Theory of the screened Coulomb field generated by impurity ions in semiconductors

Young D. Kwon
California State University, Fullerton, California 92834, USA
(Received 3 August 2005; revised manuscript received 6 February 2006; published 19 April 2006)

This theory examines the effect of the screened Coulomb field generated by impurity ions on various semiconductor phenomena. The exact equation for the screening length, which is derived for semiconductors, is a function of temperature, doping density and intrinsic carrier density. In silicon the screening length varies from a fraction of a nanometer to thousands of nanometers depending on the temperature and the doping density. The overlap of these impurity fields causes both crystal potential shift and repulsion among impurity ions. The repulsion among the impurity ions gives rise to a pressure build-up, which is attributed to such phenomena as solid solubility and diffusion-induced defect generation. The experimental solubilities of lithium in germanium, and boron, phosphorus, and arsenic in silicon are compared with the theoretical values. Also, the optically enhanced solubility during rapid thermal processing is explained.

DOI: 10.1103/PhysRevB.73.165210 PACS number(s): 61.72.Bb, 61.80.Az, 78.70.−g

I. INTRODUCTION

The impurity doping of a semiconductor has been known only as a process by which the host material is converted into either *n*-type or *p*-type. However, little attention has been given to the interaction among impurity ions. As long as the overall charge neutrality is maintained, these ions have been treated as noninteracting point charges. The prevailing view has been that after these ions generate electrons or holes, their only role is as space charge. Consequently their other contributions to various semiconductor phenomena have been completely overlooked.

In 1969, this author published a paper explaining the effect of impurity field on the Fermi-level shift and the built-in electric field in semiconductors.[1] Although the screened Coulomb field around an isolated impurity ion has been discussed in many textbooks[2–4] and some research papers,[5–8] these studies have failed to look into the collective behavior of these ions.

A mathematical analysis shows that an impurity ion in a semiconductor is shielded by oppositely charged mobile carriers as part of the neutralization process. Because of the screening effect of the shielding electrons or holes, the resultant Coulomb field around the ionic charge becomes short range. However the sphere of influence extends to a certain distance before it becomes negligible. Therefore, if another impurity ion is brought close enough, they begin to see each other's Coulomb field through their shielding electron (hole) clouds. When the doping is very light, the inter-ionic distance is far enough not to cause any Coulomb interaction among neighboring impurity ions. However, as the semiconductor's doping becomes heavier, the ions come closer to experience each other's Coulomb field. Depending on their proximity, they can repulse strongly.

The screening effect comes from free charge carriers that are attracted to the ionic charge. These free carriers neutralize the ionic charge collectively by staying closer to the ion. Since none of these charge carriers are bound to any ion, numerous free electrons or holes participate in the shielding. As it turns out, the shielding is much more effective if the carrier density is high. The thicker the electron (or hole) cloud, the more effective is the shielding. However, what differs for the impurity field in a semiconductor in comparison to a similar Coulomb field in a vacuum is its dynamic nature. A Coulomb field inside a vacuum is steady and constant. However, the impurity field in a semiconductor fluctuates depending on the temperature and instantaneous carrier density variation. The impurity field in a semiconductor shrinks or expands as the electron (hole) density fluctuates.

In wafer fabrication, the processes like diffusion, oxidation and annealing take place at high temperatures in the order of 900–1100 °C. At this high temperature range, the carrier density in the wafer is high and the screening length becomes very short. Therefore, the impurity ions' interaction is minimal. However, as the wafers are cooled after a high temperature process, the carrier density decreases rapidly. This in turn causes the screening length to increase. As the sphere of the impurity field expands, the impurity ions see each other's Coulomb field and either attract or repulse strongly. Hence, the attraction promotes ion-pairing (compensation), and the repulsion affects such phenomena as solid solubility, impurity precipitation, diffusion induced defect generation, etc.

II. SCREENED COULOMB POTENTIAL OF AN IMPURITY ION

For simplicity, let us assume a nondegenerate *N*-type semiconductor in which all the donors are ionized. Then from the charge neutrality condition

$$n_o - p_o = N_D, \tag{1}$$

where n_o, p_o, and N_D are electron, hole, and impurity concentrations at thermal equilibrium respectively. The electron and hole concentrations are given by

$$n_o = n_i e^{q\phi_b/kT}, \tag{2}$$

$$p_o = n_i e^{-q\phi_b/kT}, \tag{3}$$

where n_i, ϕ_b, q, and k are the intrinsic electron concentration at temperature T, the Fermi-level shift, the electronic charge,

and the Boltzmann constant, respectively. Substituting Eqs. (2) and (3) into (1), N_D can be expressed as

$$N_D = 2n_i \sinh \frac{q\phi_b}{kT}. \tag{4}$$

In order to understand the behavior of the Coulomb field around an ionized impurity, we have to solve the Poisson's equation at the vicinity of an ionic charge q,

$$\nabla^2 V = -\frac{q}{\varepsilon}(p - n + N_D), \tag{5}$$

where n and p are the local carrier densities which correspond to the potential perturbation V near the point charge. Substituting Eqs. (2)–(4), into (5),

$$\nabla^2 V = \frac{2n_i q}{\varepsilon}\left\{\sinh \frac{q(\phi_b + V)}{kT} - \sinh \frac{q\phi_b}{kT}\right\}. \tag{6}$$

$$\text{For } V \ll \frac{kT}{q}, \tag{7}$$

$$\nabla^2 V \approx \frac{2n_i q^2}{\varepsilon kT}\left(\cosh \frac{q\phi_b}{kT}\right)V. \tag{8}$$

The condition set by Eq. (7) can be satisfied at the fringe of the Coulomb field away from the point charge. The solution of the above equation for a point charge, q, in the polar coordinate system is

$$V = \frac{q}{4\pi\varepsilon r}e^{-r/r_s}, \tag{9}$$

where

$$r_s = \sqrt{\frac{\varepsilon kT}{2n_i q^2 \cosh \frac{q\phi_b}{kT}}}. \tag{10}$$

Equation (9) shows that the Coulomb potential due to an ionic charge, $+q$, in a semiconductor is the normal Coulomb field, $q/4\pi\varepsilon r$, modified by the screening factor, $\exp(-r/r_s)$. Here r_s is the well-known screening length or Debye length. In a semiconductor, the screening of the Coulomb potential is due to the neutralizing electron and hole clouds that shield the impurity ions. As a result the Coulomb field attenuates rapidly as the distance from the point charge is increased. According to Eq. (10), the screening length r_s is a function of temperature, carrier density, and the Fermi-level shift.

Equation (9) is identical to the Debye-Hückel[9] equation for an ionic charge in electrolytic solution. However, the main difference is in the parameters that determine the screening length. In electrolytes, there are cations and anions in equal numbers. In semiconductors there are electrons, holes, and fixed ions.

From the denominator of Eq. (10),

$$2n_i \cosh \frac{q\phi_b}{kT} = n_i(e^{q\phi_b/kT} + e^{-q\phi_b/kT}) = n_0 + p_0. \tag{11}$$

Substituting Eq. (11) into (10),

$$r_s = \sqrt{\frac{\varepsilon kT}{q^2(n_0 + p_0)}}. \tag{12}$$

The above equation gives a better insight into the nature of the screening effect. The screening length is now a function of temperature and the carrier densities only. That is, the screening length is proportional to the square-root of temperature and inversely proportional to the square-root of the mobile carrier densities.

The physical interpretation of the above equation is that, as temperature rises, mobile carriers acquire more kinetic energy and the shielding cloud tends to spread farther out, thus increasing the screening length. On the other hand, the raised temperature generates more electron-hole pairs, thickening the shielding electron cloud, and thus shortening the screening length. Therefore, as the temperature is raised, two physical phenomena compete to decide the final screening length. One is the increased kinetic energy of carriers, which tends to increase the screening length, and the other is the generation of more electron-hole pairs, which tends to make the screening length shorter by thickening the shielding electron cloud. Depending on which process dominates, the screening length either shrinks or expands.

From Eq. (1)

$$N^2 = n_0^2 - 2n_0 p_0 + p_0^2 = n_0^2 - 2n_i^2 + p_0^2. \tag{13}$$

Here we drop the subscript from N_D to generalize the equation. Manipulation of Eq. (13) yields

$$n_0 + p_0 = \sqrt{N^2 + 4n_i^2}. \tag{14}$$

Substituting Eq. (14) into (12),

$$r_s = \sqrt{\frac{\varepsilon kT}{q^2(N^2 + 4n_i^2)^{1/2}}}. \tag{15}$$

The above equation relates the screening length to the doping density and the intrinsic carrier density.

III. CHARACTERISTICS OF SCREENING LENGTH, r_s

One can visualize an impurity ion surrounded by the sphere of a shielding electron cloud. Within the sphere of screening radius r_s, the Coulomb field is strong. But it decays exponentially outside and becomes negligible. The screening radius fluctuates as the temperature or the carrier density is varied.

From Eq. (18), the screening length can be estimated for a given temperature. The accuracy of the n_i value more or less dictates the accuracy of the calculated screening length. In this study, the empirical equation proposed by Morin *et al.*[10] for silicon is used. It was derived from experimental data measured up to 1000 °K (727 °C).

$$n_i = \left(1.5 \times 10^{33} T^3 \exp\left(\frac{-1.21 + \Delta E}{kT}\right)\right)^{1/2}, \tag{16}$$

$$\text{where } \Delta E = 7.1 \times 10^{-10}\left(\frac{n_i}{T}\right)^{1/2}. \tag{17}$$

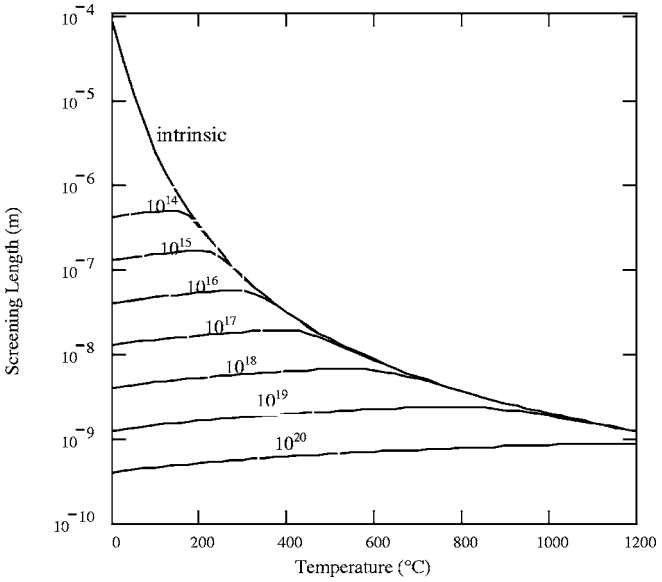

FIG. 1. Screening length vs temperature.

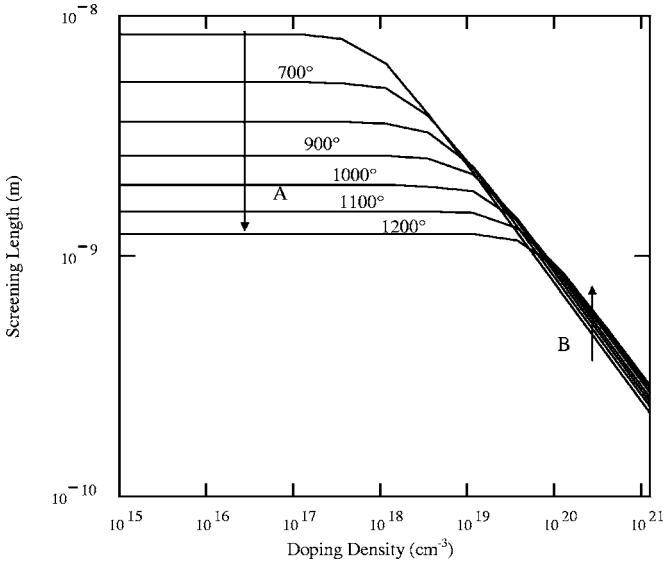

FIG. 2. Screening length as a function of the doping density.

Although the validity of the above equation is very good for the temperature range up to 700 °C at which the electron-hole pair generation is purely thermal, its accuracy at higher temperature is somewhat questionable. The reason why the above empirical equation may not be accurate at a high temperature is because it represents only the thermally generated carriers. However, at the high temperature ranges above 900 °C, the photo-electric generation of carriers cannot be ignored. In the red-hot furnace where the samples are heated, photons from the heating element, which would naturally have a higher temperature than the furnace temperature, would continuously bombard the samples. Also, the energy gap at this temperature range shrinks by about 0.3–0.5 eV,[11,12] making the photo-generation more pronounced. Therefore, the appreciable portion of the carrier generation must be photo-electric. Consequently when the photo-electric effect is taken into account the carrier density in this temperature range would be much higher than the value calculated by Eq. (16). McCaldin[13] reported the measured n_i value of GaAs at 1000 °C to be 4×10^{18} cm^3 which is about six times greater than the value predicted by the extrapolation of lower temperature Hall measurement data. Also this value is about ten times greater than the value calculated by the Morin-Maida's equation for GaAs,[11] 4×10^{17} cm^3. The wide discrepancy can only be explained by the unaccounted photo-generated carriers. In the absence of a reliable equation which would take the photo-effect into account, the above empirical equation will be used, keeping in mind that the carrier density at a high temperature is much higher due to the photo-effect.

The screening length r_s for silicon calculated from Eq. (15) is plotted as a function of temperature and doping density in Figs. 1 and 2. In Fig. 1, the screening length increases as the temperature is raised, initially indicating the effect of the kinetic energy being predominant over that of thermal generation. After peaking, the screening length decreases as the temperature is further raised, probably because the effect of the thermal generation becomes predominant.

Also, note that, as the temperature is raised higher, the curves converge on the one that corresponds to the intrinsic carriers only, because the thermally generated carrier density is higher than that of the doping density. Figure 2 shows the screening length versus the impurity concentration for various temperature values. The screening length is almost constant up to the point where the doping density approaches the intrinsic carrier density, and thereafter it decreases rapidly as the doping density is further increased. It shows, at a given temperature, the screening length remains constant until the doping density approaches to that of the intrinsic carriers. As the doping density approaches and exceeds the intrinsic carrier density, the screening length starts to shrink because overall carrier density is increased. Another way to look at Fig. 2 is that, on the left side of the curves, shown by the arrow "*A*," the effect of the carrier generation is predominant, so the screening length decreases as the temperature is raised. The right side, shown by arrow "*B*," is where the effect of kinetic energy predominates over the carrier generation.

IV. FERMI-LEVEL SHIFT AND IMPURITY FIELD

It is worthwhile to test the validity of the theory by calculating one of the well-known semiconductor parameters. One such test is that if the impurity field truly exists in the semiconductor as described in Eq. (9), the superposition of all such potentials from individual impurity ions would add up to modify the original crystal potential. The amount of total potential shift thus calculated can very well be equal to the well-known Fermi-level shift[14] derived from Eq. (4),

$$\phi_b = \frac{kT}{q}\sinh^{-1}\frac{N_D}{2n_i}. \tag{18}$$

Physically, the doping of a semiconductor is equivalent to the injection of fixed ions and the same number of oppositely charged carriers into the crystal lattice. Since the equal and

opposite charges are injected, the total system remains neutral. However, the Fermi level shifts when the impurity atoms are ionized. The impurity ions give rise to individual impurity fields inside a semiconductor. Depending on the density of the ions in the semiconductor, the tail portions of the impurity fields overlap to change the electro-static potential inside the semiconductor. Since every ionized impurity atoms has its own Coulomb field, the superposition of these fields should add up to an average potential shift that must be identical to the Fermi-level shift of Eq. (18).

For a uniformly doped semiconductor, let us use a jellium model in which the charge density, qN_D, is uniformly distributed. Then, the potential perturbation dV at an arbitrary point which is assumed to be the center of a spherical shell element $4\pi r^2 dr$ can be obtained by use of Eq. (9). The spherical charge density at radius r from the center is $4\pi r^2 qN_D dr$. Because of the spherical symmetry, dV can be found as

$$dV = \frac{qN_D}{\varepsilon} re^{-r/r_s} dr. \tag{19}$$

Now integrating the above equation over all space, we obtain the total potential shift due to the surrounding impurity ions.

$$V = \frac{qN_D}{\varepsilon} \int_0^\infty re^{-r/r_s} dr, \tag{20}$$

or

$$V = \frac{qr_s^2 N_D}{\varepsilon}. \tag{21}$$

Equation (18) is the Fermi-level shift predicted by the energy band theory and Eq. (21) is the potential shift calculated from the screened Coulomb fields of impurity ions. Although the equations are not identical, they could still represent the same entity within the confines of Eq. (7). If they are identical, then the Fermi-level shift can now be attributed to the screened Coulomb field of the ionized impurities.

The numerical comparison of Eqs. (18) and (21) is made in Fig. 3 by plotting the computed values of each equation. The two sets of curves coincide perfectly within the limit imposed by Eq. (7).[15] It is evident from the figure that indeed the Fermi level shift in a doped semiconductor is caused by the screened Coulomb potential of individual impurity ions. Also this result validates the impurity field theory of this paper.

Figure 4 depicts the effect of the impurity field on the Fermi-level shift. The Fermi potential before the doping is shown by the solid line which is equivalent to the intrinsic level. After the doping, the impurity ions set up their screened Coulomb field to modify the electronic potential by "V" as shown.

V. IMPURITY-IMPURITY INTERACTION

In a lightly doped semiconductor, the impurity ions do not see each other's Coulomb field if the inter-impurity distance is much farther than the screening length. However, as the doping density is increased, the impurity fields from adjacent

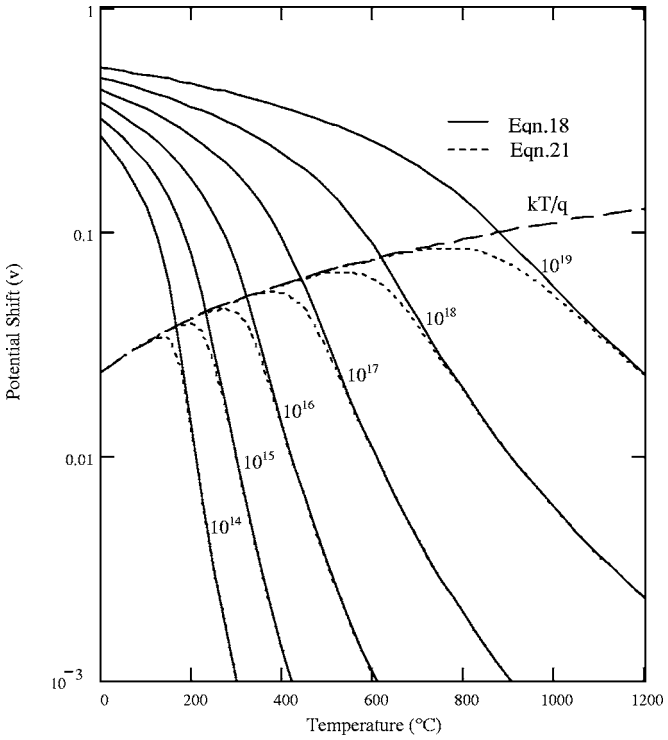

FIG. 3. Comparison of Eqs. (18) and (21).

impurity ions start to overlap and the ions start either to repulse or attract each other depending on the polarities. Since the normal doping of a semiconductor involves the same type (N or P) at a time, the impurity interaction is repulsive.

Figure 5 is a simplified depiction of the ionic interaction. Each circle represents an elastic ball within which the Coulomb field is strong enough to affect mutual repulsion or attraction. Within the ball, the impurity field is strong, but outside it is negligible. Although it is somewhat an oversimplification, it helps us to visualize the impurity interaction. In Fig. 5(a), the balls touch and push each other indicating overlap of the impurity fields, which would lead to a strong repulsion. In Fig. 5(b), the balls are somewhat separated so that the interaction is negligible. Also, note that the size of the balls expands or shrinks depending on the temperature and the local carrier density based on Eq. (15). For example, if a sample depicted in Fig. 5(a) is illuminated by a strong beam of light, then the balls shrink to a small size and they do not touch each other. This is because the light beam generates a high density of electron-hole pairs which make the screening length shrink. The new, thick electron and hole clouds shield the ions effectively to make the impurity field within a very short range. However, if the light beam is turned off, then the excess carriers recombine quickly mak-

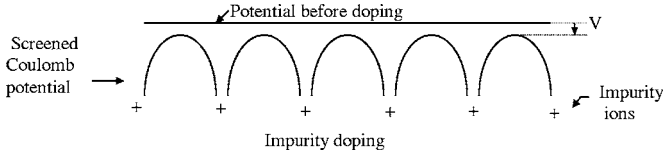

FIG. 4. Impurity field and potential shift.

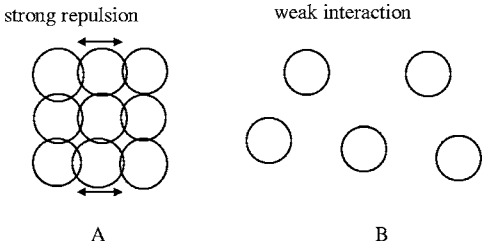

FIG. 5. Impurity repulsion.

ing the balls to expand to their original size. This simple picture is useful in understanding the behavior of impurity ions during high temperature processing.

A. Internal pressure

If the impurity density is high enough to cause strong repulsion, then the repulsion leads to a pressure build up in the host crystal. In Sec. IV, it was established that the impurity fields add up to boost the crystal potential by the amount of the Fermi-level shift ϕ_b. Assuming that the impurity ions are brought from infinity to the crystal one by one to raise the potential from zero to ϕ_b, then the total energy required to do the job is

$$W = \frac{1}{2} q \phi_b N V, \tag{22}$$

where N is the density of impurity ions and V is the volume of crystal. Therefore, NV is the total number of ions in the crystal. The pressure generated by the repulsion can be derived by differentiating the energy by volume V,

$$P = \frac{dW}{dV},$$

$$= \frac{1}{2} q \phi_b N. \tag{23}$$

From the above equation, the pressure is a function of the impurity density and the Fermi-level shift. Substituting Eq. (18) into (23),

$$P = \frac{1}{2} kTN \sinh^{-1} \frac{N}{2n_i}. \tag{24}$$

Now the pressure is a function of doping and the intrinsic carrier densities. The above equation is plotted as a function of the doping density for several temperature parameters in Fig. 6. It is clear from the figure that the pressure is negligible for a low to moderate doping density, but increases beyond the doping density of about 10^{19} atoms/cm^3. High temperature pressure curves in Fig. 6 would shift right, as shown by the dotted line in the figure, if the n_i value was modified to take into account the optically generated electron-hole pairs. The dotted line is a hypothetical curve when n_i value is increased by tenfold due to optical generation of carriers at 1100 °C (see note in Sec. III).

It is important to note that the inter-impurity repulsion gives rise to tensile pressure which induces more vacancy generation than without the pressure. It also implies that the highly doped region of a crystal has a higher pressure and more vacancies than the rest of the region.

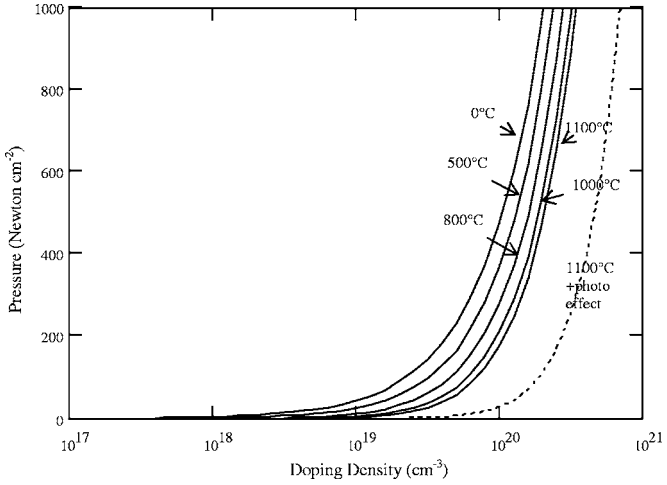

FIG. 6. Pressure vs active impurity density.

B. Ion-pairing and impurity compensation

When both donor and acceptor impurities are present in a crystal, the oppositely charged impurity ions attract each other during the diffusion process. If they are close enough they form ion-pairs, either in the form of dipoles or chemical compounds.[16,17] In other words, the ion pairs can have direct chemical-bonds to form a compound or they can be just dipoles without chemical bonds. As soon as the ions are paired, their electric field becomes localized between the two ions and the original radial impurity field of each ion disappears. The paired ions automatically neutralize each other making the shielding electron-cloud unnecessary. This is why the compensated impurities do not generate free carriers.

It is proven fact that the ion-pairing increases the solid solubility of the impurities involved.[16,18] Reiss *et al.*[16] discovered that the solubility of lithium (donor) increased when the host crystal (germanium) was pre-doped with known quantity of gallium (acceptor). In another experiment, they reported the increase of mobility in compensated crystal compared to one without compensation. That is, they measured Hall mobility of a germanium sample doped with known quantities of gallium and then measured the mobility after the same sample was partially compensated by lithium. The latter showed marked improvement in Hall mobility despite the fact that it contained a higher density of total impurities (gallium plus lithium). They attributed this to the disappearance of scattering centers. These findings are consistent with the present theory that the ion-pairing eliminates the original screened Coulomb fields of each ion.

VI. SOLID SOLUBILITY

Trumbore[19] studied and compiled the solid solubility of various impurities in silicon and germanium. His solubility chart is still being used universally. He proposed three factors that might affect the solubility, namely, chemical com-

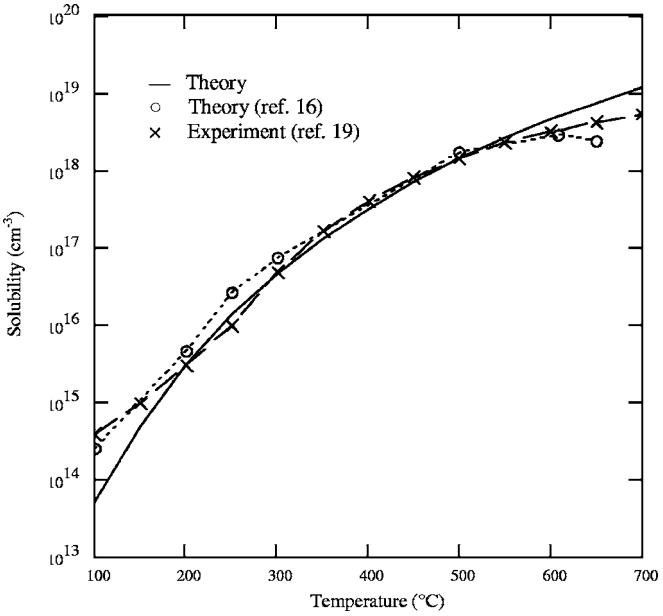

FIG. 7. Li solubility in germanium.

patibility, atomic size, and crystal structure. He showed correlations of the solubility with the atomic sizes and the heats of sublimation of the impurity elements. He also suggested that the solubility is determined by the thermodynamic free energy of the crystal at a given temperature. Shockley and Moll[20] studied the solubility of charged impurities in heavily doped semiconductors and proposed the correlation between the solubility and the Fermi-level.

According to the Trumbore chart, the substitutional impurities in silicon can be divided into two groups. The first group, which shows the highest solubility, includes P, B, and As. These impurities are chemically compatible with silicon and their atomic sizes are either about the same or smaller than silicon atom. For these impurities, the structural mismatch is not much of a problem. The next group, which includes Al, Sb, Sn, and Ga, has lower solubility values, possibly due to their atomic sizes being bigger than silicon. If the impurity atoms are bigger in size, then they have difficulty in fitting into the host crystal. In this study, the discussion will be limited to the former group to study the effect of the screened Coulomb field on the solubility.

The solubility is defined as the maximum concentration of a given impurity atom that can be incorporated into a host crystal at a given temperature. It is a well accepted concept that the solubility is the value that minimizes the thermodynamic free energy of the host system. Reiss *et al.*[16] carried out an extensive study on Li diffusion in Ge and Si. Also, they developed a theoretical solubility formula which was based on the law of mass action and the Fermi statistics. Their calculated solid solubility agreed very well with the experimental values, which is reexamined in Fig. 7.

However, there has been no basic understanding of the microscopic behavior of individual impurity atoms in relation to the solubility. In light of this theory, we can now visualize the situation much better. Under the light doping condition, the impurity ions do not see each other's Coulomb field and, therefore, they behave like noninteracting point charges. When the doping density approaches that of solid solubility, the inter-impurity distance becomes short and the impurity field of neighboring ions overlap, which gives rise to a repulsive force. Because of the exponential factor in the screened Coulomb field, the repulsive force increases drastically as two ions are pushed closer. The solubility is determined by the critical inter-impurity distance at which the repulsive force becomes so strong that further introduction of impurities is rejected.

The same phenomenon can be explained by using the pressure build-up shown in Fig. 6. The repulsion among the impurity ions gives rise to internal pressure as shown in the figure. At an elevated temperature the pressure is negligible up to about 10^{19} atoms/cm^3. However, the pressure curve rises beyond this point and the slope becomes steeper at about 10^{20} atoms/cm^3 and beyond. At a certain concentration, the pressure build-up becomes so excessive that no more impurity can be introduced substitutionally beyond this limit. Note, however, that the curves in Fig. 6 have to be modified if there is an optical generation of carriers, which is the case with a temperature above 900 °C and also with the RTP process. The dotted curve in Fig. 6 is such a modification assuming that the carrier density increased tenfold by photo-generation. Tenfold is arbitrarily chosen in reference to McCaldin's finding on GaAs mentioned in Sec. III of this article.

If the aforementioned elastic ball model is used, the radius of the ball is determined by the balance of the repulsive force among the impurity ions and the counter retaining force composed of the chemical bond and the potential barrier of the crystal lattice. If the repulsive force is stronger than the retaining force imposed by the crystal lattice, then the impurity ions would be dispersed by force, either to precipitate, by breaking chemical bonds, or to undergo enhanced diffusion to be pushed out. Therefore, the solubility is simply the maximum number of the elastic balls that can be safely squeezed into a unit volume at a given temperature.

As a first order of approximation, let us assume that the radius of the elastic ball is the same as the screening length. This is a reasonable approximation given that the Coulomb field increases exponentially if the radius becomes shorter than the screening length. Also if another ion is brought within this proximity, then a strong repulsion is to be expected. Further, it is assumed that the solubility is the number of these balls which can be close-packed into a unit volume of the host crystal. The close-packing of these balls is achieved by assuming that the packed balls form tetrahedral pyramids, with one ball sitting on top of three triangularly packed balls. The corresponding number of balls that can be packed this way in a unit volume (/cm^3) is

$$N = \frac{1}{5.657 r_s^3} \tag{25}$$

Note that this close packing density is higher than the packing density of cubes with size $(2r_s)^3$.

To test this theory, the solubility of Li in germanium crystal is analyzed. The atomic size of germanium (atomic number 32) is much bigger than that of lithium (atomic number

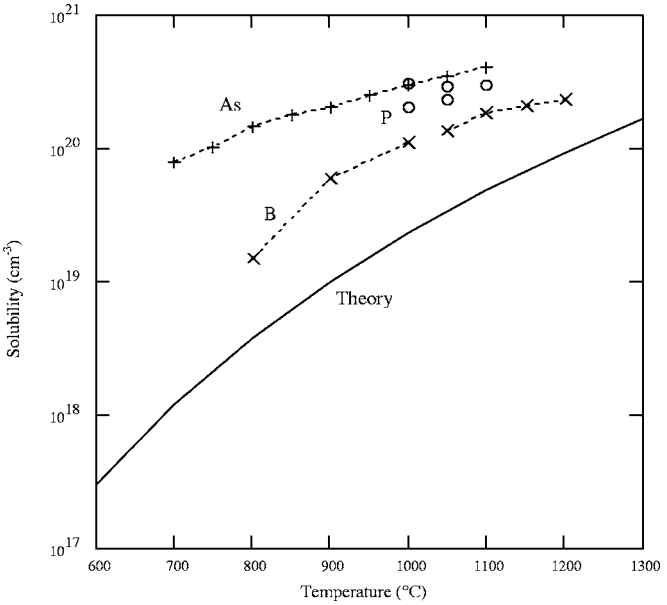

FIG. 8. Solubility of active B, P, and As in silicon.

3), and these two atoms do not form chemical bonds. However, the lithium atoms ionize to become donors. As a result, the lithium ions diffuse relatively easily through the interstitial voids of germanium crystal. Fortunately, we have experimental solubility data in the temperature range of 100–700 °C.[19] Also, for this temperature range, the accurate carrier density can be calculated by using the Morin-Maita's equation,

$$n_i = \left(3.1 \times 10^{32} T^3 \exp\left(\frac{-0.785 + \Delta E}{kT}\right)\right)^{1/2}, \qquad (26)$$

$$\text{where } \Delta E = 4.61 \times 10^{-10}\left(\frac{n_i}{T}\right)^{1/2} \qquad (27)$$

The packing density of these balls with radius r_s is calculated using Eqs. (10), (25), and (26), and is compared with the experimental data on Li solubility in germanium. Although it may seem an oversimplification, at least this method is the first step toward estimating the theoretical solubility which would shed some light on the relationship between the screened Coulomb field and the solubility. The solubility calculated from the packing density is plotted in Fig. 7, together with the theoretical values of Reiss *et al.*[16] and the experimental values of Ref. 19. The solid line is for the calculated solubility and the dotted line having "∘" represents the theoretical values of Reiss *et al.* The experimental solubility of Li from the Trumbore[19] curve is retraced with "×" markings. A remarkable agreement is found between the two different set of theoretical values and the experimental solubility, although it deviates somewhat at the high and low ends of the temperature range.

A similar comparison is made for active B, P, and As solubility in silicon for the 800–1200 °C range. In Fig. 8, the solid line represents the theoretical values calculated from Eqs. (10), (16), and (25). The dotted lines having "×" markings representing B are from the experimental data of Ryssel *et al.*[21] and Maekawa and Oshida.[22] The upper dotted line with "+" markings represents the experimental solubility of As from Guerrero *et al.*[23] The cluster of "∘" markings represents experimental values for P from Refs. 28–30. As expected the theoretical curve is fairly lower than that of the experimental curves. This discrepancy can be attributed to the underestimation of carrier density in Eq. (16). This equation does not account for the carriers generated by the photoeffect, which would be substantial at the temperature range of interest. Also, the discrepancies between the solubility of B, P, and As are noteworthy. This may be due to the fact that the size of the ions in the host crystal could hinder the mobility of the ions under Coulomb repulsion. The bigger ions are harder to move once they settle into a crystal lattice.

Figures 7 and 8 strongly suggest that, indeed, there is a strong correlation between the screened Coulomb field and the solid solubility of electrically active impurities. However, more research is needed to clarify the relationship between the solubility and various other factors, such as atomic size, chemical bond strength, and activation energy.

A. Optically enhanced solubility

Rapid thermal process (RTP) has been used to anneal implanted impurities. Regolini *et al.*[24] reported activation of implanted As to 10^{21} cm^3 by electron beam and laser annealing at the wafer temperature of 350 °C. Also Lietoila *et al.*[25] reported the activation of implanted As in silicon to 5×10^{20} cm^3 by RTP annealing at 560 °C. These values are way above the solid solubility limit. Singh *et al.*[26,27] reported that they were able to carry out shallow high density phosphorus diffusion at 700 °C by use of UV light activated RTP process. They achieved the surface concentration of about 3×10^{20} atom/cm^3 with the junction depth of 0.125 μ after 30 min diffusion. This value also exceeds the solid solubility at the temperature and is about the same order of magnitude as those reported for 1050 and 1100 °C with the conventional furnace diffusion.[28,29,31,32]

The RTP process relies on optical radiation for heating the wafers. The illumination of strong light onto the wafer surface will generate numerous electron-hole pairs in the surface area by the photo-electric effect. From Eq. (10) or (12), it is obvious that the increased carrier density will make the screening length very short. As the impurities are diffused under the continuous illumination, the high density of electron-hole cloud shields the impurity ions more effectively. As a result, a higher density of impurity ions can be introduced into the wafer beyond the conventional solubility limit. This fact is illustrated by a dotted line in Fig. 6.

VII. DIFFUSION INDUCED DEFECT GENERATION

Tannenbaum[28] first reported the anomalous diffusion of phosphorus into silicon and discovered that the diffusion profile she obtained not only deviated from the ideal profile predicted by the simple diffusion theory but also an appreciable portion of the diffused impurities were electrically inactive. The diffusion profile of the phosphorus determined by a radio-tracer showed much greater density than that deter-

mined by conductivity data. She attributed these electrically inactive phosphorus atoms to precipitation. Her finding was confirmed by many other researchers.[22,29,31,32]

When the impurities are introduced into the host crystal from an infinite source, the surface area of the crystal becomes saturated with impurities to a solid-solubility limit. At the high furnace temperature of about 1000 °C, the furnace heating element had to be heated to a higher temperature to compensate for the heat loss of the open-tube system. In other words, the heating element and its immediate surroundings would be at higher temperature than that of wafers to maintain the necessary temperature gradient needed to support required wafer temperature. These overheated elements radiate visible and infrared lights just like an incandescent light bulb. When the wafers are irradiated with these light beams at high temperature, numerous electron-hole pairs are generated in addition to the normal intrinsic carrier density due to thermal generation. As a result the carrier density in the wafers is much higher than that predicted by thermal generation alone.

Naturally, the solubility in the wafer is boosted by the optically generated electron-hole pairs as discussed in the previous section. Thus, the wafer absorbs higher density of impurities than it would in the absence of photo-generated carriers. After the diffusion process, when the wafers are pulled out quickly, the excess carrier density due to the optical generation drops abruptly because the wafers are out of the optical zone. There was no motorized loading system in the old days. Wafers were pushed in and out manually and quickly. As the wafers are cooled down, the carrier density drops even further though not as quickly as the optical portion. As the carrier density drops, the screening length expands causing a sudden increase in ionic repulsion and internal pressure.

The diffusion-induced defects and crystal damage have been well studied in the past.[29,31,32,34] The heavy doping of impurities by the diffusion process not only produces a high density of inactive impurities in the diffused layer, but also generates numerous defects such as dislocations, slips, and precipitation clusters. Although atomic misfit stress[30] could cause some defects, it is not likely to be the main cause in the case of boron and phosphorus. The tetrahedral radii of both boron (0.88A) and phosphorus (1.1A) are less than that of silicon (1.17A). When smaller impurity atoms replace bigger host atoms, their small size might induce some tensile stress locally. However, the inter-impurity repulsion could counter and override the size related stress at high impurity concentration. The smaller impurity ions tend to reduce volume while the repulsion among them tends to increase the volume.

To relieve sudden rise of internal pressure, numerous stress-induced crystal defects are generated. At the same time, some of the excess impurity ions are squeezed out to precipitate. The strong repulsion among the impurity ions at the time of cooling must have produced enormous pressure in the heavily doped region of the wafers. From Fig. 6, the internal pressure for 10^{20} atoms/cm^3 is estimated to be about 17 atm. at 1100 °C, although the actual pressure should be lower if the effect of the photo-generated electron-hole pairs is taken into account. Also note that the pressure increases exponentially beyond this point. Czaja[31] reported the critical surface concentration for the generation of slip during phosphorus diffusion to be in the range of 2×10^{20} to 2.5×10^{20} atoms/cm^3. He carried out the diffusion at 1050 °C. A similar result was reported by McDonald *et al.*,[29] but they went a step further to prove that the defect generation is not just a function of surface concentration but also a function of diffusion depth. They were able to minimize the defect density by using a shallower diffusion. At any rate, all these findings are consistent with the implication of Fig. 6.

Levine *et al.*[33] discovered that the diffusion induced dislocation density was maximal at the layer where the impurity concentration gradient was maximum. If the entire crystal is under uniform pressure, then nothing happens. Only when there is a pressure differential, the crystal planes will be shifted or pushed. Such a pressure differential occurs at the surface and this is where the impurity concentration gradient is maximal. Therefore, it is only natural to find more dislocations at the region where the impurity concentration gradient is maximal. Similarly, numerous slip patterns are found on the surface of heavily doped wafers.

VIII. CONCLUSION

The impurity field theory provides greater insight into the hitherto unexplained phenomena of semiconductor physics. The most important concept of this theory is that the impurity ions carry their own screened Coulomb fields and interact with each other to cause various phenomena observed. The superposition of the impurity fields is proven to be identical to the Fermi level shift in the semiconductor. The screening length is a function of temperature and fluctuating carrier density. When the impurity ions are brought close enough, the ions start to repulse or attract each other depending on the impurity types. Those of opposite type attract each other either to form an ion-pair or an outright chemical bond. These pairings nullify individual screened Coulomb fields for the participating ions, because their fields become localized between two ions.

The repulsion among the same type impurity ions plays a more important role in high temperature processing technology. At a given temperature, the solid solubility of substitutional impurity which has comparable atomic size to the host crystal is determined mainly by the Coulomb repulsion among the impurity ions. The impurity ions can be introduced into the crystal as long as the repulsion is not strong enough to be rejected. In other words, because of the strong repulsion from the ions already in the crystal, any new impurity ion cannot substitutionally enter into the crystal. In a way, the solid solubility can be defined as the maximum number of impurity ions whose mutual repulsion can be countered by the binding force of the crystal lattice. The first order of approximation presented in Figs. 7 and 8 proves that, indeed, the mutual repulsion of impurity ions plays an important role in determining the final solubility.

The optically enhanced solubility of impurities reported for the rapid thermal processing (RTP) process can be explained by the shielding effect of those optically generated numerous electron-hole pairs. The thicker electron-hole

cloud shields impurity ions more effectively to enhance the solubility beyond the normal limit without the optical generation. The same explanation can be given for the shallow and high surface concentration accomplished by the low temperature RTP diffusion reported in Refs. 26 and 27.

When the heavily diffused wafers are cooled at the end of diffusion cycle, the impurity fields of each ion expand as the excess electrons and holes recombine reducing the total carrier density. At the early stage of cooling, the carrier density decreases abruptly by recombination of optically generated carriers. Then the normal recombination of thermally generated carriers follows as the cooling process continues. The sudden decrease in carrier density gives rise to a strong repulsion among the impurity ions and causes pressure build-up. When the pressure build-up becomes excessive, various crystal defects such as dislocation, slip, precipitation cluster, etc. are generated to release the pressure. However, the quick cooling would also freeze the impurity ions and vacancies in the lattice above the equilibrium values.

As pointed out, we do not currently have an empirical or theoretical equation to predict the exact carrier density at an elevated temperature which takes into account the photo-generated carriers. Also, the effect of band-gap narrowing for heavy doping[30,35,36] is not taken into account for intrinsic carrier calculation. Therefore, the calculations presented in this paper would deviate somewhat from the true values at the high temperature range where the photo-electric effect becomes appreciable.

[1] Y. D. Kwon, J. Korean Phys. Soc. **2**(2), 57 (1969).
[2] J. M. Ziman, *Principle of the Theory of Solid* (Cambridge University Press, Cambridge, 1964).
[3] C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).
[4] C. M. Wolfe, N. Holonyak Jr., and G. E. Stillman, *Physical Properties of Semiconductors* (Prentice Hall, Englewood Cliffs, NJ, 1989).
[5] V. B. Glasko and A. G. Mironov, Sov. Phys. Solid State **4**(2), 241 (1962).
[6] D. I. Thouless, Phys. Rev. **107**, 1162 (1957).
[7] V. L. Bonch-Bruevich and S. M. Kogan, Sov. Phys. Solid State **1**, 1118 (1959).
[8] W. Brandt and J. Reinheimer, Can. J. Phys. **46**, 607 (1968).
[9] P. Debye and E. Huckel, Phys. Z. **24**(9), 185 (1923).
[10] F. J. Morin and J. P. Maita, Phys. Rev. **96**, 28 (1954); **94**, 1525 (1954).
[11] C. D. Thurmond, J. Electrochem. Soc. **122**, 1133 (1975).
[12] H. D. Barber, Solid-State Electron. **10**, 1039 (1967).
[13] J. O. McCaldin, J. Appl. Phys. **34**, 1748 (1963).
[14] This section is repeated from this author's paper in Ref. 1 for completeness.
[15] Analytical proof is possible by expanding Eq. (18) into a series under the condition $N/2n_i \ll 1$.
[16] H. Reiss, C. S. Fuller, and F. J. Morin, Bell Syst. Tech. J. **35**(3), 535 (1956).
[17] F. Williams, Phys. Status Solidi **25**, 493 (1968).
[18] R. N. Hall and J. H. Racette, J. Appl. Phys. **35**, 379 (1964).
[19] F. A. Trumbore, Bell Syst. Tech. J. **39**, 205 (1960).
[20] W. Shockley and J. L. Moll, Phys. Rev. **119**, 1480 (1960).
[21] H. Ryssel, K. Muller, K. Haberger, R. Henkellmann, and F. Jahael, Appl. Phys. **22**, 35 (1980).
[22] S. Maekawa and T. Oshida, J. Phys. Soc. Jpn. **19**(3), 253 (1964).
[23] E. Guerrerro, H. Potzl, R. Grasserbauer, and G. Stingeder, J. Electrochem. Soc. **129**, 1826 (1982).
[24] J. L. Regolini, T. W. Sigmon, and J. F. Gibbons, Appl. Phys. Lett. **35**, 114 (1979).
[25] A. Lietoila, R. B. Gold, J. F. Gibbons, and T. W. Sigmon, J. Appl. Phys. **52**, 230 (1981).
[26] R. Singh, K. C. Cherukuri, L. Vedula, A. Rohatgi, and S. Narayanan, Appl. Phys. Lett. **70**, 1700 (1997).
[27] R. Singh, S. V. Nimmagadda, V. Parihar, Yuanning Chen, and K. F. Poole, IEEE Trans. Electron Devices **45**, 643 (1998).
[28] E. Tannenbaum, Solid-State Electron. **2**, 1 (1961).
[29] R. A. McDonald, G. G. Ehlenberger, and T. R. Huffman, Solid-State Electron. **9**, 807 (1966).
[30] R. B. Fair, J. Appl. Phys. **50**, 860 (1979).
[31] M. L. Joshi and S. Dash, IBM Syst. J. **11**, 446 (1966).
[32] W. Czaja, J. Appl. Phys. **37**, 3441 (1966).
[33] E. Levine, J. Washburn, and G. Thomas, J. Appl. Phys. **38**, 81 (1967); **38**, 87 (1967).
[34] M. C. Duffy, F. Barson, J. M. Fairfield, and G. H. Schwuttke, J. Electrochem. Soc. **115**, 84 (1968).
[35] G. D. Mahan, J. Appl. Phys. **51**, 2634 (1980).
[36] P. T. Lansberg, A. Neugroschel, F. A. Lindholm, and C. T. Sah, Phys. Status Solidi B **130**, 255 (1985).