*SIMULATION OF SEMICONDUCTOR DEVICES AND PROCESSES Vol. 3*
*Edited by G. Baccarani, M. Rudan - Bologna (Italy) September 26-28, 1988 - Tecnoprint*

# ON EFFICIENCY OF MULTIGRID METHODS IN TWO-DIMENSIONAL IMPURITY REDISTRIBUTION SIMULATION

S. Mijalković, D. Pantić, N. Stojadinović

Faculty of Electronic Engineering, University of Niš
Niš, Yugoslavia

SUMMARY

A comprehensive study on efficency of multigrid methods in two-dimensional impurity redistribution simulation is presented. Impurity diffusion equation in non-conformally transformed rectangular simulation domain is considered as a mathematical model for impurity redistribution process. The theoretical part of the study is based on the smoothing efficiency factor predicted by local mode analysis. On the other hand two practical impurity redistribution examples from VLSI technology are used for numerical experiments with an actual multigrid program to predict global efficiency through convergence efficiency factor and multigrid gain factor.

## INTRODUCTION

Increasingly it is recognized that accurate simulation of impurity redistribution represents, as far as computer resources are considered, the most severe bottlneck in existing two-dimensional process simulation programs (Yeager, 1985). As a result of the importance of the continuing advance in physical models for impurity and defect diffusion processes an increasing amount of attention is being devoted to much more efficient and parallelizable numerical tools for the next generation of process simulation programs.

In spite of the fact that multigrid methods are fully parallelizable and presently represent the most efficient solvers of eliptic boundary value problems, efficiency and guiding principles for application of multigrid methods to nonlinear evolution problems like those appearing in impurity redistribution simulation have not yet been studied systematically.

The two major questions with which one who attempts to apply multigrid methods to solution of two-dimensional impurity diffusion equation is confronted are: (1) what efficiency is really

obtainable by multigrid methods and (2) which multigrid components give this efficiency. The purpose of this work is to give clear and as quantitative answers as possible to these questions.

Quantitative analysis of efficiency of multigrid method in impurity redistribution simulation has been performed in two steps. The first step is local analysis of efficiency which is based on the worst case local mode analysis and smoothing efficiency factor. The second step is global analysis of efficiency based on numerical experiments with an actual multigrid program. The global efficiency of the simulation of two practical impurity redistribution processes typical for fabrication of VLSI NMOS transistors is estimated using the convergence efficiency factor and the multigrid gain factor.

## PROBLEM DEFINITION

The redistribution of an arbitrary impurity A in a semiconductor is usually described by the nonlinear diffusion equation

$$\frac{\partial C_A}{\partial t} - div(D_A \cdot grad\ C_A) = 0 \qquad (1)$$

where $C_A$ is the impurity concentration and $D_A$ is the diffusion coefficient which is a complex function of temperature and of concentrations of all impurities present in the semiconductor as well as point defects concentration. For the sake of simplicity we have analysed here only the decoupled impurity redistribution with concentration dependent diffusion coefficient $D_A = D_A(C_A)$. We have assumed sophisticated phenomenological model for $D_A$ which in our numerical experiments includes both vacancy- and interstitial-assisted diffusion mechanisms as well as the field enhanced diffusion (Mijalković, 1988).

By ignoring diffusion in oxide, the boundary condition which accounts for impurity redistribution at the curved and moving $Si-SiO_2$ boundary is (Seidl, 1983a):

$$D_A \cdot grad\ C_A \cdot \hat{n} = K_A\ C_A\ \vec{v} \cdot \hat{n} \qquad (2)$$

where $K_A$ is the temperature dependent term which describes impurity segregation at the interface, $\vec{v}$ is the velocity of the moving interface and $\hat{n}$ is the unit normal to boundary. Zero flux boundary condition is used at all other boundaries of simulation domain. It should be noted that problems posed by curved and moving boundaries are much more emphasized in multigrid then in single-grid methods. Namely, curved and moving boundaries should also be properly dealt with on very coarse grid levels.

If we do not want to deprive ourselves of the simplicity of uniform finite-difference grids, which are desirable in the multigrid contex, there are two possible approaches already used to overcome the problem of the curved and moving boundaries. First, less common approach is to use a special local discretization at the boundary (Joppich, 1987). However, stability of such special discretization sometimes can be critical. The second approach which is much safer in view of discretization

instabilities at the curved and moving boundaries is to bypass the geometry problems using some coordinate transformation.

Let us consider a coordinate transformation from an arbitrary time-dependent physical domain $P$ in $(u,v)$ plane onto rectangular simulation domain $Q$ defined in $(x,y)$ plane. The diffusion equation (1) in transformed domain is

$$\frac{\partial C_A}{\partial t}+\frac{\partial x}{\partial t}\frac{\partial C_A}{\partial x}+\frac{\partial y}{\partial t}\frac{\partial C_A}{\partial y} - Ł_{xx}C_A(x_u^2+y_v^2)-Ł_{yy}C_A(y_u^2+y_v^2)- \tag{3}$$

$$- (Ł_{xy}C_A+ Ł_{yx}C_A)(x_u y_u+x_v y_v)- Ł_x C_A \Delta x- Ł_y C_A \Delta y = 0$$

where $Ł_{pq}C=\frac{\partial}{\partial p}(D(C)\frac{\partial C}{\partial q})$ and $Ł_p C = D(C)\frac{\partial C}{\partial p}$ are nonlinear operators, $D$ is the diffusion coefficient and $C$ is the impurity concentration. The transformation is carried out by finding the mapping functions $u(x,y,t)$ and $v(x,y,t)$ or their inverse functions.

In process simulation it is sometimes possible to define the $Si\text{-}SiO_2$ interface by expression $v = V(u,t)$. Then the simple mapping functions (Penumalli, 1983; Seidl, 1983b):

$$\begin{aligned} x &= u \\ y &= v - V(u,t) \end{aligned} \tag{4}$$

give a closed form coordinate transformation which is very often used in various single-grid and multigrid programs.

More sophisticated coordinate transformations give simpler forms for the diffusion equation than (4) but closed form expressions for the mapping functions are not always available. For example, in conformal transformation the mapping functions should satisfy Cauchi-Rieman equations

$$\begin{aligned} u_x &= v_y \\ u_y &= -v_x \end{aligned} \tag{5}$$

It should be noted that our decision to analyze the efficiency of multigrid methods using transformed diffusion equation (3) with non-conformal mapping functions (4) is not at the expense of generality since this transformation contains almost all stalling processes present in other coordinate transformations.

The second-order spatial derivatives of (3) with (4) are discretized by 9-point central differences. It is well known that the usage of central differences for first-order convection terms could cause unstable discretization which is much more pronaunced in the multigrid context because of coarser grid levels. We ensured stability of discretization by adding the numerical viscosity in discretization using the one-sided (upwind) discretization for convection terms of (3). In order to avoid an unwanted influence of the boundary relaxation on the smoothness of interior it is advizable to have the same consistency order of discretization inside the domain and at the boundary. In our opinion, the most convenient way tosatisfy this requirement is to use "mirror imaging" method (Selberherr, 1984).

## LOCAL ANALYSIS OF EFFICIENCY

The most crucial and problem-dependent stage in the development of fast multigrid solvers is design of interior relaxation scheme with efficient smoothing property. Relaxation in multigrid algorithms could be regarded as a local process with a local task: to reduce high-frequency error components (smooth error components are easy to remove on the coarser levels). This is why the efficiency of relaxations and consequently of the hole multigrid algorithm can accurately be measured by local mode analysis introduced by Brandt (1977).

The local mode analysis is based on a simplified Fourier analysis which can be applied to difference schemes with constant (frozen) coefficients in unbounded domain. Therefore, preparation of the diffusion equation for local mode analysis has two steps: (1) linearization and (2) freezing of variable coefficients.

For the purpose of linearization the so-called "principal linearization" is a good as full linearization in multigrid relaxation. Principally linearization term which correspond to the nonlinear term $L_{pq}C$ in (3) is

$$L_{pq}C = \frac{\partial}{\partial p}\left(D(\tilde{C})\cdot\frac{\partial C}{\partial q}\right) \tag{6}$$

which is obtained just by using the diffusion coefficient $D(\tilde{C})$ from the previous iteration. Expanding (6) and replacing the function $C$ in the diffusion coefficient by the constant $\bar{C}$ the "frozen" operator is obtained:

$$\bar{L}_{pq}C = \frac{\partial D(\bar{C})}{\partial p}\frac{\partial C}{\partial q} + D(\bar{C})\frac{\partial^2 C}{\partial p \partial q} \tag{7}$$

Accordingly, the discrete "frozen" operator of the diffusion equation (3) with mapping functions (4) is

$$\begin{aligned}
&\bar{L}C = T_{xx}C + a_1 T_{yy}C + a_2 h^{-1} T_x C + a_3 h^{-1} T_y C + a_4 T_{xy}C + a_5 h^{-2} C \\
&a_1 = 1 + V'^2 \\
&a_2 = hD(\bar{C})^{-1}\frac{\partial D(\bar{C})}{\partial x} - V' hD(\bar{C})^{-1}\frac{\partial D(\bar{C})}{\partial y} \\
&a_3 = hD(\bar{C})^{-1}\dot{V} - hV'' + (1+V'^2)hD(\bar{C})^{-1}\frac{\partial D(\bar{C})}{\partial y} - V' hD(\bar{C})^{-1}\frac{\partial D(\bar{C})}{\partial x} \\
&a_4 = -2V' \\
&a_5 = -h^2 / (D(\bar{C})\,\Delta t)
\end{aligned} \tag{8}$$

where $T_p$ and $T_{pq}$ are the discretized forms of differential operators $\partial C/\partial p$ and $\partial^2 C/\partial p \partial q$, respectively. $h$ is the grid-step size and $\Delta t$ is the time-step size. The typical maximum values of the most important normalized "frozen" physical parameters in

Table 1. Typical worst case values of "frozen" parameters

| $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ |
|---|---|---|---|---|
| $\lvert V' \rvert$ | $\lvert hV'' \rvert$ | $h\dot{V}D(\bar{C})^{-1}$ | $hD(\bar{C})^{-1}\left\lvert \frac{\partial D(\bar{C})}{\partial y} \right\rvert$ | $-h^2/(D(\bar{C})\Delta t)$ |
| 2 | 2 | >5 | 1 | 0 |

(8) are given in Table 1. Note from Table 1 that the "frozen" diffusion equation (8) can be localy regarded as a singular perturbation problem because of large anisotropy and large coefficients in front of the convection terms.

As a quantitative measure of local efficiency we have used the smoothing efficiency factor (Brandt, 1977):

$$E_s = w_0^{-1} \log(1/\mathring{\mu}) \tag{9}$$

where $w_0$ is the computational work per grid point and relaxation sweep. $\mathring{\mu}$ is the multigrid convergence factor which is in a two-dimensional case defined as

$$\mathring{\mu} = \bar{\mu}^{3/4} \tag{10}$$

where $\bar{\mu}$ is the smoothing factor. The smoothing factor is the worst amplification factor in the range of high-frequency error components. As a local computational work unit we have chosen 1000 floating point operations.

The smoothing efficiency factor is a very useful quantitative tool for comparation of different relaxation schemes. For example let us consider the two relaxation schemes which are potentionally for transformed diffusion equation: pointwise Gauss-Seidel with lexicogtaphical ordering of points (GS-lex+) and y-line relaxation with lexicographical ordering of lines. Zero computational work for computation of diffusion coefficient should be assumed in the local analysis of efficiency since the worst case comparation of smoothing efficiency factors for these two relaxations implies the largest difference in computational works for each of the relaxations.

Fig. 1 shows the dependence of the GS-lex+ and y-line smoothing efficiency factors on parameter $p_1$ assuming $p_2=p_3=p_4=p_5=0$. Smaller smoothing efficiency factor of the GS-lex+ for large values of parameter $p_1$ is caused by anisotropy when the GS-lex+ relaxation smoothes only with respect to y-direction. However, in spite of the anisotropy, the GS-lex+ relaxation has larger smoothing efficiency factor up to the certain value of $p_1$. Beyond this point the y-line relaxation dominates because it compensates for the loss of ellipticity. It should be noted that the anisotropy problem can be avoided by using conformal mapping functions (5).

Fig. 2 shows the influence of parameter $p_3$ on the smoothing efficiency factor assuming $p_1=p_2=p_4=p_5=0$ for three different types of relaxation: the GS-lex+ relaxation, Gauss-Seidel relax-

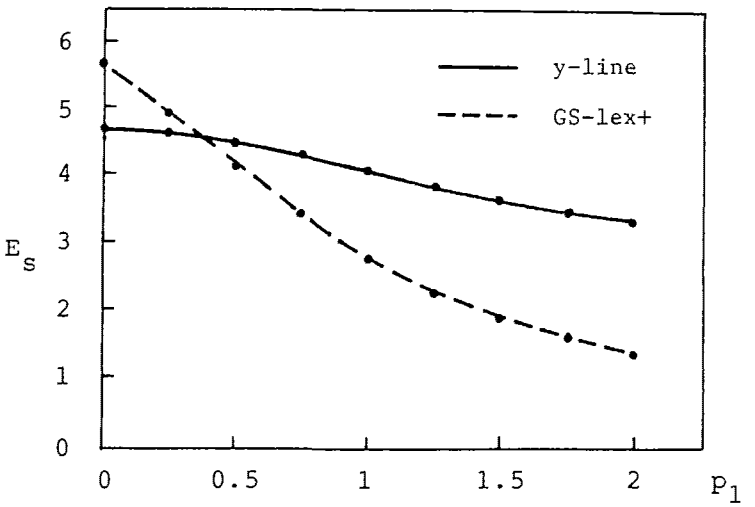


Fig. 1. Dependence of the smoothing efficiency factors on parameter $p_1$.

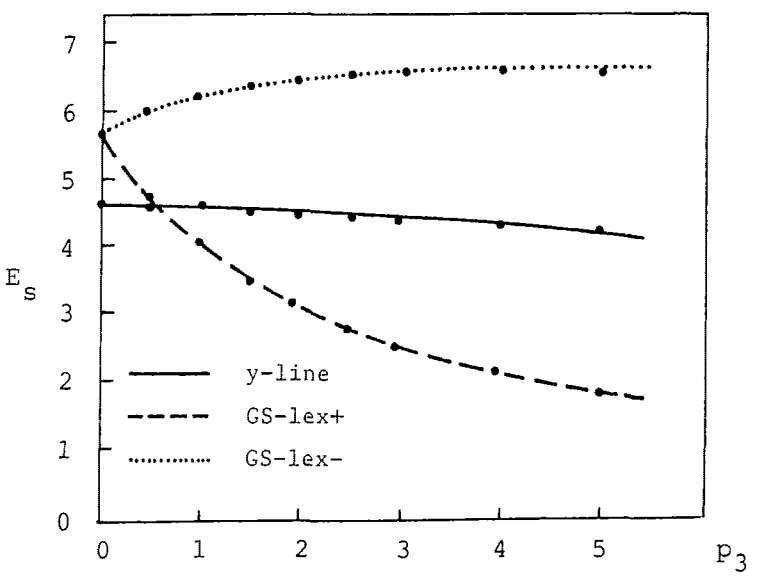


Fig. 2. Influence of parameter $p_3$ on the smoothing efficiency factors.

ation with inverted lexicographical ordering of points (GS-lex-) and y-line relaxation. It is obvious from fig. 2 that the GS-lex- relaxation has the best smoothing efficiency factor in the hole range of parameter $p_3$ values. This is not difficult to understand since this type of relaxation tends to reach the smoothing factor of y-line relaxation i.e. to become an exact solver inside a single line of grid points when $p_3 \to \infty$ while, on the other hand it needs less computational work then the y-line relaxation. Note that the GS-lex- relaxation has the same smoothing efficiency factor in the case when only parameter $p_1$ is nonzero. When the GS-lex+ and the y-line relaxations are considered their qualitative relationship is similar to that shown in fig. 1. The GS-lex+ relaxation dominates over the y-line relaxation for small values of parameter $p_3$ because of the smaller computational work. For larger values of $p_3$ the GS-lex+ relaxation becomes powerless and only shifts the high-frequency error over the grid without reducing it.

Fig. 3 shows the dependence of smoothing efficiency factor of GS-lex+ relaxation on parameter $p_5$ for two worst case values of parameters $p_1$ and $p_3$. It is obvious that the presence of parameter $p_5$, which originates from the implicit Euler discretization in time, improves the smoothing property of the GS-lex+ relaxation as well as of all other types of relaxation. For $p_5 \to \infty$ each relaxation tends to become an exact solver. Accordingly, in the contex of multigrid methods, evolution problems are easier to solve than an equivalent stationary problem.

It can be concluded that the only relaxation working as an efficient smoother in the wide range of parameters $p_1$ and $p_3$ is the y-line relaxation. This fact recommendts the y-line relaxation for application in the general impurity redistribution simulator.

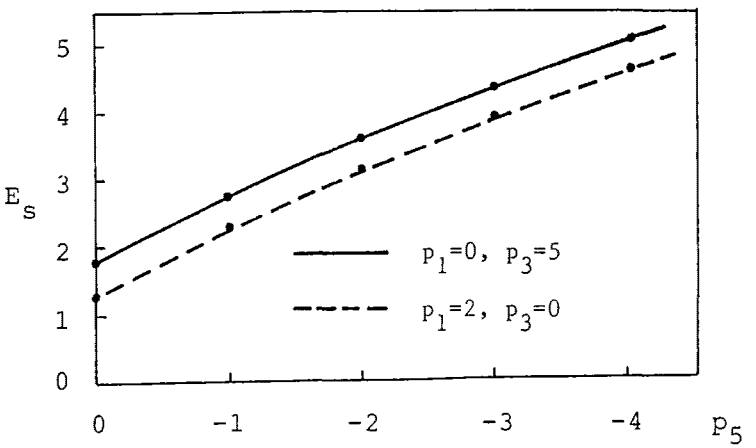


Fig. 3. Dependence of the smoothing efficiency factor on parameter $p_5$.

## GLOBAL ANALYSIS OF EFFICIENCY

We consider here a really obtainable efficiency of an actual multigrid program in simulation of impurity redistribution processes under practical processing conditions.

The two impurity redistribution processes typical for fabrication of VLSI NMOS transistors are chosen as practical examples for numerical experiments. The first process is the high-concentration arsenic redistribution for the source/drain formation The second process is the low-concentration boron distribution during the field oxidation. Distribution of the initial as-implanted arsenic and boron profiles as well as the thermal oxidation process which determines the shape of the silicon surface are both modeled using analytical expresions. Process parameters used in numerical experiments are given in table 2. All components of the actual multigrid program are selected so as to ensure approaching to the ideally obtainable efficiency and to make the program as problem independent as possible (Mijalković, 1987):

(1) *sequence of grid*. The grids are uniform, non-staggered and rectangular. The grid size for the level l are $h_l=d^{-l+1}h_1$ $1<l<M$.
(2) *multigrid version*. We sugest a widely used version of non-linear multigrid algorithm - Full Approximation Scheme (FAS) and V-shaped cycles with 2 pre- and 1 post-relaxation.
(3) *relaxation*. It is possible to use the GS-lex+ relaxation, Gauss-Seidel relaxation with red-black ordering of points (GS-rb) and y-line relaxation.
(4) *restriction*. The nine point restriction (full weighting).
(5) *prolongation*. The nine point prolongation (bilinear interpolation).
(6) *solution of the coarsest level*. We apply 5 y-line relaxations for the solution on the level l=1.
(7) *time-step size selection*. An automatic time-step selection based on the Milne's device is used.

The starting simulation parameters are given in table 3.

As a quantitative measure of global efficiency obtained in numerical experiments we have considered two global efficiency factors. The first is convergence efficiency factor

Table 2. Process parameters in numerical experiments

| process | parameter | arsenic | boron |
|---|---|---|---|
| implantation | energy | 100 keV | 160 keV |
| | dose | $6\cdot10^{15}$ cm$^{-2}$ | $5.5\cdot10^{13}$ cm$^{-2}$ |
| redistribution | temperature | 1000°C | 1000°C |
| | time | 30 min. | 180 min. |
| | ambient | inert | $H_2O$ |

Table 3. The starting simulation parameters

| parameter | arsenic | boron |
|---|---|---|
| the coarsest grid size | 3x3 | 4x3 |
| number of grid levels | 5 | 5 |
| grid step size | $1.35 \cdot 10^{-6}$cm | $5.46 \cdot 10^{-6}$ cm |
| initial time-step size | 15s | 60s |

(11) $$E_c = w_g^{-1} \log(1/\eta)$$

where $w_g$ is the amount of computational work given in global computational work units and $\eta$ is the empirical convergence rate. The empirical convergence rate

(12) $$\eta = |r(w_g)| / |r_o|$$

is the residual error norm reduction on the finest grid level after multigriding which spends computational work $w_g$. For the global work unit we have assumed computational work needed for 10 relaxation sweeps over the finest grid level.

However, the convergence efficiency parameter $E_c$ is usualy not sufficient to estimate the real benefit from using multigrid algorithm. Namely, carefully chosen relaxation with high-error smoothing rate is effective only if low-frequency components of error is really present in a problem. Therefore, we have introduced another supplementing efficiency factor called multigrid gain factor:

(13) $$E_g = E_c^{mg} / E_c^{sg} .$$

where $E_c^{mg}$ and $E_c^{sg}$ are the convergence efficiency factors for a multigrid method and an equivalent single-grid method.

Table 4 shows average convergence efficiency factors for arsenic and boron redistribution simulation using three diferent relaxation types. It can be noted that average convergence efficiency factor does not strongly depend on the choice of the

Table 4. Average convergence efficiency factors

| | boron | arsenic |
|---|---|---|
| GS-lex+ | 2.95 | 2.34 |
| GS-rb | 3.41 | 3.68 |
| y-line | 3.58 | 3.00 |

relaxation type and simulation example. As was predicted by local mode analysis, the y-line relaxation is the best in the case of boron simulation since oxidation process influences a large part of the simulation domain. However, in the case of arsenic redistribution example in inert ambient the GS-rb is advantageous. Apart from having the best average convergence efficiency factor, the GS-rb relaxation is very suitable for full parallelization of a multigrid algorithm.

Fig. 4 shows distributions of the time-step size and convergence efficiency factor during arsenic redistribution simulation. Once more, one can notice the robustness of a multigrid method because the convergence efficiency factor for each of the relaxations remains almost unaffected by variations of the time-step size chosen so as to ensure the same local truncation error in time throughout the entire simulation. It is also interesting to note that the GS-rb relaxation, apart from being the most efficient comparing to the two other relaxation types, is also the least sensitive to variations of the time-step size.

Finally, fig. 5 shows dependence of multigrid gain factor on the time-step size and grid-step size for the first time-step in arsenic redistribution simulation. We expected to obtain relation $E \sim \Delta t/h^2$ but it can be seen from fig. 5 that in practical simulation examples this relation is more complicated. It is obvious that the time-step increase and/or grid-step size decrease enlarge the amount of low-frequency error components and which in turn increases the effectivness of multigrid algorithm which is measured by multigrid gain factor.

## CONCLUSION

In this paper we have presented the study of efficiency of multigrid methods in two-dimensional impurity redistribution problems. We have tried to use the most quantitative measures for local and global efficiency as possible by using the smoothing efficiency factor, convergence efficiency factor and multigrid gain factor. This efficiency factors allow comparation of different relaxation processes which are the most responsible for efficient work of multigrid algorithms.

As the most important conclusions of this efficiency analyse we select the following:

(1) Multigrid program based on Full Approximation Scheme (FAS) version of nonlinear multigrid, "principal linearization" for relaxation operator, bilinear interpolation for prolongation and full weighting for restriction gives the optimum convergence efficiency of impurity redistribution simulation in a non-conformally time-independent rectangular simulation domain.

(2) Smoothing efficiency factor and convergence efficiency factor almost do not depend on the grid- and time-step size variations as well as variations of physical parameters of impurity redistribution problems.

(3) The block relaxations (y-line in our case) are the best choice for the general impurity redistribution simulation pro-

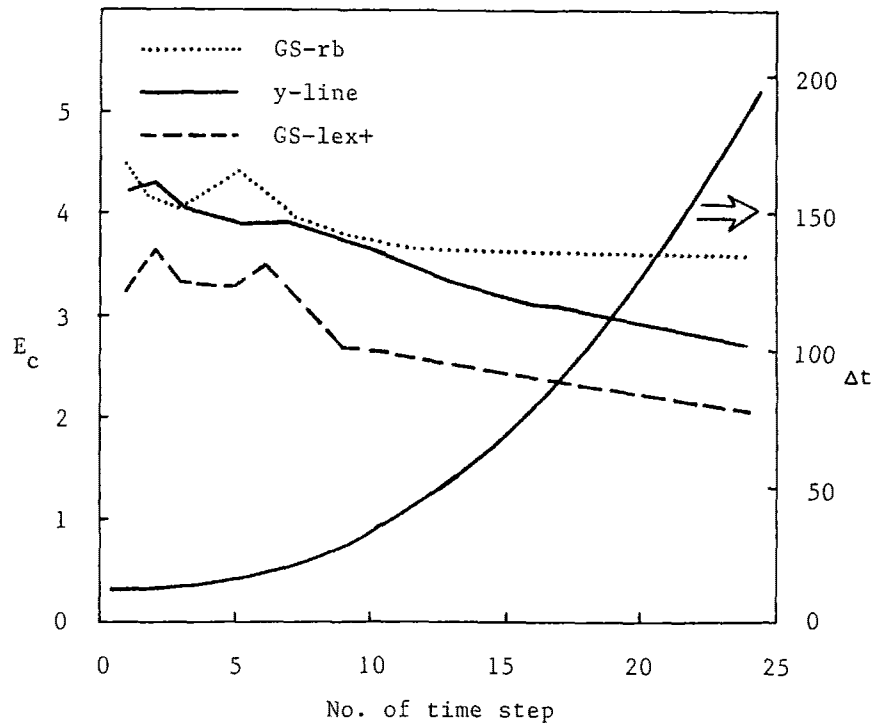


Fig. 4. Distribution of the time-step size and the convergence efficiency factor during arsenic redistribution simulation.

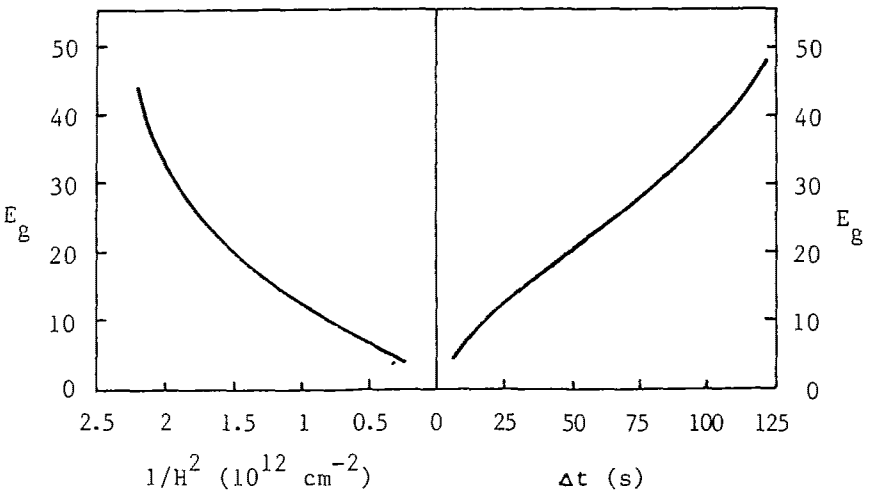


Fig. 5. Dependence of the multigrid gain factor on time-step size and grid-step size.

grams and their efficiency and dominates especially when local oxidation process is present. On the other hand, the GS-rb relaxation is the best for high-concentration inert impurity redistribution simulations and can be easilly parallelized. (4) Multigrid gain factor strongly depends on the maximum allowable local truncation error.

With respect to the high convergence efficiency factors and multigrid gain factors obtainable by FAS multigrid algorithm and having in mind possibilities for further extensions of FAS algorithm so as to include local grid refinement it can be concluded that multigrid methods are very auspicious for the next generation of process simulation programs.

## REFERENCES


Brandt A. (1977). Multi-Level Adaptive Solution to Boundary-Value Problems. *Mathematics of Computation, 31,* 333-390.

Joppich W. (1987). A Multigrid Method for Solving the Nonlinear Diffusion Equation on a Time-Dependent Domain Using Rectangular Grids in Cartesian Coordinates. In J. J. Miller (Ed.), *Proceedings of NASECODE V Conference,* Trinity College, Dublin, pp. 243-248.

Mijalković S. and N. Stojadinović. (1987). Multigrid Method: An Efficient Numerical Tool in VLSI Process Modeling. In W. E. Proebster and H. Reiner (Ed.), *Proceedings of COMPEURO 87 Conference,* Vol. 1, Hamburg, pp. 508-509.

Mijalković S. and N. Stojadinović. (1988). Efficent Simulation of Impurity Redistribution in VLSI Fabrication Processes. Accepted for publication in *Solid-State Electronics.*

Penumalli B. (1983). A Comprehensive Two-Dimensional VLSI Process Simulation Program BICEPS. *IEEE Transaction on Electron Device, ED-30,* 986-992.

Seidl A. (1983a). A Mathematical Implementation of Segregation Model for Two-Dimensional Process Simulation. *IEEE Transaction on Electron Device, ED-30,* 722-723.

Seidl A. (1983b). A Multigrid Method for Solution of the Diffusion Equation in VLSI Process Modeling. *IEEE Transaction on Electron Device, ED-30,* 999-1004.

Selberherr S. (1984). The Discretization of the Basic Semiconductor Equations. *Analysis and Simulation of Semiconductor Devices,* Springer-Verlag, Wien, pp. 149-201.

Yeager H. and R. Dutton (1985). An Approach to Solving Multiparticle Diffusion Exhibiting Nonlinear Stiff Coupling. *IEEE Transaction on Electron Device, ED-32,* 1964-1975.