Simulation Of Multilayer Structures For VLSI
Using The SUPREM-III Process Simulation Program

S. E. Hansen, J. D. Shott, P. M. Fahey, J. D. Plummer, and R. W. Dutton

Integrated Circuits Laboratory
Stanford University
Stanford, CA 94305 USA

Summary

With the advent of VLSI, device structures have become more complex, utilizing new materials and processing techniques. The recently released version III of the process modeling program SUPREM has been designed to model many of these new materials and processing techniques. Up to ten layers and ten different materials can be modeled by the most recent version of the program. Default models and coefficients for silicon, polycrystalline silicon, silicon-dioxide, and silicon-nitride are included in the program. As the coefficients which define the material characteristics are definable by the user, additional materials may be added to the default ones. New models for oxidation, diffusion, epitaxy, and ion implantation are included. Many of these new models have a much more physical basis than those included in previous versions of the program.

This paper presents some of the capabilities of the latest version of the SUPREM-III program and demonstrates their use through a detailed example which simulates a state-of-the-art 2 µm CMOS process. The aspects of effective process simulation are discussed, with particular emphasis on understanding the limitations of process simulators and measurement techniques as well as the need for calibration of the program's coefficients. A comparison of the simulations with physical and electrical data from measured devices will be shown. This comparison will point out both the modeling capabilities in SUPREM-III and areas in which further improvement and better physical understanding are required. In addition, we conclude with a discussion of work in progress, recent modeling efforts on new materials and processes that are to be added to future releases of SUPREM-III.
1. Introduction

Today, Computer Aided Design and Computer Aided Engineering tools are used extensively over the entire range of VLSI design. For many years now each additional innovation has been achieved with the assistance of increasingly sophisticated computer based tools. In the area of processing technology, the availability of computer simulation programs has given the process design engineer a tool that can significantly reduce or even eliminate the number of costly experimental runs. While modern processes have become more complex, process simulators have benefited from increasingly accurate physical models, providing an attractive alternative to the iterative empirical approach of designing or optimizing processes.

One of these simulation programs, SUPREM [1-4], has been under continuous development since 1976. The latest version, SUPREM-III, is a one dimensional process simulator capable of modeling most of the common steps in silicon integrated circuit processing. It is possible with SUPREM-III to simulate a multilayer structure of up to ten materials. Coefficients and models for silicon, silicon-dioxide, polysilicon, and silicon-nitride are included in the program, as is the capability for defining other materials. The program can simultaneously model the introduction and redistribution of all four of the common impurities, B, P, As, and Sb.

This paper presents some of the capabilities of the SUPREM-III program and demonstrates their use through example. Specifically, a state-of-the-art 2μm CMOS process is simulated and compared to measured data. The object of this example is to gain an understanding of the sensitivity of critical points in the process, and at the same time to calibrate the default model coefficients in the SUPREM program with the equipment and procedures used in the fabrication facility.

2. The Process

Stanford has developed a fully implanted 2μm CMOS fabrication sequence which will be described in this section. This technology has become the most sophisticated baseline technology in our laboratory and has provided a rigorous set of challenges to SUPREM-III. Figure 1 shows a cross-section of the n-channel and p-channel devices along with vertical lines running through the cross-section at channel, drain/source, and field regions of each type of device.

In order to achieve the lithographic control required of a 2μm technology, we are using an Ultratech 900 wafer-stepper which features automatic, dark-field, site-by-site alignment on all masking levels. Furthermore, we are using plasma etching of polysilicon, silicon nitride, and deposited oxide layers. Conventional local oxidation techniques are used to isolate active regions.

The decision to use an n-well approach was based largely on the compatibility with our existing NMOS process. As a result, our starting material is p-type, (100) oriented silicon with a doping concentration of \( \approx 9 \times 10^{14} \text{cm}^{-3} \). The n-well is implanted with a P\(^{31} \) dose of \( 2.5 \times 10^{12} \text{cm}^{-2} \) at an energy of 100 keV. The n-well is then annealed, oxidized for 32 minutes at 1000°C, and
then driven-in for 960 minutes at 1150°C in an inert ambient. The surface concentration of the n-well ($\approx 1 \times 10^{16}$ cm$^{-3}$) is determined primarily by the need to maintain a sufficiently high surface concentration to prevent field inversion of the n-well — we wanted to avoid the use of an additional channel-stop implant into the n-well regions. The depth of the n-well ($\approx 4 \mu$m) is then determined by the need to prevent punch-through of the parasitic vertical $pnp$ transistor under worst-case bias conditions.

After the n-well drive-in we strip all the oxide, grow a thin ($\approx 400\AA$) layer of "pad" oxide for 48 minutes at 1000°C in a dry O$_2$ ambient and then deposit LPCVD Si$_3$N$_4$ to begin the locally oxidized field region processing. After patterning the active regions in the Si$_3$N$_4$ (and while the active region photoresist is still in place) we apply and pattern a second layer of photoresist which is used to protect the n-well regions from the upcoming n-channel field implant. After the B$^{11}$ implant of $1 \times 10^{13}$ cm$^{-2}$ at 100 KeV, we then grow the thick field oxide in a pyrogenic steam ambient (partial pressure of H$_2$O $\approx 0.80$) for 190 minutes at 1000°C. The initial thickness of this field region is $\approx 7500\AA$, but will be reduced to $\approx 6000\AA$ by various unmasked SiO$_2$ etches prior to polysilicon deposition.

For a gate oxide thickness of 400Å, we have found it impractical to use a single threshold shifting implant to shift the thresholds of the n- and p-channel devices by the desired amount. We therefore perform an unmasked B$^{11}$ implant of $4 \times 10^{11}$ cm$^{-2}$ at 35 KeV to shift the threshold of the p-channel devices to the desired level $V_{TP} \approx -1.0$ V followed by an additional masking step which protects the p-channel devices while the n-channel devices receive an additional B$^{11}$ implant of $3.5 \times 10^{11}$ cm$^{-2}$ at 35 KeV to bring their threshold up to a
designed value of $V_{T_n} \approx 0.85 \text{ V}$. These threshold voltages have been selected to insure adequate protection from sub-threshold leakage currents. The $p$-channel device with an $n^+$ polysilicon gate electrode becomes a buried-channel device if we try to shift its threshold to a symmetrical value of $V_{T_p} \approx -0.85 \text{ V}$. Buried channel devices have been shown to possess inferior sub-threshold and drain-induced barrier lowering characteristics [5]. Because neither the $n$-channel nor the $p$-channel device will be operated with “substrate” potentials other than 0 V and $V_{DD}$, respectively, we did not find it necessary to use an additional deep implant to prevent drain/source punch-through problems. In order to avoid potential damage of the gate oxide due to either the implants and/or resist processing, we actually implant the threshold shifting implants through a sacrificial 400Å oxide. Once this gate oxide is removed, we then grow the actual gate oxide in a dry $O_2$ ambient for 48 minute at $1000^\circ C$. Redistribution and segregation of the boron during the gate oxidation must be carefully considered in order to achieve good threshold control during this step.

LPCVD polysilicon is next deposited at 620°C, doped using $POCl_3$ (1100 ppm partial pressure at 950°C for 30 minutes), and patterned in a plasma reactor using a mixture of $SF_6$ and $C_2ClF_5$. A photoresist layer defining the $n^+$ source/drain and $n$-well contact regions is applied and a high dose ($6 \times 10^{15} \text{ cm}^{-2}$ at 100 KeV) arsenic implant is performed, followed by an anneal for 20 minutes at 950°C. The $p^+$ regions are then defined with a photoresist layer and the wafers are implanted with $B^{11}$ ($1 \times 10^{16} \text{ cm}^{-2}$ at 35 KeV) followed by a low-temperature anneal/oxidation of 60 minutes at 850°C (30 minute in an $N_2$ ambient and 30 minutes in an $O_2$ ambient). This oxidation prevents the penetration of unwanted phosphorus into the $p$-channel drain/source regions from the P-glass which is deposited in an LTO reactor at 450°C. After depositing the P-glass, we place the wafers in a pyrogenic steam ambient for 30 minutes at 900°C which more fully activates the $p$-channel drain/source regions and densifies the P-glass. Even in a steam ambient we get little, if any, reflow of the P-glass layer; we rely on our sputtered aluminum alloy deposition to provide good step coverage in the process.

Following the densification of the P-glass, we pattern the contact holes with photoresist and etch them in a plasma etcher. Because the drain/source junctions are quite shallow and the selectivity of the plasma etch is only $\approx 3 : 1$ ($SiO_2 : Si$), the use of end-point detection at this step is particularly important in order to insure that we don’t exacerbate the contacting of shallow junctions during the metalization procedures.

Following contact hole etch, we selectively deposit a thin ($\approx 1000Å$) layer of tungsten in an LPCVD reactor [6] and then sputter 1.0 $\mu$m of aluminum alloy for the final metalization layer.

3. Simulation

While the state of silicon process modeling has improved dramatically over the past several years, much remains to be done. As long as this remains true, process simulation — which is built upon process modeling — will continue to have its weak points. In addition, anything less than complete
three dimensional modeling and simulation will depend upon certain implicit approximations. The process engineer must at all times be aware of these weak points and approximations in order to extract full value from the simulation program. With a one-dimensional simulator such as SUPREM-III, care must be taken not to simulate regions of a device structure where multi-dimensional effects are significant. Equally important, the user of any process simulator must be aware of the weaknesses or limitations of the various models used.

The simulation and evaluation of the CMOS process described in Section 2 has identified a few points that should be mentioned in regard to the use of SUPREM-III (and possibly other process simulators). The first and most important point is that the models and coefficients used in any process simulator should be benchmarked against each processing line. Since any two processing lines using the same processing sequence will almost certainly give more or less different results, it is reasonable to assume that at least some of the default coefficients provided can not be applied accurately to any given line. Coefficients, such as those for diffusion and segregation, have more of a physical basis and seldom need modification for use within their range of applicability. Those that are sensitive to the procedures or equipment used may need to be adjusted.

Certain oxidation coefficients can be considered process or equipment sensitive. While the Deal-Grove parameters [7], B and B/A, may be used with a fair amount of confidence, others, such as the oxidant partial pressure [8] and thin-oxide rate coefficients [9], are likely candidates for calibration. The default B, B/A, and partial pressure values used in SUPREM were derived from experimental work done at Fairchild Semiconductor [10], and SUPREM can accurately reproduce the experimental results of that work. However, SiO₂ layers grown in the Stanford Integrated Circuit Laboratory are consistently thinner than those grown at Fairchild, and SUPREM’s default coefficients will give oxide thicknesses that are \( \approx 10\% \) thicker than those measured experimentally at Stanford. By reducing the default partial pressure for pyrogenic steam oxidation from 0.92 to 0.80 atmospheres, the oxide thicknesses predicted by SUPREM are in close agreement with measurement as shown in Table 1.

The coefficients controlling SiO₂ growth for layer thicknesses of less than approximately 200Å are expected to be sensitive to the cleaning procedures.
used [11-12]. The default coefficients are derived from the work of Massoud [9] which utilized an in situ cleaning not common in standard processing. The difference in measured vs. simulation for the pad oxide shown in Table 1 may be due to differences in cleaning procedures. If a high degree of accuracy is needed for any thin SiO₂ layers, then the applicability of these coefficients should be investigated.

One of the most common mistakes made in the use of a process simulation program (or in the use of any simulation program for that matter) is to apply it to a problem outside the valid range of its models or coefficients. It is unsafe to assume that if the program gives accurate results for oxide thicknesses grown at three atmospheres and 950 degrees centigrade, it will also give accurate results at ten atmospheres and 800 degrees. It will depend largely on the range of experimental data from which the coefficients and the models are derived. In SUPREM-III, much emphasis has been placed on physical, rather than empirical models, so that they can be used with some confidence outside the range of conditions over which they were derived. However, the user must realize that these models have ranges of operation outside of which their results will be based on extrapolation, and that if operated in these areas, other unexpected physical effects may become significant.

For example, the initial simulations of the above process were affected adversely by the limited range of the mobility data. The short times and low temperatures of the anneal and drive-ins resulted in peak carrier concentrations above $2.0 \times 10^{20} \text{cm}^{-3}$ in the arsenic source/drains. The highest carrier concentration data point used by the mobility model [13] in extracting the coefficients is $2.0 \times 10^{20} \text{cm}^{-3}$. The mobility coefficients are used in a fourth-order polynomial to extract the mobility as a function of the carrier concentration. When used outside of the valid range, the polynomial can cause SUPREM-III, without warning, to predict sheet resistances that will vary widely with small variations in concentration. Future versions of the program will make use of the carrier vs. mobility data in tabular form which will allow users to both determine the range of the current mobility data, as well as to add to or modify the existing data. While in a few instances the program will warn the user when an attempt is made to use coefficients outside of their valid range, the majority of the coefficients may be employed in situations far from the conditions under which they were derived. Users must be aware of this when using the program, especially for processes that utilize very high concentrations, very high or low temperatures or pressures, or other non-standard conditions. In the future, efforts will be made to have the program recognize more of these problems itself and warn the user.

Another caveat that must be recognized for effective use of a process simulation program is that of model interaction. When investigating a particular physical effect, it is desirable to develop experiments and measurement techniques that will isolate the details of that physical process from any others. This may present a problem when models are incorporated into a complete process simulation program such as SUPREM. The way in which the physical processes interact may be very different from their uncoupled behavior. Fortunately, the more physically based the model, the less important this should become.
In the simulation of the CMOS process described above, five different cross-sections were selected (see Figure 1). To reduce both the CPU and real time required for the simulation, the processing sequence was divided into several pieces. The processing for the early stages of many cross-sections is identical, and up to the point where the processing for the cross-sections diverge, a common input sequence is used. The resulting structure serves as the starting point in the simulation inputs that will complete the various cross-sections' processing.

4. Comparison and Evaluation

The structures and impurity distributions resulting from SUPREM-III simulations at each cross-section indicated in Figure 1 are shown in Figures 2 through 6. The corresponding measured concentration profiles derived from spreading resistance measurements [14] are overlaid in each of these figures.

Figures 2 and 3 show the impurity distributions for cross-sections 1 and 2, the p- and n-channel gate regions respectively. The agreement in both cases between simulation and measurement is excellent. As mentioned in the previous section, the SUPREM-III-predicted gate oxide thickness, a critical parameter in these devices, was also in excellent agreement with the value measured by ellipsometry techniques. The desired threshold voltage for each of these devices was \( V_{T_p} = -1.0 \text{V} \) and \( V_{T_n} = +0.85 \text{V} \). The Figures 7a and 7b show the measured and simulated channel conductances vs. gate voltage for the p-channel device at three n-well biases, while Figures 8a and 8b have the equivalent curves for the n-channel device at three substrate biases. Table 2 shows the SUPREM-III-predicted threshold voltages and the values subsequently extracted from the measured curves in Figures 7a and 8a. As can be seen, SUPREM predicted both the p- and n-channel device threshold voltages to within a reasonable margin, exactly for the p-channel device, though shifted \( \approx -0.35 \text{V} \) for the n-channel device. Considering the excellent agreement between the impurity distributions and oxide thicknesses, differences in the slopes of the curves in Figures 7 and 8 may be attributable to differences between the mobility values used and those present in the gate region of the devices. In the case of the p-channel device, using \( \frac{1}{2} \) of the bulk mobility value brings the 0 V bias curve into almost complete alignment and the slopes of the other curves compare favorably. For the n-channel device, using mobility values of \( \frac{5}{9} \) bulk mobility gives similar results, though it appears that the SUPREM electrical simulation does not deal completely with some aspect of the body bias effect. At present, SUPREM-III electrical simulations use only bulk mobility values.

In the source/drain regions, the simulated junction depths, shown in Table 3, agree well with measured values, especially when considering the possible error in the measurement technique at these shallow depths. Figures 4 and 5 show the structure and impurity profiles in the p- and n-channel source/drain regions. The impurity profiles show significant differences at the higher concentrations, especially near the Si/SiO\(_2\) interface. The discrepancy near the surface can be expected in profiles extracted from spreading resistance measurements.
Figure 2. SUPREM-III (solid) and measured (dashed) impurity distributions through the $p$-channel device gate region.

Figure 3. SUPREM-III (solid) and measured (dashed) impurity distributions through the $n$-channel device gate region.
Figure 5. SUPREM-III (solid) and measured (dashed) impurity distributions through the n-channel device source/drain region.

Figure 6. SUPREM-III (solid) and measured (dashed) impurity distributions through the field region.
Table 2
Threshold Voltage Comparison

<table>
<thead>
<tr>
<th>Device</th>
<th>Measured (Volts)</th>
<th>SUPREM-III (Volts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Channel</td>
<td>-1.35</td>
<td>-1.35</td>
</tr>
<tr>
<td>N-Channel</td>
<td>0.8</td>
<td>0.45</td>
</tr>
<tr>
<td>Field Region</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3
Junction Depth Comparison

<table>
<thead>
<tr>
<th>Diffusion</th>
<th>Measured (μm)</th>
<th>SUPREM-III (μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Channel Source/Drain</td>
<td>0.4</td>
<td>0.36</td>
</tr>
<tr>
<td>N-Channel Source/Drain</td>
<td>0.27</td>
<td>0.257</td>
</tr>
<tr>
<td>N-Well</td>
<td>4.41</td>
<td>4.82</td>
</tr>
</tbody>
</table>

Table 4
Sheet Resistance Comparison

<table>
<thead>
<tr>
<th>Diffusion</th>
<th>Measured (Ω/□)</th>
<th>SUPREM-III (Ω/□)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Channel Source/Drain</td>
<td>240</td>
<td>129</td>
</tr>
<tr>
<td>N-Channel Source/Drain</td>
<td>57</td>
<td>19</td>
</tr>
<tr>
<td>N-Well</td>
<td>3000</td>
<td>3218</td>
</tr>
</tbody>
</table>

Figure 4. SUPREM-III (solid) and measured (dashed) impurity distributions through the p-channel device source/drain region.
and is thought to be due primarily to bevel rounding. However, even neglecting the surface discrepancy, the difference between the implanted dose and the integrated dose for both of the measured profiles is more than can be explained from either measurement error or out-diffusion into the oxide or ambient during processing. Simple box profile calculations of equivalent doses and junction depths suggest that, if $\sim 80\%$ of the implanted arsenic were present and activated, the sheet resistivities should be on the order of 25 $\Omega/\square$. Table 4 shows the measured and simulated sheet resistances for the source/drain diffusions.
Figure 8. Comparison of SUPREM-III (solid) and measured (dashed) n-channel conductivity vs. gate voltage for three different $V_{DD}$ biases.

The differences in active dopant levels are reflected in the sheet resistance values shown. This indicates that the difference is almost certainly due to incomplete activation of the implanted dose, even with the thermal processing subsequent to implantation. SUPREM-III assumes that annealing and dopant activation is instantaneous.

The structure and impurity profiles for the cross-section through the field region are shown in Figure 6. Table 2 shows both the measured and simulated threshold voltage for a polysilicon gate device at this cross-section. Again the
results are in close agreement.

5. SUPREM-III Work In Progress

As stated in Section 3, while the state of process modeling has shown significant advancement over the past several years, much work remains to be done. Currently, efforts are underway to improve SUPREM-III’s capabilities in the following areas:

1. Diffusion, particularly for high concentrations, OED and ORD;
2. The effects of thermal nitridation on impurity diffusion;
3. Improved mobility/resistivity values;
4. A new silicon epitaxy model based on the work of Reif [15-18];
5. Polysilicon deposition, grain growth, resistivity, and diffusivity.

Additional efforts are underway to improve the speed and accuracy of the numerical solution methods and the program’s output capabilities.

6. Conclusions

The capabilities of the SUPREM-III process simulation program have been presented. A state-of-the-art 2μm CMOS process was described and the reasons for several of the process design decisions were discussed. The results from both measurements and SUPREM-III simulations have been presented and compared. It has been shown that the results of this comparison, whether or not they are in good agreement, provide insight into both the limitations of the program and previously unknown or unexplained aspects of the process. Through this comparison, certain limitations in the program’s coefficients were discovered and their effects on the resulting simulation results were discussed. The necessity of calibrating the programs default parameters was emphasized and it was shown that once calibrated SUPREM-III can be used to successfully predict the results of a given processing facility.

We have shown that with an understanding of both its strengths and limitations, SUPREM-III can be used to predict silicon IC structural, metalurgical, and electrical characteristics resulting from a given processing sequence. Care must be taken so that the parameters and coefficients used by the program are appropriate for the particular processing equipment and techniques being modeled.

When calibrated to the processing equipment and procedures of a particular IC facility, SUPREM-III can be a useful design tool in many modes of operation. As a predictive tool it can allow a device or process designer to develop a new process or modify an existing one with many fewer trial processing runs. As a diagnostic tool it may point out where complex physical processes or subtle interactions are causing undesirable device performance. It can also be used to identify sensitive points in the processing where slight, perhaps unavoidable, variations give significant variations in device performance.
References


