ELSEVIER

Thin Solid Films 365 (2000) 368–375

thin solid films

www.elsevier.com/locate/tsf

Multiple scale integrated modeling of deposition processes

Tushar P. Merchant[a,*], Matthias K. Gobbert[b], Timothy S. Cale[c], Leonard J. Borucki[a]

[a]*Predictive Engineering Laboratory, Motorola Inc., 2200 W. Broadway Road, Mesa, AZ 85202, USA*
[b]*Department of Mathematics and Statistics, University of Maryland–Baltimore County, Baltimore, MD 21250, USA*
[c]*Department of Chemical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

Abstract

The ability to predict feature profile evolution across wafers during processing using equipment scale operating conditions is one important goal of process engineers. We present an integrated approach for simulating the multiple length scales needed to address this problem for thermal chemical vapour deposition (CVD) processes. In this approach, continuum models on the reactor scale and mesoscopic scales are coupled tightly with ballistic transport models on the feature scale to predict micro and macro loading effects in a transient environment. As an example of this approach, the transient simulation results for thermal deposition of silicon dioxide from tetraethoxysilane (TEOS) are presented. The efficiency of the approach presented and extensions to more complex systems are briefly discussed.

Keywords: Chemical vapour deposition; Tetraethoxysilane

1. Introduction

Conventional TCAD simulation is extensively used in the semiconductor industry to determine what needs to be done to manufacture a particular device. However, the means to achieve it in silicon has always been left to the process engineers, who have to attain those objectives within the limits of their equipment. This often leads to expensive experiments using silicon, which not only leads to additional cost, but also increases the overall development cycle time. Simulation of the process tool, also known as equipment simulation, is gaining acceptance as the method to assist in converting the 'what' needs of the device engineer to the 'how' needs of the process engineer.

Traditionally, equipment simulation has been used to address issues of reactor design, optimization and prediction of blanket wafer scale properties such as growth rates [1]. Feature scale simulations, on the other hand, have been predominantly used to predict film topography and composition in deposition or etch processes, based on flux distributions at the feature surface [2]. The major drawback with these two approaches has been the disconnect, where neither approach adequately addresses the other scale, resulting in limited predictive capability for deposition or etch processes over patterned wafers. The basic difficulty in merging these two approaches has been the inherent disparity in length scales, which span more than six orders of magnitude, from about a meter at the equipment scale to submicrometer at the feature scale. This disparity is expected to get larger as the wafer size increases and the line width decreases. The smaller margin of error required in these newer processes, combined with the additional expense of experiments, makes integrated simulation over multiple length scales increasingly attractive. These simulations could be used to predict both wafer scale uniformity as well as feature profile evolution as a function of position on the wafer, from equipment scale operating conditions, thereby reducing the development cost of a new process.

There have been only a limited number of attempts to resolve patterned wafer effects such as micro and macro loading from traditional equipment scale models. This is because, in addition to the grid resolution issues associated with the disparate length scales, the traditional continuum models of the equipment scale cannot be extended to the molecular description necessary in the Knudsen regime of the feature scale. These simulations have tried to resolve this issue by assuming an effective area approximation, wherein an effective area associated with the densely packed features is given a higher deposition rate than surrounding field areas [3]. Within this framework, Holleman and coworkers have been able to investigate loading issues associated with tungsten silicide deposition in a single wafer LPCVD reactor. Another approach used by Cale et al. [4] consists of using a reactor scale simulator to first predict the conditions near the wafer surface based on the operating conditions. These local conditions are then

* Corresponding author.

0040-6090/00/$ - see front matter
PII: S0040-6090(99)01055-X

used in a feature scale simulator to predict deposition profiles at various points on the wafer surface. While the approach by Holleman et al. attempted to determine the effect of feature scale on the reactor scale, the primary focus of the second approach was the feature scale. Thus in both of these approaches, there was no feedback of information of one scale to the other. A mesoscopic scale model has also been introduced by Gobbert et al. [5], to simulate deposition processes at the scale of a few dies (mm). It was used to provide understanding of deposition processes at a scale inaccessible by both traditional equipment and feature scale models. This model too has limited predictive capability because of its dependence on input parameters from the two other scales. A truly predictive simulator must have the capability to couple phenomena occurring at the reactor scale, mesoscale, as well as the feature scale, where information from each scale is transferred correctly and coupled tightly to the other scales.

The first approach towards full integration of reactor scale and feature scale simulations was by Gobbert et al. [6]. In their technique, the species concentrations of the reactor scale were given as input to a feature scale simulator, which then returned a homogenized net flux of each species back to the reactor scale. The process is iterated between the multiple scales till a fully consistent solution is obtained. They also introduced a mesoscopic scale in between the reactor and feature scale, which is used to provide further information regarding variations of species concentrations and fluxes at the die scale and on the scale of feature clusters. This approach was demonstrated for pseudo-steady state conditions, where the slow change in topography evolution on the feature scale did not impact the reactor scale. Details of this approach are elaborated on somewhat in subsequent sections.

Another approach used to link up reactor and feature scale simulations has been the effective reactivity function formulation described by Rogers and Jensen [7]. In this technique, the reactor and feature scale simulations are linked together using this effective reactivity ε, which includes effects of both surface variations as well as feature scale transport. A Monte Carlo-based ballistic transport scheme is used to calculate the effective reactivity of a single type of feature. The reactivity of each set of features is then linearly superimposed to obtain ε. The key to superimposition lies in ensuring that the source plane for the Monte Carlo simulations is at a sufficient height to resolve the gradients in ε. They empirically found that a height of about one third of the mean free path was sufficient to satisfy this assumption. This effective reactivity is then fed into the reactor scale simulation as an enhancement factor to the flux boundary condition over a blanket area. The reactor and feature scale simulations are then iterated to arrive at a consistent solution. They applied this technique to the simulation of tungsten deposition, showing deposition variations across the wafer due to depletion effects, as well as snapshots of reactor scale concentration variation at the beginning and end of deposition. They also illustrate that simple exposed area approximations do not always give the same results as the more detailed calculations. While this technique can be used to couple reactor and feature scale simulations, the prohibitive time requirements for each feature scale simulation and calculation of ε, and the iterative nature of this process, make it difficult to use it effectively in its current form.

The approach presented here illustrates an integrated model spanning multiple length scales. The results from higher order scales are fed into the lower order scales, and the results from the lower orders are fed back up to form a tightly coupled solution. It builds on the steady state model presented by Gobbert et al. [6] to include transients. These transient simulations are carried out at all length scales to capture time variation of species concentrations on the reactor scale, and the corresponding effect at the lower scales. Thermal deposition of silicon dioxide from tetraethoxysilane (TEOS) is used as the example for illustrating this approach.

The organization of the paper is as follows. After the individual models are presented, there is a description of how the multiple scales are integrated and the simulation technique used. The physical model and some illustrative results in a steady and transient environment are presented next. Finally there is a discussion on implementation issues, and future trends as well as challenges for application to other semiconductor manufacturing processes.

2. Multiple scale integrated model

2.1. Description of individual models

The multiple scale integrated model is realized by coupling together individual models at different length scales. This approach not only avoids the excessive grid resolution that would be necessary for a single model, but also allows for the capture of the underlying physics by varying the model description according to the scale. Thus the model physics is changed from continuum to the molecular flow (ballistic transport) regime at length scales where the mean free path becomes comparable to geometry dimensions, i.e. the Knudsen number becomes large. The integrated model used here involves a transient reactor scale simulator for continuum mechanics, which solves for the governing equations of mass, momentum, heat, and species transport to compute the flow, temperature and species concentrations as functions of position and time. In this case, the commercial finite element based fluid dynamics package FIDAP (FIDAP 7.6, Fluent Inc., 500 Davis St. Suite 600, Evanston, IL 60201, 1996) is used to solve the governing equations. As is typical for fluid flow solutions, the velocities, temperature and species concentrations are interpolated using quadratic basis functions, and pressure is interpolated with linear basis functions. On the reactor scale,

the wafer topography is not taken into account explicitly, and the information from the other scales is only incorporated as a net flux boundary condition on all nodes of the finite element grid representing the wafer. At the next level of detail, which corresponds to the mesoscale, continuum equations are still valid and the same solver is used as for the reactor scale. At the feature scale however, the continuum equations are no longer valid, and transport of individual molecules need to be taken into account to accurately represent the underlying physics. In our case, the ballistic transport and reaction model incorporated in EVOLVE (EVOLVE is a deposition, etch, and reflow process simulator developed by T.S. Cale with funding from the Semiconductor Research Corporation, the National Science Foundation, and Motorola) is used for simulating the feature scale. This feature scale simulator uses the incoming concentrations and angular distributions of individual species, and couples it with very general surface chemistry representations to deterministically obtain the net flux of individual species, and move the surface according to the growth kinetics. While EVOLVE can be used to simulate topography evolution for ionic (non-maxwellian) species, in the case of thermal deposition as illustrated here, the incoming angular flux of species is always maxwellian, and hence simplifies the coupling process. One key advantage of using a deterministic simulator such as EVOLVE over the Monte Carlo approach is the much shorter simulation times required, it takes only a few seconds to do a simulation as opposed to the few tens of minutes needed in the latter approaches. Another big advantage of the deterministic approach is the relative ease with which reaction chemistries can be incorporated.

Fig. 1 depicts the transition between three models at different scales, and shows a representative axisymmetric finite element grid of the reactor and mesoscale, as well as the feature geometry and pitch. When using a two-scale model, which includes the reactor and feature scale, the region between the grid nodes of the reactor scale are implicitly assumed to have a uniform pattern corresponding to the feature density as simulated on the feature scale. This implies that feature size and pitch are uniform on a length scale associated with the nodal distance, and results in a simplified representation of the prescribed pattern density. In the three-scale model on the other hand, the reactor scale nodes give a representation of the prescribed density associated with the mesoscale. The grid nodes at the mesoscale get net fluxes that correspond to the feature density. Thus, additional information about the changes in feature density is obtained in the three scale model that is absent in the two-scale model.

2.2. Coupling of individual models

In principle, combining the models to form an integrated model is fairly straightforward. The initial guess of species concentrations over an element on the reactor scale is interpolated onto the mesoscale grid using the finite element basis functions. The mesoscale model is then solved in a coupled fashion with multiple feature scale simulations at each node representing a patterned area. The guesses for concentrations at the individual nodes are fed into the feature scale simulator, which returns a net flux of each species at that node. The mesoscale and feature scale models are then iterated until convergence is obtained. The net flux of each species within the converged mesoscale solution is homogenized to the reactor scale grid. This is then fed back along with the net fluxes associated with the flat regions of the wafer into the reactor scale simulator, which proceeds iteratively to obtain the next guess. Additional details of the homogenization process can be obtained from the paper by Gobbert et al. [6]. In the case of a fully coupled transient simulation, the matter is further complicated, since the feature scale simulator should move the surface corresponding to the time step taken in the reactor scale simulation, and give the resulting fluxes at the end of the time step. In addition to that, the current profile at every node has to be stored so that only an update is made to the current profile at a subsequent time. While this is not difficult to conceptualize, formulation of the problem for arbitrary features and nodes is not a trivial bookkeeping task.

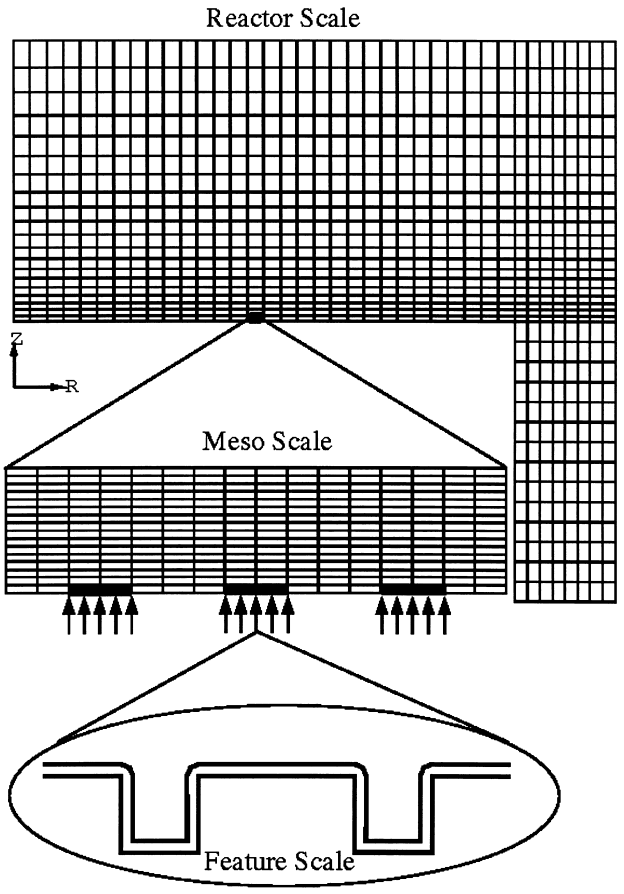

Fig. 1. Reactor, mesoscale and feature scale simulation grid and transition scheme. Arrows indicate position of lower level models.

The addition of other intermediate simulation scales is straightforward as long as the continuum equations are valid. On the feature scale, the particulate nature of the species transport is taken into account and is 'driven' by transferring the flux distribution of each species at the feature surface. For chemical vapor deposition (CVD) it suffices to estimate the local temperature and species concentrations. Fig. 2 schematically shows the iterative algorithm associated with the coupling process.

3. Simulation specifics

The thermal deposition of silicon dioxide from TEOS is used as an example to demonstrate the use of the multiple scale integrated model. The kinetic model for the gas phase involves six gas phase species involved in four gas phase reactions [8]. The major gas phase intermediates are triethoxysilanol, water, ethylene, and ethanol. In addition, there are six surface species involved in eight surface reactions, whose primary byproducts are water, ethylene and ethanol. Details of the chemical mechanism and reaction kinetics can be obtained from the paper of Gobbert et al. [6]. Improved reaction kinetic parameter estimates for TEOS decomposition are discussed in another paper in this volume.

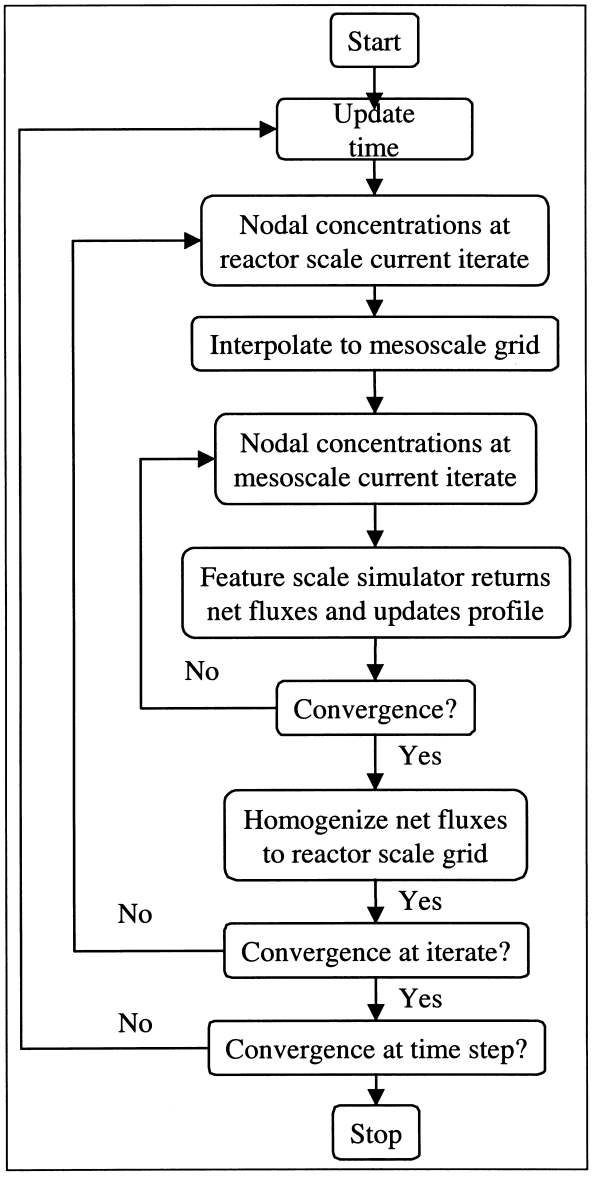

Fig. 2. Algorithm for combining multiple scale simulators for transient simulations.

The reactor model used for the simulations is a generic axisymmetric single wafer reactor with the reactive gases entering from the top showerhead and impinging on the susceptor heated wafer. The gases enter at room temperature about 5 cm from the susceptor at 1000 K. The gases leave the water cooled reactor from an annular outlet at the bottom of the susceptor. For the simulations shown here Argon is the inert carrier gas flowing at 2 slm at a pressure of 0.01 atm. All the transport properties, such as diffusion coefficients, thermal diffusion, viscosity etc. are determined using the CHEMKIN database [9] which was coupled into FIDAP. The forward and reverse reaction rates are also automatically computed using the CHEMKIN formulation at the nodal points in the reactor. Some of the grid nodes corresponding to elements of the wafer surface are flagged to have patterns, where the reactor scale model is coupled to either a mesoscopic scale model (three-scale approach) or a feature scale model (two-scale approach). On other wafer nodes, the flux corresponding to a flat topography is returned to the reactor scale model. In all cases, the desired local heterogeneous reaction rates are computed by EVOLVE using CHEMKIN. The reactor scale solution is obtained by solving the governing equations on a cross-section of the cylindrical chamber as shown in Fig. 1. The transient simulations are started under the pseudo steady state conditions described above. This would physically correspond to a situation where the reactor flow, concentration and temperature fields are stabilized at the operating specified conditions with the patterned wafer exposed to the incoming reactive gases, but little growth has taken place to alter the original feature topography. Since the growth rates are much smaller than residence times for the flow this is a reasonable approximation.

For most of the simulations shown here, a single patterned area about 3.3-mm long is placed at about the halfway point of the wafer radius. The mesoscopic scale model spans this distance, and in some cases contains three clusters of features of width 0.4 mm each. The height of the mesoscopic scale model is taken to be 1 mm, which is more than three times the mean free path for all species under those conditions. The grid for the mesoscale simulation and the feature scale geometry are also shown in Fig. 1. For steady state feature scale simulations, the individual features are infinite trenches of 1 μm height, 1 μm width, and a pitch of 3 μm. For the transient calculations, the

feature width is 0.5 μm, which corresponds to an aspect ratio of 2, and the pitch is 2.5 μ in order to increase the effects of loading; i.e. the impact of local feature density on local deposition rate.

4. Results

The validation of the multiple scale approach has already been documented in the paper by Gobbert et al. [6]. In that paper it was shown that introducing the additional mesoscale model to a two scale model did not change the computed concentrations of reactive intermediates for the two extreme cases of a blanket wafer, and a uniform die. Fig. 3 is a representative sample of the pseudo steady-state results from two- and three-scale models. It shows the mass fraction of the reaction byproduct water for three different cases, as functions of radial position on the wafer. The three cases shown in the graph correspond to the case of a blanket wafer that has no topography; a uniform die case where the feature density extends uniformly throughout the 3.3-mm patterned area, and the case of three clusters of features across the die. As is intuitive, the mass fraction of water increases as the feature density increases. The water mass fraction for the clustered die case falls in between the other two scenarios since its average feature density is in between the blanket wafer and uniform die cases. The point to note here is that the effect is fairly local, affecting only a very small region near the die. Also, there is no information about the effect of individual clusters that can be seen at the reactor scale. The information about the individual clusters can be clearly seen on the mesoscale, which can capture these variations. This is illustrated in Fig. 4. The variations in the mass fraction of water associated with the clustered die are clearly visible at this scale. The variations decrease in amplitude quickly as a function of distance away from the surface, and only an effective value is observed at the reactor scale. Thus the mesoscale model can capture effects that are not captured at any other scale. Since the amplitudes of the variations over each cluster at the mesoscale are about the same, the effect of variations within a die do not seem to be important for TEOS deposition under these conditions, although it could be important for other systems.

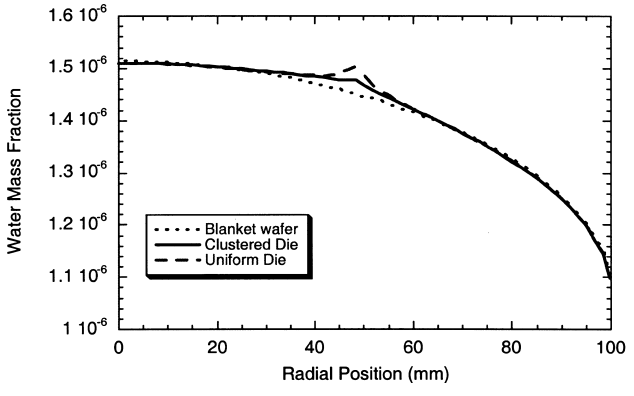

Fig. 3. Mass fraction of water at the wafer surface as functions of radial position on the wafer at the reactor scale, for three cases; a blanket (unpatterned) wafer, a uniformly patterned die, and a die with three clusters of features, as described in the text.

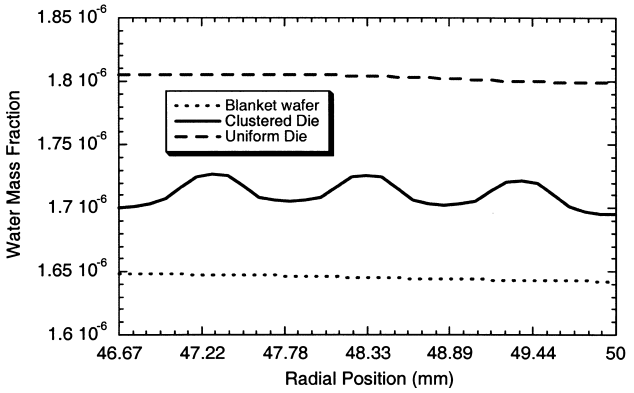

Fig. 4. Mass fraction of water at the wafer surface as functions of radial position on the mesoscale, for three cases; a blanket (unpatterned) wafer, a uniformly patterned die, and a die with three clusters of features, as described in the text.

Fig. 5 shows the mass fraction of water radially across the wafer for different patterns on the wafer. In the case of five patterns, the patterns were placed uniformly every 10 mm from the center of the wafer and are 10 mm wide. The effect of just one pattern is small and local, however the effect of having multiple large patterns is quite substantial. The mass fraction of water increases by 22% over the blanket wafer case, and the variations in the water mass fraction due to the patterns are more significant. The figure also shows differences in the amplitudes of the variations across the wafer, indicating that these loading effects are dependent on the position and density of the patterns on the wafer. Similar behavior is observed for other intermediates and byproducts as well, although the amplitudes are quantitatively different for each species. While a simple exposed area approximation may give a similar overall behavior for the mass fractions, it would not be able to correctly predict the differences

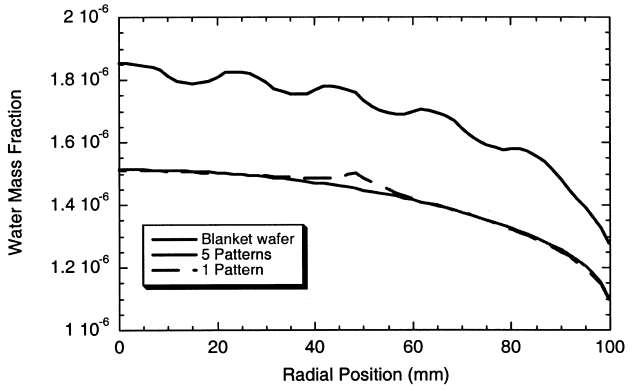

Fig. 5. Mass fraction of water at the wafer surface as functions of radial position across the wafer, for three cases; an unpatterned wafer, a wafer with five patterns, and a wafer with one pattern, as described in the text.

in the amplitude of variations across the wafer. These results imply that under certain conditions and chemistries the effect of loading can be significant to the overall performance of the process.

Fig. 6 shows an example of the evolution of the silicon dioxide film profile in a representative feature at the center of the die on the wafer, during a transient simulation. The profile contours are 100 s apart and hence this aspect ratio two feature closes somewhere between 600 and 700 s. The result shows that for this particular chemistry and conditions, the TEOS deposition is conformal, and a smooth uniform film is deposited without any void formation.

Fig. 7 shows the transient behavior of the key reactive intermediate triethoxysilanol on the reactor scale. The results are scaled with the mass fraction of triethoxysilanol at the center of the wafer at zero time. There is a sharp drop in its mass fraction right near the die, because it has a near-unity sticking coefficient. The aspect ratio of two for the features and the higher packing density contribute to the sharp drop. From the figure it can be observed that there is an initial drop in this intermediate fraction which extends all across the wafer. This can be attributed to the change in transport of the intermediate caused by the sudden additional depletion at the die. The time scale of one second corresponds to the residence time within the reactor. At the time of 10 s, the mass fraction stabilizes close to the initial state. The major effects of the features on the reactor scale only become observable close to feature closure and beyond it. As closure of the local features is approached, the depletion of the triethoxysilanol due to deposition in the features is reduced and is clearly observed on the reactor scale. Once the features are completely closed, the behavior of the intermediate mass fraction is just like deposition on a blanket wafer.

The transient behavior of the byproduct water is shown in Fig. 8. It shows the scaled mass fractions of water as functions of radius for selected deposition times. Unlike the heavier triethoxysilanol, the impact on water is more pronounced initially. The water mass fraction continues to decrease across the entire wafer and goes through a minimum around 10 s into the deposition. The mass fraction then starts stabilizing and rising back up again to its original state. The initial reduction at the wafer surface is associated with momentum and diffusive transport of water from the wafer surface into the reactor. Once the gradients in the water concentration are stabilized, the surface fraction starts increasing due to the continued generation of water from the surface and stabilization of the transport and reaction dynamics. Towards feature closure, the rate of generation exceeds the transport and the mass fraction is slightly higher. After closure, the mass fraction stabilizes at the new level of a blanket wafer. Also seen from this figure is that the change in water mass fraction is much smaller near the features than triethoxysilanol. This is because water is predominantly generated from the wafer surface as a byproduct, and being lighter will diffuse more easily throughout the reactor.

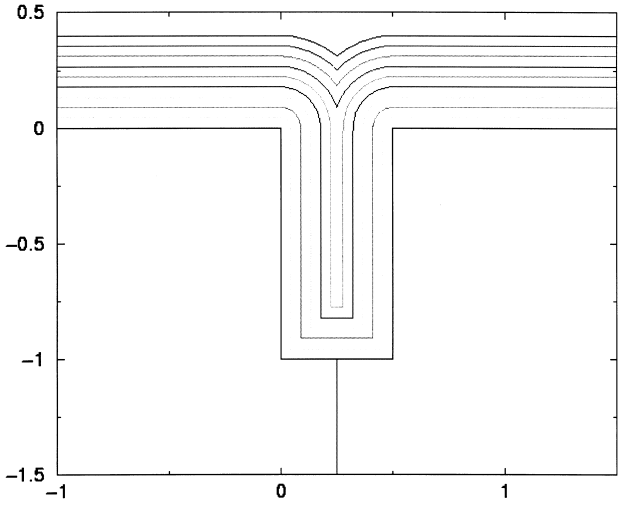

Fig. 6. Transient evolution due to silicon dioxide deposition in an aspect ratio two feature at the center node of the die on the wafer surface. Profile contours are every 100 s and extend to 900 s.

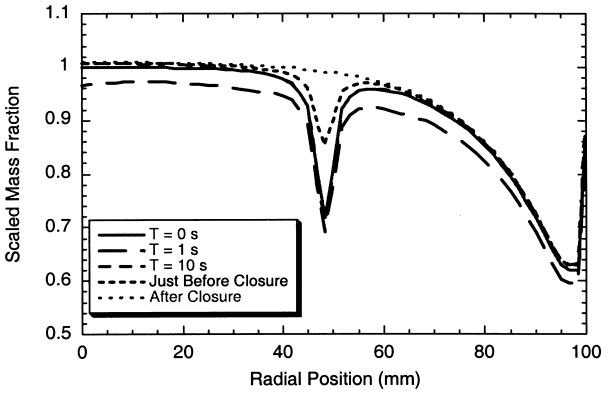

Fig. 7. Mass fraction of triethoxysilanol at the wafer surface as functions of radial position, for selected deposition times to demonstrate transient evolution. The results are scaled to the mass fraction at radius = 0, and time $t = 0$.

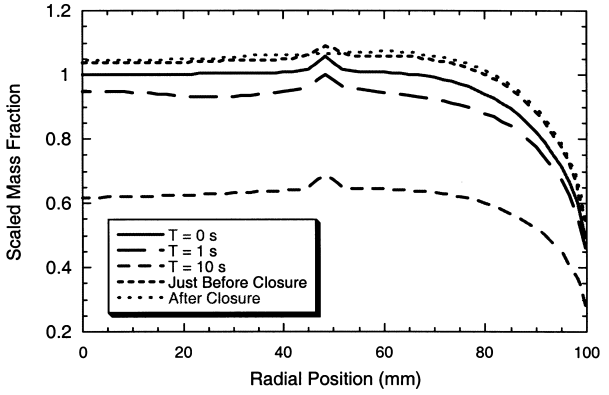

Fig. 8. Mass fraction of water at the wafer surface as functions of radial position, for selected deposition times to demonstrate transient evolution. The results are scaled to the mass fraction at radius = 0, and time $t = 0$.

These results show that there are different time scales associated with the transport and reaction of reactants and byproducts in the reactor, which could have an impact on feature scale deposition across the wafer. Another point of consideration is that despite the fact that the time scale for filling a feature is much larger than the reactor transport time, some variation in species fluxes at other portions of the wafer occur just because of transport effects. While this is not significant for this particular chemistry and conditions, it could be of importance for other processes.

5. Discussion and future directions

The brute force technique of local refinement of the grid size to deal with the multiple scale simulation problem described in this paper is fraught with computational efficiency and algorithm convergence issues. Also, it is difficult to extend the brute force technique over scales where the governing equations change, such as the transition from the continuum to molecular regimes. The techniques and results presented here just represent a starting point in investigating the interaction of multiple-scale phenomena typical of semiconductor manufacturing. The concept of using fluxes and concentrations at an interface to bridge different simulations at multiple scales enables the coupled simulation of distinct phenomena. While this technique is straightforward, actual implementation is not easy. As mentioned above, a considerable amount of bookkeeping is needed to ensure seamless transition of information from one scale to another. Not only does the coding have to be general enough to account for multiple nodes, the geometry of the feature scale should also be stored at the current time and some previous times for correct time integration. Storage of the feature profile at earlier times is necessary to allow for non-convergence of the reactor scale solution at a predicted time step. In addition, since the particular implementation of the coupling of scales involves two independent and stand-alone codes (FIDAP and EVOLVE), there is minimal coupling of the time scales and exchange of information. For every lower scale simulation, there is valuable time lost associated with the standard startup sequence of the simulator. While the actual time per simulation is small (it is about 1 s), the number of feature scale simulations, at each node on the wafer for every reactor scale iteration and every time step, adds up quickly. Also, since the reactor scale can take a lot of iterations to converge at a particular time step, the overall time for a simulation increases rapidly. Efficiency of the process is achieved by performing mesoscale and feature scale simulations only once over many reactor scale iterations. This is possible because the change in the current guesses of the reactor scale mass fractions has a relatively small impact on the resulting fluxes. While this reduces the overall solution time considerably, the optimal ratio of reactor scale iterates to lower scale simulations will depend on the particular chemistry and operating conditions. Efficiency of the process can also be improved dramatically with parallelization of the lower scale simulations, which involves additional coding for resource management.

There is also the issue of robustness of the codes with respect to integration with other scales. This is especially true for transient simulations, where the time scale associated with one level is vastly different from the time scale of the other level. For instance, the time scale in the reactor scale simulator is associated with the flow characteristics and is on the order of seconds. The time scale on the feature scale is however the time to fill the feature, which can be a few hundred seconds. Both simulators should be robust enough to handle these widely varying time scales in situations where the codes are coupled. Thus a robust reactor scale simulator should not give erroneous answers at long times, and the feature scale simulator should not diverge for very short time steps. A lot of this can be overcome by having a closer coupling of the simulators, with information such as perhaps the gradients in time also being shared. Other relevant information which might be needed to ensure close coupling between these scales needs to be investigated further.

While methods for integration of reactor and feature scale simulations for thermal processes are known, the extension to non-thermal processes such as plasma processes also needs to be carried out. In addition to the concentration information typical of thermal processes, the ion angle and energy distributions also have to be transferred to the lower scales. In this area, there has already been considerable progress in feeding reactor scale information to feature scale. Hoekstra et al. [10] have combined their plasma reactor code with a sheath model, and a Monte Carlo based feature scale simulator to simulate feature profile evolution from reactor scale inputs. Coronell et al. [11,12] have used a similar technique for simulation of ionized physical vapor deposition processes. They have also used ab-initio calculations of the metal surface to determine ion angle and energy dependent sputter kinetics, and incorporated it into their feature scale simulations. In both of these cases though, information has not been fed back to the higher scales. The incorporation of this feedback mechanism will result in truly integrated simulations for plasma processes.

Multiple scale integrated models are also needed for other semiconductor manufacturing processes such as electroplating and CMP. In these cases, continuum models are typically valid from the reactor to the feature scale, but other relevant information such as current density and stress have to be transferred. Appropriate homogenization techniques will have to be established for transfer of information between scales.

The other obvious area of future work is in 3D simulations at the multiple scales. While the reactor scale simulators can handle 3D well, the major changes are occurring at the feature scale. Algorithms based on direct evolution of meshed surfaces are difficult to implement in 3D. Monte Carlo based feature scale simulators easily allow simple

physics such as ion angle and energy dependent kinetics to be incorporated: however, it is very expensive to include chemistry of any detail. In addition, surface evolution is complicated because of the discrete nature of the simulation. Improvement of speed and the ability to simulate detailed chemistry will have to be made before integration with other scales becomes straightforward. Another approach is a level set-based algorithm, which is robust, handles 3D evolution of features easily, and is currently being used to predict 3D surface topography with detailed kinetics [13]. For integration of multiple scales in 3D, simulation time will become the major issue, and novel schemes or simplifications such as an effective reactivity may be needed to alleviate the problem.

The smallest scale of all the simulations discussed here has been the feature scale. As device sizes continue to shrink, the influence of operating conditions on the size of individual grains and morphology become important to performance. Simulations will have to proceed to the next level of grain growth to address this issue. The combination of discrete and continuum models becomes important, and techniques will be needed not only to predict the evolution of heterogeneous grains into a continuum film, but its integration with the higher scales.

6. Conclusions

Techniques that can predict feature scale behavior from reactor scale operating conditions are very relevant to the semiconductor industry, as they can reduce development time as well as expedite process optimization. A multiple scale integrated model that combines reactor scale to feature scale in a transient environment was presented here. Within this framework, the reactor scale concentrations are fed to mesoscale and feature scale simulations to obtain the net fluxes at the deposition surface. These net fluxes are fed back to the reactor scale simulation and iterated to achieve self consistent solutions at all length scales. The transient thermal deposition of silicon dioxide from TEOS was used as an illustrative example of this technique. While the framework of this methodology is established, there is scope for improvement in both the efficiency and robustness of this approach. Extensions of this technique to three dimensions and other semiconductor manufacturing processes will make its use even more prevalent in the industry.

Acknowledgements

Timothy Cale and Matthias Gobbert acknowledge support for part of the work presented by the Semiconductor Research Corporation, DARPA and the National Science Foundation.

References

[1] W.L. Holstein, Prog. Cryst. Growth Charact. 24 (1992) 111.
[2] T.S. Cale, V. Mahadev, Modeling of Film Deposition for Microelectronic Applications, Thin Films 2 (1996) 175.
[3] J. Holleman, A. Hasper, C.R. Kleijn, J. Electrochem. Soc. 140 (1993) 818.
[4] T.S. Cale, J.-H. Park, T.H. Gandy, G.B. Raupp, M.K. Jam, Chem. Eng. Commun. 122 (1993) 197.
[5] M.K. Gobbert, C.A. Ringhofer, T.S. Cale, J. Electrochem. Soc. 143 (1996) 2624.
[6] M.K. Gobbert, T.P. Merchant, L.J. Borucki, T.S. Cale, J. Electrochem. Soc. 144 (1997) 3945.
[7] S. Rogers, K.F. Jensen, J. Appl. Phys. 83 (1998) 524.
[8] M.E. Coltrin, P. Ho, H.K. Moffat, R.J. Buss, Thin Solid Films, (1999) submitted.
[9] R.J. Kee, F.M. Rupley, E. Meeks, J.A. Miller, CHEMKIN-III: a Fortran chemical kinetics package for the analysis of gas-phase chemical and plasma kinetics, Sandia National Laboratories, Livermore, CA, 1996.
[10] R.J. Hoekstra, M.J. Kushner, V. Sukharev, P. Schoenborn, J. Vac. Sci. Tech. B 16 (4) (1998) 2102.
[11] D.G. Coronell, P. Ventzek, V. Arunachalam, C.-L. Liu, D. Hansen, J. Kress, A. Voter, Presented at the 45th AVS Int. Symp., Baltimore MD, Nov.1998,
[12] D.G. Coronell, D.E. Hansen, A.F. Voter, C.-L. Liu, X.Y. Liu, J.D. Kress, Appl. Phys. Lett. 73 (1998) 3860.
[13] T.S. Cale, B.R. Rogers, T.P. Merchant, L.J. Borucki, Comput. Mater. Sci. 12 (1998) 333.