ON THE TOPOGRAPHY SIMULATION OF MEMORY CELL TRENCHES FOR SEMICONDUCTOR MANUFACTURING DEPOSITION PROCESSES USING THE LEVEL SET METHOD

Clemens Heitzinger and Siegfried Selberherr
Institute for Microelectronics
TU Wien
Gußhausstraße 27–29,
A-1040 Vienna, Austria
Email: heitzinger@iue.tuwien.ac.at

KEYWORDS
Computer aided design, semiconductor processes, chemical vapor deposition, surface evolution, level set method.

ABSTRACT

Etching and deposition of Silicon trenches is an important semiconductor manufacturing step for state of the art memory cells and other semiconductor devices like, e.g., Power MOSFETs. Understanding and simulating the transport of gas species and surface evolution enables to achieve voidless filling of deep trenches, to predict the resulting profiles, and thus to optimize the process parameters with respect to manufacturing throughput and the resulting memory cells.

An accurate and fast method for surface evolution has been combined with the simulation of the transport of gas species above the wafer surface and applied to a SiO_2 deposition process. In experiments a SiO_2 layer was deposited into trenches roughly $4\mu\mathrm{m}$ deep and $2\mu\mathrm{m}$ wide, where the final layer thickness was in the range of $1\mu\mathrm{m}$ for the flat wafer surface. Simulation results are discussed and compared to scanning electron microscope pictures, where all effects were reproduced in the simulation and good quantitative agreement was achieved as well.

INTRODUCTION

When simulating etching and deposition processes for semiconductor manufacturing, an accurate description of moving boundaries is crucial in addition to proper treatment of the chemical and physical processes. In these applications the moving boundaries usually are the surface of a wafer or the surfaces between wafers and deposited layers. One approach is to use a cellular format, where the simulation domain is divided into cubic or cuboid cells and each cell either belongs to the exterior vacuum above the wafer or to its interior (Pyka 2000, Pyka et al. 2001). One disadvantage of this method is that computing surface normals and tangents leads to accuracy problems. Surface normals are, e.g., crucial for computing fluxes to the surface for simulating transport phenomena via the radiosity approach.

The level set method (Sethian 1999b) provides an interesting alternative and a solution to the above mentioned problem. This method is a relatively new method for describing boundaries, i.e., curves, surfaces or hypersurfaces in arbitrary dimensions, and their evolution in time. Applying this method means solving a certain partial differential equation and extracting the zero level set of its solution.

Firstly, the basic ideas of the level set method are presented. Next extending the so called speed function and a narrow banding algorithm for accelerating simulations are discussed. The two seemingly unconnected concepts are combined in our implementation in order to enable efficient surface evolutions. For the case of radiosity simulations an algorithm for coalescing certain surface elements is presented. Its purpose is to increase accuracy where it is needed most and reduce simulation time by reducing the size of the radiosity matrix.

Secondly, mass transport in the diffusion and radiosity regimes above the wafer is shortly discussed and the modeling approach is described.

Finally, an example where SiO_2 is deposited from TEOS (Tetraethoxysilane) is presented. Simulation results are shown and discussed comparing with SEM (scanning electron microscope) pictures of fabricated test structures. In the experiments a SiO_2 layer was deposited into trenches roughly $4\mu\mathrm{m}$ deep and $2\mu\mathrm{m}$ wide, where the final layer thickness was in the range of $1\mu\mathrm{m}$ for the flat wafer surface. Simulation results for the diffusion and the radiosity regime are presented and discussed, and the algorithms mentioned at the beginning of the paper are illustrated as well.

THE LEVEL SET METHOD

The level set method (Sethian 1999a, Sethian 1999b) provides means for describing boundaries, i.e., curves, surfaces or hypersurfaces in arbitrary dimensions, and their evolution in time, where the evolution is caused by forces or fluxes normal to the surface. The basic idea is to view the curve or surface in question at a certaim time t as the zero level set (with respect to the space variables) of a certain function $u(t, \mathbf{x})$, the so called level set function (Adalsteinsson and Sethian 1995, Adalsteinsson and Sethian 1999, Sethian 1996). Thus the initial surface is the set $\{\mathbf{x} \mid u(0, \mathbf{x}) = 0\}$.

Each point on the surface is moved with a certain speed normal to the surface and this determines the time evolution of the surface. The speed normal to the surface will be denoted by $F(t, \mathbf{x})$. For points on the zero level set it is usually determined by physical models and in our case by the etching and deposition processes, or more precisely by the fluxes of certain gas species and subsequent surface reactions. The speed function $F(t, \mathbf{x})$ generally depends on time and the space variables, and we assume for now that it is defined on the whole simulation domain and for the time interval considered.

The surface at a later time t_1 shall also be considered as the zero level set of the function $u(t, \mathbf{x})$, namely $\{\mathbf{x} \mid u(t_1, \mathbf{x}) = 0\}$. This leads to the level set equation

$$\begin{aligned} &u_t + F(t, \mathbf{x}) \|\nabla_{\mathbf{x}} u\| = 0, \\ &u(0, \mathbf{x}) \quad \text{given} \end{aligned}$$

in the unknown variable u, where $u(0, \mathbf{x})$ determines the initial surface. Having solved this equation the zero level set of the solution is the seeked curve or surface at all later times.

Now in order to apply the level set method a suitable initial function $u(0, \mathbf{x})$ has to be determined first. There are two requirements: first it goes without saying that its zero level set has to be the surface given by the application, and second it should essentially be a linear function so that in the final surface extraction step linear interpolation can be applied. A beneficial choice is the signed distance function of a point from the given surface. This function is the common distance function multiplied by minus or plus one depending on which side of the surface the point lies in. The common distance function of a point x from a set M is then defined by $d(x, M) := \inf_{y \in M} d(x, y)$, where d is a metric, usually the Euclidean distance.

In summary, first the initial level set grid is calculated as the signed distance function from a given initial surface. Then the speed function values on the whole grid are used to update the level set grid in a finite difference or finite element scheme. Usually the values of the speed function are not determined on the whole domain by the physical models and therefore have to extrapolated suitably from the values provided on the boundary, i.e., the zero level set. This will be discussed in the next section. In the last step, the surface extraction step, the curve or surface is reconstructed from the function values on the grid, where the zero level set is approximated by lines or triangles using linear interpolation along grid lines.

Of course the use of linear interpolation seems arbitrary and must be justified. It can be shown that if the extension velocity F satisfies $\nabla F \cdot \nabla u = 0$, then the level set function u remains the signed distance function for all time. Hence it has to be ensured that the speed function used fulfills this condition. This is the case with the implementation developed. Now since u remains the signed distance function, which is essentially a linear function, linear interpolation is indeed the best method. Consequently, small changes in the level set function result in small movements of the surface, whereas in the case of a cellular format a cell is either part of the material or not. Thus, although in the numerical application the level set function is eventually calculated on a grid, the resolution achieved is in fact much higher than the resolution of the grid, and hence higher than the resolution achieved using a cellular format on a grid of same size.

The advantages of the level set method are twofold: the resolution achieved is higher than the resolution of the grid where the calculations take place (cf. (Pyka et al. 2000, Hössinger et al. 2001)) and calculating surface normals, crucial for radiosity simulations, is straightforward and much more precise than when using a cellular format.

EXTENDING THE SPEED FUNCTION AND NARROW BANDING

Before discussing the details of the various simulation steps, the outline of topography simulations for semiconductor processes is given in Figure 1.

In applications linking to physical models the speed function is not known on the whole simulation domain, but only at the surface. In order to use the level set method it has to be suitably extended from the known values to the whole simulation domain. This can be done iteratively by starting from the points nearest to the surface. Mathematical arguments show that the signed distance functions can by maintained from on time step to the next by choosing a suitable extension.

The idea leading to fast level set algorithms stems from observing that only the values of the level set function near its zero level set are essential, and thus only the values at

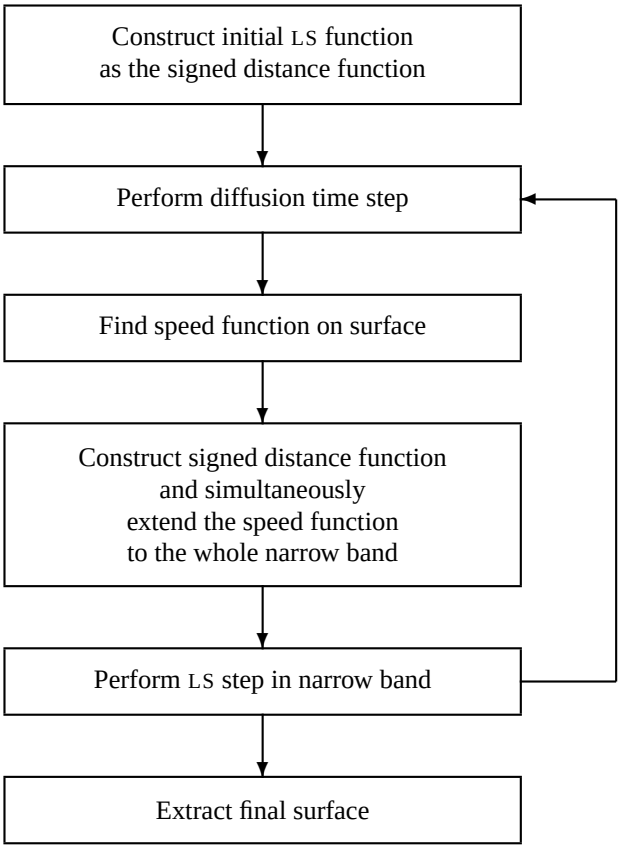


Figure 1: Overview of the simulation flow combining transport by diffusion and surface evolution using the level set (LS) method. The simulation stops when a prescribed time is reached on when a layer of prescribed thickness has been deposited.

the grid points in a narrow band around the zero level set have to be calculated. As the zero level set moves, the signed distance function in the narrow band must be maintained.

Both, extending the speed function and narrow banding require constructing the distance function from the zero level set in the order of increasing distance. But calculating the exact distance function from a curve or surface consisting of a large number of small line segments or triangles is computationally expensive and can only be justified for the initialization. An approximation to the distance function can be computed by a special fast marching method (Sethian 1999b).

For the first time narrow banding and extending the speed function were combined into one algorithm. This algorithm provides several benefits. First, the speed function is retained as the signed distance function throughout the simulation, which assures good accuracy till the end of the simulation. Second, narrow banding reduces the number of active points that have to be updated from $O(n^2)$ to $O(n)$. By retaining the signed distance function the width of the narrow band can be kept down to two points on each side (cf. Figure 6) without decreasing accuracy. Third, time consuming calculations (cf. (Adalsteinsson and Sethian 1999)) are reduced to a minimum by intertwining the computations necessary for narrow banding and extending the speed function. Finally the width of the narrow band can be adjusted if desired.

An outline of the algorithm is as follows. First the initial points near the zero level set, where the speed function is known, and the neighboring trial points are determined. In the main loop it is checked if there is still a trial point to be considered in the narrow band. All trial points are stored in a heap ordered by their distance to the zero level set. If there is a point to be considered, both its distance is approximated and its extension speed calculated, and its neighbors are updated accordingly. Finally after the main loop, bookkeeping information for the narrow band points is updated using distance information just computed. The computation time consumed by this algorithm is negligible compared to that required for the physical models, while it provides high accuracy.

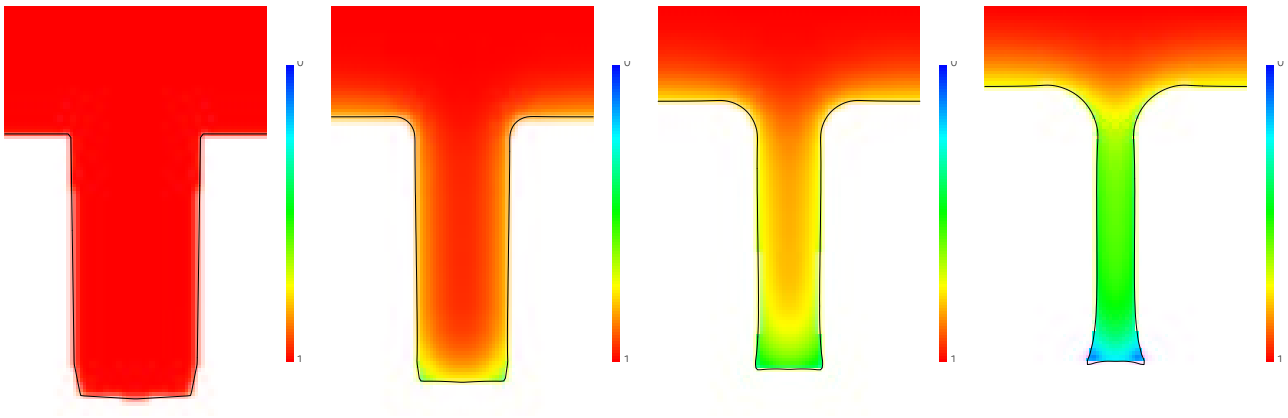
By intertwining both, extending the speed function and narrow banding, in our implementation expensive calculations are kept to a minimum. Although the level set method is a seemingly computationally expensive method, since it requires solving a partial differential equation for describing surface evolutions, the computation time consumed for the surface evolution by narrow banding is negligible compared to that required for the physical models, e.g., diffusion.

TRANSPORT

In an integrated simulation of transport phenomena and surface evolution the transport phenomena above the wafer surface specify the etch and deposition rates. They can broadly be divided into two classes according to the mean free path length, although this distinction is only a rough classification and the suitable model in each case depends on other considerations as well:

- If the mean free path length is much larger than the diameter of the simulation domain, the collision of single particles can be neglected and the transport can be simulated using the radiosity approach.
- If, on the other hand, the mean free path length is much smaller than the simulation domain, the collisions between single particles play a major role and their concentration is determined by the diffusion equation.

Of course, the question which case to choose for a certain deposition process depends on the conditions of the deposi-



(a) Step 1. **(b)** Step 31. **(c)** Step 61. **(d)** Step 91.

Figure 2: The figures show the surface evolution of a typical trench structure during a deposition processes. The concentration of the diffusing particles above the wafer is between 0 and 1 and it is constantly 1 at the top of the simulatin domain.

tion or etching process to be simulated and more details cannot be presented here. In order to provide a general purposes simulator, both approaches are indispensable.

In the following the modeling of both cases as used in the simulations presented later is described.

Transport in the Diffusion Regime

Here transport is governed by the well-known diffusion equation

$$\frac{\partial c}{\partial t} = \nabla \cdot (D \nabla c),$$

where c is the concentration and D the diffusion coefficient. The boundary conditions (cf. Figure 2) are usually as follows: at the top of the simulation domain a Dirichlet boundary condition is assumed, i.e., a constant concentration is supplied by the reactor; on the left and right hand side a Neumann boundary concentration is assumed, i.e., the fluxes are zero; and finally the fluxes on the wafer surface are determined by the deposition rates.

Transport in the Radiosity Regime

In this case particles are injected into the simulation domain from sources above the wafer and their way is traced according to reflections from the surface until it attaches itself at a certain location or leaves the simulation domain. How particles are reflected mostly depends on their energy. This model is similar to ray tracing in computer graphics.

A formulation of the radiosity method for the transport of particles of low energy only, where luminescent reflection is assumed, which excludes the case of high energetic particles (i.e., ions), can be found, e.g., in (Sethian 1999b):

$$\text{Flux} = \frac{\beta - \beta_0}{1-\beta} I_S + \\ + \frac{\beta(1-\beta_0)}{1-\beta} \underbrace{L^{-1}(L^{-1} - (1-\beta)\Psi)^{-1}}_{T:=} I_S,$$

where I_S is the vector of fluxes coming from the sources to the surface elements, β_0 the sticking coefficient for particles coming directly from the source, β the one for secondary bounces, L the diagonal matrix containing the lengths of the surface elements, and

$$\Psi_{ij} = \frac{n_i \cdot (t_j - t_i) n_j \cdot (t_i - t_j)}{\pi |t_j - t_i|^3} [i \text{ visible } j],$$

where t_i are the centroids of the surface elements, n_i their unit normal vectors, and $[i \text{ visible } j]$ is 1 or 0 if surface element j is visible from i respectively not.

In the case of multiple, low energy species the calculation of the visibility matrix and the inverse T only depends on topographic information and thus does not have to be repeated for each species.

COALESCING SURFACE ELEMENTS

When using radiosity models for simulating the transport of particles above the wafer in the case where the length of the mean free path is greater than the size of the feature, two operations consume most of the computation time. The first

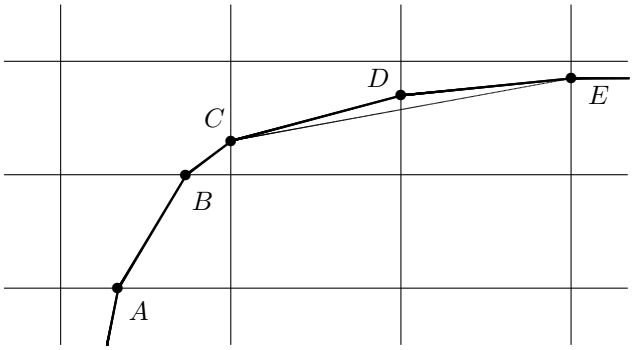


Figure 3: Illustration of the coalescing algorithm. Since the angles at A, B, and C are above the threshold value, no replacement takes place here. The angle at D is below the threshold value, and thus the segments CD and DE are replaced by a new segment CE.

Coarsening Steps	Visibility Test	Flux Calculation
0	1	1
1	0.29	0.10
2	0.12	0.02

Table 1: Comparison of the speed of the visibility test and of the calculation of the fluxes on surface elements by radiosity both with the coalescing algorithm and without. The computation time relative to the conventional algorithm, equaling 1, is shown.

operation is determining the visibility between all surface elements, which is an $O(n^2)$ operation where n denotes the number of surface elements extracted from the level set grid. The second operation is solving a certain system of linear equations, which leads to calculating the inverse of a matrix with n^2 elements, which is an $O(n^3)$ operation.

Obviously increasing the number of surface elements is not a remedy in cases where high resolution is required. High resolution is needed, e.g., near the trench opening, and the bottom of the trench, and for the simulation of microtrenching and side wall push back. One approach is to devise a refinement and coarsening strategy for unstructured grids on which the level set equation is numerically solved. This, however, complicates the fast marching algorithm necessary for extending the speed function. A different approach was taken in this work by coarsening the surfaces after having been extracted from the level set grid.

The algorithm works by walking down the list of surface elements extracted as the zero level set and calculating the angle α between two neighboring surface elements. Whenever $|\pi - \alpha|$ is below a certain threshold value of a few degrees, the neighboring elements are coalesced into one. After one sweep through the list, the algorithm can be reapplied for further coarsening. After k coarsening sweeps, at most 2^k surface elements are coalesced into one. The resulting longer surface elements are used for the radiosity calculation, after which the fluxes are translated back from the coarsened elements to the original ones.

SIMULATION RESULTS

Example of Deposition in the Diffusion Regime

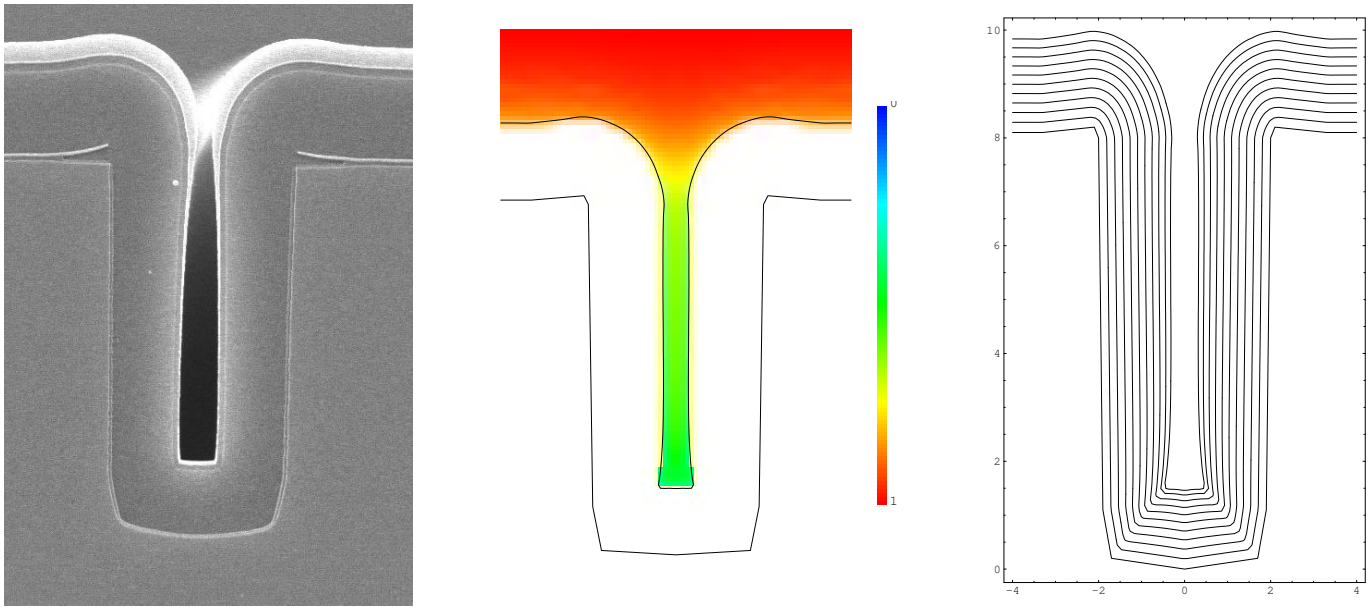
Before discussing the simulations and the effects of the speeding up strategies, the physical modeling approach is shortly described.

In order to calculate the thickness Δd of the film deposited during a time interval of length Δt, we observe that Δd is proportional to Δt, to an Arrhenius term, and to the deposition rate R_i corresponding to the deposition model chosen. This implies $\Delta d = \Delta t \cdot k_e \mathrm{e}^{-E/kT} \cdot R_i$. Here $k_e \mathrm{e}^{-E/kT}$ is the Arrhenius term with activation energy E, absolute temperature T, and preexponential constant k_e. R_i is the deposition rate of the deposition model chosen, where two heterogeneous deposition models, a homogeneous intermediate-mediated deposition model, and a heterogeneous deposition with byproduct inhibition model are available (Raupp et al. 1992). This setup also provides a way to determine the actual chemical reaction, which is a non-trivial problem and can only be solved indirectly by comparing measurements and simulation results.

As noted from the characteristic shape of the measurements provided (cf. Figure 4) and the process conditions employed, the transport of TEOS in the boundary layer above the wafer happens in the diffusion regime and thus is governed by the diffusion equation. The boundary conditions are as follows: at the top of the simulation domain a Dirichlet boundary condition is assumed, i.e., a constant concentration is supplied from the convective zone in the reactor; on the left and right hand side a Neumann boundary concentration is assumed, i.e., the fluxes are zero; and finally the fluxes at the wafer surface are determined by the amount of particles deposited.

The simulation flow for arriving at the results shown in Figure 2 is depicted in Figure 1. In Figure 2 the results of SiO_2 deposition from TEOS are shown. The initial structures are roughly rectangular trenches $4\mu\mathrm{m}$ deep and $2\mu\mathrm{m}$ wide. Initially the TEOS concentration equals 1 everywhere and the boundary conditions from the previous section were applied, where the concentration at the top was constantly 1.

The depletion of TEOS during the deposition can clearly be seen. This leads to the narrowing of the upper part of the trench, where TEOS is still supplied. Furthermore the



(a) SEM image. **(b)** Simulated trench. **(c)** Zero level sets.

Figure 4: (a): SEM (scanning electron microscope) image of a cross section through a trench about $4\,\mu\mathrm{m}$ deep and $2\,\mu\mathrm{m}$ wide. **(b)**: Simulation result showing the final trench geometry. The TEOS concentration above the wafer in the boundary layer is shown as well. **(c)**: Some intermediate zero level set of the simulation shown in (b).

stronger depletion in the corners at the bottom of the trench causes the edges to become concave, although they were initially convex. Finally at the upper corners the deposited layer becomes thicker than on the flat layer surface, which might not seem intuitive at first, but is also observed in the SEM pictures.

Example of Deposition in the Radiosity Regime

In experiments SiO_2 was deposited under different process conditions in a different set of trenches for Power MOSFETs with a higher aspect ratio than the trench shown in the SEM image in Figure 4. For the following simulation results, the radiosity module of the topography simulator was employed.

For adjusting the parameters of the model the topography simulator was used in combination with the optimization framework SIESTA (Heitzinger et al. 2001, Heitzinger and Selberherr 2002). Hence extracting the model parameters is performed automatically and can be immediately applied to different measurements and structures produced under different process conditions.

A simulation result is shown in Figure 5, where the surface coalescing algorithm described above was employed. First, the computation time of the level set algorithm with narrow banding as described above (cf. Figures 6, 7, and 8) is negligible compared to the computation time of the physical models. This, however, is not the case when narrow banding is not employed. Table 1 lists the relative computation time of testing for visibility and the actual radiosity calculation both with and without the coalescing algorithm. The simulation result with coarsening in Figure 5 is nearly identical to the one yielded when no coarsening was applied. Accuracy is hardly affected, but the simulation time considerably decreased.

CONCLUSION

State of the art algorithms for surface evolution processes like etching and deposition processes used for manufacturing semiconductor components have been developed and implemented. The simulator developed consists of three independent modules, namely the level set module, a reaction module, and a diffusion module, which can be used for simulating all common deposition and etching processes as well. Step coverages measured in several SEM images were used for extracting model parameters, where good quantitative agreement was achieved. Hence the process conditions have been optimized with respect to the quality of the trenches and manufacturing throughput.

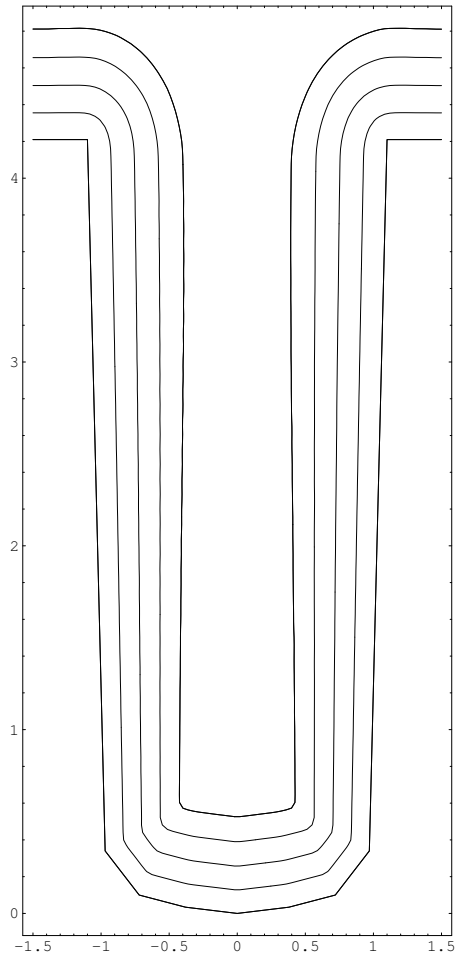


Figure 5: A simulation result showing initial, intermediate, and final surfaces. The resolution of the underlying level set grid was $80 \cdot 160$. The coarsening algorithm was applied twice, coalescing at most four surface elements into one, and the threshold angle was $3°$. This result is nearly identical to the one achieved when no coarsening was applied.

Two strategies for increasing the accuracy of radiosity simulations are presented and compared to measurements of a deposition process. The first method is an algorithm which performs three level set computations in parallel: calculating the signed distance function via a fast marching algorithm, extending the speed function, and moving the narrow band according to the new zero level set. This gives rise to a fast and accurate level set algorithm.

The second method is a coarsening algorithm which ensures fine resolution of the surface in parts of the boundary with relatively high curvature, i.e., where it is needed most. These parts are typically the opening of the trench, its bottom, and places where microtrenching and side wall push

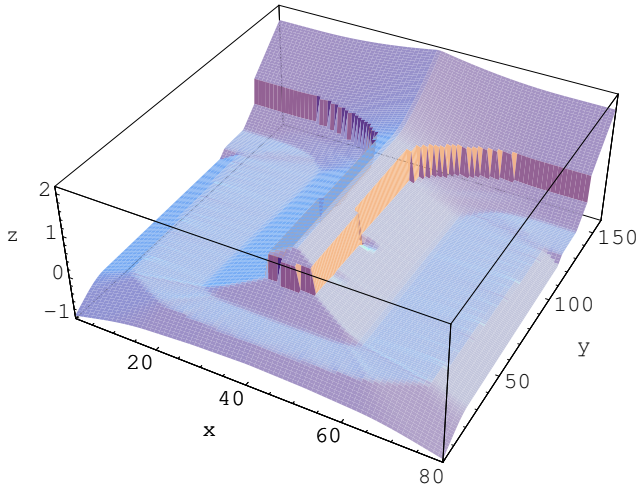


Figure 6: The level set function after the last step of the simulation whose result is shown in Figure 5. The active narrow band around the zero level set retains the signed distance function, whereas other grid points have not been updated.

back take place. At the same time the resolution is lowered where possible which reduces the demand on computational resources significantly.

In order to verify the models and simulator developed it was applied to the deposition of SiO_2 from TEOS in different Silicon trenches under different process conditions. All observed effects match well comparing the SEM pictures and the simulation results.

ACKNOWLEDGEMENTS

The authors acknowledge support from the “Christian Doppler Forschungsgesellschaft”, Vienna, Austria, and from Infineon Technologies, Villach, Austria.

REFERENCES

Adalsteinsson, D. and J.A. Sethian, 1995, ”A Fast Level Set Method for Propagating Interfaces”. *J. Comput. Phys.*, 118(2):269–277.

Adalsteinsson, D. and J.A. Sethian, 1999, ”The Fast Construction of Extension Velocities in Level Set Methods”. *J. Comput. Phys.*, 148(1):2–22.

Heitzinger, C.; T. Binder; and S. Selberherr, June 2001, ”Parallel TCAD Optimization and Parameter Extraction for Computationally Expensive Objective Functions”. In *Proc. 15th European Simulation Multiconference: Modelling and Simulation 2001 (ESM 2001)*, pages 534–538, Prague.

Heitzinger, C. and S. Selberherr, 2002, ”An Extensible TCAD Optimization Framework Combining Gradient Based and Genetic

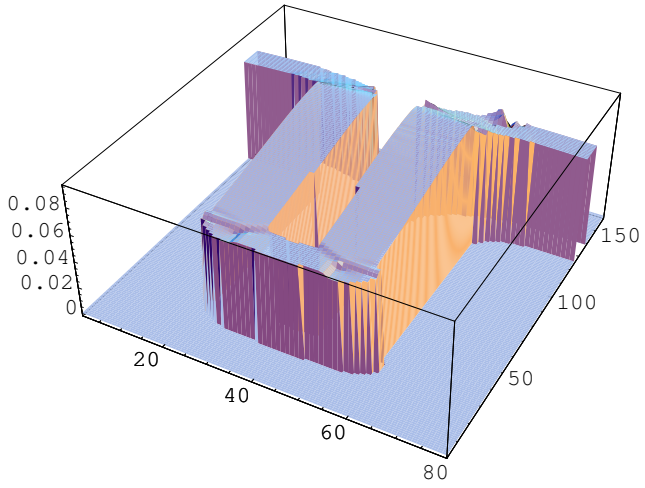


Figure 7: The extended speed function in the narrow band in the last step of a simulation. In this simulation, no coarsening was performed, but apart from that it is identical to the one leading to Figure 5.

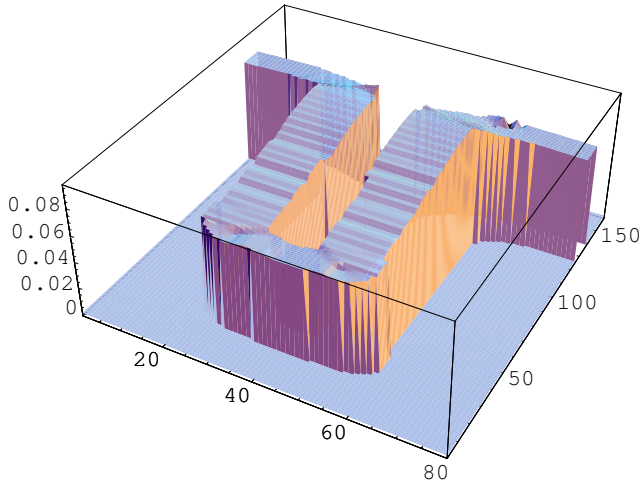


Figure 8: The extended speed function in the narrow band in the last step of the simulation leading to Figure 5. The coarsening can clearly be seen at the side walls of the trench.

Optimizers". *Microelectronics Journal (Design, Modeling and Simulation in Microelectronics and MEMS; Smart Electronics and MEMS)*, 33(1-2):61–68.

Hössinger, A.; T. Binder; W. Pyka; and S. Selberherr, September 2001, "Advanced Hybrid Cellular Based Approach for Three-Dimensional Etching and Deposition Simulation". In *Simulation of Semiconductor Processes and Devices*, pages 424–427, Athens, Greece.

Pyka, W.; R. Martins; and S. Selberherr, 2000, "Optimized Algorithms for Three-Dimensional Cellular Topography Simulation". *IEEE J.Technology Computer Aided Design*, 20. `http://www.ieee.org/products/online/journal/tcad/accepted/Pyka-March00/`.

Pyka, W.; C. Heitzinger; N. Tamaoki; T. Takase; T. Ohmine; and S. Selberherr, September 2001, "Monitoring Arsenic In-Situ Doping with Advanced Models for Poly-Silicon CVD". In *Simulation of Semiconductor Processes and Devices*, pages 124–127, Athens, Greece.

Pyka, W., 2000, *Feature Scale Modeling for Etching and Deposition Processes in Semiconductor Manufacturing*. Dissertation, Technische Universität Wien. http://www.iue.tuwien.ac.at/phd/pyka.

Raupp, G.B.; F.A. Shemansky; and T.S. Cale, November 1992, "Kinetics and Mechanism of Silicon Dioxide Deposition Through Thermal Pyrolysis of Tetraethoxysilane". *J.Vac.Sci.Technol.B*, 10(6):2422–2430.

Sethian, J.A., 1996, "A Fast Marching Level Set Method for Monotonically Advancing Fronts". *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1591–1595.

Sethian, J.A., 1999, "Fast Marching Methods". *SIAM Rev.*, 41(2):199–235.

Sethian, J.A., 1999, *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge.

AUTHOR BIOGRAPHY

CLEMENS HEITZINGER was born in Linz, Austria, in 1974. After the compulsory military service he studied Technical Mathematics at the Technische Universität Wien, where he received the degree of Diplom-Ingenieur (with honors) in 1999. He joined the Institut für Mikroelektronik in February 2000, where he is currently working for his doctoral degree. From March to May 2001 he held a position as visiting researcher at the Sony Technology Center in Hon-Atsugi (Tokyo, Japan). His scientific interests include modeling and simulationn of semiconductor processes and devices and symmetry methods for partial differential equations.

SIEGFRIED SELBERHERR was born in Klosterneuburg, Austria, in 1955. He received the degree of Diplom-Ingenieur in Electrical Engineering and the doctoral degree in Technical Sciences from the Technische Universität Wien in 1978 and 1981, respectively. Dr. Selberherr has been holding the venia docendi on Computer-Aided Design since 1984. Since 1988 he has been the head of the Institut für Mikroelektronik and since 1999 he has been Dean of the Fakultät für Elektrotechnik und Informationstechnik. His current topics are modeling and simulation of problems for microelectronics engineering.