

Work Function, Photoelectric Threshold, and Surface States of Atomically Clean Silicon

F. G. ALLEN AND G. W. GOBELI
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received February 23, 1962)

The position of the Fermi level at the surface has been determined over the entire available range of bulk dopings for atomically clean (111) silicon surfaces cleaved in a vacuum of 10^{-10} mm Hg. Surface potential is found from the difference in work function and photoelectric threshold. The latter equals the energy difference between the vacuum level and the top of the valance band only for a range of high resistivities where the bands are flat to the surface over the escape depth of emitted electrons (about 25 Å). This photoelectric threshold is 5.15 ± 0.08 eV. The work function varies only from 4.9 to 4.7 eV in going from extreme *p* to extreme *n* type, being close to 4.83 eV throughout most of the doping range. The flat band or neutral surface condition occurs for high-resistivity ($\sim 500 \Omega\text{-cm}$) *p*-type bulk doping, for which the Fermi level lies ~ 0.23 eV below the mid-gap. The results indicate a surface-state density approximately equal to the surface atom density of $8 \times 10^{14}/\text{cm}^2$ with the states located in two groups, one normally empty just below the center of the gap and another normally filled just above the valence-band edge.

INTRODUCTION

THE question of how the potential at the surface of a semiconductor varies as the energy levels in the interior are moved with respect to the Fermi level has often been considered. To explain early experimental results^{1,2} showing that the surface potential moved much less rapidly than the bulk potential and was insensitive to metal contacts Bardeen³ pointed out that surface states of sufficient density would control surface potential. While there have been many theoretical predictions on surface states,⁴ due to the mathematical complexity of describing any real surface, even an atomically flat and clean one, little can be concluded from the theory except that there are likely to be as many surface states in the forbidden gap as surface atoms. Experiments with germanium and silicon surfaces in high vacuum cleaned by heating or sputtering have attempted to measure various properties of surface states. Thus, external potential measurements⁵⁻⁷ examine their ability to determine surface potential when bulk potential is changed by temperature and doping, field-effect experiments⁸⁻¹⁰ measure their screening effect on an external field; channel conductance and change in contact potential or photoconductivity with illumination experiments measure their ability to act as recombination centers. While some of these results may still be open to reinterpretation, and no one set of experiments has been able to control the parameters

well, they have generally agreed that surface state densities must be within an order of magnitude of that of the surface atoms. Recent results on cleaved germanium¹¹ and silicon¹² have again indicated high surface-state densities, though with a somewhat different neutral surface level for silicon from that found here.

Field-effect experiments on "real" or gas-covered surfaces, on the other hand, have indicated fast surface-state densities on silicon¹³ and germanium¹⁴ to be about three to four orders of magnitude below the surface atom density, respectively. Theoretical predictions for such surfaces are, of course, even more difficult than for clean ones.

PRINCIPLE OF THE EXPERIMENT

The present experiment was designed to provide a system of reproducible silicon surfaces in ultrahigh vacuum ($< 10^{-10}$ mm Hg) that were atomically clean, nearly atomically smooth, and for which chemical doping was constant up to and including the surface. The purpose was then to vary this doping over the complete range available from degenerate *n* to degenerate *p* type—one surface for each doping desired—and to determine both surface and bulk potential for each sample. From these results information on the surface states present at the clean surface is then derived.

The true work function, ϕ , is the difference in energy of an electron at the Fermi level just inside the surface and at rest just outside the image potential barrier in the vacuum [see Fig. 1(a)]. The value of ϕ for each silicon surface is determined using the Kelvin contact potential difference (C. P. D.) method against a metal reference surface. The work function of the reference surface is determined by a photoelectric measurement,

¹ W. E. Meyerhof, Phys. Rev. **71**, 727 (1947).

² W. H. Brattain and W. Shockley, Phys. Rev. **72**, 345 (1947).

³ J. Bardeen, Phys. Rev. **71**, 717 (1947).

⁴ See, for example, I. Tamm, Physik Z. Sowjetunion **1**, 733 (1932); W. Shockley, Phys. Rev. **56**, 317 (1939); J. Koutecký and M. Tomásek, *ibid.* **120**, 1212 (1960).

⁵ A. H. Smith, Phys. Rev. **75**, 953 (1949).

⁶ F. G. Allen and A. B. Fowler, J. Phys. Chem. Solids **3**, 107 (1957).

⁷ A. B. Fowler, J. Appl. Phys. **30**, 556 (1959); R. E. Simon, Bull. Am. Phys. Soc. **4**, 410 (1959).

⁸ J. T. Law and C. G. B. Garrett, J. Appl. Phys. **27**, 656 (1956).

⁹ P. Handler, *Semiconductor Surface Physics* (University of Pennsylvania Press, Philadelphia, Pennsylvania, 1957).

¹⁰ J. T. Law, J. Phys. Chem. Solids **14**, 9 (1960).

¹¹ D. R. Palmer *et al.*, J. Phys. Chem. Solids **14**, 27 (1960).

¹² D. R. Palmer, S. R. Morrison, and C. E. Dauenbaugh, Phys. Rev. Letters **6**, 170 (1961).

¹³ Litovchenko and Snitko, Fizika Tverdogo Tela U.S.S.R. [translation Soviet Phys.—Solid State **2**, 554 (1960)].

¹⁴ H. C. Montgomery and W. L. Brown, Phys. Rev. **103**, 865 (1956).

using the fact that the photoelectric threshold of a clean metal equals its work function. To locate the electronic energy levels of the semiconductor surface with respect to the Fermi level, the photoelectric threshold of the semiconductor is then measured. This threshold, Φ , is the lowest photon energy able to excite electrons from filled levels over the surface barrier. In general, photoemission from a semiconductor could take place from the conduction band, surface states, or the valence band.

For the simple case shown in Fig. 1(a)—flat bands and no surface states—emission from the valence band dominates and we have $\Phi = \chi + E_G$. For such a model, assuming that the electron affinity χ and band gap E_G are independent of doping, we see that as the Fermi level is moved across the band gap by bulk doping, the measured value of φ should change as rapidly as $(E_F - E_I)_{\text{Bulk}}$ as shown in Fig. 1(b). The measured photoelectric threshold would be constant and equal $\chi + E_G$ until the Fermi level came within a few kT of the band edges. It would then become equal to the work function for degenerate n - and p -type samples, since the highest filled states would be at the Fermi level in either case.

The actual case differs from this simple model principally in the presence of surface states in the forbidden gap and, to a small degree, in possible variations in χ and E_G with doping. Surface states will, in general, prevent φ from varying as rapidly as $(E_F - E_I)_B$ as a result of which the energy bands must bend just beneath the surface. Such bending, as shown below, causes the measured or *apparent* photoelectric threshold to differ from $\chi + E_G$. The measured photoelectric threshold will still equal $\chi + E_G$, however, on samples doped to give negligible band bending over the escape depth of emitted electrons provided that on such samples E_F lies well inside the band edges and that emission from the valence band rather than from surface states dominates the yield near threshold. Once measured, $\Phi = \chi + E_G$ will remain constant for all other bulk dopings except for small variations in χ and E_G to be expected at heavy dopings. Using this Φ and the measured value of φ for

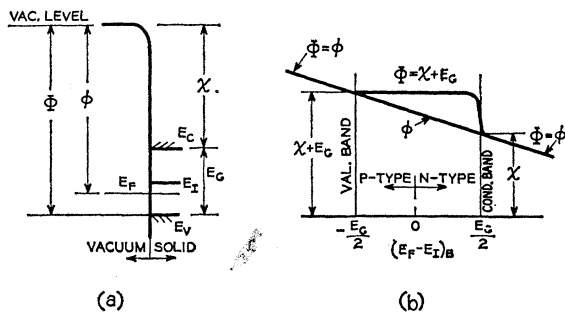


FIG. 1. Predicted semiconductor surface behavior for the case of no surface states. (a) Energy levels at the surface. (b) Variation of work function, φ , and photoelectric threshold, Φ , with bulk doping. E_F =Fermi level, E_I =intrinsic level, E_G =bottom of conduction band, E_V =top of valence band.

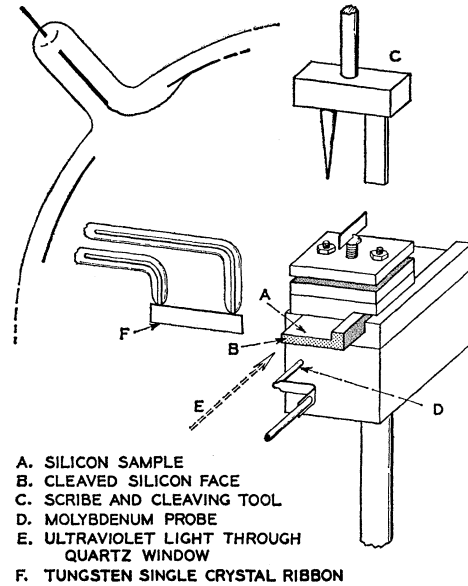


FIG. 2. Arrangement of elements in the experimental high-vacuum tube.

each sample, the surface potential is determined by the relation

$$(E_F - E_I)_S = (E_V - E_I) + (\Phi - \varphi). \quad (1)$$

The difference between surface and bulk potential determines the bending of the bands, and hence the charge in the space-charge layer, Q_{SC} . Since this is balanced at a free surface by an equal and opposite surface-state charge, $Q_{SC} = -Q_{SS}$, the latter is now known over the range of surface potential values which can be explored by the maximum possible variation in bulk doping. Finally, assuming that the surface states are not seriously altered by bulk doping, considerable information can be obtained about their density and distribution since these must satisfy the Q_{SS} vs $(E_F - E_I)_S$ relation mentioned above.

EXPERIMENTAL

To avoid the surface and bulk resistivity changes that are difficult to avoid with any cleaning treatment involving heating of the sample,¹⁵ a cleavage technique was used to expose fresh surfaces for each measurement. Hence, only the (111) surface—the cleavage plane—was studied.

The arrangement of elements in the experimental bakeable high-vacuum tube is indicated in Fig. 2. The silicon sample A having an L-shaped cross section to promote good cleavage¹⁶ is shown clamped to the supporting block. All elements of the tube can be manipulated from outside the tube through syphon bellows.

¹⁵ F. G. Allen, T. M. Buck and J. T. Law, J. Appl. Phys. **31**, 979 (1960).

¹⁶ G. W. Gobeli and F. G. Allen, J. Phys. Chem. Solids **14**, 23 (1960).

The tool C is used to unclamp the sample, advance it, clamp it, scribe a nick in its smooth upper surface, and then cleave it. This can be repeated up to ten or fifteen consecutive times on one sample in the same vacuum. Work function measurements are made by vibrating the polycrystalline 0.040-in.-diam molybdenum probe D extremely close first to the portion of the silicon cleaved face of interest and then close to the single-crystal tungsten ribbon F. The dc potential of the probe is adjusted for an ac null from sample to ground. By subtracting the two C. P. D. values found, the difference in the areal average work function of clean tungsten and silicon, measured over a 0.040-in.-diam spot, is found and the work function of the molybdenum probe need not be known. Actually, the latter was purposely not cleaned to ensure greater stability of its potential.

Null accuracy was ± 0.005 V and the variation in C. P. D. values over the surfaces was $\sim \pm 0.020$ V, although on some good cleavage faces readings were everywhere constant within a few millivolts. On some rough cleavages this variation became as high as ± 0.040 V with rougher surfaces giving higher work functions. The latter result is in accord with previous measurements that indicate that $\phi(111)$ for both silicon and germanium is lower than for other principal faces.^{6,17,18}

Photoelectric thresholds and photoelectric yield of both silicon and tungsten are measured by directing a well-focused 1×2 mm spot of monochromatic ultraviolet light on the region of interest and measuring the saturated emission current emitted by the surface, when all other surfaces in the tube are positively biased as collectors. (The sample is insulated from its support block and clamps.) Reproducibility of the photoelectric threshold for the tungsten ribbon, determined as the intercept on a (yield)^{1/2} vs $h\nu$ plot, was $\sim \pm 0.020$ V. The value of this threshold for several freshly flashed ribbons, which had close to (113) surfaces, varied from 4.6 to 4.8 V, depending on the ribbon and its flashing history. Work function values of silicon depend directly on this threshold and on two C. P. D. readings and are probably reliable in absolute value to ± 0.07 V. Differences in work function across a p - n junction or between cleavages on the same sample are accurate to ± 0.02 V. Reliability of the silicon photoelectric threshold, discussed below; was ± 0.08 V.

Further details on the cleavage technique and on the quality of the surfaces produced have been published elsewhere.¹⁶ The smooth portion of the cleaved surface was large enough to contain either the molybdenum probe or the uv light spot on either the n or p half (see below), and thus spurious results connected with the dirty or rough surfaces were avoided.

Two types of samples were used to cover the entire doping range. First, samples were cleaved with a p - n

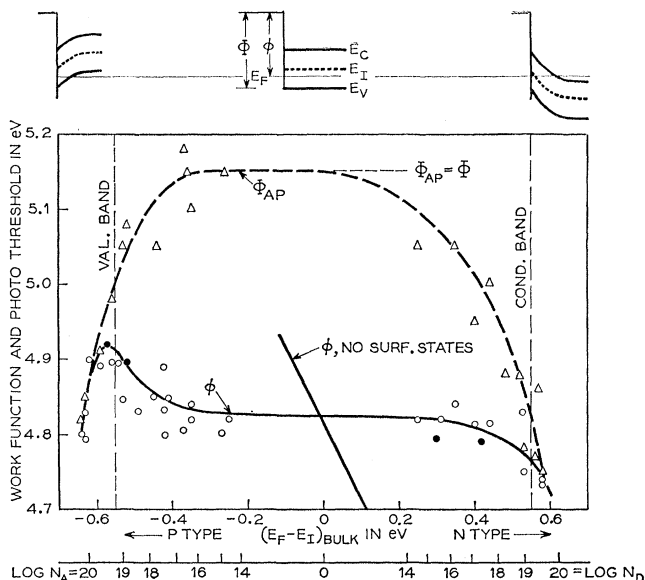


FIG. 3. Work function, ϕ , and apparent photoelectric threshold, Φ_{AP} , vs bulk doping, $(E_F - E_I)_B$ for cleaved Si (111) faces. Energy level diagrams at the surface for degenerate p type, n type and for the flat-band condition are shown at the appropriate $(E_F - E_I)_B$ values. The predicted work function variation for the case of no surface states is indicated. The logarithm of the acceptor and donor concentration, N_A and N_D per cm^3 , is shown at the bottom.

junction running down their length, so that the cleavage exposed both a p - and n -type surface side by side. This method gave the most reliable results for direct comparison of photoelectric yield and work function of n - vs p -type. Three or four repeated cleavages on the same sample allowed a multiple check on all measurements being made for the same two resistivities. Second, long samples of a single conductivity type were graded in resistivity from high purity to degenerate along their length. These permitted a good determination of the progressive change with doping of photoelectric yield characteristics and work function. Boron and arsenic were used for p - and n -type samples throughout. Resistivities from $0.002 \Omega \text{ cm}$ n to $0.0015 \Omega \text{ cm}$ p type were covered. Crystals were grown both by drawing from the melt and by pyrolytic deposition. High degrees of compensation were avoided. There was no evidence of difference in work function or photoelectric yield characteristics for samples from different sources but of the same doping.

The resistivity and, hence, bulk potential of each sample was measured after the completion of the run by four-point probe measurements on a slice 0.040×0.040 in. in cross section cut from immediately behind each cleavage face with all surfaces lapped. Hence, there was no heating of the samples between the run and the resistivity measurement.

The effect of oxygen on the cleaved surfaces was determined by admitting it to the system from AIRCO flasks through Granville-Phillips bakeable metal valves. Photoelectric threshold and C. P. D. were then moni-

¹⁷ F. G. Allen, J. Phys. Chem. Solids 8, 119 (1959).

¹⁸ J. A. Dillon and H. E. Farnsworth, J. Appl. Phys. 29, 1195 (1958).

tored as a function of time and oxygen pressure. Both quantities rose by ~ 0.5 eV and then leveled off after a few minutes at $\sim 10^{-6}$ mm Hg of O_2 pressure. Results were not accurate enough to derive information on the difference in these two quantities during O_2 exposure, which would give information on changes in surface states. The principal effect was a large increase in the electron affinity, χ .

Change in C. P. D. of the silicon, due to illumination with both visible and ultraviolet light, was looked for on most samples but never found. Since bulk lifetime was very low in most cases due to the high doping levels used, this result is not surprising and does not permit conclusions to be drawn on the surface recombination velocity at the cleaved surface. It does indicate that no corrections for this effect were needed when measuring photoelectric thresholds.

Tentative results using an intense light source focused on the junction of a high-lifetime sample indicated that junction photoelectric voltages could be generated both before and after cleavage. This implies that the surface recombination velocity at the cleaved surface is not excessive, and also that the junctions are not short circuited by a highly conducting surface layer.

Direct evidence that the n and p sides of the junctions were actually n -type and p -type material at the surface was found in one case where the molybdenum C. P. D. probe was heated slightly, then brought into contact with both n and p sides of a freshly cleaved junction. Thermoelectric voltages of the correct sign were indicated on a galvanometer for each side. This result does not contradict the finding that the free surface is weakly p type on both sides of a junction due to surface states and a resulting space-charge layer, since the thermoelectric voltage of the material directly beneath the surface can presumably over-ride such a surface layer.

Throughout all observations on these cleaved surfaces no slow drifting of potential, photoelectric yield, or photo-voltage at the cleaved junction mentioned above was observed, such as would result if slow traps were present on the surface. It is therefore believed that the surface states are entirely of the "fast" type.

RESULTS

1. Photoelectric Threshold of Silicon

If (yield) $^{1/2}$ is plotted vs $h\nu$ for all samples, a rough extrapolation can be made to zero yield and the $h\nu$ intercept will be called the apparent photoelectric threshold, Φ_{AP} . (The actual extrapolation can be made more accurately once the form of Y vs $h\nu$ is established which, as discussed elsewhere,¹⁹ depends strongly on the doping of the silicon sample.) If this apparent photoelectric threshold is now plotted for all samples against the bulk potential, $(E_F - E_T)_B$, the upper curve shown in Fig. 3 results. The work function measurements on

the same samples are shown in the lower curve and the energy diagrams at the surface are indicated at three dopings at the top. It is immediately seen that Φ_{AP} depends strongly on doping, and that throughout the high-resistivity range its values is ~ 0.3 V higher than the work function. It drops down and becomes equal to the work function at both degenerate n and p extremes, as predicted for the simple model of Fig. 1(b). The question is then what value of this apparent photoelectric threshold, if any, is also equal to the true photoelectric threshold as defined in Fig. 1(a).

The requirements that we measure the true photoelectric threshold are that (1) E_F lie several kT units away from the band edges at the surface, (2) the valence band near the surface must not bend appreciably over the escape depth of emitted electrons, and (3) emission from the valence band must dominate the yield near threshold. It can be shown that all three requirements are met in the high-resistivity range where Φ_{AP} reaches its maximum. (1) must be satisfied here because $\Phi_{AP} > \varphi$, yet if E_F were close to either band edge at the surface we should have $\Phi_{AP} \cong \varphi$, since there would then be an appreciable density of filled levels just below E_F . (2) is satisfied for near intrinsic silicon, since there the space-charge potential changes by only $kT = 0.025$ eV in a Debye length of $L = (KkT/2\pi n_i e^2)^{1/2} \cong 2 \times 10^{-3}$ cm, whereas, as will be shown later, the escape depth for emitted electrons is only of the order of 25 Å. (Actual space-charge calculations¹⁹ show that band bending should not start to affect Φ_{AP} until $|E_F - E_T|_B > 0.2$ eV. Finally, that (3) is satisfied is seen by the following arguments: the drop in Φ_{AP} going to the left of the maximum in Fig. 3 can be accurately explained as due to a progressive bending down of the valence band to meet the surface for more strongly p -type doping. This gives rise to a lower threshold for electrons beneath the surface and thus a higher yield, as further discussed elsewhere.¹⁹ If yield were dominated by emission from surface states below the Fermi level at the surface, the progressive raising of the levels at the surface relative to E_F , due to more strongly p -type doping, could only cause the apparent threshold to remain constant or increase.

We conclude, then, that the apparent photoelectric threshold becomes equal to the true photoelectric threshold near the center of the diagram where it flattens out at its maximum. The value at this maximum can now be more accurately determined in light of the theory^{19,20} by fitting the yield curves near threshold to a 3 or $\frac{5}{2}$ power law in $(h\nu - \Phi)$ for extrapolation to zero. Note that Φ corresponds to the indirect or lower threshold in reference 19. Such fits made on several of the high-resistivity samples give a value of $\Phi = 5.15 \pm 0.08$ eV.

The fact that Φ_{AP} approaches φ at the p extreme of the diagram can be explained by the fact that the bands bend downward to the surface so sharply that the top

¹⁹ G. W. Gobeli and F. G. Allen, Phys. Rev. **127**, 141 (1962).

²⁰ E. O. Kane, Phys. Rev. **127**, 131 (1962).

of the valence band crosses the Fermi level within an electron's escape depth of the surface. This makes filled levels just beneath the Fermi level available for excitation into the vacuum. Similarly, at the n -type extreme, the conduction band crosses the Fermi level within an escape depth of the surface, again making $\Phi_{AP} = \varphi$.

The gradual drop in Φ_{AP} going from intrinsic toward n type still requires explanation. Since the drop begins while the Fermi level is still several tenths of a volt below the conduction band, it seems likely that emission from surface states rather than the conduction band is the explanation. Such states, normally lying above E_F , are progressively filled as the bands are drawn downward relative to E_F at the surface in going to more n -type bulk, so that their yield should rise, lowering the apparent threshold, just as is observed. It is possible that some emission from surface states is present even in the high-resistivity range where Φ is determined, though analysis of the data indicate such an effect is probably negligible.¹⁹ If so, the measured value of Φ would be somewhat too low. At worst, the value of 5.15 might be increased to something less than 5.4 V, since the direct threshold there certainly originates in the valence band. Such an error would merely result in shifting the neutral surface level (see below) closer to the center of the gap.

In conclusion, the true photoelectric threshold $\Phi = \chi + E_G$ can be measured for silicon at high resistivities, its value is 5.15 ± 0.08 eV, and it can now be assumed constant for all other resistivities except for changes in χ at extreme doping, which are discussed in the Appendix.

2. Work Function

The lower curve on Fig. 3 represents a weighted average of all work-function measurements made. The dark circled points are data taken directly across p - n junctions and, hence, are given the most weight in determining the relative value of φ for highly p vs highly n type. The absolute value of φ was not measured for these points; they are located on the φ scale so that their average matches the average of absolute φ measurements. The other (open circle) data are absolute φ measurements on graded n - or p -type samples and are used to determine the shape of the curve on either the n or p side, as well as the average absolute value. While data at high resistivities are scarce, there is no reason to expect that φ values should depart significantly from the smooth curve shown.

The averaged curve shown in Fig. 3 indicates that as in previous work, the surface potential of a clean semiconductor is very insensitive to bulk doping. While the value of $(E_F - E_I)_B$ varies by 1.2 V in going from degenerate n to p type, φ varies only by about 0.2 V going from ~ 4.7 to 4.9 V. This is in marked contrast to the results predicted in Fig. 1(b) for no surface states and indicated again on Fig. 3, where φ would have changed by 1.2 V. Since from Eq. (1) $\Delta\varphi =$

$-\Delta(E_F - E_I)_S$, the actual shape of the φ vs $(E_F - E_I)_B$ curve must be attributed to properties of the surface states as they resist the tendency of the bulk potential to determine surface potential. At extreme dopings where χ can no longer be assumed constant (see Appendix), the shape of the φ vs $(E_F - E_I)_B$ curve will include this variation in χ , in addition to changes in $(E_F - E_I)_S$. The apparent drop in φ at extreme p type samples in Fig. 3 must, if real, be such an effect, since no real distribution of surface states could cause $(E_F - E_I)_S$ to vary in the opposite direction from $(E_F - E_I)_B$.

DERIVATION OF SURFACE-STATE DENSITY AND DISTRIBUTION

Once the true photoelectric threshold $\Phi = \chi + E_G$, as well as the work function φ are known, the surface potential $(E_F - E_I)_S$ is determined from Eq. (1) above. The bulk potential $(E_F - E_I)_B$ is determined for each surface from its resistivity, assuming no compensation, using impurity concentration vs resistivity data²¹ and a machine solution of the generalized charge density Eq. (16) of reference 19, setting $\rho = 0$. With surface potential, bulk potential, and impurity concentration known, the charge in the space-charge region Q_{SC} is then computed from Eq. (15) of reference 19, using the relation

$$Q_{SC} = -\frac{K}{4\pi} \frac{kT}{e} \left(\frac{du}{dx} \right)_{x=0}, \quad (2)$$

where $u = (E_F - E_I)/kT$ and K is the dielectric constant. Since at a free surface the charge in the space-charge layer is balanced by an equal and opposite charge in the surface states, Q_{SS} , we have $Q_{SS} = -Q_{SC}$, and hence, a plot can be made for all samples of $Q_{SS}(u_S, u_B)$ vs u_S , each sample giving one point on the curve.²² The results are shown in Fig. 4, where surface-state charge, Q_{SS}/e , in electronic units is plotted vertically and surface potential, u_S , in kT units horizontally. The curve is S-shaped and surface-state charges of plus and minus $\sim 6 \times 10^{13} \text{ cm}^{-2}$ are found at extreme p - and n -type bulk dopings, respectively. The reverse curve shown at extreme p -type doping could not actually occur and is thus attributed to variation in χ . The flat-band condition of the surface, where $Q_{SS} = 0$, occurs at $u_S \cong -9kT$ or at ~ 0.22 eV below the center of the gap. As discussed above, since absolute values of u_S depend upon $\Phi - \varphi$ the uncertainty in the u_S position of the entire curve is $\sim \pm 0.1$ eV. The uncertainty in the shape of the curve is $\sim \pm 0.025$ eV.

The density and energy distribution of surface states is probably not seriously altered by bulk doping (see Appendix), so that information on both may be derived

²¹ J. C. Irvin, Bell System Tech. J. (to be published).

²² For ease in studying the whole system a family of curves of Q_{SC} vs u_S was computed with the IBM 7090, one for each u_B value, as suggested by Fowler (see reference 7).

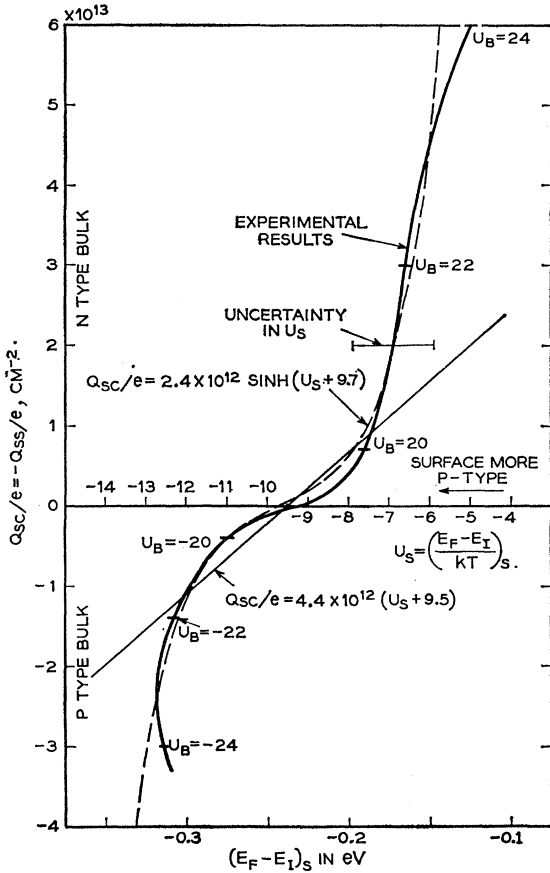


FIG. 4. Surface-state charge, Q_{SS}/e in electronic units/cm², vs surface potential, u_s , for cleaved Si (111) faces. The heavy curve shows the experimental results and the other curves show the fits given by a linear and a sinh charge vs potential relation.

from the Q_{SS} vs u_s plot. Wolff²³ has shown that the determination can be exact at any temperature, if Q_{SS} vs u_s is known over the entire range of u_s of interest, through the solution of an integral equation. For a smaller known range of u_s the determination of $n(u_s)$ will not be unique.

For the present data, we discuss the fit to the data given by two assumed surface charge-surface potential relations. The first of these is the linear one, $Q_{SS}/e = 4.4 \times 10^{12} (u_s + 9.5)$, shown as a straight line on Fig. 4. Since all states below E_F at the surface will be filled, all above empty (except for small corrections at $T > 0$ near E_F), the surface-state distribution yielding such a relation will be continuous and of constant density. It must extend through the neutral level out to at least ± 0.11 eV. We assume throughout that each surface state can be filled by only one electron so that each state filled contributes one negative electronic charge to Q_{SS} . Hence, the required density is 1.76×10^{14} states/(cm²-eV), with a neutral level at $-9.5 kT$. It is seen on Fig. 4 that at best only a poor fit can be made to the central portion

²³ P. A. Wolff (unpublished).

of the curve by this constant distribution, and that it falls well outside the experimental uncertainty at higher dopings.

A much better fit over the entire data range is given is given by the relation, also shown on Fig. 4,

$$Q_{SS}/e = -2.44 \times 10^{12} \sinh(u_s + 9.7). \quad (3)$$

Note that no arbitrary constant is needed in the argument of the hyperbolic sine function here, unlike the case reported for real surfaces.¹⁴ In judging the fit between this curve and the data, it should be noted that the exact match depends on the details of the ϕ vs u_B data curve of Fig. 3. The general S-shaped feature of Fig. 4 remains unchanged, however, for any reasonable curve through the ϕ data of Fig. 3, including a straight line through the dark circled points.

The hyperbolic sine relation between Q_{SS} and u_s of (3) can be provided by at least three different surface-state distributions: (1) two bands of surface states, one above and the other below the neutral level, E_0 ; (2) two discrete levels, one above and one below E_0 ; and (3) a hyperbolic cosine distribution of states centered at E_0 .

Considering (1) first, just as theory predicts a state density proportional to $(E - E_E)^{1/2}$ near a band edge E_E in three dimensions, it predicts a constant density near the edge of a surface band in two dimensions. Hence, for the case of two surface bands, assume a constant density n_1 states/(cm²-eV) for the band above and a density n_2 for the band below the neutral level, E_0 , see Fig. 5(a). Let the band edges be E_1 and E_2 , respectively; assume for the moment that the extent of each band is many kT units, and assume that each state holds only one electron with a degeneracy factor in the Fermi occupation probability of unity. Then the net surface state charge will be

$$Q_{SS}/e = n_2 \int_{-\infty}^{E_2} \frac{dE}{1 + e^{(E_F - E)/kT}} - n_1 \int_{E_1}^{\infty} \frac{dE}{1 + e^{(E_F - E)/kT}}. \quad (4)$$

Assume now that the Fermi level remains several kT inside E_1 and E_2 , justifying Boltzmann statistics. Using $\Delta E_1 = E_1 - E_0$, $\Delta E_2 = E_0 - E_2$, and $E_F = E_0 + \Delta E_F$ where E_0 is the neutral level which makes $Q_{SS}/e = 0$, we define

$$n' = n_2 e^{-\Delta E_2/kT} = n_1 e^{-\Delta E_1/kT} \quad (5)$$

and we have

$$Q_{SS}/e = -2n'kT \sinh(\Delta E_F/kT)_S, \quad (6)$$

which is the same form as the best fit to the data, relation (3). To match (3) and (6), we set $2n'kT = 2.44 \times 10^{12}$ or

$$n_1 e^{-\Delta E_1/kT} = n_2 e^{-\Delta E_2/kT} \cong 4.9 \times 10^{13} \text{ states/(cm}^2\text{-eV)}. \quad (7)$$

Thus, the band densities are only determined implicitly together with the energy separations ΔE_1 and ΔE_2 . Since the hyperbolic sine relation will not result for this

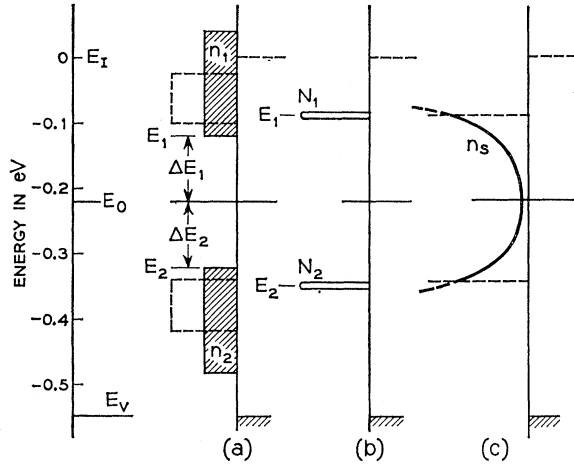


FIG. 5. Three surface-state distributions yielding surface charge vs surface potential relations of the sinh form required to fit the present data.

model unless the band edges E_1 and E_2 are at least kT beyond the largest excursion of $(E_F)_S$ which is ± 0.1 eV, the smallest values of n_1 and n_2 that will fit the data are given from (7) for $\Delta E_1 = \Delta E_2 = 0.1$ eV $= 4kT$, or

$$n_1 = n_2 \geq 2.6 \times 10^{15} \text{ states}/(\text{cm}^2 - \text{eV}). \quad (8)$$

The density of surface atoms in the (111) face of silicon is $8 \times 10^{14}/\text{cm}^2$. Since theory (Shockley⁴ and Kouřek⁴) predicts not more than one surface state per surface atom for the diamond structure, we find that the two bands cannot extend farther than ~ 0.16 eV beyond the edges E_1 and E_2 . Each band then contains 4×10^{14} states. The assumption in (4) above that the bands extended many kT units beyond E_1 and E_2 is thus justified. The results for this case are shown in Fig. 5(a). If E_1 and E_2 are assumed further out from E_0 than 0.1 eV, n_1 and n_2 must increase exponentially to satisfy (7) and to maintain the total number of states equal to $8 \times 10^{14}/\text{cm}^2$ their extent must decrease.

The limiting case of this shrinking in the extent of the two surface bands is shown in Fig. 5(b), where the bands have become two discrete levels, or delta functions. The densities are N_1 and N_2 states/ cm^2 located at E_1 above and E_2 below E_0 . The integration in (4) above over the delta functions now results in

$$Q_{SS}/e = -2N_S \sinh(\Delta E_F/kT)_S, \quad (9)$$

where, to fit our data,

$$N_S = N_1 e^{-\Delta E_1/kT} = N_2 e^{-\Delta E_2/kT} = 1.22 \times 10^{12} \text{ states}/\text{cm}^2. \quad (10)$$

In this case, it is even more important than for the case of two bands that the levels lie outside the largest excursion of $(E_F)_S$ of the data if (9) is to result. Otherwise, the curve of Q_{SS} vs u_S would become horizontal after E_F crossed the level. The smallest values of N_1 and N_2

are thus found by setting $\Delta E_1 = \Delta E_2 = 5kT$, so that

$$N_1 = N_2 \geq 1.8 \times 10^{14} \text{ states}/\text{cm}^2. \quad (11)$$

From theory we expect, again, that $N_1 + N_2 \leq 8 \times 10^{14}$ so that for the simple case of $N_1 = N_2 \leq 4 \times 10^{14}$ we have, by (10), the largest separation for the two levels to be

$$\Delta E_1 = \Delta E_2 \leq 6kT = 0.15 \text{ eV}. \quad (12)$$

The third distribution of surface states that fulfills the experimentally determined relation (3) is found approximately (exactly at $T=0$) by differentiating (3) to give

$$n_S(u_S) = 2.44 \times 10^{12} \cosh(u_S + 9.7) \text{ states}/(\text{cm}^2 - kT) \quad (13)$$

This distribution, shown in Fig. 5(c), must extend at least as far as the maximum excursion of $(E_F)_S$ or ± 0.1 eV about E_0 . The minimum number of states required is then given by integrating (13) over this range which gives $N_{\text{tot}} \geq 1.3 \times 10^{14}$ states/ cm^2 . The distribution could not extend more than ± 0.13 eV from E_0 without exceeding the number of surface atoms.

In conclusion, all models that fit the data well have a density of surface states which is low at the neutral level, rises rapidly at about $E_0 \pm 0.1$ eV and which includes a total number of levels of between one-quarter to one times the surface atom density. The data do not discriminate between discrete levels and surface bands.

It is noted that a measurement of the temperature dependence of the work function going to low temperatures might resolve several questions about this surface-state distribution, if experimental difficulties can be overcome.

In deciding which of the above three distributions is most likely, it is expected that an ideal surface with surface states caused only by a regular system of dangling bonds should exhibit well-defined energy levels. If surface atoms are sufficiently close to each other, these should be broadened into bands of constant density. Hence, for good cleaved surfaces such as used here, it is believed that either the two-band or the two-discrete-level model could apply depending on the strength of interaction between neighboring surface atoms. There seems to be no reason to expect the hyperbolic cosine distribution, unless the surfaces are inhomogeneous and this happens to approximate the result of many different levels.

If the interaction between surface states is large, one might expect conduction via surface states or surface bands. Since the separation between the lower filled states and the empty upper states is of the order of 0.25 eV, one should expect appreciable thermal excitation from lower to upper states. Thus, at room temperature one might find $n \cong p \cong 4 \times 10^{14} e^{-5} \cong 3 \times 10^{12}/\text{cm}^2$, as the carrier density for conduction. An upper limit may be set on the mobility of electrons or holes in these surface states from results of channel measurements on

cleaved n - p - n or p - n - p structures, such as reported by Palmer *et al.*¹¹ and further discussed by Handler.²⁴ Assuming, as claimed, that the total surface conductivity of both surface states and space-charge region was $<10^{-8}$ mho, we find $\mu=\sigma_s/ne\leq 0.02$ cm²/V-sec, a surprisingly low value if there is indeed any interaction between states.

DISCUSSION

The detailed nature of cleaved surfaces studied here as to atomic arrangement and perfection is unknown. They are certainly atomically clean and free of chemical contamination. The electron microscope and x-ray diffraction results published earlier¹⁶ offer hope that the surface consists principally of atomically flat (111) regions completely undisturbed except for the single broken bond of each surface atom. Haneman²⁵ has published evidence that cleaved bismuth telluride surfaces look identical to sputtered and annealed BiTe surfaces in slow electron diffraction, suggesting that cleaved surfaces need not be in a state of stress.

Four types of data exist that make possible a comparison of the properties of the cleaved silicon surfaces used here with those of silicon surfaces cleaned either by heating or by sputtering and annealing in high vacuum. (1) Photoelectric emission data on silicon surfaces taken by Eisinger²⁶ after high-temperature heating, where it is known that a strong p layer formed due to boron from the system, are identical in spectral yield and apparent threshold to the highly p -type cleaved samples in this study; (2) the sticking coefficient of oxygen on the cleaved surfaces found in this work of between 10^{-1} and 10^{-2} agrees within rough experimental accuracy with the values found for heated and for sputtered and annealed silicon,²⁷ and its effect in increasing Φ_{AP} by ~ 0.5 eV is similar to that found by Eisinger; (3) the range of Fermi level position at the surface of from 0.1 eV for extreme n type to ~ 0.35 eV for extreme p type below the mid-gap found in this work over the entire doping range just includes, within the uncertainty of ± 0.1 eV, the value reported by Law¹⁰ of 0.35 to 0.45 eV for sputtered and annealed surfaces when no chemical p layer was present. However, the results would not agree well unless Law's sample actually had a p -type chemical layer, since for a high-resistivity sample such as used by Law, this value should be close to the neutral level of 0.22 eV below midgap (actually, the orientation of the surface in Law's paper is not specified, and the neutral level on different faces may well be different); and (4) the work function values found here for cleaved (111) surfaces are ~ 0.08 eV higher than values previously published for sputtered and annealed (111) surfaces that were

believed to have no p layer present,¹⁸ and are ~ 0.05 eV higher than those reported for (111) surfaces cleaned by heating alone where a strong p layer did exist,¹⁷ though both comparisons are uncertain due to lack of exact knowledge of surface doping in previous work. The differences are small in any case.

Judging from these four areas of comparison, it would seem unlikely that the present cleaved surfaces are greatly different from previously studied clean silicon surfaces. Direct observation of these surfaces reported earlier¹⁶ and analysis of the photoelectric yield¹⁹ indicate, however, that the present surfaces are physically much more perfect.

ACKNOWLEDGMENTS

The authors wish to thank A. Studna and F. Eyler for their technical assistance and E. O. Kane, P. A. Wolff, and other colleagues for helpful discussions.

APPENDIX

For many of the results in this paper and in reference 19 it has been assumed that E_G , χ , and the density and distribution of surface states are not altered by doping until near degenerate levels are reached. Variation in E_G with doping has been measured and since it amounts only to millivolts at degenerate levels it can be ignored here.

χ , the electron affinity, includes the volume contribution from the ion core potential and exchange and correlation terms, as well as surface contributions from the electrostatic dipole layer of the electron cloud that spreads beyond the positive ion cores of the surface atoms. Impurity atoms can thus affect χ either by changing the volume potential terms or the surface dipole. Also, when an electron is localized in a surface state it will presumably have some effect on the electron cloud outside the surface so that as the filling of the surface states is changed by bulk doping an additional change in χ is expected.

The volume effects on χ of an impurity atom are very difficult to estimate.²⁸ However, let us first assume that the impurity effect on the average value of χ is simply proportional to the atomic fraction of impurity N_A/N_{Si} in the host lattice and is such that for 100% impurity the change in χ would be $\Delta\chi_{\max}<\chi$. Then we have $\Delta\chi=(N_A/N_{Si})\Delta\chi_{\max}<(N_A/N_{Si})\chi$. Since the highest doping level used was $N_A\sim 10^{20}$ cm⁻³, whereas $N_{Si}=5\times 10^{22}$ cm⁻³, we estimate that $\Delta\chi<0.008$ V and would thus be negligible. (N_D could be used in place of N_A in this discussion.)

Since the average potential of a uniform charge cloud or radius r_0 and average dielectric constant K about a charged ion varies as $1/Kr_0$ and since the average distance between impurity ions varies as $(1/N_A)^{1/3}$, a somewhat better approximation to the volume effects of impurities on χ may result from the relation $\Delta\chi\sim\text{const}$

²⁴ P. Handler (to be published).

²⁵ D. Haneman, Phys. Rev. **119**, 563 (1960).

²⁶ F. G. Allen, J. Eisinger, H. D. Hagstrum, and J. T. Law, J. Appl. Phys. **30**, 1563 (1959).

²⁷ H. D. Hagstrum, J. Appl. Phys. **32**, 1020 (1961).

²⁸ P. A. Wolff, Phys. Rev. **126**, 405 (1962).

$(N_A/K)^{1/3}$. Such a calculation has been made,²⁹ and the constant is such that one might expect $\Delta\chi$ to be ~ 0.1 eV when $N_A = 10^{18} \text{ cm}^{-3}$. Hence, as seen from the impurity level scale of Fig. 3, appreciable corrections to our data would be necessary for samples for which $|E_F - E_I| > 0.45$ eV. Even for this case, however, the data over the central 0.9 eV of the energy gap would remain unchanged.

For changes in the surface dipole due to an impurity atom, we are justified in taking areal averages since work function was measured by the C. P. D. method. Assume that the work function of individual boron atoms on the surface is 1 V lower than that of silicon, as would be naively predicted from published values. Then, again, for this case $\Delta\chi = (N_A/N_{\text{Si}})(1 \text{ V})$ and at the heaviest doping $\Delta\chi \sim 0.002$ V and is negligible.

For changes in χ due to electrons in surface states, the important variable will be the ratio of filled states to host surface electrons. This ratio is $\sim 1/16$ at the extremes of our data, assuming two surface electrons/silicon atom, and $(Q_{\text{SS}})_{\text{max}}$ of $1 \times 10^{14} \text{ cm}^{-2}$. Hence, if the dipole change due to each filled state is comparable to that of each surface electron, whose total contribution to χ is on the order of 2–3 V, we might expect a change in χ due to filling of states of as much as 0.15 V at extreme dopings. This will drop to a negligible value, however, when N_A is reduced to 10^{19} cm^{-3} and remain negligible

²⁹ C. C. Wang, Technical Report No. 1654-1, 1960, Stanford Electronics Laboratory (unpublished).

for all other dopings. An actual calculation of $\Delta\chi$ due to an electron in a surface state must find the mean position of the electron relative to the positively charged surface ions. The wave function drops off as e^{-2kr} , where k is determined from $\hbar^2 k^2 / 2m^* = -E - E_V$, $E - E_V$ is the energy of the surface state level above the valence band edge, and where m^* will be the free electron mass outside the surface but an effective (light hole) mass inside. Hence, the wave function will decay faster outside the surface than inside, indicating that for the fraction of charge found in the tail of the wave function a large *negative* contribution in χ can result. The actual change in χ will, however, probably be determined by the shape of the potential well at $r=0$, where most of the charge resides.

Finally, the assumption that the surface states themselves are not altered seriously by the doping is believed justified since the energy of the states is presumably determined by the way the lattice terminates and how broken bonds are adjusted after cleavage. The effect of impurity atoms in the surface plane should be no greater than providing one new surface state for each impurity surface atom. But, as shown above, the ratio of impurity to host surface atoms never exceeded 1/500 so this effect should also be negligible.

We conclude that of all these effects only the change in χ from the volume potential change due to the charged ion cores of impurities is likely to be important and this only for near-degenerate samples.