EFFECTS OF BANDGAP NARROWING ON THE CAPACITANCE OF SILICON AND GaAs pn JUNCTIONS

S. C. Jain*, R. P. Mertens, P. Van Mieghem, M. G. Mauk
M. Ghannam, G. Borghs and R. Van Overstraeten**
IMEC: Interuniversity Microelectronics Center
Kapeldreef 75, Leuven, Belgium
*Former address: Solid State Physics Laboratory, Delhi- 110007, India
**Also with Electronics and Communications Dept., Cairo University, Ghiza, Egypt

Abstract: The effect of heavy doping on the capacitance-voltage relation of abrupt and linearly-graded p-n junctions is studied by computer simulations. An estimate of bandgap narrowing in compensated silicon is given for linearly-graded junctions. Capacitance-voltage curves of abrupt p-n GaAs junctions grown by MBE are investigated and compared to the theoretical curves.

1. Introduction. Recently, the importance of junction capacitance has been emphasized by Lindholm and Liou [1] in the design of advanced thin-base bipolar transistors. Since the built-in voltage of a diode depends on the bandgap of the material, the junction capacitance will be influenced by bandgap narrowing effects occuring in heavily-doped semiconductors.

Van Overstraeten and Nuyts [2] have studied the C^{-3}-V behavior of some forty linearly-graded p-n junctions fabricated on heavily-doped silicon substrates. The results are shown in Figure 1. The difference between the theoretical intercept (or "gradient") voltage and its experimental value is up to 350 mV. On account of the experimental intercept voltage being smaller than its theoretical value, the observed C^{-3}-V line is displaced along the voltage axis from the predicted line. Mertens et al. [3, 4] attributed this discrepancy to bandgap narrowing in the quasi-neutral regions which results in a lower value of the built-in voltage as well as a reduced value of the intercept voltage. In this regard, Lowney has published two interesting papers [5, 6]. He confirmed independently the findings of Van Overstraeten et al. [2, 4], namely that the difference between the observed intercept and its theoretical value was very large. However, he attributed the large shift in the C^{-3}-V curve to the lack of screening in the space-charge region. Thus, the same observation was explained by two distinct physical phenomena: bandgap narrowing in the quasi-neutral regions [3, 4] and reduced free carrier screening in the depletion region [5, 6].

In this paper, we first use the depletion approximation to derive expressions for the shifts of the C^{-2}-V curve (for abrupt junctions) and the C^{-3}-V curve (for linearly-graded junctions) due to bandgap narrowing (section 2). We then calculate the bandgap narrowing ΔE_g using a theory based on the depletion layer approximation and the experimental results of Van Overstraeten and Nuyts [2]. We attempt to interpret ΔE_g using self-consistent calculations based on a model proposed by Mock [7] using Morgan's theory of impurity bands [8] and Kane's theory of band tails [9]. In section 4 we investigate by computer simulation the effect of majority carrier Debye tails in the depletion region edge on the junction capacitance, including the effect of bandgap narrowing . In section 5 we discuss the computer results from general physical considerations and describe methods which can be used to determine ΔE_g from C-V measurements. Finally, experimental results on abrupt GaAs p-n junctions are given and discussed in section 6.

2. Effect of Bandgap Narrowing on the C-V Relations in the Depletion Region Approximation. Computer simulations show that the effect of Fermi-Dirac statistics on the C-V relation is small in silicon even at doping concentrations as high as 10^{19} cm^{-3}. We first confine our discussion to the case where Boltzmann statistics are valid. The case of GaAs with Fermi-Dirac statistics will be considered in section 6. In heavily-doped silicon, the pn product is given by

$$pn = n_{ie}^2 = n_i^2 \exp(\Delta E_g/kT) \tag{1}$$

where n_{ie} is the *effective* intrinsic carrier concentration including bandgap narrowing. The built-in potential for an abrupt junction is

$$V_{bi} = \frac{kT}{q} \ln \frac{N_d N_a}{n_{ie}(N) n_{ie}(P)} \tag{2}$$

where $n_{ie}(N)$ and $n_{ie}(P)$ are the values of n_{ie} on the N and P sides, respectively and N_d and N_a are the donor and acceptor concentrations. For the linearly-graded junction [10],

$$V_{bi} = \frac{kT}{q} \ln \frac{a^2 W^2}{4 n_{ie}(-\frac{W}{2}) n_{ie}(+\frac{W}{2})} \tag{3}$$

where W is the width of the depletion region and a is the *net* doping gradient. $n_{ie}(-\frac{W}{2})$ and $n_{ie}(+\frac{W}{2})$ are the values of n_i at the edges $(\pm\frac{W}{2})$ of the depletion region. The C-V relations are given by [10]

$$C^{-2} = \frac{2}{q\epsilon_s} \frac{(N_d + N_a)}{N_d N_a} (V_{bi} - V_A) \tag{4a}$$

and [10]

$$C^{-3} = \frac{12(V_{bi} - V_A)}{q a \epsilon_s^2} \tag{4b}$$

for abrupt and linearly-graded junctions, respectively. In the depletion region approximation these C-V relations remain essentially unaltered when heavy doping effects are included; the one modification being that n_{ie} is used in place of n_i when calculating the built-in potential V_{bi}.

3. Calculation of ΔE_g from the Experiments of Van Overstraeten and Nuyts. Chawla and Gummel [11] have shown that in linearly-graded junctions, the *intercept* or *gradient* voltage V_g is different from the *built-in* voltage V_{bi} and is given by [10, 11]

$$V_g = \frac{2}{3} \frac{kT}{q} \ln \frac{a^2 \epsilon_s kT}{8 q n_i^3} \tag{5}$$

As discussed in the previous section, if bandgap narrowing is taken into account, n_i is replaced by n_{ie}, with the intercept voltage now given as

$$V_g' = \frac{2}{3}\frac{kT}{q}\ln\frac{a^2\epsilon_s kT}{8qn_{ie}^3} \tag{6}$$

The C-V relation becomes

$$C^{-3} = \frac{12}{qa\epsilon_s^2}(V_g' - V_a) \tag{7}$$

which implies

$$\Delta E_g = V_g - V_g' \tag{8}$$

The effect of bandgap narrowing is to significantly increase the value of C and displace the C^{-3}-V plot by an amount ΔE_g along the voltage axis. The values of ΔE_g can therefore be calculated from Figure 1. The values thus determined are plotted in Figure 2. The results of Figure 2 are interesting. For small values of N_a, ΔE_g increases as $(N_d - N_a)$ increases. For large values of N_a, ΔE_g decreases with $(N_d - N_a)$. These are in qualtitative agreement with the predictions of Mock[7]. For large N_a, the junction is strongly compensated and the screening length will be small, resulting in a ΔE_g that decreases with $(N_d - N_a)$.

Our attempt to fit the ΔE_g data quantitatively with Mock's theory using the Thomas-Fermi or Stern [12] formula for screening length was not successful. We could, however, fit the data if the screening length was taken as an adjustable parameter. The screening lengths which fit the data are given in Table 1.

4. Numerical Calculations of the Effect of Bandgap Narrowing on the Majority Carrier Tails and Junction Capacitance. The effect of the majority carrier Debye tails on the junction capacitance and C-V relation has been discussed by, among others, Chawla and Gummel [11], Gummel and Scharfetter [13], Kennedy [14] and, more recently, Tuckevich and Frenkel [15]. The effect of bandgap narrowing has not been included in any of the papers so far. The three cases of practical importance are 1) symmetrical abrupt junctions, 2) asymmetrical abrupt junctions and 3) linearly-graded junctions. Since we are interested in understanding the underlying physics with regard to how ΔE_g modifies the majority carrier tails and junction capacitance, we have investigated extensively, using SEDAN [16] computer simulations, the C-V relation of silicon abrupt and linearly-graded junctions. For abrupt junctions, doping concentrations ranging from 10^{17} cm^{-3} to 10^{20} cm^{-3} were used. Typical results for four junctions are shown in Figure 3. The straight lines represented by the squares give the value of Δ defined as

$$\Delta = \frac{1}{2}[\Delta E_g(N) + \Delta E_g(P)] \tag{9}$$

We made C^{-2}-V plots *with bandgap narrowing* (BGN) and *with no bandgap narrowing* (NBGN) for different values of applied voltage V_A. In all cases, the plots were almost exact straight lines. The plots with ΔE_g included in the SEDAN simulation were lower, i.e. the values of C^{-2} were smaller. We define the ***shift*** $S(V_a)$ as the displacement along the voltage axis of the C^{-2}-V line when bandgap narrowing effects are included, i.e.

$$S(V_A) = C^{-2}_{(\text{NBGN})}(V_A) - C^{-2}_{(\text{BGN})}(V_A) \tag{10}$$

The shift $S(V_A)$ should be equal to Δ if the depletion theory outlined in the previous section is correct. The difference δ

$$\delta = \Delta - S(V_A) \tag{11}$$

is plotted as the solid lines in Figure 1. From curve a and from other data not shown in the figure, we find that if the asymmetry (defined as N_d/N_a) is between 0.1 and 10, δ is very nearly zero. However, as the value of N_d/N_a increases above 10 (or decreases less than 0.1), δ starts increasing rapidly for low reverse biases (see curves c and d). The increase becomes smaller for larger values of reverse bias.

Similar computer calculations were also made for the linearly-graded junctions. The value of the concentration gradient a was varied from 10^{20} to 10^{25} cm^{-4}. In all cases, $\delta(V_A)$ was found to be about 20 meV. In the case of the graded junction, Δ was calculated for N_d and N_a at the edges of the depletion region and thus was a function of the applied voltage.

5. Discussion of the SEDAN Simulations. We have seen in the earlier sections, that in the depletion approximation Δ equals $S(V_A)$ at all applied voltages V_A. However, when the effect of free carriers is taken into account, δ is not zero and is a function of applied voltage for asymmetrical and linearly-graded junctions. These results can be understood with the analysis of Chawla and Gummel [11], and Gummel and Scharfetter [13] (see also ref. [15]). For symmetrical junctions, the effect of majority carriers is to reduce the intercept voltage by $(2kT/q)$. Since this quantity is independent of n, p or n_i^2, it does not depend on the value of bandgap narrowing ΔE_g. The term $2kT/q$ is common in both cases and does not affect the shift $S(V_A)$. This result is in complete agreement with our computer simulations. This result is of great practical importance because it provides a method of determining Δ using symmetrical p-n junctions. This method is superior because it does not involve the knowledge of minority carrier lifetime or mobility.

We now consider asymmetrical junctions. Gummel and Scharfetter [13] have extensively studied the effect of free carrier "tails" on the offset and intercept voltage in asymmetrical junctions. If the asymmetry factor N_d/N_a is more than about 50, the reduction is not an additive term (as in symmetrical junctions where the intercept voltage is decreased by an amount $2kT/q$, see ref. 10) but larger and depends both on doping concentrations and on the applied voltage. For small applied voltage and large N_d/N_a, i.e. greater than 50, the offset voltage has a value close to $(kT/q)\ln(N_d^2/n_i^2)$. As the reverse bias increases, the offset voltage increases and approaches the value $(kT/q)\ln(N_dN_a/n_i^2)$. This result can be understood from general physical considerations. The mobile carrier profile in a highly doped asymmetrical p$^+$n$^-$ junction is shown in Figure 4 for zero bias and a small reverse bias. It can be seen that the depletion region charge and dQ/dV is determined mainly by the lightly doped side of the junction in both cases. In this case dQ/dV is also mainly determined by N_d/n_i. If a large reverse bias is now applied, the holes recede back into the p$^+$ region and the intercept voltage is determined by both N_a/n_i and N_d/n_i. In the limit of very large reverse bias and small asymmetry, it again approaches the value of $[(kT/q)\ln(N_dN_a/n_i^2) - 2kT/q]$ as mentioned earlier.

We are now in a position to understand the computer results of Figure 3 from general physical considerations. For a symmetrical or nearly symmetrical junction, the free carrier effect is independent of bandgap narrowing and the shift due to ΔE_g is exactly the same as in the case of the depletion approximation. Using equations (2) and (3), it can be easily shown that $S(V_A)$ equals Δ which is in agreement with the SEDAN simulations (Figure 3). For highly asymmetrical junctions, the offset voltage is determined by n_{ie} of the lightly doped side and $S(V_A)$ approaches the value of $\Delta E_g(N)$ (lightly doped N side) if reverse bias is not large. This is clearly seen in curves c and d of Fig. 3. It is also seen from these curves that as the reverse bias V_A increases, $S(V_A)$ approaches a value Δ equal to $\frac{1}{2}[\Delta E_g(N) + \Delta E_g(P)]$.

It is clear by the appropriate combination of doping concentrations and values of reverse bias, the bandgap narrowing values can be determined without necessarily using computer simulations. The best results should be obtained using symmetrical doping profiles (N_a equal N_d). Even this case is not exactly symmetrical, as assumed in the Boltzmann approach, on account of differences in the electron and hole effective masses and modifications due to Fermi-Dirac statistics.

Qualitatively, the SEDAN results for the linearly-graded junction can be understood from general physical considerations. As shown by Chawla and Gummel [11], the effect of majority carrier tails is to reduce the intercept voltage by about 100 meV. This reduction is only weakly dependent on the gradient a. If this reduction were independent of ΔE_g, $S(V_A)$ should be equal to Δ and δ should be zero. The computer results show that δ is approximately 20 meV which implies that Δ does not correspond to the impurity concentrations at the edges of the depletion layer in the quasi-neutral region but is the weighted average of the impurity concentration "seen" by the free carriers at the edges of the space-charge region. Since the Δ values at the edges of the depletion layer range from 80 to 150 meV, the error in attributing S to the concentration at the edges in the quasi-neutral region neglecting δ as being small causes an error of 15 to 25% in the value of Δ determined from the C^{-2}-V plots.

We now discuss briefly the case of the linearly-graded junction in highly compensated silicon. The main difficulty in calculating theoretically the C^{-3}-V plots for diodes with compensated silicon is that a reliable model for bandgap narrowing is not available. Mock's theory [7] is based on Morgan's model [8] for compensated semiconductors and it is known that classical models do not give accurate results [17]. Serre and Ghazali [17] and Lowney [6] have calculated ΔE_g for compensated silicon using quantum mechanical methods. However, both have used Thomas-Fermi screening which is known to be inaccurate at low concentrations of free carriers as in the case of compensated silicon. There are practically no measured values of bandgap narrowing in compensated silicon reported in the literature. In view of the fact that in many diffused and implanted junctions the emitter and/or base are compensated, the results reported in Figure 2 and Table I are important. The high values of ΔE_g in Figure 2 are due to compensation [7, 15] and not due to lack of screening in the space charge region as suggested by Lowney [6].

6. Experimental Results on Heavily-doped Abrupt GaAs p-n Junctions. To study capacitance-voltage relations in GaAs, we fabricated p-n junctions in GaAs using Molecular Beam Epitaxy (MBE). The sharp abruptness of the junctions is indicated by the straightness of the C^{-2}-V plot which is shown in Figure 5. In Table 2, C^{-2}-V intercept values for three different GaAs p-n junctions are given. In the table, the first set of intercepts is obtained by simulating the experimental devices including Fermi-Dirac statistics but without bandgap narrowing. The second set of intercepts is obtained from the measured data. Finally, the difference between these two intercepts gives an indication of the "average" bandgap narrowing Δ (eqn. 9). Detailed simulations show that the effect of the p^+-p and n^+-n junctions (resulting from the inclusion of heavily-doped contact layers in the structure) on the measured capacitance is negligible.

These experiments suggest that bandgap narrowing can be determined from C-V measurements, providing the doping profile is known with sufficient accuracy. To interpret the data an analytical expression for the intercept voltage of an abrupt junction, which incorporates Fermi-Dirac statistics and includes bandgap as a parameter, would be highly useful.

References.

1. F. Lindholm and J. J. Liou, *J. Applied Physics* **63, 2** (1988) pp. 561-564.
2. R. Van Overstraeten and W. Nuyts, *J. Applied Physics* **43, 10** (1982) pp. 4040-4050.
3. R. P. Mertens, *Ph.D. Dissertation* Katholike Universiteit Leuven (1972).
4. R. Van Overstraeten, H. J. DeMan and R. Mertens, *IEEE Trans. Electron Devices* **ED-20, 3** (1973) pp. 290-298.
5. J. Lowney and W. R. Thurber, *Electronics Letters* **20, 3** (1984) pp. 142-143.
6. J. Lowney, *Solid-State Electronics* **28, 1-2** (1985) pp. 187-191.
7. M. S. Mock, *Solid-State Electronics* **16** (1973) pp. 1251-1259.
8. T. N. Morgan, *Physical Review* **139, 1A** (1965) pp. A343-A348.
9. E. O. Kane, *Physical Review* **131, 1** (1963) pp. 79-88.
10. S. M. Sze, *Physics of Semiconductor Devices* 2 ed., (New York: Wiley, 1981) pp. 74-84.
11. B. R. Chawla and H. K. Gummel, *IEEE Trans. Electron Devices* **ED-18, 3** (1981) pp. 178-195.
12. F. Stern, *Physical Review, B* **3, 10** (1971) pp. 3559-3560.
13. H. K. Gummel and D. L. Scharfetter, *J. Applied Physics* **38, 5** (1967) pp. 2148-2153.
14. D. P. Kennedy, *IEEE Trans. Electron Devices* **ED-22, 11** (1975) pp. 988-994.
15. V. M. Tuckevich and V. Y. Frenkel, in *Semiconductor Physics* ed. N. Mott (New York: Consultants Bureau, 1987) pp. 491-517.
16. W. Dutton and Zhiping Yu, *SEDAN III, A Generalized Electronic Material Device Analysis Program* (July 1985) Integrated Circuits Laboratory, Stanford University.
17. J. Serre and A. Ghazali, *Physical Review B* **28** (1983) pp. 4704-4715.

Table 1
Screening Lengths for Compensated Silicon

Base doping: $N_a = 2.5e19$					
$\log(N_d - N_a)$:	18.6	18.8	19.0	19.2	19.4
screening length:	2.9e-7	2.3e-7	1.7e-7	1.1e-7	0.5e-7
Base doping: $N_a = 2.5e18$					
$\log(N_d - N_a)$:	18.0	18.2	18.4	18.6	18.8
screening length:	8.8e-7	6.7e-7	4.9e-7	3.6e-7	1.3e-7
Base doping: $N_a = 4.0e16$					
$\log(N_d - N_a)$:	17.2	17.4	17.6	17.8	18.0
screening length:	2.1e-6	2.5e-6	2.6e-6	2.6e-6	2.5e-6
Base doping: $N_a = 5.0e15$					
$\log(N_d - N_a)$:	16.6	16.8	17.0	17.2	17.4
screening length:	2.7e-5	2.6e-5	2.1e-5	1.9e-5	1.6e-5

Note: 1. screening length in centimeters, substrate base doping N_a in cm^{-3}, net doping $N_d - N_a$ in cm^{-3}; 2. screening lengths were derived by fitting the data of Figure 2 and using the model of Mock [7] for the effective intrinsic carrier concentration n_{ie}.

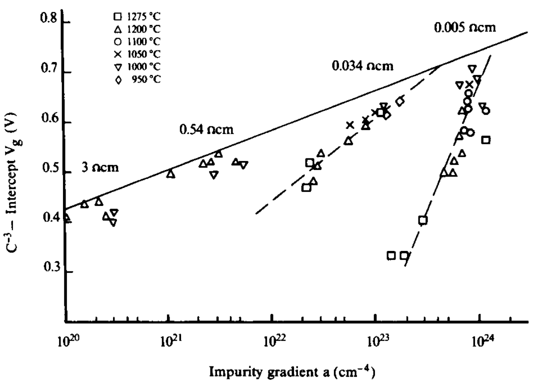

fig 1. C^{-3}– V_g for linearly graded, diffused silicon diodes for 4 different substrate resistivities [2]

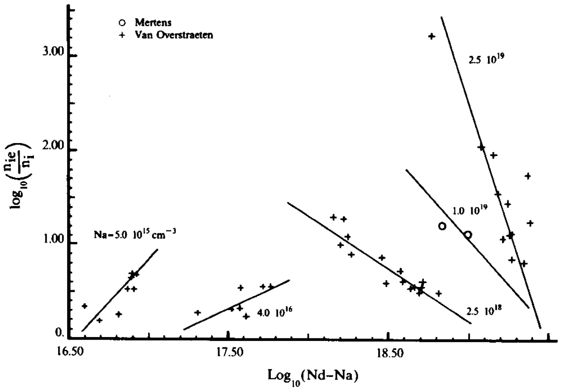

fig 2. Values of effective n_{ie} as a function of $\log_{10}$(Nd-Na), derived from the data of [2] (see fig 1) and [3]

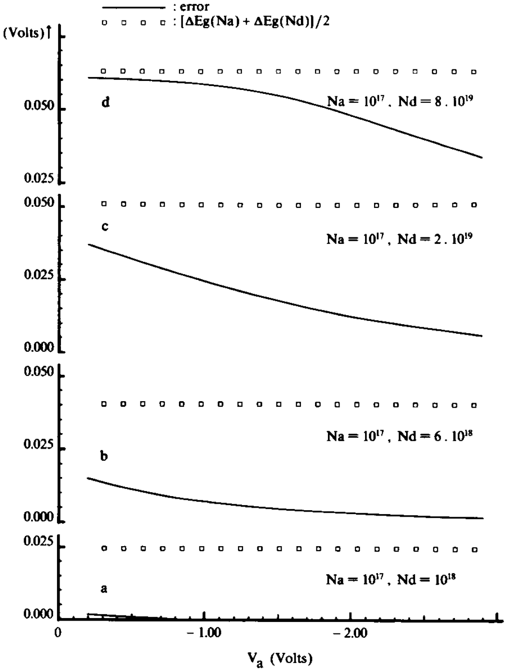

fig 3. The shift of C^{-2}– V plots due to bandgap narrowing (see text)

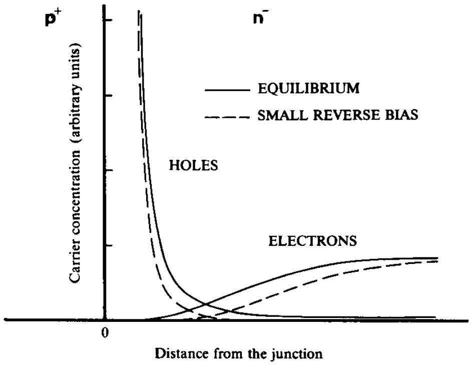

fig 4. Free carrier profiles (schematic) in a highly asymmetrical junction.

Table 2 C-V Plot Intercepts for GaAs p-n Junctions			
Sample	291	292	294
Donor Conc. (cm^{-3})	1.6e18	3.7e18	1.6e18
Acceptor Conc. (cm^{-3})	2.8e18	7.0e18	2.0e19
Calculated(V)	1.39	1.45	1.45
Measured (V)	1.34	1.40	1.34
Difference (V)	0.05	0.05	0.10

MBE-grown abrupt junctions; approximate layer thickness: 0.4 microns.

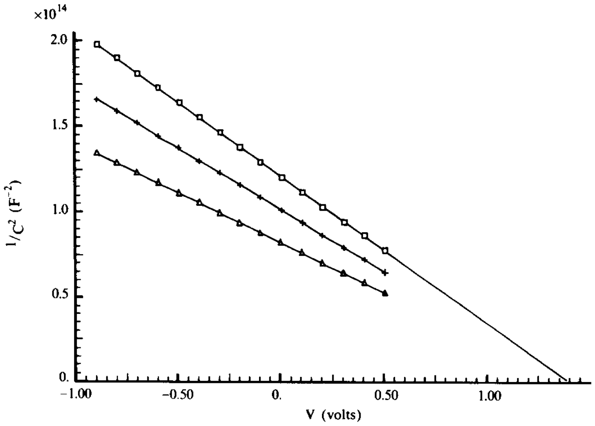

fig 5. C^{-2}– V plot of a heavily doped abrupt GaAs diode with different areas but with the same doping levels (sample 1, Table 2)