Diffusion and Segregation in Inhomogeneous Media and the Ge_xSi_{1-x} Heterostructure

S. M. Hu
IBM General Technology Division, East Fishkill Facility, Hopewell Junction, New York 12533
(Received 21 August 1989)

A theoretical groundwork has been laid for modeling doping diffusion and distribution in semiconductor heterostructures. The driving force for the diffusion of a dopant is no longer simply given by its concentration gradient as in homogeneous semiconductors, but by the gradient of its chemical potential that derives from several contributions to the free energy of the solid solution, including two sources of strain energy and the variation of band gap with the composition of the heterostructure. Expressions for dopant flux and segregation are given.

PACS numbers: 66.30.Dn

More than forty years ago the heterojunction transistor was conceived for performance advantages.[1-3] New developments in processing technologies in the last several years have brought a surge of industry-wide activities aimed at its realization. Much attention has been given to the important issues of pseudomorphic epitaxial strain and misfit dislocations, and band offsets at heterojunctions. An important topic that has not received any attention concerns the diffusion and segregation of dopants in the heterostructures. Here we will address this issue by setting up a theoretical basis for the understanding of such behavior. This theoretical basis is essential for the prediction and the control of dopant distribution in the fabrication of heterojunction devices. We will make special reference to the heterostructure of a germanium-silicon alloy, Ge_xSi_{1-x}, grown epitaxially on a silicon substrate, a heterostructure that accommodates the desire, for compelling reasons, to stay with the silicon technology. As the analysis is quite basic and the results physically intuitive, readers with interests more general than the heterojunction may also benefit from reading on.

The segregation and the driving force for the diffusion of a dopant in a heterostructure are determined, respectively, by its electrochemical potential (hereafter simply called "chemical potential") and by the gradient of this chemical potential. Let N be the number of atoms per unit volume of the solid solution. The chemical potential of the dopant, solute species 1, is given by the partial derivative of the free energy of the solution with respect to N_1. In a homogeneous semiconductor, the chemical potential of a dopant derives mainly from two contributions to the free energy of the solid solution: the usual configurational entropy, and the electronic energy through the ionization of the dopant atoms; it is given by[4]

$$\mu_1 = kT\ln(N_1/N_L) - Z_1 kT\ln[(1+\zeta)/(1+\zeta^*)] - kT\ln(1-N_1/N_L)\,, \quad (1)$$

where N_L is the concentration of lattice sites. The last term in Eq. (1) is usually negligible. Neglected from Eq. (1) is the effect of dopant-atom clustering,[5,6] which becomes appreciable at concentrations exceeding a few 10^{20} atoms/cm^3.[5] The entropy from the effect of solute atoms on lattice vibrations[7] has also been neglected. In the second term, due to dopant ionization, Z_1 is the charge state for the dopant ion, and ζ is given by

$$\zeta = g_1 \exp[(E_1 - E_F)/kT]\,, \quad (2)$$

where g_1 and E_1 are, respectively, the degeneracy factor and the energy level of the dopant electron state, and E_F is the Fermi level. The asterisk denotes the intrinsic condition. For simplicity, we will consider here systems in the nondegenerate condition, which is appropriate for the base region at diffusion temperatures. Then the second term on the right-hand side of Eq. (1) is approximately given by $Z_1(E_F - E_i)$, where E_i is the intrinsic Fermi level; it may then be considered the electrostatic part of the chemical potential of the fully ionized dopant ion. For $N_1 \gg n_i$, where n_i is the intrinsic carrier concentration, we have $E_F - E_i \cong Z_1 kT\ln(N_1/n_i)$, and Eq. (1) reduces to

$$\mu_1 = kT\ln(N_1/N_L) + kT\ln(N_1/n_i)\,. \quad (3)$$

When there is more than one dopant present, the second term on the right-hand side of Eq. (3) should, in the local charge-neutrality approximation, be replaced with $Z_1\,\mathrm{sgn}(Z_nN_n)kT\ln f$, where f is given by

$$f = \frac{|Z_nN_n|}{2n_i}\left\{1 + \left[1 + \left(\frac{2n_i}{Z_nN_n}\right)^2\right]^{1/2}\right\}, \quad (4)$$

where Z_n is the charge state of the nth dopant ion, and $\mathrm{sgn}(y)$ denotes the sign of argument y. Expression (4) also takes care of the situation in which $|Z_nN_n| \ll n_i$, such as in the neighborhood of a pn junction.

In a heterostructure containing a layer of, for example, Ge_xSi_{1-x}, there are two complications. The first is that the band gap[8] and the effective densities of states (or the carrier effective masses) in the conduction and the valence bands vary with the alloy composition x, and hence with location. This affects n_i in Eq. (3). Furthermore, the Fermi level is usually evaluated in reference to some other energy level, such as the valence-band edge, of the same material. In a heterostructure, a material-

© 1989 The American Physical Society

independent reference, such as the vacuum level, must be used. A proper reference is the ideal work function which is defined as the energy difference between the vacuum level and the Fermi level in an uncharged solid. In practice the work function is usually defined operationally as the work required to remove an electron from a solid in vacuum to a remote point; it is thus affected by the charge distribution at the solid surface. We shall assume that the ideal work function is available, and define an intrinsic work function W_i for a material in which the Fermi level is E_i. The electron chemical potentials in two isolated intrinsic materials A and B differ by an amount equal to $W_i(A)-W_i(B)$. When they come into contact, some electrons will change hand, until an electrostatic potential difference between them builds up to $\phi=[W_i(B)-W_i(A)]/q$, thereby equilibrating the electron chemical potentials between them. Doping adds another contribution to the electrostatic potential; this is already accounted for by the second term on the right-hand side of Eq. (1). To be general, we include a term $\Delta E_{b1}(N_2)$ representing the difference between the bond energy of the unionized dopant atom in the local phase and that in a reference phase (in which $N_2=0$, e.g.). With these considerations, Eq. (3) is modified to

$$\mu_1=kT\ln\left(\frac{N_1}{N_L}\right)-Z_1\Delta W_i(N_2)-\Delta E_{b1}(N_2)+kT\ln N_1 + \tfrac{1}{2}E_g(N_2)-\tfrac{1}{2}kT\ln[N_C(N_2)N_V(N_2)]\,. \qquad (5)$$

In Eq. (5), only E_{b1} is dopant specific. W_i, E_{b1}, the band gap E_g, the conduction- and valence-band densities of states N_C and N_V are all functions of the germanium concentration N_2 (or the alloy composition x). They will also be affected by the epitaxial strain; the strain effect will be considered separately. The work-function difference can be taken between the local phase and the phase with $N_2=0$. The work functions for silicon and germanium, measured with an accuracy no better than ± 0.2 eV, appear not significantly different. One would expect, from the trend of work functions of group-IV elements, that the germanium work function is slightly smaller, perhaps by some 0.2 eV. For our purpose, the work function may be assumed to vary linearly with the alloy composition. This would track the reported[9,10] trend of valence-band offset in the Ge_xSi_{1-x} alloy. This term affects donors and acceptors in opposite directions. No data are available on bond energy versus alloy composition. In general, we would expect the bond energy to be smaller in a germanium-rich phase. Hence, the bond-energy term, in contrast to the band-gap term, tends to drive both donors and acceptors away from a germanium-rich phase. The sum of the last two terms in Eq. (5) is just $-kT\ln(n_i)$, expressing $n_i(N_2)$ in more easily measured parameters. E_g decreases rather linearly with x between $x=0$ and 0.3, with $\partial E_g/\partial N_2=-(0.45\text{ eV})/N_L$ $(N_2\simeq xN_L)$.[8] E_g also decreases with temperature; the decrease is only slightly dependent on x between 0 and 0.3. In a system having more than one dopant, Eq. (5) can be modified by replacing the last three terms with $Z_1\,\mathrm{sgn}(Z_nN_n)kT\ln f$, and relating $n_i(N_2)$ through the last two terms of Eq. (5). When $|Z_nN_n|\gg n_i$, we need only to replace the fourth term on the right-hand side of Eq. (5) with $Z_n\,\mathrm{sgn}(Z_nN_n)kT\ln|Z_nN_n|$.

The second complication is the strain energy of the solid solution. Two kinds of strain exist in the pseudomorphic epitaxial films: the microscopic internal strain and the macroscopic external strain. The internal strain arises from the replacement of some host atoms with solute atoms of a different size. A number of seminal elasticity-theoretic analyses[11-13] are applicable for the continuum approximation of the problem: A solute atom, idealized as a spherical inclusion of a radius $r_0(1+\delta)$, squeezes into a spherical cavity vacated by a host atom of radius r_0, and the two settle on a common radius $r_0(1+\delta')$. The strain energy associated with an inclusion consists of two parts: the uniform deformation of the inclusion and the nonuniform deformation of the matrix surrounding it. Both parts can be easily calculated.[13] The difficulty arises, when the density of solute atoms is very high, that the strain energy due to each solute atom is no longer additive. This is because the strain originating at individual solute atoms is linearly additive, while the local strain energy is quadratic with the local strain. The local strain energy in a volume element dv is given by

$$dW_L=\tfrac{1}{2}c_{ij}\epsilon_i\epsilon_j dv\,, \qquad (6)$$

where ϵ_i are the strain-tensor components in the one-suffix matrix notation, c_{ij} are the stiffness constants, and the right-hand side represents a sum over repeated subscripts. Equation (6) must be integrated over the entire volume. If the matrix and the inclusion share the same elastic constants, the displacement at position $\mathbf{R}$ exterior to the inclusion at $R=0$ is given by[14]

$$\mathbf{u}(\mathbf{R})=\left[\delta r_0^3\frac{1+\nu}{3(1-\nu)}\right]\frac{\mathbf{R}}{R^3}\,, \qquad (7)$$

where ν is Poisson's ratio (0.27 for silicon), and the bracketed term is the "strength" of the elastic potential. An appropriate image field needs to be added to take care of the free boundary condition.[12,13] The strain field can easily be obtained from Eq. (7). However, in a solid solution containing many solute atoms, there exists the problem of integrating over a domain riddled with singularities.[12] A second, but less serious, problem is that the continuum theory of elasticity will surely break down in the immediate neighborhood of the atomic-sized inclusion. Thus, from Eq. (7), the radial displacement of the nearest host atom, centered at $R=(2+\delta)r_0$, is about $0.145\delta r_0$. This is substantially less than $0.41\delta r_0$ calculated from a first-principles wave-function approach.[15] The discrepancy is anticipated since the elastic constants of all materials rise with compression.[16] But there is more to the origin of this discrepancy. It is merely an

expediency to name the solute atom as the inclusion. We could just as well consider the inclusion to be, for example, a cluster comprising the solute atom and its four nearest-neighbor solvent atoms, or just the halves of these nearest-neighbor atoms. In this latter choice, we would have an inclusion of radius $(1+\delta/2)2r_0$, forced into a cavity of radius $2r_0$. The displacement of the nearest neighbor is then $0.58\delta r_0$, a factor of 4 increase. The strain energy is also affected by this conceptualization of the inclusion, but by a factor of 2.

The problem simplifies at high solute densities, whence the strain fields around individual solute atoms overlap substantially, giving rise to a substantially uniform change of the lattice constant,[12,13,17-19] a phenomenon known empirically as Vegard's law.[20] Since the deformation is uniform in all directions, $\epsilon_1=\epsilon_2=\epsilon_3=\text{const}$, and the nondiagonal tensor components vanish. The work done by the solute atoms in expanding the lattice is then easily obtained from Eq. (6): $W_L=\frac{3}{2}(c_{11}+2c_{12})\epsilon_1^2$ for cubic crystals, or $\frac{9}{2}K\epsilon_1^2$ for isotropic media (K is the bulk modulus, about 1.1×10^{12} dynes/cm^2 for silicon). A simple extension of Vegard's law gives ϵ_1 as $\beta_n N_n$, where β is the lattice contraction (negative) or expansion (positive) coefficient, the subscript n denotes the solute species (say, 1 for the dopant and 2 for germanium), and Einstein's summation convention again applies. Some experimental data on β for a number of relevant solutes in silicon are as follows: for boron, -4.5×10^{-24} cm^3/atom;[21] for phosphorus, -1.8×10^{-24} cm^3/atom;[21] and for germanium, $+0.71\times10^{-24}$ cm^3/atom[22] (see also Ref. 23). It has been reported[24] that doping with arsenic causes a contraction of the silicon lattice, contrary to the prediction based on covalent bond length; but the magnitude is small and can be neglected.

A heterostructure, for example, a Ge_xSi_{1-x} layer grown epitaxially on a silicon substrate, additionally contracts an externally induced strain. The Ge_xSi_{1-x} layer, below a critical thickness, is forced to adopt the lattice constant of the silicon substrate as its lateral lattice constant, while the vertical direction is allowed to expand freely. If all three dimensions of the epitaxial layer are constrained, the strain energy of uniform expansion as derived above would vanish, and microscopic internal strain would take its place, resulting in a higher internal energy. The microscopic strain energy is difficult to calculate. However, since elastic deformation, by definition, is a nondissipative, completely reversible process, we can construct a sequence of process steps to arrive at the final state, and sum up the energy expenditures of all steps involved. The energy of uniformly deforming the lattice by the solute atoms has already been given in the preceding section. The solid solution is next compressed along the two lateral dimensions until these match that of an imagined substrate. This produces an external lateral strain of $-\beta_n N_n$, and an external vertical strain of $-(2c_{12}/c_{11})\beta_n N_n$, or $(\beta_n N_n)2\nu/(1-\nu)$. The energy associated with this external straining is $(c_{11}+2c_{12})\times(1-c_{12}/c_{11})(\beta_n N_n)^2$, or $[3K(1-2\nu)/(1-\nu)]\times(\beta_n N_n)^2$. Combining the strain energies from the internal and the external origins, one obtains the strain contribution to μ_1 as

$$(\mu_1)_{\text{str}}=\beta_1(c_{11}+2c_{12})(5-2c_{12}/c_{11})\beta_n N_n\,, \tag{8}$$

or $\beta_1 3K[(5-7\nu)/(1-\nu)]\beta_n N_n$ for isotropic media. In deriving Eq. (8), we assumed $\partial c_{ij}/\partial N_1$ to be negligible.

An important question must now be asked and answered: Are there any other contributions from the lattice strain to the free energy of the solid solution? The first to come to mind is the deformation potential which affects the chemical potential of electrons, and which has been proposed to vary linearly with strain.[25] Indeed, significant band-gap narrowing, additional to that caused by the alloy composition, has been reported for epitaxially strained Ge_xSi_{1-x}.[26-28] We shall argue below that the correct answer is no.

As are all causal relationships of physical laws, the free energy of a defined part of the lattice as a function of relative atomic displacements must be analytic within the domain in which such a functional relationship exists. Therefore, we can express the free energy of the part of the lattice of interest in a Taylor series in terms of strain. Let $F(\epsilon_i)$ be the free energy of the defined part of lattice (for given N_1 and N_2). We can write

$$F(\epsilon_1,\epsilon_2,\ldots)=F(0)+\left(\frac{\partial F}{\partial\epsilon_i}\right)\epsilon_i+\frac{1}{2}\left(\frac{\partial^2 F}{\partial\epsilon_i\,\partial\epsilon_j}\right)\epsilon_i\epsilon_j+\cdots\,, \tag{9}$$

where all derivatives are evaluated at $\epsilon_i=0$ ($i=1$ to 6). For a given crystalline structure, $F(\epsilon_i)$ is minimum at $\epsilon_i=0$; otherwise the atoms would spontaneously rearrange themselves. Therefore, all the first derivatives of F with respect to ϵ_i must be zero, and all the second derivatives must be positive. All terms higher than the second order drop out for sufficiently small ϵ_i. One can identify the second derivatives in Eq. (9) as the elastic stiffness constants c_{ij}.[29] The sum of the second-order terms is hence just the strain energy, and it is inclusive of all kinds of energy affected by strain.

We can write the chemical potential of species 1 as

$$\mu_1=kT\ln\left(\frac{N_1^2}{N_L(N_C N_V)^{1/2}}\right)-Z_1\Delta W_i(N_2)-\Delta E_{b1}(N_2)+\tfrac{1}{2}E_g(N_2)+\theta\beta_1\beta_2 N_2+\theta\beta_1^2 N_1\,, \tag{10}$$

where $\theta=9K$ for free standing Ge_xSi_{1-x} films, and $\theta=3K(5-7\nu)/(1-\nu)$ for epitaxially strained films. To find the segregation coefficient of dopant 1 between the alloy film and the silicon substrate, we equate μ_1 between the two phases. The last term in Eq. (10) introduces a concentration dependence of the segregation coefficient. For $N_1<1\times10^{20}$ cm^{-3}, this affects the segregation by a

factor $<\exp[(-0.009\text{ eV})/kT]$ for boron, and even less for phosphorus. Therefore, provided $|\beta_2 N_2| \gg |\beta_1 N_1|$, as is appropriate for the system under consideration, the last term in Eq. (10) can be dropped. Then we obtain the segregation coefficient k_{seg} and the diffusion flux J_1 as

$$k_{seg} = \left[\frac{N_C(N_2)N_V(N_2)}{N_C(0)N_V(0)}\right]^{1/4} \exp\left[\frac{2Z_1\Delta W_i(N_2)+2\Delta E_{b1}(N_2)+E_g(0)-E_g(N_2)-2\theta\beta_1\beta_2 N_2}{4kT}\right], \tag{11}$$

$$J_1 = -\left(\frac{D_1 N_1}{kT}\right)\frac{\partial \mu_1}{\partial x}$$
$$= -2D_1\frac{\partial N_1}{\partial x} + \frac{D_1 N_1}{2}\left[\frac{1}{N_C}\frac{\partial N_C}{\partial N_2} + \frac{1}{N_V}\frac{\partial N_V}{\partial N_2} + \frac{2}{kT}\left(Z_1\frac{\partial W_i}{\partial N_2} + \frac{\partial E_{b1}}{\partial N_2} - \frac{1}{2}\frac{\partial E_g}{\partial N_2} - \theta\beta_1\beta_2\right)\right]\frac{\partial N_2}{\partial x}, \tag{12}$$

where D_1 is the diffusivity of the dopant. The first term on the right-hand side of Eq. (12) is the familiar diffusion term for a single dopant, with the factor 2 coming from the internal field (for $N_1 \gg n_i$).[30] In a system containing more than one dopant, one can write in place of Eq. (12)

$$J_1 = -D_1\frac{\partial N_1}{\partial x} - \frac{D_1 N_1}{f}\left[\sum_{n\neq 2}\frac{\partial f}{\partial N_n}\frac{\partial N_n}{\partial x} + \frac{\partial f}{\partial n_i}\left(\frac{\partial n_i}{\partial N_C}\frac{\partial N_C}{\partial N_2} + \frac{\partial n_i}{\partial N_V}\frac{\partial N_V}{\partial N_2} + \frac{\partial n_i}{\partial E_g}\frac{\partial E_g}{\partial N_2}\right)\frac{\partial N_2}{\partial x}\right]$$
$$+ \frac{D_1 N_1}{kT}\left(Z_1\frac{\partial W_i}{\partial N_2} + \frac{\partial E_{b1}}{\partial N_2} - \theta\beta_1\beta_2\right)\frac{\partial N_2}{\partial x}. \tag{13}$$

The use of Eq. (13) [or Eq. (12)] for the calculation of diffusion profiles will automatically take care of the effect of dopant segregation in heterostructures.

In summary, we have analyzed the problem of dopant segregation and diffusion driving force in inhomogeneous media, in general, and in Ge_xSi_{1-x} heterostructres, in particular. Expressions have been derived which can be used for modeling dopant diffusion in such media, an urgent need in the development of heterostructure devices. Several factors affect the chemical potential of a dopant in a heterostructure. Among these only the band-gap change and the lattice strain versus alloy composition have been measured with sufficient accuracy. The remaining parameters will have to be deduced from modeling and diffusion data. It should be noted that important issues concerning diffusion processes in heterostructures also include the understanding of how D_1 is affected by the alloy composition. We leave this topic for a future study.

I thank Dr. Sherry Gillespie for her enlightened support.

[1]W. Shockley, U.S. Patent No. 2569346 (25 September 1951; filed 26 June 1948).

[2]H. Kroemer, Proc. IRE **45**, 1535 (1957).

[3]H. Kroemer, Proc. IEEE **70**, 13 (1982).

[4]S. M. Hu, Phys. Rev. **180**, 773 (1969).

[5]S. M. Hu, in *Atomic Diffusion in Semiconductors,* edited by D. Shaw (Plenum, London, 1973), p. 217ff.

[6]A. Erbil, W. Weber, G. S. Cargill, III, and R. F. Boehme, Phys. Rev. B **34**, 1392 (1986).

[7]P. Cappalletti, G. F. Cerofonlini, and G. U. Pignatel, Philos. Mag. **46**, 863 (1982).

[8]R. Braunstein, A. R. Moore, and F. Herman, Phys. Rev. **109**, 695 (1958).

[9]A. Munoz, L. Brey, and C. Tejedor, Solid State Commun. **67**, 445 (1988).

[10]J. F. Gibbons, C. A. King, J. L. Hoyt, D. B. Noble, C. M. Gronet, M. P. Scott, S. J. Rosner, G. Reid, S. Laderman, K. Nauka, J. Turner, and T. I. Kamins, Tech. Dig. Int. Electron Devices Meet. **88**, 566 (1988).

[11]N. F. Mott and F. R. N. Nabarro, Proc. Roy. Soc. London **52**, 86 (1940).

[12]J. D. Eshelby, J. Appl. Phys. **25**, 255 (1954).

[13]J. D. Eshelby, in *Solid State Physics,* edited by F. Seitz and D. Turnbull (Academic, New York, 1956), p. 79ff.

[14]J. Goodier, Philos. Mag. **7**, 1017 (1937).

[15]A. Baldereschi and J. J. Hopfield, Phys. Rev. Lett. **28**, 171 (1972).

[16]F. Seitz, Rev. Mod. Phys. **18**, 384 (1946).

[17]K. Huang, Proc. Roy. Soc. London A **190**, 102 (1947).

[18]B. J. Pines, J. Phys. (U.S.S.R.) **3**, 308 (1940).

[19]J. Friedel, Philos. Mag. **46**, 514 (1955).

[20]L. Vegard, Z. Phys. **5**, 17 (1921).

[21]G. Celotti, D. Nobili, and P. Ostoja, J. Mater. Sci. **9**, 821 (1974).

[22]J. P. Dismukes, L. Ekstrom, and R. J. Paff, J. Phys. Chem. **68**, 3021 (1964).

[23]M. Rubenstein, J. Cryst. Growth **17**, 149 (1972).

[24]G. S. Cargill, III, J. Angilello, and K. L. Kavanagh, Phys. Rev. Lett. **61**, 1748 (1988).

[25]J. Bardeen and W. Shockley, Phys. Rev. **80**, 72 (1950).

[26]R. People, Phys. Rev. B **32**, 1405 (1985).

[27]D. V. Lang, R. People, J. C. Bean, and A. M. Sergent, Appl. Phys. Lett. **47**, 1333 (1985).

[28]J. C. Sturm, E. J. Prinz, P. M. Garone, and P. V. Schwartz, Appl. Phys. Lett. **54**, 2707 (1989).

[29]For example, see F. Seitz, *Modern Theory of Solids* (McGraw-Hill, New York, 1940), p. 95.

[30]S. M. Hu, in *VLSI Science and Technology 1985,* edited by W. M. Bullis and S. Broydo (Electrochemical Society, Pennington, NJ, 1985), p. 465ff.