Theory of the junction capacitance of an abrupt diode

P. Van Mieghem, R. P. Mertens, and R. J. Van Overstraeten
Interuniversity Micro Electronics Center, Kapeldreef 75, B3030 Leuven, Belgium

(Received 18 September 1989; accepted for publication 11 January 1990)

A new theory for the junction capacitance of mathematically abrupt diodes is presented. In contrast with previous theories, Fermi–Dirac statistics are applied, and instead of using a parabolic density of states, a more appropriate function can be taken into account as this is required for heavily doped material. The main approximation is that of constant quasi-Fermi levels. Besides the study of symmetrical junctions, the behavior of asymmetrical abrupt junctions is both analyzed and explained. Anomalies resulting from the calculations are shown to be due to the mathematical discontinuity.

I. INTRODUCTION

Many reports about capacitances of diodes[1–12] have been published. Mainly linearly graded junctions were studied. As very abrupt junctions were only theoretically attractive, but practically impossible to fabricate, less effort was devoted to them. However, all the published work on the capacitance of an abrupt junction shows the inability to handle highly asymmetrical junctions. The dissimilarity between processed junctions, which were supposed to be abrupt, and their mathematical description was never put forward as the main reason for the observed discrepancy between experiment and theory.

Here we present a new theory and explain the behavior of highly asymmetrical mathematically abrupt junctions. This work was initiated while studying band-gap narrowing from the capacitance intercept voltage of an abrupt junction. A very accurate intercept voltage formula was required as this turns out to be a crucial point in determining band-gap narrowing from capacitance data. Results of the band-gap shrinkage in GaAs using an intercept voltage formula derived from the proposed theory are presented in Ref. 13.

In Sec. II, we develop a general capacitance formula for both continuous and discontinuous doping profiles. Moreover, we present a new $\wp$ transformation that generalizes the Fermi Dirac integrals to an arbitrary density of states in order to describe a heavily doped material properly. An outline of the calculation is given in Sec. III and Sec. IV evaluates the new formula numerically. These numerical results are explained by examining the behavior of a simplified analytical capacitance formula which is derived from the general procedure.

II. DERIVATION OF A GENERAL JUNCTION CAPACITANCE FORMULA

A. Continuous doping profile

The charge density $\rho(x)$ in a diode is given by

$$\rho(x) = q[\, p(x) - n(x) + N(x)\,] \tag{1}$$

with $p(x) - n(x)$ the number of positive free charges and $N(x)$ the net doping concentration of ionized impurities. We concentrate on a one-dimensional *pn* diode with the origin of the coordinate axis x at the metallurgical junction and we call the border points in the n-type region and p-type region $x = -L$ and $x = K$, respectively. If the doping profile is continuous, the net doping concentration at the metallurgical junction $N(0)$ equals zero. The number of electrons $n(x)$ and the number of holes $p(x)$ are related to the electric potential $\psi(x)$ and to the quasi-Fermi levels $\phi_n(x)$ and $\phi_p(x)$ through the Fermi Dirac statistics. In general, one can write (see the Appendix)

$$n = N_C \wp[\, g_c, (\zeta_n - E_c)/kT\,] , \tag{2}$$

$$p = N_V \wp[\, g_v, (E_v - \zeta_p)/kT\,] , \tag{3}$$

with $g_c(E)$ and $g_v(E)$ the density of states in the conduction and valence band; N_C and N_V the effective density of states in the conduction and in the valence band, respectively; ζ_n and ζ_p the quasi-Fermi energy levels for electrons and holes;

$$\wp[\, g(\xi), y] = \int_{-\infty}^{\infty} \frac{g(\xi)}{1 + \exp(\xi - y)}\, d\xi ;$$

and E_c and E_v the nondistorted band edges (see the Appendix).

The energy quantities can be converted into potential quantities, using the relations

$$E_c = -q[\psi - E_\gamma] ,$$

$$E_v = -q[\psi + E_\gamma] ,$$

$$E_\gamma = E_g/2q , \tag{4}$$

with E_γ the half-band gap expressed in volts and

$$\zeta_n = -q\phi_n ,$$

$$\zeta_p = -q\phi_p . \tag{5}$$

It can be shown[14] that the charge density $\rho(x)$ is continuous only if the band gap and the doping profile are continuous functions of the distance x. The net amount of charge $Q(x)$ between the n-type border and a point x in the diode reads

$$Q(x) = \int_{-L}^{x} \rho(t)\, dt \tag{6}$$

and due to the overall neutrality, we find

$$Q(K) = 0 . \tag{7}$$

Invoking the Poisson equation

$$\frac{d^2\psi}{dx^2} = -\frac{\rho(x)}{\epsilon} , \tag{8}$$

where the dielectric permittivity ϵ is assumed constant over

 0021-8979/90/094203-09$03.00

the diode structure, and with the definition of the electric field $E(x)$ as

$$E(x) = -\frac{d\psi(x)}{dx}, \tag{9}$$

we can rewrite (6)

$$\int_{-L}^{x} \rho(t)\,dt = \epsilon E(x) - \epsilon E(-L)\,. \tag{10}$$

Both the electric field $E(x)$ and its gradient at $x = -L$ and $x = K$ are zero because the structure is assumed to be long compared to the depletion regions. This assumption causes the electric field $E(x)$ to be proportional to $Q(x)$:

$$E(x) = \frac{1}{\epsilon}\int_{-L}^{x} \rho(t)\,dt = \frac{1}{\epsilon}Q(x)\,. \tag{11}$$

For $\rho(x)$ being continuous, the electric field achieves its maximum at X_0 which satisfies $\rho(X_0) = 0$. Since the electric field in an arbitrary diode structure has only one maximum, the charge density $\rho(x)$ changes sign only once inside $(-L,X)$, and, consequently

$$\begin{aligned} &\rho(x) > 0\,, \quad \text{for } x < X_0\,,\\ &\rho(x) < 0\,, \quad \text{for } x > X_0\,. \end{aligned} \tag{12}$$

The junction capacitance C is defined as the absolute value of the change with the applied voltage V_a of the total positive (or negative) charge. Hence, combining Eqs. (7), (11), and (12),

$$\begin{aligned} C &= \left|\frac{d}{dV_a}\left(\int_{-L}^{X_0} \rho(x)\,dx\right)\right| = \left|\frac{d}{dV_a}\left(\int_{X_0}^{K} \rho(x)\,dx\right)\right| \\ &= \left|\epsilon\frac{dE(X_0)}{dV_a}\right|\,. \end{aligned} \tag{13}$$

This is the general definition for the junction capacitance of a homojunction if the doping profile is continuous at all points x of the structure. As this is in reality always the case, (13) is thus always applicable.

B. Discontinuous doping profile

Let us concentrate on the nonrealistic doping profile of an abrupt junction that is discontinuous at $x = 0$. In general, using Heavyside's step function $\theta(x)$ defined by

$$\theta(x) = \begin{cases} 0, & x < 0,\\ \frac{1}{2}, & x = 0,\\ 1, & x > 0, \end{cases}$$

we have

$$N(x) = N_D(x)\theta(-x) + N_A(x)\theta(x)\,, \tag{14}$$

with $N_D(x)$ and $N_A(x)$ differentiable functions of x. Since $N_A(x)$ and $N_D(x)$ are continuous in their defined region, arguments from Sec. II A let us conclude that the charge density function $\rho(x)$ is still a continuous function with a single discontinuity at the metallurgical junction $(x = 0)$ where the doping profile $N(x)$ is discontinuous.

In the following, we study the special case where both $N_A(x)$ and $N_D(x)$ are constant functions. Only under these restrictions can an analytical calculation be performed. We list some properties of an abrupt junction with constant doping profiles.

(1) In low injection [i.e., $N_A > p(x)$ and $N_D > n(x)$] the charge density $\rho(x)$ has no zero between $x = -L$ and $x = K$ and sign $[\rho(-x)] = -\text{sign}[\rho(x)]$.

(2) In low injection the derivative of the charge density $\rho(x)$ with respect to x does not change sign except at the metallurgical junction $(x = 0)$.

Finally, the junction capacitance formula for abrupt homojunctions results in

$$C = \epsilon\frac{dE(0)}{dV_a}\,. \tag{15}$$

Although the properties above are easy to accept, their proofs can be found in Ref. 14. (For plots of the charge density of an abrupt junction we refer to Fig. 5 and Ref. 15.)

III. A CAPACITANCE FORMULA FOR AN ABRUPT JUNCTION

A. General procedure

Applied to solid-state physics, the Poisson equation is a nonlinear differential equation that can only be solved analytically under simplifying conditions. The most difficult quantities appearing in the Poisson equation are the quasi-Fermi potentials ϕ_n and ϕ_p because their values can only be obtained by simultaneously solving both the current equations and the Poisson equation.

The simplifying conditions listed below will be discussed briefly.

(1) The band gap E_g is assumed only to be a function of the semiconductor and of the doping concentration.

(2) The quasi-Fermi potentials ϕ_n and ϕ_p are taken constant over the diode structure, which enables analytical integration of the Poisson equation.

For constant doping profiles in both the n- and p-type regions, the first assumption implies constant band gaps, E_{gn} and E_{gp}, respectively. A consequence of this is the high degree of linearity of the function $C^{-2}(V)$ in the voltage region corresponding to low injection. For, if we perform the derivation of the capacitance procedure below formally (which means that we assume the band gap to be a certain function of applied voltage, electrostatic potential, and position without knowing the precise dependence), we find the slope of the function $C^{-2}(V)$ dependent on dE_g/dV_a and dE_g/dx, both functions of the applied voltage. Consequently, in reality, the slope is not constant. If this band-gap dependence on applied voltage were actually pronounced, we would find deviations from a straight line in the experimental results. As the experimental C-V data for GaAs symmetrical abrupt diodes[13] remain very linear with voltage, the first approximation is well justified.

The study of Sah[16] shows that the assumption of constant quasi-Fermi levels is a good approximation for the voltage range we are interested in. This is confirmed by Nuyts and Van Overstraeten.[6] Moreover, it can be demonstrated[14] that the approximation of constant quasi-Fermi levels is best suited for symmetrical junctions.

Since $\phi_p(K) - \phi_n(-L) = V_a$ we can choose the following potential reference:

$$\phi_p = V_a/2\,,$$
$$\phi_n = -V_a/2\,. \tag{16}$$

The Poisson equation deals with potentials and not with energies; therefore, we transform all quantities in the proper form, employing (16), (4), and (5) to obtain

$$\frac{E_v - \zeta_p}{kT} = \frac{q(-\psi - E_\gamma + V_a/2)}{kT} = -u + V,$$
$$\frac{\zeta_n - E_c}{kT} = \frac{q(\psi - E_\gamma + V_a/2)}{kT} = u + V,$$

with

$$u = q\psi/kT\,,$$
$$V = \frac{q(V_a/2 - E_\gamma)}{kT}\,. \tag{17}$$

Here we introduce the normalized potential u and a new formalism for the expression of the free carriers (2) and (3) as explained in the Appendix. After normalization, the Poisson equation (8) for an abrupt junction reads

$$\frac{d^2u}{dx^2} = \frac{q^2}{\epsilon kT}\,[N_C \wp(g_c, V_D + u) - N_V \wp \times (g_v, V_D - u) - N_D]\,, \quad \text{for } x < 0\,,$$
$$\frac{d^2u}{dx^2} = \frac{q^2}{\epsilon kT}\,\{N_C \wp(g_c, V_A + u) - N_V \wp \times (g_v, V_A - u) - N_A\}, \quad \text{for } x > 0\,, \tag{18}$$

where $V_\lambda = q(V_a/2 - E_{\gamma\lambda})/kT$ for λ referring to the donor side (D) or acceptor side (A). The boundary conditions expressed in the normalized potential u become

$$\frac{du(-L)}{dx} = \frac{du(K)}{dx} = 0\,,$$
$$\frac{d^2u(-L)}{dx^2} = \frac{d^2u(K)}{dx^2} = 0\,. \tag{19}$$

Involving (18) and (19), implicit equations for the normalized potential at the border points u_L and u_K are obtained,

$$N_C \wp(g_c, V_D + u_L) - N_V \wp(g_v, V_D - u_L) - N_D = 0\,,$$
$$N_C \wp(g_c, V_A + u_K) - N_V \wp(g_v, V_A - u_K) + N_A = 0\,. \tag{20}$$

In low injection, the minority carriers far enough from the metallurgical junction are negligible with respect to the number of majority carriers,

$$n(K) = N_C \wp(g_c, V_A + u_K) \approx 0$$

and

$$p(-L) = N_V \wp(g_v, V_D - u_L) \approx 0\,. \tag{21}$$

Including (21) and putting $\wp(g_\lambda, y) = R_\lambda(y)$, (20) can be solved explicitly:

$$u_L = -V_D + R_C^{-1}(N_D/N_C)\,,$$
$$u_K = V_A - R_V^{-1}(N_A/N_V)\,, \tag{22}$$

where $R^{-1}(y)$ denotes the inverse of the transformed function with respect to y as explained in the Appendix.

One obtains the electric field by integrating (18). Concentrating on the first equation (an analogous computation can be performed for the second) and using the boundary conditions (19) and the property (A10), we have

$$\left(\frac{du}{dx}\right)^2 = \frac{2q^2}{\epsilon kT}\left[N_C \wp\left(\int_{-\infty}^{\xi} g_c(t)\,dt, V_D + u\right) - N_C \wp \times \left(\int_{-\infty}^{\xi} g_c(t)\,dt, V_D + u_L\right) + N_V \wp\left(\int_{-\infty}^{\xi} g_v(t) \times dt, V_D - u\right) - N_V \wp\left(\int_{-\infty}^{\xi} g_v(t)\,dt, \times V_D - u_L\right) - N_D(u - u_L)\right]. \tag{23}$$

Before continuing, we simplify our notation:

$$\wp\left(\int_{-\infty}^{\xi} g_\lambda(t)\,dt, y\right) = \mathcal{J}_\lambda(y) \tag{24}$$

then (23) becomes

$$\left(\frac{du}{dx}\right)^2 = \frac{2q^2}{\epsilon kT}[N_C \mathcal{J}_c(V_D + u) - N_C \mathcal{J}_c(V_D + u_L) + N_V \mathcal{J}_v(V_D - u) - N_V \mathcal{J}_v(V_D - u_L) - N_D(u - u_L)]\,. \tag{25}$$

The continuity of the electric field at the junction requires

$$\lim_{\zeta \to 0} \frac{du(-\zeta)}{dx} = \lim_{\zeta \to 0} \frac{du(+\zeta)}{dx} \tag{26}$$

such that a useful coupling equation (27), which determines the potential $\mathbf{u_0}$ for every applied voltage V, is obtained:

$$N_C[\mathcal{J}_c(V_D + \mathbf{u_0}) - \mathcal{J}_c(V_D + u_L) - \mathcal{J}_c(V_A + \mathbf{u_0}) + \mathcal{J}_c(V_A + u_K)] + N_V[\mathcal{J}_v(V_D - \mathbf{u_0}) - \mathcal{J}_v(V_D - u_L) - \mathcal{J}_v(V_A - \mathbf{u_0}) + \mathcal{J}_v(V_A - u_K)] - (N_A + N_D)\mathbf{u_0} + (N_D u_L + N_A u_K) = 0\,. \tag{27}$$

We are now in the position to use the capacitance formula for an abrupt junction (15). We concentrate on the n-type region ($x < 0$) and normalize (15) to obtain

$$C = \frac{\epsilon}{2}\frac{d}{dV}\left(\frac{du(0)}{dx}\right),$$

and notice that $d/dV_A = d/dV_D = d/dV$.

The calculation is rather cumbersome, but systematic derivation with respect to V of the square root of (25) evaluated for $u = \mathbf{u_0}$ gives

$$C = \sqrt{\frac{q^2\epsilon}{8kT}}\,\frac{\mathbf{G}}{du^*(0)/dx}\,, \tag{28}$$

where

$$\frac{du^*(0)}{dx} = \sqrt{[N_C \mathcal{J}_c(V_D + v) + N_V \mathcal{J}_v(V_D - v) - N_D v]\Big|_{v = u_L}^{v = \mathbf{u_0}}} \tag{29}$$

with $f(x)|_{x=a}^{x=b} = f(b) - f(a)$ and $\mathbf{G}$ is the derivative of the expression under the square root sign with respect to V or

$$\mathbf{G} = \{N_C R_c(V_D + v) + N_V R_v(V_D - v)\}|_{v=u_L}^{v=\mathbf{u}_0} + [N_C R_c(V_D + \mathbf{u}_0) - N_V R_v \times (V_D - \mathbf{u}_0) - N_D] \frac{d\mathbf{u}_0}{dV}. \quad (30)$$

The only unknown quantity is the derivative of the normalized potential $u(x)$ at the junction with respect to the normalized applied voltage V. By differentiating the coupling equation (27) with respect to V, one obtains

$$\frac{d\mathbf{u}_0}{dV} = \frac{1}{\mathbf{H}}\left(N_C R_c(\eta)\Big|_{\eta = V_A + u_K}^{\eta = V_D + u_L} + N_V R_v(\eta)\Big|_{\eta = V_A - u_K}^{\eta = V_D - u_L} + N_C R_c(\eta)\Big|_{\eta = V_D + \mathbf{u}_0}^{\eta = V_A + \mathbf{u}_0} + N_V R_v(\eta)\Big|_{\eta = V_D - \mathbf{u}_0}^{\eta = V_A - \mathbf{u}_0}\right) \quad (31)$$

and

$$\mathbf{H} = (N_A + N_D) - N_C R_c(\eta)\Big|_{\eta = V_A + \mathbf{u}_0}^{\eta = V_D + \mathbf{u}_0} + N_V R_v(\eta)\Big|_{\eta = V_A - \mathbf{u}_0}^{\eta = V_D - \mathbf{u}_0}.$$

A similar formula for C, based on the p-type region, can be derived analogously.

B. Low injection simplification

In order to be able to evaluate the capacitance of an abrupt junction, we now assume low injection conditions (22) and neglect the very small quantities

$$R_{v(c)}(\eta)\Big|_{\eta = V_A \pm \mathbf{u}_0}^{\eta = V_D \pm \mathbf{u}_0}$$

and

$$\mathcal{J}_{v(c)}(\eta)\Big|_{\eta = V_A \pm \mathbf{u}_0}^{\eta = V_D \pm \mathbf{u}_0}.$$

The normalized potential at the junction $\mathbf{u}_0$ reads

$$\mathbf{u}_0 = \left(\frac{N_A V_A - N_D V_D}{N_A + N_D}\right) + \frac{1}{N_D + N_A}\left\{N_D R_C^{-1}\left(\frac{N_D}{N_C}\right) - N_A R_V^{-1}\left(\frac{N_A}{N_V}\right) - N_C \mathcal{J}_c\left[R_C^{-1}\left(\frac{N_D}{N_C}\right)\right] + N_V \mathcal{J}_v\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right]\right\}. \quad (32)$$

The square of the normalized electric field at $x = 0$ rewrites as

$$\left(\frac{du(0)}{dx}\right)^2 = \frac{2q^2}{\epsilon kT}\left(N_C \mathcal{J}_c(V_D + \mathbf{u}_0) + N_V \mathcal{J}_v(V_D - \mathbf{u}_0) + \frac{N_C N_A}{N_D + N_A}\left\{-V_A - V_D + R_V^{-1}\left(\frac{N_A}{N_V}\right) + R_C^{-1}\left(\frac{N_D}{N_C}\right) - \frac{N_C}{N_D}\mathcal{J}_c\left[R_C^{-1}\left(\frac{N_D}{N_C}\right)\right] - \frac{N_V}{N_A}\mathcal{J}_v\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right]\right\}\right). \quad (33)$$

Let us define u_B as

$$u_B = R_V^{-1}\left(\frac{N_A}{N_V}\right) + R_C^{-1}\left(\frac{N_D}{N_C}\right) - \frac{N_C}{N_D}\mathcal{J}_c \times \left[R_C^{-1}\left(\frac{N_D}{N_C}\right)\right] - \frac{N_V}{N_A}\mathcal{J}_v\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right]; \quad (34)$$

then the capacitance of an abrupt junction reads

$$C = \sqrt{\frac{q^2\epsilon}{8kT}}\,\frac{2/(N_D + N_A)\,[N_A N_C R_c(V_D + \mathbf{u}_0) + N_D N_V R_v(V_D - \mathbf{u}_0) - N_D N_A]}{\sqrt{N_C \mathcal{J}_c(V_D + \mathbf{u}_0) + N_V \mathcal{J}_c(V_D - \mathbf{u}_0) + [N_D N_A/(N_D + N_A)](-V_A - V_D + u_B)}}. \quad (35)$$

Note that since the band gaps at both sides of the metallurgical junction are not equal in general, a small discontinuity in free carriers occurs at the junction. Substituting the expressions for the free carriers (in the n-type region) at the junction, i.e.,

$$n_D(0) = N_C R_c(V_D + \mathbf{u}_0),$$
$$p_D(0) = N_V R_v(V_D + \mathbf{u}_0),$$

and multiplying both numerator and denominator by $(N_D + N_A)/N_D N_A$, we finally obtain the simplified capacitance formula, which will be used to explain the numerical results of Sec. IV:

$$C = \frac{\sqrt{N_D N_A/(N_D + N_A)}\,\sqrt{q^2\epsilon/2kT}\,\{1 - [p_D(0)/N_A] - [n_D(0)/N_D]\}}{\sqrt{(N_D + N_A)/N_D N_A\,[N_C \mathcal{J}_c(V_D + \mathbf{u}_0) + N_V \mathcal{J}_v(V_D - \mathbf{u}_0)] + (-V_D - V_A + u_B)}}. \quad (36)$$

In comparison, the expression of the capacitance obtained by the depletion layer approximation (DLA) [in the normalized voltage (17)] is given by

$$C=\frac{\sqrt{N_D N_A/(N_D+N_A)}\,\sqrt{q^2\epsilon/2kT}}{\sqrt{-V_D-V_A+R_C^{-1}(N_D/N_C)+R_V^{-1}(N_A/N_V)}}. \tag{37}$$

IV. EVALUATION OF THE GENERAL CAPACITANCE FORMULA

The general capacitance procedure, derived in Sec. III A, has been programmed for abrupt junctions in GaAs assuming a parabolic density of states and compared to the Sedan3[17] results, which we regard as exact, and to DLA (37).

Since the C^{-2} vs V_a is expected to be a straight line,

$$C^{-2}(V_a)=aV_a+b, \tag{38}$$

the behavior of the slope a, the coefficient b, and the correlation coefficient r of the $C^{-2}(V_a)$ function has been studied for nearly all interesting combinations of N_A and N_D in the voltage region [-3 V, 0 V]. The correlation coefficient r is a measure for the linearity of the $C^{-2}(V_a)$ function. Using the auto-grid and diode lengths of ten space charge layers, the Sedan3 results coincide with DLA.

Typical results for $N_A=10^{19}\,\mathrm{cm}^{-3}$ and varying N_D are shown in Figs. 1–3. Two doping ratio N_D/N_A regions can be distinguished.

(1) In the doping ratio region where N_D/N_A lies roughly between 1 and 0.1, all methods converge.

(2) In the doping ratio N_D/N_A range smaller than 0.1, deviations from Sedan3 (and DLA) occur.

We first will discuss symmetrical junctions. In Sec. IV B, asymmetrical junctions are analyzed while their anomalous behavior is explained in Sec. IV C.

A. Symmetrical junctions ($N_A=N_D=N$)

The results for symmetrical junctions are in excellent agreement with the Sedan3 results as observed from Figs. 1–3. The perfect agreement with Sedan3 is explained by examining (36) and (32). The absolute value of the normalized potential at the metallurgical junction $\mathbf{u_0}$ (32) achieves its minimum for symmetrical junctions and equals approximately zero. This implies that in (36), the transformation terms [i.e., $\mathcal{J}$ terms and $n_D(0)$ and $p_D(0)$], which refer to free carrier effects, are negligible and consequently that the DLA is excellent. Hence, both numerator and denominator simplify and (36) almost equals (37) except from u_B. This will cause a different intercept voltage which is discussed in Ref. 13.

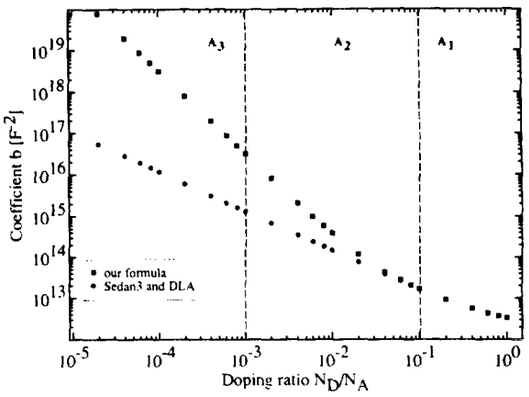

FIG. 2. The capacitance intercept b of $C^{-2}(V_a)$(38) in function of the doping ratio N_D/N_A.

In contrast to the DLA, our calculations (Fig. 4) do not exhibit the unphysical pole in the forward bias region. Due to the assumed boundary conditions (19), the influence of the diffusion capacitance is excluded. As a consequence of the growing importance of the Fermi–Dirac statistic for high doping concentrations, the maximum of the normalized capacitance decreases for increasing doping concentration, while the width of the bell shape increases.

B. Asymmetrical junctions

As long as the doping ratio N_D/N_A is roughly between 0.1 and 1, the agreement of our theory with Sedan3 and with the DLA is good. This region is marked as zone A_1 in Figs. 1–

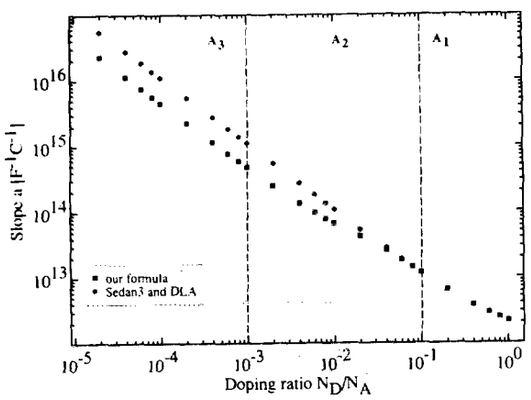

FIG. 1. The slope a of $C^{-2}(V_a)$(38) in function of the doping ratio N_D/N_A.

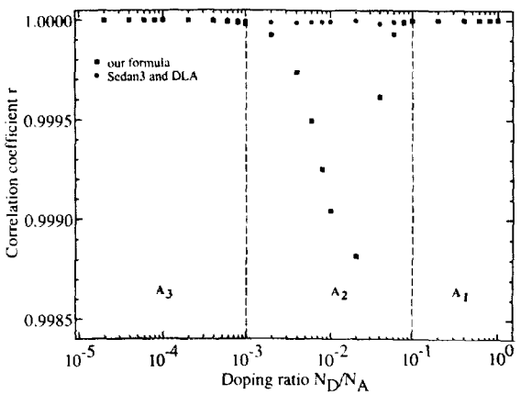

FIG. 3. The correlation coefficient r of $C^{-2}(V_a)$(38) in function of the doping ratio N_D/N_A.

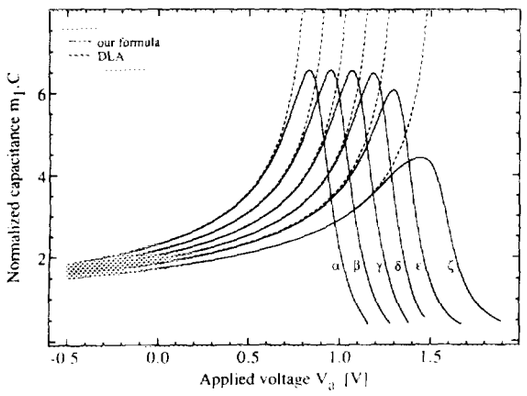

FIG. 4. The normalized capacitances m_1C versus applied voltage V_a for symmetrical abrupt diodes (curves α to ζ) with doping concentration varying from $N = 10^{14}$ to 10^{19} cm^{-3}. The normalization factor $m_1 = 10^{16}/\sqrt{N}$.

3. Increasing the asymmetry in doping concentration forces $\mathbf{u}_0$ (32) to deviate from zero and to tend to the potential of the more highly doped neutral region (i.e., $\mathbf{u}_0 \to u_K$ for $N_D < N_A$). However, not until the doping ratio N_D/N_A reaches about 0.1 does the first term in (32) dominate the other to some extent and thus $\mathbf{u}_0$ remains small. The influence of the free carriers [i.e., the transformation terms in (36)] is still very small.

For doping ratios N_D/N_A between 10^{-1} and 10^{-3} (indicated as zone A_2 in Figs. 1–3), the relationship (38) is not linear anymore, as observed in Fig. 3. This is explained by considering (32). The first term in (32) may be approximated by V_A, while the free carrier terms now increase slowly. For small values of the applied voltage V_a, the normalized voltage V_A is small compared to the free carrier terms and $\mathbf{u}_0$ becomes significantly negative. Hence, free carrier effects dominate and our $C^{-2}(V_a)$ bends upwards with respect to the classical (Sedan3 and DLA) linear behavior. For larger values of the applied voltage, V_A starts to dominate again and forces the capacitance to return to the classical regime. Notwithstanding the nonlinearity of our $C^{-2}(V_a)$, a straight line has been fitted and, from Figs. 1 and 2, we notice that in this zone A_2 the set of slopes a and the set of b coefficients of our formula start following another straight line.

If the asymmetry in doping concentration is very high, i.e., $N_A \gg N_D$ (shown as zone A_3 in Figs. 1–3), then

$$\mathbf{u}_0 \approx V_A - R_V^{-1}\left(\frac{N_A}{N_V}\right) + \frac{N_V}{N_A}\mathcal{F}_c\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right]. \tag{39}$$

Defining $u(N_A)$ as

$$u(N_A) = -R_V^{-1}\left(\frac{N_A}{N_V}\right) + \frac{N_V}{N_A}\mathcal{F}_c\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right], \tag{40}$$

it is easily verified that $u(N_A)$ is a continuous decreasing function of N_A, which tends to $u_K - V_A$ (22) if N_A approaches infinity. Consequently, in the numerator of (36), $n_D(0) \approx 0$, while

$$\frac{p(0)}{N_A} \approx \frac{1}{N_A} R_V[u(N_A)] < 1 \tag{41}$$

and (41) is almost independent of applied voltage. The denominator D of (36) tends to

$$D = \sqrt{-V_A - V_D + R_C^{-1}\left(\frac{N_D}{N_C}\right) + R_V^{-1}\left(\frac{N_A}{N_V}\right) + \frac{N_V}{N_D}\mathcal{F}_c\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right]} \tag{42}$$

and, as a consequence, the intercept voltage can be approximated as

$$V_{\text{int}} \approx \frac{kT}{q}\frac{N_A}{N_D}. \tag{43}$$

This explains the behavior of the slope a and the coefficient b in Figs. 1 and 2, since $V_{\text{int}} = -b/a$. As the last term under the square root in (42) is very large, the denominator D changes very slowly with voltage in [-3 V,0V]. Since the numerator is almost constant, the $C^{-2}(V_a)$ function is again a straight line (Fig. 3). Hence, for increasing doping ratios the numerator of (36) decreases, while the denominator increases such that the capacitance decreases stronger than the DLA capacitance (37). Considering (33) and (36), the observed discrepancy for high doping asymmetries can mainly be reduced to the discrepancy in electric field which increases rapidly due to the increasing spill over of free carriers. The electric field $E(0)$ at the metallurgical junction reads

$$E(0) = \frac{kT}{q}\frac{du(0)}{dx} \approx \sqrt{\frac{2kT}{\epsilon}}\sqrt{N_V\mathcal{F}_c\left[R_V^{-1}\left(\frac{N_A}{N_V}\right)\right] + N_D(-V_A - V_D + u_B)}. \tag{44}$$

In (44), the last term under the square root sign is negligible compared to the first which illustrates the almost independence on applied voltage. Combining this with numerical values for GaAs (44) simplifies to

$$E(0) \approx 8.36\times 10^{-5}\sqrt{N_C\mathcal{F}_c\left[R_C^{-1}(N_D/N_C)\right]}\ [\text{V/cm}]. \tag{45}$$

For an asymmetrical junction ($N_D = 10^{15}$ cm^{-3} and N_A

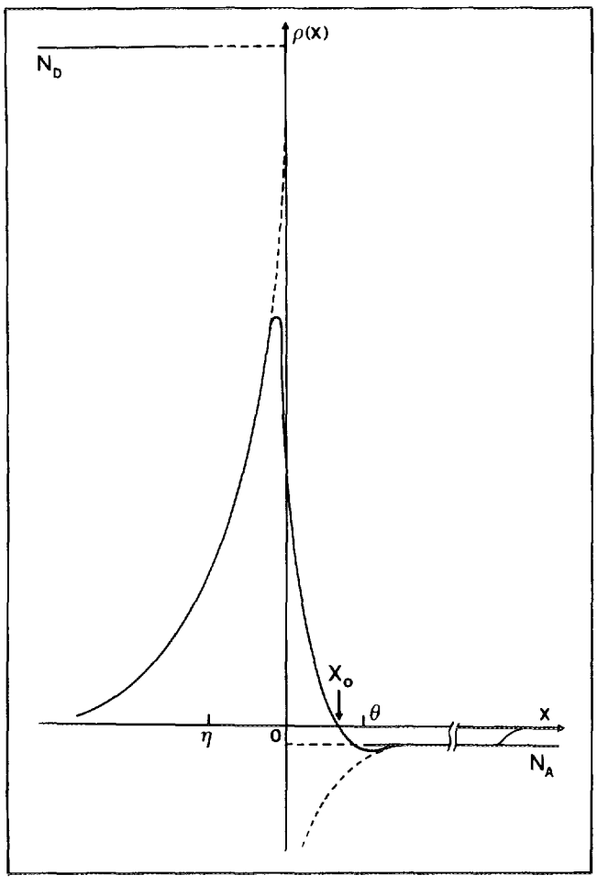

FIG. 5. Schematic picture of the charge density $\rho(x)$ (not scaled and with $q = 1$) in the neighborhood of the junction at $x = 0$, for the case of a mathematically abrupt junction (dashed line) and a realistic abrupt junction (full line). The doping profiles N_D $(x < 0)$ and N_A $(x > 0)$ for a mathematically abrupt junction, and N_D $(x < \eta)$ and N_A $(x > \theta)$ for a realistic abrupt junction are also shown. The interval $[\eta,\theta]$ indicates a transition region. Note that the realistic charge density must have a neutral point X_0.

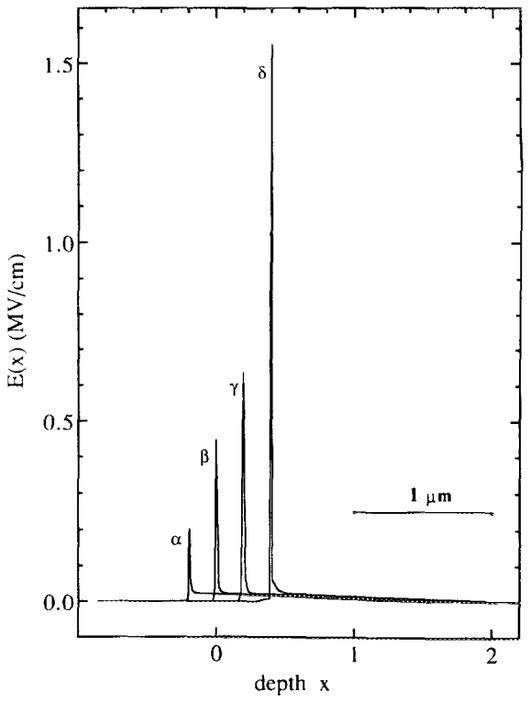

FIG. 6. The electric field of a particular diode ($N_A = 10^{15}$, $N_D = 10^{19}$ and $V_{applied} = -1$ V) simulated in Sedan3 with decreasing (from curve α to δ) grid spacings. The curves are horizontally shifted for clarity; the peak position of their electric field corresponds with the metallurgical junction $(x = 0)$.

$= 10^{19}\,\mathrm{cm}^{-3}$) in GaAs one volt reverse biased, $E(0)$ equals 1.665×10^{5} V/cm while the maximum electric field calculated in the DLA, $E_{\mathrm{DLA}}(0)$, is 7.135×10^{3} V/cm. Our capacitance values are, in this case, about 23 [the ratio between $E(0)$ and $E_{\mathrm{DLA}}(0)$] times less than those obtained by the DLA.

C. Explanation of the anomalies in asymmetrical junctions

For large asymmetries in doping, the discrepancy between the DLA (and Sedan3) and our description is explained as follows. The assumption of constant quasi-Fermi levels leads only to a minor effect because the calculations performed by Kennedy[18] were done in equilibrium (both quasi-Fermi levels equal the constant Fermi level) and show the same discrepancy. Consequently, we might argue that all approximations made were excellent and that our capacitance calculation should give excellent results. Notwithstanding this fact, from capacitance measurements of an asymmetrical abrupt junction in GaAs ($N_D = 2\times10^{16}$ cm^{-3} and $N_A = 2\times10^{18}$ cm^{-3}), grown with MBE, the $C^{-2}(V_a)$ function is a straight line with an intercept voltage of 1.2 V, while the presented theory predicts a value of almost 6.5 V, due to the nonlinear $C^{-2}(V_a)$ relationship in $[-3\text{ V},0.5\text{ V}]$ for $N_A/N_D = 0.01$.

For large asymmetries in doping concentration, the charge density $\rho(x)$ schematically shown in the neighborhood of the metallurgical junction in Fig. 5 in dashed line, keeps increasing faster towards the junction at $x = 0$ and causes a sharply peaked electric field at this point. The experiments do not show evidence for such an electric field although Sedan3 simulations for a doping profile tending to a mathematically abrupt junction confirm the existence of a peaked electric field (Fig. 6). Decreasing the grid spacing causes the increase of the peak value of the electric field. In spite of the numerical values which exceeds $E(0)$ (45) due to numerical differentiation with extreme small gridwidths, the importance of the discontinuity in doping profile is demonstrated. This discontinuity in highly asymmetrical diodes is responsible for a voltage insensitive spill-over of free carriers, causing an almost constant capacitance with respect to applied voltage, which is experimentally not observed. This implies that mathematically abrupt junctions do not exist and that a kind of transition region (shown as the interval $[\eta,\theta]$ in Fig. 5) must occur in which the dependence of the doping concentration on the distance x is not well defined. In conclusion, perfect physical reasoning on a mathematically

abrupt structure must bring us, irrevocably, to nonreality and hence the consequences drawn in Ref. 18 are highly doubtful.

V. CONCLUSION

A formalism to derive the capacitance of an abrupt junction has been presented and evaluated. For symmetrical junctions, excellent agreement with numerical results of a device simulator is achieved. A growing discrepancy between our theory and the numerical results occurs for increasing asymmetries. This phenomenon is explained as the consequence of the assumption of perfect abruptness.

APPENDIX: THE ENERGY DENSITY OF STATES

The study of band-gap narrowing mainly describes the behavior of two quantities, the width of the forbidden energy gap E_g and the energy density of states g_v in the valence and g_c in the conduction band. In this Appendix we briefly discuss the relation between the carrier concentration and the energy density of states employing a new transformation formalism.

The number of electrons is given by

$$n = \int_{E_{c\min}}^{E_{c\max}} \frac{g_c(E - E_c)}{1 + \exp[(E - \zeta_n)/kT]}\, dE . \tag{A1}$$

We substitute

$$\xi = (E - E_c)/kT ,$$

where E_c means the unperturbed band edge similar to E_0 in Refs. 19 and 20. We extend the upperlimit $E_{c\max}$ to infinity while the underlimit $E_{c\min}$ is elongated to minus infinity. Since the density of states can have tails, then

$$n = kT \int_{-\infty}^{\infty} \frac{g_c(kT\xi)}{1 + \exp[\xi - (\zeta_n - E_c)/kT]}\, d\xi . \tag{A2}$$

The number of electrons calculated assuming a parabolic density of states reads

$$n = N_C F_{1/2}[(\zeta_n - E_c)/kT] , \tag{A3}$$

where

$$F_{1/2}(y) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{\sqrt{\xi}}{1 + \exp(\xi - y)}\, d\xi$$

and the effective density of states N_C,

$$N_C = 8\sqrt{2}\,\pi (m_c^* kT)^{3/2}/h^3 .$$

Analogous to (A3), (A2) can be written as

$$n = N_C \wp[g(\xi), (\zeta_n - E_c)/kT] , \tag{A4}$$

where

$$g(\xi) = (kT/N_C)\, g_c(kT\xi) \tag{A5}$$

and where the $\wp$ transformation is defined as

$$\wp[g(\xi), y] = \int_{-\infty}^{\infty} \frac{g(\xi)}{1 + \exp(\xi - y)}\, d\xi . \tag{A6}$$

Hence, if the density of states is parabolic, i.e., $g(E) = \sqrt{E}$ for $E > 0$ and $g(E) = 0$ for $E < 0$, (A3) written in the formalism of (A6), yields

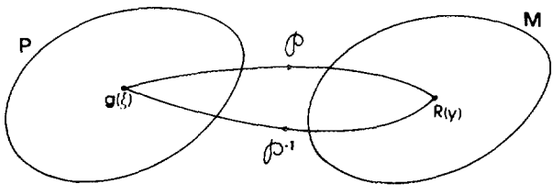

FIG. 7. Group representation of the function classes P and M.

$$n = N_C \wp[2/\sqrt{\pi}\,\sqrt{\xi}\;\theta(\xi), (\zeta_n - E_c)/kT] . \tag{A7}$$

An analogous derivation for the holes can be performed.

As in other integral transformations (e.g., the Laplace transformation) we propose the following notation. We define P as the class of density functions $g(\xi)$ for which the integral (A6) converges and M as the class of functions $R(y)$ that are related to the number of carriers according to $g(\xi)$ (see Fig. 7).

Here, $\wp$ is a transformation from P to M, such that

$$\forall g(\xi) \in P, \exists R(y) \in M : \wp[g(\xi), y] = R(y) ,$$

while $\wp^{-1}$ is the inverse transformation from M to P, such that

$$\forall R(y) \in M, \exists g(\xi) \in P : \wp^{-1}[R(y), \xi] = g(\xi) .$$

We frequently use the inverse function $R^{-1}(z)$, defined as

$$\forall R(y) \in M, \forall z, y \in \mathbb{R} : z = R(y) \Leftrightarrow y = R^{-1}(z) .$$

Notice that $R^{-1}(z)$ does not necessarily belong to M.

Finally, we list some properties[14] of the $\wp$ transformation:

(1) The $\wp$ transformation is linear:

$$\wp[\alpha f(\xi) + \beta g(\xi), y] = \alpha \cdot \wp[f(\xi), y] + \beta \cdot \wp[g(\xi), y] . \tag{A8}$$

$$(2)\ \wp\left[\frac{dg(\xi)}{d\xi}, y\right] = \frac{d}{dy}\{\wp[g(\xi), y]\} . \tag{A9}$$

This equation even holds if $g(0) = 0$ and $g(\xi) = 0$ for $\xi < 0$.

$$(3)\ \int_a^y \wp[g(\xi), y]\, dy = \wp\left(\int_{-\infty}^{\xi} g(\tau)\, d\tau, y\right) - \wp\left(\int_{-\infty}^{\xi} g(\tau)\, d\tau, a\right) . \tag{A10}$$

[1] W. Shockley, Bell Syst. Tech. J. **28**, 435 (1949).
[2] S. P. Morgan and F. M. Smits, Bell Syst. Tech. J. **39**, 1573 (1960).
[3] Y. F. Chang, J. Appl. Phys. **37**, 2337 (1966).
[4] H. K. Gummel and D. L. Scharfetter, J. Appl. Phys. **38**, 2148 (1967).
[5] B. R. Chawla and H. K. Gummel, IEEE Trans. Electron Devices **ED-18**, 178 (1985).
[6] W. Nuyts and R. Van Overstraeten, J. Appl. Phys. **42**, 5109 (1971).
[7] R. Van Overstraeten and W. Nuyts, J. Appl. Phys. **42**, 4041 (1972).
[8] J. E. Parrott and L. Ph. Leonidou, Phys. Status Solidi A **25**, 231 (1974).
[9] D. Redfield, Appl. Phys. Lett. **35**, 182 (1979).
[10] J. J. H. Van Den Biesen, Philips J. Res. **40**, 2 (1985).
[11] F. A. Lindholm and J. J. Liou, J. Appl. Phys. **63**, 561 (1988).
[12] O. V. Konstantinov and O. A. Merzin, *Semiconductor Physics*, edited by V. M. Tuchkevich and V. Ya Frenkel (Consultants Bureau, New York, 1985), p. 491.

[13]P. Van Mieghem, R. P. Mertens, G. Borghs, and R. J. Van Overstraeten, Phys. Rev. B **41**, 5952 (1990).
[14]P. Van Mieghem, IMEC internal report.
[15]*Physical Electronics and Circuit Models of Transistors*, SEEC (Wiley, New York, 1964), Vol. 2, Appendix A, pp. 245–253.
[16]C. T. Sah, IEEE Trans. Electron Devices **ED-13**, 839 (1966).
[17]Z. Yu and R. W. Dutton, *Sedan III–A Generalized Electronic Material Device Analysis Program, Integrated Circuits Laboratory* (Stanford University, Stanford, CA, 1985).
[18]D. P. Kennedy, IEEE Trans. Electron Devices **ED-22**, 998 (1975).
[19]B. I. Halperin and M. Lax, Phys. Rev. **148**, 722 (1966).
[20]V. Sa-yakauit and H. R. Glyde, Phys. Rev. B **22**, 6222 (1980).