Physical Model of Incomplete Ionization for Silicon Device Simulation

Andreas Schenk
Integrated Systems Laboratory
ETH Zurich, Gloriastr. 35
CH-8092 Zürich, Switzerland
and
Synopsys Switzerland LLC.
Affolternstrasse 52
CH-8050 Zürich, Switzerland
Email: schenk@iis.ee.ethz.ch

Pietro P. Altermatt
University of Hannover
Inst. Solid-State Physics,
Dep. Solar Energy, Appelstr. 2
30167 Hannover, Germany
Email: altermatt@solar.uni-hannover.de

Bernhard Schmithüsen
Integrated Systems Laboratory
ETH Zurich, Gloriastr. 35
CH-8092 Zürich, Switzerland
Email: bernhard@iis.ee.ethz.ch

***Abstract*— An empirical model of incomplete ionization (ii) in phosphorus-, arsenic-, and boron-doped crystalline silicon is derived from photoluminescence, conductance, and mobility measurements. It is found that up to 25% of phosphorus and boron atoms and up to 35% of arsenic atoms are non-ionized at room temperature near the Mott transition, whereas there is no significant amount of ii at dopant densities far above the Mott transition. Simplified equations of ii suitable for implementation in device simulators are exploited to study the effect of ii on the performance of bipolar and MOS devices. It is demonstrated that ii can increase the current gain of bipolar transistors by up to 25%.**

I. Introduction

The effect of incomplete ionization (ii) of dopants is often neglected in silicon device simulation due to the common belief that it is unimportant at room temperature. Moreover, in recent papers ii was assumed to increase continuously with increasing doping concentration N_{dop} [1]–[3]. Here we show that ii is only pronounced in a narrow N_{dop} range around the Mott density, but can change e.g. the maximum current-gain of bipolar transistors by 25%. Based on a thorough analysis of published experimental data and suitable expressions for the density of states (DOS), a parameterized ii model for phosphorus-, arsenic-, and boron-doped Si is derived, implemented into a device simulator, and applied to different devices.

II. The Physics of Incomplete Ionization

Fig. 1 shows photoluminescence and conductance measurements [4], [5] of the Si DOS near the conduction band edge at conditions of increasing phosphorus concentration. One observes a (nearly symmetrical) broadening of the impurity band due to cluster formation, the shift of this broadening band towards the conduction band edge due to increasing screening of the binding potential, the starting merge of the impurity band with the conduction band at roughly the Mott density ($N_{\text{crit}} = 3.74 \times 10^{18}\,\text{cm}^{-3}$), and the formation of a strong DOS tail when the doping concentration is further increased, due to remaining states bound to dopant clusters (triangles in Fig. 1f) and disorder effects. We use a commonly accepted theoretical model for the combined DOS [6], which is the sum of a tail DOS $D_k(E)$ and a Gaussian DOS of the impurity band $D_{\text{dop}}(E)$. The latter is multiplied by a doping-dependent attenuation factor

$$b(N_{\text{dop}}) = [1 + (N_{\text{dop}}/N_b)^d]^{-1} \quad (1)$$

to account for the gradual disappearance of bound states in favor of extended states as the doping exceeds N_{crit}. Upper and lower bounds for $b(N_{\text{dop}})$ are found from published magnetic susceptibility and DOS measurements, and its parameterization (i.e. the determination of the fit parameters N_b and d) is the result of a detailed comparison between calculated ii (see below) and extracted ii from mobility measurements as described next.

III. Extraction of Incomplete Ionization from Experimental Data

The amount of ii can be extracted from the ratio between conductivity mobility $\mu_{\text{cond}} = \sigma/qn(N_{\text{dop}}, T)$ and Hall mobility $\mu_H = \sigma R_H/r(N_{\text{dop}}, T)$. The former is obtained by measuring the conductivity σ and assuming that $n(N_{\text{dop}}, T) = N_{\text{dop}}$. To obtain consistency, each mobility value was derived from the published σ value using the N_{dop}-σ relationship of Mousty et al. [7]. The Hall mobility follows from measurements of both σ and the Hall factor R_H. For the Hall correction factor $r(N_{\text{dop}}, T)$ the empirical relation $r(N_{\text{dop}}, 300K) = 1.18 + 0.18/(1 + N_{\text{dop}}/10^{17})^{0.8}$ was used for phosphorus and arsenic, whereas $r = 0.8$ was found to be the best choice for boron. From this procedure one therefore obtains $n/N_{\text{dop}} = \mu_{\text{cond}}/\mu_H$, as shown in Fig. 2 for the case of phosphorus, since μ_H contains the information on the carrier density, whereas μ_{cond} contains the information on the doping concentration.

IV. Device Model for Phosphorus, Arsenic, and Boron

We compute ii by solving self-consistently the set of equations for the concentration of non-ionized dopants, the free

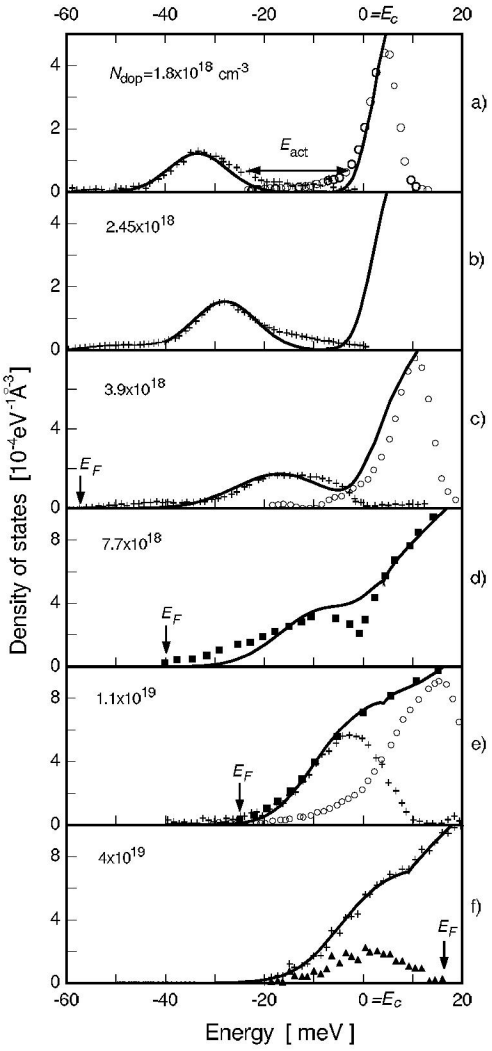

Fig. 1. The measured Si photoluminescence intensity at 4.2 K [4] (arbitrary units) at low injection power (crosses) and at high injection power (open circles), and the DOS derived from conductance measurements [5] (squares), at various phosphorus densities. The solid lines represent the theoretical DOS with Eq. (1).

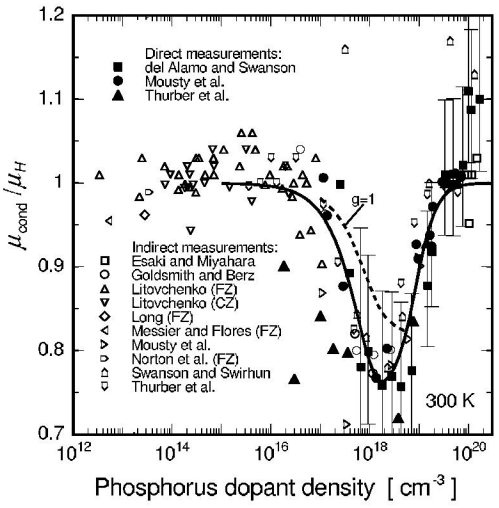

Fig. 2. The fraction between measured conductivity mobility and Hall mobility (symbols) reflecting incomplete ionization, which is compared with our full model (solid line).

carrier density, and the total (electrically active) concentration, using the combined DOS with $b(N_{\rm dop})$. In the case of donors it reads

$$N^0_{\rm dop} = \int_{-\infty}^{0} [D_{\rm dop}(E) + D_k(E)] f_D(E, E_F) dE \quad (2a)$$

$$N^+_{\rm dop} = \int_{0}^{\infty} [D_{\rm dop}(E) + D_k(E)] f(E, E_F) dE \quad (2b)$$

$$N_{\rm dop} = N^0_{\rm dop} + N^+_{\rm dop}. \quad (2c)$$

where $N^0_{\rm dop}$ is the neutral (non-ionized) dopant density. The unknown quantities in this equation system are $N^0_{\rm dop}$, $N^+_{\rm dop}$ $(= n)$, and E_F. The occupation probability of the dopant states, f_D, differs from the Fermi-Dirac function f by the degeneracy factor g. The parameters for $b(N_{\rm dop})$ are adjusted such that the experimental data of ii are reproduced within their precision as shown by the solid line in Fig. 2. One has to keep in mind that some of the dopant states may be extended within dopant clusters which can cause g to increase from 1/2 towards unity. Interestingly, the dashed line in Fig. 2 turns out if the bound-state occupation probability is computed with the degeneracy factor $g = 1$ instead of $g = 1/2$ for donors. As stated above, this could reflect delocalization of states within the growing clusters, but it leads to an underestimation of ii below $N_{\rm crit}$. This suggests that the dopant band contains mainly localized states, which is also consistent with published magnetic susceptibility and specific heat measurements.

To obtain an ii model for device simulation, we neglect the DOS tails. Furthermore, the Gaussian of the impurity DOS is replaced by a delta function and quasi-Fermi levels are expressed by densities using Maxwell-Boltzmann statistics, which both has a negligible effect, since the ionization rate quickly tends towards 1 as the carriers become degenerate. Then

$$N^+_{\rm don}/N_{\rm don} = 1 - b(N_{\rm dop})\, n/[n + g\, n_1]\ , \quad (3a)$$

$$N^-_{\rm acc}/N_{\rm acc} = 1 - b(N_{\rm dop})\, p/[p + g\, p_1] \quad (3b)$$

with the densities

$$n_1 = N_c\, e^{-E_{\rm dop}/kT}, \qquad p_1 = N_v\, e^{-E_{\rm dop}/kT}. \quad (4)$$

The screening-dependent binding energies $E_{\rm dop}$ are found from their doping dependence (see Fig. 1) and are modelled as

$$E_{\rm dop} = \frac{E_{\rm dop,0}}{1 + \left(\frac{N_{\rm dop}}{N_{\rm ref}}\right)^c}. \quad (5)$$

The results for the degree of ii of phosphorus dopants are shown in Fig. 2 in comparison to experimental values, in Fig. 3 in comparison to various theoretical models from the literature [1], [2], [8]–[10] and in Fig. 4 (denoted as "full model" there) for P, As, and B.

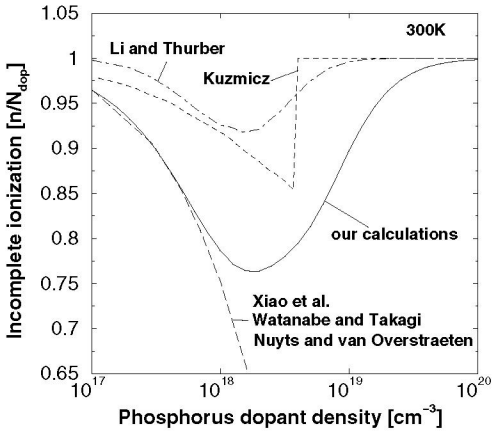

Fig. 3. The fraction of non-ionized phosphorus atoms from various published models compared to the full model of the present paper.

The recommended device model for ii is given by Eqs. (3a) – (3b) with density-dependent functions b, Eq. (1), and $E_{\rm dop}$, Eq. (5), together with the parameters given in Table I. We intended to implement this model into the device simulator DESSIS–Synopsys [11] via the 'Physical Model Interface' (PMI). The PMI allows to use own models, but requires to follow *pre-defined* dependencies in order not to degrade the convergence behaviour. The most serious restrictions apply to the dependence on the carrier densities. In case of ii the pre-defined functionality in DESSIS–Synopsys is given by

$$\frac{N^+_{\rm don}}{N_{\rm don}} = \frac{1}{1+g_{\rm D}n/n_1}, \tag{6a}$$

$$\frac{N^-_{\rm acc}}{N_{\rm acc}} = \frac{1}{1+g_{\rm A}p/p_1}, \tag{6b}$$

where the "degeneracy factors" $g_{\rm D}$ and $g_{\rm A}$ can depend on temperature and doping, but not on the free carrier densities. Transforming our ii model (3a) – (3b) into the form (6a) – (6b), we obtain "effective degeneracy factors"

$$g_{\rm D}(T, N_{\rm don}, n, p) = \frac{b}{g+(1-b)\,n/n_1}, \tag{7a}$$

$$g_{\rm A}(T, N_{\rm acc}, n, p) = \frac{b}{g+(1-b)\,p/p_1}, \tag{7b}$$

TABLE I

THE PARAMETERS USED IN EQS. (3A) AND (3B), CONTAINING EQS. (5) AND (1), TO CALCULATE THE AMOUNT OF INCOMPLETE IONIZATION IN DEVICE SIMULATIONS.

Parameter	Si:P	Si:As	Si:B
$E_{\rm dop,0}$ [meV]	45.5	53.7	44.39
$N_{\rm ref}$ [cm^{-3}]	2.2×10^{18}	3×10^{18}	1.3×10^{18}
c	2	1.5	1.4
N_b [cm^{-3}]	6×10^{18}	9×10^{18}	4.5×10^{18}
d	2.3	1.8	2.4
g	1/2	1/2	1/4

Besides their explicit dependence on the densities n (p), an additional implicit dependence is given, if the screening effect on the binding energies in the functions n_1 and p_1 is expressed as function of $n+p$. Replacing the densities by the ionized dopant concentration in (7a) – (7b) is only exact in neutral regions, but even then would make our ii model an implicit relation. For simplicity, we thus replace all densities in the "effective degeneracy factors" (7a) – (7b) by their respective total doping concentration:

$$g_{\rm D}(T, N_{\rm don}) = \frac{b}{g+(1-b)\,N_{\rm don}/n_1}, \tag{8a}$$

$$g_{\rm A}(T, N_{\rm acc}) = \frac{b}{g+(1-b)\,N_{\rm acc}/p_1}. \tag{8b}$$

The difference to the original model Eqs. (3a) – (3b) in neutral, uncompensated regions is shown in Fig. 4. These deviations are minor compared to the scattering of the experimental data. Note, that for low doping concentration the effective degeneracy factors converge to the inverse g:

$$g_{\rm D,A}(T, N_{\rm dop}\to 0) \to g^{-1}\ . \tag{9}$$

When using Eqs. (6a) – (6b) in a device simulator with *constant* $g_{\rm D,A} = g^{-1}$ leads to an unphysical monotonous increase of ii with rising doping. Sometimes, as in DESSIS–Synopsys, a hard transition to complete ionization is enforced above $N_{\rm crit}$. Our doping-dependent degeneracy factors (8a) – (8b) describe a smooth transition to complete ionization

$$g_{\rm D,A}(T, N_{\rm dop}\to\infty) \to 0 \tag{10}$$

based on the physical process that bound states gradually disappear in favor of extended states as the doping exceeds $N_{\rm crit}$.

The behaviour of a physical model developed for homogeneous conditions has to be carefully checked in inhomogeneous regions of a device. Replacing densities in the effective degeneracy factors (7a) – (7b) by the respective total doping concentration will cause a certain error in space charge regions. Inside depleted p-n junctions one should better replace

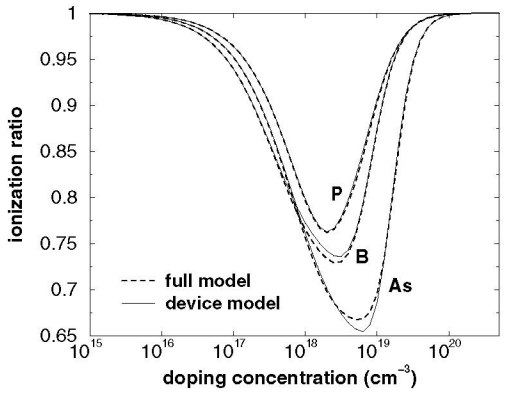

Fig. 4. Ionization level for P, B, and As using the original model Eqs. (3a) – (3b) (dashed lines) and the approximated device model Eqs. (6a) – (6b) with (8a) – (8b) (solid lines).

them by zero in accord with the Schottky approximation (depletion approximation). However, as is obvious from Eqs. (3a) – (3b), and (6a) – (6b), the degree of ionization becomes equal to 1 when the free carrier density vanishes. Therefore, the form of the effective degeneracy factors $g_{\mathrm{D,A}}$ does not influence the degree of ionization as long as the absolute value of the energetic separation between dopant level and quasi-Fermi level is large compared to kT (and positive for donors, negative for acceptors). The only slight error then occurs at the boundaries of depletion regions. In the case of high injection, the excess density strongly screens the Coulomb potential and would rather make the ionization complete. However, a large error is not to be expected, if the local density takes the value of the injected density, but not that of the local ionized doping. This is indeed the case in the base of a bipolar transistor under high injection conditions.

V. Application Examples

Bipolar Transistor

Fig. 5 shows the simulated Gummel characteristics and the common-emitter current-gain of a double poly-emitter npn bipolar transistor of a $0.3\,\mu$m BiCMOS process, where the geometry and doping information was taken from Ref. [12]. One observes an increase of the maximum gain by 25% due to the effect of ii. This increase, which is related to the increase of the collector current at moderate V_{BE}, is directly caused by the lower active boron doping in the base, since the gain is proportional to $N^{+}_{\mathrm{don,emitter}}/N^{-}_{\mathrm{acc,base}}$.

Solar Cell

To demonstrate the maximum possible impact of ii on the operation of crystalline Si thin-film solar cells, which are asymmetrical pn-diodes, we chose $N_{\mathrm{dop}} = 2{\times}10^{18}$ cm^{-3} in the boron-doped base and a $1\,\mu$m deep phosphorus diffusion to form a Gaussian emitter layer with a peak dopant density of $5{\times}10^{18}$ cm^{-3}. The simulated open-circuit voltage V_{oc} decreases due to ii from 624 mV to 614 mV, because the band bending across the pn junction is reduced. The reduction of the hole majority carrier density increases the electron minority carrier density and hence the recombination rate in the base.

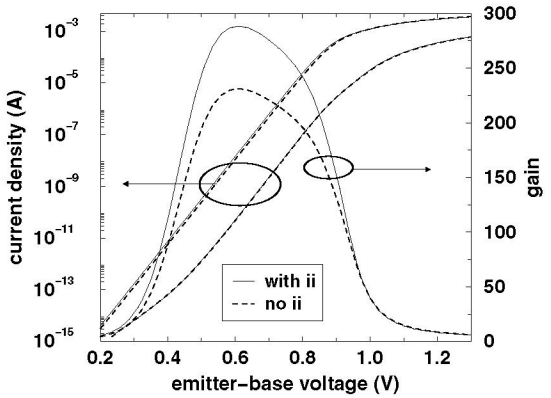

Fig. 5. Gummel characteristics and gain of a state-of-the-art double poly-emitter npn bipolar transistor.

In precise cell simulations, a reduction of V_{oc} by 10 mV is a noticeable effect. The impact on the energy conversion efficiency of the solar cell depends on the cell's optical and surface properties. In the case simulated here, cell efficiency drops from 14.3% to 13.5%. If the surface is well passivated, the efficiency drop is larger, because ii is a bulk property, which dominates in devices with good surface passivation.

MOS Devices

Besides the RF bipolar transistor and the thin-film solar cell, the impact of ii was also studied for other devices and phenomena: the IV characteristics of a PDSOI MOSFET, the CV characteristics of a MOS capacitor, and the gate direct tunnel current in a MOSFET with ultra-thin gate oxide. In all examples, the active device regions had a doping concentration near the Mott density. Although ii had a visible effect in all cases, its magnitude was found not to be of practical interest. In a MOSFET, even with a boron substrate concentration of $2 \times 10^{18}\,\mathrm{cm}^{-3}$, ii becomes negligible, since in the depletion region under the gate the effect of ii is not present (large distance between boron level and Fermi level), and in the inverted channel the density is pinned by the amount of gate charge due to the constraint of global charge neutrality.

VI. Conclusion

It is possible to derive an empirical model of ii in phosphorus-, arsenic-, and boron-doped crystalline silicon as function of free carrier densities based on experimental data and a physical DOS model. Deviations caused by necessary simplifications to achieve compatibility with the PMI of DESSIS–Synopsys can be fully covered by the re-adjustment of some parameters. The model yields the correct behaviour in depletion zones, whereas the failure in regions of high injection becomes ineffectual, because the local density is determined by the injected density there. Simulations show that ii can change the common-emitter current gain of a bipolar transistor by 25% and the open-circuit voltage of a solar cell by 10 mV. MOS devices are hardly affected by ii.

Acknowledgment

The first author is grateful for the financial support by Fujitsu Laboratories Ltd.

References

[1] H. Watanabe and S. Takagi, *Jour. Appl. Phys.* **90**, 1600 (2001).
[2] G. Xiao, J. Lee, J. J. Liou, and A. Ortiz-Conde, *Microelectronics Reliability* **39**, 1299 (1999).
[3] Y. Yue and J. J. Liou, Solid-State Electronics **39**, 318 (1996).
[4] B. Bergersen, J. A. Rostworowski, M. Eswaran, R. R. Parsons, and P. Jena, *Phys. Rev.* **B14**, 1633 (1976).
[5] K. P. Abdurakhmanov, S. Mirakhmedov, and A. T. Teshabaev, *Sov. Phys. Sem.* **12**, 457 (1987).
[6] E. O. Kane, Phys. Rev. **131**, 79 (1963).
[7] F. Mousty, P. Ostoja, and L. Passari, *Jour. Appl. Phys.* **45**, 4576 (1974).
[8] S. S. Li and W. R. Thurber, Solid-State Electronics **20**, 609 (1977).
[9] W. Kuzmicz, Solid-State Electronics **29**, 1223 (1986).
[10] W. Nuyts and R. van Overstraeten, *Jour. Appl. Phys.* **42**, 3988 (1971).
[11] DESSIS 10.0 User's Manual, Synopsys, 2005.
[12] H. Nii, C. Yoshino, S. Yoshitomi, K. Inoh, H. Furuya, H. Nakajima, H. Sugaya, H. Naruse, Y. Katsumata, and H. Iwai, *IEEE TED* **46**, 712 (1999).