J. A. Swanson

Clarification of First-Order Semiconduction Effects through Use of Electrochemical Potentials

Abstract: When deviations from equilibrium conditions are small, the net rate at which electrons are interchanged between two groups of electrons which are separately in states of thermodynamic equilibrium is proportional to the difference of the Fermi levels applying to the two groups. With the help of this principle the first-order treatment of conduction effects in semiconductors (appropriate when carrier concentration deviations are small) is considerably simplified. Poisson's equation is shown to be ignorable in the first-order treatment of steady-state effects. Application is given to the Hall effect and to the characterization of probes.

Introduction

The complete set of equations governing carrier flow in semiconductors is exceedingly complex. This complexity derives not only from the number of simultaneous equations involved,* but also from the nonlinear character of some of the equations.

One method of dealing with these nonlinearities is to remove them *ab initio* by restricting oneself to effects of first order in the "causative" physical variables. Mathematically, this corresponds to neglecting the products of quantities which are zero at equilibrium. There are, of course, many problems (such as those involving rectification) in which the nonlinearities are essential. Nevertheless certain quantities of basic physical interest can be determined precisely by first-order considerations. It is the theory of such first-order considerations which concerns us in this paper.

The linearized theory of conduction effects in semiconductors has been previously treated many times in connection with problems of varied nature. This paper is not intended primarily to furnish new results, but rather to clarify the conceptual picture. The nature of this conceptual clarification will be more evident if we first review briefly the conventional approach to the problem.

In the nonlinearized theory the following simplification is often introduced: Poisson's equation is replaced by the "charge neutrality condition." The latter is an assumed relation of equality between the deviation of the electron concentration at any point from the normal equilibrium concentration at that point and the corresponding deviation in the hole concentration. In general, this assumed relation is not precisely fulfilled, since by Poisson's equation it is the discrepancy between these concentration deviations which provides the charge acting as the "source" of the electric field. The replacement is justified, however, as an approximation, if the required discrepancy is much smaller than the individual concentration deviations. Whether or not this is so can be checked, after a solution of the set of equations has been obtained, by substitution of the solved electric field in Poisson's equation and calculation of the required charge density. The approximation is found to be good in the interior of a semiconductor except near reverse-biased junctions, for which other approximations are appropriate.

The conventional first-order treatment has also incorporated the neutrality condition. One of the principal points of this paper is to show that the mathematics of first-order steady-state effects is as simple without the neutrality assumption as with it. Hence, if a problem can be solved approximately by assuming space-charge neutrality, it can also be solved exactly. In fact, we shall be able to show that the exact solution can be obtained from the approximate solution by a simple reinterpretation of the variables.

When the problem is formulated in terms of the electrochemical potentials, Poisson's equation is seen to be ignorable. It is this unusual feature of the first-order, steady-state problem which gives rise to the above properties. The fact that the electrochemical potentials are also operationally significant (connected with the voltage assumed by the probes) means that the electrostatic potential can be altogether ignored in this type of semiconductor problem.

This investigation was first undertaken in order to make more rigorous the results of some previous work, in

*In the simplest case, there are five: two transport equations, two equations of continuity, and Poisson's equation.

which the author participated, on the effect on the Hall voltage of carrier concentration deviations induced by the magnetic field. Accordingly, some discussion of this problem will be included.

The first-order system in terms of electrochemical potentials

Electrochemical potentials, which are related directly to the Fermi level or "quasi-Fermi level" of the one-electron band theory approximation, have often been employed in connection with semiconductor problems.[1] A separate electrochemical potential can be defined for any group of electrons which is in thermodynamic equilibrium with itself (but not necessarily with other electrons). The advantage here gained by the use of electrochemical potentials consists in the fact that the net rate of interchange of electrons between two such groups is, so far as first-order considerations are involved, simply proportional to the electrochemical potential difference, or, in the continuous case, to the potential gradient.

Thus, the current densities of holes and electrons are given to first order by the equations:

$$i_p = -e\mu_p p_o \nabla\phi_p \tag{1}$$

$$i_n = -e\mu_n n_o \nabla\phi_n \tag{2}$$

where p_o and n_o are the equilibrium concentrations, not necessarily uniform, ϕ_p and ϕ_n are the electrochemical potentials for holes and electrons, respectively, and μ_p and μ_n are the hole and electron mobilities. The electrochemical potential of holes is defined as that of electrons in the valence band, and that of "electrons" as the electrochemical potential appropriate to the conduction band electrons. The sign convention adopted here is such as to render Eqs. (1) and (2) formally identical to the usual pure conduction equations written in terms of the electrostatic potential. These equations, however, also include diffusion effects.

First-order concentration deviations (deviations small compared to equilibrium concentrations) result in a steady-state recombination of holes and electrons proportional to the difference of the electrochemical potentials. Thus to equations (1) and (2) may be added:

$$\nabla\cdot i_p = -\nabla\cdot i_n = K_1(\phi_p - \phi_n) \tag{3}$$

where K_1 is a positive constant determined by the lifetime. Equation (3) is valid for a steady-state process even if the deviations in hole and electron concentrations are not equal; i.e., even if there are trapping effects. Recombination at the surface is also proportional to the difference $(\phi_p - \phi_n)$. In this case we have

$$\hat{n}\cdot i_p = K_2(\phi_p - \phi_n) \tag{4}$$

where $\hat{n}$ is the unit outward normal to the surface and K_2 is a positive constant related to the surface recombination velocity. This is a complete set of equations. The current densities may be eliminated by substituting the right members of equations (1) and (2) in equation (3) for i_p and i_n in equations (3) and (4). Thus the current densities and electrochemical potentials can be found through the use of equations (1) to (4) and suitable boundary conditions at contacts, without inquiring into the behavior of the electrostatic field, or introducing an arbitrary relationship between hole and electron concentration deviations. The electrostatic potential may subsequently be calculated, if desired, from the first-order Poisson's equation, which takes the form

$$\nabla^2\psi = \varepsilon_o e^2[(n_o + p_o)\psi - n_o\phi_n - p_o\phi_p]/kT$$

in the absence of trapping levels. However, the electrostatic potential is intrinsically less interesting than the electrochemical potentials, since the voltage of a probe placed on the semiconductor surface is sensitive to the latter rather than to the former. Poisson's equation is thus of little interest in connection with first-order, steady-state effects. (This result cannot be generalized to non-steady-state problems, since in this case the equation of continuity involves the concentration deviations explicitly.)

Application to the weak-field Hall effect

The preceding discussion shows in particular that the semiconductor Hall constant(s) can be calculated without reference to Poisson's equation. We have (to first order in the magnetic field)

$$\frac{\partial\phi_p}{\partial z} = \frac{\mu_{pH}}{e}EH + \frac{i_z}{e\mu_p p_o} \tag{5}$$

$$\frac{\partial\phi_n}{\partial z} = -\frac{\mu_{nh}}{e}EH - \frac{i_z}{e\mu_n n_o} \tag{6}$$

where E is the primary longitudinal electric field, H is the perpendicular component of magnetic field, i_z is the current of holes flowing perpendicularly to E and H (i_z is also the negative of the corresponding electron current, since the net transverse current is zero), and μ_{pH} and μ_{nH} are the hole and electron Hall mobilities. The z-axis has been chosen to lie in the direction of magnetic deflection of holes and electrons.

When the lifetime is infinite, i_z is zero, and ϕ_n and ϕ_p have independent values appropriate to the particle concerned. When the lifetime is finite, i_z may be calculated using equations (3) to (6) and appropriate boundary conditions. In the special case of zero lifetime, ϕ_p and ϕ_n are equal and equations (5) and (6) determine the conventional Hall voltage for mixed carriers.

Probes may in principle be constructed to measure either ϕ_p or ϕ_n. Thus it is possible to measure a positive

[1] For a discussion see W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1951) Sec. 12.4.

Hall constant on n-type material or a negative Hall constant on p-type material. The discrepancy between ϕ_p and ϕ_n may be used as a measure of lifetime. An ohmic probe introduces the boundary condition $\phi_n = \phi_p$ at the surface of the probe and thus insures the measurement of the conventional Hall effect in a uniform sample.

A comprehensive analysis of the effects of lifetime on the Hall constant has been made by Landauer and Swanson.[2] In that paper, the usual approximation

$$p_1 = n_1 \tag{7}$$

is employed, where p_1 is the deviation in the concentration of holes from the equilibrium value, and n_1 is the same quantity for electrons. Equation (7) is not fulfilled exactly, since a net difference $(p_1 - n_1)$ in the concentration deviations is required to satisfy Poisson's equation. However, the fraction $(p_1 - n_1)/p_1$ is usually very small, and the approximation (7) is therefore a reasonable one. It will now be shown in fact that the previous results are *exact* insofar as the electrochemical potentials are concerned. This circumstance results essentially from the ignorability of Poisson's equation in first-order, steady-state processes.

Let us introduce the new quantities

$$\phi = \frac{p_o \phi_p + n_o \phi_n}{p_o + n_o} \tag{8}$$

$$p' = \frac{e}{kT} \frac{p_o n_o}{p_o + n_o} (\phi_p - \phi_n) \tag{9}$$

It is easily verified that if equation (7) holds, then

$$\phi = \psi \qquad (\text{if } p_1 = n_1) \tag{10}$$

$$p' = p_1 = n_1 \quad (\text{if } p_1 = n_1) \tag{11}$$

Let us introduce also the quantities τ and s:

$$K_1 \tau \equiv \frac{e^2}{kT} \frac{p_o n_o}{p_o + n_o} \equiv \frac{K_2}{s} \tag{12}$$

If the concentration deviations are equal, τ is the usual volume lifetime, and s is the surface recombination velocity. In terms of ϕ, p', τ, and s the equations (1) to (4) take the same form as those for ψ and p_1 employed in Ref. 2. Thus, substituting the interpretations (10) to (12) for the occurrences of ψ, p_1, τ, and s in the equations of Ref. 2, we achieve an exact treatment of the Hall effect. It is interesting to note that even should trapping effects occur, resulting in a relatively large value of $(p_1 - n_1)$, the treatment is still exact, providing that τ and s are reinterpreted in accordance with equation (12). Trapping effects do not invalidate equations (3) and (4), but necessitate the use of a more general form of Poisson's equation, in which p_o and n_o are replaced by other concentrations determined by the relative number of trapped carriers.

A correction to the Hall *electrostatic* potential for the special case of uniform, extrinsic, short-lifetime material has been calculated by Banbury, Henisch, and Many[3] taking into account the deviations from space charge neutrality which, in the absence of surface states, constitute the source of the Hall electric field. The present discussion, while affording a means of calculating the correction in the more general case, shows this correction to be relatively uninteresting, since it does not lead to a corresponding alteration of the measured Hall constant. The Hall voltage is in fact measured by probes, and the voltage assumed by probes is a characteristic average of the *electrochemical* potentials in the neighborhood of the probe (*cf.* next section).

In treating higher order effects, such as the variation of the Hall constant with field, the electrostatic potential can no longer be ignored. However, approximate treatments which reject Poisson's equation in favor of an approximate relationship between hole and electron concentrations should give satisfactory results in view of the minuteness of the space charge effects. Higher order effects should be unimportant as long as the Hall voltage remains less than kT/e.

The characterization of probes

In this section, it will be shown that the voltage of a floating probe of given character is completely determined by the electrochemical potentials in the neighborhood of the probe, provided that the concentration deviations in the neighborhood of the probe are small compared to the equilibrium concentrations. This latter condition is usually fulfilled when dealing with linear or "first order" effects, such as the Hall effect at low magnetic field strengths. It has been shown above that the electrochemical potentials may be determined without reference to the electrostatic potential or to Poisson's equation. It will be established that the electrostatic potential is also ignorable when considering the actual voltages assumed by probes.

The obtained result is the following:

$$V_{pf} = k_p \phi_p + k_n \phi_n \tag{13}$$

where V_{pf} is the voltage assumed by a floating probe, ϕ_p and ϕ_n are the electrochemical potentials which would exist at the position of the probe were the probe absent, and k_p and k_n are dimensionless constants characteristic of the probe. In most cases

$$k_p \doteq \gamma \tag{13a}$$

$$k_n \doteq (1 - \gamma) \tag{13b}$$

where γ is the hole injection efficiency of the probe. These results follow essentially from the linearity of equations

[2] R. Landauer and J. Swanson, Phys. Rev. **91**, 555 (1953).
[3] P. C. Banbury, H. K. Henisch, and A. Many, Proc. Phys. Soc. **66**, 753 (1953).

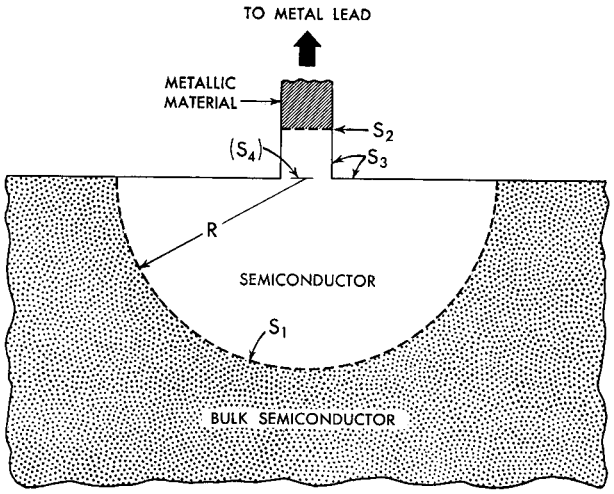

Figure 1

A probe and the associated surfaces S_1, S_2, and S_3 employed in the analysis. The surface S_4 is the continuation within the probe of the surface of the bulk semiconductor. The surface S_2 may be above S_4 as shown (filamentary contact), at S_4 (point contact or alloy contact), or possibly somewhat beneath S_4. There may be a *p-n* junction in the neighborhood of S_2.

(1) to (4). A rigorous proof of equation (13) will now be given.

Consider a volume of semiconductor enclosed by a surface, S. Let S be composed of three sections S_1, S_2, and S_3. Along S_1 and S_2 let the electrochemical potentials have constant values, and furthermore let them have equal values along S_2. Along S_3 the values are not predetermined, but condition (4) is to apply. The argument is general, but S_1, S_2, and S_3 will be interpreted in the present case as follows: S_1 is a hemisphere totally contained by the semiconductor with edge on the semiconductor surface, such that this edge is centered about the probe under consideration (see Fig. 1). We shall assume that the probe is small compared with distances over which the electrochemical potentials in the absence of the probe would change appreciably; thus a radius R exists for the hemisphere such that the electrochemical potentials are substantially constant along the hemisphere and equal to the values they would have were the probe absent. This radius is less than the distance over which the potentials would change appreciably were the probe absent, but several times larger than a typical linear dimension of the probe. The electrochemical potentials at S_1, ϕ_{p1} and ϕ_{n1}, are hence in this case essentially the potentials which would exist at the surface of the semiconductor were the probe absent. Whatever the nature of the probe, at some surface there is a transition from the semiconductor to an essentially metallic material. At this surface, which we shall identify with S_2, the electrochemical potentials are both equal to the voltage of the probe, V_p,

$$\phi_{p2} = \phi_{n2} = V_p \tag{14}$$

The surface S_3 is now interpreted as the free surface of the semiconductor which joins S_1 and S_2. Let I_p and I_n be the total hole and electron currents, respectively, that flow through the surface S_1. Because of the linearity of equations (1) to (4), we may write the relations

$$I_p = C_{11}\Delta\phi_p + C_{12}\Delta\phi_n \tag{15a}$$

$$I_n = C_{21}\Delta\phi_p + C_{22}\Delta\phi_n \tag{15b}$$

where

$$\Delta\phi_p = \phi_{p2} - \phi_{p1} = V_p - \phi_{p1} \tag{16a}$$

$$\Delta\phi_n = \phi_{n2} - \phi_{n1} = V_p - \phi_{n1} \tag{16b}$$

and the C_{ij}'s are constants.

(The validity of equations (15a and b) is easily confirmed by observing that the general solution is a superposition of solutions having $\Delta\phi_p = 0$ and solutions having $\Delta\phi_n = 0$. But all solutions with $\Delta\phi_p = 0$ are constant multiples of each other; hence all currents in this case are linearly related to $\Delta\phi_n$. The symmetrical statement holds when $\Delta\phi_n = 0$).

If the probe is floating, $I_p + I_n = 0$. Hence,

$$(C_{11} + C_{21})\Delta\phi_p = -(C_{12} + C_{22})\Delta\phi_n \tag{17}$$

from which it readily follows that a relationship of type (13) holds for ϕ_{p1} and ϕ_{n1} with

$$k_p = \frac{C_{11} + C_{21}}{\Sigma_{ij} C_{ij}} \tag{18a}$$

$$k_n = \frac{C_{12} + C_{22}}{\Sigma_{ij} C_{ij}} \tag{18b}$$

This completes the more important part of this section. It is, however, of interest also to justify the semiquantitative relations (13). To do this we must find γ in terms of the C_{ij}'s. Imagine a current passing through the probe, and let this current be the sole source of excess carriers. Then, if the radius of the hemispherical surface S_1 is truly large compared to probe dimensions, we may take $\phi_{p1} = \phi_{n1} =$ voltage of the bulk of the semiconductor, and consequently $\Delta\phi_p = \Delta\phi_n$. Then

$$(C_{21} + C_{22})\, I_p = (C_{11} + C_{12})\, I_n \tag{19}$$

from which it follows that

$$\gamma = (C_{11} + C_{12})/\Sigma_{ij} C_{ij} \tag{20}$$

if γ is defined as

$$\gamma = I_p/(I_p + I_n) \qquad (\text{at } S_1) \tag{21}$$

If we could take $C_{12} = C_{21}$, equations (13a) and (13b) would follow. This cannot be done in general. However, in most cases, the off-diagonal elements of C are much smaller than the diagonal elements,[4] so that relations (13) are approximately fulfilled. These relations are not of great practical value, since it would be difficult to measure γ independently. However, it is true that a probe with unity γ would certainly measure ϕ_p, and a probe with zero γ would measure ϕ_n. Such probes can be constructed by suitable forming, that is, impurity diffusion or alloying in the immediate vicinity of the probe.

[4] The presence of the off-diagonal terms C_{12} and C_{21} is due to the recombination of holes and electrons. If the bulk lifetime were infinite, and if the surface recombination velocity were zero except at S_1, we should have $C_{12} = C_{21} = 0$. There is little contribution to C_{12} or C_{21} from the bulk semiconductor if surface recombination is negligible, and if $R \ll 1$, where L is the diffusion length. In the general case equations (2) are more nearly fulfilled if we define γ as the injection efficiency along the surface S_4 (a continuation of the external semiconductor surface) or, if the probe structure includes a junction extending beyond S_4, the injection efficiency along this junction. However, since we are interested principally in equation (1), which is always valid, and not the qualitative equivalence (2), it does not seem desirable to treat these questions in greater detail.

Received August 21, 1956