Stress effects on impurity solubility in crystalline materials: A general model and density-functional calculations for dopants in silicon

Chihak Ahn,[1,2] Nick Bennett,[1] Scott T. Dunham,[2,3] and Nick E. B. Cowern[1]
[1]*School of Electrical, Electronic and Computer Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom*
[2]*Department of Physics, University of Washington, Seattle, Washington 98195, USA*
[3]*Department of Electrical Engineering, University of Washington, Seattle, Washington 98195, USA*
(Received 20 January 2009; published 27 February 2009)

We present a general theory of stress effects on the solid solubility of impurities in crystalline materials, including the effects of ionization and the Fermi level in semiconductors. Critical errors and limitations in previously proposed theory are discussed, and a rigorous accurate treatment incorporating charge-carrier-induced lattice strain and correct statistics is presented. Considering all contributing effects, we find that the strain compensation energy is the primary contribution to solubility enhancement in essentially all material systems of interest. An exception is the case of low-solubility charged impurities in semiconductors, where a Fermi-level contribution is also found. We present explicit calculations for a range of dopant impurities in Si, utilizing this system as a model example and vehicle for comparison with experiment. Our results agree closely with experimental solubilities for dopants with widely different ionic sizes.

DOI: 10.1103/PhysRevB.79.073201 PACS number(s): 61.72.sd, 64.75.Bc, 61.72.uf, 85.40.Ry

The effect of stress on impurity solubility is one of the simplest examples of a mechanical effect on the phase diagram of a material system. It influences mechanical properties such as ductility[1] and crack evolution[2] electronic properties such as carrier density in semiconductors[3] and other phenomena, such as superconductivity.[4] Consequently it has a role to play in research topics ranging from materials science, geophysics, and superconductivity to semiconductor technology.

Despite this very broad significance, our basic understanding of impurity solubility under stress is incomplete and no general physical model has yet emerged. In the case of dopant solubility in silicon, where the phenomenon has been closely studied, previously reported theoretical works have shown either inconsistencies[5,6] or important differences in formulation.[7–9] Until we understand the origin of these weaknesses it is uncertain whether solubility theory can be reliably applied to key technology problems, for example, dopant activation in nanoelectronic devices where high stress levels are introduced for band gap and mobility engineering, or stress-driven fracture[2] during the lifetime of critical engineering components. Moreover, since a small set of accurate experimental measurements has recently been reported for the case of impurities in Si,[3,10,11] it is especially worthwhile to develop an accurate comprehensive model to compare with these data and test theoretical understanding.

Early theoretical work investigated stress-dependent B solubility in Si (Refs. 5 and 6) and suggested that the main contributing effect is a stress-induced change in the Fermi level. It was concluded that Fermi-level change produces a strong enhancement of B solubility in compressively strained silicon, while a relatively small additional enhancement occurs as a result of strain compensation arising from the size mismatch between B and Si. Adey *et al.*[6] later predicted that Fermi level and strain compensation contributions are both important in the case of As in Si but are of nearly equal magnitude and opposite in sign, and thus cancel. However, these previous works are based on the use of Maxwell-Boltzmann (M-B) statistics in which the Fermi level varies with doping concentrations at the same rate near and beyond the band edges as at midgap, leading to unphysical results.

In another approach used by Diebel *et al.*[7] and Ahn and Dunham,[8] the stress energy was calculated simply via strain × stress ($\Delta\vec{\epsilon}\cdot\vec{\sigma}$), where the strain is obtained from the energy vs strain curve for the relevant dopant (Fig. 1) using neutral supercells rather than the charged supercells that are conventionally used. As we shall show, this approach is reasonable for the case of heavily doped semiconductors where the Fermi level is pinned to the band edge but is not of general validity.

Finally, Höglund *et al.*,[9] also assuming the Fermi level at the band edge, predicted dramatic stress effects on solubility in pure and multilayered alloy structures using charged supercells. However, in multilayered structures, the band offsets can build an electric potential causing dopant segregation.[12,13] Therefore, in such structures, changes in solubility are due not only to the stress effect but also to the band offsets.

Up to now, there has been little or no discussion of the relative accuracy of the different approaches that have been

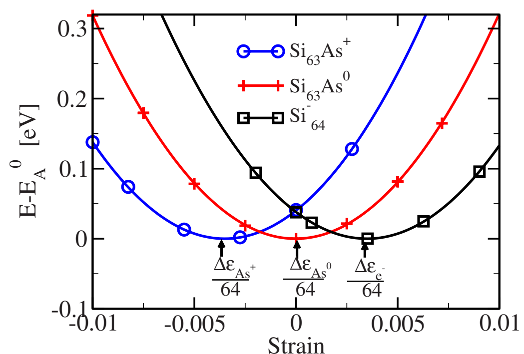

FIG. 1. (Color online) Energy vs hydrostatic strain curve for 64 atom supercell with As$^+$, As^0, and e^-. The normalized induced strain is defined as $\Delta\epsilon = 64\times(a-a_0)/a_0$, where a is the equilibrium lattice constant of the supercell with a dopant or an electron and a_0 is the Si lattice constant.

 ©2009 The American Physical Society

used, the relationships between them, and their ability or otherwise to reproduce pertinent experimental data. Furthermore, so far no general model has been presented.

In this Brief Report, we develop a general theory for the effect of stress on impurity solubility in crystalline solids. In order to describe charged or neutral impurities in semiconductors, metals, or insulators, the formulation is developed for the general case of a semiconducting band gap. We use neutral supercell methods and account for the relative position of the Fermi level and defect energy levels within Fermi-Dirac (F-D) statistics. Our methodology has much in common with established methods for modeling diffusion via the concept of "activation volume" ($V_0\Delta\vec{\epsilon}$) but is more general and in the future will also be applied to more accurate modeling of stress-dependent impurity diffusion.

At thermal equilibrium, the solid solubility of an impurity is determined by the condition minimizing the Gibbs free energy, and in the dilute limit it is given by[14]

$$C_A^{ss} = C_s \exp\left(\frac{-E_A^f}{kT}\right), \tag{1}$$

where C_s is the lattice concentration (5×10^{22} cm^{-3} for Si) and E_A^f is the impurity formation energy, frequently calculated using density-functional theory (DFT) supercell methods.[15,16] In a binary phase system the formation energy is determined by the energy change during the phase transition and is customarily given by[5,9,14]

$$E_A^f = E_{A^q}^{\mathrm{Tot}} - E_{\mathrm{ref}}^{\mathrm{Tot}} - \mu_A + q(E_F + E_{v/c}) \tag{2}$$

where $E_{A^q}^{\mathrm{Tot}}$ is the total free energy of the charged supercell with one impurity, q is the charge state, $E_{\mathrm{ref}}^{\mathrm{Tot}}$ is the total free energy of the reference supercell without an impurity, μ_A is the chemical potential of an impurity in solubility-limiting clusters/precipitates, E_F is the Fermi energy in reference to the band edge, and $E_{v/c}$ is the band edge. To predict the stress effects, $\Delta E^f(\vec{\epsilon})$ is calculated by evaluating Eq. (2) at various lattice constants, which requires extensive DFT calculations, especially under anisotropic stress conditions. In contrast, when the neutral supercell method is used within the linear elasticity limit, $\Delta E^f(\vec{\epsilon})$ can be simplified to strain$\times$stress similar to the first term in Eq. (5), which requires less computing resources than the charged supercell method and effectively removes finite-size errors, even in 64 atom supercells. This is critical for predicting stress effects since the magnitude of finite-size errors is usually comparable to the stress energy.

In neutral supercells, charge carriers occupy dopant levels (E_d) which coincide with the Fermi level at absolute zero temperature. At finite temperature, the Fermi level can deviate from the dopant level, and thus Eq. (2) can be rewritten as

$$E_A^f = E_A^{\mathrm{Tot}} - E_{\mathrm{ref}}^{\mathrm{Tot}} - \mu_A + qfE_{\mathrm{act}}, \tag{3}$$

where f is the ionized fraction of the dopant and $E_{\mathrm{act}}=E_F - E_d$ is the activation energy. In principle this method requires a correction for the underestimated band gap when the energy level is above E_v; however, when the solubility enhancement factor $C_A^{ss}(\vec{\epsilon})/C_A^{ss}(0)$ is of interest only the change in the formation energy $\Delta E_A^f(\vec{\epsilon})$ is important and the DFT band-gap problem disappears. Equation (3) is the generalized form of Eq. (2) such that it can be applied to a partially ionized system where $f<1$.

Within the linear elasticity limit, the free energy of the M-atom supercell containing a single impurity is determined from the generalized Hooke's law,[7]

$$E(\vec{\epsilon}) = E_A^0 + \frac{MV_0}{2}\left(\vec{\epsilon} - \frac{\Delta\vec{\epsilon}_A}{M}\right)\cdot \mathbf{C}\cdot\left(\vec{\epsilon} - \frac{\Delta\vec{\epsilon}_A}{M}\right), \tag{4}$$

where E_A^0 is the energy of the fully relaxed supercell, V_0 is the unit atomic volume of the lattice, $\vec{\epsilon}$ is the applied strain, $\Delta\vec{\epsilon}_A$ is the normalized induced strain due to the impurity, and $\mathbf{C}$ is the elastic stiffness tensor of the host material. The $\Delta\vec{\epsilon}_A$ is a generalized tensor representation of the volume expansion coefficient, which is conveniently applicable to all types of strain. For an isolated impurity in an isotropic material, there is only one independent element due to the symmetry. Following the usual convention, tensile applied strain and outward relaxations are defined as positive.

Applying Eq. (4) to Eq. (3), the change in the formation energy becomes

$$\Delta E_A^f(\vec{\epsilon}) = -V_0\Delta\vec{\epsilon}_A\cdot\mathbf{C}\cdot\vec{\epsilon} - \Delta\mu_A(\vec{\epsilon}) + qf\Delta E_{\mathrm{act}}(\vec{\epsilon}). \tag{5}$$

The first term is the strain compensation energy which includes all volumetric effects due to an isolated impurity, and the key factor $\Delta\vec{\epsilon}_A \approx \Delta\vec{\epsilon}_{A^q} + \Delta\vec{\epsilon}_{h/e}$ can be obtained from the lattice constant vs energy curve as shown in Fig. 1 or measured with high-resolution x-ray reciprocal-lattice mapping.[17] It is interesting to note in passing that a comparison between Eqs. (2) and (5) shows that $q\Delta E_{v/c}(\vec{\epsilon}) = -V_0\Delta\vec{\epsilon}_{h/e}\cdot\mathbf{C}\cdot\vec{\epsilon}$ where the right-hand side is the stress energy due to a charge carrier. This identity has not previously been reported and potentially gives an efficient method to calculate the deformation potential of semiconductors.

The second term in Eq. (5) $[\Delta\mu_A(\vec{\epsilon}) = -V_0\Delta\vec{\epsilon}_A^{\mathrm{cl}}\cdot\mathbf{C}\cdot\vec{\epsilon}]$ is often ignored because the induced strain due to a single atom in a large cluster/precipitate ($\Delta\vec{\epsilon}_A^{\mathrm{cl}}$) is generally very small. The induced strain due to the cluster can be obtained with the same methods applied to the isolated impurity by using the supercell with a cluster instead of an impurity. For B, the stable cluster $B_{12}I_7$, known as the building block of SiB_3,[7,18,19] produces an induced strain per clustered B atom that is at least 1 order of magnitude smaller than that of the isolated B atom, and thus effectively negligible. Likewise, the solubility-limiting cluster As_4V (Ref. 20) causes minimal lattice distortion.[8] Table I lists the induced strain due to solubility-limiting clusters for B and As. For other elements, although the structure is not known, the energetically favorable cluster is still unlikely to involve a large distortion as this would result in a large stress energy. In the exceptional case where a small impurity cluster becomes the solubility-limiting cluster and results in a significant volume change, this term should be taken into account to predict the stress effects on solubility. In addition, if the dominant cluster is charged, Fermi-level effects may need to be added to the expression for $\Delta\mu_A$ as discussed for the case of dopants below.

TABLE I. The normalized induced strain for dopants, charge carriers, dopant-defect clusters, and isovalent elements in Si. Note that the value of dopants includes both ion and electron/hole components, and cluster-induced strains are given per dopant. The values in parentheses are experimental values.

	P	As	Sb	Bi		B	Ga	In
$\Delta\epsilon$	−0.084 (−0.095[a])	0.013 (−0.02[b])	0.16	0.23		−0.30 (−0.315[c])	0.066	0.21

	e	h	As_4V	$B_{12}I_7$	C	Ge	Sn	Pb
$\Delta\epsilon$	0.22	−0.26	−0.018	−0.020	−0.42	0.05	0.21	0.26

[a]Reference 29.
[b]Reference 30.
[c]Reference 17.

The third term in Eq. (5) $[qf\Delta E_{\text{act}}(\vec{\epsilon})]$ varies depending on the type of the impurity and the position of dopant impurity energy levels. For isovalent elements or deep dopants, q or f vanishes, respectively. Thus this term does not contribute to the formation energy. For shallow dopants, or rather, dopants which are within $\sim 2kT$ of the band edge at the annealing temperature, two important extreme cases are possible.

(i) *Solubility in the intrinsic doping range*: for low-solubility dopants giving intrinsic conditions at the annealing temperature, the Fermi level is at midgap, leading to $\Delta E_{\text{act}}(\vec{\epsilon}) \sim -\Delta E_g(\vec{\epsilon})/2$. In this case $\Delta E_g(\vec{\epsilon})$ is calculated by determining thermodynamic averages of the conduction and valence band positions. This is necessary since at high temperature a carrier may occupy any one of the split subbands resulting from applied strain. Bi-doped Si is an example of this type of system, having an equilibrium solubility of $\sim 10^{17}$ cm^{-3}.[21] Likewise, In-doped Si is well described for annealing temperatures above 700 °C since the solubility limit of In, 1.8×10^{18} cm^{-3},[22] is below the intrinsic carrier concentration in this range, and the In level (considered deep at room temperature) lies within $2kT$ of the valence band edge.

(ii) *Solubility in the degenerate doping range*: in heavily doped semiconductors, the Fermi level converges to the conduction/valence band edge and becomes independent of dopant concentration.[23] At this high doping, the dopant band is merged to the conduction/valence band due to the metal-insulator (M-I) transition.[24] The M-I transition originates at doping concentrations below 10^{19} cm^{-3} for B, P, and As in Si (Ref. 25) as the dopant band touches the relevant band edge, and for example, in the case of P in Si, the dopant band has completely merged into the conduction band by 2×10^{19} cm^{-3},[26] above which concentration the doped material becomes metallic. Therefore $\Delta E_{\text{act}}(\vec{\epsilon})$ is negligible for the common shallow dopants in Si.

For dopants whose solubility lies between the above extremes, it becomes necessary to consider Fermi-level effects explicitly, requiring full numerical calculation of the effective density of states and the intrinsic carrier density as a function of strain at high temperature. Likewise, for dopants with deep levels at the annealing temperature, the solubility is complicated by partial ionization which can also vary with stress and temperature. Since neither of these cases is of direct interest for doping in semiconductor technology, we shall skip detailed discussion of them for the sake of simplicity.

When we apply the theory to commonly used shallow dopants with high solubility in Si, the solubility enhancement factor is primarily determined by the stress energy due only to an isolated dopant,

$$\frac{C^{ss}(\vec{\epsilon})}{C^{ss}(0)} = \exp\left(-\frac{\Delta E_A^f(\vec{\epsilon})}{kT}\right) \approx \exp\left(\frac{V_0 \Delta\vec{\epsilon}_A \cdot \mathbf{C} \cdot \vec{\epsilon}}{kT}\right) \quad (6)$$

Equation (6) can be used for any type of stress and predicts that the solubility of small size dopants is enhanced under compressive stress and suppressed under tensile stress and vice versa for large size dopants.

For various impurities in Si, $\Delta\vec{\epsilon}_A$ was calculated using the DFT code VASP (Refs. 15 and 16) in the generalized gradient approximation (GGA) with the PW91 functional.[27] Calculations were done using a 64 atom supercell and 2^3 Monkhorst $\vec{k}$-point sampling[28] at relatively high energy cutoff, 340 eV. We found the finite-size effects to be minimal (less than 3% change in $\Delta\vec{\epsilon}_A$) by repeating the same calculations using a 216 atom supercell for B and P. The calculated values are compared with available experimental measurements and are presented in Table I. Theoretical values for B and P agree well with experimental values[17,29] and the modest difference for As is ascribed to As-V clusters.[8,31] The temperature dependence of V_0 and $\mathbf{C}$ due to thermal expansion is ignored since they tend to compensate each other and the stress energy is nearly temperature independent.[13]

Figure 2 shows the effects of biaxial stress on the equilibrium solubility of all common dopants in Si. A large enhancement in B solubility is possible under compressive strain. For example, under −1.5% strain, B solubility at 700 °C is 1 order of magnitude higher than in conventional Si, 50% greater than the enhancement predicted by Adey *et al.*,[6] and caused entirely by atomic size, not atomic charge. The solubility of As is nearly stress independent due to the small induced strain, which agrees well with the experimental results of Sugii *et al.*[10] and Bennett *et al.*[11] Although both experiments were done under metastable conditions associated with epitaxial regrowth, we believe that the amorphous layer with high dopant concentrations acts as a stress-

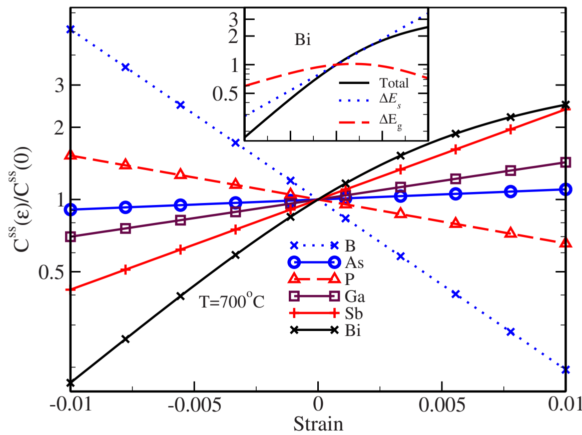

FIG. 2. (Color online) Stress effects on the solubility of shallow dopants under biaxial stress. Inset shows the components contributing to Bi solubility: the strain compensation energy ($\Delta E_s = -V_0 \Delta\vec{\epsilon} \cdot \mathbf{C} \cdot \vec{\epsilon}$) and the activation energy ($\Delta E_{\rm act} \approx -\Delta E_g$). The axis scale of the inset is the same as in the main figure.

independent reservoir. Bennett *et al.*[3] also reported a metastable Sb solubility enhancement factor of ≥ 2 at 0.7% biaxial tensile strain, in good agreement with our results. For the Bi line in Fig. 2, the term $\Delta E_{\rm act}(\vec{\epsilon})$ is nonzero and approximates to $-\Delta E_g(\vec{\epsilon})/2$ as explained in paragraph (i) above. E_g was calculated by taking the thermodynamic average of the split subbands presented in Ref. 32. To determine the averaged valence band edge E_v, a value for the ratio between the carrier-concentration effective masses of heavy and light holes was estimated from Ref. 33. Although this ratio is a function of temperature and strain, a constant value, 3, was used for the calculations since this is a reasonable approximation at high temperature[33] and any variation results in only minor changes to the final solubility. The reduced band gap tends to make Bi less soluble; however, the stress-energy factor is stronger than the band-gap factor as shown in the inset of Fig. 2.

In summary, a general theory for predicting the solubility enhancement for impurities in crystalline materials under applied strain has been developed. For semiconductors doped with high-solubility shallow dopant impurities, the strain compensation energy is the dominant factor and the strength of the stress energy is determined by the dopant-induced lattice strain, the sum of induced strain due to an ion plus a charge carrier. When the dopant impurity has low solubility, the stress-induced band-gap narrowing tends to suppress solubility. Using induced strains from DFT calculations, we have predicted changes in solubility for dopant impurities in Si, in good agreement with experimental data for dopants of widely different sizes. Key messages from this work are the importance of atom size for impurity solubility in crystalline materials, in general, and the impact this will have on the choice of high-solubility dopants in the case of strained semiconductor systems.

The authors thank Bob Jones for useful discussions. Part of this work was funded by the FP6 European projects (Grant No. 027152) ATOMICS and (Grant No. 026828) PULLNANO, and part by the Semiconductor Research Corporation (SRC) and NSF (Grant No. EIA-0101254).

[1] P. Cordier and J.-C. Doukhan, Eur. J. Mineral. **1**, 221 (1989).
[2] Y. Kim, S. Ahn, and Y. Cheong, J. Alloys Compd. **429**, 221 (2007).
[3] N. S. Bennett *et al.*, Appl. Phys. Lett. **89**, 182122 (2006).
[4] H. Su, D. O. Welch, and Winnie Wong-Ng, Phys. Rev. B **70**, 054517 (2004).
[5] B. Sadigh *et al.*, Appl. Phys. Lett. **80**, 4738 (2002).
[6] J. Adey, R. Jones, and P. R. Briddon, Phys. Status Solidi C **2**, 1953 (2005).
[7] M. Diebel, S. Chakravarthi, S. T. Dunham, and C. F. Machala, Simulation of Semiconductor Processes and Devices: SISPAD 2004 (unpublished), p. 37.
[8] C. Ahn and S. T. Dunham, Appl. Phys. Lett. **93**, 022112 (2008).
[9] A. Höglund, O. Eriksson, C. W. M. Castleton, and S. Mirbt, Phys. Rev. Lett. **100**, 105501 (2008).
[10] N. Sugii, S. Irieda, J. Morioka, and T. Inada, J. Appl. Phys. **96**, 261 (2004).
[11] N. S. Bennett *et al.*, J. Vac. Sci. Technol. B **26**, 391 (2008).
[12] P. Boguslawski, N. G. Szwacki, and J. Bernholc, Phys. Rev. Lett. **96**, 185501 (2006).
[13] C. Ahn and S. T. Dunham, Phys. Rev. B **78**, 195303 (2008).
[14] C. G. Van de Walle, D. B. Laks, G. F. Neumark, and S. T. Pantelides, Phys. Rev. B **47**, 9425 (1993).
[15] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[16] VASP the GUIDE, http://ms.mpi.univie.ac.at/vasp
[17] M. R. Sardela, Jr. *et al.*, Semicond. Sci. Technol. **9**, 1272 (1994).
[18] T. L. Aselage, J. Mater. Res. **13**, 1786 (1998).
[19] J. Yamauchi, N. Aoki, and I. Mizushima, Phys. Rev. B **55**, R10245 (1997).
[20] K. C. Pandey *et al.*, Phys. Rev. Lett. **61**, 1282 (1988).
[21] F. A. Trumbore, Bell Syst. Tech. J. **39**, 205 (1960).
[22] S. Solmi *et al.*, J. Appl. Phys. **92**, 1361 (2002).
[23] R. J. van Overstraeten, H. J. Deman, and R. P. Mertens, IEEE Trans. Electron Devices **20**, 290 (1973).
[24] N. F. Mott, Rev. Mod. Phys. **40**, 677 (1968).
[25] P. Dai, Y. Zhang, and M. P. Sarachik, Phys. Rev. B **45**, 3984 (1992).
[26] P. P. Altermatt, A. Schenk, and G. Heiser, J. Appl. Phys. **100**, 113714 (2006).
[27] J. P. Perdew *et al.*, Phys. Rev. B **46**, 6671 (1992).
[28] A. Baldereschi, Phys. Rev. B **7**, 5212 (1973); D. J. Chadi and M. L. Cohen, *ibid.* **8**, 5747 (1973); H. J. Monkhorst and J. D. Pack, *ibid.* **13**, 5188 (1976).
[29] A. Fukuhara and Y. Takano, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **33**, 137 (1977).
[30] G. S. Cargill, J. Angilello, and K. L. Kavanagh, Phys. Rev. Lett. **61**, 1748 (1988).
[31] G. Borot *et al.*, J. Appl. Phys. **102**, 103505 (2007).
[32] L. Yang *et al.*, Semicond. Sci. Technol. **19**, 1174 (2004).
[33] T. Manku and A. Nathan, J. Appl. Phys. **69**, 8414 (1991).