Two-dimensional process modeling: A description of the SAFEPRO program

by R. R. O'Brien
C. M. Hsieh
J. S. Moore
R. F. Lever
P. C. Murley
K. W. Brannon
G. R. Srinivasan
R. W. Knepper

This paper describes the development, testing, and application of a finite element program which simulates the processes used in manufacturing transistors. The profiles calculated by the program can be input directly into a device analysis program. The paper includes a description of the physical phenomena modeled and explains the choice of the particular numerical methods used to solve the resulting equations. It shows an example of the application of the program to the design and sensitivity study of a submicrometer shallow-junction bipolar transistor and presents results obtained when an oxide is grown on boron-doped silicon.

1. Introduction

The SAFEPRO program (Semiconductor Applications of Finite Elements to PROcessing) was developed to provide two-dimensional doping profiles for use in modeling advanced bipolar transistors. Because of their large areas and deep junctions, early semiconductor devices were essentially one-dimensional structures. The effect of lateral diffusions on their behavior was limited to a small percentage increase in capacitance and a decrease in junction breakdown voltage. As a result, early two-dimensional process simulators were relatively simple programs which used first-order approximations to model diffusion processes and solved the resulting equations analytically by series expansions [1] or finite difference methods [2]. However, with the present trend towards miniaturized submicrometer shallow bipolar devices, a two-dimensional process simulator of greater sophistication has become essential to accurately model and optimize the process and the device.

In today's advanced bipolar transistors, in order to achieve a low-resistance base contact, a narrow base width, and a low junction capacitance, the extrinsic base and the intrinsic base of a transistor are often formed separately. In these cases, the lateral diffusions become crucial in determining the base resistance and the junction capacitance. Furthermore, the interaction between donor and acceptor species significantly affects both the vertical and lateral profiles of the transistor. This species interaction, which is included in one-dimensional process simulators [3], was not developed for two-dimensional cases until recently. The need to model this interaction accurately was one of the primary reasons for the development of SAFEPRO.

SAFEPRO is based on the existing finite element program SAFE [4]. Since SAFE solves coupled systems of partial differential equations, it is easily modified to treat the interaction between two different doping species in process simulation. SAFE was designed so that the code determining

the particular equations being solved appears only in a few specific subroutines. This makes it possible to change the equations by changing only those subroutines. This approach was retained in SAFEPRO. The code containing the diffusion coefficients and other information pertaining to the models used in SAFEPRO is also restricted to a few specific subroutines. This program structure greatly simplifies the task of changing the diffusion models used in the program. This approach is important because the physical understanding of diffusion is advancing rapidly and a general-purpose program such as SAFEPRO must be capable of continual modification. Since SAFEPRO is a finite element program, it has the ability to use irregular meshes that can be customized to a particular device instead of being limited to the quasi-rectangular grids characteristic of finite difference programs.

SAFEPRO is part of a series of three programs developed and linked together at the IBM General Technology Division laboratory, East Fishkill, New York, to perform advanced bipolar device modeling and equivalent circuit model generation. The present paper is part of a series of four papers published in this issue that describe the application and the details of this modeling methodology. The two following papers describe the two-dimensional device physics modeling program 2DP [5] and the three-dimensional equivalent circuit model generation program MGP [6]. The previous paper gives an overview of the complete modeling methodology, including the software links among the three programs, and also describes a sample application of the modeling procedure for an advanced bipolar technology [7].

2. Description of the program

• *Program operation*

To run SAFEPRO a user must create both a finite element mesh that discretizes the device geometry and an input dataset describing the various process steps to be simulated. A finite element mesh generator program is used to generate the meshes. The program accepts a high-level description of the desired mesh in terms of boundary lines and the number of subdivisions of these lines. It creates a trial mesh which the user may accept or further modify by changing the high-level description. It also has the capability to assign an identifying integer to regions of the mesh. This integer is used by SAFEPRO to define the material in that region of the mesh. The mesh is chosen to provide as much resolution as possible in the regions where the dopant profile is expected to change most rapidly. Examples of such meshes are shown in Section 3, which illustrates the simulation of a device using SAFEPRO.

The input dataset allows the user to describe process parameters and request that solutions be printed or saved for future use as starting values for a new calculation. SAFEPRO has defaults for all process parameters, such as diffusion and segregation coefficients, and the input dataset need contain only those parameters which are to be modified. The ability to save and later reuse solutions as starting values may be utilized to simulate the etching of a structure or the deposition of material on a structure in specific areas. A saved solution may be used in combination with a finite element mesh which represents either the addition or the deletion of a region from the solution's original mesh, thus enabling the user to simulate a deposition or an etching of material. The solution in an "etched" region is ignored. The solution in a "deposited" region is set to zero, to be specified later by a SAFEPRO input distribution option. A deposition of material of any desired doping may in this way be added to any previously calculated structure.

Since a user is often interested in the integrated base doping of the intrinsic transistor, SAFEPRO will print out this value calculated along any vertical cross section that the user specifies.

• *Solution transfer to device analysis programs*

After a SAFEPRO simulation has been completed, it is necessary to transfer the solution to a device analysis program. In general, a mesh suitable for process simulation is not suitable for use in a device analysis program because the different programs need different degrees of resolution in different regions of the structure. Usually it is not practical to avoid this problem by choosing a mesh which simultaneously has adequate resolution for both types of programs.

To solve this problem, an auxiliary program named LINK was created. LINK reads a finite element mesh and a saved solution obtained by SAFEPRO using that mesh. It also accepts a list of x-y coordinates at which the device analysis program requires dopant concentrations. The LINK program finds the element containing the requested point and performs a logarithmic interpolation in that element to obtain the interpolated dopant value. The dopant values are stored in a dataset that can be read by the device analysis program. If an x-y coordinate is outside the saved solution mesh, a negative dopant value is returned as an indicator to the device analysis program.

• *Input distribution options*

SAFEPRO provides four different options with which a dopant distribution can be added to a solution. These options may be used for simulating an ion implantation or for some other initial distribution. They may be used at the beginning of any process step. These options are also useful in allowing the user to bypass process steps which produce known profiles, such as a low-concentration drive-in which gives rise to a Gaussian distribution. They may also be used to replace the initial stages of a diffusion from a constant source into a low doped substrate.

The first implant option requires as input the energy and dose of the ion implantation. SAFEPRO then constructs the implanted dopant profile based on the LSS statistics [8]. The arsenic profile is a joined half-Gaussian and the boron profile a Pearson Type IV with an exponential tail to simulate channeling [3]. Boron may be implanted as BF_2 if desired. The second implant option provides a Gaussian profile when the range, dose, and sigma values are supplied. The third gives joined half-Gaussians when the range, dose, and the left and right sigma values are supplied. The fourth option allows the user to supply an arbitrary doping profile in the form of a list of depths and associated dopant concentrations. This option gives the user the ability to create a profile that might be needed to model experimental data which could not be simulated by one of the first three options.

With each of these options the user can supply mask information so that the implant will be restricted to portions of the structure. This allows the simulation of implantation through a mask. When any of these options is used to model implantation through a mask, the profile is extended laterally beyond the mask edge by the use of a complementary error function approximation [9]. Default values for this spreading are provided, but may be overridden if the user desires.

• *Physics of process simulation*

The physical models describing the diffusion process used in SAFEPRO are largely those used in SUPREM II, the one-dimensional process modeling program widely distributed by Stanford University. This allowed SAFEPRO to be readily compared with SUPREM II during its development by comparing essentially one-dimensional SAFEPRO runs with SUPREM II runs. The major difference between the SUPREM II and SAFEPRO approaches is that SUPREM II does not use the electric field directly but modifies Fick's Law to account for the electric field. That approach is exact for single-species diffusion with the diffusivity proportional to the concentration, but otherwise is an approximation with an unknown error.

The motion of doping atoms during semiconductor processing is conventionally described by a generalized transport equation of the form

$$\vec{\nabla}\cdot\vec{F} = -\frac{\partial N}{\partial t}, \tag{1}$$

where $\vec{F}$ is the flux of the doping atoms, which can be expressed as

$$\vec{F} = -D\vec{\nabla}N \pm \mu N\vec{E}, \tag{2}$$

with D the diffusion coefficient, N the density of the dopant atoms, μ the mobility, and $\vec{E}$ the electric field. The electric field term appears because the diffusing atoms are electrically charged at diffusion temperatures; its sign is positive if the dopant is an acceptor and negative if the dopant is a donor.

For boron the diffusion coefficient has the form

$$D_B = \left(D_o e^{-\frac{Q}{kT}}\right)\cdot\left(\frac{1+\beta\frac{p}{n_i}}{1+\beta}\right), \tag{3}$$

where the first factor is the intrinsic diffusivity of the dopant in the particular material. The values of D_o and Q are reasonably well known in silicon. The second factor is the enhanced diffusivity at high doping concentrations attributed to charged point defects in silicon. In the second factor p is the local hole concentration, n_i the intrinsic carrier density, and β the vacancy-enhanced diffusion parameter. The parameter β is generally adjusted to allow calculated results to match experimental results for the particular process being modeled. While β is in principle temperature dependent, its value is not sufficiently well known to justify attempting to include that dependence in this simulation. If oxide is being grown on silicon, the diffusivity of boron is greatly increased [10]. SAFEPRO has a feature that allows the user to increase the boron intrinsic diffusivity in appropriate portions of the structure being modeled. This is necessary because the oxidation-enhanced diffusion region extends far into a silicon structure and is expected to affect diffusion even when small amounts of oxide are grown. The electric field is calculated from the charge-neutrality assumption and has the form

$$\vec{E} = -\frac{kT}{q}\frac{1}{p}\vec{\nabla}p. \tag{4}$$

For arsenic, the situation is more complicated because the diffusion coefficient does not continue to rise with concentration as in Eq. (3), but reaches a maximum and then declines [11]. This decrease is attributed to the formation of clusters at high concentrations. The clusters are immobile and do not participate in the diffusion directly, but may have an electric charge which would affect the local carrier concentration and the local electric field. The clustering process is probably not instantaneous on the diffusion time scale, but there is little information on the clustering dynamics [12]. Nevertheless, SAFEPRO treats clustering as instantaneous and absorbs any resulting errors into the value of β. The use of a finite clustering time would simplify the actual calculation because the doping profiles would change more slowly with time and the program would converge more readily. As a result of clustering, the arsenic flux will be

$$\vec{F} = -D_{As}\vec{\nabla}A - \mu A\vec{E}, \tag{5}$$

where A, the mobile active arsenic concentration, replaces the total dopant concentration N and the diffusion coefficient has the same form as the boron diffusion coefficient. However, in the second factor of the arsenic diffusion coefficient, the hole density p appearing in the boron expression is replaced by the electron density n. The

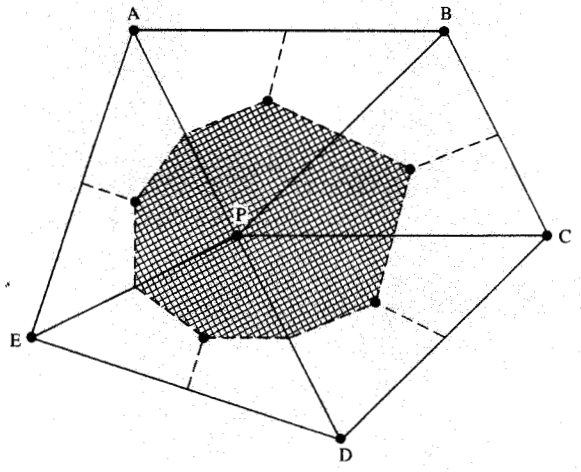

Figure 1

Spatial discretization of the continuity equation. The node P and the nodes A, B, C, D, and E, which together with P define the elements containing P, are shown. The nodal region associated with P is shown shaded.

relation between the concentration of active arsenic atoms and the total concentration depends on the number of atoms per cluster and the charge on each cluster [13]. By using the SUPREM II clustering model of a three-atom cluster carrying a charge of two electrons, the active arsenic is related to the total arsenic by the relation

$$N_{As} = A + KA^3(A + 2N_{As}), \tag{6}$$

where K is the empirically determined clustering coefficient. Given the total concentration N_{As}, it is possible to determine the active concentration by using a Newton-Raphson iteration.

The simultaneous diffusion of arsenic and boron is described by the coupled equations

$$\vec{\nabla}\cdot(-D_B\vec{\nabla}N_B + \mu N_B\vec{E}) = -\frac{\partial N_B}{\partial t},$$

$$\vec{\nabla}\cdot(-D_{As}\vec{\nabla}A - \mu A\vec{E}) = -\frac{\partial N_{As}}{\partial t}. \tag{7}$$

Note that these equations are coupled in two ways. First, the diffusion coefficients depend on the carrier densities. The presence of arsenic will reduce the hole density and thus the diffusivity of boron. Similarly, boron will reduce the diffusivity of arsenic. The electric field appears in both equations and is a function of both the arsenic and boron concentrations. It provides a second way for the different species to interact. When the electric field is calculated, each boron atom and each active arsenic atom is considered to be singly charged. The arsenic clusters in the models used here each have two units of charge. After some algebraic manipulation, the expression for the electric field becomes

$$\vec{E} = -\frac{kT}{q}\frac{\vec{\nabla}N_B - \vec{\nabla}\left(\frac{2}{3}N_{As} + \frac{1}{3}A\right)}{\sqrt{\left(N_B - \left(\frac{2}{3}N_{As} + \frac{1}{3}A\right)\right)^2 + 4n_i^2}}. \tag{8}$$

Another phenomenon of importance in process modeling that has to be taken into account is the possible appearance of boundary layer films of very low diffusivity between silicon and polysilicon regions. This fits naturally into the finite element approach. The low-diffusivity film is modeled by defining a new type of element, a "barrier" element possessing very low diffusivity. These elements can easily be defined during mesh generation. Treatment of this phenomenon using a finite difference approach would be much more involved.

Numerical methods of solving the transport equations

The first step in solving the transport equations is to discretize Eq. (1) in both space and time. Space discretization is dealt with first. SAFEPRO uses triangular elements with linear basis functions within each element. With each node of the finite element mesh is associated a region made up of parts of each element containing that node. These parts of each element are defined by the lines connecting the center of the element to the midpoints of the sides of the element. This method is shown in **Figure 1**. This particular approach has the advantage that obtuse triangles do not cause any numerical problems, as can be the case for other subdivision methods. Since automatic mesh generators often include some obtuse triangles whenever an irregular mesh is created, this property of the discretization method is advantageous.

SAFEPRO also has an option that allows it to interpolate linearly in the logarithm of the concentration instead of in the concentration itself. This procedure has the advantage of obtaining an improved fit to the solution within each element and therefore increased accuracy for a given mesh fineness. The log-linear interpolation increases computation time slightly because additional logarithm and exponential calculations are required.

An evaluation of the relative advantages of the two different methods is continuing. Preliminary results indicate that the log-linear interpolation will be cost-effective. Simulations of a one-dimensional constant diffusivity diffusion process which has a Gaussian as its exact solution gave a relative error of 19% for the log-linear interpolation versus 43% for the linear interpolation at 0.005 of the peak doping of the Gaussian. The CPU time for the log-linear calculation increased by 29% over that for the linear calculation. To obtain the same relative accuracy for the linear interpolation method would require an increase of approximately 60% in the number of nodes in the finite element mesh and therefore of approximately 60% in the

CPU time for a one-dimensional calculation. For a two-dimensional calculation the time increase would be approximately $1.60^{1.8} = 2.33$ or about 133%.

The flux density in each element is calculated at the center of the element, and the flux into each nodal region is defined to be the integral of the flux density component perpendicular to the nodal region boundary. As a result, the flow of dopant atoms within each element is locally conservative. That is, the sum of the fluxes between the parts of the three nodal regions in each element is 0. This local flux conservation implies global flux conservation, since the total flux into any collection of nodal regions must equal the flux out of the nodal regions bounding them. The global conservation of the flux of the dopant atoms is important because ion implantation establishes an initial dose of dopant atoms in the structure. The numerical technique used to solve the diffusion equations must conserve this initial dose if an acceptable solution is to be obtained. In order to compensate for the discretization error caused by the finite element grid, the implanted dose calculated by SAFEPRO is adjusted to equal the desired implanted dose immediately after each implantation and before any calculation takes place. If an oxide-silicon boundary is involved, no difficulty arises as long as the elements are chosen so that the element boundaries lie along the interface. The fluxes in each element are calculated using the nodal concentrations appropriate to oxide in an oxide element and silicon in a silicon element.

If the set of elements bounding a node is denoted by S and the area of the node by a, then the continuity equation can be written in spatially discretized form as

$$\sum_S FLUX = -a \cdot \frac{\partial N}{\partial t} , \tag{9}$$

where $FLUX$ is the integrated flux density out of the nodal regions of each element.

There are several possible ways to discretize Eq. (9) in time. Let t be the present time, at which all nodal concentrations are assumed to be known, and let $t + \Delta t$ be the next time value at which the new nodal concentrations are to be calculated. SAFEPRO uses primarily the scheme

$$\sum_S FLUX\,(t + \Delta t) = -a \cdot \frac{N_{t+\Delta_t} - N_t}{\Delta t} . \tag{10}$$

This is known as the backward Euler method. All of the fluxes are calculated at the next time value $t + \Delta t$ and hence depend on the concentrations at this new time, which are not known. This requires the solution of a large system of nonlinear equations at each time step. The method is termed implicit because the new concentrations cannot be obtained explicitly from the old but must be obtained implicitly by solving a nonlinear system of equations. As compensation for its computational complexity, the backward Euler method allows large time steps to be taken. The limit on the size of the time step is accuracy, not stability. Indeed, the backward Euler method is the most stable of the conventional time discretization schemes. The backward Euler method is first-order-accurate in time; that is, the truncation error due to time discretization is proportional to Δt.

A second possible approach is the forward Euler or explicit method. In this method all of the fluxes are calculated at the present time t. Then Eq. (9) becomes

$$\sum_S FLUX(t) = -a \cdot \frac{N_{t+\Delta_t} - N_t}{\Delta t} , \tag{11}$$

which can be explicitly solved for the nodal concentrations at time $t + \Delta t$. These concentrations can even be written explicitly in terms of the nodal values at time t. Although this method requires much less computation than the backward Euler, it has the disadvantage that the time step is limited by stability requirements, not accuracy. A particularly interesting method is due to DuFort and Frankel. This method is both explicit and stable. Unfortunately, as the time and space steps are made smaller, it does not converge to Eq. (1) but to that equation plus an additional term proportional to the time step divided by the mesh spacing. Therefore it is very difficult to evaluate the accuracy of the scheme as a function of the time step. Another implicit method, known as the Crank-Nicholson method, which uses an average of the fluxes at time t and time $t + \Delta t$, is very attractive. It is more complicated than the backward Euler but is second-order-accurate in time. An extensive discussion of time discretization methods is available [14].

The decision was made to use an implicit method because an analysis of typical calculations indicated that taking fewer but more expensive time steps would be more efficient than taking many inexpensive ones. This ruled out the forward Euler scheme. The earliest SAFEPRO version used the backward Euler method because it is the simplest stable scheme known and it was expected to give the fewest problems during the debugging stages of program development. A Crank-Nicholson scheme has been developed and is now available in SAFEPRO.

All of the implicit schemes require the solution of a large set of algebraic equations which are usually nonlinear. Many different methods are known by which equations of this type can be solved, and several excellent books are available [15, 16]. These methods can be divided into two classes. The first kind, commonly called linearization methods, provide very accurate solutions by a method such as Newton-Raphson iteration, which, starting with an initial solution, produces a global correction. These methods are computationally expensive but generally converge rapidly. Once the error becomes small, the next error will be approximately the square of its predecessor. This property is called quadratic convergence. It is very desirable, because to conserve the

amount of dopant in the structure, Eq. (10) must be solved very precisely, and a quadratically convergent algorithm makes this possible.

The other methods are known as iterative methods and gradually improve an initial solution by reducing local errors. These methods tend to converge linearly to the exact solution. In many cases it is possible to speed up the convergence by the use of acceleration techniques and even be more efficient than by using a linearization method. However, these acceleration techniques are usually specific to particular problems. Thus, the time required to solve a problem cannot be easily estimated in advance, since experimentation with acceleration parameters may be required. Also, it is often difficult to decide when the solution that has been obtained is accurate enough that the iteration process can be halted. Since SAFEPRO is intended for users who are not primarily experts in computation, but rather in devices, it was decided to use a direct solution method to minimize the burden on the user of specifying convergence controls or estimating computer time required for a solution.

A multidimensional Newton-Raphson method was chosen. The derivatives of the fluxes at each node are calculated with respect to the concentrations at all other nodes. An initial guess at the solution is obtained, and the amount by which Eq. (10) fails to equal zero is calculated for this initial solution. Then the Newton corrections ΔC to the nodal concentrations are the solutions of

$$\mathrm{J} \cdot \Delta C = r, \tag{12}$$

where J is the matrix of the flux derivatives with respect to the nodal concentrations and r is the vector of the residual imbalances at each node. J is also known as the Jacobian matrix of the fluxes with respect to the nodal concentrations. For finite element (or finite difference) problems the Jacobian matrix is usually sparse (i.e., contains many zeros) because each node has relatively few immediate neighbors.

The IBM Scientific Subroutine Library–Mathematics [17] contains subroutines designed to solve such sparse matrix problems efficiently. Since SAFEPRO handles the interaction between two species, its Jacobian matrix is not symmetric. Therefore, SAFEPRO uses the general sparse matrix routines. The subroutine DMOOP is used to order the matrix for efficient solution and the subroutines DMNSP and DMBSP to obtain the actual solution.

In order to increase program efficiency it is necessary to provide a way in which the time step can be automatically varied as the solution changes with time. The concentrations at the next time step are predicted and then compared with the actual values calculated. When the predicted values agree well with the actual solution, the time step is small enough and perhaps can be increased. When the prediction differs substantially from the calculated solution, the time steps are becoming too large and either the next time step should be smaller or the current time step is unacceptable and the calculation must be repeated with a reduced time step. This predicted solution is also useful since it provides a more accurate initial guess for the solution at the next time step than merely using the solution at the last time step.

A detailed theory exists as to the relation between the predicted and corrected values of the solution and the changes which should be made in the time step [18]. This theory allows the truncation error on each time step—the error due to approximating the variation of the nodal concentration with time by a polynomial—to be related to the size of the time step and the difference between the predicted and final nodal concentrations at the new time value. Once the allowable truncation error has been specified, SAFEPRO will calculate the maximum time step that can be taken. If the actual time step is less than the allowable value, the time step will be increased. As a precaution, the time step is never allowed to more than double. This helps to avoid time-step cycling. If the actual time step is too large, it will be reduced. A time step is never reduced by more than 50%; if the algorithm calls for a greater reduction than 50% in the next time step, the previous time step is deemed unacceptable and is retaken with a smaller step.

The SAFEPRO user specifies the allowable truncation error by giving two parameters, *REL* and *ABS*, for each dopant. They are the relative and absolute errors allowed per time step, and the truncation error *TRUN* for each species at each node must satisfy

$$TRUN < REL \cdot N + ABS,$$

where N is the concentration at the node. Typically, SAFEPRO uses $REL = 0.001$ and $ABS = 10^{16}$. A quite small value for the *REL* parameter is chosen so that the relative error will be small at large dopant concentrations. The *ABS* parameter value is chosen to be approximately the lowest dopant concentration of physical interest. This prevents the relative error restriction from unduly limiting the time step at low concentrations.

The time step control allows the user to specify three parameters—*STINIT*, *STMIN*, and *STMAX*. *STINIT* is the initial time step that SAFEPRO tries to take. If that time step is too large, SAFEPRO will reduce the time step to an acceptable value. *STMIN* is the minimum allowable time step. If the time step is reduced below *STMIN*, SAFEPRO will halt. This is an effective means of preventing the program from looping if an input error has been made which presents SAFEPRO with a nonphysical simulation problem. *STMAX* is the maximum time step SAFEPRO is allowed to take. This allows the user the option of forcing a minimum number of steps per process run.

This automatic time step control is very important in SAFEPRO, because the solution will change slowly during the initial stages of a typical calculation when the extrinsic

base diffusion is not interacting with the emitter diffusion. As the two dopants diffuse into each other, the interaction between them will cause the solution to change rapidly and smaller time steps will be required. As the diffusion process continues, and the peak concentrations drop, the solution may then slow in its time variation. A user cannot be asked to prescribe a suitable sequence of time steps, nor can he afford to use the smallest time step required at any time for the entire calculation. The usual approach in process simulation has been to start with a minimum time step and gradually increase it, based on the argument that diffusion is a smoothing process, and that the longer the diffusion proceeds the larger the time step can be. In our primary application, the modeling of lateral diffusions, the acceptor and donor species diffuse towards each other, and the smallest time step required will occur not at the start of the simulation but later, when the donors and acceptors have met and are interacting.

• *Estimation of diffusion parameters*

While SAFEPRO was being developed, a simultaneous effort was carried out to determine the fitting parameters for our particular processes. This was accomplished by matching one-dimensional SIMS profile measurements with SUPREM II. Since SUPREM II does not simulate polysilicon as such, an oxide layer with polysilicon-type diffusion coefficients was used in SUPREM to model the polysilicon; an interfacial barrier was simulated by using the oxide/silicon segregation or mass transfer coefficients of SUPREM II. By fine-tuning SUPREM II to available experimental data, it was possible to obtain parameters for use in SAFEPRO.

The SUPREM II runs were very useful as test cases for SAFEPRO. In general, it was possible to match SAFEPRO (on one-dimensional problems) exactly to SUPREM II except in regions of high electric field. An example of the agreement is given in **Figure 2**. The differences such as those shown are attributed to SUPREM II's use of a modified version of Fick's Law, which is not exact for two interacting dopants and which caused slight differences in regions of high electric field. (SUPREM III [19] now uses the standard form of Fick's Law.)

3. Application of SAFEPRO to an advanced transistor design

The cross section of an advanced bipolar transistor [20] analyzed with SAFEPRO is shown in **Figure 3**. The transistor is fabricated on an n-type epitaxial layer with a 500-nm flat zone. The base contact of the device is made by a polysilicon layer heavily doped with boron. The extrinsic base is formed by the out-diffusion of boron from the polysilicon into the epitaxial layer. A critical element in the design is achieving the proper amount of out-diffusion vertically and horizontally. If the boron does not diffuse far enough into the epitaxial region, good electrical contact to the intrinsic base will not be made. If the boron diffuses too far into the epitaxial region, it will meet the emitter diffusion at a high concentration. This will result in large depletion capacitance and possibly a poor-quality junction with a low avalanche breakdown. The combined sheet resistance of the extrinsic base and the p^+ polysilicon is of the order of 60 ohms/sq.

A typical dielectric sidewall is formed by conformal CVD deposition and directional RIE [21]. The emitter is self-aligned to the base polysilicon and defined by the sidewall spacer. A second polysilicon layer is deposited for the emitter and is heavily doped with As. The completed transistor has an emitter junction at a depth of 50 nm below the epitaxial/polysilicon interface and a 150-nm-wide intrinsic base region.

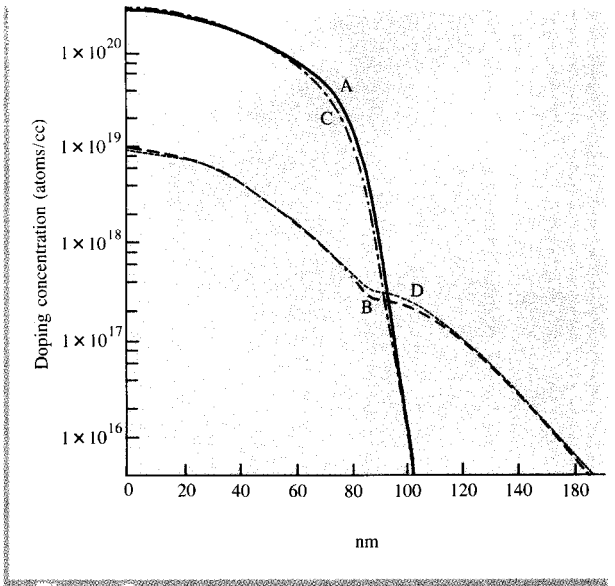

Figure 2

Comparison of SAFEPRO and SUPREM II. SAFEPRO–Curve A: arsenic, Curve B: boron; SUPREM II–Curve C: arsenic, Curve D: boron.

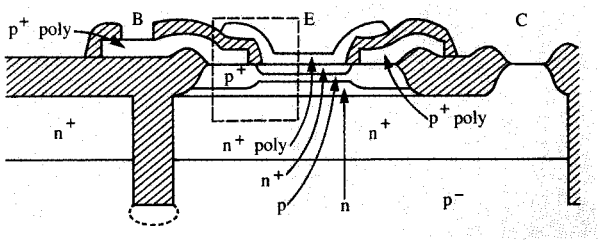

Figure 3

Schematic of a deep-groove-isolated self-aligned npn bipolar transistor.

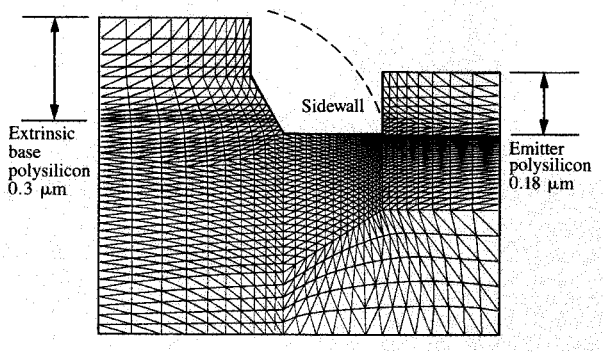

Figure 4

Finite element mesh used in the analysis.

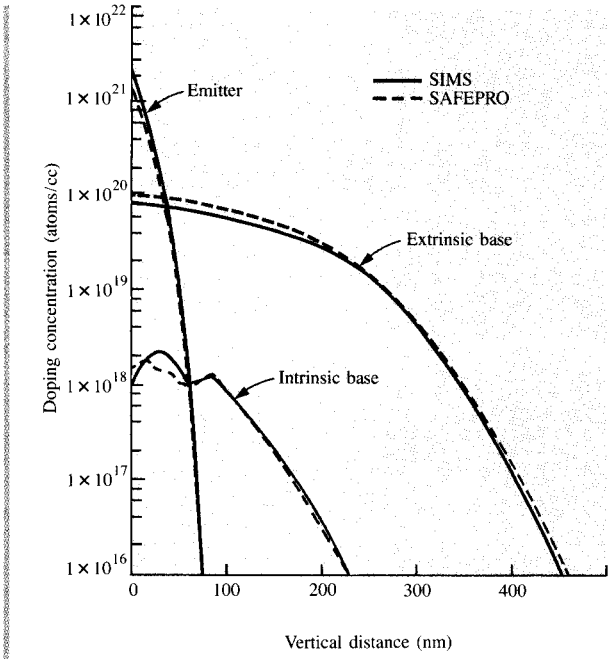

Figure 5

Vertical profiles at the center of the device from SAFEPRO and from SIMS.

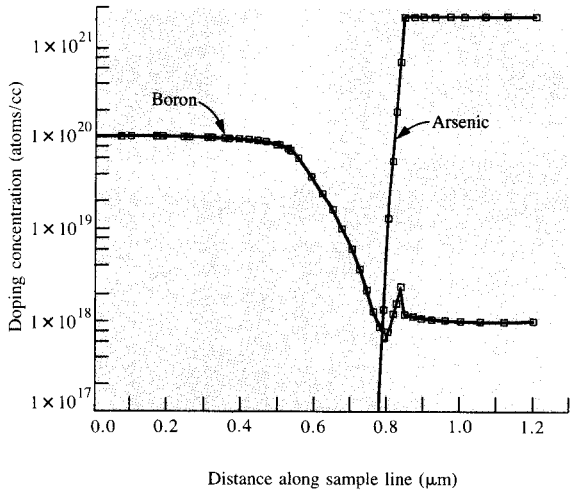

Figure 6

Calculated lateral profiles of arsenic and boron along the surface.

In order to reduce the computation time required and to provide the desired degree of resolution, only the portion of the structure within the rectangle of dotted lines in Fig. 3 was simulated. This region includes enough of the structure to allow the critical lateral base diffusion and its interaction with the emitter diffusion to be properly modeled. The triangular mesh used in the analysis is shown in **Figure 4**. Since the sidewall was formed before the extrinsic base drive-in, the oxide-silicon interface was modeled as an insulating boundary in the computation. The areas of interest, such as the emitter, the intrinsic base, and the area under the sidewall, were designed to have the densest mesh. The minimum mesh spacing was approximately 5 nm in the vertical direction and 8 nm in the horizontal direction. Results obtained using SAFEPRO in previous calculations indicated that the mesh spacings used here are adequate to resolve the diffusions involved. The total number of nodes in the mesh is 1157. The nodes form a total of 2156 elements. Of these elements, 220 are used to model the extrinsic base polysilicon and 144 are used for the emitter polysilicon.

In preparation for the two-dimensional calculation, SAFEPRO was run on two essentially one-dimensional meshes having the same vertical spacings as those in Fig. 4 for both the emitter and the extrinsic base regions. The diffusion parameters were adjusted slightly to fit the SAFEPRO calculations to the SIMS experimental results. **Figure 5** shows the vertical profiles of both the SAFEPRO results and the SIMS measurements. The integrated base doping was also matched to the SIMS measurement of 6×10^{12} atoms per cm^2.

The two-dimensional calculations were then run. Three major process steps, namely the extrinsic base diffusion, the intrinsic base diffusion, and the emitter diffusion, were simulated. The lateral profiles of boron and arsenic along the silicon surface are shown in **Figure 6** for a typical sidewall thickness. The boron concentration decreases as one moves along the silicon surface away from the extrinsic base polysilicon until one meets the lateral diffusion of intrinsic base near the emitter window. The interaction of boron and arsenic results in a boron pileup near the intersection of these two species. This can be seen in both the vertical and

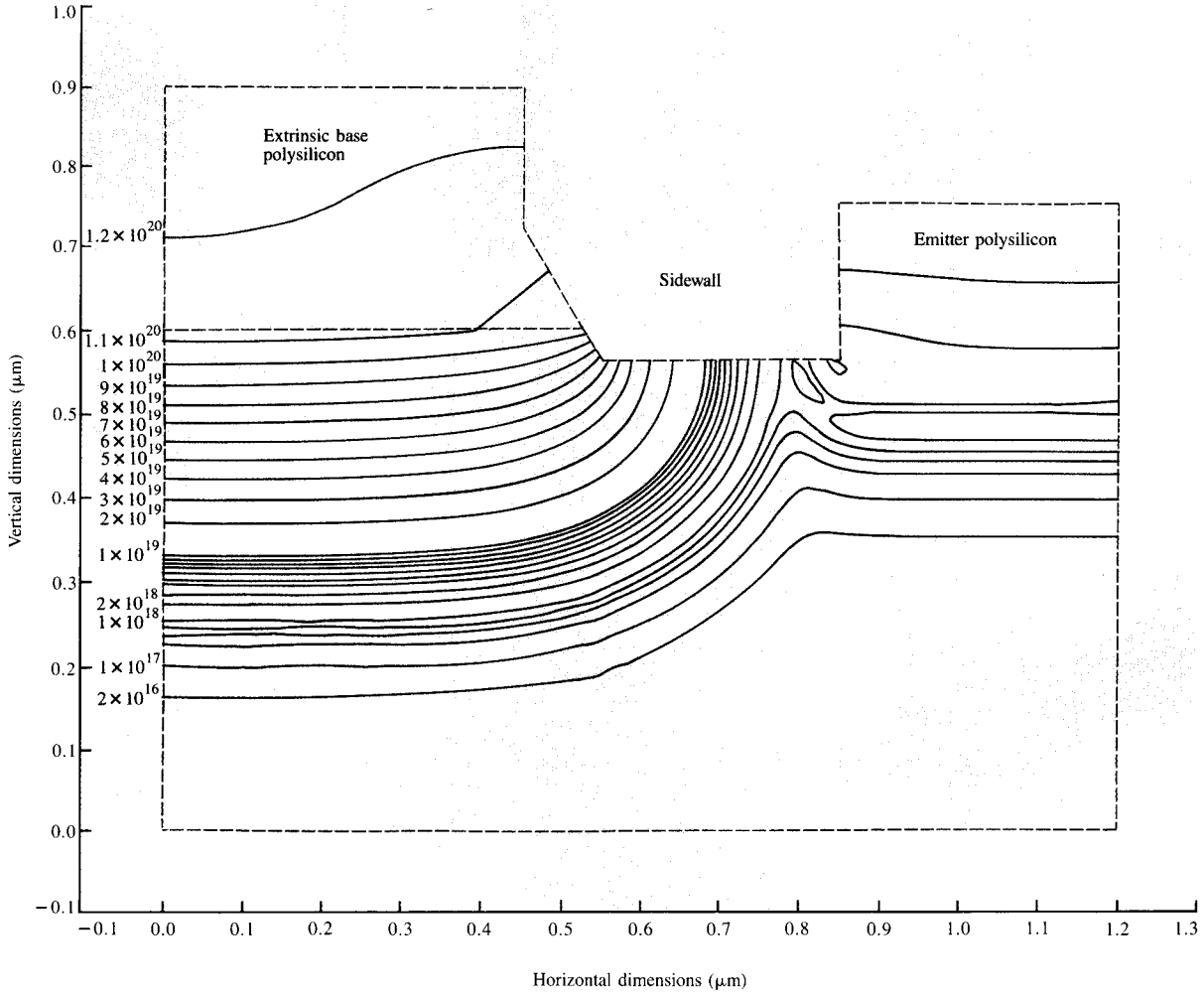

Figure 7

Contour plot of the boron distrubution over the entire mesh.

the lateral profiles. **Figure 7** shows a contour plot of boron for the entire mesh. It includes the boron from both the extrinsic base and the intrinsic base. **Figure 8** is an enlarged plot near the emitter window edge. The junctions can be located by plotting the net concentration. This plot shows the strong interaction between boron and arsenic near the base/emitter junction. The arsenic profile (as shown in Fig. 5) drops steeply from emitter polysilicon, and its contours (not shown for the sake of clarity in Fig. 8) extend only to about 50 nm in depth.

Since the questions of primary interest are the quality of the intersection between the extrinsic base and the intrinsic base, and the concentration at the lateral emitter-base junction, a number of two-dimensional runs were made for several sidewall thicknesses in order to determine how these items depend on the sidewall thickness. The results obtained are summarized in the following two figures. **Figure 9** shows the minimum concentration of boron at the silicon surface along the sidewall as a function of sidewall thickness. The flat portion of the curve at 10^{18} represents the boron concentration of the intrinsic base at the epitaxial surface in the emitter window. **Figure 10** shows the maximum surface concentration of boron at the emitter junction as a function of sidewall thickness. If the sidewall thickness is very great, the extrinsic base and the intrinsic base will have high resistance or be disconnected, and the base/emitter breakdown voltage and current gain will usually be very high, but punch-through will occur. (The variation of base resistance under the sidewall is discussed in the next paper of this issue.) If the sidewall thickness is too narrow, the extrinsic base extends well within the emitter window and the boron and arsenic intersect at very high doping

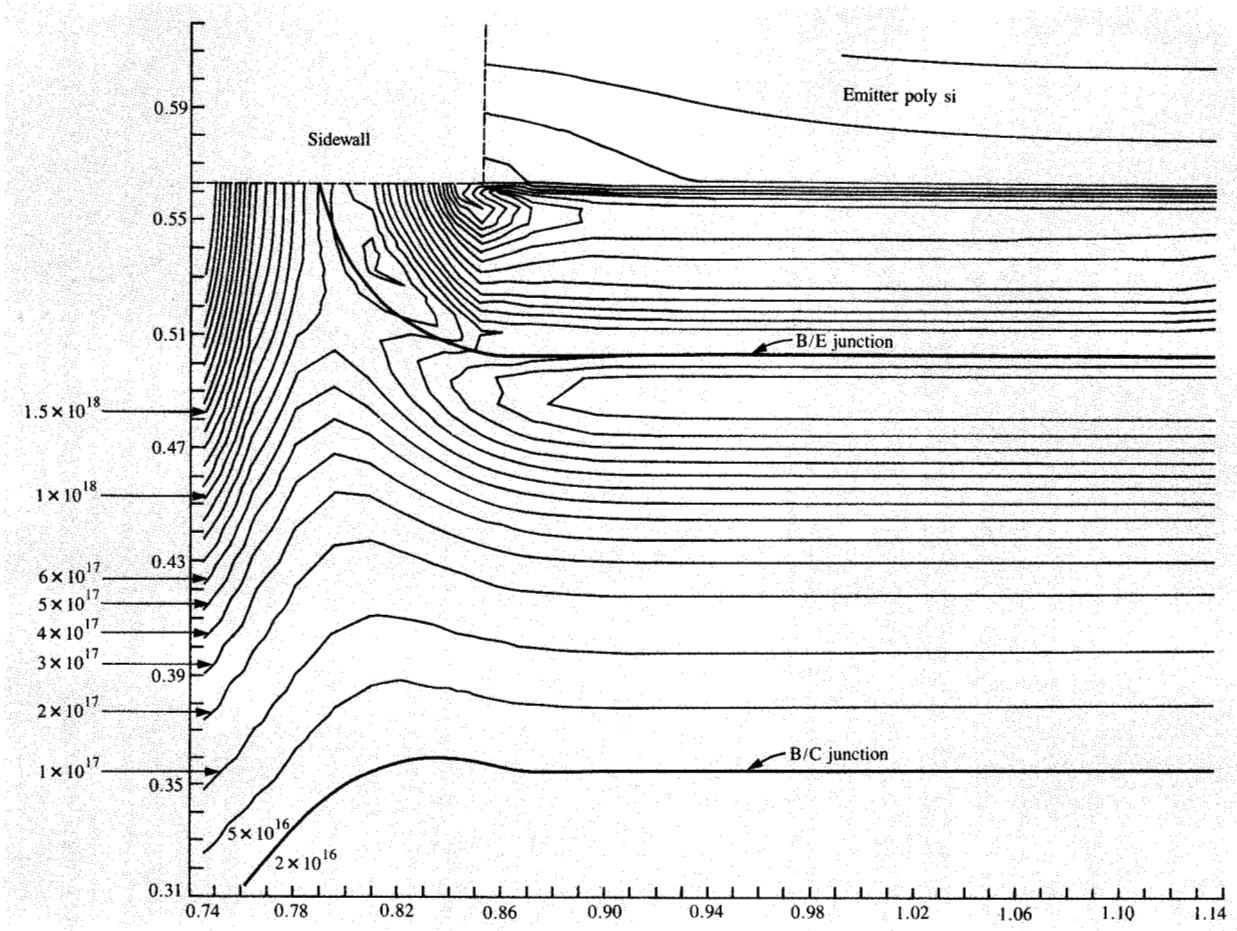

Figure 8

Enlarged portion of Fig. 7 near the emitter window edge. Scale is in micrometers.

concentrations. For the latter case, although the base contact resistance is very low, the effective base length decreases and the base/emitter junction tends to exhibit low breakdown voltage and high leakage current due to tunneling effect. Figures 9 and 10 allow the process designer to find the window of tolerance between these undesirable effects.

4. Extension of SAFEPRO to include oxide growth

When silicon dioxide is grown on silicon doped with boron, part of the boron in the silicon segregates into the oxide. This segregation can have a significant effect on the boron distribution in the silicon which remains unoxidized. This problem has been treated for one-dimensional cases by SUPREM [3]. The capability to treat the problem of oxide growth has been included in SAFEPRO. Because of the increased complexity of the problem in two dimensions, only very simple geometrical problems can be handled at present. We believe, however, that they include cases of great practical importance. A user may specify a region of the finite element mesh which is to be converted to oxide. This is done by assigning the elements to be turned to oxide a distinctive element property when the finite element mesh is generated. Since SAFEPRO must be able to locate the element boundaries which will serve as boundaries between the growing oxide and the remaining silicon, this region must at present have a simple shape.

SAFEPRO will scan this array of elements and locate the horizontal lines within it. The oxide silicon boundary will first be located at the top horizontal line (max y value) of the region. As the process simulation proceeds, the boundary is moved in steps down the rectangular region from one horizontal line to the next lower one. The velocity of motion of the boundary is normal to the boundary and by default equals the thickness of the region to be oxidized divided by the total process time for the oxidation step. This uniform

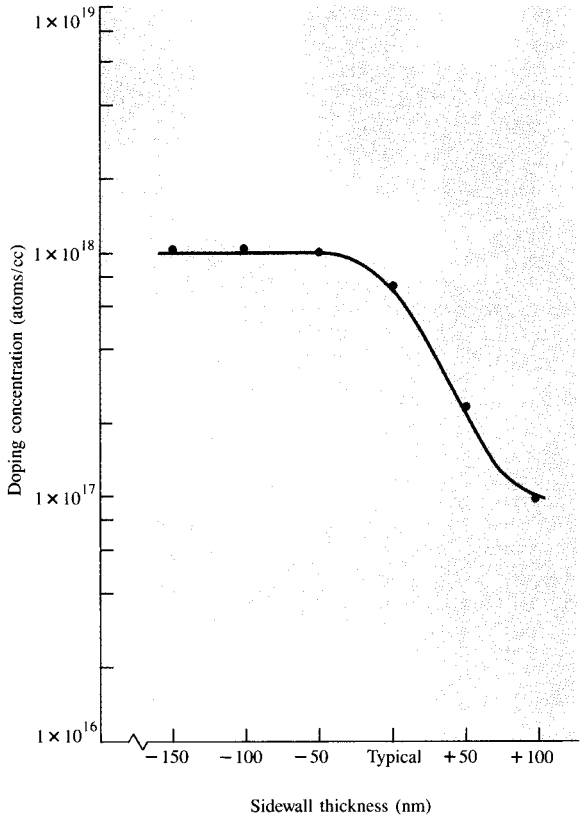

Figure 9

Minimum concentration of boron at the silicon surface along the sidewall as a function of sidewall thickness.

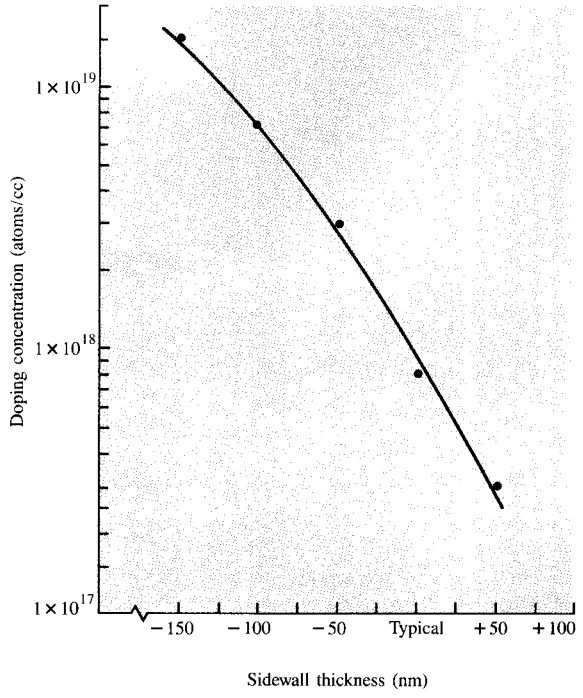

Figure 10

Maximum surface concentration of boron at the emitter junction as a function of sidewall thickness.

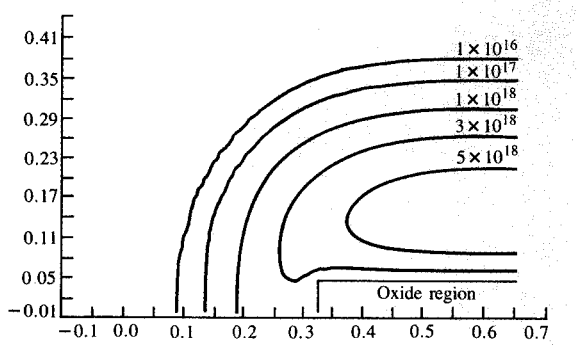

Figure 11

Boron diffusion with oxide growth. Scale is in micrometers.

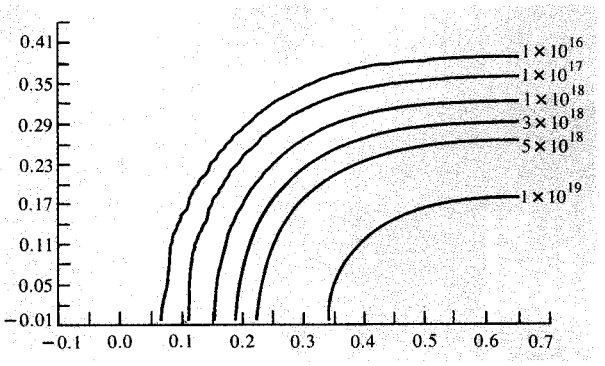

Figure 12

Boron diffusion without oxide growth. Scale is in micrometers.

velocity is appropriate for low-temperature oxidations in which the oxide growth is essentially uniform. A user may specify that the boundary motion follow the Grove-Deal Law by specifying the linear rate constant B/A and the parabolic rate constant B.

To illustrate the use of this oxide growth simulation, the two-dimensional boron profile in a silicon structure which had boron ions implanted and then driven in with the simultaneous growth of 100 nm of oxide is presented in **Figure 11**. For comparison a similar calculation without the

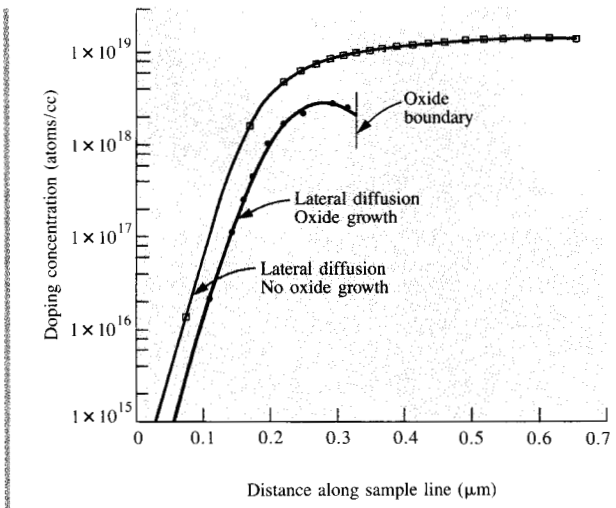

Figure 13

Lateral boron diffusion at the surface (oxide growth and no oxide growth).

oxide growth is shown in **Figure 12**. The oxidation-enhanced diffusion coefficient was used in both calculations to include the effect of the oxide in increasing the diffusion coefficient. The difference in the results reflects the additional effect of a growing oxide in depleting boron from the silicon. In the portions of both figures where the diffusion is one-dimensional, the results agree with those obtained from SUPREM II. In **Figure 13** the lateral boron distributions in the silicon at the surface are shown for both cases.

5. Conclusions

This paper described a 2D process simulator that uses the finite element method and includes many important effects previously available only in one-dimensional programs. These effects include the interaction between two different diffusing species, the presence of polysilicon or "barrier" regions, and the depletion of boron by a growing oxide. The paper also demonstrated the successful incorporation of dynamic time step selection in a diffusion program. This procedure, which is commonly used in circuit analysis programs, is now needed in diffusion programs, since in the problems of current interest, the maximum allowable time step varies widely throughout the course of the calculation. The results presented show the application of SAFEPRO to realistic problems of current interest in the design of advanced transistors.

Acknowledgments

To generate their finite element meshes and plot their results, the authors used a mesh generation and plotting program which was written at IBM in Burlington by E. M. Buturla, M. B. Lawlor, and C. M. McMullen.

References

1. D. P. Kennedy and R. R. O'Brien, "Analysis of the Impurity Atom Distribution Near the Diffusion Mask for a Planar p-n Junction," *IBM J. Res. Develop.* **9**, 179–186 (1965).
2. D. P. Kennedy and P. C. Murley, "Calculations of Impurity Atom Diffusion Through a Narrow Diffusion Mask Opening," *IBM J. Res. Develop.* **10**, 6–12 (1966).
3. D. A. Antoniadis and R. W. Dutton, "Models for Computer Fabrication of Complete IC Fabrication Process," *IEEE Trans. Electron Devices* **ED-26**, 490–500 (1979).
4. G. D. Hachtel, M. H. Mack, R. R. O'Brien, and B. Speelpenning, "Semiconductor Analysis Using Finite Elements—Part I: Computational Aspects," *IBM J. Res. Develop.* **25**, 232–245 (1981).
5. S. P. Gaur, P. A. Habitz, Y.-J. Park, R. K. Cook, Y.-S. Huang, and L. F. Wagner, "Two-Dimensional Device Simulation Program: 2DP," *IBM J. Res. Develop.* **29**, 242–251 (1985, this issue).
6. F.-Y. Chang and L. F. Wagner, "The Generation of Three-Dimensional Bipolar Transistor Models for Circuit Analysis," *IBM J. Res. Develop.* **29**, 252–262 (1985, this issue).
7. R. W. Knepper, S. P. Gaur, F.-Y. Chang, and G. R. Srinivasan, "Advanced Bipolar Transistor Modeling: Process and Device Simulation Tools for Today's Technology," *IBM J. Res. Develop.* **29**, 218–228 (1985, this issue).
8. J. Lindhard, M. Scharff, and H. E. Schiott, *Mat. Phys. Medd. Dan. Vid. Selek.* **33**, No. 14, 1–42 (1963).
9. S. Furukawa, H. Matsumura, and H. Ishiwara, "Theoretical Considerations on Lateral Spread of Implanted Ions," *Jpn. J. Appl. Phys.* **11**, 134–142 (1972).
10. S. M. Hu, "Formation of Stacking Faults and Enhanced Diffusion in the Oxidation of Silicon," *J. Appl. Phys.* **45**, 1567–1573 (1974).
11. D. P. Kennedy and P. C. Murley, "Concentration Dependent Diffusion of Arsenic in Silicon," *Proc. IEEE* **59**, 335–336 (1971).
12. R. O. Schwenker, E. S. Pan, and R. F. Lever, "Arsenic Clustering in Silicon," *J. Appl. Phys.* **42**, 3195–3200 (1971).
13. E. Guerrero, H. Pötzl, R. Tielert, M. Grasserbauer, and G. Stingeder, "Generalized Model for the Clustering of Arsenic Dopants in Silicon," *J. Electrochem Soc.* **129**, 1826–1831 (1982).
14. R. D. Richtymer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd Ed., Interscience Publishers, New York, 1967.
15. J. R. Bunch and D. J. Rose, *Sparse Matrix Computations*, Academic Press, Inc., New York, 1976.
16. Alan George and Joseph W. H. Liu, *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980.
17. *IBM System/360 and System/370, Subroutine Library–MATHematics, User's Guide*, Order No. SH12-5300, available through IBM branch offices.
18. D. A. Zein, C. W. Ho, and A. T. Gruodis, "A New Interactive Circuit Design Program in APL," *Proceedings of the 1980 IEEE International Symposium on Circuits and Systems*, Houston, TX, April 28–30, 1980, pp. 913–917.
19. G. P. Ho, J. D. Plummer, S. E. Hansen, and R. W. Dutton, "VSLI—Process Modeling—SUPREM III," *IEEE Trans. Electron Devices* **ED-27**, 1438–1453 (1983).
20. S. F. Chu, G. R. Srinivasan, H. Bhatia, B. M. Kemlage, F. Barson, J. Maurer, and J. Riseman, "A Self-Aligned Bipolar Transistor," *Proceedings of the First International Symposium on VLSI Science and Technology*, Detroit, MI, October 18–21, 1983.
21. J. Riseman, "Method for Forming an Insulator Between Layers of Conductive Material," U.S. Patent No. 4,234,362, November 18, 1980.

Received September 27, 1984

Karen W. Brannon *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Brannon is a staff engineer in the Computer Aided Device Design Department. She joined IBM at East Fishkill in 1983 and is currently working on modeling ion implantation processes using Monte Carlo techniques. Prior to joining IBM, Dr. Brannon was employed from 1980 to 1983 by JAYCOR in San Diego, California, where she worked on modeling of coal gasification processes and on problems in radiation effects. Dr. Brannon received a Ph.D. in chemical physics from the University of California–San Diego in 1980. She is a member of the American Chemical Society and the American Physical Society.

C. M. Hsieh *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Hsieh received the B.S. in electrical engineering from the National Taiwan University, Taipei, in 1958, the M.S. in electronics from the National Chiao Tung University in 1960, and the M.S. and Ph.D., both in electrical engineering, from the Polytechnic Institute of Brooklyn, New York, in 1964 and 1966, respectively. From 1962 to 1966, he was a research Fellow at the Polytechnic Institute of Brooklyn. Since joining IBM in 1966, Dr. Hsieh has worked in the areas of logic and memory circuit diagnostics. He is currently a senior engineer in the Computer Aided Device Design Department at the East Fishkill laboratory, working on the process simulation of advanced bipolar transistors.

Ronald W. Knepper *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Knepper is a senior engineering manager for the exploratory devices project in the semiconductor laboratory at East Fishkill. He received a B.A. in physics at Juniata College, Huntingdon, Pennsylvania, and a B.S., an M.S., and a Ph.D., all in electrical engineering, at Carnegie-Mellon University, Pittsburgh, Pennsylvania. He joined IBM in 1969 at East Fishkill in the advanced FET circuit design area. He worked on both bipolar and FET circuit design and modeling projects before managing an advanced bipolar array design department and later a process model development team. Currently, he is interested in applying computer-aided device design and modeling techniques to high-performance bipolar technology development. Dr. Knepper is a member of the Institute of Electrical and Electronics Engineers, Sigma Xi, and Tau Beta Pi. He has received an IBM Outstanding Technical Achievement Award for his work on bipolar array design and has also received three IBM Invention Achievement Awards.

R. F. Lever *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Mr. Lever is an advisory engineer in the Computer Aided Device Design Department at East Fishkill. He received his M.A. in physics from Oxford University, England, in 1951, joining IBM in 1960. His work has been in the field of materials science, including chemical vapor deposition, Rutherford backscattering analysis, and semiconductor processing fundamentals. His current interest is in developing semiconductor process simulation models which do not violate the laws of physics, while using the results of such modeling to further our understanding of processing fundamentals.

J. Scott Moore *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Moore is a staff engineer in the Computer Aided Device Design Department. He joined IBM in 1982. He is currently involved in modeling of semiconductor processes, especially ion implantation. Dr. Moore received a B.S. in physics in 1979 and an M.S. and a Ph.D. in materials engineering in 1977 and 1981, respectively, all from Rensselaer Polytechnic Institute in Troy, New York. His thesis work included integrated optics and hot electron transport.

P. C. Murley *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Mr. Murley is an advisory mathematician in the Computer Aided Device Department. He received the B.A. degree in mathematics from Northwestern University, Evanston, Illinois, and the M.S. degree in mathematics from the University of Illinois, Urbana, where he worked on cylindrical shell buckling problems, using the Illiac-II computer. In 1957 he joined IBM at the product development laboratory in Poughkeepsie, where he helped develop magnetic character recognition logic and software for large-scale linearized network analysis. Since 1961, he has concentrated on mathematical analysis of semiconductor devices and processing.

Redmond R. O'Brien *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. O'Brien received the B.S., M.S., and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology, Cambridge. From 1958 to 1960, he worked for Sylvania on communication and detection problems. In 1960, he joined IBM to work on the theoretical analysis of semiconductor devices. At present, he is an advisory mathematician in the Computer Aided Device Design Department. Dr. O'Brien is a member of the American Mathematical Association, the American Statistical Association, the Institute of Electrical and Electronics Engineers, and the Society for Industrial and Applied Mathematics.

G. R. Srinivasan *IBM General Technology Division, East Fishkill facility, Route 52, Hopewell Junction, New York 12533.* Dr. Srinivasan received B.Sc. and B.Sc. (with honors) degrees in physics from the University of Mysore, India, a Diploma in metallurgy from the Indian Institute of Science, an M.S. degree in metallurgy from the Colorado School of Mines, and a Ph.D. degree in physical metallurgy from the University of Illinois. He was on the research faculty at Cornell University, where he worked on the crystallography of phase transformations in metals and alloys. He was Associate Professor of Materials Science at the Catholic University of America, Washington, DC, where he established and directed the electron microscope and X-ray diffraction laboratory for the study of glasses and ceramics. Since 1974, he has been with IBM at the East Fishkill laboratory, where he is presently the manager of the Computer Aided Device Design Department. He has published over 80 papers in the areas of crystallography of phase transformations, composition fluctuation dynamics, spinodal decomposition, chalcogenide glass ceramics, silicon epitaxy, device physics, process modeling, and process design, and has obtained several patents in device and process design. Dr. Srinivasan has served on many committees in the Metallurgical Society and the Electrochemical Society and is listed in *American Men and Women of Science.*