PHYSICAL REVIEW B VOLUME 16, NUMBER 12 15 DECEMBER 1977

Quantum statistical theory of semiconductor junctions in thermal equilibrium

Oldwig von Roos*

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

(Received 13 July 1977)

By means of a quantum-mechanical phase-space distribution function and its corresponding Boltzmann equation, the free-carrier and electric-field distributions of one-dimensional semiconductor junctions (n-p, p-p^+, etc.) are evaluated. It is shown that quantum and exchange corrections, which have been neglected in the past, play an important role in the determination of the built-in electric field within the transition region, the region in which the doping concentration changes rapidly (from n-type to p-type material for instance). This is particularly true in cases of high doping concentrations, i.e., when carrier densities become degenerate. Exact expressions will be given for the maximum built-in electric field in case of abrupt junctions. It is also shown that the exchange effect induces a slight change in the position of the band edges which persists through the homogeneous (neutral) part of the junction far away from the transition region. A numerical example is given and the quantitative differences between heavily doped (degenerate) and nondegenerate (classical) junction characteristics (maximum electric field, built-in voltage and carrier concentration within the transition region) are determined. The theory is briefly generalized to encompass high-low junctions.

I. INTRODUCTION

Before going into the details of the theory let us delineate the model of the semiconductor junction which will be analyzed in the main body of this paper. Later, generalizations will be made at appropriate points.

From the outset we are considering an abrupt one-dimensional n-p junction in thermal equilibrium. The junction is located at $x=0$. The uniformly doped n-type material extends into the half-space $x<0$, whereas the uniformly doped p-type material extends into the half-space $x>0$. The n-type material is so heavily doped that degenerate statistics for the electrons apply. By necessity then the hole concentration or minority-carrier number density is nondegenerate in the n material.[1] We must emphasize that current mechanisms which may be due to a varity of causes (majority-carrier tunneling, minority-carrier diffusion, and other mechanisms) need not concern us since we are dealing with thermal equilibrium throughout. Also, the assumption of a *uniform* impurity number density precludes the incorporation of the effect of band-gap tailing,[2] an effect which is due to spatial fluctuations in the impurity density and affects both the electron and hole band edges and is not particularly dependent on the average impurity density although spatial fluctuations tend to be larger the larger the average impurity concentrations are. We shall, however, see that a slight rise of the conduction-band edge occurs which is exclusively due to degeneracy. Furthermore, we consider the semiconductor crystal to be ideal, while grain boundaries and other faults are ignored. The mechanism of trapping is severely affected by these faults, but fortunately trapping plays no role in considerations of thermal equilibrium. In the language of transport theory the collision integral is absent in thermal equilibrium.

In a by now classical paper[3] Slater has given a survey of the Thomas-Fermi approach adapted to the electron-hole distribution in semiconductors. Not many authors[4] have used this approach for a determination of n-p junction characteristics in spite of the large literature devoted to semiconductor-device theory. The reason for this is the fact that there exists an excellent, approximate and therefore much simpler theory than the self-consistent-field theory of Thomas-Fermi. This is the theory developed by Shockley[5] which contains as central part the assumption of a "depletion layer." The depletion-layer assumption is dealt with in all textbooks which expatiate on semiconductor junction theory.[6] Under this assumption an n-p junction is visualized as consisting of three parts. The n part, the depletion layer between n- and p-type material in which the free-carrier concentration is assumed to be negligibly small, and the p part. Both n and p parts are assumed to be electrically neutral and contain no electric field whether or not an external voltage is applied across the junction. The depletion layer itself is endowed with a certain width which may be determined by the requirement of overall space-charge neutrality and the magnitude of the donor and acceptor concentrations on either side of the depletion layer. In reality of course the free-carrier number density is a continuous function of the distance from the junction and does not change abruptly to zero. As the dopant concentration increases and the carrier concentration becomes more and more de-

generate the depletion-layer assumption becomes more and more questionable even under conditions of thermal equilibrium. It is of course true that in a situation far off equilibrium, under conditions of high-level injection, the depletion-layer approximation breaks down completely whether or not the carrier concentration is degenerate.[7] We will therefore refrain from making any approximations in addition to the Thomas-Fermi self-consistent-field approximation and consider the junction or transition layer a region in which carrier concentrations change continuously, albeit rapidly.

If the doping levels stay moderately high a classical Maxwell-Boltzmann velocity distribution for free carriers can be assumed. But as the doping level increases and the impurity density becomes larger and larger, quantum-mechanical effects become more and more important.

These effects can be classified as due to the Pauli principle (exclusion principle) and the Heisenberg principle (uncertainty principle). The former will give rise to the necessity of introducing Fermi-Dirac statistics and exchange effects mediated via Coulomb forces. The latter will give rise to diffraction or inhomogeneity effects of the wave packets constituting the free carriers.

Incorporation of these effects into an analysis of highly doped junctions makes the use of a quantum-mechanical formalism mandatory. Since the effects to be described are largely statistical in nature, it is advantageous to employ a quantum-mechanical analog to the classical Boltzmann equation rather than Schrödinger's equation.

In Sec. II then, the quantum-mechanical Boltzmann equation as it applies to semiconductors will be displayed and connection with previous work will be established. In Sec. III the theory of an n-p junction will be derived with the tool provided for in Sec. II and in Sec. IV some of the implications will be worked out with the aid of a numerical example. Also generalizations to other junction configurations (high-low junctions for instance) will be given. To conclude this introduction a final remark is in order: as the dopant concentration increases and the semiconductor becomes degenerate two effects will play a progressively more important role. One effect is the previously mentioned band-gap tailing due to spatial fluctuations of dopant concentrations. As long as we are dealing with thermal-equilibrium conditions this effect is unimportant since it changes only the density of states near the band edge insignificantly changing the electrostatic-potential distribution. The effect is however important under *non*equilibrium conditions (enhanced tunneling in Esaki diodes, etc.). Another effect is the broadening of donor (or acceptor) levels.[8] But as long as the overlap of wave functions corresponding to this broadening is confined to only a few impurity centers (≈10 say), electrons residing there can hardly be called free.[9] Henceforth we neglect these effects.

II. PRELIMINARIES

Some time ago the author developed a quantum-mechanical Boltzmann equation[10] which proved to be particularly simple in dealing with exchange effects.[11] Although the reader is referred to these papers for details, we give here those properties of the distribution function which are subsequently utilized in the development of the junction theory.

For a one-electron system define

$$\tilde{F}(\vec{r},\vec{k}) = (2\pi)^{-3/2}\sum_n w_n \psi_n(\vec{r})\, c_n^*(\vec{k})\, e^{-i\vec{k}\cdot\vec{r}} . \qquad (1)$$

Here $\vec{k}=(m/\hbar)\vec{v}$ is the wave vector, ψ_n is a complete orthonormal set of wave functions, c_n^* is their complex-conjugate Fourier transforms, and w_n are the statistical weights (probability of occupancy of state n). The equation satisfied by expression (1) is

$$\left(\frac{\partial}{\partial t}+\frac{\hbar}{m}\vec{k}\cdot\nabla_{\vec{r}} - i\,\frac{\hbar}{2m}\nabla_r^2\right)\tilde{F} = \frac{i}{\hbar}\left[\exp(-i\nabla_{\vec{r}}\cdot\nabla_{\vec{k}})-1\right] V(r)\tilde{F} . \qquad (2)$$

where V is the potential energy acting on the electron and the operator on the right-hand side of Eq. (2) is defined by

$$\left[\exp(-i\nabla_{\vec{r}}\cdot\nabla_{\vec{k}})-1\right] V(\vec{r})\tilde{F}(r,k) = \sum_{n=1}^{\infty}\frac{(-i)^n}{n!}(\nabla_{\vec{r}}\cdot\nabla_{\vec{k}})^n V(\vec{r})\tilde{F}(\vec{r},\vec{k}) , \qquad (3)$$

with the understanding that the gradient $\nabla_{\vec{r}}$ *only* operates on $V(\vec{r})$. Remember that $\vec{k}=(m/\hbar)\vec{v}$ we see that in the limit $\hbar=0$: Eq. (2) goes over into the classical Liouville equation

$$\left(\frac{\partial}{\partial t}+\vec{v}\cdot\nabla_{\vec{r}}\right)F = \frac{1}{m}\nabla_{\vec{r}}V(\vec{r})\cdot\nabla_{\vec{v}}F(\vec{r},\vec{v}) . \qquad (4)$$

Some other important properties of $\tilde{F}$ are

$$\int \tilde{F}(\vec{r},\vec{k})\, d^3k = N(\vec{r}) , \qquad (5)$$

or the integration over k (or velocity) space gives the number density of electrons. But since $\tilde{F}$ as defined by Eq. (1) is in general complex it is not an observable, although observable quantities may be extracted from it quite similar to the way it is done with classical statistical distribution functions. An example was already provided by Eq. (5). Another example is the mean kinetic energy. It is simply given by

$$E_{\text{kin}} = \frac{\hbar^2}{2m}\int k^2 \tilde{F}(\vec{r},\vec{k})\, d^3k\, d^3r \,. \tag{6}$$

Sometimes one should realize that $\tilde{F}$ is not an observable. This circumstance is for instance reflected in the fact that the mass current density (flux) of electrons is given by

$$\vec{j} = \hbar\, \text{Re}\left(\int \vec{k}\tilde{F}(\vec{r},\vec{k})\, d^3k\right), \tag{7}$$

rather than by the integral itself (Re means real part).

In order to proceed, and introduce the spirit of the Thomas-Fermi self-consistent approach, Eq. (2) has to be augmented by the Coulomb interaction and exchange terms mediated by the Coulomb forces. In short, the many body counterpart of Eq. (2) has to be invoked. But this has been done by Levine and this author many years ago.[12] There, the many electron formalism[10] was applied to the electron cloud of an atom or ion.[13] It is conceptually very easy to apply the theory developed in Ref. 12 to semiconductor junctions with due regard to necessary modifications imposed by the fact that the free carriers consist of electrons and holes rather than electrons alone, the electron and hole masses are effective masses determined by the curvature of the bottom (top) of the conduction (valence) bands and that finally the electrostatic forces between the charged particles are effectively weakened by the static dielectric constant ϵ. A further simplification arises from the assumed one dimensionality of the junction considered. If $U(x)$ defines the potential energy normalized to zero at $x=-\infty$ (deep inside the n-type material), the Boltzmann equation for the one particle distribution function for electrons $F_e(x,\vec{v})$ may be written

$$\begin{aligned}\left(v\frac{\partial}{\partial x} - \frac{1}{m_e}\frac{\partial U}{\partial x}\frac{\partial}{\partial v}\right)\tilde{F}_e(x,\vec{v}) = {} & \frac{i\hbar}{m_e}\left[\frac{\partial^2}{\partial x^2} - \frac{1}{m_e}\sum_{n=0}^{\infty}\left(-\frac{i\hbar}{m_e}\right)^n \frac{\partial^{n+2}U}{\partial x^{n+2}}\frac{2}{(n+2)!}\frac{\partial^{n+2}}{\partial v^{n+2}}\right]\tilde{F}_e(x,\vec{v}) \\ & + \frac{ie^2\hbar}{2m_e^2\epsilon}\int d^3l \exp(i\vec{l}\cdot\vec{v})\tilde{f}_e(x,\vec{l}) \\ & \times\left[\frac{1}{|\vec{l}|}\tilde{F}_e\left(\vec{r}+\frac{\hbar}{m_e}\vec{l},\vec{v}\right) - \frac{1}{2\pi^2}\int d^3v'\,|\vec{v}-\vec{v}'|^{-2}\tilde{F}_e\left(\vec{r}+\frac{\hbar}{m_e}\vec{l},\vec{v}'\right)\right].\end{aligned} \tag{8}$$

The last term on the right-hand side of Eq. (8) is due to exchange.[14] ϵ is the static dielectric constant of the medium. The vector $\vec{r}$ in the arguments of $\tilde{F}_e$ is $\vec{r}\equiv(x,0,0)$ and the vector $\vec{v}$ is defined as $\vec{v}\equiv(v,v_y,v_z)$. There are of course no exchange terms arising from electron-hole interactions since electrons and holes are distinguishable.[15] The quantity $\tilde{f}_e$ defined by

$$\tilde{f}_e = \int d^3v \exp(-i\vec{l}\cdot\vec{v})\tilde{F}_e(\vec{r},\vec{v}) \tag{9}$$

is the Fourier transform of $\tilde{F}_e$. An equation completely analogous to Eq. (8) can be written for holes almost immediately. We merely have to replace the effective electron mass m_e by the effective hole mass m_h, the electrostatic energy U by $-U$ and finally the electron distribution function $\tilde{F}_e$ by the corresponding hole distribution function $\tilde{F}_h$.[16] In other words

$$\begin{aligned}\left(v\frac{\partial}{\partial x} + \frac{1}{m_h}\frac{\partial U}{\partial x}\frac{\partial}{\partial v}\right)\tilde{F}_h(x,\vec{v}) = {} & \frac{i\hbar}{m_h}\left[\frac{\partial^2}{\partial x^2} + \frac{1}{m_h}\sum_{n=0}^{\infty}\left(-\frac{i\hbar}{m_h}\right)^n \frac{\partial^{n+2}U}{\partial x^{n+2}}\frac{2}{(n+2)!}\frac{\partial^{n+2}}{\partial v^{n+2}}\right]\tilde{F}_h(x,\vec{v}) \\ & + \frac{ie^2\hbar}{2m_h^2\epsilon}\int d^3l \exp(i\vec{l}\cdot\vec{v})\tilde{f}_h(x,\vec{l}) \\ & \times\left[\frac{1}{|\vec{l}|}\tilde{F}_h\left(\vec{r}+\frac{\hbar}{m_h}\vec{l},\vec{v}\right) - \frac{1}{2\pi^2}\int d^3v\,|\vec{v}-\vec{v}'|^{-2}\tilde{F}_h\left(\vec{r}+\frac{\hbar}{m_h}\vec{l},\vec{v}'\right)\right].\end{aligned} \tag{10}$$

The system of equations (8) and (10) is closed by the self-consistent requirement (Poisson's equation)

$$\frac{d^2U}{dx^2} = \frac{4\pi e^2}{\epsilon}\left(\int d^3v\,[\tilde{F}_h(x,\vec{v}) - \tilde{F}_e(x,\vec{v})] + N_D(x) - N_A(x)\right). \tag{11}$$

N_D and N_A are the effective donor and acceptor number densities, i.e., the number density of positively charged donors and negatively charged acceptors, respectively, and ϵ is the dielectric constant of the medium. For the case of monovalent donors and acceptors we have

$$N_D = N_{D0}(x)\,[1+\beta_D \exp(\epsilon_F - \epsilon_D)]^{-1} \tag{12a}$$

and

$$N_A = N_{A0}(x)\,[1+\beta_A^{-1}\exp(\epsilon_A - \epsilon_F)]^{-1}. \quad (12b)$$

Here β_D and β_A are the spin degeneracies of the donor and acceptor ground states, respectively.[17] ϵ_D, ϵ_A, and ϵ_F are the energies of the donor, acceptor, and Fermi levels in units of kT, respectively. N_{D0} and N_{A0} represent the total donor and acceptor densities which are allowed to be functions of x, the distance from the junction. In our model, introduced in Sec. I,

$$N_{D0}=0 \ \text{for } x>0, \quad N_{A0}=0 \ \text{for } x<0. \quad (13)$$

Equation (12) must not be taken too seriously since in any realistic case several monovalent as well as multivalent impurities may be present. Also compensating acceptor levels in the n-type material as well as compensating donor levels in the p-type material may not be overlooked. $N_D(x)$ and $N_A(x)$ are considered to be known functions of x given by the type of semiconductor material under consideration.

Equations (8), (10), and (11) constiture the starting point of the junction theory to be developed in Sec. III. Before entering into that development an additional remark is in order. Both Eqs. (8) and (10) do not contain any exchange contributions arising from bound-free interactions between free electrons and electrons residing on acceptor sites for instance. But these contributions are negligible since the overlap of the plane waves associated with free electrons (or holes) with the bound (localized) electrons (or holes) is in fact very small.

III. DEVELOPMENT OF THE THEORY

Equations (8), (10), and (11) are formidable equations which can only be solved by approximations. Since for nondegenerate semiconductors quantum effects will be negligible and with increasing doping levels these effects become more and more important, we propose to solve Eq. (8) by an expansion in $\hbar$ so that for $\hbar=0$ classical results prevail. Faithful to our introductory remarks that we only consider degenerate electron concentrations (for the n-type material $x<0$) of the junction we may write

$$\tilde{F}_e = F_{e0} + \hbar F_1 + \hbar^2 F_2 + \cdots \quad (14)$$

and

$$\left(v\frac{\partial}{\partial x} + \frac{1}{m_h}\frac{\partial U}{\partial x}\frac{\partial}{\partial v}\right) F_{h0}(x,\vec{v}) = 0. \quad (15)$$

We shall keep only terms to order $\hbar^2$ [in Eq. (14)] because we know that quantum and exchange corrections are only contributing at fairly high densities and are likely to become progressively smaller as the order of expansion is increased. An excellent approximation for the distribution function in zeroth order is already given by,[18] assuming local equilibrium,

$$F_{e0} = \frac{M_c}{4\pi^3}\left(\frac{m_e}{\hbar}\right)^3\left[1+\exp\left(\frac{m_e}{2kT}(v^2+v_y^2+v_z^2) + \frac{U(x)-E_F}{kT}\right)\right]^{-1}. \quad (16)$$

In analogy, the solution of Eq. (15) is taken to be

$$F_{h0} = \frac{1}{4\pi^3}\left(\frac{m_h}{\hbar}\right)^3\left[1+\exp\left(\frac{m_h}{2kT}(v^2+v_y^2+v_z^2) + \frac{E_F-U(x)+E_G}{kT}\right)\right]^{-1}. \quad (17)$$

E_G is the energy of the band gap. As stated earlier the potential energy U is normalized to zero at $x=-\infty$. This signifies that all energies are measured from the bottom of the conduction band in the uniform n-type material far away from the junction. It also means that the conduction band edge moves up or down depending on the functional dependence of $U(x)$, the Fermi level and the band gap being constant throughout the junction as they must in thermal equilibrium. Adopting largely the terminology of Blakemore[20] we have with

$$\eta = \eta(x) = (E_F - U)/kT = \epsilon_F - u(x), \quad (18a)$$

$$\epsilon_G = E_G/kT \quad (18b)$$

for the number density of electrons and holes,

$$N_e = \int d^3v\, F_{e0}(x,\vec{v}) = N_c\,\mathfrak{F}_{1/2}(\eta) \quad (19a)$$

and

$$N_h = \int d^3v\, F_{h0}(x,\vec{v}) = N_v\,\mathfrak{F}_{1/2}(-\eta-\epsilon_G), \quad (19b)$$

respectively. The Fermi-Dirac integrals are defined by Blakemore.[20] N_c and N_v are the well-known effective density of states.[20] For the convenience of the reader we give here two properties of the Fermi-Dirac integrals which will be needed later. They are

$$\mathfrak{F}_j(\eta) = \exp(\eta) \ \text{for } \eta\to-\infty \ \text{(classical limit)}, \quad (A)$$

$$\frac{d\mathfrak{F}_j(\eta)}{d\eta} = \mathfrak{F}_{j-1}(\eta). \quad (B)$$

For more details the reader is referred to Blakemore.[20] Before we turn to Eq. (14) and a determination of the quantum corrections F_1 and F_2, we would like to point out that the set of equations (8), (10), and (11) has been reduced to a determination of only one function, i.e., $U(x)$ or equivalently $\eta(x)$, the conduction-band edge.[21] To proceed

now, we have to solve Eq. (8) with the help of Eq. (14). Indicating the derivative with respect to x by means of a prime, we have first from Eq. (18a) the important relationship for the electric-field vector $\vec{E}\equiv(E,0,0)$:

$$E=-(kT/e)\eta'=U'. \tag{20}$$

To find F_1 and F_2 as defined by Eq. (14), all one has to do is to insert Eq. (14) into Eq. (8) and equate like powers of $\hbar$ and solve the ensuing equations for F_1 and F_2 keeping in mind Eq. (16) for F_{e0} which obviously satisfies the equation

$$\left(v\frac{\partial}{\partial x}-\frac{1}{m_e}\frac{\partial U}{\partial x}\right)F_{e0}=0. \tag{21}$$

The details of such a calculation have been given in Ref. 12. With the definitions

$$F_{e0}^{(n)}=\frac{d^n}{dy^n}F_{e0}\left(\frac{m}{kT}y\right)\Big|_{y=v^2/2+U}, \tag{22}$$

we merely quote the result

$$F_1=-(i/2m_e^2)EvF_{e0}^{(2)} \tag{23a}$$

and

$$F_2=-\frac{e}{4m_e^3}E'F_{e0}^{(2)}-\frac{e}{6m_e^3}\left(E'v^2+\frac{e}{m_e}E^2\right)F_{e0}^{(3)}-\frac{e^2}{8m_e^4}E^2v^2F_0^{(4)}-\frac{2\pi e^2}{m_e^3\epsilon}g_0F_{e0}^{(1)}, \tag{23b}$$

with

$$g_0=\int d^3v'\,|\vec{v}-\vec{v}'|^{-2}F_{e0}(x,\vec{v}'). \tag{23c}$$

Equation (23b) is identical with Eq. (35) of Ref. 12 if the angular velocity of the Fermi gas considered there is set equal to zero.

From Eq. (11) it is clear that we need to evaluate the integrals of the distribution functions over velocity space. Using the fact that

$$F_{e0}^{(n)}=\left(-\frac{m_e}{kT}\right)^n\frac{d^nF_{e0}}{dy^n}, \tag{24}$$

which follows from definition (22), and using Eq. (19a), it is not difficult to discover that (the term F_1 does not contribute)

$$\int d^3v\,(F_{e0}+\hbar^2F_2)=N_c\mathfrak{F}_{1/2}(\eta)-\frac{e\hbar^2E'}{12m_e(kT)^2}N_c\mathfrak{F}_{-3/2}(\eta)+\frac{e^2\hbar^2E^2}{24m_e(kT)^3}N_c\mathfrak{F}_{-5/2}(\eta)+\cdots, \tag{25}$$

where the dots represent the exchange term which can easily be evaluated using the definition (22) rather than (24) and is given by

$$[\pi e^2\hbar^2/m_e(kT)^3\epsilon]N_c^2[\mathfrak{F}_{-1/2}(\eta)]^2. \tag{26}$$

The fundamental equation for the electric field E or for that matter the potential U follows now from Eqs. (20) and (11). The latter may be written in terms of η as follows:

$$\eta''=\frac{4\pi e^2}{kT}\left(N_c\mathfrak{F}_{1/2}(\eta)-N_v\mathfrak{F}_{1/2}(-\epsilon_G-\eta)+\frac{\hbar^2N_c}{24m_ekT\eta'}(\eta'^2\mathfrak{F}_{-3/2})'+\frac{\pi e^2\hbar^2}{m_e(kT)^2\epsilon}N_c^2[\mathfrak{F}_{-1/2}(\eta)]^2+N_A-N_D\right). \tag{27}$$

We now introduce two characteristic lengths, the Debye length

$$\lambda_D=(\epsilon kT/8\pi e^2N_c)^{1/2}, \tag{28a}$$

and the "thermal" de Broglie length

$$\lambda_{\text{th}}=(\hbar^2/2m_ekT)^{1/2}. \tag{28b}$$

The equation for η can now be written more concisely

$$\eta''=\tfrac{1}{2}\lambda_D^{-2}\left[\mathfrak{F}_{1/2}(\eta)-\frac{N_v}{N_c}\mathfrak{F}_{1/2}(-\eta-\epsilon_G)+\frac{\lambda_{\text{th}}^2}{3\eta'}[\eta'^2\mathfrak{F}_{-3/2}(\eta)]'+\frac{1}{4}\left(\frac{\lambda_{\text{th}}}{\lambda_D}\right)^2[\mathfrak{F}_{-1/2}(\eta)]^2+\frac{N_A-N_D}{N_c}\right]. \tag{29}$$

The third term on the right-hand side of Eq. (29) is due to quantum-mechanical diffraction effects (sometimes called inhomogeneity correction) and the fourth term is due to exchange mediated by the electron-electron interaction. The differential equation (29) possesses a unique solution for η if appropriate boundary conditions are imposed. Far away from the junction or transition layer charge neutrality prevails. For any junction this means that for $x=-\infty$ (deep inside the n-type material we must have with $\eta(-\infty)=\eta_-$:

$$N_c\mathfrak{F}_{1/2}(\eta_-)+\frac{1}{4}\left(\frac{\lambda_{\text{th}}}{\lambda_D}\right)^2N_c[\mathfrak{F}_{-1/2}(\eta_-)]^2=N_D(-\infty). \tag{30a}$$

In Eq. (30a) hole and acceptor contributions have been neglected. Also the electric field vanishes [$\eta'=0$ according to Eq. (20)]. For $x=+\infty$ (deep inside the p-type material), setting $\eta(+\infty)=\eta_+$, neglecting electron as well as donor contributions, and using the fact that the hole distribution is nondegenerate, we have

$$N_v\exp(-\epsilon_G-\eta_+)=N_A(+\infty). \tag{30b}$$

Both $N_D(-\infty)$ and $N_A(+\infty)$ are given quantities.

With both N_A and N_D as given functions of x, Eq.

(29) together with the boundary conditions (30) provide for a unique solution of η and therefore, the electric field and the carrier concentrations throughout the junction. Unfortunately, Eq. (29) cannot be integrated analytically. However if $N_A - N_D$ is constant or a simple function of η over a wide range of x, a first integral of Eq. (29) can readily be found. In the following we shall assume an abrupt n-p junction. Looking back to Eqs. (12) and noting that the hole concentration is nondegenerate, so that $\epsilon_A - \epsilon_F \ll 1$, we have

$$N_A = \begin{cases} N_{A0} & \text{for } x>0\,, \\ 0 & \text{for } x<0\,, \end{cases} \tag{31}$$

where N_{A0} is constant. However for the n side of the junction $x<0$ we must be careful. Since the electron concentration may be degenerate, $\epsilon_F - \epsilon_D$ may even be greater than one. The energy levels of the donors E_D follow the conduction band edge.[22] In other words

$$\epsilon_F - \epsilon_D = (E_F - U + U - E_D)/kT = \eta - \epsilon_D'\,, \tag{32}$$

with a constant ϵ_D'. We therefore have for the donor number density

$$N_D = \begin{cases} 0 & \text{for } x>0\,, \\ N_{D0}[1+\beta_D \exp(\eta - \epsilon_D')]^{-1} & \text{for } x<0\,, \end{cases} \tag{33}$$

with constant N_{D0}.

It is now easy to integrate Eq. (29) by standard methods. For $x>0$ (p-type material)

$$\eta'^2 = \lambda_D^{-2}\left[\mathfrak{F}_{3/2}(\eta) + \frac{N_v}{N_c}\mathfrak{F}_{3/2}(-\eta - \epsilon_G) + \tfrac{1}{3}\lambda_{\text{th}}^2(\eta')^2\mathfrak{F}_{-3/2}(\eta) + \frac{1}{4}\left(\frac{\lambda_{\text{th}}}{\lambda_D}\right)^2\phi_+(\eta) + \frac{N_{A0}}{N_c}\eta + \frac{C_+}{N_c}\right]. \tag{34}$$

Here $\phi(\eta)$ is defined by

$$\phi_+(\eta) = \int_{\eta_+}^{\eta}[\mathfrak{F}_{-1/2}(u)]^2\,du\,. \tag{35}$$

The integration constant C_+ is determined by the requirement that at $x=+\infty$ (for $\eta=\eta_+$) the electric field vanishes, which signifies that $\eta'=0$ at $x=+\infty$ [see Eq. (20)]. It follows that

$$C_+ = -[N_{A0}\eta_+ + N_c\,\mathfrak{F}_{3/2}(\eta_+) + N_v\,\mathfrak{F}_{3/2}(-\eta_+ - \epsilon_G)]\,. \tag{36}$$

For $x<0$ (n-type material)

$$\eta'^2 = \lambda_D^{-2}\left[\mathfrak{F}_{3/2}(\eta) + \frac{N_v}{N_c}\mathfrak{F}_{3/2}(-\eta - \epsilon_G) + \tfrac{1}{3}\lambda_{\text{th}}^2(\eta')^2\mathfrak{F}_{-3/2}(\eta) + \frac{1}{4}\left(\frac{\lambda_{\text{th}}}{\lambda_D}\right)^2\phi_-(\eta) - \frac{N_{D0}}{N_c}\{\eta - \ln[1+\beta_D\exp(\eta - \epsilon_D')]\} + \frac{C_-}{N_c}\right]. \tag{37}$$

ϕ_- is given by

$$\phi_-(\eta) = \int_{\eta_-}^{\eta}[\mathfrak{F}_{-1/2}(u)]^2\,du\,, \tag{38}$$

and the integration constant C_- is again obtained from the requirement of a vanishing electric field at $x=-\infty$ and is given by

$$C_- = N_{D0}\{\eta_- - \ln[1+\beta_D\exp(\eta_- - \epsilon_D')]\} - N_c\,\mathfrak{F}_{3/2}(\eta_-) - N_v\,\mathfrak{F}_{3/2}(-\eta_- - \epsilon_G)\,. \tag{39}$$

The solutions (34) and (37) for the two separate sides of the junction must be joined together at $x=0$. With the aid of Eq. (20) we may rewrite η in terms of the electric field E and have

$$\eta = \eta_1(x) = \eta_- - \frac{e}{kT}\int_{-\infty}^{x}E(y)\,dy \quad (x<0) \tag{40a}$$

and

$$\eta = \eta_2(x) = \eta_+ - \frac{e}{kT}\int_{-\infty}^{x}E(y)\,dy \quad (x>0)\,. \tag{40b}$$

Since η determines the carrier concentration and the latter must be continuous at $x=0$, we have

$$\eta_1(0) = \eta_2(0) = \eta_0\,. \tag{41a}$$

The electric field must also be continuous at $x=0$, therefore

$$\eta_1'(0) = \eta_2'(0) = \eta_0'\,. \tag{41b}$$

From Eqs. (40) and (41a) it now follows that the built in voltage is simply given by

$$V_b = \int_{-\infty}^{+\infty}E(y)\,dy = \frac{kT}{e}(\eta_- - \eta_+)\,, \tag{42}$$

where, of course, η_- and η_+ are determined from Eqs. (30). Conditions (41) together with Eqs. (34) and (37) give

$$(N_{A0}+N_{D0})\eta_0 - N_{D0}\ln[1+\beta_D\exp(\eta_0 - \epsilon_D')] = C_- - C_+ + \frac{1}{4}\left(\frac{\lambda_{\text{th}}}{\lambda_D}\right)^2 N_c\int_{\eta_-}^{\eta_+}[\mathfrak{F}_{-1/2}(u)]^2\,du\,, \tag{43}$$

a transcendental equation for the determination of η_0. Once η_0 is known, η_0' or equivalently the electric field at $x=0$ can be determined via Eq. (34) or (38). The electric field at $x=0$ is the maximum electric field because E monotonically increases in going to $x=0$ from either side of the junction and E' is discontinuous at $x=0$.

Unfortunately a second integration of Eqs. (34) and (37) cannot be performed analytically even in the completely nondegenerate case. We therefore confine ourselves to a determination of the maximum electric field and the carrier concentrations at $x=0$ in Sec. IV.

IV. DISCUSSION AND GENERALIZATIONS

With the help of the equations derived in Sec. III, we determine various quantities for nondegenerate junctions. In this case quantum effects can safely be neglected. This is formally achieved by putting $\lambda_{th}=0$. Also $\mathfrak{F}_j(\eta)=e^{\eta}$ throughout. Equations (3) then yield

$$\eta_- = \ln(N_{D0}/N_c),\ \eta_+ = \ln(N_v/N_{A0}) - \epsilon_G, \quad (44)$$

so that the built-in voltage is given by [from Eq. (42)]

$$V_b = (kT/e)\ln(N_{D0}N_{A0}/n_i^2), \quad (45)$$

since $n_i^2 = N_c N_v \exp(-\epsilon_G)$. Furthermore, the electron number density at $x=0$ is given by

$$N_{e0} = N_{D0}\left(\frac{n_i^2}{N_{D0}N_{A0}}\right)^{N_{A0}/(N_{D0}+N_{A0})} \times \exp\left(-\frac{N_{D0}-N_{A0}}{N_{D0}+N_{A0}}\right). \quad (46)$$

This follows with a little algebra from Eqs. (36), (39), and (43). The maximum electric field computes now from Eq. (37), for instance, to be

$$E_{\max} = E(0) = (8\pi kTN_{D0}/\epsilon)^{1/2} \times(\kappa - 1 - \ln\kappa)^{1/2}\ (\text{statvolt/cm}), \quad (47)$$

where $\kappa = N_{e0}/N_{D0}$. Equation (45) is well known, but Eqs. (46) and (47) are not. The depletion-layer theory gives

$$\overline{E}_{\max} = (8\pi kTN_{A0}/\epsilon)^{1/2}[\ln(N_{A0}N_{D0}/n_i^2)]^{1/2}. \quad (48)$$

A quick comparison with typical values $n_i = 1.6\times10^{10}\ \text{cm}^{-3}$, $N_{A0}=10^{15}\ \text{cm}^{-3}$, $N_{D0}=10^{18}\ \text{cm}^{-3}$, $\epsilon = 11.8$, $T=300\ ^\circ\text{K}$ (typical for Si) gives

$$E_{\max} = 5.53\times10^4\ \text{V/cm},\quad \overline{E}_{\max} = 8.17\times10^4\ (\text{V/cm}), \quad (49)$$

a difference by a factor of 1.5 between the two values. Since the built in voltage is the same in both theories, the higher value of $E_{\max}$ signifies a narrower transition region for the depletion-layer theory.

A nondegenerate abrupt high-low junction may be treated in the same manner. A junction consisting of p-type material in which the side $x<0$ is more heavily doped (acceptor concentration N_{A+}) than the other side ($x>0$) can be evaluated, using the formalism developed in Sec. III, by merely replacing Eq. (30a) by an equation analogous to Eq. (30b), namely,

$$N_v \exp(-\epsilon_G - \eta_-) = N_{A+}, \quad (50)$$

and ignoring the electron concentration throughout. The built in voltage V_b is now given by[23]

$$V_b = (kT/e)\ln(N_{A0}/N_{A+}), \quad (51)$$

a well-known expression. From Eq. (43) it follows after some algebra that

$$N_{h0} = N_h(0) = \exp^{-1} N_{A0}(N_{A0}/N_{A+})^{N_{A+}/(N_{A0}-N_{A+})}. \quad (52)$$

The maximum electric field is given by

$$E_{\max} = (8\pi kTN_{A0}/\epsilon)^{1/2} \times(\kappa - 1 - \ln\kappa)^{1/2}\ (\text{statvolt/cm}), \quad (53)$$

where κ is now given by $\kappa = N_{h0}/N_{A0}$ [N_{h0} from Eq. (52)]. There is no counterpart for Eq. (53) in the depletion-layer theory for the simple reason that a high-low junction does not possess a depletion layer. With $N_{A+}=10^{18}\ \text{cm}^{-3}$ and all other pertinent values unchanged it turns out that the maximum electric field is $E_{\max}=5.46\times10^4$ (V/cm) for the p-p^+ junction, a value pretty much the same as that calculated for the n-p junction *with* the assumption of a depletion layer [$\overline{E}_{\max}$ of Eq. (49)] and therefore 50% larger than the value for the n-p junction assuming the validity of the Thomas-Fermi approach of this paper.

We turn now to a discussion of degenerate n-p junctions. As the dopant level of the n side of the junction increases, quantum effects become more and more important.

One of the boundary conditions [Eq. (30a)] now reads

$$\mathfrak{F}_{1/2}(\eta_-) + \frac{1}{4}\left(\frac{\lambda_{th}}{\lambda_D}\right)^2[\mathfrak{F}_{-1/2}(\eta_-)]^2 = \frac{N_{D0}}{N_c}[1+\beta_D\exp(\epsilon_F - \epsilon_D)]^{-1}, \quad (54)$$

using Eq. (12a) and neglecting hole and acceptor contributions because of their smallness. The second term on the left-hand side of Eq. (54) is due to exchange and is missing in the customary treatment.[20] The significance of this term is the following. Equation (54) is valid deep inside the n material far away from the junction. If we call $\eta_-^{(0)}$ the solution of Eq. (54) without the exchange term, or

$$\mathfrak{F}_{1/2}(\eta_-^{(0)}) = (N_{D0}/N_c)[1+\beta_D\exp(\eta_-^{(0)} - \epsilon'_D)]^{-1} \quad (55)$$

and the solution for η_- of Eq. (54) $\eta_-^{(1)}$, then we have

$$\eta_-^{(0)} = \epsilon_F - \epsilon_C^{(0)},\quad \eta_-^{(1)} = \epsilon_F - \epsilon_c^{(1)} \quad (56)$$

and

$$\epsilon_c^{(0)} - \epsilon_c^{(1)} = \eta_-^{(1)} - \eta_-^{(0)}$$

or (57)

$$E_c^{(0)} - E_c^{(1)} = kT(\eta_-^{(1)} - \eta_-^{(0)}) < 0.$$

In fact, assuming for the moment that $\eta_-^{(1)} - \eta_-^{(0)}$ is small, we have

$$\mathfrak{F}_{1/2}(\eta_-^{(0)}+\eta_-^{(1)}-\eta_-^{(0)})=\mathfrak{F}_{1/2}(\eta_-^{(0)})+(\eta_-^{(1)}-\eta_-^{(0)})\mathfrak{F}_{-1/2}(\eta_-^{(0)})\,.$$

Treating the exchange term as a perturbation in this case, solving for $\eta_-^{(1)}-\eta_-^{(0)}$, we obtain

$$kT\delta E_c=\eta_-^{(1)}-\eta_-^{(0)}=-\frac{1}{4}\left(\frac{\lambda_{th}}{\lambda_D}\right)^2\mathfrak{F}_{-1/2}(\eta_-^{(0)})\,,\quad \times\left(1-\frac{N_{D0}}{N_c}[1+\beta_D\exp(\eta_-^{(0)}-\epsilon_D')]^{-2}\times\beta_D\exp(\eta_-^{(0)}-\epsilon_D')\mathfrak{F}_{-1/2}^{-1}\right)^{-1} \qquad (58)$$

which is usually negative. We note an increase of the conduction-band edge with *increasing* doping level. The consequences are twofold. The built-in voltage V_b *decreased* as seen from Eq. (42),

$$V_b=(kT/e)(\eta_-^{(0)}-\eta_+)+(1/e)\delta E_c<V_b^{(0)}\,, \qquad (59)$$

since δE_c is negative and η_+ is not changed from its value of Eq. (30b) [the p side of the junction ($x>0$) being nondegenerate]. Another consequence is that tunneling probabilities may be altered (Ref. 17, p. 159). But since this is a nonequilibrium effect we are not pursuing this aspect here any further, in any case the inhomogeneity effect will overshadow the exchange effect right at the edge of the junction where tunneling occurs if the p side of the junction is also degenerate. Although the position of the conduction-band edge is raised relative to its position for the nondegenerate case so is the position of the valence-band edge. This follows from the fact that

$$N_h=N_v F_{1/2}(-\epsilon_G-\eta_-)=N_v\exp(-\epsilon_G-\eta_-)\,, \qquad (60)$$

so that the minimum energy for an electron to jump into a hole by emitting radiation is still $kT\epsilon_G$. No band-gap narrowing occurs. This is not surprising since the band-gap narrowing is primarily due to a broadening of the donor levels into a band which eventually is incorporated into the conduction band thus producing a narrowing of the gap. As to the magnitude of δE_c, suffice it to say that for $N_{D0}\approx 10^{19}\ \mathrm{cm}^{-3}$ in silicon at 300°K it turns out to be of the order of 5 meV which is about an order of magnitude smaller than the ionization energy of the donors phosphorus or arsenic.

TABLE I. Input parameters for the calculation of the maximum electric field.

Input parameter	Value
N_{A0}	10^{16} cm^{-3}
N_{D0}	Variable
N_c	2.8×10^{19} cm^{-3}
N_v	1.05×10^{19} cm^{-3}
kT	0.0259 eV
β_D	2 (spin multiplicity)
ϵ_D'	1.7 (donor level)
ϵ_G	43.24 (band gap)
ϵ	11.8 (dielectric constant)
λ_{th}	12.2 Å (using free-electron mass)
λ_D	5.5 Å

TABLE II. Maximum electric field as a function of donor concentration computed with and without quantum corrections.

N_{D0} (cm^{-3})	E_{max} (V/cm) classical	E_{max} (V/cm) with quantum corrections	% difference
0.43×10^{19}	1.08×10^5	0.90×10^5	16%
0.74	1.48	1.33	10%
1.36	1.80	1.65	8%
2.50	2.27	2.12	7%
4.60	2.83	2.66	6%
8.55	3.33	3.13	6%

From the preceding pages it is easy to see how the maximum electric field in the transition region between n- and p-type material is computed. From the given values of donor and acceptor concentrations, Eqs. (30) and (54), we compute the quantities η_- and η_+. Then turning to Eqs. (36) and (39) the integration constants C_- and C_+ are determined. Once this is done, we turn to Eq. (43) to obtain η_0, the conduction-band edge at $x=0$. Finally η_0' is calculated via Eq. (34) or (37) and from η_0' the value of E_{max}, the maximum electric field is extracted through the definition (20). We have performed such a calculation with the input parameters for a model of a silicon n-p junction listed in Table I. The result of the calculations is listed in Table II. From Table II it is seen that the maximum electric field is significantly smaller when quantum corrections are taken into account. The quantum corrections consist of two parts, the inhomogeneity correction [the third term on the right-hand side of Eq. (29)] and the exchange correction [the fourth term of the right-hand side of Eq. (29)]. Both tend to increase as the dopant concentration is increased. But their effect on the electric field is in opposite directions. The inhomogeneity contribution tends to increase the field, whereas the exchange contribution tends to decrease the built-in electric field.

This decrease of the electric field is a manifestation of the Pauli principle, the electrons, being screened from each other, occupy a larger volume and lower their mutual interaction energy thus giving rise to a smaller potential energy difference between n and p material [see Eq. (59)] and consequently a lowering of the maximum electric field.

On the other hand, the larger the donor concentration leaving the acceptor concentration unchanged on the other side of the junction, the larger the inhomogeneity across the junction and consequently the larger the contribution of the inhomogeneity correction becomes compared to the exchange contribution and indeed, as the last column shows the difference between the classical and quantum-mechanical calculations becomes smaller with an increasing donor concentration. But this does not mean that ultimately quantum and classical calculations coalesce. It simply means that the expansion (14) has been terminated too early, only terms up to order h^2 having been considered here. But going to higher orders in h is not warranted since at higher built-in electrical fields the very concept of an abrupt junction becomes questionable. The core electronic wave function will spread through the whole transition region and the simple self-consistent Thomas-Fermi approximation, at least in its present form, will break down. Also, at the necessarily high dopant concentrations the host lattice will be strained in such a manner that the effective mass equations underlying this theory[3] become invalid.

V. SUMMARY

On the preceding pages the Thomas-Fermi self-consistent approximation within the context of a quantum-mechanical phase-space distribution function was applied to a semiconductor junction in thermal equilibrium. Even in the simplest case of a one-dimensional abrupt junction, the differential equation for the determination of the free-carrier number densities and the electric field as a function of x, the distance from the (discontinuous) transition from n-type to p-type material, proved not to be amenable to an analytic solution whether or not the carrier concentration is degenerate. However, the expressions for the maximum electric field induced by the polarization of the junction have been shown to be simple enough to lie within easy reach of a pocket calculator. The most important result of the analysis of Sec. IV is the fact that as soon as degeneracy sets in ($N_D > 10^{18}$ cm^{-3} for silicon) quantum effects, although not all important, are by no means negligible (see Table II). In tracing back the origin of these quantum effects (inhomogeneity and exchange contributions), we find the pivotal equation (27), Poisson's equation, to contain these effects. But Poisson's equation constitutes an integral part of all *non*equilibrium analyses. It is therefore obvious that quantum effects must be incorporated into any device analysis which uses a high degree of sophistication.[24]

As already stated in Sec. I certain effects which become important at very high dopant concentrations have not been incorporated into the present analysis. The theory does not take into account donor-level broadening and the ensuing interaction with the conduction band (band-gap narrowing and tailing). It is therefore confined in its application to donor number densities less than 10^{20} cm^{-3} (for silicon). It is felt that this is no serious drawback since in many applications (for instance solar cells) doping concentrations of 10^{20} cm^{-3} are rarely encountered.

---

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by NASA.

[1]This is true for wide-band-gap meterials like Si, Ge, and GaAs, for instance.

[2]E. O. Kane, Phys. Rev. 131, 79 (1963).

[3]J. C. Slater, Phys. Rev. 76, 1592 (1949), Sec. IV.

[4]J. B. Gunn, J. Electron. Control 4, 17 (1958); H. K. Gummel, IEEE Trans. Electron Devices ED-455 (1964); A. DeMari, Solid State Electron. 2, 1021 (1968); S. C. Choo, IEEE Trans. Electron Devices ED-18, 574 (1971); R. D. Middlebrook, Solid State Electron. 6, 555 (1963); W. van Roosbroeck, Phys. Rev. 123, 474 (1961); J. G. Fossum, Solid State Electron. 19, 269 (1976). This list is not exhaustive but contains some of the more important papers on the subject.

[5]W. Shockley, Bell Syst. Tech. J. 28–29, 435 (1949–1950).

[6]It is impossible to quote all the literature on the subject of the depletion-layer approximation. Instead one representative textbook must suffice: W. R. Beam, *Electronics of Solids* (McGraw-Hill, New York, 1965), p. 231.

[7]See, for instance, A. K. Jonscher, J. Electron. Control 5, 1 (1958).

[8]T. P. Brody, J. Appl. Phys. 33, 100 (1962).

[9]See, in this context, M. H. Cohen, H. Fritzsche, and S. R. Ovshinsky, Phys. Rev. Lett. 22, 1065 (1969).

[10]O. von Roos, Phys. Rev. 119, 1174 (1960); 124, 71 (1961).

[11]O. von Roos and J. S. Zmuidzinas, Phys. Rev. 121, 941 (1961).

[12]P. H. Levine and O. von Roos, Phys. Rev. 125, 207 (1962).

[13]E. H. Lieb and B. Simon, Phys. Rev. Lett. 31, 681 (1973). This paper elucidates the Thomas-Fermi approach still further.

[14]Equation (8) does not contain any correlation effects. As pointed out in Ref. 12 the incorporation of correlations (deviation of the two body distribution function from a product of one body functions) is exceedingly difficult and furthermore it has not been done for an

inhomogeneous Fermi-Dirac gas as of this writing. Physically the incorporation of correlation effects signifies the consideration of three-body collisions via screened Coulomb potentials in the approximation to be adopted in this paper. Three-body collisions are ignored in the following.

[15]Comparing Eq. (8) with its analog equation (1) of Ref. 12 we note the term $(N-1)/N$ missing. This is because $N >> 1$ in the case contemplated here.

[16]The best way to see this is via the formalism of second quantization. Another, less formal but more cumbersome approach is given in W. Shockley, *Electrons and Holes in Semiconductors* (D. van Nostrand, New York, 1950), p. 440.

[17]See, for instance, S. M. Sze, *Physics of Semiconductor Devices* (Wiley-Interscience, 1969), p. 34.

[18]This is really no restriction. It can easily be removed and is only motivated by the fact that most solar cells consist of very highly doped n-type material diffused into p-type material for no other reason than ease of manufacture.

[19]M_c is the number of equivalant bands (6 for silicon).

[20]J. S. Blakemore, *Semiconductor Statistics* (Pergamon, New York, 1962).

[21]We see now clearly that the reduction from three equations to essentially one for the determination of the only unknown function $U(x)$ is not possible under *nonequilibrium* conditions, for in that case, Eqs. (8) and (10) contain additional collision integrals and the functional $\frac{1}{2}mv^2 \pm U$ is not a characteristic of the first-order partial differential equations any more.

[22]This fact is not generally realized. However it is immaterial for nondegenerate junctions for which most calculations have been performed.

[23]V_b is negative because the electric field points into the negative direction, from the N_A side to the N_{A+} side of the junction.

[24]J. R. Hauser and P. M. Dunbar, NASA report NASA-CA-138828 (North Carolina State University Grant No. NGR34-002-195, 1974) (unpublished).