Physical integrated diffusion-oxidation model for implanted nitrogen in silicon

Lahir Shaik Adam and Mark E. Law
Department of Electrical Engineering, Software for Advanced Materials Processing (SWAMP) Center, University of Florida, Gainesville, Florida 32611

Omer Dokumaci and Suri Hegde
Semiconductor Research and Development Center, IBM Corporation, East Fishkill, New York 12533

Received 2 July 2001; accepted for publication 1 November 2001

Scaling the gate oxide thickness is one of many process development challenges facing device engineers today. Nitrogen implantation has been used to control gate oxide thickness. By varying the dose of the nitrogen implant, process engineers can have multiple gate oxide thicknesses in the same process. Although it has been observed that nitrogen retards gate oxidation kinetics, the physics of how this occurs is not yet well understood. Since the retardation in oxide growth is due to the diffusion of nitrogen and its subsequent incorporation at the silicon/silicon oxide interface, the study of the diffusion behavior of nitrogen in silicon becomes important. Further, it is also necessary to study how this diffusion behavior impacts oxide growth. Models have been developed to explore these issues. The diffusion model is based on ab initio results and is compared to experimental results at two temperatures. The oxide reduction model is based on the diffusion of nitrogen to the surface. The surface nitrogen is coupled to the surface reaction rate of silicon and oxygen to moderate oxide growth. © 2002 American Institute of Physics. [DOI: 10.1063/1.1430537]

I. INTRODUCTION

Nitrogen implantation has been used to control the gate oxide thickness. This implies that by controlling the dose and energy of the nitrogen implant, one can vary the thickness of the gate oxide across the wafer. Therefore, the nitrogen implantation process is especially useful for processes that require multiple gate oxide thicknesses for example, System On A Chip technologies. Although it has been observed that nitrogen retards gate oxidation kinetics, the physics of how this occurs is not yet well understood. Since the retardation of oxide growth is due to the diffusion of nitrogen to the surface and its subsequent incorporation at the silicon/silicon–oxide interface, the study of the diffusion behavior of nitrogen becomes important. Furthermore, it is also important to study how this diffusion behavior impacts oxide growth. This article discusses models that deal with these issues. Section II of this article summarizes the experimental results previously reported by us. These experimental results form the basis of the diffusion model. The diffusion model is described in detail in Sec. III. As mentioned above, the motivation behind implanting nitrogen in silicon is to retard gate oxide growth. Therefore, if the diffusion model can be coupled to an oxidation model to moderate the gate oxide growth, then this integrated diffusion–oxidation model can be of considerable use in gate oxide thickness optimization.

II. PREVIOUS EXPERIMENTAL RESULTS

Experimental results of nitrogen implanted at $5 \times 10^{13}$ N$_2$ ions/cm$^2$, 40 keV have been reported previously in Ref. 2. The key observations in Ref. 2 are as follows: Nitrogen diffuses toward the surface rapidly with time at 750 °C. However, the nitrogen barely moved at 650 °C. Furthermore, the peaks of the nitrogen implants shifted toward the surface with anneal time. No extended defects were observed in any of the annealed samples.

The last observation mentioned above is of particular interest. This is because silicon implants at this dose and energy annealed at 750 °C show large numbers of extended defects. However, transmission electron microscopy did not reveal any extended defects for the nitrogen implants. This motivates the fact that any explanation of the diffusion behavior of implanted nitrogen in silicon cannot rely on extended defects. A physical model that relies only on submicroscopic defects and nitrogen complexes has been used to explain the observed diffusion behavior. This will be explained in detail below.

III. DIFFUSION MODEL

A. Theory

The following reactions have been used in this model:

$$N_S + I = N_I,$$  \hspace{1cm} (1)

$$N_I + V = N_S,$$  \hspace{1cm} (2)
\[ N_S + V = NV, \]
\[ NV + I = N_S, \]
\[ I + I = I_2, \]
\[ V + V = V_2, \]
\[ N_I + V_2 = NV, \]
\[ I + V = 0. \]

In the above equations \( N_S, N_I, N_V, I, V, I_2, \) and \( V_2 \) refer to substitutional nitrogen, interstitial nitrogen, nitrogen vacancy pairs, self-interstitials, vacancies, di-interstitials, and divacancies, respectively. The surface flux is determined according to the expression

\[ J_s = -k_r C_{X=0}, \]

where \( J_s \) corresponds to the flux of the mobile species at the interface, \( k_r \) is the surface recombination velocity of the mobile species, and \( C_{X=0} \) corresponds to the concentration of the mobile species at the interface. The \( I \) and \( V \) profiles have been obtained through a \textsc{ut-marlowe}\textsuperscript{5} simulation where the simulation implant energies were scaled according to the boron to nitrogen mass ratio.

The process of determining the reaction rates will be discussed generally for any dopant–defect pair. This is intended only as a brief review and the reader is invited to read the seminal paper by Fahey and Plummer\textsuperscript{5} for an extensive review on the subject. This formalism can then be applied to any of the reactions mentioned above.

Let us consider a reaction of the form

\[ A + X = AX. \]

In Eq. (10), \( A \) refers to a substitutional dopant atom, \( X \) refers to a point defect (\( I \) or \( V \)), and \( AX \) refers to a dopant–defect pair. Assuming local equilibrium, the concentration of the dopant–defect pair \( C_{AX} \) can be written as

\[ C_{AX} = (k_f / k_r) C_A C_X. \]

In Eq. (11), \( k_f \) corresponds to the forward reaction rate, \( k_r \) corresponds to the reverse reaction rate, and \( C_A \) and \( C_X \) correspond to the concentrations of the substitutional dopant and point defect, respectively.

In thermodynamic equilibrium, \( C_{AX} \) can also be written as

\[ C_{AX} = \theta_{AX} \exp(-S_{AX} / k)(C_A C_X / C_L) \exp(E_{bAX} / kT). \]

In Eq. (12), \( \theta_{AX} \) refers to the degree of freedom of the dopant defect pair (for example, if \( AX \) is a dopant–vacancy pair, then, there are four equivalent positions where a vacancy can be placed and \( \theta_{AV} = 4 \)). \( S_{AX} \) corresponds to the entropy of reaction (10), \( C_L \) is the lattice concentration (\( 5 \times 10^{22}/\text{cm}^3 \) in silicon), \( E_{bAX} \) refers to the binding energy of the dopant–defect pair, \( k \) is the Boltzmann constant, and \( T \) refers to the absolute temperature in K.

Equating Eqs. (11) and (12), we can express the ratio of the forward and the reverse reaction rates as

\[ (k_f / k_r) = (C_L / \theta_{AX}) \exp(S_{AX} / k) \exp(-E_{bAX} / kT). \]

However, in order to determine unique reaction rates, the forward reaction has been determined assuming diffusion limited kinetics.\textsuperscript{5} The forward reaction rate can be expressed as

\[ k_f = 4 \pi d_X \exp(-\Delta E / kT). \]

In Eq. (14), \( a \) refers to the capture radius, \( d_X \) refers to the diffusivity of the point defect, and \( \Delta E \) refers to the barrier height. The other symbols have their usual meanings.

Once the reaction rates have been determined, the dopant evolution can be determined numerically by solving a set of differential equations. For example, the evolution of \( N_I \) in Eq. (1) can be determined as

\[ \partial C_{NI} / \partial t = (\partial / \partial X)(d_{NI}(\partial C_{NI} / \partial X)) + k_{f1} C_{NS} C_I - k_{r1} C_{NI} - k_{f2} C_{NS} C_{V2} + k_{r2} C_{NS} C_{V2} + k_{r7} C_{NV}. \]

In Eq. (15), \( k_{f1} \) represents the forward reaction rate of the \( i \)th reaction and \( k_{r1} \) represents the reverse reaction rate of the \( i \)th reaction. For example, \( k_{f1} \) is the forward reaction rate of reaction (1) and \( k_{r1} \) is the reverse reaction rate of reaction (1). \( k_{fi} \) and \( k_{ri} \) are determined using the formalism developed in Eqs. (10)--(14). All the other symbols have their usual meanings as described above. Similar equations can be written for the other reactions. The diffusivity of the impurity–defect pair can be expressed as

\[ d = d_0 \exp[-(E_f + E_m) / kT]. \]

In Eq. (16), \( d \) refers to the diffusivity of the mobile impurity–defect specie and \( E_f \) and \( E_m \) refer to their formation and migration energies, respectively. To avoid any confusion of notation, it must be noted that the diffusivity \( d \) as mentioned in Eq. (16) is the diffusivity of the impurity–defect pair and NOT the aggregate diffusivity \( D \) of the impurity itself.

In this model, the binding, migration, and formation energies are determined through \textit{ab initio} calculations\textsuperscript{6,7} or through experimental observation.\textsuperscript{5} It must also be mentioned that in Ref. 6, it has been reported that the exchange barrier for NV is 4.4 eV. This means that the energy required to exchange a nitrogen atom with a vacancy at the first neighbor site is 4.4 eV. This energy barrier is high. Therefore, the probability of the nitrogen–vacancy pair being mobile is very low and hence is considered immobile in this model. Thus, the nitrogen interstitials are the only nitrogen related mobile species in the set of reactions considered. The effect of entropy was neglected (assumed to be 0) as there is no consensus in the literature on the entropy values. The capture radius \( a \) was assumed to be one lattice spacing (5.43 Å) and the barrier heights were extracted to fit the measured data.

The parameter set is shown in Table I. The only fitting parameters in this model are the reaction barriers. Although there are seven fitting parameters as shown in Table I, it has to be appreciated that if every parameter related to nitrogen were used as a fitting parameter in the model, then one would end up with about 35 fitting parameters for this model (8 reactions×2 reaction directions×2 temperatures used to
TABLE I. Parameter set used in the diffusion model.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Binding energy (eV)</th>
<th>Reaction barrier (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_1^+ + I = N_1)</td>
<td>3.5 (ab initio)</td>
<td>0.55</td>
</tr>
<tr>
<td>(N_1^+ + V = N_2)</td>
<td>3.1 (ab initio)</td>
<td>0.07</td>
</tr>
<tr>
<td>(N_1^+ + V = NV)</td>
<td>1.73 (ab initio)</td>
<td>0</td>
</tr>
<tr>
<td>(NV + I = N_1)</td>
<td>4.87 (ab initio)</td>
<td>0.202</td>
</tr>
<tr>
<td>(I + I = I_2)</td>
<td>2.38 (ab initio)</td>
<td>0.18</td>
</tr>
<tr>
<td>(V + V = V_2)</td>
<td>3.5 (deep level transient spectroscopy)</td>
<td>0.06</td>
</tr>
<tr>
<td>(N_1^+ + V = NV)</td>
<td>1.33 (ab initio)</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Simulate the diffusion as shown later + diffusivity values. Therefore, the number of fitting parameters have been drastically reduced. Further, although the values of the reaction barriers have been used to fit measured data, still, the concept of the reaction barrier is physical and the same parameter set has been used at all temperatures. As can be seen from Table I, the values of the reaction barriers are quite low and well within physical realms. \(N_i\) is used as the reference state in the energy calculations. Care has also been taken to ensure that all reaction rates are energetically self-consistent. For an example, please refer to the Appendix.

The surface site density was fixed at \(1 \times 10^{15}/\text{cm}^2\) in the simulations. The surface bonding configuration of implanted nitrogen in silicon is not yet known. Therefore, it was decided not to arbitrarily introduce another species at the surface. Hence, it was decided to accumulate the nitrogen interstitials that reach the surface simply as nitrogen atoms at the Si/SiO\(_2\) interface. It is valid to accumulate nitrogen at the Si/SiO\(_2\) interface as x-ray photoelectron spectroscopy studies on nitrogen implanted silicon show that nitrogen is either at or very close to the Si/SiO\(_2\) interface upon annealing. However, the nitrogen atoms that accumulate at the interface have to occupy the surface sites. Therefore, as the nitrogen accumulates at the interface, the number of available sites for further nitrogen incorporation decreases. This can be simply represented by the equation below:

\[
dC_N/\text{interface}/dt = k_s(10^{15} - C_N/\text{interface})C_N,\tag{17}\]

In Eq. (17), \(k_s\) corresponds to the surface recombination velocity, \(10^{15}\) is the surface site density, \(C_N/\text{interface}\) corresponds to the concentration of nitrogen at the interface, and \(C_N\) refers to the concentration of nitrogen interstitials.

**B. Qualitative description of the diffusion model**

The as-implanted profile was read as \(N_i\). The profiles were then evolved based on a room temperature simulation for 1 day, which allowed all of the room temperature diffusion to settle. As explained above, any nitrogen that reaches the surface is assumed to be trapped there. \(I_2\) and \(V_2\) serve as the storage clusters that store \(I\) and \(V\). Since the as-implanted profiles were read as \(N_i\), the \(N_i\) at room temperature is quickly converted to \(N_2\) and NV through reactions (2) and (3). At diffusion temperatures, the NV thus formed is then converted to \(N_2\) through reaction (4) and the \(N_2\) to \(N_i\) through the kickout reaction, reaction (1). The self-interstitials released from \(I_2\) provide the necessary self-interstitials for reactions (1) and (4) to take place at diffusion temperatures. The reaction rates were such that they produced a nitrogen interstitial gradient toward the surface over the depth range of the total nitrogen profile. Since nitrogen interstitials are the only nitrogen related mobile species in the set of reactions considered, the total nitrogen profile shifts toward the surface with time.

One could argue that varying the surface recombination velocity of the self-interstitials or nitrogen at the surface can also produce a gradient toward the surface. However, varying the surface recombination velocity varies the gradient of the self-interstitials and/or the nitrogen interstitials only over the first 5–10 nm of the total nitrogen profile. Hence, a simple change in this surface recombination velocity at the Si/SiO\(_2\) interface is not enough to produce the necessary gradients of the nitrogen interstitials that are necessary to move the nitrogen toward the surface as observed through secondary ion mass spectrometry. The reaction rates have to be manipulated such that the gradients of the nitrogen interstitials toward the surface are produced over the depth of the total nitrogen profile.

As shown in Fig. 1, good fits were obtained at 750 °C. The model predicts that about 90% of the implanted nitrogen dose is at the interface after a 750 °C, 120 min anneal. Based on the calculations at 750 °C, there were no adjustable parameters to account for the temperature dependence at 650 °C. That is, once the parameter set was decided based on the 750 °C simulations, they were not changed subsequently for the 650 °C anneal simulations. The simulations were performed of the measured data at 650 °C. Good fits were obtained as shown in Fig. 2.

**IV. OXIDATION MODEL**

**A. Background**

In order to predict the oxide thickness as a function of time, the Deal–Grove equation has been historically used. This equation is shown below:

\[
[X^2/B]+[X/(B/A)] = t + \tau,\tag{18}\]
In Eq. (18), \( X \) represents the oxide thickness and \( t \) represents the time. \( \tau \) is a parameter that has the units of time and takes into account the initial oxide thickness (native oxide). In Eq. (18), \( B \) is associated with the quadratic term and is hence called the parabolic rate constant. Similarly, \( (B/A) \) is associated with the linear term and is hence called the linear rate constant. When the oxide thickness is small, the linear reaction rate \( (B/A) \) is dominant and as the thickness increases, the parabolic rate constant term gradually takes over.

As device dimensions were scaled, the gate oxide thickness also decreased. It was found that the linear regime of the Deal–Grove model \(^{12} \) (characterized by the \( B/A \) term in Eq. (18) \(^{18} \)) underpredicted the oxide thickness when the oxide thickness was less than about 100 Å. In order to model the thin oxide regime accurately (i.e., when the oxide thickness is \(<100 \text{ Å}\) ), Massoud and Plummer \(^{13} \) proposed a model for the initial rapid oxidation. The heart of this model can be briefly summarized by the following equation:

\[
x = X_0 + \int \alpha_0 \exp(-\Delta E/kT)\exp(-X/L)dt.
\]

In Eq. (19), \( X_0 \) corresponds to the thickness calculated from the Deal–Grove model characterized by Eq. (18) above, \( \alpha_0 \) and \( \Delta E \) represent an Arrhenius relationship that is temperature dependent, and \( L \) represents a decay length. \( \alpha_0 \) has units of \( \text{s}^{-1} \). The limits of the integral in Eq. (19) are from \( t = 0 \) to the time under consideration. As can be seen from Eq. (19), as the oxide thickness \( X \) increases, the effect of the second term in Eq. (19) reduces, and the Deal–Grove model gradually takes over. It must also be mentioned that it is generally believed that molecular oxygen is the primary oxidant species in the Deal–Grove regime of the oxide growth while atomic oxygen is the primary oxidant species in the initial rapid oxidation regime.

**B. Theory**

A semiempirical model that couples the diffusion model has been developed to predict the oxide thickness. The oxide thickness measured (explained in detail in Sec. IV C) fall in the thin oxide regime and we model the initial rapid oxidation present with the Massoud model as described in Sec. IV A. The value of the parameters used in the Massoud model are as shown in Table II. The model developed for oxidation includes two different factors for the reduction in oxidation rate due to nitrogen. The first applies only to the regular \( B/A \) term in the Deal–Grove equation, and the second applies to the rapid oxidation effect. This is quite reasonable, since the rapid oxidation is generally thought to occur through the diffusion of atomic oxygen as mentioned above. Since a different species is involved in accounting for the rapid oxidation, it is quite possible that it can be effected differently by nitrogen at the surface. An empirical relationship has been developed that predicts the oxide thickness as a function of implanted nitrogen dose. Based on the amount of nitrogen incorporated at the interface using the diffusion model explained in Sec. III, the surface reaction rate of the oxidant \( k_S \) and the oxide growth velocity have been scaled (reduced) continuously with time according to Eq. (20) in the Deal–Grove parameter regime. In the Massoud parameter regime, the oxide growth velocity has been scaled according to Eq. (21). Equations (20) and (21) are shown below:

\[
B/A \text{ reduction factor: } 1 - \frac{(\text{dose}(N_{\text{interface}})/1 \times 10^{15})}{10^{15}}, \tag{20}
\]

Massoud reduction factor:

\[
[b/(b+(\text{dose}(N_{\text{interface}})/1 \times 10^{15}))]^m. \tag{21}
\]

In Eqs. (20) and (21), \( b \) and \( m \) are fitting parameters. The factor of \( 1 \times 10^{15} \) corresponds to the surface site density and \( \text{dose}(N_{\text{interface}}) \) corresponds to the nitrogen dose incorporated at the interface through the diffusion model described in Sec. III. From Eq. (20) and (21), we can see that as the nitrogen interfacial dose increases, the growth velocity and the surface reaction rate reduces. As described in the diffusion model in Sec. III, as nitrogen accumulates at the interface, it fills up available lattice sites at the surface. Therefore, the number of available surface sites for the oxidant to attach itself to reduces and hence the oxide thickness decreases. Since the accumulation of nitrogen at the interface is limited to the surface site density (\( 1 \times 10^{15}/\text{cm}^2 \)) in the diffusion model, Eq. (20) has a lower limit of zero. Furthermore, Eq. (20) scales linearly with dose \( (N_{\text{interface}}) \). This adds further strength to the model because Eq. (20) impacts the linear growth regime of the Deal–Grove equation. Figure 3 shows the plots of Eq. (20) and (21) with dose. The best fits were obtained with \( b = 0.25 \) and \( m = 1.8 \). The same values of \( b \) and \( m \) were used for all the oxide simulation cases that will be discussed in Sec. IV C.
C. Simulation predictions and comparison to experiment

Simulations have been performed and the predictions compared to measured oxide thickness. The measured oxide thickness has been reported in Ref. 14 and we have also measured some data at 900 °C. Figure 4 shows the oxide thickness as a function of furnace oxidation time at 900 °C. We obtained some of the data in Fig. 4 and the rest was obtained from Ref. 14. These are the conditions that were primarily used to develop the parameters in the model. All fits are within a 5 Å measurement error or about 15%, whichever is higher. Figure 5 shows the oxide thickness as a function of nitrogen implant dose at 900 °C for three different times. Figure 6 shows the fits of the oxide thickness as a function of nitrogen implant dose data at 900 °C as published in Ref. 14. Accurate predictions of the oxide thickness are obtained in both figures.

Figure 7 shows the oxide thickness as function of implanted nitrogen dose at 800 °C for three different oxidation times. Figure 8 shows the oxide thickness as a function of rapid thermal oxidation (RTO) time for two different implanted nitrogen doses at 1050 °C. Good predictions have been obtained in both figures. Therefore, we can conclude that the oxidation model predicts the oxide thickness over a reasonably wide range of temperature—from 800 °C furnace anneals to 1050 °C RTO anneals. In addition, the fact that the same diffusion model can be used effectively to predict the oxide thickness for various doses shows that the diffusion model is quite robust.

It may be beneficial to point out that the oxidation model as presented in this article, consists only of two equations and two fitting parameters. One equation [Eq. (19)] decreases the linear growth rate of the Deal–Grove equation in a linear fashion while the other equation [Eq. (20)] moderates the growth rate of the initial rapid oxidation. Using just these two equations and two fitting parameters accurate predictions of oxide thickness data from different sources have been obtained in the wide temperature range of 800 °C furnace an-
V. CONCLUSIONS

A model for nitrogen diffusion has been presented. This model is compared to the experiment and is based on 
\textit{ab initio} calculations. A set of reactions has been manipulated such that they produce a nitrogen interstitial gradient towards the surface. Since nitrogen interstitials are the only nitrogen related mobile species in this diffusion model, the entire nitrogen profile shifts towards the surface with time as observed experimentally. In addition, the diffusion model is coupled to a model for oxide growth reduction. This model is developed using two growth reduction factors. One factor impacts the linear growth regime of the Deal–Grove equation while the other modifies the oxide growth velocity in the initial rapid oxidation regime. The reduction is based on the amount of nitrogen that diffuses to the surface as predicted by the diffusion model. This model offers accurate predictions in a wide temperature range of 800 °C furnace anneals to 1050 °C rapid thermal oxidation anneals using the same set of parameters. This integrated diffusion–oxidation model will allow prediction and optimization of multitickness gate oxide process using nitrogen diffusion.

APPENDIX

According to thermodynamic theory, the energy gained or lost by the system should depend only on the initial and final states of the system and not upon the path taken from the initial to the final state. This concept can be used to determine the energetics of the reactions that are not with respect to the reference state \( N_S \). This will be illustrated with an example below:

\[
\begin{align*}
\text{step 1} & \quad N_i + V + I \rightarrow 0 + 3 + 3.6 = 6.6 \text{ eV}, \\
\text{step 2} & \quad N_i + V \rightarrow 6.6 - 3.5 = 3.1 \text{ eV}, \\
\text{step 3} & \quad N_i \rightarrow 0.
\end{align*}
\]

In the above reaction steps, 3 and 3.6 eV correspond to the formation energies of the vacancy and the self-interstitial, respectively, as built into the Florida Object Oriented Process Simulator. The value of 3.5 eV corresponds to the binding energy of the nitrogen interstitial as obtained through \textit{ab initio} calculations. To retrace steps 2 and 3, when the nitrogen atom is at a substitutional site and the interstitial and vacancy are separated from each other, the total energy of the system is given by the sum of the individual formation energies of the respective species. This amounts to 6.6 eV. In step 2, the nitrogen substitutional atom and the self-interstitial react to form the nitrogen interstitial \( N_i \). Therefore, in this step, the system energy will decrease by an amount equal to the binding energy of the nitrogen interstitial. Hence the system energy at the end of step 2 will be the system energy at the end of step 1 minus the binding energy of the nitrogen interstitial. This amounts to 3.1 eV. In step 3, \( N_i \) and \( V \) recombine to form \( N_S \). Since \( N_S \) is the reference state of the system, the system energy at the end of step 3 will be 0. Therefore, in proceeding from step 2 to step 3 the system gains 3.1 eV. Hence, the binding energy for the Frank–Turnbull reaction \( N_i + V = N_S \) is 3.1 eV.

4. UT-MARLOWE, version 5.0, Manual, University of Texas at Austin.
7 P. A. Schultz (private communication).
8 Venezia, MRS Meeting, April 1999.
11 Florida Object Oriented Process Simulator (FLOOPS) manual, 2000 release and references therein. (Manual can be obtained from the website: www.swamp.tec.ufl.edu)