

Solubility of Flaws in Heavily-Doped Semiconductors*

W. SHOCKLEY

Shockley Transistor, Unit of Clevite Transistor, Palo Alto, California

AND

J. L. MOLL

Stanford University, Stanford, California

(Received April 21, 1960)

The solubility of a charged impurity in a semiconductor depends upon the Fermi level. This dependence may be understood in terms of a conceptual model in which an impurity is allowed to diffuse in a specimen containing a p - n junction, so that the Fermi level varies in respect to the band edges. If the impurity can exist in many states of charge (i.e., is a "flaw"), then the concentration of flaws with charge r times the electronic charge varies as the r th power of the hole density. Summing the concentrations for the different states of charge gives the solubility and its dependence upon hole concentration, and, hence, Fermi level.

IT is the purpose of this article to present a simple model on the basis of which the dependence of the solubility of a chemical imperfection can be understood in regard to its dependence upon the donor and acceptor concentration. The essential feature of the argument is that thermal equilibrium is not affected by the introduction of hypothetical processes which permit equilibrium to be established but do not violate the principle of detailed balance.

In Fig. 1 the potential energy distribution for an electron in a semiconductor containing a p - n junction is shown. The situation corresponds to thermal equilibrium with F representing the Fermi level and E_i the intrinsic level. By definition the electrostatic potential and Fermi potential are

$$V = -E_i/q, \quad \phi = -F/q, \quad (1)$$

where $q = |q|$ is the electronic charge and E_i is the level of energy for the Fermi level in intrinsic material. Levels for the first donor ($1d$), first acceptor ($1a$), and second acceptor ($2a$) for a hypothetical flaw, such as gold in germanium, in the notation of Shockley and

Last¹ are also shown. The potential that is shown corresponds to a linear graded junction with local space charge neutrality, which will occur if the Debye length for the intrinsic material is short enough at the temperature involved compared to the concentration gradient. The criteria for shortness is that given by Shockley.²

Figure 1(b) represents the predominant charge condition, which is substantially the same as the average charge on the flaw. In accordance with the Shockley-Last model, this charge changes when the Fermi level passes through any of the characteristic energy levels for the flaw. Thus, the breaks in the curve of Figure 1(b) correspond to the points at which the flaw levels cross the Fermi level in Fig. 1(a).

The problem with which we are confronted is to determine the actual concentration of flaws at each point in Fig. 1 under thermal equilibrium conditions. This is a problem which has been considered from a different viewpoint by Riess, Fuller, and Morin.³

In Fig. 2 a hypothetical system is represented. Here the p - n junction specimen of Fig. 1 is surrounded by an impermeable membrane except for the attachment of an i -type region which extends out of the original specimen. The i -type region is separated from the source of flaws by a semipermeable membrane through which only neutral flaws may pass. When this system has come to thermal equilibrium, the concentration of flaws in all parts of the system will be the same as if the source of flaws with its characteristic chemical potential for flaws were put in contact with any portion of the body of silicon. This is a consequence of the general theorem that any mechanism of establishing the equilibrium leads to the same final equilibrium state. In the intrinsic region there will be a certain concentration of neutral flaws and these are in equilibrium with the source of flaws.

We now imagine that a neutral flaw can diffuse. This

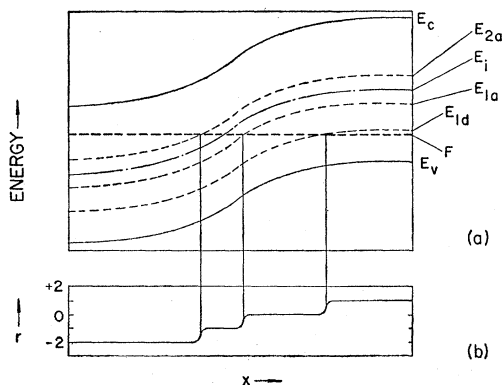


FIG. 1. Distribution of charge on flaws in a p - n junction; (a) Energy levels as a function of position; (b) Average charge on a flaw in units of $q = |q|$.

* Work supported in part by a contract with the Air Force Cambridge Research Center.

¹ W. Shockley and J. Last, Phys. Rev. **107**, 392 (1957).

² W. Shockley, Bell System Tech. J. **28**, 435-439 (1949).

³ H. Riess, C. S. Fuller, F. J. Morin, Bell System Tech. J. **35**, 535-636 (1956).

certainly can occur to a very limited degree by simple interchange of a neutral flaw with an adjoining silicon atom. Although this may not be an important process in the actual diffusion of flaws, it is a possible conceptual process. If we consider this process we see that no force acts to cause a neutral flaw to move. Consequently its equilibrium distribution is determined solely by diffusion. In other words, when equilibrium is reached there can be no concentration gradient of neutral flaws throughout the system. This neglects certain second order effects such as a change in lattice constant with doping level which will exert a mechanical force upon a flaw. It also neglects electrical polarization effects on flaws which may cause a neutral flaw to be concentrated at the point where it is in the highest electrical field.

If the concentration of neutral flaws is known at any point in the semiconductor under thermal equilibrium, then the total concentration is also known from the Shockley-Last Theory, which shows that the ratio of concentrations of flaws in any one of the charge con-

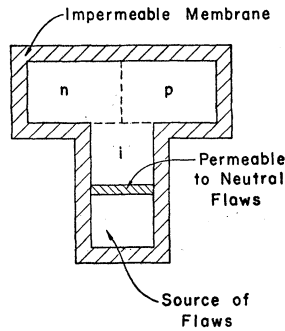


FIG. 2. A hypothetical system with diffusing flaws.

ditions is given simply by the ratio of Boltzmann factors:

$$\begin{aligned} f_{1a}:f_0:f_{1a}:f_{2a} &= \exp[(E_{1a}-F)/kT]:1: \\ &\exp[(F-E_{1a})/kT]: \\ &\exp[(2F-E_{1a}-E_{2a})/kT], \end{aligned} \quad (2)$$

where the f 's are the fractions of the flaws in the various states of charge. From Eq. (2) it follows, for example, that N_{-2} , the concentration of flaws in charge state -2 , i.e., condition 2a, varies with position as

$$N_{-2} = N_f f_{2a} = N_0 \exp[(2F-E_{1a}-E_{2a})/kT], \quad (3)$$

where N_f is the total concentration of flaws at any point and

$$N_0 = N_f f_0 \quad (4)$$

is the concentration of neutral flaws.

Equation (3) may be rewritten in the form:

$$N_{-2} = \exp(2qV/kT) N_0 \times \exp[(2F+E_i-E_{1a}+E_i-E_{2a})/kT], \quad (5)$$

using Eq. (1) to introduce the potential V . In Eq. (5), only the first exponential varies with position in Fig. 2

because N_0 , $2F$, E_i-E_{1a} , and E_i-E_{2a} are all independent of position. Hence the distribution of flaws in charge state $-2q$ is given by the Boltzmann distribution for particles of that state of charge.

Equation (5) can also be reached simply by assuming that the flaws of different states of charge diffuse without changing their state of charge. Under this assumption they will evidently also reach the same Boltzmann distribution.

The result just obtained can be generalized to give the solubility of the flaw as a function of the Fermi level. In a saturated crystal, the density of neutral flaws denoted by N_{0s} , will be independent of $F-E_i$ for the reasons discussed in connection with Fig. 2. The density of flaws of charge rq is $N_{rs}(F-E_i)$ and the saturation solubility N_{fs} is thus

$$\begin{aligned} N_{fs}(F-E_i) &= \sum_r N_{rs}(F-E_i) \\ &= N_{0s} \sum_r \exp[(-rF - \sum^* E_r)/kT], \end{aligned} \quad (6)$$

where

$$\begin{aligned} \sum^* E_r &\equiv E_{1a} + E_{2a} + \cdots + E_{-ra} & r < 0 \\ &\equiv -(E_{1a} + E_{2a} + \cdots + E_{ra}) & r > 0 \\ &\equiv 0 & r = 0. \end{aligned} \quad (7)$$

For levels separated by several kT , only the predominant N_{rs} term in the sum of Eq. (6) will be important and consequently $N_{fs}(F-E_i)$ will depend upon F , as shown by the heavy line of Fig. 3. In each energy interval $N_{fs}(F-E_i)$ varies as $\exp[-r(F-E_i)/kT]$ where the charge on the flaw is $rq = r|q|$.

The charge density due to the flaws is

$$\rho_{fs}(F-E_i) = \sum_r qr N_{rs}(F-E_i) = q \langle r \rangle N_{fs}, \quad (8)$$

where $\langle r \rangle$ is a function of $F-E_i$ and is the average number of electron charges on the flaw. The line $|\langle r \rangle N_{fs}|$, which represents the density of electronic charges is not shown on Fig. 3. It would lie on the higher of the lines N_{1s} and N_{-1s} for $F < E_{2a}$ and would lie above N_{2a} by a factor of 2 for $F > E_{2a}$. (There will, of course, be transition regions of curvature near at $(E_{1a}+E_{1d})/2$ and E_{2a} .)

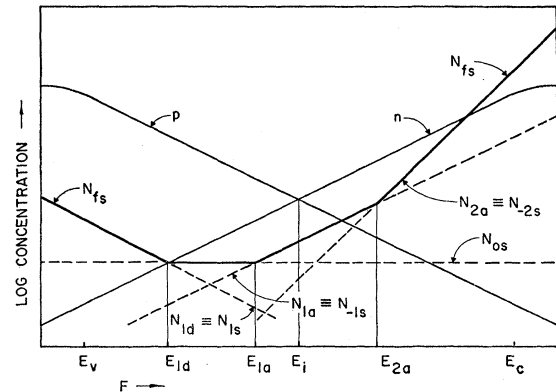


FIG. 3. Saturation concentration of the flaw in various charge states and solubility of the flaw.

In a specimen with N_D ionized donors and N_A ionized acceptors, the electrical charge density is

$$\rho = q\{N_D - N_A - 2n_i \sinh[(F - E_i)/kT] + \langle r \rangle N_{fs}(F - E_i)\}. \quad (9)$$

For a given value of $N_D - N_A$, Eq. (9) with $\rho = 0$ gives the value of F which leads to neutrality and hence to the solubility of the flaw for that doping level. If the carrier charge density is large compared to the flaw density, as is represented for most parts of Fig. 3, then the Fermi level is effectively unaltered by the presence of the flaws and may be calculated in the usual way. On the other hand, if the flaw charge density is larger than the electron density, then the flaw charge density is equal and opposite to the chemical charge density $q(N_D - N_A)$ and the solubility of the flaw is

$$N_{fs} = |N_D - N_A| / |\langle r \rangle|. \quad (10)$$

This same conclusion is reached for the case of lithium with $\langle r \rangle = +1$ by Reiss, Fuller, and Morin³ using a different method of attack.

As an example of a multiply-charged flaw, we consider copper in germanium at 900°C. The levels, as measured near room temperature, are as follows⁴:

$$\begin{aligned} E_{1a} - E_i &= -0.35 \text{ ev}, \\ E_{2a} - E_i &= -0.07 \text{ ev}, \\ E_{3a} - E_i &= 0.13 \text{ ev}. \end{aligned} \quad (11)$$

(These values are probably temperature dependent and possibly should be scaled in proportion to the energy gap at higher temperatures. This would raise the predicted saturation concentration somewhat on Fig. 4.) The value of n_i is⁵

$$n_i = 4 \times 10^{18} \text{ cm}^{-3}. \quad (12)$$

The solubility of copper in intrinsic germanium at 900°C is⁶

$$N_{Cu,s}(O) = 3 \times 10^{16} \text{ cm}^{-3}. \quad (13)$$

⁴ See H. H. Woodbury and W. W. Tyler, Phys. Rev. **105**, 84 (1957), who give $E_{1a} - E_v = 0.04$ ev, $E_{2a} - E_v = 0.32$ ev and $E_c - E_{3a} = 0.26$ ev; F. J. Morin and J. P. Maita, Phys. Rev. **94**, 1525 (1954) who give 0.785 ev for $E_c - E_v$. See also Semiconductors, edited by N. B. Hannay, (Reinhold Publishing Corporation, New York, 1959), p. 341.

⁵ F. J. Morin and J. P. Maita, Phys. Rev. **94**, 1525 (1954). See also reference 4, Hannay, p. 350.

⁶ F. A. Trumbore, Bell System Tech. J. **39**, 208 (1960).

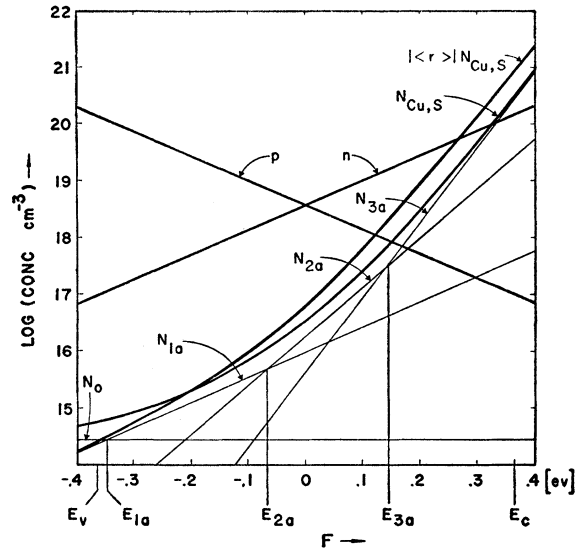


Fig. 4. Solubility and charge density for saturated solid solution of copper in germanium at 900°C.

On the basis of these values, Fig. 4 has been constructed. Since $kT/q = 0.101$ volt (or 0.232 volts per decade) at 900°C, there is conspicuous curvature at the intersections of the line segments. For values of $N_D - N_A > 10^{20} \text{ cm}^{-3}$, the charge on the Cu^{3-} ions dominates the electrons, and

$$N_{Cu,s} \cong (N_D - N_A)/3. \quad (14)$$

The solubility of Cu in a heavily-doped n -type specimen is seen to be more than 10^4 times larger than in an intrinsic specimen.

Copper is thought to be soluble chiefly in substitutional sites, but also to be present at about 10^{-2} less concentration in interstitial sites, where it has a positive charge.⁷ The diffusion is thought to occur chiefly as positive interstitial ions. The solubility in the positive ion form should increase in proportion to the hole density, if the charge is one electronic unit. The ratio of concentrations of the two different forms of dissolved flaws, i.e., interstitial and substitutional, will depend on temperature. Each form will have a diagram like Fig. 3; however, the two diagrams will shift relative to each other along the concentration axis as the temperature changes.

⁷ F. C. Frank and David Turnbull, Phys. Rev. **104**, 617 (1956) and reference 4, Hannay, p. 238.