

Amorphous Silicon Dioxide

So far, both electronic and material properties of single crystal silicon have been considered in some detail. In addition, effects of defects and impurities have also been considered. All of these properties are essential to modern solid-state electronics; however, if the characteristics of the semiconductor material itself were all that was important, silicon would actually present little (if any) advantage over germanium or gallium arsenide. (Indeed, some other semiconductor might very well be better suited from the point of view of carrier mobility, *etc.*) Accordingly, there is another material, quite different from single crystal silicon, which is also of essential importance, *viz.*, amorphous silicon dioxide. Within this context, it is worthwhile to compare the most obvious characteristics of single crystal silicon and amorphous silicon dioxide, *i.e.*, quartz glass: 1) silicon is crystalline, quartz glass is amorphous, 2) silicon conducts heat and electricity reasonably well, quartz glass is a poor conductor of both, 3) silicon is an opaque, metallic appearing material (although it is transparent at infrared wavelengths), quartz glass is very transparent well into the ultraviolet region of the spectrum.

Indeed, the success of silicon solid-state electronics is due, in no small part, to the fact that high quality amorphous silicon dioxide thin films are easily produced by direct oxidation of silicon. Therefore, even though germanium was commercialized earlier than silicon and, moreover, although it has higher intrinsic electron and hole mobilities, because a high quality, chemically stable germanium dioxide (GeO_2) layer cannot be formed on a germanium surface by direct oxidation (germanium monoxide (GeO) sublimates at 710°C) represents a serious limitation. As a consequence, silicon is the material of choice for industrial production of the vast majority of solid-state electronic devices (although germanium and especially silicon-germanium alloys have undergone somewhat of a renaissance in recent years, but generally in combination with silicon). Similar issues also exist for compound semiconductors such as gallium arsenide (GaAs), silicon carbide (SiC), *etc.* Indeed, as a practical matter, a semiconductor material other than silicon will be used only if it has some unique property that silicon does not have. For example, because of higher carrier mobilities GaAs and more recently indium phosphide (InP) have found some commercial use for fabrication of high speed, high frequency devices, such as amplifiers for cell telephones and wireless information networks. In addition, GaAs and other III-V materials are direct band gap semiconductors and, thus, useful for optoelectronic devices such as lasers and light emitting diodes (LEDs), which are applications for which silicon is not well suited. Similarly, silicon carbide may be useful if high temperature operation is required since it has a much larger band gap than silicon. (Diamond also has similar semiconductor characteristics.) Other semiconductors, such as indium antimonide (InSb), cadmium selenide (CdSe), *etc.*, may find use as specialty optical detectors or emitters; however, the production volume remains small and integration level low. Consequently, silicon successfully competes with (*e.g.*, in device speed) or surpasses (*e.g.*, in integration level) all other semiconductor materials for all but a few specific applications. In any case, the silicon/silicon dioxide material system is dominant and is likely to remain so for the foreseeable future. This remains true regardless of any consideration that essentially without exception, all other semiconductor materials are much rarer than silicon and, consequently, inherently more expensive. (However, in practice the cost of the substrate

is generally only a small part of the cost of a finished integrated circuit or other solid-state electronic device.)

Direct oxidation of the surface of a silicon wafer at high temperature in an oxidizing atmosphere is known conventionally as *thermal oxidation*. The resulting thin quartz glass film is known as *thermal oxide*. As observed at the outset, quartz glass is not crystalline, but is amorphous with an open random network structure. This is in distinct contrast with silicon, which, of course, has a very well-defined crystal structure. The fundamental unit of the network structure is the SiO_4 tetrahedron. A diagrammatic representation of an SiO_2 network is shown below:

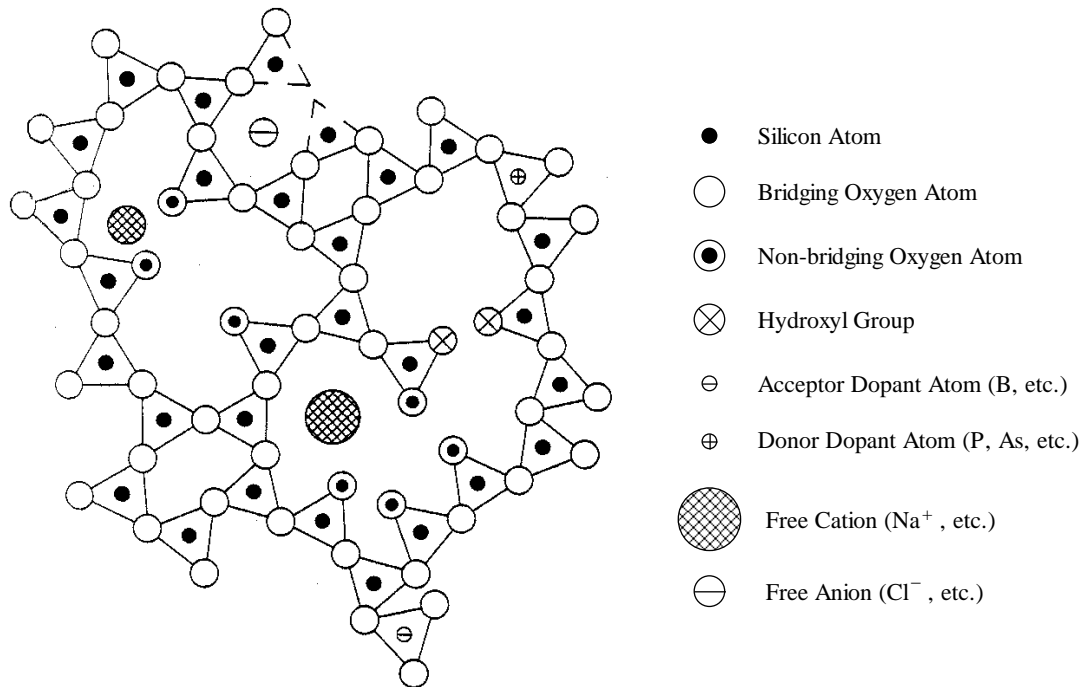


Fig. 29: Diagrammatic representation of quartz glass network structure

(Here, for convenience SiO_4 tetrahedrons are represented two dimensionally as triangles.) Indeed, thermal oxide has characteristics of both a liquid (*e.g.*, short-range order) and a solid (*e.g.*, rigidity and elasticity). Although the network structure of quartz glass is thermodynamically unstable below 1710°C , the rate of *devitrification*, *i.e.*, crystallization, is negligible below 1000°C . Therefore, once formed, thermal oxide is very stable under normal conditions.

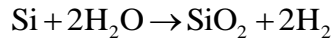
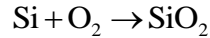
Amorphous silicon dioxide has a well-defined refractive index of 1.46 and density of 2.27 g/cm^3 . In a perfect structure, each SiO_4 tetrahedron is joined to four other tetrahedra, one at each apex. This implies that oxygen atoms must bridge between silicon atoms. Thus, in an ideal structure, each oxygen atom is bonded to two silicon atoms and each silicon atom is bonded to four oxygen atoms (hence, the stoichiometric formula SiO_2). This results in a much less dense structure than single crystal silicon; therefore, the network structure of SiO_2 includes voids of various shape and size. Furthermore, the exact structure of these voids is generally process dependent. Additionally, some of the

tetrahedra in the network may not be attached at all apexes. In this case, the oxygen atom must be bound to some other type of atom since two bonds are required. This is commonly hydrogen resulting in the incorporation of a hydroxyl (OH) group into the network structure. It is also possible for the silicon to become trigonally coordinated with only three oxygen atoms attached. Two of these are attached to other tetrahedra, the third one is unattached, *i.e.*, non-bridging, and is, in principle, doubly bonded to the silicon atom.

For the purposes of integrated circuit fabrication, thermal oxidation is a very effective process. It produces thin films of amorphous SiO_2 having a dense uniform network structure in comparison to other methods of thin film fabrication such as evaporation or chemical vapor deposition (CVD). The material properties of thermal oxide are quite uniform and invariant over time. Furthermore, even though it has an open structure, diffusion rates of many species in amorphous SiO_2 are quite low. Of particular importance are the usual shallow level impurities, B, P, As, Sb, Ga, *etc.* These species typically form oxides that themselves become strongly bound within the network (as illustrated in the preceding figure). Thus, SiO_2 is a very good mask for doping particular regions on the wafer surface. (This will be considered in more detail in later treatment of diffusion and ion implant processes.) Other species, which do not become bound in the network structure, diffuse quite rapidly in SiO_2 . In particular, hydrogen diffuses quite readily as does oxygen, water, and a number of small inorganic anions and cations. All of these species diffuse through the voids in the network structure. (As will become evident, the fact that SiO_2 is permeable to H_2 , O_2 , and H_2O is of essential significance.)

Thermal Oxidation of Clean Silicon

As indicated previously, thermal oxidation of a clean silicon surface in an ambient oxidizing atmosphere is, perhaps, the most fundamental of all integrated circuit fabrication processes. Physically, it is an example of a heterogeneous (gas-solid) chemical reaction. In conventional practice, either dry oxygen or pyrogenic steam is used as an oxidant. (Pyrogenic steam is produced by burning hydrogen and oxygen inside the oxidation furnace.) The two overall reactions are as follows:



Clearly, so-called *dry oxidation* in oxygen produces no gaseous products; however, *wet oxidation* in steam produces hydrogen as a byproduct.

The Deal-Grove Model of Thermal Oxidation

In general, an overall heterogeneous chemical reaction can be separated into several transport and reaction steps. First of all, the gaseous reactant must be transported from the bulk of the ambient gas atmosphere to the substrate surface. Accordingly, the flux of reactant to the substrate surface can be described by a simple mass transport equation:

$$F_1 = h_G(C_G - C_S)$$

Here, F_1 is oxidant flux to the substrate surface, C_G is bulk concentration of oxidant, C_S is the concentration of oxidant in proximity to the wafer surface, and h_G is a linear mass transport coefficient. This expression accounts for depletion effects in the gas phase due to consumption of oxidant by the reaction. Second, oxidant is dissolved in the surface of the thermal oxide film and diffuses to the Si/SiO₂ interface, hence:

$$F_2 = \frac{D}{x}(C_o - C_i)$$

Here, F_2 is oxidant flux diffusing through the growing thermal oxide film, C_o is dissolved oxidant concentration at the oxide surface, C_i is the dissolved oxidant concentration in the oxide at the Si/SiO₂ interface, D is the oxidant diffusivity in thermal oxide, and x is the thermal oxide layer thickness. Third, assuming first order kinetics, the oxidation reaction at the Si/SiO₂ interface corresponds to the expression:

$$F_3 = k_s C_i$$

In this case, F_3 is the oxidant flux (or, more correctly, a pseudo-flux) due to consumption of reactant by the oxidation reaction and k_s is a first order rate constant for the reaction.

Of course, oxidant concentrations, C_s and C_o , cannot be expected to be equal, but rather, to satisfy a heterogeneous distribution equilibrium across the gas-solid interface, *viz.*, Henry's Law:

$$H = \frac{C_o}{C_s}$$

Here, H is a distribution coefficient and is defined in analogy to distribution coefficients associated with crystal growth (except that the heterogeneous phases are gas and solid rather than liquid and solid). Clearly, H is closely related to the equilibrium solubility of the gaseous oxidant species in quartz glass and, naturally, is dependent on temperature and the microstructure of the glass. As has been noted previously, in the case of wet oxidation, a gaseous product, namely hydrogen, is formed. For generality, the diffusion flux of hydrogen back out of the oxide should also be considered since Le Chatelier's Principle implies that any local increase of hydrogen in proximity of the Si/SiO₂ interface must reduce the reaction rate, *i.e.*, favor the back reaction. However, since hydrogen is a small molecule and diffuses rapidly, it does not build up and its effects can be ignored. Clearly, for dry oxidation no gaseous products are formed and preceding expressions are entirely sufficient. Furthermore, assuming that wafer dimensions are much larger than film thickness, a one dimensional picture of thermal oxidation is satisfactory and is illustrated by the following figure:

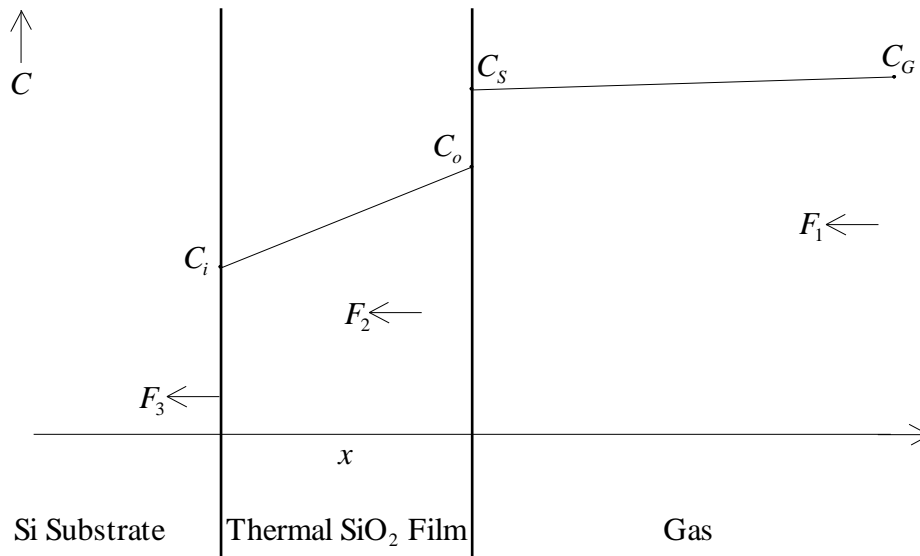


Fig. 30: Diagrammatic representation of the thermal oxidation of a clean silicon surface

Here, oxidant concentration, C , is plotted versus perpendicular dimension relative to the wafer surface. By definition, oxidant concentration within the silicon substrate is negligible.

Thus, assuming conditions of quasi-steady state, *i.e.*, assuming that any transients are small, all fluxes are taken to be equal. Accordingly, if one applies the distribution equilibrium and identifies F_1 as equal to F_3 , one obtains:

$$k_s C_i = h_G \left(C_G - \frac{C_o}{H} \right)$$

Naturally, one solves this expression for C_G :

$$C_G = \frac{k_s C_i}{h_G} + \frac{C_o}{H}$$

Equivalently, one can identify F_2 as equal to F_3 :

$$k_s C_i = \frac{D}{x} (C_o - C_i)$$

Consequently, this expression is solved for C_o :

$$\frac{D}{x} C_o = \left(k_s + \frac{D}{x} \right) C_i$$

$$C_o = \left(1 + \frac{k_s x}{D} \right) C_i$$

The two preceding expressions can be combined by substitution of this equation into the previous formula for C_G :

$$C_G = \frac{k_s C_i}{h_G} + \frac{1}{H} \left(1 + \frac{k_s x}{D} \right) C_i$$

$$C_G = \left[\frac{k_s}{h_G} + \frac{1}{H} + \frac{k_s x}{HD} \right] C_i$$

Inverting this equation to obtain an explicit form for C_i yields the desired result:

$$C_i = H C_G \left[\frac{H k_s}{h_G} + \frac{k_s x}{D} + 1 \right]^{-1}$$

Thus, the oxidant concentration at the Si/SiO₂ interface has been formally related to the concentration of oxidant in the gas phase. Of course, the concentration, C_G , is just fixed by gas pressure inside the furnace.

Naturally, the reaction flux, F_3 , must be proportional to the thermal oxide growth rate; hence, one can write:

$$F_3 = N \frac{dx}{dt} = k_s H C_G \left[\frac{H k_s}{h_G} + \frac{k_s x}{D} + 1 \right]^{-1}$$

Here, N is a proportionality constant relating the number of oxidant species arriving at the interface per unit area to the thickness of SiO_2 grown on that same area if all oxidant species react with the substrate. Of course, N is determined directly by consideration of the reaction stoichiometry and the density of the thermal oxide film. Clearly, this first order differential equation is easily integrated to give:

$$x^2 + 2D \left(\frac{1}{k_s} + \frac{H}{h_G} \right) x = \frac{2DHC_G}{N} (t + t_0)$$

Here, t_0 represents an initial condition, which in principle corresponds to some pre-existing thermal oxide layer thickness of x_0 ; therefore, one has:

$$t_0 = \frac{N}{2DHC_G} \left(x_0^2 + 2D \left(\frac{1}{k_s} + \frac{H}{h_G} \right) x_0 \right)$$

Thus, t_0 is the time necessary to pre-grow a thermal oxide layer of thickness, x_0 , under prevailing conditions, *i.e.*, growth conditions defined by current values of h_G , C_G , D , k_s , H , and N . Of course, in actual processing, the pre-existing oxide layer may be grown under different conditions; however, the properties of thermal oxide are sufficiently uniform so that only the thickness, x_0 , is relevant to subsequent processing. Clearly, if x_0 equals 0, then t_0 equals 0.

Rather than expressing t as a function of x , it is desirable to express x as a function of t . This is easily accomplished by means of the quadratic formula:

$$x = D \left[\sqrt{\left(\frac{1}{k_s} + \frac{H}{h_G} \right)^2 + \frac{2HC_G}{ND} (t + t_0)} - \left(\frac{1}{k_s} + \frac{H}{h_G} \right) \right]$$

Thus, one obtains the general relationship between thermal oxide film thickness and growth time characteristic of the Deal-Grove model. From this formula, two important asymptotic expressions can be obtained. The first of these corresponds to the limit that t tends toward ∞ . In this case, only the second term within the radical remains significant, hence:

$$x = \sqrt{\frac{2DHC_G}{N} t}$$

This defines the so-called *parabolic growth regime*. The second form is obtained if $t + t_0$ vanishes. In this case, one expands the radical as a Taylor series from which it follows that:

$$x = \frac{HC_G}{N \left(\frac{1}{k_s} + \frac{H}{h_G} \right)} (t + t_0)$$

This defines the so-called *linear growth regime*. Clearly, unless t_0 is small, *i.e.*, the initial oxide thickness is very small or absent, the linear growth regime cannot be realized.

Physically, the parabolic growth regime corresponds to the classical case of a *diffusion limited process* for which the rate limiting step is diffusion of oxidant through a relatively thick oxide film. Conversely, the linear growth regime corresponds to the case of a *reaction limited process* for which the rate limiting step is the interfacial reaction between oxidant species and the silicon substrate. In passing, one observes that another limiting regime, that of a *mass transport limited process*, is possible in principle. This situation would occur if oxidant became depleted in the gas phase in close proximity to the substrate surface. Clearly, this requires very rapid consumption of oxidant species by the oxidation process. However, in practice, oxidant transport in the gas phase is much more rapid than either diffusion of oxidant through the growing oxide film or the interfacial reaction itself. Therefore, a mass transport limited regime is never realized in conventional thermal oxidation processes, *i.e.*, for practical purposes, the coefficient, h_G , can be treated as indefinitely large.

In practice, one does not usually know (or care to know) all of the values of the various transport, equilibrium, and reaction rate coefficients. However, they can be collected into two aggregate rate constants, A and B , defined as follows:

$$A = 2D \left(\frac{1}{k_s} + \frac{H}{h_G} \right) \quad ; \quad B = \frac{2DHC_G}{N}$$

Therefore, in terms of A and B , the previous results can be recast as follows:

$$x^2 + Ax = B(t + t_0) \quad ; \quad t_0 = \frac{1}{B}(x_0^2 + Ax_0)$$

$$x = \frac{\sqrt{A^2 + 4B(t + t_0)} - A}{2}$$

Similarly, the parabolic and linear limiting forms are:

$$x = \sqrt{Bt} \quad ; \quad x = \frac{B}{A}(t + t_0)$$

By convention, B is known as the *parabolic rate constant* and B/A as *the linear rate constant*. Values for A and B (and or B and B/A) have been determined over a variety of conditions. Using these values, it is found that the Deal-Grove model describes thermal

oxidation very well over a wide temperature range, viz., 700°-1300°C. This is illustrated by the following figure:

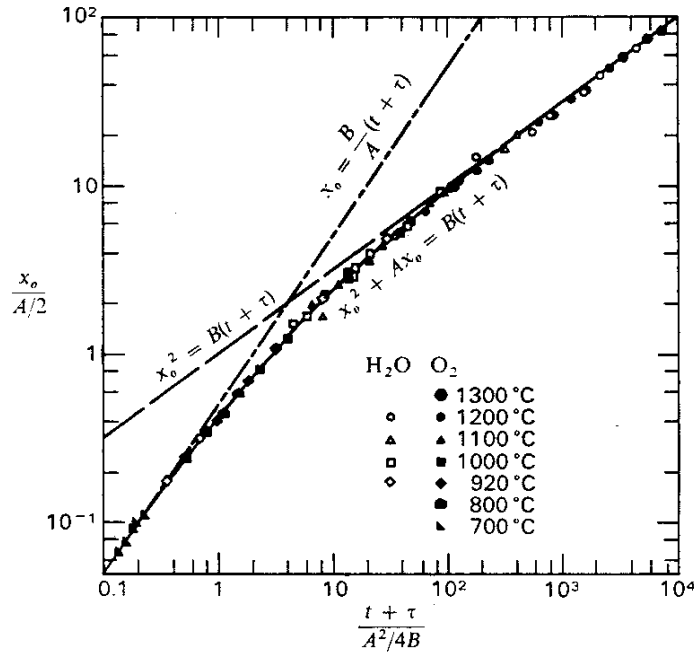


Fig. 31: Scaled thickness vs time for thermal oxidation (solid curve: Deal-Grove model; broken curves: linear and parabolic limits)

In practice, any conventional oxidation process used for integrated circuit fabrication will almost certainly be included within this temperature range.

Temperature Dependence of Oxidation Rate

Although, the Deal-Grove model is applicable over a wide range of temperatures, oxidation rate is strongly temperature dependent. As might be expected, this dependence has a classical *Arrhenius form*:

$$\kappa = \kappa_o \exp\left(-\frac{E_a}{kT}\right)$$

Here, κ can be identified as either the parabolic or linear rate constant. By definition, any reaction or process that is characterized by an Arrhenius form is said to be thermally activated and, accordingly, E_a is identified as *activation energy*. To understand the precise meaning of E_a , one should think of any process (chemical reaction, diffusion, etc.) as a transition from some stable *reactant* state to a stable *product* state. In order, for both the reactant and the product state to be stable, the system must pass through some unstable “high energy” *transition state* (conventionally indicated by the symbol, ‡) during the process. Clearly, the transition state provides a “barrier” to free conversion of

reactants into products. Such a “chemical” description of thermally activated processes can be represented pictorially as follows:

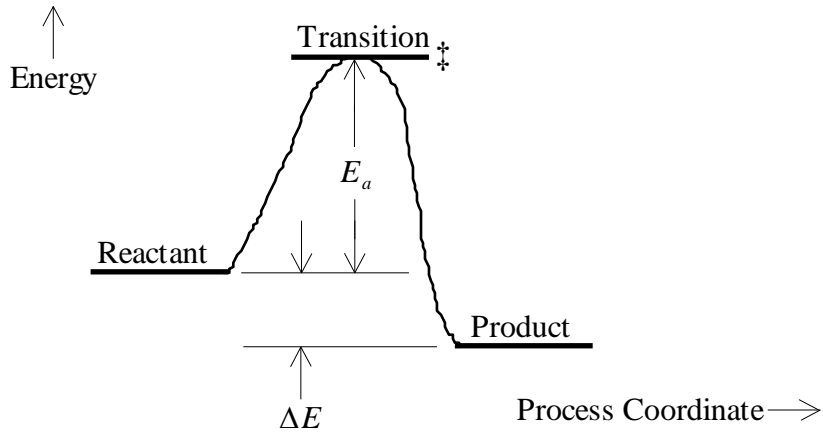


Fig. 32: Energetic relationship of transition, reactant, and product states

The horizontal dimension is defined as *process coordinate*, which is just a symbolic representation of the aggregate dynamics of the process. The vertical dimension represents thermodynamic internal energy. Thus, ΔE is the thermodynamic internal energy change for the overall process and E_a is the energy change taken between the reactant state and the transition state. (Thus, E_a can be thought of as a kind of formation energy for the transition state.) Clearly, because the energy of the transition state is higher than either the reactant or product states, it forms an *energy barrier* for the process and reactant and product states tend to be stable once they are formed. However, if thermal fluctuations randomly generate some of the transition state from the reactant state, then the product state is easily formed. Of course, the reverse process can also occur.

Digressing briefly, the case for which the product state has a lower internal energy than the reactant state is called an *exoenergetic* process since energy is released to the environment during the process. This situation is illustrated by the preceding figure. Conversely, if the product state has a higher internal energy than the reactant state, energy must be absorbed and the process is called *endoenergetic*. Obviously, the reverse of an exoenergetic process must be endoenergetic and vice-versa. (The terms exoenergetic and endoenergetic are analogous to the more common terms, *exothermic* and *endothermic*, except that they refer specifically to internal energy, rather than enthalpy.) Obviously, in the endoenergetic case, the activation energy must be larger than the internal energy change, ΔE , since it must be the sum of the activation energy for the reverse exoenergetic process, E_a , and ΔE . Of course, thermally activated processes for which reactant and product states are of equal internal energy, *e.g.*, diffusion, can be called *aenergetic*. However, the activation energy for an aenergetic process does not necessarily vanish and, clearly, is the same in both “forward” and “reverse” directions.

It is found that oxidation of clean silicon is a thermally activated, exoenergetic, *i.e.*, exothermic, process. Therefore, Arrhenius forms can be expected to represent temperature dependence of the linear and parabolic rate constants satisfactorily.

However, before providing specific values for activation energies and pre-exponential factors, it is important to note that oxidation rate is also found to depend on the orientation of the wafer surface, that is to say, that one finds that oxidation rates differ on [100] and [111] surfaces. Various models have been formulated to explain orientation dependence, however, in all of these it is attributed to differences in surface atom concentration and specific activation energy (derived from steric effects, *etc.*) For substrates commonly used in integrated circuit fabrication, one invariably finds that [111] wafers oxidize faster than [100] wafers under the same conditions. Since orientation is a property of the substrate only and does not affect the structure of the oxide once it is grown, *i.e.*, thermal oxide grown on [111] substrates is essentially identical to oxide grown on [100] substrates, one expects that orientation dependence enters the Deal-Grove model only through the specific rate constant, k_s . Therefore, it is to be expected that only the linear rate constant depends on orientation and that the parabolic rate constant is independent of orientation, as is, indeed, the case. Arrhenius forms for various process conditions and orientations appear in the following table:

Process	B/A for [100]	B/A for [111]	B
Dry Oxidation	$1.03(10^3)e^{-2.00/kT}$	$1.73(10^3)e^{-2.00/kT}$	$0.214e^{-1.23/kT}$
Steam Oxidation	$2.70(10^4)e^{-2.05/kT}$	$4.53(10^4)e^{-2.05/kT}$	$0.107e^{-0.79/kT}$

Note: Activation energies are in eV's, B/A is in $\mu\text{m}/\text{sec}$, B is in $\mu\text{m}^2/\text{sec}$

Table 2: Arrhenius forms for thermal oxidation rate constants

Clearly, steam oxidation is much faster than dry oxidation. Therefore, steam oxidation is advantageous for the growth of relatively thick oxide layers. These are typically field or isolation oxides, which surround devices and insulate the substrate from overlying wiring, *etc.* However, for oxides, usually thin, that are used as integral parts of devices, such as a gate insulator (or gate oxide), dry oxidation is generally used because it produces a higher quality Si/SiO₂ interface. The quality of this interface is critical for good electrical performance. (In the case of very thin oxides, this distinction breaks down. Indeed, the fabrication of ultrathin oxide layers is currently of great interest.)

Another process variable that is available to change oxidation rate is oxidant pressure. It is evident from the Deal-Grove model that B is proportional to oxidant concentration, C_G . Of course, C_G is just proportional to pressure (or partial pressure) through the usual gas laws. Therefore, both the linear and parabolic rate constants simply scale linearly with pressure. This provides several advantages. First of all, one can grow thick oxides much more rapidly at elevated pressure. However, "thermal budget", *i.e.*, the total exposure of the substrate to elevated temperature, rather than process time itself, is often a more important consideration in practical integrated circuit fabrication. Therefore, it is also advantageous to reduce the thermal budget without adversely affecting process time by lowering temperature and compensating the resulting lowered growth rate by increasing process pressure. An added benefit is that at lower temperature thermally activated defect generation is also reduced. Finally, for very thin oxides, the growth rate at normal atmospheric pressure may be too fast for adequate process control of oxide

thickness. In this case, pressure (or partial pressure) can be reduced to a sub-atmospheric value to lower the growth rate and provide a more controllable process.

Deviations from the Deal-Grove Model

Before proceeding further, it is necessary to observe that there is one important deviation from the Deal-Grove model. In particular, the Deal-Grove model is unable to explain the kinetics associated with very thin oxide growth in dry oxygen. Specifically, a very rapid initial growth phase is observed. After this initial phase, the process follows the Deal-Grove model. Empirically, it is found that for dry oxide films of thickness greater than about 20 nm, the Deal-Grove model can be applied by assuming an initial fictitious oxide thickness of about this thickness, *i.e.*, one assumes that this initial film grows so rapidly that it can be taken as an initial condition for the Deal-Grove model. Of course, fabrication by dry oxidation of thin SiO₂ films having thickness on the order of 20 nm or less requires careful experimental characterization of growth kinetics in any initial growth regime. In contrast, a rapid initial growth phase is not observed for wet oxidation and the Deal-Grove model can be used to describe all stages of the process. (Wet oxidation is generally used for thick oxides anyway; however, recently there has been a renewed interest in using wet oxidation at very low temperature for fabrication of very thin oxide layers.)

The initial rapid growth phase in dry oxygen may be explained by observing that since no pre-existing oxide layer is present, the oxygen concentration is initially very high at the substrate surface. In this case, it is plausible that oxygen dissolves appreciably in the substrate itself to create a thin oxygen rich surface layer or *oxygen-diffused zone*. Of course, one expects the solubility of oxygen in silicon to be much less than in silicon dioxide, however, since little or no surface oxide is present, the concentration may still become significant. Thus, one can regard oxidation in the oxygen-diffused zone as more of a volume reaction than a surface reaction, *i.e.*, oxidation is occurring at an appreciable rate throughout the whole thickness of the oxygen-diffused zone. Therefore, since the whole thickness of the oxygen-diffused zone is rapidly converted to oxide, the apparent surface oxidation rate is “abnormally” high. Of course, once an initial oxide layer of sufficient thickness is formed, the concentration of oxygen at the interface falls and the oxygen-diffused zone disappears. Obviously, this must correspond to the onset of Deal-Grove kinetics. Alternatively, the initial rapid growth phase might be a consequence of deviation of the surface reaction kinetics from first order when the oxygen concentration is very high. (This is not necessarily inconsistent with the existence of an oxygen-diffused zone.)

For wet oxidation two observations can be made. First of all, water does not appear to dissolve or diffuse appreciably into silicon, hence, if any diffused zone is formed, it must be very thin. Secondly, the surface reaction in wet oxidation is inherently more rapid than in dry oxidation, which also serves to reduce the relative importance of any initial oxidation phase.

Oxidation Induced Defects

Under some conditions, thermal oxidation can produce *oxidation induced stacking faults* aligned with [111] planes. These stacking faults are typically extrinsic and, of course, are bounded by dislocations. Moreover, it is thought that oxidation induced stacking faults occur because thermal oxidation generates interstitial defects. Indeed, during normal oxidation, about one out of a thousand silicon atoms at the interface does not become incorporated into the growing oxide layer, but instead, diffuses back into the silicon lattice as an interstitial defect. Clearly, if the oxide growth rate is sufficiently high, these interstitials cannot come to equilibrium with vacancies, but rather “condense” as extrinsic stacking faults. (One recalls that an extrinsic stacking fault can be regarded as insertion of an extra plane of atoms.)

Within this context, it is found that stacking fault growth is thermally activated and is characterized by an Arrhenius form up to about $\sim 1200^\circ\text{C}$. Above this temperature, stacking faults no longer grow larger, but shrink (a process called “retrogrowth”). This behavior can be understood if one recalls that the melting point of silicon is nominally 1414°C . Naturally, at a temperature near the melting point, one expects that lattice defects will be rapidly “annealed out” due to high atomic mobility. Furthermore, growth of oxidation induced stacking faults is found to be dependent on substrate orientation, majority carrier type, and defects. Accordingly, the growth rate of stacking faults is greater for [100] than for [111] substrates and stacking fault density is greater on *n*-type rather than *p*-type. Generally, the distribution of surface nucleated stacking fault lengths is very narrow. Furthermore, it is found that even for thick oxides, stacking fault growth is almost completely suppressed if oxidation temperature is reduced below 950°C . However, if it is desirable to oxidize silicon substrates at higher temperature (perhaps to obtain a good Si/SiO₂ interface), subsequent high temperature annealing in an inert ambient can substantially reduce stacking faults.

Empirical observations indicate that for oxidation at the same temperature and time, the average length of oxidation induced stacking faults is greater for wet oxidation than for dry oxidation. This suggests that stacking fault length depends on oxidation rate as, indeed, is found to be the case. (However, if the same thickness of oxide is grown at a given temperature, wet oxidation will produce shorter stacking faults than dry oxidation since the oxidation time is much shorter.) Within this context, an empirical formula has been proposed to characterize dependence of stacking fault growth rate on oxidation rate as follows:

$$\frac{dl}{dt} = K_1 R_{ox}^n - K_2$$

Here, l is stacking fault length, R_{ox} is oxidation rate, and n , K_1 , and K_2 are constant parameters. The exponent, n , is found to have a value of about 0.4. Therefore, dependence of stacking fault growth rate on oxidation rate is sub-linear and as indicated above, at some fixed temperature and oxide thickness, smaller stacking faults will be formed by a higher growth rate oxidation process. This suggests that high pressure oxidation should be useful for reduction of oxidation induced defects.

Kinetic Effects of Defects, Dopants, Chlorine, *etc.*

Defects in the silicon substrate are invariably associated with disruption in lattice bonding. Therefore, since lattice bonds are already broken, one expects that both wet and dry oxidation rates should be increased by the presence of defects. Although difficult to characterize quantitatively, this phenomenon is frequently observed. Furthermore, the effect of shallow level dopant concentration on oxidation rate is an important consideration for silicon integrated circuit fabrication. Indeed, it is well known that high dopant levels ($>10^{18} \text{ cm}^{-3}$) tend to accelerate both dry and wet thermal oxidation. The underlying cause of this is imperfectly understood; however, it may be a consequence of changes within the oxide structure itself due to the presence of dopants or enhanced defect generation within the substrate. Of course, any effect on oxide structure, hence, on oxidant diffusion coefficient, can be expected to change the parabolic rate constant. Accordingly, it has long been known that boron preferentially segregates into the oxide; therefore, since boron is trivalent rather than tetravalent, one may plausibly suppose that the oxide network structure should be weakened and oxidant diffusion enhanced. Conversely, defect generation within the substrate increases the surface reaction rate, but should not substantially affect oxide structure. Accordingly, the linear rate constant should be affected, but, not the parabolic rate. This evidently is the case for phosphorus and arsenic, which do not preferentially segregate into oxide. Clearly, oxidation of doped or defected silicon can be expected to deviate substantially from the Deal-Grove model. (Interaction between oxidation and dopant diffusion will be treated in more detail later.)

Chlorine (Cl_2) and chlorine containing species (*e.g.*, hydrogen chloride (HCl), trichloroethane (TCA), *etc.*) can be added to an oxidizing ambient with beneficial effects. Empirically, the presence of chlorine is found to improve the quality of the Si/SiO₂ interface. This may be partially a consequence of increased volatilization of metallic impurities. In addition, an increase in both linear and parabolic rate constants is also observed with the addition of chlorine or chlorine containing species to the oxidizing ambient. This may be due to two factors: First, enhanced vacancy generation at the Si/SiO₂ interface due to direct reaction of chlorine with silicon to produce volatile silicon chlorides, which allows more silicon migration to the surface or oxygen entrapment at the surface. Both of these effects should serve to enhance the rate. Second, chlorine incorporation into the oxide opens and expands the network structure resulting in an increase in the oxidant diffusion coefficient.