Automatica 44 (2008) 2241–2247

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Maximum a posteriori estimation of activation energies that control silicon self-diffusion☆

Charlotte T.M. Kwok, Kapil Dev, Edmund G. Seebauer, Richard D. Braatz *

University of Illinois at Urbana-Champaign, Urbana, IL 61801-3602, USA

ARTICLE INFO

Article history:
Received 25 September 2006
Received in revised form
21 June 2007
Accepted 4 January 2008
Available online 23 May 2008

Keywords:
Bayesian estimation
Parameter estimation
Semiconductor processes
Rapid thermal processing

ABSTRACT

Self-diffusion in crystalline silicon is controlled by a network of elementary steps whose activation energies are important to know in a variety of applications in microelectronic fabrication. The present work employs maximum a posteriori (MAP) estimation to improve existing values for these activation energies, based on self-diffusion data collected at different values of the loss rates for interstitial atoms to the surface. Parameter sensitivity analysis shows that for high surface loss fluxes, the energy for exchange between an interstitial and the lattice plays the leading role in determining the shape of diffusion profiles. At low surface loss fluxes, the dissociation energy of large-atom clusters plays a more important role. Subsequent MAP analysis provides significantly improved values for these parameters.

1. Introduction

The mechanism of diffusion in silicon has been extensively investigated over the past few decades (Fahey, Griffin, & Plummer, 1989). One of the major driving forces comes from the technological need to reduce the transient enhanced diffusion (TED) (Jain et al., 2002; Shao, Liu, Chen, & Chu, 2003) of implanted dopants during annealing while also increasing dopant activation — a dual problem that poses a major impediment to the continuous miniaturization of semiconductor devices. Although it is generally accepted that point defects such as interstitials and vacancies serve as primary mediators of diffusion in silicon (Bracht, Haller, & Clark-Phelps, 1998), these defects are difficult to monitor directly owing to their low concentrations, making the study of specific diffusion mechanisms very challenging. To circumvent this problem, models have been employed together with experimental diffusion data to estimate the activation energies for the diffusion and reaction of point defects and defect clusters in silicon (Bracht, Stolwijk, & Mehrer, 1995; Chakravarthi & Dunham, 2001; Cowern, Janssen, van de Walle, & Gravesteijn, 1990; Stolk et al., 1997). Some of the key parameters have emerged from silicon self-diffusion data.

Progress toward the solution of the technological problem depends heavily on the development of suitable diffusion-reaction models for point defects in silicon. In the most data-efficient implementations, model identification should be an iterative process involving the steps of optimal experimental design, experimental data collection, parameter estimation, and hypothesis mechanism selection. These steps need to be repeated until sufficiently accurate parameter estimates are obtained (Braatz et al., 2006; Emery & Nenarokomov, 1998; Walter & Pronzato, 1990). However, even with the application of optimal experimental design, the accuracy of the parameter estimates may still be limited unacceptably by the physicochemical behavior of the process and various practical constraints in a particular experimental configuration. In such cases, significantly improved estimates can often be obtained by studying additional experimental configurations.

In past work, this laboratory has developed a model to describe the transient enhanced diffusion (TED) of boron in silicon (Gunawan, Jung, Seebauer, & Braatz, 2003; Gunawan, Jung, Braatz, & Seebauer, 2003; Jung, Gunawan, Braatz, & Seebauer, 2004a). Values of the model parameters were determined by parametric sensitivity analysis, maximum likelihood (ML) estimation, and maximum a posteriori (MAP) estimation. Although the model provided good fits to boron diffusion and activation data that existed at the time (Murto, 1999), the predictions for diffusion profiles annealed at slow ramp rates left room for significant improvement (Jung, Gunawan, Braatz, & Seebauer, 2003).

In other prior work (Kwok, Dev, Braatz, & Seebauer, 2005; Seebauer et al., 2006), we studied the effect of surface adsorption state on the surface generation and annihilation rates of self-interstitials, using self-diffusion as a marker. Both the generation and annihilation rates could be modulated though adsorption of small quantities of atomic nitrogen. Those experiments provided

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Edward Gatzke under the direction of Editor Frank Allgöwer.

* Corresponding author. Tel: +1 217 333 5073; fax: +1 217 333 5052.
E-mail address: braatz@uiuc.edu (R.D. Braatz).

0005-1098/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2008.01.020

a significantly expanded data set for examining the activation energies of the rate processes that control self-diffusion. This paper uses those data to refine the relevant model parameters through parameter sensitivity analysis and MAP estimation.

Parameter sensitivity coefficients quantify the variations in the model output in response to perturbations in its parameters. The relative importance of various parameters in the model indicates which parameters can be estimated most accurately and should be included in the parameter estimation algorithm. MAP estimation determines the most likely values of model parameters when prior information is available. It utilizes Bayes theorem to determine the refined probability density function of the parameters by combining prior statistical information and additional experimental data. By utilizing MAP estimation, the prior information from previously studied systems were augmented to new experimental data to obtain refined estimates of various parameters.

2. Experimental methods

The new data employed here have already been described elsewhere (Seebauer et al., 2006), but a brief description of experimental methods is given here for convenience. Diffusion data were collected using an isotopically-labeled ^{30}Si tracer implanted into an epitaxial matrix depleted in the ^{30}Si isotope. Experiments were performed in a turbomolecularly pumped ultrahigh vacuum chamber with a base pressure in the low 10^{-9} torr range. The Si-isotope heterostructure consisted of a 0.5 μm thick slightly p-type ([B] = 10^{15} cm^{-3}) epitaxial Si depleted in ^{30}Si to 0.002% (compared to the natural abundance of 3.10%) grown on natural silicon, which was also doped with boron, but to a level of 10^{19} cm^{-3}. SiH_4 enriched in the ^{30}Si isotope to 90% was implanted into the Si matrix. Nitrogen adsorption was accomplished by exposure to ammonia (99.99%) at 800 °C. The degree of nitrogen coverage was measured by Auger Electron Spectroscopy (AES) and was controlled by exposure time. The as-implanted and annealed profiles were measured *ex situ* by secondary ion mass spectroscopy (SIMS).

Three specimens were prepared, with nitrogen coverages ranging between 0 and 1.2 monolayer. Annealing was accomplished by resistive heating at one of two conditions: 1000 °C for 120 min or 980°C for 90 min.

3. Silicon Self-diffusion Model

The model employed in this work describes the diffusion of the implanted ^{30}Si tracer in an epitaxial Si matrix depleted in the ^{30}Si isotope (Kwok et al., 2005). The model utilizes continuum equations to describe the reaction and diffusion of interstitial atoms and related defects in silicon. These equations have the general form for species i:

$$\frac{\partial C_i}{\partial t} = -\frac{\partial J_i}{\partial x} + G_i, \tag{1}$$

where C_i, J_i, and G_i denote the concentration, flux, and net generation rate of species i, respectively. The net generation G_i incorporates terms associated with cluster formation and exchange between interstitials and the lattice. The reactions for the clustering of interstitials in the model are

$$\mathrm{I}_{m-1} + \mathrm{I} \leftrightarrow \mathrm{I}_m, \quad 2 \le m \le 5, \tag{2}$$

where the index m denotes the number of atoms in the cluster. The exchange between interstitial silicon and the lattice obeys the reaction

$$^{30}\mathrm{Si}_i + {}^{28}\mathrm{Si}_l \leftrightarrow {}^{30}\mathrm{Si}_l + {}^{28}\mathrm{Si}_i. \tag{3}$$

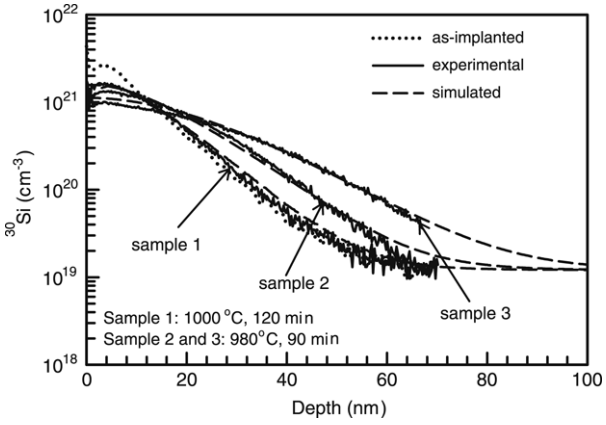

Fig. 1. Experimental and simulated ^{30}Si profiles using the surface loss probability S as the only adjustable parameter determined by maximum likelihood (ML) estimation.

Including the Poisson equation, the model consists of 21 highly stiff coupled partial differential equations (PDEs), with the maximum cluster size restricted to five due to limitations of the FLOOPS simulator (Mark Law, 2005). The five-atom dissociation energy is equated to that for very large defects, with the advantages and limitations of this approximation discussed elsewhere (Gunawan, Jung, Braatz, et al., 2003).

The effect of adsorbate coverage on the surface annihilation rate of interstitials is quantified by the surface loss probability S, which is incorporated in the boundary condition as

$$-D\left.\frac{\partial C}{\partial x}\right|_{x=0} = J_{\text{total}} S, \tag{4}$$

where J_{total} denotes the total impinging flux of interstitials. The actual flux at the surface is the product of the total impinging flux and the surface loss probability.

The initial condition for each simulation run was an experimental as-implanted profile, with the assumption that 20% of the implanted ^{30}Si enters as substitutional sites. The initial condition for the interstitial ^{28}Si was set based on the "+1 model" (Giles, 1991). Values of surface loss probability S at different degrees of nitrogen coverage were determined by maximum likelihood (ML) estimation with S as the only adjustable parameter. Fig. 1 shows the experimental as-implanted and annealed profiles as well as the simulation fit for three sets of conditions. The sample with highest nitrogen coverage (sample 3) shows the most profile spreading, while the clean surface sample (sample 1) shows the least. The ML estimates of the surface loss probability S and the values of key activation energies are summarized in Table 1. The more diffused profile corresponds to a smaller surface loss probability S.

4. Description of analytical methods

4.1. Parameter sensitivity analysis

Parameter sensitivity analysis quantifies the influence of perturbations in model parameters on the model outputs, and has been widely applied in the analysis and design of chemical systems (Leis, Gallagher, & Kramer, 1987; Van Voorhees & Bahill, 1995). The analysis determines which model parameters need to be estimated or calculated most accurately for the model to have maximum predictive value, and which parameters can be largely ignored. Fortunately, the parameters having high sensitivities to measurements can also be determined most accurately by MAP techniques.

This laboratory has applied sensitivity analysis to the activation energies of the elementary kinetic steps governing in the TED of

Table 1
Parameter estimates with maximum likelihood estimation of the surface loss probability S

Parameters[a]	Value	Standard deviation	95% confidence interval half-width	Method	Reference[b]
E_2 (cluster dissociation–size 2)	1.4 eV	0.03	0.06	ML	G03a
E_3 (cluster dissociation–size 3)	2.192 eV	0.012	0.023	MAP	G03a
E_4 (cluster dissociation–size 4)	3.055 eV	0.002	0.004	MAP	G03a
E_{large} (cluster dissociation–large)	3.7 eV	0.1	0.2	ML	G03a
E_{diff} (Si interstitial diffusivity)	0.72 eV	0.03	0.06	ML	G03a
E_{ex} (exchange between interstitial and substitutional Si)	1.02 eV	0.30	0.59	ML	S06[c]
S_1 (surface loss probability, sample 1)	4.15×10^{-2}	2.6×10^{-3}	5.2×10^{-3}	ML	K05
S_2 (surface loss probability, sample 2)	7.11×10^{-4}	7.4×10^{-6}	1.4×10^{-5}	ML	K05
S_3 (surface loss probability, sample 3)	2.36×10^{-4}	1.4×10^{-6}	2.7×10^{-6}	ML	K05

[a] Past estimates of activation energies relied on density functional theory calculations and data from isolated experiments reported in the literature, as well as experimental profile data collected by the authors and their collaborators.
[b] G03a =Gunawan, Jung, Seebauer, et al. (2003); S06 =Seebauer et al. (2006); K05 =Kwok et al. (2005).
[c] Value not stated explicitly in the reference.

boron in silicon (Gunawan, Jung, Braatz, et al., 2003). As the silicon self-diffusion system shares similar elementary reactions with the boron model (e.g., interstitial diffusion and cluster dynamics), some of the activation energies in the silicon self-diffusion model are the same as for the boron system (see Table 1). However, the sensitivity coefficients can be functions of the particular system and the corresponding experimental conditions. Hence, parameters that exhibit relatively small sensitivity coefficients in one system may be important in another. In such cases, the accuracy of the aggregated parameter estimates can be improved by combining the information from multiple experimental systems.

Both the boron TED system and the silicon self-diffusion system entail the diffusion of species under supersaturation of defects. However, the two systems are different in several ways. One major difference lies on the boundary condition. In the boron system, the silicon surface was covered with screen oxide during experiment, giving a boundary condition reasonably close to a perfect reflector (i.e., $S = 0$) (Jung, Gunawan, Braatz, & Seebauer, 2004b). For the self-diffusion model, the surface dangling bonds were partially saturated with adsorbate to give a boundary condition lying between the two extremes of perfect sink and perfect reflector.

The annealing program employed was different, too. The boron experiment utilized a "spike anneal" program. The sample was heated up and cooled down rapidly such that it was kept at high temperature (> 600°C) for short time (~20 s). "Soak annealing" was used in the self-diffusion experiment, which kept the sample at high temperature for much longer time (~90 min).

The differences between the two systems may result in large differences in the sensitivities of the model outputs to the same model parameter. For example, the soak anneal program employed in the self-diffusion model would be expected to improve the sensitivities of model outputs to parameters involved in the reaction and diffusion that have larger time constants, while the difference in the boundary condition may result in a change of the dominating reaction/diffusion mechanism in the system.

The matrix F of sensitivity coefficients includes the partial derivatives of the variables β_k with respect to the dependent variables P_j (Tomović & Vukobratović, 1972):

$$F_{j,k} = F(P_j; \beta_k) = \frac{\partial P_j(\beta_k)}{\partial \beta_k}, \tag{5}$$

where $F_{j,k}$ denotes the sensitivity coefficient of the jth measurement to the kth parameter. In this work, the sensitivity coefficients were estimated by the finite difference method:

$$F(P_j; \beta_k) \approx \frac{\Delta P_j}{\Delta \beta_k} = \frac{P_j(\beta_k + \Delta\beta_k) - P_j(\beta_k - \Delta\beta_k)}{2\Delta\beta_k}. \tag{6}$$

The total sensitivity for the kth parameter is the sum of squares of the sensitivity coefficients over the entire depth of the ^{30}Si profile:

$$\Phi_k = \sum_{p=1}^{N_d} \frac{1}{\sigma_p^2} \left(\frac{C^{^{30}\mathrm{Si}}_{p,\beta_k+\Delta\beta_k} - C^{^{30}\mathrm{Si}}_{p,\beta_k-\Delta\beta_k}}{2\Delta\beta_k} \right)^2, \tag{7}$$

where N_d is the total number of data points in the ^{30}Si profile, and $\Delta\beta_k/\beta_k = 0.1$. The individual values of F are weighted by the corresponding measurement covariance at the pth depth, σ_p^2. A higher value of the total sensitivity Φ_k implies a stronger influence of the corresponding model parameter β_k on the final ^{30}Si profile. The data points with smaller measurement covariance are weighted more strongly in determining the total sensitivity. The standard deviation at each depth σ_p was estimated by the measurement of n different SIMS profiles on the same specimen, yielding

$$\sigma_p = \frac{1}{n-1} \sqrt{\sum_{q=1}^{n} (C_{p,q} - \hat{C}_p)^2}, \tag{8}$$

where $\hat{C}_p$ is the average concentration over the n measurements for the pth depth. The standard deviation on a relative basis (i.e., normalized by concentration) obeyed the following square-root relation:

$$\frac{\sigma_p}{\hat{C}_p} = a_1 \frac{1}{\sqrt{\hat{C}_p}} + a_2, \tag{9}$$

where a_1 and a_2 denote constants equal to 4.4×10^8 and 2.7×10^{-2}, respectively (Kwok et al., 2005). Eq. (9) indicates that the relative measurement error decreased with the square root of the signal strength.

As experiments were performed under various degrees of adsorbate coverage, the sensitivity analysis was repeated for various values of the surface loss probability S. This helped elucidate the effect of surface boundary condition on the governing diffusion/reaction mechanism in the system. The results also can be used to guide experimental design if particular parameters need to be refined in future studies.

4.2. Maximum a posteriori (MAP) estimation

A closer inspection of Fig. 1 indicates that the profile obtained by the ML estimate of the surface loss probability S is *twice* as far as the experimental profile from the as-implanted profile for sample 1, for a depth between 30 and 55 nm. This large relative error for sample 1 suggests that improved estimation may be required for the activation energies in Table 1, for the model to describe profile spreading for high values of the surface loss probability. This motivates the application of maximum a posteriori (MAP) estimation, which determines the most likely values of parameters when prior information is available (Beck & Arnold, 1977; Sparacino, Tombolato, & Cobelli, 2000). MAP estimation optimally combines prior statistical information of the parameter estimates with additional experimental data to

obtain improved *a posteriori* estimates. For this application MAP estimation can be equivalently posed as a minimization problem:

$$\min_{\substack{\bar{\beta},S_j \\ j=1,\cdots,d}} (\bar{\beta}-\mu)^{\mathrm{T}} V_{\mu}^{-1}(\bar{\beta}-\mu) + \sum_{j=1}^{d} \left(Y_j - P_j(\bar{\beta},S_j)\right)^{\mathrm{T}} V_{\varepsilon,j}^{-1} \left(Y_j - P_j(\bar{\beta},S_j)\right), \tag{10}$$

where $\bar{\beta}$ denotes the vector of estimated parameters that are the same for all profiles, μ is the vector of corresponding *a priori* parameter estimates, V_{μ} is the prior parameter covariance matrix, d is the total number of SIMS profiles, S_j is the surface loss probability, Y_j is the vector of experimental observations, P_j is the vector of model predictions, and $V_{\varepsilon,j}$ is the measurement covariance matrix for the jth profile. The *a priori* parameter estimates are shown in Table 1, and were estimated from density functional theory calculations, isolated experiments from the literature, and previous experimental concentration profiles as cited in the table. Assuming the measurement errors at different depths are uncorrelated, the covariance matrix of the jth measurement is diagonal with nonzero elements given by

$$[V_{\varepsilon,j}]_{pp} = \sigma_{p,j}^2, \tag{11}$$

where $\sigma_{p,j}$ is the standard deviation of the jth depth profile determined by (9). A complete set of as-implanted and annealed profiles at different surface conditions was collected at the same location on different specimens. A systematic shift in each set of profiles was identified for each spatial location, which was corrected before applying ML and MAP estimation. This did not correct biases within 1 nm of the surface, which were sufficiently large for that data to be of limited value for parameter estimation purposes, so these data points were excluded from the parameter estimation (Kwok et al., 2005). The remaining data points indicated that the measurement error depended upon isotope concentration as described by (9).

Assuming the accuracies of the individual prior parameter estimates are independent of one another, the prior parameter covariance matrix can be expressed as a diagonal matrix with nonzero entries given by

$$V_{\mu,kk} = \sigma_{\bar{\beta}_k}^2, \tag{12}$$

where $\sigma_{\bar{\beta}_k}$ is the standard deviation in the prior estimate of the kth parameter. No prior information was assumed for the surface loss probability S_j, which varies with the adsorbate coverage that is different for each profile. In the case where only the surface loss probabilities are estimated, (10) gives maximum likelihood estimates.

For computational efficiency, the elements within $\bar{\beta}$ are chosen to exclude activation energies to which the model has little sensitivity (based on the results of sensitivity analysis). The objective function in (10) is formulated as the sum of weighted squared differences between the experimental profile concentrations and the model predictions for *all* available profiles. The values of various activation energies in $\bar{\beta}$ are independent of the experimental conditions, while the values of S_j are a function of nitrogen coverage.

All of the estimated parameters can be embedded into a single vector,

$$\beta^{\mathrm{T}} = \left[\bar{\beta}^{\mathrm{T}}, S_1, \cdots, S_d\right], \tag{13}$$

with an estimate of the covariance of $(\beta^* - \beta_{\mathrm{true}})$ given by

$$\mathrm{cov}(\beta^* - \beta_{\mathrm{true}}) = V_{\beta^*} \approx \left(F^{\mathrm{T}} \begin{pmatrix} V_{\varepsilon,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & V_{\varepsilon,d} \end{pmatrix}^{-1} F + \begin{pmatrix} V_{\mu}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1}, \tag{14}$$

where β^* and β_{true} denote the best estimate and the true value of the vector of parameters, respectively, and each 0 is a matrix of zeros of compatible dimensions. The matrix F is the sensitivity matrix of

$$P = [P_1, \cdots, P_d]^{\mathrm{T}} \tag{15}$$

with respect to the vector of parameters β, computed from (6).

The accuracy of the parameter estimates are quantified by a hyperellipsoidal confidence region given by

$$E_{\beta} = \left\{\beta : (\beta - \beta_{\mathrm{true}}) V_{\beta^*}^{-1} (\beta - \beta_{\mathrm{true}}) \le \chi_{\alpha}^2(p)\right\}, \tag{16}$$

where α denotes the confidence level and χ denotes the chi-squared distribution with p degrees of freedom. The confidence region can be visualized by approximate confidence intervals (Matthews, 1997):

$$\beta_k^* - \sqrt{\chi_{\alpha}^2(p) V_{\beta^*,kk}} \le \beta_{\mathrm{true},k} \le \beta_k^* + \sqrt{\chi_{\alpha}^2(p) V_{\beta^*,kk}}. \tag{17}$$

The confidence intervals are a poorer representation of the hyperellipsoidal confidence region when the off-diagonal elements in $V_{\beta*}$ are large relative to its diagonal elements, in which case the correlation between the parameters is significant.

5. Results and discussion

5.1. Sensitivity analysis

Values of the total sensitivities for different surface loss probabilities are shown in Table 2. The results are:

(1) The energetics of exchange between interstitial and substitutional silicon, E_{ex}, have a large influence on the final concentration profile for high values of the surface loss probability S (10^{-2} and 10^{-4}). The total sensitivity decreases as S decreases.
(2) The total sensitivities of cluster energetics mostly increase as the surface loss probability S decreases. For small S (10^{-6} and 10^{-8}), the profile is most sensitive to the cluster energetics for large clusters (E_4 and E_{large}).
(3) The profile is sensitive to the activation energy for Si interstitial diffusivity, E_{diff}, over the entire range of the surface loss probability S.

In Si implanted to significant levels of interstitial supersaturation as described above, the self-diffusion behavior is determined primarily by the interplay between the interstitials, the surface, and the reservoirs that render interstitials immobile, i.e., lattice sites and interstitial clusters. A high value of the surface loss probability S implies a high flux at the surface. As interstitials of both ^{28}Si and ^{30}Si move toward the surface, interstitial ^{30}Si will be selectively immobilized by the exchange reaction with lattice sites described in (3), owing to a higher concentration of substitutional ^{28}Si than ^{30}Si. Continuous removal of the interstitial ^{28}Si at the surface drives the exchange reaction (3) to the right hand side, resulting in more ^{30}Si kicked into lattice sites. Owing to this selective removal of interstitial ^{28}Si over ^{30}Si, a high value of the surface loss probability S results in a less diffused final ^{30}Si profile (see Fig. 1).

The results of the sensitivity analysis can be rationalized as follows. For all values of the surface loss probability, the highest reaction rates are associated with the forward and reverse exchange reaction (3) for nearly the entire annealing process (see Fig. 2). As discussed above, for high values of the surface loss probability S, the forward exchange reaction (3) that moves interstitial ^{30}Si into lattice sites dominates during the early stages of annealing (Fig. 2(c)). As this is the highest reaction rate, for high S the profile spreading is very sensitive to the activation energy for the kick-in reaction (E_{ex}). For low values of the surface loss probability S, the lack of an efficient surface sink leads to

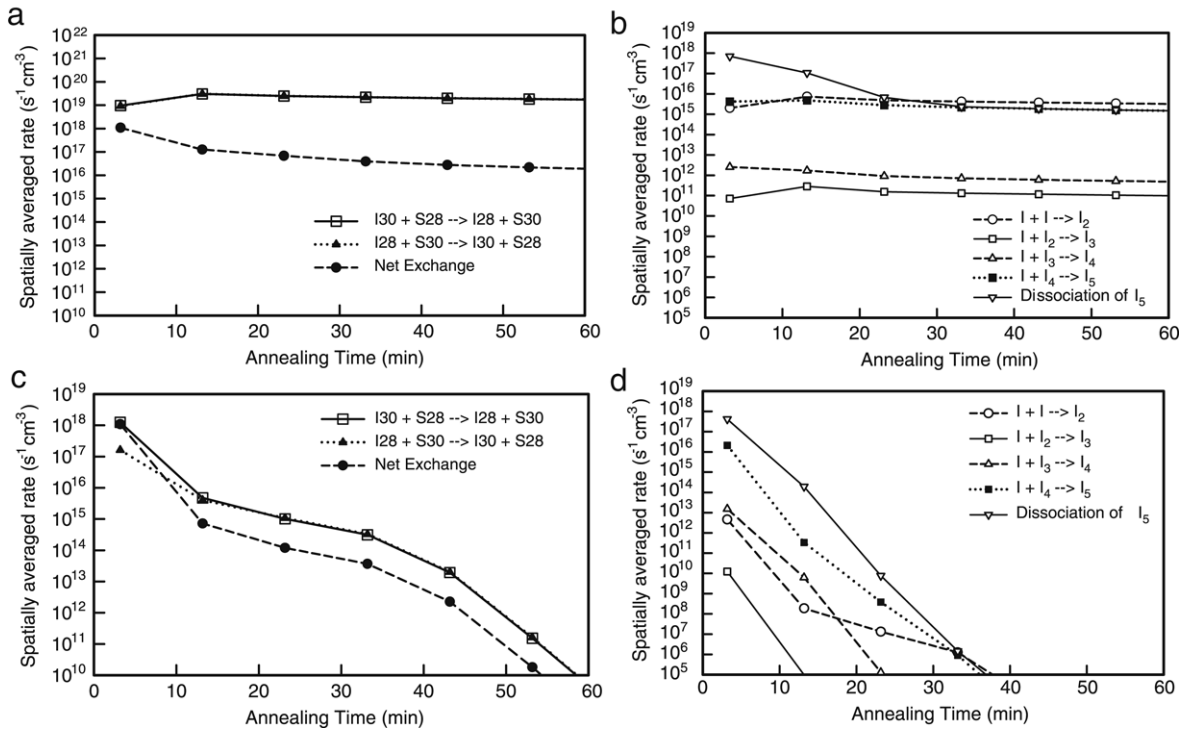

Fig. 2. Simulated spatially-averaged reactions rates for the (a) exchange reaction between interstitial and lattice atoms and (b) cluster dissociation and association for $S = 10^{-6}$. The corresponding plots for $S = 10^{-2}$ are (c) and (d). The reaction rate is integrated over the entire profile and divided by the corresponding depth to obtain the average. The net exchange rate is the absolute value of the difference between the forward and backward exchange reactions in (3).

Table 2
Total sensitivity coefficients Φ of activation energies at various surface loss probabilities

Parameter	Total sensitivity, Φ			
	$S = 10^{-2}$	$S = 10^{-4}$	$S = 10^{-6}$	$S = 10^{-8}$
E_2	2.9	2.2	4.3×10^3	6.1×10^4
E_3	0.83	7.4	8.7×10^3	6.4×10^4
E_4	11.	6.0×10^3	2.5×10^5	8.4×10^4
E_{large}	2.0×10^2	7.9×10^4	9.3×10^5	1.1×10^6
E_{diff}	2.2×10^5	1.5×10^4	4.1×10^3	6.3×10^4
E_{ex}	2.3×10^5	1.7×10^4	39.	27.

concentrations of interstitials that are orders-of-magnitude higher in the bulk silicon, such that the forward and reverse reaction rates (3) very quickly reach quasi-steady-state values that are nearly equal and constant throughout the annealing (Fig. 2(a)), with the net reaction rate being ~1000 × smaller than the absolute rates for most of the annealing. For reactions that occur at a quasi-steady-state, the species concentrations depend primarily on the ratio of the forward and reverse reaction rate constants rather than their absolute values (Fogler, 2004). Varying the activation energy for the kick-in reaction (E_{ex}) has the same effect on the forward and reverse exchange reaction rates (3); thus changing the value of E_{ex} has a low effect on profile spreading for low values of the surface loss probability S.

The rates for all reactions, including the cluster reactions, are much higher for low values of the surface loss probability S (see Fig. 2(a), (b)). This is consistent with the total sensitivities for the activation energies associated with cluster reactions (E_2, E_3, E_4, E_{large}) being higher for low S.

The magnitude of the total sensitivity for E_{diff} agrees with physical intuition. It is reasonable to expect that the diffusivity of the silicon interstitial would play an important role on the final annealed profile regardless of the mechanism that render interstitials immobile. Hence the total sensitivity of E_{diff} is high for all values of the surface loss probability S.

Based on the results of ML estimation with the surface loss probability S as the only adjustable parameter (see Table 1), the available experimental data gave values of the surface loss probability S lying between 10^{-4} and 10^{-1}. In this regime, the concentration profiles are relatively insensitive to the dissociation energy of small clusters (see Table 2). As a result, E_2 and E_3 were not included in the parameter set for maximum a posteriori estimation, resulting in the vector of parameter estimates

$$\beta^{\mathrm{T}} = \left[E_4, E_{\text{large}}, E_{ex}, E_{\text{diff}}, S_1, S_2, S_3\right]. \tag{18}$$

According to the results of the sensitivity analysis, if the values of E_2 and E_3 need to be determined more accurately in the future, then the experiments should be performed under conditions in which the surface boundary condition is close to that of a perfect reflector, that is, experimental conditions in which the surface loss probability S is small.

The concentration profiles were not very sensitive to E_{large} and E_{diff} in previous experimental systems (which is why the standard deviations from past parameter estimation studies are relatively large in Table 1), but are very sensitive to these parameters in the present system (see Table 2). This indicates that incorporating the experimental results from the present system with prior information should result in more accurate estimates of these parameters.

The dissociation energy of the second largest cluster, E_4, shows moderate sensitivity for $S \approx 10^{-4}$ (see Table 2). Although a prior parameter estimation study for boron diffusion experiments estimated a standard deviation of 0.002 eV for the estimate of this parameter, this value is too small for use in the MAP estimation for this silicon self-diffusion study. In the boron system, there are three kinds of clusters, namely pure B clusters, pure Si clusters, and mixed B-Si clusters. The dissociation energy for clusters of size bigger than two was assumed to be solely size-dependent. This assumption is probably only an approximation. Results from first-principles calculations

Table 3
Maximum a posteriori (MAP) estimates

Parameter	MAP estimate	Standard deviation	95% confidence interval half-width
E_4	3.1087 eV	0.012	0.044
E_{large}	3.7749 eV	0.0034	0.013
E_{diff}	0.7642 eV	0.0031	0.012
E_{ex}	0.9636 eV	0.0031	0.012
S_1	6.72×10^{-2}	3.0×10^{-4}	9.9×10^{-4}
S_2	5.77×10^{-4}	4.8×10^{-6}	1.6×10^{-5}
S_3	2.19×10^{-4}	1.2×10^{-6}	4.1×10^{-6}

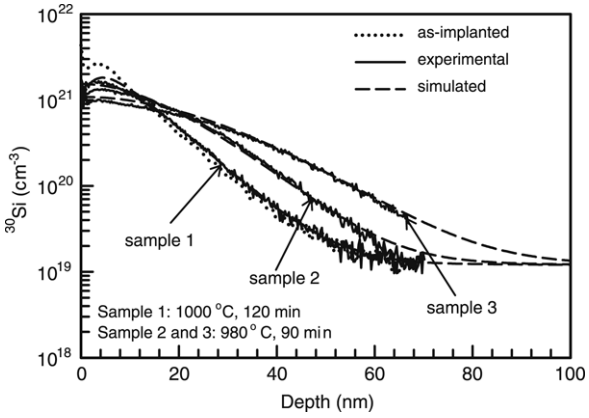

Fig. 3. Experimental and simulated ^{30}Si profiles using refined estimates of the surface loss probability S and various energetics determined by maximum a posteriori (MAP) estimation. Fits are noticeably better than in Fig. 1.

(Lenosky, Sadigh, Theiss, Caturla, & Rubia, 2000) and the tight-binding method (Luo & Clancy, 2001) indicate that the identity of atoms in a cluster affects the value of the formation energy. In addition, it is unlikely that the dissociation of these three groups of clusters is of same importance in affecting the final boron profile. The estimated high accuracy of the estimate of E_4 may have been due to a high sensitivity of the concentration profiles to the dissociation of a *certain* group of clusters. These issues suggest that the prior standard deviation for E_4 used in the current MAP estimation should be increased; we used 0.1 eV, which is the same as that of E_{large}. This makes the refined *a posteriori* estimate of E_4 a more defensible value for pure Si clusters. In the next section the refined value is compared to the prior estimate to quantify the validity of the "size-dependent-only" assumption.

5.2. Maximum a posteriori (MAP) estimation

MAP estimates of the model parameters and the associated 95% confidence intervals are shown in Table 3. Fig. 3 shows the simulated profiles using the refined parameters. The simulated profiles with MAP estimates are closer to the experimental profiles than when *a priori* estimates are used (see Fig. 1), especially for the atomically clean surface (sample 1, for depth between 20 and 55 nm). Values of the MAP estimates of the activation energies are reasonably close to the prior estimates, but with confidence intervals significantly tightened. With the exception of E_4, each activation energy estimate lies within the 95% confidence region of the prior estimate, suggesting that the values of the model parameters are consistent in the various experimental systems studied. Note that the MAP estimates of S lie outside the 95% confidence region of the prior estimate. This suggests that the estimated values of S are sensitive to the values of other parameters in the model. The modification of other parameters by MAP estimation affects the best estimate of S.

As discussed above, the error bound in the prior estimate of E_4 (the dissociation energy of the second largest clusters) was loosened in the present study. Even so, the refined MAP estimate is larger than the prior estimate and is outside the original confidence interval. This discrepancy suggests that the dissociation of clusters may not depend solely on the number of atoms in the cluster, as discussed earlier, but also on the composition of four-atom clusters.

Another possible cause is the limitation of the model to clusters of no more than five atoms. This limitation seemed to cause little difficulty when modeling the spike annealing of boron profiles (Jung et al., 2004a). With the maximum cluster size restricted to five, the dissociation energies of clusters are widely discretized with $E_2 = 1.4$ eV and $E_5 = E_{\text{large}} = 3.7$ eV. This wide discretization causes a huge number of cluster dissociation events to occur at a few specific temperatures. The model may therefore be insufficiently able to simulate cluster dissociation at intermediate temperatures, especially in soak anneal temperature programs. Moreover, when four-atom clusters start to dissociate at around 700°C, the released interstitials cannot re-associate to clusters since five-atom clusters are not allowed to accrete interstitials. This situation does not represent the actual physical picture where the biggest size clusters can still accrete interstitials which results in the formation of {311} defects (Cowern et al., 1999). In order to obtain more widely applicable values for cluster activation energies, refinement in the diffusion model to include more cluster sizes and additional experiments may be necessary.

6. Conclusions

The energetics in silicon self-diffusion were identified using a combination of parameter sensitivity analysis and maximum a posteriori (MAP) estimation. Refined parameter estimates were obtained by combining information from multiple experimental systems. The much higher sensitivity of the concentration profiles in the new experimental system to the dissociation energy of the largest clusters E_{large} and the energetics of Si interstitial diffusion E_{diff} enabled an order-of-magnitude improvement in the accuracy of the estimates of these parameters, upon application of MAP estimation. The greatest improvement in accuracy was in the activation energy for the exchange between interstitial and substitutional silicon, E_{ex}, which had a very strong effect on the concentration profiles for high values of the surface loss probability S. The results also suggested that refinement in the model may be necessary to obtain a more widely applicable parameter set for cluster dissociation.

Acknowledgments

This work was partially supported by NSF (CTS 02-03237). SIMS was performed at the Center for Microanalysis of Materials, University of Illinois, which is partially supported by the US Department of Energy under grant DEFG02-96-ER45439. We thank Steve Burdin at Isonics Corporation for overseeing the successful creation of isotopically-labeled specimens. We also thank Joel Ager at Lawrence Berkeley National Laboratories for arranging the supply of isotopically-labeled silane.

References

Beck, J. V., & Arnold, K. J. (1977). *Parameter estimation in engineering & science*. New York: Wiley.

Braatz, R. D., Alkire, R. C., Seebauer, E. G., Rusli, E., Gunawan, R., Drews, T. O., et al. (2006). Perspectives on the design and control of multiscale systems. *Journal of Process Control*, *16*, 193–204.

Bracht, H., Haller, E. E., & Clark-Phelps, R. (1998). Silicon self-diffusion in isotope heterostructures. *Physical Review Letters*, *81*, 393–396.

Bracht, H., Stolwijk, N. A., & Mehrer, H. (1995). Properties of intrinsic point defects in silicon determined by zinc diffusion experiments under nonequilibrium conditions. *Physical Review B*, *52*, 16542–16560.

Chakravarthi, S., & Dunham, S. T. (2001). A simple continuum model for boron clustering based on atomistic calculations. *Journal of Applied Physics*, *89*, 3650–3655.

Cowern, N. E. B., Janssen, K. T. F., van de Walle, G. F. A., & Gravesteijn, D. J. (1990). Transient diffusion of ion-implanted B in Si - Dose, time, and matrix dependence of atomic and electrical profiles. *Physical Review Letters*, *65*, 2434–2437.

Cowern, N. E. B., Mannino, G., Stolk, P. A., Roozeboom, F., Huizing, H. G. A., van Berkum, J. G. M., et al. (1999). Energetics of self-interstitial clusters in Si. *Physical Review Letters*, *82*, 4460–4463.

Emery, A. F., & Nenarokomov, A. V. (1998). Optimal experiment design. *Measurement Science & Technology*, *9*, 864–876.

Fahey, P. M., Griffin, P. B., & Plummer, J. D. (1989). Point-defects and dopant diffusion in silicon. *Reviews of Modern Physics*, *61*, 289–384.

Fogler, H. S. (2004). *Elements of chemical reaction engineering* (4th edition). Upper Saddle River, NJ: Prentice Hall.

Giles, M. D. (1991). Transient phosphorus diffusion below the amorphization threshold. *Journal of the Electrochemical Society*, *138*, 1160–1165.

Gunawan, R., Jung, M. Y. L., Seebauer, E. G., & Braatz, R. D. (2003). Maximum a posteriori estimation of transient enhanced diffusion energetics. *AIChE Journal*, *49*, 2114–2123.

Gunawan, R., Jung, M. Y. L., Braatz, R. D., & Seebauer, E. G. (2003). Parameter sensitivity analysis of boron activation and transient enhanced diffusion in silicon. *Journal of the Electrochemical Society*, *150*, G758–G765.

Jain, S. C., Schoenmaker, W., Lindsay, R., Stolk, P. A., Decoutere, S., Willander, M., et al. (2002). Transient enhanced diffusion of boron in Si. *Journal of Applied Physics*, *91*, 8919–8941.

Jung, M. Y. L., Gunawan, R., Braatz, R. D., & Seebauer, E. G. (2003). Ramp-rate effects in transient enhanced diffusion and dopant activation. *Journal of the Electrochemical Society*, *150*, G838–G842.

Jung, M. Y. L., Gunawan, R., Braatz, R. D., & Seebauer, E. G. (2004a). Pair diffusion and kick-out: Contributions to diffusion of boron in silicon. *AIChE Journal*, *50*, 3248–3256.

Jung, M. Y. L., Gunawan, R., Braatz, R. D., & Seebauer, E. G. (2004b). Effect of near-surface band bending on dopant profiles in ion-implanted silicon. *Journal of Applied Physics*, *95*, 1134–1139.

Kwok, C. T. M., Dev, K., Braatz, R. D., & Seebauer, E. G. (2005). A method for quantifying annihilation rates of bulk point defects at surfaces. *Journal of Applied Physics*, *98*, 013524.

Mark Law, (2005). University of Florida at Gainesville. http://www.swamp.tec.ufl.edu/.

Matthews, H. B. III. (1997). Ph.D. thesis, Madison, University of Wisconsin.

Leis, J. R., Gallagher, S. A., & Kramer, M. A. (1987). Parametric sensitivity analysis of complex process flowsheets using sequential modular simulation. *Computers & Chemical Engineering*, *11*, 409–421.

Lenosky, T. J., Sadigh, B., Theiss, S. K., Caturla, M., & Rubia, T. D. (2000). Ab initio energetics of boron-interstitial clusters in crystalline Si. *Applied Physics Letters*, *77*, 1834–1836.

Luo, W., & Clancy, P. (2001). Identification of stable boron clusters in c-Si using tight-binding statics. *Journal of Applied Physics*, *89*, 1596–1604.

Murto, B. (1999). Recent advances and continuing challenges in ultra-shallow junctions. In Proceedings of the third national implant users meeting, Austin, TX, October.

Seebauer, E. G., Dev, K., Jung, M. Y. L., Vaidyanathan, R., Kwok, C. T. M., Ager, J. W., et al. (2006). Control of defect concentrations within a semiconductor through adsorption. *Physical Review Letters*, *97*, 055053.

Shao, L., Liu, J., Chen, Q. Y., & Chu, W. (2003). Boron diffusion in silicon: the anomalies and control by point defect engineering. *Materials Science & Engineering R*, *42*, 65–114.

Sparacino, G., Tombolato, C., & Cobelli, C. (2000). Maximum-likelihood versus maximum a posteriori parameter estimation of physiological system models: The C-peptide impulse response case study. *IEEE Transactions on Biomedical Engineering*, *47*, 801–811.

Stolk, P. A., Gossmann, H.-J., Eaglesham, D. J., Jacobson, D. C., Rafferty, C. S., Gilmer, G. H., et al. (1997). Physical mechanisms of transient enhanced dopant diffusion in ion-implanted silicon. *Journal of Applied Physics*, *81*, 6031–6050.

Tomović, R., & Vukobratović, M. (1972). *General sensitivity theory*. New York: Elsevier.

Van Voorhees, F. D., & Bahill, A. T. (1995). Parametric sensitivity analysis: a tool for robust design. Proceedings of the IEEE international conference on systems, man & cybernetics: Vol. 2 (pp. 971–976). Vancouver, Canada.

Walter, E., & Pronzato, L. (1990). Qualitative and quantitative experiment design for phenomenological models — a survey. *Automatica*, *26*, 195–213.

Charlotte T.M. Kwok is an engineer in research and development at Taiwan Semiconductor Manufacturing Company. She received the B.S. degree from Hong Kong University of Science and Technology and M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign. Her interests are in semiconductor processing including the modeling and simulation of dopant diffusion in silicon after ion implantation.

Kapil Dev is an analysis engineer at Intel. He received the B.S. degree from Indian Institute of Technology of New Delhi and M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign. His interests are in semiconductor processing including the engineering of defects for controlling dopant diffusion in silicon.

Edmund G. Seebauer is James W. Westwater Professor at the University of Illinois at Urbana-Champaign. He received the B.S. degree from the University of Illinois at Urbana-Champaign and Ph.D. degree from the University of Minnesota. His interests are in the engineering of semiconductor defects for nanoscale devices. He was a Distinguished Lecturer for the Institute for Electrical and Electronics Engineers and is a Fellow of the American Vacuum Society, the American Physical Society, and the American Association for the Advancement of Science. His interactions with semiconductor companies span more than 20 years.

Richard D. Braatz is Professor and Millennium Chair at the University of Illinois at Urbana-Champaign. He received the B.S. degree from Oregon State University and M.S. and Ph.D. degrees from the California Institute of Technology. His interests are in the modeling and control of complex and multiscale systems. Honors include the Donald P. Eckman Award of the American Automatic Control Council, ASEE Curtis W. McGraw Research Award, Antonio Ruberti Young Researcher Prize, and IEEE Fellow. Professor Braatz has consulted/collaborated with more than a dozen companies including IBM, UTC Power, and Merck.