal–semiconductor junction one is up against an even more complicated system, certainly from the point of view of a detailed calculation of the RPA type. Even for the bulk semiconductor the agreement between theoretical and experimental dielectric functions leaves a lot to be desired. There are a number of reasons however for wishing to know the properties of the plasmons at the metal–semiconductor junction. Phillips (1970) has suggested that low energy plasmons ($\lesssim$ the direct band gap energy) at the interface may be responsible for the charge accumulation which results in the pinning of the Fermi level in Schottky barrier formation. If these low energy plasmons exist they may also be important in electron tunnelling experiments. For instance Tsui (1969) has observed the increase in conductance due to the onset of surface plasmon excitations in forward biased lead–GaAs tunnel junctions. These plasmons due to the collective oscillation of the conduction electrons were at very low energies ($\simeq 100$ meV). In order to achieve excitation of plasmons of the order of an electron volt or so in this sort of experiment one would have to consider much more intrinsic semiconductors to obtain the electron energies needed without excessive currents.

If we are to calculate the self energy and hence the exchange and correlation potential from (Hedin 1965, Hedin and Lundqvist 1969)

$$\sum (x, x', \omega) = \frac{i}{2\pi} \int \exp (-i\delta\omega') \, G(x, x', \omega - \omega') \, W(x, x', \omega') \, d\omega'$$

where G is the Green function and W the interaction, we must know the poles of the interaction. These poles correspond directly to the plasmon oscillations so that in a two component system there will be poles in the integrand corresponding to the surface excitations of the system. This causes changes in the self energy and in fact for the free metal surface the image potential asymptote comes from the surface plasmon pole in the interaction (Inkson 1971b).

We have examined the surface poles of the interaction and find that there are two basic types of surface plasmons at the metal–semiconductor interface. One is characteristic of a bi-metal and the other of a metal–dielectric interface. The bi-metal type of plasmon lies between the bulk plasmon energies of the two media while the metal–dielectric type always has an energy less than the direct band gap of the semiconductor. The different dispersion relations for the two modes are related to the interaction at the interface as a function of energy. Finally we discuss the suggestion by Phillips (1970) in the light of these new results.

2. Dielectric functions

In order to calculate the screened interaction we need to have the bulk dielectric function of the media on either side of the interface. Unfortunately these are not known in any

great detail so we use model dielectric functions which have the correct long wavelength high energy limits. For the metal we use

$$\epsilon_M(q, \omega) = 1 + \frac{k^2}{q^2 - (\omega^2/\omega_p^2)\, k^2} \tag{1}$$

with k being the Thomas–Fermi wavevector given by

$$k = \left(\frac{6\pi n e^2}{E_F}\right)^{1/2}$$

n is the electron density and E_F the Fermi energy, and ω_p the plasmon energy. This is a simplified version of the dielectric function used by Lundqvist (1967, 1968) for the calculation of the self energy of the electron gas, which gave good agreement with the Lindhard dielectric function. It amounts to letting the plasmon absorption peak exhaust the sum rule (Hedin and Lundqvist 1969, Pines 1963) so that we ignore the single particle excitations which form the only means of plasmon decay at low energies.

For the semiconductor we use

$$\epsilon_s(q, \omega) = 1 + \frac{(\epsilon_0 - 1)}{1 + \epsilon_0\,(q^2/\gamma^2) - (\omega^2/\omega_R^2)\,\epsilon_0} \tag{2}$$

where

$$\gamma^2 = \left(\frac{\epsilon_0}{\epsilon_0 - 1}\right) k_S^2 \tag{3}$$

and

$$\omega_R^2 = \left(\frac{\epsilon_0}{\epsilon_0 - 1}\right) \omega_S^2 \tag{4}$$

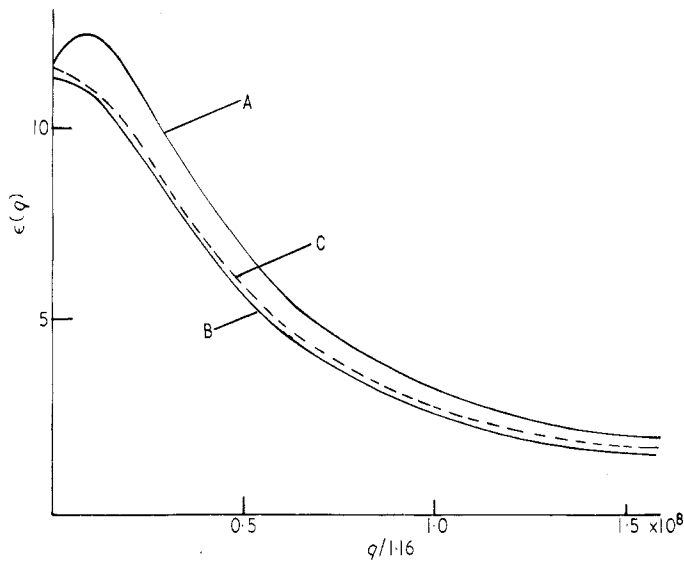k_S, ω_S are the Thomas–Fermi wavevector and plasmon energy of the corresponding

Figure 1. Comparison of theoretical static dielectric functions: A, Srinivasan; B, Walter and Cohen; C, model.

density electron gas and ϵ_0 is the static dielectric constant of the semiconductor. This dielectric function corresponds to the same approximation in the sum rule. In figure 1 we plot the static dielectric functions as calculated by Walter and Cohen (1970) and Srinivasan (1969). We see that in this limit the model is at least as good as the more sophisticated calculations. (Note that this limit can be used to produce a Thomas–Fermi type of equation for the semiconductor (Inkson 1971a).)

In figure 2 we plot the experimental dielectric function variation with energy (Ehrenreich and Phillipp 1963) for two common semiconductors and compare it with the model
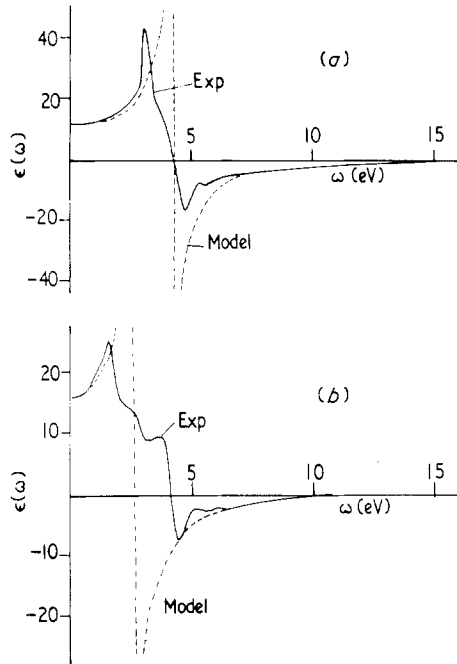

Figure 2. Comparison of the long wavelength model dielectric function with the experimental results of Ehrenreich and Phillipp: (a) silicon, (b) indium antinomide.

dielectric function. The structure around the direct band gap energy is due to the high density of states on either side of the gap. The model function includes this in a simplified form. In fact we will find that as far as the surface plasmons are concerned these lie at energies where the agreement with the experimental function is very good and for the calculation of the self energy we are interested in integrations over the dielectric function. We do not expect the use of these model dielectric functions to be a major limitation of this work. The bulk plasmon dispersion relations for the metal and semiconductor are given by

$$\omega_M(q) = \omega_p(1 + q^2/k^2)^{1/2} \tag{5a}$$

$$\omega_s(q) = \omega_R(1 + q^2/\gamma^2)^{1/2} \tag{5b}$$

3. The screened interaction

The development of the screened interaction has been described previously (Inkson 1971b, referred to as I hereafter) so we will only briefly review it here. When we have an interface the interaction becomes nonlocal, depending upon absolute rather than relative coordinates, and, in addition, new structure appears in the energy dependence corresponding to the new interface excitations. Nonlocality will arise in two ways: firstly we have at the interface an abrupt change in the screening properties; secondly, because of charge redistribution and the new boundary conditions on the wavefunctions at the interface, there are additional changes in the bulk dielectric functions and they also become nonlocal. In our model this second contribution is ignored and we obtain all the nonlocality from the change in the screening properties across the interface. This is reasonable because the second form of nonlocality is brought about by the changes in the density of the electrons near the interface but as we see from the simple Thomas–Fermi equation the bulk screening is not very sensitive to the electron density (ie $k \propto n^{1/\gamma}$). In addition of course this type of nonlocality will be very sensitive to the boundary conditions at the interface; we will have a self consistency problem. Within this model it is thus possible to calculate the interaction between two electrons throughout all space. When they are on the same side it splits naturally into the direct interaction just as in the bulk, and an indirect interaction by way of the induced potential due to the presence of the other medium. We have for an electron at $(a, 0, 0)$ and another at (z, ρ, ϕ) (cylindrical coordinates) in medium 1 (Inkson 1971b)

$$
W(z, \rho; a, \omega) = \frac{4\pi e^2}{(2\pi)^2} \int m \, dl \, dm \left(\frac{\exp(ial) + f_1(a, m, \omega)}{(l^2 + m^2) \epsilon_1 (l^2 + m^2, \omega)} \right) \exp(-izl) J_0(m\rho) \tag{6}
$$

$f_1(a, m, \omega)$ gives the effect of having medium 2 on the other side of the interface; that is, used in equation 6 it gives the induced potential arising from the change in screening properties across the interface. This is discussed in detail in I. (l, m, ψ) are the momentum variables; because of the cylindrical symmetry ϕ and ψ do not appear. The interaction splits into bulk and surface parts.

$$
W_s(z, \rho, a, \omega) = \frac{4\pi e^2}{(2\pi)^2} \int \frac{m \, dl \, dm \, f_1(a, m, \omega) \exp(-izl) J_0(m\rho)}{(l^2 + m^2) \epsilon_1 (l^2 + m^2, \omega)}. \tag{7a}
$$

$$
W_b(z, \rho, a, \omega) = \frac{4\pi e^2}{(2\pi)^2} \int \frac{m \, dl \, dm \exp\{-i(z - a)l\} J_0(mp)}{(l^2 + m^2) \epsilon_1 (l^2 + m^2, \omega)} \tag{7b}
$$

In the bulk case the vanishing of the denominator, that is

$$
\epsilon(l^2 + m^2, \omega) = 0
$$

gives an equation for the bulk plasmon dispersion relation (Pines 1963). A pole in the integrand of the surface part will in the same way describe the interface plasmons. There are two possible ways in which the integrand can explode, by the vanishing of the bulk dielectric constant, which tells us nothing new, or when

$$
f_1(a, m, \omega) \to \infty
$$

We can calculate $f_1(a, m, \omega)$ by the techniques described in I. It involves evaluating integrals like

$$\int_{-\infty}^{+\infty} \frac{l \exp{(ial)}\, \mathrm{d}l}{(l^2 + m^2)\, \epsilon(l^2 + m^2, \omega)}$$

as a function of m and ω. When the energy is high enough so that we have zeros of the dielectric function on the real l axis (physically corresponding to the possibility of bulk plasmon excitation) this integral needs to be further defined. By considering the energy as a complex function the ambiguity is resolved, we find that there are cuts along the real ω axis starting from the energies given by

$$\epsilon(m^2, \omega) = 0$$

and going out to infinity. In the self energy calculation we will see that the contribution of these cuts is quite important. For the case of the two model dielectric functions we obtain for $f_1(a, m, \omega)$

$$f_1(a, m, \omega) = \frac{1}{D(m, \omega)}\left[\exp{(-am)}\left(1 - \frac{\epsilon_0 \omega^2}{\omega_R^2}\right)\left(\frac{im}{M} - 1\right) + (\epsilon_0 - 1)\exp{(iaS)}\right.$$

$$\left. \times \left\{\left(\frac{im}{M} - \frac{im}{S}\right) + \frac{\omega^2}{\omega_p^2}\left(\frac{im}{S} - 1\right)\right\}\right] \tag{8}$$

$$D(m, \omega) = \left(1 - \frac{\omega^2}{\omega_p^2}\right)\left(1 - \epsilon_0\frac{\omega^2}{\omega_R^2} + \frac{i(\epsilon_0 - 1)m}{S}\right) + \left(1 - \frac{\omega^2}{\omega_R^2}\right)\left(\frac{i\epsilon_0 m}{M} - \frac{\epsilon_0 \omega^2}{\omega_p^2}\right)$$

$$\tag{9}$$

where

$$M = i\left\{k^2\left(1 - \frac{\omega^2}{\omega_p^2}\right) + m^2\right\}^{1/2} \qquad \omega^2 < \omega_p^2\left(1 + \frac{m^2}{k^2}\right) \tag{10a}$$

$$= \pm\left\{k^2\left(\frac{\omega^2}{\omega_P^2} - 1\right) - m^2\right\}^{1/2} \qquad \omega^2 > \omega_p^2\left(1 + \frac{m^2}{k^2}\right) \tag{10b}$$

$$S = i\left\{\gamma^2\left(1 - \frac{\omega^2}{\omega_R^2}\right) + m^2\right\}^{1/2} \qquad \omega^2 < \omega_R^2\left(1 + \frac{m^2}{k^2}\right) \tag{11a}$$

$$= \pm\left\{\gamma^2\left(\frac{\omega^2}{\omega_R^2} - 1\right) - m^2\right\}^{1/2} \qquad \omega^2 > \omega_R^2\left(1 + \frac{m^2}{k^2}\right) \tag{11b}$$

The $\pm$ refers to the values on either side of the cut. The analytic structure of $f_1(a, m, \omega)$ is shown schematically in figure 3. In the present paper we are concerned primarily with the surface plasmon poles, that is when

$$D(m, \omega) = 0 \tag{12}$$

Note that this denominator does not change as z or a cross the interface so although we have introduced it by way of $f_1(a, m, \omega)$ we are not restricted in that sense (cf equations 8 in I).

Let us first look at the long wavelength plasmons ($m = 0$). We have from equations 9 and 12

$$\left(1 - \epsilon_0\frac{\omega^2}{\omega_R^2}\right)\left(1 - \frac{\omega^2}{\omega_p^2}\right) - \epsilon_0\frac{\omega^2}{\omega_p^2}\left(1 - \frac{\omega^2}{\omega_R^2}\right) = 0 \tag{13}$$
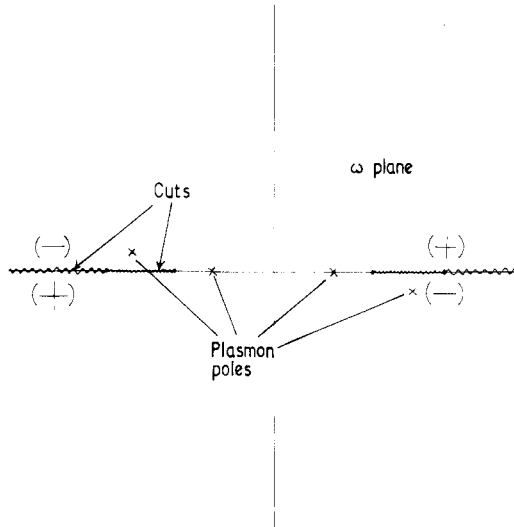
Figure 3. Analytic structure of $f_1(a,m,\omega)$ in the complex energy plane.

There are two useful limits which will tell us the type of plasmon we have

(i) $\epsilon_0 \to \infty$, the metal–metal interface, gives the well known solution

$$\omega^2 = \frac{\omega_p^2 + \omega_R^2}{2}$$

(ii) $\omega_R \to \infty$; this approximates to a metal–classical dielectric interface in this case since then for all energies of interest $\epsilon_s(\omega) \simeq \epsilon_0$ (cf equation 2) and then

$$\omega^2 = \frac{\omega_p^2}{1 + \epsilon_0}$$

The other reference energy is the direct band gap of the semiconductor which is related to the static dielectric constant by the well known relation (Ziman 1965)

$$\epsilon_0 = 1 + \frac{E_g^2}{\omega_S^2} \tag{14}$$

which means that

$$E_g = \frac{\omega_R}{\sqrt{\epsilon_0}} \tag{15}$$

Now the right hand side of equation 13 is positive for zero energy but negative for an energy equal to the direct band gap. This means that no matter what the relative values of the metal and semiconductor plasmon energies there will always be one interface plasmon with an energy less than the direct band gap energy. Such a low energy plasmon should be fairly easily excited and could be important in some tunnelling processes. Figures 4 and 5 show the long wavelength plasmon energies for various values of the static dielectric constant ϵ_0. (ω_S like ω_p is determined by the electron density and will not vary very much between semiconductors, in contrast to ϵ_0.) We see that there

are two interface plasmons at the metal–semiconductor junction: one metal–metal and the other metal–dielectric like. The metal dielectric plasmon stays well within the (direct) band gap energy being approximately $\frac{1}{2}$ to $\frac{2}{3}E_g$. In silicon for instance this amounts to a plasmon of energy between 2 and 3 eV.
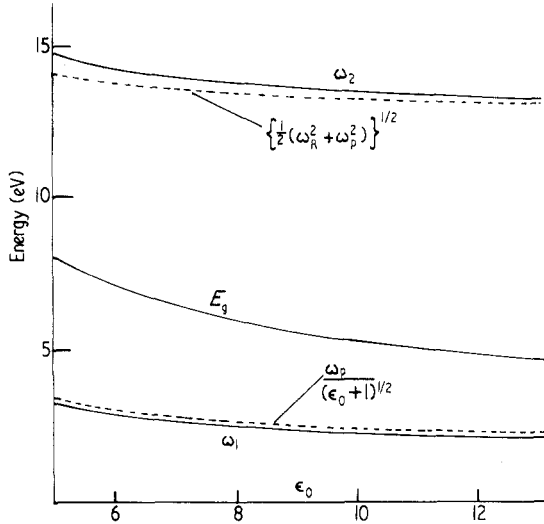

Figure 4. Long wavelength surface plasma energies as a function of the semiconductor dielectric constant. ($\omega_p = 8$ eV, $\omega_s = 16$ eV).
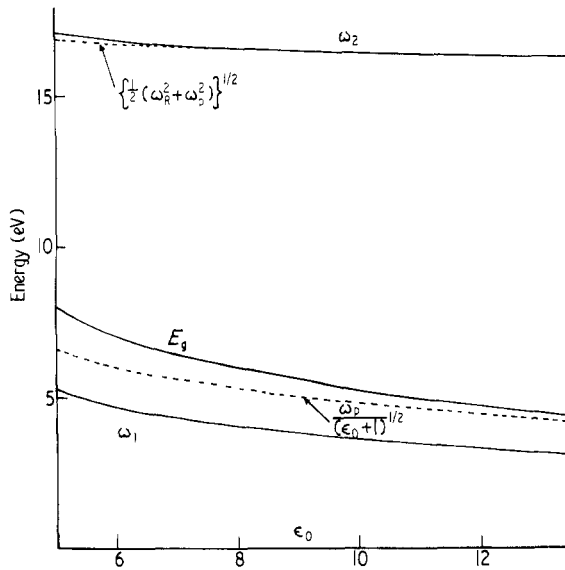

Figure 5. Long wavelength surface plasmon energies as a function of the semiconductor dielectric constant ($\omega_p = 16$ eV, $\omega_s = 16$ eV).

For finite m, that is in order to obtain the dispersion relation, we have to solve equation 12. Even for our model dielectric functions this can not be done analytically
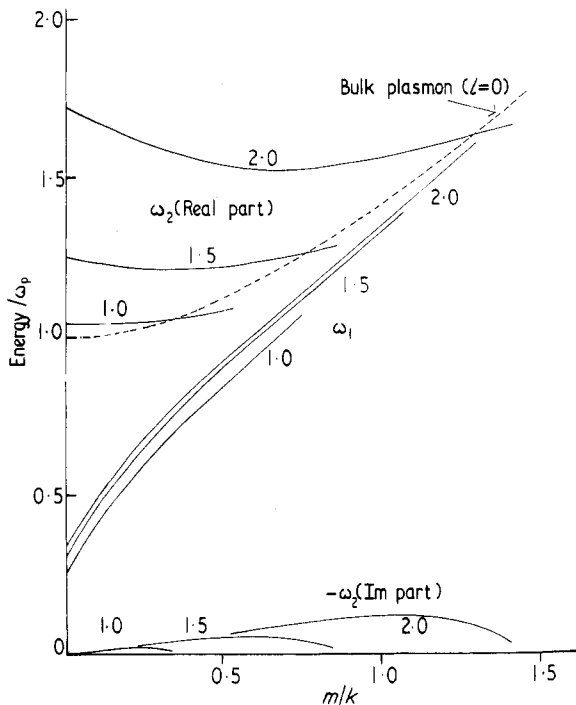
Figure 6. Surface plasmon dispersion relations as a function of the bulk plasmon energies: ω_1, metal–dielectric like plasmon; ω_2, metal–metal like plasmon. Numbers on the dispersion curves correspond to ω_s/ω_p.
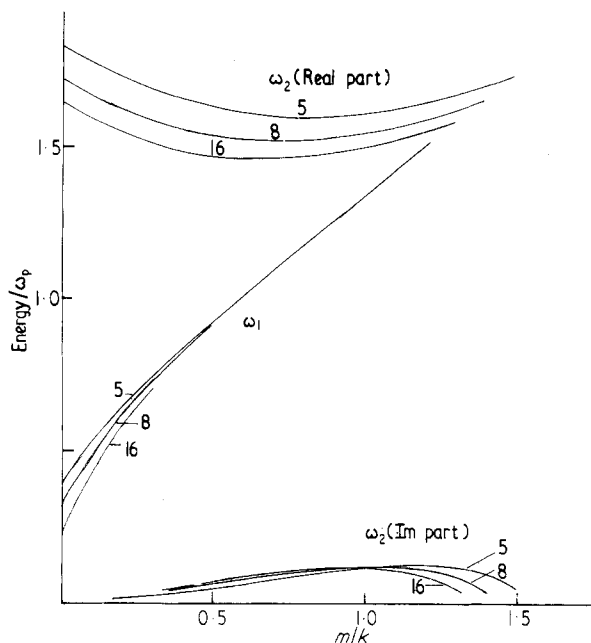

Figure 7. Surface plasmon dispersion relations as a function of the dielectric constant of the semiconductor.

and, since the values of ω may very well be complex, the solution was performed numerically. Starting from the $m = 0$ solutions, m was incremented and the new solutions found using Newton's method in complex algebra taking care as to which side of the cut we were on. This procedure was satisfactory until the solutions came close together when the method broke down. This usually occurred at fairly large m however so was not too restrictive.

Summaries of the solutions obtained are shown in figures 6 and 7. The most striking features are the initial decrease in the metal–metal plasmon energy with increasing m and the appearance of a negative imaginary part to this same plasmon energy showing that the excitation has a finite lifetime. This is maintained for large variations in ϵ_0, k, γ, only vanishing as the two bulk plasmon energies come close together as in figure 6. The finite lifetime must come about through decay into bulk plasmons of the metal since this is the only process available in this energy range (we specifically neglected the single particle excitations in constructing our model dielectric functions). The decay process will take place since the plasmons will be able to couple through the electric field. In figure 6 we have plotted the bulk metal dispersion relation (with the l component
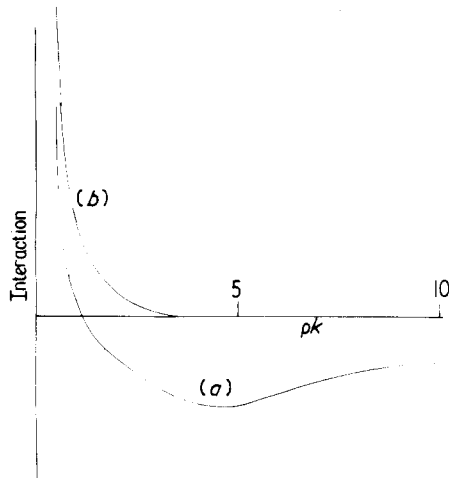

Figure 8. The electron–electron interaction at the interface: (a) below the metal–metal surface plasmon energy; (b) below the metal–dielectric surface plasmon energy

zero). There is certainly some correlation between the movement of the peak of the imaginary part and the position of the bulk and surface plasmons with the same (real) energy and momentum. The effect of a damping mechanism on the plasmon frequency has been discussed by Mendelowitz (1960) for the infinite crystal.

The negative slope in the metal–metal plasmon dispersion relation is interesting. It can not have anything to do with the coupled surface plasmon behaviour in thin films (Economou 1969) since there is only one surface. Bennet (1970) has found that the single surface plasmon dispersion relation for the metal–vacuum system shows a minimum under certain conditions. In this case it is due to a variation in the background and electron charge distributions in the surface region which are absent here.

To investigate the phenomena further we obtain the interaction between electrons at the interface. From equation 6 we obtain

$$W(\rho, \omega) = \frac{4\pi e^2}{(2\pi)^2} \int \frac{m \, dl \, dm(f_1(0, m, \omega) + 1) J_0(m\rho)}{(l^2 + m^2) \epsilon_1(l^2 + m^2, \omega)}$$

where $f_1(0, m, \omega)$ is evaluated at $a = 0^+$.

This simplifies to

$$W(\rho,\omega) = 2e^2 \int \frac{\{(im/M) - (\omega^2/\omega_p^2)\}\{1 - (\omega^2/\omega_R^2)\epsilon_0 + (\epsilon_0 - 1)im/S\} J_0(m\rho) \, dm}{\{(im/M) - \omega^2/\omega_p^2\}\epsilon_0(1 - \omega^2/\omega_R^2) + (1 - \omega^2/\omega_p^2)\{1 - \epsilon_0\omega^2/\omega_R^2 + (\epsilon_0 - 1)(im/s)\}}$$

the real part of which can be evaluated numerically. In figure 8 we show the interaction for energies below the metal–metal and metal–dielectric dispersion relation. We see that the interaction is attractive at the higher energy so that the uniform electron density, that is $m = 0$ configuration, is not stable and finite wavelength density fluctuations will have a lower energy. This would explain the minimum in the higher energy dispersion relation but obviously further work is required on this subject.

The lower metal–dielectric plasmon is by comparison much better behaved. It has a positive slope dispersion relation and an infinite lifetime. If we included some low energy absorption due to single particle excitations, as in Ritchie and Marusak (1966) for the metal–vacuum case, the lifetime would become finite of course.

4. Discussion

We have shown how the screened interaction at the metal–semiconductor interface gives rise to two surface plasmon modes. One is similar to the interface plasmons of a bimetal junction and the other more characteristic of a metal–classical dielectric system. The metal–dielectric plasmon has a very low long wavelength energy always lying well within the direct band gap energy of the semiconductor concerned. With such a low energy the plasmon could be important in tunnelling experiments where excitation of surface plasmons could be expected to cause changes in the current voltage characteristics (Tsui 1969). In this energy range however there are other competing energy loss mechanisms, particularly impact ionization (Crowell 1967), which will complicate the situation. Because of its restricted lifetime the higher energy plasmon will not be such a well defined excitation at finite wavelength but we shall see that it has an important part to play in the calculation of the electron self energy.

Finally we would like to discuss the suggestion of Phillips (1970) as to the part played by stable plasmons in Schottky barriers. His assumptions are that the charge accumulation which pins the Fermi level at the surface of some covalent semiconductors is due to the excitation of surface plasmons and that these will only be stable if the energy of the plasmon is smaller than that required to break a bond in the semiconductor ($\approx$ direct energy gap). These are applied to the case of covalent and ionic semiconductors and he shows that only in covalent semiconductors will the plasmons be stable and hence pinning occur in agreement with experimental evidence (Kurtin *et al* 1969). Our work shows however that the surface plasmon energy is always within the direct band gap energy. Thus on Phillips' criterion we would always have stable surface plasmons and there would be no difference between covalent and ionic semiconductors in Schottky barrier behaviour. He also considers the metal–oxide–semiconductor system but we cannot say anything about this at present.

Acknowledgments

I would like to thank Professor P W Anderson for discussions during the course of this work and the Science Research Council for their financial support.

References

Bennett A J 1970 *Phys. Rev.* B **1**, 203–7
Bohm D and Pines D 1952 *Phys. Rev.* **85** 338–53
—— 1953 *Phys. Rev.* **92** 609–25
Crowell E M 1967 *Solid St. Phys.* (New York: Academic Press) Suppl 9
Economou E N 1969 *Phys. Rev.* **182** 539–54
Ehrenreich H and Phillipp H R 1963 *Phys. Rev.* **129** 1550–60
Fedders P A 1967 *Phys. Rev.* **153** 438–43
Feibleman P J 1968 *Phys. Rev.* **176** 551–5
Harris J and Griffin A 1971 *Phys. Lett.* **34A** 51–2
Hedin L 1965 *Phys. Rev.* **139** A796–823
Hedin L and Lundqvist S 1969 *Solid St. Phys.* (New York: Academic Press) **23** 1–181
Inkson J C 1971a *J. Phys. C: Solid St. Phys.* **4** 591–7
—— 1971b *Surf. Sci.* **28** 69–76
Kurtin S, McGill T C and Mead C A 1969 *Phys. Rev. Lett.* **22** 1433–6
Lundqvist B I 1967 *Phys. kondens. Mater.* **6** 193–205
—— 1968 *Phys. konders. Mater.* **7** 117–23
Mendelowitz 1960 *J. Opt. Sci. Am.* **50** 739–44
Newns D M 1970 *Phys. Rev.* B **1** 3304–22
Phillips J C 1970 *Phys. Rev.* B **1** 593–5
Powell C J and Swan J B 1960 *Phys. Rev.* **118** 640–3
Pines D 1955 *Adv. Solid St. Phys.* (New York: Academic Press) **1** 368–450
—— *Elementary Excitations in Solids* (New York: Benjamin)
Ritchie R H 1957 *Phys. Rev.* **106** 874–81
Ritchie R H and Marusak A L 1966 *Surf. Sci.* **4** 234–40
Srinivasan G 1969 *Phys. Rev.* **178** 1244–51
Stern E A and Ferrell R A 1960 *Phys. Rev.* **120** 130–6
Tsui D C 1969 *Phys. Rev. Lett.* **22** 293–5
Walter J P and Cohen M L 1970 *Phys. Rev.* B **2** 1821–6
Ziman J M 1965 *Principles of the Theory of Solids* (London: Cambridge University Press)