Theory of Surface States

Volker Heine

Bell Telephone Laboratories, Murray Hill, New Jersey

(Received 22 January 1965)

The properties of metal-to-semiconductor junctions and of free semiconductor surfaces are usually explained on the basis of surface states. The theory of the metal contacts is discussed critically, because strictly speaking localized surface states cannot exist in such junctions. However, it is shown that virtual or resonance surface states can exist which behave for practical purposes in the same way. They are really the tails of the metal wave functions rather than separate states. In the past, the length of this tail has often been ignored. Some estimates of its length are made and its consequences pointed out. A semi-quantitative discussion is given of various recent data, including the effect of an oxide layer on barrier height, the variation of barrier height with the metal, the work function of a free surface at high doping, and the effect of a cesium layer on the work function.

1. INTRODUCTION

This paper is in the nature of a commentary and elaboration of the theory of metal-to-semiconductor junctions and of free semiconductor surfaces, with a discussion of some of the more recent data. As is well known, when a contact is made between a metal and a semiconductor, there is a potential barrier $\Phi$. The Fermi level $E_F$ at the surface of a semiconductor lies in the energy gap at a point,$^1$

$$\xi = E_F - \Phi$$

above the top of the valence band [Fig. 2(a)] where $E_g$ is the gap width. The remarkable thing is that $\xi$ is constant to within about 0.2 eV independent of:

(a) the degree of doping$^2,3$ which swings the bands in the bulk material by something of the order of 1 eV;
(b) the metal that is used,$^4,5$ some differing from the semiconductor by several volts in the work function $\phi$;
(c) the crystallographic orientation of the surface$^6,7$;
(d) whether there is a monolayer or so of oxygen or oxide between the metal and semiconductor$^8,9$;
(e) any bias voltage applied across the junction$^6,8$;
(f) whether one dispenses with the metal completely and has a free surface, at least in the case of silicon, again either atomically clean or with an oxide layer.$^{10,8}$

Furthermore, $\xi$ is roughly a constant fraction $\sim 0.3$ of the energy gap in a whole range of group IV and III–V semiconductors.$^5,7$ Of course there are exceptions to all these statements, but by and large this is the story.

All these facts receive a natural explanation in terms of Bardeen's theory,$^1$ namely, that there are surface states on the semiconductor in the band gap. A density of even a fraction of a surface state per surface atom per eV is quite sufficient to absorb any required extra charge into the band of surface states without the Fermi level in this band moving very much. Thus $E_F$ at the surface of the semiconductor is effectively pegged to some fixed value of $\xi$. A few of the data are shown in Fig. 1.

At the time Bardeen's theory was first advanced, the semiconductor-metal contacts were made by pressing the metal against the semiconductor, each coated with an oxide film. Now, however, many of the junctions are made by cleaving the semiconductor in a stream of the metal vapor used to deposit the metal. The effective partial pressure of the condensing metal is ten or more times the residual partial pressure of gas$^7,8$ so that there is intimate atomic contact between the metal and semiconductor without any intervening oxygen or other impurity layer. Under such conditions it appears at first

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig1.png}
\caption{Position of Fermi level at the surface ($\xi$) above the valence-band edge for metal-to-silicon contacts. $E_g$ is the value for a free-silicon surface. $\xi$ is given by $E_g - \Phi$, where $\Phi$ is the measured barrier height. White circles refer to intimate contact between metal and semiconductor, black circles to the case of a thin intervening oxide layer. Circles half-black are used where the two values coincide within experimental error. The data points second from the left are for magnesium. (After Ref. 8. Data from Refs. 2–8.)}
\end{figure}

A 1689

© Editorial Jaca Book spa, Milano 1990
sight unreasonable to talk about surface states at all. For an energy $E$ below $E_F$, the solutions of the Schrödinger equation will decay exponentially in the semiconductor but propagate as Bloch states on the metal side of the junction to form the ordinary volume states of the metal.\(^{11,12}\) This follows from simple considerations of matching the wave functions at the boundary as shown in Fig. 2. Let us assume the $x$ axis perpendicular to the surface. For some value of $k=k_{\parallel}$ parallel to the surface, e.g., $k_y=k_z=0$, we have the bands shown in Fig. 2(b). At energy $E$, the exponential solution in the semiconductor can always be joined onto the two Bloch states with wave vector $k_1 = k_{\parallel}, -k_1$, in the metal, the existence of the two $k$'s ensuring that both $\psi$ and its derivative can be matched at the boundary. Thus for energies in the semiconductor band gap the volume states of the metal all have tails in the semiconductor; there are actually no surface states.

Our first task is therefore to discuss in Sec. II how the Bardeen explanation in terms of surface states may be rescued. Although the exact form of the band of surface states is not known, there are some semiquantitative statements one can make. For example, the form of the exponential tail in the semiconductor is determined by the ordinary volume band structure. In Sec. III we make some estimates of its length and discuss its relation to some of the experimental data.

II. MATCHING THE WAVE FUNCTION

Let us first assume that $k_{\parallel}$, the component of $k$ parallel to the boundary plane, is conserved as an electron crosses the junction. In reality there is sufficient disorder at the boundary for diffuse scattering to be probably a better approximation than the specular assumption made here. But the effect of scattering is to scramble states of the same energy $E$, leaving invariant any total property such as the density of states or the charge distribution $\psi^*\psi$ summed over all states in some energy interval $dE$. Thus the device of considering states of definite $k_{\parallel}$ leads to correct results for total or average properties.

As a first step we treat the valence states of the semiconductor and the conduction electrons of the metal as free electron gases, which is a good approximation in view of the small band gap.\(^{11,12,13}\) Because of the higher density of four electrons per atom in the semiconductor than the metal, the potential and Fermi surfaces in the two regions are as shown in Fig. 3. The important point is that in this simple model all semiconductor states with $k_{\parallel}$ greater than $A' A''$ do not penetrate into the metal. Their $k_{\parallel}$ is sufficiently large that their perpendicular kinetic energy

$$E_1 = E - \frac{\hbar^2 k_{\parallel}^2}{2m} \quad (2)$$

is negative in the metal, i.e., they decay exponentially on the metal side of the boundary. The value of the Fermi momentum $k_F$ for silicon is compared with that of some typical metals in Table I. We have not counted the $d$ electrons in the case of transition metals because they have such different wave functions with small value at the cell edge that one does not expect them to match at all well onto the silicon wave functions. Let us concentrate for the present on the top part of the table. We note that a large fraction $f$ of the high-energy states of the silicon, given by

$$f = 1 - k_{\parallel,m}/k_{\parallel,sc} \quad (3)$$

do not propagate into the metal. The same applies to any surface states in the band gap formed when the

\begin{table}[h]
\centering
\caption{Fermi momenta $k_F$ of metals (m) compared with silicon (sc).}
\begin{tabular}{ll}
\hline
 & $k_{\parallel,m}/k_{\parallel,sc}$ \\
\hline
sodium & 0.50 \\
palladium & 0.55 \\
calcium & 0.61 \\
silver & 0.66 \\
magnesium & 0.75 \\
antimony & 0.93 \\
aluminum & 0.97 \\
\hline
\end{tabular}
\end{table}
