



FILE COPY

UNCLASSIFIED
EXCLUDED FROM AUTOMATIC
DOWNGRADING AND
DECLASSIFICATION

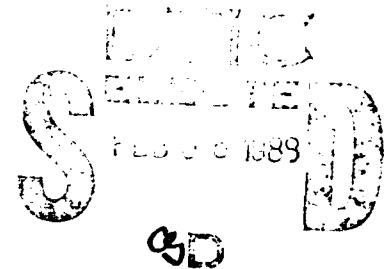
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

VLSI PUBLICATIONS

VLSI Memo No. 88-489
November 1988

BOUNDARY ELEMENT TECHNIQUES FOR MODELING THERMAL OXIDATION OF SILICON

Thye-Lai Tung



Abstract

This thesis advances boundary element techniques to model thermal oxidation of silicon in two dimensions. At temperatures encountered in thermal oxidation, silicon dioxide flows viscoelastically. A reduced-dimension, generalized boundary element method for modeling such a problem has been developed. With a Laplace transform technique, a viscoelastic kernel function is derived from Kelvin's solution, which is the fundamental solution to linear elasticity. Constant-velocity loading is chosen to operate with a wide range of stress relaxation times. This scheme is capable of replacing boundary element methods developed for slow viscous flow and elastic deformation. The oxidant diffusion problem is solved using a standard potential method for Laplace problems. Generated by oxide growth, stress affects both oxidant diffusion and oxide flow. In particular, it changes the diffusivity of oxidants and viscosity of oxide, rendering the diffusion and flow problems nonhomogeneous. Domain solutions are needed for both processes. A unified initial stress/built-in field formulation has been developed to account for the nonlinear effects, using interior cells that are placed where stress is significant. The interior solutions are realized with an interfacial source method, whereby an area integral for a cell is transformed to a line integral on the perimeter of the cell. It has been found that kernel functions based on Kelvin's solution are deficient in modeling incompressible materials with a "hole". A correction method that uses a source term placed at the center of the hole has been implemented to overcome the numerical problem.

Acknowledgements

Submitted to the Department of Electrical Engineering and Computer Science, MIT, June 1988, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. This work was sponsored in part by the Defense Advanced Research Projects Agency under contract number N00014-85-K-0213 and the Microelectronics and Computer Technology Corporation.

Author Information

Tung, current address: Intel Corporation, MS SC9-35, P.O. Box 58125, Santa Clara, CA 95052.

Copyright© 1988 MIT. Memos in this series are for use inside MIT and are not considered to be published merely by virtue of appearing in this series. This copy is for private circulation only and may not be further copied or distributed, except for government purposes, if the paper acknowledges U. S. Government sponsorship. References to this work should be either to the published version, if any, or in the form "private communication." For information about the ideas expressed herein, contact the author directly. For information about this series, contact Microsystems Research Center, Room 39-321, MIT, Cambridge, MA 02139; (617) 253-8138.

Chairman, Departmental Committee on Graduate Students

BOUNDARY ELEMENT TECHNIQUES FOR MODELING THERMAL OXIDATION OF SILICON

by

THYE-LAI TUNG

Submitted to the Department of Electrical Engineering
and Computer Science on June 17, 1988 in partial fulfillment
of the requirements for the Degree of Doctor of Philosophy
in Electrical Engineering and Computer Science

ABSTRACT

This thesis advances boundary element techniques to model thermal oxidation of silicon in two dimensions. At temperatures encountered in thermal oxidation, silicon dioxide flows viscoelastically. A reduced-dimension, generalized boundary element method for modeling such a problem has been developed. With a Laplace transform technique, a viscoelastic kernel function is derived from Kelvin's solution, which is the fundamental solution to linear elasticity. Constant-velocity loading is chosen to operate with a wide range of stress relaxation times. This scheme is capable of replacing boundary element methods developed for slow viscous flow and elastic deformation. The oxidant diffusion problem is solved using a standard potential method for Laplace problems. Generated by oxide growth, stress affects both oxidant diffusion and oxide flow. In particular, it changes the diffusivity of oxidants and viscosity of oxide, rendering the diffusion and flow problems nonhomogeneous. Domain solutions are needed for both processes. A unified initial stress/built-in field formulation has been developed to account for the nonlinear effects, using interior cells that are placed where stress is significant. The interior solutions are realized with an interfacial source method, whereby an area integral for a cell is transformed to a line integral on the perimeter of the cell. It has been found that kernel functions based on Kelvin's solution are deficient in modeling incompressible materials with a "hole". A correction method that uses a source term placed at the center of the hole has been implemented to overcome the numerical problem.

Thesis Supervisor: Dr. Dimitri A. Antoniadis

Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

I wish to express my appreciation to Prof. Dimitri A. Antoniadis for his guidance, encouragement and support throughout the course of this work. I am grateful to Prof. Jerome Connor for his valuable discussions on boundary element techniques. His enthusiasm has always helped to push me along in my work. I would also like to thank Prof. Jesus del Alamo for his willingness to try to understand my thesis, which deals with mechanics and integral equations mostly.

I am grateful to Connor Rafferty, Len Borucki, and Pantas Sutardja - workers in the same field - for valuable discussions and exchange of information. Discussions on oxidation issues with John Hui, D.-B. Kao, and K. Tanaguchi are duly acknowledged.

This work would not have materialized without the special computer resources provided by various groups at MIT. Many formulae are derived and analyzed with Macsyma, a symbolic algebraic manipulator. The conversion of computer codes from Pascal to C was done rather painlessly with a translator provided by R.C. Huang.

Contents

Abstract	2
Acknowledgment	3
Table of Contents	4
List of Figures	6
List of Tables	8
1 Introduction	10
1.1 Survey of Oxidation Techniques	10
1.2 Stress Problems and Measurements	14
1.3 Thesis Objective	15
1.3.1 Modeling Considerations	17
1.3.2 History of Finite Element Approaches	18
1.3.3 Boundary Element Formulations	19
1.3.4 Other BEM Developments	20
1.4 Report Outline	20
2 Thermal Oxidation Process	22
2.1 Diffusion Mechanism	23
2.1.1 One-Dimensional Diffusion	23
2.1.2 Oxidant Diffusion in Two Dimensions	27
2.1.3 Effects of Stress on Oxidant Transport	28
2.2 Oxide Motion	29
2.2.1 Material Properties	30
2.2.2 Mechanics of Solids and Fluids	32
2.2.3 Stress and Strain Tensors	34
2.2.4 Maximum Shear Stress	36
2.2.5 Linear Elasticity	37

2.2.6	Two-Dimensional Plane Strain	39
2.2.7	Incompressible Viscous Flow	40
2.2.8	Viscoelastic Deformation	41
2.2.9	Nonlinear Effects of Stress	43
3	Boundary Element Formulations	47
3.1	Laplace Problems	48
3.1.1	Indirect Method	51
3.1.2	Direct Method	54
3.1.3	Periodic Kernels	55
3.1.4	Source Formulation for Oxidant Diffusion	56
3.2	Biharmonic Problems	58
3.2.1	Biharmonic Kernel Functions	59
3.2.2	Generic BEM for Oxide Motion	61
3.2.3	Incompressible Slow Viscous Flow	63
3.2.4	Elastostatic Deformation	64
3.2.5	Elastostatic Kernels	65
3.3	Laplace Transform and Viscoelastic BEM	66
3.3.1	Laplace Transform	66
3.3.2	Viscoelastic Kernels	67
3.4	Modeling Nonlinear Bulk Behaviors	72
3.4.1	Viscoplastic BEM	73
3.5	Nonlinear Diffusion BEM	77
3.6	Nonlinear Viscoelastic BEM	78
3.7	Incompressible Materials	79
4	Numerical Solutions	83
4.1	Boundary and Domain Segmentation	84
4.1.1	Classes of geometries	86
4.1.2	Interior-Cell Generation	89
4.2	Evaluation of Integrals	90
4.2.1	Numerical Quadratures	91
4.2.2	Analytical Integrations	94
4.2.3	Line Integrals Over Collocation Points	96
4.2.4	General Analytical Line Integrals	97
4.3	Matrix solutions	97
4.3.1	Matrix System for Diffusion	98
4.3.2	Oxide Flow	99
4.3.3	Multiple Domains	99
4.3.4	General Domain contributions	101

4.4	Computation for Nonhomogeneous Domain	101
4.4.1	Dipole and Couplet Approximations	102
4.4.2	Interfacial Source Formulation	105
4.5	Nonlinear Solutions	108
4.5.1	Scalar Iterative Methods	109
4.5.2	Multi-Variable Solutions	111
4.5.3	Determination of Relaxation Parameters	113
4.5.4	Iteration on Reaction Rate	114
5	Simulations	120
5.1	Viscoelastic Stress Relaxation	120
5.2	Nonhomogenous Bulk Parameters	124
5.3	Modeling Kao's Experiments	128
5.3.1	Arbitrary Scaling of Activation Volumes	129
6	Conclusions	132
6.1	Future BEM in Thermal Oxidation	133
6.2	Proposed Experimental Studies	134
A	Miscellaneous Formulations	135
A.1	Principal Stresses	135
A.2	Stress and Strain Transformation	136
A.3	Special Analytical Line Integrals	136
A.4	General Analytical Line Integrals	137
	Bibliography	139

List of Figures

1.1	Cross Section of a MOSFET Structure	11
1.2	Sidewall Masked Isolation Structure (SWAMI)	12
1.3	Trench Capacitor Cell	13
1.4	Schematic of a Structure with Dislocations	14
1.5	Experimental Data for Cylindrical Structures	16
2.1	One-Dimensional Oxidation System	24
2.2	Growth of Silicon Dioxide	30
2.3	Stress Tensor Diagram	34
2.4	Two-Dimensional Elasticity	40
2.5	Nonlinear Elasticity and Plasticity	44
3.1	Source Distribution Technique for Potential Problems	52
3.2	Vector Potential Formulation	62
3.3	Elastoplasticity	74
3.4	Modeling Nonlinear Elasticity	76
3.5	Radial loading forces	80
3.6	Source of incompressible singularity	81
4.1	Boundary Segmentation Techniques	85
4.2	Source distributions	86
4.3	Geometry Types	88
4.4	Domain Partition	90
4.5	Dipole Sources	102
4.6	Couplet sources	103
4.7	Conversion of Dipole Area Integral into Contour Integral	104
4.8	One-Dimensional Nonhomogeneous Diffusion	106
4.9	Boundary Sources.	108
5.1	Semi-Recessed LOCOS.	121
5.2	Simulation of LOCOS with 0.1 μ m Nitride.	123
5.3	Relaxation Time - Stress Curves.	125

5.4	Modeling Stress Effect on Viscosity	126
5.5	Effect of Stress on Diffusivity	127
5.6	Simulations of Oxide Growth on Cylindrical Structures	130

List of Tables

2.1	Coefficients for Linear and Parabolic Rates.	26
2.2	Equilibrium viscosities for various silica glasses	31
2.3	Elastic properties of oxide, nitride, and silicon.	39
2.4	Viscoelastic Models	42
3.1	Laplace Transform Table	68
4.1	Newton-Cotes Coefficients	93
4.2	Gaussian Quadrature Coefficients	93

Chapter 1

Introduction

Thermal oxidation is a principal process in the fabrication of silicon integrated circuits (ICs). In this process, silicon wafers are exposed to oxidizing gases at elevated temperatures to produce silicon dioxide on the surface of the wafers. Because of its many desirable properties, such as low surface states, electrical insulation, and masking ability against certain dopants, thermal oxide is used in many applications, ranging from gate insulation for metal-oxide-semiconductor field-effect transistors (MOSFETs) to dielectric isolation between devices. Together with some other processing characteristics, thermal oxidation currently makes silicon the most widely-used material for solid state electronic devices.

1.1 Survey of Oxidation Techniques

Oxidation on a silicon wafer can be localized by patterning a layer of masking material such as silicon nitride in regions where no oxide is desired. Most MOSFET ICs are produced with the local oxidation of silicon (LOCOS) technique. A typical MOSFET structure is shown in Fig. 1.1 [1]. To avoid excessive stress exerted on the silicon substrate during the field oxidation process, a thin layer of pad oxide (relief oxide) is typically sandwiched between the silicon nitride mask and the substrate. Unfortu-

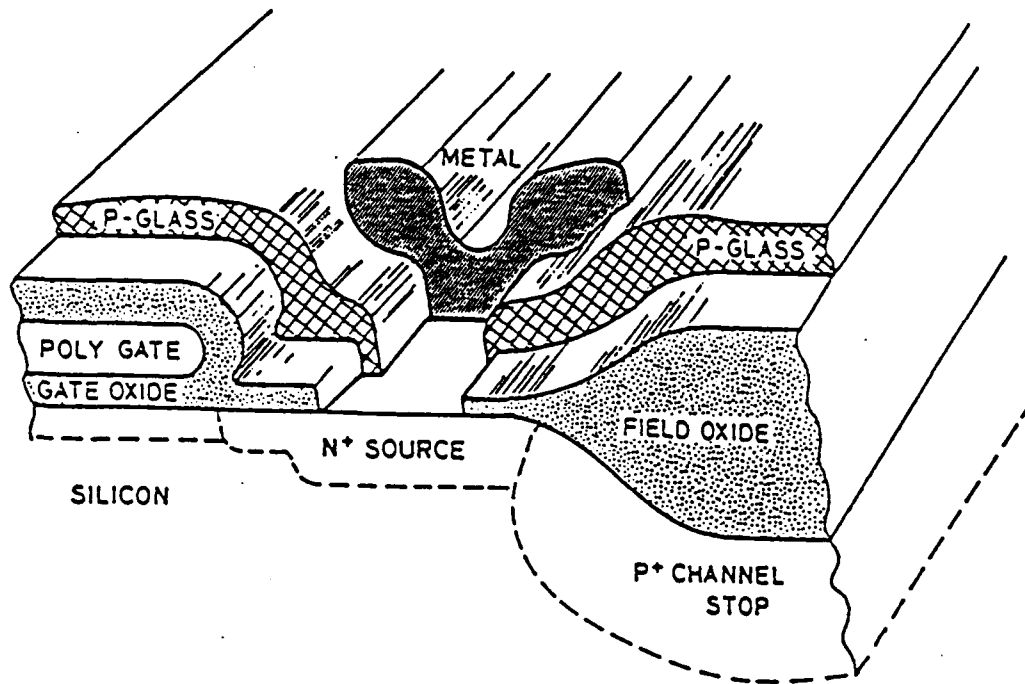
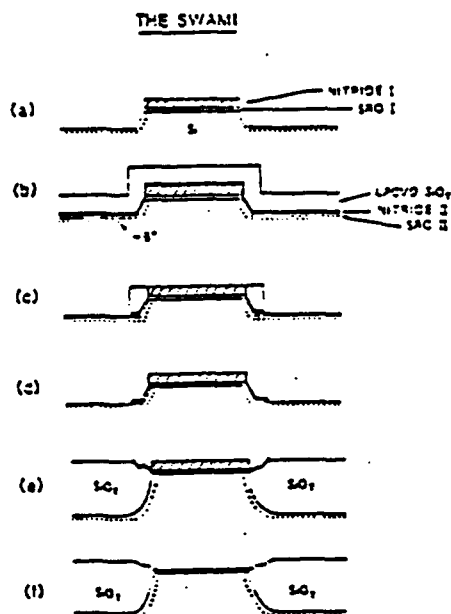


Figure 1.1: Cross Section of a MOSFET Structure. Only source and gate of the device is shown. The field oxide region that becomes narrower as it gets closer to the source is commonly referred to as the bird's beak.

nately, the pad oxide also promotes lateral oxidation at the edge of the nitride mask. As a result, the termination region of the field oxide, commonly referred to as the bird's beak, is widened. Due to physical constraints and processing considerations, the bird's beak may not be made arbitrarily narrow without introducing serious side-effects.

In order to increase IC system complexity and functionality, devices are scaled down to improve packing density and operation speed. Consequently the bird's beak becomes a critical issue as it takes up valuable real estate. Its impact on device characteristics also becomes equally important as, for a small device, its encroachment into the active area has a large effect on the local dopants concentration and on the fringing electric field. Thus an accurate modeling of the formation of the bird's beak, which is two-dimensional in geometry, will help both process design and device modeling. One-dimensional process simulation programs cannot perform such a task;



The major processing steps of the SWAMI.

- (a). After island patterning and etching.
- (b). After nitride II and LPCVD oxide deposition.
- (c). After plasma oxide/nitride etching.
- (d). Island structure before field oxidation.
- (e). After field oxidation.
- (f). Final isolation structure.

Figure 1.2: Sidewall Masked Isolation Structure (SWAMI). Basic processing steps and evolution of the oxide shape.

two-dimensional simulators must be used.

One technique that does not suffer from the formation of the bird's beak is the sidewall masked isolation (SWAMI) [2]. To create a sloped trench in the silicon substrate, it utilized an anisotropic wet chemical etch that attacks in certain preferred crystallographic directions. Then, a few more process steps are involved to deposit and pattern silicon nitride layers to cover all surfaces except the bottom of the trench. After the field-oxidation step, the trench is filled with oxide. The key steps are outlined in Fig. 1.2. Good planarity is maintained, as it can be seen. A bird's beak is formed near the bottom of the trench, but it does not intrude into the active device area, which is located at the top.

Trench isolation is emerging as an important isolation technique for high performance circuitry. This process uses plasma etching to form trenches in the silicon

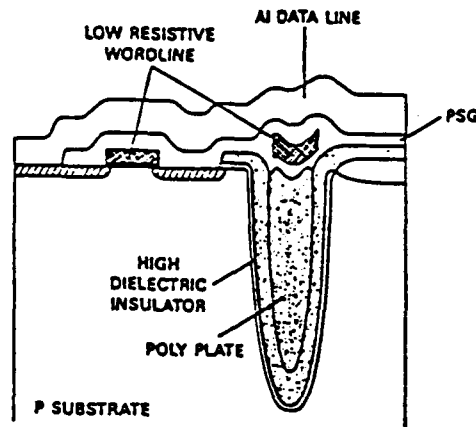


Figure 1.3: Trench Capacitor Cell. After trench is created by plasma etching, a thermal oxide layer is grown. Then a layer of dielectric (such as silicon nitride) is deposited. The hole is filled with polysilicon to serve as a capacitor plate.

substrate. Due to directional selectivity of this etching technique, a high aspect ratio of trench depth to opening width can be achieved, as shown in Fig. 1.3 [3]. For a small substrate real estate, the trench wall can provide a large surface area suitable for use as a storage capacitor. After thermal oxide is grown and some dielectric material is deposited, the trench is filled with chemical-vapor deposition (CVD) polysilicon to serve as one plate for the capacitor. The substrate acts as the other plate. The trench isolation is used extensively in high-density random-access memory (RAM) such as one megabit dynamic RAM. This structure is also suitable as a barrier in device isolation and is being used in bipolar circuits.

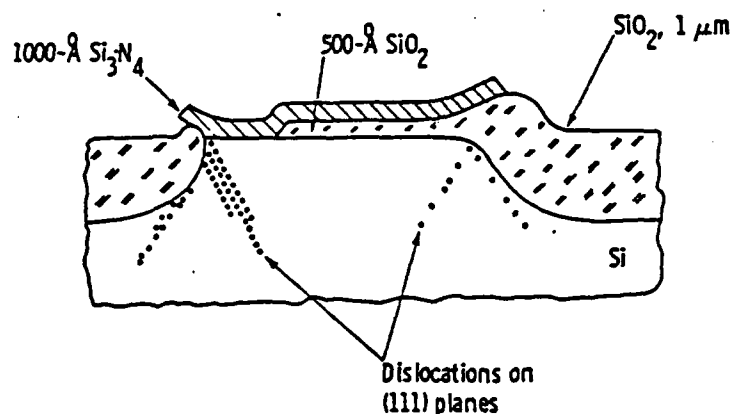


Figure 1.4: Schematic of a Structure with Dislocations. The slips are formed on (111) planes. The lower defects density on the right-hand side is due to stress relief provided by the pad oxide.

1.2 Stress Problems and Measurements

In the LOCOS process, the function of the pad oxide is to reduce (1) stress due to thermal expansion mismatch between silicon nitride mask and the substrate and (2) stress generated during the oxidation process. When high enough, stress damages the underlying silicon substrate, and severely degrades its electrical properties. Defects usually appear in the form of plane slips on {111} planes, as shown in Fig. 1.4. Stress also affects the oxidation process itself. The phenomenon of oxide thinning has been observed in many structures. The reduction in oxide thickness occurs in regions where stress is significant. Experimental characterizations include gate oxide growth [4] and oxidation of step-shaped silicon structures [5].

Possibly the most comprehensive experiments on quantifying stress effects on ox-

idation were done by Kao [6]. Cylindrical structures of different radii were oxidized at different temperatures. Oxide thickness reductions on both convex and concave surfaces were measured and plotted against the curvatures, as shown in Fig. 1.5. Because cylindrical geometry is simple and smooth, the oxide growth behaviors are more disposed to numerical analysis than other approaches. Kao showed that a model based on incompressible creeping flow of oxide can be fitted to his experimental data, for temperatures as low as 800°C[7].

Although many processing techniques have been tried to reduce stress level, one thing has yet to be accomplished is the measurement of actual stress in local oxidations. Intrinsic stress, which is due to volumetric expansion of oxidized silicon, has been deduced by measuring the degree of bending of silicon wafers *in situ* [8]. The viscoelastic relaxation of stress can be detected with X-ray topography that has a resolution as fine as 1 μm [9]. However, it does not provide stress values. Raman scattering [10,11] and electron channeling pattern [12] are used to measure residual stress in the silicon lattice, after oxidation. The level of stress generated during local oxidations is still a matter of speculation. This is problematic because stress values produced by numerical simulations are scalable, as explained in Chapter 5.

1.3 Thesis Objective

The goal of undertaking this thesis is to advance numerical techniques for modeling thermal oxidation in two dimensions. The boundary element method (BEM), which is also known as the boundary- integral equation method (BIEM), deals with an integral form of the governing equations. It is particularly attractive for certain classes of linear physical problems because it offers reduced dimensionality in the system solutions. Only boundaries are segmented, the enclosed domain does not need any partitioning as in the finite element method (FEM). Hence the simulation setup process is simple.

Tasks of the thesis include the extension of the incompressible viscous flow model,

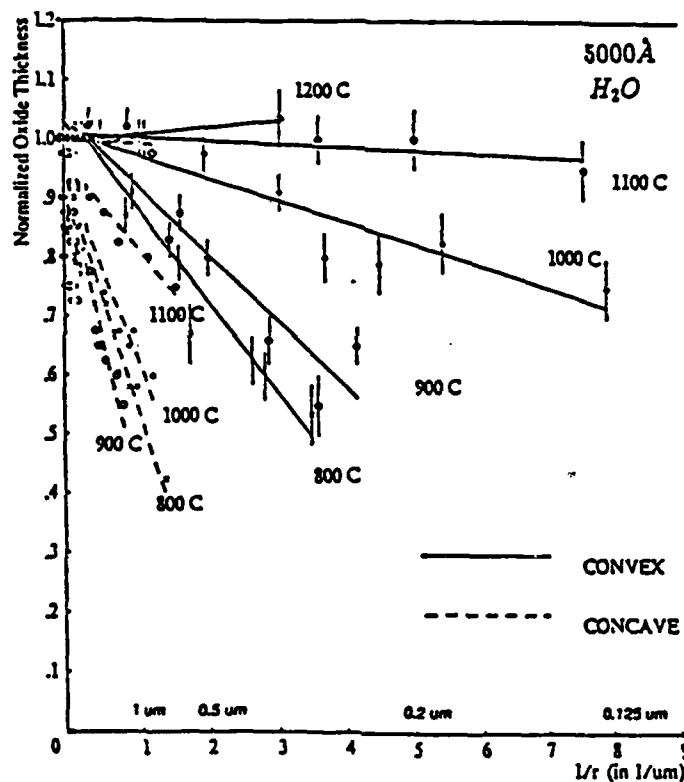


Figure 1.5: Experimental Data for Cylindrical Structures. (a) A cylindrical silicon ring. (b) Top view of the ring. After oxidation, oxide thickness is measured on both the concave and convex surfaces of the cylinder. (c) The vertical axis is the oxide thickness normalized to that of a flat surface. The horizontal axis is the inverse of the radius of the silicon boundary.

developed previously by the author for his Master's thesis, to include models such as elastic deformation and viscoelastic flow. A method based on Laplace transform technique is finally developed. It can deal with such a wide range of shear stress relaxation times that it has supplanted the elastic and viscous flow formulations as the default model in this project.

Another objective is to investigate methods for incorporating nonlinear effects in the boundary element method. Unlike the finite element or finite difference methods, the BEM does not readily model nonlinear problems. Enhancements or approximations are needed to deal with them. Several attempts had made to approximate the nonlinear behaviors so as to avoid domain calculations, but they did not provided satisfactory results. Finally, a unified method utilizing interior cells is developed for both the diffusion and viscoelastic flow problems. The range of nonlinearity it can handle is adequate for oxidation modeling.

1.3.1 Modeling Considerations

The dynamics that sustain thermal oxidation are (1) the supply of oxidizing species to react with silicon to form silicon dioxide, and (2) the motion of the newly created oxide. The main transport mechanism for oxidant is diffusion by which the oxidant moves from the free oxide surface, through the oxide, to the oxide-silicon interface where it reacts with silicon. Due to volume expansion, newly formed oxide partly fills the void left behind by consumed silicon, and partly displaces existing oxide towards the surface. The role of masking materials like silicon nitride is to block the supply of oxidants, thus halting the oxidation process locally.

A simple one-dimensional oxidation model has been around for about twenty years. Developed by Deal and Grove [13], this is a diffusion model that considers only the transport of oxidants, while omitting that of the oxide. Its accuracy in predicting oxide thicknesses has been verified for many oxidation conditions. In one-dimensional

problems, mechanics of oxide motion is not considered because it is uniform everywhere and therefore can be treated as a translation in the space dimensions. This uniformity condition is violated in the bird's beak region, or in any place that is not planar. Under such circumstances, one must consider the non-uniform motion of oxide. Viscous flow or viscoelastic flow are two models that are commonly used for describing the oxide motion.

Research efforts in modeling thermal oxidation of silicon can be divided in two camps according to the numerical technique – the finite element method and the boundary element method. For the oxidant diffusion problem, all formulations are based on Deal-Grove model. For the oxide motion problem, all three possible models, namely elastic deformation, viscoelastic flow, and viscous flow, have been considered. The finite element efforts concentrate on viscous flow model whereas the boundary element works emphasize on viscoelastic flow. Finite element approaches will be reviewed first, followed by the BEM. A survey of applications of the BEM method on related disciplines will be given at the end.

1.3.2 History of Finite Element Approaches

The finite element method (FEM) is used widely to analyze mechanical structures, fluid flows, and other continua. Thermal oxidation poses a special problem for the FEM in that the oxide bulk deforms and expands continuously. The mesh that divides the bulk into elements must be generated successively with an automatic algorithm. Thus an important task in applying the FEM to thermal oxidation is to develop a robust mesh generator.

In using the FEM to simulate local oxidation, Poncet considered a linear elastic model and a large-strain nonlinear viscoelastic flow model to calculate oxide motion [14]. Unfortunately, no stress values were provided to judge the numerical aspect of his approach. Poncet had also investigated the effect of "mechanical stress potential"

on diffusion. Rafferty incorporates the incompressible viscous model for the oxidation part of a general-purpose process simulator, SUPREM-IV [15]. Sutardja treated oxide as an incompressible nonlinear viscous fluid in the oxidation/reflow simulator CREEP [16]. Sutardja also proposed modifications to Kao's model for modeling stress effects on oxidation.

1.3.3 Boundary Element Formulations

The boundary element method has gained the acceptance as a tool for studying structural problems. Its greatest claim to fame is its reduced dimensionality feature on modeling linear problems. Because it only requires the outline definition of a structure, the analysis process is greatly simplified, especially for three-dimensional problems. However, a major drawback of the BEM is that it is primarily intended for linear systems. Techniques for dealing with nonlinearity are not fully developed, and their efficiencies have yet to be fully demonstrated.

The first general-purpose two-dimensional oxidation simulator was developed by Chin [17]. Chin used the boundary element method to model both diffusion and flow. However, because the viscous flow problem was solved as a coupled Poisson systems, domain computations were needed. Later, the author of this report showed that viscous incompressible flow is a biharmonic system that is subject to the same reduced-dimension treatment as the diffusion problem [18]. Matsumoto investigated a viscoelastic BEM that, like Chin's approach, needs a mesh to subdivide the domain [19]. The bending of the silicon nitride mask was modeled as an elastic process. But despite the large deformation of the thick nitride mask, reported stress values were relatively small. This suggests the possibility that the stress history is not updated correctly. Although its constraints are less stringent than those of a FEM mesh, a dense BEM mesh, as used by Chin and Matsumoto, severely degrades the speed performance. To avoid domain calculations, Isomae used a Laplace transform technique to model viscoelastic flow of oxide [20]. However, numerical approximations to the

inverse transform are needed.

1.3.4 Other BEM Developments

Thermal oxidation is a rather limited field for the BEM. In other areas, many advances have been made. The following are relevant to our research efforts. Kaneko considered the application of the Laplace transform technique to modify Kelvin's solution [21] for viscoelastic deformation, much the same as our approach. However, because of the narrow interpretation of the constitutive equations, the particular solution is not adaptable to viscoelastic flow. A new solution needs to be derived. In a different approach, Tanaka uses the original Kelvin's solution, and makes the elastic modulus and Poisson's ratio time-dependent functions [22]. For nonhomogeneous Laplace problems, Butterfield suggested a "direct method" technique [23]. Our method for nonhomogeneous Laplace problems, which is modeled after elastoplastic techniques, turns out to be similar to Butterworth's, except that ours is an "indirect" formulation.

1.4 Report Outline

Other than the fact that oxide motion is a "higher" order system compared to oxidant diffusion, these processes share many common features. Therefore they will be presented in parallel, from the physics, the mathematics, to the numerical approximation.

Chapter 2 describes the basics of thermal oxidation. The oxidant diffusion and oxide motion issues will be dealt with in order; the physics will be described, but the mathematics will not be covered extensively. Fundamental concepts, especially those of solids, will be introduced. Modeling aspects of stress effects will be discussed.

The mathematical backgrounds are extended in Chapter 3 to bring out the characteristics of the systems. Their relation to the boundary element will be studied. The foundation of Laplace and biharmonic problems will be covered. The use of Laplace

transform on deriving the viscoelastic kernels will be illustrated. A unified approach on modeling nonlinearities in diffusion and viscoelastic flow will be presented at the end of the chapter.

Chapter 4 is on the numerical realization of the integral equations. Basics such as boundary and domain segmentation, methods for evaluating line integrals, and matrix solution for multiple domains will be covered. Adaptive iterative techniques required for solving nonlinear behaviors will also be treated.

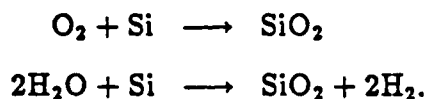
Simulations of various problems are shown in Chapter 5. Stress behavior is studied as a function of relaxation time. Parametric fittings on Kao's experimental data will be presented, along with a discussion on the scalability of parameter values. Effects of stress on viscosity of oxide and diffusivity of oxidants will be demonstrated.

Finally, Chapter 6 summarizes work that has been performed and proposes possible improvements and directions for future investigations.

Chapter 2

Thermal Oxidation Process

Thermal oxidation is typically carried out at 800–1100 °C. Silicon wafers are exposed to an oxidizing ambient such as dry oxygen, “wet” oxygen, or steam. Wet oxygen is prepared by passing oxygen through heated water to saturate it with water vapor or by direct burning of hydrogen gas in an oxygen ambient. From the following chemical reactions, silicon dioxide is formed:



Because the oxidation rate of H_2O is much higher than O_2 , thick oxides are typically grown in wet oxygen and steam ambients. Dry oxygen is used for growing thin layers such as gate insulator, where good oxide quality is of utmost importance.

In a one-dimensional thermal oxidation model, the processes of oxidant diffusion and oxide movement are often lumped together to form a general expression that relates oxide thickness to oxidation time. Extension of such an approach to two or three dimensions is not feasible. The direction of oxidant diffusion has no bearing on the oxide movement, therefore the two processes must be considered separately, even though they are tightly coupled. Their mathematical models will be presented to serve as a foundation for the boundary element method described in the next chapter. A review

on the one-dimensional oxidation model is given, followed by an examination of the two-dimensional oxidant diffusion model. In the second part of this chapter, a brief survey of the mechanics of solids and fluids will be given. Elastic deformation and incompressible viscous flow will be covered, followed by a generalization to viscoelasticity. The influence of stress on the oxidation process will also be discussed.

2.1 Diffusion Mechanism

The most widely-used oxidation model was formulated by Deal and Grove [13]. This model primarily concerns with the transport of oxidants from the gas phase to the oxide and within the oxide, and the chemical reaction that takes place at the silicon interface. The kinetics of oxide motion is largely ignored. The Deal-Grove model has been found to be generally valid for temperatures between 700 and 1300 °C, in atmospheric or reduced pressure [24]. In this model, steady-state molecular diffusion is assumed to be the main transport mechanism for the oxidizing species. The steady-state approximation is justified by the low solid solubility of oxidants in silicon dioxide. As a consequence of their low concentrations, oxidants must move many times faster than the silicon interface in order to maintain oxidation; thus their relaxation time is much smaller than that of the moving boundaries. To the oxidants, the silicon interface appears to be stationary; to the silicon interface, the oxidant flux appears to be in steady state. A dimensional analysis has been performed by Chin to prove that the steady-state approximation on the oxidant diffusion process is indeed valid [1].

2.1.1 One-Dimensional Diffusion

Fig. 2.1 shows a one dimensional oxidant profile. Oxidizing species in the gas phase enter the oxide layer at the free surface; the incoming flux is given by

$$F_1 = h(C^* - C_o) \quad (2.1)$$

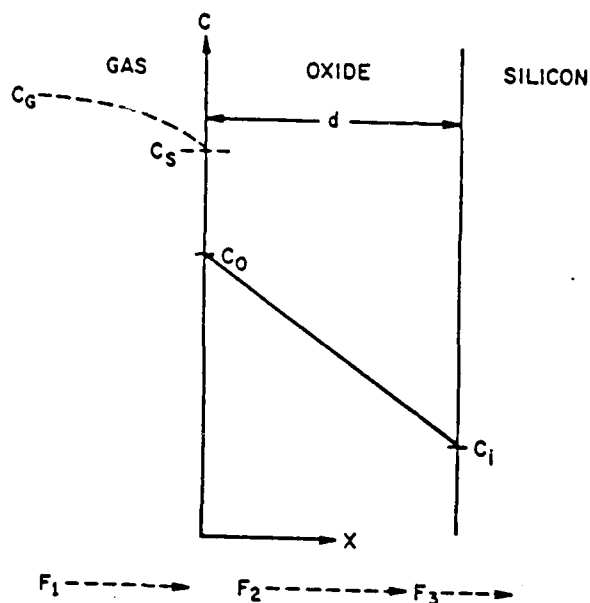


Figure 2.1: One-Dimensional Oxidation System. Oxidizing species enter from the gas phase, diffuse through the oxide, and react with silicon at the silicon interface. The bulk concentrations at the surface and at the silicon interface are C_o and C_i , respectively.

where h , C^* , and C_o are the gas-oxide mass-transfer coefficient, the equilibrium bulk concentration (when there is no diffusion), and the surface concentration, respectively.

The flux in the oxide is given by

$$\begin{aligned} F_2 &= -D \frac{dC}{dx} \\ &= D \frac{(C_o - C_i)}{d_o} \end{aligned} \quad (2.2)$$

where D , C_i , and d are the diffusion coefficient, the oxidant concentration at the silicon interface, and the oxide thickness respectively.

F_3 is the flux reaching the silicon interface where chemical reaction takes place to form oxide. The quantity of incoming flux must equal the amount consumed by the oxidation process. Oxide growth is assumed to be given by $k_s C_i$, where k_s is the

rate of surface reaction and C_i is the interface concentration. We thus have the third relationship:

$$F_3 = k_s C_i. \quad (2.3)$$

In steady state, $F_1 = F_2 = F_3$ as fluxes have to be conserved. Solving the three simultaneous equations, expressions for C_i and C_o are obtained,

$$C_i = \frac{C^*}{1 + \frac{k_s}{h} + \frac{k_s d_0}{D}} \quad (2.4)$$

$$C_o = \frac{\left(1 + \frac{k_s d_0}{D}\right) C^*}{1 + \frac{k_s}{h} + \frac{k_s d_0}{D}} \quad (2.5)$$

Combining Eqs. 2.3 and 2.4, the following differential equation that determines the oxide growth rate results

$$\begin{aligned} N \frac{dd_0}{dt} &= F_3 \\ &= \frac{k_s C^*}{1 + \frac{k_s}{h} + \frac{k_s d_0}{D}} \end{aligned} \quad (2.6)$$

where N is the number of oxidant molecules required to form a unit volume of oxide. The solution of Eq. 2.6 is the well-known linear-parabolic oxidation equation which relates oxidation time, t , to oxide thickness, d_0 :

$$d_0^2 + A d_0 = B(t + \tau) \quad (2.7)$$

where

$$\begin{aligned} A &= 2D \left[\frac{1}{k_s} + \frac{1}{h} \right], \\ B &= \frac{2DC^*}{N}, \\ \tau &= \frac{A d_i + d_i^2}{B}. \end{aligned}$$

τ is the equivalent time required to produce the existing oxide thickness d_i on bare silicon. Usually h is removed from the expression for A because it is much larger than k_s .

	Parabolic Linear	$B = C_1 e^{-E_1/kT}$ $B/A = C_2 e^{-E_2/kT}$
	(111) Silicon	(100) Silicon
Dry O ₂	$C_1 = 7.72 \times 10^2 \mu\text{m}^2/\text{h}$ $C_2 = 6.23 \times 10^6 \mu\text{m}/\text{h}$ $E_1 = 1.23 \text{ eV}$ $E_2 = 2.0 \text{ eV}$	$C_2 = 3.71 \times 10^6 \mu\text{m}/\text{h}$ All other parameters same as for (111) silicon
H ₂ O (640 torr)	$C_1 = 3.86 \times 10^2 \mu\text{m}^2/\text{h}$ $C_2 = 1.63 \times 10^8 \mu\text{m}/\text{h}$ $E_1 = 0.78 \text{ eV}$ $E_2 = 2.05 \text{ eV}$	$C_2 = 0.97 \times 10^8 \mu\text{m}/\text{h}$ All other parameters same as for (111) silicon

Table 2.1: Coefficients for Linear and Parabolic Rates.

There are two limiting cases to the general relationship, Eq. 2.7. When the oxide is thin, the growth rate is approximately constant since it is limited by the reaction rate at the silicon interface; d_0 is then given by

$$d_0 \approx \frac{B}{A}(t + \tau). \quad (2.8)$$

As the oxide grows thicker, the growth rate becomes limited by the diffusion rate of the oxidant in the oxide. In this case, the growth rate is inversely proportional to the oxide thickness; the relationship between d and t is then given by the parabolic equation

$$d_0^2 \approx B(t + \tau). \quad (2.9)$$

Hence B/A and B are known as the linear and the parabolic oxidation rate constants respectively. The values of B/A and B in dry and wet oxygen are given in Table 2.1 [25].

The Deal-Grove model has been found to apply very well to a wide range of conditions. However, it does not account for the rapid initial growth phase in dry oxygen: when oxide thicknesses is less than 300Å, anomalous high oxidation rate is observed.

Another behavior not predicted by the Deal-Grove model is the nonlinear power-law dependence of the linear rate rate on the ambient pressure in high-pressure dry oxidation. Various modifications to the theory have been proposed to explain these phenomena [26,27,28].

2.1.2 Oxidant Diffusion in Two Dimensions

The extension of the diffusion model to two and three dimensions is straight forward. The flux vector given by Fick's first law is ¹

$$\mathbf{F} = -D\nabla C \quad (2.10)$$

At the oxide free surface, Eq. 2.1 is replaced by

$$\mathbf{F} \cdot \hat{\mathbf{n}} = -h(C^* - C),$$

or equivalently by

$$D \frac{\partial C}{\partial n} = h(C^* - C).$$

where $\hat{\mathbf{n}}$ is the unit normal vector at the surface pointing away from the bulk. As mentioned earlier, the mass-transport coefficient h is large; hence C is approximately C^* at the free surface. For all practical purposes, Eq. (2.8) can be replaced by

$$C = C^*.$$

At the silicon interface, Eq. 2.3 is generalized to

$$\mathbf{F} \cdot \hat{\mathbf{n}} = k_s C$$

or

$$D \frac{\partial C}{\partial n} = -k_s C. \quad (2.11)$$

¹From now on, we do not use subscripts on C to denote the boundary type. This change is to be consistent with other expressions,

To complete the model, the boundary condition at the silicon nitride interface is considered. Being highly refractory and impermeable to oxidants, silicon nitride is used extensively as a masking material for local oxidation. Since no significant amount of oxidants diffuse through the nitride layer, the normal component of the oxidant flux at the nitride interface is assumed to be zero:

$$\mathbf{F} \cdot \hat{\mathbf{n}} = 0,$$

or simply

$$\frac{\partial C}{\partial n} = 0. \quad (2.12)$$

2.1.3 Effects of Stress on Oxidant Transport

Stress affects the diffusion of oxidants by way of several mechanisms. The ones that Kao implemented in his oxidation model are on the reaction rate and the equilibrium concentration [29]. According to Kao's theory, the normal surface traction (i.e. the force acting perpendicular to the surface) changes the reaction kinetics at the silicon interface. A positive force impedes the flow of oxide. Since extra energy is required to overcome the force, the likelihood of an oxidation event for this energy-activated process is reduced. Similarly, hydrostatic pressure can alter the equilibrium concentration, C^* , by reducing the concentration when it is compressive. Hydrostatic pressure is known to interfere with diffusivity because it removes vacancies needed for diffusion [30]. Although described but not used in Kao's original model, the diffusivity model was incorporated in Sutardja's simulation model [16]. Finally, it should be noted that an elastic potential gradient can drive oxidants from a high pressure area to a low pressure area [30,14]; this process is related to that on C^* .

Followed Sutardja's model, only effects on k , and D are considered in the stress study. Although other mechanisms may be included, they only make the situation worse by introducing more unknown variables than it is currently possible to resolve.

The stress formulae are given below:

$$k_s = k_{s0} \exp\left(\frac{p_n V_{k_s}}{kT}\right) \quad (2.13)$$

$$D = D_0 \exp\left(\frac{\sigma_M V_d}{kT}\right) \quad (2.14)$$

where k_s and D_0 are the original unstressed reaction rate and diffusion coefficient, p_n is the normal component of the surface traction, σ_M is the pressure, V_{k_s} and V_d are the activation volumes for the reaction rate and diffusivity. The equations are given in the activation-energy form, to be consistent with other temperature-dependent expressions. Note that k_{s0} and D_0 are functions of temperature as well.

2.2 Oxide Motion

In oxidation, the silicon material undergoes a phase change, from single-crystal (or sometimes polycrystalline) silicon into an amorphous glass which has an open network structure and which does not show any long-range crystal order. When a silicon atom is oxidized to form a silicon dioxide molecule, its volume increases from $20 \text{ \AA}^3$ to $45 \text{ \AA}^3$; the expansion is approximately 2.25 times. Part of the oxide molecule fills in the void of the silicon atom, the remaining squeezes into the existing oxide network. On an atomic level, a high stress is produced. This intrinsic stress is responsible for the bending of silicon substrate under planar oxidation conditions [31].

As oxidation progresses, the silicon interface recedes into the silicon substrate, while the free surface moves away from its original position, as shown in Fig. 2.2. From a macroscopic view, the motion of oxide in one dimension can be treated as a coordinate translation that does not produce stress and therefore does not require any knowledge about the mechanical properties of oxide.

For nonplanar surfaces and nonuniform oxidation rate, the mechanical properties of oxide are needed to define a model for the oxide motion. Interactions with other

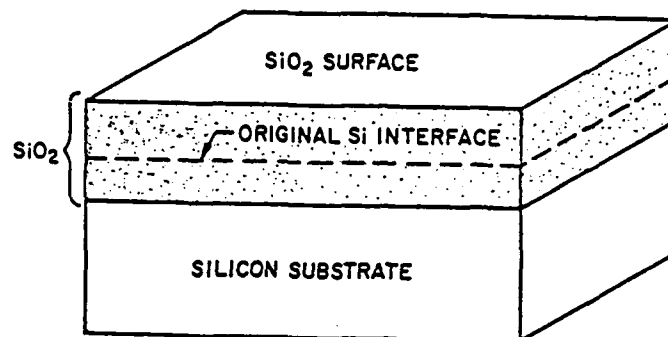


Figure 2.2: Growth of Silicon Dioxide.

materials present in the system, namely the silicon substrate and the silicon nitride mask, must also be contemplated in two and three dimensional models.

2.2.1 Material Properties

The two solid materials that oxide comes in contact are the silicon substrate and the nitride mask. The properties of these three materials are dissimilar and may require different numerical treatments.

Silicon dioxide has been studied extensively for its mechanical properties [32,33]. (But most of these properties are not relevant to thermal oxidation.). Silicon dioxide is also known as silica; it is usually found in amorphous state. At high temperatures, it has a tendency to devitrify into crystalline form. Like any other materials, silica exhibits nonlinear behavior in its elastic properties. The bulk modulus changes with applied pressure. Under suitably high pressure, silica can be permanently densified.

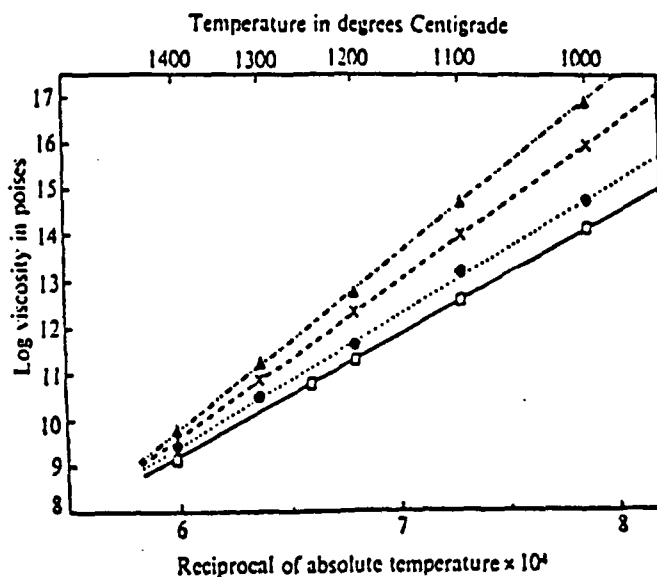


Table 2.2: Equilibrium viscosities for various silica glasses. (Δ) type I (IR); ($\times$) type II (OG with 0.027wt% OH); ($\bullet$) type II (OG with 0.04wt% OH); ($\square$) type III (Spectrosil) silica glass.

This phenomenon is also observed in low-temperature oxidation [34].

Oxide is a viscoelastic material that flows readily at high temperatures. Its viscosity is a strong function of the temperature. Shown in tables 2.2 are viscosities for silica glasses with different OH concentrations [33,35]. As it is evident from the table, viscosity is also very sensitive to the presence of water vapor. Water molecules combine with bridge oxygen ions in the oxide to form stable nonbridging hydroxyl groups [25], thereby weakening the network structure and greatly reducing the viscosity. Impurities such as Na and P show similar effects. However, as mentioned in Chapter 1, oxide is frequently approximated as a viscous incompressible fluid for simplicity.

Silicon nitride is usually deposited by means of chemical vapor deposition. The chemical structure of pure silicon nitride is Si_3N_4 . Like oxide, nitride is amorphous. Its high-temperature behavior is not well characterized, although it is known to crack readily. In LOCOS, the nitride mask bends during the process. When the oxide is

removed by etching, the mask only returns partially to the original position. It is suspected that silicon nitride deforms elastoplastically (plastically). As a first cut, nitride is modeled as a linear elastic element. If more information is available, it may be treated as a nonlinear viscoelastic or plastic material.

The silicon substrate on which oxide is grown is a single-crystal material. For obvious reasons, we would not like to allow the substrate to "flow", deform, or otherwise be damaged structurally. Defects in the crystal structure is detrimental to the electrical characteristics of devices built on it. Silicon is therefore treated as an elastic material. However, having a long-range and well-defined atomic arrangement in the structure, single-crystal silicon exhibits anisotropy in its mechanical properties — different crystallographic planes have different elastic modulus and Poisson's ratio. For simplicity, silicon may be treated as if it is an isotropic material.

At this point, it is clear that a general-purpose viscoelastic method that can handle a wide range of stress relaxation is highly desirable. As we will see later in the next chapter, viscous incompressible flow and elastic deformation can be considered special cases of viscoelasticity in which the relaxation time is small or ∞ respectively. Nonetheless, all the three models will be examined in the following sections to show their interrelationship.

2.2.2 Mechanics of Solids and Fluids

For many physical systems, it is often adequate to consider only two (or even one) dimensional models to understand their fundamentals and to perform analysis work. Their generalization to two or three dimensions is straight forward, as we have seen earlier in the diffusion process. For elasticity, the converse is true. One has to start with three-dimensional concepts, and work the way down to one or two dimensions. All one or two-dimensional problems are based on a particular simplified three-dimensional behavior.

Mathematically, diffusion is simpler than elasticity. The former deals with scalars and vectors, whereas the latter works with vectors and tensors. Vectors specify displacement fields and forces; tensors defines the strain and stress quantities in the materials. Because of the increased "rank" of parameters, indicial notation is employed in this report to represent components of interest and to define repeated operations. When an index letter appears only once in a term, it implicitly takes on a value range of 1 to N where N is 2 or 3 — the dimension of the system. Thus

$$a_i = \mu b_i$$

is a compact representation for a set of equations:

$$a_1 = \mu b_1;$$

$$a_2 = \mu b_2;$$

$$a_3 = \mu b_3.$$

If an index letter appears twice in a term, it means summation over the range of 1 to N . Hence

$$a_i = b_i c_{jj}$$

is understood to be

$$a_i = b_i (c_{11} + c_{22} + c_{33})$$

repeated for all possible values of i .

To be compatible with the indicial notation, the principal axis of the global Cartesian coordinate, x , y , and z are also referred to as x_1 , x_2 , and x_3 so that they can be indexed appropriately. Subscripts and superscripts are used throughout the report for different meanings and purposes. In order to avoid excessive confusion, an index letter (i , j , or k) appears only as a subscript and only in lower case. The indicial notation, although concise, is inconvenient at times; hence, the vector notation will be retained wherever it is more suitable. Note that f and f_i are equivalent.

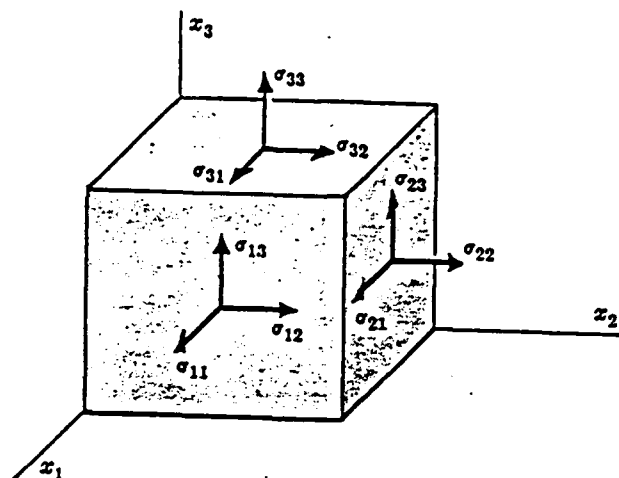


Figure 2.3: Stress Tensor Diagram.

2.2.3 Stress and Strain Tensors

When a force is applied to a material, the material deforms or reacts accordingly. Like the diffusion flux, the force is a vector that has a magnitude and a direction. The resulting stress inside the material is a second order tensor, however. A stress tensor or dyadic has 9 components:

$$\sigma_{ij} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \quad (2.15)$$

Their spatial orientations are illustrated in Fig. 2.3. σ_{11} , σ_{22} , and σ_{33} acts in the direction perpendicular to the principal planes (whose normals are $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$); they are called the normal stress. σ_{12} , σ_{13} , σ_{21} , σ_{23} , σ_{31} , and σ_{32} are shear stress acting tangent to the planes. In summary, σ_{ij} is a force acting in the j direction on plane i . Due to moment conservation, the stress tensor is a symmetric matrix, i.e. $\sigma_{ij} = \sigma_{ji}$.

The forces on the 3 principal planes are therefore given by

$$\begin{aligned} t_1 &= \sigma_{11}\hat{x}_1 + \sigma_{12}\hat{x}_2 + \sigma_{13}\hat{x}_3 \\ t_2 &= \sigma_{21}\hat{x}_1 + \sigma_{22}\hat{x}_2 + \sigma_{23}\hat{x}_3 \\ t_3 &= \sigma_{31}\hat{x}_1 + \sigma_{32}\hat{x}_2 + \sigma_{33}\hat{x}_3. \end{aligned}$$

For a plane of arbitrary orientation, the force acting on it is a linear combination of the 3 t 's:

$$p = n_1 t_1 + n_2 t_2 + n_3 t_3$$

which is, in the indicial notation,

$$\begin{aligned} p_i &= n_j \sigma_{ij} \\ &= n_j \sigma_{ji}. \end{aligned} \tag{2.16}$$

The n 's are the "direction cosines" associated with the plane. n_i is defined to be the cosine of the angle between the normal of the plane and the axis $\hat{x}_i$. Eq. 2.16 is one of the fundamental expressions in mechanics which relates surface traction (i.e. surface force) to the stress tensor of the continuum.

As diffusion is subject to flux conservation, a mechanics problem obeys force conservation, which is given by the equilibrium condition:

$$\frac{\partial \sigma_{ij}}{\partial x_j} = b_i \tag{2.17}$$

where b is the body force, which may be due to gravity.

Finally, we have the following definition for the strain or deformation state:

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \tag{2.18}$$

where ϵ is the strain tensor and $u = u_1 \hat{x}_1 + u_2 \hat{x}_2$ is the displacement. This statement reveals that strain is a unitless parameter quantifying the distortion of a material. As a linear equation, Eq. 2.18 is for small distortions.

It is useful at this point to separate the stress and strain tensors into a spherical part and a deviatoric part. Given that the spherical (or mean) stress and strain are rotationally invariant, they are often expressed as scalars:

$$\sigma_M = \frac{1}{3}\sigma_{ii}; \quad (2.19)$$

$$\epsilon_M = \frac{1}{3}\epsilon_{ii}, \quad (2.20)$$

where subscript M is capitalized to indicate that it is not an index. The deviatoric stress and strain tensors are obtained from the regular expressions:

$$s_{ij} = \sigma_{ij} - \delta_{ij}\sigma_M, \quad (2.21)$$

$$e_{ij} = \epsilon_{ij} - \delta_{ij}\epsilon_M, \quad (2.22)$$

where the definition for the Kroncker delta is:

$$\begin{aligned} \delta_{ij} &= 1, & (i = j); \\ &= 0, & (i \neq j). \end{aligned}$$

The main reason for separating the strain and stress tensor into these two components is that the spherical part is involved in volume change of the material whereas the deviatoric part is not. Their behavior are expected to be fundamentally different.

2.2.4 Maximum Shear Stress

A vector, having a direction and a magnitude, is easy to picture regarding its translation from one coordinate system to another. However, the transformation for tensor is more difficult to visualize. Consider a rod aligned in the $\hat{x}_1$ direction and under tension. It only has a nonzero stress tensor component, namely σ_{11} . With this particular global reference, it is not clear that there is any shear stress. In fact, maximum shear stress does occur on any planes that are inclined at 45° with respect to the axis of the rod.

Maximum shear stress is often used as a criterion for determining the plastic yield threshold and other nonlinear behaviors. Its value may be obtained from the principal stresses. For any stress state, it is possible to find a local coordinate frame such that only σ'_{11} , σ'_{22} , and σ'_{33} are nonzero. If these three components are sorted and relabeled such that $\sigma_I \geq \sigma_{II} \geq \sigma_{III}$, the maximum shear stress, σ_S (of another frame) is given by

$$\sigma_S = \frac{1}{2}(\sigma_I - \sigma_{III}). \quad (2.23)$$

Another parameter used for calculating nonlinear effects is the maximum distortion energy given by

$$C_Y = \frac{1}{6} [(\sigma_I - \sigma_{II})^2 + (\sigma_I - \sigma_{III})^2 + (\sigma_{II} - \sigma_{III})^2]. \quad (2.24)$$

Both parameters have been used in this investigation. The calculation for the principal stresses is given in the Appendix.

Eqs. 2.16, 2.17, and 2.18 are some basic concerns that do not depend on the details of the material types. What decides whether a material is fluid or solid is the constitutive equations that relates the stress tensor to the strain tensor. The three basic property models, namely elastic, viscous, and viscoelastic, are examined in the following sections.

2.2.5 Linear Elasticity

A purely elastic material has perfect memory. When a force is applied to it, it deforms. However, it will return to its original shape after the force is removed. Expressed in terms of the deviatoric and spherical components, the constitutive equation for linear elasticity is:

$$s_{ij} = 2Ge_{ij} \quad (2.25)$$

$$\sigma_M = 3K\epsilon_M \quad (2.26)$$

where G and K are the shear and bulk moduli. These moduli may be given in terms of Young's modulus, E , and Poisson's ratio, ν :

$$K = \frac{E}{3(1-2\nu)} \quad (2.27)$$

$$G = \frac{E}{2(1+\nu)}. \quad (2.28)$$

Young's modulus is more suitable for uniaxial analysis than for 2 and 3 dimensional problems. Poisson's ratio is a measure of the compressibility of a material; it ranges from 0 to 0.5. When ν is 0, the material is said to be totally compressible. At this point, the decomposition of stress and strain into spherical and deviatoric parts is not necessary because $2G = 3K$. At the other extreme, the material is totally incompressible when ν is 0.5. Because the bulk modulus is infinite, the relationship between spherical stress and strain as specified in Eq. 2.28 breaks down. ϵ_M must be zero at all times, but σ_M can take on any finite value. In other words, deformation cannot involved volume change in the material.

In a compact form, the relation between stress and strain is expressed as follows:

$$\sigma_{ij} = \lambda \delta_{ij} \epsilon_{kk} + 2\mu \epsilon_{ij} \quad (2.29)$$

where λ and μ are Lamé's constants defined as:

$$\lambda = E \frac{\nu}{(1+\nu)(1-2\nu)} \quad (2.30)$$

$$\begin{aligned} \mu &= \frac{E}{2(1+\nu)} \\ &= G. \end{aligned} \quad (2.31)$$

When the body force, b , is zero, Eqs. 2.18, 2.17 and 2.29 establish a constraint for the displacement field u :

$$\mu \nabla^2 u + (\lambda + \mu) \nabla (\nabla \cdot u) = 0. \quad (2.32)$$

Eq. 2.32 is known as the Cauchy-Navier equation.

Listed in the table below are the mechanical properties of silicon dioxide, silicon, and silicon nitride:

Material	Silicon Dioxide	Silicon Nitride	Silicon
Young's Modulus (E), dynes/cm ⁻²	8×10^{11}	3.29×10^{11}	3.69×10^{12}
Poisson's ratio (ν)	0.194	0.266	0.42
Bulk Modulus (K), dynes/cm ⁻²	4.35×10^{11}	2.34×10^{11}	7.68×10^{12}
Shear Modulus (G), dynes/cm ⁻²	3.35×10^{11}	1.3×10^{11}	1.3×10^{12}

Table 2.3: Elastic properties of oxide, nitride, and silicon.

2.2.6 Two-Dimensional Plane Strain

In making a two-dimensional approximation for a three-dimensional problem, two approaches may be taken. In plane stress, the domain is assumed to be thin and free to expand in the third dimension, say $\hat{x}_3$, and loading on the boundary does not vary in the third dimension. This results in the following constraints on the stress and strain tensors:

$$\epsilon_{13} = \epsilon_{23} = \epsilon_{32} = \epsilon_{31} = 0 \quad (2.33)$$

$$\sigma_{13} = \sigma_{23} = \sigma_{32} = \sigma_{31} = \sigma_{33} = 0 \quad (2.34)$$

$$\epsilon_{33} = -\frac{\nu}{E}(\sigma_{11} + \sigma_{22}). \quad (2.35)$$

In plane strain, the domain is assumed to be long and uniform in the $\hat{x}_3$ direction. Because it cannot expand in $\hat{x}_3$ direction, the following constraints are imposed on the stress and strain tensors:

$$\epsilon_{13} = \epsilon_{23} = \epsilon_{33} = \epsilon_{32} = \epsilon_{31} = 0 \quad (2.36)$$

$$\sigma_{13} = \sigma_{23} = \sigma_{32} = \sigma_{31} = 0 \quad (2.37)$$

$$\sigma_{33} = \nu(\sigma_{11} + \sigma_{22}). \quad (2.38)$$

The respective configurations for plane stress and plain strain are illustrated in Fig. 2.4.

For our simulation efforts, we assume all structures are long in the third dimension and therefore treat them as 2-dimensional plane strain problems. The Cauchy-Navier equation Eq. 2.32 and other essential relationships remain unchanged for plane strain. Unless otherwise noted, we will refer to all parameters and functions as if they are for two-dimensional systems from here on.

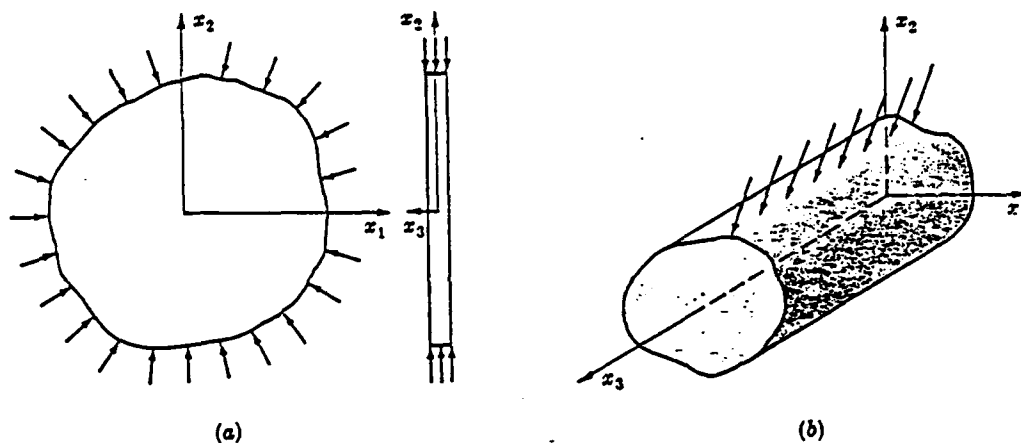


Figure 2.4: Two-Dimensional Elasticity. (a) Plane stress; (b) plane strain.

2.2.7 Incompressible Viscous Flow

Viscous flow is one model totally different from elasticity. It is totally memoryless – any deformation is permanent. Stress is only sustained only the material deforms continuously, *i.e.* when it flows as a fluid. The dynamical equation for slow incompressible viscous flow is:

$$\eta \nabla^2 \mathbf{v} = -\nabla P \quad (2.39)$$

where η is the viscosity, P the hydrostatic pressure, and $\mathbf{v} = \partial \mathbf{u} / \partial t$ is the velocity. This parameter is the same as the spherical stress in elasticity but with opposite sign: $P = -\sigma_M$. A positive σ_M implies tensile pressure whereas a positive P is compressive. The flow is subject to the incompressibility condition:

$$\nabla \cdot \mathbf{v} = 0 \quad (2.40)$$

which is the equivalent to the $\epsilon_M = 0$ statement for incompressible elasticity.

2.2.8 Viscoelastic Deformation

Showing intermediate behavior between elastic deformation and viscous flow is a wide range of viscoelastic materials. They are further classified as Maxwell fluid or Voigt solid. Like an elastic material, the Voigt solid will return to its original shape after the applied force is removed. It differs in one aspect – the strain is not only related to stress but also to stress rate. The Maxwell element behaves like a viscous fluid but it retains some memory. If the applied force is removed fast enough, it will return partially to its original position. On the other hand, if it is kept in a distorted state for an extended period, it will lose all its stress and remain in the new shape.

Formally, viscoelastic stress-strain relationship is given by a time-differential equation of the form

$$p_0\sigma + p_1\dot{\sigma} + p_2\ddot{\sigma} = q_0\epsilon + q_1\dot{\epsilon} + q_2\ddot{\epsilon} \quad (2.41)$$

or by the more compact statement:

$$P\sigma = Q\epsilon \quad (2.42)$$

where

$$P = \sum_{i=0}^m p_i \frac{\partial^i}{\partial t^i},$$

$$Q = \sum_{i=0}^n q_i \frac{\partial^i}{\partial t^i}.$$

Table 2.2 show some fluid and solid models [36]. In general, the material is a solid if q_0 is nonzero. We note that the elastic constitutive equations Eq. 2.26 is a special case of Eq. 2.42. The time-differential stress-strain relationship can be converted to different forms to suit the applications – hereditary integral, complex variable, Laplace transform, and so forth.

Although one may incline to use a high-order differential equation to describe a viscoelastic behavior as completely as possible, such an overkill approach may not

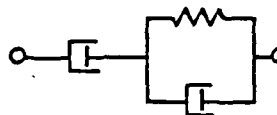
Model	Name	Differential equation
		Inequalities
	elastic solid	$\sigma = q_0 \epsilon$
	viscous fluid	$\sigma = q_1 \dot{\epsilon}$
	Maxwell fluid	$\sigma + p_1 \dot{\sigma} = q_1 \dot{\epsilon}$
	Kelvin solid	$\sigma = q_0 \epsilon + q_1 \dot{\epsilon}$
	3-parameter solid	$\sigma + p_1 \dot{\sigma} = q_0 \epsilon + q_1 \dot{\epsilon}$
		$q_1 > p_1 q_0$
	3-parameter fluid	$\sigma + p_1 \dot{\sigma} = q_1 \dot{\epsilon} + q_2 \ddot{\epsilon}$
		$p_1 q_1 > q_2$

Table 2.4: Viscoelastic Models

serve to illuminate the problem. In fact, there may even be difficulties in gathering sufficient data to estimate the basic p 's and q 's. Factors such as nonlinearity deserve more attention in advanced approximations.

For our simulation, we only consider the simplest viscoelastic flow representation. Here, we assume that the spherical component is purely elastic:

$$\sigma_M = 3K\epsilon_M.$$

And the deviatoric part is a Maxwell fluid governed by a first order differential equation:

$$\frac{1}{G} \frac{\partial \sigma}{\partial t} + \frac{\sigma}{\eta} = \frac{\partial \epsilon}{\partial t}.$$

Its corresponding mechanical model is a linear spring in series with a viscous dashpot, as depicted in Table 2.2. Note that the spherical component must display a solid behavior – a fluid loses stress and allows indefinite compression, producing a mass conservation problem.

2.2.9 Nonlinear Effects of Stress

All the models described earlier are idealized linear approximation. They may be augmented to include certain nonlinear stress effects. Consider a elastic problem whose stress-strain curve is shown in Fig. 2.5. For small stress (or strain), the relationship is linear from O up to P ; it starts to depart from the straight line and become nonlinear for larger stress. If the material is able to return to O after the applied stress is removed, it is nonlinear elasticity that we are working with. If the material has experienced irreversible slip on the atomic level, it may only return to the position R . In this case, we are dealing with elastoplasticity or plasticity.

For our modeling efforts, we ignore the nonlinear elastic moduli or plasticity — we are only concerned with the stress effects on viscosity in viscous and viscoelastic flow. Note that the derivation of the nonlinear viscoelasticity BEM is based on plasticity.

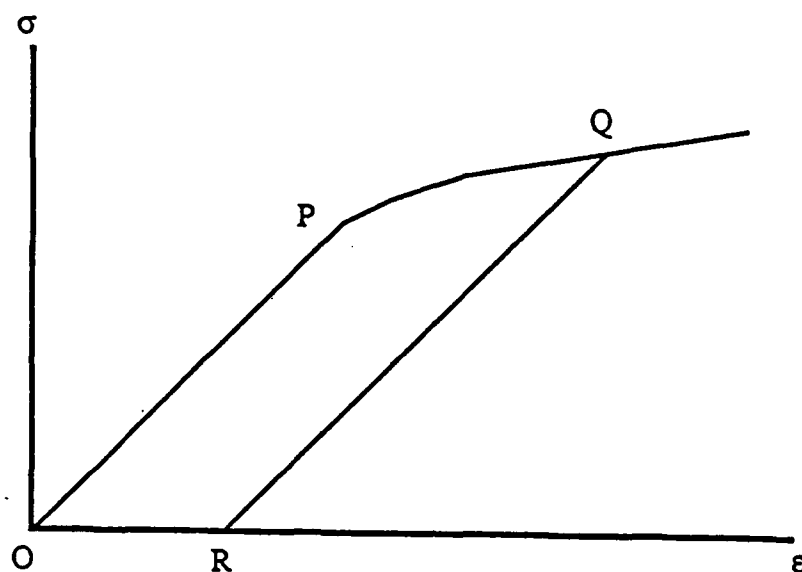


Figure 2.5: Nonlinear Elasticity and Plasticity. The material is linear from O to P; it exhibits a nonlinear behavior from P to Q. If it is able to return to O when the applied stress is removed, the material is elastic. If it can only return to R, it is said to suffer from permanent plastic deformation.

In Kao's viscous flow model, the following formulation is used to modify the viscosity [29]:

$$\eta = \eta_0 \exp \left(-\frac{\sigma_M V_\eta}{kT} \right). \quad (2.43)$$

However, as pointed out by Rafferty [15], this model can induce positive feedback for certain geometries. What happens is that a large negative (compressive) stress causes the viscosity to increase, this in turn drives the stress more negative, until the solution blows up². Using the same viscous flow model, Sutardja eliminated the instability by changing Eq. 2.43 to [16]

$$\eta = \eta_0 \frac{\sigma_s V_0 / 2kT}{\sinh(\sigma_s V_0 / 2kT)}. \quad (2.44)$$

Because the maximum shear stress σ_s is defined to be a positive quantity, the viscosity will always be equal or less than its nominal value. Thus the system is stable. However, in order to get a good fit with Kao's experimental data, Sutardja was required to use a truncated function for the diffusivity:

$$D = D_0 \times \min \left(\exp \left(\frac{\sigma_M V_D}{kT} \right), D_{\max} \right). \quad (2.45)$$

The value for $D_{\max}$ is sometimes as low as D_0 . To constraint D so that it does not exceed its nominal value is not physical. If compressive stress can reduce diffusivity, tensile stress should enhance it. In Deal-Grove oxidation model, the retardation effect of intrinsic stress on diffusivity is probably incorporated in the various parameters. When the stress level is reduced, the actual diffusivity can be higher than its nominal value. Of course, there is a limit as to how much diffusivity can be boosted by tensile stress before other effects, such as fracturing, set in.

We propose the following model for viscosity:

$$\eta = \eta_0 \exp \left(-\frac{\sigma_M V_{\eta 1}}{kT} \right) \frac{\sigma_s V_{\eta 2} / kT}{\sinh(\sigma_s V_{\eta 2} / kT)} \quad (2.46)$$

which is a combination of Eqs. 2.43 and 2.44. We retain the pressure dependency for the following two reasons. First, in Kao's experiments, the main difference between a

²If a viscoelastic flow model is used instead, the solution will always be stable because the worst that can happen is that the material becomes elastic when the viscosity goes to infinity. But now it suffers from a different misbehavior: a sudden solidification from a viscoelastic fluid to an elastic solid.

concave surface and a convex one is the sign of the pressure term, as far as stress effects are concerned. The pressure is tensile in the convex case, yet no significant overshoot in oxide thickness is observed for any curvatures. This suggests that the viscosity of oxide is less than its normal value for the convex structures. Second, researchers have proposed that shear action help to densify silica under pressure by interlocking the SiO_2 network [32]. The same interlocking mechanism, which requires the assistance of compressive pressure, may make it more difficult for Si-O bonds to break, hence increasing the viscosity. Because of geometric considerations, shear stress is always present in oxide. By choosing a suitable V_{n1}/V_{n2} ratio, we can assure that Eq. 2.46 is always stable.

Chapter 3

Boundary Element Formulations

The equations for diffusion and mechanics are presented in Chapter 2. In this chapter we shall examine and understand their mathematical characteristics and significance. These properties serve as the foundation for the boundary element method (BEM). The development of the boundary element technique will be laid out in detail. Variations of implementation techniques will be described.

It will be shown that both oxidant diffusion and oxide flow are simple boundary value problems (BVP's). The definition of a BVP is that the solution within the enclosed domain is uniquely determined by the boundary conditions. Many physical systems are BVP's in nature. The DC operation of a bipolar or field-effect transistor is one such example. The terminals or boundary contact points are the emitter, base, and collector for the bipolar device, or source, gate, and drain for the field effect transistor. For a given set of terminal voltages, there will be some well-defined current flowing through these terminals. However, the modeling of semiconductor devices is complicated and involved because the systems are highly nonlinear. Linear BVP's are easier to solved; in particular, the so-called Laplace and biharmonic problems are especially amenable to the boundary element method.

As its other more generic name - the boundary-integral equation method - im-

plies, the BEM involves the solution of some integral equations at the boundary. No domain calculations are necessary, exception for nonlinear problems. This reduced dimension feature, *i.e.* solving a three-dimensional system as a surface problem (which is two-dimensional), and a two-dimensional system as a line problem (which is one-dimensional) is very attractive for many applications.

3.1 Laplace Problems

The solutions for many physical problems of interest, such as electrostatic potential in insulators, steady-state heat flow and steady-state diffusion, satisfy Laplace's equation:

$$\nabla^2 \Phi = 0 \quad (3.1)$$

where Φ denotes the primary parameter of the system. For the three examples, Φ is the potential, the temperature distribution, or the concentration of diffusing species, respectively. The secondary parameter of these systems is the "flux", a vector element defined as

$$\mathbf{F} = -D \nabla \Phi \quad (3.2)$$

where $\mathbf{F}$ is interpreted to be the electric field, thermal energy flux, or particle flux, and D is the dielectric constant, thermal conductivity, or diffusivity.

The Laplacian condition arises from a conservation law on the flux

$$\nabla \cdot \mathbf{F} = 0 \quad (3.3)$$

which states that the divergence of $\mathbf{F}$ is zero, *i.e.* electric field, heat energy, or diffusing particles must conserve.

Substituting Eq. 3.2 into Eq. 3.3, the following expression is obtained,

$$\nabla D \cdot \nabla \Phi + D \nabla^2 \Phi = 0. \quad (3.4)$$

This equation becomes Laplace's equation (Eq. 3.1) if D is uniform so that $\nabla D = 0$. In other words, Φ is Laplacian if the system is homogeneous.

The solution to a Laplace problem is uniquely determined by its boundary conditions, which may be given in the following forms:

- Φ itself — a Dirichlet boundary condition.
- $\frac{\partial \Phi}{\partial n}$, the normal derivative of Φ — a Neumann boundary condition.
- $\alpha\Phi + \beta\frac{\partial \Phi}{\partial n} = \gamma$ where α, β , and γ are constants. This generalized expression is called a Robin condition.

A function $f(x, y)$ that satisfies Laplace's equation for two dimensions is also known as a harmonic function, a term ostensibly comes from complex analysis¹. The mean value theorem states that the value of a harmonic function at point p is equal to the average of its values over the area of any circle centered on p [37]. From this theory, we obtain the statement that the maximum (or minimum) of $f(x, y)$ in the domain Ω must occur on the enclosing boundary Γ . This consequently guarantees the uniqueness of the solution: if two functions $f_1(x, y)$ and $f_2(x, y)$ are harmonic inside Ω and are found to match along Γ , they must be identical within Ω : if the difference $f_1(x, y) - f_2(x, y)$ (which is also harmonic) is zero on Γ , it has to be zero everywhere in Ω , as dictated by the maximum principle.

We have thus seen the foundation of the BEM Laplace problem: if we can somehow find a certain function that is harmonic inside Ω and that matches the prescribed condition on Γ , we know that we have obtained the solution for Ω . Indeed, we never need to figure out what is happening inside Ω during the computation process, although we can if we want to.

The determination of a solution that fits the prescribed boundary condition can be accomplished in several ways. First, a polynomial series in x and y may be used. As

¹The periodic logarithmic kernel mentioned later in this chapter is in fact obtained from a complex function, as it will be shown later.

an illustration, five terms are shown below:

$$F(x, y) = A + Bx + Cy + Dxy + E(x^2 - y^2) \quad (3.5)$$

This particular harmonic function has five unknown coefficients, therefore it can be used to match "five" boundary conditions. If there are fewer boundary conditions, some terms in Eq. 3.5 have to be removed. Conversely, if there are more, higher polynomial terms such as $x^3 - xy^2$ must be added. If the problem is periodic, say in the $\hat{x}$ direction, the following trigonometric series may be used:

$$G(x, y) = \sum_{n=1}^N [a_n \cos(nx) + b_n \sin(nx)][c_n \exp(ny) + d_n \exp(-ny)]. \quad (3.6)$$

Eqs. 3.5 and 3.6 are distinctly different – they do not seem to be able to yield the same result. This is because the number of "boundary conditions" is assumed to be finite. In actuality, the boundary conditions vary continuously along Γ , therefore these series should be infinite and should produce identical results. In a numerical solution, only a finite number of samples of the boundary conditions are used, in conjunction with a finite series approximation.

The two techniques mentioned above have a common drawback, namely, the order of the series depends on the number of sampling points. One would like to use an approximating function that does not have this dependency so that the solution can be generated in a straightforward manner. Green's function fulfills such a requirement. It is a characteristics solution to the governing equation of the system, Laplace's equation in this case. All BEM formulations rely on Green's function for solutions.

The BEM may be classified into two different categories according to its formulations: the direct method and the indirect method. The indirect method is also known as the potential, source, or classic method. The direct method also known as Green's third identity for Laplace problems and Somigliana's formula for elasticity [38].

The direct method is so called because the unknown boundary parameters are expressed directly in terms of the known parameters, namely the boundary conditions.

For the indirect method, an intermediary solution must be explicitly obtained before the unknown boundary parameters can be computed. On computational efficiency, both methods are approximately the same – they both require a matrix multiplication and a matrix inversion (solution), only the order of these operations is different.

The direct method is preferred by many users because its error-weighting method is similar to that of the finite element method and also because of its alleged better accuracy. With the absence of intermediary sources, it is readily interfaced with the finite element method or the finite difference method. The indirect method, on the other hand, is conceptually simpler. Also, parameters of various types are readily extractable or implementable, as it will be shown.

From a physical standpoint, the indirect method is actually more “direct” than the direct method because it models what goes on physically at the boundary, such as the application of electrical charges or loading forces. Since it is the method that we use, we shall begin with a description of the indirect method.

3.1.1 Indirect Method

Consider a two-dimensional potential problem. Suppose that the potential distribution in the enclosed region, Ω , shown in Fig.3.1 is to be modeled by distributed charges on the boundary, Γ , [39]. The potential generated by these sources will be

$$\Phi(p) = \int_{\Gamma} \rho(q) \phi^*(p - q) d\Gamma \quad (3.7)$$

where Φ is the potential at p , ρ is the charge distribution and $\phi^*(p - q) = \log |p - q|$ is the logarithmic potential at p due to a point charge at q . $\phi^*(p - q)$ is Green's function for Laplace problems because it satisfies Laplace's equation everywhere except at the source point:

$$\nabla^2 \phi^*(q) = 2\pi \delta(q) \quad (3.8)$$

where $\delta(q)$ is a Dirac delta function. Eq. 3.7 is commonly referred to as Fredholm's integral equation of the first kind. The function ϕ^* is called a kernel or a kernel function

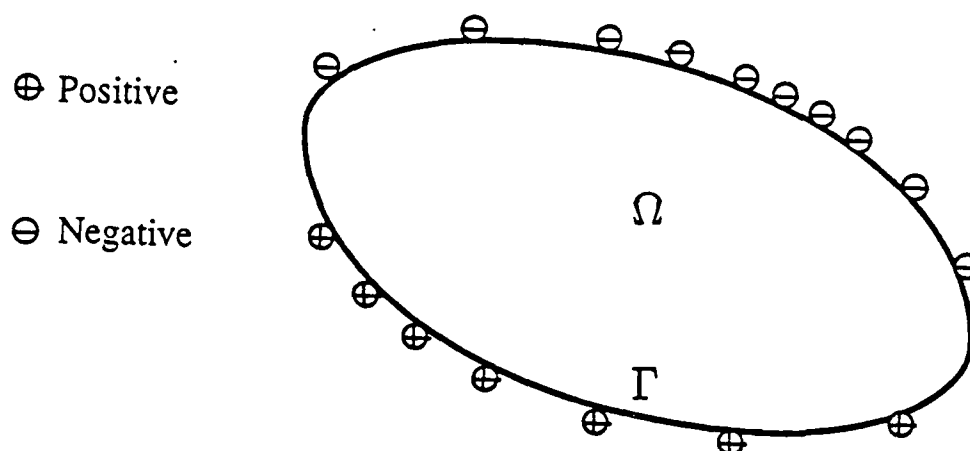


Figure 3.1: Source Distribution Technique for Potential Problems

and ρ is an unknown function.

Let's assume that a solution to the source distribution ρ has been determined. The potential it generates will be correct on Γ and inside Ω . Moreover the electric field can be readily obtained, as shown below in both vectorial and indicial forms:

$$\begin{aligned}
 \mathbf{E} &= -\nabla\Phi \\
 &= -\int_{\Gamma} \rho(\mathbf{q}) \nabla\phi^*(\mathbf{p} - \mathbf{q}) d\Gamma \\
 E_i &= -\frac{\partial}{\partial x_i} \Phi \\
 &= -\int_{\Gamma} \rho(\mathbf{q}) \frac{\partial}{\partial x_i} \phi^*(\mathbf{p} - \mathbf{q}) d\Gamma
 \end{aligned} \tag{3.9}$$

where we have omitted the dielectric constant for simplicity.

If, instead of potential, electric field is specified at the boundary, we have a Neumann problem. As the boundary condition is given by $\partial\Phi/\partial n$, the equation for deter-

mining the source distribution is

$$\frac{\partial \Phi}{\partial n} = -\mathbf{E} \cdot \hat{\mathbf{n}} \quad (3.10)$$

$$= \int_{\Gamma} \rho(\mathbf{p}) \frac{\partial}{\partial n} \phi^*(\mathbf{p} - \mathbf{q}) d\Gamma \quad (3.11)$$

Note that a Neumann problem is ill-specified in the sense that if $\Phi(\mathbf{p})$ is a solution to the potential distribution, $\Phi(\mathbf{p}) + c$ where c is an arbitrary constant is also a permissible solution. (As a homogeneous solution, c does not contribute to the boundary condition $\partial\Phi/\partial n$.) To guarantee a unique solution, an additional normalization equation on ρ is required:

$$\int_{\Gamma} \rho(\mathbf{q}) d\Gamma = 0. \quad (3.12)$$

It has the effect of preassigning a value for c .

What has been described is basically the boundary element method of simple-source distribution [38,40]. It is also called the single-layer formulation. For Dirichlet problems, the double-layer method of the following form is sometimes employed [38]:

$$\Phi(\mathbf{p}) = \int_{\Gamma} \rho(\mathbf{q}) \frac{\partial}{\partial n_q} \phi^* \log(\mathbf{p} - \mathbf{q}) d\Gamma \quad (3.13)$$

$$= \int_{\Gamma} \rho(\mathbf{q}) \frac{\partial}{\partial n_q} \phi^*(\mathbf{p} - \mathbf{q}) + \pi \rho(\mathbf{p}) \quad (\text{on } \Gamma) d\Gamma \quad (3.14)$$

where $\partial/\partial n_q$ is the normal derivative at $\mathbf{q}$ with respect to $\mathbf{q}$. The π term in Eq. 3.14 results from the integration over the singularity at $\mathbf{q} = \mathbf{p}$. The name "double layer" comes from the observation that $\partial\phi^*/\partial n_q$ is potential due to a dipole charge whose axis is normal to the surface – hence it appears that the domain is enclosed by two layers of boundary charges.

Eq. 3.14 is known as a Fredholm integral equation of the second kind. It produces a diagonally dominant matrix ideal for numerical calculations. It can also be solved very efficiently using the Liouville-Neumann series [41]. However the kernel functions become highly singular if we attempt to compute $\nabla\Phi$ numerically. Thus our application uses the single source formulation.

Poisson's Equation

The Laplace condition implies the domain is source-free; i.e. there are no electrical charges, heat sources, or diffusing species sources for the respective systems. If the domain has sources, Poisson's equation applies:

$$\nabla^2 \Phi(p) = \rho(p) \quad (3.15)$$

where ρ is the source density. Consequently, Eqs. 3.7 and 3.14 will have an additional expression to account for the domain contributions:

$$\frac{1}{2\pi} \int_{\Omega} \rho(q) \phi^*(p - q) d\Omega. \quad (3.16)$$

The Poisson condition often arises when the domain is nonhomogeneous. If that is the case, the domain source ρ is dependent on the boundary condition.

3.1.2 Direct Method

To derive the direct method for potential problems, we start with the divergence theorem:

$$\int_{\Omega} \nabla \cdot \mathbf{F} d\Omega = \int_{\Gamma} \mathbf{F} \cdot \hat{\mathbf{n}} d\Gamma. \quad (3.17)$$

If we subtract the solution for $\mathbf{F} = \psi \nabla \phi$ from the solution for $\mathbf{F} = \phi \nabla \psi$, where ϕ and ψ are scalars, we obtain Green's second identity:

$$\int_{\Omega} (\phi \nabla^2 \psi - \psi \nabla^2 \phi) d\Omega = \int_{\Gamma} \left(\phi \frac{\partial \psi}{\partial n} - \psi \frac{\partial \phi}{\partial n} \right) d\Gamma. \quad (3.18)$$

If we replace ϕ and ψ with ϕ^* and Φ defined earlier for the indirect method, we obtain the direct formulation:

$$-2\pi \Phi(p) = \int_{\Gamma} \left[\phi^*(q - p) \frac{\partial \Phi(q)}{\partial n_q} - \frac{\partial \phi^*(q - p)}{\partial n_q} \Phi(q) \right] d\Gamma - \int_{\Omega} \phi^*(q - p) \nabla^2 \Phi(q) d\Omega, \quad (3.19)$$

which is better known as Green's third identity or Green's formula. If the domain is source-free, the domain integral on the right-hand side of Eq. 3.19 can be removed.

To determine the unknown boundary parameters, p is set to Γ and Φ is expressed in terms of $\partial\Phi/\partial n$ or vice versa ². One advantage of Green's formula is that Eq. 3.19 does not contain a source function. On the other hand, it is tedious to determine $\nabla\Phi$ inside Ω - it requires the second derivatives of ϕ^* to do so. These highly singular second derivatives are not needed by the indirect method.

3.1.3 Periodic Kernels

So far, the form of domain that we are concerned with is a simple, closed region. There are many situations in which one would like to work with a periodic structure that extends from $x = -\infty$ to $x = \infty$. In this configuration, the upper boundary Γ_u and the lower boundary Γ_l are separated and never touch. The solution for Φ is given by

$$\Phi(p) = \int_{\Gamma_u, -\infty}^{\infty} \rho_u(p) \phi^*(q - p) d\Gamma + \int_{\Gamma_l, -\infty}^{\infty} \rho_l(p) \phi^*(q - p) d\Gamma \quad (3.20)$$

Assuming the dimensions are normalized, the following conditions will always be satisfied:

$$\begin{aligned} \Phi(p) &= \Phi(p + 2n\pi\hat{x}) \\ \rho(p) &= \rho(p + 2n\pi\hat{x}) \end{aligned}$$

for any integer n . Given its periodic nature, the indefinite integral can be reduced to an integral over a period:

$$\int_{-\infty}^{\infty} \rho(q) \phi^*(p - q) d\Gamma = \int_{-\pi}^{\pi} \rho(q) \tilde{\phi}^*(p - q) d\Gamma$$

by introducing a periodic kernel:

$$\begin{aligned} \tilde{\phi}^*(p) &= \sum_{n=-\infty}^{\infty} \phi^*(p + 2n\pi\hat{x}) \\ &= \frac{1}{2} \log [2 (\cosh(y) - \cos(x))], \end{aligned}$$

²This description is oversimplified. The left-hand side of Eq. 3.19 actually contains a scale factor that depends on the shape of the boundary. See [38,40,42] for details.

which is obtained from the real part of this complex-variable expression [41]:

$$\Re \left\{ \log \left[\sin \frac{1}{2}(x + iy) \right] \right\}.$$

The singularities of $\tilde{\phi}^*$ are located at $2n\pi x$ for any integer n . By applying Taylor series expansion on $\cosh(y)$ and $\cos(x)$,

$$\cosh(y) \approx 1 + y^2/2;$$

$$\cos(x) \approx 1 - x^2/2,$$

it can be shown that $\tilde{\phi}^*$ reduces to the non-periodic version near the origin:

$$\tilde{\phi}(p \approx 0) \approx \phi^*(p).$$

For $|y| \gg 0$, we have this approximation:

$$\tilde{\phi} \approx \frac{1}{2}|y|$$

which is Green's function for one-dimensional Laplace problems. The far-field analysis is a very useful tool because it deals with a simplified one-dimensional behavior that reveals the source activities at $y = 0$.

3.1.4 Source Formulation for Oxidant Diffusion

Having covered the principles of the BEM for Laplace problems, we now describe its application on modeling oxidant diffusion.

First, it will be shown that diffusion is a Laplace problem. Consider Fick's second law of diffusion:

$$\frac{\partial C}{\partial t} = -\nabla \cdot \mathbf{F}$$

In steady state, $\partial C / \partial t = 0$. Thus the flux is divergence-free:

$$\nabla \cdot \mathbf{F} = -\nabla(D\nabla C) \quad (3.21)$$

$$= 0. \quad (3.22)$$

If D is uniform throughout the oxide, as it will be initially assumed, Eq. (2.12) is reduced to Laplace's equation:

$$\nabla^2 C = 0. \quad (3.23)$$

With the indirect BEM method, the oxidant distribution is given by this statement,

$$C(p) = \int_{\Gamma} \rho(q) \phi^*(p - q) d\Gamma, \quad (3.24)$$

where ρ is interpreted as the 'source' or 'sink' for oxidants.

In the literature, parameters for oxidation are commonly given in terms of B and B/A , not in C^* , k_s , and D . By multiplying all C 's with k_s/N and redefining ρ in Eq. 3.24 to be $\rho k_s/N$, the oxidant diffusion problem may be strictly expressed in terms of B/A and B . Eq. 3.24 now becomes an oxide 'growth rate' distribution equation

$$G(p) = \int_{\Gamma} \rho(q) \phi^*(p - q) d\Gamma \quad (3.25)$$

where $G(p)$ is the oxide growth rate if silicon is present at p .

The boundary conditions for the free oxide surface and the oxide-silicon interface are replaced by

$$G = \frac{B}{A} \quad (3.26)$$

and

$$\frac{A}{2} \frac{\partial G}{\partial n} = G. \quad (3.27)$$

respectively. The Neumann boundary condition for the nitride interface remains unchanged:

$$\frac{\partial G}{\partial n} = 0. \quad (3.28)$$

Oxidant diffusion is a mixed boundary value problem involving three different types of boundary conditions. The source distribution technique is flexible in dealing with such a problem. We observe that both Eqs. 3.7 and 3.11 share the same source ρ , thus either one of them may be used on different parts of the boundary to describe the problem, depending on whether the boundary conditions are Dirichlet or Neumann. A linear combination of both produces the Robin boundary condition.

3.2 Biharmonic Problems

The oxide motion process, be it elastic deformation or viscous flow, can be shown to be a boundary value problem related to the Laplace problem. The solution to an elastic deformation or incompressible viscous flow in two-dimension is biharmonic or Bi-Laplacian:

$$\nabla^4 \Psi = 0 \quad (3.29)$$

where the bi-Laplace operator is given as

$$\begin{aligned} \nabla^4 &= (\nabla^2)^2 \\ &= \frac{\partial^4}{\partial x_1^4} + 2 \frac{\partial^4}{\partial x_1^2 \partial x_2^2} + \frac{\partial^4}{\partial x_2^4} \end{aligned}$$

for two dimensions.

Like a Laplace problem, a biharmonic problem is also a BVP whose solution is uniquely determined by the boundary conditions. The specification of the solution requires a pair of boundary conditions from the list: $\Psi, \partial\Psi/\partial n, \nabla^2\Psi, \partial(\nabla^2\Psi)/\partial n$. Once again, one must be careful in picking the proper pair. If Ψ is not uniquely specified anywhere on the boundary, the system becomes ill-defined because $\Psi + c$, where c is constant, is also a permissible solution.

Laplace's equation can be considered a special case of the biharmonic problem. If a function f is harmonic:

$$\nabla^2 f = 0,$$

it always satisfies the bi-Laplace equation:

$$\begin{aligned} \nabla^4 f &= \nabla^2(\nabla^2 f) \\ &= \nabla^2 0 \\ &\equiv 0 \end{aligned}$$

A Laplace problem may also be solved as a biharmonic problem by using $\nabla^2\Psi = 0$ as

one of the boundary conditions on Γ . Since $\nabla^2 \Psi$ is Laplacian, $\nabla^2 \Psi$ is zero everywhere inside Ω , thus Ψ also satisfies Laplace's equation.

At one time, nonlinear oxidant diffusion was implemented as a biharmonic problem. The idea is that nonhomogeneity in the diffusion coefficient results in pseudo domain-sources. By assuming the pseudo domain-source density to be Laplacian, the system becomes biharmonic, hence domain calculations can be avoided. However, only passable results were obtained.

3.2.1 Biharmonic Kernel Functions

Biharmonic kernel functions can be readily generated from the harmonic counterparts. We note that if $f(x, y)$ is harmonic, then

- $f(x, y)$ itself is also biharmonic, as noted earlier.
- $xf(x, y)$ is biharmonic, since $\nabla^4(xf(x, y)) = 2\nabla^2 \frac{\partial f}{\partial x} = 0$.
- Similarly, $yf(x, y)$ is biharmonic.
- $(x^2 + y^2)f(x, y)$ is also biharmonic, because $\nabla^4(x^2 + y^2)f(x, y) = \nabla^2[4f + 4(x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y})] = 0$.

There are more varieties than the Laplace system. As it will be shown later, we will only dealing with this pair:

$$\begin{aligned}\phi_A^* &= x \log |r| - x \\ &= x\phi - x \\ \phi_B^* &= y \log |r| \\ &= y\phi,\end{aligned}$$

where the term x is added to ϕ_A^* to make it compatible with the period version. The corresponding periodic version is given by:

$$\begin{aligned}\bar{\phi}_A^* &= \sum_{n=-\infty}^{\infty} (x + 2n\pi) \log |r + 2n\pi\hat{x}_1| \\ \bar{\phi}_B^* &= \sum_{n=-\infty}^{\infty} y \log |r + 2n\pi\hat{x}_1| \\ &= y\bar{\phi}^*.\end{aligned}$$

There is no simple functional form for $\bar{\phi}_A^*$. But as discussed in a previous article [18], its derivatives do:

$$\begin{aligned}\frac{\partial \bar{\phi}_A^*}{\partial x} &= \frac{1}{2} \log[2(\cosh(y) - \cos(x))] - \frac{1}{2} y \frac{\sinh(y)}{\cosh(y) - \cos(x)} \\ &= \bar{\phi} - y \frac{\partial \bar{\phi}}{\partial y}\end{aligned}\tag{3.30}$$

$$\begin{aligned}\frac{\partial \bar{\phi}_A^*}{\partial x} &= \frac{1}{2} y \frac{\sin(x)}{\cosh(y) - \cos(x)} \\ &= y \frac{\partial \bar{\phi}}{\partial x}.\end{aligned}\tag{3.31}$$

We note that, for both periodic and aperiodic versions, the first and second derivatives of ϕ_A , ϕ_B , and ϕ share many common terms. Thus it is appropriate at this point to introduce a set of basis functions consisting of

$$\phi, \quad \phi_{,1}, \quad \phi_{,2}, \quad x_2\phi_{,1}, \quad x_2\phi_{,2}, \quad x_2\phi_{,11}, \quad x_2\phi_{,12}$$

where indicial notation is used and the asterisk (*) is dropped for simplicity. The subscript “,*i*” denotes derivatives in the *i* direction; thus,

$$x_2\phi_{,12} \equiv y \frac{\partial^2 \phi}{\partial x \partial y}.$$

These expressions are combined in various ways to generate the desired Green's functions for diffusion, viscous flow, elastic deformation and viscoelastic flow.

3.2.2 Generic BEM for Oxide Motion

We are working with three oxide motion models. Except for the kernels, their biharmonic BEM formulation are similar in most respects. For this reason, we will present a template BEM framework here, delaying the derivation of the kernels until we discuss the mathematics of the appropriate motion model.

Biharmonic BEM resembles Laplace BEM in many ways. The displacement vector u parallels the scale potential Φ . The surface traction p is similar to $\partial\Phi/\partial n$. The stress tensor σ_{ij} is analogous to $\nabla\Phi$. There are also the direct and indirect methods for biharmonic BEM. Somigliana's formula parallels Green's formula. The vector potential method described below is an extension to the scalar source used in the oxidant diffusion modeling.

Vector Potential Method

In indicial notation, the displacement field produced by boundary sources is given by

$$u_i(p) = \int_{\Gamma} \rho_j(q) u_{ij}^*(q - p) d\Gamma \quad (3.32)$$

where σ 's are the "loading" sources and u_{ij}^* is the displacement in the x_i direction due to a loading source in the x_j direction.

Similarly, the formula for stress is

$$\sigma_{ij}(p) = \int_{\Gamma} \rho_k(q) \sigma_{ijk}^*(q - p) d\Gamma. \quad (3.33)$$

Using the definition $p_i = n_j \sigma_{ij}$ (Eq. 2.16), we obtained the statement for surface traction:

$$p_i(p) = \int_{\Gamma} \rho_j(q) p_{ij}^*(q - p) d\Gamma. \quad (3.34)$$

Conceptually, ρ_j defines loading forces at the boundary, as shown in Fig. 3.2.

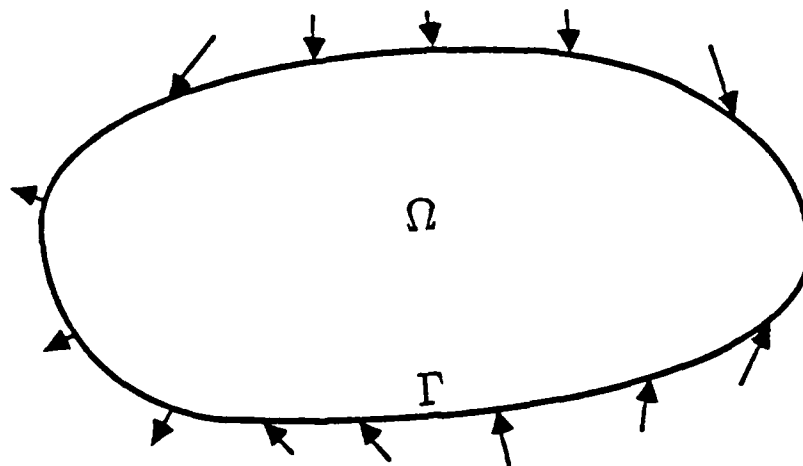


Figure 3.2: Vector Potential Formulation. The source density ρ_i defines loading forces applied at the boundary Γ .

Somigliana's Formula

Somigliana's formula is the foundation for the direct method on elasticity. It is of the form:

$$u_i(\mathbf{p}) = \int_{\Gamma} u_{ij}^* p_j(\mathbf{p} - \mathbf{q}) + p_{ij}^* u_j(\mathbf{p} - \mathbf{q}) d\Gamma + \int_{\Omega} u_{ij}^* (\mathbf{p} - \mathbf{q}) b_j(\mathbf{p} - \mathbf{q}) d\Omega \quad (3.35)$$

where the normal direction $\hat{n}$ (embedded in p_{ij}^*) refers to the dummy boundary point $\mathbf{q}$. As we can see, the procedure to obtain stress information for interior points is rather cumbersome – it requires the derivatives of u_i , and that means a new set of kernel functions: u_{ijk}^* and p_{ijk}^* .

3.2.3 Incompressible Slow Viscous Flow

In solving an incompressible flow problem, it is customary to express the velocity vector in terms of a scale parameter called stream function. The x and y components of the velocity vector $\mathbf{v} = v_x \hat{\mathbf{x}} + v_y \hat{\mathbf{y}}$ is related to the stream function Ψ by

$$v_x = \frac{\partial \Psi}{\partial y} \quad (3.36)$$

$$v_y = -\frac{\partial \Psi}{\partial x} \quad (3.37)$$

By this arrangement, $\bar{\mathbf{v}}$ always satisfies the incompressibility equation:

$$\begin{aligned} \nabla \cdot \mathbf{v} &= \frac{\partial^2 \Psi}{\partial x \partial y} - \frac{\partial^2 \Psi}{\partial x \partial y} \\ &\equiv 0, \end{aligned}$$

provided Ψ is a smooth function in the domain of interest.

It is known that the stream function for incompressible viscous flow is biharmonic [43]:

$$\nabla^4 \Psi = 0$$

and that the hydrostatic pressure is a harmonics conjugate function to the vorticity $\omega = -\nabla^2 \Psi$:

$$\begin{aligned} \eta \frac{\partial \omega}{\partial x} &= \frac{\partial P}{\partial y} \\ \eta \frac{\partial \omega}{\partial y} &= -\frac{\partial P}{\partial x}. \end{aligned} \quad (3.38)$$

Thus the stress tensor is given by:

$$\sigma_{ij} = \eta \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) - \delta_{ij} P. \quad (3.39)$$

Now we select $-\phi_B^*$ and ϕ_A^* as Green's functions corresponding to force loading in the $\hat{x}_1$ and $\hat{x}_2$ directions respectively. The complete set of kernel functions are given below:

$$v_{11} = -(x_2 \phi_{,2} + \phi) \quad (3.40)$$

$$v_{12} = x_2 \phi_{,1} \quad (3.41)$$

$$\sigma_{111} = -2(x_2 \phi_{,12} + 2\phi_{,1})\eta \quad (3.42)$$

$$\sigma_{112} = 2(x_2 \phi_{,11} - \phi_{,2})\eta \quad (3.43)$$

$$\sigma_{122} = 2x_2 \phi_{,12}\eta \quad (3.44)$$

$$v_{21} = x_2 \phi_{,1} \quad (3.45)$$

$$v_{22} = (x_2 \phi_{,2} - \phi) \quad (3.46)$$

$$\sigma_{211} = 2(x_2 \phi_{,11} - \phi_{,2})\eta \quad (3.47)$$

$$\sigma_{212} = 2x_2 \phi_{,12}\eta \quad (3.48)$$

$$\sigma_{222} = -2(x_2 \phi_{,11} + \phi_{,2})\eta \quad (3.49)$$

where the first subscript digit denotes the direction of the loading force.

3.2.4 Elastostatic Deformation

According to Papkovitch and Neuber [38], a general solution u that satisfies the Cauchy-Navier equation (Eq. 2.32) can be generated from

$$u(p) = h(p) - \frac{1}{4(1-\nu)} \nabla(p \cdot h(p) + f(p)) \quad (3.50)$$

where

$$\nabla^2 h(p) = 0, \quad (3.51)$$

$$\nabla^2 f(p) = 0. \quad (3.52)$$

If we let $f = 0$ and $\vec{h} = |\mathbf{p}|^{-1} \hat{x}_1$, we obtain the fundamental solution for the elastostatic problem. Also known as Kelvin's solution, it consists of the stress and displacement fields generated by a point force in an infinite 3-dimensional medium. The body force is applied at the origin $\vec{r} = 0$ in the $\hat{x}_1$ direction, as it is evident from the fact that the Laplacian is violated by a Dirac delta function centered at that point:

$$\nabla^2 |\mathbf{p}|^{-1} \hat{x}_1 = -4\pi \delta(p) \hat{x}_1. \quad (3.53)$$

To get the solutions for loading forces applied in the other two principal directions, we change $\vec{h}$ to $|\mathbf{p}|^{-1}\hat{x}_2$ and $|\mathbf{p}|^{-1}\hat{x}_3$ accordingly.

3.2.5 Elastostatic Kernels

The two-dimensional Kelvin's solution is obtained by setting $f = 0$ and $\mathbf{h} = \log(|\vec{p}|)\hat{x}_1$ or $\log(|\vec{p}|)\hat{x}_2$. It is used as the kernel or Green's function for elastostatic BEM applications [42,40,18]. In indicial notations, the displacement and stress fields generated by point forces in the $\hat{x}_1$ and $\hat{x}_2$ directions are given below:

$$u_{11} = \frac{1}{4(1-\nu)}[(4\nu-3)\phi - x_2\phi_{,2}] \quad (3.54)$$

$$u_{12} = \frac{1}{4(1-\nu)}x_2\phi_{,1} \quad (3.55)$$

$$\sigma_{111} = \frac{E}{4(1-\nu)(1+\nu)}[(2\nu-3)\phi_{,1} - x_2\phi_{,12}] \quad (3.56)$$

$$\sigma_{112} = \frac{E}{4(1-\nu)(1+\nu)}[x_2\phi_{,11} - 2(1-\nu)\phi_{,2}] \quad (3.57)$$

$$\sigma_{122} = \frac{E}{4(1-\nu)(1+\nu)}[x_2\phi_{,12} + (1-2\nu)\phi_{,1}] \quad (3.58)$$

$$u_{21} = \frac{1}{4(1-\nu)}x_2\phi_{,1} \quad (3.59)$$

$$u_{22} = \frac{1}{4(1-\nu)}[(4\nu-3)\phi + x_2\phi_{,2}] \quad (3.60)$$

$$\sigma_{211} = \frac{E}{4(1-\nu)(1+\nu)}[x_2\phi_{,11} - 2\nu\phi_{,2}] \quad (3.61)$$

$$\sigma_{212} = \frac{E}{4(1-\nu)(1+\nu)}[x_2\phi_{,12} - (1-2\nu)\phi_{,1}] \quad (3.62)$$

$$\sigma_{222} = \frac{E}{4(1-\nu)(1+\nu)}[2(\nu-1)\phi_2 - x_2\phi_{11}]. \quad (3.63)$$

Again, the first digit in the subscript denotes the direction of the loading force, as in the viscous incompressible flow kernels.

3.3 Laplace Transform and Viscoelastic BEM

Elastostatic deformation is a static problem in which the solution does not change if the inputs are held constant. Incompressible viscous creeping flow is a steady-state problem that also exhibits similar behaviors. The main difference between them is the primary variable (that may be used as an input condition) is displacement for elasticity and rate of displacement (*i.e.* velocity) for flow. On the other hand, viscoelastic flow is a transient problem. Whether constant displacement or constant velocity is specified as a boundary condition, the stress field still changes with time.

At first glance, it seems that linear viscoelasticity may be not be solved in a reduced-dimension manner because of its transient nature. However, according to the correspondence principle on viscoelasticity, a viscoelastic problem may be modeled using an equivalent elastostatic problem. In simple terms, it says that the time-dependent stress-strain behavior in the bulk is related to the boundary conditions in a simple manner.

3.3.1 Laplace Transform

In initial value problems (IVPs), the solutions are defined by the initial conditions imposed on the systems. If the problems are given in terms of linear differential equations with constant coefficients, they are known as linear time-invariant (LTI) systems. The Laplace transform method is a convenient and indispensable tool for working with LTI systems. This technique converts a differential equation into a polynomial function of a transform parameter. After the unknown coefficients are calculated algebraically, the inverse transform produces the solution for the differential equation. The importance of the Laplace transform is highlighted by its conversion of time-domain convolution to transform space multiplication. This simplification assists the analysis of complex interconnected subsystems.

The definition of Laplace transform and its inverse are given in the following notation:

$$\begin{aligned}
 F(s) &= \mathcal{L}\{f(t)\} \\
 &= \int_{0^-}^{\infty} f(\tau) e^{-s\tau} d\tau \\
 f(t) &= \mathcal{L}^{-1}\{F(s)\} \\
 &= \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} F(s) e^{st} ds
 \end{aligned} \tag{3.64}$$

where i is the imaginary number. The lower limit on the Laplace transform is set to 0^- so that impulse behaviors at $t = 0$ in $f(t)$ are included in the integration.

Important characteristics relevant to our application are given below:

$\mathcal{L}\{a + b\} = \mathcal{L}\{a\} + \mathcal{L}\{b\}$	Linear
$\mathcal{L}\{da/dt\} = s\mathcal{L}\{a\} + \mathcal{L}\{a(0)\}$	Time Differentiation
$f_1(t) * f_2(t) = F_1(s) \times F_2(s)$	Time Convolution
$f(0^+) = \lim_{s \rightarrow \infty} sF(s)$	Initial Value Theorem
$f(\infty) = \lim_{s \rightarrow 0} sF(s)$	Final Value Theorem

where the definition of convolution integral is

$$f_1(t) * f_2(t) = \int_{-\infty}^{\infty} f_1(\tau) f_2(t - \tau) d\tau.$$

Actual evaluations of transform integrals are seldom required since table lookup is more convenient. Given in Table 3.1 are a list of formulae useful for our applications.

3.3.2 Viscoelastic Kernels

As stated in Chapter 2, the constitutive equation relating stress and strain is usually given in time-derivative form, as in Eq. 2.42. Using the Laplace transform technique, it can be converted to a polynomial expression of the transform parameter s :

$$P(s)\sigma = Q(s)\epsilon \tag{3.65}$$

Inverse Transform $F(s)$	Function $f(t)$
1	$\delta(t)$
$1/s$	$u(t)$
$1/s^2$	$u(t)t$
$1/(s+a)$	$e^{-at}u(t)$

Table 3.1: Laplace Transform Table

where

$$P = \sum_{i=0}^m p_i s^m$$

$$Q = \sum_{i=0}^n q_i s^n.$$

Note that we have made an implicit assumption that the system is initially at rest, i.e. $\sigma(t=0) = \epsilon(t=0) = \dot{\sigma}(t) = \dot{\epsilon}(t) \dots = 0$.

The strategy for finding the viscoelastic kernels is to employ expressions already derived for linear elasticity. Since most of them are stated in terms of Young's modulus, E , and Poisson's ratio, ν , it is desirable to relate the Laplace polynomials to these two parameters. The relevant constitutive equations given in Chapter 2 are reproduced here for comparison; they are

$$s_{ij} = 2G e_{ij}$$

$$\sigma_M = 3K \epsilon_M$$

for elasticity and

$$P_1 s_{ij} = Q_1 e_{ij}$$

$$P_2 \sigma_M = Q_2 \epsilon_M$$

for viscoelasticity. From them, the following relationships are obtained:

$$2G = \frac{Q_1(s)}{P_1(s)}$$

$$3K = \frac{Q_2(s)}{P_2(s)}.$$

Since $G = E/2(1 + \nu)$ and $K = E/3(1 - 2\nu)$, the expressions for E and ν are therefore obtained[36]:

$$E(s) = \frac{3Q_1(s)Q_2(s)}{Q_1(s)P_2(s) + 2P_1(s)Q_2(s)} \quad (3.66)$$

$$\nu(s) = \frac{P_1(s)Q_2(s) - Q_1(s)P_2(s)}{Q_1(s)P_2(s) + 2P_1(s)Q_2(s)}. \quad (3.67)$$

With these modified definitions $E(s)$ and $\nu(s)$, the solution for an elastic problem can be converted to a viscoelastic form by applying the inverse Laplace transform.

This conversion procedure also applies to the elastic kernels Eqs. 3.54–3.63 to generate a set of viscoelastic kernels. Notice that we can premultiply all these expressions with a scale factor $\mathcal{W}(s)$ before the inverse Laplace transform. This effectively modifies the behavior of the loading force. Hence we can tailor $\mathcal{W}(s)$ to generate a solution with desirable characteristics for our modeling effort. Typically, we set either the displacement or the force (but not both) at the load point to follow a step change or linear change in time. The driving force of oxidation is the oxide growth at the oxide-silicon interface. As a first order approximation, we assume that the growth rate is constant within each time step, therefore the injection velocity is also constant. This suggests that a constant-velocity formulation is best suited for our application. For some other applications, a constant-force loading may be more appropriate.

To obtain a constant-velocity loading, we note that at the load point ($\bar{r} \rightarrow 0$), u_{111} (or u_{222}) is dominated by the term in $\log |\bar{r}|$, which has a coefficient of

$$\frac{4\nu(s) - 3}{4(1 - \nu(s))}.$$

The premultiplier is set to

$$\mathcal{W}(s) = \frac{4(1 - \nu(s))}{4\nu(s) - 3} \cdot \frac{1}{s^2} \quad (3.68)$$

to convert the coefficient to $1/s^2$, which is the transform of a constant velocity ($\partial u / \partial t = c$). Putting it in Eqs. 3.54–3.63 and taking the inverse Laplace transform results in

the following expressions:

$$u_{11} = 7K_\alpha(1 - 2\nu)x_2\phi_{,2} - (x_2\phi_{,2} + \phi)t \quad (3.69)$$

$$u_{12} = -7K_\alpha(1 - 2\nu)x_2\phi_{,1} + x_2\phi_{,1}t \quad (3.70)$$

$$\sigma_{111} = [-K_\alpha(2x_2\phi_{,12} + 3\phi_{,1}) - K_\beta(x_2\phi_{,12} + 5\phi_{,1})]E \quad (3.71)$$

$$\sigma_{112} = [K_\alpha(2x_2\phi_{,11} - \phi_{,2}) + K_\beta(x_2\phi_{,11} - 4\phi_{,2})]E \quad (3.72)$$

$$\sigma_{122} = [K_\alpha(2x_2\phi_{,12} - \phi_{,1}) + K_\beta(x_2\phi_{,12} + 3\phi_{,1})]E \quad (3.73)$$

$$u_{21} = -7K_\alpha(1 - 2\nu)x_2\phi_{,1} + x_2\phi_{,1}t \quad (3.74)$$

$$u_{22} = -7K_\alpha(1 - 2\nu)x_2\phi_{,2} + (x_2\phi_{,2} - \phi)t \quad (3.75)$$

$$\sigma_{211} = [K_\alpha(2x_2\phi_{,11} - 3\phi_{,2}) + K_\beta(x_2\phi_{,11} + 2\phi_{,2})]E \quad (3.76)$$

$$\sigma_{212} = [K_\alpha(2x_2\phi_{,12} + \phi_{,1}) + K_\beta(x_2\phi_{,12} - 3\phi_{,1})]E \quad (3.77)$$

$$\sigma_{222} = [K_\alpha(2x_2\phi_{,11} + \phi_{,2}) + K_\beta(x_2\phi_{,11} + 4\phi_{,2})]E \quad (3.78)$$

where

$$\begin{aligned} K_\alpha &= \frac{6}{7} \frac{\eta}{E} [1 - \exp(-\frac{t}{\tau_\alpha})] \\ K_\beta &= \frac{2}{7} \frac{\eta}{E} [1 - \exp(-\frac{t}{\tau_\beta})] \\ \tau_\alpha &= \frac{3(3 - 4\nu)\eta}{E} \\ \tau_\beta &= \frac{2(1 + \nu)\eta}{E} \\ &= \frac{\eta}{G} \end{aligned}$$

K_α and K_β are analogous to the charging of the capacitor C in a simple R-C circuit. Like the capacitor that discharges when the applied voltage is removed, K_α and K_β change to decaying exponential functions with time constants τ_α and τ_β respectively when the loading is stopped.

In general, the two relaxation time constants are different. The second one, τ_β , is readily identified as shear relaxation. The origin of τ_α is less clear; probably it has to

do with combined volume-shear relaxation as evident from the fact that it becomes the same as the shear relaxation time when the material is incompressible ($\nu = 0.5$). Stress fields associated with K_α and K_β satisfy the force conservation equation Eq. 2.17 independently. This condition arises from the fact that K_α and K_β decay differently.

We expect that, as the ratio η/G approaches ∞ or 0, the viscoelastic kernels reduce to that of elastic deformation or viscous incompressible flow respectively.

For the first case, the limit $\eta/G \rightarrow \infty$ is obtained by keeping G at a finite value and letting η go to infinity. The exponential functions are approximated with Taylor series and the following expressions are obtained

$$\begin{aligned} \frac{\eta}{E}(1 - e^{At}) &\approx \frac{\eta}{E}At \\ &= \frac{1}{3(3 - 4\nu)}t \\ \frac{\eta}{E}(1 - e^{Bt}) &\approx \frac{\eta}{E}Bt \\ &= \frac{1}{2(1 + \nu)}t. \end{aligned}$$

for the appropriate terms in Eqs. 3.69–3.78. In this regime, the viscoelastic kernels are the elastic kernels given in Eqs. 3.54–3.63 scaled by $4(1 - \nu)t/(4\nu - 3)$.

In the second case, the limit $\eta/G \rightarrow 0$ is reached by fixing η at a finite value and letting G go to infinity. For any nonzero t , no matter how small, the exponentials are zero. We find that the viscoelastic kernels simplify to the viscous incompressible flow kernels (Eqs. 3.40–3.49). Note that ν does not appear in the final expressions. The reason is that the elastic moduli are set to be infinitely large in this simplification ($G \rightarrow \infty$), thus the material is incompressible irrespective of ν .

The viscoelastic kernels are different from those of elasticity and viscous flow in that they have an additional parameter t . This parameter is set to the time-step size Δt in numerical simulations.

As we can see, the constant-velocity (linear-displacement) loading viscoelastic ker-

nels approach is a flexible formulation. Because the stress relaxation time can be pushed to the extremes, this viscoelastic formulation is also suitable for modeling elastostatic deformation and viscous incompressible flow.

3.4 Modeling Nonlinear Bulk Behaviors

The integral equation in the BEM is nothing more than a continuous summation of contributions from boundary sources. For this superposition technique to be valid, the domain has to be homogeneous, *i.e.* bulk parameters such as diffusivity and viscosity have to be uniform throughout the bulk, as originally assumed in the derivations of those Green's functions. It has been mentioned earlier in Chapter 2 that stress alters those parameters, rendering the domain nonhomogeneous. This section describes a formalism for handling the nonlinear effects with the BEM.

The concept comes from the observation that both the diffusion flux and stress-strain equations are of the form

$$A = cB. \quad (3.79)$$

where

	Diffusion	Elasticity
A	F	σ
B	∇C	ϵ
c	D	E

For notational simplicity, the strain and stress parameters are assumed to be scalars.

The similarity suggests that both processes can be treated in a unified manner. For generality, we will discuss the issues in terms of the representative parameters A , B , and c .

Nonhomogeneity or nonlinearity implies the parameter c is not uniform within the

domain. c can be rewritten as

$$c = c_0 k(x, A, \dots). \quad (3.80)$$

As shown, the scale factor k may be a function of many parameters, including A , B and position x . To conform with the linear system format, Eq 3.79 is rewritten as

$$A = c_0 B + A^0, \quad (3.81)$$

where A^0 is $c_0(k - 1)B$. In a normal setting, A^0 is interpreted as a initial condition of A that does not have a associated B^0 value. If B is expressed in terms of A , this alternate statement is obtained:

$$B = \frac{1}{c} A \quad (3.82)$$

$$= \frac{1}{c_0} A + B^0 \quad (3.83)$$

where B^0 is the initial condition of B .

One can find familiar situations in mechanics where there is A without the accompany B , and vice versa. For instance, consider a freely-supported material undergone thermal expansion: it sees a nonzero thermal strain ϵ^T . But it has zero thermal stress σ^T because it is freely supported. On the other hand, if the material is clamped on all sides, then its ϵ^T is zero because it is not free to expand. Consequently, its σ^T is non-zero. It is said to have initial stress. For the diffusion problem, similar analogies are hardy to come up with. We note that in insulating materials, the increase in dielectric constant (beyond ϵ_0) is due to electronic polarization. If we can freeze the polarization when we remove the applied field, we will see a residual built-in field. Because dielectric constant in potential is analogous to diffusion coefficient in diffusion, we coin the term "built-in field" as the initial condition for Laplace problems.

3.4.1 Viscoplastic BEM

This unified framework for nonlinear diffusion and viscoelastic flow evolves from considerations on elastoplastic BEM method. Elastoplasticity is the permanent deformation

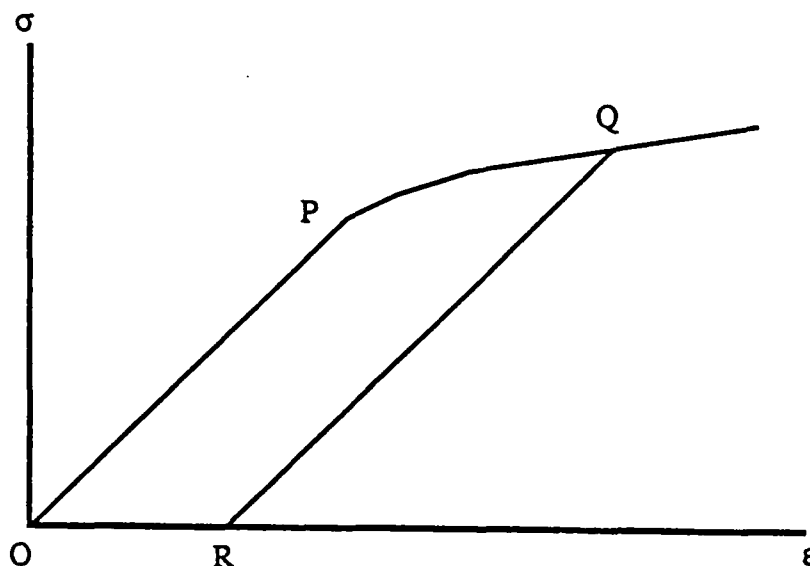


Figure 3.3: Elastoplasticity. Beyond the point P , the system is nonlinear. The effective E' is different from E of the linear regime.

in an elastic material due to large stress or strain. Typically, such an effect is observed only when stress exceeds the yield limit of the material. Fig. 2.5 from Chapter 2 is reproduced here as Fig. 3.3 to show a representative plastic stress-strain curve ³.

An unmodified BEM can only model the initial linear regime, as defined by OP in the figure. Beyond the point P , the stress-strain relationship is nonlinear. At Q , the effective E' is given by σ_Q/ϵ_Q . Because stress (or strain) is not likely to be uniform throughout the domain, different regions will have different effective E' . Thus the domain is not homogeneous.

The three common techniques for nonlinear elasticity are the modified body-force, initial stress, and initial strain formulations. Only the concepts of initial strain and initial stress will be described here as they are relevant to our work. For details on all these methods, readers are referred to the book authored by Banerjee and Butterfield

³In this discussion, we are not concerned with how the system returns to the unstressed state.

[40]. The treatment by Brebbia *et al* [42] is not as in-depth as the former.

In the "initial stress" formulation, the nonlinear stress components is expressed as [40]:

$$\sigma_{ij}^0 = \sigma_{ij}^e - \sigma_{ij}^{ep} \quad (3.84)$$

where σ^0 is the initial stress, σ_e is the elastic stress, and σ^{ep} is the elastoplastic stress that we seek. The equilibrium equation then becomes

$$\frac{\partial \sigma_{ij}^{ep}}{\partial x_j} = 0$$

or

$$\frac{\partial \sigma_{ij}^e}{\partial x_j} = \frac{\partial \sigma_{ij}^0}{\partial x_j}. \quad (3.85)$$

As far as the elastic component is concerned, there is a pseudo body force due to initial stress σ^0 . The displacement formula is modified to include this contribution:

$$u_i(p) = \int_{\Gamma} \rho_j u_{ij}^*(p - q) d\Gamma + \int_{\Omega} \frac{\partial \sigma_{jk}^0}{\partial x_k} u_{ij}^*(p - q) d\Omega. \quad (3.86)$$

The statement for elastoplastic stress is accordingly given by

$$\sigma_{ij}(p) = \int_{\Gamma} \rho_k \sigma_{ijk}^*(p - q) d\Gamma + \int_{\Omega} \frac{\partial \sigma_{kl}^0}{\partial x_l} \sigma_{ijk}^*(p - q) d\Omega - \sigma_{ij}^0. \quad (3.87)$$

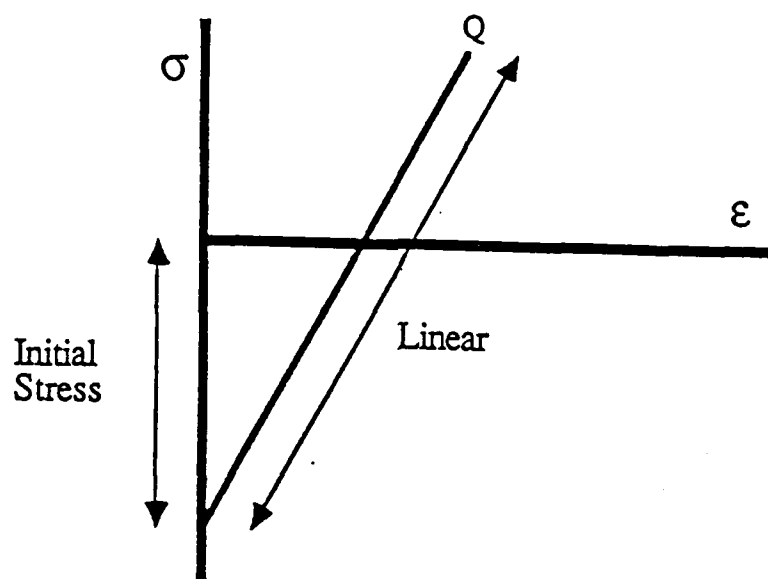
Note that the stress expression has a term in σ^0 that has no counterpart in the displacement formula. The domain integrals in both expressions are usually converted, using the divergence theorem, to a different form that utilizes σ_{ij}^0 as the domain source, instead of $\partial \sigma_{ij}^0 / \partial x_j$. With this transformation, we avoid taking derivatives of $\partial \sigma_{ij}^0$, which may have to be done numerically. However there is no need to do so in our particular formulation; hence we stay with the original expressions.

In the initial strain formulation, we modify strain instead of stress:

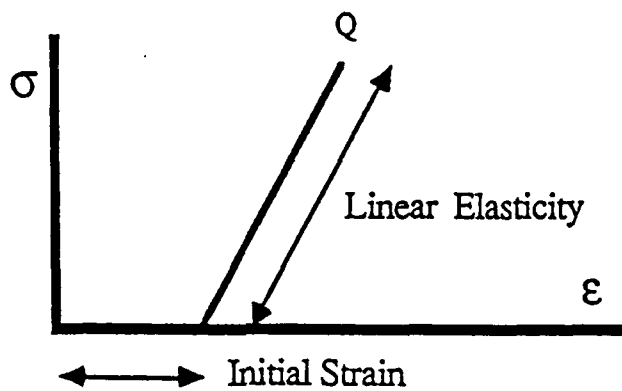
$$\epsilon_{ij}^0 = \epsilon_{ij} - \epsilon_{ij}^e \quad (3.88)$$

where ϵ_{ij}^0 is the initial strain, ϵ_{ij} is the total strain, and ϵ_{ij}^e is the elastic strain. Stress is given in terms of the elastic strain:

$$\sigma_{ij} = \frac{1}{E} \epsilon_{kl}$$



(a)



(b)

Figure 3.4: Modeling nonlinear elasticity with (a) the initial stress approach and (b) the initial strain approach.

The initial strain ϵ_{ij}^0 is then used in domain integrals similar to those for σ_{ij}^0 .

Fig. 3.4 illustrates how nonlinear elasticity is modelled with the initial stress and initial strain formulations. To reach the point Q defined in Fig. 3.3, the initial stress approach uses a vertical translation (stress without strain - initial stress) in conjunction with a linear deformation. Mathematically this is given by Eq. 3.81. Likewise, the initial strain formulation employs a horizontal translation (strain without stress - initial strain) and a linear deformation, as represented by Eq. 3.83.

3.5 Nonlinear Diffusion BEM

The nonlinear diffusion follows the initial stress method closely. The oxidant flux

$$\begin{aligned} \mathbf{F} &= -D\nabla C \\ &= -D_0 k_d \nabla C \\ &= \mathbf{F}^L + \mathbf{F}^0 \end{aligned} \tag{3.89}$$

is split into a linear (or homogeneous) part

$$\mathbf{F}^L = -D_0 \nabla C$$

and a built-in field

$$\mathbf{F}^0 = D_0(1 - k_d) \nabla C$$

where k_d is the diffusivity scale factor.

The flux conservation is on $\mathbf{F}$:

$$\nabla \cdot \mathbf{F} = \nabla \cdot \mathbf{F}^L + \nabla \cdot \mathbf{F}^0$$

Hence the linear part sees a pseudo domain source due to the built-in field:

$$\rho = -\nabla \cdot \mathbf{F}^L \tag{3.90}$$

$$= \nabla \cdot \mathbf{F}^0. \tag{3.91}$$

$$\tag{3.92}$$

The integral equation for C now becomes

$$C(p) = \int_{\Gamma} \rho(q) \phi^*(q-p) d\Gamma + \int_{\Omega} (\nabla_q \cdot F^0(q)) \phi^*(q-p) d\Omega \quad (3.93)$$

where the subscript in ∇_q indicates the operation is done with respect to the dummy parameter q , rather than with p . The integral equation for flux is accordingly given by

$$\begin{aligned} F &= D_0 \nabla C + F^0 \\ &= D_0 \int_{\Gamma} \rho(q) \nabla \phi^*(q-p) d\Gamma + D_0 \int_{\Omega} (\nabla_q \cdot F^0(q)) \nabla \phi^*(q-p) d\Omega + F^0 \\ &= D_0 k_d \left[\int_{\Gamma} \rho(q) \nabla \phi^*(q-p) d\Gamma + \int_{\Omega} (\nabla_q \cdot F^0(q)) \nabla \phi^*(q-p) d\Omega \right] \end{aligned} \quad (3.94)$$

3.6 Nonlinear Viscoelastic BEM

Our nonlinear viscoelastic BEM formulation is different from the elastoplastic or viscoplastic formulations because of its use of the Laplace transform technique. In the conventional approach, the deviatoric components in stress (or strain) is singled out for plastic deformation treatment, leaving the spherical components unchanged. In our problem, we are only concerned with the effect of stress on the viscosity. The viscosity parameter appears in the exponential functions K_A and K_B of the viscoelastic kernels Eqs. 3.69—3.78. Given that

$$\eta = \eta_0 k_{\eta} \quad (3.95)$$

where

$$k_{\eta} = \exp \left(\frac{-\sigma_M V_{\eta 1}}{kT} \right) \frac{\sigma_S V_{\eta 2} / kT}{\sinh(\sigma_S V_{\eta 2} / kT)} \quad (3.96)$$

as described in Chapter 2, we can expressively show the influence of k_{η} by writing the kernels of displacement and stress. For simplicity, we introduce two sets of variables to represent the effective scaling of k_{η} on the displacement and stress:

$$\begin{aligned} u_i(p) &= \int_{\Gamma} \rho_j(q) u_{ij}^*(q-p, \Delta t, k_{\eta}) d\Gamma \\ &= k_{ui}(p) u_i^L(p) \end{aligned}$$

$$\begin{aligned}\sigma_{ij}(p) &= \int_{\Gamma} \rho_k(q) \sigma_{ijk}^*(q - p, \Delta t, k_\eta) d\Gamma \\ &= k_{\sigma ij}(p) \sigma_{ij}^L\end{aligned}$$

where the superscript L denotes the solution to the linear system in which $k_\eta \equiv 0$.

The nonlinear stress must satisfy the force equilibrium condition:

$$\begin{aligned}\frac{\partial \sigma_{ij}}{\partial x_j} &= \frac{\partial}{\partial x_j} (k_\sigma \sigma_{ij}^L) \\ &= \frac{\partial k_\sigma}{\partial x_j} \sigma_{ij}^L + k_\sigma \frac{\partial \sigma_{ij}^L}{\partial x_j}.\end{aligned}$$

This results in a pseudo body force of

$$\begin{aligned}b_i &= \partial \sigma_{ij}^L / \partial x_j \\ &= \frac{1}{k_\sigma} \sigma_{ij}^L \frac{\partial k_\sigma}{\partial x_j}\end{aligned}\tag{3.97}$$

for the linear part.

The presentation so far has been analytical and general. How well this formulation performs depends critically on the numerical approximation technique.

3.7 Incompressible Materials

In certain numerical tests on ring structures, the loading force is arranged in a radial direction, as shown in Fig. 3.5. An anomalous behavior was found when Poisson's ratio is set to 0.5, *i.e.* when the material is incompressible. Outside the ring, no displacement and stress is seen (except for residual numerical errors). Inside the ring, only a high pressure is observed, the displacement remains zero. The inability to generate a radial symmetric solution reveals that Kelvin's solution that we have been using is deficient – it cannot model a certain legitimate behavior or geometric configuration.

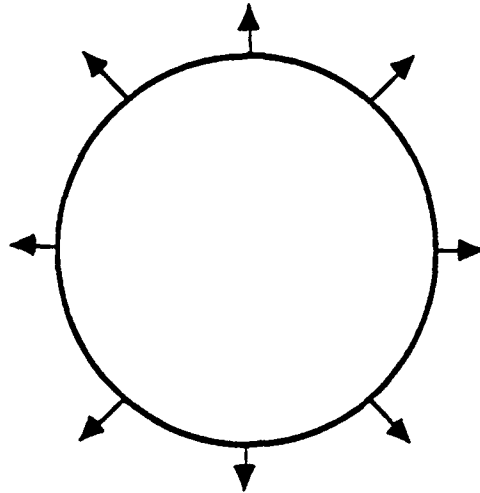


Figure 3.5: Radial loading forces. Neither stress nor displacement is observed outside the ring.

This is problematic for modeling any annulus structure because it has an outer circular boundary and an inner circular boundary. The inability of the inner boundary sources to influence the domain behavior means the matrix solution is singular or nearly so, producing very poor results. Granted, one would use a more realistic Poisson's ratio of 0.125, for instance. However, as we noted in a previous section, the material will automatically become incompressible as the viscosity decreases.

Note that the inner boundary needs not be circular for singular behavior to show up – abnormal results are also obtained when the inner boundary is rectangular.

The physical cause of the singular behavior can be understood by examining Fig. 3.6. Shown in part (a) is an incompressible annulus structure that is to be modeled. The inner hole and the exterior, which are not shaded, are of a different material that is totally compressible – free space. The way the indirect BEM han-

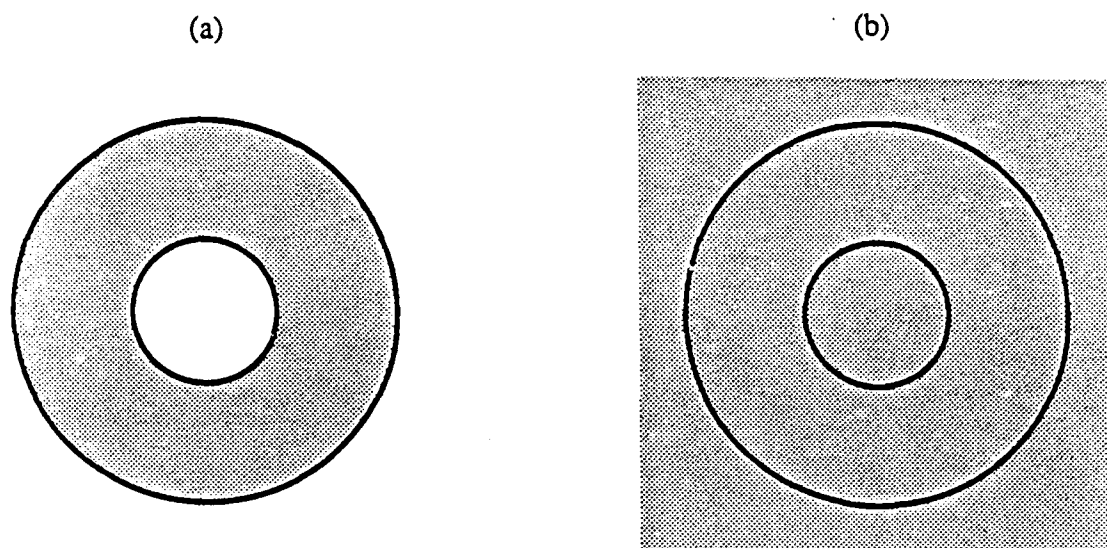


Figure 3.6: Source of incompressible singularity. (a) Actual problem. (b) In BEM implementation, the hole in the center is filled with the same incompressible material.

dealing with this problem is illustrated in part (b). The whole space is filled with the same incompressible material, as required by Kelvin's solution. The BEM applies loading force on the two circular lines that map to the actual boundaries in (a) so that the "boundary conditions" on these lines match the real ones. The problem now is that the hole is filled with an incompressible material. The Kelvin's solution only produces distortion, it does not introduce any new material or otherwise change the volume. Hence it cannot change the area of the hole, and this behavior is not compatible with the real situation. Thus a special solution is needed in order to be able to alter the size of the hole.

The special solution is an elastic potential function:

$$\Phi^o = \frac{1}{2} \log(x^2 + y^2).$$

The displacement field is given by

$$\mathbf{u}^{\circ} = \nabla \Phi^{\circ} \quad (3.98)$$

$$= \frac{x}{x^2 + y^2} \hat{x} + \frac{y}{x^2 + y^2} \hat{y} \quad (3.99)$$

and hence the stress is

$$\sigma_{11}^{\circ} = 2G \frac{y^2 - x^2}{(x^2 + y^2)^2} \quad (3.100)$$

$$\sigma_{12}^{\circ} = 0 \quad (3.101)$$

$$\sigma_{22}^{\circ} = 2G \frac{x^2 - y^2}{(x^2 + y^2)^2} \quad (3.102)$$

As we can see, there is no spherical deformation or stress associated with this function, except at the origin:

$$\epsilon_{11} + \epsilon_{22} = \nabla \cdot \mathbf{u}(\mathbf{p}) \quad (3.103)$$

$$= 2\pi\delta(\mathbf{p}). \quad (3.104)$$

where there is a source of material. The corresponding viscoelastic functions are obtained by scaling the displacement field with t and replacing $2G$ in the stress expressions with $2\eta[1 - \exp(-Gt/\eta)]$.

To include this special feature in the numerical solution, an additional condition on the system is needed. After experimenting with several options, it is decided that the best constraint is to specify the net hydrostatic pressure generated by the boundary sources be zero at the origin. With this enhancement, it is found that the solutions are very well behaved.

Chapter 4

Numerical Solutions

Solving a boundary value problem with the BEM entails the following steps. First, the boundary is suitably divided into segments or elements. The sources are assumed to take a certain shape distributions on these elements. A matrix is set up to define the sources in terms of the known boundary conditions. After the source values are obtained from the matrix solution, the unknown boundary and interior parameters can then be calculated.

In modeling thermal oxidation, the solution of oxidant diffusion is first computed to determine the oxide growth rate along the silicon interface. The injection of newly-created oxide into the bulk is calculated from the growth rate and desired time-step size. This information is then fed into the flow matrix equation that may represent oxide motion as elastic deformation, incompressible viscous flow, or viscoelastic flow. From the flow solution, the oxide displacement at the free surface is deduced. At this point the motion pattern of all boundary points are known, hence their locations can be updated. A new step can then begin. This loop continues until the desired oxide thickness or oxidation time is reached.

When the mechanical interaction of oxide with the nitride layer and effects of stress are considered, the numerical procedures become more complex and intertwined - ma-

trix solutions are more complicated, domain calculations are required, and iterations are embedded in other iterations. The flow chart described in the previous paragraph serves to give a clear outline of the underlying activities. The presentation of this chapter does not necessarily follow the order of program execution. Rather, the description is organized around logical groupings of functionalities. The segmentation of boundary and domain will first be discussed. After issues of line integration, matrix synthesis and solution will be dealt with. Methods for solving nonlinear equations will also be presented.

4.1 Boundary and Domain Segmentation

The first step in BEM numerical approximation is to discretize the curved boundaries by subdividing them into small segments or elements. This subdivision serves two purposes. The piece-wise straight-line approximation makes integral calculations more tractable. It also provides a suitable mechanism to prescribe a shape function for the source distribution on each segment.

Three basic boundary segmentation techniques are shown in Fig. 4.1; they are (a) constant, (b) linear, and (c) quadratic. In the constant element representation, the collocation point at which the boundary condition is specified is located in the middle of an segment. For the indirect method, the "constant" term refers to the fact that the source density along a segment is uniform. Its values takes a step jump from one segment to another, as depicted in Fig. 4.2a. For the direct method, the "constant" term refers to the weighting function on the boundary condition. In the linear element configuration, the collocation points are the endpoints of the segments. Source density is assumed to vary linearly from one collocation point to the adjacent one, as shown in Fig. 4.2b. Both "constant" and "linear" segments are straight lines. In the quadratic case, a boundary element is a curve defined by 3 nodes points. Since there are 3 collocation points, it is therefore possible to make a quadratic fit for the

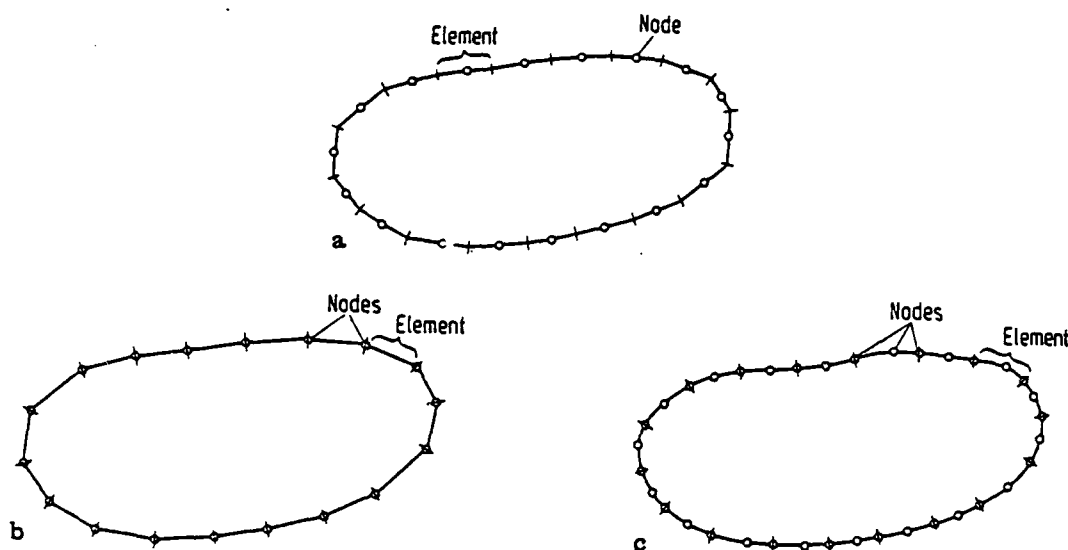


Figure 4.1: Boundary segmentation techniques: (a) constant element, (b) linear element, and (c) quadratic element. The "nodes" are the collocation points.

source distribution.

In principle, the smoother the distribution approximation is, the better the results will be. However, only the constant element is practical for the indirect method; the other two are actually only relevant to the direct method. The reason is that these segmentation schemes, with the exception of the quadratic approach, are just a polygonization of a geometry. Where two boundary segments are joined together, the surface curvature is infinite. In other words, we have a Liapunov surface that is not smooth [38]. Whenever the surface orientation of a boundary point is discontinuous, the indirect-method solutions are singular. No meaningful boundary condition can be imposed or extracted if a collocation point is placed there. Therefore collocation points are set at the center of each segments, as in the constant element scheme.

A linear-element approximation had been tried. To avoid surface normal discontinuity, the edge joining two boundary segment was replaced by an arc. The arc

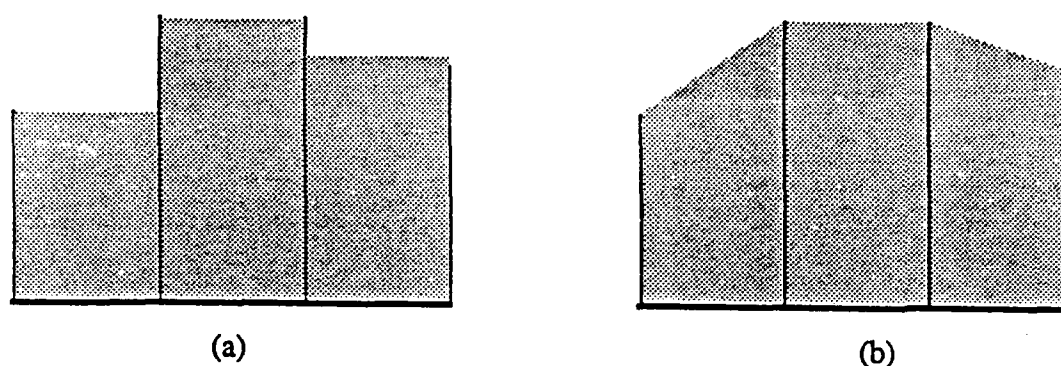


Figure 4.2: Source distributions: (a) constant element, and (b) linear element.

curvature, which is the reciprocal radius of a circle, is of a predefined value. An attempt to improve the “mid-point” collocation method had also been experimented. The source distribution was made to vary linearly from one collocation point to the next. Both approaches were abandoned because the minor improvement in accuracy could not justify the increased complexity in the solution technique. This point is particularly applicable to our oxidation modeling effort because we are trying to explore the usability of the boundary element technique, rather than to tune it to get improved computation performance.

4.1.1 Classes of geometries

As mentioned in Chapter 3, it is desirable to mirror a simulation structure in such a way that the upper boundary (*i.e.* the free surface and the silicon nitride mask) never touches the lower boundary (*i.e.* the silicon substrate). This is to avoid artificial boundary closure that can introduce local numerical errors in regions next to the

artificial wall. Two types of mirroring formats are supported. First, the mirroring of a structure at the left- and right-hand edges results in periodic symmetry. This form requires periodic kernels. The other mirrors one horizontal edge and one vertical edge to obtain a 2-fold symmetry; it uses the normal non-periodic kernels. A variation on the 2-fold symmetry is the 4-fold one that has an additional folding at a 45° line. All these configurations are depicted in Fig. 4.3.

Reduction in number of unknowns is achieved by exploiting symmetry features of the structures. The solution is only calculated for half a period, a quadrant, or half a quadrant respectively. Typically, a total of 40 to 60 segments are used for periodic and 2-fold symmetric structures, and 20 to 30 for the 4-fold symmetric case. The lower limit is due to accuracy consideration of the discrete approximation to continuous quantities while the upper limit is constrained by the computational speed.

Depending on the nature of the kernel functions, the source distributions are either symmetric or antisymmetric with respect to a symmetry line. The kernels for the oxide-flow problem work in pairs to generate a loading force. The ones specifying loading forces in the x direction produce a source distribution that is antisymmetric with respect to a vertical fold line ($\sigma(x) = -\sigma(-x)$) but symmetric with respect to a horizontal fold line ($\sigma(y) = \sigma(-y)$). The converse is true for the y -loading kernels. Moreover, in the 4-fold symmetry case, the source distribution of a "x" kernel is mirrored to that of a "y" kernel along the 45° line. For the oxidant problem, the kernel is scalar, therefore the source distribution is always symmetric with respect to the fold lines.

The periodic symmetric geometry is used for analyzing LOCOS and SWAMI structures. The trench structure has a long vertical perimeter, solving it as a periodic structure will require many boundary segments and slow down the computation speed significantly. Therefore we take advantage of the vertical wall and cut it in the middle into an upper-half and lower-half, and solve each of them as a 2-fold symmetric object. The 4-fold symmetry is primarily used for modeling Kao's experiments on cylindrical

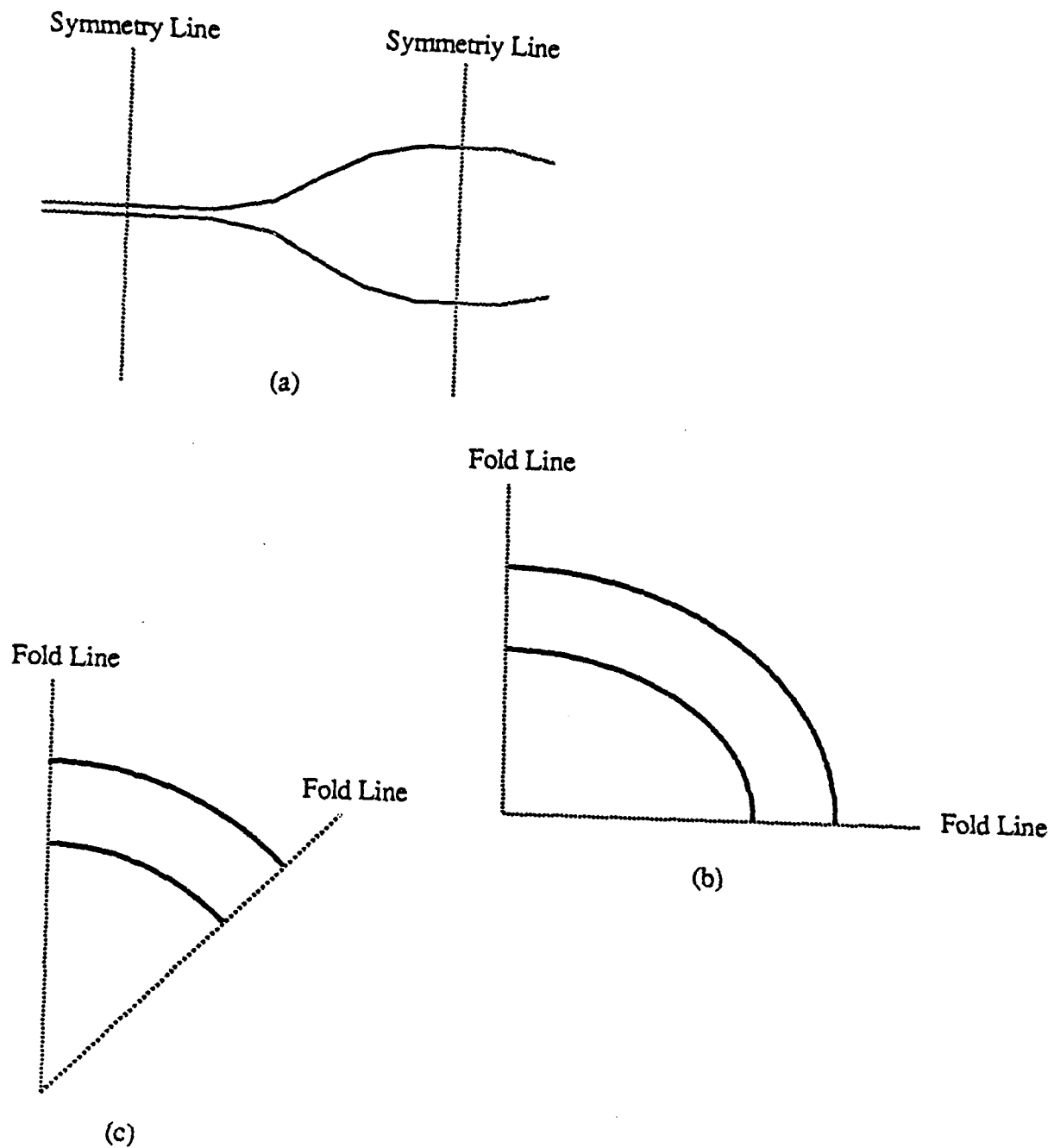


Figure 4.3: G

Geometry Types: (a) periodic symmetry, (b) 2-fold symmetry, (c) 4-fold symmetry.

structures.

4.1.2 Interior-Cell Generation

It is possible to subdivide a nonhomogeneous domain into piece-wise homogeneous partitions and solve them as a group of interconnected systems. This subregion technique is effective when the bulk parameters are different in a few large regions. It is used for modeling the mechanical interaction between oxide and nitride bodies. However, it is not suitable for modeling stress effects because many partitions may be needed to model the variations in bulk parameters accurately; this complicates the matrix solution techniques. Thus perturbation techniques based on interior cells are explored.

Because of the ever-changing domain, cell partitions are not specified in the input data but are determined by the program on the fly at every time step. To write a fully automated generator is a difficult task; only a rudimentary scheme has been used. This method takes advantage of the fact that upper and lower boundary segments are aligned and specified in pairs. Where interior cells are needed, it creates a column joining the segment pairs, as illustrated in Fig. 4.4a. Then it looks at the dimension of the column to determine the number of cells that should be created, using this rule:

$$n = \frac{\text{effective height}}{\text{effective width}}.$$

where the effective height is the average length of the two sides of the column, and the effective width is the average length of the top and bottom. The resulting cells are mostly quadrilaterals, as shown in Fig. 4.4b. Some fill-in triangular cells are created when the number of cells changes from one column to another.

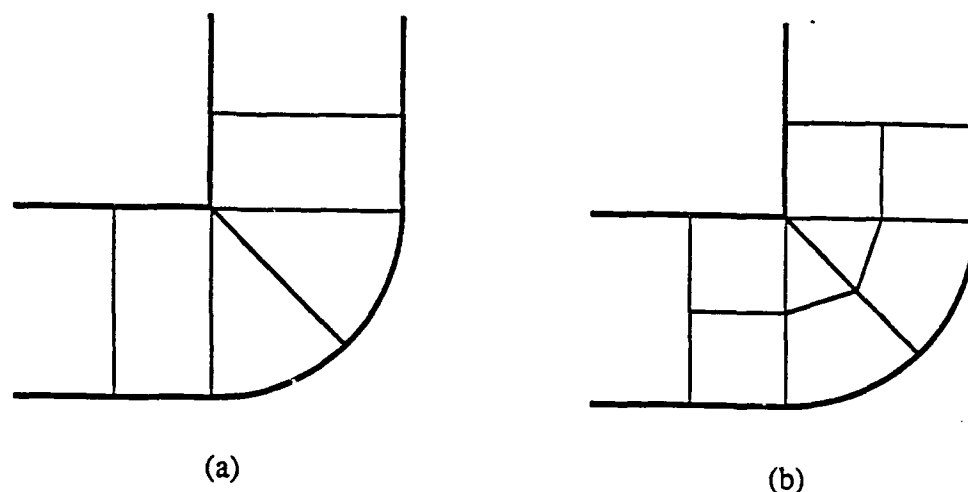


Figure 4.4: Domain Partition. (a) First step – column division; (b) second step – cell division.

4.2 Evaluation of Integrals

The boundary element method depends on the evaluation of integral equations to model a system. Many types of integrands are involved, and practically all of them exhibit singular behaviors, creating difficulties for numerical implementations. As the heart of the BEM, integration techniques warrant a detailed examination.

The division of the boundary into segments facilitates the estimation of source distribution and the prescription of boundary conditions. Similarly, the representation of segments by straight lines simplifies the evaluation of line integrals. Even so, these expressions are still difficult to derive, as the line path defining the integration limit are arbitrarily placed in a two-dimensional space. The predicament becomes worse for area integrals on quadrilaterals or triangular elements¹.

¹Fortunately, the pseudo-sources that are used for modeling nonlinear bulk parameters can be com-

In two-dimensional Laplace problems that employ $\phi^*(x) = \log(x)$ as the kernel functions, line integral involving ϕ^* and its first derivatives are routinely done in analytical form. This approach is usually abandoned in biharmonic systems because the functions are more complicated. For periodic problems that use $\phi^*(x) = \log[2(\cosh(y) - \cos(x))]$ kernel, such an analytical evaluation actually becomes infeasible. The Lobachevski functions family $L(x) = \int \log[\sin(x)]dx$, to which the periodic kernel yields after integration, only exists in integral form and cannot be reduced further – just like the error function $\text{erf}(x)$. But unlike the error function, no library routines are available to provide numerical results for Lobachevski functions. Thus numerical integration is used whenever possible and convenient.

Consider the integral equations stated previously in Eqs. 3.7 and 3.11:

$$\begin{aligned}\Phi(p) &= \int_{\Gamma} \rho(q) \phi^*(p - q) d\Gamma \\ \frac{\partial \Phi(p)}{\partial n} &= \int_{\Gamma} \rho(q) \frac{\partial \phi^*(p - q)}{\partial n} d\Gamma\end{aligned}$$

After boundary segmentation, they become

$$\Phi(p) \approx \sum_{i=1}^N \rho_i \int_{\Gamma_i} \phi^*(p - q) dq \quad (4.1)$$

$$\frac{\partial \Phi(p)}{\partial n} \approx \sum_{i=1}^N \rho_i \int_{\Gamma_i} \frac{\partial \phi^*(p - q)}{\partial n} dq \quad (4.2)$$

where N is the number of segments and σ_i is the source distribution on Γ_i , the i^{th} boundary segment. The σ 's are factored out from the integral expressions because they are constants. Methods of evaluating straight-line segment integrals are discussed in the following sections.

4.2.1 Numerical Quadratures

Numerical integration, or quadrature as it is often called, seeks to approximate the integration $\int f(x)dx$ by evaluating the expression $f(x)$ at discrete points x_i , weighing puted by integrating around the perimeter of a cell, as shown later. No area integrals are needed.

them with w_i and summing up the results:

$$\int_{-1}^1 f(x) dx \approx \sum_{i=0}^N f(x_i) w_i.$$

The two most common schemes to determine the spacings of x_i and values for w_i are the Gaussian quadrature and the Newton-Cotes formulae. In the following discussions, it is assumed that the integration goes from -1 to 1 .

Newton-Cotes Formulae

In the Newton-Cotes formulae, the nodes x_i 's are equally spaced. If the endpoints of the line path are also quadrature points, the formula is said to be closed, otherwise it is open. The accuracy or the "order" of the formula depends on the number of quadrature points. Of the low-order closed approximations, the 2nd order is more commonly known as the trapezoidal rule, and the 3rd order Simpson's rule. The closed Newton-Cotes formulae that have been used in this investigation are given in Table 4.1. Observe that the 7th order approximation has an undesirable property in that w_i fluctuates wildly. In fact the 9th (not shown) and higher order approximations have undesirable negative weighing factors.

Gaussian Quadrature

The Newton-Cotes approximation was later replaced with Gaussian quadrature. The Gaussian scheme uses irregularly spaced nodes. The coefficients for Gaussian quadrature are given in Table 4.2.

The reason for the need of high order approximations is that fields have fast varying fields near a singularity. This situation arises when the top and bottom boundaries lie very close together, as oxide is thin at the beginning of oxidation. The selection of quadrature order is determined by the closeness the collocation point is to the

n	$\pm x_i$	w_i
3	0	4/3
	1	1/3
5	0	12/45
	1/2	32/45
	1	7/45
7	0	272/420
	1/3	27/420
	2/3	216/420
	1	41/420

Table 4.1: Newton-Cotes Coefficients

n	$\pm x_i$	w_i
3	0.0000000000	0.8888888888
	0.7745966692	0.5555555555
5	0.0000000000	0.5688888888
	0.5384693101	0.4786286704
	0.9061798459	0.2369268850
7	0.0000000000	0.4179591836
	0.4058451513	0.3818300505
	0.7415311855	0.2797053914
	0.9491079123	0.1294849661
9	0.0000000000	0.3302393550
	0.3242534234	0.3123470770
	0.6133714327	0.2606106964
	0.8360311073	0.1806481606
	0.9681602395	0.0812743883

Table 4.2: Gaussian Quadrature Coefficients

boundary segment².

4.2.2 Analytical Integrations

Source contributions from all segments are summed up to determine the boundary condition of a particular segment. When the source coincides with the collocation point, the integrand becomes indefinite. Under such a condition, analytical integration must be used as numerical quadrature performs rather poorly. If a quadrature node coincides with the collocation point, the numerical solutions will blow up as we attempt to evaluate $\log(0)$ or divide something by 0. On the other hand, if the singular point is omitted, the approximation will be in great error because the contribution of many constant come from the singular point, as we will see later. Only analytical integration can deal with singularity adequately.

Periodic kernels, as noted that in Chapter 3, can be reduced to non-periodic counterparts around the singularities by using Taylor series expansions of the trigonometric and hyperbolic functions. All analytical integrations over singularities are done in terms of those non-periodic versions. Because of the limitations of Taylor series approximation, the integral must be done close to the singular point, and the path cannot be too long. Typically the segment length is limited to be shorter than 0.2 radian. For a longer boundary segment, It is possible to use quadrature on the edges and analytical integration at the center.

Coordinate Transformation

In numerical quadrature, it does not matter in what order the contributions from the quadrature points are summed, the results are the same. Such is not the case

²Note that only odd number of nodes are employed in either Newton-Cotes or Gaussian quadrature. This has to do with the programming feature that checks the central node before deciding which order of quadrature should be used.

for analytical integration – the solution depends on the direction of integration path. Incorporating all these information will make the resulting expressions unnecessarily complicated. Thus a modular approach is chosen to breakdown the integration procedure into two parts. First, the integral expressions are solved in a local $x' - y'$ coordinate system. This local frame is chosen such that the integration path always runs in a predefined direction of either x' or y' . The solutions are then transformed back to the global reference frame using the following rules:

$$x = \cos \theta x' - \sin \theta y' \quad (4.3)$$

$$y = \sin \theta x' + \cos \theta y' \quad (4.4)$$

$$x^2 = \cos^2 \theta x'^2 - \cos \theta \sin \theta x' y' + \sin^2 \theta y'^2 \quad (4.5)$$

$$xy = \cos^2 \theta x' y' + \cos \theta \sin \theta (x'^2 - y'^2) - \sin^2 \theta x' y' \quad (4.6)$$

$$y^2 = \cos^2 \theta y'^2 + \cos \theta \sin \theta x' y' + \sin^2 \theta x'^2 \quad (4.7)$$

$$x^3 = \cos^3 \theta x'^3 - 3 \cos^2 \theta \sin \theta x'^2 y' + \cos \theta \sin^2 \theta x' y'^2 - \sin^3 \theta y'^3 \quad (4.8)$$

$$\begin{aligned} x^2 y = & \cos^3 \theta x'^2 y' + \cos^2 \theta \sin \theta (x'^3 - 2x' y'^2) + \cos \theta \sin^2 \theta (y'^3 - 2x'^2 y') \\ & + \sin^3 \theta x' y'^2 \end{aligned} \quad (4.9)$$

$$\begin{aligned} xy^2 = & \cos^3 \theta x' y'^2 + \cos^2 \theta \sin \theta (2x'^2 y' - y'^3) + \cos \theta \sin^2 \theta (x'^3 - 2x' y'^2) \\ & - \sin^3 \theta x'^2 y' \end{aligned} \quad (4.10)$$

$$y^3 = \cos^3 \theta y'^3 + 3 \cos^2 \theta \sin \theta x' y'^3 + 3 \cos \theta \sin^2 \theta x'^2 y' + \sin^3 \theta x'^3 \quad (4.11)$$

where θ is the rotation angle from the x axis to the x' axis, measured counterclockwise. Note that these rules also applies to differentiation operators by substituting parameters with differentiation operators. For instance:

$$\frac{\partial^2}{\partial x^2} = \cos^2 \theta \frac{\partial^2}{\partial x'^2} - \cos \theta \sin \theta \frac{\partial^2}{\partial x' \partial y'} + \sin^2 \theta \frac{\partial^2}{\partial y'^2}.$$

Working in a local frame where the integration path runs either in the x' or y' direction greatly simplifies the analytical expressions.

4.2.3 Line Integrals Over Collocation Points

It turns out that the integrals over the collocation point usually yield simple results due to the midpoint collocation method and the step function approximation of the sources. Great care must be taken, however, when dealing with singularities. The integral path must be chosen carefully to be consistent with the source-free assumption in the boundary integral equations.

Three types of integrands are encountered. First, for non-singular or weak singular functions such as logarithm $\log(r)$, we have a well-defined and well-behaved solution. For strong singularity, we get two other types of behaviors. The second one comes from expressions of the form

$$\begin{aligned}\lim_{y \rightarrow 0} \int_{-a}^a \frac{x}{x^2 + y^2} dx &= \int_{-a}^a \frac{1}{x} dx \\ &= [\log(x)]_{-a}^a \\ &= 0.\end{aligned}$$

Although the overall result is zero, it is due to cancellation of two logarithm terms that become singular as $a \rightarrow 0$. This integration must be interpreted in the Cauchy principle value sense. As we can see from this expression, the result shows logarithmic singular behavior if one of the endpoints is 0 instead of $\pm a$. This is the cause of singular behaviors when the surface orientation is discontinuous. The third integration type is of the form

$$\int_{-a}^a \frac{y}{x^2 + y^2} dx; \quad y \rightarrow 0,$$

which yields $0, \pi$, or $-\pi$, depending $y = 0, 0^+$, or 0^- . We must choose y to be 0^+ or 0^- such that the source at the singularity just lies 'outside' the simulation region B . We don't want any source inside the region³. For this expression, the jump in values is

³Actually, y is chosen to be exactly zero in many boundary element references [38,42,40]; the integral is 0. But now the half of the source point is lying inside the domain, therefore the π term is picked up by the divergence operator on the source. We see that these two approaches are consistent with one another.

due to the term $\tan^{-1}(x/y)$, which is problematic in many applications, as it suddenly changes values.

The integral expressions for the basis functions defined in Chapter 3 are given in the Appendix.

4.2.4 General Analytical Line Integrals

Over the course of investigating different methods of modeling nonlinearity, it was found that high accuracy in the integral technique is desirable, not to avoid numerical errors, but to ensure that integration errors do not play a role in the ill behavior of the solutions. A set of analytical integration formulae is also given in the Appendix. However they are only applicable to the non-periodic basis functions.

4.3 Matrix solutions

In Eqs. 4.1–4.2, σ 's are the unknowns. Values for σ 's must be calculated before the domain solution or the unknown boundary parameters can be determined. The procedure is to compute σ 's in terms of the "known" boundary conditions. Suppose we are dealing with a Dirichlet problem in which Φ at the boundary is specified. Φ_i , the potential at the collocation point p_i of segment i , is given as

$$\begin{aligned}\Phi(p_i) &= \sum_{j=1}^N \sigma_j \int_{\Gamma_j} \phi^*(p_i - \vec{q}) d\Gamma \\ &= a_{ij} \sigma_j\end{aligned}$$

where $a_{ij} := \int_{\Gamma_j} \phi^*(p_i - \vec{q}) d\Gamma$. Combining this expression from all boundary elements, the following matrix representation is arrived:

$$A\vec{x} = \vec{b}. \quad (4.12)$$

A is a $N \times N$ matrix whose entry a_{ij} is defined above, and x and b are vectors whose elements are $\{\sigma_1 \dots \sigma_N\}$ and $\Phi_1 \dots \Phi_N$ respectively. The solution of x is given by

$$\bar{x} = A^{-1}\bar{b}. \quad (4.13)$$

With σ 's solved, we can proceed to calculate $\frac{\partial \Phi}{\partial n}$ at the boundary.

Suppose the problem is Neumann, Eq. 4.2 is used to assemble a similar matrix equation to solve for σ 's:

$$C\bar{x} = \bar{d}. \quad (4.14)$$

where $c_{ij} = \int_{\Gamma} \partial \phi^*(p_i - q) / \frac{\partial}{\partial n} n dq$, and $d = \partial \Phi_1 / \partial n, \dots, \partial \Phi_N / \partial n$. For simplicity, we have ignored the required normalization equation.

4.3.1 Matrix System for Diffusion

The oxidant diffusion problem is a mixed boundary-value problem consisting of Dirichlet, Neumann and Robin boundary condition, as mentioned in Chapter 2. A new hybrid matrix equation is assembled from Eqs. 4.12 and 4.14, given symbolically as

$$[A|C]\bar{x} = \bar{b}|\bar{d}. \quad (4.15)$$

Depending on the boundary condition of a particular segment i , the i^{th} row of A or C , and i^{th} element of b or d , or a linear combination of both, is used to form the entries of this matrix:

$$\begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ c_{l1} & \dots & c_{lN} \\ \vdots & \ddots & \vdots \end{bmatrix} x = \begin{bmatrix} b_1 \\ \vdots \\ d_l \\ \vdots \end{bmatrix} \quad (4.16)$$

where, for illustration, boundary elements 1 and l are shown to have a Dirichlet condition and a Neumann condition respectively.

Note that all those matrices A , C and the hybrid, are fully populated. This is due to the global influence of the kernels. The source σ on any one segment has a direct

effect on the boundary condition of on all segments. In contrast, the finite difference and finite element methods always produce sparsely populated matrix because of the short-ranged interaction between elements. Their nodes or elements only interact with nearest neighbors.

The solution for the unknown x is obtained using Gaussian or Gauss-Jordan elimination with partial pivoting. Note that an explicit inversion of the matrices is seldom required.

4.3.2 Oxide Flow

The boundary conditions for oxide flow are specified in vector form. For a two-dimensional problem, each boundary segment requires two entries in the matrix. Thus, the resulting matrix size is $2N \times 2N$. Otherwise, the matrix synthesis process is similar to that of diffusion.

4.3.3 Multiple Domains

The discussion so far concerns with single-domain problems. This treatment is adequate for oxidant diffusion but not for the oxide motion. The oxide bulk is in contact with silicon nitride mask and silicon substrate (and also the ambient). For the diffusion problem, the boundary conditions at these interfaces are $\vec{F} \cdot \hat{n} = 0$ and $k_s C = D \vec{F} \cdot \hat{n}$, due to the fact that silicon nitride is impermeable to oxidant and that chemical reaction is a interfacial activity. The mechanical interaction between these bodies are more involved. All materials are flexible. But to make the problem simpler, we assume that the silicon substrate, which is very thick with respect to the oxide, is totally rigid. However, the silicon nitride mask is of comparable thickness to oxide and deforms during oxidation, therefore its interaction with the oxide motion may not be overlooked.

At the interface between oxide and nitride, neither the displacement nor surface traction can be preassigned. The actual values for these parameters depends on the detail interactions of these bodies. A general approach for dealing with multiple domains is to combine their matrix equations into one and match the boundary conditions at the common interface. Let's assume that the matrix equations for two domains, A and B, are given as

$$\begin{bmatrix} G_1^A \\ G_I^A \end{bmatrix} \bar{\sigma}^A = \begin{Bmatrix} U_1^A \\ U_I^A \end{Bmatrix} \quad (4.17)$$

$$\begin{bmatrix} H_1^A \\ H_I^A \end{bmatrix} \bar{\sigma}^A = \begin{Bmatrix} P_1^A \\ P_I^A \end{Bmatrix} \quad (4.18)$$

$$\begin{bmatrix} G_2^B \\ G_I^B \end{bmatrix} \bar{\sigma}^B = \begin{Bmatrix} U_2^B \\ U_I^B \end{Bmatrix} \quad (4.19)$$

$$\begin{bmatrix} H_2^B \\ H_I^B \end{bmatrix} \bar{\sigma}^B = \begin{Bmatrix} P_2^B \\ P_I^B \end{Bmatrix} \quad (4.20)$$

where U 's, P 's, and σ 's are the surface displacement, traction, and sources respectively. The domains are identified by the superscripts; the boundaries are denoted by the subscripts.

The requirements at the common interface Γ_I are (i) displacements be continuous:

$$U_I^A = U_I^B$$

and (ii) surface tractions be equal and opposite:

$$P_I^A = -P_I^B.$$

The first ensures no overlapping of materials is possible, the second conserves forces. Note that these vectors are defined with respect to the global reference frame for consistency, instead of the local frames. For the combined systems, we thus have the following matrix:

$$\begin{bmatrix} [G|H]_1^A & 0 \\ [G]_I^A & -[G]_I^B \\ [H]_I^A & [H]_I^B \\ 0 & [G|H]_2^B \end{bmatrix} \begin{bmatrix} \sigma^A \\ \sigma^B \end{bmatrix} = \begin{pmatrix} \{[U|P]\}^A \\ 0 \\ 0 \\ \{[U|P]\}^B \end{pmatrix}. \quad (4.21)$$

This matrix is larger but contains many zero entries. It has been found that the matrix solution can be sped up significantly by taking advantage of the blocked nature of the matrix.

4.3.4 General Domain contributions

In the interior-cell formulation for nonlinear diffusion and oxide motion, there is no difference between the contributions from the domain sources and those from the boundary sources, as far as the matrix setup is concerned. For an equivalent system, the interior-cell approach produces a denser and less efficient matrix than the subregion technique, however it is more readily extended.

4.4 Computation for Nonhomogeneous Domain

By inducing variations in the diffusivity and viscosity, stress causes pseudo domain-sources to appear in the integral equations for both oxidant diffusion and oxide motion processes. Simplifications and techniques for evaluating the domain contribution are outlined below.

We have mentioned in Chapter 3 that we use $\partial\sigma_{ij}^0/\partial x_j$ and $\nabla \cdot \mathbf{F}$ instead of the nonderivative forms in our domain calculations. The reason has to do with the domain integral approximation. In our scheme, the domain is partitioned into triangular or quadrilateral cells, as described earlier in this chapter. Within each cell, it is assumed that the scale factors are constant; they take finite jumps from one cell to another. Consequently the values for those pseudo sources are zero inside a cell but become an impulse at the cell boundaries. This features reduces domain area integrals into line integrals along the perimeter of the cells.

The values for the scale factors are computed using stress values at the center of the cell. For the source strength of the cells, two different estimation techniques have

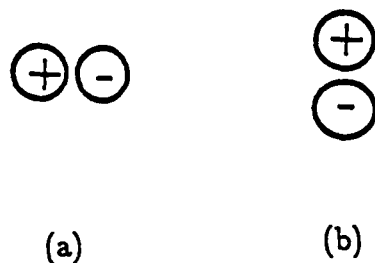


Figure 4.5: Dipole sources. Unlike a simple source, a dipole source can have spatial orientation.

been considered.

4.4.1 Dipole and Couplet Approximations

As a vehicle to realize nonlinearities in the bulk parameters, the pseudo domain-sources are supposed to perturb the solution, but not to introduce real body-force or oxidant source in the bulk. Hence the net values of the domain sources should be zero.

One way to automatically satisfy the zero average-value constraint is to employ dipole and couplets. As the kernel function for the double-layer method mentioned in Chapter 3, a dipole is obtained from ϕ^* by taking derivative in an appropriate direction. Shown in Fig. fg:dipoles are the two basic dipoles, aligned in the x and y directions, respectively. Because positive and negative charges (source and sink) always appear in pair, the flux contribution to an enclosing boundary is zero. The couplets are a

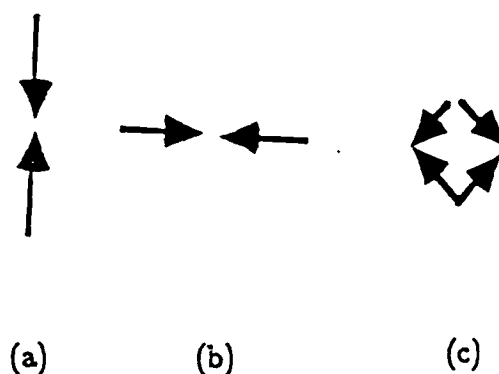


Figure 4.6: Couplet sources

pair of force acting in opposite directions. The three basic couplet configurations are shown in Fig. 4.6.

When one dipole or couplet source is distributed uniformly over a domain Ω , the positive and negative charges overlaps and cancels in the bulk, leaving unbalanced ones on the perimeter. This situation is shown graphically in Fig. 4.7 for a dipole source. It can be shown mathematically that, with integration by parts, an area integral is reduced to a line integral around the perimeter.

In the first approach for determining pseudo domain-source, the flux and stress are estimated at the cell center and assumed to be uniform across the cell, just like the scale factors k_d and k_η . Each cell is considered individually without regards to neighbor cells. The resulting source distributions resemble the dipoles and couplets, thus the name for for this correct method.

When used on simple test structures where diffusion flux variation is smooth (for

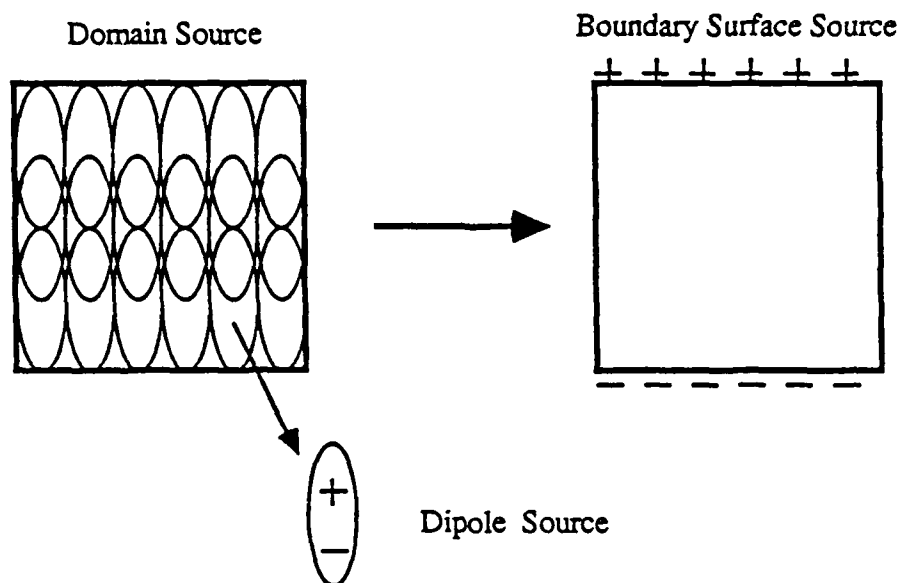


Figure 4.7: Conversion of Dipole Area Integral into Contour Integral. The transformation is possible due to the fact the dipoles in the interior cancel one another.

the homogeneous case), the performance of this method is adequate, i.e. it can block the diffusion to a good extent if the scale factors k_d 's for the test cells are reduced significantly. Some residual flux is present unavoidably as this is a perturbation technique. Unfortunately for real oxidation structures such as LOCOS, the performance is rather dismal. What happens is that the oxidant flux near the edge of the nitride mask is highly nonuniform. Because of the thinness of the relief oxide, interior cells there are only single-layer. Sources determined at the cell center cannot provide adequate correction to the flux distribution at the cell boundaries – it is not sufficient to counteract the oxidant flux reaching the silicon side, and on the other hand, oversupply oxidant to the nitride side, which is reflected back. Consequently, it loses its ability to block the diffusion. In fact, it produces terrible results – where the cell partition ends, the supply of oxidant may increase. The performance of the doublets on the oxide flow problem is equally lackluster – it can never reduce the stress level by more than half regardless how small the scale factor k_η is.

4.4.2 Interfacial Source Formulation

One major weakness with the dipole/couplet scheme is that the collocation point for the domain sources is located at the cell center - flux and stress are assumed to be uniform across the cell. This problem is circumvented in the interfacial source formulation. Major activities such as oxidation take place on the sides of a cell. By having the source computed individually at each side of a cell, wider variations of source strength can be achieved to cope with the highly nonuniform flux.

To decide how the interfacial source is computed, we consider the domain integral of Eq. 3.93:

$$\int_{\Omega} \frac{1}{2\pi} (\nabla_{\mathbf{q}} \cdot \mathbf{F}^0(\mathbf{q})) \phi^*(\mathbf{q} - \mathbf{p}) d\Omega.$$

At the interface between two cells, the value of k_d changes abruptly from k_1 to k_2 , resulting in non-conservation of flux. Given the fact that only the normal component of flux matters and that the interface is a straight line, at locations very close to the interface, the behavior is one dimensional.

Hence we examine an equivalent one-dimensional model, using physical arguments instead of mathematics. As shown in Fig. 4.8, this model has two regions - 1 and 2 - having different diffusivity k_1 and k_2 . (For simplicity, we drop D_0 as it never appears in the final solution.) The flux in each region is given by

$$F_1 = k_1 C$$

and

$$F_2 = k_2 C$$

respectively. Thus, at the interface, there is a build-up of oxidant of

$$(k_1 - k_2)C.$$

which has to be dissipated. Green's function for one-dimensional diffusion is $-|x|$, which has a gradient of ± 1 , depending on whether x is smaller or larger than zero.

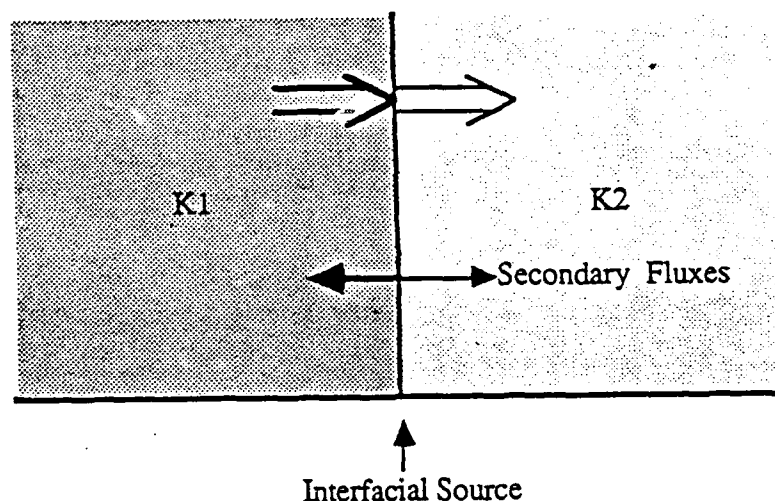


Figure 4.8: One-Dimensional Nonhomogeneous Diffusion. The secondary fluxes are created due to the mismatch in diffusivity between the two regions.

Hence the fluxes generated by the interface are

$$F'_1 = -k_1 D$$

$$F'_2 = k_2 D$$

where D is the source strength yet to be determined and the minus sign in F'_1 indicates that the flux goes in the negative direction, as illustrated at the bottom of Fig. 4.8. Therefore net outgoing flux is given by $(F'_2 - F'_1)$, and it must equal to net incoming flux, which is $(F_1 - F_2)$. Thus we have the following expression for the interfacial source:

$$D = \frac{k_1 - k_2}{k_1 + k_2} C. \quad (4.22)$$

Simply speaking, when a flux impedes on a barrier (that has lower diffusivity), it goes through partially. Due to build-up of materials at the boundary, two additional fluxes are created: one is reflected back (F'_1), the other continues forwards (F'_2). These fluxes seem to be created by a pseudo domain-source. Note that even if k_2 is zero, i.e. when

region 2 is impermeable, the oxidant concentration in it is nonzero and show a finite gradient. This is due to the steady-state diffusion assumption. This technique has been verified to work in one dimension analytically, and two dimensions, numerically.

The extension to viscoelastic flow is straight forward. The effective stress scale factor is $k_{\sigma 1}$ and $k_{\sigma 2}$ for Region 1 and 2 respectively. The unscaled stress distribution is continuous across the interface. Each region contributes towards the surface traction on the interface. If $k_{\sigma 1} = k_{\sigma 2}$, the contributions will cancel each other. On the other hand, if $k_{\sigma 1} \neq k_{\sigma 2}$, a net force will be exerted on the interface. It is nullified by a pseudo body-force which generates its own stress and displacement fields. The value of the body-force vector is given by

$$\frac{k_{\sigma 1} - k_{\sigma 2}}{k_{\sigma 1} + k_{\sigma 2}} \sigma_{ij} n_j. \quad (4.23)$$

As we can see, the nonlinear viscoelasticity formulation parallels that of nonlinear diffusion.

A last note, the calculation for a particular pseudo source must not include the flux or stress generated by the source itself. This is can be done easily in the program codes.

Cells in Contact with Boundary Segments

With the formulae given by Eqs. 4.22 and 4.23, we can calculate the interfacial source between two cells and between a cell and a homogeneous region. One remaining question is on the proper treatment for a cell connected to a boundary segment. Recall the boundary source is located on the outside of a boundary collocation. The interfacial source is by definition on the inside. Shown schematically in Fig. 4.9a. are the positions of the sources with respect to the collocation point. The distances, δ_1 and δ_2 , are infinitesimally small. Consider the diffusion problem. The sandwich area containing the collocation point is outside a cell, and therefore assumed to belong to the homogeneous confine. Its diffusivity is the nominal value $D = D_0$, but this really does

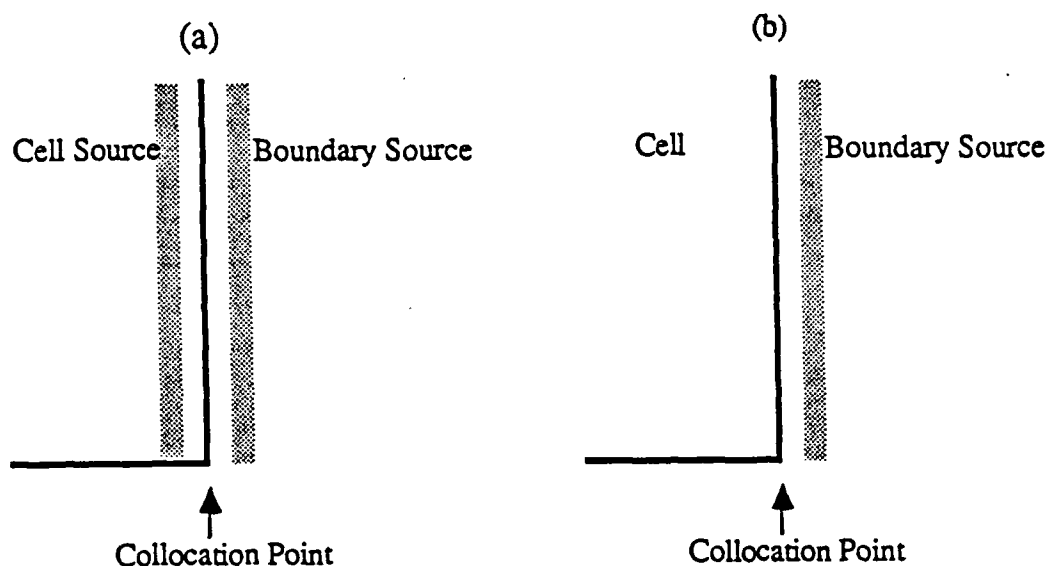


Figure 4.9: Boundary Sources.

(a) Original; (b) merged.

not matter because the layer is so thin that the transport process is dominated by the cell. Given the close proximity of the two sources ($\delta_1 + \delta_2 \rightarrow 0$), one would expect that the value of one source would track the other in some specific way. Thus it is useful and beneficial to recode the mathematics so that one source (*i.e.* the interfacial one) can be eliminated. Through derivations that are not presented here, indeed, the interfacial source can be eliminated and in place, and assign the boundary segment the effective diffusivity $D = D_0$. The new configuration is shown in Fig. 4.9b. The same approach works for the nonlinear viscoelasticity problem.

4.5 Nonlinear Solutions

The effects of stress on thermal oxidation may be divided into two categories, depending on whether they act on the boundary conditions or on the bulk parameters. Either

case, they are nonlinear functions that have to be solved iteratively. The two boundary parameters influenced by stress are the surface reaction rate k , and the equilibrium concentration C^* of the diffusion system. Because boundary conditions do not directly involve the domain flux conservation law, the system remains Laplacian and is still subject to reduced dimensionality treatment. Bulk parameters affected by stress include the diffusion coefficient of oxidants and the viscosity of oxide. For them, domain corrections outlined in the previous section are used.

The stress functions used in this investigation are of this form:

$$\exp\left(\frac{PV}{kT}\right)$$

where P is a specific component of the stress tensor, V is an activation volume, k is Boltzmann's constant, and T is the temperature. In descriptions that follow, it is assumed that the system is stable (i.e. it does not have positive feedback) and that there is only one solution.

4.5.1 Scalar Iterative Methods

All the stress effects requires the solution to a nonlinear function, conceptually of the following form:

$$f(x) = x \tag{4.24}$$

where, for a simple case, x and f are scalars. A few iterative techniques are available for obtaining the solution for x . One basic approach is to use this updating scheme:

$$x^{k+1} = f(x^k).$$

where the superscripts indicate the iteration numbers. x^0 is a starting guess value. This technique shows poor convergence, nonetheless it is simple and does not require additional information of the system.

Newton's Method

For faster and more stable convergence, Newton's method is often employed. To illustrate this technique, we recast Eq. 4.24 in a different form:

$$g(x) = f(x) - x.$$

where $g(x)$ is the residue. $g(x)$ is 0 if x is the root to Eq. 4.24. Suppose x_0 is the solution (which we don't know yet), and x is close to x_0 , taking the Taylor series expansion of $g(x_0)$ to the 1st order at x yields

$$\begin{aligned} g(x_0) &\approx g(x) + g'(x)(x_0 - x) \\ &\approx 0. \end{aligned} \tag{4.25}$$

where the prime (') denotes the first derivative with respect to x . Accordingly, $x_0 \approx x - g(x)/g'(x)$. We have thus derived the Newton iteration scheme:

$$x^{k+1} = x^k - \frac{g(x^k)}{g'(x^k)} \tag{4.26}$$

This scheme shows a quadratic convergence, i.e. error is reduced by half in every iteration. However, it still may fail to converge under some conditions. For instance, the solution may oscillate between two regions for certain $f(x)$'s and initial conditions. Also the solution can blow up if $g'(x)$ happens to be zero.

It is difficult to come up with a general iterative method that is always robust and converges quickly for any type of $f(x)$. Although not always possible, one should analyze the problem nature and determine the best possible solution approach. Indeed the main reason that the Newton scheme has a fast convergence is that it considers the direction the system is heading, namely $f'(x)$. In case $f'(x)$ cannot be obtained directly, one may approximate it by

$$f'(x^k) \approx \frac{f(x^k) - f(x^{k-1})}{x^k - x^{k-1}}. \tag{4.27}$$

The iteration scheme thus become:

$$x^{k+1} = x^k - g(x^k) \frac{x^k - x^{k-1}}{g(x^k) - g(x^{k-1})} \tag{4.28}$$

which is known as the secant or false position technique [44].

Modified Iterations

Sometimes it is necessary to "damp" the iterations in order to stabilize the solution. This is achieved by adding a relaxation parameter ($0 \leq \lambda \leq 1$) to scale down the change in x at each iteration. The modification is

$$x^{k+1} = (1 - \lambda)x^k + \lambda f(x^k) \quad (4.29)$$

for the simple updating scheme and

$$x^{k+1} = x^k - \lambda g(x^k) \frac{x^k - x^{k-1}}{g(x^k) - g(x^{k-1})} \quad (4.30)$$

for Newton's method.

4.5.2 Multi-Variable Solutions

The scalar iteration methods can be extended to deal with multiple variables, but there are complications. For one thing, the variables in general are not decoupled. In other words, changing the condition on one variable affects the rest. It is difficult to predict in what direction the the system is moving. Consequently schemes such as the simple updating may easily become unstable if they update the solution in the wrong direction. The solution may need to be damped, resulting in slow convergence.

Updating Scheme

With the damping factor λ incorporated, the multi-variable updating scheme is

$$\bar{x}^{k+1} = (1 - \lambda^k)\bar{x}^k + \lambda^k \bar{f}(\bar{x}^k). \quad (4.31)$$

which is a straightforward extension of the scalar case.

Newton's Method

For Newton's method, the vector Taylor series expansion becomes

$$\begin{aligned}\bar{g}(\bar{x}_0) &\approx \bar{g}(\bar{x}) + J(\bar{x})(\bar{x}_0 - \bar{x}) \\ &\approx 0.\end{aligned}\tag{4.32}$$

where $\bar{g}(\bar{x}) = \bar{f}(\bar{x}) - \bar{x}$, and J is the Jacobian of $\bar{g}$.

Assuming that the vector size of $\bar{g}$ and $\bar{x}$ is N , the definition for J is

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_N}{\partial x_1} & \cdots & \frac{\partial g_N}{\partial x_N} \end{bmatrix}\tag{4.33}$$

where the element number is given by the subscript. The Jacobian is a measure of how $\bar{g}$ will respond to a small change in $\bar{x}$.

Newton's method thus becomes

$$\bar{x}^{k+1} = \bar{x}^k - J^{-1}(\bar{x}^k)\bar{g}(\bar{x}^k)\tag{4.34}$$

The determination of the Jacobian is not trivial – the array is large ($N \times N$) and its components are not readily obtained from $\bar{g}$ in analytical form. Using the false position technique is almost out of the question because enormous computation is required; only rough approximation is used here. Even when an accurate Jacobian is available, the method often fails converge. One often-cited reason is that it tends to overshoot in correction [45]. The Taylor series expansion is good only when $\bar{x}$ is very close to the actual solution. If it is not close, the Jacobian produces a poor correction term. In our applications, an error can result in larger deviations due to the exponential terms embedded in $\bar{g}$.

Newton's scheme is modified, as suggested in [45]:

$$\bar{x}^{k+1} = \bar{x}^k + t_k[J(\bar{x}^k) + s_k I]^{-1}\bar{g}(\bar{x}^k)\tag{4.35}$$

to include two parameters — s_k and t_k — to increase damping. The function of t_k is the same as λ^k in the scalar case — to reduce the changes. s_k makes the Jacobian more diagonally dominant, effectively reducing the interaction between parameters.

In both iteration techniques, there is no algorithm for calculating the optimum values for the relaxation parameters. They are determined on a trial and error basis. A two-level iteration scheme is adopted. The purpose of having the sub-iteration is to find large relaxation parameters that does not cause the system to go unstable.

4.5.3 Determination of Relaxation Parameters

There is no formula to determine optimum values for the relaxation parameters λ^k , s^k , and t^k *a priori*. They are obtained on a trial-and-error basis at each iteration. If these parameters are too small, then the speed of convergence will suffer. Conversely, if they are too large, the solutions will become unstable.

These parameters are adjusted according to the convergence rate of the solutions. Two types of error definition are employed in our iteration for control purposes; they are the maximum norm ($\|\cdot\|_\infty$) and the normalized Euclidean norm ($\|\cdot\|_2$) defined below:

$$\|\bar{g}\|_\infty = \max |g_i| \quad (4.36)$$

$$\|\bar{g}\|_2 = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N g_i^2}. \quad (4.37)$$

Note that our definition for the Euclidean norm has been normalized by $1/\sqrt{N}$ so that it yields the same result as the maximum norm if all elements of $\bar{g}$ are the same magnitude.

The maximum norm is applied to a residual error vector $\bar{g}$ to determine whether an iteration loop should be terminated because the solution has converged, i.e. the maximum error has dropped below a certain threshold.

The Euclidean norm is used for determining the relaxation parameters. Since it weights all elements in $\vec{g}$, the Euclidean norm is less sensitive to fluctuations of individual elements than the maximum norm. This feature is higher desirable because in the iteration process, some errors may arise locally and momentarily.

The value for a relaxation parameter is adjusted by comparing $\|\vec{g}^{k-1}\|_2$ and $\|\vec{g}^*\|_2$ where $*$ denotes the current trial step for iteration k . If the current trial error is smaller, $\vec{x}^*$ is made $\vec{x}^k$ and the relaxation parameter is increased slightly for the next iteration. On the other hand, if the current trial error is larger, the relaxation parameter is reduced to increase damping, and x^* and everything else is recomputed. This sub-iteration is repeated until the current trial error becomes smaller. The sub-iteration is also terminated after a few trials if it seems impossible to make the error smaller at the current iteration.

4.5.4 Iteration on Reaction Rate

The stress effect on reaction rate, k_s , is a self-regulating, negative feedback system that tends to make oxide growth rate more uniform along the silicon interface. Nonlinear k_s is a surprisingly difficult problem to solve numerically. With simple iterative schemes, a small local error can easily transform into a global instability.

Let's suppose the value of k_s of a particular boundary segment is higher than its equilibrium. Being more reactive, the segment therefore generates more oxide than its two neighbors; it also deplete to a certain degree the supply of oxidants to them. (We ignore segments further away.) Because it injects more oxide into the bulk, the segment therefore experiences a compressive surface traction that reduces its k_s . Meanwhile the slower-moving oxide of the two neighboring segments is being pulled, producing a tensile surface traction on the segment. In the next iteration step of a simple updating scheme, the k_s of the segment in question will be lowered, while those of the two neighbors will be raised: a reversal process is in progress. As

iteration proceeds, a ripple of alternating stress and k_s will spread through all the segments. If damping is insufficient, the errors become a self-reinforcing wave that grows in magnitude, resulting in global oscillation. For large damping factors, the errors will still propagate to adjacent segments, but with decaying magnitude; the solution eventually converges after many iterations.

The reason for the poor performance is that the technique does not consider the behavior of the system. In particular it does not anticipate that a change in k_s of a segment will affect others as well. Better stability is only ensured when the system response is factored into the updating scheme. This brings us back to Newton's method.

Accordingly, $\vec{f}$ for normalized k_s is of the form:

$$\vec{f} = \begin{bmatrix} \exp\left(\frac{\mathbf{p}_1 \cdot \hat{\mathbf{n}}_1 V_{k_s}}{kT}\right) \\ \vdots \\ \exp\left(\frac{\mathbf{p}_M \cdot \hat{\mathbf{n}}_M V_{k_s}}{kT}\right) \end{bmatrix},$$

where $\mathbf{p}_i \cdot \hat{\mathbf{n}}_i$ is the normal surface traction at boundary segment i . As shown, $\vec{f}$ has entries for M silicon interface segments. The surface traction list $\vec{p} = \{p_1, p_2, \dots, p_M\}$ may be explicitly defined as a function of the normalized k_s list: $\vec{p} = \vec{p}(\vec{f})$. As a combined product of the oxidant diffusion and oxide flow systems, the feedback mechanism on k_s is indirect. Its analytical expressions are practically not obtainable.

Due to the Robin boundary condition type, the growth rate k_s is embedded in the matrix equation that determines the source distribution for diffusion, as shown schematically below:

$$A(k_s)\vec{\sigma} = \vec{b}.$$

It requires an implicit or explicit inversion of the matrix A to obtain $\vec{\sigma}$. Given the largeness of A , it is thus impossible to maintain analyticity of $\vec{k}_s$ in the solution and perform differentiation on it later to get the Jacobian. For now, assume somehow the Jacobian can be obtained, the change in oxide injected into bulk due to a small change

in k_s is given by

$$\Delta \bar{u} = k J_D \Delta \bar{k}_s. \quad (4.38)$$

where k is an appropriate scale factor.

The situation with oxide flow is simpler. The relationship between the injection $\bar{u}$ and $\bar{P}$ is linear and is given by

$$\bar{P} = A \bar{U} + \bar{P}_0 \quad (4.39)$$

where $\bar{P}_0$ is contributions from nitride or free surface segments. Hence its Jacobian is given by $J_P = A$. The change in $\bar{P}$ due to $\bar{k}_s$ is therefore given by

$$\Delta \bar{P}_n = J_P J_D \Delta \bar{k}_s. \quad (4.40)$$

Approximation on Diffusion Jacobian

As mentioned earlier, the diffusion Jacobian J_D cannot be obtained by analytical means. A numerical check on a typical oxidation problem has show that it can be approximated with a diagonal matrix. The oxide growth rate of a one-dimensional problem is given by Eq. 2.6:

$$\begin{aligned} G &= \frac{dd_0}{dt} \\ &= \frac{1}{N} \frac{k_s C^*}{1 + \frac{k_s}{h} + \frac{k_s d_0}{D}} \\ &\approx \frac{1}{N} \frac{k_s C^*}{1 + \frac{k_s d_0}{D}} \end{aligned}$$

Its sensitivity to k_s is defined as

$$\frac{dG}{dk_s} = \frac{C^*}{N} \frac{D^2}{(k_s d_0 + D)^2} \quad (4.41)$$

When the oxide is thin, the growth rate is reaction-rate limited. In that case, $\frac{dG}{dk_s}$ is almost C^*/N , the maximum possible. As the oxide gets thicker, the growth rate

is diffusion limited and the sensitivity becomes weak: $(C^*/N) \cdot (D^2/k_s^2 d_0^2)$. Such a prediction does not generalize to a two-dimensional system. An increase in k_s of a particular boundary location does not necessitate an increased supply of oxidant from the free surface to enhance its growth rate. Rather, it can absorb oxidants that originally bound for neighboring points. Thus the sensitivity of growth rate on k_s is higher than the one-dimensional case for thick oxide. Through numerical investigations, it has been found to be close to C^*/N in most cases. More surprisingly and perplexing, the growth rate of neighboring points are not affected much. Thus a good approximation for the Jacobian of the diffusion problem is just a diagonal matrix.

Approximation on Jacobian for Oxide Flow

A change of k_s in one segment can in principle affect the rest. However, due to geometric and feedback considerations, strong influence is limited to a few neighboring segments; the rest experience very little changes. The implication on the Jacobian is that terms far away from the diagonal are approximately 0 (provided the k_s 's are properly ordered). In our tridiagonal approximation, entries that are not on the diagonal or the two off-diagonals are set to zero.

The oxide flow problem is not subject to the isolation treatment as the diffusion. Numerical calculations of stress due to the increased growth rate of a segment yield a $-/+/-$ pattern distributed over a few segments. Compressive stress is experienced by the modifying segment, and tensile stress by its two adjacent segments. The next two further out also see tensile stress, but with much less magnitude. In consideration of the approximation employed previously in the diffusion system, there is no strong reason to model stress change beyond the two neighboring segments. Thus the approximated Jacobian is a tridiagonal matrix.

The remaining problem is to calculate the entries of the matrix, that is the estimation of stress due to a unit change in the injection rate of each silicon segment.

Suppose among the N boundary segments, M of them are for the silicon interface. In the original solution scheme, we solve for the source S in terms of the boundary condition $\vec{b}$:

$$A\vec{S} = \vec{b}.$$

The method is extended to find simultaneously the solutions for the "impulse" responses associated with the M silicon segments. The single-column vectors, S and b , are expanded to $(M + 1)$ columns. After the normal displacement components from the silicon segments are removed, the modified $\vec{b}$ is put in the zeroth column of the new multi-column b . Unit vectors corresponding to the missing normal displacement component are placed in each of the remaining columns. After solving for S , we can reconstruct the desired source distribution

$$\vec{S} = \vec{S}[0] + \sum_{i=1}^N u_n[i] \vec{S}[i].$$

$\vec{S}[1]$ to $\vec{S}[N]$ are unit-impulse solutions from which we can obtain the stress contribution due to each silicon segments, that is, the entries for the Jacobian. Although it requires significant computation time to obtain S , considerable savings are later achieved for each subsequent iteration because the matrix equation is not solved again.

Given that the Jacobians for diffusion and oxide flow are a diagonal matrix and a tridiagonal matrix respectively, the combined Jacobian is still a tridiagonal matrix. Because the Jacobians are crudely approximated, the parameter s_k is utilized to make the system more diagonally prominent. Typically s_k is chosen to be less than the largest entry in $J_F J_D$.

Nonlinear Diffusivity and Viscosity

The feedback mechanism for the bulk parameters diffusivity and viscosity is more indirect than the reaction-rate. Consider the diffusion problem. Suppose the diffusivity of a region is perturbed and made smaller than its equilibrium value. This reduces

the oxidant flux to a certain section of the silicon interface. The oxide growth rate at that section is altered, and that consequently generates a new stress field that reaches the perturbed region and changes its diffusivity. The details depend on how far the perturbed region is from the silicon interface. The situation with nonlinear viscosity is also complicated. Without any knowledge of their Jacobians, these two parameters are solved using the simple iterative method.

Chapter 5

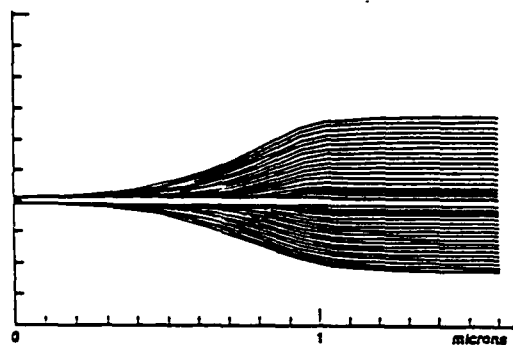
Simulations

Simulation results for various geometries and conditions are presented in this chapter.

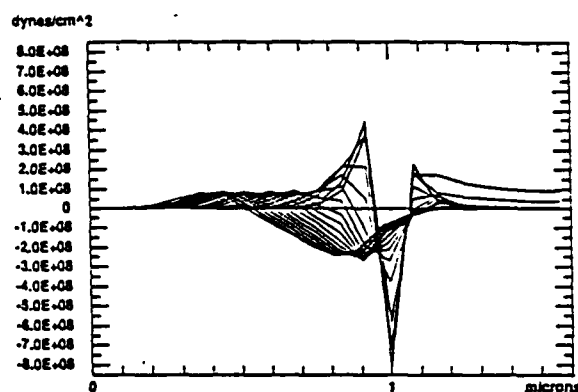
5.1 Viscoelastic Stress Relaxation

We will consider simple cases of local oxidation to demonstrate how stress is relieved through viscoelastic flow. A wide range of stress relaxation times are tested to see how different relaxation speed changes the stress behaviors and distributions. No stress effects are included in this study.

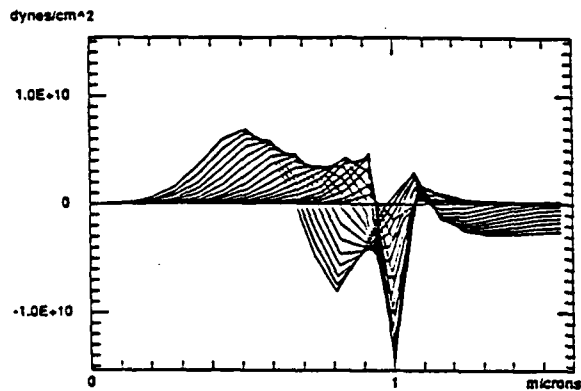
Fig. 5.1 shows the simulation results for a semi-recessed LOCOS structure for different stress relaxation times. In (a), the outline of the structure is plotted at every time step. In this simulation window of width $1.6\mu\text{m}$, the silicon nitride mask extends from $x = 0$ to $x = 0.96\mu\text{m}$ and is assumed to be totally flexible. The pad oxide thickness is $200\text{ \AA}$. Oxidation is carried out at 900°C in a wet ambient for 4.6 hours to obtain a final field oxide thickness of $5000\text{ \AA}$. Young's modulus and Poisson's ratio are taken to be $8 \times 10^{11}\text{ dynes}\cdot\text{cm}^{-2}$ and 0.194. Shown in b, c and d is stress distribution corresponding to relaxation times, η/G , of 1 minute, 1 hour, and 10 hours respectively. The normal component of the surface traction at the oxide-silicon interface is plotted



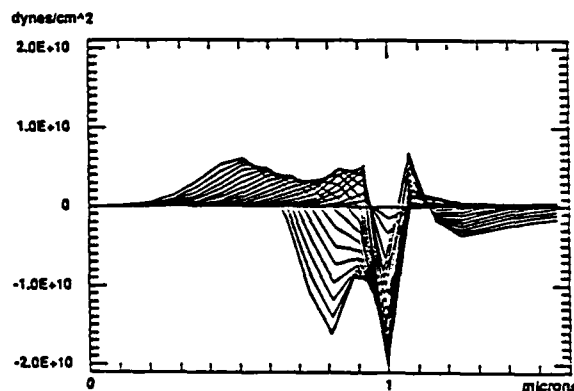
(a)



(b)



(c)



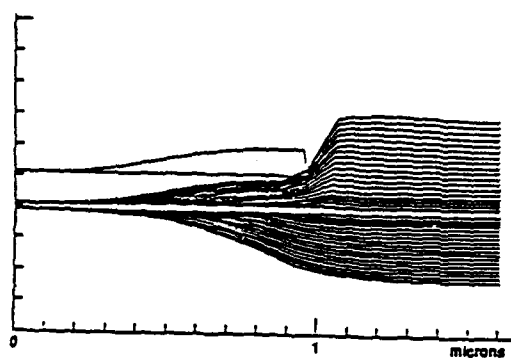
(d)

Figure 5.1: The nitride mask is assumed to be totally flexible. (a) Boundary outline. Stress behavior for stress relaxation time of (b) 1 min; (c) 1 hour; and (d) 10 hours.

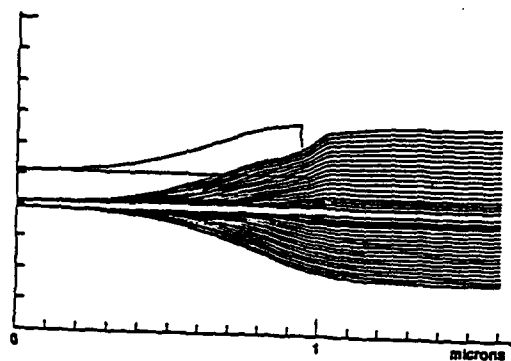
at every time step, like the outline of the oxide. In all three stress plots, there are two peaks in the compressive stress region (the negative value range). The early peak occurs at the edge of the nitride mask. This peak is due to the highly nonuniform oxidation rate in that region. As time progresses, the peak shifts to the left, further under the nitride mask, and gives rise to a late peak. As one would expect, stress reduces as viscosity decreases. Maximum stress attained for the 3 relaxation times is 8.1×10^8 , 1.5×10^{10} , and 2×10^{10} dynes-cm⁻² respectively. The oxide shape is not affected much by the relaxation times, therefore only one is shown.

We now repeat the same simulations with the silicon nitride mask modeled as an elastic material. The thickness of the nitride layer is $0.1\mu\text{m}$. The final shapes of the oxide and nitride layer are shown in Figs. 5.2a, b, and c. As the oxide becomes less viscous and flows more readily, the nitride layer bends less. The corresponding values for peak stress are 3.8×10^{10} , 1.2×10^{11} , and 1.6×10^{11} dynes-cm⁻² respectively. The last two values are unrealistically high as they are of the same order of magnitude as Young's modulus of oxide or nitride. We have found that in general we need these high stress values in order to deflect the nitride mask to a degree comparable to what is found in experiments. To keep stress down to a realistic level, the nitride mask must deform plastically or viscoelastically. Unfortunately we don't have any models for those behaviors nor any data to determine a model. In all likelihood, the stress value is inaccurate if it is a few percent of Young's modulus.

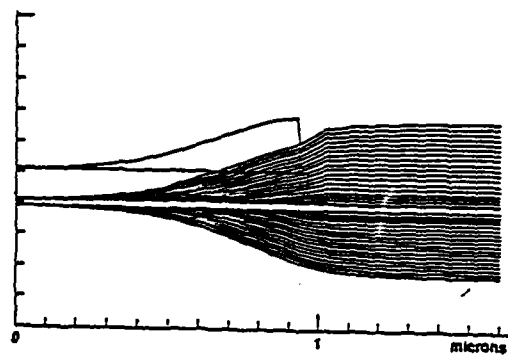
Viscoelasticity is a relative concept. If the time period involved is much shorter than the time constant of stress relaxation, a viscoelastic material deforms elastically. If the converse is true, the material flows viscously. The viscoelastic behavior is most profound when the oxidation time and relaxation time are comparable. Shown in Fig. 3a is a plot of maximum stress versus the relaxation time. The simulation conditions are the ones used earlier, without modeling the elastic bending of nitride mask. For long relaxation times ($\tau > 10$ hours), the stress curve is constant as the material is essentially elastic. For short relaxation times ($\tau < 1$), the curve bends down



(a)



(b)



(c)

Figure 5.2: The nitride is assumed to be elastic. Final oxide profile is shown for oxide relaxation time of: (a) 1 min; (b) 1 hour; and (c) 10 hours.

with a constant slope. The drop is a result of stress being inversely proportional to the relaxation time. For all the data points, maximum stress is observed at around 35 minutes into oxidation; that value roughly corresponds to the transition region in Fig. 5.3a. An alternate way of assessing the significance of viscoelastic flow is to check the sensitivity of stress to a change in viscosity. If the oxide is elastic, stress will not be affected. On the other hand, if it is viscous, one would see the same change in stress. Shown in Fig. 5.3b is the stress change due to a 10% increase in viscosity. There is little change in stress value for large relaxation times and an almost full 10% change for short relaxation times. This essentially corresponds to the gradient of the curve in Fig. 3a. As we can see, the stress value levels off after the viscosity reaches a certain threshold. If a viscous flow model was used instead, stress would be strictly proportional to the viscosity.

5.2 Nonhomogenous Bulk Parameters

Here we demonstrate the capability of the initial stress/built-in field formulation for modeling stress effects on viscosity and diffusivity. Concave and convex silicon corners are oxidized at 950°C in a wet oxidation; approximately 1100 Å of oxide is grown.

Fig. 5.4 shows the oxidation of a convex silicon corner and the domain partition used for computing pseudo domain sources. When the stress effect on viscosity is not modeled, the peak stress goes as high as 2.7×10^{11} dynes/cm⁻². With a stress model included, the peak stress drops down to 2.2×10^9 dynes/cm⁻².

The retardation effect of stress on diffusivity is illustrated in Fig. 5.5. Oxidation is done on an concave silicon corner. A large compressive hydrostatic pressure is present at the corner. In (a), no stress effect on diffusivity is considered – no growth retardation at the corner is observed. In (b), stress effect is modeled. The oxide thinning observed at the corner is due to reduced diffusivity of oxidant.

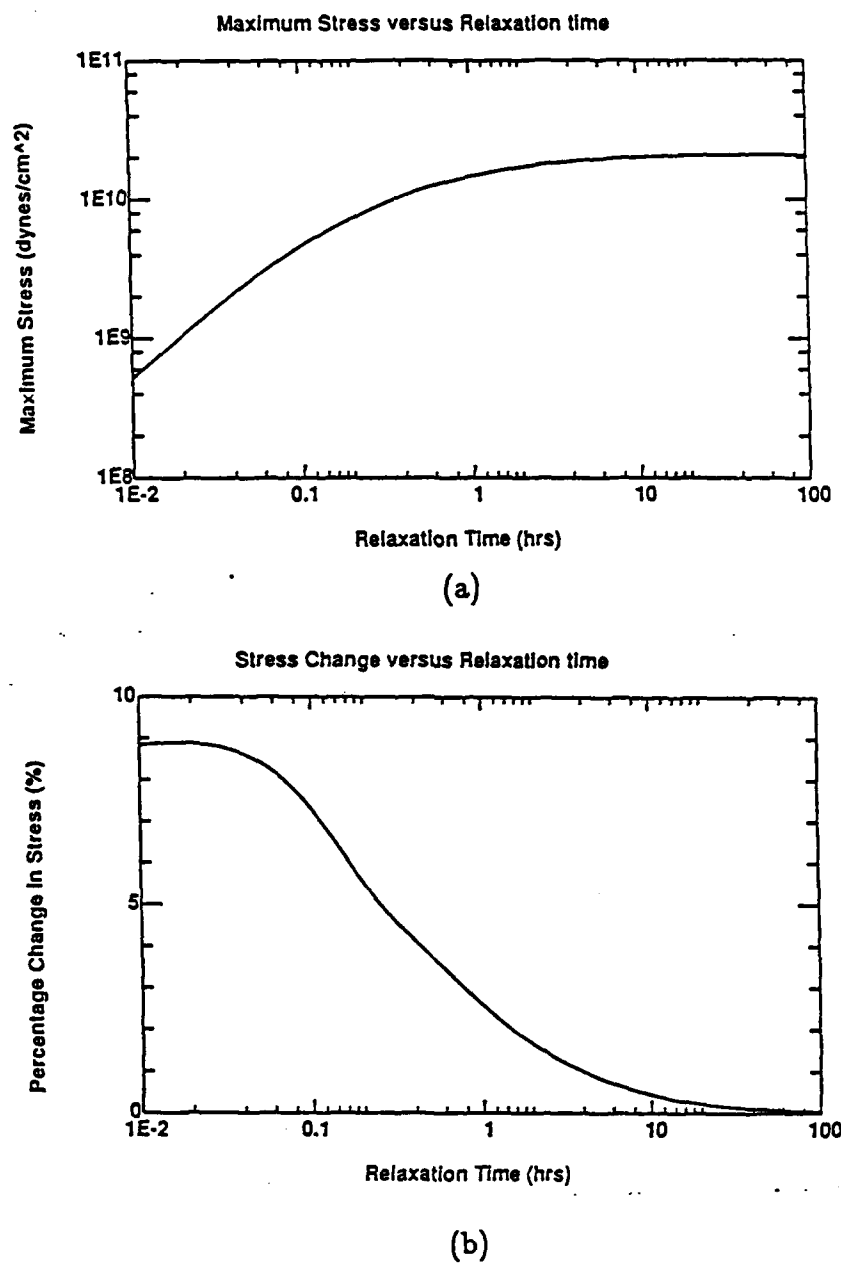
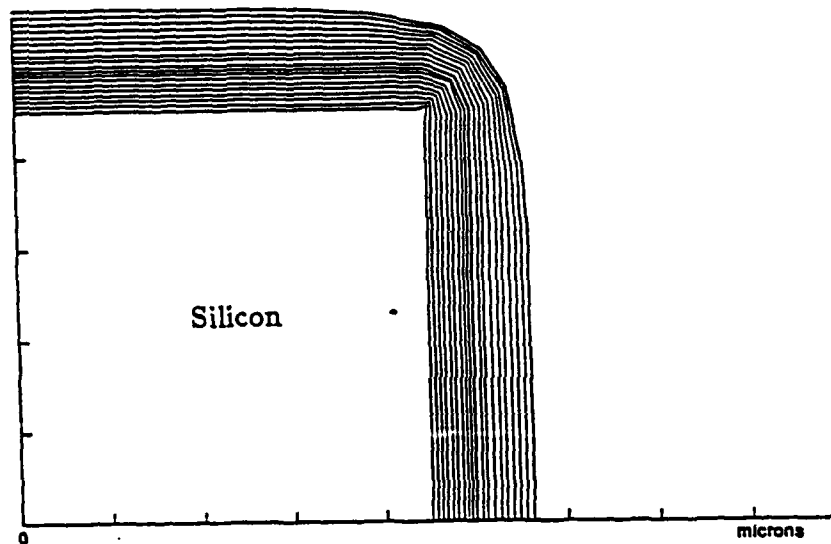
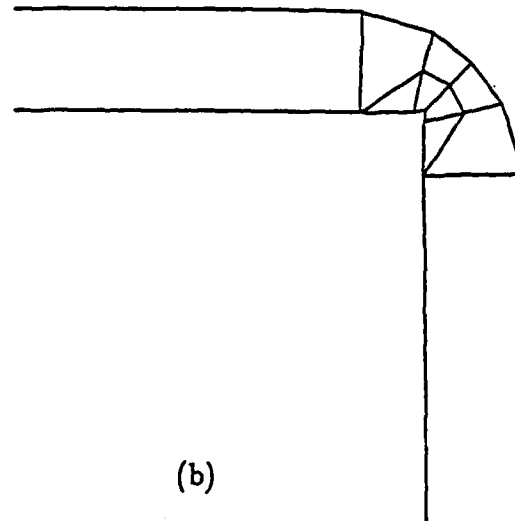


Figure 5.3: (a) Maximum stress versus relaxation times. (b) Stress change due to a 10% increase in viscosity at different relaxation times.



(a)



(b)

Figure 5.4: Modeling Stress Effect on Viscosity. (a) shows the outlines of oxide on a convex silicon corner. (b) shows the domain partition for modeling nonlinear viscosity. Cells are placed at the corner where high stress is expected.

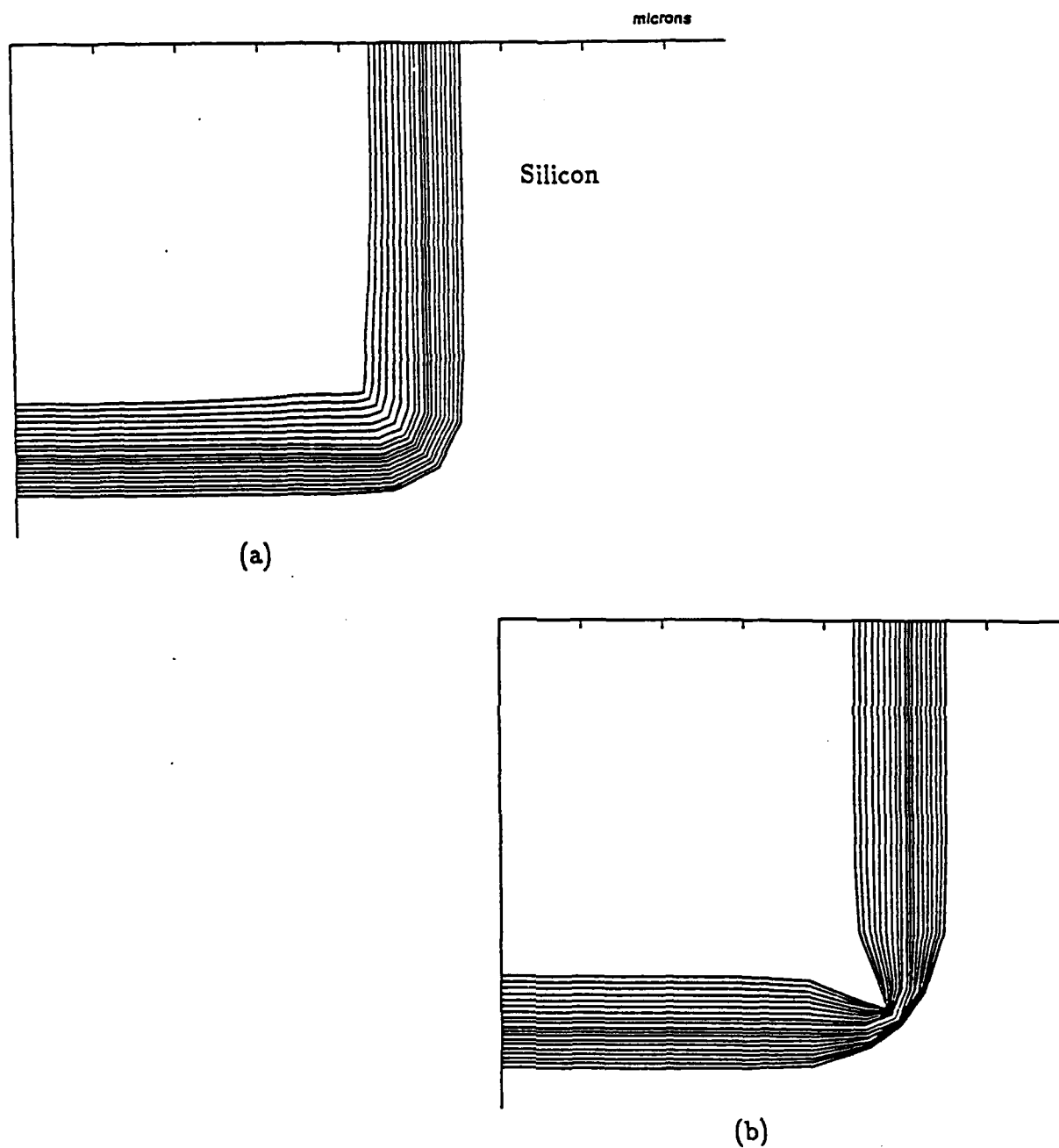


Figure 5.5: Effect of Stress on Diffusivity. (a) No stress effect is considered. (b) The effect of hydrostatic pressure on diffusivity is modeled. The oxide thinning in the corner is due to a large compressive stress in that region.

5.3 Modeling Kao's Experiments

In this section, we demonstrate a simplified viscoelastic flow model for fitting Kao's experimental data. This model ignores the crystallographic dependency of the reaction rate k_s , it solves radial symmetric problems whose solutions are pseudo one-dimensional in nature. The oxide is assumed to be viscoelastic. An effective diffusivity and viscosity is assumed for the domain. Shown below are the stress models:

$$\begin{aligned} k_s &= k_{s0} \exp\left(\frac{p_n V_{ks}}{kT}\right) \\ D &= D_0 \exp\left(\frac{\sigma_M V_{ks}}{kT}\right) \\ \eta &= D_0 \exp\left(-\frac{\sigma_M V_{\eta 1}}{kT}\right) \frac{\sigma_Y V_{\eta 2}/kT}{\sinh(\sigma_Y V_{\eta 2}/kT)} \end{aligned}$$

where $\sigma_Y = \sqrt{C_Y}$. (C_Y is the maximum distortion energy.) No maximum limit is placed on any parameters. With the following fitting values,

Temp(°C)	η_0 (poise)	$V_{\eta 1}$ (Å ³)	$V_{\eta 2}$ (Å ³)	V_{ks} (Å ³)	V_D (Å ³)
800	2×10^{15}	120	180	45	50
900	5×10^{14}	145	220		
1000	1×10^{14}	170	340		
1100	2.5×10^{13}	400	800		
1200	5×10^{11}	200	800		

reasonable simulation results have been obtained. They are compared with Sutardja's in Fig. 5.6. As it can be seen, they all agree very well. However, the fittings values are not close. For instance, our V_{ks} and V_D are 45 Å³ and 50 Å³ respectively, as compared to 12.5 Å³ and 75 Å³ in Sutardja's model. Besides the differences in formulation and numerical approximations, the disagreement may also be attributed to the scaling problem of activation volumes, which is explained in the next section.

5.3.1 Arbitrary Scaling of Activation Volumes

We have more unknown parameters than data points can fit for thermal oxidation processes. Part of the problem is that we are unable to measure local stress *in situ* or other activities. Experimental measurements are usually given in terms of oxide thickness – a collective parameter due to many stress effects. It is difficult to come up with a unique set of activation volumes for the stress effects.

A problem inherit in incompressible viscous flow methods is that the activation volumes for stress parameters are scalable – their values can be changed without altering the magnitudes of retardation effects. The root of the problem is that stress is proportional to the viscosity and that all the stress equations are of the form

$$f\left(\frac{\sigma V}{kT}\right),$$

which is restructured to show the dependency on η :

$$f\left(\frac{A\eta V}{kT}\right).$$

we realize that one important consideration in this expression is the product consisting of viscosity and activation volume, ηV . As long as this product is kept constant, the value of f will be the same, and hence the stress effects will be the same too. Hence, doubling all activation volumes and halving the viscosity will produce the same reduction in oxide thickness, although the stress values are reduced by half. We are free to choose any viscosity/activation volumes combination to get any stress values we want. Kao eliminates this degree of freedom by fixing V_{k_s} , the activation volume for k_s , to be $25\text{\AA}$, which is the net volume increase of an oxidized Si atom. In Sutardja's model, V_{k_s} is $12.5\text{\AA}$. Probably it has to do with the silicon monoxide phrase that an Si atom has to go through in the oxidation event. In any case, unless we have better measurement of oxide viscosity or stress, the values for activation volumes will still be debatable.

We have also attempted to model Kao's experiments in true two dimensions, taking in account the orientation dependency of the reaction rate. Unfortunately, the results

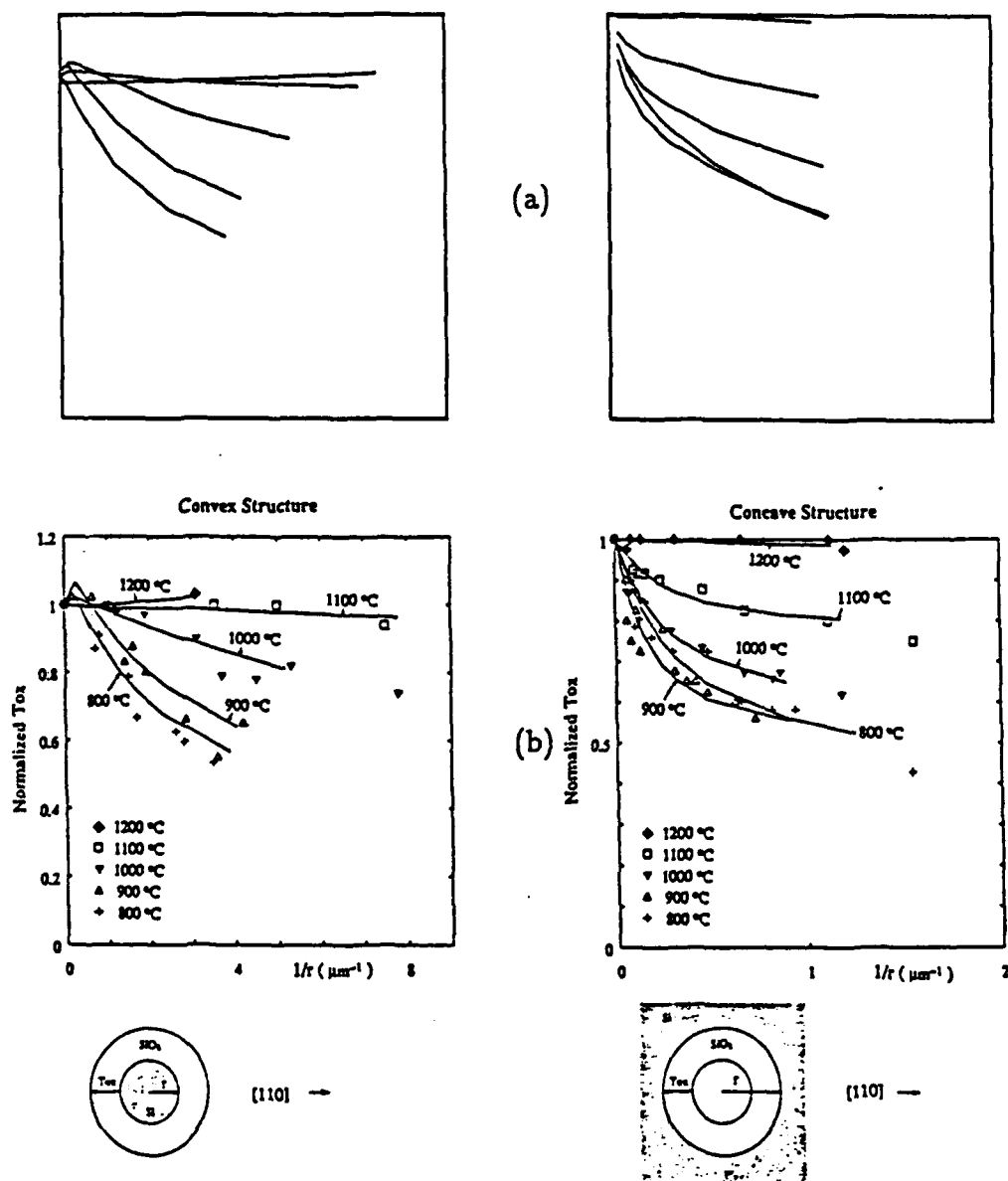


Figure 5.6: Simulations of Oxide Growth on Cylindrical Structures. (a) Simplified viscoelastic flow model. (b) Incompressible viscous flow model.

do not matched very well with the experimental data. Possibly, part of the problem has to do with the fact that stress history is not propagated from one timestep to another when operating under nonlinear viscosity condition, due to programing constraints.

Chapter 6

Conclusions

Thermal oxidation of silicon involves two parallel processes, namely the motion of oxide caused by volume expansion and the diffusion of oxidants such as H_2O or O_2 . An "indirect" boundary element approach for modeling oxidation has been introduced. The diffusion problem is modeled as a standard Laplace problem using the scalar potential formulation; the motion of oxide is solved with a vector potential formulation.

Three different motion models, namely, viscous flow, elastic deformation, and viscoelastic flow, are considered. The BEM formulation for these models are essentially identical, with the only difference in the kernel functions (or Green's functions). A Laplace transform technique is used to transform Kelvin's solution, which is a fundamental solution for elasticity, into a viscoelastic kernel function. A constant-velocity loading function is chosen to ensure a wide range of operation conditions. Essentially, it can model elastic deformation as well as incompressible viscous flow problems.

Stress generated during oxidation steps affects many variables – some boundary parameters, some bulk parameters. Diffusivity of oxidants and viscosity of oxide are the bulk parameters rendered nonhomogeneous and nonlinear by stress. For them, domain calculations are needed. A unified method for dealing with nonhomogeneity in both diffusion and viscoelastic flow has been developed. This initial stress/built-

in field scheme utilizes pseudo domain sources to generate correction terms to the solutions.

6.1 Future BEM in Thermal Oxidation

For what it is designed for, namely, linear systems, the BEM performs very well. In fact, it has been found to be superior to the finite element method in fracture mechanics. In applications where stress is highly concentrated in a small region, the BEM is able to generate good results without much refinement on the boundary segments. In contrast, the finite element method requires fine-tuning of the mesh to get accurate results. The superior performance of the BEM comes from Green's function that is singular – it is difficult to deal with, but it can model large change in stress field accurately.

The situation with nonlinear problems is less clearly defined. In dealing with non-homogeneous materials, the BEM requires domain calculations with either interior cells or subregions. In places where strong variations in the bulk parameters are observed, large number of interior cells or subregions may be required. Granted, no cells or subregions are needed in regions where stress is insignificant. However, for high stress areas, the cell generation requirements may be similar to those of the FEM.

Given the fact that the matrix setup operation is $O(n^2)$ and the solution is $O(n^3)$, where n is total number of boundary elements and interior cells (or interfacial sources), a severe computational penalty is imposed on the BEM when there is a sizable population of cells.

To determine the viability of BEM in thermal oxidation modeling, more investigative works are needed on iterative methods for nonlinear solutions. Currently, the domain pseudo sources and boundary sources are solved simultaneously. In other BEM

applications, the domain perturbation is not incorporated in the system matrix but calculated externally in an iterative manner. Such schemes help to reduce the matrix solution time significantly.

6.2 Proposed Experimental Studies

In order to function effectively, process simulators need accurate models and parameter values. To determine the effects of stress on various parameters in an oxidation model, stress must be introduced to the test samples in a controlled manner so that the results can be quantified. So far, Kao's experimental technique of using cylindrical structures represents one of the best approaches.

More experimental studies and measurements can be done with cylindrical structures. Consider the crystal orientation on the reaction rate. Only the thinning effects on the $[110]$ surfaces have been presented. The $[100]$ surfaces should be able to provide another set of data for simulation verifications. Before that is done, we need to know how the reaction rate changes as the crystal orientation varies from $[110]$ to $[100]$. This is important because stress distribution is sensitive to the orientation dependency of the reaction rate. Different oxide thicknesses should also be tried to extract more information. Consider the stress behavior. Pressure is primarily a function of the free surface curvature; however, the surface traction on the silicon interface depends strongly on the thickness of the oxide. Thus, growing different oxide thicknesses may help to differentiate the contributions of these two mechanisms to the growth retardation phenomenon. The activation volumes in stress functions are scalable for an oxidation model based on incompressible viscous flow of oxide. Unless stress or viscosity is measured independently, there will be uncertainty in the range of values for activation volumes.

Appendix A

Miscellaneous Formulations

The formulae for principal stresses, stress transformation, and analytical line integrations are given in this appendix.

A.1 Principal Stresses

The values for the principal stresses are given by the roots of this determinant:

$$\begin{vmatrix} \sigma_{11} - \sigma & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} - \sigma & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} - \sigma \end{vmatrix} = 0. \quad (\text{A.1})$$

In plane strain, $\sigma_{13} = \sigma_{23} = \sigma_{32} = \sigma_{31} = 0$. Thus, the determinant is reduced to the following equation:

$$(\sigma - \sigma_{33}) [(\sigma_{11} - \sigma)(\sigma_{22} - \sigma) - \sigma_{12}\sigma_{21}] = 0.$$

The roots are given by:

$$\begin{aligned} \sigma &= -\sigma_{33}, \\ \sigma &= \frac{1}{2} \left[\sigma_{11} + \sigma_{22} \pm \sqrt{(\sigma_{11} - \sigma_{22})^2 + 4\sigma_{12}^2} \right]. \end{aligned}$$

They are ordered so that $\sigma_I \geq \sigma_{II} \geq \sigma_{III}$.

A.2 Stress and Strain Transformation

It is often needed to transform stress and strain field from the global coordinate frame to a local frame. The expression for two-dimensional stress transformation is

$$\begin{Bmatrix} \sigma'_{11} \\ \sigma'_{22} \\ \sigma'_{12} \end{Bmatrix} = \begin{bmatrix} \cos^2 \theta & \sin^2 \theta & \sin 2\theta \\ \sin^2 \theta & \cos^2 \theta & -\sin 2\theta \\ -\frac{1}{2} \sin 2\theta & \frac{1}{2} \sin 2\theta & \cos 2\theta \end{bmatrix} \begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix}. \quad (\text{A.2})$$

where θ is the angle measured counterclockwise from the x axis of the old frame to the x' axis of the new frame. For strain, the transformation rule is slightly different:

$$\begin{Bmatrix} \epsilon'_{11} \\ \epsilon'_{22} \\ \epsilon'_{12} \end{Bmatrix} = \begin{bmatrix} \cos^2 \theta & \sin^2 \theta & \frac{1}{2} \sin 2\theta \\ \sin^2 \theta & \cos^2 \theta & -\frac{1}{2} \sin 2\theta \\ -\sin 2\theta & \sin 2\theta & \cos 2\theta \end{bmatrix} \begin{Bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{12} \end{Bmatrix}. \quad (\text{A.3})$$

For comparison, the rotation expression for a vector is show here:

$$\begin{Bmatrix} f'_1 \\ f'_2 \end{Bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix}. \quad (\text{A.4})$$

With reference to Eq. A.2, one notes that, for the uniaxial stress problem, a maximum σ'_{12} occurs when $\theta = 45^\circ$, as mentioned in Chapter 2.

A.3 Special Analytical Line Integrals

In determining the influence of a source distribution on the collocation point of the same segment, analytical technique must be employed as we are integrated over a singularity. The results are simple because the collocation point is located at the center and the source distribution is assumed to be constant. The integration path is chosen such that the source does not lie on the boundary but just outside the domain (0^+). Shown below are the results for the seven basis functions

$$\int_{-l/2}^{l/2} \phi dl = l(\log(\frac{l}{2} - 1)) \quad (\text{A.5})$$

$$\int_{-l/2}^{l/2} \phi_{,1} dl = -\pi \sin \theta \quad (\text{A.6})$$

$$\int_{-l/2}^{l/2} \phi_{,2} dl = \pi \cos \theta \quad (\text{A.7})$$

$$\int_{-l/2}^{l/2} x_2 \phi_{,1} dl = l \sin \theta \cos \theta \quad (\text{A.8})$$

$$\int_{-l/2}^{l/2} x_2 \phi_{,2} dl = l \sin^2 \theta \quad (\text{A.9})$$

$$\int_{-l/2}^{l/2} x_2 \phi_{,11} dl = 2\pi \sin^2 \theta \cos \theta \quad (\text{A.10})$$

$$\int_{-l/2}^{l/2} x_2 \phi_{,12} dl = \pi \sin \theta (\sin^2 \theta - \cos^2 \theta) \quad (\text{A.11})$$

where l is the length of the segment, and θ the tilt angle of the segment with respect to the $\hat{x}_1$ axis, measured counter-clockwise.

A.4 General Analytical Line Integrals

In this section, we consider the integral results for the case when the location of the collocation point is arbitrary with respect to the source segment. For simplicity, the solutions are given in terms of a local reference frame in which the source segment (*i.e.* the integration path) appears to orientate vertically (in the $\hat{y}$ direction), going from (a_0, b_0) to (a_0, b_1) .

$$\int \phi = b_1 \left(\frac{1}{2} \log(a_0^2 + b_1^2) - 1 \right) - b_0 \left(\frac{1}{2} \log(a_0^2 + b_0^2) - 1 \right) \quad (\text{A.12})$$

$$-a_0 \left(\log \left(\frac{a_0^2 + b_1^2}{a_0^2 + b_0^2} \right) \right) \quad (\text{A.13})$$

$$\int \phi_{,1} = -a_0 \left(\tan^{-1} \left(\frac{b_1}{a_0} \right) - \tan^{-1} \left(\frac{b_0}{a_0} \right) \right) \quad (\text{A.14})$$

$$\int \phi_{,2} = -\log \left(\frac{a_0^2 + b_1^2}{a_0^2 + b_0^2} \right) \quad (\text{A.15})$$

$$\int x_2 \phi_{,1} = a_0 \log \left(\frac{a_0^2 + b_1^2}{a_0^2 + b_0^2} \right) \quad (\text{A.16})$$

$$\int x_2 \phi_{,2} = a_0 \tan^{-1} \left(\frac{b_1}{a_0} \right) - \tan^{-1} \left(\frac{b_0}{a_0} \right) + \frac{1}{2}(b_0 - b_1) \quad (\text{A.17})$$

$$\int x_2 \phi_{,12} = \tan^{-1} \left(\frac{b_1}{a_0} \right) - \tan^{-1} \left(\frac{b_0}{a_0} \right) + 2a_0 \left(\frac{b_0}{a_0^2 + b_0^2} - \frac{b_1}{a_0^2 + b_1^2} \right) \quad (\text{A.18})$$

$$\int x_2 \phi_{,22} = \log \left(\frac{a_0^2 + b_1^2}{a_0^2 + b_0^2} \right) + 2a_0^2 \left(\frac{1}{a_0^2 + b_1^2} - \frac{1}{a_0^2 + b_0^2} \right). \quad (\text{A.19})$$

Bibliography

- [1] D. Chin, *Two Dimensional Oxidation Modeling and Applications*. PhD thesis, Stanford University, 1983.
- [2] K. Y. Chiu, R. Fang, J. Lin, C. Lage, S. Angelos, and R. Tillman, "The SWAMI - a defect free and near-zero bird's-beak local oxidation process and its application in VLSI technology," in *1982 IEDM Tech. Dig.*, p. 224, IEEE, 1982.
- [3] P. Chatterjee and 4Mb dRAM Team, "Trench and compact structures for drams," in *1986 IEDM Tech. Dig.*, p. 128, IEEE, 1986.
- [4] T. T. Sheng and R. B. Marcus, "Gate oxide thinning at the isolation oxide wall," *J. Electrochem. Soc.*, vol. 125, p. 432, 1978.
- [5] Y. Sakina, T. Ohno, and S. Matsumoto, "Two-dimensional analysis of thermal oxidation of silicon," *Jap. J. Appl. Phys.*, vol. 22, p. L514, 1983.
- [6] D. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-dimensional thermal oxidation of silicon - i. experiments," *IEEE Trans. Elec. Dev.*, vol. ED-34, p. 1008, 1987.
- [7] D. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-dimensional thermal oxidation of silicon - ii. modeling stress effects in wet oxides," *IEEE Trans. Elec. Dev.*, vol. ED-35, p. 25, 1988.
- [8] E. P. EerNisse, "Stress in thermal SiO_2 during growth," *Appl. Phys. Lett.*, vol. 35, p. 8, 1979.
- [9] W. Hartmann and G. Franz, "Real-time x-ray topography studies of the viscoelastic behavior of SiO_2 in the system Si/SiO_2 ," *Appl. Phys. Lett.*, vol. 37, p. 1004, 1980.
- [10] K. Kobayashi, Y. Inoue, T. Nishimura, T. Nishioka, H. Arima, M. Hirayama, and T. Matsukawa, "Stress measurement of LOCOS structure using microscopic

- raman spectroscopy," in *19th Conference on Solid State Devices and Materials, Tokyo*, p. 323, 1987.
- [11] S. Kambayashi, T. Hamasaki, T. Nakakubo, M. Watanabe, and H. Tango, "Investigation on microarea stress in silicon by microprobe raman spectroscopy," in *19th Conference on Solid State Devices and Materials, Tokyo*, p. 323, 1987.
- [12] G. E. Davis and M. E. Taylor, "A new method for the investigation of stress in MIS devices using SEM electron channeling patterns," *J. Vac. Sci. Technol.*, vol. 19, p. 1024, 1981.
- [13] B. E. Deal and A. S. Grove, "General relationship for the thermal oxidation of silicon," *J. Appl. Phys.*, vol. 36, p. 3770, 1965.
- [14] A. Poncet, "Finite-element simulation of local oxidation of Silicon," *IEEE Trans. Comp. Aided Design*, vol. CAD-4, p. 41, 1985.
- [15] C. Rafferty and R. Dutton, "Two dimensional oxidation," in *SRC Topical Research Conference - Advanced MOS Process and Device Models*, p. ??, SRC, 1986.
- [16] P. Sutardja, W. G. Oldham, and D. B. Kao, "Modeling of stress-effects in silicon oxidation including the non-linear viscosity of oxide," in *1987 IEDM Tech. Dig.*, p. 264, IEEE, 1987.
- [17] D. Chin, S. Y. Oh, and R. Dutton, "A general solution method for two-dimensional non-planar oxidation," *IEEE Trans. Elec. Dev.*, vol. ED-30, p. 993, 1983.
- [18] T. Tung and D. A. Antoniadis, "A boundary integral equation approach to oxidation modeling," *IEEE Trans. Elec. Dev.*, vol. ED-32, p. 1954, 1985.
- [19] H. Matsumoto and M. Fukuma, "Numerical modeling of nonuniform Si thermal oxidation," *IEEE Trans. Elec. Dev.*, vol. ED-32, p. 132, 1985.
- [20] S. Isomae and S. . Yamamoto, "A new two-dimensional silicon oxidation model," *IEEE Trans. Comp. Aided Design*, vol. CAD-6, p. 410, 1987.
- [21] N. Kaneko, T. Shinokawa, N. Yoshida, and M. Kawahara, "Numerical analysis of viscoelasticity using boundary element method," in *Proceedings of 4th International Conference on Applied Numerical Modeling*, p. 473, 1985.
- [22] M. Tanaka, "New boundary element methods for viscoelastic problems," in *BETECH 85. Proceedings of the 1st Boundary Element Technology Conference, South Australian Institute of Technology, Adelaide, November 1985*, (C. A. Brebbia and B. J. Noye, eds.), p. 129, Springer-Verlag: New York, 1985.

- [23] R. Butterfield, "An application of the boundary element method to potential flow problems in generally inhomogeneous bodies," in *Recent Advances in Boundary Element Methods*, (C. Brebbia, ed.), p. 123, Pentech Press, 1978.
- [24] S. M. Sze, *VLSI Technology*. Wiley and sons, 1983.
- [25] S. K. Ghandi, *VLSI Fabrication Principles*, p. 382. John Wiley and Sons, 1983.
- [26] J. Blanc, "A revised model for the oxidation of Si by oxygen," *Appl. Phys. Lett.*, vol. 33, p. 424, 1978.
- [27] C. Han and R. Helms, "Parallel oxidation mechanism for Si oxidation in dry O₂," *J. Electrochem. Soc.*, vol. 134, p. 1297, 1987.
- [28] W. A. Tiller, "On the kinetics of the thermal oxidation of silicon (i. a theoretical perspective)," *J. Electrochem. Soc.*, vol. 127, p. 619, 1980.
- [29] D. Kao, *Two-Dimensional Oxidation Effects in Silicon - Experiments and Theory*. PhD thesis, Stanford University, 1986.
- [30] P. G. Shewmon, *Diffusion in Solids*. McGraw Hill, 1963.
- [31] E. P. EerNisse, "Viscous flow of thermal SiO₂," *Appl. Phys. Lett.*, vol. 30, p. 290, 1977.
- [32] R. Bruckner, "Properties and structure of vitreous silica. i.," *J. Non-Crystalline Solids*, vol. 5, p. 123, 1970.
- [33] R. Bruckner, "Properties and structure of vitreous silica. ii.," *J. Non-Crystalline Solids*, vol. 5, p. 177, 1970.
- [34] E. A. Irene, E. Tierney, and J. Angilello, "A viscous flow model to explain the appearance of high density thermal SiO₂ at low oxidation temperatures," *J. Electrochem. Soc.*, vol. 129, p. 2594, 1982.
- [35] G. Hetherington, K. H. Jack, and J. C. Kennedy, "The viscosity of vitreous silica," *J. Phys. Chem. Glasses*, vol. 5, p. 130, 1964.
- [36] W. Flugge, *Viscoelasticity*, ch. 8, p. 112. Blaisdell Publishing Co., 1967.
- [37] G. F. Carrier and C. E. Pearson, *Partial Differential Equations*. Academic Press, 1976.
- [38] M. A. Jaswon and G. T. Symm, *Integral Equation Methods in Potential Theory and Elastostatics*. Academic Press, 1977.

- [39] A. Wexler, "Some applications of the boundary element method to electrical engineering problems," in *Recent Advances in Boundary Element Methods*, (C. A. Brebbia, ed.), p. 405, Pentech Press, 1978.
- [40] P. K. Banerjee and R. Butterfield, *Boundary Element Methods in Engineering Science*. MacGraw Hill, London, 1981.
- [41] G. R. Baker, D. I. Meiron, and S. A. Orszag, "Generalized vortex methods for free-surface flow problems," *J. Fluid Mech.*, vol. 123, p. 477, 1982.
- [42] C. A. Brebbia, J. C. F. Telles, and L. C. Wrobel, *Boundary Element Techniques*. Springer-Verlag, 1984.
- [43] W. E. Langlois, *Slow Viscous Flow*. The Macmillan Co., 1964.
- [44] G. Dahlquist and Åke Björck, *Numerical Methods*. Prentice-Hall, 1974.
- [45] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag: New York, 1984.